Analysis of Genetic Host Response Risk Factors in Severe COVID-19 Patients


PrecisionLife Ltd, Long Hanborough, Oxford, UK

* These authors contributed equally.
† Corresponding author: steve@precisionlife.com

ABSTRACT

BACKGROUND

Epidemiological studies indicate that as many as 20% of individuals who test positive for COVID-19 develop severe symptoms that can require hospitalization. These symptoms include low platelet count, severe hypoxia, increased inflammatory cytokines and reduced glomerular filtration rate. Additionally, severe COVID-19 is associated with several chronic co-morbidities, including cardiovascular disease, hypertension and type 2 diabetes mellitus.

The identification of genetic risk factors that impact differential host responses to SARS-CoV-2, resulting in the development of severe COVID-19, is important in gaining greater understanding into the biological mechanisms underpinning life-threatening responses to the virus. These insights could be used in the identification of high-risk individuals and for the development of treatment strategies for these patients.

METHODS

As of June 6, 2020, there were 976 patients who tested positive for COVID-19 and were hospitalized, indicating they had a severe response to SARS-CoV-2. To overcome the limited number of patients with a mild form of COVID-19, we used similar control criteria to our previous study looking at shared genetic risk factors between severe COVID-19 and sepsis, selecting controls who had not developed sepsis despite having maximum co-morbidity risk and exposure to sepsis-causing pathogens.

RESULTS

Using a combinatorial (high-order epistasis) analysis approach, we identified 68 protein-coding genes that were highly associated with severe COVID-19. At the time of analysis, nine of these genes have been linked to differential response to viral pathogens including SARS-CoV-2. We also found many novel targets that are involved in key biological pathways associated with the development of severe COVID-19, including production of pro-inflammatory cytokines, endothelial cell dysfunction, lipid droplets, neurodegeneration and viral susceptibility factors.

CONCLUSION

The variants we found in genes relating to immune response pathways and cytokine production cascades, were in equal proportions across all severe COVID-19 patients, regardless of their co-morbidities. This suggests that such variants are not associated with any specific co-morbidity, but are common amongst patients who develop severe COVID-19. This is consistent with being able to find and validate severe disease biomarker signatures when larger patient datasets become available.

Several of the genes identified relate to lipid programming, beta-catenin and protein kinase C signalling. These processes converge in a central pathway involved in plasma membrane repair, clotting and wound healing. This pathway is largely driven by Ca\(^{2+}\) activation, which is a known serum biomarker associated with severe COVID-19 and ARDS. This suggests that aberrant calcium ion signalling may be responsible for driving severe COVID-19 responses in patients with variants in genes that regulate the expression and activity of this ion. We intend to perform further analyses to confirm this hypothesis.

Among the 68 severe COVID-19 risk-associated genes, we found several druggable protein targets and pathways. Nine are targeted by drugs that have reached at least Phase I clinical trials, and a further eight have active chemical starting points for novel drug development.

Several of these targets were particularly enriched in specific co-morbidities, providing insights into shared pathological mechanisms underlying both the development of severe COVID-19, ARDS and these predisposing co-morbidities. We can use these insights to identify patients who are at greatest risk of developing severe COVID-19 and develop targeted therapeutic strategies for them, with the aim of improving disease burden and survival rates.

NOTES

† The corresponding parties would like to acknowledge and thank the following individuals for their support: Krystyna Taylor*, Sayoni Das*, Matthew Pearson*, James Kozubek, Marcin Pawlowski, Claus Erik Jensen, Zbigniew Skowron, Gert Lykke Møller, Mark Strivens*, Steve Gardner**. This work was supported by a grant from the European Commission (Horizon 2020).
Introduction

Coronavirus disease 2019 (COVID-19), caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), is a major threat to public health. As of 16th June 2020, there are estimated to be over 8 million confirmed cases globally, resulting in approximately 437,000 deaths worldwide. Although many who develop the disease present with only mild symptoms, reports from multiple international health systems have shown that up to 20% of individuals testing positive for COVID-19 go on to develop severe forms of the disease that may require hospitalization.

Significant associations of disease severity risk have been observed with epidemiological factors including age, sex, blood group and ethnicity in addition to co-morbidity with many common conditions such as cardiovascular disease, diabetes, hypertension and chronic pulmonary diseases including asthma. It would be clinically useful to be able to identify the features that result in differential host responses, particularly those that predispose some patients to developing severe COVID-19. In the context of management of at-risk populations, prior to an effective and widely available vaccine, such insights could have great utility in developing new detection, protection and treatment strategies targeted specifically at these high-risk individuals.

In our previous paper, we explored the shared risk factors between sepsis (a major clinical feature in hospitalized COVID-19 patients) and severe COVID-19 disease. We observed that 59% of hospitalized COVID-19 cases also have sepsis, and that the two diseases share similar co-morbidity risks. In that study we identified 70 risk-associated genes in a sepsis population and found significant overlap in genetic risk variants between sepsis patients and those hospitalized with severe COVID-19.

As more data from COVID-19 patients has become available in the UK Biobank, we are now able to investigate the host response genetic risk factors directly, using genotype datasets from 779, 877 and 929 patients hospitalized with severe COVID-19 and comparing them against healthy controls. In this study, we have sought to identify the risk variants associated directly with severe COVID-19 patients, to gain insight of the underlying pathalogy and disease mechanisms in relation to this patient group.

Methods

COVID-19 test records were downloaded from the UK Biobank (data releases 18 May, 26 May and 6 June, 2020). The 18 May 2020 dataset included 5,657 test records relating to 3,002 individuals in UK Biobank. Of these, 1,073 patients had at least one positive COVID-19 test record, including 818 patients who were hospitalized and 255 who tested positive but were not hospitalized. In accordance with guidance from UK Biobank, we classified those 818 patients who were hospitalized after testing positive with COVID-19 as having a severe form of the disease and the 255 who were not hospitalized as having the mild form of the disease.

The analysis was conducted as a case-control study. After quality control and removal of samples with missing data, we assigned as cases 779 patients (442 males, 337 females) who had been hospitalized with severe COVID-19. Of these severe cases, 62% had one or more of the most common co-morbidities associated with high severe COVID-19 risk (Figure 1).

The most prevalent co-morbidity was hypertension (50%), followed by chronic respiratory disease (22%), diabetes (20%) and cardiovascular disease (18%). 248 (32%) patients were reported having two or more co-morbidities. 291 severe COVID-19 cases had none of these co-morbidities. In comparison, these co-morbidities were found to be less prevalent in the mild cases and an age and gender-matched random selection of the UK Biobank population (Figure 1).

Due to the limited number of patients with confirmed mild disease we were unfortunately not able to use this cohort directly as our control set due to its small size and lack of statistical power. We therefore adopted similar control cohort criteria to our...
previous sepsis study\textsuperscript{4}, based on evidence of shared risk factors and that severe COVID-19 patients often present with a similar pattern of co-morbid chronic conditions to those with sepsis. We selected patients who had not developed sepsis in spite of having been exposed to the most common sepsis-causing pathogens as well as having at least one of the most common chronic comorbidities known to increase a patient’s risk of developing sepsis and COVID-19 (Figure 2).

Figure 2: Incidence of co-morbidities by ICD-10 code from UK Biobank in the May 18, 2020 severe COVID-19 dataset (n=2,332, 1:2 case control ratio with 779 cases and 1,553 controls).

We selected the oldest possible patients (as age is also a critical phenotypic risk factor for the severe form of COVID-19) with the highest number of co-morbidities. Controls were therefore selected for a lack of disease given maximum risk and exposure, and were gender-matched against the severe COVID-19 cases in the ratio 2:1. The exact control criteria and distribution of the co-morbidities in cases are described in the Appendix.

This set of patients represents the most similar available control cohort who could reasonably be expected to have some genetic protective effect against developing severe COVID-19 (or to lack any such risk factors). By selecting controls with a higher prevalence of similar chronic co-morbidities, we seek to ensure that the severe COVID-associated signal observed is not simply caused by the co-morbidities represented in the patient set but has the potential to be a true COVID-19 related enrichment.

After quality control (removal of SNPs with <95% coverage across subjects), the genotype data for the two cohorts contained 542,245 SNPs. The age distribution of cases and controls is shown in Figure 3.

Figure 3: Age distribution of cases vs controls from UK Biobank in the May 18, 2020 severe COVID-19 dataset (779 cases and 1,553 controls).

We performed a SNP-based blood group analysis of the UK Biobank cohort by determining the blood groups (A, B, O and AB) of the cohort based on allele combinations of three SNPs (rs8176747, rs8176746, rs8176719) in the ABO gene\textsuperscript{5}. The blood group
frequencies of the cases and controls were calculated and two-sided Fisher’s exact tests were used to calculate blood-group specific odds ratios against the other blood groups (see Appendix).

Having generated our case-control dataset, we analyzed it using the precisionlife platform to identify risk-associated SNPs and genes that were strongly associated in the severe COVID-19 disease cohorts. This platform identifies high-order, disease-associated combinations of multi-modal (e.g. SNPs, transcriptomic, epidemiological or clinical) features at whole genome resolution in large patient cohorts. It has been validated across multiple different disease populations\textsuperscript{10,11,12}. This type of analysis is intractable to other existing methods due to the combinatorial explosion posed by the analysis of large numbers of patients with combinatorial non-linear additive combinations of features per patient.

When applied to genomic data, precisionlife finds high-order epistatic interactions (multi-SNP genotype signatures - typically of combinatorial order between 3-8) that are significantly more predictive of patients’ phenotype than those identified using existing single SNP based methods. When the individual SNPs making up these signatures are assessed individually across the whole population, they may fall below the GWAS (genome-wide association study) significance thresholds. However, we have demonstrated that when evaluated in combination with each other using multiple statistical validation techniques, these SNPs can be highly significant in particular disease sub-populations. The phenotype to which the signatures are associated in this context might be disease status, progression rate, therapy response or other, depending on the data available and study design.

We used the platform to find and statistically validate combinations of SNPs that together are strongly associated with the severe COVID-19 disease diagnosis. Analysis and annotation of these COVID-19-associated combinatorial genomic signatures took less than a day to complete, running on a dual CPU, 4 GPU compute server. The signatures identified by the analysis were then mapped to the human reference genome\textsuperscript{13} to identify disease-associated and clinically relevant target genes. A semantic knowledge graph derived from multiple public and private data sources was used to annotate the SNP and gene targets, including relevant tissue expression, chemical activity/tractability for gene targets, functional assignment and disease-associated literature. This provides contextual information to test the targets against the 5Rs criteria of early drug discovery\textsuperscript{14} and allows us efficiently to form strong, testable hypotheses for their mechanism of action in driving severe, life-threatening host responses to COVID-19 infection.

All of the significant disease signatures were traced back to the cases in which they were found and were associated with selected attributes such as case co-morbidities and COVID-19 test records. This generated a high-resolution stratification of severe COVID-19 patient subgroups and enabled further analysis of the different underlying factors relating to their specific forms of the disease.

Subsequently, UK Biobank added the COVID-19 test records of 1,508 individuals in data release 26 May, 2020 and 1,607 individuals in data release 6 June, 2020\textsuperscript{8}. Of these, 401 patients were tested positive and 158 of them were hospitalized. We repeated our analysis on these two updated datasets with a higher number of controls to make the analysis more robust (see Appendix). 50 (~70%) of the genes identified in our initial analysis were identified in more than one subsequent analysis (highlighted in bold in the gene tables) and therefore have an additional replication.

### Results

Using the severe COVID-19 dataset to perform a standard PLINK\textsuperscript{15} GWAS analysis revealed no significant SNPs (Figure 4) using a genome-wide significance threshold of \(p<5e-08\). The lowest SNP PLINK \(p\)-value reported was 9.02e-08.

![Figure 4: Manhattan plot generated using PLINK of genome-wide \(p\)-values of association for the May 18, 2020 severe COVID-19 UK Biobank cohort. The horizontal red and blue lines represent the genome-wide significance threshold at \(p<5e-08\) and \(p<1e-05\) respectively.](image-url)
When the same May 18, 2020 severe COVID-19 patient dataset was run using the precisionlife platform, we identified 3,515 combinations of SNP genotypes representing different combinations of SNP genotypes that were highly associated with the severe COVID-19 patient cohort (Table 1). The majority (n = 3,494) of SNPs were found in combinations with 5 or more SNP genotypes, and as such could not have been found using standard GWAS analysis methods.

**Table 1: Summary of May 18, 2020 Severe COVID-19 Cases vs Sepsis Controls disease study.**

<table>
<thead>
<tr>
<th></th>
<th>Severe COVID19/Sepsis Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>FDR</td>
<td>5%</td>
</tr>
<tr>
<td>Disease signatures</td>
<td>3,515</td>
</tr>
<tr>
<td>SNPs in all Disease signatures</td>
<td>5,402</td>
</tr>
<tr>
<td>Penetrance (cases represented by all signatures)</td>
<td>100%</td>
</tr>
<tr>
<td>RF scored SNPs</td>
<td>156</td>
</tr>
<tr>
<td>RF scored Genes</td>
<td>71</td>
</tr>
</tbody>
</table>

All of the SNP genotypes and their combinations were scored using a Random Forest (RF) algorithm based on a 5-fold cross-validation method to evaluate the accuracy with which the SNP genotype combinations predict the observed case: control split. 156 SNPs were scored by the RF algorithm, indicating that they accurately predict the differences between cases and controls (Figure 5A). The chromosomal distribution of the critical SNPs is shown in aggregate in Figure 5B.

Figure 5: Distributions of (A) RF scores and (B) chromosomal locations for critical disease associated SNPs

Clustering the SNPs by the patients in whom they co-occur allows us to generate a disease architecture of severe COVID-19 patients, providing useful insights into patient stratification. We can use this to find genes and biological pathways that are associated with patient sub-populations and co-morbidities, enabling the development of disease biomarkers and precision medicine strategies.
Figure 6: Disease architecture of the severe COVID-19 patient population generated by the precisionlife platform. Each circle represents a disease-associated SNP genotype, edges represent co-association in patients, and colours represent distinct patient sub-populations.

Figure 7: Disease architecture of the severe COVID-19 patient population highlighting (A) the critical disease SNPs (green) and (B) showing SNP genotypes (right homozygous major allele = blue, heterozygous = green, homozygous minor allele = gold).
Our analysis identified 68 protein-coding genes that were strongly associated with the disease phenotype in patients who developed severe COVID-19. To date, nine of these genes have already been linked to differential host responses to viral infections, including SARS-CoV-2 in various studies, providing validation for our hypothesis-free approach comparing severe COVID-19 patients against sepsis-free controls. We found several biological pathways and processes that were common across the 68 COVID-associated genes, including T cell regulation and host pathogenic responses, inflammatory cytokine production, and lipid formation and endothelial cell function (Figure 9).

Figure 9: Biological pathways and processes known to be associated with some of the genes replicated across the datasets used in this study.
We identified twelve genes that are associated with host immune response to viral infections, including SARS-CoV-2 (Table 2). Variants in several of these genes have been associated with increased infectivity to these strains. Genes marked in bold represent those that were further validated in subsequent analyses using the updated severe COVID-19 cases.

Table 2: Severe COVID-19 risk-associated genes associated with host immune response to virus infection, listed alphabetically. Bold text indicates genes found in multiple study populations.

<table>
<thead>
<tr>
<th>GENE</th>
<th>FUNCTION</th>
<th>MECHANISM OF ACTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANTXR1</td>
<td>Anthrax toxin receptor 1 (TEM8), type I transmembrane protein</td>
<td>ANTXR1 is used by anthrax toxin and Seneca valley virus (SVV) to gain entry into host cells</td>
</tr>
<tr>
<td>ATRNL1</td>
<td>Attractin-like 1</td>
<td>Phenotypes associated with decreased Salmonella invasion, hepatitis C virus replication and vaccinia virus infection</td>
</tr>
<tr>
<td>CTR9</td>
<td>Paf/RNA polymerase II complex component</td>
<td>Acts as a host restriction factor, suppressing HIV replication</td>
</tr>
<tr>
<td>DNAH20</td>
<td>Dynein heavy chain</td>
<td>Expressed in lung cilia, may be responsible for effective pathogenic clearance</td>
</tr>
<tr>
<td>GFRA1</td>
<td>GDNF family receptor alpha</td>
<td>SNPs associated with susceptibility to hepatitis C and Staphylococcus aureus infection</td>
</tr>
<tr>
<td>HOPX</td>
<td>Hop homeodomain protein, cofactor</td>
<td>High expression in expanded CD8+ T cells, improving SARS-CoV-2 viral clearance</td>
</tr>
<tr>
<td>IKZF2</td>
<td>Ikaros family member, Helios</td>
<td>Helios expression is high in regulatory T cells, necessary for self-tolerance and prevention of autoimmunity</td>
</tr>
<tr>
<td>ITK</td>
<td>IL-2 inducible T cell kinase, Th2 differentiation</td>
<td>Altered expression associated with SARS-CoV-2 infectivity</td>
</tr>
<tr>
<td>MEP1B</td>
<td>Meprin A subunit beta, membrane metallopeptidase</td>
<td>Macrophages in Mep1b−/− mice have greater phagocytic function. Biomarker for high tuberculosis burden.</td>
</tr>
<tr>
<td>NRDE2</td>
<td>RNA interference</td>
<td>RNA interference is a mechanism utilized by the innate immune response. NRDE2 is used by Kaposi’s sarcoma-associated herpesvirus (KSHV) for late gene expression</td>
</tr>
<tr>
<td>SPEF2</td>
<td>Sperm flagellar 2</td>
<td>Loss of Spef2 in mice results in increased inflammatory response to Streptococcus pneumonia infection due to defects in pulmonary cilia.</td>
</tr>
<tr>
<td>TBC1D2</td>
<td>TBC1 domain family member, regulates Rab7A activation</td>
<td>SNP associated with severe H1N1 infection</td>
</tr>
</tbody>
</table>

Our analysis also revealed five genes regulating pro-inflammatory pathways such as necroptosis, reactive oxygen species (ROS) production and cytokine signalling (Table 3).

Table 3: Severe COVID-19 risk-associated genes that have a role in regulation of inflammatory cytokine production

<table>
<thead>
<tr>
<th>GENE</th>
<th>FUNCTION</th>
<th>MECHANISM OF ACTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLKL</td>
<td>Mixed lineage pseudokinase</td>
<td>Activator of TNF-induced necroptosis, high expression results in necrotic cell death following coronavirus infection</td>
</tr>
<tr>
<td>NRROS</td>
<td>Negative regulator of reactive oxygen species (ROS)</td>
<td>Limits phagocytic ROS and pro-inflammatory cytokine production during host immune response</td>
</tr>
<tr>
<td>RAB3C</td>
<td>Rab GTPase</td>
<td>RAB3C increases cells’ ability to release IL-6 and activate the JAK2-STAT3 pathway.</td>
</tr>
<tr>
<td>RBM47</td>
<td>RNA binding motif protein</td>
<td>Promotes IL-10 production in B cells, repressed by TGF-β</td>
</tr>
<tr>
<td>SEMASA</td>
<td>Semaphorin 5A</td>
<td>High levels of SemaSA are associated with high IFN-Y and low IL-4 expression</td>
</tr>
</tbody>
</table>

There were also six genes associated with severe COVID-19 patients that play central roles in lipid droplet biology, as well as having high expression in adipose tissue, correlating with serum lipid levels and coronary artery disease (Table 4).
Amongst other cancer-associated genes, we identified nine that directly interact with the Wnt/β-catenin signaling pathway (Table 6). Except for KRAS and SLC9A9, all of the proteins encoded by these genes act as endogenous inhibitors of the pathway.

### Table 4: Severe COVID-19 risk-associated genes that are highly expressed in adipose tissue and relate to lipid storage and signalling

<table>
<thead>
<tr>
<th>GENE</th>
<th>FUNCTION</th>
<th>MECHANISM OF ACTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIDEA</td>
<td>Cell death inducing DFFA like effector</td>
<td>Co-localizes with perilins, highly expressed in brown adipose tissue</td>
</tr>
<tr>
<td>MACROD230</td>
<td>Mono-ADP ribosylhydrolase</td>
<td>Highly associated with VAP-1 in human adipose tissue</td>
</tr>
<tr>
<td>PLIN4</td>
<td>Perilipin 4</td>
<td>Lipid droplet protein, highly expressed in white adipose tissue, high expression results in cardiac steatosis</td>
</tr>
<tr>
<td>SLIT3</td>
<td>Slit guidance ligand 3</td>
<td>High expression in adipose tissue controlled by Prdm16 expression. Upregulated in obese patients</td>
</tr>
<tr>
<td>TMEM159</td>
<td>Promethin</td>
<td>Activated by peroxisome proliferator-activated receptor γ1 (PPARγ1), partners with seipin to regulate lipid droplet organization</td>
</tr>
<tr>
<td>XKR6</td>
<td>XK-related 6</td>
<td>SNP association with serum lipid levels and risk of CAD</td>
</tr>
</tbody>
</table>

In addition, our analysis revealed 12 more genes that may be implicated in the vascular complications seen in COVID-19 (Table 5). These are all highly expressed in the cardiovascular system, modulating cardiac function and endothelial cell homeostasis.

### Table 5: Severe COVID-19 risk-associated genes that are implicated in cardiovascular and endothelial cell function

<table>
<thead>
<tr>
<th>GENE</th>
<th>FUNCTION</th>
<th>MECHANISM OF ACTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANTXR1</td>
<td>Anthrax toxin receptor 1, type I transmembrane protein (TEM8)</td>
<td>Loss of ANTXR1 results in endothelial basement membrane loss and leaky blood vessels</td>
</tr>
<tr>
<td>HOPX</td>
<td>Hop homeodomain protein, cofactor</td>
<td>Acts via serum response factor (SRF), modulating the expression of cardiac genes and stress-induced cardiac hypertrophy</td>
</tr>
<tr>
<td>MCUB</td>
<td>Mitochondrial calcium uniporter</td>
<td>Role in cardiac homeostasis, response to cardiac stress including ischemia</td>
</tr>
<tr>
<td>PIGX</td>
<td>Phosphatidylinositol glycan anchor biosynthesis class X</td>
<td>Associated with hypertension. Forms a complex with PIG-M that regulates several processes, including maintenance of arterial blood pressure. Acts via apelin receptor signalling.</td>
</tr>
<tr>
<td>PRKCB</td>
<td>Protein kinase C beta form, protein kinase</td>
<td>Variety of different cellular functions, including endothelial cell proliferation and insulin signaling</td>
</tr>
<tr>
<td>PLS3</td>
<td>Plastin-3</td>
<td>Induced by angiotensin II in endothelial cells to promote cell migration</td>
</tr>
<tr>
<td>RAP1GAP</td>
<td>GTPase activator for RAP-1A</td>
<td>Regulates platelet granule secretion and aggregation</td>
</tr>
<tr>
<td>RBM47</td>
<td>RNA binding motif protein</td>
<td>Missense variant is associated with hypertension</td>
</tr>
<tr>
<td>SEMASA</td>
<td>Semaphorin 5A</td>
<td>High Semasa5A expression increased endothelial cell proliferation and angiogenesis and decreased apoptosis</td>
</tr>
<tr>
<td>SLIT3</td>
<td>Slit guidance ligand 3</td>
<td>Upregulation associated with endothelial cell dysfunction in pulmonary hypertension</td>
</tr>
<tr>
<td>SORCS2</td>
<td>Sortilin-related VPS10 domain-containing receptor 2 (Sortilin2), oxidative stress response gene</td>
<td>SorCS2 releases endostatin, an endogenous inhibitor of endothelial cell proliferation and angiogenesis</td>
</tr>
<tr>
<td>SRD5A1</td>
<td>Steroid 5 alpha-reductase 1, catalyzes the production of androgen dihydrotestosterone</td>
<td>Androgens have been linked to peripheral artery disease and macrophage modulation</td>
</tr>
</tbody>
</table>

### Table 6: Severe COVID-19 risk-associated genes that directly interact with the Wnt/β-Catenin signalling pathway

<table>
<thead>
<tr>
<th>GENE</th>
<th>FUNCTION</th>
<th>MECHANISM OF ACTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>HOPX</td>
<td>Hop homeodomain protein, cofactor</td>
<td>Promotes BMP-mediated inhibition of Wnt signaling pathway</td>
</tr>
<tr>
<td>KRAS</td>
<td>KRAS proto-oncogene, GTPase</td>
<td>Activates the Wnt/β-catenin pathway</td>
</tr>
<tr>
<td>MACROD2</td>
<td>Mono-ADP ribosylhydrolase</td>
<td>Suppresses GSK-3β/β-catenin signaling</td>
</tr>
<tr>
<td>PCDH17</td>
<td>Protocadherin 17, member of the cadherin superfamily</td>
<td>Acts as a tumour suppressor in breast cancer, inhibiting the Wnt signaling pathway</td>
</tr>
<tr>
<td>PRKCB</td>
<td>Protein kinase C beta form, protein kinase</td>
<td>Promotes the phosphorylation of β-catenin</td>
</tr>
<tr>
<td>PTPRK</td>
<td>Protein tyrosine phosphatase (PTP) receptor, type K</td>
<td>Redistributes and inhibits the transcriptional activity of β-catenin</td>
</tr>
<tr>
<td>RRM2</td>
<td>Ribonucleotide reductase regulatory subunit M2</td>
<td>Downregulation of RRM2 suppresses the activity of β-catenin and the Wnt signaling pathway</td>
</tr>
<tr>
<td>SLC9A9</td>
<td>Solute carrier family 9</td>
<td>Upregulates beta-catenin</td>
</tr>
<tr>
<td>SLIT3</td>
<td>Slit guidance ligand 3</td>
<td>Suppresses GSK3β/β-catenin pathway</td>
</tr>
</tbody>
</table>
Finally, our analysis identified four genes that have previously been associated with increased risk of developing Alzheimer’s disease (AD), including \textit{MAPT}, the key gene underlying tau pathology (Table 7).

**Table 7:** Severe COVID-19 risk-associated genes that have also been found to confer increased Alzheimer’s disease risk

<table>
<thead>
<tr>
<th>GENE</th>
<th>FUNCTION</th>
<th>MECHANISM OF ACTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{ATXN1}</td>
<td>Ataxin-1, unknown function</td>
<td>Loss of ataxin-1 causes increased BACE1 expression and amyloid precursor protein (APP) cleavage</td>
</tr>
<tr>
<td>\textit{MAPT}</td>
<td>Microtubule-associated protein tau, stabilizes microtubules</td>
<td>Tau pathology is a hallmark of AD and other neurodegeneration diseases</td>
</tr>
<tr>
<td>\textit{SORCS2}</td>
<td>Sortilin-related VPS10 domain-containing receptor 2, oxidative stress response gene</td>
<td>SNPs are associated with AD, result in altered APP processing</td>
</tr>
<tr>
<td>\textit{STH}</td>
<td>Saitohin, unknown function</td>
<td>Q7R polymorphism increases risk of AD and other neurodegenerative diseases. Interacts with tau.</td>
</tr>
</tbody>
</table>

Out of the total 68 genes, nine are targeted by clinical candidates that have been evaluated in Phase I clinical trials and beyond (see Appendix). These could potentially form the basis of repurposing therapies, after evaluating factors such as safety and pharmacology data. A further eight targets have active chemistry in ChEMBL\textsuperscript{57}, meaning they have active chemical starting points for novel drug development (see Appendix). We can also extend our search for repurposing candidates by looking into known gene interactions to find other more tractable targets in implicated in the same biological pathway.

![Figure 10: Stacked bar plots showing the number of cases with different co-morbidities (cardiovascular disease, hypertension, diabetes, chronic respiratory diseases and Alzheimer’s disease) associated with the severe form of COVID-19 who are affected by the risk-associated genes identified by the \textit{precision}life platform. The line plot shows the total unique number of cases who are affected for each gene.](image-url)
Discussion

Grouping the 68 genes by common biological functions revealed that many are involved in processes that have also been linked to aberrant host responses to COVID-19, such as the pro-inflammatory cytokine storm and immune system dysregulation, as well as lipid droplet formation and endothelial cell dysfunction.

ABO BLOOD-GROUP ASSOCIATION

We found that blood group A was significantly overrepresented in the cases compared to the sepsis-free controls (p=0.037) and all UK Biobank individuals (p=0.096), indicating that it confers a risk factor for developing severe COVID-19. On the other hand, blood group O was found in fewer severe COVID-19 cases (p=0.154) than the controls and the UK Biobank as a whole (p=0.05). This supports the findings that having 0 blood group conferred a protective effect against developing respiratory failure during COVID-1958.

HOST RESPONSE FACTORS AND INFLAMMATION

The host immune response must maintain a balance between effective viral clearance and limiting the immune response to prevent chronic inflammation and collateral tissue damage. In many patients who develop severe COVID-19 reactions and acute respiratory distress syndrome (ARDS) there is evidence of dysregulated cytokine production, resulting in increased levels of pro-inflammatory mediators including IL-1, IL-6, IL-8, CXCL-1059, causing pathological features such as inflammatory cell infiltration, pulmonary edema and sepsis60.

From the number of genes we found relating to immune response pathways and cytokine production cascades, it is clear that patients who develop severe COVID-19 and ARDS may have innate genetic variants that prevent this balance from being struck. These variants were found in equal proportions across all severe COVID-19 patients, regardless of their co-morbidities. This suggests that variants in these genes may not be associated with any specific co-morbidity, but are common amongst many patients who develop severe COVID-19.

HOPX regulates a variety of different cellular processes, including cardiac development and myogenesis37. However, it was reported in a recent COVID-19 study as part of a selection of genes upregulated in expanded CD8+ T cells in patients with mild COVID-1961. These expanded CD8+ effector T cells likely represent SARS-CoV-2-specific T cells, indicating greater efficiency in viral clearance in those patients. However, it is also necessary for Th-1 persistence and resistance to apoptosis, driving chronic inflammation and autoimmune mechanisms62. This exemplifies the delicate balance between establishing an effective host immune response necessary for viral clearance and preventing the development of chronic inflammation that results in collateral tissue damage. It seems likely that patients who develop severe COVID-19 may have an inherent imbalance in these factors.

Similarly, we identified SNP variants in ITK that were highly associated with severe COVID-19. High ITK expression is associated with Th2-driven diseases such as allergic asthma, causing pro-inflammatory cytokine production, eosinophil infiltration and production of mucus63. A range of potent and selective inhibitors (both small molecule and biologic) of ITK have been developed by several pharmaceutical companies64,65. No selective ITK inhibitors have progressed beyond preclinical testing to date, although ibrutinib – a joint ITK and BTK inhibitor – is currently licensed for use in B cell malignancies and as also demonstrated efficacy in the treatment of leishmaniasis. Data collected from several clinical trials indicate that ibrutinib is reasonably well-tolerated by patients66. Inhibition of ITK diminishes lung injury, cytokine production and inflammation in a mouse model of asthma67. Many of the pro-inflammatory cytokines, such as IFNγ, IL-2, and IL-17, blocked by ITK inhibition are raised in patients with ARDS68,69. Furthermore, use of a selective ITK inhibitor blocked HIV cell entry, transcription and particle formation, effectively reducing viral replication70.

A study has found that MLKL is implicated in necrotic cell death following infection with a neuroviral human coronavirus (HCoV)71. Whilst MLKL-induced necrosis may be useful in limiting viral replication, increased necroptosis can also lead to increased inflammation and tissue damage. SNP variants in MLKL could be predisposing patients to severe COVID-19 in one of two ways; with under-functioning protein resulting in poor viral clearance, or through over-expression and over-activation of the necroptotic inflammatory response resulting in organ damage. The SNP variant identified in MLKL was found in the highest scoring 20% of the significant genes in the severe COVID-19 dataset and is present in 251 cases. It is not highly associated with any of the most common co-morbidities found in this dataset. Necrosulfonamide (NSA) is a specific inhibitor of MLKL that potently suppresses necroptosis72. Inhibition of this pathway using NSA decreases pro-inflammatory cytokines such as IL-1β, IL-6 and IL-17A in a way that was protective in a model of psoriasis. High levels of IL-1β and IL-6 have both been observed in cases with severe COVID-1973. Unfortunately, NSA has only ever been evaluated in preclinical trials, so there is no safety or toxicology data in humans available.
NRROS (LRRC3) is an inhibitor of multiple toll-like receptors (TLR) and subsequent NF-κB signalling, acting as a brake on pro-inflammatory cytokine production. NRROS also limits the amount of reactive oxygen species (ROS) produced by phagocytes during the innate immune response, thereby limiting tissue damage caused whilst defending against invading pathogens. It could therefore be hypothesized that SNP variants in this gene, limiting its activity, could make individuals more susceptible to COVID-19-induced inflammatory damage. This is supported by the evidence that Lrrc33-/- mice suffered from increased organ damage as a result of greater pro-inflammatory signalling when challenged with LPS. There are currently no specific small molecule activators of NRROS in publicly-available databases, however a number of chemical agents have been shown to increase the expression of both the protein and mRNA of this gene which may provide chemical starting points for novel drug discovery.

RAB3C may be driving inflammation by inducing the release of IL-6 and activation of the IL-6/JAK2/STAT3 pathway, driving the production of pro-inflammatory cytokines. Ruxolitinib – a JAK2 inhibitor – has been shown to mitigate the effects of RAB3C-induced IL-6 release. Ruxolitinib is currently being trialled as a treatment for respiratory distress caused by SARS-CoV-2.

VASCULAR INFLAMMATION, AUTOPHAGY AND CARDIOVASCULAR DYSFUNCTION

Endothelial dysfunction and vascular inflammation, resulting in neutrophil infiltration, endothelial cell apoptosis and tissue edema, is seen in patients who develop ARDS from SARS-CoV-2 infection. Many of the common co-morbidities associated with severe COVID-19 development, such as diabetes, hypertension and cardiovascular disease, are already associated with vascular inflammation and endothelial cell dysfunction pathological features. It is therefore unsurprising that these patients are at higher risk of developing severe COVID-19.

Many of the risk-associated SNPs that we found that were related to these pathological pathways were found in severe COVID-19 patients who had at least one of these co-morbidities. However, we controlled for this effect by co-morbidity-matching these cases against the sepsis-free controls, meaning that these SNPs are not likely to be just be artefacts of the co-associated co-morbidity, but a differentiating factor in the development of severe COVID-19 and ARDS.

Helios (IZKF2) expression is used as a marker for regulatory T cells (Tregs), and therefore has an important role in self-tolerance and regulating autoimmunity. Decreased levels of Helios+ Tregs have been observed in patients with hypertension and rheumatoid arthritis, with lower expression levels correlating with increased inflammatory markers. Tregs may protect against hypertension by limiting vascular inflammation through suppression of effector T cells. Furthermore, our analysis revealed that variants in IZKF2 were disproportionately co-associated with patients with hypertension (Figure 10). This adds further evidence to the association between vascular inflammation, severe COVID-19 and cardiovascular hypertensive co-morbidities.

MCUB encodes one of the pore-forming subunits of the mitochondrial Ca²⁺ uniporter (MCU). MCUB is a necessary part of a protective response against mitochondrial Ca²⁺ overload during cardiac injury and ischemia. Mcub-/- mice displayed increased cardiac remodelling and ischemic injury as a result of increased mitochondrial Ca²⁺ uptake. Although more research is required to fully understand the role of MCUB in severe COVID-19 pathogenesis, a decreased level of serum calcium has been suggested as a biomarker for increased COVID-19 severity and ARDS.

PKCβ (PRKCB) is a serine/threonine protein kinase that is activated by calcium. It has a range of functions, including B cell activation, endothelial cell proliferation and activation of apoptosis. However, PKCβ has also been linked to a number of different vascular diseases, including atherosclerosis, diabetes and hypertension. We find that in our case population, the risk-associated SNP found in PRKCB was found in 165 severe COVID-19 cases (penetrance = 20.2%), and was present in 45% of patients with cardiovascular disease and 51% of patients with hypertension. High levels of PKCβ result in increased vascular inflammation, endothelial dysfunction and oxidative stress, all of which have been found in patients with severe COVID-19. PKCβ also drives the accumulation of cholesterol in macrophages, leading to foam cell development and macrophage dysregulation. Therefore, inhibition of PKCβ could help to reverse some of the vascular-related pathology, contributing to sepsis development and multi-organ failure, that is seen in severe COVID-19 patients. Ruboxistaurin (LY333531) is a selective PKCβ inhibitor that has been trialled as an anti-diabetic drug to reduce vascular and retinal complications that certain diabetic patients develop. It has reasonable pharmacokinetic and toxicology data, as it can be orally administered and is well-tolerated by patients. However, it was discontinued and has failed to progress beyond Phase III clinical trials.

We also found genes, including CIDEA and PLIN4, that protect against insulin resistance (IR). IR results in increased cardiovascular inflammation and oxidative stress, contributing to atherosclerotic plaque formation and hypertension. High Cidea expression decreases circulating fatty acid levels by increasing the level of triglycerides stored in lipid droplets. This helps to protect against insulin resistance. The anti-diabetic drug rosiglitazone may have insulin sensitising effects by increasing the expression of both Cidea and perilipins. Our analysis identified CIDEA and a perilipin, PLIN4, as being highly associated with severe COVID-19 patients. SNPs in CIDEA were found in one of the highest proportions in severe COVID-19 patients, particularly in
patients with cardiovascular and hypertensive diseases (Figure 10). This indicates that insulin resistance is a contributing factor to vascular inflammatory pathologies that may predispose patients to developing ARDS in response to SARS-CoV-2.

Finally, variants in **GFRα1** are within the highest 10% of genes most associated with hypertensive patients in our case population. GFRα1, encoded by **GFRα1**, plays a key role in glial cell line-derived neurotrophic factor (GDNF)-mediated signalling. **GFRα1** induces autophagy via activation of the RET/AMPK signalling pathway independently from its interaction with GDNF. In the early stages of sepsis, autophagy is a protective mechanism employed by cells to remove damaged proteins, reduce mitochondrial dysfunction-induced inflammation and eliminate pathogens. Furthermore, inhibition of autophagy in mouse models of sepsis results in increased mortality. Selective agonists of GFRα1 are under development, which may help to protect against sepsis-induced tissue damage in the early stages of the disease.

**LIPID DROPLET BIOLOGY**

There is increasing evidence that lipid dysfunction plays a role in COVID-19 pathogenesis. Lipid droplets (LD) play a key role in viral pathogenesis, as intracellular lipid stores are crucial for viral replication and assembly. LD proteins have been associated with particle load and pathogenicity in other virus strains, including hepatitis C, dengue and rotavirus. It has also been shown that coronaviruses use host cellular lipid machinery during replication.

We identified six genes associated with lipids and adipose tissue, with three genes directly involved in lipid droplet formation (Table 4). A different member of the perilipin family (PLIN3) is used by hepatitis C virus (HCV) for steatosis development and viral assembly, with PLIN3 inhibition resulting in decreased viral particle release from host cells. Promethin (TMEM159) has also been recently identified as lipid droplet organiser (LDO) in partnership with seipin. Although promethin has not been studied in the context of viral disease, overexpression of its co-associated protein seipin results in significantly reduced viral particle secretion and infectivity in a model of hepatitis C.

**WNT/β-CATENIN SIGNALLING PATHWAY**

Our analysis identified nine genes that regulate either the Wnt or β-catenin components of the Wnt/β-catenin pathway.

There is a long-established role of the Wnt gene-family and the associated signally pathway in embryogenesis and therefore its relationship to developmental disorders and carcinogenesis. However, recent work shows an emerging and complex picture of the role of Wnt genes in host cell defence mechanisms, the modulation of inflammatory cytokine production and connections between innate and adaptive immune systems. The role of these ligands varies within the family members and their relation to β-catenin.

Wnt/β-catenin signalling has both anti- and pro-inflammatory effects in different contexts, dependent on its interaction with NF-κB. In several studies, activation of β-catenin was shown to inhibit IL-1β and the subsequent production of IL-6 and matrix metalloproteinases (MMPs), resulting in an anti-inflammatory effect. Although its role in sepsis pathogenesis is yet to be fully elucidated, patients with severe sepsis and sepsis-driven ARDS had increased levels of WntSA in lung tissue and serum.

However, in bronchial epithelial cells, inhibition of beta-catenin’s activation of NF-κB resulted in fewer proinflammatory cytokines and cell injury. In a previous unpublished analysis using the same combinatorial approach on a Sjögren’s Syndrome cohort found in the UK Biobank, we identified a significant number of genes involved in the Wnt/β-catenin pathway. As with severe COVID-19, patients with Sjögren’s syndrome also present with elevated cytokines and inflammation-driving pathology.

The Wnt/β-catenin pathway has also been implicated in promoting viral replication. In a model of influenza (H1N1), activation of this pathway resulted in higher viral replication and inhibition of Wnt/β-catenin signalling using iCRT14 reduced viral production and improved clinical symptoms. It has been demonstrated that Rift Valley Virus exploits the proliferative cell state that activation of Wnt signalling promotes to enable viral replication through easier trafficking of viral proteins within the cell. It has also been shown that coronaviruses use this mechanism in proliferative cells. Therefore, inhibition of the Wnt/β-catenin pathway may help limit SARS-CoV-2 replication and reduce viral load in this way. This theory is supported by the fact that niclosamide – an anti-helminthic drug – limits coronavirus replication in a model of SARS. The study did not investigate how this drug inhibited viral replication, but a subsequent study has revealed that niclosamide is an inhibitor of the Wnt signalling pathway. Due to the significant number of genes found to regulate this pathway in our analysis, we believe that modulation of this pathway could be of benefit to a large number of patients who develop severe COVID-19.

**CALCIUM SIGNALLING PATHWAY**

Several of the processes mentioned above, including lipid programming, beta-catenin and protein kinase C signalling, converge in a central pathway involved in plasma membrane repair, clotting and wound healing. This pathway is largely driven by Ca²⁺

---

*Note: The preprint details, including the DOI and licensing information, are provided for context and are not part of the natural text content.*
activation, which is a known serum biomarker associated with severe COVID-19 and ARDS. Of the 68 risk-associated genes found in our analysis, at least 16 of them having calcium-binding domains or are dependent on Ca^{2+} signalling.

Calcium (Ca^{2+}) drives wound constriction and clotting via F-actin. Beta-catenin binds to F-actin and plays a role in endothelial barrier function. Calcium also plays a role in lipid patterning, altering the plasma membrane composition in response to cell injury. Finally, activation of diacylglycerol (DAG and protein kinase c (including the β isoform) results in vesicle replenishment and wound repair.

The pathological observations seen in severe COVID-19 of thrombocytopenia, leaky blood vessels and multi-nucleated fusion cells, support the hypothesis that these Ca^{2+}/lipid/F-actin pathways may converge into a pathological process that drives life-threatening reactions to COVID-19 such as sepsis and ARDS.

Furthermore, the gene cluster at 3p21.31 identified as a susceptibility locus for respiratory failure in COVID-19 contains several genes that are involved in calcium ion signalling; SLC6A29 is a regulator of calcium-dependent amino acid uptake, and CXCR9 and CCR9 act as a signalling molecules by increasing the level of intracellular Ca^{2+}. This suggests that aberrant calcium ion signalling may be responsible for driving severe COVID-19 responses in patients with variants in genes that regulate the expression and activity of this ion.

We intend to perform further analyses to confirm this hypothesis.

**NEURODEGENERATIVE DISEASE ASSOCIATIONS**

There is emerging evidence that key genes associated with Alzheimer’s Disease (AD) risk such as APOE4 are associated with increased risk of severe COVID-19 disease, although the reasons for this link are still unclear. Our analysis identified a strong conserved signal for four genes that also confer greater risk of developing AD, including MAPT.

We also found three additional genes that were significant in a previous neurodegenerative disease study we performed. This unpublished study identified several gene variants in which were highly associated in cases with sporadic amyotrophic lateral sclerosis (ALS). We have also shown that increasing the activity of two of these genes in a mixed cell neuronal assay containing SOD1 reactive astrocytes significantly improves motor neuron survival. This could indicate shared biological pathways that drive both severe COVID-19 and neurodegeneration, potentially through pro-inflammatory, neurotoxic mechanisms. However, the SNP variants found in these genes were not highly enriched in the few severe COVID-19 patients who had been diagnosed with AD. This is likely due to the relatively low numbers of cases diagnosed with AD in UK Biobank as a whole.

**Conclusion**

We performed this analysis using only genetic data from patients found in the UK Biobank, identifying 68 protein coding genes that are highly associated with the development of severe COVID-19. These targets would not have been found using standard analytical approaches such as GWAS on the same population. In addition to this, we have identified 29 drugs and clinical candidates and a further eight targets with chemical starting points that could be used in the development of treatment strategies that improve clinical outcomes in severe COVID-19 patients.

It appears that the variants we found in genes relating to immune response pathways and cytokine production cascades were in equal proportions across all severe COVID-19 patients, regardless of their co-morbidities. This suggests that such variants are not associated with any specific co-morbidity, but are common amongst patients who develop severe COVID-19. In contrast there were small deviations in the penetrance of some of the SNPs identified in different co-morbidity cohorts. While the pattern of these deviations were similar across cardiovascular, diabetes and hypertension co-morbidity cohorts, the pattern across the respiratory co-morbidity cohort was somewhat different. These differences, while suggestive of the need for further study, were not yet significant enough to be reported in detail.

As more test records and additional medical data become available in the UK Biobank and other data sources, we will be able to fully ascertain the severe and mild COVID-19 cases and provide an additional layer of validation to the results from this study.

One limitation of the UK Biobank dataset is that the ethnicity distribution of the participants is heavily skewed to white British participants and it has consequently not been possible to fully investigate additional risk factors in BAME patients. We are actively seeking to investigate severe COVID-19 risk factors in other datasets with more ethnically diverse populations. The addition of more phenotypic and clinical data in our analyses may also be used to gain greater understanding into the association of other observed epidemiological risk factors such as ethnicity, socioeconomic status and prescription medication history with development of severe COVID-19.
Acknowledgements

We would like to acknowledge UK Biobank for providing us access to the data under Application ID 44288 and Bugbank for linking infection data from Public Health England to improve the study of infection in the UK Biobank cohort.
References


49. Lemieux E, Cagnol S, Beaudry K, Carrier J, Rivard N. Oncogenic KRAS signalling promotes the Wnt/β-catenin pathway through LRP6 in colorectal cancer. Oncogene. 2015;34(38):4914-4927. doi:10.1038/onc.2014.416


88. Huo J, Lu S, Kwong JQ et al. MCub Induction Protects the Heart from Post-Ischemic Remodelling. Circ Res. 2020;0. https://doi.org/10.1161/CIRCRESAHA.119.316369

89. Sun JK, Zhang WH, Zou L. Serum calcium as a biomarker of clinical severity and prognosis in patients with coronavirus disease 2019: a retrospective cross-sectional study. 2020;0. Critical Care & Emergency Medicine. doi: 10.21203/rs.3.rs.17575/v1


Appendix

SEPSIS CONTROL CRITERIA

1. Controls to exclude any patients with the following ICD-10 codes:

<table>
<thead>
<tr>
<th>ICD-10</th>
<th>Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>A02.1</td>
<td>Salmonella septicemia</td>
</tr>
<tr>
<td>A22.7</td>
<td>Anthrax septicemia</td>
</tr>
<tr>
<td>A40.x</td>
<td>Streptococcal septicemia</td>
</tr>
<tr>
<td>A41.x</td>
<td>Other septicemia</td>
</tr>
<tr>
<td>B37.7</td>
<td>Candidal septicemia</td>
</tr>
<tr>
<td>O35</td>
<td>Puerperal sepsis</td>
</tr>
<tr>
<td>R57.2</td>
<td>Septic shock</td>
</tr>
</tbody>
</table>

2. Controls to include at least one of the following ICD-10 codes:

<table>
<thead>
<tr>
<th>ICD-10</th>
<th>Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>J12-J18</td>
<td>Pneumonia</td>
</tr>
<tr>
<td>J20-J22</td>
<td>Lower respiratory infection</td>
</tr>
<tr>
<td>B95</td>
<td>Streptococcus</td>
</tr>
<tr>
<td>B95</td>
<td>Staphylococcus</td>
</tr>
</tbody>
</table>

3. Controls to include least one of the following ICD-10 codes:

<table>
<thead>
<tr>
<th>ICD-10</th>
<th>Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>E10-E14</td>
<td>Diabetes</td>
</tr>
<tr>
<td>N00-N19</td>
<td>Kidney Disease</td>
</tr>
<tr>
<td>K70-K77</td>
<td>Liver Disease</td>
</tr>
<tr>
<td>J40-J44</td>
<td>COPD</td>
</tr>
</tbody>
</table>

Table 8: Pairwise comparison of co-morbidities prevalent in severe COVID-19 patient population

<table>
<thead>
<tr>
<th></th>
<th>All Severe COVID-19</th>
<th>Cardiovascular Disease</th>
<th>Hypertension</th>
<th>Respiratory Disease</th>
<th>Diabetes</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Severe COVID-19</td>
<td>779</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cardiovascular Disease</td>
<td>136</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypertension</td>
<td>385</td>
<td>115</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Respiratory Disease</td>
<td>170</td>
<td>39</td>
<td>105</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diabetes</td>
<td>143</td>
<td>55</td>
<td>117</td>
<td>38</td>
<td></td>
</tr>
<tr>
<td>Alzheimer's Disease</td>
<td>10</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

IDENTIFICATION OF CASES WITH SPECIFIC CO-MORBIDITIES

Cases with co-morbidities were identified using the following disease codes:

<table>
<thead>
<tr>
<th>Disease</th>
<th>ICD-10</th>
<th>Self-Reported codes</th>
<th>Operation codes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cardiovascular disease</td>
<td>I20-I25</td>
<td>1070, 1075, 1095</td>
<td>K40-41, K45, K49, K50.2, K75</td>
</tr>
<tr>
<td>Hypertension</td>
<td>I10, I11, R03</td>
<td>1065</td>
<td></td>
</tr>
<tr>
<td>Diabetes</td>
<td>E10-14</td>
<td>1220, 1222, 1223</td>
<td></td>
</tr>
<tr>
<td>Chronic respiratory diseases</td>
<td>J40-47</td>
<td>1111, 1112</td>
<td></td>
</tr>
<tr>
<td>Alzheimer's disease</td>
<td>G30</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 9: Table of severe COVID-19-associated genes with chromosomal location and RF score (genes in bold have been validated in further studies)

<table>
<thead>
<tr>
<th>GENE</th>
<th>CHROMOSOME</th>
<th>RF SCORE</th>
</tr>
</thead>
<tbody>
<tr>
<td>RPL7AP57</td>
<td>11</td>
<td>7.80</td>
</tr>
<tr>
<td>TMEM159</td>
<td>16</td>
<td>2.41</td>
</tr>
<tr>
<td>CDC7</td>
<td>1</td>
<td>1.83</td>
</tr>
<tr>
<td>REX1BD</td>
<td>19</td>
<td>1.26</td>
</tr>
<tr>
<td>PSMC1</td>
<td>14</td>
<td>1.16</td>
</tr>
<tr>
<td>STH</td>
<td>17</td>
<td>1.10</td>
</tr>
<tr>
<td>KRAS</td>
<td>12</td>
<td>0.95</td>
</tr>
<tr>
<td>CIDEA</td>
<td>18</td>
<td>0.82</td>
</tr>
<tr>
<td>PLS3</td>
<td>X</td>
<td>0.74</td>
</tr>
<tr>
<td>COQ6</td>
<td>14</td>
<td>0.72</td>
</tr>
<tr>
<td>PLIN4</td>
<td>19</td>
<td>0.67</td>
</tr>
<tr>
<td>HOPX</td>
<td>4</td>
<td>0.67</td>
</tr>
<tr>
<td>PCDH17</td>
<td>13</td>
<td>0.67</td>
</tr>
<tr>
<td>SRD5A1</td>
<td>5</td>
<td>0.61</td>
</tr>
<tr>
<td>MLKL</td>
<td>16</td>
<td>0.57</td>
</tr>
<tr>
<td>TBC1D2</td>
<td>9</td>
<td>0.53</td>
</tr>
<tr>
<td>MEP1B</td>
<td>18</td>
<td>0.49</td>
</tr>
<tr>
<td>B3GLCT</td>
<td>13</td>
<td>0.48</td>
</tr>
<tr>
<td>CRLF1</td>
<td>19</td>
<td>0.46</td>
</tr>
<tr>
<td>MCUB</td>
<td>4</td>
<td>0.45</td>
</tr>
<tr>
<td>SAMD3</td>
<td>6</td>
<td>0.44</td>
</tr>
<tr>
<td>ENTPD5</td>
<td>14</td>
<td>0.42</td>
</tr>
<tr>
<td>LINC02210-CHR1</td>
<td>17</td>
<td>0.39</td>
</tr>
<tr>
<td>L3MBTL3</td>
<td>6</td>
<td>0.36</td>
</tr>
<tr>
<td>PEX14</td>
<td>1</td>
<td>0.36</td>
</tr>
<tr>
<td>SPEF2</td>
<td>5</td>
<td>0.35</td>
</tr>
<tr>
<td>SLC16A10</td>
<td>6</td>
<td>0.35</td>
</tr>
<tr>
<td>LINC02210</td>
<td>17</td>
<td>0.34</td>
</tr>
<tr>
<td>ERICH1</td>
<td>8</td>
<td>0.29</td>
</tr>
<tr>
<td>MAPT</td>
<td>17</td>
<td>0.28</td>
</tr>
<tr>
<td>SMCHD1</td>
<td>18</td>
<td>0.26</td>
</tr>
<tr>
<td>RRM2</td>
<td>2</td>
<td>0.26</td>
</tr>
<tr>
<td>NRD2E2</td>
<td>14</td>
<td>0.25</td>
</tr>
<tr>
<td>ANTXR1</td>
<td>2</td>
<td>0.25</td>
</tr>
<tr>
<td>STAC</td>
<td>3</td>
<td>0.23</td>
</tr>
<tr>
<td>ITK</td>
<td>5</td>
<td>0.22</td>
</tr>
<tr>
<td>ZNF106</td>
<td>15</td>
<td>0.21</td>
</tr>
<tr>
<td>DNAH2</td>
<td>17</td>
<td>0.20</td>
</tr>
<tr>
<td>CTR9</td>
<td>11</td>
<td>0.19</td>
</tr>
<tr>
<td>NRROS</td>
<td>3</td>
<td>0.19</td>
</tr>
<tr>
<td>RBM47</td>
<td>4</td>
<td>0.18</td>
</tr>
<tr>
<td>RAB3C</td>
<td>5</td>
<td>0.17</td>
</tr>
<tr>
<td>GFA1</td>
<td>10</td>
<td>0.15</td>
</tr>
<tr>
<td>KANSL1</td>
<td>17</td>
<td>0.14</td>
</tr>
<tr>
<td>C9orf92</td>
<td>9</td>
<td>0.14</td>
</tr>
<tr>
<td>I2C2F2</td>
<td>2</td>
<td>0.13</td>
</tr>
<tr>
<td>XR6</td>
<td>8</td>
<td>0.11</td>
</tr>
<tr>
<td>EIF3E</td>
<td>8</td>
<td>0.11</td>
</tr>
<tr>
<td>MAST4</td>
<td>5</td>
<td>0.10</td>
</tr>
<tr>
<td>PIQX</td>
<td>3</td>
<td>0.10</td>
</tr>
<tr>
<td>GAREM1</td>
<td>18</td>
<td>0.10</td>
</tr>
<tr>
<td>IQCM</td>
<td>4</td>
<td>0.09</td>
</tr>
<tr>
<td>RAP1GAP2</td>
<td>17</td>
<td>0.09</td>
</tr>
<tr>
<td>PRKCB</td>
<td>16</td>
<td>0.09</td>
</tr>
<tr>
<td>AACSP1</td>
<td>5</td>
<td>0.09</td>
</tr>
<tr>
<td>ATXN1</td>
<td>6</td>
<td>0.08</td>
</tr>
<tr>
<td>VTI1A</td>
<td>10</td>
<td>0.08</td>
</tr>
<tr>
<td>C4orf5S0</td>
<td>4</td>
<td>0.07</td>
</tr>
<tr>
<td>SEMA5A</td>
<td>5</td>
<td>0.07</td>
</tr>
<tr>
<td>SLC9A9</td>
<td>3</td>
<td>0.07</td>
</tr>
<tr>
<td>SPPL2C</td>
<td>17</td>
<td>0.07</td>
</tr>
<tr>
<td>PTPTK</td>
<td>6</td>
<td>0.06</td>
</tr>
<tr>
<td>ATRN1L</td>
<td>10</td>
<td>0.06</td>
</tr>
<tr>
<td>LHFPL2</td>
<td>5</td>
<td>0.05</td>
</tr>
<tr>
<td>PRUNE2</td>
<td>9</td>
<td>0.05</td>
</tr>
<tr>
<td>CNTNS</td>
<td>11</td>
<td>0.05</td>
</tr>
<tr>
<td>SLIT3</td>
<td>5</td>
<td>0.05</td>
</tr>
<tr>
<td>SORCS2</td>
<td>4</td>
<td>0.04</td>
</tr>
<tr>
<td>KNCNP4</td>
<td>4</td>
<td>0.03</td>
</tr>
<tr>
<td>MACROD2</td>
<td>20</td>
<td>0.02</td>
</tr>
<tr>
<td>CNTNAP2</td>
<td>7</td>
<td>0.02</td>
</tr>
</tbody>
</table>

Table 10: Table showing drug repurposing candidates for 10 target genes identified as being COVID-19 related

<table>
<thead>
<tr>
<th>GENE</th>
<th>COMPOUND CHEMBL ID</th>
<th>DRUG</th>
<th>PHASE</th>
<th>MOLECULE TYPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDC7</td>
<td>CHEMBL3544943</td>
<td>bms-863233</td>
<td>2</td>
<td>Small molecule</td>
</tr>
<tr>
<td>CDC7</td>
<td>CHEMBL3545090</td>
<td>rxdx-103</td>
<td>1</td>
<td>Small molecule</td>
</tr>
<tr>
<td>CDC7</td>
<td>CHEMBL3545321</td>
<td>nms-1116354</td>
<td>1</td>
<td>Small molecule</td>
</tr>
<tr>
<td>PSMC1</td>
<td>CHEMBL325041</td>
<td>bortezomib</td>
<td>4</td>
<td>Small molecule</td>
</tr>
<tr>
<td>PSMC1</td>
<td>CHEMBL451887</td>
<td>carfilzomib</td>
<td>4</td>
<td>Protein</td>
</tr>
<tr>
<td>PSMC1</td>
<td>CHEMBL2103884</td>
<td>oprozomib</td>
<td>1</td>
<td>Small molecule</td>
</tr>
<tr>
<td>PSMC1</td>
<td>CHEMBL3545432</td>
<td>ixazomib citrate</td>
<td>4</td>
<td>Small molecule</td>
</tr>
<tr>
<td>SRD5A1</td>
<td>CHEMBL254328</td>
<td>abiraterone</td>
<td>4</td>
<td>Small molecule</td>
</tr>
<tr>
<td>SRD5A1</td>
<td>CHEMBL1200969</td>
<td>dutasteride</td>
<td>4</td>
<td>Small molecule</td>
</tr>
<tr>
<td>RRM2</td>
<td>CHEMBL467</td>
<td>hydroxyurea</td>
<td>4</td>
<td>Small molecule</td>
</tr>
<tr>
<td>RRM2</td>
<td>CHEMBL1637</td>
<td>gemcitabine hydrochloride</td>
<td>4</td>
<td>Small molecule</td>
</tr>
<tr>
<td>RRM2</td>
<td>CHEMBL1750</td>
<td>clofarabine</td>
<td>4</td>
<td>Small molecule</td>
</tr>
<tr>
<td>RRM2</td>
<td>CHEMBL1096882</td>
<td>fludarabine phosphate</td>
<td>4</td>
<td>Small molecule</td>
</tr>
<tr>
<td>RRM2</td>
<td>CHEMBL1200983</td>
<td>gallium nitrate</td>
<td>4</td>
<td>Small molecule</td>
</tr>
<tr>
<td>RRM2</td>
<td>CHEMBL3544910</td>
<td>motexafin gadolinium</td>
<td>3</td>
<td>Small molecule</td>
</tr>
<tr>
<td>RRM2</td>
<td>CHEMBL3989496</td>
<td>tezacetabine</td>
<td>2</td>
<td>Small molecule</td>
</tr>
<tr>
<td>ITK</td>
<td>CHEMBL1201733</td>
<td>pazopanib hydrochloride</td>
<td>4</td>
<td>Small molecule</td>
</tr>
<tr>
<td>ITK</td>
<td>CHEMBL4085457</td>
<td>pf-06651600</td>
<td>3</td>
<td>Small molecule</td>
</tr>
</tbody>
</table>
Table 11. Table showing 8 target genes identified as being COVID-19 related that have active compounds in ChEMBL

<table>
<thead>
<tr>
<th>GENE</th>
<th>NO. OF ACTIVE COMPOUNDS (ChEMBL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MACROD2</td>
<td>3</td>
</tr>
<tr>
<td>MAST4</td>
<td>6</td>
</tr>
<tr>
<td>MLKL</td>
<td>11</td>
</tr>
<tr>
<td>SLC16A10</td>
<td>70</td>
</tr>
<tr>
<td>MEP1B</td>
<td>90</td>
</tr>
<tr>
<td>KRAS</td>
<td>120</td>
</tr>
<tr>
<td>L3MBTL3</td>
<td>122</td>
</tr>
<tr>
<td>MAPT</td>
<td>9,589</td>
</tr>
</tbody>
</table>

Table 12. Table showing all the UK Biobank COVID-19 datasets and numbers of cases:controls (post-QC) used in this study

<table>
<thead>
<tr>
<th>UK Biobank Data Release</th>
<th>18 May 2020</th>
<th>26 May 2020</th>
<th>6 June 2020</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cases</td>
<td>779</td>
<td>877</td>
<td>929</td>
</tr>
<tr>
<td></td>
<td>(442 males, 337 females)</td>
<td>(492 males, 385 females)</td>
<td>(524 males, 405 females)</td>
</tr>
<tr>
<td>Controls</td>
<td>1,553</td>
<td>5,438</td>
<td>5,563</td>
</tr>
</tbody>
</table>

Table 13. Results of the ABO blood group analysis. Two-sided Fisher’s exact tests were used to calculate blood-group specific odds ratios against the other blood groups for the analyses. P-values <0.05 are shown in bold.

<table>
<thead>
<tr>
<th>Blood group analysis</th>
<th>Severe cases (779) vs Sepsis controls (1553)</th>
<th>Severe cases (779) vs all UK Biobank individuals (488,295)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Odds Ratio</td>
<td>P-value</td>
</tr>
<tr>
<td>A vs AB/B/O</td>
<td>1.201</td>
<td>0.037</td>
</tr>
<tr>
<td>O vs A/B/AB</td>
<td>0.880</td>
<td>0.154</td>
</tr>
<tr>
<td>B vs A/B/O</td>
<td>1.010</td>
<td>1.000</td>
</tr>
<tr>
<td>AB vs A/B/O</td>
<td>0.750</td>
<td>0.296</td>
</tr>
</tbody>
</table>

Table 11. Table showing 8 target genes identified as being COVID-19 related that have active compounds in ChEMBL

Table 12. Table showing all the UK Biobank COVID-19 datasets and numbers of cases:controls (post-QC) used in this study

Table 13. Results of the ABO blood group analysis. Two-sided Fisher’s exact tests were used to calculate blood-group specific odds ratios against the other blood groups for the analyses. P-values <0.05 are shown in bold.