

Exploration of Immune Cell Type Composition and How It Relates to Biological Age

Master's Thesis in Bioinformatics (30 ECTS)

Sofie Poder Jørgensen, 201905630

Aarhus University



Supervisor: Nicolai Juul Birkebæk
Associate Professor
Department of Molecular Medicine
Aarhus University

Abstract

Biological age is a better indicator of health and life expectancy than chronological age. However, determination of biological age is complicated, as many factors such as environment and lifestyle affect it in different ways. The aim of this study was to explore the correlation between immune cell type composition and age, using scRNA-seq data. Seven datasets with diverse diseases and tissue types, were collected. Findings highlighted the crucial step of normalization for downstream analysis. Notably, despite multiple attempts of advanced normalization techniques, a batch effect persisted in the data, making it necessary to conduct the downstream analysis on each dataset rather than the collection of the data. An age-related decline in naive cytotoxic T cells, was found to be the most significant cell type in multiple datasets. Other cell types, such as NK cells and Plasma cells were also found to be correlated with age. However, the individual datasets showed different results. Plasma cells had the most significant correlation and change in a dataset containing COVID-positive patients, indicating that COVID-19 may disrupt age-related patterns. This study found sex-specific differences in immune cell composition, highlighting the necessity of taking sex into account when investigating immune cell type composition in relation to age. Finally, machine learning models using cell types and ratios between cell types were compared. Findings indicate that interactions between cell types should be considered as the Random Forest model performed better than the Lasso model.

Contents

Abstract	2
1. Introduction	4
1.1 Biological age	4
1.2 The immune system and age-related changes	4
1.3 Single Cell Sequencing	6
1.4 Previous Studies	7
2. Methods	9
2.1 Data	9
2.2 Quality Control	10
2.3 Cell Type Annotation	10
2.4 Normalization and batch correction	10
2.5 UMAPs	11
2.6 Immune Cell Type Composition	11
2.7 Immune Cell Type Ratios	12
2.8 Machine Learning	12
3. Results	14
3.1 Normalization results	14
3.2 Exploration of Immune Cell Types Composition and correlation with age	16
3.1.1 Sex and Age Correlation	17
3.1.2 Inspection of individual datasets	18
3.2 Differences in cell fractions across age thresholds	21
3.2.1 Immune Cell type	21
3.2.2 Immune Cell type Ratios	24
3.3 Prediction of age	28
3.3.1 Regression	28
3.3.2 Classification	29
3.3.3 Regression: Lung Tissue and cancer	30
4. Discussion and Future Works	32
4.1 Normalization challenges	32
4.2 Diversity in immune cell type composition	33
4.3 Age threshold and Wilcoxon Rank Sum Test	34
4.4 Ratios vs individual cell types	35
4.5 Prediction of age	35
4.6 Limitations	36
Conclusion	37
References	38
Supplementary	40