Predicting the sensitivity of transcription factors to DNA methylation using deep learning

Multicellular organisms establish and maintain different transcriptional states in disparate cell types through complex and specific **regulation of gene expression**. This regulation is mediated by the cooperative **binding of transcription factors (TFs) to regulatory elements** through the recognition of specific DNA sequence motifs. Additionally, the physical access of transcription factors to DNA can be modulated by **epigenetic regulation**, such as nucleosome positioning and **DNA methylation**. Failure to maintain this tight regulation of gene expression results in developmental defects and various diseases including cancer.

Recent **deep learning models** have been developed to predict transcription factor binding sites but their performance still remains limited and they do not take DNA methylation into account. In this project, we propose to improve transcription factor binding site predictions by integrating DNA methylation into deep learning models, investigate which transcription factors are affected by DNA methylation and learn the genomic features underlying good predictions. This project will shed light on **fundamental principles underlying gene regulation by transcription factors and chromatin**.

During this internship, you will **develop a deep learning architecture** inspired from BPNet (Avsec et al. Nature Genetics 2021) and maxATAC (Cazares et al. PLOS Computational Biology 2023) to better predict transcription factor binding sites in the context of DNA methylation by **integrating genomic data** such as TF binding by ChIP-seq, chromatin accessibility by ATAC-seq and DNase-seq, DNA methylation by WGBS) and sequence data such as TF motifs.

You will work in an **international environment at the IGBMC**, a leading European institute. We are a team of computational and experimental biologists and you will benefit from our expertise in genomic data analysis (Anaïs Bardet) and deep learning (Nacho Molina) for the success of the project.

The student recruited should be motivated to continue this work as a PhD project by applying to PhD fellowships (additionally to laboratory grant applications).

Skills:

- Education in Computational Biology, Bioinformatics, Biostatistics or a related field
- Good programming skills (e.g. bash, python, R) in a linux environment
- Good Knowledge of statistics
- Good knowledge of biology and interest in genomics and gene regulation
- Previous experience in machine learning is a plus
- Previous experience analyzing sequencing data is a plus
- Ability to work in a team with both computational and experimental biologists
- Good level in spoken and written english

Contact: Anaïs Bardet, CR CNRS
Please send a cover letter, CV, grades and ranking to anais.bardet@igbmc.fr
The recruitment process is open until one candidate is selected. Deadline: November 30th 2023

Laboratory: IGBMC, Institut de génétique et de biologie moléculaire
https://anaisbardet.cnrs.fr/ - https://www.igbmc.fr/anais-bardet
Address: 1 rue Laurent Fries, 67400 Illkirch (Strasbourg area)

Relevant group publications

- *Preprint*: **Balaramane D, Spill YG,** Weber M, **Bardet AF**. MethyLasso: a segmentation approach to analyze DNA methylation patterns and identify differentially methylation regions from whole-genome datasets. BioRxiv (2023)
- **Detilleux D, Spill YG, Balaramane D**, Weber M, **Bardet AF**. Pan-cancer predictions of transcription factors mediating aberrant DNA methylation. **Epigenetics & Chromatin** (2022)
- Riba A, Oravecz A, Durik M, Jiménez S, Alunni V, Cerciat M, Jung M, Keime C, Keyes WM, **Molina N.** Cell cycle gene regulation dynamics revealed by RNA velocity and deep-learning. (2022) **Nature Communications** 13(1):2865
- **Leporcq C, Spill Y, Balaramane D, Toussaint C,** Weber M, **Bardet AF**. TFmotifView: a webserver for the visualization of transcription factor motifs in genomic regions. **Nucleic Acid Research** (2020)
- *Review*: **Héberlé É**, **Bardet AF**. Sensitivity of transcription factors to DNA methylation. **Essays in Biochemistry** (2019)
- Domcke S*, **Bardet AF***, Ginno P, Hartl D, Burger L, Schübeler D. Competition between DNA methylation and transcription factors determines binding of NRF1. **Nature** (2015)