

Contents

Abstract	iii
Chapter 1. Introduction	
1.1. The dangers of tuberculosis	1
1.2. Macrophages and dendritic cells in tuberculosis	2
1.3. BCG vaccine against tuberculosis mechanism	4
1.4. Limitations of BCG vaccine	5
1.5. Project preparation and goals	7
Chapter 2. Methodology	
2.1. CellRanger	8
2.2. Quality control and cell filtration	10
2.3. Data normalizing, Data scaling and Doublets removing	10
2.4. Dimensionality reduction	11
2.5. Clustering and annotating cells	13
2.6. Differential Expression Analysis	14
2.7. Enrichment analysis	14
2.8. Code and Data Availability	15
Chapter 3. Results	
3.1 Quality control and selection filtering of eight cell datasets based on single-cell RNA-seq	15
3.2 Dimensionality reduction	19
3.3 Cluster analysis and annotate cell types	24
3.4 Differential expression analysis	30
3.4.1 Identification of differentially expressed genes in dendritic cells group 1 (cluster 1, 3, 4, 6 and 13)	30
3.4.2 Identification of differentially expressed genes in dendritic cells group 2 (cluster 8 and 16)	34
3.4.3 Identification of differentially expressed genes in dendritic cells group 3 (cluster 2)	35
3.4.4 Identification of differentially expressed genes in dendritic cells group 4 (cluster 0, 10 and 11)	39
3.4.5 Identification of differentially expressed genes in dendritic cells group 5 (cluster 18)	42
3.4.6 Identification of differentially expressed genes in macrophages (cluster 17)	43
3.5 Defining cell types of dendritic cell groups	46
3.6 Functional enrichment analysis	48
3.6.1 Enrichment analysis in dendritic cells group 1 (cluster 1, 3, 4, 6 and 13)	48
3.6.2 Enrichment analysis in dendritic cells group 3 (cluster 2)	52
3.6.3 Enrichment analysis in dendritic cells group 4 (cluster 0, 10 and 11).57	57
3.6.4 Functional enrichment analysis in macrophages (cluster 17)	62
Chapter 4. Discussion	

4.1 Analysis of dendritic cell results	66
4.2 Analysis of macrophage results	69
4.3 Limitations of the experiment	70
5. Conclusions	71
6. Workflow	72
7. Acknowledgement	72
8. Bibliography and references	73
9. Appendix	80

Abstract

Tuberculosis (TB) is a recurrent respiratory infectious disease resulting from the human body becoming infected with *Mycobacterium tuberculosis*. It continues to be a significant social and public health issue, seriously threatening people's health. Bovine tuberculosis (BTB), caused by *Mycobacterium bovis*, is also a dangerous disease, causing approximately US\$3 billion economic losses yearly (1). Both *Mycobacterium bovis* and *Mycobacterium tuberculosis*, the human form, can cause tuberculosis in humans. The worldwide pandemic of tuberculosis has stimulated research into related treatments and vaccines. For the time being, Bacille Calmette-Guérin (BCG) is a vaccine that is widely used worldwide to prevent tuberculosis, while the immunisation effect of BCG is not satisfactory. One drawback of BCG is that the vaccine protects against tuberculosis in children but not adults (2), suggesting that BCG does not induce long-term protection. Currently, researchers hope to improve the immune effect and memory of BCG or develop new vaccines to address this issue. This makes the study of the principle of BCG immunity against *Mycobacterium tuberculosis* or *Mycobacterium bovis* an important research direction, which is the aim of this experiment. The immune response of BCG against tuberculosis is mainly initiated and controlled via the function of dendritic cells and macrophages. *Mycobacterium tuberculosis* or *Mycobacterium bovis* is the first to infect alveolar macrophages and BCG exerts its immune effect in the presence of dendritic cells. In this experiment, a bovine model's immunological response induced by macrophages and dendritic cells will be the main subject because cattle are also partially protected by BCG and are good models for human disease and vaccine studies. The experimental results can also be applied to humans. The samples we studied were including both dendritic cells subjected to fluorescence-activated cell sorting and immune cells from whole lymph nodes. By using bioinformatics packages such as CellRanger (version 2.2.0)

and Seurat (version 2.3.4) to analyze the single cell transcriptomic data, we compared differentially expressed genes in dendritic cells and macrophages after BCG injection. We also looked for the relevant cellular pathways affected using enrichment analysis. The experimental results can provide guidance and insights for subsequent research on tuberculosis control methods and the development of novel vaccines.

Chapter 1 Introduction

1.1 The dangers of tuberculosis

Bovine tuberculosis (BTB) is a serious zoonotic disease that poses a significant risk to the cattle industry, producing yearly economic losses of approximately US\$ 3 billion (1). Bovine tuberculosis is similar to tuberculosis in humans and has a similar immune response. The bacteria that causes tuberculosis in cattle can be transmitted from cattle to humans. For example, if a cow has tuberculosis, its milk will also contain tuberculosis bacteria, and a person can contract tuberculosis by drinking raw milk. Cough, fever, and weight loss are signs and symptoms of tuberculosis, which primarily affects the lungs in humans (3). Although the majority of humans with tuberculosis are affected by *Mycobacterium tuberculosis*, around 10-20% of cases are caused by *Mycobacterium bovis* and understanding the immune response could lead to improved vaccines and control measures for both human and bovine tuberculosis. In spite of the fact that *Mycobacterium tuberculosis* and *Mycobacterium bovis* can infect most organs, pulmonary tuberculosis is the most frequent (4). These bacilli are slow-growing, acid-fast, and transmit predominantly through the respiratory system (4).

Despite the fact that *Mycobacterium tuberculosis* or *Mycobacterium bovis* is thought to be present in one-third of the global populace, active disease is not frequently caused by infection (5). Immune responses to infections can often be successfully controlled, but pathogens cannot. Every year, tuberculosis in humans leads to ten million new cases and one million plus fatalities globally (6). However, exposure to *Mycobacterium tuberculosis* or *Mycobacterium bovis* does not always cause active disease. The death rate for tuberculosis patients admitted to hospitals is still high (>5–10%), even in the middle- and high-income nations where anti-tuberculosis medication is more easily accessible (7). Although the immunogenic mechanisms at play have not been

fully understood, it is believed that excessive inflammation has a role in the severity of the disease and its consequences in active tuberculosis (8). Once *Mycobacterium tuberculosis* or *Mycobacterium bovis* infects the lungs, an initial host immune response is initiated that may kill and eliminates the bacilli. These individuals do not develop tuberculosis as a result of this exposure event. Otherwise, the bacteria will grow and multiply immediately after infection, causing primary tuberculosis. A state of "latent infection" can also exist, in which *Mycobacterium tuberculosis* or *Mycobacterium bovis* persists in a subclinical form. The immune system can stop bacteria from reproducing uncontrollably, allowing them to be dormant or remain at trace levels without causing disease. These dormant bacteria may reactivate under certain conditions or escape from quiescence and cause disease. At any time in the life of an infected person, *Mycobacterium tuberculosis* or *Mycobacterium bovis* may remain dormant or persistent to evade host defences or may proliferate within the cells of the infected host and remain in unstable equilibrium with the host (9).

1.2 Macrophages and dendritic cells in tuberculosis

Mycobacterium tuberculosis persists in macrophages within granulomas of infected host organs. Granulomas are composed of macrophages, giant cells, T cells, B cells, and fibroblasts (10). Granulomas are a powerful barrier against the invasion of *Mycobacterium tuberculosis*, but *Mycobacterium tuberculosis* has a way of evading phagocytosis by macrophages. Antimicrobial defences mediated by phagocytic oxidase and inducible nitric oxide synthase are overcome by *Mycobacterium tuberculosis* by circumvention methods (11). Additionally, *Mycobacterium tuberculosis* inhibits the fusion of infected phagosomes with lysosomes, which would kill *Mycobacterium tuberculosis* and create *Mycobacterium tuberculosis* peptides and lipids for antigen presentation pathways (12). *Mycobacterium tuberculosis* down-regulates

antigen presentation by macrophages and dendritic cells (12). Similar mechanisms take place after infection with *Mycobacterium bovis* in the human and bovine hosts. Understanding the early events in macrophages and dendritic cells may pinpoint mechanisms that can be manipulated for improved immunity.

One of the key cells linking innate and adaptive immune reactions is the dendritic cell, which is crucial in collecting, processing and presenting antigens (13). Dendritic cells are also involved in immunophenotypes that are generated by BCG. CD4 and CD8 T cells receive antigens from antigen-presenting cells, which in turn trigger specialised cellular defence against intracellular infection (14). Immediately after coming into contact with and recognising *Mycobacterium tuberculosis* or *Mycobacterium bovis*, dendritic cells go through a process called maturation that modifies their phenotypic characteristics and triggers the production of cytokines that regulate the immune response in various microenvironments and different leukocytes, in turn (15). Mature dendritic cells go from draining lymph nodes to peripheral organs, where they encourage the growth of memory T cells and the differentiation of effector T cells that participate in the infection-specific adaptive immune response (16). Antigen-carrying dendritic cells stimulate T cells. However, the consequences of the interaction of *Mycobacterium tuberculosis* or *Mycobacterium bovis* with dendritic cells are not fully understood. Existing reports are conflicting. It has been demonstrated in numerous investigations that dendritic cells improve the cellular immunological reaction to *Mycobacterium tuberculosis* or *Mycobacterium bovis* infection (17). Dendritic cells strongly resemble the infection site of *Mycobacterium tuberculosis* during the beginning of the inflammatory response against it (17). The lung mucosa contains immature dendritic cells that have been trained to take up and digest antigens (17,18). They develop and move in lymphoid organs after contact with pathogens, where they excite T cells by expressing

MHC and costimulatory molecules on their cell surfaces and secreting immunoregulatory cytokines (17,18). However, the synthesis of surface molecules and anti-inflammatory cytokines has not been found to be significantly increased in dendritic cells infected with *Mycobacterium tuberculosis* or *Mycobacterium bovis*, according to research (18). There have been claims that *Mycobacterium tuberculosis* or *Mycobacterium bovis* hinders dendritic cell maturation, lessens IL-12 release from dendritic cells, and prevents the stimulation of T cell proliferation (19). Dendritic cells are frequently infected by *Mycobacterium tuberculosis* or *Mycobacterium bovis*, which inhibits their ability to operate in vivo. Dendritic cells carry *Mycobacterium tuberculosis* or *Mycobacterium bovis* from the lungs to the nearby draining lymph nodes when the infection is still in its early stages (20). However, despite expressing costimulatory molecules and MHC class II on their surfaces, these cells are ineffective at promoting *Mycobacterium tuberculosis* or *Mycobacterium bovis* immunity (20). Dendritic cells infected with *Mycobacterium tuberculosis* or *Mycobacterium bovis* are less effective at presenting mycobacterial antigens and activating CD4+ T cells specific to *Mycobacterium tuberculosis* or *Mycobacterium bovis* in vivo (21). In addition, *Mycobacterium tuberculosis* or *Mycobacterium bovis* infection induces the production of several cytokines, such as IL-10, that reduce dendritic cell trafficking to draining lymph nodes (22). From the current research, it is not certain whether *Mycobacterium tuberculosis* or *Mycobacterium bovis* stimulates or inhibits the cellular immunity elicited by dendritic cells.

1.3 BCG vaccine against tuberculosis mechanism

The only vaccine that has been widely used to prevent tuberculosis to yet is Bacille Calmette-Guérin (BCG) (23). Its antigenic composition is very close to *Mycobacterium tuberculosis* and *Mycobacterium bovis* (24). Although the exact processes by which BCG triggers immunity are not entirely understood,

it is known that CD4+ and CD8+ T cells that are specific to the antigen are activated (25). There is growing proof that BCG has nonspecific beneficial effects, especially in the neonatal (infant and calf) period. In addition to preventing tuberculosis, it can also help protect newborns from other infections. Study confirms BCG is effective in preventing tuberculous meningitis and miliary tuberculosis, which are highly lethal to infants and young children (26). This proposed that the immunological protection brought on by BCG may involve innate immune cells (26). The immune protection produced by BCG belongs to innate immunity after training. The main mechanism is that the acetylation and methylation of innate immune cell genes are changed after BCG inoculation, and this innate immune response is amplified by antigen stimulation. The freshly absorbed BCG is destroyed via the phagosome-lysosome pathway, and its byproducts can stimulate CD4+ T cells via the MHC-II molecular pathway (27). BCG can improve the ability of macrophages to attack tumour cells, their activity, the activation of T lymphocytes, and the cellular immunity of the body (28).

1.4 Limitations of BCG vaccine

However, although BCG can relieve symptoms caused by *Mycobacterium tuberculosis* or *Mycobacterium bovis* and it can protect humans from infection as a child. A significant problem with BCG is that BCG vaccination does not induce long-term protection because it does not adequately induce long-term T cell memory. This is evident in humans, where the excellent protection against childhood tuberculosis cannot be replicated in adults. Similarly, in cattle, the duration of BCG-induced immunity against Bovine tuberculosis is approximately two years (29). The ability of vaccination to induce long-term memory is an important aspect in evaluating and measuring the merits of vaccines (30). And vaccines with long-term T-cell memory are more effective and appropriate for societal needs (30). Despite the fact that the BCG vaccine

has been administered to more people than any other vaccine, a large number of people around the world contract *Mycobacterium tuberculosis* or *Mycobacterium bovis* each year. Therefore, the protective effect of BCG vaccine remains questionable. Over the last few decades, many new vaccine approaches have been elucidated, such as live attenuated vaccines, subunit vaccinations, and naked DNA vaccines (31,32). Unfortunately, none of the vaccines has proved effective in the long term against tuberculosis because they do not produce durable memory cells (31,32). This means that there is still a need for research into new and unique tuberculosis vaccine technologies.

It has been shown that the T-cell memory-enhancing cytokines IL-1+IL-6+TNF- α and IL-7+IL-15 can be used to promote enhanced long-term immunity to vaccines (33). These cytokines promote the retention of antigen-presenting cells such as dendritic cells at the vaccination site, as well as the survival and growth of memory T cells (33). By using IM-1.6.alpha, this strategy elicited a robust immune response, producing durable memory CD4 and CD8 T cells, as well as a significant Th1 immune response, reducing mycobacterial load and attenuating lung lesions, providing better protection than the conventional BCG vaccine (34). This finding has further stimulated researchers to investigate new vaccines and therapies to control tuberculosis. Currently, three main methods of developing vaccines have been developed: a completely new tuberculosis vaccine, a new recombinant vaccine derived from an existing BCG vaccine, and a booster vaccine that enhances the current BCG vaccination (35). We used BCG in cattle and found similar limitations. In a recent study, protection was induced in calves that had received the BCG vaccine at the age of one month and again 12 months later. However, no discernible protection was seen in another cohort that had received the vaccine at the 24-month mark. This suggests that BCG immunity may not last for a long time (36). It is also not used routinely because it interferes with the bovine tuberculosis diagnostic tests. However, BCG remains the gold standard vaccine. Understanding how it induces immunity and affects macrophages and

dendritic cells can be used to determine areas for further research and vaccine development.

1.5 Project preparation and goals

To accelerate the future creation of fresh vaccinations and treatments, Hope's team used a cow model to explore the pathogenesis of tuberculosis and the BCG vaccine's immune response. The team used a cattle model with an afferent lymphatic cannula (37). Using this model, it is more convenient for us to study the immune response of dendritic cells. Many veterinary and human vaccines, such as BCG, are given intradermally, from bovine skin into the lymph to carry the vaccine from the injection site to the site of induction of an immune response (37). This has given us a unique understanding of the gene expression changes that occur *in vivo* during vaccination, which plays an important role in vaccination research (37). Samples were taken immediately and 24h after vaccination of BCG into the skin. Six samples were from samples enriched for dendritic cells and two were whole, unfractionated lymph samples taken pre and post BCG injection. A total of 8 single cell transcriptome sequencing samples were prepared. Single cell RNA sequencing (scRNA-seq) allows for the concurrent analysis of over 10,000 single cell transcriptomes, unlike bulk RNA-seq, which can only supply the average expression signal of millions of cells. This allows for the description of new cell subpopulations (38). Therefore, even closely related cell groups can be accurately distinguished from one another using scRNA-seq and it can show the alterations that give each type of cell its own identity. Four of the eight single-cell transcriptome sequencing samples were pre-BCG vaccination samples and four were post-BCG vaccination samples. Three of these sample sets from animals 602791, 303215 and 402740 were transcriptome data from dendritic cells after fluorescence-activated cell sorting, for a total of six samples. Sample set LNode was transcriptomic data from two whole lymph

node cells from cattle before and after BCG injection, for a total of two samples. The first part of the protocol for this experiment used the CellRanger package. First, we built the bovine reference genome, an index to find the gene of interest, imported it into CellRanger for each sample and compared it with the bovine reference genome. Then we performed quantitative analysis on the sample data to obtain quality control results and related files for downstream analysis. The second part of this experiment was using the Seurat package. We performed quality control, dimensionality reduction, cluster analysis, differential expression analysis and enrichment analysis on the datasets. We compared the differential expression of dendritic cells and macrophages in samples before and after BCG injection and searched for differentially expressed genes-related signalling pathways through enrichment analysis. This allows the study of the role and impact of specific dendritic cells and macrophages in the immune process that follows vaccination with BCG. This will provide data and analysis pipelines for follow-up studies of vaccines and immune responses to pathogens and help develop new vaccines and treatments to control tuberculosis.

Chapter 2 Methodology

2.1 CellRanger

CellRanger was a collection of analysis pipelines for single-cell data processing that could be used to align reads, create feature barcode matrices, cluster data, do other secondary studies and more (39). CellRanger could be used by researchers to create a unique reference genome for a species or to add unique marker genes to a reference gene. Finding the reference genome FASTA and GTF data for cattle is the initial step in this investigation (40). This species (*Bos taurus*) could be obtained from the Ensembl database. The data source websites were http://ftp.ensembl.org/pub/release-106/gtf/bos_taurus and http://ftp.ensembl.org/pub/release-106/fasta/bos_taurus. Next, the

reference genome data was uploaded to Eddie (The ECDF [Edinburgh Compute and Data Facility] Linux Compute Cluster), and then the GTF files were filtered using CellRanger '*mkgtf*' function. We then used the CellRanger package to filter the transcripts and genes to find genes of interest, and as Ensembl's GTF files contain selectable tags, filtering is easy. FASTA files and GTF files were then copied using CellRanger '*mkref*' function to index these files in a variety of formats. In this experiment, there were eight groups of samples, four of which had been injected with BCG and the other four had not. These were 602791_Pre_BCG_DC, 303215_Pre_BCG_DC, 402740_Pre_BCG_DC, Pre_BCG_Lymph_node and 602791_Post_BCG_DC, 303215_Post_BCG_DC, 402740_Post_BCG_DC, Post_BCG_Lymph_node. The two samples, Post_BCG_Lymph_node and Pre_BCG_Lymph_node, were taken from full lymph nodes, which contained all cell types in lymph nodes. The remaining six samples were sorted by fluorescence-activated cell sorting to isolate the dendritic cells, but other cell types were still present at low numbers. Finally, CellRanger '*count*' function was used to perform quantitative analysis, including alignment, filtering, barcode count and UMI count. The most visible result was web_summary.html, which showed metrics including the number of cells sequenced and the number of genes per cell. It contained quality control metrics, including mapping metrics, screening thresholds, estimated cell counts after the screening, the information on the number of reads and genes per cell after screening. In the 'filtered_gene_bc_matrix' directory, three important files were stored: barcodes.tsv, features.tsv and matrix.mtx. Barcodes.tsv was a file containing all the cell barcodes for the sample. Features.tsv was a file containing the identifiers of the quantified genes. The origin of the identifiers could vary depending on the reference used in the quantification method (i.e. Ensembl, NCBI, UCSC), but in most cases, these were the official gene markers. Matrix.mtx was a file containing a matrix of count values. These files were the input files for downstream Seurat analysis.

2.2 Quality control and cell filtering

Cells were filtered, and quality was checked using Seurat packages (41). Whether they received BCG injections or not, the eight datasets were marked. They were then integrated into a single Seurat object. The '*PercentageFeatureSet*' function was used to update the metadata to reflect the proportion of ribosomal and mitochondrial genes in each cell. Seurat automatically calculated some quality control statistics, such as the number of genes found in each cell and the total amount of molecules found inside a cell, which were stored in the metadata nCount_RNA and nFeature_RNA columns. And both violin plots and scatter plots were used to visualise these quality control features. The standard filtering method was to filter genes and remove those expressed in the low number of cells. The rationale for removing low gene count was because they were likely to be of low quality, such as empty droplets contaminated with highly expressed genes from other cells or dead/dying cells. High nCount_RNA or nFeature_RNA suggest the cell might be a doublet. Moreover, we need to remove cells with low nFeature_count. This was because if too few genes were detected, there might be low-quality or empty cells, while if too many genes were detected, there might be more than two cells in one droplet. Regarding the filtering of mitochondria and ribosomes, we removed genes with low percentage readings and high mitochondrial readings, as mitochondrial gene expression was elevated in low-quality or dying cells (42). In general, cells with less than 200 genes detected needed to be filtered out and each selected gene needed to be expressed in at least three cells.

2.3 Data normalizing, Data scaling and Doublets removing

The data were then normalised using a global scaling normalisation technique called "LogNormalize," which transformed the count data for each cell into a logarithmic scale and could make a matrix from the normalised and

logarithmically transformed data after eliminating any extraneous cells from the dataset. The confusing impact of varying read depths across cells was eliminated by scaling the data by library size, followed by a log2 transformation (43). After that, during pre-processing, cell cycle phase scores based on standard markers were calculated and regressed from the data, giving each cell a score based on how well G2/M and S phase markers were expressed and removing sources of heterogeneity from the data to lessen the impact of cell cycle heterogeneity in scRNA-seq data (44). The data were then scaled, a standard pre-processing step before performing dimensionality reduction. The features in the dataset were scaled and centred for each feature before regression, and the residuals were also scaled and centred. Since each gene is expressed at a distinct level, PCA will be able to capture the larger variation that naturally exists in genes with higher expression values (45). This meant that each gene needed to be given a similar weighting when PCA was performed. It was common practice to place each gene in the centre and scale it prior to performing PCA. The Z-score normalisation method was used for this exact scaling. It showed how much a value differs from the average across multiple samples, or how much the read count for one cell of a specific gene differs from the average across all cells (46). Z scores were usually performed based on log-normalized counts and were useful for PCA, clustering and drawing heat maps (46). Additionally, as the presence of one or more cells in the same droplet was a common problem during trials, DoubletFinder was used to foresee uncertainties and remove any common worries from the dataset (47).

2.4 Dimensionality reduction

The following analysis steps necessitated grouping by determining variations in gene expression between cells and examining the diversity of distinct cell types. This involves dimensionality reduction in single cell RNA data processing and data visualisation. The first step in dimensionality reduction

was feature selection. Here, a feature would be used to refer to a gene. The goal of feature selection was to eliminate highly variable genes that can be utilised to represent the main variability in the data because the majority of cellular expression was identical between cells and only a small number of genes changed in expression (48). Typically 1k to 5k genes were selected, with the exact number varying depending on the complexity of the data (48). In this experiment, 2k genes were selected as highly variable genes. By counting the subset of features in the dataset that exhibit high intercellular variability, the biological signal in the single-cell dataset could be discovered to be highlighted by concentrating on these genes in downstream studies. These different genes with high intercellular variation could identify which genes were important in the dataset to distinguish between cell types and also separate cell populations well (49). Due to the fact that each gene was expressed at a variable level, genes with higher expression values would inevitably have more variations, which PCA would be able to detect. Direct clustering with gene down-keeping frequently introduced significant noise and wasteful computation because single-cell RNA-seq dataset was noisy and data-heavy (50). So the second step was to use linear dimensionality reduction with PCA to reduce the noise by merging all the information and then processing it.

Seurat grouped the cells based on their PCA scores to combat the significant technical noise associated with any one aspect of the scRNA-seq data. There are three criteria for determining the PC threshold: the cumulative contribution of the principal components is greater than 90%, the contribution of the PCs to the variance is less than 5% and the difference between two consecutive PCs is less than 0.1%. Elbow plots were used to determine the contribution of each PC axis to the variance. According to the elbow plot, PCs would be 30. To visualise the cell clustering afterwards, there are a number of different techniques that can help. The most popular methods include t-SNE and UMAP techniques. Both strategies were intended to put cells with comparable local neighbourhoods in a low-dimensional space (51). More importantly, unlike

t-SNE, UMAP is not limited by the number of dimensions of the data and PCA, UMAP and t-SNE can be plotted side-by-side for comparison (52). The comparison was used to confirm whether the data set needs to be corrected for batch effects prior to clustering and differential gene expression analysis. Suppose there were batch effects in the expression data between samples, demonstrating that some subgroups that can be clustered together were shifted between samples. In that case, this requires the integration of multiple single-cell RNA-seq datasets to correct for batch effects across datasets (53). Seurat offered a variety of self-contained methods for integrating single-cell datasets: first looking for anchor points between different datasets that reflect the correlation between cells in different datasets and then using this correlation to correct the expression matrix to integrate multiple datasets or to act as a predictor.

2.5 Clustering and annotating cells

Seurat employed a graph-based clustering technique that embeds cells into a graph using a K-nearest neighbour (KNN) graph and connects cells with similar gene expression patterns by drawing edges (54). It then aimed to cluster the cells in this graph into groups that were closely related to one another. This experiment used the '*FindClusters*' function to perform graph-based clustering. The two closest clusters were found and these clusters were merged to generate a new distance matrix where the distance between clusters was now the shortest structure-structure distance of the cluster, and this was repeated until the shortest distance was greater than the tolerance (55). Then, to visualise these cell clusters, this experiment used the UMAP method for visual clustering analysis. Next, based on the marker genes that have been defined, researchers can identify each cell type observed through literature mining (56). However, this experiment attempted to use the automatic annotation of SingleR (56), which can help us obtain relevant clustering information for dendritic cells and macrophages.

2.6 Differential Expression Analysis

Differential expression analysis could be achieved via '*FindMarkers*' function, which did a non-parametric Wilcoxon rank sum test for the investigation of differential expression (57). This experiment was to perform differential analysis on two specific groups of cell classes, dendritic cells and macrophages before BCG injection and those two classes of cells after BCG injection, and two specific cell classes can be specified by setting the ident.1 and ident.2 parameters. It was finding different expressed features by finding markers that define clusters via differential expression. It was assigning cell type identity to clusters use '*FindMarkers*' function to find genes that were differentially expressed before and after BCG injection. '*FindMarkers*' function would find the marker between two different identity groups. Another useful way to visualise changes in gene expression was to use the '*FeaturePlot*' and '*VlnPlot*' functions. This would display a feature plot of a given list of genes, split by grouped variables. Moreover, for each of the eight samples, the data were integrated separately for each control and control group. Differentially expressed genes were identified using '*Findmarkers*' function and visualized using '*FeaturePlot*' and '*VlnPlot*' functions to explore the differential genes.

2.7 Enrichment analysis

Finally, the differential genes were subjected to GO enrichment analysis. Using the chosen differential genes, it was possible to determine the hypergeometric distribution link between these distinct genes and some specific GO categorization branches, according to GO enrichment analysis (58). GO analysis will return a hypothetical value p-value, and the analysis results include GO functional classification results and GO functional enrichment results(58). GO enrichment analysis was classified into three categories at different levels, describing the cellular components, molecular functions and biological processes of genes (59). GO enrichment analysis

screened the differential genes and then determined which annotation pathways the differential genes were enriched in. We could focus on the GO terms specifically for the specificity of the biological question. The important thing was to combine the biological question and then annotate it with the gene's function to determine whether its genetic changes were biologically important.

2.8 Code and Data Availability

The Supplementary data contained the code for this project. Single cell transcriptomic data was stored in the Github (<https://github.com/B196466-2021/single-cell-transcriptomic-data>).

Chapter 3 Results

3.1 Quality control and selection filtering of eight cell datasets based on scRNA-seq

After CellRanger analysis and loading of the eight datasets, a single test revealed 21880 features across 30573 cells for this experiment. The percentage data for mitochondrial and ribosomal genes was then calculated for each cell and added to the metadata, which would aid in visualising them within the various metadata parameters. The violin plot (Figure 1 (a)) showed the number of feature RNA, the number of count RNA, the percentage of mitochondrial genes and the percentage of ribosomal genes. As Figure 1 (a) shows, there were differences among the eight datasets. In addition, the different quality control measures could be plotted as scatter plots (Figure 1 (b)), and the differences between each sample could also be clearly identified by the scatter plots.

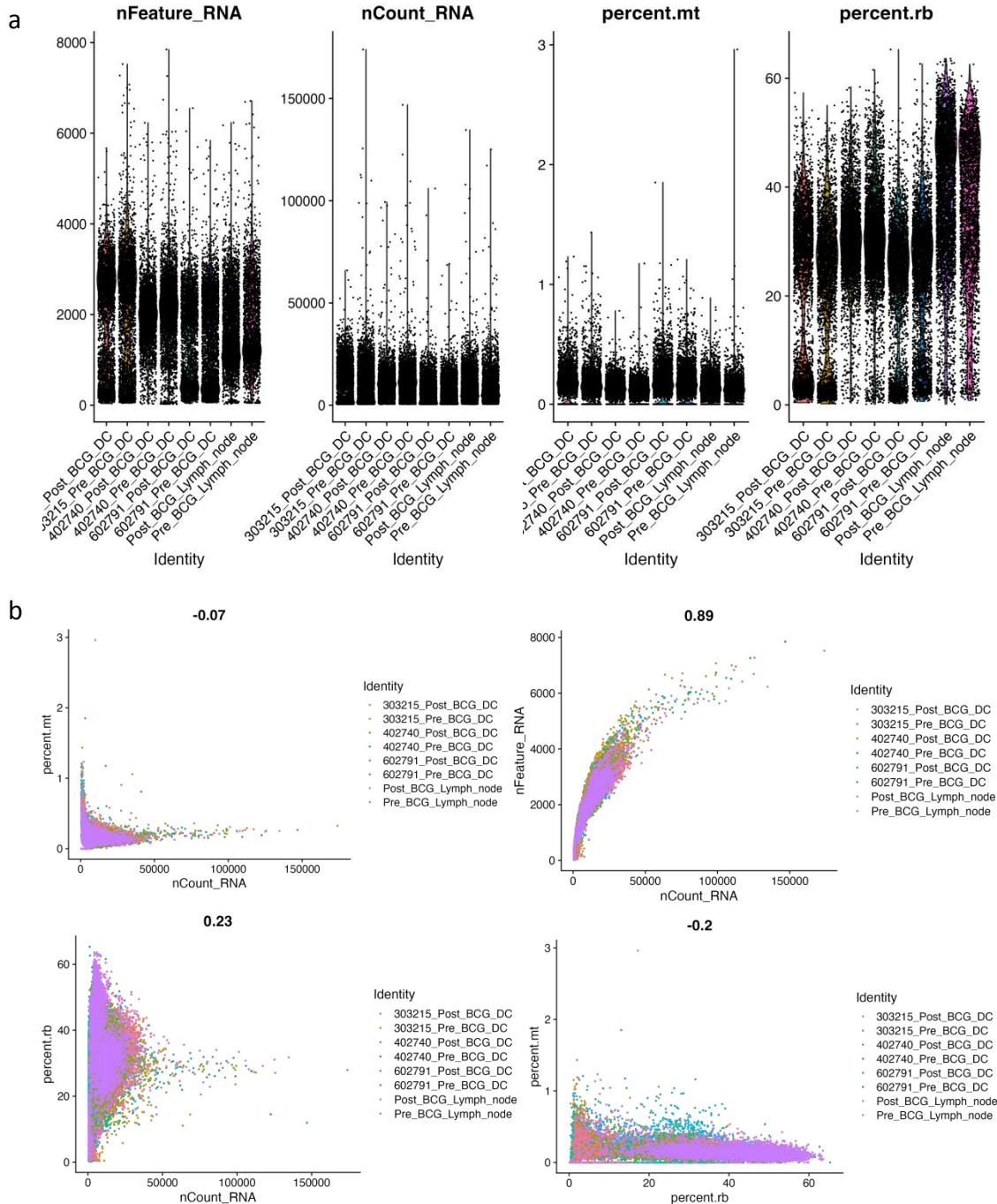


Figure1 (a) Violin plot showed gene counts (nCount_RNA), number of characteristic genes (nFeature_RNA), percentage of mitochondrial counts (percent.mt) and percentage of ribosomal counts (percent.rb) in each cell. This figure was used for quality control and selection filtering to remove low quality cells. The Y-axis of the first two figures was the number, and that of the last two figures was the percentage. (b) Intercomparison plot of metadata characteristics. Metadata features were compared to each other to show how the features are related. The numbers above each graph were the Pearson correlation coefficients. The four graphs showed the comparison between percent.mt and nCount_RNA, the comparison between nFeature_RNA and nCount_RNA, the comparison between percent.rb and nCount_RNA, and the comparison between percent.mt and percent.rb.

Next, sample-based filtering was performed. A common strategy was filtering out cells with few reads and genes present in at least a given percentage of cells. According to Figure 1(a), the filtering condition for this experiment was that the genes had to be expressed in at least three cells and be found in cells containing at least 200 identified genes. Cells needed a mitochondrial readout higher than 5% and lower than 20%. We also removed cells with ribosomal reads below 5% and above 60%. The total number of count RNA detected in the cells was limited to 39,000. These steps also aimed to remove low-quality/dying cells in which it was unlikely that too many cell type characteristics would remain. The same violin plot (Figure 2) was plotted again after filtering was completed to check the quality control results, and it could be seen that after filtering, there was little difference between the eight datasets.

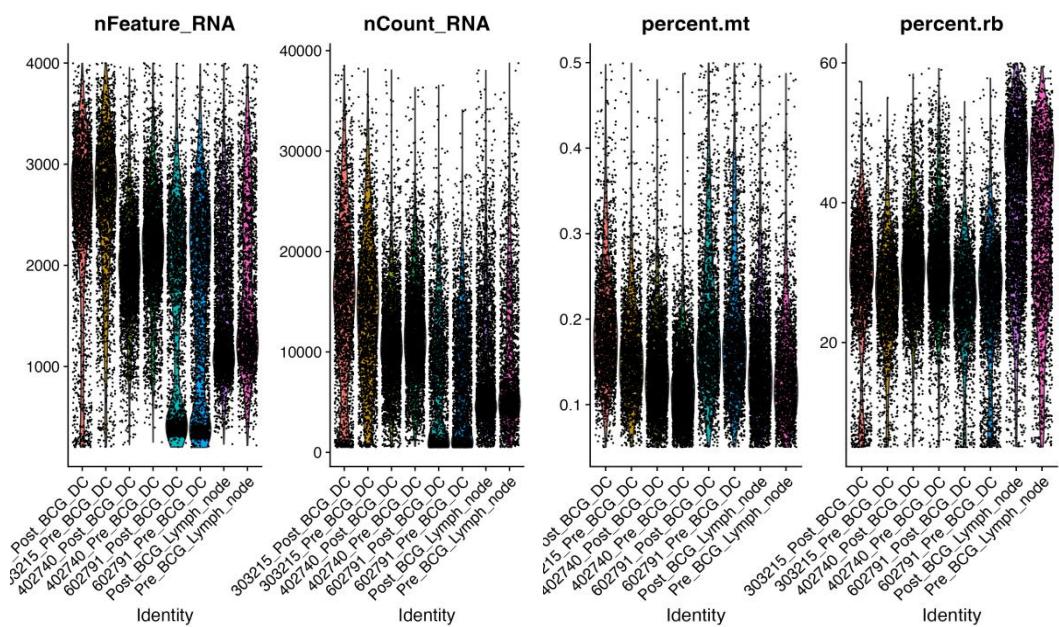
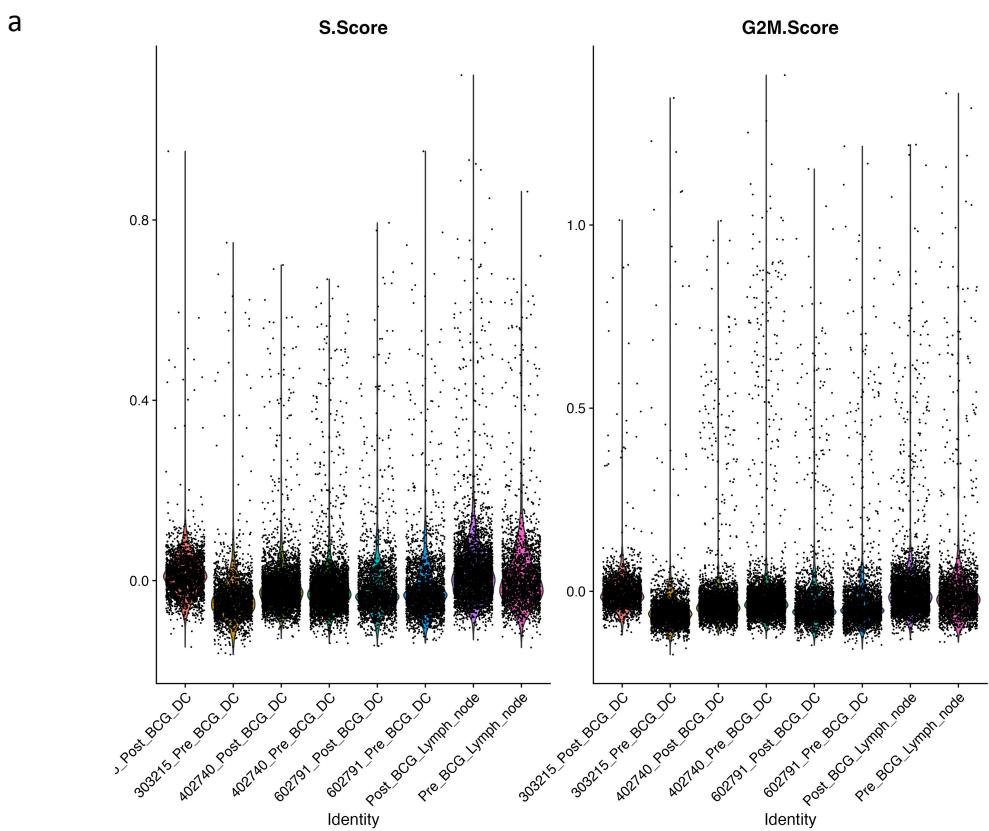


Figure 2 Violin plot showed gene counts (nCount_RNA), number of characteristic genes (nFeature_RNA), percentage of mitochondrial counts (percent.mt) and percentage of ribosomal counts (percent.rb) in each cell after quality control and selection filtering. The Y-axis of the first two figures was the number, and that of the last two figures was the percentage.

In this experiment, the impact of cell cycle on clustering was examined by scoring the cell cycle of these data in order to remove the interference of cell cycle on dimensionality reduction and clustering (Figure 3 (a)), and it was

discovered that there was no clear change in the expression of the cell cycle across these samples. Finally, two cells were sequenced as a pair in a doublet under the same cell barcode (e.g. captured in the same droplet). Doublets created by distinct cell types or states could be misclassified and cause downstream studies to be distorted, even though doublets of the same cell type were generally harmless in downstream analyses because the relative gene expression was preserved. In order to filter out Doublets in a single droplet, we used the DoubletFinder package to predict doublets and filtered out the doublets so that the retained cells were singlets(Figure 3 (b)).



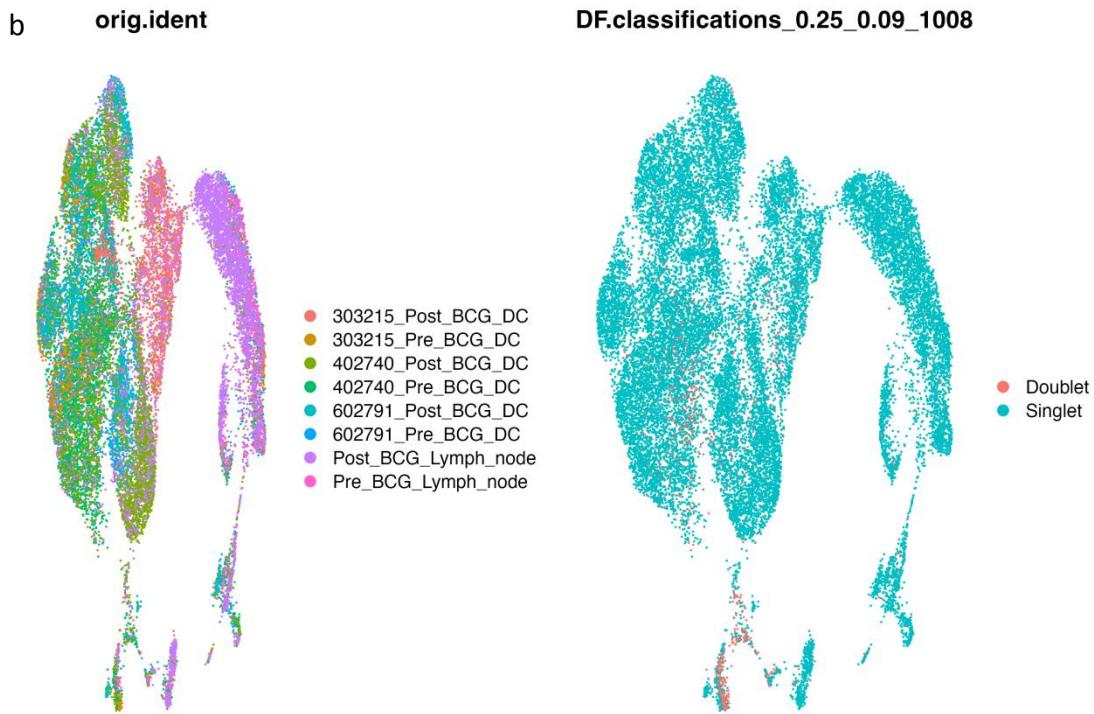


Figure 3 (a) S and G2M Cell cycle scoring chart. Cell Cycle Score provided an S-phase score and a G2M-phase score for each sample to compare differences in cell cycle per sample. (b) Distribution of doublets and singlets in the datasets. The left picture showed the specific distribution of each sample, and the right picture showed the specific distribution of singlet and doublet in each sample.

3.2 Dimensionality reduction

Then, to identify genes that were highly variable across cells and provide a clear separation of cell clusters, we found 2000 highly variable genes and 14153 non-highly variable genes in a single cell using the '*FindVariableFeatures*' function. The top 10 highly variable genes were CXCL2, S100A12, CCL2, TAP, CXCL9, CCL8, CRABP1, ENSBTAG00000048980, ENSBTAG00000010155 and GRO1, and they are shown in Figure 4.

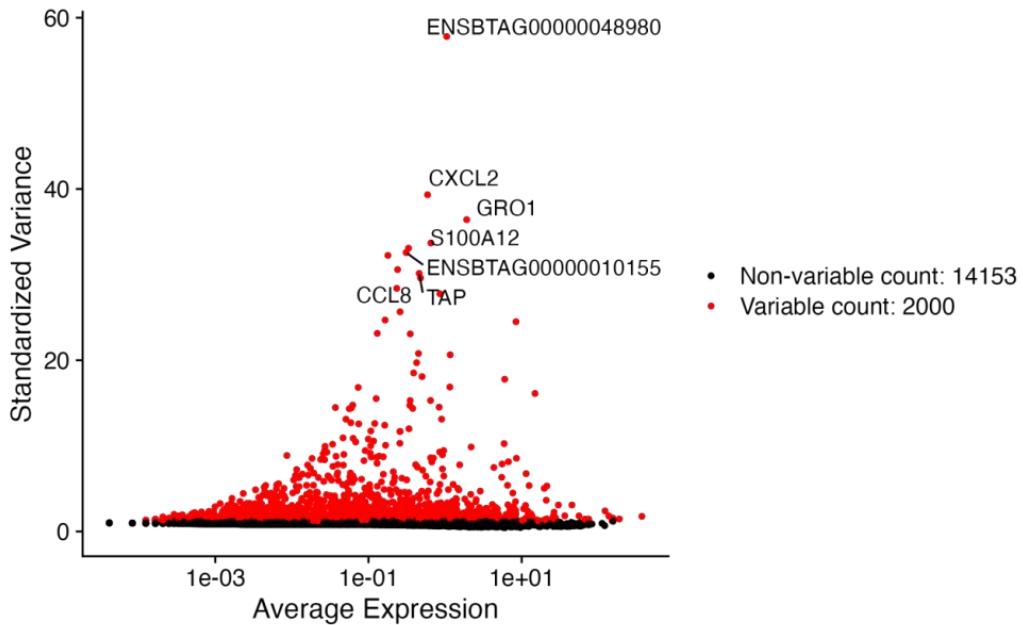
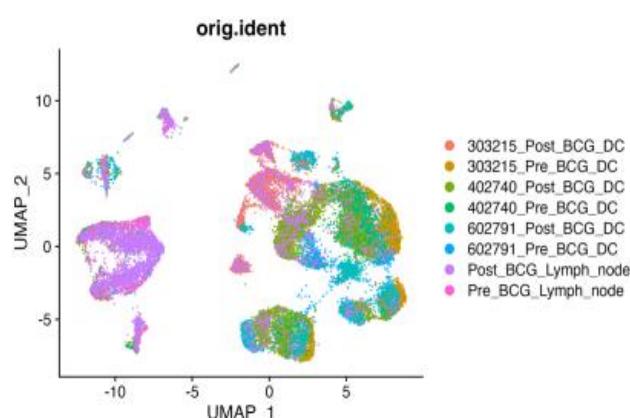
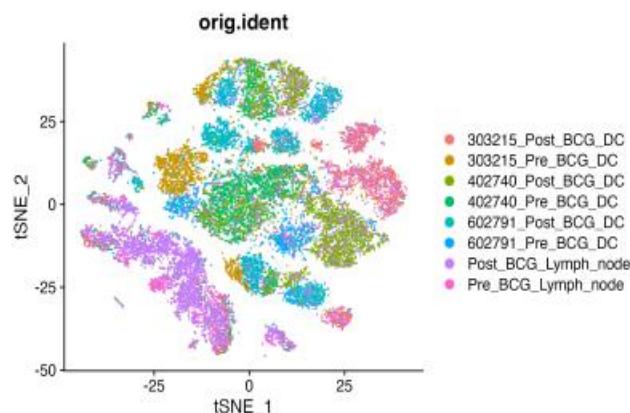
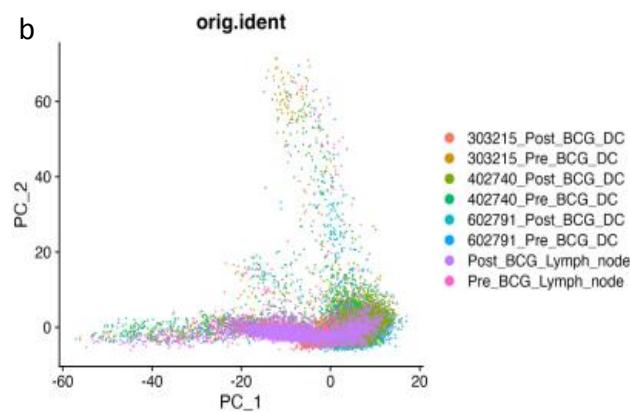
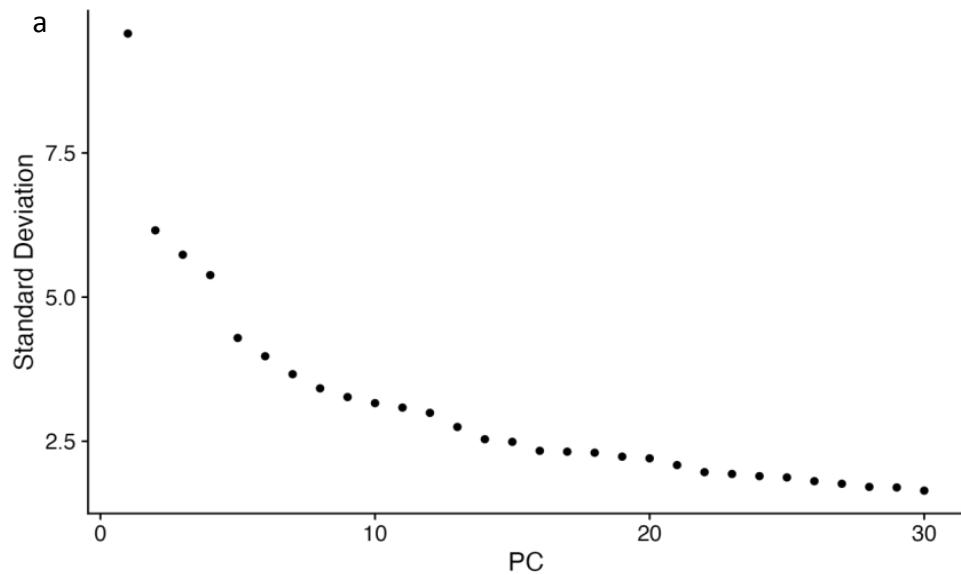


Figure 4 Cross-cell detection plot of highly variable genes. 2000 variable genes were found in the dataset to distinguish between cell types.

With the help of the Seurat package, cells were clustered together based on their PCA scores to eliminate the different technical noises in each feature of the scRNA-seq data. ElbowPlot was used to rank PCs according to standard deviation (Figure 5 (a)). According to the elbow plot, we set PCs as 30. The cell clusters were then visualised utilising two distinct functions ('UMAP' and 'tSNE'), and rows of PCA, UMAP and tSNE were plotted for comparison (Figure 5 (b)). It was concluded here that the dataset for this experiment needed to be corrected for batch effects before clustering and differential gene expression analysis could be performed. After integration, the Seurat package contained a new matrix with a matrix of batch-corrected expressions. The "RNA" assay object still had uncorrected values, which could be switched back and forth. Then, downstream analysis and visualisation were performed using this new integrated matrix. Here, we scaled the integrated data, performed PCA, and then used UMAP and TSNE to visualise the findings (Figure 5 (c)), and the integrated dataset was to be clustered by cell type.



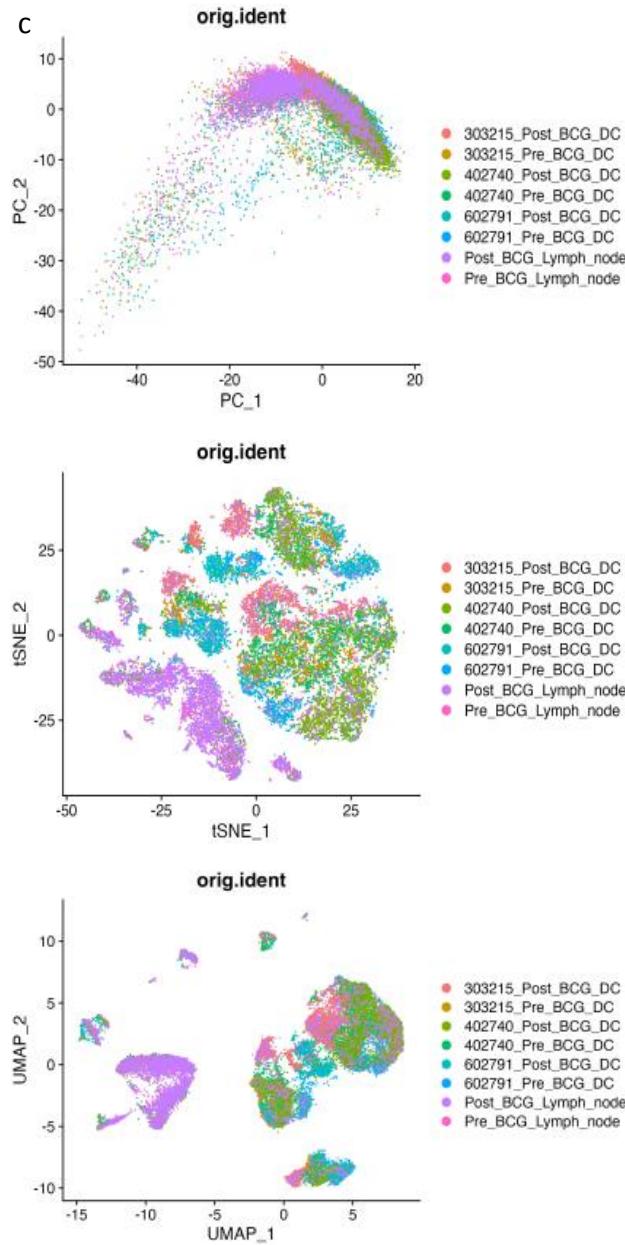


Figure 5 (a) Elbow plots sorting PCs by standard deviation (b) Distributions of all eight samples were displayed in PCA, tSNE and UMAP plots before integration. (c) Distributions of all eight samples were displayed in PCA, tSNE and UMAP plots after integration.

3.3 Cluster analysis and annotate cell types

Followed by quality control, doublet removing, dimensionality reduction, elimination of batch effects and other steps, each cell cluster contained cells from each sample distributed equally. We clustered the dataset using the

'FindNeighbors' and 'FindClusters' functions and obtained a total of 18 cell clusters, and then the 'UMAP' function was applied to visualise these 18 cell clusters (Figure 6 (a)). Ten cell types were annotated using the SingleR package (Figure 6 (b)), representing different kinds of cells, including dendritic cells, NK cells, CD4+ T cells, CD8+ T cells, monocytes, basophils, T cells, B cells, progenitor cells, and neutrophils. This was done to show the identity of the cells represented by these clusters.

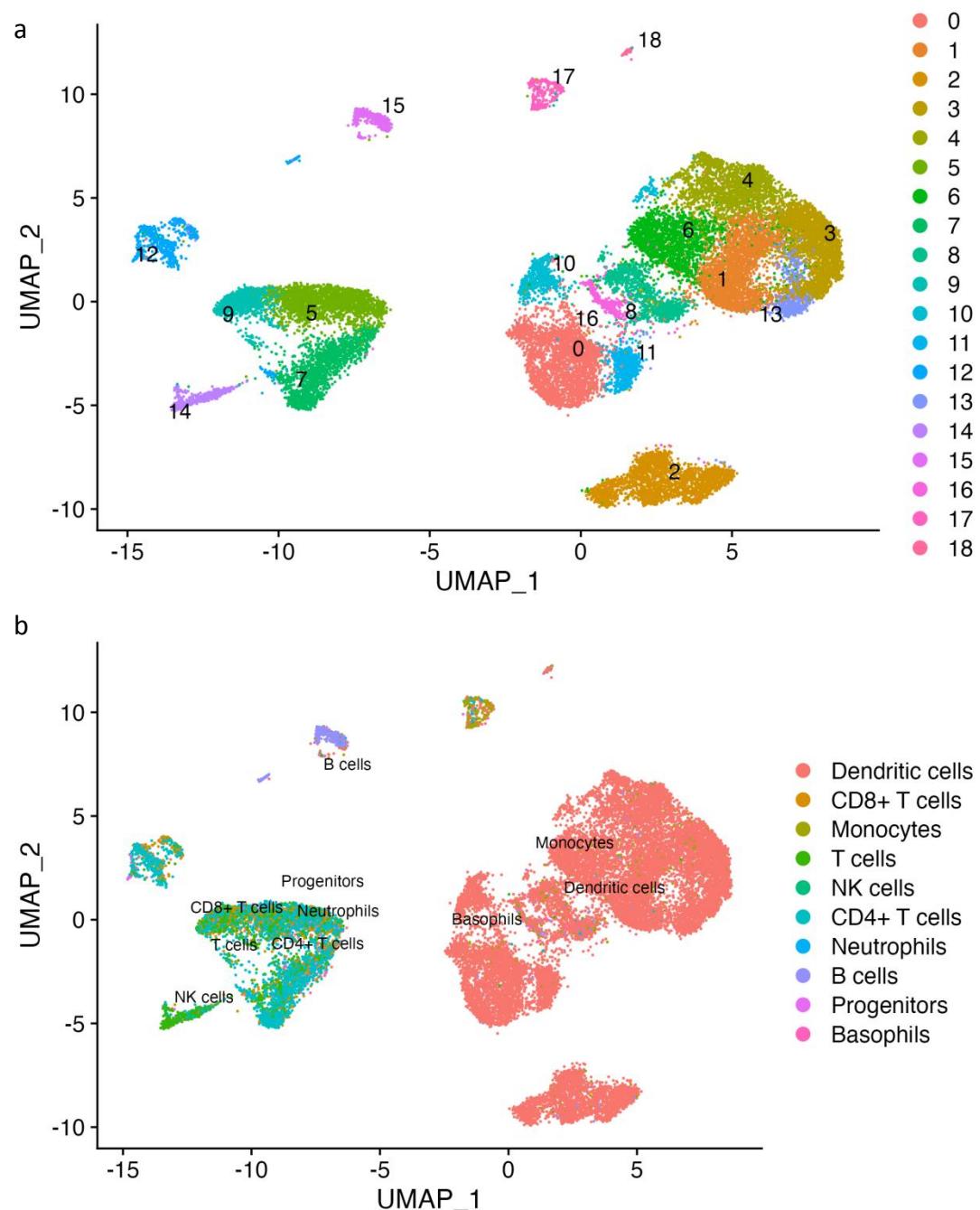
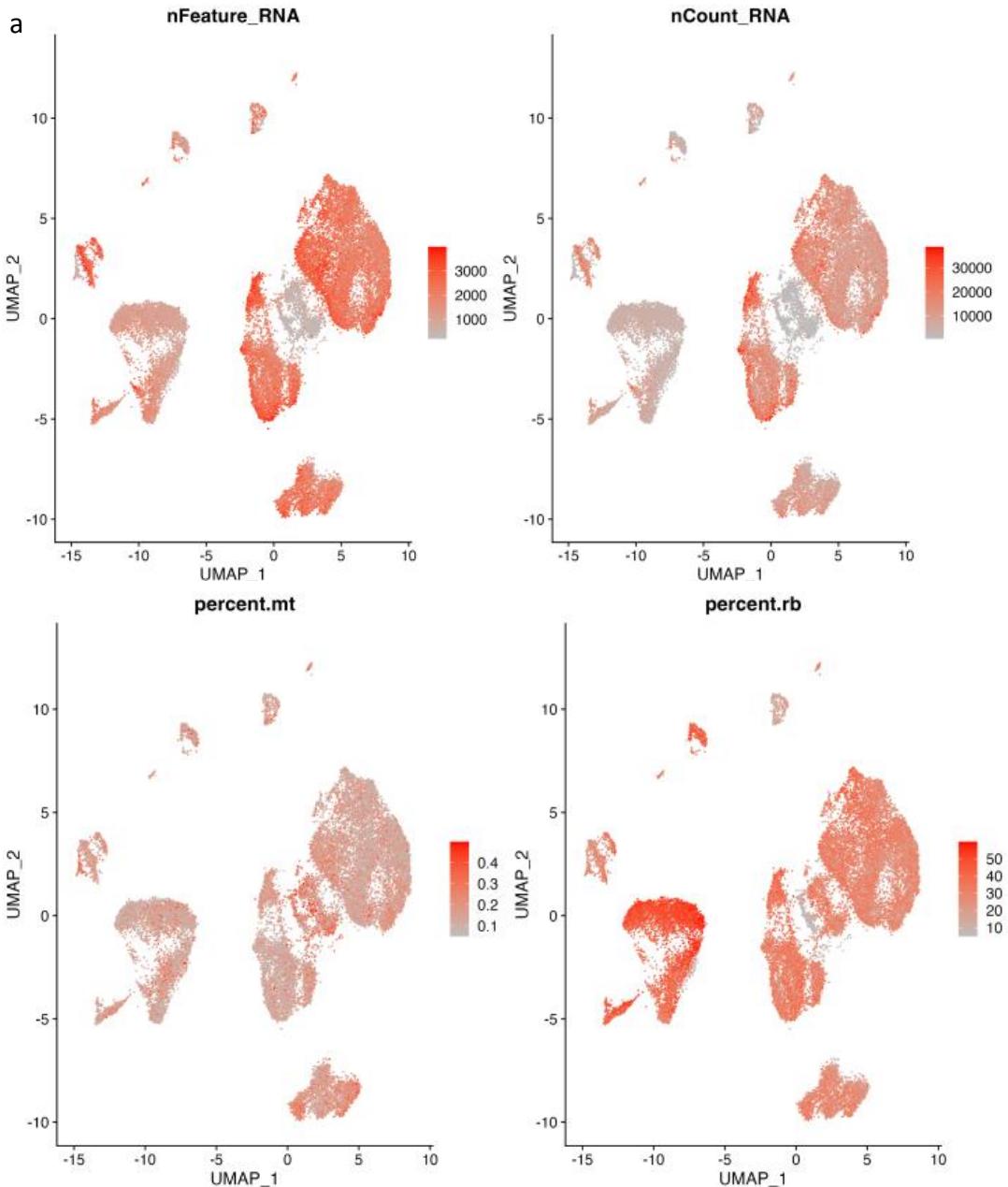
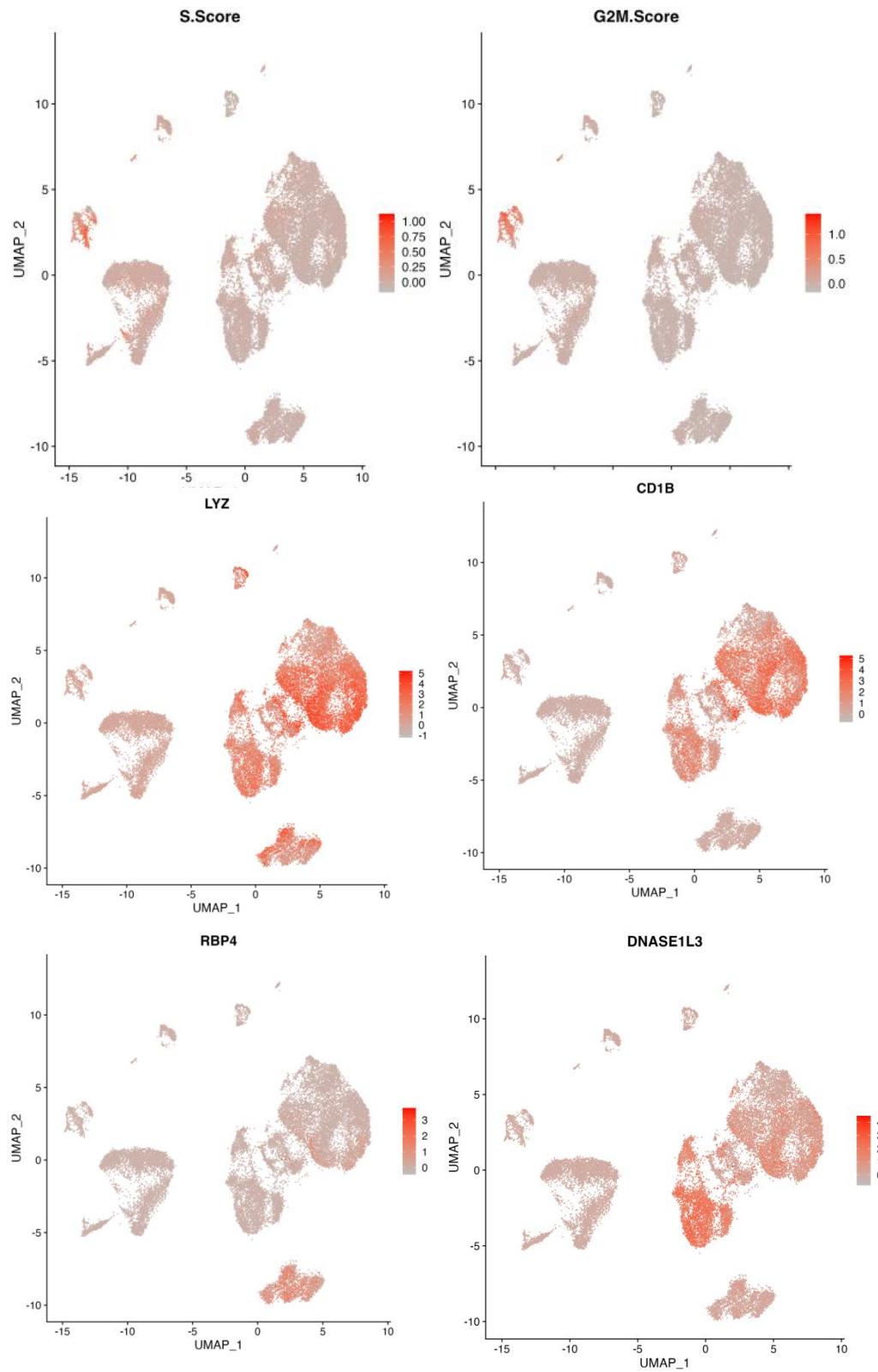


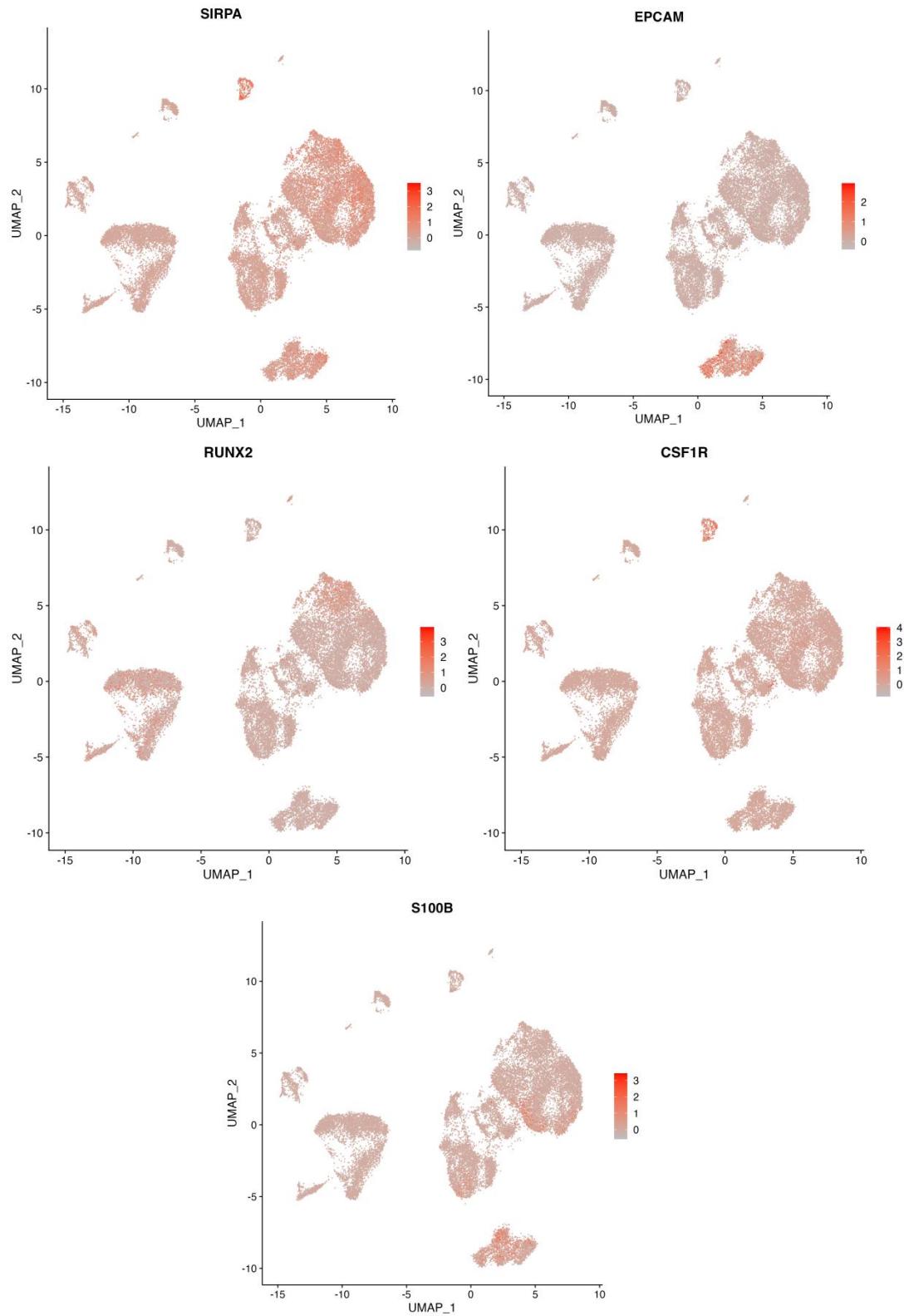
Figure 6 (a) UMAP plot of the scRNA-seq data for 18 cell clusters. The entire dataset was

clustered into 18 cell clusters. (b) UMAP plots of 10 immune cell types annotated by SingleR
The entire dataset was defined by various cell types.

Among these, clusters 0, 1, 2, 3, 4, 6, 8, 10,11,13,16 and 18 were the major cell clusters assigned as dendritic cells. By analyzing FeaturePlot of nFeature_count, nRNA_count, ribosomal percentage, mitochondrial percentage, cell cycle, LYZ, CD1B, RBP4, DNASE1L3, SIRPA, EPCAM, RUNX2, CSF1R, and S100B (Figure 7 (a)), we split the clusters like this: cluster 1, 3, 4, 6 and 13 as dendritic cells group 1, cluster 8 and 16 as dendritic cells group 2, cluster 2 as dendritic cells group 3, cluster 0, 10 and 11 as dendritic cells group 4 and cluster 18 as dendritic cells group 5 (Figure 7 (b)). We believed that each dendritic cell group represented a type of dendritic cell, from immature to mature, like conventional dendritic cell (cDC) and plasmacytoid dendritic cell (pDC). After that, we used '*FeaturePlot*' function to plot macrophage marker gene CD14 and the result is shown in Figure 7 (c). It showed that cluster 17 was the major cell cluster associated with macrophage (Figure 7 (b)). The identity of each cell cluster was then ascertained by differential gene expression analysis through the screening of cluster-specific marker genes.







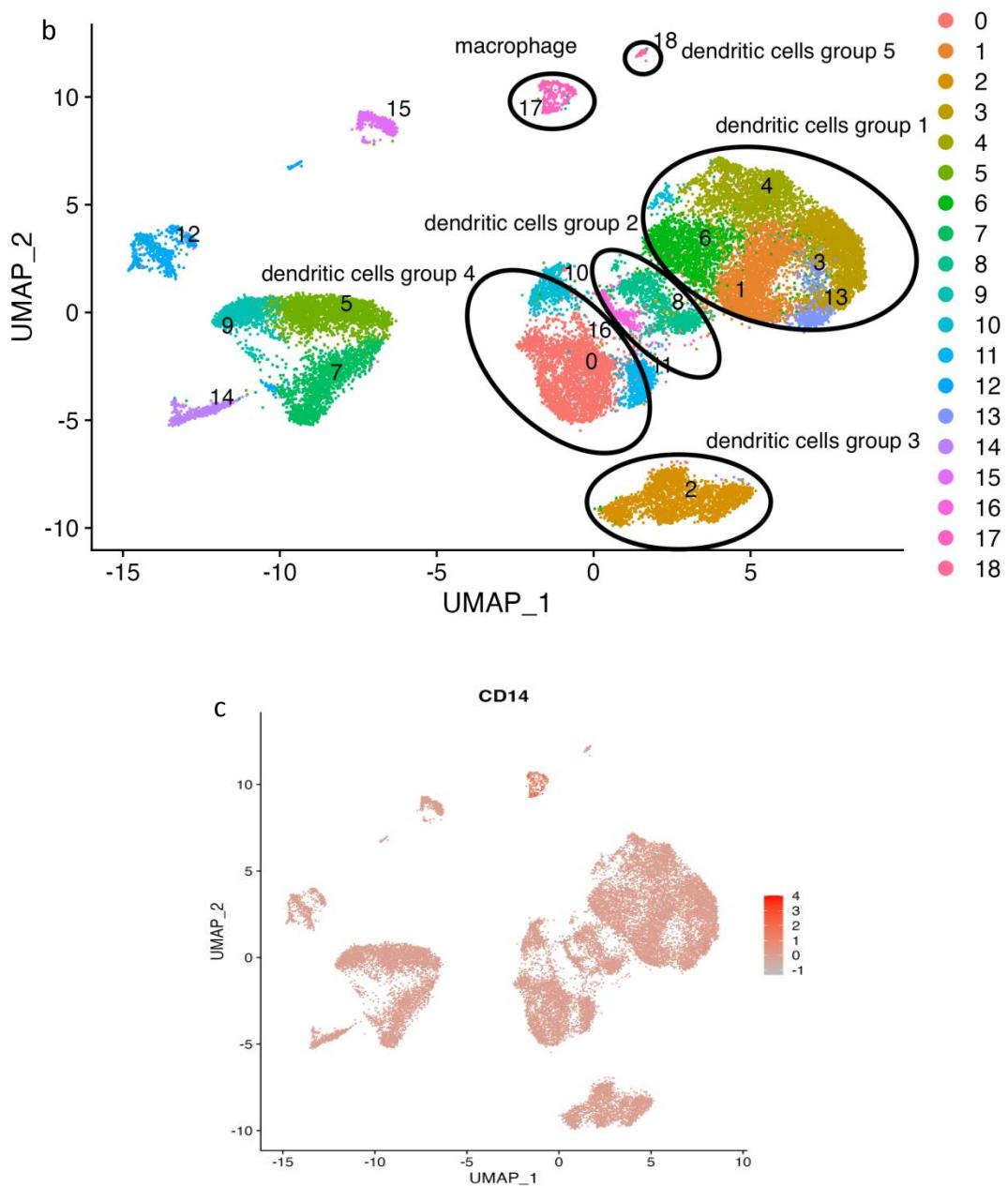


Figure 7 (a) FeaturePlot of nFeature_count, nRNA_count, ribosomal percentage, mitochondrial percentage, cell cycle, LYZ, CD1B, RBP4, DNASE1L3, SIRPA, EPCAM, RUNX2, CSF1R, S100B. These figures were used to find five dendritic cell groups (b) UMAP plot of five dendritic cell groups and macrophage group (c) FeaturePlot of CD14. This figure was used to find macrophages.

3.4 Differential expression analysis

3.4.1 Identification of differentially expressed genes in dendritic cells group 1 (cluster 1, 3, 4, 6 and 13)

On the clustered and annotated datasets, We used the Seurat package to

analyse differential expressions to discover markers that characterise clusters. Differentially expressed genes of dendritic cells group 1 (cluster 1, 3, 4, 6 and 13) were shown in Table 1. 1021 differentially expressed genes in total were found. The count of -log10P went up 200, which meant some genes were extremely significant. Volcano plots (Figure 8 (a)) allowed for more visual identification of up-and-down-regulated genes. We set ident.1 as after BCG injection and ident.2 as before BCG injection. Ident.1 was up-regulated relative to ident.2 when avg_log2FC was positive. When avg_log2FC was negative, ident.1 was down-regulated relative to ident.2. To more visually detect differences in the location and expression of these differentially expressed genes within different types of immune cells, some differentially expressed genes which show the biggest differences were next visualised using ‘*VlnPlot*’ and ‘*FeaturePlot*’ function (Figure 8 (b)(c)(d)). It could be seen from Figures 8 (b) and (c) that genes SAMS1, SIAH2, RNF19B and RAB9A were up-regulated and genes PAK2 and LY75 were down-regulated in dendritic cells. Figure 8 (d) shows the expression levels of these differential genes in other dendritic cell groups.

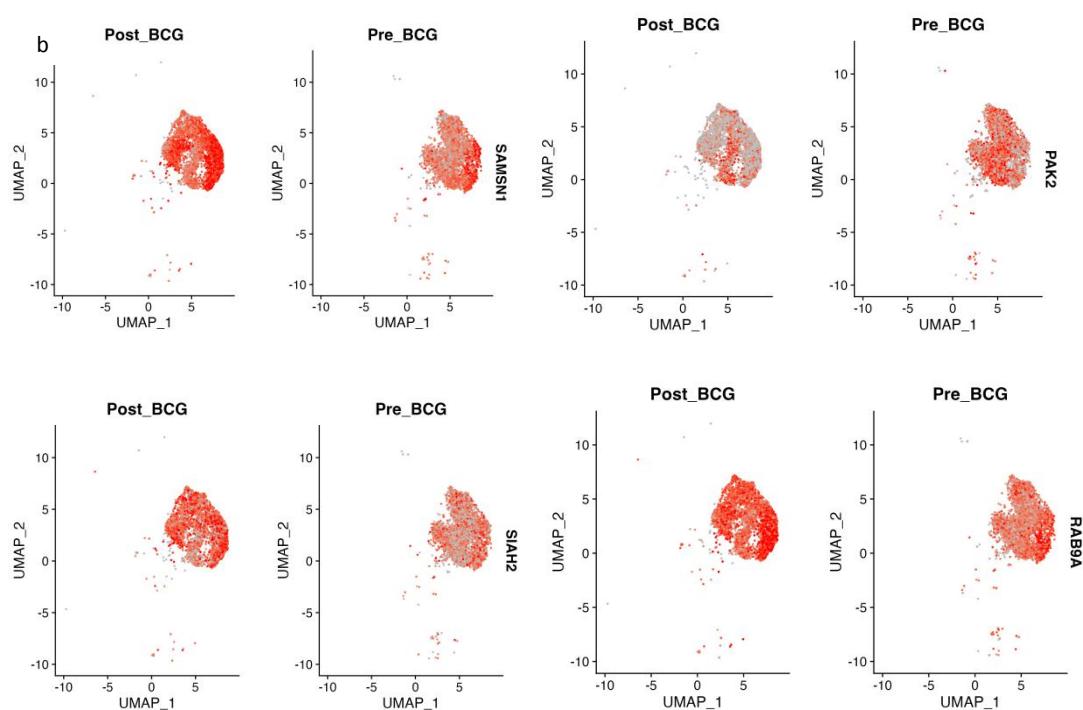
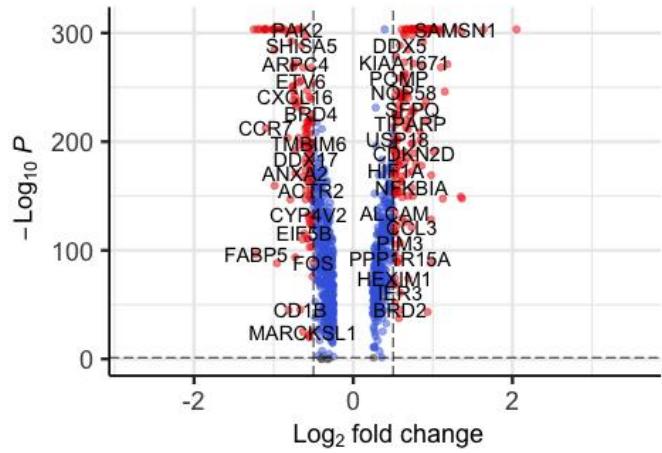
	p_val	avg_log2FC	pct. 1	pct. 2	p_val_adj
SAMS1	0	1.288418583	0.977	0.891	0
PAK2	0	-0.698444521	0.451	0.761	0
SIAH2	0	0.725907281	0.868	0.653	0
RAB9A	0	0.852371532	0.97	0.909	0
LY75	0	-0.997550617	0.421	0.745	0
RNF19B	0	0.890338225	0.978	0.894	0
JAK1	0	-0.770568299	0.847	0.952	0
SMAP2	0	-1.105588028	0.735	0.931	0
BTG1	0	0.75580865	0.998	0.995	0
PTGES3	0	0.910372779	0.981	0.91	0

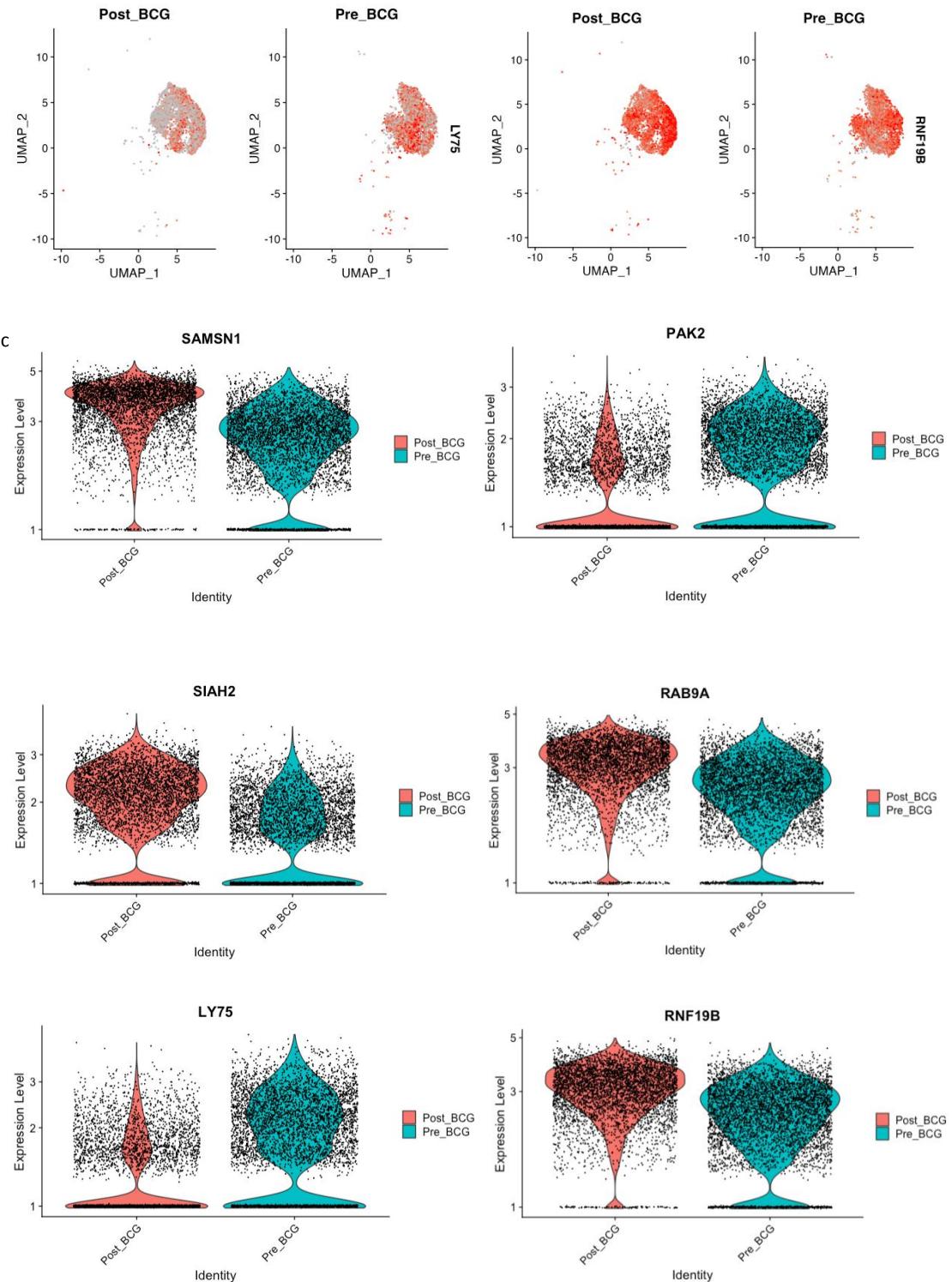
Table 1. Top 10 list of genes differentially expressed in dendritic cells group 1 (cluster 1, 3, 4, 6 and 13)

Volcano plot

a
EnhancedVolcano

● NS ● Log₂ FC ● p-value ● p – value and log₂ FC





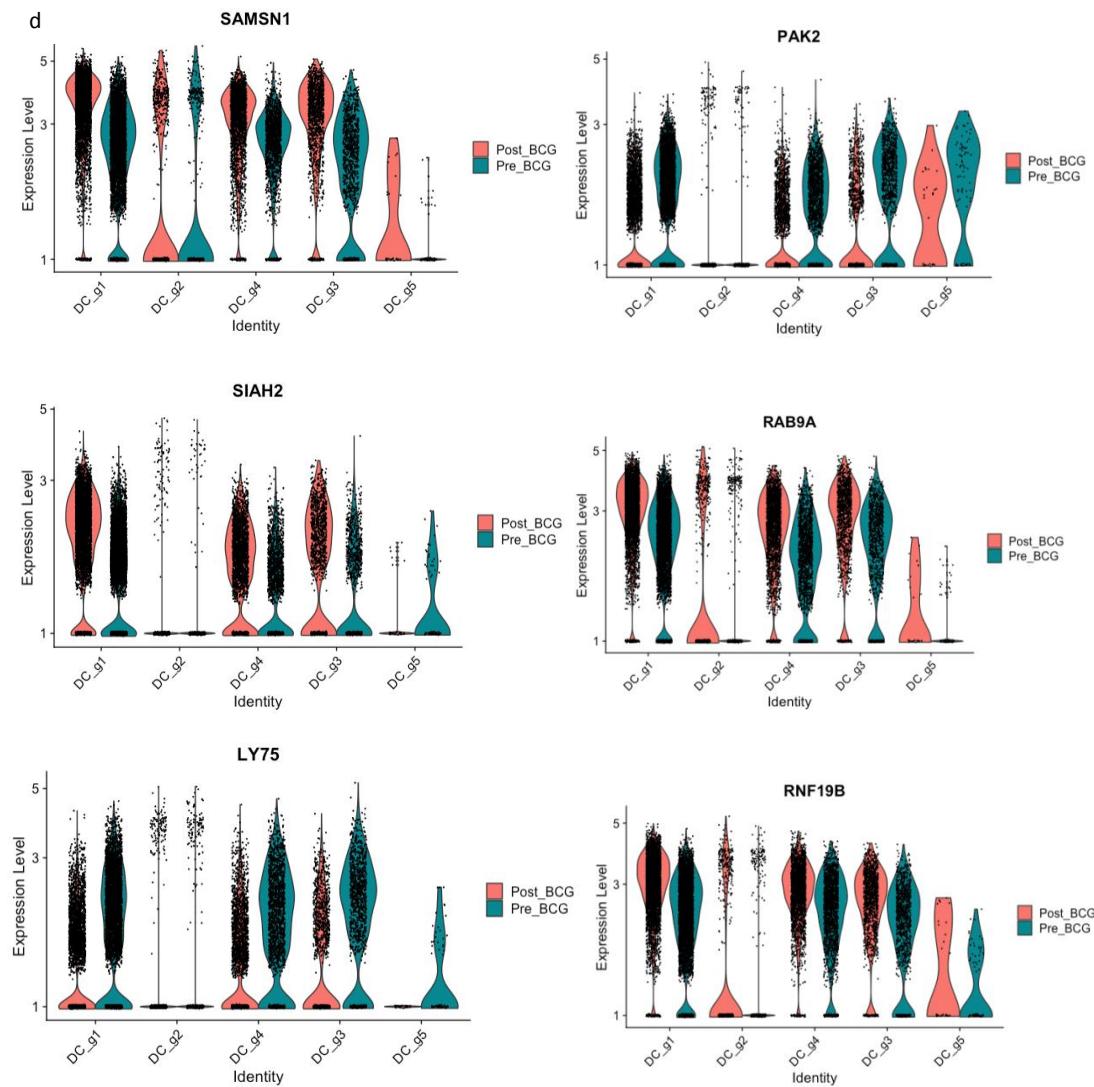


Figure 8 (a) Volcano plot of differential gene expression in dendritic cells group 1 (cluster 1, 3, 4, 6 and 13)
 (b) FeaturePlot of differential genes in dendritic cells before and after BCG injection (c) VInPlot of differential genes in dendritic cells group 1 (cluster 1, 3, 4, 6 and 13) before and after BCG injection (d) VInPlot of differential genes in all dendritic cells groups before and after BCG injection

3.4.2 Identification of differentially expressed genes in dendritic cells group 2 (cluster 8 and 16)

Differential expression genes of dendritic cells group 2 (cluster 8 and 16) before and after BCG injection were shown in Table 2. 288 differentially expressed genes in total were found. Volcano plots (Figure 9) allowed for more visual identification of up-and-down-regulated genes. Here we suspected cluster 8 was inferior because of its low nFeature_count and nRNA_count (Figure 7 (a)). We did not focus on cluster 8 because of its low quality. This

group's up/down-regulated genes might be due to their contamination.

	p_val	avg_log2FC	pct. 1	pct. 2	p_val_adj
RBM25	2.56E-15	1.325747469	0.415	0.242	4.13E-11
SRSF11	9.34E-13	0.994849338	0.306	0.147	1.51E-08
SREK1	1.98E-10	0.890755678	0.2	0.079	3.21E-06
RPS2	3.87E-10	-0.286372903	0.964	0.975	6.24E-06
CD74	1.26E-09	-0.476528026	0.781	0.828	2.04E-05
SFSWAP	2.92E-09	0.745404215	0.133	0.04	4.72E-05
SAFB2	4.06E-09	0.833177244	0.304	0.167	6.57E-05
RBBP6	4.18E-09	0.795431247	0.542	0.406	6.75E-05
RPS8	2.37E-08	-0.272184225	0.933	0.97	0.000383552
TRRAP	3.04E-08	0.858118991	0.114	0.033	0.000491845

Table 2. Top 10 list of genes differentially expressed in dendritic cells group 2 (cluster 8 and 16)

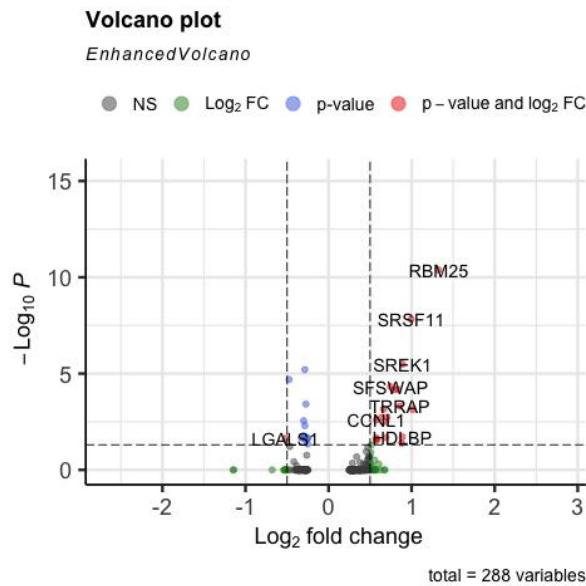


Figure 9 Volcano plot of differential gene expression in dendritic cells group 2 (cluster 8 and 16)

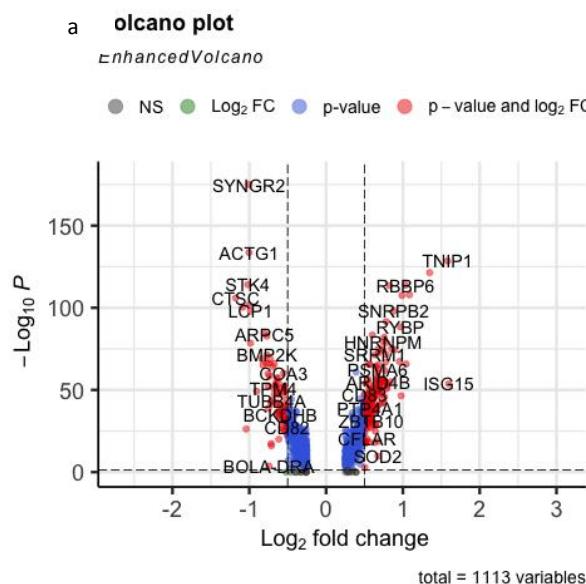
3.4.3 Identification of differentially expressed genes in dendritic cells group 3 (cluster 2)

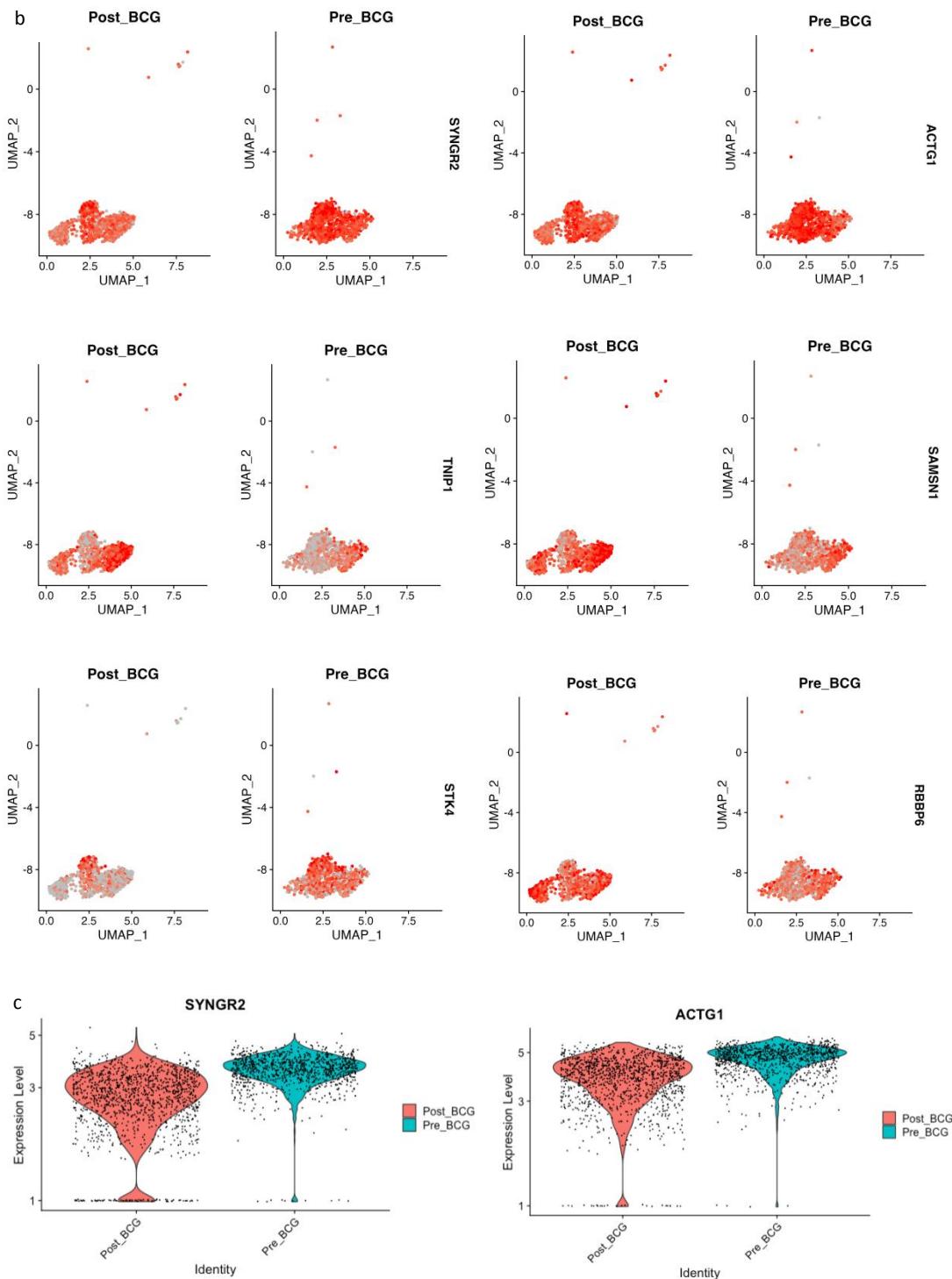
Differential expression genes of dendritic cells group 3 (cluster 2) before and after BCG injection are shown in Table 3. Volcano plots showed 1113 differentially expressed genes in Figure 10 (a). Some differentially expressed

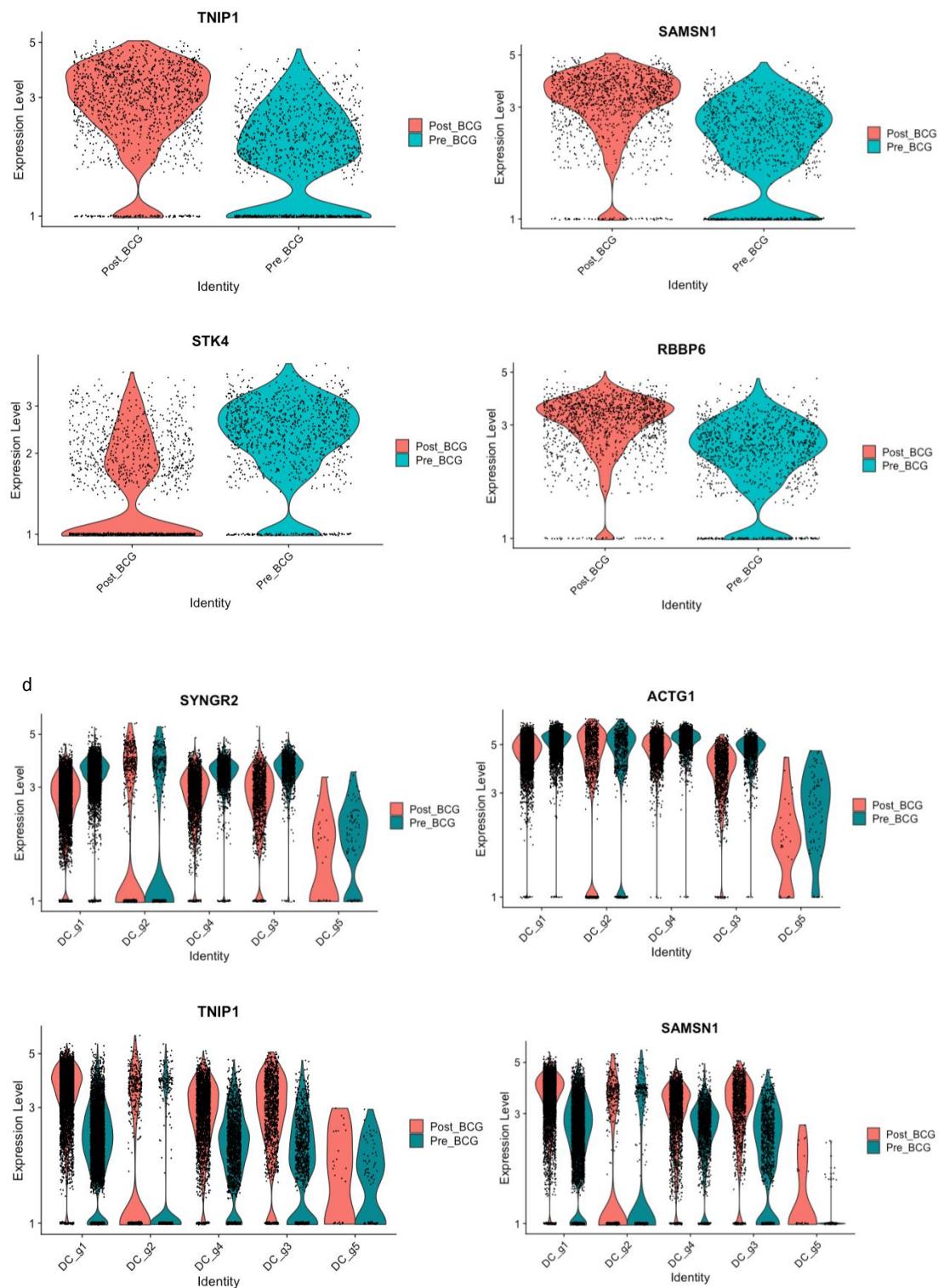
genes were next visualised using ‘*VlnPlot*’ and ‘*FeaturePlot*’ functions (Figure 10 (b)(c)). In dendritic cells group 3, genes SYNGR2, ACTG1, and STK4 were down-regulated, and genes TNIP1, SAMSN1 and RBBP6 were up-regulated. Figure 10 (d) shows the expression levels of these differential genes in other dendritic cell groups.

	p_val	avg_log2FC	pct. 1	pct. 2	p_val_adj
SYNGR2	6.08E-180	-1.00523417	0.941	0.991	9.82E-176
ACTG1	1.47E-138	-1.003721583	0.98	0.996	2.38E-134
TNIP1	1.96E-133	1.582677026	0.921	0.727	3.16E-129
SAMSN1	2.53E-126	1.348193729	0.949	0.799	4.09E-122
STK4	5.05E-119	-1.020973175	0.527	0.891	8.16E-115
RBBP6	8.75E-119	1.035007119	0.965	0.879	1.41E-114
RBM39	1.79E-118	0.826257051	0.98	0.929	2.90E-114
SAFB2	7.37E-113	1.084120945	0.861	0.611	1.19E-108
USP12	1.91E-112	0.988285316	0.963	0.839	3.08E-108
CTSC	6.77E-111	-1.177377154	0.622	0.906	1.09E-106

Table 3. Top 10 list of genes differentially expressed in dendritic cells group 3 (cluster 2)







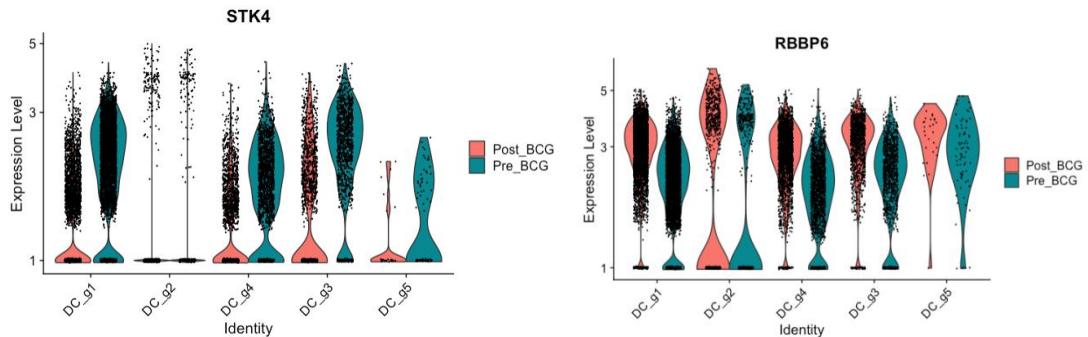


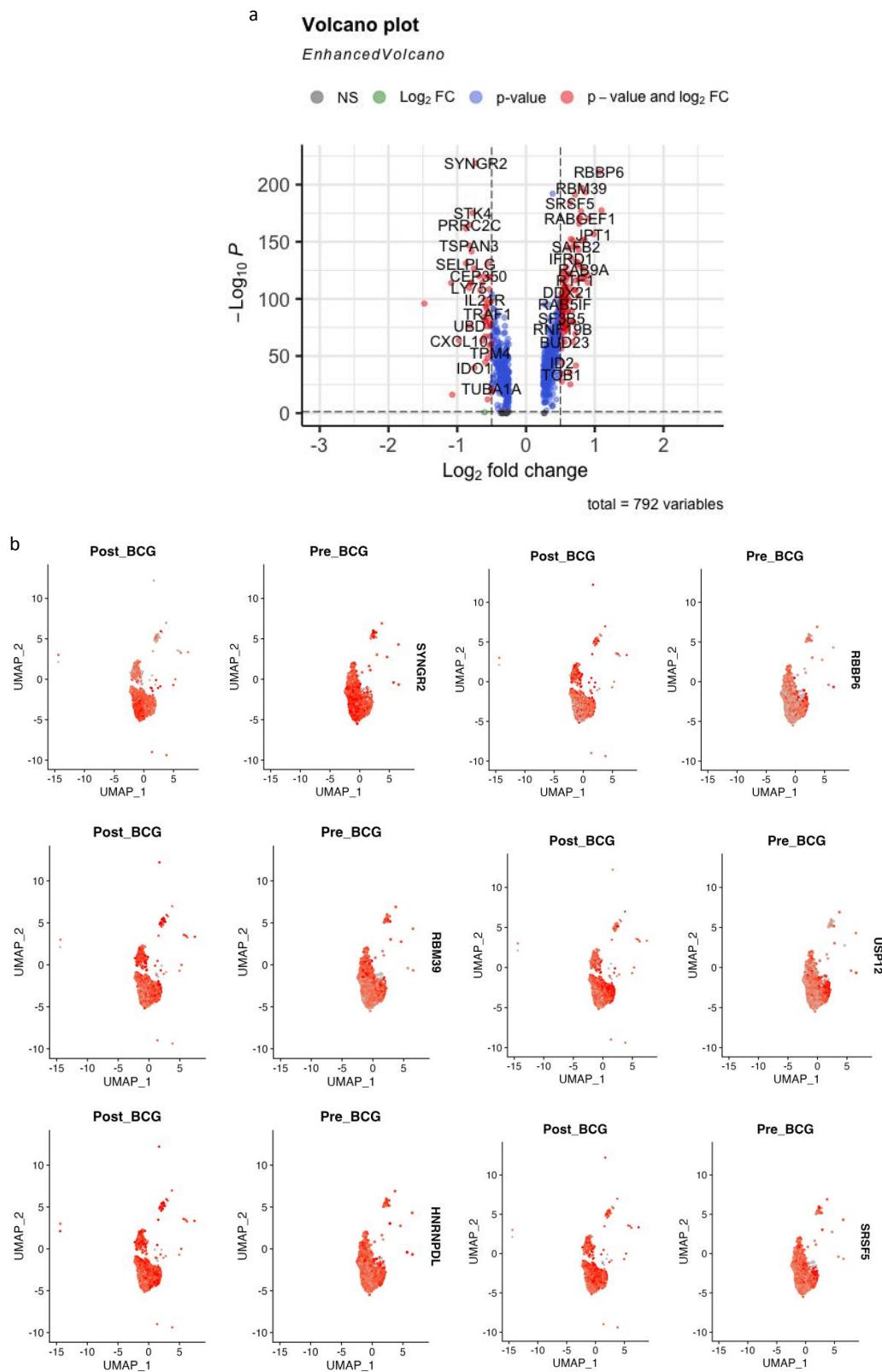
Figure 10 (a) Volcano plot of differential gene expression in dendritic cells group 3 (cluster 2) (b)
 FeaturePlot of differential genes in dendritic cells group 3 (cluster 2) before and after BCG injection (c)
 VinPlot of differential genes in dendritic cells group 3 (cluster 2) before and after BCG injection (d)
 VinPlot of differential genes in all dendritic cells groups before and after BCG injection

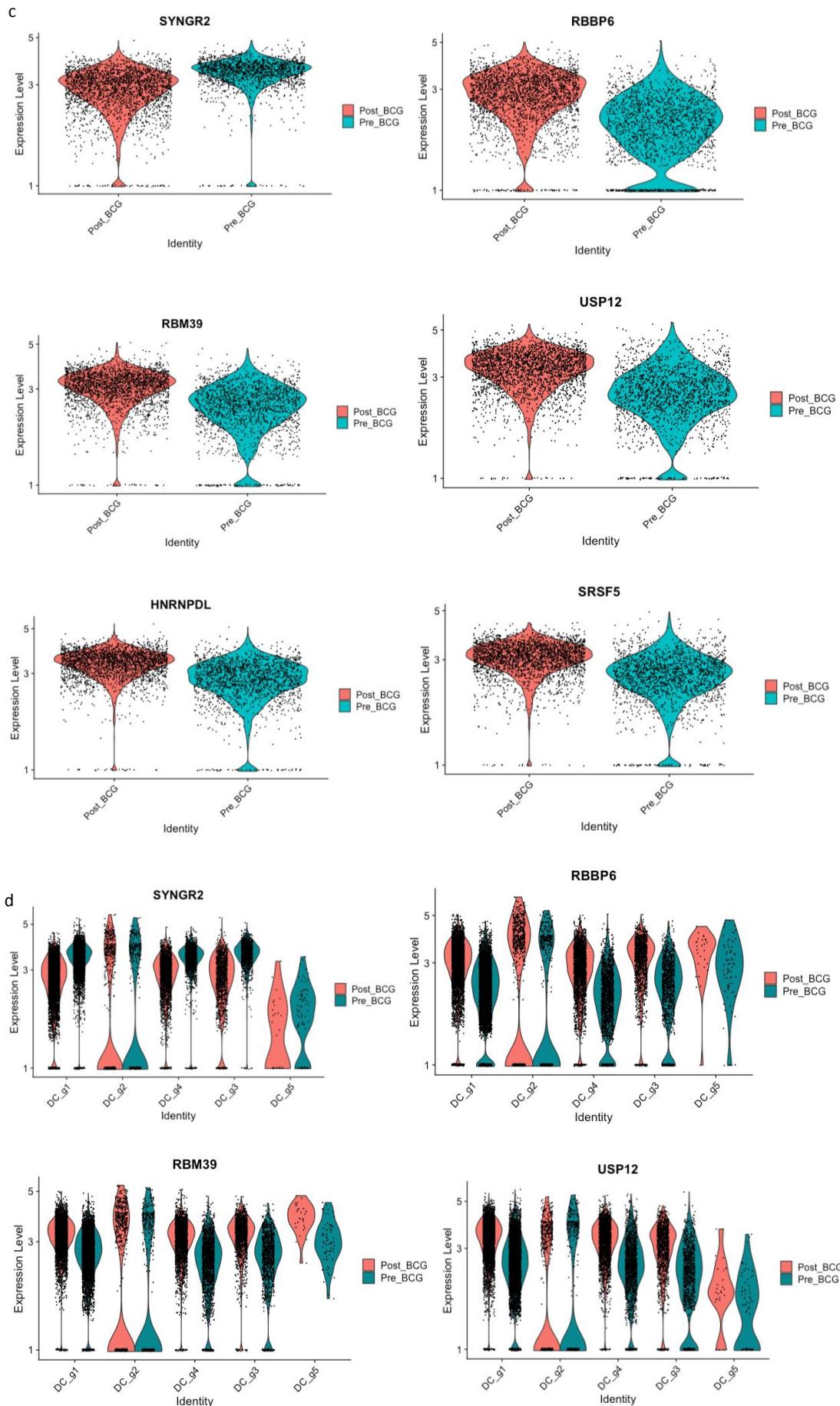
3.4.4 Identification of differentially expressed genes in dendritic cells group 4 (cluster 0, 10 and 11)

Differential expression genes of dendritic cells group 4 (cluster 0, 10 and 11) before and after BCG injection was shown in Table 4. Volcano plots showed 792 differentially expressed genes in total in Figure 11 (a). Using the ‘*VInPlot*’ and ‘*FeaturePlot*’ functions, certain differentially expressed genes were visualised (Figure 11 (b)(c)). Genes RBBP6, RBM39, USP12, HNRNPDL, and SRSF5, were up-regulated, while genes SYNGR2 were down-regulated, as seen in Figures 11 (b) and (c). The expression levels of these distinct genes in other dendritic cell groups are depicted in Figure 11 (d).

	p_val	avg_log2FC	pct. 1	pct. 2	p_val_adj
SYNGR2	1.15E-223	-0.729721494	0.977	0.981	1.86E-219
RBBP6	5.59E-216	1.061933422	0.972	0.873	9.03E-212
RBM39	5.22E-202	0.806751495	0.988	0.951	8.43E-198
USP12	3.51E-198	0.866099812	0.991	0.955	5.68E-194
RPS28	5.07E-197	0.387344396	0.999	0.998	8.19E-193
HNRNPDL	4.59E-196	0.712501168	0.992	0.969	7.42E-192
SRSF5	1.25E-188	0.640562367	0.994	0.963	2.02E-184
TNIP1	1.78E-182	1.097035703	0.972	0.888	2.87E-178
RYBP	1.29E-181	0.80057724	0.832	0.539	2.08E-177
STK4	3.27E-180	-0.776329188	0.397	0.784	5.29E-176

Table 4. Top 10 list of genes differentially expressed in dendritic cells group 4 (cluster 0, 10 and 11)





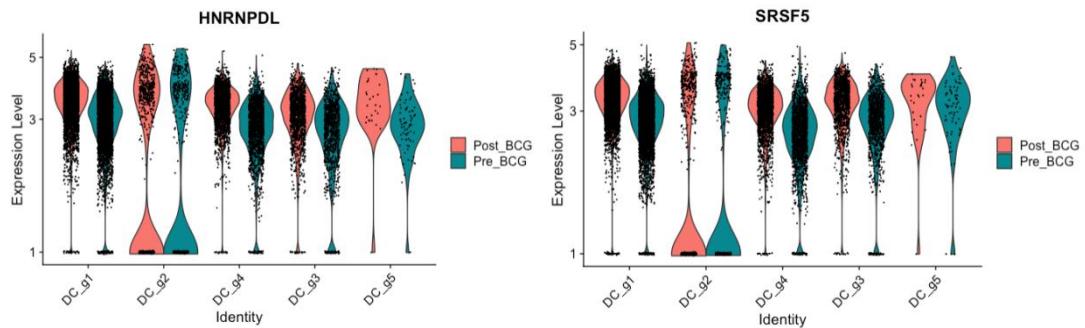


Figure 11 (a) Volcano plot of differential gene expression in dendritic cells group 4 (cluster 0, 10 and 11)
 (b) FeaturePlot of differential genes in dendritic cells group 4 (cluster 0, 10 and 11) before and after BCG injection
 (c) VinPlot of differential genes in dendritic cells group 4 (cluster 0, 10 and 11) before and after BCG injection
 (d) VinPlot of differential genes in all dendritic cells groups before and after BCG injection

3.4.5 Identification of differentially expressed genes in dendritic cells group 5 (cluster 18)

Differential expression genes of dendritic cells group 5 (cluster 18) before and after BCG injection was shown in Table 5. Volcano plots showed 2150 differentially expressed genes (Figure 12). However, as the number of genes in this cluster was not large enough and the different level of the differential gene was also not significant enough, we would not continue to focus on this cluster.

	p_val	avg_log2FC	pct. 1	pct. 2	p_val_adj
TRA2B	5.42E-09	1.248741223	0.964	0.986	8.75E-05
EPST11	5.68E-09	-1.613063656	0.393	0.959	9.18E-05
HNRNPK	2.81E-08	0.873206861	1	0.946	0.000454027
EID1	1.07E-07	-1.038691283	0.393	0.838	0.001731516
DCTN6	1.07E-07	0.986638495	0.714	0.189	0.001732143
HNRNPM	1.27E-07	1.008959378	0.964	0.973	0.002055207
RBM39	1.91E-07	1.050602949	1	1	0.003087456
PLAC8B	2.03E-07	-1.221632179	0.964	0.986	0.003279881
EIF4G2	3.48E-07	1.024875904	0.964	0.973	0.005623695
IFRD1	4.88E-07	1.293231948	0.929	0.581	0.007875657

Table 5. Top 10 list of genes differentially expressed in dendritic cells group 5 (cluster 18)

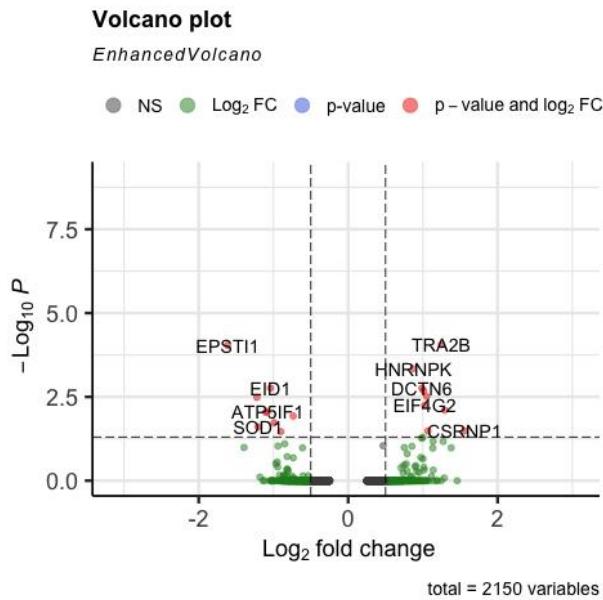


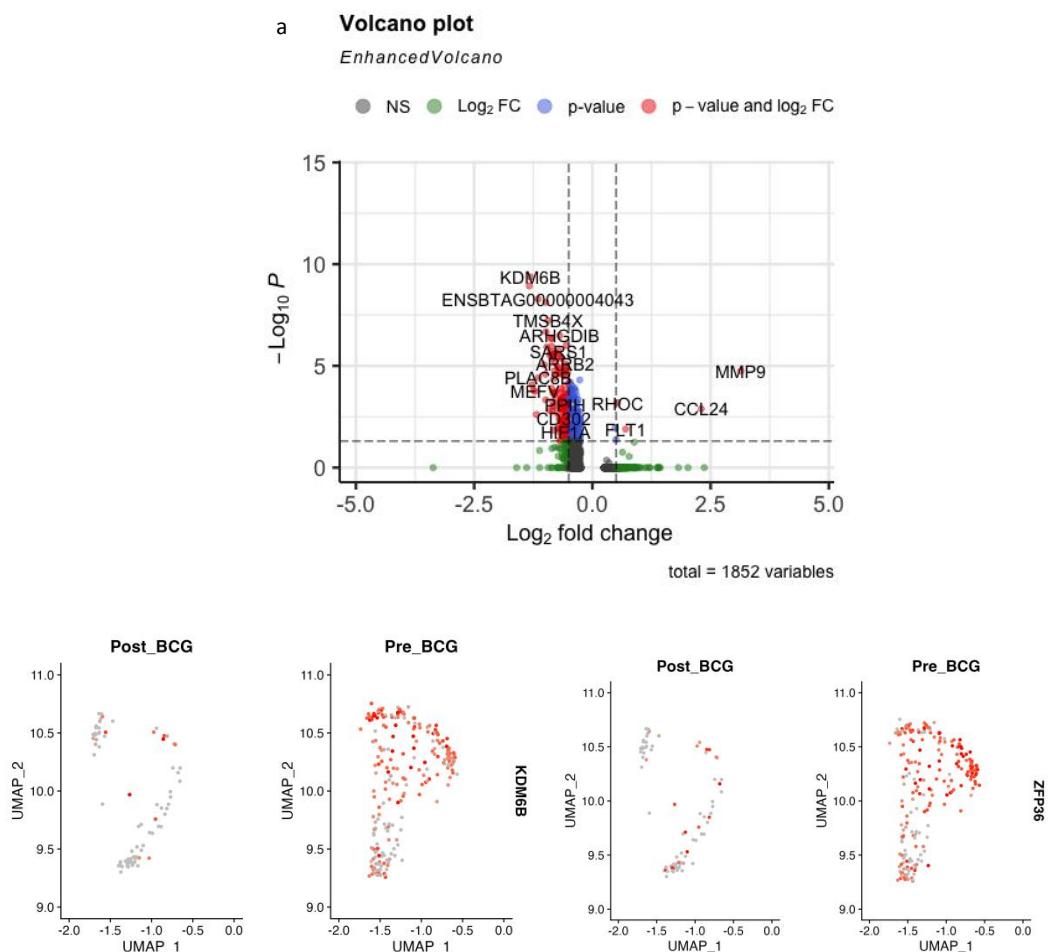
Figure 12 Volcano plot of differential gene expression in dendritic cells group 5 (cluster 18)

3.4.6 Identification of differentially expressed genes in macrophages (cluster 17)

The differential expression gene of macrophages (cluster 17) before and after BCG injection is shown in Table 6. 426 genes were differentially expressed in total. The volcano plot in Figure 13 (a) revealed that nearly all of the differently expressed genes were down-regulated, and some of the differentially expressed genes were displayed in Figures 13 (b) and (c) by the '*VlnPlot*' and '*FeaturePlot*' functions. Differentially down-regulated genes included KDM6B, ZFP36, ENSBTAG00000004043, BOLA-DMA, RAMP2, and ARHGDIb, as illustrated in Figures 13 (b) and (c).

	p_val	avg_log2FC	pct. 1	pct. 2	p_val_adj
KDM6B	2.64E-14	-1.307439214	0.215	0.741	4.27E-10
ZFP36	7.29E-14	-1.332915907	0.266	0.807	1.18E-09
ENSBTAG00000004043	3.17E-13	-1.149241164	0.304	0.807	5.11E-09
BOLA-DMA	4.67E-13	-0.987095075	0.203	0.724	7.54E-09
TMSB4X	3.71E-12	-0.938703079	0.937	0.987	5.99E-08
RAMP2	1.32E-11	-0.977880767	0.165	0.645	2.14E-07
ARHGDIB	1.99E-11	-0.690834812	0.253	0.776	3.22E-07
MEF2A	2.60E-11	-0.875497087	0.228	0.675	4.20E-07
CNP	2.93E-11	-0.860812618	0.038	0.469	4.73E-07
TET2	6.02E-11	-0.551176549	0.051	0.496	9.73E-07

Table 6. Top 10 list of genes differentially expressed in macrophages (cluster 17)



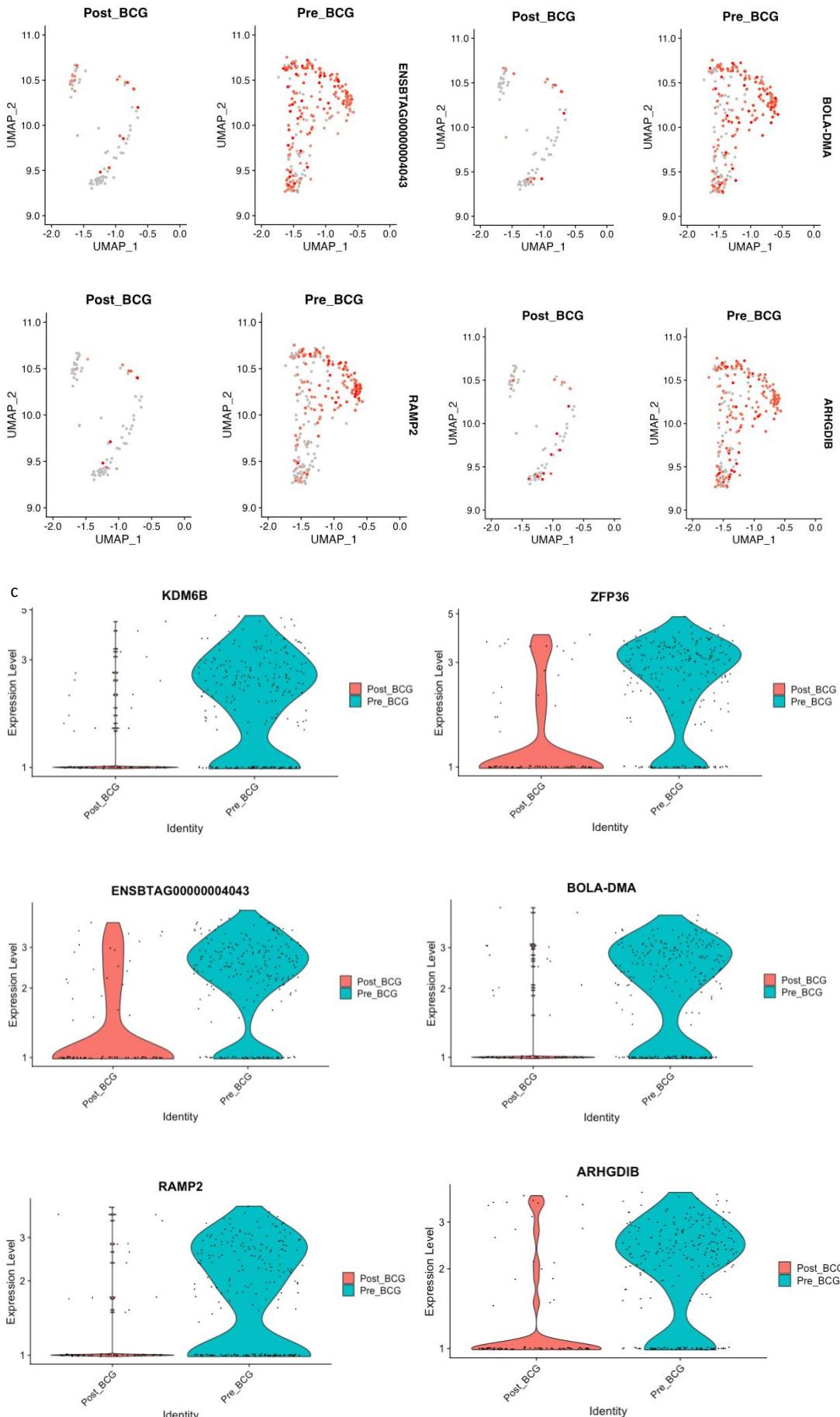


Figure 13 (a) Volcano plot of differential gene expression in macrophages group (cluster 17) (b) FeaturePlot of differential genes in macrophages group (cluster 17) (c) VinPlot of differential genes in macrophages group (cluster 17)

3.5 Defining cell types of dendritic cell groups

We used the FindAllMarkers package to find marker genes for each dendritic cell group, and the result is shown in Figure 14. Table 7 list the top 20 marker gene of each dendritic cell group. Hematopoietic stem cells (HSCs), which start as a multi-step process in the bone marrow, give rise to dendritic cells. The majority of dendritic cells were generated by a common DC progenitor (CDP) that differentiates into conventional DC precursor cells (pre-cDC) and plasmacytoid DC precursor cells (pre-pDC) (60). Both conventional dendritic cells were produced from pre-cDCs in peripheral lymphoid tissue, although E2-2-dependent pDC development was still present in the bone marrow (60). Basic leucine zipper ATF-like transcription factor 3 (BATF3) was necessary for the formation of cDC1, and many transcription factors, including interferon regulatory factor 4 (IRF4), controlled the development of cDC2 (60). Under certain conditions, monocytes could differentiate into macrophages to complement classical dendritic cells. In Table 7, we found gene BATF3 in dendritic cells group 4 and gene IRF4 in dendritic cells group 1, so dendritic cells group 4 could be defined as cDC1 and dendritic cells group 1 as cDC2. In addition, Ly6D had been identified as the key marker of pDC differentiation (61). We found Ly6D in dendritic cells group 5 and presumed it to be pDC precursors.

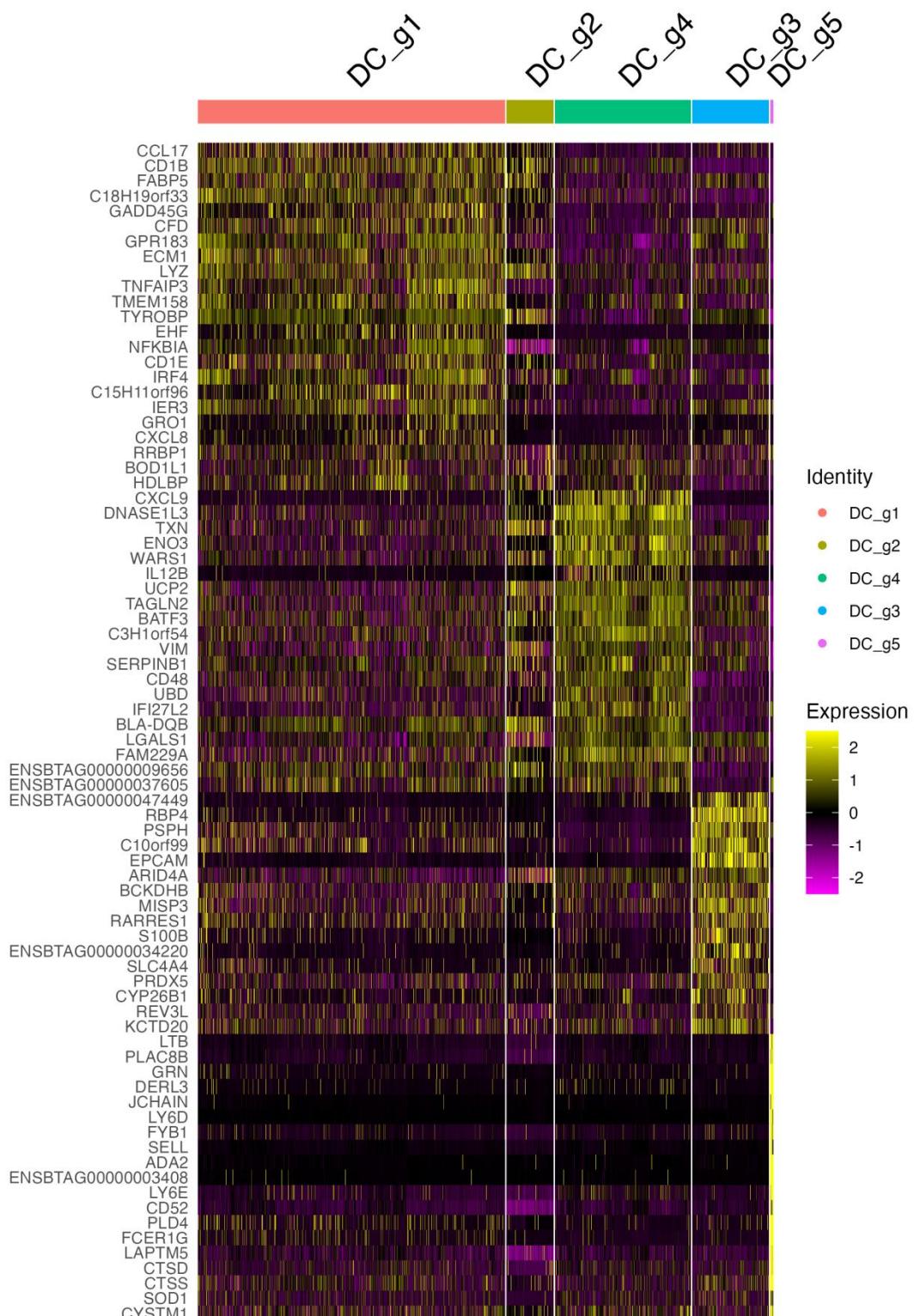


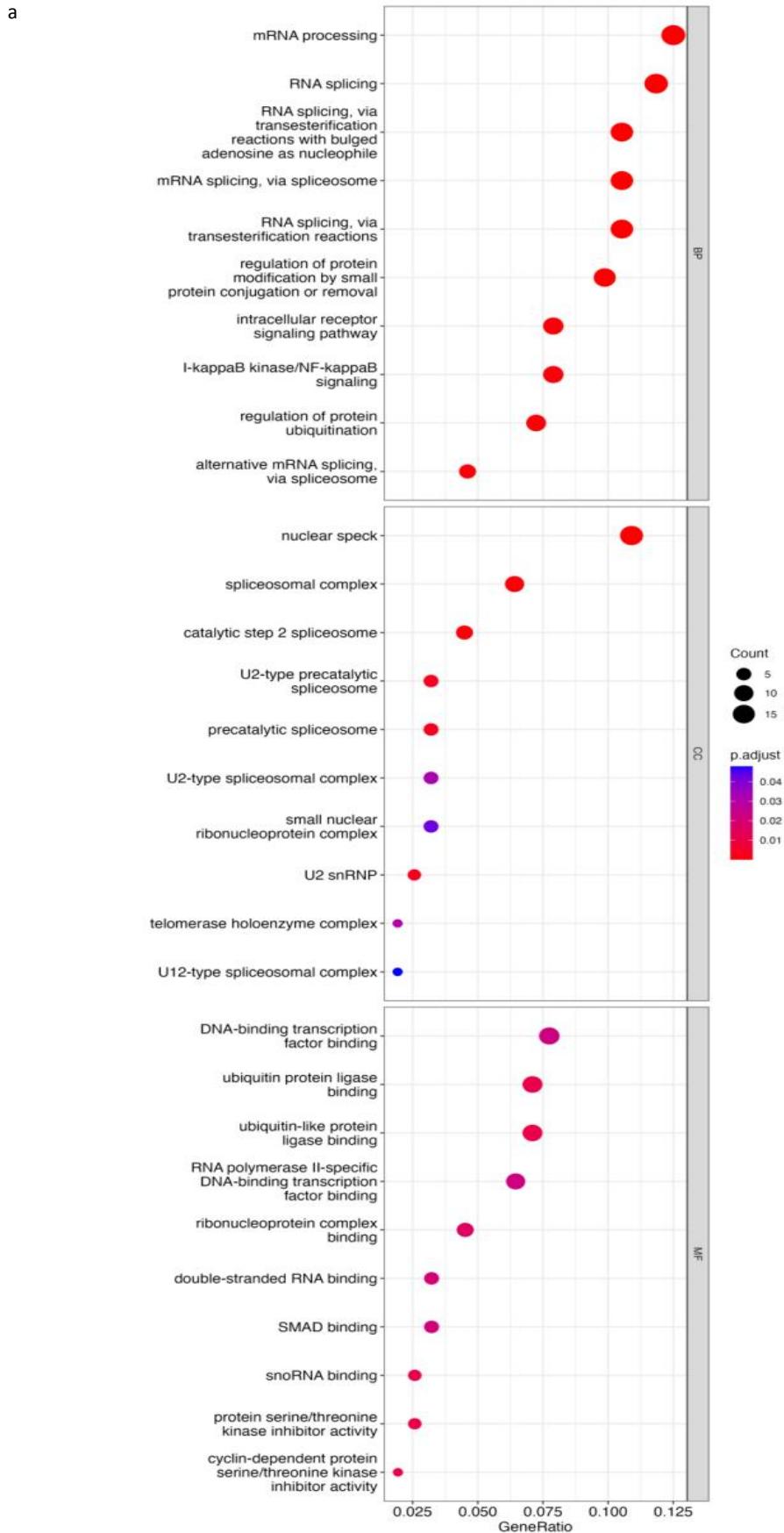
Figure 14 Heatmap of marker genes for each dendritic cell group

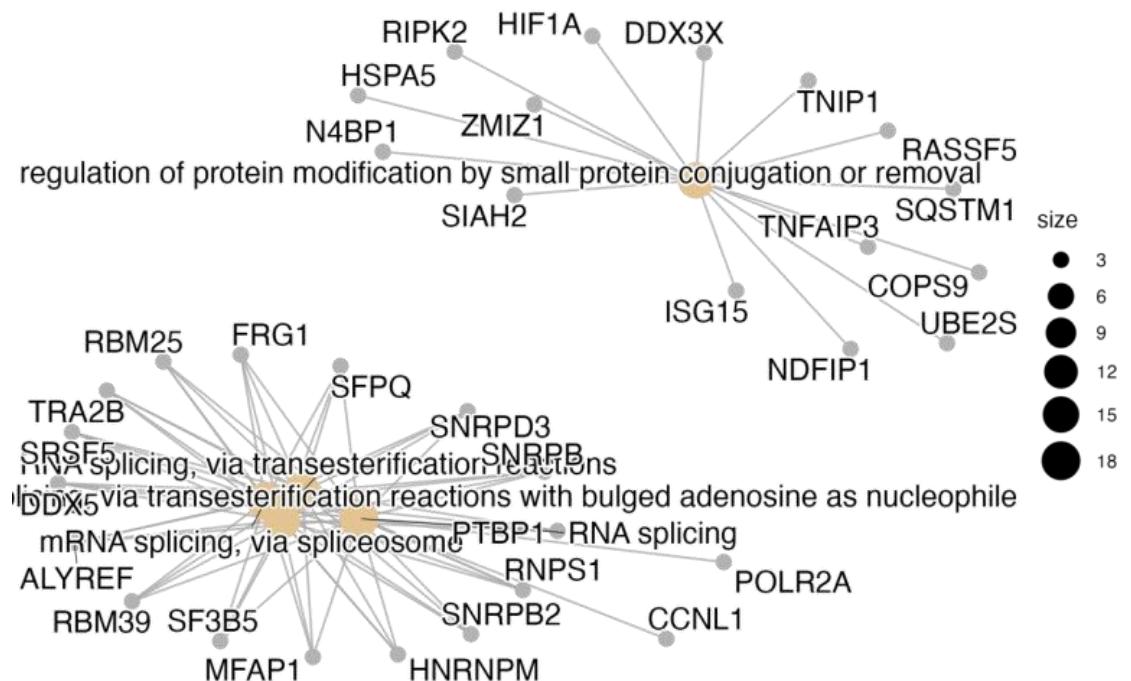
3.6 Functional enrichment analysis

GO enrichment analysis divided the annotated genes into cellular composition, biological process and molecular function, and cell localization. Here, we did enrichment analysis in three dendritic cell groups, 1, 3 and 4, because the other two dendritic cell groups were unsuitable for continued analysis.

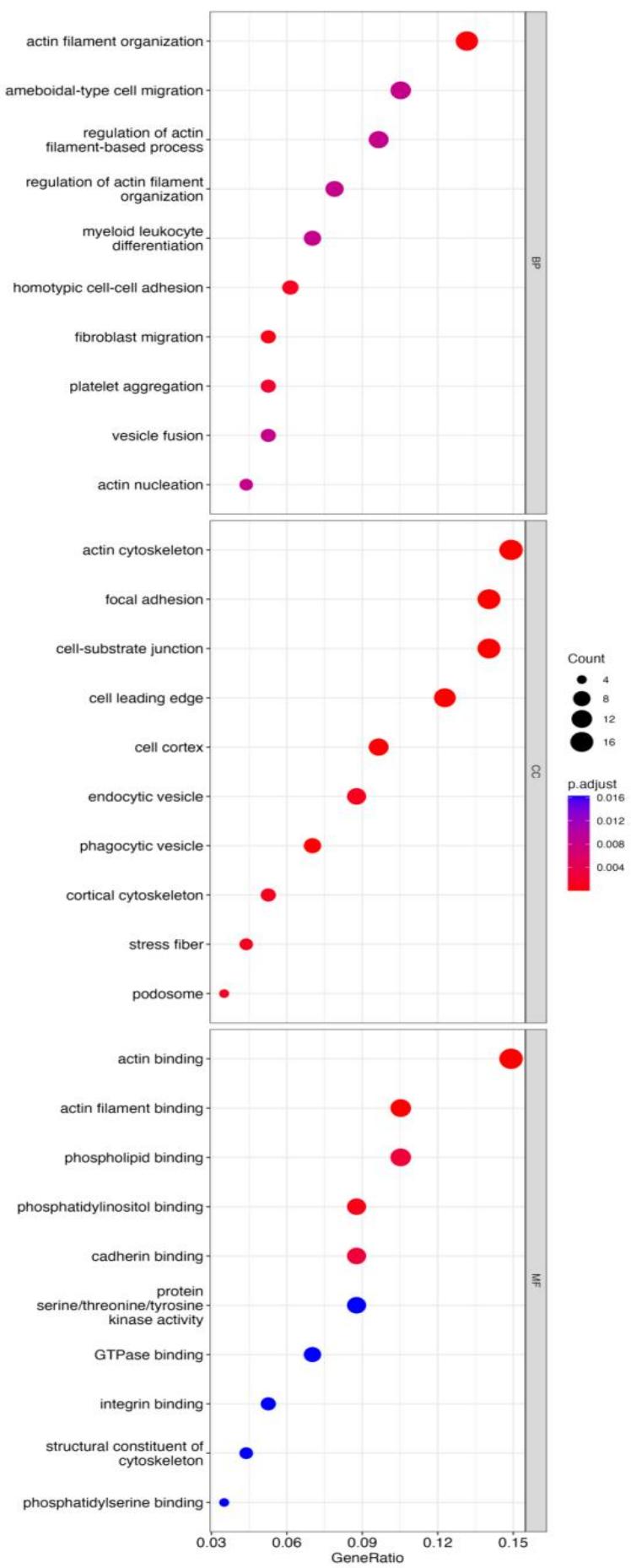
3.6.1 Enrichment analysis in dendritic cells group 1 (cluster 1, 3, 4, 6 and 13)

For dendritic cells in group 1, GO enrichment analysis results obtained 160 up-regulated genes and 120 down-regulated genes, respectively. Among up-regulated genes, differentially expressed genes were primarily enriched in protein alterations through small protein conjugation or elimination, RNA splicing via transesterification using bulge adenosine as a nucleophile, splicing of mRNA by the spliceosome, and RNA by transesterification splicing, mRNA processing and other biological processes. These genes were mainly enriched in cellular components such as nuclear speckle, spliceosome complex, second-stage catalytic spliceosome, U2 snRNP and spliceosomes of the U2-type precatalytic (Figure 15 (a)). Among down-regulated genes, differentially expressed genes were primarily enriched in actin filament organization, fibroblast migration, homotypic cell-cell adhesion, platelet aggregation, actin nucleation and other biological processes. These genes were primarily enriched in cellular components such as focal adhesion, cell-substrate intersection, the cytoskeleton of actin, cellular leading edge and phagocytic vesicle. These genes were primarily enriched in molecular functions such as actin binding, actin filament binding, phosphatidylinositol binding and cadherin binding(Figure 15 (b)).





b



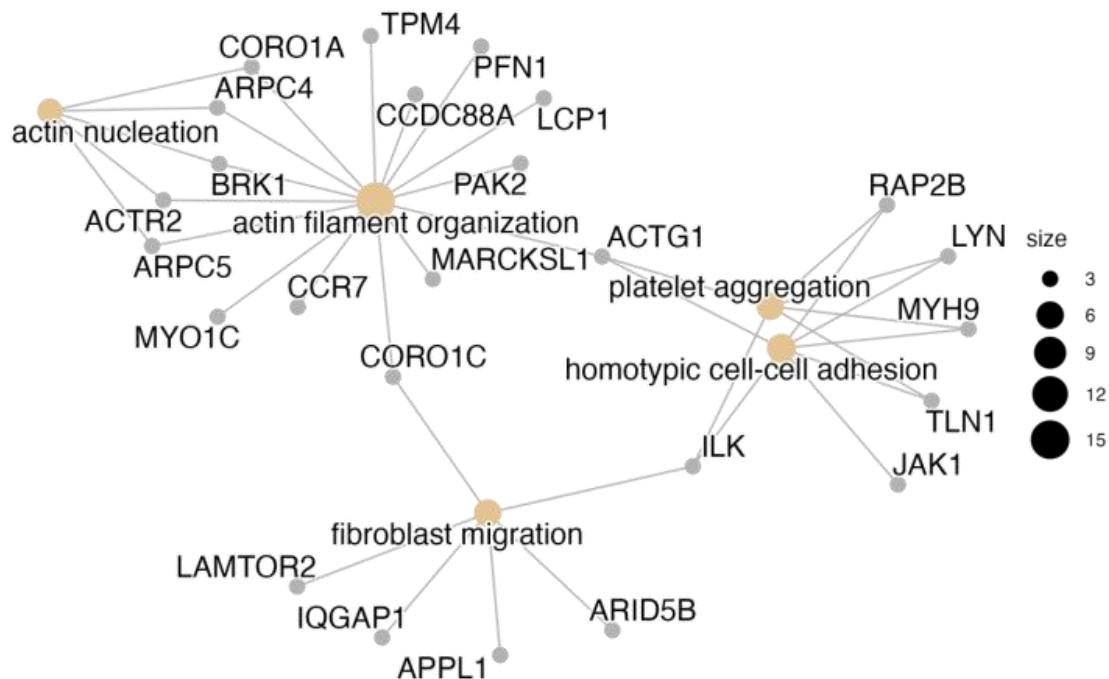


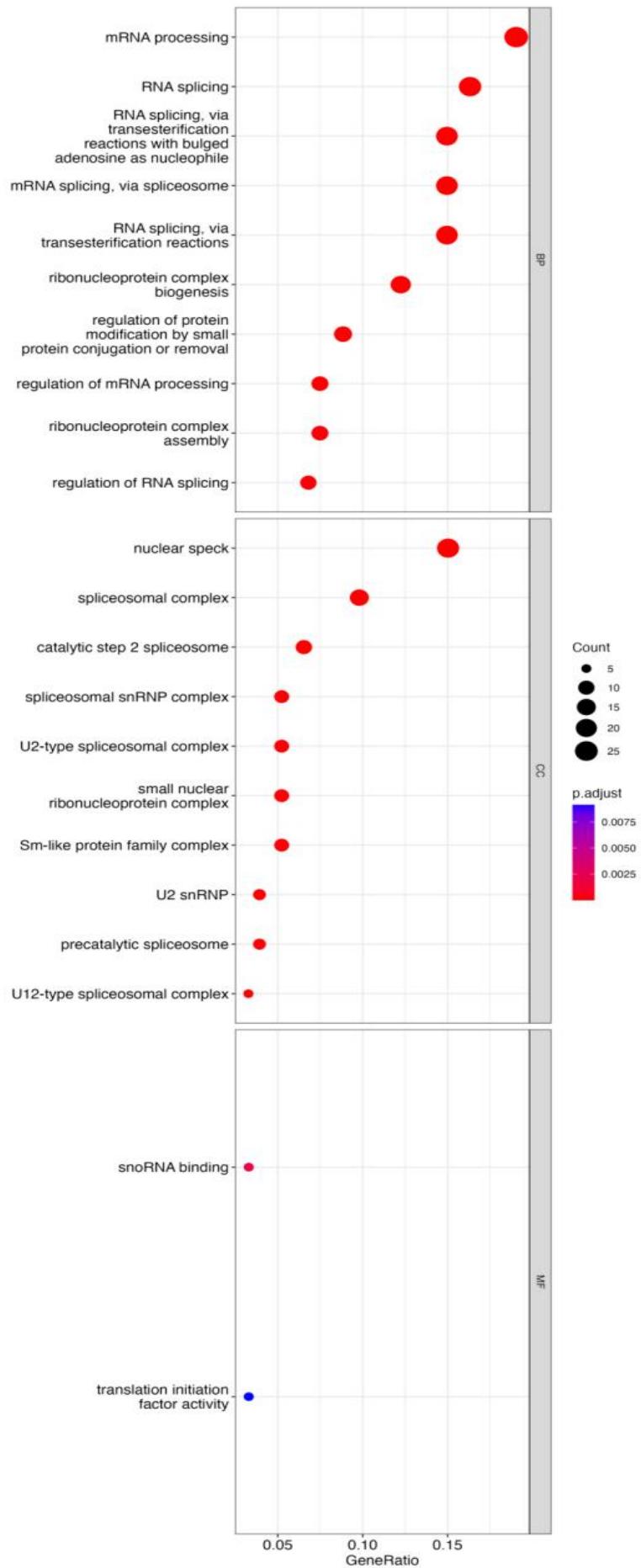
Figure 15 (a) GO enrichment analysis result of up-regulated genes in dendritic cells group 1 (cluster 1, 3, 4, 6 and 13) (b) GO enrichment analysis result of down-regulated genes in dendritic cells group 1 (cluster 1, 3, 4, 6 and 13)

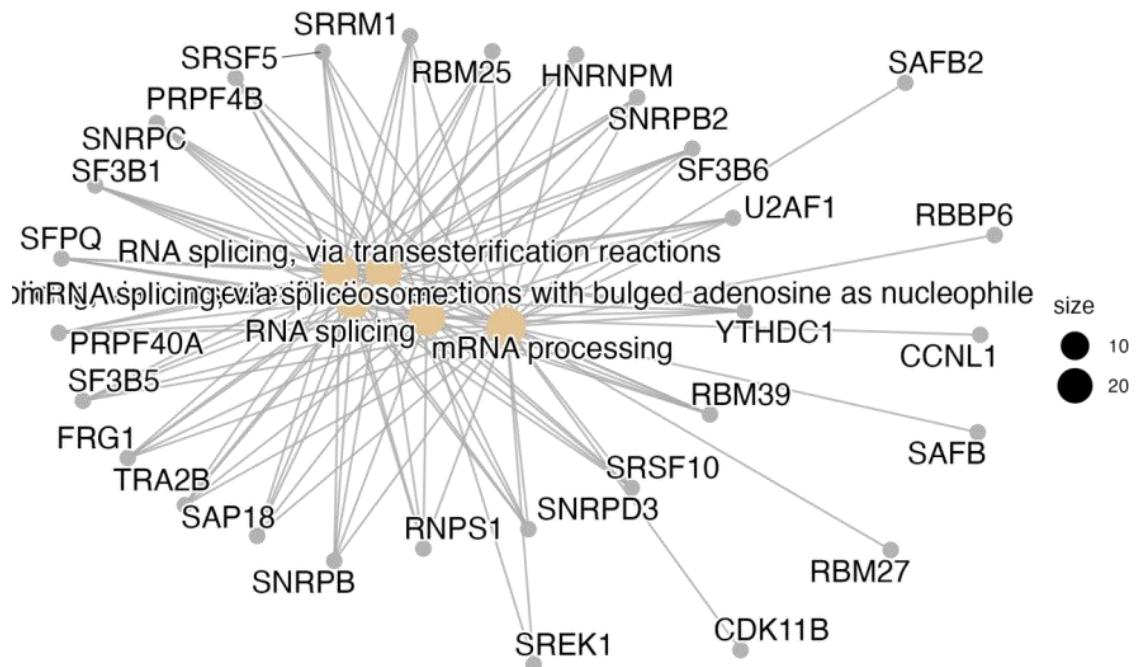
3.6.2 Enrichment analysis in dendritic cells group 3 (cluster 2)

For dendritic cells in group 3, GO enrichment analysis results obtained 156 up-regulated and 83 down-regulated genes, respectively. Among up-regulated genes, some cellular components related to them were atomic speck, the complex of spliceosomes, second-stage catalytic spliceosome and spliceosomal snRNP complex. The main biological processes involved were mRNA processing, transesterification processes for RNA splicing using bulged adenine as the nucleophile, spliceosome-mediated mRNA splicing and regulation of mRNA processing. The main molecular function involved were snoRNA binding and translation initiation factor activity (Figure 16 (a)). Among down-regulated genes, differentially expressed genes were mainly enriched in actin filament organization, control of supramolecular fibre structure, organisational control of actin filaments, homotypic cell-cell adhesion, control of actin filament length and other biological processes. Some cellular components related to them were focal adhesion, cell-substrate intersection,

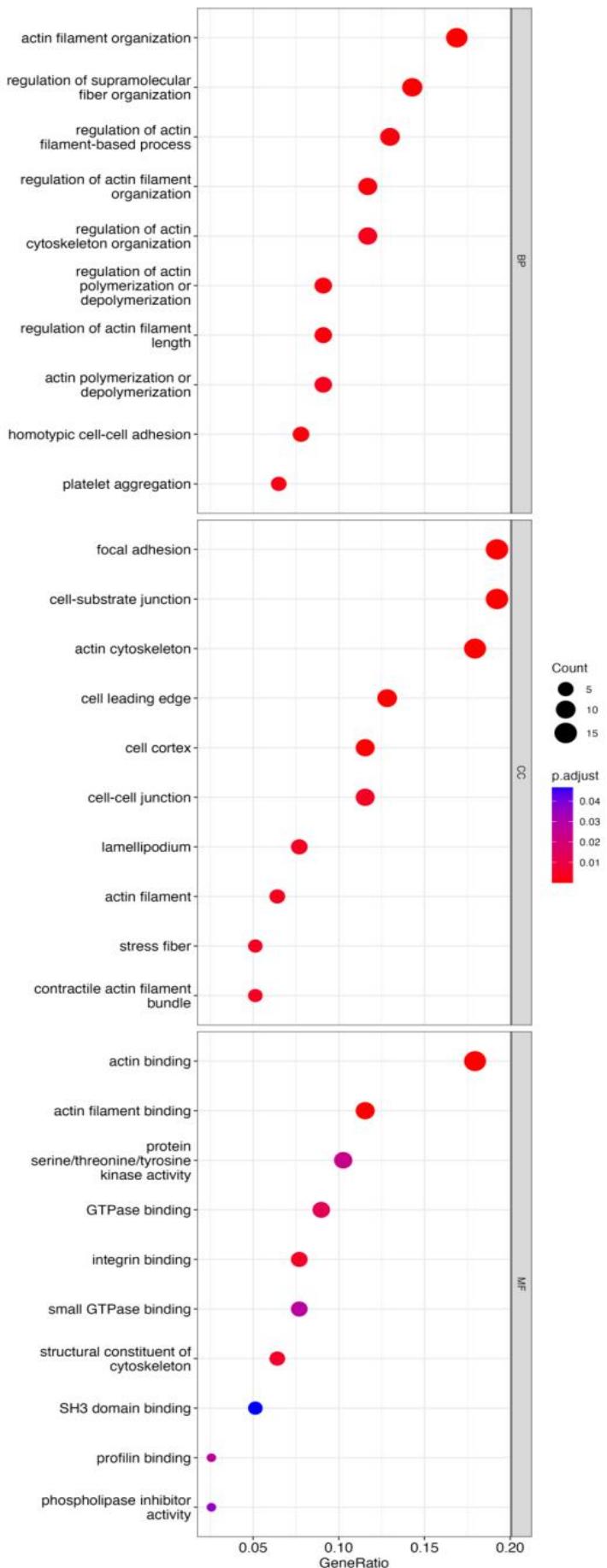
the cytoskeleton of actin, cell cortex and the leading edge of a cell. The principal molecular activities involved were binding of actin, actin filaments, integrin binding and structural constituent of the cytoskeleton (Figure 16 (b)).

a





b



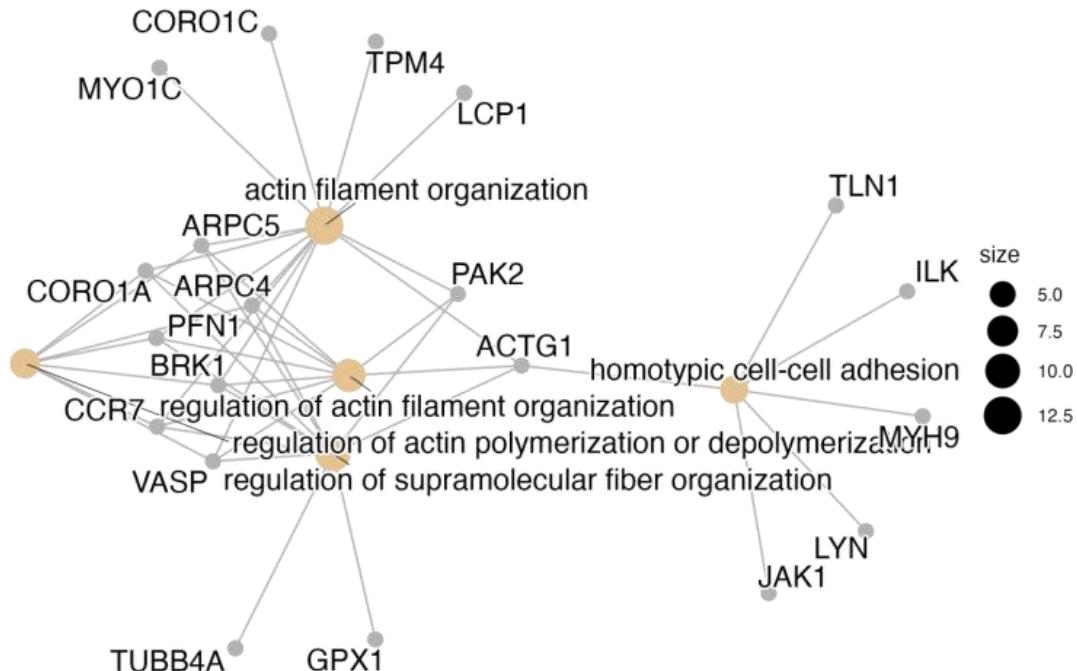


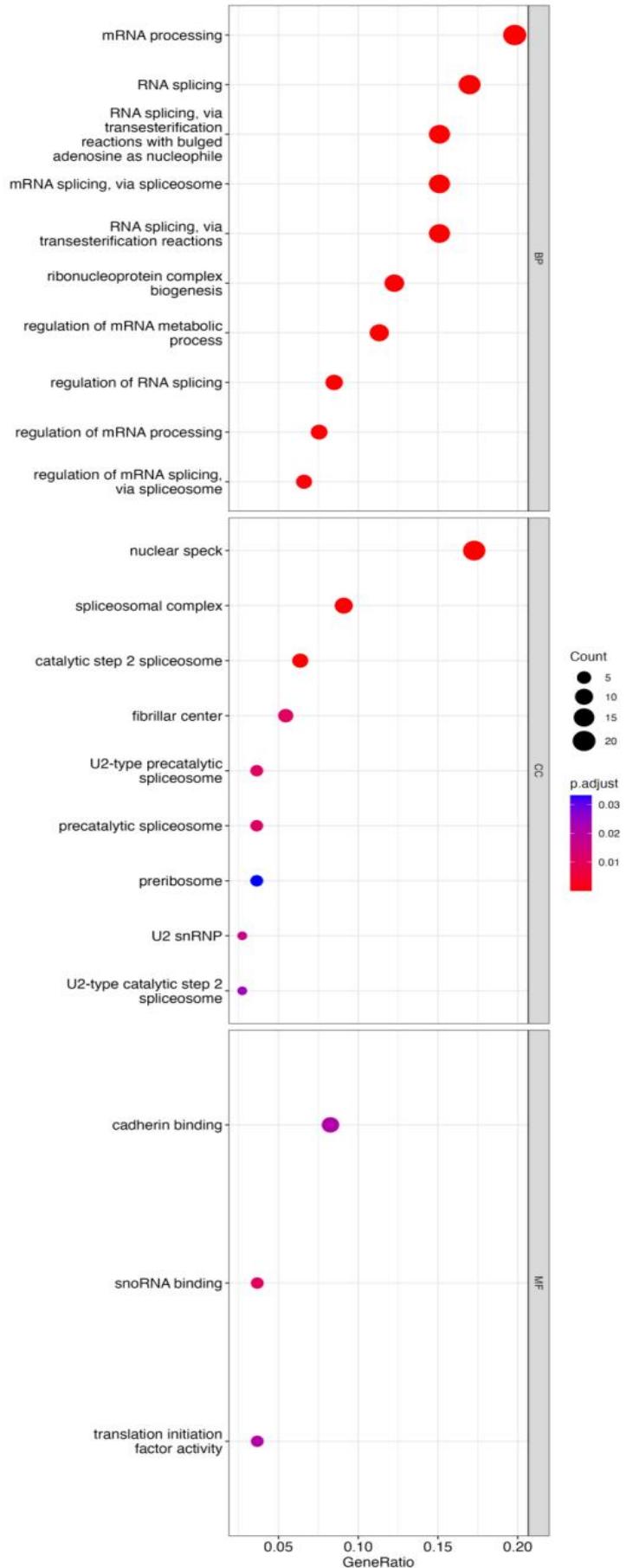
Figure 16 (a) GO enrichment analysis result of up-regulated genes in dendritic cells group 3 (cluster 2) (b) GO enrichment analysis result of down-regulated genes in dendritic cells group 3 (cluster 2)

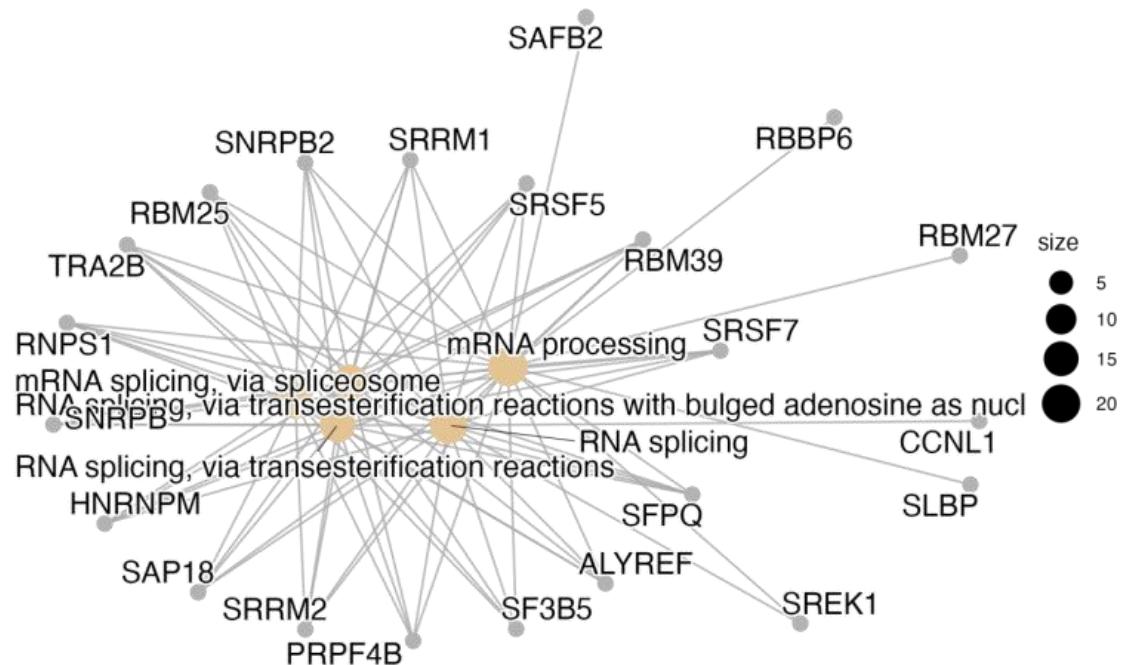
3.6.3 Functional enrichment analysis in dendritic cells group 4 (cluster 0, 10 and 11)

For dendritic cells n group 4, GO enrichment analysis results obtained 112 up-regulated genes and 49 down-regulated genes, respectively. Among up-regulated genes, differentially expressed genes were largely enriched in mRNA processing, transesterification processes for RNA splicing using bulged adenosine as the nucleophile, spliceosome for mRNA splicing, transesterification processes for RNA splicing, control of the mRNA metabolism and other biological processes. These genes primarily appeared in cellular components such as nuclear speckle, spliceosome complex and catalytic step 2 spliceosome (Figure 17 (a)). Among down-regulated genes, differentially expressed genes are mainly enriched in actin filament organization, organisational control of actin filaments, positive control of interleukin-12 synthesis, lymphocyte migration, regulation of actin cytoskeleton organization, T cell mediated cytotoxicity and other biological processes. These genes primarily appeared in cellular components such as focal

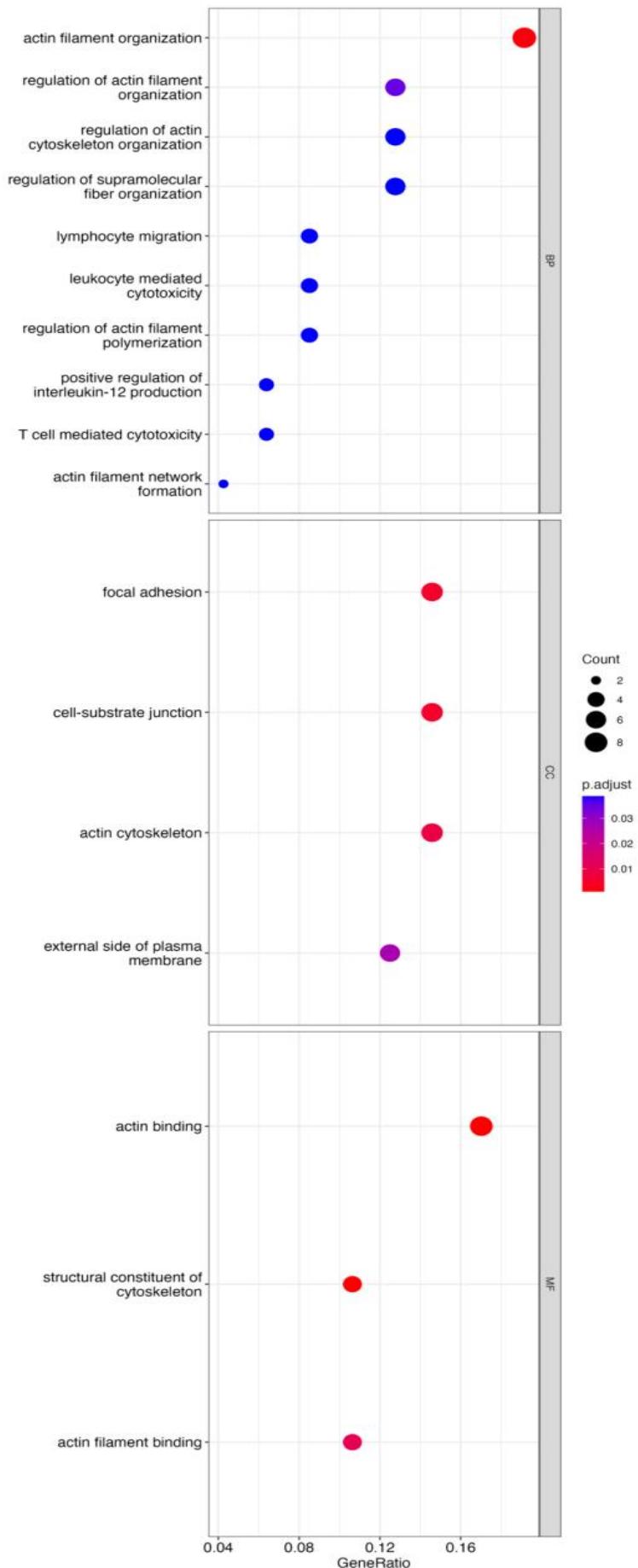
adhesion, cell-substrate junction and actin cytoskeleton (Figure 17 (b)).

a





b



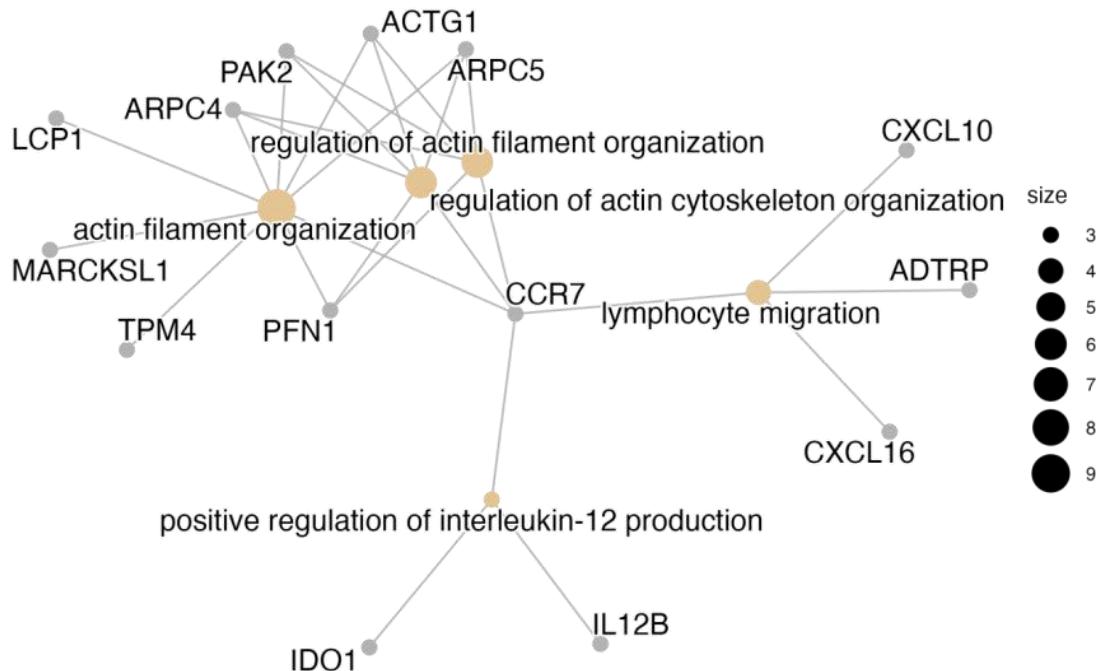
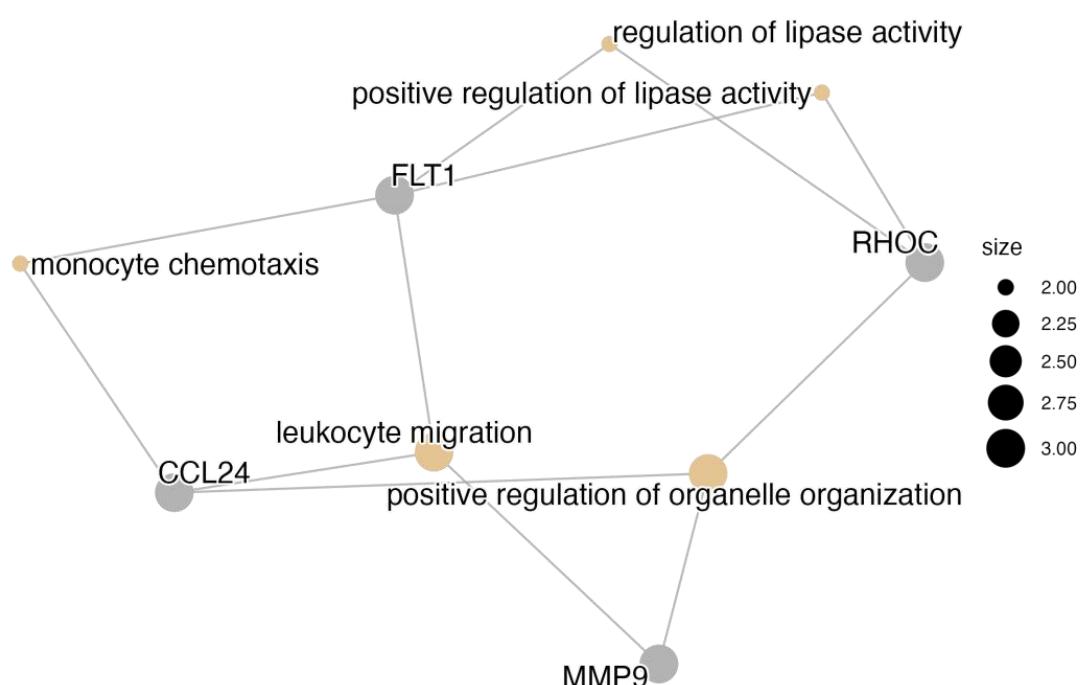
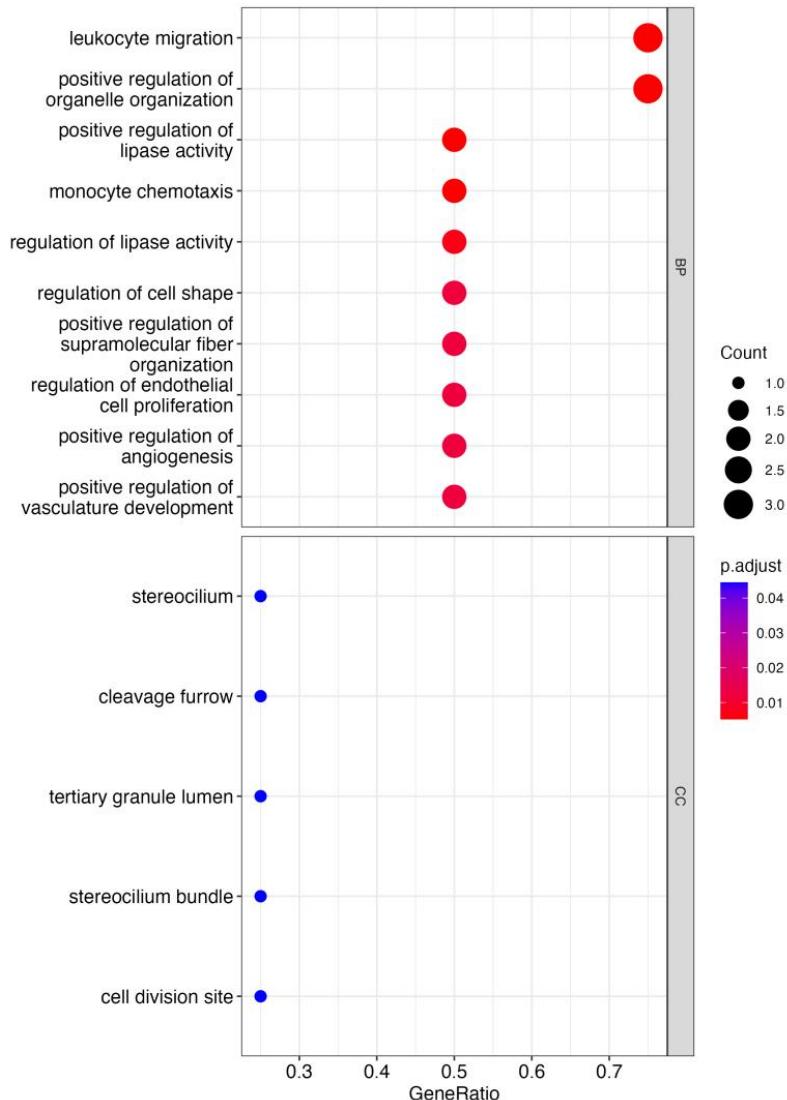


Figure 17 (a) GO enrichment analysis result of up-regulated genes in dendritic cells group 4 (cluster 0,10 and 11) (b) GO enrichment analysis result of down-regulated genes in dendritic cells group 4 (cluster 0, 10 and 11)

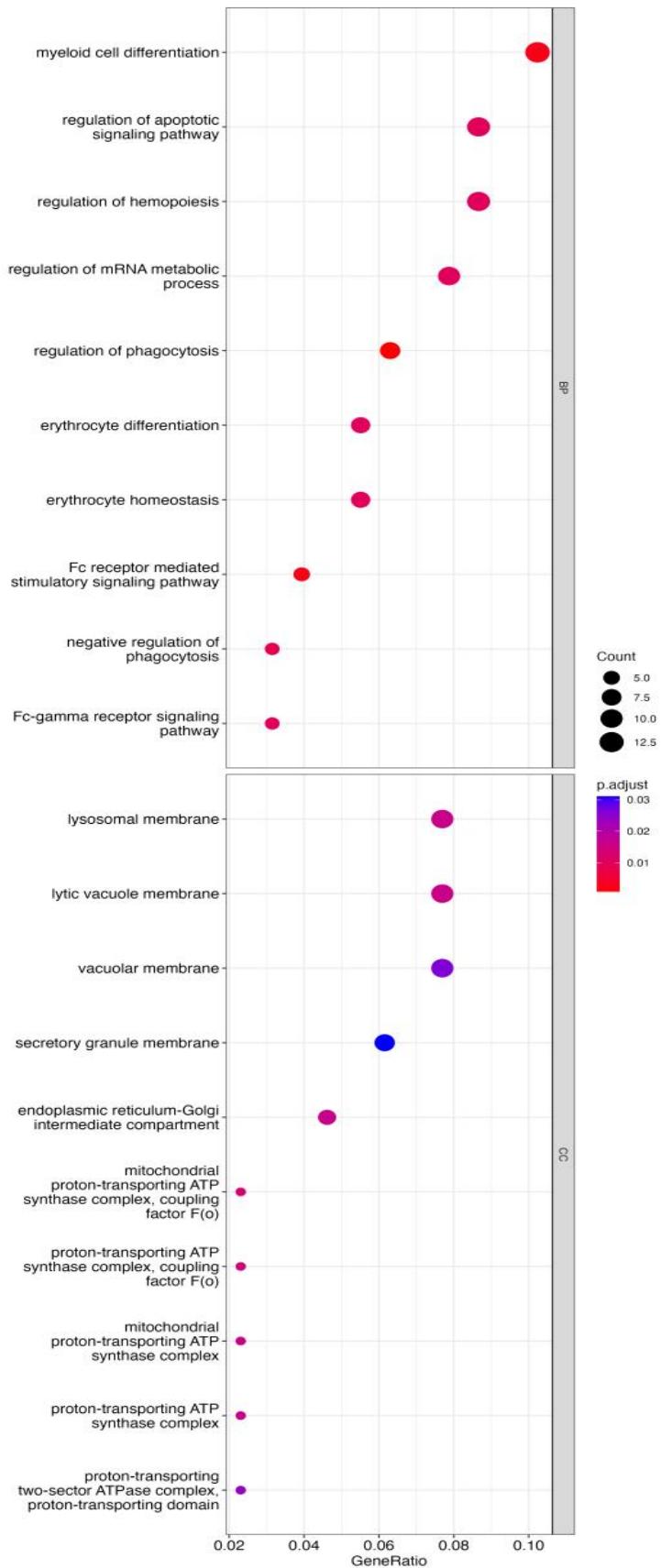
3.6.4 Functional enrichment analysis in macrophages (cluster 17)

For macrophages, GO enrichment analysis results obtained 4 up-regulated genes and 142 down-regulated genes, respectively. Among up-regulated genes, differentially expressed genes were primarily enriched in leukocyte migration and positive regulation of organelle organization (Figure 18 (a)). Down-regulated genes were mainly enriched in phagocytosis, myeloid cell morphology, Fc receptor-mediated mechanism for stimulating signalling, negatively controlling phagocytosis, erythrocyte differentiation, control of the mRNA metabolism and other biological processes (Figure 18 (b)).

a



b



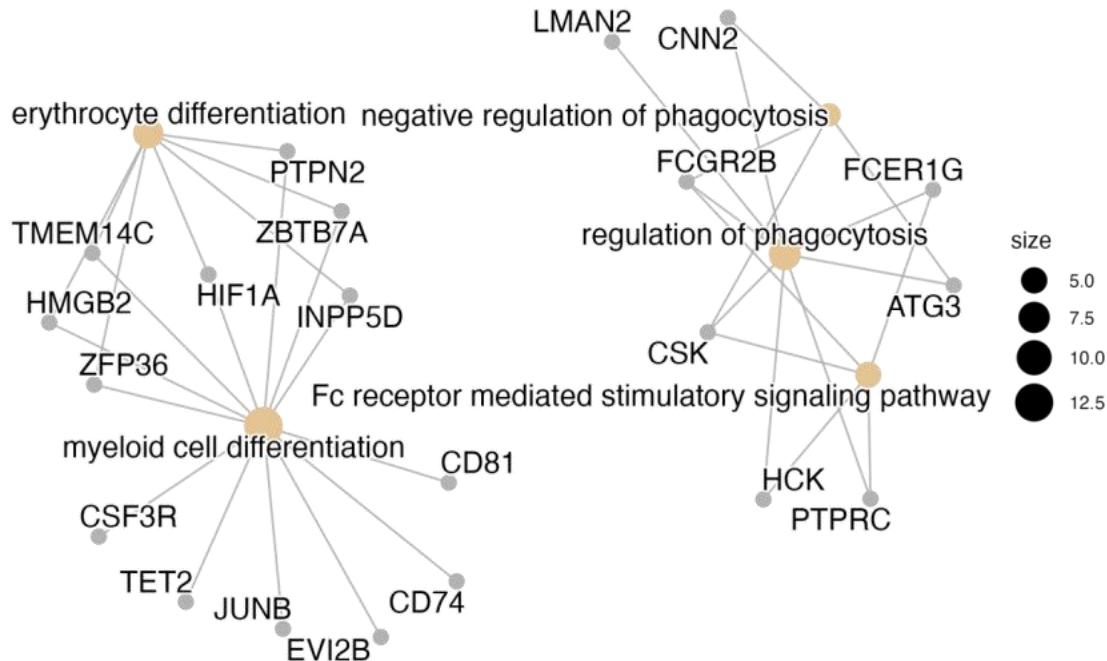


Figure 18 (a) GO enrichment analysis result of up-regulated genes in macrophages group (cluster 17) (b) GO enrichment analysis result of down-regulated genes in macrophages group (cluster 17)

Chapter 4 Discussion

Because BCG does not provide long-term protection in humans or cattle, researchers want to study the immune response to BCG against *Mycobacterium tuberculosis* or *Mycobacterium bovis* to improve or develop new vaccines. When *Mycobacterium tuberculosis* or *Mycobacterium bovis* infect the host, it first binds to alveolar macrophages, and these cells act as cellular hosts for the bacteria. In vaccination-induced responses, both dendritic cells and macrophages play key roles. Therefore, this study focused on how dendritic cells and macrophages are affected by BCG vaccination. This experiment used the bovine afferent lymphatic vessel model to collect large amounts of afferent lymph and provided the opportunity to study lymphocytes transported from the periphery directly ex vivo (37). This model improves the understanding of dendritic cells and aids in BCG vaccination studies. In addition, Single-cell RNA-sequencing techniques can explore cell-to-cell heterogeneity and reconstruct lineage hierarchies at the single-cell level. This method enables unbiased analysis of heterogeneity distribution within cell

groups because it exploits transcriptome reorganization in individual cells. In this experiment, we compared the four groups of samples before and after BCG injection by analyzing single cell transcriptomic data. A total of 30,573 cells were included in our study, of which 21,880 were featured. We used the Seurat package for quality control to remove low-quality and empty cells. Doublets were then removed using the doubletFinder package, and batch effects were removed by data integration. The PCA, tSNE, and UMAP functions in the Seurat package were used for dimensionality reduction, and SingleR was used to provide annotations for each cell type. Differential expression and GO enrichment analysis were used to identify differentially expressed genes and their corresponding cellular pathways. We identify some genes associated with BCG immune effects in dendritic cells and macrophages, look for reliable biological markers and develop appropriate targeted therapies.

4.1 Analysis of dendritic cell results

The differential analysis results were performed based on the results of cell groupings to find significantly different genes for each cell group. According to the results, significantly different genes can be found in dendritic cell groups 1, 3 and 4.

In dendritic cell group 1, deduced to be cDC2, we found 1021 differentially expressed genes such as SAMS1, SIAH2, RNF19B, RAB9A, PAK2 and LY75. The dendritic cell group 1 had the most significant differential expression. Some genes with significant differences in expression were analyzed and discussed below. SAMS1 was the most significantly up-regulated gene in dendritic cell group 1 in this experiment, and this gene was also up-regulated in dendritic cell group 3. B cell activation is negatively regulated by SAMS1. In addition to promoting HDAC1 activity, it controls cell polarisation and

spreading (62). Studies have also pointed out that SAMSN1 acts as a route to activate B cells, one of the feature sets that are differentially up-regulated in childhood tuberculosis (63). PAK2 was a down-regulated gene, and it functions in several signalling pathways, including regulation of the cytoskeleton, cell cycle progression, cell motility, apoptosis or proliferation. When cells undergo caspase-mediated apoptosis, the protein that this gene produces is activated by proteolytic cleavage, and it may control the apoptotic processes that occur during cell death (64). HIV-1 Nef is acknowledged to down-regulate CD1a lipid antigen presentation in dendritic cells through the actions of Hck and PAK2 (65). SIAH2 can help target proteins become ubiquitinated and then degraded by proteasomes. It is also involved in apoptosis, tumour suppression, cell cycle, transcription and signal transduction processes, but its role in BCG is unknown (66). GO enrichment annotation found that these up-regulated differentially expressed genes were mainly concentrated in biological processes such as mRNA splicing, mRNA processing and protein modification. These genes were mainly present in the spliceosome. These down-regulated differentially expressed genes were mainly concentrated in biological processes such as actin filament organization regulation, fibroblast migration, and homotypic cell-cell adhesion. These down-regulated differential genes appeared in some molecular functions such as actin filament binding, phosphatidylinositol binding and phospholipid binding.

In dendritic cell group 3, we found 1113 differentially expressed genes, such as SYNGR2, ACTG1, STK4, TNIP1, SAMSN1 and RBBP6. SYNGR2 was significantly down-regulated in dendritic cell populations 3 and 4, but the specific function of this gene is unclear and may play a role in regulating exocytosis. ACTG1 is part of the actin family. The actin cytoskeleton, made up of the proteins in this family, is a web of fibres that serves as the cell's structural framework (67). ACTG1 was down-regulated in dendritic cell group 3,

and the expression of ACTG1 increases the motility of macrophages and enhances their interaction with T cells, thereby enhancing the antigen presentation ability (68). TNIP1 is one of the main regulators of the NF- κ B signalling pathway and is associated with cellular inflammation (69). Related research has shown that more widely than BCG, rBCG-LTAK63-infected M2 macrophages up-regulate the transcripts of the inflammatory genes TAP1, GBP1, SLAMF7, TNIP1, and IL6, leading to increased levels of inflammation and tissue repair (70). From this, we know that TNIP1 is a cytokine involved in inflammation and tissue repair. In addition, GO enrichment results showed that up-regulated differentially expressed genes were mainly concentrated in biological processes such as mRNA splicing and mRNA processing. These genes are also mainly present in the spliceosome. Down-regulated differentially expressed genes are also enriched in actin filament organization.

In dendritic cell group 4, deduced to be cDC1, we found 792 differentially expressed genes, such as RBBP6, RBM39, USP12, HNRNPDL, SRSF5 and SYNGR2. RBBP6 can restrict cell growth and inactivate, resulting in apoptosis. It can also function in the 3' processing of mRNA (71). RBM39 is an RNA-binding protein involved in alternative RNA splicing and transcriptional co-regulation. Loss of RBM39 impairs cell cycle progression and results in tumour regression in many preclinical models by causing aberrant splicing events and differential gene expression (72). In addition to being a component of the spliceosome, SRSF5 is required for mRNA splicing and has been linked to mRNA export and translation from the nucleus (73). Up-regulated differentially expressed genes were mainly concentrated in mRNA processing and RNA splicing and appeared in nuclear speck and spliceosomal complex. Down-regulated differentially expressed genes were enriched in actin filament organization and appeared in focal adhesion, cell-substrate junction and the actin cytoskeleton.

We found that the enrichment analysis result of the three dendritic cell groups was very similar. We found them involved in RNA splicing in all three groups of up-regulated genes. RNA splicing can regulate and enhance anti-tumour immunity. Some studies have pointed out that the RNA-binding protein RBM39 is involved in transcriptional co-regulation, and alternative RNA splicing can stimulate antigen-presenting cells to stimulate T cells (74). RBM39 was also a significantly up-regulated gene differentially expressed in dendritic cell group 3. We found that all were related to actin among the three groups of down-regulated genes in dendritic cells. In dendritic cell group 3, ACTG1 is part of the actin family. Actin forms the cytoskeleton, participates in cell movement, and transmits cell signals (75). Actin mediates inflammatory mediators leading to cellular dysfunction (75). At the site of inflammation, the products of pathogenic microorganisms can directly damage the body cells and mediate further damage to the body by pathogens through changes in cellular actin filaments (75). After BCG injection, the related pathways corresponding to actin were down-regulated.

In summary, in dendritic cells groups 1, 3 and 4, we found 2926 differential genes. Dendritic cells can mediate both innate and adaptive immune responses. Among these differentially expressed genes, we found not only innate immunity-related genes, such as TNIP1 involved in the inflammatory response and Actin ACTG1 mediates inflammatory mediators, but also adaptive immunity-related genes, such as SAMSN1 involved in activating B cells and RBM39 involved in RNA splicing activation of T cells.

4.2 Analysis of macrophage results

Differences in gene changes in macrophages were not as significant in dendritic cells. In macrophages, differential expression analysis revealed that there were 426 differential expression genes, like KDM6B, ZFP36,

ENSBTAG00000004043, BOLA-DMA, RAMP2 and ARHGDI β . Almost all of these differential expression genes were down-regulated, but these differences were insignificant. KDM6B is involved in macrophage differentiation in inflammatory responses by regulating gene expression and macrophage differentiation (76). ZFP36 plays a role in the anti-inflammatory response, inhibiting tumour necrosis factor- α production by stimulating ARE-mediated decay of TNF- α mRNA and interferon-induced mRNA of several other inflammatory AREs in macrophages (77). It also regulates dendritic cell maturation at the post-transcriptional level, thus limiting inflammatory responses as part of a negative feedback loop (77). GO enrichment annotation found that these down-regulated differentially expressed genes were mainly concentrated in biological processes such as regulation of phagocytosis, myeloid cell differentiation, Fc receptor-mediated stimulation signalling pathway, erythrocyte differentiation, and regulation of mRNA metabolic processes. Macrophages are involved in phagocytosis and promote the differentiation of bone marrow cells into T/B immune cells. One study found that Fc receptor-mediated stimulatory signalling pathways enhance immunity and suppress inflammatory stress (78). Up-regulated differentially expressed genes were mainly enriched in leukocyte migration and positive regulation of organelle organization. Leukocyte migration promotes a rapid innate immune response and is involved in the inflammatory response. One possible factor contributing to the insignificant macrophage difference, with few up-regulated cells, might be that most samples were dendritic cells containing few macrophages. The main samples containing macrophages were lymph node samples.

4.3 Limitations of the experiment

The way to perform the analysis was to treat each cell as an independent observation. However, it might be more accurate to use pseudobulk analysis to

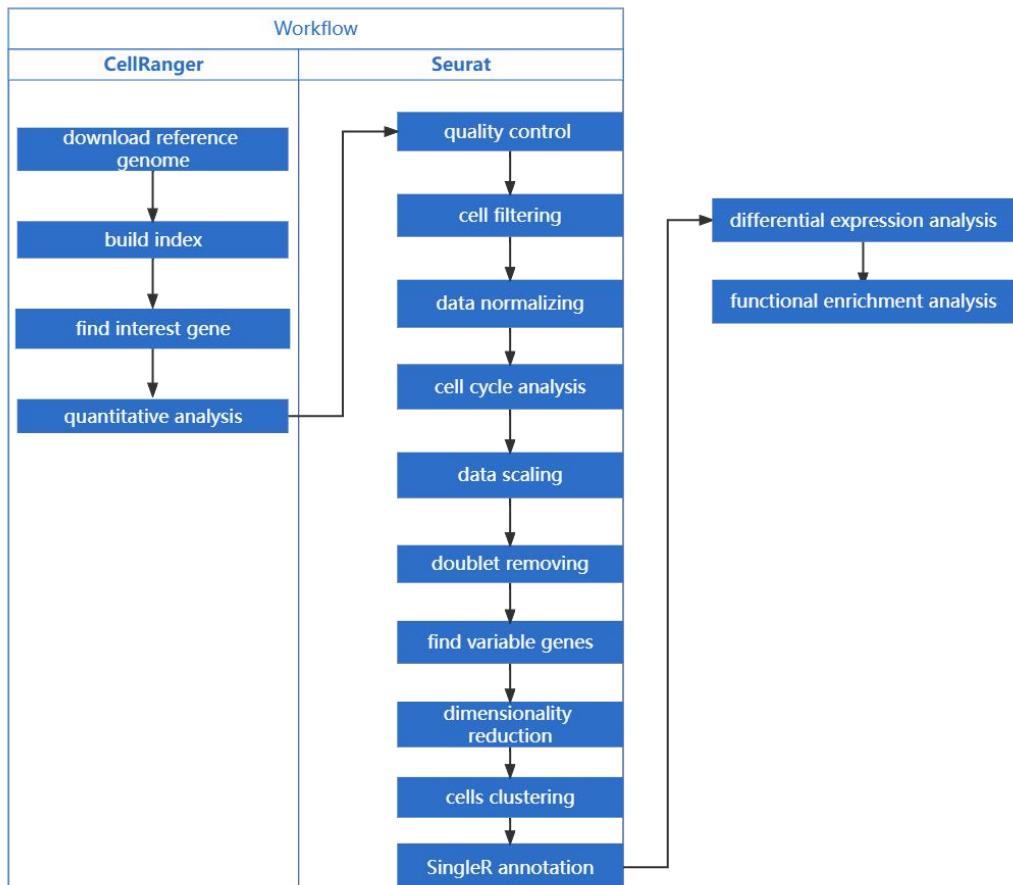
collapse the expression for all cells within each sample and each gene cluster. That way, stats could be carried out on the four samples. However, this was not possible for macrophages, as most macrophages were derived from lymph node samples. During the experiment, we also found that if we studied other cell types, there would be significantly different gene products, such as neutrophils, T cells and B cells. We speculated that the detection time after BCG injection affected the number of dendritic cells or macrophages and the expression of related genes in each sample since innate and adaptive immunity did not occur simultaneously. The first was innate immunity. At this time, the expression difference of innate immunity-related genes would be more significant, and then adaptive immunity-related genes would be expressed one after another. In addition to this, another limitation was that we did not know if any differential genes were up/down-regulated in all dendritic cell groups, as experimental results can only show if these genes are up/down-regulated in specific dendritic cell groups. A potential future direction would be to compare the list of important genes pre/post BCG in all dendritic cell groups to see commonality/differentiation.

5. Conclusion

In the present study, we used the Seurat package to analyze single cell transcriptomic data, mainly by differential expression analysis and enrichment analysis to compare gene expression differences in dendritic cells and macrophages before and after BCG injection. Our study identified many differentially expressed genes in dendritic cells and macrophages, and we identified cellular pathways associated with these differential genes. Gene differences in macrophages were not significant. Based on the experimental results, we found significant differential expression of some innate immune-related genes after BCG treatment in dendritic cells, such as up-regulation of TNIP1 to promote inflammation and tissue healing and

down-regulation of the actin-related gene ACTG1 to alleviate the damage caused by pathogens at the site of inflammation. Adaptive immunity-related genes were also significantly differentially expressed, such as up-regulation of SAMSN1 in the activation of B cells and up-regulation of RBM39 involved in RNA splicing to activate T cells. These adaptive immune-related genes stimulate T/B cell production via dendritic cells. We believe new vaccines could be designed to specifically switch on these gene expression pathways to enhance protection and help address the issue of long-term protection.

6. Workflow



7. Acknowledgements

I wish to express my sincere gratitude to Prof. Jayne Hope and Dr Barbara Shih for supervision. And I wish to thank my parents for funding me to study.

8. Bibliography and references

1. Maggioli, Mayara F. "A bloody evidence: Is *Mycobacterium bovis* bacteraemia frequent in cattle?!" *Virulence* 7.7 (2016): 748-750.
2. Rodrigues, Laura Cunha, Punam Mangtani, and Ibrahim Abubakar. "How does the level of BCG vaccine protection against tuberculosis fall over time?." *Bmj* 343 (2011).
3. Mondal, M. N. I., et al. "Socio-demographic factors affecting knowledge level of Tuberculosis patients in Rajshahi City, Bangladesh." *African health sciences* 14.4 (2014): 855-865.
4. Chu, H. Q., et al. "Chest imaging comparison between non-tuberculous and tuberculosis mycobacteria in sputum acid fast bacilli smear-positive patients." *Eur Rev Med Pharmacol Sci* 19.13 (2015): 2429-2439.
5. Lawn, Stephen D., Robin Wood, and Robert J. Wilkinson. "Changing concepts of "latent tuberculosis infection" in patients living with HIV infection." *Clinical and Developmental Immunology* 2011 (2011).
6. Black, Robert E., Saul S. Morris, and Jennifer Bryce. "Where and why are 10 million children dying every year?." *The lancet* 361.9376 (2003): 2226-2234.
7. Caminero, J. A. "Management of multidrug-resistant tuberculosis and patients in retreatment." *European Respiratory Journal* 25.5 (2005): 928-936.
8. Murdoch, Jenna R., and Clare M. Lloyd. "Chronic inflammation and asthma." *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* 690.1-2 (2010): 24-39.
9. Tufariello, JoAnn M., John Chan, and JoAnne L. Flynn. "Latent tuberculosis: mechanisms of host and bacillus that contribute to persistent infection." *The Lancet infectious diseases* 3.9 (2003): 578-590.
10. Williams, Geraint T., and W. Jones Williams. "Granulomatous inflammation--a review." *Journal of clinical pathology* 36.7 (1983): 723-733.
11. Chakravortty, Dipshikha, and Michael Hensel. "Inducible nitric oxide

- synthase and control of intracellular bacterial pathogens." *Microbes and Infection* 5.7 (2003): 621-627.
12. Baena, Andres, and Steven A. Porcelli. "Evasion and subversion of antigen presentation by *Mycobacterium tuberculosis*." *Tissue antigens* 74.3 (2009): 189-204.
13. Schleimer, Robert P., et al. "Epithelium: at the interface of innate and adaptive immune responses." *Journal of Allergy and Clinical Immunology* 120.6 (2007): 1279-1284.
14. Mihret, Adane. "The role of dendritic cells in *Mycobacterium tuberculosis* infection." *Virulence* 3.7 (2012): 654-659.
15. Henderson, Robert A., Simon C. Watkins, and J. L. Flynn. "Activation of human dendritic cells following infection with *Mycobacterium tuberculosis*." *The Journal of Immunology* 159.2 (1997): 635-643.
16. Banchereau, Jacques, and A. Karolina Palucka. "Dendritic cells as therapeutic vaccines against cancer." *Nature Reviews Immunology* 5.4 (2005): 296-306.
17. Mihret, Adane. "The role of dendritic cells in *Mycobacterium tuberculosis* infection." *Virulence* 3.7 (2012): 654-659.
18. Guermonprez, Pierre, et al. "Antigen presentation and T cell stimulation by dendritic cells." *Annual review of immunology* 20.1 (2002): 621-667.
19. Giacomini, Elena, et al. "Infection of human macrophages and dendritic cells with *Mycobacterium tuberculosis* induces a differential cytokine gene expression that modulates T cell response." *The Journal of Immunology* 166.12 (2001): 7033-7041.
20. Giri, Pramod K., and Jeffrey S. Schorey. "Exosomes derived from *Mycobacterium bovis* BCG infected macrophages activate antigen-specific CD4+ and CD8+ T cells in vitro and in vivo." *PLoS one* 3.6 (2008): e2461.
21. Grace, Patricia S., and Joel D. Ernst. "Suboptimal antigen presentation contributes to virulence of *Mycobacterium tuberculosis* in vivo." *The Journal of Immunology* 196.1 (2016): 357-364.

22. Marino, Simeone, et al. "Dendritic cell trafficking and antigen presentation in the human immune response to *Mycobacterium tuberculosis*." *The Journal of Immunology* 173.1 (2004): 494-506.
23. Lange, Christoph, et al. "100 years of *Mycobacterium bovis* bacille Calmette-Guérin." *The Lancet Infectious Diseases* (2021).
24. Andersen, Peter, and T. Mark Doherty. "The success and failure of BCG—implications for a novel tuberculosis vaccine." *Nature Reviews Microbiology* 3.8 (2005): 656-662.
25. Behar, Samuel M. "Antigen-specific CD8+ T cells and protective immunity to tuberculosis." *The New Paradigm of Immunity to Tuberculosis* (2013): 141-163.
26. Behar, Samuel M. "Antigen-specific CD8+ T cells and protective immunity to tuberculosis." *The New Paradigm of Immunity to Tuberculosis* (2013): 141-163.
27. Ankley, Laurisa, Sean Thomas, and Andrew J. Olive. "Fighting persistence: how chronic infections with *Mycobacterium tuberculosis* evade T cell-mediated clearance and new strategies to defeat them." *Infection and immunity* 88.7 (2020): e00916-19.
28. Anzengruber, Florian, et al. "T-cell mediated anti-tumor immunity after photodynamic therapy: why does it not always work and how can we improve it?" *Photochemical & Photobiological Sciences* 14.8 (2015): 1492-1509.
29. Thom, M. L., et al. "Duration of immunity against *Mycobacterium bovis* following neonatal vaccination with bacillus Calmette-Guérin Danish: significant protection against infection at 12, but not 24, months." *Clinical and Vaccine Immunology* 19.8 (2012): 1254-1260.
30. Lindenstrøm, Thomas, et al. "Tuberculosis subunit vaccination provides long-term protective immunity characterized by multifunctional CD4 memory T cells." *The Journal of Immunology* 182.12 (2009): 8047-8055.
31. Kaufmann, Stefan HE, and Andrew J. McMichael. "Annulling a dangerous liaison: vaccination strategies against AIDS and tuberculosis." *Nature*

- medicine 11.4 (2005): S33-S44.
32. Titball, Richard W., and E. Diane Williamson. "Yersinia pestis (plague) vaccines." Expert opinion on biological therapy 4.6 (2004): 965-973.
 33. Pahari, Susanta, et al. "Bolstering immunity through pattern recognition receptors: a unique approach to control tuberculosis." Frontiers in Immunology 8 (2017): 906.
 34. 27.Wille-Reece, Ulrike, et al. "Immunization with HIV-1 Gag protein conjugated to a TLR7/8 agonist results in the generation of HIV-1 Gag-specific Th1 and CD8+ T cell responses." The Journal of Immunology 174.12 (2005): 7676-7683.
 35. Thom, M. L., et al. "Duration of immunity against *Mycobacterium bovis* following neonatal vaccination with bacillus Calmette-Guérin Danish: significant protection against infection at 12, but not 24, months." Clinical and Vaccine Immunology 19.8 (2012): 1254-1260.
 36. Parlane, Natalie A., et al. "Revaccination of cattle with bacille Calmette-Guerin two years after first vaccination when immunity has waned, boosted protection against challenge with *Mycobacterium bovis*." PLoS One 9.9 (2014): e106519.
 37. Marzo, Sara, et al. "Characterisation of dendritic cell frequency and phenotype in bovine afferent lymph reveals kinetic changes in costimulatory molecule expression." Veterinary Immunology and Immunopathology 243 (2022): 110363.
 38. Li, Wei Vivian, and Jingyi Jessica Li. "An accurate and robust imputation method scImpute for single-cell RNA-seq data." Nature communications 9.1 (2018): 1-9.
 39. Tian, Luyi, et al. "scPipe: A flexible R/Bioconductor preprocessing pipeline for single-cell RNA-sequencing data." PLoS computational biology 14.8 (2018): e1006361.
 40. Nguyen, Tho Huu, et al. "Single-Cell RNA Sequencing Analysis of the Drosophila Larval Ventral Cord." Current Protocols 1.2 (2021): e38.

41. Grandi, Francesco, et al. "popsicleR: An R Package for Pre-processing and Quality Control Analysis of Single Cell RNA-seq Data." *Journal of Molecular Biology* (2022): 167560.
42. Guantes, Raul, et al. "Global variability in gene expression and alternative splicing is modulated by mitochondrial content." *Genome research* 25.5 (2015): 633-644.
43. Yang, Chen. Rapid Drug Adaptation to MAPK Pathway Inhibition in Single Melanoma Cells. Diss. University of Colorado at Boulder, 2020.
44. Biddy, Brent A., et al. "Single-cell mapping of lineage and identity in direct reprogramming." *Nature* 564.7735 (2018): 219-224.
45. Hafemeister, Christoph, and Rahul Satija. "Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression." *Genome biology* 20.1 (2019): 1-15.
46. Brechtmann, Felix, et al. "OUTRIDER: a statistical method for detecting aberrantly expressed genes in RNA sequencing data." *The American Journal of Human Genetics* 103.6 (2018): 907-917.
47. Xi, Nan Miles, and Jingyi Jessica Li. "Benchmarking computational doublet-detection methods for single-cell RNA sequencing data." *Cell systems* 12.2 (2021): 176-194.
48. Lê Cao, Kim-Anh, Simon Boitard, and Philippe Besse. "Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems." *BMC bioinformatics* 12.1 (2011): 1-17.
49. Marco, Eugenio, et al. "Bifurcation analysis of single-cell gene expression data reveals epigenetic landscape." *Proceedings of the National Academy of Sciences* 111.52 (2014): E5643-E5650.
50. Hu, Hang, et al. "ScCAEs: deep clustering of single-cell RNA-seq via convolutional autoencoder embedding and soft K-means." *Briefings in Bioinformatics* 23.1 (2022): bbab321.
51. Kobak, Dmitry, and Philipp Berens. "The art of using t-SNE for single-cell transcriptomics." *Nature communications* 10.1 (2019): 1-14.

52. Becht, Etienne, et al. "Dimensionality reduction for visualizing single-cell data using UMAP." *Nature biotechnology* 37.1 (2019): 38-44.
53. Büttner, Maren, et al. "A test metric for assessing single-cell RNA-seq batch correction." *Nature methods* 16.1 (2019): 43-49.
54. Zand, Maryam, and Jianhua Ruan. "A completely parameter-free method for graph-based single cell RNA-seq clustering." *bioRxiv* (2021).
55. Qi, Guoying, Richard Lee, and Steven Hayward. "A comprehensive and non-redundant database of protein domain movements." *Bioinformatics* 21.12 (2005): 2832-2838.
56. Huang, Qianhui, et al. "Evaluation of cell type annotation R packages on single-cell RNA-seq data." *Genomics, proteomics & bioinformatics* 19.2 (2021): 267-281.
57. Breitling, Rainer, and Pawel Herzyk. "Rank-based methods as a non-parametric alternative of the T-statistic for the analysis of biological microarray data." *Journal of bioinformatics and computational biology* 3.05 (2005): 1171-1189.
58. Newton, Michael A., et al. "Random-set methods identify distinct aspects of the enrichment signal in gene-set analysis." *The Annals of Applied Statistics* 1.1 (2007): 85-106.
59. Merico, Daniele, et al. "Enrichment map: a network-based method for gene-set enrichment visualization and interpretation." *PloS one* 5.11 (2010): e13984.
60. Schlitzer, Andreas, et al. "Recent advances in understanding dendritic cell development, classification, and phenotype." *F1000Research* 7 (2018).
61. Dress, Regine J., et al. "Plasmacytoid dendritic cells develop from Ly6D+ lymphoid progenitors distinct from the myeloid lineage." *Nature immunology* 20.7 (2019): 852-864.
62. Jaufmann, Jennifer, et al. "The emerging and diverse roles of the SLy/SASH1-protein family in health and disease—Overview of three multifunctional proteins." *The FASEB Journal* 35.4 (2021): e21470.

63. Zhong, Ming, et al. "Identification of key genes involved in type 2 diabetic islet dysfunction: a bioinformatics study." *Bioscience reports* 39.5 (2019).
64. Stamenkovic, Ivan, and Qin Yu. "Merlin, a “magic” linker between the extracellular cues and intracellular signaling pathways that regulate cell motility, proliferation, and survival." *Current Protein and Peptide Science* 11.6 (2010): 471-484.
65. Otsuka, Yohei, et al. "Differentiation of Langerhans cells from monocytes and their specific function in inducing IL-22–specific Th cells." *The Journal of Immunology* 201.10 (2018): 3006-3016.
66. Kocaturk, Nur Mehpore, and Devrim Gozuacik. "Crosstalk between mammalian autophagy and the ubiquitin-proteasome system." *Frontiers in cell and developmental biology* (2018): 128.
67. Liu, Jiahui, et al. "Regulation of Actg1 and Gsta2 is possible mechanism by which capsaicin alleviates apoptosis in cell model of 6-OHDA-induced Parkinson's disease." *Bioscience Reports* 40.6 (2020).
68. Bunnell, Tina M., et al. "β-Actin specifically controls cell growth, migration, and the G-actin pool." *Molecular biology of the cell* 22.21 (2011): 4047-4058.
69. Cheng, Yujing, et al. "TNIP1 polymorphisms with the risk of hepatocellular carcinoma based on chronic hepatitis B infection in Chinese Han Population." *Biochemical genetics* 57.1 (2019): 117-128.
70. dos Santos, Carina C., et al. "Recombinant BCG-LTAK63 Vaccine Candidate for Tuberculosis Induces an Inflammatory Profile in Human Macrophages." *Vaccines* 10.6 (2022): 831.
71. Moela, Pontsho, and Lesetja Raymond Motadi. "RBBP6: a potential biomarker of apoptosis induction in human cervical cancer cell lines." *OncoTargets and therapy* 9 (2016): 4721.
72. Xu, Yuwei, Anke Nijhuis, and Hector C. Keun. "RNA-binding motif protein 39 (RBM39): An emerging cancer target." *British Journal of Pharmacology* 179.12 (2022): 2795-2812.
73. Wagner, Rebecca E., and Michaela Frye. "Noncanonical functions of the

- serine-arginine-rich splicing factor (SR) family of proteins in development and disease." *Bioessays* 43.4 (2021): 2000242.
74. Xu, Caipeng, et al. "RNA-binding protein 39: a promising therapeutic target for cancer." *Cell death discovery* 7.1 (2021): 1-9.
75. Chen, Grace Y., and Gabriel Nuñez. "Sterile inflammation: sensing and reacting to damage." *Nature Reviews Immunology* 10.12 (2010): 826-837.
76. Essandoh, Kobina, et al. "MiRNA-mediated macrophage polarization and its potential role in the regulation of inflammatory response." *Shock (Augusta, Ga.)* 46.2 (2016): 122.
77. Dolicka, Dobrochna, et al. "mRNA Post-transcriptional regulation by au-rich element-binding proteins in liver inflammation and cancer." *International Journal of Molecular Sciences* 21.18 (2020): 6648.
78. Qiu, Quan, et al. "Toll-like receptor-mediated IRE1 α activation as a therapeutic target for inflammatory arthritis." *The EMBO journal* 32.18 (2013): 2477-2490.

9. Appendix

List of materials in the Appendix

Table 1: Top 20 list of genes differentially expressed in dendritic cells group 1 (cluster 1, 3, 4, 6 and 13)

Table 2: Top 20 list of genes differentially expressed in dendritic cells group 2 (cluster 8 and 16)

Table 3: Top 20 list of genes differentially expressed in dendritic cells group 3 (cluster 2)

Table 4: Top 10 list of genes differentially expressed in dendritic cells group 4 (cluster 0, 10 and 11)

Table 5: Top 20 list of genes differentially expressed in dendritic cells group 5 (cluster 18)

Table 6: Top 20 list of genes differentially expressed in macrophages (cluster

17)

Table 7 list of top 20 marker gene of each dendritic cell group

Table 1: Top 20 list of genes differentially expressed in dendritic cells group 1 (cluster 1, 3, 4, 6 and 13)

	p_val	avg_log2FC	pct. 1	pct. 2	p_val_adj
SAMSN1	0	1.288418583	0.977	0.891	0
PAK2	0	-0.698444521	0.451	0.761	0
SIAH2	0	0.725907281	0.868	0.653	0
RAB9A	0	0.852371532	0.97	0.909	0
LY75	0	-0.997550617	0.421	0.745	0
RNF19B	0	0.890338225	0.978	0.894	0
JAK1	0	-0.770568299	0.847	0.952	0
SMAP2	0	-1.105588028	0.735	0.931	0
BTG1	0	0.75580865	0.998	0.995	0
PTGES3	0	0.910372779	0.981	0.91	0
IFRD1	0	0.800903351	0.928	0.75	0
STK17A	0	0.905851764	0.838	0.534	0
DAPP1	0	-0.901617123	0.424	0.72	0
BMP2K	0	-1.00639882	0.408	0.754	0
ZFAND5	0	0.726408015	0.943	0.817	0
RPS28	0	0.395166399	0.999	0.998	0
SAFB2	0	0.873161417	0.93	0.762	0
TNIP1	0	1.638071501	0.98	0.887	0
ELL2	0	0.748410775	0.792	0.407	0
HSPA5	0	1.083824377	0.986	0.958	0

Table 2: Top 20 list of genes differentially expressed in dendritic cells group 2 (cluster 8 and 16)

	p_val	avg_log2FC	pct. 1	pct. 2	p_val_adj
RBM25	2.56E-15	1.325747469	0.415	0.242	4.13E-11
SRSF11	9.34E-13	0.994849338	0.306	0.147	1.51E-08
SREK1	1.98E-10	0.890755678	0.2	0.079	3.21E-06
RPS2	3.87E-10	-0.286372903	0.964	0.975	6.24E-06
CD74	1.26E-09	-0.476528026	0.781	0.828	2.04E-05
SFSWAP	2.92E-09	0.745404215	0.133	0.04	4.72E-05
SAFB2	4.06E-09	0.833177244	0.304	0.167	6.57E-05

RBBP6	4.18E-09	0.795431247	0.542	0.406	6.75E-05
RPS8	2.37E-08	-0.272184225	0.933	0.97	0.000383552
TRRAP	3.04E-08	0.858118991	0.114	0.033	0.000491845
ZC3H13	4.33E-08	0.662080587	0.102	0.027	0.000699295
DYNC1H1	4.52E-08	1.012958375	0.282	0.167	0.000729601
UBR4	1.14E-07	0.704792499	0.178	0.08	0.001834676
PCNT	1.20E-07	0.632722815	0.154	0.062	0.001943082
CCNL1	1.60E-07	0.575480204	0.451	0.319	0.002582981
FTH1	1.73E-07	-0.299996821	0.949	0.938	0.002791076
MDN1	1.93E-07	0.688779782	0.103	0.03	0.003116675
RPS24	3.14E-07	-0.283801825	0.857	0.9	0.005075904
RPL7A	1.10E-06	-0.289568608	0.849	0.868	0.017820047
HDLBP	1.19E-06	0.887418469	0.262	0.161	0.019283661

Table 3: Top 20 list of genes differentially expressed in dendritic cells group 3 (cluster 2)

	p_val	avg_log2FC	pct. 1	pct. 2	p_val_adj
SYNGR2	6.08E-180	-1.00523417	0.941	0.991	9.82E-176
ACTG1	1.47E-138	-1.003721583	0.98	0.996	2.38E-134
TNIP1	1.96E-133	1.582677026	0.921	0.727	3.16E-129
SAMSN1	2.53E-126	1.348193729	0.949	0.799	4.09E-122
STK4	5.05E-119	-1.020973175	0.527	0.891	8.16E-115
RBBP6	8.75E-119	1.035007119	0.965	0.879	1.41E-114
RBM39	1.79E-118	0.826257051	0.98	0.929	2.90E-114
SAFB2	7.37E-113	1.084120945	0.861	0.611	1.19E-108
USP12	1.91E-112	0.988285316	0.963	0.839	3.08E-108
CTSC	6.77E-111	-1.177377154	0.622	0.906	1.09E-106
PRRC2C	1.08E-106	-1.006242507	0.463	0.843	1.75E-102
SELPLG	2.89E-105	-1.088935179	0.471	0.834	4.67E-101
SNRPB2	6.43E-103	0.878808296	0.898	0.707	1.04E-98
LCP1	6.58E-103	-0.984221412	0.82	0.956	1.06E-98
COPS9	1.60E-96	0.782882478	0.925	0.806	2.58E-92
RYBP	2.81E-93	0.967755523	0.761	0.402	4.54E-89
ARPC5	4.14E-89	-0.799763929	0.882	0.953	6.68E-85
OSTF1	2.50E-88	0.602715313	0.989	0.974	4.04E-84
PAK2	1.05E-87	-0.771010146	0.404	0.759	1.70E-83

Table 4: Top 20 list of genes differentially expressed in dendritic cells group 4 (cluster 0, 10 and 11)

	p_val	avg_log2FC	pct. 1	pct. 2	p_val_adj
SYNGR2	1. 15E-223	-0. 729721494	0. 977	0. 981	1. 86E-219
RBBP6	5. 59E-216	1. 061933422	0. 972	0. 873	9. 03E-212
RBM39	5. 22E-202	0. 806751495	0. 988	0. 951	8. 43E-198
USP12	3. 51E-198	0. 866099812	0. 991	0. 955	5. 68E-194
RPS28	5. 07E-197	0. 387344396	0. 999	0. 998	8. 19E-193
HNRNPDL	4. 59E-196	0. 712501168	0. 992	0. 969	7. 42E-192
SRSF5	1. 25E-188	0. 640562367	0. 994	0. 963	2. 02E-184
TNIP1	1. 78E-182	1. 097035703	0. 972	0. 888	2. 87E-178
RYBP	1. 29E-181	0. 80057724	0. 832	0. 539	2. 08E-177
STK4	3. 27E-180	-0. 776329188	0. 397	0. 784	5. 29E-176
RABGEF1	1. 14E-175	0. 788345119	0. 84	0. 518	1. 85E-171
STK17A	1. 99E-175	0. 769013459	0. 906	0. 664	3. 22E-171
HMOX2	2. 18E-175	0. 914151225	0. 743	0. 383	3. 52E-171
MTDH	1. 33E-170	0. 772974717	0. 973	0. 932	2. 15E-166
PRRC2C	9. 86E-170	-0. 819866073	0. 379	0. 759	1. 59E-165
ARPC5	1. 67E-166	-0. 874466632	0. 916	0. 953	2. 69E-162
JPT1	1. 23E-161	0. 985775051	0. 982	0. 946	1. 99E-157
ZFAND5	2. 67E-157	0. 651693069	0. 97	0. 894	4. 32E-153
ACTN1	5. 57E-157	0. 853431345	0. 923	0. 7	8. 99E-153
SRRM1	5. 68E-156	0. 681248127	0. 967	0. 9	9. 17E-152

Table 5: Top 20 list of genes differentially expressed in dendritic cells group 5 (cluster 18)

	p_val	avg_log2FC	pct. 1	pct. 2	p_val_adj
TRA2B	5. 42E-09	1. 248741223	0. 964	0. 986	8. 75E-05
EPSTI1	5. 68E-09	-1. 613063656	0. 393	0. 959	9. 18E-05
HNRNPK	2. 81E-08	0. 873206861	1	0. 946	0. 000454027
EID1	1. 07E-07	-1. 038691283	0. 393	0. 838	0. 001731516
DCTN6	1. 07E-07	0. 986638495	0. 714	0. 189	0. 001732143
HNRNPM	1. 27E-07	1. 008959378	0. 964	0. 973	0. 002055207
RBM39	1. 91E-07	1. 050602949	1	1	0. 003087456
PLAC8B	2. 03E-07	-1. 221632179	0. 964	0. 986	0. 003279881
EIF4G2	3. 48E-07	1. 024875904	0. 964	0. 973	0. 005623695
IFRD1	4. 88E-07	1. 293231948	0. 929	0. 581	0. 007875657
ATP5IF1	5. 32E-07	-1. 076550999	0. 464	0. 838	0. 008591094
ARPC5	5. 68E-07	-1. 100378562	0. 429	0. 946	0. 009182039
DEK	7. 28E-07	-0. 735483815	0. 536	0. 919	0. 011756708

STK26	1. 16E-06	-0. 992114764	0. 107	0. 649	0. 018791928
SOD1	1. 54E-06	-1. 206672558	0. 714	0. 959	0. 024896051
CSRNP1	2. 02E-06	1. 556490104	0. 75	0. 378	0. 032669607
UBB	2. 03E-06	1. 066675109	1	1	0. 03280436
COA3	2. 12E-06	-0. 90153544	0. 071	0. 622	0. 034319201
ARPP19	3. 14E-06	0. 998548194	0. 893	0. 581	0. 050785605
PTBP1	3. 37E-06	0. 982743851	0. 964	0. 77	0. 054438866

Table 6: Top 20 list of genes differentially expressed in macrophages (cluster 17)

	p_val	avg_log2FC	pct. 1	pct. 2	p_val_adj
KDM6B	2. 64E-14	-1. 307439214	0. 215	0. 741	4. 27E-10
ZFP36	7. 29E-14	-1. 332915907	0. 266	0. 807	1. 18E-09
ENSBTAG00000004043	3. 17E-13	-1. 149241164	0. 304	0. 807	5. 11E-09
BOLA-DMA	4. 67E-13	-0. 987095075	0. 203	0. 724	7. 54E-09
TMSB4X	3. 71E-12	-0. 938703079	0. 937	0. 987	5. 99E-08
RAMP2	1. 32E-11	-0. 977880767	0. 165	0. 645	2. 14E-07
ARHGDIB	1. 99E-11	-0. 690834812	0. 253	0. 776	3. 22E-07
MEF2A	2. 60E-11	-0. 875497087	0. 228	0. 675	4. 20E-07
CNP	2. 93E-11	-0. 860812618	0. 038	0. 469	4. 73E-07
TET2	6. 02E-11	-0. 551176549	0. 051	0. 496	9. 73E-07
HMGB2	6. 84E-11	-0. 860510292	0. 241	0. 728	1. 11E-06
SELL	7. 25E-11	-0. 964490999	0. 19	0. 689	1. 17E-06
SARS1	1. 11E-10	-0. 71270775	0. 266	0. 706	1. 80E-06
EIF5A	1. 28E-10	-0. 879181705	0. 367	0. 798	2. 06E-06
GALNT1	1. 35E-10	-0. 781365404	0. 215	0. 68	2. 18E-06
GLIPR2	1. 74E-10	-0. 873904093	0. 316	0. 759	2. 81E-06
APRT	1. 74E-10	-0. 794327998	0. 405	0. 86	2. 82E-06
RBM3	1. 76E-10	-0. 902713729	0. 304	0. 711	2. 85E-06
XDH	1. 77E-10	-0. 848687503	0. 291	0. 75	2. 86E-06
P4HB	2. 15E-10	-0. 721245126	0. 354	0. 794	3. 47E-06

Table 7 list of top 20 marker gene of each dendritic cell group

p_val	avg_log2FC	pct. 1	pct. 2	p_val_adj	cluster	gene
0	1. 47889968	0. 595	0. 251	0	DC_g1	CCL17
0	1. 28965682	0. 815	0. 508	0	DC_g1	CD1B
0	1. 261183915	0. 803	0. 58	0	DC_g1	FABP5
0	1. 192390703	0. 83	0. 385	0	DC_g1	C18H19orf33
0	1. 182537223	0. 413	0. 167	0	DC_g1	GADD45G

0	1.158503617	0.745	0.48	0	DC_g1	CFD
0	1.156954516	0.945	0.638	0	DC_g1	GPR183
0	1.025850768	0.744	0.428	0	DC_g1	ECM1
0	0.972515034	0.896	0.717	0	DC_g1	LYZ
0	0.970778019	0.856	0.672	0	DC_g1	TNFAIP3
0	0.942148102	0.681	0.409	0	DC_g1	TMEM158
0	0.916635974	0.989	0.84	0	DC_g1	TYROBP
0	0.88831548	0.235	0.01	0	DC_g1	EHF
0	0.869623381	0.997	0.916	0	DC_g1	NFKBIA
0	0.856442839	0.6	0.319	0	DC_g1	CD1E
0	0.851588128	0.839	0.568	0	DC_g1	IRF4
0	0.802853025	0.397	0.156	0	DC_g1	C15H11orf96
3.12E-297	0.861654729	0.829	0.62	5.04E-293	DC_g1	IER3
6.61E-219	1.197075576	0.247	0.074	1.07E-214	DC_g1	GR01
4.18E-105	0.814462588	0.257	0.132	6.75E-101	DC_g1	CXCL8
6.34E-109	1.075273112	0.257	0.798	1.02E-104	DC_g2	ND4L
1.73E-103	1.070609182	0.232	0.723	2.80E-99	DC_g2	RRBP1
1.22E-88	1.045181702	0.235	0.694	1.98E-84	DC_g2	DYNC1H1
1.41E-81	1.142164826	0.214	0.634	2.28E-77	DC_g2	ATP8
9.04E-81	0.977154761	0.212	0.627	1.46E-76	DC_g2	BOD1L1
3.01E-73	0.989105321	0.344	0.91	4.85E-69	DC_g2	RBM25
1.07E-66	1.270643127	0.672	0.996	1.73E-62	DC_g2	ND3
4.20E-66	1.157191616	0.22	0.602	6.78E-62	DC_g2	HDLBP
1.62E-50	1.213650044	0.328	0.799	2.62E-46	DC_g2	ANKRD11
2.21E-49	1.230248071	0.241	0.6	3.56E-45	DC_g2	SRSF11
9.77E-43	1.828989376	0.837	0.996	1.58E-38	DC_g2	COX2
1.91E-40	0.910962779	0.385	0.897	3.09E-36	DC_g2	SFPQ
9.31E-32	1.75427079	0.872	0.997	1.50E-27	DC_g2	CYTB
2.70E-17	1.762176136	0.871	0.998	4.35E-13	DC_g2	ATP6
4.21E-14	1.675065434	0.871	0.998	6.79E-10	DC_g2	ND4
2.31E-12	1.836173483	0.947	0.999	3.73E-08	DC_g2	COX3
2.74E-07	1.8022305	0.886	0.998	0.004424303	DC_g2	COX1
4.51E-06	1.783098115	0.869	0.998	0.072803678	DC_g2	ND2
8.93E-06	0.925273034	0.487	0.926	0.144240332	DC_g2	RBBP6
0.008364024	1.367179752	0.565	0.992	1	DC_g2	ND5
0	3.687391526	0.608	0.081	0	DC_g4	CXCL9
0	1.773436852	0.953	0.454	0	DC_g4	DNASE1L3
0	1.643224485	0.98	0.81	0	DC_g4	TXN
0	1.504747348	0.864	0.443	0	DC_g4	EN03
0	1.306776853	0.971	0.781	0	DC_g4	WARS1
0	1.252828236	0.383	0.021	0	DC_g4	IL12B
0	1.154198083	0.977	0.803	0	DC_g4	UCP2
0	1.102207298	0.977	0.818	0	DC_g4	TAGLN2
0	0.997328426	0.965	0.78	0	DC_g4	BATF3

0	0.957622455	0.906	0.519	0	DC_g4	C3H1orf54
0	0.957252471	0.991	0.898	0	DC_g4	VIM
0	0.955265227	0.976	0.833	0	DC_g4	SERPINB1
0	0.948626377	0.97	0.771	0	DC_g4	CD48
0	0.94487019	0.817	0.35	0	DC_g4	UBD
0	0.894246823	0.792	0.314	0	DC_g4	IFI27L2
0	0.888843247	0.983	0.885	0	DC_g4	BLA-DQB
0	0.853767373	0.991	0.904	0	DC_g4	LGALS1
0	0.823839684	0.847	0.393	0	DC_g4	FAM229A
2.44E-256	0.809669065	0.93	0.746	3.93E-252	DC_g4	ENSBTAG00000009656
1.90E-223	0.973771708	0.741	0.467	3.07E-219	DC_g4	ENSBTAG00000037605
0	2.213799802	0.51	0.068	0	DC_g3	ENSBTAG00000047449
0	1.430253245	0.684	0.063	0	DC_g3	RBP4
0	1.394439157	0.717	0.294	0	DC_g3	PSPH
0	1.374244894	0.823	0.3	0	DC_g3	C10orf99
0	1.184896337	0.581	0.014	0	DC_g3	EPCAM
0	1.150777422	0.998	0.983	0	DC_g3	RPS6
0	0.968594585	0.947	0.764	0	DC_g3	ARID4A
0	0.939131145	0.735	0.423	0	DC_g3	BCKDHB
0	0.886005158	0.702	0.342	0	DC_g3	MISP3
0	0.864583057	0.627	0.254	0	DC_g3	RARRES1
0	0.843688282	0.405	0.089	0	DC_g3	S100B
0	0.720379639	0.378	0.032	0	DC_g3	ENSBTAG00000034220
3.65E-302	0.78992328	0.907	0.749	5.90E-298	DC_g3	AKAP13
2.29E-265	0.753898248	0.942	0.838	3.70E-261	DC_g3	CMTM6
4.59E-256	0.70917023	0.376	0.121	7.42E-252	DC_g3	SLC4A4
9.18E-248	0.765527137	0.767	0.532	1.48E-243	DC_g3	CLINT1
3.48E-205	0.714204256	0.77	0.605	5.62E-201	DC_g3	PRDX5
2.36E-175	0.823369859	0.357	0.138	3.82E-171	DC_g3	CYP26B1
3.55E-172	0.86231657	0.608	0.388	5.74E-168	DC_g3	REV3L
2.07E-170	0.717116476	0.82	0.642	3.34E-166	DC_g3	KCTD20
0	4.486687599	0.951	0.017	0	DC_g5	LTB
0	4.412269497	0.98	0.018	0	DC_g5	PLAC8B
0	3.910947962	0.99	0.057	0	DC_g5	GRN
0	3.633378741	0.961	0.045	0	DC_g5	DERL3
0	3.450686432	0.765	0.006	0	DC_g5	JCHAIN
0	3.357100634	0.627	0	0	DC_g5	LY6D
0	3.31509029	0.961	0.058	0	DC_g5	FYB1
0	2.808236975	0.912	0.004	0	DC_g5	SELL
0	2.597791524	0.902	0.007	0	DC_g5	ADA2
0	2.487977962	0.863	0.006	0	DC_g5	ENSBTAG00000003408
8.46E-216	3.654290997	1	0.109	1.37E-211	DC_g5	LY6E
3.63E-215	4.761811181	0.99	0.108	5.87E-211	DC_g5	CD52
2.27E-203	3.714373413	1	0.118	3.67E-199	DC_g5	PLD4

2. 25E-180	4. 345064515	1	0. 136	3. 64E-176	DC_g5	FCER1G
1. 30E-92	4. 095662516	1	0. 347	2. 11E-88	DC_g5	LAPTM5
1. 50E-81	2. 622485026	0. 98	0. 366	2. 43E-77	DC_g5	CTSD
6. 64E-79	2. 566446267	0. 99	0. 412	1. 07E-74	DC_g5	CHID1
1. 44E-75	4. 814131664	1	0. 516	2. 32E-71	DC_g5	CTSS
8. 45E-74	3. 148704986	0. 892	0. 287	1. 36E-69	DC_g5	SOD1
7. 70E-58	2. 688729003	0. 961	0. 62	1. 24E-53	DC_g5	CYSTM1