AARHUS UNIVERSITY

regionmidtjylland

Master's Thesis in Bioinformatics

Ribosomal DNA Copy Number Variation and its Multi-Omic Implications in Cancer

Author:

Diego Núñez Martínez

Student Number: 202402136

Supervisor:

Nicolai J. Birkbak

Professor

Department of Molecular Medicine

Cancer Evolution & Immunology Group

June, 2026

Acknowledgements

There are many people that I would like to thank for these last two years of my Master's, but, as I only have one page for this, I will try to keep it brief.

First of all, I would like to thank Nicolai J. Birkbak for giving me the opportunity to work on a such interesting topic for my Master's Thesis, as well as for his guidance and support during these last five months.

Second, I would like to thank the Cancer Evolution and Immunology group for welcoming me for my Master's Thesis and for giving me the opportunity to work with such an amazing team. Furthermore, I would like to extend my gratitude to my colleagues in the Department of Molecular Medicine and the Bioinformatics Research Centre, as working in such an inspiring environment has deeply reinforced my passion for research.

On a more personal note, I would like to thank all of the new friends and colleagues that I have met during my Master's studies. They have filled these two years with wonderful experiences that I'll treasure for the rest of my life.

Last but not least, I would like to thank all of the people in my hometown who are dearest to me: my parents, my sister and all of my friends. Moving to a new country is always a challenge, but their support from day one made the transition much easier. I really feel that these last two years would not have existed without their encouragement to take this leap, and I am incredibly grateful for their love and support.

Generative AI Use Declaration

I have used generative artificial intelligence (GAI) tools in the completion of this project.

Tools Used:

- Perplexity (version: Sonar 1) by Perplexity AI.
- Gemini (version: 3.5) by Google.

Use Cases:

- **Code generation and correction:** GAI was used for writing, debugging and refining Python and R code related to data processing and analysis, like drawing plots or calculating specific estimations.
- **Obtaining feedback on written content:** GAI was used to give feedback on the grammar, coherence and clarity of the text that I wrote, so that the quality and structure of my writing could be improved.
- **Understanding a topic better and aiding in reading process:** GAI was used for understanding complex topics of the project and for going more in depth in the search of relevant references.

In all cases, the output generated by these GAI tools served as a starting point or reference, as it was subsequently reviewed, evaluated and modified to ensure correctness, relevance and alignment with the project's objectives. The final responsibility for the content, analysis and conclusions of this thesis remains my own.

Abstract

Cancer is a disease driven by genomic structural alterations that reprogram cellular behaviour and genomic architecture. While ribosomal DNA (rDNA) arrays are critical for maintaining genomic stability and driving ribosome biogenesis, the functional impact of somatic rDNA Copy Number (CN) fluctuations in carcinogenesis and its implications in the tumor immune microenvironment remain uncharacterized.

To address this gap, this thesis developed a bioinformatic pipeline to estimate rDNA CN from germline and tumor whole-genome sequencing data across six distinct tumor cohorts. By tracking somatic copy number alterations driven by tumor proliferation, this analytical framework integrated genomic, transcriptomic and clinical data to evaluate the metabolic and immune consequences of rDNA instability on the tumor microenvironment. Multivariate Cox Proportional Hazards models demonstrated that 45S rDNA CN predicts lineage-dependent patient survival phenotypes and immune microenvironment variations. Furthermore, this analysis also revealed that certain 45S CN somatic fluctuation trajectories generate metabolic upregulations that cause replication stress, inflammatory cascades and immune responses, particularly within skin cutaneous melanoma and luminal A breast cancer.

This project establishes baseline 45S CN and its somatic alterations as significant factors that influence the transcriptomic, immunogenic and clinical aspects of tumor proliferation. Although the application of the pipeline revealed tissue-specific heterogeneity and resolution limitations, these findings demonstrate the potential of rDNA copy number as a key component of pan-cancer proliferation, immunogenicity and patient survivability.

Abbreviations

- ASCAT: Allele-Specific Copy Number Analysis of Tumors
- BAM: Binary Alignment Map
- BCR: B-Cell Receptor
- BED: Browser Extensible Data
- BLCA: Bladder Urothelial Carcinoma
- BRCA: Breast Invasive Carcinoma
- BRD: Background Read Depth
- BWA-MEM: Burrows-Wheeler Aligner – Maximal Exact Match
- CN: Copy Number
- CIN: Chromosomal Instability
- DSB: Double Strand Break
- EVT4: ETS Variant Transcription Factor 4
- FASTA: FAST-All
- FDR: False Discovery Rate
- GBM: Glioblastoma Multiforme
- GC: Guanine-Cytosine
- GDC: Genomic Data Commons
- GFF: General Feature Format
- GISTIC: Genomic Identification of Significant Targets in Cancer
- GSTM1: Glutathione S-Transferase Mu1
- GSVA: Gene Set Variation Analysis
- HR: Hazard Ratio
- IGH: Immunoglobulin Heavy Locus
- IGS: Intergenic Spacer
- MAPK: Mitogen-Activated Protein Kinase (pathway)
- MSigDB: Molecular Signatures Database
- mTOR: Mechanistic Target of Rapamycin (pathway)
- NAHR: Non-Allelic Homologous Recombination
- NCBI: National Centre for Biotechnology Information
- NCI: National Cancer Institute
- NHEJ: Non-Homologous End Joining
- NHGRI: National Human Genome Research Centre
- NK: Natural Killer (cells)
- OV: Ovarian Serous Cystadenocarcinoma
- QC: Quality Filtering
- rDNA: ribosomal Deoxyribonucleic Acid
- RNA: Ribonucleic Acid
- RNA-seq: RNA sequencing

- RP: Ribosomal Protein
- SCNA: Somatic Copy Number Alteration
- scRNA-seq: Single-Cell RNA Sequencing
- SKCM: Skin Cutaneous Melanoma
- SnoRNA: Small nucleolar RNA
- SnoRNP: Small nucleolar Ribonucleoprotein
- SNP: Single Nucleoid Polymorphism
- TCGA: The Cancer Genome Atlas
- TCR: T-Cell Receptor
- TGCT: Testicular Germ Cell Tumor
- TIME: Tumor Immune Micro-Environment
- TLS: Tertiary Lymphoid Structure
- TP: Tumor Protein
- wGII: weighted Genomic Instability Index
- WGS: Whole Genome Sequencing

Contents: