

Complex rearrangements and extrachromosomal circular DNA in Colorectal Cancer

Master's Thesis in Bioinformatics (30 ECTS)

Elías Calandra Jiménez, Studienr: 202303411

Supervisors:

Claus Lindbjerg Andersen, Professor Department of Molecular Medicine, Aarhus University

Nicolai Juul Birkbak, Professor Department of Molecular Medicine, Aarhus University

ABSTRACT

Genomic instability in cancer often manifests as structural variants (SVs) and extrachromosomal circular DNAs (ecDNAs), which can drive oncogene amplification and therapeutic resistance. In this study three SV callers (SvABA, Manta, Delly), the genome-graph tool JaBbA, and AmpliconArchitect (AA) were benchmarked on the NCI-H2009 cell line sequenced by two library preparations. Caller concordance, karyotype support, and ecDNA detection were assessed. The tools were then applied to ten colorectal cancer (CRC) tumors. A difference in the number of structural variants called was detected between the two sequencing strategies, while AmpliconArchitect and JaBbA were not highly affected by the sequencing. In CRC, unstable tumors harbored more SVs (mean 432 vs. 198, $p=0.06$), and one contained a ecDNA on 6p21 amplifying several oncogenes. Recurrent deletions were found disrupting the gene *PRKN* in 5 of 10 tumors. Overall, integrating SV calling with JaBbA and AA provides a great framework for complex rearrangements and ecDNA detection for short-read WGS, although long-read sequencing and empirical validations would help in the ecDNA reconstruction.

ABBREVIATIONS

AA: AmpliconArchitect

CRC: Colorectal cancer

ecDNA: Extrachromosomal DNA/Extrachromosomal circular DNA

dm: double minute/extrachromosomal DNA

BFB: Breakage-fusion-bridge

SVs: Structural Variants

SGA: String Graph Assembly

LOESS: locally weighted regression regression

INDEX

1. INTRODUCTION.....	4
1.1. The human genome.....	4
Variation in the genome.....	4
Source of variation in the genome.....	4
1.2. Genomic instability and tumorigenesis.....	4
Genomic instability in cancer.....	5
1.3. Structural variants in the genome.....	5
Leveraging genomic instability and structural variants.....	6
1.4. Extrachromosomal circular DNA.....	7
Challenges when identifying ecDNAs and complex rearrangements.....	8
2. AIMS.....	10
3. PROJECT STRATEGY.....	11
4. METHODS.....	12
4.1. Data.....	12
NCI-H2009 cell line.....	12
CRC tumor samples.....	13
4.2. Structural variant calling.....	14
Structural variant analysis in colorectal cancer tumor samples.....	15
4.3. Analysis of complex structural rearrangements.....	16
4.4. Analysis of ecDNAs.....	17
4.5. Assessment of the tools performance.....	17
Structural variants assessment.....	17
Assessment of the complex rearrangements called by JaBbA.....	18
Assessment of the amplicons identified by AmpliconArchitect.....	18
5. RESULTS.....	19
5.1. Assessment of the tools to identify SVs, complex rearrangements and ecDNAs.....	19
Structural variants in NCI-H2009.....	19
Focally amplified regions in NCI-H2009.....	21
Complex rearrangements in NCI-H2009 by JaBbA.....	23
5.2. Complex rearrangements and ecDNAs in colorectal cancer tumors.....	26
Structural variants in colorectal cancer tumors.....	26
Focally amplified regions in colorectal cancer tumors.....	26
Complex rearrangements in colorectal cancer tumors by JaBbA.....	28
6. DISCUSSION.....	30
7. CONCLUSIONS.....	35
8. FUTURE PERSPECTIVES.....	36
9. REFERENCES.....	37
10. SUPPLEMENTARY MATERIAL.....	41