



Association of *CXCR6* with COVID-19 severity: delineating the host genetic factors in transcriptomic regulation

Yulin Dai¹ · Junke Wang² · Hyun-Hwan Jeong¹ · Wenhao Chen^{3,4} · Peilin Jia¹ · Zhongming Zhao^{1,2,5}

Received: 10 March 2021 / Accepted: 13 June 2021 / Published online: 21 June 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

The coronavirus disease 2019 (COVID-19) is an infectious disease that mainly affects the host respiratory system with ~80% asymptomatic or mild cases and ~5% severe cases. Recent genome-wide association studies (GWAS) have identified several genetic loci associated with the severe COVID-19 symptoms. Delineating the genetic variants and genes is important for better understanding its biological mechanisms. We implemented integrative approaches, including transcriptome-wide association studies (TWAS), colocalization analysis, and functional element prediction analysis, to interpret the genetic risks using two independent GWAS datasets in lung and immune cells. To understand the context-specific molecular alteration, we further performed deep learning-based single-cell transcriptomic analyses on a bronchoalveolar lavage fluid (BALF) dataset from moderate and severe COVID-19 patients. We discovered and replicated the genetically regulated expression of *CXCR6* and *CCR9* genes. These two genes have a protective effect on lung, and a risk effect on whole blood, respectively. The colocalization analysis of GWAS and *cis*-expression quantitative trait loci highlighted the regulatory effect on *CXCR6* expression in lung and immune cells. In the lung-resident memory CD8⁺ T (T_{RM}) cells, we found a 2.24-fold decrease of cell proportion among CD8⁺ T cells and lower expression of *CXCR6* in the severe patients than moderate patients. Pro-inflammatory transcriptional programs were highlighted in the T_{RM} cellular trajectory from moderate to severe patients. *CXCR6* from the *3p21.31* locus is associated with severe COVID-19. *CXCR6* tends to have a lower expression in lung T_{RM} cells of severe patients, which aligns with the protective effect of *CXCR6* from TWAS analysis.

Abbreviations

BALF Bronchoalveolar lavage fluid
BIOS Biobank-based Integrative Omics Studies

ChromHMM Chromatin-state hidden Markov model
COVID-19 Coronavirus disease 2019
CLPP Colocalization posterior probability
CSEA-DB Cell-type-specific expression database
DAP Deterministic approximation of posteriors
DEG Differentially expressed gene
DICE Database of immune cell expression
DrivAER Driving transcriptional programs based on AutoEncoder derived Relevance scores
eQTL Expression quantitative trait
GRex Genetically regulated expression
GWAS Genome-wide association study
HGI Host Genetics Initiative
Hi-C High-throughput chromatin interaction
LD Linkage disequilibrium
MAF Minor allele frequency
MASHR Multivariate adaptive shrinkage in R
Mb Million base pairs
MSigDB Molecular signatures database
PIP Posterior inclusion probability
PWM Position weight matrix

Yulin Dai and Junke Wang contributed equally to this work.

✉ Zhongming Zhao
Zhongming.Zhao@uth.tmc.edu

¹ Center for Precision Health, School of Biomedical Informatics, The University of Texas Health Science Center at Houston, 7000 Fannin St. Suite 600, Houston, TX 77030, USA

² MD Anderson Cancer Center UTHHealth Graduate School of Biomedical Sciences, Houston, TX 77030, USA

³ Immunobiology and Transplant Science Center, Department of Surgery, Houston Methodist Research Institute and Institute for Academic Medicine, Houston Methodist Hospital, Houston, TX 77030, USA

⁴ Department of Surgery, Weill Cornell Medicine, Cornell University, New York, NY 10065, USA

⁵ Human Genetics Center, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA

SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
RCP	Regional colocalization probability
SCGG	Severe COVID-19 GWAS Group
scRNA-seq	Single-cell RNA sequencing
tSNE	T-Distributed Stochastic Neighbor Embedding
TF	Transcription factor
T _{RM} cells	Resident memory CD8 ⁺ T cells
TWAS	Transcriptome-wide association study

Introduction

The coronavirus disease 2019 (COVID-19) pandemic has already infected over 1.6 million people and caused numerous morbidities and over 3.4 million death worldwide as of May 2021. The virus is evolving fast with new variants being emerged in the world (Liu et al. 2020a; Zhao et al. 2004). A huge disparity in the severity of symptoms in different patients has been observed. In some of the patients, only mild symptoms or even no symptoms are shown, and little treatment or interventions are required, while a subset of patients experience rapid disease progression to respiratory failure and need urgent and intensive care (Wu and McGoogan 2020). Although age and sex are major risk factors of COVID-19 disease severity (Bhopal and Bhopal 2020), it remains largely unclear about the factors leading to the variability on COVID-19 severity and which group of individuals confer intrinsic susceptibility to COVID-19.

Several genome-wide association studies (GWAS) have been carried out, and one genomic risk locus, *3p21.31*, has been replicated to be associated with the critical illness. One recent study by the Severe COVID-19 GWAS Group identified *3p21.31* risk locus for the susceptibility to severe COVID-19 with respiratory failure (Severe Covid et al. 2020). This GWAS signal was then replicated in a separate meta-analysis comprising total 2972 cases from nine cohorts by COVID-19 Host Genetics Initiative (HGI) round 4 alpha. However, there is a cluster of six genes (*SLC6A20*, *LZTFL1*, *CCR9*, *FYCO1*, *CXCR6*, and *XCRI*) nearby the lead SNP rs35081325 within a complex linkage disequilibrium (LD) structure, which makes the “causal” gene and functional implication of this locus remain elusive (Pairo-Castineira et al. 2020; Severe Covid et al. 2020).

The majority of GWAS variants are located in non-coding loci, many of which are in the enhancer or promoter regions, playing roles as *cis*- or *trans*-regulatory elements to alter gene expression (GTEx Consortium 2020). Although the function of non-coding variants could not be directly interrupted by their locations, their mediation effect on gene expression could be inferred by the expression quantitative trait loci (eQTL) analysis. In recent years, large consortia

like GTEx (Genotype-Tissue Expression), eQTLGen Consortium, and DICE (database of immune cell expression) have generated rich eQTLs resources in diverse tissues and immune-related cell types (GTEx Consortium 2020; Schmiedel et al. 2018; Vösa et al. 2018a). A variety of statistical approaches, such as transcriptome-wide association study (TWAS) analysis and colocalization analysis, have successfully interpreted the target genes of non-coding variants by integrating the context-specific eQTLs (Dai et al. 2020; Dai et al. 2019; Gamazon et al. 2015; Giambartolomei et al. 2014).

Recent advances in single-cell transcriptome sequencing provide unprecedented opportunities to understand the biological mechanism underlying disease pathogenesis at the single-cell and cell type levels (Dai et al. 2021; Mathys et al. 2019; Papalexi and Satija 2018). The recent generation of single-cell RNA sequencing (scRNA-seq) data from the bronchoalveolar lavage fluid (BALF) of moderate and severe COVID-19 patients has revealed the landscape of the gene expression changes in major immune cells. However, the transcriptome alteration in specific subpopulations remains mostly unexplored (Liao et al. 2020).

In this study, we aimed to connect the genetic factors with the context-specific molecular phenotype in COVID-19 patients. As illustrated in Fig. 1, we designed a multi-level workflow to dissect the genetically regulated expression (GR_{EX}) that contributed to severe COVID-19. We performed TWAS and colocalization analyses with a broad collection of eQTL datasets at the tissue and cellular levels. We further integrated the BALF single-cell transcriptome dataset to explore the cellular transcriptome alterations in severe and moderate COVID-19 patients. Lastly, we proposed a hypothetical mechanism, connecting our multi-layer evidence in host genetic factors, risk genes, and single-cell transcriptome features with the severity of COVID-19.

Materials and methods

GWAS dataset

We obtained GWAS summary statistics for the phenotype “severe COVID-19 patients vs population” (severe COVID-19) from two separate meta-analyses carried out by the COVID-19 Host Genetics Initiative (HGI, <https://www.covid19hg.org/>) and the Severe COVID-19 GWAS Group (SCGG) (Severe Covid et al. 2020). The GWAS_{HGI} A2 round 4 (alpha) cohort consists of 12,816,037 SNPs from the association study of 2972 very severe respiratory confirmed COVID-19 cases and 284,472 controls with unknown SARS-CoV-2 infection status from nine independent studies in a majority of the European Ancestry population. The GWAS_{SCGG} dataset is from the first GWAS of severe

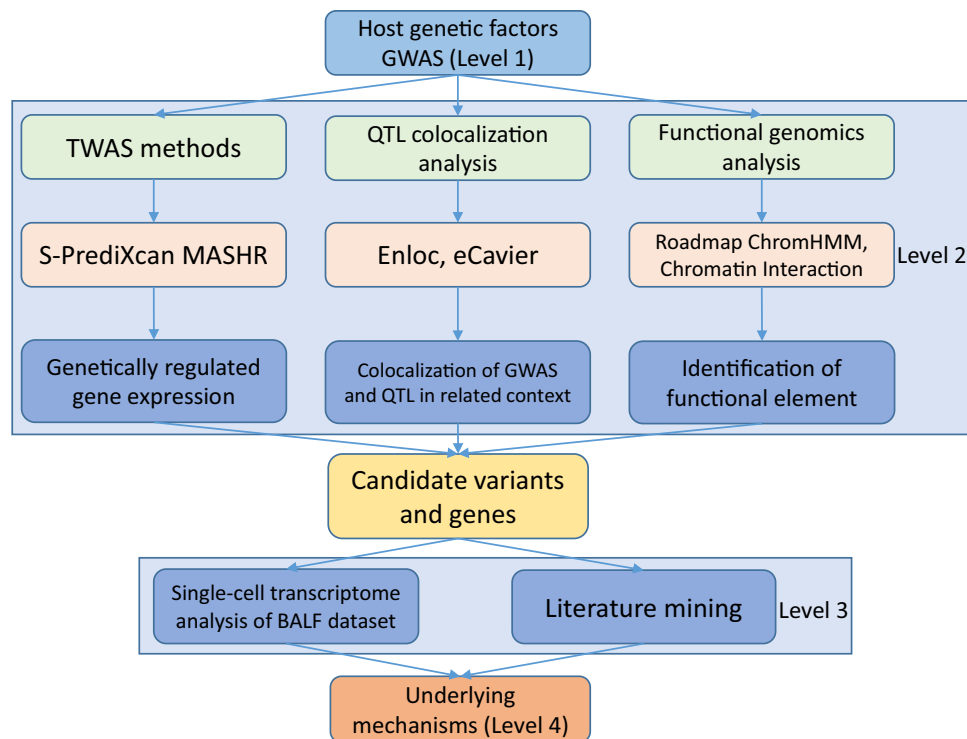


Fig. 1 Workflow of a data-driven study: from genetic factor to molecular phenotype. The study has four major levels. Level 1: We collected the current largest COVID-19 genome-wide association study (GWAS) datasets and a non-duplicated replicate of the severe COVID-19 GWAS dataset. Level 2: We utilized the cutting-edge statistical approaches (transcriptome-wide association study and colocalization analysis) and public functional genomics annotations to dissect the genetic effects on gene expression (“[Materials and methods](#)”). Then, we cross-validated our findings of these methods to

ensure robustness. Level 3: We adapted single-cell RNA sequencing dataset from COVID-19 bronchoalveolar lavage fluid samples. We applied differentially expressed gene analysis and machine learning methods to characterize the molecular changes of candidate genes at single-cell level from COVID-19 moderate and severe patients. We conducted an extensive literature review to explain our observations. Level 4: We proposed a mechanism for explaining the “causal” association of genetic factors and the severity of COVID-19 patients

COVID-19 (Severe Covid et al. 2020), including 8,431,427 SNPs from the association study conducted from 1980 COVID-19 confirmed patients with severe disease status and 2205 control participants from two separate cohorts in Europe.

Transcriptome-wide association analysis

We performed TWAS analyses of severe COVID-19 using S-PrediXcan (Barbeira et al. 2018) to prioritize GWAS findings and identify eQTL-linked genes. S-PrediXcan is a systematic approach that integrates GWAS summary statistics with publicly available eQTL data to translate the evidence of association with a phenotype from the SNP level to the gene level. Briefly, prediction models were built by a flexible and generic approach multivariate adaptive shrinkage in R package (MASHR) using variants with a high probability of being causal for QTL and tissue expression profiles from the GTEx version 8 (GTEx Consortium 2020; Urbet et al. 2019). We chose three tissues that were relevant to SARS-CoV-2 infection, including lung, whole blood, and spleen.

Then, we ran S-PrediXcan scripts (downloaded from <https://github.com/hakyimlab/MetaXcan>, accessed on 10/10/2020) with each of the three tissue-specific models in two severe COVID-19 GWAS datasets, respectively. The threshold used in TWAS significance was adjusted by Bonferroni multiple-test correction with the ~10,000 genes. We defined the strict significance as $p < 5 \times 10^{-6}$ ($|z| > 4.56$) and suggestive significance as $p < 5 \times 10^{-5}$ ($|z| > 4.06$).

Colocalization analysis

Colocalization was performed to validate significant TWAS associations using two recent and cutting-edge statistical approaches: eCAVIAR (Hormozdiari et al. 2016) and fastENLOC (Wen et al. 2017), which aim to identify a single genetic variant that has shared causality between expression and GWAS trait. Both eCAVIAR and fastENLOC could assess the colocalization posterior probability (CLPP) for two traits at a locus, while eCAVIAR allows for multiple causal variants and fastENLOC features accountability for allelic heterogeneity in expression traits and high sensitivity

of the methodology. We ran eCAVIAR between significant TWAS genes and GWAS trait with a maximum of five causal variants per locus and defined a locus as 50 SNPs up- and down-stream of the tested causal variant, following the recommendation in the original paper. The eCAVIAR was downloaded from <https://github.com/fhormoz/caviar/> (accessed on 10/25/2020). The biallelic variants from the 1000 Genomes Project phase III in European ancestry were used as an LD reference (Genomes Project et al. 2015). We defined $CLPP > 0.5$ as having strong colocalization evidence.

To run fastENLOC, we first prepared probabilistic eQTL annotations to generate the cis-eQTL's posterior inclusion probability (PIP). Specifically, we applied the tissue-specific data from GTEx and T follicular cell-specific data from the DICE database (Schmiedel et al. 2018) using the integrative genetic association analysis with the deterministic approximation of posteriors (DAP-G) package (Lee et al. 2018). Then, GWAS summary statistics were split into approximately LD-independent regions defined by reference panel from European ancestry, and z-scores were converted to PIP. We downloaded the fastENLOC from <https://github.com/xqwen/fastenloc> (accessed on 10/25/2020) and followed the guideline to yield regional colocalization probability (RCP) for each independent GWAS locus using each tissue- or cell type-specific eQTL annotation. We defined $RCP > 0.5$ as having strong colocalization evidence.

Functional genomics annotations

To better understand the potential function of the variants identified by GWAS analyses and how they mediate the regulatory effect, we annotated significant SNPs using publicly available data. We obtained the tissue and cellular level eQTL data from the following resources: 1) the eQTL-Gen consortium (Võsa et al. 2018b) eQTLs generated from 30,912 whole blood samples; 2) Biobank-based Integrative Omics Studies (BIOS) eQTLs generated from 2116 healthy adults (Zhernakova et al. 2017); 3) The GTEx v8 (GTEx Consortium 2020) eQTLs of the lung, whole blood, and spleen tissues; 4) DICE database (Schmiedel et al. 2018) with cellular eQTLs of nine available T cell subpopulations. To identify the genomic annotation of the significant SNPs, we downloaded the multivariate hidden Markov model (ChromHMM) (Ernst and Kellis 2012) processed chromatin-state data of 17 lung and T cell lines from the Roadmap Epigenomics project (Roadmap Epigenomics et al. 2015). To explore the potential chromatin looping of GWAS locus, we used publicly available chromatin interaction (Hi-C) data (Dixon et al. 2012) at a resolution of 40 Kb on IMR-90, a normal lung fibroblast cell line. The Hi-C data have been used to identify specific baits and targets from distant chromatin regions that frequently interact with each other. Variants within the regulatory regions can be connected to

the potential gene targets and thus mediate the gene expression. Statistical tests of bait-target pairs were conducted to define significant bait interaction regions and their targets. The eQTL associations and chromatin-state information and Hi-C interactions were processed and plotted using the R Bioconductor package *gviz* in R version 4.0.3 (Hahne and Ivanek 2016).

Resident memory CD8⁺ T cells selection and differentially expressed gene analysis

We use the recently published scRNA-seq dataset of bronchoalveolar lavage fluids (BALF) samples from eight patients (three moderate and five severe) with COVID-19 (Liao et al. 2020; Liu et al. 2020b). We adapted the original annotation (Liao et al. 2020) and followed their method to calculate the resident memory CD8⁺ T (T_{RM}) cells signature score using 31 markers (14 positive markers and 17 negative markers) for all annotated 2677 CD8⁺ T cells in moderate and severe samples (Kumar et al. 2017; Pont et al. 2019). We defined the CD8⁺ T cells with top 50% scores as T_{RM} -like cells. We normalized these 1339 cells and conducted a principal component analysis (PCA) to filter out the outliers assuming that samples in the first two principal components follow a multivariate normal distribution, leaving 1282 cells within 95% confidence interval (Supplementary Figure S1). We specifically checked the cell surface marker of CD8⁺ T cells. We only kept cells with either CD8A or CD8B expression. To distinguish from the CD4⁺ T_{RM} , we excluded the cells (9 cells) with CD4 expression. Overall, we kept 1,101 CD8⁺ T_{RM} cells (765 from moderate patients and 336 from severe patients). Lastly, we conducted a non-parametric Wilcoxon rank-sum test by the function of “FindAllMarkers” from R package Seurat (Stuart et al. 2019) (version 3.1.5 in R version 3.5.2) to perform the differentially expressed genes (DEG) analysis between moderate and severe patients.

Cellular trajectory and transcriptional program analysis in T_{RM} cells

We used the R package Slingshot (Street et al. 2018) to infer cell transition and pseudotime from the scRNA-seq data. Specifically, we first used the expression data to generate the minimum spanning tree of cells in a reduced-dimensionality space [t-Distributed Stochastic Neighbor Embedding (tSNE) project from top 30 principal components of top 3000 variable genes] assuming there are two major clusters (moderate and severe T_{RM} cells). We then applied the principal curve algorithm (Hastie and Stuetzle 1989) to infer an one-dimensional variable (pseudotime) representing each cellular trajectory along with the transcriptional progression. We used our in-house machine learning tool, DrivAER (Driving transcriptional programs based on AutoEncoder derived

relevance scores) (Simon et al. 2020), to identify potential transcriptional programs [e.g., gene sets of pathways or transcription factors (TFs)] that potentially regulate the inferred cellular trajectory between the moderate and severe patients. To avoid the potential noise from the low expression genes, we excluded those genes expressed in < 10% cells. DrivAER took gene expression and pseudotime inferred from previous cellular trajectory results (Slingshot) and calculated each gene's relevance score by performing cellular manifold using Deep Count AutoEncoder (Eraslan et al. 2019) and a random forest model with out-of-bag score calculation as the relevance score. The transcriptional program annotations were from the hallmark pathway gene sets from MSigDB (Liberzon et al. 2015) and transcription factor (TF) target gene sets from TRRUST (Han et al. 2018). To calculate the relevance score, we used the “calc_relevance” function with the following parameters: min_targets = 10, ae_type = “nb-conddisp”, epoch = 100, early_stop = 3, and hidden_size = “(8,2,8)”. The relevance score (R^2 coefficient of determination) indicates the proportion of variance in the pseudotime explained by target genes of transcription factors or genes in the hallmark pathways.

DNA motif recognition analysis of genome-wide significant SNPs

We used the function “variation-scan” of the online tool RSAT (<http://rsat.sb-roscoff.fr/index.php>, accessed on 01/15/2020) (Nguyen et al. 2018) to predict the binding effect of all the significant SNPs in the *3p21.31* locus. We defined the TF with Bonferroni-corrected $p < 0.05$ as the significant TF. Later, we compared them with the TF with high relevance score from the DrivAER analysis above. The position weight matrices (PWMs) for all the TFs were downloaded from cis-BP Database (<http://cisbp.cabr.utoronto.ca/>) version 2019-06_v2.00) (Weirauch et al. 2014) and sequence logos representing motif-binding sites were generated using R package seqLogo version 1.54.3 in R version 3.5.2.

Results

TWAS analysis identified and replicated two chemokine receptor genes

We utilized the latest S-PrediXcan MASHR models trained with GTEx v8 data for TWAS analyses in lung and whole blood on two GWAS datasets of susceptibility to severe COVID-19 (Urbut et al. 2019). In the HGI cohort, we found that a decreased expression of *CXCR6*, which encodes C–X–C chemokine receptor type 6, in the lung was associated with an increased risk for the development of severe COVID-19 symptoms ($p = 1.57 \times 10^{-17}$, $z = -8.53$),

and this result was then replicated in the SCGG cohort ($p = 2.84 \times 10^{-5}$, $z = -4.19$, suggestive significant) (Fig. 2 and Table 1). Likewise, an increased expression of *CCR9*, which encodes C–C chemokine receptor type 9, in whole blood was associated with an increased risk for the development of severe COVID-19 complications in GWAS_{HGI} cohort ($p = 7.90 \times 10^{-11}$, $z = 6.50$) and this result was replicated in the other GWAS_{SCGG} cohort, ($p = 3.78 \times 10^{-10}$, $z = 6.26$) (Fig. 2 and Table 1). Whole blood and lung transcriptome models also identified two additional significant TWAS genes that are specific to one of the two cohorts. Increased expression of *ABO* gene in the lung was associated with risk for the development of severe COVID-19 symptoms in GWAS_{SCGG} data set ($p = 5.98 \times 10^{-7}$, $z = 4.99$). Similarly, increased expression of *GAS7* gene (Growth Arrest-Specific 7) in whole blood was associated with an increased risk for the development of COVID-19 symptoms in the GWAS_{HGI} data set ($p = 8.46 \times 10^{-7}$, $z = 4.92$). Lastly in spleen, we identified lower expression of C–C Motif Chemokine Receptor 3 (*CCR3*) was associated with the severe COVID-19 symptoms (Supplementary Table S1) in GWAS_{HGI} as previous TWAS mentioned (Pairo-Castineira et al. 2020). However, we failed to replicate the *CCR3* signal in GWAS_{SCGG}. Overall, *CXCR6* and *CCR9* were found and replicated to be associated with COVID-19.

Colocalization analysis validated the mediation effect of CXCR6 between GWAS 3p21.31 locus and severe COVID-19

The TWAS findings might be driven by pleiotropy or linkage effect by the LD structure in the GWAS loci instead of the true mediation effect (Wainberg et al. 2019) (Fig. 3A). To rule out the linkage effect and find further evidence of true colocalization of causal signals in the variants that were significant in both GWAS and eQTL analyses, we performed colocalization analyses on all the genes in the *3p21.31* region including *CCR1*, *CCR2*, *CCR3*, *CCRL2*, *CCR5*, *XCRI*, *CCR9*, *CXCR6*, *FYCO1*, *LZTFL1* by eCAVIAR and fastENLOC using several tissue- and cell-line- specific eQTL datasets. eCAVIAR with the eQTL data in lung tissue revealed that the severe COVID-19 association could be mediated by the variants that were associated with the expression of *CXCR6* (CLPP = 0.79) (Table 1). And the colocalized SNP rs34068335 (GWAS_{HGI} $p = 5.02 \times 10^{-22}$) is also related to the increased monocyte percentage of white cells in a blood-trait GWAS study using PhenoScanner (Kamat et al. 2019). *CCR3* in spleen shows colocalization with CLPP = 0.580 and *CCR1* in T follicular helper cell shows marginal colocalization with CLPP = 0.10 (Supplementary Table S2). The fastENLOC analysis showed a high RCP between the expression of *CXCR6*

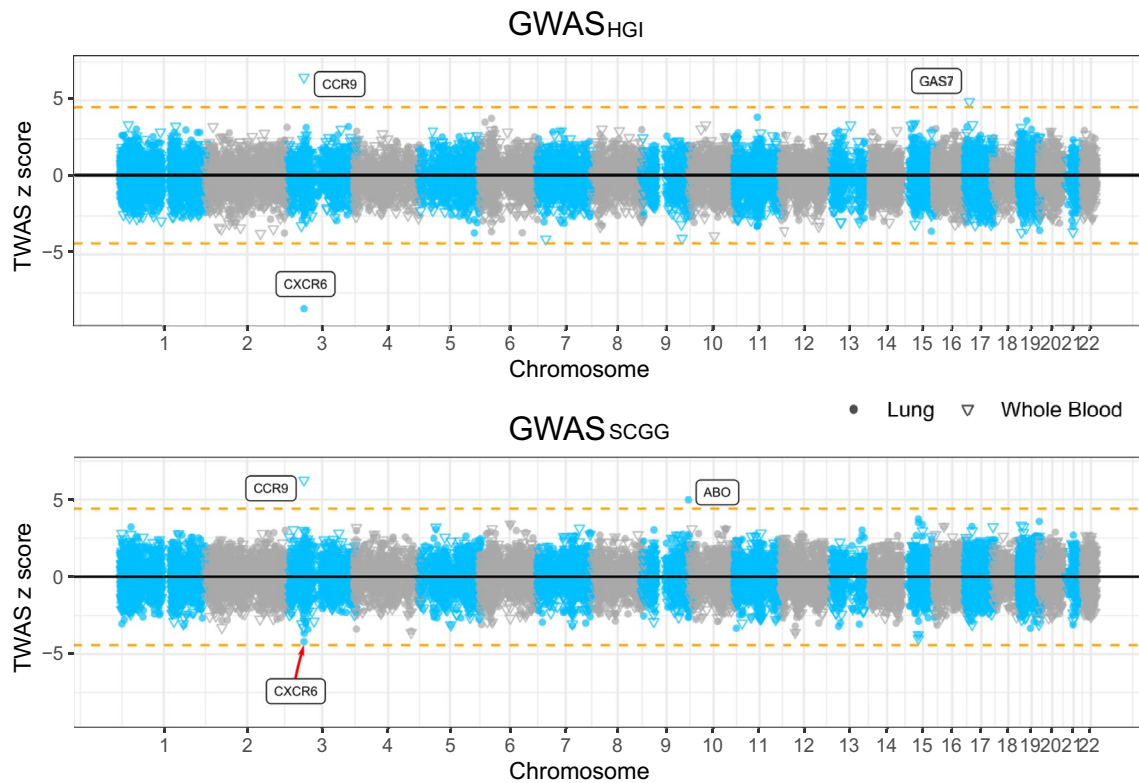


Fig. 2 Manhattan plots illustrating the z-scores of transcriptome-wide association study (TWAS) genes. TWAS z-scores for two genome-wide association study (GWAS) datasets of susceptibility to severe COVID-19 using lung and whole blood tissue models. The upper panel shows the results from GWAS_{HGI} and the lower panel from

GWAS_{SCGG} (see “Materials and methods”). The round and triangle points denote lung and whole blood tissues, respectively, in the TWAS analysis. Dashed horizontal lines denote the Bonferroni-corrected significance threshold ($|z|=4.56$, $p < 5 \times 10^{-6}$). Significant genes were highlighted with their gene symbol

Table 1 Summary of TWAS and colocalization analyses in tissues and cell line

Gene symbol	Tissue	Discovery: GWAS _{HGI}				Validation: GWAS _{SCGG}			
		TWAS z	TWAS p	PP	Colocalized SNP p	TWAS z	TWAS p	PP	Colocalized SNP p
CXCR6	Lung	-8.53	1.57×10^{-17}	0.79*	rs34068335 5.02×10^{-22}	-4.19	2.84×10^{-5}	ns	-
	T follicular helper cells	-	-	0.99**	rs35081325 3.82×10^{-39}	-	-	0.99**	rs35081325 2.49×10^{-10}
CCR9	Whole blood	6.50	7.90×10^{-11}	ns	-	6.26	3.78×10^{-10}	ns	-

GWAS_{HGI} denotes the GWAS dataset from the host genetics initiative

GWAS_{SCGG} represents the GWAS dataset from the severe COVID-19 GWAS Group

PP posterior probability, z z score, p p value, ns no significant colocalization from either eCAVIAR or fastENLOC, “-” no available data

*Statistically significant by the colocalization posterior probability (CLPP) from eCAVIAR

**Statistically significant by the regional colocalization probability (RCP) from fastENLOC

in T follicular helper cells and GWAS signal in both the GWAS_{HGI} cohort (RCP = 0.99) and the GWAS_{SCGG} cohort (RCP = 0.99) (Table 1). *CCRL2* shows colocalization with RCP = 0.5 in T follicular helper cells in both cohorts. *CCR5* shows colocalization with RCP = 0.88 in monocytes in SCGG cohort and RC $p = 0.49$ in monocytes

in HGI cohort (Supplementary Table S3). However, colocalization analysis of *CCR9* or other genes did not suggest strong colocalization evidence (CLPP < 0.1 and RCP < 0.1).

Multi-level functional annotations linked 3p21.31 locus with CXCR6 and CCR9 functions

To explore the potential functions linked with the GWAS risk variants, we examined the functional genomic annotations in this locus. Specifically, we found a consistent decreasing effect of *CXCR6* expression in T cells and whole blood from the two large-scaled eQTL datasets (Fig. 3B). The Hi-C cell line data from lung fibroblast (Dixon et al. 2012) also showed a significant interaction between the *3p21.31* locus and both promoter regions of *CXCR6* and *CCR9* (Fig. 3C). Furthermore, multiple SNPs in the *3p21.31* locus reside in the annotated regulatory elements across blood, T cell, and lung cell lines (Fig. 3D, see “Materials and methods”). Overall, these results from the multiple lines of evidence all supported the potential regulatory effects of the *3p21.31* locus on *CXCR6* and *CCR9* expression.

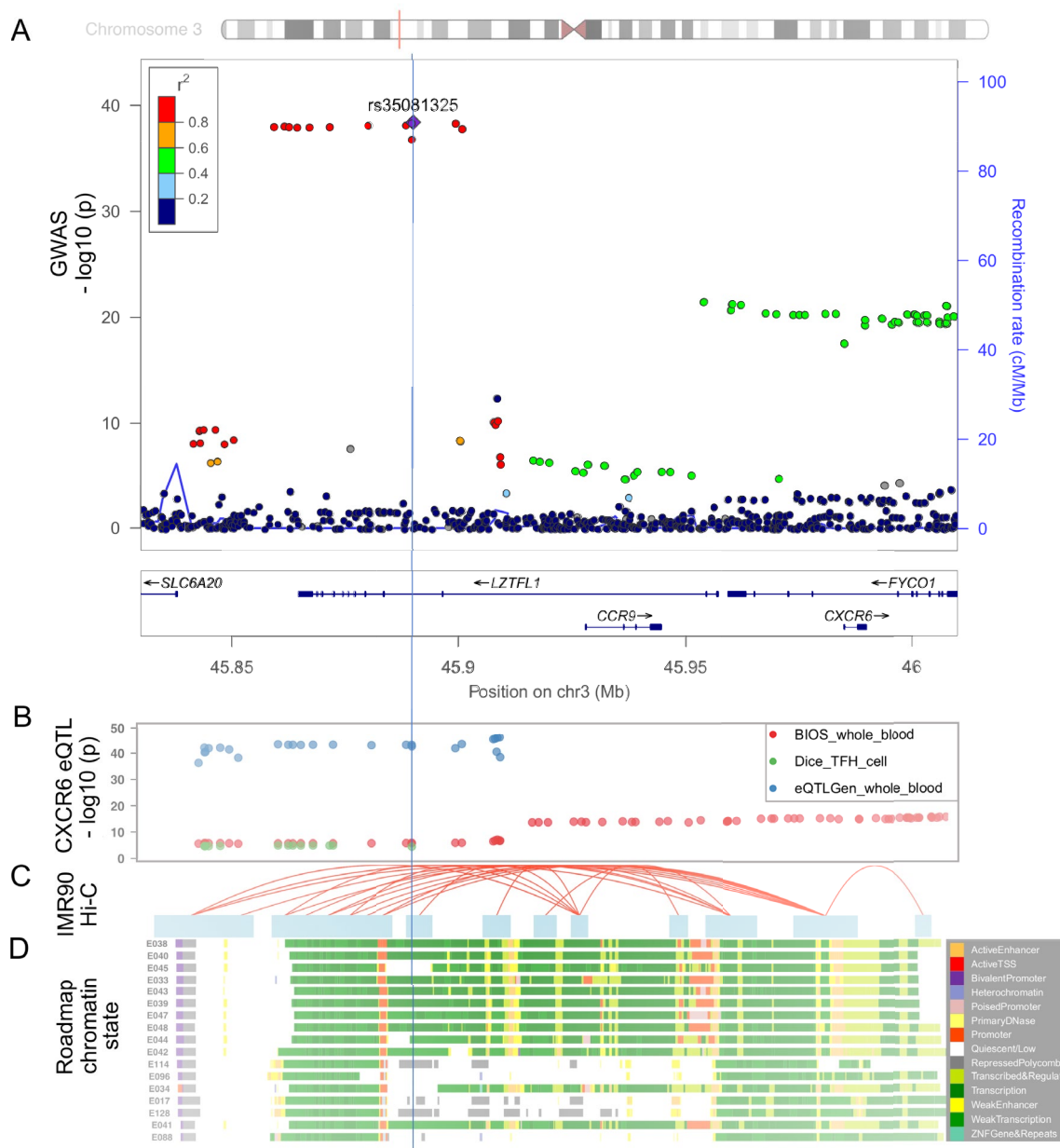
CXCR6 differentially expressed in T_{RM} cells of severe and moderate patients

According to our tissue cell-type-specific expression database (CSEA-DB), *CXCR6* is mainly expressed in immune cells in human lung tissue (e.g., T cell and NK cell) (Dai et al. 2021). In Liao et al.’s work, the authors reported that *CXCR6* had lower expression in severe patients than moderate patients, indicating a potential protective effect in T cells of human respiratory systems (Liao et al. 2020). However, T cells have various resident and circulating subtypes with diverse functions (Hombrink et al. 2016). To understand which subpopulation(s) of T cells might be associated with the severity of COVID-19, we used the BLAF scRNA-seq data of severe and moderate patients. The data included 2677 CD8⁺ T-cells (1,329 from five severe patients and 1349 from three moderate patients). We further used a set of 31 T_{RM} cell marker genes to distinguish the T_{RM} cells and conventional CD8⁺ T cells (see “Materials and methods”). As shown in Fig. 4A, B, the T_{RM} cells and conventional T cells could be distinguished in both moderate and severe patients with the classic T_{RM} cells markers (*CXCR6* (Kumar et al. 2017), CD69 (Martin and Badovinac 2018), *ITGAE* (the gene encoding CD103) (Martin and Badovinac 2018; Wauters et al. 2021), *ZNF683* (Wauters et al. 2021), and *XCL1* (Hombrink et al. 2016)) and three negative-control markers (*SELL* (the gene encoding CD62L) (Martin and Badovinac 2018), *KLF2*, and *S1PR1* (Skon et al. 2013)) from a previous study (Kumar et al. 2017). After removing the outlier cells and CD4 T cells (see “Materials and methods”), we obtained 1101 lung-resident T_{RM} cells, including 765 cells from moderate patients and only 336 cells from severe patients. This represented a 2.24-fold decrease for the expected number of T_{RM} cells in severe patients in CD8⁺ T cells (Table 2). We used the non-parametric Wilcoxon

rank-sum test to identify the DEGs in the T_{RM} cells between severe and moderate patients and found *CXCR6* had significantly lower expression in the severe patients than the moderate patients ($p < 2.2 \times 10^{-16}$, fold change = 1.50, Fig. 4C). We further checked the *CXCR6* expression in T_{RM} cells by sample to exclude the possibility that *CXCR6* DEG analyses were disrupted by excessively high or low expression of certain samples (Supplementary Table S4).

Inferring the transcriptional programs that drive the cell status transition

Previous COVID-19 studies demonstrated that performing pseudotime analysis could infer a cellular trajectory that indicates the disease progression from scRNA-seq data (Bernardes et al. 2020; Jeong et al. 2021). Therefore, to understand the transition between moderate and severe T_{RM} cells, we constructed the cellular trajectory/pseudotime along with T_{RM} cells using Slingshot (Fig. 4D) (Street et al. 2018). Next, we applied our DrivaER approach (Driving transcriptional programs based on AutoEncoder derived Relevance scores) (Simon et al. 2020) to identify the potential transcriptional programs that were most likely involved in the cellular trajectory/pseudotime. Figure 4D shows a scaled heatmap to demonstrate the relative expression of naïve and effector markers of T cells in the order of pseudotime generated by Slingshot (Han et al. 2018; Street et al. 2018). We identified that the severe T_{RM} cells were mainly gathered in the later stage of the pseudotime. The naïve markers (*IL7R*, *BCL2*) were higher expressed in moderate patients than in severe patients (except *SELL*). On the contrary, some effector markers (*GZMB*, *HAVCR2*, *LAG3*, *IFNG*) were lower expressed in moderate patients than in severe patients. Other effector markers (*IRF4*, *PRF1*) had higher expression in the middle of the transition than their expression at the start and end sides. Furthermore, we also found that there was a negative correlation between *CXCR6* expression and the pseudotime (Pearson correlation coefficient $\rho = -0.36$, Supplementary Figure S2). These results indicated the T_{RM} cells in severe patients still in pro-inflammatory status although the T_{RM} cells status was more heterogeneous in severe patients than in moderate patients (Fig. 4a, b, e). As shown in Fig. 4f, g, the top five molecular signatures (relevance score > 0.25) identified by DrivaER included T cell pro-inflammatory actions (interferon gamma response, allograft rejection (Benichou et al. 2017), and complement system) as well as apoptosis (Cizmecioglu et al. 2021) and hypoxia (Jahani et al. 2020) pathways. Among the top TFs (relevance score > 0.2) that drove this trajectory, the DNA-binding RELA-NFKB1 complex is involved in several biological processes, such as inflammation, immunity, and cell growth initiated by external stimuli. *MYC* gene is related to the cell cycle and apoptosis pathway and upstream transcription



factor 1 (*USF1*) is related to metabolic alteration associated inflammation process (Ruuth et al. 2018). In summary, the TF results are well consistent with our previous hallmark pathway findings (Supplementary Table S5 and S6).

Several genome-wide significant SNPs might change the TF binding-site affinity

To understand the potential TF binding affinity changes of genome-wide significant SNPs, we conducted the DNA motif recognition analysis of the five TFs related to the transcriptional program between moderate and severe T_{RM} cells (relevance score > 0.2, Supplementary Table S6).

We identified SNP rs10490770 [T/C, minor allele frequency (MAF)=0.097, $GWAS_{HGI}=9.53 \times 10^{-39}$] and SNP rs67959919 (G/A, MAF=0.097, $GWAS_{HGI}=8.83 \times 10^{-39}$) that were predicted to alter the binding affinity of TFs RELA and SP1, respectively (Supplementary Fig. S3a and S3b). Moreover, these two SNPs were in the high LD region ($r^2 > 0.8$) with several significant lead eQTLs (SNP rs35896106 and rs17713054) of *CXCR6* in whole blood ($p=5.03 \times 10^{-37}$) and T follicular helper cell ($p=1.30 \times 10^{-5}$) (Fig. 3B). In summary, the genome-wide significant SNPs were predicted to change the binding affinity of those TFs highly related to T_{RM} cells status

Fig. 3 Functional genomic annotation in the *3p21.31* locus with signals from GWAS_{HGI}. **a** LocusZoom view of the association signals of SNPs in the *3p21.31* locus of GWAS_{HGI}. The x-axis is the chromosome position in million base pairs (Mb) on GRCh37 reference genome and y-axis represents the $-\log_{10}(p \text{ value})$ from GWAS_{HGI} dataset. The color indicates the strength of linkage disequilibrium from the lead SNP rs35081325. The genes within the region are annotated in the lower panel. A vertical blue line labels the position of the lead SNP rs35081325 to denote the relationship of GWAS variants to other datasets: expression quantitative trait (eQTL) (**b**), chromatin interaction (**c**), and imputed Roadmap functional elements (**d**). **b** The significant eQTLs associated with *CXCR6* expression in this region. The *cis*-eQTL datasets include two whole blood datasets [Biobank-based Integrative Omics Studies (BIOS) QTL and eQTLGen] and one T follicular helper cell dataset (DICE). The y-axis represents the $-\log_{10}(p \text{ value})$ from the eQTL studies. **c** The significant Hi-C interactions in normal lung fibroblast cell line (IMR-90). Blue blocks denote the target and bait regions, and red arcs indicate the interactions between functional elements. **d** The region annotated with the chromatin-state segmentation track (ChromHMM) from the Roadmap Epigenomics data for T cell and lung tissue. The Roadmap Epigenomics cell line IDs are shown on the left side: E017 (IMR-90 fetal lung fibroblasts Cell Line), E033 (Primary T Cells from cord blood), E034 (Primary T Cells from blood), E038 (Primary T help naïve cells from peripheral blood), E039 (Primary T helper naïve cells from peripheral blood), E040 (Primary T helper memory cells from peripheral blood), E041 (Primary T helper cells PMA-Ionomycin stimulated), E042 (Primary T helper 17 cells PMA-Ionomycin stimulated), E043 (Primary T helper cells from peripheral blood), E044 (Primary T regulatory cells from peripheral blood), E045 (Primary T cells effector/memory enriched from peripheral blood), E047 (Primary T CD8 naïve cells from peripheral blood), E048 (Primary T CD8 memory cells from peripheral blood), E088 (Fetal lung), E096 (Lung), E114 (A549 EtOH 0.02pct Lung Carcinoma Cell Line), and E128 (NHLE Human Lung Fibroblast Primary Cells). The colors denote chromatin states imputed by ChromHMM, with the color key in the gray box (“Materials and methods”)

transition, (Supplementary Table S6), suggesting their potential regulation of *CXCR6* expression.

Discussion

In this work, we developed a multi-level, integrative genetic and functional analysis framework to explore the host genetic factors on the expression change of GWAS-implicated genes for COVID-19 severity. Specifically, we conducted TWAS analysis for two independent COVID-19 GWAS datasets. We identified and replicated three chemokine receptor genes, *CXCR6* and *CCR9*, with a protective effect in the lung and a risk effect in whole blood, respectively. *CXCR6* is expressed in T lymphocytes and essential genes in CD8⁺ T_{RM} cells, mediating the homing of T_{RM} cells to the lung along with its ligand *CXCL16* (Payne et al. 2021; Wein et al. 2019). *CCR9* was reported to regulate chemotaxis in response to thymus-expressed chemokines in T cells (Lee et al. 2012). The colocalization analysis identified that both GWAS and eQTLs of *CXCR6* had high colocalization probabilities in the lung, whole blood, and T follicular helper cells, which

confirms the genetic regulation roles at this locus. At the single-cell level, our DEG analysis identified *CXCR6* gene had lower expression in the COVID-19 severe patients than the moderate patients in both T cells and T_{RM} cells, supporting its protective effect identified in TWAS analysis in lung and whole blood. The expected proportion of T_{RM} cells in CD8⁺ T cells also decreased by 2.24-fold (Table 2). We identified the major transition force from moderate T_{RM} cells to severe T_{RM} cells are pro-inflammatory pathways and TFs.

From the TWAS and colocalization analysis in lung and immune cells, we successfully replicated that *CXCR6* was centered in the GWAS signal in the *3p21.31* locus. Previous studies have reported that *CXCR6*^{-/-} significantly decreases airway lung T_{RM} cells due to altered trafficking of *CXCR6*^{-/-} cells within the lung of the mice (Wein et al. 2019), which could explain a much less proportion of T_{RM} cells in severe patients than moderate patients. The lung T_{RM} cells provide the first line of defense against infection and coordinate the subsequent adaptive response (Ardain et al. 2020). The previous study has reported that T_{RM} cells constitutively expressed surface receptors (PD-1 and CTLA-4) that are associated with inhibition of T cell function, which might prevent excessive activation or inflammation in the tissue niche (Szabo et al. 2019).

We are aware of several recent works have been decoding the genetic factors in the *3p21.31* locus. In Pairo-Castineira et al. (Pairo-Castineira et al. 2020), their TWAS analyses highlighted *CCR2*, *CCR3*, and *CXCR6* genes in using the Meta-TWAS method MultiXcan for all GTEx v8 MASHR prediction model. We successfully replicated their *CXCR6* in lung for both HGI and SCGG cohorts and *CCR3* in spleen for HGI cohort only. The function of *CCR3* at the single-cell level remained investigated since there is virtually absent *CCR3* expression in our BALF dataset. We failed to detect any significant TWAS signals for *CCR2* in any tissues or cohorts. Differences in associations identified in our TWAS compared to the previous study could be due to different TWAS prediction models used and different GWAS sample compositions. We also identified several novel candidates *CCR9*, *CCR1*, *CCR5*, and *CCRL2* that held a line of evidence from either TWAS or colocalization analysis (Supplementary Table S1, S2, S3).

We further explored the known function of *CXCR6* associated with COVID-19 and other infectious diseases in case-control studies, phenome-wide association study (PheWAS) and GWAS. The association of the single-cell level phenotype (lower expression of *CXCR6* and decreased proportion of CD8⁺ CXCR6⁺ T cells) and the severe COVID-19 has been observed in another work in circulating T cells with flow cytometry experiments (Payne et al. 2021), suggesting the significant role of the chemokine receptor *CXCR6* and its ligand *CXCL16* in the immunopathogenesis of severe COVID-19. Another recent CRISPR screen

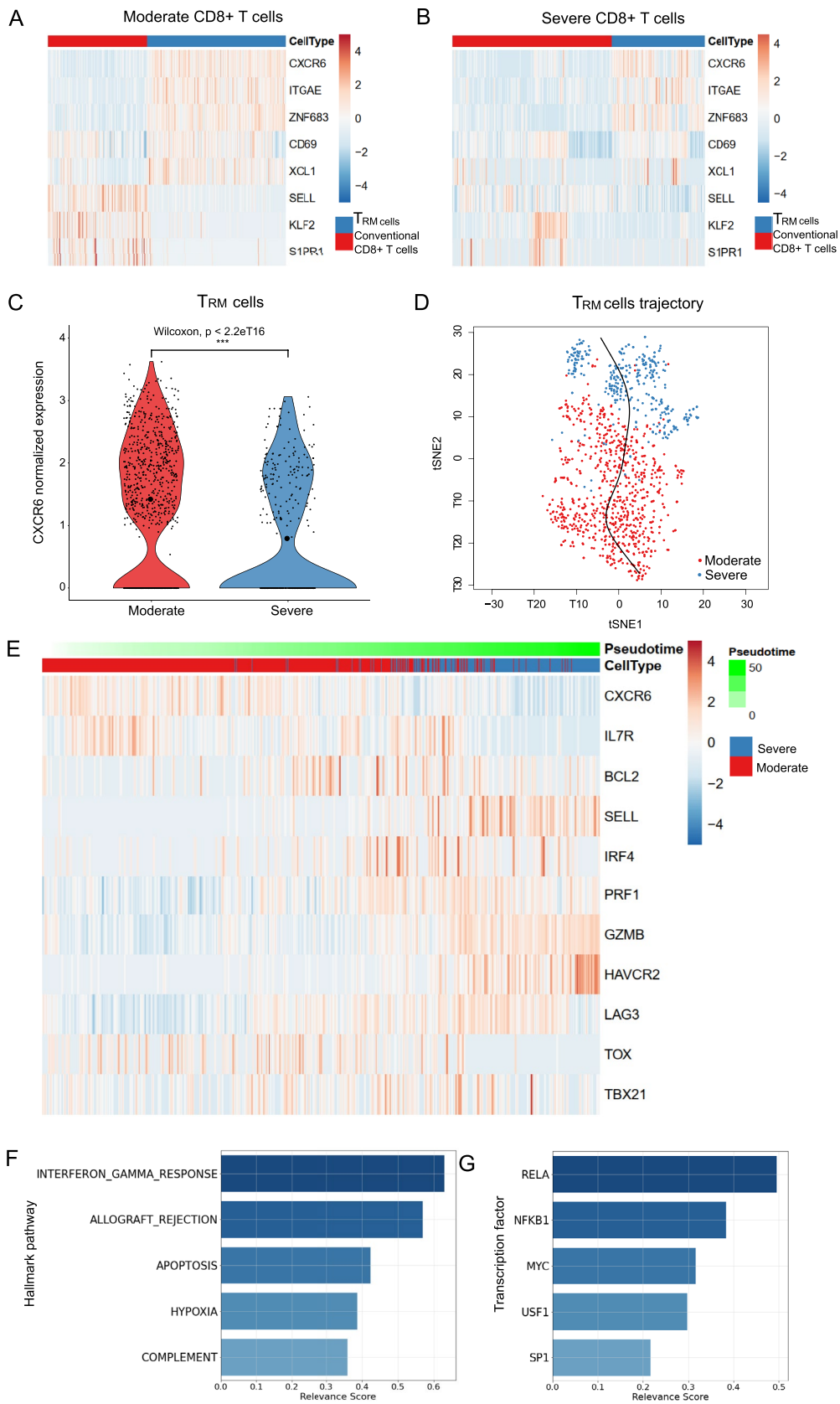


Fig. 4 Single-cell transcriptome analysis of the severe and moderate COVID-19 patients. **a** Relative expression of the lung-resident memory CD8⁺ T (T_{RM}) signature genes in T_{RM} cells and conventional CD8⁺ T cells in moderate patients. **b** Relative expression of the T_{RM} featured genes in T_{RM} cells and conventional CD8⁺ T cells in severe patients. **c** *CXCR6* expression in the T_{RM} cells of moderate and severe patients. We split the T_{RM} cells from the annotation of the original paper with 31 marker genes (“Materials and methods”). We conducted a two-sided non-parameter Wilcoxon rank-sum test to test whether *CXCR6* was differentially expressed in moderate (red) and severe (blue) groups of T_{RM} cells. “***” indicates it is genome-wide significant after multiple-test correction of all expressed genes. The small points denote the normalized expression in each cell. Mean normalized expression of *CXCR6* in each group is highlighted with the largest circle in black. **d** Pseudotime inference for the moderate and severe T_{RM} cells. The red and blue points on t-Distributed Stochastic Neighbor Embedding (tSNE) projection denote the T_{RM} cells from moderate and severe patients, respectively. The *x*-axis and *y*-axis are the first and second dimensions of the tSNE, respectively. **e** Relative expression of the *CXCR6* and naïve and effector T cell markers along the pseudotime proportional to the green color. The gene expressions are scaled by cells. Cells from moderate and severe groups are annotated in blue and red. **f** Relevance score for hallmark pathways from the molecular signatures database (MSigDB) along the pseudotime. The relevance score (R^2 coefficient of determination) indicates the proportion of variance in the pseudotime explained by the genes in the hallmark pathways. **g** Relevance score for transcription factors and their target genes along the pseudotime. The relevance score denotes the proportion of variance in the pseudotime explained by the target genes regulated by the transcription factor

study also suggested *CXCR6* is one of the top-ranked genes, whose loss reduces SARS-CoV-2 infection in lung epithelial-like cells (Kasela et al. 2021). Moreover, in Ashhurst et al.’s work (Ashhurst et al. 2019), they identified *CXCR6* was redundant for T-lymphocyte recruitment to airway and lung. The *CXCR6*-deficiency mice could increase the host control of tuberculosis and influenza through other alternative receptors (*CXCR3*, *CCR5*, and *CCR4*) to activate the inflammatory response, suggesting that other activated pathways might compensate *CXCR6*-deficiency in COVID-19 severe patients. We applied PhenoScanner to conduct the PheWAS analysis (Kamat et al. 2019) and identified several phenotypes “monocyte percentage of white cells” and “monocyte count” related to SNPs in *CXCR6* gene regions, indicating potential association with *CXCR6*. However, those lead SNPs in *CXCR6* might be *cis*-regulators of nearby genes due to the high linkage disequilibrium (LD) in this *3p21.31* locus. In GWAS, we failed to find any GWAS risk variants annotated to *CXCR6* from the GWAS catalog using keywords “pulmonary tuberculosis”, “influenza”, “malaria” or “HIV” (Buniello et al. 2019).

To understand the status of T_{RM} cells, we further used nine classic naïve markers (e.g., *BCL2*, *SELL*, *TCF7*, and *IL7R*) and ten classic effector markers (e.g., *GZMB*, *PRF1*, *IFNG*, *LAG3*, and *PDCDI*) to quantify the naïve and effector status of the T_{RM} cells (Supplementary Figure S4). T_{RM} cells in severe patients had a much higher median of effector

Table 2 Counts and ratio of T_{RM} cells in moderate and severe patients relevant to CD8⁺ T cells

Patient group (sample size)	#CD8 ⁺ T cells	# T_{RM} cells	Expected T_{RM} cell proportion ratio (moderate/severe)
Moderate (3)	1348	765	2.24
Severe (5)	1329	336	

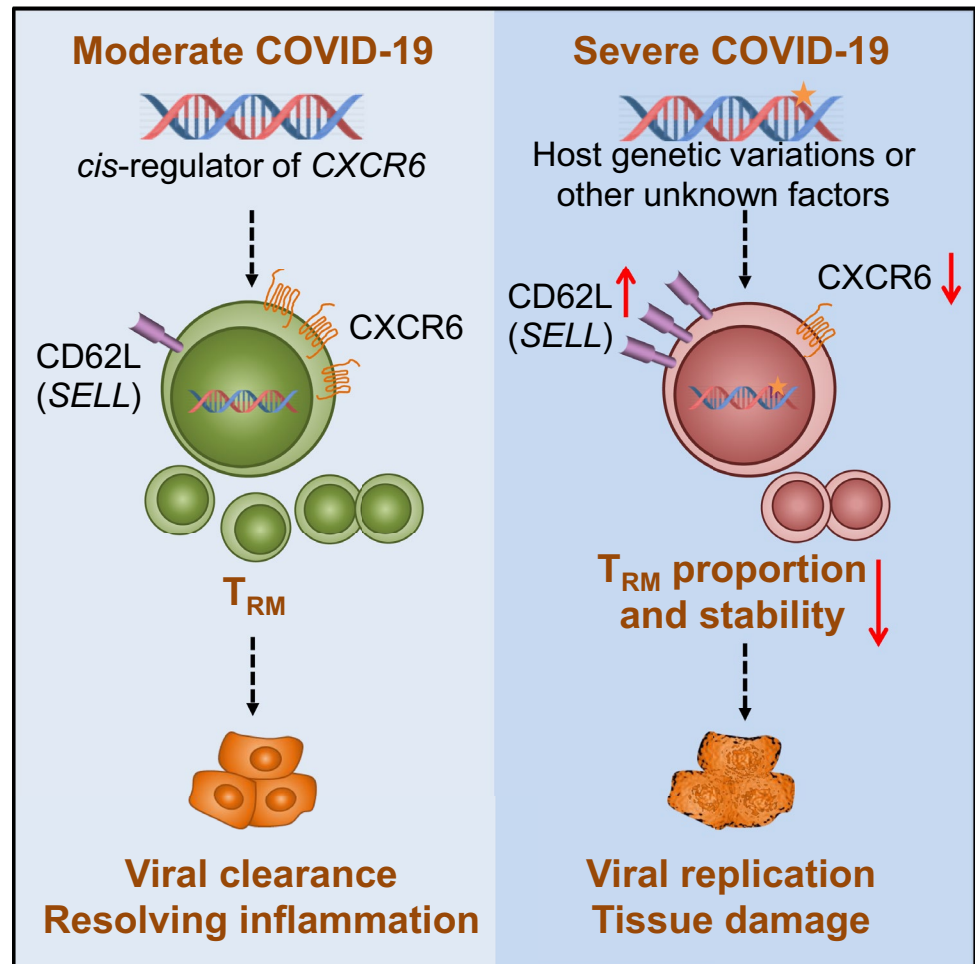
T_{RM} cells the resident memory CD8⁺ T cells as defined in “Materials and methods”

#The counted number

marker score (0.48 in severe group and 0.19 in moderate group) than T_{RM} cells in moderate patients did, suggesting that the severe T_{RM} cells had much higher activities in inflammation as we discovered in Fig. 4F despite their proportion decrease. For the naïve score (Supplementary Figure S4), both moderate and severe T_{RM} cells had limited expressions (median: 0.032 in the severe group; median: 0.027 in the moderate group). Interestingly, if we removed the lymph node homing receptor *SELL* (Kumar et al. 2017) from the naïve markers list, we would find the median naïve markers in the severe group would drop to 0 (Supplementary Figure S4). This indicated that *SELL* expression contributed greatly to the naïve status of T_{RM} severe patients. Consistently in Fig. 4E, we could also observe that a large proportion of T_{RM} cells had higher *SELL* expression in severe patients than in moderate patients, suggesting the T_{RM} cells in severe patients might not be in a stable cell status due to the expression of lymph node homing signal (*SELL*). As illustrated in Fig. 5, we hypothesized that host genetic variants or other unknown factors would lead to lower expression of *CXCR6*, which might decrease the proportion of T_{RM} cells residing in the lung through the *CXCR6/CXCL16* axis (Payne et al. 2021; Wein et al. 2019), impairing the first-line defense. Moreover, the lower expression of *CXCR6* might be related to the “unstable” residency of T_{RM} cells in lung. Since T_{RM} cells play essential roles in orchestrating the immune system, the lack of which would lead to severe COVID-19 symptoms, such as acute respiratory distress syndrome, cytokine storm, and major multi-organ damage (Tay et al. 2020). Lastly, the *CXCR6*-deficiency might also lead to the compensatory inflammatory response (Ashhurst et al. 2019).

The GWAS_{HGI} dataset used in this study was HGI round 4 (alpha), which was the largest GWAS by the access date of October 20, 2020. However, it was not the currently largest GWAS meta-analysis for severe COVID-19 when we prepared the manuscript. This research field is evolving very fast, due to the urgent demand for public health. Currently, the largest GWAS HGI round 4 (freeze) contained more samples (4336 cases/353,891 controls), and it included two independent datasets we used in this study. Considering that the GWAS_{HGI} dataset included ~ 10% control samples from

Fig. 5 The proposed *CXCR6* regulation mechanism on COVID-19 severity. We proposed one pathogenesis mechanism using current knowledge to explain how the lower expression of *CXCR6* could be associated with the outcome of severe COVID-19 symptoms, which was supported by our findings of the genetic factors on decreasing the *CXCR6* expression and aligned with our observations from single-cell transcriptome analysis. The star on the DNA indicates the host genetic effects or other unknown factors



the Asian population, we checked the LocusZoom plot of the chr3: 45.80–46.40 million base pairs (Mb) region on GRCh37 reference genome. We found a consistent tendency in GWAS round 4 alpha and freeze version (Supplementary Figure S5). Another limitation is that the scRNA-seq data only had eight COVID-19 patient samples (five severe and three moderate samples), which might not provide enough statistical power at the sample level as it is commonly considered each scRNA-seq data acts as a population. Thus, our single-cell findings remain further investigated in larger sample-size studies. Finally, the TF binding-site affinity alterations were assessed based on computational prediction, therefore, the *in vivo* effects require experimental validation. We anticipate more and larger datasets will be released in the near future. We will apply our integrative analysis approach to such new data.

Our work systematically explored the genetic effect on gene expression in the chromosome *3p21.31* locus and pinpointed the multi-evidence validated gene *CXCR6* might be involved in the severity of COVID-19. Several genome-wide significant SNPs were within the LD block of *CXCR6* eQTLs in immune-related cells. In a scRNA-seq

COVID-19 BALF dataset, we characterized that *CXCR6* (T_{RM} cells marker gene) had a lower expression in severe patients than in moderate patients. Moreover, the T_{RM} cells in severe patients had a 2.24-fold proportion decrease in $CD8^+$ T cells and much higher pro-inflammatory activity than T_{RM} cells in moderate patients. Based on these observations, we proposed a potential mechanism on how the lower expression of *CXCR6* regulated by the endogenous factors or other unknown factors could progress to severe COVID-19 outcomes.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00439-021-02305-z>.

Acknowledgements We appreciate Drs. Teng Liu and Dawei Zou for the valuable comments. We thank all members of the Bioinformatics and Systems Medicine Laboratory for the discussion.

Authors' contributions YD, ZZ, JW and WC contributed to the conception and design of the study; YD, JW, and HHJ collected the data and performed the analysis; YD, JW, WC, PJ and ZZ interpreted the results; YD, JW, HHJ, ZZ, and PJ wrote the manuscript. All authors read and approved the final manuscript.

Funding Dr. Zhao was partially supported by National Institutes of Health grants (R01LM012806 and R01DE030122) and Chair Professorship for Precision Health funds. We thank the technical support from the Cancer Genomics Core funded by the Cancer Prevention and Research Institute of Texas (CPRIT RP180734). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data availability All datasets analyzed in this study were publicly available. Other data could be accessed from public resources described in Methods.

Code availability Available upon reasonable request.

Declarations

Conflict of interest The authors declare that they have no competing interests.

Animal research ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

References

- Ardain A, Marakalala MJ, Leslie A (2020) Tissue-resident innate immunity in the lung. *Immunology* 159:245–256. <https://doi.org/10.1111/imm.13143>
- Ashhurst AS, Florido M, Lin LCW, Quan D, Armitage E, Stifter SA, Stambas J, Britton WJ (2019) CXCR6-deficiency improves the control of pulmonary mycobacterium tuberculosis and influenza infection independent of T-lymphocyte recruitment to the lungs. *Front Immunol* 10:339. <https://doi.org/10.3389/fimmu.2019.00339>
- Barbeira AN, Dickinson SP, Bonazzola R, Zheng J, Wheeler HE, Torres JM, Torstenson ES, Shah KP, Garcia T, Edwards TL, Stahl EA, Huckins LM, Aguet F, Ardlie KG, Cummings BB, Gelfand ET, Getz G, Hadley K, Handsaker RE, Huang KH, Kashin S, Karczewski KJ, Lek M, Li X, MacArthur DG, Nedzel JL, Nguyen DT, Noble MS, Segrè AV, Trowbridge CA, Tukiainen T, Abell NS, Balliu B, Barshir R, Basha O, Battle A, Bogu GK, Brown A, Brown CD, Castel SE, Chen LS, Chiang C, Conrad DF, Damani FN, Davis JR, Delaneau O, Dermitzakis ET, Engelhardt BE, Eskin E, Ferreira PG, Frésard L, Gamazon ER, Garrido-Martín D, Gewirtz ADH, Gliner G, Gloudemans MJ, Guigo R, Hall IM, Han B, He Y, Hormozdiari F, Howald C, Jo B, Kang EY, Kim Y, Kim-Hellmuth S, Lappalainen T, Li G, Li X, Liu B, Mangul S, McCarthy MI, McDowell IC, Mohammadi P, Monlong J, Montgomery SB, Muñoz-Aguirre M, Ndungu AW, Nobel AB, Oliva M, Ongen H, Palowitch JJ, Panousis N, Papasaiakas P, Park YS, Parsana P, Payne AJ, Peterson CB, Quan J, Reverter F, Sabatti C, Saha A, Sammeth M, Scott AJ, Shabalín AA, Sodaei R, Stephens M, Stranger BE, Strober BJ, Sul JH et al (2018) Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat Commun* 9:1–20. <https://doi.org/10.1038/s41467-018-03621-1>
- Benichou G, Gonzalez B, Marino J, Ayasoufi K, Valujskikh A (2017) Role of memory T cells in allograft rejection and tolerance. *Front Immunol* 8:170. <https://doi.org/10.3389/fimmu.2017.00170>
- Bernardes JP, Mishra N, Tran F, Bahmer T, Best L, Blase JJ, Bordoni D, Franzenburg J, Geisen U, Josephs-Spaulling J, Kohler P, Kuntstner A, Rosati E, Aschenbrenner AC, Bacher P, Baran N, Boysen T, Brandt B, Bruse N, Dorr J, Dräger A, Elke G, Ellinghaus D, Fischer J, Forster M, Franke A, Franzenburg S, Frey N, Friedrichs A, Fuss J, Gluck A, Hamm J, Hinrichsen F, Hoepfner MP, Imm S, Junker R, Kaiser S, Kan YH, Knoll R, Lange C, Laue G, Lier C, Lindner M, Marinos G, Markewitz R, Nattermann J, Noth R, Pickkers P, Rabe KF, Renz A, Rocken C, Rupp J, Schaffarzyk A, Scheffold A, Schulte-Schrepping J, Schunk D, Skowasch D, Ulas T, Wandinger KP, Wittig M, Zimmermann J, Busch H, Hoyer BF, Kaleta C, Heyckendorf J, Kox M, Rybniker J, Schreiber S, Schultze JL, Rosenstiel P, Network HCALB, Deutsche C-OI (2020) Longitudinal multi-omics analyses identify responses of megakaryocytes, erythroid cells, and plasmablasts as hallmarks of severe COVID-19. *Immunity* 53:1296–1314 e9. <https://doi.org/10.1016/j.immuni.2020.11.017>
- Bhopal SS, Bhopal R (2020) Sex differential in COVID-19 mortality varies markedly by age, vol 396. *Lancet Publishing Group*, pp 532–533
- Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, McMahon A, Morales J, Mountjoy E, Sollis E, Suveges D, Vrousou O, Whetzel PL, Amode R, Guillen JA, Riat HS, Trevanion SJ, Hall P, Junkins H, Flicek P, Burdett T, Hindorf LA, Cunningham F, Parkinson H (2019) The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res* 47:D1005–D1012. <https://doi.org/10.1093/nar/gky1120>
- Cizmecioglu A, Akay Cizmecioglu H, Goktepe MH, Emsen A, Korkmaz C, Esenkaya Tasbent F, Colkesen F, Artac H (2021) Apoptosis-induced T-cell lymphopenia is related to COVID-19 severity. *J Med Virol* 93:2867–2874. <https://doi.org/10.1002/jmv.26742>
- Dai Y, Pei G, Zhao Z, Jia P (2019) A convergent study of genetic variants associated with crohn’s disease: evidence from GWAS, gene expression, methylation, eQTL and TWAS. *Front Genet* 10:318. <https://doi.org/10.3389/fgene.2019.00318>
- Dai Y, Hu R, Pei G, Zhang H, Zhao Z, Jia P (2020) Diverse types of genomic evidence converge on alcohol use disorder risk genes. *J Med Genet* 57:733–743. <https://doi.org/10.1136/jmedgenet-2019-106490>
- Dai Y, Hu R, Manuel AM, Liu A, Jia P, Zhao Z (2021) CSEA-DB: an omnibus for human complex trait and cell type associations. *Nucleic Acids Res* 49:D862–D870. <https://doi.org/10.1093/nar/gkaa1064>
- Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485:376–380. <https://doi.org/10.1038/nature11082>
- Eraslan G, Simon LM, Mircea M, Mueller NS, Theis FJ (2019) Single-cell RNA-seq denoising using a deep count autoencoder. *Nat Commun* 10:390. <https://doi.org/10.1038/s41467-018-07931-2>
- Ernst J, Kellis M (2012) ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods* 9:215–216. <https://doi.org/10.1038/nmeth.1906>
- Gamazon ER, Wheeler HE, Shah KP, Mozaffari SV, Aquino-Michaels K, Carroll RJ, Eyer AE, Denny JC, Consortium GT, Nicolae DL, Cox NJ, Im HK (2015) A gene-based association method for mapping traits using reference transcriptome data. *Nat Genet* 47:1091–1098. <https://doi.org/10.1038/ng.3367>
- Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR (2015) A global reference for human genetic variation. *Nature* 526:68–74. <https://doi.org/10.1038/nature15393>
- Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, Plagnol V (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>
- GTEC Consortium (2020) The GTEC Consortium atlas of genetic regulatory effects across human tissues. *Science* 369:1318–1330. <https://doi.org/10.1126/science.aaz1776>
- Hahne F, Ivanek R (2016) Visualizing genomic data using Gviz and bioconductor, vol 1418. Humana Press Inc., pp 335–351
- Han H, Cho JW, Lee S, Yun A, Kim H, Bae D, Yang S, Kim CY, Lee M, Kim E, Lee S, Kang B, Jeong D, Kim Y, Jeon HN, Jung H, Nam S, Chung M, Kim JH, Lee I (2018) TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic Acids Res* 46:D380–D386. <https://doi.org/10.1093/nar/gkx1013>
- Hastie T, Stuetzle W (1989) Principal curves. *J Am Stat Assoc* 84:502–516. <https://doi.org/10.1080/01621459.1989.10478797>
- Hombink P, Helbig C, Backer RA, Piet B, Oja AE, Stark R, Brassier G, Jongejan A, Jonkers RE, Nota B, Basak O, Clevers HC, Moerland PD, Amsen D, van Lier RA (2016) Programs for the persistence, vigilance and control of human CD8(+) lung-resident memory T cells. *Nat Immunol* 17:1467–1478. <https://doi.org/10.1038/ni.3589>
- Hormozdiari F, van de Bunt M, Segrè AV, Li X, Joo JWJ, Bilow M, Sul JH, Sankararaman S, Pasaniuc B, Eskin E (2016) Colocalization of GWAS and eQTL signals detects target genes. *Am J Hum Genet* 99:1245–1260. <https://doi.org/10.1016/j.ajhg.2016.10.003>
- Jahani M, Dokaneheifard S, Mansouri K (2020) Hypoxia: a key feature of COVID-19 launching activation of HIF-1 and cytokine storm. *J Inflamm (lond)* 17:33. <https://doi.org/10.1186/s12950-020-00263-3>
- Jeong HH, Jia J, Dai Y, Simon LM, Zhao Z (2021) Investigating cellular trajectories in the severity of COVID-19 and their transcriptional programs using machine learning approaches. *Genes (basel)*. <https://doi.org/10.3390/genes12050635>
- Kamat MA, Blackshaw JA, Young R, Surendran P, Burgess S, Danesh J, Butterworth AS, Staley JR (2019) PhenoScanner V2: An expanded tool for searching human genotype-phenotype associations. *Bioinformatics* 35:4851–4853. <https://doi.org/10.1093/bioinformatics/btz469>
- Kasela S, Daniloski Z, Jordan TX, tenOever BR, Sanjana NE, Lapalain T (2021) Integrative approach identifies SLC6A20 and CXCR6 as putative causal genes for the COVID-19 GWAS signal in the 3p21.31 locus. *medRxiv*. <https://doi.org/10.1101/2021.04.09.21255184>
- Kumar BV, Ma W, Miron M, Granot T, Guyer RS, Carpenter DJ, Senda T, Sun X, Ho SH, Lerner H, Friedman AL, Shen Y, Farber DL (2017) Human tissue-resident memory T Cells are defined by core transcriptional and functional signatures in lymphoid and mucosal sites. *Cell Rep* 20:2921–2934. <https://doi.org/10.1016/j.celrep.2017.08.078>
- Lee HS, Kim HR, Lee EH, Jang MH, Kim SB, Park JW, Seoh JY, Jung YJ (2012) Characterization of CCR9 expression and thymus-expressed chemokine responsiveness of the murine thymus, spleen and mesenteric lymph node. *Immunobiology* 217:402–411. <https://doi.org/10.1016/j.imbjo.2011.10.014>
- Lee Y, Luca F, Pique-Regi R, Wen X (2018) Bayesian Multi-SNP genetic association analysis: control of FDR and use of summary statistics. *bioRxiv*:316471–316471
- Liao M, Liu Y, Yuan J, Wen Y, Xu G, Zhao J, Cheng L, Li J, Wang X, Wang F, Liu L, Amit I, Zhang S, Zhang Z (2020) Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. *Nat Med* 26:842–844. <https://doi.org/10.1038/s41591-020-0901-9>
- Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P (2015) The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* 1:417–425. <https://doi.org/10.1016/j.cels.2015.12.004>
- Liu S, Shen J, Fang S, Li K, Liu J, Yang L, Hu CD, Wan J (2020a) Genetic spectrum and distinct evolution patterns of SARS-CoV-2. *Front Microbiol*. <https://doi.org/10.3389/fmicb.2020.593548>
- Liu T, Jia P, Fang B, Zhao Z (2020b) Differential expression of viral transcripts from single-cell RNA sequencing of moderate and severe COVID-19 patients and its implications for case severity. *Front Microbiol*. <https://doi.org/10.3389/fmicb.2020.603509>
- Martin MD, Badovinac VP (2018) Defining memory CD8 T cell. *Front Immunol* 9:2692. <https://doi.org/10.3389/fimmu.2018.02692>
- Mathys H, Davila-Velderrain J, Peng Z, Gao F, Mohammadi S, Young JZ, Menon M, He L, Abdurrob F, Jiang X, Martorell AJ, Ransohoff RM, Hafler BP, Bennett DA, Kellis M, Tsai LH (2019) Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* 570:332–337. <https://doi.org/10.1038/s41586-019-1195-2>
- Nguyen NTT, Contreras-Moreira B, Castro-Mondragon JA, Santana-Garcia W, Ossio R, Robles-Espinoza CD, Bahin M, Collombet S, Vincens P, Thieffry D, van Helden J, Medina-Rivera A, Thomas-Chollier M (2018) RSAT 2018: regulatory sequence analysis tools 20th anniversary. *Nucleic Acids Res* 46:W209–W214. <https://doi.org/10.1093/nar/gky317>
- Pairo-Castineira E, Clohisey S, Klaric L, Bretherick AD, Rawlik K, Pasko D, Walker S, Parkinson N, Fourman MH, Russell CD, Furniss J, Richmond A, Gountouna E, Wrobel N, Harrison D, Wang B, Wu Y, Meynert A, Griffiths F, Oosthuizen W, Kousathanas A, Moutsianas L, Yang Z, Zhai R, Zheng C, Grimes G, Beale R, Millar J, Shih B, Keating S, Zechner M, Haley C, Porteous DJ, Hayward C, Yang J, Knight J, Summers C, Shankar-Hari M, Klennerman P, Turtle L, Ho A, Moore SC, Hinds C, Horby P, Nichol A, Maslove D, Ling L, McAuley D, Montgomery H, Walsh T, Pereira A, Renieri A, Gen OI, Investigators I, Initiative C-HG, and Me I, Investigators B, Gen CI, Shen X, Ponting CP, Fawkes A, Tenesa A, Caulfield M, Scott R, Rowan K, Murphy L, Openshaw PJM, Semple MG, Law A, Vitart V, Wilson JF, Baillie JK (2020) Genetic mechanisms of critical illness in Covid-19. *Nature*. <https://doi.org/10.1038/s41586-020-03065-y>
- Papalexi E, Satija R (2018) Single-cell RNA sequencing to explore immune cell heterogeneity. *Nat Rev Immunol* 18:35–45. <https://doi.org/10.1038/nri.2017.76>
- Payne DJ, Dalal S, Leach R, Parker R, Griffin S, McKimmie CS, Cook GP, Richards SJ, Hillmen P, Munir T, Arnold L, Riley K, McKinley C, Place S, Baretto RL, Newton DJ (2021) The CXCR6/CXCL16 axis links inflamm-aging to disease severity in COVID-19 patients. *bioRxiv*. <https://doi.org/10.1101/2021.01.25.428125>
- Pont F, Tosolini M, Fournie JJ (2019) Single-cell signature explorer for comprehensive visualization of single cell signatures across scRNA-seq datasets. *Nucleic Acids Res* 47:e133. <https://doi.org/10.1093/nar/gkz601>
- Roadmap Epigenomics C, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, Ziller MJ, Amin V, Whitaker JW, Schultz MD, Ward LD, Sarkar A, Quon G, Sandstrom RS, Eaton ML, Wu YC, Pfenning AR, Wang X, Claussnitzer M, Liu Y, Coarfa C, Harris RA, Shores N, Epstein CB, Gjoneska E, Leung D, Xie W, Hawkins RD, Lister R, Hong C, Gascard P, Mungall AJ, Moore R, Chuah E, Tam A, Canfield TK, Hansen RS, Kaul R, Sabo PJ, Bansal MS, Carles A, Dixon JR, Farh KH, Feizi S, Karlic R, Kim AR, Kulkarni A, Li D, Lowdon R, Elliott G, Mercer TR, Neph SJ, Onuchic V, Polak P, Rajagopal N, Ray P, Sallari RC, Siebenthall KT, Sinnott-Armstrong NA, Stevens M, Thurman RE, Wu J, Zhang B, Zhou X, Beaudet AE, Boyer LA, De Jager PL, Farnham PJ, Fisher SJ, Haussler D, Jones SJM, Li W, Marra MA, McManus MT, Sunyaev S, Thomson JA, Tlsty TD, Tsai LH, Wang W, Waterland RA, Zhang MQ, Chadwick LH, Bernstein BE, Costello JF, Ecker JR, Hirst M, Meissner A, Milosavljevic A, Ren

- B, Stamatoypoulos JA, Wang T, Kellis M (2015) Integrative analysis of 111 reference human epigenomes. *Nature* 518:317–329. <https://doi.org/10.1038/nature14248>
- Ruuth M, Soronen J, Kaiharju E, Merikanto K, Perttinen J, Metso J, Lee-Rueckert M, Taskinen MR, Kovanen PT, Oorni K, Olkkonen VM, Jauhainen MS, Laurila PP (2018) USF1 deficiency alleviates inflammation, enhances cholesterol efflux and prevents cholesterol accumulation in macrophages. *Lipids Health Dis* 17:285. <https://doi.org/10.1186/s12944-018-0930-2>
- Schmiedel BJ, Singh D, Madrigal A, Valdovino-Gonzalez AG, White BM, Zapardiel-Gonzalo J, Ha B, Altay G, Greenbaum JA, McVicker G, Seumois G, Rao A, Kronenberg M, Peters B, Vijayanand P (2018) Impact of genetic polymorphisms on human immune cell gene expression. *Cell* 175:1701–1715 e16. <https://doi.org/10.1016/j.cell.2018.10.022>
- Severe Covid GG, Ellinghaus D, Degenhardt F, Bujanda L, Buti M, Albillos A, Invernizzi P, Fernandez J, Prati D, Baselli G, Asselta R, Grimsrud MM, Milani C, Aziz F, Kassens J, May S, Wendorff M, Wienbrandt L, Uellendahl-Werth F, Zheng T, Yi X, de Pablo R, Chercoles AG, Palom A, Garcia-Fernandez AE, Rodriguez-Frias F, Zanella A, Bandera A, Protti A, Aghemo A, Lleo A, Biondi A, Caballero-Garralda A, Gori A, Tanck A, Carreras Nolla A, Latiano A, Fracanzani AL, Peschuck A, Julia A, Pesenti A, Voza A, Jimenez D, Mateos B, Nafria Jimenez B, Quereda C, Paccapelo C, Gassner C, Angelini C, Cea C, Solier A, Pestana D, Muniz-Diaz E, Sandoval E, Paraboschi EM, Navas E, Garcia Sanchez F, Ceriotti F, Martinelli-Boneschi F, Peyvandi F, Blasi F, Tellez L, Blanco-Grau A, Hemmrich-Stanisak G, Grasselli G, Costantino G, Cardamone G, Foti G, Aneli S, Kurihara H, ElAbd H, My I, Galvan-Femenia I, Martin J, Erdmann J, Ferrusquia-Acosta J, Garcia-Etxebarria K, Izquierdo-Sanchez L, Bettini LR, Sumoy L, Terranova L, Moreira L, Santoro L, Scudeller L, Mesonero F, Roade L, Ruhlemann MC, Schaefer M, Carrabba M, Riveiro-Barciela M, Figuera Basso ME, Valsecchi MG, Hernandez-Tejero M, Acosta-Herrera M, D'Angio M, Baldini M, Cazzaniga M, Schulzky M, Cecconi M, Wittig M et al (2020) Genomewide Association Study of Severe Covid-19 with respiratory failure. *N Engl J Med* 383:1522–1534. <https://doi.org/10.1056/NEJMoa2020283>
- Simon LM, Yan F, Zhao Z (2020) DrivAER: Identification of driving transcriptional programs in single-cell RNA sequencing data. *Gigascience*. <https://doi.org/10.1093/gigascience/giaa122>
- Skon CN, Lee JY, Anderson KG, Masopust D, Hogquist KA, Jameson SC (2013) Transcriptional downregulation of S1pr1 is required for the establishment of resident memory CD8+ T cells. *Nat Immunol* 14:1285–1293. <https://doi.org/10.1038/ni.2745>
- Street K, Rizzo D, Fletcher RB, Das D, Ngai J, Yosef N, Purdom E, Dudoit S (2018) Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genom* 19:477. <https://doi.org/10.1186/s12864-018-4772-0>
- Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, 3rd, Hao Y, Stoerckius M, Smibert P, Satija R (2019) Comprehensive integration of single-cell data. *Cell* 177:1888–1902 e21. <https://doi.org/10.1016/j.cell.2019.05.031>
- Szabo PA, Miron M, Farber DL (2019) Location, location, location: tissue resident memory T cells in mice and humans. *Sci Immunol*. <https://doi.org/10.1126/sciimmunol.aas9673>
- Tay MZ, Poh CM, Renia L, MacAry PA, Ng LFP (2020) The trinity of COVID-19: immunity, inflammation and intervention. *Nat Rev Immunol* 20:363–374. <https://doi.org/10.1038/s41577-020-0311-8>
- Urbat SM, Wang G, Carbonetto P, Stephens M (2019) Flexible statistical methods for estimating and testing effects in genomic studies with multiple conditions. *Nat Genet* 51:187–195. <https://doi.org/10.1038/s41588-018-0268-8>
- Vösa U, Claringbould A, Westra H-J, Bonder MJ, Deelen P, Zeng B, Kirsten H, Saha A, Kreuzhuber R, Kasela S, Pervjakova N, Alvaes I, Fave M-J, Agbessi M, Christiansen M, Jansen R, Seppälä I, Tong L, Teumer A, Schramm K, Hemani G, Verlouw J, Yaghoobkar H, Sönmez R, Brown A, Kukushkina V, Kalnapenkis A, Rieger S, Porcu E, Kronberg-Guzman J, Kettunen J, Powell J, Lee B, Zhang F, Arindrarto W, Beutner F, Brugge H, Dmitreva J, Elansary M, Fairfax BP, Georges M, Heijmans BT, Kähönen M, Kim Y, Knight JC, Kovacs P, Krohn K, Li S, Loeffler M, Marigorta UM, Mei H, Momozawa Y, Müller-Nurasyid M, Nauck M, Nivard M, Penninx B, Pritchard J, Raitakari O, Rotzchke O, Slagboom EP, Stehouwer CDA, Stumvoll M, Sullivan P, Hoen PACt, Thiery J, Tönjes A, van Dongen J, van Iterson M, Veldink J, Völker U, Wijmenga C, Swertz M, Andiappan A, Montgomery GW, Ripatti S, Perola M, Kutalik Z, Dermitzakis E, Bergmann S, Frayling T, van Meurs J, Prokisch H, Ahsan H, Pierce B, Lehtimäki T, Boomsma D, Psaty BM, Gharib SA, Awadalla P, Milani L, Ouwehand W, Downes K, Stegle O, Battle A, Yang J, Visscher PM, Scholz M, Gibson G, Esko T, Franke L (2018a) Unraveling the polygenic architecture of complex traits using blood eQTL meta-analysis. *bioRxiv*. <https://doi.org/10.1101/447367>
- Vösa U, Claringbould A, Westra HJ, Deelen P, Zeng B, Kirsten H, Saha A, Kreuzhuber R, Kasela S, Pervjakova N, Alvaes I, Fave MJ, Agbessi M, Christiansen M, Jansen R, Seppälä I, Tong L, Teumer A, Schramm K, Hemani G, Verlouw J, Yaghoobkar H, Sönmez R, Brown A, Kukushkina V, Kalnapenkis A, Rieger S, Porcu E, Kronberg-Guzman J, Kettunen J, Powell J, Lee B, Zhang F, Arindrarto W, Beutner F, Brugge H, Dmitreva J, Elansary M, Fairfax BP, Georges M, Heijmans BT, Kähönen M, Kim Y, Knight JC, Kovacs P, Krohn K, Li S, Loeffler M, Marigorta UM, Mei H, Momozawa Y, Müller-Nurasyid M, Nauck M, Nivard M, Penninx B, Pritchard J, Raitakari O, Rotzchke O, Slagboom EP, Stehouwer CDA, Stumvoll M, Sullivan P, 't Hoen PAC, Thiery J, Tönjes A, van Dongen J, van Iterson M, Veldink J, Völker U, Wijmenga C, Swertz M, Andiappan A, Montgomery GW, Ripatti S, Perola M, Kutalik Z, Dermitzakis E, Bergmann S, Frayling T, van Meurs J, Prokisch H, Ahsan H, Pierce B, Lehtimäki T, Boomsma D, Psaty BM, Gharib SA, Awadalla P, Milani L, Ouwehand W, Downes K, Stegle O, Battle A, Yang J, Visscher PM, Scholz M, Gibson G, Esko T, Franke L (2018b) Unraveling the polygenic architecture of complex traits using blood eQTL meta-analysis. *bioRxiv* 18:10
- Wainberg M, Sinnott-Armstrong N, Mancuso N, Barbeira AN, Knowles DA, Golan D, Ermel R, Ruusalepp A, Quertermous T, Hao K, Bjorkegren JLM, Im HK, Pasaniuc B, Rivas MA, Kundaje A (2019) Opportunities and challenges for transcriptome-wide association studies. *Nat Genet* 51:592–599. <https://doi.org/10.1038/s41588-019-0385-z>
- Wauters E, Van Mol P, Garg AD, Jansen S, Van Herck Y, Vanderbeke L, Bassez A, Boeckx B, Malengier-Devlies B, Timmerman A, Van Brussel T, Van Buyten T, Schepers R, Heylen E, Dauwe D, Doms C, Gunst J, Hermans G, Meersseman P, Testelmans D, Yersby J, Tejpar S, De Wever W, Matthys P, collaborators C, Neys J, Wauters J, Qian J, Lambrechts D (2021) Discriminating mild from critical COVID-19 by innate and adaptive immune single-cell profiling of bronchoalveolar lavages. *Cell Res*. <https://doi.org/10.1038/s41422-020-00455-9>
- Wein AN, McMaster SR, Takamura S, Dunbar PR, Cartwright EK, Hayward SL, McManus DT, Shimaoka T, Ueha S, Tsukui T, Masumoto T, Kurachi M, Matsuhashima K, Kohlmeier JE (2019) CXCR6 regulates localization of tissue-resident memory CD8 T cells to the airways. *J Exp Med* 216:2748–2762. <https://doi.org/10.1084/jem.20181308>
- Weirauch MT, Yang A, Albu M, Cote AG, Montenegro-Montero A, Drewe P, Najafabadi HS, Lambert SA, Mann I, Cook K, Zheng H, Goity A, van Bakel H, Lozano JC, Galli M, Lewsey MG, Huang E, Mukherjee T, Chen X, Reece-Hoyes JS, Govindarajan S, Shaulsky G, Walhout AJM, Bouget FY, Ratsch G, Larrondo LF, Ecker JR, Hughes TR (2014) Determination and inference of eukaryotic

- transcription factor sequence specificity. *Cell* 158:1431–1443. <https://doi.org/10.1016/j.cell.2014.08.009>
- Wen X, Pique-Regi R, Luca F (2017) Integrating molecular QTL data into genome-wide genetic association analysis: probabilistic assessment of enrichment and colocalization. *PLoS Genet* 13:e1006646–e1006646. <https://doi.org/10.1371/journal.pgen.1006646>
- Wu Z, McGoogan JM (2020) Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72314 cases from the Chinese Center for Disease Control and Prevention. *JAMA* 323:1239–1242. <https://doi.org/10.1001/jama.2020.2648>
- Zhao Z, Li H, Wu X, Zhong Y, Zhang K, Zhang YP, Boerwinkle E, Fu YX (2004) Moderate mutation rate in the SARS coronavirus genome and its implications. *BMC Evol Biol* 4:21. <https://doi.org/10.1186/1471-2148-4-21>
- Zhernakova DV, Deelen P, Vermaat M, van Itersson M, van Galen M, Arindrarto W, van 't Hof P, Mei H, van Dijk F, Westra HJ, Bonder MJ, van Rooij J, Verkerk M, Jhamai PM, Moed M, Kielbasa SM, Bot J, Nooren I, Pool R, van Dongen J, Hottenga JJ, Stehouwer CD, van der Kallen CJ, Schalkwijk CG, Zhernakova A, Li Y, Tigchelaar EF, de Klein N, Beekman M, Deelen J, van Heemst D, van den Berg LH, Hofman A, Uitterlinden AG, van Greevenbroek MM, Veldink JH, Boomsma DI, van Duijn CM, Wijmenga C, Slagboom PE, Swertz MA, Isaacs A, van Meurs JB, Jansen R, Heijmans BT, t Hoen PA, Franke L (2017) Identification of context-dependent expression quantitative trait loci in whole blood. *Nat Genet* 49:139–145. <https://doi.org/10.1038/ng.3737>

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