RESEARCH ARTICLE SUMMARY

IMMUNOLOGY

Uneven TCR chain pairing constraints govern epitope recognition

Anastasia A. Minervina*, Mikhail V. Pogorelyy, Koshlan Mayer-Blackwell, Ricky Tirtakusuma, Stefan A. Schattgen, Andrew Fiore-Gartland, Aleksandra M. Walczak, Thierry Mora, Philip Bradley, Paul G. Thomas*

Full article and list of author affiliations: https://doi.org/10.1126/science.adx3863

INTRODUCTION: Conventional T cell recognition is mediated by the paired alpha and beta chains of the T cell receptor (TCR). Predicting target antigens from receptor sequences, and vice versa, has many diagnostic and therapeutic applications. Because TCRs are generated by a stochastic gene-shuffling process, their vast potential diversity far exceeds the limited number of T cells in any one individual. As a result, different people responding to the same antigen usually use distinct receptors, complicating efforts to derive universal rules for specificity. Even so, shared sequence motifs are commonly observed in one or both receptor chains within responses to the same molecular target.

RATIONALE: In this study, we investigated the relative roles of each chain in TCR recognition, particularly where sequence motifs are concentrated in one chain. We developed a chain replacement system to selectively remove the endogenous TCR chain across a large cell population and replace it with a chain carrying a strong motif against a known target. In effect, fixing one chain creates a biological search engine, scanning a healthy donor's own T cells to fish out the partners compatible with target binding. This allows us to observe a level of diversity in a handful of individuals that would otherwise require sampling millions of people.

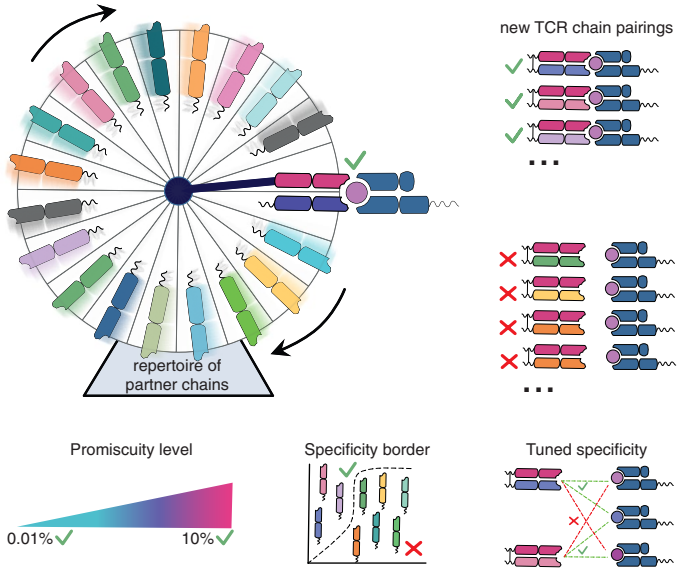
RESULTS: We found that nearly all chain combinations (>94%) can form a stable receptor on the cell surface, but only a small fraction (from ~0.1 to 10%) are able to recognize the intended target. Across 10 antigenic targets, we recovered more than 70,000 specific receptors, substantially expanding the known catalog. As most (90 to 99.9%) of the random candidate partner chains failed to maintain epitope recognition, our data highlight the limitations of relying on single-chain data alone to infer specificity.

Our results pose a problem for specificity prediction because every TCR we recovered contains one chain with a potent antigen-associated motif. Consequently, the nonbinding receptors were routinely scored as positive by leading specificity prediction algorithms that are easily deceived by single-chain matches. Retraining these algorithms with the closely related but nonbinding negatives sharpened the specificity borders of these models. As the field moves toward structural models, we found that promiscuous chains tend to make more predicted contacts with the epitope target than selective chains, and that predicted structures can resolve some, but not all, features of the specificity border.

Finally, although developed to probe the rules of recognition, partner chain exchange has practical value for tuning the specificity of TCRs to distinguish between closely related targets. We demonstrated this by engineering TCRs with improved recognition of a cancer-specific neoantigen, but with reduced self-reactivity against the nonmutated epitope.

CONCLUSION: TCR chains collaborate through an asymmetric logic, with pairing constraints compatible with recognition spanning over three orders of magnitude across different epitope targets. Mapping these constraints in sequence and structure is essential to understand how chain pairing facilitates target recognition, a prerequisite for predicting specificity from sequence alone. These insights will improve our computational models and advance predictive algorithms for diagnostic and therapeutic use. □

*Corresponding author. Email: aminervi@fredhutch.org (A.A.M); pthomas@fredhutch.org (P.G.T.) Cite this article as: A.A. Minervina *et al.*, *Science* **393**, eadx3863 (2026). DOI: 10.1126/science.adx3863



Mapping the rules of immune recognition. An experimental approach fixes one T cell receptor chain to identify compatible partners from a diverse repertoire. Screening thousands of pairings against a molecular target identifies both binding and nonbinding receptors. These data quantify the compatibility of different chains ("promiscuity"), define the precise border of immune recognition, and allow for the engineering of receptors with variable cross-reactivity ("tuned specificity").

IMMUNOLOGY

Uneven TCR chain pairing constraints govern epitope recognition

Anastasia A. Minervina[1]*, Mikhail V. Pogorelyy[1], Koshlan Mayer-Blackwell[1], Ricky Tirtakusuma[2], Stefan A. Schattgen[1], Andrew Fiore-Gartland[1], Aleksandra M. Walczak[3], Thierry Mora[3], Philip Bradley[4,5], Paul G. Thomas[1]*

Combinatorial pairing of independently recombined T cell receptor (TCR) α- and β-chains is central to diversifying the TCR repertoire. Although sequence motifs in one chain correlate with epitope recognition, the extent to which a single chain dictates specificity remains unclear. Here, we systematically tested TCR chain coupling constraints by enforcing the pairing of individual chains with hundreds of thousands of partners. Although most chains paired stably, the preservation of epitope specificity was rare and highly variable, with the frequency of compatible partners ranging from ~10 to <0.1%. This approach identified >70,000 epitope-specific TCRs across 10 epitopes. Our work illuminates the distinct contributions of TCR chains, highlights the limitations of single-chain data, and provides an experimental and analytical framework for refining TCR-peptide-major histocompatibility complex specificity inference.

Living systems frequently face the challenge of generating broad functional diversity from a finite number of genetically encoded building blocks. One prevalent solution is the modular design of heteromeric receptors which enables the detection of an extended range of signals and fine-tuning of cellular responses. This strategy is utilized by N-methyl-D-aspartate (NMDA)-receptors in neuronal signaling (*1*) and G-protein coupled receptors in cellular communication (*2*), and is central to the immune system in shaping the repertoire of both antigen-presenting major histocompatibility complex (MHC) class II molecules (*3*) and the antigen-recognizing receptors of B and T cells.

Antigen-recognizing B and T cell receptors (BCRs and TCRs, respectively) are heterodimeric receptors formed by heavy and light chains for BCRs, and α- and β-chains for conventional TCRs. The diversity of these receptors is largely attributed to somatic V(D)J-recombination, the rearrangement of genomic variable (V), diversity (D), and joining (J) segments with random addition and deletion of nucleotides at the junctions. This process yields a repertoire of ~10^{17} BCR heavy chains and ~10^{11} TCR β-chains (*4*). Random pairing of these chains with their partners amplifies this diversity, with αβTCR combinations reaching a lower-bound estimate of 2×10^{19}. Chain pairing is thus a dominant, yet relatively understudied, force in shaping the functional repertoire.

The BCR heavy chain is the primary determinant of B cell specificity, as exemplified in camelid species, where the heavy chain is the sole component of a large proportion of the antibody repertoire (*5*). Although traditional BCR heavy chains can pair with multiple light chains while retaining antigen binding (*6*–*8*)—a strategy exploited in antibody engineering (*9*–*11*)—this flexibility is not absolute. Functional antibodies often exhibit a constrained pairing termed "light chain coherence" (*12*).

In contrast to the dominant role of the BCR heavy chain, the contribution of the TCR chains to antigen recognition are more balanced, with both α and β making vital contacts with the peptide-MHC (pMHC). Computational analyses suggest that pairing within epitope-specific repertoires is highly constrained (*13*–*15*). However, "public" TCRβ chains in common immune responses (*16*, *17*), along with in vitro studies showing that a fixed TCR chain can pair with a library of modified partners to improve functional avidity (*18*–*22*), suggest that a degree of pairing promiscuity may exist for TCRs as well. This study sought to quantitatively and qualitatively define the requirements for epitope-specific TCR chain pairing. Mapping these constraints is essential to understand how TCR chains collaborate to recognize a target, a prerequisite for predicting TCR specificity from sequence alone.

The majority of TCR chains can pair

We developed an in vitro approach to quantitatively estimate the role of TCR chain pairing in constraining or facilitating epitope recognition. By introducing RNA encoding a specific exogenous TCR chain (either α or β, hereafter called exo-TCR chain) into healthy donor peripheral blood mononuclear cells (PBMCs), we produced a pool of dual-specific T cells (Fig. 1A). These modified T cells have both an endogenous αβTCR and a hybrid TCR formed by the pairing of endogenous and exogenous TCR chains. To enhance this system, we adapted a double sequential electroporation protocol (*23*). This protocol uses silencing RNA to target the constant region of the endogenous α- or β-chain to selectively knock down endogenous TCR expression while sparing the codon-optimized exo-TCR RNA. Although this partial knockdown (fig. S1, A and B) was not strictly necessary as hybrid and endogenous TCRs coexist on the cell surface (fig. S1C), it improved detection efficiency (fig. S1, D and E), likely by reducing competition between TCRs for limited CD3 molecules.

To quantify exo-TCR RNA transfection efficiency, we encoded green fluorescent protein (GFP) into the transcript. More than 90% of cells expressed GFP (mean 93.3 ± 0.54%, across five technical replicates) 18 to 24 hours post-RNA introduction (Fig. 1B). To verify the surface expression of the hybrid TCR, we tested five exo-TCR chains (two α-chains and three β-chains) followed by staining with the corresponding anti-V segment antibody. Hybrid TCRs were present on the surface of more than 94% of CD4 and CD8 cells for all tested exo-TCR chains (Fig. 1C and fig. S1F), indicating no major constraints in pairing for stable TCR expression. This observation aligns with previous computational analysis of paired αβTCR repertoires showing that chain pairing is largely independent (*24*, *25*).

Generation of epitope-specific TCRs in unexposed donors

Epitope-specific TCRs often demonstrate sequence convergence, in which different T cells use similar amino acid motifs, on one or both TCR chains (*26*, *27*). For instance, TCRs targeting the immunodominant epitopes human leukocyte antigen (HLA)-A02 $NS4B_{214\text{-}222}$ from yellow fever virus (A02-LLW) and HLA-A02 $M1_{58\text{-}66}$ from influenza virus (A02-GIL) exhibit strong sequence convergence in α- and β-chains, respectively (*27*, *28*). By contrast, their partner chains display greater diversity (Fig. 1D). We selected the TCR chain with the highest number of sequence neighbors as the "prototypical" binding chain for each epitope: TCRα for A02-LLW ($TCR\alpha_{LLW}$) and TCRβ for A02-GIL ($TCR\beta_{GIL}$).

We introduced these exogenous $TCR\alpha_{LLW}$ or $TCR\beta_{GIL}$ chains into PBMCs from four healthy individuals and tested for reactivity against the corresponding viral pMHC. After the exo-TCR chain introduction, we observed an increase in pMHC-tetramer–positive cells for all donors (Fig. 1, E and F, and fig. S2A). This demonstrates that the natural T cell pool contains many partners capable of pairing with a single exo-TCR chain to form an epitope-specific receptor. Prior to manipulation, only

[1]Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Center, Seattle, WA, USA. [2]Department of Host-Microbe Interactions, St. Jude Children's Research Hospital, Memphis, TN, USA. [3]Laboratoire de physique de l'École Normale Supérieure, CNRS, PSL Université, Sorbonne Université, and Université de Paris Cité, Paris, France. [4]Computational Biology Program, Fred Hutchinson Cancer Center, Seattle, WA, USA. [5]Institute for Protein Design, University of Washington, Seattle, WA, USA. ***Corresponding author. Email: aminervi@fredhutch.org (A.A.M); pthomas@fredhutch.org (P.G.T.)**

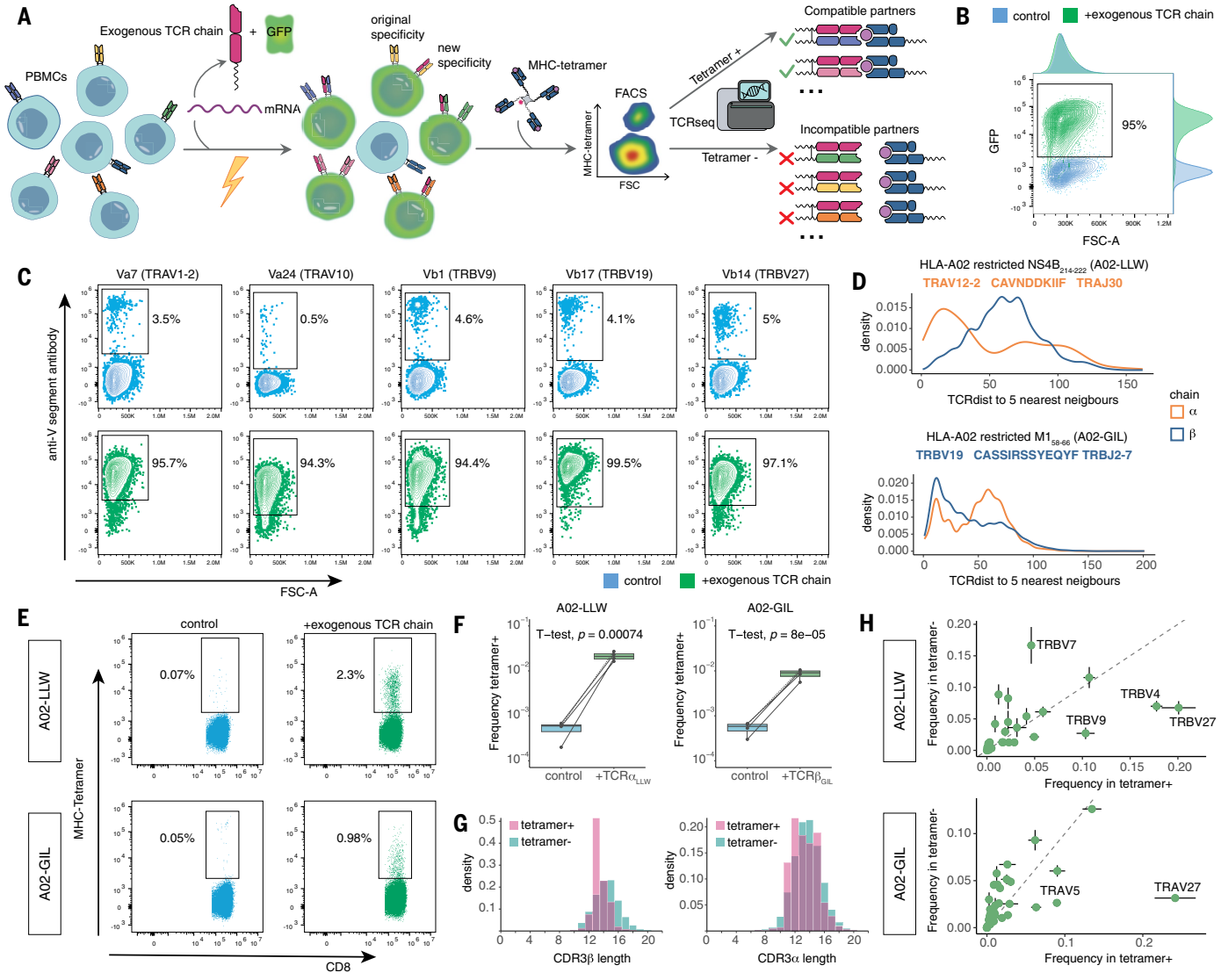


Fig. 1. Generation of epitope-specific TCRs in unexposed donors. (**A**) A schematic representation of the method. (**B**) Representative flow plot showing GFP expression in live cells 24 hours after mRNA introduction. (**C**) Representative flow plot showing the expression of the hybrid TCR on the cell surface of CD4/CD8 cells. Blue represents cells prior to exogenous chain introduction, green postintroduction. Both populations were stained with an anti-V segment antibody corresponding to the exo-TCR chain. (**D**) The distribution of TCRdist to the five nearest neighbors within α (orange) and β (blue) chains from A02-LLW (top) and A02-GIL (bottom) epitope-specific TCR repertoires from VDJdb. (**E**) Representative A02-LLW (top) and A02-GIL (bottom) MHC-tetramer staining of CD8+ cells from donor d41 before (blue) and after (green) exo-TCR chain introduction. The exo-TCR chain in both cases corresponds to a prototypical chain from epitope-specific repertoires as shown in the plot titles of (D). (**F**) Frequency of the CD8+ MHC-tetramer+ population (shown in log) before and after introduction of exo-TCR chain for four donor PBMCs. HLA-A*02 status is indicated by line type (solid, positive; dashed, negative). Two-sided paired *t*-test, $n = 4$. (**G**) The distribution of CDR3 amino acid lengths in CD8+ tetramer-negative (teal) and tetramer-positive (pink) partner chains of $TCR\alpha_{LLW}$ (left) and $TCR\beta_{GIL}$ (right). (**H**) The distribution of V-gene usage frequencies (mean ± SEM, $n = 4$) in CD8+ tetramer- (*y* axis) and tetramer+ (*x* axis) partner chains.

~0.05% of healthy donor CD8 cells were specific for either epitope. Exo-TCR chain introduction increased this frequency to 0.85% of CD8 cells for A02-GIL and 2% for A02-LLW. These percentages reflect the frequency of compatible partner chains for these motif-enriched exo-TCR chains. No increase was observed for an unrelated epitope (fig. S2B), indicating that the introduction of the exo-TCR chain did not generate broadly HLA-reactive, peptide-agnostic TCRs. We also observed a small increase in tetramer-positive CD4 cells (fig. S2C), suggesting that compatible partner chains can mediate coreceptor-independent TCR recognition.

Antigen exposures often drive substantial clonal expansions, resulting in T cell populations dominated by only a few epitope-specific clonotypes (cells sharing identical TCR sequences). By contrast, sequencing tetramer-positive hybrid TCR repertoires identified 3470 distinct compatible TCRα chains for $TCR\beta_{GIL}$ and more than 21,000 TCRβ chains for $TCR\alpha_{LLW}$. The distribution of these clonotypes across donors correlated with the number of collected tetramer-positive cells (fig. S2D and table S1).

Sequence constraints imposed on chains of epitope-specific TCRs

Although some hybrid TCRs recognized the epitope, most (>98%) failed to bind despite having a "prototypical" chain. This suggested specific constraints that govern the sequence of a compatible partner chain. We compared sequences of tetramer-positive (compatible) and tetramer-negative (incompatible) partner chains. Tetramer-positive A02-LLW TCRβ chains demonstrated a bias toward CDR3 length 13

(Fig. 1G) and a preferential usage of V-gene families TRBV27, TRBV4, TRBV9 (Fig. 1H), and J-gene TRBJ2-7 (fig. S2E). Tetramer-positive A02-GIL TCRα were enriched in V-family TRAV27 (Fig. 1H), and J-genes TRAJ42 and TRAJ37 (fig. S2E). Similar biases in length (fig. S2F), V (fig. S2G), and J-usage (fig. S2H) were observed in CD4 tetramer-positive cells which bind without CD8 coreceptor help.

We then asked whether tetramer-positive hybrid TCR repertoires were reproducible across donors. A TCR sequence similarity network for tetramer-positive chains showed clusters shared across all four donors (Fig. 2, A and B). This suggested that a donor's HLA-haplotype does not affect the single-chain repertoire because no major clusters were exclusive to a single donor. We compared our hybrid TCRs with TCRs from natural antigen exposures reported in a public database (*28*, *29*). Among 233 natural LLW-specific αβTCRs having α-chains similar to our exo-TCR, 11 matched our partner chains exactly, whereas 114 displayed high similarity (no more than two amino acid mismatches in CDR3 and the same V/J-usage; dark blue in Fig. 2C). Similarly, for 542 natural GIL-specific TCRs, 66 had exact matches, and 302 were highly similar (dark blue in Fig. 2D). Recovery of these natural partners was directly connected to their predicted frequencies in the repertoire. Public database partner chains identified in exo-TCR experiments had higher probabilities of generation (P_{gen}) compared with those not found; exact matches showed the highest P_{gen} values (fig. S2I). Thus, hybrid TCR repertoires followed the same rules as natural ones and effectively recapitulated epitope-specific sequences observed in nature.

Reversed partner chain search identifies variants of prototypical epitope-specific chains

To test the reciprocity of the system, we reversed the search. We tested $TCR\beta_{LLW}$, (instead of $TCR\alpha_{LLW}$) and $TCR\alpha_{GIL}$ (instead of $TCR\beta_{GIL}$). We selected $TCR\beta_{LLW}$ (Fig. 2E) and $TCR\alpha_{GIL}$ (Fig. 2F) from the largest clusters—both matching the motif consensus—as new exo-TCR chains to investigate the breadth of partner chains in this reversed scenario. The frequencies of tetramer-positive cells reached only 0.045% for A02-GIL and 0.19% for A02-LLW (fig. S2J), indicative of a low frequency of compatible partner chains. This suggested that the β-chain for A02-GIL and the α-chain for A02-LLW are strictly constrained,

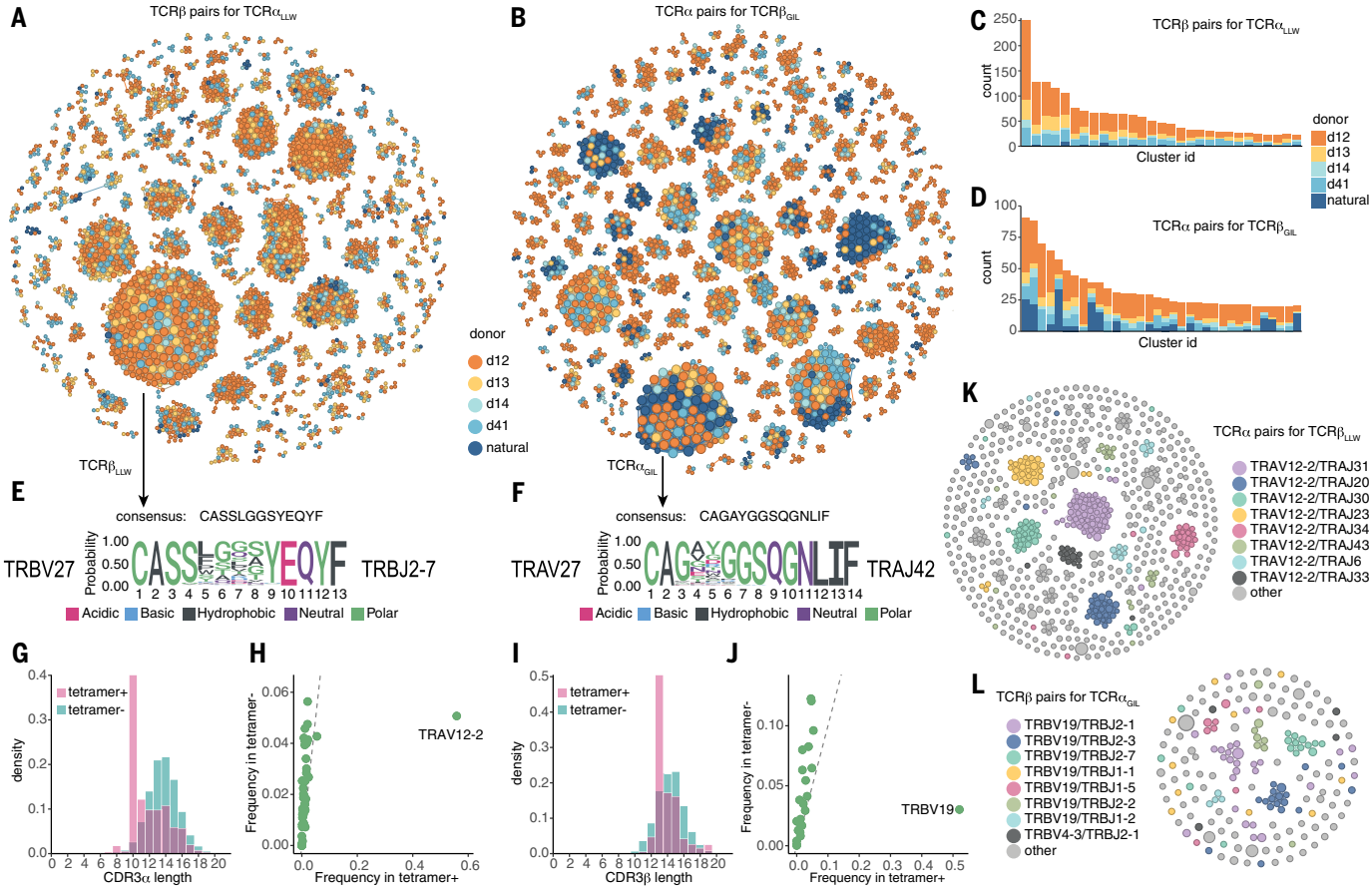


Fig. 2. Reversed partner chain search identified variants of prototypical epitope-specific chains. (**A** and **B**) The TCR similarity networks for TCRβ partners of the prototypical $TCR\alpha_{LLW}$ (A) and TCRα partners of the prototypical $TCR\beta_{GIL}$ (B). Each vertex represents a TCR clonotype and is connected with an edge if CDR3 amino acid sequences have fewer than three mismatches and V and J-genes are identical. Colors represent different donors, except for dark blue, which corresponds to natural epitope-specific sequences from VDJdb. The size of the vertex corresponds to the number of sequence neighbors. For the A02-LLW network, the repertoire was downsampled to 5000 TCRβ. Only clusters with three or more members are shown. (**C** and **D**) Cluster sharing between different donors and naturally occurring receptors for A02-LLW (C) and A02-GIL (D). The largest clusters with at least 20 members are shown as bars. (**E** and **F**) CDR3 sequence logos corresponding to the largest partner TCR chain clusters. TCR sequences corresponding to members from these clusters matching the consensus were used as exo-TCR chains in a reversed scenario. (**G** and **I**) The distribution of CDR3 amino acid lengths in tetramer- (teal) and tetramer+ (pink) partner chains of $TCR\beta_{LLW}$ (G) and $TCR\alpha_{GIL}$ (I). (**H** and **J**) The distribution of V-gene usage frequencies in CD8+ tetramer- and tetramer+ partner chains of $TCR\beta_{LLW}$ (H) and $TCR\alpha_{GIL}$ (J). (**K** and **L**) The TCR similarity networks for TCRα partners of $TCR\beta_{LLW}$ (K) and for TCRβ partners of $TCR\alpha_{GIL}$ (L). Color represents different V/J-gene segment combinations. Size of the vertex corresponds to a fraction of a clonotype in the tetramer-positive repertoire.

aligning with the initial sequence convergence analysis (Fig. 1D). Consistent with strict structural requirements, 40% of A02-LLW CDR3α had length 10 (Fig. 2G) and 55% used the TRAV12-2 gene (Fig. 2H). Approximately 50% of A02-GIL CDR3βs had length 13 (Fig. 2I), and 52% used the TRBV19 gene (Fig. 2J). The J-genes remained diverse, with a slight preference for TRAJ30 and TRAJ31 in the case of A02-LLW (fig. S2K). Among the unique A02-LLW TCRβ sequences recovered ($n = 682$), 35 matched naturally occurring A02-LLW-specific β-chains exactly (*28*, *29*) and 214 differed by up to two amino acids in CDR3, while sharing V- and J-genes. Similarly, of the A02-GIL TCRα sequences recovered ($n = 186$), 17 matched VDJdb exactly and 72 were highly similar. The reversed approach yielded a narrow repertoire closely resembling the original prototypic chains (Fig. 1D), but also revealed both subtle variations and highly divergent compatible chains (Fig. 2, K and L, and fig. S2L).

The TCRβ motifs from TCRs that recognize A02-GIL include dominant (tested here) and subdominant features (*29*). To investigate whether these shared similar constraints, we analyzed four additional GIL-specific exo-TCRβ chains from subdominant motifs, all with similar TRBV19 CDR3 sequences but different J-gene usage. These chains showed 4 to 5 times lower frequencies of MHC-tetramer–positive cells (fig. S3A) and distinct V- and J-gene pairing compared with the dominant TCRβ$_{GIL}$ (J2-7), yet shared features such as TRAV5 enrichment and TRAJ20 depletion (fig. S3B). This suggests that although each exo-TCRβ has a distinct repertoire of compatible partner chains, there are also shared solutions (fig. S3C).

Overall, our results suggest that TCR chains contribute unequally to epitope recognition. For both A02-GIL and A02-LLW epitopes, the TCR chain with a highly conserved epitope-recognizing sequence was paired with a diverse array of complementary chains. By contrast, the less sequence-constrained chain required specific partner chains to retain epitope specificity.

Compatible partner chains are defined by quantitative predictive features

The exo-TCR method generates a large set of TCRs that do not bind the tested epitope. By design, these nonbinding αβTCRs are nearly identical to binding pairs, differing only by their partner chain. Distinguishing these "close negative" examples is a challenging benchmark for TCR specificity prediction algorithms. To evaluate supervised machine learning (ML) algorithm performance, we used data from one donor as a test set and trained models using public datasets and/or exo-TCR data from the remaining donors.

A TCRdist-based K-nearest neighbor (KNN) classifier, trained on both positive (binding) and negative (nonbinding) αβTCRs from public databases, poorly differentiated between positive and negative examples from the hold-out donor [Fig. 3, A and B; TCRαβ$_{GIL}$ area under the curve (AUC) = 0.63, TCRαβ$_{LLW}$ AUC = 0.65]. We also observed low discriminatory power when using a recently developed advanced transformer-based model, mixTCRpred (*30*) (fig. S4, A and B; TCRαβ$_{GIL}$ AUC = 0.68, TCRαβ$_{LLW}$ AUC = 0.69). The high rate of misclassification was expected given that "close negatives" are absent from the public training TCR data. Thus, with KNN, more than 98% of test TCRs were classified as positive (fig. S4, C and D). Only when both positive and negative training data from our experiment were combined with the labeled TCRs from public databases did the model performance improve substantially (Fig. 3, A and B, AUC = 0.88 for TCRαβ$_{GIL}$, AUC = 0.88 for TCRαβ$_{LLW}$). Using our data alone—without any publicly available examples—yielded similarly high classification ROC AUC (Fig. 3, A and B). These results highlight the importance of training models on both positive and close-in-sequence negative TCRs to define a clear "decision boundary."

To avoid bias from the fixed exo-TCR chain, we tested whether KNN-TCRdist could predict the compatibility of the single, nonmotif enriched partner chain (β for TCRα$_{LLW}$ and α for TCRβ$_{GIL}$). A KNN-TCRdist classifier trained only on public TCRα or TCRβ data had modest predictive power (Fig. 3, C and D; AUC = 0.75 for TCRα$_{GIL}$, AUC = 0.63 for TCRβ$_{LLW}$, fig. S4, E and F). Augmenting the training set with data from our experiments substantially boosted the performance (Fig. 3, C and D; AUC = 0.88 for TCRα$_{GIL}$, AUC = 0.89 for TCRβ$_{LLW}$). The performance remained high even with smaller subsets of our data (fig. S4, G and H), indicating that our hybrid repertoire provides a dense, informative dataset for refining TCR specificity predictions.

Defining recognition determinants at single amino acid resolution

In the exo-TCR experiment we retrieved both distant and highly similar sequences in the tetramer-positive and tetramer-negative partner chains (Fig. 3E). In the extreme cases, positive and negative partner chain sequences differed at a single position (Fig. 3, F and G). For TCRβ$_{LLW}$, substitutions often occurred at positions 5 through 9 of the CDR3 and were enriched in charged amino acids. Most of the substitutions for TCRα$_{GIL}$ occurred at position 4. This suggested that tolerated substitutions in epitope-specific TCR sequences were nonrandom, likely indicating crucial contacts within specific TCR:pMHC.

We selected five sets of single-mismatch positive and negative TCRs for both epitopes to test reactivity in vitro. Among these, TCRα1β1$_{LLW}$+ and TCRα1β1$_{GIL}$+ had exact matching partners in natural repertoires. All 20 TCRs were expressed on the surface of Jurkat cell lines (fig. S4I). All TCRs identified as positive recognized the cognate epitope across a range of concentrations (0.01 to 10μM) (Fig. 3H). Only one selected negative TCR showed any reactivity, and only at the highest peptide concentration (fig. S5A). Positive TCRs had varying functional avidity, defined by sensitivity to decreasing peptide concentrations, with TCRs with partners matching natural repertoires demonstrating the highest avidity.

To understand how a single amino acid change can disrupt TCR:pMHC binding, we generated structural models using TCRdock (an AlphaFold3-based TCR modeling pipeline) (*31*) for A02-TCRα2β1$_{GIL}$ with a representative P→F in position 4 of CDR3α (Fig. 3I). Modeling suggested that the mutation alters the conformation of the CDR3α loop and introduces additional nonpolar atoms that desolvate polar groups near the TCR:pMHC interface. The region of lowest model confidence was concentrated near the mutated residue (Fig. 3J), and the interchain confidence score across the interface (ipTM score) decreased from 0.88 (high confidence) to 0.775 (medium confidence); however, this trend was not consistent across all TCR pairs (fig. S5B). Thus, although some mutations can be structurally rationalized, distinguishing closely related nonbinders remains challenging even for structure-based models with solved structure inputs. A simpler classification task, distinguishing LLW- and GIL-specific TCRs by docking to both cognate and cross-epitopes, showed differences in ipTM scores (fig. S5C).

We next investigated whether single-mutation negative TCRs could recognize any mutant variations of the peptide. We screened all 10 negative TCRs (five for each epitope) against 133 variants of A02-LLW or A02-GIL peptides, covering every possible mutation at all positions except HLA-A02 anchors p2 and p9. Nine out of 10 negative TCRs showed no reactivity to any of the peptide variants. Tetramer-negative TCRα1β3$_{LLW}$ regained its function against the M7G version of the peptide, showing that a specific mutation on the peptide side could rescue recognition.

To explore the cross-reactivity profiles, we tested 16 tetramer-positive TCRs with varying levels of similarity to naturally occurring sequences (fig. S5D) against 133 variants of LLW or GIL peptides. All tested TCRs exhibited reactivity to numerous peptide variants, with some variants eliciting a stronger response than the wild-type (WT) sequence (fig. S5E). All LLW-specific TCRs were largely tolerant to mutations at peptide positions 6, 7, and 8, whereas GIL-specific TCRs recognized nearly all mutations at position 1 (Fig. 3K).

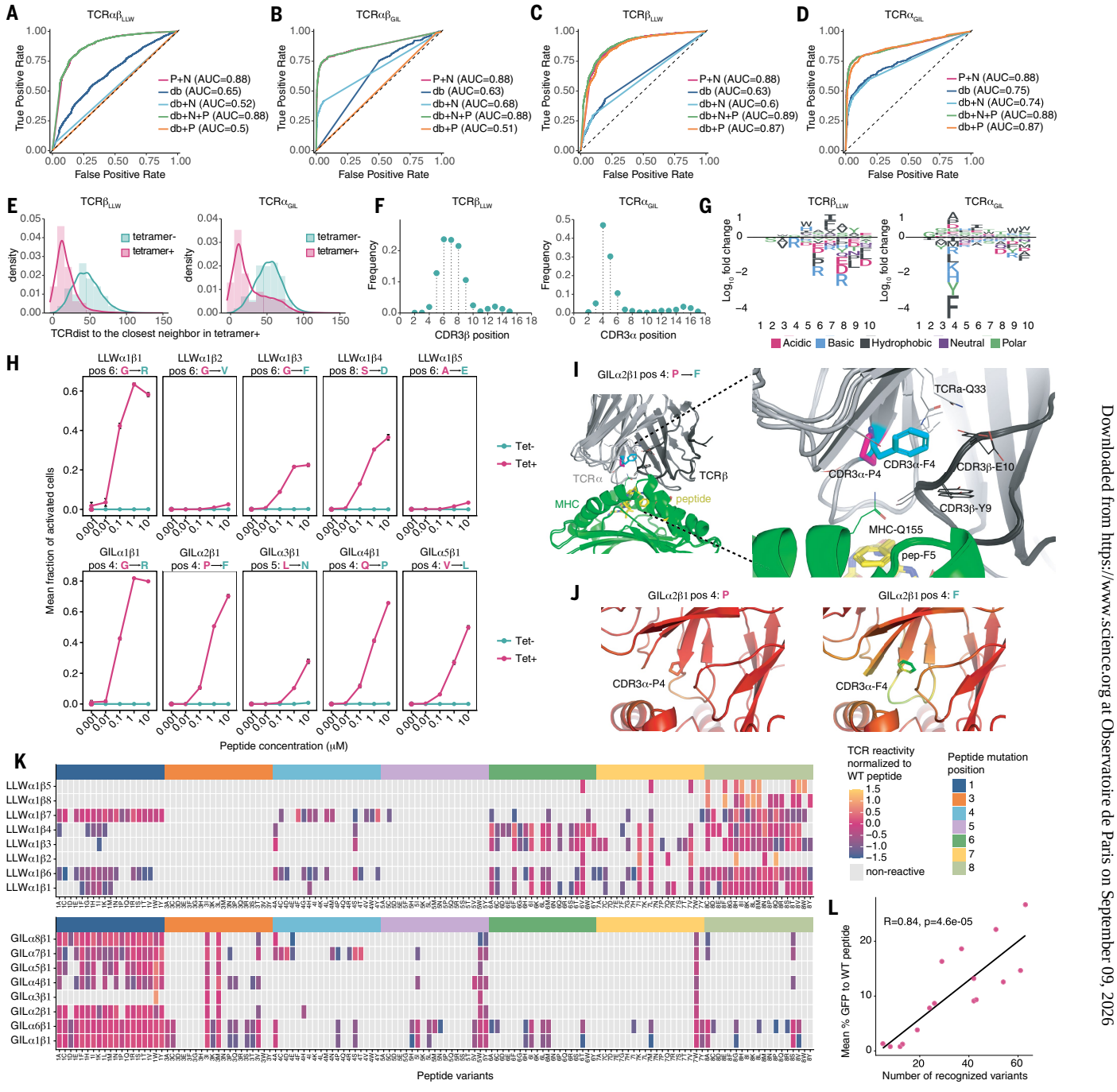


Fig. 3. Defining recognition determinants at single amino acid resolution. (**A** to **D**) ROC curves for LLW-specific TCRαβ (A), GIL-specific TCRαβ (B), LLW-specific TCRβ (C), GIL-specific TCRα (D) TCRdist-based KNN-classifiers trained on: database (db) [positive n(LLW) = 237, n(GIL) = 1818 and negative data from public databases; n(LLW) = 11073, n(GIL) = 9492], P [positive data from our experiment; n(LLW) = 16118, n(GIL) = 3196], N [negative data from our experiment; n(LLW) = 12641, n(GIL) = 27145]. (**E**) The distribution of TCRdist to the closest neighbor in the tetramer+ set for each tetramer+ sequence (pink) and each tetramer- sequence (teal). (**F**) Frequency of amino acid substitutions in different positions of CDR3 between close-in-sequence tetramer+ and tetramer-. (**G**) Sequence logo visualizing log foldchange in amino acid frequencies at the first ten positions of CDR3 between close-in-sequence tetramer+ and tetramer-. (**H**) Results of in vitro stimulation of transgenic reporter Jurkat cell lines expressing TCRs with one amino acid mismatch from tetramer+ and tetramer- sets. The reactivity was assessed through an endogenous NFAT-GFP reporter. Data points represent the mean ± SEM GFP of three technical replicates following stimulation with varying concentrations of cognate peptide in the presence of artificial APCs. (**I**) Theoretical TCRdock models of TCR:pMHC complexes of GILα2β1, differing by a single position in the CDR3α. The MHC (green) and TCR (light gray for the α-chain and dark gray for the β-chain) are shown in cartoon representation, the peptide (yellow), the mutant position as thick lines, and atoms within 6Å of the mutated side chains as thin lines. The mutant side chains are colored magenta for the binding variant and cyan for the nonbinding variant. (**J**) Theoretical TCRdock models of TCR:pMHC complexes for binding (left) and nonbinding (right) GILα2β1, differing by a single position in the CDR3α. Models are colored by AF3 confidence score (pLDDT) with red indicating high and green indicating low confidence. (**K**) Heatmap showing the reactivity of TCRs to single-position mutations of the original peptide (colored bar on top). Gray indicates peptides with no reactivity, and colors indicate reactivity normalized to that of the original peptide. TCRs are ordered by increasing TCRdist from known natural receptors (bottom to top). (**L**) Correlation between the number of peptide variants recognized by a given TCR (*x* axis) and reactivity to the original WT peptide (*y* axis). Spearman correlation R = 0.84, $P = 4.6 \times 10^{5}$.

Although the TCR chain combinations retrieved in this experiment did not undergo thymic selection, they did not exhibit a higher cross-reactivity potential compared with TCRs matching naturally occurring sequences. We observed a positive correlation between the number of recognized peptide variants and the magnitude of response to the WT peptide (Fig. 3L). This suggests that avidity for the WT epitope plays a role in determining the breadth of cross-reactivity to closely related mutants. Since TCRs matching naturally occurring sequences had higher avidity for WT peptides (TCRs α1β1 on Fig. 3H), they generally showed broader cross-reactivity profiles.

Epitope-specific TCR chain pairing promiscuity varies over orders of magnitude

We found that $TCR\alpha_{LLW}$ and $TCR\beta_{GIL}$ displayed comparable levels of TCR partner chain promiscuity, defined as the fraction of all partner chains that retain binding, at around 1% (Fig. 1F). To explore whether the degree of promiscuity varies, we examined two extreme cases of unconventional T cells: Mucosal-associated invariant T (MAIT) and invariant natural killer T (iNKT) cells. Both MAITs and NKTs exhibit highly restricted ("invariant") TCRα chains recognizing nonpeptide antigens bound to MR1 and CD1d, respectively. We hypothesized that these "invariant" α-chains would have the highest level of TCR promiscuity.

We also selected six epitopes across five HLA alleles that showed evidence of sequence convergence within one of the two TCR chains (Fig. 4A and fig. S6A). This includes four SARS-CoV-2 epitopes: HLA-A02 $S_{269\text{-}277}$ (A02-YLQ), HLA-A01 $ORF1ab_{1637\text{-}1646}$ (A01_TTD), $S_{865\text{-}873}$ (A01_LTD), and HLA-DPB1*04 $S_{167\text{-}180}$ (DPB1-TFE). Two other selected epitopes—HLA-A24 $Tax_{301\text{-}309}$ from human T cell lymphotropic virus-1 (A24-SFH) and HLA-B08 $Gag_{128\text{-}135}$ from HIV-1 (B08-EIY)—only had reported TCRβ sequences in public databases, allowing us to test whether exo-TCRβ chain introduction could retrieve their missing partners.

For all eight epitopes we observed an increase in the number of tetramer-positive cells following the introduction of the exo-TCR chain (Fig. 4B and fig. S6B). The tetramer frequency was consistent across donors, independent of HLA haplotypes (fig. S6C), but varied over several orders of magnitude between epitopes (Fig. 4C). As expected, $TCR\alpha_{MAIT}$ exhibited the highest promiscuity with almost 9% of compatible partner chains. A $TCR\beta_{EIY}$ (B08-EIY) was compatible with more than 8% of α-chains. The predicted TCRdock structure for this TCR suggests that particular contacts between the β-chain and pMHC are especially important: R10 of CDR3β (stabilized by D8) forms hydrogen bonds with peptide backbone positions 4 and 5 whereas V6 packs against W6 and I7 of the peptide (fig. S6D). Similar contact patterns were observed with three additional modeling approaches (fig. S6E). Peptide position K4 forms multiple hydrogen bonds to CDR3α, but these are to backbone atoms (fig. S6D), which may explain the higher frequency of compatible partners.

Conversely, invariant $TCR\alpha_{NKT}$ promiscuity was around 0.8%. The $TCR\alpha_{TFE}$ and $TCR\beta_{TTD}$ were the most restricted, with only 0.06% of compatible chain partners. To ensure that these measurements were not skewed as a result of the clonal structure of the repertoire, we confirmed that the fraction of tetramer-positive cells (Fig. 4D) strongly correlated with the fraction of tetramer-positive clones. The same clone size distribution in tetramer-positive populations can be obtained by random sampling (Fig. 4E). Thus, the promiscuity of chain pairing in epitope-specific TCRs varied across more than three orders of magnitude, ranging from approximately 10 to 0.06%, and likely lower across different TCRs.

The considerable difference stems from sequence constraints imposed on the compatible chain partners. All of the generated partner chain repertoires showed sequence selection as evident by preferences in CDR3 length (Fig. 4F), V-usage (Fig. 4G), J-usage (fig. S7A), and amino acid position bias (fig. S7B). Tetramer-positive MAIT TCRβs preferred V-gene families TRBV6, TRBV20, and TRBV4 (Fig. 4F), aligning with previous reports of natural MAIT TCRβ repertoires (*32*–*34*). Consistent with existing research (*35*), approximately 70% of NKT TCRβ chains used TRBV25 gene compared with a mere 0.4% in the tetramer-negative group. For all exo-TCR chains present in the public databases of naturally occurring epitope-specific TCRs (*29*), our method identified partner chains with both identical and highly similar sequences (table S2). The TCRdist-based KNN-classifier trained on positive and negative exo-TCR data from all but one donor showed a good performance for 9 out of 10 epitopes (including A02-LLW and A02-GIL) with AUCs 0.78 to 0.92 (Fig. 4H). The TCRβ classifier for DPB4-TFE recognition showed poor performance with an AUC = 0.67.

In summary, we generated hundreds to thousands of paired TCRs for specific epitopes from antigen-naive individuals. We obtained more than 70,000 unique TCRs binding classical pMHCs and ~25,000 specific for MR1 and CD1d. This substantially expands the known space of paired epitope-specific TCRs (fig. S8A).

We hypothesized that chain promiscuity depends on the distribution of interactions between TCR chains and the pMHC. To explore this, for each exo-TCR chain tested, we generated 200 TCRdock models (*31*). More promiscuous exo-TCR chains had, on average, more contacts with the peptide compared with less promiscuous chains (Fig. 4I). This trend persisted when both peptide and MHC contacts were considered (fig. S8B). Furthermore, the more promiscuous chains formed a higher number of contacts with pMHC relative to their partner chains (fig. S8, C and D). These observations suggest a link between a chain's promiscuity and its contribution to epitope binding.

Generating TCRs with rare specificity starting from a single example

Our approach succeeded with epitopes that have numerous known TCRs. However, certain TCR specificities—including neoantigen-specific TCRs—are rare, often with only one known example. We tested whether a single TCR can be used in an exo-TCR chain approach to identify more TCRs binding the neoantigen.

As a proof of concept, we chose a rare TCR specific to a shared A68-restricted neoepitope derived from the characteristic Fibrolamellar carcinoma (FLC) gene fusion driver mutation (*36*). We used TCRα or TCRβ as exo-TCR chains and stained cells with a pMHC-tetramer loaded with the neoantigen peptide. Although the introduction of the $TCR\alpha_{FLC}$ failed to increase the frequency of tetramer-binding cells, $TCR\beta_{FLC}$ increased the number of binding cells from 0.01 to 0.047% (Fig. 5A). The tetramer-positive repertoire consisted of ~390 α-chains with distinct V- and J-gene usage (Fig. 5, B and C), suggesting epitope-driven selection. By contrast, the 65 identified β-chains lacked distinct sequence features (fig. S8, E and F), suggesting that they likely represent background noise.

The tetramer-positive TCRα sequences formed clusters of highly similar clones, with one major cluster being similar to the original receptor's α1-chain (Fig. 5D). We selected 11 representative α-chains from different sequence clusters for in vitro validation (Fig. 5D, pink-outlined vertices). All 11 TCRα chains paired with the exo-TCRβ were activated in a peptide stimulation assay (Fig. 5E and fig. S8G).

The original FLC-specific α1β1 TCR had weak discriminatory power between the neoepitope and the self-peptide, which differs by a single amino acid (*36*). All tested hybrid TCRs recognized the neoepitope but showed reduced or no activity against the self-peptide (Fig. 5E). Therefore, introducing a specific exo-TCR chain can expand a single TCR into a pool of binders with altered recognition of epitope variants. This approach reduces the randomness of natural recombination to achieve a discovery scale that would otherwise require sampling thousands or even millions of individuals (Fig. 5F).

Discussion

Despite advancements in ML, decoding TCR specificity from sequence remains elusive, with predictive accuracy limited by the amount of

A

Epitope		Source	Exo-TCR chain
MR1	5-OP-RU	MAIT	TRAV1-2 CAVMDSNYQLIW TRAJ33
B08:01	EIYKRWII	HIV-1	TRBV9 CASSVVGDSRETQYF TRBJ2-5
A02:01	LLWNGPMAV	YFV	TRAV12-2 CAVNDDKIIF TRAJ30
A02:01	GILGFVFTL	Influenza	TRBV19 CASSIRSSYEQYF TRBJ2-7
CD1d	α-GalCer	NKT type I	TRAV10 CVVSDRGSTLGRLYF TRAJ18
A01:01	LTDEMIAQY	SARS-CoV-2	TRBV29-1 CSVGTGWPYEQYF TRBJ2-7
A24:02	SFHSLHLLF	HTLV-1	TRBV7-9 CASSPDRAQTQYF TRBJ2-5
A02:01	YLQPRTFLL	SARS-CoV-2	TRBV7-9 CASSLDIEAFF TRBJ1-1
DPB1:04:01	TFEYVSQPFLMDLE	SARS-CoV-2	TRAV25 CAGQNYGGSQGNLIF TRAJ42
A01:01	TTDPSFLGRY	SARS-CoV-2	TRBV27 CASSLRSGTDTQYF TRBJ2-3

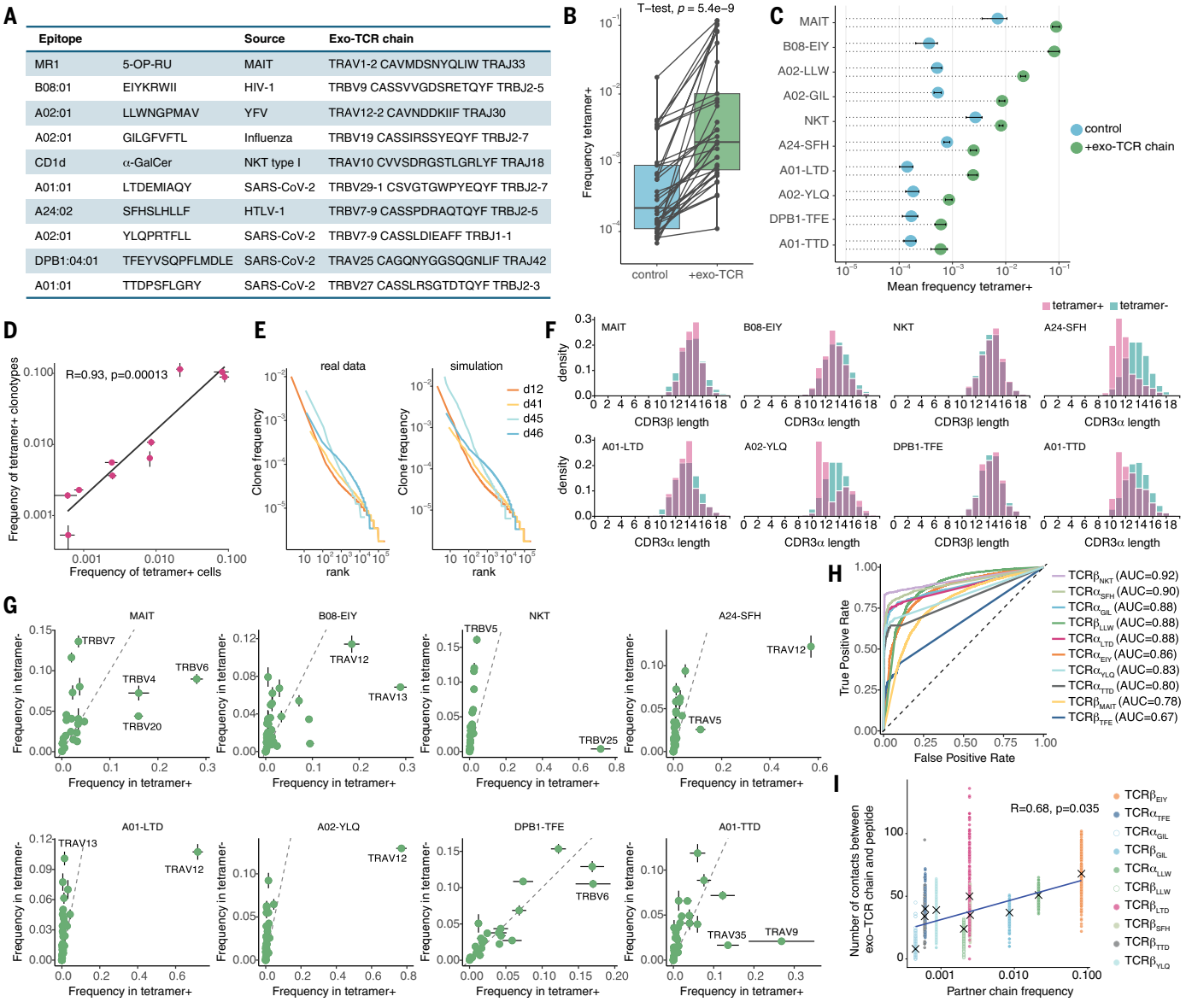


Fig. 4. Epitope-specific TCR chain pairing promiscuity varies over orders of magnitude. (**A**) Exo-TCR chain sequences and specificities. (**B**) Frequency of tetramer-positive population (shown in log) is significantly increased after exo-TCR chain introduction (two-sided paired *t*-test, $n = 33$, error bars: SEM). (**C**) Mean tetramer-positive frequency (of CD4+ cells for NKT and DPB1-TFE and of CD8+ cells for the rest, shown in log) varies over orders of magnitude for different exo-TCR chains. Error bars show SEM; $n = 5$ for DPB1-TFE, rest $n = 4$. (**D**) Correlation between chain promiscuity measured at the level of cells (*x* axis, shown in log; mean ± SEM) and clonotypes (*y* axis, shown in log; mean ± SEM). $n = 5$ for DPB1-TFE, rest $n = 4$. Spearman correlation R = 0.93, p = 0.00013. (**E**) Representative clone size distribution of compatible TCRα partner chains for $TCR\beta_{EIY}$. (Left) real data; (right) random assignment of clonotype to tetramer+ and tetramer- at the same frequencies as real data. (**F**) The distribution of CDR3 amino acid lengths in tetramer- (teal) and tetramer+ (pink) partner chains. (**G**) The distribution of V-gene usage frequencies (mean ± SEM; $n = 5$ for DPB1-TFE, rest $n = 4$) in tetramer- (*y* axis) and tetramer+ (*x* axis) partner chains. (**H**) ROC curve for the TCR partner chain classifier trained on our data and tested on an unseen donor. (**I**) Correlation between the number of contacts between exo-TCR chains and peptides in TCRdock TCR:pMHC models and chain promiscuity (Spearman correlation for median values indicated by crosses, $n = 200$, R = 0.68, p = 0.035).

available data (*37*, *38*). Retrieving epitope-specific TCRs from natural repertoires is challenging as a result of the scarcity of exposed donors and the low frequencies of antigen-specific clones. Consequently, the field has increasingly shifted to synthetic screening. Numerous in vitro methods have been developed for TCR screening against epitope libraries (*39*–*50*) or to increase TCR diversity for known epitopes using artificial CDR3 sequences (*18*–*20*) or TCR chain libraries (*21*, *22*). Here, we present an approach that combines natural TCR recombination rules with the high-throughput capabilities of synthetic platforms.

An individual TCR was sufficient to generate hundreds of binding TCRs with altered specificity profiles through the exo-TCR method. Introducing an exo-TCR chain bypasses the randomness of the V(D)J-recombination process by "fixing" one of the two chains. Consequently, identifying the same TCR sequences in natural repertoires would require a sample size 10^4 to 10^{10} times larger. Although the exo-TCR method explores sparse TCR sequence spaces, it is still limited by the natural V(D)J-recombination constraints. Thus, certain TCR chain partners may remain extremely rare in natural repertoires even with targeted sampling. Synthetic libraries of CDR3-only mutated chains designed for a specific epitope could overcome these limits (*18*, *20*), albeit at the cost of determining the natural frequency of partner chains.

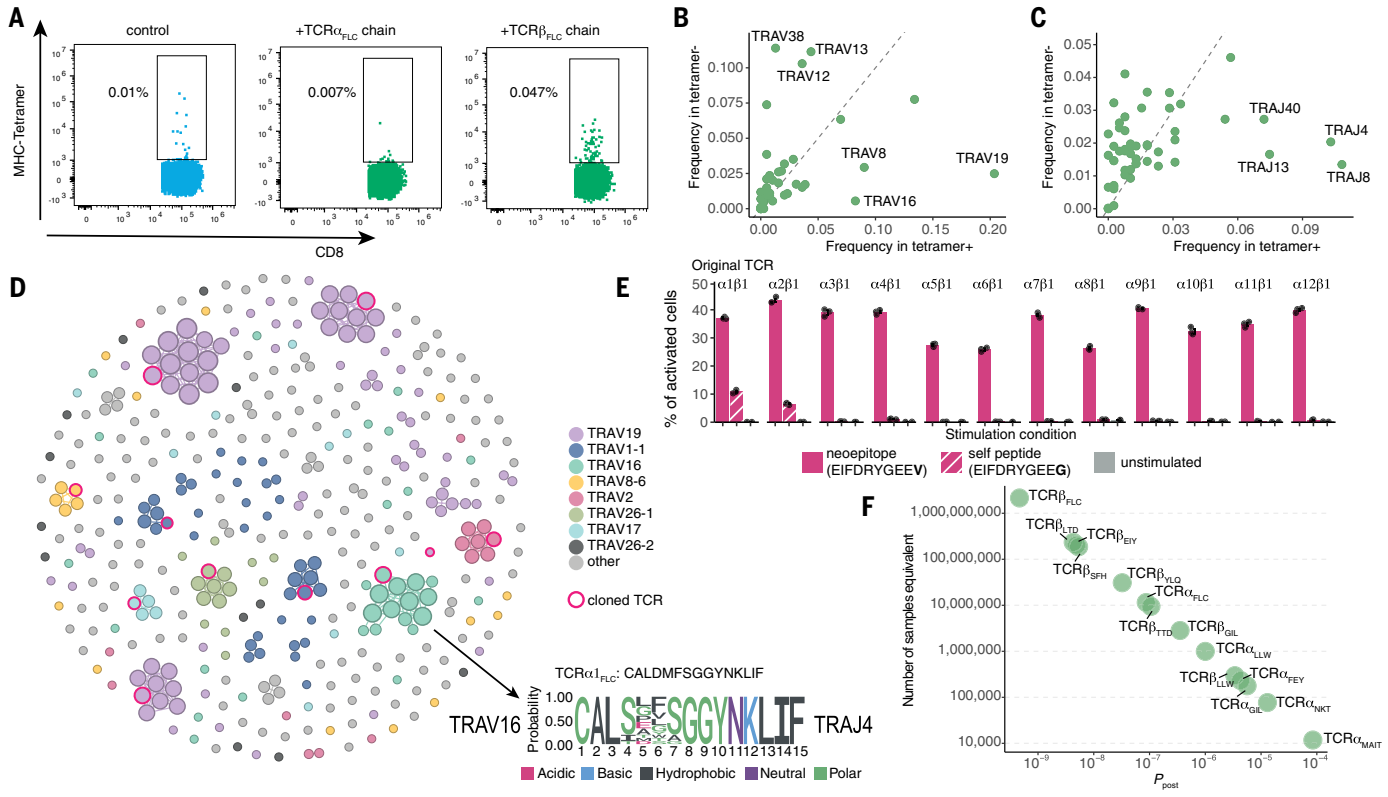


Fig. 5. Generating novel TCRs with rare specificity starting from a single example. (**A**) Frequency of FLC neoepitope-loaded A68-tetramer+ populations before (blue) and after introduction of the TCRα or TCRβ chain (green). (**B**) The distribution of V-gene usage frequencies in tetramer- (*y* axis) and tetramer+ (*x* axis) partner α-chains of $TCR\beta_{FLC}$. (**C**) The distribution of J-gene usage frequencies in tetramer- (*y* axis) and tetramer+ (*x* axis) partner α-chains of $TCR\beta_{FLC}$. (**D**) TCR similarity networks for TCRα partners of $TCR\beta_{FLC}$. Each vertex represents a TCR clonotype and is connected with an edge if CDR3 amino acid sequences have fewer than three mismatches and V-genes are identical. Color represents different V-genes. The size of the vertex corresponds to the number of sequence neighbors. Pink outlined TCRs were selected for cloning. Teal cluster CDR3 sequence motif showed close similarity to the original $TCR\alpha1_{FLC}$ chain. (**E**) Results of in vitro stimulation of transgenic reporter Jurkat cell lines expressing newly identified TCRα paired with the original $TCR\beta_{FLC}$. The reactivity is assessed through an endogenous NFAT-GFP reporter. Data points represent the mean ± SEM GFP of three technical replicates following stimulation with 0.1 μM of cognate neoepitope peptide or self-peptide in the presence of artificial APCs. TCR α1β1 is the original FLC-specific TCR isolated from the patient. (**F**) Estimation of the sample size equivalent for experiments introducing exo-TCR chains with varying generation probabilities.

Most TCRs produced through exo-TCR chain introduction closely resemble naturally occurring sequences. However, while the individual chains have undergone thymic selection, their specific pairs have not. Although surface expression confirmed that most chains form stable heterodimers, it remains unclear whether some tetramer-negative TCRs would fail positive selection due to an inability to bind HLA. Addressing this requires large-scale, naturally paired TCR datasets, which are becoming accessible through high-throughput paired TCR sequencing (*51*, *52*).

Conversely, tetramer-positive hybrid TCRs may represent receptors that would fail negative selection. We did not observe peptide-agnostic recognition in hybrid TCRs, as they showed no binding to unrelated pMHC during tetramer staining or peptide stimulation experiments. Although generated hybrid TCRs may carry a lower risk as a result of generation from natural sequences, they would still require rigorous testing to assess potential self-reactivity.

The data presented here largely relied on pMHC-tetramer staining, which may not always translate to activation in vivo because of reverse docking (*53*), insufficient functional avidity (*54*), and other factors. Hybrid TCRs resembling natural ones exhibited higher avidity for cognate epitopes and broader cross-reactivity toward similar ligands, suggesting that the exo-TCR method samples a wider avidity range, including weak responders. Although the hybrid repertoire is governed by recombination statistics, naturally selected TCRs are further shaped by avidity thresholds that are enforced by antigen-driven selection. Notably, none of our functionally characterized hybrid TCRs failed to activate with cognate peptide exposure.

We observed that although nearly all TCR chain pairings can form a stable TCR (>94%), the proportion of compatible partner chains that maintain pMHC binding to the original target varies greatly from 0.06 to 10% for the TCR:pMHCs examined in this study. Even the most promiscuous chains lost specificity in 90% of pairings. Thus, knowing only partial information from a single chain is insufficient to accurately predict the specificity of an individual TCR. Still, single-chain sequences can be predictive for large cohorts with known antigen exposures, in which antigen-driven selection enriches well-curated motifs, such as in immunodominant cytomegalovirus (CMV) (*17*) or SARS-CoV-2 epitopes (*16*). Disease-status prediction from single-chain data could benefit from incorporating the binding chains from our study. By contrast, adding large numbers of negative partner chains is likely detrimental, as they may dilute or obscure true positive signals in single-chain classifiers. These complexities demonstrate the need for careful curation of the training and test datasets for predictive applications (*37*), in which the context of the targeted problem is key.

As we aim for accurate TCR target decoding, paired chain TCR analyses will be required. Current specificity prediction algorithms are susceptible to prototypical chain bias, assuming a single binding chain guarantees specificity. This limitation derives from additive

scoring frameworks, in which the similarity score of a paired receptor is simply a sum of individual chain scores. Here, we demonstrate that functional recognition is a nonlinear property of the αβ heterodimer and propose that future models adopt a more chain-balanced approach that accounts for interchain interactions. Overall, positive and negative data generated in this study provide a resource for future efforts in defining the requirements for de novo prediction success in a real-world scenario, where the same TCR chains can be specific to different epitopes depending on their partner.

These data similarly provide a rigorous challenge to structurally based TCR-decoding efforts. Algorithms such as TCRdock (*31*) are demonstrating impressive predictive potential for specific epitopes and classifying broad binding versus nonbinding behavior (fig. S5C). The ultimate solution to the TCR-decoding problem will necessarily involve structure-level resolution modeled in some form. The ambiguous resolution for some of our single mutant binder/nonbinder pairs demonstrates that there are still gains to be made in the current state-of-the-art models, but the difficulty of this problem (resolving single point mutants that destroy recognition) provides a true test of decoding progress.

Highly promiscuous TCR chains are more likely to find a match during the natural recombination process, resulting in a higher precursor frequency of T cells with the same specificity. Their promiscuity is likely mediated at least in part by their ability to make many stabilizing contacts, freeing their partner chains to explore greater diversity due to fewer structural constraints, as observed in the B08-EIY β-chain analysis and mutagenesis experiments for A02-LLW and A02-GIL. This raises questions as to what extent the immunodominance of certain epitopes (*55*) may be explained by their recognition by TCRs with promiscuous chains and how the immune system maintains a balanced naive repertoire without overrepresenting particular specificities. The interplay of highly promiscuous chains and their generation probabilities (*56*)—which spans more than 10 orders of magnitude—is likely a key mechanism underlying why only some epitopes are highly recognized by the immune system. Addressing these questions could offer valuable insights into the organization of the immune repertoire.

Materials and Methods

Primary cells

PBMCs from healthy donors were obtained from apheresis rings through the St. Jude Blood Donor Center under Department of Pathology protocol BDC035. Donors provided informed consent for research use of the anonymized, discarded apheresis ring samples.

HLA typing

HLA typing was performed using the AllType NGS 11-Loci Amplification Kit (One Lambda) according to manufacturer's instructions. The resulting libraries were sequenced on Illumina at 150×150bp. HLA types were called using the TypeStream Visual Software (One Lambda). HLA typing results can be found in table S3.

Generation of silencing RNAs

For endogenous TCR chain knockdown we designed dsiRNAs–a 27-mer duplex siRNA optimized for Dicer processing–targeting constant segments of TCRα (GrArGrArArArArGrCrUrUrUrGrArArArCrArGrArUrArCGArUrCrGrUrArUrCrUrGrUrUrUrCrArArArGrCrUrUrUrUrCrUrCrGrA) and TCRβ (CrArCrUrUrCrCrGrCrUrGrUrCrArArGrUrCrCrArGrUrUCTrArGrArArCrUrGrGrArCrUrUrGrArCrArGrCrGrGrArArGrUrGrGrU) chains using IDT online custom dsiRNA design tool. The target site for TCRβ was selected to be the same in TRBC1 and TRBC2. To test the designed dsiRNAs we used a Jurkat cell line with overexpression of human noncodon-optimized TCR. The TCR and CD3 expression was substantially down-regulated in comparison to the control after 48 hours following dsiRNA (1μM per million cells) electroporation with a more pronounced effect for dsiRNA against TRAC (fig. S8H). We then electroporated healthy donor PBMCs, with dsiRNA aimed at the endogenous TCR chain that we intended to replace with the exo-TCR chain. For instance, we used dsiRNA against the TRAC segment when replacing the endogenous TCRα with a fixed one. Following 48 hours, these cells were further electroporated with 8μg of the desired exogenous TCR chain RNA.

Exo-TCR chain RNA generation

Amino acid sequences of CDR3 and V/J regions for exo-TCR chains are provided in table S4. Nucleotide sequences for exo-TCR chains were generated using stitchR (*57*) and were codon-optimized to ensure the absence of dsiRNA target sites. Additional 5′ sequence containing Kozak sequence was added to the fragment (GCTAGCAGACCCAGTCTTGGTACCGAGCTCGGATCCACTAGTAACGGCCGCCAGTGTGCTGGAATTCTGCAGATATCACAAGTTTGTACAAAAAAGCAGGCTTTAAAGGAACCAATTCAGTCGACTGGATCCGGTACCGAATTCGGCGCGCCAGATATCACACGTGCCAAGGGGCTGGACAAGCTTGCCGCCACC). Gene fragments were synthesized by Genscript and cloned into the pcDNA3.1(+)-P2A-eGFP vector using the NheI and XbaI restriction sites. 10μg of the plasmid was linearized using the PspOMI restriction enzyme (NEB) for 1 hour at 37°C in a 50μl volume with 40 units of enzyme. The linearized plasmid was purified using a QIAquick PCR purification kit (Qiagen) according to the manufacturer's protocol and quantified on a NanoDrop (Thermo Scientific). 1μg of the linearized plasmid was used as template for in vitro transcription (IVT) using the HiScribe T7 ARCA mRNA Kit with tailing (NEB) according to the manufacturer's protocol. In brief, the IVT reaction was held at 37°C for 2 hours, followed by a 15 min DNA digestion at the same temperature, and then followed by a 40-min polyA-tailing at 37°C. RNA was precipitated using standard LiCl protocol following a 2-hour to overnight incubation at -20°C. The RNA-containing solution was precipitated at 15,000 g at 4°C for 20 min. The RNA pellet was washed with 80% ethanol and air-dried. The RNA pellet was then resuspended in nuclease-free water. Its concentration was measured using NanoDrop (Thermo Scientific).

Cell culture and transfections of primary T cells

On day 1 of the developed protocol, PBMCs from healthy individuals were thawed and plated in 12-well tissue culture plates, with up to 10 million cells in a well. These were placed in 2 mL RP-10 media (RPMI 1640 (Gibco, 22400-089) supplemented with 10% heat-inactivated human AB serum (Gemini Bio-Products), 1% non-essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), and 100 U/mL penicillin-streptomycin). The media was further supplemented with 3000 IU/mL human recombinant IL-2 (Stemcell Technologies, 78036) and 0.05 μg/mL of anti-CD3 (clone OKT3, Miltenyi Biotec). On day 4, cells were washed with 50 mL DPBS (Gibco) and resuspended in 100μl Opti-Mem (Gibco) for electroporation. A fraction of the cells was set aside as a non-transfected control. Up to 10 million cells in a volume of 100μl, along with dsiRNA targeting one of the TCR chains (at a concentration of 1μM per million cells), were then transferred into a 2 mm-gap electroporation cuvette (BTX). Electroporation was done at a single square wave of 400V for 0.5 ms using either a BTX ECM830 or Bio-rad Gene Pulser Xcell Total Electroporation system. Following electroporation, the cells were immediately transferred to a 12-well plate containing 2 mL warm RP-10 supplemented with 3000 IU/mL IL-2. Control non-electroporated PBMCs were cultured under the same conditions. On day 6, approximately 48 hours after dsiRNA electroporation, up to 10 million cells were electroporated with 8μg of exo-TCR chain RNA using the same electroporation protocol. On day 7, cells were stained and used for flow cytometry analysis and cell sorting (FACS).

Cell staining and sorting

Before staining, the cells were washed with DPBS (Gibco) and then incubated for 40 min in 1 mL DPBS containing 50nM dasatinib (Sigma-Aldrich) at 37°C, 5% CO_2 to increase tetramer staining efficiency (*58*).

Following this, the cells were pelleted at 500 g for 5 min and resuspended in 50µl DPBS with 50 nM dasatinib containing 2 µl TruStain FcX (Biolegend, 422302) and 1µg of the corresponding MHC-tetramer for up to 5 million cells. Samples larger were scaled accordingly. They were then incubated for 30 min at room temperature in the dark. After this, additional surface antibodies diluted in DPBS were added until the final volume reached 100 µl. This mixture was then incubated for another 30 min at room temperature in the dark. We used a 1:100 dilution of Ghost Dye Violet 510 Viability Dye (Tonbo Biosciences, 13-080-T100), anti-human CD4 PerCP-Cy5.5 (clone OKT4, 317428, Biolegend), anti-human CD8 Brilliant Violet 785 (clone SK1, 344740, Biolegend) or anti-human CD8 APC (clone SK1,344721,Biolegend) antibodies to stain up to 10 million cells. A 1:50 dilution of PE-conjugated or APC-conjugated anti-human Vβ1 (clone REA662, 130-126-568, Miltenyi Biotec), Vβ17 (clone REA915, 130-115-347, Miltenyi Biotec), Vβ14 (clone REA557, 130-127-511, Miltenyi Biotec), Vα24 (clone REA948, 130-115-786, Miltenyi Biotec), and a 1:10 dilution of PE-conjugated anti-human Vα7.2 (clone 3C10, 351705, Biolegend) were used according to the manufacturer's recommendations. After staining, cells were washed twice with 1 mL DPBS and strained through CellTrics 50µm filters (Sysmex) before FACS. A Sony SY3200 or Cytoflex SRT was used for cell sorting. Cells were collected into tubes containing complete RPMI (RPMI 1640 (Gibco, 22400-089), with 10% FBS, 2mM L-glutamine and 100 U/mL penicillin/streptomycin). Sorted cells were then pelleted at 500 g for 5 min at 4°C. The supernatant was discarded and the cell pellet was resuspended in 1mL of Trizol reagent (Invitrogen) and stored at -80°C. Flow cytometry data analysis was performed using the FlowJo version 10.8.1 software (BD Biosciences). The number of tetramer-positive cells collected per donor is provided in table S1.

MHC-tetramer generation

The Brilliant Violet 421-conjugated MR1-tetramer loaded with 5-OP-RU and the Alexa Fluor 647-conjugated CD1d-tetramer loaded with PBS-57 were obtained from the NIH Tetramer Core Facility. The biotinylated DPB1*04:01 monomer loaded with the TFEYVSQPFLMDLE peptide from the SARS-CoV-2 spike was a generous gift from Jamie Rossjohn and was prepared as described in (*59*). For tetramerization, one volume of PE-conjugated streptavidin (Biolegend) was added to one volume of HLA-DP4-monomer (1 mg/mL). The volume of PE-streptavidin (0.2 mg/ml) was divided into four parts and added in four consecutive steps, each followed by a 10-min incubation period. After all required amounts of PE-streptavidin were added, the mixture was incubated for at least 1 hour on ice before staining.

Biotinylated easYmer HLA-monomers (Immudex) were loaded with minimal peptides as per the manufacturer's protocol. The corresponding peptides, synthesized commercially with over 90% purity (Genscript), were diluted to 1mM in DMSO. To tetramerize 25µl of loaded easYmer, we used 1µl of a 0.2 mg/mL Streptavidin-PE (Biolegend, 405204) or 1µl of a 0.2 mg/mL Streptavidin-APC (Biolegend, 405207) and incubated the mixture for at least an hour on ice before staining.

Artificial antigen-presenting cell lines

The full-length coding sequences of HLA-A*02:01 and HLA-A*68:02 were downloaded from the IMGT/HLA database (*60*). Gene fragments were then synthesized by Genscript and cloned into the lentiviral pLVX-EF1α-IRES-Puro (Clontech). 293T packaging cells (ATCC CRL-3216) were transfected with HLA-encoding lentiviral vectors, the psPAX2 packaging plasmid (Addgene plasmid #12260), and the pMD2.G envelope plasmid (Addgene plasmid #12259) using Lipofectamine 3000 Transfection Reagent (Invitrogen). Lentivirus was collected 24 and 48 hours post-transfection, strained through a 0.45µm SFCA filter (Thermo Scientific), and then concentrated using Lenti-X Concentrator (Takara Bio) according to the manufacturer's protocol. To generate monoallelic HLA class I antigen-presenters, HLA-deficient K562 cells (ATCC CCL-243) were transduced with lentivirus. At 48 hours post-transduction, cells were transferred to Iscove's Modified Dulbecco's Medium media containing 10% FBS, 2mM L-glutamine, 100 U/mL penicillin/streptomycin, and 2µg/mL puromycin (Gibco). The antibiotic selection continued for one week. Surface HLA expression was confirmed by staining with a pan-HLA class I antibody (clone W6/32, PE-conjugated, 311406, Biolegend; or APC-conjugated, 311410).

TCR-expressing Jurkat cell lines

Amino acid sequences of CDR3 and V/J regions for selected αβTCRs are provided in table S5. Nucleotide sequences of selected TCRα and TCRβ chains were generated using stitchr (*57*). Both chains were modified to incorporate mouse constant segments (TRAC*01 and TRBC2*01) to enhance TCR surface expression. Given that the same chains are often reused by many of our tested TCRs, gene fragments for TCRα and TCRβ chains were separately synthesized by Genscript and cloned into either the lentiviral pLVX-EF1a-P2A-mCherry-IRES-Puro (Clontech) or pLVX-EF1α-IRES-G418 (modified in-house from pLVX-EF1α-IRES-Puro). We then transfected 293T cells (ATCC CRL-3216) with lentiviral vectors, psPAX2 packaging plasmid (Addgene plasmid #12260), and pMD2.G envelope plasmid (Addgene plasmid #12259) using Lipofectamine 3000 Transfection Reagent (Invitrogen). Lentivirus was collected 24 and 48 hours post-transfection, then concentrated using Lenti-X Concentrator (Takara Bio) according to the manufacturer's protocol. To generate Jurkat cells expressing the paired TCR, TCR-null Jurkat 76.7 cells with NFAT-eGFP reporter (a kind gift from Fumihiro Fujiki) were co-transduced with lentiviruses encoding the TCRα and TCRβ chain. At 48 hours post-transduction, cells were transferred to RPMI media containing 10% FBS, 2mM L-glutamine, 100 U/mL penicillin/streptomycin, 1µg/mL puromycin (Gibco), and 500µg/mL Genectin (Gibco). The antibiotic selection with both antibiotics continued for one week. The surface expression of abTCR was confirmed by mCherry expression and staining with a 1:100 dilution of anti-mouse TCRβ APC-Fire750 (clone H57-597, 109246, Biolegend) and anti-human CD3 Brilliant Violet 421 (clone SK7, 344834, Biolegend).

Specificity validation of TCR-expressing Jurkat

To validate the specificity of the generated TCR-expressing Jurkat cell lines, 1–1.5 × 10^5 Jurkat cells were co-cultured with equal amounts of aAPCs expressing the corresponding HLA allele. Cells were cultured in RPMI media containing 10% FBS, 2mM L-glutamine, 100 U/mL penicillin/streptomycin, and 1µg/mL each of anti-human CD28 (BD Biosciences, 555725) and CD49d (BD Biosciences, 555501) in round-bottom 96-well tissue culture plates. The co-cultures were pulsed with LLWNGPMAV or GILGFVFTL peptides at 0.001-10µM in tenfold dilutions. For A68-FLC specific Jurkat cell lines, the co-culture was carried out with 0.1µM of either EIFDRYGEEV neoepitope or EIFDRYGEEG self-peptide. All cultures were done in triplicates. DMSO, at concentrations equal to those found in peptide stimulation wells, was added to the unstimulated control wells (no peptide). Cells were incubated for 16-18 hours at 37°C and with a 5% CO_2. After incubation, cells were washed once in DPBS (Gibco) and resuspended in 300µl DPBS for flow cytometry analysis on a custom-configured BD Fortessa with FACSDiva software (Beckton Dickenson). Flow cytometry data analysis was done using FlowJo version 10.8.1 software (BD Biosciences). Figure 3H and fig. S5A show GFP expression after subtracting GFP background from unstimulated control wells.

Evaluating TCR-expressing Jurkat reactivity to peptide variants

LLWNGPMAV and GILGFVFTL peptide variants were synthesized as a crude library by Genscript and diluted to a 1 mM concentration in DMSO. Each peptide included 133 variants with all possible amino acid mutations at positions 1, 3, 4, 5, 6, 7, and 8. Mutations in HLA anchor positions (2 and 9) were excluded. To examine the specificity of transgenic Jurkat cell lines against the peptide variants, we co-cultured 50 000 Jurkat cells with equal amounts HLA-A*02:01 expressing K562 in

96-well round-bottom tissue culture plates in RPMI media containing 10% FBS, 2mM L-glutamine, 100 U/mL penicillin/streptomycin. Peptide variants of LLWNGPMAV or GILGFVFTL were added to the co-cultures at a final concentration of 1 μM. To evaluate GFP reactivity we used a fluorescence microscopy-based acquisition assay that showed strong correlation with the flow cytometry results (fig. S8I). After a 20 hour long co-culture we measured GFP percentage using Cellcyte X (Echo) with 4x objective in spheroid mode, acquisition in green channel (800 ms exposure 3dB gain). Image analysis was done with CellCyte Studio software with "total spheroid area" recipe and standard defaults (50 a.u. Contrast Sensitivity, 2 a.u Smoothing, 100 um^2 filled hole size 100 um Min. object size), relative GFP+ spheroid area was used as the output. Triplicates of "no peptide" negative controls as well as non-mutated peptide stimulations were included to establish positivity thresholds. Positivity was defined as either achieving a signal at least three standard deviations above the mean of the negative control or reaching a minimum relative spheroid area of 0.5% GFP. Additionally, all images generated by Cellcyte X were manually reviewed for quality assurance. Wells identified as GFP-positive were manually reviewed for quality assurance to exclude GFP-positivity due to image artifacts ("false positives"). "False positive" wells were re-analyzed using stricter thresholds. The relative reactivity against peptide variants was determined by calculating the log10 ratio of the reactivity of each variant to the mean reactivity of the corresponding unmutated peptide.

TCR sequencing

The TCRα and TCRβ repertoires for endogenous TCRs were generated using the 5′RACE RNA-based protocol as previously described (*61*). In brief, RNA was isolated using the Trizol reagent (Invitrogen) as per the manufacturer's protocol. All RNA isolated from tetramer-positive fractions, and up to 1200 ng of RNA from tetramer-negative fractions (quantified using the Qubit RNA HS kit from Invitrogen), was used for cDNA synthesis with the SmartScribe kit (Takara). This process used a template switch oligonucleotide containing unique molecular identifiers (UMIs) and primers specific to TCR constant segments of alpha and beta chains. The cDNA was then purified with a 1xAmpure XP beads sample ratio (Beckman Coulter) and amplified in two rounds of PCR using Q5 high-fidelity polymerase (NEB). Adapters necessary for sequencing on the Illumina platform were introduced with the KAPA HyperPrep kit (Roche). Libraries were sequenced on the Illumina platform with 2×150bp length.

Raw TCR data analysis

Demultiplexing of raw sequencing reads by sample barcodes and building unique molecular identifier (UMI) consensuses were performed with *migec* v. 1.2.9 (*62*) using CheckoutBatch routine with *-ute and AssembleBatch with –force-collision-filter* parameters respectively.

Alignment of TCR V, D, J segments to UMI consesuses were done with *mixcr* v.3.0.13 (*63*) with *analyze amplicon* preset and *--species hs–starting-material rna–5-end no-v-primers–3-end c-primers–adapters adapters-present* parameters.

Resulting mixcr clonotype tables (provided in zenodo (*64*) were converted to VDJtools format with *VDJtools* v.1.2.1 (*65*).

TCR repertoire analysis

Analysis of TCR repertoire clonotype tables was performed using R language (*66*) and RStudio (*67*), with merging and subsetting of data performed using *data.table* (*68*) and *dplyr* packages (*69*). We performed the following in silico clean-up procedure to remove biases introduced by staining, sorting, and sequencing, and identify high-confidence tetramer-positive and tetramer-negative TCR chains:

(i) Elimination of non-productive CDR3 sequences and erroneous sequences that did not begin with cysteine.

(ii) Determination of the frequency ratio for TCR clonotypes (defined by the identical nucleotide sequence of CDR3 and V/J segments) found in both tetramer-positive and tetramer-negative repertoires of the same donor. Clonotypes were assigned to positive or negative sets if they displayed at least a 10-fold increase in either the positive or negative-tetramer repertoire, respectively.

(iii) Clonotypes found exclusively in one fraction were assigned appropriately.

(iv) Convergent recombination, both within and across individuals, can produce identical amino acid sequences for TCRs from distinct nucleotide sequences. To ensure consistent and unambiguous assignment of identical amino acid clonotypes (defined by identical CDR3 amino acid sequences and V/J-gene usage) to either the positive or negative fraction across all donors, we implemented an additional refinement step. Rare clonotypes resulting from convergent recombination that appeared in both positive and negative fractions were assigned to the label based on a minimum 2-fold difference in cumulative UMI counts across all donors. Clonotypes detected in both tetramer fractions but failing to meet this 2-fold UMI count difference threshold were excluded from both datasets.

As a result of this procedure, we obtained a high-confidence list of tetramer-positive and tetramer-negative TCRs for each epitope zenodo (*64*), where each amino acid TCR could only be classified as either positive or negative but never both. The in silico purified data was used for all downstream analysis. Additionally, raw clonotype tables for the tetramer-positive and tetramer-negative fractions, prior to any filtering, are also provided in zenodo (*64*) to allow alternative filtering approaches. To calculate the proportion of the TCR repertoire taken up by tetramer-positive clonotypes, we analyzed bulk repertoires from CD4 and CD8 cells from six individuals. We searched for amino acid CDR3 sequences (with fixed V/J) of both tetramer-positive and tetramer-negative TCRs in matched compartments (CD4 for NKT and DPB1-TFE, CD8 for others) across all samples. We then divided the total number of matched tetramer-positive TCRs by the total number of matched clonotypes (both tetramer-positive and tetramer-negative) within each donor. This calculation gave us the proportion of the repertoire occupied by positive TCRs (See Fig. 4D).

TCRdist calculations

TCRdist is a weighted multi CDR distance metric (*27*, *70*). To implement K-nearest neighbor (KNN) with TCRdist with large datasets generated by the exo-TCR chain introduction, we developed a GPU-enabled version of TCRdist based on (i) precomputation of BLOSUM62 or other amino acid substitution penalty derived distances aligned germline encoded invariant CDR1, CDR2, CDR2.5 regions of the α and β-chains, (ii) combined integer vectorization of V-gene with the variant CDR3 region (center-padded to accommodate TCRdist approximation between sequences of any length), and (iii) chunked computation of all-vs-all results, with compression of intermediate results to only retain k-closest neighbors or a maximum radius neighbors to avoid memory overflow. TCRdist GPU closely approximates brute-force TCRdist particularly at relevant TCRdistances. Software to compute KNN with TCRdist is freely available (*71*).

Sequence similarity analysis and comparison to the public database

Sequence similarity analysis based on Hamming distance was performed in R using the *stringdist* package (*72*). The *igraph* R package (*73*) was used to build TCR similarity networks where each vertex represents a TCR clonotype and is connected with an edge if CDR3 amino acid sequences have no more than 3 mismatches and V-genes are identical.The networks were visualized using Gephi software v.0.10.1 (*74*). VDJdb (*29*) was downloaded in August 2024. For the comparative analysis of naturally occurring and hybrid TCRs generated in this study VDJdb was filtered to only include paired αβTCRs, additionally data from (*28*) not present in VDJdb was added to increase the number of LLW-specific paired chain sequences. To match the features of the exo-TCR chains used to generate the hybrid repertoire data, the LLW-specific TCRβ used in the network were subset to pairs with a TRAV12-2

TCRα, while the GIL-specific TCRα had to pair with a TCRβ using TRBV19 and TRBJ2-7. This resulted in 233 unique LLW-specific TCR β-chains and 542 unique GIL-specific α-chains. Unique amino acid paired natural TCR chains and unique chains from each donor from exo-TCR data were used to construct the network on Fig. 2A and Fig. 2B. Probability of generation (*P*gen) was estimated using OLGA software (*75*) with a default human model.

To calculate the TCRdist to the five closest unique neighbors for each chain of epitope-specific TCRs from VDJdb (Fig. 1D, fig. S6A), we used a GPU-enabled version of TCRdist.

Classification of positive and negative TCRs

For KNN-based classification, in addition to data generated in this study, we used freely available IMMREP23 dataset (*38*), a curated library of 11,310 unique labeled paired chain TCRs drawn from public databases annotated to bind 808 distinct epitopes, with 1,818 and 237 unique receptors sequence assigned to A02-GIL and A02-LLW epitopes, respectively. KNN-based classification was performed with and without data augmentation with labeled binding and nonbinding receptors from exo-TCR experiments. Ten percent of all nonbinders were used. The prediction task was based on predicting binder status of TCR from a fixed-exogenous chain hold out donor (GIL nonbinders n = 13895, binders n = 510; LLW nonbinders n = 20284, binders n = 6968). KNN classification used 10 nearest neighbor TCRs from the reference set and/or exo-TCR data from 3 independent donors (GIL nonbinders n = 27145, binders n = 3196; LLW nonbinders n = 12641, binders n = 16118). Probability was based on weighted sum class labels in the training set, inversely weighted by distance and class label prevalence in the training set. In all experiments, only the unique set of donor TCRs deduplicated at the amino acid level were used. Implementation of weighted KNN-TCRdist is provided in (*71*).

Predictions with default MixTCRpred A02-GIL and A02-LLW models (*30*) were performed on the same test set as used for KNN-TCRdist.

TCRdock models and contact scoring

TCR:pMHC structural models were built as described in (*31*) using the updated pipeline implemented in the TCRdock repository (*76*). TCRdock is a TCR:pMHC docking pipeline that uses a fine-tuned version of the Alphafold2 (*77*) model and template docking information from experimentally determined ternary complexes. For the B08-EIY epitope, we also built structural models with Alphafold3 (*78*), Boltz-2 (*79*), and RoseTTAFold3 (*80*) and saw good agreement with the TCRdock models (fig. S6E). To assess whether there was a correlation between exo-TCR chain contacts with pMHC and exo-TCR chain partner frequency, we built TCRdock models for 200 high-confidence positive αβTCR pairings for each epitope (the partner chain sequences were selected at random from the top quartile by on-target read count). Heavy atom-heavy atom contacts in the models were counted using PyRosetta (*81*) and a distance threshold of 4.5 Å.

Estimating effective sample size for exo-TCR approach

To estimate the effective size for TCR detection using the exo-TCR approach we calculated the post-selection probability of finding TCR in the periphery (P_{post}) using SONIA software (*82*) with a default human model. The effective sample size was then defined as the reciprocal of this probability ($1/P_{\text{post}}$). This value represents the theoretical number of individuals that would need to be screened—given the same blood sample size used in the exo-TCR experiment—to approximate the results for this specific exo-chain (Fig. 5F).

Statistical analysis

Descriptive and comparative statistics were calculated in the manuscript as described in the figure legends with the number of replicates indicated. Statistical analysis was performed in R. The *ggplot2* R package (*83*) was used for visualizations. Boxes on box plots represent the median and 25th to 75th percentiles, and whiskers indicate the minimum to maximum values but no further than 1.5 times the IQR. Errorbars show standard error of the mean.

REFERENCES AND NOTES

1. P. Paoletti, C. Bellone, Q. Zhou, NMDA receptor subunit diversity: Impact on receptor properties, synaptic plasticity and disease. *Nat. Rev. Neurosci.* **14**, 383–400 (2013). doi: 10.1038/nrn3504; pmid: 23686171
2. P. S.-H. Park, K. Palczewski, Diversifying the repertoire of G protein-coupled receptors through oligomerization. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 8793–8794 (2005). doi: 10.1073/pnas.0504016102; pmid: 15956197
3. P. A. Roche, K. Furuta, The ins and outs of MHC class II-mediated antigen processing and presentation. *Nat. Rev. Immunol.* **15**, 203–216 (2015). doi: 10.1038/nri3818; pmid: 25720354
4. T. Mora, A. Walczak, Quantifying lymphocyte receptor diversity. arXiv:1604.00487 [q-bio.PE] (2016). doi: 10.1101/046870
5. S. Muyldermans, Nanobodies: Natural single-domain antibodies. *Annu. Rev. Biochem.* **82**, 775–797 (2013). doi: 10.1146/annurev-biochem-063011-092449; pmid: 23495938
6. X. Wu *et al.*, Focused evolution of HIV-1 neutralizing antibodies revealed by structures and deep sequencing. *Science* **333**, 1593–1602 (2011). doi: 10.1126/science.1207532; pmid: 21835983
7. M. Yuan *et al.*, Structural basis of a shared antibody response to SARS-CoV-2. *Science* **369**, 1119–1123 (2020). doi: 10.1126/science.abd2321; pmid: 32661058
8. I. St-Amour, C. Proulx, R. Lemieux, R. Bazin, Modulations of anti-D affinity following promiscuous binding of the heavy chain with naïve light chains. *Transfusion* **43**, 246–253 (2003). doi: 10.1046/j.1537-2995.2003.00309.x; pmid: 12559021
9. J. Lou *et al.*, Affinity maturation of human botulinum neurotoxin antibodies by light chain shuffling via yeast mating. *Protein Eng. Des. Sel.* **23**, 311–319 (2010). doi: 10.1093/protein/gzq001; pmid: 20156888
10. T. Tulika *et al.*, Engineering of pH-dependent antigen binding properties for toxin-targeting IgG1 antibodies using light-chain shuffling. *Structure* **32**, 1404–1418.e7 (2024). doi: 10.1016/j.str.2024.07.014; pmid: 39146931
11. J. D. Marks *et al.*, By-passing immunization: Building high affinity human antibodies by chain shuffling. *Biotechnology (N. Y.)* **10**, 779–783 (1992). doi: 10.1038/nbt0792-779;pmid: 1368267
12. D. B. Jaffe *et al.*, Functional antibodies exhibit light chain coherence. *Nature* **611**, 352–357 (2022). doi: 10.1038/s41586-022-05371-z; pmid: 36289331
13. M. Milighetti *et al.*, Systematic analysis of CDR contacts and sequence constraints between T cell receptor *αβ* chains. *Immunology* (2024). doi: 10.1101/2024.05.24.595718
14. J. Henderson, Y. Nagano, M. Milighetti, A. Tiffeau-Mayer, Limits on inferring T cell specificity from partial information. *Proc. Natl. Acad. Sci. U.S.A.* **121**, e2408696121 (2024). doi: 10.1073/pnas.2408696121; pmid: 39374400
15. A. Mayer, C. G. Callan Jr., Measures of epitope binding degeneracy from T cell receptor repertoires. *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2213264120 (2023). doi: 10.1073/pnas.2213264120; pmid: 36649423
16. T. M. Snyder *et al.*, Magnitude and dynamics of the T-cell response to SARS-CoV-2 infection at both individual and population levels. *Front. Immunol.* **15**, 1488860 (2025). doi: 10.3389/fimmu.2024.1488860; pmid: 39840037
17. R. O. Emerson *et al.*, Immunosequencing identifies signatures of cytomegalovirus exposure history and HLA-mediated effects on the T cell repertoire. *Nat. Genet.* **49**, 659–665 (2017). doi: 10.1038/ng.3822; pmid: 28369038
18. G. Croce *et al.*, Phage display enables machine learning discovery of cancer antigen-specific TCRs. *Sci. Adv.* **11**, eads5589 (2025). doi: 10.1126/sciadv.ads5589; pmid: 40498839
19. R. Vazquez-Lombardi *et al.*, High-throughput T cell receptor engineering by functional screening identifies candidates with enhanced potency and specificity. *Immunity* **55**, 1953–1966.e10 (2022). doi: 10.1016/j.immuni.2022.09.004; pmid: 36174557
20. N. S. Abdelfattah, T. Kula, S. J. Elledge, T-Switch: A specificity-based engineering platform for developing safe and effective T cell therapeutics. *Immunity* **57**, 2945–2958.e5 (2024). doi: 10.1016/j.immuni.2024.11.009; pmid: 39631392
21. T. M. Schmitt *et al.*, Generation of higher affinity T cell receptors by antigen-driven differentiation of progenitor T cells in vitro. *Nat. Biotechnol.* **35**, 1188–1195 (2017). doi: 10.1038/nbt.4004; pmid: 29106410
22. M. Nakatsugawa *et al.*, Specific roles of each TCR hemichain in generating functional chain-centric TCR. *J. Immunol.* **194**, 3487–3500 (2015). doi: 10.4049/jimmunol.1401717; pmid: 25710913
23. D. Campillo-Davo *et al.*, Efficient and Non-genotoxic RNA-Based Engineering of Human T Cells Using Tumor-Specific T Cell Receptors With Minimal TCR Mispairing. *Front. Immunol.* **9**, 2503 (2018). doi: 10.3389/fimmu.2018.02503; pmid: 30464762
24. D. S. Shcherbinin, V. A. Belousov, M. Shugay, Comprehensive analysis of structural and sequencing data reveals almost unconstrained chain pairing in TCRαβ complex. *PLOS Comput. Biol.* **16**, e1007714 (2020). doi: 10.1371/journal.pcbi.1007714; pmid: 32163410
25. T. Dupic, Q. Marcou, A. M. Walczak, T. Mora, Genesis of the αβ T-cell receptor. *PLOS Comput. Biol.* **15**, e1006874 (2019). doi: 10.1371/journal.pcbi.1006874; pmid: 30830899
26. J. Glanville *et al.*, Identifying specificity groups in the T cell receptor repertoire. *Nature* **547**, 94–98 (2017). doi: 10.1038/nature22976; pmid: 28636589

27. P. Dash *et al.*, Quantifiable predictive features define epitope-specific T cell receptor repertoires. *Nature* **547**, 89–93 (2017). doi: 10.1038/nature22383; pmid: 28636592
28. A. A. Minervina *et al.*, Primary and secondary anti-viral response captured by the dynamics and phenotype of individual T cell clones. *eLife* **9**, e53704 (2020). doi: 10.7554/eLife.53704; pmid: 32081129
29. M. Goncharov *et al.*, VDJdb in the pandemic era: A compendium of T cell receptors specific for SARS-CoV-2. *Nat. Methods* **19**, 1017–1019 (2022). doi: 10.1038/s41592-022-01578-0; pmid: 35970936
30. G. Croce *et al.*, Deep learning predictions of TCR-epitope interactions reveal epitope-specific chains in dual alpha T cells. *Nat. Commun.* **15**, 3211 (2024). doi: 10.1038/s41467-024-47461-8; pmid: 38615042
31. P. Bradley, Structure-based prediction of T cell receptor:peptide-MHC interactions. *eLife* **12**, e82813 (2023). doi: 10.7554/eLife.82813; pmid: 36661395
32. R. Reantragoon *et al.*, Antigen-loaded MR1 tetramers define T cell receptor heterogeneity in mucosal-associated invariant T cells. *J. Exp. Med.* **210**, 2305–2320 (2013). doi: 10.1084/jem.20130958; pmid: 24101382
33. F. Tilloy *et al.*, An invariant T cell receptor α chain defines a novel TAP-independent major histocompatibility complex class Ib-restricted α/β T cell subpopulation in mammals. *J. Exp. Med.* **189**, 1907–1921 (1999). doi: 10.1084/jem.189.12.1907; pmid: 10377186
34. M. Lepore *et al.*, Parallel T-cell cloning and deep sequencing of human MAIT cells reveal stable oligoclonal TCRβ repertoire. *Nat. Commun.* **5**, 3866 (2014). doi: 10.1038/ncomms4866; pmid: 24832684
35. J. Rossjohn, D. G. Pellicci, O. Patel, L. Gapin, D. I. Godfrey, Recognition of CD1d-restricted antigens by natural killer T cells. *Nat. Rev. Immunol.* **12**, 845–857 (2012). doi: 10.1038/nri3328; pmid: 23154222
36. A. M. Kirk *et al.*, DNAJB1-PRKACA fusion neoantigens elicit rare endogenous T cell responses that potentiate cell therapy for fibrolamellar carcinoma. *Cell Rep. Med.* **5**, 101469 (2024). doi: 10.1016/j.xcrm.2024.101469; pmid: 38508137
37. Y. Lu *et al.*, Assessment of computational methods in predicting TCR-epitope binding recognition. *Nat. Methods* **23**, 248–259 (2026). doi: 10.1038/s41592-025-02910-0; pmid: 41315816
38. M. Nielsen *et al.*, Lessons learned from the IMMREP23 TCR-epitope prediction challenge. *Immunoinformatics (Amst.)* **16**, 100045 (2024). doi: 10.1016/j.immuno.2024.100045
39. M. Cetin *et al.*, T-FINDER: A highly sensitive, pan-HLA platform for functional T cell receptor and ligand discovery. *Sci. Adv.* **10**, eadk3060 (2024). doi: 10.1126/sciadv.adk3060; pmid: 38306432
40. A. V. Joglekar *et al.*, T cell antigen discovery via signaling and antigen-presenting bifunctional receptors. *Nat. Methods* **16**, 191–198 (2019). doi: 10.1038/s41592-018-0304-8; pmid: 30700902
41. G. Li *et al.*, T cell antigen discovery via trogocytosis. *Nat. Methods* **16**, 183–190 (2019). doi: 10.1038/s41592-018-0305-7; pmid: 30700903
42. M. N. Lee, M. Meyerson, Antigen identification for HLA class I- and HLA class II-restricted T cell receptors using cytokine-capturing antigen-presenting cells. *Sci. Immunol.* **6**, eabf4001 (2021). doi: 10.1126/sciimmunol.abf4001; pmid: 33483338
43. T. Kula *et al.*, T-Scan: A Genome-wide Method for the Systematic Discovery of T Cell Epitopes. *Cell* **178**, 1016–1028.e13 (2019). doi: 10.1016/j.cell.2019.07.009; pmid: 31398327
44. M. H. Dezfulian *et al.*, TScan-II: A genome-scale platform for the de novo identification of $CD4^+$ T cell epitopes. *Cell* **186**, 5569–5586.e21 (2023). doi: 10.1016/j.cell.2023.10.024; pmid: 38016469
45. M. E. Birnbaum *et al.*, Deconstructing the peptide-MHC specificity of T cell recognition. *Cell* **157**, 1073–1087 (2014). doi: 10.1016/j.cell.2014.03.047; pmid: 24855945
46. G. Sharma, C. M. Rive, R. A. Holt, Rapid selection and identification of functional $CD8^+$ T cell epitopes from large peptide-coding libraries. *Nat. Commun.* **10**, 4553 (2019). doi: 10.1038/s41467-019-12444-7; pmid: 31591401
47. L. Wang, X. Lan, Rapid screening of TCR-pMHC interactions by the YAMTAD system. *Cell Discov.* **8**, 30 (2022). doi: 10.1038/s41421-022-00386-2; pmid: 35379810
48. J. Kisielow, F.-J. Obermair, M. Kopf, Deciphering $CD4^+$ T cell specificity using novel MHC-TCR chimeric receptors. *Nat. Immunol.* **20**, 652–662 (2019). doi: 10.1038/s41590-019-0335-z; pmid: 30858620
49. Z. Moravec *et al.*, Discovery of tumor-reactive T cell receptors by massively parallel library synthesis and screening. *Nat. Biotechnol.* **43**, 214–222 (2025). doi: 10.1038/s41587-024-02210-6; pmid: 38653798
50. A. C. Kohlgruber *et al.*, High-throughput discovery of MHC class I- and II-restricted T cell epitopes using synthetic cellular circuits. *Nat. Biotechnol.* **43**, 623–634 (2025). doi: 10.1038/s41587-024-02248-6; pmid: 38956325
51. M. R. Hart *et al.*, REFLEX, a Novel Immune Profiling Assay, Combining TCR Repertoire and Multiome at Massively Scalable Single-cell Resolution to Catapult Exploration of T-cell Derived Immunity. *Immunology* 2025.10.24.684243 (2025). doi: 10.1101/2025.10.24.684243; pmid: 41394686
52. M. V. Pogorelyy *et al.*, TIRTL-seq: Deep, quantitative and affordable paired TCR repertoire sequencing. *Nat. Methods* **23**, 56–64 (2026). doi: 10.1038/s41592-025-02907-9; pmid: 41286199
53. P. Zareie *et al.*, Canonical T cell receptor docking on peptide-MHC is essential for T cell signaling. *Science* **372**, eabe9124 (2021). doi: 10.1126/science.abe9124; pmid: 34083463
54. A. Straub *et al.*, Recruitment of epitope-specific T cell clones with a low-avidity threshold supports efficacy against mutational escape upon re-infection. *Immunity* **56**, 1269–1284.e6 (2023). doi: 10.1016/j.immuni.2023.04.010; pmid: 37164014
55. M. K. Jenkins, J. J. Moon, The role of naive T cell precursor frequency and recruitment in dictating immune response magnitude. *J. Immunol.* **188**, 4135–4140 (2012). doi: 10.4049/jimmunol.1102661; pmid: 22517866
56. A. Murugan, T. Mora, A. M. Walczak, C. G. Callan Jr., Statistical inference of the generation probability of T-cell receptors from sequence repertoires. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 16161–16166 (2012). doi: 10.1073/pnas.1212755109; pmid: 22988065
57. J. M. Heather *et al.*, Stitchr: Stitching coding TCR nucleotide sequences from V/J/CDR3 information. *Nucleic Acids Res.* **50**, e68–e68 (2022). doi: 10.1093/nar/gkac190; pmid: 35325179
58. G. Dolton *et al.*, More tricks with tetramers: A practical guide to staining T cells with peptide-MHC multimers. *Immunology* **146**, 11–22 (2015). doi: 10.1111/imm.12499; pmid: 26076649
59. P. A. Mudd *et al.*, SARS-CoV-2 mRNA vaccination elicits a robust and persistent T follicular helper cell response in humans. *Cell* **185**, 603–613.e15 (2022). doi: 10.1016/j.cell.2021.12.026; pmid: 35026152
60. J. Robinson *et al.*, IMGT/HLA and IMGT/MHC: Sequence databases for the study of the major histocompatibility complex. *Nucleic Acids Res.* **31**, 311–314 (2003). doi: 10.1093/nar/gkg070; pmid: 12520010
61. E. S. Egorov *et al.*, Quantitative profiling of immune repertoires for minor lymphocyte counts using unique molecular identifiers. *J. Immunol.* **194**, 6155–6163 (2015). doi: 10.4049/jimmunol.1500215; pmid: 25957172
62. M. Shugay *et al.*, Towards error-free profiling of immune repertoires. *Nat. Methods* **11**, 653–655 (2014). doi: 10.1038/nmeth.2960; pmid: 24793455
63. D. A. Bolotin *et al.*, MiXCR: Software for comprehensive adaptive immunity profiling. *Nat. Methods* **12**, 380–381 (2015). doi: 10.1038/nmeth.3364; pmid: 25924071
64. A. Minervina, M. Pogorelyy, P. Thomas, Uneven TCR chain pairing constraints govern epitope recognition, Zenodo (2025); https://doi.org/10.5281/ZENODO.14885005.
65. M. Shugay *et al.*, VDJtools: Unifying Post-analysis of T Cell Receptor Repertoires. *PLOS Comput. Biol.* **11**, e1004503 (2015). doi: 10.1371/journal.pcbi.1004503; pmid: 26606115
66. 52R Core Team, R: A language and environment for statistical computing (R Foundation for Statistical Computing, 2021); www.R-project.org/.
67. Posit team, RStudio: Integrated Development Environment for R (Posit Software, 2025); https://www.posit.co/.
68. T. Barrett *et al.*, Data.Table: Extension of "data.Frame" (2025); https://r-datatable.com.
69. H. Wickham, R. François, L. Henry, K. Müller, D. Vaughan, Dplyr: A Grammar of Data Manipulation (2025); https://dplyr.tidyverse.org.
70. K. Mayer-Blackwell *et al.*, TCR meta-clonotypes for biomarker discovery with *tcrdist3* enabled identification of public, HLA-restricted clusters of SARS-CoV-2 TCRs. *eLife* **10**, e68605 (2021). doi: 10.7554/eLife.68605; pmid: 34845983
71. K. Mayer-Blackwell, kmayerb/tcrdistgpu: Manuscript version 1.0, version v1.0-manuscript, Zenodo (2026); https://doi.org/10.5281/ZENODO.20944403.
72. M. P. J. van der Loo, The stringdist package for approximate string matching. *R J.* **6**, 111–122 (2014). doi: 10.32614/RJ-2014-011
73. G. Csárdi, T. Nepusz, The igraph software package for complex network research. *InterJournal. Complex Syst.* **1695**, 1–9 (2006).
74. M. Bastian, S. Heymann, M. Jacomy, Gephi: An Open Source Software for Exploring and Manipulating Networks. *Proc. Int. AAAI Conf. Web Soc. Media* **3**, 361–362 (2009). doi: 10.1609/icwsm.v3i1.13937
75. Z. Sethna, Y. Elhanati, C. G. Callan Jr., A. M. Walczak, T. Mora, OLGA: Fast computation of generation probabilities of B- and T-cell receptor amino acid sequences and motifs. *Bioinformatics* **35**, 2974–2981 (2019). doi: 10.1093/bioinformatics/btz035; pmid: 30657870
76. P. Bradley, E. Ball, phbradley/TCRdock: v2.0.1–zenodo DOI, version v2.0.1, Zenodo (2026); https://doi.org/10.5281/ZENODO.20938245.
77. J. Jumper *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021). doi: 10.1038/s41586-021-03819-2; pmid: 34265844
78. J. Abramson *et al.*, Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024). doi: 10.1038/s41586-024-07487-w; pmid: 38718835
79. S. Passaro *et al.*, Boltz-2: Towards Accurate and Efficient Binding Affinity Prediction. bioRxiv 2025.06.14.659707 [Preprint] (2025); doi: 10.1101/2025.06.14.659707v1.
80. N. Corley *et al.*, Accelerating Biomolecular Modeling with AtomWorks and RF3. bioRxiv 2025.08.14.670328 [Preprint] (2025); doi: 10.1101/2025.08.14.670328v1.
81. S. Chaudhury, S. Lyskov, J. J. Gray, PyRosetta: A script-based interface for implementing molecular modeling algorithms using Rosetta. *Bioinformatics* **26**, 689–691 (2010). doi: 10.1093/bioinformatics/btq007; pmid: 20061306
82. Z. Sethna *et al.*, Population variability in the generation and selection of T-cell repertoires. *PLOS Comput. Biol.* **16**, e1008394 (2020). doi: 10.1371/journal.pcbi.1008394; pmid: 33296360
83. H. Wickham, *Ggplot2: Elegant Graphics for Data Analysis* (Springer, 2016); https://ggplot2.tidyverse.org.

ACKNOWLEDGMENTS

We thank E. K. Sliger and Z. Scott for apheresis bank sample curation and HLA typing, A. Kirk for creation of artificial antigen presenting cell lines, and G. Lennon from St. Jude Immunology flow core for his help with FACS. **Funding:** This work was supported by grants AI136514, AI144616, AI165077, AI187063, and American Lebanese Syrian Associated Charities (ALSAC) at St. Jude. Scientific Computing Infrastructure at Fred Hutchinson Cancer Research Center was funded by ORIP grant 1S10OD028685-01. P.B. was supported in part by R35GM141457. A.M.W. and T.M. were supported by Agence Nationale de la Recherche grant ANR-19-CE45-0018, H2020 European Research Council Proof of Concept Grant "RESPOND" (101185627), and Foundation pour la Recherche Medical (EQU202503019997). A.A.M. was supported in part by Paul Barrett endowed fellowship. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health, and the funders had no role in experimental design or interpretation. **Author contributions:** Conceptualization: A.A.M., M.V.P., and P.G.T. Methodology: A.A.M., M.V.P., R.T., K.M.B., T.M., A.M.W., and P.B. Software: M.V.P., K.M.B., A.F.G., and P.B. Validation: A.A.M., M.V.P., K.M.B., and P.B. Formal analysis: A.A.M., M.V.P., S.A.S., K.M.B., and P.B. Investigation: A.A.M. Resources: K.M.B., P.B., and P.G.T. Data curation: A.A.M., M.V.P., K.M.B., and P.B. Writing - original draft: A.A.M. Writing, review & editing: all authors; Visualization: A.A.M. and P.B. Supervision: A.F.G. and P.G.T.; Project administration: A.A.M. and P.G.T.; Funding acquisition: A.M.W., T.M., P.B., and P.G.T. **Competing interests:** P.G.T. is on the Scientific Advisory Board of Immunoscape, INTcRON, and Shennon Bio, has received research support and personal fees from Elevate Bio, and consulted for 10X Genomics, Illumina, Pfizer, Cytoagents, Sanofi, Merck, and JNJ. P.G.T., A.A.M., and M.V.P. have patents related to TCR amplification, cloning, and/or applications thereof. The other authors declare that they have no competing interests. **Data, code, and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials. Raw clonotype tables and data frames of filtered epitope-positive and epitope-negative repertoires are available at Zenodo (*64*). The code used for TCRdist calculations is available at (*71*). Source data for figures are provided in data S1. Plasmids and/or transgenic Jurkat cell lines generated in this study are available from P.G.T. upon request under a Uniform Biological Material Transfer Agreement (UBMTA) with Fred Hutchinson Cancer Center.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adx3863
Figs. S1 to S8; Tables S1 to S5; MDAR Reproducibility Checklist; Data S1
Submitted 11 March 2025; resubmitted 6 March 2026; accepted 30 June 2026
10.1126/science.adx3863

Uneven TCR chain pairing constraints govern epitope recognition

Anastasia A. Minervina, Mikhail V. Pogorelyy, Koshlan Mayer-Blackwell, Ricky Tirtakusuma, Stefan A. Schattgen, Andrew Fiore-Gartland, Aleksandra M. Walczak, Thierry Mora, Philip Bradley, and Paul G. Thomas

Science **393** (6815), eadx3863. DOI: 10.1126/science.adx3863

Editor's summary

Sequencing T cell receptors (TCRs) has the potential to reveal the infection history and disease status of an individual. However, it is difficult to predict from the sequence of a single TCR chain how it pairs with another chain to recognize specific antigens. Minervina *et al*. used primary human T cells to generate "exogenous" TCRs that contained one TCR chain endogenous to the T cell, paired with a single # or # TCR chain of known epitope recognition that had been transfected into the cell. This approach revealed principles contributing to chain pairing and epitope recognition. In addition, these data provide a resource of a diverse array of TCRs of defined specificity and a set of very similar TCRs with different binding properties that can be used to train machine learning models for TCR antigen prediction. —Sarah H. Ross