香港浸會大學理學院
HKBU Faculty of Science

DEPARTMENT OF COMPUTER SCIENCE

PhD Degree Oral Presentation

PhD Candidate:	Mr. XU Ke
Date	15 May 2025 (Thursday)
Time:	9:00 am – 10:30 am (35 mins presentation and 15 mins Q & A)
Venue:	ZOOM (Meeting ID: 982 0549 0889) (The password and direct link will only be provided to registrants)
Registration:	https://tinyurl.com/bucs-tdreg (Deadline: 12:00 nn, 14 May 2025)

Computational Methods for Single-cell Omics Data Analysis

Abstract

The discovery of the structure of DNA and the formulation of the central dogma of molecular biology illuminated the rich information contained in cellular molecules. Advances in high-throughput sequencing technologies allow us to explore molecular activities in cells at an unprecedented resolution. The single-cell omics data from these technologies provide a new perspective to understand cellular heterogeneity and the underlying mechanisms of biological processes such as organogenesis, immune response, and drug resistance. These technologies have also played a critical role in addressing public health challenges such as creating COVID-19 transcriptomic atlas or recording host transcriptomic response for infection of Ebola virus.

Among the various single-cell omics technologies, our study focuses primarily on single-cell RNA sequencing (scRNA-seq) and spatially resolved omics. Numerous scRNA-seq technologies have been developed to measure gene expression levels at single-cell resolution. However, their data suffer from low sensitivity of capturing RNA molecules. To address these issues, researchers have developed gene expression imputation methods to estimate missing gene expression values in scRNA-seq datasets. Existing imputation methods only adopt global feature selection, i.e., selecting the same set of genes to perform imputation for all cells. These gene sets may contain outlier values in certain cells, which would affect imputation accuracy. Although imputation methods can predict missing gene expression values, they cannot recover spatial information lost during tissue dissociation. Spatially resolved omics technologies instead retain both spatial information and expression profiles. The two types of information together enable exploration of spatial domains like the L1 layer of the mouse cortex, where cells that perform similar or synergistic functions are located in close proximity. As clustering methods for scRNA-seq data rely solely on gene expression data, there is an urgent need to develop novel spatial domain clustering methods tailored for spatially resolved omics data analysis. Moreover, certain spatially resolved omics datasets include additional morphological images, which pose another challenge to integrating image information into clustering analysis.

We developed three computational methods to address the challenges in single-cell omics data analysis. First, we developed IGSimpute, an accurate and interpretable gene expression imputation method for scRNA-seq data. IGSimpute employs a gene selection strategy to identify the most informative genes for each cell to impute missing gene expression values. Building on the growing availability of spatially resolved omics technologies, we then developed stDyer to identify spatial domains for such datasets. stDyer leverages spatial information and gene expression information to construct dynamic graphs, capturing intercellular relationships. Additionally, stDyer identifies spatially variable genes (SVGs) that exhibit spatial patterns associated with corresponding spatial domains. Finally, we developed stDyer-image, a clustering analysis method with image feature links for spatially resolved omics data. stDyer-image utilizes image information to improve clustering analysis by integrating image information with gene expression data. The developed methods have been evaluated on real datasets from various technologies, demonstrating superior performance compared with state-of-the-art methods in terms of imputation accuracy and clustering performance.

***** ALL INTERESTED ARE WELCOME *****