Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2021-2022

Personal Information

Supervisor	Anthony Mathelier and Vipin Kumar
Email	anthony.mathelier@ncmm.uio.no
Institution	Centre for Molecular Medicine Norway, University of Oslo
Website	https://mathelierlab.com
Group	Computational Biology & Gene Regulation / Mathelier group

Project

Computational genomics

Project Title:

Identification of positively selected breast cancer somatic variants in 3D

Keywords:

Cancer mutation; positive selection; non-coding mutation

Summary:

Context Most cancer alterations occur in the noncoding portion of the human genome, which contains important regulatory DNA segments acting as genetic switches to ensure gene expression occurs at correct times and intensities in correct tissues. However, the identification of critical noncoding cancer drivers has been hampered by the lack of accurate mapping and functional characterization of these DNA segments. Available methods that aim to detect noncoding cis-regulatory cancer driver variants in clinical data rely on a narrow representation of genomic disruption exclusively considering hotspots of recurring mutations in an abstract 1D description of the genome. These hotspots are likely cancer-drivers as they represent signals of positive selection in tumor mutations. Unfortunately, such 1D-centric approach omits mutations affecting DNA regions whose combined enrichment occurs in 3D notably through multiple enhancer-promoter coupling. Strategy More specifically, this project looks for enrichment across patients of noncoding variants enriched within 3D chromatin aggregates. To locate these enrichments, we will take advantage of the genome-wide 3D description produced by HiC to capture mutations whose aggregate effect in 3D gets scattered along the conventional 1D representation of the genome. We will first chart the tridimensional organisation of chromosomes for “normal” breast tissue cells using Hi-C data from the HMEC cells. This conformation will be used to derive a reference architecture that will contextualise the identification of driver mutations in breast cancer patients. The accurate reconstitution of chromatin 3D aggregates will be pivotal to capture these mutation patterns. To achieve this we will be using a state-of-the-art clustering method developed in-house that dynamically adjusts the resolution of HiC data to faithfully reflect chromosome architecture at all its scales. We will then use statistical models to evaluate how the spatial proximity between mutation events within the 3D chromatin aggregates coincides with coordinated mutation patterns among these mutations. The innovative part of this project is to join structural and functional characterisation of chromatin aggregates to pinpoint driver mutations. Finally, we will assess how the 3D-clustered mutations coincide with the dysregulation of their potential target genes in patients. Expected Outcome This project will contribute to the characterisation of noncoding alterations by examining the interplay between chromosome architecture and cancer associated variants to produce a more realistic contextualisation for the interpretation of noncoding alterations than current approaches. This multi-omics strategy will culminate with the development of computational tools identifying cancer-driving DNA alterations enriched in specific 3D chromatin aggregates.

Expected skills::

Python, R, or bash

Possibility of funding::

To be discussed

Possible continuity with PhD: :

To be discussed