Universitat Pompeu Fabra Barcelona | Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Mireya Plass
Email	mplass@idibell.cat
Institution	Idibell
Website	
Group	Gene regulation of Cell Identity

Project

Computational systems biology

Project Title:

Functional characterization of RNA binding proteins in neural differentiation

Keywords:

single-cell trabscriptomics, gene regulation, RNA binding proteins, alternative polyadenylation

Summary:

Background The Plass lab, located at Bellvitge Biomedical Research Institute (IDIBELL), investigates the role of RNA binding proteins (RBPs) in the regulation of cell identity using a combination of computational and high-throughput experimental approaches. In particular, the lab is interested in understanding how RNA processing mechanisms, including splicing and alternative polyadenylation (APA), impact neuronal cell differentiation and how this process is altered in the development of neurodegenerative diseases. We known that around 70% of human genes are regulated by APA. Despite the growing amount of evidence showing that APA changes according to the proliferation and differentiation status of a cell, it is still not clear if APA contributes to this process and which are RBPs involved. Goals The project proposed will focus on developing new computational approaches to identify RBPs involved in the differentiation of neural cell types in single-cell transcriptomics data (scRNA-seq) and characterize their target genes across different cell types. During this project, the student will learn how to analyze scRNA-seq datasets using existing tools and will develop new computational methods to identify regulatory interactions in single-cell datasets. Approach 1.- Characterize cell and tissue specific RBPs Using a collection of published scRNA-seq datasets, the student will characterize the expression of RBPs across human and mouse cell types from different tissues and differentiation states using existing packages for single-cell transcriptomic analyses. She/he will use these data to make an expression atlas of RBPs and identify cell type and tissue specific RBPs, by comparing the expression of the RBPs across datasets. 2.- Identify RBPs target genes across cells Using a combination of published and newly-developed methods, the student will identify RBP-RBP and RBP-target interactions across cell types that could suggest a functional relation. Assuming that we find interactions in which the expression of an RBP affects the expression of a gene, the student will develop a computational method to recover them from single-cell datasets. We will benchmark the method developed by assessing the ability to identify known interactions between RBPs and genes. Next, we will use a computational pipeline developed in the lab (Plass et al. unpublished), that allows quantifying the expression of individual APA isoforms. Using a similar approach as described before, we will now identify robust associations between specific RBP and individual APA isoforms across cells. 3. – Identify RBP – gene/isoform interactions relevant for neural differentiation The student will use a computational lineage-reconstruction to understand the relationships between different neural cell types in a time-dependent manner, i.e. understand which cell types give rise to other cell types. These methods can also be used to order cells according to their differentiation status. Once she/he has obtained a ordering of cells, she/he will develop a new method to identify interactions between an RBP and a target gene or isoform in time, i.e., identify cases in which the expression of an RBP in time n

affects the expression of a target gene/isoform in time n + t. In this way, we will be able to identify new correlations that may be related to the cellular differentiation process.

References:

Derti, A. et al. A quantitative atlas of polyadenylation in five mammals. Genome Res 22, 1173–1183 (2012). Ji, Z., Lee, J. Y., Pan, Z., Jiang, B. & Tian, B. Progressive lengthening of 3’ untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. Proc Natl Acad Sci U S A 106, 7028–7033 (2009). Miura, P., Shenker, S., Andreu-Agullo, C., Westholm, J. O. & Lai, E. C. Widespread and extensive lengthening of 3’ UTRs in the mammalian brain. Genome Res 23, 812–825 (2013). Plass, M. et al. Cell type atlas and lineage tree of a whole complex animal by single-cell transcriptomics. Science (80-) 360, eaaq1723 (2018). Wolf, F. A. et al. PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. Genome Biol 20, 59 (2019).

Expected skills::

We are looking for a master student with experience in high-throughput sequencing data analyses. Candidates are expected to have experience in R and a scripting language such as Python or Perl. Prior knowledge on post-transcriptional regulation or method development will be a plus. Interest in gene regulation and working in a multidisciplinary team will be valued, as interaction with experimental researchers will be required for the success of the project.

Possibility of funding::

To be discussed

Possible continuity with PhD: :

To be discussed