Bioinformatics, 38(13), 2022, 3488–3489
https://doi.org/10.1093/bioinformatics/btac350
Advance Access Publication Date: 23 May 2022
Applications Note

Databases and ontologies

webSCST: an interactive web application for single-cell RNA-sequencing data and spatial transcriptomic data integration

Zilong Zhang [1,2,3], Feifei Cui [1,2,3], Wei Su[4], Lijun Dou [2,3], Anqi Xu[5], Chen Cao [6] and Quan Zou [2,3,*]

[1]School of Computer Science and Technology, Hainan University, Haikou 570228, China, [2]Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China, Chengdu 610054, China, [3]Yangtze Delta Region Institute (Quzhou), University of Electronic Science and Technology of China, Quzhou 324000, China, [4]Yahoo Japan Corporation, Tokyo 102-8282, Japan, [5]The First School of Clinical Medicine, Shandong University of Traditional Chinese Medicine, Jinan 250014, China and [6]School of Biomedical Engineering and Informatics, Nanjing Medical University, Nanjing 211166, China

*To whom correspondence should be addressed.
Associate Editor: Peter Robinson

Received on March 27, 2022; revised on April 26, 2022; editorial decision on May 14, 2022; accepted on May 18, 2022

Abstract

Summary: Integrative analysis of single-cell RNA-sequencing (scRNA-seq) data with spatial data for the same species and organ would provide each cell sample with a predictive spatial location, which would facilitate biological study. However, publicly available spatial sequencing datasets for specific species and organs are rare and are often displayed in different formats. In this study, we introduce a new web-based scRNA-seq analysis tool, webSCST, that integrates well-organized spatial transcriptome sequencing datasets categorized by species and organs, provides a user-friendly interface for raw single-cell processing with popular integration methods and allows users to submit their raw scRNA-seq data once to obtain predicted spatial locations for each cell type.

Availability and implementation: webSCST implemented in shiny with all major browsers supported is available at http://www.webscst.com. webSCST is also freely available as an R package at https://github.com/swsoyee/webSCST.

Contact: zouquan@nclab.net

1 Introduction

With the rapid development of next-generation sequencing technology, RNA-sequencing allows biologists to quantify gene expression in tissues or cell lines. However, knowledge of the spatial location and cell heterogeneity is lacking (Wang *et al.*, 2009). Single-cell RNA-sequencing (scRNA-seq) characterizes the transcriptome at the single-cell level and can show a snapshot of the gene expression profiles in a single cell. However, spatial location information is lost during the creation of single-cell suspension in dissociated steps (Zhang *et al.*, 2021). To address this problem, various technologies have been developed to quantify gene expression while preserving spatial information (Dries *et al.*, 2021). Among them, spatial labeling technology has become the most popular protocol because it can detect gene expression at the whole-transcriptome level. Therefore, integrating the analysis of scRNA-seq data and spatial sequencing data could allow single-cell level gene expression, together with spatial information within the native tissue, to be obtained (Longo *et al.*, 2021).

However, as an emerging technology, spatial sequencing is usually either expensive or time-consuming, and the format of the results varies greatly for different studies. Also, there is no gold standard for it as integration tools are developing rapidly. Consequently, users generally need to try several different tools to obtain meaningful biological results. Moreover, because different tools often request different input formats, the process takes biologists considerable time and effort (Fan *et al.*, 2020).

To address these challenges, we propose a user-friendly web application, webSCST (http://www.webscst.com). webSCST consists of a spatial database organized by different species and tissues, and can be utilized by either checking spatially differentially expressed genes or via integration with users' scRNA-seq data. webSCST also performs a full bioinformatics analysis of scRNA-seq data. Users upload raw scRNA-seq data for processing; it is then matched to possible spatial data from the webSCST database, incorporating the most popular integration methods for scRNA-seq and spatial data.

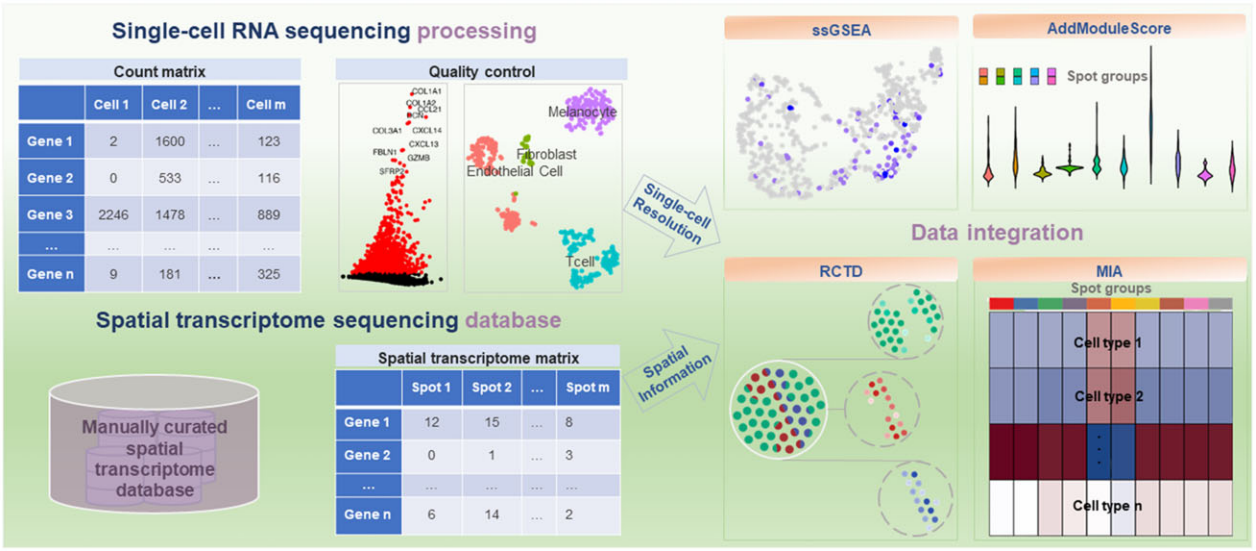

Fig. 1. Overall architecture of the webSCST platform. webSCST consists of three sections: scRNA-seq data processing, a spatial transcriptomic database and data integration

2 Materials and methods

Data for the spatial sequencing database were collected from newly published spatially resolved transcriptome datasets from publications and resources including the 10X Genomics website (https://www.10xgenomics.com/resources/datasets). Raw spatial sequencing data, downloaded from the literature, were processed using the R package Seurat (Hao *et al.*, 2021). Following the general spatial data processing pipeline using default parameters, we manually extracted gene expression values and spatial position information that facilitated subsequent integration with scRNA-seq data. In addition, all the spatial datasets were grouped by different species and tissues. Users can view the spatial expression of any gene of interest after specifying the species and tissue. Currently, there are 43 manually curated spatial transcriptome datasets (136 subdatasets) that will be regularly updated as further spatial transcriptome datasets become publicly available.

More importantly, webSCST allows users to easily obtain predicted special gene expression from raw scRNA-seq data. By uploading raw scRNA-seq datasets (gene expression matrix in MTX format, gene and cell information in TSV format and cell type annotation in TXT format), users can perform quality control in an interactive web application, choosing parameters such as maximum gene numbers, percentage of mitochondria and principal component numbers in the interactive interface.

After scRNA-seq processing, users can choose to match their scRNA-seq data automatically or manually. Currently, webSCST supports the integration of human and mouse datasets using the popular integration methods: AddModuleScore (Stuart *et al.*, 2019), multimodal intersection analysis (Moncada *et al.*, 2020), ssGSEA (Hänzelmann *et al.*, 2013) and robust cell type decomposition (Cable *et al.*, 2021). The webSCT pipeline is shown in Figure 1.

3 Discussion

webSCST is a freely available spatial transcriptomic database and web service that aims to conveniently allow users to obtain predicted spatial information using scRNA-seq data. Users with raw scRNA-seq data can utilize webSCST by quality-controlling the scRNA-seq data, matching with spatial datasets in the webSCST database and finally obtaining the integrated results through four popular integrating tools. In the future, we plan to add more publicly available published spatial sequencing data to our database and add newly developed integration tools to maintain webSCST as an up-to-date resource.

Funding

This work was supported by the National Natural Science Foundation of China [62131004, 62102064]; and the Special Science Foundation of Quzhou [2021D004].

Conflict of Interest: none declared.

References

Cable,D.M. *et al.* (2021) Robust decomposition of cell type mixtures in spatial transcriptomics. *Nat. Biotechnol.*, **40**, 517–526.

Dries,R. *et al.* (2021) Advances in spatial transcriptomic data analysis. *Genome Res.*, **31**, 1706–1718.

Fan,Z. *et al.* (2020) SpatialDB: a database for spatially resolved transcriptomes. *Nucleic Acids Res.*, **48**, D233–D237.

Hänzelmann,S. *et al.* (2013) GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics*, **14**, 7.

Hao,Y. *et al.* (2021) Integrated analysis of multimodal single-cell data. *Cell*, **184**, 3573–3587.e3529.

Longo,S.K. *et al.* (2021) Integrating single-cell and spatial transcriptomics to elucidate intercellular tissue dynamics. *Nat. Rev. Genet.*, **22**, 627–644.

Moncada,R. *et al.* (2020) Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas. *Nat. Biotechnol.*, **38**, 333–342.

Stuart,T. *et al.* (2019) Comprehensive integration of single-cell data. *Cell*, **177**, 1888–1902.e1821.

Wang,Z. *et al.* (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.*, **10**, 57–63.

Zhang,Z. *et al.* (2021) Goals and approaches for each processing step for single-cell RNA sequencing data. *Brief. Bioinformatics*, **22**, bbaa314.