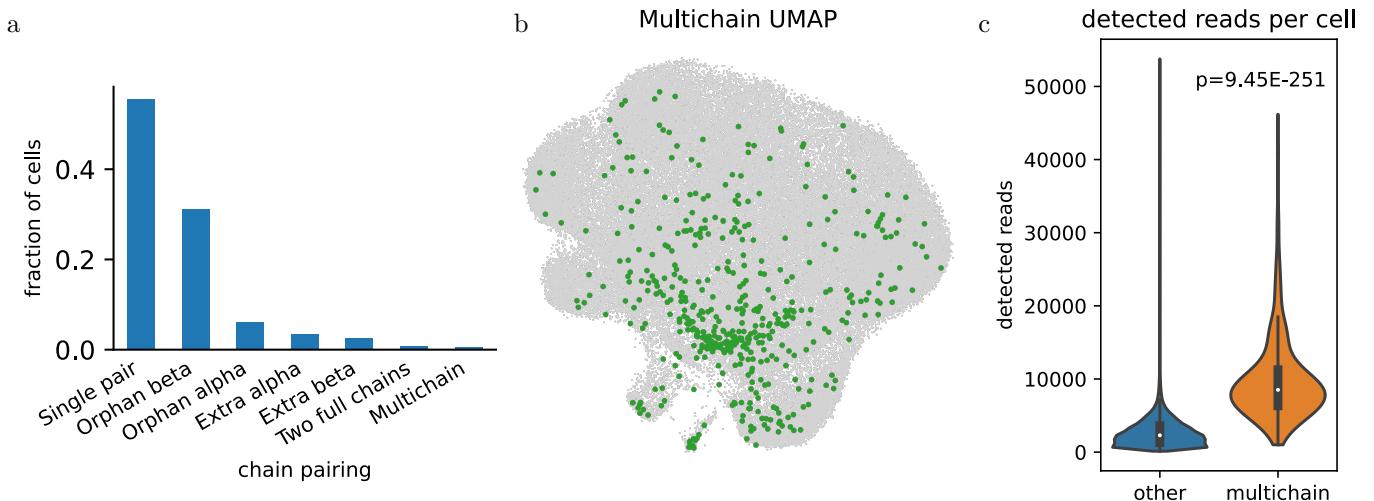
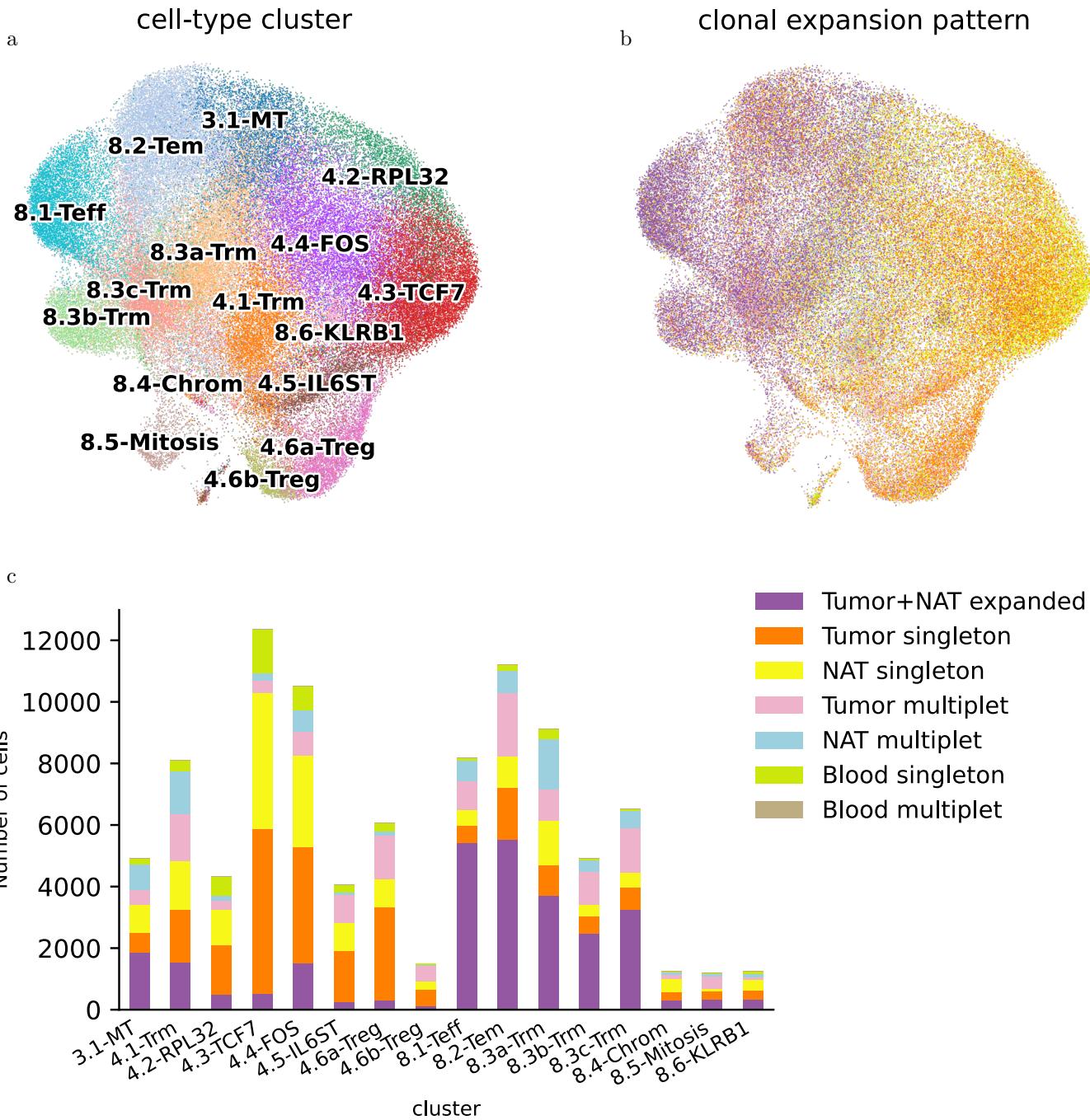


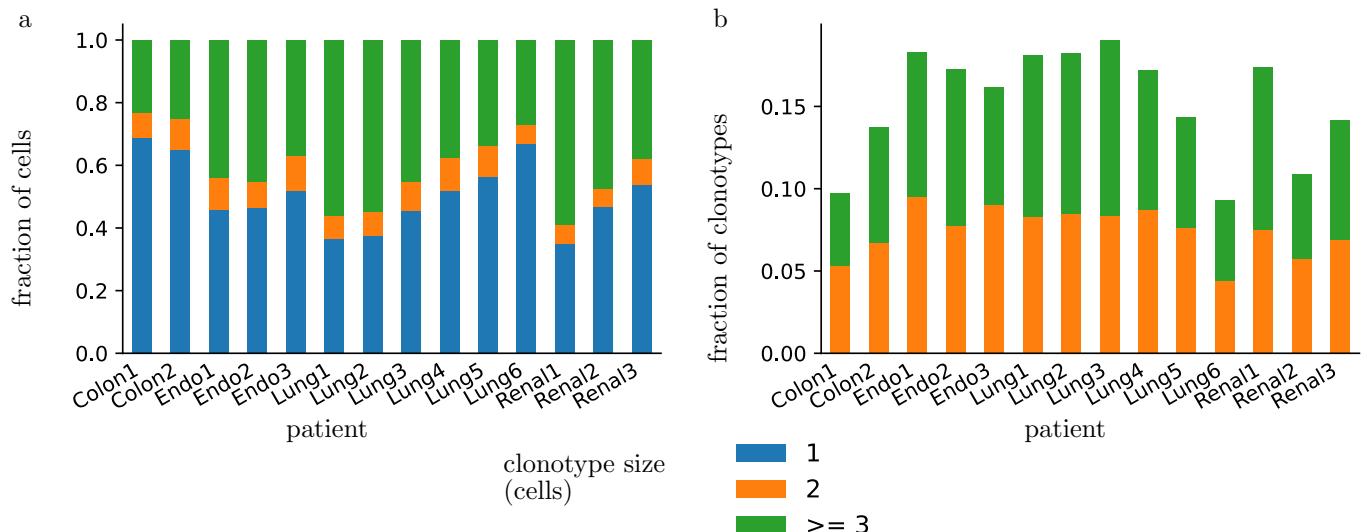
Supplementary Information



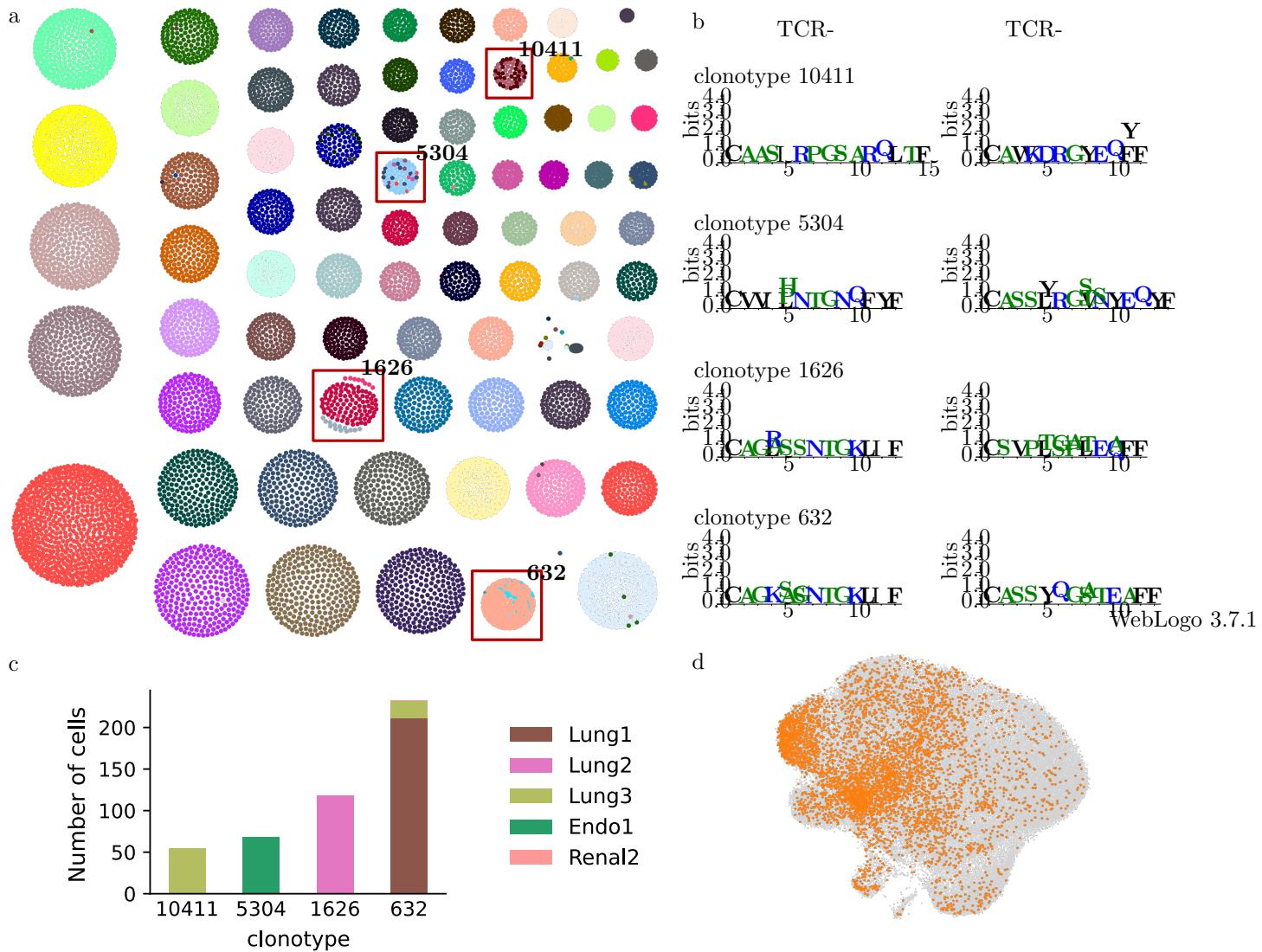
Supplementary Figure 1: Scirpy flags cells with more than two pairs of TCR-chains as “multichain” cells. (a) Quality control of chain pair recovery showing the fraction of cells with unconventional receptor configuration. Orphan alpha and beta designates cells where only a single alpha or beta chain sequence could be recovered, respectively. Extra in this context means that in addition to a valid pair of receptor chains, an additional alpha or beta sequence was found. (b) UMAP plot of 96,000 cells from Wu et al. [1] with at least one detected CDR3 sequence with multichain-cells ($n=474$) highlighted in green. (c) Comparison of detected reads per cell in multichain-cells and other cells. Multichain cells comprised significantly more reads per cell ($p = 9.45 \times 10^{-251}$, Wilcoxon-Mann-Whitney-test), supporting the hypothesis that (most of) multichain cells are technical artifacts arising from cell-multiplets [2].



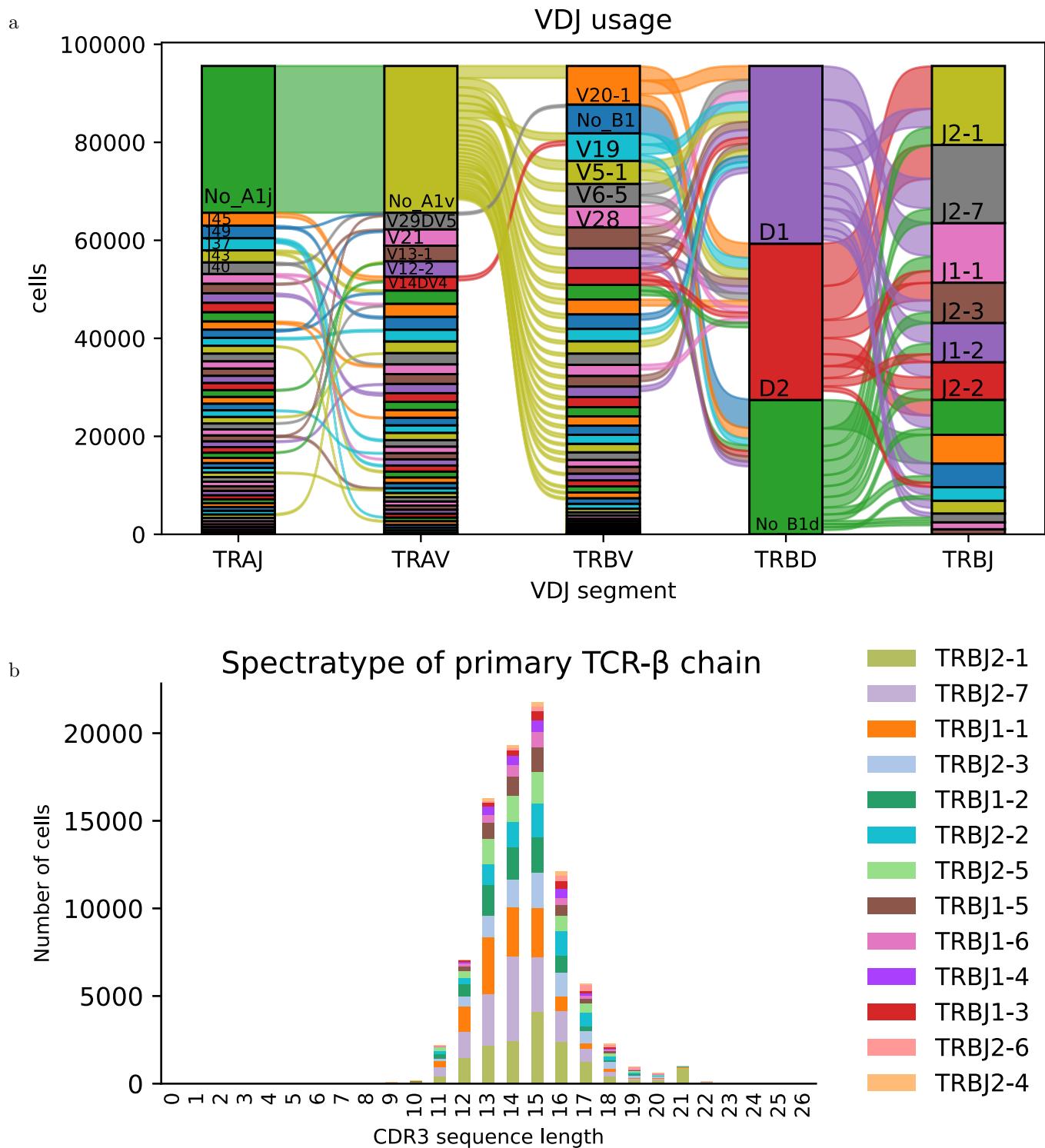
Supplementary Figure 2: Differential tissue expansion patterns by T-cell clusters. (a) UMAP plot of 96,000 cells from [1] with at least one detected CDR3 region, colored according to the cell-type clusters described in the original study. (b, c) Distribution of tissue expansion patterns by clustered T-cell subpopulations visualized as (b) UMAP plot and (c) bar chart. While CD8+ effector, effector memory, and tissue-resident T-cells tended to be expanded in both tumor and normal adjacent tissue (NAT), CD4+ T-cell clusters primarily comprised singleton clonotypes.



Supplementary Figure 3: T cell clonal expansion per patient. (a) Fraction of cells belonging to clonotype singletons (blue), clonotypes with two cells (orange), or clonotypes with more than two cells (green). (b) Fraction of clonotypes comprising two, or more than two cells, respectively. Overall, between 8% and 19% of clonotypes were expanded, depending on the patient.



Supplementary Figure 4: The comparison of clonotypes based on either sequence similarity or identity yields evidence of convergent evolution. (a) Clonotypes with more than 50 cells visualized as a network plot. Each cluster represents a sub-network corresponding to a clonotype based on amino-acid sequence similarity; each dot represents an individual cell and is colored according to the clonotype assigned based on nucleotide-sequence identity. Four clonotypes with heterogeneous nucleotide sequences are highlighted. (b) Sequence logos of the primary TCR- and TCR- chains of the clonotypes highlighted in (a). (c) Patient composition of the four clonotypes highlighted in (a). Clonotypes number 10411, 5304, and 1626 comprised cells from a single patient only and were likely the result of convergent clonotype evolution. Clonotype 632 comprised cells from different patients and potentially represents an epitope targeting a common disease epitope (e.g. a viral epitope). (d) Cells from convergent clonotypes highlighted in the UMAP plot from [1]. We defined convergent clonotypes as subnetworks including different nucleotide sequences from the same patient.



Supplementary Figure 5: VDJ gene usage of 96,000 T cells from Wu et al. [1] (a) VDJ gene usage of primary TCR- γ chains visualized as a Sankey plot. Each bar refers to a specific VDJ segment, and each bar section to a specific V, D, or J gene. (b) Spectratype plot of the primary TCR- β chain, colored by J gene usage. The bars show the distribution of TCR- CDR3 sequence length.

Tool name	Website	Language	Receptor type	Support for 10x	Support for Smart-seq2	Paired chains	Post-processing of clonotypes	Basic IR visualization	Advanced IR visualization	Clonotype abundance analysis	Clonotype clustering of single cells	Integrated Gex Ref.	
			TCR	x	x	x	✓	✓	✓	x	x	[3]	
Immcantation (pRESTO)	https://immcantation.readthedocs.io	Python (and R)	TCR BCR	x	x	x	✓	✓	✓	x	x	[3]	
ImmuneArch (tcR)	https://immanarch.com/	R	TCR BCR	✓	x	x	✓	✓	✓	x	x	[4]	
iMonitor	https://github.com/zhangwei2015/IMonitor	Perl (and R)	TCR BCR	x	x	x	✓	✓	x	x	x	[5]	
VDJtools	https://github.com/mikessh/vdjtools	Java (Python)	TCR BCR	x	x	x	✓	✓	✓	x	x	[6]	
TRUST4	https://github.com/liulab-dfc1/TRUST4	Perl	TCR	(✓) ¹	(✓) ¹	x	x	x	x	x	x	[7]	
scTCR Seq	https://github.com/ElementoLab/scTCRseq	Python	TCR	x	✓	✓	✓	✓	✓	x	x	x	[8]
TraCeR	https://github.com/teichlab/tracer	Python	TCR	x	✓	✓	✓	✓	✓	x	✓	(✓) ²	[9]
VDJ Puzzle	https://github.com/simone-rizzetto/VDJPuzzle	Java	TCR	x	✓	x	✓	✓	✓	x	x	(✓) ²	[10]
TRAPes	https://github.com/YosefLab/TRAPeS	C++	TCR	x	✓	x	x	x	x	x	x	x	[11]
Scirpy	https://github.com/icbi-lab/scirpy	Python	TCR	✓	✓	✓	✓	✓	✓	✓	✓	✓	

Supplementary Table 1: Comparison of Scirpy with currently available tools supporting immune repertoire analysis at the single-cell level, or offering immune repertoire visualization. Basic immune repertoire (IR) visualization includes clonotype abundance, diversity, and VDJ usage. Advanced IR visualization also includes repertoire overlap, clustering as well as specialized analysis of individual clonotypes and publication ready figures. “Gex” indicates gene expression data.

¹Requires extra settings.

²Requires special skills and extra coding.

1. Supplementary Note

A clonotype designates a collection of T or B cells that bear the same adaptive immune receptors, and thus recognize the same epitopes. Generally, these cells are also descendants of a common, antecedent cell and belong to the same cell clone. In single-cell RNA-sequencing (scRNA-seq) data, T cells sharing identical complementarity-determining regions 3 (CDR3) sequences of both α and β TCR chains make up a clonotype.

Contrary to what would be expected based on the previously described mechanism of allelic exclusion [12], scRNA-seq datasets can feature a considerable number of cells with more than one TCR pair. Since cells with more than one productive CDR3 sequence for each chain did not fit into our understanding of T cell biology, most TCR analysis tools ignore these cells [13, 14] or select the CDR3 sequence with the highest expression level [11]. 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 bone fide dual-TCR population [15, 16].

Scirpy allows investigating the composition and phenotypes of both single- and dual-TCR T cells by leveraging a T cell model similar to the one proposed in [9], where T cells are allowed to have a primary and a secondary pair of α and β chains. For each cell, the primary pair consists of the α- and β-chain with the highest read count. Likewise, the secondary pair is the pair of α/β-chains with the second highest expression level. Based on the assumption that each cell has only two copies of the underlying chromosome set, if more than two variants of a chain are recovered for the same cell, the excess TCR chains are ignored by Scirpy and the corresponding cells flagged as “multichain” (Supplementary Figure 1). This filtering strategy leaves the choice of discarding or including multichain cells in downstream analyses.

Scirpy implements a network-based clonotype definition that enables clustering cells into clonotypes based on the following options:

- (a) identical CDR3 nucleotide sequences;
- (b) identical CDR3 amino acid sequences;
- (c) similar CDR3 amino acid sequences based on pairwise sequence alignment.

The latter approach is inspired by studies showing that similar TCR sequences also share epitope targets [13, 17, 18]. Based on these approaches, Scirpy constructs a global, epitope-focused cell similarity network. While convergence of the nucleotide-based clonotype definition to the amino acid based one hints at selection pressure, sequence alignment based networks offer the opportunity to identify cells that might recognize the same epitopes.

Supplementary References

1. Wu, T. D. *et al.* Peripheral T cell expansion predicts tumour infiltration and clinical response. en. *Nature* (Feb. 2020).
2. Ilicic, T. *et al.* Classification of low quality cells from single-cell RNA-seq data. *Genome Biol.* **17**, 29 (Feb. 2016).
3. 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).
4. Nazarov, V. I. *et al.* tcr: an R package for T cell receptor repertoire advanced data analysis. en. *BMC Bioinformatics* **16**, 175 (May 2015).
5. Zhang, W. *et al.* IMonitor: A Robust Pipeline for TCR and BCR Repertoire Analysis. en. *Genetics* **201**, 459–472 (Oct. 2015).
6. Shugay, M. *et al.* VDJtools: Unifying Post-analysis of T Cell Receptor Repertoires. en. *PLoS Comput. Biol.* **11**, e1004503 (Nov. 2015).
7. Li, B. *et al.* Landscape of tumor-infiltrating T cell repertoire of human cancers. en. *Nat. Genet.* **48**, 725–732 (July 2016).
8. 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).
9. Stubbington, M. J. T. *et al.* T cell fate and clonality inference from single-cell transcriptomes. en. *Nat. Methods* **13**, 329–332 (Apr. 2016).
10. Rizzetto, S. *et al.* B-cell receptor reconstruction from single-cell RNA-seq with VDJPuzzle. en. *Bioinformatics* **34**, 2846–2847 (Aug. 2018).
11. 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).
12. Brady, B. L., Steinle, N. C. & Bassing, C. H. Antigen receptor allelic exclusion: an update and reappraisal. en. *J. Immunol.* **185**, 3801–3808 (Oct. 2010).
13. Fischer, D. S., Wu, Y., Schubert, B. & Theis, F. J. Predicting antigen-specificity of single T-cells based on TCR CDR regions.
14. Zhang, L. *et al.* Lineage tracking reveals dynamic relationships of T cells in colorectal cancer. en. *Nature* **564**, 268–272 (Dec. 2018).
15. Schuldt, N. J. & Binstadt, B. A. Dual TCR T Cells: Identity Crisis or Multitaskers? en. *J. Immunol.* **202**, 637–644 (Feb. 2019).
16. Ji, Q., Perchellet, A. & Goverman, 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).
17. Glanville, J. *et al.* Identifying specificity groups in the T cell receptor repertoire. en. *Nature* **547**, 94–98 (July 2017).
18. Dash, P. *et al.* Quantifiable predictive features define epitope-specific T cell receptor repertoires. en. *Nature* **547**, 89–93 (July 2017).