Massively parallel interrogation and mining of natively paired human TCRαβ repertoires

Matthew J. Spindler¹, Ayla L. Nelson¹, Ellen K. Wagner¹, Natasha Oppermans², John S. Bridgeman³, James M. Heather⁴, Adam S. Adler⁵, Michael A. Asensio⁶, Robert C. Edgar¹, Yoong Wearn Lim¹, Everett H. Meyer⁵,⁶, Robert E. Hawkins²,³, Mark Cobbold⁴,⁷ and David S. Johnson⁰¹,²

T cells engineered to express antigen-specific T cell receptors (TCRs) are potent therapies for viral infections and cancer. However, efficient identification of clinical candidate TCRs is complicated by the size and complexity of T cell repertoires and the challenges of working with primary T cells. Here we present a high-throughput method to identify TCRs with high functional avidity from diverse human T cell repertoires. The approach used massively parallel microfluidics to generate libraries of natively paired, full-length TCRαβ clones, from millions of primary T cells, which were then expressed in Jurkat cells. The TCRαβ–Jurkat libraries enabled repeated screening and panning for antigen-reactive TCRs using peptide major histocompatibility complex binding and cellular activation. We captured more than 2.9 million natively paired TCRαβ clonotypes from six healthy human donors and identified rare (<0.001% frequency) viral-antigen-reactive TCRs. We also mined a tumor-infiltrating lymphocyte sample from a patient with melanoma and identified several tumor-specific TCRs, which, after expression in primary T cells, led to tumor cell killing.
expression library. The TCRαβ–Jurkat libraries were repeatedly screened and enriched for antigen reactivity by pMHC binding as well as by TCR-mediated cellular activation.

Results

Analysis of diverse, natively paired TCRαβ repertoires. Human TCRαβ repertoires comprise millions of single cells that express different TCRα and TCRβ sequences. To capture this diversity, we adapted our existing microfluidic technology to isolate millions of single T cells into droplets in 1 h (Supplementary Fig. 1). T cells were lysed within the droplets, and mRNA from single cells was captured on oligo(dT) beads. The mRNA-bound beads were reattached into droplets containing reagents for overlap extension RT–PCR, such that the TCRα and TCRβ transcripts were amplified and physically linked into a single TCRαβ amplon (Supplementary Fig. 2). The physically linked TCRαβ amplon libraries were further processed for deep sequencing and cloning into expression constructs (Supplementary Fig. 3).

We used our technology to generate natively paired TCRαβ libraries for three HLA-A*02:01+ and three HLA-A*24:01+ human donors who were seropositive for cytomegalovirus (CMV) and Epstein–Barr virus (EBV) (Supplementary Table 1). We ran 2.7–3.4 million T cells through the microfluidic system for each donor and processed the TCRαβ amplons for deep sequencing (Illumina) and cloning into the lentiviral expression system.

We analyzed the six linked TCRαβ repertoires with deep sequencing (Supplementary Table 1). Illumina sequencing errors can be difficult to differentiate from bona fide TCRαβ sequence diversity, so we applied conservative informatics to eliminate low-quality sequence reads. We defined a TCRαβ clonotype as a unique combination of CDR3 α and CDR3 β amino acid sequences with specific V- and J-gene usage. The libraries contained an average of 488,601 clonotypes (range, 339,707–711,589), for a total of 2,931,604 TCRαβ clonotypes, which comprehensively covers the diversity of possible TRAV–TRBV gene combinations (Fig. 2a). The TCRαβ libraries were highly diverse, with the top 20 clonotypes accounting for 10.77% of the total sequencing reads on average (range, 8.48–14.62%) and a long tail of low-frequency clonotypes. The average median clonotype read frequency was 0.000021% across our six libraries (Supplementary Figs. 4 and 5).

To characterize proper TCRα–TCRβ pairing within our linked TCRαβ libraries, we spiked Jurkat cells into primary T cells from healthy donors at a frequency of 0.1% or 0.001% and ran the samples through our microfluidic platform. We found that 92.9% and 98.6% of Jurkat–TCRα chains paired with Jurkat–TCRβ chains for the 0.1% and 0.001% mixtures, respectively.

To broadly assess the TCRα–TCRβ pairing fidelity generated by our microfluidic system, we compared the number of unique TCRβ sequences paired to known invariant and noninvariant TCRα sequences within our linked TCRαβ libraries (Fig. 2b and
Supplementary Fig. 6). Invariant TCRα sequences are expressed by specialized T cell populations, such as natural killer T cells and mucosal-associated invariant T cells\cite{28,29}. These T cell populations express a constrained set of invariant TCRα CDR3 sequences, such that repertoire diversity is driven by the TCRβ chain\cite{30,31}. Previously, single-cell analysis demonstrated that identical TCRα CDR3
sequences pair with numerous TCRβ sequences. We therefore reasoned that known invariant TCRα sequences would pair with a variety of TCRβ sequences, whereas noninvariant TCRα sequences would generally pair with only one TCRβ, providing a method to quantify our TCRαβ pairing fidelity. In agreement with this, we observed that a high proportion (81.45%) of invariant TCRα sequences target stage

CSS-930

Dextramer Activation

CSS-938

Target stage

CSS-944

Target stage

CSS-948

Target stage

A2/pp65 (NLV)

A2/IE-1 (VLE)

A2/BMLF1 (GLC)

A2/LMP2 (CLG)

Read frequency (%) 100 80 60 40 20 0

A2/pp65 (NLV) A2/IE-1 (VLE) A2/BMLF1 (GLC) A2/LMP2 (CLG) Irrelevant control

Antiviral panning of CSS-930 library

Round 1 Round 2 Round 3 Round 4 Round 5

Irrelevant Target

A2/pp65(NLV)

A2/IE-1(VLE)

A2/BMLF1(GLC)

A2/LMP2(CLG)

Dextramer-APC

CD3-FITC

Dextramer stage

Input Round 1 Round 2 Round 3 Round 4 Round 5

Activation stage

Input Activated

Cellular activation

CD69-PE

CD62L-APC

Enrichment of antiviral TCRβ sequences

Target stage

Activation

CSS-930

CSS-938

CSS-944

CSS-948

Fig. 3 | Identification of virus-specific TCRαβ clones by library panning and cellular activation screens. a, Panning the donor CSS-930 TCRαβ–Jurkat library with pMHC dextramers for four different viral antigens: A2/pp65(NLV), A2/IE-1(VLE), A2/BMLF1(GLC) and A2/LMP2(CLG). pMHC dextramer signal is on the x axis and CD3 staining is on the y axis. The inset indicates the gate used to sort for pMHC dextramer-positive cells. Panning experiments were completed one time. b, Cellular activation of the fully pMHC dextramer-enriched cell populations co-cultured with peptide-pulsed T2 cells. T2 cells were loaded with four different viral antigens, right column (target), from top to bottom: pp65(NLV), IE-1(VLE), BMLF1(GLC) and LMP2(CLG). The left column comprises irrelevant peptide targets loaded onto T2 cells. CD69 staining is on the x axis and CD62L staining is on the y axis. Cellular activation experiments were independently repeated twice with similar results. c, Enrichment of antiviral TCRβ sequences for donors CSS-930, CSS-938, CSS-944 and CSS-948. The legend indicates color coding for each target, pMHC dextramer panning and cellular activation stage. TCRβ frequencies are indicated by a heat map, from light to dark blue. TCRβ clonotypes labeled with a star and in bold font were engineered into monoclonal Jurkat cell lines.
sequences paired with more than one TCRβ, whereas a significantly smaller fraction (18.90%; one-sided binomial proportion test, z statistic, $P < 2.2 \times 10^{-16}$) of noninvariant TCRα sequences paired with more than one TCRβ (Fig. 2b and Supplementary Fig. 6). Similarly, most TCRβ sequences (79.88%) paired with a single TCRα. These data suggest that our cloned TCR libraries retained high TCRαβ pairing fidelity. The TCRβ and noninvariant TCRα sequences that were found to pair to multiple partners could represent natural features of these T cell repertoires or could have resulted from mispairing during TCRβ linkage owing to multiple cells in droplets, PCR-based artifacts or other technical errors.

Recent work using single-chain TCR sequencing methods has shown considerable overlap between TCR repertoires from unrelated individuals, both in terms of shared TCR sequences and common sequence motifs. Globally, we observed very little overlap between donors’ TCRαβ repertoires, with only one TCRαβ clonotype shared among 5 donors and 12 shared among at least 4 donors (Fig. 2c). No clear global distinctions were observed between HLA-A*02:01 and HLA-A*24:02 genotypes, even though all six donors were positive for CMV and EBV and therefore were likely to have some antiviral T cells in the periphery. Notably, we observed higher cross-donor sharing of unlinked TCRs, with 1,735 TCRα and 17 TCRβ sequences shared among at least 5 repertoires and 4,866 TCRα and 128 TCRβ sequences shared among at least 4 repertoires (Supplementary Figs. 7–9). This suggests that conventional single-chain TCR sequencing methods might overestimate functional repertoire overlap between unrelated individuals.

Generation of natively paired TCRαβ–Jurkat expression libraries. To understand how well our method captured TCR diversity, we compared the TCRβ clonotype diversity present in our natively paired TCRαβ amplicon libraries to the TCRβ clonotype diversity obtained using single-chain TCRβ amplification of RNA from a matched aliquot of primary T cells (Supplementary Table 2). We controlled for the number of input cells and sequenced to a similar depth (average of 5.4 million and 5.0 million reads from single-chain TCRβ and natively paired TCRαβ libraries, respectively) across all samples. We identified an average of 303,599 unique TCRβ clonotypes from single-chain amplification and 210,249 TCRβ clonotypes from the paired TCRαβ libraries, suggesting a clonotype yield of about 71%. Furthermore, the natively paired TCRαβ libraries captured, on average, 97% of the top 100 clonotypes and 91% of the top 250 clonotypes identified in the single-chain TCRβ libraries (Supplementary Fig. 10). These data confirmed that our natively paired TCRαβ libraries efficiently captured the clonotypes present in the primary T cell population, and, as expected, the rate of clonotype overlap between single-chain TCRβ and paired TCRαβ libraries decreased with decreasing clonotype frequency.

We next developed a workflow to clone these natively paired TCRαβ amplicon libraries into expression libraries. An important innovation was to optimize the Gibson assembly process for cloning the libraries of TCRαβ amplicons en masse into a lentiviral cell expression system also provided the ability to test these donors with fluorescently labeled pMHC multimers and activation by artificial antigen-presenting cells (aAPCs).

To improve Jurkat TCR-mediated cellular activation in response to HLA class I-presented peptides, we engineered a TCRβ-deficient (ΔTCRβ) Jurkat cell line (J.RT3-T3.5) to express human CD8 (Supplementary Fig. 11). We then transduced the six full-length TCRβ expression libraries into the CD8+ ΔTCRβ Jurkat cells at low infection rates (estimated 10–20%) to obtain single integration events. We obtained CD3+ TCRβ+ surface expression in about 10% of preselected cells, which increased to 42–56% after selection for stable lentiviral integration (Supplementary Fig. 12). Assuming a Poisson probability distribution, we calculated that a 20% transduction efficiency should result in about 90% of our selected cells expressing a single TCRβ. To determine the actual frequency of cells expressing single TCRs, we sorted single CD3+ TCRβ+ cells from two postselected TCRβ–Jurkat libraries and amplified the expressed TCRβ mRNA. This analysis confirmed that more than 80% of the TCRβ–Jurkat library cells expressed a single TCR (Supplementary Fig. 13).

Finally, we confirmed that the TCRβ–Jurkat expression library build maintained the clonotype diversity present in the starting natively paired TCRαβ amplicon libraries. Deep TCRβ sequencing demonstrated that the various molecular cloning steps and the fully selected TCRβ–Jurkat libraries were highly correlated with their respective starting TCRαβ amplicon library, with Pearson correlation coefficients ranging from 0.77 to 0.97 (Supplementary Figs. 14 and 15). In addition, more than 95% of the 100 most frequent TCRβ clonotypes were maintained through the full process (Supplementary Fig. 16). In total, the six healthy donor TCRαβ–Jurkat libraries expressed over 1 million TCRβ clonotypes, with an average of 186,117 unique TCRβ clonotypes per library (Supplementary Table 2).

Identification of rare antiviral TCRs. Antigen-specific T cells are conventionally identified by staining primary T cells with pMHC multimers, but any further functional analysis requires single-cell cloning. Unlike primary T cells, Jurkat cells are easy to culture and expand indefinitely, enabling application of a ‘panning’ approach. In this method, TCRβ–Jurkat libraries are stained with dextramer and sorted by flow cytometry. These ‘round 1 dextramer’ cells are expanded, restained with dextramer and sorted a second time (‘round 2 dextramer’); this cycle can be repeated as necessary to enrich for rare antigen-specific clones. To test the method, we first performed control experiments with a previously reported HLA-A*02:01/MART-1(ELA)-reactive TCRβ clone, MEL5 (ref. 13), spiked into the donor CSS-930 TCRβ–Jurkat library at a frequency range of 0.001–1%. Panning recovered the clone quantitatively at all levels tested (Supplementary Figs. 17 and 18).

To identify virus-specific TCRβ clones, we panned the six TCRβ–Jurkat libraries with fluorescently labeled pMHC dextramers against eight well-characterized antigens from CMV and EBV (Supplementary Table 3). Multiple rounds of panning enriched for dextramer-binding cells to four HLA-A*02:01 targets and one HLA-A*24:02 target across our donor libraries (Fig. 3a and Supplementary Figs. 19–21). Two of the HLA-A*24:02 donors yielded no binders. Dextramer-binding cells were rare (<0.1%) in the starting TCRβ–Jurkat libraries and increased to 47–98% with panning. The Jurkat cell expression system also provided the ability to test these dextramer-enriched TCRs for functional activity by staining for activation markers CD69 and CD62L (nonactivated cells are CD69+CD62L+ and activated cells are CD69−CD62L−). We used an aAPC activation assay with peptide-pulsed T2 cells to functionally validate the fully panned Jurkat samples (Fig. 3b and Supplementary Figs. 19–21). This confirmed strong cell activation in six of eight samples, weak activation for the A24/LMP2(TYG) sample and no activation for the CSS-930 Jurkat library enriched for A2/p65(NLV) binding.

To benchmark our dextramer panning method against conventional sorting of primary T cells, we stained peripheral blood
mononuclear cells (PBMCs) from donor CSS-930 with the four viral HLA-A*02:01 dextramers, sorted for CD8+ T cells, and conducted TCRβ clonotype analysis. The primary PBMCs displayed a low frequency of viral pMHC-binding T cells (range, 0.055–0.32%), in agreement with our TCRβ-Jurkat libraries (Supplementary Fig. 22). The dextramer-binding TCRβ clonotypes were typically quite rare in the primary T cell population (Supplementary Table 4), with a median read frequency of 0.0003%, suggesting that many of these clonotypes would have limited overlap across independent cell aliquots. Six of the TCRβ clonotypes were present at a frequency of more than 0.01% in the primary T cell repertoire, and three (50%) of these were also identified by panning of
Fig. 4 | Experimental validation of viral-specific TCRαβ clones. a, Validation data for the NLV.8 monoclonal TCR. The top panels show dextramer binding: pMHC dextramer signal is on the x axis and CD3 staining is on the y axis. The bottom panels show Jurkat cell activation upon co-culture with peptide-loaded aAPCs; CD69 staining is on the x axis and CD62L staining is on the y axis. The left column is irrelevant and the right column is target peptide. Dextramer-binding and cellular-activation experiments were independently repeated twice with similar results. b, Functional avidity analysis of clone NLV.8 using T2 cells pulsed with a dilution series of peptide. Peptide-loading concentration is on the x axis and EC50 values calculated using peptide-pulsed T2 cells as described in the legend. Two independent experiments were conducted for each clone, and the calculated EC50 values from each experiment are plotted. c, Functional avidity of viral-specific clones, using alanine scanning and IL-2 secretion. Clones NLV.8 and NLV.10 and the positive control (C7) are shown. T2 cells were loaded with wild-type, single-alanine-mutation or irrelevant peptides, and IL-2 secretion was measured (on the basis of data from Table 1). d, Integration of pMHC dextramer panning and activation data. Each point represents a unique clone, with clones colored according to their respective validation result. The dextramer enrichment of each clone is on the x axis, and the activation enrichment (ratio of activated-to-input TCR read count frequency) is on the y axis (on the basis of data from Table 1). e, Functional specificity of viral-specific clones, using alanine scanning and IL-2 secretion. Clones NLV.8 and NLV.10 and the positive control (C7) are shown. T2 cells were loaded with wild-type, single-alanine-mutation or irrelevant peptides, and IL-2 secretion was measured by ELISA (x axis). Replicate data (circles) and means (bars) are plotted for two independent experiments.

The TCRαβ-Jurkat library, indicating that the TCRαβ-Jurkat libraries were able to capture common dextramer-binding TCRs.

A major advantage of the TCRαβ-Jurkat library system is the ability to quickly assess the cellular activity of the dextramer-enriched TCRs and reduce the false-positive rate. To identify functional TCRs, we performed in vitro aAPC activation screens on partially enriched TCRαβ-Jurkat populations and sorted for activated (CD69+CD62L-) cells (Supplementary Fig. 23). We used deep sequencing to quantify clonotype enrichment in the activated cell population versus the starting population and integrated these data with the pMHC dextramer panning data (Fig. 3c). We found that only 19 of the 44 (43.2%) clonotypes most enriched by pMHC dextramer panning were also enriched in the activation screen. For example, A2/pp65(NLV) dextramer panning of the CSS-930 library strongly enriched for two TCRs; however, these TCRs did not display cellular activation in response to NLV-pulsed T2 cells. These data suggest that combining an activation screen with pMHC dextramer panning could reduce false-positive rates.

Characterization of individual antiviral TCRs. To quantify pMHC dextramer panning false-positive rates, we selected 34 of the most frequent TCRαβ clonotypes after pMHC dextramer panning (mean, 24.55%; median, 7.3%; range, 0.2–97.6%) for monoclonal expression in Jurkat cells (Table 1 and Supplementary Table 5). These clonotypes were generally very rare in the initial TCRαβ-Jurkat libraries (mean, 0.036%; median, 0.003%), with multiple clonotypes present at a frequency of less than 0.0001%. On average, pMHC dextramer panning enriched the clonotypes 136,672-fold (median, 3,640-fold), with two clonotypes, NLV.1 and NLV.8, enriching more than 1 million-fold. Notably, the NLV.8 clonotype was previously identified by other groups and shown to have high functional avidity and cytotoxic activity.4,5
Three of the 34 monoclonal TCRαβ constructs were poorly expressed in Jurkat cells, leaving 31 clones for downstream analysis. We demonstrated that 11/31 (35.5%) of the monoclonal TCRαβ cell lines bound their cognate pMHC, and 9/11 (81.8%) of the validated binders induced cellular activation (Fig. 4a, Supplementary Figs. 24–28, Table 1 and Supplementary Table 5). We also built two negative-control monoclonal TCRαβ cell lines from high-frequency clonotypes in the starting donor CSS-930 TCRαβ library that persisted throughout pMHC dextramer pan- ning. These high-frequency negative-control cell lines did not bind any of the four HLA-A*02:01 pMHC dextramers, as expected (Supplementary Fig. 29).

We further characterized the functional avidities of the 11 TCRαβ monoclonal cell lines that displayed target-specific pMHC binding. We co-cultured each TCRαβ–Jurkat line with peptide-pulsed T2 cells (dilution series, 0.0001–10,000 nM) and quantified cell activation by flow cytometry for surface CD69 staining (Fig. 4b and Supplementary Fig. 30). The previously reported α2A/pp65(MLV) ‘C7’ clone was used as a positive control40. The NLV.2 and NLV.9 cell lines failed to induce cellular activation, whereas functional avidity half-maximal effective concentration (EC50) values for the other nine lines ranged from 0.3 nM to 565 nM (median, 5.62 nM) (Fig. 4c, Table 1 and Supplementary Table 5).

We retrospectively assessed the benefit of combining pMHC dextramer panning with activation enrichment a priori in the initial screen (Fig. 4d). If we had applied a 100-fold pMHC dextramer enrichment cutoff, 15/26 clones (58%) would have been false positives for pMHC dextramer binding, and 17/26 clones (65%) would have been false positives for activation. If we had additionally applied an activation enrichment requirement of more than onefold, the overall false-positive rate would have decreased to 2/11 (18%). Seven of the nine monoclonal TCRαβ cell lines that validated as activators were the highest-frequency clonotypes after pMHC dextramer panning, indicating that dextramer panning can efficiently enrich for functional TCRs. However, this approach also enriched for two strong pMHC dextramer-binding clones (NLV.2 and NLV.9) that did not induce cellular activation—a phenomenon reported previously for some high-affinity TCRs41. Including the activation enrichment requirement would have eliminated these nonactivating TCRs from further analysis. Additionally, the activation enrichment would have verified functional activity for two low-frequency (<4%) post-panning TCRs, NLV.10 and CLG.4. Given these results, combining dextramer and cellular activation enrichment measurements decreased the false-positive rate and could also decrease the false-negative rate.

Finally, we investigated the specificity of eight TCR clones by measuring interleukin (IL)-2 secretion upon incubation with a series of peptides with single-alanine substitutions (Fig. 4e and Supplementary Fig. 31). Alanine mutations that reduced peptide presentation or TCR interaction induced lower IL-2 secretion than wild-type cognate peptides. These data provide some evidence of contact residues and antigen specificity, which varied between TCRs and pMHC targets. Although a larger library of pMHCs would better characterize specificity42, these alanine scanning data provide evidence that we identified TCRs with varied specificity toward their cognate pMHC, rather than spurious TCRs that are artifacts of our methodology.

**Identification and characterization of antitumor TCRs from tumor-infiltrating lymphocytes.** We obtained informed consent from an HLA-A*02:01+ patient with stage IV melanoma who was treated unsuccessfully with conventional therapy but who achieved a partial response to autologous TIL ACT43. We ran 1.36 million of the post-expanded therapeutic TILs through the microfluidic system and generated sequencing libraries from the natively paired TCRαβ amplicons. Deep sequencing revealed a diverse TCRαβ repertoire with 395,464 unique TCRαβ clonotypes in our natively paired TCRαβ amplicon library (Supplementary Table 1, Supplementary Fig. 32 and Fig. 4a). We also confirmed a high estimated TCRαβ clonotype yield of 96.6%, with 86,068 unique TCRαβ clonotypes present in the starting cell population, 83,172 in the linked TCRαβ amplicon library and 71,753 in the Jurkat expression library (Supplementary Table 2). Furthermore, the linked TCRαβ amplicon library captured 100% of the top 250 TCRαβ clonotypes present in the expanded TIL cell population (Supplementary Fig. 32e).

We built a TIL TCRαβ–Jurkat expression library and screened for pMHC dextramer binding to four well-characterized melanoma antigens expressed in patient-matched tumor tissue (Supplementary Tables 2 and 3, and Supplementary Fig. 33). The TIL TCRαβ–Jurkat library displayed clear binding only to the A2/PMEL(KTW) dextramer. This binding was enriched with panning (Fig. 5b), and cellular activation was confirmed by co-culture with peptide-pulsed T2 cells (Fig. 5c). We used deep sequencing to identify five TCRαβ clonotypes that were enriched by dextramer panning and APC-induced cellular activation (Fig. 5d, Supplementary Fig. 34, Table 1 and Supplementary Table 5). As we observed previously in the virus-seropositive libraries, the anti-A2/PMEL(KTW) binders were initially rare (≤0.3%) in the TIL TCRαβ–Jurkat library. This low level of antitumor antigen reactivity in TILs is consistent with recent reports using analogous methods44.
We first engineered the anti-A2/PMEL(KTW) TCRαβ clonotypes into Jurkat cells that lacked the CD8 co-receptor. All five properly paired clones bound pMHC dextramer and showed CD8-independent cellular activation against peptide-loaded T2 cells as well as endogenously expressed antigen on the HLA-A*02:01+ melanoma tumor cell line SK-Mel-5 (Supplementary Fig. 35). We also tested mispaired TCRα–TCRβ cell lines as negative controls, and none of these showed strong dextramer binding or cellular activation (PMEL.5-4 showed weak dextramer binding). These results suggest that the anti-PMEL TCR clones have high functional activity. We then reengineered the anti-PMEL TCRαβ clones into CD8+ Jurkat cells and determined their functional
These TCRs also demonstrated targeted tumor cell killing of HLA-A*02:01+ melanoma cell line SK-Mel-5 and unrelated HLA-A*02:01+ stage IV melanoma patient-derived CTMM4.1 tumor cells but not when co-cultured with the HLA-A*02:01+ melanoma cell line SK-Mel-28 (Fig. 5f and Supplementary Figs. 37 and 38). These TCRs also demonstrated targeted tumor cell killing of HLA-A*02:01+ SK-Mel-5 and CTMM4.1 cells (Fig. 5g and Supplementary Fig. 39), suggesting that cell killing was A2 restricted.

Discussion

In a recent query of the VDJdb website (https://vdjdb.cdr3.net/), we found fewer than 1,000 high-confidence paired TCRαβ sequences with known binding specificity and functional avidity\(^1\), even though a single person’s T cell repertoire comprises millions of clonotypes. Here we introduced a method to greatly accelerate the pace of identifying natively paired TCRαβ sequences and their cognate antigen specificities. Each immortalized T cell repertoire can be mined for TCRs against large panels of pMHC targets. Our method has the potential to scale, providing an opportunity to comprehensively profile human TCR reactivity to a diverse set of antigens.

We captured over 2.9 million natively paired TCRαβ clonotypes from six healthy virus-seropositive human donors. Compared to previously published reports using single-cell microfluidic and other methods\(^2\), we obtained 10- to 100-fold more natively paired TCRαβ sequences. Similarly, two recently published single-cell studies each profiled a total of about 30,000 TILs across multiple patient samples, obtaining 2,000–10,000 unique TCRαβ clonotypes per patient\(^3\), whereas our study profiled 1.36 million post-expanded TIL cells used for a therapeutic infusion product, and we identified 395,464 unique TCRαβ clonotypes. We then took the seven libraries of TCRαβ amplicons and built highly diverse, natively paired expression libraries in Jurkat cells. We identified and validated nine high-avidity antiviral TCRs and five high-avidity autotumor TCRs from these seven libraries. We also showed that integrating activation-based screening with dextramer binding is critical for reducing the false-positive rate and could also be useful for decreasing the false-negative rate, thus differentiating our system from other TCR discovery methods. Additionally, we identified three anti-PMEL TCRs that induced strong HLA-restricted CD8+ T cell degranulation and cell killing in response to endogenously presented antigen on the established melanoma cell line SK-Mel-5 and unrelated patient-derived cancer cells in vitro, demonstrating that this approach can efficiently identify functional TCRs from natively paired T cell repertoires.

Despite our progress, there remain many areas for improvement and further investigation. Although we benchmarked our technology against pMHC dextramer sorting of primary T cells, we could not compute the false-negative rate, because the number of ‘true positives’ in the repertoires was unknown. Further, our technology still requires low-throughput methods to validate functional avidity for each individual TCR identified in a screen. Also, the pMHC dextramer panning approach described here led to a highly oligoclonal enrichment of a small number of TCRs, whereas the cellular activation enrichment was able to identify several lower- frequency TCRs in partially pMHC dextramer-panned populations. Thus, a key area of continued optimization is in our flow sorting approach. For example, conducting the cellular activation enrichment after a single round of pMHC dextramer binding could improve false-positive and false-negative rates. Alternatively, implementing an initial selection step on the basis of activation followed by pMHC dextramer sorting may be beneficial. To identify even rarer clones, future work could make TCRαβ–Jurkat libraries from T cells expanded ex vivo with peptide-loaded APCs. The Jurkat activation system could also be improved, for example, by using artificial reporter systems or alternative endogenous markers for activation. Also, our methods were assessed for identification of TCRs against only 12 pMHC targets across only two HLA types. Different targets and HLA types will certainly require further protocol development, for example by panning with activation rather than pMHC binding, when pMHC multimer reagents are not available. Finally, to simultaneously screen for new pMHC targets and their cognate TCRs, we are actively working on methods for screening large pMHC target libraries against TCRαβ–Jurkat libraries.

The functional TCRs identified in this study have the potential for translation to clinical applications. Pioneering groups use virus-specific primary T cells as a way to restore long-term viral immunity after hematopoietic stem cell transplantation\(^4\). The EBV- and CMV-targeted TCRs identified in our study, plus TCRs for adenovirus, BK virus and respiratory syncytial virus, could one day be used for this clinical ACT application, perhaps offering a more streamlined manufacturing protocol and higher potency. Recently, chimeric antigen receptor T cell (CAR-T) therapies were approved by the US Food and Drug Administration for leukemia and lymphoma\(^5\), however, no CAR-T or ACTs have been approved for solid tumors. The anti-PMEL TCRs identified in this study could be assessed for potential development into TCR-engineered T cells for advanced melanoma. Because our TCRs are not artificially affinity matured, they might be less likely to show off-target toxicity\(^6\). In conclusion, our technology has removed a substantial bottleneck in functional TCR discovery and the development of new ACT.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41587-020-0438-y.

Received: 2 April 2019; Accepted: 27 January 2020; Published online: 16 March 2020

References

1. Yee, C. Adoptive T cell therapy: points to consider. Curr. Opin. Immunol. 51, 187–203 (2018).
2. Barrett, A. J., Procop, S. & Bolland, C. M. Virus-specific T cells: broadening applicability. Biol. Blood Marrow Transplant. 24, 13–18 (2018).
3. Romano, M., Fanelli, G., Albay, C. J., Giganti, G. & Lombardi, G. Past, present, and future of regulatory T cell therapy in transplantation and autoimmunity. Front. Immunol. 10, 43 (2019).
4. Chodon, T. et al. Adoptive transfer of MART-I T-cell receptor transgenic lymphocytes and dendritic cell vaccination in patients with metastatic melanoma. Clin. Cancer Res. 20, 2457–2465 (2014).
5. Rapoport, A. P. et al. NY-ESO-1-specific TCR-engineered T cells mediate sustained antigen-specific antitumor effects in myeloma. Nat. Med. 21, 914–921 (2015).
6. Tendeiro Rego, R., Morris, E. C. & Lowdell, M. W. T-cell receptor gene-modified cells: past promises, present methodologies and future challenges. Cytotherapy 21, 341–357 (2019).
7. Zhang, J. & Wang, L. The emerging world of TCR-T cell trials against cancer: a systematic review. Technol. Cancer Res. Treat. 18, 153033819831068 (2019).
8. Sadelan, M., Rivière, I. & Rubelli, S. Therapeutic T cell engineering. Nature 545, 423–431 (2017).
9. Dawson, N. A. J. & Leving, M. K. Antigen-specific regulatory T cells: are police CARs the answer? Transl. Res. 187, 53–58 (2017).
10. Limmemann, C. et al. High-throughput identification of antigen-specific TCRs by TCR gene capture. Nat. Med. 19, 1534–1541 (2013).
11. Schepers, W. et al. Low and variable tumor reactivity of the intratumoral TCR repertoire in human cancers. Nat. Med. 25, 89–94 (2019).
12. Yossef, R. et al. Enhanced detection of neoantigen-reactive T cells targeting unique and shared oncoenes for personalized cancer immunotherapy. JCI Insight 3, 122467 (2018).

13. 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).

14. Li, Y. et al. Directed evolution of human T-cell receptors with picomolar affinities by phage display. Nat. Biotechnol. 23, 349–354 (2005).

15. Wagner, E. K. et al. Human cytomegalovirus-specific T cell receptor engineered for high affinity and soluble expression using mammalian cell display. J. Biol. Chem. 294, 5790–5804 (2019).

16. 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 MAEG-A10: strategy for selection of an optimal candidate. Oncoimmunology 8, e1532759 (2019).

17. 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).

18. Guo, X.-Z. J. et al. Rapid cloning, expression, and functional characterization of paired αβ and γδ T-cell receptor chains from single-cell analysis. Mol. Ther. Methods Clin. Dev. 3, 15054 (2016).

19. Azizi, E. et al. Single-cell map of diverse immune phenotypes in the breast tumor microenvironment. Cell 174, 1293–1308 (2018).

20. Adler, A. S. et al. Rare, high-affinity anti-pathogen antibodies from human repertoire, discovered using microfluidics and molecular genomics. mAbs 9, 1282–1296 (2017).

21. Wang, B. et al. Functional interrogation and mining of natively paired human V_{\alpha}V_{\beta} antibody repertoires. Nat. Biotechnol. 36, 152–155 (2018).

22. Kieke, M. C. et al. Selection of functional T cell receptor mutants from a yeast surface-display library. Proc. Natl Acad. Sci. USA 96, 5651–5656 (1999).

23. Smith, S. N., Harris, D. T. & Kranz, D. M. in Yeast Surface Display 1319, 95–141 (Springer, 2015).

24. Kuhns, M. S., Davis, M. M. & Garcia, K. C. Deconstructing the form and function of the TCR/CD3 complex. Immunity 24, 133–139 (2006).

25. Tsuji, T. et al. Rapid construction of anitumor T-cell receptor vectors from frozen tumors for engineered T-cell therapy. Cancer Immunol. Res. 6, 594–604 (2018).

26. Shagay, M. et al. Towards error-free profiling of immune repertoires. Nat. Methods 11, 653–655 (2014).

27. Heath, J. M., Ismail, M., Oakes, T. & Chain, B. High-throughput sequencing of the T-cell receptor repertoire: pitfalls and opportunities. Brief. Bioinformatics 19, 554–565 (2018).

28. Exley, M., Porcelli, S., Furman, M., Garcia, J. & Balk, S. CD161 (NKR-P1A) costimulation of CD1d-dependent activation of human T cells expressing invariant V_{\alpha}24J_{\alpha}Q T cell receptor α chains. J. Exp. Med. 188, 867–876 (1998).

29. Kjer-Nielsen, L. et al. MR1 presents microbial vitamin B metabolites to MAIT cells. Nature 491, 717–723 (2012).

30. Gold, M. C. et al. MR1-restricted MAIT cells display ligand discrimination and pathogen selectivity through distinct T cell receptor usage. J. Exp. Med. 211, 1601–1610 (2014).

31. Dellabona, P., Padovan, E., Casorati, G., Brockhaus, M. & Lanzavecchia, A. An invariant V_{\alpha}24–I_{\beta}Q/V_{\beta}11 T cell receptor is expressed in all individuals by clonally expanded CD4-8+ T cells. J. Exp. Med. 180, 1171–1176 (1994).

32. Porcelli, S., Yockey, C. E., Brenner, M. B. & Balk, S. P. Analysis of T cell antigen receptor (TCR) expression by human peripheral blood CD4-8+ αβ T cells demonstrates preferential use of several V_{\alpha}J_{\beta} genes and an invariant TCRβ chain. J. Exp. Med. 178, 1–16 (1993).

33. Held, K., Beltrán, E., Moser, M., Hohlfeld, R. & Dornmair, K. T-cell receptor repertoire of human peripheral CD161^{+}TRA1-2+ MAIT cells revealed by next generation sequencing and single cell analysis. Hum. Immunol. 76, 607–614 (2015).

34. Dash, P. et al. Quantifiable predictive features define epitope-specific T cell receptor repertoires. Nature 547, 89–93 (2017).

35. Glavivic, J. et al. Identifying specificity groups in the T cell receptor repertoire. Nature 547, 94–98 (2017).

36. Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. Nat. Methods 11, 783–784 (2014).

37. 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).

38. 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).

39. 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).

40. Yang, X. et al. Structural basis for clonal diversity of the public T cell response to a dominant human cytomegalovirus epitope. J. Biol. Chem. 290, 29106–29119 (2015).

41. Siember, 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).

42. Birnbaum, M. E. et al. Deconstructing the peptide–MHC specificity of T cell recognition. Cell 157, 1073–1087 (2014).

43. Rosenberg, S. A. et al. Durable complete responses in heavily pretreated patients with metastatic melanoma using T-cell transfer immunotherapy. Clin. Cancer Res. 17, 4550–4557 (2011).

44. Shugay, M. et al. VDJdb: a curated database of T-cell receptor sequences with known antigen specificity. Nucleic Acids Res. 46, D419–D427 (2018).

45. Yost, K. E. et al. Clonal replacement of tumor-specific T cells following PD-1 blockade. Nat. Med. 25, 1251–1259 (2019).

46. Gerdemann, U. et al. Rapidly generated multivirusspecific cytoytic T lymphocytes for the prophylaxis and treatment of viral infections. Mol. Ther. 20, 1622–1632 (2012).

47. Barrett, A. J. & Bollard, C. M. The coming of age of adoptive T-cell therapy for viral infection after stem cell transplantation. Ann. Transl. Med. 3, 62 (2015).

48. O’Leary, M. C. et al. FDA approval summary: tisagenlecleucel for treatment of patients with relapsed or refractory B-cell precursor acute lymphoblastic leukemia. Clin. Cancer Res. 25, 1142–1146 (2019).

49. Rouchkov, N. et al. FDA approval summary: axicabtagene ciloleucel for treatment of patients with relapsed or refractory B-cell precursor acute lymphoblastic leukemia. Clin. Cancer Res. 25, 1702–1708 (2019).

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
© The Author(s), under exclusive licence to Springer Nature America, Inc. 2020
Methods

Sourcing and processing human materials. Deidentified PBMCs in leukopacks were obtained from HLA-A*02:01 and HLA-A*24:02* healthy donors (AllCells) under an institutional review board-approved protocol. T cells were isolated from the PBMCs using EasySep Human CD8+ T Cell Enrichment Kit (Stemcell Technologies). After isolation, T cells were cryopreserved using Cryostor CS10 (Stemcell Technologies). For downstream single-cell TCR–TCRβ linkage, cells were thawed, washed and resuspended at 5,000–6,000 cells per μl in cold DPBS and 0.5% BSA with 12% OptiPrep Density Gradient Medium (Sigma). The resuspended cells were then used for microfluidic encapsulation as described in the next section.

Human TILs were generated from tumor biopsies obtained from consenting individuals undergoing TIL clinical therapy for metastatic melanoma using a proprietary method that, in brief, involves the following. Tumor biopsies were disaggregated to a single-cell suspension using collagenase and DNase and then plated into culture flasks in the presence of IL-2 for up to 21 d. After this initial outgrowth, the TILs underwent a rapid expansion protocol (REP) in which the derived TIL cells were mixed with donor-doped irradiated PBMC feeder cells with the addition of OKT3 and IL-2 for 14 d before cryopreservation. Patient-matched tumor lines were grown by plating digested tumor material onto 24-well tissue culture plates in RPMI supplemented with 10% FBS (Sigma), 0.01 M HEPES (Sigma) and 1% penicillin and streptomycin (Sigma). After 24 h, the supernatant was removed from the plate and fresh medium was added to the original well. Wells were observed every few days, medium was changed and wells were split as necessary. The CTMMA1 patient-derived cell line was derived from an adherent cell population of an HLA-A*02:01 patient with melanoma. The CTMMA1 melanoma cells and C3S-673 TIL samples were from unrelated patients. Wels were observed every few days, medium was changed and wells were split as necessary. The CTMMA1 patient-derived cell line was derived from an adherent cell population of an HLA-A*02:01 patient with melanoma. The CTMMA1 melanoma cells and C3S-673 TIL samples were from unrelated patients.

For TCR-mediated tumor cell killing assays, human CD8+ T cells were isolated from normal buffy coats, obtained from the National Health Service Blood Transfusion service. PBMCs were isolated by diluting 50:50 in PBS with 0.5% FBS and 2 mM EDTA and then layering onto Ficoll-Paque (GE Healthcare) before centrifuging for 20 min. Red blood cells were removed using 1× Pharrmalyte RBC lysis buffer (BD Biosciences), and CD8+ T cells were isolated from PBMCs using the EasySep Human CD8+ T Cell Isolation Kit (Stemcell Technologies).

Generating paired TCR–TCRβ linkage libraries. Library generation was divided into three steps: (1) poly(A)+ mRNA capture, (2) multiplexed overlap extension RT–PCR (OE RT–PCR) and (3) nested PCR to remove artifacts and add adaptors for deep sequencing or expression libraries17. For poly(A)+ mRNA capture, we used a custom-designed co-flow emulsion droplet microfluidic chip fabricated from glass (Dolomite Microfluidics). The microfluidic chip has two input channels for fluorocarbon oil (Dolomite Microfluidics) and an input channel for the cell suspension mix and one input channel for oligo(dT) beads (New England Biolabs) in 0.5 M NaCl, 0.5% Tween-20 and 20 mM dithiothreitol. The input channels are etched to 50 μm x 150 μm for most of the chip’s length, narrow to 55 μm at the droplet junction and are coated with hydrophilic Pico-Glide (Dolomite Microfluidics). Three Mitos P-Pump pumps (Dolomite Microfluidics) are used to pump the liquids through the chip. Droplet size depends on pressure, but typically we found that droplets of ~45 μm in diameter were optimal. Emulsions were collected into 1.5-microcentrifuge tubes and incubated at 40 °C for 30 min to capture mRNA onto oligo(dT) beads. Emulsions were then broken using Pico-Break (Dolomite Microfluidics), and mRNA-bound beads were magnetically isolated. Droplets containing cell nuclei were physically linked by overlapping primer sequences included on the TRAC and TRBV primers. The amplified DNA was recovered from the droplets using a proprietary droplet breaking solution (GigaMune) and purified using a QIAquick PCR Purification Kit (Qiagen). For nested PCR, the OE RT–PCR product was first run on a 1.7% agarose gel and the band of interest was excised and purified using NucleoSpin Gel and a PCR Clean-up Kit (Macherey-Nagel). Nested PCR was performed using NEBNext amplification mix (New England Biolabs) to add adaptors for Illumina sequencing or cloning into a mammalian expression construct. PCR products were run on a 1.2% agarose gel, and the band at 800–1,000 bp was excised and purified using NucleoSpin Gel and a PCR Clean-up Kit (Macherey-Nagel). We ran 2.7–3.4 million live T cells from each of our six healthy PBMC donors through this workflow to generate our six viral-seropositive TCRβ repertoire libraries (Supplementary Table 1). We ran 1.4 million live T cells from the post-REP TIL sample to generate the TIL TCRβ library (Supplementary Table 1).

Linked TCRβ repertoire sequencing. Deep TCRβ repertoire sequencing libraries were quantified using the quantitative PCR Illumina Library Quantitation Kit (Kapa Biosystems) and diluted to 8.5–10 pm. Libraries were sequenced on a MiSeq Illumina using v3 600-cycle MiSeq Reagent Kits, according to the manufacturer’s instructions. To identify the paired sequences from the TCRβ libraries, we obtained forward reads of 357 cycles that covered the TCRβ V gene and CDR3 and reverse reads of 162 cycles that covered the TCRβ J gene and enough of the TCRβ V gene for accurate calling. Repertoire sequencing data are available in the Sequence Read Archive, using the Sequence ID PRJNA585353.

To exclude base call errors, we used the expected error filtering method18. The expected number of errors (E) for a read was calculated from its Phred scores. By default, reads with E greater than 3 were discarded. Except where noted, statistics for TCR clonotype counts included singleton sequencing reads. We classified a unique clonotype by the CDR3 α amino acid sequence plus V- and J-gene usage with ImMunoGeneTics (IMGt) nomenclature. Exceptions included the pairing fidelity density plots (Fig. 2b and Supplementary Figs. 6 and 32c), the cross-sample comparison UpSet plots (Fig. 2c and Supplementary Figs. 7 and 8), the percent sharing plots (Supplementary Fig. 9) and the library-build cloning correlation and heat maps (Supplementary Figs. 14 and 16), for which we excluded singleton sequence reads from the analysis to reduce the risk of Illumina barcode sequence miscalls.

To identify reading frame and CDR3 α amino acid sequences generated by V(D)J rearrangements, we first processed a database of well-curated TCR sequences19 (IMGt, http://www.imgt.org/download/LIGM-D/B/) to generate position-specific sequence matrices (PSSMs) for the 5′ and 3′ CDR3 junctions (Supplementary Fig. 40). Each nucleotide sequence from the Illumina sequencing runs was translated into all reading frames. We then used the PSSMs to identify the FR3–CDR3 (5′) and CDR3–FR4 (3′) junctions and the protein reading frame for each of the nucleotide sequences. To report a CDR3 sequence, we required 5′ and 3′ PSSMs to be identified. Additionally, sequences that had low PSSM identity scores were marked with an exclamation point. These steps allowed us to predict valid, functional CDR3 sequences with high confidence. We queried TCRα and TCRβ nucleotide sequences against the IMGt database of reference V- and J-gene germline sequences using UBLAST (https://www.drive5.com/usearch/manual/ublast.algo.html) and V and J genes were identified based on the UBLAST alignments with the best lowest (E) values.

Single-chain TCRα and TCRβ repertoire sequencing. For TCRβ analysis of primary T cells, RNA was isolated from 3 million T cells from healthy donors C3S-930, CSS-938, CSS-944, CSS-948, CSS-962 and CSS-963 and from 2 million T cells from the TIL infusion product CSS-673 using the Nucleospin RNA Plus kit (Macherey-Nagel). TCRβ (V) regions were amplified using a mixture of primers directed against the V genes and a common primer within the TRBC region (Supplementary Fig. 2); these primers contained adaptors for Illumina sequencing. RT–PCR was conducted using SuperScript III reverse transcriptase (Thermo Fisher) and Platinum Taq (Thermo Fisher). These amplicons were run on 1.7% agarose gels, and the bands at approximately 480 bp were excised and purified using NucleoSpin Gel and a PCR Clean-up Kit (Macherey-Nagel). Samples were separately quantified using an Illumina Library Quantification Kit (Kapa Biosystems). After diluting to 9 pm, libraries were sequenced on a MiSeq (Illumina) using v2 300-cycle and 500-cycle MiSeq Reagent Kits, according to the manufacturer’s instructions.

Single-chain sequencing of the recombinant TCRα–TCRβ expression libraries, TCRα and TCRβ V(DFJ) regions were amplified separately using universal primers that contained adaptors for Illumina sequencing, within the TRAV, Leader and TRAC regions for TCRα and within the TRBV,Leader and TRBC regions for TCRβ. We conducted RT–PCR of RNA samples using SuperScript III reverse transcriptase (Thermo Fisher) and Platinum Taq (Thermo Fisher). These amplicons were run on 1.7% agarose gels, and the bands at 500–600 bp were excised and purified using NucleoSpin Gel and a PCR Clean-up Kit (Macherey-Nagel). Samples were separately quantified using an Illumina Library Quantitation Kit (Kapa Biosystems). After diluting to 9 pm, libraries were sequenced on a MiSeq (Illumina) using v2 500-cycle and 500-cycle MiSeq Reagent Kits according to the manufacturer’s instructions.

TCRα and TCRβ sequences were analyzed separately using the same methods as described above for paired sequences.

Visualization of TCRα and TCRβ repertoire. To generate the V-β gene usage heat maps, we tallied the number of unique clonotypes (CDR3 α amino acid sequence + V-gene + J-gene calls) for each α-chain variable gene and β-chain variable gene pair. The heat maps were generated in R 3.4.2 with ggplot2 v3.1.0. To evaluate TCRβ pairing precision, we first annotated each TCRβ as nonvariant or invariant, on the basis of previously published invariant TCRβ sequences20. For each TCRβ sequence, we classified it as invariant or variable, based on the number of unique partners with which they were found and generated density plots (Fig. 2b and Supplementary Figs. 6 and 32c).

To compare TCR α clonotype sharing among donors, we first curated a list of all TCRα, TCRβ and TCRβ α clonotypes (defined as unique CDR3 amino acid sequences + α-gene and β-gene usage) across libraries from all donors. We then asked whether each clonotype was observed with at least two reads in a given library. The UpSet plots were generated in R 3.4.2 with UpSetR v1.3.3. To illustrate clonotype enrichment through the panning rounds, we selected the five most abundant clones from the fully dextramer-enriched population and
display their sequencing read frequency as heat maps for each round of dextramer panning. Similarly, we present sequencing read frequency heat maps for these clonotypes from the activation assays.

Generating recombinant TCRβ-Jurkat expression libraries. We developed a subcloning workflow to convert the linked TCRβ amplicons into full-length lentiviral expression constructs (Supplementary Fig. 3). In this workflow, we first used NEBuilder to add overlap adaptors to the 5′ and 3′ ends of the linked TCRβ amplicons for downstream Gibson assembly. Then, we used NEBuilder HiFi DNA Assembly Master Mix (New England Biolabs) to insert the linked TCRβ library into a p2A-derived lentiviral vector (GeneCopoeia) that contained the EF1α promoter, a TCRβ leader sequence, the TCRβ constant region and a puromycin resistance gene. We transformed this intermediate library into E. coli. The TCRβ sequence was confirmed by Sanger sequencing. We selected 2–5 seropositive donor TCRβ-Jurkat library cells with 10 μl of APC- or PE-conjugated dextramer at room temperature for 10 min. Cells were then stained with an anti-CD3 antibody (FITC or APC-conjugated; clone UCHT1, BioLegend) for 30 min at 4°C and DAPI (BioLegend) to assess cell viability. Cells were then sorted on a FACSMelody, BD Influx or CytoFLEX LX (Beckman Coulter) as a positive control and irrelevant-peptide-pulsed T2 cells as a negative control. T2 cells natively expressed HLA-A*02:01, and we generated a stable T2 cell line that expressed HLA-A*02:01 and GFP (data not shown) for additional peptide presentation.

To identify TCRs present in peptide-activated Jurkat cells, we co-cultured partially enriched TCRβ-Jurkat cell populations with peptide-pulsed T2 cells, stained with the activation markers described above and sorted for activated (HLA-A2* CD69+CD2L+) cells. These sorted cells were lysed and RNA was isolated using a NucleoSpin RNA Plus XS kit (Macherey-Nagel) and viral titers were quantified.

Lentiviral transduction of Jurkat cells. To identify antigen-reactive TCRβ clones, we converted TCRβ-Jurkat cell lines via retroviral transduction using lentiviral constructs using the p2A sequence (Integrated DNA Technologies). The TIL TCRβ–Jurkat library was stained with pMHC dextramers (Immudex) and TCRβ clonotype counts. V- and J-gene usage were classified as having one TCR clone. All other wells were analyzed for positive control, irrelevant and irrelevant-peptide-pulsed T2 cells as a negative control. T2 cells natively expressed HLA-A*02:01, and we generated a stable T2 cell line that expressed HLA-A*02:01 and GFP (data not shown) for additional peptide presentation.

To identify TCRs present in peptide-activated Jurkat cells, we co-cultured partially enriched TCRβ–Jurkat cell populations with peptide-pulsed T2 cells, stained with the activation markers described above and sorted for activated (HLA-A2* CD69+CD2L+) cells. These sorted cells were lysed and RNA was isolated using a NucleoSpin RNA Plus XS kit (Macherey-Nagel) and viral titers were quantified.
These monoclonal TCRβ expression constructs followed the same layout as the TCRβ libraries. Lentiviral plasmids were sequence verified by Sanger sequencing, packaged into YSV-G pseudotyped lentiviral particles and transduced into ΔTCRβ Jurkat cells, and stable cell lines were selected. Antibody TCRβ clones were introduced into CD8+ ΔTCRβ Jurkat cells, and anti-PMEL TCRβ clones were first introduced into CD8+ ΔTCRβ Jurkat cells and then into CD8+ ΔTCRβ Jurkat cells. TCRβ–Jurkat clones can be made available to nonprofit researchers, subject to approval by the ATCC.

Monoclonal TCRβ–Jurkat cell lines were assessed for pMHC binding and cellular activation. We stained 0.5–1 million cells with 5μl of pMHC dextramer and anti-CD3 antibodies and analyzed binding as described above. We then ran co-culture assays with the monoclonal TCRβ–Jurkat cell lines that showed pMHC binding. As described above, we pulsed T2 cells with 10μg per ml peptide, mixed 200,000 peptide-pulsed T2 cells with 200,000 TCRβ–Jurkat cells per well and measured cell activation by staining for CD69 and CD62L. We then assessed TCRβ–mediated cell activation in response to tumor cells by co-culturing 500,000 anti-PMEL TCRβ–Jurkat cells with a confluent monolayer of melanoma cells from the SK-MEL-5 (HLA-A*0201*PMEL+; ATCC, HTB-70) and SK-MEL-28 (HLA-A*0201*PMEL+; ATCC, HTB-72) lines in a 24-well plate.

We measured the functional avidity of the monoclonal TCRs that showed cellular activation using T2 cells pulsed with a dilution series of peptide (0.1 pM–10 μM). Cells were co-cultured in duplicate wells for 20 h and then harvested, pooled and stained for the CD69 and CD62L activation markers as described above. Cells were run on the CytoFLEX LX, and Jurkat cell CD69 median fluorescence intensity (MFI) was calculated using FlowJo (Treestar/BD Biosciences) and analyzed in Prism (GraphPad). A nonlinear three-parameter curve was fit to the dose response data. Two independent experiments were conducted, and the means were plotted for the dose response curves. EC50 values were calculated for each independent experiment and plotted.

We analyzed a set of viral antigen-reactive TCRs for peptide specificity using alanine scanning mutagenesis as previously described34. Synthetic peptides with alanine substitutions were obtained from Pepscan (purity = crude), resuspended in DMSO to 5 mg ml−1, aliquoted and stored at −20 °C. T2 cells were pulsed with 10 μg ml−1 peptide and co-cultured with monoclonal ΔTCRβ TCRβ–Jurkat cell lines for 24 h as described above. After co-culture, the cell supernatant was removed and stored at −80 °C. IL-2 levels were measured in the cell supernatant by sandwich ELISA (BioLegend), following the manufacturer’s protocol. MaxiSorp ELISA plates were coated with antibody at 4 °C overnight. ELISA wells were blocked with 1% BSA in PBS, 50 μl of cell supernatant was added to duplicate wells and plates were incubated at room temperature for 2 h. IL-2 levels were measured using an anti-IL-2 detection antibody followed by Avidin–HRP incubation and signal development with a TMB substrate solution. Signal absorbance was read at 450 nm on a SpectraMax 13x (Molecular Devices). IL-2 levels were calculated using a standard curve run on the same plate. Two independent experiments were conducted, and data points were plotted with the mean in Prism (GraphPad).

Monoclonal TCRβ characterization in human T cells. For enhanced expression, detection and proper pairing of the recombiant TCRβ clones in human T cells, we generated human–mouse fusion TCRαβ (described above), and the resulting post-expanded cells were used fresh or cryopreserved after the WST-1 assay using anti-mouse TCRβ (clone H57-597) antibody and run on a MACSQuantVYB (Milleniy Biotech) (Supplementary Fig. 39). Data were analyzed using Flowjo.

Recombinant TCR-mediated T cell degranulation was measured by culturing TCR-transduced cells with melanoma cell lines (SK-Mel-5, SK-Mel28 and CTM44.1) and staining for CD107a. Transduced T cells were co-cultured with tumor cells at a 1:2 effector-to-target (E:T) ratio with 1:1,000 dilutions of monensin and brefeldin A with a 1:50 dilution of anti-human CD107a–PE (clone REA792, Mileniy Biotech) antibody for 4 h. The cells were then stained with eFluor 780 Fixable Viability Dye (eBioscience), along with extracellular staining with anti-CD8–APC (clone SK1, BioLegend) and anti-CD4–FITC (clone SK5, BD Biosciences), before fixing in 4% paraformaldehyde. Cells were resuspended in 200 μl DPBS with 2 mM EDTA and 0.5% FBS before the samples were run on a MACSQuant Flow Cytometer. Triplicate wells were run for each condition per human T cell donor. Three independent experiments were conducted using three different human T cell donors. Data were analyzed using Flowjo and Prism (GraphPad) software, and statistical analysis was conducted using a two-way ANOVA with Dunnett’s multiple-comparisons test. Data from the three independent experiments are plotted as the mean and s.d. in Fig. 5f. The three technical replicates (dots) for each of the three human donors are plotted with the mean (bars) in Supplementary Fig. 37.

TCR-directed tumor cell killing was measured using WST-1 assays on co-cultured tumor and T cells. Twenty thousand tumor cells were co-cultured with varying amounts of T cells at E:T ratios of 1:10, 1:5, 1:2, 1:1 and 2:1 in a total volume of 200 μl per well in 96-well round-bottom plates and incubated overnight. The next day, cells were washed twice with Flowjo, the cell supernatant was removed and 10 μl of Cell Proliferation Reagent WST-1 (Roche) was added to the cells. After incubation for 30 min, 80 μl of the supernatant was transferred to a flat-bottom plate for analysis on a FLUOStar Omega plate reader (BMG Labtech). For the TCR-specific killing signal, we performed the following transformation:

\[
\text{Relative lyso} = \left( \frac{\text{[TCR}}{-\text{TCR}}} \right) \times 100
\]

Triplicate wells were run for each co-culture condition, and mean and s.d. values from the resulting WST-1 measurements were plotted for one experiment using Prism (GraphPad).

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability
Presort TCR–TCRβ repertoire fasta sequence files are deposited at the Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra/), under BioProject ID PRJNA541985. All other data are available from the corresponding author upon reasonable request.

References
50. Edgar, R. C. & Flyvbjerg, H. Error filtering, pair assembly and error correction for next-generation sequencing reads. Bioinformatics 31, 3476–3482 (2015).
51. Lefranc, M.-P. et al. IMGT, the international ImMunoGeneTics information system. Nucleic Acids Res. 37, D1006–D1012 (2009).
52. Edgar, R. C. Search engine chaining and clustering orders of magnitude faster than BLAST. Bioinformatics 26, 2460–2461 (2010).
53. Exley, M., Garcia, J., Balk, S. P. & Porcelli, S. Requirements for CD1d recognition by human invariant Vα24*CD4–CD8− T cells. J. Exp. Med. 186, 109–120 (1997).
54. Han, M. et al. Invariant or highly conserved TCRβ clones are expressed on double-negative (CD3−CD4−CD8−) and CD8+ T cells. J Immunol 163, 301–311 (1999).
55. Démoulin, T., Gachelin, G., Bequet, D. & Dormont, D. A biased Vα24+ T-cell repertoire leads to circulating NKT-cell defects in a multiple sclerosis patient at the onset of his disease. Immunol. Lett. 90, 223–228 (2003).
56. Greeneway, H. Y. et al. The clinical and biologic impact of TCRβ sequences can be produced efficiently by VI gene recombination. Immunochemistry 218, 213–224 (2013).
57. Van Rhijn, I. et al. A conserved human T cell population targets mycobacterial antigens presented by CD1β. Nat. Immunol. 14, 706–713 (2013).
58. Lepore, M. et al. Parallel T-cell cloning and deep sequencing of human MALT cell lines reveal stable oligo-TCRβ repertoire data. Sci. Immunol. 5, eaav3866 (2014).
59. Funston, G. M., Kallioinen, S. E., de Felipe, P., Ryan, M. D. & Igo, R. D. Expression of heterologous genes in oncolytic adenoviruses using picornaviral 2A sequences that trigger ribosome skipping. J. Gen. Virol. 89, 389–396 (2008).
60. Lyons, G. E. et al. Influence of human CD8 on antigen recognition by T-cell receptor-transduced cells. Cancer Res. 66, 11455–11461 (2006).
61. Thakral, K., Dobbins, J., Devine, L. & Kavathas, P. B. Differential expression of the human CD8β splice variants and regulation of the M-2 isoform by ubiquitination. J. Immunol. 180, 7431–7442 (2008).

NATURE BIOTECHNOLOGY | www.nature.com/naturebiotechnology
62. Zufferey, R. et al. Self-inactivating lentivirus vector for safe and efficient in vivo gene delivery. *J. Virol.* **72**, 9873–9880 (1998).

63. Han, A., Glanville, J., Hansmann, L. & Davis, M. M. Linking T-cell receptor sequence to functional phenotype at the single-cell level. *Nat. Biotechnol.* **32**, 684–692 (2014).

64. Chheda, Z. S. et al. Novel and shared neoantigen derived from histone 3 variant H3.3K27M mutation for glioma T cell therapy. *J. Exp. Med.* **215**, 141–157 (2018).

65. Goff, S. L. et al. Enhanced receptor expression and in vitro effector function of a murine–human hybrid MART-1-reactive T cell receptor following a rapid expansion. *Cancer Immunol. Immunother.* **59**, 1551–1560 (2010).

Acknowledgements
This work was partially funded by National Institute of Allergy and Infectious Diseases grant R43AI120313-01 to D.S.J. and National Cancer Institute grant R43CA232942 to M.J.S. N.O. is supported with funding from the Medical Research Council, Engineering and Physical Sciences Research Council Centre for Doctoral Training (EP/L014904/1).
R. Mizrahi, M. Adams, R. Leong and J. Leong (GigaGen) assisted with the development of Gibson assembly protocols. E. Stone and S. Keating (GigaGen) provided useful discussions about general T cell immunology. R. Guest (Immetacyte) provided assistance with isolation and expansion of TILs.

Author contributions
Conceptualization: M.J.S., J.S.B., A.S.A., E.H.M., R.E.H., M.C. and D.S.J.; methodology: M.J.S., E.K.W., J.M.H., N.O., J.S.B., A.S.A., E.H.M. and D.S.J.; software: R.C.E. and Y.W.L.; validation: M.J.S., A.L.N. and E.K.W.; investigation: M.J.S., A.L.N., N.O., J.S.B. and E.K.W.; data curation: M.J.S., E.K.W., J.M.H., Y.W.L., A.S.A. and D.S.J.; writing—original draft preparation: M.J.S. and D.S.J.; writing—review and editing: M.J.S., E.K.W., N.O., J.S.B., J.M.H., A.S.A., Y.W.L., E.H.M., M.C. and D.S.J.; visualization: M.J.S., E.K.W., J.M.H., Y.W.L., A.S.A. and D.S.J.; supervision: M.J.S., J.S.B., A.S.A. and D.S.J.; project administration: M.J.S. and D.S.J.; funding acquisition: M.J.S. and D.S.J.

Competing interests
M.J.S., A.L.N., E.K.W., A.S.A., Y.W.L., R.C.E. and D.S.J. are salaried employees of GigaGen, which is an affiliate of GigaMune. GigaMune pays cash to GigaGen for research services. M.J.S., A.L.N., E.K.W., A.S.A., R.C.E., Y.W.L., E.H.M., M.C. and D.S.J. are holders of equity shares in GigaMune. J.M.H. and M.C. hold research positions at the Massachusetts General Hospital. Massachusetts General Hospital has entered into a research collaboration with GigaMune. M.C. is currently an employee of AstraZeneca. M.C. owns equity in Revitope Oncology and Gristone Oncology. M.C. received consultant fees from Merck Laboratories. J.S.B. and R.E.H. are salaried employees of Immetacyte. R.E.H. is a holder of equity shares in Immetacyte. The viral TCRs and TCR repertoire mining methods are described in US Patent and Trademark Office (USPTO) provisional patent application 62/821808, assigned to GigaMune (M.J.S., A.L.N., E.K.W., Y.W.L., A.S.A., M.A.A. and D.S.J.). The PMEL TCRs are described in USPTO provisional patent application 62/842691, assigned to GigaMune (M.J.S., A.S.A., M.A.A. and D.S.J.). Methods for generating TCR libraries are described in patents WO2012083225A2, US20160362470A1, US20170247684A1 and US20170247683A1, assigned to GigaGen or GigaMune (M.J.S., A.S.A., E.H.M. and D.S.J.).

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41587-020-0438-y.
Correspondence and requests for materials should be addressed to D.S.J.
Reprints and permissions information is available at www.nature.com/reprints.
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
- Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted
- Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

- Data collection: FlowJo (10.5.3), MiSeq 3.1, BaseSpace
- Data analysis: usearch (v11), ublast (v11), R (3.4.2), UpSetR (1.3.3), ggplot2 (3.1.0), igraph (1.2.4), GraphPad Prism (v8)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Paired TCRab repertoire data (Figure 2) has been deposited in the NCBI short read archive under BioProject ID PRJNA541985. Complete TCRab sequences for functionally characterized TCRab clones (Table 2) will be deposited in Genbank prior to publication.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences
- Behavioural & social sciences
- Ecological, evolutionary & environmental sciences
Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| Sample size | No sample-size calculation was performed |
| Data exclusions | No data were excluded |
| Replication | All attempts at replication were successful |
| Randomization | This was not relevant to this study |
| Blinding | This was not relevant to this study |

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

- **n/a** Involved in the study
  - Antibodies
    - □ □ Antigen
    - ■ □ Eukaryotic cell lines
    - □ □ Palaeontology
    - □ □ Animals and other organisms
    - □ □ Human research participants
    - □ □ Clinical data

### Methods

- **n/a** Involved in the study
  - □ □ ChIP-seq
  - □ □ Flow cytometry
  - □ □ MRI-based neuroimaging

### Antibodies

**Antibodies used**

All antibodies were obtained from BioLegend unless indicated: anti-human CD3 [clone: UCHT1, cat#s 300439 & 300440], anti-human TCRalpha/beta [clone: IP26, cat# 306706], anti-human CD69 [clone: FN50, cat# 310906], anti-human CD62L [clone: DREG-56, cat# 304810], anti-human HLA-A2 [clone: BB7.2, cat# 343304], anti-human CD8B [clone: REA715, Miltenyi Biotec cat# 130-110-509], anti-human CD8A [clone: SK1, BD Pharmingen cat# 560179], anti-human CD4 [clone: SK3, BD Biosciences cat# 566320], anti-human CD107a [clone: REA792, Miltenyi Biotec cat# 130-111-621], and anti-mouse TCRb [clone: H57-597, cat# 109212]

Dextramers were obtained from Immudex and are listed in Supplementary Table 1.

**Validation**

Antibodies were validated by commercial vendors and this information is available on their websites.

### Eukaryotic cell lines

**Policy information about cell lines**

**Cell line source(s)**

- Wild type Jurkat cells [clone: E6-1, from ATCC TIB-152]
- TCRbeta deficient Jurkat cells [clone: J.RT3-T3.5, from ATCC TIB-153]
- T2 cells [174 x CEM.T2, from ATCC CRL-1992]
- Lenti-Pac 293Ta (HEK-293) cells (from GeneCopoeia, cat# LT008)

**Authentication**

Product data sheets and certificates of analysis were provided by vendors. We validated the TCRbeta deficient Jurkat cell line by staining for surface CD3 and TCRalpha/beta expression and confirmed loss of expression by flow cytometry. Cell lines have also been authenticated by morphology checking and through use in other experiments.

**Mycoplasma contamination**

Tested negative for mycoplasma contamination.

**Commonly misidentified lines**

(See ICLAC register)

- No commonly misidentified cell lines were used.
# Flow Cytometry

## Plots

Confirm that:
- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a ‘group’ is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

## Methodology

| Sample preparation | Tissue culture samples were resuspended to single cells for staining. PBMCs were obtained from Leuko Paks collected by AllCells, cryopreserved, and thawed fresh for use. |
|--------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Instrument         | Data were acquired using a BD FACS Melody (analysis and sorting) and Beckman Coulter CytoFLEX LX (analysis).                                                                                           |
| Software           | Manual gating was used in FlowJo (10.5.3) for analysis.                                                                                                                                              |
| Cell population abundance | The sorted cell populations were limiting and prevented us from re-analyzing their sort purity. However, we routinely measured sorting purity using control samples and obtained >95% purity upon re-running the sorted samples when using the Purity sort precision mode on the FACS Melody. |
| Gating strategy    | The gating strategies used are provided in Supplementary Fig 41 and all applicable figures. In brief, our gating strategy was: FSC-A x SSC-A to capture the cell population of interest -> FSC-A x FSC-H to identify single cells -> DAPI x SSC-A to gate live cells. |

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.