 DOI: https://doi.org/10.5114/ceji.2018.80049

Comparison of T-cell receptor β chain variable region 23-1 open reading frame and 3-1 F CDR3 repertoire in cDNA and gDNA from peripheral blood mononuclear cells of healthy volunteers by high-throughput sequencing

XIAOYAN HE[1], BIN SHI[2*], LONG MA[1*], RUI MA[1], AIHUA PENG[1], SUHONG SUN[3], XINSHENG YAO[1]*

[1]Department of Immunology, Research Center for Medicine & Biology, Innovation & Practice Base for Graduate Students Education, Zunyi Medical University, Zunyi, China
[2]Department of Laboratory Medicine, Zunyi Medical University, Zunyi, China
[3]Department of Breast Surgery, the First Affiliated Hospital of Zunyi Medical University, Zunyi, China
[*]*These authors contributed equally to this work.*

Abstract

The TRBV germline gene can be classified into three types: open reading frame (ORF), functional gene (F), and pseudogene (P); however, the differences and connections among ORF, functional gene, and pseudogene rearrangements and expression characteristics remain unclear. Here, we compared the characteristics of the CDR3 repertoire that was rearranged by TRBV23-1(ORF) and TRBV3-1(F) in four cDNA samples and six gDNA samples from healthy volunteers. In this study, we show that the frequencies of in-frame sequences in the TRBV23-1 complementarity determining region 3 (CDR3) repertoire were significantly lower than those in the TRBV3-1 CDR3 repertoire. The TRBV23-1 CDR3 repertoire, which differed from the TRBV3-1 in-frame and out-of-frame CDR3 repertoire, consisted of 1/3 in-frame sequences and 2/3 out-of-frame sequences. The usage of TRBD1 was higher than that of TRBD2 in the TRBV23-1 in-frame and out-of-frame CDR3 repertoire. In four cDNA samples from PBMCs, the TRBV23-1 in-frame and out-of-frame CDR3 repertoire showed a longer N2 relative to that of N1. Therefore, our results demonstrate that the mechanisms of rearrangement in the TRBV23-1 and TRBV3-1 CDR3 repertoire were different, and thus offered new ideas and methods to resolve the discrepancies in reports on the functionality of TRBV23-1 and to better understand the mechanisms underlying VDJ recombination.

Key words: *T-cell receptor, T-cell receptor β variable, complementarity-determining region 3 repertoire, open reading frame, high-throughput sequencing.*

(Centr Eur J Immunol 2018; 43 (3): 295-305)

Introduction

T-cell receptors (TCRs) are either TCR αβ or TCR γδ heterodimers that are formed by four types of peptide chains α, β, γ, and δ. Each polypeptide chain in the thymus rearranges from the TCR germline gene to form a diverse CDR3 repertoire. According to the definition and annotation of the IMGT and EMBL, the TCR germline gene entity can be defined according to three types: open reading frame (*ORF*), functional gene (F) and pseudogene (P). An open reading frame is a portion of a gene's sequence that contains a sequence of bases, uninterrupted by stop sequences, that could potentially encode a protein. A functional gene's coding region has an *ORF* without a stop codon, and for which there is no described defect in the splicing sites, recombination signals, or regulatory elements. Pseudogenes are the remnants of genomic sequences of genes which are no longer functional; they are frequent in most eukaryotic genomes, and an important resource for comparative genomics [1-4].

In the process of TCR creation, according to the principle of allelic exclusion, the β chain is rearranged earlier than the α chain, so detecting a β chain could reflect the

Correspondence: Dr. Xinsheng Yao, Department of Immunology, Zunyi Medical University, 563000 Zunyi, China, e-mail: immunology01@126.com
Submitted: 27.03.2017; Accepted: 31.07.2017

character of TCR. According to the homology of the human *TRBV* gene, 65 *TRBV* genes can be divided into 24 conservative families, which can cover more than 85% of human TCR β chain genes. The complementarity-determining region 3 (CDR3) of a specific TCR β chain could be amplified through designing sense primers and *TRBC* antisense primers. According to the definition and annotation of the IMGT, the human TCR β locus consists of 48 *TRBV* functional genes, 19 *TRBV* pseudogenes, and 7 *ORFs*. However, some *TRBV* genes are difficult to classify. For example, *TRBV7-3* is classified into either a functional gene or an *ORF*, namely, *TRBV7-3*01*, *TRBV7-3*04* and *TRBV7-3*05* (accession numbers X61440, X74843 and M13550, respectively), whereas *TRBV7-3*02* and *TRBV7-3*03* are classified into *ORFs* (accession numbers M97943 and AF009660). *TRBV16* is classified into either a functional gene or a pseudogene: *TRBV16*01* and *TRBV16*03* are classified into a functional gene (accession numbers L26231 and L26054), whereas *TRBV16*02* is classified into a pseudogene (accession number U03115) [3, 4]. Identification of the functions of these *TRBVs* is highly complex.

There is currently a debate on how to define and classify human *TRBV19S1* (*TRBV23-1*), and the human T-cell receptor β chain variable region (clone HVB10.1) was initially defined as a pseudogene, then later described as an *ORF* because nucleotide bases *GT* were replaced by *AT* in the donor splice site, whereas only a portion of the V-gene has been identified [3]. In 1996, Currie *et al*. [5] discovered a single extra base in the *TRBV19S1* leader sequence, which leads to reading frame changes after mRNA splicing, thereby resulting in premature termination of mRNA translation. Therefore, the *TRBV19S1* would be non-functional after rearrangement. However, in 1999, while conducting a research study involving children infected with HIV, Than *et al*. [6] discovered that *TRBV19S1* had dominant usage expression, thus prompting him to suggest that these T-cells participate in the immune response. In 1999, Manfras *et al*. [7] sequenced more than 500 *TRBV20S1* and *TRBV19S1* genes. Their study identified different CDR3 lengths between the productive sequences (*TRBV20S1*) and in-frame nonproductive sequences (*TRBV19S1*), although no significant differences in *TRBD* and *TRBJ* usage and nucleotide additions in junctional sequences were observed.

These analyses were based on the intrinsic defects in gene sequences or by cloning a specific CDR3 region to indirectly demonstrate the in-frame and out-of-frame rearrangements of *TRBV23-1*. The mechanism underlying the participation of *TRBV23-1* in the VDJ rearrangement, and the characteristics of CDR3 repertoires rearranged by *TRBV23-1* as well as differences in CDR3 repertoires rearranged by a *TRBV* functional gene have not been elucidated to date. In this study, we compared the characteristics of the CDR3 repertoire that has been rearranged by *TRBV23-1* (*ORF*) with that by *TRBV3-1* (F) in the cDNA samples from four healthy volunteers by laser capillary electrophoresis scanning and 454 GS FLX high-throughput sequencing (HTS). We also compared the characteristics of the CDR3 repertoire that has been rearranged by *TRBV23-1* with that by *TRBV3-1* in the gDNA samples from six healthy volunteers by Illumina sequencing. Our results show that the frequencies of in-frame sequences in the *TRBV23-1* CDR3 repertoire were significantly lower than those in the *TRBV3-1* CDR3 repertoire. The *TRBV23-1* CDR3 repertoire, which differed from the *TRBV3-1* in-frame and out-of-frame CDR3 repertoire, consisted of 1/3 in-frame sequences and 2/3 out-of-frame sequences. The usage of *TRBD1* was higher than that of *TRBD2* in the *TRBV23-1* in-frame and out-of-frame CDR3 repertoire. In four cDNA samples from PBMCs, the *TRBV23-1* in-frame and out-of-frame CDR3 repertoire showed a longer N2 relative to that of N1. The mechanisms of rearrangement in the *TRBV23-1* (*ORF*) and *TRBV3-1* (F) CDR3 repertoire were different.

Material and methods

The CDR3 repertoire of *TRBV23-1* and *TRBV3-1* in the cDNA samples from PBMCs of four healthy volunteers

Ten-milliliter peripheral blood samples were collected from each of four healthy volunteers (H-1: male, 14 years old; H-2: female, 40 years old; H-3: male, 13 years old; H-4: male, 13 years old). H-3 and H-4 were a pair of twins. H-2 was a mother of twins. H-1 was the same age as the twins. Peripheral blood mononuclear cells (PBMCs) were separated by Ficoll-Hypaque density gradient centrifugation. The study protocol was approved by the local ethics committee. The volunteers had signed the informed consent, and in the case of the younger subjects under the age of 15, they also had the signed informed consent from their parents. Total RNA was extracted from each of the four PBMCs samples, and cDNA was prepared. Using the *TRBC* downstream primer with the FAM label and the individual *TRBV* upstream primers, the *TRBV* CDR3 regions were amplified by PCR (1 cycle of 94°C for 3 min, 1 cycle; 35 cycles each of 94°C for 60 s, 55°C for 60 s, and 72°C for 60 s; 1 cycle of 72°C for 10 min). The PCR products were stored at –20°C.

Laser capillary electrophoresis DNA scan (ABI3730) was used to analyze each PCR product of the *TRBV* gene CDR3. Peak Scanner Software v1.0 was used to analyze the CDR3 spectratyping. The *TRBV23-1* and *TRBV3-1* CDR3 PCR products of 4 healthy volunteers were chosen for cloning and sequence analysis of in-frame or out-of-frame rearrangements that consisted of the CDR3-region (Fig. 1).

Using the four sets of upstream and downstream primers with specific base labels (Table 1), the CDR3 regions of the *TRBV* genes were separately amplified by polymerase chain reaction (PCR; using the same conditions as earlier described). Roche 454 GS FLX high-throughput

sequencing was used to determine the genomic sequences of the CDR3 repertoire (the PCR products) of each sample by Majorbio Bio-pharm Technology Co., Ltd. (Shanghai, China). IMGT High V-QUEST was used to screen the original HTS data (Fig. 1). First, "Unproductive", "Warnings", "Unknown functionality", and "No results" sequences were removed. Then, the in-frame and out-of-frame CDR3 sequences with a V-gene consistency rate of > 90% were selected for analysis [8-11].

The CDR3 repertoire of *TRBV23-1* and *TRBV3-1* in the gDNA samples from PBMCs of six healthy volunteers

Ten-milliliter samples of peripheral blood were collected from each of six healthy volunteers (H-5: male, 46 years old; H-6: male, 46 years old; H-7: male, 35 years old; H-8: male, 35 years old; H-9: male, 22 years old; H-10: female, 22 years old). The study protocol was approved by the lo-

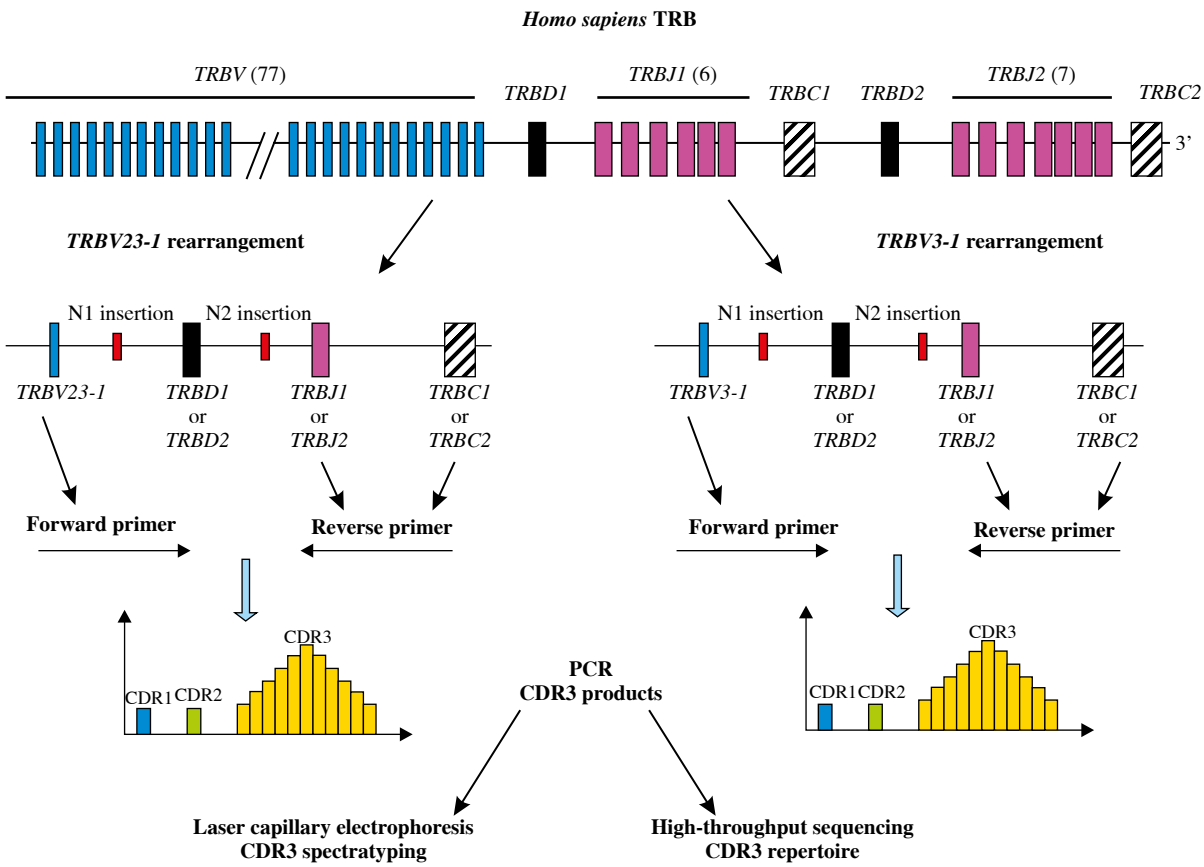

Fig. 1. The *TRBV* CDR3 rearrangements principle and the experimental procedure of the CDR3 repertoire spectratyping analysis by laser capillary electrophoresis and HTS

Table 1. Primer sequences of TCR-β chain *TRBV3-1* and *TRBV23-1* and *TRBC* in human for 454 GS FLX high-throughput sequencing

	TRBV3-1	*TRBV23-1*	*TRBC*
H-1	GAGTGCGTCT ctaaatctccagacaaagctcac	GAGTGCGTCT caatgccccaagaacgcaccctgc	AGACGCACTC ttctgatggctcaaacac
H-2	CTACAGTGCT ctaaatctccagacaaagctcac	CTACAGTGCT caatgccccaagaacgcaccctgc	AGCACTGTAG ttctgatggctcaaacac
H-3	CGTGTCTGAT ctaaatctccagacaaagctcac	CGTGTCTGAT caatgccccaagaacgcaccctgc	ATCAGACACG ttctgatggctcaaacac
H-4	CTCGCGATAT ctaaatctccagacaaagctcac	CTCGCGATAT caatgccccaagaacgcaccctgc	ATATCGCGAG ttctgatggctcaaacac

cal ethics committee. The subjects had signed the informed consent. PBMCs were isolated by Ficoll-Hypaque density gradient centrifugation and genomic DNA was extracted from the six samples, yielding the following individual total amounts: H-5, 1,800 ng; H-6, 1,990 ng; H-7, 3,350 ng; H-8, 5,090 ng; H-9, 2,380 ng; and H-10, 1,730 ng.

The CDR3 repertoire of human TCR beta chain preparation and Illumina HTS sequencing of the gDNA samples were completed by Adaptive Biotechnologies Corp [12]. Adaptive Biotechnologies has developed a novel method that amplifies rearranged TCR CDR3 sequences and exploits the capacity of high-throughput sequencing technology to sequence tens of thousands of TCR CDR3 chains simultaneously (Fig. 1). Because the technology utilizes gDNA, the frequency of sequenced CDR3 chains is representative of the relative frequency of each CDR3 sequence in the starting population of T-cells [13-15]. The CDR3 sequences were screened and analyzed using the ImmuneSEQ assay [16] and the IMGT High V-QUEST [3]. The CDR3 sequences of *TRBV3-1* and *TRBV23-1* were selected from all the CDR3 repertoire sequences for analysis. First, 'Unproductive', 'Warnings', 'Unknown functionality', and 'No results' sequences were removed. Then, the in-frame and out-of-frame CDR3 sequences with a V-gene consistency rate of > 90% were selected for analysis [14, 15].

Experimental reagents and statistical analysis

Total RNA and total DNA extraction reagent kits, the gel extraction kit, the PCR purification kit, the Taq PCR master mix, and the DNA extraction and freeze-dry kit were obtained from Qiagen (Germany). DNA marker, DNA loading buffer, and cDNA synthesis kits were obtained from MBI Fermentas (Canada).

TRBD gene and *TRBJ* gene usage were compared using the χ^2 test; N1 and N2 lengths of nucleotides were compared using ANOVA. $p < 0.05$ was considered statistically significant. All statistically significant differences are indicated; $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$ [12, 17-19].

Results

CDR3 spectratyping of *TRBV23-1* and *TRBV3-1* in cDNA samples from PBMCs of four healthy volunteers

CDR3 spectratyping of *TRBV23-1* showed a 1-bp or 2-bp (out-of-frame) and 3-bp (in-frame) interval length and peak. On the other hand, the CDR3 spectratyping of *TRBV3-1* showed a 3-bp (in-frame) interval length and peak (Fig. 2, Table 2). In the nine cloned sequencing results of the *TRBV23-1* CDR3 region PCR products that were randomly collected from H-4, four were in-frame with 3-bp insertions, three were out-of-frame with a 1-bp insertion, and two were out-of-frame with 2-bp insertions (Table 3).

CDR3 repertoire characterization of *TRBV23-1* and *TRBV3-1* in healthy volunteers by HTS

The cDNA samples were extracted from PBMCs of four healthy volunteers. A total of 509 reads of the *TRBV23-1* CDR3 complete sequences of the four samples, including 142 in-frame and 367 out-of-frame sequences,

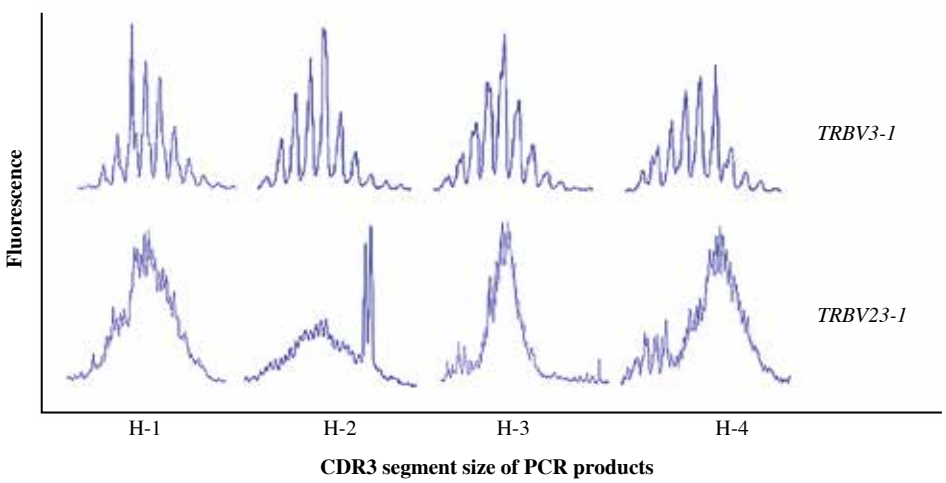

Fig. 2. The CDR3 repertoire (PCR products) spectratyping of *TRBV3-1* and *TRBV23-1* in the cDNA samples from PBMCs of four healthy volunteers by capillary electrophoresis DNA scanning. Note: *TRBV3-1* expressed as regular CDR3 spectratype with 3-bp intervals; *TRBV23-1* presented as irregular CDR3 spectratype of small peaks and sloping peaks with 1-bp intervals

Table 2. The CDR3 repertoire (segment sizes of PCR products) of *TRBV23-1* and *TRBV3-1* in the cDNA samples from PBMCs of four healthy volunteers

***TRBV3-1* (bp)**				***TRBV23-1* (bp)**			
H-1	**H-2**	**H-3**	**H-4**	**H-1**	**H-2**	**H-3**	**H-4**
185	186	189	189	180	189190	170171	171 173
188	189	192	192	183184	193	173	174 176
191	192	195	195	186187	196197	176	177 179
194195	195	198	198	189190191	198199200	179180	180181
197	198	201	201	192	202203	182183	183184
200	201	204	204	195196197	206	185186187	
203	204	207	207	198199200	207208	188189190	189190191
206	207	210	210	201202203		191192193	192193
209	210	213	213	204205206		194195196	196197
212	213	216	216	209		197198199	198199200
215	216	219	219	212		200201202	201
	219					203204205	204205
						206 208	207208
						209210	
						212213	

Only the PCR products with > 200 fluorescent units bands were analyzed by capillary electrophoresis DNA scanning.
The black numbers of TRBV3-1 correspond to CDR3 PCR products of 3bp interval (according to DNA marker in the experiment), which can show the standard CDR3 spectrum peak pattern having 3bp interval in Figure 2.
The red numbers of TRBV23-1 correspond to CDR3 PCR products of 3bp interval (according to DNA marker in the experiment). TRBV23-1 black numbers corresponding to CDR3 PCR are larger than the red numbers 1bp. TRBV23-1 blue numbers are larger than the red numbers 2bp, which can show an abnormal CDR3 spectrum peak pattern having 1bp interval in Figure 2.

were identified. A total of 2,678 reads of the *TRBV3-1* CDR3 complete sequences were identified, which included 1,322 in-frame and 1,356 out-of-fame sequences (Table 4). The gDNA samples were extracted from PBMCs of six healthy volunteers. A total of 1,091 CDR3 complete sequence reads were detected in *TRBV23-1*, which included 350 in-frame reads and 741 out-of-frame reads. A total of 5,043 CDR3 complete sequence reads were detected in *TRBV3-1*, which included 4,254 in-frame reads and 789 out-of-frame reads (Table 5).

For both cDNA and gDNA, the CDR3 total sequences of *TRBV23-1* were significantly smaller than those of the CDR3 sequences of *TRBV3-1* (Tables 4 and 5, Fig. 3); the total number of CDR3 sequences of *TRBV23-1* out-of-frame was significantly higher than that of the *TRBV23-1* in-frame sequences (Tables 4 and 5, Fig. 4).

Comparative analysis of the CDR3 repertoire with *TRBD* and *TRBJ* usage of *TRBV23-1* and *TRBV3-1* by HTS

For both cDNA and gDNA, the usage of *TRBD1* in the CDR3 repertoire of *TRBV3-1* in-frame sequences was higher than that of *TRBD2* in terms of mean percentage; however, it was not statistically significant ($p > 0.05$) (Fig. 5). The usage of *TRBD1* in the CDR3 repertoire of *TRBV23-1* in-frame and out-of frame sequences was higher than that of *TRBD2* and was statistically significant ($p < 0.001$) (Fig. 5). The usage of *TRBJ2* was significantly higher than that of *TRBJ1* in terms of the CDR3 repertoire of all *TRBV23-1* and *TRBV3-1* ($p < 0.001$) (Fig. 6).

The usage of *TRBD1* in the CDR3 repertoire of *TRBV3-1* out-of-frame sequences was higher than that of *TRBD2* and was statistically significant ($p < 0.01$) in cDNA samples (Fig. 5A). The usage of *TRBD1* in the CDR3 repertoire of *TRBV3-1* out-of-frame sequences was lower than that of *TRBD2* and was statistically significant ($p < 0.001$) for the gDNA samples (Fig. 5B).

Comparative analysis of the CDR3 repertoire lengths with N1 an N2 of *TRBV23-1* and *TRBV3-1* by HTS

For the cDNA samples, the CDR3 repertoire of *TRBV3-1* in-frame sequences showed that N2 was longer than N1 in terms of mean length; however, this was not

Table 3. The *TRBV23-1* in-frame and out-of-frame CDR3 gene and amino acid sequences in the cDNA samples from PBMCs of volunteer H-4

Number *TRBV23-1*			TCR V β CDR3 junctional sequence (BV-N1-BD-N2-BJ)	*TRBJ*	
1/9	ALYL	CASS	Q T N Y S N Q P Q H (in-frame) CAG ACA AAT TAT AGC AAT CAG CCC CAG CAT (in-frame)	FG	DGTR
1/9	ALYL	CASR	H D S N Y G Y T (in-frame) CAC GAC AGT AAC TAT GGC TAC ACC (in-frame)	FG	SGTR
1/9	ALYL	CASS	Q V G S G N T I Y (in-frame) CAG GTG GGA AGT GGA AAC ACC ATA TAT (in-frame)	FG	EGSW
1/9	ALYL	CASS	Q S P W D L A N Y G Y T (in-frame) CAA AGC CCC TGG GAT TTA GCT AAC TAT GGC TAC ACC (in-frame)	FG	SGTR
1/9	TVSP	VASS	Q * W G K R A V (out-of-frame) CAA TAG TGG GGA AAG CGA GCA GTA C (out-of-frame)	FG	PGTR
1/9	ALYL	CASR	Q S G V V E T S S X (out-of-frame) CAA TCG GGG GTC GTA GAA ACG AGC AGT AG (out-of-frame)	FG	PGTR
1/9	ALYL	CASS	Q S T S G P * T P G S C X (out-of-frame) CAA TCG ACT AGC GGC CCC TAA ACA CCG GGG AGC TGT TT (out-of-frame)	FG	EGSR
1/9	ALYL	CASS	Q S Y H L G Q G E G R D P V X (out-of-frame) CAATCGTATCACCTGGGACAGGGAGAGGGAAGAGACCCAGTAC (out-of-frame)	FG	PGTR
1/9	ALYL	CASS	Q D * R E D R G A X (out-of-frame) CAA GAC TAG CGG GAG GAC CGG GGA GCT T (out-of-frame)	FG	EGSR

The TRBV23-1 CDR3 gene sequences in the PCR products were sequenced by clonal sequencing

Table 4. The CDR3 repertoire total gene sequence and proportion of *TRBV3-1* and *TRBV23-1* in-frame and out-of-frame in the cDNA samples from PBMCs of four healthy volunteers by 454 GS FLX HTS

Samples name		*TRBV3-1* CDR3 repertoire		*TRBV23-1* CDR3 repertoire		
	Total sequences	In-frame (%)	Out-of-frame (%)	Total sequences	In-frame (%)	Out-of-frame (%)
H-1	447	222 (49.7)	225 (53.3)	107	33 (30.8)	74 (69.2)
H-2	787	379 (48.2)	408 (51.8)	152	31 (20.4)	121 (79.6)
H-3	493	261 (52.9)	232 (47.1)	89	29 (32.6)	60 (67.4)
H-4	951	460 (48.4)	491 (51.6)	161	49 (30.4)	112 (69.6)
Total	2678	1322 (49.4)	1356 (50.6)	509	142 (27.9)	367 (72.1)

Table 5. The CDR3 repertoire total gene sequence and proportion of *TRBV3-1* and *TRBV23-1* in-frame and out-of-frame in the gDNA samples from PBMCs of six healthy volunteers by Illumina HTS

Sample name		*TRBV3-1* CDR3 repertoire		*TRBV23-1* CDR3 repertoire		
	Total sequences	In-frame (%)	Out-of-frame (%)	Total sequences	In-frame (%)	Out-of-frame (%)
H-5	757	625 (82.6)	132 (17.4)	161	48 (29.8)	113 (70.2)
H-6	747	615 (82.3)	132 (17.7)	150	44 (29.3)	106 (70.7)
H-7	1318	1157 (87.8)	161 (12.2)	259	95 (36.7)	164 (63.3)
H-8	968	856 (88.4)	112 (11.6)	200	66 (33)	134 (67)
H-9	428	341 (79.7)	87 (20.3)	111	36 (32.4)	75 (67.6)
H-10	825	660 (80.0)	165 (20.0)	210	61 (29.0)	149 (71.0)
Total	5043	4254 (84.4)	789 (15.6)	1091	350 (32.1)	741 (67.9)

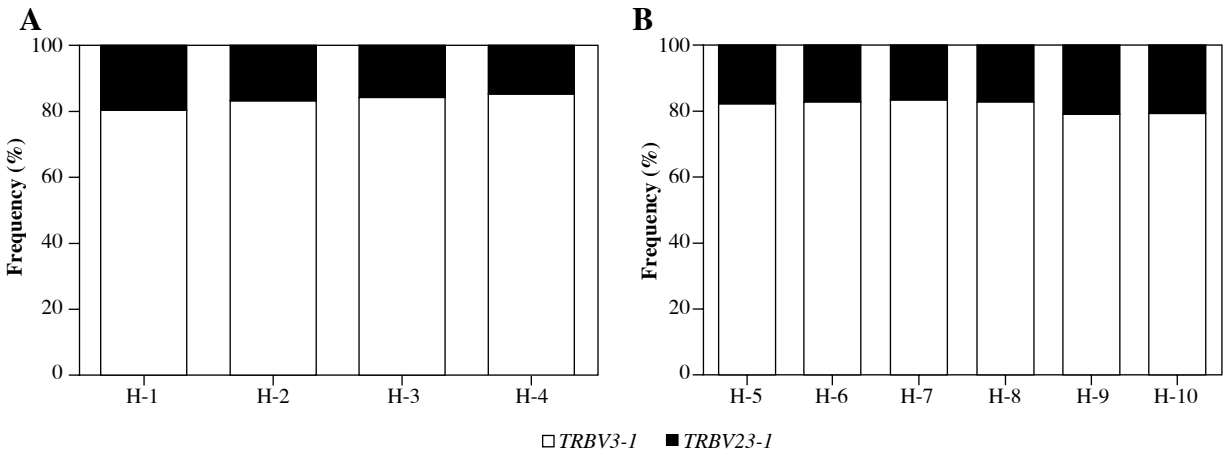

Fig. 3. The CDR3 repertoire sequences proportion of *TRBV3-1* and *TRBV23-1* by HTS. **A)** The cDNA samples from PBMCs of four volunteers. **B)** The gDNA samples from PBMCs of six volunteers

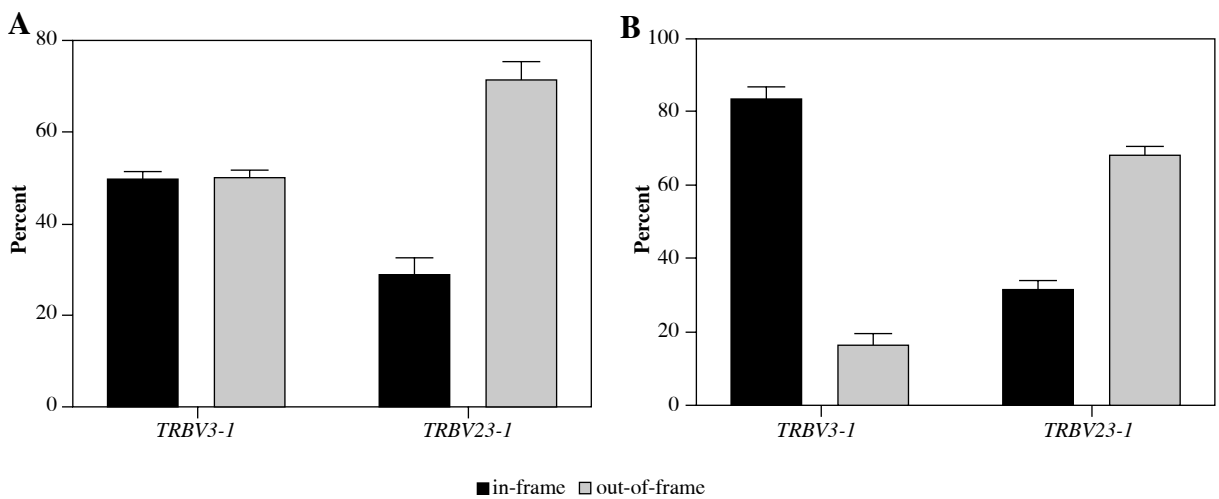

Fig. 4. The CDR3 repertoire sequences proportion of *TRBV3-1* and *TRBV23-1* in-frame and out-of-frame by HTS. **A)** The cDNA samples from PBMCs of four volunteers. **B)** The gDNA samples from PBMCs of six volunteers

statistically significant ($p > 0.05$). The CDR3 repertoire of *TRBV23-1* in-frame sequences showed that N2 was longer than N1 and was statistically significant ($p < 0.05$). The CDR3 repertoire of *TRBV3-1* and *TRBV23-1* out-of-frame sequences consistently showed that N2 was longer than N1 and was statistically significant ($p < 0.001$) (Fig. 7A). However, the CDR3 repertoire of all *TRBV23-1* and *TRBV3-1* sequences showed no differences in the lengths of N1 and N2 for the gDNA samples ($p > 0.05$) (Fig. 7B).

Discussion

Several studies have shown that human *TRBV19S1* (*TRBV23-1*) is a pseudogene [5, 7]; however, it is also possible that this gene may have undergone rearrangements as a functional TCR [6]. The IMGT and EMBL initially defined the T-cell receptor beta chain variable region (clone HVB10.1) as a pseudogene, which was later described as an *ORF* because the nucleotide bases *GT* were replaced by *AT* in the donor splice site, whereas only a portion of the V-GENE has been identified [3, 4]. The characteristics of CDR3 repertoire rearrangement by *TRBV23-1* and the differences in the function of *TRBV* genes require further analysis.

We compared the CDR3 repertoire rearranged by *TRBV23-1* to those rearranged by a random selecting functional gene (*TRBV3-1*) in cDNA and gDNA samples from PBMCs of healthy volunteers. We determined that: (1) the CDR3 spectratyping of *TRBV23-1* (cDNA samples from four volunteers) showed with 1-bp or 2-bp (out-of-frame) and 3-bp (in-frame) interval lengths and peaks (Fig. 2, Table 2). However, the CDR3 spectratyping of *TRBV3-1* showed with 3-bp (in-frame) interval length and

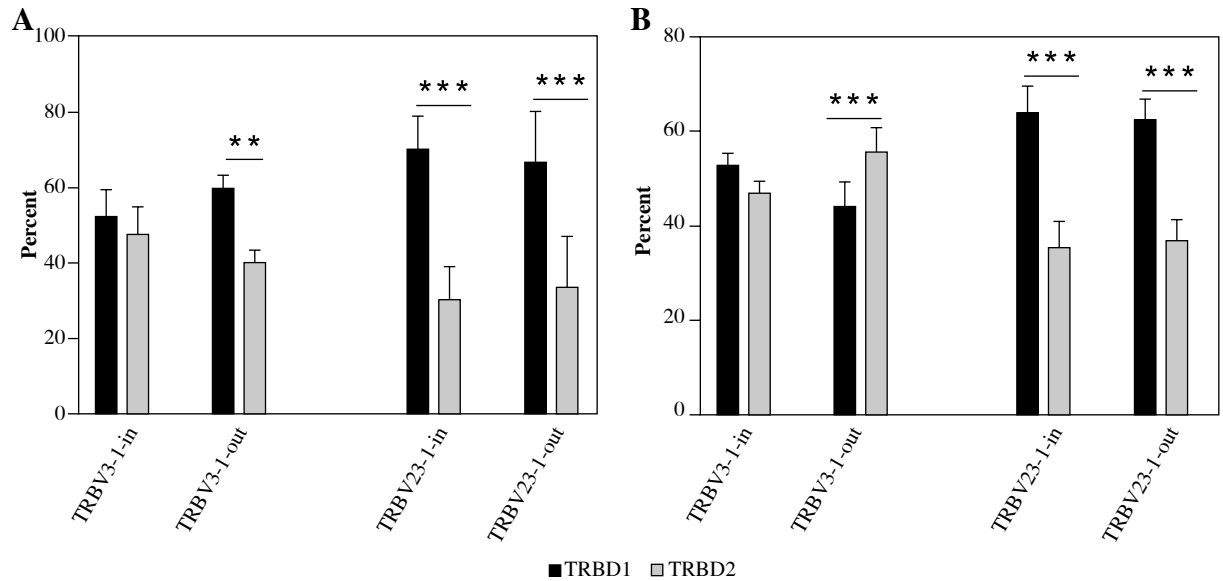

Fig. 5. The CDR3 repertoire with *TRBD1* and *TRBD2* usage of *TRBV3-1* and *TRBV23-1* in-frame and out-of-frame by HTS **A)** The cDNA samples from PBMCs of four volunteers. **B)** The gDNA samples from PBMCs of six volunteers. Note: The *p*-values were determined using the χ^2 test. All statistically significant differences are indicated. $^{**}p < 0.01$, $^{***}p < 0.001$

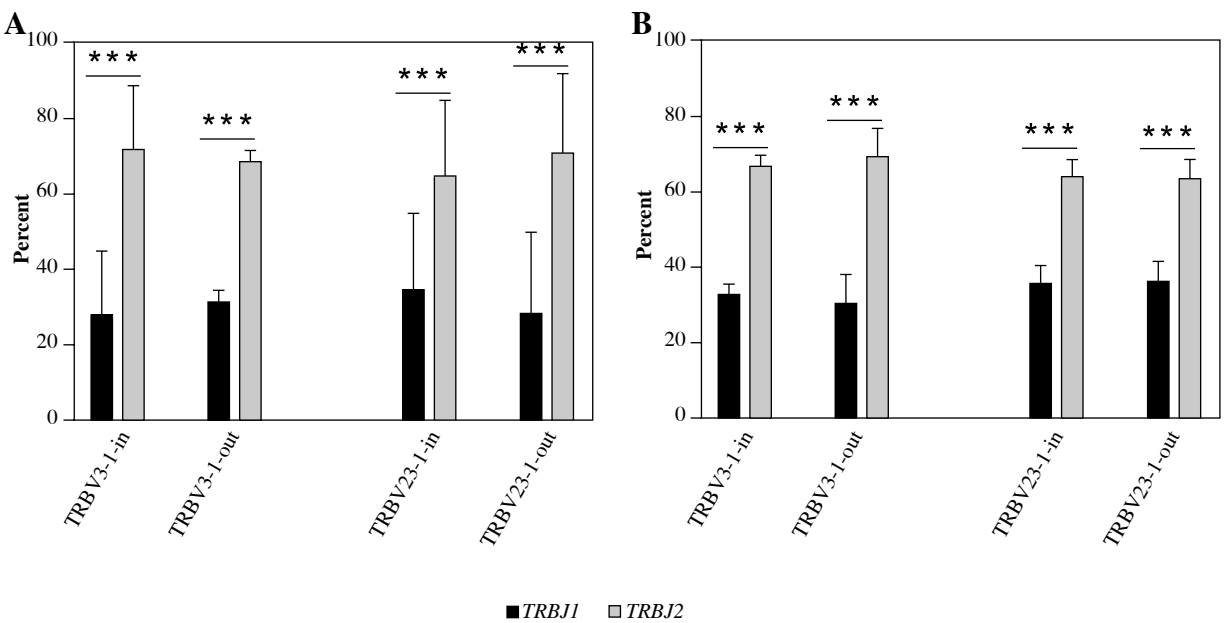

Fig. 6. The CDR3 repertoire with *TRBJ* usage of *TRBV3-1* and *TRBV23-1* in frame and out frame by HTS. **A)** The cDNA samples from PBMCs of four volunteers. **B)** The gDNA samples from PBMCs of six volunteers. Note: The *p*-values were determined using the χ^2 test. All statistically significant differences are indicated. $^{***}p < 0.001$

peak (Fig. 2, Table 2); (2) the *TRBV23-1* CDR3 region PCR products were randomly selected from H-4 (cDNA samples), four sequences in the CDR3 region were in-frame with 3-bp insertions, three sequences in the CDR3 region were out-of-frame with a 1-bp insertion, and two were out-of-frame with 2-bp insertions (Table 3); (3) the CDR3 repertoire of *TRBV23-1* was expressed at significantly lower levels than that of *TRBV3-1* (Tables 4 and 5);

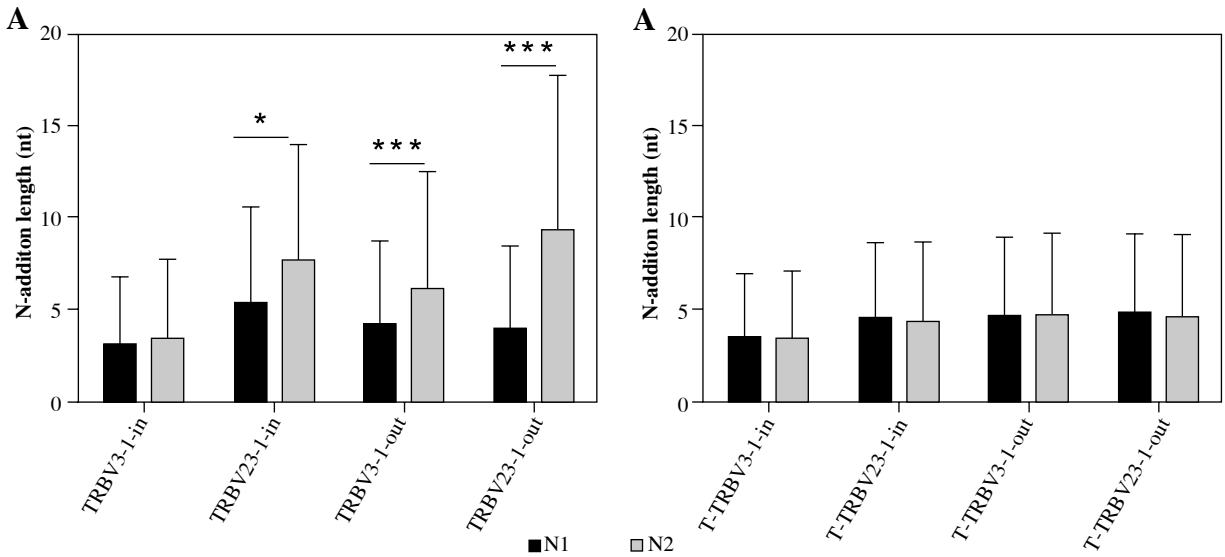

Fig. 7. The CDR3 repertoire N1 and N2 lengths of *TRBV3-1* and *TRBV23-1* in-frame and out-of-frame by HTS. **A)** The cDNA samples from PBMC of four volunteers. **B)** The gDNA samples from PBMC of six volunteers. Note: The *p*-values were determined by using ANOVA. All statistically significant differences are indicated. $^{*}p < 0.05$, $^{***}p < 0.001$

(4) the CDR3 repertoire of *TRBV23-1* consisted of 2/3 out-of-frame sequences and 1/3 in-frame sequences, which was exactly the opposite to that observed of the *TRBV3-1* in-frame and out-of-frame CDR3 repertoire. These findings suggest that the CDR3 repertoire that was rearranged by *TRBV23-1* mainly emerged from random rearrangements (2/3 sequences were 1-bp and 2-bp insertions, whereas 1/3 of the sequences harbored 3-bp insertions), and the CDR3 repertoire rearranged by *TRBV23-1* may have rendered non-functionality, as well as in the absence of positive and negative selection. The CDR3 repertoire rearranged by *TRBV3-1* was mainly composed of an in-frame CDR3, and a few out-of-frame CDR3 regions. In the gDNA samples, the percent of in-frame CDR3 regions was > 2/3, and the percentage of in-frame CDR3 regions was < 2/3, which may be attributable to cDNA samples that were not amplified by multiplex PCR. In our other experiments on functional TCR repertoires, the percentage of in-frame CDR3 in both human and mice was > 2/3 (cDNA samples and gDNA samples). The results from these massive CDR3 repertoire sequences from *TRBV23-1* and *TRBV3-1* are in agreement with those of Burkhard's research, which used a small amount of cloning sequencing data to compare and analyze *TRBV19* and *TRBV20* [7].

The maintenance of *TRBV* functional gene and *ORF* (or pseudogene) out-of-frame CDR3 sequences at a certain proportion in peripheral blood might be possibly due to the TCR β chain undergoing rearrangements, with the V-D-J combination rearrangement occurring first in one chromosome. Theoretically, the rearrangement of the *TRBV* functional gene, pseudogene and *ORF* is random [20]. In the CDR3 repertoire formed by rearrangements, the CDR3 region insertion/splicing is random, where the ratio of 1-bp, 2-bp, and 3-bp change is 1/3, respectively. Because *TRBV ORFs* or pseudogenes do not have functional protein expression after rearrangement, the cells that undergo *TRBV ORF* or pseudogene rearrangements would either die or initiate a second V-D-J rearrangement in another chromosome. If the second rearrangement were successful, the cells would enter the subsequent selection and maturation stages. If not, then the cells would eventually die. Hence, because *TRBV ORFs* or pseudogenes are not functionally expressed after rearrangement, these will not be subjected to positive and negative selections to mature. In theory, the total CDR3 repertoire of the *TRBV ORF* or pseudogene is eventually formed after initiation of rearrangements in another chromosome and successful functional selections. Therefore, the total CDR3 repertoire of the *TRBV ORF* or pseudogene follows the rearrangement rules of base insertion/deletion, thereby yielding the in-frame ratio of approximately 1/3 (3-bp insertion/deletion) and the out-of-frame ratio of approximately 2/3 (1-bp and 2-bp insertion/deletion). At the same time, because *TRBV ORF* or pseudogenes have no functionality after rearrangement, the formation of their peripheral CDR3 repertoire depends on successful rearrangements involving the *TRBV* functional gene in another chromosome. Consequently, the number of CDR3 sequences of the *TRBV ORF* or pseudogene with no functionality should be much lower than that of *TRBV* functional genes. However, *TRBV* functional genes would only mature after random rearrangement and positive and negative selections. Theoretically, the ratio of *TRBV* func-

tional gene in-frame CDR3 and out-of-frame CDR3 sequences is extremely complex and closely related to individualized selection [5, 21]. The pattern of rearrangement and selection of *TRBV23-1* CDR3 repertoire detected by HTS is consistent with the theory, and these results suggest that we can detect the CDR3 repertoires of the *ORF* or pseudogene by HTS as a control to analyze the mechanism underlying rearrangements and positive/negative selection of functional genes.

We then compared the usage of *TRBD* and *TRBJ* in *TRBV23-1* and *TRBV3-1 CDR3* repertoires, and the CDR3 repertoire with *TRBD1* usage in the *TRBV23-1* in-frame and out-of-frame was higher than that of *TRBD2*, whereas *TRBV3-1* in-frame and out-of-frame CDR3 repertoires differed. These findings suggest that the mechanism underlying *TRBD* rearrangement with *TRBV3-1* (F) and *TRBV23-1* (*ORF*) were different, and subsequently yielded variable CDR3 repertoires. However, the usage of *TRBJ* was similar among these repertoires, and that of *TRBJ2* was significantly higher than that of *TRBJ1* in the *TRBV23-1* and *TRBV3-1 CDR3* repertoire in all the cDNA and gDNA samples from PBMCs of volunteers. Meanwhile, at the TCR loci, the 3' end of *TRBV* or *TRBD* in TCR β chains is a heptamer (CACAGTG)-23 base pair (bp)-nonamer (ACAAAAACC) rearrangement of signal sequences (3' *TRBV* 23 RSS and 3' *TRBD* 23 RSS), whereas the 5' end of *TRBD* or *TRBJ* is a nonamer-12bp-heptamer rearrangement of signal sequences (5' *TRBD* 12 RSS and 5' *TRBJ* 12 RSS). V-D-J recombination occurs only between segments with the 23 RSS terminal and segments with the 12 RSS terminal, and this restriction is called the 12/23 rule. V-D-J rearrangement of the TCR-beta chain follows the 12/23 rule and the beyond 12/23 restriction. The mechanism of differential gene rearrangement in the *TRBV23-1* and *TRBV3-1* CDR3 repertoire may be related to the 12/23 rule and the beyond 12/23 restriction, but that needs further study.

The diversity of the TCR CDR3 repertoire is the result of random combinations, insertions, and splicing of the germline VDJ gene fragment. The random nucleotide insertion (N region) was very important for composition of TCR β CDR3 diversity during the TCR β gene combination process. The region in which the nucleotide fragments randomly added into the Vβ-Dβ junction was called the N1 region. The N2 region was where the nucleotide fragments randomly added into the Dβ-Jβ junction. We also compared the length of N1 and N2 in *TRBV23-1* and *TRBV3-1 CDR3* repertoires. In the four cDNA samples from PBMCs, the *TRBV23-1* in-frame and out-of-frame CDR3 repertoire showed that N2 was longer than N1 and similar to the *TRBV3-1* out-of-frame CDR3 repertoire, whereas in the gDNA samples from PBMC of six volunteers, the *TRBV23-1* and *TRBV3-1* in-frame CDR3 and out-of-frame CDR3 repertoire showed no difference in the lengths of N1 and N2. These results show that the insertions during rearrangement of *TRBV23-1* and *TRBV3-1* (gDNA samples) were similar, whereas in the cDNA samples, the length of N in unproductive repertoires, namely, *TRBV3-1* out-of-frame, *TRBV23-1* in-frame and *TRBV23-1* out-of-frame, showed the same N2 > N1, while the length of N2 and N1 in the productive repertoire *TRBV3-1* was similar, thereby suggesting that the long N2 was closely related to the unproductive TCR receptor.

Taken together, the results of the present study generate the following conclusions: (1) The *TRBV23-1* (*ORF*) CDR3 repertoire expression frequency is lower than that of the *TRBV3-1* (F) CDR3 repertoire, indicating that the non-functional rearrangements of *TRBV23-1* limit the probabilities of being selected for taking part in rearrangement and development-maturation. (2) Due to the lack of functional maturation, the ratio of *TRBV23-1* (*ORF*) in-frame CDR3 and out-of-frame CDR3 completely follows the theoretical rules of random rearrangement. In contrast, the *TRBV3-1* (F) in-frame CDR3 and out-of-frame CDR3 ratio, although following the theoretical rules of random rearrangement, must reach maturity through individual negative and positive selection, thereby resulting in complicated and variable ratio characteristics. (3) In the cDNA of the CDR3 repertoire, due to the lack of selection by functional maturation of *TRBD* gene usage, N-region insertion/splicing of *TRBV23-1* (*ORF*) in-frame and out-of-frame CDR3 present characteristics that are similar to the *TRBV3-1* (F) out-of-frame CDR3 repertoire (non-functional).

In summary, these results indicate that analysis of the composition and characteristics of the *TRBV23-1* (*ORF*) CDR3 repertoire could provide the basis for research on the rules and mechanisms of TCR random rearrangement and the positive or negative selection process. At the same time, it could provide experimental quality control and research references for the analysis of the *TRBV* functional gene CDR3 repertoire.

Acknowledgements

We are grateful to the ten healthy volunteers for supporting this study. We thank Adaptive Biotechnologies Corporation (Seattle, WA, US) and Majorbio Bio-pharm Technology Co., Ltd (Shanghai, China) for help with human TCR β chain CDR3 repertoire sequencing and analysis.

The work was supported by grants from the National Prophase Project on Basic Research of China (973 pre-Program, 2008CB517310), the Natural Science and International Cooperation Program of Guizhou Province (2008-700105 & 2007-2122) and the National Natural Science Foundation of China (31160195 and 81441048).

The authors declare no conflict of interest.

References

1. Tonegawa S (1988): Antibody and T-cell receptors. JAMA 259: 1845-1847.
2. Rowen L, Koop BF, Hood L (1996): The complete 685-kilobase DNA sequence of the human beta T cell receptor locus. Science 272: 1755-1762.
3. http://www.imgt.org/
4. http://www.embl.org/
5. Currier JR, Yassai M, Robinson MA, Gorski J (1996): Molecular defects in TCRBV genes preclude thymic selection and limit the expressed TCR repertoire. J Immunol 157: 170-175.
6. Than S, Kharbanda M, Chitnis V, et al. (1999): Clonal dominance patterns of CD8 T cells in relation to disease progression in HIV-infected children. J Immunol 162: 3680-3686.
7. Manfras BJ, Terjung D, Boehm BO (1999): Non-productive human TCR beta chain genes represent V-D-J diversity before selection upon function: insight into biased usage of TCRBD and TCRBJ genes and diversity of CDR3 region length. Hum Immunol 60: 1090-1100.
8. Pommié C, Levadoux S, Sabatier R, et al. (2004): IMGT standardized criteria for statistical analysis of immunoglobulin V-REGION amino acid properties. J Mol Recognit 17: 17-32.
9. Alamyar E, Duroux P, Lefranc MP, Giudicelli V (2012): IMGT(R) 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 882: 569-604.
10. Alamyar E, Giudicelli V, Li S, et al. (2012): IMGT/HighV-QUEST: the IMGT® web portal for immunoglobulin (IG) or antibody and T cell receptor (TR) analysis from NGS high throughput and deep sequencing. Immun Res 8: 26.
11. Li S, Lefranc MP, Miles JJ, et al. (2013): IMGT/HighV QUEST paradigm for T cell receptor IMGT clonotype diversity and next generation repertoire immunoprofiling. Nat Commun 4: 2333.
12. Robins HS, Campregher PV, Srivastava SK, et al. (2009): Comprehensive assessment of T-cell receptor beta-chain diversity in alpha beta T cells. Blood 114: 4099-4107.
13. Robins HS, Srivastava SK, Campregher PV, et al. (2010): Overlap and effective size of the human CD8+ T cell receptor repertoire. Sci Transl Med 2: 47ra64.
14. Sherwood AM, Desmarais C, Livingston RJ, et al. (2011): Deep sequencing of the human TCRγ and TCRβ repertoires suggests that TCRβ rearranges after αβ and γδ T cell commitment. Sci Transl Med 3: 90ra61.
15. Wu D, Sherwood A, Fromm JR, et al. (2012): High-throughput sequencing detects minimal residual disease in acute T lymphoblastic leukemia. Sci Transl Med 4: 134ra63.
16. http://www.immunoseq.com
17. Stewart JJ, Lee CY, Ibrahim S, et al. (1997): A Shannon entropy analysis of immunoglobulin and T cell receptor. Mol Immunol 34: 1067-1082.
18. Yousfi Monod M, Giudicelli V, Chaume D, Lefranc MP (2004): IMGT/JunctionAnalysis: the first tool for the analysis of the immunoglobulin and T cell receptor complex V-J and V-D-J JUNCTIONs. Bioinformatics 20 Suppl 1: i379-385.
19. Lefranc MP (2011): IMGT unique numbering for the variable (V), constant (C), and groove (G) domains of IG, TR, MH, IgSF, and MhSF. Cold Spring Harb Protoc 2011; 633-642.
20. Meyer-Olson D, Brady KW, Blackard JT, et al. (2003): Analysis of the TCR beta variable gene repertoire in chimpanzees: identification of functional homologs to human pseudogenes. J Immunol 170: 4161-4169.
21. Murphy KM (2011): Janeway's immunobiology. Garland Science, New York.