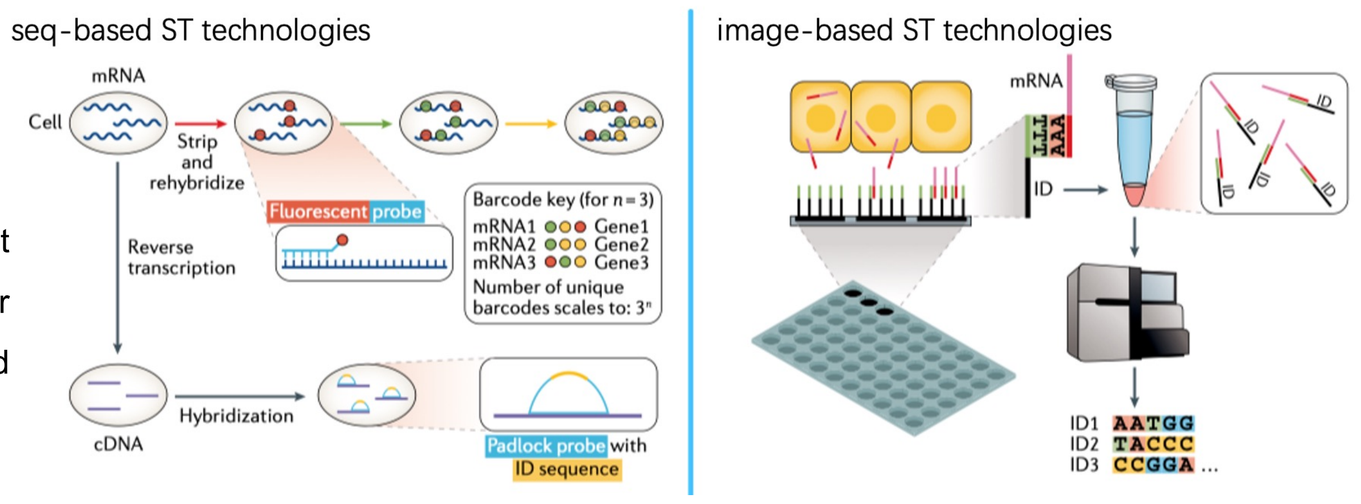
MIIS 2022
International Workshop On Mathematical Issues In Information Sciences
数据科学国际研讨会
17th-18th, Dec. 2022

SpatialScope: A unified toolbox for integrative analysis of spatial and single-cell transcriptomics data

Introduction

- **Single-cell RNA sequencing (scRNA-seq)**

scRNA-seq characterizes the whole transcriptome of individual cells but the spatial location of cells are lost. In fact, spatial information is crucial for understanding the spatial distribution, interaction, communication, and physiological functions of different cell types.



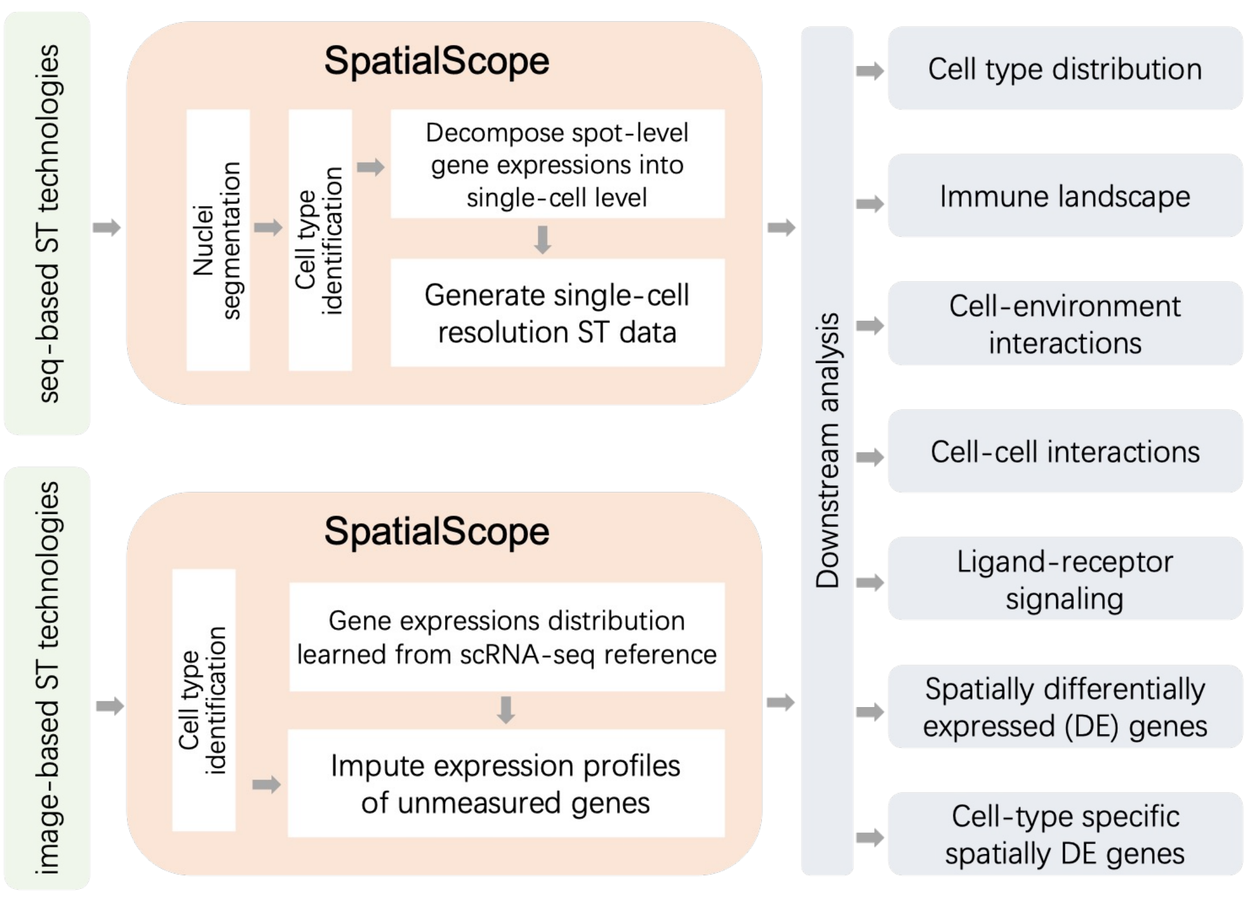
- **Spatial transcriptomics (ST)**

As a newly developed technology in recent years, ST technologies have been developed to quantify spatially localized transcriptomes, which accelerated the capacity to elucidate the development of healthy tissue, and tumor microenvironment of cancers. However, current two major ST technologies have their own limitations: seq-based ST technology cannot achieve single-cell resolution, one spatial location may contain multiple cells; image-based ST technology suffers from limited gene throughput, only a few hundred to a few thousand genes can be detected.

- **SpatialScope**

We developed an unified toolbox, SpatialScope, that can integrate multi-platform ST and scRNA-seq data. For seq-based ST data, we decompose low-resolution ST data into single-cell level by combining deep learning and multimodality data (HE staining imaging, spatial location, spatial transcriptomics, scRNA-seq) , thus achieving single-cell resolution. For the image-based ST data, we impute the gene expressions of unmeasured genes with a deep generative model, thus achieving whole transcriptome coverage.

Method

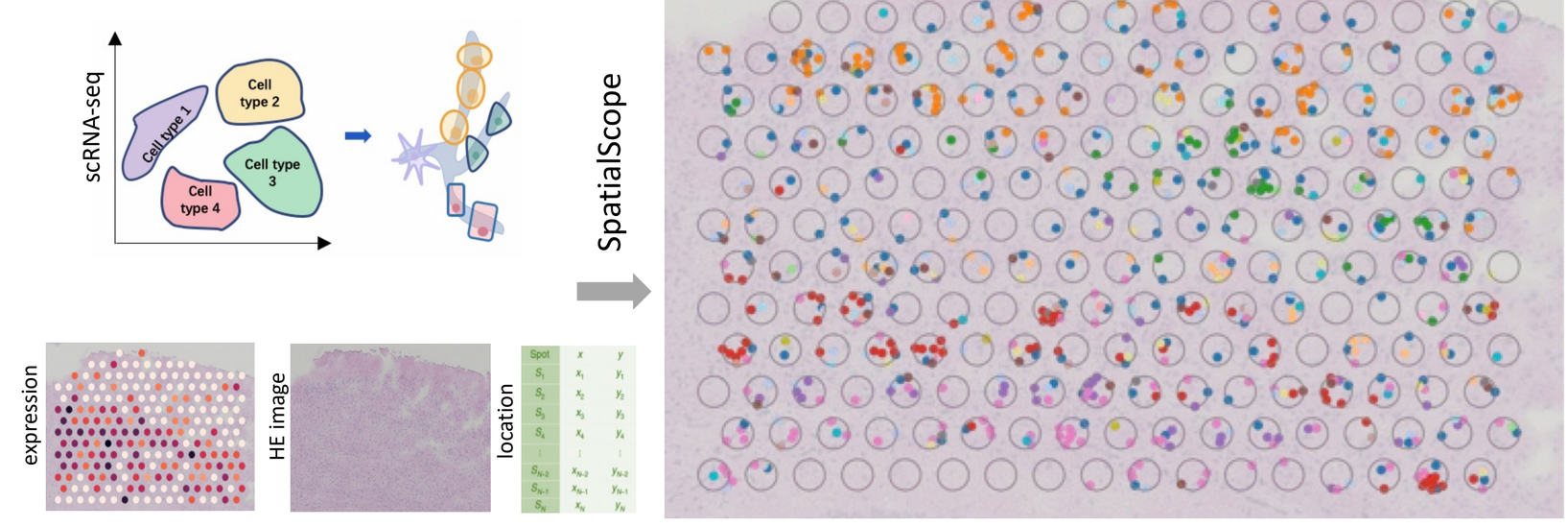


Seq-based ST data: We first obtain the spatial locations of individual cells through deep nuclei segmentation tool, and then decompose the ST data measured at low resolution into single-cell resolution by combining a hierarchical statistical model and a gene expression separation model. Finally, we assign the generated single-cell level gene expressions to the segmented cells to achieve single-cell resolution ST data.

Image-based ST data: We first learn the gene expression data distribution of scRNA-seq through a deep generative model, and then impute the expressions of unmeasured genes by taking samples from the posterior distribution given the measured gene expressions to achieve whole transcriptome coverage.

After obtaining ST data with single-cell resolution and whole transcriptome coverage, we can perform comprehensive and in-depth downstream analysis. For example, the spatial distribution and interactions of different cell types, the interactions between cells and the environment, the ligand-receptor signaling, and the spatially differentially expressed genes, etc.

Demonstration



Research members

- Xiao Jiashun, Research scientist, SRIBD
- Wan Xiaomeng, Ph.D candidate, HKUST
- Wan Xiang, Director of medical big data laboratory, SRIBD
- Yang Can, Associate professor, HKUST
- Wu Angela, Associate professor, HKUST
- Lin Zhixiang, Assistant professor, CUHK

深圳市大数据研究院 Shenzhen Research Institute of Big Data
香港中文大學(深圳) The Chinese University of Hong Kong, Shenzhen
TBSI 清华-伯克利深圳学院 Tsinghua-Berkeley Shenzhen Institute
雷达信号处理国家级重点实验室 National Key Laboratory of Radar Signal Processing
综合业务网理论及关键技术国家重点实验室 State Key Laboratory of Integrated Services Networks
鹏城实验室 PENGCHENG LABORATORY
广东省大数据计算基础理论与方法重点实验室 GuangDong Key Laboratory of Big Data Computing
国家健康医疗大数据研究院(深圳) NATIONAL HEALTH DATA INSTITUTE, SHENZHEN