



# Multi-omics highlights ABO plasma protein as a causal risk factor for COVID-19

Ana I. Hernández Cordero<sup>1</sup> · Xuan Li<sup>1</sup> · Stephen Milne<sup>1,2,3</sup> · Chen Xi Yang<sup>1</sup> · Yohan Bossé<sup>4</sup> · Philippe Joubert<sup>4</sup> · Wim Timens<sup>5</sup> · Maarten van den Berge<sup>6</sup> · David Nickle<sup>7,8</sup> · Ke Hao<sup>9</sup> · Don D. Sin<sup>1,2</sup>

Received: 23 September 2020 / Accepted: 4 February 2021 / Published online: 19 February 2021  
© The Author(s), under exclusive licence to Springer-Verlag GmbH, DE part of Springer Nature 2021

## Abstract

SARS-CoV-2 is responsible for the coronavirus disease 2019 (COVID-19) and the current health crisis. Despite intensive research efforts, the genes and pathways that contribute to COVID-19 remain poorly understood. We, therefore, used an integrative genomics (IG) approach to identify candidate genes responsible for COVID-19 and its severity. We used Bayesian colocalization (COLOC) and summary-based Mendelian randomization to combine gene expression quantitative trait loci (eQTLs) from the Lung eQTL ( $n = 1,038$ ) and eQTLGen ( $n = 31,784$ ) studies with published COVID-19 genome-wide association study (GWAS) data from the COVID-19 Host Genetics Initiative. Additionally, we used COLOC to integrate plasma protein quantitative trait loci (pQTL) from the INTERVAL study ( $n = 3,301$ ) with COVID-19 loci. Finally, we determined any causal associations between plasma proteins and COVID-19 using multi-variable two-sample Mendelian randomization (MR). The expression of 18 genes in lung and/or blood co-localized with COVID-19 loci. Of these, 12 genes were in suggestive loci ( $P_{\text{GWAS}} < 5 \times 10^{-05}$ ). *LZTFL1*, *SLC6A20*, *ABO*, *IL10RB* and *IFNAR2* and *OAS1* had been previously associated with a heightened risk of COVID-19 ( $P_{\text{GWAS}} < 5 \times 10^{-08}$ ). We identified a causal association between *OAS1* and COVID-19 GWAS. Plasma ABO protein, which is associated with blood type in humans, demonstrated a significant causal relationship with COVID-19 in the MR analysis; increased plasma levels were associated with an increased risk of COVID-19 and, in particular, severe COVID-19. In summary, our study identified genes associated with COVID-19 that may be prioritized for future investigations. Importantly, this is the first study to demonstrate a causal association between plasma ABO protein and COVID-19.

## Introduction

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is responsible for the coronavirus disease 2019 (COVID-19) and the current world pandemic.

The SARS-CoV-2 was first identified in Wuhan, Hubei province, China in late 2019 and since then, it has spread across the world, affecting more than 180 countries (Zhu et al. 2020). To date, the COVID-19 health crisis has resulted in the loss of more than a million human lives

✉ Ana I. Hernández Cordero  
Ana.Hernandez@ubc.hli.ca

<sup>1</sup> Centre for Heart Lung Innovation, University of British Columbia, St. Paul's Hospital, Vancouver, BC, Canada

<sup>2</sup> Division of Respiratory Medicine, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada

<sup>3</sup> Faculty of Medicine and Health, University of Sydney, Sydney, NSW, Australia

<sup>4</sup> Institut Universitaire de Cardiologie Et de Pneumologie de Québec-Université Laval, Québec City, QC, Canada

<sup>5</sup> Department of Pathology and Medical Biology, University of Groningen, University Medical Centre Groningen, Groningen, The Netherlands

<sup>6</sup> Department of Pulmonary Diseases, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

<sup>7</sup> Global Health, University of Washington, Seattle, WA, USA

<sup>8</sup> Gossamer Bio, 3013 Science Park Road, San Diego, CA, USA

<sup>9</sup> Department of Genetics and Genomic Sciences and Icahn Institute for Data Science and Genomic Technology, Icahn School of Medicine At Mount Sinai, New York, NY, USA

(Dong et al. 2020). While the disease has mild effects in most individuals, severe COVID-19 is more likely in the elderly population and individuals with comorbidities such as cardiovascular diseases and diabetes (Zhou et al. 2020a). Why these populations are at a higher risk of adverse COVID-19 outcomes is still unclear.

Genetic variations in the host may explain some of the heterogeneity in COVID-19 outcomes. Understanding the role of genetic variants may also provide critical insights into COVID-19 pathogenesis. However, the genes and pathways that contribute to SARS-CoV-2 infection are poorly understood. Recently, a pooled genome-wide association study (GWAS) revealed novel susceptibility loci, including 3p21.31 and 9q34.2, for severe COVID-19 leading to respiratory failure (Ellinghaus et al. 2020). These loci encompassed several genes, but which (if any) of the individual genes are causally associated with COVID-19 was not explored.

GWAS leverages genetic variants to determine associations between regions of the genome and a particular trait (e.g. disease). However, traditional GWAS has several limitations that may prevent true gene–trait associations from being identified. For example, due to correction for multiple comparisons, GWAS results necessarily have a stringent threshold for statistical significance and relevant associations below this threshold may be missed. Second, disease-associated loci are typically thousands of base pairs wide and thus contain multiple genes, which may obscure the causal gene contributing to the disease trait. One emerging approach to identify genes within susceptibility loci is integrative genomics (IG). By combining genomic information with transcriptomic, proteomic and/or methylation data, IG is able to fine-map genetic susceptibility loci and identify genes and proteins most likely to have a causal association with disease. This method has previously elucidated specific protein coding genes and mechanisms that contribute to complex traits (Cano-Gamez and Trynka 2020). In the present study, we harnessed the power of IG to identify several susceptibility genes for COVID-19 and investigate the causal relationship between the plasma protein levels of candidate genes and COVID-19. Notably, here, we demonstrate a causal association of the ABO protein (this protein is responsible for the ABO blood groups) with both the risk for COVID-19 and its severity.

## Methods

### Study pipeline

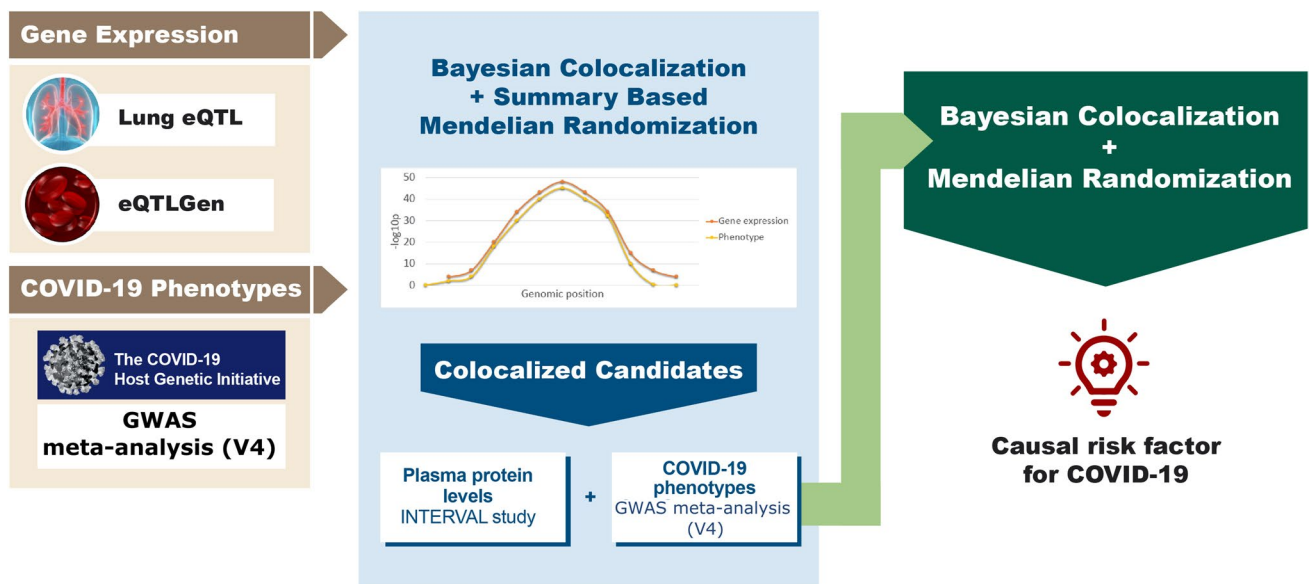
The COVID-19 Host Genetic Initiative (COVID-19 HG) GWAS identified four genetic loci within three chromosomes

(chromosome 3, 9 and 21), which were associated with susceptibility to COVID-19 (defined as a positive COVID-19 diagnosis versus the general population) and seven genetic loci (chromosome 3, 6, 7, 9, 12, 19 and 21), which were associated with severe COVID-19 (defined as hospitalization for COVID-19 versus the general population) at a genome-wide significant threshold of  $P_{\text{GWAS}} < 5 \times 10^{-08}$  (The COVID-19 Host Genetics Initiative 2020). In addition to these loci, we also explored regions below this stringent threshold, particularly those with GWAS  $P_{\text{GWAS}} < 5 \times 10^{-05}$ . We integrated these GWAS results with gene expression data from both lung tissue (Lung eQTL study) (Hao et al. 2012) and blood (eQTLGen) (Võsa et al. 2018) using two statistical methodologies (Fig. 1). Bayesian colocalization (Coloc) assessed whether two genetic association signals are consistent with a shared causal variant (Giambartolomei et al. 2014). We defined colocalization as the posterior probability of this hypothesis ( $PP_{H4}$ ) being  $> 0.80$ . Summary Based Mendelian Randomization (SMR) integrates summary data from GWAS and eQTLs in order to identify genes whose expression levels are associated with a trait due to the effects of a common genetic variant (either by direct causal or pleiotropic effects) rather than due to a genetic linkage (Zhu et al. 2016). Significance of SMR estimate was set at  $P_{\text{SMR}} < 0.001$  with no significant heterogeneity ( $P_{\text{HEIDI}} > 0.05$ ).

### Study cohorts

#### COVID-19 host genetics initiative (COVID-19 HG)

For our study, we obtained publicly available summary statistics from the COVID-19 HG GWAS meta-analysis V4 (<https://www.covid19hg.org/>) (The COVID-19 Host Genetics Initiative 2020). We obtained the summary statistics for two case–control analyses: (1) “susceptibility to COVID-19” [where cases were all individuals with a diagnosis of COVID-19 ( $n = 17,965$ ), and controls were all individuals without a COVID-19 diagnosis ( $n = 1,370,547$ )]; and (2) “Severe COVID-19” [where cases were individuals with a COVID-19 diagnosis and who were hospitalized ( $n = 7,885$ ), and controls were all individuals without a COVID-19 diagnosis and no hospitalization ( $n = 961,804$ )]. In brief, the COVID-19 HG is an ongoing initiative that aims to facilitate COVID-19 genetic host research through international collaboration. Contributing investigators submit individual level data or summary results from GWASs performed according to pre-specified methodological standards to ensure quality and consistency of the data. For imputation, each individual study used their own reference panel or existing imputation panel; the association analyses were adjusted for age, sex and population structure. The COVID-19 HG combines this data in an inverse variance weighted meta-analysis using variants with a minor allele frequency



**Fig. 1** Study overview. The diagram summarizes the genomics datasets and analytic pipeline of the study. First, publicly available omics datasets were obtained, which were later processed using integrative genomics (IG) methods (Bayesian Colocalization and Summary-based Mendelian Randomization) to identify potential candidate

genes for COVID-19 phenotypes. Lastly, using a Bayesian Colocalization and Mendelian Randomization approaches, we explored the causal association between the plasma protein levels of the most promising candidate gene and the risk of COVID-19

(MAF) > 0.001 and imputation quality ( $r^2$ ) > 0.6. Note that the summary statistics used for our analysis did not include the 23andMe cohort due to data privacy issues. Further details are provided by the COVID-19 HG (The COVID-19 Host Genetics Initiative 2020).

### Lung eQTL study

For our study, we obtained lung expression quantitative loci (eQTL) from the Lung eQTL study (Hao et al. 2012). In summary, the Lung eQTL study consisted of 1038 participants from three institutions, the University of British Columbia (UBC), Laval University and University of Groningen. At UBC and Laval University, the studies were approved by the ethics committees within each institution. For University of Groningen, the lung specimens were provided by the local tissue bank of the Department of Pathology; the study protocol that was used in this study was consistent with the Research Code of the University Medical Center Groningen and Dutch national ethical and professional guidelines (“Code of conduct; Dutch federation of biomedical scientific societies”, <http://www.federa.org>). The Lung eQTL study determined the gene expression of non-tumour lung tissue samples using 43,466 non-control probe sets (see GEO platform GPL10379). The participants were also genotyped using the Illumina Human 1 M Duo BeadChip, and after imputation a total of 7,640,142 single nucleotide polymorphisms (SNPs) for Laval,

7,610,179 for UBC, and 7,741,505 for Groningen were kept for an eQTL analysis. Data for each site were evaluated, separately, using a robust linear regression model adjusted for age, sex and smoking status, which assumed an additive genotype effect. No population structure adjustment was applied to the individual analyses because of the homogeneity of the cohort within each site (white European ancestry) and the small sample size (Hao et al. 2012). Site-specific results were then combined by a meta-analysis using a fixed effects model with inverse variance weighting to account for any heterogeneity between sites. Significant eQTLs were defined at a false discovery rate (FDR) < 0.10; *cis* expression quantitative trait loci (*cis*-eQTLs) were defined by a 2 Mb window ( $\pm 1$  Mb probe to SNP distance). Full details of the cohort and genotyping quality control, and eQTL analysis are provided by Hao and colleagues (Hao et al. 2012).

### eQTLGen

We obtained blood *cis*-eQTL summary statistics from eQTLGen. The eQTLGen cohort consisted of 31,684 whole blood (85%) and peripheral blood mononuclear cell (15%) samples from 37 datasets. Gene expression profiles and genotypes were obtained for the eQTLGen cohort. The full details on the participants, gene expression measurements and genotyping for each dataset are described by the eQTLGen consortium (Võsa et al. 2018). Based on the first two

principal components of gene expression, the outliers were removed from the eQTLGen analysis. To account for population structure, the eQTLGen study regressed the first four multidimensional scaling (MDS) components of the expression matrix. The *cis*-eQTL analysis was performed in each separate dataset and were estimated within a 2 Mb window ( $\pm 1$  Mb probe to SNP distance) as previously described by Westra and colleagues (Westra et al. 2013); later, the results were combine by a meta-analysis using a weighted Z-score method (Westra et al. 2013; Vösa et al. 2018). Significant eQTLs were defined at a false discovery rate (FDR)  $< 0.05$ .

### INTERVAL study

We used plasma protein quantitative trait loci (pQTL) obtained from the INTERVAL study (Sun et al. 2018). In brief, the INTERVAL study was a randomized trial of approximately 50,000 participants across 25 static donor centres of the NHS Blood and Transplant (NHSBT) Network (Di Angelantonio et al. 2017; Sun et al. 2018). Blood was collected from the participants using standard venepuncture. The Affymetrix Axiom UK Biobank array was used for the genotyping of 830,000 SNPs and genotypes were later imputed using the 1000 Genome phase 3 UK10K reference panel. After quality control, 10,572,788 SNPs were retained. A randomly selected subset of 3,301 participants were used for the plasma pQTL analyses of 3,622 proteins (Sun et al. 2018). Plasma protein levels were measured using an expanded version of an aptamer-based multiplex protein assay (SOMAscan), which was previously described by Sun and colleagues (2018). The protein levels were adjusted for confounding variables (age, sex, waiting period between blood collection and processing and the first three genetic principal components) and the residuals were extracted and rank-inverse normalized; pQTL analysis involved testing the association between plasma protein levels and genetic variants with a linear regression using an additive genetic model. The results from each donor centers were combined using a fixed-effect inverse-variance meta-analysis. Significant pQTLs were identified at a meta-analysis  $P$  value  $< 1.5 \times 10^{-11}$  (Sun et al. 2018). Further details on the study cohort, genotyping protocol and quality control are described by Sun and colleagues (Sun et al. 2018).

### Integrative-omics methods

#### Bayesian colocalization test (Coloc)

We first conducted Coloc tests to determine the probability that SNPs associated with COVID-19 phenotypes and gene expression (eQTLs) were shared genetic causal variants (colocalization). This IG method estimates the ‘posterior probabilities’ (PP) of five hypotheses: (1) a genetic locus has

no associations with either of the two traits (i.e. gene expression and a complex trait) investigated ( $H_0$ ); (2) the locus is associated only with gene expression ( $H_1$ ); (3) the locus is associated only with the complex trait ( $H_2$ ); (4) the locus is associated with both traits via independent SNPs ( $H_3$ ); and (5) the locus is associated with both traits through shared SNPs ( $H_4$ ) (i.e.: a SNP is associated with COVID-19 and is also a *cis*-eQTL). Colocalization is, therefore, indicated by a high PP of  $H_4$  being true. For these analyses, we used the coloc package (Giambartolomei et al. 2014) implemented in R. We tested only *cis*-eQTL regions ( $\pm 1$  Mb probe to SNP distance). As required by the method and recommended by Giambartolomei and colleagues (Giambartolomei et al. 2014), we set the ‘prior probability’ of the various configurations ( $H_1$ ,  $H_2$ , and  $H_4$ ). For the eQTL dataset, we used  $1 \times 10^{-04}$  prior probability for a *cis*-eQTL ( $H_1$ ). We also used  $1 \times 10^{-04}$  prior probability for COVID-19 associations ( $H_2$ ). Finally, we set a prior probability that a single variant affects both traits ( $H_4$ ) at  $1 \times 10^{-06}$ . We set significant colocalization (posterior probability) at  $PP_{H_4} > 0.80$ , in addition we further prioritized colocalized genes based on the  $P_{\text{GWAS}} < 5 \times 10^{-05}$ . We executed coloc between the loci associated with COVID-19 (susceptibility and disease severity) and *cis*-eQTLs that were associated with gene expression in both lung and blood tissues, retaining genes whose expression colocalized with COVID-19 in both compartments as ‘candidate genes’. The number of the SNPs included in the Coloc analysis was determined based on the *cis*-QTLs width as defined by Hao and colleagues (Hao et al. 2012). If the corresponding proteins of the candidate genes were present in the plasma protein dataset (INTERVAL study), we executed Coloc analysis for the plasma protein levels and COVID-19 phenotypes; the number of the SNPs included in this Coloc analysis was determined based on a 1 Mb window (pQTL SNP bp position  $\pm 500,000$  bp).

#### Summary-based mendelian randomization (SMR)

The SMR method was specifically built to test the association between gene expression and a complex trait using a SNP as the instrument (Zhu et al. 2016). SMR is based on the standard Mendelian Randomization (MR) analysis, where the effect of genetic variants is linked to the trait of interest via an exposure (gene expression) (Swerdlow et al. 2016). Therefore, we employed an SMR method to identify genes whose effect on COVID-19 was mediated by their expression in lung and blood. For the SMR, the lung and blood *cis*-eQTLs and COVID-19 HG GWAS meta-analysis summary statistics were used. We selected the 1000G phase 3 EUR (1000 Genomes Project Consortium et al. 2015) as the reference panel for linkage disequilibrium (LD) estimation. Significant SMR was defined at  $P_{\text{SMR}} < 0.001$ . The significant SMR by itself does not necessarily indicate that the



same variants are associated with the gene expression and the phenotype; the association could be a consequence of the LD between independent causal variants, rather than pleiotropy of a single causal variant or causality. To determine whether the associations were related to LD we used the heterogeneity in dependent instruments test (HEIDI) test (Zhu et al. 2016), the null hypothesis of which is that the effect of a variant is shared in gene expression and the phenotype; rejecting the null hypothesis based on the  $P$  value ( $P_{\text{HEIDI}}$ ) is interpreted as evidence of heterogeneity (linkage) (Zhu et al. 2016). We reported the significant SMR associations that also showed  $P_{\text{HEIDI}} \geq 0.05$ .

### Mendelian randomization: ABO and OAS1 plasma proteins and COVID-19

To test causality of the ABO plasma protein, we conducted MR tests of the plasma protein on COVID-19 phenotypes. MR is based on the unidirectional flow of genetic information, which assumes that genetic variants will have an effect on downstream phenotypes (gene  $\rightarrow$  protein  $\rightarrow$  phenotype). MR uses SNPs as instrumental variables (IVs) to link a risk factor ('exposure') to a health trait ('outcome'). The MR assumptions are as follow: IVs are associated with the exposure, and only affect an outcome via the exposure, and are independent of confounders.

A pQTL on chromosome 9q34.2 was identified for ABO plasma protein (Sun et al. 2018). We performed stepwise conditional analysis, using GCTA 1.92.0beta3 (Yang et al. 2011), within each 2 Mb region on chromosome 9 to identify independently associated SNPs. UK Biobank genotypes (Bycroft et al. 2018) were used as the reference sample, and excluded SNPs with  $\text{MAF} < 0.01$ . We later extracted the effect size (beta) and standard error (SE) for each independent variant. Likewise, we obtained the beta and SE for each of these SNPs on the two COVID-19 phenotypes in the COVID-19 HG meta-analysis. We used stepwise removal to identify the variants with homogeneous effects to fulfil the MR assumption of homogeneity of the IVs. We then performed an inverse variance weighted (IVW) MR (IVW-MR) (implemented in 'MendelianRandomization' R package (Staley 2020)) to link the allelic effects on the exposure (plasma protein levels) to their effects on the outcome (severe COVID-19 and COVID-19 diagnosis). We adjusted for possible correlations between SNPs by incorporating the estimated LD on 1000G phase 3 EUR using PLINK 1.9 (Chang et al. 2015) into the MR input. The IVW-MR assumes no directional pleiotropy (i.e. genetic variant associated with multiple unrelated phenotypes) (Bowden et al. 2015). Significance of the MR estimate was set at  $P < 0.05$ . To assess the presence of pleiotropy, we then executed MR-Egger analysis (Bowden et al. 2015) which, unlike IVW-MR, allows the estimation of directional pleiotropy.

The presence of pleiotropy is suggested by a significant ( $P_{\text{Egger Intercept}} < 0.05$ ), non-zero, intercept term.

We assessed heterogeneity of the IVs using the Cochran's Q test obtained from the IVW-MR output (Staley 2020); significant heterogeneity ( $P < 0.05$ ) suggests that the variability of the IVs estimates is greater than would be expected by chance alone, which may indicate bias due to invalid IVs. Although a MR with multiple IVs has greater power compared to a single IV MR, it is possible that the average causal effect may be mainly driven by the SNP with the strongest association with the outcome. To evaluate the reliance of the multi-variable MR on individual SNPs we performed a sensitivity analysis by excluding each SNP at a time and re-calculating the IVW-MR estimate. If removing a SNP significantly changes the MR estimate, it indicates that the multi-variable MR relied mostly on the removed SNP.

To test causality of the OAS1 plasma protein, we also conducted an IVW-MR of the plasma protein on COVID-19 phenotypes as described before. First, we identified pQTLs for the protein and found two signals [*trans* (chromosome 19) and *cis* (chromosome 12)]. No additional independent variants were identified within the two pQTLs; therefore, only two variants were used for the initial MR. The initial MR result was significant ( $P < 0.05$ ); however, the two IVs showed heterogeneity ( $P < 0.05$ ). Thus, we only used the *cis*-pQTL for a single IVW-MR.

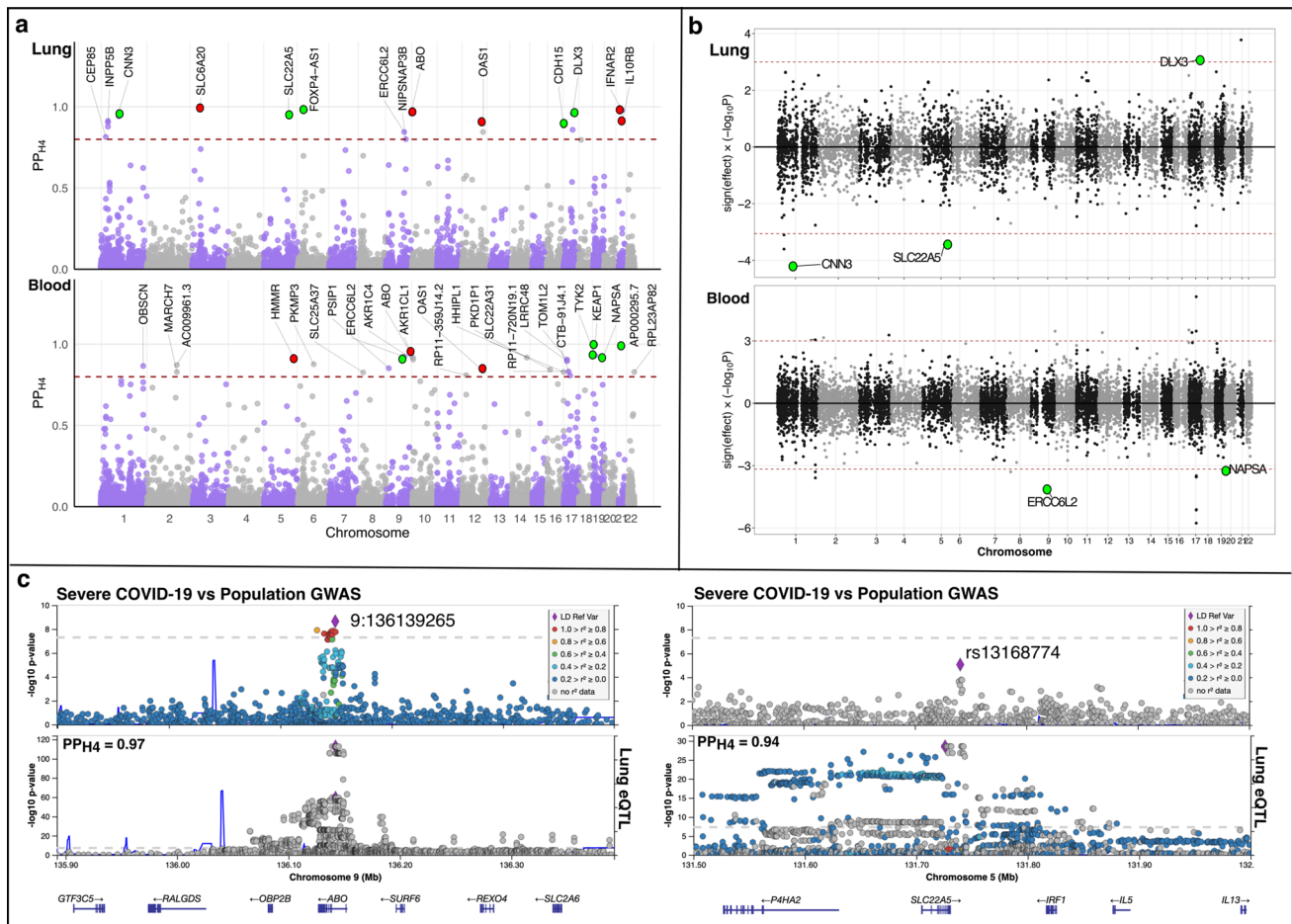
### Additional analyses

We followed the pipeline described in Fig. 1 and repeated the IG analyses using the COVID-19 test positive versus test negative phenotype (The COVID-19 Host Genetics Initiative 2020). The results corresponding to these analyses are shown in Supplementary Fig. S4.

## Results

### Lung gene expression

The expression of ten unique genes in lung tissue co-localized ( $PP_{H4}$  being  $> 0.80$ ) with severe COVID-19-associated loci (Fig. 2a) and 8 with the COVID-19 trait associated loci ( $P_{\text{GWAS}} < 5 \times 10^{-05}$ ) (Supplementary Table S1). Six unique gene associations identified in lung tissue were novel (i.e. not been identified by GWAS) (The COVID-19 Host Genetics Initiative 2020), (*FOXP4-AS1*, *CNN3*, *DLX3*, *SLC22A5*, *CDH15* and *SEPWI*) and met the suggestive statistical significance ( $P_{\text{GWAS}} < 5 \times 10^{-05}$ ) rather than the genome-wide significance ( $P_{\text{GWAS}} < 5 \times 10^{-08}$ ) (Supplementary Table S1). For example: *SLC22A5* is a novel association for COVID-19 that is located in a suggestive locus (Fig. 2c) on chromosome 5 [sentinel SNP rs13168774,  $P_{\text{GWAS}} = 8.6 \times 10^{-06}$ ].



**Fig. 2** COVID-19 genomics and gene expression integration. **a** Colocalization of COVID-19 (hospitalization vs population) with gene expression in lung and blood tissues, respectively. The circles represent the probability ( $PP_{H4}$ ) that a gene colocalizes with COVID-19 plotted against its chromosomal position ( $x$  axis). The red dashed horizontal line represents the threshold of significance ( $PP_{H4} > 0.80$ ). Red and green circles highlight the genes within previously identified COVID-19 loci, and those within suggestive COVID-19 loci ( $P_{GWAS} < 5 \times 10^{-05}$ ), respectively. **b** Results from the Summary Based Mendelian Randomization are displayed in this mirror Manhattan plot. The circles represent the association between susceptibility to COVID-19 (diagnosis vs population) and the gene expression (lung and blood) multiplied by the direction of the effect ( $y$  axis) plotted against the genes chromosomal position ( $x$  axis). The

dotted horizontal lines (red) represent the threshold of significance ( $P < 0.001$ ). SMR plot only shows the results that passed the heterogeneity test (see methods). Labels represent those genes that were also co-localized. **c** Regional plots for severe COVID-19 (COVID-19 hospitalization vs population) and lung tissue *cis*-eQTL at the *ABO* (left panel) and *SLC22A5* (right panel) loci. Severe GWAS signals (top of the panel **c**) colocalize with the *cis*-QTL region (bottom of the panel **c**) at *ABO* (chromosome 9) and *SLC22A5* (chromosome 5). The circles represent the  $-\log_{10}$  association  $P$  values ( $y$  axis) of SNPs plotted against their chromosomal position ( $x$  axis). The co-localized SNP is shown inside each plot and  $PP_{H4}$  is at the left top corner of the lung eQTL regional plots. Pairwise linkage disequilibrium ( $r^2$ , calculated from the 1000 Genomes European population) in each region is colored with respect to the lead GWAS SNP within each region

(COVID-19 hospitalization vs population)]; this gene has been associated with asthma (Moffatt et al. 2010), IgG glycosylation (Lauc et al. 2013), and systemic carnitine deficiency (Mutlu-Albayrak et al. 2015). Furthermore, based on the SMR analysis, we determined that *SLC22A5* expression in lung tissue is associated to a decreased risk of COVID-19 (Fig. 2b). SMR also showed that *CNN3* gene expression in the lung decreases the risk of severe COVID-19, while *DLX3* increases the risk. In addition to these novel gene associations, we identified three co-localized genes (*LZTFL1*, *SLC6A20* and *ABO*) that were within loci, which

had been previously associated with severe COVID-19 by GWAS (Ellinghaus et al. 2020) (Fig. 2a and Supplementary Table S1). *ABO* gene expression co-localized with susceptibility and severe COVID-19 Fig. 2c.

Other candidates that have been associated with COVID-19 included *IL10RB* and *IFNAR2* (Fig. 2a) (Ma et al. 2020, p. 19; Pairo-Castineira et al. 2020), our study expands on these by associating lung gene expression with COVID-19 risk and providing the direction of the effect for *IFNAR2* on COVID-19. *IL10RB* and *IFNAR2* are interferon (IFN) receptor genes that co-localized with COVID-19 ( $PP_{H4} > 0.80$ );

these genes were located in the same locus on chromosome 21 [sentinel SNP rs13050728,  $P_{\text{GWAS}} = 1.9 \times 10^{-11}$  (COVID-19 hospitalization vs population) (The COVID-19 Host Genetics Initiative 2020)]. In addition, results from the SMR showed that the increased expression of *IFNAR2* in lung tissue was associated with decreased risk for COVID-19 (Supplementary Table S2) ( $P_{\text{SMR}} < 0.001$ ). We also identified colocalization of COVID-19 risk and severity with the *OAS1* gene locus (Fig. 2 and Supplementary Table S1), *OAS1* is an interferon stimulated gene involved in the cellular response to viral infection.

## Blood gene expression

In blood, the expression levels of 8 genes co-localized with COVID-19 severity (Fig. 2a); whereas only two genes co-localized to the risk of COVID-19 ( $P_{\text{GWAS}} < 5 \times 10^{-05}$ ) (Supplementary Table S1). The expression of *ABO* in blood co-localized with the risk for COVID-19 ( $PP_{H4} = 0.86$ ) as well as that for severe COVID-19 ( $PP_{H4} = 0.94$ ) associated loci (Sentinel SNP 9:136,146,597,  $P_{\text{GWAS}} = 7.3 \times 10^{-08}$ ). Based on the Coloc test, *OAS1* was also highlighted as a candidate in blood for COVID-19 severity ( $PP_{H4} = 0.83$ ) and susceptibility ( $PP_{H4} = 0.95$ ). Other first-time associations within COVID-19 GWAS suggestive loci included *KEAP1*, *AP000295.7*, *TYK2*, *ERCC6L2*, *NAPSA* and *HMM*. SMR results showed that the expression of *ERCC6L2* and *NAPSA* in blood was associated with decreased risk of severe COVID-19. The full list of SMR results and its details are presented in Supplementary Table S2.

## ABO plasma protein level is a risk factor for COVID-19

The functional effects of genes are generally imparted through their translation into proteins. To strengthen the mechanistic association between the identified genes and COVID-19, we determined which of these genes were associated with both blood protein levels and COVID-19 phenotypes. We integrated the COVID-19 HG GWAS and blood protein GWAS from the INTERVAL study (which includes genome-wide associations between genetic variants and 2,995 blood proteins) using Coloc. Additionally, we applied MR to determined causal associations between protein and COVID-19.

The expression of two protein-coding candidate genes (*ABO* and *OAS1*) co-localized with COVID-19 phenotypes in both tissues ( $P_{\text{GWAS}} < 5 \times 10^{-05}$  and  $PP_{H4} > 0.80$ ). Both genes were present in the INTERVAL study. We found that ABO plasma protein levels co-localized with susceptibility ( $PP_{H4} = 0.99$ ) to COVID-19 and its severity ( $PP_{H4} = 0.99$ ) (Fig. 3a). Likewise, *OAS1* plasma protein levels co-localized

with the COVID-19 phenotypes (Supplementary Table S3), although at a lower probability than ABO [susceptibility ( $PP_{H4} = 0.88$ ), severity ( $PP_{H4} = 0.81$ )].

An IVW-MR test was conducted using three independent variants as instrumental variables (IVs) to investigate the causal relationship between ABO plasma protein and COVID-19. Based on the Cochran's Q and MR-Egger intercept p values the IV's average effect on COVID-19 phenotypes did not demonstrate significant heterogeneity or horizontal pleiotropy amongst the IVs (Supplementary Table S4). The results indicated a significant causal association between ABO plasma protein levels and COVID-19 phenotypes ( $P < 0.05$ ) (Fig. 3b and Supplementary Figure S1). ABO-increasing allele increased the odds of COVID-19 by 0.06 and the odds of severe COVID-19 by 0.10. In addition, we performed a sensitivity analysis to assess if the MR results were driven by a single SNP. Supplementary Figures S2 and S3 show that the exclusion of any of the IVs used for the MR did not significantly change the overall IVW-MR estimate.

A single variant (rs4767027) IVW-MR (see "Methods") also showed that *OAS1* plasma protein level is causally associated with susceptibility to COVID-19 ( $P_{\text{MR}} = 4.00 \times 10^{-05}$ , MR estimate = -0.21) and its severity ( $P_{\text{MR}} = 7.08 \times 10^{-07}$ , MR estimate = -0.42). The MR effect direction shows that plasma *OAS1*-increasing allele (rs4767027) decreases the odds of COVID-19.

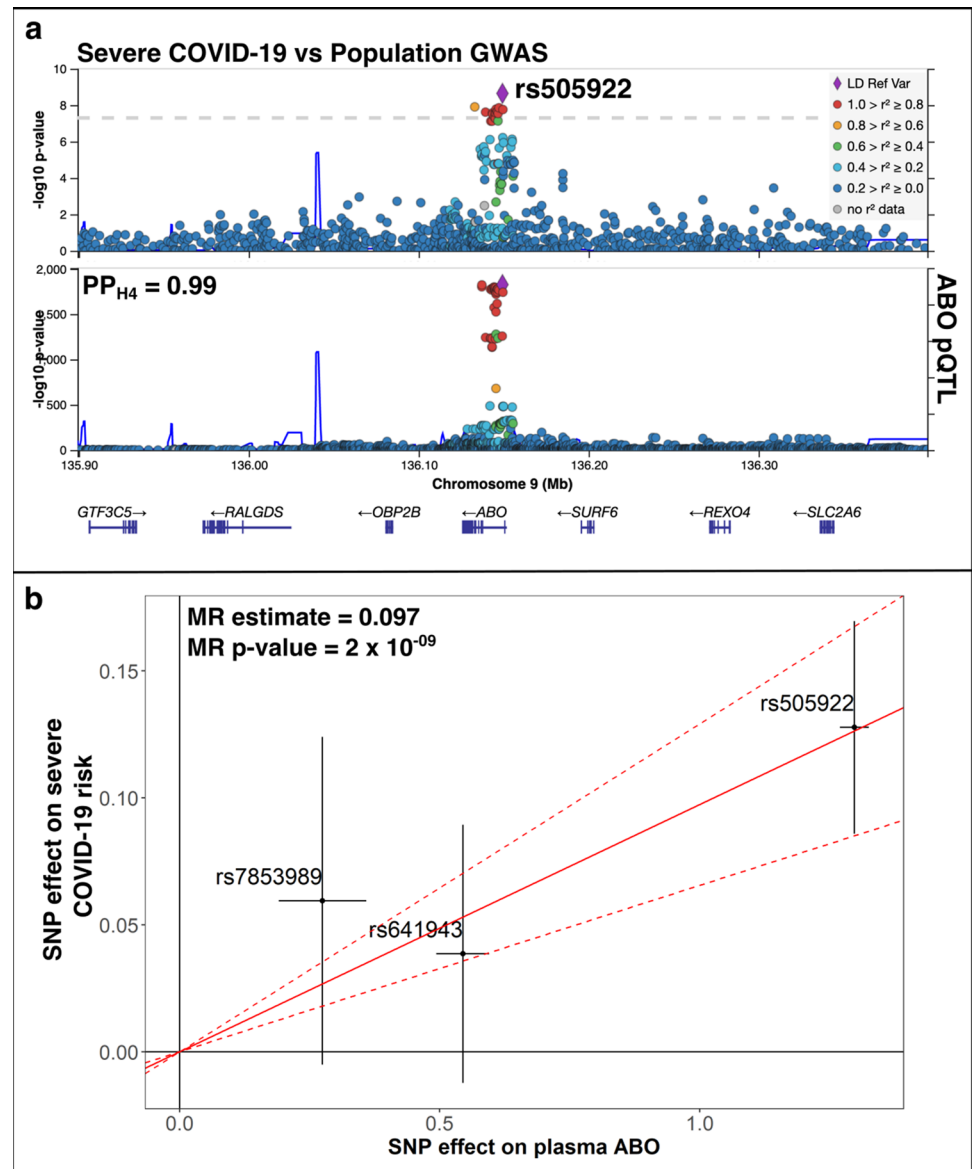
## Additional analyses

We analysed an additional phenotype (COVID-19 test positive versus test negative) to check for the consistency of the different COVID-19 phenotypes. Five genes (*DNP1*, *FCER1G*, *ABO*, *SLC6A20* and *SEPWI*) identify with this COVID-19 test positive versus test negative GWAS coincided with our findings on COVID-19 susceptibility and severity (Supplementary Figure S4).

## Discussion

Recent GWAS have revealed genetic loci that are significantly associated with COVID-19 (The COVID-19 Host Genetics Initiative 2020; Ellinghaus et al. 2020); however, since these loci often harbor multiple genes, the underlying genes or pathways responsible for COVID-19 remain unknown. To address this crucial gap in knowledge, we used integrative genomic methods, which unveiled several novel findings. First, we identified gene associations whose physiology plausibly relates to COVID-19. Second, we associated the effect of genetic variants to the gene expression levels of previously identified COVID-19-related-genes. Third, we showed that genetically-determined plasma ABO and

**Fig. 3** Bayesian Colocalization (a) and Mendelian Randomization (MR) (b) of ABO plasma protein and COVID-19. **a** Regional plot describing severe COVID-19 GWAS (COVID-19 hospitalization vs population) (top of **a**) and plasma protein pQTL (bottom of **a**) at the *ABO* locus. The circles represent the  $-\log_{10}$  association  $P$  values (y axis) of SNPs plotted against their chromosomal position (x axis). The co-localized SNP is shown inside each plot and  $PP_{H4}$  is at the left top corner of the plasma protein pQTL regional plot. Pairwise linkage disequilibrium ( $r^2$ , calculated from the 1000 Genomes European population) in each region is colored with respect to the lead GWAS SNP within each region. **b** Inverse variance weighting (IVW) MR (IVW-MR) of ABO plasma proteins on risk of severe COVID-19 (COVID-19 hospitalization vs population). The x axis represents the SNP effect on the plasma protein levels, and the y axis the SNP effect on severe COVID-19. The variants used for the MR are shown inside the plot with error bars that represent the 95% confidence intervals. The slope of the solid red line is the instrumental variables regression estimate of the effect of the protein on severe COVID-19, with dashed red lines representing the 95% confident interval. IVW-MR  $P$  value and estimate are shown in the top left corner



*OAS1* proteins levels are causally associated with the risk of COVID-19 and its severity.

Although the chromosome 3p21 region has been previously associated with severe COVID-19, this locus encompasses six genes, making it hard to identify the precise gene responsible for the association with COVID-19 (Ellinghaus et al. 2020). We found that two genes within this locus (*SLC6A20* and *LZTFL1*) co-localized with COVID-19 phenotypes. The *SLC6A20* protein product (sodium- and chloride-dependent transporter XTRP3) is involved in amino acid transport, with a role in the regulation of thymocyte selection (Simeoni et al. 2005) and negative regulation of T-cell activation (Arndt et al. 2011). This protein may also interact with angiotensin converting enzyme 2 (*ACE2*), which is the putative SARS-CoV-2 receptor (Vuille-dit-Bille et al. 2015; Zhou et al. 2020b). Interestingly, both

*ACE2* and *SLC6A20* gene and protein expression levels increase with age (Meier et al. 2018; Bunyavanich et al. 2020; Vuille-dit-Bille et al. 2020) with the lowest concentrations noted in children (small intestinal and nasal epitheliums). Although children can develop SARS-CoV-2 infections, they appear to be at very low risk of developing severe COVID-19 (Jordan et al. 2020). Consistent with these findings, another recent study has shown that a risk variant (rs11385942) for *SLC6A20* expression in human lung cells conferred increased risk for severe COVID-19 (Ellinghaus et al. 2020, p. 19). *LZTFL1* encodes leucine zipper transcription factor-like protein 1, a protein involved in intracellular cargo trafficking that is linked to congenital ciliopathies; the mechanism of its association with COVID-19 is unclear.

Of the described gene associations with susceptibility to COVID-19 and its severity, *IL10RB*, *IFNAR2* and *OAS1* are



notable given their role in the IFN pathways. Variants within the vicinity of these genes, particularly within *IFNAR2* and *OAS1* have been associated with COVID-19 phenotypes (Ma et al. 2020, p. 19; Pairo-Castineira et al. 2020). In the case of *OAS1*, previous research has failed to provide a potential mechanistic link. Our analyses suggest that genetic variants within *IL10RB*, *IFNAR2* and *OAS1* exert their effect on COVID-19 through lung and blood gene expression; in addition, this is the first time that *OAS1* plasma protein levels have been associated to COVID-19 susceptibility and its severity. The protein products of *IFNAR2* and *IL10RB* are components of the receptor complexes for type I and III IFNs, respectively, which are critical in the early host responses to a viral infection. The protein product of *OAS1* (2'-5'-oligoadenylate synthase 1), which is induced by IFNs, indirectly promotes viral RNA degradation and inhibition of viral replication. Our findings suggest that increased *OAS1* gene expression decreases the susceptibility to COVID-19 and, in particular, severe COVID-19. This is consistent with the concept of IFN dysregulation in severe COVID-19, in which patients mount a delayed and often blunted IFN responses to the virus (Acharya et al. 2020; Made et al. 2020). Our findings support the results of a recent clinical trial showing that inhaled IFN-beta decreased the risk of severe COVID-19 (Balfour 2020).

Our results also show that a COVID-19 locus was associated with *ABO* gene expression in lung and blood tissues and plasma protein levels in blood. The *ABO* gene encodes for a protein responsible for the ABO blood groups. While the A and B allele carriers express glycosyltransferase activities that convert H antigen into A or B antigen, the O group protein lacks this enzymatic activity [due to a deletion in the gene (frameshift)]. Previous studies have shown that blood type O individuals demonstrated a lower susceptibility to COVID-19 (Zhao et al. 2020; Zietz and Tatonetti 2020; Ellinghaus et al. 2020), and variants in the *ABO* have been associated with increased risk of severe COVID-19 (Ellinghaus et al. 2020). Recent findings reported that the rhesus negative (Rh<sup>-</sup>) blood group may also be protective for SARS-CoV-2 infection and severity, particularly in the O-negative blood group, again showing a potential link between the blood groups and COVID-19 (Ray et al. 2020). While others have previously reported an association between COVID-19 and ABO, here, we showed for the first time that the ABO plasma protein level is likely a causal risk factor for susceptibility to COVID-19 and its severity. The mechanism by which ABO protein modifies COVID-19 risk is unclear. The protein-decreasing allele (T) of the top SNP (rs505922) used in the IVW-MR tests is in high LD ( $r \sim 1$ ) with the O blood group SNP genotype (rs8176719) (Melzer et al. 2008; Paré et al. 2008; Wolpin et al. 2010; Groot Hilde E. et al. 2020). This raises the possibility that the apparent protective effect of blood

group O is a consequence of lower ABO protein level. Our results cannot distinguish between these mechanisms, and it is unknown whether individuals with blood group O actually have lower plasma ABO levels. Regardless, how and why lower ABO protein levels would be protective against COVID-19 is a matter for speculation. In SARS-CoV, a naturally occurring anti-A antibodies can inhibit spike protein-mediated cellular entry via the ACE2 receptor (Guillon et al. 2008) (this is also the putative entry mechanism for SARS-CoV-2). It has been speculated that this effect may also be found in SARS-CoV-2 (Zhao et al. 2020). Another possible explanation is that the A blood type is associated with increased risk of cardiovascular disease (Wu et al. 2008), which is a known risk factor for severe COVID-19, while those with the O blood type are less likely to develop cardiovascular diseases. Furthermore, a previous report found an incidence of venous thromboembolism (VTE) of 27% in critically ill COVID-19 patients (Klok et al. 2020, p. 19). ABO blood type has been previously associated with increased risk of VTE (Wang et al. 2017); interestingly, the protein-increasing allele (C) of the top SNP (rs505922) used in the MR tests is also a risk variant for VTE (Trégouët et al. 2009). Further investigation is needed to understand the physiological role of ABO in the pathophysiology of COVID-19.

Our study had several limitations. First, our analyses were limited by the number of genetic variants that overlapped between the COVID-19 HG meta-analysis and the datasets that were used for this study. Thus, we could not test some genes that may be of importance. Second, replication of our result in an independent dataset was not possible due to the lack of COVID-19 GWAS data available at the present time; first-time associations identified by our analyses should be considered with caution. Third, although between (meta-analysis) and within cohort differences were taken into consideration on the COVID-19 GWAS analyses, the controls represented the general population and no exposure status was reported; therefore, it is possible that some GWAS associations that were used in our study may reflect factors associated with exposure, rather than a response to the SARS-CoV-2. Lastly, the lung and blood -omics data that were used reflect transcriptomic and proteomic profiles under normal conditions; it is likely that these associations change under stimulation by viral infection or acute inflammation.

## Conclusions

We used a multi-omics approach to identify several candidate genes that may be involved in the pathogenesis of COVID-19. The analyses presented here associated COVID-19 genomics to gene expression in lung and blood tissues. This approach revealed specific genes within previously

reported COVID-19 loci, and also identified new genes whose biology is consistent with COVID-19. Importantly, our analysis suggests that the ABO protein is a causal risk factor for severe COVID-19 and COVID-19 susceptibility.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s00439-021-02264-5>.

**Acknowledgements** The eQTL data from Laval University were generated from tissues obtained through the Quebec Research Respiratory Network Biobank, IUCPQ site. We acknowledge Compute Canada, which provided computational resources to conduct this research. We acknowledge MITACS Accelerate grant and Providence Airway Centre for supporting A.I.H.C. and S.M.

**Author contributions** DDS and AIHC designed the study. XL and AIHC conducted the data analyses. AIHC wrote the first draft of the manuscript. SM was involved in the interpretation of the results and critical review of the manuscript. CXY revised the manuscript. YB, PJ, WT, MB, DN, KH provided the lung *cis*-eQTLs dataset and revised the manuscript. DDS supervised this study, contributed to the interpretation of the results and revised the manuscript. All authors read and approved the final manuscript.

**Funding** A.I.H.C. and S.M. are funded by MITACS Accelerate grant and Providence Airway Centre. D.D.S. holds the De Lazzari Family Chair at HLI and a Tier 1 Canada Research Chair in COPD. Y.B. holds a Canada Research Chair in Genomics of Heart and Lung Diseases.

**Data availability** The results associated with this manuscript are available in the additional file. Specific datasets used to conduct this research are available as follow: The COVID-19 summary statistic can be accessed through the COVID-19 HG website (<https://www.covid19hg.org/>). Lung eQTL study expression and genotype data can be obtained through the GEO platform (accession number GSE23546) and the dbGaP (Study Accession phs001745.v1.p1), respectively. Lung *cis*-eQTL summary data can be obtained upon request to the Lung eQTL Consortium. Blood *cis*-eQTLs summary statistics can be accessed through the eQTLGen website (<https://www.eqtlgen.org/cis-eqtl.html>). The INTERVAL study pQTLs summary statistic are available through the European Genotype Archive (accession number EGAS00001002555).

## Compliance with ethical standards

**Conflict of interest** S.M. reports personal fees from Novartis and Boehringer-Ingelheim, outside the submitted work. W.T. reports fees to Institution from Roche-Ventana, AbbVie, Merck-Sharp-Dohme and Bristol-Myers-Squibb, outside the submitted work. M.v.d.B. reports research grants paid to Institution from AstraZeneca, Novartis, outside the submitted work. D.D.S. reports research funding from AstraZeneca and received honoraria for speaking engagements from Boehringer Ingelheim and AstraZeneca over the past 36 months, outside of the submitted work.

## References

Acharya D, Liu G, Gack MU (2020) Dysregulation of type I interferon responses in COVID-19. *Nat Rev Immunol* 20:397–398. <https://doi.org/10.1038/s41577-020-0346-x>

- Arndt B, Krieger T, Kalinski T et al (2011) The transmembrane adaptor protein SIT inhibits TCR-mediated signaling. *PLoS ONE* 6:e23761. <https://doi.org/10.1371/journal.pone.0023761>
- Balfour H (2020) Inhaled interferon beta therapy shows promise in COVID-19 trial. In: *European Pharmaceutical Review*. <https://www.europeanpharmaceuticalreview.com/news/123994/inhaled-interferon-beta-therapy-shows-promise-in-covid-19-trial/>. Accessed 7 Aug 2020
- Bowden J, Davey Smith G, Burgess S (2015) Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol* 44:512–525. <https://doi.org/10.1093/ije/dyv080>
- Bunyavanich S, Do A, Vicencio A (2020) Nasal gene expression of angiotensin-converting enzyme 2 in children and adults. *JAMA* 323:2427–2429. <https://doi.org/10.1001/jama.2020.8707>
- Bycroft C, Freeman C, Petkova D et al (2018) The UK Biobank resource with deep phenotyping and genomic data. *Nature* 562:203–209. <https://doi.org/10.1038/s41586-018-0579-z>
- Cano-Gamez E, Trynka G (2020) From GWAS to function: using functional genomics to identify the mechanisms underlying complex diseases. *Front Genet*. <https://doi.org/10.3389/fgene.2020.00424>
- Chang CC, Chow CC, Tellier LC et al (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. <https://doi.org/10.1186/s13742-015-0047-8>
- Di Angelantonio E, Thompson SG, Kaptoge S et al (2017) Efficiency and safety of varying the frequency of whole blood donation (INTERVAL): a randomised trial of 45 000 donors. *Lancet* 390:2360–2371. [https://doi.org/10.1016/S0140-6736\(17\)31928-1](https://doi.org/10.1016/S0140-6736(17)31928-1)
- Dong E, Du H, Gardner L (2020) An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis* 20:533–534. [https://doi.org/10.1016/S1473-3099\(20\)30120-1](https://doi.org/10.1016/S1473-3099(20)30120-1)
- Ellinghaus D, Degenhardt F, Bujanda L et al (2020) Genomewide association study of severe Covid-19 with respiratory failure. *N Engl J Med*. <https://doi.org/10.1056/NEJMoa2020283>
- 1000 Genomes Project Consortium, Auton A, Brooks LD et al (2015) A global reference for human genetic variation. *Nature* 526:68–74. <https://doi.org/10.1038/nature15393>
- Giambartolomei C, Vukcevic D, Schadt EE et al (2014) Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet* 10:e1004383. <https://doi.org/10.1371/journal.pgen.1004383>
- Groot HE, Villegas Sierra Laura E, Abdullah SM et al (2020) Genetically determined ABO blood group and its associations with health and disease. *Arterioscler Thromb Vasc Biol* 40:830–838. <https://doi.org/10.1161/ATVBAHA.119.313658>
- Guillon P, Clément M, Sébille V et al (2008) Inhibition of the interaction between the SARS-CoV Spike protein and its cellular receptor by anti-histo-blood group antibodies. *Glycobiology* 18:1085–1093. <https://doi.org/10.1093/glycob/cwn093>
- Hao K, Bossé Y, Nickle DC et al (2012) Lung eQTLs to help reveal the molecular underpinnings of asthma. *PLoS Genet* 8:e1003029. <https://doi.org/10.1371/journal.pgen.1003029>
- Jordan RE, Adab P, Cheng KK (2020) Covid-19: risk factors for severe disease and death. *BMJ*. <https://doi.org/10.1136/bmj.m1198>
- Klok FA, Kruip MJHA, van der Meer NJM et al (2020) Incidence of thrombotic complications in critically ill ICU patients with COVID-19. *Thromb Res* 191:145–147. <https://doi.org/10.1016/j.thromres.2020.04.013>
- Lauc G, Huffman JE, Pučić M et al (2013) Loci associated with *N*-glycosylation of human immunoglobulin G show pleiotropy with autoimmune diseases and haematological cancers. *PLoS Genet* 9:e1003225. <https://doi.org/10.1371/journal.pgen.1003225>
- Ma Y, Huang Y, Zhao S et al (2020) Integrative genomics analysis reveals a novel 21q22.11 locus contributing to susceptibility of COVID-19. *medRxiv*. <https://doi.org/10.1101/2020.09.16.20195685>

- Meier C, Camargo SM, Hunziker S et al (2018) Intestinal IMINO transporter SIT1 is not expressed in human newborns. *Am J Physiol Gastrointest Liver Physiol* 315:887–895. <https://doi.org/10.1152/ajpgi.00318.2017>
- Melzer D, Perry JRB, Hernandez D et al (2008) A genome-wide association study identifies protein quantitative trait loci (pQTLs). *PLoS Genet* 4:e1000072. <https://doi.org/10.1371/journal.pgen.1000072>
- Moffatt MF, Gut IG, Demenais F et al (2010) A large-scale, consortium-based genomewide association study of asthma. *N Engl J Med* 363:1211–1221. <https://doi.org/10.1056/NEJMoa0906312>
- Mutlu-Albayrak H, Bene J, Oflaz MB, et al (2015) Identification of SLC22A5 gene mutation in a family with carnitine uptake defect. In: *Case reports in genetics*. <https://www.hindawi.com/journals/crig/2015/259627/>. Accessed 9 Nov 2020
- Pairo-Castineira E, Clohisey S, Klaric L et al (2020) Genetic mechanisms of critical illness in Covid-19. *Nature*. <https://doi.org/10.1038/s41586-020-03065-y>
- Paré G, Chasman DI, Kellogg M et al (2008) Novel association of ABO histo-blood group antigen with soluble ICAM-1: results of a genome-wide association study of 6,578 women. *PLoS Genet* 4:e1000118. <https://doi.org/10.1371/journal.pgen.1000118>
- Ray JG, Schull MJ, Vermeulen MJ, Park AL (2020) Association between ABO and Rh blood groups and SARS-CoV-2 infection or severe COVID-19 illness. *Ann Intern Med*. <https://doi.org/10.7326/M20-4511>
- Simeoni L, Posevitz V, Kölsch U et al (2005) The transmembrane adapter protein SIT regulates thymic development and peripheral T-cell functions. *Mol Cell Biol* 25:7557–7568. <https://doi.org/10.1128/MCB.25.17.7557-7568.2005>
- Staley OYJ (2020) MendelianRandomization: Mendelian Randomization Package. Version 0.4.3URL <https://CRAN.R-project.org/package=MendelianRandomization>
- Sun BB, Maranville JC, Peters JE et al (2018) Genomic atlas of the human plasma proteome. *Nature* 558:73–79. <https://doi.org/10.1038/s41586-018-0175-2>
- Swerdlow DI, Kuchenbaecker KB, Shah S et al (2016) Selecting instruments for Mendelian randomization in the wake of genome-wide association studies. *Int J Epidemiol* 45:1600–1616. <https://doi.org/10.1093/ije/dyw088>
- The COVID-19 Host Genetics Initiative (2020) 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. <https://doi.org/10.1038/s41431-020-0636-6>
- Trégouët D-A, Heath S, Saut N et al (2009) Common susceptibility alleles are unlikely to contribute as strongly as the FV and ABO loci to VTE risk: results from a GWAS approach. *Blood* 113:5298–5303. <https://doi.org/10.1182/blood-2008-11-190389>
- van der Made CI, Simons A, Schuurs-Hoeijmakers J et al (2020) Presence of genetic variants among young men with severe COVID-19. *JAMA*. <https://doi.org/10.1001/jama.2020.13719>
- Võsa U, Claringbould A, Westra H-J et al (2018) Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis. *bioRxiv*. <https://doi.org/10.1101/447367>
- Vuille-dit-Bille RN, Camargo SM, Emmenegger L et al (2015) Human intestine luminal ACE2 and amino acid transporter expression increased by ACE-inhibitors. *Amino Acids* 47:693–705. <https://doi.org/10.1007/s00726-014-1889-6>
- Vuille-dit-Bille RN, Liechty KW, Verrey F, Guglielmetti LC (2020) SARS-CoV-2 receptor ACE2 gene expression in small intestine correlates with age. *Amino Acids*. <https://doi.org/10.1007/s00726-020-02870-z>
- Wang Z, Dou M, Du X et al (2017) Influences of ABO blood group, age and gender on plasma coagulation factor VIII, fibrinogen, von Willebrand factor and ADAMTS13 levels in a Chinese population. *PeerJ*. <https://doi.org/10.7717/peerj.3156>
- Westra H-J, Peters MJ, Esko T et al (2013) Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet* 45:1238–1243. <https://doi.org/10.1038/ng.2756>
- Wolpin BM, Kraft P, Gross M et al (2010) Pancreatic cancer risk and ABO blood group alleles: results from the pancreatic cancer cohort consortium. *Cancer Res* 70:1015–1023. <https://doi.org/10.1158/0008-5472.CAN-09-2993>
- Wu O, Bayoumi N, Vickers MA, Clark P (2008) ABO(H) blood groups and vascular disease: a systematic review and meta-analysis. *J Thromb Haemost* 6:62–69. <https://doi.org/10.1111/j.1538-7836.2007.02818.x>
- Yang J, Lee SH, Goddard ME, Visscher PM (2011) GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* 88:76–82. <https://doi.org/10.1016/j.ajhg.2010.11.011>
- Zhao J, Yang Y, Huang H et al (2020) Relationship between the ABO blood group and the COVID-19 susceptibility. *medrxiv*. <https://doi.org/10.1101/2020.03.11.20031096>
- Zhou F, Yu T, Du R et al (2020a) Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* 395:1054–1062. [https://doi.org/10.1016/S0140-6736\(20\)30566-3](https://doi.org/10.1016/S0140-6736(20)30566-3)
- Zhou P, Yang X-L, Wang X-G et al (2020b) A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579:270–273. <https://doi.org/10.1038/s41586-020-2012-7>
- Zhu Z, Zhang F, Hu H et al (2016) Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet* 48:481–487. <https://doi.org/10.1038/ng.3538>
- Zhu N, Zhang D, Wang W et al (2020) A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med* 382:727–733. <https://doi.org/10.1056/NEJMoa2001017>
- Zietz M, Tatonetti NP (2020) Testing the association between blood type and COVID-19 infection, intubation, and death. *medRxiv*. <https://doi.org/10.1101/2020.04.08.20058073>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.