

# Gene-regulatory network study of rheumatoid arthritis in single-cell chromatin landscapes of peripheral blood mononuclear cells

**Cantong Zhang**

Shenzhen People's Hospital, Jinan University, Southern University of Science and Technology

**Xiaoping Hong**

Shenzhen People's Hospital, Jinan University, Southern University of Science and Technology

**Haiyan Yu**

Shenzhen People's Hospital, Jinan University, Southern University of Science and Technology

**Hongwei Wu**

the First Affiliated Hospital of Jinan University

**Huixuan Xu**

Shenzhen People's Hospital, Jinan University, Southern University of Science and Technology

**Xiaofen Qiu**

Shenzhen People's Hospital, Jinan University, Southern University of Science and Technology

**Wanxia Cai**

Shenzhen People's Hospital, Jinan University, Southern University of Science and Technology

**Berthold Hocher**

Mannheim Heidelberg University

**Weier Dai**

University of Texas at Austin

**Donge Tang**

Shenzhen People's Hospital, Jinan University, Southern University of Science and Technology

**Dongzhou Liu**

Shenzhen People's Hospital, Jinan University, Southern University of Science and Technology

**Yong Dai** (✉ [daiyong22@aliyun.com](mailto:daiyong22@aliyun.com))

Shenzhen People's Hospital, Jinan University, Southern University of Science and Technology

---

## Research Article

**Keywords:** rheumatoid arthritis, peripheral blood mononuclear cells, single-cell chromatin accessible assay, transcription factor

**Posted Date:** September 17th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-899470/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

Rheumatoid arthritis is a chronic autoinflammatory disease with an elusive etiology. Assays for transposase-accessible chromatin with single-cell sequencing (scATAC-seq) contribute to the progress in epigenetic studies. However, the impact of epigenetic technology on autoimmune diseases has not been objectively analyzed. Therefore, scATAC-seq was performed to generate a high-resolution map of accessible loci in peripheral blood mononuclear cells (PBMCs) of RA patients at the single-cell level. The purpose of our project was to discover the transcription factors (TFs) that were involved in the pathogenesis of RA at single-cell resolution. In our research, we obtained 22 accessible chromatin patterns. Then, 10 key TFs were involved in the RA pathogenesis by regulating the activity of MAP kinase. Consequently, two genes (PTPRC, SPAG9) regulated by 10 key TFs were found that may be associated with RA disease pathogenesis and these TFs were obviously enriched in RA patients ( $p < 0.05$ ,  $FC > 1.2$ ). With further qPCR validation on PTPRC and SPAG9 in monocytes, we found differential expression of these two genes, which were regulated by eight TFs (ZNF384, HNF1B, DMRTA2, MEF2A, NFE2L1, CREB3L4 (var. 2), FOSL2::JUNB (var. 2), MEF2B). What is more, the eight TFs showed highly accessible binding sites in RA patients. These findings demonstrate the value of using scATAC-seq to reveal transcriptional regulatory variation in RA-derived PBMCs, providing insights on therapy from an epigenetic perspective.

## Introduction

Rheumatoid arthritis (RA) is a chronic autoinflammatory disease distinguished by autoantibodies to citrullinated proteins and affecting approximately 1% of the world's population. Without sufficient treatment, RA can contribute to disabling joint damage and systemic disorders due to chronic synovial inflammation with synovial cell proliferation and pannus formation<sup>[1, 2]</sup>.

Mounting evidence indicates that epigenetic modifications play significant roles in regulating RA pathogenesis<sup>[3]</sup>. We currently know that more than 100 loci are associated with RA risks, such as HLA RB1, TRAF1, PSORS1C1, and microRNA 146a, which are variously related to joint damage<sup>[4]</sup>. Most recently, Christopher Loh successfully applied an assay for transposase-accessible chromatin by high-throughput sequencing (ATAC-seq) to discover a novel mechanism in which some motifs in approximately 80 genes were more accessible, which facilitated sustained chromatin activation by tumor necrosis factor (TNF), leading to continuous uninterrupted synovitis in RA<sup>[5]</sup>. Simultaneously, using ATAC-seq, Krishna V confirmed that JQ1 can inhibit the proliferation and activation of fibroblast-like synoviocytes (FLS) from RA patients in vitro, which indicated an alternative treatment for RA<sup>[6]</sup>.

To better understand RA etiology, we need to study the gene regulatory network of RA, which will reveal which genes and cell types play roles in RA<sup>[4]</sup>. There are some loose regions in chromatin in eukaryotic genomes, including those containing active regulatory elements, which control gene activity and are relatively accessible, thereby playing roles in disease<sup>[7]</sup>. By analyzing active regulatory elements, such as enhancers, promoters, and other regulatory sequences, we can profile a map of open, active DNA. In

addition, a map of open, active DNA in a range of cell types can be profiled by techniques such as DNase-seq and ATAC-seq, which can identify both active and open chromatin<sup>[4]</sup>. Here, we introduce an advanced technique called assay for transposase-accessible chromatin in single cells (scATAC-seq). Using Tn5 transposase, accessible regions of DNA fragments are acquired, and scATAC-seq can capture individual cells using a programmable microfluidics platform (Fluidigm). Then, libraries are produced via PCR in a specific amplification protocol. As a result, a high-resolution map of accessible loci of single cells is profiled<sup>[8]</sup>. Hence, researchers can discover cell-type-specific biological functions and biomarkers of diseases, which are often obscured by the average results of the overall cell population, using this technique<sup>[9]</sup>.

Peripheral blood mononuclear cells (PBMCs), mainly lymphocytes and monocytes, are related to [oxidative stress](#) and [inflammatory response](#). Hence, we collected and purified PBMCs from RA patients and healthy donors. Subsequently, we used the scATAC-seq method to profile a high-resolution map of accessible chromatin. Then, we performed GO and KEGG analyses to explore RA-related target genes and significant signaling pathways to provide a better understanding of the pathogenesis of RA.

## Methods

### Sample collection

According to the 2010 ACR/EULAR criteria for classifying RA, a total score of more than 6 points can be classified and diagnosed as RA. As a disease involving multiple systems, peripheral blood mononuclear cells (PBMCs), which are associated with oxidative stress and inflammatory response, can be readily accessed for assaying<sup>[34]</sup>. Seven patients with RA were enrolled at Shenzhen People's Hospital, and 8 ml of fresh fasting peripheral blood was collected intravenously and injected into heparin anticoagulant tubes for 8 h. In addition, seven healthy individuals of matched age and sex were selected, and peripheral blood was collected from these participants to serve as controls (Table 1). None of the patients were treated with standard treatment or immunosuppressive therapy within 3 months; furthermore, potential subjects were excluded for infections, tumors, diabetes, and other rheumatic diseases. This study was approved by the Medical Ethics Committees of Shenzhen People's Hospital, China (LL-KY 2019514), and all donors signed informed written consent.

### Cell isolation

An equal volume of normal saline was added to the peripheral blood at room temperature and mixed and diluted. By adding [equal proportions](#) of solution (GE Healthcare, 17-1440-03), PBMCs were separated and then subjected to density-gradient centrifugation (2700 g, 25 minutes, 25°C). Then, red blood cell lysis buffer was used to remove the remaining red blood cells. PBS containing 0.04% BSA was added, and the PBMCs of the RA group and the healthy group were mixed. Each concentration was reduced to 1 ml. Blood cell counts were used, and samples were cryopreserved at 4°C until use.

## Raw Data of scATAC-seq Processing

All protocols for data processing were available on the following internet site:

<https://support.10xgenomics.com/single-cell-atac/software/pipelines/latest/algorithms/overview>. The major steps were as follows:

## Droplet-based single-cell ATAC-seq (scATAC-seq)

Nuclear suspensions were incubated in a transposition mix that included Tn5 transposase, which enabled fragmentation of DNA in open regions of the chromatin and marking them with adapter sequences. Then, GEMs were produced by mixing barcoded gel beads, transposed nuclei, a master mix, and partitioning oil on Chromium Chip E (10x Genomics, CG000168). The nuclei were delivered at limiting dilution, such that the majority (90-99%) of generated GEMs contained no nuclei, to obtain single nucleus resolution, while the remainder largely contained a single nucleus. Then, silane magnetic beads were utilized to remove excess biochemical reagents from the post-GEM reaction mixture. Solid phase reversible immobilization (SPRI) beads were used to remove unutilized barcodes from the sample. During material amplification by PCR, the libraries were indexed by P7 (a sample index) and Read 2 (Read 2N) sequences. Therefore, the final libraries contained the P5 and P7 primers used in Illumina® bridge amplification.

## Barcode processing and genome alignment

The occasional sequencing error in barcodes obtained from the 'I2' index read was fixed to improve data quality. For example, if barcodes outside the whitelist had a Hamming distance less than two and the probability that they are real barcodes was higher than 90%, they were corrected to be whitelist barcodes<sup>14</sup>. The Cell Ranger ATAC pipeline performs reference-based analysis and requires adapter and primer oligo sequences to be trimmed off before confident mapping. We utilized the cutadapt<sup>[35]</sup> tool to identify the reverse complement of the primer sequence, which could exist at the end of a read sequence. We trimmed it from the read before alignment. Then, the trimmed read pairs were aligned to a specified reference using BWA-MEM<sup>[36]</sup> with default parameters. Then, duplicate reads were identified as groups of read pairs across all barcodes, where the 5' ends of both R1 and R2 had identical mapping positions as the reference, while other read pairs were detected only as fragments.

## Peak Analysis, Peak-Barcode Matrix and t-SNE projection

As each fragment end was indicative of regions of open chromatin, we analyzed the combined signals from these fragments to determine regions of the genome enriched for open chromatin and hence have putative regulatory and functional significance. Peak calling was performed through followed method: First, we counted the number of transposition events at each base pair along the genome;

second, a smoothed profile of these events with a 401 bp moving window around each base pair was produced and combined with a ZINBA-like mixture model, which includes a binarization algorithm; third, a region is identified as a peak signal (enriched for open chromatin) or noise, which depends on the signal threshold that is set at an odds ratio of 1/5. Consequently, peaks 500 bp apart are combined to generate a peak-containing BED file for position classification.

According to the number of peak areas with overlapping fragments, we separated the signal and noise for each barcode. To simulate and identify whitelist contamination, we subtracted the fixed count relative to a depth from all the bar code counts. Setting an odds ratio of 1000, we separated the barcodes that corresponded to real cells from noncell barcodes. Subsequently, a raw peak barcode matrix containing each barcode of chromatin open richness was produced and then used for subsequent analysis, such as dimensional reduction, clustering, and visualization.

For binarization of the scATAC-seq data, all cells merge the peak calling which was performed on reads, if at least one reading with overlapping peaks, marks the peak '1' (open), or for the '0' (closed) by first by term frequency-inverse document frequency (TF - IDF) to a normalized binary counter matrix, and then to perform Singular-Value Decomposition (SVD) to form the Latent semantic indexing analysis only after the SVD of 2-50 d was passed to the t-SNE visualization. We also provided graph-based clustering and visualization via t-SNE. We normalized the data to the unit norm before performing graph-based clustering and t-SNE projection.

## **Peak-Related Gene Identification by GO Enrichment and KEGG Enrichment Analyses**

As a gene functional classification system, Gene Ontology (GO) reports annotations in three categories: molecular function, cellular component, and biological process. GO enrichment analysis reports all GO terms that are significantly enriched in peak-related genes compared to their levels in the genome backgrounds and filters the peak-related genes that correspond to biological functions. First, all peak-related genes were mapped to GO terms in the Gene Ontology database (<http://www.geneontology.org/>), gene numbers were calculated for every term, and significantly enriched GO terms in peak-related genes compared to the genome background were defined by hypergeometric testing. To study gene biological functions, we also performed KEGG enrichment analysis to identify vital signal transduction pathways with peak-related genes.

## **TF Motif Enrichment Analysis and Differential Accessibility Analysis**

As transcription factors (TFs) bind through specific motifs, we performed enrichment analysis for these motifs in the peaks. To identify their accessibility and specific TF motifs for each cluster, Cell Ranger

ATAC determines whether the average within a cluster is different than the average outside the cluster for each motif and each cluster. To find differentially accessible motifs between cell groups, Cell Ranger ATAC uses the fast-asymptotic beta test, which is utilized in edgeR. For each cluster, the algorithm ran on the cluster with all other cells, producing a list of genes that are differentially expressed in the cluster relative to the remainder of the sample.

## Validation by qPCR

Three healthy donors (male: female=1:2, mean age  $28\pm 3$  years) and three RA patients (male: female=1:2, mean age  $56\pm 11$  years) were involved in the qPCR validation experiments. Then, we collected 10 ml of peripheral blood from each donor for isolating PBMCs. After that, using CD14 microbeads (microbeads conjugated to monoclonal anti-human CD14 antibodies, Miltenyi), monocytes were isolated from PBMCs. Using RNAiso Plus (TAKARA 9109), chloroform, and isopropanol, the RNA was separated from monocytes. Then, total RNA was reversed transcribed into cDNA. Then, the cDNA was detected by real-time PCR. The relative expression of genes in monocytes was used by Student's t-test.

The primer sequences of target genes were as follows: SPAG9 forward, GCTGGAGATGGATTGCTTACAC, SPAG9 reverse, TTCAGGCTGGTATGAGAACGTG; PTPRC forward, GCCCTGCTTGTGTTCTCTAC, PTPRC reverse, TGGCTCCACATTCATCAGTTG; T- $\beta$ -actin forward, CGAGGCCAGAGCAAGAGAG, T- $\beta$ -actin reverse, GGCCACACGCAGCTCATTG.

# Results

## Quality Control of scATAC-seq

RA group and NC group were acquired and analyzed by scATAC-seq as the workflow shown. After quality control, 16,407 cells and 5,344 unique fragments were retained. At single-cell resolution, RA\_PBMC enriched TSS fragments four times in the proximal region relative to the distal region, and NC\_PBMC enriched TSS fragments five times, which reflected that the proportion of fragments captured in the open chromatin was higher than that in the closed chromatin. Besides, we noted distribution characteristics of nucleosome fragment between the RA group and NC group. The recovered library from RA\_PBMC was sequenced to an average depth of 22,743 raw reads per cell, generating chromatin accessibility profiles for 8014 cells with a median of 5,345 unique fragments per cell. Simultaneously, the library of NC\_PBMC generates profiles of 8,393 cells with 23,035 original reads per cell and 5,344 unique fragments per cell.

## Identification of primary cell types among PBMCs

The RA group and NC group were acquired and analyzed by scATAC-seq as shown in the workflow (Figure 1a). Quality control of the scATAC-seq profile was showed in figure 1b and figure 1c. In detail, we observed the signal distribution map around TTS after normalization of NC\_PBMCs and RA\_PBMCs library and the length distribution of corresponding inserts for each sample. Here, we used the activity of known cell marker genes to identify and annotate 9 clusters (Figure 1b), including T lymphocytes (T

cells; clusters 0, 1, 3, and 5), natural killer cells (NK cells; cluster 2), monocytes (cluster 4), B lymphocytes (B cells; cluster 6) and dendritic cells (DCs; cluster 8). More specifically, T cells can be identified by CD3D, CD3G, LEF1, CD8A and IL2RA<sup>[10, 11]</sup>; NK cells by GZMB, NKG7, and KLRD1<sup>[10, 12]</sup>; monocytes by CD68, CD14, and ITGAM<sup>[10, 13]</sup>; B cells by CD79A, CD19 and MS4A1<sup>[10, 12]</sup>; and DCs by IL3RA<sup>[10]</sup>. Select marker genes and the corresponding expression quantities are shown in Figure 1c.

In addition, scATAC-seq enables cell type identification by cell type-specific transcription factor motifs<sup>[14]</sup>. For the fragments that overlapped the list of TF motifs, annotations of cell-type specificity were created for the most significantly enriched TF motifs ( $P$ -value < 0.01, fold change value, FC>1.5) in each cluster with no apparent difference between the RA\_PBMC and NC\_PBMC groups to serve as an alternative identifying feature for each cluster. Notably, more than 20 TF motifs in B cells were identified. Thus, we selected the top 10 TF motifs: POU2F3, IRF1, POU5F1B, STAT1:: STAT2, POU2F, POU1F1, POU2F2, POU3F3, POU5F, and POU3F1 (Figure 2a). In addition, we found 13 TF motifs in NK cells, including EOMES, TBX2, TBR1, TBX20, and TBX21 (Figure 2a). Furthermore, we found 95 TF motifs in monocytes. The top 10 TF motifs in monocytes were FOSL2::JUN, FOSL1::JUN, FOS, FOSL1, FOSL1::JUND, FOSL1::JUNB, JUN(var.2), JUNB, FOS::JUND, and JUND (Figure 2a), while there were 20 TF motifs including GATA3, ELF2, ELK1, CEBPB, ELK3, GATA5, FEV, CEBPE, CEBPA, ELF5 in DCs. Interestingly, 7 TF motifs were identical in DCs and monocytes: CEBPB, CEBPE, CEBPA, ELF5, NFIL3, CEBPG, and HLF (Figure 2b), because some DCs differentiate from monocytes. For T cells, there was no TF motif with an FC value higher than 1.5. In summary, we concluded the identified markers genes and type-specific transcription factor motifs in the scATAC-seq experiments (Table 2).

### **Differential chromosome accessibility analysis in rheumatoid arthritis**

The distribution of immune cells and their regulatory functions are variously related to the occurrence of autoimmune diseases. Comparing the ratios of immune cells in the PMBC\_RA with PBMC\_NC groups (Table 3, Figure 2c), a significant difference was observed between B cells and T cells (Chi-square test,  $P$ -value < 0.01, FDR<0.05). Notably, the ratio of T cells in the PBMC\_RA group was lower than that in the PBMC\_NC group, implying a possible mechanism by which fewer T cells curb inflammation in the RA environment, a finding that is consistent with the hypothesis suggesting that regulatory CD8+ T cells play a role in preventing the development of autoimmune diseases<sup>[15]</sup>. In contrast, there were more B cells in the RA group than in the NC group ( $P$ -value < 0.01, FDR<0.05), suggesting that B cells in peripheral blood can promote RA pathogenesis by changing the number of cells. When calculating the different peaks ( $P$ -value < 0.05, FC>1.2) in the PBMC\_NC and PBMC\_RA groups, we found no different peaks in T cells, 51 in NK cells, 11 in monocytes, 149 in B cells, and 665 in DCs (Figure 2d). Moreover, the differential accessibility of TF motifs ( $P$ -value < 0.05, FC>1.2) for each immune cell type was selected. Consequently, we identified 56 TF motifs in RA, including three motifs (CEBPG (var. 2), NFE2L1, MAF:: NFE2) in monocytes, six motifs (NRF1, PHOX2A, TCFL5, ZBTB14, PHOX2B, PROP1) in B cells, one motif (TCFL5) in T cells and 46 motifs in NK cells (Figure 2e). However, no differential TF motifs were observed in DCs, indicating that the transcriptome signatures of these cells did not change in the RA environment.

We identified a total of 22 subclusters among B cells, DCs, monocytes, NK cells, and T cells (Figure 3a) with further cluster analysis. The results show that the cell number ratios of subcluster 0 B cells (B-0), subcluster 3 NK cells (NK-3), and subcluster 3 T cells (T-3) in the PBMC\_RA group were more eight-fold greater than those in the PBMC\_NC group. In addition, the cell number ratios in B-2, monocyte-1, NK-1, and T-5 in the PBMC\_RA group were also higher than in the PBMC\_NC group. However, the cell number ratio of subcluster 3 B cells (B-3), subcluster three monocyte cells (monocyte-3), subcluster 0 natural killer cells (NK-0), subcluster four natural killer cells (NK-4), subcluster 0 T cells (T-0), subcluster 4 T cells (T-4) and subcluster 6 T cells (T-6) in the PBMC\_RA group was smaller than that in the PBMC\_NC group, especially for NK-0 and NK-4, in which it was decreased by more than three-fold. Furthermore, no obvious difference in the number of cells was observed in the remaining subclusters in the PBMC\_RA and PBMC\_NC groups. We profiled a figure for the cell ratio, as mentioned above (Figure 3b). Furthermore, the differential accessibility of TF motifs ( $P$ -value < 0.05, FC>1.2) in each subgroup of PBMCs is shown in Figure 3c.

### **Functional analysis of significantly enriched peaks in the PBMC\_RA group**

According to the epigenomic analysis, we selected peaks with an obvious fold change value higher than 1.2 for the GO and KEGG analysis. Based on the KEGG analysis, NKs participate and mediate RA progression through virus infection-related pathways, such as human CMV, human immunodeficiency virus 1, Kaposi sarcoma-associated herpesvirus, Epstein-Barr virus, and human T-cell leukemia virus (Figure 3d). In addition, NKs were also active in Th17 cell differentiation. This result was related to another study showing that helps T (Th) cells participate in RA pathogenesis, especially Th17 cells, which produce IL-21, IL-22, and TNF $\alpha$ <sup>[16]</sup>. Similarly, we found that DCs are regulated through human cytomegalovirus (CMV) infection (Figure 3e), a finding consistent with research showing that RA patients usually have CMV<sup>[17]</sup>.

Similarly, after further clustering of five major types of PBMCs, we found cell-type-specific functions through GO enrichments considering peak-related genes in subclusters, except for the B-2 and T-1 subclusters. Here, the B-0 and B-3 subcluster genes were found to play roles in T cell activation. At the same time, B-1 genes are involved in B cell differentiation and activation. In addition to T cell activation, B-3 genes also participate in regulating cell-cell adhesion, negative regulation of protein phosphorylation, regulation of MAP kinase activity, and T cell activation regulation. In monocytes, M-0 genes are involved in the negative regulation of phosphorylation, regulation of MAP kinase activity, neutrophil degranulation, activation, and mediation of immunity. M-1 genes are regulated through DNA-binding transcription activator activity; RNA polymerase II-specific. M-2 genes are active in histone modification, T cell activation, and differentiation.

Similarly, the M-3 subcluster genes participate in dephosphorylation, regulation of GTPase activity, regulation of small GTPase-mediated signal transduction, positive regulation of cell adhesion, and T cell activation. In NK cells, peak-related genes in the NK-0 subcluster are mainly involved in the endosome membrane; NK-1 genes participate in the regulation of mitotic cell cycle phase transition and kinase

regulator activity; NK-2 genes take part in Ras protein signal transduction, regulation of small GTPase-mediated signal transduction, and T cell activation; NK-3 genes have effects on the cellular response to UV; NK-4 genes are mediated through T cell activation, regulation of lymphocyte activation, positive regulation of cytokine production and regulation of GTPase activity. With regard to subtypes of T cells, genes are abundant in the ubiquitin ligase complex in the T-0 subcluster, focal adhesion in T-2, and the Cul4-RING E3 ubiquitin ligase complex in T-3. In addition, considering biological process, the T-2 subcluster genes are active in cell-substrate adherens junction; T-4 in autophagy; T-5 in the regulation of cellular amide metabolic process and p38 MAPK cascade; T-6 in T cell activation and positive regulation of cytokine production; and T-7 in I-kappaB kinase/NF-kappaB signaling (supplemental figure 4g).

Specifically, B-3 and M-0 genes are active in the same pathway of regulation as MAP kinase activity (Figure 4a and Figure 4b), which is enriched by 33 genes in the B-3 subcluster and 25 genes in the M-0 subcluster. Among these genes, five genes are in both the B-3 and M-0 subclusters: MAP3K3, SPAG9, PTPRC, PTPN1, and SRC.

### **The seminal discovery of TFs in the PBMC\_RA group**

A total of 434 significantly TF-enriched TF motifs were identified in the RA\_PBMC group (P-value < 0.05, FC>1.2) through further cell clustering based on the scATAC-seq analysis. We discovered two (PTPRC and SPAG9) of the 5 intersecting genes with differential accessible TF-binding sites in B-3 and M-0 subclusters of RA patients compared to healthy controls (P-value < 0.05, FC>1.2). In addition, PTPRC related to RA was identified from Gene Data Mining to Disease Genome Sequence Analysis ([www.genecards.org](http://www.genecards.org)), and the relevance scores were higher than 9. More specifically, 71 TFs could be involved in regulating two genes (PTPRC and SPAG9), and only 10 TFs had highly accessible binding sites (GATA6, IRF2, ZNF384, HNF1B, DMRTA2, MEF2A, NFE2L1, CREB3L4 (var. 2), FOSL2::JUNB (var. 2), and MEF2B). Therefore, 2 TFs (GATA6 and IRF2) in B-3 and the 8 TFs in M-0 (ZNF384, HNF1B, DMRTA2, MEF2A, NFE2L1, CREB3L4 (var. 2), FOSL2::JUNB (var. 2), and MEF2B) contributed for RA pathogenesis by regulating corresponding target genes and MAP kinase activity (Figure 4c and 4d). To figure out the cell traits of these subcluster, we further identified the enriched motifs in these subcategories in PBMC\_(p<0.05, FC>1.2) (Table 4).

To validate the expression of two genes in monocytes (PTPRC, SPAG9), we took the peripheral blood of another three RA patients and healthy donors. Using these samples, monocytes were isolated and used for extracting their RNA for qPCR experiments. Compared to healthy controls, differential expression of two genes (SPAG9 and PTPRC) was observed in RA patients (SPAG9,  $1.673 \pm 0.8945$  vs.  $1.013 \pm 0.1673$ ,  $p = 0.0045$ ; PTPRC,  $0.5878 \pm 0.3841$  vs.  $1.007 \pm 0.1203$ ,  $p = 0.0066$ ) (Figure 4e).

## **Discussion**

With the rise of single-cell sequencing technology, single-cell epigenomics sequencing has attracted more and more attention, and scATAC-seq is also the absolute new favorite in the field of epigenetics. The research identified 85 accessible chromatin patterns and 400,000 regulatory elements from 13 mouse

tissues by scATAC-seq and found the cell type reflected by chromatin accessibility varies according to anatomical coordinates<sup>[18]</sup>. More biological information can be revealed at the cell level by studying single-cell heterogeneity. And scATAC-seq can be used to study migratory behavior in cancer biology. Meanwhile, this technology on autoimmune disease has not been profoundly applied.

To better understand RA epigenetics [pathogenesis](#), we performed scATAC-seq on peripheral blood mononuclear cells (PBMCs) from RA subjects to profile an epigenetic landscape of active regulatory DNA, which can be used to distinguish different cell subtypes, identify cell-type-specific TF enrichment, and reveal a regulatory pattern. In this way, T cells, B cells, NK cells, monocytes, and DCs were identified without antibodies. Furthermore, as part of the epigenetic signature identifies the aforementioned PBMC types, cell-type-specific TF motifs were also identified. The discovery of TF motifs ( $P$ -value < 0.01, FC>1.5) may promote parsing of cellular heterogeneity by identifying immune cell types.

As above description (figure 3d and 3e), NK cells and DC cells both played a role in RA progression, which was associated with some viruses. We hypothesized that as extrinsic antigens these viruses worked through molecular simulation mechanisms, inducing an autoimmune response in RA patients. And this hypothesis was consistent with other research. In this way, we confirmed the importance of using the single-cell sequencing technique and its reliability to reveal the traits of immune cells in RA patients.

With further cell reclustering, we identified a total of 22 subclusters among the five main cell types. In this way, we obtained more detailed and differential data compared with the five central immune cells. For example, little information could be obtained from T cells through the GO enrichment analysis of peak-related genes ( $P$ -value < 0.05, | FC |>1.2). In contrast, GO analysis showed that T cell subtypes showed varied functions, biological processes, and cellular components. Furthermore, no TF motifs were found in T cells with an FC value higher than 1.2, but we observed 89 and 11 enriched motifs ( $P$ -value < 0.05, FC>1.2) in the T-6 and T-7 subclusters, respectively. Single-cell sequencing has been shown to provide a higher resolution of cellular differences and enables a more in-depth exploration of the function of an individual cell<sup>[19]</sup>.

Specifically, according to GO analysis, we found that 32 genes in the T-2 subcluster cells showed an effect on focal adhesion, cell-substrate junctions, and cell-substrate adherens junctions. Considering previous research results<sup>[20, 21]</sup>, these three pathways contribute to the migration of fibroblast-like synoviocytes (FLSs) in RA. In addition, T-4 subcluster cells participate in autophagy with 20 genes associated with various autoimmune diseases, including RA<sup>[22]</sup>. Dysregulation of the autophagy pathway plays a vital role in RA pathogenesis through peptide citrullination, osteoclast activation, T/B cell activation, and RA FLS survival<sup>[22, 23]</sup>. If we better understand the autophagy pathway, we would be able to apply compounds that modulate the autophagy pathway to in vitro tests and develop new drugs. In addition, 12 genes in the T-5 subcluster cells were involved in regulating cellular amide metabolic processes, which may facilitate the metabolic regulation of the T cell response in RA<sup>[24]</sup>.

Furthermore, we observed that the pathway of the p38 MAPK cascade was enriched by 12 genes in T-5. Moreover, the p38 mitogen-activated protein kinase (MAPK) pathway has been strongly suggested to be associated with the pathology of RA<sup>[20]</sup>. In the T-6 subcluster, a reduced cell number ratio was evident, and 59 genes and 54 genes participated in the T cell activation pathway and positive cytokine production pathway, respectively. Therefore, we speculated that T-6 promotes the development of inflammation in RA in these two ways.

In the subtypes of B cells, we found that B-0 and B-3 both played a role in T cell activation with 26 genes and 46 genes, consistent with a previous study showing that the T cell activation in RA was B cell dependent<sup>[21]</sup>. However, the cell number changed in an opposite trend, which showed B cell dysregulation. While B-0 and B-2 cells were obviously increased in the RA group compared to the NC group, B-3 cells were more profoundly decreased. Our results may indicate that protective (B-3) and pathogenic (B-0) B cell populations exist in vivo. Hence, past research has shown that the depletion of B-3 cells and the proliferation of B-0 cells both promote the development of RA<sup>[25]</sup>. Given this finding, anti-CD20 drugs may not be an advisable choice for use in treating RA, as they may not be able to distinguish protective B cells.

In monocytes, the M-0 subcluster genes were mainly embodied in immune-inflammatory pathways (e.g. neutrophil degranulation, neutrophil activation involved in immune response, neutrophil activation, neutrophil-mediated immunity, and regulation of MAP kinase activity). According to the literature, M-0 is interconnected with other cells by producing proinflammatory cytokines<sup>[26]</sup>. In addition, this finding suggests that M-0 genes may mediate the activation and degranulation of neutrophils to play a role in RA pathological processes. In the M-2 subcluster, the biological processes identified were mainly T cell differentiation, lymphocyte differentiation, T cell activation, and histone modification. This finding suggests that M-2 cells have more significant local chromatin accessibility changes through histone modification than other RA cells, contributing to additional, different peaks.

Furthermore, immune cells can undergo epigenetic changes through histone modification<sup>[3]</sup>. M-3 subcluster cells participated in immune-inflammatory pathways (e.g. regulation of GTPase activity, regulation of small GTPase-mediated signal transduction, positive regulation of cell adhesion, T cell activation, and dephosphorylation). As reported by [Amanda Chan](#), Rac small GTPases mediate JNK activation to facilitate FLS proliferation and invasion, similar to tumor cells<sup>[27]</sup>, which are associated with M-3.

According to KEGG analysis, our findings focused on virus infection-related pathways and tumor-associated pathways in NK cells of RA patients. NK-0 showed activity in proteoglycan pathways in cancer and Salmonella infection. In addition, NK-1 participated in Epstein–Barr virus infection, human immunodeficiency virus 1 infection, human cytomegalovirus infection, and colorectal cancer. Infections and a high risk of virus-associated cancers increased RA susceptibility because most of the infectious agent proteins are similar to human polypeptides, as indicated by the molecular mimicry hypothesis<sup>[28]</sup>. In addition, our results showed that NK-1 was involved in apoptosis and the p53 signaling pathway, which was consistent with a study showing that antiapoptotic molecules, including the

transcriptional regulator (p53), facilitate resident synoviocyte hyperplasia and proliferation<sup>[29]</sup>. Furthermore, NK-2 and NK-4 both played a role in Th17 cell differentiation and Th1 and Th2 cell differentiation; NK-4 was also active in two inflammatory pathways, leukocyte transendothelial migration, and the TNF signaling pathway. These findings support the idea that T cells are critical to the pathogenesis of RA<sup>[16]</sup>.

Finally, B-3 and M-0 subcluster genes both participated in regulating MAP kinase activity. Mitogen-activated protein kinase (MAPK) signaling pathways are critical for regulating the production of proinflammatory cytokines, ultimately culminating in joint inflammation and destruction<sup>[30,31]</sup>. Thus, five genes that participate in this pathway are in both of these two subclusters (MAP3K3, SPAG9, PTPRC, PTPN1, and SRC), which may be key elements of RA disease. In addition, we discovered 10 TFs with highly accessible binding sites (GATA6, IRF2, ZNF384, HNF1B, DMRTA2, MEF2A, NFE2L1, CREB3L4 (var. 2), FOSL2::JUNB (var. 2), MEF2B) near two of these genes (PTPRC, SPAG9). The literature indicated that PTPRC has a strong association with the response to TNF inhibitors in RA<sup>[32]</sup>, meanwhile, European studies also supported that PTPRC was RA risk gene locus<sup>[4]</sup>. This result confirmed the reality of our findings. However, SPAG9 was found to be related to prostate cancer<sup>[33]</sup> and abnormally regulate the protein expressed in systemic sclerosis. To validate the expression of these genes in RA patients, using qPCR, we observed the existence of differential expression in monocytes. Thus, if the accessibility of these transcription factor binding sites (ZNF384, HNF1B, DMRTA2, MEF2A, NFE2L1, CREB3L4 (var. 2), FOSL2::JUNB (var. 2), MEF2B) were transformed, they could regulate PTPRC and SPAG9 and mediate MAP kinase activity. These results enhance our understanding of gene transcription regulatory roles played by different TF motifs. These cell type-specific regulatory patterns can drive cell identity and function. Moreover, these findings provide insights into epigenetic changes in RA\_PBMCs, the pathogenesis of RA, and the identity of potential therapeutic targets.

In the future, single-cell RNA sequencing for further study of RA and potential drug molecules targeted to RA pathogenesis will be possible. In other words, single-cell multimodal omics is a suitable method for further study to reveal a complicated relationship in different dimensions.

## **Declaration**

### **ETHICS STATEMENT**

The study was performed in accordance with the ethical standards of the Declaration of Helsinki (1964) and its subsequent amendments. The studies involving human participants were approved by the Medical Ethics Committee of Shenzhen People's Hospital. The participants provided their written informed consent to participate in this study.

### **Consent for publication**

Not applicable.

## Availability of data and materials

The datasets supporting the conclusions of this article are available in the GEO repository, [unique persistent hyperlink to dataset in <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE183970>].

## Competing interests

The authors declare that they have no competing interests.

## Funding

This study was supported by grants from the National Science Foundation of Young Scientists of China (No. 31700795), the Key Research and Development Program of Guangdong Province (No. 2019B020229001), and the Sanming Project of Medicine in Shenzhen (No. SYJY201704 and No. SYJY201705), Shenzhen Key Medical Discipline Construction Fund (No. SZXK011), Guangxi Key Laboratory of Metabolic Diseases Research (20-065-76).

## ACKNOWLEDGMENTS

We are grateful to thank Li Chen from Hangzhou LC-BIO Co., Ltd. for assisting the authors in technical issues and analysis guidance.

## Author Contributions

Cantong Zhang interpreted the data and wrote the main manuscript text. Xiaoping Hong prepared all the figures. Haiyan Yu and Wu Hongwei both contributed to the interpretation of the data. Xiaofen Qiu and Wanxia Cai reviewed the manuscript. Huixuan Xu prepared figures 1-4. Berthold Hocher revised the manuscript. Weier Dai performed the data analysis. Donger Tang, Dongzhou Liu, and Yong Dai supervised the experiments and contributed to the interpretation of the data.

## References

1. Cribbs AP, Kennedy A, Penn Het *al*: **Methotrexate Restores Regulatory T Cell Function Through Demethylation of the FoxP3 Upstream Enhancer in Patients With Rheumatoid Arthritis.** *Arthritis & rheumatology (Hoboken, NJ)*2015, **67**(5):1182-1192.
2. Smolen JS, Aletaha D, Barton A, *et al*:**Rheumatoid arthritis.** *Nature reviews Disease primers*2018, **4**:18002.
3. Nemtsova MV, Zaletaev DV, Bure IV *et al*: **Epigenetic Changes in the Pathogenesis of Rheumatoid Arthritis.** *Frontiers in genetics*2019, **10**:570.
4. Okada Y, Eyre S, Suzuki A *et al*: **Genetics of rheumatoid arthritis: 2018 status.** *Annals of the rheumatic diseases*2019, **78**(4):446-453.

5. Loh C, Park S-H, Lee A *et al*: **TNF-induced inflammatory genes escape repression in fibroblast-like synoviocytes: transcriptomic and epigenomic analysis.** *Annals of the rheumatic diseases*2019, **78**(9):1205-1214.
6. Krishna V, Yin X, Song Q *et al*: **Integration of the Transcriptome and Genome-Wide Landscape of BRD2 and BRD4 Binding Motifs Identifies Key Superenhancer Genes and Reveals the Mechanism of Bet Inhibitor Action in Rheumatoid Arthritis Synovial Fibroblasts.** *Journal of immunology (Baltimore, Md : 1950)*2021, **206**(2):422-431.
7. Chen X, Shen Y, Draper W *et al*: **ATAC-seq reveals the accessible genome by transposase-mediated imaging and sequencing.** *Nature methods*2016, **13**(12):1013-1020.
8. Buenrostro JD, Wu B, Litzenburger U *et al*: **Single-cell chromatin accessibility reveals principles of regulatory variation.** *Nature*2015, **523**(7561):486-490.
9. Saliba A-E, Westermann AJ, Gorski S *et al*: **Single-cell RNA-seq: advances and future challenges.** *Nucleic acids research*2014, **42**(14):8845-8860.
10. Sinha D, Kumar A, Kumar H *et al*: **dropClust: efficient clustering of ultra-large scRNA-seq data.** *Nucleic acids research*2018, **46**(6):e36.
11. Zheng C, Zheng L, Yoo J-K *et al*: **Landscape of Infiltrating T Cells in Liver Cancer Revealed by Single-Cell Sequencing.** *Cell*2017, **169**(7).
12. Schelker M, Feau S, Du J *et al*: **Estimation of immune cell content in tumour tissue using single-cell RNA-seq data.** *Nature communications*2017, **8**(1):2032.
13. Sugita S, Makabe K, Iwasaki Y *et al*: **Natural Killer Cell Inhibition by HLA-E Molecules on Induced Pluripotent Stem Cell-Derived Retinal Pigment Epithelial Cells.** *Investigative ophthalmology & visual science*2018, **59**(5):1719-1731.
14. Satpathy AT, Granja JM, Yost K *et al*: **Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion.** *Nature biotechnology*2019, **37**(8):925-936.
15. Saligrama N, Zhao F, Sikora M *et al*: **Opposing T cell responses in experimental autoimmune encephalomyelitis.** *Nature*2019, **572**(7770):481-487.
16. Kotake S, Yago T, Kobashigawa T *et al*: **The Plasticity of Th17 Cells in the Pathogenesis of Rheumatoid Arthritis.** *Journal of clinical medicine*2017, **6**(7).
17. Bassyouni RH, Dwedar RA, Ezzat E *et al*: **Elevated Cytomegalovirus and Epstein-Barr virus burden in rheumatoid arthritis: A true pathogenic role or just a coincidence.** 2019, **41**(4).

18. Cusanovich DA, Hill AJ, Aghamirzaie D *et al*: **A Single-Cell Atlas of InVivo Mammalian Chromatin Accessibility.** *Cell*2018, **174**(5).
19. Eberwine J, Sul J-Y, Bartfai T *et al*: **The promise of single-cell sequencing.** *Nature methods*2014, **11**(1):25-27.
20. Schett G, Zwerina J, Firestein G: **The p38 mitogen-activated protein kinase (MAPK) pathway in rheumatoid arthritis.** *Annals of the rheumatic diseases*2008, **67**(7):909-916.
21. Takemura S, Klimiuk PA, Braun A *et al*: **T cell activation in rheumatoid synovium is B cell dependent.** *Journal of immunology (Baltimore, Md : 1950)*2001, **167**(8):4710-4718.
22. Karami J, Masoumi M, Khorramdelazad H *et al*: **Role of autophagy in the pathogenesis of rheumatoid arthritis: Latest evidence and therapeutic approaches.** *Life sciences*2020, **254**:117734.
23. Vomero M, Barbati C, Colasanti T *et al*: **Autophagy and Rheumatoid Arthritis: Current Knowledges and Future Perspectives.** *Frontiers in immunology*2018, **9**:1577.
24. Colamatteo A, Micillo T, Bruzzaniti S *et al*: **Metabolism and Autoimmune Responses: The microRNA Connection.** *Frontiers in immunology*2019, **10**:1969.
25. Marston B, Palanichamy A, Anolik JH: **B cells in the pathogenesis and treatment of rheumatoid arthritis.** *Current opinion in rheumatology*2010, **22**(3):307-315.
26. Rana AK, Li Y, Dang Q *et al*: **Monocytes in rheumatoid arthritis: Circulating precursors of macrophages and osteoclasts and, their heterogeneity and plasticity role in RA pathogenesis.** *International immunopharmacology*2018, **65**:348-359.
27. Chan A, Akhtar M, Brenner M *et al*: **The GTPase Rac regulates the proliferation and invasion of fibroblast-like synoviocytes from rheumatoid arthritis patients.** *Molecular medicine (Cambridge, Mass)*2007, **13**(5-6):297-304.
28. Arleevskaya MI, Kravtsova OA, Lemerle J *et al*: **How Rheumatoid Arthritis Can Result from Provocation of the Immune System by Microorganisms and Viruses.** *Frontiers in microbiology*2016, **7**:1296.
29. Baier A, Meineckel I, Gay S *et al*: **Apoptosis in rheumatoid arthritis.** *Current opinion in rheumatology*2003, **15**(3):274-279.
30. Son Y, Kim S, Chung H-T *et al*: **Reactive oxygen species in the activation of MAP kinases.** *Methods in enzymology*2013, **528**:27-48.
31. Thalhamer T, McGrath MA, Harnett MM: **MAPKs and their relevance to arthritis and inflammation.** *Rheumatology (Oxford, England)*2008, **47**(4):409-414.

32. Ferreiro-Iglesias A, Montes A, Perez-Pampin E *et al*: **Replication of PTPRC as genetic biomarker of response to TNF inhibitors in patients with rheumatoid arthritis.** *The pharmacogenomics journal* 2016, **16**(2):137-140.
33. Xiao C, Li M, Huang Q *et al*: **SPAG9 promotes prostate cancer proliferation and metastasis via MAPK signaling pathway.** *American journal of translational research* 2019, **11**(8):5249-5260.
34. Alexandre-Ramos DS, Silva-Carvalho AÉ, Lacerda M *et al*: **LL-37 treatment on human peripheral blood mononuclear cells modulates immune response and promotes regulatory T-cells generation.** *Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie* 2018, **108**:1584-1590.
35. Martin MJEJ: **Cutadapt removes adapter sequences from high-throughput sequencing reads.** 2011, **17**(1).
36. Li HJapa: **Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.** 2013.

## Tables

Table 1. Clinical characteristics of NC group and RA group.

Clinical characteristic		NC	RA
Number of cases		7	7
Age *		34±9	49±12
Gender	Male	1	1
	Female	6	6
Inflammatory marks	CRP (mg L <sup>-1</sup> ) *	NA	9±8
	ESR(mm h <sup>-1</sup> )*	NA	54±32
	Complement C3(g L <sup>-1</sup> ) *	NA	1.0±0.4
	RF(IU ml <sup>-1</sup> ) *	NA	344±314

NC, healthy controls; RA, rheumatoid arthritis; PBMC, Peripheral blood mononuclear cells; NA, not applicable; ESR, Erythrocyte sedimentation rate; CRP, C-reactive protein; anti-CCP, anti-cyclic peptide containing citrulline; RF, rheumatoid factor; \*, mean ± standard deviation.

Table 2. Identified markers genes and type-specific transcription factor motifs in the scATAC-seq experiments

Cell types	Markers genes	Transcription factor motifs
B cells	CD79A, CD19, MS4A1	POU2F3, IRF1, POU5F1B, STAT1:: STAT2, POU2F, POU1F1, POU2F2, POU3F3, POU5F, POU3F1
T cells	CD3D, CD3G, LEF1, CD8A ,IL2RA	None
Monocytes	CD14, CD68, ITGAM	FOSL2::JUN, FOSL1::JUN, FOS, FOSL1, FOSL1::JUND, FOSL1::JUNB, JUN(var.2), JUNB, FOS::JUND, JUND
DCs	IL3RA	GATA3, ELF2, ELK1, CEBPB, ELK3, GATA5, FEV, CEBPE, CEBPA, ELF5
NK cells	GZMB, NKG7, KLRD1	EOMES, TBX2, TBR1, TBX20, TBX21

Table 3. Comparison of five major cells in both PBMC\_RA and PBMC\_NC libraries.

Cell types	Cell number ratio		<i>P</i> -value	FDR	Number of differential peaks ( <i>P</i> < 0.05, FC > 1.2)
	RA	NC			
T cells	43.02	56.98	0.00	0.01	0
NK cells	59.55	40.45	0.69	1.00	51
Monocytes	54.78	45.22	0.50	1.00	11
B cells	71.77	28.23	0.00	0.01	149
DCs	59.77	40.23	0.65	1.00	665

NC, healthy controls; RA, rheumatoid arthritis; PBMC, Peripheral blood mononuclear cells; FDR, false discovery rate

Table 4. Type-specific transcription factor motifs in subclusters for scATAC-seq dataset

Subclusters	Enriched motifs in PBMCs (p<0.05, FC>1.2)		Motifs ID (p<0.05,FC>1.5)
	Top 3 enriched motifs in each subclusters	Transcription factors	
B-0	MA0817.1,MA0506.1, MA0713.1	BHLHE23, NRF1, PHOX2A	MA0817.1,MA0506.1,MA0713.1, MA1549.1,MA0912.2,MA0675.1, MA0890.1,MA0793.1,MA0707.1, MA0027.2,MA0132.2,MA0902.2, MA0103.3,MA0618.1,MA1519.1, MA1132.1,MA1495.1,MA0700.2
B-1	MA0506.1	NRF1	/
B-2	MA1572.1,MA1571.1, MA0014.3	TGIF2LY, TGIF2LX, PAX5	/
B-3	MA0877.2,MA1104.2, MA0036.3	BARHL1, GATA6, GATA2	MA0877.2,MA1104.2,MA0036.3, MA0482.2,MA0700.2,MA1650.1, MA0131.2,MA0025.2,MA0644.1, MA0119.1,MA0632.2,MA0833.2, MA1518.1,MA1569.1,MA1476.1, MA1563.1,MA1107.2,MA1118.1, MA0666.1
M-0	MA0139.1,MA1125.1, MA0052.4	CTCF,ZNF384,MEF2A	MA0139.1
M-1	MA1476.1	DLX5	/
M-2	MA1112.2,MA0509.2, MA0738.1	NR4A1,RFX1, HIC2	/

# Figures

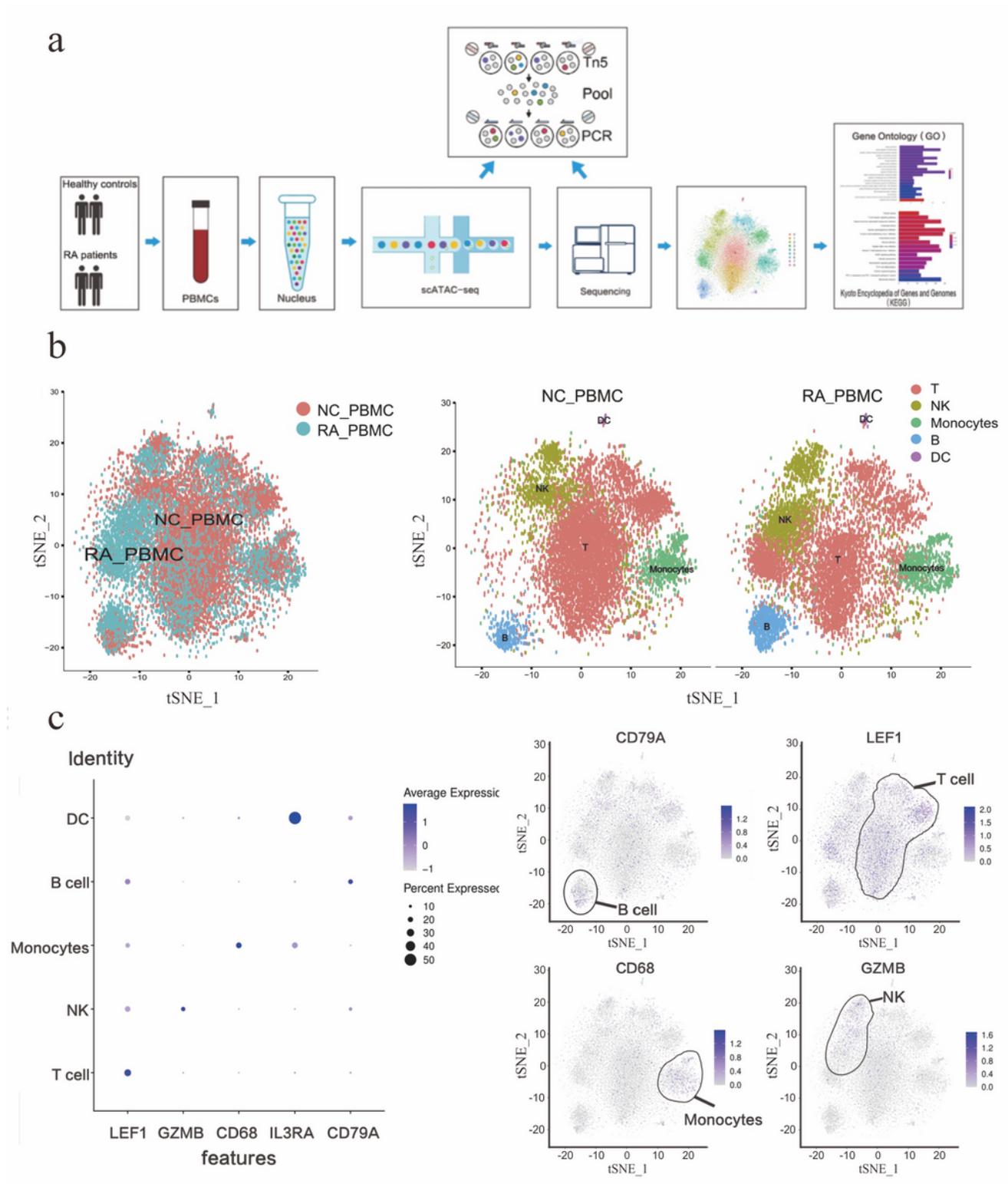


Figure 1

Chromatin structure reflects cellular specialization via the scATAC-seq (a) workflow of the assay for transposase accessible chromatin in single-cell sequencing in human PBMCs; (b) clustering of single nuclei accessibility profiles of the five cell groups in healthy controls and rheumatoid arthritis patients; (c) Open chromatin signals around marker genes of four clusters, including natural killer cells (NK), T lymphocytes (T cell), B lymphocytes (B-cell) and monocytes.

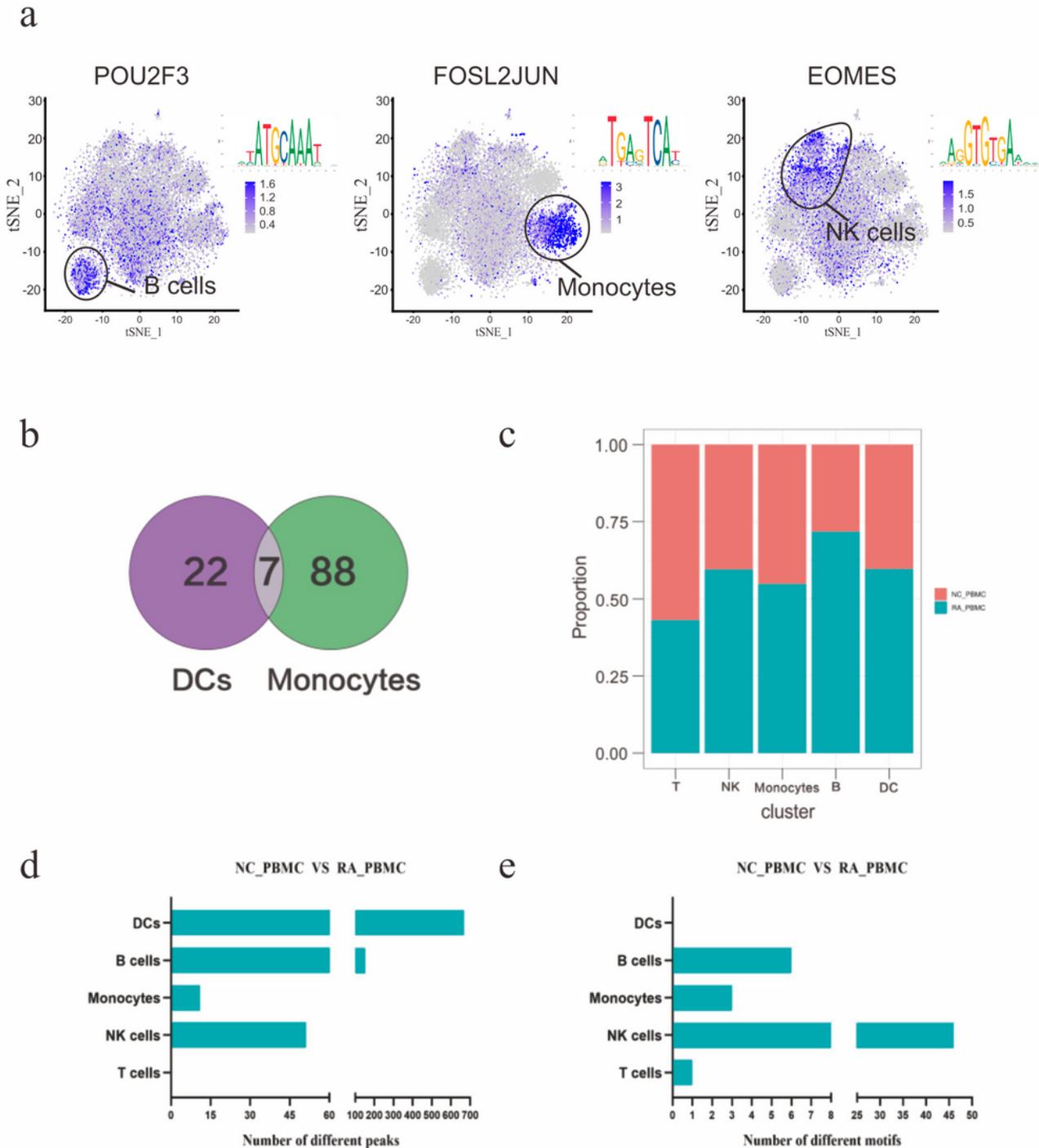
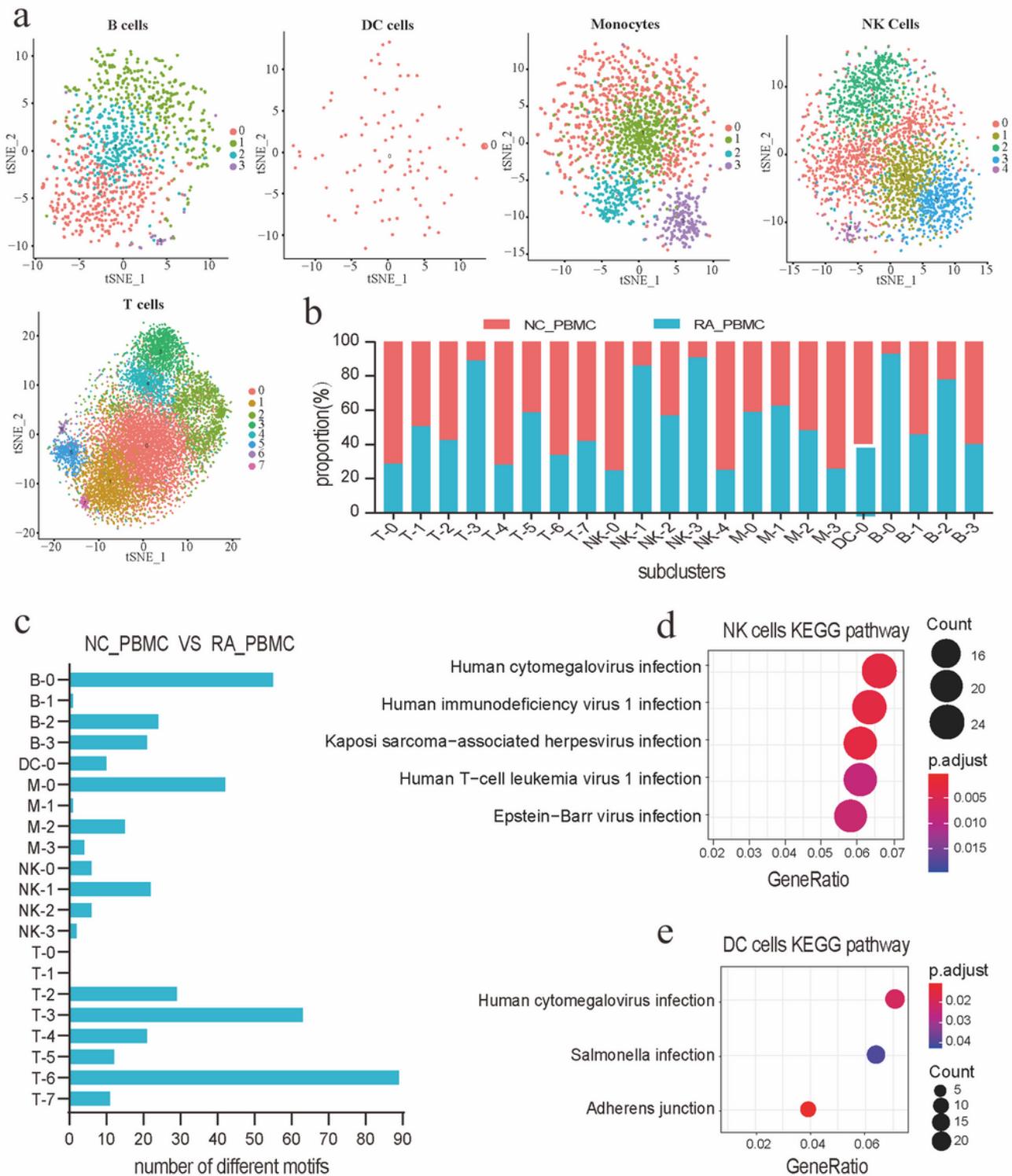


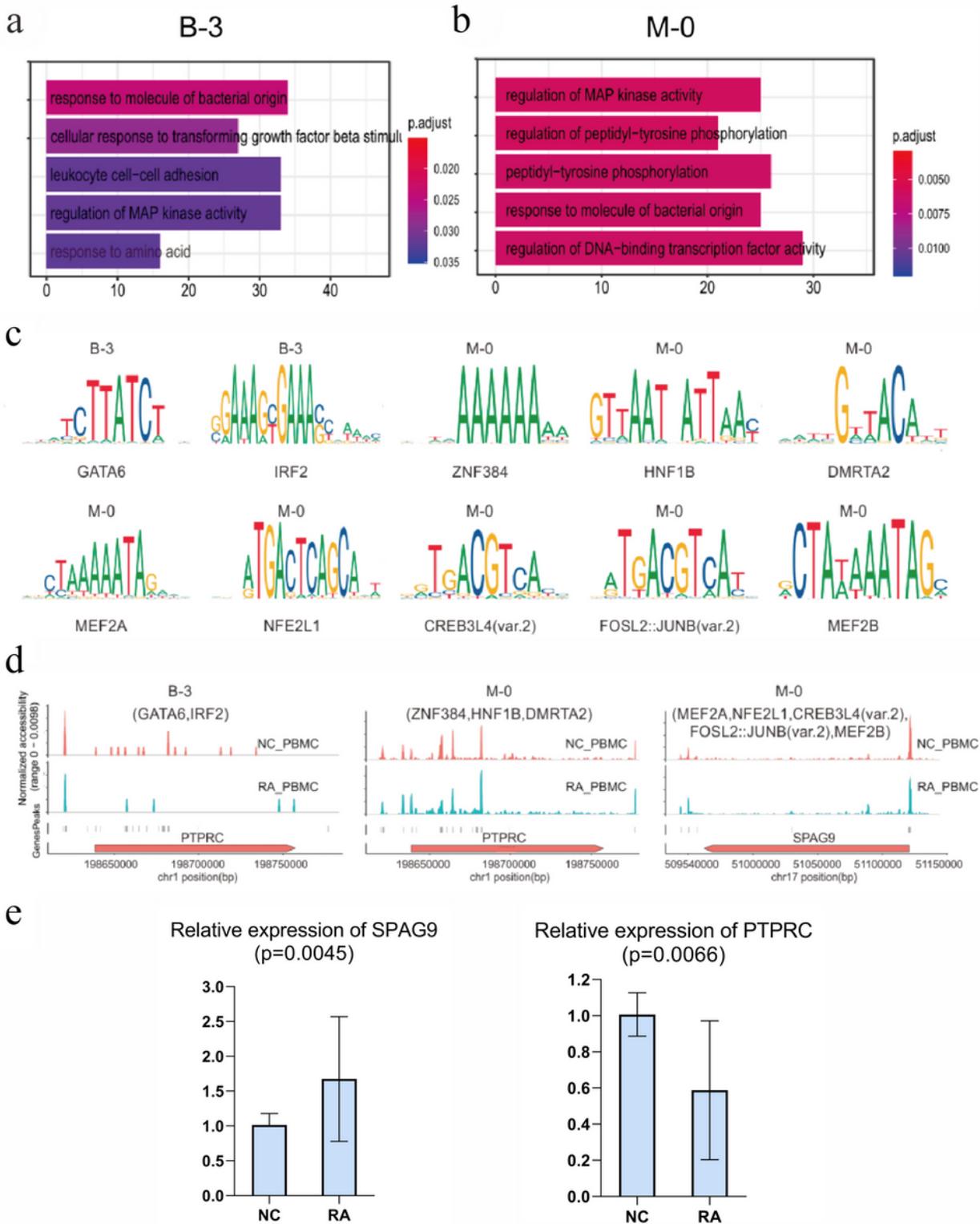
Figure 2

Differential chromosome accessibility analysis with human PBMCs. (a) t-SNE visualization of accessibility deviations at cell-type-specific TF motifs across three clusters. (POU2F3 for B cells; FOSL2JUN for monocytes; EOMES for Natural killer cells); (b) intersection of TF motifs between DCs and monocytes shown in a Venn diagram; (c) the percentage of cell subclusters for comparison of the cell number ratio in the NC\_PBMC and RA\_PBMC groups; (d) the number of different peaks in the NC\_PBMC and RA\_PBMC groups; (e) number of different motifs in the subcluster of cells in the NC\_PBMC and RA\_PBMC groups.



**Figure 3**

Epigenomic analysis of subclusters in human PBMCs. (a) Subclusters of five major cell types of PBMCs in t-SNE; (b) the percentage of cells subclusters for comparison of cell number ratio in the NC\_PBMC and RA\_PBMC groups; (c) Number of different motifs in the subcluster of cells; (d) significant biological functions and signal pathways within natural killer (NK) cells, according to KEGG analysis; (e) significant biological functions and signaling pathways within dendritic cells (DCs), according to KEGG analysis.



## Figure 4

Identifying significant TF motifs and their corresponding pathways in the regulation of MAP kinase activity. (a) GO analysis of the differential genes between RA\_PBMC and NC\_PBMC libraries shown in B-3; (b) GO analysis of the differential genes between RA\_PBMC and NC\_PBMC libraries in M-0; (c) Position weight matrices (PWMs) for the key TFs in regulating target genes; (d) binding sites near PTPRC and SPAG9 with differential accessibility in the RA\_PBMC and NC\_PBMC libraries across B-3 and M-0. (e) Relative expression of SPAG9 and PTPRC using qPCR.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplementalfigure1.png](#)
- [supplementalfigure2.png](#)
- [supplementalfigure3.png](#)
- [supplementalfigure4.png](#)