Journal for ImmunoTherapy of Cancer

Mapping the TCR landscape: computational tools empowering translational immunology and therapy design

Pâmella Borges ,[1] Martiela Vaz de Freitas,[1] Jinkyung Yoo,[2] Finn Beruldsen,[1] Jaila Lewis,[1] Francisca Joseli Freitas de Sousa,[3] Sae Hee Choi,[1] Duy Bao Nguyen ,[4] Geancarlo Zanatta,[5,6] Jeong Hoon Jang,[7] Eduardo Donadi,[8] Houda Alachkar ,[9] Steven P Wolf ,[10,11] Maurício Menegatti Rigo,[12] Hyeongseon Jeon,[2,13] Dinler Amaral Antunes [1]

To cite: Borges P, Vaz de Freitas M, Yoo J, *et al*. Mapping the TCR landscape: computational tools empowering translational immunology and therapy design. *Journal for ImmunoTherapy of Cancer* 2026;**14**:e014184. doi:10.1136/jitc-2025-014184

► Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/jitc-2025-014184).

Accepted 13 May 2026

For numbered affiliations see end of article.

Correspondence to
Dr Dinler Amaral Antunes; dinler@uh.edu

ABSTRACT

T cell receptors (TCRs) are central to adaptive immunity, yet their vast sequence and structural diversity present a significant challenge to fully understand immune responses. The application of high-throughput sequencing technologies, including bulk and single-cell approaches, generates vast datasets of TCR repertoire information, requiring advanced computational tools for meaningful analysis. Here, we provide a comprehensive overview of the state-of-the-art in silico tools developed to enable diverse TCR repertoire analyses. We categorize over 40 computational tools into six primary analytical stages creating a workflow for TCR analysis in the context of cancer immunotherapy: (1) data acquisition, including differences between TCR sequencing technologies and databases; (2) TCR reconstruction and inference, which focuses on accurately extracting from raw sequencing data the V(D)J gene usage, including complementarity-determining region sequences, and the α/β pairing; (3) TCR clustering, which groups receptors based on similarity, helping characterize repertoire shifts, therapy responses and identify cancer-associated TCR clones; (4) structural modeling of TCRs and TCR–peptide-major histocompatibility complex (MHC), which is used to predict the three-dimensional structures of TCRs with or without their targets; (5) TCR specificity prediction, which predicts whether a given TCR can bind to a given peptide-MHC complex; and finally (6) functional and clinical integration, addressing the breakthroughs and bottlenecks for wider clinical application of these methods. For each category, we discuss the underlying methodologies, representative tools and their key applications, details about usability and accessibility, and comments on their strengths and limitations. With this overview, we offer a critical perspective on the current state of the field, providing an overall framework and guidance for new users and developers of these technologies. We also highlight open challenges and key future directions, particularly regarding the integration of multi-omics data and next-generation artificial intelligence approaches to unlock the full potential of TCR repertoire analysis for clinical immunotherapy applications.

INTRODUCTION

Cancer immunotherapy harnesses the body's own immune system to eliminate malignant cells by enhancing or redirecting the immune response, with a particular focus on the cytotoxic activity of T cells. Central to their cellular response is the T cell receptor (TCR), which initiates T cell activation through recognition of antigenic peptides presented by major histocompatibility complex (MHC) molecules on the surface of target cells. Understanding the selective recognition of peptide-MHC (pMHC) complexes (ie, TCR specificity) and how TCR–pMHC engagement governs downstream signaling is crucial for identifying therapeutic TCRs and their targets, predicting and monitoring treatment efficacy, and designing next-generation cancer immunotherapies.[1–3]

Due to their significance for adaptive immunity, TCRs are one of the most extensively studied receptors in modern biology.[4] These heterodimeric protein complexes are typically composed of α/β chains (≈95% of peripheral T cells) or, less commonly, γ/δ chains (≈5%). Each chain contains variable (V) and constant (C) domains and is generated through somatic V(D)J recombination during T cell development in the thymus.[5] The β chain undergoes ordered D–J then V-DJ recombination, whereas the α chain forms through successive V–J rearrangements without D segments (figure 1). While α chain recombination is not genetically allelically excluded, positive selection enforces functional allelic exclusion by stabilizing only the productive α chain, ensuring that most mature T cells express a single, clonally unique αβ receptor. This developmental

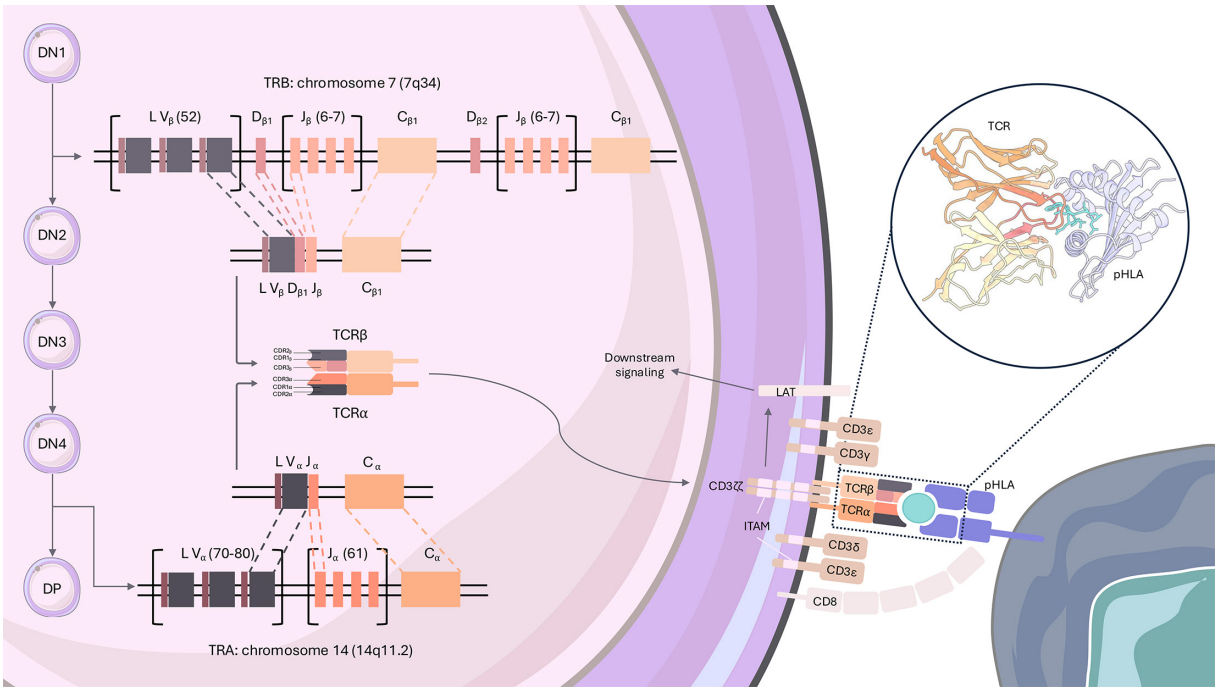

Figure 1 Human TCR biology: somatic recombination, assembly of the TCR during thymic development, and antigen recognition. Each unique TCR is generated through tightly regulated V(D)J recombination of variable (V), diversity (D), joining (J), and constant (C) gene segments during intrathymic T-cell maturation. Between the DN1 and DN2 stages, thymocytes initiate rearrangement of the TRD, TRG, and TRB loci, whereas TRA rearrangement starts in the DN4 thymocytes, resulting in DP development following successful selection and pre-TCR signaling. This developmental timing reflects the human genomic architecture of the loci: TRB (7q34) and TRG (7p14) reside on chromosome 7, whereas TRA (14q11.2) spans the TRD locus on chromosome 14, a configuration that contributes to early lineage decisions between αβ and γδ T cells. The TRA locus contains 70–80 V segments (each preceded by a leader exon), a cluster of 61 J segments, and a single C gene. The TRB locus includes 52 functional V segments and two D–J clusters, each followed by a single C gene. The β chain is assembled through sequential D–J and V–DJ recombination mediated by RAG1/2 and resolved by NHEJ. Productive β chain rearrangements induce allelic exclusion and allow thymocytes to pass the DN3 checkpoint, whereas non-productive rearrangements lead to apoptosis. In contrast, the α chain is generated by direct V–J recombination without D segments and can undergo multiple successive rearrangements across its extensive V and J arrays. Although α chain rearrangement is not strictly allelically excluded at the genetic level, positive selection enforces phenotypic allelic exclusion by stabilizing only the chain that supports productive TCR signaling. TCR antigen specificity is dictated by six CDRs, three per chain. Germline-encoded CDR1 and CDR2 loops primarily contact and stabilize binding to the MHC molecule, whereas the hypervariable CDR3 loop, formed at the V-J (α) or V-D-J (β) junction, provides the principal specificity for the antigenic peptide. Extensive junctional diversification through nucleotide trimming and addition generates the remarkable diversity needed for T-cell recognition across tissues. On engagement of pMHC, conformational changes and mechanical forces transmitted through the TCR promote phosphorylation of ITAMs within associated CD3 complexes (CD3εγ, CD3εδ, CD3ζζ). Lck phosphorylates these ITAMs, creating docking sites for ZAP-70, which is subsequently activated and phosphorylates LAT (Linker for Activation of T cells) (not shown). LAT then propagates downstream signaling pathways that control T-cell activation, proliferation, differentiation, and effector functions. CDRs, complementarity-determining regions; DN, double negative; DP, double positive; ITAMs, immunoreceptor tyrosine-based activation motifs; MHC, major histocompatibility complex; NHEJ, non-homologous end joining; pMHC, peptide-MHC; TCR, T cell receptor; TRA, T-cell receptor α; TRB, T-cell receptor β; TRD, T-cell receptor β; TRG, T-cell receptor γ.

architecture forms the basis for a TCR repertoire of extraordinary breadth (estimated to be $\geq 10^{15}$ possible different TCRs[6]), capable of recognizing an estimated 10^7–10^8 distinct antigens,[7] while only ≈10–20% of T cells express dual-specificity.[8–10]

TCR specificity is determined by six complementarity-determining regions (CDRs), three contributed by each chain. CDR1 and CDR2 loops, encoded within the germline V gene segment, predominantly contact the MHC helices, anchoring the overall binding orientation. In contrast, the hypervariable CDR3 loop, formed by V(D)J junctional diversity and extensive nucleotide trimming/addition, dictates peptide specificity and confers the majority of TCR repertoire diversity. The structural interplay of these loops enables TCRs to maintain broad antigen recognition while preserving sensitivity to subtle biochemical differences in peptide presentation, a balance essential for immune surveillance and self-tolerance.[11 12]

On pMHC engagement, TCR signaling is transduced through associated CD3 complexes (CD3εγ, CD3εδ, CD3ζζ), which contain immunoreceptor tyrosine-based activation motifs (ITAMs).[13] Conformational changes and mechanical forces applied during TCR–pMHC binding

facilitate ITAM phosphorylation, initiating a downstream signaling cascade. These early signaling events govern T cell proliferation, cytokine production, differentiation, and cytotoxic activity, processes that are frequently dysregulated in cancer due to immune suppression, antigen loss, or MHC downregulation, thereby limiting effective antitumor immunity.[14]

The advent of high-throughput TCR sequencing has transformed our ability to profile TCR repertoires at unprecedented scale and depth. However, these datasets often contain millions of clonotypes that recognize non-cancer-related antigens, introducing background noise.[15] Effective TCR analysis depends on sophisticated computational pipelines designed to filter background sequences, identify expanded or antigen-driven clones, reconstruct chain pairing, and delineate the functional fraction of the repertoire.

Beyond descriptive profiling, advanced computational approaches for structure-based modeling, TCR–pMHC interaction prediction, and functional interpretation of TCR repertoires have become indispensable for rational T cell engineering. Tumors often evade immune surveillance by creating an immunosuppressive environment that reduces T cell responses. In this context, engineered TCR-T cell therapies aim to restore or enhance tumor-specific immune responses by introducing defined tumor-reactive TCRs selected and optimized through computational and experimental approaches.[16 17] However, the rational design of TCR-T cell therapies requires precise control over TCR binding properties to maximize tumor specificity while minimizing cross-reactivity and toxicity, and remains an area of active investigation. The precision of engineered TCR-T cell therapy design, therefore, relies heavily on computational models capable of predicting affinity, specificity, cross-reactivity, and off-target risks, establishing bioinformatics as a core component of next-generation therapeutic design in oncology, infectious disease, and autoimmunity.[18–21]

TCR repertoire analysis also functions as a dynamic biomarker for monitoring treatment response. For instance, local tumor ablation methods, including radiofrequency, microwave, and cryoablation, induce immunogenic cell death, releasing damage-associated molecular patterns that prime innate immunity and promote epitope spreading.[22] This leads to the expansion of a more diverse T cell repertoire both intratumorally and systemically, a phenomenon strongly associated with improved outcomes, particularly when ablation is combined with immune checkpoint inhibitor treatment.[22] Computational tools are essential for quantifying these immunological dynamics and linking them to clinical benefit. The convergence of artificial intelligence (AI) with TCR analytics further enhances clinical translation. Machine learning augments data filtering, structural prediction, antigen discovery, and therapy design. These advances establish TCR-focused tools as the indispensable foundation of precision immuno-oncology, providing the means to systematically identify, engineer, and track tumor-specific T cell responses for curative cancer therapies, while also extending these critical analytical principles to the study of systemic immune dysregulation.

Here, we present a comprehensive review of bioinformatics tools for TCR analysis, structured into a six-step workflow encompassing data acquisition, TCR reconstruction and inference, TCR clustering, structural modeling of TCR and TCR–pMHC complexes, specificity prediction, and functional or clinical integration. For each step, we summarize core methodologies, use cases, strengths, limitations, and practical considerations, providing a united overview for leveraging TCR analytics in cancer immunotherapy.

TCR ANALYSIS: FROM SEQUENCING TO TCR-BASED THERAPY

TCR data acquisition

TCR repertoire studies aim to catalog the sequence diversity and clonal architecture of T cell populations and to connect those repertoires to function, phenotype, and disease. In cancer immunotherapy, repertoire profiling has been instrumental for assessing patient-specific immune landscapes, monitoring treatment-induced shifts in T cell populations, identifying therapeutic targets, and gaining insight into tumor heterogeneity and neoantigen recognition.[23–25] High-throughput sequencing has further advanced the field by enabling deep sampling of TCR repertoires from diverse cell sources, producing datasets spanning millions to hundreds of millions of receptor sequences across various cohorts and conditions, and transforming our ability to study T cell biology.[26 27] These data now serve both descriptive and mechanistic purposes, allowing the quantification of diversity, clonality, and the distribution of public vs private clones, while also supporting analyses of antigen-driven clonal expansions, phenotypic associations, and, when possible, antigen specificity.[28] Importantly, sampling across tissues and time points enables precise measurements of immune dynamics in both health and disease. In the context of cancer immunotherapy, TCR repertoire analysis has emerged as a valuable biomarker for predicting patient response to immune checkpoint inhibitors and adoptive T cell therapies.[27 29]

At a high level, TCR sequencing data can be broadly classified by scale into bulk and single-cell approaches. Bulk TCR sequencing assays profile TCR transcripts (RNA-based) or genomic V(D)J rearrangements (DNA-based) from a population of cells *en masse*, producing high depth and excellent sensitivity for quantifying clonotype frequencies and repertoire diversity. This makes

bulk approaches especially powerful for detecting low-frequency or rare T cell clonotypes, such as those associated with early tumor emergence or minimal residual disease, but it lacks information on α/β pairing and phenotypic or transcriptional context, which limits downstream interpretation of TCR specificity and function.[30]

On the other hand, single-cell approaches, like single-cell RNA sequencing (scRNA-seq) and single-cell T-cell receptor sequencing (scTCR-seq), allow identification of TCR clonotypes and paired α/β chain recovery with gene expression profiles when combined with scRNA-seq, enabling the detection of receptor sequences, clonotypes, and cell states within the same cell. This enables lineage/phenotype mapping and more directly links the TCR sequence to function, although at the expense of throughput, cell-capture efficiency, and cost per cell.[28 31 32] Another distinction among sc-seq is the transcript end targeted for sequencing. 5′ end is generally preferred for TCR analysis because the V(D)J region is located at this end of the transcript, enabling comprehensive recovery of TCR sequences together with gene expression profiles when combined with 5' scRNA-seq. In contrast, 3′ end sequencing primarily captures the constant region via poly (A)-based barcoding, preserving transcriptomic information but losing essential V(D)J diversity. Taken together, there is no single "best" sequencing or analysis pipeline, mainly because the choice depends on the purpose of the study and decisions regarding platform, multiplexing, unique molecular identifier (UMI) use, read length, sequencing depth, and downstream computational steps that can affect inferred repertoire metrics and influence sensitivity, accuracy, and coverage.[32–34]

Beyond technical limitations, TCR repertoire analysis also faces biological and interpretive challenges. A central obstacle is the lack of ground-truth data. The number of experimentally validated TCR-antigen pairs remains small relative to the enormous sequence diversity, constraining efforts to annotate specificity directly from sequencing data, a key resource for training AI models. The biological complexity of TCR recognition, marked by cross-reactivity, context-dependent immunodominance, and dynamic clonal selection, further complicates interpretation.[27] Even state-of-the-art algorithms struggle to predict antigen specificity with high confidence, given the sparsity of training data and the structural complexity of TCR–pMHC interactions.[33 34] Recent initiatives aim to solve these gaps by generating larger and more diverse datasets using multiple sequencing technologies and sample sources, and by assembling them into comprehensive databases that incorporate sequence-level and structural information for both TCRs and TCR–pMHC complexes (table 1, online supplemental table 1).[35 36] While these resources will allow the development of more precise tools, fully addressing these challenges will require integrating repertoire sequencing with complementary approaches such as immunopeptidomics, functional assays, and structure-guided modeling, alongside the development of interoperable databases and shared benchmarking standards.[27 34] Such efforts are crucial for advancing the field beyond descriptive cataloging toward a predictive and mechanistic understanding of TCR function.

TCR reconstruction and inference

High-quality sequence reconstruction and accurate inference of α/β chain pairing are essential components for reliable TCR repertoire analysis, as they form the foundation for identifying the receptor sequences present within patient samples. The resulting repertoire profile offers an interpretable view of the immune landscape, enabling correlations between repertoire dynamics and factors such as age, sex, diet, treatment response, biomarkers, and antigen exposure.[23 37–39] In cancer research, these insights are particularly valuable as they support investigations into mechanisms of immunotherapy responsiveness, assist in monitoring therapeutic efficacy, and help guide the development and refinement of targeted immunotherapies.[1–3]

Reconstruction focuses on accurately recovering TCR sequences, from raw sequencing data, particularly by extracting and annotating V(D)J segments. Inference builds on this information to reconstruct or assign paired TCR α and β chain sequences, and link them to individual clonotypes (ie, clonotype calling) and associated transcriptomic features. Although these steps can occur within the same pipeline, the sequencing technology used strongly influences how inference and clonotype calling are performed, ultimately shaping both the quality and depth of the resulting data.[40 41] Primarily, the tools differ in their compatibility with bulk and single-cell sequencing, targeted TCR, and whole-transcriptome sequencing and 5′ or 3′ protocols. These factors collectively determine the completeness of the recovered information and, critically, the ability to resolve chain pairing, an essential requirement for translating TCR discovery into clinical applications.

The primary output of reconstruction tools consists of annotated V(D)J gene assignments and CDR3 sequences, which form the basis for clonotype calling and broader repertoire analyses. These data provide critical insights into a patient's immune status and can highlight opportunities for improved or personalized therapeutic strategies.[42 43] Reconstruction tools use a range of alignment and assembly methods, such as guided assembly, de novo assembly, and realignment (table 2, (online supplemental table 2). In turn, these methodological differences can affect accuracy, computational runtime, and performance.

Reference-guided methods generally achieve higher accuracy by mapping sequencing reads to curated germline references, enabling the identification of read overlaps, contig assembly, and correction of sequencing errors. A major limitation of this approach, however, lies in the reference databases themselves. Most tools rely on the International ImMunoGeneTics (IMGT) information system for V(D)J gene assignment, which has historically

Table 1 Overview of publicly available databases and datasets that provide TCR-related information

Abbreviation	Name	Webpage	Paper	Information	Data volume	Source
Databases						
General						
IMGT	The international ImMunoGeneTics	https://www.imgt.org/	[1]*	It provides databases and tools based on immunogenetics information		
				IMGT/LIGM-DB	Nucleotide sequences of IG and TR from 368 species (251,616 entries)	
				IPD-IMGT/HLA-DB	Sequences of the human MH (HLA)	
				IMGT/PRIMER-DB	Oligonucleotides (primers) of IG and TR from 11 species (1,864 entries)	
				IMGT/GENE-DB (doc) LIGM, Montpellier, France	International nomenclature for IG and TR genes from 41 species (12,185 genes, 17,300 alleles)	PDB, INN, Kabat
				IMGT/3Dstructure-DB and IMGT/2Dstructure-DB (doc) LIGM, Montpellier, France	3D structures (IMGT Colliers de Perles) of IG antibodies, TR, MH and RPI (9,141 entries)	
IEDB	The Immune Epitope Database	https://www.iedb.org/	[2]*	A resource that houses experimental data related to adaptive immune epitopes. TCRs are limited to T cell assay types	T cell assays: 562,821	Manually curates' data from the literature.
CEDAR	The Cancer Epitope Database and Analysis Resource	https://cedar.iedb.org/	[3]*	Catalog epitope and receptor data in the context of cancer. It has a Receptor table with TCR sequences and generating 3D structures.	T cell assays: 151,096	Curated from the literature.
PDB	RCSB Protein Data Bank	https://www.rcsb.org/	[4]*	Archive of 3D structure data for large biological molecules (proteins, DNA, and RNA)	958 TCR from experimental method, being 873 from X-ray diffraction, 67 from electron microscopy and 18 from solution NMR	Data depositors submit the results of their structural studies of biological macromolecules.
Specific to TCR						

Continued

Table 1 Continued

Abbreviation	Name	Webpage	Paper	Information	Data volume	Source
VDJdb		https://vdjdb.cdr3.net/	[5]*	Curated database of TCR sequences with known antigen specificities.	HomoSapiens: TRA: 43,964, Paired: 36,899 TRB:87610, Paired: 66,829	
TCRdb		https://guolab.wchscu.cn/TCRdb/#/	[6]*	It contains processed sequences of TCRs CDR3 beta chain of human with different phenotypes.	277,439,349 TCR sequences from a total of 131 studies, 8,265 samples from 41 tissues, 54 cell types, and 113 clinical conditions	From public TCR-Seq datasets from NCBI SRA, and other TCR sequences databases, including iReceptor, VDJServer, and immuneACCESS with known sample attribute of disease/cell type/clinical condition.
TCRdb2.0		https://guolab.wchscu.cn/TCRdb2/#/	[7]*	It contains processed sequences of TCRs CDR3 beta chain of human with different phenotypes.	Contain 691,744,135 sequences from 269 projects, 19,701 samples, 147 conditions, 46 sources, and 16 cell types	From public TCR-Seq datasets from NCBI SRA, and other TCR sequences databases, including iReceptor, VDJServer, and immuneACCESS with known sample attribute of disease/cell type/clinical condition.
TCR3d	TCR structural repertoire database	https://tcr3d.ibbr.umd.edu/	[8]*	Easy-to-use interface to view all experimentally determined structures of TCRs and their complexes.	241 Class I complex structures and 90 Class II complex structures	A curated collection of experimentally determined X-ray TCR structures from the PDB, analyzed for structure, sequence, and antigen recognition, as well as TCR germline gene sequences from IMGT and TCR sequencing data from various studies.
STCRDab	The Structural TCR Database	https://opig.stats.ox.ac.uk/webapps/stcrdab-stcrpred	[9]*	A database of TCR structures that automatically collects and curates' information from the PDB on a weekly basis, and annotates the MHC molecule that is bound by a TCR.		PDB

Continued

Table 1 Continued

Abbreviation	Name	Webpage	Paper	Information	Data volume	Source
McPAS-TCR	A manually curated catalog of pathology associated TCR sequences	https://friedmanlab.weizmann.ac.il/McPAS-TCR/	[10]*	Sequences of TCRs associated with various pathologic conditions (including pathogen infections, cancer and autoimmunity) and their respective antigens in humans and in mice.	~5,100 entries, curated from 118 publications of TCR sequences; ~75% of the sequences are from human data, and the rest come from mouse data.	Manually curated database of TCR sequences published in literature.
TRAIT	TCR antigen interactions	https://pgx.zju.edu.cn/traitdb/	[11]*†	TCR-antigen pairs, integrating sequences, structures, and affinities.	3,393,826 TCR-antigen pairs, 1,184 antigen records, 51,601 interactive pairs, 3,342,225 non-interactive pairs, 223 published structures, 1,269 mutation records, 336 publications, and 34 clinical trials	Experimentally verified.
UcTCRs	unconventional TCRs	http://uctcrdb.cn/#/	[12]*	Focus on unconventional TCRs involved in unique immune responses	669,900 unconventional TCRs	Systematically collected from published literature (PubMed) corresponding studies in human, mouse and cattle.
TCRStructDB		https://ai4s.tencent.com/tcr	[13]*†	Online platform offering extensive data, including structures of known TCR–pMHC complexes, paired unliganded TCRs and unbound pMHCs, and a suite of search and analytical tools.	4,517,910 paired TCR sequences, including 2,226,260 unique TCR fragment variable region sequences, 817,584 pMHC complexes, comprising 531,578 entries for MHC I and 286,006 for MHC II, and 45,000 TCR–pMHC complexes	Data from existing databases such as STCRDab, TCR3d, IEDB, HuARdb, OTS, with manually curated metadata. When experimental structures were not available, it used tFold-TCR predictions.
Datasets						
10x Genomics	https://www.10xgenomics.com/datasets			Datasets created by 10x Genomics using their technologies		
Parse	https://www.parsebiosciences.com/resources/datasets/			Datasets created by Parse using their technologies		

*The completed table, including all resources and references, is available in the online supplemental material.
†Preprints from bioRxiv.
CDR, complementarity-determining region; 3D, three-dimensional; INN, international nonproprietary name; MH, major histocompatibility; MHC, major histocompatibility complex; NCBI, National Center for Biotechnology Information; PDB, Protein Data Bank; pMHC, peptide-MHC; RPI, related proteins of the immune system; SRA, Sequence Read Archive; TCR, T cell receptor; TR, T-cell receptor; TRA, T-cell receptor alpha; TRB, T-cell receptor beta.

Table 2 Non-extensive review of tools for TCR analyses

Tool	Input data	Output	Primary methodology	Strengths	Limitations	Implementation details	Paper
TCR reconstruction and inference							
MiXCR	TCR repertoire sequencing data	Annotated V(D)J gene alignments, CDR3 sequences, clonotypes, repertoire summary, and scα/β paired clonotypes.	Multi-stage V(D)J alignment and assembly pipeline.	Available for several methodologies, the pipeline includes paired-end read merging, quality trimming, and barcode/UMI extraction.	Parameter tuning needed for non-standard libraries; dependent on reference accuracy; not designed for de novo assembly; license needed.	https://mixcr.com/ https://github.com/milaboratory/mixcr	[14]*
TRUST4	Reads or prealigned BAMs	De novo-assembled TCR contigs, annotated V(D)J gene, CDR3 sequences, clonotypes, and optional single-cell barcode-linked chain reconstructions.	Filters out candidate receptor reads via k-mer matching, assembles them de novo into contigs, and realigns contigs to reference genes.	Reconstruction of full-length receptor sequences, realigns the contigs to IMGT reference gene sequences to identify the corresponding gene and CDR3 details.	Slower on very large datasets; assembly may fail for extremely low coverage, moderate learning curve.	https://github.com/liulab-dfci/TRUST4	[15]*
TraCeR	scRNA-seq data	Assembled and paired TCR α and β chain sequences per cell; clonotype; V/J gene annotations and CDR3 sequences.	Read extraction followed by de novo per-cell assembly, V/J annotation and α/β pairing, and clonotype inference.	Reconstructs paired α/β chains per cell, compatible with unenriched scRNA-seq, integrates clonotype with gene expression.	Only single-cell data; partial or single-chain reconstruction possible; computationally intensive.	https://github.com/Teichlab/tracer	[23]*
Clustering							
tcrdist3	Paired or unpaired TCR α/β chain sequences (typically CDR1, CDR2, CDR3 regions for each chain, V/J gene)	Pairwise distance matrix (used for identifying nearest neighbors, performing clustering, and defining TCR meta-clonotypes for biomarker discovery).	Pairwise distance metrics by summing the weighted sequence differences across TCR CDRs and incorporating V/J gene usage.	Effective TCR similarity quantification and computationally efficient for pairwise comparisons. Identifies highly similar clonotypes.	Potential gap in complex shared specificity capture (non-linear patterns). Requires existing knowledge/large datasets for contextualization; lacks explicit modeling of 3D structure and binding kinetics.	https://github.com/kmayerb/tcrdist3	[27]*
GLIPH2 (Grouping of Lymphocyte Interactions by Paratope Hotspots)	Unpaired TCR β chain CDR3 sequences for clustering. Paired TCR α/β chain sequences to reconstruct the TCR heterodimer	TCR clusters with enriched motifs, consensus motifs, predicted HLA restriction.	Graph-based Clustering followed by Motif Enrichment to identify shared motifs among clustered TCRs.	Powerful for discovering shared specificities (public clonotypes); identifies antigen-associated motifs (even without labels). Predicts HLA restriction associated with TCR clusters	High computational intensity for very large datasets; sensitive to parameter tuning. Potential to overlook specificity from other TCR or V/J gene features.	https://rdrr.io/github/HetzDra/turboGliph/man/gliph2.html	[30]*
DeepTCR	Paired or unpaired TCR α/β chain sequences, V(D)J gene, Antigen/epitope-label data	TCR/repertoire embeddings, antigen-specificity prediction score.	DNN learning a joint representation that combines CDR3 sequence patterns with V(D)J gene usage information.	Enhanced performance in pattern-recognition and TCR classification. Improved TCR featurization; generalizes well across human and murine datasets.	Requires large, accurately labeled datasets. Potential lack of direct mechanistic interpretability.	https://github.com/sidhomj/DeepTCR	[33]*

Continued

Table 2 Continued

Tool	Input data	Output	Primary methodology	Strengths	Limitations	Implementation details	Paper
Structural Modeling							
AlphaFold3	Sequences (Proteins, DNA, RNA), ligands (SMILES), ions, modifications	3D complex structure.	Deep learning: diffusion model (Pairformer).	Multimodal (DNA, RNA, ligands, ions) with better interface accuracy.	Hallucinations in disordered regions.	https://github.com/google-deepmind/alphafold3	[38]*
TCRdock	TCR and pMHC sequences	Docked structure.	Hybrid AF/template: specialized docking sampling.	Specialized in capturing correct docking orientation (often difficult for standard AF); open source.	Can be complex to set up locally; focused specifically on the docking problem, not just folding.	https://github.com/phbradley/TCRdock	[44]*
TCRBuilder2	TCR α/β chain sequence	Fv region 3D model.	Deep Learning: error prediction network (ImmuneBuilder).	Speed, accuracy, and accessibility, achieved through modifications to the AlphaFold framework.	Focused on the isolated TCR variable region.	https://github.com/oxpig/ImmuneBuilder?tab=readme-ov-file#tcr-structure-prediction	[47]*
TCR Specificity							
NetTCR-2.0	CDR3alfa, CDR3beta, or both+three peptides available in the webserver	TCR–pMHC binding prediction.	One-dimensional CNN model.	Leverages paired chains (improves performance vs single-chain models); code+web server available; widely used baseline for paired-TCR specificity prediction.	Performance depends strongly on peptide/HLA coverage; the public web server focuses on limited peptide sets (notably HLA-A02:01) and does not directly output biophysical affinity.	https://services.healthtech.dtu.dk/services/NetTCR-2.0/ •Github: https://github.com/mnielLab/NetTCR-2.0	[53]*
ERGO/ERGO-II	ERGO: peptide+TCRbeta CDR3. ERGO-II: can use TCRalfa, V/J genes, MHC allele, and T cell type	TCR–peptide binding prediction.	Sequence models for ERGO; ERGO-II extends the feature set and provides a flexible multi-input predictor.	Feature-flexible (ERGO-II); supports richer immunogenetic context; open-source+web server; good for ablation studies on which features help.	Quality depends on availability/accuracy of auxiliary features; still limited by dataset bias and negative sampling; generalization to unseen epitopes remains challenging.	ERGO: https://github.com/louzounlab/ERGO; ERGO-II: https://github.com/IdoSpringer/ERGO-II	[54]*
ImRex	TCR CDR3 (commonly beta) + peptide epitope sequence; labeled binding pairs.	Binding prediction/probability; can also output interpretable interaction maps/attributions (model-dependent).	CNN on interaction maps derived from amino-acid pairwise features between TCR and peptide (image-classification perspective).	Interpretable via interaction-map saliency/filters; well-suited to seen-epitope settings; open-source implementation.	Generalization to unseen epitopes is difficult (explicitly discussed by the authors); depends on negative sampling and training split; typically CDR3-centric.	https://github.com/pmoris/ImRex	[56]*

The table presents general information for the three most-cited papers for each of the four steps in the TCR analysis workflow: TCR reconstruction and inference, TCR clustering, structural modeling of TCR and TCR–pMHC complexes, and specificity prediction.
*The completed non-extensive review of tools for TCR analysis table, including all tools and references, is available in the online supplemental material.
AF, AlphaFold; BAMs, Binary Alignment/Map files; CDRs, complementarity-determining regions; CNN, Convolutional Neural Network; 3D, three-dimensional; DNN, Deep Neural Network; ERGO, pEptide tcR matchinG predictiOn; Fv, fragment variable; IMGT, International ImMunoGeneTics; pMHC, peptide-major histocompatibility complex; scRNA-seq, single-cell RNA-sequencing; TCR, T cell receptor; UMI, Unique Molecular Identifier.

exhibited population bias due to underrepresentation of diverse ancestries in the sampled datasets.[44] This limitation can reduce reconstruction accuracy, particularly for individuals carrying non-canonical or population-specific alleles. In contrast, de novo assembly reconstructs TCR sequences without reliance on a reference database, instead overlapping reads directly to form contigs. While this approach offers greater flexibility and avoids reference bias, it may suffer from reduced accuracy in regions of high sequence similarity or low read coverage. De novo methods are often faster and require less memory than reference-guided approaches, but this computational efficiency can come at the cost of precise gene assignment.

Beyond assembly strategy, recovery of TCR sequences can be substantially improved by targeting the amplification of TCR α and β chains with pre-designed primers during library preparation, an approach recommended in most TCR reconstruction pipelines.[45] Regardless of the reconstruction method, at present, only single-cell-based approaches can directly resolve paired α and β chains, as they employ cell-specific barcodes to link receptor chains originating from the same lymphocyte.[42 46 47] Ultimately, the choice of tool for a given analysis depends on the sequencing technology used, the chosen platform, the available computational resources, and time constraints. Another practical consideration is the level of coding required. Some tools, such as MiXCR[42] and Vidjil,[48] provide user-friendly platforms that allow analyses to run with minimal coding, facilitating broader accessibility for researchers.

TCR clustering

While TCR reconstruction and inference can facilitate integration of clinical or biological metadata, providing translational insights into immune dynamics,[34] repertoire-level summaries alone cannot fully capture patterns of shared antigen recognition or functional convergence across clonotypes. This gap motivates the need for TCR clustering algorithms, which can identify groups of TCRs with shared sequence features across diverse TCR repertoires. Unlike typical proteins, TCRs produce extensive somatic diversity independent of evolutionary relationship, therefore limiting the utility of overall sequence similarity or evolutionary conservation for inferring function.[49 50] For instance, TCRs that recognize the same antigen can vary widely outside the CDR3 region but frequently share similar CDR3 motifs imposed by biochemical constraints (ie, convergent recognition). This property provides a powerful signal for identifying antigen-specific T cells and can be leveraged to detect tumor-reactive clones, making it a valuable biomarker for monitoring or predicting responses to cancer immunotherapy.[51]

TCRs sharing similar sequences are expected to exhibit similar arrangements of key antigen-contacting residues, which are reflected by small pairwise distances or higher similarity within an appropriate sequence or embedding representation space. This assumption underlies TCR clustering, and diverse methodologies have emerged depending on how differences between sequences are defined. Early approaches predominantly relied on simple string-based metrics, such as Hamming or Levenshtein distance, which treat all amino acid substitutions equally. However, their inability to capture observed amino acid substitution preferences led to the widespread adoption of substitution matrices, which assign scores to amino acid replacements based on their observed frequencies or biochemical similarity. Among these, BLOSUM62[52] has become the primary substitution matrix for TCR comparison.[53 54]

TCR clustering algorithms can be broadly grouped into distance-based, graph-based, and integrative algorithms (table 2, online supplemental table 3). Distance-based methods compute pairwise distance matrices and combine them with various clustering algorithms; notably, TCRdist3[53] assigns differential weights across CDR1–3 regions, with greater emphasis on CDR3, and integrates BLOSUM62-based substitution penalties to quantify amino-acid substitutions and insertions/deletions within a weighted Levenshtein framework, yielding biologically informed distances for hierarchical clustering. Graph-based approaches construct networks in which nodes represent individual TCRs and edges encode pairwise similarity derived from sequence-based similarity matrices, often using thresholded or weighted connections. Clustering is then performed by analyzing network connectivity patterns, with iSMART,[55] clusTCR,[56] and GLIPH2,[54] serving as representative examples. Integrative approaches incorporate additional biological information beyond TCR sequence to improve interpretability and statistical robustness. Representative examples include deep learning–based embedding methods such as DeepTCR,[57] TCR-BERT,[58] BertTCR,[58] and TouCAN,[59] which learn low-dimensional representations of TCR sequences and perform clustering in the resulting embedding space, as well as TCRMatch, which clusters TCRs using similarity scores informed by reference antigen-specific databases. LRT (Lineage inference by integrative analysis of scRNA-seq and scTCR-seq data)[60] represents a complementary, integrative strategy that clusters TCRs based on similarities in their lineage distribution patterns rather than on direct sequence similarity.

Antigen-driven TCR motif convergence has been documented in cancer studies, reflecting repeated immune selection of T cells recognizing shared or closely related antigens. For example, across virus-associated cancers, including cervical cancer and Epstein-Barr virus (EBV)-positive nasopharyngeal carcinoma, subsets of TCRs exhibit recurrent CDR3 motifs that distinguish disease states or predict clinical outcomes, consistent with antigen-driven convergent immune selection.[61 62] However, the detectability of such convergence varies substantially across tumor types, disease stages, and anatomical compartments. In settings where immune selection is relatively focused, such as within a single cancer type, intrapatient analyses, or selected stages of

disease progression, motif-level similarity can provide an informative, although partial, signal of shared antigen specificity.

In these contexts, effective TCR clustering in cancer requires method selection aligned with the expected strength of motif convergence, which may depend on factors such as tumor type, MHC restriction, viral etiology, and whether analyses are performed within or across patients. Importantly, because TCR recognition is inherently MHC-restricted, meaningful detection of shared specificity across individuals is most reliable in MHC-matched settings, as identical antigens presented by different MHC molecules may generate distinct TCR solutions. When convergence is anticipated, purely sequence-based approaches that explicitly model similarity at the level of CDR3 motifs or global CDR sequence architecture, including distance-based and graph-based methods, provide an interpretable and computationally tractable framework for identifying antigen-associated TCR clusters. Benchmark evaluations also indicate that representative methods like TCRdist3, iSMART, and GLIPH2 occupy different regions of the purity-retention spectrum, where purity reflects antigen homogeneity within clusters and retention denotes the fraction of TCRs assigned to clusters. This analysis positions TCRdist3 in the intermediate range, with GLIPH2 and iSMART achieving higher purity at the cost of lower retention.[63]

By contrast, in settings with limited or weak motif convergence, such as cancers with broad antigenic diversity, substantial interpatient heterogeneity, or diffuse immune selection, sequence similarity alone may be insufficient to support reliable antigen-specific TCR clustering. In these scenarios, reliance on sequence-level convergence can lead to fragmented or unstable clusters that do not generalize across datasets. In such cases, integrative approaches that incorporate external information become essential, particularly when robustness to noise and broader applicability across datasets are prioritized. Representative examples include DeepTCR and TCRMatch, which leverage supervised learning or database-informed matching to infer antigen specificity, as well as LRT, which infers antigen-driven selection from clonal expansion and differentiation patterns rather than explicit sequence convergence. Collectively, these methods are widely used and well established in the field, representing distinct methodological paradigms rather than interchangeable solutions, and therefore requiring a context-dependent strategy for TCR clustering in cancer immunology.

Structural modeling

Going beyond sequence-based analysis of TCR repertoires, some of the most advanced and computationally intensive TCR prediction methods make use of structural information directly and represent an important frontier not only in immunotherapy, but also in computational biology (table 2, online supplemental table 4). Cancer immunotherapy, particularly TCR-engineered adoptive T cell therapy, has transformed the treatment of solid tumors and hematologic malignancies.[2 3] Rational design of therapeutic TCRs critically depends on understanding TCR–pMHC recognition in atomic resolution,[64] making structural prediction an essential step for the next generation of immunotherapies. X-ray crystallography and cryo-electron microscopy have now resolved hundreds of human and mouse TCR structures, thousands of pMHC structures, and hundreds of TCRs in complex with MHC or MHC-like molecules (figure 2a).[65–67] These advances have substantially expanded the available structural landscape and support current structure-based approaches. However, this amount of data remains insufficient to train advanced AI-based models, especially given the complexity of the TCR–pMHC system. As expected, current predictors tend to perform better on the conserved framework of TCRs and MHCs, than on the MHC-restricted peptide-ligand[68] or the CDR loops.[69 70]

Early TCR modeling tools, such as LYRA,[71] relied on classical homology modeling. Given a TCR amino acid sequence, these methods search the Protein Data Bank for the closest structural template and build models by "grafting" CDR loops onto conserved frameworks. Homology modeling has helped bridge the gap between rapidly expanding TCR sequencing datasets and the much smaller set of experimentally solved structures, but accuracy is strongly dependent on high-quality templates and often fails to properly account for alternative CDR loop sequences.[70]

This landscape changed with attention-based deep learning methods such as AlphaFold2 (AF2)[72] and AlphaFold3 (AF3),[73] and the use of large multiple-sequence alignments (MSAs) to compensate for limited experimentally-determined structures.[74] These advances enable structure-guided TCR engineering for immunotherapy, allowing computational screening and optimization of therapeutic candidates.[75] On standard benchmarks, MSA-based models like AF2/AF3 still set the bar for global structural accuracy, particularly for multimeric assemblies, whereas most "MSA-free" protein language model (pLM)-based folding methods trade some accuracy for both speed and scalability.[35 69] Nevertheless, pLM-based architectures are improving rapidly and already offer attractive throughput for large-scale applications,[35] including high-throughput TCR repertoire screening for neoantigen-reactive clones in personalized vaccines.[76]

Given the extreme diversity of TCRs, specialized tools adapted these general architectures to immunological data. TCRmodel2,[77] for example, builds on AlphaFold-Multimer but restricts sequence and structural templates to TCRs and MHC molecules. This improves both speed and accuracy for TCR–pMHC complexes relative to general-purpose models. TCRBuilder2[78] and TCRBuilder2+[79] emphasize high-throughput TCR modeling and achieve CDR3 accuracies that approach AlphaFold-Multimer[80] while being optimized for large repertoires. More recently, tFold-TCR uses a pretrained

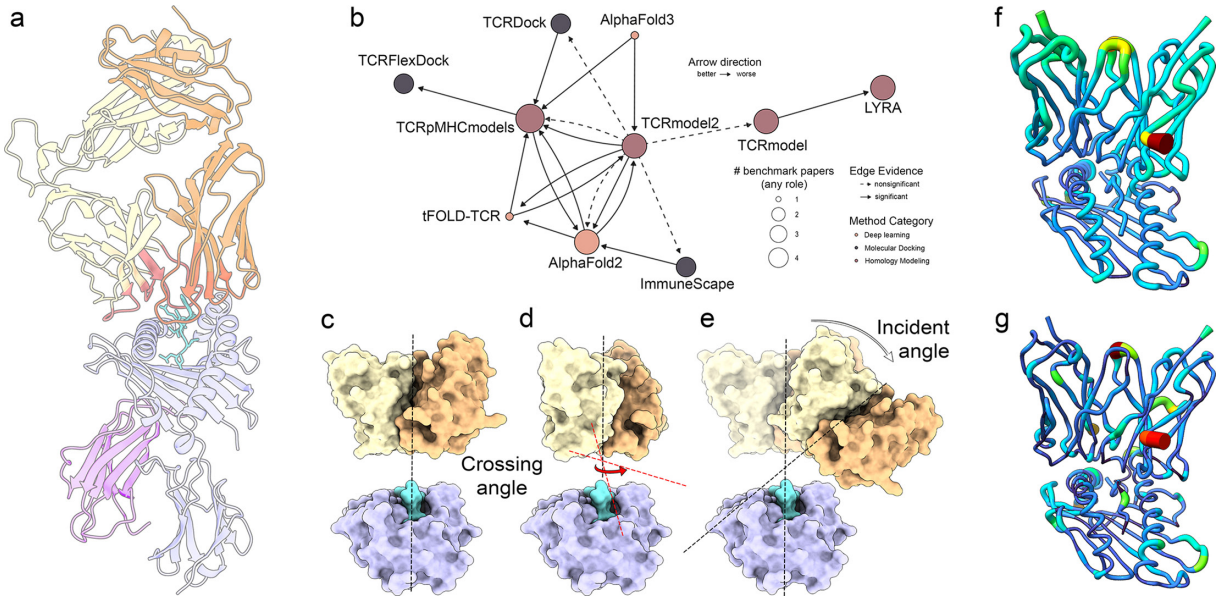

Figure 2 Structural features and modeling approaches of TCR–pMHC interactions. (a) Representative structure of a TCR bound to a pMHC, PDB: 2BNQ. (b) Dominance network of structural TCR and TCR–pMHC modeling tools summarizing cross-study benchmarking results. Nodes represent methods (size indicates number of benchmark studies), and directed edges indicate reported performance superiority (winner → loser). Solid edges denote statistically significant differences, while dashed edges represent qualitative claims. The network highlights widely benchmarked tools and patterns reported as superior across studies. (c–d) Illustration of crossing angle, rotational movement of the TCR relative to the pMHC along an axis perpendicular to the binding interface, highlighting the rotational shift (red arrow) of the TCR over the peptide-binding groove. (e) Illustration of incident angles, describing angular displacement of the TCR relative to the pMHC surface. The dashed lines indicate changes in the TCR binding mode's inclination. (f–g) Dynamics of TCR–pMHC complexes. Flipbook representations of conformational change over a short period of molecular dynamics simulation. The ribbon width and the color range depict low (deep blue) to high (red) flexibility per amino acid residue, as captured by RMSX (f) or LDDT (g). All protein images were rendered with UCSF ChimeraX. PDB, Protein Data Bank; pMHC, peptide-major histocompatibility complex; TCR, T cell receptor.

pLM (Evolutionary Scale Modeling for Protein-Protein Interactions, ESM-PPI) with separate TCR-specific and pMHC-specific encoders (ie, domain-specific training), and combines these models with an AlphaFold-style structure "docking" module, to predict full TCR–pMHC complexes without an explicit MSA step. This method yields DockQ scores comparable to AlphaFold-3[81] on the STCRDab-22 benchmark,[82] while running tens of times faster.[35]

Independent benchmarks focusing specifically on TCRs support several general trends in structural prediction accuracy and docking performance (figure 2b). Shi *et al*[69] demonstrated that for isolated TCRs, AF2, AF3, and tFold-TCR achieve the highest overall structural accuracy, whereas for TCR–pMHC complexes, AF2 and TCRmodel2 often performed best, with tFold-TCR providing a favorable accuracy-throughput trade-off. Note that performance may also differ between complexes restricted to MHC class I or MHC class II (ie, $CD8^+$ and $CD4^+$ T cell recognition, respectively). Although MHC II complexes are often considered more challenging to model due to their open peptide-binding groove and longer peptides, benchmarking results across tools show heterogeneous trends. For example, tFold-TCR reported relatively lower prediction accuracy for TCR–pMHC II complexes than for TCR–pMHC I, attributing this discrepancy to structural complexity and the limited representation of MHC II structures in training datasets.[35] In contrast, AF-TCR,[83] a specialized AlphaFold-based docking framework, and TCRmodel2,[77] found improved predictions and higher-confidence scores of class II binding modes relative to class I. These findings suggest that while peptide length and docking variability can complicate TCR–pMHC II modeling, current AI-based frameworks do not uniformly underperform on class II systems, and their performance is rather more influenced by the composition of the training data.

Interestingly, one of the remaining challenges for accurate TCR–pMHC structural prediction relates to the proper "docking" orientation of the TCR (ie, crossing and incident angles) (figure 2c–e).[70] The use of a TCR–pMHC template can give a docking-based approach (eg, ImmuneScape) an edge over AI-based approaches (eg, TCRmodel2) when considering the accuracy of crossing and incident angles, regardless of TCRmodel2 being consistently better at predicting the CDR loops. The most recent tFold-TCR implementation seems to have achieved the best of both worlds by separately modeling the TCR and the pMHC, and having a dedicated model to predict the "docked" TCR–pMHC structure.[35]

Note that the modeling tools described in this section can potentially generate structures for any TCR–pMHC complex, but even a high-confidence model is not a strong indication of TCR–pMHC engagement. First,

current limitations on the accuracy of scoring functions for protein–protein interactions prevent us from deriving accurate binding energies directly from modeled complexes. Second, a static structure still lacks dynamic properties important for binding, which might be better approximated by an ensemble of conformations.[84]

Despite remaining limitations, TCR–pMHC structural modeling remains a key research focus, with many new computational tools currently being developed to advance the field. Notably, structure-based models that leverage explicit or predicted TCR–pMHC structures (eg, NetTCR-struct[85] or STAG[86]) have been shown to outperform purely sequence-based methods in benchmark studies, provided that the structural inputs are of sufficient quality (figure 2f–g).[85 87] In this context, there is growing interest in methods that can perform better quality assessment of three-dimensional models produced by different tools,[85 88] and may help refine these models using more accurate biophysics-based features and advanced sampling methods.

These advances are beginning to translate into clinical applications. For instance, structure-guided engineering has produced enhanced-affinity TCRs targeting shared tumor antigens like NY-ESO-1 and MAGE-A4, with candidates in clinical trials for melanoma and sarcoma.[16] Integrating structural predictions with TCR repertoire sequencing from patient tumors enables the identification of therapeutically relevant tumor-infiltrating lymphocytes, accelerating personalized TCR-T cell therapy development.[27]

Specificity prediction

The repertoire analyses discussed in previous sections can identify shifts in clonal proportions and even identify particular TCR sequences that are more prevalent or potentially associated with a phenotype or clinical outcome. However, the missing piece of the puzzle remains the identification of the exact antigen-specificities for these TCRs of interest. Clustering methods have often been used to provide an indirect answer to that question by identifying similarities between TCRs of interest and TCRs with verified specificity. This approach, however, provides a very coarse approximation, due to limitations of clustering methods, lack of extensive public datasets for reference, and the non-trivial impacts of alternative CDRs or Vα/Vβ pairings. Here, we focus on a different group of tools that try to bridge this gap and directly predict TCR specificity (table 2, online supplemental table 5). For instance, tools like PanPep[89] and TPepRet[90] are set to predict TCR–peptide binding using as input only the sequences of the TCR and peptide of interest (ie, predict a "match"). Despite being sequence-based tools, these methods implicitly capture aspects of structural constraint patterns derived from known contacts within the TCR–pMHC interface,[58 89] without the overhead cost of three-dimensional structural modeling. However, they are usually limited to the CDR3β interactions, and do not explicitly account for the MHC allele.

Another category of tools requires only the TCR sequence as input, outputting a single or multiple putative peptide-target sequences (eg, TCR-BERT[91] and TCRconv, respectively). These tools represent an important step towards the desired goal of predicting antigen-specificity for unknown TCRs. However, they are still limited to the peptides used in the training dataset and therefore cannot predict novel peptide-targets. The inverse task of predicting epitope-specific TCRs can be performed with TCR-epiDiff.[92] The generative capabilities of TCR-epiDiff also enable additional downstream analysis, as provided by the tool's TCR–peptide binding prediction pipeline. As for limitations, TCR-epiDiff is computationally expensive and also limited to the TCR-β chain.

Finally, methods such as TCRcost[88] and STAG-LLM[86] require structural models as input for TCR–peptide binding prediction. TCRcost even enables correcting structural models prior to prediction, but does not include the MHC structure. On the other hand, STAG-LLM[86] accounts for the structural information of the complete TCR–pMHC interface, potentially capturing more of the biophysical and biochemical properties driving TCR engagement with pMHC complexes. As expected, reliance on structural data comes with some trade-offs, including the additional cost of the modeling step, which could limit scalability for large-scale TCR repertoire analysis. Given the complementary capabilities of the tools described in this section, it is already possible to overcome some of their individual limitations by integrating these methods into more comprehensive pipelines, from initial filtering with sequence-based methods,[93 94] to three-dimensional modeling, quality assessment, and structure-guided specificity prediction for selected complexes.[85 87 95]

Functional and clinical integration

The design of effective personalized cancer immunotherapies relies fundamentally on the precise identification of tumor-associated antigens and, ideally, the potentially therapeutic TCRs that recognize these MHC-bound peptide-targets. While traditional approaches have mostly focused on tumor variability for neoantigen discovery, through next-generation sequencing and immunopeptidomics, there has been a recent shift towards TCR-guided neoantigen discovery, powered by novel high-throughput experimental approaches such as Yeast-displayed pMHC libraries, signaling and antigen-presenting bifunctional receptors, Biomechanically-Assisted T cell Triggering for Large-scale Exogenous-pMHC Screening, and TCR-epitope scanning (T-Scan).[96]

Approaches based on TCR-sequencing have also been developed, such as the 10x Genomics Barcode Enabled Antigen Mapping (BEAM-T), now discontinued, and the U-Load dCODE Dextramer from Immudex. However, the cost of these experimental approaches remains a major challenge for translational applications, especially when considering personalized therapies. In this context, the emerging computational methods and databases reviewed in

previous sections provide an additional avenue for accelerating translational applications in cancer immunotherapy, through neoantigen-specific TCR-discovery, and computational TCR-guided neoantigen discovery. For instance, new methods for TCR data acquisition, TCR inference, and TCR clustering, combined with scalable approaches for structural modeling and specificity prediction, can allow identification of the best existing TCRs in a patient's repertoire to drive cellular immunity against a specific neoantigen displayed by the patient's tumor. Conversely, knowledge of the patient's TCR repertoire and the capacity to computationally predict or test neoantigen-specificities can actually help overcome some of the limitations of neoantigen discovery pipelines.

The use of the patients' own TCRs for cellular therapy or TCR-guided neoantigen discovery has the additional advantage of reducing the risk for off-target toxicities, given the central tolerance mechanisms.[19] However, the great expansion of a selected clonal population prior to adoptive T cell transfer, in combination with lymphodepletion, can still push the system beyond a natural tolerance threshold by breaking the boundaries of biological constraints imposed by central and peripheral tolerance mechanisms. This may increase the risk of cross-reactivity and off-target toxicity, as regulatory checkpoints that normally limit autoreactive potential are reduced or bypassed. This risk can be further minimized by computational screening of the most likely off-targets for a given therapeutic TCR or neoantigen, and subsequent experimental validation.[97 98] In turn, all this information can be used to tailor the best therapy, or combination of therapies, considering available resources and clinical details. Of course, several limitations are still hindering the broader adoption of these computational approaches into clinical use.

In the regulatory context, computational tools currently play a supportive rather than determinative role in the approval process of TCR-engineered T cell therapies. Regulatory agencies such as the US Food and Drug Administration require extensive experimental validation of specificity, safety, and manufacturability, and computational predictions are primarily used to prioritize candidates, assess potential off-target risks, and guide preclinical study design. Although in silico modeling, cross-reactivity screening, and structural prediction can substantially de-risk candidate selection, these approaches are not yet standardized or sufficiently benchmarked to serve as standalone evidence for regulatory decision-making. Improving those steps and having more preclinical evaluation pipelines will be essential for broader regulatory acceptance in the future.

The recent AI-driven revolution in structural modeling has enabled increasingly accurate predictions of TCR–pMHC structures.[74 77] However, reliably deriving TCR binding affinity from these complexes remains challenging. Moreover, key determinants of T cell activation are still incompletely captured, particularly those related to molecular dynamics and the mechanical forces governing TCR engagement and downstream signal transduction.[99–102] An orthogonal strategy for TCR specificity prediction involves deep learning models trained on validated TCR–pMHC complexes.[86] Yet, limited availability of ground-truth data, including "true negative" TCR–antigen pairs, constrains both training and robust validation, leading to poor extrapolation to unseen datapoints, especially novel peptides.[95] Addressing these limitations will require deeper integration of experimental and computational approaches, alongside the development of shared benchmarking standards and interoperable databases. Expanding training datasets and feature representations, coupled with more efficient algorithms and an improved mechanistic understanding of TCR specificity, holds significant promise for transforming the design and broadening the clinical impact of cancer immunotherapies.

CONCLUSION

In this review, we reflect on the current understanding of TCR biology and how its complexity creates challenges for TCR repertoire analyses. In spite of that, we highlight how advances in high-throughput sequencing and bioinformatics have transformed TCR repertoire analysis into a central pillar of translational cancer immunotherapy. We outline a unified, end-to-end analytical workflow, from data acquisition and TCR reconstruction to clustering, structural modeling, specificity prediction, and clinical integration. We discuss the strengths and limitations of existing computational approaches, including emerging deep learning and structure-informed methods, and underscore persistent challenges such as limited ground-truth datasets, TCR cross-reactivity, and limited interpretability. Collectively, these advances position TCR-focused analyses as indispensable tools for precision immuno-oncology, enabling predictive, mechanistic, and clinically actionable approaches to immunotherapy design.

Author affiliations
[1]Biology and Biochemistry, University of Houston, Houston, Texas, USA
[2]Mathematics, University of Houston, Houston, Texas, USA
[3]Biochemistry and Molecular Biology, Universidade Federal do Ceará, Fortaleza, Brazil
[4]Computer Science, University of Houston, Houston, Texas, USA
[5]Postgraduate program in Biochemistry, Universidade Federal do Ceará, Fortaleza, Brazil
[6]Biophysics, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil
[7]Biostatistics and Data Science, The University of Texas Medical Branch at Galveston, Galveston, Texas, USA
[8]Medicine, Universidade de São Paulo, São Paulo, Brazil
[9]Clinical Pharmacy, University of Southern California, Los Angeles, California, USA
[10]Department of Medicine, The University of Chicago, Chicago, Illinois, USA

[11]David and Etta Jonas Center for Cellular Therapy, The University of Chicago, Chicago, Illinois, USA
[12]Department of Molecular Biology and Biotechnology, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil
[13]Health Systems and Population Health Sciences, University of Houston, Houston, Texas, USA

Acknowledgements We thank Jared Slone for his careful review and valuable feedback on the binding chapter, which helped improve the clarity and accuracy of this work.

Contributors PB, MVDF, JY, JL, FB, FJFDS, SHC, MR, HJ, and DAA contributed to writing and editing the manuscript. DBN contributed to reviewing the tools listed on the clustering table. GZ, ED, HA, JHJ, and SPW contributed to reviewing and editing the manuscript. All authors made direct and intellectual contributions to the work and approved the final manuscript. DAA will serve as a guarantor. The text of the manuscript was originally created by the authors. As part of the revision process, when needed, short segments of the text were submitted to ChatGPT in order to improve conciseness and clarity. All those revisions were further revised and modified by the authors until the final version of the manuscript was produced.

Funding This work was supported by the National Institutes of Health under grant number 1R21CA289333-01.

Competing interests No, there are no competing interests.

Patient consent for publication Not applicable.

Ethics approval Not applicable.

Provenance and peer review Commissioned; externally peer reviewed.

ORCID iDs
Pâmella Borges https://orcid.org/0000-0003-2250-7112
Duy Bao Nguyen https://orcid.org/0009-0001-7336-4644
Houda Alachkar https://orcid.org/0000-0001-5567-5521
Steven P Wolf https://orcid.org/0009-0008-2723-695X
Dinler Amaral Antunes https://orcid.org/0000-0001-7947-6455

REFERENCES

1 Aran A, Garrigós L, Curigliano G, *et al*. Evaluation of the TCR Repertoire as a Predictive and Prognostic Biomarker in Cancer: Diversity or Clonality? *Cancers (Basel)* 2022;14:1771.
2 Waldman AD, Fritz JM, Lenardo MJ. A guide to cancer immunotherapy: from T cell basic science to clinical practice. *Nat Rev Immunol* 2020;20:651–68.
3 Chandran SS, Klebanoff CA. T cell receptor-based cancer immunotherapy: Emerging efficacy and pathways of resistance. *Immunol Rev* 2019;290:127–47.
4 Mariuzza RA, Agnihotri P, Orban J. The structural basis of T-cell receptor (TCR) activation: An enduring enigma. *J Biol Chem* 2020;295:914–25.
5 Aversa I, Malanga D, Fiume G, *et al*. Molecular T-Cell Repertoire Analysis as Source of Prognostic and Predictive Biomarkers for Checkpoint Blockade Immunotherapy. *Int J Mol Sci* 2020;21:2378.
6 Davis MM, Bjorkman PJ. T-cell antigen receptor genes and T-cell recognition. *Nature New Biol* 1988;334:395–402.
7 Weng N. Numbers and odds: TCR repertoire size and its age changes impacting on T cell functions. *Semin Immunol* 2023;69:101810.
8 Rodríguez-Caparrós A, Álvarez-Santiago J, Del Valle-Pastor MJ, *et al*. Regulation of T-cell Receptor Gene Expression by Three-Dimensional Locus Conformation and Enhancer Function. *Int J Mol Sci* 2020;21:8478.
9 Abarrategui I, Krangel MS. Regulation of T cell receptor-alpha gene recombination by transcription. *Nat Immunol* 2006;7:1109–15.
10 Padovan E, Casorati G, Dellabona P, *et al*. Expression of two T cell receptor alpha chains: dual receptor T cells. *Science* 1993;262:422–4.
11 Wang F, Ono T, Kalergis AM, *et al*. On defining the rules for interactions between the T cell receptor and its ligand: A critical role for a specific amino acid residue of the T cell receptor β chain. *Proc Natl Acad Sci USA* 1998;95:5217–22.
12 Hughes MM, Yassai M, Sedy JR, *et al*. T cell receptor CDR3 loop length repertoire is determined primarily by features of the V(D)J recombination reaction. *Eur J Immunol* 2003;33:1568–75.
13 Love PE, Hatzihristidis T, Bryant J, *et al*. Early events in TCR signaling - the evolving role of ITAMs. *Front Immunol* 2025;16:1563049.
14 Lo W-L, Weiss A. Adapting T Cell Receptor Ligand Discrimination Capability *via* LAT. *Front Immunol* 2021;12:673196.
15 Soto C, Bombardi RG, Branchizio A, *et al*. High frequency of shared clonotypes in human B cell receptor repertoires. *Nature New Biol* 2019;566:398–402.
16 Mo G, Lu X, Wu S, *et al*. Strategies and rules for tuning TCR-derived therapy. *Expert Rev Mol Med* 2023;26:e4.
17 Blankenstein T, Leisegang M, Uckert W, *et al*. Targeting cancer-specific mutations by T cell receptor gene therapy. *Curr Opin Immunol* 2015;33:112–9.
18 Wong WK, Leem J, Deane CM. Comparative Analysis of the CDR Loops of Antigen Receptors. *Front Immunol* 2019;10:2454.
19 Antunes DA, Rigo MM, Freitas MV, *et al*. Interpreting T-Cell Cross-reactivity through Structure: Implications for TCR-Based Cancer Immunotherapy. *Front Immunol* 2017;8:1210.
20 Pretti MAM, Vieira GF, Boroni M, *et al*. Unveiling cross-reactivity: implications for immune response modulation in cancer. *Brief Bioinformatics* 2024;26:bbaf012.
21 Sarén T, Saronio G, Marti Torrell P, *et al*. Complementarity-determining region clustering may cause CAR-T cell dysfunction. *Nat Commun* 2023;14:4732.
22 Chikovani T, Magen E. In Situ Vaccination by Tumor Ablation: Principles and Prospects for Systemic Antitumor Immunity. *Vaccines (Basel)* 2025;13:1114.
23 Desai PN, Wang B, Fonseca A, *et al*. Single-Cell Profiling of CD8+ T Cells in Acute Myeloid Leukemia Reveals a Continuous Spectrum of Differentiation and Clonal Hyperexpansion. *Cancer Immunol Res* 2023;11:1011–28.
24 Kalaora S, Wolf Y, Feferman T, *et al*. Combined Analysis of Antigen Presentation and T-cell Recognition Reveals Restricted Immune Responses in Melanoma. *Cancer Discov* 2018;8:1366–75.
25 Rosenberger L, Hansmann L, Anastasopoulou V, *et al*. Selection of therapeutically effective T-cell receptors from the diverse tumor-bearing repertoire. *J Immunother Cancer* 2025;13:e011351.
26 Rosati E, Dowds CM, Liaskou E, *et al*. Overview of methodologies for T-cell receptor repertoire analysis. *BMC Biotechnol* 2017;17:61.
27 Frank ML, Lu K, Erdogan C, *et al*. T-Cell Receptor Repertoire Sequencing in the Era of Cancer Immunotherapy. *Clin Cancer Res* 2023;29:994–1008.
28 Pai JA, Satpathy AT. High-throughput and single-cell T cell receptor sequencing technologies. *Nat Methods* 2021;18:881–92.
29 Shorer O, Pinhasi A, Yizhak K. Single-cell meta-analysis of T cells reveals clonal dynamics of response to checkpoint immunotherapy. *Cell Genom* 2025;5:100842.
30 Yassai MB, Naumov YN, Naumova EN, *et al*. A clonotype nomenclature for T cell receptors. *Immunogenetics* 2009;61:493–502.
31 Yamawaki TM, Lu DR, Ellwanger DC, *et al*. Systematic comparison of high-throughput single-cell RNA-seq methods for immune cell profiling. *BMC Genomics* 2021;22:66.
32 Zhang Y, Yang X, Zhang Y, *et al*. Tools for fundamental analysis functions of TCR repertoires: a systematic comparison. *Brief Bioinformatics* 2020;21:1706–16.
33 Mazzotti L, Gaimari A, Bravaccini S, *et al*. T-Cell Receptor Repertoire Sequencing and Its Applications: Focus on Infectious Diseases and Cancer. *Int J Mol Sci* 2022;23:8590.
34 Arunkumar M, Zielinski CE. T-Cell Receptor Repertoire Analysis with Computational Tools—An Immunologist's Perspective. *Cells* 2021;10:3582.

35 Wu F, Zhao Y, Xiao Y, *et al*. Fast and accurate modeling of tcr-peptide-mhc complexes using tfold-tcr. *Bioinformatics* [Preprint] 2025.
36 Yue T, Chen SY, Shen WK, *et al*. TCRdb 2.0: an updated T-cell receptor sequence database. *Nucleic Acids Res* 2026;54:D504–10.
37 Ershler WB. Biomarkers of aging: immunological events. *Exp Gerontol* 1988;23:387–9.
38 Ullrich F, Bröckelmann PJ, Turki AT, *et al*. Impact of immunological aging on T cell-mediated therapies in older adults with multiple myeloma and lymphoma. *J Immunother Cancer* 2024;12:e009462.
39 Huang Z, Chen B, Liu X, *et al*. Effects of sex and aging on the immune cell landscape as assessed by single-cell transcriptomic analysis. *Proc Natl Acad Sci USA* 2021;118.
40 Peng K, Nowicki TS, Campbell K, *et al*. Rigorous benchmarking of T-cell receptor repertoire profiling methods for cancer RNA sequencing. *Brief Bioinformatics* 2023;24:bbad220.
41 Mahdy AKH, Fuß J, Franzenburg S, *et al*. Quantifying the impact of bulk tcr-seq methodological choices on the profiled t cell repertoire. *Immunology* [Preprint].
42 Bolotin DA, Poslavsky S, Mitrophanov I, *et al*. MiXCR: software for comprehensive adaptive immunity profiling. *Nat Methods* 2015;12:380–1.
43 Marquez S, Babrak L, Greiff V, *et al*. Adaptive Immune Receptor Repertoire (AIRR) Community Guide to Repertoire Analysis. *Methods Mol Biol* 2022;2453:297–316.
44 Khatri I, Berkowska MA, van den Akker EB, *et al*. Population matched (pm) germline allelic variants of immunoglobulin (IG) loci: Relevance in infectious diseases and vaccination studies in human populations. *Genes Immun* 2021;22:172–86.
45 Benichou J, Ben-Hamo R, Louzoun Y, *et al*. Rep-Seq: uncovering the immunological repertoire through next-generation sequencing. *Immunology* 2012;135:183–91.
46 Song L, Cohen D, Ouyang Z, *et al*. TRUST4: immune repertoire reconstruction from bulk and single-cell RNA-seq data. *Nat Methods* 2021;18:627–30.
47 Stubbington MJT, Lönnberg T, Proserpio V, *et al*. T cell fate and clonality inference from single-cell transcriptomes. *Nat Methods* 2016;13:329–32.
48 Giraud M, Salson M, Duez M, *et al*. Fast multiclonal clusterization of V(D)J recombinations from high-throughput sequencing. *BMC Genomics* 2014;15:409.
49 Garboczi DN, Ghosh P, Utz U, *et al*. Structure of the complex between human T-cell receptor, viral peptide and HLA-A2. *Nature New Biol* 1996;384:134–41.
50 Rossjohn J, Gras S, Miles JJ, *et al*. T cell antigen receptor recognition of antigen-presenting molecules. *Annu Rev Immunol* 2015;33:169–200.
51 Pan M, Li B. T cell receptor convergence is an indicator of antigen-specific T cell response in cancer immunotherapies. *Elife* 2022;11:e81952.
52 Henikoff S, Henikoff JG. Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci USA* 1992;89:10915–9.
53 Dash P, Fiore-Gartland AJ, Hertz T, *et al*. Quantifiable predictive features define epitope-specific T cell receptor repertoires. *Nature New Biol* 2017;547:89–93.
54 Huang H, Wang C, Rubelt F, *et al*. Analyzing the Mycobacterium tuberculosis immune response by T-cell receptor clustering with GLIPH2 and genome-wide antigen screening. *Nat Biotechnol* 2020;38:1194–202.
55 Zhang H, Liu L, Zhang J, *et al*. Investigation of Antigen-Specific T-Cell Receptor Clusters in Human Cancers. *Clin Cancer Res* 2020;26:1359–71.
56 Valkiers S, Van Houcke M, Laukens K, *et al*. ClusTCR: a python interface for rapid clustering of large sets of CDR3 sequences with unknown antigen specificity. *Bioinformatics* 2021;37:4865–7.
57 Sidhom JW, Larman HB, Pardoll DM, *et al*. DeepTCR is a deep learning framework for revealing sequence concepts within T-cell repertoires. *Nat Commun* 2021;12:1605.
58 Wu KE, Yost K, Daniel B, *et al*. TCR-bert: learning the grammar of t-cell receptors for flexible antigen-binding analyses. In: *Machine learning in computational biology*. PMLR, 2024: 194–229.
59 Pertseva M, Follonier O, Scarcella D, *et al*. TCR clustering by contrastive learning on antigen specificity. *Brief Bioinform* 2024;25:bbae375.
60 Xie J, Jeon H, Xin G, *et al*. LRT: Integrative analysis of scRNA-seq and scTCR-seq data to investigate clonal differentiation heterogeneity. *PLoS Comput Biol* 2023;19:e1011300.
61 Cui J-H, Lin K-R, Yuan S-H, *et al*. TCR Repertoire as a Novel Indicator for Immune Monitoring and Prognosis Assessment of Patients With Cervical Cancer. *Front Immunol* 2018;9:2729.
62 Wang G, Mudgal P, Wang L, *et al*. TCR repertoire characteristics predict clinical response to adoptive CTL therapy against nasopharyngeal carcinoma. *Oncoimmunology* 2021;10:1955545.
63 Jouannet C, Vantomme H, Gouge KL, *et al*. Benchmarking unsupervised methods for inferring TCR specificity. *NAR Genom Bioinform* 2025;7:lqaf150.
64 Crean RM, MacLachlan BJ, Madura F, *et al*. Molecular Rules Underpinning Enhanced Affinity Binding of Human T Cell Receptors Engineered for Immunotherapy. *Mol Ther Oncolytics* 2020;18:443–56.
65 Quast NP, Abanades B, Guloglu B, *et al*. T-cell receptor structures and predictive models reveal comparable alpha and beta chain structural diversity despite differing genetic complexity. *Commun Biol* 2025;8:362.
66 Lin V, Cheung M, Gowthaman R, *et al*. TCR3d 2.0: expanding the T cell receptor structure database with new structures, tools and interactions. *Nucleic Acids Res* 2025;53:D604–8.
67 Lab P. TCR3d: T cell receptor structural repertoire database. Class I and Class II complex statistics (273 and 96 structures, respectively). 2025.
68 Antunes DA, Vieira GF, Rigo MM, *et al*. Structural allele-specific patterns adopted by epitopes in the MHC-I cleft and reconstruction of MHC:peptide complexes to cross-reactivity assessment. *PLoS One* 2010;5:e10353.
69 Shi Y, Parks JM, Smith JC. Comparative Analysis of TCR and TCR-pMHC Complex Structure Prediction Tools. *J Chem Inf Model* 2025;65:7156–73.
70 Le HN, de Freitas MV, Antunes DA. Strengths and limitations of web servers for the modeling of TCRpMHC complexes. *Comput Struct Biotechnol J* 2024;23:2938–48.
71 Klausen MS, Anderson MV, Jespersen MC, *et al*. LYRA, a webserver for lymphocyte receptor structural modeling. *Nucleic Acids Res* 2015;43:W349–55.
72 Jumper J, Evans R, Pritzel A, *et al*. Highly accurate protein structure prediction with AlphaFold. *Nature New Biol* 2021;596:583–9.
73 Abramson J, Adler J, Dunger J, *et al*. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature New Biol* 2024;630:500:493–500:.
74 Sapoval N, Aghazadeh A, Nute MG, *et al*. Current progress and open challenges for applying deep learning across the biosciences. *Nat Commun* 2022;13:1728.
75 Wu D, Yin R, Chen G, *et al*. Structural characterization and AlphaFold modeling of human T cell receptor recognition of NRAS cancer neoantigens. *Sci Adv* 2024;10:eadq6150.
76 Joshi K, Milighetti M, Chain BM. Application of T cell receptor (TCR) repertoire analysis for the advancement of cancer immunotherapy. *Curr Opin Immunol* 2022;74:1–8.
77 Yin R, Ribeiro-Filho HV, Lin V, *et al*. TCRmodel2: high-resolution modeling of T cell receptor recognition using deep learning. *Nucleic Acids Res* 2023;51:W569–76.
78 Abanades B, Wong WK, Boyles F, *et al*. ImmuneBuilder: Deep-Learning models for predicting the structures of immune proteins. *Commun Biol* 2023;6:575.
79 Quast NP, Abanades B, Guloglu B, *et al*. T-cell receptor structures and predictive models reveal comparable alpha and beta chain structural diversity despite differing genetic complexity. *Immunology* [Preprint] 2024.
80 Evans R, O'Neill M, Pritzel A, *et al*. Protein complex prediction with alphafold-multimer. *Bioinformatics* [Preprint] 2021.
81 Basu S, Wallner B. DockQ: A Quality Measure for Protein-Protein Docking Models. *PLoS One* 2016;11:e0161879.
82 Leem J, de Oliveira SHP, Krawczyk K, *et al*. STCRDab: the structural T-cell receptor database. *Nucleic Acids Res* 2018;46:D406–12.
83 Bradley P. Structure-based prediction of T cell receptor:peptide-MHC interactions. *Elife* 2023;12:e82813.
84 Beruldsen F, de Freitas MV, Antunes DA. High resolution mapping of protein motions in time and space with RMSX and Flipbook. *Sci Rep* 2026;16:10035.
85 Deleuran SN, Nielsen M. NetTCR-struc, a structure driven approach for prediction of TCR-pMHC interactions. *Front Immunol* 2025;16:1616328.
86 Slone JK, Zhang M, Jiang P, *et al*. STAG-LLM: Predicting TCR-pHLA binding with protein language models and computationally generated 3D structures. *Comput Struct Biotechnol J* 2025;27:3885–96.
87 Slone JK, Conev A, Rigo MM, *et al*. TCR-pMHC Binding Specificity Prediction from Structure Using Graph Neural Networks. IEEE Transactions on Computational Biology and Bioinformatics; 2025. Available: http://ieeexplore.ieee.org/document/ 10819951

88 Li F, Qian X, Zhu X, *et al*. TCRcost: a deep learning model utilizing TCR 3D structure for enhanced of TCR–peptide binding. *Front Genet* 2024;15:1346784.
89 Gao Y, Gao Y, Fan Y, *et al*. Pan-Peptide Meta Learning for T-cell receptor–antigen binding recognition. *Nat Mach Intell* 2023;5:236–49.
90 Wang M, Fan W, Wu T, *et al*. TPepRet: a deep learning model for characterizing T-cell receptors-antigen binding patterns. *Bioinformatics* 2024;41:btaf022.
91 Zhang M, Cheng Q, Wei Z, *et al*. BertTCR: a Bert-based deep learning framework for predicting cancer-related immune status based on T cell receptor repertoire. *Brief Bioinformatics* 2024;25.
92 Seo SY, Rhee JK. TCR-epiDiff: solving dual challenges of TCR generation and binding prediction. *Bioinformatics* 2025;41:i125–32.
93 Pham M-DN, Nguyen T-N, Tran LS, *et al*. epiTCR: a highly sensitive predictor for TCR-peptide binding. *Bioinformatics* 2023;39:btad284.
94 Myronov A, Mazzocco G, Król P, *et al*. BERTrand-peptide:TCR binding prediction using Bidirectional Encoder Representations from Transformers augmented with random TCR pairing. *Bioinformatics* 2023;39:btad468.
95 Nielsen M, Eugster A, Jensen MF, *et al*. Lessons learned from the IMMREP23 TCR-epitope prediction challenge. *ImmunoInformatics* 2024;16:100045.
96 Xie N, Shen G, Gao W, *et al*. Neoantigens: promising targets for cancer therapy. *Signal Transduct Target Ther* 2023;8:9.
97 Fonseca AF, Antunes DA. CrossDome: an interactive R package to predict cross-reactivity risk using immunopeptidomics databases. *Front Immunol* 2023;14:1142573.
98 Banerjee A, Pattinson DJ, Wincek CL, *et al*. T cell receptor cross-reactivity prediction improved by a comprehensive mutational scan database. *Cell Syst* 2025;16:101345.
99 Antunes DA, Baker BM, Cornberg M, *et al*. Editorial: Quantification and prediction of T-cell cross-reactivity through experimental and computational methods. *Front Immunol* 2024;15.
100 Yan M, Wang J, Zhang JZH. A kinetic gatekeeper: TCR-pMHC dissociation dynamics determine T cell activation efficacy. *Protein Sci* 2025;34:e70346.
101 Qin R, Zhang Y, Shi J, *et al*. n.d. TCR catch bonds nonlinearly control CD8 cooperation to shape T cell specificity. *Cell Res*35:265–83.
102 Wu X, Ye Z. Mechanoimmunology of T-Cell Activation. *Scand J Immunol* 2025;101:e70009.