Altered blood cell traits underlie a major genetic locus of severe COVID-19

Jingqi Zhou, PhD[1,2], Yitang Sun, MPH[1], Weishan Huang, PhD[3,4], Kaixiong Ye, PhD[1,5*]

1. Department of Genetics, Franklin College of Arts and Sciences, University of Georgia, Athens, GA, USA

2. School of Public Health, Shanghai Jiao Tong University School of Medicine, Shanghai, P.R. China

3. Department of Pathobiological Sciences, School of Veterinary Medicine, Louisiana State University, Baton Rouge, LA, USA

4. Department of Microbiology and Immunology, College of Veterinary Medicine, Cornell University, Ithaca, NY, USA

5. Institute of Bioinformatics, University of Georgia, Athens, GA, USA

* **Address for Correspondence:** Dr. Kaixiong Ye

Department of Genetics

University of Georgia

C220 Davison Life Sciences

120 East Green Street, Athens, GA 30602

Office: 706-542-5898

Fax: 706-542-3910

Email: Kaixiong.Ye@uga.edu

ORCID ID: JZ (0000-0002-9961-199X); YS (0000-0002-6564-8815); WH (0000-0002-1330-1131); KY (0000-0003-4658-7292)

Abstract

Background: The genetic locus 3p21.31 has been associated with severe coronavirus disease 2019 (COVID-19), but the underlying pathophysiological mechanism is unknown.

Methods: To identify intermediate traits associated with the 3p21.31 locus, we first performed a phenome-wide association study (PheWAS) with 923 phenotypes in 310,999 European individuals from the UK Biobank. For genes potentially regulated by the COVID-19 risk variant, we examined associations between their expression and the polygenic score (PGS) of 1,263 complex traits in a meta-analysis of 31,684 blood samples. For the prioritized blood cell traits, we tested their associations with age and sex in the same UK Biobank sample.

Results: Our PheWAS highlighted multiple blood cell traits to be associated with the COVID-19 risk variant, including monocyte count and percentage ($p = 1.07\times10^{-8}$, 4.09×10^{-13}), eosinophil count and percentage ($p = 5.73\times10^{-3}$, 2.20×10^{-3}), and neutrophil percentage ($p = 3.23\times10^{-3}$). The PGS analysis revealed positive associations between the expression of candidate genes and genetically predicted counts of specific blood cells: *CCR3* with eosinophil and basophil ($p = 5.73\times10^{-21}$, 5.08×10^{-19}); *CCR2* with monocytes ($p = 2.40\times10^{-10}$); and *CCR1* with monocytes and neutrophil ($p = 1.78\times10^{-6}$, 7.17×10^{-5}). Additionally, we found that almost all examined white blood cell traits are significantly different across age and sex groups.

Conclusions: Our findings suggest that altered blood cell traits, especially those of monocyte, eosinophil, and neutrophil, may represent the mechanistic links between the genetic locus 3p21.31 and severe COVID-19. They may also underlie the increased risk of severe COVID-19 in older adults and men.

Key Words: COVID-19; blood cells; monocyte; eosinophil; phenome-wide association study

Introduction

The coronavirus disease 2019 (COVID-19), caused by infection of the severe acute respiratory syndrome coronavirus 2 (SARS- CoV- 2), affects individuals differently, with clinical manifestations ranging from asymptomatic infection, to mild flu-like symptoms, to severe respiratory failure (1-4). While some demographic factors and pre-existing conditions, especially older age and male sex, are well-established risk factors for severe COVID-19, the exact mechanisms are still elusive (5, 6). Genetic variation is partly responsible for varying individual responses (7-9). The first genome-wide association study (GWAS) for COVID-19 was published in June 2020, comparing 1,610 severe patients with respiratory failure to 2,205 healthy controls from Italy and Spain. It identified two genetic loci, with the most signal at locus 3p21.31 and the other at locus 9q34.2 coinciding with the *ABO* blood group locus (8). The association at locus 3p21.31 was independently replicated by the COVID-19 Host Genetics Initiative (9). The peak signal at this locus spans multiple chemokine receptor genes (e.g., *CCR9, CXCR6, XCR1* and *CCR1*) and risk variants are associated with the expression of *CXCR6, CCR1* and *SLC6A20* (8). However, the underlying pathophysiological process is unknown.

Phenome-wide association study (PheWAS) is an unbiased approach that evaluates the associations of a disease-associated genetic variant (e.g., a COVID-19 risk variant) with a wide range of phenotypes (i.e., the phenome). PheWAS may identify intermediate traits or biomarkers residing in the causal physiological route from the genetic variant to the disease of interest, or it may reveal unexpected comorbidities that indicate shared biological mechanisms (10, 11). Similarly, expression quantitative trait locus (eQTL) analysis for a trait-associated genetic variant across the transcriptome can identify candidate causal genes that are either close (in *cis*) or remote (in *trans*) to the variant (12, 13). From the perspective of a candidate gene, insights could be gained into its physiological pathways and downstream functional effects by examining the associations of its expression level

with phenotypes across the phenome, or even with the genetically predicted phenotypic status if measured ones are unavailable (13).

This project aims to explore the mechanistic link between the genetic locus 3p21.31 and severe COVID-19. We first leveraged the deep phenotyping and genomic data in the UK Biobank (N = 310,999) and performed a PheWAS for a severe COVID-19 risk variant across 923 disease phenotypes, biomarkers and blood cell traits. Moreover, for genes potentially regulated by the COVID-19 risk variant, associations between their expression levels and the polygenic scores (PGS) of 1,263 traits were evaluated in 31,684 blood samples. These two unbiased phenome-wide approaches converged on blood cell traits, especially counts of monocyte, eosinophil and neutrophil, as the possible intermediate link between the genetic locus 3p21.31 and severe COVID-19. These blood cell traits are significantly associated with age and sex in UK Biobank, calling for future studies into their roles in the increased risk of severe COVID-19 in older adults and men.

Method

Ethics statement

UK Biobank is a large population-based prospective study that recruited more than 500,000 individuals aged 40-70 years between 2006 and 2010. It was approved by the North West Multi-Centre Research Ethics Committee (11/NW/ 0382) and proper informed consent was obtained. All participants received baseline measurements, donated biological materials, and provided access to their medical records (14). Data for this project was accessed through an approved application to UK Biobank (Application ID: 48818).

Phenome-wide association study for the COVID-19 risk variant

Among all UK Biobank participants, only those fulfilling the following criteria were included in our PheWAS analyses: 1) genetic ancestry is Caucasian; 2) included in the genetic principal component analysis; 3) no sex chromosome aneuploidy; 4) no high degree of genetic kinship, and 5) for relative pairs (kinship coefficient > 0.0884), a minimum number of participants were removed so that all those remaining are unrelated. A total of 310,999 unrelated individuals passed this quality control and filtering procedure.

Three sets of phenotypes were examined: binary disease outcomes, blood and urine biomarkers, and blood cell traits. Binary disease status was defined by mapping ICD9/ICD10 diagnosis codes in the hospital episode statistics to phecodes in the PheCODE grouping system (15). To ensure sufficient statistical power, only phecodes with a minimum of 200 cases were retained. A total of 858 phecodes were included in our analysis. For continuous traits, our PheWAS included 30 blood and 4 urine biochemistry markers, and 31 blood cell traits (14, 16, 17). The full list of phenotypes could be found in our additional supporting materials (18). Statistical association analyses were performed with the R PheWAS package (19). Logistic regression was performed for binary disease outcomes and linear regression for continuous blood and urine biomarkers, adjusting for age, sex, genotyping array, assessment center, and the first 10 genetic principal components. For statistical significance, we applied Bonferroni correction for the total number of phenotypes tested (i.e., $p < 0.05 / 923 = 5.42\times10^{-5}$), although we note that this is a conservative approach because the phenotypes are not independent.

For blood cell traits whose associations with the severe COVID-19 risk variant are nominally significant ($p < 0.05$), we examined their associations in two existing studies. The first is GeneATLAS,

a large database of associations based on 452,264 UK Biobank White British individuals (20). The other is a GWAS meta-analysis for 36 blood cell traits in 173,480 European ancestry individuals across three cohorts: UK Biobank (N = 87,265), UK BiLEVE (N = 45,694), and INTERVAL (N = 40,521) (21). Summary statistics for this study were retrieved from the IEU OpenGWAS database (22). We note that these two studies were also mainly based on UK Biobank samples and thus should not be considered as independent replications.

eQTL and polygenic score association analysis

To identify genes whose expression levels are associated with the COVID-19 risk variant, we inquired eQTL analysis results from GTEx and eQTLGen (13, 23). The GTEx project studies tissue-specific gene expression and regulation in 54 non-diseased tissue sites from about 1,000 individuals (23). The eQTLGen Consortium conducted eQTL meta-analysis in 31,684 samples of blood and peripheral blood mononuclear cells from 37 datasets (13). It also performed polygenic score association analysis to evaluate the associations between the expression level of most genes and the PGS of 1,263 traits (13). The majority of samples in both studies are of European ancestry. In the PGS association analysis, multiple PGS were calculated for each trait with different GWAS, sample ancestry, and p value cutoffs (p = 0.01, 1×10^{-3}, 1×10^{-4}, 1×10^{-5}, 5×10^{-8}). For blood cell traits, three previous GWAS were used and designated as study 1 (24), study 2 (25), and study 3 (26), respectively. Statistical significance was defined with the false discovery rate approach (FDR < 0.05) (13).

Association analysis for the age and sex effects on blood cell traits

We evaluated the age and sex effects on 31 blood cell traits in the UK Biobank dataset with a linear regression model that included three variables, a continuous variable for age, a categorical variable for sex (female = 0 and male = 1), and a third term for the interaction between sex and age. For

statistical significance, we applied Bonferroni correction for the total number of blood cell traits tested (i.e., $p < 0.05 / 31 = 1.61\times10^{-3}$). For each blood cell trait, we also reported their mean and standard deviation in all samples and also in men and women separately.

Resources

eQTLGen: https://www.eqtlgen.org/index.html

Gene ATLAS: http://geneatlas.roslin.ed.ac.uk/

GTEx: https://www.gtexportal.org/home/

The COVID-19 GWAS Results Browser: https://ikmb.shinyapps.io/COVID-19_GWAS_Browser/

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

The IEU OpenGWAS database: https://gwas.mrcieu.ac.uk/

Additional Supporting Materials: All summary statistics for our PheWAS of the COVID-19 risk variant, polygenic score association analysis, PrediXcan and MultiXcan analysis are available at https://doi.org/10.6084/m9.figshare.13637984.v1 (18).

Results

The severe COVID-19 risk variant is associated with blood cell traits

The severe COVID-19 risk variant examined in this study is rs67959919 (G/A), whose risk allele A has an odds ratio (OR) of 2.07 (95% confidence interval (CI): 1.66-2.56, $p = 4.69\times10^{-11}$) for severe COVID-19 after adjustment for genetic principal components, age and sex (8). It is in perfect linkage disequilibrium (LD, $r^2 = 1$) with the lead variant, rs11385942 (A/GA, OR = 2.11, 95% CI: 1.70-2.61, p =

9.46×10^{-12}) in European populations (27). The lead variant is an insertion-deletion polymorphism and is not found in some existing datasets. To identify phenotypes associated with rs67959919, we performed a PheWAS in a subset of 310,999 unrelated European individuals from the UK Biobank after quality control and filtering (Supplementary Table 1 for baseline characteristics). A total of 923 phenotypes were investigated, including 858 binary disease outcomes, 34 blood and urine biomarkers, and 31 blood cell traits (Figure 1).

With the conservative Bonferroni correction for the total number of phenotypes tested ($p < 5.42\times10^{-5}$), we observed that the severe COVID-19 risk variant is associated with monocyte percentage ($p = 4.09\times10^{-13}$) and monocyte count ($p = 1.07\times10^{-8}$). None of the binary disease outcomes or biomarkers passes this significance cutoff. The top three binary phenotypes were all related to the digestive system: sialolithiasis ($p = 4.76\times10^{-4}$), periodontitis ($p = 1.53\times10^{-3}$) and its subcategory, chronic periodontitis ($p = 2.43\times10^{-3}$). At the nominal significance level ($p < 0.05$), associations were observed with additional blood cell traits: eosinophil count ($p = 5.73\times10^{-3}$), eosinophil percentage ($p = 2.20\times10^{-3}$), neutrophil count ($p = 0.032$), neutrophil percentage ($p = 3.23\times10^{-3}$), mean corpuscular hemoglobin ($p = 0.042$), and mean corpuscular hemoglobin concentration ($p = 0.042$). In GeneATLAS, a large database of associations based on 452,264 UK Biobank White British individuals (20), consistent and even more significant associations were observed (Table 1). It also revealed three additional suggestive negative associations: basophil count ($p = 1.40\times10^{-3}$), basophil percentage ($p = 1.87\times10^{-3}$), and red blood cell (RBC) count ($p = 0.022$). Another meta-analysis of 173,480 European-ancestry individuals also reported similar trends (Table 1) (21). Overall, our unbiased PheWAS revealed associations of the severe COVID-19 risk variant with multiple blood cell traits.

Expression of candidate target genes is associated with genetically predicted blood cell traits

Candidate target genes of the COVID-19 risk variant affected through regulation of gene expression could be identified with eQTL analysis. Based on *cis*-eQTL analysis in 54 tissues from GTEx (23), the COVID-19 risk variant is associated with the expression of *CXCR6, SLC6A20, CCR1, CCR9, RP11-697K23.3*, and *LZTFL1* in a total of 9 tissues (Supplementary Table 2). Moreover, eQTLGen (13), a meta-analysis for *cis*-eQTL in 31,684 blood samples additionally identified the following genes: *FLT1P1, CCR3, SACM1L, CCR5, CCR2* and *RP11-24F11.2* (Supplementary Table 3). *Trans*-eQTL analysis in both studies did not identify any genes.

For all these potential target genes of the COVID-19 risk variant, we interrogated if their expression levels are associated with the PGS of 1,263 traits examined in eQTLGen. A significant association indicates that the gene is implicated in pathways contributing to the trait (13). Multiple significant associations were identified after correction for multiple testing (Figure 2). Genetically predicated higher monocyte count is positively associated with the expression of *CCR1* ($p = 1.78\times10^{-6}$) and *CCR2* ($p = 2.40\times10^{-10}$). Genetically predicated higher eosinophil count and basophil count are positively associated with *CCR3* expression ($p = 5.73\times10^{-21}$ and $p = 5.08\times10^{-19}$, respectively). At the nominal significance level ($p < 0.05$), additional associations were observed for neutrophils, lymphocytes, RBC, and white blood cells (WBC). Predicted neutrophil count is positively associated with the expression of *CCR1* ($p = 7.17\times10^{-5}$), *FLT1P1* ($p = 1.46\times10^{-5}$), *SACM1L* ($p = 5.69\times10^{-3}$) and *CCR3* ($p = 0.011$). A negative association was observed between predicted lymphocyte count and *CCR1* ($p = 9.41\times10^{-4}$). Genetically predicted higher RBC count is associated with lower expression of *CXCR6* ($p = 3.74\times10^{-4}$) and *RP11-24F11.2* ($p = 2.09\times10^{-3}$), while for WBC count, a positive association was observed with *SACM1L* ($p = 8.26\times10^{-4}$). Notably, these associations were consistent across different GWAS datasets and *p* value cutoffs used in the PGS calculation (18). These results suggest a

possibility that the target gene of the COVID-19 risk variant is involved in hematologic processes and regulates blood cell counts.

Integrating and reconciling association signals in PheWAS, eQTL, and PGS analysis, three candidate blood cell traits and their corresponding candidate genes were prioritized (Figure 3). First, the severe COVID-19 risk allele inhibits the expression of *CCR1* and *CCR2*, subsequently reducing the monocyte count. Second, the risk allele downregulates *CCR3* expression and further diminishes the eosinophil count. Third, the risk allele downregulates *CCR3* expression and relieves its inhibition on the neutrophil count.

Blood cell traits are significantly different across age and sex groups

Given the well-established role of older age and male sex as risk factors for severe COVID-19, we examined if blood cell traits are significantly different across age and sex groups by testing their associations in the UK Biobank. The basic statistics for 31 blood cell traits were reported in Supplementary Table 4. Among the 11 traits related to white blood cells, all of them are significantly associated with age (Table 2, $p < 1.61\times10^{-3}$). All but one, eosinophil count, are significantly different in men and women. The interaction between sex and age is highly significant across all these white blood cell traits. Significant age and sex effects were also observed for traits related to red blood cells and platelets (Supplementary Table 5).

Discussion

With an unbiased phenome-wide scan approach, our study established two pairs of relationships: 1) associations of the severe COVID-19 risk variant with blood cell traits; and 2) associations between expression levels of candidate target genes and the PGS of blood cell traits. Integrating association

signals across multiple analyses prioritizes three blood cell traits, the counts of monocyte, eosinophil and neutrophil, and their candidate target genes, *CCR1*, *CCR2*, and *CCR3*. Taken together, our results proposed blood cell traits as the probable mechanistic link between the risk variant at 3p21.31 and severe COVID-19. We further showed that these blood cell traits are drastically different across age and sex groups, calling for future investigation into their roles in the increased risk of severe COVID-19 in older adults and men.

Hematologic manifestations are common in COVID-19 patients, especially elevated WBC and neutrophil counts but decreased lymphocyte and platelet counts (28-30). Leukocytosis, neutrophilia, lymphopenia, thrombocytopenia, and neutrophil-to-lymphocyte ratio have been repeatedly associated with worse COVID-19 outcomes and could serve as prognostic biomarkers (1, 4, 31-33). Reducing basophil count or percentage was generally observed in patients (34-36). For monocyte, its total number in circulation does not change dramatically in COVID-19 patients, with reports of no change or only slight increase (29, 36, 37). However, its composition exhibits a pronounced shift, with a significant expansion of its inflammatory subsets, which are not typically seen in healthy individuals (6, 29, 30, 37-40). The pattern of eosinophil is less well-established. Some studies observed diminished and even undetectable eosinophil counts (i.e., eosinopenia) in COVID-19 patients (34, 36, 41-45), and it was also shown that eosinophil counts are positively associated with lymphocyte counts in both severe and non-severe cases (45). However, others did not find a significant difference (46), and there is also an report of an expanded eosinophil percentage among the total viable leukocyte CD45+ population (29). These changes in circulating blood cells are closely related to the infiltration and accumulation of lymphocyte, neutrophil, eosinophil, and inflammatory monocyte-macrophage in the lung and other organs, leading to neutrophil extracellular trap and cytokine release syndrome (30, 47-49). Notably, an immuno-monitoring study of COVID-19 patients

from acute to recovery phages observed gradual reduction of neutrophil and replenishment of basophil, eosinophil and non-inflammatory monocyte (50).

Our PheWAS in UK Biobank for the severe COVID-19 risk variant revealed that the risk allele is associated with decreased monocyte count and percentage, eosinophil count and percentage, but with increased neutrophil percentage. These significant associations remained almost unchanged when the white blood cell count was included as an additional covariate. GeneATLAS reported even more significant associations for these relationships, probably due to its different quality control procedures and a larger sample size (20). It also reports suggestive evidence of negative associations between the risk allele and basophil count and percentage. These association directions are consistent with the observed blood cell count changes in COVID-19 patients, as discussed above. Of note, our associations were identified in the generally healthy population samples. On the other hand, the vast majority of existing studies measured blood cell counts at hospital admission or during hospitalization, which likely reflect immune responses to SARS-CoV-2 infection. Future studies are warranted to evaluate if before-infection differences in blood cell counts play a role in modulating the risk of developing severe COVID-19, especially in the context of age and sex differences.

Our PGS analysis for the potential target genes of the COVID-19 risk variant further unraveled associations with multiple blood cell counts. It is important to stress that these associations are consistent across analyses with PGS calculated with different GWAS datasets, p value cutoffs, and sample ancestries. Intersecting and reconciling association signals across PheWAS, eQTL, and PGS analysis yielded multiple possible pathways for the COVID-19 risk allele. Strong and consistent evidence was found on the pathways through monocyte and eosinophil. On the other hand, support

for the role of neutrophil is weaker. The association with neutrophil count was not significant in a meta-analysis of three UK cohorts (21). Also, the negative associations between *CCR3* expression and PGS of the neutrophil count were only suggestive (p = 0.011). In addition to these three blood cells, basophil may serve as another candidate pathway: the risk allele downregulates *CCR3* expression, reduces its stimulatory effect on basophil count, and thus leads to a reduction of basophil. Additional evidence for the potential importance of these candidate genes could be drawn from their cell-type-specific expression patterns (Supplementary Figure 1). *CCR3* has highly specific expression in eosinophil and basophil and only slight expression in neutrophil, *CCR2* has high expression in basophil and medium expression in classical monocyte, while *CCR1* has medium to high expression across all types of granulocytes and monocyte. Notably, this pathway prioritization analysis utilized eQTL association signals in blood samples, but the regulatory effects could be different across tissues (23). Also, the eQTL analyses were based on generally healthy samples (13, 23). The regulatory effects of the risk variant may be different under the SARS-CoV-2 infection. Further studies are needed to examine its functional effects in patients and to identify the most relevant tissue. Nonetheless, our study prioritized hematologic processes as the downstream pathophysiology of a major genetic locus of severe COVID-19.

To further confirm and examine the associations between the expression of candidate target genes and blood cell traits, we additionally performed association analyses between genetically predicted gene expression levels and observed blood cell traits in 310,999 unrelated European individuals from the UK Biobank, the same dataset used in the PheWAS analysis (See Supplementary Note for references). Applying the PrediXcan pipeline and leveraging MASHR-based models of the GTEx V8 reference dataset, we predicted the expression levels of each candidate target gene in 49 tissues and cells, and further tested their associations with each of the 31 observed blood cell traits (18). To summarize evidence of an association between a gene and a specific trait, we applied the MultiXcan

tool to integrate information across all available tissues. First, these additional analyses confirmed the associations between the expression of candidate target genes and blood cell traits. For instance, across all examined tissues, observed monocyte count is associated with genetically predicted *CCR1* ($p = 2.79\times10^{-19}$) and *CCR2* expression ($p = 1.83\times10^{-28}$); observed eosinophil count is associated with genetically predicted *CCR3* expression ($p = 4.35\times10^{-4}$); observed neutrophil percentage is associated with genetically predicted *CCR3* expression ($p = 3.35\times10^{-5}$). Second, the tissue-level analysis in 49 tissues revealed tissue-specific association patterns. For instance, the association between *CCR1* and monocyte count is positive in some tissues (e.g., amygdala, cerebellar hemisphere, and hippocampus), but negative (e.g., esophagus, lung, and colon) and non-significant in others (e.g., stomach, heart atrial appendage, and putamen basal ganglia). These tissue-dependent patterns indicate that there are more possible pathways connecting the COVID-19 risk variant, blood cell traits, and COVID-19, in addition to those presented in Figure 3 and based on associative patterns in blood samples. The relatively large number of significant associations in different tissues makes it challenging to further narrow down to specific tissues. Additionally, since the phenotypes of special interest are blood cell counts and percentages, it will be especially informative if we could evaluate gene expression in specific subsets of blood cells. However, these are currently unavailable in the GTEx reference dataset. Nevertheless, our additional analysis confirms the link between a major genetic locus of severe COVID-19 and blood cell traits.

The strengths of our study include the unbiased phenome-wide approach at two levels of analysis, the genetic variant (923 phenotypes) and the gene expression (1,263 phenotypes). The large sample sizes in UK Biobank (N = 310,999) and eQTLGen (N = 31,684) increase the statistical power to identify associations. The two phenome-wide analyses converged on multiple blood traits and the association directions are consistent with existing studies in COVID-19 patients, further bringing credibility to our findings. This study also has some weaknesses. First, there are still phenotypes not

covered in our analyses, such as monocyte subsets, T cells and B cells. It is possible that the COVID-19 risk variant affects other unexplored intermediate traits. The phenotypes used in this study, including biochemistry markers, blood cell traits, and medical conditions, were either measured at recruitment or retrieved from medical records. They may not reflect the most up-to-date status, and some cases may be classified as healthy controls, reducing the power of our analysis. Although the associations between blood cell traits and the COVID-19 risk variant are consistent across three separate studies, they all relied on the UK Biobank and thus could not represent independent replications. Moreover, future fine-mapping studies are needed to identify causal variants at locus 3p21.31 for both severe COVID-19 and blood cell traits. The possibility could not be ruled out at present that they are different variants in strong LD (51). Our analyses were restricted to healthy individuals and may not reflect patterns in COVID-19 patients. Additionally, the associations between gene expression and genetically predicted blood cell counts should be further confirmed with direct analysis of their measured counts.

While this manuscript was under review, tremendous research progress was made on understanding the host genetics and disease etiology of COVID-19 (See Supplementary Note for references). First, more genetic association studies of COVID-19 have confirmed the association of locus 3p21.31 with the severity of and susceptibility to COVID-19. Second, transcriptome-wide association analysis (TWAS) similarly highlighted genes at locus 3p21.31. The genetically predicted expression levels of these genes in specific tissues have been associated with COVID-19 susceptibility or severity, such as the association of *CCR2*, *CCR3,* and *CXCR6* expression in lung tissue with critical COVID-19. Third, potential mechanistic insights were obtained. *CCR1* and the canonical ligands for *CCR1, CCR2,* and *CCR3*, such as *CCL2, CCL3, CCL4, CCL7,* and *CCL8*, have upregulated expression in bronchoalveolar lavage fluid of COVID-19 patients. The blood level of *CCL2* was significantly elevated in both mild and severe COVID-19 patients. Additionally, the potentially causal roles of white blood cells in the

development of severe COVID-19 were evaluated in a two-sample Mendelian randomization study. Genetically predicted lower counts of white blood cells, myeloid white blood cells, and granulocytes, and higher eosinophil percentage were found to be associated with an increased risk of severe COVID-19. These recent discoveries further strengthen the link from locus 3p21.31 to blood cell traits and then to severe COVID-19.

In conclusion, our phenome-wide association study for the severe COVID-19 risk variant at locus 3p21.31 and its candidate target genes identified altered blood cell traits, especially counts of monocyte, eosinophil, and neutrophil, as the probable mechanistic links between the genetic locus and severe COVID-19. These blood cell traits, together with their candidate acting genes, *CCR1*, *CCR2* and *CCR3*, represent compelling and testable hypothesis that call for follow-up studies into their roles in COVID-19 pathogenesis, especially in elevating the risk in the older adults and men.

Conflicts of interest

The authors declare that they have no conflict of interest.

Funding

This work is supported by the University of Georgia Research Foundation.

Acknowledgments

We would like to thank UK Biobank participants and administrators for data access. We also want to thank all other Ye lab members for helpful discussions. JZ and KY initiated and designed this study. JZ and YS performed data analysis and prepared visualizations. JZ, WH, and KY interpreted the results. JZ and KY wrote the first draft of the manuscript. All authors read, edited and approved the final version.

References

1. Zhou F, Yu T, Du R, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet. 2020;395:1054-1062. doi: 10.1016/S0140-6736(20)30566-3
2. Williamson EJ, Walker AJ, Bhaskaran K, et al. OpenSAFELY: factors associated with COVID-19 death in 17 million patients. Nature. 2020. doi: 10.1038/s41586-020-2521-4
3. Oran DP, Topol EJ. Prevalence of Asymptomatic SARS-CoV-2 Infection: A Narrative Review. Ann Intern Med. 2020. doi: 10.7326/M20-3012
4. Shi S, Qin M, Shen B, et al. Association of Cardiac Injury With Mortality in Hospitalized Patients With COVID-19 in Wuhan, China. JAMA Cardiol. 2020. doi: 10.1001/jamacardio.2020.0950
5. Nikolich-Zugich J, Knox KS, Rios CT, Natt B, Bhattacharya D, Fain MJ. SARS-CoV-2 and COVID-19 in older adults: what we may expect regarding pathogenesis, immune responses, and outcomes. Geroscience. 2020;42:505-514. doi: 10.1007/s11357-020-00186-0
6. Pence BD. Severe COVID-19 and aging: are monocytes the key? Geroscience. 2020;42:1051-1061. doi: 10.1007/s11357-020-00213-0
7. van der Made CI, Simons A, Schuurs-Hoeijmakers J, et al. Presence of Genetic Variants Among Young Men With Severe COVID-19. JAMA. 2020. doi: 10.1001/jama.2020.13719
8. Ellinghaus D, Degenhardt F, Bujanda L, et al. Genomewide Association Study of Severe Covid-19 with Respiratory Failure. N Engl J Med. 2020. doi: 10.1056/NEJMoa2020283
9. Initiative C-HG. 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. 2020;28:715-718. doi: 10.1038/s41431-020-0636-6
10. Bush WS, Oetjens MT, Crawford DC. Unravelling the human genome-phenome relationship using phenome-wide association studies. Nat Rev Genet. 2016;17:129-145. doi: 10.1038/nrg.2015.36
11. Hebbring SJ. The challenges, advantages and future of phenome-wide association studies. Immunology. 2014;141:157-165. doi: 10.1111/imm.12195
12. Westra HJ, Peters MJ, Esko T, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat Genet. 2013;45:1238-1243. doi: 10.1038/ng.2756
13. Võsa U, Claringbould A, Westra H-J, et al. Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis. bioRxiv. 2018:447367. doi: 10.1101/447367
14. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018;562:203-209. doi: 10.1038/s41586-018-0579-z
15. Wu P, Gifford A, Meng X, et al. Mapping ICD-10 and ICD-10-CM Codes to Phecodes: Workflow Development and Initial Evaluation. JMIR Med Inform. 2019;7:e14325. doi: 10.2196/14325
16. Sinnott-Armstrong N, Tanigawa Y, Amar D, et al. Genetics of 38 blood and urine biomarkers in the UK Biobank. bioRxiv. 2019:660506. doi: 10.1101/660506
17. Xu Y, Vuckovic D, Ritchie SC, et al. Learning polygenic scores for human blood cell traits. bioRxiv. 2020:2020.2002.2017.952788. doi: 10.1101/2020.02.17.952788
18. Zhou J, Sun Y, Huang W, Ye K. Additional Supporting Materials for Zhou et al 2021 Gerontology. figshare. 2021. doi: 10.6084/m9.figshare.13637984.v1

19. Carroll RJ, Bastarache L, Denny JC. R PheWAS: data analysis and plotting tools for phenome-wide association studies in the R environment. Bioinformatics. 2014;30:2375-2376. doi: 10.1093/bioinformatics/btu197
20. Canela-Xandri O, Rawlik K, Tenesa A. An atlas of genetic associations in UK Biobank. Nat Genet. 2018;50:1593-1599. doi: 10.1038/s41588-018-0248-z
21. Astle WJ, Elding H, Jiang T, et al. The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. Cell. 2016;167:1415-1429 e1419. doi: 10.1016/j.cell.2016.10.042
22. Hemani G, Zheng J, Elsworth B, et al. The MR-Base platform supports systematic causal inference across the human phenome. Elife. 2018;7. doi: 10.7554/eLife.34408
23. Consortium GT, Laboratory DA, Coordinating Center -Analysis Working G, et al. Genetic effects on gene expression across human tissues. Nature. 2017;550:204-213. doi: 10.1038/nature24277
24. Tajuddin SM, Schick UM, Eicher JD, et al. Large-Scale Exome-wide Association Analysis Identifies Loci for White Blood Cell Traits and Pleiotropy with Immune-Mediated Diseases. Am J Hum Genet. 2016;99:22-39. doi: 10.1016/j.ajhg.2016.05.003
25. van der Harst P, Zhang W, Mateo Leach I, et al. Seventy-five genetic loci influencing the human red blood cell. Nature. 2012;492:369-375. doi: 10.1038/nature11677
26. Gieger C, Radhakrishnan A, Cvejic A, et al. New gene functions in megakaryopoiesis and platelet formation. Nature. 2011;480:201-208. doi: 10.1038/nature10659
27. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics. 2015;31:3555-3557. doi: 10.1093/bioinformatics/btv402
28. Behzad S, Aghaghazvini L, Radmard AR, Gholamrezanezhad A. Extrapulmonary manifestations of COVID-19: Radiologic and clinical overview. Clin Imaging. 2020;66:35-41. doi: 10.1016/j.clinimag.2020.05.013
29. Kuri-Cervantes L, Pampena MB, Meng W, et al. Comprehensive mapping of immune perturbations associated with severe COVID-19. Sci Immunol. 2020;5. doi: 10.1126/sciimmunol.abd7114
30. Vabret N, Britton GJ, Gruber C, et al. Immunology of COVID-19: Current State of the Science. Immunity. 2020;52:910-941. doi: 10.1016/j.immuni.2020.05.002
31. Liu J, Liu Y, Xiang P, et al. Neutrophil-to-lymphocyte ratio predicts critical illness patients with 2019 coronavirus disease in the early stage. J Transl Med. 2020;18:206. doi: 10.1186/s12967-020-02374-0
32. Liu Y, Sun W, Guo Y, et al. Association between platelet parameters and mortality in coronavirus disease 2019: Retrospective cohort study. Platelets. 2020;31:490-496. doi: 10.1080/09537104.2020.1754383
33. Lippi G, Plebani M, Henry BM. Thrombocytopenia is associated with severe coronavirus disease 2019 (COVID-19) infections: A meta-analysis. Clin Chim Acta. 2020;506:145-148. doi: 10.1016/j.cca.2020.03.022
34. Sun DW, Zhang D, Tian RH, et al. The underlying changes and predicting role of peripheral blood inflammatory cells in severe COVID-19 patients: A sentinel? Clin Chim Acta. 2020;508:122-129. doi: 10.1016/j.cca.2020.05.027
35. Qian GQ, Yang NB, Ding F, et al. Epidemiologic and clinical characteristics of 91 hospitalized patients with COVID-19 in Zhejiang, China: a retrospective, multi-centre case series. QJM. 2020;113:474-481. doi: 10.1093/qjmed/hcaa089
36. Qin C, Zhou L, Hu Z, et al. Dysregulation of Immune Response in Patients With Coronavirus 2019 (COVID-19) in Wuhan, China. Clin Infect Dis. 2020;71:762-768. doi: 10.1093/cid/ciaa248

37. Zhang D, Guo R, Lei L, et al. COVID-19 infection induces readily detectable morphological and inflammation-related phenotypic changes in peripheral blood monocytes, the severity of which correlate with patient outcome. medRxiv. 2020:2020.2003.2024.20042655. doi: 10.1101/2020.03.24.20042655
38. Zhou Y, Fu B, Zheng X, et al. Pathogenic T-cells and inflammatory monocytes incite inflammatory storms in severe COVID-19 patients. National Science Review. 2020;7:998-1002. doi: 10.1093/nsr/nwaa041
39. Giamarellos-Bourboulis EJ, Netea MG, Rovina N, et al. Complex Immune Dysregulation in COVID-19 Patients with Severe Respiratory Failure. Cell Host Microbe. 2020;27:992-1000 e1003. doi: 10.1016/j.chom.2020.04.009
40. Merad M, Martin JC. Pathological inflammation in patients with COVID-19: a key role for monocytes and macrophages. Nat Rev Immunol. 2020;20:355-362. doi: 10.1038/s41577-020-0331-4
41. Gong J, Dong H, Xia SQ, et al. Correlation Analysis Between Disease Severity and Inflammation-related Parameters in Patients with COVID-19 Pneumonia. medRxiv. 2020:2020.2002.2025.20025643. doi: 10.1101/2020.02.25.20025643
42. Tanni F, Akker E, Zaman MM, Figueroa N, Tharian B, Hupart KH. Eosinopenia and COVID-19. The Journal of the American Osteopathic Association. 2020;120:504-508. doi: 10.7556/jaoa.2020.091
43. Li Q, Ding X, Xia G, et al. Eosinopenia and elevated C-reactive protein facilitate triage of COVID-19 patients in fever clinic: a retrospective case-control study. EClinicalMedicine. 2020:100375. doi: 10.1016/j.eclinm.2020.100375
44. Xie G, Ding F, Han L, Yin D, Lu H, Zhang M. The role of peripheral blood eosinophil counts in COVID-19 patients. Allergy. 2020. doi: 10.1111/all.14465
45. Zhang JJ, Dong X, Cao YY, et al. Clinical characteristics of 140 patients infected with SARS-CoV-2 in Wuhan, China. Allergy. 2020;75:1730-1741. doi: 10.1111/all.14238
46. Lippi G, Henry BM. Eosinophil count in severe coronavirus disease 2019. QJM. 2020;113:511-512. doi: 10.1093/qjmed/hcaa137
47. Barnes BJ, Adrover JM, Baxter-Stoltzfus A, et al. Targeting potential drivers of COVID-19: Neutrophil extracellular traps. J Exp Med. 2020;217. doi: 10.1084/jem.20200652
48. Matheson NJ, Lehner PJ. How does SARS-CoV-2 cause COVID-19? Science. 2020;369:510-511. doi: 10.1126/science.abc6156
49. Liao M, Liu Y, Yuan J, et al. Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. Nat Med. 2020;26:842-844. doi: 10.1038/s41591-020-0901-9
50. Rodriguez L, Pekkarinen PT, Lakshmikanth T, et al. Systems-Level Immunomonitoring from Acute to Recovery Phase of Severe COVID-19. Cell Rep Med. 2020;1:100078. doi: 10.1016/j.xcrm.2020.100078
51. Giambartolomei C, Vukcevic D, Schadt EE, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. 2014;10:e1004383. doi: 10.1371/journal.pgen.1004383

Figure Legends

Figure 1. A Manhattan plot showing the associations between the severe COVID-19 risk variant and 923 phenotypes in UK Biobank. Each triangle represents one phenotype. Triangles pointing up indicate increasing effects of the COVID-19 risk allele on the phenotypes, while those pointing down indicate decreasing effects. The size of the triangle is proportional to the effect size. The significance threshold with Bonferroni correction ($p < 0.05 / 923 = 5.42\times10^{-5}$) is represented by the red dashed line. Logistic regression was performed for all binary traits, while linear regression was done on continuous traits. Summary statistics, including the sample size, for each trait are in additional supporting materials.

Figure 2. Associations between the expression of candidate genes and the polygenic score of blood cell traits. Each column corresponds to a gene. Each row corresponds to a polygenic score of a blood cell trait. Row names are organized as the combination of trait name, study number for the GWAS providing summary statistics, and the sample ancestry. EUR refers to European ancestry, while ALL refers to multi-ancestry. If no ancestry label is present, the study used only European samples. All PGS shown in this figure were calculated with a *p*-value cut off of 5×10^{-8}. Complete association results for PGS calculated with other *p*-value cutoffs could be found in additional supporting materials. Blood cell traits are categorized into three groups: platelet, red blood cells, and white blood cells. The effects of association, Z-score, are shown as the heatmap. The statistical significance is indicated with "*" ($p < 0.05$) or "**" (FDR < 0.05).

Figure 3. Schematics of possible pathways between the severe COVID-19 risk variant and three blood cell traits. (A) monocyte count, (B) eosinophil count, and (C) neutrophil count. The directions of effects, either up-regulating or down-regulating, were inferred from PheWAS, eQTL analysis in blood samples, and PGS association analysis in blood samples.

Tables

Table 1. Associations between the severe COVID-19 risk variant and blood cell traits

Trait	This Study in UKB N = 310,999			GeneATLAS in UKB[a] N = 452,264		Meta-analysis of three UK cohorts N = 173,480		
	Beta ($\times10^{-3}$)	SE ($\times10^{-4}$)	*p*	Beta ($\times10^{-2}$)	*p*	Beta ($\times10^{-2}$)	SE ($\times10^{-3}$)	*p*
MONO#	-3.78	6.61	1.07e-08	-0.53	1.15e-21	-4.19	7.01	2.34e-09
MONO%	-4.81	6.63	4.09e-13	-7.80	8.79e-26	-4.48	6.99	1.45e-10
EO#	-1.80	6.52	5.73e-03	-0.20	3.81e-06	-2.39	7.00	6.29e-04
EO%	-2.00	6.52	2.20e-03	-2.66	1.54e-05	-2.38	6.99	6.66e-04

NEUT#	1.39	6.50	0.032	1.04	0.035	0.42	7.04	0.55
NEUT%	1.91	6.50	3.23e-03	13.18	1.03e-05	1.47	7.01	0.036
MCH	1.33	6.56	0.042	1.29	0.023	1.44	6.95	0.039
MCHC	1.34	6.61	0.042	0.64	0.070	-0.55	6.80	0.42

Notes: Risk allele A is the effect allele. MONO# = Monocyte count; MONO% = Monocyte percentage; EO# = Eosinophil count; EO% = Eosinophil percentage; NEUT# = Neutrophil count; NEUT% = Neutrophil percentage; MCH = Mean corpuscular hemoglobin; MCHC = Mean corpuscular hemoglobin concentration.

[a] SE is not available in GeneATLAS.

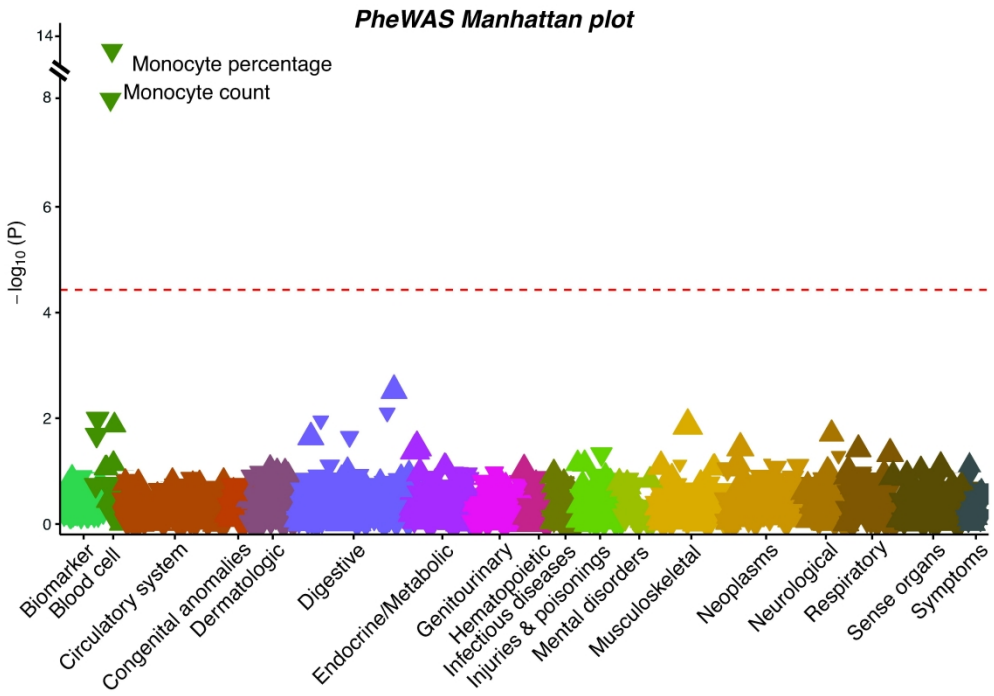
Table 2. The effect of age and sex on white blood cell traits

Trait	Age			Sex			Sex×Age		
	Beta ($\times10^{-3}$)	SE ($\times10^{-4}$)	*p*	Beta ($\times10^{-1}$)	SE ($\times10^{-2}$)	*p*	Beta ($\times10^{-3}$)	SE ($\times10^{-4}$)	*p*
WBC#	-1.95	2.45	$< 1.61\times10^{-3}$	-6.03	2.04	$< 1.61\times10^{-3}$	11.18	3.58	$< 1.61\times10^{-3}$
MONO#	3.90	2.43	$< 1.61\times10^{-3}$	-1.23	2.02	$< 1.61\times10^{-3}$	7.08	3.54	$< 1.61\times10^{-3}$
MONO%	6.91	2.41	$< 1.61\times10^{-3}$	3.04	2.01	$< 1.61\times10^{-3}$	1.56	3.51	$< 1.61\times10^{-3}$
EO#	-1.23	2.45	$< 1.61\times10^{-3}$	0.16	2.04	0.42	2.87	3.57	$< 1.61\times10^{-3}$
EO%	-0.87	2.45	$< 1.61\times10^{-3}$	2.71	2.04	$< 1.61\times10^{-3}$	-1.47	3.57	$< 1.61\times10^{-3}$
NEUT#	-6.32	2.45	$< 1.61\times10^{-3}$	-9.97	2.04	$< 1.61\times10^{-3}$	18.50	3.57	$< 1.61\times10^{-3}$
NEUT%	-8.72	2.45	$< 1.61\times10^{-3}$	-11.01	2.04	$< 1.61\times10^{-3}$	20.61	3.57	$< 1.61\times10^{-3}$
BASO#	-4.63	2.46	$< 1.61\times10^{-3}$	-4.70	2.05	$< 1.61\times10^{-3}$	7.05	3.58	$< 1.61\times10^{-3}$

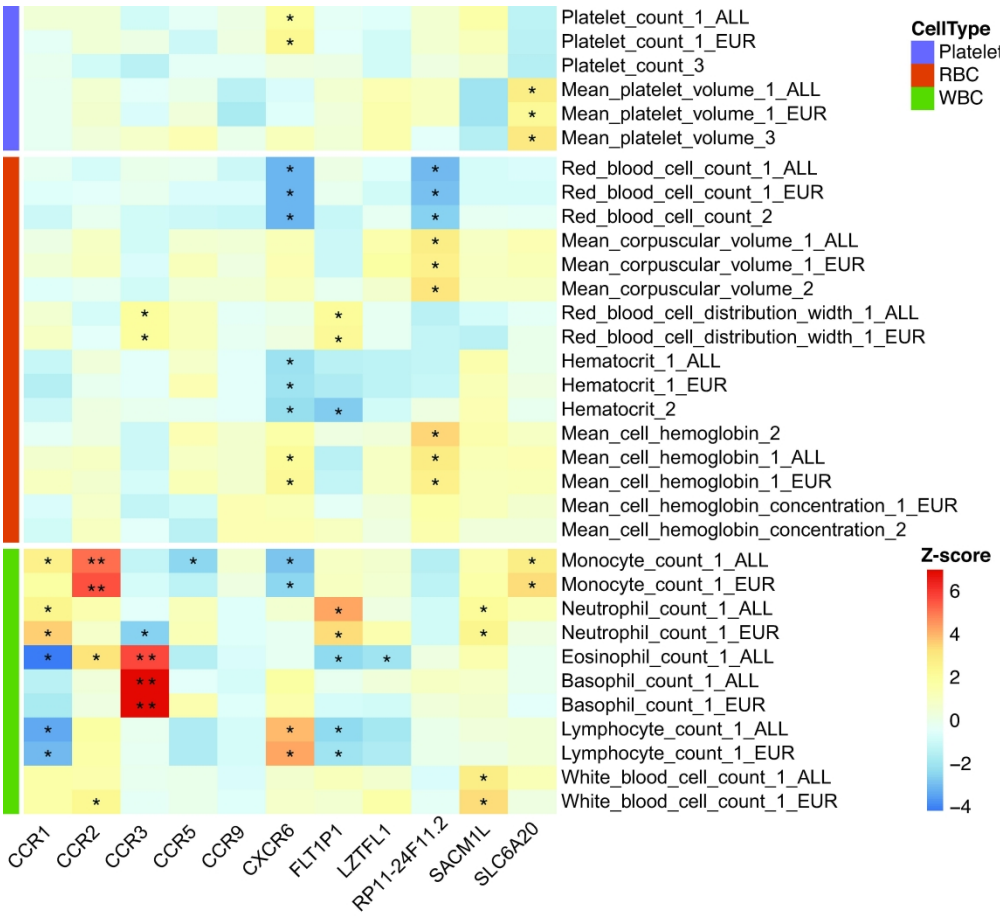
BASO%	-4.68	2.46	$< 1.61\times10^{-3}$	-3.39	2.05	$< 1.61\times10^{-3}$	4.49	3.58	$< 1.61\times10^{-3}$
LYMPH#	3.59	2.46	$< 1.61\times10^{-3}$	1.62	2.05	$< 1.61\times10^{-3}$	-4.48	3.58	$< 1.61\times10^{-3}$
LYMPH%	8.02	2.43	$< 1.61\times10^{-3}$	11.02	2.02	$< 1.61\times10^{-3}$	-23.99	3.54	$< 1.61\times10^{-3}$

Notes: The full results for all blood cell traits, including those related to red blood cells and platelets are available in Supplementary Table 5.

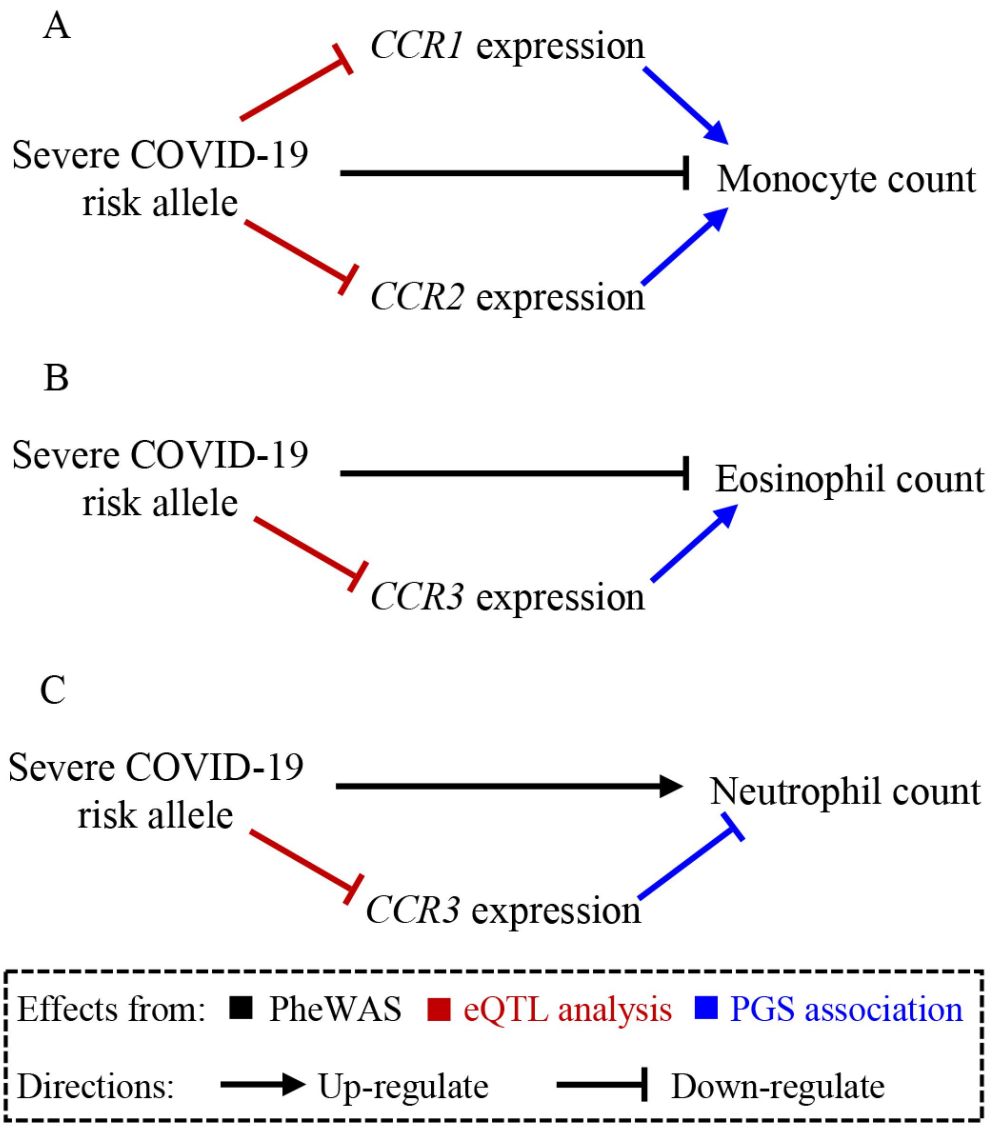
WBC# = White blood cell count; BASO# = Basophil count; BASO% = Basophil percentage; LYMPH# = Lymphocyte count; LYMPH% = Lymphocyte percentage.



A Manhattan plot showing the associations between the severe COVID-19 risk variant and 923 phenotypes in UK Biobank.



Associations between the expression of candidate genes and the polygenic score of blood cell traits.



Schematics of possible pathways between the severe COVID-19 risk variant and three blood cell traits.