BITS :: Call for Abstracts 2023 - Oral communication

Type	Oral communication
Session	Single-cell data analysis
Title	CClens: effective and efficient differential cellular communication analysis of large-scale single-cell RNA sequencing data
All Authors	Cesaro G(1), Baruzzo G(1), Di Camillo B(1,2)

Affiliation

(1) Department of Information Engineering, University of Padova, Padova, Italy
(2) Department of Comparative Biomedicine and Food Science, University of Padova, Padova, Italy

Motivation

The recent advance in single-cell transcriptomics has enabled the study of cellular communication, a complex multicellular mechanism that governs many biological processes. Several computational tools have been proposed to infer ligand-receptor interactions from single-cell RNA sequencing (scRNA-seq) data, but very few of them also investigate and quantify the downstream intracellular signaling and identify differences in communication across distinct experimental contexts (1,2). Moreover, as scRNA-seq has become cheaper, widespread and accessible, the availability of large-scale studies and cell atlases of growing complexity (different conditions and/or subjects and time series studies) has been increased, posing the challenges of computational demand, visualization and interpretation of cell-cell communication analysis.
Therefore, there is the need of a generalizable and scalable workflow to perform and support the interpretation of cellular communication analysis from large-scale scRNA-seq data in a user-friendly, efficient, and effective way. For this reason, we developed CClens, a bioinformatics pipeline that allows to i) quantify and characterize cell-cell communication in distinct contexts (i.e. multiple experimental condition or multiple patient scenarios) at both inter- and intra-cellular level; ii) infer differential cellular crosstalk across contexts through statistical methods; iii) perform fast and memory-efficient analysis of complex and multi-gigabyte dataset exploiting cross-language interoperability (i.e. R and C++); iv) support interpretation and exploration of cell-cell communication data through an interactive and user-friendly interface.

Methods

The method requires as input a scRNA-seq normalized gene expression matrix, cell cluster assignment (i.e. cell type annotation) and, if available, data about samples and/or patients (i.e. multi-condition and/or multi-patient scenario, respectively). To quantify cellular communication in each scenario, users can choose their preferred intercellular scoring scheme among scSeqComm, SingleCellSignalR, Zhou and Skelly score (3–6). scSeqComm intracellular score, the only available score to the best of our knowledge, is implemented to characterize the effect of intracellular signaling in terms of: i) association between receptors and transcription factors (TF) based on Personalized PageRank algorithm applied to biological pathways; ii) activity of TFs based on the expression levels of their target genes.
Significant differential intercellular communication is assessed using a permutation test in multi-condition scenario, by randomly shuffling the experimental label of each cell to obtain the null hypothesis, or, in case of multiple patient's data, using Wilcoxon Mann-Whitney test on patient-specific scores. Differential intracellular crosstalk is assessed by measuring, in each pathway, differentially expressed TF's target genes, exploiting already existing approaches (e.g. MAST or pseudobulk approach (7,8)).
To handle the increased computational burden of the analysis, the proposed method implements in-memory computation through the bigmemory R package (9) and exploits C++ efficiency through the library Rcpp (10). The shared-memory framework from the bigmemory package and its interface with R and C++ through Rcpp allow a very effective use of the memory and enable shared-memory parallelism to improve speed.
Moreover, an interactive R/shiny platform assists user in the interpretation and exploration of multi-dimensional cell-cell communication data. It includes i) multiple filtering options to dynamically and interactively inspect data, ii) a powerful and effective visualization framework for summarizing and interpreting communication data, and iii) advanced visualization tools to analyse multi-condition and multi-patient datasets on all their dimensions (e.g. inter- and intra-cellular signaling).

Results

To appreciate the various facets of the method, we applied it to a large publicly available (GSE174332) single-nucleus dataset of human primary motor cortex (~242,000 nuclei) from a cohort of 23 amyotrophic lateral sclerosis and 17 pathologically normal patients (11). The use of the R/shiny app enables the characterization of cellular crosstalk in each distinct condition (or patient), and quantitative and qualitative comparison of communication patterns, revealing alterations which might be associated to neurodegeneration.
Moreover, the efficient in-memory implementation and the shared-memory parallelism allow to process dataset as big as entire cells atlas, i.e. matrixes that cannot even be represented with vanilla R matrix due to R memory limit. For example, we successfully perform differential cellular communication analysis of "LUCA Single-cell Lung Cancer Atlas" (12) (a cell atlas comprising ~900k cells from 318 patients with 14 distinct cell types) with relative low execution time and small memory footprint (~200 minutes and ~110 GB with 4 cores @2.1 GHz).

Info

References

1. Liu Z, Sun D, Wang C. Evaluation of cell-cell interaction methods by integrating single-cell RNA sequencing data with spatial information. Genome Biol [Internet]. 2022 Dec 1 [cited 2023 Apr 27];23(1):1–38. Available from: https://genomebiology.biomedcentral.com/articles/10.1186/s13059-022-02783-y
2. Armingol E, Officer A, Harismendy O, Lewis NE. Deciphering cell–cell interactions and communication from gene expression. Nat Rev Genet 2020 222 [Internet]. 2020 Nov 9 [cited 2023 Apr 20];22(2):71–88. Available from: https://www.nature.com/articles/s41576-020-00292-x
3. Baruzzo G, Cesaro G, Di Camillo B. Identify, quantify and characterize cellular communication from single cell RNA sequencing data with scSeqComm. Bioinformatics [Internet]. 2022 Mar 28 [cited 2022 Apr 21];38(7):1920–9. Available from: https://pubmed.ncbi.nlm.nih.gov/35043939/
4. Cabello-Aguilar S, Alame M, Kon-Sun-Tack F, Fau C, Lacroix M, Colinge J. SingleCellSignalR: inference of intercellular networks from single-cell transcriptomics. Nucleic Acids Res [Internet]. 2020 Jun 4 [cited 2022 Apr 26];48(10):e55–e55. Available from: https://academic.oup.com/nar/article/48/10/e55/5810485
5. Zhou JX, Taramelli R, Pedrini E, Knijnenburg T, Huang S. Extracting Intercellular Signaling Network of Cancer Tissues using Ligand-Receptor Expression Patterns from Whole-tumor and Single-cell Transcriptomes. Sci Rep [Internet]. 2017 Dec 1 [cited 2022 Apr 28];7(1). Available from: https://pubmed.ncbi.nlm.nih.gov/28821810/
6. Skelly DA, Squiers GT, McLellan MA, Bolisetty MT, Robson P, Rosenthal NA, et al. Single-Cell Transcriptional Profiling Reveals Cellular Diversity and Intercommunication in the Mouse Heart. Cell Rep [Internet]. 2018 Jan 16 [cited 2022 Apr 28];22(3):600–10. Available from: https://pubmed.ncbi.nlm.nih.gov/29346760/
7. Finak G, McDavid A, Yajima M, Deng J, Gersuk V, Shalek AK, et al. MAST: A flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. Genome Biol [Internet]. 2015 Dec 10 [cited 2022 Oct 28];16(1):1–13. Available from: https://genomebiology.biomedcentral.com/articles/10.1186/s13059-015-0844-5
8. Squair JW, Gautier M, Kathe C, Anderson MA, James ND, Hutson TH, et al. Confronting false discoveries in single-cell differential expression. Nat Commun 2021 121 [Internet]. 2021 Sep 28 [cited 2022 Nov 4];12(1):1–15. Available from: https://www.nature.com/articles/s41467-021-25960-2
9. Kane MJ, Emerson JW, Weston S. Scalable Strategies for Computing with Massive Data. J Stat Softw [Internet]. 2013 Nov 20 [cited 2022 Oct 28];55(14):1–19. Available from: https://www.jstatsoft.org/index.php/jss/article/view/v055i14
10. Eddelbuettel D, Balamuta JJ. Extending R with C++: A Brief Introduction to Rcpp. Am Stat [Internet]. 2018 Jan 2 [cited 2022 Nov 4];72(1):28–36. Available from: https://amstat.tandfonline.com/doi/abs/10.1080/00031305.2017.1375990
11. Pineda SS, Lee H, Fitzwalter BE, Mohammadi S, Pregent LJ, Gardashli ME, et al. Single-cell profiling of the human primary motor cortex in ALS and FTLD. bioRxiv [Internet]. 2021 Jul 7 [cited 2023 Apr 20];2021.07.07.451374. Available from: https://www.biorxiv.org/content/10.1101/2021.07.07.451374v1
12. Salcher S, Sturm G, Horvath L, Untergasser G, Kuempers C, Fotakis G, et al. High-resolution single-cell atlas reveals diversity and plasticity of tissue-resident neutrophils in non-small cell lung cancer. Cancer Cell. 2022 Dec 12;40(12):1503-1520.e8.

filename	-
Figure	
-	
Availability	https://gitlab.com/sysbiobig/cclens

Corresponding Author

Name, Surname	Barbara, Di Camillo
Email	barbara.dicamillo@unipd.it
Submitted on	27.04.2023