

## Applied Machine Learning/ Bioinformatics Workshop

### Core Bioinformatics [Mohorianu] Group

09:30 - 10:00\* Friedrich-Maximilian Weberling

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

#### **Design of biocatalysts through machine learning guided directed evolution**

*High-throughput deep mutational scanning of enzymes enables mapping the sequence-function relationship in directed evolution campaigns. Machine learning can further guide the search for higher functional variants by predicting beneficial mutations for future campaigns.*

10:00 - 10:30 Dorin Miron

(Bioinformatics intern, jointly supervised with Prof Liviu Ciortuz, UAIC)

#### **Molecular HLA typing can improve transplant assignment.**

*Transplant inequity as well as high rejection rates can be addressed using molecular HLA typing. Machine Learning and Data Science approaches are used for extracting discriminative features at amino-acid and protein structure levels.*

10:30 - 11:00 Liviu Pirvan

(Research Associate, jointly supervised with Dr Srinjan Basu, Imperial College)

#### **The 3D shape of DNA and its predictiveness gene expression**

*The 3D shape of the DNA has a wide-ranging impact on which and how genes are expressed, ultimately impacting how organisms work. In this study, we consider how chromatin information can be integrated with expression data at single-cell level, to enhance biological insights.*

11:00 – 11:30 Coffee break

11:30 - 12:00 Andi Munteanu

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

#### **StarLng: 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.*

12:00 – 12:30 Rafael Kollyfas (MSc student and Research Assistant)

#### **Flufftail: 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.*

12:30 - 13:00 Cameron Crawford

(BSc student in Computer Science, jointly supervised with Prof Pietro Lio, U Cambridge)

#### **RobIn: causality embedded in co-variation in the context of Gene Regulatory Networks**

*Current methods for inferring gene regulatory networks are based on standard (symmetric) distances; the symmetry of these evaluations hinders the cause/effect interpretation of outputs and masks biologically relevant patterns. RobIn (**R**obust **I**nference) uses explained variance to predict causality.*

\*all times are BST + 2 [Romanian time]

13:30 – 15:00 Lunch

15:00 - 15:30 Cristian Bulgaru

(MSc student and Bioinformatics intern, jointly supervised with Prof Liviu Ciortuz, UAIC)

**Canard: 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.*

16:00 - 16:30 Eleanor Williams

(MRC-iCASE PhD student, jointly supervised with Dr Jennifer Tan, Astra Zeneca)

**Smew & Magpie: Unsupervised learning for spatially-resolved biology**

*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.*

16:30 - 17:00 Miguel Larraz (Research Assistant)

**Wren: Integrating genome sequencing with RNA expression**

*The availability of large multi-omics datasets presents opportunities for groundbreaking scientific discoveries, but also introduces challenges in terms of data integration. The WREN pipeline proposes a robust interactive approach for seamlessly linking single-cell RNAseq with whole genome sequencing, and assessing the predictive power of somatic or germline mutations on gene expression.*

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]