Discovery of tumor-reactive T cell receptors by massively parallel library synthesis and screening

Moravec, Z.; Zhao, Y.; Voogd, R.; Cook, D.R.; Kinrot, S.; Capra, B.; ... ; Scheper, W.

nature biotechnology

Article

https://doi.org/10.1038/s41587-024-02210-6

Discovery of tumor-reactive T cell receptors by massively parallel library synthesis and screening

Received: 26 May 2023

Accepted: 18 March 2024

Published online: 23 April 2024

Check for updates

Ziva Moravec[1], Yue Zhao[2,13], Rhianne Voogd[1,13], Danielle R. Cook[3], Seon Kinrot[3], Benjamin Capra[3], Haiyan Yang[2], Brenda Raud[1], Jiayu Ou[2], Jiekun Xuan[3,4], Teng Wei[5], Lili Ren[5], Dandan Hu[6,7], Jun Wang[8,9], John B.A.G. Haanen[1,10,11], Ton N. Schumacher[1,11,12], Xi Chen[2,3,4] ✉, Ely Porter[3] ✉ & Wouter Scheper[1] ✉

T cell receptor (TCR) gene therapy is a potent form of cellular immunotherapy in which patient T cells are genetically engineered to express TCRs with defined tumor reactivity. However, the isolation of therapeutic TCRs is complicated by both the general scarcity of tumor-specific T cells among patient T cell repertoires and the patient-specific nature of T cell epitopes expressed on tumors. Here we describe a high-throughput, personalized TCR discovery pipeline that enables the assembly of complex synthetic TCR libraries in a one-pot reaction, followed by pooled expression in reporter T cells and functional genetic screening against patient-derived tumor or antigen-presenting cells. We applied the method to screen thousands of tumor-infiltrating lymphocyte (TIL)-derived TCRs from multiple patients and identified dozens of $CD4^+$ and $CD8^+$ T-cell-derived TCRs with potent tumor reactivity, including TCRs that recognized patient-specific neoantigens.

T cell receptor (TCR) gene therapy is a form of cellular immunotherapy in which peripheral blood T cells of patients with cancer are genetically engineered with tumor-specific TCRs in vitro before being reinfused into the patient. In contrast to immunotherapies such as immune checkpoint blockade[1] or adoptive transfer of ex vivo expanded tumor-infiltrating lymphocyte (TIL) therapy[2,3], the genetic transfer of therapeutic TCRs enables complete control over the specificity of the engineered T cell response, which depends strictly on the selected TCRs. Furthermore, generation of TCR gene-engineered T cells in vitro allows additional genetic modifications to endow T cell products with maximized in vivo activity and persistence[4]. However, the identification of patient-derived, therapeutically relevant TCRs is complicated by the fact that tumor-specific TCRs in many patients make up only a minority of peripheral or intratumoral TCR repertoires[5,6]. In addition, many of the most immunogenic tumor antigens, such as cancer neoantigens that arise as a consequence of patient-specific tumor mutations[7,8], are unique to individual patients, highlighting the need for truly personalized technologies for the discovery of therapeutic

[1]Department of Molecular Oncology and Immunology, Netherlands Cancer Institute, Amsterdam, The Netherlands. [2]RootPath, Inc. (Guangzhou), Guangzhou, China. [3]RootPath, Inc. (US), Watertown, MA, USA. [4]RootPath, Inc. (Hangzhou), Hangzhou, China. [5]Cytotherapy Laboratory, People's Hospital, Shenzhen, Guangdong, China. [6]Department of Liver Surgery, Sun Yat-sen University Cancer Center, Guangzhou, China. [7]State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangzhou, China. [8]Department of Pathology, New York University Grossman School of Medicine, New York, NY, USA. [9]Laura and Isaac Perlmutter Cancer Center, New York University Langone Health, New York, NY, USA. [10]Department of Medical Oncology, Netherlands Cancer Institute, Amsterdam, The Netherlands. [11]Department of Hematology, Leiden University Medical Center, Leiden, The Netherlands. [12]Oncode Institute, Utrecht, The Netherlands. [13]These authors contributed equally: Yue Zhao, Rhianne Voogd. ✉e-mail: xi@rootpath.com; elyporter@gmail.com; w.scheper@nki.nl

TCRs. Moreover, although the role of major histocompatibility complex (MHC) class I–restricted $CD8^+$ T cells in anti-tumor immunity is well established[9–11], recent data highlight important roles for MHC class II–restricted $CD4^+$ T cells in tumor control and response to immunotherapy, both through direct anti-tumor cytotoxicity and by boosting the activity of tumor-specific $CD8^+$ T cells[12–15]. Thus, technologies that enable the high-throughput discovery of both MHC class I–restricted and MHC class II–restricted TCRs on a per-patient basis are needed.

In the present study, we developed a personalized TCR discovery pipeline that couples large-scale assembly of synthetic TCR libraries with high-throughput genetic screening, thereby enabling the functional analysis of the specificities of thousands of individual TCRs in a single experiment. Notably, by generating synthetic libraries of patient-derived TCRs, this approach permits the assessment of TCR specificities in a manner that is not biased by the phenotypic fitness or clonal abundance of patient T cells in situ. We leveraged our technology to screen the tumor specificities of intratumoral T cell repertoires in multiple patients with cancer, and we identified MHC class I–restricted and MHC class II–restricted, patient-derived TCRs that recognized tumor-associated antigens, including neoantigens.

Results

Ultra-high-throughput pooled TCR gene synthesis

Methods for the generation of synthetic TCRs have traditionally relied on de novo synthesis of full-length TCRs, assembly of TCRs in individual reactions (for instance, using synthetic CDR3α/β fragments and reusable plasmid stocks for germline TRAV and TRBV[16–18]) or amplification of individual TCRs from complex TCR sequencing libraries[19]. However, both the low throughput and considerable costs (>$50 per TCR) prohibit the generation of synthetic TCR libraries at large scale. Emulsion-based approaches that amplify and clone native TCRs directly from polyclonal T cell populations have considerably increased the scale of TCR library generation[20,21], but they lack control over the presence and abundance of individual TCRs in resulting libraries and do not enable linking the function of a TCR to the phenotypic state of the T cell that it was isolated from. To address these limitations, we devised a new approach that takes advantage of both the reusability of germline-encoded TCR variable genes and the availability of inexpensive synthetic oligonucleotide pools. With this new approach, up to 1,000 unique TCRs are synthesized in a one-pot reaction (Fig. 1a and Extended Data Fig. 1a,b), making the effective material cost of synthesizing an individual TCR as low as $1. Fragments encoding the CDR3Jα and CDR3Jβ sequences of individual TCRs are prepared as separate single-stranded DNA (ssDNA) oligo pools. Each oligonucleotide in these pools contains a unique, orthogonal hybridization sequence (Zip) that enables correct pairing of CDR3Jα and CDR3Jβ oligonucleotides of individual TCRs. In addition, approximately 20-nucleotide (nt) connector sequences (ConnA and ConnB) are included to guide pairing of CDR3Jα and CDR3Jβ oligonucleotides with the relevant germline TRAV and TRBV alleles, respectively. Through a multi-step assembly process, this strategy results in a library of TCRs, where each molecule is composed of TRBV, TRBC and TRAV domains (termed 'VCV' fragments) coupled to the TCR-specific Zip sequence. The complete VCV pool can subsequently be amplified by common forward and reverse primers, or single TCRs can be selectively amplified using primers targeting their respective Zip sequences (Extended Data Fig. 1c). TCRs are then subcloned into an expression vector containing the TRAC domain directly downstream of the cloning site, such that the inserted VCV fragment is in frame with the TRAC sequence to form a complete TCR.

We validated the accuracy of our method by producing a proof-of-concept VCV pool encoding 553 unique TCRs (Supplementary Table 1) and characterized the assembly intermediates and products. We pre-designed 1,000 unique Zip sequences with sufficient orthogonality to enable highly specific, multiplexed hybridization of correct CDR3Jα and CDR3Jβ pairs (Supplementary Table 2), and we used these to assemble the proof-of-concept CDR3Jα and CDR3Jβ oligo pools. Analysis of the resultant CDR3Jα–CDR3Jβ pool by next-generation sequencing (NGS) revealed an average pairing accuracy (termed 'α–β pairing accuracy') of 97.4% (Fig. 1b and Supplementary Table 3), demonstrating that Zip sequences enable fragment assembly with extremely high fidelity. Of note, α–β pairing accuracy and frequency were not influenced by oligo sequence features, such as GC content or length, but assembly did lead to a slight decrease in library uniformity compared to the initial oligo pools, with the difference between the frequencies of resulting pairs at the 5th and 95th percentile increasing approximately six-fold (Extended Data Fig. 1d–f and Supplementary Table 4).

Although the sequence-specific incorporation of TRAV and TRBV genes is of lower plex (~50–150), the germline amino acid sequences of TRAV and TRBV heavily restrict the nucleotide sequence space that can be used to design the ConnA and ConnB sequences. This is further complicated by the highly homologous amino acid sequences across different variable gene alleles (Fig. 1c). We, therefore, developed a thermodynamics-driven codon diversification algorithm for the design of ConnA and ConnB sequence sets for both human and mouse TCRs that are predicted to be highly orthogonal under these constraints (Fig. 1d,e, Extended Data Fig. 2a,b and Supplementary Tables 5 and 6). NGS analysis after assembly of the proof-of-concept Vα–CDR3Jα–CDR3Jβ and VCV pools demonstrated exceptionally accurate acquisition of correct TRAV and TRBV genes in assembled sequences (Fig. 1f,g and Extended Data Fig. 2c,d). Specifically, we observed correct ConnA-mediated TRAV acquisition (termed 'TRAV acquisition accuracy') in, on average, 99.7% of resulting sequences and an average ConnB-mediated 'TRBV acquisition accuracy' of 99.8%. Similar to CDR3Jα and CDR3Jβ pairing, successive assembly steps modestly decreased library uniformity, with the difference between the frequencies of resulting Vα–CDR3Jα–CDR3Jβ and VCV sequences at the 5th and 95th percentile increasing to ~36-fold and ~48-fold, respectively (Supplementary Tables 7 and 8).

Finally, characterization of single-point mutations, insertions and deletions introduced during oligonucleotide synthesis and polymerase chain reaction (PCR) amplification demonstrated that 83.3% of TRAV_GL and TRBV_GL regions, 98.3% of ConnA and ConnB regions, 71.2% of CDR3Jα regions and 75.2% of CDR3Jβ regions were mutation free (Fig. 1h). The higher error rates for CDR3Jα and CDR3Jβ regions likely derive from the fact that these were synthesized as oligonucleotide pools, whereas TRAV_GL and TRBV_GL fragments were amplified from sequence-verified plasmids. By multiplying the different assembly accuracies and mutation-free rates, we estimate a 'perfect rate' of approximately 43% of all correctly assembled VCV molecules. This percentage is sufficient for the applications described in this work and will likely improve in the coming years due to continuous improvements in DNA synthesis technologies.

NGS-based functional profiling of synthetic TCR libraries

Conventional methods to test the specificity and functionality of single TCRs generally rely on measuring cytokine secretion, cytotoxicity or cellular activation of TCR-engineered T cells in individual experimental conditions[6,16–19,22] and, as such, are not suitable for differentiating responding TCRs from bystander TCRs within complex TCR libraries. Recent work demonstrated the feasibility of NGS-based identification of antigen-specific TCRs from large pools of TCR-modified T cells, subsequent to isolation of relevant cells based on expression of T cell activation markers after co-culture with antigen-presenting cells (APCs)[23,24]. Building on these approaches, we designed and benchmarked a pooled screening platform that couples the functional analysis of T cell activation to sensitive quantification of responding TCRs by NGS (Fig. 2a). In this method, TCR libraries are delivered in bulk to reporter T cells, and the resultant T cell library is incubated with APCs of interest. Activated T cells are isolated by fluorescence-activated cell sorting (FACS)

https://doi.org/10.1038/s41587-024-02210-6

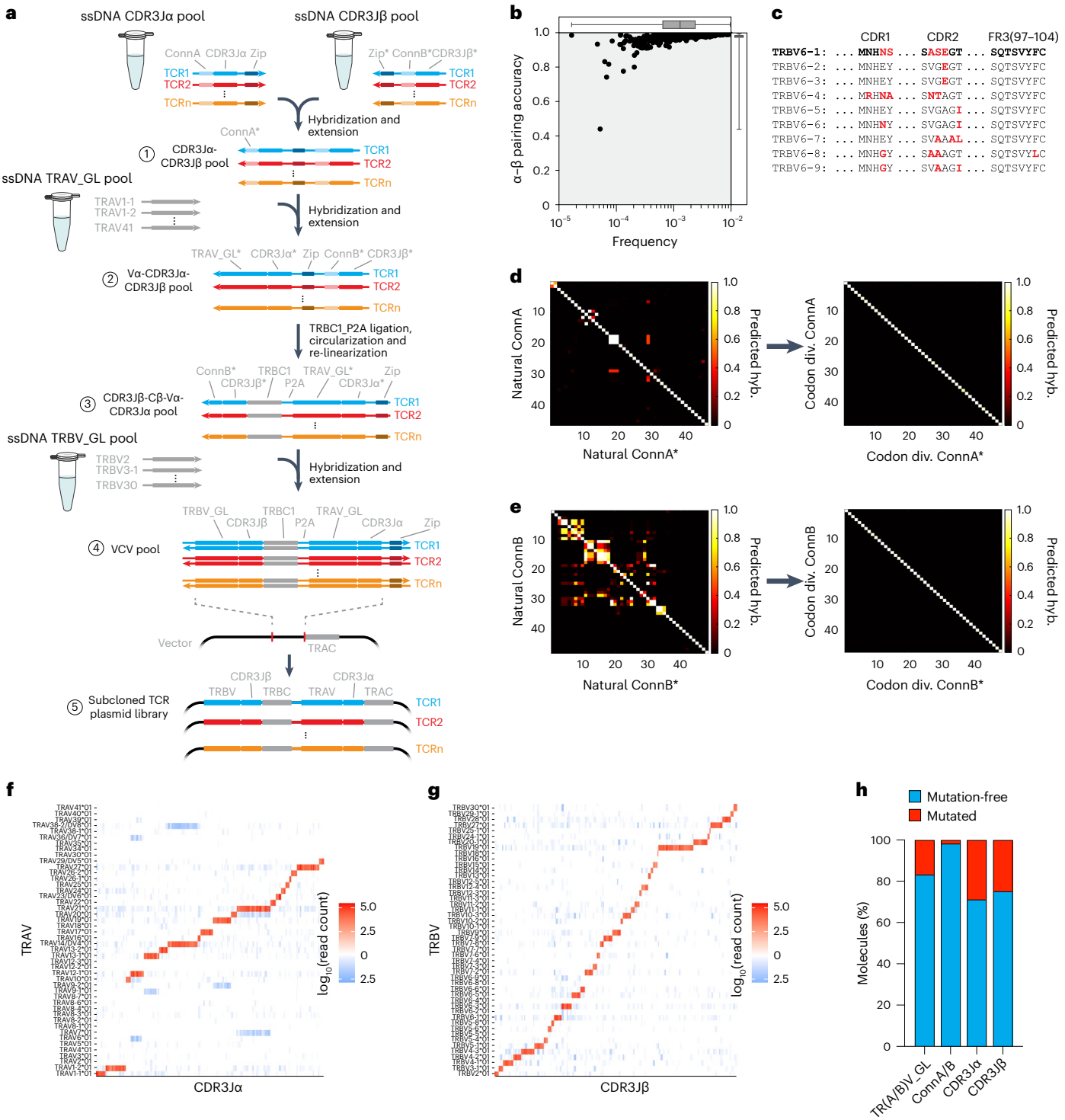


Fig. 1 | Schematic overview and performance of massively parallel TCR gene synthesis. a, Pooled TCR gene assembly scheme. Gray fragments indicate reusable TRBC1, TRAV_GL or TRBV_GL gene sequences (GL, germline). Arrows indicate 3′ of DNA strands. ConnA, TRAV gene-specific connector sequences; ConnB, TRBV gene-specific connector sequences; Zip, TCR-specific barcode sequence. Asterisks indicate reverse complement sequences. The products of individual assembly steps are indicated by circled numbers. **b**, Frequency and α–β pairing accuracy of Zip-mediated CDR3Jα–CDR3Jβ fragments of the proof-of-concept TCR library (n = 553 TCRs). Dots represent individual CDR3Jα–CDR3Jβ fragments. Box plots depict the median, the interquartile range and whiskers extending to minimal and maximal values. **c**, Framework 3 (FR3) regions of TRAV and TRBV protein sequences may be highly homologous. CDR1, CDR3 and FR3 sequences of human TRBV6-1 to TRBV6-9 are shown as examples. **d**, Computer-predicted orthogonality of natural (left) and codon-diversified (right) FR3 connector sequences for TRAV (ConnA). **e**, Computer-predicted orthogonality of natural (left) and codon-diversified (right) FR3 connector sequences for TRBV (ConnB). **f**, Accuracy of TRAV ligation to individual CDR3Jα sequences within the proof-of-concept library. The heatmap depicts $\log_{10}$-transformed read counts of all possible TRAV–CDR3Jα combinations within the library. Red and blue colors indicate on-target and off-target ligations, respectively. **g**, Accuracy of TRBV ligation to individual CDR3Jβ sequences within the proof-of-concept library. Data are depicted as in **f**. **h**, Fractions of mutation-free sequences for the indicated assembly fragments of the proof-of-concept library. hyb., hybridization; div., diversified.

https://doi.org/10.1038/s41587-024-02210-6

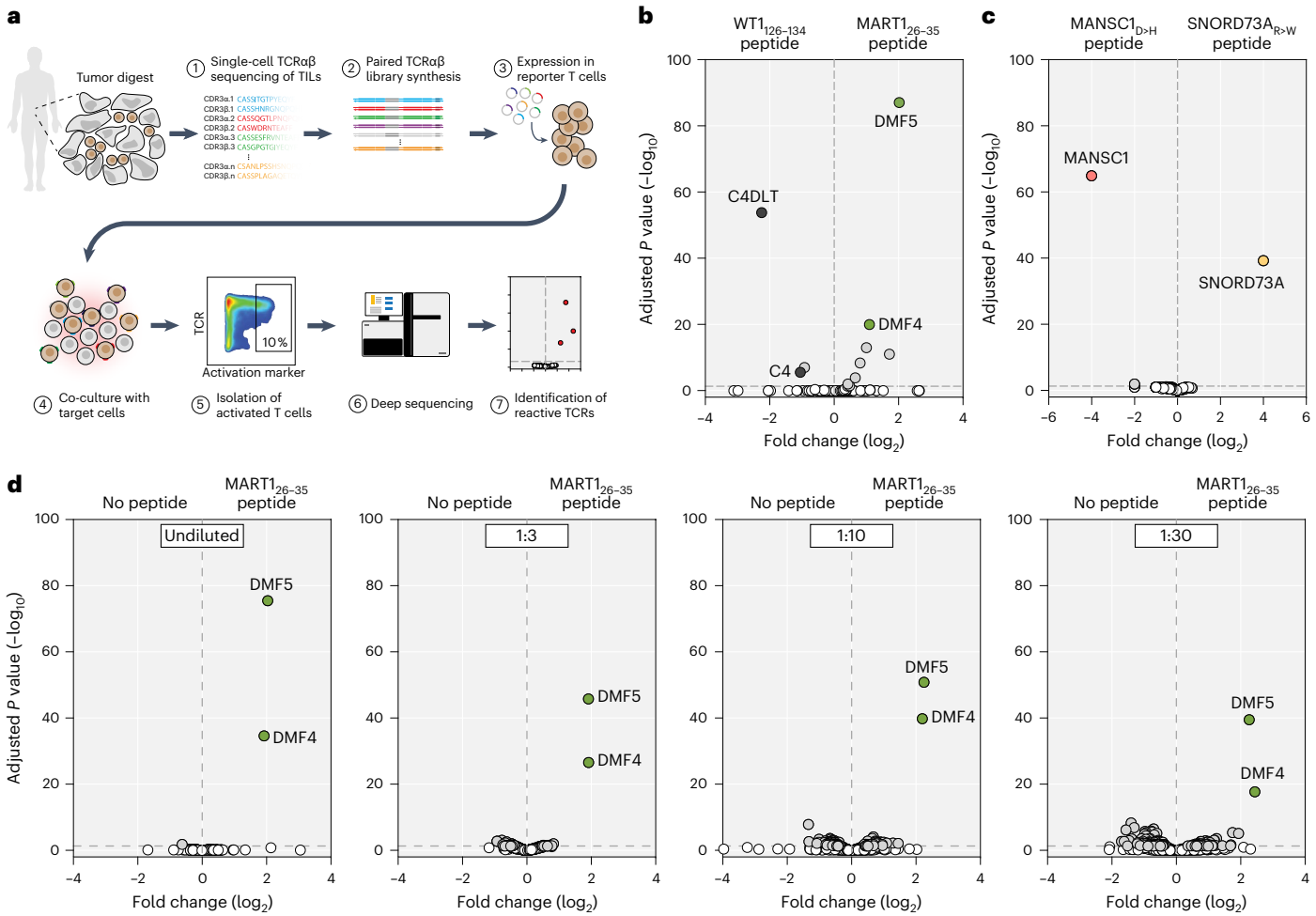


Fig. 2 | Overview and validation of large-scale functional profiling of synthetic TCR libraries. a, Schematic overview of the screening methodology. **b**, $CD8^+$ Jurkat cells were engineered to express the NKIRTIL063 $CD8^+$ TIL-derived TCR library ($n = 935$ TCRs), which included control TCRs specific for $MART1_{26–35}$ (TCRs DMF4 and DMF5) and $WT1_{126–134}$ (TCRs C4 and C4DLT). Library-expressing Jurkat cells were screened against immortalized HLA-A*02:01$^+$ B cells pulsed with $MART1_{26–35}$ or $WT1_{126–134}$ peptide. Dots represent individual TCRs. Fold change represents the relative abundance of TCRs in $WT1_{126–134}$ and $MART1_{26–35}$ peptide-pulsed conditions. **c**, $CD4^+$ Jurkat cells were transduced with the OVC190 $CD4^+$ TIL-derived TCR library (2,900 unique TCR sequences) that included MHC class II–restricted $MANSC1_{D85H}$ and $SNORD73A_{R165W}$ neoantigen-specific TCRs as internal controls. The Jurkat library was screened against patient-matched immortalized B cells pulsed with $MANSC1_{D85H}$ or $SNORD73A_{R165W}$ peptide. Data are plotted as in **b**. **d**, NKIRTIL063 library-expressing $CD8^+$ Jurkat cells were diluted 3-fold, 10-fold and 30-fold with mock-transduced cells and screened against $MART1_{26–35}$ peptide-pulsed HLA-A*02:01$^+$ B cells. Data are plotted as in **b**. In all plots, screen hits were defined as outlined in the Methods and are marked by colored dots. *P* values were generated using the DESeq2 Wald test and adjusted for several comparisons.

or magnetic-activated cell sorting (MACS) based on the surface expression of activation markers, and transgenic TCRs of isolated cells are then amplified by PCR and quantified using NGS. To determine the feasibility of this approach, we subjected $CD8^+$ TILs from the tumor of melanoma patient NKIRTIL063 to single-cell TCR sequencing (Extended Data Fig. 3), and we designed and synthesized a personalized TCR library that encoded all identified TCRs (928 unique TCRs among 4,542 sequenced $CD8^+$ TILs; Supplementary Table 9). In addition, several well-characterized HLA-A*02:01-restricted TCRs specific for common tumor-associated antigens and of diverse affinities were included in the library as internal controls. To allow the one-step analysis of TCR assembly quality of the final VCV pool, rather than characterizing assembly intermediates as above, we developed a custom NGS method to enable analysis of full-length VCV fragments (Extended Data Fig. 4a and Methods). Applying this method to the NKIRTIL063 VCV pool showed average overall assembly accuracy (as defined in the Methods) and mutation-free rates of 89.1% and 53.9%, respectively (Extended Data Fig. 4b,c and Supplementary Table 10). The NKIRTIL063 TCR library was retrovirally transduced into $CD8^+$ Jurkat cells, and 99.9% of successfully assembled TCR sequences yielded cell-surface-expressed TCRs (Extended Data Fig. 4d). To assess the sensitivity of our screening method, we first took advantage of the included control TCRs and evaluated screening performance against an HLA-A*02:01-positive immortalized B cell line pulsed with the cognate peptide of either the MART1-specific model TCRs DMF4 and DMF5 or the WT1-specific TCRs C4 and C4DLT. Screening the NKIRTIL063 TCR library against the two pulsed B cell lines revealed clear enrichment of TCRs DMF4 and DMF5 among activated T cells after incubation with MART1 peptide-pulsed cells and enrichment of TCRs C4 and C4DLT in response to WT1 peptide-pulsed cells (Fig. 2b). Notably, the high-affinity DMF5 and C4DLT TCRs displayed more pronounced enrichment as compared to the lower-affinity DMF4 and C4 TCRs, demonstrating that the strength of T cell–target cell interactions may be gauged from screening data.

Next, to evaluate the suitability of our screening platform for the discovery of MHC class II–restricted TCRs, we synthesized a library of 1,341 unique TCRs that encoded all TCRs that were identified by single-cell sequencing of $CD4^+$ TILs isolated from an ovarian tumor (patient OVC190; Extended Data Fig. 5 and Supplementary Tables 11–16)

and expressed the library in CD4$^+$ Jurkat cells. Two MHC class II–restricted neoantigen-specific TCRs that were previously isolated from TILs of a patient with melanoma[25] were included in this library as internal controls. Screening the library against patient-matched immortalized B cells pulsed with either of the cognate peptides of both control TCRs demonstrated clear and exclusive enrichment of the two neoantigen-specific TCRs in response to their cognate neoantigens (Fig. 2c).

Given that abundances of individual TCRs in synthesized libraries may vary, we next assessed the sensitivity of our screening method in conditions where antigen-specific TCRs are represented at lower-than-average frequencies. To this end, Jurkat cells expressing the NKIRTIL063 library were mixed with mock-transduced cells such that individual NKIRTIL063 TCRs were present at frequencies of, on average, 1×10^{-3}, 0.3×10^{-3}, 1×10^{-4} or 0.3×10^{-4} of total cells. Screening of resulting populations against MART1 peptide-pulsed B cells clearly identified both MART1-specific TCRs, including the low-affinity DMF4 TCR, demonstrating that our method enables detection of antigen-specific TCRs even in settings where these are present at low abundance (Fig. 2d). Finally, we validated that screening performance was not affected by the type of reporter T cells used (Jurkat cells or primary T cells) nor by alternative methods for isolation of activated T cells (FACS or MACS; Supplementary Fig. 1), demonstrating the flexibility and robustness of our method across different experimental procedures. Collectively, these data demonstrate that our pooled TCR screening technology enables the identification of antigen-specific TCRs from complex TCR libraries. Moreover, the method allows one to distinguish TCRs of different affinities and to screen the specificities of both MHC class I–restricted and MHC class II–restricted TCRs.

Personalized tumor-specific TCR discovery

To assess the feasibility of our TCR screening method for the personalized identification of therapeutically relevant TCRs, we next set out to identify tumor-specific TCRs within the NKIRTIL063 library. As a first approach, we aimed to broadly survey the anti-tumor repertoire within the patient library in an antigen-agnostic manner. We, therefore, screened the TCR library against two allogeneic melanoma cell lines that share expression of the HLA-A*02:01 allele with patient NKIRTIL063. Screening against cell lines RPMI7951 and Malme3M identified multiple TCRs (RPMI7951: 48 TCRs; Malme3M: 81 TCRs) that were significantly enriched (Fig. 3a), suggesting a broadly tumor-reactive TCR repertoire among TILs of this patient. To validate these observations, a broad array of responding and non-responding TCRs ($n = 62$ and $n = 33$, respectively) from the Malme3M screen were PCR amplified from the library using their unique Zip barcode and expressed in donor T cells. Functional testing of resulting single TCR-engineered T cells confirmed MHC class I–dependent recognition of Malme3M cells by 31 of 62 (50%) responding TCRs (Fig. 3b,c). To further validate the feasibility of this approach, we profiled tumor reactivity of TIL-derived TCRs in an additional patient. TILs from cervical cancer patient CV19 were subjected to single-cell TCR sequencing, and a library encoding all identified TCRs ($n = 1{,}501$) was synthesized (Extended Data Fig. 6 and Supplementary Tables 17–20). In parallel, the human papillomavirus (HPV)-positive cervical cancer cell line SiHa was engineered to express all six MHC class I alleles of patient CV19. Screening the library against MHC-modified SiHa cells identified 34 responding TCRs (Extended Data Fig. 7a). In addition, screening against HPV E7-expressing or E6-expressing cells revealed reactivity to E7 for three TCRs (Extended Data Fig. 7b,c). Functional testing of four individual hit TCRs, including two E7-reactive TCRs, against SiHa and K562-E7 cells confirmed reactivity of two TCRs against HPV E7 and of two TCRs against an unidentified SiHa-expressed tumor-associated antigen (Extended Data Fig. 7d–f). Notably, T cells engineered with SiHa-reactive, but not HPV E7-reactive, TCRs were capable of recognizing autologous tumor digest of patient CV19 (Extended Data Fig. 7g).

Identification of neoantigen-specific TCRs

We next leveraged our screening method for the personalized identification of neoantigen-specific TCRs within the NKIRTIL063 library. We previously identified non-synonymous tumor mutations of patient NKIRTIL063 and designed and expressed tandem minigene (TMG) constructs that collectively encode the 200 most highly expressed mutations (out of a total of 685) in immortalized patient NKIRTIL063 B cells[25]. Notably, because such patient-matched B cells are fully MHC proficient, the use of these cells enables the screening of TCRs across the patient's complete MHC haplotype. The resulting 20 TMG-expressing B cell lines were combined into five pools and used to screen the patient TCR library. This yielded 32 significantly enriched patient TCRs responding to diverse TMG pools (Fig. 3d and Supplementary Fig. 2). Expression of isolated hit TCRs in CD8$^+$ Jurkat cells and co-culture with individual TMG lines confirmed reactivity of 22 out of 32 (69%) TCRs against TMGs 3, 7, 8 and 16 (Fig. 3e). In line with our sensitivity benchmarking (Fig. 2d), reactive TCRs could be identified irrespective of their frequency and assembly accuracy within the total library (Extended Data Fig. 8). Deconvolution of recognized TMG constructs identified four neoantigens ($GFPT2_{A676V}$, $CCSER_{P329L}$, $TNFAIP2_{P348A}$ and $ADAM10_{P164L}$) as the cognate targets of screen-identified TCRs (Fig. 3f). Moreover, incubation of selected TCR-engineered primary T cells with the patient's melanoma line demonstrated the capacity of five out of six tested TCRs to recognize and kill patient tumor cells (Fig. 3g).

To evaluate whether the throughput of our screening method permits neoantigen-specific TCR discovery in a cancer type in which tumor-specific TILs are expectedly less frequent compared to melanoma, we focused on ovarian cancer patient OVC190. In addition to the library encoding TCRs identified among CD4$^+$ TIL from this patient (1,341 TCRs; see above), we synthesized a TCR library encoding all 274 CD8$^+$ TIL clonotypes identified in the patient's tumor. Paired exome and RNA sequencing revealed 61 expressed non-synonymous tumor variants, which were encoded in eight different TMG constructs and expressed in immortalized patient-matched B cells. Screening of the CD4$^+$ T-cell-derived library against pooled TMG-expressing B cell lines yielded two putatively neoantigen-reactive TCRs (TCR 198.1 and TCR 576.3; Fig. 4a). Although TCR 97.2 and TCR 184.1 did not reach statistical significance, they were included for validation. Expression of hit TCRs in CD4$^+$ Jurkat cells and incubation with single TMG-expressing B cells demonstrated reactivity of three out of four identified TCRs toward TMG 5 (Fig. 4b), and subsequent TMG deconvolution experiments using single minigene-expressing B cells demonstrated that all three TCRs recognized an $MTREX_{D398A}$ neoantigen (Fig. 4c and Extended Data Fig. 9a). In addition, TCR-engineered donor CD4$^+$ T cells displayed cytotoxicity toward $MTREX_{D398A}$-expressing B cells but not cells expressing the wild-type MTREX sequence (Fig. 4d). No neoantigen-specific TCRs were identified among CD8$^+$ TIL-derived TCRs in this patient's tumor (Extended Data Fig. 9b).

Taken together, these data demonstrate the feasibility of using our TCR screening technology for the personalized discovery of tumor-specific and neoantigen-specific TCRs in patients with cancer, including patients with cancers in which tumor-specific TILs are generally rare.

Phenotype of tumor-reactive T cells

Recent work proposed gene expression signatures for the prospective identification of tumor-specific CD4$^+$ and CD8$^+$ T cells[26], and we, therefore, questioned whether these signatures would effectively identify the tumor-specific TCRs identified using our screening platform. Because the method of processing NKIRTIL063 tumor material precluded reliable profiling of TIL transcriptomes, we initially focused our analysis on the OVC190 tumor. NeoTCR4 and NeoTCR8 signature scores of OVC190 CD4$^+$ and CD8$^+$ TIL clonotypes were derived from single-cell gene expression data. Neoantigen-specific clonotypes and NeoTCR signature-positive cells mapped to clusters of CD4$^+$ TILs with

https://doi.org/10.1038/s41587-024-02210-6

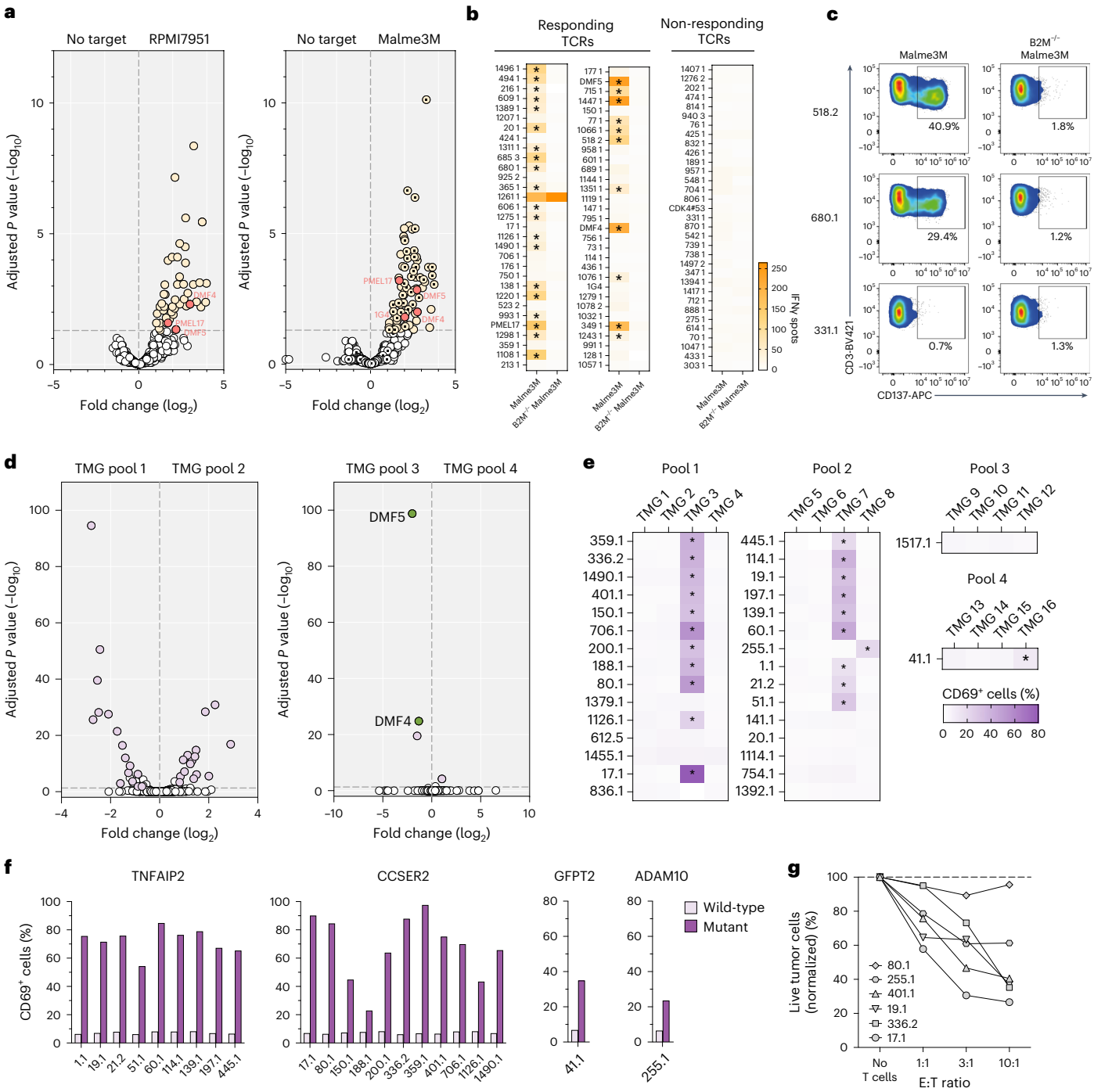


Fig. 3 | Personalized mining of tumor-reactive TCRs from melanoma TILs. **a**, The NKIRTIL063 TCR library (n = 935 TCRs) was expressed in CD8$^+$ T cells and screened against the melanoma lines RPMI7951 and Malme3M. Fold change represents the relative abundance of TCRs after incubating T cells with or without the lines. Screen hits were defined as in Fig. 2. TCRs selected for validation are marked with dots. **b**, Malme3M-reactive TCRs identified in **a** were individually expressed in CD3$^+$ T cells, and their activation in response to Malme3M cells was measured by IFNγ ELISpot. β_2-microglobulin-deficient Malme3M served as negative control. Heatmap depicts the number of IFNγ spots. Responding TCRs are ordered according to increasing adjusted P values. Asterisks mark Malme3M-reactive TCRs (as defined in the Methods). **c**, Flow cytometric data of a strong, medium and non-responder TCR after co-culture with Malme3M cells, gated on live CD3$^+$ cells. **d**, NKIRTIL063 immortalized B cells were transduced with TMG constructs encoding the patient's 200 most highly expressed non-synonymous tumor mutations. TMG B cell lines were combined into five pools of four lines and used to screen the NKIRTIL063 TCR library. MART1$_{26–35}$ minigene-expressing B cells were included in TMG pool 3. Dots represent individual TCRs. Fold change represents the relative abundance of TCRs in cultures with the indicated TMG pools. **e**, TMG-reactive TCRs identified in **d** were individually expressed in CD8$^+$ Jurkat cells and cultured with TMG-expressing NKIRTIL063 B cells. Activation of TCR-transduced cells was assessed by measuring CD69 expression using flow cytometry. Heatmap depicts the percentage of CD69$^+$ cells within live mTCRβ$^+$ cells. Within TMG pools, TCRs are ordered according to increasing adjusted P values. Asterisks mark TMGs with $z > 1$. **f**, Neoantigen specificity of TMG-reactive TCRs was determined by incubating TCR-transduced Jurkat cells with TMG constructs in which indicated mutant minigenes were reverted to their wild-type sequence. T cell activation was evaluated by measuring CD69 expression using flow cytometry. **g**, Cytotoxicity toward the NKIRTIL063 cell line was measured after incubating TCR$^+$ T cells with the tumor line at the indicated effector-to-target ratios.

https://doi.org/10.1038/s41587-024-02210-6

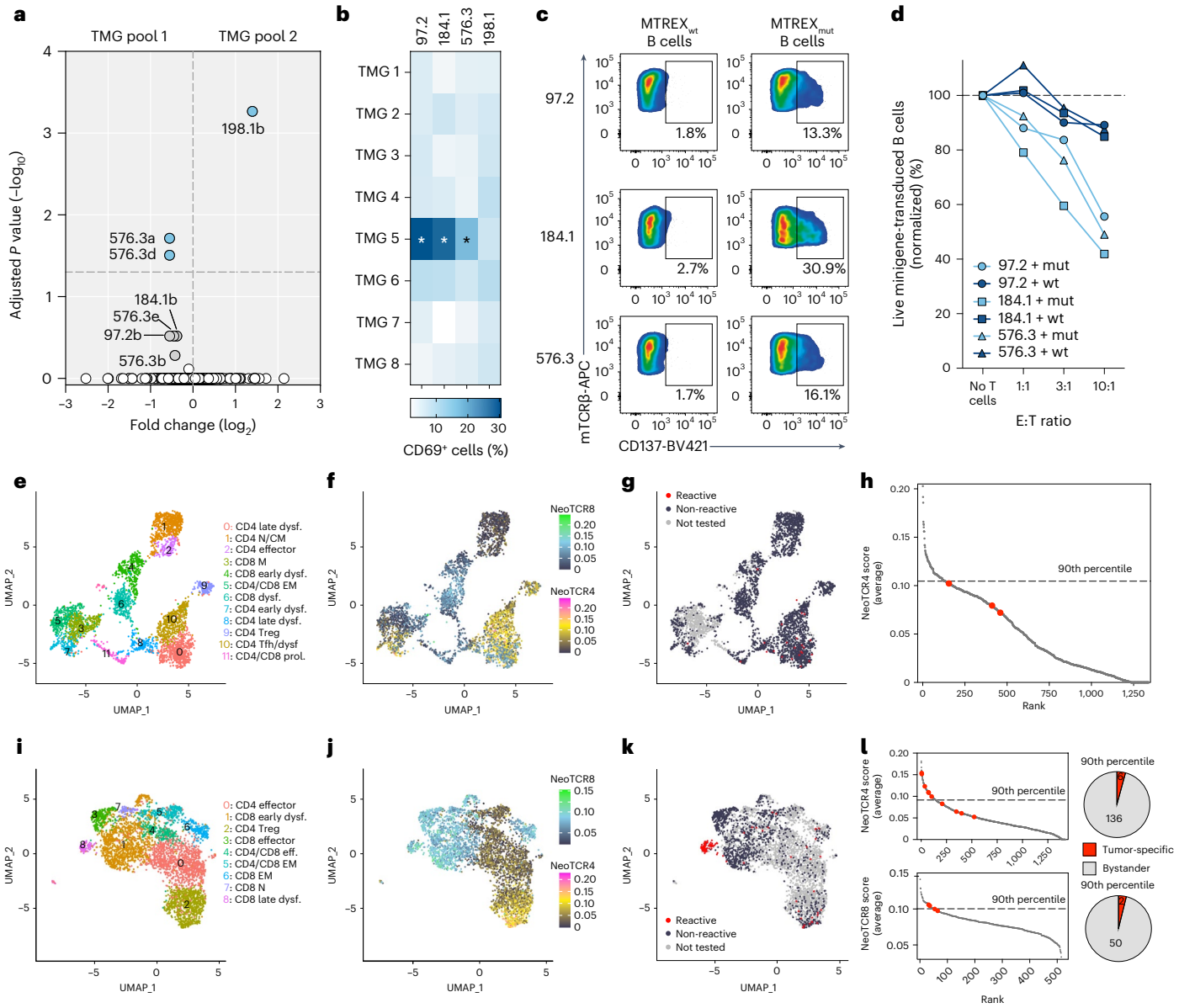


Fig. 4 | Phenotype of tumor-specific T cells. a, CD4$^+$ Jurkat cells were transduced with the CD4$^+$ TIL-derived library of patient OVC190 (n = 1,341 TCRs encoded as 2,900 unique nucleotide sequences). Patient immortalized B cells were engineered to express TMGs encoding all expressed non-synonymous tumor mutations (n = 61) and used to screen the OVC190 TCR library. Dots represent individual TCRs. Codon-diversified replicates of individual TCRs are designated by suffixes 'a–e'. Screen hits and P values were defined as in Fig. 2. **b**, Hit TCRs identified in **a** were individually expressed in CD4$^+$ Jurkat cells and co-cultured with TMG-expressing patient B cells. Heatmap depicts the percentage of CD69$^+$ cells within live mTCRβ$^+$ cells. Asterisks indicate conditions with $z > 2$. **c**, CD4$^+$ T cells transduced with neoantigen-specific OVC190 TCRs were incubated with patient B cells expressing the MTREX$_{D398A}$ neoantigen or its wild-type sequence, and T cell activation was assessed by measuring CD137 expression by flow cytometry. **d**, OVC190 TCR-engineered donor CD4$^+$ T cells were incubated with patient B cells expressing the mutant or wild-type MTREX sequence at the indicated effector-to-target ratios. The fraction of remaining live B cells after co-culture was measured by flow cytometry. **e**, UMAP plot of single-cell expression data from OVC190 TILs. **f**, Projection of NeoTCR4 and NeoTCR8 scores of OVC190 TILs onto the transcriptomic map. **g**, Projection of neoantigen-specific clonotypes onto the transcriptomic map. **h**, Ranking of OVC190 TIL clonotypes based on NeoTCR4 scores. Neoantigen-specific clonotypes are highlighted in red. **i**, UMAP of single-cell transcriptomic data from HC25 TILs. **j**, Projection of NeoTCR4 and NeoTCR8 scores of HC25 TILs onto the transcriptomic map. **k**, Projection of tumor-reactive clonotypes onto the transcriptomic map. **l**, Ranking of HC25 CD4$^+$ (top) and CD8$^+$ (bottom) TIL clonotypes based on NeoTCR4 and NeoTCR8 scores, respectively. Tumor-specific clonotypes are highlighted in red. Pie charts depict the fraction of tumor-specific CD4$^+$ (top) and CD8$^+$ (bottom) TIL-derived clonotypes among the 90th percentile of NeoTCR4 and NeoTCR8 scoring clonotypes, respectively. CM, central memory; dysf., dysfunctional; eff., effector; EM, effector memory; mut, mutant; N, naive-like; M, memory; prol., proliferating; wt, wild-type.

dysfunctional and Treg phenotypes (Fig. 4e–g). Surprisingly, however, only one of the three neoantigen-specific TCRs (TCR 576.3) ranked within the top 250 scoring clonotypes, and none ranked within the top 10% (90th percentile) of signature scoring clonotypes (Fig. 4h). Lack of MHC class II expression on OVC190 tumor cells (Extended Data Fig. 9c) did not allow us to functionally test whether CD4$^+$ TIL-derived TCRs may recognize tumor-expressed antigens other than neoantigens (such as non-mutated differentiation or cancer testis antigens). Although it, therefore, cannot be excluded that top signature scoring clones may be specific against other classes of tumor antigens, these data suggest that neoantigen-specific TCRs may be obtained from TILs that express only low levels of reactivity-associated signatures and, thus, may only

be identified using approaches that enable a comprehensive screening of TIL repertoires, as described here. Of note, screening the OVC190 $CD8^+$ T-cell-derived library against primary OVC190 tumor cells did not identify tumor-reactive TCRs (Extended Data Fig. 9d).

To more directly assess the capacity of these transcriptional signatures to identify tumor-reactive T cells, we additionally profiled the tumor specificities of TILs from a hepatocellular carcinoma specimen (patient HC25; Supplementary Table 21). Here, $PD1^+$ TILs were isolated to enrich for T cells presumed to be tumor reactive[9,27], followed by single-cell RNA and TCR sequencing. NeoTCR4 and NeoTCR8 scores were derived for $CD4^+$ and $CD8^+$ TIL-derived clonotypes, respectively, revealing highest levels of signature expression in dysfunctional $CD8^+$ T cells and $CD4^+$ regulatory T cells (Tregs) (Fig. 4i,j). Next, functional screening of the 382 highest-scoring clonotypes against dissociated autologous tumor cells yielded as few as 22 tumor-reactive clonotypes (Extended Data Fig. 10), suggesting that a large majority of signature-positive TILs in this specimen are non-tumor-specific bystanders. Interestingly, the transcriptional states of tumor-reactive clonotypes converged to a late dysfunctional phenotype (Fig. 4k). Notably, within the 90th percentile of signature scoring MHC class I–restricted and MHC class II–restricted clonotypes, only two (out of 52) and six (out of 142) TCRs, respectively, displayed tumor reactivity (Fig. 4l). Taken together, these observations underline the value of comprehensive, functional TCR screening approaches for the unbiased identification of bona fide tumor-reactive TCRs.

Discussion

The identification of therapeutic TCRs for personalized TCR gene therapy is complicated by the patient-specific and highly diverse nature of TCR repertoires in individual patients with cancer. Moreover, tumor-specific T cells may be exceedingly rare, in particular in tumors that lack robust endogenous T cell reactivity[5,6]. To date, antigen-specific TCR discovery methods have generally relied on the low-throughput assembly and expression of single TCRs, followed by analysis of their reactivity in separate experimental conditions[6,16–19,22]. However, such approaches are costly and poorly scalable and are, therefore, not suited for the screening of complex TCR repertoires. More recent strategies capable of isolating natively paired TCRαβ chains from polyclonal T cells[20,21] provide the opportunity to screen reactivities of large numbers of TCRs, for example when combined with NGS-based analysis of responding TCRs[23,24], but lack control over the fidelity or frequencies of individual TCRs within the generated libraries.

To address these limitations, we developed a personalized TCR discovery pipeline that couples large-scale assembly of synthetic TCR libraries with functional genetic screening, enabling the profiling of thousands of individual TCRs in a single experiment. Our 'sequence-and-synthesize' approach results in TCR libraries with relatively uniform representation of individual TCRs and, in contrast to approaches that rely on primary patient-derived T cells[28], enables the functional analysis of TCR repertoires in a manner that is not biased by the phenotype or clonal abundance of intratumoral T cells. A detailed analysis of the TCR sequence fidelity of synthesized libraries demonstrated correct assembly rates and 'perfect rates' of, on average, approximately 90% and approximately 40% across libraries, respectively, which is sufficient for successful identification of antigen-specific TCRs, including those of relatively low affinity. Because the highest fraction of sequence errors originated from the synthesized CDR3Jα and CDR3Jβ oligonucleotide pools, we expect that the sequence fidelity of TCR libraries synthesized using our method will improve further with future advances in commercial DNA synthesis technology. Although it is, in principle, feasible to design well more than 1,000 orthogonal Zip sequences to enable pairing of more complex CDR3Jα–CDR3Jβ oligonucleotide pools, such increased complexities would require undesirably prolonged reaction times. To enable scalability while safeguarding rapid turnaround times, we, therefore, limit the number of unique TCRs per one-pot assembly to 1,000 TCRs and straightforwardly synthesize libraries of multiple thousands of TCRs in parallel one-pot reactions.

We leveraged our technology to synthesize and screen approximately 5,000 unique synthetic TCRs across four patient samples. In addition, we demonstrate the utility of our NGS-based pooled functional screening approach for the identification of both MHC class I– and MHC class II–restricted tumor-specific TCRs from tumor-infiltrating T cell repertoires, including TCRs specific for cancer neoantigens. Recent efforts that compared the transcriptional phenotypes of tumor-specific and bystander T cells have led to the identification of tumor reactivity-associated gene programs[26], and such signatures could facilitate the relatively straightforward isolation of tumor-specific T cells and their TCRs. However, the lack of scalable and affordable TCR screening technologies has so far made it difficult to validate the performance of these signatures. In an initial effort in two tumor specimens, we benchmarked the identification of tumor-specific T cell clonotypes using such gene signatures against our functional screening method. In doing so, we observed low frequencies of tumor-specific TCRs among clonotypes that expressed high levels of reactivity-associated signatures, suggesting that bystander T cells may still be common among signature-positive TILs. More notably, our data indicate that truly tumor-reactive TCRs may be isolated from TILs that express such transcriptional signatures at only low levels, indicating that a comprehensive understanding of the tumor-reactive TCR pool requires the use of technologies that allow functional profiling of TCRs at scale, as described here. Of note, given the exploratory nature of this effort, it will be of interest to corroborate these findings in larger patient cohorts.

Looking forward, our method should facilitate the identification of tumor antigen-specific TCRs for use in the next generation of personalized TCR gene therapies. Using carefully streamlined, Good Laboratory Practice (GLP)-compliant workflows, we project that our pipeline will enable therapeutic TCR discovery within 4–5 weeks (from patient tissue onboarding to validated TCRs; Supplementary Fig. 3). Within the landscape of recent efforts to design personalized (neo) antigen-specific immunotherapies, in which the TCR or tumor antigen discovery so far may take from 3 months to up to 5 months[29–31], this constitutes an important advance. Thus, the method presented here represents a powerful tool for the design of future cellular immunotherapies.

Online content

Any methods, additional references, Nature Portfolio 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/s41587-024-02210-6.

References

1. Ribas, A. & Wolchok, J. D. Cancer immunotherapy using checkpoint blockade. *Science* **359**, 1350–1355 (2018).
2. Rohaan, M. W. et al. Tumor-infiltrating lymphocyte therapy or ipilimumab in advanced melanoma. *N. Engl. J. Med.* **387**, 2113–2125 (2022).
3. Rosenberg, S. A. & Restifo, N. P. Adoptive cell transfer as personalized immunotherapy for human cancer. *Science* **348**, 62–68 (2015).
4. Klebanoff, C. A., Chandran, S. S., Baker, B. M., Quezada, S. A. & Ribas, A. T cell receptor therapeutics: immunological targeting of the intracellular cancer proteome. *Nat. Rev. Drug Discov.* **22**, 996–1017 (2023).
5. Simoni, Y. et al. Bystander $CD8^+$ T cells are abundant and phenotypically distinct in human tumour infiltrates. *Nature* **557**, 575–579 (2018).

6. Scheper, W. et al. Low and variable tumor reactivity of the intratumoral TCR repertoire in human cancers. *Nat. Med.* **25**, 89–94 (2019).
7. Schumacher, T. N., Scheper, W. & Kvistborg, P. Cancer neoantigens. *Annu. Rev. Immunol.* **37**, 173–200 (2019).
8. Schumacher, T. N. & Schreiber, R. D. Neoantigens in cancer immunotherapy. *Science* **348**, 69–74 (2015).
9. Gros, A. et al. Prospective identification of neoantigen-specific lymphocytes in the peripheral blood of melanoma patients. *Nat. Med.* **22**, 433–438 (2016).
10. Gubin, M. M. et al. Checkpoint blockade cancer immunotherapy targets tumour-specific mutant antigens. *Nature* **515**, 577–581 (2014).
11. Rizvi, N. A. et al. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* **348**, 124–128 (2015).
12. Alspach, E. et al. MHC-II neoantigens shape tumour immunity and response to immunotherapy. *Nature* **574**, 696–701 (2019).
13. Oh, D. Y. et al. Intratumoral $CD4^+$ T cells mediate anti-tumor cytotoxicity in human bladder cancer. *Cell* **181**, 1612–1625 (2020).
14. Tran, E. et al. Cancer immunotherapy based on mutation-specific $CD4^+$ T cells in a patient with epithelial cancer. *Science* **344**, 641–645 (2014).
15. Borst, J., Ahrends, T., Babala, N., Melief, C. J. M. & Kastenmuller, W. $CD4^+$ T cell help in cancer immunology and immunotherapy. *Nat. Rev. Immunol.* **18**, 635–647 (2018).
16. 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).
17. Guo, X. Z. 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).
18. Zong, S. et al. Very rapid cloning, expression and identifying specificity of T-cell receptors for T-cell engineering. *PLoS ONE* **15**, e0228112 (2020).
19. Genolet, R. et al. TCR sequencing and cloning methods for repertoire analysis and isolation of tumor-reactive TCRs. *Cell Rep. Methods* **3**, 100459 (2023).
20. Fahad, A. S. et al. Immortalization and functional screening of natively paired human T cell receptor repertoires. *Protein Eng. Des. Sel.* **35**, gzab034 (2022).
21. Spindler, M. J. et al. Massively parallel interrogation and mining of natively paired human TCRαβ repertoires. *Nat. Biotechnol.* **38**, 609–619 (2020).
22. Muller, T. R. et al. A T-cell reporter platform for high-throughput and reliable investigation of TCR function and biology. *Clin. Transl. Immunol.* **9**, e1216 (2020).
23. Fahad, A. S. et al. Cell activation-based screening of natively paired human T cell receptor repertoires. *Sci. Rep.* **13**, 8011 (2023).
24. 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 (2022).
25. Cattaneo, C. M. et al. Identification of patient-specific $CD4^+$ and $CD8^+$ T cell neoantigens through HLA-unbiased genetic screens. *Nat. Biotechnol.* **41**, 783–787 (2023).
26. Lowery, F. J. et al. Molecular signatures of antitumor neoantigen-reactive T cells from metastatic human cancers. *Science* **375**, 877–884 (2022).
27. Gros, A. et al. PD-1 identifies the patient-specific $CD8^+$ tumor-reactive repertoire infiltrating human tumors. *J. Clin. Invest.* **124**, 2246–2259 (2014).
28. Arnaud, M. et al. Sensitive identification of neoantigens and cognate TCRs in human solid tumors. *Nat. Biotechnol.* **40**, 656–660 (2022).
29. Foy, S. P. et al. Non-viral precision T cell receptor replacement for personalized cell therapy. *Nature* **615**, 687–696 (2022).
30. Hilf, N. et al. Actively personalized vaccination trial for newly diagnosed glioblastoma. *Nature* **565**, 240–245 (2019).
31. Ott, P. A. et al. An immunogenic personal neoantigen vaccine for patients with melanoma. *Nature* **547**, 217–221 (2017).

Methods

Patient material

Tumor tissue and peripheral blood mononuclear cells (PBMCs) were collected from patients treated at the Netherlands Cancer Institute–Antoni van Leeuwenhoek Hospital (NKI-AVL; Amsterdam, The Netherlands), Shenzhen People's Hospital (Shenzhen, China) or Sun Yat-sen University Cancer Center (Guangzhou, China), with written informed consent and in accordance with guidelines of the respective local medical ethical committees at those institutions. For patients NKIRTIL063 and OVC190, fresh tumor tissue was obtained by surgical resection, mechanically disrupted and digested for 30 min (for patient OVC190 tissue) or overnight (for patient NKIRTIL063 tissue) at 37 °C in RPMI 1640 medium (Life Technologies) supplemented with penicillin–streptomycin (Roche), 0.01 mg ml^{-1} pulmozyme (Roche) and 1 mg ml^{-1} collagenase type IV (BD Biosciences). The resulting single-cell suspensions were cryopreserved in liquid nitrogen. For patients CV19 and HC25, fresh tumor tissue was cut into fine pieces and digested into single-cell suspension in DMEM supplemented with Accutase (STEMCELL Technologies) and DNaseI (Sigma-Aldrich). After digestion, an aliquot of single-cell suspension of patient HC25 was used to sort $PD1^+$ TILs by FACS, as described below, which were immediately subjected to RNA and TCR single-cell sequencing. The remainder of the single-cell suspension was cryopreserved in liquid nitrogen.

Antibodies

The following antibodies were used for flow cytometry: CD3-APC (clone SK7, BD Biosciences, dilution 1:50); CD3-FITC (clone SK7, BD Biosciences, dilution 1:50); CD3-FITC (clone OKT3, BioLegend, dilution 1:30); CD3-BV421 (clone UCHT1, BioLegend, dilution 1:100), CD4-FITC (clone RPA-T4, BD Biosciences, dilution 1:200), CD14-APC-H7 (clone MoP9, BD Biosciences, dilution 1:100); CD16-APC-H7 (clone 3G8, BD Biosciences, dilution 1:100); CD19-FITC (clone 4G7, BD Biosciences, dilution 1:200); CD45-APC (clone HI30, BD Biosciences, dilution 1:50); CD8α-PE (clone RPA-T8, BD Biosciences, dilution 1:30); CD8α-PacBlue (clone SK1, BioLegend, dilution 1:100); CD8α-Alexa Fluor 700 (clone 3B5, Thermo Fisher Scientific, dilution 1:50); CD8β-BB700 (clone 2ST8.5H7, BD Biosciences, dilution 1:400); CD69-BV421 (clone FN50, BioLegend, dilution 1:50); CD137-PE (clone 4B4, BioLegend, dilution 1:100); CD137-APC (clone 4B4, BioLegend, dilution 1:100); CD107a-PE (clone H4A3, BioLegend, dilution 1:50); and PE-conjugated anti-mouse TCRβ constant domain (clone H57-597, BD Biosciences, dilution 1:150). Live/Dead Fixable Near-IR Dead Cell Stain (Thermo Fisher Scientific), 7-AAD (Thermo Fisher Scientific) or DAPI was used to identify live cells. Data from flow cytometry experiments were acquired using FACSDiva software (version 8.0.2, BD Biosciences) and analyzed using FlowJo (version 10.7.1, BD Biosciences).

Single-cell RNA and TCR sequencing

For patient NKIRTIL063, cryopreserved tumor digest was thawed and stained with DAPI and antibodies against CD45, CD3 and CD8. Live single $CD45^+CD3^+CD8^+$ T cells were then isolated by flow cytometry and loaded into a Chromium single-cell sorting system (10x Genomics). For patient OVC190, tumor digest was thawed and stained with antibodies against CD45, CD3, CD4, CD8 and CD25, and live $CD45^+CD3^+CD4^+$ or $CD8^+$ T cells were isolated by flow cytometry and loaded into two separate lanes of a Chromium system. For patient HC25, live $CD45^+CD3^+PD\text{-}1^+$ TILs were sorted to enrich for tumor-reactive cells, followed by loading into a Chromium system. Cells were kept on ice during all steps of the workflow. Gene expression (for all samples except NKIRTIL063) and TCR library preparation were performed using the Chromium Next GEM Single Cell 5′ and V(D)J reagent kits (10x Genomics) according to the manufacturer's instructions, and libraries were subsequently sequenced on Illumina HiSeq 2500 and MiSeq sequencing systems. Sequencing reads were aligned to the human transcriptome and assigned to individual cells using Cell Ranger (version 6.1.2). Unique molecular identifiers (UMIs) were merged and counted according to Cell Ranger default parameters, and cells were filtered so that cells with fewer than 10,000 detected UMIs, less than 10% of reads mapping to mitochondrial genes or more than 10% of reads mapping to ribosomal protein-coding genes were retained. Further analysis was performed using Seurat[32] (version 4.0.1) in R (version 4.1.1). In brief, gene expression data (excluding TCR variable genes) were normalized using the SCTransform function, simultaneously regressing out the effects of mitochondrial content and total UMI count in each cell. Then, Seurat functions were used for principal component analysis (PCA), followed by uniform manifold approximation and projection (UMAP), nearest neighbor identification and graph-based clustering (neighbors were identified using the first 30 principal components, and clusters were identified with the resolution set to 0.8 (for OVC190 data) or 0.4 (for HC25 data)). Mean NeoTCR4 and NeoTCR8 scores per T cell clonotype were calculated as defined in ref. 26.

Design and synthesis of TCR libraries

Synthetic TCR libraries were produced as in a multi-step assembly process (Fig. 1a and Extended Data Fig. 1a,b). First, germline (GL)-encoded, sequence-verified, reusable human TRAV_GL and TRBV_GL fragments were amplified from sequence-verified plasmids (covering 52 human TRAV alleles and 53 human TRBV alleles). Each TRBV_GL fragment contains a 5′ primer binding site (termed CF), followed by the coding sequence of the *TRBV* gene (including the leader peptide), ending at the cysteine immediately preceding the CDR3 ('2nd-CYS', following IMGT nomenclature[33]). Similarly, each TRAV_GL fragment contains a 5′ region encoding a P2A peptide followed by the coding sequence of the *TRAV* gene (including the leader peptide) ending at the 2nd-CYS. The last 20–30 bp of each TRAV_GL or TRBV_GL allele (that is, the ConnA or ConnB region, respectively) is codon diversified (as described in ref. 34) to maximize DNA sequence diversity. Each TRAV_GL or TRBV_GL fragment is then PCR amplified using a 5′ phosphothioate-modified forward primer and a 5′ phosphate-modified reverse primer, followed by treatment of the PCR product with lambda exonuclease to generate ssDNA TRAV_GL and TRBV_GL fragment pools. In parallel, ssDNA CDR3Jα and CDR3β pools were gene synthesized (Twist Bioscience, Agilent Technologies or Integrated DNA Technologies; see Supplementary Tables 22–28 for oligo sequences of synthesized pools) and PCR amplified using the universal 5′ phosphothioate-modified GQ1 and unmodified GQ4* primers (for ssDNA CDR3Jα pools) or unmodified GQ5 and 5′ phosphothioate-modified BCD* (for ssDNA CDR3β pools; see Extended Data Fig. 1a,b and Supplementary Table 29 for primer sequences). GQ4 and GQ5 segments were removed after PCR, for which two different methods were used with similar results. One method involved the introduction of a type IIS restriction site (for example, BsaI or BbsI) between Zip and GQ4 on CDR3Jα oligos and between GQ5 and Zip on CDR3β oligos, followed by digestion using the corresponding type IIS restriction enzyme to remove the GQ4 and GQ5 segments. Alternatively, one or two dT bases were used instead of the type IIS restriction sites. Accordingly, one or two dU bases were added to the 3′ end of primers GQ4* and GQ5, and GQ4 and GQ5 segments were removed after PCR using the New England Biolabs End Repair kit, as described previously[35]. After either method, ssDNA was generated using lambda exonuclease. The two resultant ssDNA pools (~10 nM each) were incubated in PCR buffer at 50 °C for 24 h to allow cognate pairs of CDR3Jα and CDR3β oligos to hybridize through their Zip and Zip* domains, after which double-stranded DNA (dsDNA) was generated by addition of Taq DNA polymerase to extend the 3′ ends. Zip sequences were pre-designed to have GC content of 50% and similar melting temperatures (Supplementary Table 2). This 'CDR3Jα–CDR3β' pool was then PCR amplified using 5′-unmodified GQ1 and 5′-phosphothioate-modified BCD*. The GQ1 segment was removed using type IIS restriction or dU-based DNA repair as described above, and the top strand of the dsDNA amplicon was removed using lambda

exonuclease to expose the 3′ ConnA* sequence for hybridization to the TRAV_GL fragment pool. The resultant single-stranded paired CDR3Jα–CDR3β pool (at 10 nM) was annealed with the TRAV_GL pool (at 50 nM) via hybridization of ConnA (on TRAV_GL) and ConnA* (on CDR3Jα–CDR3β) in PCR buffer at 50 °C for 3 h, and dsDNA was generated using Taq DNA polymerase to extend the 3′ ends. The dsDNA pool was subsequently ligated to a dsDNA fragment encoding the murine TRBC1 domain directly followed by a short linker (GSG) and P2A sequence using Golden Gate assembly, making use of a type IIS restriction site in the BCD and murine TRBC1 fragments. Next, the identical P2A sequence at the 5′ of TRAV_GL fragments and the 3′ of the TRBC1 fragment were exploited to circularize the ligation product using Gibson assembly. The circularized product was re-linearized by PCR using primers GQ3 and 5′-phosphothioate-modified GQ2*. The GQ3 domain was removed by type IIS restriction or dU-based DNA repair as described above, and ssDNA was generated using lambda exonuclease. The ssDNA pool was annealed to the TRBV_GL fragment pool via hybridization between ConnB and ConnB* sequences, and dsDNA was generated by extension of the 3′ ends using Taq DNA polymerase, producing the final VCV pool. Of note, the NKIRTIL063 VCV pool was synthesized using an oligonucleotide pool encoding pre-paired CDR3Jα–CDR3Jβ sequences (~350 nt in length; Integrated DNA Technologies), and the dropout rate within the resulting VCV pool was relatively high (10.9%; Supplementary Table 10). All other VCV pools were synthesized using separate CDR3Jα and CDR3Jβ pools (each ~230 nt in length; Agilent Technologies or Twist Bioscience), and dropout rates were maintained at 2–3%.

The following HLA-A*02:01-restricted TCRs were included in selected VCV pools as internal controls: 1G4 (specific for NY-ESO-$1_{157–165}$), DMF4 and DMF5 (both specific for $MART1_{26–35}$), PMEL17 (specific for $gp100_{280–288}$), C4 and C4DLT (both specific for $WT1_{126–134}$) and CDK4 53 (specific for $CDK4_{R24L}$). TCR C4DLT has a higher affinity toward the WT1 epitope compared to C4 (refs. 36,37), and DMF5 has a higher affinity toward the MART1 epitope compared to DMF4 (refs. 38,39). In addition, for the OVC190 VCV pools, selected TCRs were encoded in replicate. To this end, the CDR3Jβ nucleotide sequences of individual TCRs were codon diversified without changing the encoded amino acid sequence.

Sequencing analysis of library assembly intermediates

The assembly accuracy and sequence fidelity of each synthetic VCV library were assessed by NGS. In brief, UMI-containing Illumina adapters were appended to VCV fragments by PCR, and libraries were sequenced using the 600-cycle MiSeq Reagent Kit v3 to retain TCRα/TCRβ chain pairing information. Reads mapping to individual VCV sequences and having UMIs with ≤2 mismatched nucleotides were grouped to generate consensus sequences (referred to as molecules). Molecules with ≤2 mapped reads were excluded from analysis. Each molecule was next mapped to the TRAV_GL, ConnA, CDR3Jα, TRBV_GL, ConnB and CDR3Jβ segment sequences, allowing for insertions/deletions and mismatches at this stage. The frequency of each molecule within a library was calculated as the ratio of molecules with the correct combination of segment sequences divided by the total molecule count. To determine the 'α–β pairing accuracy' for individual TCRs (that is, to evaluate the accuracy of Zip hybridization), the number of molecules with both its correct CDR3Jα and CDR3Jβ regions was divided by the number of molecules with its correct CDR3Jα region. Similarly, the ConnB-mediated assembly accuracy between TRBV_GL and CDR3Jβ (termed 'TRBV acquisition accuracy') for individual TCRs was calculated by dividing the number of molecules with both its correct TRBV_GL and CDR3Jβ regions by the number of molecules with its correct CDR3Jβ region. The same method was applied for the quantification of the TRAV_GL–CDR3Jα assembly accuracy (termed 'TRAV acquisition accuracy'). Finally, insertions/deletions and nucleotide mismatches were counted for each molecule identified to be correctly assembled, and the ratio of the number of error-free sequences and the number of correctly assembled sequences for each TCR was used to evaluate the final percentage of molecules with the desired sequence (termed 'perfect rate') within individual libraries.

One-step sequencing analysis of full-length VCV pools

The one-step characterization of full-length VCV products was enabled by the in-house development of a custom NGS method on the Illumina MiSeq platform. In this method, which we termed '2-reads-on-1-strand MiSeq' (Extended Data Fig. 4a), paired reads generate sequence information that covers the TRAV_GL–CDR3αJα and TRBV_GL–CDR3βJβ regions within individual VCV molecules, which can then be mapped to reference sequences of library TCRs. To generate sequencing libraries, UMIs were incorporated into VCV fragment libraries by single-cycle extension using Q5 Hot Start High-Fidelity DNA Polymerase (New England Biolabs) and primer MEA-UMI(N12–15)-ACDrc, the latter being an equimolar mix of primers that differ only in the length of their UMI (ranging from 12 nt to 15 nt). Free primers were removed using Exonuclease I (New England Biolabs), and UMI-extended VCV fragment libraries were purified using AMPure XP beads (Beckman Coulter) and eluted in EDTA TE buffer. The Illumina Rd2 sequence was subsequently appended by 20 cycles of PCR using Q5 Hot Start High-Fidelity DNA Polymerase (New England Biolabs) and primers Rd2-CF and Tn5MEA_5end. After purification of Rd2-extended VCV fragment libraries using AMPure XP beads (Beckman Coulter), Illumina P5 and P7 sequences were adapted by 12 cycles of PCR using Q5 Hot Start High-Fidelity DNA Polymerase (New England Biolabs) and primers P5-Tn5MEA_5end and P7-index(N7)-Rd2, where the 7-nt stretch of N nucleotides indicates a unique sample index sequence used to enable the multiplexed preparation of sequencing libraries. Final sequencing libraries were purified using AMPure XP beads (Beckman Coulter), eluted in nuclease-free water, and amplicon quantities were measured by Qubit (Thermo Fisher Scientific). Sequencing libraries were sequenced on an Illumina MiSeq system (paired end 300 bp) using PhiX as spike-in control and using a customized MiSeq sequencing program that omits resynthesis of the bottom strand on the flow cell. Primers Tn5MEA (identical to the Illumina Nextera Rd1 sequencing primer, for reading of TRAV–CDR3α–Jα) and Illumina_Rd1 (for PhiX reading) were used in cycle 1; primer Illumina_Rd2rc was used in cycle 2 (for sample index reading); and primers BCErc (for TRBV–CDR3β–Jβ reading) and Illumina_Rd1 (for PhiX reading) were used in cycle 3. For each TCR, we measured (1) the count of reads of which the CDR3Jβ region mapped to this TCR's CDR3Jβ reference sequence and (2), among these reads, the count of reads of which TRAV_GL, TRBV_GL and CDR3Jα all mapped to this TCR's reference sequence. We then divided the latter read count by the former to obtain the 'overall assembly accuracy' for each TCR. The mean and median total overall assembly accuracy was consistently between 85% and 95%.

TCR library cloning

VCV fragment libraries were amplified by 15 cycles of PCR using Phusion High-Fidelity DNA Polymerase (New England Biolabs) and universal primers CF and GQ2rc. PCR products were purified using the Monarch PCR & DNA Cleanup Kit (New England Biolabs), digested using BtgZI (New England Biolabs) and purified again using the Monarch PCR & DNA Cleanup Kit. In parallel, a pMX retroviral vector that also encoded murine *TRAC* and the puromycin N-acetyltransferase resistance gene was digested using BbsI (New England Biolabs) and run on a 1% agarose gel. Digested pMX vector DNA was extracted from gel using the Monarch DNA Gel Extraction Kit, and VCV libraries were subsequently ligated into the digested pMX vector, such that the inserted VCV fragment was in frame with the TRAC-PuroR cassette. Subcloned TCR libraries were amplified using Endura electrocompetent cells (Lucigen), and library DNA was isolated using the PureLink HiPure Maxiprep Kit (Invitrogen). A library coverage of at least 100× was maintained during all cloning steps.

Expression of TCR libraries in reporter T cells

For expression in Jurkat T cells, TCR libraries were retrovirally transduced into a TCRα$^-$/TCRβ$^-$ Jurkat T cell line previously modified to express either CD8αβ or CD4, respectively[6]. In brief, retroviral supernatant was produced by transfection of FLY-RD18 packaging cells with pMX-TCR library DNA using X-tremeGENE 9 transfection reagent (Roche). After 48 h, supernatants were collected and used to infect CD8$^+$ or CD4$^+$ TCRα$^-$/TCRβ$^-$ Jurkat cells by spinoculation in 24-well plates precoated with RetroNectin (Takara). Jurkat cells were transduced at infection rates of less than 20% to ensure single retroviral integration in the majority of cells. Transduction efficiency was measured after 72–96 h by flow cytometry. TCR-expressing cells were subsequently sorted using a FACSAria Fusion cell sorter (BD Biosciences) to greater than 80% purity and cultured in RPMI 1640 medium supplemented with 10% FBS (Sigma-Aldrich) until use in TCR discovery screens.

For expression in primary T cells, TCR libraries were cloned into an AAV6 plasmid to be packaged into AAV viral particles (PackGene). Healthy donor T cells, previously isolated and cryopreserved from leukopaks using CD3 isolation kits (STEMCELL Technologies), were thawed and activated with CD3/CD28 beads (Thermo Fisher Scientific) in ImmunoCult Human T Cell Expansion media (STEMCELL Technologies) and 200 U ml^{-1} IL-2 (PeproTech). Two days after bead activation, T cells were electroporated (program EO-115) with Cas9 ribonucleoprotein (RNP) using P3 buffer and a Nucleofector 4D Electroporation System (Lonza). RNPs consisted of SpyFi Cas9 nuclease (Aldevron) duplexed with HPLC-purified single guide RNAs (Integrated DNA Technologies) targeting the human TRAC and TRBC1/2 loci. After electroporation, TCR knockout T cells were treated with TCR library AAV (5×10^4 genome copies per cell) and cultured for two additional days, after which beads were removed and cells were kept in culture for 6–10 d until use in TCR discovery screens or cryopreservation.

Generation of immortalized B cell lines

Immortalized patient B cell lines were generated as previously described[40]. In brief, patient PBMCs were isolated from peripheral blood by Ficoll-Paque density gradient separation and stained with IR-Dye and anti-CD3, anti-CD14, anti-CD16 and anti-CD19 antibodies. Single live B cells (CD3$^-$CD14$^-$CD16$^-$CD19$^+$) were sorted using a FACSAria Fusion cell sorter (BD Biosciences) and stimulated by incubation with irradiated (55 Gy) CD40L$^+$ mouse L cells in IMDM medium (Gibco) supplemented with 10% heat-inactivated FBS (Sigma-Aldrich), penicillin-streptomycin and 50 ng ml^{-1} IL-21 (BioLegend). After 36 h, stimulated B cells were retrovirally transduced with Bcl-6 and Bcl-xL. Bcl-6/xL-immortalized B cells were cultured in IMDM medium supplemented with 10% heat-inactivated FBS, penicillin–streptomycin and 25 ng ml^{-1} IL-21 and were restimulated every 10 d by addition of irradiated CD40L$^+$ L cells. Medium and IL-21 were refreshed every 3–4 d.

Exome and RNA sequencing

Genomic DNA and RNA were isolated from formalin-fixed paraffin-embedded tumor material using the AllPrep DNA/RNA Kit (Qiagen). Genomic DNA of patient PBMCs was isolated using the DNeasy Blood & Tissue Kit (Qiagen). Whole-exome and RNA sequencing was subsequently performed as described previously[6,25]. Patient-specific non-synonymous mutations were identified by mapping DNA reads to genome build GRCh38 (release 88) using BWA (version 0.7.10), followed by variant calling and annotation using MuTect2 (version 3.7.0) and SnpSift (version 4.3p) and validation of expression from RNA sequencing data.

TMG design and transduction

TMG constructs encoding NKIRTIL063 tumor mutations were previously designed and generated[25]. In brief, 200 tumor mutations (of a total of 685 expressed non-synonymous mutations) were selected based on mutation clonality and gene expression level. Individual NKIRTIL063 TMG constructs encoded 10 mutation-encoding minigenes in which the mutant codon was flanked on either side by 45 nt of wild-type gene sequence. Sequences were codon optimized, synthesized (Twist Bioscience) and subcloned into a pMSCV vector directly downstream of the signal sequence of CD74. The pMSCV-CD74 vector also encodes mCherry and the puromycin N-acetyltransferase resistance gene to allow evaluation of transduction efficiencies and selection, respectively. For patient OVC190, eight TMG constructs encoding all 61 expressed non-synonymous tumor variants were designed, synthesized and subcloned into the pMSCV-CD74-Puro-mCherry vector. NKIRTIL063 and OVC190 TMG constructs were retrovirally transduced into patient-matched immortalized B cells as described above.

TCR discovery screens

In brief, TCR discovery screens were performed by incubating TCR library-expressing T cells with target epitope-expressing APCs (allogenic tumor cell lines, antigen-modified B cells or patient tumor cells), followed by isolation of activated T cells using FACS or MACS and deep sequencing to identify responding TCRs. For screens using Jurkat T cells, TCR library-expressing Jurkat cells were incubated with immortalized B cells in RPMI 1640 medium (Life Technologies) supplemented with 10% FBS (Sigma-Aldrich) and penicillin–streptomycin (Roche) for 8 h at a density of 600,000 total cells per cm^2. When peptide-pulsed B cells were used in screens, immortalized B cells were harvested and resuspended in 1 µg ml^{-1} peptide for 1 h at 37 °C. After peptide pulsing, B cells were washed twice with RPMI 1640 medium and incubated in duplicate with TCR library-expressing Jurkat cells at an effector-to-target ratio of 1:1. Screens using TMG-expressing B cells were performed by pooling equal cell numbers of four TMG B cell lines per screening condition, followed by incubation in triplicate with TCR library-expressing Jurkat cells at an effector-to-target ratio of 1:4. TCR library coverage of at least 500× was maintained in all experiments. After co-culture, cells were harvested, washed and stained with Live/Dead Fixable Near-IR Dead Cell Stain (Thermo Fisher Scientific) and antibodies against mTCRβ and CD69. Activated Jurkat cells (defined as the top 10% CD69$^+$ fraction of single live mTCRβ$^+$ cells) were isolated by flow cytometry using a FACSAria Fusion cell sorter (BD Biosciences), washed and lysed in DirectPCR Lysis Reagent (Viagen) containing 500 µg ml^{-1} proteinase K by incubation at 55 °C for 60 min, 85 °C for 30 min and 94 °C for 5 min. Transgenic TCRβ chains were PCR amplified using NEBNext High-Fidelity 2× PCR Master Mix (New England Biolabs) using a P7-containing forward primer annealing to the retroviral backbone sequence directly 5′ of the TCRβ chain (5′-CAAGCAGAAGACGGCATACGAGATGGGGT GGACCATCCTCTAGACTGC-3′) and barcoded reverse primers annealing to the TRBC domain directly 3′ of the CDR3β, containing P5 and Illumina read 1 sequences (5′- AATGATACGGCGACCACCGAGAT CTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNG GGACACTTTTGGAGGTGTGACGTTCC-3′). The N nucleotides enable demultiplexing of samples during sequencing analysis. To prepare sequencing library pools, PCR products were pooled equimolarly and by correcting for the number of sorted cells per screening condition. Pooled PCR products of the expected amplicon size were purified from agarose gel using the Monarch DNA Gel Extraction Kit (New England Biolabs) and sequenced on an Illumina NextSeq 550 sequencing system (single read 150 bp).

To analyze the influence of the abundance of individual TCRs within libraries on screening sensitivity, NKIRTIL063 library-expressing CD8$^+$ Jurkat cells were diluted three-fold, 10-fold or 30-fold with mock-transduced Jurkat cells, followed by use in screens against $MART1_{26–35}$ peptide-pulsed HLA-A*02:01-positive immortalized B cells. To simulate a screening setting in which antigen-specific TCRs are represented at reduced frequencies, identical numbers of total Jurkat cells were used across screening conditions.

For screens using primary donor T cells, TCR library-engineered T cells were incubated overnight in duplicate with allogenic tumor lines (Malme3M, RPMI7951, SiHa, K562 or HMCB (American Type Culture Collection)) or patient tumor cells at an effector-to-target ratio of 1:1. After co-culture, cells were harvested, washed and stained with an APC-conjugated anti-CD137 antibody. Activated CD137$^+$ T cells were subsequently isolated by flow cytometry using an MA900 cell sorter (Sony). Alternatively, CD137$^+$ cells were isolated by MACS using anti-APC magnetic beads (Miltenyi Biotec) following the manufacturer's instructions. Transgenic TCRs were subsequently PCR amplified from isolated cells and submitted to deep sequencing.

Screen sequencing analysis

Quality of sequencing reads was determined by FastQC (version 0.11.9), and reads were de-multiplexed using je-demultiplex (version 1.2), allowing one mismatch in the sample barcode and up to five bases of lower quality toward the end of reads. The barcode sequences were trimmed, and reads were further filtered using fastp (version 0.32.2), allowing only reads with less than 10% of bases with quality lower than Q28 and fewer than five N bases. Filter-passing reads were then aligned to sequences of library TCRs with no mismatches allowed and counted using the MAGeCK count algorithm (version 0.5.8). To identify reactive TCRs, per-sample count tables were normalized and compared using DESeq2 (version 1.34.0) and dplyr (version 1.0.9) in R (version 4.1.1), and hit TCRs were determined using the DESeq2 Wald test and an adjusted $P < 0.05$, unless stated otherwise. For the NKIRTIL063 TCR library screen against TMG-expressing B cell pools, TCRs were defined as reactive to an individual TMG B cell pool only when they reached statistical significance in comparison to all other four TMG pools. Sequencing data were deposited in the National Center for Biotechnology Information's Sequence Read Archive under accession codes PRJNA1068078 (ref. 41), PRJNA1068299 (ref. 42), PRJNA1068301 (ref. 43) and PRJNA1068303 (ref. 44) and are publicly available.

Expression of individual TCRs

For validation of TCRs identified in pooled screens, individual TCRs were expressed in Jurkat or healthy donor T cells using retroviral transduction or electroporation of in vitro transcribed mRNA. Individual TCRs were first amplified from VCV pools by 26 cycles of PCR using Phusion High-Fidelity DNA Polymerase and a universal forward primer in combination with a reverse primer targeting the TCR's unique Zip barcode sequence. For retroviral transduction, amplified individual TCRs were subcloned into the pMX-TRAC-PuroR retroviral vector and subsequently transduced into Jurkat cells or healthy donor T cells as described above. When TCRs were transduced into primary T cells, healthy donor PBMCs (Sanquin Blood Bank) were separated into CD8$^+$ cells (when transducing MHC class I–restricted TCRs) and CD8$^-$ cells (when transducing MHC class II–restricted TCRs) using the CD8$^+$ T Cell Isolation Kit (Miltenyi Biotec) and stimulated with anti-CD3/anti-CD28 Dynabeads (Life Technologies) in RPMI 1640 medium supplemented with 10% human AB serum (Life Technologies), penicillin–streptomycin (Roche) and 150 U ml^{-1} IL-2 (Proleukin, Novartis). Jurkat cells or pre-stimulated CD8$^-$/CD8$^+$ PBMCs were infected with retroviral supernatants by spinoculation (2,000*g* for 90 min) in RetroNectin-coated 24-well plates. Transduction efficiency was measured 72–96 h later by staining transduced cells with an anti-mouse TRBC antibody and analysis by flow cytometry. Transduced cells were selected with 2.5 µg ml^{-1} puromycin for 48 h and received fresh IL-2 (when using primary T cells) and medium every 3–4 d.

For electroporation of TCR mRNA, amplified individual TCRs were subcloned by Golden Gate assembly into an RNA expression vector containing the *TRAC* domain. mRNA was then generated using a HiScribe T7 High Yield RNA Synthesis Kit (New England Biolabs). TCR knockout T cells, generated as described above, were subsequently electroporated with 1 µg of TCR mRNA using P3 buffer and a Nucleofector 4D Electroporation System (Lonza). After electroporation, cells were cultured overnight in ImmunoCult Human T Cell Expansion media (STEMCELL Technologies) supplemented with 200 U ml^{-1} IL-2 (PeproTech). TCR expression was measured by flow cytometry the next day, and cells were used in functional assays.

T cell activation assays

Reactivity of single TCR-engineered T cells was analyzed by flow cytometry or IFNγ ELISpot. For flow cytometry, TCR-engineered T cells were cultured with target cells overnight at an effector-to-target ratio of 1:1 in U-bottom 96-well plates. Incubation of T cells alone, or in the presence of phorbol 12-myristate 13-acetate (50 ng ml^{-1}; Sigma-Aldrich) and ionomycin (1 µg ml^{-1}; Sigma-Aldrich), served as negative and positive controls, respectively. Cells were stained with IR-Dye and antibodies against anti-mouse TRBC antibody, CD3, CD4, CD8 and CD137 (when using primary T cells) or CD69 (when using Jurkat cells) and analyzed by flow cytometry. When applicable, z-scores of the percentage of activated cells were calculated across conditions per individual TCR, and conditions were defined as reactive when z-scores exceeded the indicated thresholds. For IFNγ ELISpot, TCRs were defined as reactive when (1) TCR-engineered T cells produced at least 20 spots and (2) target cells elicited at least four-fold more IFNγ spots than the respective negative control condition.

The cytotoxic capacity of TCR-transduced T cells was assessed by culturing T cells and target cells for 48 h in flat-bottom 96-well plates at the indicated effector-to-target cell ratios. Target cells cultured in the absence of T cells served as negative control. After co-incubation, cells were harvested and subsequently stained with DAPI and anti-CD3 antibody and analyzed by flow cytometry. Data from functional T cell assays were analyzed using GraphPad Prism version 9 software.

Reporting summary

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

Data availability

DNA sequencing data of TCR discovery screens have been deposited in the National Center for Biotechnology Information's Sequence Read Archive under accession codes PRJNA1068078 (ref. 41), PRJNA1068299 (ref. 42), PRJNA1068301 (ref. 43) and PRJNA1068303 (ref. 44).

References

32. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
33. Lefranc, M. P. et al. IMGT®, the international ImMunoGeneTics information system® 25 years on. *Nucleic Acids Res.* **43**, D413–D422 (2015).
34. Chen, X. & Porter, E. Compositions and methods for T-cell receptor gene assembly. Patent WO2020206238A2 (2020).
35. Klein, J. C. et al. Multiplex pairwise assembly of array-derived DNA oligonucleotides. *Nucleic Acids Res.* **44**, e43 (2016).
36. Chapuis, A. G. et al. T cell receptor gene therapy targeting WT1 prevents acute myeloid leukemia relapse post-transplant. *Nat. Med.* **25**, 1064–1072 (2019).
37. Schmitt, T. M., Greenberg, P. D. & Nguyen, H. N. T cell immunotherapy specific for WT-1. US patent US20160083449A1 (2015).
38. Borbulevych, O. Y., Santhanagopolan, S. M., Hossain, M. & Baker, B. M. TCRs used in cancer gene therapy cross-react with MART-1/Melan-A tumor antigens via distinct mechanisms. *J. Immunol.* **187**, 2453–2463 (2011).
39. Johnson, L. A. et al. Gene transfer of tumor-reactive TCR confers both high avidity and tumor reactivity to nonreactive peripheral blood mononuclear cells and tumor-infiltrating lymphocytes. *J. Immunol.* **177**, 6548–6559 (2006).

40. Kwakkenbos, M. J. et al. Generation of stable monoclonal antibody–producing B cell receptor–positive human memory B cells by genetic programming. *Nat. Med.* **16**, 123–128 (2010).
41. Moravec, Z. et al. Functional discovery of tumor-reactive T cell receptors by massively parallel library synthesis and screening: validation in CD8 T cells. *NCBI Sequence Read Archive* https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1068078 (2024).
42. Moravec, Z. et al. Functional discovery of tumor-reactive T cell receptors by massively parallel library synthesis and screening: validation in CD4 T cells and OVC190 TCR screen. *NCBI Sequence Read Archive* https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1068299 (2024).
43. Moravec, Z. et al. Functional discovery of tumor-reactive T cell receptors by massively parallel library synthesis and screening: NKIRTIL063 titration screen. *NCBI Sequence Read Archive* https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1068301 (2024).
44. Moravec, Z. et al. Functional discovery of tumor-reactive T cell receptors by massively parallel library synthesis and screening: NKIRTIL063 neoantigen screen. *NCBI Sequence Read Archive* https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1068303 (2024).

Acknowledgements

We would like to thank P. Kaptein and J. Urbanus for valuable help with developing the screening technology; S. Ketelaars for bioinformatic support; R. Tissier for support with statistical analyses; the Netherlands Cancer Institute–Antoni van Leeuwenhoek Hospital (NKI-AVL) Flow Cytometry Facility for flow cytometric support; the NKI-AVL Core Facility Molecular Pathology & Biobanking for supplying NKI-AVL Biobank material and laboratory support; and the NKI-AVL Genomics Core Facility for support with next-generation sequencing. This work was supported by the Dutch Cancer Society Young Investigator Grant (no. 2020-1/12977) (to W.S.); ZonMw Translational Research Program 2 (no. 446002001) (to W.S. and J.B.A.G.H.); the Stevin Prize and the Louis-Jeantet Prize (to T.N.S.); National Institutes of Health grants (no. R01CA269898 and no. R37CA273333-01) (to J.W.); the V Foundation (to J.W.); the Mark Foundation ASPIRE award (to J.W.); the Melanoma Research Alliance Young Investigator award (to J.W.); and a generous donation from Florry Vyth (to J.B.A.G.H.). Research at the Netherlands Cancer Institute is supported by institutional grants from the Dutch Cancer Society and the Dutch Ministry of Health, Welfare and Sport.

Author contributions

Z.M., R.V., D.R.C., Y.Z., B.R., B.C., H.Y., J.O., J.X., T.N.S., X.C., E.P. and W.S. designed, performed, analyzed and/or interpreted experiments. Z.M. and S.K. analyzed sequencing data. J.B.A.G.H. and D.H. supplied patient tumor material. T.W. and L.R. supplied patient tumor material and isolated and cultured patient TIL lines. J.W. helped conceptualize the methodology. E.P., X.C. and W.S. wrote the manuscript. All authors reviewed the manuscript.

Competing interests

D.R.C., Y.Z., B.C., S.K., H.Y., J.O., J.X. and E.P. are employees and stock option holders of RootPath, Inc. or its affiliates. X.C. is a director and shareholder of RootPath, Inc. or its affiliates. J.W. is a paid consultant for RootPath Genomics, Bristol Myers Squibb (Relatlimab Advisory Council) and Hanmi Pharmaceutical and is a founder and equity holder of and a consultant to Remunix, Inc. J.B.A.G.H. is an advisor to Achilles Therapeutics, BioNTech, Instil Bio, Neogene Therapeutics, PokeAcell, Scenic Biotech, T-Knife and Third Rock Ventures; is a recipient of research grant support from BioNTech; and is a stock option holder of Neogene Therapeutics. T.N.S. serves as an advisor for Allogene Therapeutics, Merus, Neogene Therapeutics and Scenic Biotech and is a stockholder in Allogene Therapeutics, Cell Control, Celsius, Merus and Scenic Biotech. T.N.S. is also a venture partner at Third Rock Ventures, all outside the submitted work. W.S. is an advisor to BD Biosciences and Lumicks. The TCR library synthesis method is described in patent application WO2020206238A2, assigned to a subsidiary of RootPath, Inc. (X.C. and E.P.). The other authors declare no competing interests.

Additional information

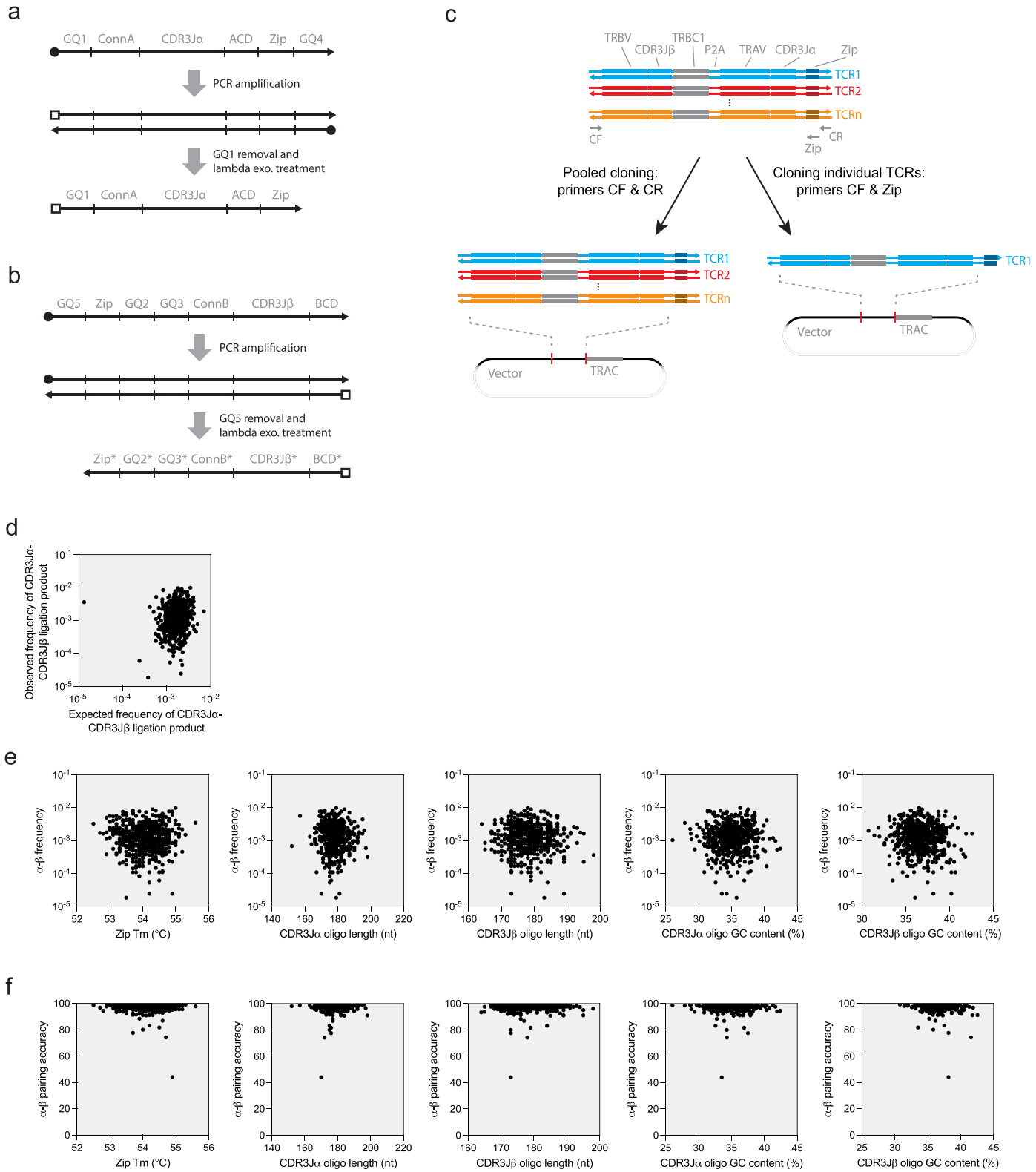
Extended data is available for this paper at https://doi.org/10.1038/s41587-024-02210-6.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41587-024-02210-6.

Correspondence and requests for materials should be addressed to Xi Chen, Ely Porter or Wouter Scheper.

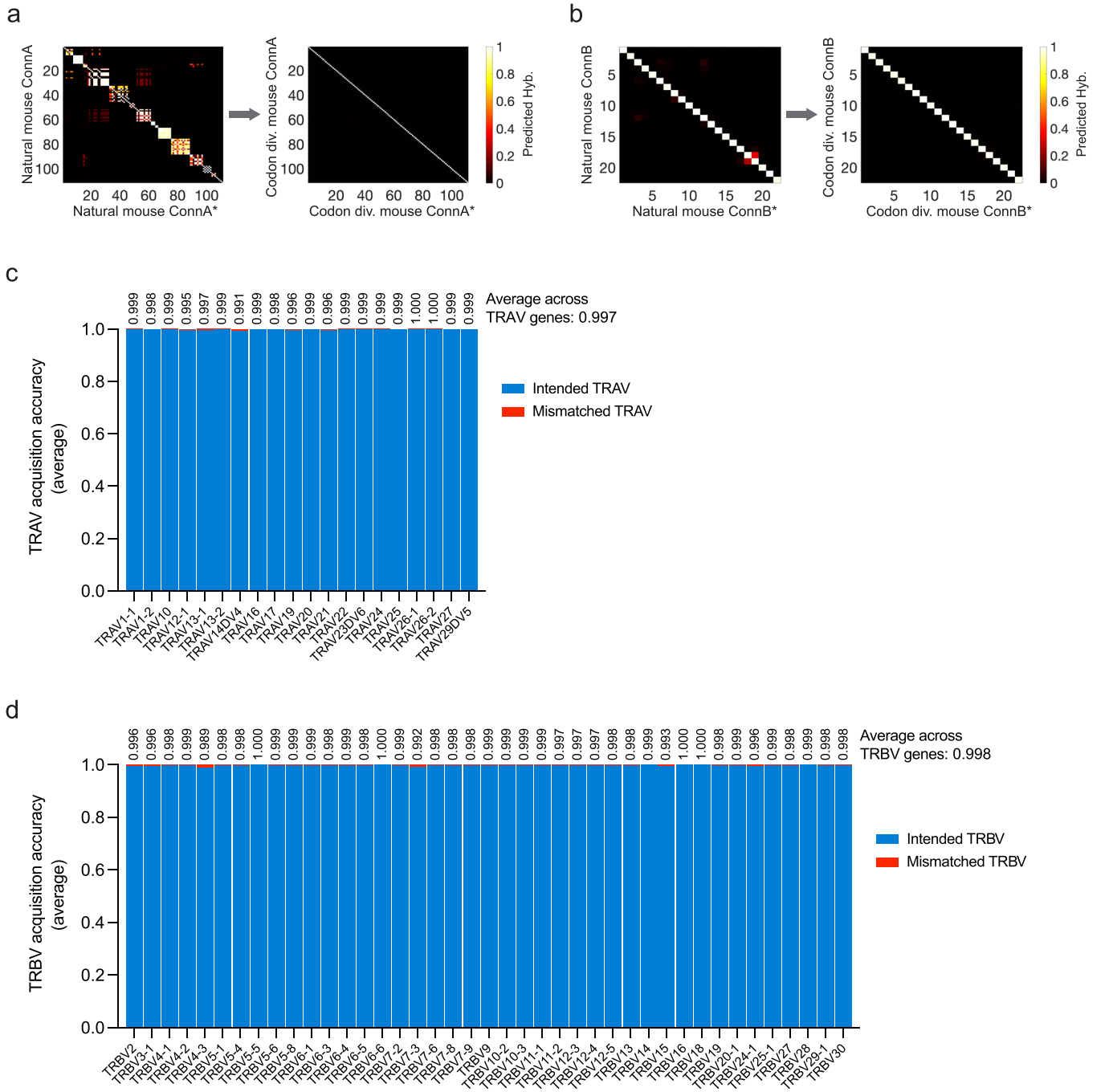
Peer review information *Nature Biotechnology* thanks Michael Birnbaum and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

https://doi.org/10.1038/s41587-024-02210-6



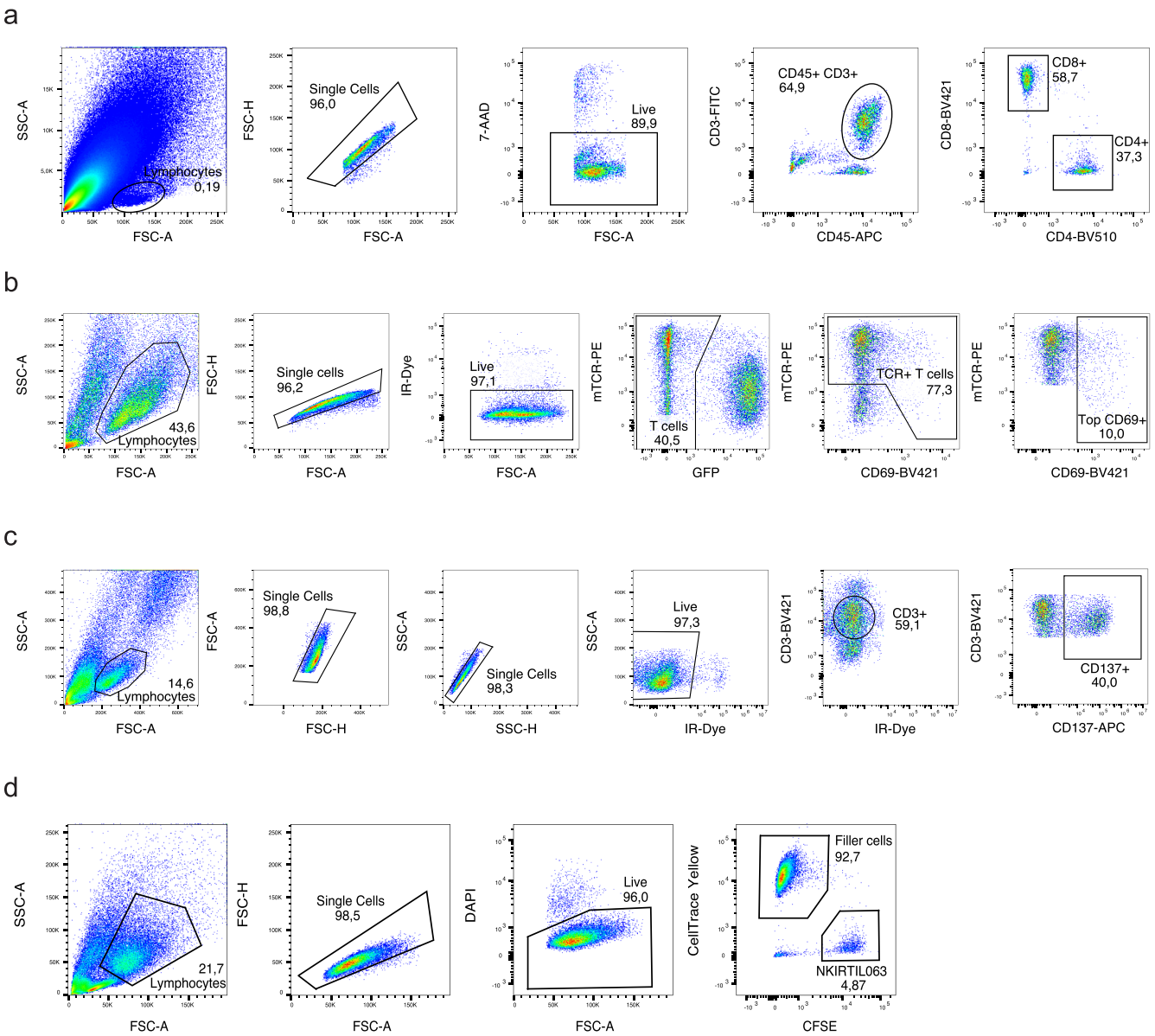
Extended Data Fig. 1 | Pooled TCR library subcloning and accuracy. (**a, b**) Schemes to produce ssDNA CDR3Jα pools (a) and ssDNA CDR3Jβ pools (b). Squares indicate 5′ phosphothioate modification. See Supplementary Table 29 for primer sequences. (**c**) Complete VCV pools can be PCR-amplified prior to subcloning using the common forward (CF) and reverse (CR) primers. Individual TCRs may be selectively amplified from VCV pools using the common forward primers (CF) and reverse primers targeting their respective Zip barcode sequences. (**d**) For the proof-of-concept TCR library (553 TCRs), the frequencies of individual CDR3Jα and CDR3Jβ sequences within the commercially synthesized CDR3Jα and CDR3Jβ ssDNA oligo pools and the assembled CDR3Jα-CDR3Jβ product were assessed by deep sequencing. The expected frequency of each CDR3Jα-CDR3Jβ pair was derived by multiplication of the observed frequencies of its respective CDR3Jα and CDR3Jβ sequences in the original, unassembled oligo pools. (**e**) Relation between Zip and CDR3Jα/CDR3Jβ oligo characteristics and the frequencies of resulting CDR3Jα-CDR3Jβ pairs after hybridization. (**f**) Relation between Zip and CDR3Jα/CDR3Jβ oligo characteristics and the accuracy of CDR3Jα-CDR3Jβ pairing ('α-β pairing accuracy').

https://doi.org/10.1038/s41587-024-02210-6



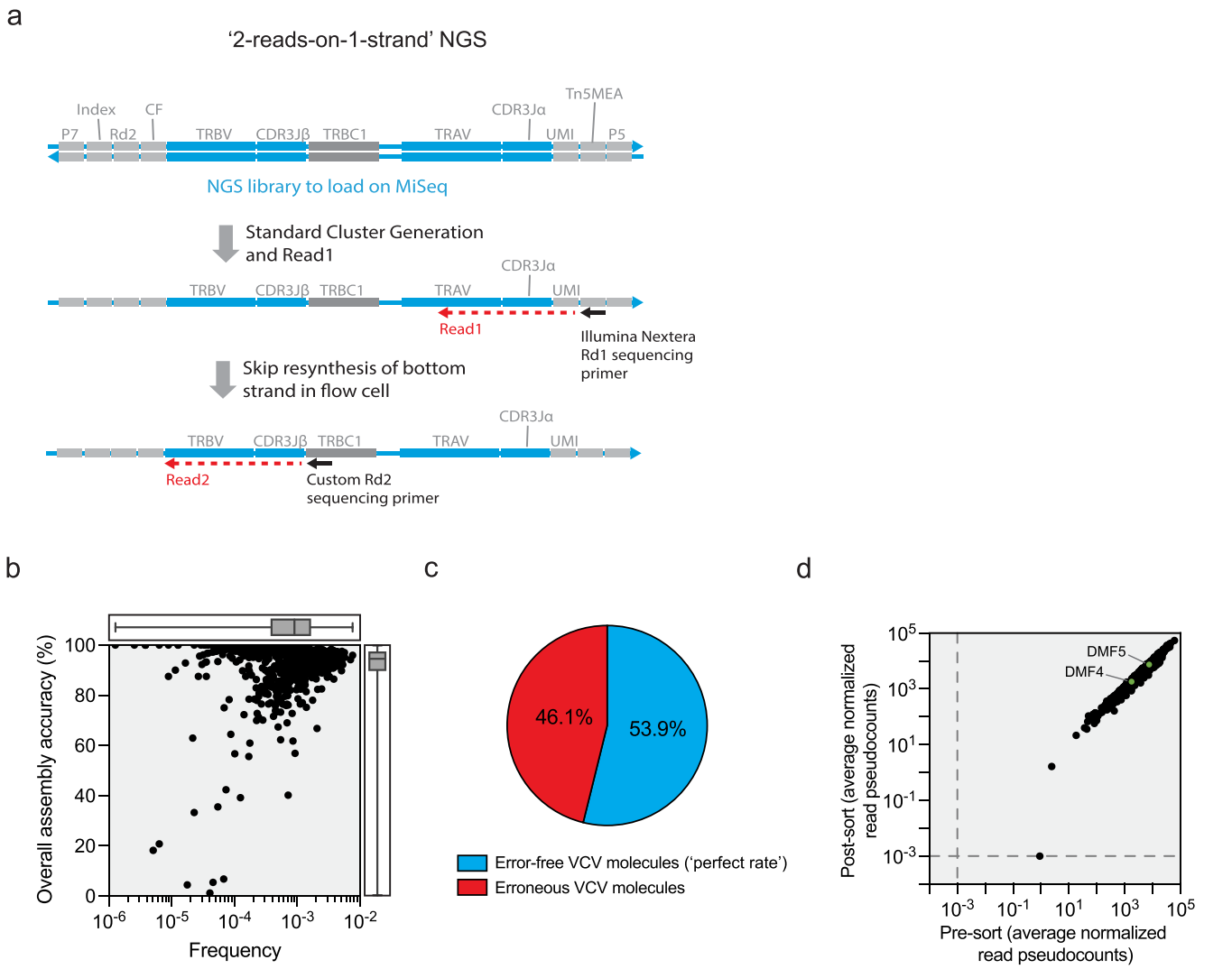
Extended Data Fig. 2 | ConnA and ConnB orthogonality for murine TRAV and TRBV genes and acquisition accuracies. (**a**) Computer-predicted orthogonality of natural (left) and codon-diversified (right) FR3 connector sequences (ConnA) for murine TRAV. (**b**) Computer-predicted orthogonality of natural (left) and codon-diversified (right) FR3 connector sequences (ConnB) for murine TRBV. (**c, d**) Acquisition accuracies for TRAV (c) and TRBV (d) genes used in the proof-of-concept TCR library. The average acquisition accuracy for each TRAV/TRBV gene is shown above the bars.

https://doi.org/10.1038/s41587-024-02210-6



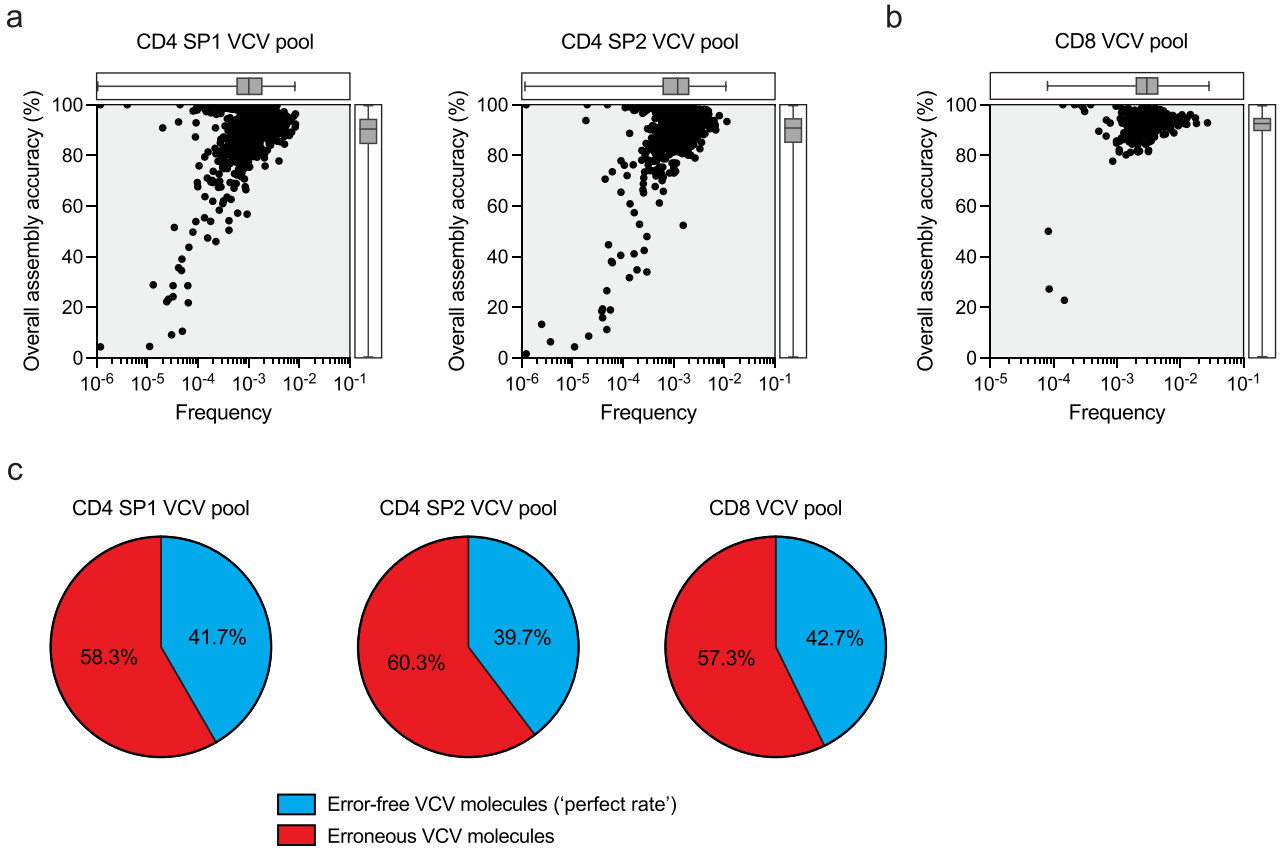
Extended Data Fig. 3 | Gating strategies. (**a**) Gating strategy for the isolation of CD4+ and/or CD8+ T cells from patient tumor material, for the purpose of identifying patient-derived TCRs by single-cell TCR sequencing (relating to the patient TCR libraries used in Figs. 2 to 4). (**b**) Gating strategy for the isolation of activated reporter T cells in TCR library screens (relating to Figs. 2b–d, 3a,d and 4a, and Extended Data Figs. 7a–c, and 9b,d. (**c**) Gating strategy for T cell activation assays (relating to Figs. 3c,e,f and 4b,c, and Extended Data Figs. 7d–f and 9a. CD137 served as activation marker when using primary T cells, and CD69 served as activation marker when using Jurkat cells. (**d**) Gating strategy for T cell cytotoxicity assays (relating to Figs. 3g and 4d).

https://doi.org/10.1038/s41587-024-02210-6



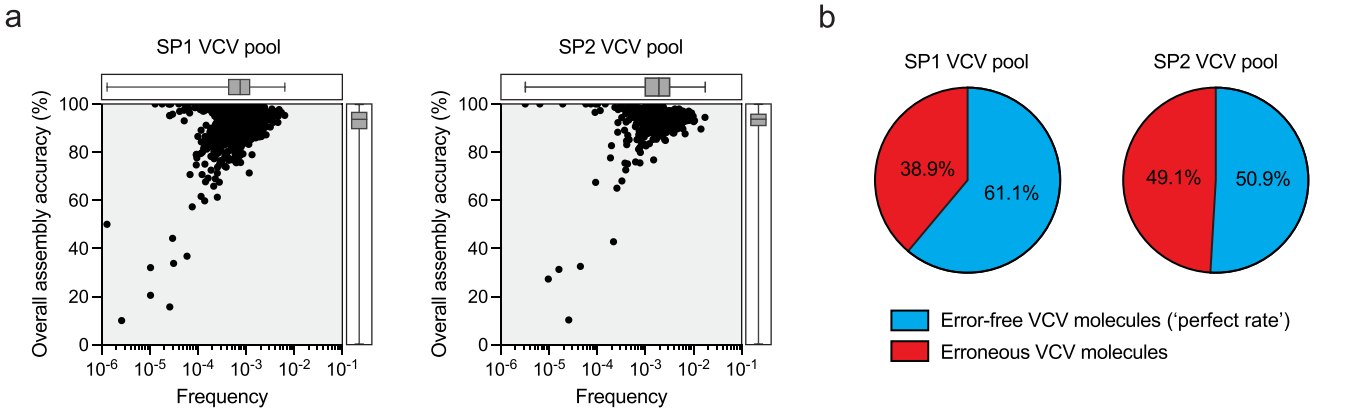
Extended Data Fig. 4 | Quality control of the NKIRTIL063 TCR library. (**a**) Schematic overview of the custom '2-reads-on-1-strand' sequencing method. (**b**) Overall assembly accuracies and frequencies of TCRs in the fully assembled NKIRTIL063 VCV library (n = 935 TCRs). Dots represent individual TCRs. Box plots depict the median, the interquartile range and whiskers extending to minimal and maximal values. (**c**) Pie chart depicting the overall fraction of NKIRTIL063 VCV molecules with and without sequence errors (taking both assembly accuracy and sequence mutations into account). (**d**) The fraction of NKIRTIL063 library TCRs successfully expressed at the cell surface of library-transduced cells was determined by isolation of mouse TCRβ+ Jurkat cells by flow cytometry, followed by deep sequencing. NKIRTIL063 library-expressing Jurkat cells that were not subjected to cell sorting were used as reference. Internal control TCRs DMF4 and DMF5 are highlighted for reference.

https://doi.org/10.1038/s41587-024-02210-6



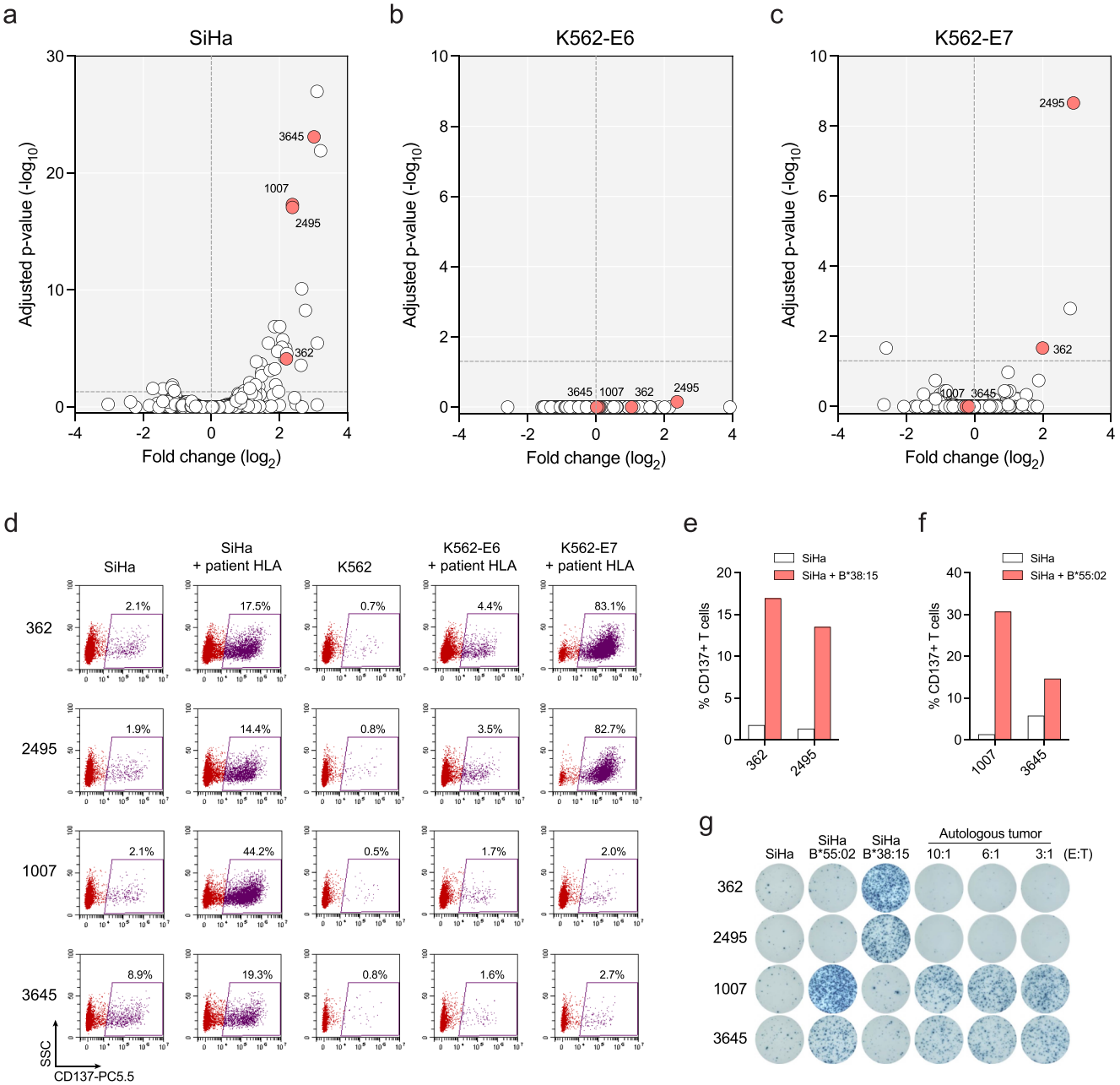
Extended Data Fig. 5 | Quality control of the OVC190 TCR libraries. (**a**) Overall assembly accuracies and frequencies of TCRs in the fully assembled OVC190 CD4+ TIL-derived VCV libraries (combined n = 1,341 TCRs). Selected TCRs were encoded in replicate, resulting in a total of 2,899 unique sequences that were assembled in two separate reactions (SP1 and SP2). Dots represent individual TCRs. Box plots depict the median, the interquartile range and whiskers extending to minimal and maximal values. (**b**) Overall assembly accuracies and frequencies of TCRs in the fully assembled OVC190 CD8+ TIL-derived VCV library (n = 274 TCRs). (**c**) Pie chart depicting the overall fraction of OVC190 CD4+ and CD8+ TIL-derived VCV molecules with and without sequence errors (taking both assembly accuracy and sequence mutations into account).

https://doi.org/10.1038/s41587-024-02210-6



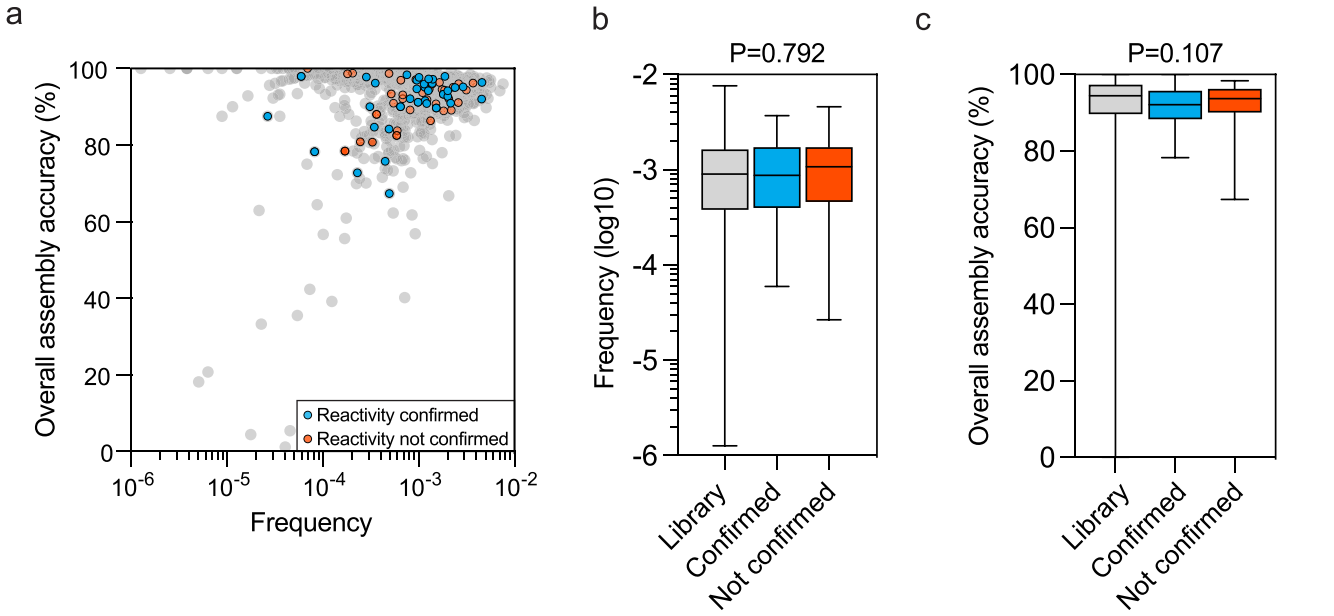
Extended Data Fig. 6 | Quality control of the CV19 TCR library. (**a**) Overall assembly accuracies and frequencies of TCRs in the fully assembled CV19 TIL-derived VCV libraries (combined n = 1,501 TCRs). Sequences were assembled in two separate reactions (SP1 and SP2). Dots represent individual TCRs. Box plots depict the median, the interquartile range and whiskers extending to minimal and maximal values. (**b**) Pie chart depicting the overall fraction of CV19 VCV molecules with and without sequence errors (taking both assembly accuracy and sequence mutations into account).

https://doi.org/10.1038/s41587-024-02210-6



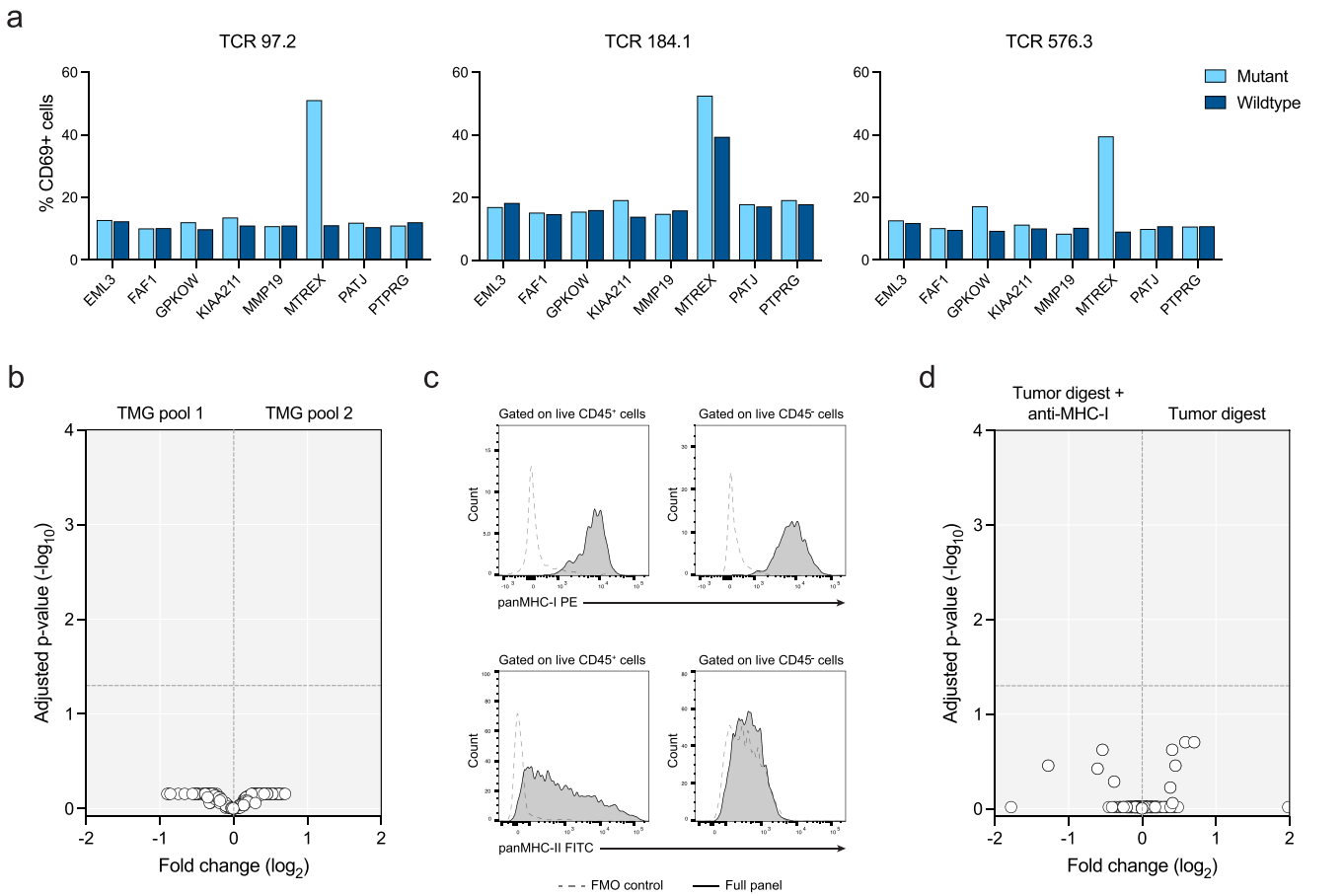
Extended Data Fig. 7 | Functional screening of intratumoral TCR repertoire of patient CV19. (**a**) The SiHa cervical cancer cell line was engineered to express the entire MHC class I haplotype of patient CV19 (A*24:02, A*33:03, B*38:15, B*55:02, C*03:02 and C*12:03). The CV19 TCR library (n = 1,501 TCRs) was subsequently expressed in donor T cells and screened against either the unmodified or MHC-modified SiHa line. Fold change represents the relative abundance of TCRs after incubating T cells with unmodified or MHC-modified SiHa cells. TCRs selected for validation are highlighted in red. (**b, c**) To assess TIL reactivity against HPV-derived antigens, the patient TCR library was screened against K562 cells that were modified to express the patient's MHC class I alleles as well as the full ORF of either HPV E6 (b) or E7 (c) oncoproteins. Data are depicted as in (a). Two of four selected TCRs (2495 and 362) responded to E7-expressing K562 cells, while no TCRs responded against E6-expressing cells. (**d**) The reactivity of selected TCRs was validated by amplifying TCRs from the CV19 VCV pool, followed by expression in donor T cells and co-incubation with the indicated cell lines. T cell activation was assessed by measuring CD137 expression using flow cytometry. (**e, f**) MHC restriction of selected TCRs was assessed by expressing individual patient MHC alleles in SiHa cells and incubating resulting cells with donor T cells engineered to express selected TCRs. T cell activation was assessed by measuring CD137 surface expression. (**g**) SiHa-reactive TCRs 1007 and 3645, but not E7-specific TCRs 362 and 2495, recognize autologous patient tumor cells. TCR-modified donor T cells were incubated with unmodified, HLA-B*55:02-modified or HLA-B*38:15-modified SiHa cells, or autologous dissociated tumor tissue at the indicated effector to target ratios. Activation of TCR-engineered T cells in measured by IFNγ ELISpot.

https://doi.org/10.1038/s41587-024-02210-6



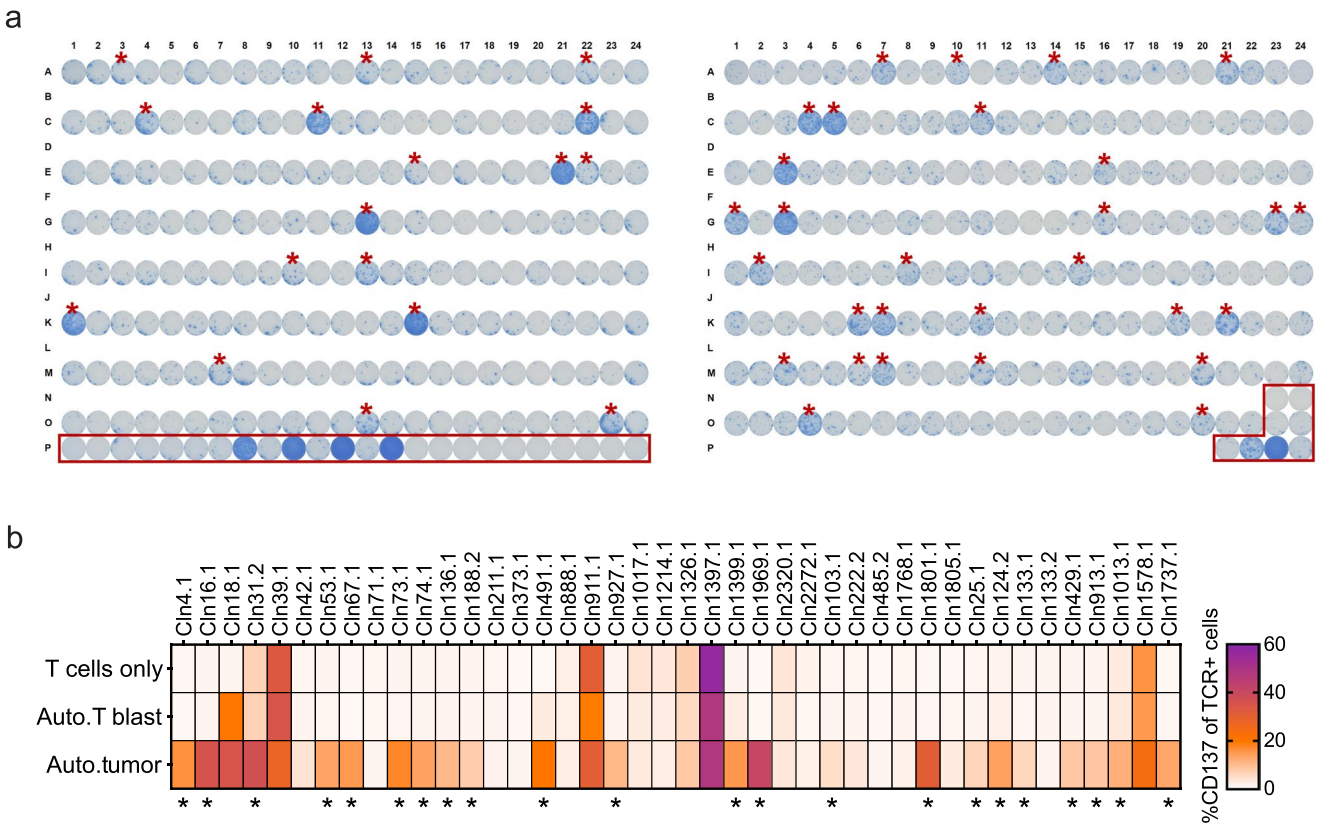
Extended Data Fig. 8 | Sensitive TCR discovery across variable TCR frequencies and sequence fidelities. (**a**) Projection of all NKIRTIL063 screen hit TCRs with confirmed (blue, n = 44) and unconfirmed (red, n = 36) reactivity (as reported in Fig. 3) onto the quality control data of the NKIRTIL063 TCR library (see Extended Data Fig. 4). (**b, c**) Comparison between the frequencies (b) and assembly accuracies (c) of the overall NKIRTIL063 TCR library, TCRs with confirmed reactivity and TCRs with unconfirmed reactivity. Box plots depict the median, the interquartile range and whiskers extending to minimal and maximal values. P values were determined using the Kruskal-Wallis test.

https://doi.org/10.1038/s41587-024-02210-6



Extended Data Fig. 9 | Neoantigen specificities of OVC190 TCRs. (**a**) TCR-modified CD4+ Jurkat cells were incubated with patient B cells expressing either the mutant or wildtype sequence of the individual minigenes of TMG 5. T cell activation was determined by measuring CD69 expression on T cells by flow cytometry. (**b**) CD8+ Jurkat cells were transduced with the CD8+ TIL-derived TCR library of patient OVC190 (n = 274 unique TCRs). Patient immortalized B cells were transduced with TMGs encoding all expressed non-synonymous mutations (n = 61) of the patient's tumor, combined in pools and used to screen the OVC190 TCR library. Dots represent individual TCRs. Fold change represents the relative abundance of TCRs in cultures with the indicated TMG pools. (**c**) Flow cytometry analysis of MHC class I (top panels) and MHC class II (bottom panels) expression on CD45+ and CD45- cells within the OVC190 tumor. Fluorescence minus one (FMO) stains with antibody panels that lacked either the panMHC-I or panMHC-II antibody served as negative control. (**d**) CD8+ Jurkat cells expressing the OVC190 CD8+ TIL-derived TCR library were screened against single cell suspension of OVC190 tumor. Screening the TCR library against tumor cells in the presence of MHC class I blocking antibody (clone W6-32) served as negative control. Data are depicted as in (b).

https://doi.org/10.1038/s41587-024-02210-6



Extended Data Fig. 10 | Patient HC25 tumor-specific TCR identification. (**a**) Patient HC25 PD-1+ TIL were sorted by flow cytometry, and subjected to paired single cell RNA and TCR sequencing. NeoTCR4 and NeoTCR8 transcriptional signatures were derived for CD4+ and CD8+ clonotypes, respectively, and clonotypes with the 382 highest scores were gene-synthesized. Individual TCRs were expressed in donor T cells and reactivity to autologous dissociated tumor tissue was assessed by IFNγ ELISpot. Wells with responding TCRs are marked by red asterisks. Red boxes indicate HC25 TCR-independent experimental controls. (**b**) Selective reactivity of hit TCRs to patient tumor cells, but not non-malignant cells, was validated by incubating TCR-engineered donor T cells with either medium, patient HC25 activated T blasts, or patient HC25 dissociated tumor tissue. T cell activation was assessed by measuring CD137 expression using flow cytometry. Asterisks indicate TCRs with selective tumor-reactivity.