Bioinformatics, 36(14), 2020, 4217–4219
doi: 10.1093/bioinformatics/btaa532
Advance Access Publication Date: 21 May 2020
Applications Note

Gene expression

scTPA: a web tool for single-cell transcriptome analysis of pathway activation signatures

Yan Zhang[1,†], Yaru Zhang[1,†], Jun Hu [2,†], Ji Zhang[1], Fangjie Guo[1], Meng Zhou[1], Guijun Zhang [2,*], Fulong Yu[1,*] and Jianzhong Su[1,*]

[1]School of Biomedical Engineering, School of Ophthalmology and Optometry and Eye Hospital, Wenzhou Medical University, Wenzhou 325011, Zhejiang, China and [2]College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, Zhejiang, China

*To whom correspondence should be addressed.
†The authors wish it to be known that, in their opinion, the first three authors should be regarded as Joint Authors.
Associate Editor: Anthony Mathelier

Received on March 9, 2020; revised on May 13, 2020; editorial decision on May 14, 2020; accepted on May 15, 2020

Abstract

Motivation: At present, a fundamental challenge in single-cell RNA-sequencing data analysis is functional interpretation and annotation of cell clusters. Biological pathways in distinct cell types have different activation patterns, which facilitates the understanding of cell functions using single-cell transcriptomics. However, no effective web tool has been implemented for single-cell transcriptome data analysis based on prior biological pathway knowledge.

Results: Here, we present scTPA, a web-based platform for pathway-based analysis of single-cell RNA-seq data in human and mouse. scTPA incorporates four widely-used gene set enrichment methods to estimate the pathway activation scores of single cells based on a collection of available biological pathways with different functional and taxonomic classifications. The clustering analysis and cell-type-specific activation pathway identification were provided for the functional interpretation of cell types from a pathway-oriented perspective. An intuitive interface allows users to conveniently visualize and download single-cell pathway signatures. Overall, scTPA is a comprehensive tool for the identification of pathway activation signatures for the analysis of single cell heterogeneity.

Availability and implementation: http://sctpa.bio-data.cn/sctpa.

Contact: sujz@wmu.edu.cn or yufulong421@gmail.com or zgj@zjut.edu.cn

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Single-cell RNA sequencing (scRNA-seq) technology has been widely used to characterize cell-to-cell heterogeneity. Single-cell transcriptome analysis enables novel and unexpected biological discoveries compared with traditional 'bulk' cell methods (Hwang *et al.*, 2018; Trapnell and Liu, 2016). Various computational methods have been developed for cell clustering, marker gene identification and visualization of single-cell RNA-sequencing (scRNA-seq) data (DeTomaso *et al.*, 2019; Stegle *et al.*, 2015). However, functional interpretation of cell clustering remains a challenge for scRNA-seq data analysis.

Pathways provide biological insights for disease subtype classification, functional annotation of cellular diversity and drug discovery (Gatza *et al.*, 2010; Pollen *et al.*, 2014). In single-cell studies, pathway activation analysis has become a powerful approach for the extraction of biologically relevant signatures to uncover the potential mechanisms of cell heterogeneity and dysfunction in human diseases (Moffitt *et al.*, 2018; Xiao *et al.*, 2019). For example, pathway signatures exhibit an activation difference in breast cancer (Chung *et al.*, 2017) and Alzheimer's disease cells (Grubman *et al.*, 2019) according to gene set enrichment analysis. However, there is a lack of online tools for the comprehensive analysis and visualization of single-cell transcriptome data based on prior biological pathway knowledge.

The present study developed scTPA, a web tool for single-cell transcriptome analysis of pathway activation signatures. scTPA provides an easy-to-use interface to view and download pathway activity scores (PASs), cell clustering, pathway signatures and the associated gene expression data, enabling an improved understanding of their potential functions from a pathway-oriented perspective.

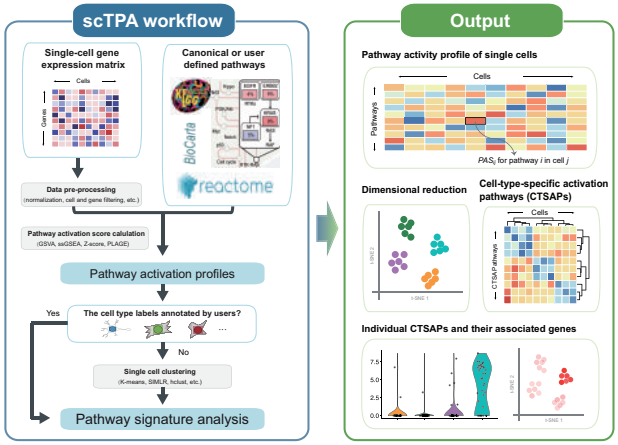

Fig. 1. Overview of the scTPA web tool

2 Materials and methods

2.1 ScTPA workflow

As shown in Figure 1, an input single-cell gene expression profile is preprocessed and the PAS matrix is calculated based on pathway signatures of interest using statistical methods (ssGSEA, GSVA, PLAGE and Z-scores). The cell type label can be provided either by the user or unsupervised clustering analysis of the PAS matrix, and the pathway signature analysis is performed for each cell type. The resulting data are passed on for single cell transcriptome dimension reduction, clustering, pathway signature identification and visualization. In addition, users can download various files, including the pathway activity profile, tSNE and UMAP plots in 2- and 3-dimensions and the gene expression heatmap of marker pathways.

2.2 Input and data processing

scRNA-seq profile. The required input of the scTPA tool is a single-cell gene expression matrix where columns represent cells and rows represent genes. The input file is the UMI or read count/RPKM/FPKM/TPM/CPM of single cells generated from different platforms, such as 10X genomics and Smart-seq. Four data normalization methods are available (Butler *et al.*, 2018; Lun *et al.*, 2016; Satija *et al.*, 2015). Additionally, users can choose to remove poor cells and genes. Furthermore, the scTPA tool provides the option to impute the missing values of genes.

Biological pathways. To facilitate the evaluation of pathway activation at single-cell resolution, the 'Canonical pathway' and 'Extended pathway' options (details can be seen in the Supplementary Material) were provided to enable users to select literature-curated pathways of interest. In addition, users can simultaneously upload their pathways of interest which are not cataloged by scTPA for specific scRNA-seq analysis.

PAS calculation. Four widely used methods, namely ssGSEA, GSVA, PLAGE and Z-scores, were incorporated into the scTPA tool to measure the activation of pathway signatures for single-cell transcriptomes. For a given pathway, these methods calculate the enrichment scores based on expression level rank statistics using the improved R package GSVA. To increase the computational efficiency, we rewrote the main loop function of GSVA to decrease run time by massive parallel scoring of pathway activation from a processed gene expression matrix (Supplementary Table S1).

Unsupervised cluster analysis. Unsupervised cluster analysis is a useful exploratory tool to analyze the heterogeneity of complex populations. If the cell type label file is not pre-defined by the user, scTPA provides six different clustering methods, such as K-means and hierarchical clustering, to cluster cells based on the PAS matrix. In addition, we provided the options of the main parameters for these clustering methods, including the number of clusters, resolution and dimensions of principal component analysis in the advanced setting panel.

Identification of cell-type-specific activation pathways (CTSAPs). Pathways signatures are important to unveil and characterize cell types and their functional states. Based on the PAS matrix of individual cells, multiple non-parametric statistical methods and fold-change analysis are available for CTSAPs identification. These may distinguish cell populations into the case-control groups consisting of the cell type of interest and other cell types, which helps the user to identify CTSAPs with statistically significant activation among different cell types.

3 Results

The web tool provides the visualization of the PAS matrix, dimensional reduction and cluster analysis and pathway signatures. Specifically, interactive tSNE and UMAP plots in 2D and 3D were provided to visualize the differences among cell populations. In addition, a heatmap plot was used to reveal the PAS profile of the significant activated pathway signatures in each cell type. For each CTSAP in the corresponding cell type, the tool provides a UMAP plot and a box plot to visualize the PAS distribution in different cell types. Furthermore, the heatmap for gene expression in the CTSAPs was provided to explore how transcriptional changes affect the pathway activation of various cell types. All software, algorithm and database resources that are integrated into scTPA are presented in Supplementary Table S2.

To illustrate the function and utility of scTPA, we used a dataset from a melanoma study (GSE72056) (Tirosh *et al.*, 2016) which covers a variety of non-malignant cell types. The results of the analysis demonstrated that scTPA could not only classify, annotate and interpret the functional outcomes for single-cell transcriptome data with known cell types (Supplementary Fig. S1), but also that scTPA could decipher the complex heterogeneity of cell populations with unknown cell types (Supplementary Fig. S2).

4 Discussion

Pathway activity analysis methods are useful for the extraction of novel functional and mechanistic insights from scRNA-seq data (Ding *et al.*, 2019). The web-based tool developed here, scTPA, provides an option for the computation of pathway activation scores of single cells using several well-established functional analysis methods. The performance of pathway analysis methods is sensitive to the selection of gene sets (Holland *et al.*, 2020). The user needs to manually optimize the suitable pathway gene set from the collected 'Canonical pathway' and 'Extended pathway' datasets according to the biological context of the sample set and their research interest. In addition, the integrated analysis of a large number of cell populations (>50 000 cells) is recommended using our stand-alone program (https://github.com/sulab-wmu/scTPA) due to long upload and run time of the web service. Our tool for in-depth single-cell transcriptome analysis will be continuously updated and improved in the future to make it easily accessible via a web interface.

Funding

This work was supported by National Natural Science Foundation of China [61903277, 61871294, 61873193, 61773346 and 61902352] and Science Foundation of Zhejiang Province [LR19C060001 and LZ20F030002].

Conflict of Interest: none declared.

References

Butler,A. *et al.* (2018) Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.*, **36**, 411–420.

Chung,W. *et al.* (2017) Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer. *Nat. Commun.*, **8**, 15081.

DeTomaso,D. *et al*. (2019) Functional interpretation of single cell similarity maps. *Nat. Commun*., **10**, 4376.

Ding,H. *et al*. (2019) Biological process activity transformation of single cell gene expression for cross-species alignment. *Nat. Commun*., **10**, 1–6.

Gatza,M.L. *et al*. (2010) A pathway-based classification of human breast cancer. *Proc. Natl. Acad. Sci. USA*, **107**, 6994–6999.

Grubman,A. *et al*. (2019) A single-cell atlas of entorhinal cortex from individuals with Alzheimer's disease reveals cell-type-specific gene expression regulation. *Nat. Neurosci*., **22**, 2087–2097.

Holland,C.H. *et al*. (2020) Robustness and applicability of transcription factor and pathway analysis tools on single-cell RNA-seq data. *Genome Biol*., **21**, 19.

Hwang,B. *et al*. (2018) Single-cell RNA sequencing technologies and bioinformatics pipelines. *Exp. Mol. Med*., **50**.

Lun,A.T.L. *et al*. (2016) A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor [version 2; referees: 3 approved, 2 approved with reservations]. *F1000Research*, **5**, 2122.

Moffitt,J.R. *et al*. (2018) Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. *Science*, **362**, 18.

Pollen,A.A. *et al*. (2014) Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. *Nat. Biotechnol*., **32**, 1053–1058.

Satija,R. *et al*. (2015) Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol*., **33**, 495–502.

Stegle,O. *et al*. (2015) Computational and analytical challenges in single-cell transcriptomics. *Nat. Rev. Genet*., **16**, 133–145.

Tirosh,I. *et al*. (2016) Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science*, **352**, 189–196.

Trapnell,C. and Liu,S., (2016) Single-cell transcriptome sequencing: recent advances and remaining challenges. *F1000Research*, **5**, 182.

Xiao,Z. *et al*. (2019) Metabolic landscape of the tumor microenvironment at single cell resolution. *Nat. Commun*., **10**, 3763.