

# Supplementary Information

## 1. Supplementary Note

A clonotype designates a collection of T or B cells that bear the same adaptive immune receptors, and thus can be regarded as descendants of a common, antecedent cell, that recognize the same epitopes. In single-cell sequencing data, T cells sharing identical complementarity-determining regions 3 (CDR3) sequences of both alpha and beta TCR chains make up a clonotype.

Contrary to what would be expected based on the previously described mechanism of allelic exclusion [1], single-cell sequencing datasets feature a considerable number of cells with more than one TCR alpha and beta pair. Since their existence violates current canon, most TCR analysis tools ignore cells with more than one productive CDR3 sequence [2, 3] or select the CDR3 sequence with the highest expression level [4]. While in some cases these double-TCR cells might represent artifacts (e.g. doublets of a CD8+ and a CD4+ T cell engaged in an immunological synapse), there is an increasing amount of evidence in support of a bona fide dual-TCR population [5, 6].

Scirpy allows to investigate the composition and phenotypes of both, single- and dual-TCR T cells by leveraging on a T cell model similar to the one suggested by clonotype networks in TraCeR [7], where T cells are allowed to have a primary and a secondary pair of alpha and beta chains. The primary pair consists of the alpha chain and beta chain with the highest read count. Likewise, the secondary pair is the pair of alpha-beta chains with the second highest expression level. If more than two variants of a chain are recovered for the same cell, those are ignored based on the assumption that each cell has only two copies of the underlying chromosome set. Scirpy flags these cells as 'multichain', allowing the user to possibly discard them from downstream analysis. The user can also choose if secondary TCRs shall be included in the analysis or not.

Another decision when implementing a clonotype definition is whether it should be based on the nucleotide or amino acid sequence. Although nucleotide sequence reflects common origin, epitope reactivity is determined by the amino acid sequence. This has made the amino acid sequence of the CDR3 region a common choice for clonotype definition by existing tools and Scirpy conforms to this consensus. Given both the nucleotide and amino acid sequences of the CDR3 regions, it is also possible to analyse the probability of convergent evolution of TCR sequences toward a given amino acid sequence as a result of selection pressure by epitopes.

Individual cells can be grouped together into clonotype networks based on:

- (i) shared CDR3 amino acid sequences (despite differences at the DNA level);
- (ii) shared primary or secondary chain pairs;
- or (iii) shared alpha or shared beta chains.

The biological relevance of such networks would be the clustering of cells that recognize the same epitope. Inspired by recent papers stating the similar TCR sequences also share epitope targets [2, 8, 9], we aimed at implementing a more general approach, where different layers of TCR sequence similarity can be integrated into a global, epitope-focused cell similarity network. The possibility of linking similar clonotypes together also provides an opportunity to limit the impact of sequencing bias resulting in the loss of an alpha or a beta chain for some cells.

Tool name	Website	Language	Receptor type	Recommended for scSeq	Paired chains	Post-processing of clonotypes	Basic IR visualization	Advanced IR visualization	Clonotype clusters of single cells	Integrated with Gex
Immcantation (pRESTO) [10]	<a href="https://immcantation.readthedocs.io">https://immcantation.readthedocs.io</a>	Python (and R)	TCR	No	No	Yes	Yes	Yes	No	No
ImmuneArch (tcR) [11]	<a href="https://immunarch.com/">https://immunarch.com/</a>	R	TCR	Yes	No	Yes	Yes	Yes	No	No
iMonitor [12]	<a href="https://github.com/zhangwei2015/IMonitor">https://github.com/zhangwei2015/IMonitor</a>	Perl (and R)	TCR	No	No	Yes	Yes	No	No	No
MIXCR [13]	<a href="https://github.com/milaboratory/mixcr">https://github.com/milaboratory/mixcr</a>	Java	TCR	Yes	No	No	No	No	No	No
VDJtools [14]	<a href="https://github.com/mikessh/vdjtools">https://github.com/mikessh/vdjtools</a>	Java (Python)	TCR	No	No	Yes	Yes	Yes	No	No
scTCR Seq [15]	<a href="https://github.com/ElementoLab/scTCRseq">https://github.com/ElementoLab/scTCRseq</a>	Python	TCR	Yes	Yes	Yes	Yes	No	No	No
TraCeR [16]	<a href="https://github.com/teichlab/tracer">https://github.com/teichlab/tracer</a>	Python	TCR	Yes	Yes	Yes	Yes	No	Yes	Some
VDJ Puzzle [17]	<a href="https://github.com/simone-rizzetto/VDJPuzzle">https://github.com/simone-rizzetto/VDJPuzzle</a>	Java	TCR	Yes	No	Yes	Yes	No	No	Some
TRAPes [16]	<a href="https://github.com/YosefLab/TRAPes">https://github.com/YosefLab/TRAPes</a>	C++	TCR	Yes	No	No	No	No	No	No
scirpy	<a href="https://github.com/icbi-lab/scirpy">https://github.com/icbi-lab/scirpy</a>	Python	TCR	Yes	Yes	Yes	Yes	Yes	Yes	Yes

**Supplementary Table 1:** Comparison of scirpy with currently available tools for immune repertoire analysis and visualization at the single cell level. Basic Immune Repertoire visualization includes clonotype abundance, diversity and VDJ usage. Advanced visualization adds repertoire overlap, clustering and specialized analysis of individual clonotypes.

## Supplementary References

1. Brady, B. L., Steinell, N. C. & Bassing, C. H. Antigen receptor allelic exclusion: an update and reappraisal. en. *J. Immunol.* **185**, 3801–3808 (Oct. 2010).
2. Fischer, D. S., Wu, Y., Schubert, B. & Theis, F. J. Predicting antigen-specificity of single T-cells based on TCR CDR regions.
3. Zhang, L. *et al.* Lineage tracking reveals dynamic relationships of T cells in colorectal cancer. en. *Nature* **564**, 268–272 (Dec. 2018).
4. Afik, S. *et al.* Targeted reconstruction of T cell receptor sequence from single cell RNA-seq links CDR3 length to T cell differentiation state. en. *Nucleic Acids Res.* **45**, e148 (Sept. 2017).
5. Schuldt, N. J. & Binstadt, B. A. Dual TCR T Cells: Identity Crisis or Multitaskers? en. *J. Immunol.* **202**, 637–644 (Feb. 2019).
6. Ji, Q., Perchellet, A. & Gorman, J. M. Viral infection triggers central nervous system autoimmunity via activation of CD8+ T cells expressing dual TCRs. en. *Nat. Immunol.* **11**, 628–634 (July 2010).
7. Stubbington, M. J. T. *et al.* T cell fate and clonality inference from single-cell transcriptomes. en. *Nat. Methods* **13**, 329–332 (Apr. 2016).
8. Glanville, J. *et al.* Identifying specificity groups in the T cell receptor repertoire. en. *Nature* **547**, 94–98 (July 2017).
9. Dash, P. *et al.* Quantifiable predictive features define epitope-specific T cell receptor repertoires. en. *Nature* **547**, 89–93 (July 2017).
10. Vander Heiden, J. A. *et al.* pRESTO: a toolkit for processing high-throughput sequencing raw reads of lymphocyte receptor repertoires. en. *Bioinformatics* **30**, 1930–1932 (July 2014).
11. Nazarov, V. I. *et al.* tcR: an R package for T cell receptor repertoire advanced data analysis. en. *BMC Bioinformatics* **16**, 175 (May 2015).
12. Zhang, W. *et al.* IMonitor: A Robust Pipeline for TCR and BCR Repertoire Analysis. en. *Genetics* **201**, 459–472 (Oct. 2015).
13. Bolotin, D. A. *et al.* MiXCR: software for comprehensive adaptive immunity profiling. en. *Nat. Methods* **12**, 380–381 (May 2015).
14. Shugay, M. *et al.* VDJtools: Unifying Post-analysis of T Cell Receptor Repertoires. en. *PLoS Comput. Biol.* **11**, e1004503 (Nov. 2015).
15. Redmond, D., Poran, A. & Elemento, O. Single-cell TCRseq: paired recovery of entire T-cell alpha and beta chain transcripts in T-cell receptors from single-cell RNAseq. en. *Genome Med.* **8**, 80 (July 2016).
16. Joshi, K. *et al.* Spatial heterogeneity of the T cell receptor repertoire reflects the mutational landscape in lung cancer. en. *Nat. Med.* (Oct. 2019).
17. Rizzetto, S. *et al.* B-cell receptor reconstruction from single-cell RNA-seq with VDJPUZZLE. en. *Bioinformatics* **34**, 2846–2847 (Aug. 2018).