APackOfTheClones: Visualization of clonal expansion with circle packing

Qile Yang [1]

1 University of California, Berkeley, Berkeley, CA 94720, United States of America

Summary

T and B lymphocytes exhibit diverse surface receptor repertoires that interact with antigens to communicate and protect the body from foreign invaders. (Haan et al., 2014) Modern methods allows the experimental characterization of the collective identity of the immune cells in tissue samples by sequencing RNA transcripts and cell receptors with scRNA-seq and scTCR-seq / scBCR-seq. Such methods generate high-dimensional data which are analyzed with the aid of dimensional reduction techniques to allow annotation of cell subsets. (Huang et al., 2022)

The R (R Core Team, 2023) package APackOfTheClones implements and *extends* a novel method to visualize clonal expansion (abbreviated as CE) at a single-cell (abbreviated as SC) resolution using circle packing, along with many clonal analysis utilities. Clonotype frequencies for each cell subset is counted and the values are used as radii and packed into one circular cluster with the largest circles near the center, and overlaid onto a the centroids of each cell subset on the corresponding 2D dimensional reduction plot. The package easily integrates into existing analysis pipelines using the *Seurat* (Hao et al., 2023) and *scRepertoire* (Borcherding & Bormann, 2020) packages. The original implementation was created in the julia language (Bezanson et al., 2017) by Christian Murray and Ben Murrell and utilized in two immunology papers. (Ma et al., 2021; Stark et al., 2022)

Statement of need

Cells in scRNA-seq experiments are conventionally visualized on a reduction scatterplot where each point represents a cell after all its features have been projected into two dimensions. The points can be colored by different factors to display useful information, including its assigned identity. (Andrews et al., 2021)

Understanding the role of specific cells in various contexts requires understanding of relationships between cell populations and their behaviors. One attribute is the clonal expansion dynamics, which are inferred from downstream analyses. (Adams et al., 2020) While it is simple to visualize aggregate CE information (for example with bar plots of the frequencies of different clonotypes), overlaying it on a by-clonotype basis onto the reduction scatterplot of the cells of an experiment allows for a more intuitive understanding of how clonal dynamics relate to the identified cellular subsets. For example, it can help gauge the presence of hyper expanded clones for each cell type; compare potential changes in frequencies after certain therapeutic treatments, etc.

There is no standardized convention to visualize this on the dimensional reduction. Some of the current approaches include 1. Using a color gradient corresponding to each frequency to highlight each individual point by the CE, implemented in *scRepertoire* and *scirpy* (Sturm et al., 2020) 2. Overlaying a 2D contour where points representing clones with higher frequencies

have elevated levels ([Andreatta et al., 2023](#)) 3. Increasing sizes of points based on clonal frequency, used in figure 2c of Wang et al. ([2021](#)).

From a visual standpoint, these approaches capture approximate trends that lack precision in depicting the true diversity and abundance of clonal populations for every cell subset. For instance, using a color gradient or contour lines to depict clones fails to directly differentiate between different clonotypes within the same cell subset. This is particularly problematic in scenarios where identifying rare clonotypes is essential, such as detecting the presence of emerging clones in response to a therapy or tracking the evolution of immune responses in different contexts.

APackOfTheClones offers an approach to address these issues by representing exact sizes of each clonotype, in a manner that corresponds exactly to the relevant cell subset. This level of granularity is helpful uncovering hidden patterns, identifying rare clonal populations, and precisely quantifying the impact of therapeutic interventions on immune responses. The original julia implementation's successful utilization in two publication is promising evidence in its potential ([Ma et al., 2021](#); [Stark et al., 2022](#)) - however, language choice and the lack of a user-friendly interface limited its accessibility and integration into existing pipelines. The integration of this approach within Seurat and scRepertoire in the popular R ecosystem bridges this gap, increasing accessibility to the wider bioinformatics community through a straightforward, familiar interface.

APackOfTheClones also offers a suite of methods for visualizing and analyzing SCCE data. Novelty features include functions for highlighting certain clones, and the filtering and visualization of clonotypes shared between subsets by linking circles on the APackOfTheClones clonal expansion plot.

Results

An example of the main CE visualization that APackOfTheClones can produce is shown in the following [Figure 1](#), using scRNA-seq + scTCR-seq data from Borcherding et al. ([2021](#)). The steps needed to generate the plot can be found in the [supplemental repository](#).
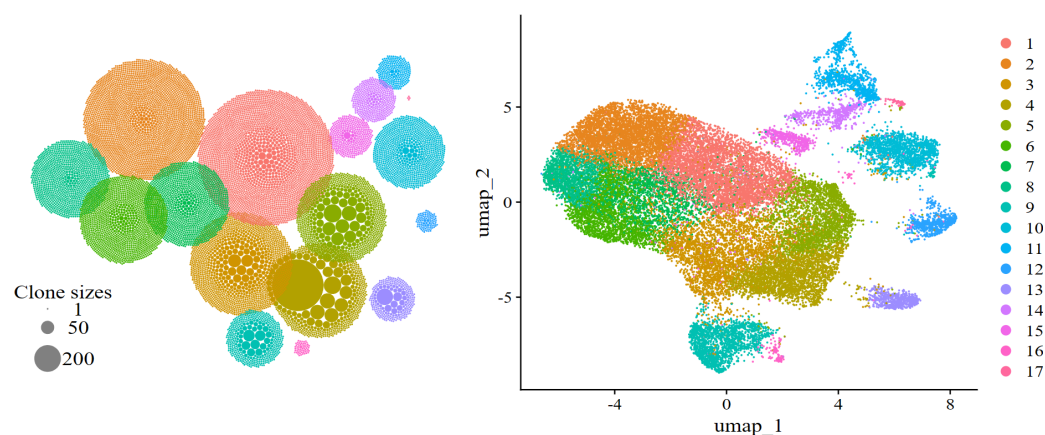

Figure 1: A scRNA-seq experiment's dimensional reduction projection and its corresponding APackOfTheClones clonal expansion plot. The plot on the right is the projection of all cells of its *Seurat* object based on Uniform Manifold Approximation and Projection (UMAP), colored by unknown identities 1 through 17. On the left is the clonal expansion plot generated by APackOfTheClones on the same data. Each cell identity on the seurat object corresponds to a cluster of circles with a similar geometric placement, and the size of each individual circle is the clonotype frequency of some clonotype within that cell subset. Note that the largest clones are packed near the origins of each cluster to accentuate their difference with the rest of all clonotypes.

The visualization gives the immediate insight that certain cell subsets such as those in cluster four contains many more expanded clones both by quantity and frequency.

The package extends objects and the functionality of the *Seurat* and *scRepertoire* package, and given a correctly processed seurat object of scRNA-seq data that was combined with paired TCR/BCRs, only a few functions need to be used to as little or as much customization of function arguments as needed to produce a `ggplot` object (Wickham, 2016) that fits into the conventional plotting ecosystem of R. Functions are accelerated with a `c++` layer via the *Rcpp* package (Eddelbuettel & François, 2011) to deliver all plots and R objects quickly in time complexity roughly linearly proportional to the number of cells, with the main time bottleneck being the plot display time.

Conclusion

APackOfTheClones offers a fast, and simple interface to produce an intuitive, easily extendible, and *publication ready* visualization of CE on a cell-by-cell basis, and slots seamlessly into existing analysis pipelines. It can be a useful sub-figure in any immunological/therapeutic study involving single-cell omics and immune repertoire to provide an additional degree of understanding for readers and researchers like. However, it should be noted that precise statistical/biological statements about clonal dynamics still obviously require rigorous analysis.

Acknowledgements

Supplemental Materials

https://github.com/Qile0317/APackOfTheClonesAppendix

References

Adams, N. M., Grassmann, S., & Sun, J. C. (2020). Clonal expansion of innate and adaptive lymphocytes. *Nature Reviews Immunology*, *20*(11), 694–707. https://doi.org/10.1038/s41577-020-0307-4

Andreatta, M., Gueguen, P., Borcherding, N., & Carmona, S. J. (2023). T cell clonal analysis using single-cell RNA sequencing and reference maps. *Bio-Protocol*. https://doi.org/10.21769/BioProtoc.4735

Andrews, T. S., Kiselev, V. Y., McCarthy, D., & Hemberg, M. (2021). Tutorial: Guidelines for the computational analysis of single-cell RNA sequencing data. *Nature Protocols*, *16*(1), 1–9. https://doi.org/10.1038/s41596-020-00409-w

Bezanson, J., Edelman, A., Karpinski, S., & Shah, V. B. (2017). Julia: A fresh approach to numerical computing. *SIAM Review*, *59*(1), 65–98. https://doi.org/10.1137/141000671

Borcherding, N., & Bormann, N. L. (2020). *scRepertoire: An r-based toolkit for single-cell immune receptor analysi [version 2; peer review: 2 approved]*. https://doi.org/10.12688/f1000research.22139.2

Borcherding, N., Vishwakarma, A., Voigt, A. P., Bellizzi, A., Kaplan, J., Nepple, K., Salem, A. K., Jenkins, R. W., Zakharia, Y., & Zhang, W. (2021). Mapping the immune environment

in clear cell renal carcinoma by single-cell genomics. *Communications Biology*, *4*(1), 122. https://doi.org/10.1038/s42003-020-01625-6

Eddelbuettel, D., & François, R. (2011). Rcpp: Seamless R and C++ integration. *Journal of Statistical Software*, *40*(8), 1–18. https://doi.org/10.18637/jss.v040.i08

Haan, J. M. den, Arens, R., & Zelm, M. C. van. (2014). The activation of the adaptive immune system: Cross-talk between antigen-presenting cells, t cells and b cells. *Immunology Letters*, *162*(2), 103–112. https://doi.org/10.3389/fimmu.2019.00360

Hao, Y., Stuart, T., Kowalski, M. H., Choudhary, S., Hoffman, P., Hartman, A., Srivastava, A., Molla, G., Madad, S., Fernandez-Granda, C., & others. (2023). Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nature Biotechnology*, 1–12. https://doi.org/10.1038/s41587-023-01767-y

Huang, Y. V., Waliany, S., Lee, D., Galdos, F. X., Witteles, R. M., Neal, J. W., Fan, A. C., Maecker, H. T., Nguyen, P. K., Wu, S. M., & others. (2022). The role of single-cell profiling and deep immunophenotyping in understanding immune therapy cardiotoxicity. *Cardio Oncology*, *4*(5), 629–634. https://doi.org/10.1016/j.jaccao.2022.08.012

Ma, J., Tibbitt, C. A., Georén, S. K., Christian, M., Murrell, B., Cardell, L.-O., Bachert, C., & Coquet, J. M. (2021). Single-cell analysis pinpoints distinct populations of cytotoxic CD4+ t cells and an IL-10+ CD109+ TH2 cell population in nasal polyps. *Science Immunology*, *6*(62), eabg6356. https://doi.org/10.1126/sciimmunol.abg6356

R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Stark, J. M., Liu, J., Tibbitt, C. A., Christian, M., Ma, J., Wintersand, A., Dunst, J., Kreslavsky, T., Murrell, B., Adner, M., & others. (2022). Recombinant multimeric dog allergen prevents airway hyperresponsiveness in a model of asthma marked by vigorous TH2 and TH17 cell responses. *Allergy*, *77*(10), 2987–3001. https://doi.org/10.1111/all.15399

Sturm, G., Szabo, T., Fotakis, G., Haider, M., Rieder, D., Trajanoski, Z., & Finotello, F. (2020). Scirpy: A scanpy extension for analyzing single-cell t-cell receptor-sequencing data. *Bioinformatics*, *36*(18), 4817–4818. https://doi.org/10.37473/dac/10.1101/2020.04.10.035865

Wang, P., Yao, L., Luo, M., Zhou, W., Jin, X., Xu, Z., Yan, S., Li, Y., Xu, C., Cheng, R., & others. (2021). Single-cell transcriptome and TCR profiling reveal activated and expanded t cell populations in parkinson's disease. *Cell Discovery*, *7*(1), 52. https://doi.org/10.1038/s41421-021-00280-3

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. ISBN: 978-3-319-24277-4