Universitat Pompeu Fabra Barcelona

Master in Bioinformatics for Health Sciences

Master project 2024-2025

Personal Information

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

Project

Computational genomics

Project Title:

Mapping and characterizing cis-regulatory activities in breast cancer at single-cell resolution

Keywords:

single-cell, cancer genomics, gene regulation, deep learning

Summary:

We observe different clinical behaviors between breast cancer patients in response to the same treatment regime. Hence, it is critical to stratify patients to provide them with optimal personalized clinical treatment options (1). Current strategies rely on the expression of gene panels for patient stratification (1). As cancer is a disease of disrupted regulation, it is critical to characterize how the gene regulatory program is (re-)wired in cancer cells. Accordingly, studies recurrently showed the importance of cancer alterations occurring outside of genes, specifically in regulatory regions that act as gradual regulatory switches to express genes at the correct time, in the correct tissue, and at the correct intensity (2). Unfortunately, these regulatory regions have been relatively understudied, particularly in the context of cancer patients. Our goal is to characterize at single-cell resolution regulatory regions which may explain sensitivity or resistance to therapy in breast cancer patients. Single cell methods are perceived as 'leading a revolution in our ability to systematically dissect intratumor heterogeneity' (3). Single cell technologies have led to outstanding research explaining both the phenotype of the cell types found in tumors (4) and the mechanisms of action of therapies (5). During the last two years, we have generated single-cell RNA sequencing data for breast tumors from patients involved in clinical trial(s) in Oslo to assess response to neoadjuvant therapy (6). We have been able to identify the main cell types found in breast tumors by relying on gene expression patterns and to pinpoint at the cell types sensitive or resistant to the given therapies. In this project, the candidate will leverage these data to study the activity of transcribed cis-regulatory regions (tCRRs) in cancer cells and cells from the tumor micro-environment. Specifically, the candidate will • make themselves familiar with the SCAFE tool to to generate a tCRR-by-cell count matrix for downstream analyses (7). • use a deep learning approach based on convolutional neural networks to predict single-cell tCRR activities across samples directly from their DNA sequences. Such deep learning frameworks have successfully deciphered the cis-regulatory motif syntax from ATAC-seq or CAGE data (8–10). The selected candidate will use and adapt the scBasset framework (10), initially developed for scATAC-seq data, to automatically identify both the TF binding profiles/motifs and their combinatorics acting upon the tCRRs to control their activities. • investigate the longitudinal patterns of activity of the tCRRs across patients during the course of neoadjuvant therapy and how changes in tCRRs activity or pre-existing tCRRs patterns correlate with response to treatment and the sensitive or resistant malignant phenotypes. As opposed to approaches solely relying on gene expression, our approach will assess the activity of transcriptional regulatory regions at single-cell resolution, thus providing an innovative concept to study the impact of neoadjuvant therapy. Moreover, our approach will offer nucleotide-resolution feature discovery to shed light on the TF binding code at tCRRs specific to the breast cancer cells.

References:

1. Russnes,H.G., Lingjærde,O.C., Børresen-Dale,A.-L. and Caldas,C. (2017) Breast Cancer Molecular Stratification: From Intrinsic Subtypes to Integrative Clusters. The American Journal of Pathology, 187, 2152–2162. 2. Cuykendall,T.N., Rubin,M.A. and Khurana,E. (2017) Non-coding genetic variation in cancer. Current Opinion in Systems Biology, 1, 9–15. 3. Gonzalez Castro,L.N., Tirosh,I. and Suvà,M.L. (2021) Decoding Cancer Biology One Cell at a Time. Cancer Discovery, 11, 960–970. 4. Qian,J., Olbrecht,S., Boeckx,B., Vos,H., Laoui,D., Etlioglu,E., Wauters,E., Pomella,V., Verbandt,S., Busschaert,P., et al. (2020) A pan-cancer blueprint of

the heterogeneous tumor microenvironment revealed by single-cell profiling. Cell Res, 30, 745–762. 5. Schmidt,F., Marx,A., Baumgarten,N., Hebel,M., Wegner,M., Kaulich,M., Leisegang,M.S., Brandes,R.P., Göke,J., Vreeken,J., et al. (2021) Integrative analysis of epigenetics data identifies gene-specific regulatory elements. Nucleic Acids Research, 49, 10397–10418. 6. Bahrami,N., Sauer,T., Engebretsen,S., Aljabri,B., Bemanian,V., Lindstrøm,J., Lüders,T., Kristensen,V., Lorentzen,A., Loeng,M., et al. (2019) The NEOLETEXE trial: a neoadjuvant cross-over study exploring the lack of cross resistance between aromatase inhibitors. Future Oncol, 15, 3675–3682. 7. SCAFE: a software suite for analysis of transcribed cis-regulatory elements in single cells | Bioinformatics | Oxford Academic. 8. Maslova,A., Ramirez,R.N., Ma,K., Schmutz,H., Wang,C., Fox,C., Ng,B., Benoist,C., Mostafavi,S. and Project,I.G. (2020) Deep learning of immune cell differentiation. PNAS, 117, 25655–25666. 9. Einarsson,H., Salvatore,M., Vaagensø,C., Alcaraz,N., Lange,J.B., Rennie,S. and Andersson,R. (2021) Promoter sequence and architecture determine expression variability and confer robustness to genetic variants Genomics. 10. Yuan,H. and Kelley,D.R. (2022) scBasset: sequence-based modeling of single-cell ATAC-seq using convolutional neural networks. Nat Methods, 10.1038/s41592-022-01562-8.

Expected skills:

experience in programming in Python, basic knowledge of gene regulation, interest in cancer genomics, interest in single cell data analysis

Possibility of funding:

To be discussed

Possible continuity with PhD:

No

Comments:

Possibility of covering housing expenses.