Single-cell Analysis on Peripheral Blood Mononuclear Cells for Revealing the Landscape and its Adaptive Immunity Heterogeneity

(no AI correction, all are original from myself)

Abstract

It is well-known that the immune system plays an important role in defending human health from the attack of virus and other outside pathogens. There are 2 parts consisting of this system. One is the innate immunity that has been treated as the first barrier in order to prevent the pathogen away from the body. Another is the adaptive immunity whose core elements are B cells and T cells full of different functions. The latter is what we focus on the in this report due to its higher complexity. In this report, our primary goal is to explore the cell type landscape from a single-cell RNA-seq dataset. What is more, it is valuable to gain insights from the cellular heterogeneity in certain cell types in order to find potential differences in terms of functions. By using dataset from peripheral blood mononuclear cells (PBMC) which is a reference available in the Seurat, results show that after applying batch effect correction, the data reveals different cell clusters, in which B cells, CD4+ T cells, cytotoxic T cells are the main parts in this dataset, thus highlighting their core roles in the immune system especially the adaptive immunity. Further heterogeneity analysis identifies sub-clusters among these cell types by certain markers. Even some of them cannot be certainly determined, the results still suggest that the heterogeneity exists among them and can be classified due to their markers in most cases. Therefore, this report provides a foundation for future investigations of immune cell diversity and functionality.

Introduction

The immune system is essential for protecting the body against infections and maintaining health. It consists of two parts: The first one is innate immunity acting as the body's first barrier of defense which can provide non-specific responses and the ability to kill pathogens outside. The other is the adaptive immune system, which is apparently offers a more specialized and complex defense mechanism since it can only be acquired after birth and changed based on surrounding environments. In most cases, it is mediated by T cells and B cells which only exist in adaptive immune system and have the ability to recognize specific antigens, consequently generating targeted antibodies to kill target pathogens which is the function of B cells and CD8+ T cells (Rastogi I, et al.2022; Raskov H, et al. 2021) as well as remember them for future encounters and instant generation which is the function of memory B cells (Inoue T, Kurosaki T. 2024). Therefore, comparing with innate counterpart, the adaptive one is of more value for studies as it involves with a long-lasting immune response that can be altered based on different environments in terms of controlling infections, eliminating abnormal cells, and providing immune memory, which is full of complexity for deeper analysis.

Dysregulation of immune cells can lead to a variety of diseases obviously, including autoimmune disorders, infections, and cancers. So it is ever essential to study immune cells in order to understand the mechanism of defending against disease as well as maintaining immune homeostasis. In this study, our primary goal is to explore the immune landscape across different platforms. What is more, by gaining insights from the cellular heterogeneity, we may find different subclusters with certain functions in certain cell types, which will obviously provide a clearer understanding of the composition and functional diversity of immune cells in humans in the future.

Methods

Data

The PBMC single-cell atlas (pbmcsa) dataset was downloaded from publicly available resources by using the SeuratData package (Slovin S, et al. 2021). This dataset has been treated as a reference for immune cell landscapes, that is why it is stored in the SeuratData package. In this case, we only focused on samples from 10x Chromium (v3), CEL-Seq2, Drop-seq, Seq-Well, inDrops, Seq-Well and Smart-seq2, which are known for their robust sequencing capabilities and ability to capture cellular heterogeneity. We extract them as ‘pbmc’ for analysis here.

Preprocessing

Normalization: Ensure comparability between cells.

Feature Selection: Find the most variable features and record.

Scaling: Ensure that data is centered and standardized for downstream analyses like principal component analysis (PCA) and Uniform Manifold Approximation and Projection (UMAP).

Batch Effect Correction: To address the batch effects coming from multiple sequencing platforms, here we used ‘IntegrateLayers’. Rather than utilizing Seurat's reference-based integration approach, this integration method avoids manually selecting a reference dataset and instead considers all batches to provide a comprehensive correction in order to prevent ourselves away from potential prejudice.

What is more, to quantify the result of batch correction, here we used ‘Silhouette’ scores (Li J, et al. 2020). This metric assesses the clustering of cells within their respective cell types and their separation from other clusters. And there are 3 main boundaries for this evaluation: ‘+1’ means a perfect score indicating the perfect clustering. ‘0’ means it's ambiguous whether the point belongs to one cluster or another. ‘-1’ indicate that the data point might be incorrectly clustered.

Visualization

UMAP: UMAP is utilized here in order to visualize the results. UMAP is well-known for its ability to provide a two-dimensional representation of high-dimensional data while still keeping relative distances among different clusters, which can help us uncover the relationships between different immune cell subsets.

Heatmaps: Heatmaps were used to visualize the expression of high expressed genes across different cell types, which can highlight the heterogeneity and distinct subpopulations in each major immune cell types.

Results

Global Landscape of Cell Types and Their Markers

After batch effects correction, we can see the UMAP of pbmc gathered from multiple platforms. The main cell types include: B cells, CD4+ T cells, cytotoxic T cells, CD14+ monocytes, CD16+ monocytes, dendritic cells and plasmacytoid dendritic cells (Figure 1). As for the count in each cell type, B cells and 2 kinds of T cells as well as CD14+ monocytes rank higher than others, which matches the research facts that these are the main part in the immune system in the blood.

What is more, it is shown that some cell types overlap with each other, especially cytotoxic T cells which clearly overlap with CD4+ T cells and Natural killer cells. It is reasonable since CD4+ T cells even own a different function from cytotoxic T cells, they are still T cells so many genes will share the similar expression pattern that cause the overlap. While for NK cells, it is due to its function. Bothe NK cells and cytotoxic T cells have the ability to kill the abnormal cells or virus by secreting cytotoxic molecular to let targets to undergo apoptosis, which makes the pattern share the overlap parts for both cell types.

Another significant overlap occurs in 2 kinds of monocytes. Monocytes, defined as a type of white blood cell that play a crucial role in the immune system (Guilliams M, et al. 2018), can be classified into two main subsets based on their expression of surface markers: CD14+ monocytes and CD16+ monocytes. CD14+ monocytes are typically considered the classical monocytes, while CD16+ monocytes, on the contrary, are often referred to as non-classical or intermediate monocytes (Narasimhan PB et al. 2019). CD14+ monocytes are critical for the initiation and modulation of immune responses, known for their ability to phagocytose pathogens, present antigens, and secrete pro-inflammatory cytokines like Tumor Necrosis Factor-alpha (TNF-α) and Interleukin-6 (IL-6) (Tang C, et al. 2014; Kitamura H, et al. 2017), which demonstrates their vital roles in the early stages of immune defense.

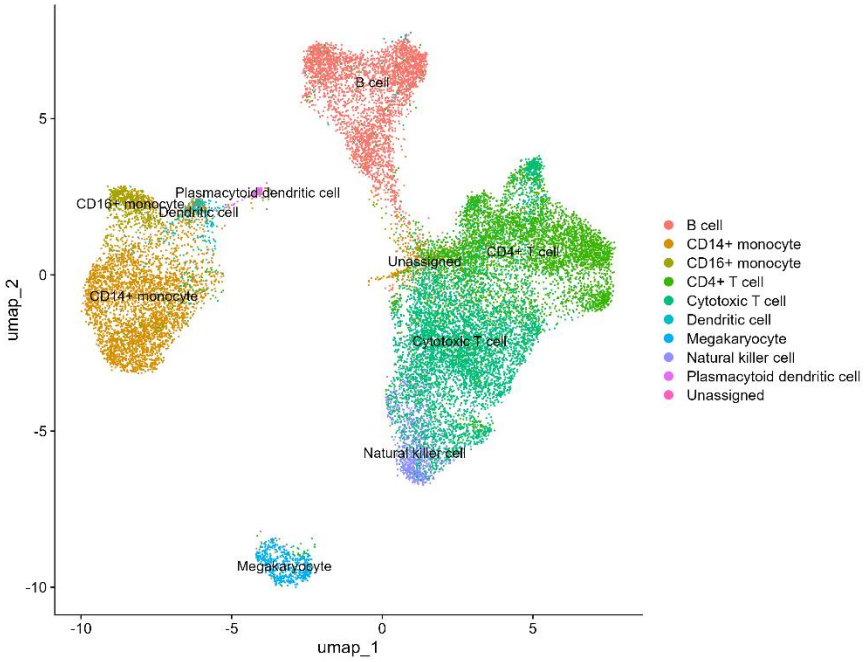

Figure 1. UMAP for pbmc Cell Types after Batch Effect Correction

When turning to the highly expressed genes (Figure 2), for cytotoxic T cell, the Granzyme H (GZM H) shows the highest expression in the figure 2 as well as in NK cells, suggesting the reasonability of overlap. GZMH is involved in the process of inducing apoptosis (programmed cell death) in target cells. When cytotoxic T cells or NK cells recognize and bind to infected or cancerous cells, they release granules containing granzymes, including GZM belongs to granzyme (Gzm) families which have the ability to enter the target cell and activate pathways

leading to apoptosis (Shresta S, et al. 19956; Susanto O, et al. 2013). When scanning the data, it is found that most of CD8A and CD8B are expression in cytotoxic T cell, suggesting there are more marker gene expressing higher levels than CD8 even CD8 is one of the markers for marking CD8+ T cells (Figure 3).

While for CD4+, it is shown that CD14+ monocytes own a more expression level of CD4+ compared with the former. Additionally, dendritic cells that play a vital role in linking the innate and adaptive immune systems by capturing and presenting antigens to T cells also express CD4 (Figure 3), suggesting that CD4 is more likely to separate CD4+ T cells from CD8+ T cells. Another fact is that the expression of CD4 in these cells underscores its significance in facilitating T cell activation, but it also reveals complexity in this immune system. As for B cells, it is demonstrated that may genes are related with immunoglobulin (Ig) genes, suggesting function of B cells: generating antibodies, memorizing antigen for generating antibodies as fast as it can next time when the antigen is going through human bodies.

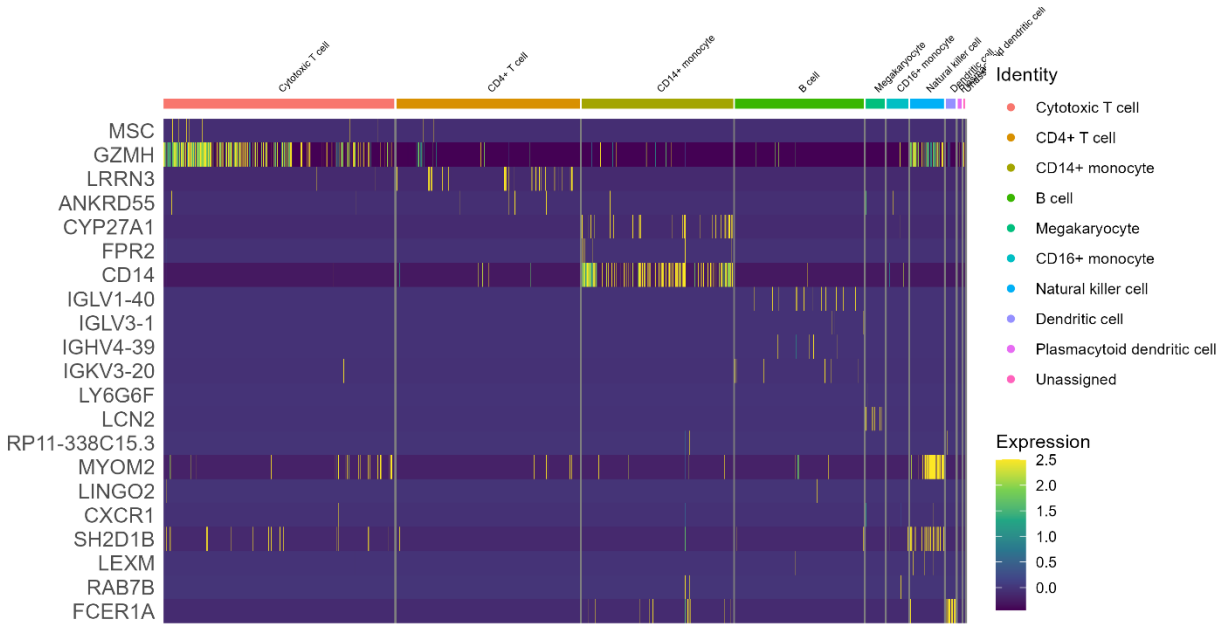

Figure 2. Heatmap of High Expression Genes after Batch Effect Correction

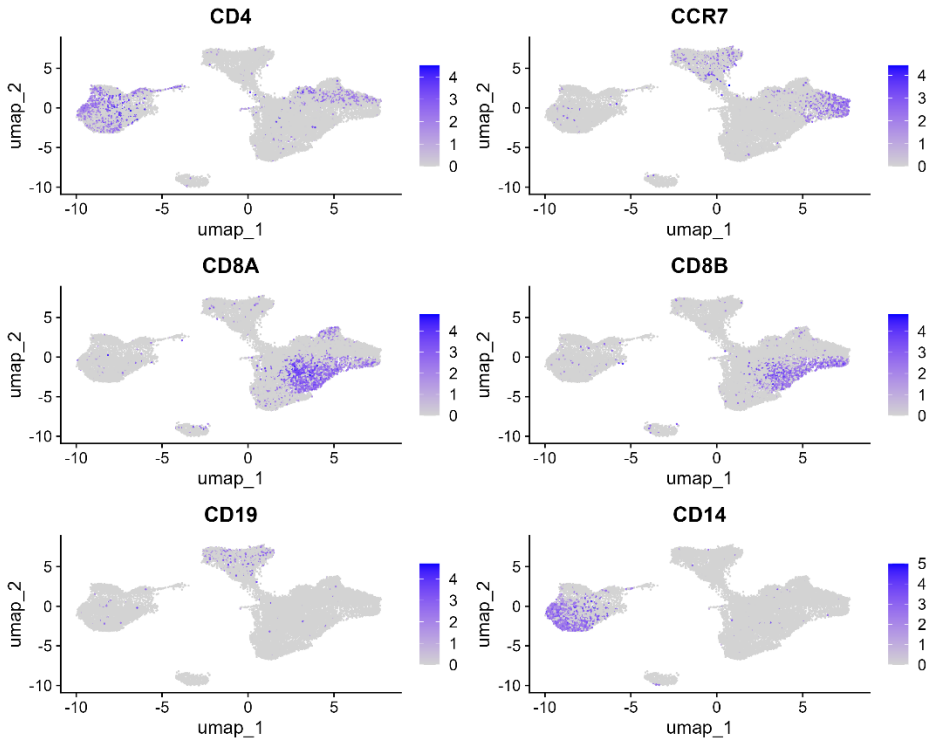

Figure 3. Feature Plots for CD4, CCR7(CD4+ T cells), CD8A, CD8B (Cytotoxic T cells), CD19 (B cell) and CD14 (CD14+ monocytes)

Heterogeneity in Adaptive Immune Cells

In order to specify sub-clusters in adaptive immune cells such as CD4+ T cells, we decide to analyze the clusters of B cells, CD4+ T cells, cytotoxic T cells solely (Figure 3). For selecting the highly expressed genes, we choose to select every top 10 ranked genes in each subset. In this way, we have tried our best to find the markers for marking each cell type in every subset and most of parts are successful to set the markers from Figure 4 to Figure 9, demonstrating the heterogeneity among the adaptive immune cells. Some markers overlap especially turning to CD4+ T cells and cytotoxic T cells as some clusters share the same genes expression, suggesting that same origin for T cells. Unfortunately, some of clusters like cluster 6 in B cells, own the highly expression data with long-noncoding RNA (lncRNA), which cannot be treated as a valid result. Another dilemma is just like cluster 3 in B cells whose highly expressed genes are also seen in other subclusters so it is hard to classify it from other subclusters like 4 which owns the NFKBID expression, which indicates that this cluster cannot simply treat as a unique group. In the contrast, it can somehow belong to other cluster in different situations.

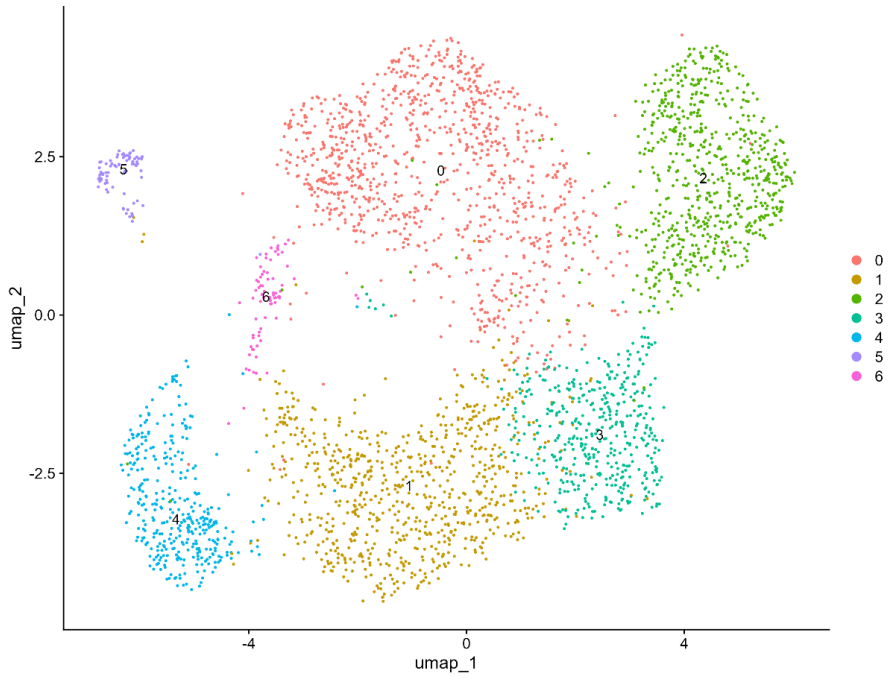

Figure 4. UMAP of B Cell Subclusters

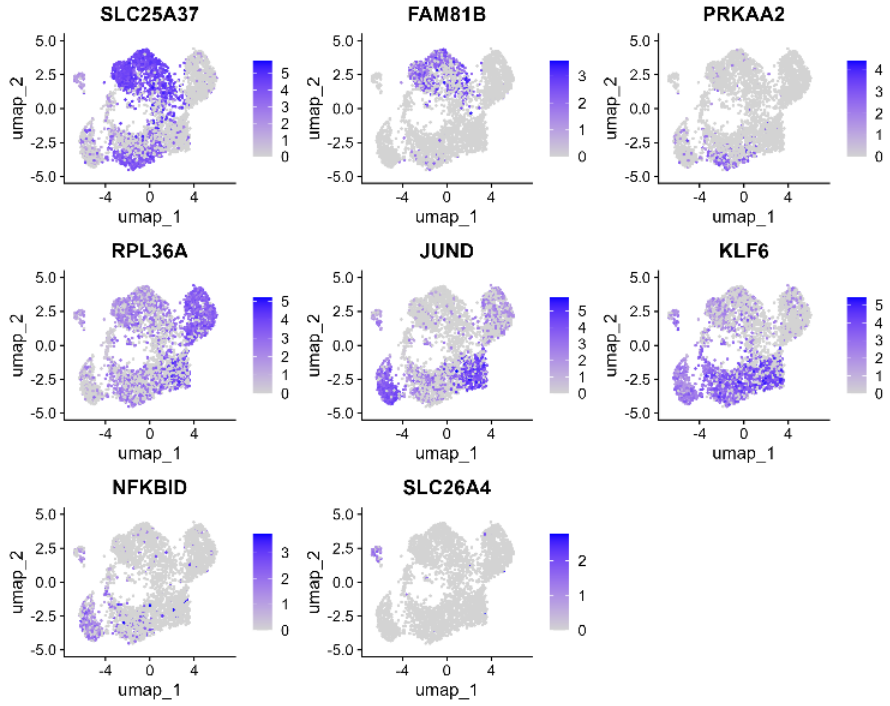

Figure 5. Markers Expressed Genes in B Cell Subclusters. FAM81B for cluster 0, PRKAA2 for cluster 1, RPL36A for cluster 2, NFKBID for cluster 4, SLC26A4 for cluster 5.

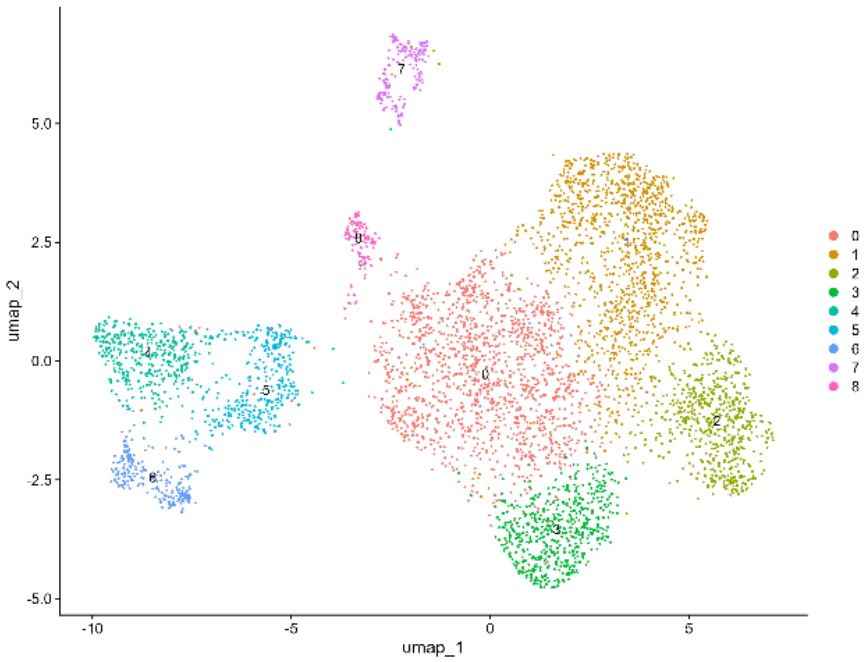

Figure 6. UMAP of CD4+ T Cell Subclusters

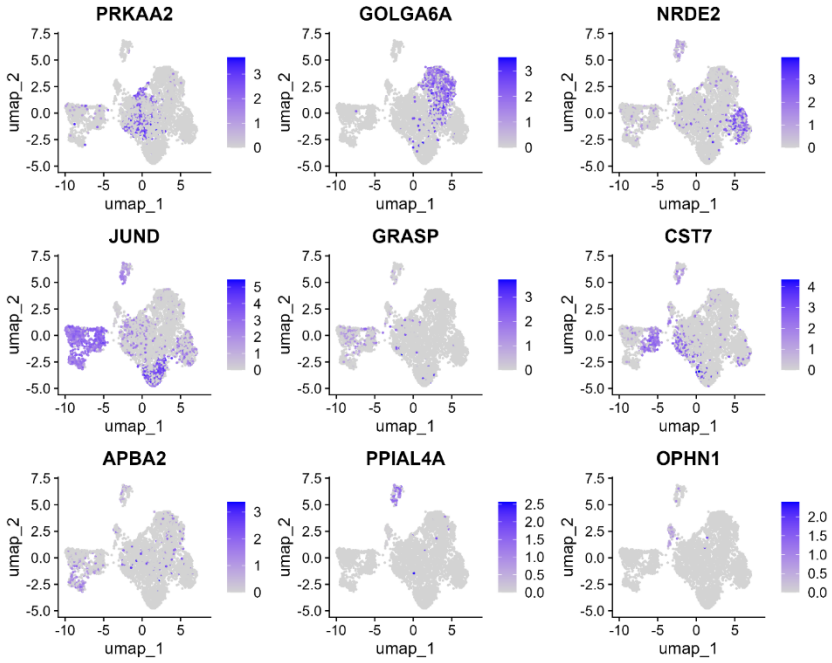

Figure 7. Markers Expressed Genes in CD4+ T Cell Subclusters. PRKAA2 for cluster0, GOLGA6A for cluster 1, NRDE2 for cluster 2, GRASP for cluster4, CST7 for cluster5, APBA2 for cluster 6, PPIAL4A for cluster 7, OPHN1 for cluster 8.

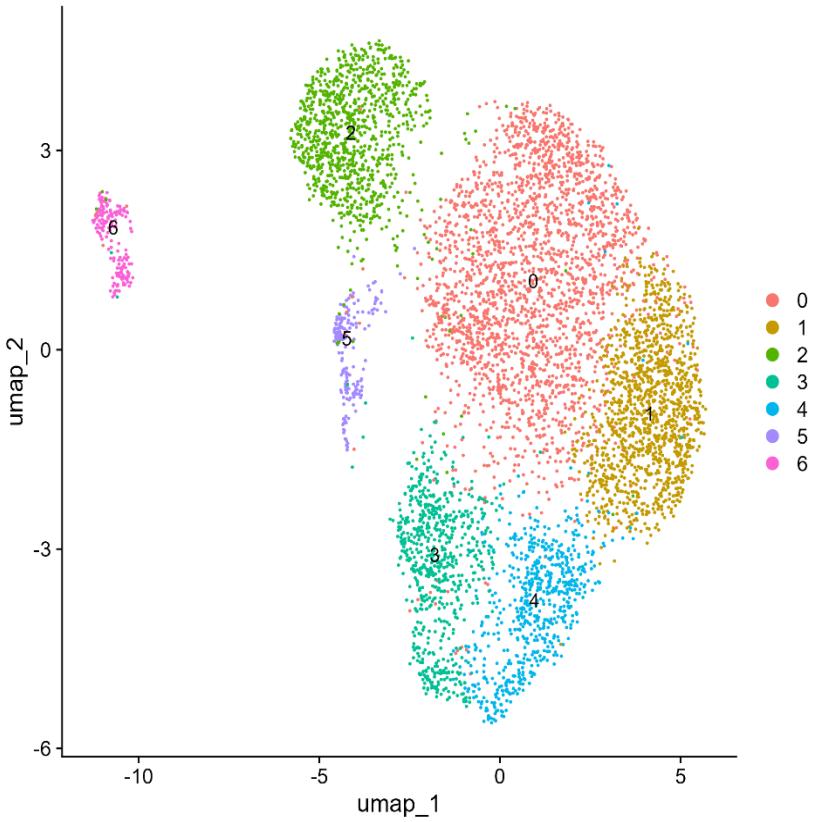

Figure 8. UMAP of cytotoxic T Cell Subclusters

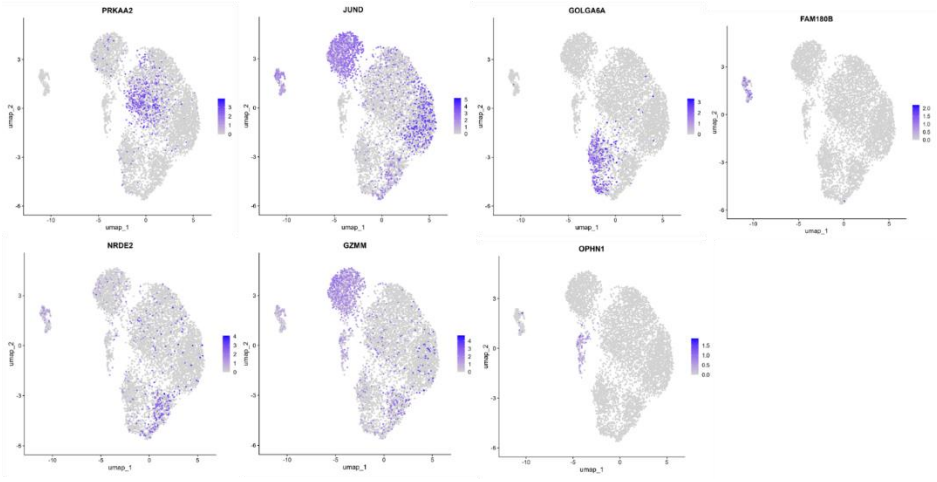

Figure 9. Markers Expressed Genes in cytotoxic T Cell Subclusters. PRKAA2 for cluster 0, intersection of JUND except GZMM for cluster 1, GZMM for cluster 2, GOLGA6A for cluster 3, NRDE2 for cluster 4, OPHN1 for cluster 5, FAM180B for cluster 6.

Discussion

When considering the results of batch effect correction in terms of cell type clustering, the average silhouette width is 0.21281, which indicates moderate clustering quality overall. Specifically, one cluster owns a silhouette width of 0.85764 indicating that certain cell types are well-defined. However, the average silhouette width in batch correction is only 0.06413, indicating that the clusters do not separate very well and some points may be misclassified but it is reasonable as unlike the ability of providing more straightforward paths for analysis and correction in the field of cell type clustering, the correcting the batch effects of Methods/Platforms depends on a series of factors like technical difference, which suggests that we underestimated the complexity of batch effects that can exist even after integration.

Since we mainly focus on the adaptive immunity, so the absence of specific immune cell types such as neutrophils and macrophages, which limits our comprehensiveness of the innate immune landscape represented in the dataset may not undermine our chances to see more information about adaptive immunity.

In terms of heterogeneity, our primary goal to is to determine the existence of these subclusters, which has been achieved. However, the next goal is to know if these subclusters can be matched to some known cell types. For example, in B clusters, we hope to find memory B cells but fail. What we know is some unique functions in most of subclusters, which means that additional analysis and database combination is required.

Another is that without ligand-receptor pairs, the analysis by using CellChat will be constrained as interactions between ligands and receptors are quite important for understanding cell communication and signaling pathways. Therefore, their absence hampers the chance to elucidate potential key pathways that could uncover cellular dynamics within the immune system.

Conclusion

In this study, we performed a comprehensive analysis of the immune cell by using high-dimensional single-cell RNA sequencing data despite the limitations including the relatively low

score of batch effect correction in terms of multiple platforms, the absence of specific immune cell types as well as the lack of data about ligand-receptor interactions for further analysis like CellChat, we successfully characterize several unique immune cell clusters as well as reveal heterogeneity within adaptive immune cells. The analysis highlighted significant clusters, enabling us to understand the roles of various immune cell types.

Reference

Guilliams M, Mildner A, Yona S. Developmental and Functional Heterogeneity of Monocytes. Immunity. 2018 Oct 16;49(4):595-613. doi: 10.1016/j.immuni.2018.10.005.

Inoue T, Kurosaki T. Memory B cells. Nat Rev Immunol. 2024 Jan;24(1):5-17. doi: 10.1038/s41577-023-00897-3.

Kitamura H, Ohno Y, Toyoshima Y, Ohtake J, Homma S, Kawamura H, Takahashi N, Taketomi A. Interleukin-6/STAT3 signaling as a promising target to improve the efficacy of cancer immunotherapy. Cancer Sci. 2017 Oct;108(10):1947-1952. doi: 10.1111/cas.13332.

Li J, Yu C, Ma L, Wang J, Guo G. Comparison of Scanpy-based algorithms to remove the batch effect from single-cell RNA-seq data. Cell Regen. 2020 Jul 6;9(1):10. doi: 10.1186/s13619-020-00041-9.seq batch correction. Nat Methods. 2019 Jan;16(1):43-49. doi: 10.1038/s41592-018-0254-1.

Narasimhan PB, Marcovecchio P, Hamers AAJ, Hedrick CC. Nonclassical Monocytes in Health and Disease. Annu Rev Immunol. 2019 Apr 26;37:439-456. doi: 10.1146/annurev-immunol-042617-053119.

Raskov H, Orhan A, Christensen JP, Gögenur I. Cytotoxic CD8+ T cells in cancer and cancer immunotherapy. Br J Cancer. 2021 Jan;124(2):359-367. doi: 10.1038/s41416-020-01048-4.

Rastogi I, Jeon D, Moseman JE, Muralidhar A, Potluri HK, McNeel DG. Role of B cells as antigen presenting cells. Front Immunol. 2022 Sep 8;13:954936. doi: 10.3389/fimmu.2022.954936.

Shresta S, MacIvor DM, Heusel JW, Russell JH, Ley TJ. Natural killer and lymphokine-activated killer cells require granzyme B for the rapid induction of apoptosis in susceptible target cells. Proc Natl Acad Sci U S A. 1995 Jun 6;92(12):5679-83. doi: 10.1073/pnas.92.12.5679.

Slovin S, Carissimo A, Panariello F, Grimaldi A, Bouché V, Gambardella G, Cacchiarelli D. Single-Cell RNA Sequencing Analysis: A Step-by-Step Overview. Methods Mol Biol. 2021;2284:343-365. doi: 10.1007/978-1-0716-1307-8_19.

Susanto O, Stewart SE, Voskoboinik I, Brasacchio D, Hagn M, Ellis S, Asquith S, Sedelies KA, Bird PI, Waterhouse NJ, Trapani JA. Mouse granzyme A induces a novel death with writhing morphology that is mechanistically distinct from granzyme B-induced apoptosis. Cell Death Differ. 2013 Sep;20(9):1183-93. doi: 10.1038/cdd.2013.59.

Tang C, Li Y, Lin X, Ye J, Li W, He Z, Li F, Cai X. Prolactin increases tumor necrosis factor alpha expression in peripheral CD14 monocytes of patients with rheumatoid arthritis. Cell Immunol. 2014 Jul;290(1):164-8. doi: 10.1016/j.cellimm.2014.06.005.