Detection of circulating tumor DNA using structural variants

Master's Thesis in Bioinformatics

Candidate: Andrea Zauli

Supervisors: Lasse Marrety and Soren Besenbacher

Study program: Bioinformatics

Affiliation: Bioinformatics research center

Place: Aarhus university, June 2021

Target of the thesis

Standard approach in cancer screening fail to full fill the necessary characteristics to tackle the problem of tumor detection: the available biomarkers overlap with other disease or even physiological condition, and similar issues apply to imaging technique. Biopsy is the only technique capable to be definitive, but it still high invasive.

cfDNA has proven to be a suitable alternative, allowing the use of liquid samples directly linked to the intimate nature of the disease: ctDNA, the portion derived from cancer tissue, show significant level of correlation with cancer volume and metabolism and provide the possibility to inspect directly the alterations caused by the disease at a genomic level.

In the last two decades, structural variants emerged as a source of genetic variability as much as single nucleotide polymorphisms. They can actively drive the development of tumor and anomalous patterns of structural variations have been found to strongly associate with the disease.

High coverage NGS procedure to analyze cfDNA show high level of sensibility and accuracy in detecting the presence of ctDNA in the form of abnormal levels of structural variants. However, the high costs associated to the high coverages limit their employment at a population scale due to the economic burden.

The target of my thesis was analyze data from shallow WGS performed on cfDNA samples from groups of cancer cases and healthy controls. The low coverage (0.1-1.0x) don't allow the detection of structural variants. However, a possibility is to search for signs of structural variants in the form of anomalous patterns between read-pairs. With raised levels of ctDNA in cancer samples, the amount of evidence supporting S.V. is expected to increase, and could work as an indicator for the presence of the disease. Furthermore, signs of structural variants can arise also in case of mapping issues or for the systematic presence of germline S.V. in the population. Attempts to mitigate this problem were implemented in the form of filters based on the frequencies of the collected evidence.