Modeling and Exploring Immune Aging from Single Cell Transcriptomic Data

Master's Thesis in Bioinformatics (30 ECTS)

June 2026

Author:

Elise Ledet Jensen

Student number: 202108599

Supervisor:

Nicolai Juul Birkbak

Cancer Evolution & Immunology Group

Department of Molecular Medicine (MOMA), Aarhus University

Abstract

With advancing age, immune function declines in a process known as immunosenescence, characterized by altered immune cell composition, impaired responsiveness, and chronic inflammation. Although chronological age is a major determinant of immune decline, substantial interindividual variation exists. This variability highlights the need for quantitative measures of immune aging beyond chronological age. In this thesis, I developed cell type specific transcriptomic aging clock models for major immune cell populations within PBMCs, including $CD4^+$ T cells, $CD8^+$ T cells, monocytes, NK cells, and B cells, using scRNA-seq data from healthy individuals in the OneK1K cohort. Models were trained using lasso regression to predict chronological age from gene expression measurements and were subsequently evaluated in the independent Sound Life cohort.

The T cell aging clock models showed the strongest correlations between predicted transcriptomic age and chronological age and generalized well to the Sound Life cohort (r = 0.70–0.81), indicating that the T cell models capture robust age-associated transcriptional signals that are reproducible across independent cohorts. Application of the aging clock models to longitudinal samples from the Sound Life cohort further revealed that predicted transcriptomic age was relatively stable over the 2–3 year study period despite modest temporal fluctuations, with interindividual variability exceeding intraindividual variability, and showed no significant change following seasonal influenza vaccination. Furthermore, CMV+ individuals exhibited substantially higher predicted transcriptomic age in $CD8^+$ T cells than CMV− individuals of the same chronological age, suggesting that CMV infection may accelerate the acquisition of a transcriptomic profile normally associated with immune aging. At the cohort mean age, CMV+ individuals were predicted to be approximately 10 years older than CMV− individuals, an effect that was specific to the $CD8^+$ T cell aging clock.

Together, these findings demonstrate that T cell based transcriptomic aging clocks capture reproducible age-associated transcriptional signatures and can be used to study interindividual variation in immune aging dynamics. The results further identify CMV infection as an important contributor to accelerated transcriptomic aging in $CD8^+$ T cells.

Preface and Acknowledgements

This master's thesis was completed under the supervision of Professor Nicolai Juul Birkbak and PhD student David Martin-Pestana at the Department of Molecular Medicine (MOMA), Aarhus University. The work presented in this thesis was carried out between January and May 2026.

The project focused on (i) developing immune cell type specific aging clocks based on lasso regression models trained on single cell RNA sequencing data, (ii) evaluating the performance, robustness, and biological relevance of these models, and (iii) applying the aging clocks to investigate the dynamics of immune aging.

I would like to express my sincere gratitude to my supervisors, Nicolai Juul Birkbak and David Martin-Pestana, for their guidance, support, and expertise throughout this project. Their insights and feedback have been invaluable in shaping both the direction and quality of this work. Additionally, I would like to thank David for carefully reading the thesis manuscript and providing constructive feedback. I would also like to extend my thanks to the members of the Cancer Evolution and Immunology Group and my colleagues at MOMA for creating such a stimulating and supportive research environment.

Some of the computing for this project was performed on the GenomeDK cluster. I would like to thank GenomeDK and Aarhus University for providing computational resources and support that contributed to this research project.

I would like to thank Nicolai Juul Birkbak for giving me the opportunity to continue in the group as a PhD candidate over the next three years. I am excited to remain part of this inspiring research environment and look forward to continuing to work with everyone in the group and to the challenges and opportunities ahead.

Finally, I would like to thank my family and friends for their unwavering encouragement, patience, and support throughout my studies. I am especially grateful to Julia Karen Demtröder, whose friendship, insightful conversations, and many shared runs have provided both renewed energy and fresh perspectives throughout this journey.

Generative AI use declaration

I have used generative artificial intelligence (GAI) tools to assist in completing this project.

Tools used: ChatGPT by OpenAI (free version; various GPT-5-series models used during the period in which this thesis was completed).

Use cases:

- For programming tasks: I used ChatGPT to assist in writing and debugging R and Python code for the workflow and analyses.
- For feedback on my own text: I used ChatGPT to provide feedback on how to improve the clarity, structure, and flow of specific parts of the thesis.

In all cases, I carefully reviewed and critically evaluated the output generated by GAI tools. The final responsibility for the content, analyses, and conclusions of this thesis remains solely my own.

Abbreviations

BH	Benjamini–Hochberg (procedure for controlling the false discovery rate)
BMI	Body mass index
CMV	Cytomegalovirus
CMV−/+	CMV serostatus, either negative (−) or positive (+)
CRP	C-reactive protein
DEGs	Differentially expressed genes
FDR	False discovery rate
GZMH	Granzyme H
HDL	High-density lipoprotein
ICC	Intraclass correlation coefficient
Lasso	Least absolute shrinkage and selection operator
LDL	Low-density lipoprotein
Limma	Linear models for microarray and RNA-seq data
LM	Linear model
LME	Linear mixed effects model
LRRN3	Leucine rich repeat neuronal 3
MAE	Mean absolute error
MI	Mutual information
NLR	Neutrophil-to-lymphocyte ratio
PBMCs	Peripheral blood mononuclear cells
PCA	Principal component analysis
QC	Quality control
RP	Rank product
scRNA-seq	Single cell RNA sequencing
SD	Standard deviation
Tcm	Central memory T cells
Tem	Effector memory T cells
Temra	Terminally differentiated effector memory T cells re-expressing CD45RA
Trm	Tissue-resident memory T cells
UMAP	Uniform manifold approximation and projection

Table of contents