1. Extended Data

Complete the Inventory below for all Extended Data figures.

- Keep Figure Titles to one sentence only
- Upload your files as 'Figure Files' in our Manuscript Tracking system
- File names should include the Figure Number. i.e.: *Smith_ED_Fig1.jpg*
- Please be sure to include the file extension in the Filename. Note that Extended Data files must be submitted as .jpg, .tif or .eps files *only*, and should be approximately 10MB
- All Extended Data figure legends must be provided in the Inventory below and should not exceed 300 words each *(if possible)*
- Please include Extended Data *ONLY* in this table

Figure #	Figure title One sentence only	Filename This should be the name the file is saved as when it is uploaded to our system. Please include the file extension. i.e.: *Smith_ED_Fig1.jpg*	Figure Legend If you are citing a reference for the first time in these legends, please include all new references in the main text Methods References section, and carry on the numbering from the main References section of the paper. If your paper does not have a Methods section, include all new references at the end of the main Reference list.
Extended Data Fig. 1	Colocalization of the Genetic Determinants of ABO Plasma Protein Levels and COVID-19 Outcomes	Richards_FigS1.jpg	Colored circles represent the LD of SNPs within 1Mb region around the leading *cis*-pQTL of ABO rs505922. PP: posterior probability of sharing the same genetic signal with ABO level, estimated by coloc.
Extended Data Fig. 2	Colocalization of the expression QTL (eQTL), splicing QTL (sQTL) and protein QTL (pQTL) of OAS1 with COVID-19 outcomes	Richards_FigS2.jpg	(A) R2 was calculated using the rs4767027 SNP in each COVID-19 GWAS for 503 European individuals in the 1000 Genomes Project. (B) A table illustrating the pairwise colocalization results between OAS1 expression (eQTL), splicing (sQTL), protein (pQTL) and COVID-19 outcomes. Colocalization analyses were performed using GWAS summary from GTEx by restricting to the regions within 1 Mb of rs4767027.
Extended Data	Colocalization of	Richards_FigS3.	R2 was calculated using the SNP with the

Fig. 3	the expression QTL (eQTL) of OAS3 and COVID-19 outcomes	jpg	lowest p-value in each COVID-19 GWAS for 503 European individuals in 1000 Genomes Project. Colocalization analysis was performed using GWAS summary from GTEx by restricting to the region within 1 Mb of rs4767027.
Extended Data Fig. 4	OAS1 Level Trajectory with Days Since Symptom Onset in different COVID-19 Outcomes	Richards_FigS4.jpg	Trajectory plot showing OAS1 level (in normalized RFUs) of 38 COVID-19 patients with their blood samples collected at four different time points, represented by days since symptom onset. Each line represents one patient. Blue shows COVID-19 patient without very severe outcomes and red shows COVID-19 patient with very severe outcome.
Extended Data Fig. 5	Correlation of log OAS1 level with Processing Time in Samples from a Non-infectious State and During Infection	Richards_FigS5.jpg	Pearson correlation (two-sided) of sample processing time and log OAS1 level of the two groups (During infection and Non-infectious state) after QC. The regression line was added for each of the two groups separately, with shaded parts showing the 95% confidence interval.
Extended Data Fig. 6	Distribution of log transformed of OAS1 in samples during non-infectious state and during infection	Richards_FigS6.jpg	Figure includes samples in each group before QC. X-axis is the natural log transformed OAS1 level. Showing 10 Quantiles of the log OAS1 level of each group.
Extended Data Fig. 7	OAS1 level association with Age and Sex in samples during non-infectious state	Richards_FigS7.jpg	Pearson correlation (two-sided) of age and log OAS1 level after the removal of outliers of OAS1 level and extended sample processing time, stratified by sex. The regression line was added for males and females separately, with shaded parts showing the 95% confidence interval.
Extended Data Fig. 8	Distribution of Days from Symptom Onset and Sample Processing Time	Richards_FigS8.jpg	A: Distribution of Days from Symptom Onset in all 934 samples from COVID-19 patients with SomaScan® measurement. B. Distribution of Sample Processing Time in

	in BQC-19 Samples		all 1039 samples with SomaScan® measurement. C. Distribution of Days from Symptom Onset in 421 samples from COVID-19 patients included in the analyses after QC. D. Distribution of Sample Processing Time in 524 samples included in the analyses after QC.
Extended Data Fig. 9	Violin plot of OAS1 Level Measured by SomaScan® Technology in BQC19 cohort in Groups of Different COVID-19 Outcomes, in samples taken from a Non-infectious State and During Infection	Richards_FigS9.jpg	A violin plot with box plot showing the distribution and median value of natural log transformed OAS1 level (Y-axis) in each group (During infection and Non-infectious state) with each outcome (Very severe COVID-19, Hospitalization Due to COVID-19 and COVID-19 Susceptibility). The lower and upper hinges of the boxplot correspond to the first and third quartiles. The lower and upper whiskers of the boxplot extend from the hinge to the smallest and largest value at most 1.5 * IQR of the hinge. The sample size included in each group can be found in Figure 4 and Extended Data 10. Controls of each group include COVID negative individuals.

Delete rows as needed to accommodate the number of figures (10 is the maximum allowed).

2. Supplementary Information:

A. Flat Files

Complete the Inventory below for all additional textual information and any additional Supplementary Figures, which should be supplied in one combined PDF file.

- **Row 1:** A combined, flat PDF containing any Supplementary Text, Discussion, Notes, Additional Supplementary Figures, Supplementary Protocols, simple tables, and all associated legends. Only one such file is permitted.

- **Row 2:** Nature Research's Reporting Summary; if previously requested by the editor, please provide an updated Summary, fully completed, without any mark-ups or comments. **(Reporting Summaries are not required for all manuscripts.)**

Item	Present?	Filename This should be the name the file is saved as when it is uploaded to our system, and should include the file extension. The extension must be .pdf	A brief, numerical description of file contents. i.e.: *Supplementary Figures 1-4, Supplementary Discussion, and Supplementary Tables 1-4.*
Supplementary Information	No		
Reporting Summary	Yes	NMED-A109062A-nr-reporting-summary-NM-0128	

B. Additional Supplementary Files

Complete the Inventory below for all additional Supplementary Files that cannot be submitted as part of the Combined PDF.

- Do not list Supplementary Figures in this table (see section 2A)
- Where possible, include the title and description within the file itself
- Spreadsheet-based tables and data should be combined into a workbook with multiple tabs, not submitted as individual files.
- Compressed files are acceptable where necessary. ZIP files are preferred.
- Please note that the *ONLY* allowable types of additional Supplementary Files are:
  - Supplementary Tables
  - Supplementary Audio
  - Supplementary Videos
  - Supple
  - Supplementary Data, for example: raw NMR Data, Cryo-EM Data, Computational Data, Crystallo

Type	Number If there are multiple files of the same type this should be the numerical indicator. i.e. "1" for Video 1, "2" for Video 2, etc.	Filename This should be the name the file is saved as when it is uploaded to our system, and should include the file extension. i.e.: *Smith_ Supplementary_Video_1.mov*	Legend or Descriptive Caption Describe the contents of the file
Supplementary Table	Supplementary Tables 1-12	Extended_Data_0128.xlsx	Supplementary tables in multiple tabs of the spreadsheet

Editor summary: A variant of OAS1 gene, which encodes an enzyme that is critical for the innate immune response to viral infections, is associated with increased risk of death in patients with COVID-19.

Editor recognition statement: Joao Monteiro was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Reviewer Recognition Statement: Nature Medicine thanks Luis Barriero, Andre Franke, and Adam Butterworth for their contribution to the peer review of this work.

A Neanderthal OAS1 Isoform Protects European Ancestry Individuals Against COVID-19 Susceptibility and Severity

Sirui Zhou[1,2],* Guillaume Butler-Laporte[1,2],* Tomoko Nakanishi[1,3,4,5],* David Morrison[1], Jonathan Afilalo[1,2,6], Marc Afilalo[1,7], Laetitia Laurent[1], Maik Pietzner[8], Nicola Kerrison[8], Kaiqiong Zhao[1,2], Elsa Brunet-Ratnasingham[9,10], Danielle Henry[1], Nofar Kimchi[1], Zaman Afrasiabi[1], Nardin Rezk[1], Meriem Bouab[1], Louis Petitjean[1], Charlotte Guzman[1], Xiaoqing Xue[1], Chris Tselios,[1] Branka Vulesevic[1], Olumide Adeleye[1], Tala Abdullah[1], Noor Almamlouk[1], Yiheng Chen[1,3], Michaël Chassé[9], Madeleine Durand[9], Clare Paterson[11], Johan Normark[12], Robert Frithiof[13], Miklós Lipcsey[13,14], Michael Hultström[13,15], Celia M T Greenwood[1,2,16], Hugo Zeberg[17], Claudia Langenberg[6,18], Elin Thysell[19], Michael Pollak[1,20], Vincent Mooser[3], Vincenzo Forgetta[1], Daniel E. Kaufmann[8,21], J Brent Richards[1,2,3,22]

Affiliations:

1) Lady Davis Institute, Jewish General Hospital, McGill University, Montréal, Québec, Canada
2) Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montréal, Québec, Canada
3) Department of Human Genetics, McGill University Montréal, Québec, Canada
4) Kyoto-McGill International Collaborative School in Genomic Medicine, Graduate School of Medicine, Kyoto University, Kyoto, Japan
5) Research Fellow, Japan Society for the Promotion of Science
6) Department of Medicine, Division of Cardiology, McGill University, Montréal, Québec, Canada
7) Department of Emergency Medicine, Jewish General Hospital, McGill University, Montréal, Québec, Canada
8) MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Cambridge, UK
9) Research Centre of the Centre Hospitalier de l'Université de Montréal, Montréal, Canada
10) Department of Microbiology, Infectiology and Immunology, Université de Montréal, Montréal, Canada
11) SomaLogic, Inc., Boulder, Colorado, USA
12) Molecular Infection Medicine Sweden (MIMS) and Wallenberg Center for Molecular Medicine, Department of Clinical Microbiology, Section of Infection and Immunology, Umeå University, Umeå, Sweden
13) Anaesthesiology and Intensive Care Medicine, Department of Surgical Sciences, Uppsala University, Uppsala, Sweden
14) Hedenstierna Laboratory, CIRRUS, Anaesthesiology and Intensive Care Medicine, Department of Surgical Sciences, Uppsala University, Uppsala, Sweden
15) Integrative Physiology, Department of Medical Cell Biology, Uppsala University, Uppsala, Sweden
16) Gerald Bronfman Department of Oncology, McGill University
17) Department of Neuroscience, Karolinska Institutet, Stockholm, Sweden
18) Computational Medicine, Berlin Institute of Health, Charité University Medicine Berlin, Germany
19) Department of Medical Biosciences, Pathology, Umeå University, Umeå, Sweden
20) Departments of Medicine and Oncology, McGill University
21) Department of Medicine, Université de Montréal, Montréal, Canada
22) Department of Twin Research, King's College London, London, United Kingdom

***These authors contributed equally to this study**

Corresponding author:

Brent Richards, Professor of Medicine, McGill University
Senior Lecturer, King's College London (Honorary)
Pavilion H-413, Jewish General Hospital
3755 Côte-Ste-Catherine Montréal, Québec, Canada, H3T 1E2
T: +1 514 340 8222 x24362 F: +1 514 340 7529
E: brent.richards@mcgill.ca www.mcgill.ca/genepi

Key words: Genome-wide association studies, COVID-19, Mendelian randomization, proteomics, OAS1

Abstract: Currently 148 words

To identify circulating proteins influencing COVID-19 susceptibility and severity we undertook a two-sample Mendelian randomization (MR) study, rapidly scanning hundreds of circulating proteins while reducing bias due to reverse causation and confounding. In up to 14,134 cases and 1.2M controls, we found that a standard deviation increase in OAS1 levels was associated with reduced COVID-19 death or ventilation (OR = 0.54, P = $7x10^{-8}$), hospitalization (OR = 0.61, P = $8x10^{-8}$) and susceptibility (OR = 0.78, P = $8x10^{-6}$). Measuring OAS1 levels in 504 patients, we found that higher plasma OAS1 levels in a non-infectious state were associated with reduced COVID-19 susceptibility and severity. Further analyses suggested that a Neanderthal isoform of OAS1 in Europeans ancestry individuals affords this protection. Thus, evidence from MR and a case-control study support a protective role for OAS1 in COVID-19 adverse outcomes. Available pharmacological agents that increase OAS1 levels could be prioritized for drug development.

(Main text: 4211 words)

Introduction

To date, the COVID-19 pandemic has caused more than 2 million deaths worldwide, and infected approximately 100 million individuals.[1] Despite the scale of the epidemic, there are at present few disease-specific therapies[2] to reduce the morbidity and mortality of SARS-CoV-2 infection. Apart from dexamethasone therapy in oxygen-dependent patients[3], most clinical trials have shown, at most, mild or inconsistent benefits on disease outcomes.[4–6] Therefore, validated targets are needed for COVID-19 therapeutic development.

One source of such targets is circulating proteins. Recent advances in large-scale proteomics have enabled the measurement of thousands of circulating proteins—and when combined with evidence from human genetics, such targets greatly improve the probability of drug development success.[7–9] While *de novo* drug development will take time, repurposing of currently available molecules targeting those proteins could provide an accelerated opportunity to deliver new therapies to patients.

Nevertheless, since confounding and reverse causation often bias traditional circulating protein studies, methods are needed to dissect causal relationships. This is especially the case in COVID-19, where exposure to SARS-CoV-2 unleashes profound changes in circulating protein levels[10]. One way to address these limitations is by using Mendelian randomization (MR), a genetic epidemiology method that uses genetic variants as instrumental variables to test the effect of an exposure (here protein levels) on an outcome (here COVID-19 outcomes). The process of random assignment of alleles at conception greatly reduces bias from confounding. Since genotypes are always assigned prior to disease onset, MR studies are not influenced by reverse causation. However, MR rests on several assumptions[11], the most problematic being horizontal pleiotropy of the genetic instruments (wherein the genotype influences the outcome, independently of the exposure). One way to help avoid this bias is to use genetic variants that influence circulating protein levels which are adjacent to the gene which encodes the circulating protein

through the use of *cis*-protein quantitative trait loci (*cis*-pQTLs).[9] *cis*-pQTLs are likely to influence the level of the circulating protein by directly influencing its transcription or translation, and therefore less likely to affect the outcome of interest through pleiotropic pathways. Nevertheless, a causal genetic association between the exposure and outcome may be confounded by linkage disequilibrium (LD),[12] which can be detected through colocalization testing.

Understanding the etiologic role of circulating proteins in infectious diseases is challenging because the infection itself often leads to large changes in circulating protein levels[10]. Thus, it may appear that an increase in a circulating protein, such as a cytokine, is associated with a worsened outcome, when in fact, the cytokine may be the host's response to this infection and help to mitigate this outcome. It is therefore important to identify genetic determinants of the protein levels in the non-infected state, which would reflect a person's baseline predisposition to the level of a protein.

MR studies can be complemented by traditional case-control studies, where the protein is longitudinally measured in COVID-19 patients and controls, allowing for an estimation of the association between the protein level and COVID-19 outcomes. However, MR studies tend to predict the effect of the protein in the non-infectious state when the genetic determinants of such proteins are measured in the non-infected population. Since MR and case-control studies rely on different assumptions, and may be influenced by different biases, concordant results between the two study designs can strengthen the cumulative evidence.[13]

In this study, we therefore undertook two-sample MR and colocalization analyses to combine results from large-scale genome-wide association studies (GWAS) of circulating protein levels and COVID-19 outcomes.[14] We began by identifying the genetic determinants of circulating protein levels in large-scale proteomic GWAS, then used MR to assess whether these *cis*-pQTLs were associated with COVID-19 outcomes GWAS. Next, we investigated expression QTL (eQTL) and splice QTL (sQTL) effects of lead

proteins. We then measured the most promising protein, OAS1, in subjects ascertained for SARS-CoV-2 infection, followed for longitudinal sampling during and after their infection.

Results

MR using cis-pQTLs, and pleiotropy assessment

The study design is illustrated in **Figure 1**. We began by obtaining the genetic determinants of circulating protein levels from six large proteomic GWAS of European-ancestry individuals (Sun *et al*[15] N=3,301; Emilsson *et al*[16] N=3,200; Pietzner *et al*[17] N=10,708; Folkersen *et al*[18] N=3,394; Yao *et al*[19] N=6,861 and Suhre *et al*[20] N=997). A total of 931 proteins from these six studies had genome-wide significant *cis*-pQTLs, or highly correlated LD proxies ($R^2 > 0.8$), in the meta-analyses of data from COVID-19 Host Genetics Initiative[21] which included results from the GenOMICC program[22]. We then undertook MR analyses using 1,425 *cis*-pQTLs and 39 LD proxies as genetic instruments for circulating proteins in three COVID-19 outcomes: 1) Very severe COVID-19 disease (defined as individuals experiencing death, mechanical ventilation, non-invasive ventilation, high-flow oxygen, or use of extracorporeal membrane oxygenation. 99.7% of these individuals were of European ancestry) using 4,336 cases and 623,902 controls; 2) COVID-19 disease requiring hospitalization using 6,406 cases and 902,088 controls of European ancestry and 3) COVID-19 susceptibility using 14,134 cases and 1,284,876 controls of European ancestry. In all outcomes, cases required evidence of SARS-CoV-2 infection. For the very severe COVID-19 and hospitalization outcomes, COVID-19 cases were defined as laboratory confirmed SARS-CoV-2 infection based on nucleic acid amplification or serology tests. For the COVID-19 susceptibility outcome, cases were also identified by review of health records (using International Classification of Disease codes or physician notes).

MR analyses revealed that the levels of three circulating proteins, 2'-5'-oligoadenylate synthetase 1 (OAS1), interleukin-10 receptor beta subunit (IL10RB) and ABO were associated with at least two COVID-19 outcomes after Benjamini & Hochberg FDR correction (**Table 1, Extended Data 1-6**). Notably,

increased OAS1 levels were strongly associated with protection from all three COVID-19 outcomes. Further, these effect sizes were more pronounced with more severe outcomes, such that each standard deviation increase in OAS1 levels was associated with decreased odds of very severe COVID-19 (OR=0.54; 95% CI: 0.44-0.68, P=7.0x10^{-8}), hospitalization (OR=0.61; 95% CI: 0.51-0.73, P=8.3x10^{-8}) and susceptibility (OR=0.78; 95% CI: 0.69-0.87, P=7.6x10^{-6}) (**Figure 2A**). We also identified OAS1 *cis*-pQTLs in Emilsson *et al*[16] and Pietzner *et al*[17] which were not included in the initial MR due to lack of genome-wide significance for their association with OAS1 levels[16] or not included in their COVID-19 discovery panel.[17] MR analyses of using these additional *cis*-pQTLs, yielded concordant results (**Extended Data 7**).

We next assessed whether the *cis*-pQTL for OAS1 levels (rs4767027) was associated with over 5,000 other diseases, traits or protein levels, as catalogued in PhenoScanner.[23] rs4767027 was not associated with any other traits or protein levels (P<5.0x10^{-5}). These findings reduce the possibility that the MR estimate of the effect of OAS1 on COVID-19 outcomes is due to horizontal pleiotropy. Finally, except for COVID-19 susceptibility, the effect of rs4767027 did not demonstrate evidence of heterogeneity across COVID-19 Host Genetics Initiative GWAS meta-analyses (**Table 1).**

Using a *cis*-pQTL for IL10RB (rs2834167), we found that a one standard deviation increase in circulating IL10RB level was associated with decreased odds of very severe COVID-19 (OR=0.47; 95% CI: 0.32-0.68, P=7.1x10^{-5}) and hospitalization (OR = 0.53; 95% CI: 0.39-0.73, P=8.8x10^{-5}), but not susceptibility (**Figure 2A**). Using PhenoScanner, we did not find evidence of pleiotropic effects of the *cis*-pQTL for IL10RB. A one standard deviation increase in circulating ABO level was associated with increased odds of adverse COVID-19 outcomes (**Table 1**), however, we found that a *cis*-pQTL for ABO (rs505922) was strongly associated with the levels of several other proteins, suggesting potential horizontal pleiotropic effects (**Extended Data 8**). Given ABO's known involvement in multiple physiological processes, these results were expected, but highlight that MR analyses may suffer from significant bias from horizontal pleiotropy.

Colocalization Studies

To test whether confounding due to LD may have influenced the estimated effect of circulating OAS1 on COVID-19 outcomes, we tested the probability that the genetic determinants of OAS1 circulating protein level were shared with the three COVID-19 outcomes using colocalization analyses, as implemented in *coloc*.[12] The posterior probability that OAS1 levels and COVID-19 outcomes shared a single causal signal in the 1Mb locus around the cis-pQTL, rs4767027, was 0.72 for very severe COVID-19, 0.82 for hospitalization due to COVID-19, and 0.89 for COVID-19 susceptibility (**Figure 3)**. This colocalization result was also replicated using *OAS1 cis*-pQTL identified by Pietzner *et al*[17] (**Extended Data 7**). This suggests that there is likely a single shared causal signal for OAS1 circulating protein levels and COVID-19 outcomes.

Colocalization of ABO levels and different COVID-19 outcomes also showed colocalization between ABO level and different COVID-19 outcomes (posterior probability of single shared signal = 0.90, 0.98 and 1 for ABO level and very severe COVID-19, hospitalization due to COVID-19 and susceptibility, respectively) (**Extended Data Fig. 1)**. We were unable to perform colocalization analyses for IL10RB due to a lack of genome-wide summary level data from the original proteomic GWAS[16].

Aptamer Binding Effects

Protein altering variants (PAVs)[15] may influence binding of affinity agents, such as aptamers or antibodies, that are used to quantify protein levels. We thus assessed if the *cis*-pQTLs for the MR-prioritized proteins were PAVs, or in LD (R^2>0.8) with PAVs. rs2834167 (*IL10RB*) is a nonsense variant and could therefore be subject to potential binding effects. rs505922 (*ABO*) is not in LD with known missense variants. rs4767027 (*OAS1*) is an intronic variant, which is in LD with a missense variant rs2660 (R^2=1) in European ancestry. However, since expression studies derived from RNA sequencing are not

subject to potential effects of missense variants that could influence aptamer binding, we next explored whether rs4767027 also influences OAS1 expression and/or splicing.

sQTL and eQTL studies for OAS genes

Splicing QTLs (sQTLs) are genetic variants that influence the transcription of different isoforms of a protein. The aptamer that targets OAS1 was developed against a synthetic protein comprising the amino acid sequence 1-364 of NP002525.2[24], which is common to the two major *OAS1* isoforms: p46 and p42. Hence the aptamer may identify both, or either isoforms. rs10774671 is a known sQTL for OAS1 that induces alternate splicing, and creating p46 and p42 isoforms. A majority of present-day European ancestry individuals carry the alternative variant (rs10774671-A). The ancestral variant (rs10774671-G) is the major allele in African populations and became fixed in Neanderthal and Denisovan genomes[25,26]. However, the ancestral variant, with its increased expression of the p46 isoform, was reintroduced into the European population via gene flow from Neanderthals[27]. Previous analyses suggest that individuals with either the GG or GA genotype at rs10774671 express higher amounts of p46[27] which is also the predominant isoform found in circulating blood[28]. Differences in anti-viral activity has been observed between isoforms, with p46 being more active in certain viral infections[29]. Interestingly, the *OAS1* pQTL, rs4767027, is in high LD (R^2=0.97) with rs10774671[27] in European populations. Functional studies support that the G allele at rs10774671 increases expression of the p46 isoform but decreases expression of the p42 isoform[28]. This G allele at the sQTL rs10774671 reflects the T allele at pQTL rs4767027, which itself is associated with higher measured OAS1 levels and reduced odds of COVID-19 severity and susceptibility. These separate lines of evidence suggest that OAS1 levels, as measured by the SomaScan® platform predominantly identifies the p46 isoform, which may protect against COVID-19 outcomes.

Undertaking MR studies of *OAS1* splicing, we found that increased expression of the p46 isoform (as defined by normalized read counts of the intron cluster defined by LeafCutter[30,31]) was associated with

reduced odds of COVID-19 outcomes (OR = 0.29; 95% CI: 0.17-0.49, P=4.1x10^{-6} for susceptibility, OR = 0.09; 95% CI: 0.04-0.21, P=2.0x10^{-8} for hospitalization and OR = 0.05; 95% CI: 0.02-0.13, P=3.1x10^{-9} for very severe COVID-19) (**Figure 2B**). Colocalization analyses also supported a shared causal signal between the sQTL for *OAS1,* the pQTL and COVID-19 outcomes (**Extended Data Fig. 2**). Interestingly, the colocalization analyses supported a stronger probability of a shared signal with the sQTL, than the pQTL, suggesting that the p46 isoform may be the driver of the association of OAS1 levels with COVID-19 outcomes.

Next, we tested whether increased expression of OAS1 levels, without respect to isoform, were associated with COVID-19 outcomes using expression QTL (eQTL) MR analyses. We identified an eQTL for total OAS1, rs10744785, from GTEx v8.[32] Total *OAS1* expression levels were not associated with COVID-19 susceptibility and hospitalization (**Figure 2B**). We also found that increased *OAS3* expression in whole blood was positively associated with COVID-19 outcomes in MR analyses and support for colocalization of their genetic signal (**Extended Data 9, Extended Data Fig. 3**).

Taken together, these pQTL, sQTL and eQTL studies suggest that increased levels of the p46 isoform of OAS1 seem to protect against COVID-19 adverse outcomes.

Association of measured OAS1 protein level with COVID-19 outcomes

Since MR studies were derived from protein levels measured in a non-infected state, we tested the hypothesis that increased OAS1 protein levels in a non-infected state would be associated with reduced odds of COVID-19 outcomes. To do so, we undertook a case-control study, measuring OAS1 protein levels using the SomaScan® platform in 1,039 longitudinal samples from 399 SARS-CoV-2 PCR positive patients collected at multiple time points during their COVID-19 infection and 105 individuals who presented with COVID-19 symptoms but had negative SARS-CoV-2 PCR nasal swabs from the

Biobanque Quebecoise de la COVID-19 cohort (www.BQC19.ca). Individuals were recruited prospectively who had undergone nasal swabs for SARS-CoV-2 infection (**Table 2**).

We defined non-infectious samples as those collected from convalescent SARS-CoV-2 patients at least 31 days after onset of their symptoms (N=115), or samples collected from SARS-CoV-2 PCR negative patients (N=105). We also measured OAS1 levels in individuals with samples from SARS-CoV-2 positive patients <14 days after symptom onset (N=313), which showed increased OAS1 levels during infection (**Extended Data Fig. 4-6**). OAS1 levels are not associated with age and sex in non-infectious samples (**Extended Data Fig. 7**). After sample QC (see **Methods**), 308 patients with at least one sample collected during infection and 113 patients with at least one sample collected during a non-infectious state and 103 COVID-19 negative controls were included in the analyses (**Extended Data Fig. 8**).

To test whether OAS1 levels in a non-infectious state were associated with COVID-19 outcomes we undertook logistic regression controlling for age, sex, age^2, plate, recruitment center and sample processing time. OAS1 levels were log-transformed and standardized to match the transformation procedure of the MR study. We found that in the non-infectious samples, each standard deviation increase in OAS1 levels on the log-transformed scale was associated with reduced odds of COVID-19 outcomes (OR = 0.20 [95% CI: 0.08 – 0.53]; P = 0.001 for very severe COVID-19, OR = 0.46 [95% CI: 0.28 – 0.76], P = 0.002 for hospitalization and OR = 0.69 [95% CI: 0.49 – 0.98], P = 0.04 for susceptibility) (**Figure 4, Extended Data 10, Extended Data Fig. 9**). These results are consistent with our findings from MR, where increased circulating OAS1 levels in a non-infectious state were associated with protection against all of these adverse COVID-19 outcomes.

In samples drawn during active infection we found that increased OAS1 levels were associated with increased odds of adverse COVID-19 outcomes (OR = 1.50 [95% CI: 1.19 – 1.90]; P = 0.0007 for very severe COVID-19, OR = 1.93 [95% CI: 1.46 – 2.56], P = 4.8×10^{-6} for hospitalization and OR = 4.39 [95% CI: 2.87 – 6.73], P = 1.09×10^{-11} for susceptibility) (**Figure 4**).

Taken together, these findings suggest that increased OAS1 levels in a non-infectious state are associated with better COVID-19 outcomes, and that during infection, SARS-CoV-2 exposure likely causes OAS1 levels to increase, as interferon pathways are stimulated, which are known to increase OAS1 levels[33].

Discussion:

Disease-specific therapies are needed to reduce the morbidity and mortality associated with COVID-19 outcomes. In this large-scale two-sample MR study of 931 proteins assessed for three COVID-19 outcomes in up to 14,134 cases and 1.2 million controls with European ancestry, we provide evidence that increased OAS1 levels in the non-infectious state are strongly associated with reduced risks of very severe COVID-19, hospitalization and susceptibility. The protective effect size was particularly large, such that a 50% decrease in the odds of very severe COVID-19 was observed per standard deviation increase in OAS1 circulating levels. OAS proteins are part of the innate immune response against RNA viruses. They are induced by interferons and activate latent RNase L, resulting in direct viral and endogenous RNA destruction, as demonstrated in *in-vitro* studies.[34] Thus OAS1 has a plausible biological activity against SARS-CoV-2. Since therapies exist that activate OAS1, repositioning them as potential COVID-19 treatments should be prioritized.

In populations outside South-Saharan Africa, the protective alleles at both rs4767027-T (the OAS1 pQTL) and rs10774671-G (the OAS1 sQTL) are found on a Neanderthal haplotype[35] which was passed on to modern humans ~50-60,000 years ago[36]. The correspondence between the previously described gene flow[36] from Neandertals at this locus and the haplotype associated with protection against COVID-19 in the GWAS[22] was recently demonstrated[35]. Even though these two SNPs share a haplotype, their evolutionary histories differ. The rs4767027-T allele is derived from the Neanderthal lineage, whereas for the rs10774671-G allele, Neanderthals preserved the ancestral state. *OAS1* alternative splicing regulated by the rs10774671-G allele increases the isoform p46, which has a higher enzymatic activity against

viruses than the p42 isoform[37] and is the only OAS1 isoform robustly upregulated during infection[27]. Although further studies are needed to fully elucidate the functional relevance of the pQTL and sQTL for OAS1, the antiviral activity of the gene products is higher for the Neandertal haplotype than the common haplotype in Europeans[29]. In Europeans the Neandertal haplotype has undergone positive selection[27] and the rs4767027-T allele reaches an allele frequency of 0.32. Using MR and measurements of circulating proteins, we demonstrated here that increased OAS1 levels of the Neandertal haplotype in modern-day European ancestry individuals confers this protective effect.

Our MR evidence indicated that higher p46 isoform levels of OAS1 and higher OAS1 total protein levels, as measured by the SomaScan® assay had protective effects on COVID-19 outcomes. These results were strongly supported by colocalization analysis. Given the consistent colocalization between the sQTL and pQTL for OAS1, the lack of colocalization between the eQTL and pQTL for OAS1, and the evidence that the SomaScan® assay likely measures p46 isoforms, it seems probable that the protective effect of OAS1 is derived from the p46 isoform. However, further investigations are required to specifically measure each isoform in circulation and isoform activity assays will be required to better understand if the p46 isoform, rather than total OAS1 levels are most protective against COVID-19 outcomes.

The ancestral *OAS1* splice variant encoding the more active p46 isoform was lost in the modern human population that left Africa. Several scenarios might explain this loss-of-function, e.g., loss of purifying selection during the out-of-Africa exodus, which may be due to changes in environmental pathogens, or potential harm induced by OAS1 antiviral activity.[38] Unfortunately, we do not have sufficient data to test if the OAS1 p46 ancestral allele in Africans south of Sahara also offers protection against COVID-19. Nevertheless these findings further emphasize the importance of the Neanderthal genome in COVID-19 risk modulation, since a risk locus on chromosome 3 has also been reported to be inherited from Neanderthals.[39]

OAS1, *OAS2* and *OAS*3 share significant homology. As an interferon stimulated gene[40], *OAS1* polymorphisms have been associated with the host immune response to several classes of viral infection.[41,42,43,44,45] Given that OAS1 is an intracellular enzyme activating RNase L leading to viral RNA degradation, it is probable that the circulating levels of this enzyme reflect intracellular levels of this protein. However, there is experimental evidence that extracellular OAS1 may also be important in the viral immune response[34].

Molecules currently exist which can influence *OAS1* expression. Interferon beta-1b, which activates a cytokine cascade leading to increased *OAS1* expression,[46] is currently used to treat multiple sclerosis and has been shown to induce *OAS1* expression in blood cells.[47] Interferon-based therapy has also been used in other viral infections[48]. However, recent randomized trials have shown inconsistent results. While intravenous interferon beta-1b combined with lopinavir-ritonavir reduced mortality due to MERS-CoV infections,[49] in the unblinded SOLIDARITY trial,[50] there was no demonstrated benefit of intravenous interferon-beta-1b. On the other hand, a recent phase II trial testing the effect of inhaled nebulized interferon beta-1a (which is closely related to interferon beta-1b) showed improved COVID-19 symptoms in the treatment arm.[51] While this study was not powered to show a difference in mortality, all deaths occurred in the placebo group. Inhaled nebulized interferon beta results in a much higher tissue availability in the lung and may result in improved anti-viral activity. Moreover, timing of administration is likely to play a role, as the administration of a pro-inflammatory cytokine may not provide benefit during the inflammation driven phase of the disease. However, data on timing of administration is currently unavailable in the SOLIDARITY trial, and conclusions cannot yet be drawn. Lastly the effect of interferon supplement may vary across ancestral population, as different ancestries have different amounts of the more active p46 isoform of OAS1. Our study was limited to individuals of European ancestry, a population with higher expression of the p46 isoform. Interestingly, the SOLIDARITY trial enrolled 78% of its patients in South Asia, Middle East, North Africa, and Latin America, populations that may have higher expression of the p42 OAS1 isoform, while the study on inhaled interferon beta was comprised of 80% White patients

from the United Kingdom. It is possible that interferon beta-1b may have different effects in populations of different ancestry, due to different frequency of genetic variant in different populations.

There is *in-vitro* evidence that pharmacological inhibition of phosphodiesterase-12, which degrades 2'-5'-oligoadenylate synthesized by OAS1, potentiates OAS-mediated antiviral activity.[52,53] Interestingly, coronaviruses in the same family as SARS-CoV-2 have been shown to produce viral proteins that degrade 2'-5'-oligoadenylate and reduce RNase-L activity, leading to evasion of the host immune response.[54,55] Our findings are also consistent with recent experimental work[56], showing that there are situations where SARS-CoV-2 is sensitive to OAS1-related antiviral defenses. Our findings motivate pharmacologic strategies to increase OAS1 levels or activity, as well as further evaluation of the possible antiviral activity of extracellular OAS1.[34] Thus existing pre-clinical molecules that lead to increased OAS1 levels[52] could be optimized and tested their effect upon COVID-19 outcomes.

Our MR analyses found that higher levels of *OAS3* expression is associated with worse COVID-19 outcomes, which is an opposite direction of effect compared to OAS1. The discordant effects of the p46 isoform for OAS1 and OAS3 were also reported by a previous study[27] , which might reflect complex biology of *OAS* genes for innate immune response. In a recent transcription-wide association study from the GenOMICC program[22], genetically-predicted high expression of OAS3 in lungs and whole blood were associated with higher risk of becoming critically ill COVID-19 patients. Although further studies to assess the roles of *OAS* genes specific to SARS-CoV-2 are needed, it is likely that OAS1 is the main driver of the protective effect of p46 isoform for COVID-19 outcomes given prior functional studies demonstrating the antiviral effect of *OAS* genes[27].

This study has limitations. First, we used MR to test the effect of circulating protein levels measured in a non-infected state since the effect of the *cis*-pQTLs upon circulating proteins was estimated in individuals who had not been exposed to SARS-CoV-2. Once a person contracts SARS-CoV-2 infection, levels of

circulating proteins could be altered, and this may be especially relevant for cytokines such as IL10 (which binds to IL10RB) and OAS1. Thus, the MR results presented in this paper should be interpreted as an estimation of the effect of circulating protein levels, when measured in the non-infected state. On-going studies will help to clarify if the same *cis*-pQTLs influence circulating protein levels during infection. Second, this type of study suffers a high false-negative rate. Our goal was not to identify *every* circulating protein influencing COVID-19 outcomes, but rather to provide evidence for few proteins with strong *cis*-pQTLs since these proteins are more likely to be robust to the assumptions of MR studies. Future large-scale proteomic studies with more circulating proteins properly assayed should help to overcome these limitations. Third, most MR studies assume a linear relationship between the exposure and the outcome. Thus, our findings would not identify proteins whose effect upon COVID-19 outcomes has a clear threshold effect. Fourth, the overall OAS1 levels measured by RNA-seq (not only p46) may be biased by the effect of alternative splicing, and the role of overall OAS1 and OAS3 levels indicated by the association of the *cis*-pQTL of OAS1 in protection against COVID-19 are possible and not yet explored. We also could not completely exclude the possibility that measurement of OAS1 levels may be influenced by aptamer-binding effects. Last, all data presented in the manuscript pertain to individuals of European ancestry only—once again underlining the importance of genotyping efforts in other populations.

In conclusion, we have used genetic determinants of circulating protein levels and COVID-19 outcomes obtained from large-scale studies and found compelling evidence that OAS1 has a protective effect on COVID-19 susceptibility and severity. Measuring plasma OAS1 levels in a case-control study demonstrated that higher circulating levels of this protein in a non-infectious state are strongly associated with reduced risk of adverse COVID-19 outcomes. Interestingly, the available evidence suggests that the protective effect from OAS1 in European ancestry individuals is likely due to the Neanderthal introgressed p46 OAS1 isoform. Known pharmacological agents that increase OAS1 levels[52] could be explored for their effect on COVID-19 outcomes.

Methods:

pQTL GWAS

We systematically identified pQTL associations from six large proteomic GWASs.[15–20] Each of these studies undertook proteomic profiling using either SomaLogic® SomaScans, or O-link proximal extension assays.

COVID GWAS and COVID-19 Outcomes

To assess the association of *cis*-pQTLs with COVID-19 outcomes, we used COVID-19 meta-analytic GWAS (data-freeze 4) from the COVID-19 Host Genetics Initiative[21]. For our study, we used three of these GWAS meta-analyses which included 25 cohorts of European ancestry and 1 cohort of admixed American ancestry. The outcomes tested were very severe COVID-19, hospitalization due to COVID-19, and susceptibility to COVID-19 (named A2, B2, and C2, respectively by the COVID-19 Host Genetics Initiative).

Very severe COVID-19 cases were defined as hospitalized individuals with COVID-19 as the primary reason for hospital admission with laboratory confirmed SARS-CoV-2 infection (nucleic acid amplification tests or serology based), and death or respiratory support (invasive ventilation, continuous positive airway pressure, Bilevel Positive Airway Pressure, or continuous external negative pressure, high-flow nasal or face-mask oxygen). Simple supplementary oxygen (e.g. 2 liters/minute via nasal cannula) did not qualify for case status. Controls were all individuals in the participating cohorts who did not meet this case definition.

Hospitalized COVID-19 cases were defined as individuals hospitalized with laboratory confirmed SARS-CoV-2 infection (using the same microbiology methods as for the very severe phenotype), where

hospitalization was due to COVID-19 related symptoms. Controls were all individuals in the participating cohorts who did not meet this case definition.

Susceptibility to COVID-19 cases was defined as individuals with laboratory confirmed SARS-CoV-2 infection, health record evidence of COVID-19 (international classification of disease coding or physician confirmation), or with self-reported infections (e.g. by questionnaire). Controls were all individuals who did not meet this case definition.

Two-sample Mendelian randomization

We used two-sample MR analyses to screen and test potential circulating proteins for their role influencing COVID-19 outcomes. In two-sample MR, the effect of SNPs on the exposure and outcome are taken from separate GWASs. This method often improves statistical power, because it allows for larger sample sizes for the exposure and outcome GWAS.[57]

Exposure definitions: We conducted MR using six large proteomic GWAS studies.[15–20] Circulating proteins from Sun *et al*, Emilsson *et al* and Pietzner *et al* were measured on the Somalogic platform, Suhre *et al*, Yao *et al* and Folkersen *et al* used protein measurements on the O-link platform. We selected proteins with only *cis-pQTL*s to test their effects on COVID-19 outcomes, because they are less likely to be affected by potential horizontal pleiotropy. The *cis-pQTL*s were defined as the genome-wide significant SNPs ($P < 5 \times 10^{-8}$) with the lowest P value within 1 Mb of the transcription start site (TSS) of the gene encoding the measured protein.[9] For proteins from Emilsson *et al*, Pietzner *et al*, Suhre *et al*, Yao *et al* and Folkersen *et al,* we used the sentinel *cis*-pQTL per protein per study as this was the data available. For proteins from Sun *et al,* we used PLINK 1.9[58] and the 1000 Genome[59] European population reference panels to clump and select LD-independent *cis*-pQTL (R2<0.001, distance 1000 kb) with the lowest P-value from reported summary statistics for each SOMAmer® bound proteins. We included the same proteins represented by different *cis*-pQTLs from different studies in order to cross examine the findings.

For *cis*-pQTLs that were not present in the COVID-19 GWAS, SNPs with LD R^2>0.8 and with minor allele frequency (MAF) < 0.42 were selected as proxies, MAF > 0.3 was used for allelic alignment for proxy SNPs. *cis*-pQTLs with palindromic effects and with minor allele frequency (MAF) > 0.42 were removed prior to MR to prevent allele-mismatches. Benjamini & Hochberg correction was used to control for the total number of proteins tested using MR. MR analyses were performed using the TwoSampleMR package in R.[60] For proteins with a single (sentinel) *cis*-pQTL, we used the Wald ratio to estimate the effect of each circulating protein on each of the three COVID-19 outcomes. For any proteins/SOMAmer® reagents with multiple independent *cis*-pQTL, an inverse variance weighted (IVW) method was used to meta-analyze their combined effects. After harmonizing the *cis*-pQTLs of proteins with COVID-19 GWAS, a total of 566 SOMAmer® reagents (529 proteins, 565 directly matched IVs and 26 proxies) from Sun *et al*, 760 proteins (747 directly matched IVs and 11 proxies) from Emilsson *et al*, 91 proteins (90 directly matched IVs and 2 proxies) from Pietzner *et al*, 74 proteins (72 directly matched IVs) from Suhre *et al*, 24 proteins (24 directly matched IVs) from Yao *et al* and 13 proteins (13 directly matched IVs) from Folkersen *et al* were used as instruments for the MR analyses across the three COVID-19 outcomes (**Extended Data 11-12**).[15–20]

Pleiotropy assessments

A common pitfall of MR is horizontal pleiotropy, which occurs when the genetic variant affects the outcome via pathways independent of the exposure. The use of circulating protein *cis*-pQTLs greatly reduces the possibility of pleiotropy, for reasons described above. We also searched in the PhenoScanner[23] database, a large catalogue of observed SNP-outcome relationships involving > 5,000 GWAS done to date to assess potentially pleiotropic effects of the *cis*-pQTLs of MR prioritized proteins, by testing the association of *cis*-pQTLs with other circulating proteins (i.e. if they were *trans*-pQTLs to other proteins, or significantly associated with other unrelated disease or traits). For *cis*-pQTLs of MR prioritized proteins measured on the SomaLogic® platform, we assessed the possibility of potential aptamer-binding effects (where the presence of protein altering variants may affect protein

546 measurements). We also checked if *cis*-pQTLs of MR prioritized proteins had significantly heterogeneous
547 associations across COVID-19 populations in each COVID-19 outcome GWAS.

548

549 ***Colocalization analysis***

550 Next, we tested colocalization of the genetic signal for the circulating protein and each of the three
551 COVID-19 outcomes using colocalization analyses, which assess potential confounding by LD.
552 Specifically, for each of these MR significant proteins with genome-wide summary data available, for the
553 proteomic GWASs, a stringent Bayesian analysis was implemented in *coloc*[12] R package to analyze all
554 variants in 1MB genomic locus centered on the *cis*-pQTL. Colocalizations with posterior probability for
555 hypothesis 4 (PP4, that there is an association for both protein level and COVID-19 outcomes and they
556 are driven by the same causal variant) > 0.5 were considered likely to colocalize (which means the
557 highest posterior probability for all 5 *coloc* hypotheses), and PP4 > 0.8 was considered to be highly likely
558 to colocalize.

559

560 ***sQTL and eQTL MR and colocalization studies for OAS genes***

561 We performed MR and colocalization analysis using GTEx project v8[32] GWAS summary data to
562 understand the effects of expression and alternative splicing of *OAS* genes in whole blood. The genetic
563 instruments were conditionally independent (R^2 < 0.001) sQTL and eQTL for *OAS1*, eQTL for *OAS2* and
564 *OAS3* identified by using stepwise regression in GTEx[32]. The sQTL SNP for *OAS1* (rs10774671), was
565 originally identified for the normalized read counts of LeafCutter[30] cluster of the last intron of p46 isoform
566 (chr12:112,917,700-112,919,389 GRCh38) in GTEx[31], and was used to estimate the effect of p46
567 isoform. Colocalization analysis was performed using GWAS summary from GTEx by restricting the
568 regions within 1 Mb of each QTL.

569

570 ***Measurement of plasma OAS1 protein levels associated with COVID-19 outcomes in BQC19***

BQC19 is a Québec-wide initiative to enable research into the causes and consequences of COVID-19 disease. The patients included in this paper were recruited at the Jewish General Hospital (JGH) and Centre hospitalier de l'Université de Montréal (CHUM) in Montréal, Québec, Canada.

COVID-19 case – control status was defined to be consistent with the GWAS study from COVID-19 HGI, from which the MR results were derived. Namely, we tested the association of OAS1 protein levels with the three different COVID-19 outcome definitions both in samples procured from non-infected stages and from samples during the acute phase of the infection. The three outcomes were: 1) Very severe COVID-19—defined as hospitalized individuals with laboratory confirmed SARS-CoV-2 infection (nucleic acid amplification tests or serology based), and death or respiratory support (invasive ventilation, continuous positive airway pressure, Bilevel Positive Airway Pressure, or continuous external negative pressure, high flow nasal or face-mask oxygen). Controls were all individuals who did not meet this case definition; 2) Hospitalized COVID-19 cases—defined as individuals hospitalized with laboratory confirmed SARS-CoV-2 infection. Controls were all who did not meet this case definition; 3) Susceptibility to COVID-19—cases were defined as individuals with laboratory confirmed SARS-CoV-2 infection, and controls were all individuals who underwent PCR testing for SARS-CoV-2, but were negative. The date of symptom onset for COVID-19 patients was collected from patients' charts or estimated from their first positive COVID-19 tests if missing. Case inclusion criteria was not exclusive, which means that some individuals who were cases in the susceptibility analyses were also included in the hospitalization and very severe COVID-19 if they met case definitions.

A total of 125 individuals were recruited from CHUM and 379 individuals were recruited from the JGH. Individuals had blood sampling done at up to five different time points (200 individuals had one measurement, 113 individuals had two measurements, 152 individuals had three measurements, 38 individuals had four measurements and 1 individual had five measurements). Days from symptom onset were calculated for each sample based on the date of symptom and blood draw date (T1). For COVID-19 negative individuals, T1 was set to 0. Sample processing time (in hours) for each sample was also

598 calculated to measure the duration of time from sample collection to processing to account for the
599 increase in the amount of protein released from cell lysis due to extended sample handling time.

600

601 Protein levels in citrated (ACD) plasma samples were measured using the SomaScan® assay.1,039
602 samples from 399 SARS-CoV-2 positive patients and 105 SARS-CoV-2 negative patients of mainly
603 European descent underwent SomaScan® assays, which included 5,284 SOMAmer reagents, targeting
604 4,742 proteins. The SomaScan® assay uses single-stranded DNA aptamers ("SOMAmers"), which are
605 designed to selectively bind to a particular protein target[61]. SOMAmer reagent binding is quantified by
606 microarray, measuring abundance in Relative Fluorescent Units (RFU). The RFUs for each protein
607 underwent four normalization processes including hybridization control, intraplate median signal
608 normalization, plate scaling and calibration and median signal normalization to a reference generated
609 from internal data across all samples. All normalizations were conducted by SomaLogic® and detailed in
610 their Technical Note[62].

611

612 Among SARS-CoV-2 positive participants, we defined samples procured from patients during the
613 infectious state as those sampled within 14 days (including the 14th day) from the first date of symptoms[63].
614 For patients with more than one sample within 14 days of symptom onset, the earliest sample was used.
615 We defined samples procured from patients who were non-infectious as samples from SARS-CoV-2
616 positive patients taken at least 31 days after symptom onset.We selected 31 days, as this is the upper
617 limit of the interquartile range of the duration of SARS-CoV-2 positivity in a recent systematic review and
618 coincided with the first scheduled outpatient follow-up blood test in the BQC19[64]. For individuals with
619 more than one sample at least 31 days of symptom onset, the latest sample was used.

620

621 OAS1 level was measured by one SOMAmer reagent (OAS1.10361.25). Within each group, median
622 signal normalized OAS1 levels were natural log transformed, adjusted for sample processing time and the
623 residuals were further standardized. For each group, we removed samples that were outliers with long

sample processing time (sample processing time > 50 hrs) or high OAS1 level (log OAS1 level > 8)**.** Logistic regression was performed to test the association standardized OAS1 level with the three COVID-19 outcomes including age, sex, age^2, center of recruitment and plates as covariates.

Acknowledgements:

We thank Dr. Luis Barriero for his comments on the introgressed Neanderthal p46 isoform, and Dr. Nicolas Chomont for the discussion on the clinical phenotype of BQC19. We would like to also thank the MI4 and the MUHC Foundation that contributed for the SomaLogic® panel.

Funding:

The Richards research group is supported by the Canadian Institutes of Health Research (CIHR: 365825; 409511), the Lady Davis Institute of the Jewish General Hospital, the Canadian Foundation for Innovation (CFI), the NIH Foundation, Cancer Research UK, Genome Québec, the Public Health Agency of Canada, the McGill Interdisciplinary Initiative in Infection and Immunity and the Fonds de Recherche Québec Santé (FRQS). SZ is supported by a CIHR fellowship and a FRQS fellowship. GBL is supported by the a CIHR scholarship, and a joint FRQS and Québec Ministry of Health and Social Services scholarship. TN is supported by Research Fellowships of Japan Society for the Promotion of Science (JSPS) for Young Scientists. JBR is supported by a FRQS Clinical Research Scholarship. Support from Calcul Québec and Compute Canada is acknowledged. TwinsUK is funded by the Welcome Trust, Medical Research Council, European Union, the National Institute for Health Research (NIHR)-funded BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation

649 Trust in partnership with King's College London. BQC19 is funded by FRQS, Genome Québec and the
650 Public Health Agency of Canada. These funding agencies had no role in the design, implementation or
651 interpretation of this study. VM is supported by a Canada Excellence Research Chair. MD is supported by
652 a FRQS Clinical Research Scholarship. RF is funded by the Swedish Research Council (2014-02569 and
653 2014-07606). The Kaufmann lab COVID-19 work is supported by the CIHR/CITF (VR2-173203), AmFAR
654 (110068-68-RGCV) the CFI and FRQS. RF is supported by Swedish Research Council (2014-02569 and
655 2014-07606). MH is supported by a SciLifeLab/KAW national COVID-19 research program project grant
656 (KAW 2020.0182).
657

658 **Ethics declarations:**

659 All cohorts contributing cohorts to COVID-19 HGI received ethics approval from their respective research
660 ethics review boards. This study was approved by the research ethics boards of the Jewish General
661 Hospital and the Centre Hospitalier du Université de Montréal. Informed consent was obtained from the
662 participants of BQC19.

663 **Disclosures:**

664 JBR has served as an advisor to GlaxoSmithKline and Deerfield Capital. CP is an employee of
665 SomaLogic Inc. MP has consulted for Astellas Pharma, Ansh Labs, Immunomet, Zymeworks and Pfizer
666 on unrelated projects. All other authors declare that there is no conflict of interest.
667

668 **Data availability:**

669 Data from proteomics studies and GTEx consortium (GTEx project v8[32]) are available from the referenced
670 peer-reviewed studies[15–20] or their corresponding authors, as applicable. PhenoScanner online database:
671 http://www.phenoscanner.medschl.cam.ac.uk/. Summary statistics for the COVID-19 outcomes are
672 publicly available for download on the COVID-19 HGI website (www.covid19hg.org). The BQC19 is an
673 Open Science biobank. Instructions on how to access data for individuals from the BQC19 at the Jewish
674 General Hospital site is available here: https://www.mcgill.ca/genepi/mcg-covid-19-biobank. Instructions

on how to access data from other sites of the BQC19 is available here: https://www.bqc19.ca/en/access-data-samples.

Author contributions:

Conception and design: SZ, GBL and JBR. Data analyses: SZ and TN. Data acquisition: TN, GBL, DM, DEK, JA, MA, LL, EBR, DH, NK, ZA, NR, MB, LP, CG, XX, CT, BV, OA, TA, NA, MC, MD, VF, DEK and JBR. Interpretation of data: SZ, GBL, TN, MP, YC, DEK, VF and JBR. Funding acquisition: DM, VM, VF, JBR. Methodology: SZ, KZ, CMTG and JBR. Project administration: DM, VF and JBR. Validation: SZ, TN, MP, NK, MP, JN, ET, CL, DEK and JBR. Visualization: SZ, TN and VF. Writing-original draft: SZ, GBL, TN and JBR. Writing-review & editing: SZ, GBL, TN, MP, HZ, VM, MP, RF, ML, MH, CP, DEK and JBR. All authors were involved in further drafts of the manuscript and revised it critically for content. All authors gave final approval of the version to be published. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

687 **Table 1. MR-Identified Circulating Protein Levels Affecting COVID-19 Outcomes**

688

Protein	*cis*-pQTL	Source	Very Severe COVID-19 (99.7% European Ancestry)				COVID-19 Hospitalization (European Ancestry Only)				COVID-19 Susceptibility (European Ancestry Only)			
			OR	95%CI	P value	P het	OR	95%CI	P value	P het	OR	95%CI	P value	P het
OAS1	rs4767027	Sun	0.54	0.44-0.68	7.0×10^{-8}	0.37	0.61	0.51- 0.73	8.3×10^{-8}	0.16	0.78	0.69-0.87	7.6×10^{-6}	0.005
ABO	rs505922	Sun, Emilsson	1.09	1.05-1.14	6.4×10^{-5}	0.10	1.11	1.07-1.15	6.8×10^{-9}	0.06	1.07	1.05-1.10	1.1×10^{-9}	0.10
IL10RB	rs2834167	Emilsson	0.47	0.32-0.68	7.1×10^{-5}	0.02	0.53	0.39-0.73	8.8×10^{-5}	0.11	0.87	0.72-1.07	0.18	0.006

689

690 OR: represents the estimated effect of a standard deviation on the natural log scale (for Sun et
691 al) or one unit (for Emilsson et al) increase in protein levels on the odds of the three COVID-19
692 outcomes. P het: P value of heterogeneity for each *cis*-pQTLs across the cohorts in the GWAS
693 summary-level meta-analysis from COVID-19 Host Genomic Initiative.
694

695

Table 2. Participant Demographics of the BQC19 Cohort Included in this Study

Sample Demographics	Number of Individuals (%) (Total N=504)
Sex	
Female	250 (49.6%)
Male	254 (50.4%)
Age (years) *	65.4 (18.0)
BMI*	28.6 (6.18)
Missing	225 (44.6%)
SARS-CoV-2 PCR test	
Positive	399 (79.2%)
Negative	105 (20.8%)
Hospitalization	
Hospitalized	406 (80.6%)
Outpatient treatment only	98 (19.4%)
Hospitalization duration (days)†	14.0 [6.00, 27.0]
Death	
Deceased	43 (8.5%)
Survived	461 (91.5%)
Respiratory Support	
No oxygen	233 (46.2%)
Oxygen supplement	143 (28.4%)
Mechanical Ventilation	128 (25.4%)
Days on ventilator†	14.0 [6.75, 23.5]

*Mean (SD) and † Median (25%QR, 75%QR), which was calculated amongst those who were hospitalized and those on ventilator, respectively.

Figure Legends:

Figure 1. Flow Diagram of Study Design

Figure 2. Association of Circulating Protein Levels of OAS1, ABO and IL10RB and mRNA levels of OAS1 with COVID-19 Outcomes from MR

Forest plot showing odds ratio and 95% confidence interval from two sample Mendelian Randomization analyses (two-sided). P values are unadjusted. A: MR estimates of proteins influencing COVID-19 outcomes, unit: standard deviation of log normalized value; B. MR estimates of OAS1 mRNA influencing COVID-19 outcomes, unit: standard deviation of normalized read counts.

Figure 3. Colocalization of the Genetic Determinants of OAS1 Plasma Protein Levels and COVID-19 Outcomes

Colocalizatinon of genetic signal of 1MB region around OAS1 pQTL rs4767027 of OAS1 level (top plot) and COVID-19 outcomes (three bottom plots), color shows SNPs in the region in LD (r^2) to rs4767027 (purple). The posterior probability (PP) of a shared single signal between OAS1 levels and three COVID-19 ouctomes are estimated by *coloc*.

Figure 4. Association of OAS1 levels with COVID-19 Outcomes from the Case-Control Study in BQC19

Forest plot showing odds ratios and 95% confidence intervals from logistic regression analyses (two-sided). P values are unadjusted. During Infection: Patient samples that were collected within 14 days from the date of symptom onset. For individuals with two or more samples collected within 14 days of symptom onset, the earliest time point was used.

Non-Infectious State: Patient samples that were collected at least 31 days from the date of symptom onset. For individuals with two or more samples collected at different time points at least 31 days from symptom onset, the latest time point was used.

Additional information is also described in Extended Data 10.

References

1. Johns Hopkins. Corona Virus Resource Center. (2020).

2. David M. Weinreich, M.D., Sumathi Sivapalasingam, M.D., Thomas Norton, M.D., Shazia Ali, Pharm.D., Haitao Gao, Ph.D., Rafia Bhore, Ph.D., Bret J. Musser, Ph.D., Yuhwen Soo, Ph.D., Diana Rofail, Ph.D., Joseph Im, B.S., Christina Perry, M.B.A., Cynthia Pan, B. P. REGN-COV2, a Neutralizing Antibody Cocktail, in Outpatients with Covid-19. *N. Engl. J. Med.* (2020) doi:10.1056/NEJMoa2035002.

3. Dexamethasone in Hospitalized Patients with Covid-19 — Preliminary Report. *N. Engl. J. Med.* 1–11 (2020) doi:10.1056/nejmoa2021436.

4. Sterne, J. A. C. *et al.* Association between administration of systemic corticosteroids and mortality among critically ill patients with COVID-19: A meta-analysis. *JAMA - J. Am. Med. Assoc.* (2020) doi:10.1001/jama.2020.17023.

5. Beigel, J. H. *et al.* Remdesivir for the Treatment of Covid-19 — Preliminary Report. *N. Engl. J. Med.* 2019–2020 (2020) doi:10.1056/nejmoa2007764.

6. Cavalcanti, A. B. *et al.* Hydroxychloroquine with or without Azithromycin in Mild-to-Moderate Covid-19. *N. Engl. J. Med.* (2020) doi:10.1056/nejmoa2019014.

7. Nelson, M. R. *et al.* The support of human genetic evidence for approved drug indications. *Nat. Genet.* **47**, 856–860 (2015).

8. Cook, D. *et al.* Lessons learned from the fate of AstraZeneca's drug pipeline: a five-dimensional framework. *Nat. Rev. Drug Discov.* **13**, 419–31 (2014).

9. Zheng, J. *et al.* Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. *Nat. Genet.* (2020) doi:10.1038/s41588-020-0682-6.

10. Filbin, M. R. *et al.* Plasma proteomics reveals tissue-specific cell death and mediators of cell-cell interactions in severe COVID-19 patients. *bioRxiv* (2020).

11. Davey Smith, G., Ebrahim, S., Smith, G. D. & Ebrahim, S. *'Mendelian randomization': Can genetic epidemiology contribute to understanding environmental determinants of disease? International Journal of Epidemiology* vol. 32 1–22 (2003).

12. Giambartolomei, C. *et al.* Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. *PLoS Genet.* **10**, (2014).

13. Lawlor, D. A., Tilling, K. & Smith, G. D. Triangulation in aetiological epidemiology. *Int. J. Epidemiol.* (2016) doi:10.1093/ije/dyw314.

14. The COVID-19 Host Genetics Initiative, a global initiative to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic. *Eur. J. Hum. Genet.* **28**, 715–718 (2020).

15. Sun, B. B. *et al.* Genomic atlas of the human plasma proteome. *Nature* **558**, 73–79 (2018).

16. Emilsson, V. *et al.* Co-regulatory networks of human serum proteins link genetics to disease. *Science (80-.).* **1327**, 1–12 (2018).

17. Pietzner, M. *et al.* Genetic architecture of host proteins interacting with SARS-CoV-2. *bioRxiv Prepr. Serv. Biol.* (2020) doi:10.1101/2020.07.01.182709.

18. Folkersen, L. *et al.* Mapping of 79 loci for 83 plasma protein biomarkers in cardiovascular disease. *PLoS Genet.* **13**, e1006706 (2017).

19. Yao, C. *et al.* Genome-wide mapping of plasma protein QTLs identifies putatively causal genes and pathways for cardiovascular disease. *Nat. Commun.* **9**, 3268 (2018).

20. Suhre, K. *et al.* Connecting genetic risk to disease end points through the human blood plasma proteome. *Nat. Commun.* **8**, 14357 (2017).

21. COVID-19 Host Genetics Initiative. https://www.covid19hg.org/results/.

22. Pairo-Castineira, E. *et al.* Genetic mechanisms of critical illness in Covid-19. *Nature* (2020) doi:10.1038/s41586-020-03065-y.

23. Staley, J. R. *et al.* PhenoScanner: a database of human genotype–phenotype associations. *Bioinformatics* **32**, 3207–3209 (2016).

24. ATGen. https://nkmaxbio.com/sub/shop.php?ptype=view&prdcode=2004011435.

25. Prüfer, K. *et al.* A high-coverage Neandertal genome from Vindija Cave in Croatia. *Science (80-.).* (2017) doi:10.1126/science.aao1887.

26. Meyer, M. *et al.* A high-coverage genome sequence from an archaic Denisovan individual. *Science (80-.).* (2012) doi:10.1126/science.1224344.

27. Sams, A. J. *et al.* Adaptively introgressed Neandertal haplotype at the OAS locus functionally impacts innate immune responses in humans. *Genome Biology* (2016) doi:10.1186/s13059-016-1098-6.

28. Li, H. *et al.* Identification of a Sjögren's syndrome susceptibility locus at OAS1 that influences isoform switching, protein expression, and responsiveness to type I interferons. *PLoS Genet.* (2017) doi:10.1371/journal.pgen.1006820.

29. Liu, X. *et al.* A functional variant in the OAS1 gene is associated with Sjögren's syndrome complicated with HBV infection. *Sci. Rep.* (2017) doi:10.1038/s41598-017-17931-9.

30. Li, Y. I. *et al.* Annotation-free quantification of RNA splicing using LeafCutter. *Nat. Genet.* (2018) doi:10.1038/s41588-017-0004-9.

31. Aguet, F. *et al.* The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science (80-.).* (2020) doi:10.1126/SCIENCE.AAZ1776.

32. Aguet, F. *et al.* Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).

33. Hornung, V., Hartmann, R., Ablasser, A. & Hopfner, K. P. OAS proteins and cGAS: Unifying concepts in sensing and responding to cytosolic nucleic acids. *Nat. Rev. Immunol.* (2014) doi:10.1038/nri3719.

34. Kristiansen, H. *et al.* Extracellular 2′-5′ Oligoadenylate Synthetase Stimulates RNase L-Independent Antiviral Activity: a Novel Mechanism of Virus-Induced Innate Immunity. *J. Virol.* **84**, 11898–11904 (2010).

35. Zeberg, H., and Pääbo, S. A genomic region associated with protection against severe

COVID-19 is inherited from Neandertals. *Proc. Natl. Acad. Sci. U. S. A.* In Press (2021).

36. Mendez, F. L., Watkins, J. C. & Hammer, M. F. Neandertal origin of genetic variation at the cluster of OAS immunity genes. *Mol. Biol. Evol.* (2013) doi:10.1093/molbev/mst004.

37. Bonnevie-Nielsen, V. *et al.* Variation in antiviral 2′,5′-oligoadenylate synthetase (2′5′AS) enzyme activity is controlled by a single-nucleotide polymorphism at a splice-acceptor site in the OAS1 gene. *Am. J. Hum. Genet.* (2005) doi:10.1086/429391.

38. Carey, C. M. *et al.* Recurrent Loss-of-Function Mutations Reveal Costs to OAS1 Antiviral Activity in Primates. *Cell Host Microbe* (2019) doi:10.1016/j.chom.2019.01.001.

39. Zeberg, H. & Pääbo, S. The major genetic risk factor for severe COVID-19 is inherited from Neanderthals. *Nature* (2020) doi:10.1038/s41586-020-2818-3.

40. Schneider, W. M., Chevillotte, M. D. & Rice, C. M. Interferon-Stimulated Genes: A Complex Web of Host Defenses. *Annu. Rev. Immunol.* **32**, 513–545 (2014).

41. Min, J.-Y. & Krug, R. M. The primary function of RNA binding by the influenza A virus NS1 protein in infected cells: Inhibiting the 2'-5' oligo (A) synthetase/RNase L pathway. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 7100–7105 (2006).

42. Hu, B. *et al.* Cellular responses to HSV-1 infection are linked to specific types of alterations in the host transcriptome. *Sci. Rep.* **6**, 28075 (2016).

43. Lim, J. K. *et al.* Genetic Variation in OAS1 Is a Risk Factor for Initial Infection with West Nile Virus in Man. *PLoS Pathog.* **5**, e1000321 (2009).

44. Simon-Loriere, E. *et al.* High Anti–Dengue Virus Activity of the OAS Gene Family Is Associated With Increased Severity of Dengue. *J. Infect. Dis.* **212**, 2011–2020 (2015).

45. Hamano, E. *et al.* Polymorphisms of interferon-inducible genes OAS-1 and MxA associated with SARS in the Vietnamese population. *Biochem. Biophys. Res. Commun.* **329**, 1234–1239 (2005).

46. Cheng, G. *et al.* Pharmacologic Activation of the Innate Immune System to Prevent Respiratory Viral Infections. *Am. J. Respir. Cell Mol. Biol.* **45**, 480–488 (2011).

47. Harari, D., Orr, I., Rotkopf, R., Baranzini, S. E. & Schreiber, G. A robust type I interferon gene signature from blood RNA defines quantitative but not qualitative differences between three major IFNβ drugs in the treatment of multiple sclerosis. *Hum. Mol. Genet.* **24**, 3192–3205 (2014).

48. Lin, F. ching & Young, H. A. Interferons: Success in anti-viral immunotherapy. *Cytokine and Growth Factor Reviews* (2014) doi:10.1016/j.cytogfr.2014.07.015.

49. Arabi, Y. M. *et al.* Interferon Beta-1b and Lopinavir–Ritonavir for Middle East Respiratory Syndrome. *N. Engl. J. Med.* NEJMoa2015294 (2020) doi:10.1056/NEJMoa2015294.

50. Pan, H. *et al.* Repurposed antiviral drugs for COVID-19 –interim WHO SOLIDARITY trial results. *medRxiv* 2020.10.15.20209817 (2020) doi:10.1101/2020.10.15.20209817.

51. Monk, P. D. *et al.* Safety and efficacy of inhaled nebulised interferon beta-1a (SNG001) for treatment of SARS-CoV-2 infection: a randomised, double-blind, placebo-controlled, phase 2 trial. *Lancet Respir. Med.* (2020) doi:10.1016/S2213-2600(20)30511-7.

52. Wood, E. R. *et al.* The role of phosphodiesterase 12 (PDE12) as a negative regulator of

the innate immune response and the discovery of antiviral inhibitors. *J. Biol. Chem.* **290**, 19681–19696 (2015).

53. Silverman, R. H. & Weiss, S. R. Viral phosphodiesterases that antagonize double-stranded RNA signaling to RNase L by degrading 2-5A. *Journal of Interferon and Cytokine Research* (2014) doi:10.1089/jir.2014.0007.

54. Zhao, L. *et al.* Antagonism of the interferon-induced OAS-RNase L pathway by murine coronavirus ns2 protein is required for virus replication and liver pathology. *Cell Host Microbe* (2012) doi:10.1016/j.chom.2012.04.011.

55. Zhang, R. *et al.* Homologous 2',5'-phosphodiesterases from disparate RNA viruses antagonize antiviral innate immunity. *Proc. Natl. Acad. Sci. U. S. A.* (2013) doi:10.1073/pnas.1306917110.

56. Li, Y. *et al.* SARS-CoV-2 induces double-stranded RNA-mediated innate immune responses in respiratory epithelial derived cells and cardiomyocytes. *bioRxiv Prepr. Serv. Biol.* (2020) doi:10.1101/2020.09.24.312553.

57. Burgess, S., Davies, N. M. & Thompson, S. G. Bias due to participant overlap in two-sample Mendelian randomization. *Genet. Epidemiol.* 1–12 (2016) doi:10.1002/gepi.21998.

58. Chang, C. C. *et al.* Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* (2015) doi:10.1186/s13742-015-0047-8.

59. Auton, A. *et al.* A global reference for human genetic variation. *Nature* (2015) doi:10.1038/nature15393.

60. Hemani, G. *et al.* The MR-Base platform supports systematic causal inference across the human phenome. *Elife* **7**, e34408 (2018).

61. Candia, J. *et al.* Assessment of Variability in the SOMAscan Assay. *Sci. Rep.* (2017) doi:10.1038/s41598-017-14755-5.

62. Technical Note SOMAscan Data Standardization. https://somalogic.com/wp-content/uploads/2017/06/SSM-071-Rev-0-Technical-Note-SOMAscan-Data-Standardization.pdf.

63. Duration of Isolation and Precautions for Adults with COVID-19. https://www.cdc.gov/coronavirus/2019-ncov/hcp/duration-isolation.html.

64. Cevik, M. *et al.* SARS-CoV-2, SARS-CoV, and MERS-CoV viral load dynamics, duration of viral shedding, and infectiousness: a systematic review and meta-analysis. *The Lancet Microbe* (2020) doi:10.1016/s2666-5247(20)30172-5.