SCANNER: A Web Server for Annotation, Visualization and Sharing of Single Cell RNA-seq Data

Guoshuai Cai[1], Feifei Xiao[2]

[1]Department of Environmental Health Sciences, [2]Department of Epidemiology and Biostatistics, Arnold School of Public Health, University of South Carolina, Columbia, SC 29208

Short title: A Web Server for scRNA-seq Data

Corresponding Authors:

Guoshuai Cai, PhD
Department of Environmental Health Sciences
Arnold School of Public Health
University of South Carolina
921 Assembly Street
Public Health Research Center 401C
Columbia, SC 29204
GCAI@mailbox.sc.edu
Phone: 803-777-4120

Feifei Xiao, PhD
Department of Epidemiology and Biostatistics
Arnold School of Public Health
University of South Carolina
915 Greene Street, Room 449
Columbia, SC 29204
xiaof@mailbox.sc.edu
Phone: 803-777-8936

Abstract

Motivation

In recent years, efficient single cell RNA sequencing (scRNA-seq) methods have been developed, enabling the transcriptome profiling of each single cell massively in parallel. Meanwhile, its high dimensionality brought in challenges in data modeling, analysis, visualization and interpretation. Various tools have been developed for fast browsing the transcriptome gene expression in single cells. However, their applications require that users have extensive knowledge of data properties, statistical modeling, data dimension reduction techniques and substantial training of computational skills. This brings obstacles for biologists to efficiently view, browse and interpret the data. Also, currently available tools are either missing the scalability for accommodating multiple datasets, or not offering easy data sharing, or ignoring group information for comparison or providing limited annotation capacity on gene functions and involved pathways.

Results

Here we published a user-friendly interactive web application, Single Cell Transcriptomics Annotated Viewer (SCANNER), as a public resource to equip the biologists and bioinformatician to share, analyze, visualize and interpret scRNA-seq data in a comprehensive, flexible and collaborative manner. It is effort-less without requirement on software setup or coding skills and enables an easy way to annotate, visualize and compare ontologies, pathways and functions in experimental groups on single cell basis. Also, it provides a user-friendly layout with side-by-side group-split view to compare experimental groups and equipped with multiple data interfaces for easy data sharing. In summary, SCANNER provides a useful way to share, visualize scRNA-seq data, as well as to annotate and interpret the analysis results.

Availability and implementation

SCANNER is available at https://www.thecailab.com/scanner/.

Contact

GCAI@mailbox.sc.edu or XIAOF@mailbox.sc.ecu

In recent years, efficient single cell RNA sequencing (scRNA-seq) methods have been developed, enabling the transcriptome profiling of each single cell massively in parallel (Kolodziejczyk, et al., 2015). It provides opportunities to identify cell types or clusters and their specific markers. It also provides insights into capturing of cell development trajectory, estimation of gene bursting parameters, investigation of communications among cells and solutions of other important biological questions (Hwang, et al., 2018). Meanwhile, its high dimensionality brought in challenges in data modeling, analysis, visualization and interpretation. A widely used strategy is to reduce the dimension in a low-dimension space of two- or three-dimensions using methods such as T-distributed Stochastic Neighbor Embedding (t-SNE) (van der Maaten and Hinton, 2008), Uniform Manifold Approximation and Projection (UMAP) (Diaz-Papkovich, et al., 2019) or Principle Component Analysis (PCA) (Jolliffe and Cadima, 2016). In this low dimension space, visualization and further interpretation of the data could be further performed. Tools such as SCV (Wang, et al., 2019) and Cerebro (Hillje, et al., 2019) have been developed for fast browsing and analyze transcriptome gene expression in single cells. However, these tools require local installation and implementation, as well as a multi-step data pre-processing including raw data processing, dimension reduction and data format standardization. Therefore, these local applications requires extensive knowledge of data properties, statistical modeling, dimension reduction techniques and substantial training of computational skills. This brings obstacles for biologists to freely browse the data and efficiently derive biological insights from it. Moreover, due to the rapid development and application of new technologies, the volume and complexity of data increase fast which requires a new effective and efficient gateway for storage, annotation and visualization. The Broad Institute Single Cell Portal (https://portals.broadinstitute.org/single_cell) and scRNASeqDB (Cao, et al., 2017) are available databases for scRNA-seq studies. However, their visualization and analysis functions

are limited. A new flexible web server with a user-friendly interface and comprehensive analysis and visualization functions is highly demanded for convenient data sharing and efficient communication between data analysis unit and biological experiment unit.

SCV is a useful local R Shiny application with five visualization modules, for data overview of identified clusters (Overview), gene expression (Expression), the distribution of gene expression in each cluster (Distribution), expression detection rate and size (Detection) and expression heatmap with hierarchical clustering (Similarity). Each module allows users to set parameters for plot generation and visualization. Based on SCV, we developed an user-friendly interactive web application, Single Cell Transcriptomics Annotated Viewer (SCANNER, https://www.thecailab.com/scanner/), as a public resource to equip the biologists and bioinformatician to share, analyze, visualize and interpret scRNA-seq data in a comprehensive, flexible and collaborative manner. SCANNER has a unique set of functions that are not available in existing tools, which is highlighted in Figure 1A. Specifically,

(1) SCANNER is a web server application, which is effort-less without requirement on software setup or coding skills. It enables the location-free and host-free work, which fulfills the demand for efficient communication and seamlessly collaboration. Also, it provides a framework of online database for scRNA-seq data. Currently, a published melanoma single cell RNA seq dataset (Tirosh, et al., 2016) is available in SCANNER for demonstration. Further, a large number of publicly available datasets will be processed and categorized in SCANNER to build a comprehensive database, which will provide a valuable online resource of single cell transcriptomics for public use.

(2) SCANNER annotates pathways or functions of gene sets and visualizes their activation status. The activation status of a particular pathway can be inferred by averaging expression abundance or expression ranks of its involved genes. Alternatively, it can be represented by an eigengene which is the first principle component of the expression matrix of

the gene set (Langfelder and Horvath, 2008). This function provides the opportunities to identify dysregulation that are significantly associated with a phenotype of interest on the level of pathway and ontology. Current analysis is available on MSigDB v.7.0 (Liberzon, et al., 2011) gene sets of Hallmark Collection, KEGG Pathway, Biocarta Pathway, Reacome Pathway, GO Biological Process, GO Cellular Component, GO Molecular Function, Oncogenic Signatures and Immunological Signatures.

(3) SCANNER provides a user-friendly layout with side-by-side group-split view to easily visualize and compare the single cell data from experimental groups. This function is highly demanded as many studies aim to detection difference between groups. For example, we analyzed the melanoma dataset (Tirosh, et al., 2016) and found that tumors from female melanoma patients had significantly larger proportions of cancer-associated fibroblasts (CAF) and endothelial cells than that from male melanoma patients (Fig. 1B). Correspondingly, in CAF and endothelial cells of the women patients, the fibroblast growth factor binding function was highly activated (Fig. 1C) and the fibroblast growth factor genes, FGF1 and FGF2 were highly expressed (Fig. S1). Given that CAF is a promise target to treat melanoma, this gender difference in CAF may provide a new hint to the reason for that male patients have a worse survival outcome compared to females (Joosse, et al., 2013; Scoggins, et al., 2006). Further large-scale studies are needed to understand the melanoma disparity of gender together with other factors including the mitotic rates of skin lesions (Farahi, et al., 2018), delay in seeking medical attention (Richard, et al., 1999), sex hormones (Smalley, 2018) and mutation burden (Gupta, et al., 2015).

(4) SCANNER provides multiple data interfaces which enables easy data sharing. Three options are available, by (1) SCANNER data object: users can generate SCANNER data object (which is based on Seurat (Butler, et al., 2018) object and compatible with SCV data object) following instructions as described in the SCV paper (Wang, et al., 2019) and then uploaded into SCV; (2) access credentials: users can contact the authors to load data into the

sever, with password access control option available (an demo account “auser” with password “nopass” has been generated); and (3) database: a comprehensive database which is the most efficient method for public data analysis will be available soon.

In summary, SCANNER provides an efficient way to share, visualize, annotate and interpret scRNA-seq data.

Acknowledge:

We thank Ben Torkian and Jun Zhou from Research Computing program of University of South Carolina for the kind assistant on gateway application, allocation and implementation. This study was supported by the NSF XSEDE Startup Allocation Award.

References

Butler, A.*, et al.* Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* 2018;36(5):411-420.
Cao, Y.*, et al.* scRNASeqDB: A Database for RNA-Seq Based Gene Expression Profiles in Human Single Cells. *Genes (Basel)* 2017;8(12).
Diaz-Papkovich, A.*, et al.* UMAP reveals cryptic population structure and phenotype heterogeneity in large genomic cohorts. *PLoS Genet* 2019;15(11):e1008432.
Farahi, J.M.*, et al.* Gender differences in melanoma prognostic factors. *Dermatol Online J* 2018;24(4).
Gupta, S.*, et al.* Gender Disparity and Mutation Burden in Metastatic Melanoma. *J Natl Cancer Inst* 2015;107(11).
Hillje, R., Pelicci, P.G. and Luzi, L. Cerebro: Interactive visualization of scRNA-seq data. *Bioinformatics* 2019.
Hwang, B., Lee, J.H. and Bang, D. Single-cell RNA sequencing technologies and bioinformatics pipelines. *Exp Mol Med* 2018;50(8):96.
Jolliffe, I.T. and Cadima, J. Principal component analysis: a review and recent developments. *Philos Trans A Math Phys Eng Sci* 2016;374(2065):20150202.
Joosse, A.*, et al.* Sex is an independent prognostic indicator for survival and relapse/progression-free survival in metastasized stage III to IV melanoma: a pooled analysis of five European organisation for research and treatment of cancer randomized controlled trials. *J Clin Oncol* 2013;31(18):2337-2346.
Kolodziejczyk, A.A.*, et al.* The technology and biology of single-cell RNA sequencing. *Mol Cell* 2015;58(4):610-620.

Langfelder, P. and Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008;9:559.
Liberzon, A.*, et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* 2011;27(12):1739-1740.
Richard, M.A.*, et al.* Melanoma and tumor thickness: challenges of early diagnosis. *Arch Dermatol* 1999;135(3):269-274.
Scoggins, C.R.*, et al.* Gender-related differences in outcome for melanoma patients. *Ann Surg* 2006;243(5):693-698; discussion 698-700.
Smalley, K.S. Why do women with melanoma do better than men? *Elife* 2018;7.
Tirosh, I.*, et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* 2016;352(6282):189-196.
van der Maaten, L. and Hinton, G. dimensionality_reduction visualization. *Journal of Machine Learning Research* 2008;9:2579-2605.
Wang, S.*, et al.* Single Cell Viewer (SCV): An interactive visualization data portal for single cell RNA sequence data. *bioRxiv* 2019:664789.

FIGURES

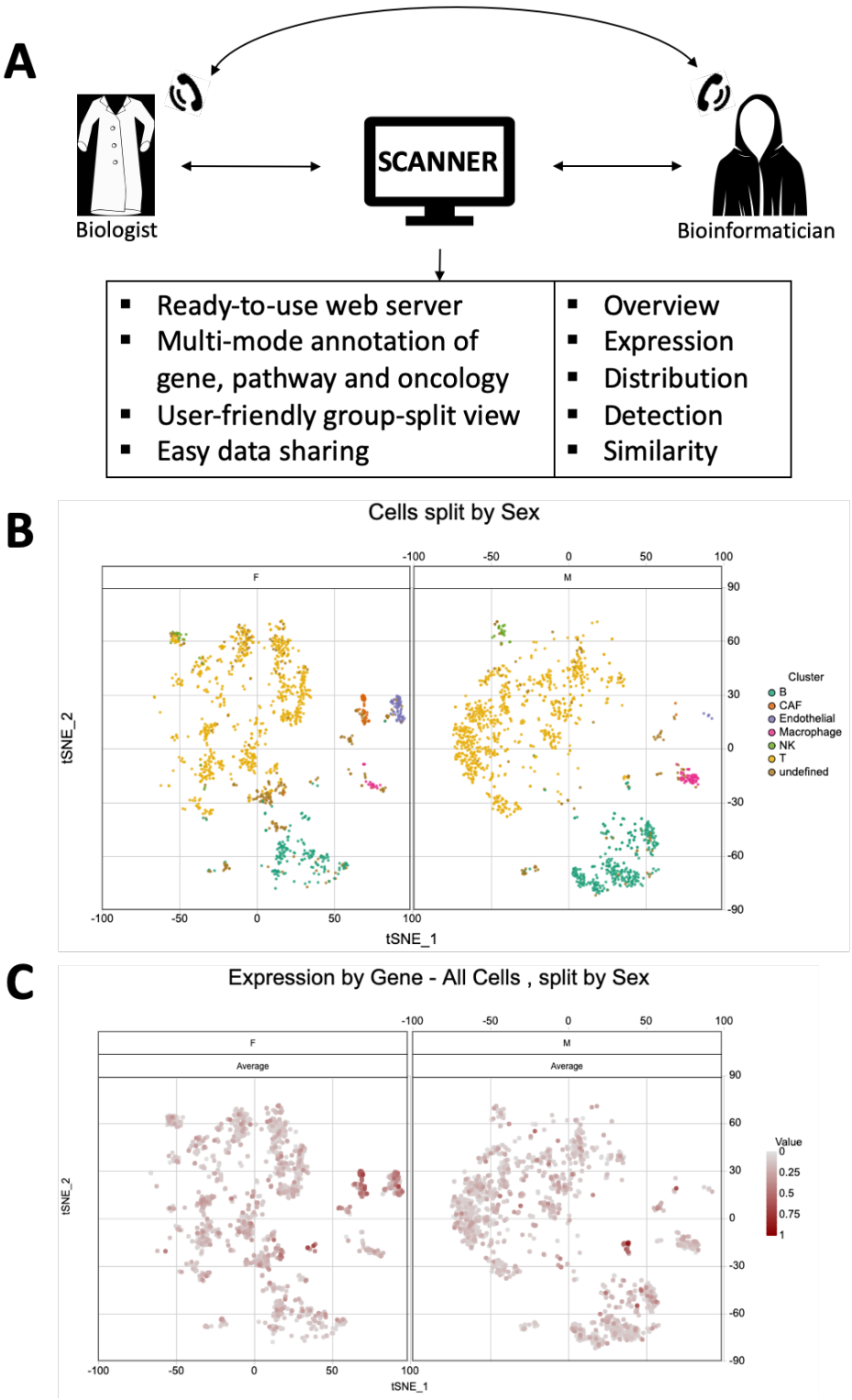

Figure 1. SCANNER application scene and example. A. Scanner facilitates the sharing, visualization, analysis and interpretation of scRNA-seq data and the communication between biologists and bioinformatician in a flexible, multi-functional and user-friendly manner. **B.** the side-by-side view of cell population from samples of male and female patients. **C.** the side-by-side view of average expression of

the GO: fibroblast growth factor binding function related genes from samples of male and female patients.

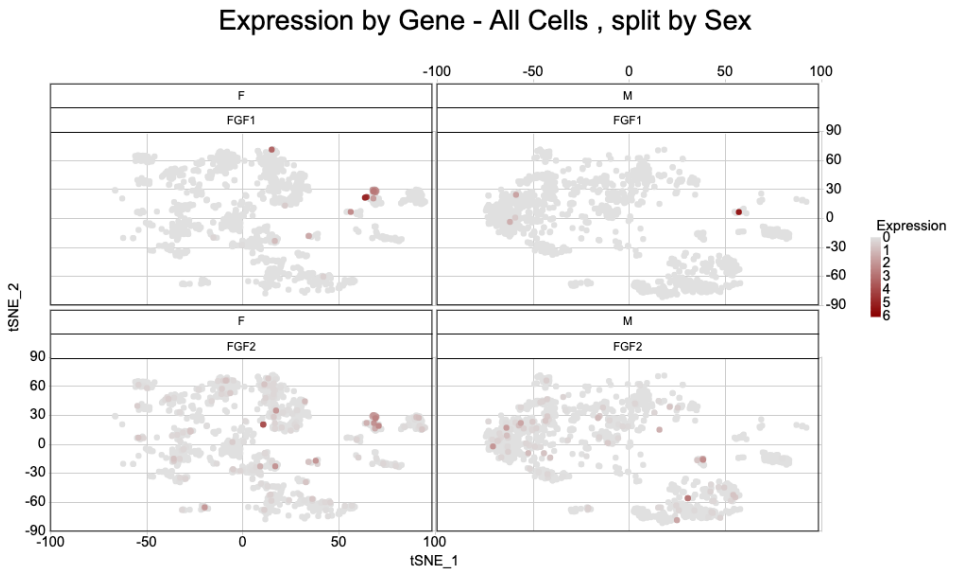

Figure S1. The side-by-side view of expression of FGF1 and FGF2 genes from samples of male and female patients.