Universitat Pompeu Fabra Barcelona | Master in Bioinformatics for Health Sciences

Master project 2024-2025

Personal Information

Supervisor	Maria Colome Tatche
Email	maria.colome@bmc.med.lmu.de
Institution	Institute of Computational Biology, Helmholtz Center Munich and Faculty of Medicine, Ludwig Maximilian University Munich (Germany)
Website	https://www.physiolchemie.abi.med.uni-muenchen.de/research/maria_colome_tatche/index.html
Group	Computational Epigenomics

Project

Computational genomics

Project Title:

Calling copy number variations with single-cell resolution

Keywords:

single-cell ATAC-seq, single-cell RNA-seq, copy number variations, binary segmentation, cancer

Summary:

Aneuploidy is defined as the state where a cell has an abnormal number of chromosomes. More generally, copy number variations (CNVs), also called copy number alterations (CNAs), describe the deletion and duplication events that affect a given number of base pairs in the genome. It is possible to leverage the read coverage information from single-cell sequencing experiments to determine the most likely copy number state of every cell, genome-wide. In our group, we have so far succeeded at doing that reliably from single cell DNA sequencing data as well as from single-cell ATAC-seq data [1,2]. Recently, a new single-cell sequencing technique has been developed, where for every single cell both the RNA and the open chromatin (i.e. ATAC-seq) are measured (multiomic measurements) [3,4]. This allows to combine both layers of information to study the functional state of the cell. For CNV calling, it should be possible to combine both layers of information for every individual cell to determine the most likely copy number state, more reliably than with only one layer of information. In this project, we will develop methods that call CNVs from single-cell multiomic data. These methods will take into account the complementary information from scRNA and scATAC-seq measurements to construct the most likely CNVs in the individual single cells. Apart from read coverage information, we will also take into account allelic imbalances resulting from both gain and loss of genomic regions, which can be identified in both RNA and ATAC sequencing data. This enables the calling of allele-specific Copy Number Variations (CNVs) and copy-neutral Loss of Heterozygosity (LOH). However, these features pose a greater challenge for detection in scATAC-seq data compared to scRNA-seq, primarily due to the inherent data sparsity in scATAC-seq. In this project, we will combine the allelic information per gene from the scRNA-seq data with the count data information from the scATAC-seq to call CNVs. For that, we will use binary segmentation approaches. Binary segmentation is a methodology employed for detecting change points in signals, aiming to identify positions where there are shifts in data distribution. In the context of CNV calling, the underlying assumption is that the distribution of reads mapping per genomic segment, as well as the allelic imbalances, vary across gained, lost, and normal copy number regions. The objective is to pinpoint locations in the genome where these distribution changes occur. This is achieved by scanning the genome of each individual cell per chromosome, and calculating a distance between the read and allelic distributions on the left and right sides. We will study what is the best multivariate distance to use to maximize the detection of the breakpoints. In this project, we aim to capitalize on the strengths of both scRNA-seq and scATAC-seq data modalities to develop innovative algorithms for more accurate single-cell CNV calling and to explore the functional implications of CNVs on cellular fitness.

References:

[1] B. Bakker, A. Taudt, M.E. Belderbos, D. Porubsky, D.C.J. Spierings, T.V. de Jong, N. Halsema, H.G. Kazemier, K. Hoekstra-Wakker, A. Bradley, E.S.J.M. de Bont, A. van den Berg, V. Guryev, P.M. Lansdorp, M. Colomé-Tatché, F. Foijer. Single cell sequencing reveals karyotype heterogeneity in murine and human malignancies. Genome Biology 17:115 (2016). [2] A. Ramakrishnan, A. Symeonidi, P. Hanel, K. T. Schmid, M. L. Richter, M. Schubert, M. Colomé-Tatché. epiAneufinder identifies copy number alterations from single-cell ATAC-seq data. Nat. Commun. 14, 5846 (2023). [3] A. Ma, B. Zhang, et al. Chromatin Potential Identified by Shared Single-Cell Profiling of RNA and Chromatin. Cell, 183, Issue 4 (2020). [4] AE. Trevino, F. Müller, et al.

Chromatin and gene-regulatory dynamics of the developing human cerebral cortex at single-cell resolution. Cell. 184(19):5053-5069.e23 (2021).

Expected skills:

Good programming skills in R and Python, knowledge of single-cell data analysis

Possibility of funding:

No

Possible continuity with PhD:

To be discussed

Comments:

Please see our recent publication "EpiAneufinder: identifying copy number variations from single-cell ATAC-seq data" in Nature Communications 2023 (Nat Commun 14, 5846 (2023))