TriState Data Analysis Core: Advancements in Data Submission, Analysis, and Tool Development

TriState Data Analysis Core team, Dongmei Li[1, *], Jose Lugo-Martinez[2, *], Qin Ma[3, *]

1 Clinical and Translational Science Institute, University of Rochester, Rochester, NY, USA; 2 Computational Biology Department, Carnegie Mellon University, Pittsburgh, PA, USA;
3 Department of Biomedical Informatics, Ohio State University, Columbus, OH, USA;
* Corresponding author: Dr. Dongmei Li <dongmei_li@urmc.rochester.edu>, Dr. Jose Lugo-Martinez <jlugomar@andrew.cmu.edu>, Dr. Qin Ma <qin.ma@osumc.edu>

Introduction

The TriState Data Analysis Core (DAC) has made significant progress over the past year, contributing to the TriState SenNet Consortium's efforts to map senescent cells in the human heart and lung.

Core Mission: Construct and analyze high-fidelity biomarkers and map datasets from heart and lung tissues

1. **Data Delivery and Collaborative Standards Development**: Robust and standardized data delivery to the SenNet Consortium Organization and Data Coordinating Center (CODCC). Establish open data and metadata standards network-wide.
2. **Storage & Analysis**: Employing advanced computational tools to store, analyze, and model data from Biospecimen and Biological Analysis Cores.
3. **Computational and AI Tools Development**: Innovating and refining advanced and robust computational tools and pipelines in assisting constructing a lung and heart senescence atlas.

Data Submission

1. **Virtual Coordination**: Hosted focused Zoom meetings with site coordinators for metadata alignment.
2. **Submission Toolkits**: Developed and shared site-specific data submission folders with integrated metadata templates.
3. **Training Resources**: Provided recorded training sessions to facilitate data submissions at each site.
4. **Data Tour Participation**: Engaged in CODCC weekly meetings and workshops to expedite data submission and transfer processes.

Table 1: 2023 Data Coordination Commitment from TMC-TriState

Site Contact	Site	Organ	Assay	# Datasets
Lorena Rosas	Ohio State U.	Human Lung	scRNA-seq H&E	34
Irfan Rahman	U. of Rochester	Human Lung	FACS snRNA-seq H&E	34
Nayra Cardenes	U. of Pittsburgh	Human Lung	Bulk RNA-seq LC-MS	38
Nayra Cardenes	U. of Pittsburgh	Human Heart	Bulk RNA-seq	24

Data Analysis

Senescence gene signature enrichment in scRNA-seq data

PIs: Ana Mora and Mauricio Rojas (Ohio State U.)
Analyzed by: Hao Cheng, Cankun Wang, and Qin Ma (Ohio State U.)
Datasets: scRNA-seq of Idiopathic pulmonary fibrosis (IPF) and healthy lungs samples (66,189 cells)

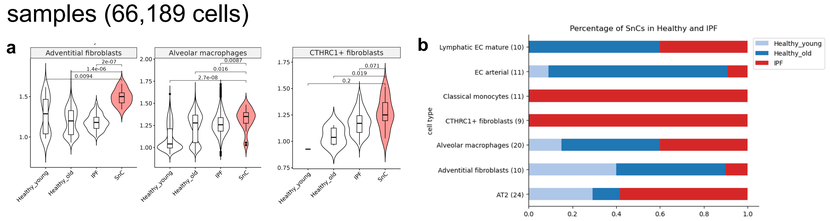

Figure 1. (a) Violin plots of SenMayo marker gene signature enrichment score between healthy and IPF cells. **(b)** Barplots of the distribution of cell-type-specific senescent cells by disease groups. Results were identified by DeepSAS.

Tissue architecture inference from spatial transcriptomics data analysis

PIs: Ana Mora and Mauricio Rojas (Ohio State U.)
Analyzed by: Hyeongseon Jeon and Dongjun Chung (Ohio State U.)
Datasets: 10X Visium Spatial Transcriptomics of IPF and healthy lungs (11,423 spots)

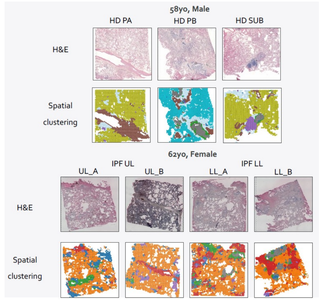

Figure 2. H&E images and spatial clustering results for healthy and IPF tissues using MAPLE, our novel in-house algorithm for multi-sample spatial omics data.

Dissecting trigger-specific cellular senescence in human precision-cut lung slices (PCLS)

PIs: Melanie Königshoff, Oliver Eickelberg (U. Pittsburgh)
Analyzed by: Cankun Wang, Qin Ma (Ohio State U.)
Datasets: bleomycin or doxorubicin-induced snRNA-seq of PCLS lungs (142,847 cells)

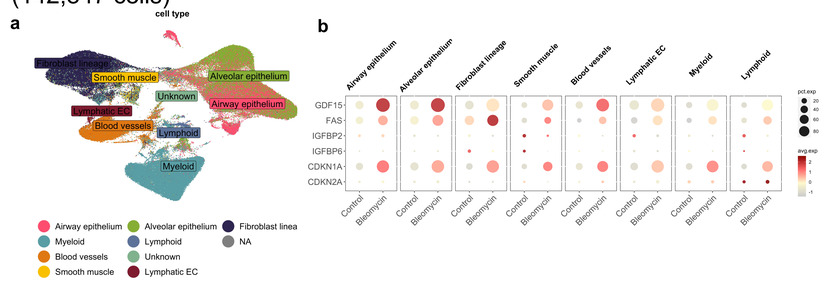

Figure 3. (a) UMAP plot colored by predicted cell type annotations. **(b)** Dot plots of SASP factors gene expression between control and Bleomycin treatment.

Tool Development

DeepSAS: Identifying cell-type-specific senescent cells and signature genes using heterogeneous graph contrastive learning
MarsGT: Multi-omics analysis for rare population inference using single-cell graph transformer
Lead by: Qin Ma (Ohio State U.)

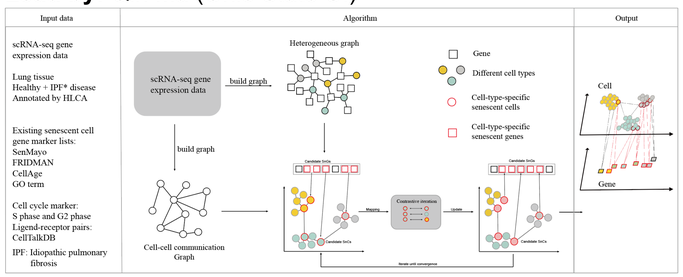

Figure 4. DeepSAS consists of the development of robust contrastive learning and graph representation learning frameworks for the discovery of Senescent cells and Senescent cell marker genes.

MAPLE: A hybrid framework for multi-sample spatial transcriptomics data
Lead by: Hyeongseon Jeon and Dongjun Chung (Ohio State U.)

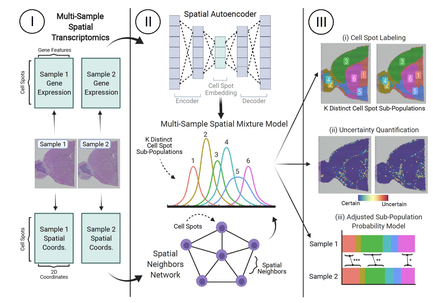

Figure 5. MAPLE is a hybrid deep learning and Bayesian modeling framework for the detection of spatially informed cell spot sub-populations, uncertainty quantification, and inference of group effects in multi-sample experiments.

scDOT: Integration of single-cell and spatial transcriptomics data to infer spatial organization of senescent cells
Lead by: Jose Lugo-Martinez and Ziv Bar-Joseph (Carnegie Mellon U.)

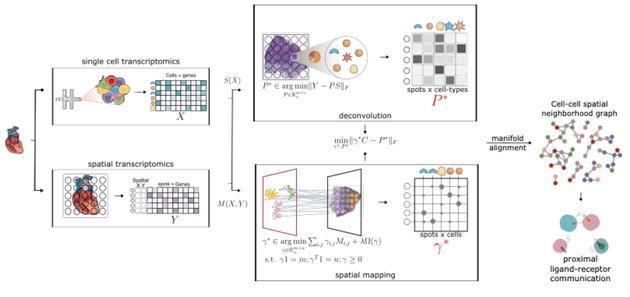

Figure 6. scDOT integrates optimal transport and expression deconvolution to learn non-linear couplings between cells and spots and to infer cell placements.

Acknowledgment: This work was supported by awards and U54-AG075931 from the National Institutes of Health.

BMBL Bioinformatics and Mathematical Biosciences Lab
The Ohio State University Wexner Medical Center
TriState SenNet