Comparing Expression Differences Between B-lymphocytes and Granulocytes

Introduction: The human immune system consists of two broad classes of immune cells, granulocytes and lymphocytes. Granulocytes are a nonspecific defense against pathogens such as bacteria and include cells such as eosinophils and basophils. These cells can be identified in a blood sample under a microscope by the presence of distinct granules in their cytoplasm[1]. Lymphocytes, which include B- and T-lymphocytes, are part of the adaptive immune system that enables targeted attack of viruses or infected cells if a specific antigen is present[2]. However, both classes of immune cells are derived from hematopoietic stem cells[1]. This raises the question, what expression differences between these cell populations lead to their distinct functions? To answer this question, analysis was conducted on single-cell RNA (scRNA) sequencing of bone marrow cells from the Tabula Muris Consortium dataset which used mice. While the mouse immune system is not identical to the human immune system given an evolutionary gap of around 65 million years, the system for the differentiation of immune cells is largely similar[4]. Studying the differences in expression between granulocytes and lymphocytes in this dataset could give understanding on how these cell groups differentiate in humans.

Dataset and Methods: The dataset used for this analysis is from the Tabula Muris Consortium. In that analysis, the authors collected 100,000 cells from 20 organ systems of mice. They then conducted scRNA sequencing of those cells and organized the data into count matrices. The paper also includes metadata and annotations files which have additional information about each cell including such as the cell ontology class, tissue, subtissue, plate number, and other information regarding each individual cell[4]. The marrow count matrix was chosen as the source for this analysis to look into the cell populations closer to the developmental stage. Subtissue labels of the cells from the metadata file included in the marrow count matrix included B-cells, T-cells, Granulocytes, and KLS. To simplify the analysis, the Granulocyte subtissue category was compared only against the B-cell subtissue category. Initial data acquisition from the marrow count matrix was conducted using pandas through searching of the plate numbers that corresponded to B-cells and granulocytes. Further filtering and analysis was performed using the scprep library[5].

From the marrow matrix, 1484 B-cells and 1373 granulocytes were found. After the batches were combined, library sizes were plotted. Library size was then filtered to keep cells above the 6th percentile and below the 90th percentile. These cutoffs removed the zero peak as

well as large outliers. Genes with expression in less than 10 cells were filtered out as well, which removed rare genes from further analysis. This combined filtering reduced the total number of cells to 2399 and kept the data for 14313 genes. After this, library size was normalized and mean expression was transformed using a square root transformation. The square root transformation was chosen to make sure events where the count was zero on one side were still recognized in the final analysis. After transformation, principal component analysis (PCA) was performed with 50 components. Following visualization of principal components one (PC1) and two (PC2), the cells were clustered using spectral clustering with three clusters. Three clusters were chosen based on the graphical visualization of PC1 vs PC2, where three distinct clusters were visible in the data. Spectral clustering, which uses linear algebra methods on edge matrices to identify similar points[6], was chosen over K Means and phenograph clustering methods because it visually distinguished the three groups of cells the best. After labelling the cells based on the metadata to determine which cluster isolated either B-cells or granulocytes best, a cluster was found that contained most of the B-cells. Differential expression was then assessed using T-test comparing the expression in the majority B-cell cluster versus the other clusters.

Results:

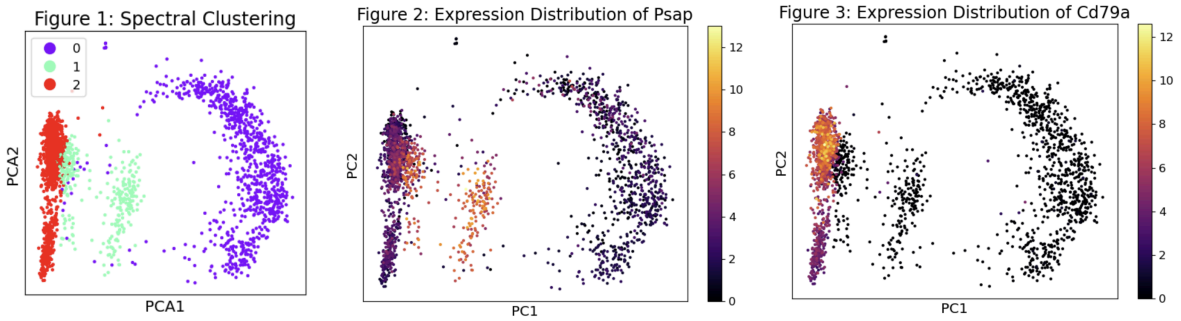

Fig 1. Cluster visualization of the spectral clusters **Fig 2.** Distribution of Psap expression, the gene with the most significant difference in expression between cluster 2 and the other clusters. Brighter color indicates greater expression **Fig 3.** Distribution of Cd79a expression, a previously known marker present on B-cells[7]. Brighter color indicates greater expression.

The main finding found from the data was that the B-cells and granulocytes from the marrow data organized into three subregions on the PC1 vs. PC2 graph as visible in Figures 1-3. While the granulocytes demonstrated a more spread out distribution, the B-cells were localized mostly to cluster 2 as seen in the comparison between Figure 1 and Figure 3. This demonstrates that spectral clustering effectively segregated the B-cells into a single cluster. While the Cd79a distribution looked the most localized toward cluster 2, the gene with the most significant

difference in expression between cluster 2 and the other clusters was Psap whose expression distribution is visualized in Figure 2. This gene showed the highest expression in cluster 1, with lower expression in clusters 0 and 2. The top five genes with the most statistically significant differences between cluster 2 and the other clusters were Psap, Cd79a, Ahnak, Mpeg1, and Pld4.

Discussion: From the statistical analysis comparing the majority B-cell cluster and the other clusters, five genes were found to have the most significant difference. While a difference in a B-cell development gene such as Cd79a is to be expected, expression changes in a couple of the other genes were interesting. Psap stands for Prosaposin, which is a gene that regulates lysosomal function[8]. Expression of this gene was highest in the cluster 1 population, which mostly consisted of granulocytes. This indicates that the formation of the granules in granulocytes may be related to altered lysosome function. Another interesting gene from that list was Ahnak. Ahnak is a gene that is involved in various processes, but a key function is involvement in the inflammatory response[9]. This gene had high expression in cluster 1 as well, which demonstrates that the gene regulation with respect to the immune system's inflammatory response is also different between B-cells and granulocytes.

Limitations and Future Research: One limitation of this analysis is that it uses the metadata plate numbers to identify cell types in the marrow count matrix. This limits the categories of cells that can be isolated from the marrow dataset. An alternative approach to this would be to use the cell ontology field in the annotations file of the dataset to identify cells to compare. This document includes identification of cells in categories such as granulocytopoietic cell and precursor B cell. By studying the differences between those two populations of cells, a clearer picture of the developmental differences between granulocytes and B-lymphocytes could be visible. Another limitation of this analysis is limiting the study to only granulocytes and B-cells. The metadata for the marrow cell population also contained labels for KLS cells and T-cells, while the annotations file also had many more cell ontology categories that could be examined. Future analysis of the differential expression between granulocytes and lymphocytes could compare differences between different cell ontology categories of cells, or include a rerun of the original scRNA sequencing to focus on bone marrow only.

References

1. Görgens, A., Radtke, S., Möllmann, M., Cross, M., Dürig, J., Horn, P. A., & Giebel, B. (2013). Revision of the human hematopoietic tree: granulocyte subtypes derive from distinct hematopoietic lineages. Cell reports, 3(5), 1539–1552. https://doi.org/10.1016/j.celrep.2013.04.025
2. Cano RLE, Lopera HDE. Introduction to T and B lymphocytes. In: Anaya JM, Shoenfeld Y, Rojas-Villarraga A, et al., editors. Autoimmunity: From Bench to Bedside [Internet]. Bogota (Colombia): El Rosario University Press; 2013 Jul 18. Chapter 5. Available from: https://www.ncbi.nlm.nih.gov/books/NBK459471/
3. Tabula Muris Consortium, Overall coordination, Logistical coordination, Organ collection and processing, Library preparation and sequencing, Computational data analysis, Cell type annotation, Writing group, Supplemental text writing group, & Principal investigators (2018). Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. Nature, 562(7727), 367–372. https://doi.org/10.1038/s41586-018-0590-4
4. Mestas, J., & Hughes, C. C. (2004). Of mice and not men: differences between mouse and human immunology. Journal of immunology (Baltimore, Md. : 1950), 172(5), 2731–2738. https://doi.org/10.4049/jimmunol.172.5.2731
5. Krishnaswamy Lab, Yale University. (n.d.). scprep. Read the Docs. https://scprep.readthedocs.io/en/stable/index.html
6. Huang, G. T., Cunningham, K. I., Benos, P. V., & Chennubhotla, C. S. (2013). Spectral clustering strategies for heterogeneous disease expression data. Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing, 212–223.
7. Mason, D. Y., Cordell, J. L., Brown, M. H., Borst, J., Jones, M., Pulford, K., Jaffe, E., Ralfkiaer, E., Dallenbach, F., & Stein, H. (1995). CD79a: a novel marker for B-cell neoplasms in routinely processed tissue samples. Blood, 86(4), 1453–1459.
8. He, Y., Kaya, I., Shariatgorji, R., Lundkvist, J., Wahlberg, L. U., Nilsson, A., Mamula, D., Kehr, J., Zareba-Paslawska, J., Biverstål, H., Chergui, K., Zhang, X., Andren, P. E., & Svenningsson, P. (2023). Prosaposin maintains lipid homeostasis in dopamine neurons and counteracts experimental parkinsonism in rodents. Nature communications, 14(1), 5804. https://doi.org/10.1038/s41467-023-41539-5
9. Zhang, S., Cai, Z., & Li, H. (2023). AHNAKs roles in physiology and malignant tumors. Frontiers in oncology, 13, 1258951. https://doi.org/10.3389/fonc.2023.1258951