The characterization of bone marrow cells by applying Single-cell RNA sequencing data analysis

Maria Uzun, Alisa Fedorenko

1. Preprocessing

This work was conducted using Scanpy tutorial [1]. The preprocessing step includes filtering out cells that have less than 200 genes expressed and genes that are detected in less than 3 cells. At the initial stage, we have got **33454**. We filtered out 6 cells and 1 gene based on those parameters, respectively, so further steps were applied to the **33448** cells after filtering out. We also filtered out some mitochondrial genes.

The highly expressed genes were identified by those parameters: min_mean=0.0125, max_mean=3, min_disp=0.5.

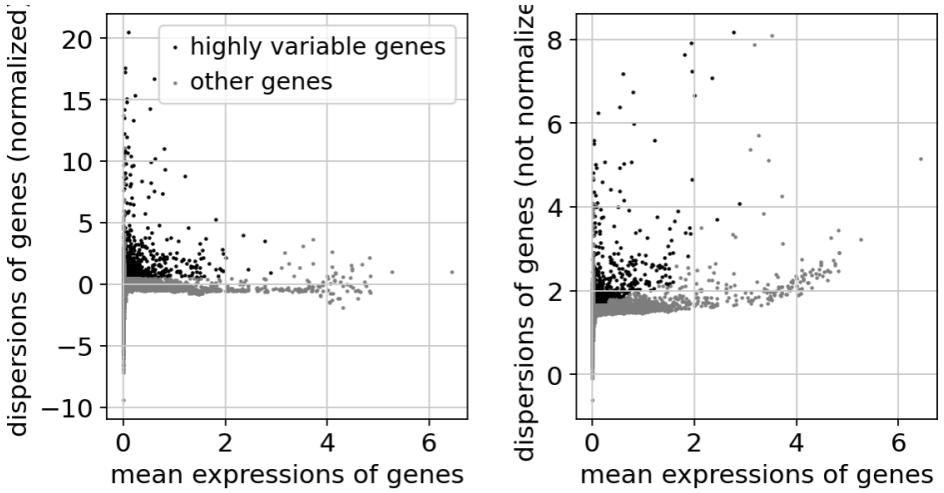


Figure 1. The result of the highly-variable-genes detection

2. Principal component analysis

The principal component analysis revealed **20 PCs** we should consider in order to compute the neighborhood. Considering more than 20 PCs is pointless.

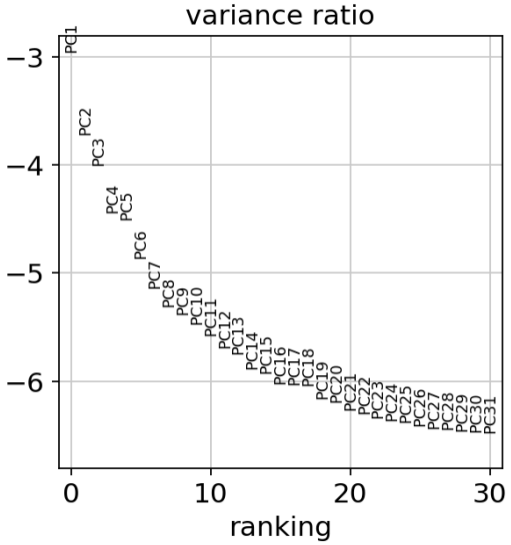


Figure 2. The contribution of single PCs to the total variance in the data.

3. Clustering the neighborhood graph

Leiden graph-clustering method and **resolution** equals to **0.08**. As a result, 6 clusters were identified. The optimal resolution parameter selection was based on the gene markers “CST3”, and “NKG7” (the basic markers from the scanpy tutorial) and the visual distribution of clusters.

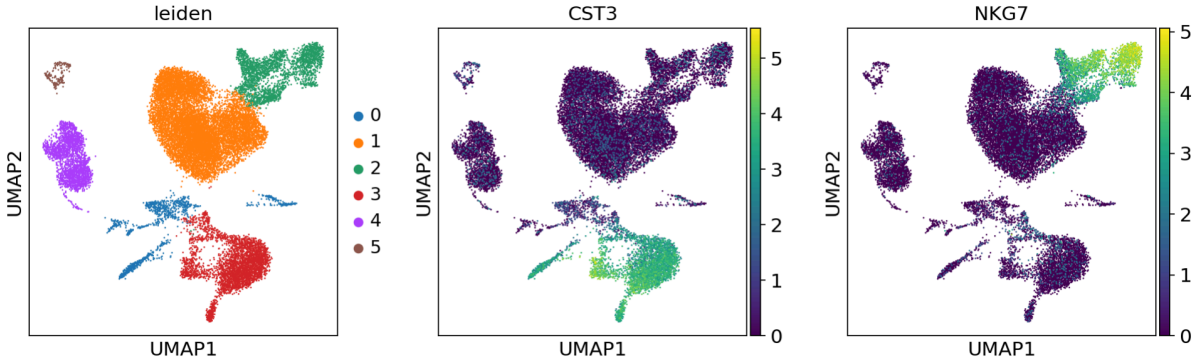
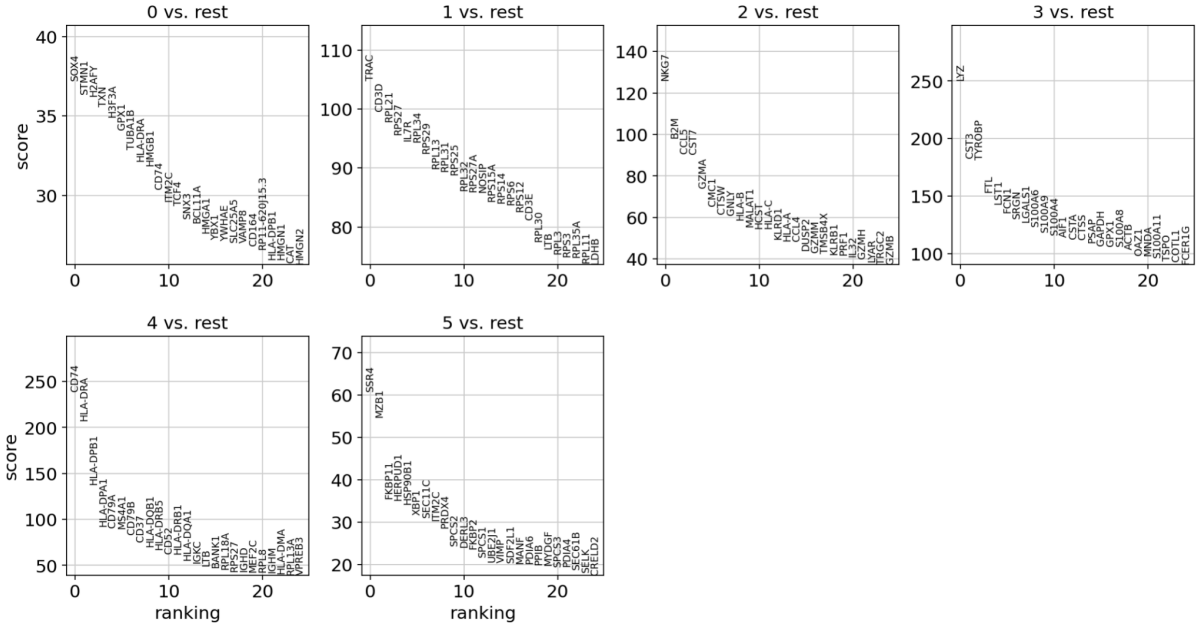


Figure 3. UMAP clustering of single-cell data with defined 6 clusters.

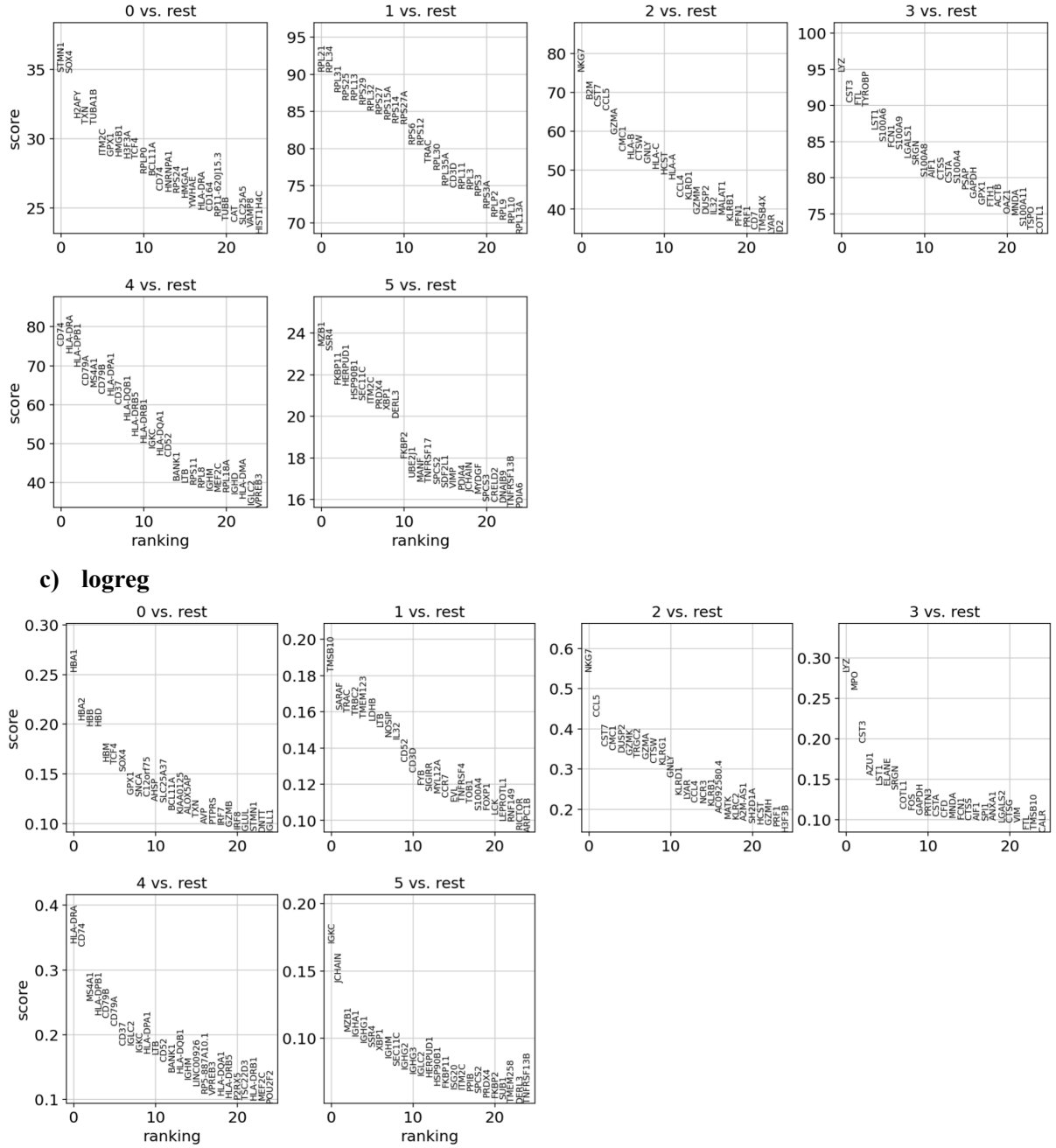
4. Finding marker genes

To identify highly differential genes in each cluster we applied 3 methods to do so.

a) t-test



b) Wilcoxon



c) logreg

Figure 4. Ranking for the highly differential genes in each cluster. a)

All marker genes are recovered in almost all approaches and presented in **Table 1**. The cell types were defined by special gene markers represented in sources [2, 3].

Table 1. List of marker genes and corresponding cell types for each cluster.

Leiden group	Markers	Cell Type
0	STMN1, TUBA1	B cells proliferative
1	TRAC, CD3D, IL7R	CD4+ T cells
2	NKG7	NK cells
3	CST3, LYZ	CD14+ Monocytes
4	CD74	Macrophages
5	MZB1, IGKC	Plasma B cells

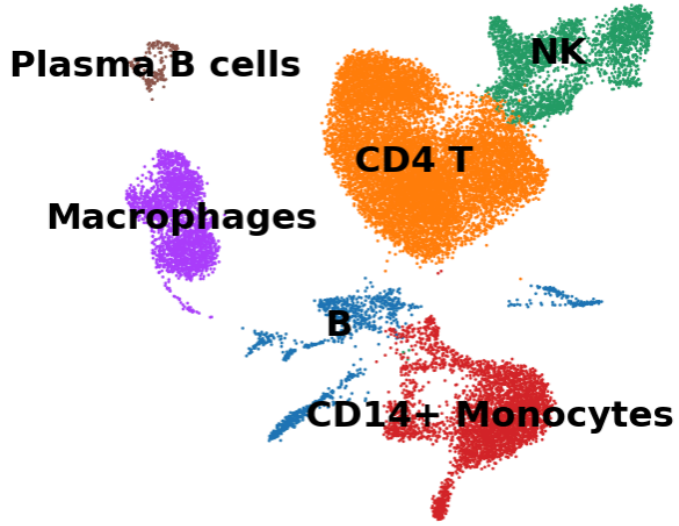


Figure 5. UMAP clustering of single-cell data.

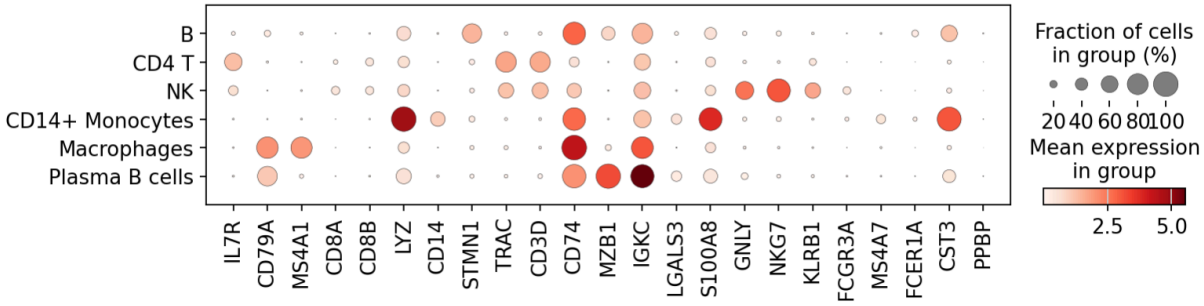


Figure 6. Expression of marker genes across clusters. The markers are used to identify the cell types constituting each cluster.

References:

1 - https://scanpy-tutorials.readthedocs.io/en/latest/pbmc3k.html
2 - Nirmal, A. J., Regan, T., Shih, B. B., Hume, D. A., Sims, A. H., & Freeman, T. C. (2018). Immune cell gene signatures for profiling the microenvironment of solid tumors. Cancer immunology research, 6(11), 1388-1400.
3 - https://genome.cshlp.org/content/suppl/2021/09/21/gr.273300.120.DC1/Supplemental_Table_S2.pdf