Development and characterization of human T-cell receptor (TCR) alpha and beta clones' library as biological standards and resources for TCR sequencing and engineering

Yu-Chun Wei[1,†], Mateusz Pospiech[1,†], Yiting Meng[1], Houda Alachkar [1,2,*]

[1]Department of Clinical Pharmacy, USC Alfred E. Mann School of Pharmacy and Pharmaceutical Sciences, University of Southern California, Los Angeles, CA, 90089, United States
[2]USC Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, CA, 90089, United States

*Correspondence address. USC Alfred E. Mann School of Pharmacy and Pharmaceutical Sciences, University of Southern California, 1985 Zonal Avenue John Stauffer Pharmaceutical Sciences Center Room 608, Los Angeles, CA 90089, United States. Tel: 323-442-2696; E-mail: alachkar@usc.edu

†These authors contributed equally to the work.

Abstract

Characterization of T-cell receptors (TCRs) repertoire was revolutionized by next-generation sequencing technologies; however, standardization using biological controls to facilitate precision of current alignment and assembly tools remains a challenge. Additionally, availability of TCR libraries for off-the-shelf cloning and engineering TCR-specific T cells is a valuable resource for TCR-based immunotherapies. We established nine human TCR α and β clones that were evaluated using the 5′-rapid amplification of cDNA ends-like RNA-based TCR sequencing on the Illumina platform. TCR sequences were extracted and aligned using MiXCR, TRUST4, and CATT to validate their sensitivity and specificity and to validate library preparation methods. The correlation between actual and expected TCR ratios within libraries confirmed accuracy of the approach. Our findings established the development of biological standards and library of TCR clones to be leveraged in TCR sequencing and engineering. The remaining human TCR clones' libraries for a more diverse biological control will be generated.

Keywords: T-cell receptor (TCR); next-generation sequencing (NGS); biological standard and resources; computational tools; cloning, TCR engineering

Highlights

Here, we report a biological library of TCR clones that allows more accurate evaluation, optimization, and increased accuracy of current technical and computational methodologies. Our TCR library preparation method utilizes innovative features, such as plasmid controls to ensure consistency across experiments, is composed of specific *TRAV* and *TRBV* gene segments relevant to various cancer types to enhance its applicability, and generates paired-end reads crucial for obtaining full-length TCR sequences, ensuring accurate and comprehensive analysis.

Introduction

The interaction of the T-cell receptor (TCR) with antigen-major histocompatibility complex (MHC) molecules is essential for T-cell-mediated antigen recognition [1]. The human genome encodes two distinct heterodimers: TCRα/TCRβ (95–99.5%) and TCRγ/TCRδ (0.5–5%) [2–4]. The third complementarity-determining region (CDR3) resulting from the recombination of variable (V), joining (J), and constant (C) segments and an additional diversity (D) segment for TCRβ and TCRγ leads up to 10^{20} unique TCR sequences enabling the recognition of a wide range of tumor neoantigens [5–10]. Naïve T-cell activation, triggered by specific antigen-MHC binding, prompts clonal expansion while preserving TCR sequences for an effective immune response, including anti-tumor activity. After pathogen clearance, most expanded T cells undergo apoptosis, with some maturing into long-term memory T cells, ensuring swift responses upon antigen re-exposure [11–13]. The TCR repertoire mirrors an individual's antigen encounters, shaping their immune history.

Studies from the last few decades led to the discovery of a range of antigens that are recognized by patients' tumor-reactive T cells [14, 15]. Determination of antigen specificity can be deduced from TCR sequence using advanced computational tools to predict peptide bound to MHC (pMHC) specificity [16–20]. Identification of tumor-specific T cells with their specificity toward neoantigens resulted in the design of a new generation of cell therapies for several malignancies [21, 22].

Next-generation sequencing (NGS) technology is commonly employed to analyze the TCR repertoire with standard workflow

starting with DNA or RNA as an input following library preparation methods that include multiplex polymerase chain reaction (PCR), target enrichment, and 5′-rapid amplification of cDNA ends (RACE)-switch-oligonucleotide nested PCR with each method bringing own set of systematic biases [23, 24]. Several computational methods have been developed for the analysis of the TCR sequences [25–30]. While benchmarking studies have been conducted, the majority of these studies were based on simulations with *in silico* generated datasets, or synthetic sequences, or samples with controlled clonality by mixing cell lines with T cells without introducing the ground truth based on biological controls [31–34]. Because the immune repertoires are associated with significant intra- and inter-individual differences, biological controls are needed to provide sensitivity, accuracy, and reliable quantification levels required for disease-related associations [23, 24, 31]. The accuracy of current computational methods remains unknown due to the lack of a control framework for direct comparison.

The availability of TCR libraries for off-the-shelf cloning and the engineering of TCR-specific T cells represent a critical resource for the advancement of TCR-based immunotherapies. These libraries offer a diverse range of pre-characterized TCRs that can be rapidly cloned and transduced into T cells, facilitating the development of personalized treatments with the potential to target a wide array of antigens specific to different types of tumors.

Here we describe a method for generating human TCRα (TRA) and TCRβ (TRB) libraries developed based on TCR clones obtained from healthy donors. This library will serve as a positive control with known composition to quantitatively and qualitatively analyze TCR sequencing as well as off-the-shelf clones for future development of TCR-specific T cells.

Materials and methods
Workflow for the generation of human TCRs library

Here we report the workflow for the generation of a human TCR library to serve as a biological standard for TCR-targeted sequencing analyses to assess the composition and quantities of the TCR repertoire. Peripheral blood mononuclear cells (PBMCs) were isolated from healthy volunteers, followed by RNA extraction and cDNA synthesis for further PCR amplification. Second, primers were designed, targeting nine TRA V regions, and nine TRB V regions, along with the TRA and TRB constant regions for amplification and cloning into pCR2.1 TOPO vector. Third, the human TCR plasmids were segmented into three strategies to develop the most optimal method for generating control libraries of known composition both quantitatively and qualitatively, while maintaining the simplicity in the preparation. Strategy A involved conducting PCR individually on each plasmid to add the T7 promoter sequence at the 5′-end followed by pooling the PCR products together. Strategy B involved pooling all plasmids first and then subjecting the pool to PCR amplification to add the T7 promoter sequence at 5′-end. The next step of these two strategies included transcription using T7 RNA polymerase to transcribe DNA into RNA and polyadenylation to add a poly-A tail at the 3′-end for further library preparation. Strategy C involved conducting PCR individually on each plasmid to add the T7 promoter sequence at the 5′-end, followed by transcription using T7 RNA polymerase to transcribe DNA into RNA and polyadenylation to add a poly-A tail at the 3′-end on each TCR sequence separately, and then pooling resulting RNAs together. After the addition of the T7 sequence to TCR sequences, a few PCR products at random were sent for additional Sanger sequencing to ensure no PCR errors were introduced at this step. Lastly, library preparation for NGS was performed using SMARTer Human TCR a/b Profiling Kit (#635016, Takara, CA, USA), including first-strand cDNA synthesis, PCR1, PCR2, purification, and qPCR followed by sequencing utilizing MiSeq System (Fig. 1).

Each strategy included the preparation of four libraries with different ratios of the TCR sequences. The first strategy encompasses libraries A1–A4, the second strategy comprises libraries B1–B4, and the third strategy includes libraries C1–C4. The ratio of plasmids differs within each library as indicated in Table 1.

Library preparation strategies rationale

The rationale behind the three different library preparation strategies is to identify the optimal method for generating the control libraries. Strategy A assesses the feasibility of performing transcription on pooled PCR products, focusing on potential biases in transcription toward certain clones while excluding biases from PCR that add T7 promoter sequences. Strategy B involves using the plasmid pool, introducing possible biases from both PCR and transcription. Strategy C is designed to avoid all these possible biases in Strategy A and Strategy B by performing each step separately. The primary consideration in preparing libraries with different ratios between TCR clones was to evaluate whether the resulting ratios from the sequencing reads obtained would reflect the input ratios, aiming to investigate whether any clonotype would be disproportionately amplified based on their input during library preparation steps. Similarly, different ratios between TCR clones would exist within an actual sample obtained from a healthy donor or a patient; therefore, it is crucial to investigate the accuracy of quantification of clonotypes at both low and high abundance.

PBMCs isolation

PBMCs from ∼10 ml of whole blood samples of healthy donors that volunteered to participate in the study according to IRB protocol HS-16-00274 at the University of Southern California School of Pharmacy were isolated by density gradient centrifugation using Ficoll-Paque (#17144003, Cytiva Life Sciences, MA, USA) [35]. All subjects were informed of the present research project and signed written informed consent.

RNA extraction and cDNA synthesis

Total RNA was isolated from the patient's PBMCs using the RNeasy Mini Kit (#74104, Qiagen, Germany) according to the manufacturer's protocol. RNA sample quality was assessed by Thermo Scientific NanoDrop OneC. SuperScript™ IV First-Strand Synthesis cDNA was performed according to the manufacturer's protocol (#18091050, ThermoFisher, MA, USA).

Primers and PCR

Sequences of human TCR genes were acquired from the International Immunogenetics (IMGT) Information System [6]. A set of forward primers for the TRA and TRB V-region with the L-part of the TCR gene (L′-TRAV and L′-TRBV) was designed (Fig. 2). Reverse primer for TRAC was designed on chromosome 14 of GRCh38.p14 (22550645–22550661). Reverse primer for TRBC was designed to be located on chromosome 7 of GRCh38.p14 of human genome assembly (142793120–142793138). In addition, two leading nucleotides (-CT- or -AT-) and restriction enzymes-EcoRI (-GAATTC-) or BamHI (-GGATCC-) were added to each primer. To prepare TCR sequences for transcription, PCR using plasmid as a template to add the T7 promoter sequence at the
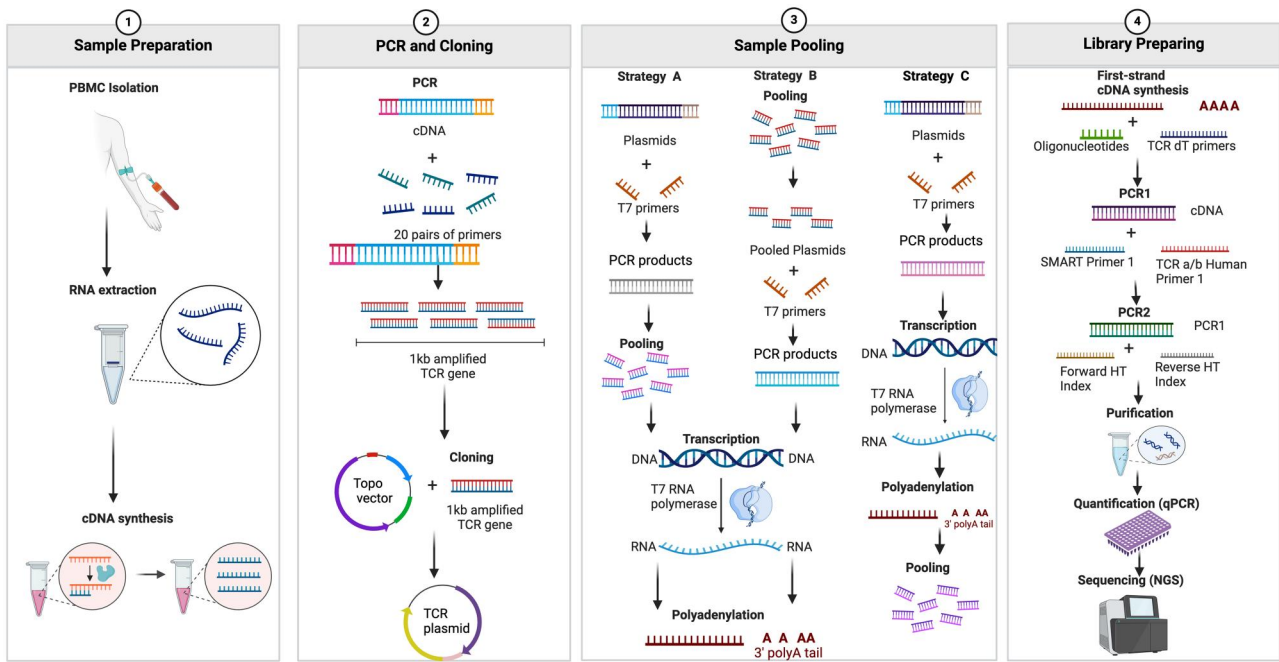
Figure 1 Overview of human TCR library generation.

Table 1. Plasmid ratios within libraries.

Plasmids		Ratio of TRA:TRB plasmids in each library			
TRA	**TRB**	**Library 1**	**Library 2**	**Library 3**	**Library 4**
L-TRAV5	L-TRBV29-1	1:1	16:16	1:1	1:2
L-TRAV10	L-TRBV28	1:1	16:16	2:2	1:2
L-TRAV12-2	L-TRBV27	1:1	8:8	2:2	1:2
L-TRAV13-1	L-TRBV20-1	1:1	8:8	4:4	1:2
L-TRAV17	L-TRBV12-4	1:1	4:4	4:4	1:2
L-TRAV20	L-TRBV7-2	1:1	4:4	8:8	1:2
L-TRAV21	L-TRBV5-1	1:1	2:2	8:8	1:2
L-TRAV26-1	L-TRBV4-1	1:1	2:2	16:16	1:2
L-TRAV29/DV5	L-TRBV2	1:1	1:1	16:16	1:2

Figure 2 Primers design. Forward primers were designed from V regions of TCR and reverse primers were designed from C regions of TCR. (A) TRA and (B) TRB.

5′-end was performed. The full list of primers used is available in Supplementary Table S1. Amplification was performed in a 25 µl reaction volume consisting of 12.5 µl Taq 2X Master Mix (#M0270L, New England BioLabs Inc., MA, USA), 1.25 µl of 10 µM forward primer, 1.25 µl of 10 µM reverse primer, 1 µl template, and 9 µl nuclease-free water. The PCR reactions consisted of an initial denaturation at 95°C for 2 min, followed by 35 cycles of denaturation at 95°C for 30 s, annealing at 62°C for 30 s, and extension at 68°C for 1 min. A final extension step was performed at 68°C for 10 min.

TOPO cloning

TOPO vector obtained from TOPO^{TM} TA Cloning^{TM} Kit for Sequencing (ThermoFisher Scientific Inc., MA, USA) was utilized for cloning reactions following the manufacturer's protocol for TOP10 *Escherichia coli* cells. A quantity of 20 ng of PCR product was used and a 30-min incubation period was implemented to enhance colony yield. Successful cloning was confirmed by colony PCR with M13 Forward Primer (5′-GTAAAACGACGGCCAG-3′) and M13 Reverse Primer (5′-CAGGAAACAGCTATGAC-3′) and positive clones were expanded, plasmids were isolated by ZymoPURE Plasmid Miniprep Kit according to manufacturer's protocol (#D4209, Zymo Research, CA, USA) and confirmed by Sanger sequencing (Azenta Life Science, MA, USA). The V, D, J, and CDR3 regions were aligned through the National Center for Biotechnology Information IGBLAST [36].

DNA transcription

Transcription was performed by HiScribe® T7 Quick High Yield RNA Synthesis Kit (#E2050S, NEB, MA, USA) following the manufacturer's protocol including DNase treatment. Resulting RNA was purified by LiCl precipitation method and 1 µg of the resulting RNA was subjected to polyadenylation by *E. coli* Poly(A) Polymerase (#M0276L, NEB, MA, USA) with additional purification step by LiCl precipitation.

Library preparation

Twelve libraries were prepared starting from a total of 200 ng polyadenylated RNA following SMARTer human TCR a/b profiling kit (#635016, Takara, CA, USA) according to the manufacturer's instructions. Library preparation included reverse transcription followed by two PCR steps, in the first PCR, TCR sequences are amplified from the first-strand cDNA and the second PCR amplifies those sequences covering the full-length TCR variable and joining regions with part of constant region while integrating sequencing adapters through the semi-nested approach. Unique dual index primers (one forward, one reverse) are incorporated into the process resulting in pair-end reads. Because of the length of TCR sequences (~700 bp), this approach allows for obtaining full TCR sequence that covers the CDR3 region without the need for specific J primers. Each library's quality and expected size of
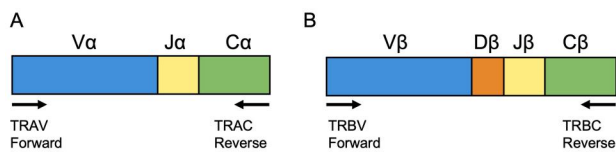
approximately 700 bp were assessed by Agilent high sensitivity DNA ScreenTape (#5067-5584, Agilent Technologies, CA, USA). Purified library concentration was estimated by Qubit dsDNA HS assay kit (#QS32854, Thermofisher Scientific, MA, USA) and further determined by qPCR using NEBNext library quant kit for Illumina (#E7630L, NEB, MA, USA). The library was pooled based on the qPCR results, and then approximately 2 nM library was denatured and diluted to 10 pM according to Illumina instructions and then loaded onto the MiSeq sequencing system as part of a bigger run utilizing 2 × 300 bp paired-end reads (#MS-102-3003, Illumina, CA, USA).

TCR sequencing data analysis

FASTQ files obtained from the sequencing were deposited in the SRA database under the accession number PRJNA1102507. Extraction and alignment of the TCR repertoire from raw .fastq files of TCR-seq data were performed from paired-end reads using MiXCR tool version 4.1.2 according to the default build-in presets for SMARTer human TCR a/b profiling kit with -cdr3 preset to assemble CDR3 region (#635016, Takara, CA, USA) [29]. The following code is utilized to perform the analysis: *mixcr analyze Takara-human-tcr-V1-cdr3/input_R1.fastq.gz/input_R2.fastq.gz/result.* MiXCR is publicly available on its developer's GitHub repository at https://github.com/milaboratory/mixcr.

To validate the results from MiXCR, we performed an additional analysis of the raw .fastq files using the TRUST4 tool processing the data with "trust-smartseq.pl" wrapper [30]. The following code was used to perform the analysis: *perl trust-smartseq. pl -1 read1_list.txt -2 read2_list.txt -t 8 -f hg38_bctcr.fa—ref human_ IMGT+C.fa -o TRUST.* TRUST4 is publicly available on its developer GitHub repository at https://github.com/liulab-dfci/TRUST4.

We also incorporated analysis performed by CATT [37] using the default settings following the command *catt —f1 inputFile1 —f2 inputFile2 -o outputName.* CATT is publicly available on its developer GitHub repository at https://github.com/GuoBioinfoLab/CATT.

Annotated output of CDR3 clonotypes with their corresponding V(D)J regions, clonotype count, and frequency was used for downstream analysis. The percentage of correctly assembled CDR3s was calculated under the criteria of CDR3 amino acid sequence with correctly assigned V(D)J genes matching the results of the Sanger sequencing from the input clonotypes.

VDJtools 1.2.1 [38] was used to visualize the clonality within each library in the form of a three-layer donut utilizing the PlotQuantileStats functionality of the software. TCR power [34] was used to estimate the detection limit of generated libraries through TCR detection power which is a function of true (expected) TCR frequency, TCR sample count, sequencing depth, and cut-off applied to the number of reads considered. We used 10^{-2} as the target T-cell frequency with read threshold of three reads.

Statistical analysis

Data from each computational tool for correlative analysis were prepared by calculating fraction of each clone across all four libraries for each strategy, separately for TRA and TRB. Calculated fractions were subjected to -log(Y) transformation. Pearson correlation was calculated for four libraries of strategy B and separately for four libraries of strategy C for TRA and TRB on transformed data to determine the relationship between expected (based on the input clonotypes) and actual (as determined by computational tool count analysis) ratios of TRA and TRB clones in all libraries prepared in strategy B and strategy C with $P < 0.05$ considered significant. Receiver operating characteristic (ROC) curves were generated using Prism 10 version

10.1.1 using counts of clonotypes according to whether they were correctly assembled (correct CDR3 amino acid sequence, *V* gene, and *J* gene) which were considered true positive; and incorrectly assembled by the computational tool were considered false positive. Sensitivity is defined as a measure of the proportion of true-positive clones that are correctly identified by the tool, while specificity is a measure of the negatively identified clones by the computational tool. ROC curve is plotted with sensitivity on the *y*-axis, while 1-specificity (the false-positive rate) is plotted on the *x*-axis. Quantile statistics for the three-layer donut plot were performed using PlotQuantileStats module of VDJ tools based on the analysis output from MiXCR. TCRpower tool was used to calculate the minimum number of sampled TCR and sequencing reads required for a 95% probability of detecting a target TCR with clonal frequency 10^{-2}.

TCR cloning

Previously described paired TRA and TRB construct of CAVVGGSQGNLIF and CASSLSPGGGYGYFT were designed to serve as a positive control for the transduction experiments [39], while a construct consisting of CAVRDNYGQNFVF and CASSPRGNTGELFF was designed based on our in-house TCR sequencing data. Briefly, inserts containing entire TRA and TRB fragments containing variable, joining, diversity (for TRB), and mouse counterparts of constant regions were codon-optimized, synthesized, and cloned (Azenta Life Science, MA, USA) into pMIG-II vector (#52107, Addgene, MA, USA).

Cell culture

Experiments used Jurkat and Phoenix-Ampho cells. Phoenix Ampho cells were cultured in 75 cm^2 tissue culture flasks with Dulbecco's Modified Eagle Medium (DMEM) culture medium (#11965118, Thermofisher, MA, USA) supplemented with 10% fetal bovine serum (FBS, #10437028, Thermofisher, MA, USA). Jurkat cells were cultured in 75 cm^2 suspension culture flasks with RPMI1640 medium (#11875119, Thermofisher, MA, USA). Cells were incubated at 37°C in humidified air with 5% CO_2.

Retroviral transduction of TCR clones into Jurkat cells

Empty pMIG-II vector and TCR-containing plasmids were transfected into Phoenix-Ampho cells using jetPRIME transfection reagent (#101000015, Polyplus) to produce retrovirus. Virus-containing supernatant was collected at 48- and 72-h post-transfection, then centrifuged at 500 g for 10 min at 32°C to remove any remaining cell debris, and then sterile-filtered. Fresh retroviral supernatant was placed in 24-well plates pre-coated with retronectin as per manufacturer instruction (#T100A, Takara, CA, USA) and placed into a centrifuge pre-warmed to 32°C for 2 h at 2000 g. Supernatant was discarded and 500,000 Jurkat cells in RPMI1640 supplemented with 10% FBS were added to respective wells containing retronectin-bound virus and centrifuged for 5 min at 500 g. Second infection was performed the following day with fresh viral supernatant. Cells were then expanded, and the transduction efficiency was determined 7 days post first infection by staining the cells with anti-mouse TCR-β chain antibody (PE, #12-5961-82, Thermofisher Scientific, MA, USA). Samples were analyzed immediately after staining on Fortessa X20 Cell Analyzer (BD Biosciences, NJ, USA) and processed using FlowJo software (BD Biosciences, NJ, USA).

Results

Generation of plasmid libraries of TRAV and TRBV segments

PCR products were generated using primers corresponding to nine *TRAV* and *TRBV* genes [40, 41] using cDNA generated from the RNA of a healthy donor. Five of each of the *TRAV* and *TRBV* were selected based on their prevalence in the top 10 most common segments reported in our previous study [41] as well as other reported studies [40], while others were chosen randomly. Resulting PCR products were subsequently cloned into pCR2.1 Topo vector following standardized workflows. Positive clones were identified by *lacZ* gene disruption, which prevented β-galactosidase activity, keeping colonies white, while clones with functional enzymes turned blue from X-gal metabolism. Positive clones selected based on blue–white colony screening were further confirmed by colony PCR (Fig. 3A and B).

Selected positive clones were expanded in ampicillin-containing medium and plasmids were isolated and purified from the cell pellets by performing miniprep. The concentration of each plasmid was quantified by Nanodrop, and each isolated plasmid was verified for the presence of the desired band corresponding to the inserted gene size by PCR followed by gel electrophoresis. Results indicated that the amplified region was approximately 1 kb in size, revealing the successful amplification of the insert gene and confirming the presence of the target insert in all plasmids (Fig. 3C and D). The isolated plasmids were sent for Sanger sequencing to confirm specificity of primer to the specific *V* and *C* genes and allowed for further identify *J* region and CDR3 sequence of resulting TCR sequence. Plasmid sequences were aligned to the reference *V*, *D*, and *J* genes from IMGT to determine the exact genes in each of the plasmids. The Sanger sequencing results are presented in Table 2, column 1 indicates variable region according to the IMGT classification where forward primer was designed to generate an insert that was cloned into a vector.

To mimic standard 5′-RACE-like workflow following SMARTer TCR a/b profiling kit protocol (Takara) for TCR repertoire library preparation, we synthesized RNA of selected TCR receptors.

Similar-sized bands were observed after the addition of T7 promoter sequences to each *V* gene by PCR across all individual TCR clones (Fig. 3E and F), while multiple bands with similar sized and varying intensities can be observed for plasmids pooled in one tube at ratios listed in Table 1 indicating possible PCR bias between sequences (Fig. 3G). PCR products were purified and transcribed using the same amount of DNA as an input. Transcription efficiencies done on the pools of PCR products in ratios specified in Table 1 have ranged from 1478.9 ng/µl to 8574.7 for strategy A and from 1973.1 ng/µl to 15027.2 ng/µl for strategy B. Transcription efficiencies for individual TCRs in strategy C ranged from 5846.8 ng/µl for TRAV26-1 to 12180.9 ng/µl for TRAV21 (TRA), and from 7084 nl/µl for TRBV12–4 to 10073.5 ng/µl for TRBV7-2 (TRB), suggesting some variability in transcription efficiencies between TCRs. RNAs generated by transcription from strategy A and strategy B were diluted to 1000 ng/µl and 5 µg of each RNA pool were polyadenylated, while individual TCR's RNA from strategy C were polyadenylated separately resulting in products ranging from 129.7 ng/µl for TRAV5 to 233 for TRAV12-2 (TRA), and 193.4 ng/µl for TRBV7-2 to 252.1 for TRBV27 (TRB). Resulting polyadenylated RNAs were pooled together in ratios indicated in Table 1. Twelve libraries were prepared with an average size ∼765 bp (Fig. 4A and B). Small differences in sizes between libraries might be related to a technical issue in migration of control markers. No significant impurities were observed in any of the library preparations (Fig. 4A and B). Quantification by qPCR showed ranging concentrations of prepared libraries (Fig. 4C) with strategy C having the lowest concentrations of 71.59 nM on average and strategy B having the highest concentrations of 131.73 nM on average. The samples were further pooled together in amount of 2 nmol of each library and further pooled with other unrelated samples for sequencing to contribute to 2.85% of the final pool, which was denatured and diluted to 10 pM before being sequenced on the Miseq system.

MiXCR and TRUST4 correctly assemble input clonotypes

Raw FASTQ files were processed by MiXCR, TRUST4, and CATT to extract TCR clones following default settings specific to SMARTer
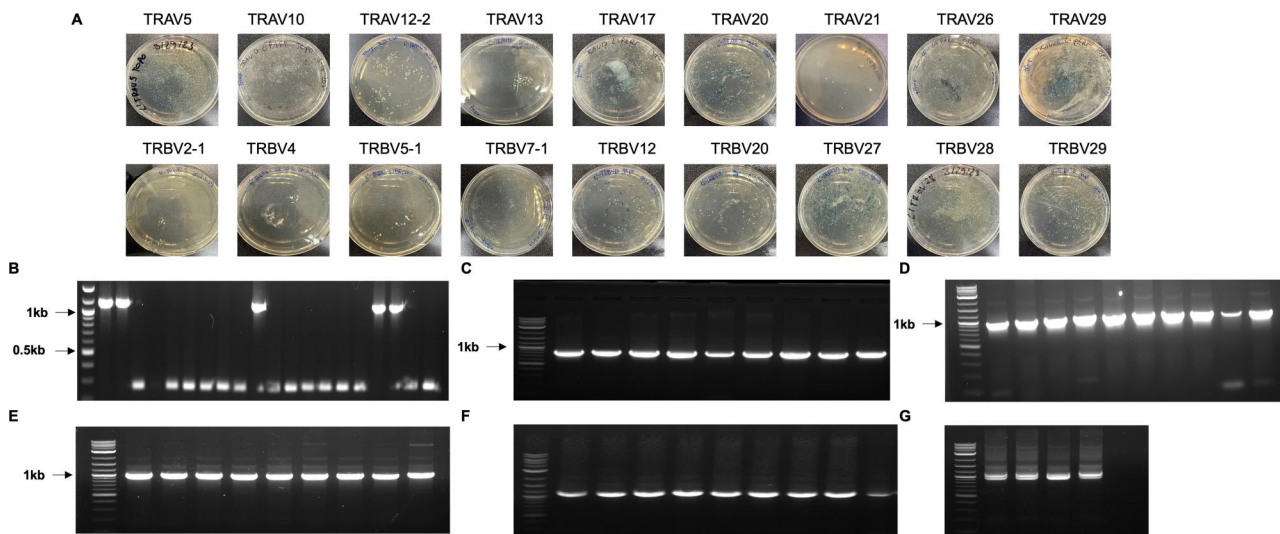

Figure 3 Successfully cloned TCR receptors were amplified with the addition of T7 sequence at 5′-end. PCR products with full-length TCR sequences were cloned into TOPO vectors, and positive clones were selected based on blue–white colony screening. (A) Representative pictures of blue and white colony growth on agar plates. (B) Representative 1% agarose gel of colony PCR screening for positive colonies with product ∼ 1.2 kb showing correct sized clones. (C) TRA and (D) TRB isolated plasmid PCR product. (E) TRA and (F) TRB PCR products with added T7 sequence at 5′–end and (G) PCR products of a pool of plasmids to add T7 sequence for Library B1 − 4.

Table 2. Unique set of V(D)J genes and CDR3s identified by Sanger sequencing corresponding to the different TCR clones that were generated.

Plasmids	V region	J region	C region	CDR3 Sequences	CDR3aa
L-TRAV5	TRAV5*01	TRAJ 37*02	TRAC	GCAGAGAGTAAGAGCAACACAGGCAAACTAATC	AESKSNTGKLI
L-TRAV10	TRAV10*01	TRAJ 18*01	TRAC	GTGGTGAGCGACAGAGGCTCAACCCTGGGGAGGCTATAC	VVSDRGSTLGRLY
L-TRAV12-2	TRAV12-2*01	TRAJ 42*01	TRAC	GCCGTGAACAAGGGCGGGCATGGGAGGAAGCCAAGGAAATCTCATC	AVNKAGMGGSQGNLI
L-TRAV13-1	TRAV13-1*02	TRAJ 53*01	TRAC	GCAGCAAAAAATGGAGGTAGCAACTATAAACTGACA	AAKNGGSNYKLT
L-TRAV17	TRAV17*01	TRAJ 50*01	TRAC	GCTACGGACGTAACCTCCTACGACAAGGTGATA	ATDVTSYDKVI
L-TRAV20	TRAV20*02	TRAJ 42*01	TRAC	GCCGTGCAGGGAGGAAGCCAAGGAAATCTCATC	AVQGGSQGNLI
L-TRAV21	TRAV21*01	TRAJ 8*01	TRAC	GCTCCCTATACTGGAGGCTTCAAAACTATC	LCPWGQAFRNL
L-TRAV26-1	TRAV26-1*01	TRAJ 29*01	TRAC	ATCCCACACGATTCAGGAAAACACACCTCTTGTC	IPHDSGNTPLV
L-TRAV29/DV5	TRAV29/DV5*04	TRAJ 26*01	TRAC	GCAGCAAGCGGCGCGTAACTATGGTCAGAATTTTGTC	AASARNYGQNFV

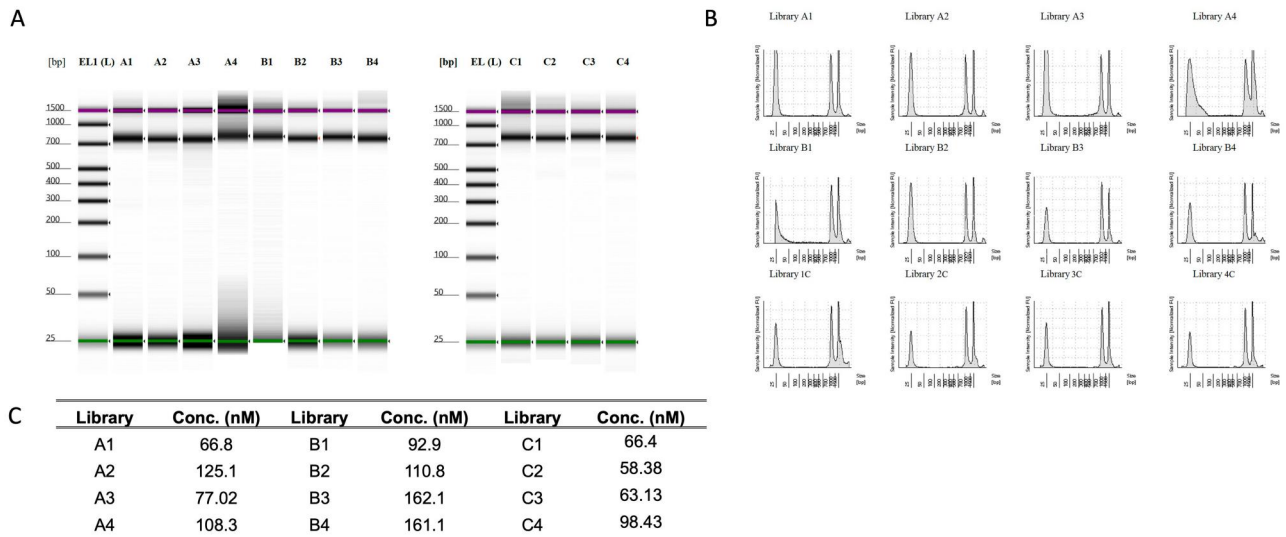| Plasmids	V region	D Region	J region	C region	CDR3 Sequences	CDR3aa
L-TRBV2	TRBV2*01	TRBD01*01	TRBJ 1-2*01	TRBC1	TGTGCCAGCAGCTCTGCTCGGGACAGGGGCCCTGGCTACACCTTC	ASSSARDRGPGYT
L-TRBV4-1	TRBV4-1*01		TRBJ 1-1*01	TRBC1	TGCGCCAGCAACGACGGGAACACTGAAGCTTTCTTT	ASNDGNTEAF
L-TRBV5-1	TRBV5-1*01	TRBD2*01	TRBJ 2-7*01	TRBC1	TGCGTCAGCAGCTTGGCTAGCGTGACCTACGAGCAGTACTTC	VSSLASVTYEQY
L-TRBV7-2	TRBV7-2*02	TRBD2*02	TRBJ 2-7*01	TRBC1	TGTGCCAGCAGCTTAACTAGCGGGAAGGTTAGGAGCTCCTACGAGCAGTACTTC	ASSLTSGKVRSSYEQY
L-TRBV12-4	TRBV12-4*01	TRBD1*01	TRBJ 1-1*01	TRBC1	TGTGCCAGCAGCCTTGGACAGTTAATGAACACTGAAGCTTTCTTT	ASSLGQLMNTEAF
L-TRBV20-1	TRBV20-1*01		TRBJ 2-1*01	TRBC1	TGCAGTGCTAGAGGGTTGCTCAATGAGCAGTTCTTC	SARGLLNEQF
L-TRBV27	TRBV27*01	TRBD2*02	TRBJ 2-5*01	TRBC1	TGTGCCAGCAGTTTGTAGCGGGAGGCAAGAGACCCAGTACTTC	ASSL*R_GQETQY
L-TRBV28	TRBV28*01	TRBD1*01	TRBJ 1-5*01	TRBC1	TGTGCCAGCAGTTCCACGGACAGACGCCGTCCAAAGCATTTT	ASSPRDRRRPKH
L-TRBV29-1	TRBV29-1*01	TRBD2*01	TRBJ 2-1*01	TRBC1	TGCAGGCGTTGAACCTAGCGACTCCTACGAATGAGCAGTTCTTC	SVEPSDSYNEQF

Figure 4 Quantification and assessment of generated libraries. (A) Gel electrophoresis image with an electronic ladder (EL) with a band indicating successful library generation of a size slightly higher than 700 bp. (B) Library A1–C4 size distribution with bottom and top marker from Tapestation. (C) qPCR quantification of each generated library.

Human TCR a/b Profiling Kit library preparation as described in Materials and Methods section. The MiSeq sequencing resulted in 290208 of passed filter reads, which contributed to 1.12% of the entire sequencing run.

Clonotypes were considered correct when CDR3 amino acid sequence and assigned V–J genes from MiXCR, TRUST4, or CATT were identical to those from Sanger sequencing of input clonotypes. We performed qualitative analysis on all libraries utilizing MiXCR, TRUST4, and CATT and found that libraries A1–A4 were missing TRAV12-2 and one TRBV12-4 in the output of all three computational tools; therefore, we have excluded it from further quantitative analyses. Libraries B1–4 and C1–4 had successfully resulted in the identification of all input clonotypes using either MiXCR or TRUST4. While CATT was not able to identify correctly TRAV21 or TRBV27 in any of the libraries, in library L4B none of TRA or TRB were correctly identified, similarly in L1C and L4C none of TRB clones were identified correctly. Library L2C only identified two reads of TRAV10 among TRA. Under the selected criteria (matching CDR3aa, V gene and J gene), MiXCR and TRUST4 were able to identify more correct sequences than CATT (Supplementary Tables S2–S4).

Expected number of total TCR reads is presented in Supplementary Fig. S1. The number of aligned reads by the computational tools is lower than reads obtained on the Illumina platform; however, MiXCR and TRUST4 correctly aligned a similar number of reads across the libraries, while CATT exhibited a much lower number of correctly aligned reads for most of the prepared libraries. We calculated the percentages of correct sequencing reads within each library by dividing the total number of reads from correctly assembled clonotypes by the total number of reads from given library. MiXCR analysis resulted in 98.5% correct sequences on average for TRA in libraries B1–4 and C1–4 (range: 96.1–99.6%), TRUST4 had 97.9% correct sequences on average for TRA in libraries B1–4 and C1–4 (range: 95.9–99.2%). CATT analysis had 73.0% correct sequences on average for TRA in libraries B1–4 and C1–C4 (range: 0–98%).

For TRB, MiXCR analysis correctly identified 98.3% of sequences in libraries B1–4 and C1–4 (range: 97.3–99.3%), while TRUST4 had 97.4% correct sequences on average for TRB in libraries B1–4 and C1–4 (range: 96.3–98.6%). The number of reads corresponding to TRBV28 and TRBV4-1 clones was consistently and significantly lower than expected based on the input material across all libraries generated for both MiXCR and TRUST4. CATT analysis resulted in 45.0% correct sequences on average for TRB in libraries B1–4 and C1–4 (range: 0–92.9%).

While the number of correctly assigned sequences is much lower for CATT, this is mainly because consistently TRAV21 and TRBV27 clones were missing in all libraries; and for some libraries, CATT was not able to simultaneously assign correct V and J segments.

Individual transcription of each TCR clone results in more accurate quantitative sequencing result

Experimental ratios between all true-positive TCR clones were calculated by dividing the number of reads of each of the clones by the number of a reference clone. For TRA, we normalized the ratios of each TRA clone to TRAV10 as a reference, while for TRB, we used TRBV29-1. The TRAV10 and TRBV29-1 were chosen as a reference giving consistent results across all libraries generated. For libraries B4 and C4, we further normalized the data by calculating ratio of TRB to TRA. To compare the expected ratios that were based on the input material for each library preparation strategy to the actual ratios based on the experimental read number values, we combined all four libraries from each strategy separately for TRA and TRB and calculated the fraction of the total for each clone and transformed the data by −log(Y) and performed Pearson correlation. Pearson correlation for TRA obtained by strategy B shows positive correlation between the expected and actual ratios for all computational tools (MiXCR: $R^2 = 0.79$, $P < 0.001$; TRUST4: $R^2 = 0.76$, $P < 0.001$; CATT: $R^2 = 0.83$, $P < 0.001$, Fig. 5A), while low positive correlation for TRB is found for MiXCR and TRUST4, but not significant for CATT (MiXCR: $R^2 = 0.17$, $P = 0.012$; TRUST4: $R^2 = 0.22$, $P = 0.004$; CATT: $R^2 = 0.23$, $P = 0.063$, Fig. 5B). Pearson correlation for TRA obtained by strategy C shows strong positive correlation between the expected and actual ratios for all computational tools (MiXCR: $R^2 = 0.88$, $P < 0.001$; TRUST4: $R^2 = 0.90$, $P < 0.001$; CATT: $R^2 = 0.83$, $P < 0.001$, Fig. 5C). Lower pearson correlation is found for TRB (MiXCR: $R^2 = 0.31$, $P < 0.001$; TRUST4: $R^2 = 0.38$, $P < 0.001$; CATT: $R^2 = 0.33$, $P = 0.021$, Fig. 5D). Pearson correlation for strategy C shows stronger
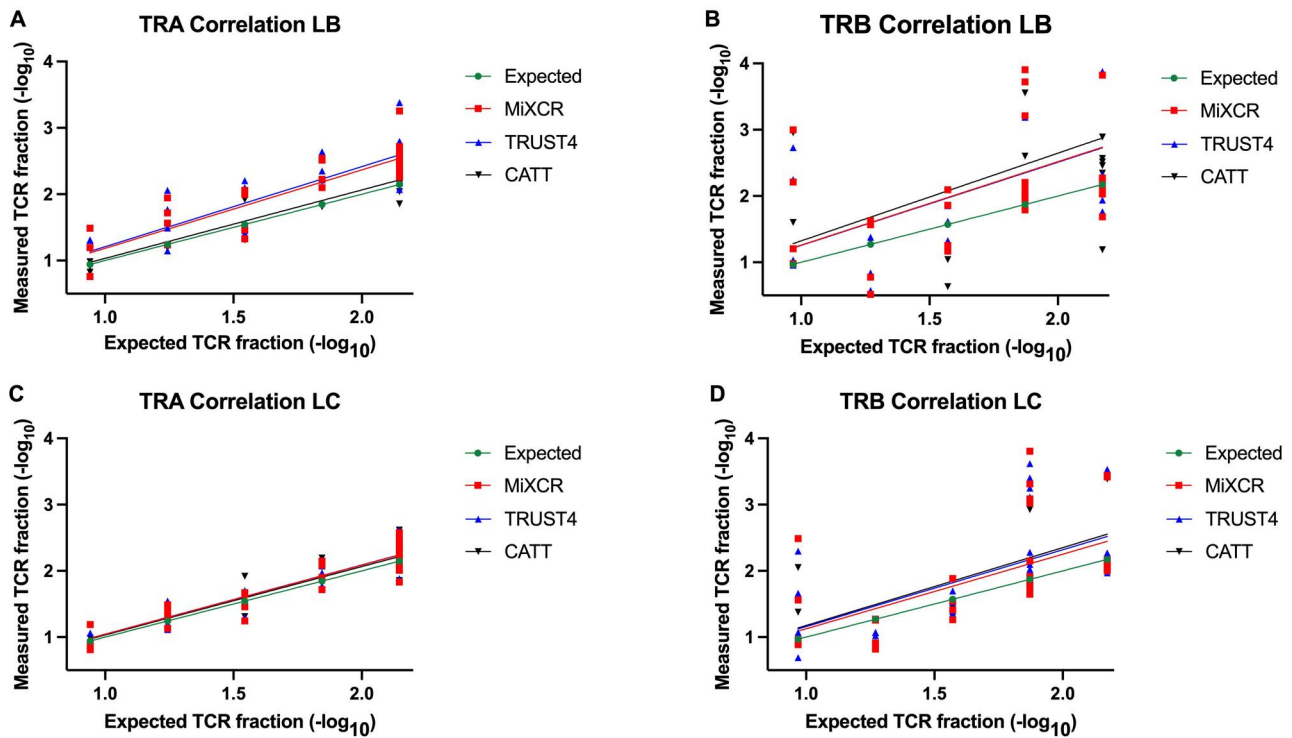
Figure 5 Actual versus expected TCR ratios. Plots represent measured fraction of each clone in the libraries versus expected fraction (ground truth) of each clone based on its input for library preparation for Strategy B TRA (A), Strategy B TRB (B), strategy C TRA (C), Strategy C TRB (D). Solid lines on the graphs represent linear regression.

correlations between the actual and expected ratios indicating that strategy C is more suitable for preparation of control libraries.

Further CDR3 ratios from MiXCR are displayed in Fig. 6 by plotting quantile statistics that resemble clonality of the repertoire from each of the libraries for TRA and TRB (Fig. 6). The plots show input clonotypes in a high-order layer (clonotypes represented by three or more reads). The size of the segment per clonotype corresponds to the frequency of given clone in a library showing varying ratios between different TRA and TRB clones in each set as per the study design.

Control libraries are characterized by high detection sensitivity

To determine the detection limit of TCRs for our libraries, we estimated the minimal TCR frequency that can be detected with 95% probability of data from libraries B1–4 and C1–4 combined, when aligned using MiXCR (Fig. 7A and B), TRUST4 (Fig. 7C and D), or CATT (Fig. 7E and F). We utilized the TCR power tool [34] that follows the negative binominal model component 2 separately for TRA and TRB. Figure 7 depicts the 95% interval of the calibrated TCR read count models as compared with the experimental read counts for each clone (Fig. 7A–F). Furthermore, we utilized the functionality of the Power Calculator to estimate the minimum number of sampled TCR and sequencing reads that are needed for 95% probability of detecting a target TCR with clonal frequency 10^{-2} in a sample. The detection read cutoff was set to three reads. TRA has better detection efficiency than TRB, requiring a smaller number of reads and the same number of sampled TCR to achieve 95% detection power (Fig. 7G–L).

We plot the ROC curves for TCR in all libraries including each correct clonotype read number from each library and each incorrectly built clonotype read number according to the Sanger

sequencing (Fig. 7M–R). The sensitivity of detected clonotypes seems to be very high in the given library set for both TRA and TRB for MiXCR and TRUST4 and lower for CATT (MiXCR: TRA: AUC = 0.99, TRB: AUC = 0.99, Fig. 7M and N; TRUST4: TRA: AUC = 0.93, TRB: 0.80, Figure 7O and P; CATT: TRA: AUC = 0.68, TRB: AUC = 0.51, Fig. 7Q and R).

Control libraries can be used for TCR-T cells' engineering

TCR sequences that were cloned into pCR2.1 Topo vector can further be utilized as templates to produce paired TRA and TRB inserts to use for bioengineered T-cell production and testing. A schematic of the TCR clone design in mammalian retroviral vector to transduce human T cells is described in Fig. 8A. To confirm the validity of such construct, we designed a construct based on TCR previously published in the literature and another construct based on our in-house TCR sequencing data. Retroviral particles containing each of the TCR clones were used to transduce Jurkat cells. Transduction efficiency was measured by staining the cells with anti-mouse TCR beta constant region antibody to confirm the presence of expressed TCR receptor. Successful transduction can be observed by FITC-positive cells indicating the presence of GFP, while TCR expression is noted by the shift in PE-positive cells (Fig. 8B).

Discussion

NGS has revolutionized the characterization of TCR repertoire, facilitating the discovery of antigen-specific TCR through grouping TCRs with similar CDR3 peptide sequences and analyzing these clusters further grouped into networks [42, 43]. Determination of antigen specificity can be deduced from the TCR sequence using advanced computational tools, which
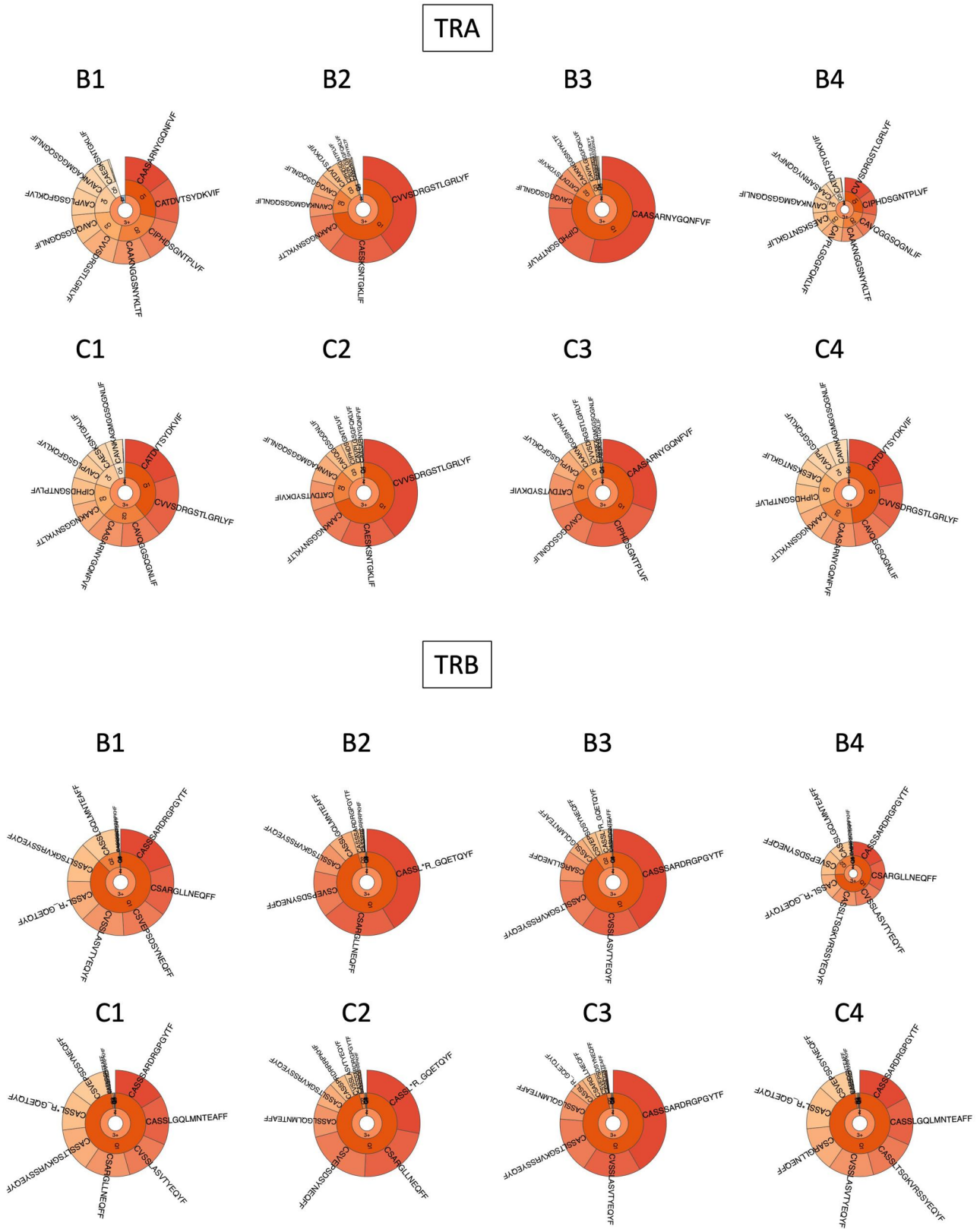
Figure 6 Repertoire clonality plots. The VDJ tools PlotQuantileStats following default parameters was applied on converted MiXCR output to generate three-layer donut charts. The first layer indicates frequency as singleton (1), doubleton (2), or high-order (3+) clonotypes. The second layer is divided into quantiles displaying abundance of the top 20% (Q1), next 20% (Q2) up to Q5 clonotypes for high-order set of clonotypes. The last layer shows individual abundances of the top 10 clonotypes.
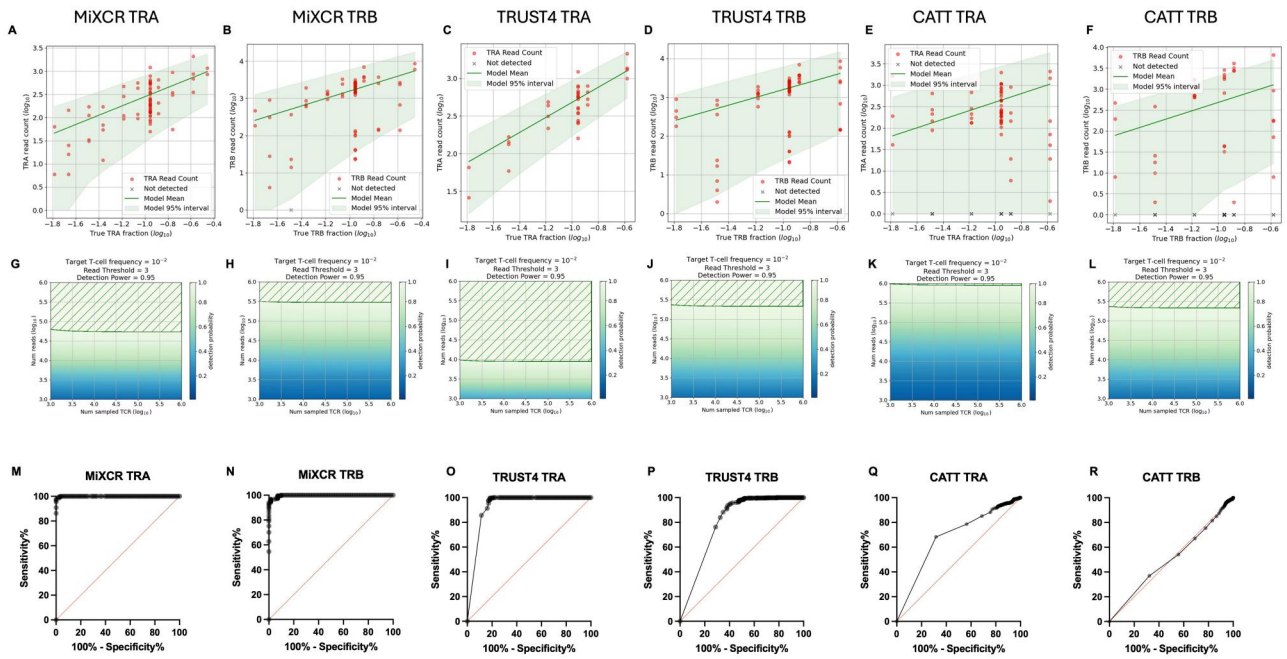
Figure 7 Negative binomial modeling of TCR read counts and detection limit and power estimation. TCR read versus frequency. The dots represent measured counts, while the solid line and shaded areas are the respective mean and 95% prediction intervals of the model for MiXCR (A and B), TRUST4 (C and D), and CATT (E and F). Probability of detecting a TCR with frequency 10^{-2} and read count threshold 3, as a function of the number of sampled TCR and sequencing reads for MiXCR (G and H), TRUST4 (I and J), and CATT (K and L). 95% detection probability is marked by the area shaded by diagonal lines. ROC curve for MiXCR (M and N), TRUST4 (O and P), and CATT (Q and R).
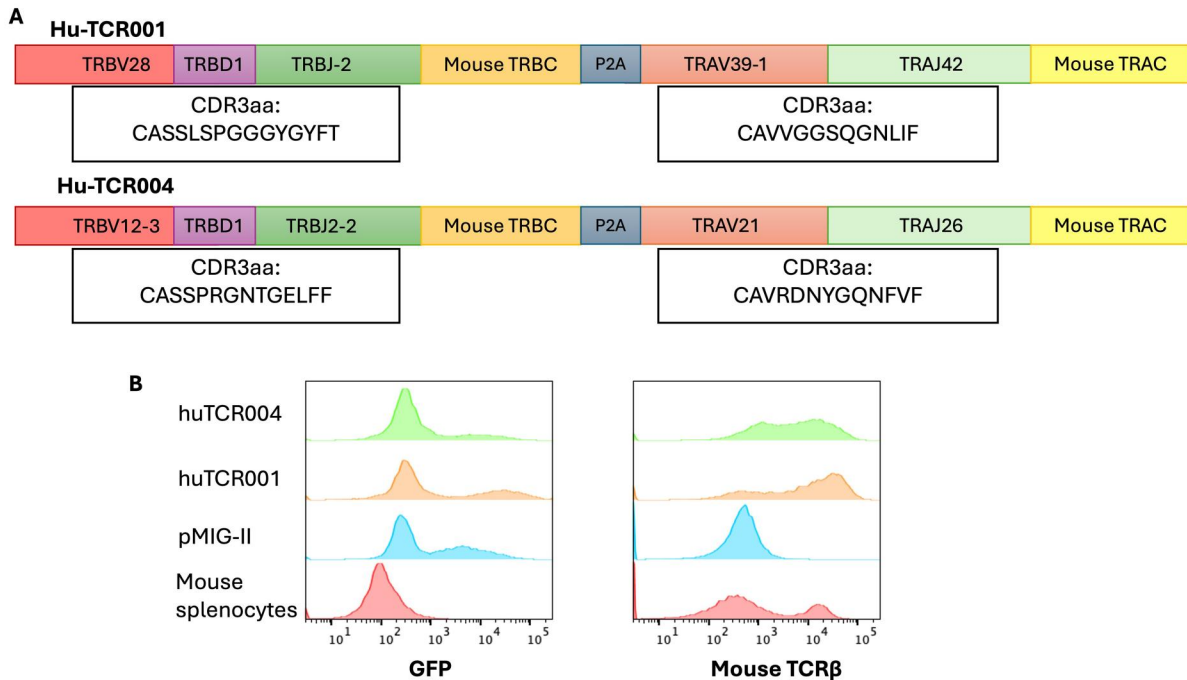

Figure 8 Murinized TCR can be detected by flow cytometry. Schematic of the TCR clone design in mammalian retroviral vector to transduce human T cells (A) Histograms represent expression of GFP and mouse TCR beta constant region (B).

predict pMHC antigen specificity of individual TCRs, mostly achieved through machine learning approaches [16–20, 44]. Accurate and qualitative sequencing methods are crucial for correctly determining antigen specificity [45].

Limitations of current library preparation techniques involve amplification bias affecting the final counts of TCR clones and PCR errors that could lead to incorrect assignment of specific TCR genes [46]. Benchmarking study involving polyclonal CD4 T cells from healthy donors established that multiplex methods show lower reproducibility than RACE-based methods for library preparations. RACE methods had lower efficiency in detecting TRA diversity in comparison to TRB diversity [24]. While those

are reported using commonly used computational tools, none of them were optimized and standardized against a biological control with known composition, which would warrant further studies to identify and eliminate potential errors [47, 48].

In this study, three distinct library preparation strategies were implemented to generate controls and to identify sources of variations in the TCR-sequencing results. While the three strategies led to RNA preparations resembling patient samples, they differed in the approach. Strategy A and strategy B can introduce biases from the transcription step that was done on a pool of clones. While strategy A included PCR that adds the T7 sequence separately to each sequence with subsequent pooling and transcription in bulk. Out of 12 prepared libraries, A2-3, B2-3, C2-3 are more likely to resemble real library preparation from a patient sample considering the rather variable ratio of different TCR receptors. Strategy C would be considered the most relevant as a biological control of known composition since it avoids transcription bias in the preparation of input RNAs allowing for variability/bias introduction only in the library preparation steps that are included in processing patient samples. Strategy B could additionally introduce bias in the transcription step affecting the quantitative TCR analyses as shown in the lack of significant correlation between actual and expected ratios for TRB. Slight discrepancies exist in the TCR repertoire analysis when comparing TRUST and MiXCR while significantly different result is obtained by CATT to the input clonotypes of strategy B and strategy C mainly in the number of reads per clonotype in each of the libraries as well as specific ratios between certain TCR clones. In several libraries, CATT exhibited difficulties in assigning either the correct V or J gene. TRAV or TRBV genes may exhibit differences in their transcriptional efficiencies leading to their different levels in the final pool. TRBV4-1 and TRBV28 are the two clones that consistently exhibit disparities within the strategies implemented in the study. Additionally, while 5′-RACE technology requires less PCR cycles compared to the multiplex PCR approach, this methodology still can introduce biases from PCR, template switching step in reverse transcription, or sequencing itself. Several approaches have been implemented in TCR-sequencing studies to reduce these errors such as optimization of primer concentration or addition of UMIs. While UMIs can resolve PCR and sequencing errors by grouping reads tagged by the same UMIs generating a consensus, it is unable to account for reverse transcription-generated errors. Additionally, a study has shown that non-UMI methods can detect clonotypes at a 10-fold lower frequency than UMI methods [24]. Optimization of computational tools based on biological control would allow to further address the biases and detection issues from different preparation strategies.

Importantly, the generated libraries serve as a repository of TCRs that have been pre-identified and characterized, enabling researchers to quickly select and clone these receptors for the purpose of engineering T cells with specific TCRs. By utilizing such libraries, researchers can bypass the time-consuming process of TCR discovery and validation, accelerating the pace at which new therapies can be moved into clinical testing. Moreover, the pre-engineered TCRs can be optimized for higher affinity and specificity to enhance their therapeutic efficacy. Moreover, the presence of such libraries allows for high-throughput screening of TCRs against a broad range of antigens, enhancing the chances of finding a match for unique or less common antigens presented by tumor cells. This is especially useful for rare cancers or for patients with an uncommon human leukocyte antigen (HLA) type, who may otherwise have limited treatment options. The TCR library generated will serve as a valuable tool for synthetic biology applications, enabling the design and testing of novel T-cell-based constructs against different antigens in different diseases, such as cancer, autoimmune diseases, and viral infections. Researchers can utilize the library to engineer T cells with enhanced functionalities, such as improved targeting, persistence, and resistance to immunosuppressive environments. A trivalent TCR vaccine based on a library of TCR peptides from V genes has been shown to have promising effects in treating multiple sclerosis and other autoimmune diseases [49–51]. Another group has developed a library of virus-specific TCRs composed of five different TCRs restricted to four HLA-class I molecules commonly found in the general population. This approach increased expression of TCR and cytokine production, serving as a foundation for effective treatments for viral infections [52]. Additionally, TCR clones could be found useful in diagnostics to characterize autoreactive T cells [53].

Limitations of study

While the present study acknowledges the limitation of constructing only nine TRA and nine of TRB plasmids, which restricts the comprehensive coverage of TRA and TRB variants particularly involving more rare V genes, we presented a feasible method for biological control preparation that could lead to comprehensive evaluation and benchmarking of computational tools, specifically focusing on sequencing errors and alignment accuracy. Additionally, a standardized approach involving a biological control would promote consistency and reproducibility in TCR analysis, enhancing the reliability of results across different studies and laboratories.

Acknowledgements

Author contributions

Yu-Chun Wei (Data curation [lead], Formal analysis [lead], Methodology [lead], Visualization [lead], Writing—original draft [lead], Writing—review & editing [supporting]), Mateusz Pospiech (Conceptualization [equal lead], Data curation [equal lead], Formal analysis [equal lead], Investigation [equal lead], Methodology [equal lead], Validation [equal lead], Visualization [equal lead], Writing—original draft [equal lead], Writing—review & editing [equal lead]), Yiting Meng (Data curation [supporting], Methodology [supporting], Writing—review & editing [supporting]), and Houda Alachkar (Conceptualization [equal], Data curation [equal], Formal analysis [equal], Funding acquisition [equal], Investigation [equal], Methodology [equal], Project administration [equal], Supervision [equal], Validation [equal], Visualization [equal], Writing—original draft [equal], Writing—review & editing [equal]).

Data availability

The data underlying this article are available in the SRA database and can be accessed with the accession number PRJNA1102507.

Supplementary data

Supplementary data is available at *Biology Methods and Protocols* online.

References

1. Lo Presti E, Dieli F, Meraviglia S. Tumor-infiltrating γδ T lymphocytes: pathogenic role, clinical significance, and differential programing in the tumor microenvironment. *Front Immunol* 2014;**5**:607. https://doi.org/10.3389/fimmu.2014.00607.

2. Rast JP, Anderson MK, Strong SJ *et al.* Litman, alpha, beta, gamma, and delta T cell antigen receptor genes arose early in vertebrate phylogeny. *Immunity* 1997;**6**:1–11. https://doi.org/10.1016/s1074-7613(00)80237-x.

3. Gaulard P, Bourquelot P, Kanavaros P *et al.* Expression of the alpha/beta and gamma/delta T-cell receptors in 57 cases of peripheral T-cell lymphomas. Identification of a subset of gamma/delta T-cell lymphomas. *Am J Pathol* 1990;**137**:617–28.

4. Bruno L, Fehling HJ, von Boehmer H. The alpha beta T cell receptor can replace the gamma delta receptor in the development of gamma delta lineage cells. *Immunity* 1996;**5**:343–52. https://doi.org/10.1016/s1074-7613(00)80260-5.

5. Covacu R, Philip H, Jaronen M *et al.* System-wide Analysis of the T Cell Response. *Cell Rep* 2016;**14**:2733–44. https://doi.org/10.1016/j.celrep.2016.02.056.

6. Lefranc M-P, Giudicelli V, Kaas Q *et al.* IMGT, the international ImMunoGeneTics information system. *Nucleic Acids Res* 2005;**33**:D593–597. https://doi.org/10.1093/nar/gki065.

7. Miles JJ, Douek DC, Price DA. Bias in the αβ T-cell repertoire: implications for disease pathogenesis and vaccination. *Immunol Cell Biol* 2011;**89**:375–87. https://doi.org/10.1038/icb.2010.139.

8. Bassing CH, Swat W, Alt FW. The mechanism and regulation of chromosomal V(D)J recombination. *Cell* 2002;**109 (Suppl)**:S45–55. https://doi.org/10.1016/s0092-8674(02)00675-x.

9. Glanville J, Huang H, Nau A *et al.* Identifying specificity groups in the T cell receptor repertoire. *Nature* 2017;**547**:94–8. https://doi.org/10.1038/nature22976.

10. Turner SJ, Doherty PC, McCluskey J *et al.* Structural determinants of T-cell receptor bias in immunity. *Nat Rev Immunol* 2006;**6**:883–94. https://doi.org/10.1038/nri1977.

11. Huse M, Klein LO, Girvin AT *et al.* Spatial and temporal dynamics of T cell receptor signaling with a photoactivatable agonist. *Immunity* 2007;**27**:76–88. https://doi.org/10.1016/j.immuni.2007.05.017.

12. Hansen SG, Piatak M, Ventura AB *et al.* Immune clearance of highly pathogenic SIV infection. *Nature* 2013;**502**:100–4. https://doi.org/10.1038/nature12519.

13. Shah K, Al-Haidari A, Sun J *et al.* T cell receptor (TCR) signaling in health and disease. *Sig Transduct Target Ther* 2021;**6**:412. https://doi.org/10.1038/s41392-021-00823-w.

14. van der Bruggen P, Traversari C, Chomez P *et al.* A gene encoding an antigen recognized by cytolytic T lymphocytes on a human melanoma. *Science* 1991;**254**:1643–7. https://doi.org/10.1126/science.1840703.

15. Morgan RA, Dudley ME, Yu YYL *et al.* High efficiency TCR gene transfer into primary human lymphocytes affords avid recognition of melanoma tumor antigen glycoprotein 100 and does not alter the recognition of autologous melanoma antigens. *J Immunol* 2003;**171**:3287–95. https://doi.org/10.4049/jimmunol.171.6.3287.

16. Springer I, Besser H, Tickotsky-Moskovitz N *et al.* Prediction of Specific TCR-Peptide Binding From Large Dictionaries of TCR-Peptide Pairs. *Front Immunol* 2020;**11**:1803. https://doi.org/10.3389/fimmu.2020.01803.

17. Sidhom J-W, Larman HB, Pardoll DM *et al.* DeepTCR is a deep learning framework for revealing sequence concepts within T-cell repertoires. *Nat Commun* 2021;**12**:1605. https://doi.org/10.1038/s41467-021-21879-w.

18. Tong Y, Wang J, Zheng T *et al.* SETE: sequence-based Ensemble learning approach for TCR Epitope binding prediction. *Comput Biol Chem* 2020;**87**:107281. https://doi.org/10.1016/j.compbiolchem.2020.107281.

19. Borrman T, Pierce BG, Vreven T *et al.* High-throughput modeling and scoring of TCR-pMHC complexes to predict cross-reactive peptides. *Bioinformatics* 2021;**36**:5377–85. https://doi.org/10.1093/bioinformatics/btaa1050.

20. Lanzarotti E, Marcatili P, Nielsen M. Identification of the cognate peptide-MHC target of T cell receptors using molecular modeling and force field scoring. *Mol Immunol* 2018;**94**:91–7. https://doi.org/10.1016/j.molimm.2017.12.019.

21. Tawara I, Kageyama S, Miyahara Y *et al.* Safety and persistence of WT1-specific T-cell receptor gene-transduced lymphocytes in patients with AML and MDS. *Blood* 2017;**130**:1985–94. https://doi.org/10.1182/blood-2017-06-791202.

22. Hong DS, Van Tine BA, Biswas S *et al.* Autologous T cell therapy for MAGE-A4+ solid cancers in HLA-A*02+ patients: a phase 1 trial. *Nat Med* 2023;**29**:104–14. https://doi.org/10.1038/s41591-022-02128-z.

23. Rosati E, Dowds CM, Liaskou E *et al.* Overview of methodologies for T-cell receptor repertoire analysis. *BMC Biotechnol* 2017;**17**:61. https://doi.org/10.1186/s12896-017-0379-9.

24. Barennes P, Quiniou V, Shugay M *et al.* Benchmarking of T cell receptor repertoire profiling methods reveals large systematic biases. *Nat Biotechnol* 2021;**39**:236–45. https://doi.org/10.1038/s41587-020-0656-3.

25. Kuchenbecker L, Nienen M, Hecht J *et al.* IMSEQ—a fast and error aware approach to immunogenetic sequence analysis. *Bioinformatics* 2015;**31**:2963–71. https://doi.org/10.1093/bioinformatics/btv309.

26. Zhang W, Du Y, Su Z *et al.* IMonitor: a Robust Pipeline for TCR and BCR Repertoire Analysis. *Genetics* 2015;**201**:459–72. https://doi.org/10.1534/genetics.115.176735.

27. Yu Y, Ceredig R, Seoighe C. LymAnalyzer: a tool for comprehensive analysis of next generation sequencing data of T cell receptors and immunoglobulins. *Nucleic Acids Res* 2016;**44**:e31. https://doi.org/10.1093/nar/gkv1016.

28. Gerritsen B, Pandit A, Andeweg AC *et al.* RTCR: a pipeline for complete and accurate recovery of T cell repertoires from high throughput sequencing data. *Bioinformatics* 2016;**32**:3098–106. https://doi.org/10.1093/bioinformatics/btw339.

29. Bolotin DA, Poslavsky S, Mitrophanov I *et al.* MiXCR: software for comprehensive adaptive immunity profiling. *Nat Methods* 2015;**12**:380–1. https://doi.org/10.1038/nmeth.3364.

30. Song L, Cohen D, Ouyang Z *et al.* TRUST4: immune repertoire reconstruction from bulk and single-cell RNA-seq data. *Nat Methods* 2021;**18**:627–30. https://doi.org/10.1038/s41592-021-01142-2.

31. Afzal S, Gil-Farina I, Gabriel R *et al.* Systematic comparative study of computational methods for T-cell receptor sequencing data analysis. *Brief Bioinform* 2019;**20**:222–34. https://doi.org/10.1093/bib/bbx111.

32. Zhang Y, Yang X, Zhang Y *et al.* Tools for fundamental analysis functions of TCR repertoires: a systematic comparison. *Brief Bioinform* 2020;**21**:1706–16. https://doi.org/10.1093/bib/bbz092.

33. Peng K, Nowicki TS, Campbell K *et al.* Rigorous benchmarking of T-cell receptor repertoire profiling methods for cancer RNA sequencing. *Brief Bioinform* 2023;**24**:bbad220. https://doi.org/10.1093/bib/bbad220.

34. Dahal-Koirala S, Balaban G, Neumann RS *et al.* TCR power: quantifying the detection power of T-cell receptor sequencing with a novel computational pipeline calibrated by spike-in sequences. *Brief Bioinform* 2022;**23**:bbab566. https://doi.org/10.1093/bib/bbab566.

35. Böyum A. Isolation of mononuclear cells and granulocytes from human blood. Isolation of monuclear cells by one centrifugation, and of granulocytes by combining centrifugation and sedimentation at 1 g. *Scand J Clin Lab Invest Suppl* 1968;**97**:77–89.

36. IgBlast tool, (n.d.). https://www.ncbi.nlm.nih.gov/igblast/igblast.cgi?CMD=Web&SEARCH_TYPE=TCR&LINK_LOC=igtab (3 March 2024, date last accessed).

37. Chen S-Y, Liu C-J, Zhang Q *et al.* An ultra-sensitive T-cell receptor detection method for TCR-Seq and RNA-Seq data. *Bioinformatics* 2020;**36**:4255–62. https://doi.org/10.1093/bioinformatics/btaa432.

38. Shugay M, Bagaev DV, Turchaninova MA *et al.* VDJtools: unifying post-analysis of T cell receptor repertoires. *PLOS Comput Biol* 2015;**11**:e1004503. https://doi.org/10.1371/journal.pcbi.1004503.

39. Giannakopoulou E, Lehander M, Virding Culleton S *et al.* A T cell receptor targeting a recurrent driver mutation in FLT3 mediates elimination of primary human acute myeloid leukemia in vivo. *Nat Cancer* 2023;**4**:1474–90. https://doi.org/10.1038/s43018-023-00642-8.

40. Li B, Li T, Pignon J-C *et al.* Landscape of tumor-infiltrating T cell repertoire of human cancers. *Nat Genet* 2016;**48**:725–32. https://doi.org/10.1038/ng.3581.

41. Pospiech M, Tamizharasan M, Wei Y-C *et al.* Features of the TCR repertoire associate with patients' clinical and molecular characteristics in acute myeloid leukemia. *Front Immunol* 2023;**14**:1236514. https://doi.org/10.3389/fimmu.2023.1236514.

42. Huang H, Wang C, Rubelt F *et al.* Analyzing the Mycobacterium tuberculosis immune response by T-cell receptor clustering with GLIPH2 and genome-wide antigen screening. *Nat Biotechnol* 2020;**38**:1194–202. https://doi.org/10.1038/s41587-020-0505-4.

43. Singh NK, Abualrous ET, Ayres CM *et al.* Geometrical characterization of T cell receptor binding modes reveals class-specific binding to maximize access to antigen. *Proteins* 2020;**88**:503–13. https://doi.org/10.1002/prot.25829.

44. Dash P, Fiore-Gartland AJ, Hertz T *et al.* Quantifiable predictive features define epitope-specific T cell receptor repertoires. *Nature* 2017;**547**:89–93. https://doi.org/10.1038/nature22383.

45. Scheper W, Kelderman S, Fanchi LF *et al.* Low and variable tumor reactivity of the intratumoral TCR repertoire in human cancers. *Nat Med* 2019;**25**:89–94. https://doi.org/10.1038/s41591-018-0266-5.

46. Lefranc M-P. Immunoglobulin and T Cell Receptor Genes: IMGT (®) and the Birth and Rise of Immunoinformatics. *Front Immunol* 2014;**5**:22. https://doi.org/10.3389/fimmu.2014.00022.

47. Carlson CS, Emerson RO, Sherwood AM *et al.* Using synthetic templates to design an unbiased multiplex PCR assay. *Nat Commun* 2013;**4**:2680. https://doi.org/10.1038/ncomms3680.

48. Peng Q, Vijaya Satya R, Lewis M *et al.* Reducing amplification artifacts in high multiplex amplicon sequencing by using molecular barcodes. *BMC Genomics* 2015;**16**:589. https://doi.org/10.1186/s12864-015-1806-8.

49. Shang X, Xi N, Wang T *et al.* The change of periphery and centra CD4(+);CD25(+);Treg, CD8(+);CD28(-);Treg in the MOG induced model of experimental autoimmune encephalomyelitis. *Xi Bao Yu Fen Zi Mian Yi Xue Za Zhi* 2010;**26**:746–9.

50. Fantin RF, Fraga VG, Lopes CA *et al.* New highly antigenic linear B cell epitope peptides from PvAMA-1 as potential vaccine candidates. *PLoS One* 2021;**16**:e0258637. https://doi.org/10.1371/journal.pone.0258637.

51. Bemani P, Jalili S, Hassanpour K *et al.* Designing and characterization of Tregitope-based multi-epitope vaccine against multiple sclerosis: an immunoinformatic approach. *Curr Drug Saf* 2023;**18**:79–92. https://doi.org/10.2174/1574886316666220429105439.

52. Banu N, Chia A, Ho ZZ *et al.* Building and optimizing a virus-specific T cell receptor library for targeted immunotherapy in viral infections. *Sci Rep* 2014;**4**:4166. https://doi.org/10.1038/srep04166.

53. Dornmair K, Goebels N, Weltzien H-U *et al.* T-cell-mediated autoimmunity: novel techniques to characterize autoreactive T-cell receptors. *Am J Pathol* 2003;**163**:1215–26. https://doi.org/10.1016/S0002-9440(10)63481-5.