ImmunoInformatics 5 (2022) 100009

Contents lists available at ScienceDirect

ImmunoInformatics

journal homepage: www.elsevier.com/locate/immuno

# Recent advances in T-cell receptor repertoire analysis: Bridging the gap with multimodal single-cell RNA sequencing

Sebastiaan Valkiers [a,b], Nicky de Vrij [a,b,c], Sofie Gielis [a,b], Sara Verbandt [d], Benson Ogunjimi [b,e,f,g], Kris Laukens [a,b], Pieter Meysman [a,b,*]

[a] Adrem Data Lab, Department of Computer Science, University of Antwerp, Antwerp, Belgium
[b] Antwerp Unit for Data Analysis and Computation in Immunology and Sequencing (AUDACIS), University of Antwerp, Antwerp, Belgium
[c] Clinical Immunology Unit, Department of Clinical Sciences, Institute of Tropical Medicine, Antwerp, Belgium
[d] Molecular Digestive Oncology, Department of Oncology, Katholieke Universiteit Leuven, Leuven, Belgium
[e] Centre for Health Economics Research & Modeling Infectious Diseases (CHERMID), Vaccine & Infectious Disease Institute (VAXINFECTIO), University of Antwerp, Antwerp, Belgium
[f] Antwerp Center for Translational Immunology and Virology (ACTIV), Vaccine and Infectious Disease Institute (VAXINFECTIO), University of Antwerp, Antwerp, Belgium
[g] Department of Paediatrics, Antwerp University Hospital, Antwerp, Belgium

ARTICLE INFO

*Keywords:*
Single-cell sequencing
Single-cell immune profiling
Single-cell gene expression
T-cell receptor repertoire
Multi-omics

ABSTRACT

T cells exercise a multitude of functions such as cytotoxicity, secretion of immunomodulating cytokines or regulation of tolerance, collectively resulting in an effective control of immune-related disease. Through the unique mechanism of V(D)J recombination, T cells express a highly specific receptor complex known as the T-cell receptor (TCR). Single-cell sequencing technologies have paved the road for interrogating the transcriptome and the paired $\alpha\beta$ TCR repertoire of a single T cell in tandem. In contrast, conventional bulk methods are restricted to only one layer of information. This combination of transcriptomic- and repertoire information can provide novel insight into the functional character of T cell immunity. Recently, single-cell technologies have gained in popularity due to improvements in throughput, decrease in cost and the ability for multimodal experiments that integrate different information layers. Consequently, this prompts the need for the development of novel computational tools that integrate transcriptomic profiles and corresponding features of the TCR repertoire. Here we discuss the current progress in the field of single-cell T cell sequencing, with a focus on the multimodality of new approaches that allow the paired profiling of cellular phenotype and clonotype information. In addition, this review provides detailed descriptions of recent computational developments for analyzing single-cell TCR sequencing data in an integrative manner using novel computational approaches. Finally, we present an overview of the available software tools that can be used to perform integrative analysis of gene expression and TCR profiles.

## 1. Introduction

T cells are central players of the adaptive immune system, and play a crucial role in the control of immune-related diseases. In addition, T cells are indispensable mediators of vaccination and immunotherapy response. By generating a highly diverse T cell repertoire, the adaptive immune system is equipped with a powerful toolkit to protect against a broad range of pathogenic organisms and cancer. This diversity in the T cell repertoire is achieved by the generation of a multitude of different T-cell receptor (TCR) complexes through V(D)J recombination. The TCR, expressed on the cellular surface, recognizes small peptides (epitopes), derived from foreign or self-antigens, and presented by Major Histocompatibility Complex (MHC) molecules on antigen-presenting cells. Upon binding of a TCR to its cognate peptide-MHC (pMHC) complex, a T-cell-mediated immune response is triggered [1]. The TCR complex of most T cells consists of an $\alpha$ and $\beta$ chain, and the diversity of the complex is produced by the recombination process of V and J gene segments in the $\alpha$ chain, and an additional D gene segment in the $\beta$ chain. This is the first level of diversity, known as combinatorial diversity. During this process of recombination, non-templated nucleotides are added and deleted at the junctions of the segments, drastically increasing the potential diversity of the TCR repertoire [2]. This is known as junctional diversity. Finally, an additional level of diversity is established through the near unconstrained pairing of $\alpha$ and $\beta$ chains [3]. The total theoretical diversity of the TCR repertoire remains largely disputed but estimations range from $10^{15}$ up to $10^{61}$, although the true diversity is limited to the

* Corresponding author.
*E-mail address:* pieter.meysman@uantwerpen.be (P. Meysman).

https://doi.org/10.1016/j.immuno.2022.100009
Received 12 November 2021; Received in revised form 7 January 2022; Accepted 24 January 2022
2667-1190/© 2022 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)

total number of T cells in the human body ($3•10^{11}$) and restricted by selection processes in the thymus [4–8].

TCR sequencing has become an invaluable tool for understanding the complex TCR repertoire dynamics. A clonotype is typically defined by the unique combination of consecutive V gene, CDR3 amino acid sequence and J gene. In general, the number of unique clones within a typical bulk repertoire sample may vary between $10^3$ to $10^6$, depending on clonality of the repertoire, sampling conditions and sequencing depth [9,10]. However, even though TCR sequencing has become indispensable, typical bulk approaches provide only one layer of information on the T cell, as they only capture the TCR characteristics. Considering T cells exhibit a wide array of immune phenotypes, enabling various functionalities ranging from secreting (anti-)inflammatory cytokines to releasing cytotoxic effector molecules to induce cell death, the receptor characteristics do not fully capture the functionality of T cell they originate from. Traditionally, immunologists rely on targeted techniques, such as flow cytometry, to characterize these distinct immune cell phenotypes. Flow cytometry separates cells based on the expression of certain marker proteins, using fluorescently-tagged antibodies to label these targets [11]. However, the number of available fluorophores is limited by emission spectra overlap, restricting the number of measurable parameters [12,11]. Although the number of measurable parameters can be increased to 50 by using altered and improved techniques, such as mass cytometry (also known as CyTOF), this still limits users to a specific number of pre-defined markers [13]. In contrast, RNA sequencing can capture the phenotype of cellular subsets in an unbiased manner as it is not restricted to a limited and targeted selection of markers. However, gene expression at the transcriptional level can be insufficient for discriminating between certain cellular subsets, whereas a protein marker might be more descriptive. For example, the expression of distinct CD45 isoforms, enabling us to discriminate between naive and memory T cells, cannot be identified on the transcriptional level [14]. Additionally, bulk RNA sequencing typically results in a composite mix of gene expression profiles derived from all cells in the sample, which does not sufficiently reflect the cellular diversity. Thus, bulk RNA sequencing often requires prior cell sorting with fluorescent-labeled antibodies to target protein markers, in order to purify cell types within a sample.

A more promising alternative to these conventional techniques is single-cell RNA sequencing, which leverages the power of combining multiple information layers such as paired sequencing of the gene expression and TCR sequences within a single-cell. This feature of multimodality is not restricted to the transcriptional level. For example, the addition of antibodies linked to specific oligonucleotide barcodes (Feature Barcoding) enables the characterization of surface proteins, similar to flow cytometry. Nonetheless, despite its promise, several aspects of single-cell sequencing remain challenging. Compared to conventional (bulk) techniques, single-cell sequencing is still costly and labor-intensive. Consequently, sample sizes are typically lower. However, as the single-cell sequencing field is rapidly growing [15], more recent developments have allowed easy sample multiplexing by including oligonucleotide-labeled antibodies that allow discrimination between samples (Cell hashing), reducing costs and enabling greater sample sizes [16]. The number of cells that can be sequenced with single-cell platforms is several orders of magnitude lower than with bulk methods as well. For example, most single-cell sequencing technologies only allow the assessment of up to $10^4$ cells, whereas bulk approaches can typically confidently assess $> 10^5$ cells [17]. However, this number is rising for single-cell sequencing as technology advances.

There are several platforms that enable the characterization of T cells at the single-cell level, each of them differing in how the cells are prepared and how the genetic material is subsequently enriched for sequencing. These factors have substantial impact on the sequencing throughput, depth, costs, and even the ability of generating data from multiple modalities. As prior reviews have extensively compared multiple single-cell sequencing methods, we will not further discuss the respective methods [18,19,20,21,17]. Instead, in this review, we will focus on the data analysis of the paired T cell gene expression profile and their TCR sequences [22–25]. Next to the previously described advantages, the sequencing of the TCR of a single cell has the benefit of conveniently pairing $\alpha$ and $\beta$ chains. This is difficult to achieve with conventional bulk methodologies, due to the unknown origin of TCR molecules in bulk experiments. In addition, technical limitations of bulk techniques, and the larger heterogeneity introduced by the additional recombination of the D segment, have resulted in a preferential interest in the TCR$\beta$ chain. As a result, most of our understanding of TCR recognition has been based on the sequencing of the $\beta$ chain alone [26]. However, it has been shown that the $\alpha$ chain partially mediates the recognition of a peptide-MHC complex (pMHC) to varying degrees as well [27–32]. Moreover, the power of multimodality provided by single-cell sequencing also allows the inclusion of T cell targeting peptide-MHC dextramers, enabling the identification of antigen-specific T cells, their TCR sequences and their functional phenotypes. This information is vital not only for elucidating the immunopathology of immune-mediated disease, but can be utilized to identify potential immunotherapeutic targets or help guide immunomonitoring in clinical trials [33–35].

Collectively, the combination of single-cell RNA and TCR sequencing offers multiple benefits over conventional (bulk) techniques. In this review, we will outline several of these benefits. Furthermore, we aim to introduce experimental or computational immunologists working on bulk TCR sequencing data or bulk RNA-seq data to common high-resolution and unbiased multimodal single-cell workflows that are able to generate new biologically relevant insights. Due to emerging applications in both targeted enrichment of TCR transcripts and single-cell sequencing of TCR repertoires, a large number of different tools are being developed for the downstream analysis of this data, in an attempt to integrate gene expression with clonotype information. The final purpose of this review is to provide an overview of the current state-of-the-art methods and software tools for the integration and downstream analysis of single-cell TCR and single-cell gene expression data. Although this review will primarily focus on the analysis of T cells and their receptors, many of the discussed methods and technologies are also applicable to B cells.

To fully introduce the reader to the benefits of paired single-cell RNA and TCR sequencing, we will first outline the typical workflows for unpaired single-cell RNA sequencing and TCR repertoire analysis. Next, we discuss the advantages of integrating these two layers of information, and the available tools that enable this integration. Finally, we identify current challenges in the field of TCR repertoire analysis and provide novel perspectives on how to bridge the research gaps in this field.

## 2. General workflow for single-cell transcriptomics

### *2.1. Power calculation for single-cell transcriptomics*

Power calculation is an important component of experimental design. While power calculation approaches for bulk sequencing can be applied on the single-cell level, these often fail to take single-cell specific characteristics such as data sparsity into account. Several factors determine the statistical power of a single-cell sequencing experiment, including the depth of sequencing (# of reads per cell), the number of cells per sample and the number of samples. These factors are typically influenced by budgetary restraints, technical limits of the chosen sequencing platform and sample availability. Although information on the recommended sequencing depths is typically provided by the assay manufacturer (e.g. 10x Genomics), deciding on adequate sample sizes and the required number of cells per sample are more challenging. In addition, other prior knowledge, depending on the research question, might be required to calculate the power. For example, when trying to identify a rare cell population in a sample, prior knowledge on the proportion of cells within that sample type may be required to deter-

**Table 1**
Pre-processing steps for single-cell RNA sequencing data.

| Pre-processing step | Goal | Method |
|---|---|---|
| Acquisition of initial gene expression matrix | Transformation of raw sequencing data to raw gene expression matrix. | Cell Ranger (10x Genomics) [159]<br>Kalisto - bustools [160,161]<br>STARsolo [162]<br>Alevin-fry [163] |
| Quality control | Removal of low quality cells. | Using combinations of 4 metrics per barcode:<br>(1) number of UMIs<br>(2) number of detected genes (features)<br>(3) fraction of reads that map to mitochondrial genes,<br>(4) fraction of reads that map to ribosomal genes |
| | Identification and removal of multiplets | Manual inspection: Removal of multiplets when visualizing the distribution of the number of detected genes for all cells (Caution: potential loss of relevant cells)<br>Computational doublet-detection methods (benchmarked in [164]) |
| Normalization and scaling | Normalization to relative gene expression levels in order to adequately compare expression across cells. | Log normalization (most common)<br>SCTransform [140] (shown to improve performance of downstream analysis) |
| Integration of samples | Combining samples for analysis while removing technical variability such as batch- or sample -specific sequencing effects. | Manual inspection: Regressing out sources of variability.<br>Batch correction algorithms: Harmony [141], LIGER [165], Seurat [166], scGen [167] and scVI [168] |
| Dimensionality reduction | Limit the amount of noise by reducing the high-dimensional gene expression matrix to a low- dimensional space, retaining the bulk of the information captured by the data. | Selecting the top n most variable genes (generally 1000 to 5000).<br>Dimensionality reduction techniques: PCA followed by t-SNE [42] or UMAP [43]) |

mine how many cells need to be sequenced for adequate power [36]. As different prior knowledge is required depending on the research question, several single-cell power analysis tools have been developed with varying purposes, such as scPower [36], SCEED [37] or SCOPIT [38]. While SCEED is focused on power calculation for cell type identification, scPower can be rather wheeled for the power calculation for differential gene expression testing and expression quantitative trait loci analysis. Finally, SCOPIT uses a multinomial distribution to calculate the required cell number, based on the minimum number of cells that must be sequenced per subpopulation, the desired probability of sampling that number of cells for each population, and the frequency of the rarest subpopulation. The model can be used as an intuitive web interface or as an R package.

### 2.2. Pre-processing of single-cell RNA sequencing data

Sequencing data generated by single-cell platforms, similar to standard (bulk) sequencing applications, requires some processing before proceeding to downstream analysis. The main pre-processing steps are summarized in Table 1. While we will briefly discuss some of these in this section, readers are referred to the supplementary text for a more detailed explanation, including popular software tools that enable pre-processing of (single-cell) RNA-seq data. Some of these steps are also further explained in an excellent review by Luecken and Theis [39]. In addition, a recent benchmark demonstrated that the choice between pre-processing tools is relatively unimportant, observing minor differences after downstream processing [40]. Nonetheless, for V(D)J profiling using the 10x Genomics platform, CellRanger is recommended as it simultaneously processes both the gene expression and paired TCR data.

The first step in pre-processing consists of the conversion of raw sequencing output data to a more interpretable gene expression matrix. The acquisition of the raw gene expression matrix includes read assignment to individual samples and cells based on cellular barcodes (demultiplexing), genome and/or transcriptome alignment, and read quantification. Even after assigning all reads to a cellular barcode, not every barcode will correspond to living cells of sufficient quality. Additionally, a small but non-negligible proportion of the data will consist of distinct multiple cells that have been captured as one single cell (multiplets), potentially confounding downstream analysis. Thus, quality control, involving the removal of low quality cells and multiplets, is typically performed directly after acquisition of the raw gene expression matrix. Next, the gene expression level will be normalized to capture the relative gene expression level between cells, accounting for potential technical variation, such as sequencing depth. Some unwanted variability might still exist after normalization, including both technical and biological factors. Consequently, these factors may need to be corrected for. Generally, technical variation originate from batch effects (cells sequenced in different runs or different sequencing lanes) or the integration of data from multiple experiments, among others. Batch effects can be clearly observed during downstream visualization, as similar cells will separate based on their batch or donor origin. While correcting for technical variation is often warranted, correcting for biological factors, such as cell cycle variability, is not always advisable as it can mask relevant biological information [41]. Lastly, a single-cell RNA-seq gene expression matrix is very high dimensional and prone to noise. The high dimensional matrix will be reduced to a low-dimensional space during dimensionality reduction, thereby reducing noise in the data while retaining essential information.

### 2.3. Common downstream analysis for single-cell transcriptomics

In this section, the most widely used analysis methods for single-cell gene expression data will be covered. However, available downstream analyses are not restricted to only these popular methods, and include, for example, the identification of gene regulatory networks or the inference of cell-cell communication as well. The techniques discussed in the upcoming section are summarized in Fig. 1.

#### 2.3.1. Clustering and cluster annotation

Non-linear dimensionality reduction techniques like t-distributed Stochastic Neighbor Embedding (t-SNE) [42] and Uniform Manifold Approximation and Projection (UMAP) [43] create representations where cells that are highly similar in gene expression tend to be grouped together (Fig. 1A). The lowdimensional embeddings created by these algorithms can be used to identify distinct cell clusters or communities based on similarity scores or distance metrics. To achieve this, classical machine learning clustering techniques or graph-based algorithms (community detection) can be applied on either a distance matrix (e.g. k-means clustering) or a graph-based representation, respectively. The Louvain community detection algorithm [44] is the most popular graph-based method, offering great computational performance. Several improvements of the Louvain algorithm have been suggested, contributing towards enhancements in modularity, speed and scalability. These improvements include the smart local move [45], fast local move [46,47] and random neighbor move [48] algorithms. More recently, the Leiden algorithm was introduced as an augmentation of the Louvain com-

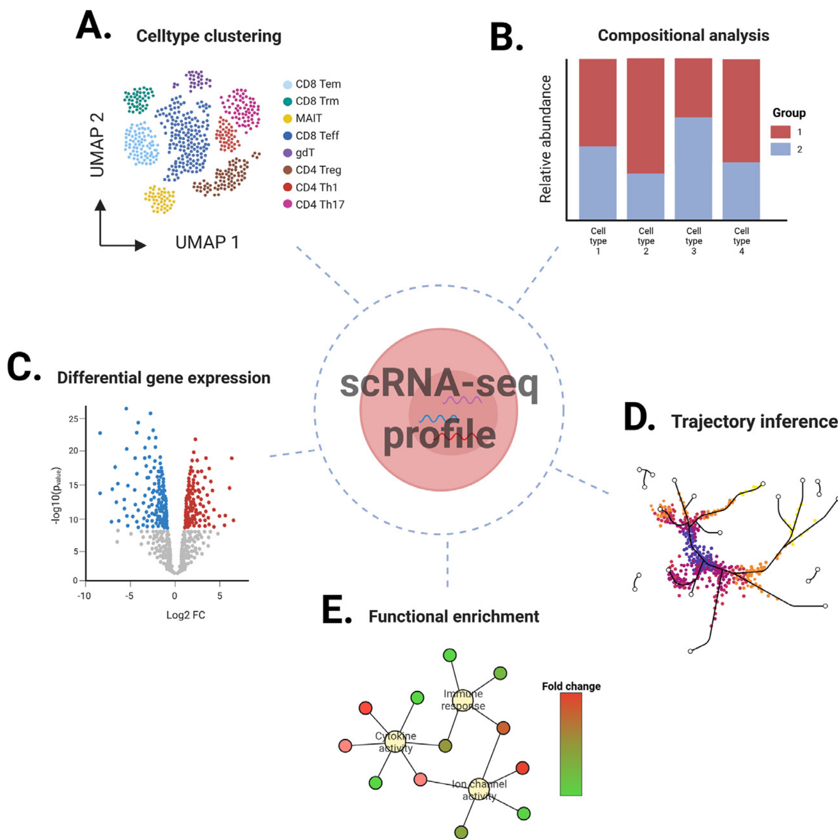


**Fig. 1.** Frequently used analyses for scRNA-seq data. A. Clustering of cells based on gene expression profiles and subsequent labeling of clusters with suspected cell types. B. Compositional analysis to compare the proportion of particular cell types across conditions. C. Volcano plots can be used to visualize differentially expressed genes identified through statistical testing. D. Ligand-receptor pair analysis or similar analyses based on correlative co-expression of particular genes. E. Functional enrichment analysis using publicly available databases determines up- or downregulated biological processes. F. Cell trajectory analysis to infer pseudotemporal evolution of cells in a dynamic process.

munity detection algorithm, by integrating these earlier improvements [49].

Once clusters have been established, they can be annotated with relevant biological information such as cell type identity. Generally, this is either performed manually or by using reference-based mappings. Manual cluster annotation requires a priori knowledge of marker genes, genes that are typically only (abundantly) expressed by a particular cell type. Cells are then compared for differential gene expression. Next, annotation labels are provided based on differential expression of the marker genes. In contrast, reference-based mappings make use of reference 'atlases' to transfer annotation labels to the dataset in question. Commonly employed reference mapping tools include SingleR [50], Azimuth [51], and within the context of T cells, ProjecTILs [52]. Aside from differences between tools in underlying methodologies for mapping, these tools use different reference atlases as well. For example, SingleR makes use of bulk RNAseq datasets of pure cell types to map the query dataset on. In sharp contrast, Azimuth uses atlases consisting of multimodal single-cell datasets that leverage the combination of transcriptome profiling with antibodies targeting known cell markers. Reference-based mapping, however, remains challenging. As these reference atlases consist of datasets that have been generated under different conditions than the query dataset of interest, they may contain an incorrect label for the cells under investigation. However, new datasets that could facilitate novel reference atlases are continuously generated, potentially providing an extensive catalog of high-quality reference atlases that better fit the dataset under investigation. In addition to this dataset generalization problem, droplet-based single-cell methods have a low recovery rate for transcription factors and cytokines [53]. As immune cells are notoriously heterogeneous, subtle differences in the expression of transcription factors or cytokines may pose additional difficulties for reference-based mapping tools to correctly assign T cell subsets [54,55]. It should be mentioned that this problem also affects manual annotation. However, considering these problems, it is generally recommended to combine reference-based mapping with manual annotation to confirm the validity of the annotations.

### 2.3.2. Differential cell type composition analysis

Differential cell type composition analysis includes the comparison of the proportions of particular cell types, relative to the total number of cells, between conditions or state (Fig. 1B). Some pathogenic organisms or diseases are known to affect the abundance of certain cell types. For example, CD4+ T-cell depletion is a general hallmark of HIV infection [56]. Cell type composition analysis can thus be a crude way to identify affected cell types in a particular disease, without a priori knowledge. However, the composition of a sample is heavily dependent on the preparation protocol, potentially confounding the compositional analysis [57]. For instance, certain cells may be more prone to stress and damage during library preparation, potentially skewing the proportions as these cells may be depleted in the sample as a result [58].

As a recent example, researchers applied compositional analysis to study the immune system's response against SARS-CoV-2 infections. In accordance with the observation that COVID-19 patients experience a cytokine storm induced by inflammatory monocytes and pathogenic T cells, compositional analysis was used to illustrate that proliferative T cells and CD14+ monocytes are significantly enriched in patients with severe symptoms [35,59].

### 2.3.3. Differential gene expression and functional enrichment

Although differential gene expression analysis and functional enrichment have long been employed for bulk gene expression profiling, the single-cell environment confers several advantages (Fig. 1C). In contrast to bulk gene expression profiling, resulting in a homogeneous average gene expression profile, single-cell data consists of gene expression profiles for each individual cell, offering greater resolution. Moreover, for each particular cell type cluster, the fraction of cells expressing a certain gene can be calculated. Within the context of differential gene expres-

sion testing in single cells, both conventional bulk methods and methods specifically developed for single-cell data are employed. Soneson & Robinson (2018) demonstrated that both techniques perform equally well [60]. However, a more recent benchmark shows increased fidelity to ground truth using pseudobulk methods [61]. Popular methods include the non-parametric Wilcoxon test, pseudobulk methods DESeq2 and edgeR, and single-cell method MAST among others [62–66]. In a recent paper, Bassez et al. introduce the concept of applying differential expression testing on expanded and non-expanded T cells to illustrate that the expanding T cells pre-anti-PD1-treatment are tumor-reactive, showing higher expression of activation, effector and immune-checkpoint markers [34]. In another example, Zhang et al. applied differential gene expression analysis to identify differences in the transcriptional profile of different T cell types in colorectal tumor samples [67].

Differential gene expression testing typically outputs a large number of differentially expressed genes that require additional biological knowledge to interpret. A common aid herein are pathway analysis methods. Here, annotated genes are grouped into certain sets based on biological features, and computational algorithms test whether any sets are enriched in the differential gene list (over/under-representation analysis) or in the extremes of the ranked log-fold change list (gene set enrichment analysis) [68] (Fig. 1E). These methods rely on databases of annotated gene sets to test against, such as the Molecular Signatures Database (MSigDB) [69], Reactome [70] or Gene Ontology (GO) [71,72].

#### 2.3.4. Trajectory analysis

scRNA-seq provides a static snapshot of the cells at a particular time point. However, some of these cells will be involved in a dynamic process, such as cellular differentiation, the cell cycle or a gradual change in biological function. Thus, clusters annotated with just cell type labels do not fully capture the heterogeneity of the clusters, as they may contain a mixture of cells at different stages along a trajectory of a particular dynamic process. With trajectory analysis, cells are ordered along a path or trajectory based on transcriptional similarity (Fig. 1D). An inferred pseudotime variable represents the progression along this trajectory, starting from a particular cell type that is designated as the root cell. Trajectory analysis thus enables the interpretation of distinct dynamic processes, and the identification of gene expression profiles responsible for branching off along the trajectory. Differential gene expression along the trajectory is also possible [73]. Popular methods that allow trajectory analysis include Monocle [74] and Slingshot [75]. Trajectory analysis may be combined with RNA velocity to quantify the speed by which cells transition between different states [76,77]. In the context of T cells, integrating RNA velocity with cell trajectories may help to unravel the dynamics of the T cell response and reveal phenotypic transition between clonotypes. The choice of method generally depends on the dataset and trajectory topology, and interested readers are recommended to follow guidelines for method selection as proposed by Saelens et al. [78]. A very unique single-cell transcriptomic profiling study in supercentenarians used trajectory analysis to demonstrate that T cells of these supercentenarians were more terminally differentiated as compared to the T cells of healthy donors [79].

## 3. Extracting knowledge from TCR repertoires

The TCR repertoire is the collection of clonotypes constituting an individual's T cell landscape. TCR repertoire data can be generated through targeted enrichment strategies or the computational reconstruction of RNA-seq reads. Similarly to gene expression profiling, TCR sequencing data also requires some processing before downstream analysis. In brief, raw sequencing reads are first aligned to a reference set of V, D and J gene sequences, after which identical sequences are grouped into single clonotypes. Subsequently, poor quality reads are removed, and PCR and sequencing errors are corrected, resulting in quantitative clonotype information. There is a wide collection of tools available for processing TCR sequencing reads from bulk experiments [80–90], of which MiXCR remains the most popular choice [91]. The difference between these methods, their advantages and disadvantages have been extensively discussed by Bradley & Thomas [92].

In recent years, the post-processing of TCR repertoire data to reveal biologically relevant insight has gained more attention. These analyses can be roughly categorized into three major sections: analyzing the repertoire's diversity, specificity and clonal composition. Various methodologies have been developed for analyzing each of these aspects of the TCR repertoire, which have been summarized in Table 2. Additionally, Fig. 2 provides an overview of the different techniques discussed in this section. For each of the methods listed in Table 2 an extensive description can be found in the supplementary materials to this article. Researchers have developed several software tools that cover most of the functionalities discussed in Table 2. These allow the calculation of repertoire statistics like diversity (Fig. 2A), clonal composition or gene usage (Fig. 2C). Some tools provide additional functionalities for comparing different repertoires, for example through the quantification of clonal overlap (Fig. 2C). Lastly, more specific tools exist for advanced analyses of TCR repertoire data such as network analysis (Fig. 2H), clonotype clustering, enrichment analysis or the prediction of epitope specificity (Fig. 2G). For a more comprehensive overview of the available methodologies for analyzing T-cell clonotype data we refer to two excellent reviews by Bradley & Thomas [93], and Brown et al. [94].

### 3.1. Basic repertoire analysis

There is a plethora of software tools that can be used to for exploratory analysis of TCR repertoires. The Immcantation Portal1 hosts a range of different Python and R packages [95–102], leveraging an ecosystem for end-to-end analysis of TCR-seq data, from mapping raw sequencing reads up to advanced post-analysis (e.g. clustering of clonotypes). Moreover, the Immcantation framework is certified as being compliant with the adaptive immune receptor repertoire (AIRR) Standards2 guidelines for software tools [103]. Another software package, immunarch [104], provides an extensive suite of tools for the analysis of TCR data, which includes quantitation of clonotype abundance, repertoire diversity (Fig. 2A), repertoire overlap (Fig. 2B), gene usage estimation (Fig. 2C), clonotype tracking (Fig. 2D), CDR3 spectratyping (Fig. 2F), calculation of k-mer distributions and annotation of clonotypes with database information from VDJdb [105], McPAS-TCR [106] and TBAdb (PIRD) [107]. Finally, another popular package is VDJtools. This command line tool provides similar functionality to Immcantation and immunarch. VDJtools integrates a TCR neighborhood enrichment test (TCRNET) [108,109], which can be used to identify enriched clonotypes within a single repertoire as compared to a background distribution.

### 3.2. Generation probability

One of the most crucial advances in the field of immunoinformatics has been the development of probabilistic models for the V(D)J recombination process [110]. This is known to be a stochastic process that favours the generation of certain TCR sequence conformations over others. These models provide the opportunity to assign a probability of generation ($P_{gen}$) to any specific TCR sequence [111,112]. This $P_{gen}$ is calculated by explicitly modeling the probabilities of selecting a V, J or D gene (in case of TRB), and the potential nucleotide insertions and deletions at the junctions of these gene segments. This value provides an indication of whether a specific TCR sequence is rare or common. For example, longer TCR sequences tend to have a lower $P_{gen}$ (i.e. they are rarer) due larger numbers of insertions. In addition, probabilistic models of V(D)J rearrangement permit the generation of large synthetic repertoires that mimic the TCR repertoire of healthy individuals. Based on this concept, Pogorelyy et al. developed ALICE, an approach similar to TCRNET that can be used to identify enriched clones from single reper-

**Table 2**
Analysis of TCR repertoire data. The table provides an overview of common methods used to extract knowledge from TCR sequencing profiles. The Description columns provides a brief overview of what the analysis involves. Output describes the potential insight that the analysis permits. AIRR tools gives an overview of open-source software tools that can be used to perform the analysis. The Examples column provides an overview of different studies where the analysis was used to gain novel insight.

| Analysis | Description | Output | AIRR tools | Examples |
|---|---|---|---|---|
| Diversity analysis | Quantification of the diversity within a lymphocyte population. Diversity is typically expressed using Hill numbers. Alternatively, different estimation approaches exist to quantify true repertoire diversity (e.g. Chao2, ICE, DivE and Recon). | Combination of summary metrics that describe different aspects of the underlying frequency distribution. | Immcantation immunarch [104] VDJtools [135] | [169–178] |
| Clonal overlap | Quantification of the overlap between TCR repertoires. Popular metrics are the Jaccard or Morisita index. | Quantitative metric describing the degree of overlap between two repertoires. | Immcantation immunarch [104] VDJtools [135] | [179,180] |
| V(D)J gene usage | Evaluation of the distribution and co-occurrence of different V, D and J germline gene segments. | Potential biases in the use of specific V, D or J genes. | Immcantation immunarch [104] VDJtools [135] | [181–184] |
| Clonotype tracking | Track the frequency of a limited set of clonotypes across different time points or samples | Potentially expanding clones across time or treatment. | Immcantation immunarch [104] VDJtools [135] | [185–187] |
| CDR3 spectratyping | Evaluation of the CDR3 amino acid sequence length distribution. | Potential aberrations in the distribution of CDR3 length indicating expanded populations of clones with a bias in CDR3 length. | Immcantation immunarch [104] VDJtools [135] | [188,189,190,136] |
| K-mer & motif analysis | Decompose CDR3 sequences into overlapping k-mers and identify enriched subsequences that may represent important signatures contributing to the specificity of a TCR. | K-mers or sequence motifs enriched in one repertoire (or group of repertoires) versus another (group of) repertoire(s). | immunarch [104] | [10,191,192,193] |
| TCR clustering | Epitope-specific grouping of TCRs based on properties of the TCR sequence. | Clusters of TCRs targeting similar epitopes. | TCRdist [194] GLIPH2 [195] iSMART [196] ClusTCR [197] GIANA [198] | [199–205] |
| Enrichment analysis | Perform statistical enrichment tests to identify clonotypes that are significantly enriched in one or a group of repertoires versus another (e.g. healthy versus disease). | A list of TCRs that are statistically enriched in a group of individuals. | VDJtools [135] ALICE [113] immuneML [117] | [9,109,206] |
| TCR-epitope specificity | Identify the epitope specificity of a TCR by matching against a database with known TCR-epitope interactions or predict the specificity of a TCR using machine learning models. | A list of matched or predicted TCR-epitope interactions. | VDJdb [105] TCRex [115] immuneML [117] TCRGP [30] NetTCR [207] ERGO [26] DeepTCR [116] ImRex [157] | [155,156,30,26,31,157,158] |
| Network analysis | Represent the TCR repertoire as a network where nodes represent TCRs and the edges between them indicate similarity (typically hamming distance = 1). The repertoire architecture can be analyzed using various graph theoretic metrics. | Visualization of the repertoire architecture. Quantitative metrics of the repertoire architecture. | igraph [142] networkx [143] Cytoscape [208] | [154,209,210] |

toire snapshots, using a synthetic repertoire as background distribution [113].

### 3.3. Receptor specificity

Understanding which TCRs target which epitopes is arguably the most important challenge in repertoire analysis. This allows the identification of T cells responsible for the neutralization of pathogens. Consequently, this knowledge helps us understand why certain individuals may be susceptible to infection or cancer, while others are able to mount an effective immune response. In the context of autoimmune diseases, identifying TCRs that target self-antigens may leverage potential therapeutic targets. As indicated, immunarch and VDJtools provide functionalities for annotating clonotypes with epitope-specificity using databases of experimentally verified TCR-epitope interactions or associations, such as VDJdb [105], McPAS-TCR [106] and IEDB [114]. Other tools like TCRex [115] predict the specificity of any TCR towards a finite number of epitopes, based on epitope-specific machine learning models. For this application, TCR sequences are typically transformed into a numerical encoding. Popular types of encoding include the use of physicochemical properties or one-hot-encoding. The recent DeepTCR provides a deep learning framework for generating numerical representations of TCR sequences, which can be used for downstream machine learning applications such as the prediction of TCR-epitope specificity [116]. The immuneML platform [117] also provides functionalities to train and assess receptor-level machine learning classifiers using various encodings.

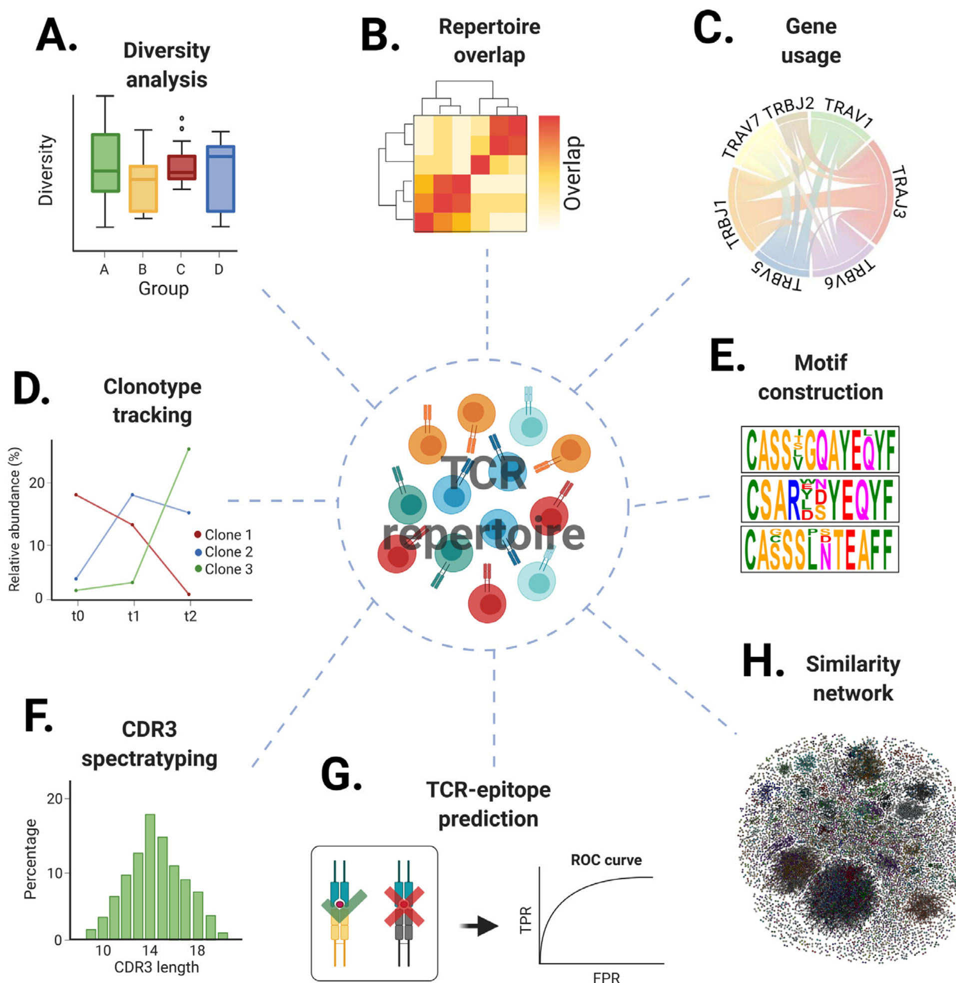


**Fig. 2.** Frequently used analyses for TCR sequencing data. A. Repertoire diversity compared across different groups or time points. B. Repertoire overlap describes the pairwise similarity of a range of TCR repertoires. A heatmap is a common type of visualization for repertoire overlap. C. Using chord diagrams, the co-occurrence of V and J genes can be visualized. D. Clonotype tracking is often used to evaluate the abundance of a limited set of TCRs across time points. E. Sequence logos can be constructed from a set of TCR sequences, revealing shared motifs that may contribute to the recognition of common antigens. F. CDR3 spectratyping is used to identify biases in the distribution of CDR3 lengths, which may disclose expansions of oligoclonal populations with similar lengths of the CDR3 region. G. Based on data of known TCR-pMHC interactions, epitope-specific machine learning models can be built for predicting whether an unknown TCR will recognize a specific epitope or not. H. Repertoire architecture can be presented by a network where the nodes represent unique TCR sequences, and the edges connecting them describe the similarity between two TCRs (typically defined as HD = 1).

ImmuneML offers models like K-Nearest Neighbours (KNN), logistic regression, random forests, TCRDist classifier, etc. Additional methods for predicting the epitope specificity of a TCR are presented in Table 2.

## 4. Generating matched single-cell gene expression and TCR data

### 4.1. Targeted enrichment of the V(D)J locus

Combined single-cell transcriptomics and adaptive immune profiling data is typically acquired through a targeted enrichment of the V(D)J region in conjunction with gene expression profiling. Amplifying the TCR locus can be performed using three main strategies. The first involves a multiplex PCR amplification using a set of primers that target all V and J gene segments. Alternatively, V(D)J sequences can be purified by targeting them with tagged TCR-specific oligos. These baits will anneal to the target regions, and can therefore easily be captured once the sample has been fragmented. Lastly, the most popular method for V(D)J amplification of cDNA samples is the 5′ RACE strategy. To effectively pair enriched V(D)J sequences and the rest of the gene expression profile, two main methods can be distinguished. Droplet-based methods using microfluidic devices are among the most popular strategies. Commercial examples of droplet-based approaches for isolating and barcoding individual cells are the Chromium device offered by 10x Genomics, ddSEQ by Bio-Rad, Nadia by Dolomite Bio, and inDrop by Illumina. There are also approaches that apply flow cytometric-cell sorting in 96 or 384 well plates to isolate individual cells. However, this approach limits the analysis to only one cell per well, per run. A commercial example of this approach is the C1 Single-Cell Auto Prep system by Fluidigm. A comprehensive overview of library preparation methods and sequencing strategies for paired sequencing are discussed in [118] and [17]. Generally, however, these sequencing protocols differ only in the method of amplification.

### 4.2. Computational reconstruction of TCR sequences

Aside from targeted enrichment, it is also possible to reconstruct TCRs from scRNA-seq data using computational methods. In contrast to targeted approaches, computational reconstruction methods provide lower coverage of the TCR repertoire, but allow re-analysis of existing scRNA-seq datasets, potentially providing additional insights. Moreover, conventional immunoprofiling kits typically contain only $\alpha/\beta$ amplification primers, resulting in minimal recovery of $\gamma\delta$ TCRs. However, the reconstruction of $\gamma\delta$ TCRs from the gene expression profile is possible, provided that the data was amplified from the 5′ end. There is a broad range of tools designed for recovering TCR sequences from scRNA-seq data, summarized in Table 3. For a more detailed description of each of the tools listed in Table 3, we refer to the supplementary materials of this article. TCR reconstruction tools generally use a combination of reference-based and de novo assembly, enabling the reconstruction of a considerable subset of V(D)J sequences from transcriptomic data. Although not competitive to targeted amplification methods, recent developments of TCR reconstruction tools have shown significant recovery of TCR sequences from scRNA-seq profiles. For example, the TRUST4 software was able to recover about 70% of all V(D)J sequences from scRNA-seq data [119]. The authors of MiXCR illustrated the recovery of around 3000 TRBs from lymph node metastasis samples, 1700–3000

**Table 3**
Overview of software tools for reconstructing TCR and BCR clonotypes from single-cell RNA sequencing data. A more detailed description of each method can be found in the supplementary information.

| Tool | Ref. | Version | Latest release | Source | Platform | Receptor type | Type of assembly | Assembler | Compatible with SMART-seq | Compatible with 10x |
|---|---|---|---|---|---|---|---|---|---|---|
| BASIC | [211] | 1.5.1 | 2019–07–12 | https://github.com/akds/BASIC | Python | B + T | Anchor-guided | Native | Yes | No |
| MiXCR | [91] | 3.0.13 | 2020–04–15 | https://github.com/milaboratory/mixcr | Java | B + T | Mapping | Native | Yes | No |
| scTCRseq | [212] | N.A. | 2016–06–09 | https://github.com/ElementoLab/scTCRseq | Python | T | *de novo* | GapFiller | Yes | No |
| TraCeR | [213] | 0.6.0 | 2018–03–08 | https://github.com/Teichlab/tracer | Python | T | *de novo* | Trinity | Yes | No |
| TRAPeS | [214] | N.A. | 2019–10–13 | https://github.com/YosefLab/TRAPeS | Python | T | Anchor-guided | Native | Yes | No |
| TRUST4 | [119] | 1.0.4 | 2021–05–13 | https://github.com/liulab-dfci/TRUST4 | C(++), Python | B + T | *de novo* | Native | Yes | Yes |
| VDJPuzzle | [215, 147] | 3.0 | 2020–03–19 | https://bitbucket.org/kirbyvisp/vdjpuzzle2 | Python | B + T | *de novo* | Trinity | Yes | No |

**Table 4**
Advantages and disadvantages of bulk and single-cell approaches to TCR and gene expression profiling. [1]: Repertoire coverage here refers to the total number of unique TCR sequences that can be identified. Depending on the scale of the experiment, single-cell approaches can reach similar repertoire coverage as bulk methods, but this will drastically increase the cost of the experiment. [2]: It is possible to study various modalities (e.g. TCR profile, gene expression profile, antigen-specificity, etc.) using bulk approaches, but they cannot be integrated. [3]: Generally, bulk approaches are more appropriate for large sample sizes, mainly due to the lower cost, efficiency and duration of the protocol.

| | Bulk | Single-cell |
|---|---|---|
| Repertoire coverage[1] | High | Low |
| Chain pairing | No | Yes |
| Gene drop-out rate[2] | Low | High |
| Multimodality | No | Yes |
| Cost per cell | Low | High |
| Sample size[3] | High | Low |

TRBs from CD4 T cells isolated from spleen, and around 400–1000 TRBs from central nervous system tissue [120]. However, the effectiveness of TCR recovery from scRNA-seq data is highly dependent on the sequencing depth and expression level of the TCR locus, which may vary considerably between cells [121]. Hence, this may introduce a substantial bias which should be considered when analyzing TCR diversity and clonality [17]. In conclusion, reconstruction of TCRs from scRNA-seq samples may be desirable if the aim of the experiment is the identification of expanded or dominant clones in the repertoire Table 3.

## 5. When to opt for single-cell over bulk approaches: features of single-cell T cell profiling

Whether the TCR data accompanying gene expression profiles is generated through dedicated enrichment of the V(D)J region or reconstructed from scRNA-seq data, having both layers of information can provide multiple advantages over conventional bulk technologies. Table 4 provides a brief comparison of the main features of bulk and single-cell sequencing approaches.

### *5.1. Single-cell sequencing allows integration of functional and immune receptor characteristics*

While bulk sequencing of TCRs provides a clear representation of the breadth of the antigenic response, it does not offer information about the functional characteristics of the T cells it originates from. Such information, provided by scRNA-seq, may help elucidate the mechanism of action for T cell subsets associated with pathology. For example, straightforward analyses include overlaying clonal properties on a UMAP or t-SNE generated from the gene expression matrix of the cells (Fig. 3A). Such analyses are especially interesting when performed in parallel to the cell type annotations described in the subsection on Clustering and cluster annotation. This may reveal certain biases in specific groups of cells, such as hyperexpansion of distinct phenotypic subsets [67,122]. However, these analyses are not restricted to just overlaying clonality, and may also include the application of previously described TCR-specific analyses (Fig. 2) on distinct T cell subsets. Such analyses may include, but are not limited to, calculating subset-specific TCR diversity (Fig. 3B), annotation with (imputed) antigen-specificity or analyzing bias in the use of germline genes in specific cell subpopulations. For example, Bilate et al. identified a clonally restricted CD4+CD8$\alpha\alpha$+ population whose differentiation was dependent on local antigen challenge [123]. As another example, Lee and colleagues identified reduced T cell clonality in patients with hereditary chronic pancreatitis (CP), as a result of CD4+ Th cells replacing tissue-resident CD8+ T cells [124]. In addition they studied clonotype sharing between different phenotypic subsets, and discovered interactions between CCR6+ Th and Th1 populations, in combination with significant upregulation of CCR6 ligand. These findings may indicate a role of a CCR6-CCL20 signaling pathway in hereditary CP. Tools such as immunarch, Immcantation or VDJtools can be used to evaluate the clonal distribution and diversity of T cells within a specific subset or to compare the overlap between multiple cell subsets. Conversely, the combination of gene expression and TCR repertoire profiles enables direct interrogation of the functional response of specific T cells using conventional transcriptomic analyses. For example, differential gene expression analysis can also be applied on distinct T cell subsets, on expanding versus non-expanding T cells, or on specific T cell clones or clusters (Fig. 3C). This functionality is offered by various software packages, including Scirpy [125], VDJView [126], scRepertoire [127], and Platypus [128]. The practical execution of this analysis is specific for each individual package. We therefore instruct the reader to carefully study the documentation provided with each of the described packages.

Vice versa, information obtained from the gene expression profile may be mapped onto a TCR similarity network (Fig. 3D), which is something that has been explored to a far lesser extent with existing tools. This type of analysis may reveal clusters of highly similar clones (therefore likely targeting identical epitopes) that belong to the same or related cell subsets, revealing expansions of T cell sets both on the phenotypic as well as the clonotype level.

### *5.2. Power of multimodality: antigen-specificity profiling*

Novel modalities have been developed for single-cell sequencing that allow researchers to unambiguously determine the antigen specificity of T cells. In these approaches, scTCR-seq and scRNA-seq are combined with epitopeloaded MHC multimers that epitope-specific T cells will interact with [129]. For example, TetTCR-seq, as described by Zhang et al., used such pMHC tetramers to profile the antigen-specificity of T cells [130]. This introduces a third, and very important layer of information

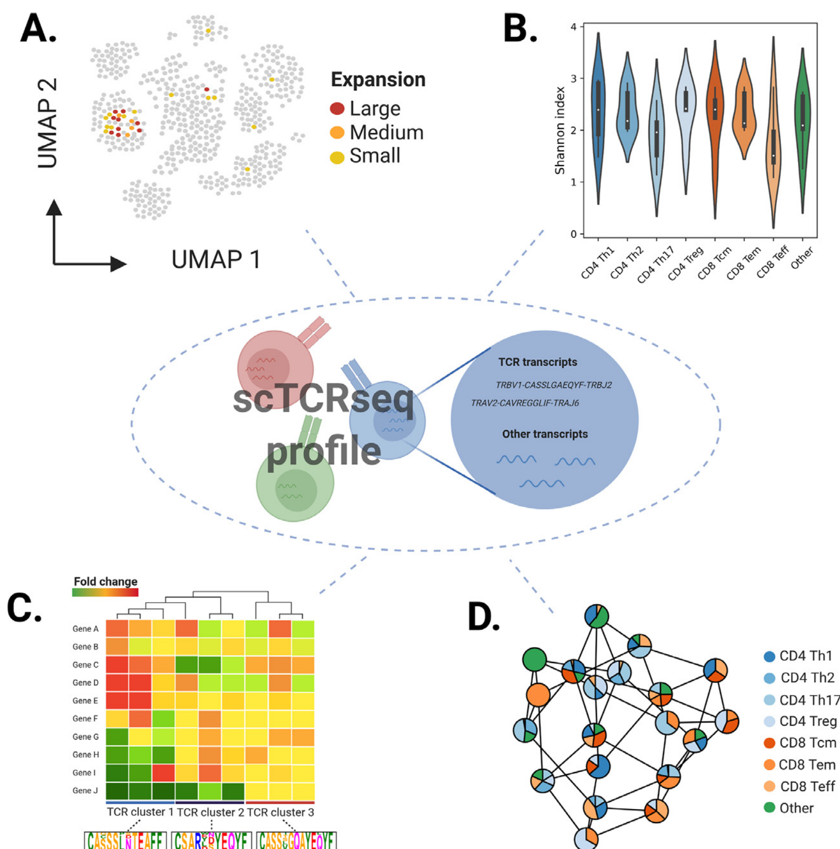


**Fig. 3.** Integrative methods for studying TCR and gene expression profiles in tandem. A. Clonotype information, such as clonal expansion, can be mapped onto a gene expression-based UMAP. B. TCR-specific metrics, such as diversity, can be evaluated on the level of different celltypes. C. Evaluating the gene expression profile of clonotypes within a TCR cluster. D. Celltype information can be projected onto TCR ype similarity networks in order to identify clonotype clusters with convergent or divergent celltypes.

that enables the complete characterization of T cell functionality, providing information about its cellular phenotype, receptor sequence and the peptide-MHC complex that it can recognize. In cancer research, for example, tumor-specific T cells can be identified, and subsequently used for adoptive T cell therapies, by capturing them using barcoded MHC-multimers loaded with the tumor epitope of interest [131,132]. Moreover, single-cell methodologies allow for the pairing of $\alpha$ and $\beta$ chain. This provides additional resolution, by including information about both TRA and TRB. In contrast, bulk approaches typically only offer single chain information.

## 6. Software packages for analyzing T cells at the single-cell level

With the emerging availability of analytical techniques for studying T cells at the single cell level, there has been a need for the development of tools to analyze the growing amount of data that accompanies this technological revolution. There is a plethora of tools available for analyzing either transcriptomics [74,133,134,75] or TCR [135,97,99,102,104,113,117] data individually, but less attention has been spent on the combination of both layers of information. Recently, researchers have shown increased interest in the development of such tools aiming at the integrative analysis of TCR and gene expression profiles. In the upcoming section, we will discuss the current landscape of computational tools specific for the analysis of scTCR-seq data. We included all tools, to the best of our knowledge, that had an associated peer-reviewed publication or pre-print article before the 1st of October, 2021. These tools build on the foundations of the rapidly progressing field of TCR repertoire analysis and provide a giant leap towards the system-based analysis of T cell immunity, thereby providing a deeper mechanistic understanding of T cell biology. Table 5 provides an overview of the different functionalities provided by the packages discussed in this review.

### 6.1. CoNGA

Integration of gene expression and TCR data typically involves mapping TCR sequence properties to cell subsets, the latter being defined by these cells' gene expression profile. This approach impedes the identification of new cell subsets by defining them up front. Schattgen and colleagues developed a graph-theoretic approach, Clonotype Neighbor Graph Analysis (CoNGA)3 [136], which aims to identify correlations between gene expression and TCR profiles in an unbiased way. CoNGA builds a similarity graph based on TCR sequence similarity (as defined by the TCRdist measure) and one based on the gene expression data. As a Python package, CoNGA is built on top of the scanpy package [134] and therefore makes use of the AnnData object to store integrated gene expression and TCR sequence data. Additionally, the package integrates an implementation of TCRdist for distance calculations between TCRs. CoNGA provides a graph-vs-graph and a graph-vs-feature analysis. The first analysis involves correlating the gene expression with the TCR sequence similarity graph by identifying clonotypes whose neighbours significantly overlap in both graphs. For each clonotype, CoNGA evaluates all components that are directly connected to this clonotype (graph neighbours) in both the TCR and gene expression graphs. A score is assigned to each clonotype, reflecting the probability that observing this degree of overlap between both graphs is greater than or equal to the overlap that would be expected by chance. To limit the number of false positives, this score is multiplied by the total number of clonotypes. In the second type of analysis, graph-vs-feature, numerical features from either property are mapped onto the similarity graph of the complementary property, thereby aiming to identify graph neighborhoods with a bias in the score distribution.

By applying CoNGA to a collection of publicly available T cell datasets, the authors identified a population of HOBIT+ expressing T cells with long CDR3s enriched for hydrophobic residues. Moreover,

**Table 5**
Tools for analyzing single-cell TCR profiles. Asterisks indicate the availability of multiple metrics. A single asterisk (*) corresponds to a single metric (e.g. only Shannon index for measuring diversity), while double asterisks (**) reflect the availability of multiple diversity or clonality metrics. Advanced visualizations may include graph representations, UMAP projections, circos plots... The Clustering column only accounts for receptor-based clustering. Clustering of samples is covered by the Repertoire overlap column. The Integration with GE column additionally indicates the single-cell RNA-seq analysis environment each tool interacts with. GE, gene expression; AIRR, adaptive immune receptor repertoire; B, BCR; T, TCR; Se, Seurat; *Sc*, Scanpy; N, native.

| Tool | Ref. | Version | Latest release | Platform | Receptor type | Integration with GE | 10x support | Pre-processing | Paired chains | Diversity/ clonality | V(D)J usage | Repertoire overlap | Advanced visualization | Clustering | Documentation | AIRR-compliant |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CoNGA | [136] | 0.1.1 | 2021–08–23 | Python | B + T | *Sc* | Yes | Yes | Yes | No | No | No | Yes | Yes | Yes | No |
| mvTCR | [95,96] | 4.2.0 | 2021–02–07 | Python, R | B + T | N | Yes | Yes | Yes | No | No | No | No | No | No | No |
| Platypus | [128] | 0.6.1 | 2021–01–30 | R | B + T | Se | Yes | No | Yes | Yes (*) | Yes | Yes | Yes | Yes | Yes | Yes |
| Scirpy | [125] | 0.6.1 | 2021–01–30 | Python | T | *Sc* | Yes | No | Yes | Yes (*) | Yes | Yes | Yes | Yes | Yes | Yes |
| scRepertoire | [127] | 1.14 | 2021–02–25 | R | B + T | Se | Yes | Yes | Yes | Yes (**) | Yes | Yes | Yes | Yes | Yes | No |
| Tessa | [144] | 1.0.0 | 2020–10–30 | Python, R | T | N | Yes | Yes | Yes | No | No | No | Yes | Yes | Yes | No |
| VDJView | [126] | N.A. | 2021–05–17 | R | B + T | Se | Yes | No | Yes | No | Yes | Yes | Yes | Yes | No | No |

they observed strong correlation between the usage of the TRBV30 gene segment and the expression of the conserved EPHB6 gene.

### 6.2. mvTCR

Supplementing gene expression profiles with functional information from the TCR provides a more detailed understanding of the behavior of different T cell subsets. Typically, this data is processed and analyzed parallel from one another, thereby impeding the identification of novel T cell phenotypes. An et al [137] developed a multiview variational autoencoder, termed mvTCR, that creates a joint embedding of gene expression and TCR sequence data at the level of individual cells. By integrating both modalities, it is possible to capture groups of T cells that are correlated on both a phenotypic and functional level. mvTCR applies two types of mixture models to integrate the transcriptomic and TCR embedding into a joint latent distribution. The authors show that this joint embedding improves separation of epitopespecific clusters in the UMAP, as compared to either gene expression or TCR embeddings individually. Consequently, mvTCR-generated embeddings for multimodal single-cell data may be used to improve existing models for predicting TCR-epitope specificity, by integrating an additional layer of phenotypic information. Alternatively, sub-clustering of epitope-specific clusters may reveal epitope-specific expansions in certain T cell subsets.

### 6.3. Platypus

Platypus [128] is an R-based software package dedicated to the analysis of single-cell immune repertoires. Platypus is optimized for data generated through the 10x Genomics Chromium platform, but it is also compatible with other barcode-based scRNA-seq approaches like RAGE-seq [25] or SplitSeq [138]. Platypus uses the Seurat platform to integrate transcriptomic profiles with V(D)J sequencing data [133]. By default, scaling and normalization of the gene expression data is performed using the default Seurat parameters, although the software also supports alternative normalization methods like SCTransform [139,140] or Harmony [141]. Platypus provides a method for extracting V(D)J sequences from Cell Ranger output and it contains a range of functions for pre-processing and calculating basic repertoire statistics. The latter include calculating the number of isotypes per clones (BCRs), CDR3 length distributions and constructing sequence logos. An interesting feature of the Platypus package is its ability to automate Seurat workflows. The results of this gene expression analysis can be subsequently integrated with clonotype information using a custom function. This allows users to project clonotype information onto the UMAP plots generated by clustering the gene expression profiles. For example, the visualize_clones_gene expression can be used to highlight expanded clones within the gene expression clusters. Finally, Platypus provides a feature for evaluating repertoire topology through the construction of sequence similarity networks.

### 6.4. Scirpy

Scirpy [125] is a Python library built on top of the Scanpy toolkit for analyzing scRNA-seq data in Python. Data can be directly imported from various sources, including Cell Ranger, TraCeR and standardized AIRR formats. Similar to Scanpy and CoNGA, Scirpy leverages the AnnData format [134], which stores a matrix of values along with annotations of observations and variables. This data structure also keeps track of additional unstructured annotations. Moreover, Scirpy follows the practices of the intuitive API of Scanpy. To integrate V(D)J and gene expression profiles, Scirpy provides functions for merging AIRR and gene expression data into a single AnnData object. Scirpy offers tools for pre-processing and analyzing TCR repertoire and gene expression data in tandem. The pre-processing procedure allows up to two $\alpha$ and $\beta$ chains per T cell, flagging any cell containing more than two of either chains as potential doublets and discarding them during the process. Analysis tools include calculation of clonotype abundance within a specific group

of samples, clonal expansion, diversity, imbalance, as well as repertoire overlap. However, the only available diversity metric is Shannon entropy. Next to this, the package provides graph visualizations of clonotype clusters with high sequence similarity, using igraph [142] or networkx [143]. The package also provides clustering where similarity is based on pairwise alignments, but also offers other distance metrics.

### 6.5. scRepertoire

scRepertoire [127] is an R package designed for the post-analysis of filtered contigs generated from the Cell Ranger pipeline. The package interacts with Seurat and SingleCellExperiment (SCE), allowing for integration of gene expression data. Various functions are provided for the visualization of T cell contigs, which includes abundance, length, gene usage and clonotype sharing plots. scRepertoire also offers more advanced types of analysis, such as clonal homeostasis (visualization of different levels of expansion) or clonal proportions (the proportions of clone sizes). Other analysis include the calculation of repertoire overlap, sample diversity and clustering of clonotypes based on the amino acid edit-distance (the number of mismatched amino acids between two sequences).

Previously described features can also be calculated for gene expression clusters. Integration with Seurat also allows projection of clonotype information on the UMAP plots. Additional advanced visualizations include alluvial plots displaying clonotypes shared across different categories. Lastly, shared clonotype gene usage patterns across cell type clusters can be analyzed using chord diagrams.

### 6.6. Tessa

Tessa is a tool that generates numerical embeddings for TCR sequence and integrates it with the gene expression profiles of T cells [144]. Numerical encoding of the TCR is based on Atchley factors [145] of the amino acids in the CDR3$\beta$ region. Tessa uses a stacked autoencoder to reduce the size of the numeric vector, while maintaining its intrinsic structural features. For the gene expression matrix, only the top 10% genes with the highest variation in expression are kept. Tessa then uses a parametric Bayesian model to identify the influence of the TCR on the gene expression profile of matched clones. In addition, tessa uses weighted TCR embeddings to cluster clones into groups that represent their antigen specificity. The algorithm alternates between both processes (correlating TCR and gene expression matrix, and antigen-specific grouping), updating the weights of the embeddings, until the model reaches convergence. Using tessa, Zhang and colleagues firstly showed that clonotypes sharing similar TCRs are more likely to also share similar gene expression profiles, as determined from the correlation between embeddings from the TCR and transcriptomic profiles. Moreover, the correlation was stronger in PBMCs from healthy donors than tumor samples of different cancer types. This may indicate a proportionally smaller influence of the TCR on the gene expression profile in tumor samples, which may be a consequence of high cyto- and chemokine secretion in the tumor microenvironment, influencing T cells transcriptionally [146].

### 6.7. VDJView

VDJView [126] integrates various R packages for analyzing scRNA (Scater, Seurat, SC3, Monocle & MAST) and V(D)J sequencing data (immunarch) into an easy-to-use R Shiny web application. As input, the software allows 3′- as well as 5′-generated scRNA-seq data (both 10x and SmartSeq2). Moreover, TCR sequences can be directly reconstructed from the input scRNA-seq data, using the VDJPuzzle software [147]. The tool offers various features for analyzing clonotype abundance, CDR3 length distributions, V(D)J gene usage and clonotype sharing. For the analysis of gene expression levels, the tool includes common dimensionality reduction techniques such as PCA, t-SNE and UMAP. Additionally, cell clustering (both supervised and unsupervised) is provided based on gene expression values. Finally, the software offers pseudo-time analysis for determining single-cell state trajectories based on the Monocle package [74].

## 7. Challenges that remain for single-cell over bulk approaches

The size and constitution of the repertoire is strongly influenced by the applied repertoire profiling technique [118]. In addition, the cell population may also influence the potential number of identified clonotypes, as some cell types may be rarer than others. Therefore, researchers must carefully evaluate the choice of single-cell approaches versus bulk, depending on the research question to be answered. Deep sampling approaches that allow the capture of large numbers of cells (e.g. leukapheresis) have revealed up to $2{\bullet}10^7$ unique clonotypes within a single sample [148]. From a practical perspective, the analysis of this amount of cells is only possible using bulk sequencing approaches. For single-cell experiments, the number of uniquely identified clonotypes is typically lower. Consequently, bulk sequencing approaches may be more appropriate when the goal of the study is to characterize the full repertoire from whole blood samples. However, when interested in the functional characteristics and the phenotypes of specific (sub)populations, one may opt for single-cell technologies. These may include situations where the number of profiled clonotypes is less relevant. For example, when studying certain epitope-specific T cells and the immune responses they elicit.

## 8. Perspective

Single-cell technologies have opened up novel opportunities for identifying specific $\alpha\beta$TCR pairs along with the functional profile of the cells they originate from. The information obtained through the use of these technologies simultaneously delivers gene expression profiles, TCR sequence information and optionally other modalities (such as peptide specificity, epigenetic modifications, chromatin accessibility, etc.). While a plethora of techniques have been established to individually analyze these information layers, the use of single-cell technologies has offered a new way to integrate this information at the level of individual cells. This imposes a tremendous challenge from the perspective of data analysis. In this review we have discussed several excellent software modules that offer tools aimed at the integrative analysis of paired single-cell gene expression and TCR repertoire data. Although these packages provide a comprehensive toolkit for exploring and analyzing gene expression and TCR profiles, several issues remain. scTCRseq allows chain pairing, granting information about both $\alpha$ and $\beta$ chain. While this is considered a major advantage, there remains an unresolved problem with the pairing of $\alpha$ and $\beta$ chains, even in the case of single-cell sequencing. Occasionally, a single cell may express multiple productive $\alpha$ and/or $\beta$ chains [149,150]. In this case it is not possible to know which of the $\alpha\beta$ pairs is functional. It has long been known that post-translational silencing mechanisms exist that result in allelic exclusion [151,152]. Nonetheless, one question AIRR researchers should address is: what determines functional chain pairing in TCRs? Additionally, although rare, there is always the possibility of sequencing only one TRA and one TRB, while the cell may in fact express multiple TRAs and/or TRBs. Moreover, the identified TRA and TRB may not even match as they might only pair the other, unidentified chains. Consequently, this raises the question: whether identified $\alpha\beta$ pairs in single-cell experiments are truly functional rearrangements? Another consideration is the fact that gene expression and TCR data obtained during single-cell experiments are typically processed and analyzed individually using established approaches from the fields of scRNA-seq and TCR-seq analysis. Integration is often limited to the projection of clonotype features on gene expression-based UMAP. Therefore, we advocate for the development of novel methodologies that integrate information from both sources into a reciprocal metric. Several approaches already adopt this

idea, including CoNGA [136], mvTCR [137], and tessa [144]. Integrative approaches like this may reveal distinct subpopulations of T cells that display similar gene expression and TCR sequence characteristics. Such observations may be explained by the expansion of certain T cell subpopulations, triggered by an immunogenic peptide. Similarly, UMAP is commonly applied to gene expression matrices to project distinct cell subpopulations based on a set of highly variable genes. Little attention has been focused on applying UMAP to a combination of gene expression and TCR features. Such an approach may reveal distinct clusters of epitope-specific cells that could not be identified from gene expression or TCR profile features individually. This idea has been outlined by An et al. [137] who developed a variational autoencoder, mvTCR, to generate a joint embedding for gene expression and TCR sequence information, thereby improving separation of epitope-specific clusters in the UMAP.

Another major challenge for scTCR-seq is the development of improved visualizations. Currently, the most common approach for visualizing scRNAseq data is a UMAP. UMAPs can be annotated with additional layers of information such as clonal expansion amongst others. For TCR sequences, the similarity network representation is one of the most common visualizations. Although this representation provides a general overview of the repertoire architecture and highlights clonal expansions [153], network representations become infeasible when the number of nodes is very large as is often the case for AIRR-seq data [154]. It is imperative to extract relevant subsets of the TCR network (e.g. expanded clonotype clusters with low generation probability), thus enabling visualization. In addition, features from gene expression space can be mapped onto clonotype similarity networks. Such network representations enable the identification of clonotype clusters with similar expression profiles, potentially indicating common origin or preferential differentiation of a cellular subtype. Conversely, the observation of clonotype clusters with distinct transcriptomic profiles may indicate phenotypic plasticity between celltypes. These visualization strategies map features from one modality (TCR or gene expression) onto the other, but do not truly integrate both layers. Hence, there is a need for improved visualization techniques that captures both gene expression and TCR features by integrating them.

With the advent of novel experimental and computational approaches for determining the specificity of T cells, scTCR-seq combined with scRNA-seq will be an indispensable tool for fully characterizing the complete molecular profile of a T cell. Several methods exist that accurately predict the binding of any TCR with a known epitope, based on an epitope-specific model [30,155,115,156,26]. These are often referred to as seen epitopes. A major downside to these models is that they require sufficient data for a single epitope in order to make accurate prediction about which TCRs bind to it. In addition, these models are typically trained using only $\beta$ chain information, thereby neglecting potential contribution of the less diverse $\alpha$ chain. Predicting the binding of a TCR with an unseen epitope is a problem that is considerably more difficult. Nonetheless, multiple studies have illustrated the possibility of solving this problem using deep neural networks [157,158]. A common conclusion is that predictions for epitopes similar to known epitopes are superior to vastly different ones. One of the current limitations is the low amount of known high-quality TCR-epitope pairs. However, due to the introduction of high-throughput methods for TCR-antigen screening, more data will become available which will allow the construction of more accurate models for predicting the specificity of any TCR sequence. Finally, we encourage the use of standardized pipelines for processing and analyzing scTCR-seq data, which will result in more transparency and improved comparability between scTCR-seq studies.

## Declaration of Competing Interest

BO, KL and PM hold shares in ImmuneWatch BV, an immunoinformatics company.

## CRediT authorship contribution statement

**Sebastiaan Valkiers:** Conceptualization, Visualization, Writing – original draft, Writing – review & editing. **Nicky de Vrij:** Conceptualization, Writing – original draft, Writing – review & editing. **Sofie Gielis:** Writing – review & editing. **Sara Verbandt:** Writing – review & editing. **Benson Ogunjimi:** Writing – review & editing. **Kris Laukens:** Supervision, Writing – review & editing. **Pieter Meysman:** Conceptualization, Supervision, Writing – review & editing.

## Acknowledgments

This work was supported by the Research Foundation Flanders (FWO: 1S40321N to SV 1S71721N to NDV and 1S48819N to SG), the iBOF Modulating Immunity and the Microbiome for Effective CRC Immunotherapy (MIMICRY) Project, and the Flemish Government under the "Onderzoeksprogramma Artifici¨ele Intelligentie (AI) Vlaanderen" Program.

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.immuno.2022.100009.

## References

[1] Shah Kinjal, Al-Haidari Amr, Sun Jianmin, Kazi Julhash U. T cell receptor (tcr) signaling in health and disease. Signal Transduct Target Ther 2021;6(1):1–26.
[2] Davis Mark M, Bjorkman Pamela J. T-cell antigen receptor genes and t-cell recognition. Nature 1988;334(6181):395–402.
[3] Shcherbinin Dmitrii S, Belousov Vlad A, Shugay Mikhail. Comprehensive analysis of structural and sequencing data reveals almost unconstrained chain pairing in tcr$\alpha\beta$ complex. PLoS Comput Biol 2020;16(3):e1007714.
[4] Nikolich-Zugich Janko, Slifka Mark K, Messaoudi Ilhem. The many ˇ important facets of t-cell repertoire diversity. Nat Rev Immunol 2004;4(2):123–32.
[5] Zarnitsyna Veronika, Evavold Brian, Schoettle Louie, Blattman Joseph, Antia Rustom. Estimating the diversity, completeness, and cross-reactivity of the t cell repertoire. Front Immunol 2013;4:485.
[6] Mora Thierry, Walczak Aleksandra M. Quantifying lymphocyte receptor diversity. In: Systems immunology. CRC Press; 2018. p. 183–98.
[7] Qi Qian, Liu Yi, Cheng Yong, Glanville Jacob, Zhang David, Lee Ji-Yeun, Olshen Richard A, Weyand Cornelia M, Boyd Scott D, Goronzy J¨org J. Diversity and clonal selection in the human t-cell repertoire. Proc Natl Acad Sci 2014;111(36):13139–44.
[8] Mora Thierry, Walczak Aleksandra M. How many different clonotypes do immune repertoires contain? Curr Opin Syst Biol 2019;18:104–10.
[9] Emerson Ryan O, DeWitt William S, Vignali Marissa, Gravley Jenna, Hu Joyce K, Osborne Edward J, Desmarais Cindy, Klinger Mark, Carlson Christopher S, Hansen John A, et al. Immunosequencing identifies signatures of cytomegalovirus exposure history and hla-mediated effects on the t cell repertoire. Nat Genet 2017;49(5):659–65.
[10] Amoriello Roberta, Greiff Victor, Aldinucci Alessandra, Bonechi Elena, Carnasciali Alberto, Peruzzi Benedetta, Repice Anna Maria, Mariottini Alice, Saccardi Riccardo, Mazzanti Benedetta, et al. The tcr repertoire reconstitution in multiple sclerosis: comparing one-shot and continuous immunosuppressive therapies. Front Immunol 2020;11:559.
[11] Picot Julien, Guerin Coralie L, Le Van Kim Caroline, Boulanger Chantal M. Flow cytometry: retrospective, fundamentals and recent instrumentation. Cytotechnology 2012;64(2):109–30.
[12] Perfetto Stephen P, Chattopadhyay Pratip K, Roederer Mario. Seventeen–colour flow cytometry: unravelling the immune system. Nat Rev Immunol 2004;4(8):648–55.
[13] Bandura Dmitry R, Baranov Vladimir I, Ornatsky Olga I, Antonov Alexei, Kinach Robert, Lou Xudong, Pavlov Serguei, Vorobiev Sergey, Dick John E, Tanner Scott D. Mass cytometry: technique for real time single cell multitarget immunoassay based on inductively coupled plasma time-of-flight mass spectrometry. Anal Chem 2009;81(16):6813–22.
[14] Devi Mani, Vijayalakshmi Dhanaraj, Dhivya Kumar, Janane Murali. Memory t cells (cd45ro) role and evaluation in pathogenesis of lichen planus and lichenoid mucositis. J Clin Diagn Res: JCDR 2017;11(5):ZC84.
[15] Zappia Luke, Theis Fabian J. Over 1000 tools reveal trends in the single-cell RNA-Seq analysis landscape. Genome Biol 2021;22(1):1–18.
[16] Stoeckius Marlon, Zheng Shiwei, Houck-Loomis Brian, Hao Stephanie, Yeung Bertrand Z, Mauck William M, Smibert Peter, Satija Rahul. Cell hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. Genome Biol 2018;19(1):1–12.
[17] Pai Joy A, Satpathy Ansuman T. High-throughput and single-cell T cell receptor sequencing technologies. Nat Methods 2021.

[18] Hwang Byungjin, Lee Ji Hyun, Bang Duhee. Single-cell rna sequencing technologies and bioinformatics pipelines. Exp Mol Med 2018;50(8):1–14.

[19] Kashima Yukie, Sakamoto Yoshitaka, Kaneko Keiya, Seki Masahide, Suzuki Yutaka, Suzuki Ayako. Single-cell sequencing techniques from individual to multiomics analyses. Exp Mol Med 2020;52(9):1419–27.

[20] Chen Wanqiu, Zhao Yongmei, Chen Xin, Yang Zhaowei, Xu Xiaojiang, Bi Yingtao, Chen Vicky, Li Jing, Choi Hannah, Ernest Ben, et al. A multicenter study benchmarking single-cell rna sequencing technologies using reference samples. Nat Biotechnol 2021;39(9):1103–14.

[21] Pasetto Anna, Lu Yong-Chen. Single-cell tcr and transcriptome analysis: an indispensable tool for studying t-cell biology and cancer immunotherapy. Front Immunol 1972;12:2021.

[22] Zemmour David, Zilionis Rapolas, Kiner Evgeny, Klein Allon M, Mathis Diane, Benoist Christophe. Single-cell gene expression reveals a landscape of regulatory t cell phenotypes shaped by the tcr. Nat Immunol 2018;19(3):291–301.

[23] Neal James T, Li Xingnan, Zhu Junjie, Giangarra Valeria, Grzeskowiak Caitlin L, Ju Jihang, Liu Iris H, Chiou Shin-Heng, Salahudeen Ameen A, Smith Amber R, et al. Organoid modeling of the tumor immune microenvironment. Cell 2018;175(7):1972–88.

[24] Tu Ang A, Gierahn Todd M, Monian Brinda, Morgan Duncan M, Mehta Naveen K, Ruiter Bert, Shreffler Wayne G, Shalek Alex K, Christopher Love J. TCR sequencing paired with massively parallel 3' RNA-Seq reveals clonotypic t cell signatures. Nat Immunol 2019;20(12):1692–9.

[25] Singh Mandeep, Al-Eryani Ghamdan, Carswell Shaun, Ferguson James M, Blackburn James, Barton Kirston, Roden Daniel, Luciani Fabio, Phan Tri Giang, Junankar Simon, et al. High-throughput targeted long-read single cell sequencing reveals the clonal and transcriptional landscape of lymphocytes. Nat Commun 2019;10(1):1–13.

[26] Springer Ido, Besser Hanan, Tickotsky-Moskovitz Nili, Dvorkin Shirit, Louzoun Yoram. Prediction of specific tcr-peptide binding from large dictionaries of tcr-peptide pairs. Front Immunol 2020;11:1803.

[27] Kamga Larisa, Gil Anna, Song Inyoung, Brody Robin, Ghersi Dario, Aslan Nuray, Stern Lawrence J, Selin Liisa K, Luzuriaga Katherine. Cdr3$\alpha$ drives selection of the immunodominant epstein barr virus (ebv) brlf1-specific cd8 t cell receptor repertoire in primary infection. PLoS Pathog 2019;15(11):e1008122.

[28] Carter Jason A, Preall Jonathan B, Grigaityte Kristina, Goldfless Stephen J, Jeffery Eric, Briggs Adrian W, Vigneault Francois, Atwal Gurinder S. Single t cell sequencing demonstrates the functional role of $\alpha\beta$ tcr pairing in cell lineage and antigen specificity. Front Immunol 2019;10:1516.

[29] Gil Anna, Kamga Larisa, Chirravuri-Venkata Ramakanth, Aslan Nuray, Clark Fransenio, Ghersi Dario, Luzuriaga Katherine, Selin Liisa K. Epstein-barr virus epitope–major histocompatibility complex interaction combined with convergent recombination drives selection of diverse t cell receptor $\alpha$ and $\beta$ repertoires. mBio 2020;11(2) e00250–e00220.

[30] Jokinen Emmi, Huuhtanen Jani, Mustjoki Satu, Heinonen Markus, L¨ahdesm¨aki Harri. Predicting recognition between t cell receptors and epitopes with tcrgp. PLoS Comput Biol 2021;17(3):e1008814.

[31] Springer Ido, Tickotsky Nili, Louzoun Yoram. Contribution of t cell receptor alpha and beta cdr3, mhc typing, v and j genes to peptide binding prediction. Front Immunol 2021;12.

[32] Zhang Wen, Hawkins Peter G, He Jing, Gupta Namita T, Liu Jinrui, Choonoo Gabrielle, Jeong Se W, Chen Calvin R, Dhanik Ankur, Dillon Myles, et al. A framework for highly multiplexed dextramer mapping and prediction of t cell receptor sequences to antigen specificity. Sci Adv 2021;7(20):eabf5835.

[33] Spindler Matthew J, Nelson Ayla L, Wagner Ellen K, Oppermans Natasha, Bridgeman John S, Heather James M, Adler Adam S, Asensio Michael A, Edgar Robert C, Lim Yoong Wearn, et al. Massively parallel interrogation and mining of natively paired human tcr$\alpha\beta$ repertoires. Nat Biotechnol 2020;38(5):609–19.

[34] Bassez Ayse, Vos Hanne, Dyck Laurien Van, Floris Giuseppe, Arijs Ingrid, Desmedt Christine, Boeckx Bram, Bempt Marlies Vanden, Nevelsteen Ines, Lambein Kathleen, et al. A single-cell map of intratumoral changes during anti-pd1 treatment of patients with breast cancer. Nat Med 2021;27(5):820–32.

[35] Zhang Ji-Yuan, Wang Xiang-Ming, Xing Xudong, Xu Zhe, Zhang Chao, Song Jin-Wen, Fan Xing, Xia Peng, Fu Jun-Liang, Wang Si-Yu, et al. Single-cell landscape of immunological responses in patients with covid-19. Nat Immunol 2020;21(9):1107–18.

[36] Schmid Katharina T, H¨ollbacher Barbara, Cruceanu Cristiana, B¨ottcher Anika, Lickert Heiko, Binder Elisabeth B, Theis Fabian J, Heinig Matthias. scpower accelerates and optimizes the design of multi-sample single cell transcriptomic studies. Nat Commun 2021;12(1):1–18.

[37] Abrams Douglas, Kumar Parveen, Krishna Murthy Karuturi R, George Joshy. A computational method to aid the design and analysis of single cell rna-seq experiments for cell type identification. BMC Bioinf 2019;20(11):1–6.

[38] Davis Alexander, Gao Ruli, Navin Nicholas E. Scopit: sample size calculations for single-cell sequencing experiments. BMC Bioinf 2019;20(1):1–6.

[39] Luecken Malte D, Theis Fabian J. Current best practices in single cell RNA-Seq analysis: a tutorial. Mol Syst Biol 2019;15(6):e8746.

[40] You Yue, Tian Luyi, Su Shian, Dong Xueyi, Jabbari Jafar S, Hickey Peter F, Ritchie Matthew E. Benchmarking umi-based single-cell RNA-Seq preprocessing workflows. Genome Biol 2021;22(1):1–32.

[41] Barron Martin, Li Jun. Identifying and removing the cell-cycle effect from single–cell rna-sequencing data. Sci Rep 2016;6(1):1–10.

[42] Van der Maaten Laurens, Hinton Geoffrey. Visualizing data using t-sne. J Mach Learn Res 2008;9(11).

[43] Leland McInnes, John Healy, and James Melville. Umap: Uniform manifold approximation and projection for dimension reduction. arXiv preprint arXiv:1802.03426, 2018.

[44] Blondel Vincent D, Guillaume Jean-Loup, Lambiotte Renaud, Lefebvre Etienne. Fast unfolding of communities in large networks. J Stat Mech: Theory Exp 2008;2008(10):P10008.

[45] Waltman Ludo, Eck Nees Jan Van. A smart local moving algorithm for large-scale modularity-based community detection. Eur Phys. J. B 2013;86(11):1–14.

[46] Ozaki Naoto, Tezuka Hiroshi, Inaba Mary. A simple acceleration method for the louvain algorithm. Int. J. Comput Electr Eng 2016;8(3):207.

[47] Bae Seung-Hee, Halperin Daniel, West Jevin D, Rosvall Martin, Howe Bill. Scalable and efficient flow-based community detection for large-scale graph analysis. ACM Transac Knowl Discov Data (TKDD) 2017;11(3):1–30.

[48] Traag Vincent A. Faster unfolding of communities: Speeding up the Louvain algorithm. Phys Rev E 2015;92(3):032801.

[49] Traag Vincent A, Waltman Ludo, Eck Nees Jan Van. From Louvain to Leiden: guaranteeing well-connected communities. Sci Rep 2019;9(1):1–12.

[50] Aran Dvir, Looney Agnieszka P, Liu Leqian, Wu Esther, Fong Valerie, Hsu Austin, Chak Suzanna, Naikawadi Ram P, Wolters Paul J, Abate Adam R, et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. Nat Immunol 2019;20(2):163–72.

[51] Hao Yuhan, Hao Stephanie, Andersen-Nissen Erica, Mauck William M III, Zheng Shiwei, Butler Andrew, Lee Maddie J, Wilk Aaron J, Darby Charlotte, Zager Michael, et al. Integrated analysis of multimodal single-cell data. Cell 2021.

[52] Andreatta Massimo, Corria-Osorio Jesus, M¨uller S¨oren, Cubas Rafael, Coukos George, Carmona Santiago J. Interpretation of t cell states from single-cell transcriptomics data using reference atlases. Nat Commun 2021;12(1):1–19.

[53] Hughes Travis K, Wadsworth Marc H II, Gierahn Todd M, Do Tran, Weiss David, Andrade Priscila R, Ma Feiyang, Silva Bruno J de Andrade, Shao Shuai, Tsoi Lam C, et al. Second-strand synthesisbased massively parallel scRNA-seq reveals cellular states and molecular features of human inflammatory skin pathologies. Immunity 2020;53(4):878–94.

[54] Villani Alexandra-Chlo´e, Satija Rahul, Reynolds Gary, Sarkizova Siranush, Shekhar Karthik, Fletcher James, Griesbeck Morgane, Butler Andrew, Zheng Shiwei, Lazo Suzan, et al. Single-cell rna-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. Science 2017;356(6335).

[55] Dutertre Charles-Antoine, Becht Etienne, Irac Sergio Erdal, Khalilnezhad Ahad, Narang Vipin, Khalilnezhad Shabnam, Ng Pei Y, Hoogen Lucas L van den, Leong Jing Yao, Lee Bernett, et al. Singlecell analysis of human mononuclear phagocytes reveals subset-defining markers and identifies circulating inflammatory dendritic cells. Immunity 2019;51(3):573–89.

[56] Vidya Vijayan KK, Karthigeyan Krithika Priyadarshini, Tripathi Srikanth P, Hanna Luke Elizabeth. Pathophysiology of cd4+ t-cell depletion in hiv-1 and hiv-2 infections. Front Immunol 2017;8:580.

[57] Wohnhaas Christian T, Leparc Germ´an G, Fernandez-Albert Francesc, Kind David, Gantner Florian, Viollet Coralie, Hildebrandt Tobias, Baum Patrick. Dmso cryopreservation is the method of choice to preserve cells for droplet-based single-cell rna sequencing. Sci Rep 2019;9(1):1–14.

[58] Ilicic Tomislav, Kim Jong Kyoung, Kolodziejczyk Aleksandra A, Bagger Frederik Otzen, McCarthy Davis James, Marioni John C, Teichmann Sarah A. Classification of low quality cells from single-cell rna-seq data. Genome Biol 2016;17(1):1–15.

[59] Zhou Yonggang, Fu Binqing, Zheng Xiaohu, Wang Dongsheng, Zhao Changcheng, Qi Yingjie, Sun Rui, Tian Zhigang, Xu Xiaoling, Wei Haiming. Pathogenic t-cells and inflammatory monocytes incite inflammatory storms in severe covid-19 patients. Natl Sci Rev 2020;7(6):998–1002.

[60] Soneson Charlotte, Robinson Mark D. Bias, robustness and scalability in single-cell differential expression analysis. Nat Methods 2018;15(4):255–61.

[61] Jordan W Squair, Matthieu Gautier, Claudi Kathe, Mark A Anderson, Nicholas D James, Thomas H Hutson, R´emi Hudelle, Taha Qaiser, Kaya JE Matson, Quentin Barraud, et al. Confronting false discoveries in single-cell differential expression. bioRxiv, 2021.

[62] Robinson Mark D, McCarthy Davis J, Smyth Gordon K. Edger: a bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 2010;26(1):139–40.

[63] McCarthy Davis J, Chen Yunshun, Smyth Gordon K. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res 2012;40(10):4288–97.

[64] Love Michael I, Huber Wolfgang, Anders Simon. Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. Genome Biol 2014;15(12):1–21.

[65] Finak Greg, McDavid Andrew, Yajima Masanao, Deng Jingyuan, Gersuk Vivian, Shalek Alex K, Slichter Chloe K, Miller Hannah W, Juliana McElrath M, Prlic Martin, et al. Mast: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell rna sequencing data. Genome Biol 2015;16(1):1–13.

[66] Andrews Tallulah S, Kiselev Vladimir Yu, McCarthy Davis, Hemberg Martin. Tutorial: guidelines for the computational analysis of singlecell rna sequencing data. Nat Protoc 2021;16(1):1–9.

[67] Zhang Lei, Yu Xin, Zheng Liangtao, Zhang Yuanyuan, Li Yansen, Fang Qiao, Gao Ranran, Kang Boxi, Zhang Qiming, Huang Julie Y, et al. Lineage tracking reveals dynamic relationships of t cells in colorectal cancer. Nature 2018;564(7735):268–72.

[68] Subramanian Aravind, Tamayo Pablo, Mootha Vamsi K, Mukherjee Sayan, Ebert Benjamin L, Gillette Michael A, Paulovich Amanda, Pomeroy Scott L, Golub Todd R, Lander Eric S, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genomewide expression profiles. Proc Natl Acad Sci 2005;102(43):15545–50.
[69] Liberzon Arthur, Birger Chet, Thorvaldsdóttir Helga, Ghandi Mahmoud, Mesirov Jill P, Tamayo Pablo. The molecular signatures database hallmark gene set collection. Cell Syst 2015;1(6):417–25.
[70] Jassal Bijay, Matthews Lisa, Viteri Guilherme, Gong Chuqiao, Lorente Pascual, Fabregat Antonio, Sidiropoulos Konstantinos, Cook Justin, Gillespie Marc, Haw Robin, et al. The reactome pathway knowledgebase. Nucleic Acids Res 2020;48(D1):D498–503.
[71] Ashburner Michael, Ball Catherine A, Blake Judith A, Botstein David, Butler Heather, Michael Cherry J, Davis Allan P, Dolinski Kara, Dwight Selina S, Eppig Janan T, et al. Gene ontology: tool for the unification of biology. Nat Genet 2000;25(1):25–9.
[72] The Gene Ontology Consortium. The gene ontology resource: enriching a gold mine. Nucleic Acids Res 2021;49(D1):D325–34.
[73] Berge Koen Van den, Bezieux Hector Roux De, Street Kelly, Saelens Wouter, Cannoodt Robrecht, Saeys Yvan, Dudoit Sandrine, Clement Lieven. Trajectory-based differential expression analysis for singlecell sequencing data. Nat Commun 2020;11(1):1–13.
[74] Trapnell Cole, Cacchiarelli Davide, Grimsby Jonna, Pokharel Prapti, Li Shuqiang, Morse Michael, Lennon Niall J, Livak Kenneth J, Mikkelsen Tarjei S, Rinn John L. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat Biotechnol 2014;32(4):381–6.
[75] Street Kelly, Risso Davide, Fletcher Russell B, Das Diya, Ngai John, Yosef Nir, Purdom Elizabeth, Dudoit Sandrine. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. BMC Genom 2018;19(1):1–16.
[76] Manno Gioele La, Soldatov Ruslan, Zeisel Amit, Braun Emelie, Hochgerner Hannah, Petukhov Viktor, Lidschreiber Katja, Kastriti Maria E, L¨onnerberg Peter, Furlan Alessandro, et al. RNA velocity of single cells. Nature 2018;560(7719):494–8.
[77] Bergen Volker, Lange Marius, Peidli Stefan, Alexander Wolf F, Theis Fabian J. Generalizing rna velocity to transient cell states through dynamical modeling. Nat Biotechnol 2020;38(12):1408–14.
[78] Saelens Wouter, Cannoodt Robrecht, Todorov Helena, Saeys Yvan. A comparison of single-cell trajectory inference methods. Nat Biotechnol 2019;37(5):547–54.
[79] Hashimoto Kosuke, Kouno Tsukasa, Ikawa Tomokatsu, Hayatsu Norihito, Miyajima Yurina, Yabukami Haruka, Terooatea Tommy, Sasaki Takashi, Suzuki Takahiro, Valentine Matthew, et al. Single-cell transcriptomics reveals expansion of cytotoxic cd4 t cells in supercentenarians. Proc Natl Acad Sci 2019;116(48):24242–51.
[80] Giudicelli Veronique, Chaume Denys, Lefranc Marie-Paule. Imgt/v-quest, an integrated software program for immunoglobulin and t cell receptor v–j and v–d–j rearrangement analysis. Nucleic Acids Res 2004;32(suppl 2):W435–40.
[81] Alamyar Eltaf, Giudicelli V´eronique, Duroux Patrice, Lefranc Mp. Imgt/highv-quest: the imgt® web portal for immunoglobulin (ig) or antibody and t cell receptor (tr) analysis from ngs high throughput and deep sequencing. Immun Res 2012;8(1):26.
[82] Ye Jian, Ma Ning, Madden Thomas L, Ostell James M. Igblast: an immunoglobulin variable domain sequence analysis tool. Nucleic Acids Res 2013;41(W1):W34–40.
[83] Thomas Niclas, Heather James, Ndifon Wilfred, Shawe-Taylor John, Chain Benjamin. Decombinator: a tool for fast, efficient gene assignment in T-cell receptor sequences using a finite state machine. Bioinformatics 2013;29(5):542–50.
[84] Giraud Mathieu, Salson Mika¨el, Duez Marc, Villenet C´eline, Quief Sabine, Caillault Aur´elie, Grardel Nathalie, Roumier Christophe, Preudhomme Claude, Figeac Martin. Fast multiclonal clusterization of v (d) j recombinations from high-throughput sequencing. BMC Genom 2014;15(1):1–12.
[85] Zhang Wei, Du Yuanping, Su Zheng, Wang Changxi, Zeng Xiaojing, Zhang Ruifang, Hong Xueyu, Nie Chao, Wu Jinghua, Cao Hongzhi, et al. Imonitor: a robust pipeline for tcr and bcr repertoire analysis. Genetics 2015;201(2):459–72.
[86] Kuchenbecker Leon, Nienen Mikalai, Hecht Jochen, Neumann Avidan U, Babel Nina, Reinert Knut, Robinson Peter N. Imseq—a fast and error aware approach to immunogenetic sequence analysis. Bioinformatics 2015;31(18):2963–71.
[87] Yu Yaxuan, Ceredig Rhodri, Seoighe Cathal. Lymanalyzer: a tool for comprehensive analysis of next generation sequencing data of t cell receptors and immunoglobulins. Nucleic Acids Res 2016;44(4) e31–e31.
[88] Yang Xi, Liu Di, Lv Na, Zhao Fangqing, Liu Fei, Zou Jing, Chen Yan, Xiao Xue, Wu Jun, Liu Peipei, et al. Tcrklass: a new k-string–based algorithm for human and mouse tcr repertoire characterization. J Immunol 2015;194(1):446–54.
[89] Gerritsen Bram, Pandit Aridaman, Andeweg Arno C, Boer Rob J De. RTCR: a pipeline for complete and accurate recovery of t cell repertoires from high throughput sequencing data. Bioinformatics 2016;32(20):3098–106.
[90] Hung Sheng-Jou, Chen Yi-Lin, Chu Chia-Hung, Lee Chuan-Chun, Chen Wan-Li, Lin Ya-Lan, Lin Ming-Ching, Ho Chung-Liang, Liu Tsunglin. Trig: a robust alignment pipeline for non-regular t-cell receptor and immunoglobulin sequences. BMC Bioinf 2016;17(1):1–9.
[91] Bolotin Dmitriy A, Poslavsky Stanislav, Mitrophanov Igor, Shugay Mikhail, Mamedov Ilgar Z, Putintseva Ekaterina V, Chudakov Dmitriy M. Mixcr: software for comprehensive adaptive immunity profiling. Nat Methods 2015;12(5):380–1.
[92] Heather James M, Ismail Mazlina, Oakes Theres, Chain Benny. High-throughput sequencing of the t-cell receptor repertoire: pitfalls and opportunities. Briefings Bioinf 2018;19(4):554–65.
[93] Bradley Philip, Thomas Paul G. Using t cell receptor repertoires to understand the principles of adaptive immune recognition. Annu Rev Immunol 2019;37:547–70.
[94] Brown Alex J, Snapkov Igor, Akbar Rahmad, Pavlovi´c Milena, Miho Enkelejda, Sandve Geir K, Greiff Victor. Augmenting adaptive immunity: progress and challenges in the quantitative engineering and analysis of adaptive immune receptor repertoires. Mol Syst Des Eng 2019;4(4):701–36.
[95] Heiden Jason A Vander, Yaari Gur, Uduman Mohamed, Stern Joel NH, O'Connor Kevin C, Hafler David A, Vigneault Francois, Kleinstein Steven H. presto: a toolkit for processing high-throughput sequencing raw reads of lymphocyte receptor repertoires. Bioinformatics 2014;30(13):1930–2.
[96] Gupta Namita T, Heiden Jason A Vander, Uduman Mohamed, Gadala-Maria Daniel, Yaari Gur, Kleinstein Steven H. Change-o: a toolkit for analyzing large-scale b cell immunoglobulin repertoire sequencing data. Bioinformatics 2015;31(20):3356–8.
[97] Gadala-Maria Daniel, Yaari Gur, Uduman Mohamed, Kleinstein Steven H. Automated analysis of high-throughput b-cell sequencing data reveals a high frequency of novel immunoglobulin v gene segment alleles. Proc Natl Acad Sci 2015;112(8):E862–70.
[98] Nouri Nima, Kleinstein Steven H. A spectral clustering-based method for identifying clones from high-throughput b cell repertoire sequencing data. Bioinformatics 2018;34(13):i341–9.
[99] Bolen Christopher R, Rubelt Florian, Heiden Jason A Vander, Davis Mark M. The repertoire dissimilarity index as a method to compare lymphocyte receptor repertoires. BMC Bioinf 2017;18(1):1–8.
[100] Peres Ayelet, Gidoni Moriah, Polak Pazit, Yaari Gur. Rabhit: R antibody haplotype inference tool. Bioinformatics 2019;35(22):4840–2.
[101] Hoehn Kenneth B, Gall Astrid, Bashford-Rogers Rachael, Fidler SJ, Kaye S, Weber JN, McClure MO, Trial Investigators SPARTAC, Kellam Paul, Pybus Oliver G. Dynamics of immunoglobulin sequence diversity in HIV-1 infected individuals. Philos Trans R Soc B: Biol Sci 2015;370(1676):20140241.
[102] Olson Branden J, Moghimi Pejvak, Schramm Chaim A, Obraztsova Anna, Ralph Duncan, Heiden Jason A Vander, Shugay Mikhail, Shepherd Adrian J, Lees William, Matsen Frederick A IV. Sumrep: a summary statistic framework for immune receptor repertoire comparison and model validation. Front Immunol 2019;10:2533.
[103] Heiden Jason Anthony Vander, Marquez Susanna, Marthandan Nishanth, Bukhari Syed Ahmad Chan, Busse Christian E, Corrie Brian, Hershberg Uri, Kleinstein Steven H, Matsen IV, Frederick A, et al. Airr community standardized representations for annotated immune repertoires. Front Immunol 2018;9:2206.
[104] ImmunoMind Team. Immunarch: an R Package for painless bioinformatics analysis of T-Cell and B-cell immune repertoires, August 2019.
[105] Bagaev Dmitry V, Vroomans Renske MA, Samir Jerome, Stervbo Ulrik, Rius Cristina, Dolton Garry, Greenshields-Watson Alexander, Attaf Meriem, Egorov Evgeny S, Zvyagin Ivan V, et al. Vdjdb in 2019: database extension, new analysis infrastructure and a t-cell receptor motif compendium. Nucleic Acids Res 2020;48(D1):D1057–62.
[106] Tickotsky Nili, Sagiv Tal, Prilusky Jaime, Shifrut Eric, Friedman Nir. Mcpas-tcr: a manually curated catalogue of pathology-associated t cell receptor sequences. Bioinformatics 2017;33(18):2924–9.
[107] Zhang Wei, Wang Longlong, Liu Ke, Wei Xiaofeng, Yang Kai, Du Wensi, Wang Shiyu, Guo Nannan, Ma Chuanchuan, Luo Lihua, et al. Pird: Pan immune repertoire database. Bioinformatics 2020;36(3):897–903.
[108] Pogorelyy Mikhail V, Shugay Mikhail. A framework for annotation of antigen specificities in high-throughput t-cell repertoire sequencing studies. Front Immunol 2019;10:2159.
[109] Ritvo Paul-Gydeon, Saadawi Ahmed, Barennes Pierre, Quiniou Valentin, Chaara Wahiba, Soufi Karim El, Bonnet Benjamin, Six Adrien, Shugay Mikhail, Mariotti-Ferrandiz Encarnita, et al. High-resolution repertoire analysis reveals a major bystander activation of tfh and tfr cells. Proc Natl Acad Sci 2018;115(38):9604–9.
[110] Murugan Anand, Mora Thierry, Walczak Aleksandra M, Callan Curtis G. Statistical inference of the generation probability of t-cell receptors from sequence repertoires. Proc Natl Acad Sci 2012;109(40):16161–6.
[111] Marcou Quentin, Mora Thierry, Walczak Aleksandra M. Highthroughput immune repertoire analysis with IGoR. Nat Commun 2018;9(1):1–10.
[112] Sethna Zachary, Elhanati Yuval, Jr Curtis G Callan, Walczak Aleksandra M, Mora Thierry. Olga: fast computation of generation probabilities of b-and t-cell receptor amino acid sequences and motifs. Bioinformatics 2019;35(17):2974–81.
[113] Pogorelyy Mikhail V, Minervina Anastasia A, Shugay Mikhail, Chudakov Dmitriy M, Lebedev Yuri B, Mora Thierry, Walczak Aleksandra M. Detecting t cell receptors involved in immune responses from single repertoire snapshots. PLoS Biol 2019;17(6):e3000314.
[114] Vita Randi, Mahajan Swapnil, Overton James A, Dhanda Sandeep Kumar, Martini Sheridan, Cantrell Jason R, Wheeler Daniel K, Sette Alessandro, Peters Bjoern. The immune epitope database (iedb): 2018 update. Nucleic Acids Res 2019;47(D1):D339–43.
[115] Gielis Sofie, Moris Pieter, Bittremieux Wout, De Neuter Nicolas, Ogunjimi Benson, Laukens Kris, Meysman Pieter. Detection of enriched t cell epitope specificity in full t cell receptor sequence repertoires. Front Immunol 2019;10:2820.
[116] Sidhom John-William, Larman H Benjamin, Pardoll Drew M, Baras Alexander S. Deeptcr is a deep learning framework for revealing sequence concepts within t-cell repertoires. Nat Commun 2021;12(1):1–12.
[117] Milena Pavlovi´c, Lonneke Scheffer, Keshav Motwani, Chakravarthi Kanduri, Radmila Kompova, Nikolay Vazov, Knut Waagan, Fabian L.M. Bernal, Alexandre Almeida Costa, Brian Corrie, Rahmad Akbar, Ghadi S. Al Hajj, Gabriel Balaban, Todd M. Brusko, Maria Chernigovskaya, Scott Christley, Lindsay G. Cowell, Robert Frank, Ivar Grytten, Sveinung Gundersen, Ingrid Hobæk Haff, Sepp Hochreiter, Eivind Hovig, Ping-Han Hsieh, G¨unter Klambauer, Marieke L. Kuijjer, Christin Lund-Andersen, Antonio Martini, Thomas Minotto, Johan Pensar, Knut Rand, Enrico Riccardi, Philippe A. Robert, Artur Rocha, Andrei Slabodkin, Igor Snapkov,

Ludvig M. Sollid, Dmytro Titov, Cédric R. Weber, Michael Widrich, Gur Yaari, Victor Greiff, and Geir Kjetil Sandve. Immuneml: an ecosystem for machine learning analysis of adaptive immune receptor repertoires. bioRxiv, 2021.
[118] Barennes Pierre, Quiniou Valentin, Shugay Mikhail, Egorov Evgeniy S, Davydov Alexey N, Chudakov Dmitriy M, Uddin Imran, Ismail Mazlina, Oakes Theres, Chain Benny, et al. Benchmarking of T cell receptor repertoire profiling methods reveals large systematic biases. Nat Biotechnol 2021;39(2):236–45.
[119] Song Li, Cohen David, Ouyang Zhangyi, Cao Yang, Hu Xihao, Liu X Shirley. Trust4: Immune repertoire reconstruction from bulk and single-cell RNA-Seq data. Nat Methods 2021:1–4 pages.
[120] Bolotin Dmitriy A, Poslavsky Stanislav, Davydov Alexey N, Frenkel Felix E, Fanchi Lorenzo, Zolotareva Olga I, Hemmers Saskia, Putintseva Ekaterina V, Obraztsova Anna S, Shugay Mikhail, et al. Antigen receptor repertoire profiling from rna-seq data. Nat Biotechnol 2017;35(10):908–11.
[121] Rizzetto Simone, Eltahla Auda A, Lin Peijie, Bull Rowena, Lloyd Andrew R, Ho Joshua WK, Venturi Vanessa, Luciani Fabio. Impact of sequencing depth and read length on single cell RNA sequencing data of t cells. Sci Rep 2017;7(1):1–11.
[122] Yang Hu-Qin, Wang Yi-Shan, Zhai Kan, Tong Zhao-Hui. Single-cell TCR sequencing reveals the dynamics of t cell repertoire profiling during pneumocystis infection. Front Microbiol 2021;12.
[123] Bilate Angelina M, London Mariya, Castro Tiago BR, Mesin Luka, Bortolatto Juliana, Kongthong Suppawat, Harnagel Audrey, Victora Gabriel D, Mucida Daniel. T cell receptor is required for differentiation, but not maintenance, of intestinal cd4+ intraepithelial lymphocytes. Immunity 2020;53(5):1001–14.
[124] Bomi Lee, Hong Namkoong, Yan Yang, Huang Huang, David Heller, Greg L Szot, Mark M Davis, Stephen J Pandol, Melena D Bellin, and Aida Habtezion. Single-cell sequencing unveils distinct immune microenvironment with ccr6-ccl20 crosstalk in human chronic pancreatitis. bioRxiv, 2021.
[125] Sturm Gregor, Szabo Tamas, Fotakis Georgios, Haider Marlene, Rieder Dietmar, Trajanoski Zlatko, Finotello Francesca. Scirpy: a scanpy extension for analyzing single-cell t-cell receptor-sequencing data. Bioinformatics 2020;36(18):4817–18.
[126] Samir Jerome, Rizzetto Simone, Gupta Money, Luciani Fabio. Exploring and analysing single cell multi-omics data with vdjview. BMC Med Genet 2020;13(1):1–9.
[127] Borcherding Nicholas, Bormann Nicholas L, Kraus Gloria. screpertoire: an r-based toolkit for single-cell immune receptor analysis. F1000Research 2020;9.
[128] Yermanos Alexander, Agrafiotis Andreas, Kuhn Raphael, Robbiani Damiano, Yates Josephine, Papadopoulou Chrysa, Han Jiami, Sandu Ioana, Weber C´edric, Bieberich Florian, et al. Platypus: an open-access software for integrating lymphocyte single-cell immune repertoires with transcriptomes. NAR Genom Bioinform 2021;3(2):lqab023.
[129] Segaliny Aude I, Li Guideng, Kong Lingshun, Ren Ci, Chen Xiaoming, Wang Jessica K, Baltimore David, Wu Guikai, Zhao Weian. Functional tcr t cell screening using single-cell droplet microfluidics. Lab Chip 2018;18(24):3733–49.
[130] Zhang Shu-Qi, Ma Ke-Yue, Schonnesen Alexandra A, Zhang Mingliang, He Chenfeng, Sun Eric, Williams Chad M, Jia Weiping, Jiang Ning. High-throughput determination of the antigen specificities of t cell receptors in single cells. Nat Biotechnol 2018;36(12):1156–9.
[131] Bentzen Amalie Kai, Marquard Andrea Marion, Lyngaa Rikke, Saini Sunil Kumar, Ramskov Sofie, Donia Marco, Such Lina, Furness Andrew JS, McGranahan Nicholas, Rosenthal Rachel, et al. Large-scale detection of antigen-specific t cells using peptide-mhc-i multimers labeled with dna barcodes. Nat Biotechnol 2016;34(10):1037–45.
[132] Manfredi Francesco, Cianciotti Beatrice Claudia, Potenza Alessia, Tassi Elena, Noviello Maddalena, Biondi Andrea, Ciceri Fabio, Bonini Chiara, Ruggiero Eliana. Tcr redirected t cells for cancer treatment: achievements, hurdles, and goals. Front Immunol 2020;11:1689.
[133] Satija Rahul, Farrell Jeffrey A, Gennert David, Schier Alexander F, Regev Aviv. Spatial reconstruction of single-cell gene expression data. Nat Biotechnol 2015;33:495–502.
[134] Wolf F Alexander, Angerer Philipp, Theis Fabian J. Scanpy: largescale single-cell gene expression data analysis. Genome Biol 2018;19(1):1–5.
[135] Shugay Mikhail, Bagaev Dmitriy V, Turchaninova Maria A, Bolotin Dmitriy A, Britanova Olga V, Putintseva Ekaterina V, Pogorelyy Mikhail V, Nazarov Vadim I, Zvyagin Ivan V, Kirgizova Vitalina I, et al. Vdjtools: unifying post-analysis of t cell receptor repertoires. PLoS Comput Biol 2015;11(11):e1004503.
[136] Stefan A. Schattgen, Kate Guion, Jeremy Chase Crawford, Aisha Souquette, Alvaro Martinez Barrio, Michael J.T. Stubbington, Paul G. Thomas, and Philip Bradley. Linking t cell receptor sequence to transcriptional profiles with clonotype neighbor graph analysis (conga). bioRxiv, 2020.
[137] Yang An, Felix Drost, Fabian Theis, Benjamin Schubert, and Mohammad Lotfollahi. Jointly learning t-cell receptor and transcriptomic information to decipher the immune response. bioRxiv, 2021.
[138] Rosenberg Alexander B, Roco Charles M, Muscat Richard A, Kuchina Anna, Sample Paul, Yao Zizhen, Graybuck Lucas T, Peeler David J, Mukherjee Sumit, Chen Wei, et al. Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. Science 2018;360(6385):176–82.
[139] Butler Andrew, Hoffman Paul, Smibert Peter, Papalexi Efthymia, Satija Rahul. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol 2018;36(5):411–20.
[140] Hafemeister Christoph, Satija Rahul. Normalization and variance stabilization of single-cell Rna-Seq data using regularized negative binomial regression. Genome Biol 2019;20(1):1–15.
[141] Korsunsky Ilya, Millard Nghia, Fan Jean, Slowikowski Kamil, Zhang Fan, Wei Kevin, Baglaenko Yuriy, Brenner Michael. Po-ru Loh, and Soumya Raychaudhuri. Fast, sensitive and accurate integration of single-cell data with harmony. Nat Methods 2019;16(12):1289–96.
[142] Csardi Gabor, Nepusz Tamas, et al. The igraph software package for complex network research. Int J Complex Syst 2006;1695(5):1–9.
[143] Hagberg Aric, Swart Pieter, Chult Daniel S. Exploring network structure, dynamics, and function using networkx, Los Alamos, NM (United States): Los Alamos National Lab(LANL); 2008. Technical report.
[144] Zhang Ze, Xiong Danyi, Wang Xinlei, Liu Hongyu, Wang Tao. Mapping the functional landscape of T cell receptor repertoires by single-t cell transcriptomics. Nat Methods 2021;18(1):92–9.
[145] Atchley William R, Zhao Jieping, Fernandes Andrew D, Dr¨uke Tanja. Solving the protein sequence metric problem. Proc Natl Acad Sci 2005;102(18):6395–400.
[146] Burkholder Brett, Huang Ren-Yu, Burgess Rob, Luo Shuhong, Jones Valerie Sloane, Zhang Wenji, Lv Zhi-Qiang, Gao Chang-Yu, Wang BaoLing, Zhang Yu-Ming, et al. Tumor-induced perturbations of cytokines and immune cell networks. Biochim Biophys Acta (BBA)-Rev Cancer 2014;1845(2):182–201.
[147] Rizzetto Simone, Koppstein David NP, Samir Jerome, Singh Mandeep, Reed Joanne H, Cai Curtis H, Lloyd Andrew R, Eltahla Auda A, Goodnow Christopher C, Luciani Fabio. B-cell receptor reconstruction from single-cell rna-seq with vdjpuzzle. Bioinformatics 2018;34(16):2846–7.
[148] Soto Cinque, Bombardi Robin G, Branchizio Andre, Kose Nurgun, Matta Pranathi, Sevy Alexander M, Sinkovits Robert S, Gilchuk Pavlo, Finn Jessica A, Crowe James E. High frequency of shared clonotypes in human B cell receptor repertoires. Nature 2019;566(7744):398–402.
[149] Padovan Elisabetta, Casorati Giulia, Dellabona Paolo, Meyer Stefan, Brockhaus Manfred, Lanzavecchia Antonio. Expression of two t cell receptor alpha chains: dual receptor t cells. Science 1993;262(5132):422–4.
[150] Schuldt Nathaniel J, Binstadt Bryce A. Dual tcr t cells: identity crisis or multitaskers? J Immunol 2019;202(3):637–44.
[151] Niederberger Nathalie, Holmberg Kaisa, Munir Alam S, Sakati Wayne, Naramura Mayumi, Gu Hua, Gascoigne Nicholas RJ. Allelic exclusion of the TCR $\alpha$-chain is an active process requiring tcr-mediated signaling and c-cbl. J Immunol 2003;170(9):4557–63.
[152] Steinel Natalie C, Brady Brenna L, Carpenter Andrea C, Yang-Iott Katherine S, Bassing Craig H. Posttranscriptional silencing of v$\beta$dj$\beta$c$\beta$ genes contributes to tcr$\beta$ allelic exclusion in mammalian lymphocytes. J Immunol 2010;185(2):1055–62.
[153] Bashford-Rogers Rachael JM, Palser Anne L, Huntly Brian J, Rance Richard, Vassiliou George S, Follows George A, Kellam Paul. Network properties derived from deep sequencing of human b-cell receptor repertoires delineate b-cell populations. Genome Res 2013;23(11):1874–84.
[154] Miho Enkelejda, Greiff Victor, Reddy Sai T, et al. Large-scale network analysis reveals the sequence space architecture of antibody repertoires. Nat Commun 2019;10(1):1–11.
[155] Neuter Nicolas De, Bittremieux Wout, Beirnaert Charlie, Cuypers Bart, Mrzic Aida, Moris Pieter, Suls Arvid, Tendeloo Viggo Van, Ogunjimi Benson, Laukens Kris, et al. On the feasibility of mining cd8+ T cell receptor patterns underlying immunogenic peptide recognition. Immunogenetics 2018;70(3):159–68.
[156] Tong Yao, Wang Jiayin, Zheng Tian, Zhang Xuanping, Xiao Xiao, Zhu Xiaoyan, Lai Xin, Liu Xiang. Sete: Sequence-based ensemble learning approach for TCR epitope binding prediction. Comput Biol Chem 2020;87:107281.
[157] Moris Pieter, De Pauw Joey, Postovskaya Anna, Gielis Sofie, De Neuter Nicolas, Bittremieux Wout, Ogunjimi Benson, Laukens Kris, Meysman Pieter. Current challenges for unseen-epitope tcr interaction prediction and a new perspective derived from image classification. Briefings Bioinf 2020.
[158] Anna Weber, Jannis Born, and María Rodríguez Martínez. Titan: T cell receptor specificity prediction with bimodal attention networks. arXiv preprint arXiv:2105.03323, 2021.
[159] Zheng Chunhong, Zheng Liangtao, Yoo Jae-Kwang, Guo Huahu, Zhang Yuanyuan, Guo Xinyi, Kang Boxi, Hu Ruozhen, Huang Julie Y, Zhang Qiming, et al. Landscape of infiltrating t cells in liver cancer revealed by single-cell sequencing. Cell 2017;169(7):1342–56.
[160] Bray Nicolas L, Pimentel Harold, Melsted Páll, Pachter Lior. Near-optimal probabilistic RNA-Seq quantification. Nat Biotechnol 2016;34(5):525–7.
[161] Melsted Páll, Booeshaghi A Sina, Liu Lauren, Gao Fan, Lu Lambda, Min Kyung Hoi Joseph, Beltrame Eduardo da Veiga. Kristján Eldjárn Hjörleifsson, Jase Gehring, and Lior Pachter. Modular, efficient and constant-memory single-cell rna-seq preprocessing. Nat Biotechnol 2021:1–6.
[162] Benjamin Kaminow, Dinar Yunusov, and Alexander Dobin. Starsolo: accurate, fast and versatile mapping/quantification of single-cell and single-nucleus rna-seq data. bioRxiv, 2021.
[163] Dongze He, Mohsen Zakeri, Hirak Sarkar, Charlotte Soneson, Avi Srivastava, and Rob Patro. Alevin-fry unlocks rapid, accurate, and memory-frugal quantification of single-cell rna-seq data. bioRxiv, 2021.
[164] Xi Nan Miles, Li Jingyi Jessica. Benchmarking computational doublet-detection methods for single-cell rna sequencing data. Cell Syst 2021;12(2):176–94.
[165] Welch Joshua D, Kozareva Velina, Ferreira Ashley, Vanderburg Charles, Martin Carly, Macosko Evan Z. Single-cell multi-omic integration compares and contrasts features of brain cell identity. Cell 2019;177(7):1873–87.
[166] Stuart Tim, Butler Andrew, Hoffman Paul, Hafemeister Christoph, Papalexi Efthymia, Mauck William M III, Hao Yuhan, Stoeckius Marlon, Smibert Peter, Satija Rahul. Comprehensive integration of single-cell data. Cell 2019;177(7):1888–902.

[167] Lotfollahi Mohammad, Wolf F Alexander, J Theis Fabian. scgen predicts single-cell perturbation responses. Nat Methods 2019;16(8):715–21.

[168] Lopez Romain, Regier Jeffrey, Cole Michael B, Jordan Michael I, Yosef Nir. Deep generative modeling for single-cell transcriptomics. Nat Methods 2018;15(12):1053–8.

[169] Messaoudi Ilhem, Patino Jose A Guevara, Dyall Ruben, LeMaoult Joël, Nikolich-Zugich Janko. Direct link between mhc polymorphism, t-cell avidity, and diversity in immune defense. Science 2002;298(5599):1797–800.

[170] Farmanbar Amir, Kneller Robert, Firouzi Sanaz. Rna sequencing identifies clonal structure of T-cell repertoires in patients with adult t-cell leukemia/lymphoma. NPJ Genom Med 2019;4(1):1–9.

[171] Aversa Ilenia, Malanga Donatella, Fiume Giuseppe, Palmieri Camillo. Molecular t–cell repertoire analysis as source of prognostic and predictive biomarkers for checkpoint blockade immunotherapy. Int J Mol Sci 2020;21(7):2378.

[172] Naylor Keith, Li Guangjin, Vallejo Abbe N, Lee Won-Woo, Koetz Kerstin, Bryl Ewa, Witkowski Jacek, Fulbright James, Weyand Cornelia M, Goronzy Jörg J. The influence of age on t cell generation and TCR diversity. J Immunol 2005;174(11):7446–52.

[173] Yager Eric J, Ahmed Mushtaq, Lanzer Kathleen, Randall Troy D, Woodland David L, Blackman Marcia A. Age-associated decline in t cell repertoire diversity leads to holes in the repertoire and impaired immunity to influenza virus. J Exp Med 2008;205(3):711–23.

[174] Boyd Scott D, Liu Yi, Wang Chen, Martin Victoria, Dunn-Walters Deborah K. Human lymphocyte repertoires in ageing. Curr Opin Immunol 2013;25(4):511–15.

[175] Britanova Olga V, Putintseva Ekaterina V, Shugay Mikhail, Merzlyak Ekaterina M, Turchaninova Maria A, Staroverov Dmitriy B, Bolotin Dmitriy A, Lukyanov Sergey, Bogdanova Ekaterina A, Mamedov Ilgar Z, et al. Age-related decrease in TCR repertoire diversity measured with deep and normalized sequence profiling. J Immunol 2014;192(6):2689–98.

[176] Rohit Arora, Harry M Burke, and Ramy Arnaout. Immunological diversity with similarity. BioRxiv, page 483131, 2018.

[177] Laydon Daniel J, Bangham Charles RM, Asquith Becca. Estimating t-cell repertoire diversity: limitations of classical estimators and a new approach. Philos Transac R Soc B: Biol Sci 2015;370(1675):20140291.

[178] Kaplinsky Joseph, Arnaout Ramy. Robust estimates of overall immune-repertoire diversity from high-throughput measurements on samples. Nat Commun 2016;7(1):1–10.

[179] Jiang Xu, Wang Shiyu, Zhou Chen, Wu Jinghua, Jiao Yuhao, Lin Liya, Lu Xin, Yang Bo, Zhang Wei, Xiao Xinyue, et al. Comprehensive TCR repertoire analysis of cd4+ t-cell subsets in rheumatoid arthritis. J Autoimmun 2020;109:102432.

[180] Zheng Ming, Zhang Xin, Zhou Yinghui, Tang Juan, Han Qing, Zhang Yang, Ni Qingshan, Chen Gang, Jia Qingzhu, Yu Haili, et al. Tcr repertoire and cdr3 motif analyses depict the role of $\alpha\beta$ t cells in ankylosing spondylitis. EBioMedicine 2019;47:414–26.

[181] Serana Federico, Sottini Alessandra, Caimi Luigi, Palermo Belinda, Natali Pier Giorgio, Nisticò Paola, Imberti Luisa. Identification of a public cdr3 motif and a biased utilization of t-cell receptor v beta and j beta chains in hla-a2/melan-a-specific t-cell clonotypes of melanoma patients. J Transl Med 2009;7(1):1–14.

[182] Dahal-Koirala Shiva, Risnes Louise Fremgaard, Christophersen Asbjørn, Sarna Vikas K, Lundin K EA, Sollid Ludvig M, Qiao ShuoWang. Tcr sequencing of single cells reactive to dq2. 5-glia-$\alpha$2 and dq2. 5-glia-$\omega$2 reveals clonal expansion and epitope-specific v-gene usage. Mucosal Immunol 2016;9(3):587–96.

[183] Greenshields-Watson Alexander, Attaf Meriem, MacLachlan Bruce J, Whalley Thomas, Rius Cristina, Wall Aaron, Lloyd Angharad, Hughes Hywel, Strange Kathryn E, Mason Georgina H, et al. Cd4+ t cells recognize conserved influenza a epitopes through shared patterns of v-gene usage and complementary biochemical features. Cell Rep 2020;32(2):107885.

[184] Gao Kai, Chen Lingyan, Zhang Yuanwei, Zhao Yi, Wan Ziyun, Wu Jinghua, Lin Liya, Kuang Yashu, Lu Jinhua, Zhang Xiuqing, et al. Germline-encoded tcr-mhc contacts promote tcr v gene bias in umbilical cord blood t cell repertoire. Front Immunol 2019;10:2064.

[185] Pogorelyy Mikhail V, Minervina Anastasia A, Touzel Maximilian Puelma, Sycheva Anastasiia L, Komech Ekaterina A, Kovalenko Elena I, Karganova Galina G, Egorov Evgeniy S, Komkov Alexander Yu, Chudakov Dmitriy M, et al. Precise tracking of vaccine-responding t cell clones reveals convergent and personalized response in identical twins. Proc Natl Acad Sci 2018;115(50):12704–9.

[186] George Elias, Pieter Meysman, Esther Bartholomeus, Nicolas De Neuter, Nina Keersmaekers, Arvid Suls, Hilde Jansens, Aisha Souquette, Hans De Reu, Evelien Smits, Eva Lion, Paul G. Thomas, Geert Mortier, Pierre Van Damme, Philippe Beutels, Kris Laukens, Viggo Van Tendeloo, and Benson Ogunjimi. Preexisting memory cd4 t cells in na¨ıve individuals confer robust immunity upon hepatitis b vaccination. bioRxiv, 2021.

[187] Chapuis Aude G, Desmarais Cindy, Emerson Ryan, Schmitt Thomas M, Shibuya Kendall, Lai Ivy, Wagener Felecia, Chou Jeffrey, Roberts Ilana M, Coffey David G, et al. Tracking the fate and origin of clinically relevant adoptively transferred cd8+ t cells *in vivo*. Sci Immunol 2017;2(8).

[188] Kim Giok, Tanuma Naoyuki, Kojima Takashi, Kohyama Kuniko, Suzuki Yoko, Kawazoe Yoko, Matsumoto Yoh. Cdr3 size spectratyping and sequencing of spectratype-derived tcr of spinal cord t cells in autoimmune encephalomyelitis. J Immunol 1998;160(1):509–13.

[189] Pickman Yishai, Dunn-Walters Deborah, Mehr Ramit. Bcr cdr3 length distributions differ between blood and spleen and between old and young patients, and tcr distributions can be used to detect myelodysplastic syndrome. Phys Biol 2013;10(5):056001.

[190] Sankar Kannan, Hoi Kam Hon, Hötzel Isidro. Dynamics of heavy chain junctional length biases in antibody repertoires. Commun Biol 2020;3(1):1–10.

[191] Greiff Victor, Weber Cédric R, Palme Johannes, Bodenhofer Ulrich, Miho Enkelejda, Menzel Ulrike, Reddy Sai T. Learning the high-dimensional immunogenomic features that predict public and private antibody repertoires. J Immunol 2017;199(8):2985–97.

[192] Ostmeyer Jared, Christley Scott, Toby Inimary T, Cowell Lindsay G. Biophysicochemical motifs in t-cell receptor sequences distinguish repertoires from tumor-infiltrating lymphocyte and adjacent healthy tissue. Cancer Res 2019;79(7):1671–80.

[193] Glanville Jacob, Huang Huang, Nau Allison, Hatton Olivia, Wagar Lisa E, Rubelt Florian, Ji Xuhuai, Han Arnold, Krams Sheri M, Pettus Christina, et al. Identifying specificity groups in the t cell receptor repertoire. Nature 2017;547(7661):94–8.

[194] Dash Pradyot, Fiore-Gartland Andrew J, Hertz Tomer, Wang George C, Sharma Shalini, Souquette Aisha, Crawford Jeremy Chase, Bridie Clemens E, Nguyen Thi HO, Kedzierska Katherine, et al. Quantifiable predictive features define epitope-specific t cell receptor repertoires. Nature 2017;547(7661):89–93.

[195] Huang Huang, Wang Chunlin, Rubelt Florian, Scriba Thomas J, Davis Mark M. Analyzing the mycobacterium tuberculosis immune response by t-cell receptor clustering with gliph2 and genome-wide antigen screening. Nat Biotechnol 2020:1–9 pages.

[196] Zhang Hongyi, Liu Longchao, Zhang Jian, Chen Jiahui, Ye Jianfeng, Shukla Sachet, Qiao Jian, Zhan Xiaowei, Chen Hao, Wu Catherine J, et al. Investigation of antigen-specific t-cell receptor clusters in human cancers. Clin Cancer Res 2020;26(6):1359–71.

[197] Valkiers Sebastiaan, Van Houcke Max, Laukens Kris, Meysman Pieter. ClusTCR: a Python interface for rapid clustering of large sets of CDR3 sequences with unknown antigen specificity. Bioinformatics 2021;06:btab446.

[198] Zhang Hongyi, Zhan Xiaowei, Li Bo. Giana allows computationally-efficient tcr clustering and multi-disease repertoire classification by isometric transformation. Nat Commun 2021;12(1):1–11.

[199] Servaas NH, Zaaraoui-Boutahar Fatiha, Wichers CGK, Ottria Andrea, Chouri Eleni, Affandi AJ, Silva-Cardoso Sandra, Kroef Maarten van der, Carvalheiro T, Wijk Femke van, et al. Longitudinal analysis of t-cell receptor repertoires reveals persistence of antigen-driven cd4+ and cd8+ t-cell clusters in systemic sclerosis. J Autoimmun 2021;117:102574.

[200] Smith Natasha L, Nahrendorf Wiebke, Sutherland Catherine, Mooney Jason P, Thompson Joanne, Spence Philip J, Cowan Graeme JM. A conserved tcr$\beta$ signature dominates a highly polyclonal t-cell expansion during the acute phase of a murine malaria infection. Front Immunol 2020;11:3055.

[201] Schultheiß Christoph, Paschold Lisa, Simnica Donjete, Mohme Malte, Willscher Edith, Wenserski Lisa von, Scholz Rebekka, Wieters Imke, Dahlke Christine, Tolosa Eva, et al. Next-generation sequencing of t and b cell receptor repertoires from covid-19 patients showed signatures associated with severity of disease. Immunity 2020;53(2):442–55.

[202] Chiou Shin-Heng, Tseng Diane, Reuben Alexandre, Mallajosyula Vamsee, Molina Irene S, Conley Stephanie, Wilhelmy Julie, McSween Alana M, Yang Xinbo, Nishimiya Daisuke, et al. Global analysis of shared t cell specificities in human non-small cell lung cancer enables hla inference and antigen discovery. Immunity 2021;54(3):586–602.

[203] Beshnova Daria, Ye Jianfeng, Onabolu Oreoluwa, Moon Benjamin, Zheng Wenxin, Fu Yang-Xin, Brugarolas James, Lea Jayanthi, Li Bo. De novo prediction of cancer-associated t cell receptors for noninvasive cancer detection. Sci Transl Med 2020;12(557).

[204] Wang Zhongfang, Zhu Lingyan, Nguyen Thi HO, Wan Yanmin, Sant Sneha, Quiñones-Parra Sergio M, Crawford Jeremy Chase, Eltahla Auda A, Rizzetto Simone, Bull Rowena A, et al. Clonally diverse cd38+ hla-dr+ cd8+ t cells persist during fatal h7n9 disease. Nat Commun 2018;9(1):1–12.

[205] Sant Sneha, Grzelak Ludivine, Wang Zhongfang, Pizzolla Angela, Koutsakos Marios, Crowe Jane, Loudovaris Thomas, Mannering Stuart I, Westall Glen P, Wakim Linda M, et al. Single-cell approach to influenza-specific cd8+ t cell receptor repertoires across different age groups, tissues, and following influenza virus infection. Front Immunol 2018;9:1453.

[206] Smith Neal P, Ruiter Bert, Virkud Yamini V, Tu Ang A, Monian Brinda, Moon James J, Christopher Love J, Shreffler Wayne G. Identification of antigen-specific tcr sequences based on biological and statistical enrichment in unselected subjects. JCI Insight 2021.

[207] Montemurro Alessandro, Schuster Viktoria, Povlsen Helle Rus, Bentzen Amalie Kai, Jurtz Vanessa, Chronister William D, Crinklaw Austin, Hadrup Sine R, Winther Ole, Peters Bjoern, et al. Nettcr-2.0 enables accurate prediction of tcr-peptide binding by using paired tcr$\alpha$ and $\beta$ sequence data. Commun Biol 2021;4(1):1–13.

[208] Shannon Paul, Markiel Andrew, Ozier Owen, Baliga Nitin S, Wang Jonathan T, Ramage Daniel, Amin Nada, Schwikowski Benno, Ideker Trey. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003;13(11):2498–504.

[209] Madi Asaf, Poran Asaf, Shifrut Eric, Reich-Zeliger Shlomit, Greenstein Erez, Zaretsky Irena, Arnon Tomer, Van Laethem Francois, Singer Alfred, Lu Jinghua, et al. T cell receptor repertoires of mice and humans are clustered in similarity networks around conserved public cdr3 sequences. Elife 2017;6:e22057.

[210] Priel Avner, Gordin Miri, Philip Hagit, Zilberberg Alona, Efroni Sol. Network representation of T-cell repertoire—a novel tool to analyze immune response to cancer formation. Front Immunol 2018;9:2913.

[211] Canzar Stefan, Neu Karlynn E, Tang Qingming, Wilson Patrick C, Khan Aly A. Basic: bcr assembly from single cells. Bioinformatics 2017;33(3):425–7.

[212] Redmond David, Poran Asaf, Elemento Olivier. Single-cell tcrseq: paired recovery of entire t-cell alpha and beta chain transcripts in t-cell receptors from single-cell rnaseq. Genome Med 2016;8(1):1–12.

[213] Stubbington Michael JT, L¨onnberg Tapio, Proserpio Valentina, Clare Simon, Speak Anneliese O, Dougan Gordon, Teichmann Sarah A. T cell fate and clonality inference from single-cell transcriptomes. Nat Methods 2016;13(4):329–32.
[214] Afik Shaked, Yates Kathleen B, Bi Kevin, Darko Samuel, Godec Jernej, Gerdemann Ulrike, Swadling Leo, Douek Daniel C, Klenerman Paul, Barnes Eleanor J, et al. Targeted reconstruction of t cell receptor sequence from single cell RNA-Seq links cdr3 length to t cell differentiation state. Nucleic Acids Res 2017;45(16) e148–e148.
[215] Eltahla Auda A, Rizzetto Simone, Pirozyan Mehdi R, BetzStablein Brigid D, Venturi Vanessa, Kedzierska Katherine, Lloyd Andrew R, Bull Rowena A, Luciani Fabio. Linking the t cell receptor to the single cell transcriptome in antigen-specific human t cells. Immunol Cell Biol 2016;94(6):604–11.