# Computational Exploration of Single-Cell RNA-seq Data in Systemic Lupus Erythematosus (SLE) PBMCs

**Dataset and Study Context**

Single-cell RNA sequencing (scRNA-seq) data were analyzed from peripheral blood mononuclear cells (PBMCs) derived from two SLE patients and one healthy control (GEO accession: **GSE162577**). The dataset was generated using the 10x Genomics platform and investigates immune heterogeneity in lupus.

Systemic Lupus Erythematosus (SLE) is a chronic autoimmune disease characterized by immune dysregulation, interferon signaling activation, aberrant T and B cell responses, and inflammatory myeloid expansion. Single-cell approaches enable identification of pathogenic immune subsets and rare inflammatory populations.

Project Overview:

The Seurat pipeline included:

1. Quality control filtering
2. Log-normalization
3. Identification of highly variable genes
4. PCA and dimensionality reduction
5. Graph-based clustering
6. Marker gene identification
7. Differential expression (SLE vs Control)

# Quality Control Thresholds and Justification

**Selected thresholds:**

- nFeature_RNA > 200
- nFeature_RNA < 6000
- percent.mt < 15%

**Rationale**

**Lower bound (200 genes):**
Cells with fewer than 200 detected genes likely represent empty droplets or low-complexity captures.

**Upper bound (6000 genes):**
Cells with excessively high gene counts likely represent doublets or multiplets.

**Mitochondrial percentage (<15%):**
Elevated mitochondrial RNA content reflects stressed or apoptotic cells. SLE PBMCs often show increased stress signatures; therefore, a moderate cutoff of 15% preserved biological signal while removing dying cells.

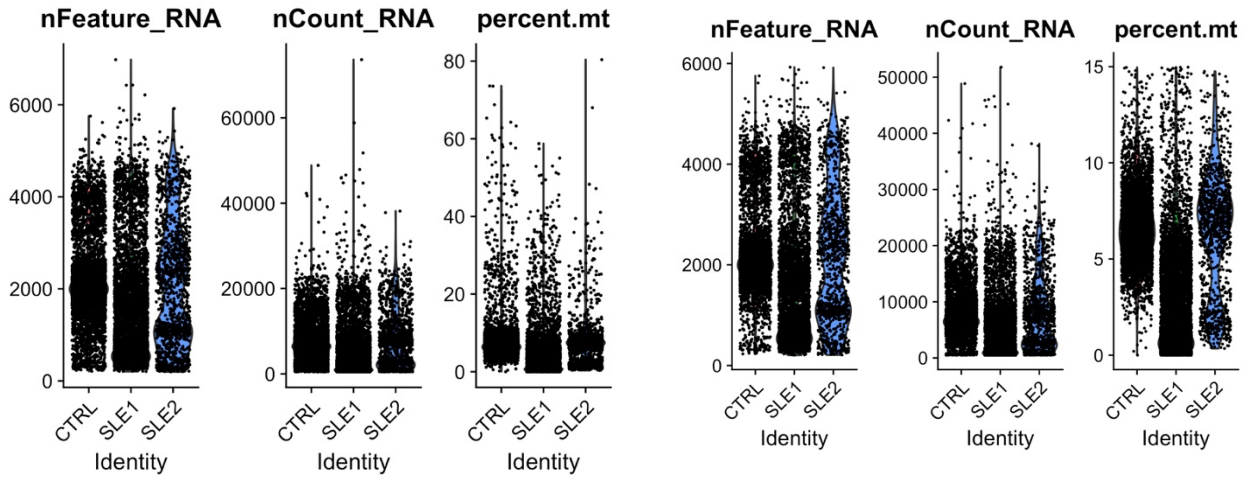


**Figure 1: QC Violin Plots (Before and After Filtering)**
Distribution of nFeature_RNA, nCount_RNA, and percent.mt across CTRL, SLE1, and SLE2 samples before and after filtering. Post-filtering plots demonstrate removal of high-mitochondrial and low-complexity outliers.

Filtering substantially reduced extreme mitochondrial values while preserving the majority of biologically relevant cells.

# Cluster Stability and Dimensionality Selection

**PCA Diagnostics:** The elbow plot (Figure 3) demonstrated a clear inflection around PCs 8–10. Variance explained plateaued beyond PC 10.

Clustering was therefore performed using:

- PCs 1–12
- Resolution = 0.5

This parameter choice balanced:

- Preservation of rare populations
- Avoidance of over-fragmentation

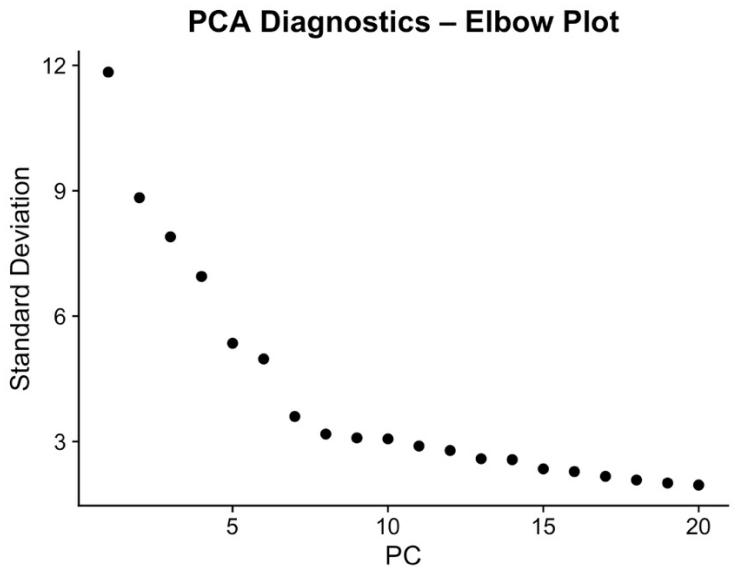


**Figure 3: PCA Elbow Plot:** Standard deviation of principal components. The elbow around PC 8–10 supports use of the first 12 PCs.

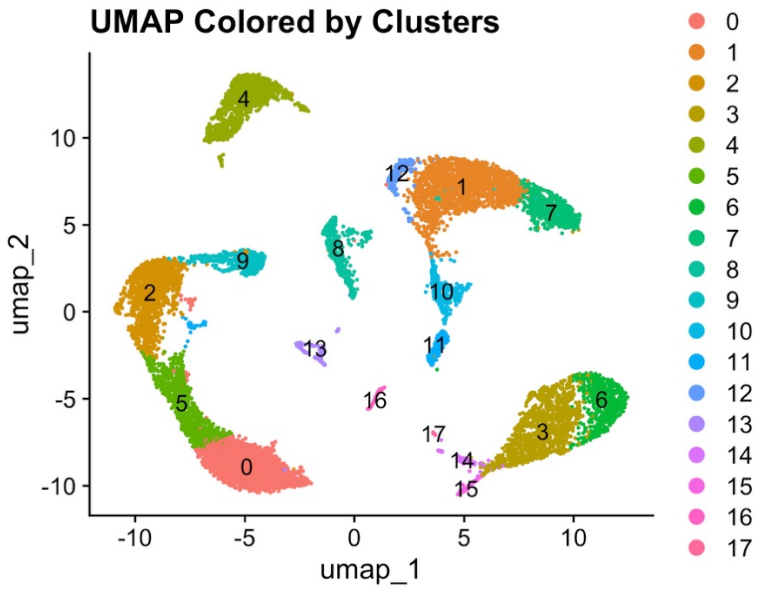


**Figure 4: UMAP Visualization:** UMAP embedding of PBMC cells colored by cluster identity. Eighteen transcriptionally distinct clusters were identified.

Clusters show clear separation with minimal overlap, indicating stable structure under chosen parameters.

# Cell Type Annotation Using Canonical Markers

Cluster annotation was performed using top markers identified via FindAllMarkers.

### Major Identified Cell Types:

| Marker Genes | Assigned Cell Type |
|---|---|
| **CD3D, CD3E, CD2, TRAC** | T cells |
| **NCR1, KLRD1, GZMB** | NK cells |
| **S100A8, S100A9, LST1** | Classical monocytes |
| **MS4A1, CD79A** | B cells |

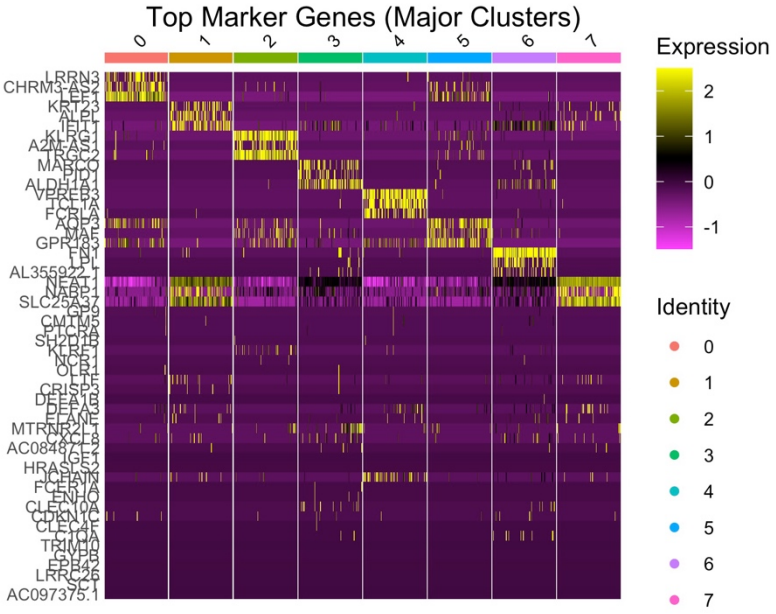


**Figure 5A: Marker Heatmap-** Scaled expression of top marker genes across major clusters. Distinct expression patterns confirm cluster specificity.

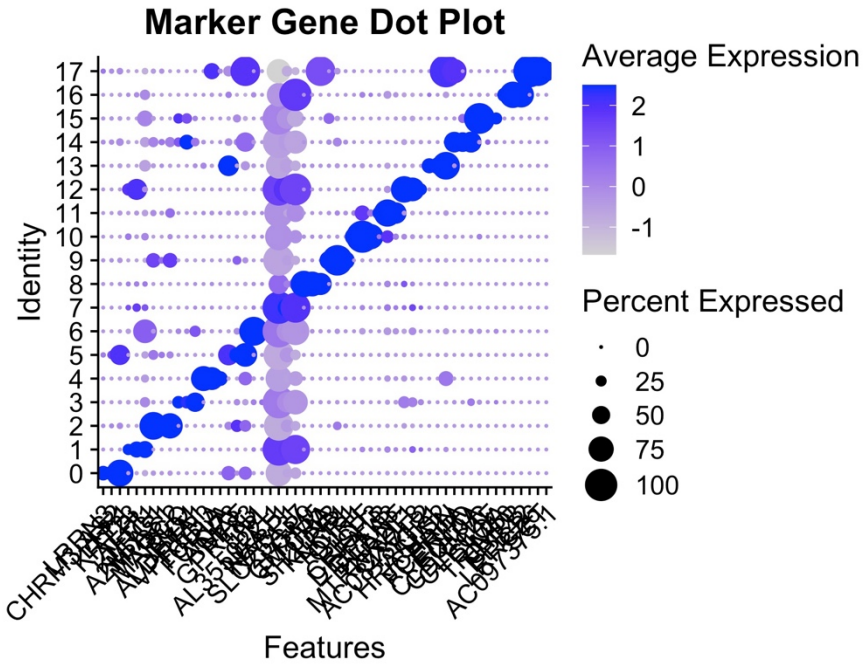


**Figure 5B: Dot Plot-** Dot size indicates percent expression; color indicates average expression. Canonical lineage markers show restricted distribution across clusters.

### Biological Interpretation

SLE PBMCs show:
Expanded inflammatory monocyte populations (S100A8/A9 high)
Activated T cell subsets
Distinct NK populations
Small rare clusters potentially representing interferon-high states

# Differential Expression: SLE vs Control

Differential expression analysis compared all SLE cells to control cells.
Parameters:

- logfc.threshold = 0.25
- min.pct = 0.1

### Figure 6: Volcano Plot

*Legend:* Differential expression between SLE and CTRL PBMCs. Each dot represents one gene; vertical dashed lines indicate |log2FC| = 0.5.

**Upregulated in SLE:**

- **S100A8** (strong positive log2FC)
    - Marker of inflammatory monocytes
    - Associated with neutrophil-like inflammatory programs
    - Known biomarker in lupus nephritis and systemic inflammation

**Downregulated in SLE:**

- **CD3D, CD3E, CD2**
    - Core T cell receptor complex components
    - Suggest altered T cell signaling or exhaustion

### Interpretation in Lupus Context

SLE is characterized by:

- Myeloid activation
- Chronic interferon signaling
- Adaptive immune dysregulation

Upregulation of S100A8 reflects inflammatory monocyte expansion, consistent with published lupus immunopathology. Downregulation of TCR-related genes suggests altered T cell functional states.

The widespread distribution of significant genes in the volcano plot indicates systemic immune remodeling in SLE PBMCs.

# Study-Linked Insight

### Rare Cluster Identification

Clusters 16 and 17 contain small cell numbers (<100 cells).

These may represent:

- Interferon-high immune subsets
- Activated transitional states
- Rare inflammatory populations

SLE pathogenesis is strongly driven by type I interferon signaling. Rare interferon-stimulated gene (ISG)-high subsets have been reported in lupus PBMC studies. These minor clusters may correspond to such populations.

Single-cell analysis uniquely enables detection of these rare but potentially pathogenic subsets.