TRAIT: A Comprehensive Database for T-cell Receptor–antigen Interactions

Mengmeng Wei (魏蒙蒙) [ID][1,2,#], **Jingcheng Wu (吴静成)** [ID][1,#], **Shengzuo Bai (白晟佐)** [ID][1,#],
Yuxuan Zhou (周宇轩) [ID][1], **Yichang Chen (陈弈菖)** [ID][1], **Xue Zhang (张雪)** [ID][1],
Wenyi Zhao (赵文艺) [ID][1,3], **Ying Chi (迟颖)** [ID][4,5], **Gang Pan (潘纲)** [ID][2], **Feng Zhu (朱峰)** [ID][1,3,5,*],
Shuqing Chen (陈枢青) [ID][1,*], **Zhan Zhou (周展)** [ID][1,3,5,6,*]

[1]State Key Laboratory of Advanced Drug Delivery and Release Systems & Zhejiang Provincial Key Laboratory of Anti-Cancer Drug Research, College of Pharmaceutical Sciences, Zhejiang University, Hangzhou 310058, China
[2]Collaborative Innovation Center of Artificial Intelligence by MOE and Zhejiang Provincial Government, College of Computer Science and Technology, Zhejiang University, Hangzhou 310027, China
[3]Innovation Institute for Artificial Intelligence in Medicine & MOE Engineering Research Center of Innovative Anticancer Drugs, Zhejiang University, Hangzhou 310018, China
[4]ZJU-UoE Institute, Zhejiang University, Haining 314400, China
[5]Alibaba-Zhejiang University Joint Research Center of Future Digital Healthcare, Hangzhou 310058, China
[6]The Fourth Affiliated Hospital, Zhejiang University School of Medicine, Yiwu 322000, China
*Corresponding authors: zhanzhou@zju.edu.cn (Zhou Z), chenshuqing@zju.edu.cn (Chen S), zhufeng@zju.edu.cn (Zhu F).
#Equal contribution.
Handling Editor: An-Yuan Guo

Abstract

Comprehensive and integrated resources on interactions between T-cell receptors (TCRs) and antigens are still lacking for adoptive T-cell-based immunotherapies, highlighting a significant gap that must be addressed to fully understand the mechanisms of antigen recognition by T cells. In this study, we present the T-cell receptor–antigen interaction database (TRAIT), a comprehensive database that profiles the interactions between TCRs and antigens. TRAIT stands out due to its comprehensive description of TCR–antigen interactions by integrating sequences, structures, and affinities. It provides millions of experimentally validated TCR–antigen pairs, resulting in an exhaustive landscape of antigen-specific TCRs. Notably, TRAIT emphasizes single-cell omics as a major reliable data source for TCR–antigen interactions and includes millions of reliable non-interactive TCRs. Additionally, it thoroughly demonstrates the interactions between mutations of TCRs and antigens, thereby benefiting affinity optimization of engineered TCRs as well as vaccine design. TCRs on clinical trials are innovatively provided. With the significant efforts made toward elucidating the complex interactions between TCRs and antigens, TRAIT is expected to ultimately contribute superior algorithms and substantial advancements in the field of T-cell-based immunotherapies. TRAIT is freely accessible at https://pgx.zju.edu.cn/traitdb.

Key words: T-cell receptor–antigen interaction; Antigen recognition; Mutation; Single-cell omics; T-cell-based immunotherapy.

Graphical abstract
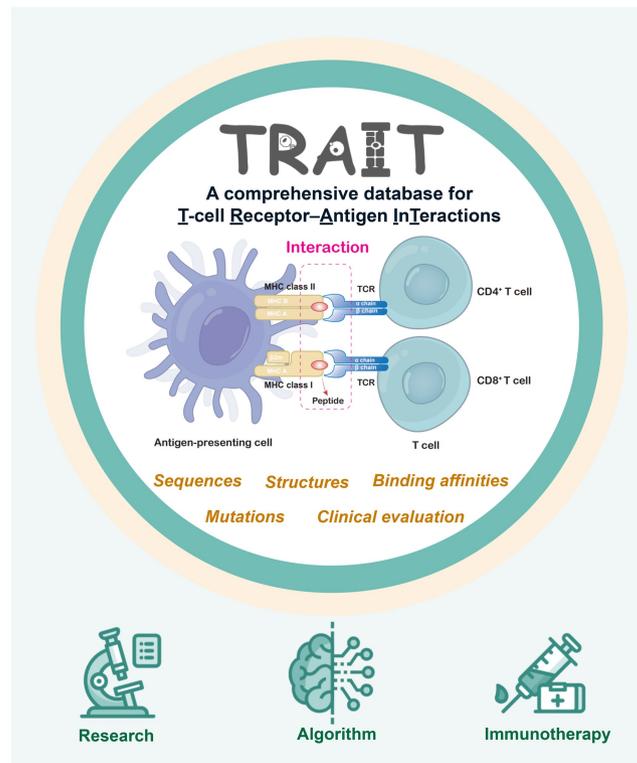
Introduction

Remarkable T-cell receptors (TCRs) play a vital role in initiating intracellular signals required for the effector response to foreign antigens in the adaptive immune system [1,2]. TCR provides the initial signal for T-cell activation by mediating the specific recognition of pathogen-derived epitopes through its interaction with the peptide–major histocompatibility complex (pMHC) multimer [3–5]. The underlying basis for the exceptional sensitivity and discriminatory recognition lies in the hypervariable complementarity-determining region 3 (CDR3) loops of TCRs, which interact mainly with the peptide, whereas the CDR1 and CDR2 regions contact the MHC α-helices [6,7]. The vast diversity of CDR3 regions and the cross-reactivity of TCRs ensure that the available TCR repertoire can recognize a wide range of pathogen-derived epitopes encountered over a lifetime. Nevertheless, the rules of how TCRs recognize cognate antigens (pMHCs) remain poorly understood [8,9]. Therefore, a precise and thorough characterization of the interactions between TCRs and antigens is important for understanding the molecular mechanisms of pathogen elimination and tumor surveillance mediated by adaptive immune responses.

Owing to their great importance, data including sequences, structures, and specificities of TCRs, TCR repertoires from different donors, and affinities of TCR–antigen pairs and their mutations have attracted extensive interest [10–12]. Furthermore, these data are essential for the fields of TCR engineering and the development of immunotherapies [13–15]. To date, several TCR-related databases have been developed and are currently active [16–26], and the majority of these databases focus on providing structural data of TCRs or TCR–pMHC complexes, such as the structural T-cell receptor database (STCRDab) [16], the T cell receptor structural repertoire database (TCR3d) [17], and the database of altered TCR ligand affinities and structures (ATLAS) [18]. Others are specialized in describing sequences of TCR repertoires, including the unconventional T cell receptor sequence database (UcTCRdb) [19], the pan immune repertoire database (PIRD) [20], and the human TCR sequence database (TCRdb) [21]. Additionally, several reputable databases have demonstrated a wealth of information on TCRs with known antigen specificity, such as the manually curated catalogue of pathology-associated T cell receptor sequences (McPAS-TCR) [22], the curated database of T-cell receptor sequences with known antigen specificity (VDJdb) [23,24], the immunoinformatic database of experimentally-supported functional neoantigen-specific TCR sequences (NeoTCR) [25], and the immune epitope database (IEDB) [26]. Although databases containing valuable TCR data are available, most existing ones concentrate on a specific aspect. Binding affinity and clinical evaluation of TCRs are not typically included in the majority of active databases (Table 1). There is noticeable gap when it comes to an integrated database that compiles a landscape of TCR–antigen interactions, highlighting the urgent need for a comprehensive database profiling TCR–antigen interactions.

In this study, we developed a novel knowledge base focused on the T-cell receptor–antigen interactions (TRAIT). First, an extensive literature review was conducted by performing keyword searches in PubMed, public datasets, and omics platform, resulting in a total of 3,393,826 antigen-specificity validated TCR sequences for 1184 unique epitopes and 112 MHC alleles. Uniquely, 3,342,225 non-interactive

Table 1 An overview of databases related to TCRs

Database	Sequence	Structure	Specificity	Binding affinity	Non-interactive pair	Mutation	Clinical evaluation	Active	Ref.
STCRDab		√						√	[16]
TCR3d		√						√	[17]
ATLAS		√	√			√		No longer updated after 2017	[18]
UcTCRdb	√							√	[19]
PIRD	√							√	[20]
TCRdb	√							√	[21]
McPAS-TCR	√		√					√	[22]
VDJdb	√	√	√					√	[23]
NeoTCR	√		√					√	[25]
IEDB	√		√					√	[26]
TRAIT	√	√	√	√	√	√	√	√	This study

Note: TCR, T-cell receptor; STCRDab, structural T-cell receptor database; TCR3d, T cell receptor structural repertoire database; ATLAS, database of altered TCR ligand affinities and structures; UcTCRdb, unconventional T cell receptor sequence database; PIRD, pan immune repertoire database; TCRdb, human TCR sequence database; McPAS-TCR, manually curated catalogue of pathology-associated T cell receptor sequences; VDJdb, curated database of T-cell receptor sequences with known antigen specificity; NeoTCR, immunoinformatic database of experimentally-supported functional neoantigen-specific TCR sequences; IEDB, immune epitope database; TRAIT, T-cell receptor–antigen interaction database.

TCR sequences verified by single-cell omics were collected and recorded. Second, structures of TCR–antigen complexes were systematically collected from the Protein Data Bank (PDB) [27], and a total of 223 structures were identified. Subsequently, valuable three-dimensional affinity data quantifying the binding ability of TCRs and antigens were manually retrieved, leading to 945 binding affinity records. Additionally, a unique dataset consisting of 641 TCR mutants and 628 antigen mutants was compiled, with each mutant linked to its respective binding affinity. Finally, TCRs with clinical evaluations were consulted exhaustively, and 34 clinical trials inferring 65 antigen-specific TCRs were identified.

In summary, TRAIT is developed to explicitly characterize the interactions between TCRs and antigens by integrating sequences, structures, and binding affinities. TRAIT also offers single-cell-based antigen-specific TCR repertoires and systematically demonstrates interactions between mutations of TCRs and antigens. Additionally, clinical evaluations of antigen-specific TCRs are uniquely provided. Given the recent prominence of T-cell-based immunotherapies in disease treatment, the comprehensive data on TCR–antigen interactions provided by TRAIT are anticipated to better facilitate the discovery and optimization of TCR candidates for immunotherapies.

Data collection and processing

Collection and processing of interaction data between TCRs and antigens

Based on literature evidence of TCR–antigen pairs including sequences, structures, and binding affinities, data were first collected by searching PubMed [28] for keywords with the combined terms "T cell receptor + repertoire", "T cell receptor + affinity", "T cell receptor + specificity", "T cell receptor + major histocompatibility complexes", and so on. Sequences of antigen-specific TCRs were also collected from public databases such as VDJdb [24], McPAS-TCR [22], and IEDB [26], as well as omics datasets available on the 10X website. The available raw data were downloaded and processed through IMGT/V-QUEST to produce annotated information for TCRs that lacked V and/or J specifications or had incomplete or overly extended CDR3 sequences. Structures of TCR–antigen complexes were further identified from PDB [27] with TCR–antigen reference sequences as queries. Only structures of TCR–MHC or TCR–pMHC complexes were collected and linked to their corresponding sequence entries in TRAIT. The binding affinities of TCR–antigen pairs experimentally determined using purified proteins by surface plasmon resonance (SPR) or isothermal titration calorimetry (ITC) were collected and recorded. Finally, the manually collected data of TCR–antigen pairs were further checked to remove duplicates.

Single-cell omics-based antigen-specific TCR identification

Datasets on the 10X website that describe TCRs extracted from pMHC multimer-labeled T cells and identified by single-cell sequencing were collected and recorded as "interactive pairs" in TRAIT. Simultaneously, TCRs extracted from non-binding T-cell populations after staining with a pMHC multimer and identified by single-cell sequencing were collected and recorded as "non-interactive pairs".

Mutation classification and binding affinity annotation

A systemic literature review was conducted using combinations of keywords such as "T cell receptors + mutations", "T cell receptors + variants", "T cell receptors + directed evolution", "T cell receptors + affinity maturation", "major histocompatibility complex + mutations", "major histocompatibility complex + variants", and so on. Mutations of TCRs or antigens with binding affinities measured experimentally were collected and recorded in TRAIT. All of the entries were divided into two categories including "mutations of TCRs" and "mutations of antigens" according to the sites of mutated amino acids. CDR sequences of TCRs were annotated through IMGT/V-QUEST by inputting their full-length sequences.

Collection of TCRs on clinical trials

A comprehensive literature review was conducted across multiple sources, including official patent organizations, various pharmaceutical company pipelines, and literature databases, using keywords "TCR-T", "TCR-transduced T cell", "TCR-engineered T cell", and so on. Detailed activity data of TCR

gene-engineered T cells (TCR-T) including objective response rates (ORRs), clinical response, and toxicities related to TCR-T in different stages of clinical trials were collected and displayed in TRAIT.

Database content and construction

As a result, nearly 3.4 million records of TCR–antigen pairs associated with cancer, autoimmune diseases, and infectious diseases were collected, and the statistical data are presented in Table 2. Large-scale TCR–antigen pairs including sequences (3,393,826 records), structures (223 records), and binding affinities (945 records) were manually retrieved. Multiple antigen species, including *Homo Sapiens* ($n = 3176$), Influenza A virus ($n = 9863$), Epstein-Barr virus (EBV, $n = 6227$), cytomegalovirus (CMV, $n = 22,296$), and human immunodeficiency virus (HIV, $n = 2663$), were involved, and a total of 112 MHC alleles were extracted. Ultimately, a total of 20,251 interactive TCRs and 3,342,225 non-interactive TCRs referring to 50 antigens from four different donors were collected and recorded in TRAIT, and a total of 1269 mutation records (including 641 TCR mutations referring to 41 wild-type antigens and 628 antigen mutations referring to 65 wild-type TCRs) were presented. Additionally, TRAIT innovatively compiles information on 34 TCR-T cell therapies that are in clinical trial stage and have published their clinical outcomes. A total of 65 TCR–antigen pairs (59 wild-type TCRs and 6 modified TCRs) and 25 cancer types were inferred.

TRAIT was built using a combination of hypertext preprocessor (PHP; https://www.php.net), Bootstrap (https://getbootstrap.com), and WordPress (https://wordpress.org) — a PHP-based blogging platform that enables users to create websites on servers supporting PHP and MySQL databases. The jQuery (https://jquery.com) and DataTables are employed to showcase the query result tables. All data in this database are stored and managed in MySQL. The overall workflow of TRAIT is depicted in Figure 1.

User interface

TRAIT has a user-friendly web interface containing seven pages: "Home", "Search", "Omics", "Mutations", "Therapeutics", "Download", and "Help" (Figure 2). The "Home" page provides an introduction to the entire database, an overview of collected data, and a visualization of total records. The "Search" page in TRAIT provides a comprehensive repository of TCR–antigen pairs. Users can query the database using a CDR3

sequence (nucleic acid or amino acid sequence), antigen information (epitope, parent gene, species, or MHC allele), interaction information (binding affinity or structure), or metadata (reference PMID, TCR sequencing method, and publication year). A built-in search box allows users to refine results to a specific subject by entering additional search terms. The retrieved results, which include brief descriptions of the TCR–antigen pairs, are presented in a new table. Clicking on "ID" in the first column opens a results page with more detailed interaction features, including inferred sequences, structures, and binding affinities for the matching term. TCR annotations [such as CDR3α, CDR3β, V(D)J recombination, species, and sequencing methods], antigen annotations (including peptide sequences, MHC allele, gene of epitope, and species of epitope), interaction information (such as binding affinity), and experimental details (such as identification and verification methods, samples, and pathology) are also presented. More importantly, the structures of the TCR–antigen complexes were linked to the interaction information described above.

The "Omics" page offers a query of TCR–antigen pairs, including interactive and non-interactive pairs verified by single-cell sequencing. Clicking on "#Pos" in the sixth column opens a separate page for the specific antigen, showcasing detailed information about interactive TCRs and visualizations of all TCRs involved with the antigen. Users can download non-interactive TCRs for certain antigens by clicking the button "#Neg". The "Mutations" page provides existing experimentally verified mutations of TCR–antigen pairs and their corresponding binding capacities. Clicking the TCR or antigen of interest opens a separate page for the list of all mutations and visualization of the affinity and sequence alignment. The "Therapeutics" page provides a detailed overview of the pharmaceutical information for each studied TCR in clinical status. Users may click on the "More Details" button to access more information, including sequences of TCR–antigen pairs, targeted disease indications, clinical trial status, and various clinical response data (*e.g.*, number of enrolled patients, administration dosage, ORRs, and toxicities related to the studied TCR-T cells). Users can download data and obtain tutorials from the "Download" and "Help" pages, respectively.

Database application
Systematic description of TCR–antigen interactions from multiple perspectives

The engagement of TCRs with antigens constitutes one of the most complex interactions in biological systems, further

Table 2 Statistics categorized by epitope species in TRAIT

Epitope species	No. of unique epitopes	No. of interactive TCRs	No. of non-interactive TCRs	No. of structures	No. of affinity records	No. of TCR mutations	No. of antigen mutations
Homo sapiens	251	3176	1,343,083	81	334	232	107
Influenza A	22	9863	65,924	16	27	23	4
EBV	30	6227	803,057	15	117	65	53
CMV	32	22,296	592,849	6	12	2	10
HIV	54	2663	268,685	23	59	28	31
HPV	1	14	134,337	0	0	0	0
HCV	15	1111	0	5	6	0	6
HTLV-1	8	176	67,144	9	137	114	24
Others	771	6075	67,146	68	253	177	393
Total	1184	51,601	3,342,225	223	945	641	628

Note: Data were collected until March 2025. EBV, Epstein-Barr virus; CMV, cytomegalovirus; HIV, human immunodeficiency virus; HPV, human papilloma virus; HCV, hepatitis C virus; HTLV-1, human T-lymphotropic virus type 1.
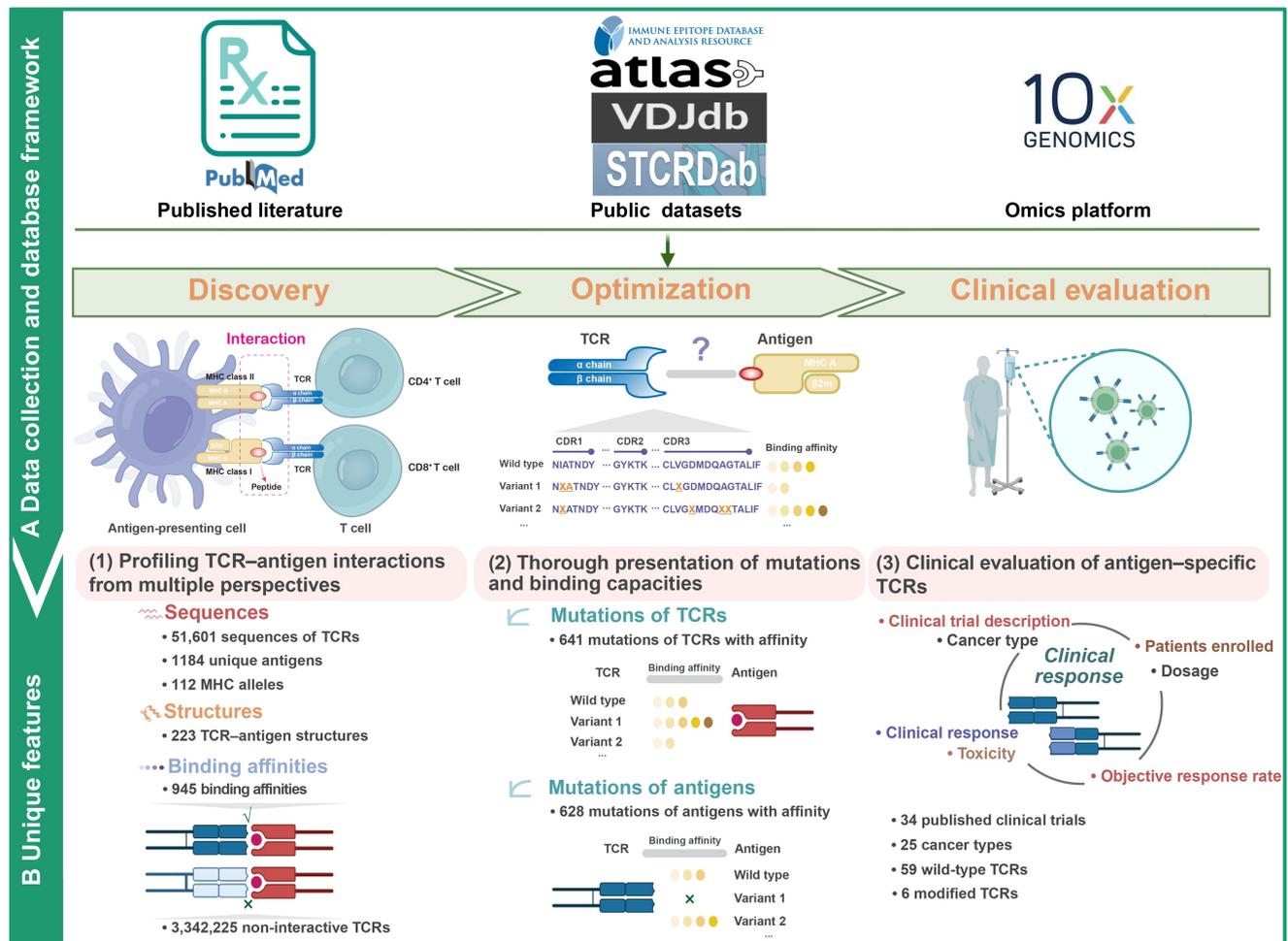
Figure 1 Overview of the TRAIT database
A. Data collection and database framework. Information of antigen-specific TCRs, including sequences, structures, binding affinities, mutations with binding capacities, and clinical response, was thoroughly collected from published literature, public databases, and 10X omics platform. And then data were reorganized by the whole process of "discovery–optimization–clinical evaluation" of antigen-specific TCRs. **B**. Unique features of TRAIT. TRAIT is unique in (1) unambiguously profiling the interactions between TCRs and antigens by integrating sequences, structures, and affinities, thus offering an exhaustively landscape of TCR–antigen interactions, (2) systematically describing mutation–affinity relationship of TCRs and antigens, benefiting affinity optimization of TCRs, and (3) thoroughly demonstrating clinical response of TCRs on clinical trials. The figure is created with BioRender.com. TCR, T-cell receptor; MHC, major histocompatibility complex; CDR1/2/3, complementarity-determining region 1/2/3; β2m, beta-2 microglobulin; TRAIT, T-cell receptor–antigen interaction database.

complicated by the intricate mechanisms of signaling and selection [29–32]. Various experimental approaches have attempted to assess the importance of TCR–antigen interactions and correlate corresponding affinities with the level of T-cell activity mediated by antigens [10,33–36]. TRAIT presents a detailed panorama of specificity and affinity, focusing on their interplay with the structural aspects of TCRs and the interfaces that they form with pMHC. As shown in Figure 2A, records of interactive TCR–antigen pairs describing antigen-specific TCR sequences are the primary source of TRAIT. A detailed annotation of the interactive pairs, including sequences of CDR3 regions, germline gene usage, sequencing methods of TCR, MHC alleles, epitope species, epitope gene of antigen, and pathologic conditions, was extracted from experimental publications. Taking mutant Kirsten rat sarcoma viral oncogene homolog (KRAS) protein as an example, TRAIT offered comprehensive knowledge of KRAS epitopes and cognate TCRs (Table S1). A total of 35 interactive entries with 6 structures and 6 affinity values were retrieved after searching with the item "Parent gene: *KRAS*" on the "Search" page, which helps with a comprehensive understanding of KRAS and rat sarcoma (RAS) drug discovery. We concluded by emphasizing how the relationships between specificity, affinity, and structure might be exploited to help provide a deeper understanding of the features of antigen recognition in T-cell immunity.

Single-cell omics-based profiling of interactions between TCRs and antigens

Single-cell sequencing provides a powerful technique to ascertain the TCR sequences of individual T cells [37–41]. As shown in Figure 2B, TRAIT offers an omnidirectional single-cell omics map of TCRs focused on certain antigens. Detailed information on TCR–antigen pairs, including sequences of CDR3 regions and germline gene usage, is shown in a separate table. A comprehensive illustration of CDR3 sequence motifs, including interactive and non-interactive pairs, is provided and can be interactively accessed online. Highly reliable
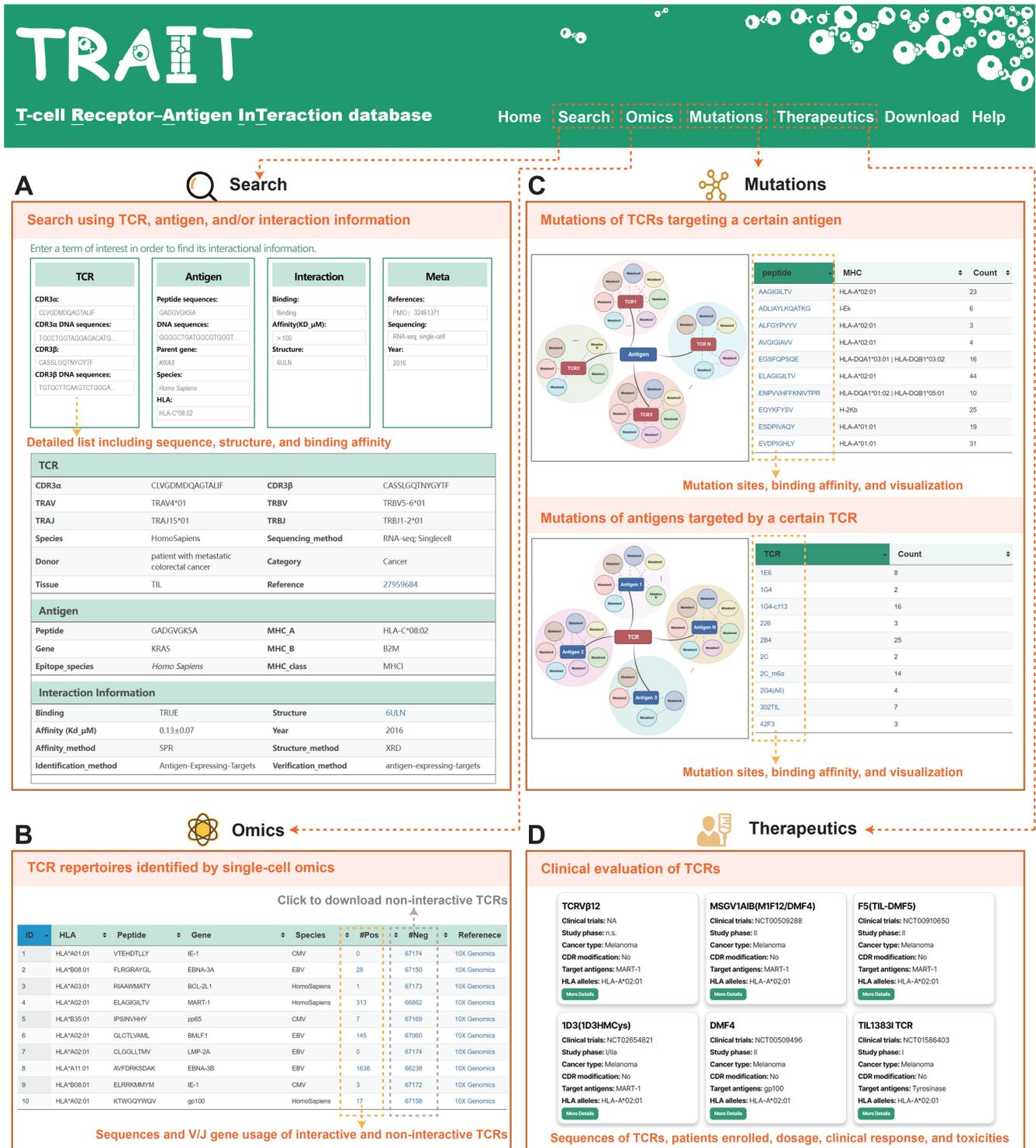
Figure 2 Schematic illustrations of the four primary functional pages in TRAIT
A. The "Search" page of TRAIT. Users can browse the detailed information including sequences, structures, and binding affinities of TCR–antigen pairs. **B**. The "Omics" page of TRAIT. Validated antigen-specific TCRs including interactive and non-interactive TCRs by single-cell sequencing are displayed on the page. **C**. The "Mutations" page of TRAIT. Mutations of TCRs and antigens were classified and linked to their binding affinities on the page. Users can visualize changes in binding affinities between wild type and mutants. **D**. The "Therapeutics" page of TRAIT. The page offers detailed clinical information of TCRs on clinical trials.

and enormous TCR–antigen pairs with both interactive and non-interactive pairs will further facilitate the development of bioinformatics algorithms [42–45] and enable sample–sample comparisons within different donors and cancer types. By incorporating non-interactive pairs, the binding characteristics of TCR–antigen pairs are illuminated in a more comprehensive perspective.

Taking the peptide "GILGFVFTL" presented by HLA-A*02:01 as an example, it has 1368 entries of interactive and 65,924 non-interactive TCR records from four different

A

V-J gene usage of α chain

V-J gene usage of β chain

B

Interactive TCRs

Non-interactive TCRs
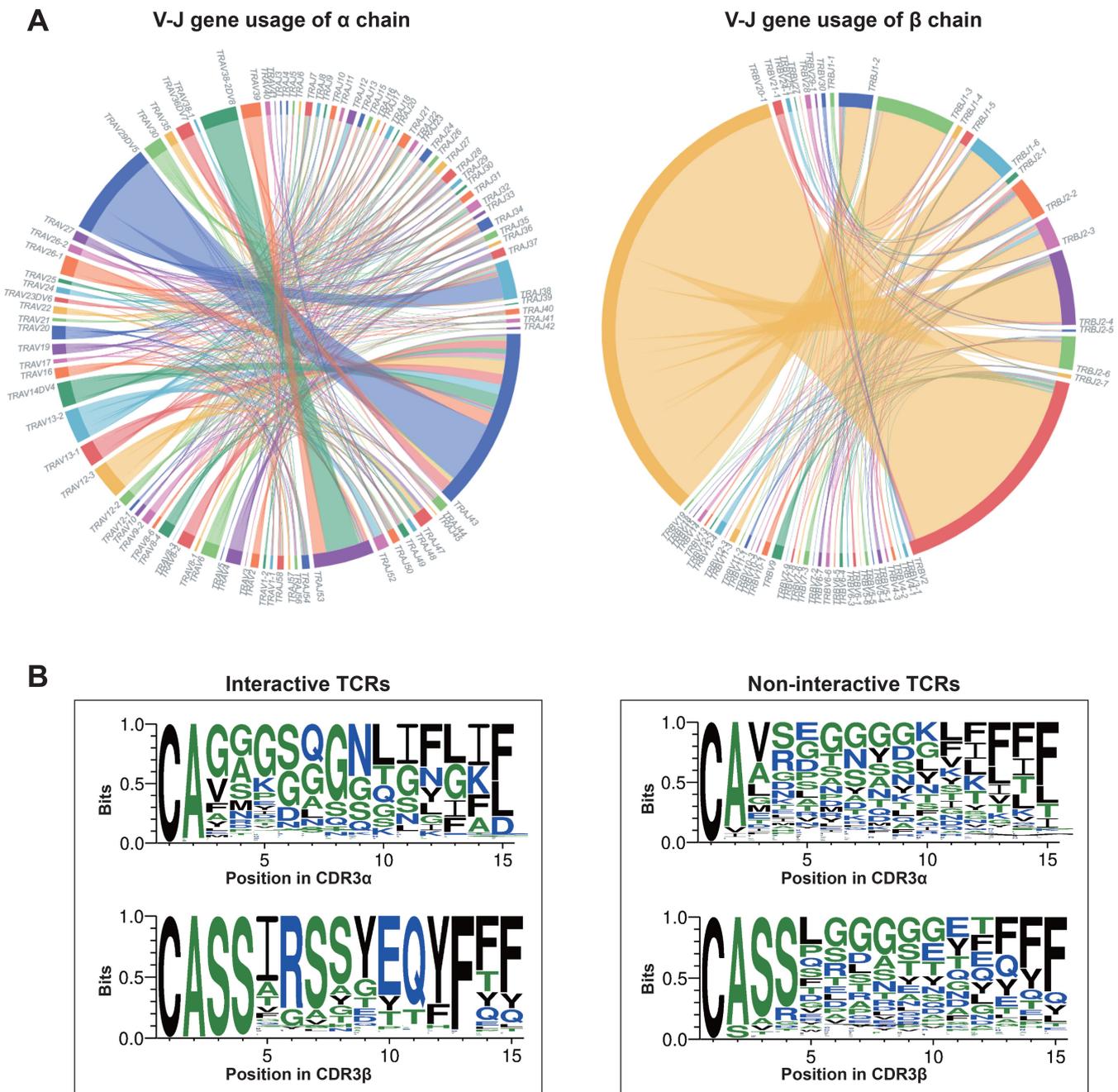
**Figure 3 Features of TCRs related to GILGFVFTL–HLA-A*02:01 on the "Omics" page**

**A**. V-J gene usage of interactive TCRs. Genes with detection frequency > 5 are used for visualization. **B**. Comparison of the CDR3 motifs between interactive and non-interactive TCRs. The visualized figures for V-J genes on the web pages were generated using the D3.js tool. Sequence logos depicting interactive and non-interactive TCR–antigen pairs were generated using the WebLogo 3 tool.

donors. As shown in Figure 3, the amino acid residues of CDR3 are relatively different between the interactive and non-interactive TCRs related to GILGFVFTL–HLA-A*02:01. CDR3 loops of most TCRs start with cysteine and alanine, and end with phenylalanine in either interactive or non-interactive CDR3 motifs. Despite the subtle similarities, notable differences between interactive and non-interactive CDR3 motifs are observed. As shown in CDR3α of interactive TCRs, glycine and asparagine are preferred separately at positions 8 and 9, and phenylalanine shows biases at position 12. Preferences of amino acids in CDR3β are more significant. Interestingly, no significant amino acid preferences are

observed for CDR3α and CDR3β for non-interactive TCRs except for the subtle enrichment of phenylalanine at positions 13 and 14.

Thorough investigation of interactions between TCR mutations and cognate antigens

The manufacturing of high-affinity TCRs has been a core strategy to improve the efficacy of adoptive cell transfer therapies [46–50]. A panorama of affinities for TCR mutations and antigens was collected and systematically described in TRAIT (Figure 2C). All antigen-specific TCR mutants that have been experimentally verified are linked to measured

A Mutants of TCR "1G4"
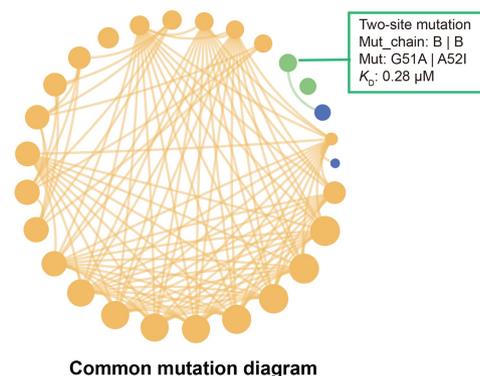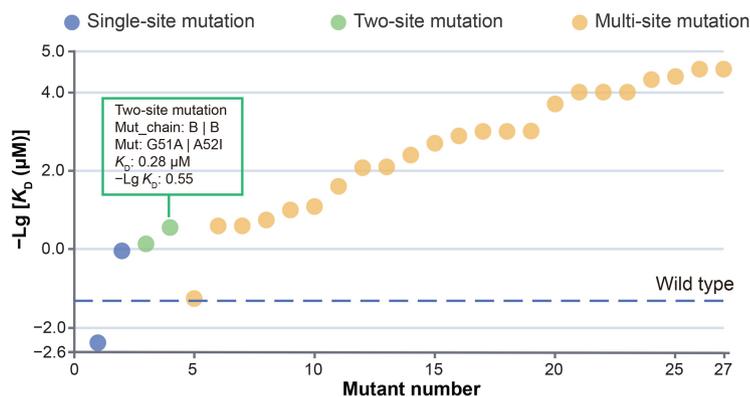
B Mutants of TCR "BC1"
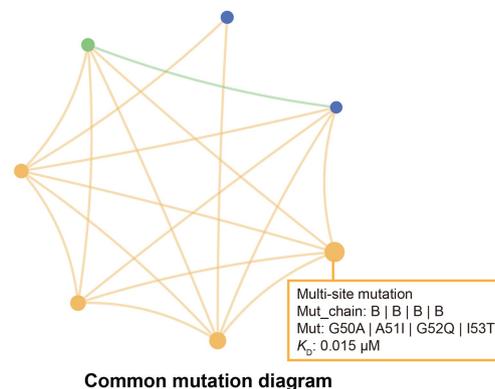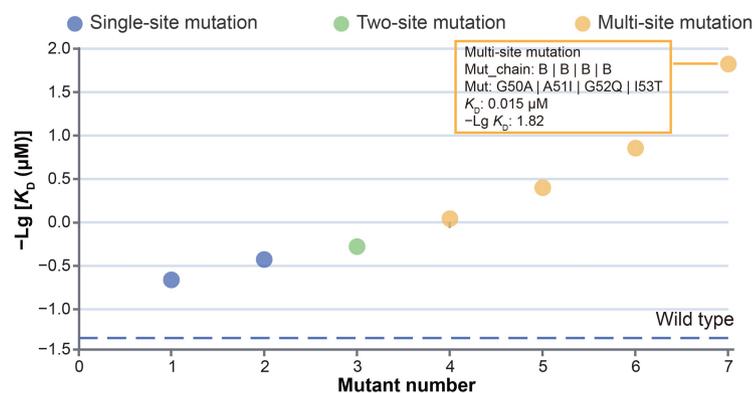
C Mutants of antigen "ANERADLIAYLKQATK−I-Ek"
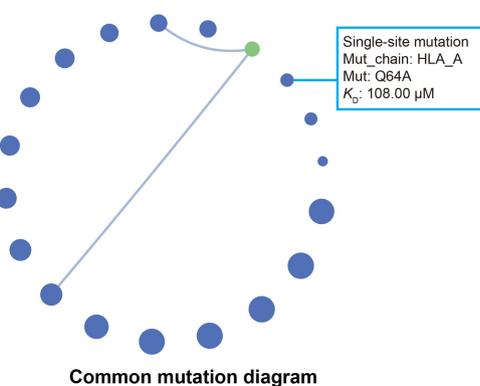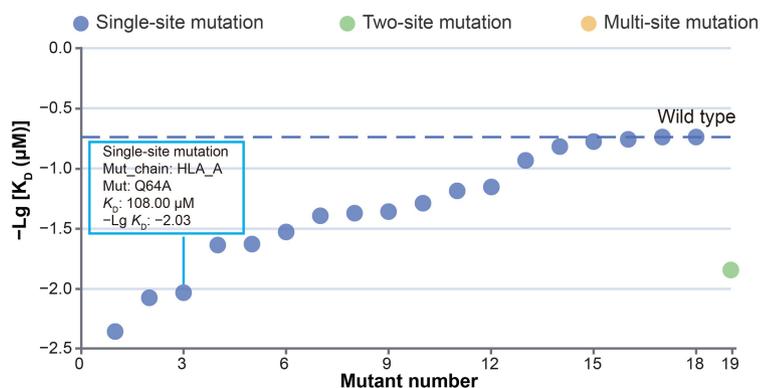
Figure 4 Description and visualization of binding affinities between mutations of TCR and antigens
A. Scatter plot and circular plot illustrating interactive capacities between mutations of TCR "1G4" targeting "SLLMWITQC–HLA-A*02:01". **B**. Scatter plot and circular plot illustrating interactive capacities between mutations of TCR "BC1" targeting "SLLMWITQC–HLA-A*02:01". **C**. Scatter plot and circular plot illustrating interactive capacities between antigen mutations and cognate TCR "2B4". Mutants with shared amino acid mutation sites were visualized by linking them using lines. Larger diameters of a mutant indicate higher affinities between TCRs and their corresponding antigens. ECharts was used to visualize the relationship between mutations and affinities in online TRAIT.

affinities, and detailed information on mutations can be further consulted. Users may click SLLMWITQC–HLA-A*02:01 as an example, and 51 entries including 42 mutation records of TCR "1G4" and 9 mutation records of TCR "BC1", are provided in the resulting list (Table S2). Mutants of 1G4 had affinities ranging from 0.026 nM to > 240 μM in four publications, while mutants of BC1 had an affinity of 0.015–21.4 μM with SLLMWITQC–HLA-A*02:01. Amino acid mutations in CDR regions were bolded and marked in red for easy track. Additionally, TRAIT enabled the

visualization of the entire panorama through a scatter plot and a circular plot between mutations of TCRs and corresponding affinities with SLLMWITQC–HLA-A*02:01 (Figure 4A and B). Different types of mutations (including single-site mutations, two-site mutations, and multi-site mutations) are indicated in the diagram using different colored circles, with the detailed amino acid residues interactively viewed by placing the mouse on any of the circles. Consequently, mutations of TCRs with enhanced or weakened binding capacities were intuitively viewed. In summary,

TRAIT is unique in providing a panorama that enables a full visualization of the mutation affinities of TCRs and is valuable for the manufacturing of high-affinity TCR profiles for research interest. Moreover, mammalian cells [51,52], yeast [53,54], and bacteriophage display [55,56] represent additional powerful, high-throughput methods capable of creating TCRs with a wide range of binding affinities, thereby providing a continuous data source for TRAIT.

Intensive explorations of interactions between mutations of antigens and corresponding TCRs

Investigating interactions between antigens and particular TCRs will lead to a more profound understanding on immunogenicity of antigens and cross-reactivity of TCRs [57,58]. TRAIT provides the binding capacities of specific TCRs with antigens or antigen mutants. As shown in Figure 2C, mutations of antigens with the capacity interacting with TCRs are summarized in the table. Detailed information can be further queried by clicking the name of TCRs. A scatter plot and a circular plot between mutants of antigens and binding capacities with certain TCRs were drawn to directly show changes in affinities to different antigens for the same TCR. Users may click 2B4 as an example, and 25 entries including 3 wild-type records, 16 mutation records of human leukocyte antigen (HLA), and 6 mutation records of peptides, are displayed in the search list. The scatter plot and circular plot between mutants of antigens and affinities with 2B4 indicated that neither a mutant of HLA nor peptide can result in higher affinities with TCR 2B4 (Figure 4C). Linking binding affinities to mutants of antigens enables us to delve deeper into the interactions between antigens and TCRs and help us trace peptides that can better stimulate T-cell immunity and promote vaccine development.

Clinical usage

TCR-driven therapies are introducing a new era in precision oncology treatments. As a promising therapeutic alternative, TCR-T cells can target epitopes derived from both surface and intracellular for tumor recognition [59]. Natural (wild-type) TCRs identified from tumor-infiltrating lymphocytes, circulating T cells in patients, healthy donors, or humanized mice are the main sources of TCR candidates [60]. Nonetheless, natural TCRs typically exhibit lower affinities due to thymic selection, which restricts their capacity to recognize antigens with low densities on tumor cells. To address this issue, a common strategy is to enhance the TCR's affinity for its target [61]. Several TCRs without CDR modification have demonstrated encouraging outcomes in clinical studies for various solid tumors, as shown detailed in Table S3. Notably, an affinity-enhanced TCR-1G4 (targeted NY-ESO-1) achieved a 44% ORR among patients with metastatic melanoma [62–64], a 45%–67% ORR in individuals with metastatic synovial sarcoma, and a 80% ORR in patients with myeloma without apparent side effects [61]. Another phase 1 clinical trial indicated that TCR-T cells engineered for high affinity to MAGE-A3 antigen led to a 56% ORR in individuals with melanoma, synovial sarcoma, and esophageal cancer [65]. In this context, the information on both natural and modified TCRs provided in TRAIT is useful for the discovery of therapeutic TCR candidates for TCR-T therapies and other immunotherapeutic approaches.

Discussion and conclusion

In this study, TRAIT is developed to provide a comprehensive TCR–antigen interaction landscape. It is unique in systematically providing TCR–antigen pairs by integrating sequences, structures, binding affinities, mutations, and clinical evaluations. More importantly, a correlation between affinity and amino acid residues is initially introduced in TRAIT to more deeply explore the interaction rules and characteristics of TCRs and antigens. Moreover, it supplies a versatile visualization platform for TCR–antigen interactions, along with an intuitive web interface for convenient data retrieval. Consequently, TRAIT will serve as an invaluable resource for researchers in both experimental and computational biologists when delving into T-cell immunology and translational medicine of immunotherapies.

Immunotherapies involving the interaction of TCRs and antigens have emerged as pivotal therapeutic strategies in oncology, with an exponentially increase expected in the future for novel TCR–antigen pairs derived from patients with different diseases [66–68]. In this regard, TRAIT will be regularly updated to provide timely data and improved web pages. Furthermore, additional annotation information encompassing detailed corresponding diseases and a focus on neoantigens, along with practical analysis tools, will be incorporated to keep pace with the rapidly evolving research landscape. These enhancements are anticipated to significantly augment the application's efficiency within the research community. Overall, TRAIT is poised to serve as a valuable resource facilitating a comprehensive understanding of molecular mechanisms underlying antigen recognition in T-cell immunology, thereby holding immense implications for the future advancements in immunotherapeutic interventions.

Data availability

TRAIT is freely available at https://pgx.zju.edu.cn/traitdb to all users without any login or registration restrictions. TRAIT has been submitted to Database Commons [69] at the National Genomics Data Center (NGDC), China National Center for Bioinformation (CNCB), which is publicly accessible at https://ngdc.cncb.ac.cn/databasecommons/database/id/9707.

CRediT author statement

Mengmeng Wei: Methodology, Data curation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing. **Jingcheng Wu:** Methodology, Data curation, Software, Investigation. **Shengzuo Bai:** Software, Investigation, Visualization, Writing – review & editing. **Yuxuan Zhou:** Data curation, Investigation. **Yichang Chen:** Investigation, Visualization. **Xue Zhang:** Visualization. **Wenyi Zhao:** Methodology, Visualization. **Ying Chi:** Investigation. **Gang Pan:** Supervision. **Feng Zhu:** Conceptualization, Methodology. **Shuqing Chen:** Conceptualization, Funding acquisition. **Zhan Zhou:** Conceptualization, Funding acquisition, Supervision, Writing – review & editing. All authors have read and approved the final manuscript.

Competing interests

The authors have declared no competing interests.

Supplementary material

Acknowledgments

ORCID

0009-0000-2166-4373 (Mengmeng Wei)
0000-0003-0194-4380 (Jingcheng Wu)
0009-0006-1503-4707 (Shengzuo Bai)
0000-0002-0643-151X (Yuxuan Zhou)
0000-0003-1697-3256 (Yichang Chen)
0000-0003-0152-1708 (Xue Zhang)
0000-0003-1834-0348 (Wenyi Zhao)
0000-0002-3020-7892 (Ying Chi)
0000-0002-4049-6181 (Gang Pan)
0000-0001-8069-0053 (Feng Zhu)
0000-0002-0792-3735 (Shuqing Chen)
0000-0002-2730-5483 (Zhan Zhou)

References

[1] van der Merwe PA, Davis SJ. Molecular interactions mediating T cell antigen recognition. Annu Rev Immunol 2003;21:659–84.

[2] Davis MM, Krogsgaard M, Huppa JB, Sumen C, Purbhoo MA, Irvine DJ, et al. Dynamics of cell surface molecules during T cell recognition. Annu Rev Biochem 2003;72:717–42.

[3] Brameshuber M, Kellner F, Rossboth BK, Ta H, Alge K, Sevcsik E, et al. Monomeric TCRs drive T cell antigen recognition. Nat Immunol 2018;19:487–96.

[4] Cai E, Marchuk K, Beemiller P, Beppler C, Rubashkin MG, Weaver VM, et al. Visualizing dynamic microvillar search and stabilization during ligand detection by T cells. Science 2017;356:eaal3118.

[5] Rossboth B, Arnold AM, Ta H, Platzer R, Kellner F, Huppa JB, et al. TCRs are randomly distributed on the plasma membrane of resting antigen-experienced T cells. Nat Immunol 2018;19:821–7.

[6] Borg NA, Ely LK, Beddoe T, Macdonald WA, Reid HH, Clements CS, et al. The CDR3 regions of an immunodominant T cell receptor dictate the "energetic landscape" of peptide–MHC recognition. Nat Immunol 2005;6:171–80.

[7] Rudolph MG, Wilson IA. The specificity of TCR/pMHC interaction. Curr Opin Immunol 2002;14:52–65.

[8] Wu P, Zhang T, Liu B, Fei P, Cui L, Qin R, et al. Mechano-regulation of peptide–MHC class I conformations determines TCR antigen recognition. Mol Cell 2019;73:1015–27.e7.

[9] Brown SD, Raeburn LA, Holt RA. Profiling tissue-resident T cell repertoires by RNA sequencing. Genome Med 2015;7:125.

[10] Rossjohn J, Gras S, Miles JJ, Turner SJ, Godfrey DI, McCluskey J. T cell antigen receptor recognition of antigen-presenting molecules. Annu Rev Immunol 2015;33:169–200.

[11] Siller-Farfán JA, Dushek O. Molecular mechanisms of T cell sensitivity to antigen. Immunol Rev 2018;285:194–205.

[12] Spear TT, Evavold BD, Baker BM, Nishimura MI. Understanding TCR affinity, antigen specificity, and cross-reactivity to improve TCR gene-modified T cells for cancer immunotherapy. Cancer Immunol Immunother 2019;68:1881–9.

[13] Greenbaum U, Dumbrava EI, Biter AB, Haymaker CL, Hong DS. Engineered T-cell receptor T cells for cancer immunotherapy. Cancer Immunol Res 2021;9:1252–61.

[14] Omer A, Peres A, Rodriguez OL, Watson CT, Lees W, Polak P, et al. T cell receptor beta germline variability is revealed by inference from repertoire data. Genome Med 2022;14:2.

[15] Zhao X, Kolawole EM, Chan W, Feng Y, Yang X, Gee MH, et al. Tuning T cell receptor sensitivity through catch bond engineering. Science 2022;376:eabl5282.

[16] Leem J, de Oliveira SHP, Krawczyk K, Deane CM. STCRDab: the structural T-cell receptor database. Nucleic Acids Res 2018;46:D406–12.

[17] Gowthaman R, Pierce BG. TCR3d: the T cell receptor structural repertoire database. Bioinformatics 2019;35:5323–5.

[18] Borrman T, Cimons J, Cosiano M, Purcaro M, Pierce BG, Baker BM, et al. ATLAS: a database linking binding affinities with structures for wild-type and mutant TCR–pMHC complexes. Proteins 2017;85:908–16.

[19] Dou Y, Shan S, Zhang J. UcTCRdb: an unconventional T cell receptor sequence database with online analysis functions. Front Immunol 2023;14:1158295.

[20] Zhang W, Wang L, Liu K, Wei X, Yang K, Du W, et al. PIRD: Pan Immune Repertoire Database. Bioinformatics 2020;36:897–903.

[21] Chen SY, Yue T, Lei Q, Guo AY. TCRdb: a comprehensive database for T-cell receptor sequences with powerful search function. Nucleic Acids Res 2021;49:D468–74.

[22] Tickotsky N, Sagiv T, Prilusky J, Shifrut E, Friedman N. McPAS-TCR: a manually curated catalogue of pathology-associated T cell receptor sequences. Bioinformatics 2017;33:2924–9.

[23] Shugay M, Bagaev DV, Zvyagin IV, Vroomans RM, Crawford JC, Dolton G, et al. VDJdb: a curated database of T-cell receptor sequences with known antigen specificity. Nucleic Acids Res 2018;46:D419–27.

[24] Bagaev DV, Vroomans RMA, Samir J, Stervbo U, Rius C, Dolton G, et al. VDJdb in 2019: database extension, new analysis infrastructure and a T-cell receptor motif compendium. Nucleic Acids Res 2020;48:D1057–62.

[25] Zhou W, Xiang W, Yu J, Ruan Z, Pan Y, Wang K, et al. NeoTCR: an immunoinformatic database of experimentally-supported functional neoantigen-specific TCR sequences. Genomics Proteomics Bioinformatics 2025;23:qzae010.

[26] Vita R, Mahajan S, Overton JA, Dhanda SK, Martini S, Cantrell JR, et al. The Immune Epitope Database (IEDB): 2018 update. Nucleic Acids Res 2019;47:D339–43.

[27] Rose PW, Beran B, Bi C, Bluhm WF, Dimitropoulos D, Goodsell DS, et al. The RCSB Protein Data Bank: redesigned web site and web services. Nucleic Acids Res 2011;39:D392–401.

[28] Sayers EW, Beck J, Brister JR, Bolton EE, Canese K, Comeau DC, et al. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 2020;48:D9–16.

[29] Riley TP, Baker BM. The intersection of affinity and specificity in the development and optimization of T cell receptor based therapeutics. Semin Cell Dev Biol 2018;84:30–41.

[30] Davis MM, Bjorkman PJ. T-cell antigen receptor genes and T-cell recognition. Nature 1988;334:395–402.

[31] Kageyama S, Tsomides TJ, Sykulev Y, Eisen HN. Variations in the number of peptide-MHC class I complexes required to activate cytotoxic T cell responses. J Immunol 1995;154:567–76.

[32] Holland CJ, Crean RM, Pentier JM, de Wet B, Lloyd A, Srikannathasan V, et al. Specificity of bispecific T cell receptors and antibodies targeting peptide–HLA. J Clin Invest 2020; 130:2673–88.

[33] Baker BM, Scott DR, Blevins SJ, Hawse WF. Structural and dynamic control of T-cell receptor specificity, cross-reactivity, and binding mechanism. Immunol Rev 2012;250:10–31.

[34] Wang Y, Singh NK, Spear TT, Hellman LM, Piepenbrink KH, McMahan RH, et al. How an alloreactive T-cell receptor achieves peptide and MHC specificity. Proc Natl Acad Sci U S A 2017; 114:E4792–801.

[35] Kula T, Dezfulian MH, Wang CI, Abdelfattah NS, Hartman ZC, Wucherpfennig KW, et al. T-Scan: a genome-wide method for the systematic discovery of T cell epitopes. Cell 2019;178: 1016–28.e13.

[36] Lee MN, Meyerson M. Antigen identification for HLA class I- and HLA class II-restricted T cell receptors using cytokine-capturing antigen-presenting cells. Sci Immunol 2021; 6:eabf4001.

[37] Zheng C, Zheng L, Yoo JK, Guo H, Zhang Y, Guo X, et al. Landscape of infiltrating T cells in liver cancer revealed by single-cell sequencing. Cell 2017;169:1342–56.e16.

[38] Gate D, Saligrama N, Leventhal O, Yang AC, Unger MS, Middeldorp J, et al. Clonally expanded CD8 T cells patrol the cerebrospinal fluid in Alzheimer's disease. Nature 2020; 577:399–404.

[39] Mogilenko DA, Shpynov O, Andhey PS, Arthur L, Swain A, Esaulova E, et al. Comprehensive profiling of an aging immune system reveals clonal GZMK$^+$ CD8$^+$ T cells as conserved hallmark of inflammaging. Immunity 2021;54:99–115.e12.

[40] Zhang SQ, Ma KY, Schonnesen AA, Zhang M, He C, Sun E, et al. High-throughput determination of the antigen specificities of T cell receptors in single cells. Nat Biotechnol 2018;36:1156–9.

[41] Wang AZ, Bowman-Kirigin JA, Desai R, Kang LI, Patel PR, Patel B, et al. Single-cell profiling of human dura and meningioma reveals cellular meningeal landscape and insights into meningioma immune response. Genome Med 2022;14:49.

[42] Lu T, Zhang Z, Zhu J, Wang Y, Jiang P, Xiao X, et al. Deep learning-based prediction of the T cell receptor–antigen binding specificity. Nat Mach Intell 2021;3:864–75.

[43] Xu Z, Luo M, Lin W, Xue G, Wang P, Jin X, et al. DLpTCR: an ensemble deep learning framework for predicting immunogenic peptide recognized by T cell receptor. Brief Bioinform 2021; 22:bbab335.

[44] Fischer DS, Wu Y, Schubert B, Theis FJ. Predicting antigen specificity of single T cells based on TCR CDR3 regions. Mol Syst Biol 2020;16:e9416.

[45] Gao Y, Gao Y, Fan Y, Zhu C, Wei Z, Zhou C, et al. Pan-peptide meta learning for T-cell receptor–antigen binding recognition. Nat Mach Intell 2023;5:236–49.

[46] Hebeisen M, Allard M, Gannon PO, Schmidt J, Speiser DE, Rufer N. Identifying individual T cell receptors of optimal avidity for tumor antigens. Front Immunol 2015;6:582.

[47] Hong DS, Van Tine BA, Biswas S, McAlpine C, Johnson ML, Olszanski AJ, et al. Autologous T cell therapy for MAGE-A4$^+$ solid cancers in HLA-A*02$^+$ patients: a phase 1 trial. Nat Med 2023;29:104–14.

[48] Schmidt J, Chiffelle J, Perez MAS, Magnin M, Bobisse S, Arnaud M, et al. Neoantigen-specific CD8 T cells with high structural avidity preferentially reside in and eliminate tumors. Nat Commun 2023;14:3188.

[49] Sim MJW, Lu J, Spencer M, Hopkins F, Tran E, Rosenberg SA, et al. High-affinity oligoclonal TCRs define effective adoptive T cell therapy targeting mutant KRAS-G12D. Proc Natl Acad Sci U S A 2020;117:12826–35.

[50] Sanderson JP, Crowley DJ, Wiedermann GE, Quinn LL, Crossland KL, Tunbridge HM, et al. Preclinical evaluation of an affinity-enhanced MAGE-A4-specific T-cell receptor for adoptive T-cell therapy. Oncoimmunology 2020;9:1682381.

[51] Ohta R, Demachi-Okamura A, Akatsuka Y, Fujiwara H, Kuzushima K. Improving TCR affinity on 293T cells. J Immunol Methods 2019;466:1–8.

[52] Dilchert J, Hofmann M, Unverdorben F, Kontermann R, Bunk S. Mammalian display platform for the maturation of bispecific TCR-based molecules. Antibodies (Basel) 2022;11:34.

[53] Sharma P, Kranz DM. Subtle changes at the variable domain interface of the T-cell receptor can strongly increase affinity. J Biol Chem 2018;293:1820–34.

[54] Yang X, Garner LI, Zvyagin IV, Paley MA, Komech EA, Jude KM, et al. Autoimmunity-associated T cell receptors recognize HLA-B*27-bound peptides. Nature 2022;612:771–7.

[55] Li Y, Moysey R, Molloy PE, Vuidepot AL, Mahon T, Baston E, et al. Directed evolution of human T-cell receptors with picomolar affinities by phage display. Nat Biotechnol 2005;23:349–54.

[56] Ch'ng ACW, Lam P, Alassiri M, Lim TS. Application of phage display for T-cell receptor discovery. Biotechnol Adv 2022; 54:107870.

[57] Wells DK, van Buuren MM, Dang KK, Hubbard-Lucey VM, Sheehan KCF, Campbell KM, et al. Key parameters of tumor epitope immunogenicity revealed through a consortium approach improve neoantigen prediction. Cell 2020;183:818–34.e13.

[58] Morazán-Fernández D, Mora J, Molina-Mora JA. *In silico* pipeline to identify tumor-specific antigens for cancer immunotherapy using exome sequencing data. Phenomics 2023;3:130–7.

[59] Baulu E, Gardet C, Chuvin N, Depil S. TCR-engineered T cell therapy in solid tumors: state of the art and perspectives. Sci Adv 2023;9:eadf3700.

[60] Klebanoff CA, Chandran SS, Baker BM, Quezada SA, Ribas A. T cell receptor therapeutics: immunological targeting of the intracellular cancer proteome. Nat Rev Drug Discov 2023; 22:996–1017.

[61] Rapoport AP, Stadtmauer EA, Binder-Scholl GK, Goloubeva O, Vogl DT, Lacey SF, et al. NY-ESO-1-specific TCR-engineered T cells mediate sustained antigen-specific antitumor effects in myeloma. Nat Med 2015;21:914–21.

[62] Robbins PF, Morgan RA, Feldman SA, Yang JC, Sherry RM, Dudley ME, et al. Tumor regression in patients with metastatic synovial cell sarcoma and melanoma using genetically engineered lymphocytes reactive with NY-ESO-1. J Clin Oncol 2011; 29:917–24.

[63] Robbins PF, Kassim SH, Tran TLN, Crystal JS, Morgan RA, Feldman SA, et al. A pilot trial using lymphocytes genetically engineered with an NY-ESO-1-reactive T-cell receptor: long-term follow-up and correlates with response. Clin Cancer Res 2015; 21:1019–27.

[64] D'Angelo SP, Melchiori L, Merchant MS, Bernstein D, Glod J, Kaplan R, et al. Antitumor activity associated with prolonged persistence of adoptively transferred NY-ESO-1^{c259}T cells in synovial sarcoma. Cancer Discov 2018;8:944–57.

[65] Morgan RA, Chinnasamy N, Abate-Daga D, Gros A, Robbins PF, Zheng Z, et al. Cancer regression and neurological toxicity following anti-MAGE-A3 TCR gene therapy. J Immunother 2013;36:133–51.

[66] Norberg SM, Hinrichs CS. Engineered T cell therapy for viral and non-viral epithelial cancers. Cancer Cell 2023;41:58–69.

[67] Puig-Saus C, Sennino B, Peng S, Wang CL, Pan Z, Yuen B, et al. Neoantigen-targeted CD8$^+$ T cell responses with PD-1 blockade therapy. Nature 2023;615:697–704.

[68] Anwaier A, Zhu SX, Tian X, Xu WH, Wang Y, Palihati M, et al. Large-scale proteomics data reveal integrated prognosis-related protein signatures and role of SMAD4 and RAD50 in prognosis and immune infiltrations of prostate cancer microenvironment. Phenomics 2022;2:404–18.

[69] Ma L, Zou D, Liu L, Shireen H, Abbasi AA, Bateman A, et al. Database Commons: a catalog of worldwide biological databases. Genomics Proteomics Bioinformatics 2023;21:1054–8.