



OPEN ACCESS

EDITED BY

Jerome Le Nours,
Monash University, Australia

REVIEWED BY

Michael Nicholas Tod Souter,
Peter Doherty Institute for Infection and
Immunity, Australia
Olivier Gasser,
Victoria University of Wellington,
New Zealand

*CORRESPONDENCE

Jian Zhang
✉ jian_zhang@tju.edu.cn

†These authors have contributed equally to
this work

SPECIALTY SECTION

This article was submitted to
T Cell Biology,
a section of the journal
Frontiers in Immunology

RECEIVED 03 February 2023

ACCEPTED 02 March 2023

PUBLISHED 13 March 2023

CITATION

Dou Y, Shan S and Zhang J (2023)
UcTCRdb: An unconventional
T cell receptor sequence database
with online analysis functions.
Front. Immunol. 14:1158295.
doi: 10.3389/fimmu.2023.1158295

COPYRIGHT

© 2023 Dou, Shan and Zhang. This is an
open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that
the original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

UcTCRdb: An unconventional T cell receptor sequence database with online analysis functions

Yunsheng Dou[†], Shiwen Shan[†] and Jian Zhang^{*}

Academy of Medical Engineering and Translational Medicine, Tianjin University, Tianjin, China

Unlike conventional major histocompatibility complex (MHC) class I and II molecules reactive T cells, the unconventional T cell subpopulations recognize various non-polymorphic antigen-presenting molecules and are typically characterized by simplified patterns of T cell receptors (TCRs), rapid effector responses and 'public' antigen specificities. Dissecting the recognition patterns of the non-MHC antigens by unconventional TCRs can help us further our understanding of the unconventional T cell immunity. The small size and irregularities of the released unconventional TCR sequences are far from high-quality to support systemic analysis of unconventional TCR repertoire. Here we present UcTCRdb, a database that contains 669,900 unconventional TCRs collected from 34 corresponding studies in humans, mice, and cattle. In UcTCRdb, users can interactively browse TCR features of different unconventional T cell subsets in different species, search and download sequences under different conditions. Additionally, basic and advanced online TCR analysis tools have been integrated into the database, which will facilitate the study of unconventional TCR patterns for users with different backgrounds. UcTCRdb is freely available at <http://uctcrdb.cn/>.

KEYWORDS

unconventional T cell, MAIT cell, NKT cell, CD8 $\alpha\alpha$ T cell, $\gamma\delta$ T cell, TCR (T cell receptor), bioinformatics and computational biology

Introduction

T cells play critical roles in immune responses. The classical T cell immune responses are primarily driven by the specific recognition between T cell receptors (TCRs) and peptide antigens presented by major histocompatibility complex (MHC) class I and II molecules. While there are several subpopulations of T cells, which work independently of this TCR-peptide-MHC paradigm, are defined as unconventional T cells. The unconventional T cells interact with molecules presented by MHC class Ib or MHC-like molecules, including lipid moieties, bacteria-derived vitamin B metabolites phosphoantigens, and butyrophilins (1, 2). These innate-like T cell subpopulations, including the MR1-restricted mucosal-associated invariant T cells (MAIT) (3, 4), CD1d-

restricted type I natural killer T cells (iNKT) (5, 6), CD1a- (7), CD1b- (8), CD1c- (9), H2-M3- (10), Qa-1- (11) or HLA-E (12)-restricted T cells, TCR $\alpha\beta$ CD8 $\alpha\alpha$ intraepithelial T lymphocytes (IELs) (13) and TCR $\gamma\delta$ T cells (14), typically exhibit relatively simplified TCR patterns, rapid effector responses, and ‘public’ antigen specificities.

Recent advances have highlighted that the recognition of antigens by unconventional TCRs is important for their early-life development and activation (15–19). However, the fine-grained patterns of antigen recognition by unconventional TCRs remain largely unknown. The most widely studied unconventional T cells, such as MAIT and iNKT cells, have been shown to have invariant TCR α gene usage and limited TCR β gene diversity (3–5, 20, 21). These gene usage biases are most likely due to the nonpolymorphic ligands they recognized, but the role of the randomly generated CDR3 sequences in the interaction is still not well clarified. Moreover, there are some unconventional T cell subpopulations such as TCR $\alpha\beta$ CD8 $\alpha\alpha$ IELs and TCR $\gamma\delta$ T cells, that were characterized by diverse TCR gene usage and CDR3 sequence, but with unknown antigenic specificity (22, 23). With the development of TCR sequencing technology and its wide use in the T cell studies (6, 24–29), further characterizing the unconventional TCR patterns is a critical requirement for understanding their responses to certain antigens and is a fundamental aspect of studying the development of such unconventional T cell subsets. While the small size and lack of standardization of the released unconventional TCR sequences strongly suggest that a comprehensive database aggregating the published data is in urgent need. To the best of our knowledge, there is still no systematic database for users to access and analyze unconventional TCR sequences.

To fill these gaps, we present UcTCRdb, a comprehensive database integrating published unconventional TCR sequences. Users can easily browse and download the data and search for sequences of their interest. To support mining the sequencing data, we also provide a series of TCR conservation analyses, including V and J gene usage, CDR3 length distribution, amino acid usage, sequence logo char, and similarity network displays. In addition, we offer users an option to analyze the datasets of interest individually, by simply entering data or uploading data files. This code-free analysis will facilitate the discovery of sequence patterns for users with different backgrounds. Since existing TCR databases mainly focus on conventional T cells (30–34), we believe that UcTCRdb will be a valuable resource to study unconventional T cells and provide new insights into T cell immunity.

Materials and methods

Data collection and preprocessing

To construct a comprehensive database of unconventional TCRs, we searched the published literature for unconventional TCR data on PubMed following the search terms: (CD8 $\alpha\alpha$ intraepithelial lymphocytes[Title/Abstract]) OR (mucosal-associated invariant T cells[Title/Abstract]) OR (MR1-restricted T

cells[Title/Abstract]) OR (type I NKT T cells[Title/Abstract]) OR (MAIT[Title/Abstract]) OR ($\gamma\delta$ T[Title/Abstract]) OR (iNKT[Title/Abstract]) OR (CD1b[Title/Abstract]) AND (T cell receptor). All results were then checked for inclusion of TCR sequences. The corresponding data listed in the literature tables or supplementary files were manually downloaded and archived. For those that do not provide processed TCR data, we downloaded the raw FASTQ files from Sequence Read Archive (SRA) (35) and extracted the TCR sequences using MiXCR (36). Besides, TCR sequences of $\gamma\delta$ T cells (dataset ID: 20-36) were collected from the immuneAccess website by setting the locus to TRAD OR TRG in the data browser page (<https://clients.adaptivebiotech.com/immuneaccess/browse>). To ensure the reliability and quality of the data, all the TCR sequences were subject to quality control by a uniform workflow, and only sequences with high confidence were retained. The sequences which lack V/J gene annotation or contain any illegal amino acid abbreviated characters were discarded. For the sequences that contain multiple V/J gene annotations, the first ones were kept as the most confident annotated gene. Duplicated sequences of identical V gene, J gene, and CDR3 amino acid sequence in each dataset were removed. By re-organizing the data format, the corresponding T cell type information and PubMed ID were retained in addition to the V gene, J gene, and TCR CDR3 amino acid sequences.

TCR analysis

UcTCRdb allows users to analyze selected data in the database or uploaded data files. V and J gene segment usage and gene-gene pairing landscapes are analyzed and shown using vertical stacks connected by curved paths whose thickness is proportional to the number of TCR clones with the respective gene pairing. The length distribution of the TCR CDR3 amino acid sequences was shown by the normalized frequency of each TCR chain. The normalized mutual information values (NMI) between each pair of CDR3 α/β residues were calculated for different CDR3 lengths to represent the association of amino acid usage at different positions. We use Shannon entropy to measure the uncertainty of amino acid usage at different CDR3 positions. The entropy $H(i)$ was calculated according to formula (1), where $P(x_i)$ represents the probability amino acid x appearing at position i , and $|X_i|$ is the set of all different amino acids in position i . Then the mutual information (MI) of amino acid distribution for position i and j was calculated according to formula (2), and $P(x_i, y_j)$ is the joint probability of x in position i and y in position j . The NMI was then calculated by using formula (3).

$$H(i) = -\sum_{x_i}^{X_i} P(x_i) \log(P(x_i)) \tag{1}$$

$$MI(i, j) = \sum_{x_i}^{X_i} \sum_{y_j}^{Y_j} P(x_i, y_j) \log\left(\frac{P(x_i, y_j)}{P(x_i)P(y_j)}\right) \tag{2}$$

$$NMI(i, j) = 2MI(i, j)/(H(i) + H(j)) \tag{3}$$

The NMI takes values in the range of 0-1, with a larger NMI indicating a higher amino acid correlation. We plotted the NMI matrix with a heatmap. The CDR3 sequence logos including amino acids as well as V and J genes were depicted by calculating the occurrence weights of different amino acids at each position. The larger the amino acid char in the sequence logo represents the higher frequency of amino acid occurrence. The amino acids were colored by their chemical properties in the sequence log plot. The scatter plot of amino acid usage was generated by calculating the proportion of each amino acid. To compare the differences between cell subsets, UcTCRdb has included 8 different datasets as references in the analysis module. To visualize the global similarity between TCR sequences, we calculated the Hamming distance between each pair of TCRs. We plotted a network graph where each node represents a TCR sequence and each edge represents a distance of less than 2 between two sequences. The visualization allows users to easily identify and explore groups of highly similar sequences.

Implementation of UcTCRdb

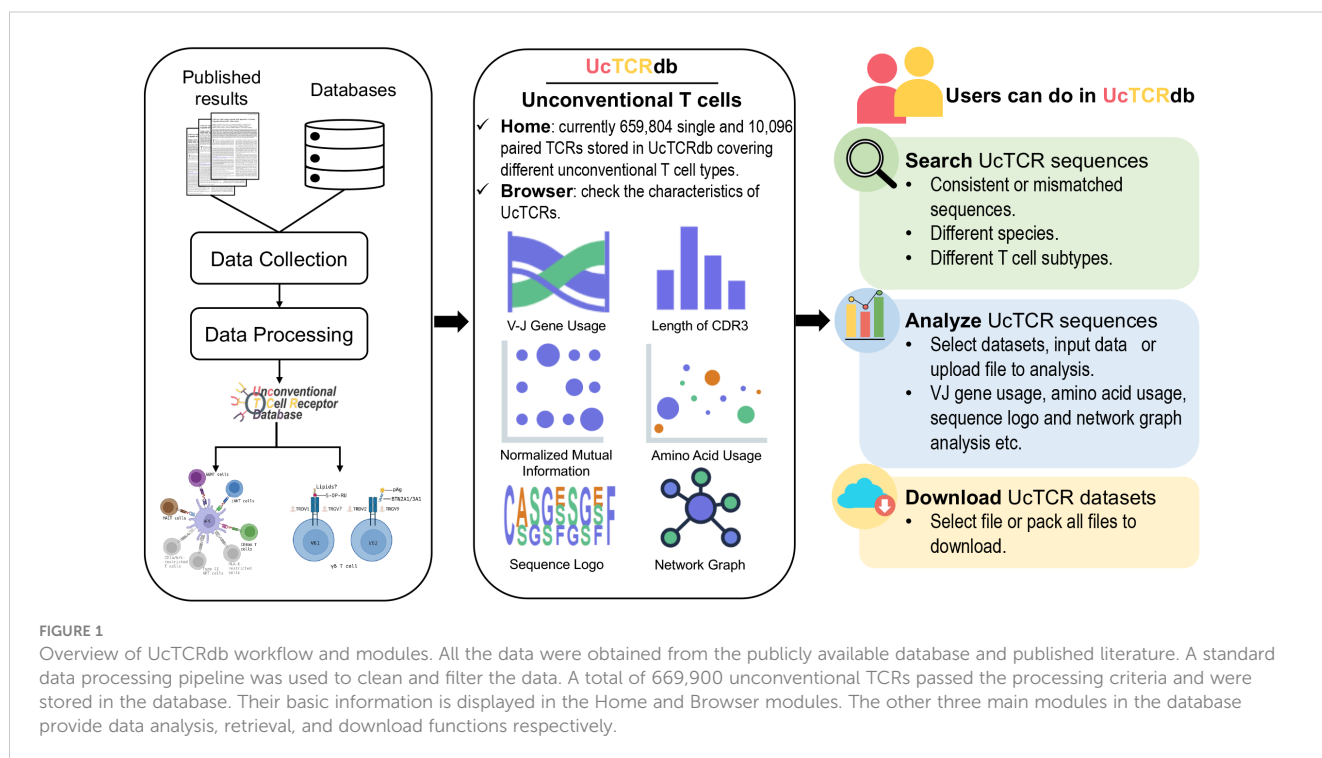
UcTCRdb was implemented using Vue (version 3.2.13), an approachable, performant, and versatile framework for building web user interfaces and Flask (version 1.1.2), a lightweight Web Server Gateway Interface (WSGI) web application framework. All statistics and figures were calculated using Python (version 3.6.9) scripts and plotted using Echarts (version 5.3.2). The website was proxied on Nginx, an HTTP and reverse proxy server. And the server was running on the AliCloud server.

Results

Overview of UcTCRdb

UcTCRdb consists of five main modules: Home, Browser, Analysis, Search, and Download (Figure 1). Users can access the introduction, significant functions, and summarization of the database from the home page. The browser module allows users to interactively look through the characteristics of TCRs in different unconventional T cell subpopulations from different species. To facilitate the visualization and comparison of the TCR characteristics between different datasets, UcTCRdb provides a codeless analysis module. Users can easily analyze the repertoire data without any coding experience by selecting the datasets in UcTCRdb or uploading their data files, which will provide a great convenience for users with no programming background to quickly analyze and grasp the features of TCRs. As a comprehensive database, one of the core applications of UcTCRdb is to allow users to search for identical or similar TCR sequences by specifying species and cell types. In addition, the download page offers a user-defined manner that allows users to directly select and download the data.

Based on the data processing criteria, a total of 659,804 single-chain TCR sequences and 10,096 paired-chain TCR sequences from 34 corresponding studies in humans, mice, and cattle were collected and integrated into UcTCRdb (Supplemental Table 1). Among them, over 85% of the data were derived from human samples (Supplementary Figure 1A) and are mostly $\gamma\delta$ TCRs (Supplementary Figures 1B, C). This was expectable since we retrieved most datasets from studies on the human $\gamma\delta$ T cells.



Besides, MR1-restricted MAIT cells, CD1d-restricted iNKT cells, and CD1b-restricted germline-encoded mycolipid-reactive (GEM) T cells, which have invariant TCR α gene usage and limited TCR β diversity, account for the majority of unconventional TCR $\alpha\beta$ datasets (Supplementary Figure 1D). Moreover, two datasets of TCR $\alpha\beta$ ⁺CD8 $\alpha\alpha$ ⁺ IELs were also included in the database.

Interactively browse the features of unconventional TCRs

The browser page focuses on the features of TCRs in the database, which allows users to look through the analysis results of the data from different views. To optimize the demonstration of the sequences data, unconventional T cells were categorized into two groups, unconventional $\alpha\beta$ T cells, and $\gamma\delta$ T cells. Here, we take human MAIT TCRs as an example to introduce the browser module (Figure 2A). Consistent with the previous studies (26, 38), we observed invariant TCR α gene usage and limited TCR β diversity in the integrated MAIT TCR data. As expected, the dominant conserved TCR α gene is TRAV1-2 and TRAJ33 (Figure 2B) and the length distribution of TCR α CDR3 is concentrated at 12 amino acids (Figure 2C). In contrast, MAIT TCR β shows more diversified gene usage and CDR3 length (Figures 2B, C). As shown in Figure 2B, the dominant TCR β V genes are TRBV20-1, TRBV6-1,2,4, and TRBV4-1,2 suggesting that the usage bias of these genes might be related to antigen recognition by MAIT cells.

To further display the conservation of MAIT TCR β sequences, we select the CDR3 sequences with a length of 15 amino acids, which represents the highest length frequency, for downstream analysis. By calculating the amino acid fraction and comparing it to the controlled conventional TCR β CDR3s, we found differences in the distribution of amino acid usage between them, particularly the higher use frequency of glycine (G) and Glutamic acid (E) in MAIT TCR β CDR3 sequences (Figure 2D). The heatmap of CDR3 NMI scores falls into three blocks, corresponding to the V gene motif region, the CDR3 highly variable region, and the J gene motif region (Figures 2E, F). In addition, multiple clusters were found in the similarity network graph of MAIT TCR β sequences, which also indicated the conserved patterns of MAIT TCR β (Figure 2G). The above examples showed that users can easily access TCR sequence features of different unconventional T cells across different species on the UcTCRdb browser page.

Analysis module in UcTCRdb

Further data mining of the unconventional TCR repertoire is critical for improving our understanding of the interaction between UcTCRs and their corresponding antigens. To simplify the analysis workflow and facilitate more users, UcTCRdb integrated several TCR analysis tools (Figure 3). For an overall analysis of the UcTCRs, the length distribution of CDR3 and Sankey graph display of VJ gene usage were provided to access the intrinsic characters of the data. In addition, the NMI heatmap and sequence

logo chart of the CDR3 sequences were provided to depict the correlation and probability of amino acids in each position. The NMI takes values from 0-1, with a larger number indicating a higher amino acid correlation.

To identify the differences between UcTCRs and conventional TCRs, we compared the amino acid usage between unconventional TCRs and background TCRs [(100,000 sequences randomly sampled from the conventional TCR CDR3 sequences (37)]. The result can be visualized by a scatter diagram with the horizontal and vertical coordinates indicating the amino acid proportion in UcTCRs and references. To obtain the global similarity of TCRs, we calculated the Hamming distance between any pair of CDR3 sequences and constructed a distance matrix. The results were visualized via a network graph, where nodes represent TCR sequences, and edges were drawn if the distance between two nodes is less than 2. Nodes were further colored by different TRBV genes. Users can click on the node to trigger all adjacent nodes and get the corresponding sequence information. The above tools were all integrated into the analysis interface of UcTCRdb, allowing for personalized analysis on both selected datasets and uploaded data files (Figure 3).

Search and download sequence data

UcTCRdb allows users to search for TCR sequences in the database that is identical or highly similar to the query sequence. Users can set different query conditions, such as specific cell types, species, or whether contain user-defined mismatches (Figure 4A). Currently, UcTCRdb provides two options for calculating sequence similarity, i.e. Hamming distance and Levenshtein (edit) distance. The search result can be further filtered by a specific V or J gene. Due to different sequencing methods, the unconventional TCR sequence data that we obtained contain both single-chain and paired chains. We use an asynchronous loading technique to fetch the corresponding pairing chain data so that the results can be reformatted before being displayed. Users can click the display button to view the result and can download the search result by directly clicking the 'Export To CSV' button.

On the download page, UcTCRdb enables users to download datasets that are specified by species and cell types (Figure 4B). For the convenience of the users who are interested in processing the unconventional TCR sequence data, UcTCRdb also allows users to download all selected datasets by clicking the 'Download' button.

Discussion

Sequencing technology has been widely used in T cell studies, which provide a deep view of the TCR repertoire in many aspects and a massive data source for integrated analysis (24, 37, 39–42). In this study, we systemically collected more than 0.6 million unconventional TCRs from corresponding studies in humans, mice, and cattle and constructed a database with online analysis tools. This database fills the gap that lacks a comprehensive aggregation of unconventional TCR sequence data. Users can easily browse and download the data and search for sequences of

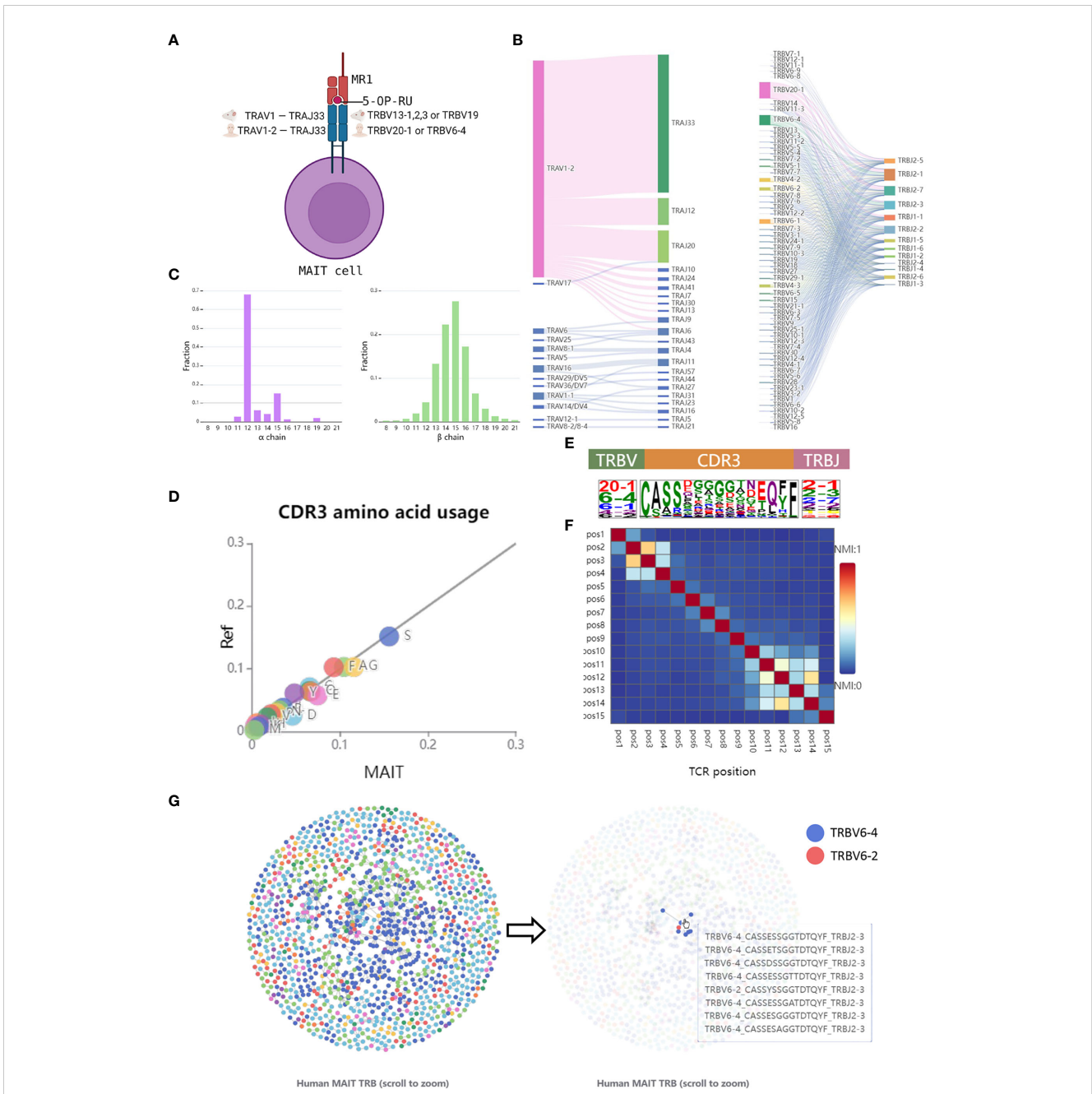


FIGURE 2

Browse the features of MAIT TCRs. **(A)** A schematic representation of MAIT cell. **(B)** V and J gene segment usage and gene-gene pairing landscapes are shown using vertical stacks (one for each V and J segment) connected by curved paths whose thickness is proportional to the number of TCR clones with the respective gene pairing. **(C)** CDR3 amino acid length distribution per chain. **(D)** The CDR3 β sequence logos including amino acids as well as TRBV and TRBJ were depicted by calculating the occurrence weights of different amino acids at the same position. The amino acids were colored by their chemical properties. **(E)** Amino acid usage between MAIT and conventional TCR β CDR3 sequences. A total of 100,000 conventional TCR β sequences were randomly selected from the dataset (37). **(F)** Heatmap showing the normalized mutual information value (NMI) between each pair of CDR3 β residues. The NMI takes values in the range of 0-1, with a larger NMI indicating a higher amino acid correlation. **(G)** The network graph is obtained by calculating the Hamming Distances between different CDR3 sequences. Nodes, which represent TCR clones, were colored by different TRBV genes. Edge was drawn if the distance between two clones is less than 2.

their interest. Given that more unconventional T cell subpopulations will be isolated and sequenced (24), we will continue collecting valid unconventional TCR sequence data and updating UcTCRdb. With the extensive unconventional TCR data and user-friendly web tools, UcTCRdb will be a powerful resource for users studying unconventional T cell immunity.

TCR signaling is the key for conventional T cells to survive positive and negative selection in the thymus, which largely depends on the binding affinity of TCR to self-peptide-MHCs. However, how unconventional T cells undergo selection and differentiation during T cell development remains unclear (43). Recent studies on MAIT cells suggested that effective antigen simulation of these cells



FIGURE 3 Analysis module in UcTCRdb. Screenshot of the search analysis page (1). UcTCRdb allows users to select a specific dataset or upload their data for analysis (2). There are six functions including TCR gene usage, CDR3 length distribution, NMI, amino acid usage, sequence conservation, and similarity network calculation in this module. Users can choose any one or several of them for data analysis.



FIGURE 4 Search module and download module in UcTCRdb. (A) Screenshot of the search module page (1). Input the query sequence and set species and cell type conditions (2). Select a preset distance calculation method for fuzzy search (3). UcTCRdb allows a secondary filter performed to the result with a specific V or J gene usage by clicking on the filter button (4). The search results can be downloaded by clicking on the 'Export To CSV' button. (B) Screenshot of the download module page (1). Select unconventional TCR datasets in different species and cell subpopulations (2). Users can download all the selected datasets by directly clicking the download button or down individual datasets one by one (3).

is essential for their thymic development and activation (44, 45), indicating that the TCRs may also determine the fate of conventional and unconventional T cells. Our study provides a systematic database of current available unconventional TCRs, which might be a valuable resource for understanding the development of unconventional T cells and how they may be related to diseases.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Author contributions

YD and SS collected the data. YD built the database and online analysis platform. JZ supervised the study and wrote the manuscript together with YD. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the National Natural Science Foundation of China (No. 32000468).

References

- Godfrey DI, Uldrich AP, McCluskey J, Rossjohn J, Moody DB. The burgeoning family of unconventional T cells. *Nat Immunol* (2015) 16:1114–23. doi: 10.1038/ni.3298
- Mayassi T, Barreiro LB, Rossjohn J, Jabri B. A multilayered immune system through the lens of unconventional T cells. *Nature* (2021) 595:501–10. doi: 10.1038/s41586-021-03578-0
- Tilloy F, Treiner E, Park SH, Garcia C, Lemonnier F, de la Salle H, et al. An invariant T cell receptor alpha chain defines a novel TAP-independent major histocompatibility complex class IIb-restricted alpha/beta T cell subpopulation in mammals. *J Exp Med* (1999) 189:1907–21. doi: 10.1084/jem.189.12.1907
- Huang S, Martin E, Kim S, Yu L, Soudais C, Fremont DH, et al. MR1 antigen presentation to mucosal-associated invariant T cells was highly conserved in evolution. *Proc Natl Acad Sci U.S.A.* (2009) 106:8290–5. doi: 10.1073/pnas.0903196106
- Borg NA, Wun KS, Kjer-Nielsen L, Wilce MCJ, Pellicci DG, Koh R, et al. CD1d-lipid-antigen recognition by the semi-invariant NKT T-cell receptor. *Nature* (2007) 448:44–9. doi: 10.1038/nature05907
- Jimeno R, Lebrusant-Fernandez M, Margreitter C, Lucas B, Veerapen N, Kelly G, et al. Tissue-specific shaping of the TCR repertoire and antigen specificity of iNKT cells. *eLife* (2019) 8:e51663. doi: 10.7554/eLife.51663
- Kasmar AG, Van Rhijn I, Magalhaes KG, Young DC, Cheng T-Y, Turner MT, et al. Cutting edge: CD1a tetramers and dextramers identify human lipopeptide-specific T cells *ex vivo*. *J Immunol Baltim Md 1950* (2013) 191:4499–503. doi: 10.4049/jimmunol.1301660
- Kasmar AG, van Rhijn I, Cheng T-Y, Turner M, Seshadri C, Schiefner A, et al. CD1b tetramers bind $\alpha\beta$ T cell receptors to identify a mycobacterial glycolipid-reactive T cell repertoire in humans. *J Exp Med* (2011) 208:1741–7. doi: 10.1084/jem.20110665
- Ly D, Kasmar AG, Cheng T-Y, de Jong A, Huang S, Roy S, et al. CD1c tetramers detect *ex vivo* T cell responses to processed phosphomycoketide antigens. *J Exp Med* (2013) 210:729–41. doi: 10.1084/jem.20120624
- Colmone A, Wang C-R. H2-M3-restricted T cell response to infection. *Microbes Infect* (2006) 8:2277–83. doi: 10.1016/j.micinf.2006.03.020
- Bian Y, Shang S, Siddiqui S, Zhao J, Joosten SA, Ottenhoff THM, et al. MHC IIb molecule qa-1 presents mycobacterium tuberculosis peptide antigens to CD8+ T cells and contributes to protection against infection. *PLoS Pathog* (2017) 13:e1006384. doi: 10.1371/journal.ppat.1006384
- Sullivan LC, Clements CS, Rossjohn J, Brooks AG. The major histocompatibility complex class IIb molecule HLA-E at the interface between innate and adaptive immunity. *Tissue Antigens* (2008) 72:415–24. doi: 10.1111/j.1399-0039.2008.01138.x
- Gapin L, Cheroutre H, Kronenberg M. Cutting edge: TCR alpha beta+ CD8 alpha alpha+ T cells are found in intestinal intraepithelial lymphocytes of mice that lack classical MHC class I molecules. *J Immunol Baltim Md 1950* (1999) 163:4100–4.
- Hayday AC. Gammadelta T cells and the lymphoid stress-surveillance response. *Immunity* (2009) 31:184–96. doi: 10.1016/j.immuni.2009.08.006
- An D, Oh SF, Olszak T, Neves JF, Avci FY, Erturk-Hasdemir D, et al. Sphingolipids from a symbiotic microbe regulate homeostasis of host intestinal natural killer T cells. *Cell* (2014) 156:123–33. doi: 10.1016/j.cell.2013.11.042
- Haas JD, Ravens S, Düber S, Sandrock I, Oberdörfer L, Kashani E, et al. Development of interleukin-17-producing $\gamma\delta$ T cells is restricted to a functional embryonic wave. *Immunity* (2012) 37:48–59. doi: 10.1016/j.immuni.2012.06.003
- Narayan K, Sylvia KE, Malhotra N, Yin CC, Martens G, Vallerkskog T, et al. Intrathymic programming of effector fates in three molecularly distinct $\gamma\delta$ T cell subtypes. *Nat Immunol* (2012) 13:511–8. doi: 10.1038/ni.2247
- Leeansyah E, Loh L, Nixon DF, Sandberg JK. Acquisition of innate-like microbial reactivity in mucosal tissues during human fetal MAIT-cell development. *Nat Commun* (2014) 5:3143. doi: 10.1038/ncomms4143
- Constantinides MG, Link VM, Tamoutounour S, Wong AC, Perez-Chaparro PJ, Han S-J, et al. MAIT cells are imprinted by the microbiota in early life and promote tissue repair. *Science* (2019) 366:eaax6624. doi: 10.1126/science.aax6624
- Eckle SBG, Birkinshaw RW, Kostenko L, Corbett AJ, McWilliam HEG, Reantragoon R, et al. A molecular basis underpinning the T cell receptor heterogeneity of mucosal-associated invariant T cells. *J Exp Med* (2014) 211:1585–600. doi: 10.1084/jem.20140484

Acknowledgments

We would like to thank Dr. Haiyang Wang for the helpful discussions.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2023.1158295/full#supplementary-material>

21. Le Nours J, Praveena T, Pellicci DG, Gherardin NA, Ross FJ, Lim RT, et al. Atypical natural killer T-cell receptor recognition of CD1d-lipid antigens. *Nat Commun* (2016) 7:10570. doi: 10.1038/ncomms10570
22. Ruscher R, Kummer RL, Lee YJ, Jameson SC, Hogquist KA. CD8 $\alpha\alpha$ intraepithelial lymphocytes arise from two main thymic precursors. *Nat Immunol* (2017) 18:771–9. doi: 10.1038/ni.3751
23. Willcox BE, Willcox CR. $\gamma\delta$ TCR ligands: the quest to solve a 500-million-year-old mystery. *Nat Immunol* (2019) 20:121–8. doi: 10.1038/s41590-018-0304-y
24. Rosati E, Rios Martini G, Pogorelyy MV, Minervina AA, Degenhardt F, Wendorff M, et al. A novel unconventional T cell population enriched in crohn's disease. *Gut* (2022) 71:2194–2204. doi: 10.1136/gutjnl-2021-325373
25. Gold MC, McLaren JE, Reistetter JA, Smyk-Pearson S, Ladell K, Swarbrick GM, et al. MR1-restricted MAIT cells display ligand discrimination and pathogen selectivity through distinct T cell receptor usage. *J Exp Med* (2014) 211:1601–10. doi: 10.1084/jem.20140507
26. Loh L, Gherardin NA, Sant S, Grzelak L, Crawford JC, Bird NL, et al. Human mucosal-associated invariant T cells in older individuals display expanded TCR $\alpha\beta$ clonotypes with potent antimicrobial responses. *J Immunol* (2020) 204:1119–33. doi: 10.4049/jimmunol.1900774
27. Wirasinha RC, Singh M, Archer SK, Chan A, Harrison PF, Goodnow CC, et al. $\alpha\beta$ T-cell receptors with a central CDR3 cysteine are enriched in CD8 $\alpha\alpha$ intraepithelial lymphocytes and their thymic precursors. *Immunol Cell Biol* (2018) 96:553–61. doi: 10.1111/imcb.12047
28. DeWitt WS, Yu KQ, Wilburn DB, Sherwood A, Vignali M, Day CL, et al. A diverse lipid antigen-specific TCR repertoire is clonally expanded during active tuberculosis. *J Immunol* (2018) 201:888–96. doi: 10.4049/jimmunol.1800186
29. Willcox CR, Davey MS, Willcox BE. Development and selection of the human V γ 9V δ 2+ T-cell repertoire. *Front Immunol* (2018) 9:1501. doi: 10.3389/fimmu.2018.01501
30. 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. doi: 10.1093/bioinformatics/btx286
31. Gowthaman R, Pierce BG. TCR3d: The T cell receptor structural repertoire database. *Bioinformatics* (2019) 35:5323–5. doi: 10.1093/bioinformatics/btz517
32. 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. doi: 10.1093/nar/gkz874
33. Zhang W, Wang L, Liu K, Wei X, Yang K, Du W, et al. PIRD: Pan immune repertoire database. *Bioinformatics* (2020) 36:897–903. doi: 10.1093/bioinformatics/btz614
34. Chen S-Y, Yue T, Lei Q, Guo A-Y. TCRdb: A comprehensive database for T-cell receptor sequences with powerful search function. *Nucleic Acids Res* (2021) 49:D468–74. doi: 10.1093/nar/gkaa796
35. Leinonen R, Sugawara H, Shumway M International Nucleotide Sequence Database Collaboration. The sequence read archive. *Nucleic Acids Res* (2011) 39: D19–21. doi: 10.1093/nar/gkq1019
36. Bolotin DA, Poslavsky S, Mitrophanov I, Shugay M, Mamedov IZ, Putintseva EV, et al. MiXCR: Software for comprehensive adaptive immunity profiling. *Nat Methods* (2015) 12:380–1. doi: 10.1038/nmeth.3364
37. Emerson RO, DeWitt WS, Vignali M, Gravley J, Hu JK, Osborne EJ, et al. Immunosequencing identifies signatures of cytomegalovirus exposure history and HLA-mediated effects on the T cell repertoire. *Nat Genet* (2017) 49:659–65. doi: 10.1038/ng.3822
38. Reantragoon R, Corbett AJ, Sakala IG, Gherardin NA, Furness JB, Chen Z, et al. Antigen-loaded MR1 tetramers define T cell receptor heterogeneity in mucosal-associated invariant T cells. *J Exp Med* (2013) 210:2305–20. doi: 10.1084/jem.20130958
39. Gomez-Tourino I, Kamra Y, Baptista R, Lorenc A, Peakman M. T Cell receptor β -chains display abnormal shortening and repertoire sharing in type 1 diabetes. *Nat Commun* (2017) 8:1792. doi: 10.1038/s41467-017-01925-2
40. Reuben A, Zhang J, Chiou S-H, Gittelman RM, Li J, Lee W-C, et al. Comprehensive T cell repertoire characterization of non-small cell lung cancer. *Nat Commun* (2020) 11:603. doi: 10.1038/s41467-019-14273-0
41. Minervina AA, Komech EA, Titov A, Bensouda Koraichi M, Rosati E, Mamedov IZ, et al. Longitudinal high-throughput TCR repertoire profiling reveals the dynamics of T-cell memory formation after mild COVID-19 infection. *eLife* (2021) 10:e63502. doi: 10.7554/eLife.63502
42. Schneider J, Kuhlmann L, Xiao Y, Raha S, Bernhardt G, Stadler M, et al. Healthy-like CD4+ regulatory and CD4+ conventional T-cell receptor repertoires predict protection from GVHD following donor lymphocyte infusion. *Int J Mol Sci* (2022) 23:10914. doi: 10.3390/ijms231810914
43. Pellicci DG, Koay H-F, Berzins SP. Thymic development of unconventional T cells: how NKT cells, MAIT cells and $\gamma\delta$ T cells emerge. *Nat Rev Immunol* (2020) 20:756–70. doi: 10.1038/s41577-020-0345-y
44. Legoux F, Bellet D, Daviaud C, El Morr Y, Darbois A, Niort K, et al. Microbial metabolites control the thymic development of mucosal-associated invariant T cells. *Science* (2019) 366:494–9. doi: 10.1126/science.aaw2719
45. Howson LJ, Awad W, von Borstel A, Lim HJ, McWilliam HEG, Sandoval-Romero ML, et al. Absence of mucosal-associated invariant T cells in a person with a homozygous point mutation in *MR1*. *Sci Immunol* (2020) 5:eabc9492. doi: 10.1126/sciimmunol.abc9492