BIOCELL
2020 44(4): 559-582

Tech Science Press

Expression profiling of immune cells in systemic lupus erythematosus by single-cell RNA sequencing

Xianliang HOU[1,2,3,#]; Donge TANG[1,#]; Fengping ZHENG[1,#]; Minglin OU[3]; Yong XU[1]; Huixuan XU[1]; Xiaoping HONG[4]; Xinzhou ZHANG[1]; Weier DAI[5]; Dongzhou LIU[4,*]; Yong DAI[1,*]

[1] Clinical Medical Research Center, The Second Clinical Medical College, Jinan University (Shenzhen People's Hospital), Shenzhen, 518020, China
[2] The First Affiliated Hospital, Jinan University, Guangzhou, 510632, China
[3] Central Laboratory, The Second Affiliated Hospital of Guilin Medical University, Guilin, 541199, China
[4] Department of Rheumatology and Immunology, the Second Clinical Medical College of Jinan University (Shenzhen People's Hospital), Shenzhen, 518020, China
[5] College of Natural Science, University of Texas at Austin, Austin, 78712, USA

Key words: B cell, Differential expression genes, Gene ontologies, Monocytes, Systemic lupus erythematosus, Single-cell RNA-sequencing

Abstract: Systemic lupus erythematosus (SLE) is a systemic autoimmune disease characterized by abnormal cellular and humoral immune responses and excessive autoantibody production. The precise pathologic mechanism of SLE remains elusive. The advent of single-cell RNA sequencing (scRNA-seq) enables unbiased analysis of the molecular differences of cell populations at the single-cell level. We used scRNA-seq to profile the transcriptomes of peripheral blood mononuclear cells from an SLE patient compared with a healthy control (HC). A total of 16,021 cells were analyzed and partitioned into 12 distinct clusters. The marker genes of each cluster and the four major immune cell types (B cells, $CD4^+$ T cells, $CD8^+$ T cells, myeloid cells, and NK cells) were determined. Moreover, several genes involved in antigen processing and presentation through MHCII were highly enriched. GO enrichment analyses revealed abnormal gene expression patterns and signaling pathways in SLE. Of note, pseudotime analysis revealed that there was a different lineage hierarchy in the peripheral blood mononuclear cells (PBMCs) of the SLE patient, indicating that the cell states were substantially altered under disease conditions. Our analysis provides a comprehensive map of the cell types and states of the PBMCs of SLE patients at the single-cell level for a better understanding of the pathogenesis, diagnosis, and treatment of SLE.

Introduction

Systemic lupus erythematosus (SLE) is a complex autoimmune disorder that mainly afflicts women of child-bearing age. This disease is characterized by the abundant production of autoantibodies to components of the cell nucleus (Maria and Davidson, 2018; Sui *et al.*, 2013). Although the precise pathologic mechanism of SLE is not yet fully understood, susceptibility genes, hormones, immunomodulatory factors, and environmental risk factors have been reported to be important for the development of the disease. Many abnormalities of the immune system leading to loss of immune tolerance have been reported in SLE (Sui *et al.*, 2013). Dysregulated immune activation and responses of both T and B cells and the abnormal production of proinflammatory cytokines play a crucial role in the occurrence and progression of tissue pathology and organ injury in SLE. The molecular basis of the abnormal immune response in SLE remains elusive (Sui *et al.*, 2015). Identification of the underlying genetic and biochemical mechanisms will enhance our understanding of the pathogenesis of SLE as well as identify novel targets for future pharmacological intervention.

Advances in sequencing technology now permit interrogation of complex sequencing targets at a reasonable cost and unprecedented depth. Bulk cell RNA-sequencing (RNA-seq) technology is widely used in transcriptome profiling to study gene and transcript expression levels, transcriptional structures, and splicing patterns (Li and Li, 2018). However, conventional RNA-seq methods process millions of cells, and cellular heterogeneity cannot be addressed since signals of variably expressed genes would be averaged across cells. The understanding of biological

*Address correspondence to: Yong Dai, daiyong22@aliyun.com; Dongzhou Liu, liu_dz2001@sina.com
#These authors contributed equally to this work
Received: 18 April 2020; Accepted: 24 August 2020

Doi: 10.32604/biocell.2020.011022

systems requires knowledge of their individual components. Fortunately, single-cell RNA sequencing (scRNA-seq) technologies are now emerging as a powerful technology for unbiased analysis of cell types and cell states based on gene activity (Rheaume *et al.*, 2018; Zeisel *et al.*, 2018). scRNA-seq allow unbiased determination of cellular heterogeneity, the discovery of previously unrecognized disease-associated cell populations or functional states, and identification of novel biomarkers and potential molecular regulators (Cochain *et al.*, 2018). Here, we used systematic scRNA-seq to generate a molecular census of cell types and states within peripheral blood mononuclear cells (PBMCs) from an SLE patient compared with a healthy volunteer. Analysis of immune cells by scRNA-seq identified B cells, $CD4^+$ T cells, $CD8^+$ T cells, myeloid cells, and NK cells. We combined this technology with clustering analysis to resolve the differential responses by cell type and characterize gene expression at the single-cell level to identify abnormal cellular subpopulations and disordered signal pathways involved in the disease initiation, maintenance, and progression of SLE.

Materials and Methods

Study group

An SLE patient and a healthy individual of the same gender and age (female, 34-year-old, Chinese) were enrolled in the study. SLE was confirmed by pathologic diagnosis and clinical evidence. SLE diagnosis was according to the 2012 Systemic Lupus International Collaborating Clinics (SLICC) classification criteria (Hartman *et al.*, 2018). The clinical characteristics of the patient were as follows: Acute cutaneous lupus erythematosus is present in the face and arms; anti-dsDNA antibodies and antinuclear antibody were positive; 24-h urine protein was 2.4 g/day; the ESR value was 77; the patient underwent renal biopsy for diagnostic evaluation, and identified as lupus nephritis (WHO class III nephritis). Twenty milliliters of whole blood samples were collected from each subject. Peripheral blood mononuclear cells (PBMCs) were isolated from whole blood using density gradient centrifugation with Ficoll-Hypaque. *This study was conducted in accordance with the tenets of the Declaration of Helsinki and was approved by the Ethics Committee of the Shenzhen People's Hospital, China (ref. no. 2015-313). All participating individuals provided written informed consent.*

Single-cell capture, library preparation, and RNA-seq

The PBMC suspension was diluted to a concentration of ~1×10^6 cells/mL in PBS plus 0.04% bovine serum albumin (BSA). The percent PBMC viability was more than 80%. PBMC count and viability were determined using trypan blue on a Countess FL II system. Then, single-cell RNA-seq (scRNA-seq) libraries were prepared following the Single Cell 3' Reagent Kit User Guide v2 (10*X* Genomics) (Xie *et al.*, 2018). Briefly, cellular suspensions were loaded on a Chromium Controller instrument (10*X* Genomics) to generate single-cell gel bead-in-emulsions (GEMs). GEM-reverse transcription (GEM-RT) reactions were performed in a 96-deep well reaction module: 55°C for 45 min, 85°C for 5 min, and hold at 4°C. After the RT step, the GEMs were broken, and barcoded cDNA was purified with DynaBeads MyOne Silane Beads (Thermo Fisher Scientific, 37002D). Subsequently, cDNAs were amplified with a 96-Deep Well Reaction Module (98°C for 3 min; cycled 12 times: 98°C for 15 s, 67°C for 20 s, and 72°C for 1 min; 72°C for 1 min; hold at 4°C) and cleaned up with the SPRIselect Reagent Kit (Beckman Coulter) (Scott *et al.*, 2018). Indexed sequencing libraries were constructed using the GemCode Single-Cell 3' Library Kit for enzymatic fragmentation, end repair, A-tailing, adaptor ligation, ligation cleanup, sample index PCR, and PCR cleanup (Xie *et al.*, 2018). Quantification of the constructed libraries was evaluated using the QubitdsDNA HS Assay Kit (Thermo Fisher), Agilent cDNA High Sensitivity Kit, and Kapa DNA Quantification Kit for Illumina platforms, according to the manufacturer's instruction. The library was sequenced on an Illumina HiSeq2500 using the paired-end 2 × 125 bp sequencing protocol. Sequencing run parameters were set up according to version 2 chemistry, and the number of cycles for each read was as follows: Read 1: 26 cycles, i7 index: 8 cycles, i5 index: 0 cycles, and Read 2: 98 cycles (D'Avola *et al.*, 2018).

Processing of scRNAseq data

Single-cell expression was analyzed using CellRanger software (version 2.1.1) to perform quality control, sample demultiplexing, barcode processing, and single-cell 3' gene counting as previously described (Zheng *et al.*, 2017). Sequencing reads were aligned to the UCSC hg38 transcriptome using the Cell Ranger suite with default parameters. As a quality control (QC) step, we first filtered out genes detected in less than three cells and cells where < 200 genes had nonzero counts. In addition, cells with a percentage of total unique molecular identifiers (UMIs) derived from the mitochondrial genome greater than 10% were removed since increased detection of mitochondrial genes can be associated with cells undergoing stress and cell death (Ilicic *et al.*, 2016). The remaining data were normalized and log-transformed, and the log-transformed matrix was used for all downstream analyses. Based on the Louvain modularity optimization algorithm, the cluster identities were defined by K-means clustering in Cell Ranger. In addition, we adopted the DDRTree method for pseudotime analysis. Briefly, genes with the top 1,500 highest standard deviations were obtained as highly variable genes. Principal component analysis (PCA) was used for dimensionality reduction using highly variable genes. Using the first ten principal components as input, we then used t-distributed stochastic neighbor embedding (t-SNE) for data visualization in two dimensions.

Cluster cell type annotation and marker gene identification

We used the feature plot function to highlight the expression of known marker genes to identify which clusters belonged to which immune cell type. Markers used to type cells included CD4, CD3D, CD3E, and CD3G ($CD4^+$ T cells); CD8A, CD8B, CD3D, CD3E, and CD3G ($CD8^+$ T cells); CD19, MS4A1, CD79A, and CD79B (B cells); and S100A8, S100A9, and LYZ (myeloid cells); CD14, CD68, and CSF1R (monocytes cell); CD4, FOXP3, IL2RA (Regulatory T cell); IL17RA, IL17RC, IL17RE (Th17 cell); NCAM1 (NK cell); FUT4 (neutrophils). Then, to identify marker genes for each of these 12 subclusters within these 5 major cell types, we used the Seurat

FindMarkers function to contrast cells from that subcluster to all other cells of that subcluster (Lambrechts *et al.*, 2018). Marker genes were required to have an average expression in that subcluster that was >2.0-fold higher than the average expression in the other subclusters from that cell type and a p-value <0.05. Violin plots for the given genes were generated using the Seurat toolkit VlnPlot function (Burl *et al.*, 2018). In addition, we used the PANTHER Overrepresentation Test to assess the Gene Ontology (GO) of DEGs (http://www.geneontology.org/), with p-values corrected by the Bonferroni adjustment for multiple comparisons.

Statistical analysis

Assessment of statistical significance was performed using unpaired t-tests or Mann–Whitney U-tests where appropriate. Statistical analyses were performed using SPSS20, and a two-tailed p-value less than 0.05 was considered significant. P values are indicated as follows: $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$; $^{****}p < 0.0001$; not significant (ns) $p > 0.05$.

Results

To generate a deep transcriptional map of the immune cell states and gene expression in the SLE patient at the single-cell level, we performed scRNA-seq analysis on total PBMCs from an SLE patient compared with an HC individual. After the aapplication of quality control filters, 8,030 cells from the SLE patient and 7,991 cells from the HC were included in the scRNA-seq analysis. The median numbers of UMIs and genes detected per cell in the SLE patient were 3,137 and 936, respectively. In the HC, we detected a median of 3,138 UMI counts and 825 genes per cell. Sequencing saturation was 84.0% and 86.7% for the SLE and HC groups, respectively, indicating a comprehensive sampling of the available transcripts. The number of detected UMIs and genes and the percentage of UMIs derived from mitochondrial and ribosomal genomes in each cell are displayed in Supplementary Figs. 1 and 2.

Single-cell profiling and unbiased clustering of PBMCs

To construct a global atlas of immune cells, we merged data across the SLE patient and the healthy controls. We then reduced the dimensionality of the expression matrix and placed individual cells into two-dimensional space using tSNE, revealing a diverse set of 12 clusters (Fig. 1A). The percentage of each cluster in the SLE and HC groups is displayed in Supplementary Tab. 1. Based on the expression of known markers for each cell type, we designated these

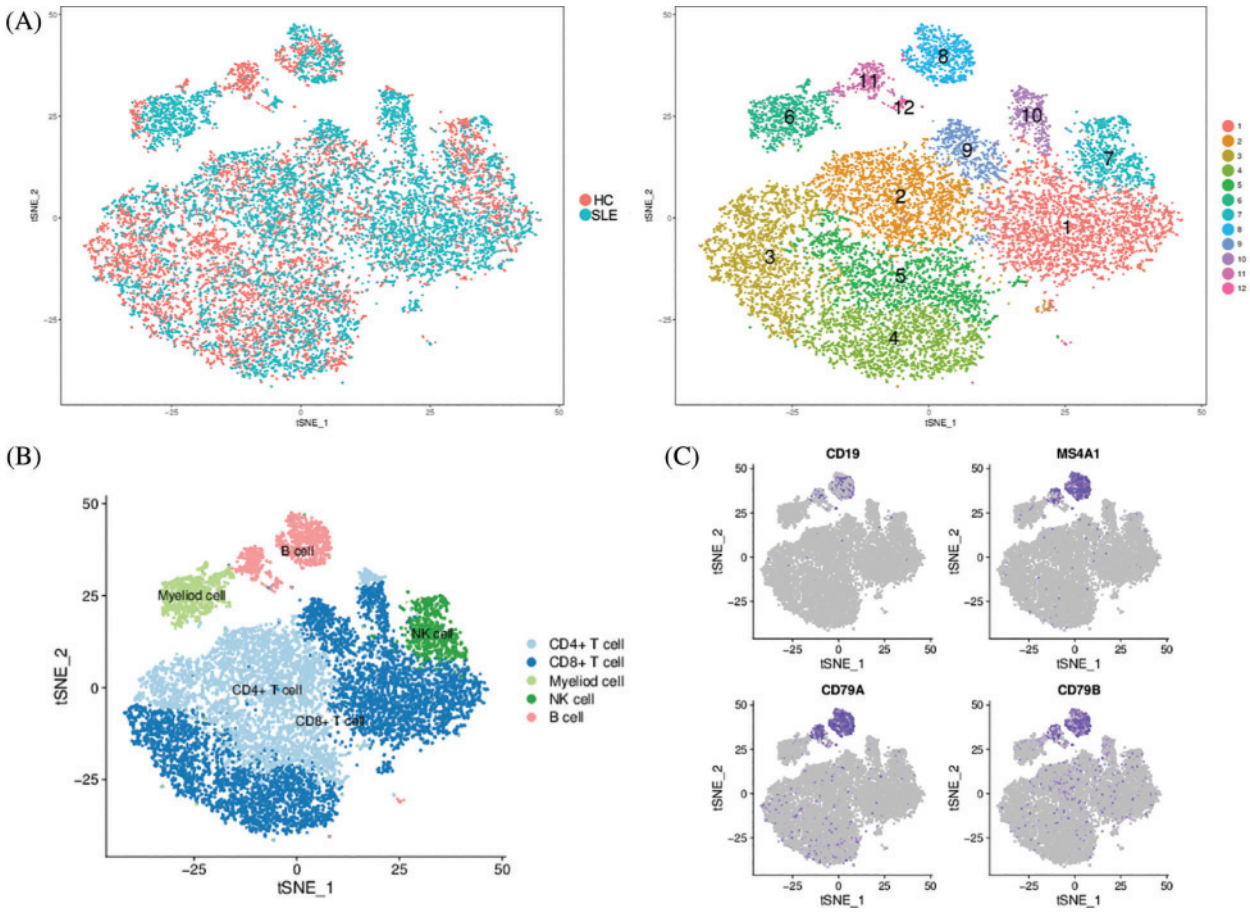


FIGURE 1. Unbiased characterization of the immune system across SLE patients and healthy controls.
(A) tSNE plot of 16,021 PBMCs (8,030 cells from the SLE patients and 7,991 cells from the HCs) passing quality control, color-coded by the sample type of origin (left) and their associated cluster (right). Each point represents a cell, and each cell is grouped into one of the 12 clusters. The complete list of DEGs that define these clusters is presented in Supplementary Tab. 2. (B) Five major immune cell types (B cells, $CD4^{+}$ T cells, $CD8^{+}$ T cells, myeloid cells, and NK cells) were identified by the expression of known markers for each cell type. (C) Feature plots of canonical markers for B cells are shown, and the gradient of purple reflects expression levels.

clusters as B cells, $CD4^+$ T cells, $CD8^+$ T cells, myeloid cells, NK cells, and monocytes cells (Figs. 1B, 1C), Supplementary Figs. 3 and 4). In addition, $CD4^+$ T cells were further subdivided into 10 clusters (Supplementary Fig. 5), and myeloid cells were further subdivided into 5 clusters (Supplementary Fig. 6). Moreover, a few scattered, non-aggregated cells were also identified, which include regulatory T cells (Supplementary Fig. 4B), Th17 cells (Supplementary Fig. 7), and neutrophils (Supplementary Fig. 8).

Major immune cell type-specific markers

To further differentiate these major immune cell types (B cells, $CD4^+$ T cells, $CD8^+$ T cells, myeloid cells, and NK cells), we then analyzed differentially expressed genes between cell types based on mean expression and covariance patterns to discover marker gene sets sufficient to uniquely identify cell types with high probability (Supplementary Tab. 2). A heatmap of the top 20 genes that were unique to each cluster based on the average log fold-change showed a high degree of heterogeneity between the clusters (Fig. 2). B cells are distinguished by harboring a unique set of significant genes, including IGLC2, IGKC, HBB, IGLC3, and HBA1, etc. A set of strong, unique signature genes, including LTB, IL7R, PPBP, MAL, and TRAT1, were found to be highly and specifically expressed in $CD4^+$ T cells from both the SLE and HC groups. CCL5, GZMH, CD8B, NKG7, CD8A, etc., were the most abundant genes expressed in $CD8^+$ T cells. GNLY, CMC1, PRF1, XCL2, and CLIC3 were significantly expressed in NK cells. LYZ, S100A9, S100A8, CST3, and FCN1 were significantly expressed in myeloid cells. In addition, the top 20 most distinct signature genes in each cell type were presented as violin plots and compared for the differences between the SLE and HC groups (Fig. 3, Supplementary Figs. 9–17).

Gene expression differences between SLE and HC groups in each immune cell type

The scRNA-seq analysis allowed us to perform a detailed comparison of the gene expression patterns between the SLE and HC groups in each immune cell type. Based on the selection criteria, fold change (FC) ≥ 2 and p-value < 0.05, we identified 93, 81, 148, 156, and 93 differentially expressed genes (DEGs) between the SLE and HC groups in B cells, $CD4^+$ T cells, $CD8^+$ T cells, NK cells, and myeloid cells, respectively. Among these DEGs, DNAJB1, MARCKSL1, RPS26, TCL1A, and HSPA1B were the genes with the strongest downregulation, and HBB, HBA2, HBA1, S100A8, and IGHG3 were the genes with the strongest upregulation in the B cells of the SLE patient. SOX4, MARCKSL1, KLF2, DNAJB1, and FHIT were the top

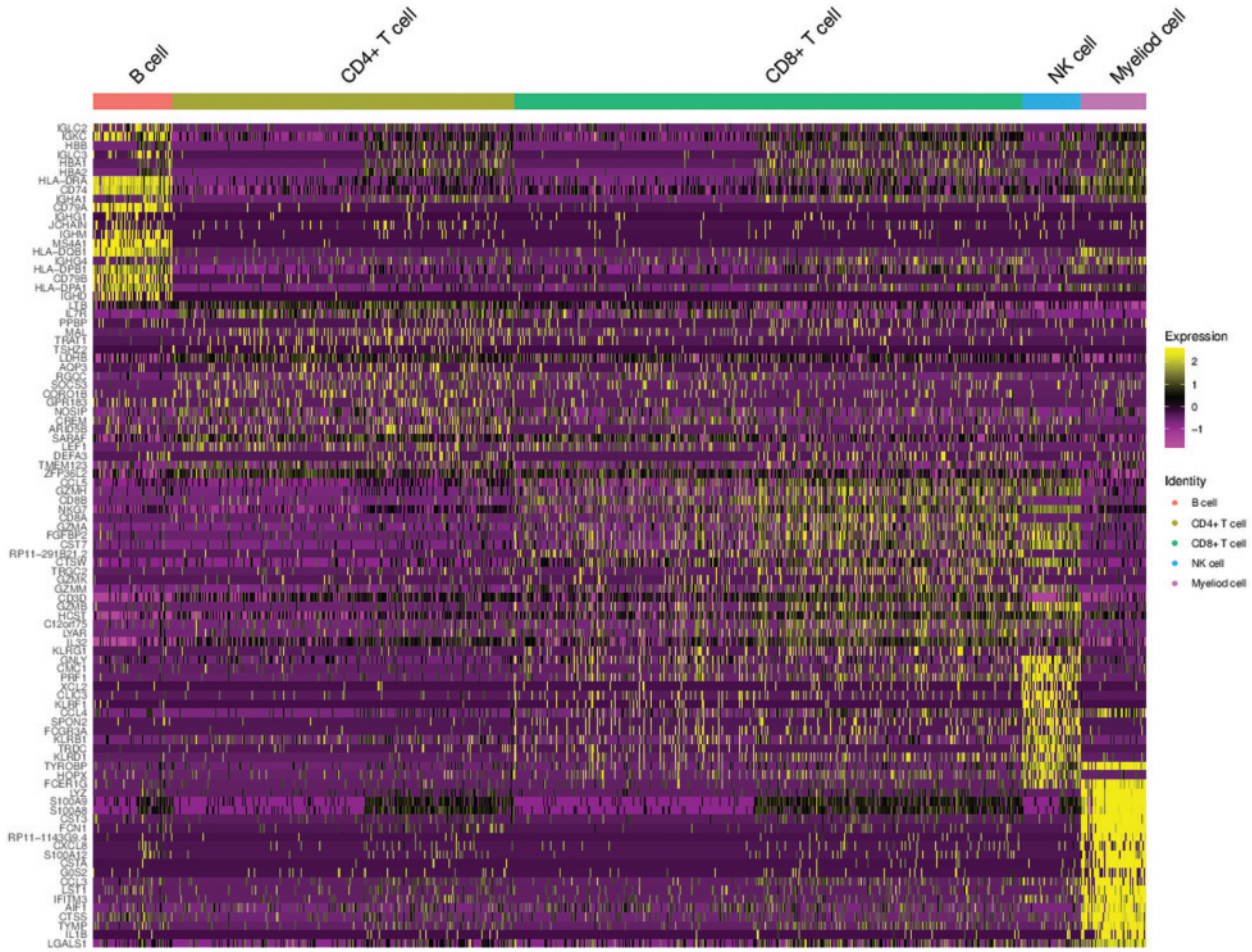


FIGURE 2. The specificity of marker gene expression.
Heatmap of standardized expression for the top 20 marker genes identified for each of the four immune cell types (B cells, $CD4^+$ T cells, $CD8^+$ T cells, myeloid cells, and NK cells). Genes are in rows, and single cells are in columns. Yellow indicates high expression of a particular gene, and purple indicates low expression.

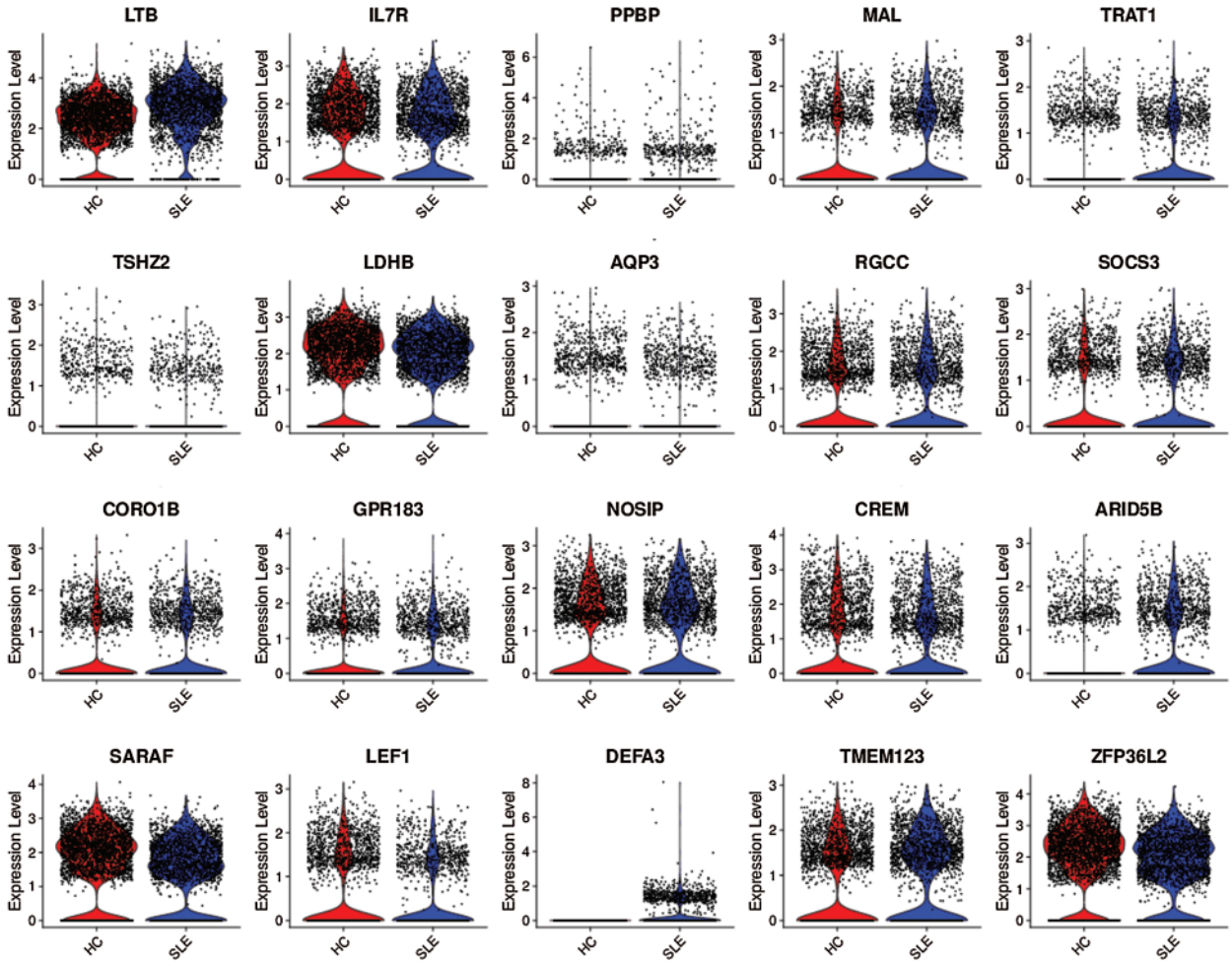


FIGURE 3. Violin diagram showing gene expression. Comparison of the expression distribution of the top 20 marker genes between the SLE patients and the healthy controls in $CD4^+$ T cells is indicated by violin plots.

downregulated genes, and HBB, HBA2, S100A8, S100A9, and HBA1 were the top upregulated genes in the $CD4^+$ T cells of the SLE patient. In the $CD8^+$ T cells of the SLE patient, the top 5 significantly downregulated genes were GNLY, DNAJB1, HSPA6, TRDC, and TYROBP; the top 5 significantly upregulated genes were HBB, S100A8, HBA2, S100A9, and IFI27. In addition, the top downregulated genes for the NK cell of the SLE patient were GADD45B, JUN, EGR1, PTGDS, and HSPA6, and the top 5 significantly upregulated genes were HBB, S100A8, CD3D, S100A9, and HBA2. Moreover, IGHG2, TMEM176B, GNLY, CEBPD, and FCER1A were the genes with the strongest downregulation, and HBB, HBA2, MT2A, FOLR3, and IFI27 were the genes with the strongest upregulation in the myeloid cells of the SLE patient.

Differentially expressed genes in GO/KEGG analysis

To further investigate SLE disease-associated cell functional states and potential molecular regulators, we performed GO analysis of the differentially expressed genes (DEGs) between the SLE and HC groups and found that biological processes such as immune system process, immune response, cell activation, defense response, leukocyte activation, immune effector process, response to stress, regulation of immune system process, leukocyte mediated immunity, etc., were highly enriched in B cells, $CD4^+$ T cells, NK cells and myeloid cells (Fig. 4, Supplementary Figs. 13–17). GO term analysis of $CD8^+$ T cells revealed enrichment of expected biological processes such as cell cycle, mitotic cell cycle process, mitotic cell cycle, cell cycle process, and cell cycle phase. Based on the DisGeNET disease database, we found that these DEGs were highly enriched in SLE or rheumatoid arthritis diseases in all four major immune cell types (B cells, $CD4^+$ T cells, $CD8^+$ T cells, myeloid cells, and NK cells), which confirmed the reliability of the experimental results and indicated that SLE is a highly complex disease. Moreover, both the KEGG pathway and protein-protein interaction network analyses were performed on DEGs between the SLE and HC groups. Notably, the IL-17 signaling pathway and Th17 cell differentiation were highly enriched in B cells, $CD4^+$ T cells, and monocytes.

Cell states were substantially altered in SLE

By means of pseudotemporal ordering of single cells, we elucidated how these cell types and states were related to each other. We reconstructed differentiation trajectories to generate a trajectory plot that can reflect both branched and linear differentiation processes (Nguyen *et al.*, 2018). Cells located on the same or adjacent branches are expected to be more hierarchically related compared to cells on the neighboring branches in a given trajectory tree. Our unbiased

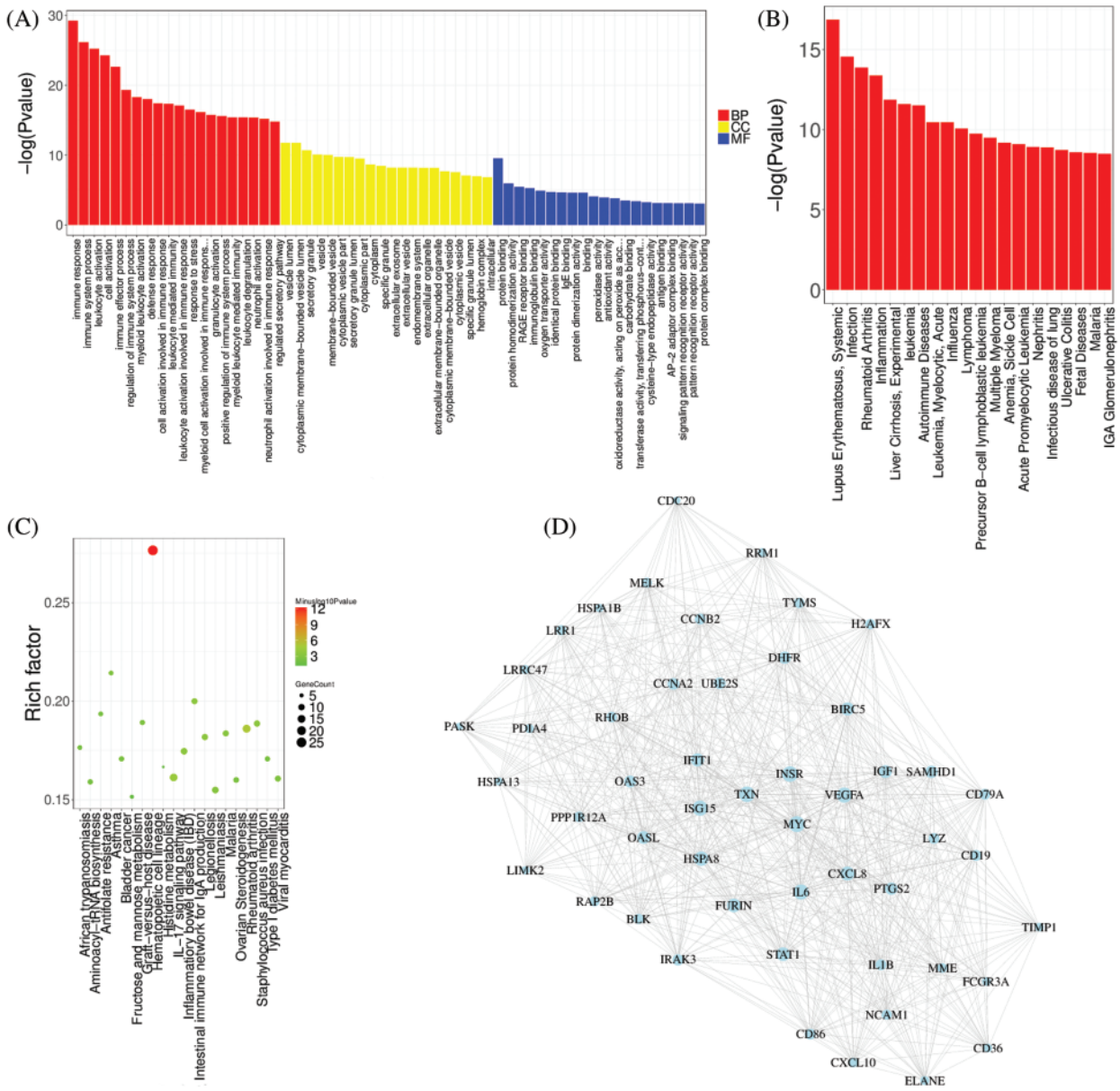


FIGURE 4. Analysis of differences in gene expression variance between the SLE patients and the healthy controls in B cells. (A) Gene ontology enrichment analysis of the differentially expressed genes (DEGs) between the SLE patients and the healthy controls in B cells. (B) DEGs between the SLE and HC groups in B cells were highly enriched in SLE and rheumatoid arthritis diseases and immune-related diseases. (C) KEGG pathway analysis of the DEGs between the SLE patients and the healthy controls in B cells. The top 20 pathways are displayed. The size of the circles depicts the gene count, and its color depicts significance levels. The number 12-16 in the label is equivalent to the -log *p*-value. (D) Protein-protein interaction network analysis of the DEGs between the SLE patients and the healthy controls in B cells. The top 50 genes were displayed based on the degree of complexity of the nodes.

clustering analysis and pseudotemporal reconstruction of differentiation trajectories revealed that five cell states were found in both the SLE and HC groups. As shown in Fig. 5, however, the morphology of differentiation trajectories was inverted between the SLE and HC groups, which suggested a different lineage hierarchy among the major immune cell types of PBMCs between these two groups. Therefore, cell states may be substantially altered in SLE disease states.

Discussion

Systemic lupus erythematosus is a complex autoimmune disease. Abnormal expression of key signaling molecules and defective immune cell functions play a significant role in the pathogenesis of SLE. However, the precise pathologic mechanisms underlying abnormal immune cell functions in SLE remain incompletely understood. Here, we report comprehensive single-cell expression profiling of PBMCs in a patient with SLE and a healthy control individual. We analyzed 16,021 cells (8,030 cells from the SLE patient and 7,991 cells from the HC). Clustering analysis identified 12 distinct cell clusters consisting of as few as 104 cells to as many as 3,080 cells per cluster. This analysis identified 12 clusters, including 5 major immune cell types: B cells, $CD4^+$ T cells, $CD8^+$ T cells, myeloid cells, and NK cells. The marker genes of each cluster and the four major immune cell types are displayed. Of note, we discovered a large number of differentially expressed genes (DEGs) between

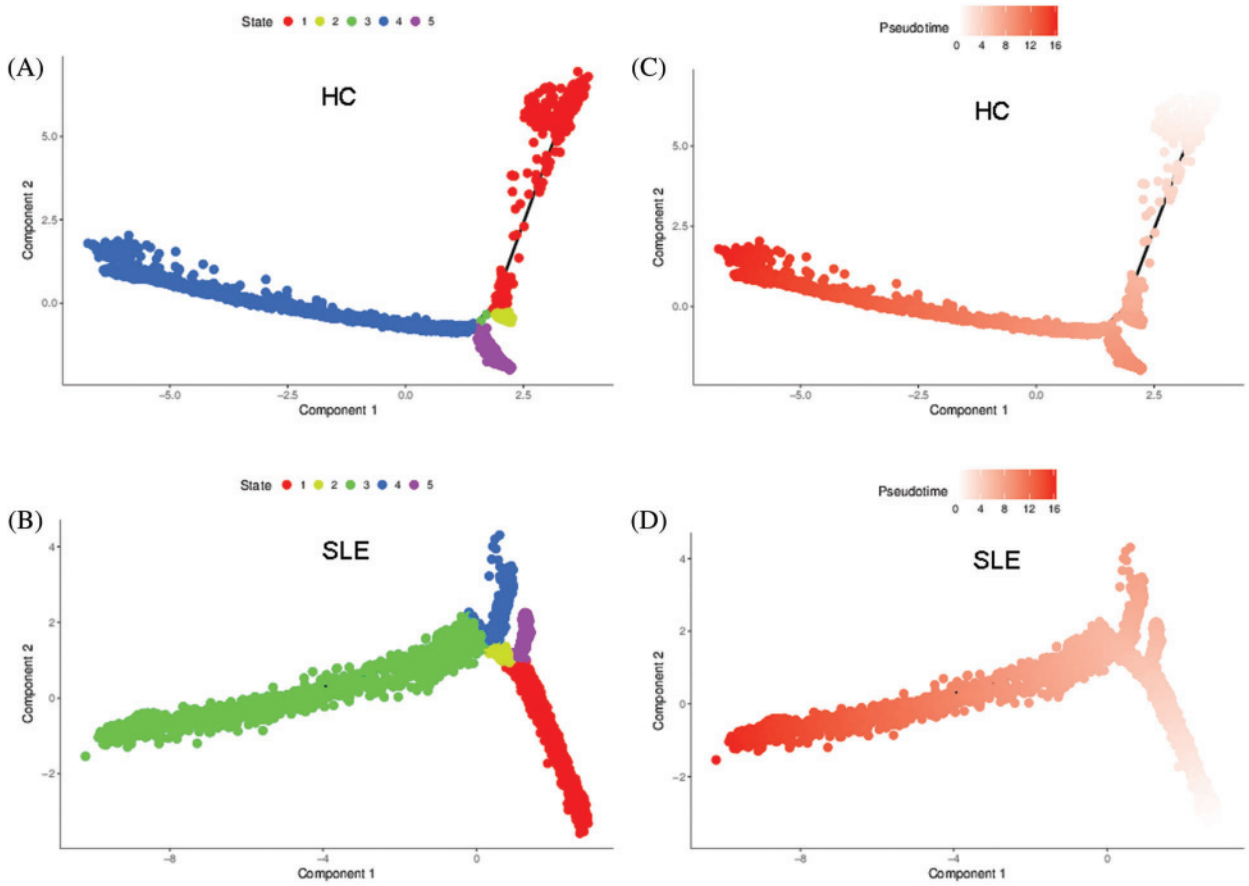


FIGURE 5. Reconstruction of differentiation and relation of cell states to immune cells in the SLE patients and the healthy controls (HCs). (A, B) Five cell states were found both in the HC (A) and SLE groups (B). (C, D) Pseudotemporal trajectory of immune cells in the HCs (C) and the SLE patients (D) is shown, which are colored by cell state designation. The morphology of differentiation trajectories was also inverted between the SLE and HC groups.

the SLE and HC groups in B cells, $CD4^+$ T cells, $CD8^+$ T cells, myeloid cells, and NK cells. Our analyses confirmed many important observations made previously *in vitro*, in bulk, or using animal models. A previous study proved that distinct transcriptional profiles were observed in B cell, $CD4^+$ T cell, and $CD33^+$ myeloid cell subsets of the SLE patients compared with the controls (Becker *et al.*, 2013). Among the top 5 significantly downregulated/upregulated expressed genes identified in each immune cell type, S100A8 and S100A9 could be found in B cells, $CD4^+$ T cells, $CD8^+$ T cells, and NK cells, which suggested that these two genes played an essential role in the pathogenesis of this SLE patient. Previous research has indicated that differential expression of S100A8/A9 was found in SLE patients, and high levels of S100A8/A9 predispose patients to cardiovascular disease (Lood *et al.*, 2016; Pavon *et al.*, 2012; Tyden *et al.*, 2013). Moreover, other differentially expressed genes between the SLE and HC groups (e.g., HSPA1B, IFI27, CD3D) were identified in this study, which was in agreement with previous studies (Bing *et al.*, 2016; Fürnrohr *et al.*, 2010; Lindén *et al.*, 2017). Reyes *et al.* utilized a low-input microfluidic system to identify subset-specific disease signatures by profiling four immune cell subsets ($CD4^+$ T cells, $CD8^+$ T cells, B cells, and $CD14^+$ monocytes) in blood from patients with SLE and matched control subjects. They found that gene sets with targets of IFN are upregulated in patients with SLE (Reyes *et al.*, 2019).

The top 20 GO terms of the differentially expressed genes (DEGs) between the SLE and HC groups strongly overlapped among B cells, $CD4^+$ T cells, myeloid cells, and NK cells and included immune system process, immune response, cell activation, defense response, leukocyte activation, immune effector process, response to stress, regulation of immune system process, and leukocyte mediated immunity, which indicated that these biological processes play a pivotal role in the pathogenesis and development of SLE. These results were further confirmed by the fact that these DEGs were highly enriched in SLE and rheumatoid arthritis diseases in all five major immune cell types. Moreover, comprehensive cell atlases of both healthy individual and the SLE patient revealed that cell states were substantially altered in disease states. This observation was consistent with previous studies. Work carried out by Sinai *et al.* reported that T/B-cell interactions are transient in response to weak stimuli in SLE-prone mice (Sinai *et al.*, 2014). A previous study by Katsuyama *et al.* (2018) showed that SLE T cells are characterized by multiple aberrant signaling pathways, such as an activated PI3K-Akt-mTORC1 pathway, calcium/calmodulin kinase IV (CaMKIV), Rho-associated protein kinase (ROCK), protein phosphatase 2A (PP2A), and decreased CD3ζ. It was also reported that Act1 has a novel function as a negative regulator in T and B cells via direct inhibition of STAT3, thereby contributing to the development of SLE (Zhang *et al.*, 2018).

In conclusion, our scRNA-seq analysis of human PBMCs from SLE patient and healthy control for the first time allows unbiased, *de novo* identification of distinct cell types and states. By identifying cell subtype alterations, differentially expressed genes, and altered pathways, this technique provides a system-level understanding that is rooted in molecular information and forms the basis for improved methods of early detection of SLE and possible strategies for SLE prevention.

Author Contributions: YD, DL, WD designed the experiments; XH, DT, and FZ performed the experiments and wrote the manuscript; MO, YX, HX performed the computational analysis; XH and XZ provided valuable resources; all authors approved the final version.

Availability of Data and Materials: Data sharing not applicable to this article as all data generated or analyzed during this study are included in this article.

Funding Statement: This work was supported by the National Natural Science Foundation of China (Grant No. 81671596), the Natural Science Foundation of Guangxi (Grant No. 2019GXNSFBA245032, and No. 2017GXNSFAA198375), the Guangxi Science and Technology Plan Project (Gui Ke AD20238021), the National Science Foundation for Young Scientists of China (Grant No. 31700795), the science and technology plan of Shenzhen (No. JCYJ20170307095606266), Shenzhen science and technology research foundation (JCYJ20160422154407256), Sanming project of medicine in Shenzhen, the group of Rheumatology and Immunology led by Xiaofeng Zeng of Peking Union medical college Hospital and Dongzhou Liu in Shenzhen People's Hospital (SYJY201704 and SYJY201705), and the open funds of the Guangxi Key Laboratory of Tumor Immunology and Microenvironmental Regulation (2019KF004), and Guilin science research and technology development project (20190218-5-5).

Conflicts of Interest: The authors declare no conflicts of interest.

References

Becker AM, Dao KH, Han BK, Kornu R, Lakhanpal S, Mobley AB, Li QZ, Lian Y, Wu T, Reimold AM, Olsen NJ, Karp DR, Chowdhury FZ, Farrar JD, Satterthwaite AB, Mohan C, Lipsky PE, Wakeland EK, Davis LS (2013). SLE peripheral blood B cell, T cell and myeloid cell transcriptomes display unique profiles and each subset contributes to the interferon signature. *PLoS One* **8**: e67003. DOI 10.1371/journal.pone.0067003.

Bing PF, Xia W, Wang L, Zhang YH, Lei SF, Deng FY (2016). Common marker genes identified from various sample types for systemic lupus erythematosus. *PLoS One* **11**: e156234.

Burl RB, Ramseyer VD, Rondini EA, Pique-Regi R, Lee YH, Granneman JG (2018). Deconstructing adipogenesis induced by β3-adrenergic receptor activation with single-cell expression profiling. *Cell Metabolism* **28**: 300–309.e4. DOI 10.1016/j.cmet.2018.05.025.

Cochain C, Vafadarnejad E, Arampatzi P, Pelisek J, Winkels H, Ley K, Wolf D, Saliba AE, Zernecke A (2018). Single-cell RNA-seq reveals the transcriptional landscape and heterogeneity of aortic macrophages in murine atherosclerosis. *Circulation Research* **122**: 1661–1674. DOI 10.1161/CIRCRESAHA.117.312509.

D'Avola D, Villacorta-Martin C, Martins-Filho S N, Craig A, Labgaa I, von Felden J, Kimaada A, Bonaccorso A, Tabrizian P, Hartmann B M, Sebra R, Schwartz M, Villanueva A (2018). High-density single cell mRNA sequencing to characterize circulating tumor cells in hepatocellular carcinoma. *Scientific Reports* **8**: 472. DOI 10.1038/s41598-018-30047-y.

Fürnrohr BG, Wach S, Kelly JA, Haslbeck M, Weber CK, Stach CM, Hueber AJ, Graef D, Spriewald BM, Manger K, Herrmann M, Kaufman KM, Frank SG, Goodmon E, James JA, Schett G, Winkler TH, Harley JB, Voll RE (2010). Polymorphisms in the Hsp70 gene locus are genetically associated with systemic lupus erythematosus. *Annals of the Rheumatic Diseases* **69**: 1983–1989. DOI 10.1136/ard.2009.122630.

Hartman E, van Royen-Kerkhof A, Jacobs J, Welsing P, Fritsch-Stork R (2018). Performance of the 2012 Systemic Lupus International Collaborating Clinics classification criteria versus the 1997 American College of Rheumatology classification criteria in adult and juvenile systemic lupus erythematosus. A systematic review and meta-analysis. *Autoimmunity Reviews* **17**: 316–322. DOI 10.1016/j.autrev.2018.01.007.

Ilicic T, Kim JK, Kolodziejczyk AA, Bagger FO, McCarthy DJ, Marioni JC, Teichmann SA (2016). Classification of low quality cells from single-cell RNA-seq data. *Genome Biology* **17**: 57. DOI 10.1186/s13059-016-0888-1.

Katsuyama T, Tsokos GC, Moulton VR (2018). Aberrant T cell signaling and subsets in systemic lupus erythematosus. *Frontiers in Immunology* **9**: 147. DOI 10.3389/fimmu.2018.01088.

Lambrechts D, Wauters E, Boeckx B, Aibar S, Nittner D, Burton O, Bassez A, Decaluwe H, Pircher A, Van den Eynde K, Weynand B, Verbeken E, De Leyn P, Liston A, Vansteenkiste J, Carmeliet P, Aerts S, Thienpont B (2018). Phenotype molding of stromal cells in the lung tumor microenvironment. *Nature Medicine* **24**: 1277–1289. DOI 10.1038/s41591-018-0096-5.

Li WV, Li JJ (2018). An accurate and robust imputation method scImpute for single-cell RNA-seq data. *Nature Communications* **9**: 57. DOI 10.1038/s41467-018-03405-7.

Lindén M, Ramírez Sepúlveda JI, James T, Thorlacius GE, Brauner S, Gómez-Cabrero D, Olsson T, Kockum I, Wahren-Herlenius M (2017). Sex influences eQTL effects of SLE and Sjögren's syndrome-associated genetic polymorphisms. *Biology of Sex Differences* **8**: 1319. DOI 10.1186/s13293-017-0153-7.

Lood C, Tyden H, Gullstrand B, Jonsen A, Kallberg E, Morgelin M, Kahn R, Gunnarsson I, Leanderson T, Ivars F, Svenungsson E, Bengtsson AA (2016). Platelet-derived S100A8/A9 and cardiovascular disease in systemic lupus erythematosus. *Arthritis & Rheumatology* **68**: 1970–1980. DOI 10.1002/art.39656.

Maria NI, Davidson A (2018). Emerging areas for therapeutic discovery in SLE. *Current Opinion in Immunology* **55**: 1–8. DOI 10.1016/j.coi.2018.09.004.

Nguyen QH, Pervolarakis N, Blake K, Ma D, Davis RT, James N, Phung AT, Willey E, Kumar R, Jabart E, Driver I, Rock J, Goga A, Khan SA, Lawson DA, Werb Z, Kessenbrock K (2018). Profiling human breast epithelial cells using single cell RNA sequencing identifies cell diversity. *Nature Communications* **9**: 747. DOI 10.1038/s41467-018-04334-1.

Pavon EJ, Garcia-Rodriguez S, Zumaquero E, Perandres-Lopez R, Rosal-Vela A, Lario A, Longobardo V, Carrascal M, Abian J, Callejas-Rubio JL, Ortego-Centeno N, Zubiaur M, Sancho J (2012). Increased expression and phosphorylation of the two S100A9 isoforms in mononuclear cells from patients with systemic lupus erythematosus: A proteomic signature for circulating low-density granulocytes. *Journal of Proteomics* **75**: 1778–1791. DOI 10.1016/j.jprot.2011.12.020.

Reyes M, Vickers D, Billman K, Eisenhaure T, Hoover P, Browne EP, Rao DA, Hacohen N, Blainey PC (2019). Multiplexed enrichment and genomic profiling of peripheral blood cells reveal subset-specific immune signatures. *Science Advances* **5**: eaau9223. DOI 10.1126/sciadv.aau9223.

Rheaume BA, Jereen A, Bolisetty M, Sajid MS, Yang Y, Renna K, Sun L, Robson P, Trakhtenberg EF (2018). Single cell transcriptome profiling of retinal ganglion cells identifies cellular subtypes. *Nature Communications* **9**: 618. DOI 10.1038/s41467-018-05134-3.

Scott CL, T'Jonck W, Martens L, Todorov H, Sichien D, Soen B, Bonnardel J, De Prijck S, Vandamme N, Cannoodt R, Saelens W, Vanneste B, Toussaint W, Bleser PD, Takahashi N, Vandenabeele P, Henri S, Pridans C, Hume DA, Lambrecht BN, Baetselier PD, Milling SWF, Ginderachter JAV, Malissen B, Berx G, Beschin A, Saeys Y, Guilliams M (2018). The transcription factor ZEB2 is required to maintain the tissue-specific identities of macrophages. *Immunity* **49**: 312–325.e5. DOI 10.1016/j.immuni.2018.07.004.

Sinai P, Dozmorov IM, Song R, Schwartzberg PL, Wakeland EK, Wulfing C (2014). T/B-cell interactions are more transient in response to weak stimuli in SLE-prone mice. *European Journal of Immunology* **44**: 3522–3531. DOI 10.1002/eji.201444602.

Sui W, Hou X, Che W, Chen J, Ou M, Xue W, Dai Y (2013). Hematopoietic and mesenchymal stem cell transplantation for severe and refractory systemic lupus erythematosus. *Clinical Immunology* **148**: 186–197. DOI 10.1016/j.clim.2013.05.014.

Sui W, Hou X, Che W, Yang M, Dai Y (2013). The applied basic research of systemic lupus erythematosus based on the biological omics. *Genes & Immunity* **14**: 133–146. DOI 10.1038/gene.2013.3.

Sui W, Hou X, Zou G, Che W, Yang M, Zheng C, Liu F, Chen P, Wei X, Lai L, Dai Y (2015). Composition and variation analysis of the TCR β-chain CDR3 repertoire in systemic lupus erythematosus using high-throughput sequencing. *Molecular Immunology* **67**: 455–464. DOI 10.1016/j.molimm.2015.07.012.

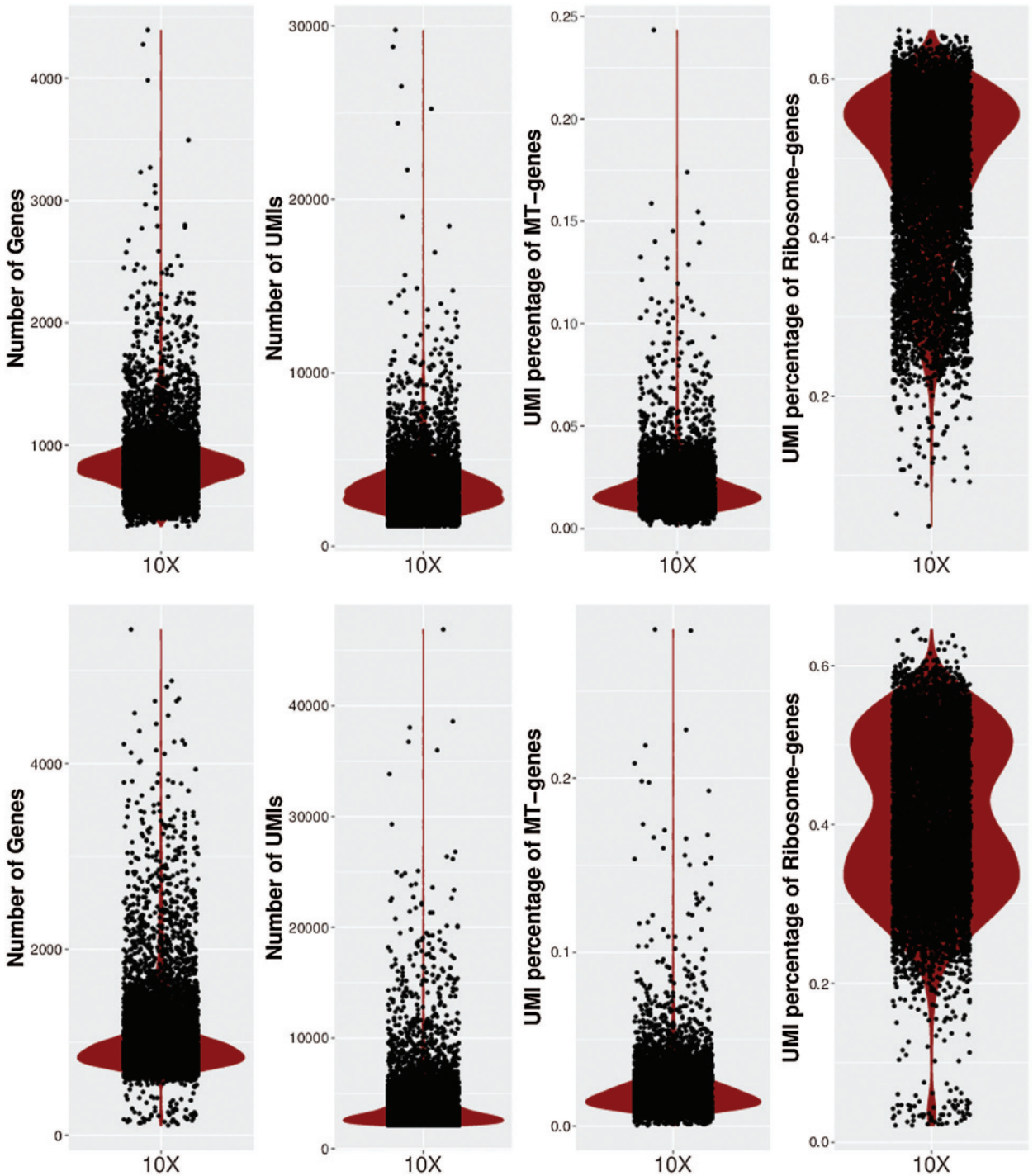
Tyden H, Lood C, Gullstrand B, Jonsen A, Nived O, Sturfelt G, Truedsson L, Ivars F, Leanderson T, Bengtsson AA (2013). Increased serum levels of S100A8/A9 and S100A12 are associated with cardiovascular disease in patients with inactive systemic lupus erythematosus. *Rheumatology* **52**: 2048–2055. DOI 10.1093/rheumatology/ket263.

Xie T, Wang Y, Deng N, Huang G, Taghavifar F, Geng Y, Liu N, Kulur V, Yao C, Chen P, Liu Z, Stripp B, Tang J, Liang J, Noble PW, Jiang D (2018). Single-cell deconvolution of fibroblast heterogeneity in mouse pulmonary fibrosis. *Cell Reports* **22**: 3625–3640. DOI 10.1016/j.celrep.2018.03.010.

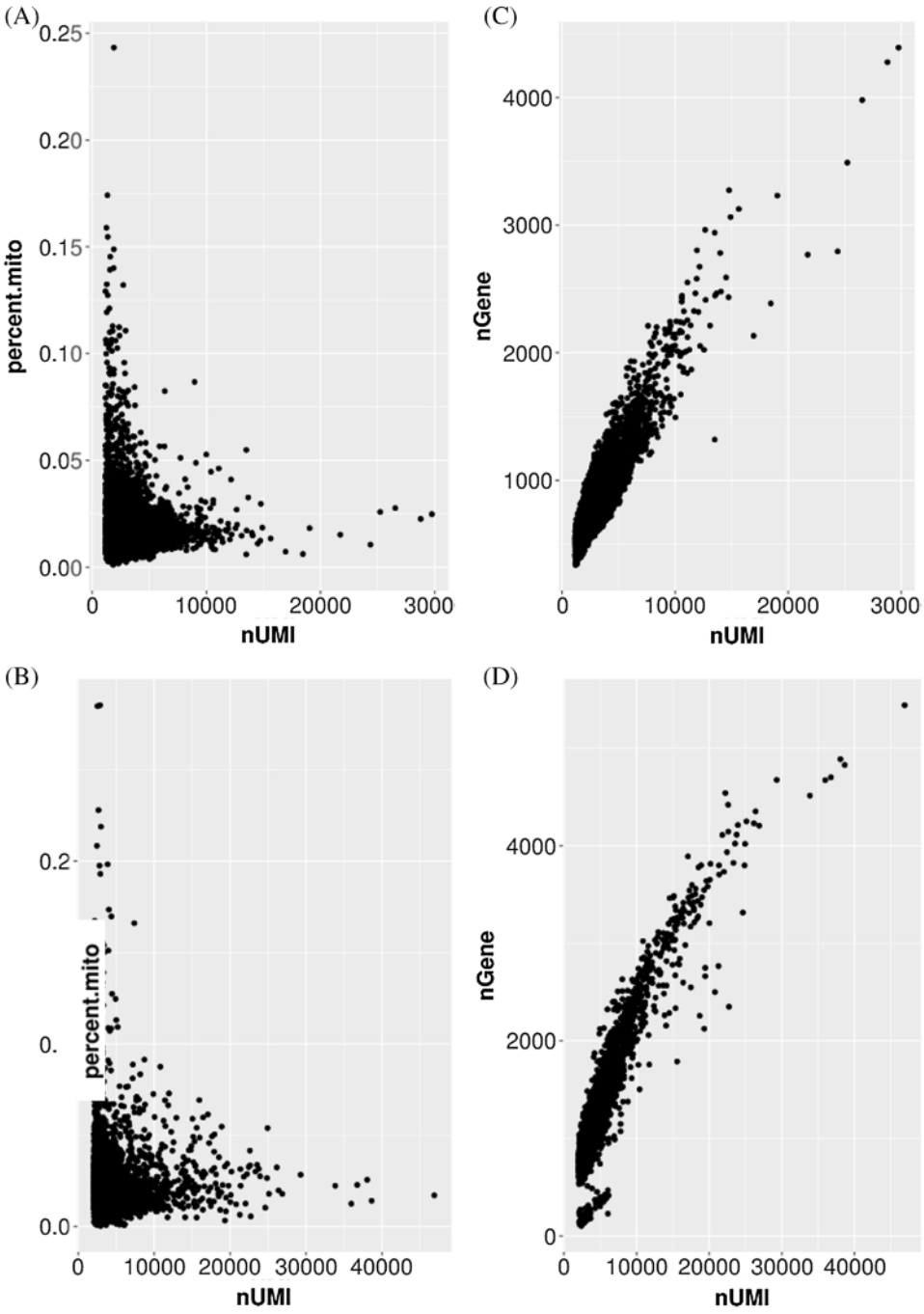
Zeisel A, Hochgerner H, Lönnerberg P, Johnsson A, Memic F, van der Zwan J, Häring M, Braun E, Borm LE, La Manno G, Codeluppi S, Furlan A, Lee K, Skene N, Harris KD, Hjerling-Leffler J, Arenas E, Ernfors P, Marklund U, Linnarsson S (2018). Molecular architecture of the mouse nervous system. *Cell* **174**: 999–1014.e22. DOI 10.1016/j.cell.2018.06.021.

Zhang CJ, Wang C, Jiang M, Gu C, Xiao J, Chen X, Martin BN, Tang F, Yamamoto E, Xian Y, Wang H, Li F, Sartor RB, Smith H, Husni ME, Shi FD, Gao J, Carman J, Dongre A, McKarns SC, Coppieters K, Jørgensen TN, Leonard WJ, Li X (2018). Act1 is a negative regulator in T and B cells via direct inhibition of STAT3. *Nature Communications* **9**: 2110. DOI 10.1038/s41467-018-04974-3.

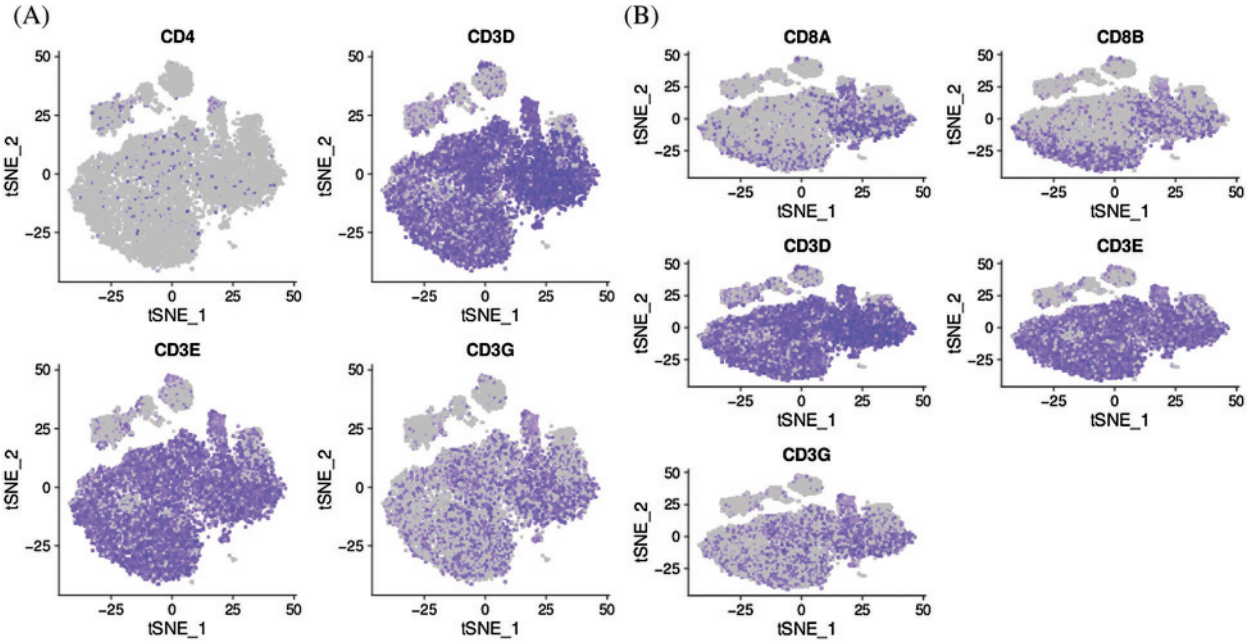
Zheng GX, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, Ziraldo SB, Wheeler TD, McDermott GP, Zhu J, Gregory MT, Shuga J, Montesclaros L, Underwood JG, Masquelier DA, Nishimura SY, Schnall-Levin M, Wyatt PW, Hindson CM, Bharadwaj R, Wong A, Ness KD, Beppu LW, Deeg HJ, McFarland C, Loeb KR, Valente WJ, Ericson NG, Stevens EA, Radich JP, Mikkelsen TS, Hindson BJ, Bielas JH (2017). Massively parallel digital transcriptional profiling of single cells. *Nature Communications* **8**: 236. DOI 10.1038/ncomms14049.



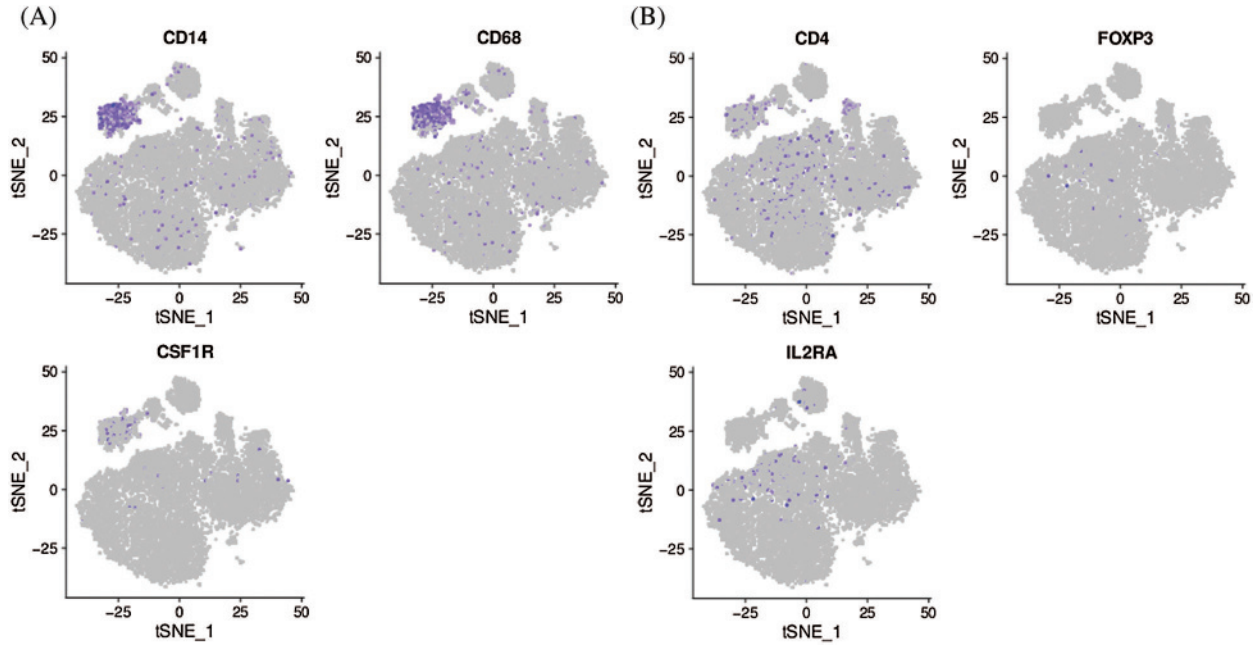
SUPPLEMENTARY FIGURE 1. Quality Control. Plots of number of genes, UMIs, and the percentage of UMIs that derived from mitochondrial and ribosomal genome detected per single cell for health control (Upper) and SLE patient (Bottom).



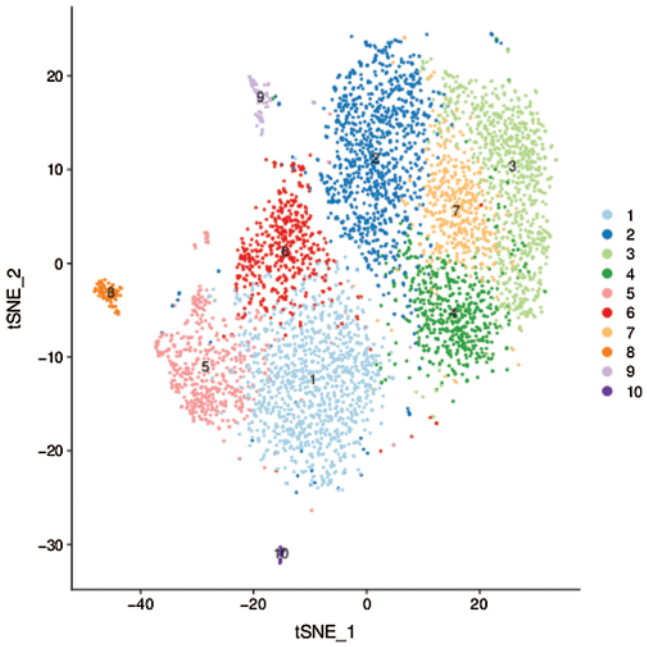
SUPPLEMENTARY FIGURE 2. Plot of the percentage of UMIs that derived from mitochondrial versus number of UMIs per single cell for all the cells from health control (A) and SLE patient (B). Plot of number of genes versus number of UMIs per single cell for all the cells from health control (C) and SLE patient (D).



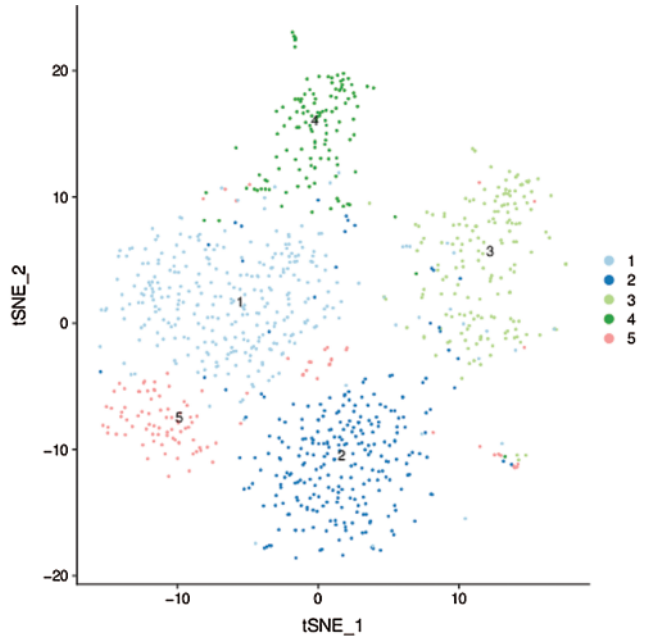
SUPPLEMENTARY FIGURE 3. Expression patterns of the canonical marker genes across the major immune cell types. t-SNE plots of canonical markers for the CD4+ T cell (A), and CD8+ T cell (B) were shown, and the gradient of purple reflects expression levels.



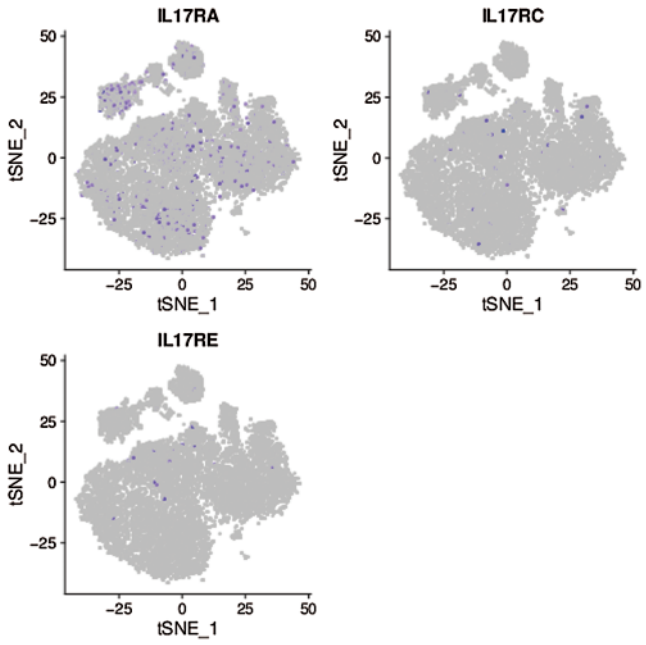
SUPPLEMENTARY FIGURE 4. Expression patterns of the canonical marker genes across the major immune cell types. t-SNE plots of canonical markers for the monocytes cell (A), and regulatory T cell (B) were shown, and the gradient of purple reflects expression levels.



SUPPLEMENTARY FIGURE 5. $CD4^+$ T cells were further subdivided into 10 clusters.

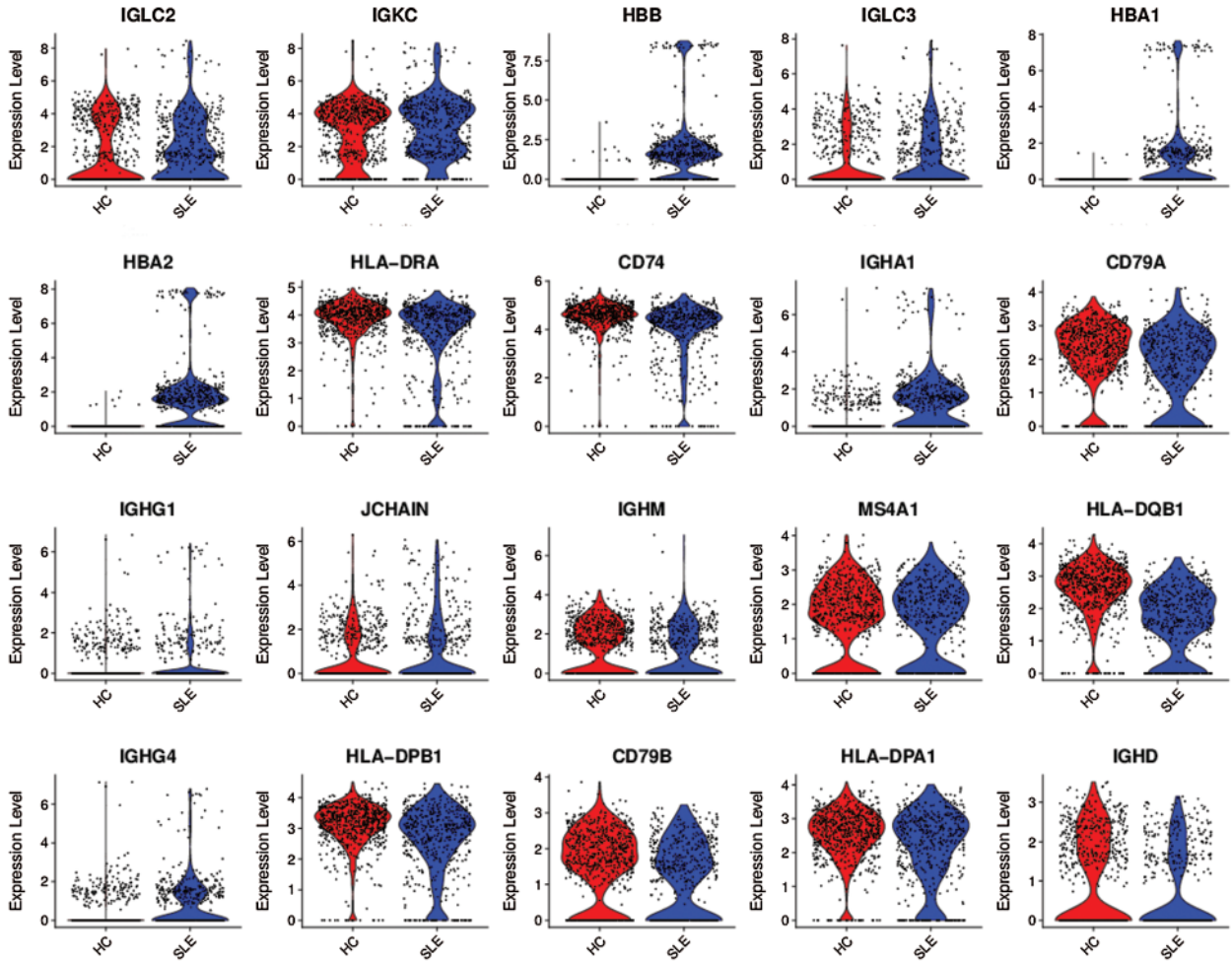


SUPPLEMENTARY FIGURE 6. Myeliod cells were further subdivided into 5 clusters.

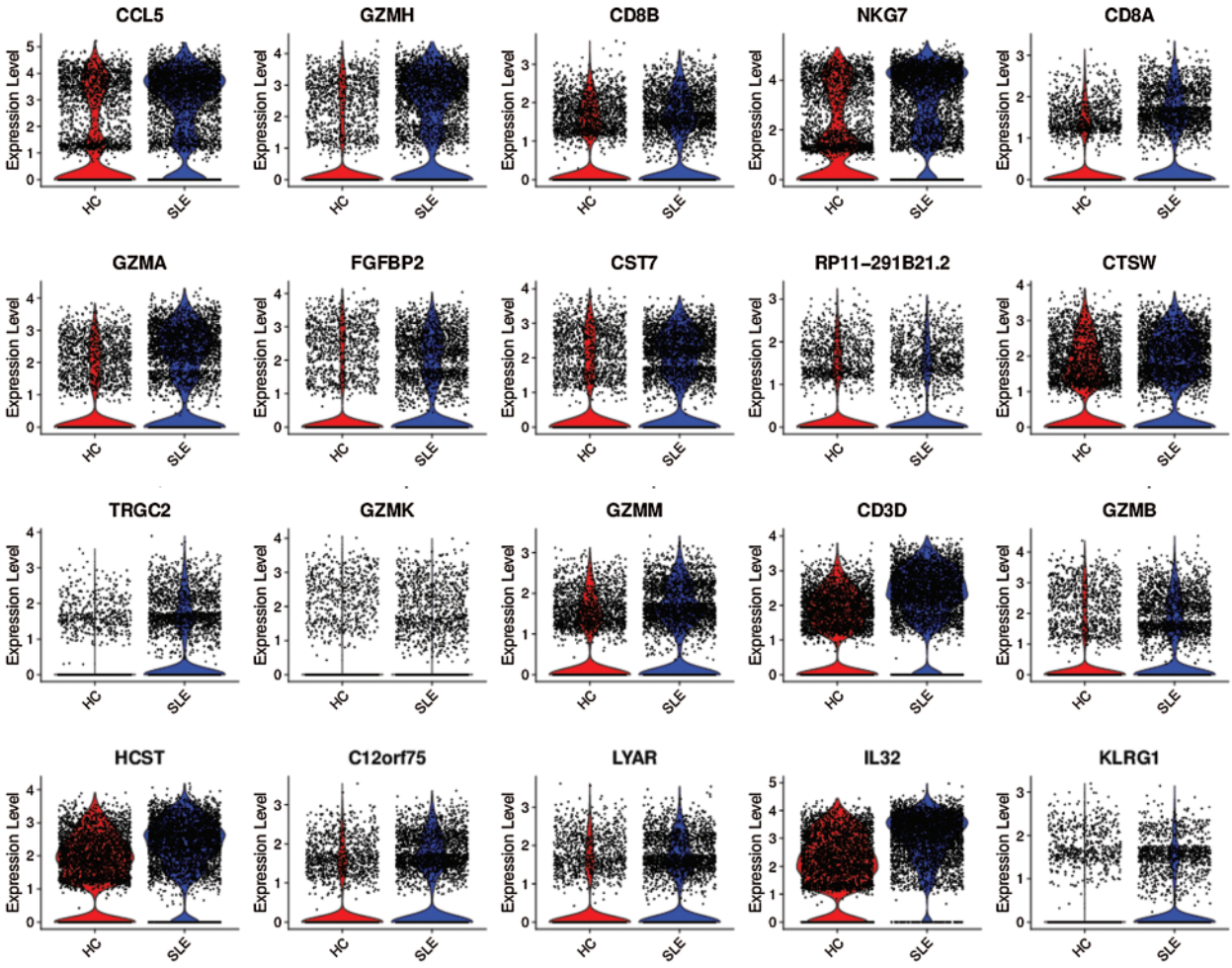


SUPPLEMENTARY FIGURE 7. Expression patterns of the canonical marker genes across the major immune cell types. t-SNE plots of canonical markers for the Th17 cell were shown, and the gradient of purple reflects expression levels.

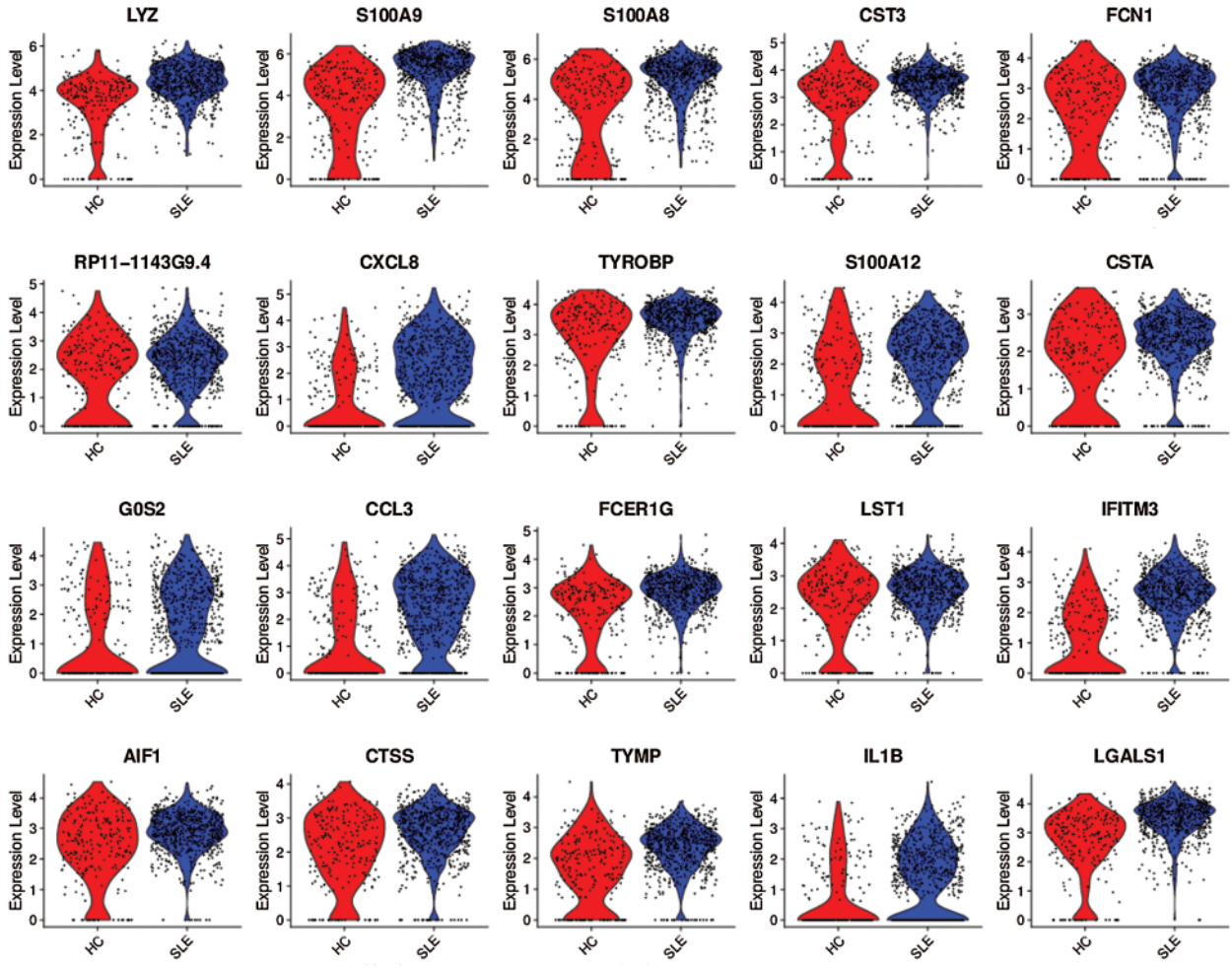
SUPPLEMENTARY FIGURE 8. Expression patterns of the canonical marker genes across the major immune cell types. t-SNE plots of canonical markers for the neutrophils were shown, and the gradient of red reflects expression levels.



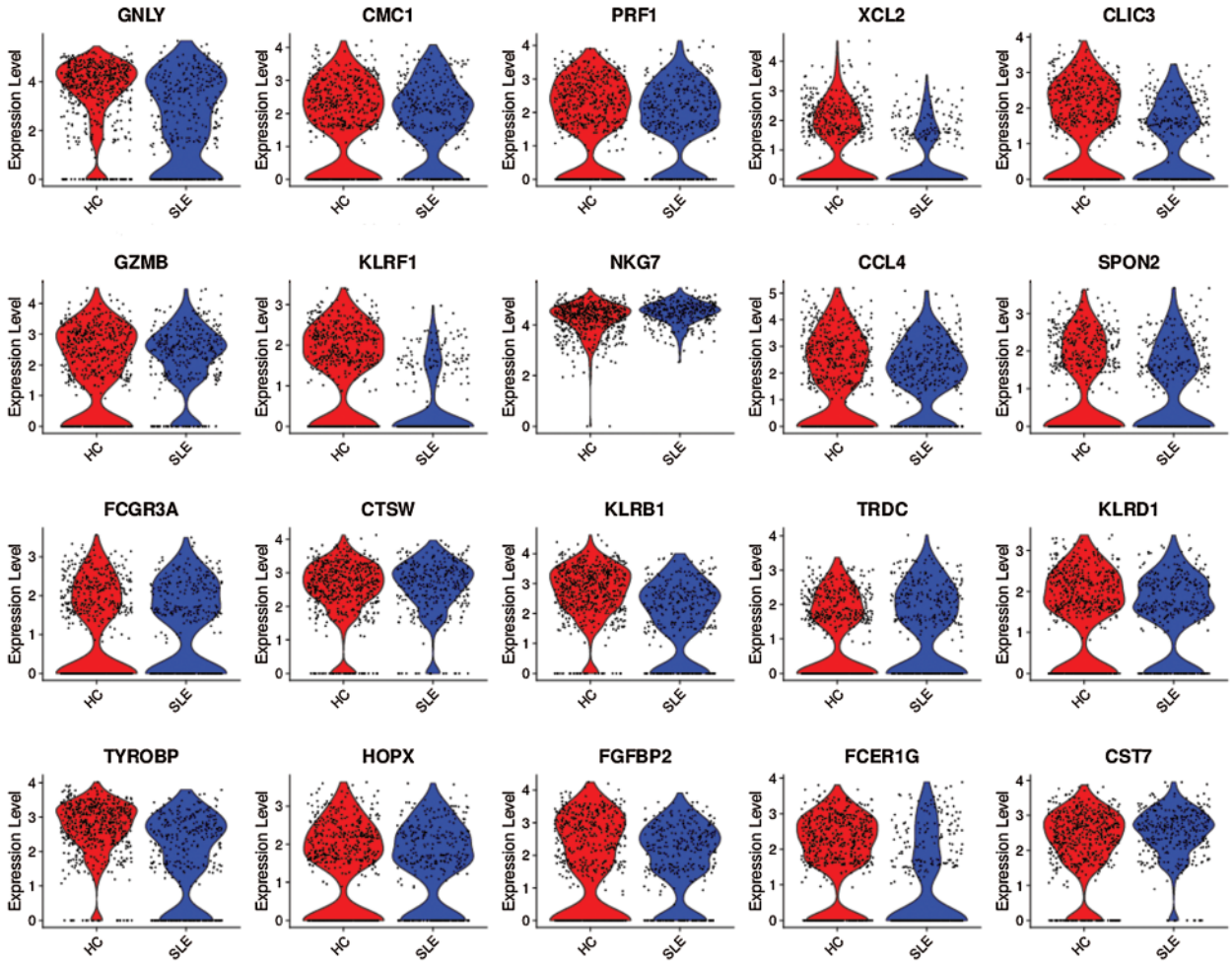
SUPPLEMENTARY FIGURE 9. Compare of the expression distribution of the top 20 marker genes between SLE patient and health control in B cell were indicated by violin plots.



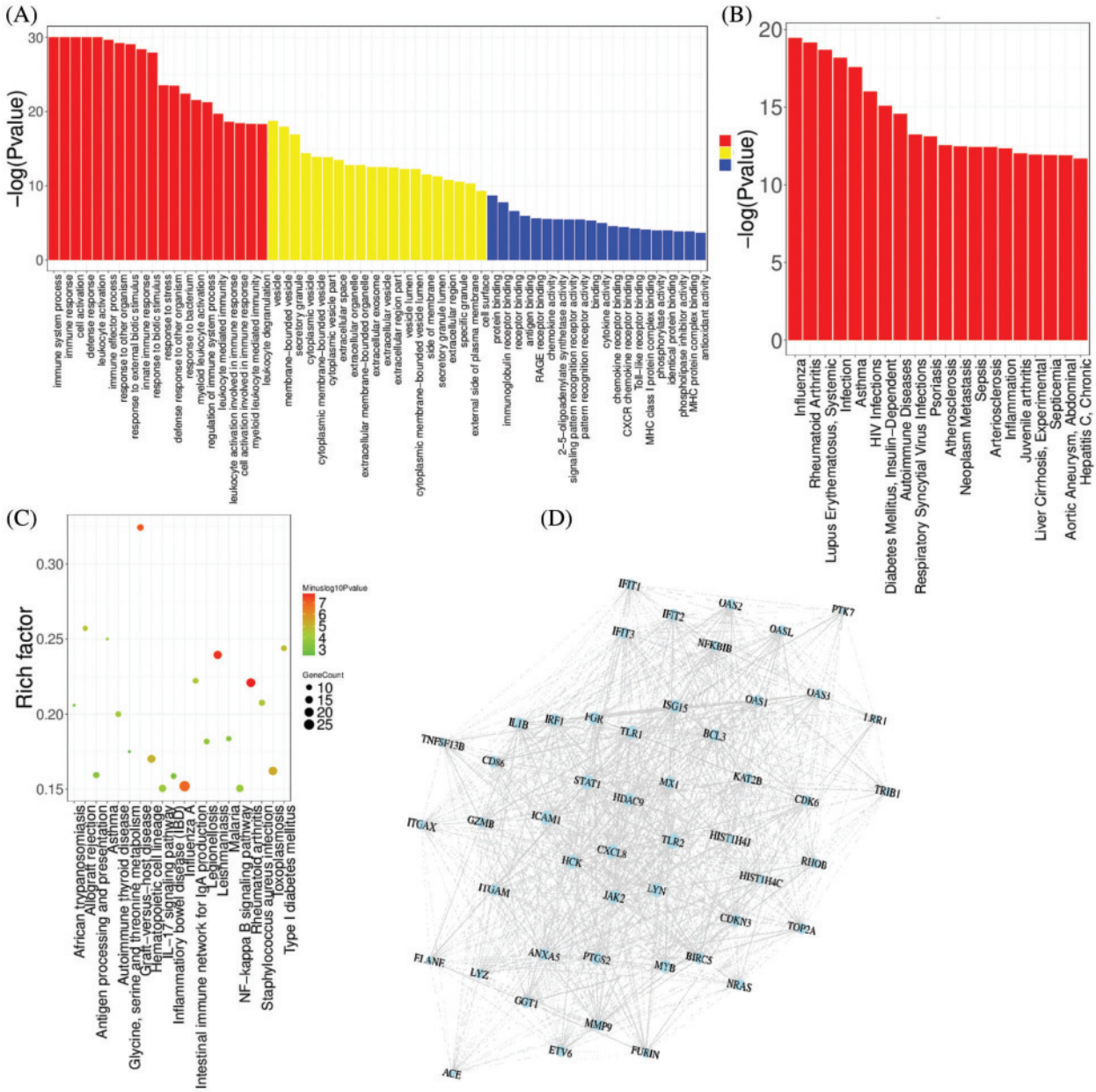
SUPPLEMENTARY FIGURE 10. Compare of the expression distribution of the top 20 marker genes between SLE patient and health control in CD8+ T cell were indicated by violin plots.



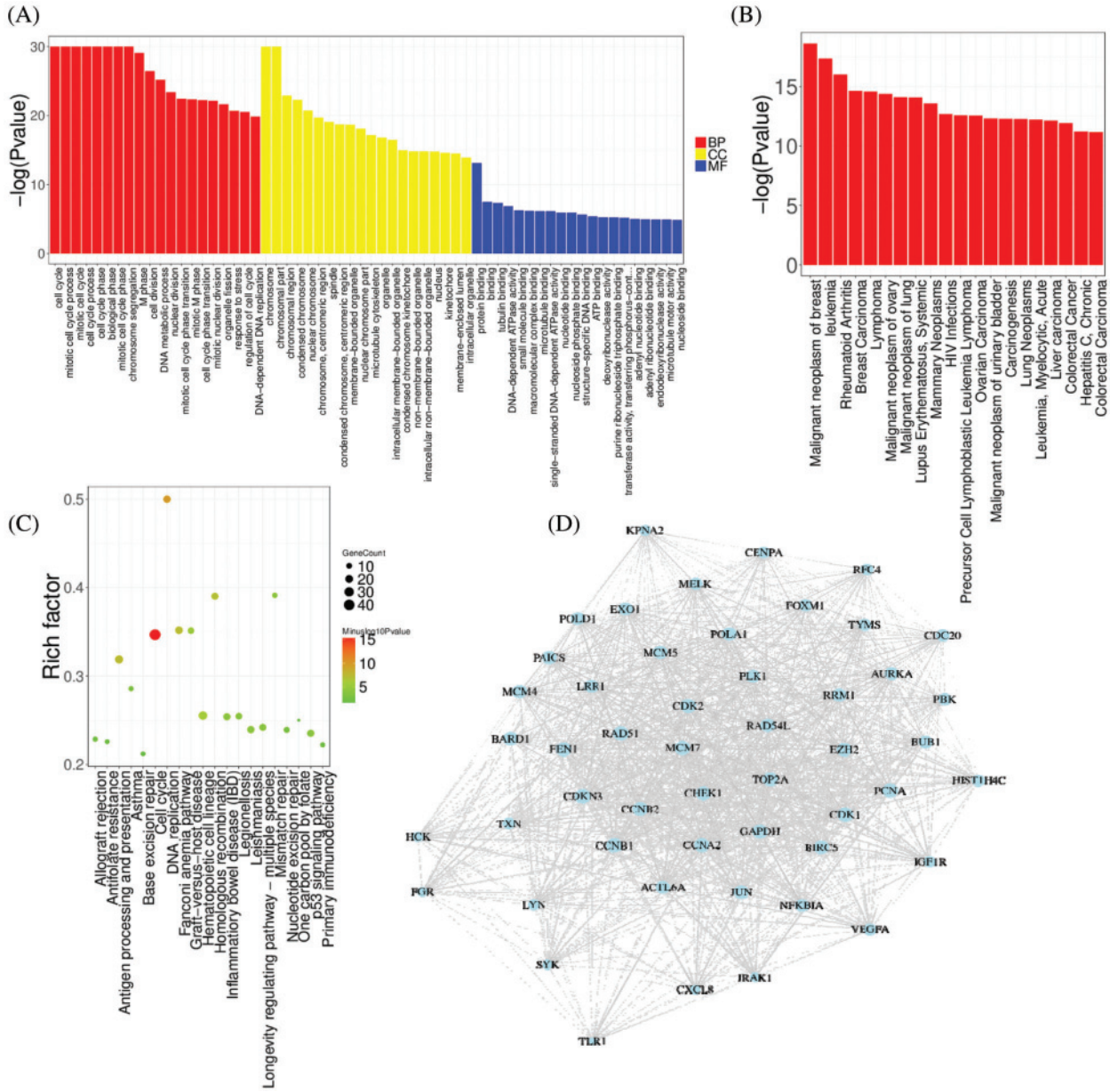
SUPPLEMENTARY FIGURE 11. Compare of the expression distribution of the top 20 marker genes between SLE patient and health control in myeliod cells were indicated by violin plots.



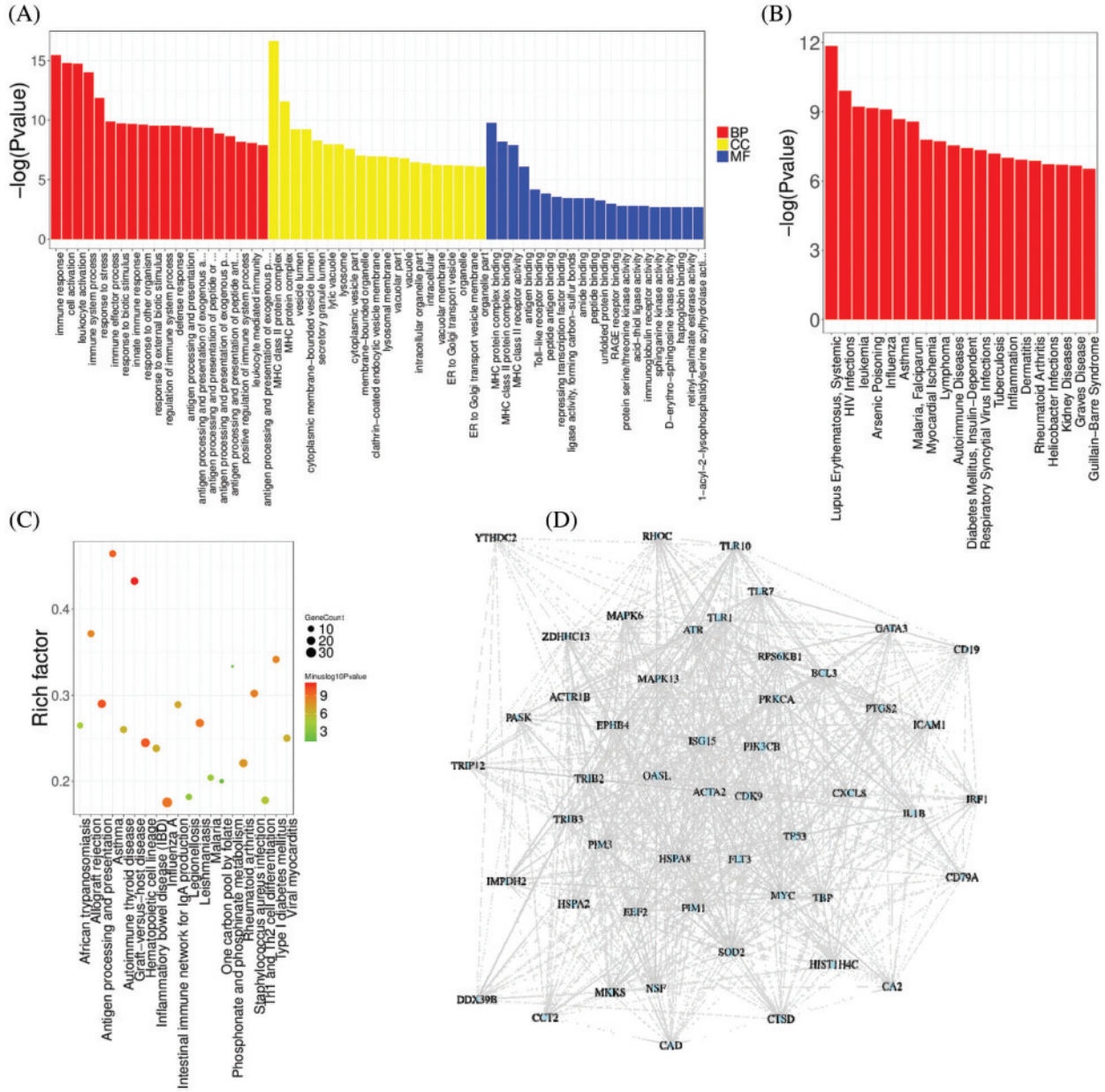
SUPPLEMENTARY FIGURE 12. Compare of the expression distribution of the top 20 marker genes between SLE patient and health control in NK cell were indicated by violin plots.



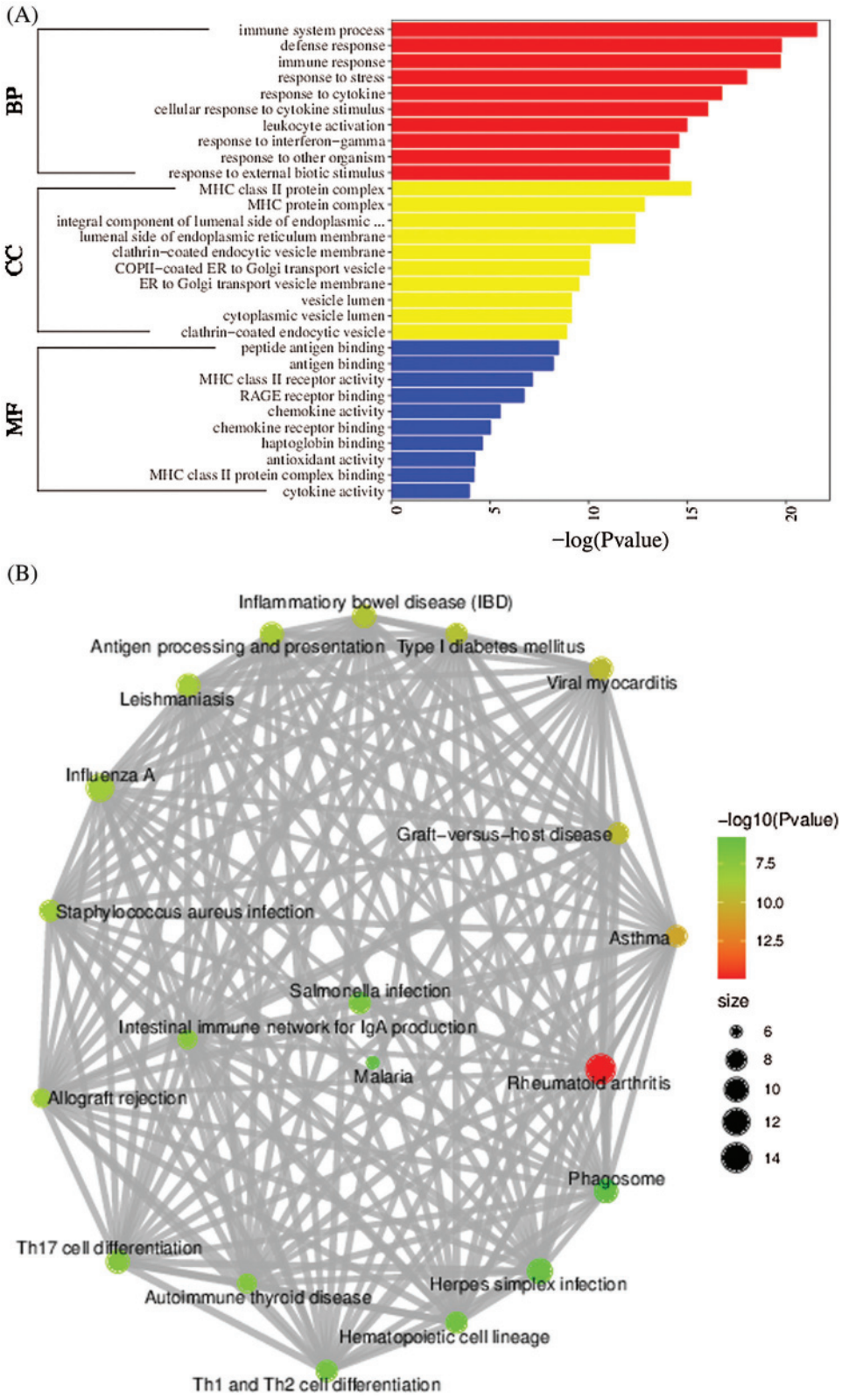
SUPPLEMENTARY FIGURE 13. Analysis of differences in gene expression variance between SLE patient and health control in the CD4+ T cells. Gene ontology enrichment analysis of the differentially expressed genes (DEG) between SLE patient and health control in CD4+ T cell (A). DEGs between SLE and HC in CD4+ T cell were highly enriched in SLE and rheumatoid arthritis diseases, etc, immune-related diseases (B). KEGG pathway analysis of the DEGs between SLE patient and health control in CD4+ T cell (C). The top 20 pathways were displayed. Size of the circles depicts the gene count, and its color depicts significance levels, the number 12–16 in the label is equivalent to -log *p*-value. Protein-protein interaction networks analysis of the DEGs between SLE patient and health control in CD4+ T cell (D). The top 50 genes were displayed base on the degree of complexity of the nodes.



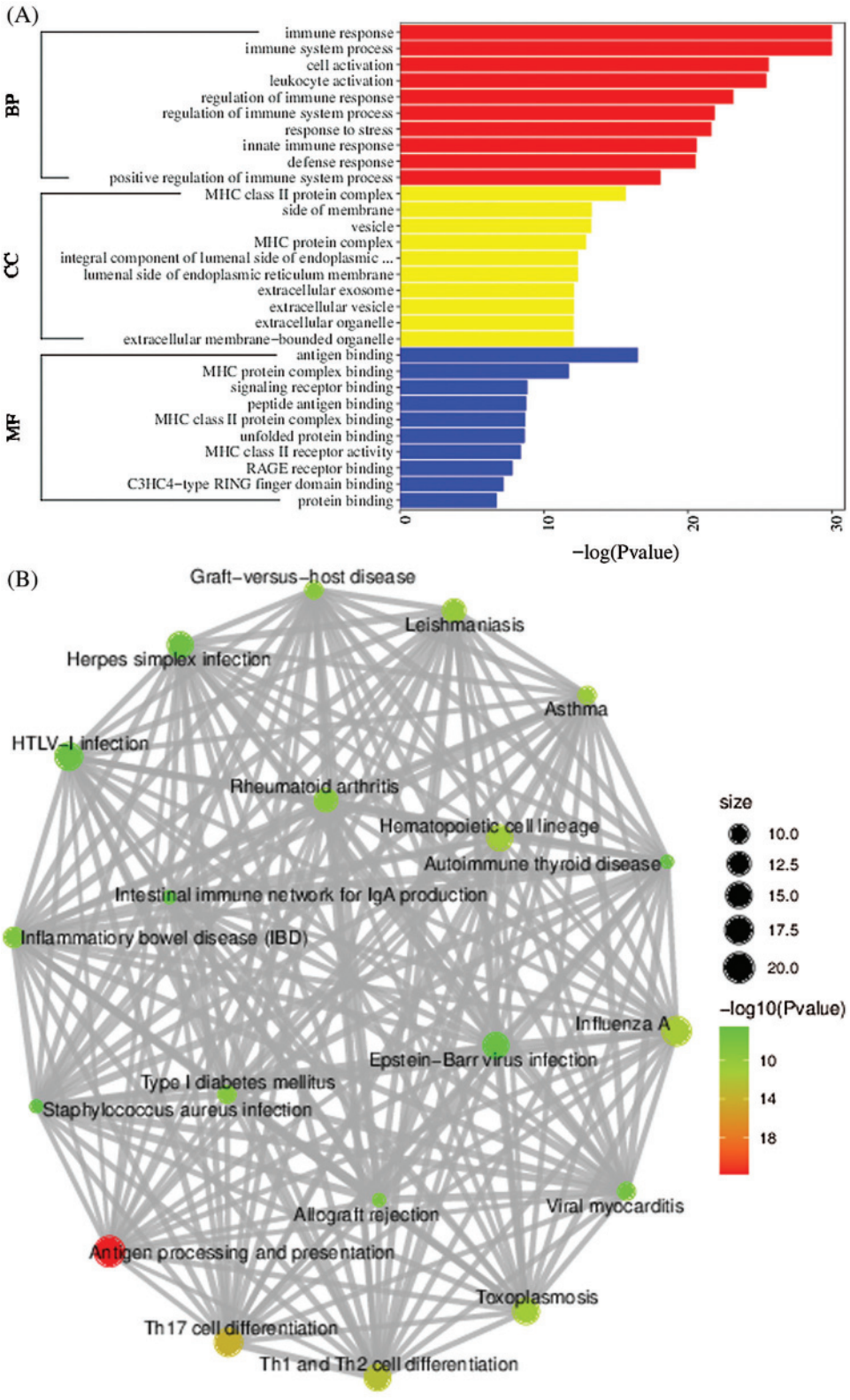
SUPPLEMENTARY FIGURE 14. Analysis of differences in gene expression variance between SLE patient and health control in the CD8+ T cells. Gene ontology enrichment analysis of the differentially expressed genes (DEG) between SLE patient and health control in CD8+ T cell (A). DEGs between SLE and HC in CD8+ T cell were highly enriched in SLE and rheumatoid arthritis diseases, etc, immune-related diseases (B). KEGG pathway analysis of the DEGs between SLE patient and health control in CD8+ T cell (C). The top 20 pathways were displayed. Size of the circles depicts the gene count, and its color depicts significance levels, the number 12-16 in the label is equivalent to -log *p*-value. Protein-protein interaction networks analysis of the DEGs between SLE patient and health control in CD8+ T cell (D). The top 50 genes were displayed base on the degree of complexity of the nodes.



SUPPLEMENTARY FIGURE 15. Analysis of differences in gene expression variance between SLE patient and health control in the monocytes. Gene ontology enrichment analysis of the differentially expressed genes (DEG) between SLE patient and health control in monocytes (A). DEGs between SLE and HC in monocytes were highly enriched in SLE and rheumatoid arthritis diseases, etc, immune-related diseases (B). KEGG pathway analysis of the DEGs between SLE patient and health control in monocytes (C). The top 20 pathways were displayed. Size of the circles depicts the gene count, and its color depicts significance levels, the number 12-16 in the label is equivalent to -log P-value. Protein-protein interaction networks analysis of the DEGs between SLE patient and health control in monocytes (D). The top 50 genes were displayed base on the degree of complexity of the nodes.



SUPPLEMENTARY FIGURE 16. Analysis of differences in gene expression variance between SLE patient and health control in the myeliod cell. Gene ontology enrichment analysis of the differentially expressed genes (DEG) between SLE patient and health control in myeliod cell (A). KEGG pathway network analysis of the DEGs between SLE patient and health control in myeliod cell (B).



SUPPLEMENTARY FIGURE 17. Analysis of differences in gene expression variance between SLE patient and health control in the NK cell. Gene ontology enrichment analysis of the differentially expressed genes (DEG) between SLE patient and health control in NK cell (A). KEGG pathway network analysis of the DEGs between SLE patient and health control in NK cell (B).

SUPPLEMENTARY TABLE 1

The cell number and percentage of each cluster in SLE and HC

Cluster	HC		SLE	
	Number	%	Number	%
1	985	12.37	2099	26.34
2	1033	12.97	1333	16.73
3	1538	19.31	662	8.31
4	1238	15.55	941	11.81
5	1312	16.48	534	6.70
6	216	2.71	710	8.91
7	550	6.91	343	4.30
8	347	4.36	425	5.33
9	340	4.27	350	4.39
10	66	0.83	420	5.27
11	322	4.04	63	0.79
12	16	0.20	88	1.10

SUPPLEMENTARY TABLE 2

The top 20 marker genes in each of the immune cell types

B cell	$CD4^+$ T cell	$CD8^+$ T cell	NK cell	Myeliod cell
IGLC2	LTB	CCL5	GNLY	LYZ
IGKC	IL7R	GZMH	CMC1	S100A9
HBB	PPBP	CD8B	PRF1	S100A8
IGLC3	MAL	NKG7	XCL2	CST3
HBA1	TRAT1	CD8A	CLIC3	FCN1
HBA2	TSHZ2	GZMA	GZMB	RP11-1143G9.4
HLA-DRA	LDHB	FGFBP2	KLRF1	CXCL8
CD74	AQP3	CST7	NKG7	TYROBP
IGHA1	RGCC	RP11-291B21.2	CCL4	S100A12
CD79A	SOCS3	CTSW	SPON2	CSTA
IGHG1	CORO1B	TRGC2	FCGR3A	G0S2
JCHAIN	GPR183	GZMK	CTSW	CCL3
IGHM	NOSIP	GZMM	KLRB1	FCER1G
MS4A1	CREM	CD3D	TRDC	LST1
HLA-DQB1	ARID5B	GZMB	KLRD1	IFITM3
IGHG4	SARAF	HCST	TYROBP	AIF1
HLA-DPB1	LEF1	C12orf75	HOPX	CTSS
CD79B	DEFA3	LYAR	FGFBP2	TYMP
HLA-DPA1	TMEM123	IL32	FCER1G	IL1B
IGHD	ZFP36L2	KLRG1	CST7	LGALS1