

## Master project 2021-2022

### Personal Information

<b>Supervisor</b>	Anthony Mathelier
<b>Email</b>	anthony.mathelier@ncmm.uio.no
<b>Institution</b>	Centre for Molecular Medicine Norway, University of Oslo
<b>Website</b>	<a href="https://mathelierlab.com">https://mathelierlab.com</a>
<b>Group</b>	Computational Biology & Gene Regulation / Mathelier Group

### Project

## Computational genomics

#### Project Title:

PREDICTING MASTER REGULATORS DISCRIMINATING BREAST CANCER SUBTYPES FROM RNA-SEQ DATA

#### Keywords:

breast cancer; gene regulation; transcription factors; multi-omics

#### Summary:

An internship for a research Master student in the field of Computational Biology / Bioinformatics is available at the Computational Biology & Gene Regulation group, Centre for Molecular Medicine Norway, University of Oslo, led by Anthony Mathelier. The group develops computational methods and tools to analyze the regulation of gene expression and the mechanisms by which it can be disrupted in human diseases such as cancers. See <https://mathelierlab.com/> for further information. The project aims at identifying transcription factors (TFs) that are discriminating breast cancer subtypes. Molecular classifications of breast cancer tumours provide intrinsic subtyping classification in addition to predictive and prognosis value, which can be helpful in assessing treatment options. The identification and understanding of such molecular classifications is paramount as they highlight biological aspects of tumour behaviour, function, and identity. Breast cancers are classically distributed across five intrinsic subtypes (luminal A and B, HER2-enriched, basal-like, and normal-like) based on gene expression profiling of 550 genes, mainly reflecting the estrogen receptor (ER) and HER2 status of the tumours. A 50-gene classifier (PAM50), approved by the US FDA (Prosigna test), was subsequently designed to assign tumours to luminal A, luminal B, HER2-enriched, or basal-like subtypes. This stratification has been useful to assess the likelihood of efficacy from neoadjuvant chemotherapy. As TFs represent special class of proteins that plays a key role in the transcriptional regulation of gene expression machinery, we hypothesize that their expressions and activities are contributing to the critical differences between breast cancer subtypes. While ER+ breast cancers are known to be driven by the TFs ESR1, GATA3, and FOXA1, the main drivers of ER- breast cancers are still unknown. In this project, the selected student will analyze large-scale omics data sets (e.g. from RNA-seq, ATAC-seq, and ChIP-seq) to highlight the TFs specifically regulating the genes differentially expressed between breast cancer subtypes. During the course of the project, the student will use state of the art computational approaches to predict where TFs bind to the DNA and methods to infer master TFs regulating the differentially expressed genes. Moreover, the work will expose the trainee to computational approaches for the management, analyses, and interpretation of large-scale, next generation sequencing data. We seek a highly motivated individual with programming skills, knowledge of computational tools development dedicated to the analysis of high throughput sequencing data. Knowledge in statistical methods and/or a biological background is a plus. We are looking for applicants excited about combining life sciences and computation to analyze gene expression regulation. The candidate will be collaborating with researchers at the Oslo University Hospital, where the group is also affiliated. The successful candidate will be collaborative, independent, with strong enthusiasm for research, and should be fluent in at least one of the following programming languages: Python, R, or bash. Being familiar with gene expression regulation in general and transcription factor binding is an advantage.

#### Expected skills::

Python, R, or bash

#### Possibility of funding::

To be discussed

**Possible continuity with PhD :**

To be discussed

---