Computational inference of heterotypic cell interactions in single cell RNA-seq data

Hanji Kim[1] & Jin-Wu Nam[1]

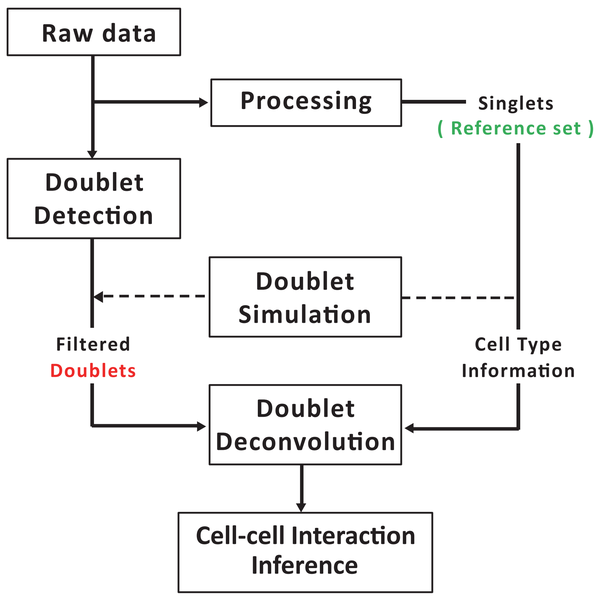
[1]Department of Life Science, College of Natural Sciences, Hanyang University, Seoul 04763, Republic of Korea

Abstract

Single-cell RNA sequencing (scRNA-seq) has become a critical technology for unraveling the complex cellular heterogeneity within a tissue. While droplet-based microfluidics system is currently most widely used among many such methodologies, a typical artifact called ‘doublets’, where two cells are caught in the same droplet, should be successfully identified and removed in downstream analysis to obtain unbiased results. Conversely, some recent studies utilized the innate interactions between those ‘physically interacting cells (PICs)’ and obtained spatial information from them. Here, we introduce a computational analysis pipeline to infer potential cell-cell interactions from scRNA-seq data. In this pipeline, for every putative cell-type pair which are forming heterotypic doublets that are computationally detected, the observed rate of doublets comprised of corresponding cell-types is compared to the expected doublet rate of them. This signal-to-noise ratio with confidence-intervals was utilized to infer the interactions between the two cell-types. We evaluated this method over simulated doublets as well as biological samples from fly blood and multiple myeloma patients, detecting immune-cell interactions with a statistical significance in scRNA-seq data. We believe our research can help gain further insight into biologically meaningful cell-cell interactions in diverse conditions and diseases.

***Keywords**: single-cell RNA sequencing, physically interacting cells, doublet, cell-cell interaction, fly blood, tumor and immune cells*

Methods

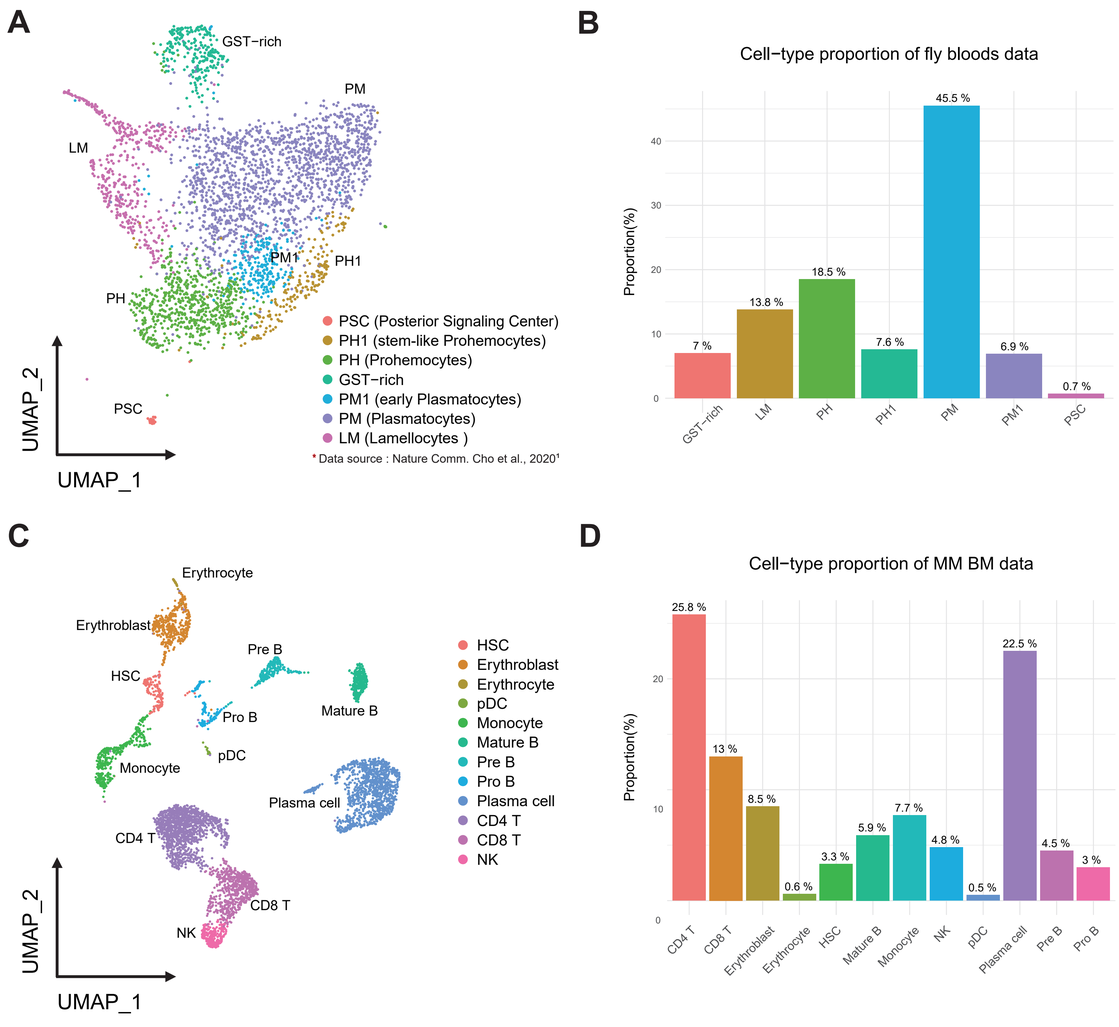


Computational analysis pipeline

- Putative singlets are obtained through processing of raw scRNA-seq data.
- Doublets detected from raw data are filtered based on the threshold which is calculated by simulation of artificial heterotypic doublets using singlets.
- Identify cell-types composing each of these doublets (Deconvolution).
- Compare the expected and observed doublet frequency using signal-to-noise ratio with confidence-intervals of all possible combinations of two cell-types.

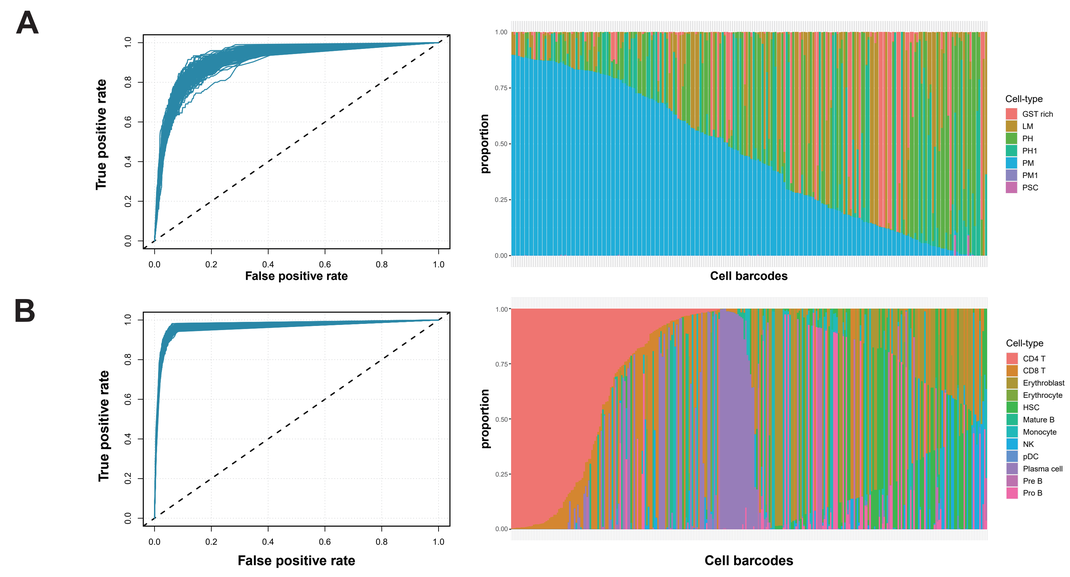
Results

Figure 1. scRNA-seq data of circulation bloods in wasp-infested fly larvae and of bone marrow from multiple myeloma



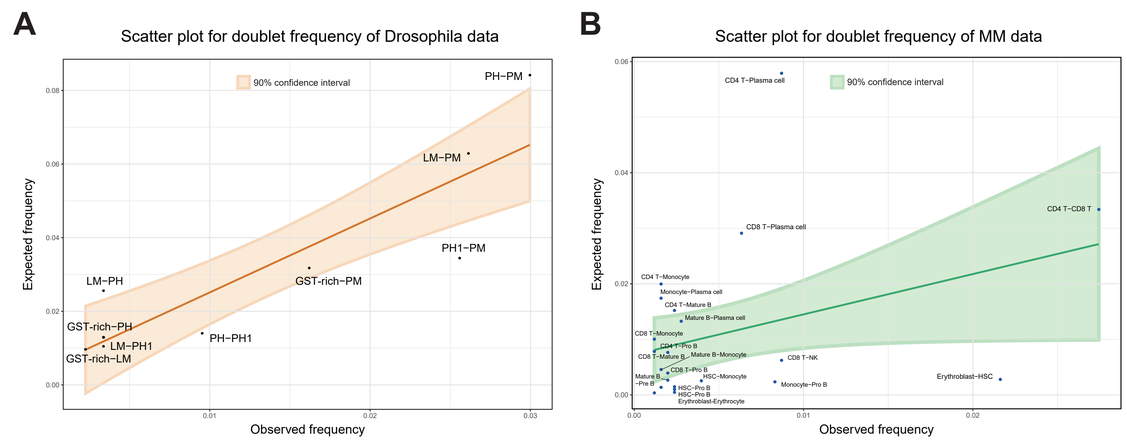
In order to generate reference-set (singlets) of doublet analysis, scRNA-seq data from circulation bloods in wasp-infested fly larvae and bone marrow(BM) of multiple myeloma(MM) patient were analyzed using Seurat. Putative singlets were obtained and clustered. UMAP and cell-type proportions of fly bloods data are represented in (A) and (B), respectively. (C) and (D) are those for the MM BM data.

Figure 2. 100 ROC curves for doublet calling simulation and deconvolution results of detected doublets



Detected doublets are filtered based on the optimal threshold for posterior probabilities of doublet prediction, calculated from results of 100 times of heterotypic doublet simulations. ROC curves of all these simulations and doublet deconvolution results are in the left and right side of (A) for fly bloods data and (B) for MM BM data, respec-

Figure 3. Comparison of expected and observed doublet frequencies



(A) is scatter plot of expected and observed doublet frequency of fly bloods data, and (B) is for MM BM data. The shaded area of each plot represents the 90% confidence interval for the regression line.

Discussion

- Potential heterotypic cell-cell interactions can be identified through this analysis pipeline which rely soley on computational analysis of expression matrix.
- Using mildly-dissociated CITE-seq data could be a suitable way to validate this pipeline.
- Accuracy of our method depends on performances of doublet detection and deconvolution algorithms.

References

1. Cho, B., et al. Single-cell transcriptome maps of myeloid blood cell lineages in Drosophila, Nat Commun (2020)
2. Yuhan Hao, et.al. Integrated analysis of multimodal single-cell data, Cell (2021)
3. Christopher S. McGinnis, et.al. DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors, Cell Systems (2019)
4. Samuel L. Wolock, et.al. Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data, Cell Systems (2019)
5. Abha S Bais, Dennis Kostka, scds: computational annotation of doublets in single-cell RNA sequencing data, Bioinformatics (2020)
6. Tim Stuart, et.al. Comprehensive Integration of Single-Cell Data, Cell (2019)

Contact info

+82 (2) 2220-2428
khj75319@hanyang.ac.kr
http://big.hanyang.ac.kr