

Potential marker genes for chronic obstructive pulmonary disease revealed based on single-cell sequencing and Mendelian randomization analysis

Gang Sun^{1,*}, Yun Zhou^{2,*}, Xiaoxiao Han^{3,*}, Xiangqian Che¹, Shuo Yu³, Di Song¹, Feifei Ma¹, Lewei Huang¹

¹General Hospital of Northern Theater Command, Shenyang 110000, Liaoning, China

²East China Normal University Wuhu Affiliated Hospital (The Second People's Hospital of Wuhu City), Wuhu 241000, Anhui, China

³The Second People's Hospital of Hefei, Hefei Hospital Affiliated to Anhui Medical University, Hefei 230022, Anhui China

*Equal contribution

Correspondence to: Lewei Huang; email: 13309880429@163.com, <https://orcid.org/0009-0000-7604-3689>

Keywords: chronic obstructive pulmonary disease, genetics, single-cell RNA sequencing, quantitative trait loci, mendelian randomization analysis

Received: December 5, 2023

Accepted: April 15, 2024

Published: May 23, 2024

Copyright: © 2024 Sun et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/) (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Background: Progress is being made in the prevention and treatment of chronic obstructive pulmonary disease (COPD), but it is still unsatisfactory. With the development of genetic technology, validated genetic information can better explain COPD.

Objective: The study utilized scRNA-seq and Mendelian randomization analysis of eQTLs to identify crucial genes and potential mechanistic pathways underlying COPD pathogenesis.

Methods: Single-cell sequencing data were used to identify marker genes for immune cells in the COPD process. Data on eQTLs for immune cell marker genes were obtained from the eQTLGen consortium. To estimate the causal effect of marker genes on COPD, we selected an independent cohort (ukb-b-16751) derived from the UK Biobank database for two-sample Mendelian randomization analysis. Subsequently, we performed immune infiltration analysis, gene set enrichment analysis (GSEA), and co-expression network analysis on the key genes.

Results: The 154 immune cell-associated marker genes identified were mainly involved in pathways such as vacuolar cleavage, positive regulation of immune response and regulation of cell activation. Mendelian randomization analysis screened four pairs of marker genes (GZMH, COTL1, CSTA and CD14) were causally associated with COPD. These four key genes were significantly associated with immune cells. In addition, we have identified potential transcription factors associated with these key genes using the Cistrome database, thus contributing to a deeper understanding of the regulatory network of these gene expressions.

Conclusions: This eQTLs Mendelian randomization study identified four key genes (GZMH, COTL1, CSTA, and CD14) causally associated with COPD, providing new insights for prevention and treatment of COPD.

INTRODUCTION

Chronic obstructive pulmonary disease (COPD) is a respiratory disease characterized by persistent airway

inflammation and is now the third leading cause of death worldwide [1]. Chronic inflammation in the periphery of the bronchi and fine bronchioles could lead to lung tissue destruction, fibrosis and

emphysema. Continued progression of COPD leads to deterioration of the patient's condition and a corresponding increase in healthcare costs and mortality [2]. However, to date, there is no satisfactory therapeutic regimen that can slow the progression of COPD or reduce mortality [3], and one important reason for this is that the underlying pathogenesis of COPD is complex. Consequently, identification of new markers is important for the prognosis and management for COPD patients.

In patients with COPD, the exacerbation of lung tissue damage and subsequent lung remodeling due to persistent chronic inflammation is closely associated with an aberrant immune response. In this process, a variety of immune cells involved in both innate and adaptive immunity play important roles [4]. For example, macrophages and neutrophils in the alveoli could secrete a variety of proteases, cytokines, and chemokines thereby leading to destruction of the alveolar wall and thus exacerbation of emphysema [5, 6]. Adaptive immunity is a process that occurs when innate immunity is activated and involves the T-cell population responsible for early warning effects and the B-cell population that produces antibodies. Relevant studies have shown that fine bronchial damage and remodeling of alveolar tissue in lung tissue were associated with over-infiltration of CD4, CD8 and B cells [7]. In conclusion, the persistence of chronic inflammation influences the course of COPD and this inflammatory response is closely associated with an increase in a variety of immune cells [8].

With the development of bioinformatics technology, there are more possibilities to understand the mechanisms of disease progression. Traditional transcriptome sequencing measures the average expression of individual genes in a large population of cells and is primarily used to study differential expression between tissues [9]. It is therefore difficult to detect molecular differences that are only relevant to specific cell types, especially when gene expression is in low abundance in some specific cell types [10]. In recent years, technologies that can be used to assess the expression of gene profiles in individual cells (single-cell RNA sequencing) have been developed [11]. Single-cell RNA sequencing (scRNA-seq) is capable of assessing the amount of gene expression within a single cell, not only to detect rare or low abundance cell populations, but also to label previously unknown cell types or subtypes [12, 13]. In addition to this, scRNA-seq could help researchers to understand cell-to-cell information transfer, thus enabling the exploration of pathological mechanisms of diseases and the identification of new diagnostic markers or

new therapeutic targets at the single-cell level [14–16]. For example, Li et al. noted in their study that scRNA-seq could be used to identify transcriptional changes and levels of individual proteins that may contribute to the development of emphysema in a cell-type specific manner [17]. Pei et al. demonstrated changes in immune cell subtypes during the progression of COPD by using peripheral blood mononuclear cells from patients with COPD for scRNA-seq [18]. However, the current types of scRNA-seq studies were mainly cross-sectional studies, which were designed to explore potential associations between genes and diseases rather than to establish causal relationships [19]. If the causal relationship between markers labeled by scRNA-seq and disease occurrence could be verified, it would not only save the cost of experiments but also provide a clearer direction for research.

In order to explore the causal associations between risk factors and diseases, Mendelian randomization (MR) analysis has attracted the attention of researchers as a new epidemiological method [20]. MR analysis is based on the principle of random assignment of genetic variants, with single nucleotide polymorphisms (SNPs) as instrumental variables (Ivs) representing the characteristics of interest, which can be used to validate causality with diseases. Thus, the combination of MR analysis with scRNA-seq results may provide greater insight into the mechanisms of development of certain diseases. For example, Wu et al. identified in their study that risk genes for schizophrenia were highly expressed in specific neuronal cells and clarified the causal relationship [21]. Indeed, expression quantitative trait loci (eQTLs) in lung tissues have gained some degree of discovery in the pathogenesis of COPD [22, 23]. However, association analyses of eQTLs based on scRNA-seq results and precision locus analysis combined with genome-wide association study (GWAS) data were still limited. Therefore, we propose in this study to utilize GWAS data on COPD with the results of scRNA-seq for a comprehensive analysis, aiming to provide new insights into the pathogenesis of COPD.

MATERIALS AND METHODS

Data sources

Exposure data

The eQTLs data in this study were obtained from the eQTLGen Consortium (<https://www.eqtlgen.org>) database. The eQTL Gen Joint Research Program aims to analyze gene expression levels in peripheral blood and gain insight into the genetic basis of complex traits. More details about the eQTLGen

Consortium data can be found in a previous report [24].

Outcome data

The GWAS data for those diagnosed with COPD come from UK Biobank. UK Biobank contains genetic and health information on over 500,000 participants of European descent. We selected an independent cohort (ID: ukb-b-16751) containing 3871 COPD patients and 459,139 controls. The outcome data are publicly available and more details can be found at <https://gwas.mrcieu.ac.uk/datasets/ukb-b-16751/>.

RNA-sequencing data

The GEO (Gene Expression Omnibus) database (<https://www.ncbi.nlm.nih.gov/geo/>) is a gene expression database established and maintained by the National Center for Biotechnology Information (NCBI). In this study, we obtained serial matrix data from this publicly available database containing lung tissue samples from COPD cases (GSE57148) annotated as GPL11154. Expression profiling data from a total of 189 groups of patients were included, including 91 controls and 98 COPD patients. In addition, we also obtained data from scRNA-seq of lung tissue samples from 3 COPD patients (GSE167295) for analysis.

Gene set

The gene set related to immunization in this study was obtained from the GeneCards (<https://www.genecards.org>) database.

Analysis methods for single-cell sequencing

Firstly, gene expression profiles of scRNA-seq samples were read using the Seurat package and screened for aberrantly expressed samples ($nFeature_RNA > 100$ and $percent.mt < 10$). Subsequently, the resulting information was normalized with homogenization and subjected to principal component analysis. The optimal number of principal components (PCs) was observed by ElbowPlot curves ($n=12$), and the positional relationship between different clusters was subsequently investigated using the t-distributed stochastic neighbor embedding (TSNE) algorithm. To add annotations to the clusters, we used the annotation file HumanPrimaryCellAtlasData, which was included in the cellDex package, to label each cluster as a cell type that is closely related to the formation of the disease. Finally, the labeled genes corresponding to each cell subtype were obtained by setting the thresholds of $logfc$ and minimum pct to 1 and 0.45, respectively, in the FindAllMarkers function. The unique marker genes associated with each cell subtype were screened using $p_val_adj < 0.05$ and $|avg_log2FC| > 1.5$ as conditions.

Gene functional enrichment analysis

Through the Metascape database (<https://www.metascape.org>), we have functionally annotated key genes in order to further explore their interrelationships. Based on this, we would conduct Genomics Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway studies. Minimum number of overlaps (Min overlap) greater than or equal to 3 and p-value less than or equal to 0.01 were judged to be statistically significant.

Mendelian randomization analysis

We performed a MR analysis using the MR Base database and GWAS database (<https://gwas.mrcieu.ac.uk/>). The Mendelian randomization analysis in this study was performed on the basis of satisfying the assumptions of association, independence and exclusivity. In the association analysis, SNPs significantly associated with each locus first needed to be screened as variable instruments ($P < 10^{-8}$). Subsequently in linkage disequilibrium (LD) analysis, we retained specific SNPs based on $R^2 < 0.001$ (using a clustering window of 10,000 kb) ($P < 5 \times 10^{-8}$). Subsequently, we assessed the causal effect between genotype and disease using four statistical methods (inverse variance weighting, MR Egger, weighted median, and weighted mode). Specifically, inverse variance weighted (IVW) is designed to utilize the Wald value for each SNP in a pooled analysis. MR-Egger is able to assess whether genetic variation has multiple effects on the outcome that are on average non-zero (directional multiple effects), as well as to provide causal effects under the assumption of instrumental strength independent of direct effect (InSIDE) estimation. The weighted median model is able to correctly estimate causality in up to 50% of cases where the Ivs are invalid, and the Weighted mode approach has greater ability to detect causal effects because of smaller biases and lower type I error rates. By evaluating each SNP, we would be able to further screen and validate whether there is a true causal relationship among them. We analyzed each SNP in the screened causal relationships by independent assessment using Wald ratio to screen and validate the results. Finally, we used the leave-one-out analysis to verify the independence of the screened causal relationships.

Co-location analysis

We performed co-localization analyses using the coloc algorithm on eQTL summary data and GWAS data from COPD. The 100-kilobase region around the index SNP was used to calculate the posterior probability. In the results, H3 indicated the posterior probability

that two traits were correlated but had different causal variants; H4 indicated the posterior probability that two traits were correlated and shared a single causal variant.

Analysis of immune cell infiltration

To assess the relative proportion of 22 types of immune infiltrating cells, we employed the CIBERSORT algorithm to analyze RNA-seq data from different sub-groups of COPD patients. Subsequently, we evaluated the correlation between gene expression levels and immune cell content using Spearman correlation analysis. P-value less than 0.05 was considered statistically significant.

Gene set enrichment analysis

To identify genes whose gene expression profiles differed between the high- and low-expression groups in COPD patients, we performed gene set enrichment analysis (GSEA; <http://www.broadinstitute.org/gsea>). The numbers of maximum and minimum gene sets were set as 500 and 15 as filtering conditions, respectively. The selected gene sets were ranked 100 times and the final results were obtained based on a *P*-value of <0.05 and a false discovery rate (FDR) value of 0.25 [25].

Regulatory network of target genes

Cistrome DB is currently a more comprehensive database for studying ChIP-seq and DNase-seq, encompassing a total of 30,451 human and 26,013 mouse samples of transcription factors, histone modifications, and chromatin. In this study, we probed the regulatory relationships between transcription factors and target genes through the Cistrome DB database, with the genome file set to hg38 and the transcription start site set to 10kb, all visualized via Cytoscape.

Statistical analysis

All data analyses were conducted in R version 4.0. MR analysis follows three fundamental assumptions: [1] the correlation assumption, stating that the instrumental variable is closely related to the exposure factor, but not to the outcome; [2] the independence assumption, meaning that the instrumental variable and the confounder have no relationship; and [3] the exclusion assumption, stating that the instrumental variable only affects the outcome through the exposure factor. If the instrumental variable affects the outcome through other pathways, it is considered to be genetic pleiotropy. A significance threshold of $P < 0.05$ was set for statistical differences.

Data availability statement

The original contribution to the study was included in the article/supplementary material. For further inquiries, please contact the corresponding author.

RESULTS

Preprocessing results of single-cell sequencing data expression profiles

This study includes the results of scRNA-seq of lung tissue from 3 patients with COPD. To proceed with the subsequent analysis, we only retained cells with nFeature_RNA greater than 100 and percent.mt less than 10. A total of 13,838 cells with expression levels of features were included for further analysis (Supplementary Figure 1A, 1B). We displayed the gene expression patterns across samples and marked the top 5 genes with the highest standardized variance (Supplementary Figure 1C).

Single-cell sample subtype clustering analysis

We utilized Principal Component Analysis (PCA) in order to reduce the dimensionality of the differential genes. The results suggested that the 20 genes that were presented scored differently on different dimensions. (Supplementary Figure 2A). However, when we analyzed the PCA dimensionality reduction among samples, we found that the overall difference was not significant (Supplementary Figure 2B). Through observing the ElbowPlot, we found that the optimal pc number was 12 (Supplementary Figure 2C, 2D), and finally obtained 18 cell subtypes through TSNE method (Supplementary Figure 2E).

Annotation of cluster subtypes

We used HumanPrimaryCellAtlasData as an annotation dataset and used the SingleR software package to annotate each cell subtype. The 18 clusters were annotated to B-cell, Endothelial cells, Epithelial cells, Macrophage, Monocyte, NK cell, T-cells, and Tissue stem cells (Supplementary Figure 2F, 2G). Finally, we extracted a total of 411 marker genes for cell subtypes from the single-cell expression profiles using the FindAllMarkers function (Supplementary Table 1).

Marker gene function analysis

In the functional analysis of marker genes, we first compared genetic differences between cell subtypes associated with the immune system (B_cells, macrophages, monocytes, NK_cells, and T_cells) and identified 154 marker genes (Figure 1A–1E). To further

explore the functions of these marker genes, we utilized the Metascape database to conduct a pathway analysis. The results of this analysis suggested that these marker genes were primarily enriched in pathways related to the lytic vacuole, the positive regulation of immune response, and the regulation of cell activation (Figure 1F). Additionally, we utilized Cytoscape software to visualize the protein interaction network of the gene sets (Figure 1G).

Mendelian randomization analysis of marker genes

We performed a Mendelian randomization analysis of marker genes to further identify key genes marked by marker genes that would have an impact on COPD. Using the COPD meta-analysis summary statistics (id: ukb-b-16751 with 459,139 controls and 3,871 cases), we extracted 120 pairs of marker genes causally associated with COPD (Supplementary Table 2). Subsequently, by MR analysis, we screened 4 pairs of marker genes that were causally associated with positive eQTL results (Figure 2A–2D, p-value of IVW less than 0.05). These 4 pairs of marker genes were GZMH, COTL1, CSTA, and CD14. The effect values corresponding to these 4 pairs of marker genes were COTL1 (OR= 0.997 ; 95% CI: 0.996 - 0.999; $P = 0.005$), GZMH (OR= 0.999; 95% CI: 0.997-0.99995; $P = 0.043$), CSTA (OR= 1.001; 95% CI: 1.000-1.002; $P = 0.032$), and CD14 (OR= 1.001; 95% CI: 1.000-1.002; $P = 0.011$), as described in more details in Supplementary

Table 3. These results suggest that CSTA and CD14 could be high-risk factors for COPD; while COTL1 and GZMH might be low-risk factors for COPD. In addition, we performed sensitivity analysis to determine the reliability of the causality for the four genes. The results showed that the exclusion of any of the SNPs had a small effect on the overall error, suggesting that the selected 4 pairs of key genes have a robust causal association with COPD (Figure 3). Finally, we performed co-localization analysis of these four genes at the eQI-GWAS level (Figure 4). We found that COTL1 was associated with COPD and shared genetic loci with single mutations (Figure 4B and Supplementary Table 4).

Clinical predictive value for key genes

The microenvironment is mainly composed of immune cells, extracellular matrix, various growth factors, inflammatory factors, and special physicochemical features, etc. The microenvironment significantly influences the diagnosis of the disease, survival outcome, and sensitivity of clinical treatment. By analyzing the relationship between gene expression levels and immune infiltration, the potential molecular mechanisms by which the expression levels of key genes influence the development of COPD can be further explored. Our results showed that the distribution of immune levels of different immune factors in the samples was not entirely consistent

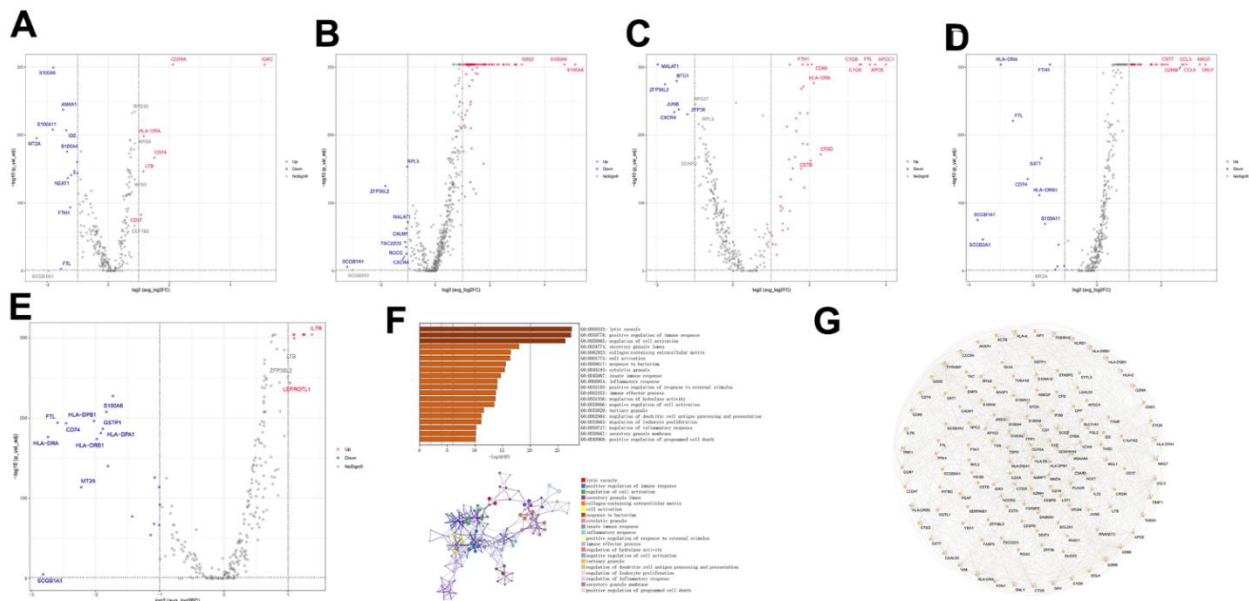


Figure 1. Differential gene expression and functional analysis of immune-related cell subtypes in COPD samples. (A–E) Volcano plots of the differential genes of the five immune cells, B_cell, Macrophage, Monocyte, NK_cell, and T_cells, respectively, in COPD samples, according to the order. **(F)** Functional enrichment analysis of the differential genes. **(G)** Differential gene protein-protein interactions network.

(Figure 5A). There were several significant correlations among the immune factors (Figure 5B), monocytes, Dendritic cells activated, etc. were significantly higher in COPD samples than in control samples, while T cells follicular helper, Macrophages M2, etc. were

significantly lower than in control samples (Figure 5C). The gene GZMH was significantly positively correlated with T cells CD8, NK cells resting, etc., and significantly negatively correlated with B cells naive, Macrophages M2; the gene COTL1 was significantly

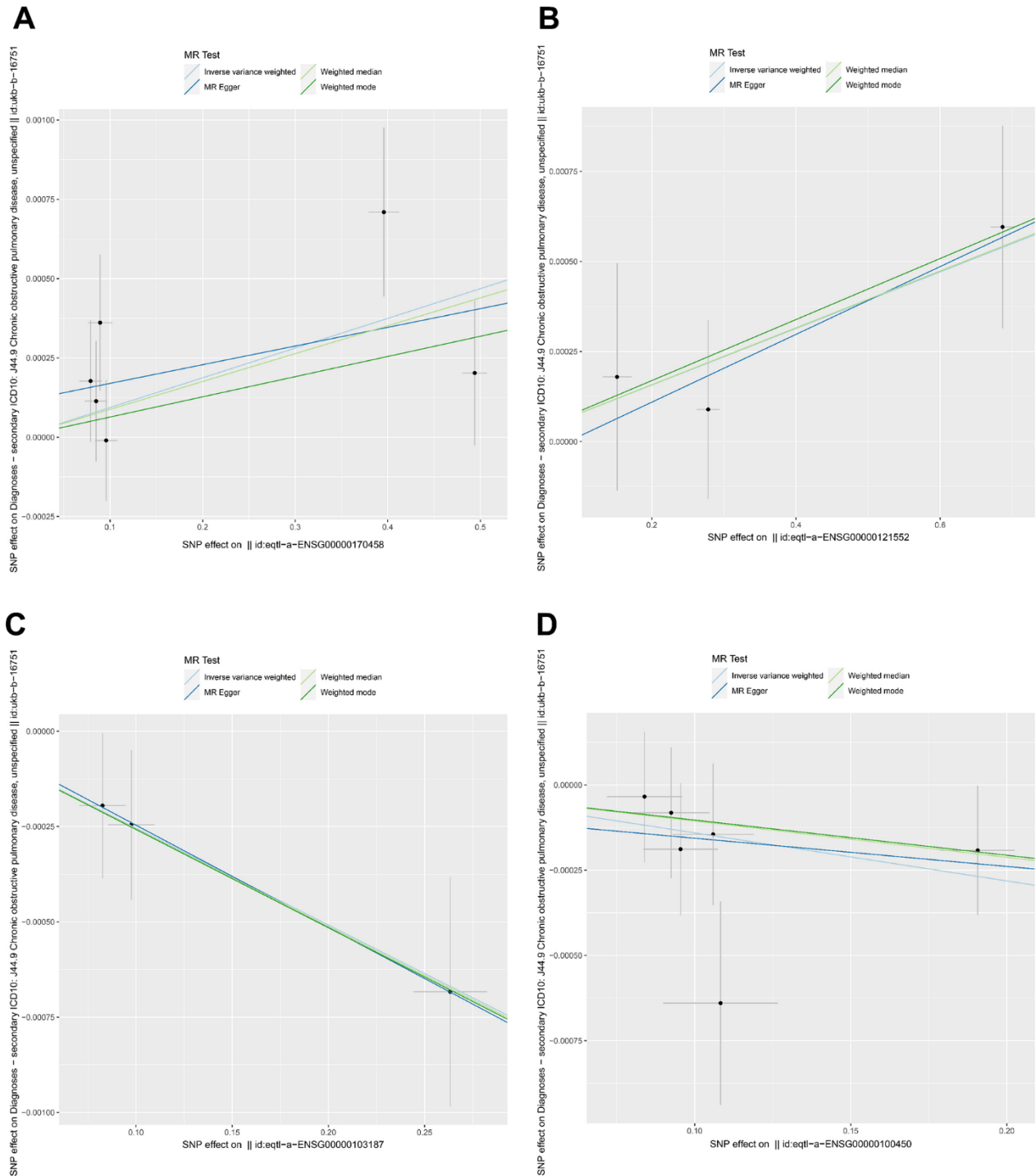


Figure 2. Mendelian randomization analysis between key genes and COPD. (A–D) Genes represented were CD14, CSTA, COTL1, and GZMH, respectively. Different colors indicated different statistical methods, and the slopes of the lines denoted the causal effect of each method, respectively.

positively correlated with Macrophages M2, Dendritic cells resting, etc., and significantly positively correlated with T cells CD4 memory resting, Dendritic cells activated, etc.; gene CSTA is significantly positively correlated with Monocytes, Macrophages M0, etc., and negatively correlated with Plasma cells, NK cells activated, etc.; gene CD14 is significantly correlated with Mast cells activated, Neutrophils significantly positively correlated, and significantly negatively correlated with Plasma cells, Mast cells resting (Figure 5D). We obtained the correlations between these key genes and different immune factors, including immunomodulators, chemokines and cellular receptors from the TISIDB database, and the results showed that the key genes were significantly correlated with several immune factors (Figure 6).

Disease gene expression levels

We obtained the disease genes associated with COPD through the GeneCards database (<https://www.genecards.org/>), and we analyzed the expression levels of four key genes (Figure 7A–7C) and the expression levels of the top 20 genes with the highest relevance score. We found that the expression levels of key genes were significantly correlated with the expression levels of several disease-related genes, including CD14, which was significantly positively correlated with HMOX1 (correlation coefficient=0.646), and COTL1, which was

significantly negatively correlated with SCGB1A1 (cor=-0.351) (Figure 7D). Subsequently, we analyzed the expression of these four key genes and the top 10 immune genes with the highest relevance score at the single-cell level, and we found that the key genes were co-expressed with several immune genes at the single-cell level (Supplementary Figures 3–6).

Potential signaling mechanisms and regulatory networks for key genes

We performed an in-depth enrichment analysis of four key genes to explore the relevant signaling pathways by which they may influence disease progression. For example, we found that the main pathways for gene ontology (GO) enrichment of GZMH gene were “LEUKOCYTE MEDIATED CYTOTOXICITY, NEGATIVE REGULATION OF CALCIUM ION TRANSPORT INTO CYTOSOL, etc” and the pathways for KEGG enrichment mainly included “ALANINE ASPARTATE AND GLUTAMATE METABOLISM, ANTIGEN PROCESSING AND PRESENTATION, etc”. The enrichment results for all four key genes were shown in Figure 8. Subsequently, we compiled these four key genes into a gene set for this analysis to further investigate the transcriptional regulatory networks encompassing these genes. Using the Cistrome DB online database, we identified the transcription factors associated with the key genes.

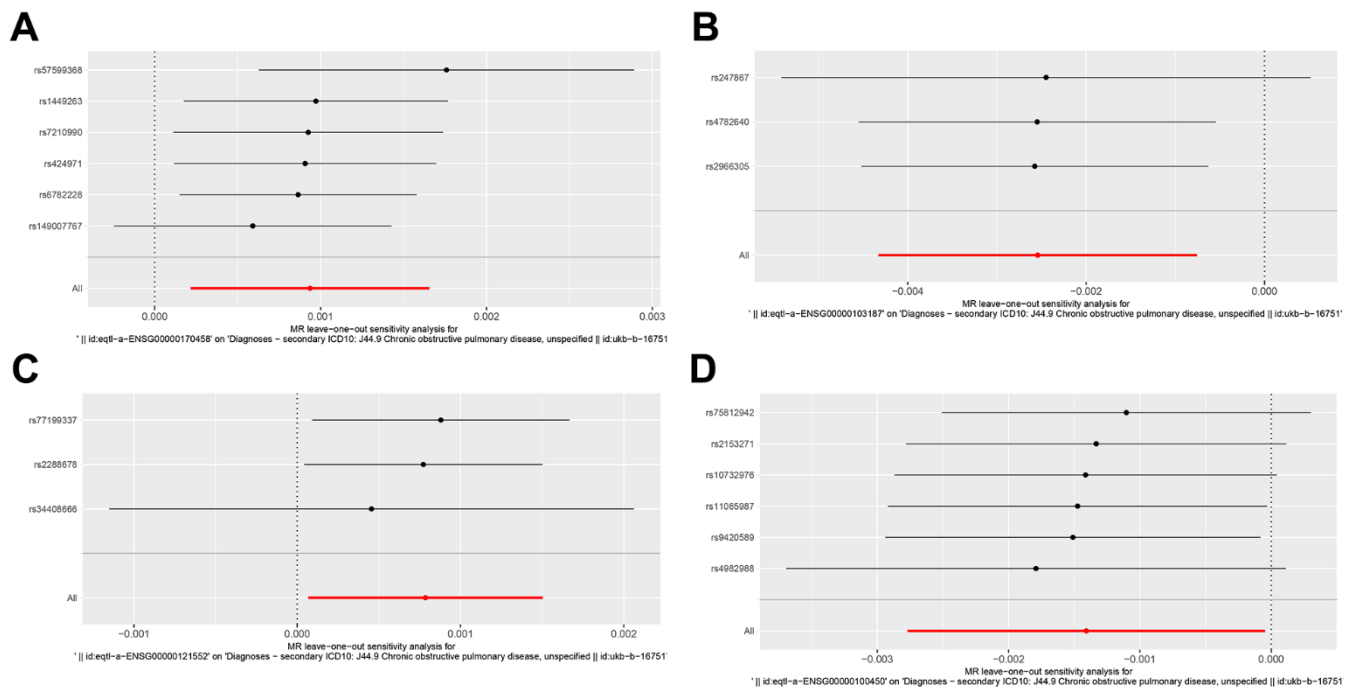


Figure 3. Forest plot of key genes corresponding to SNPs tested by leave-one-out analysis. (A–D) Respectively represented by the genes CD14, COTL1, CSTA and GZMH.

Specifically, 86 transcription factors were predicted to be associated with CD14, 93 with COTL1, 54 with CSTA, and 94 with GZMH. Finally, we employed Cytoscape software to visualize the transcriptional regulatory networks of COPD-related key genes (Figure 9).

DISCUSSION

Although smoking is widely recognized as a key factor in the development of COPD, a puzzling phenomenon is that the disease may continue to progress even when COPD patients successfully quit smoking. This phenomenon has prompted researchers to delve deeper

into the genetic factors of COPD. Currently, more and more researchers tend to regard COPD as a disease with a genetic basis. Recent breakthroughs in genomic research worldwide have led to the discovery of numerous genes associated with the development of COPD, providing new clues to the pathogenesis of the disease [26–28]. Various GWAS studies have also suggested that the pathogenesis of COPD is not caused by a single genetic variant, but rather by a disruption in the balance of the biological network consisting of genes and proteins, leading to intricate changes in the pathological processes of the disease. The exploration of the pathogenesis of COPD through genetic

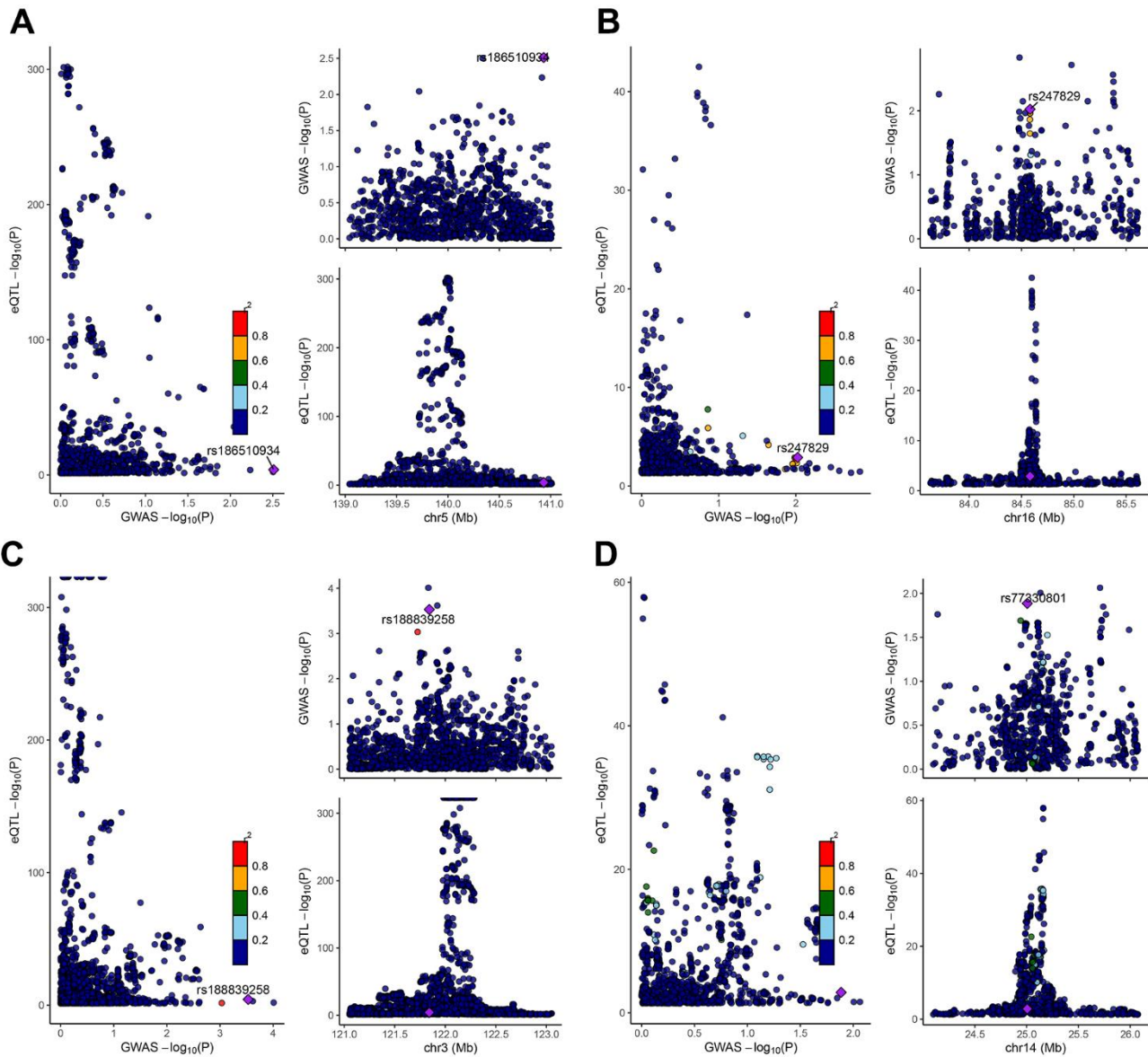


Figure 4. Co-localization analysis between SNPs of key genes and GWAS data from COPD patients. The genes represented from (A–D) were CD14, COTL1, CSTA, and GZMH, respectively. The X-axis indicated the P-value of GWAS, and the Y-axis indicated the P-value of eqtl.

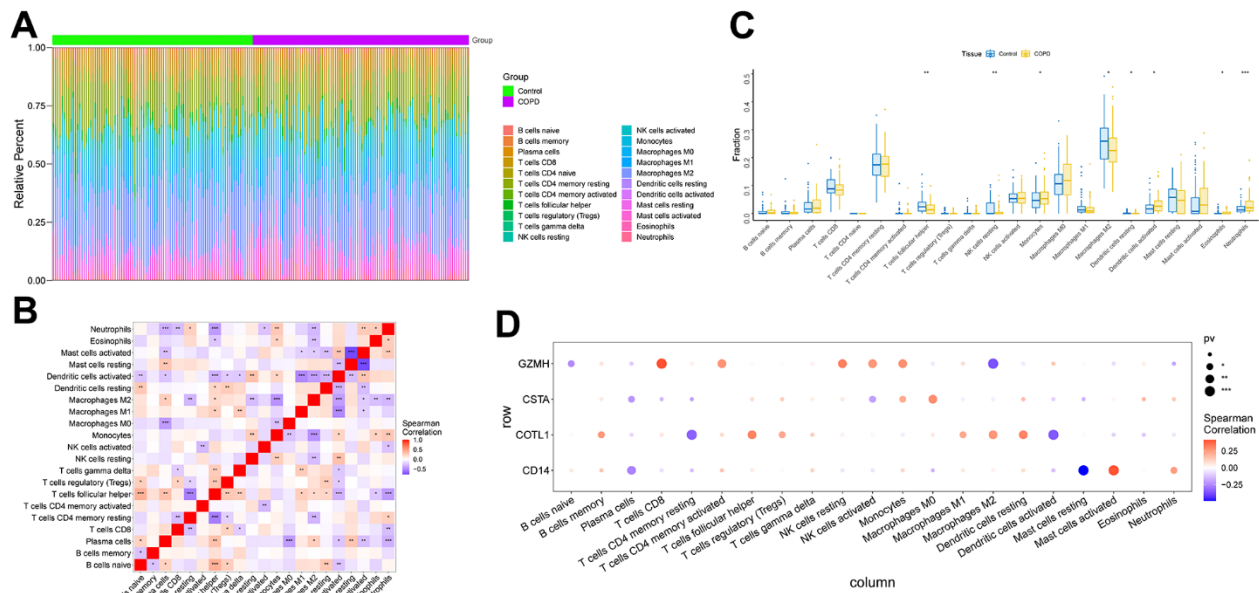


Figure 5. Analysis of immune cell infiltration in COPD samples. (A) Relative percentages of immune cell subpopulations. (B) Correlation between immune cell subpopulations, where red represents positive correlation and blue represents negative correlation. (C) Differences in immune cell content between COPD and control samples. (D) Correlation between 4 key genes and immune cells. Where red represents a positive correlation and blue represents a negative correlation, the larger black solid circle represents a more statistically significant difference.

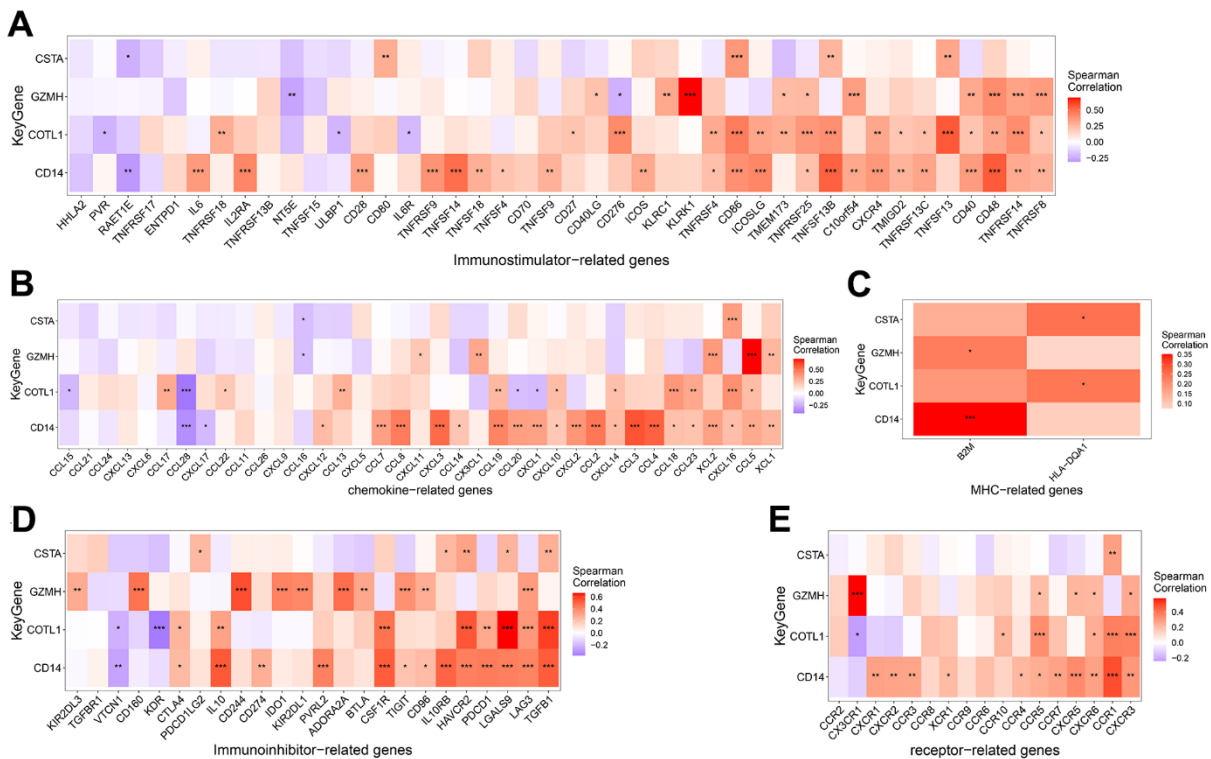


Figure 6. Correlation between key genes and Immunoinhibitor (A), Chemokine (B), MHC (C), Immunostimulator (D) and Receptor related immunogenes (E).

technology holds immense clinical significance. However, with the development of gene sequencing technology, it has been recognized that the regulation of numerous genes is specific to certain cells. Previous sequencing results based on a wide range of tissues may have resulted in the omission of information regarding expression quantitative trait loci (eQTLs) specific to certain cell subtypes [21]. Therefore, the utilization of scRNA-seq technology has become invaluable in the search for crucial COPD genes.

In our results, we identified 8 cell subtypes, 5 of which were immunologically related. The immunological profile of COPD during its progression remains unclear, so further studies of immune cell infiltration in COPD are still needed. The immune inflammation associated with COPD is at the heart of the disease, which is one of the reasons why current treatments have failed to stop the progression of the disease and the continued damage to lung tissue [29]. Previous studies have shown that the effector molecule functions of neutrophils and macrophages involved in chronic inflammation in COPD are suppressed [30, 31]. In addition, it has been shown that neutrophils and lymphocytes can have a synergistic effect to enhance migration toward chemokine receptor 3 (CXCR3) and chemokine ligand 5 (CCL5), leading to immune infiltration of macrophages and T cells in the lungs of COPD patients [32]. Our study showed that key genes of immune-related cell subtypes were mainly enriched in the positive regulation of inflammatory

response and cell activation, which is consistent with the above findings [33].

Most importantly, our study identified four key immune-related genes based on COPD patients. GZMH, also known as granzyme H, is a member of the serine protease family [34]. Granzyme H is expressed in immune cells (mainly cytotoxic T cells and natural killer cells) and promotes the release of inflammatory cytokines and cytolytic proteins from effective immune cells thereby inducing apoptosis in target cells [34, 35]. Since an important feature of COPD is the excessive deposition and remodeling of the extracellular matrix (ECM) around small airways and can provide mechanical pathways for the infiltration and migration of harmful immune cells [29]. Previous studies have shown that expression of granzyme's B can contribute to the degradation and remodeling of the ECM in the extracellular environment, thereby promoting the development of an emphysema phenotype [36]. A positive correlation between GZMH and inflammation was indirectly demonstrated in a clinical randomized controlled trial suggesting that the downregulation of the GZMH gene was more pronounced in COPD patients treated with oral prednisone than in other family members [37]. In summary, a limited number of studies have been reported explaining that high expression of the granzyme family in COPD is highly correlated with the severity of the disease. However, it has also been shown that GZMH levels are reduced in

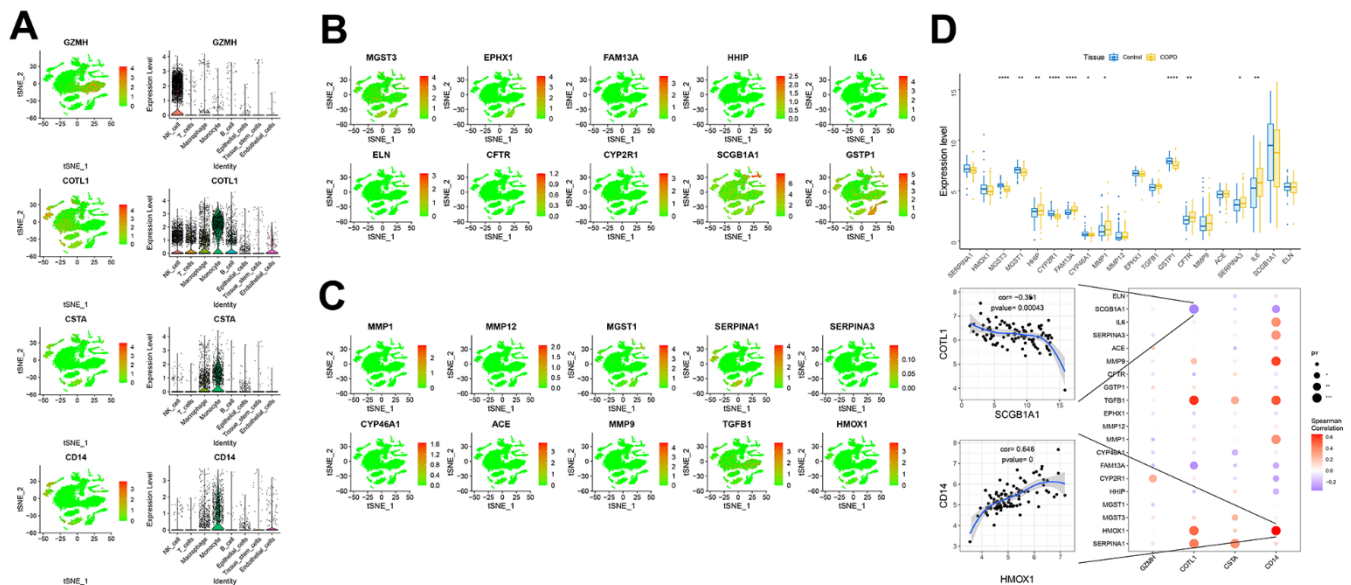


Figure 7. Expression of key genes in single cells. (A–C) Expression of key genes and disease-causing genes in cells. (D) The upper panel represented differences in the expression of disease-regulated genes, with control patients in blue and disease patients in yellow. The lower panel represented the Pearson correlation analysis between key genes and disease genes. Blue color represents negative correlation and red color represents positive correlation.

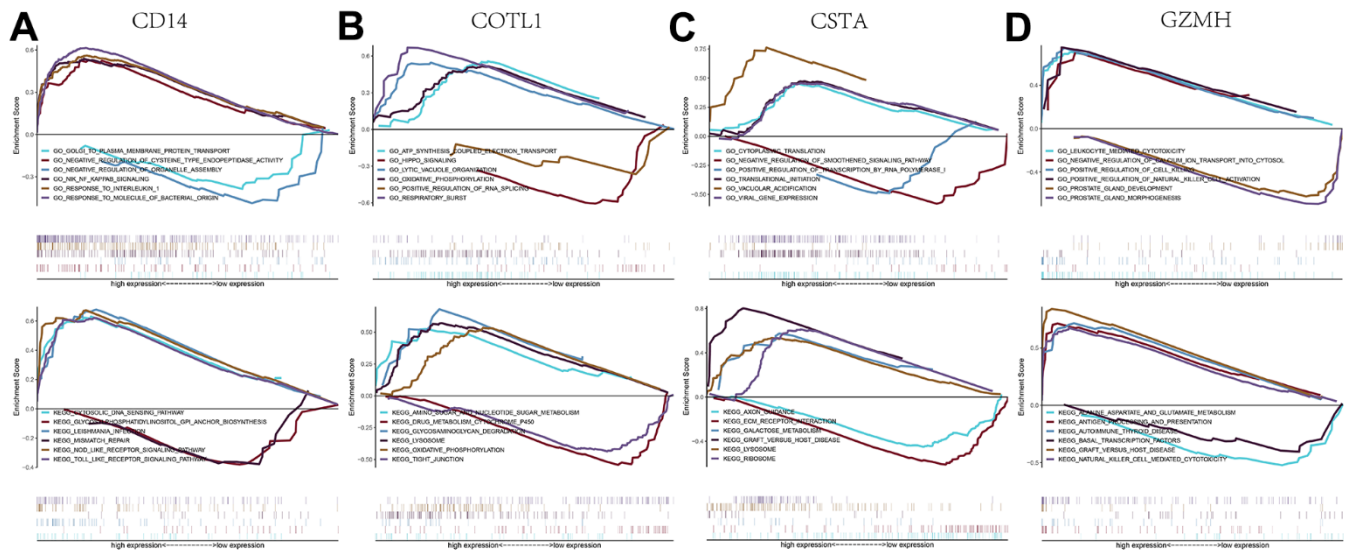


Figure 8. GSEA analysis of key genes. (A–D) GO and KEGG signaling pathways involved in different key genes.

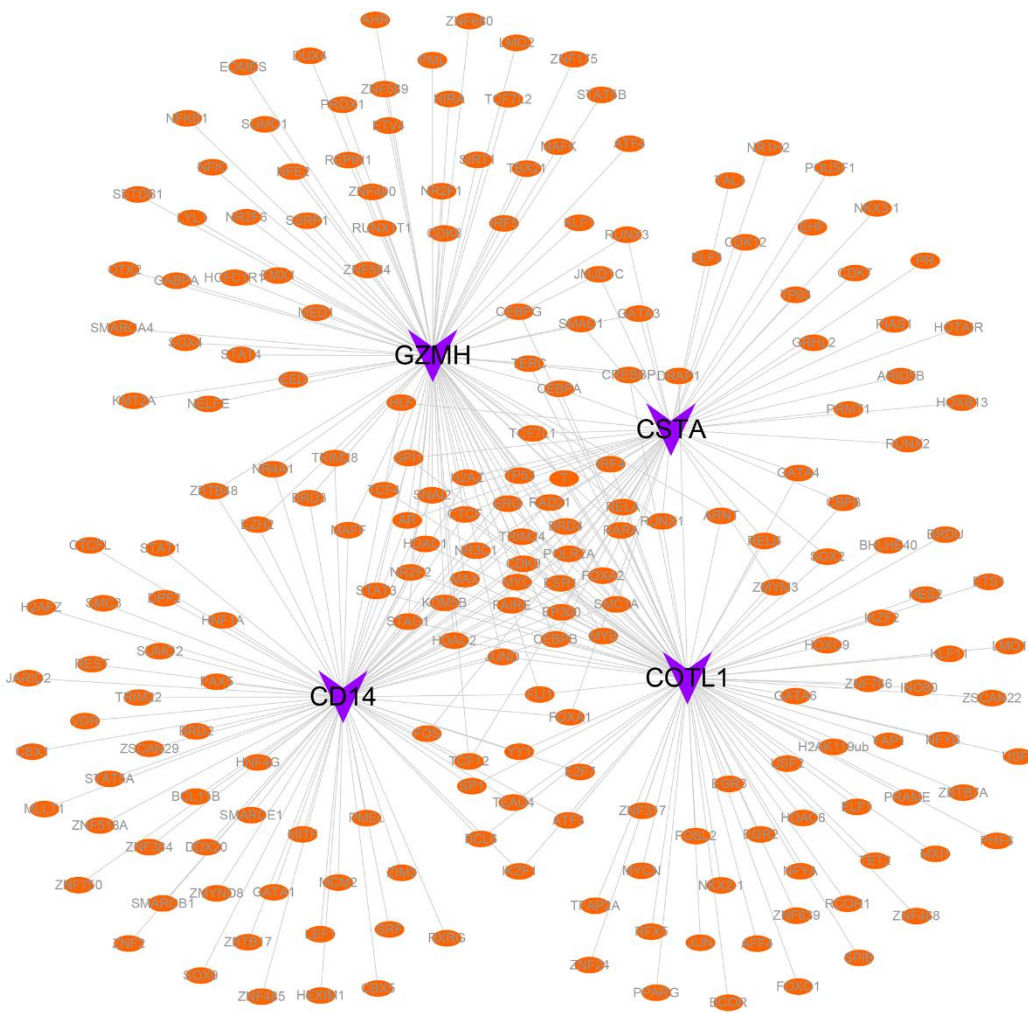


Figure 9. Transcriptional regulatory networks of key genes. Purple indicates mRNAs and orange indicates transcription factors.

patients with moderate and/or severe severity when combined with acute neocoronary pneumonia (COVID) in patients with co-morbidities, even at high levels of inflammation [38]. Our findings suggested that GZMH expression may be associated with a lower risk of developing COPD. The new puzzling question is whether GZMH expression is harmful or beneficial? Or does it play different roles at different stages of COPD? Therefore, more subsequent studies are necessary.

In our study, we found that another gene associated with a low risk of developing COPD is COTL1. COTL1 is one of the proteins that encodes for the regulation of the actin cytoskeleton, which acts mainly through the innate immune pathway [39]. In our study, COTL1 was strongly positively associated with dendritic cells as well as macrophages, which is similar to previous findings. Studies on COTL1 and immune cell infiltration in COPD samples are still unknown. COTL1 has been reported mainly in tumor samples. In a study on breast cancer, investigators demonstrated that COTL1 was highly expressed in tumor tissues and positively correlated with dendritic cells and M2 macrophages [40]. The influence of innate immunity in the COPD process is complex, with multiple types of macrophages being closely associated with pro-inflammatory responses [41], but M2 macrophages are also associated with inflammatory abatement, tissue repair, and reduction of pro-inflammatory cytokines [42]. Also not to be overlooked is the role of dendritic cells as the first barrier in the face of pathogenic microorganisms [43].

Our study demonstrated 2 key genes associated with a high risk of developing COPD, including CSTA and CD14. CSTA, also known as cystatin A, has cysteine protease inhibitory effects [44]. Butler demonstrated in their study that patients with smoking and COPD have increased levels of CSTA expression and a dose-responsive relationship with the severity of the disease [45]. Although there is still no relevant study explaining whether the high expression of CSTA gene in COPD patients is associated with the infiltration of immune cells, earlier researchers have demonstrated that the local proliferation of cystatin A-producing epidermal cells is associated with a high degree of infiltration of monocytes and granulocytes in inflammatory diseases [46]. CD14 is a co-receptor of the Toll-like Receptor (TLR)-4 in the innate immune response and plays a crucial role in its ability to recognize lipopolysaccharide, pathogens and damage-associated molecular patterns on bacterial surfaces thereby promoting an immune-inflammatory response [47]. In a study by Dewhurst, lung tissues from COPD patients were labeled for macrophage subpopulations. The results showed more infiltration of abnormally sized macrophages in the lung tissue of COPD patients

compared to normal lung tissue and confirmed the high expression of the CD14 gene [48].

Finally, there are several limitations of our study. First, based on two datasets obtained from the GEO database, we obtained gene expression profiles of COPD patients. However, all participants tested were male, and the COPD group had a greater age and number of years of smoking [49]. Although Mendelian randomization analysis can be effective in reducing this potential bias, our findings need to be treated with caution in non-smoking or female COPD patients. Second, although the dataset we utilized for single-cell sequencing was able to satisfy the cell sample size used for the analysis [50], there are still limitations in the population sample size. This leads to the possibility of a mismatch in general characteristics in COPD compared to controls [17]. In conclusion, although we identified associations between four key genes and disease genes using Mendelian randomization analysis and explored potential signaling mechanisms, whether these disease genes or signaling mechanisms exist independently or ultimately have an impact on the development of COPD through complex combinations still requires further confirmation.

CONCLUSIONS

In this study, we identified marker genes and major pathways that could be enriched for immune cell subtypes associated with COPD by analyzing scRNA-seq data. Based on Mendelian randomization analysis of eQTLs and GWAS data, the screened genes GZMH, COTL1, CSTA, and CD14 provided evidence for a causal effect on the development of COPD. The association of these four key genes with COPD-causing genes and the potential informational pathways enriched by these genes could provide additional information on the development of COPD.

AUTHOR CONTRIBUTIONS

Data analysis and manuscript writing: Gang Sun, Xiaoxiao Han, Yun Zhou; Study design and statistical advice: Gang Sun, Lewei Huang; Manuscript editing: Xiangqian Che, Shuo Yu, Di Song; Validation and review: Feifei Ma, Xiangqian Che, Shuo Yu, Di Song, Lewei Huang; Quality control: Gang Sun, Lewei Huang; All authors agreed on the journal to which the article was to be submitted and agreed to take responsibility for all aspects of the work.

CONFLICTS OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

FUNDING

No funding was provided for this study.

REFERENCES

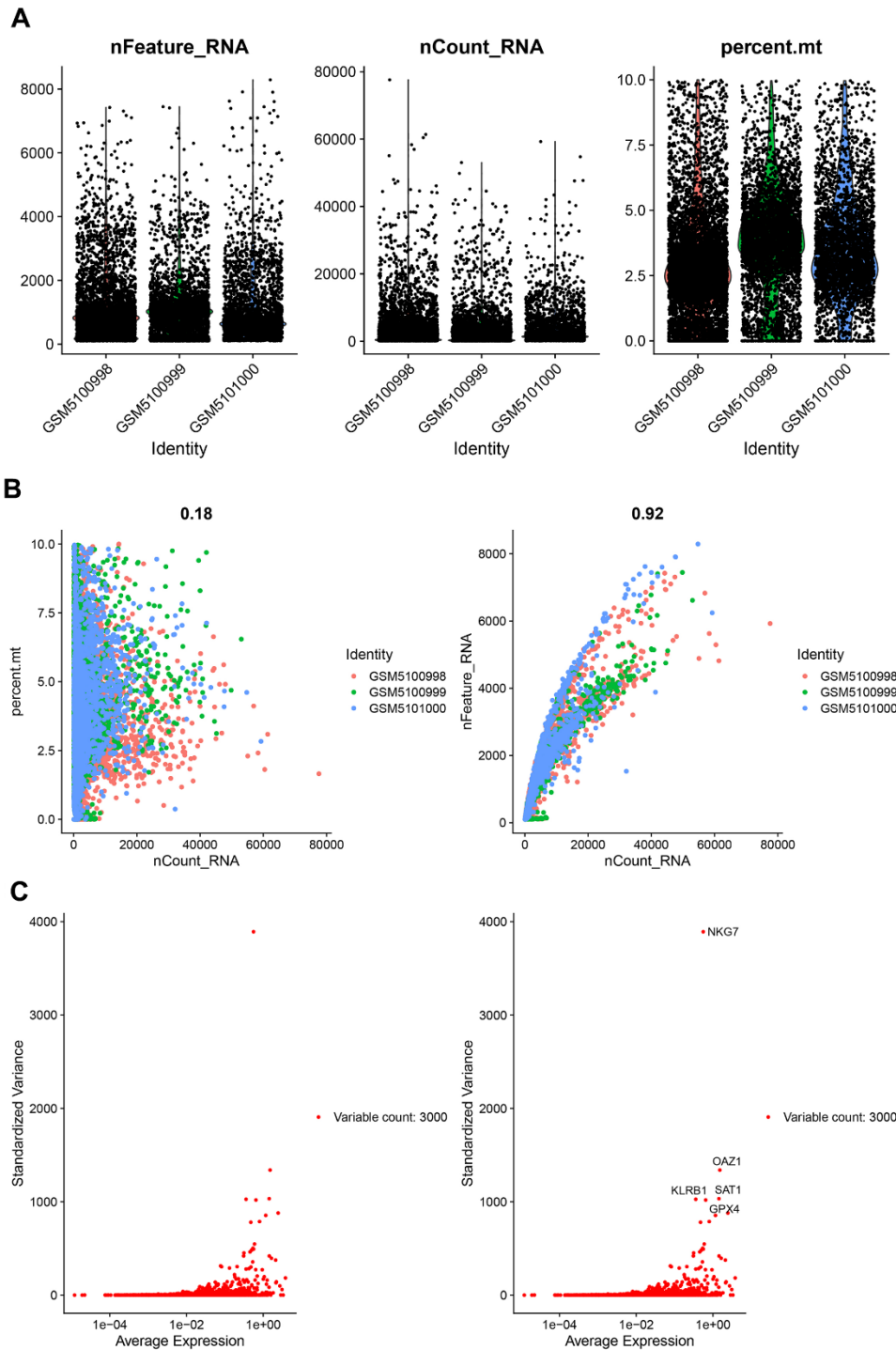
1. Vitenberga Z, Pilmane M, Babjoniševa A. The evaluation of inflammatory, anti-inflammatory and regulatory factors contributing to the pathogenesis of COPD in airways. *Pathol Res Pract*. 2019; 215:97–105.
<https://doi.org/10.1016/j.prp.2018.10.029>
PMID:[30392917](https://pubmed.ncbi.nlm.nih.gov/30392917/)
2. Vauterin D, Van Vaerenbergh F, Vanoverschelde A, Quint JK, Verhamme K, Lahousse L. Methods to assess COPD medications adherence in healthcare databases: a systematic review. *Eur Respir Rev*. 2023; 32:230103.
<https://doi.org/10.1183/16000617.0103-2023>
PMID:[37758274](https://pubmed.ncbi.nlm.nih.gov/37758274/)
3. Sakornsakolpat P, Prokopenko D, Lamontagne M, Reeve NF, Guyatt AL, Jackson VE, Shrine N, Qiao D, Bartz TM, Kim DK, Lee MK, Latourelle JC, Li X, et al, SpiroMeta Consortium, and International COPD Genetics Consortium. Genetic landscape of chronic obstructive pulmonary disease identifies heterogeneous cell-type and phenotype associations. *Nat Genet*. 2019; 51:494–505.
<https://doi.org/10.1038/s41588-018-0342-2>
PMID:[30804561](https://pubmed.ncbi.nlm.nih.gov/30804561/)
4. Bhat TA, Panzica L, Kalathil SG, Thanavala Y. Immune Dysfunction in Patients with Chronic Obstructive Pulmonary Disease. *Ann Am Thorac Soc*. 2015 (Suppl 2); 12:S169–75.
<https://doi.org/10.1513/AnnalsATS.201503-126AW>
PMID:[26595735](https://pubmed.ncbi.nlm.nih.gov/26595735/)
5. Barnes PJ. Alveolar macrophages as orchestrators of COPD. *COPD*. 2004; 1:59–70.
<https://doi.org/10.1081/COPD-120028701>
PMID:[16997739](https://pubmed.ncbi.nlm.nih.gov/16997739/)
6. Brusselle GG, Joos GF, Bracke KR. New insights into the immunology of chronic obstructive pulmonary disease. *Lancet*. 2011; 378:1015–26.
[https://doi.org/10.1016/S0140-6736\(11\)60988-4](https://doi.org/10.1016/S0140-6736(11)60988-4)
PMID:[21907865](https://pubmed.ncbi.nlm.nih.gov/21907865/)
7. Hogg JC. Pathophysiology of airflow limitation in chronic obstructive pulmonary disease. *Lancet*. 2004; 364:709–21.
[https://doi.org/10.1016/S0140-6736\(04\)16900-6](https://doi.org/10.1016/S0140-6736(04)16900-6)
PMID:[15325838](https://pubmed.ncbi.nlm.nih.gov/15325838/)
8. Barnes PJ. Inflammatory mechanisms in patients with chronic obstructive pulmonary disease. *J Allergy Clin Immunol*. 2016; 138:16–27.
<https://doi.org/10.1016/j.jaci.2016.05.011>
PMID:[27373322](https://pubmed.ncbi.nlm.nih.gov/27373322/)
9. Luo X, Tang X. Single-cell RNA sequencing in juvenile idiopathic arthritis. *Genes Dis*. 2023; 11:633–44.
<https://doi.org/10.1016/j.gendis.2023.04.014>
PMID:[37692495](https://pubmed.ncbi.nlm.nih.gov/37692495/)
10. Crespo-García E, Bueno-Costa A, Esteller M. Single-cell analysis of the epitranscriptome: RNA modifications under the microscope. *RNA Biol*. 2024; 21:1–8.
<https://doi.org/10.1080/15476286.2024.2315385>
PMID:[38368619](https://pubmed.ncbi.nlm.nih.gov/38368619/)
11. Trapnell C. Defining cell types and states with single-cell genomics. *Genome Res*. 2015; 25:1491–8.
<https://doi.org/10.1101/gr.190595.115>
PMID:[26430159](https://pubmed.ncbi.nlm.nih.gov/26430159/)
12. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM 3rd, Hao Y, Stoeckius M, Smibert P, Satija R. Comprehensive Integration of Single-Cell Data. *Cell*. 2019; 177:1888–902.e21.
<https://doi.org/10.1016/j.cell.2019.05.031>
PMID:[31178118](https://pubmed.ncbi.nlm.nih.gov/31178118/)
13. Kiselev VY, Andrews TS, Hemberg M. Challenges in unsupervised clustering of single-cell RNA-seq data. *Nat Rev Genet*. 2019; 20:273–82.
<https://doi.org/10.1038/s41576-018-0088-9>
PMID:[30617341](https://pubmed.ncbi.nlm.nih.gov/30617341/)
14. Ji Z, Zhao W, Lin HK, Zhou X. Systematically understanding the immunity leading to CRPC progression. *PLoS Comput Biol*. 2019; 15:e1007344.
<https://doi.org/10.1371/journal.pcbi.1007344>
PMID:[31504033](https://pubmed.ncbi.nlm.nih.gov/31504033/)
15. Xie S, Xie X, Zhao X, Liu F, Wang Y, Ping J, Ji Z. HNSPPI: a hybrid computational model combining network and sequence information for predicting protein-protein interaction. *Brief Bioinform*. 2023; 24:bbad261.
<https://doi.org/10.1093/bib/bbad261>
PMID:[37480553](https://pubmed.ncbi.nlm.nih.gov/37480553/)
16. Kim M, Yang SH, Han HG, Kim E, Kim S, Oh YL, Ro HS. Mitochondrial Effects on the Physiological Characteristics of *Lentinula edodes*. *Mycobiology*. 2022; 50:374–81.
<https://doi.org/10.1080/12298093.2022.2138226>
PMID:[36404899](https://pubmed.ncbi.nlm.nih.gov/36404899/)
17. Li X, Noell G, Tabib T, Gregory AD, Trejo Bittar HE, Vats R, Kaminski TW, Sembrat J, Snyder ME, Chandra D, Chen K, Zou C, Zhang Y, et al. Single cell RNA sequencing identifies IGFBP5 and QKI as ciliated epithelial cell genes associated with severe COPD. *Respir Res*. 2021; 22:100.
<https://doi.org/10.1186/s12931-021-01675-2>
PMID:[33823868](https://pubmed.ncbi.nlm.nih.gov/33823868/)
18. Pei Y, Wei Y, Peng B, Wang M, Xu W, Chen Z, Ke X,

- Rong L. Combining single-cell RNA sequencing of peripheral blood mononuclear cells and exosomal transcriptome to reveal the cellular and genetic profiles in COPD. *Respir Res.* 2022; 23:260. <https://doi.org/10.1186/s12931-022-02182-8> PMID:36127695
19. Zhang A, Wu H, Wang C, Tian S. Integrative analysis of Mendelian randomization and gene expression profiles reveals a null causal relationship between adiponectin and diabetic retinopathy. *Adipocyte.* 2023; 12:2234522. <https://doi.org/10.1080/21623945.2023.2234522> PMID:37459460
 20. Bowden J, Holmes MV. Meta-analysis and Mendelian randomization: A review. *Res Synth Methods.* 2019; 10:486–96. <https://doi.org/10.1002/jrsm.1346> PMID:30861319
 21. Wu Y, Zhang CY, Wang L, Li Y, Xiao X. Genetic Insights of Schizophrenia via Single Cell RNA-Sequencing Analyses. *Schizophr Bull.* 2023; 49:914–22. <https://doi.org/10.1093/schbul/sbad002> PMID:36805283
 22. Zhang YH, Cho MH, Morrow JD, Castaldi PJ, Hersh CP, Midha MK, Hoopmann MR, Lutz SM, Moritz RL, Silverman EK. Integrating Genetics, Transcriptomics, and Proteomics in Lung Tissue to Investigate Chronic Obstructive Pulmonary Disease. *Am J Respir Cell Mol Biol.* 2023; 68:651–63. <https://doi.org/10.1165/rcmb.2022-0302OC> PMID:36780661
 23. Saferali A, Yun JH, Parker MM, Sakornsakolpat P, Chase RP, Lamb A, Hobbs BD, Boezen MH, Dai X, de Jong K, Beaty TH, Wei W, Zhou X, et al, COPD Gene Investigators, and International COPD Genetics Consortium Investigators. Analysis of genetically driven alternative splicing identifies FBXO38 as a novel COPD susceptibility gene. *PLoS Genet.* 2019; 15:e1008229. <https://doi.org/10.1371/journal.pgen.1008229> PMID:31269066
 24. Vösa U, Claringbould A, Westra HJ, Bonder MJ, Deelen P, Zeng B, Kirsten H, Saha A, Kreuzhuber R, Yazar S, Brugge H, Oelen R, de Vries DH, et al, BIOS Consortium, and i2QTL Consortium. Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat Genet.* 2021; 53:1300–10. <https://doi.org/10.1038/s41588-021-00913-z> PMID:34475573
 25. Song J, Ren K, Zhang D, Lv X, Sun L, Deng Y, Zhu H. A novel signature combining cuproptosis- and ferroptosis-related genes in sepsis-induced cardiomyopathy. *Front Genet.* 2023; 14:1170737. <https://doi.org/10.3389/fgene.2023.1170737> PMID:37035738
 26. Pillai SG, Ge D, Zhu G, Kong X, Shianna KV, Need AC, Feng S, Hersh CP, Bakke P, Gulsvik A, Ruppert A, Lødrup Carlsen KC, Roses A, et al, and ICGN Investigators. A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. *PLoS Genet.* 2009; 5:e1000421. <https://doi.org/10.1371/journal.pgen.1000421> PMID:19300482
 27. Cho MH, Boutaoui N, Klanderman BJ, Sylvia JS, Ziniti JP, Hersh CP, DeMeo DL, Hunninghake GM, Litonjua AA, Sparrow D, Lange C, Won S, Murphy JR, et al. Variants in FAM13A are associated with chronic obstructive pulmonary disease. *Nat Genet.* 2010; 42:200–2. <https://doi.org/10.1038/ng.535> PMID:20173748
 28. Wilk JB, Shrine NR, Loehr LR, Zhao JH, Manichaikul A, Lopez LM, Smith AV, Heckbert SR, Smolonska J, Tang W, Loth DW, Curjuric I, Hui J, et al. Genome-wide association studies identify CHRNA5/3 and HTR4 in the development of airflow obstruction. *Am J Respir Crit Care Med.* 2012; 186:622–32. <https://doi.org/10.1164/rccm.201202-0366OC> PMID:22837378
 29. Joglekar MM, Nizamoglu M, Fan Y, Nemani SSP, Weckmann M, Pouwels SD, Heijink IH, Melgert BN, Pillay J, Burgess JK. Highway to heal: Influence of altered extracellular matrix on infiltrating immune cells during acute and chronic lung diseases. *Front Pharmacol.* 2022; 13:995051. <https://doi.org/10.3389/fphar.2022.995051> PMID:36408219
 30. Tan DBA, Teo TH, Setiawan AM, Ong NE, Zimmermann M, Price P, Kirkham LS, Moodley YP. Increased CTLA-4+ T cells may contribute to impaired T helper type 1 immune responses in patients with chronic obstructive pulmonary disease. *Immunology.* 2017; 151:219–26. <https://doi.org/10.1111/imm.12725> PMID:28190271
 31. Belchamber KB, Singh R, Batista CM, Whyte MK, Dockrell DH, Kilty I, Robinson MJ, Wedzicha JA, Barnes PJ, Donnelly LE, and COPD-MAP consortium. Defective bacterial phagocytosis is associated with dysfunctional mitochondria in COPD macrophages. *Eur Respir J.* 2019; 54:1802244. <https://doi.org/10.1183/13993003.02244-2018> PMID:31320451
 32. Costa C, Traves SL, Tudhope SJ, Fenwick PS, Belchamber KB, Russell RE, Barnes PJ, Donnelly LE. Enhanced monocyte migration to CXCR3 and CCR5

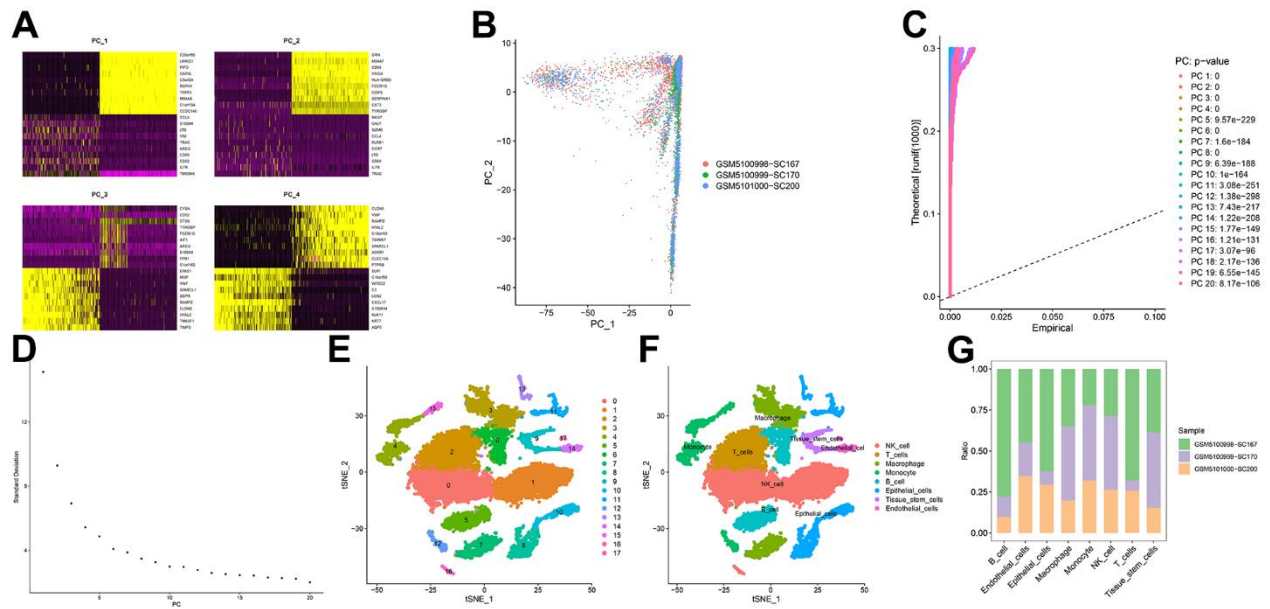
- chemokines in COPD. *Eur Respir J.* 2016; 47:1093–102.
<https://doi.org/10.1183/13993003.01642-2015>
PMID:26965295
33. Moldoveanu B, Otmishi P, Jani P, Walker J, Sarmiento X, Guardiola J, Saad M, Yu J. Inflammatory mechanisms in the lung. *J Inflamm Res.* 2009; 2:1–11.
PMID:22096348
34. Ngan DA, Vickerman SV, Granville DJ, Man SF, Sin DD. The possible role of granzyme B in the pathogenesis of chronic obstructive pulmonary disease. *Ther Adv Respir Dis.* 2009; 3:113–29.
<https://doi.org/10.1177/1753465809341965>
PMID:19638369
35. Lanier LL. Up on the tightrope: natural killer cell activation and inhibition. *Nat Immunol.* 2008; 9:495–502.
<https://doi.org/10.1038/ni1581> PMID:18425106
36. Kim WD, Chi HS, Choe KH, Kim WS, Hogg JC, Sin DD. The Role of Granzyme B Containing Cells in the Progression of Chronic Obstructive Pulmonary Disease. *Tuberc Respir Dis (Seoul).* 2020 (Supple 1); 83:S25–33.
<https://doi.org/10.4046/trd.2020.0089>
PMID:33076634
37. Takiguchi H, Chen V, Obeidat M, Hollander Z, FitzGerald JM, McManus BM, Ng RT, Sin DD. Effect of short-term oral prednisone therapy on blood gene expression: a randomised controlled clinical trial. *Respir Res.* 2019; 20:176.
<https://doi.org/10.1186/s12931-019-1147-2>
PMID:31382977
38. Kolossváry M, deFilippi C, McCallum S, Fitch KV, Diggs MR, Fulda ES, Ribaud HJ, Fichtenbaum CJ, Aberg JA, Malvestutto CD, Currier JS, Casado JL, Gutiérrez F, et al. Identification of pre-infection markers and differential plasma protein expression following SARS-CoV-2 infection in people living with HIV. *EBioMedicine.* 2023; 90:10453.
<https://doi.org/10.1016/j.ebiom.2023.104538>
PMID:36966617
39. Kumar N, So, Mazumder M, Dutta P, Maiti S, Gourinath S. EhCoactosin stabilizes actin filaments in the protist parasite *Entamoeba histolytica*. *PLoS Pathog.* 2014; 10:e1004362.
<https://doi.org/10.1371/journal.ppat.1004362>
PMID:25210743
40. Wang B, Zhao L, Chen D. Coactosin-Like Protein in Breast Carcinoma: Friend or Foe? *J Inflamm Res.* 2022; 15:4013–25.
<https://doi.org/10.2147/JIR.S362606>
PMID:35873386
41. Barnes PJ, Shapiro SD, Pauwels RA. Chronic obstructive pulmonary disease: molecular and cellular mechanisms. *Eur Respir J.* 2003; 22:672–88.
<https://doi.org/10.1183/09031936.03.00040703>
PMID:14582923
42. Martinez FO, Sica A, Mantovani A, Locati M. Macrophage activation and polarization. *Front Biosci.* 2008; 13:453–61.
<https://doi.org/10.2741/2692> PMID:17981560
43. Kim RY, Oliver B. Innate Immune Reprogramming in Chronic Obstructive Pulmonary Disease: New Mechanisms for Old Questions. *Am J Respir Cell Mol Biol.* 2023; 68:470–1.
<https://doi.org/10.1165/rcmb.2023-0030ED>
PMID:36796089
44. Barrett AJ. The cystatins: a diverse superfamily of cysteine peptidase inhibitors. *Biomed Biochim Acta.* 1986; 45:1363–74.
PMID:3555466
45. Butler MW, Fukui T, Salit J, Shaykhiev R, Mezey JG, Hackett NR, Crystal RG. Modulation of cystatin A expression in human airway epithelium related to genotype, smoking, COPD, and lung cancer. *Cancer Res.* 2011; 71:2572–81.
<https://doi.org/10.1158/0008-5472.CAN-10-2046>
PMID:21325429
46. Mlinaric-Rascan I, Asa SL, Siminovitich KA. Increased expression of the stefin A cysteine proteinase inhibitor occurs in the myelomonocytic cell-infiltrated tissues of autoimmune motheaten mice. *Am J Pathol.* 1994; 145:902–12.
PMID:7943179
47. Na K, Oh BC, Jung Y. Multifaceted role of CD14 in innate immunity and tissue homeostasis. *Cytokine Growth Factor Rev.* 2023; 74:100–7.
<https://doi.org/10.1016/j.cytogfr.2023.08.008>
PMID:37661484
48. Dewhurst JA, Lea S, Hardaker E, Dungwa JV, Ravi AK, Singh D. Characterisation of lung macrophage subpopulations in COPD patients and controls. *Sci Rep.* 2017; 7:7143.
<https://doi.org/10.1038/s41598-017-07101-2>
PMID:28769058
49. Kim WJ, Lim JH, Lee JS, Lee SD, Kim JH, Oh YM. Comprehensive Analysis of Transcriptome Sequencing Data in the Lung Tissues of COPD Subjects. *Int J Genomics.* 2015; 2015:206937.
<https://doi.org/10.1155/2015/206937> PMID:25834810
50. Morse C, Tabib T, Sembrat J, Buschur KL, Bittar HT, Valenzi E, Jiang Y, Kass DJ, Gibson K, Chen W, Mora A, Benos PV, Rojas M, Lafyatis R. Proliferating SPP1/MERTK-expressing macrophages in idiopathic pulmonary fibrosis. *Eur Respir J.* 2019; 54:1802441.
<https://doi.org/10.1183/13993003.02441-2018>
PMID:31221805

SUPPLEMENTARY MATERIALS

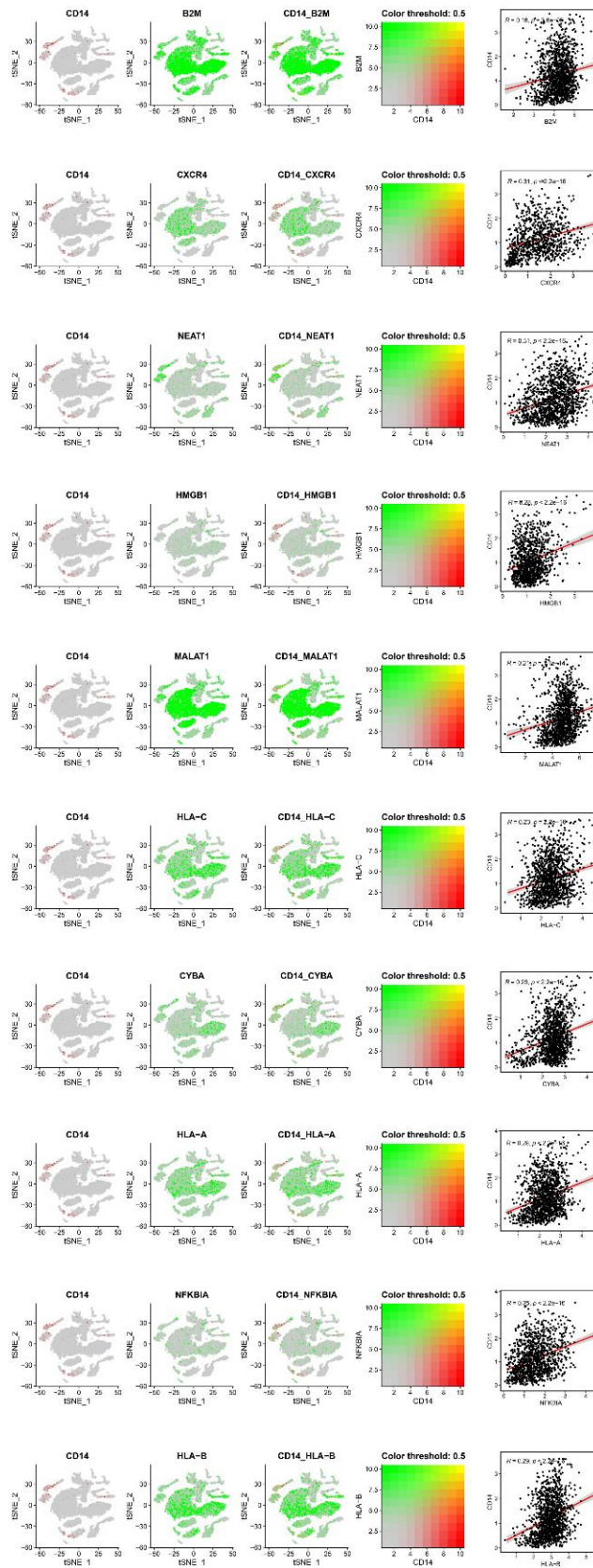
Supplementary Figures



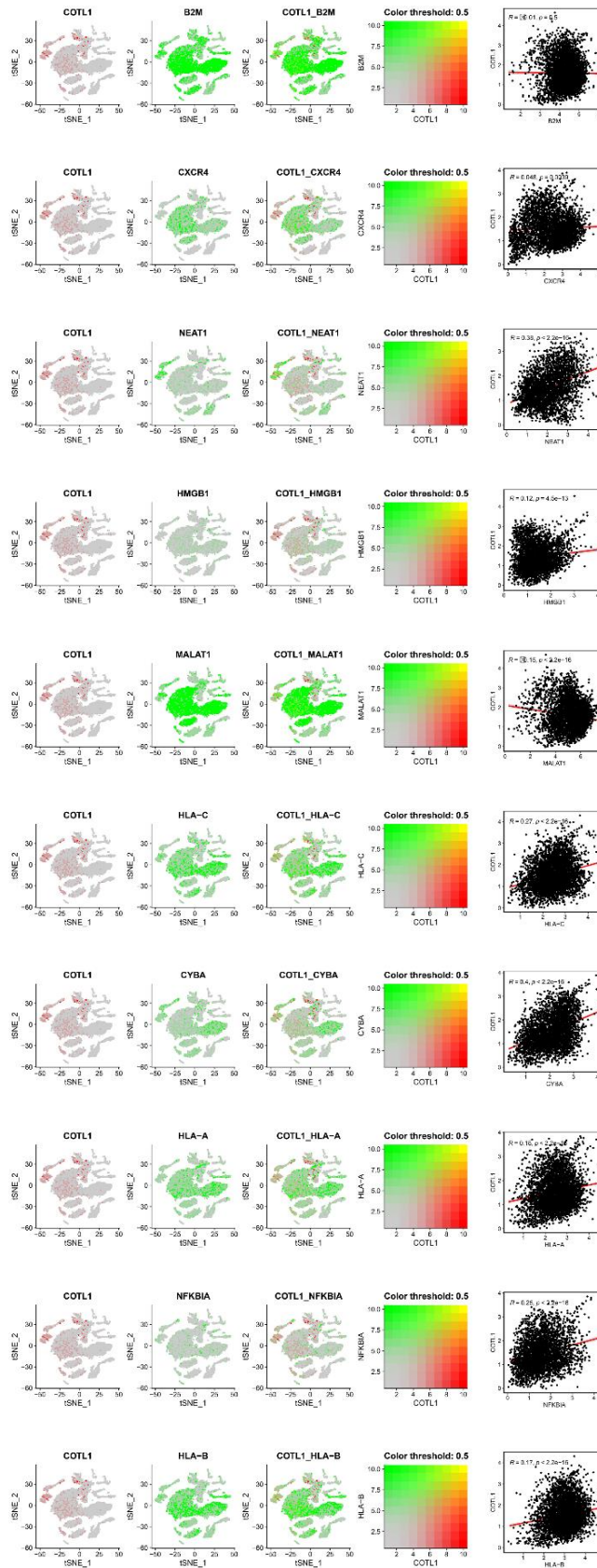
Supplementary Figure 1. Analysis of single-cell sequencing results from COPD patients. (A) Quality inspection of single-cell sequencing. Plots from left to right showed the number of cells, the number of genes and the sequencing depth for each sample, respectively. (B) Plots from left to right showed the relationship between single-cell sequencing depth and mitochondrial content (correlation coefficient of 0.18) and the number of genes (correlation coefficient of 0.92), respectively. (C) Genes that differed between cells, and the 5 genes with the highest standardized variance were flagged.



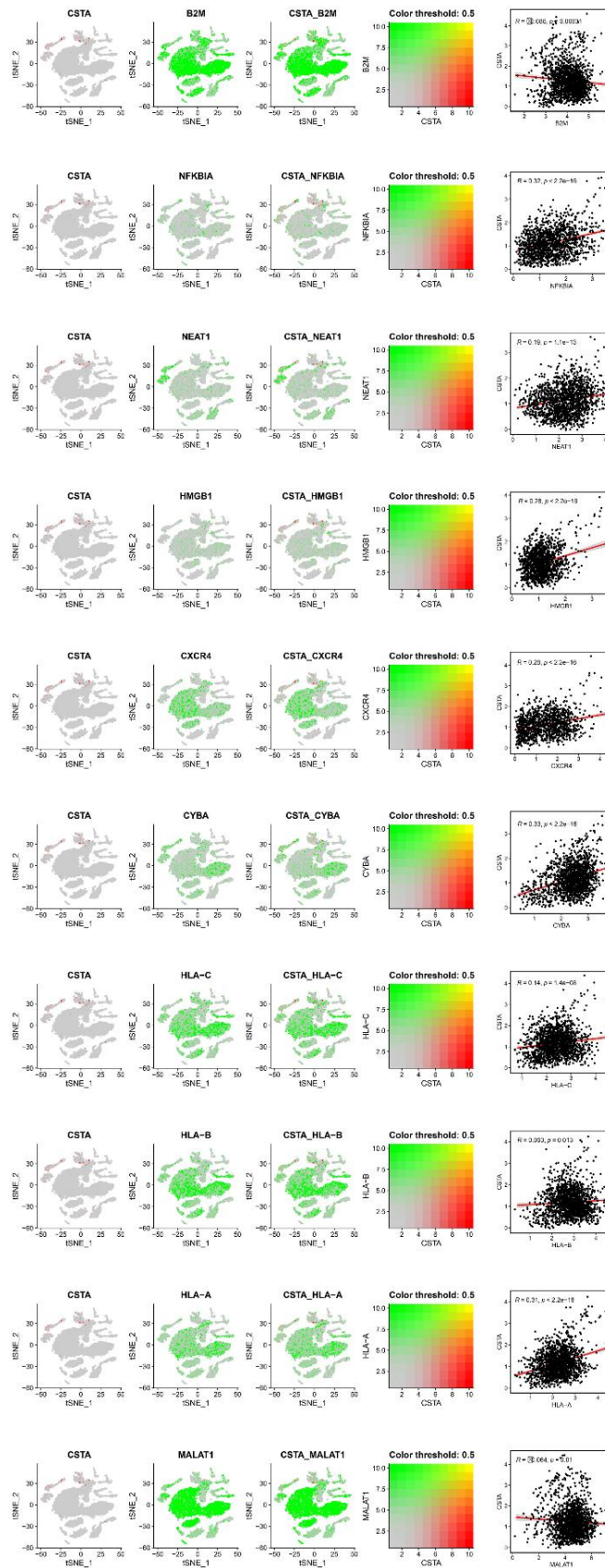
Supplementary Figure 2. Annotation of cellular subtypes of single-cell samples. (A) Heatmap of principal component analysis for genes from single-cell sequencing in COPD samples. (B) Distribution of PCs in COPD single-cell sequencing samples, where dots represent cells and different colors represent different samples. (C) Histogram of PCA results of COPD single-cell sequencing samples to show the P-value for each PC. (D) Variogram of each PC in COPD single-cell sequencing samples. (E) Cluster analysis of COPD single-cell sequencing samples. (F) Annotation of cell subtypes in COPD single-cell sequencing samples. (G) Difference in the percentage of cells in COPD samples that were annotated.



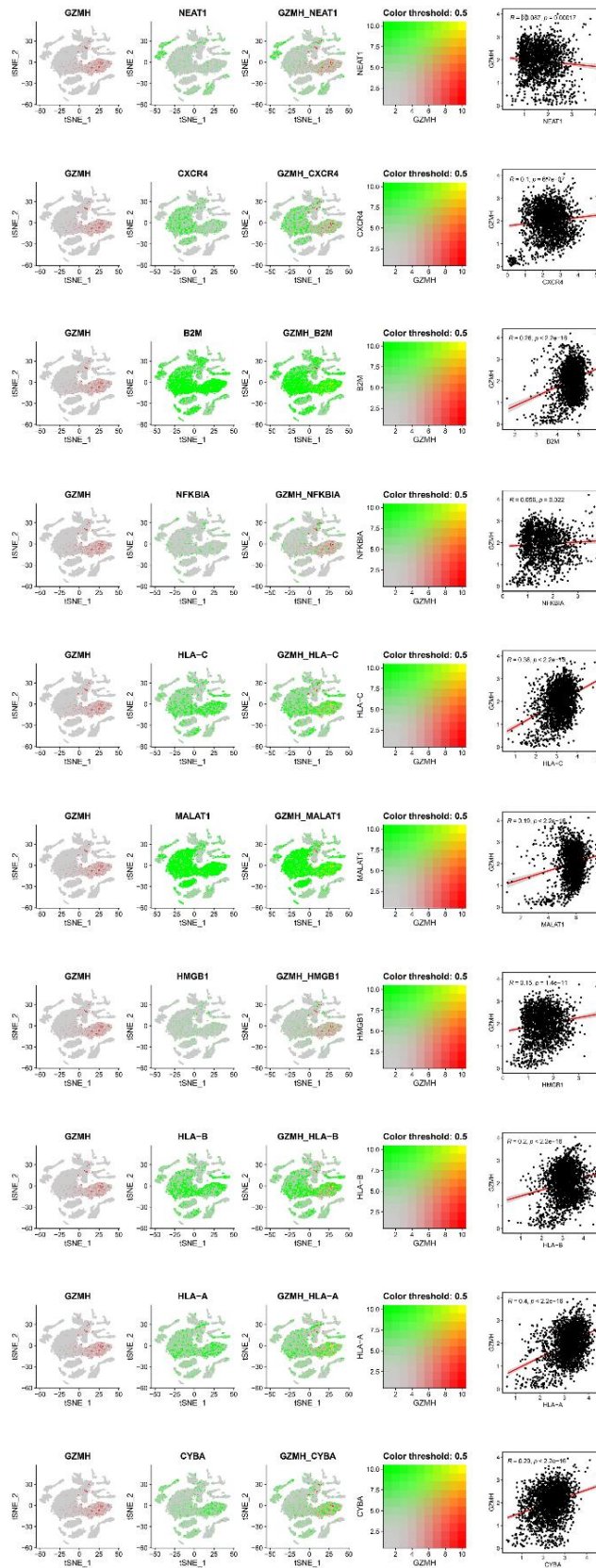
Supplementary Figure 3. Co-expression analysis of CD14 with immune genes in single-cell sequencing samples. Scatter plots represent Pearson correlation analysis between co-expressed genes.



Supplementary Figure 4. Co-expression analysis of COTL1 with immune genes in single-cell sequencing samples. Scatter plots indicate the person correlation analysis between co-expressed genes.



Supplementary Figure 5. Co-expression analysis of CSTA with immune genes in single-cell sequencing samples. Scatter plots indicate the Pearson correlation analysis between co-expressed genes.



Supplementary Figure 6. Co-expression analysis of GZMH with immune genes in single-cell sequencing samples. Scatter plots indicate the Pearson correlation analysis between co-expressed genes.

Supplementary Table

Please browse Full Text version to see the data of Supplementary Tables 1–4.

Supplementary Table 1. 411 cellular subtype marker genes extracted from single-cell expression profiles.

Supplementary Table 2. 120 pairs of marker genes causally associated with COPD.

Supplementary Table 3. Results of Mendelian randomization analysis of four key marker genes associated with COPD.

Supplementary Table 4. Results of co-localization analysis of four genes at the eQI-GWAS level.