Universitat Pompeu Fabra Barcelona | Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

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

Project

Computational genomics

Project Title:

Exploring the links between DNA methylation, transcription factor binding, and alternative splicing

Keywords:

Alternative splicing, DNA methylation, transcription factor, cancer genomics, gene regulation

Summary:

RNA splicing is a process involved in mRNA maturation that involves removal of introns from the pre-mRNA. The machinery engaged in this process, the spliceosome, recognizes conserved nucleotide sequences in the introns in order to promote their excision from the pre-mRNA. Alternative splicing is a process that allows the cells to selectively control which parts of a pre-mRNA will be represented in the mature mRNA to be translated into protein. Despite current knowledge, the mechanism of alternative splicing is still not fully understood. The CCCTC-binding factor (CTCF) is a transcription factor (TF) that has recently been shown to be relevant in this process as mutations in its binding site are linked to specific exon inclusion or exclusion [1]. As binding of CTCF to the DNA is altered by DNA methylation [2], the study of the impact of DNA methylation at CTCF binding site on alternative splicing could shed light on aberrant splicing patterns observed in cancers. In this project the Master student will explore the interplay between DNA methylation, transcription factor binding and aberrant alternative splicing in cancers. Specifically, we plan to: (1) detect differentially used exons in cohorts of tumor and normal samples obtained from The Cancer Genome Atlas (TCGA) and the International Cancer Genome Consortium (ICGC); (2) identify binding sites for CTCF and other TFs in the vicinity of the differentially used exons; and (3) characterize the effects of somatic mutations and DNA methylation at these binding sites on aberrant alternative splicing. Some details are provided below. (1) The student will use the DEXSeq [3] Bioconductor package on the RNA-seq data from normal and cancer samples to detect differentially used exons. This analysis will be performed on cohorts of samples from TCGA and/or ICGC for which RNA-seq, DNA methylation, and somatic mutations are available. (2) In the second step of the project, the student will use our UniBind database of high confident direct TF-DNA interactions [4] and TF binding analyses to highlight binding sites for CTCF and TFs that could be associated with alternative splicing. A strategy similar to what was used by Ruiz-Velasco et al. [1] for CTCF will be implemented. (3) In the final step of the project, the candidate will combine information from (1) and (2) to overlay somatic mutation and DNA methylation at TF binding sites with the observed alternative splicing events in cancer patients. This project will consolidate the student's knowledge in computational biology for the analysis of genomics data with a focus on gene expression regulation and cancer. Moreover, the student will get familiar with large cancer genomics public data sets available at TCGA and/or ICGC. She/he will learn how to develop computational workflow to analyze large-scale data, such as differential exon usage and differential methylation analyses. The student will also be exposed to different programming languages such as R, Python, and Bash.

References:

1. Ruiz-Velasco M, Kumar M, Lai MC, Bhat P, Solis-Pinson AB, Reyes A, et al. CTCF-Mediated Chromatin Loops between Promoter and Gene Body Regulate Alternative Splicing across Individuals. Cell Syst. 2017;5: 628–637.e6. 2. Shukla S, Kavak E, Gregory M, Imashimizu M, Shutinoski B, Kashlev M, et al. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. Nature. 2011;479: 74–79. 3. Reyes A, Anders S, Huber W. Inferring differential exon usage in RNA-Seq data with the DEXSeq package. 2013. Available: https://bioconductor.riken.jp/packages/3.5/bioc/vignettes/DEXSeq/inst/doc/DEXSeq.pdf 4. Gheorghe M, Sandve GK, Khan A, Chèneby J, Ballester B, Mathelier A. A map of direct TF–DNA interactions in the human genome. Nucleic Acids Res. 2018;47: e21–e21.

Expected skills::

Proficiency in Python, R, and/or bash; Previous experience in genomics data analysis; Team spirit; English proficiency

Possibility of funding::

To be discussed

Possible continuity with PhD: :

To be discussed