AARHUS
UNIVERSITY

Detection and Characterization of Clonal Hematopoiesis based on Circulating Free DNA data

Jonas Riber Jørgensen

Main Supervisor
Prof. Jakob Skou Pedersen

Daily Supervisor
postdoc. Gustav Alexander Poulsgaard

This thesis is submitted for the degree of
Master of Science in Bioinformatics

June 2025
Aarhus University
Department of Bioinformatics

ABSTRACT

Clonal hematopoiesis (CH) is a biological phenomenon characterized by the expansion of hematopoietic stem cell (HSC) clones originating from a single ancestor cell that acquires somatic mutations. Certain mutations can confer a competitive advantage for the affected cell, which leads to a clonal expansion. Although often asymptomatic, CH is increasingly recognized for its association with aging, hematological cancers, cardiovascular disease, and lowered mortality. Moreover, its presence can confound the interpretation of circulating free tumor DNA (ctDNA) in blood samples, posing a challenge for cancer detection using circulating free DNA (cfDNA).

This thesis aims to detect and characterize CH in a cohort of 146 individuals previously treated for stage III colorectal cancer (CRC), using whole-genome-sequenced cfDNA samples collected at multiple time points. The primary objective is to identify individuals exhibiting signs of CH, annotate and characterize them. To achieve this, a custom pipeline was developed to identify somatic variants within a curated set of genes associated with CH, and apply custom filters for accurate variant calling. The annotated patients were then analyzed to explore patient-level variation and used to perform an initial characterization via mutational spectra, based on the substitution types. From these methods, 312 CH variants were collected. And with a selection process, 3 patients exhibited a more significant signal of the CH variants.

Although the methods remain under development and results are limited by the project's time constraints, the findings provide a foundation for future refinement of CH detection in cfDNA. And eventually may lead to a broader characterization and a somatic mutation model for CH variant calling, potentially aiding in the distinction between CH and cancer variants, in a cancer detection setting.

ACKNOWLEDGMENTS

I want to express my deepest gratitude to my supervisor, Prof. Jakob Skou Pedersen, and my daily supervisor, postdoc. Gustav Alexander Poulsgaard, for their guidance and expertise throughout the course of this thesis. Their insights have played a crucial role in shaping this work's direction and quality.

I am also grateful to the research department at MOMA for providing the setting and environment, as well as for providing access to the data used within this thesis.

To my fellow students and friends, thank you for the shared frustrations, whiteboard sessions, and moments of humor that made this journey more bearable and enjoyable.

Lastly, I am forever grateful to my family for their support and belief in me. Without them, this achievement would not have been possible.

Jonas Riber Jørgensen, 2025

Contents

List of Figures

ABBREVIATIONS

BH Benjamini-Hochberg.

cfDNA circulating free DNA.
CH clonal hematopoiesis.
CHIP clonal hematopoiesis of intermediate potential.
CRC colorectal cancer.
ctDNA circulating free tumor DNA.

DGD deep generative decoder.

FDR false discovery rate.

HSC hematopoietic stem cell.

IGV Integrative Genomics Viewer.
indel insertion and deletion.

NGS next-generation sequencing.

ORF open reading frame.

PBMC peripheral blood mononuclear cell.
PCR polymerase chain reaction.
PoN panel of normals.

SBS sequencing-by-synthesis.

VAF variant allele frequency.

WGS whole-genome sequencing.