

A functional genomics approach to understand host genetic regulation of COVID-19 severity

Yunus Kuijpers^{1,2,†}, Xiaojing Chu^{1,2,3,†}, Martin Jaeger⁴, Simone J.C.F.M. Moorlag⁴, Valerie A.C.M. Koeken^{1,2,4}, Vera P. Mourits⁴, L. Charlotte J. de Bree⁴, Quirijn de Mast⁴, Frank L. van de Veerdonk⁴, Leo A.B. Joosten^{4,5}, Yang Li^{1,2,4}, Mihai G. Netea^{4,6,‡*}, Cheng-Jian Xu^{1,2,4},
‡*

¹Centre for Individualised Infection Medicine, CiiM, a joint venture between the Hannover Medical School and the Helmholtz Centre for Infection Research, Hannover, Germany

²TWINCORE, Centre for Experimental and Clinical Infection Research, a joint venture between the Hannover Medical School and the Helmholtz Centre for Infection Research, Hannover, Germany

³Department of Genetics, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands

⁴Department of Internal Medicine and Radboud Institute for Center for Infectious Diseases, Radboud University Medical Center, Nijmegen, the Netherlands

⁵Núcleo de Pesquisa da Faculdade da Polícia Militar (FPM) do Estado de Goiás, Goiânia, Goiás, Brasil

⁶Department for Genomics & Immunoregulation, Life and Medical Sciences Institute (LIMES), University of Bonn, Bonn, Germany

[†]shared first authors: these authors contribute equally.

[‡]shared senior authors: these authors contribute equally.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

*Corresponding authors:

Mihai.Netea@radboudumc.nl; Xu.Chengjian@mh-hannover.de

One Sentence summary

In this study, we explore the physiological significance of the COVID-19 severity associated genetic variants using detailed clinical, immunological and multi-omics data from large cohorts. Our findings allow a physiological understanding of genetic susceptibility to severe COVID-19, and indicate pathways that could be targeted for prevention and therapy.

Abstract

A recent genome-wide association study of European ancestry has identified 3p21.31 and 9q34.2 (ABO gene) to be significantly associated with COVID-19 respiratory failure (*1*). Here, we employed the detailed clinical, immunological and multi-omics data of the Human Functional Genomics Projects (HFGP) to explore the physiological significance of the host genetic variants that influence susceptibility to severe COVID-19. A functional genomics investigation based on functional characterization of individuals with high genetic risk for severe COVID-19 susceptibility identified several major patterns: i. a large impact of genetically determined innate immune responses in COVID-19, with increased susceptibility for severe disease in individuals with defective monocyte-derived cytokine production; ii. genetic susceptibility related to ABO blood groups is probably mediated through the von Willebrand factor (VWF) and endothelial dysfunction; and iii. the increased susceptibility for severe COVID-19 in men is at least partially mediated by chromosome X-mediated genetic variation. These insights allow a physiological understanding of genetic susceptibility to severe COVID-19, and indicate pathways that could be targeted for prevention and therapy.

Introduction

The novel coronavirus disease 2019 (COVID-19), caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (2)(3), firstly emerged in late December 2019 and has been spreading worldwide very quickly. The COVID-19 pandemic creates a severe disruption to the healthcare system and endangers the economy. As of the 21st of October, the World Health Organization has reported more than 40 million confirmed cases and over 1 million confirmed deaths (4). While much has been learned about the pathophysiology of the disease, treatment proven to be effective is restricted to dexamethasone (5), and there is no effective vaccine for COVID-19 yet. Therefore, there is an urgent need to better understand the exact host-pathogen interactions leading to increased severity and mortality, in order to design additional prophylactic and therapeutic strategies in future (6)(7).

The severity of SARS-CoV-2 infection is highly variable, and ranges from asymptomatic to mild disease, and even to severe Acute Respiratory Distress Syndrome with a fatal outcome (8). However, the causes for this broad variability in disease outcome between individuals are largely unknown. A recent study indicates that human host factors rather than viral genetic variation affect COVID-19 severity outcome (9). Additionally, clinical and epidemiological data have shown that old age, male sex, and chronic comorbidity are associated with higher mortality (10)(11). A recent genome-wide association study in individuals with genetic European ancestry has identified several chemokine receptor genes, including *CCR9*, *CXCR6* and *XCR1* and the locus controlling the ABO blood type to be associated with severe symptoms of COVID-19 (1). Nevertheless, little is known about the mechanisms through which these genetic variants influence COVID-19 severity. For example, several competing hypotheses may be envisaged for the involvement of immune genes in susceptibility to severe COVID-19: on the one hand, it may be hypothesized that genetic risk for severe COVID-19 is associated

with defective innate immune responses that would allow viral multiplication with high viral loads; on the other hand, the opposite hypothesis may also be true, with an exaggerated genetically-mediated cytokine production being responsible for the late phase hyperinflammation and poor outcome. A purely genetic study cannot respond to this crucial question, that would have important consequences for the approach to prophylaxis and therapy.

By making use of resources from the Human Functional Genomics Project (HFGP) (12, 13), we assessed the impact of COVID-19 associated genetic polymorphisms on variability of immune responses at the population level. This study will help us to understand how genetic variability is related to disease susceptibility through the regulation of immune responses and endothelial function.

Results

This study was conducted in two Dutch cohorts of healthy volunteers from HFGP: 451 participants from the 500 Functional Genomics (500FG) cohort (12, 13), and 313 volunteers from the 300 Bacillus Calmette–Guérin (300BCG) cohort (14). The basic characteristics of study populations are shown in Table S1.

COVID-19 loci are enriched for expression in immune organs, chemokine signaling pathways and enhancer region

To explore the functional impact of the identified COVID-19 loci, we firstly investigated if the identified independent genetic loci ($<1 \times 10^{-5}$) associated with severe COVID-19 are associated with any phenotypes available at the GWAS catalog (<https://www.ebi.ac.uk/gwas/>). We found that many of these loci are associated with immune traits such as blood protein, LDL and VLDL concentrations (Table S2). We next performed functional annotation of significant loci and

gene-mapping using FUMA(15). The SNP2GENE function identified 32 independent SNPs located in 26 different loci which reached suggestive significance in this study ($p\text{-value} < 1 \times 10^{-5}$, Table S3). Using multiple independent expression quantitative trait loci (eQTL) datasets, FUMA mapped 115 genes to these 26 genomic risk loci. Using RNA-seq data of 30 tissues from GTEx database (v8), we found significant enrichment of candidate genes in expression in immune organs such as spleen and blood (Fig. 1A), suggesting that they are important tissues contributing to the pathophysiology of COVID-19 (16)(17). Moreover, we observed the enrichment of candidate genes to be mainly expressed in small intestine and lung (Fig. 1A), suggesting that COVID-19 represents a multisystem illness with involvement of different organs, consistent with the respiratory and intestinal symptoms of the disease (18).

Pathway analysis using these 115 genes showed a strong enrichment in chemokine binding and chemokine receptors binding (Fig. 1B), which is in line with the fact that chemokines can recruit immune cells to the site of infection and are critical for the function of the immune response (19). In addition, chemokines have been reported as the most significantly elevated biomarkers in patients with severe COVID-19 on the intensive care unit (17).

Considering that all SNPs in LD with the 32 independent loci ($p\text{-value} < 1 \times 10^{-5}$) identified by the COVID-19 GWAS, were significantly enriched in the non-coding intronic region ($p\text{ value} = 0.036$, Fisher's exact test) (Table S4), we next examined whether the COVID-19 associated variants are enriched in regulatory DNA elements. We interrogated all significant SNPs ($p\text{-value} < 1 \times 10^{-5}$ and stricter thresholds) with histone marks and chromatin states of 24 blood cell-types in the Roadmap Epigenome Project (20). We found that the COVID-19 genetic loci were strongly enriched for enhancer markers and weakly enriched in promoter marker (Table S5). The strong enrichment of COVID-19 loci in enhancer marks indicates that the associated

genetic variants are likely to be involved in the regulation of immunologically related functions. This finding also suggests that epigenetic mechanisms may play an important role in the pathogenesis of COVID-19 infection.

3p21.31 loci are associated with lower production of monocyte-derived cytokines

Severe COVID-19 is characterized by complex immune dysregulation, combining immune defective features with hyperinflammatory innate immune traits (21)(22). However, these analyses in patients could be done only late during the disease, and whether genetic risk for severe COVID-19 is characterized by low or high innate immune responses in a non-infected person is not known. We therefore used the cytokine QTL data from the 500FG cohort (13) of the HFGP to test whether SNPs in 3p21.31 influence cytokine production upon stimulation. We checked all SNPs located within a 50 kilobase window of top variant rs11385942, and showed all nominal significant associations (p -value < 0.05 , Fig. 2A). Interestingly, we observed that the risk alleles for a severe course of COVID-19 are consistently associated with lower production of monocyte-dependent cytokines (IL-6, IL-1 β and TNF- α) upon various *in-vitro* stimulations (Fig. 2A). Of note, COVID-19 risk alleles also correspond to lower monocyte-derived cytokine production after influenza stimulation, a viral stimulus (Fig. 2A and B). It is thus tempting to speculate that the people who carry risk alleles may not respond properly to an initial virus infection, leading to high viral loads, subsequent systemic inflammation and poor outcome. Next, we tested whether the COVID-19 risk SNPs are associated with the levels of circulating cytokines in blood. Using the same cohort, we found that IL-18 and IL-18BP show a suggestive positive association with genetic risk of COVID-19 (Fig. S1).

Von Willebrand Factor (VWF) and lymphocytes are strongly colocalized with ABO loci

It is known that ABO blood group influences the plasma levels of von Willebrand factor (VWF)(23) and elevated VWF levels are associated with severe COVID-19 (24). We therefore tested the association of ABO locus with VWF circulating concentrations from the individuals in the 500FG cohort. Of note, we found the risk allele rs687621-G is significantly associated with elevated levels of VWF ($p\text{-value} = 9.58 \times 10^{-20}$) (Fig. 3A and B). Recent studies have reported that the VWF level is highly related to COVID-19 severity (25, 26). As VWF level in plasma is an indicator of inflammation, endothelial activation and damage (27), our results suggest that the association of VWF and COVID-19 severity is very likely mediated through genetic regulation.

We next tested if this specific locus is associated with immune functions. Interestingly, we observed consistent negative correlation of VWF and T-cell derived cytokine production in response to various *ex-vivo* stimulations (Fig. 3C and Fig. S2). In addition, the ABO locus led by the variant rs687621 also showed statistically significant co-localization with several immune-mediated traits, including cell counts of lymphocytes (Coloc analysis H4: 0.98), monocytes (Coloc analysis H4: 1), neutrophils (Coloc analysis H4: 0.8) and whole blood cells (Coloc analysis H4: 1) (Fig. 3D).

Men have a higher genetic severity risk of COVID-19 than women

Polygenic risk scores (PRS) combine multiple risk alleles and capture an individual's load of common genetic variants associated with a disease phenotype (28). Using the summary statistics provided in the GWAS study (1), we calculated the PRS for the samples from 500FG and 300BCG cohorts. Since higher mortality of COVID-19 has been reported to be associated with male sex and BMI (10, 11), we investigated whether these host factors are associated with the PRS, a predictive measure of risk for development of severe COVID-19.

We firstly assessed if males have a higher genetic risk compared to females in 500FG. Hereby, we defined people with top 10% PRS as a high-risk group and those with bottom 10% PRS as a low-risk group. As shown in Fig. 4A, male tend to have higher severe COVID-19 risk than female (odd ratio: 1.47, 95% CI: 0.98-2.22, p-value = 0.045 (Fisher's exact test)) (Table S6). We next used different percentile cut-offs (15%, 20%, 25% and 30%) to re-define low and high-risk groups. Interestingly, we observed a consistent pattern that males have higher genetic risk (PRS) than females at different percentile cut-offs. These results can be replicated in a similar, but independent, cohort (300BCG, Fig. 4B). Meanwhile, the genetic risk difference between male and female can be attenuated when a loose cut-off has been defined. The meta-analysis of two cohorts showed a significant p-value at various percentile cut-offs (10%, 15%, 20%, 25%) and marginal significant p-value of 0.051 at the percentile cut-off of 30%. Furthermore, this result persisted when PRS was computed using summary statistics from the GWAS model after age and sex correction, reported in the original GWAS study (Table S7). When computing PRS while excluding variants from the X and Y chromosomes, the enrichment of men in the higher PRS group was less pronounced at most thresholds (Fig. 4C and D). This suggested that higher genetic severity risk at least partially originates from the SNPs in the sex chromosomes.

As obesity or overweight has been reported as a risk factor for serious illness or death from COVID-19, we tested if the PRS is associated with BMI (Fig. S3). We did not observe any significant correlation between PRS and BMI.

Discussion

Understanding the pathophysiology of COVID-19 is urgently needed for designing novel preventive and therapeutic approaches against the disease. One important tool for identifying the most important mechanisms mediating a disease is genomics: genetic variants that influence susceptibility or severity to a disease are usually located in genetic loci that impact important mechanisms for that particular disease. Using the information of a recently published GWAS assessing the severity of COVID-19 (*1*), and the rich datasets available in the HFGP, we interrogated the mechanisms through which genetic variants associated with severe COVID-19 exert their effects.

Among the genetic loci associated with severe COVID-19, the 3p21.31 gene cluster has been well replicated by independent studies from the COVID-19 Host Genetics Initiative (<https://www.covid19hg.org>), and it was reported to be inherited by Neanderthals (*29*). This locus is currently regarded as a marker of COVID-19 severity, but crucial information is missing: are the risk alleles in this locus (that encode several cytokines and chemokines) associated with a lower or higher cytokine production. The answer to this question is crucial for understanding COVID-19: a genetic risk associated with low cytokine production would imply that severe COVID-19 is the consequence of a relative immunodeficiency, while a high cytokine production associated with genetic risk would mean that severe COVID-19 is a genetic hyperinflammatory disease. In our study, the 3p21.31 genetic polymorphisms associated with a high risk of severe COVID-19 were associated with lower production of monocyte-derived cytokines, especially to viral (influenza) stimuli. This important discovery has significant prophylactic and therapeutic consequences. On the one hand, it implies that improvement of innate immune responses in healthy individuals would decrease the probability that they undergo a severe form of COVID-19: this supports the rationale of clinical trials that improve innate immune responses through induction of trained immunity, e.g. by vaccination

with BCG (6). On the other hand, this also implies that the dysregulated immune responses that have been described at late time points in patients with severe COVID-19 (30)(31) are likely the consequence of accelerated viral multiplication due to defective innate immune responses, and subsequent systemic inflammation due to high viral loads.

Several studies have shown that ABO blood types are associated with COVID-19 severity (32)(33) and susceptibility (34)(35)(36). It is still not well-known how ABO gene regulates COVID-19 susceptibility. As ABO blood group are also expressed on endothelial cells and platelets, it has been speculated that this effect may manifest itself via elevated plasma VWF (37). Our results provide evidence supporting this hypothesis, by showing that the risk alleles in the ABO locus are associated with high concentrations of VWF. Moreover, interesting associations have been found between polymorphisms in this locus and the number of various immune cell populations, especially lymphocytes, since lymphopenia is also consistently associated with severe COVID-19 (38). This suggests that genetic factors are relevant to the host thrombo-inflammatory response. However, a note of caution should be mentioned, as the association between the genetics of ABO group with severity in COVID-19 Host Genetics Initiative data did not reach a genome-wide level of significance (p value $<5 \times 10^{-8}$) (Table S8) (as of 21st of October 2020, and thus the association might be population specific).

Another important observation is that an important component of the impact of genetic polymorphisms on the severity of COVID-19 is mediated through sex chromosomes, most likely chromosome X which is known to encode many genes related to the immune system. Indeed, men in both 500FG and 300BCG cohorts had a higher genetic risk than women, and this difference was largely lost when sex chromosomes were excluded from the analysis. These data strongly argue that at least part of the well-known increase of COVID-19 severity in men

is genetically determined. The recent description by our group of rare mutations in the RNA receptor TLR7 located on chromosome X as a cause of very severe COVID-19 in young men supports this hypothesis (39).

While our study sheds further light on how COVID-19 genetic risk affects the human immune system, there are several limitations of this study: firstly, due to different sets of stimuli used in measuring cytokine production to stimulations in the two healthy cohorts, we are not able to replicate all our findings of genetic associations with cytokine responses from the 500FG cohort in the 300BCG cohort. Secondly, young adults (< 30 years) are overrepresented in both healthy cohorts (500FG and 300BCG), which may lead to a biased conclusion which cannot be generalized to the whole population, especially since the severe COVID-19 cases often occur in the elderly population. Thirdly, 500FG and 300BCG cohorts are designed to understand the genetic regulation of immune function in healthy individuals. Therefore, a COVID-19 patients' cohort will be needed for better characterization of disease mechanism, which will be our future research goal.

Collectively, our data demonstrate that genetic variability explains an important component of the increased susceptibility to severe COVID-19. The genetic risk for severe COVID-19 is associated with defective innate immune responses (low cytokine production), dysregulated endothelial function, and is strongly influenced by polymorphisms in sex chromosomes. These findings may contribute to the development of novel treatment and prevention strategies for severe COVID-19.

Materials and Methods

Study cohort

The cohorts involved in this study are from the Human Functional Genomics Project (HFGP)(40). 500FG consists of 451 healthy individuals of European ancestry with genotype measurement. Within this cohort, immune cell counts, cytokine production upon stimulations, platelets, globulins, and gut microbiome were measured (for detailed information see (12, 13, 41, 42)). 300BCG consists of 313 healthy Europeans that participated in a BCG vaccination study (14)(43). Within this cohort, blood was collected before vaccination and cytokine production was measured upon *ex-vivo* stimulation of PBMCs with microbial stimuli.

Genotype quality control and imputation

Genotyping on samples from 500FG and 300BCG was performed using Illumina humanOmniExpress Exome-8 v1.0 SNP chip Calling by OptiCall 0.7.0(44) with default settings. All individuals of non-European ancestry, ambiguous sex, call rate ≤ 0.99 , excess of autosomal heterozygosity ($F < \text{mean} - 3SD$), cryptic relatedness ($\pi > 0.185$) were removed. SNPs with low genotyping rate ($< 95\%$), with low minor allele frequency (< 0.001), deviation from Hardy-Weinberg equilibrium ($p < 10^{-4}$) were excluded. The detailed QC steps have been published in reference(13). Genotype data of 500FG and 300BCG were imputed respectively. The imputation was performed on the Michigan imputation server(45). The cohorts were phased using Eagle v2.4 with the European population of HRC 1.1 2016 hg 2019 reference panel. After imputation, variants with a MAF < 0.01 , an imputation quality score $R^2 < 0.5$, or a Hardy-Weinberg-Equilibrium $P < 10^{-12}$ were excluded. All quality control steps were performed using Plink v1.9. After imputation and quality control, 451 individuals from 500FG and 313 individuals from 300BCG were available for downstream analyses.

Immune parameter quantitative trait locus (QTL) profiles

We acquired summary statistics of cytokine QTLs (13), cell proportion QTLs(41) and circulating metabolite QTLs(46) from our previous studies performed with 500FG. We performed QTL mapping for circulating mediators and platelet traits in 500FG using an R package *MatrIXeQTL*(47). The measurement of circulating mediators including IL-18BP, resistin, leptin, adiponectin, alpha-1 antitrypsin (AAT), and IL-18 have been described previously(12). Platelet traits(48) include Thrombin-Antithrombin Complex (TAT), Beta-thromboglobulin total, beta-thromboglobulin, fibrinogen binding, collagen-related peptide (CRP) P-selectin, CRP fibrinogen, ADP P-selectin, ADP fibrinogen, P-selectin, platelet-monocyte complex, total platelet count, and von Willebrand factor (VWF). The circulating mediator levels and platelet traits were log2 transformed. A linear model was applied to the platelet data and genetic data by taking age and sex as covariates. We considered $p\text{-value} < 5 \times 10^{-8}$ to be genome-wide significant.

Colocalization analysis.

We performed co-localization analysis (49) to look at the overlapping profile between molecular QTLs, COVID-19 GWAS, and other GWAS profiles using the R package ‘coloc’.

PRS calculation

Polygenic risk scores (PRS) were calculated by first intersecting the variants from the COVID-19 summary statistics(1) with the variants present in our samples. Clumping was done starting at the most significant variant. All variants within a 250kb window around that variant were excluded if they were in greater LD than 0.1 before continuing to the most significant variant outside of the previous window. For each sample specifically, we then multiplied the dosage of the effect allele with its effect size while substituting missing genotype data with the average

dosage of that variant in the entire sample set. These values were then summed to form the PRS for each specific sample. As the GWAS summary statistics for creating PRS from Eillinghaus *et.al* (1) did not correct age and sex, we also performed a sensitivity analysis with the PRS created from the GWAS model corrected for age and sex.

PRS based correlations

Linear models were constructed using the computed PRS and various phenotype data available for each cohort. Samples within the top/bottom 10% PRS were classified as high/low-risk, respectively. Using the PRS of the samples in these risk groups, we performed a Student T-test to test for significant correlation between gender and PRS. Furthermore, we tested for enrichment of any specific gender in these risk groups using a Fisher's exact test.

Functional analysis of genomic loci

We used the FUMA pipeline in order to identify genes linked to COVID-19 with severe respiratory failure. FUMA identified significant independent SNPs as variants with $P < 1 \times 10^{-5}$ that were independent from each other using an LD threshold of $r^2 < 0.6$. Within these independent significant SNPs variants lead SNPs are identified as the most significant variants that are independent using an LD threshold of $r^2 < 0.1$. We mapped Genes to these SNPs based on their genomic position allowing for a maximum distance of 10kb. In addition to this, genes were also mapped based on eQTL effects. Genes were selected based on significant SNP-gene pairs at $FDR < 0.05$ using cis- and trans-eQTLs from eQTLGen.

As part of the FUMA pipeline we used these mapped genes in order to generate gene expression heatmaps using GTEx v8 (54 tissue types and 30 general tissue types). Gene expression values with a pseudocount of 1 were normalized across tissue types using winsorization at 50 and log2

transformed. Using the hypergeometric test, we tested for significant enrichment of our input genes in DEG sets for the different tissue types using a Bonferroni corrected P value ≤ 0.05 . Finally, we tested for overrepresentation of our input genes in predetermined gene-sets using hypergeometric tests. Gene-sets were obtained from MsigDB, WikiPathways, and GWAS-catalog reported gene-sets. We used Benjamini-Hochberg FDR correction for each of the categories within these gene-set sources separately using a threshold of 0.05 for our adjusted P value.

Roadmap epigenetic state enrichment

Based on the Roadmap 15-core epigenetic state database(20), we used data obtained from 23 blood samples spanning 127 epigenomes to map the QTLs in the summary statistics to their respective epigenetic states. Epigenetic state information was available for bins of 200bp. we aggregated this information into 4 categories; active enhancer states (Enh, EnhG), active promotor states (TssA, TssAFlnk), all enhancer states (Enh, EnhG, EnhBiv), and all promotor states (TssA, TssAFlnk, TssBiv). We tested for enrichment using a Fisher's exact test based on the number of unique 200bp bins variants mapped to. This was done after filtering the QTL's down based on their p value using different thresholds (1×10^{-5} , 1×10^{-6} and 1×10^{-7}). Enrichment P values were obtained after FDR correction.

Visualization.

R package ggplot2 was used to perform bar charts, box plots and scatter plots. We applied an online tool Locus zoom to present genes within candidate loci. We used R package pheatmap to generate heat maps.

Acknowledgments:

We thank all volunteers in the 500FG and 300BCG cohorts of the Human Functional Genomics Project (HFGP) for their participation.

Funding

MGN was supported by an ERC Advanced Grant (833247) and a Spinoza Grant of the Netherlands Organization for Scientific Research. YL was supported by an ERC starting Grant (948207) and a Radboud University Medical Centre Hypatia Grant (2018). CJX was supported by Helmholtz Initiative and Networking Fund (1800167).

Author contributions

M.G.N., C.J.X. and Y.L. designed the study, Y.K. and X.C. performed statistical analyses supervised by Y.L. and C.J.X., M.J., S.M., V.K., V.M., C.d.B., Q.d.M. and L.A.B.J. established the cohorts, helped with the data analysis and interpretation of results. C.J.X., M.G.N., Y.L., Y.K. and X.C. wrote the manuscript with input from all authors.

Competing interests

The authors declare no competing interests.

Data and materials availability:

500FG data used in this project have been archived in the BBMRI-NL data infrastructure (<https://hfgp.bbmri.nl/>). This allows flexible data querying and download, including sufficiently rich metadata and interfaces for machine processing and using FAIR principles to optimize Findability, Accessibility, Interoperability and Reusability.

Ethics statement

The 300BCG (NL58553.091.16) and 500FG (NL42561.091.12) studies were approved by the Arnhem-Nijmegen Medical Ethical Committee. Experiments were conducted according to the principles expressed in the Declaration of Helsinki. Samples of venous blood were drawn after informed consent was obtained.

References

1. D. Ellinghaus, F. Degenhardt, L. Bujanda, M. Buti, A. Albillos, P. Invernizzi, J. Fernández, D. Prati, G. Baselli, R. Asselta, M. M. Grimsrud, C. Milani, F. Aziz, J. Kässens, S. May, M. Wendorff, L. Wienbrandt, F. Uellendahl-Werth, T. Zheng, X. Yi, R. de Pablo, A. G. Chercoles, A. Palom, A.-E. Garcia-Fernandez, F. Rodriguez-Frias, A. Zanella, A. Bandera, A. Protti, A. Aghemo, A. Lleo, A. Biondi, A. Caballero-Garralda, A. Gori, A. Tanck, A. Carreras Nolla, A. Latiano, A. L. Fracanzani, A. Peschuck, A. Julià, A. Pesenti, A. Voza, D. Jiménez, B. Mateos, B. Nafria Jimenez, C. Quereda, C. Paccapelo, C. Gassner, C. Angelini, C. Cea, A. Solier, D. Pestaña, E. Muñiz-Diaz, E. Sandoval, E. M. Paraboschi, E. Navas, F. García Sánchez, F. Ceriotti, F. Martinelli-Boneschi, F. Peyvandi, F. Blasi, L. Téllez, A. Blanco-Grau, G. Hemmrich-Stanisak, G. Grasselli, G. Costantino, G. Cardamone, G. Foti, S. Aneli, H. Kurihara, H. ElAbd, I. My, I. Galván-Femenia, J. Martín, J. Erdmann, J. Ferrusquía-Acosta, K. Garcia-Etxebarria, L. Izquierdo-Sanchez, L. R. Bettini, L. Sumoy, L. Terranova, L. Moreira, L. Santoro, L. Scudeller, F. Mesonero, L. Roade, M. C. Rühlemann, M. Schaefer, M. Carrabba, M. Riveiro-Barciela, M. E. Figuera Basso, M. G. Valsecchi, M. Hernandez-Tejero, M. Acosta-Herrera, M. D'Angiò, M. Baldini, M. Cazzaniga, M. Schulzky, M. Cecconi, M. Wittig, M. Ciccarelli, M. Rodríguez-Gandía, M. Boccione, M. Miozzo, N. Montano, N. Braun, N. Sacchi, N. Martínez, O. Özer, O. Palmieri, P. Faverio, P. Preatoni, P. Bonfanti, P. Omodei, P. Tentorio, P. Castro, P. M. Rodrigues, A. Blandino Ortiz, R. de Cid, R. Ferrer, R. Gualtierotti, R. Nieto, S. Goerg, S. Badalamenti, S. Marsal, G. Matullo, S. Pelusi, S. Juzenas, S. Aliberti, V. Monzani, V. Moreno, T. Wesse, T. L. Lenz, T. Pumarola, V. Rimoldi, S. Bosari, W. Albrecht, W. Peter, M. Romero-Gómez, M. D'Amato, S. Duga, J. M. Banales, J. R. Hov, T. Folseraas, L. Valenti, A. Franke, T. H. Karlsen, Genomewide Association Study of Severe Covid-19 with Respiratory Failure, *N Engl J Med* (2020), doi:10.1056/NEJMoa2020283.
2. F. Wu, S. Zhao, B. Yu, Y.-M. Chen, W. Wang, Z.-G. Song, Y. Hu, Z.-W. Tao, J.-H. Tian, Y.-Y. Pei, M.-L. Yuan, Y.-L. Zhang, F.-H. Dai, Y. Liu, Q.-M. Wang, J.-J. Zheng, L. Xu, E. C. Holmes, Y.-Z. Zhang, A new coronavirus associated with human respiratory disease in China, *Nature* **579**, 265–269 (2020).
3. P. Zhou, X.-L. Yang, X.-G. Wang, B. Hu, L. Zhang, W. Zhang, H.-R. Si, Y. Zhu, B. Li, C.-L. Huang, H.-D. Chen, J. Chen, Y. Luo, H. Guo, R.-D. Jiang, M.-Q. Liu, Y. Chen, X.-R. Shen, X. Wang, X.-S. Zheng, K. Zhao, Q.-J. Chen, F. Deng, L.-L. Liu, B. Yan, F.-X. Zhan, Y.-Y. Wang, G.-F. Xiao, Z.-L. Shi, A pneumonia outbreak associated with a new coronavirus of probable bat origin, *Nature* **579**, 270–273 (2020).
4. World Health Organization, WHO Coronavirus Disease (COVID-19) Dashboard, <https://covid19.who.int/> (2020) (available at <https://covid19.who.int/>).
5. RECOVERY Collaborative Group, P. Horby, W. S. Lim, J. R. Emberson, M. Mafham, J. L. Bell, L. Linsell, N. Staplin, C. Brightling, A. Ustianowski, E. Elmahi, B. Prudon, C. Green, T. Felton, D. Chadwick, K. Rege, C. Fegan, L. C. Chappell, S. N. Faust, T. Jaki, K. Jeffery, A. Montgomery, K. Rowan, E. Juszczak, J. K. Baillie, R. Haynes, M. J. Landray, Dexamethasone in Hospitalized Patients with Covid-19 - Preliminary Report, *N Engl J Med* (2020), doi:10.1056/NEJMoa2021436.
6. M. G. Netea, E. J. Giamarellos-Bourboulis, J. Domínguez-Andrés, N. Curtis, R. van Crevel, F. L. van de Veerdonk, M. Bonten, Trained Immunity: a Tool for Reducing Susceptibility to and the Severity of SARS-CoV-2 Infection, *Cell* **181**, 969–977 (2020).

7. M. Z. Tay, C. M. Poh, L. Rénia, P. A. MacAry, L. F. P. Ng, The trinity of COVID-19: immunity, inflammation and intervention, *Nature Reviews Immunology* **20**, 363–374 (2020).
8. R. Verity, L. C. Okell, I. Dorigatti, P. Winskill, C. Whittaker, N. Imai, G. Cuomo-Dannenburg, H. Thompson, P. G. T. Walker, H. Fu, A. Dighe, J. T. Griffin, M. Baguelin, S. Bhatia, A. Boonyasiri, A. Cori, Z. Cucunubá, R. FitzJohn, K. Gaythorpe, W. Green, A. Hamlet, W. Hinsley, D. Laydon, G. Nedjati-Gilani, S. Riley, S. van Elsland, E. Volz, H. Wang, Y. Wang, X. Xi, C. A. Donnelly, A. C. Ghani, N. M. Ferguson, Estimates of the severity of coronavirus disease 2019: a model-based analysis, *The Lancet Infectious Diseases* **20**, 669–677 (2020).
9. X. Zhang, Y. Tan, Y. Ling, G. Lu, F. Liu, Z. Yi, X. Jia, M. Wu, B. Shi, S. Xu, J. Chen, W. Wang, B. Chen, L. Jiang, S. Yu, J. Lu, J. Wang, M. Xu, Z. Yuan, Q. Zhang, X. Zhang, G. Zhao, S. Wang, S. Chen, H. Lu, Viral and host factors related to the clinical outcome of COVID-19, *Nature* **583**, 437–440 (2020).
10. G. Chen, D. Wu, W. Guo, Y. Cao, D. Huang, H. Wang, T. Wang, X. Zhang, H. Chen, H. Yu, X. Zhang, M. Zhang, S. Wu, J. Song, T. Chen, M. Han, S. Li, X. Luo, J. Zhao, Q. Ning, Clinical and immunological features of severe and moderate coronavirus disease 2019, *Journal of Clinical Investigation* **130**, 2620–2629 (2020).
11. A. B. Docherty, E. M. Harrison, C. A. Green, H. E. Hardwick, R. Pius, L. Norman, K. A. Holden, J. M. Read, F. Dondelinger, G. Carson, L. Merson, J. Lee, D. Plotkin, L. Sigfrid, S. Halpin, C. Jackson, C. Gamble, P. W. Horby, J. S. Nguyen-Van-Tam, A. Ho, C. D. Russell, J. Dunning, P. J. Openshaw, J. K. Baillie, M. G. Semple, Features of 20 133 UK patients in hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: prospective observational cohort study, *BMJ* **369**, m1985 (2020).
12. R. ter Horst, M. Jaeger, S. P. Smekens, M. Oosting, M. A. Swertz, Y. Li, V. Kumar, D. A. Diavatopoulos, A. F. M. Jansen, H. Lemmers, H. Toenhake-Dijkstra, A. E. van Herwaarden, M. Janssen, R. G. van der Molen, I. Joosten, F. C. G. J. Sweep, J. W. Smit, R. T. Netea-Maier, M. M. J. F. Koenders, R. J. Xavier, J. W. M. van der Meer, C. A. Dinarello, N. Pavelka, C. Wijmenga, R. A. Notebaart, L. A. B. Joosten, M. G. Netea, Host and Environmental Factors Influencing Individual Human Cytokine Responses, *Cell* **167**, 1111-1124.e13 (2016).
13. Y. Li, M. Oosting, S. P. Smekens, M. Jaeger, R. Aguirre-Gamboa, K. T. T. Le, P. Deelen, I. Ricaño-Ponce, T. Schoffelen, A. F. M. Jansen, M. A. Swertz, S. Withoff, E. van de Vosse, M. van Deuren, F. van de Veerdonk, A. Zhernakova, J. W. M. van der Meer, R. J. Xavier, L. Franke, L. A. B. Joosten, C. Wijmenga, V. Kumar, M. G. Netea, A Functional Genomics Approach to Understand Variation in Cytokine Production in Humans, *Cell* **167**, 1099-1110.e14 (2016).
14. V. A. C. M. Koeken, L. C. J. de Bree, V. P. Mourits, S. J. C. F. M. Moorlag, J. Walk, B. Cirovic, R. J. W. Arts, M. Jaeger, H. Dijkstra, H. Lemmers, L. A. B. Joosten, C. S. Benn, R. van Crevel, M. Netea, BCG vaccination in humans inhibits systemic inflammation in a sex-dependent manner, *J Clin Invest* (2020), doi:10.1172/JCI133935.
15. K. Watanabe, E. Taskesen, A. van Bochoven, D. Posthuma, Functional mapping and annotation of genetic associations with FUMA, *Nature Communications* **8**, 1826 (2017).

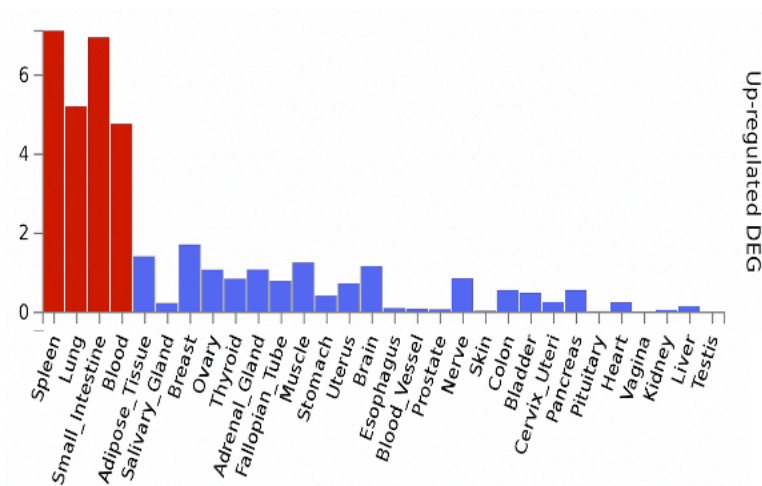
16. X. Xu, X. Chang, H. Pan, H. Su, B. Huang, M. Yang, D. Luo, M. Weng, L. Ma, X. Nie, [Pathological changes of the spleen in ten patients with coronavirus disease 2019(COVID-19) by postmortem needle autopsy], *Zhonghua bing li xue za zhi = Chinese journal of pathology* **49**, 576–582 (2020).
17. F. L. van de Veerdonk, N. A. F. Janssen, I. Grondman, A. H. de Nooijer, V. A. C. M. Koeken, V. Matzaraki, C. K. Boahen, V. Kumar, M. Kox, H. J. P. M. Koenen, R. L. Smeets, I. Joosten, R. J. M. Brüuggemann, I. J. E. Kooijzer, H. G. van der Hoeven, J. A. Schouten, T. Frenzel, M. Reijers, W. Hoefsloot, A. S. M. Dofferhoff, A. P. M. Kerckhoffs, M. J. T. Blaauw, K. Veerman, C. Maas, A. H. Schoneveld, I. E. Hoefer, L. P. G. Derde, L. Willems, E. Toonen, M. van Deuren, J. W. M. van der Meer, R. van Crevel, E. J. Giamarellos-Bourboulis, L. A. B. Joosten, M. M. van den Heuvel, J. Hoogerwerf, Q. de Mast, P. Pickkers, M. G. Netea, A systems approach to inflammation identifies therapeutic targets in SARS-CoV-2 infection, *medRxiv* , 2020.05.23.20110916 (2020).
18. C. Robba, D. Battaglini, P. Pelosi, P. R. M. Rocco, Multiple organ dysfunction in SARS-CoV-2: MODS-CoV-2, *Expert Review of Respiratory Medicine* , 1–4 (2020).
19. C. Esche, C. Stellato, L. A. Beck, Chemokines: Key Players in Innate and Adaptive Immunity, *Journal of Investigative Dermatology* **125**, 615–628 (2005).
20. Roadmap Epigenomics Consortium, A. Kundaje, W. Meuleman, J. Ernst, M. Bilenky, A. Yen, A. Heravi-Moussavi, P. Kheradpour, Z. Zhang, J. Wang, M. J. Ziller, V. Amin, J. W. Whitaker, M. D. Schultz, L. D. Ward, A. Sarkar, G. Quon, R. S. Sandstrom, M. L. Eaton, Y.-C. Wu, A. R. Pfenning, X. Wang, M. Claussnitzer, Y. Liu, C. Coarfa, R. A. Harris, N. Shores, C. B. Epstein, E. Gjoneska, D. Leung, W. Xie, R. D. Hawkins, R. Lister, C. Hong, P. Gascard, A. J. Mungall, R. Moore, E. Chuah, A. Tam, T. K. Canfield, R. S. Hansen, R. Kaul, P. J. Sabo, M. S. Bansal, A. Carles, J. R. Dixon, K.-H. Farh, S. Feizi, R. Karlic, A.-R. Kim, A. Kulkarni, D. Li, R. Lowdon, G. Elliott, T. R. Mercer, S. J. Neph, V. Onuchic, P. Polak, N. Rajagopal, P. Ray, R. C. Sallari, K. T. Siebenthall, N. A. Sinnott-Armstrong, M. Stevens, R. E. Thurman, J. Wu, B. Zhang, X. Zhou, A. E. Beaudet, L. A. Boyer, P. L. De Jager, P. J. Farnham, S. J. Fisher, D. Haussler, S. J. M. Jones, W. Li, M. A. Marra, M. T. McManus, S. Sunyaev, J. A. Thomson, T. D. Tlsty, L.-H. Tsai, W. Wang, R. A. Waterland, M. Q. Zhang, L. H. Chadwick, B. E. Bernstein, J. F. Costello, J. R. Ecker, M. Hirst, A. Meissner, A. Milosavljevic, B. Ren, J. A. Stamatoyannopoulos, T. Wang, M. Kellis, Integrative analysis of 111 reference human epigenomes, *Nature* **518**, 317–330 (2015).
21. E. J. Giamarellos-Bourboulis, M. G. Netea, N. Rovina, K. Akinosoglou, A. Antoniadou, N. Antonakos, G. Damoraki, T. Gkavogianni, M.-E. Adami, P. Katsaounou, M. Ntaganou, M. Kyriakopoulou, G. Dimopoulos, I. Koutsodimitropoulos, D. Velissaris, P. Koufargyris, A. Karageorgos, K. Katrini, V. Lekakis, M. Lupse, A. Kotsaki, G. Renieris, D. Theodoulou, V. Panou, E. Koukaki, N. Koulouris, C. Gogos, A. Koutsoukou, Complex Immune Dysregulation in COVID-19 Patients with Severe Respiratory Failure, *Cell Host & Microbe* **27**, 992-1000.e3 (2020).
22. P. Mehta, D. F. McAuley, M. Brown, E. Sanchez, R. S. Tattersall, J. J. Manson, COVID-19: consider cytokine storm syndromes and immunosuppression, *The Lancet* **395**, 1033–1034 (2020).

23. M. Franchini, F. Capra, G. Targher, M. Montagnana, G. Lippi, Relationship between ABO blood group and von Willebrand factor levels: from biology to clinical implications, *Thrombosis Journal* **5**, 14 (2007).
24. G. Goshua, A. B. Pine, M. L. Meizlish, C.-H. Chang, H. Zhang, P. Bahel, A. Baluha, N. Bar, R. D. Bona, A. J. Burns, C. S. Dela Cruz, A. Dumont, S. Halene, J. Hwa, J. Koff, H. Menninger, N. Neparidze, C. Price, J. M. Siner, C. Tormey, H. M. Rinder, H. J. Chun, A. I. Lee, Endotheliopathy in COVID-19-associated coagulopathy: evidence from a single-centre, cross-sectional study, *The Lancet Haematology* **7**, e575–e582 (2020).
25. Aksenova A.Y., Von Willebrand factor and endothelial damage: a possible association with COVID-19, *EcoGen* **18**, 135–138 (2020).
26. A. V. Rapkiewicz, X. Mai, S. E. Carsons, S. Pittaluga, D. E. Kleiner, J. S. Berger, S. Thomas, N. M. Adler, D. M. Charytan, B. Gasmi, J. S. Hochman, H. R. Reynolds, Megakaryocytes and platelet-fibrin thrombi characterize multi-organ thrombosis at autopsy in COVID-19: A case series, *EClinicalMedicine* **24** (2020), doi:10.1016/j.eclinm.2020.100434.
27. C. Kawecki, P. J. Lenting, C. V. Denis, von Willebrand factor and inflammation, *Journal of Thrombosis and Haemostasis* **15**, 1285–1294 (2017).
28. A. V. Khera, M. Chaffin, K. G. Aragam, M. E. Haas, C. Roselli, S. H. Choi, P. Natarajan, E. S. Lander, S. A. Lubitz, P. T. Ellinor, S. Kathiresan, Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations, *Nature Genetics* **50**, 1219–1224 (2018).
29. H. Zeberg, S. Pääbo, The major genetic risk factor for severe COVID-19 is inherited from Neanderthals, *Nature* (2020), doi:10.1038/s41586-020-2818-3.
30. J. Schulte-Schrepping, N. Reusch, D. Paclik, K. Baßler, S. Schlickeiser, B. Zhang, B. Krämer, T. Krammer, S. Brumhard, L. Bonaguro, E. De Domenico, D. Wendisch, M. Grasshoff, T. S. Kapellos, M. Beckstette, T. Pecht, A. Saglam, O. Dietrich, H. E. Mei, A. R. Schulz, C. Conrad, D. Kunkel, E. Vafadarnejad, C.-J. Xu, A. Horne, M. Herbert, A. Drews, C. Thibeault, M. Pfeiffer, S. Hippenstiel, A. Hocke, H. Müller-Redetzky, K.-M. Heim, F. Machleidt, A. Uhrig, L. Bosquillon de Jarcy, L. Jürgens, M. Stegemann, C. R. Glösenkamp, H.-D. Volk, C. Goffinet, M. Landthaler, E. Wyler, P. Georg, M. Schneider, C. Dang-Heine, N. Neuwinger, K. Kappert, R. Tauber, V. Corman, J. Raabe, K. M. Kaiser, M. T. Vinh, G. Rieke, C. Meisel, T. Ulas, M. Becker, R. Geffers, M. Witzernath, C. Drosten, N. Suttorp, C. von Kalle, F. Kurth, K. Händler, J. L. Schultze, A. C. Aschenbrenner, Y. Li, J. Nattermann, B. Sawitzki, A.-E. Saliba, L. E. Sander, A. Angelov, R. Bals, A. Bartholomäus, A. Becker, D. Bezdan, E. Bonifacio, P. Bork, T. Clavel, M. Colome-Tatche, A. Diefenbach, A. Diltthey, N. Fischer, K. Förstner, J.-S. Frick, J. Gagneur, A. Goesmann, T. Hain, M. Hummel, S. Janssen, J. Kalinowski, R. Kallies, B. Kehr, A. Keller, S. Kim-Hellmuth, C. Klein, O. Kohlbacher, J. O. Korb, I. Kurth, M. Landthaler, Y. Li, K. Ludwig, O. Makarewicz, M. Marz, A. McHardy, C. Mertes, M. Nöthen, P. Nürnberg, U. Ohler, S. Ossowski, J. Overmann, S. Peter, K. Pfeffer, A. R. Poetsch, A. Pühler, N. Rajewsky, M. Ralser, O. Rieß, S. Ripke, U. Nunes da Rocha, P. Rosenstiel, A.-E. Saliba, L. E. Sander, B. Sawitzki, P. Schiffer, E.-C. Schulte, J. L. Schultze, A. Sczyrba, O. Stegle, J. Stoye, F. Theis, J. Vehreschild, J. Vogel, M. von Kleist, A. Walker, J. Walter, D. Wiczorek, J. Ziebuhr, Severe COVID-19 Is Marked by a Dysregulated Myeloid Cell Compartment, *Cell* **182**, 1419-1440.e23 (2020).

31. L. Wei, W. Wang, D. Chen, B. Xu, Dysregulation of the immune response affects the outcome of critical COVID-19 patients, *Journal of Medical Virology* **92**, 2768–2776 (2020).
32. R. L. Hoiland, N. A. Fergusson, A. R. Mitra, D. E. G. Griesdale, D. V. Devine, S. Stukas, J. Cooper, S. Thiara, D. Foster, L. Y. C. Chen, A. Y. Y. Lee, E. M. Conway, C. L. Wellington, M. S. Sekhon, The association of ABO blood group with indices of disease severity and multiorgan dysfunction in COVID-19, *Blood Advances* **4**, 4981–4989 (2020).
33. M. Hultström, B. Persson, O. Eriksson, M. Lipcsey, R. Frithiof, B. Nilsson, Blood type A associates with critical COVID-19 and death in a Swedish cohort, *Critical Care* **24**, 496 (2020).
34. M. B. Barnkob, A. Pottegård, H. Støvring, T. M. Haunstrup, K. Homburg, R. Larsen, M. B. Hansen, K. Titlestad, B. Aagaard, B. K. Møller, T. Barington, Reduced prevalence of SARS-CoV-2 infection in ABO blood group O, *Blood Advances* **4**, 4990–4993 (2020).
35. Y. Wu, Z. Feng, P. Li, Q. Yu, Relationship between ABO blood group distribution and clinical characteristics in patients with COVID-19, *Clinica Chimica Acta* **509**, 220–223 (2020).
36. S. Dzik, K. Eliason, E. B. Morris, R. M. Kaufman, C. M. North, COVID-19 and ABO blood groups, *Transfusion* **60**, 1883–1884 (2020).
37. J. M. O’Sullivan, S. Ward, H. Fogarty, J. S. O’Donnell, More on ‘Association between ABO blood groups and risk of SARS-CoV-2 pneumonia,’ *British Journal of Haematology* **190**, 27–28 (2020).
38. I. Huang, R. Pranata, Lymphopenia in severe coronavirus disease-2019 (COVID-19): systematic review and meta-analysis, *Journal of Intensive Care* **8**, 36 (2020).
39. C. I. van der Made, A. Simons, J. Schuurs-Hoeijmakers, G. van den Heuvel, T. Mantere, S. Kersten, R. C. van Deuren, M. Steehouwer, S. V. van Reijmersdal, M. Jaeger, T. Hofste, G. Astuti, J. Corominas Galbany, V. van der Schoot, H. van der Hoeven, W. Hagmolen of ten Have, E. Klijn, C. van den Meer, J. Fiddelaers, Q. de Mast, C. P. Bleeker-Rovers, L. A. B. Joosten, H. G. Yntema, C. Gilissen, M. Nelen, J. W. M. van der Meer, H. G. Brunner, M. G. Netea, F. L. van de Veerdonk, A. Hoischen, Presence of Genetic Variants Among Young Men With Severe COVID-19, *JAMA* **324**, 663–673 (2020).
40. Y. Li, M. Oosting, P. Deelen, I. Ricaño-Ponce, S. Smeekens, M. Jaeger, V. Matzaraki, M. A. Swertz, R. J. Xavier, L. Franke, C. Wijmenga, L. A. B. Joosten, V. Kumar, M. G. Netea, Inter-individual variability and genetic influences on cytokine responses to bacteria and fungi, *Nature Medicine* **22**, 952–960 (2016).
41. R. Aguirre-Gamboa, I. Joosten, P. C. M. Urbano, R. G. van der Molen, E. van Rijssen, B. van Cranenbroek, M. Oosting, S. Smeekens, M. Jaeger, M. Zorro, S. Withoff, A. E. van Herwaarden, F. C. G. J. Sweep, R. T. Netea, M. A. Swertz, L. Franke, R. J. Xavier, L. A. B. Joosten, M. G. Netea, C. Wijmenga, V. Kumar, Y. Li, H. J. P. M. Koenen, Differential Effects of Environmental and Genetic Factors on T and B Cell Immune Traits, *Cell Reports* **17**, 2474–2487 (2016).
42. M. Schirmer, S. P. Smeekens, H. Vlamakis, M. Jaeger, M. Oosting, E. A. Franzosa, R. ter Horst, T. Jansen, L. Jacobs, M. J. Bonder, A. Kurilshikov, J. Fu, L. A. B. Joosten, A.

- Zhernakova, C. Huttenhower, C. Wijmenga, M. G. Netea, R. J. Xavier, Linking the Human Gut Microbiome to Inflammatory Cytokine Production Capacity, *Cell* **167**, 1125–1136.e8 (2016).
43. V. P. Mourits, V. A. C. M. Koeken, L. C. J. de Bree, S. J. C. F. M. Moorlag, W. C. Chu, X. Xu, H. Dijkstra, H. Lemmers, L. A. B. Joosten, Y. Wang, R. van Crevel, M. G. Netea, A. González, Ed. BCG-Induced Trained Immunity in Healthy Individuals: The Effect of Plasma Muramyl Dipeptide Concentrations, *Journal of Immunology Research* **2020**, 5812743 (2020).
44. T. S. Shah, J. Z. Liu, J. A. B. Floyd, J. A. Morris, N. Wirth, J. C. Barrett, C. A. Anderson, optiCall: a robust genotype-calling algorithm for rare, low-frequency and common variants, *Bioinformatics* **28**, 1598–1603 (2012).
45. P.-R. Loh, P. Danecek, P. F. Palamara, C. Fuchsberger, Y. A Reshef, H. K Finucane, S. Schoenherr, L. Forer, S. McCarthy, G. R. Abecasis, R. Durbin, A. L Price, Reference-based phasing using the Haplotype Reference Consortium panel, *Nature Genetics* **48**, 1443–1448 (2016).
46. X. Chu, M. Jaeger, O. B. Bakker, R. Aguirre-Gamboa, M. Oosting, S. P. Smeekeens, S. Withoff, R. T. Netea-Maier, H. J. P. M. Koenen, I. Joosten, R. J. Xavier, L. Franke, L. a. B. Joosten, S. Sanna, V. Kumar, C. Wijmenga, M. G. Netea, Y. Li, in *European Journal of Human Genetics*, (Nature Publishing Group, 2019), vol. 27, pp. 787–787.
47. A. A. Shabalín, Matrix eQTL: ultra fast eQTL analysis via large matrix operations, *Bioinformatics*. **28** (2012), doi:10.1093/bioinformatics/bts163.
48. R. N. Tunjungputri, Y. Li, P. G. de Groot, C. A. Dinarello, S. P. Smeekeens, M. Jaeger, M. Doppenberg-Oosting, M. Crujisen, H. Lemmers, H. Toenhake-Dijkstra, R. Aguirre-Gamboa, V. Kumar, T. N. Wijmenga, L. A. B. Joosten, M. G. Netea, A. van der Ven, Q. de Mast, The inter-relationship of platelets with interleukin-1 β -mediated inflammation in humans, *Thrombosis and Haemostasis* **118**, 2112–2125 (2018).
49. C. Giambartolomei, D. Vukcevic, E. E. Schadt, L. Franke, A. D. Hingorani, C. Wallace, V. Plagnol, Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics, *PLOS Genetics* **10**, e1004383 (2014).

A



B

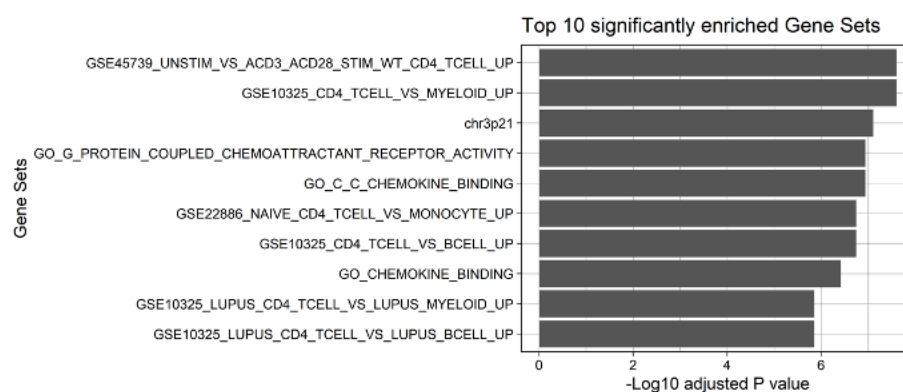


Fig.1. Functional annotation of COVID-19 loci using the FUMA pipeline. This was done based on genes identified after using FUMA to map QTLs based on their genomic location, eQTL associations, and histone activity. A) MAGMA Tissue expression results on 30 general tissues type (GTEx v8), B) The top 10 significant enriched gene sets.

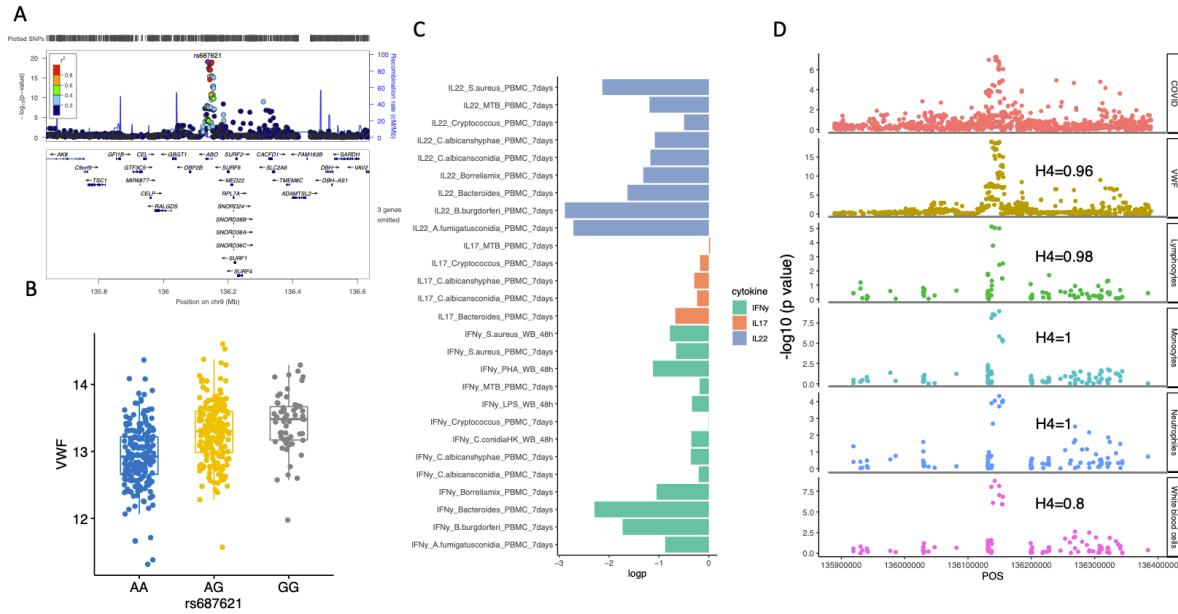


Fig 3. Functional annotation of ABO loci. A) locus zoom plot showing the significant association between ABO loci and VWF level. B) a boxplot showing COVID-19 risk allele(rs687621-G) associated with increasing VWF level ($p\text{-value} = 9.58 \times 10^{-20}$). C) a barplot showing consistent negative correlations between VWF levels and T cell-derived cytokines D) scatter plots showing colocalization between ABO loci with VWF, lymphocytes, monocytes, neutrophils and white blood cell counts.

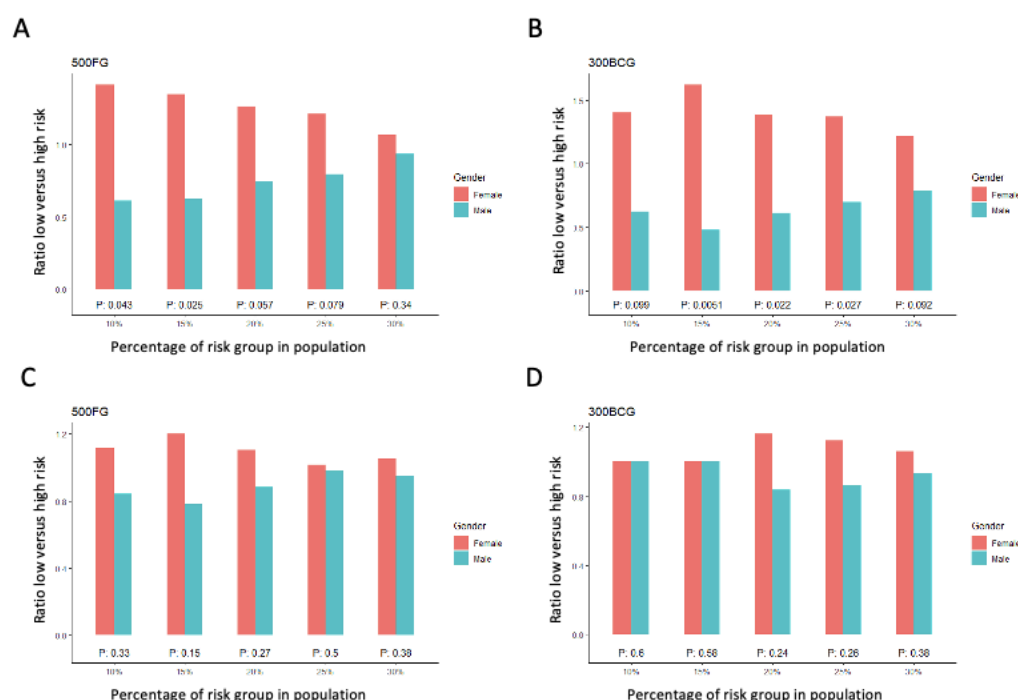


Fig. 4. Correlation of COVID-19 PRS with gender. A) bar plot representing the ratio of low risk versus high risk in 500FG. The X-axis shows the range of different quantiles (e.g., 10% corresponds to those individuals with PRS between 0th and 10th percentile of the population), and the Y-axis shows the odds ratio when comparing low PRS risk and high PRS risk in the male and female group from different quantiles. B) bar plot representing the ratio of low risk versus high risk in 300BCG. C) Bar plot representing the ratio of low versus high PRS based risk between men and women in 500FG calculated without including the sex chromosomes. D). Bar plot representing the ratio of low versus high PRS based risk between men and women in 300BCG calculated without including the sex chromosomes.

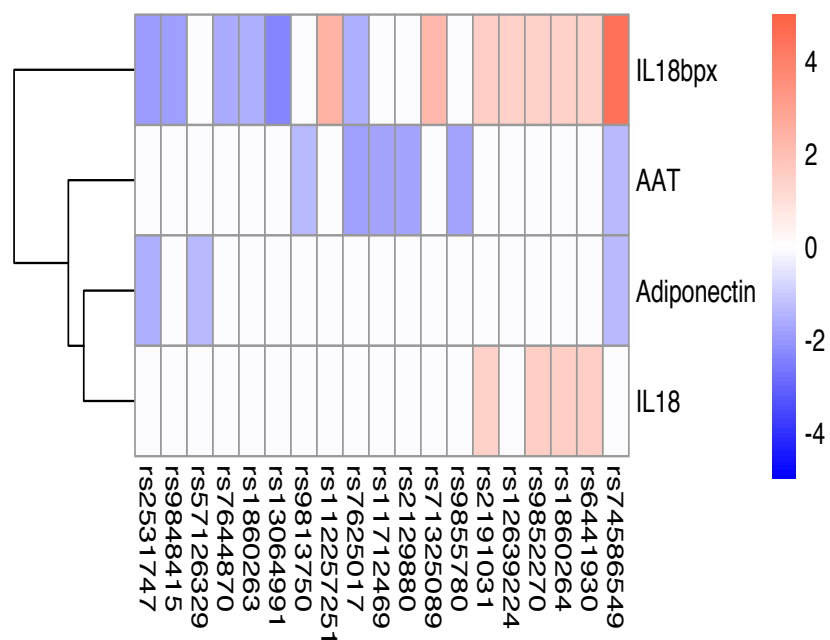


Fig. S1 Heatmap of the genetic correlations between 3p21.31 loci and circulating mediator.

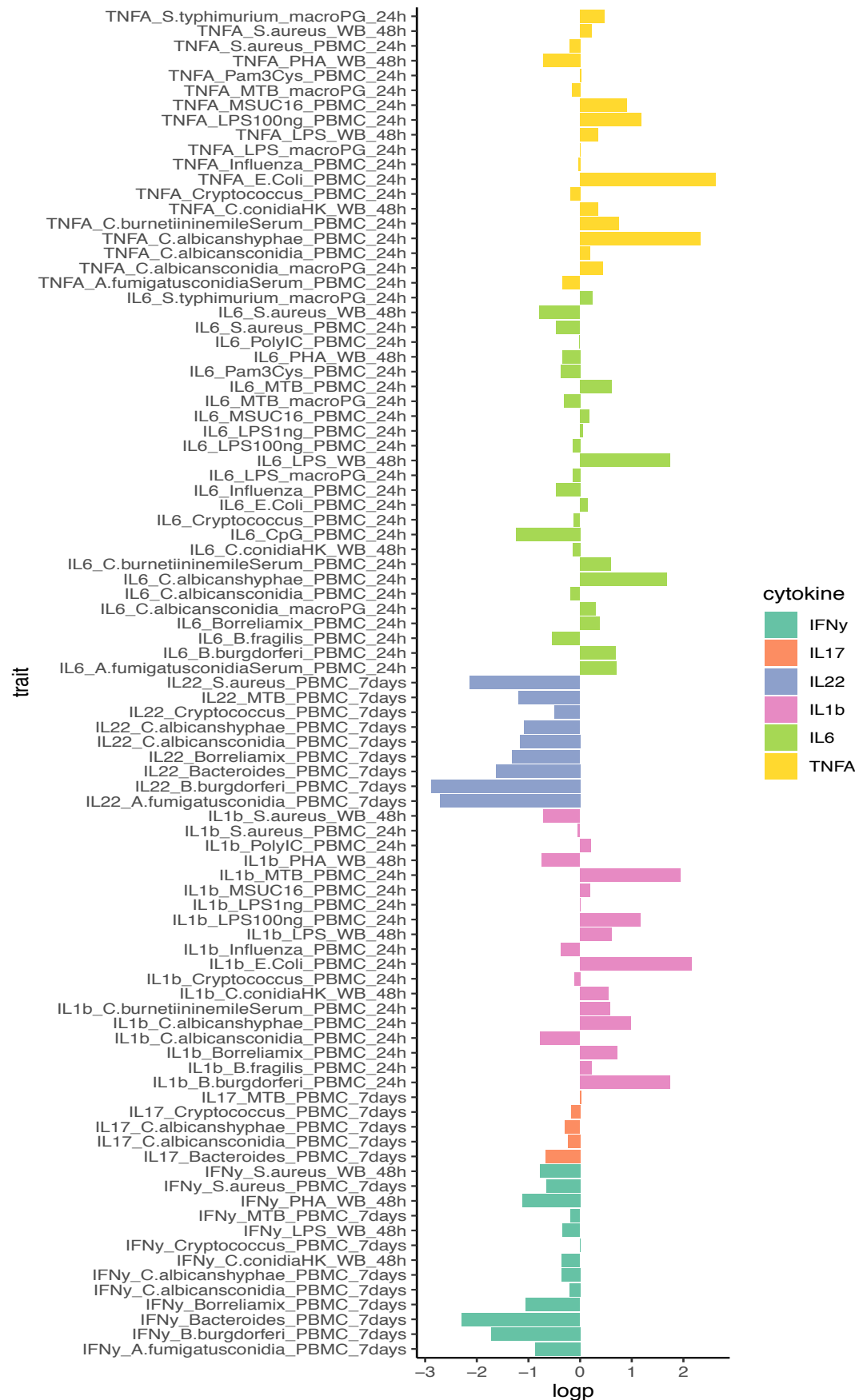


Fig. S2 A barplot showing correlations between VWF levels and cytokines.

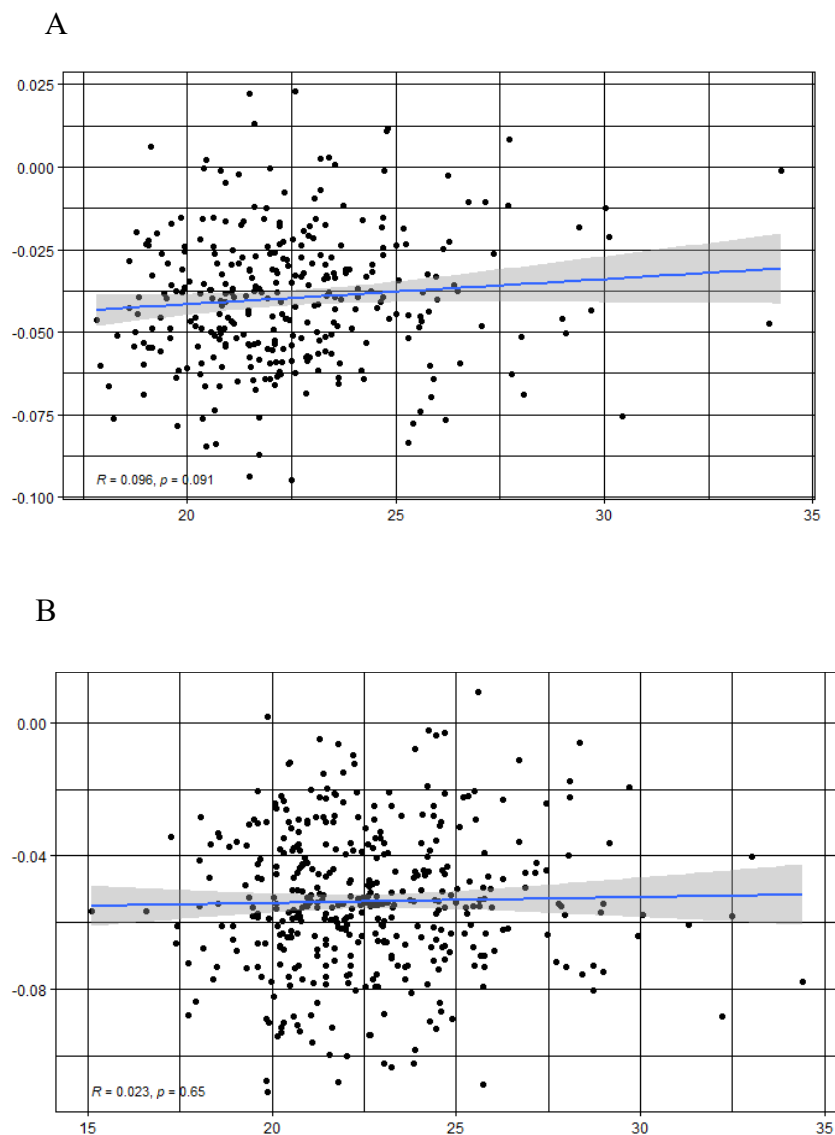


Fig. S3 Scatter plot showing the correlation between PRS with BMI A) in 500FG B) in 300BCG.

Table S1: Numbers and basic characteristics of participants included in the study.

| | 500FG (N = 451) | 300BCG (N = 313) |
|--------------------------|-------------------|-------------------|
| Age (years) | | |
| mean (SD) | 28.1 (13.5) | 25.9 (10.6) |
| median [Min, Max] | 23 [18, 75] | 23 [18, 71] |
| Sex | | |
| Male sex: N (%) | 211 (43.1%) | 141 (43.3%) |
| BMI | | |
| mean (SD) | 22.8 (2.91) | 22.5 (2.56) |
| median [Min, Max] | 22.2 [15.1, 34.6] | 22.1 [17.8, 34.2] |
| Smoking behaviour | | |
| Smoker *: N (%) | 85 (20.3%) | 63 (20.2%) |
| Missing | 32 | 0 |

SD = standard deviation

* Smokers were classified as people who regularly smoked or did so in the past.

Tabel S2 the assoication of COVID-19 risk loci with phenotypes in GWAS catalog (last accessed on 15-July-2020)

| SNP | Mapped gene | Trait(s) | Study accession |
|-------------|-------------|-----------------------------------|-----------------|
| rs505922-C | ABO | venous thromboembolism | GCST000354 |
| rs505922-C | ABO | pancreatic carcinoma | GCST000456 |
| rs8176719-G | ABO | malaria | GCST001637 |
| rs505922-T | ABO | von Willebrand factor measurement | GCST001798 |

| | | | |
|-------------|-----|---|------------|
| rs505922-? | ABO | tumor necrosis factor-alpha measurement | GCST000189 |
| rs507666-G | ABO | ICAM-1 measurement | GCST000210 |
| rs505922-? | ABO | Graves disease | GCST001982 |
| rs8176719-G | ABO | venous thromboembolism | GCST001557 |
| rs2519093-A | ABO | venous thromboembolism | GCST001557 |
| rs507666-A | ABO | ICAM-1 measurement | GCST001047 |
| rs505922-T | ABO | duodenal ulcer | GCST001433 |
| rs505922-T | ABO | Graves disease | GCST001200 |
| rs612169-G | ABO | metabolite measurement | GCST001217 |
| rs505922-C | ABO | venous thromboembolism | GCST001253 |
| rs8176685-G | ABO | platelet component distribution width | GCST004616 |
| rs505922-C | ABO | type II diabetes mellitus | GCST005414 |
| rs507666-? | ABO | e-selectin measurement | GCST004365 |
| rs2519093-C | ABO | low density lipoprotein cholesterol measurement | GCST007141 |
| rs2519093-C | ABO | low density lipoprotein cholesterol measurement | GCST007141 |
| rs2519093-C | ABO | low density lipoprotein cholesterol measurement | GCST007141 |
| rs2519093-C | ABO | low density lipoprotein cholesterol measurement | GCST007141 |
| rs507666-G | ABO | low density lipoprotein cholesterol measurement | GCST007141 |
| rs507666-G | ABO | low density lipoprotein cholesterol measurement | GCST007141 |
| rs507666-G | ABO | low density lipoprotein cholesterol measurement | GCST007141 |
| rs507666-A | ABO | low density lipoprotein cholesterol measurement | GCST002321 |

| | | | |
|---------------|-----|--|------------|
| rs507666-A | ABO | total cholesterol measurement | GCST002321 |
| rs505922-? | ABO | pancreatic carcinoma | GCST002991 |
| rs505922-? | ABO | alkaline phosphatase measurement, clinical laboratory measurement | GCST003540 |
| rs2519093-T | ABO | coronary artery disease | GCST003116 |
| rs8176685-G | ABO | basophil count, eosinophil count | GCST004624 |
| rs2519093-C | ABO | total cholesterol measurement | GCST007143 |
| rs2519093-C | ABO | total cholesterol measurement | GCST007143 |
| rs2519093-C | ABO | total cholesterol measurement | GCST007143 |
| rs507666-G | ABO | total cholesterol measurement | GCST007143 |
| rs507666-G | ABO | total cholesterol measurement | GCST007143 |
| rs507666-G | ABO | total cholesterol measurement | GCST007143 |
| rs115478735-T | ABO | blood protein measurement | GCST005806 |
| rs115478735-T | ABO | blood protein measurement | GCST005806 |
| rs2519093-? | ABO | alkaline phosphatase measurement | GCST006016 |
| rs8176719-? | ABO | factor VIII measurement | GCST007445 |
| rs8176719-? | ABO | factor VIII measurement | GCST007445 |
| rs2519093-T | ABO | e-selectin measurement | GCST008202 |
| rs2519093-T | ABO | ICAM-1 measurement | GCST008210 |
| rs2519093-? | ABO | sleep duration, low density lipoprotein cholesterol measurement | GCST009365 |
| rs2519093-T | ABO | venous thromboembolism | GCST009030 |
| rs2519093-T | ABO | low density lipoprotein cholesterol measurement | GCST006612 |
| rs507666-A | ABO | Eczema, allergic rhinitis | GCST009717 |
| rs507666-? | ABO | blood protein measurement | GCST010104 |
| rs507666-? | ABO | blood protein measurement | GCST010104 |
| rs507666-? | ABO | blood protein measurement | GCST010104 |
| rs507666-A | ABO | blood protein measurement | GCST008478 |

| | | | |
|-------------|-----------------|---|------------|
| rs2519093-? | ABO | eosinophil count | GCST007065 |
| rs2519093-? | ABO | leukocyte count | GCST007070 |
| rs495828-A | AL772161.2, ABO | angiotensin converting enzyme activity measurement | GCST000565 |
| rs495828-T | AL772161.2, ABO | erythrocyte count | GCST000588 |
| rs495828-T | AL772161.2, ABO | hematocrit | GCST000583 |
| rs495828-T | AL772161.2, ABO | alkaline phosphatase measurement | GCST000583 |
| rs495828-T | AL772161.2, ABO | hemoglobin measurement | GCST000583 |
| rs579459-C | ABO, AL772161.2 | e-selectin measurement | GCST000476 |
| rs579459-C | ABO, AL772161.2 | coronary heart disease | GCST000998 |
| rs579459-T | ABO, AL772161.2 | adhesion molecule measurement, soluble P-selectin measurement | GCST000599 |
| rs579459-T | ABO, AL772161.2 | alkaline phosphatase measurement | GCST001276 |
| rs495828-T | AL772161.2, ABO | venous thromboembolism | GCST001557 |
| rs579459-C | ABO, AL772161.2 | stroke, coronary heart disease | GCST002287 |
| rs579459-C | ABO, AL772161.2 | coronary heart disease | GCST002289 |
| rs579459-? | ABO, AL772161.2 | large artery stroke, coronary heart disease | GCST002290 |
| rs579459-T | ABO, AL772161.2 | urinary metabolite measurement | GCST002364 |
| rs579459-T | ABO, AL772161.2 | urinary metabolite measurement | GCST002364 |
| rs579459-T | ABO, AL772161.2 | total cholesterol measurement | GCST004231 |
| rs579459-T | ABO, AL772161.2 | low density lipoprotein cholesterol measurement | GCST004233 |
| rs579459-T | ABO, AL772161.2 | low density lipoprotein cholesterol measurement | GCST004233 |
| rs579459-T | ABO, AL772161.2 | low density lipoprotein cholesterol measurement | GCST007141 |
| rs579459-T | ABO, AL772161.2 | low density lipoprotein cholesterol measurement | GCST007141 |

| | | | |
|------------|-----------------|--|------------|
| rs579459-T | ABO, AL772161.2 | low density lipoprotein cholesterol measurement | GCST007141 |
| rs579459-T | ABO, AL772161.2 | low density lipoprotein cholesterol measurement | GCST007141 |
| rs495828-T | AL772161.2, ABO | blood metabolite measurement | GCST002442 |
| rs579459-T | ABO, AL772161.2 | blood metabolite measurement | GCST002442 |
| rs579459-T | ABO, AL772161.2 | erythrocyte count | GCST001765 |
| rs633862-A | AL772161.2, ABO | Malignant epithelial tumor of ovary | GCST002576 |
| rs579459-C | ABO, AL772161.2 | total cholesterol measurement | GCST003214 |
| rs579459-C | ABO, AL772161.2 | low density lipoprotein cholesterol measurement | GCST003216 |
| rs579459-C | ABO, AL772161.2 | HbA1c measurement | GCST007954 |
| rs579459-T | ABO, AL772161.2 | total cholesterol measurement | GCST007143 |
| rs579459-T | ABO, AL772161.2 | total cholesterol measurement | GCST007143 |
| rs579459-T | ABO, AL772161.2 | total cholesterol measurement | GCST007143 |
| rs495828-T | AL772161.2, ABO | triglyceride measurement, alcohol drinking | GCST008083 |
| rs495828-T | AL772161.2, ABO | low density lipoprotein cholesterol measurement, alcohol drinking | GCST008079 |
| rs495828-T | AL772161.2, ABO | triglyceride measurement, alcohol drinking | GCST008083 |
| rs495828-T | AL772161.2, ABO | low density lipoprotein cholesterol measurement, alcohol drinking | GCST008079 |
| rs495828-T | AL772161.2, ABO | low density lipoprotein cholesterol measurement, alcohol consumption measurement | GCST008078 |
| rs495828-T | AL772161.2, ABO | triglyceride measurement, alcohol consumption measurement | GCST008074 |
| rs600038-C | ABO, AL772161.2 | heart failure | GCST009541 |

[illegible]

| | | | |
|-------------|-----------------|--|------------|
| rs579459-C | ABO, AL772161.2 | blood protein measurement | GCST006585 |
| rs579459-C | ABO, AL772161.2 | blood protein measurement | GCST006585 |
| rs579459-C | ABO, AL772161.2 | blood protein measurement | GCST006585 |
| rs579459-C | ABO, AL772161.2 | blood protein measurement | GCST006585 |
| rs579459-C | ABO, AL772161.2 | blood protein measurement | GCST006585 |
| rs579459-C | ABO, AL772161.2 | blood protein measurement | GCST006585 |
| rs579459-C | ABO, AL772161.2 | blood protein measurement | GCST006585 |
| rs495828-T | AL772161.2, ABO | blood protein measurement | GCST008478 |
| rs495828-? | AL772161.2, ABO | Eczema | GCST007075 |
| rs507666-A | ABO | total cholesterol measurement | GCST002896 |
| rs2519093-? | ABO | allergy | GCST003990 |
| rs507666-G | ABO | pulse pressure measurement | GCST007096 |
| rs507666-G | ABO | diastolic blood pressure | GCST007094 |
| rs8176643-A | ABO | hemoglobin measurement | GCST004615 |
| rs8176643-? | ABO | susceptibility to childhood ear infection measurement | GCST005013 |
| rs505922-C | ABO | CD209 antigen measurement | GCST004365 |
| rs505922-C | ABO | Ischemic stroke, von Willebrand factor measurement, homocysteine measurement | GCST004598 |
| rs2519093-T | ABO | venous thromboembolism | GCST004256 |
| rs8176645-A | ABO | venous thromboembolism | GCST004256 |
| rs8176643-A | ABO | hematocrit | GCST004604 |
| rs612169-G | ABO | von Willebrand factor measurement | GCST004365 |
| rs507666-A | ABO | coronary artery disease | GCST004787 |
| rs2519093-T | ABO | blood protein measurement | GCST005806 |
| rs2519093-T | ABO | blood protein measurement | GCST005806 |
| rs2519093-T | ABO | blood protein measurement | GCST005806 |
| rs2519093-T | ABO | blood protein measurement | GCST005806 |

| | | | |
|-------------|-----|--|------------|
| rs2519093-T | ABO | blood protein measurement | GCST005806 |
| rs2519093-T | ABO | blood protein measurement | GCST005806 |
| rs2519093-T | ABO | blood protein measurement | GCST005806 |
| rs2519093-T | ABO | blood protein measurement | GCST005806 |
| rs2519093-T | ABO | blood protein measurement | GCST005806 |
| rs2519093-T | ABO | blood protein measurement | GCST005806 |
| rs505922-C | ABO | pancreatic carcinoma | GCST005434 |
| rs2519093-T | ABO | coronary artery disease | GCST005195 |
| rs2519093-? | ABO | hematocrit | GCST005994 |
| rs2519093-? | ABO | hemoglobin measurement | GCST005995 |
| rs2519093-? | ABO | erythrocyte count | GCST005996 |
| rs507666-A | ABO | coronary artery disease | GCST005196 |
| rs505922-C | ABO | peripheral arterial disease | GCST008474 |
| rs8176685-? | ABO | von Willebrand factor measurement | GCST007446 |
| rs8176719-? | ABO | von Willebrand factor measurement | GCST007446 |
| rs8176643-A | ABO | blood protein measurement | GCST005806 |
| rs8176643-A | ABO | blood protein measurement | GCST005806 |
| rs8176643-A | ABO | blood protein measurement | GCST005806 |
| rs8176643-A | ABO | blood protein measurement | GCST005806 |
| rs8176685-? | ABO | von Willebrand factor measurement | GCST007446 |
| rs505922-C | ABO | type II diabetes mellitus | GCST009379 |
| rs2519093-? | ABO | sleep duration, low density lipoprotein cholesterol measurement | GCST009366 |
| rs507666-A | ABO | e-selectin measurement | GCST009572 |
| rs507666-A | ABO | low density lipoprotein cholesterol measurement, physical activity | GCST007284 |
| rs2519093-T | ABO | allergic rhinitis | GCST006409 |
| rs505922-C | ABO | blood protein measurement | GCST005806 |
| rs505922-C | ABO | blood protein measurement | GCST005806 |

| | | | |
|-------------|-----------------|---|------------|
| rs507666-A | ABO | blood protein measurement | GCST005806 |
| rs507666-A | ABO | blood protein measurement | GCST005806 |
| rs2519093-T | ABO | HMG CoA reductase inhibitor use measurement | GCST007931 |
| rs507666-? | ABO | total cholesterol measurement | GCST008045 |
| rs2769071-A | ABO | fibroblast growth factor 23 measurement | GCST006491 |
| rs507666-? | ABO | low density lipoprotein cholesterol measurement | GCST008037 |
| rs505922-C | ABO | disposition index measurement | GCST008111 |
| rs495828-? | AL772161.2, ABO | erythrocyte count | GCST004008 |
| rs495828-T | AL772161.2, ABO | hematocrit | GCST004003 |
| rs579459-T | ABO, AL772161.2 | erythrocyte count | GCST004008 |
| rs495828-T | AL772161.2, ABO | hemoglobin measurement | GCST004005 |
| rs495828-? | AL772161.2, ABO | hemoglobin measurement | GCST004005 |
| rs579459-T | ABO, AL772161.2 | low density lipoprotein cholesterol measurement | GCST004236 |
| rs579459-T | ABO, AL772161.2 | total cholesterol measurement | GCST004235 |
| rs579459-T | ABO, AL772161.2 | total cholesterol measurement | GCST004235 |
| rs495828-T | AL772161.2, ABO | coronary artery disease | GCST005194 |
| rs495828-T | AL772161.2, ABO | low density lipoprotein cholesterol measurement, alcohol drinking | GCST008086 |
| rs495828-T | AL772161.2, ABO | low density lipoprotein cholesterol measurement, alcohol drinking | GCST008086 |
| rs579459-? | ABO, AL772161.2 | platelet reactivity measurement | GCST008457 |
| rs495828-T | AL772161.2, ABO | low density lipoprotein cholesterol measurement, alcohol drinking | GCST008086 |
| rs495828-T | AL772161.2, ABO | low density lipoprotein cholesterol measurement, alcohol drinking | GCST008086 |

| | | | |
|--------------|-----------------|--|------------|
| rs495828-T | AL772161.2, ABO | low density lipoprotein cholesterol measurement, alcohol consumption measurement | GCST008078 |
| rs495828-T | AL772161.2, ABO | low density lipoprotein cholesterol measurement, alcohol consumption measurement | GCST008078 |
| rs495828-T | AL772161.2, ABO | low density lipoprotein cholesterol measurement, alcohol consumption measurement | GCST008078 |
| rs495828-T | AL772161.2, ABO | low density lipoprotein cholesterol measurement | GCST008077 |
| rs495828-T | AL772161.2, ABO | low density lipoprotein cholesterol measurement | GCST008077 |
| rs495828-T | AL772161.2, ABO | low density lipoprotein cholesterol measurement, alcohol drinking | GCST008079 |
| rs495828-T | AL772161.2, ABO | low density lipoprotein cholesterol measurement, alcohol drinking | GCST008079 |
| rs495828-T | AL772161.2, ABO | low density lipoprotein cholesterol measurement | GCST008077 |
| rs579459-T | ABO, AL772161.2 | low density lipoprotein cholesterol measurement | GCST008676 |
| rs495828-T | AL772161.2, ABO | tissue factor measurement | GCST009731 |
| rs495828-T | AL772161.2, ABO | hemoglobin measurement | GCST010083 |
| rs579459-C | ABO, AL772161.2 | alkaline phosphatase measurement | GCST010047 |
| rs579459-T | ABO, AL772161.2 | blood protein measurement | GCST006585 |
| rs579459-? | ABO, AL772161.2 | total cholesterol measurement | GCST008045 |
| rs579459-? | ABO, AL772161.2 | low density lipoprotein cholesterol measurement | GCST008037 |
| rs17764831-? | LZTFL1 | eosinophil count | GCST007065 |

Table S3: Independent lead GWAS loci reported in NEJM (hg38) and in hg19*.

| uniqID | rsID | p | nIndSigSNPs | IndSigSNPs |
|--------------------|-------------|----------|-------------|---------------------------------|
| 1:88993151:C:T | rs75558547 | 4.49E-06 | 1 | rs75558547 |
| 2:15724310:A:G | rs2080811 | 1.12E-06 | 1 | rs2080811 |
| 3:45876459:G:GA | rs11385942 | 1.15E-10 | 3 | rs11385942;rs34901975;rs3774641 |
| 3:149805983:C:G | rs144582715 | 7.17E-06 | 1 | rs144582715 |
| 4:180344873:A:G | rs1455662 | 9.60E-06 | 1 | rs1455662 |
| 6:20314719:C:T | rs911360 | 4.09E-06 | 1 | rs911360 |
| 6:23379360:A:AT | rs5874914 | 4.64E-06 | 1 | rs5874914 |
| 6:91010824:A:ATTAC | rs3073485 | 5.72E-06 | 1 | rs3073485 |
| 7:123006727:A:G | rs12706520 | 9.63E-06 | 1 | rs12706520 |
| 7:158020147:A:G | rs6970487 | 4.89E-06 | 1 | rs6970487 |
| 8:122848829:A:G | rs10091098 | 4.55E-06 | 1 | rs10091098 |
| 8:126952467:A:T | rs28730361 | 8.11E-06 | 1 | rs28730361 |
| 9:136139265:A:C | rs657152 | 4.95E-08 | 3 | rs657152;rs550057;rs647800 |
| 11:5691474:C:G | rs12796811 | 3.62E-06 | 1 | rs12796811 |
| 12:62463979:G:T | rs10877786 | 6.89E-07 | 1 | rs10877786 |
| 14:32422746:C:G | rs7152677 | 1.52E-06 | 1 | rs7152677 |
| 16:74936031:A:G | rs2059266 | 4.01E-06 | 1 | rs2059266 |
| 16:82338406:C:T | rs114093749 | 3.86E-06 | 2 | rs114093749;rs4569267 |
| 17:10376558:G:T | rs117438562 | 7.49E-06 | 1 | rs117438562 |
| 18:3724602:A:G | rs13381043 | 1.13E-07 | 2 | rs13381043;rs4797120 |
| 18:22624707:A:G | rs56132597 | 8.94E-06 | 1 | rs56132597 |
| 19:4717672:A:G | rs12610495 | 5.20E-06 | 1 | rs12610495 |
| 19:5518492:C:T | rs183544391 | 7.29E-06 | 1 | rs183544391 |
| 19:17923554:T:TC | rs3833287 | 4.24E-06 | 1 | rs3833287 |
| 23:15138357:C:T | rs55634010 | 9.75E-06 | 1 | rs55634010 |

| | | | | |
|------------------|------------|----------|---|------------|
| 23:146325425:A:T | rs10126492 | 8.81E-06 | 1 | rs10126492 |
|------------------|------------|----------|---|------------|

Table S4 Functional annotation of SNPs on genes by ANNOVAR*

| annotation | ref.count | ref.prop | count | prop | enrichment | fisher.P |
|----------------|-----------|------------|-------|-------|------------|----------|
| UTR3 | 233824 | 0.0093225 | 13 | 0.016 | 1.694 | 0.066 |
| UTR5 | 71546 | 0.00285252 | 3 | 0.004 | 1.278 | 0.512 |
| downstream | 284177 | 0.01133006 | 8 | 0.010 | 0.858 | 0.868 |
| exonic | 254736 | 0.01015625 | 8 | 0.010 | 0.957 | 1.000 |
| intergenic | 11684523 | 0.46585868 | 378 | 0.459 | 0.986 | 0.727 |
| intronic | 9137749 | 0.36431951 | 329 | 0.400 | 1.097 | 0.036 |
| ncRNA_exonic | 259951 | 0.01036417 | 2 | 0.002 | 0.234 | 0.015 |
| ncRNA_intronic | 2884355 | 0.11499843 | 79 | 0.096 | 0.835 | 0.090 |
| ncRNA_splicing | 1313 | 5.23E-05 | 0 | 0.000 | 0.000 | 1.000 |
| splicing | 2830 | 0.00011283 | 0 | 0.000 | 0.000 | 1.000 |
| upstream | 266686 | 0.0106327 | 3 | 0.004 | 0.343 | 0.058 |

Table S5 Roadmap epigenetic state enrichment analysis

| Regulatory* elements | total Bins | 1.00E-05 | 1.00E-06 | 1.00E-07 | P. 1E-5 | P. 1E-6 | P. 1E-7 |
|---------------------------------------|-----------------------|-----------------|-----------------|-----------------|----------------|----------------|----------------|
| Active enhancer states | 2068209 | 129 | 34 | 10 | 2.96E-25 | 6.76E-12 | 0.042 |
| Active promotor states | 513015 | 31 | 15 | 0 | 4.86E-06 | 1.59E-08 | 0.63 |
| All enhancer states | 2133264 | 133 | 37 | 10 | 7.26E-26 | 1.35E-13 | 0.042 |
| All promotor states | 525189 | 31 | 15 | 0 | 8.36E-06 | 1.62E-08 | 0.63 |
| Whole Genome | 15282831 | 368 | 71 | 34 | | | |

*Enhancer and promoter states were derived from a 15-state model based on five chromatin marks; the core set of five histone modification marks was shared by 127 reference epigenomes of 24 blood cells from the Roadmap Epigenomics project. Promoter states include active promoter states (TssA , TssAFlnk) and inactive states (TssBiv); Enhancer states include active enhancer states (Enh, EnhG) and inactive states (EnhBiv). Enrichment of COVID-19 loci in regulatory elements was estimated by using Fisher's exact test.

Table S6 Comparison polygenic risk score(PRS) between male and female in 500FG (N=451) and 300BCG (N=313).

Table continued (part 1)

| Fraction | Risk group | Male -500FG | Female -500FG | OR (95% CI) | Pvalue-500FG* |
|----------|------------|-------------|---------------|-----------------------|---------------|
| 10% | high risk | 26 | 19 | 1.496 (0.980 - 2.283) | 0.045 |
| 10% | low risk | 17 | 28 | | |
| 15% | high risk | 32 | 35 | 1.235 (0.883 - 1.727) | 0.147 |
| 15% | low risk | 25 | 42 | | |
| 20% | high risk | 41 | 49 | 1.171 (0.876 - 1.566) | 0.182 |
| 20% | low risk | 34 | 56 | | |
| 25% | high risk | 53 | 59 | 1.176 (0.907 - 1.526) | 0.140 |
| 25% | low risk | 44 | 68 | | |
| 30% | high risk | 61 | 73 | 1.154 (0.909 - 1.465) | 0.149 |
| 30% | low risk | 52 | 83 | | |

Table continued (part 2)

| Male -300BCG | Female -300BCG | OR (95% CI) | Pvalue-300FG* | Meta-Pvalue** |
|--------------|----------------|-----------------------|---------------|---------------|
| 15 | 17 | 1.434 (0.894 - 2.300) | 0.115 | 0.011 |
| 9 | 22 | | | |
| 25 | 21 | 1.770 (1.182 - 2.650) | 0.005 | 0.004 |
| 12 | 34 | | | |
| 33 | 29 | 1.524 (1.075 - 2.163) | 0.014 | 0.010 |
| 20 | 42 | | | |
| 41 | 37 | 1.397 (1.022 - 1.910) | 0.026 | 0.011 |
| 28 | 50 | | | |
| 45 | 48 | 1.163 (0.874 - 1.549) | 0.188 | 0.051 |

| | | | | |
|----|----|--|--|--|
| 38 | 55 | | | |
|----|----|--|--|--|

* the p value is obtained using the fisher exact test

** the p value is obtained using the meta analyzed z score approach

Table S7 Comparision polygeneic risk score(PRS) between male and female in 500FG (N=451) and 300BCG (N=313) (age and gender corrected model)

Table continued (part 1)

| Fraction | Risk group | Male -500FG | Female -500FG | OR (95% CI) | Pvalue-500FG* |
|----------|------------|-------------|---------------|-----------------------|---------------|
| 10% | high risk | 23 | 22 | 1.498 (0.997 - 2.249) | 0.043 |
| 10% | low risk | 14 | 31 | | |
| 15% | high risk | 32 | 35 | 1.442 (1.036 - 2.006) | 0.025 |
| 15% | low risk | 20 | 47 | | |
| 20% | high risk | 43 | 46 | 1.296 (0.969 - 1.734) | 0.057 |
| 20% | low risk | 32 | 58 | | |
| 25% | high risk | 53 | 58 | 1.231 (0.948 - 1.599) | 0.079 |
| 25% | low risk | 42 | 70 | | |
| 30% | high risk | 58 | 76 | 1.070 (0.841 - 1.361) | 0.336 |
| 30% | low risk | 54 | 81 | | |

Table continued (part 2)

| Male -300BCG | Female -300BCG | OR (95% CI) | Pvalue-300FG* | Meta-Pvalue** |
|--------------|----------------|-----------------------|---------------|---------------|
| 16 | 15 | 1.477 (0.903 - 2.415) | 0.099 | 0.030 |
| 10 | 21 | | | |
| 25 | 21 | 1.770 (1.182 - 2.650) | 0.005 | 0.005 |
| 12 | 34 | | | |

| | | | | |
|----|----|-----------------------|-------|-------|
| 31 | 31 | 1.48 (1.048 - 2.090) | 0.022 | 0.008 |
| 19 | 43 | | | |
| 43 | 35 | 1.397 (1.018 - 1.917) | 0.027 | 0.039 |
| 30 | 48 | | | |
| 47 | 46 | 1.241 (0.932 - 1.652) | 0.092 | 0.214 |
| 37 | 56 | | | |

* the p value is obtained using the fisher exact test

** the p value is obtained using the meta analyzed z score approach

Table S8 The replication of ABO loci in COVID-19 Host genetics initiative(HGI) data

| RSid | Effect Allele | Other Allele | Beta | SE | P [^] | P-HGI* |
|-----------|---------------|--------------|---------|--------|----------------|----------|
| rs8176719 | T | TC | -0.2772 | 0.052 | 9.93E-08 | 8.88E-07 |
| rs633862 | T | C | 0.2534 | 0.0515 | 8.79E-07 | 2.80E-06 |

[^]Pvalue from D. Ellinghaus et.al New England Journal of Medicine 2020, DOI: 10.1056/NEJMoa2020283

*Pvalue from COVID-19 HGI (<https://www.covid19hg.org/results/>, last accessed 21st of October 2020). The effect allele has been aligned.