nature methods

REVIEW ARTICLE

https://doi.org/10.1038/s41592-021-01201-8

High-throughput and single-cell T cell receptor sequencing technologies

Joy A. Pai[1] and Ansuman T. Satpathy[1,2] ✉

T cells express T cell receptors (TCRs) composed of somatically recombined TCRα and TCRβ chains, which mediate recognition of major histocompatibility complex (MHC)–antigen complexes and drive the antigen-specific adaptive immune response to pathogens and cancer. The TCR repertoire in each individual is highly diverse, which allows for recognition of a wide array of foreign antigens, but also presents a challenge in analyzing this response using conventional methods. Recent studies have developed high-throughput sequencing technologies to identify TCR sequences, analyze their antigen specificities using experimental and computational tools, and pair TCRs with transcriptional and epigenetic cell state phenotypes in single cells. In this Review, we highlight these technological advances and describe how they have been applied to discover fundamental insights into T cell-mediated immunity.

T cells are critical effector cells of the adaptive immune system, providing protection against pathogens and cancer while maintaining immune self-tolerance. This broad range of functions is enabled by the diversity of T cell phenotypes and antigen specificities that exists in an individual. T cells can adopt a wide spectrum of phenotypes, ranging from highly cytotoxic effector T cells to anti-inflammatory regulatory T cells. This phenotypic specification is largely driven by the specific TCR expressed on each T cell's surface, which mediates antigen recognition and T cell responses and presents another element of diversity within T cell populations. The TCR is composed of two chains, TCRα and TCRβ, which are produced through combinatorial somatic rearrangement of multiple variable (V), diversity (D) (for the β-chain only), joining (J) and constant (C) gene segments[1]. This recombination process, in conjunction with nontemplated addition or deletion of nucleotides between spliced gene segments, generates a TCR repertoire with a theoretical diversity of approximately 2×10^{19} unique TCRαβ pairs[1–3], although the diversity in an individual is likely lower owing to selection pressures and the numbers of T cells present[4–6]. This vast diversity, particularly in the hypervariable complementary-determining region 3 (CDR3) of the TCR[1,7], enables the recognition of an extensive array of foreign and self-peptides presented on MHC (pMHC) complexes.

There are several aspects of the T cell response that can be best understood through analysis of the TCR repertoire. Because of the low probability of somatic recombination producing an exact V(D)J rearrangement twice in an individual[3], the TCR sequence can be used as a unique identifier of T cell clones. This feature is a useful proxy for measuring (1) antigen-driven clonal expansion of T cells as an indicator of antigen specificity and response (Fig. 1a,d) and (2) longitudinal clonal dynamics and heterogeneity of the T cell response (Fig. 1b,c). Moreover, joint assessment of the TCR and cellular phenotypes in single cells can provide information on T cell differentiation pathways and the role of TCR sequence and specificity in T cell selection, activation and phenotypic specification (Fig. 1b). This information is useful not only in understanding the etiology and pathology of immune-mediated diseases, but also in designing therapeutic strategies. For example, tracking TCR diversity in patients across time points can reveal patterns of T cell clonal dynamics that correlate with treatment response or other clinically relevant features, and pairing clonal information with cellular phenotype can identify disease-relevant cell types[8–12].

In these contexts, meaningful biological interpretation of data is enabled by the use of sensitive and accurate methods to profile the TCR repertoire. Early TCR analysis methods (reviewed previously[13–15]) revealed fundamental features of the TCR repertoire but were often limited by their ability to capture only partial sequence information. For example, CDR3 spectratyping, which measures the distribution of CDR3 sequence lengths by the separation of PCR amplicons in electrophoresis[16,17], was used to provide the first estimates of human and mouse TCR diversity[4,18]. However, this approach targets only specific V–J rearrangements and does not directly read out sequence information, which is necessary for determining T cell clonality and specificity. A second approach relied on flow cytometry using anti-TRBV monoclonal antibodies, which provided insights into repertoire diversity based on TCRβ V gene usage but did not capture sequence information or junctional diversity in the CDR3 (refs. [19,20]).

The advent of genomic sequencing enabled early TCR sequencing methods that could capture TCRs at nucleotide resolution by molecular cloning and Sanger sequencing, delivering more precise measurements of CDR3 diversity[21]. These methods were used in pioneering studies of the TCR repertoire in a wide range of contexts, including in immune-mediated diseases, infections and cancer, leading, for example, to the identification of disease-specific expansion of certain TCRβ genes[21]. Importantly, these methods also enabled the first estimations of TCRα and TCRβ sequence diversity and provided evidence that the naive T cell repertoire is highly diverse while the memory T cell compartment contributes only a small proportion of the total TCR repertoire diversity in an individual[4,18]. However, these techniques are generally low in throughput (10–1,000 TCRs per assay), which precludes extensive sampling of the TCR diversity present in a given sample or population.

As next-generation sequencing technologies have become commonplace over the past two decades, the sensitivity of TCR detection and ease of applying TCR discovery tools have greatly improved,

[1]Program in Immunology, Stanford University School of Medicine, Stanford, CA, USA. [2]Department of Pathology, Stanford University School of Medicine, Stanford, CA, USA. ✉e-mail: satpathy@stanford.edu

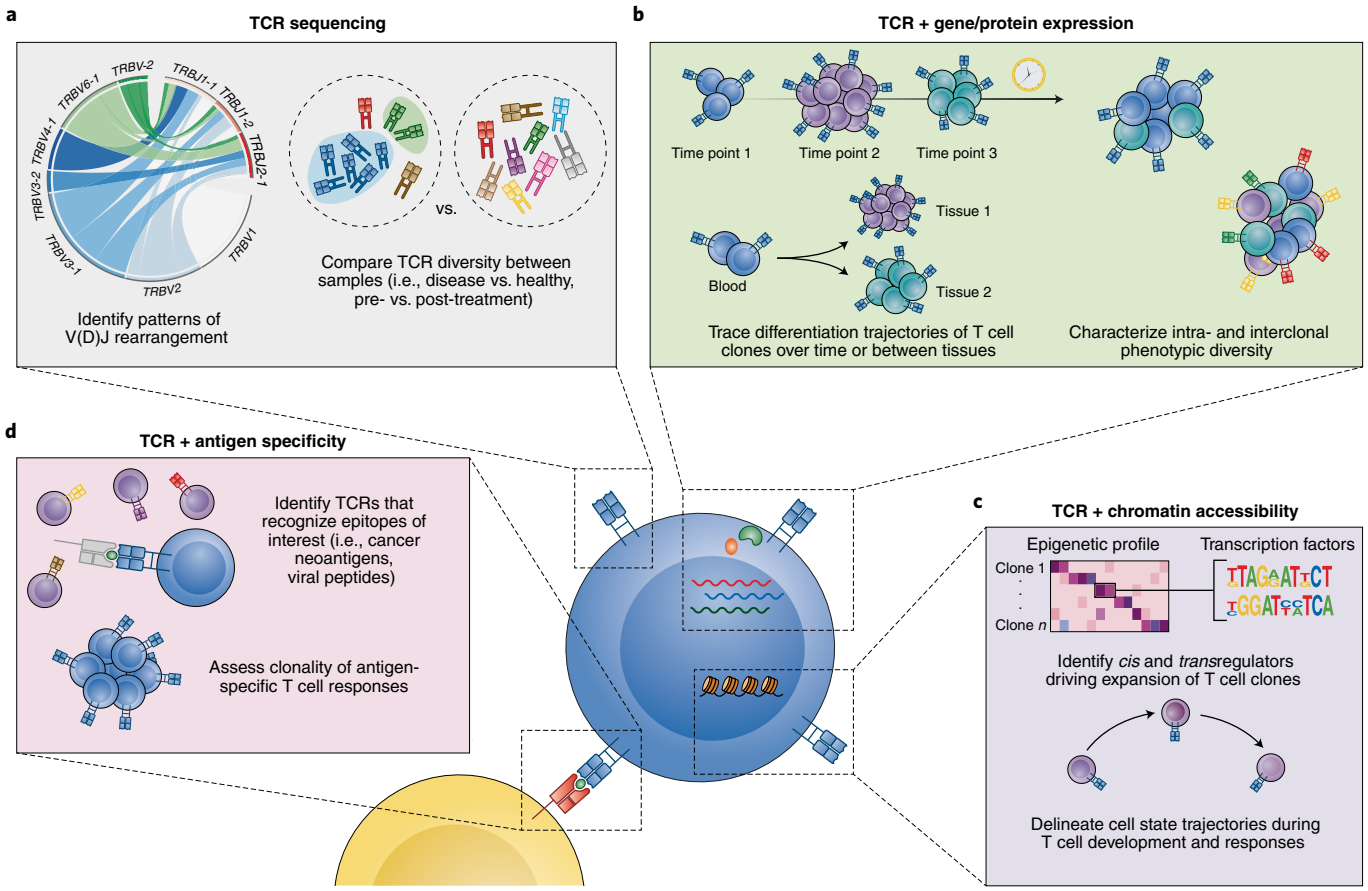

Fig. 1 | Characterization of T cell dynamics using multiomic TCR sequencing approaches. a, TCR sequencing through bulk or single-cell methods enables profiling of TCR repertoire diversity and clonality. **b**, Paired TCR sequencing and profiling of gene or protein expression allows for joint analysis of T cell clonality and phenotype and enables the tracing of clones over time or throughout tissues. **c**, TCR sequencing combined with measuring chromatin accessibility reveals clonal epigenetic signatures and uncovers developmental trajectories that may be obscured at the level of gene or protein expression. **d**, TCR epitope discovery methods identify antigens driving the T cell response.

allowing for the capture of 10^3–10^6 TCRs per assay and analysis of TCR sequences in their cellular contexts. In this Review, we discuss recent advances in high-throughput methods for interrogating the TCR repertoire at bulk and single-cell resolution, as well as techniques for pairing TCR sequences with antigen specificity, cell state and phenotype. In particular, we highlight how these approaches have enabled deeper insights into T cell clonal dynamics, differentiation and response trajectories, and function in health and disease.

Molecular strategies for TCR sequencing

Profiling the TCR repertoire using high-throughput sequencing technologies has enabled the analysis of millions of T cells in a single experiment[3] (Fig. 1a). These methods generally use one of two amplification strategies: (1) multiplex PCR[22–24] or (2) rapid amplification of 5′ complementary ends (5′ RACE)[25–29].

Multiplex PCR. Because of the large sequence diversity of TCR V genes, one pair of primers is not sufficient to capture all TCR transcripts. To address this challenge, the first approach developed for targeted sequencing of the TCR used multiplex PCR reactions[22–24]. Multiplex PCR methods use a set of forward primers complementary to all known V genes and a set of reverse primers for either J or C regions, depending on whether the starting material is genomic DNA (gDNA) or complementary DNA (cDNA), respectively. In this primer amplification strategy, gDNA or RNA is isolated from T cells and subjected to multiple rounds of PCR using these primer sets, which also contain universal primer sequences to enable subsequent sequencing on high-throughput platforms. The use of multiplex PCR methods has led to many early insights into human TCR diversity. For instance, Robins et al. performed deep sequencing of CDR3β and demonstrated that the TCR repertoire at steady state exhibits a strong bias toward certain V–J rearrangements[23]. They also observed a significantly greater degree of naive CDR3β overlap between individuals than could be expected by chance. Taken together, these results suggest that the TCR repertoire is not randomly generated but instead may be biased toward specific rearrangements, perhaps pruned through T cell selection during thymic development or common antigen experiences[23,30].

Because this approach relies on multiple rounds of PCR before sequencing, there is potential to introduce sequencing biases and errors that could skew measurements of TCR diversity and/or generate artificial TCR sequences[31–33]. However, several strategies exist to mitigate such bias. For example, because cDNA transcripts are already spliced, a smaller set of reverse primers can be used to target the C region when using mRNA instead of gDNA as starting material, thereby reducing PCR amplification bias from multiplexed J primers[34,35]. Conversely, starting from gDNA eliminates the need for reverse transcription, which also minimizes the number of errors introduced during cDNA synthesis. Additionally, the use of synthetic TCR molecules targeting the multiplexed primers allows for

the quantification of templates before and after multiplexed PCR, thereby enabling optimization of primer concentrations and correction of amplification biases[36].

5′ RACE. A second approach for TCR sequencing is based on 5′ RACE[25,27,28]. In this method, RNA is reverse transcribed by using a reverse transcriptase enzyme with terminal transferase activity that adds untemplated C nucleotides to the 3′ end of the cDNA. A template switch oligonucleotide (TSO) containing a complementary poly(G) stretch then anchors to this untemplated region, enabling the reverse transcriptase to switch templates and continue extending the cDNA to the end of the TSO, which includes a common adaptor sequence[37]. As a result, one pair of primers targeting the 5′ adaptor and the constant region is sufficient to amplify all TCR rearrangements. While 5′ RACE strategies require fewer rounds of PCR, thereby reducing amplification biases prevalent in multiplex PCR approaches, these methods are still susceptible to errors from PCR, template switching or sequencing.

Error correction strategies for TCR amplification methods. While these TCR sequencing strategies aim to provide an accurate representation of the TCR repertoire, they are subject to amplification biases and sequencing errors, as discussed above. The introduction of unique molecular identifiers (UMIs) provided a strategy to correct these errors[38]. UMIs are random DNA sequences that are appended during cDNA synthesis and uniquely tag individual cDNA molecules[32]. This strategy has two main advantages: (1) the original distribution of TCR sequences present in the sample can be recovered by counting each UMI once, allowing for more accurate quantification of TCR clonal frequencies[39,40], and (2) PCR and sequencing errors can be corrected by grouping reads tagged with the same UMI together, so the true TCR sequence can be inferred by consensus (although errors introduced during reverse transcription will remain unresolved). Because TCRs may differ by only a few nucleotides, this correction step enables the distinction between true TCR variation and error-induced artificial diversity, resulting in more accurate estimations of TCR repertoire diversity. However, this accuracy may come at the cost of lower sensitivity, as low-frequency clones may be filtered out owing to insufficient read coverage per UMI[40]. Error correction algorithms such as those using quality scores for each nucleotide have also been developed to generate more precise measurements of clonotype frequencies within the TCR repertoire[41–44].

Comparison of TCR targeting methods. Given the technical differences among TCR sequencing strategies, users should consider several factors when analyzing the TCR repertoire. First, the availability of starting material may dictate which method is applicable. Each method can be used with RNA, allowing for sensitive detection of TCR sequences owing to higher copy numbers of RNA templates as compared to gDNA. However, if the RNA quality is poor, such as in samples obtained from chemically preserved archival tissue, gDNA multiplex PCR methods are preferable, as 5′ RACE methods require high-quality RNA as starting material[34] (Table 1). Second, the purpose of the TCR analysis may also impact the choice of method. For example, RNA measurements cannot be directly correlated with cell numbers. Therefore, if the goal is to quantify clonal expansion of a T cell population, gDNA methods are preferred, because there is one genomic copy of the TCR per cell. Similarly, because multiplex PCR uses primer sets targeting different regions of the V gene, it is not always possible to capture full-length TCR sequences spanning the entire V(D)J region. This may be sufficient for analyzing the CDR3, which contains the highest sequence variability[1,7]. However, if answering a particular biological question requires assessing variability in other regions, such as CDR1 and CDR2, then 5′ RACE methods are more suitable.

Furthermore, the technical features of each step in these TCR sequencing approaches can have a substantial influence on the resultant repertoire data and, consequently, on their interpretation. In a recent comparative study of multiplex PCR and 5′ RACE TCR sequencing methods, Barennes et al. benchmarked nine TCR sequencing protocols on the same bulk T cell sample to assess their reproducibility, replicability and sensitivity[40]. They detected method-specific repertoire profiles that were largely consistent among replicates, suggesting that each protocol imposes unique biases when capturing the TCR repertoire. By comparing TRB usage with flow cytometry frequencies, they observed that RNA-based multiplex PCR methods exhibit stronger skewing toward specific V genes than gDNA-based multiplex PCR or 5′ RACE methods, likely reflecting amplification biases due to differences in RNA transcript abundance[40]. The ability to obtain a sequenced repertoire representative of the true biological repertoire is also highly dependent on the amount of starting material, regardless of method, particularly for the detection of rare TCR sequences. If the objective of a study is to capture maximal TCR diversity or detect rare clonotypes in highly diverse samples, it may be optimal to use non-UMI 5′ RACE methods, as they are more sensitive than UMI-based methods in capturing rare clones, especially for TCRα chains[33,40]. Conversely, if the objective is to construct a representative clonal structure of the TCR repertoire or identify TCRs of interest on the basis of expansion levels, then 5′ RACE with UMI correction may be most suitable because it more faithfully recapitulates TCR clonotype frequencies in comparison to other approaches, although several replicates or higher sequencing depth may be needed to consistently capture lower-frequency clones[40].

While these bulk approaches can survey the TCR repertoire at high depth, this comes at the cost of resolving the pairing of TCRα and TCRβ chains, because only one chain is captured at a time. This is an important limitation as dimerization of the two chains ultimately determines antigen specificity[1]. Additionally, clonotype analyses are more accurately performed with TCRαβ sequences, as the same TCRβ sequence may pair with different TCRα sequences and vice versa[45]. Therefore, considering each chain alone may lead to underestimates of TCR diversity, confounded analyses of intraclonal phenotypes and an inability to accurately identify T cell antigen specificities relevant to immune responses.

High-throughput methods for paired TCRαβ capture

Capturing paired TCRαβ sequences provides more accurate resolution of the clonal structure present within the TCR repertoire and is necessary for assessing TCR function and antigen specificity. This is generally achieved by (1) computational inference of TCRαβ pairing using a combinatorics-based strategy[46,47] or (2) simultaneous TCRα and TCRβ sequencing for physically isolated single cells[10,48–52].

Combinatorial inference of TCRαβ pairing. One approach for retrieving paired TCRαβ repertoires relies on combinatorics[46]. In this assay termed pairSEQ, T cells are randomly distributed across a plate such that each well contains a subset of cells tagged with the same well-specific DNA barcode (Fig. 2a). After pooling the samples for sequencing, TCRα and TCRβ sequences that have the same set of barcodes are matched computationally and inferred to be a pair, as the likelihood of two clones sharing the same set of wells is very low owing to the combinatorial nature of TCR rearrangements. In a similar approach developed by Lee et al., CDR3αβ pairs are computationally inferred from bulk-sequenced clones sorted into wells by using a repeated sampling strategy[47]. One caveat of these frequency-based pairing methods is that only expanded clones can be resolved, therefore making these assays unsuitable for the discovery of rare TCRαβ sequences.

Table 1 | Overview of TCR sequencing methods

Method	Refs.	Platform compatibility	No. of cells assessed	Paired TCRαβ	TCR regions	TCR targeting	Gene expression	Protein expression	Chromatin accessibility	No. of antigen specificities assessed[a]	TCRαβ capture efficiency	Select advantages
Multiplex PCR (bulk)	22–24	–	$>10^5$	No	CDR3	Multiplex PCR	–	–	–	–	–	Compatible with either RNA or DNA as starting material
5′ RACE (bulk)	25–29	–	$>10^5$	No	Full length	RACE PCR	–	–	–	–	–	Reduced PCR amplification bias
Han et al.	10	Plate based	10^2–10^3	Yes	CDR3	Multiplex PCR	Targeted panel	–	–	–	~80%	Can be implemented using standard laboratory reagents and equipment; high read coverage yields high-confidence TCR sequences
pairSEQ	46	Plate based	$>10^5$	Yes	CDR3	Multiplex PCR	–	–	–	–	–	Does not require physical isolation of single cells; easily implemented with standard laboratory reagents and equipment
Emulsion RT–PCR	48–52	Droplet	10^3–10^6	Yes	CDR3	Overlap extension multiplex PCR	–	–	–	–	NAR	High TCRαβ pairing efficiency due to physical linkage of transcripts; increased sensitivity and fewer artifacts as compared to standard PCR approaches
Modified InDrop	70	Droplet	10^2–10^4	Yes	CDR3	Multiplex PCR	Whole transcriptome	–	–	–	NAR	Compatible with InDrop amplified RNA libraries
Tu et al.	71	Microwell or droplet	10^3–10^4	Yes	CDR3	Multiplex PCR	Whole transcriptome	–	–	–	~35%	Compatible with most 3′ scRNA-seq methods; applicable to preexisting amplified cDNA libraries
RAGE-seq	72	Droplet	10^3–10^4	Yes	Full length	Hybridization capture	Whole transcriptome	–	–	–	~17%	Adaptable to any 3′ or 5′ scRNA-seq method
5′ V(D)J with feature barcoding	73	Droplet	10^3–10^4	Yes	Full length	RACE PCR	Whole transcriptome	Yes	–	44	~65%	Commercially available (10x Genomics) and easy to use
ECCITE-seq	75	Droplet	10^3–10^4	Yes	Full length	RACE PCR	Whole transcriptome	Yes	–	NAR	~65%	Simultaneously captures many modalities, including sgRNA perturbations; highly modular
T-ATAC-seq	74	Microwell	10^2–10^3	Yes	CDR3	Multiplex PCR	–	–	Yes	–	~70%	Pairs TCR clonality with epigenomic state
TetTCR-seq	96	Plate based	10^2–10^3	Yes	CDR3	Multiplex PCR	–	–	–	315	~70%	Sensitive detection of rare antigen-specific T cell populations; easily scalable to large peptide libraries
MATE-seq	98	Droplet	10^2–10^4	Yes[b]	CDR3	Multiplex PCR	–	–	–	4	NAR	Sensitive detection of rare antigen-specific T cell populations; compatible with lower sample inputs

[a]Demonstrated by original paper. [b]Paired TCRαβ per peptide, not per single cell. NAR, not assessed/reported.

Single-cell TCR sequencing techniques. Recent advances in single-cell isolation techniques have made it possible to capture paired TCRαβ chain information by using the TCR amplification strategies originally introduced for bulk populations (described above). The earliest methods for paired TCRαβ analysis in single cells relied on isolating individual T cells with micromanipulation techniques such as laser microdissection and micropipetting, followed by multiplex PCR and Sanger sequencing[53]. While these approaches were instrumental in the first characterizations of the paired TCRαβ repertoire, they are time-consuming and limited in both efficiency and throughput. Currently, a common method for isolating single cells is by fluorescence-activated cell sorting (FACS)[54]. An advantage of this approach is the ability to enrich for populations of interest on the basis of surface marker expression rather than sequencing all T cells in a sample, which may be useful when analyzing rare subsets. Additionally, the high read quality and coverage (~1,000 reads per chain) afforded by this method yield high-confidence TCR sequences[10], which is crucial for analyzing clonal diversity. In this method, cells are stained with fluorescent antibodies against surface proteins and individually sorted into microtiter plates, resulting in the isolation of up to several hundred single cells at a time[55]. The TCR chains are then reverse transcribed and amplified via TCR targeting strategies such as multiplex PCR. Initial studies using this strategy yielded the first characterizations of clonal diversity in antigen-specific CD8$^+$ T cell responses to viral infection in both humans and mice and revealed a markedly higher frequency of dual-allelic TCRα expression than previously appreciated[56–58]. More recent adaptations of this FACS-based isolation approach have incorporated high-throughput sequencing through the use of PCR-based barcoding strategies[10]. For example, in an approach described by Han et al., oligonucleotide barcodes unique to each well are introduced via a PCR extension step after nested TCR amplification[10] (Fig. 2a). These barcodes preserve the cellular origin of transcripts, thus enabling samples to be pooled and sequenced together on a next-generation sequencer while maintaining TCRαβ pairing.

Another strategy for performing paired TCRαβ sequencing for individual cells is through cell-based emulsion PCR approaches[48–52]. In these methods, single cells are captured within water-in-oil emulsions containing TCR primers and RT–PCR reagents. After encapsulation and cell lysis, T cells undergo overlap extension RT–PCR (OE-PCR) with primers targeting C regions and a set of V region primers containing a complementary sequence that enables linking of TCRαβ transcripts within the emulsion, as described by Turchaninova et al.[48] (Fig. 2b). These fused products containing both chains can then be sequenced while maintaining native TCRαβ pairing. For example, Munson et al. developed an emulsion OE-PCR method to identify tumor-specific TCRαβ clones that were shared among patients with breast cancer but absent in healthy samples, providing evidence that antitumor T cell responses may also be driven by shared antigens and not exclusively by patient-specific neoantigens[49].

Microfluidic platforms that partition individual cells in microwells or droplets have substantially increased the throughput of single-cell isolation[59], leading to their adoption in many single-cell TCR sequencing methods developed in the past few years. In well-based approaches, cells are loaded onto an array of nanoliter wells at low densities to confine single cells in nanowells by gravity[59–62]. Alternatively, in droplet-based approaches, cells are flowed in an aqueous medium through microchannels on a chip filled with carrier oil at a rate and pressure such that the majority of droplets contain a single cell[63]. A recent iteration of the OE-PCR approach used a high-throughput droplet microfluidic device to capture millions of T cells in a single assay[50,51] (Fig. 2b), enabling the authors to profile the TCR repertoire of 15 healthy individuals, including six pairs of monozygotic twins[6]. By using the largest dataset of paired TCRαβ sequences collected thus far, they demonstrated that the extent of TCR repertoire overlap between unrelated individuals is orders of magnitude smaller than previously appreciated by single-chain studies, highlighting the value of paired TCRαβ data in providing meaningful insights into repertoire diversity. Furthermore, the authors found that V and J gene usage was significantly more correlated in twins than in unrelated individuals, suggesting that genetic influence has a role during thymic selection, likely driven by the presentation of antigens on identical MHC molecules[6].

Single-cell methods for pairing the TCR with T cell state and function

While capturing TCR sequence information at single-cell resolution has been a powerful technical advance, studying the TCR repertoire in the context of cell state information can provide a more complete view of the functional and physiological role of T cells in immune responses. To this end, several groups have developed techniques to couple TCR sequencing with other modalities[64–75] (Fig. 1b–d). One class of methods enables the simultaneous measurement of TCR and transcriptional profiles within single cells. This is achieved by either (1) computationally reconstructing TCR chains from single-cell RNA sequencing (scRNA-seq) reads[64–69] or (2) specifically amplifying the TCR locus in conjunction with gene expression profiles[70–73] (Fig. 3).

Pairing the TCR with T cell phenotype via computational TCR reconstruction. Computational methods that extract TCR information from scRNA-seq data commonly rely on a combination of alignment and de novo assembly steps to identify TCR-originating reads and reconstruct TCR chains (Fig. 3c and Table 2). A challenge for TCR reconstruction using prior scRNA-seq methods was that they relied on amplifying the 3′ transcript end by using the poly(A) tail as an anchor[60,76–78], which does not capture the V(D)J regions of the TCR located at the 5′ transcript end. To overcome this, newer modalities use full-length cDNA amplification and sequencing strategies, in which single cells are reverse transcribed by using a template-switching mechanism similar to that of 5′ RACE (Fig. 3b)[25,26,79,80]. After fragmentation and sequencing, individual TCR sequences can be computationally reconstructed and analyzed in the context of their phenotypic profile.

One such TCR reconstruction algorithm, known as scTCRseq, constructs full-length paired TCR sequences by using nucleotide BLAST to align reads against a database of V and C genes and assembling these reads into consensus contigs (Fig. 3c, blue line)[64]. The consensus sequences for the V and C genes are then padded with 'N' nucleotides and concatenated to produce a gapped V–C scaffold against which all reads are realigned and assembled into a final TCR sequence. Stubbington et al. developed a widely used TCR reconstruction tool, TraCeR, which uses a conceptually similar process, except a database of all possible V–J combinations is used for initial read alignment[65] (Fig. 3c, gray line). Another TCR reconstruction method, VDJPuzzle, includes an additional round of alignment and assembly where the TCR reconstructions from the first round are used as an updated database to capture TCRs that differ sufficiently from the reference genome[66,67] (Fig. 3c, gold line). In contrast to these methods that require longer reads (>50 bp), the TCR reconstruction algorithm for paired-end single cell (TRAPeS) enables TCR assembly from short (25–30 bp) reads by using mate pairs to identify CDR3-originating reads followed by an iterative dynamic programming process to extend the CDR3 sequence[68] (Fig. 3c, orange line).

By combining TCR sequence with gene expression, these computational methods have provided some of the first views of T cell clonal phenotype dynamics. For example, Stubbington et al. applied TraCeR to scRNA-seq data generated from CD4$^+$ T cells during

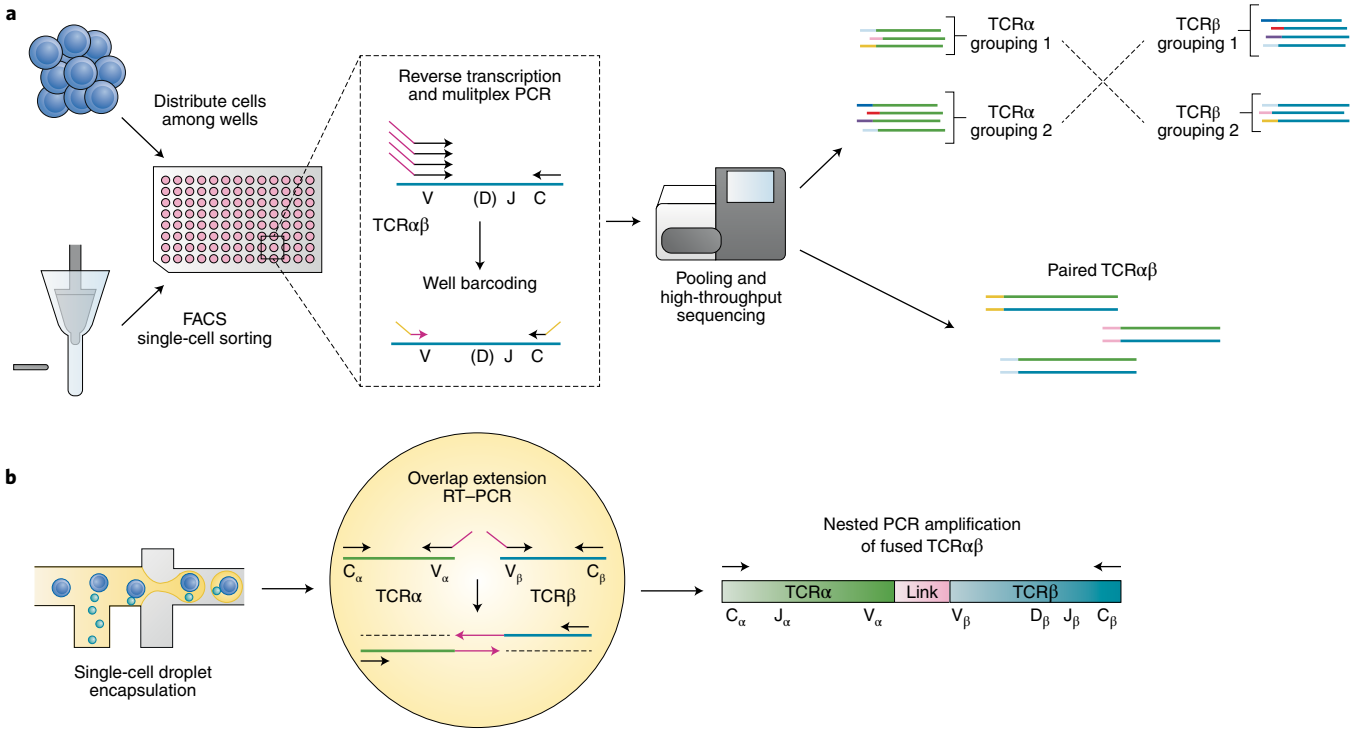

Fig. 2 | Overview of single-cell TCR sequencing approaches. a, Left, T cells are distributed among wells on a plate at a density of 10^3–10^5 cells per well in the pairSEQ method (top) or single T cells are sorted by FACS into individual wells (bottom). Middle, within each well, TCR transcripts are reverse transcribed, amplified and attached to well-specific barcodes. Right, after the amplified libraries are pooled and deep sequenced, TCRαβ pairings are computationally inferred by analyzing which TCRα and TCRβ sequences share the same set of well barcodes (top) or cellular barcodes (bottom). **b**, In emulsion-based methods, single T cells are encapsulated into droplets, where they undergo OE-PCR to enable linking of TCRα and TCRβ transcripts. Fused TCRαβ transcripts are then pooled, amplified and sequenced.

mouse *Salmonella* infection by using SMART-seq on the Fluidigm C1 system[65]. By pairing reconstructed TCRs with their cellular transcriptomes, the authors traced T cell clones along their differentiation trajectory from activated to T helper 1 (T_H1) and effector memory T cell states during infection and demonstrated that a single clone could adopt multiple cell fates[65]. This finding confirmed previous adoptive transfer studies performed in the context of viral infections that noted heterogeneity in the cell fates of antigen-reactive T cells, supporting a model of progressive differentiation during T cell responses[81–83]. Similarly, Ranasinghe et al. identified and characterized unconventional MHC class II-restricted $CD8^+$ T cell responses in human immunodeficiency virus (HIV) controllers who are able to maintain undetectable viral loads without antiretroviral therapy[84]. By using TRAPeS, they demonstrated that these antigen-specific $CD8^+$ T cells exhibit cytotoxic antiviral phenotypes and are dominated by a single expanded TCRβ clonotype within each individual, suggesting that unconventional class II-restricted $CD8^+$ T cell responses may have a previously unappreciated role in human viral infection that can be harnessed for future vaccine design.

While these approaches have the advantage of analyzing TCRs by using previously generated scRNA-seq data, successful TCR reconstruction is dependent on sequencing depth and the expression level of the TCR locus, which is variable among cells[65,85]. Consequently, incomplete capture of the TCR repertoire as a result of poor sequencing depth or limited starting material may skew biological interpretations regarding clonality and diversity.

Pairing the TCR with T cell phenotype via simultaneous RNA and targeted TCR sequencing. To generate TCR sequencing libraries with higher coverage, several groups have developed techniques that combine RNA profiling with targeted amplification of both TCR chains in single cells[10,70–73] (Fig. 3a). One of the first methods to accomplish this used a multiplex PCR strategy to simultaneously target the TCRα and TCRβ chains and a panel of 17 phenotyping genes that included cytokines and transcription factors important for T cell function and subset distinction[10]. This technique was used to characterize $CD4^+$ tumor-infiltrating lymphocytes (TILs) in colorectal cancer and revealed clonal sharing between $FOXP3^+RORC^+$ and $FOXP3^-RORC^+$ T helper 17 (T_H17) subpopulations, suggesting that interleukin (IL)-17-producing cells can originate from a common ancestor that expands in response to antigenic activation and differentiates into distinct phenotypes[10]. This finding once again supports the concept that $CD4^+$ T cells can exhibit lineage plasticity during an active immune response[65]. Moreover, this study shed light on the ambiguous role of $FOXP3^+$ T cells in tumor immunity, especially in conjunction with IL-17 expression, which has been associated with poor prognosis[86].

Subsequent studies have expanded the throughput of paired analysis by combining targeted TCR sequencing with droplet-based scRNA-seq methods that profile the whole transcriptome. For example, the InDrop method for 3′ scRNA-seq was modified by using part of the amplified RNA library to target TCR transcripts with primers specific for the *TRAV* and *TRBV* genes before fragmentation (Fig. 3a, red line)[70,78]. By applying this technique to regulatory T cells isolated from the mouse intestine, Zemmour et al. showed that cells expressing the same TCR are transcriptionally similar[70], suggesting that TCR signaling can shape the phenotype of a T cell, an observation that has since been recapitulated by other studies[11,12].

A similar method developed by Tu et al. enriches for TCR transcripts within cDNA libraries by using biotinylated oligonucleotide

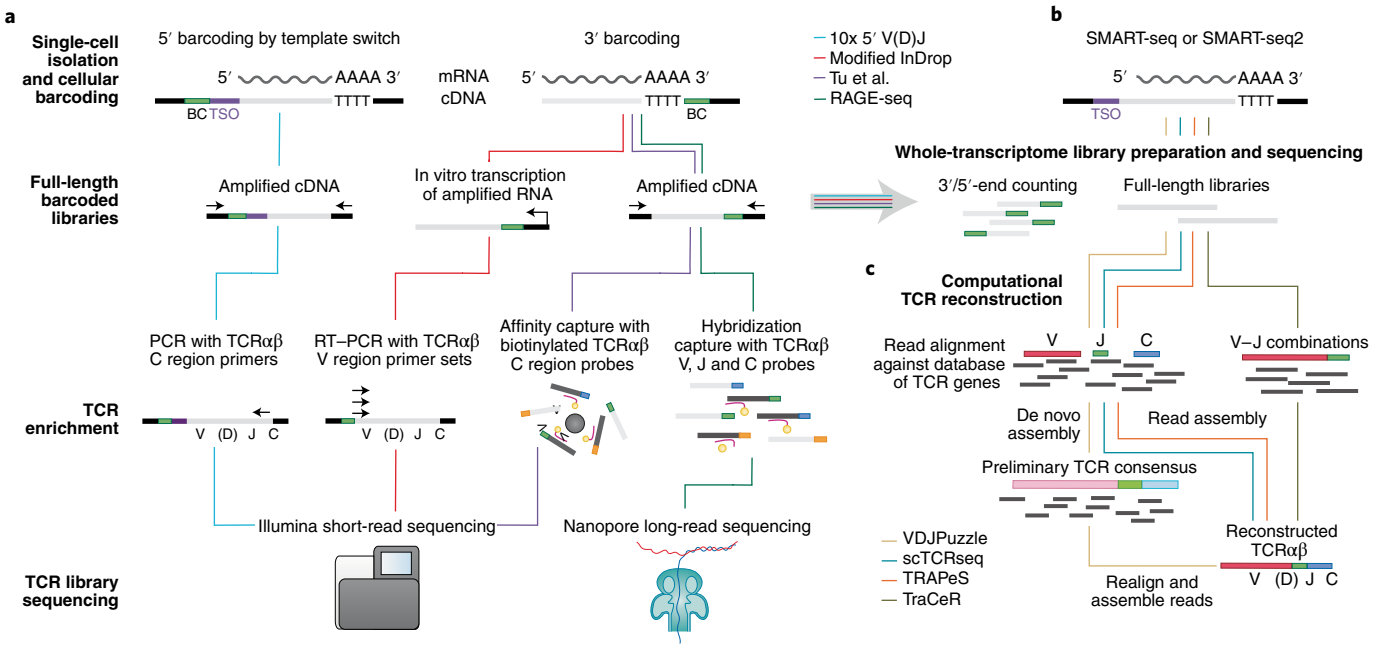

Fig. 3 | Schematic of paired scRNA-seq and TCR sequencing methods. a, Whole-transcriptome cDNA single-cell libraries barcoded at either the 3′ or 5′ end are subjected to gene expression profiling and targeted TCR capture with primers or hybridization probes. TCR and gene expression reads originating from the same cell can be linked on the basis of cellular barcodes. BC, barcode. **b**, Full-length whole-transcriptome cDNA libraries are generated from single cells by the SMART-seq or SMART-seq2 approach and sequenced. **c**, Computational methods for reconstructing TCR sequences from scRNA-seq libraries without targeted TCR amplification involve the alignment of reads to TCR gene reference databases or a 'recombinome' of possible V–J pairings. In **a**–**c**, colored lines indicate the different methods described.

probes mapping to the C regions (Fig. 3a, purple line)[71]. Importantly, this approach is compatible with standard 3′-barcoded platforms, including Seq-Well and Drop-seq[60,77], thereby bypassing the need for specific kits, which are necessary for InDrop[78] (Table 1). Another approach for performing targeted TCR capture, known as repertoire and gene expression by sequencing (RAGE-seq), uses a hybridization bait library specific for V and J genes followed by long-read Oxford nanopore sequencing to capture full-length TCR transcripts while the remaining cDNA library is subjected to standard short-read Illumina sequencing for gene expression profiling (Fig. 3a, green line)[72]. Although this method has the advantage of capturing full-length TCR sequences, relative inefficiencies in the TCR recovery rate due to error-prone nanopore sequencing and the requirement for multiple sequencing technologies have perhaps prevented the widespread adoption of this method.

Finally, simultaneous scRNA-seq and TCR sequencing has also been implemented by using 5′ RNA amplification[73]. In this approach, cells are encapsulated within droplets containing barcoded beads in a similar fashion to 3′ scRNA-seq methods; however, instead of using cellular barcodes adjacent to the poly(dT) sequence[61,76–78], the barcode in this method is adjacent to a TSO that is appended to the 5′ end of the cDNA after template extension (Fig. 3a, blue line). Full-length cDNA is then pooled and amplified with PCR primers targeting common sequences added during reverse transcription. An aliquot of the cDNA library is used for targeted enrichment of TCR transcripts via PCR with primers mapping to the TCR constant regions followed by library construction, while the remaining material is directly taken for construction of the gene expression library. This 5′ capture method has been made commercially available by 10x Genomics and enabled several key insights into T cell responses, particularly in cancer. In one of the first applications of this method, Azizi et al. analyzed the TIL compartment in human breast cancer tumors to examine the relationship between TCR repertoire diversity and T cell phenotypic diversity[11]. Integration of cell state and TCR clonotype data revealed a continuous trajectory of T cell activation that could partially be explained by TCR diversity. Furthermore, the authors observed that T cell clones are largely restricted to subsets of phenotypically related clusters that express similar environmental signatures such as anergy and gluconeogenesis, suggesting that cell state is collectively driven by TCR stimulation and microenvironmental cues[11]. Yost et al. also used this method to track T cell clonal dynamics before and after anti-PD-1 therapy in site-matched tumors from patients with basal cell carcinoma[12]. This study revealed that the post-treatment expanded CD8$^+$ T cell population largely comprises exhausted clones that were previously absent from the tumor, suggesting that checkpoint blockade acts by recruiting new T cell clones rather than by reinvigorating existing exhausted clones within the tumor, as previously thought[12,87].

Pairing the TCR with surface protein. Simultaneous analysis of the TCR and transcriptional profile of single cells can provide meaningful insights into T cell states at high resolution. However, the phenotyping of T cells using protein expression measurements may complement this analysis and more clearly define T cell subsets, as certain cell type markers are poorly detected at the transcript level[88]. Furthermore, single-cell protein measurements may be better able to resolve (1) isoform usage, which is especially important for the identification of effector and memory T cell populations that express isoforms of the *PTPRC* gene, CD45RO and CD45RA, at varying levels[89], and (2) intracellular protein activity and modifications, such as the activity of critical T cell transcription factors. To this end, Mimitou et al. recently described a method called expanded CRISPR-compatible cellular indexing of transcriptomes and epitopes by sequencing (ECCITE-seq) that extends the 5′ scRNA-seq and TCR sequencing method to capture cell-surface

Table 2 | Single-cell computational TCR reconstruction methods

Method	Refs.	TCR alignment reference	Minimum input read length (bp)[a]	Input data format	Sequencing mode	Language	Code availability	Other software dependencies
scTCRseq	64	IMGT V and C genes	50	FASTQ	Paired or single end	Python and Perl	https://github.com/ElementoLab/scTCRseq	SEQTK, Blastall, GapFiller and Vidjil
TraCeR	65	V–J recombinomes	50	FASTQ	Paired or single end	Python	https://github.com/Teichlab/tracer	Bowtie2, Trinity, IgBLAST and kallisto/salmon
VDJPuzzle	66,67	Reference genome V, D, J and C genes	150	FASTQ or BAM	Paired end	Bash, Python and Perl	https://bitbucket.org/kirbyvisp/vdjpuzzle	TopHat2 or STAR, Trinity, IgBLAST and MiGMAP
TRAPeS	68	Reference genome V, J and C genes	25	BAM	Paired end	Python and C++	https://github.com/YosefLab/TRAPeS	Bowtie2 and RSEM

[a]For a recovery rate greater than 90%, as demonstrated by the original paper.

proteins in parallel with gene expression and TCR profiles[75]. This is achieved by detecting proteins with DNA-barcoded antibodies conjugated to oligonucleotides complementary to the TSO. During reverse transcription, the cellular barcode is associated with the antibody-derived tags in addition to cDNA templates within the droplet. After amplification and sequencing, these antibody-derived tags can be read out and used to couple protein expression with TCR clonotype and RNA expression in each cell[75,88]. Notably, this method can also be adapted to assess TCR antigen specificity by using oligonucleotide-conjugated pMHC multimers, which we discuss below.

ECCITE-seq can also be extended to directly capture single guide RNAs (sgRNA) from pooled CRISPR screens in parallel with RNA, protein and TCR profiling[75]. By using an additional reverse transcription primer that anneals to the invariant 3′ region of sgRNAs, the sgRNA sequence can be captured and linked to the cellular barcode via a template-switching mechanism analogous to that used for the RNA fraction. Proof-of-concept experiments have demonstrated the ability for ECCITE-seq to capture changes at both the mRNA and protein level resulting from sgRNA-directed genetic perturbations[75]. This method presents a promising approach by which future work may investigate how CRISPR-mediated perturbations modulate T cell clonality and phenotype.

Pairing the TCR with chromatin accessibility. An orthogonal method for analyzing TCR sequences in their cellular context combines single-cell TCR capture with assay for transposase-accessible chromatin with sequencing (ATAC-seq), a technique that measures genome-wide chromatin accessibility[90,91] (Fig. 1c). Termed transcript-indexed ATAC-seq (T-ATAC-seq)[74], this method involves the microfluidic isolation of single cells followed by tagmentation of genomic DNA with Tn5, a transposase that preferentially inserts into open chromatin sites. TCR transcripts are then reverse transcribed by using primers specific for the C region and amplified via multiplex PCR with inner primers targeting the constant and variable regions, while the ATAC-seq fragments undergo 5′ extension and PCR amplification. This approach enables the interrogation of how epigenetic factors, including *cis*-acting DNA elements and *trans*-acting transcription factors, may drive the specificity and expansion of T cell clones. For example, this method was used to identify an epigenomic signature of malignant clonal T cells in lymphoma that was previously undetected by standard FACS-based approaches and may potentially be applied to improve current diagnostic abilities to distinguish cancerous from benign T cell proliferation[74] and inform clinical therapies that target the epigenome[92].

Strategies for mapping TCR sequence to antigen specificity

Linking antigen specificity to measurements of T cell clonality and phenotype is critical for understanding the drivers of the T cell response (Fig. 1d)[93–99]. Several aspects of T cell antigen recognition make mapping TCR sequence to specificity a challenging problem. First, the frequency of rare populations of antigen-specific T cells may be as low as one in a million cells[94], which makes detection of these cells difficult. Second, high levels of variability due to the polymorphic nature of the MHC, the polyspecificity of both the TCR and pMHC, and the wide variety of potential epitopes encoded by a single antigen substantially add to the complexity of resolving TCR antigen specificities[93,100]. Lastly, the relatively weak affinity and avidity of most TCR–pMHC interactions render selective isolation of antigen-specific T cell populations a nontrivial task[93,98].

While many approaches for assessing T cell antigen specificity exist, we focus our discussion here on strategies that also allow for TCR sequencing in parallel and refer interested readers to a recent review on T cell epitope discovery[93]. A classical approach for studying antigen-specific T cells involves the use of fluorescently labeled pMHC multimers to isolate T cells that recognize the specific peptide in the context of a given MHC molecule via flow cytometry[101,102]. The antigen-specific T cell populations can then be subjected to any of the methods discussed above for TCR sequencing or paired single-cell TCR and phenotype analysis. Although this approach is limited in throughput, several developments have increased the number of specificities that can be concurrently assessed. Recent advances in DNA and magnetic nanoparticle barcoding have increased the throughput of multimeric TCR specificity assays, allowing for the concurrent detection of more than 1,000 peptide specificities[95–97,103]. While many of these methods have only been used to identify the antigen specificities of T cells, several recent developments enable capturing the TCR as well, allowing for the characterization of clonality and TCR sequence variability among antigen-specific T cell populations. One such method is tetramer-associated TCR sequencing (TetTCR-seq), in which T cells are stained with fluorescently labeled DNA-barcoded pMHC tetramers and sorted into single-cell wells by FACS[96] (Fig. 4a). Following simultaneous RT–PCR amplification of TCRαβ transcripts and DNA barcodes within each well, cellular barcodes can be used to link TCR sequences with their cognate antigens at the single-cell level. Zhang et al. used this method to screen over 150 cancer neoantigen and wild-type peptide pairs concurrently, demonstrating the utility of TetTCR-seq in identifying functionally reactive neoantigen-specific TCRs in a high-throughput manner[96]. This approach can also be applied to the systematic interrogation of T cell specificity in the contexts of infection, autoimmunity and

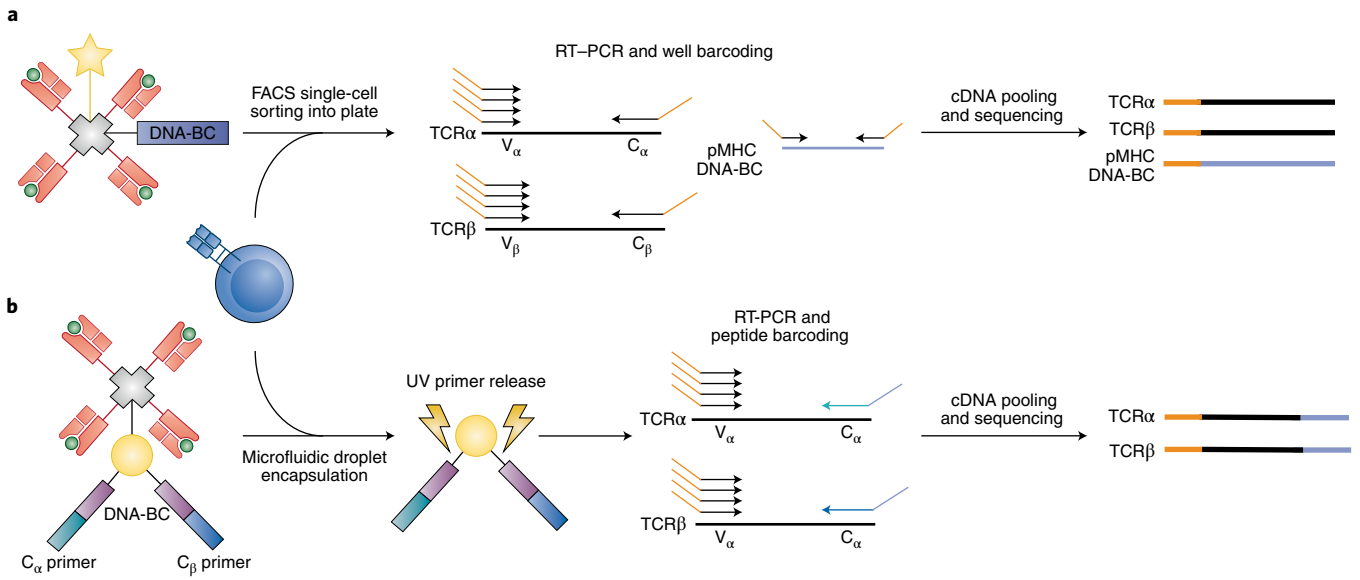

Fig. 4 | Overview of single-cell methods for linking TCR sequence to antigen specificity. a,**b**, pMHC tetramers conjugated to a fluorophore (**a**) or magnetic nanoparticle (**b**) and a unique DNA barcode (DNA-BC) are incubated with a T cell population and single-cell sorted into wells (**a**) or encapsulated in droplets (**b**). In **a**, TCR transcripts and the pMHC DNA barcode (purple) are simultaneously captured and attached to well barcodes, allowing TCR sequences to be matched to pMHC specificity. In **b**, TCR transcripts are captured by cell-barcoded V region and C region primers containing the DNA barcode. Final sequenced TCR products contain both the cell barcode and pMHC DNA barcode, enabling pairing of TCRαβ sequences to antigen specificity.

other immune-mediated diseases and may be pertinent for informing adoptive T cell therapy strategies.

Microfluidic techniques that further increase throughput have also been developed to capture antigen specificity and TCR sequence simultaneously. One method known as microfluidic antigen–TCR engagement sequencing (MATE-seq) uses magnetic nanoparticle-barcoded pMHC tetramers linked to photocleavable TCR-specific primers to capture both the TCR sequence and antigen identity within droplets[98] (Fig. 4b). Specifically, T cells are incubated with a library of nanoparticle-barcoded pMHCs and purified by magnetic separation[97]. After single cells are encapsulated into droplets and lysed, the nanoparticle-barcoded pMHCs are exposed to UV light, releasing RT–PCR primers targeting TCRαβ C regions. Because these primers are linked to a DNA barcode corresponding to the pMHC, the TCR sequence and antigen specificity can be coupled at the single-cell level even after pooling and sequencing. While this method is currently limited to relatively few pMHC tetramers owing to limitations in library construction (Table 1), future iterations may enable the analysis of T cell populations with a larger array of antigen specificities. Although these advances have greatly improved the throughput and resolution for mapping TCR sequence to specificity, these methods are only applicable if the objective is to characterize T cells with a known antigen specificity and the peptide(s) of interest can be nominated a priori, which in some cases may not be possible.

Similarly to the identification of TCR sequences that bind specific antigens, the converse problem of identifying antigen(s) that are recognized by a particular TCR is also challenging. One approach for linking antigen specificity to TCR sequences that is still relatively nascent is computational epitope prediction[104–107]. This approach involves training a supervised machine learning model on a collection of known TCR–antigen pairs and using this classifier to predict the antigen specificities of new TCR sequences. Although some models have been able to distinguish TCR specificity for a small number of epitopes[106,107], these algorithms are as yet unable to achieve high prediction accuracy on full TCR repertoires, as models must be trained separately for each epitope or set of epitopes. Improvements to these methods will require substantially larger training sets of experimentally verified TCR–pMHC pairs than are currently available through in-house and public databases such as McPAS-TCR, VDJdb and the Immune Epitope Database (IEDB)[108–111]. This could be facilitated by the continued collection of TCR–pMHC data through the strategies for concurrently capturing TCR sequence and peptide specificity described above[94,96,98], along with studies using TCR-directed methods for epitope discovery[93]. Additionally, algorithmic advancements that better model the three-dimensional complexity and flexibility of the TCR–pMHC interaction will be instrumental in improving computational epitope prediction approaches. However, even with more sophisticated algorithms that enable improved prediction of TCR epitopes, the issue of identifying T cell clones that are physiologically relevant to a given antigen-specific response is complicated by TCR polyspecificity. This cross-reactivity, whereby one TCR is able to recognize multiple unrelated peptides[112,113], makes precise linking of specificity to in vivo biological function difficult. Given these challenges, it is likely that, while these machine learning algorithms may prove helpful in nominating TCRs of interest for further investigation, they may require coupling with additional assays to provide a complete solution for TCR specificity determination.

Standards for TCR sequencing data collection

Importantly, the recent widespread adoption of the high-throughput TCR discovery methods described in this Review has led to the accumulation of large amounts of TCR sequencing data. Thus, there is an impetus for the standardization of TCR sequencing data to facilitate data reproducibility and sharing. To this end, the Adaptive Immune Receptor Repertoire (AIRR) Community has defined the Minimal Information about Adaptive Immune Receptor Repertoire (MiAIRR) standard, a set of guidelines for reporting metadata from antibody and TCR sequencing studies, ranging from details on study design and sample processing to data processing, annotation and submission to NCBI repositories[114,115]. The establishment

of a standardized schema and file format for representing annotated receptor rearrangements has also enabled streamlined access to data from repositories supporting this format[116], which are collectively organized into the AIRR Data Commons[117]. Continued efforts to define data standards and encourage adherence to these guidelines will be instrumental in promoting an ever-improving collaborative ecosystem among the research community that fosters progress through the use of TCR sequencing techniques.

Future perspectives

Although these advances in TCR sequencing have propelled the T cell field forward rapidly, there is a need for improved efficiency, sensitivity and cost, particularly for single-cell methods. Current droplet-based sequencing approaches have singlet capture rates of approximately 65% (ref. [118]). This low efficiency limits the interpretability of studies, as the cells captured in an assay may not be representative of the starting population, especially when analyzing large samples such as peripheral blood and other immunologically relevant organs. Another limitation of current droplet-based single-cell technologies is the minimum input requirement of approximately 1,000 cells per capture[119], which precludes the analysis of rare subsets, such as antigen-specific T cell populations. Additionally, while microfluidic single-cell methods are highly parallelized and carried out in nanoliter reaction volumes, leading to reduced cost in comparison to traditional plate-based methods[119], the cost to profile the TCR repertoire using single-cell methods is currently still higher than with bulk methods and may be limiting in some settings[59]. Future work that improves on these challenges will therefore be crucial for generating more faithful representations of the T cell compartment. For instance, the adaptation of microwell array techniques may provide a way forward in addressing these challenges to TCR sequencing methods, as they are compatible with lower cell inputs and more efficient than droplet-based methods[60–62].

Beyond these technical improvements, there is opportunity for combining the technologies discussed here with orthogonal methods to achieve an even finer understanding of how T cell clonality and specificity coordinately influence physiological function. One area with potential for innovation is in resolving T cell specificity and phenotype in situ. While existing single-cell methods can link the TCR repertoire with phenotype, the spatial organization of T cells within a tissue is lost. Recently developed spatial transcriptomic techniques enable high-throughput gene expression profiling while maintaining two-dimensional positional information[120,121]. Progress on these methods could enable spatially resolved TCR and phenotypic pairing. The ability to study the clonality and diversity of TCRs in their spatial context will be instrumental in dissecting how the anatomical distribution of T cells and their cognate antigens shapes immune responses.

Finally, recent advances in single-cell genetic lineage tracing methods that make use of DNA barcodes to label cells and their progeny have proven to be tremendously valuable in informing paradigms of hematopoiesis and a plethora of other biological processes[122–124]. Combining lineage tracing with TCR profiling has the potential to refine understanding of T cell developmental trajectories and phenotypic plasticity, specifically in regard to intraclonal phenotypic and functional heterogeneity. Previous efforts to track T cells using fate mapping have relied on mice with transgenic TCRs, which revealed important features of the clonal T cell response, such as the ability for a T cell to give rise to progeny exhibiting heterogeneous patterns of differentiation and growth kinetics[125,126]. However, these studies did not allow for analysis of the contribution of TCR clonality and specificity in shaping T cell responses. Integration of TCR sequencing and transcriptomic profiling with genetic tracing strategies could open the door for tracking T cell clones in vivo at an unprecedented resolution, enabling the reconstruction of lineage and phenotypic hierarchies within clones. While these approaches may require several iterations before they are available for use in physiological contexts such as human tissues, innovations such as these will nonetheless contribute to a better understanding of how T cell responses are coordinated, which can then be harnessed to develop improved therapies for cancer, infection and other immune-mediated diseases.

Received: 8 June 2020; Accepted: 7 June 2021;
Published online: 19 July 2021

References

1. Davis, M. M. & Bjorkman, P. J. The T cell receptor genes and T cell recognition. *Nature* **334**, 395–402 (1988).
2. Nikolich-Žugich, J., Slifka, M. K. & Messaoudi, I. The many important facets of T-cell repertoire diversity. *Nat. Rev. Immunol.* **4**, 123–132 (2004).
3. Bradley, P. & Thomas, P. G. Using T cell receptor repertoires to understand the principles of adaptive immune recognition. *Annu. Rev. Immunol.* **37**, 547–570 (2019).
4. Arstila, T. P. et al. A direct estimate of the human αβ T cell receptor diversity. *Science* **286**, 958–961 (1999).
5. Qi, Q. et al. Diversity and clonal selection in the human T-cell repertoire. *Proc. Natl Acad. Sci. USA* **111**, 13139–13144 (2014).
6. Tanno, H. et al. Determinants governing T cell receptor α/β-chain pairing in repertoire formation of identical twins. *Proc. Natl Acad. Sci. USA* **117**, 532–540 (2020).
7. La Gruta, N. L., Gras, S., Daley, S. R., Thomas, P. G. & Rossjohn, J. Understanding the drivers of MHC restriction of T cell receptors. *Nat. Rev. Immunol.* **18**, 467–478 (2018).
8. Cha, E. et al. Improved survival with T cell clonotype stability after anti-CTLA-4 treatment in cancer patients. *Sci. Transl. Med.* **6**, 238ra70 (2014).
9. Riaz, N. et al. Tumor and microenvironment evolution during immunotherapy with nivolumab. *Cell* **171**, 934–949 (2017).
10. Han, A., Glanville, J., Hansmann, L. & Davis, M. M. Linking T-cell receptor sequence to functional phenotype at the single-cell level. *Nat. Biotechnol.* **32**, 684–692 (2014).
This study describes the capture of paired TCRαβ and targeted gene expression in single cells.
11. Azizi, E. et al. Single-cell map of diverse immune phenotypes in the breast tumor microenvironment. *Cell* **174**, 1293–1308 (2018).
12. Yost, K. E. et al. Clonal replacement of tumor-specific T cells following PD-1 blockade. *Nat. Med.* **25**, 1251–1259 (2019).
13. Boudinot, P. et al. New perspectives for large-scale repertoire analysis of immune receptors. *Mol. Immunol.* **45**, 2437–2445 (2008).
14. Benichou, J., Ben-Hamo, R., Louzoun, Y. & Efroni, S. Rep-Seq: uncovering the immunological repertoire through next-generation sequencing. *Immunology* **135**, 183–191 (2012).
15. Six, A. et al. The past, present, and future of immune repertoire biology—the rise of next-generation repertoire analysis. *Front. Immunol.* **4**, 413 (2013).
16. Cochet, M., Pannetier, C., Regnault, A., Leclercn, C. & Immunitairesa, R. Molecular detection and in vivo analysis of the specific T cell response to a protein antigen. *Eur. J. Immunol.* **22**, 2639–2647 (1992).
17. Pannetier, C. et al. The sizes of the CDR3 hypervariable regions of the murine T-cell receptor β chains vary as a function of the recombined germ-line segments. *Proc. Natl Acad. Sci. USA* **90**, 4319–4323 (1993).
18. Casrouge, A. et al. Size estimate of the αβ TCR repertoire of naive mouse splenocytes. *J. Immunol.* **164**, 5782–5787 (2000).
19. Diu, A. et al. Fine specificity of monoclonal antibodies directed at human T cell receptor variable regions: comparison with oligonucleotide-driven amplification for evaluation of V_β expression. *Eur. J. Immunol.* **23**, 1422–1429 (1993).
20. Faint, J. M. et al. Quantitative flow cytometry for the analysis of T cell receptor V_β chain expression. *J. Immunol. Methods* **225**, 53–60 (1999).
21. Fozza, C. et al. Study of the T-cell receptor repertoire by CDR3 spectratyping. *J. Immunol. Methods* **440**, 1–11 (2017).
22. Robins, H. S. et al. Comprehensive assessment of T-cell receptor β-chain diversity in αβ T cells. *Blood* **114**, 4099–4107 (2009).
23. Robins, H. S. et al. Overlap and effective size of the human $CD8^+$ T cell receptor repertoire. *Sci. Transl. Med.* **2**, 47 (2010).
This study, along with Robins et al.[22] and Wang et al.[24], describes pioneering high-throughput TCR sequencing techniques and their use in interrogating TCR repertoire diversity.
24. Wang, C. et al. High throughput sequencing reveals a complex pattern of dynamic interrelationships among human T cell subsets. *Proc. Natl Acad. Sci. USA* **107**, 1518–1523 (2010).

25. Picelli, S. et al. Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9**, 171–181 (2014).
26. Picelli, S. et al. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* **10**, 1096–1100 (2013).
27. Anonymous. Rapid amplification of 5′ complementary DNA ends (5′ RACE). *Nat. Methods* **2**, 629–630 (2005).
28. Freeman, J. D., Warren, R. L., Webb, J. R., Nelson, B. H. & Holt, R. A. Profiling the T-cell receptor β-chain repertoire by massively parallel sequencing. *Genome Res.* **19**, 1817–1824 (2009).
29. Mamedov, I. Z. et al. Quantitative tracking of T cell clones after haematopoietic stem cell transplantation. *EMBO Mol. Med.* **3**, 201–207 (2011).
30. Murugan, A., Mora, T., Walczak, A. M. & Callan, C. G. Statistical inference of the generation probability of T-cell receptors from sequence repertoires. *Proc. Natl Acad. Sci. USA* **109**, 16161–16166 (2012).
31. Nguyen, P. et al. Identification of errors introduced during high throughput sequencing of the T cell receptor repertoire. *BMC Genomics* **12**, 106 (2011).
32. Shugay, M. et al. Towards error-free profiling of immune repertoires. *Nat. Methods* **11**, 653–655 (2014).
33. Liu, X. et al. Systematic comparative evaluation of methods for investigating the TCRβ repertoire. *PLoS ONE* **11**, e0152464 (2016).
34. Woodsworth, D. J., Castellarin, M. & Holt, R. A. Sequence analysis of T-cell repertoires in health and disease. *Genome Med.* **5**, 98 (2013).
35. De Simone, M., Rossetti, G. & Pagani, M. Single cell T cell receptor sequencing: techniques and future challenges. *Front. Immunol.* **9**, 1638 (2018).
36. Carlson, C. S. et al. Using synthetic templates to design an unbiased multiplex PCR assay. *Nat. Commun.* **4**, 2680 (2013).
37. Zhu, Y. Y., Machleder, E. M., Chenchik, A., Li, R. & Siebert, P. D. Reverse transcriptase template switching: a SMART™ approach for full-length cDNA library construction. *Biotechniques* **30**, 892–897 (2001).
38. Kivioja, T. et al. Counting absolute numbers of molecules using unique molecular identifiers. *Nat. Methods* **9**, 72–74 (2012).
39. Egorov, E. S. et al. Quantitative profiling of immune repertoires for minor lymphocyte counts using unique molecular identifiers. *J. Immunol.* **194**, 6155–6163 (2015).
40. Barennes, P. et al. Benchmarking of T cell receptor repertoire profiling methods reveals large systematic biases. *Nat. Biotechnol.* **39**, 236–245 (2020).
This study systematically compares a collection of multiplex PCR and 5′ RACE TCR sequencing protocols and provides practical guidelines for choosing an appropriate method on the basis of observations regarding technical biases between methods.
41. Bolotin, D. A. et al. Next generation sequencing for TCR repertoire profiling: platform-specific features and correction algorithms. *Eur. J. Immunol.* **42**, 3073–3083 (2012).
42. Bolotin, D. A. et al. MiTCR: software for T-cell receptor sequencing data analysis. *Nat. Methods* **10**, 813–814 (2013).
43. Bolotin, D. A. et al. MiXCR: software for comprehensive adaptive immunity profiling. *Nat. Methods* **12**, 380–381 (2015).
44. Bolotin, D. A. et al. Antigen receptor repertoire profiling from RNA-seq data. *Nat. Biotechnol.* **35**, 908–911 (2017).
45. Dupic, T., Marcou, Q., Walczak, A. M. & Mora, T. Genesis of the αβ T-cell receptor. *PLoS Comput. Biol.* **15**, e1006874 (2019).
46. Howie, B. et al. High-throughput pairing of T cell receptor α and β sequences. *Sci. Transl. Med.* **7**, 301ra131 (2015).
47. Lee, E. S., Thomas, P. G., Mold, J. E. & Yates, A. J. Identifying T cell receptors from high-throughput sequencing: dealing with promiscuity in TCRα and TCRβ pairing. *PLoS Comput. Biol.* **13**, e1005313 (2017).
48. Turchaninova, M. A. et al. Pairing of T-cell receptor chains via emulsion PCR. *Eur. J. Immunol.* **43**, 2507–2515 (2013).
49. Munson, D. J. et al. Identification of shared TCR sequences from T cells in human breast cancer using emulsion RT–PCR. *Proc. Natl Acad. Sci. USA* **113**, 8272–8277 (2016).
50. DeKosky, B. J. et al. In-depth determination and analysis of the human paired heavy- and light-chain antibody repertoire. *Nat. Med.* **21**, 86–91 (2015).
51. Mcdaniel, J. R., DeKosky, B. J., Tanno, H., Ellington, A. D. & Georgiou, G. Ultra-high-throughput sequencing of the immune receptor repertoire from millions of lymphocytes. *Nat. Protoc.* **11**, 429–442 (2016).
52. Tanno, H. et al. A facile technology for the high-throughput sequencing of the paired V_H:V_L and TCRβ:TCRα repertoires. *Sci. Adv.* **6**, 17 (2020).
53. Seitz, S. et al. Reconstitution of paired T cell receptor α- and β-chains from microdissected single cells of human inflammatory tissues. *Proc. Natl Acad. Sci. USA* **103**, 12057–12062 (2006).
54. Julius, M. H., Masuda, T. & Herzenberg, L. A. Demonstration that antigen-binding cells are precursors of antibody-producing cells after purification with a fluorescence-activated cell sorter. *Proc. Natl Acad. Sci. USA* **69**, 1934–1938 (1972).
55. Rinke, C. et al. Obtaining genomes from uncultivated environmental microorganisms using FACS-based single-cell genomics. *Nat. Protoc.* **9**, 1038–1048 (2014).
56. Dash, P. et al. Paired analysis of TCRα and TCRβ chains at the single-cell level in mice. *J. Clin. Invest.* **121**, 288–295 (2011).
57. Wang, G. C., Dash, P., McCullers, J. A., Doherty, P. C. & Thomas, P. G. T cell receptor αβ diversity inversely correlates with pathogen-specific antibody levels in human cytomegalovirus infection. *Sci. Transl. Med.* **4**, 128ra42 (2012).
58. Dash, P., Wang, G. C. & Thomas, P. G. Single-cell analysis of T-cell receptor αβ repertoire. *Methods Mol. Biol.* **1343**, 181–197 (2015).
59. Prakadan, S. M., Shalek, A. K. & Weitz, D. A. Scaling by shrinking: empowering single-cell 'omics' with microfluidic devices. *Nat. Rev. Genet.* **18**, 345–361 (2017).
60. Gierahn, T. M. et al. Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. *Nat. Methods* **14**, 395–398 (2017).
61. Han, X. et al. Mapping the mouse cell atlas by microwell-seq. *Cell* **172**, 1091–1107 (2018).
62. Yuan, J. & Sims, P. A. An automated microwell platform for large-scale single cell RNA-seq. *Sci. Rep.* **6**, 33883 (2016).
63. Mazutis, L. et al. Single-cell analysis and sorting using droplet-based microfluidics. *Nat. Protoc.* **8**, 870–891 (2013).
64. Redmond, D., Poran, A. & Elemento, O. Single-cell TCRseq: paired recovery of entire T-cell α and β chain transcripts in T-cell receptors from single-cell RNAseq. *Genome Med.* **8**, 80 (2016).
65. Stubbington, M. J. T. et al. T cell fate and clonality inference from single-cell transcriptomes. *Nat. Methods* **13**, 329–332 (2016).
This work describes and uses a key TCR reconstruction algorithm, TraCeR, to trace CD4⁺ T cell differentiation trajectories in *Salmonella* infection.
66. Eltahla, A. A. et al. Linking the T cell receptor to the single cell transcriptome in antigen-specific human T cells. *Immunol. Cell Biol.* **94**, 604–611 (2016).
67. Rizzetto, S. et al. B-cell receptor reconstruction from single-cell RNA-seq with VDJPuzzle. *Bioinformatics* **34**, 2846–2847 (2018).
68. Afik, S. 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.* **45**, e148 (2017).
69. Chen, S.-Y., Liu, C.-J., Zhang, Q. & Guo, A.-Y. An ultrasensitive T-cell receptor detection method for TCR-Seq and RNA-Seq data. *Bioinformatics* **36**, 4255–4262 (2020).
70. Zemmour, D. et al. Single-cell gene expression reveals a landscape of regulatory T cell phenotypes shaped by the TCR. *Nat. Immunol.* **19**, 291–301 (2018).
71. Tu, A. A. et al. TCR sequencing paired with massively parallel 3′ RNA-seq reveals clonotypic T cell signatures. *Nat. Immunol.* **20**, 1692–1699 (2019).
72. Singh, M. et al. High-throughput targeted long-read single cell sequencing reveals the clonal and transcriptional landscape of lymphocytes. *Nat. Commun.* **10**, 3120 (2019).
73. Neal, J. T. et al. Organoid modeling of the tumor immune microenvironment. *Cell* **175**, 1972–1988 (2018).
This study describes the use of 5′ single-cell droplet microfluidics to simultaneously capture TCR and gene expression.
74. Satpathy, A. T. et al. Transcript-indexed ATAC-seq for precision immune profiling. *Nat. Med.* **24**, 580–590 (2018).
75. Mimitou, E. P. et al. Multiplexed detection of proteins, transcriptomes, clonotypes and CRISPR perturbations in single cells. *Nat. Methods* **16**, 409–412 (2019).
76. Zheng, G. X. Y. et al. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
77. Macosko, E. Z. et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
78. Klein, A. M. et al. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187–1201 (2015).
79. Ramsköld, D. et al. Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat. Biotechnol.* **30**, 777–782 (2012).
80. Durruthy-Durruthy, R. & Ray, M. Using Fluidigm C1 to generate single-cell full-length cDNA libraries for mRNA sequencing. *Methods Mol. Biol.* **1706**, 199–221 (2018).
81. Manjunath, N. et al. Effector differentiation is not prerequisite for generation of memory cytotoxic T lymphocytes. *J. Clin. Invest.* **108**, 871–878 (2001).
82. Román, E. et al. CD4 effector T cell subsets in the response to influenza: heterogeneity, migration, and function. *J. Exp. Med.* **196**, 957–968 (2002).
83. Lanzavecchia, A. & Sallusto, F. Progressive differentiation and selection of the fittest in the immune response. *Nat. Rev. Immunol.* **2**, 982–987 (2002).
84. Ranasinghe, S. et al. Antiviral CD8⁺ T cells restricted by human leukocyte antigen class II exist during natural HIV infection and exhibit clonal expansion. *Immunity* **45**, 917–930 (2016).

85. Rizzetto, S. et al. Impact of sequencing depth and read length on single cell RNA sequencing data of T cells. *Sci. Rep.* **7**, 12781 (2017).
86. Gounaris, E. et al. T-regulatory cells shift from a protective anti-inflammatory to a cancer-promoting proinflammatory phenotype in polyposis. *Cancer Res.* **69**, 5490–5497 (2009).
87. Yost, K. E., Chang, H. Y. & Satpathy, A. T. Recruiting T cells in cancer immunotherapy. *Science* **372**, 130–131 (2021).
88. Stoeckius, M. et al. Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* **14**, 865–868 (2017).
89. Golubovskaya, V. & Wu, L. Different subsets of T cells, memory, effector functions, and CAR-T immunotherapy. *Cancers* **8**, 36 (2016).
90. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
91. Buenrostro, J. D. et al. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486–490 (2015).
92. Bolden, J. E., Peart, M. J. & Johnstone, R. W. Anticancer activities of histone deacetylase inhibitors. *Nat. Rev. Drug Discov.* **5**, 769–784 (2006).
93. Joglekar, A. V. & Li, G. T cell antigen discovery. *Nat. Methods* https://doi.org/10.1038/s41592-020-0867-z (2020).
94. Klinger, M. et al. Multiplex identification of antigen-specific T cell receptors using a combination of immune assays and immune receptor sequencing. *PLoS ONE* **10**, e0141561 (2015).
95. Bentzen, A. K. et al. T cell receptor fingerprinting enables in-depth characterization of the interactions governing recognition of peptide–MHC complexes. *Nat. Biotechnol.* **36**, 1191–1196 (2018).
96. Zhang, S. Q. et al. High-throughput determination of the antigen specificities of T cell receptors in single cells. *Nat. Biotechnol.* **36**, 1156–1159 (2018).
This study, along with Ng et al.[98], describes the use of DNA-barcoded pMHC tetramers to link TCR sequence and antigen specificity in single cells.
97. Peng, S. et al. Sensitive detection and analysis of neoantigen-specific T cell populations from tumors and blood. *Cell Rep.* **28**, 2728–2738 (2019).
98. Ng, A. H. C. et al. MATE-Seq: microfluidic antigen–TCR engagement sequencing. *Lab Chip* **19**, 3011–3021 (2019).
99. Newell, E. W. et al. Combinatorial tetramer staining and mass cytometry analysis facilitate T-cell epitope mapping and characterization. *Nat. Biotechnol.* **31**, 623–629 (2013).
100. Newell, E. W. & Davis, M. M. Beyond model antigens: high-dimensional methods for the analysis of antigen-specific T cells. *Nat. Biotechnol.* **32**, 149–157 (2014).
101. Altman, J. D. et al. Phenotypic analysis of antigen-specific T lymphocytes. *Science* **274**, 94–96 (1996).
102. Davis, M. M., Altman, J. D. & Newell, E. W. Interrogating the repertoire: broadening the scope of peptide–MHC multimer analysis. *Nat. Rev. Immunol.* **11**, 551–558 (2011).
103. Bentzen, A. K. et al. Large-scale detection of antigen-specific T cells using peptide–MHC-I multimers labeled with DNA barcodes. *Nat. Biotechnol.* **34**, 1037–1045 (2016).
104. Schneidman-Duhovny, D. et al. Predicting CD4 T-cell epitopes based on antigen cleavage, MHCII presentation, and TCR recognition. *PLoS ONE* **13**, e0206654 (2018).
105. Gielis, S. et al. Detection of enriched T cell epitope specificity in full T cell receptor sequence repertoires. *Front. Immunol.* **10**, 2820 (2019).
106. De Neuter, N. et al. On the feasibility of mining $CD8^+$ T cell receptor patterns underlying immunogenic peptide recognition. *Immunogenetics* **70**, 159–168 (2018).
107. Fischer, D. S., Wu, Y., Schubert, B. & Theis, F. J. Predicting antigen specificity of single T cells based on TCR CDR3 regions. *Mol. Syst. Biol.* **16**, e9416 (2020).
108. Tickotsky, N., Sagiv, T., Prilusky, J., Shifrut, E. & Friedman, N. McPAS-TCR: a manually curated catalogue of pathology-associated T cell receptor sequences. *Bioinformatics* **33**, 2924–2929 (2017).
109. Shugay, M. et al. VDJdb: a curated database of T-cell receptor sequences with known antigen specificity. *Nucleic Acids Res.* **46**, D419–D427 (2018).
110. Bagaev, D. V. et al. VDJdb in 2019: database extension, new analysis infrastructure and a T-cell receptor motif compendium. *Nucleic Acids Res.* **48**, D1057–D1062 (2020).
111. Vita, R. et al. The Immune Epitope Database (IEDB): 2018 update. *Nucleic Acids Res.* **47**, D339–D343 (2019).
112. Mason, D. A very high level of crossreactivity is an essential feature of the T-cell receptor. *Immunol. Today* **19**, 395–404 (1998).
113. Sewell, A. K. Why must T cells be cross-reactive? *Nat. Rev. Immunol.* **12**, 669–677 (2012).
114. Breden, F. et al. Reproducibility and reuse of adaptive immune receptor repertoire data. *Front. Immunol.* **8**, 1418 (2017).
115. Rubelt, F. et al. Adaptive Immune Receptor Repertoire Community recommendations for sharing immune-repertoire sequencing data. *Nat. Immunol.* **18**, 1274–1278 (2017).
116. Vander Heiden, J. A. et al. AIRR community standardized representations for annotated immune repertoires. *Front. Immunol.* **9**, 2206 (2018).
117. Christley, S. et al. The ADC API: a web API for the programmatic query of the AIRR Data Commons. *Front. Big Data* **3**, 22 (2020).
118. Nguyen, A., Khoo, W. H., Moran, I., Croucher, P. I. & Phan, T. G. Single cell RNA sequencing of rare immune cell populations. *Front. Immunol.* **9**, 1553 (2018).
119. Kolodziejczyk, A. A., Kim, J. K., Svensson, V., Marioni, J. C. & Teichmann, S. A. The technology and biology of single-cell RNA sequencing. *Mol. Cell* **58**, 610–620 (2015).
120. Ståhl, P. L. et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78–82 (2016).
121. Rodriques, S. G. et al. Slide-seq: a scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463–1467 (2019).
122. Rodriguez-Fraticelli, A. E. et al. Clonal analysis of lineage fate in native haematopoiesis. *Nature* **553**, 212–216 (2018).
123. Kester, L. & van Oudenaarden, A. Single-cell transcriptomics meets lineage tracing. *Cell Stem Cell* **23**, 166–179 (2018).
124. Wagner, D. E. & Klein, A. M. Lineage tracing meets single-cell omics: opportunities and challenges. *Nat. Rev. Genet.* **21**, 410–427 (2020).
125. Gerlach, C. et al. Heterogeneous differentiation patterns of individual $CD8^+$ T cells. *Science* **340**, 635–639 (2013).
126. Buchholz, V. R. et al. Disparate individual fates compose robust $CD8^+$ T cell immunity. *Science* **340**, 630–635 (2013).

Acknowledgements

This work was supported by National Institutes of Health grants K08CA230188 (A.T.S.) and 5T32AI007290 (J.A.P.), the Parker Institute for Cancer Immunotherapy, a Technology Impact Award from the Cancer Research Institute and a Career Award for Medical Scientists from the Burroughs Wellcome Fund (A.T.S.).

Competing interests

A.T.S. is a founder of Immunai and Cartography Biosciences and receives research funding from 10x Genomics, Arsenal Biosciences and Allogene Therapeutics.

Additional information

Correspondence should be addressed to A.T.S.

Peer review information *Nature Methods* thanks Benny Chain, Encarnita Mariotti-Ferrandiz and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Madhura Mukhopadhyay was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.