BMC Genomics

RESEARCH

Open Access

scTCR-seq and HTS reveal a special novel TRBD2-TRBJ1 rearrangement in mammalian TRB CDR3 repertoire

Yingjie Wu[1,3†], Fengli Wu[2,3†], Jun Li[3*], Hao Zhou[3], Long Ma[3] and Xinsheng Yao[3*]

Abstract

Mammalian T cell receptor (TCR) beta-chain (TRB) V-D-J rearrangement mainly follows the "12/23 rule", and the "D-J rearrangement preceding the V-(D-J) rearrangement". Owing to the physical position of the D-J-C cluster in the TRB locus, the TRBD2 (D2) gene cannot directly perform inversional rearrangement or deletional/loop-out rearrangement with the TRBJ1 (J1) gene. Our previous studies revealed a single reverse TRBV30 (TRBV31 in mice) gene in the mammalian TRB locus, which can cause indirect rearrangement of the D2 gene and J1 gene; however, the mechanism and proportion involved in germline gene rearrangement are unknown. We obtained TRB CDR3 repertoires of thymus and peripheral tissues from humans and mice by HTS and scTCR-seq and found that 14% of the rearrangements in which the D2 gene is involved are D2-J1 rearrangements (D2-J2 rearrangements account for approximately 86%). The mechanism is that the reverse V30 gene preferentially performs inversional rearrangement with the D2 gene (V30-D2), leading to V30-D2-J1 rearrangement in humans, or the reverse V30 gene preferentially performs inversional rearrangement with the D1 gene (V30-D1), allowing the forward V genes (Vx) to perform Vx-D2-J1 rearrangement. We further found that D2-J1 rearrangements were present in more than 24% and more than 15% of the D2 gene rearrangements in rhesus monkeys and bats, respectively. Moreover, in bovine containing D1J1C1, D3J3C3, and D2J2C2 clusters, more than 11% D3-J1 and D2-J1 rearrangements and more than 22% D2-J3 rearrangements were found. This study provides a new perspective and feasible solution for further research on the significance of the special V-D-J recombination pattern in the mammalian TRB locus and the CDR3 repertoire formed by D2-J1 rearrangement.

Keywords Inversional rearrangement of the reverse TRBV gene, CDR3 repertoire, TRBD2-TRBJ1 rearrangement, Single-cell TCR sequencing, High-throughput sequencing

†Yingjie Wu and Fengli Wu contributed equally to this work.

*Correspondence:
Jun Li
huosu2014@163.com
Xinsheng Yao
immunology@126.com

[1]The Affiliated Taizhou People's Hospital of Nanjing Medical University, Taizhou, Jiangsu, China
[2]Department of Laboratory, The Affiliated Yongchuan Hospital of Chongqing Medical University, Chongqing, China
[3]Department of Immunology, Center of Immunomolecular Engineering, Innovation & Practice Base for Graduate Students Education, Zunyi Medical University, Zunyi, China

Introduction

The diversity of T and B cell receptor is at the heart of adaptive immune responses, and V(D)J rearrangement is the primary mechanism generating this diversity. In theory, gene recombination can provide a sufficient CDR3 repertoire to match the diverse antigens found in nature. To ensure that rearrangement occurs at the correct position, the V(D)J genes are regulated by recombination signal sequences (RSSs) during the recombination process. The RSS consists of sequence conserved heptamer and nonamer, as well as spacer that is conserved in length but not in sequence. The conserved heptamer sequence is 5'-CACAGTG-3', and the conserved nonamer sequence is 5'-ACAAAAACC-3'. Depending on the length of the spacer, which is 12–23 bp, RSS is referred to as 12RSS or 23RSS, respectively. The connection between gene segments with 12RSS and 23RSS during V(D)J rearrangement is governed by the "12/23 rule".

The principle of "D-J rearrangement preceding V-(D-J) rearrangement" is a classic mechanism of V(D)J rearrangement, which posits that during evolution, the D-J rearrangement in the TRB locus recombine first, followed by the recombination of the D-J complex with the V gene. The detailed mechanism by which an organism ensures that "D-J rearrangement precedes V-(D-J) rearrangement" has not been fully elucidated, but it is likely closely related to the relatively short distance between D and J on the TRB locus and the higher RIC scores of D-J [1]. Since both V-D rearrangement and D-J rearrangement comply with the "12/23 rule," the scenario where V-D recombines first and then the V-D complex recombines with J is theoretically possible and indeed occurs in practice. Investigating the proportion, mechanism, and significance of V-D recombination first can provide additional insights into the mechanisms and implications of VDJ rearrangement. In the TCRβ chain, the forward TRBV gene possesses a 3'V-23RSS (the RSS of the reverse TRBV gene is at the 5' end on the genome), the TRBD gene has a 5'D-12RSS and a 3'D-23RSS, and the TRBJ gene has a 5'J-12RSS. According to the "12/23 rule," the classical VDJ rearrangement occurs in two stages. First, the 3'D-23RSS-D and the 5'J-12RSS-J undergo forward deletional/looping out rearrangement to form a D-J rearrangement. Subsequently, the forward V-3'-23RSS and the 5'D-12RSS-D-J gene undergo forward deletional/looping out rearrangement, while the reverse V-5'-23RSS and the 5'D-12RSS-D-J gene undergo inversional rearrangement.

The random selection rule and "12/23 rule" of classic TCR V-(D)-J rearrangement have been challenged, and special recombination patterns have been continuously added [2]. For example, the beyond 12/23 rule [3], advantageous usage of V and J gene families [4], allelic inclusion rearrangement [5, 6], TRBV-TRBJ direct rearrangement [7], and D-D fusion [8, 9], etc. These special recombination patterns were discovered early by single T cell PCR, transgenic mice and other techniques, but little is known about the proportion and significance of their occurrence under physiological conditions. The application of high-throughput sequencing (HTS) and single-cell RNA sequencing (scRNA-seq) in TCR sequencing, especially the application of single-cell RNA sequencing, provides unprecedented opportunities for in-depth study of the special patterns, proportions and significance of V-D-J rearrangements.

Among the 13 mammals whose TCRβ chains have been fully annotated, the order of gene location in the TRB gene cluster is V(1–29)--D1-J1-C1–D2-J2-C2–V30 [10]. According to the "12/23 rule" and the "D-J rearrangement preceding the V-(D-J) rearrangement", the recombination pattern of the classic TCRβ chain is that 3'D-23RSS preferentially recombines with 5'J-12RSS, and then 3'V-23RSS recombines with 5'D-RSS, which leads to the TRBD2 (D2) gene not directly performing inversional rearrangement or deletional/loop-out rearrangement with the TRBJ1 (J1) gene [2, 3, 11](Fig. 1A). Although in theory, 3'D2-23RSS can perform deletional rearrangement with 5'J 1-12RSS, this will cause the D2-J1 region to be looped out of the chromosome and no longer participate in Vx-D2-J1 rearrangement.

Our previous studies revealed a single reverse TRBV30 (TRBV31 in mice) gene in the mammalian TRB locus, which can cause indirect rearrangement of the D2 gene and J1 gene; however, the mechanism and proportion involved in germline gene rearrangement are unknown [10]. We used HTS and single-cell TCR sequencing (scTCR-seq) methods to obtain humans, mice, rhesus monkeys, bats and bovines TRB CDR3 repertoires and analyzed the mechanism of the mammalian D2-J1 recombination pattern and the proportions of D2-J1 recombination patterns involved in germline gene rearrangement, which provided a basic and new perspective for studying the special V-D-J recombination pattern and significance at the TRB locus (Fig. 1B).

Results

Location, order and recombination rule of the V, D, J, and C genes in the mammalian TRB locus

There are 13 mammals whose TRB locus has been fully annotated [10]. The physical position order of the V, D, J, and C gene clusters of TRB is forward V(1–29), D1-J1-C1, D2-J2-C2, and V30. Among them, bovines, sheep, and pigs contain unique D3-J3-C3 clusters, and the reverse TRBV gene of mice is named TRBV31 (V31).

Each recombination signal sequence (RSS) comprises a conserved heptamer (7 nucleotides (nt)) and a conserved nonamer (9 nt) separated by a 12 or 23 nt spacer. In the TRB locus, V-23RSS is located at the 3'end of the TRBV

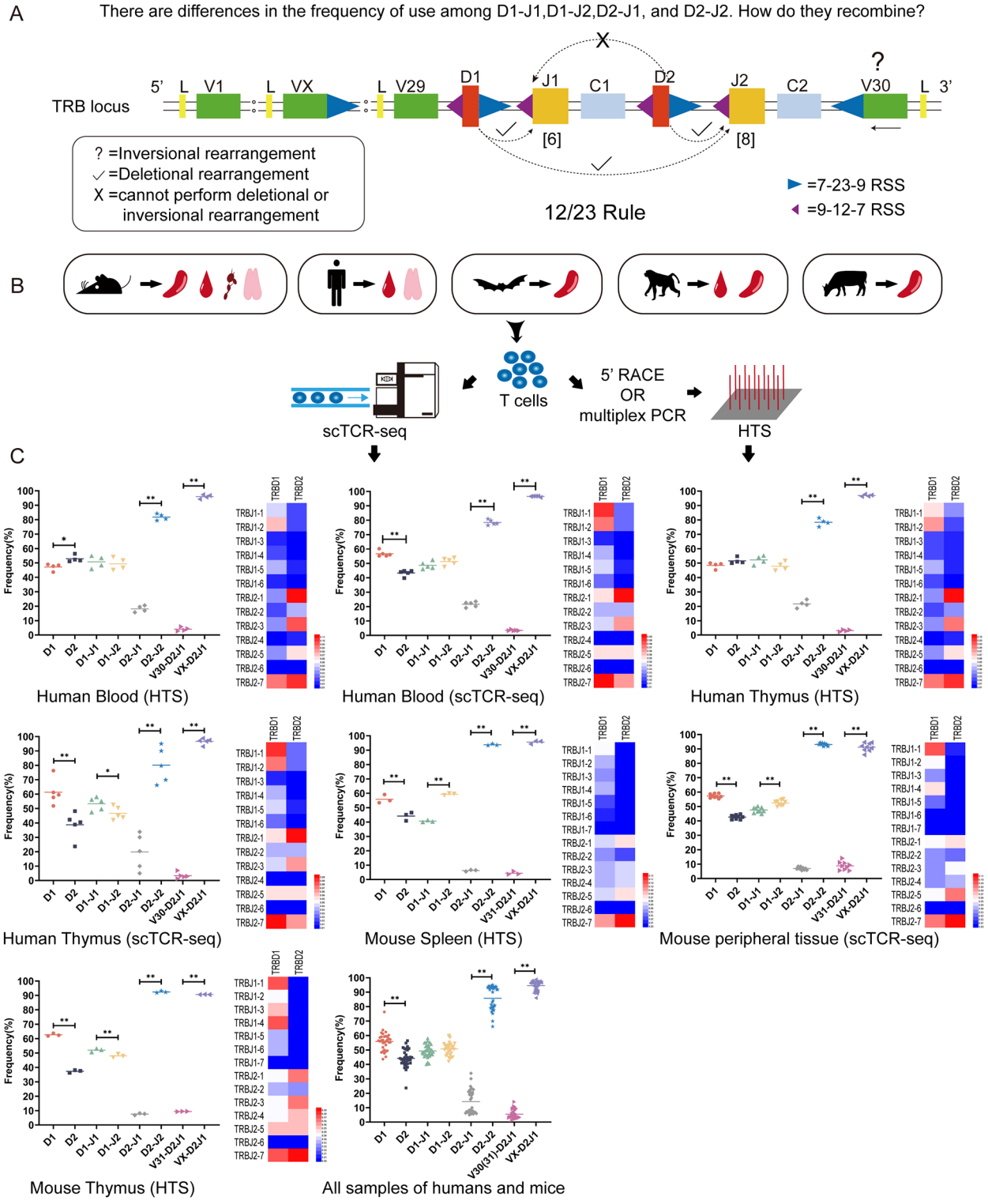

Fig. 1 (See legend on next page.)

gene, J-12RSS is located at the 5'end of the TRBJ gene, D-23RSS and D-12RSS are located at the 5'end and 3'end of the TRBD gene, respectively. The V-D-J rearrangement follows the "12/23 rule" and the "D-J rearrangement preceding the V-(D-J) rearrangement". According to these two classic rules, the D2 gene cannot directly recombine with the J1 gene in the TRB locus (Fig. 1A).

(See figure on previous page.)

Fig. 1 Schematic diagram of the TRB locus and V-D-J rearrangement; schematic diagram of the object, sample and method; and statistical analysis of the TCR V-D-J rearrangement frequency in humans and mice. (**A**) Schematic diagram of the TRB locus and V-D-J rearrangement. According to the "12/23 rule" and the "D-J rearrangement preceding the V-(D-J) rearrangement", the TRBD2 gene cannot directly undergo inversional recombination or deletional/loop-out recombination with the TRBJ1gene. (**B**) Schematic diagram of the object, sample and method. We utilized a total of 42 samples from five mammalian species, obtained TRB CDR repertoires through HTS and scTCR-seq technologies, and employed tools to perform alignment on the TRB CDR repertoires to identify V, D, and J subfamily gene sequences. (**C**) Statistical analysis of the TCR V-D-J rearrangement frequency in humans and mice. We analyzed the usage frequency of VDJ genes in the TRB CDR repertoires of human and mouse samples, presenting the data using dot plots and heatmaps. Statistical analyses were conducted between D1 and D2, between D1-J1 and D1-J2, between D2-J1 and D2-J2, and between V30(V31)-D2J1 and VX-D2J1. Note: (1) D1 is the complete sequence (complete Vx-D1-Jx) of all D1 genes involved in the rearrangement of each sample; D2 is the complete sequence (complete Vx-D2-Jx) of all D2 genes involved in the rearrangement of each sample; D1 + D2 = 100%; D1-J1 denotes total Vx-D1-J1; D1-J2 denotes total Vx-D1-J2; D1-J1 + D1-J2 = 100%; D2-J1 denotes total Vx-D2-J1; D2-J2 denotes total Vx-D2-J2; D2-J1 + D2-J2 = 100%; V30(or V31)-D2-J1 + Vx-D2-J1 = 100%. (2) The heatmap is a visual comparison of the objective existence and relatively low frequency of D2-J1 rearrangement usage. (3) Visualization of D-J (V-D-J) recombination and usage of D genes. We used a dot plot for sample groups with $n \geq 3$ and performed statistical analysis using Mann Whitney U test. All statistically significant differences are indicated as *P value < 0.05 and **P value < 0.01

Number and frequency of CDR3 sequences associated with D-J rearrangement in the human, mouse, rhesus monkey, Bat and bovine TRB CDR3 repertoires

The human sample consisted of 8 HTS samples and 10 scTCR-seq samples. The mouse sample consisted of 6 HTS samples and 9 scTCR-seq samples. The rhesus monkey sample consisted of 1 HTS sample and 5 scTCR-seq samples. The bat sample consisted of 2 HTS samples. The bovine sample consisted of 1 HTS sample. In Supplementary Table 1, we provide a detailed description of the tissue sources, sequencing methods, identification numbers, number of unique sequences (Productive/Unproductive), and the number of D-J rearrangement sequences (D1/D2; D1/D3/D2; D1-J1; D1-J3; D1-J2; D3-J1/(V30-D3J1); D3-J3; D3-J2; D2-J1/(V30-D2J1); D2-J3/(V30-D2J3); D2-J2) for the human, mouse, rhesus monkey, bat, and bovine samples (Supplementary Table 1).

In the human thymus HTS samples, D2-J1 rearrangement accounted for 21.61% of the D2 genes. In the human thymus scTCR-seq samples, D2-J1 rearrangement accounted for 19.85% of the D2 genes. In human peripheral blood HTS samples, D2-J1 rearrangement accounted for 18.19% of D2 genes. In the human peripheral blood scTCR-seq samples, D2-J1 rearrangement accounted for 21.57% of the D2 genes (Figs. 1C and 2A-D).

In the mouse spleen HTS samples, D2-J1 rearrangement accounted for 6.27% of the D2 genes. In the HTS samples from the mouse thymus, D2-J1 rearrangement accounted for 7.59% of the D2 genes. In the mouse peripheral tissue scTCR-seq samples, D2-J1 rearrangement accounted for 6.90% of the D genes (Figs. 1C and 2E-G).

In all the human and mouse samples, the D1 and D2 genes accounted for 55.86% and 44.14% of the total D genes, respectively; D1-J1 and D1-J2 rearrangements accounted for 49.33% and 50.67% of the D1 genes, respectively; D2-J1 and D2-J2 rearrangements accounted for 14.24% and 85.76% of the D2 genes, respectively; and V30(31)-D2-J1 rearrangements and Vx-D2-J1 rearrangements accounted for 5.53% and 94.47% of the D2-J1 rearrangements, respectively (Fig. 1C).

In bat spleen HTS samples (SRR21464508), D2-J1 rearrangements accounted for 21.06% of the D2 genes, and V30-D2-J1 rearrangements and Vx-D2-J1 rearrangements accounted for 3.49% and 96.51% of the D2-J1 rearrangements, respectively. In bat spleen HTS samples (SRR21464509), D2-J1 rearrangement accounted for 15.48% of D2 genes, and V30-D2-J1 rearrangement and Vx-D2-J1 rearrangement accounted for 1.19% and 98.81% of D2-J1 rearrangements, respectively (Fig. 2H, L, M).

In rhesus monkey peripheral blood HTS samples, D2-J1 rearrangement accounted for 24.14% of D2 genes, and V30-D2-J1 rearrangement and Vx-D2-J1 rearrangement accounted for 1.17% and 98.83% of D2-J1 rearrangements, respectively. In rhesus monkey peripheral blood scTCR-seq samples, D2-J1 rearrangement accounted for 29.46% of D2 genes, and V30-D2-J1 rearrangement and Vx-D2-J1 rearrangement accounted for 1.49% and 98.51% of D2-J1 rearrangements, respectively (Fig. 2I, J, N, M).

In the bovine spleen HTS samples, D3-J1 rearrangement accounted for 11.46% of the D3 gene, D2-J1 rearrangement and D2-J3 arrangement accounted for 11.51% and 22.48% of the D2 gene, respectively, V30-D3-J1 and Vx-D3-J1 accounted for 3.03% and 96.97% of the D3-J1 rearrangement, V30-D2-J1 and Vx-D2-J1 accounted for 0.98% and 99.02% of the D2-J1 rearrangement, and V30-D2-J3 and Vx-D2-J3 accounted for 3.52% and 96.48% of the D2-J3 rearrangement, respectively (Fig. 2K, P).

All reverse V30-D2-J1 sequences are listed in Supplementary Tables 2, and a complete V30-D2-J1 sequence and a complete V30-D1-J1/J2 sequence are listed in Supplementary Table 3 for comparative analysis. We selected a complete V30-D2-J1-C1 sequence and a complete V30-D1-J1-C1 sequence and compared them with the germline VDJC gene in the IMGT database to display the sequence of V30 involved in the recombination process (Fig. 2Q).

Mechanism and CDR3 sequence examples of D2-J1 rearrangement

The possible mechanisms of D2-J1 rearrangement are as follows: (1) the reverse V30 (V31 in mice) gene preferentially performs inversional rearrangement with the D2 gene, leading to V30-D2-J1 (V31-D2-J1 in mice) rearrangement (Fig. 3A); (2) the reverse V30 (V31 in mice) gene preferentially performs inversional rearrangement with the D1 gene, allowing the forward V genes (Vx) to perform inversional rearrangement with the D2 gene to form Vx-D2 rearrangement, and then performs inversional rearrangement with the J1 gene family to form Vx-D2-J1 rearrangement (Fig. 3B).

Examples of V30-D2-J1-1, V30-D2-J1-2, V30-D2-J1-3, V30-D2-J1-4, V30-D2-J1-5, and V30-D2-J1-6 CDR3 sequence in human single cell are shown in Fig. 3C. Examples of Vx-D2-J1-1, Vx-D2-J1-2, Vx-D2-J1-3, Vx-D2-J1-4, Vx-D2-J1-5, and Vx-D2-J1-6 CDR3 sequence in human single cell are shown in Fig. 3D. Examples of V31-D2-J1-1, V31-D2-J1-2, V31-D2-J1-3, V31-D2-J1-4, and V31-D2-J1-5 CDR3 sequence in mouse single cell are shown in Fig. 3E. Examples of Vx-D2-J1-1, Vx-D2-J1-2, Vx-D2-J1-3, Vx-D2-J1-4, and Vx-D2-J1-5 CDR3 sequence in mouse single cell are shown in Fig. 3F. In the human single cell TCR sequence, an example of a single T cell with both V30-D1-J2 rearrangement and Vx-D2-J1 rearrangement is shown in Fig. 3G. In the mouse single cell TCR sequence, an example of a single T cell with both V31-D1-J2 rearrangement and Vx-D2-J1 rearrangement is shown in Fig. 3H. The frequency of V30(V31)-D2-J1 rearrangement in 42 samples from 5 mammals and the average frequency of forward V1-29-D2-J1 rearrangement in 42 samples are shown in Fig. 3I.

Discussion

Mammals have evolved a highly diverse adaptive immune response that relies on random rearrangement and self-tolerant selection of germline immunoglobulin (IG) and TR genes [12]. The classical V-D-J rearrangement in the TRB locus follows the "12/23 rule", and the "D-J rearrangement preceding the V-(D-J) rearrangement". Therefore, the D2 gene cannot directly recombine with the J1 gene according to the rules and the physical location of the D and J clusters in the TRB locus [2, 3, 11].

In recent years, special V-(D)-J recombination patterns and mechanisms have consistently supplemented the classical recombination pattern [3–9]. On the basis of our own experimental data and shared data [4, 13–17], we found that a single reverse V30 gene exists at the mammalian TRB locus, which may lead to a new recombination pattern, resulting in indirect recombination between the D2 gene and the J1 gene. However, the mechanism of this recombination and the proportion of germline V-D-J recombination are not clear [10].

In this study, we selected humans, mice, rhesus monkeys, bats, and bovine from the 13 mammals whose TRB locus have been fully annotated and used HTS and scTCR-seq technologies to obtain TRB CDR3 repertoires of central and peripheral samples (42 in total) from different mammals [10]. We strictly followed the rules of IMGT/GENE-DB (https://www.imgt.org/genedb/), MiXCR (version3.0.13), IMGT/HighV-QUEST, Cell Ranger, and immunSEQ software to compare and determine the composition of the V, D, J, and C gene subfamilies in each TRB CDR3 sequence. In the scTCR-seq and HTS TCR-β CDR3 repertoires, we detected 14-24% D2-J1 rearrangement in humans, mice, rhesus monkeys and bats. Moreover, we detected 11-22% D3-J1, D2-J1 and D2-J3 rearrangements in the bovine TCR-β CDR3 repertoire, which contained three DJC clusters (D1J1C1, D3J3C3, D2J2C2).

We analyzed the usage frequencies of the V, D, and J gene subfamilies in the TRB CDR3 sequences of 42 samples and found that the usage frequency of D1 gene was higher than that of D2 gene, and the frequencies of D1-J1 rearrangements and D1-J2 rearrangements were similar. This result is consistent with the report of Teichmann et al. [6]. This result is completely consistent with the classical V-D-J recombination theory; that is, according to the recombination rules and the physical location of the genes, D1 gene can perform deletion rearrangement with J1 gene or J2 gene. In contrast, the D2 gene cannot directly recombine with J1 gene; therefore, D1 gene has more opportunities to participate in germline gene rearrangement than D2 gene does. This result was further verified in samples from bovine with three DJC clusters; that is, the usage frequency of D1 gene in bovines was higher than that of D3 gene and D2 gene.

According to classical rearrangement theory, D2 gene can only recombine with J2 gene and then participate in V-(D2-J2) rearrangement. However, we found that D2-J2 rearrangement accounted for only approximately 85% of the rearrangements involved in the D2 gene (less than 80% in bats and rhesus monkeys), and each sample had approximately 15% D2-J1 rearrangement. On the basis of the "12/23 rule", the "D-J rearrangement preceding the V-(D-J) rearrangement", the physical location of the V, D, J, and C gene clusters in the TRB locus, and a large number of V30-D2-J1 and Vx-D2-J1 sequence composition characteristics, we propose a possible mechanism for the new recombination pattern of D2-J1 rearrangement: (1) The reverse V30 gene preferentially performs inversional rearrangement with the D2 gene, resulting in the inversional rearrangement between V30-D2 and the J1 gene family, forming V30-D2-J1 (Fig. 3A). (2) The reverse V30 gene preferentially performs inversional rearrangement with the D1 gene, resulting in an inversional

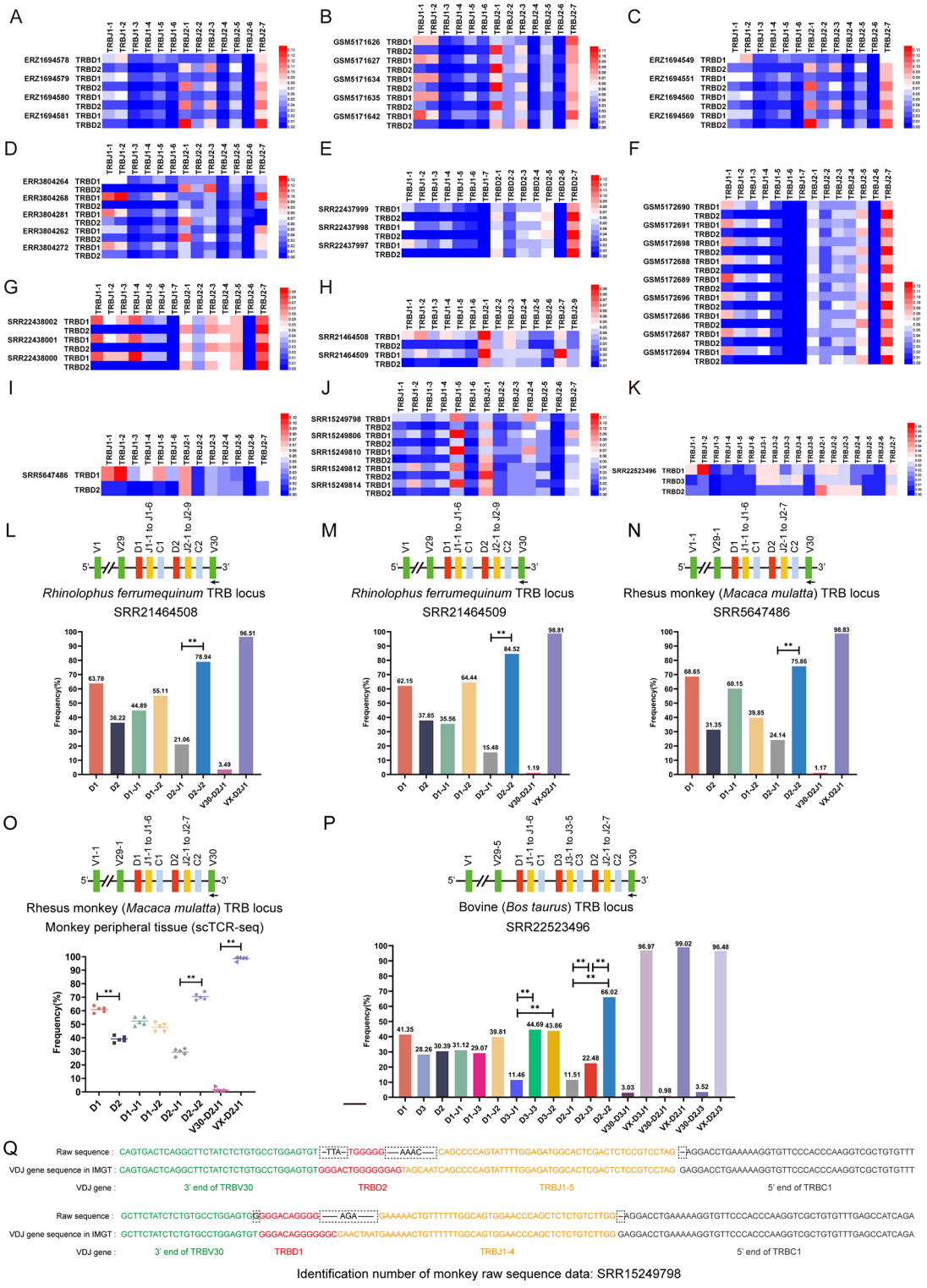

Fig. 2 (See legend on next page.)

rearrangement between the forward V gene and the D2 gene and then recombines with J1 to form Vx-D2-J1 (Fig. 3B).

On the basis of the proposed mechanism, we analyzed the proportion of V-(D-J) rearrangements that occurred after D-J rearrangement and found that the V30-D2-J1 rearrangement accounted for approximately 5%, and the

(See figure on previous page.)

Fig. 2 Heatmap of D-J rearrangement frequencies of 42 samples and statistical analysis of the TCR V-D-J rearrangement frequencies of bats, rhesus monkeys, and bovines. (**A**) HTS data of 4 human blood samples. (**B**) scTCR-seq data of 5 human blood samples. (**C**) HTS data of 4 human thymus samples. (**D**) scTCR-seq data of 5 human thymus samples. (**E**) HTS data of 3 mouse spleen samples. (**F**) scTCR-seq data of 9 mouse peripheral tissue samples. (**G**) HTS data of 3 mouse thymus samples. (**H**) HTS data of 2 bat spleen samples. (**I**) HTS data of 1 rhesus monkey blood sample. (**J**) scTCR-seq data of 5 rhesus monkey peripheral tissue samples; (**K**) HTS data of 1 bovine spleen sample. (**L**) Frequency of D genes and rearranged D-J usage in bat spleen samples (SRR21464508); (**M**) Frequency of D genes and rearranged D-J usage in bat spleen samples (SRR21464509); (**N**) rhesus monkey blood sample; (**O**) rhesus monkey peripheral tissue samples; (**P**) bovine spleen sample; (**Q**) Examples of raw and annotated sequences from rhesus monkey scTCR-seq data (SRR15249798). Note: (**1**) D1 is the complete sequence (complete Vx-D1-Jx) of all D1 genes involved in the rearrangement of each sample; D2 is the complete sequence (complete Vx-D2-Jx) of all D2 genes involved in the rearrangement of each sample; D3 is the complete sequence (complete Vx-D3-Jx) of all D3 genes involved in the rearrangement of each sample; D1-J1 denotes total Vx-D1-J1; D1-J2 denotes total Vx-D1-J2; D1-J3 denotes total Vx-D1-J3; D2-J1 denotes total Vx-D2-J1; D2-J2 denotes total Vx-D2-J2; D2-J3 denotes total Vx-D2-J3; D3-J1 denotes total Vx-D3-J1; D3-J2 denotes total Vx-D3-J2; D3-J3 denotes total Vx-D3-J3; Bat and rhesus monkey: D1 + D2 = 100%; D1-J1 + D1-J2 = 100%; D2-J1 + D2-J2 = 100%; V30(or V31)-D2-J1 + Vx-D2-J1 = 100%; Bovine: D1 + D2 + D3 = 100%; D1-J1 + D1-J2 + D1-J3 = 100%; D2-J1 + D2-J2 + D2-J3 = 100%; D3-J1 + D3-J2 + D3-J3 = 100%; V30-D3-J1 + Vx-D3-J1 = 100%; V30-D2-J1 + Vx-D2-J1 = 100%; V30-D2-J3 + Vx-D2-J3 = 100%. **(2)** A bar chart was used for sample groups with $n = 1$, and statistical analysis was performed using Pearson's chi-squared test (**L**-**N**, **P**). A dot plot was used for sample groups with $n \geq 3$, and statistical analysis was performed using the Mann-Whitney U test (**O**). All statistically significant differences are indicated as *P value < 0.05 and **P value < 0.01

rest are Vx-D2-J1 rearrangement, but the average proportion of each forward V gene subfamily was lower than that of the V30 gene (Fig. 3I). This result indirectly confirms the mechanism whereby when the V30 gene preferentially performs inversional rearrangement with the D2 gene, it has more opportunities to recombine with J1 gene, whereas when the V30 gene preferentially performs inversional rearrangement with the D1 gene, it significantly reduces the chance of each V gene subfamily from the forward V1 to the V29 gene participating in Vx-D2-J1 rearrangement. Moreover, based on the mechanism, we listed examples of both V30-D1-J2 rearrangement and Vx-D2-J1 rearrangement coexist in human single cell TCR sequence from a few T cells expressing dual TCR [18](Fig. 3G). Examples of sequences in which V31-D1-J2 rearrangement and Vx-D2-J1 rearrangement coexist in single T cells in mouse single cell TCR sequence are shown in Fig. 3H.

The proportion of T cells with a "special D2-J1 recombination pattern" in the human, mouse, rhesus monkey, bat and bovine samples analyzed in this study is based on a comparison of the gene sequences of D genes in different mammalian embryonic lines, all of which strictly follow the criteria for the D2 gene and D1 gene established by IMGT/GENE-DB, MiXCR, CellRanger, IMGT/HighV-QUEST, etc [19]. Using TRB CDR3 repertoires shared by multiple laboratories, we validated the special recombination pattern of TRBD2-TRBJ1 in mammals with two DJC clusters and in mammals with three DJC clusters with identical results. The results of this study support the special TRB V-D-J recombination pattern and provide a basis for further research on the CDR3 function of V30-D2-J1 (V31-D2-J1 in mice) and Vx-D2-J1.

Notably, the results of this study are based on evidence that the CDR3 sequences were obtained by HTS and scTCR-seq after the construction of the CDR3 repertoire by mRNA and DNA, rather than directly from V-D-J recombination sequences on each T cell chromosome. In future studies, we will consider how to isolate two chromosomes in T cells and sequence the VDJC region of a single chromosome in a single T cell to further investigate the possible mechanism of the "special recombination pattern of D2-J1". Moreover, because this study is based on the "special recombination pattern of D2-J1" under the premise of the "12/23 rule", we cannot determine whether there are other new mechanisms that do not follow the "12/23 rule"; for example, the RAG can be recombined through a special RSS (crypticRSS, whose 7-mer contains only "CAC") [20], which requires further research in specific cell and animal models. In addition, it is not clear how a single reverse TRBV30 gene evolved at the mammalian TRB locus, the mechanism and product of the "D2-J1" special recombination caused by the TRBV30 gene, and whether it is "beneficial" or "harmful" to individuals under physiological conditions need further study [10].

TCR recombination is heavily dependent on retrotransposons and transposable elements, but to date, there is no report on how the promoter and transposon of the reverse TRBV gene participate in regulation, nor on the similarities and differences between the promoter and transposon of the reverse TRBV gene and those of the forward TRBV gene. Jerzy K Kulski et al. found the SINE-VNTR-Alu (SVA) retrotransposon insertion polymorphisms (RIPs) in the major histocompatibility complex (MHC) genomic region of human, which can participate in the gene expression of innate immunity and adaptive immunity, and may regulate the progress of Parkinson's disease [21, 22]. It is suggested that it will be an important direction to explore the regulatory mechanism of retrotransposons and transposable elements of reverse TRBV gene in reverse VDJ rearrangement and transcription, as well as its advantages and transcription effects in reverse VDJ rearrangement.

The V(D)J rearrangement of TCR and BCR, along with the characteristics of the CDR3 repertoire, constitutes the foundation of adaptive immunity research. With the annotation of mammalian germline VDJC genes and

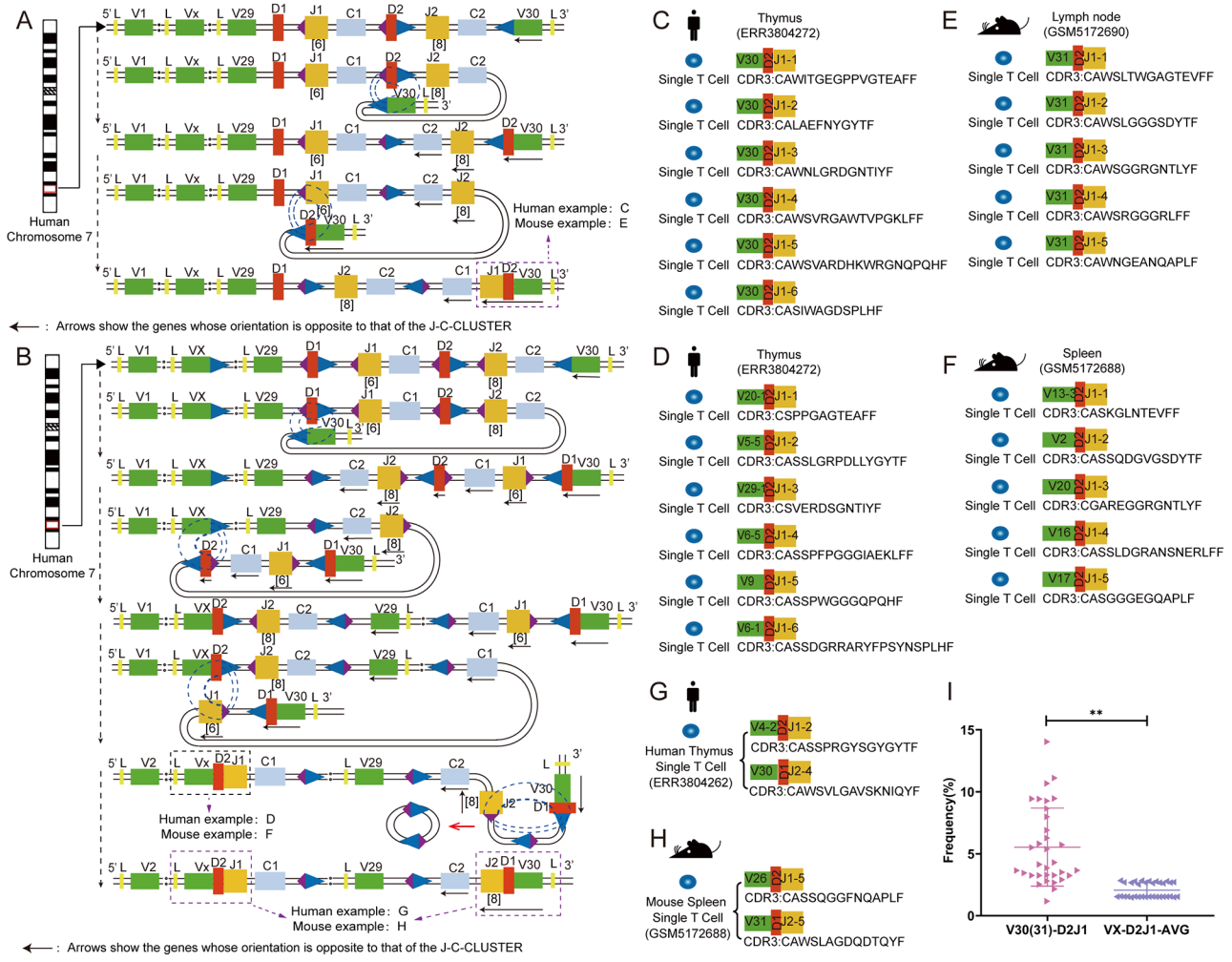

Fig. 3 TRBD2-TRBJ1 rearrangement mechanism diagram; example diagram of human and mouse D2-J1 rearrangement sequences, comparison statistics of V30(31)-D2-J1 frequency and average frequency of each V gene subfamily in forward Vx-D2-J1. (**A**) The reverse V30 first recombines with the D2 gene, resulting in inversional rearrangement of V30-D2 with the J1 family to form V30-D2-J1. (**B**) The reverse V30 first recombines with the D1 gene in inversional rearrangement, contributing to inversion of Vx-D2 and the Vx-D2-J1(x) rearrangement mechanism. (**C**) Examples of reverse V30-D2-J1 arrangement sequences in human single T cell. (**D**) Examples of forward Vx-D2-J1 arrangement sequences in human single T cell. (**E**) Examples of reverse V31-D2-J1 arrangement sequences in mouse single T cell. (**F**) Examples of forward Vx-D2-J1 arrangement sequences in mouse single T cell. (**G**) Examples of concurrent V30-D1-Jx and Vx-D2-J1(x) rearrangement sequences in a single T cell of human. (**H**) Examples of concurrent V31-D1-Jx and Vx-D2-J1(x) rearrangement sequences in a single T cell of mouse. (**I**) Statistical analysis of the frequency of V30(31)-D2-J1 in human and mouse total samples compared with the average frequency of forward Vx-D2-J1. Note. We used a dot plot for sample groups with $n \geq 3$ and performed statistical analysis using the Mann-Whitney U test. All statistically significant differences are indicated as *P value < 0.05 and **P value < 0.01

the advancement of experimental technologies, particularly the emergence of high-throughput sequencing and single-cell sequencing technologies, we are now afforded the opportunity to investigate and further elucidate V(D)J rearrangement from novel perspectives.

With an increasing number of studies revealing that V(D)J recombination not only adheres to the classical "12/23 rule" and the principle of "D-J rearrangement preceding V-(D-J) rearrangement," but also follows some special rules such as the "beyond 12/23 rule," "V-J direct rearrangement," "advantageous use of V and J gene families," "allelic inclusion rearrangement," and "D-D fusion." Based on the position and order of the D-J-C cluster on the TRB locus, we discovered that the D2 gene cannot undergo D2-J1 rearrangement through the classical deletional or inversional rearrangement. Simultaneously, we identified a single reverse TRBV gene in the mammalian TRB locus. Utilizing our research group's annotations of mammalian germline VDJC genes and the IMGT database's annotations of mammalian germline VDJC genes, combined with HTS and scTCR-seq data, we investigated the D2-J1 rearrangement. We found that the single reverse TRBV gene in the TRB locus can lead to an indirect rearrangement of the D2 and J1 genes, and we conducted a statistical analysis of its proportion. The findings of this study provide a novel perspective

and feasible solutions for further research into the special VDJ recombination patterns on the mammalian TRB locus and the expansion of the CDR3 repertoire by D2-J1 rearrangement.

Materials and methods

The collection of bat, bovine, and mouse samples conducted in our laboratory was approved by the Ethics Committee of Zunyi Medical University (the bat and mouse project were approved under permit number (2018)2-261, and the bovine project was approved under permit number ZMU21-2203-111).

HTS data of human thymus and blood samples

The data are stored in the ENA database (accession number: PRJEB41936) and can be accessed online (https://www.ebi.ac.uk/ena/browser/view/PRJEB41936), as downloaded based on the identification number in Supplementary Table 1. The basic information of data is sourced from the published article "Characterization of human T cell receptor repertoire data in eight thymus samples and four related blood samples" [14].

Thymus (ERZ1694549, ERZ1694551, ERZ1694560, and ERZ1694569) and blood (ERZ1694578, ERZ1694579, ERZ1694580, and ERZ1694581) samples were sourced from four infants aged 7 days, 52 days, 107 days, and 156 days. DNA was extracted from thymocytes and PBMCs. Multiplexed PCR was performed to amplify rearranged TCR-β CDR3 sequences. TRB sequencing was performed using the ImmunoSEQ platform.

scTCR-seq data of human blood samples

The data are stored in the NCBI database (accession number: GSE168859) and can be accessed online (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE168859), as downloaded based on the identification number in Supplementary Table 1. The basic information of data is sourced from the published article "Single-cell RNA sequencing coupled to TCR profiling of large granular lymphocyte leukemia T cells" [13].

Blood samples (GSM5171626, GSM5171627, GSM5171634, GSM5171635, and GSM5171642) were sourced from five healthy donors aged 39 years, 71 years, 55 years, 68 years, and 41 years. Peripheral blood mononuclear cells (PBMCs) were isolated by Ficoll-Hypaque density gradient centrifugation. Sequencing was performed using the Illumina HiSeq 3000 platform.

scTCR-seq data of human thymus samples

The data are stored in the Array Express database (accession number: E-EMTAB-8581) and the NCBI database (accession number: PRJEB36131), which can be accessed online (https://developmentcellatlas.ncl.ac.uk). According to the identification number in Supplementary Table 1, the data can be downloaded from the NCBI database (https://www.ncbi.nlm.nih.gov/Traces/study/?acc=ERP119282&o=acc_s%3Aa). The basic information of the data is sourced from the published article "A cell atlas of human thermal development definitions T cell repertoire formation" [4].

Thymus samples (ERR3804264, ERR3804272, ERR3804262, ERR3804268, and ERR3804281) were collected from five donors aged 11 weeks, 30 months, 10 months, 12 weeks, and 35–40 years. The thymus samples were digested using enzymes until the tissue was completely dissociated. Sequencing was performed using the Illumina HiSeq 4000 platform.

HTS data of mouse spleen and thymus samples

The data are stored in the NCBI database (accession number: PRJNA906203). It can be downloaded based on the identification number in Supplementary Table 1.

Spleen (SRR22437999, SRR22437998, and SRR22437997) and thymus (SRR22438002, SRR22438001, and SRR22438000) samples were collected from 6 different 2-month-old female BALB/c mice. RNA was extracted from splenocytes and thymocytes, and the 5'RACE method was used to amplify rearranged TCR-β CDR3 sequences. TRB sequencing was performed using the Illumina NovaSeq 6000 platform.

scTCR-seq data of mouse peripheral tissue samples

The data are stored in the NCBI database (accession number: GSE168944) and can be accessed online (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi? acc = GSE168944). The data can be downloaded based on the identification number in Supplementary Table 1. The basic information of the data is sourced from the published article "STARTRAC analyses of scTCR-seq data from tumor models reveal T cell dynamics and therapeutic targets" [16].

Total T cells in peripheral tissue samples from mice were isolated from the lymph (GSM5172690, GSM5172691, and GSM5172698), spleen (GSM5172688, GSM5172689, and GSM5172696) and blood (GSM5172686, GSM5172687, and GSM5172694) samples of three C57BL/6 tumor-bearing mice aged 6–8 weeks. The samples were pre-enriched for total T cells using the Stem-Cell Technologies Easy-Sep Mouse T-Cell Isolation Kit. Sequencing was performed using the Illumina HiSeq 4000 platform.

HTS data of Bat spleen samples from our project

The data are stored in the NCBI database (accession number: PRJNA877449), and the raw sequence data can be downloaded according to the identification number in Supplementary Table 1.

Total RNA was extracted from bat splenocytes (SRR21464508, SRR21464509). The 5'RACE method was

used to amplify rearranged TCR-β CDR3 sequences. TRB sequencing was performed using the Illumina HiSeq 1500 platform.

HTS data of rhesus monkey blood sample

The data are stored in the NCBI database (accession number: PRJNA389234), and the raw sequence data can be downloaded according to the identification number in Supplementary Table 1. The basic information of the data is sourced from the published article "A comprehensive profiling of T- and B-lymphocyte receptor repertoires from a Chinese-origin rhesus macaque by high-throughput sequencing" [15].

One blood sample (SRR5647486) from a healthy 5-year-old female Chinese-origin rhesus macaque was collected. Peripheral blood mononuclear cells (PBMCs) were isolated by Ficoll-Paque density gradient centrifugation, and total RNA was extracted from the PBMCs. The 5'RACE method was used to amplify rearranged TCR-β CDR3 sequences. TRB sequencing was performed using the Illumina HiSeq 2000 platform.

scTCR-seq data of rhesus monkey peripheral tissue samples

The data are stored in the NCBI database (accession number: PRJNA746267), and the raw sequence data can be downloaded according to the identification number in Supplementary Table 1. The basic information of the data is sourced from the published article "Single-Cell-Based High-Throughput Ig and TCR Repertoire Sequencing Analysis in Rhesus Macaques" [17].

T cells in the peripheral tissues of rhesus monkeys were collected from PBMCs (SRR15249806, SRR15249810), splenocytes (SRR15249798), and FACS-sorted stimulated nonproliferating (SRR15249814) and proliferating (SRR15249812) T cells. TRB sequencing was performed using the Illumina NextSeq 500 platform.

HTS data of bovine spleen sample from our project

The data are stored in the NCBI database (Accession number: PRJNA908273), and the raw sequence data can be downloaded according to the identification number in Supplementary Table 1.

Total RNA was extracted from bovine splenocytes (SRR22523496). The total RNA was reverse transcribed to cDNA. Multiplexed PCR was performed to amplify rearranged TCR-β CDR3 sequences. TRB sequencing was performed using the MGISEQ-2000RS platform.

Data filtering and sorting

We screened each V-D-J-C by comparing the V, D, J, and C gene sequences of the TRB locus annotated by humans, rhesus monkeys, mice and bovines in IMGT/GENE-DB (https://www.imgt.org/genedb/) [23]with those of the bat TRB locus annotated for the first time by our group [24, 25]. The basic principle of data screening and quality control is that every CDR3 sequence in every sample selected for analysis is the raw data from MiXCR (version 3.0.13), IMGT/HighV-QUEST, Cell Ranger, and immunSEQ software, and the sequence of the V, D, and J subfamily genes of the raw data is determined by homologous comparisons in IMGT. For example, every CDR3 sequence analyzed needs to be clearly identified as D2 gene or D1 gene before it is included in the analysis, and the sequences that cannot be judged as D1 or D2 genes are not included in the analysis. In addition, the analysis of this study focuses on the phenomenon and possible mechanism of special recombination of D2-J1, so all CDR3 sequences (including productive and unproductive) with clear D2 and D1 genes in the TRB CDR3 repertoire of central and peripheral tissues are included in the analysis (mainly to avoid losing the possible contribution of VDJ to the "special recombination of D2-J1" when it randomly participates in the rearrangement) [26], and the biological function of "special recombination of D2-J1" is not analyzed in this project.

To briefly show the phenomenon of "special recombination of D2-J1", we uploaded every sample of mice, bat and bovine sequenced by HTS to the shared database and also included the TRB CDR3 repertoire data shared by several laboratories in the analysis. Analysis of each complete V, D, J, and C sequence, D gene usage and D-J gene recombination pattern (D1-J1, D2-J1, D1–J2, and D2-J2) was conducted in Microsoft Excel (Supplementary Table 1).

We directly selected the V30(V31)-D2-J1 sequence of each sample analyzed from the total Excel table and added it to the supplementary material (Supplementary Table 2). Moreover, we selected one V30(V31)-D2-J1 sequence and one V30(V31)-D1-J1 (or J2) sequence from each sample for comparative analysis and display (Supplementary Table 3). In the text, a specific analysis sequence of V30-D2-J1 and V30-D1-J1 is listed (Fig. 2Q). Notably, in Supplementary Tables 2 and Supplementary Table 3, there are 24 scTCR-seq samples of humans, mice and rhesus monkeys, and 3 HTS samples of bats and bovines from our group (because there are so many CDR3 sequences from the HTS samples of humans and mice, we do not show them directly in the supplementary material, readers and peers can analyze each sample through sharing links).

Statistical analysis and tools

Dissimilarity analysis of each case was conducted using the Mann-Whitney U test and Pearson's chi-squared test. A P value < 0.05 was considered to indicate a statistically significant difference. For human and mouse samples with a sample size of $n \geq 3$, we employed the

Mann-Whitney U test for statistical analysis of usage frequency. For bat, rhesus monkey, and bovine samples with a sample size of $n = 1$, we used Pearson's chi-squared test to statistically analyze the number of unique sequences and presented the data in the form of usage frequency. In human, mouse, bat, and rhesus monkey samples, it was observed that the D1 gene can undergo deletional rearrangement with either the J1 or J2 gene, whereas the D2 gene cannot recombine with the J1 gene through deletional or inversional rearrangement. Consequently, the frequency of D1 gene involvement in recombination is significantly higher than that of the D2 gene. This phenomenon was corroborated in bovine samples, which possess three D-J-C clusters, where the D1 gene was also found to participate in recombination at a higher frequency than the D3 and D2 genes.

GraphPad Prism software 9.3 and HemI 1.0 software were used for data visualization (dot plot and heatmap); MiXCR (version 3.0.13); IMGT/HighV-QUEST; Cell Ranger; and immunSEQ software were used.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-025-11506-z.

Supplementary Material 1

Supplementary Material 2: Supplementary Table 1. Sample information and TRBD-TRBJ pairing recombination data.

Supplementary Material 3: Supplementary Table 2. The reverse V30 (mouse V31)-D2-J1 sequence of each sample was analyzed from the total CDR3 repertoire.

Supplementary Material 4: Supplementary Table 3. One V30(V31)-D2-J1 sequence and one V30(V31)-D1-J1 (or J2) sequence from each sample for comparative analysis and display.Supplementary material. Bioinformatic workflow.

Supplementary Material 5

Acknowledgements

We thank IMGT for sharing the full V-D-J-C annotation for each mammalian TRB locus and NCBI for the availability of the data. We thank Heikkilä, N., Gao, S., Park, J. E., Bhatt, D., Fu, L., and Walsh, E. S. for providing shared sequencing sample data.

Author contributions

Conceptualization: YXS; Methodology: WYJ, WFL, ZH, ML, LJ; Visualization: WYJ, WFL, LJ; Funding acquisition: YXS; Project administration: YXS; Supervision: YXS; Writing– original draft: YXS, WYJ, WFL, LJ; Writing - review & editing: YXS, WYJ, WFL, LJ.

Funding

The research was funded by the National Natural Science Foundation of China (82471630&82160279) and the Guizhou Provincial High-level Innovative Talents Project [No. (2018) 5637].

Data availability

The raw sequence data of bats, bovines, and mice used in this study were obtained from the NCBI database with accession number PRJNA906203 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA906203/), PRJNA877449 (https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA877449), and PRJNA908273 (https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA908273). In addition to our own data collection, we also utilized publicly available data from shared databases, such as EMBL-EBI, NCBI, and IMGT, for information regarding humans (https://www.ebi.ac.uk/ena/browser/view/PRJEB41936, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE168859), rhesus monkeys (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA389234, https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA746267), and mice (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE168944).

Declarations

Ethics approval and consent to participate

All experimental procedures were conducted entirely in accordance with animal welfare guidelines. The collection of samples from bats, mice, and bovines, which was carried out in our laboratory, was approved by the Ethics Committee of Zunyi Medical University. Specifically, the bat and mouse project were approved under permit number (2018)2-261, and the bovine project was approved under permit number ZMU21-2203-111. Additionally, we obtained shared data for human (https://www.ebi.ac.uk/ena/browser/view/PRJEB41936, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE168859, https://www.www.ncbi.nlm.nih.gov/Traces/study/?acc=ERP119282&o=acc_s%3Aa), rhesus monkey (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA389234, https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA746267), and additional mouse (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE168944) samples from a public database. We declare that this study is reported in accordance with ARRIVE guidelines (https://arriveguidelines.org).

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 10 December 2024 / Accepted: 19 March 2025
Published online: 04 April 2025

References

1. Wu Y, Wu F, Ma Q, Li J, Ma L, Zhou H, Gong Y, Yao X. HTS and scRNA-seq revealed that the location and RSS quality of the mammalian TRBV and TRBJ genes impact biased rearrangement. BMC Genomics. 2024;25(1):1010.
2. Jackson AM, Krangel MS. Turning T-cell receptor beta recombination on and off: more questions than answers. Immunol Rev. 2006;209:129–41.
3. Bassing CH, Alt FW, Hughes MM, D'Auteuil M, Wehrly TD, Woodman BB, Gartner F, White JM, Davidson L, Sleckman BP. Recombination signal sequences restrict chromosomal V(D)J recombination beyond the 12/23 rule. Nature. 2000;405(6786):583–6.
4. Park JE, Botting RA, Domínguez Conde C, Popescu DM, Lavaert M, Kunz DJ, Goh I, Stephenson E, Ragazzini R, Tuck E et al. A cell atlas of human thymic development defines T cell repertoire formation. *Science* 2020, 367(6480).
5. Outters P, Jaeger S, Zaarour N, Ferrier P. Long-Range control of V(D)J recombination & allelic exclusion: modeling views. Adv Immunol. 2015;128:363–413.
6. Stubbington MJT, Lonnberg T, Proserpio V, Clare S, Speak AO, Dougan G, Teichmann SA. T cell fate and clonality inference from single-cell transcriptomes. Nat Methods. 2016;13(4):329–32.
7. Ma L, Yang L, Bin S, He X, Peng A, Li Y, Zhang T, Sun S, Ma R, Yao X. Analyzing the CDR3 repertoire with respect to TCR-Beta chain V-D-J and V-J rearrangements in peripheral T cells using HTS. Sci Rep. 2016;6:29544.
8. Smirnova AO, Miroshnichenkova AM, Belyaeva LD, Kelmanson IV, Lebedev YB, Mamedov IZ, Chudakov DM, Komkov AY. Novel bimodal TRBD1-TRBD2 rearrangements with dual or absent D-region contribute to TRB V-(D)-J combinatorial diversity. Front Immunol. 2023;14:1245175.
9. Liu P, Liu D, Yang X, Gao J, Chen Y, Xiao X, Liu F, Zou J, Wu J, Ma J, et al. Characterization of human AβTCR repertoire and discovery of D-D fusion in TCRβ chains. Protein Cell. 2014;5(8):603–15.
10. Wu F, Wu Y, Yao Y, Xu Y, Peng Q, Ma L, Li J, Yao X. The reverse TRBV30 gene of mammals: a defect or superiority in evolution? BMC Genomics. 2024;25(1):705.
11. Lefranc MP, Giudicelli V, Duroux P, Jabado-Michaloud J, Folch G, Aouinti S, Carillon E, Duvergey H, Houles A, Paysan-Lafosse T, et al. IMGT®, the international immunogenetics information system® 25 years on. Nucleic Acids Res. 2015;43(Database issue):D413–422.

12. Lefranc MP. Immunoglobulin and T cell receptor genes: IMGT(®) and the birth and rise of immunoinformatics. Front Immunol. 2014;5:22.
13. Gao S, Wu Z, Arnold B, Diamond C, Batchu S, Giudice V, Alemu L, Raffo DQ, Feng X, Kajigaya S, et al. Single-cell RNA sequencing coupled to TCR profiling of large granular lymphocyte leukemia T cells. Nat Commun. 2022;13(1):1982.
14. Heikkilä N, Kleino I, Vanhanen R, Yohannes DA, Mattila IP, Saramäki J, Arstila TP. Characterization of human T cell receptor repertoire data in eight thymus samples and four related blood samples. Data Brief. 2021;35:106751.
15. Fu L, Li X, Zhang W, Wang C, Wu J, Yang H, Wang J, Liu X. A comprehensive profiling of T- and B-lymphocyte receptor repertoires from a Chinese-origin rhesus macaque by high-throughput sequencing. PLoS ONE. 2017;12(8):e0182733.
16. Bhatt D, Kang B, Sawant D, Zheng L, Perez K, Huang Z, Sekirov L, Wolak D, Huang JY, Liu X et al. STARTRAC analyses of ScRNAseq data from tumor models reveal T cell dynamics and therapeutic targets. J Exp Med 2021, 218(6).
17. Walsh ES, Tollison TS, Brochu HN, Shaw BI, Diveley KR, Chou H, Law L, Kirk AD, Gale M Jr. Single-Cell-Based High-Throughput Ig and TCR repertoire sequencing analysis in rhesus macaques. J Immunol. 2022;208(3):762–71.
18. Zhu L, Peng Q, Li J, Wu Y, Wang J, Zhou D, Ma L, Yao X. scRNA-seq revealed the special TCR B & α V(D)J allelic inclusion rearrangement and the high proportion dual (or more) TCR-expressing cells. Cell Death Dis. 2023;14(7):487.
19. Alamyar E, Duroux P, Lefranc MP, Giudicelli V. IMGT(®) tools for the nucleotide analysis of Immunoglobulin (IG) and T cell receptor (TR) V-(D)-J repertoires, polymorphisms, and IG mutations: IMGT/V-QUEST and IMGT/HighV-QUEST for NGS. Methods Mol Biol. 2012;882:569–604.
20. Hu J, Zhang Y, Zhao L, Frock RL, Du Z, Meyers RM, Meng FL, Schatz DG, Alt FW. Chromosomal loop domains direct the recombination of antigen receptor genes. Cell. 2015;163(4):947–59.
21. Kulski JK, Pfaff AL, Marney LD, Fröhlich A, Bubb VJ, Quinn JP, Koks S. Regulation of expression quantitative trait loci by SVA retrotransposons within the major histocompatibility complex. Exp Biol Med (Maywood). 2023;248(23):2304–18.
22. Kulski JK, Pfaff AL, Koks S. SVA Regulation of Transposable Element Clustered Transcription within the Major Histocompatibility Complex Genomic Class II Region of the Parkinson's Progression Markers Initiative. Genes (Basel) 2024, 15(9).
23. Giudicelli V, Chaume D, Lefranc MP. IMGT/GENE-DB: a comprehensive database for human and mouse Immunoglobulin and T cell receptor genes. Nucleic Acids Res. 2005;33(Database issue):D256–261.
24. Zhou H, Li J, Zhou D, Wu Y, Wang X, Zhou J, Ma Q, Yao X, Ma L. New insights into the germline genes and CDR3 repertoire of the TCRβ chain in chiroptera. Front Immunol. 2023;14:1147859.
25. Zhou H, Ma L, Liu L, Yao X. TR locus annotation and characteristics of rhinolophus ferrumequinum. Front Immunol. 2021;12:741408.
26. Gopalakrishnan S, Majumder K, Predeus A, Huang Y, Koues OI, Verma-Gaur J, Loguercio S, Su AI, Feeney AJ, Artyomov MN, et al. Unifying model for molecular determinants of the preselection Vβ repertoire. Proc Natl Acad Sci U S A. 2013;110(34):E3206–3215.

Publisher's note