INFERRING AND COMPARING GENE REGULATORY NETWORKS IN ARABIDOPSIS ROOTS FROM HIGH-DIMENSIONAL BIG DATA

Supervisor: **Olivier Martin**
mail: olivier.c.martin@inrae.fr

Co-supervisor: Thomas Blein
mail: thomas.blein@cnrs.fr

Institute of Plant Sciences – Paris-Saclay, Bâtiment 630, rue de Noetzlin, 91405 – Orsay, France
Web: Team REGARN: Regulatory non-coding RNAs in root plasticity

INTERNSHIP DESCRIPTION

Single-cell transcriptomics [1], i.e., the quantitative measurement of abundances of RNAs simultaneously in thousands of individual cells, have generated a revolution in our knowledge of cell types and of their associated developmental trajectories that originate in stem cells and then branch out in a cascade of differentiations until terminally differentiated states are reached. Teams working in single-cell RNA-seq (scRNA-seq) have now developed intuitive computational tools [2,3] thanks to which they have constructed expression atlases in a number of different species [4]. Developmental trajectories for gene expression patterns can now be produced for the tens of thousands of genes in the genome, both in natural conditions and when these dynamical systems are perturbed using mutations or drugs. The host team has been exploiting both bulk and scRNA-seq data to decipher regulatory interactions during root development in Arabidopsis, an ideal system because of its structure in cell files and its meristem containing the stem cell niche. It is also remarkable because roots naturally produce de novo organogenesis, specifically lateral root organs are generated through a process of spontaneous cell de-differentiation and leading to the production of new stem cells.

The internship will be mentored by O. Martin (modeler and computational biologist) and T. Blein (developmental biologist and bioinformatician). It will analyze currently available datasets (published and not) to infer GRNs (gene regulatory networks) for both primary and lateral roots of the model plant Arabidopsis. It will make use of 3 data integration methods to overcome what is commonly called the “curse of dimensionality” where the high number of genes and modest number of experiments hampers the identification of “which gene drives which other genes”. We will begin by separate GRN inference on bulk and sc RNA-seq datasets, running the dynGENIE3 and

TDCor programs for longitudinal bulk RNA-seq datasets, and running the SCENIC (Aibar et al. 2017) and MINI-EX (Ferrari et al. 2022) programs for single-cell RNA-seq datasets. The resulting predicted GRNs will then be combined using the associated weights for each interaction, a procedure that corresponds to a “post data analysis” integration. We will also perform “co-data analysis” integration in two independent ways:

(i) use the interactions predicted by dynGENIE3 and TDCor from the longitudinal bulk RNA-seq as priors within the MINI-EX analysis of the single-cell data, and

(ii) use the interactions predicted by MINI-EX from the single-cell data as priors within the dynGENIE3 and TDCor analysis of the longitudinal bulk RNA-seq.

All three of these integration methods rely on the putative interactions inferred from DAP-seq and ChIP-seq publicly-available data and known transcription factor binding motifs. We shall include genes - coding, non-coding and miRNAs - that display significant temporal variations or fold changes across these data. These pipelines will produce predicted GRNs for the primary roots and the lateral roots of Arabidopsis. The last part of the project will identify common features in these networks and extract likely evolutionary trajectories for the associated genetic interactions and GRN functional modules, of importance from an evolutionary perspective and for real world applications such as improving plant architecture.

We have ANR funding for a PhD position that will be a follow-up to this M2 internship. That ANR project, in collaboration with Patrick Laufs (IJPB, Versailles), will be focused on comparing the regulations and gene regulatory networks arising in roots, shoots and leaves of Arabidopsis.

REFERENCES

[1] Chen et al. (2019), Single-Cell RNA-Seq Technologies and Related Computational Data Analysis: https://doi.org/10.3389/fgene.2019.00317

[2] Seurat computational tools: https://satijalab.org/seurat/index.html

[3] Monocle computational tools: http://cole-trapnell-lab.github.io/monocle-release/

[4] Plant sc-Atlas, https://bioit3.irc.ugent.be/plant-sc-atlas/root

Selected publications of the team, related to the project

C Grones et al. (2024), Best practices for the execution, analysis, and data storage of plant single-cell/nucleus transcriptomics. The Plant Cell 36 (4), 812-828

SA Cervantes-Pérez et al. (2024), Single-cell transcriptome atlases of soybean root and mature nodule reveal new regulatory programs controlling the nodulation process. Plant Communications 5(8), 100984

A Subbaroyan et al. (2023), Leveraging developmental landscapes for model selection in Boolean gene regulatory networks. Briefings in Bioinformatics 24 (3), bbad160

A Subbaroyan et al. (2022), Minimum complexity drives regulatory logic in Boolean models of living systems. PNAS Nexus 1 (1), pgac017

N Gaggion et al. (2021), ChronoRoot: High-throughput phenotyping by deep segmentation networks reveals novel temporal parameters of plant root system architecture. GigaScience 10 (7), giab052