Identifying cell-type-specific senescent cells and signature genes using heterogeneous graph contrastive learning

Hao Cheng[1], Anjun Ma[1], Cankun Wang[1], Ana L. Mora[2], Mauricio Rojas[2], Dongjun Chung[1], Qin Ma[1,*]
1 Department of Biomedical Informatics, Ohio State University, Columbus, OH, USA; 2 Department of Internal Medicine Division of Pulmonary, Ohio State University, Columbus, OH, USA
* Corresponding author: Dr. Qin Ma. Email: qin.ma@osumc.edu

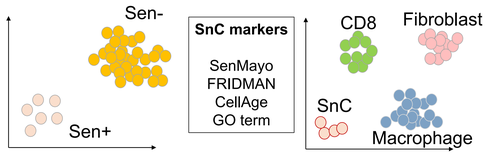
Introduction

Background and significance:
- Senescent cells (**SnCs**) are rare cells that arrest in G1 phase remain and continue to release chemicals that can trigger inflammation, mostly related to age-related diseases and cancer.
- Some hall marker, e.g., p16 and p21, have been discovered to be related to SnCs.

However, existing SnC marker genes (**SnGs**) showed exceptions in characterizing inter-cellular SnC; there exist cell type heterogeneity among SnCs and no existing tool can computationally recognize cell-type-specific SnCs.

Challenges:
- Multi-level clustering resolutions (cell type level and cell phase level) are involved in recognizing SnC.
- SnGs of each cell types are largely unknown.

Common SnC identification using hall markers



Cell-type-specific SnCs and signature gene identification

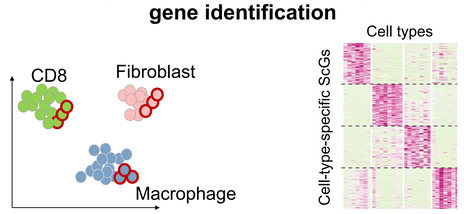


Figure 1. Theoretical identification of SnCs using hall markers and cell-type-specific SnGs

The DeepSAS Workflow

We develop **DeepSAS (Deep Learning Framework for Senescent Cell)** to identify **cell-type-specific SnCs and SnGs** from scRNA-seq data using a novel **heterogeneous graph contrastive learning model**.

Highlights:
- Using **cell-gene heterogeneous graph** to represent scRNA-seq data, so that the cell-gene influence can be involved.
- Using **contrastive learning** to amplify the cell differences between SnCs and normal cells in the same cell types.
- Involving four types of **comparisons**: intra-cluster SnC vs SnC (d1), intra-cluster SnC vs normal (d2), inter-cluster SnC vs SnC (d3), and inter-cluster SnC vs normal (d4).
- Using **attention mechanism** to calculate the importance between cells and genes and identify SnCs and SnGs in each cell type, simultaneously.

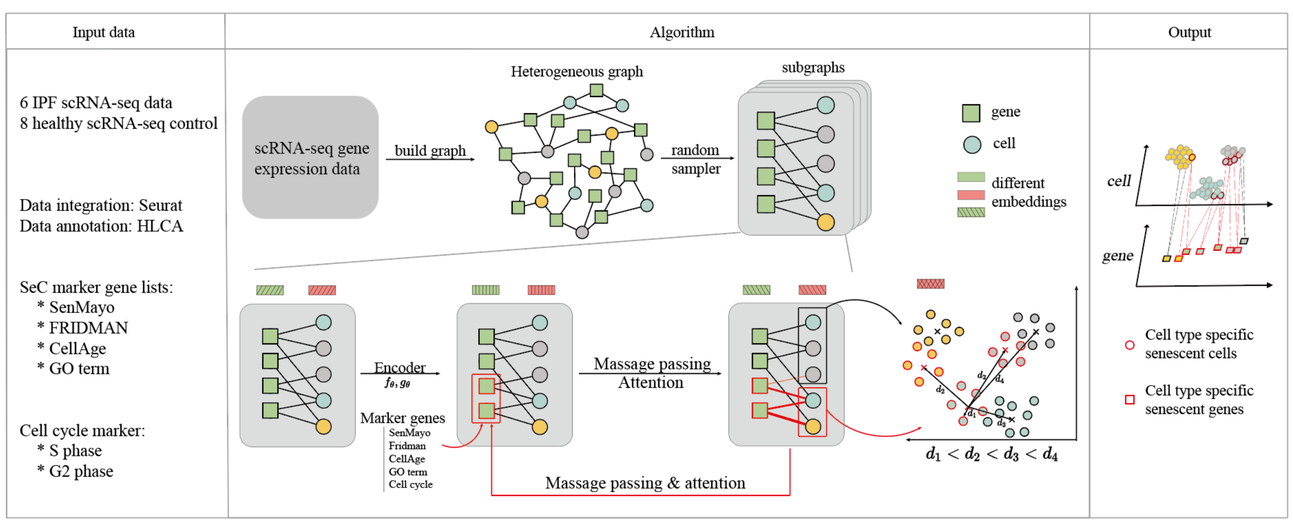


Figure 2. The workflow of DeepSAS. DeepSAS consists of the development of robust contrastive learning and graph representation learning frameworks for the discovery of SnCs and ScGs. IPF: idiopathic pulmonary fibrosis

Preliminary results

- Tested on a scRNA-Seq dataset of idiopathic pulmonary fibrosis (IPF) disease patients;
- Select four main cell types for the following analysis.

- Our model is well-trained and can successfully find the convergence point in several iterations.
- The results on different combinations of dataset demonstrate the robustness of our model. The identified SnCs do not change due to the combination of dataset scales and phenotypes.
- The distribution of SnCs varies among different clusters, and the number of SnCs is rare.
- The expression of SnGs varies among SnCs and normal cells. Some SnGs have verified by previous study (marked red).
- Most of the identified SnGs are related to pathway of cell aging and death.

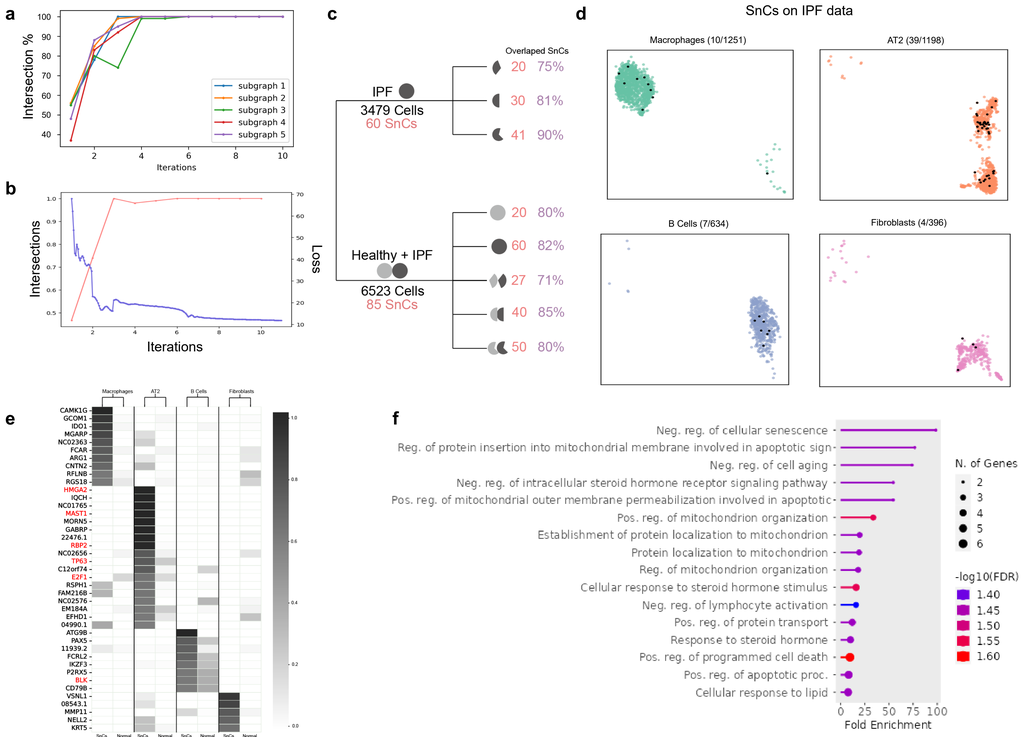


Figure 3. The experimental verification of our method. **(a)** The change of percentage of overlap during the iteration of different subgraphs. **(b)** The change of loss function values during iteration of one subgraph. **(c)** The numbers of SnCs predicted from healthy and IPF samples for different sample size and the overlap between different samples. **(d)** The UMAP plots of cell-type specific SnCs for four cell types (Macrophages, AT2, B cells and Fibroblasts). The identified SnCs are marked black. **(e)** The heatmap for cell-type-specific SnGs in the cell types showcased in **(d)**. **(f)** The GO biological process pathway enrichment analysis for identified SnGs in **(e)**.

Future study

- Integrate multiple samples to study age trends in SnGs and SnCs.
- Extend cell-type-specific SnCs and SnGs to different organs.
- Causal inference-based model to study the causal relationship SnC and diseases.
- Integrate scRNA-seq data and spatial transcriptome data for spatial SnC mapping.

Reference: Ma, A., Wang, X., Li, J. et al. Single-cell biological network inference using a heterogeneous graph transformer. Nat Commun 14, 964 (2023).
Acknowledgement: 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