



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:

Expanding the maps of direct TF-DNA interactions across species using open-chromatin data

Keywords:

gene regulation, transcription factor binding sites, computational genomics

Summary:

Transcription factors (TFs) are key proteins involved in transcriptional regulation that modulate the rate of gene transcription through their specific binding at transcription factor binding sites (TFBSs). Hence, it is critical to accurately map TF-DNA interactions to understand gene regulation. Our group has been involved in the development and maintenance of key open-access resources and new computational methods to advance the mapping of TF-DNA interactions in genomes. For instance, we maintain the JASPAR (1) and UniBind (2, 3) databases to store high-quality TF binding profiles (JASPAR) and direct TF-DNA interactions (UniBind) across species. While UniBind stores TFBSs predicted from ChIP-seq data, we aim to expand it with the prediction of TFBSs from open chromatin data (Dnase-seq and ATAC-seq). The candidate will participate in this effort by performing a benchmark of existing state-of-the-art method that predict TFBSs from DNase-seq and ATAC-seq data (e.g., HINT-ATAC (4), maxATAC (5), TAMC (6), and TRACE (7)). The results of the benchmark will identify the optimal method to process large-scale datasets of publicly available open chromatin data. The predictions obtained will expand the sets of TFBSs currently stored in UniBind. Finally, the candidate will explore the possibility of extending this project to the analysis of deep learning models applied to ATAC-seq data if time allows. Advantageously, we have expertise in the in depth analyses of high-throughput 'omics data dedicated to the analyses of transcriptional regulation and in the development of internationally renowned resources for the community. It will provide an optimal learning environment to the selected student. During the course of the project, the student will be exposed to software development and computational approaches for the management, analyses, and interpretation of large-scale, high-throughput sequencing data.

References:

1. Castro-Mondragon, J.A., Riudavets-Puig, R., Rauluseviciute, I., Berhanu Lemma, R., Turchi, L., Blanc-Mathieu, R., Lucas, J., Boddie, P., Khan, A., Manosalva Pérez, N., et al. (2022) JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Research*, 50, D165-D173.
2. Gheorghe, M., Sandve, G.K., Khan, A., Chèneby, J., Ballester, B. and Mathelier, A. (2019) A map of direct TF-DNA interactions in the human genome. *Nucleic Acids Res*, 47, e21-e21.
3. Puig, R.R., Boddie, P., Khan, A., Castro-Mondragon, J.A. and Mathelier, A. (2021) UniBind: maps of high-confidence direct TF-DNA interactions across nine species. *BMC Genomics*, 22, 482.
4. Li, Z., Schulz, M.H., Look, T., Begemann, M., Zenke, M. and Costa, I.G. (2019) Identification of transcription factor binding sites using ATAC-seq. *Genome Biology*, 20, 45.
5. Cazares, T.A., Rizvi, F.W., Iyer, B., Chen, X., Kotliar, M., Bejjani, A.T., Wayman, J.A., Donmez, O., Wronowski, B., Parameswaran, S., et al. (2023) maxATAC: Genome-scale transcription-factor binding prediction from ATAC-seq with deep neural networks. *PLOS Computational Biology*, 19, e1010863.
6. Yang, T. and Henao, R. (2022) TAMC: A deep-learning approach to predict motif-centric transcriptional factor binding activity based on ATAC-seq profile. *PLOS Computational Biology*, 18, e1009921.
7. Ouyang, N. and Boyle, A.P. (2020) TRACE: transcription factor footprinting using chromatin accessibility data and DNA sequence. *Genome Res.*, 30, 1040-1046.

Expected skills:

experience in programming in Python, basic knowledge of gene regulation, basic knowledge of one (or more) experimental technique(s) to measure molecular phenotypes of gene regulation (e.g., ATAC-seq, RNA-seq, ChIP-seq)

Possibility of funding:

To be discussed

Possible continuity with PhD:

No

Comments:

Support for housing expenses could be provided. The work will be conducted at the Centre for Molecular Medicine Norway (NCMM), Oslo, Norway, in the Computational Biology & Gene Regulation group led by Anthony Mathelier. The group combines complementary experimental and computational biology expertise and will provide additional mentorship and support. The group meets weekly for research or journal club presentations, and provides software quality and code review meetings.