

AmpliconFinder: A Python-Based Bioinformatics Pipeline for Early-Stage Cancer Diagnosis

Atia B. Amin

Department of Human Genetics
McGill University, Canada



DNA Amplicons in Cancer Cells

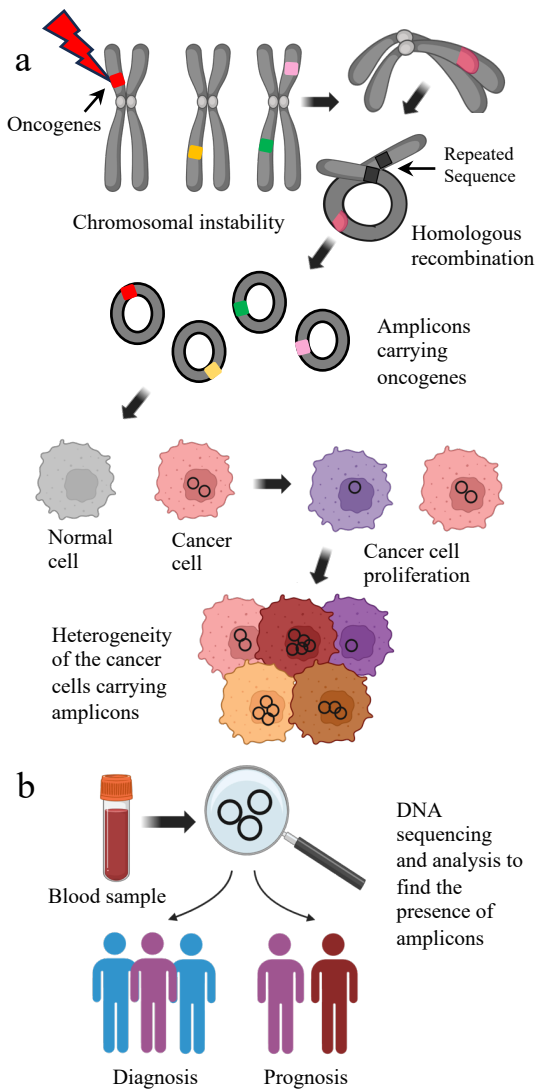


Figure 1: Extrachromosomal circular DNA amplicon generation and proliferation in cancer cells. **a)** Amplicons are formed via recombination due to chromosomal instability. **b)** These amplicons can be used as a biomarker in cancer diagnosis, prognosis, and treatment.

This research is supported by McGill Applied Genomics Innovation Core (MAGIC), Vanier Canada Graduate Scholarships (Vanier CGS) awarded by the Canadian Institutes of Health Research (CIHR).

Oxford Nanopore Technology (ONT) Sequencing of the Amplicons

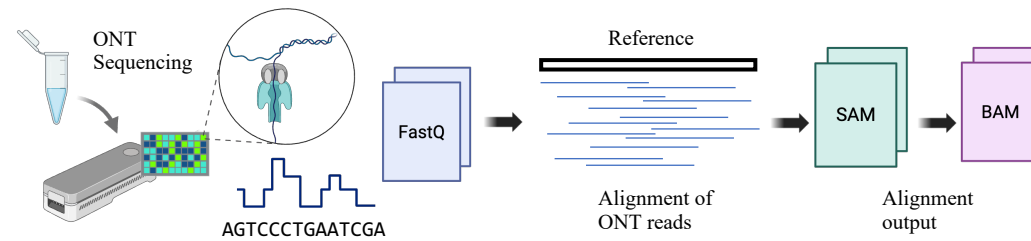
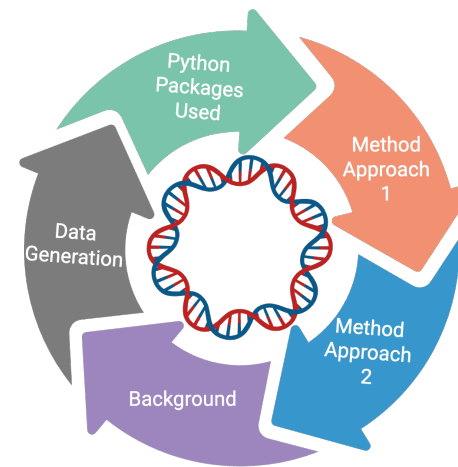


Figure 2: ONT long read sequencing of the samples and data generation. ONT reads are mapped to the reference genome. Alignment outputs are converted to BAM files.

Python Packages Used

- BioPython (Key modules used: SeqIO, SeqRecord, Seq)
- Seaborn (Used for creating statistical plots such as histograms, violin plots)
- Pandas (Used for data analysis)
- Matplotlib (pyplot was used for visualizing genome-wide coverage)
- Numpy (For data analysis)
- Pysam (For coverage calculation)
- Bioinfokit (Used for drawing Manhattan plots)



AmpliconFinder Pipeline

Targeted Approach

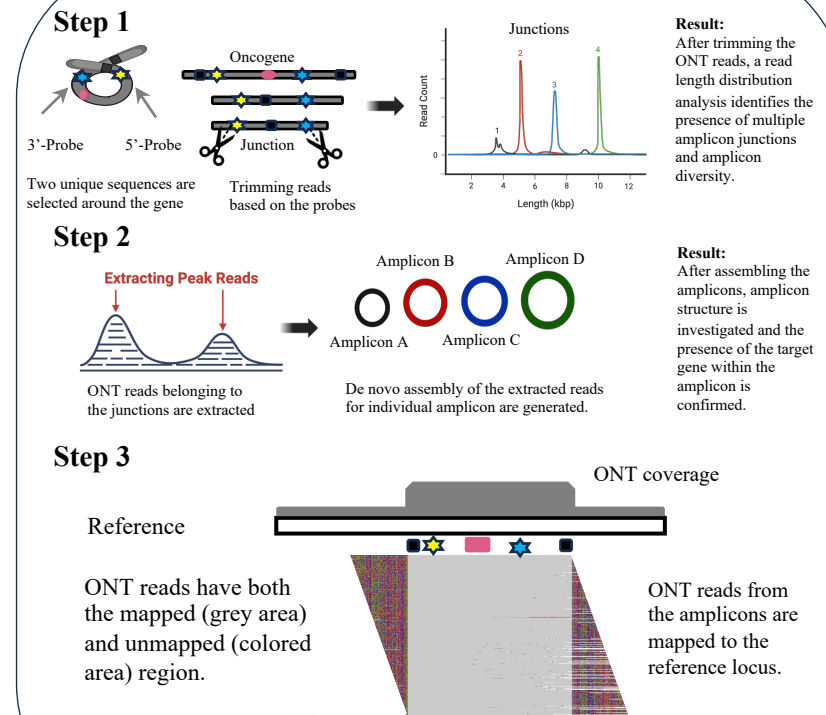


Figure 4: A probe-based amplicon searching method.

Whole Genome Based Approach

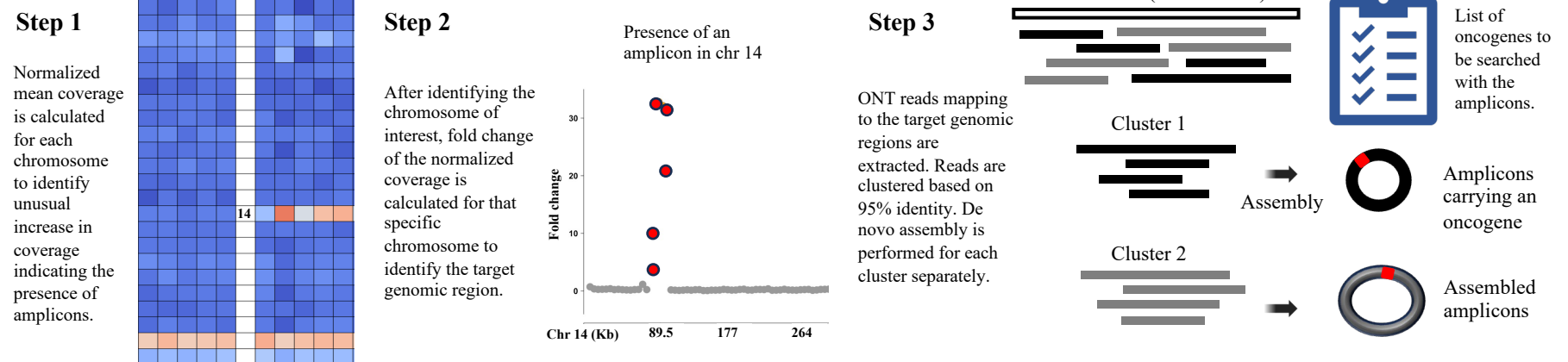


Figure 3: A whole-genome-based amplicon searching method allows for the identification of novel amplicons across all chromosomes.