

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

Jingqi Zhou<sup>1,2</sup>, Yitang Sun<sup>1</sup>, Weishan Huang<sup>3,4</sup>, Kaixiong Ye<sup>1,5\*</sup>

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](mailto:Kaixiong.Ye@uga.edu)

## Abstract

**Purpose:** 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 of the COVID-19 risk variant, we performed a phenome-wide association study (PheWAS) with 923 phenotypes in 310,999 European individuals from UK Biobank. For candidate target genes, 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.

**Results:** Our PheWAS identified and replicated multiple blood cell traits to be associated with the COVID-19 risk variant, including monocyte count and percentage ( $p = 1.07 \times 10^{-8}$ ,  $4.09 \times 10^{-13}$ ), eosinophil count and percentage ( $p = 5.73 \times 10^{-3}$ ,  $2.20 \times 10^{-3}$ ), and neutrophil percentage ( $p = 3.23 \times 10^{-3}$ ). 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 ( $p = 5.73 \times 10^{-21}$ ,  $5.08 \times 10^{-19}$ ); *CCR2* with monocytes ( $p = 2.40 \times 10^{-10}$ ); and *CCR1* with monocytes and neutrophil ( $p = 1.78 \times 10^{-6}$ ,  $7.17 \times 10^{-5}$ ).

**Conclusions:** Multiple blood cell traits, especially monocyte, eosinophil, and neutrophil numbers, are associated with the COVID-19 risk variant and the expression of its candidate target genes, representing probable mechanistic links between the genetic locus 3p21.31 and severe COVID-19.

**Key Words:** COVID-19; blood cells; monocyte; eosinophil; neutrophil; 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<sup>1-3</sup>. Besides demographic factors and pre-existing conditions, genetic variation is partly responsible for these varying individual responses<sup>4-6</sup>. The first genome-wide association study (GWAS) for COVID-19 in patients with respiratory failure identified genetic associations at locus 3p21.31<sup>4</sup>, which was independently replicated by the COVID-19 Host Genetics Initiative<sup>5</sup>. The peak signal at this locus spans multiple chemokine receptor genes (e.g., *CCR9*, *CXCR6*, *XCRI* and *CCRI*) and risk variants are associated with the expression of *CXCR6*, *CCRI* and *SLC6A20*<sup>4</sup>. However, the underlying causal variant, the target gene, and the pathophysiological process are 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. It may also reveal unexpected comorbidities that indicate shared biological mechanisms<sup>7,8</sup>. 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<sup>9,10</sup>. From the perspective of a candidate gene, insights could be gained into its physiological pathways and downstream functional effects by examining

associations of its expression level with phenotypes across the phenome, or even with the genetically predicted phenotypic status if measured ones are unavailable<sup>10</sup>.

This project aimed to explore the mechanistic link between the genetic locus 3p21.31 and severe COVID-19. We first leveraged the deep phenotyping information in the UK Biobank (N = 310,999) and performed a PheWAS for a COVID-19 risk variant with 923 disease phenotypes, biomarkers and blood cell traits. Candidate associations were further replicated in an additional dataset. 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.

## **MATERIALS AND METHODS**

### **Ethics statement**

UK Biobank is a large population-based prospective study that recruited more than 500,000 individuals aged 40-69 years between 2006 and 2010. It was approved by the North West Multi-Centre Research Ethics Committee and proper informed consent was obtained. All participants received baseline measurements, donated biological materials, and provided access to their medical records<sup>11</sup>. 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**

Discovery analysis. This was performed with data from UK Biobank. Only participants fulfilling the following criteria were included: 1) genetic ancestry is Caucasian; 2) included in the genetic principal component analysis; 3) not outliers for heterogeneity and missing genotype rate; 4) no sex chromosome aneuploidy; 5) self-reported sex matching genetic sex; 6) no high degree of genetic kinship, and 7) 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 in our PheWAS: 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<sup>12</sup>. A total of 858 phecodes with case number no less than 200 were retained in our analysis (Supplementary Table 1). For continuous traits, our PheWAS included 34 blood and

urine biochemistry markers, and 31 blood cell traits (Supplementary Table 2)<sup>11</sup>. Statistical association analyses were performed with scripts in R language and the PheWAS package<sup>13</sup>. 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. We applied Bonferroni correction for the total number of phenotypes tested, although we note that this is a conservative approach because the phenotypes are not independent. Summary statistics for nominally significant associations between the COVID-19 risk variant and blood cell traits were also retrieved from GeneATLAS, a large database of associations based on 452,264 UK Biobank White British individuals<sup>14</sup>.

Replication analysis. For blood cell traits that are associated with the COVID-19 risk variant at the nominal significance level in the discovery analysis, we tested if they could be replicated in another existing study. It was 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)<sup>15</sup>. Summary statistics for this study were retrieved from the IEU OpenGWAS database<sup>16</sup>.

### **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<sup>10,17</sup>. The GTEx project studies tissue-specific gene expression and regulation in 54 non-diseased tissue sites from about 1,000 individuals<sup>17</sup>. The eQTLGen Consortium conducted eQTL meta-analysis in 31,684 samples of blood and peripheral blood mononuclear cells from 37 datasets<sup>10</sup>. It also performed PGS analysis

to evaluate the associations between the expression level of most genes and the polygenic scores of 1,263 traits<sup>10</sup>. The majority of samples in both studies are of European ancestry. In the PGS analysis, multiple PGS were calculated for each trait with different GWAS, sample ancestry, and  $p$ -value cutoffs ( $p = 0.01, 1 \times 10^{-3}, 1 \times 10^{-4}, 1 \times 10^{-5}, 5 \times 10^{-8}$ ). For blood cell traits, three previous GWAS were used and designated as study 1<sup>18</sup>, study 2<sup>19</sup>, and study 3<sup>20</sup>, respectively. Statistical significance was defined with the false discovery rate approach ( $FDR < 0.05$ )<sup>10</sup>.

## Resources

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

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

GTEEx: <https://www.gtexportal.org/home/>

The COVID-19 GWAS Results Browser: [https://ikmb.shinyapps.io/COVID-19\\_GWAS\\_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/>



## 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,  $p = 4.69 \times 10^{-11}$ ) for severe COVID-19 after adjustment for genetic principal components, age and sex<sup>4</sup>. 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,  $p = 9.46 \times 10^{-12}$ ) in European populations<sup>21</sup>. 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 3 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 (Fig. 1, Supplementary Tables 1 and 2).

With the conservative Bonferroni correction for the total number of phenotypes tested ( $p < 5.42 \times 10^{-5}$ ), we observed that the severe COVID-19 risk variant is associated with monocyte percentage ( $p = 4.09 \times 10^{-13}$ ) and monocyte count ( $p = 1.07 \times 10^{-8}$ ). 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 ( $p = 4.76 \times 10^{-4}$ ), periodontitis ( $p = 1.53 \times 10^{-3}$ ) and its subcategory, chronic periodontitis ( $p = 2.43 \times 10^{-3}$ ). At the nominal significance level ( $p < 0.05$ ), associations were observed with additional blood cell traits (Table 1): eosinophil count and percentage, neutrophil count and percentage, mean corpuscular hemoglobin, and mean corpuscular hemoglobin concentration. In GeneATLAS, a large database of associations based

on 452,264 UK Biobank White British individuals<sup>14</sup>, consistent and even more significant associations were observed (Supplementary Table 4): monocyte percentage ( $p = 8.79 \times 10^{-26}$ ), monocyte count ( $p = 1.15 \times 10^{-21}$ ), eosinophil count ( $p = 3.81 \times 10^{-6}$ ), eosinophil percentage ( $p = 1.54 \times 10^{-5}$ ), neutrophil percentage ( $p = 1.03 \times 10^{-5}$ ), basophil count ( $p = 1.40 \times 10^{-3}$ ), basophil percentage ( $p = 1.87 \times 10^{-3}$ ), red blood cell (RBC) count ( $p = 0.022$ ), mean corpuscular hemoglobin ( $p = 0.023$ ), and neutrophil count ( $p = 0.035$ ). Notably, most of these associations were replicated in another study of 173,480 European-ancestry individuals (Table 1)<sup>15</sup>.

### **Expression of candidate target genes is associated with PGS of blood cell traits**

Candidate target genes of the COVID-19 risk variant through regulation of gene expression could be identified with eQTL analysis. Based on *cis*-eQTL analysis in 54 tissues from GTEx<sup>17</sup>, 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 5). Moreover, eQTLGen<sup>10</sup>, a meta-analysis for *cis*-eQTL in 31,684 blood samples additionally identified the following genes: *FLT1P1*, *CCR3*, *SACMIL*, *CCR5*, *CCR2* and *RP11-24F11.2* (Supplementary Table 6). *Trans*-eQTL analysis in both studies did not identify any genes.

For all these candidate 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<sup>10</sup>. Multiple significant associations were identified after correction for multiple testing (Fig. 2, Supplementary Table 7). Genetically predicted higher monocyte count is positively associated with the expression of *CCR1* ( $p = 1.78 \times 10^{-6}$ ) and *CCR2* ( $p = 2.40 \times 10^{-10}$ ). Genetically predicted

higher eosinophil count and basophil count are positively associated with *CCR3* expression ( $p = 5.73 \times 10^{-21}$  and  $p = 5.08 \times 10^{-19}$ , 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* ( $p = 7.17 \times 10^{-5}$ ), *FLTIP1* ( $p = 1.46 \times 10^{-5}$ ), *SACMIL* ( $p = 5.69 \times 10^{-3}$ ) and *CCR3* ( $p = 0.011$ ). A negative association was observed between predicted lymphocyte count and *CCR1* ( $p = 9.41 \times 10^{-4}$ ). Genetically predicted higher RBC count is associated with lower expression of *CXCR6* ( $p = 3.74 \times 10^{-4}$ ) and *RP11-24F11.2* ( $p = 2.09 \times 10^{-3}$ ), while for WBC count, a positive association was observed with *SACMIL* ( $p = 8.26 \times 10^{-4}$ ). Notably, these associations were consistent across different GWAS datasets and  $p$ -value cutoffs used in the PGS calculation (Supplementary Table 7). 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 (Fig. 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.

## 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, *CCRI*, *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.

Hematologic manifestations are common in COVID-19 patients, especially elevated WBC and neutrophil counts but decreased lymphocyte and platelet counts<sup>22,23</sup>. Leukocytosis, neutrophilia, lymphopenia, thrombocytopenia, and neutrophil-to-lymphocyte ratio have been repeatedly associated with worse COVID-19 outcomes and could serve as prognostic biomarkers<sup>1,3,24,25</sup>.

Reducing basophil count or percentage was generally observed in patients<sup>26,27</sup>. For monocyte, its total number in circulation does not change dramatically in COVID-19 patients<sup>22,27,28</sup>. However, its composition exhibits a pronounced shift, with a significant expansion of its inflammatory subsets, which are not typically seen in healthy individuals<sup>22,23,28-30</sup>. The pattern of eosinophil is less well-established. Some studies observed diminished and even undetectable eosinophil counts (i.e., eosinopenia) in COVID-19 patients<sup>26,27,31-34</sup>, and it was also shown that eosinophil counts are positively associated with lymphocyte counts in both severe and non-severe cases<sup>34</sup>. However, others did not find a significant difference<sup>35</sup>, and there is also a report of an expanded eosinophil percentage among the total viable leukocyte CD45+ population<sup>22</sup>. 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<sup>23,36-38</sup>. Notably, an immuno-monitoring study of COVID-19 patients from acute to recovery phases observed gradual reduction of neutrophil and replenishment of basophil, eosinophil and non-inflammatory monocyte<sup>39</sup>.

Our PheWAS in UK Biobank for the severe COVID-19 risk variant, with replications in another dataset, revealed that the risk allele is associated with decreased monocyte count and percentage, eosinophil count and percentage, but with increased neutrophil percentage. GeneATLAS reported even more significant associations for these relationships, probably due to its different quality control procedures and a larger sample size<sup>14</sup>. It also reports suggestive evidence of negative associations between the risk allele and basophil count and percentage (Supplementary Table 4). 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.

Our PGS analysis for the potential target genes of the severe COVID-19 risk variant 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 (Supplementary Table 7). Intersecting and reconciling association signals across PheWAS, eQTL, and PGS analysis yielded multiple possible pathways for the severe 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. While the association with neutrophil percentage was replicated, the association with neutrophil count was only significant in UK Biobank. 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<sup>17</sup>. Also, the eQTL analyses were based on generally healthy samples<sup>10,17</sup>. 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.

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. 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 severe COVID-19 risk variant affects other unexplored intermediate traits. 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<sup>40</sup>. Our analyses were restricted to population samples 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.

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 hypotheses that call for follow-up studies into their roles in COVID-19 pathogenesis.

## **ACKNOWLEDGES**

The authors would like to thank the UK Biobank participants and administrators for data access. We also want to thank all other Ye lab members for stimulating discussions. KY is supported by

the University of Georgia Research Foundation. Funding sources had no involvement in the conception, design, analysis, or presentation of this work.

## **DISCLOSURE**

The authors declare no conflicts of interest.



## 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(10229):1054-1062.
2. Oran DP, Topol EJ. Prevalence of Asymptomatic SARS-CoV-2 Infection: A Narrative Review. *Ann Intern Med*. 2020.
3. 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.
4. Ellinghaus D, Degenhardt F, Bujanda L, et al. Genomewide Association Study of Severe Covid-19 with Respiratory Failure. *N Engl J Med*. 2020.
5. 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(6):715-718.
6. Murray MF, Kenny EE, Ritchie MD, et al. COVID-19 outcomes and the human genome. *Genet Med*. 2020.
7. Bush WS, Oetjens MT, Crawford DC. Unravelling the human genome-phenome relationship using phenome-wide association studies. *Nat Rev Genet*. 2016;17(3):129-145.
8. Hebring SJ. The challenges, advantages and future of phenome-wide association studies. *Immunology*. 2014;141(2):157-165.
9. Westra HJ, Peters MJ, Esko T, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet*. 2013;45(10):1238-1243.

10. Võsa U, Claringbould A, Westra H-J, et al. Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis. *bioRxiv*. 2018:447367.
11. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 2018;562(7726):203-209.
12. 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(4):e14325.
13. 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(16):2375-2376.
14. Canela-Xandri O, Rawlik K, Tenesa A. An atlas of genetic associations in UK Biobank. *Nat Genet*. 2018;50(11):1593-1599.
15. 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(5):1415-1429 e1419.
16. Hemani G, Zheng J, Elsworth B, et al. The MR-Base platform supports systematic causal inference across the human phenome. *Elife*. 2018;7.
17. Consortium GT, Laboratory DA, Coordinating Center -Analysis Working G, et al. Genetic effects on gene expression across human tissues. *Nature*. 2017;550(7675):204-213.
18. 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(1):22-39.
19. van der Harst P, Zhang W, Mateo Leach I, et al. Seventy-five genetic loci influencing the human red blood cell. *Nature*. 2012;492(7429):369-375.

20. Gieger C, Radhakrishnan A, Cvejic A, et al. New gene functions in megakaryopoiesis and platelet formation. *Nature*. 2011;480(7376):201-208.
21. 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(21):3555-3557.
22. Kuri-Cervantes L, Pampena MB, Meng W, et al. Comprehensive mapping of immune perturbations associated with severe COVID-19. *Sci Immunol*. 2020;5(49).
23. Vabret N, Britton GJ, Gruber C, et al. Immunology of COVID-19: Current State of the Science. *Immunity*. 2020;52(6):910-941.
24. 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(1):206.
25. 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.
26. 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.
27. 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(15):762-768.
28. 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.

29. 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(6):998-1002.
30. Merad M, Martin JC. Pathological inflammation in patients with COVID-19: a key role for monocytes and macrophages. *Nat Rev Immunol*. 2020;20(6):355-362.
31. 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(8):504-508.
32. 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.
33. 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.
34. Zhang JJ, Dong X, Cao YY, et al. Clinical characteristics of 140 patients infected with SARS-CoV-2 in Wuhan, China. *Allergy*. 2020;75(7):1730-1741.
35. Lippi G, Henry BM. Eosinophil count in severe coronavirus disease 2019. *QJM*. 2020;113(7):511-512.
36. Barnes BJ, Adrover JM, Baxter-Stoltzfus A, et al. Targeting potential drivers of COVID-19: Neutrophil extracellular traps. *J Exp Med*. 2020;217(6).
37. Matheson NJ, Lehner PJ. How does SARS-CoV-2 cause COVID-19? *Science*. 2020;369(6503):510-511.
38. Liao M, Liu Y, Yuan J, et al. Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. *Nat Med*. 2020;26(6):842-844.

39. Rodriguez L, Pekkarinen P, Tadepally LK, et al. Systems-level immunomonitoring from acute to recovery phase of severe COVID-19. medRxiv. 2020:2020.2006.2003.20121582.
40. 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(5):e1004383.

## TABLES

Table 1. Associations between the COVID-19 risk variant (rs67959919) and blood cell traits

Trait	Discovery (UK Biobank) N = 310,999			Replication (Meta-analysis of three UK cohorts) N = 173,480		
	beta	SE	<i>p</i>	beta	SE	<i>p</i>
Monocyte count	-0.0038	6.61e-4	1.07e-08	-0.042	7.01e-3	2.34e-09
Monocyte percentage	-0.0048	6.63e-4	4.09e-13	-0.045	6.99e-3	1.45e-10
Eosinophil count	-0.0018	6.52e-4	5.73e-03	-0.024	7.00e-3	6.29e-04
Eosinophil percentage	-0.0020	6.52e-4	2.20e-03	-0.024	6.99e-3	6.66e-04
Neutrophil count	0.0014	6.50e-4	0.032	0.0042	7.04e-3	0.55
Neutrophil percentage	0.0019	6.50e-4	3.23e-03	0.015	7.01e-3	0.036
Mean corpuscular hemoglobin	0.0013	6.56e-4	0.042	0.014	6.95e-3	0.039
Mean corpuscular hemoglobin concentration	0.0013	6.61e-4	0.042	-0.0055	6.80e-3	0.42

NOTE: The effect allele is the COVID-19 risk allele.

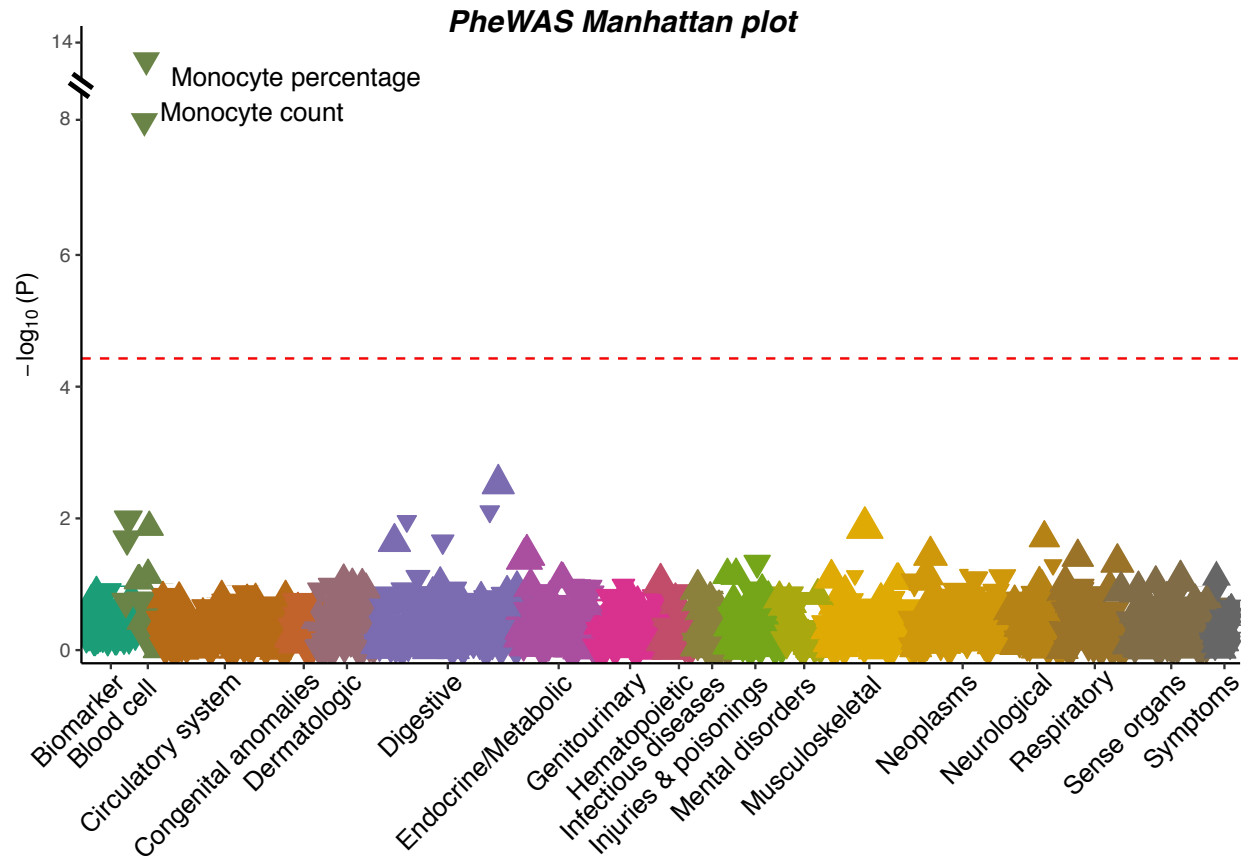
## FIGURE LEGENDS

**Fig. 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 ( $p < 0.05 / 923 = 5.42 \times 10^{-5}$ ) is represented by the red dashed line.

**Fig. 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  $5 \times 10^{-8}$ . Complete association results for PGS calculated with other  $p$ -value cutoffs could be found in Supplementary Table 7. 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 “\*\*\*” ( $FDR < 0.05$ ).

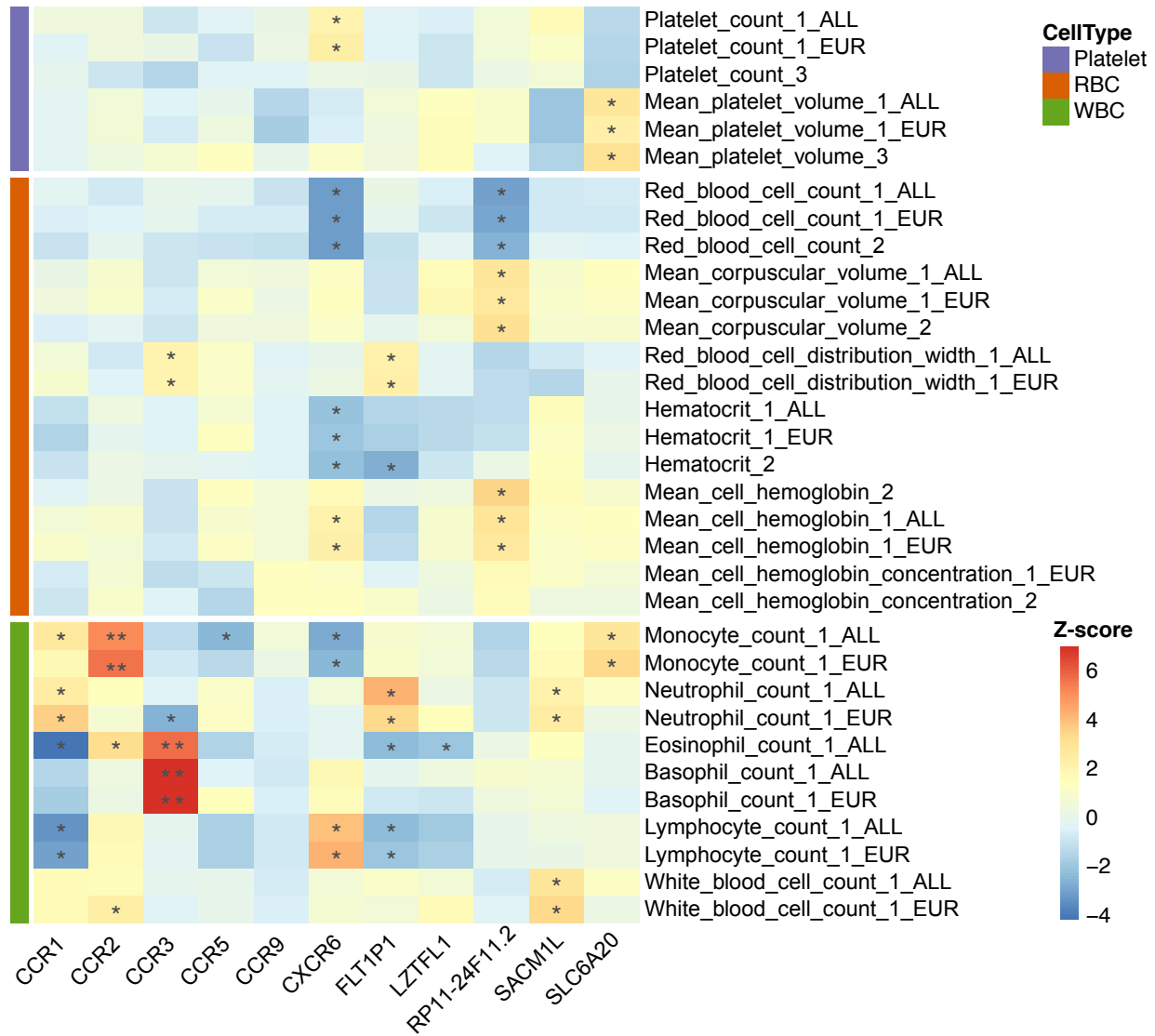
**Fig. 3** Schematics of possible pathways between the severe COVID-19 risk variant and three blood cell traits. (A) monocyte count, (B) eosinophil count, and (C) neutrophil count. The directions of effects, either up-regulating or down-regulating, were inferred from PheWAS, eQTL analysis in blood samples,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

**Figure 1**

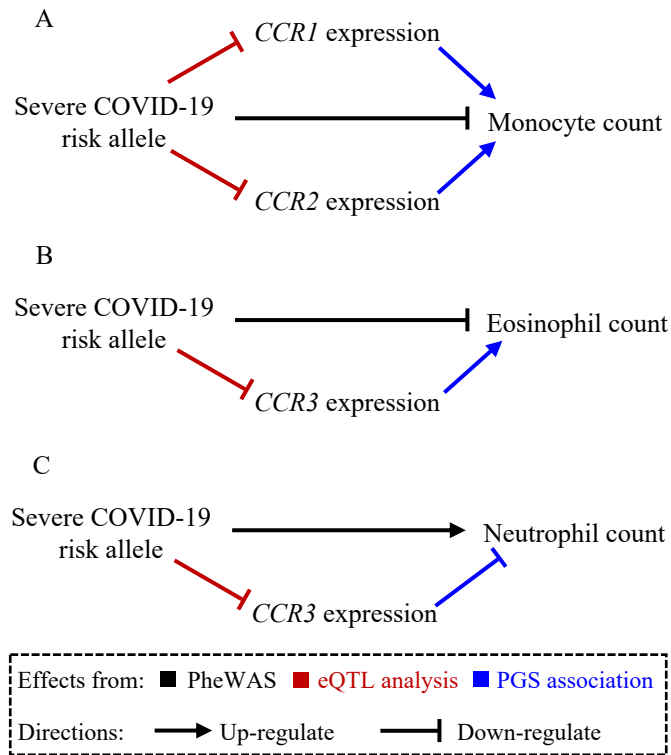




**Figure 2**



**Figure 3**



## **SUPPLEMENTARY INFORMATION**

**Supplementary Table 1.** PheWAS results for binary disease traits.

**Supplementary Table 2.** PheWAS results for continuous traits, including blood and urine biomarkers, and blood cell traits.

**Supplementary Table 3.** Baseline characteristics of UK Biobank participants.

**Supplementary Table 4.** Significant associations between rs67959919 and blood cell traits in GeneATLAS.

**Supplementary Table 5.** Significant SNP-gene expression associations for rs67959919 (G/A) in GTEx.

**Supplementary Table 6.** Significant SNP-gene expression associations for rs67959919 (G/A) in eQTLGen.

**Supplementary Table 7.** Associations between gene expression and the polygenic scores of blood cell counts. Significant p and FDR values are highlighted.

**Supplementary Figure 1.** The expression of *CCR1*, *CCR2*, and *CCR3* across blood cells.

Figures were retrieved from the Human Protein Atlas (<https://www.proteinatlas.org/>). The y axis, NX, stands for normalized expression level.