

Applied Machine Learning/ Bioinformatics Workshop Core Bioinformatics [Mohorianu] Group

June 19, Ferdinand Hall, building A
(ECODAM Session 7)

09:30 - 10:00* Andi Munteanu

(PhD student jointly supervised with Prof H Luchian and Prof L Ciortuz, UAIC and Research Assistant in the Core Bioinformatics group)

StarLng 2.0: Stability and robustness assessment of single-cell clustering

The stochasticity of graph-based clustering influences the robustness of partitions and the reproducibility of interpretations, thus affecting downstream analyses of the biological data. The robustness is assessed using Element Centric approaches which enable the focused evaluation of individual cells. ClustAssess and StarLng are two packages developed built on the assessment clustering stability at both cell and gene levels, respectively.

10:15 – 10:45 Rafael Kollyfas (Research Assistant in the Core Bioinformatics group)

Flufftail 2.0: Unveiling Gene Regulatory Network Dynamics through Fuzzy Clustering of Single-Cell RNAseq Data

State-of-the-art community detection clustering algorithms (applicable on single-cell expression matrices, across modalities) are not only stochastic but also crisp – these characteristics limit the robustness and reproducibility of results, hindering subsequent biological interpretations. Proposing a fuzzy clustering approach, linked to Element Centric Consistency, and consensus clustering, Flufftail enables the data-driven interrogation of dynamic changes in gene regulatory networks.

10:45 – 11:00 COFFEE BREAK

11:00 – 11:30 Miguel Larraz (PhD student jointly supervised with Dr V Kosmoliaptsis, U Cambridge)

Computational Analysis of Liver Organoid Self-Organization: Trajectory Modeling, Clustering Stability, and Gene Network Reconstruction

The availability of large multi-omics datasets presents opportunities for groundbreaking scientific discoveries, but also introduces challenges in terms of interpreting the data in terms of trajectory modelling, and reconstruction of gene regulatory networks, taking into account its dynamics and causality. Frameworks developed in the Mohorianu lab will be exemplified on a recent liver organoid dataset from C Gray-Wilson from the Vallier lab.

11:45 – 12:15 Friedrich-Maximilian Weberling

(PhD student, jointly supervised with Prof Florian Hollfelder, Dept Biochemistry, U Cambridge)

Crécerelle: Integrating gene expression and transcript usage analysis through multi-modal deep generative models

At single-cell resolution, gene expression data can be augmented with information on transcript usage, which refines the interpretation from the perspective of variation focused on isoforms. Using multimodal deep generative latent space models the advantage of the dual perspective is exemplified on two Tabula muris case studies.

12:15 – 12:45 Eleanor Williams (MRC-iCASE PhD student, jointly supervised with Dr Jennifer Tan, Astra Zeneca)

Integrating multiple modalities in spatial context

We present an introduction to the study of disease and injury in a spatial context and overview the current methods and challenges in effectively incorporating spatial information into the analysis of biological data. Spatial transcriptomics is a commonly used and robust method for measuring gene expression in space, and other modalities such as metabolomics can also be captured. Smew and Magpie are two data-driven frameworks for working with these types of data and for extracting more meaningful biological patterns.

*all times are BST + 2 [Romanian time]

13:00 – 14:00 LUNCH

(ECODAM Session 8)

14:30 – 15:00 Cristian Bulgaru

(Research Assistant in the Core Bioinformatics group)

Canard 2.0: Co-expression ATAC and RNA dynamics

(i.e. the challenges of assessing chromatin openness on single-cell data)

Single-cell ATAC-seq data analysis is more challenging than the RNAseq counterpart since it relies on the additional identification of peaks (corresponding to chromatin openness). Using data-driven approaches and novel normalisation and noise filtering, we illustrate how the predictive relationship between ATAC and gene expression can be improved.

15:15 - 15:45 Serban Doncean

(intern in the Core Bioinformatics group and MSc student jointly supervised with Prof L Ciortuz)

FALCON: Functional Analysis / Learning of Corrected Omics via (regulatory) Networks

Batch correction is an essential tool to ensure the integration of datasets. Strict mathematical approaches focusing on matching distributions, align both signal and noise. We propose a function approach, based on gene regulatory networks, which minimises the effect of localised noise or sequencing biases, focusing on the conserved regulatory interactions.

15:45 – 16:15 COFFEE BREAK

16:15 – 16:45 Irina Mohorianu

Head of Bioinformatics/ Scientific Computing @ CSCI, University of Cambridge

RoSignOL: robust signature of Life

The rapid development of high throughput approaches generated vast quantities of data. However, these are often underutilised due to computational or time constraints. Within the Core Bioinformatics group at CSCI, we generated a set of tools (named after birds); in this last talk, we will discuss about future projects which underline how essential a thorough computational perspective is for the analysis of biological high throughput datasets.

Please note that all information presented during the workshop should be seen as confidential. The presentations contain information submitted for publication to journals or in preparation for submission.

We acknowledge our funders, Wellcome Trust, Medical Research Council, Open Targets, Evelyn Trust, Isaac Newton Trust and our industrial collaborators Astra Zeneca and GSK. The funders did not influence the results presented in these talks.

*all times are BST + 2 [Romanian time]