scVI enables batch-effect and noise correction and highlights differences in HPSC transcriptional profiles of Aplastic Anemia patients and healthy donors

Yegor Kuznetsov[1*] **Sophia Jannetty**[1*]
[1]University of Washington
yegor@uw.edu
jannetty@uw.edu

Abstract

We should write this abstract paragraph last

1 Introduction

Aplastic anemia (AA) is a condition in which a patient's bone marrow fails to form enough red blood cells, white blood cells, and platelets, resulting in pancytopenia (Young 2018). Pathophysiological mechanisms of this disease include direct damage to bone marrow (commonly from chemotherapy), germ line loss-of-function mutations that interfere with blood-cell precursor DNA repair pathways, and autoimmune attack (Young 2018).

Current options for treating AA fall within three categories, each of which attempts to address pancytopenia in a different way. The first, bone marrow transplantation, aims to replace failing bone marrow with healthy donor marrow. This is a curative treatment, but is limited by the prevalence of graft-versus-host disease and by its reliance on tissue donors (Young 2018). The second, immunosuppression, aims to eliminate immune-mediated AA. This category includes treatment with antilymphocyte globulin (combined with cyclosporine), which has a mild lymphocyte-depleting effect. This treatment leads to improved blood production in about 66% of patients (Bacigalupo 2017), however 30% to 60% of these patients experience relapse that requires years of continued cyclosporine therapy (Scheinberg et al. 2012). Despite the success of immunosuppressant treatments, the mechanisms by which the immune system damages bone marrow remain unknown; the strongest evidence for immune-mediation of aplastic anemia is the effectiveness of immunosuppressant treatments (Young 2018). The third treatment category, stem-cell stimulation, aims to promote stem-cell regeneration within the patient directly. One such treatment involves administrating eltrombopag, a synthetic mimic of thrombopoietin (a hormone that stimulates the production of platelets). This treatment has been shown in limited clinical trials to improve blood production in 80% of treated patients, though the mechanism of this treatment remains unknown (Young 2018).

Despite the severity of this disease, the mechanisms underlying its onset and the treatment options remain obscure due to limitations in experimental techniques. It is thought that T cells may target hematopoietic stem and progenitor cells (HSPCs), which are blood cell precursor cells, in patients with immune-mediated AA (Tonglin et al. 2022). However HSPCs are severely depleated in AA patients, making them a difficult target to study (Zhu et al. 2021). Recent studies like Tonglin et al. (2022) and Zhu et al. (2021) have used single-cell RNA-sequencing (scRNA-seq) to perform differential expression analysis between healthy and AA patient HSPCs. These studies have thus far used PCA to reduce dataset dimensionality, followed by graph-based clustering and likelihood ratio tests to identify significantly differentially expressed genes between clusters (Zhu et al. 2021). These studies have successfully identified differentially expressed genes between healthy and AA HSPCs. However, their adjustments for noise and batch effects are limited to their initial dimension

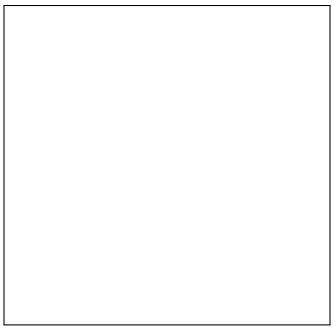

Figure 1: Sample figure caption.

reductions. This initial dimension reduction assumes a generalized linear model is sufficient for accurately mapping onto the low-dimensional manifold underlying the data. This is a particularly problematic assumption in papers like Tonglin et al. (2022) where samples were only taken from four patients (two healthy and two with AA). With such a low n of patients, it would be challenging to convincingly claim differences between AA and healthy clusters are due to disease status as opposed to being due to differences between individuals.

In this project we aimed to explore methods of learning differences between healthy and AA HSPCs using the scRNA-seq data published by Tonglin et al. (2022). We aimed to develop models of both healthy and diseased HSPC transcriptome profiles to identify perturbations capable of moving a diseased HSPC to a healthy cell or vice versa. Our current proposed approach is based on machine learning methods we have learned in previous coursework. We may adjust our methods as we learn more about transcriptomics analysis in upcoming lectures.

1.1 Citations within the text

Of note is the command `\citet`, which produces citations appropriate for use in inline text. For example,

```
\citet{giammarco_transplant_2018} investigated\dots
```

produces

> Giammarco et al. (2018) investigated…

1.2 Figures

1.3 Margins in LaTeX

Most of the margin problems come from figures positioned by hand using `\special` or other commands. We suggest using the command `\includegraphics` from the `graphicx` package. Always specify the figure width as a multiple of the line width as in the example below:

```
\usepackage[pdftex]{graphicx} ...
\includegraphics[width=0.8\linewidth]{myfile.pdf}
```

See Section 4.4 in the graphics bundle documentation (http://mirrors.ctan.org/macros/latex/required/graphics/grfguide.pdf)

A number of width problems arise when LaTeX cannot properly hyphenate a line. Please give LaTeX hyphenation hints using the `\-` command when necessary.

Acknowledgments and Disclosure of Funding

References

References follow the acknowledgments. Use unnumbered first-level heading for the references. Any choice of citation style is acceptable as long as you are consistent. It is permissible to reduce the font size to `small` (9 point) when listing the references. Note that the Reference section does not count towards the page limit.

References

Bacigalupo, Andrea. 2017. "How I treat acquired aplastic anemia | Blood | American Society of Hematology." *Blood* (March 16, 2017). Accessed October 20, 2022. https://doi.org/10.1182/blood-2016-08-693481. https://ashpublications.org/blood/article/129/11/1428/35931/How-I-treat-acquired-aplastic-anemia.

Giammarco, Sabrina, Régis Peffault de Latour, Simona Sica, Carlo Dufour, Gerard Socie, Jakob Passweg, Nicolaus Kröger, et al. 2018. "Transplant outcome for patients with acquired aplastic anemia over the age of 40: has the outcome improved?" *Blood* 131, no. 17 (April 26, 2018): 1989–1992. ISSN: 0006-4971, accessed October 20, 2022. https://doi.org/10.1182/blood-2017-09-807859. https://doi.org/10.1182/blood-2017-09-807859.

Scheinberg, Phillip, Olga Nunez, Barbara Weinstein, Priscila Scheinberg, Colin O. Wu, and Neal S. Young. 2012. "Activity of alemtuzumab monotherapy in treatment-naive, relapsed, and refractory severe acquired aplastic anemia." *Blood* 119, no. 2 (January 12, 2012): 345–354. ISSN: 0006-4971, accessed October 20, 2022. https://doi.org/10.1182/blood-2011-05-352328. https://doi.org/10.1182/blood-2011-05-352328.

Tonglin, Hu, Zhao Yanna, Yu Xiaoling, Gao Ruilan, and Yin Liming. 2022. "Single-Cell RNA-Seq of Bone Marrow Cells in Aplastic Anemia." *Frontiers in Genetics* 12 (January 3, 2022): 745483. ISSN: 1664-8021, accessed October 20, 2022. https://doi.org/10.3389/fgene.2021.745483. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8762313/.

Young, Neal S. 2018. "Aplastic Anemia." Publisher: Massachusetts Medical Society _eprint: https://doi.org/10.1056/NEJMra1413485, *New England Journal of Medicine* 379, no. 17 (October 25, 2018): 1643–1656. ISSN: 0028-4793, accessed October 20, 2022. https://doi.org/10.1056/NEJMra1413485. https://doi.org/10.1056/NEJMra1413485.

Zhu, Caiying, Yu Lian, Chenchen Wang, Peng Wu, Xuan Li, Yan Gao, Sibin Fan, et al. 2021. "Single-cell transcriptomics dissects hematopoietic cell destruction and T-cell engagement in aplastic anemia." *Blood* 138, no. 1 (July 8, 2021): 23–33. ISSN: 0006-4971, accessed October 20, 2022. https://doi.org/10.1182/blood.2020008966. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8349468/.

A Appendix

Optionally include extra information (complete proofs, additional experiments and plots) in the appendix. This section will often be part of the supplemental material.