scientific reports

OPEN

Cell activation-based screening of natively paired human T cell receptor repertoires

Ahmed S. Fahad[1,10], Cheng Yu Chung[1,10], Sheila N. López Acevedo[1], Nicoleen Boyle[1], Bharat Madan[1], Matías F. Gutiérrez-González[1], Rodrigo Matus-Nicodemos[2], Amy D. Laflin[1], Rukmini R. Ladi[1], John Zhou[1], Jacy Wolfe[1], Sian Llewellyn-Lacey[3], Richard A. Koup[2], Daniel C. Douek[2], Henry H. Balfour Jr.[4,5], David A. Price[3,6] & Brandon J. DeKosky[1,7,8,9]✉

Adoptive immune therapies based on the transfer of antigen-specific T cells have been used successfully to treat various cancers and viral infections, but improved techniques are needed to identify optimally protective human T cell receptors (TCRs). Here we present a high-throughput approach to the identification of natively paired human TCRα and TCRβ (TCRα:β) genes encoding heterodimeric TCRs that recognize specific peptide antigens bound to major histocompatibility complex molecules (pMHCs). We first captured and cloned TCRα:β genes from individual cells, ensuring fidelity using a suppression PCR. We then screened TCRα:β libraries expressed in an immortalized cell line using peptide-pulsed antigen-presenting cells and sequenced activated clones to identify the cognate TCRs. Our results validated an experimental pipeline that allows large-scale repertoire datasets to be annotated with functional specificity information, facilitating the discovery of therapeutically relevant TCRs.

Human T cell receptors (TCRs) mediate cell-based immunity against a variety of cancers and intracellular pathogens via the specific recognition of peptides bound to major histocompatibility complex molecules (pMHCs)[1,2]. Somatic recombination of germline V(D)J gene segments, along with the introduction of junctional variability and subsequent pairing of TCRα and TCRβ chains, yields a highly diverse set of TCRα:β structures[2,3], which are selected in the thymus on the basis of autologous pMHC reactivity[4]. These processes generate a theoretical diversity of $> 10^{15}$ TCRs and an expressed diversity of $\sim 2 \times 10^7$ TCRs[5–7]. Moreover, each unique clonotype, defined by the expression of a unique TCR, recognizes a unique array of pMHCs. A detailed analysis of these diverse functional and genetic landscapes is fundamentally important for our molecular understanding of adaptive immunity in health and disease, but it has proven difficult to combine both required sets of information into one experimental pipeline to enable the wholesale discovery of antigen-specific TCRs.

Single-cell analyses are typically required to capture the dual-gene nature of TCRs[7]. A series of techniques have been established to recover natively paired TCRα:β chains, including relative combinatorics-based pairing[7–9] and single-cell PCR[10–15], exemplified by the 10x Genomics platform[16]. In addition, several approaches enable the identification and isolation of antigen-specific T cells[7,17], including the use of recombinant pMHC multimers in conjunction with fluorescence-activated cell sorting (FACS)[18–22]. However, the labile nature of primary T cells precludes a complete interrogation of specificity profiles against multiple pMHCs, and activation-based approaches are further complicated by a lack of effector readouts in the naive pool and functional heterogeneity in the memory pool, including the potential for exhaustion under conditions of persistent antigen stimulation[23]. Robust methods for repertoire-scale functional screening are therefore needed to help understand the connections between antigen specificity and sequence-identified TCRs.

[1]Department of Pharmaceutical Chemistry, The University of Kansas, Lawrence, KS 66044, USA. [2]Vaccine Research Center, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892, USA. [3]Division of Infection and Immunity, Cardiff University School of Medicine, University Hospital of Wales, Cardiff CF14 4XN, UK. [4]Department of Laboratory Medicine and Pathology, University of Minnesota Medical School, Minneapolis, MN 55455, USA. [5]Department of Pediatrics, University of Minnesota Medical School, Minneapolis, MN 55455, USA. [6]Systems Immunity Research Institute, Cardiff University School of Medicine, University Hospital of Wales, Cardiff CF14 4XN, UK. [7]Department of Chemical Engineering, The University of Kansas, Lawrence, KS 66044, USA. [8]Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02142, USA. [9]The Ragon Institute of MGH, MIT, and Harvard, Cambridge, MA 02139, USA. [10]These authors contributed equally: Ahmed S. Fahad and Cheng Yu Chung. ✉email: dekosky@mit.edu

Low-throughput approaches, including single-cell sorting and limiting dilution[24-28], and even random α:β pairing combined with functional screening of small T cell populations[29], have been used effectively to discover clinically relevant TCRs. Affinity maturation has also been used to generate therapeutic TCRs, albeit with an attendant risk of serious off-target reactivity against autologous pMHCs[30-32]. Another study reported similar single-cell cloning and coculture techniques to interrogate large-scale libraries comprising natively paired TCRs, but in this case, the engagement of soluble pMHCs was used prior to coculture with donor-matched antigen-presenting cells (APCs)[21]. Due to the important differences between pMHC recognition and T cell activation[33], as well as the potential difficulties associated with generating recombinant pMHCs for every peptide antigen, a more direct screening method using cell-based coculture could be preferable for mapping functional features of the repertoire of human TCRs. For this purpose, we developed a high-throughput platform incorporating single-cell TCR sequencing and functional screening directly against cell-expressed antigens, using infectious mononucleosis (IM) as a disease model[34-36]. Our workflow comprised native TCR gene cloning into lentiviral display vectors, followed by activation-based screening of the expression libraries in SKW3 cells[37,38] and the subsequent identification of antigen-specific TCRs via high-throughput sequencing (HTS). The ability to screen thymically selected repertoires comprehensively and repeatedly in this manner should expedite a path to immune discovery and personalized medicine.

Results

High-throughput gene capture and screening of native TCRα:β pairs.

In a previous study, we used an emulsion-based sequencing platform to capture natively paired TCRα:β cDNA amplicons, which were subsequently cloned into expression vectors, displayed in Jurkat cell libraries, and screened using multimeric pMHCs in conjunction with FACS and HTS[34]. This methodology was founded on our established techniques for sequencing, cloning, and screening B cell receptors (BCRs)[39-43]. In the present study, we developed a modified protocol to display natively paired TCRα:β libraries in SKW3 cells, enabling activation-based screening and the characterization of functionally optimal TCRs. The natively paired TCRα:β amplicons interrogated here were obtained previously from two individuals with IM (Fig. 1A)[34]. Briefly, we used a flow-focusing device in the earlier study to encapsulate single T cells inside emulsion droplets containing poly-dT-coated magnetic beads to capture polyadenylated mRNAs (Fig. 1B)[34,40,41,43-45]. Magnetic beads with colocalized TCRα and TCRβ mRNAs were then recovered and used as templates in an overlap extension RT-PCR that physically linked single-cell-derived TCRα and TCRβ genes (Fig. 1B)[34,46]. The resulting TCRα:β libraries were PCR-amplified in the current study to add restriction enzyme sites for cloning into a lentiviral expression vector and subsequently transduced into SKW3 cells (Fig. 1C,D) for activation-based screening via FACS and HTS to link functionality with the gene sequences of individual TCRs (Fig. 1E,F).

Evaluation of restriction enzyme cloning sites for TCRα:β genes.

We tested and validated a set of silent and non-silent mutations that introduced restriction enzyme sites into the TCRα and TCRβ variable leader and constant regions (Fig. 2, Supplementary Tables 1 and 2). The effects of these mutations were evaluated in a mammalian lentiviral display system (Fig. 3A,B). Expression constructs encoding a single TCR (TM9) specific for the human leukocyte antigen (HLA)-B*07:02-restricted HIV-1 Nef epitope RPQVPLRPM (*RM9*) were packaged into lentiviral particles and transduced into J.RT3-T3-5/CD8⁺ (J.RT3/CD8) cells, which were subsequently purified on the basis of mCherry expression via FACS. A monoclonal anti-human TCR antibody and fluorescently labeled tetrameric complexes of *RM9*/HLA-B*07:02 were used to quantify expression of the TM9 TCR. Four optimal mutations were selected for inclusion in the final expression construct to enable direct and efficient library-scale cloning of HTS-captured TCRs (Fig. 3 and Supplementary Fig. 1).

Generation of TCRα:β libraries expressed in SKW3 cells.

TCRα:β amplicons incorporating the optimal mutations were cloned into the lentiviral pLVX-EF1α-IRES-mCherry vector as reported previously[34]. Briefly, we performed an overlap extension RT-PCR incorporating multiplex primers to amplify and fuse single-cell-derived TCRα and TCRβ chains via a linker sequence[34,46]. We also used a suppression PCR to prevent the random association of unfused TCRα and TCRβ amplicons during the bulk semi-nested PCR (Fig. 4A)[10].

As expected, agarose gel electrophoresis revealed dominant unlinked TCRα and TCRβ chain bands after the overlap extension RT-PCR (Fig. 4B), which required specific amplification of the subdominant linked TCRα:β amplicons in the mixture. In contrast to our previous fully nested PCR strategies to amplify linked heavy and light chains from expressed BCRs[34,39,41-44], our TCR cloning strategy was based on unidirectional expression, introducing the potential for unlinked TCRα and TCRβ chains to associate randomly during the bulk semi-nested PCR. To mitigate this issue, we tested a suppression PCR[10]. This approach uses blocking oligonucleotides encoding nonsense sequences at the 5' ends to prevent amplified single TCRα and TCRβ cDNAs from associating via overlap extension by eliminating sequence homology without affecting natively paired TCRα:β amplicons previously linked in the single-bead emulsion (Fig. 4A–C).

We validated this strategy using a series of control PCRs. Similar amounts of linked TCRα:β material were detected using either unlinked TCRα and TCRβ amplicons (Fig. 4C, *PCR #3*) or fully linked TCRα:β genes (Fig. 4C, *PCR #4*) as templates in standard semi-nested PCRs. In the presence of blocking primers, however, the overlap extension RT-PCR product was amplified successfully (Fig. 4C, *PCR #1*), whereas unfused amplicons no longer yielded a linked TCRα:β product (Fig. 4C, *PCR #2*). We also compared the prevalence of EBV-specific TCRs from our two donors after cloning the corresponding HTS-based libraries into J.RT3/CD8 cells using either the standard semi-nested PCR or the suppression PCR[34]. In line with our expectations based on the elimination of non-natively paired TCRs, the frequencies of J.RT3/CD8 cells that bound the cognate pMHC tetramer were an order of magnitude higher after cloning with the suppression PCR (Supplementary Fig. 2).
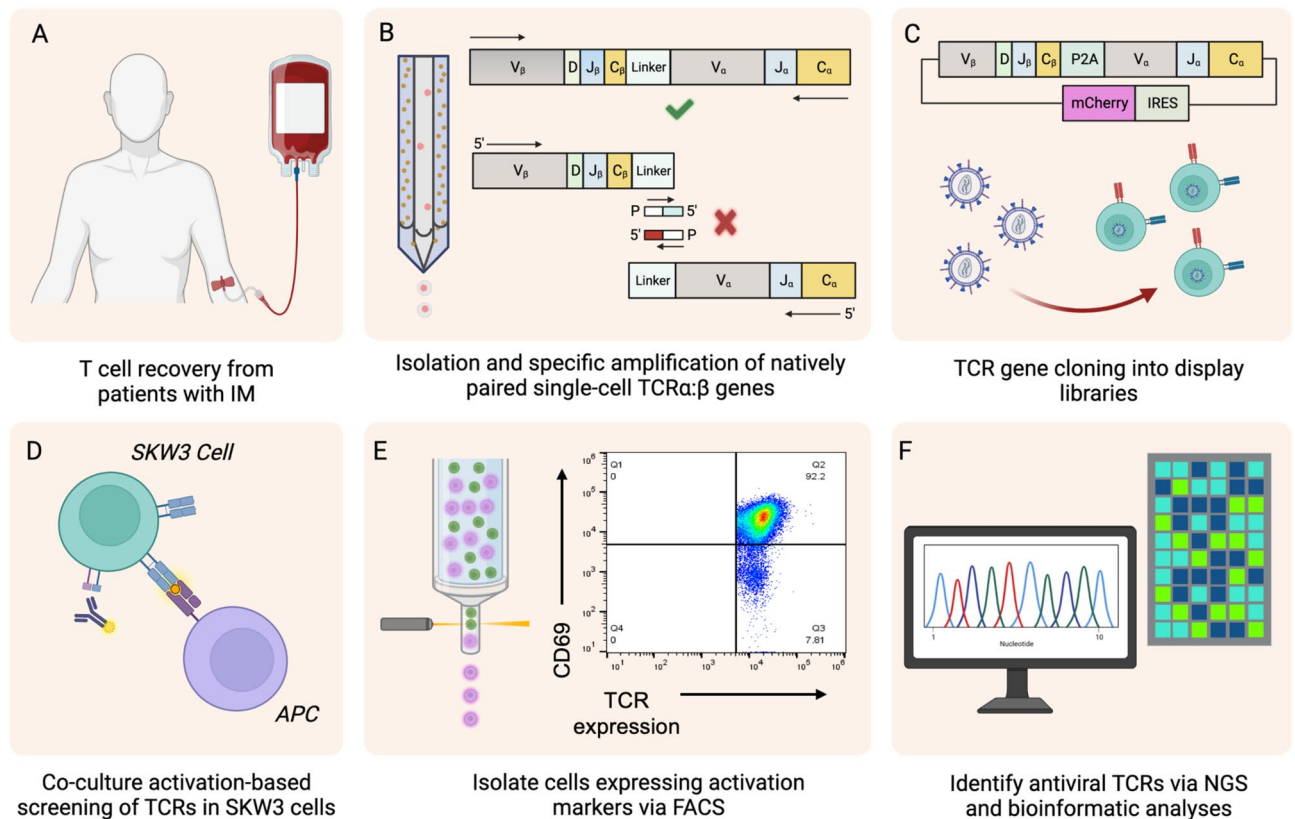
Figure 1. High-throughput TCRα:β gene cloning for cell activation-based screening of human T cell repertoires. (**A**) Peripheral blood mononuclear cells (PBMCs) were recovered from EBV-seropositive donors and expanded briefly in vitro. (**B**) T cells were encapsulated in emulsion droplets with lysis mixture and oligo (dT) beads, and beads with colocalized TCRα and TCRβ mRNAs were subjected to an overlap extension RT-PCR, linking the corresponding variable segments into the same cDNA. A suppression PCR was performed to amplify linked TCRα:β genes in the presence of blocking primers to prevent the association of unpaired TCRα and TCRβ cDNAs. (**C**) Natively paired TCRα:β amplicons were cloned into a lentiviral display construct containing a P2A translation skip motif along with IRES and mCherry elements for mammalian cell expression. TCRα:β plasmids were packaged into lentiviral particles and transduced into SKW3 cells to generate immortalized libraries. (**D**) TCRα:β-SKW3 libraries were cocultured with peptide-pulsed APCs matched to the relevant HLA. (**E**) TCRα:β-SKW3 cells that upregulated CD69 were purified via FACS. (**F**) Purified TCRα:β-SKW3 libraries were characterized via HTS, and computational analysis was used to identify antigen-specific TCRs.

Activation-based screening of TCRα:β expression libraries.

TCRα:β amplicon libraries were cloned into the lentiviral vector expression system with all the required elements for full-length TCRα:β expression on the cell surface[34], including an internal ribosomal entry site (IRES) and an mCherry marker to detect successful transduction via FACS. After cloning the amplified TCRα:β genes into the lentiviral vector backbone using AgeI and BstBI, the linker region was swapped using MluI and SpeI to incorporate a linear DNA construct containing the remaining portions of the TCRβ constant region, a ribosomal skip teschovirus-derived sequence (P2A), and a modified TRAV8-2 leader sequence incorporating an MluI site (Fig. 2, Supplementary Fig. 4) for TCRα expression[34,47]. We generated at least 10^6 transformants in each cloning step to maintain library diversity (Supplementary Figs. 5 & 6). The full-length TCRα:β expression plasmids were then packaged into lentiviral particles and transduced into SKW3 cells for functional evaluation, illustrated here with reference to Donor 1.

TCRα:β-SKW3 libraries expressing mCherry were expanded after purification via FACS. In contrast to primary T cells, these immortalized libraries can be screened over multiple rounds of panning against APCs. T2 cells transduced to express HLA-B*08:01 (T2-B8) were pulsed with the EBV BZLF1 peptide RAKFKQLL (*RAK*) at a concentration of 1 mM and cocultured with mCherry⁺ TCRα:β-SKW3 cells for 24 h. Antigen-specific TCRα:β-SKW3 cells were detected via upregulation of the activation marker CD69 (Fig. 5). Coculture parameters were optimized using the EBV EBNA3A-specific LC13 TCR, which recognizes the FLRGRAYGL (*FLR*) epitope restricted by HLA-B*08:01 (Fig. 5A)[48]. *RAK*-specific TCRα:β-SKW3 cells were enriched > 10-fold compared with the presort libraries after a single round of purification via FACS (Fig. 5B). In addition, there was minimal background activation and minimal reactivity against *FLR*-pulsed T2-B8 cells, despite initial coculture in the presence of donor-mismatched MHCs (Fig. 5B).
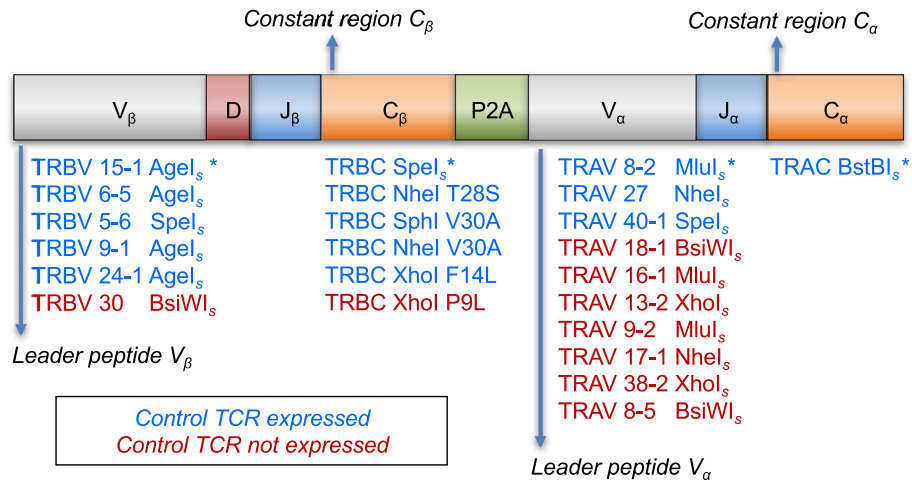
Figure 2. Validation of restriction enzyme site mutations in the TCRα and TCRβ genes. Colors indicate functional mutations (blue) and non-functional mutations (red). *, restriction enzyme site incorporated in the final primer design as reported previously[34]; s, silent mutation. Details of each sequence modification are provided in Supplementary Tables 1 and 2.
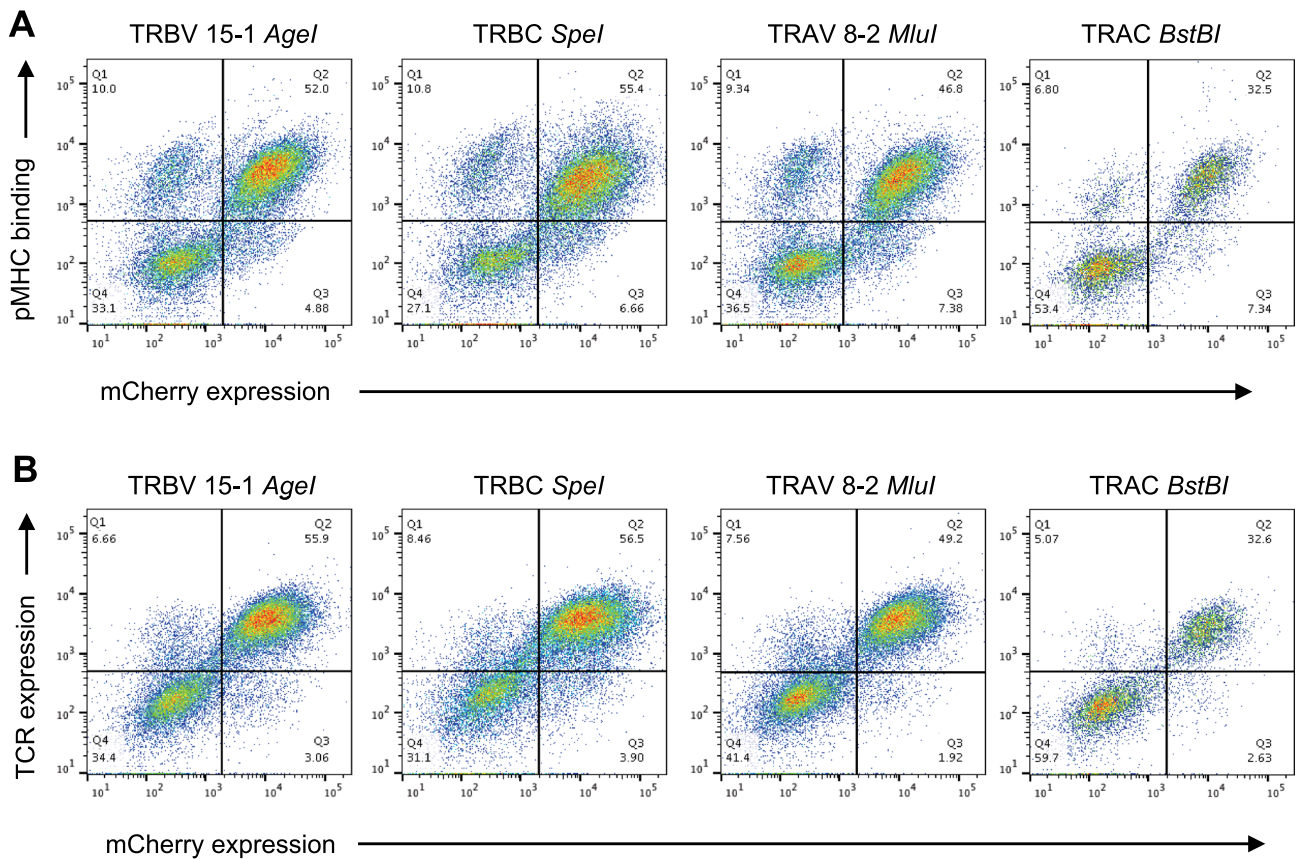

Figure 3. Functionality of restriction enzyme site mutations in the TCRα and TCRβ genes. (**A**) Functional expression of the TM9 TCR cloned via the indicated restriction enzyme sites was tested via flow cytometry using tetrameric complexes of *RM9*/HLA-B*07:02. Internal mCherry expression is shown on the x-axis, and pMHC binding is shown on the y-axis. (**B**) Functional expression of the TM9 TCR cloned via the same restriction enzyme sites was further tested via flow cytometry using anti-human TCRα/β. Internal mCherry expression is shown on the x-axis, and TCR expression is shown on the y-axis. Comparative data from less successful restriction enzyme site mutations are summarized in Supplementary Table 3.
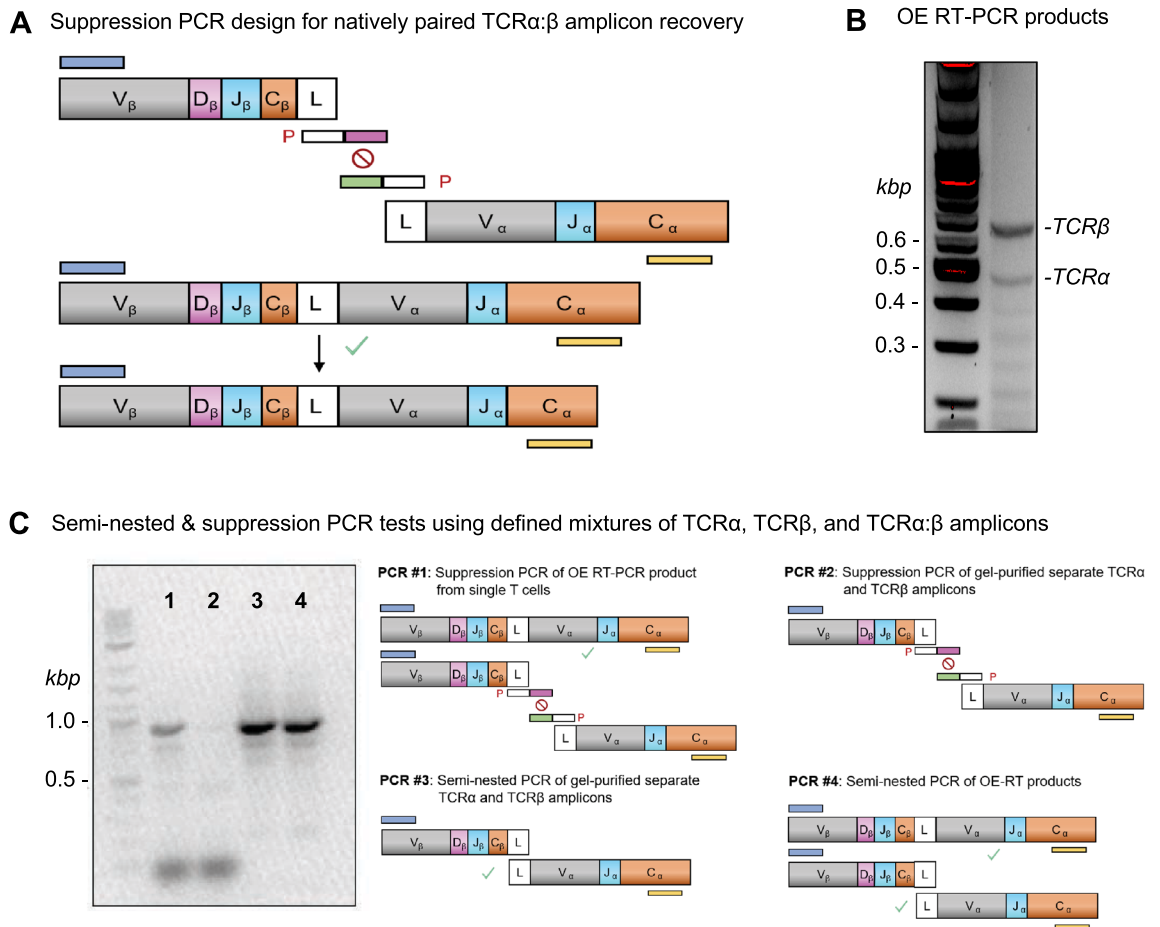
Figure 4. Elimination of non-native TCRα:β pairs via suppression PCR. (**A**) The suppression PCR strategy used to amplify natively paired TCRα:β cDNAs. (**B**) Agarose gel electrophoresis image showing amplification of unpaired TCRα and TCRβ cDNAs via overlap extension (OE) RT-PCR. Left, size ladder; right, experimental lane. (**C**) Agarose gel electrophoresis image comparing products from the first semi-nested PCR with products from various control PCRs. Original gels are presented in Supplementary Fig. 3.

Rapid bioinformatic detection of antigen-specific TCRs. HTS-based library analyses were used to identify *RAK*-specific TCRs. Data were processed as reported previously[34]. Briefly, raw FASTQ files were quality-filtered and annotated using MIXCR[49]. Out-of-frame V(D)J reads were excluded from the dataset, and productive in-frame reads were paired by Illumina ID. Reads were compiled using CDR3 and VJ gene identity. CDR3β sequences were clustered to 96% nucleotide identity after excluding singleton reads to minimize errors introduced via HTS and/or PCR. CDR3β amino acid sequences were used to track TCR clones[42]. The frequency of each individual clone in each sorted sample was calculated to evaluate the functional performance of each library. We also calculated the enrichment ratio for each CDR3β. Two known antigen-specific TCR clones identified by sequence analysis were enriched > 10-fold compared with the parental mCherry+ TCRα:β-SKW3 library and comprised the bulk of the response against cell-displayed *RAK*/HLA-B*08:01 (Supplementary Table 4).

Discussion

T cell immunotherapy holds great promise for the treatment of various cancers and infectious diseases, but comprehensive molecular platforms are required to characterize the antigen specificity, functionality, and translational potential of individual TCRs. In this study, we developed and validated a high-fidelity cloning strategy to enable activation-based screening of natively paired TCRα:β gene libraries against peptide-pulsed APCs. These efforts allowed us to link the efficacy of signal transduction in response to cognate antigen encounter with sequence information across the somatically rearranged genome via an integrated experimental and bioinformatics pipeline, facilitating the discovery of naturally selected and optimally potent TCRs.

Our screening technology was adapted from previous work that enabled the physical identification of antigen-specific TCRα:β clones based on the engagement of multimeric pMHCs[34]. We report here the design and validation of mutations enabling the introduction of restriction enzyme sites for high-throughput cloning of TCRα:β genes into mammalian display vectors and the elimination of randomly unassociated TCRα and TCRβ genes via a suppression PCR. The latter greatly enriched our libraries for natively paired TCRs[10]. Although other methods can be used to enhance the fraction of TCRα:β genes[39–43], the incorporation of a suppression PCR in our workflow proved to be compatible with the use of standard in-line cassettes for gene expression, harmonizing with
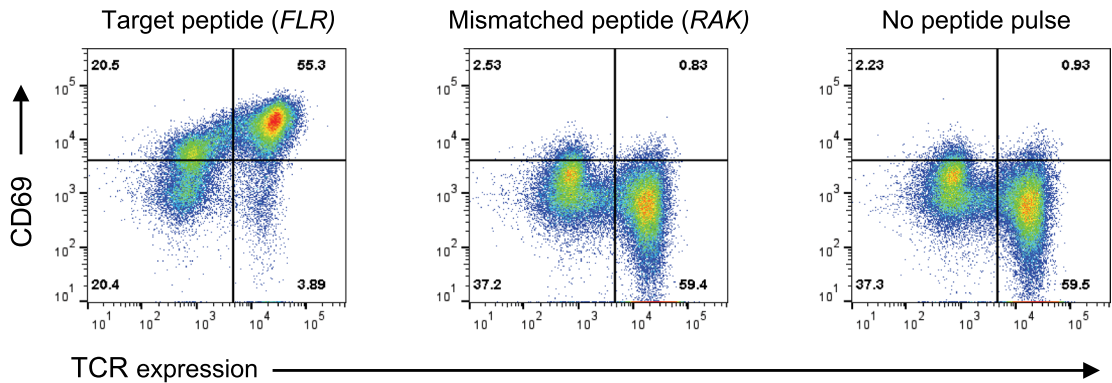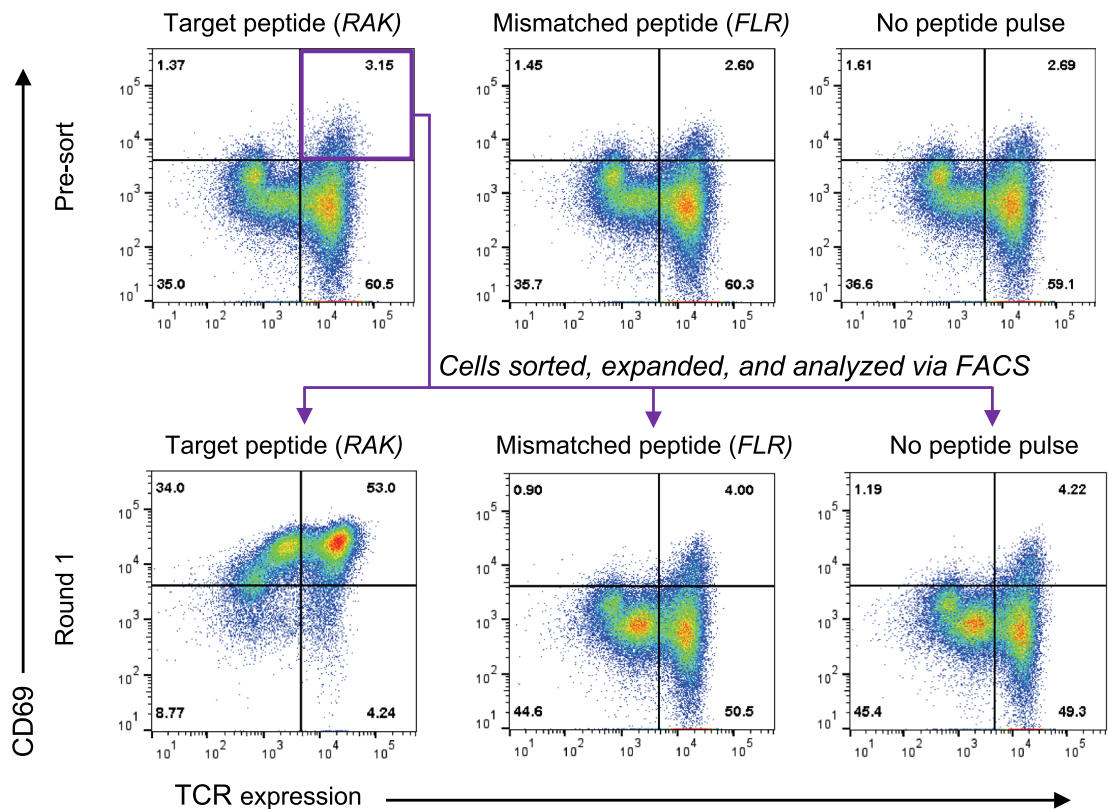
Figure 5. Cell activation-based screening of antigen-specific TCRs. (**A**) Activation of SKW3 cells expressing the LC13 TCR was measured after coculture with *FLR*-pulsed APCs (left), *RAK*-pulsed APCs (center), or unpulsed APCs (right). (**B**) TCRα:β-SKW3 cells from the Donor 1 library were enriched for CD69 expression via FACS after coculture with *RAK*-pulsed APCs (left). Peptide specificity was confirmed using *FLR*-pulsed APCs (center) and unpulsed APCs (right). TCR expression is shown on the x-axis, and CD69 expression is shown on the y-axis.

established methods for single-cell isolation[15,29,50–53] and microfluidic encapsulation[16,54]. Moreover, our libraries were genetically diverse, with each containing an average 17,241 unique clones after stringent quality filtering and clustering of homologous TCRs.

Importantly, our approach enabled the facile cloning of physically linked TCRα:β genes into various display systems, including Jurkat cells, which can be screened using multimeric pMHCs[34], and SKW3 cells, which can be screened using the activation-based method reported here. Such immortalized/renewable TCRα:β library screening techniques will be essential for the analysis of peptide specificity, MHC restriction, and the biophysical

properties of the corresponding TCRs, which ultimately govern T cell behavior in vivo. However, we found it difficult to barcode cells for the inclusion of singleton TCRα:β genes, which compromised our ability to calculate the efficiency of recovery via HTS. We are currently planning to resolve these issues by incorporating single-cell barcoding techniques and to extend the scope of our work by recovering information on gene transcription and protein expression in association with individual TCRs.

Importantly, our SKW3 cell expression system allowed us to screen TCRα:β gene libraries functionally, measuring responsiveness via the upregulation of CD69. No such activation was observed using Jurkat cell lines in our earlier study[34]. This advance is critically important for immune discovery, because the ability to deliver an activation signal is not equivalent across all antigen-specific TCRs[33]. It can also be difficult in some cases to produce large quantities of soluble pMHCs[21,29,55]. A key feature of our approach was the standardized assessment of functionality using an immortalized cell line. Library screening on this basis can eliminate potential bias arising from the heterogeneity of primary T cells, although it should be noted that other effector readouts may also afford high levels of sensitivity[56,57]. In addition, our method could be adapted for use with cancerous or infected cells rather than peptide-pulsed APCs. Accordingly, it should be feasible to screen for reactivity against target cells presenting biologically relevant densities of disease-associated pMHCs, thereby enhancing the discovery of protective and translationally efficacious TCRs.

There were some limitations to our approach. In particular, the gene capture and sequencing process revealed that repertoire diversity was reduced to around tens of thousands of TCRα:β clonal clusters after excluding singletons in each library, which are more error-prone than TCRα:β clones observed more than once). The degree of loss from the native repertoire was difficult to quantify, because our native immune libraries contained an undetermined number of TCRα:β clones, and our conservative bioinformatic filtering excluded singletons that could represent *bona fide* TCR clones. The natural occurrence of restriction enzyme sites used in the cloning process also likely resulted in the destruction of a small fraction of TCRs (estimated based on prevalence at < 1%). It is further notable that we transduced our TCRα:β libraries into SKW3 cells using viral particles at a very low exposure frequency, estimated in the region of 1–5%. This approach was designed to ensure that each cell integrated only one TCR. Our future efforts will focus on the use of CRISPR-targeted TCRα:β engineering to improve efficacy by integrating TCR genes at defined sites, which was recently demonstrated to be highly effective in similar activation-based screening assays for the analysis of libraries displaying synthetically generated TCRs[58].

In summary, we have developed a high-throughput platform for the identification of functionally responsive antigen-specific TCRs. Our workflow builds substantially on previous reports that linked somatically rearranged gene sequences with antigen specificity via library screening against soluble pMHCs[34]. In particular, we anticipate that molecular-scale functional screening will accelerate bench-to-bedside immune discovery, facilitating the clinical delivery of personalized therapies for various diseases via the rapid isolation of naturally selected and highly potent antigen-specific TCRs.

Methods

Introduction of restriction enzyme cloning sites. A monoclonal TCR (TM9) specific for the HLA-B*07:02-restricted HIV-1 Nef epitope *RM9*[59] was expressed in the lentiviral vector pLVX-EF1α-IRES-mCherry (Takara Bio, Mountain View, CA) to evaluate the functional performance of restriction enzyme cloning site mutations (Supplementary Tables 1 and 2). Leader sequences were modified for TRAV and TRBV. Restriction enzyme sites were introduced as detailed in Supplementary Tables 1 and 2. Expression of the TM9 TCR was quantified via flow cytometry using anti-human TCRα/β–Alexa Fluor 488 (clone IP26; BioLegend, San Diego, CA) and fluorescently labeled tetrameric complexes of *RM9*/HLA-B*07:02.

Human samples and cell culture. Donor 1 presented with high fever, fatigue, body aches, and headache, with a maximum illness severity of 3, as described previously[34,35]. Donor 2 presented with fever, tender cervical lymph nodes, sore throat, and fatigue. These donors were enrolled in a prospective study of primary EBV infection at the University of Minnesota (IRB 0608M90593)[35]. Venous blood samples were processed via density gradient centrifugation over ACCUSPIN System-Histopaque-1077 (Sigma-Aldrich, St. Louis, MO) to collect PBMCs, which were subsequently cryopreserved at 1×10^7 cells/mL in heat-inactivated fetal bovine serum (Thermo Fisher Scientific, Waltham, MA) containing 10% dimethyl sulfoxide (Sigma-Aldrich, St. Louis, MO). PBMCs were thawed and density-adjusted to 0.5×10^6 cells/mL in complete CTS OpTmizer T Cell Expansion SFM (Thermo Fisher Scientific, Waltham, MA) supplemented with 5% CTS Immune Cell SR (Thermo Fisher Scientific, Waltham, MA), 200 IU/mL IL-2 (National Cancer Institute Preclinical Biologics Repository, Frederick, MD), and 25 µL/mL ImmunoCult Human CD3/CD28 T Cell Activator (STEMCELL Technologies, Cambridge, MA). Cells were expanded in RPMI 1640 medium (Thermo Fisher Scientific, Waltham, MA) containing 10% heat-inactivated fetal bovine serum (Thermo Fisher Scientific, Waltham, MA), 200 IU/mL IL-2 (National Cancer Institute Preclinical Biologics Repository, Frederick, MD), and 25 µL/mL ImmunoCult Human CD3/CD28 T Cell Activator (STEMCELL Technologies, Cambridge, MA) for 7–10 days and then subjected to single-cell emulsification and overlap extension RT-PCR.

Generation of natively paired TCRα:β expression libraries. TCRα:β cDNA libraries generated in a previous study[34] were amplified and modified to incorporate restriction enzyme sites using a two-step, semi-nested PCR. An initial semi-nested suppression PCR incorporating blocking oligonucleotides complementary to the 3' ends of the unfused TCRα and TCRβ products was performed using a HotStart GoTaq Polymerase System (Promega, Madison, WI). A second semi-nested PCR was then performed using a KAPA HiFi HotStart PCR Kit (Roche, Basel, Switzerland). PCR products were recovered using agarose gel electrophoresis and purified using a 1.5% SYBR Safe Agarose Gel (Thermo Fisher Scientific, Waltham, MA).

PCR products were cloned into a modified version of the commercially available pLVX-EF1α-IRES-mCherry Vector (Takara Bio, Mountain View, CA). TCR amplicons and the expression vector were digested with BstBI and AgeI, and the digestion products were gel-purified and ligated using T4 DNA Ligase (New England Biolabs, Ipswich, MA). Ligation products were purified using a DNA Clean & Concentrator Kit (Zymo Research, Irvine, CA) and transformed via electroporation into competent MegaX DH10B T1 Electrocomp Cells (Thermo Fisher Scientific, Waltham, MA). Plasmids were purified using a ZymoPURE II Plasmid Maxiprep Kit (Zymo Research, Irvine, CA). The rest of the expression cassette was then introduced as an insert between the variable regions of the TCRβ and TCRα genes using SpeI and MluI (New England Biolabs, Ipswich, MA). The insert contained the remaining portion of the TCRβ constant region, a P2A translation skip motif, and a modified version of the TCRα leader peptide sequence containing an MluI site (Fig. 2) to enable full expression of the corresponding heterodimeric TCRs[34].

Lentiviral transduction of SKW3 cells.   Lentiviruses were packaged using an adherent 293FT cell line (Thermo Fisher Scientific, Waltham, MA). Cells were grown to a confluency of 70–90% in T75 culture flasks (Thermo Fisher Scientific, Waltham, MA) and transfected using Lipofectamine 3000 Transfection Reagent (Thermo Fisher Scientific, Waltham, MA). A master mix was prepared for each library by incubating 8 µg of PSPAX (Addgene, Watertown, MA), 2 µg of PMD2G (Addgene, Watertown, MA), 40 µL of P3000 Reagent (Thermo Fisher Scientific, Waltham, MA), and 12 µg of the pLVX-EF1α-IRES-mCherry plasmid (Takara Bio, Mountain View, CA) for 5 min in 1 mL of Opti-MEM Reduced Serum Medium per T75 flask (Thermo Fisher Scientific, Waltham, MA). Lipofectamine 3000 Transfection Reagent was diluted 1:25 in 1 mL of Opti-MEM Reduced Serum Medium (Thermo Fisher Scientific, Waltham, MA). The diluted transfection reagent was then added to the master mix and incubated for 15–20 min at room temperature. The resultant mixture was added dropwise to the flask containing 293FT cells and incubated for 3 days at 37 °C. Supernatants containing lentivirus were centrifuged at $800 \times g$ for 10 min and added at a 3:1 ratio to Lenti-X Concentrator (Takara Bio, Mountain View, CA) in polypropylene centrifuge tubes (50 mL; Thermo Fisher Scientific, Waltham, MA). Lentivirus-concentrator mixtures were incubated overnight at 4 °C and then centrifuged at $1,500 \times g$ for 45 min at 4 °C. The pellets were resuspended in 1 mL of RPMI 1640 medium and stored in two aliquots at −80 °C. For each library transduction, 3×10^6 unmodified SKW3 cells (DSMZ, Braunschweig, Germany) were seeded in 3 mL of RPMI 1640 medium in a single well of a 6-well tissue culture plate (Thermo Fisher Scientific, Waltham, MA), and 400 µL of rapidly thawed lentivirus stock was added in the presence of polybrene at a final concentration of 4 µg/mL (Sigma-Aldrich, St. Louis, MO). Cells were incubated overnight at 37 °C with 5% CO_2. After transduction, cells were centrifuged at $500 \times g$ for 5 min, resuspended in 10 mL of prewarmed RPMI 1640 medium, transferred to T25 culture flasks, and incubated for 3 days at 37 °C with 5% CO_2. Cells were then washed twice with PBS and sorted for internal mCherry expression via FACS.

Functional screening assays in SKW3 cells.   TCRα:β-SKW3 libraries were cocultured with peptide-pulsed APCs matched to the relevant HLA. Purified mCherry⁺ TCRα:β-SKW3 cells were expanded in vitro, resuspended in fresh RPMI 1640 medium, and seeded at 1×10^6 cells/well/mL in 6-well tissue culture plates (Thermo Fisher Scientific, Waltham, MA). APCs were washed twice with PBS, resuspended at 1×10^6 cells per condition in 100 µL of FACS buffer (PBS containing 0.05% BSA and 2 mM EDTA), and pulsed with the relevant peptides for 4 h at 37 °C. T2-B8 cells (a kind gift from Scott Burrows, QIMR Berghofer) were used to present epitopes restricted by HLA-B*08:01. APCs were then incubated with individual TCRα:β-SKW3 libraries overnight at 37 °C with 5% CO_2. TCRα:β-SKW3 cells were recovered after stimulation and stained with anti-CD69–FITC (clone FN50; BioLegend, San Diego, CA).

PCR-based recovery of TCRβ genes.   CD69⁺ TCRα:β-SKW3 cells were sorted via FACS. For molecular analysis of each library, mRNA was extracted from 2×10^6 purified TCRα:β-SKW3 cells using a Direct-zol RNA Kit (Zymo Research, Irvine, CA). TCRβ VDJ regions were amplified using a set of primers targeting the modified TRBV15-1 leader region and TRBC. RT-PCR was performed using SuperScript III Reverse Transcriptase and Platinum Taq (Thermo Fisher Scientific, Waltham, MA). A second primer-extension PCR contained adaptors to add a unique molecular identifier to each sample. cDNA amplicons were run on 1.5% agarose gels, and bands at ~450 bp were purified using a Gel DNA Recovery and DNA Clean & Concentrator Kit (Zymo Research, Irvine, CA). Libraries were sequenced using a 2×300 MiSeq System (Illumina, San Diego, CA).

Bioinformatic analysis.   Our bioinformatics pipeline was designed to ensure high-quality data by reducing sequence errors introduced via HTS and/or PCR[34,39–44]. First, raw sequences were quality-filtered to retain only those with a Phred Quality Score of 20 (i.e., 99% base call accuracy) in at least 50% of the reads using Fastx-toolkit/0.0.14 (http://hannonlab.cshl.edu/fastx_toolkit/). Next, V, D, and J gene annotations were performed using MiXCR/v2.1.12[49]. Out-of-frame V(D)J combinations were excluded from the dataset, and productive in-frame junction sequences were paired by Illumina read ID. Reads were then compiled based on CDR3 nucleotide exact-match sequences and V(D)J gene identity. CDR3β sequences were clustered to 96% nucleotide identity (ignoring terminal gaps) using USEARCH/v5.2.32[60], and only clusters with ≥ 2 TCRα:β nucleotide reads were included in the final dataset. Last, full-length TCRα:β sequences were recreated by stitching together the CDR3α or CDR3β sequences with the respective TRAV or TRBV genes, which had been mapped using the International ImMunoGeneTics (IMGT) Information System database (library imgt.202141-1)[61].

Ethics statement. Human subjects research was approved by the University of Minnesota Institutional Review Board. Informed consent was obtained from all donors with permission to use samples for research investigations according to the principles of the Declaration of Helsinki.

Data availability

References

1. Ishizuka, J. *et al.* Quantitating T cell cross-reactivity for unrelated peptide antigens. *J. Immunol.* **183**, 4337–4345 (2009).
2. Rossjohn, J. *et al.* T cell antigen receptor recognition of antigen-presenting molecules. *Annu. Rev. Immunol.* **33**, 169–200 (2015).
3. Miles, J. J., Douek, D. C. & Price, D. A. Bias in the αβ T-cell repertoire: implications for disease pathogenesis and vaccination. *Immunol. Cell Biol.* **89**, 375–387 (2011).
4. Gaud, G., Lesourne, R. & Love, P. E. Regulatory mechanisms in T cell receptor signalling. *Nat. Rev. Immunol.* **18**, 485–497 (2018).
5. 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).
6. Nikolich-Žugich, J., Slifka, M. K. & Messaoudi, I. The many important facets of T-cell repertoire diversity. *Nat. Rev. Immunol.* **4**, 123–132 (2004).
7. Pai, J. A. & Satpathy, A. T. High-throughput and single-cell T cell receptor sequencing technologies. *Nat. Methods* **18**, 881–892 (2021).
8. 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).
9. Howie, B. *et al.* High-throughput pairing of T cell receptor α and β sequences. *Sci. Transl. Med.* **7**, 301ra131 (2015).
10. Turchaninova, M. A. *et al.* Pairing of T-cell receptor chains via emulsion PCR. *Eur. J. Immunol.* **43**, 2507–2515 (2013).
11. Robins, H. S. *et al.* Comprehensive assessment of T-cell receptor β-chain diversity in αβ T cells. *Blood* **114**, 4099–4107 (2009).
12. Dash, P., Wang, G. C. & Thomas, P. G. Single cell analysis of T-cell receptor αβ repertoire. *Methods Mol. Biol.* **1343**, 181–197 (2015).
13. Sant, S. *et al.* Single-cell approach to influenza-specific CD8+ T cell receptor repertoires across different age groups, tissues, and following influenza virus infection. *Front Immunol.* **9**, 1453 (2018).
14. Ludwig, J., Huber, A., Bartsch, I., Busse, C. E. & Wardemann, H. High-throughput single-cell sequencing of paired TCRα and TCRβ genes for the direct expression-cloning and functional analysis of murine T-cell receptors. *Eur. J. Immunol.* **49**, 1269–1277 (2019).
15. Wahl, I. *et al.* An efficient single-cell based method for linking human T cell phenotype to T cell receptor sequence and specificity. *Eur. J. Immunol.* **52**, 237–246 (2021).
16. Azizi, E. *et al.* Single-cell map of diverse immune phenotypes in the breast tumor microenvironment. *Cell* **174**, 1293–1308.e36 (2018).
17. Tanno, H. *et al.* A facile technology for the high-throughput sequencing of the paired VH:VL and TCRβ:TCRα repertoires. *Sci. Adv.* **6**, eaay9093 (2020).
18. Rodenko, B. *et al.* Generation of peptide–MHC class I complexes through UV-mediated ligand exchange. *Nat. Protoc.* **1**, 1120–1132 (2006).
19. Dolton, G. *et al.* Optimized peptide–MHC multimer protocols for detection and isolation of autoimmune T-cells. *Front. Immunol.* **9**, 1378 (2018).
20. Bethune, M. T. *et al.* Isolation and characterization of NY-ESO-1–specific T cell receptors restricted on various MHC molecules. *Proc. Natl. Acad. Sci. USA* **115**, E10702–E10711 (2018).
21. Spindler, M. J. *et al.* Massively parallel interrogation and mining of natively paired human TCRαβ repertoires. *Nat. Biotechnol.* **38**, 609–619 (2020).
22. Dileepan, T. *et al.* MHC class II tetramers engineered for enhanced binding to CD4 improve detection of antigen-specific T cells. *Nat. Biotechnol.* **39**, 943–948 (2021).
23. Franco, F., Jaccard, A., Romero, P., Yu, Y. R. & Ho, P. C. Metabolic and epigenetic regulation of T-cell exhaustion. *Nat. Metab.* **2**, 1001–1012 (2020).
24. Scheper, W. *et al.* Low and variable tumor reactivity of the intratumoral TCR repertoire in human cancers. *Nat. Med.* **25**, 89–94 (2019).
25. Yossef, R. *et al.* Enhanced detection of neoantigen-reactive T cells targeting unique and shared oncogenes for personalized cancer immunotherapy. *JCI Insight* **3**, e122467 (2018).
26. Hu, Z. *et al.* A cloning and expression system to probe T-cell receptor specificity and assess functional avidity to neoantigens. *Blood* **132**, 1911–1921 (2018).
27. Stevanović, S. *et al.* A phase II study of tumor-infiltrating lymphocyte therapy for human papillomavirus-associated epithelial cancers. *Clin. Cancer Res.* **25**, 1486–1493 (2019).
28. Doran, S. L. *et al.* T-cell receptor gene therapy for human papillomavirus-associated epithelial cancers: a first-in-human, phase I/II study. *J. Clin. Oncol.* **37**, 2759–2768 (2019).
29. Linnemann, C. *et al.* High-throughput identification of antigen-specific TCRs by TCR gene capture. *Nat. Med.* **19**, 1534–1541 (2013).
30. Li, Y. *et al.* Directed evolution of human T-cell receptors with picomolar affinities by phage display. *Nat. Biotechnol.* **23**, 349–354 (2005).
31. Linette, G. P. *et al.* Cardiovascular toxicity and titin cross-reactivity of affinity-enhanced T cells in myeloma and melanoma. *Blood* **122**, 863–871 (2013).
32. Border, E. C., Sanderson, J. P., Weissensteiner, T., Gerry, A. B. & Pumphrey, N. J. Affinity-enhanced T-cell receptors for adoptive T-cell therapy targeting MAGE-A10: strategy for selection of an optimal candidate. *OncoImmunology* **8**, e1532759 (2019).
33. Sibener, L. V. *et al.* Isolation of a structural mechanism for uncoupling T cell receptor signaling from peptide-MHC binding. *Cell* **174**, 672–687.e27 (2018).
34. Fahad, A. S. *et al.* Immortalization and functional screening of natively paired human T cell receptor repertoires. *Protein Eng. Des. Sel.* **35**, gzab034 (2022).
35. Balfour, H. H. Jr., *et al.* Behavioral, virologic, and immunologic factors associated with acquisition and severity of primary Epstein-Barr virus infection in university students. *J. Infect. Dis* **207**, 80–88 (2013).
36. Balfour, H. H. Jr., Dunmire, S. K. & Hogquist, K. A. Infectious mononucleosis. *Clin. Transl. Immunol.* **4**, e33 (2015).

37. Mayassi, T. *et al.* Chronic inflammation permanently reshapes tissue-resident immunity in celiac disease. *Cell* **176**, 967–981.e19 (2019).
38. Nguyen, T. H. O. *et al.* Recognition of distinct cross-reactive virus-specific CD8[+] T cells reveals a unique TCR signature in a clinical setting. *J. Immunol.* **192**, 5039–5049 (2014).
39. 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).
40. DeKosky, B. J. *et al.* Large-scale sequence and structural comparisons of human naive and antigen-experienced antibody repertoires. *Proc. Natl. Acad. Sci. USA* **113**, E2636–E2645 (2016).
41. 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).
42. Fahad, A. S. *et al.* Functional profiling of antibody immune repertoires in convalescent Zika virus disease patients. *Front. Immunol.* **12**, 615102 (2021).
43. Wang, B. *et al.* Functional interrogation and mining of natively paired human VH:VL antibody repertoires. *Nat. Biotechnol.* **36**, 152–155 (2018).
44. DeKosky, B. J. *et al.* High-throughput sequencing of the paired human immunoglobulin heavy and light chain repertoire. *Nat. Biotechnol.* **31**, 166–169 (2013).
45. Banach, B. B. *et al.* Paired heavy- and light-chain signatures contribute to potent SARS-CoV-2 neutralization in public antibody responses. *Cell Rep.* **37**, 109771 (2021).
46. Boria, I., Cotella, D., Dianzani, I., Santoro, C. & Sblattero, D. Primer sets for cloning the human repertoire of T cell receptor variable regions. *BMC Immunol.* **9**, 50 (2008).
47. Liu, Z. *et al.* Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector. *Sci. Rep.* **7**, 2193 (2017).
48. Argaet, V. P. *et al.* Dominant selection of an invariant T cell antigen receptor in response to persistent infection by Epstein-Barr virus. *J. Exp. Med.* **180**, 2335–2340 (1994).
49. Bolotin, D. A. *et al.* MiXCR: software for comprehensive adaptive immunity profiling. *Nat. Methods* **12**, 380–381 (2015).
50. Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* **11**, 783–784 (2014).
51. Cole, D. K. *et al.* Germ line-governed recognition of a cancer epitope by an immunodominant human T-cell receptor. *J. Biol. Chem.* **284**, 27281–27289 (2009).
52. Trautmann, L. *et al.* Selection of T cell clones expressing high-affinity public TCRs within human cytomegalovirus-specific CD8 T cell responses. *J. Immunol.* **175**, 6123–6132 (2005).
53. Day, E. K. *et al.* Rapid CD8[+] T cell repertoire focusing and selection of high-affinity clones into memory following primary infection with a persistent human virus: human cytomegalovirus. *J. Immunol.* **179**, 3203–3213 (2007).
54. Mereu, E. *et al.* Benchmarking single-cell RNA-sequencing protocols for cell atlas projects. *Nat. Biotechnol.* **38**, 747–755 (2020).
55. 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).
56. Nesterenko, P. A. *et al.* Droplet-based mRNA sequencing of fixed and permeabilized cells by CLInt-seq allows for antigen-specific TCR cloning. *Proc. Natl. Acad. Sci. USA* **118**, e2021190118 (2021).
57. Bos, R., Marquardt, K. L., Cheung, J. & Sherman, L. A. Functional differences between low- and high-affinity CD8[+] T cells in the tumor environment. *OncoImmunology* **1**, 1239–1247 (2012).
58. Vazquez-Lombardi, R. *et al.* High-throughput T cell receptor engineering by functional screening identifies candidates with enhanced potency and specificity. *Immunity* **55**, 1953-1966.e10 (2022).
59. Kløverpris, H. N. *et al.* A molecular switch in immunodominant HIV-1-specific CD8 T-cell epitopes shapes differential HLA-restricted escape. *Retrovirology* **12**, 20 (2015).
60. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).
61. Lefranc, M. P. *et al.* IMGT®, the international ImMunoGeneTics information system® 25 years on. *Nucleic Acids Res.* **43**, D413–D422 (2015).

Acknowledgements

Author contributions

A.S.F., C.Y.C., D.C.D., H.H.B., D.A.P., and B.J.D. designed experiments; A.S.F., C.Y.C., S.N.L.A., N.B., B.M., M.F.G.-G., R.M.-N., A.D.L., R.R.L., J.Z., J.W., and S.L.-L. performed experiments; R.A.K. provided critical resources; A.S.F., C.Y.C., and B.J.D. analyzed data; A.S.F., C.Y.C., D.A.P., and B.J.D. wrote the manuscript. All authors commented on the manuscript, contributed intellectually, and approved the submitted version.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-023-31858-4.

Correspondence and requests for materials should be addressed to B.J.D.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.