

Original Article

Preliminary analysis of spatial-temporal homogeneity and heterogeneity of TCR β chain CDR3 repertoires in BALB/c mice

Yuehong Li^{1*}, Long Ma^{1*}, Xiaoheng Dong^{1*}, Yurong Pan¹, Bin Shi², Xiaoyan He¹, Teng Zhang¹, Suhong Sun³, Xinsheng Yao¹

¹Department of Immunology, Research Center for Medicine and Biology, Innovation and Practice Base for Graduate Students Education, Zunyi Medical University, Zunyi 563003, China; ²Department of Laboratory Medicine, Zunyi Medical University, Zunyi 563003, China; ³Department of Breast Surgery, The First Affiliated Hospital of Zunyi Medical College, Zunyi 563003, China. *Equal contributors.

Received December 23, 2018; Accepted January 21, 2019; Epub February 15, 2019; Published February 28, 2019

Abstract: The T-cell response and tolerance in non-lymph tissues differs from those in lymph tissues such as the spleen and thymus. The distribution and composition of the TCR repertoires in non-lymph tissues and how they differ and associate with their counterparts in lymph tissue remain unclear. Thus, we studied the thymus, spleen, blood, liver and small intestine of BALB/c mice at the ages of one, three and five months to carry out a preliminary analysis of the spatial-temporal homogeneity and heterogeneity of the total TCR β chain CDR3 repertoire using high-throughput sequencing technology and immune bioinformatics approaches. The data show that the diversity of the CDR3 repertoires was decreased as the mouse age increased, except in the small intestine. The number of low-expanded clones in the CDR3 repertoires was greatest in the thymus, followed by the spleen, blood, liver and small intestine, and highly expanded clones had an opposite trend in the different mice ages. The thymus and the spleen showed the greatest overlap of CDR3 sequences with the other tissues across the different mice ages. The distribution of the CDR3 repertoire length was normal, with a median of 14 aa in all the mouse tissues, except the small intestine of the one-month-old mice had a median of 12 aa. In summary, the composition and characteristics of the CDR3 repertoires in the thymus were similar to those in the spleen, and repertoires in the blood were similar to those in the liver; only the small intestine showed a unique composition. These results offer a novel method to explore the source, differentiation, proliferation and response of distinct T cells in different tissues at different mice ages.

Keywords: TCR β chain, CDR3 repertoires, high-throughput sequencing, homogeneity, heterogeneity

Introduction

Thymic T cells develop into a diverse TCR repertoire through germline gene rearrangements and progress to mature T cells via positive and negative selection, after which they are released to peripheral tissues that are involved in the immune response. The T-cell distribution and composition have yet to be fully described. Although many studies have focused on T-cell response and tolerance in the spleen, lymph nodes and peripheral blood, the T-cell response in non-lymphoid tissues is less well defined. Recent T-cell response studies in non-lymphoid tissues, such as the skin, lung, liver and small intestine, have indicated that res-

ponse patterns in non-lymphoid tissues differs from those in lymphoid tissue and the peripheral blood. In 2001, Masopust's group reported that CD8 T cells specific for viral and bacterial pathogens migrated to non-lymphoid tissues and were present as long-lived memory cells [1]. In 2006, this group also reported that virus-specific CD8 T cells in gut intraepithelial lymphocytes were not centralized and that effector memory CD8 T cells migrated from the spleen or peripheral blood, suggesting that memory T cells can differentiate in local tissues [2]. In 2012, Casey's group reported that effector T cell differentiation and memory T cell maintenance in the blood required constant antigen stimulation, but tissue-resident (intestinal) ef-

factor memory T cells (TRMs) do not require sustained antigen stimulation. CD69 and CD103 are expressed on the TRM surface in non-lymphoid tissues (intestine, stomach, kidney, reproductive tract, pancreas, brain and salivary glands), and this is mediated by cytokines. For example, TRMs are maintained in intestinal epithelial tissues via TGF- β -mediated CD103 expression, suggesting that the long-lasting TRM maintenance in local tissues such as the intestine uses a mechanism that differs from that of classical TEM [3]. In 2011, Hovhannisyan's group identified a distinct population of T regulatory (Treg) cells in an intestinal mucosal inflammatory environment of inflammatory bowel disease that were FoxP3⁺ and produced IL-17. These cells likely arise during differentiation of Th17 and Tregs, and the specific micro-environmental cues from tissues may have determined the balance of Th17 and Treg cell differentiation in these intestinal tissues [4]. Protective memory T cells were thought to reside in the blood and lymph nodes, but Jiang and colleagues reported that localized vaccinia virus (VACV) skin infection generated long-lived non-circulating CD8⁽⁺⁾ skin T (TRM) cells that reside within the skin that are superior to circulating central memory T cells (TCMs) for providing rapid, long-term protection against cutaneous reinfection [5]. Clark's group reported that alemtuzumab treatment of cutaneous T cell lymphoma (CTCL) did not severely compromise the immune response to infection, likely because sparing skin-resident TEMs provided local dermal immune protection by depleting circulating TCMs [6]. Purwar reported abundant resident T cells in the human lung—more than 10 billion T cells were present—which were largely effector memory T cell (TEM) phenotypes, but few TCMs and Treg cells were identified. Lung T cells had a diverse T cell receptor repertoire, and subsets produced IL-17, IL-4, IFN- γ , and TNF- α . Moreover, lung TRMs, but not TRMs resident to skin or T cells from the blood, proliferated in response to influenza virus [7]. Teijaro's group identified influenza-specific lung tissue-resident memory CD4 T cells, which did not circulate or migrate from the lung. These influenza-specific lung tissue-resident memory CD4 T cells served as *in situ* protectors to respiratory viral challenge, enhancing viral clearance and survival against lethal influenza infection. In contrast, memory CD4 T cells isolated from the spleen recirculated among multiple tissues but failed to offer protection against influenza in-

fection. Thus, tissue compartmentalization is important for immune-mediated protection at key mucosal sites for targeting local protective responses in vaccines and immunotherapies [8].

In 2013, Cebula's group analysed the T-cell receptor repertoires of CD4⁺ Foxp3⁺ and CD4⁺ Foxp3⁻ T cells in the thymus, intestine and colon using HTS. The results showed that thymus-derived Treg cells constitute the majority of Treg cells in the intestine and colon [9]. Emerson's group utilized HTS to compare the T cell CDR3 repertoire between the peripheral blood and infiltrating T cells (TILs) of ovarian carcinoma patients. The results suggest that the cellular adaptive immune response within ovarian carcinomas is spatially homogeneous and distinct from the T-cells compartment of peripheral blood [10]. Eric Shifrut's group explored the CD4⁺ T cell TCR repertoire in mice using HTS, and the results showed that the diversity of the TCR β CDR3 repertoire is reduced in the spleens of aged mice but is not reduced in the bone marrow [11]. Sathaliyawa's group analysed the distribution of human T cells in the lymph nodes, lung, spleen and intestines of individual organ donors and reported that these non-lymphoid tissues had a unique distribution within an individual but not comparatively among individual donors. They established a profile of human T cell distribution based on the compartmentalization of the debut of T cell subgroups [12]. In 2015, Bergot's group analysed the subsets of naïve and activated/memory Treg cells in mice. They reported that activated/memory (am) Treg cells are predominantly distributed in lymph nodes (LNs), and expanded clonotypes were primarily detected in deep-LN amTreg cells, accounting for 20% of the repertoire. Strikingly, these clonotypes were absent from nTreg cells, indicating different antigenic targets for naïve and amTreg cells and that amTreg cells are self-specific [13].

Although distinct T-cell populations and specific immune responses in lymphoid and non-lymphoid tissues have been confirmed, the source and activation mechanisms of these cells and their relationship with T cells in lymphoid tissues and the circulation have not been elucidated. At present, little is known regarding TCR CDR3 sequence homogeneity and heterogeneity in different tissues. The overall distribution

and compartmentalization of TCR repertoires under physiological conditions provides a foundation to explore the sources and the distinct response mechanisms of T cells in different tissues and offers new perspectives and methods to study the differences and relationships among T cells in non-lymphoid and lymphoid tissues and the circulation. Here, we sequenced the total T-cell TCR β CDR3 repertoire of the thymus, spleen, blood, liver, and intestine from BALB/c mice at the ages of one, three and five months by HTS. We preliminarily assessed the composition and characteristics of diversity, clonality, overlapping, CDR3 aa length, CDR3 aa usage, insertion or deletion of v/j/d5/d3, and TRBV/TRBJ/TRBD usage in the CDR3 repertoires.

Materials and methods

TCR β CDR3 repertoire preparation and high-throughput sequencing

Three BALB/c female mice, 4-weeks old, from the same litter, were purchased from CAVENS Experimental Animal Ltd. (China) and housed in the Zunyi Medical College SPF animal facility. All of the animal experiments were performed according to the guidelines for the Care and Use of Laboratory Animals (Ministry of Health, China, 1998). The experimental procedures were approved by the Zunyi Medical College Laboratory Animal and Use Committee (permit number 2013016). Three BALB/c female mice were sacrificed-one each at one (M1), three (M3) and five (M5) months of age-and thymus (T), spleen (S), blood (B), liver (L) and small intestine (I) samples were harvested (a total of 15 tissue samples). Genomic DNA was extracted using a QIAamp DNA MiniKit (Cat: 51304, QIAGEN, Germany) and stored in QIA safe DNA tubes (QIAGEN). The genomic DNA was identified using 1% agarose gel electrophoresis and the concentration and purity of the genomic DNA was confirmed. Mouse TCR β chain CDR3 repertoires were generated with multiplex PCR amplification consisting of 36 forward mouse T-cell receptor beta variable (TRBV) segments and 14 reverse mouse T-cell receptor beta joining (TRBJ) segment primers. The TCR β CDR3 repertoires (PCR library) were quantified and loaded on an Illumina flow cell for sequencing with an Illumina Genome Analyzer (USA). The amplification and sequencing were completed with the Adaptive Biotechnologies ImmunoSEQ

platform. Raw sequences with technical failures were removed by a complexity filter. A nearest-neighbour algorithm was used to collapse the data into unique sequences by merging closely related sequences to remove both PCR errors and GA sequencing errors. Based on unique tags that were identified during database generation, sequences of the CDR3 repertoire of 15 samples from all mice were acquired [14-18]. The TCR CDR3 repertoire preparation and the HTS were carried out by Adaptive Biotechnologies Corp. (Seattle, WA, USA).

Analysis of the composition and characteristics of TCR β CDR3 repertoires

We downloaded the original sequence of each sample from ImmunoSEQ ANALYZER (<https://clients.adaptivebiotech.com>) and analysed it further in Microsoft Excel (Microsoft). Depending on the genetic composition of the mouse TCR β CDR3 sequence status (stop/in-frame/out-of-frame), ImmunoSEQ ANALYZER was used to identify out-of-frame sequences, in-frame sequences, and stop sequences. The in-frame sequences for all samples were then used to analyse the TCR β CDR3 repertoire sequences. We identified CDR3 nucleotides, CDR3 amino acids, counts (reads), frequency count (%), CDR3 length, variable (V), diversity (D), and joining (J) gene names, v, d5, d3, and j deletions, and n1 and n2 insertions. The inverse Simpson index (1/Ds) formula was $1/Ds = 1/[1-\sum (Ni(Ni-1))/(N(N-1))]$, where Ni was defined as the frequency of the gene, and N was defined as the total number of genes [19]. In addition, the statistical analysis of the composition and characteristics of the unique and total in-frame CDR3 repertoires was analysed, including the CDR3 aa repertoire clone expansion frequency, where $<0.001\%$ = low-expanded clones (LECs), $0.1-0.001\%$ = medium-expanded clones (MECs), and $>0.01\%$ = high-expanded clones (HECs). Unique CDR3 aa overlapping sequences (<http://bioinformatics.psb.ugent.be/webtools/Venn>), the CDR3 repertoire length distribution and aa usage, the TRBV, TRBJ, TRBD usage of the CDR3 repertoires, and the insertion and deletion of the unique CDR3 repertoire [15, 17, 20, 21] were also analysed. The χ^2 test and a non-parametric one-way ANOVA and Bonferroni's post-test were used to compare differences among samples ($P<0.05$ was considered to be statistically significant).

Analysis of TCR β chain in BALB/c mice

Table 1. The number and percent of TCR β chain CDR3 gene and aa sequences in the thymus, spleen, blood, liver, and small intestine of M1, M2 and M3

	Total CDR3	Unique CDR3	Unique/Total (%)	Total in-frame CDR3	Unique in-frame CDR3	Unique/Total (%)	Total in frame CDR3 aa	Unique in frame CDR3 aa	Unique/Total (%)
M 1-T	8952783	428321	4.78	6645969	245689	3.7	6557117	241194	3.68
M 1-S	1300362	150183	11.55	972973	94203	9.68	960331	92695	9.65
M 1-B	1346868	126730	9.41	1016233	81288	8	1004675	80103	7.97
M 1-L	1099864	59910	5.45	821980	40319	4.91	812497	39762	4.89
M 1-I	336248	4694	1.4	219899	3070	1.4	216530	3011	1.39
M 1-Sum	13036125	769838	5.91	9677054	464569	4.8	9551150	376884	3.95
M 3-T	9203625	302948	3.29	6926201	176004	2.54	6844631	172855	2.53
M 3-S	10122248	254840	2.52	7533787	149276	1.98	7434851	146647	1.97
M 3-B	940691	89022	9.46	705135	58737	8.33	696185	57826	8.31
M 3-L	693019	34464	4.97	515467	23834	4.62	509102	23493	4.61
M 3-I	383327	6322	1.65	280340	3997	1.43	277737	3921	1.41
M 3-Sum	21342910	687596	3.22	15960930	411848	2.58	15762506	334698	2.12
M 5-T	10604030	259209	2.44	7935475	148079	1.87	7849764	145479	1.85
M 5-S	10547716	185315	1.76	7900736	110530	1.4	7794087	108640	1.39
M 5-B	987345	57045	5.78	732686	38612	5.27	716193	37971	5.3
M 5-L	1077880	42 632	3.96	796904	28974	3.64	775811	28546	3.68
M 5-I	577411	8672	1.5	433708	6066	1.4	428266	5967	1.39
M 5-Sum	23794382	552873	2.32	17799509	332261	1.87	17564121	270491	1.54

Results

Total T cell TCR β CDR3 repertoire preparation and the diversity of CDR3 repertoires

The band of genomic DNA, shown at a predetermined size and position by agarose gel electrophoresis, was extracted for each sample (Figure S1). The total original sequences were obtained using an Illumina Genome Analyzer. The ratio of unique CDR3 gene (or CDR3 aa) sequences to the total CDR3 repertoire was calculated, and for M1 the ratio descended in the order of spleen, blood, liver, thymus, and small intestine. For both M3 and M5 the ratio descended in the following order: blood, liver, thymus, spleen and small intestine (Table 1). To evaluate the diversity of the TCR CDR3 repertoires, we introduced the inverse Simpson's diversity index (1/Ds) [19]. The 1/DS of all 15 samples is depicted in Figure 1. The data showed that the diversity of the total CDR3 repertoire in the same tissue across M1, M3 and M5 was decreased. The diversity of the CDR3 repertoire in the thymus of M1 (youngest mouse) was significantly greater than in M3 and M5. The diversity of the CDR3 repertoire in the small intestine of M5 (oldest mouse) was

greater than in M1 and M3. The diversity of the CDR3 repertoire in the thymus, spleen and blood exceeded that in the liver and small intestine of the three mice. The diversity of the CDR3 repertoire in the spleen, blood, and liver decreased with age (M1>M3>M5).

Frequency of clonal expansion of the CDR3 aa repertoires

The clonal percent of LEC (<0.001%), MEC (0.01-0.001%) and HEC (>0.01%) CDR3 aa repertoires in the thymus, spleen, blood, liver, and intestine of M1, M3 and M5 are depicted in Figure 2A-D. LECs were greatest in the thymus, followed by the spleen, blood, liver, and small intestines in M1, M3 and M5, and HECs showed an opposite trend, except that HECs in the small intestines were significantly greater in abundance. MECs were the lowest in the thymus, followed by the spleen, blood, and liver, and the small intestines had the least MECs. The average percent of LECs, MECs and HECs of the CDR3 aa repertoires of M1, M3 and M5 are depicted in Figure 2E-G. For LECs, the small intestine and blood were not different, nor were the small intestine and liver ($P>0.05$); however, other comparisons were significantly different

Analysis of TCR β chain in BALB/c mice

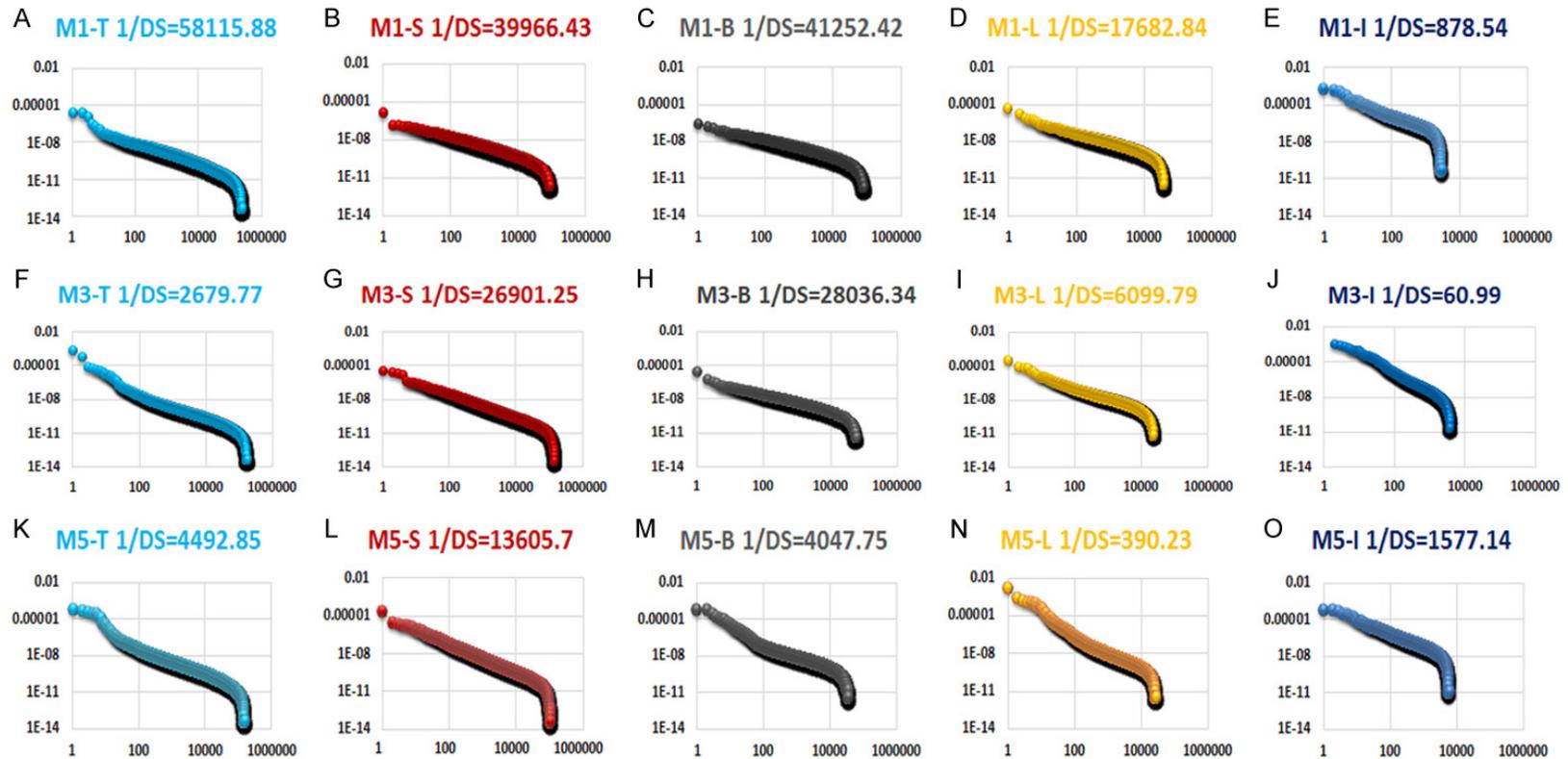


Figure 1. Statistical analysis of CDR3 aa repertoires diversity by 1/DS in the thymus, spleen, blood, liver, and small intestine of M1, M3 and M5 separately. (1) The X axis was the clonotype distribution plots of TCR CDR3 aa repertoire sequences. (2) The Y axis was the CDR3 aa sequences frequency (percent of reads). (3) Each dot represents a distinct CDR3- β aa (the dot of $1.34E-02$ for three-month-old mouse was excluded).

Analysis of TCR β chain in BALB/c mice

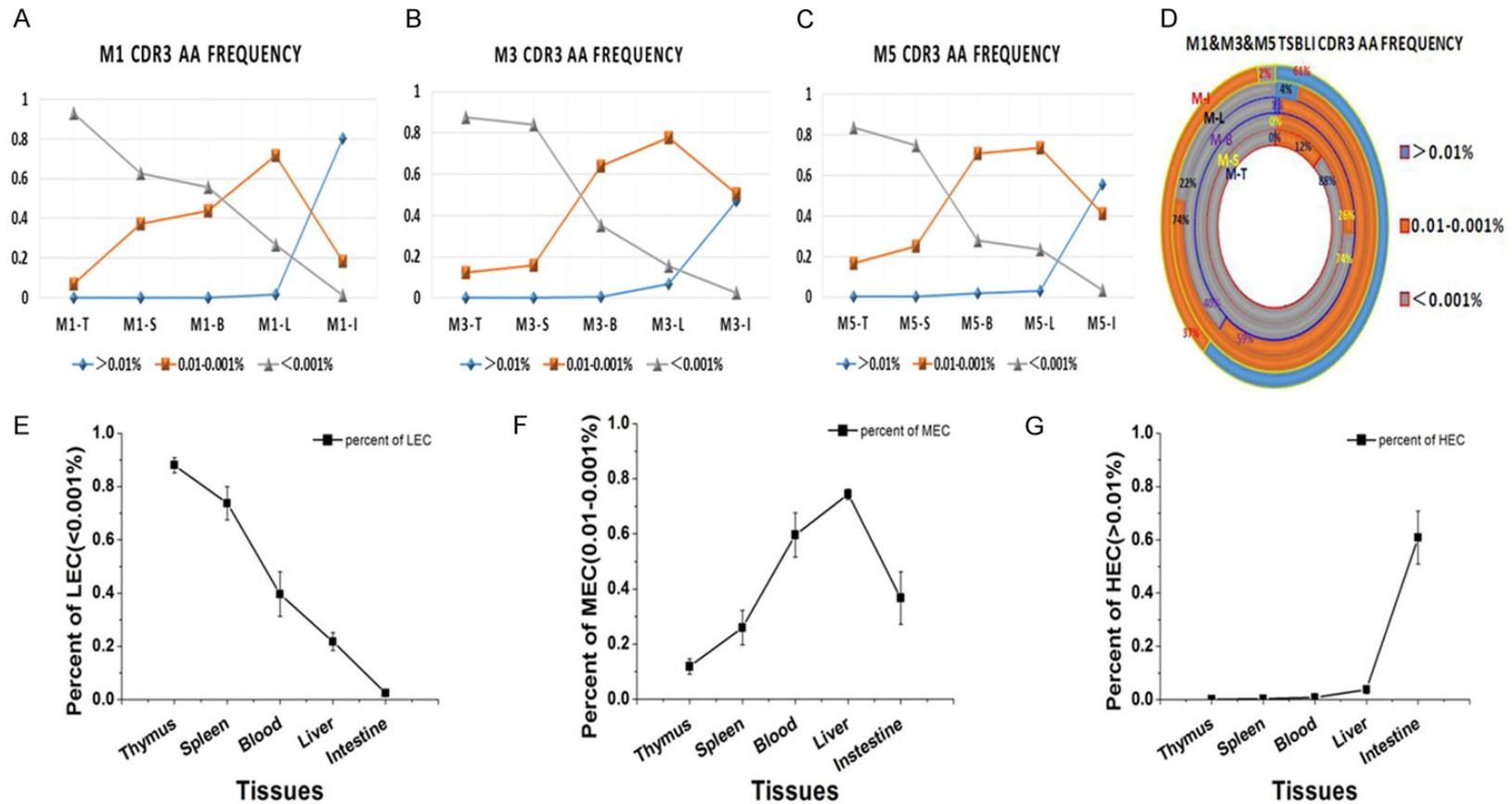


Figure 2. Clonal percent of LEC (low-expanded clones, <0.001%), MEC (medium-expanded clones, 0.01-0.001%) and HEC (high-expanded clones, > 0.01%) CDR3 aa repertoires in the thymus, spleen, blood, liver, and small intestine of M1, M3 and M5. A-C. The percent of LEC, MEC and HEC in each tissue of M1, M3 and M5 separately. D. The percent of LEC, MEC and HEC in each tissue of M1&M3&M5 in one Circular Graph. E-G. Comparing LEC, MEC and HEC clonal populations (M1+M3+M5) between any two tissues separately. LEC: small intestine and blood were not difference nor were small intestine and liver ($P>0.05$), but other comparisons were statistically significantly different ($P<0.01$). MEC: there were no difference between small intestine and thymus, or spleen, or blood, or between liver and blood ($P>0.05$); but there were significant difference for all other comparisons ($P<0.05$ or $P<0.01$). HEC: there were significant difference between small intestine and thymus, or spleen, or blood, or liver ($P<0.01$) but there were no difference for other comparisons ($P>0.05$).

Analysis of TCR β chain in BALB/c mice

	M1-T	M1-S	M1-B	M1-L	M1-I	M1-Sum	M3-T	M3-S	M3-B	M3-L	M3-I	M3-Sum	M5-T	M5-S	M5-B	M5-L	M5-I	M5-Sum
M1-T	241194	28494	24816	14180	1144		40246	37414	17874	8031	1387		35380	29468	12050	9431	2086	
M1-S	28494	92695	17509	10590	921		22229	21863	11563	5437	928		19961	17913	7961	6306	1447	
M1-B	24816	17509	80103	9031	864		19662	19626	10546	4899	857		17722	15911	7122	5690	1295	
M1-L	14180	10590	9031	39762	623		11347	11098	6184	3166	527		10317	9224	4302	3484	824	
M1-I	1144	921	864	623	3011		935	958	576	301	82		880	811	396	320	80	
M1-Sum						376884												
M3-T	40246	22229	19662	11347	935		172855	32986	15546	7882	1345		29117	24256	10047	8061	1822	
M3-S	37414	21863	19626	11098	958		32986	146647	17511	8954	1621		26923	24011	10004	7986	1780	
M3-B	17874	11563	10546	6184	576		15546	17511	57826	4929	904		13340	12267	5688	4494	1023	
M3-L	8031	5437	4899	3166	301		7882	8954	4929	23493	647		6277	5770	2696	2247	547	
M3-I	1387	928	857	527	82		1345	1621	904	647	3921		1102	968	479	379	125	
M3-Sum							334698											
M5-T	35380	19961	17722	10317	880		29117	26923	13340	6277	1102		145479	24291	10201	8873	2146	
M5-S	29468	17913	15911	9224	811		24256	24011	12267	5770	968		24291	108640	10842	9964	2608	
M5-B	12050	7961	7122	4302	396		10047	10004	5688	2696	479		10201	10842	37971	4800	1380	
M5-L	9431	6306	5690	3484	320		8061	7986	4494	2247	379		8873	9964	4800	28546	1498	
M5-I	2086	1447	1295	824	80		1822	1780	1023	547	125		2146	2608	1380	1498	5967	
M5-Sum												270491						

Figure 3. The overlap number of unique CDR3 aa sequences between any two tissues of the thymus, spleen, blood, liver, and small intestine in M1, M3 and M5. The colors in the picture are for easy to observation, and do not mean anything.

Analysis of TCR β chain in BALB/c mice

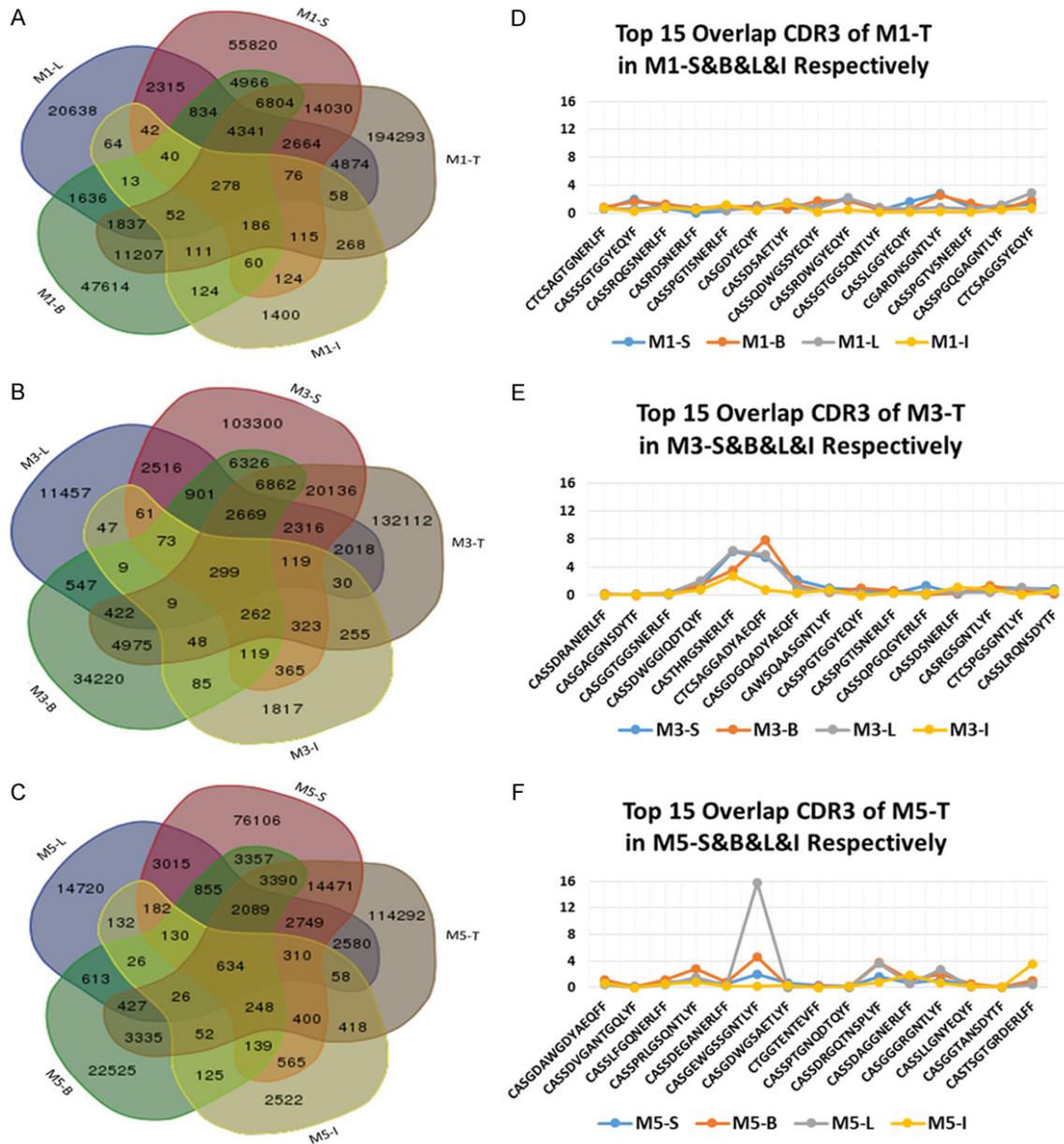


Figure 4. The number of unique CDR3 aa overlap sequences in CDR3 aa repertoires and the distribution of top 15 high-frequency CDR3 aa overlap sequences in the Thymus. A-C. The number of unique CDR3 aa overlap sequences among in the thymus, spleen, blood, liver, and small intestine of M1, M3 and M5 separately. D-F. The top 15 high-frequency CDR3 aa overlap sequences in the Thymus was distribution in the spleen, blood, live and small intestine of M1, M3 and M5 separately.

($P < 0.01$). For MECs, there were no differences between the small intestine and the thymus, or spleen, or blood, or between the liver and blood ($P > 0.05$); however, there were significant differences for all other comparisons ($P < 0.05$ or $P < 0.01$). For HECs, there were significant differences between the small intestine and the thymus, spleen, blood, or liver ($P < 0.01$), but there were no differences for the other comparisons ($P > 0.05$).

Unique CDR3 aa overlap sequences in CDR3 aa repertoires

The number and percent of unique CDR3 aa overlap sequences in the total number of unique CDR3 sequences showed differences (Figures 3, S2). The number of unique CDR3 aa overlap sequences among the thymus, spleen, blood, liver, and small intestine of M1, M3 and M5 are depicted in Figure 4A-C. The top-15

Analysis of TCR β chain in BALB/c mice

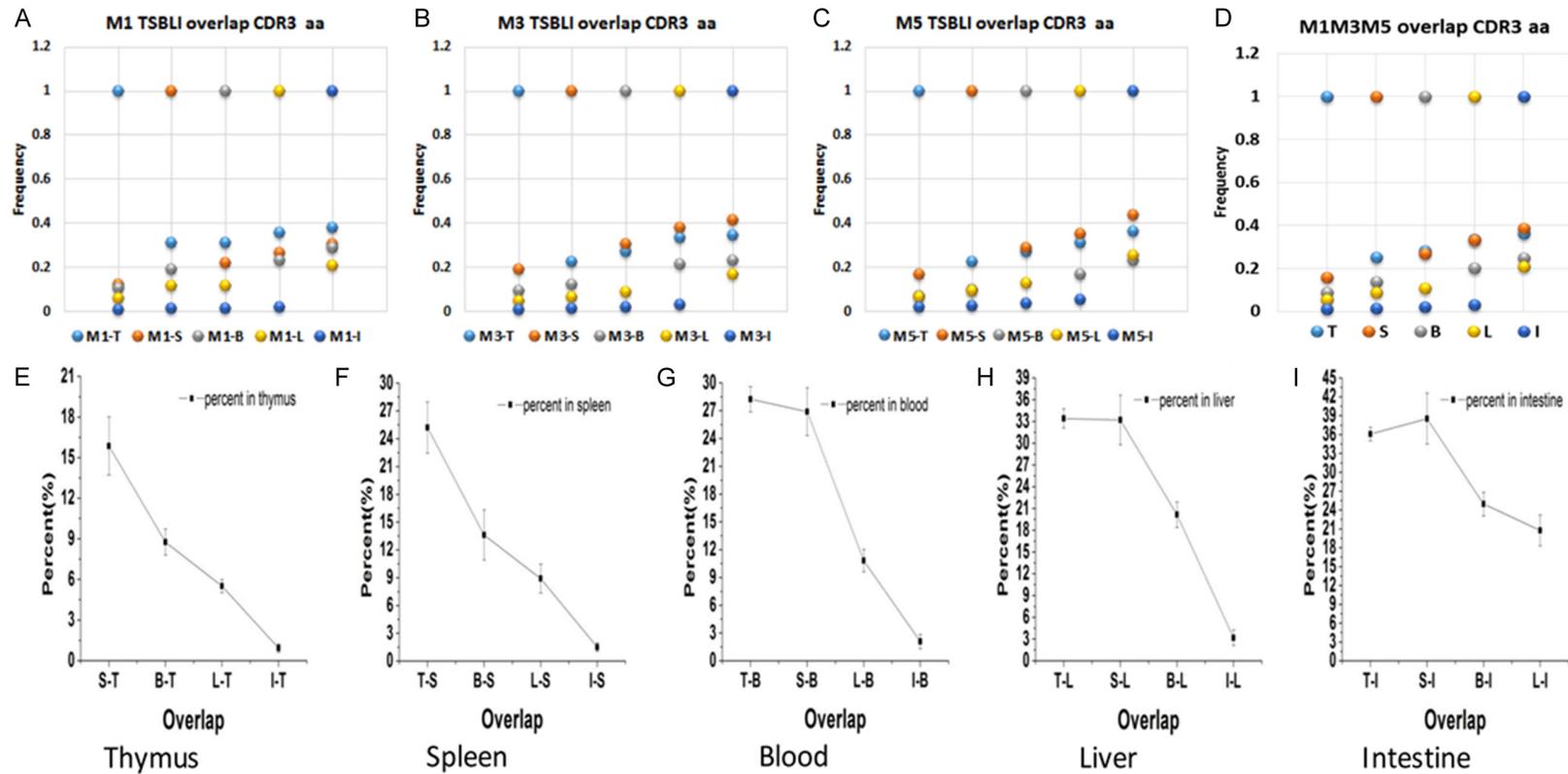


Figure 5. Analysis of unique CDR3 aa overlap sequences between any two tissues. A-D. The percent of overlapping unique CDR3 aa sequences in any tissue with other four tissues of M1, M3 and M5 separately. And the percent of unique CDR3 aa in the sum of any same one tissue (M1+M3+M5) were overlapping in the sum of four other tissue. E-I. Statistical analysis of the average percent of unique CDR3 aa in the sum of any same one tissue (M1+M3+M5) were overlapping in the sum of four other tissue separately.

Analysis of TCR β chain in BALB/c mice

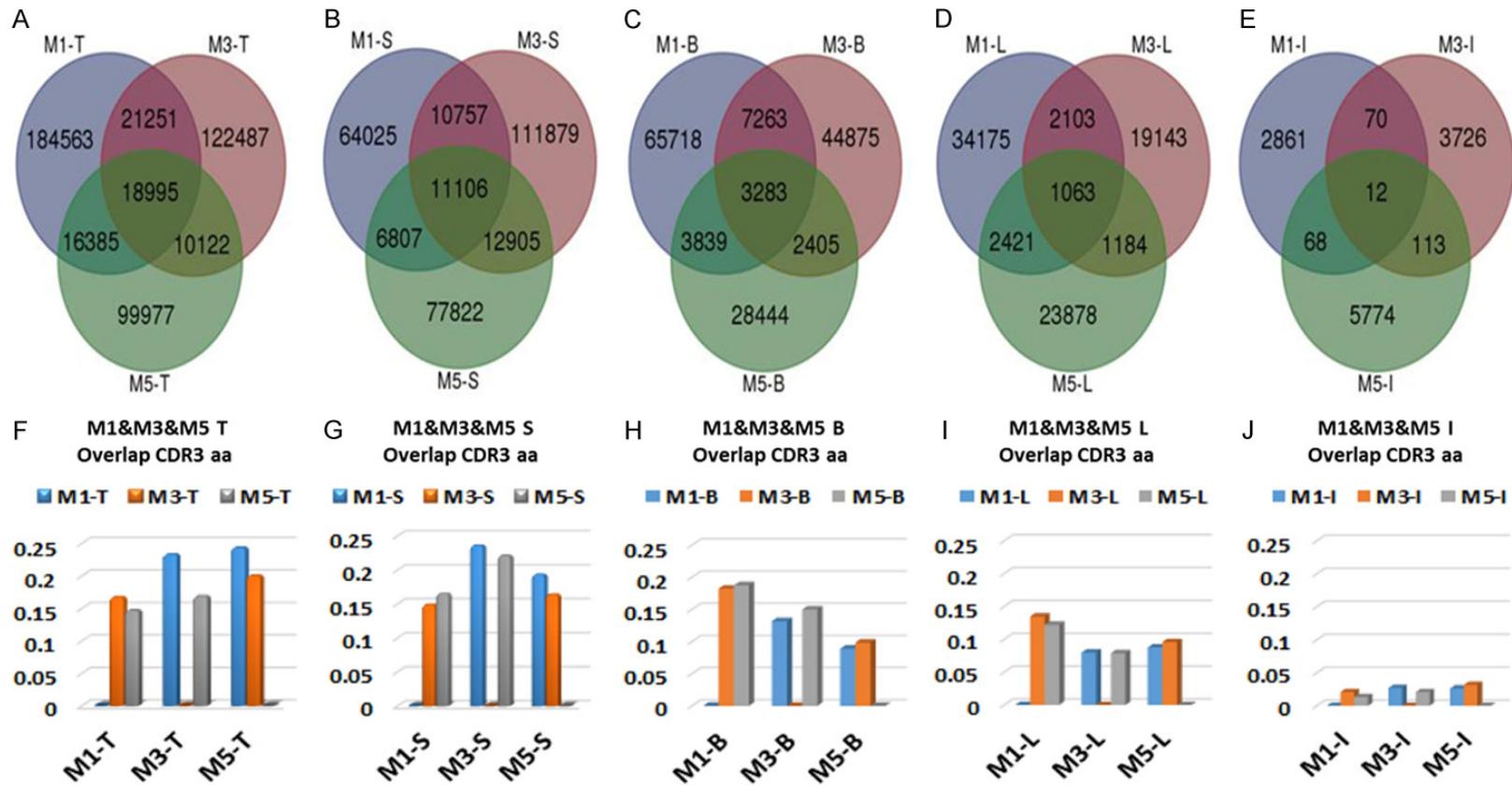


Figure 6. Analysis of unique CDR3 aa verlap sequences distribution among the same tissues. A-E. The number of unique CDR3 aa overlap sequences among the same tissues of M1, M3 and M5 separately. F-J. The percent of unique CDR3 aa overlap sequences in any one tissue of one mouse were overlapping in the other two same tissues of the other mouse separately.

Analysis of TCR β chain in BALB/c mice

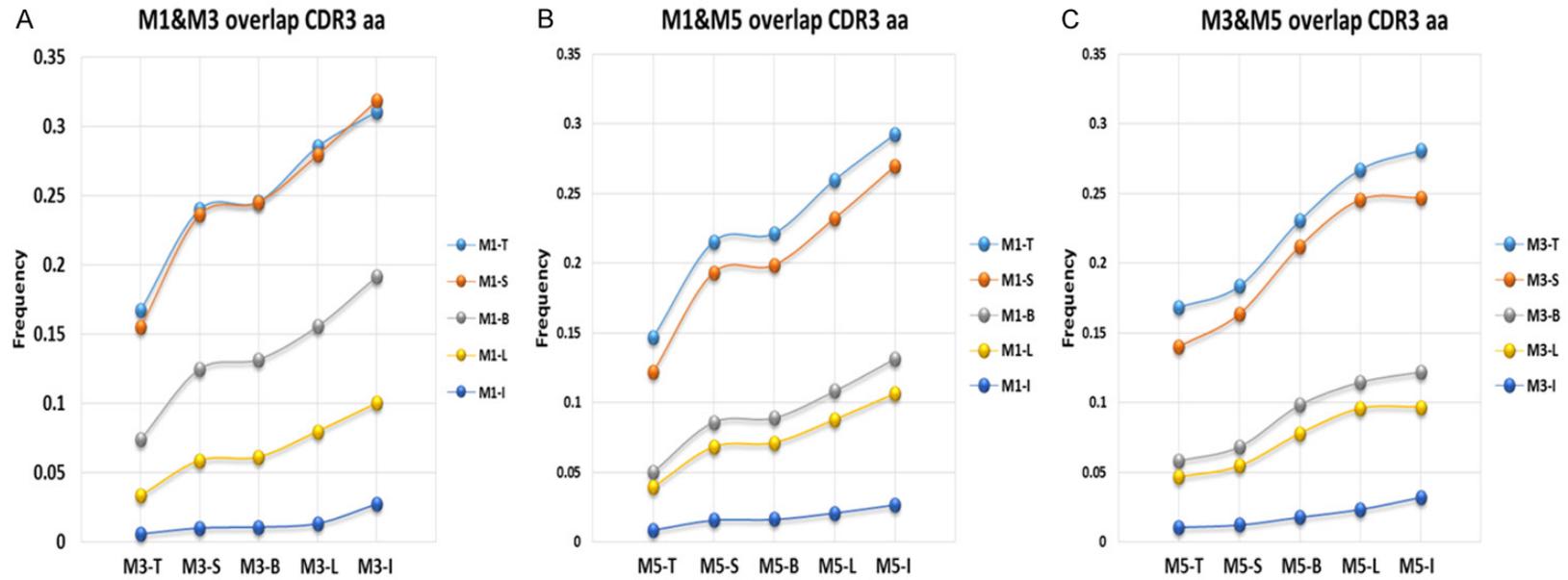


Figure 7. Comparing the percent of unique CDR3 aa overlap sequences in any ONE tissue of M1 (or M3) were overlapping in the other five tissues of M3 or M5 (A-C). Note: The data showed that the thymus and spleen had the greatest CDR3 aa overlap sequences and they were similar, followed by the blood, liver and small intestine in the M1&M3. But in the M1&M5 or M3&M5, the thymus had the greatest CDR3 aa overlap sequences, followed by the blood, liver and small intestine, and the thymus was very similar to the spleen, the blood was very similar to the liver.

high-frequency CDR3 aa overlap sequences in the thymus distributed in the spleen, blood, liver and small intestine of M1, M3 and M5 are depicted in **Figure 4D-F**. The data showed that the high proportions in the thymus had lower proportions in the other tissues of M1, but in M3 and M5, the top-15 high-frequency aa overlap sequences in the thymus also had high proportions in the other studied tissues.

The percent of unique CDR3 aa sequences in any one tissue (or the sum of the three same-tissue samples) were overlapping in the four other tissues (or the sum of the three same-tissue samples) of M1, M3 and M5 and are depicted in **Figure 5A-D**. The data show that the thymus had the least amount of overlap CDR3 aa sequences which came from the other tissues, followed by the spleen, blood, liver, small intestine. The thymus of M1 showed the greatest amount of overlap CDR3 aa sequences with the other tissues; however, for M3 and M5, the spleen showed the greatest overlap CDR3 aa sequences with the other tissues. The average percent of unique CDR3 aa in the sum of any one tissue (M1+M3+M5) overlapped in the sum of the four other tissues and is depicted in **Figure 5E-I**. The data showed that the thymus and the spleen had a similar distribution, and the blood and the liver had a similar distribution, but the small intestine had a unique distribution.

The unique CDR3 aa overlap sequence distribution among the same tissues was analysed. The number of unique CDR3 aa overlap sequences among the same tissues are depicted in **Figure 6A-E** and the percent of unique CDR3 aa overlap sequences in any one tissue of a mouse were overlapping in the same tissues of the other mice (depicted in **Figure 6F-J**). The data indicated that the thymus had the highest percent and was similar to the spleen, while the blood had the lowest percent and was similar to the liver. The small intestine had the lowest number and percent of unique CDR3 aa overlap sequences.

The number and percent of unique CDR3 aa overlap sequences between M1 and M3, M1 and M5, and M3 and M5 are presented in **Figures 3** and **S2**. We compared the percent of unique CDR3 aa overlap sequences in any one tissue of M1 (or M3) that were overlapping in the other five tissues of M3 or M5 (**Figure 7**).

The data showed that the thymus and spleen had the greatest amount of CDR3 aa overlap sequences and that they were similar, followed by the blood, liver and small intestine in M1 and M3. However, in M1 and M5 or M3 and M5, the thymus had the greatest amount of CDR3 aa overlap sequences, followed by the blood, liver and small intestine; the thymus was very similar to the spleen, and the blood was very similar to the liver.

CDR3 repertoire aa length distribution and aa usage

The unique CDR3 aa repertoire length distribution and aa usage (20 types of aa) in the thymus, spleen, blood, liver, and small intestine of M1, M3 and M5 are depicted in **Figure 8**, and the total CDR3 aa repertoire length distributions are depicted in **Figure S3**. The aa length distribution of unique CDR3 aa repertoires has a median of 14 aa in all tissues of M1, M3 and M5, except for the thymus of M1 (median 12 aa). The total CDR3 aa repertoire length distributions were similar to the unique CDR3 aa repertoires (**Figure S3**), but in the liver of M5 the median was 17 aa. The amino acid usage in all tissues was consistent with the reported anomalous low-frequency usage of K (lysine), M (methionine), H (histidine), and I (isoleucine).

Usage of TRBV, TRBJ and TRBD in unique CDR3 repertoires

There was no significant difference in the unique CDR3 repertoire for TRBV, TRBD, and TRBJ usage among all tissues of M1, M3 and M5 (**Figure 9**). TRBV1, TRBV5, TRBV13, TRBV19 and TRBV31 has the highest usage. TRBJ 02-7 usage was the greatest, and TRBJ 01-7 (ORF) and TRBJ 02-6 (P) usage was extremely low in all mouse tissues. TRBD 01 usage in all tissues was higher than that of TRBD 02.

Insertion and deletion in unique CDR3 repertoires

Comparisons of insertions from all tissues indicated that M1 (youngest) had more insertions in the thymus. For v, j, d5, d3 deletions, the thymus and the spleen comparisons among M1, M3 and M5 were different. In M1, a comparison of the blood, liver and small intestine indicated significant differences, but no differences were found in M3 and M5. The pair-wise comparison

Analysis of TCR β chain in BALB/c mice

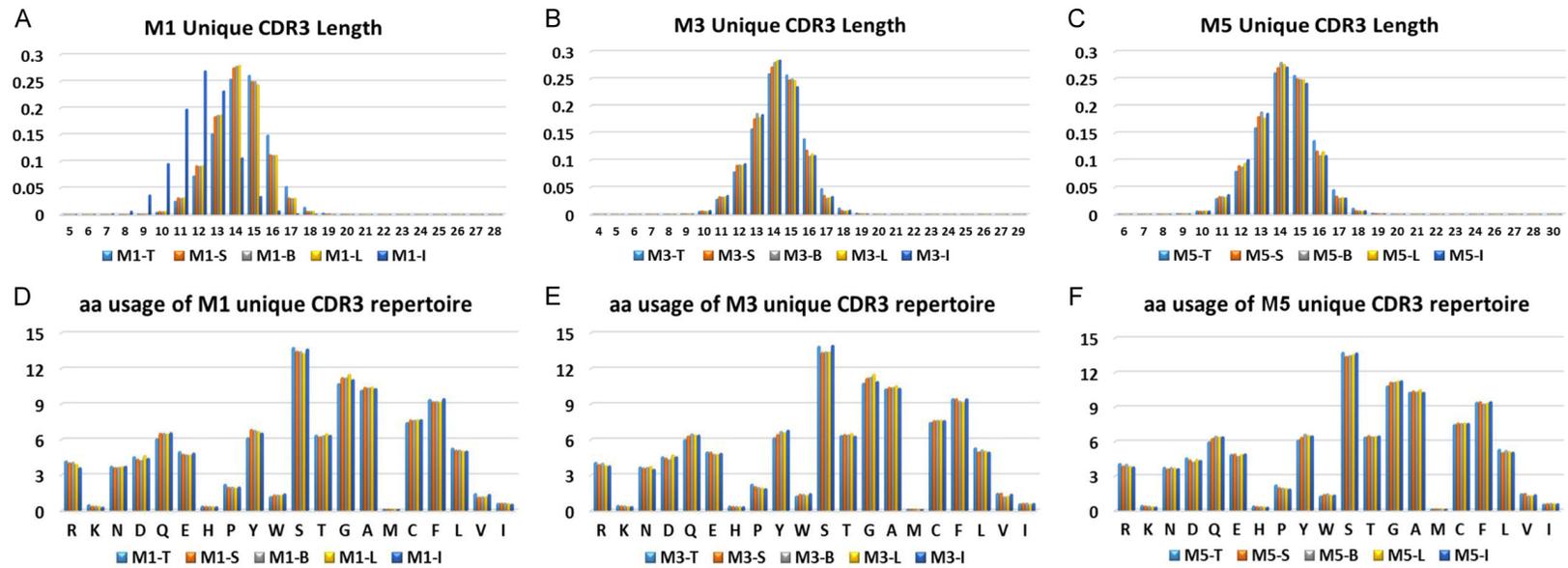


Figure 8. The unique CDR3 aa repertoires length distribution and aa usage (20 type of aa) in the thymus, spleen, blood, liver, and small intestine of M1, M3 and M5 separately (A-F). Note: (1) The X axis was the CDR3 aa sequences length or the name of 20 type aa. (2) The Y axis was the percent of CDR3 aa sequences in the CDR3 repertoires.

Analysis of TCR β chain in BALB/c mice

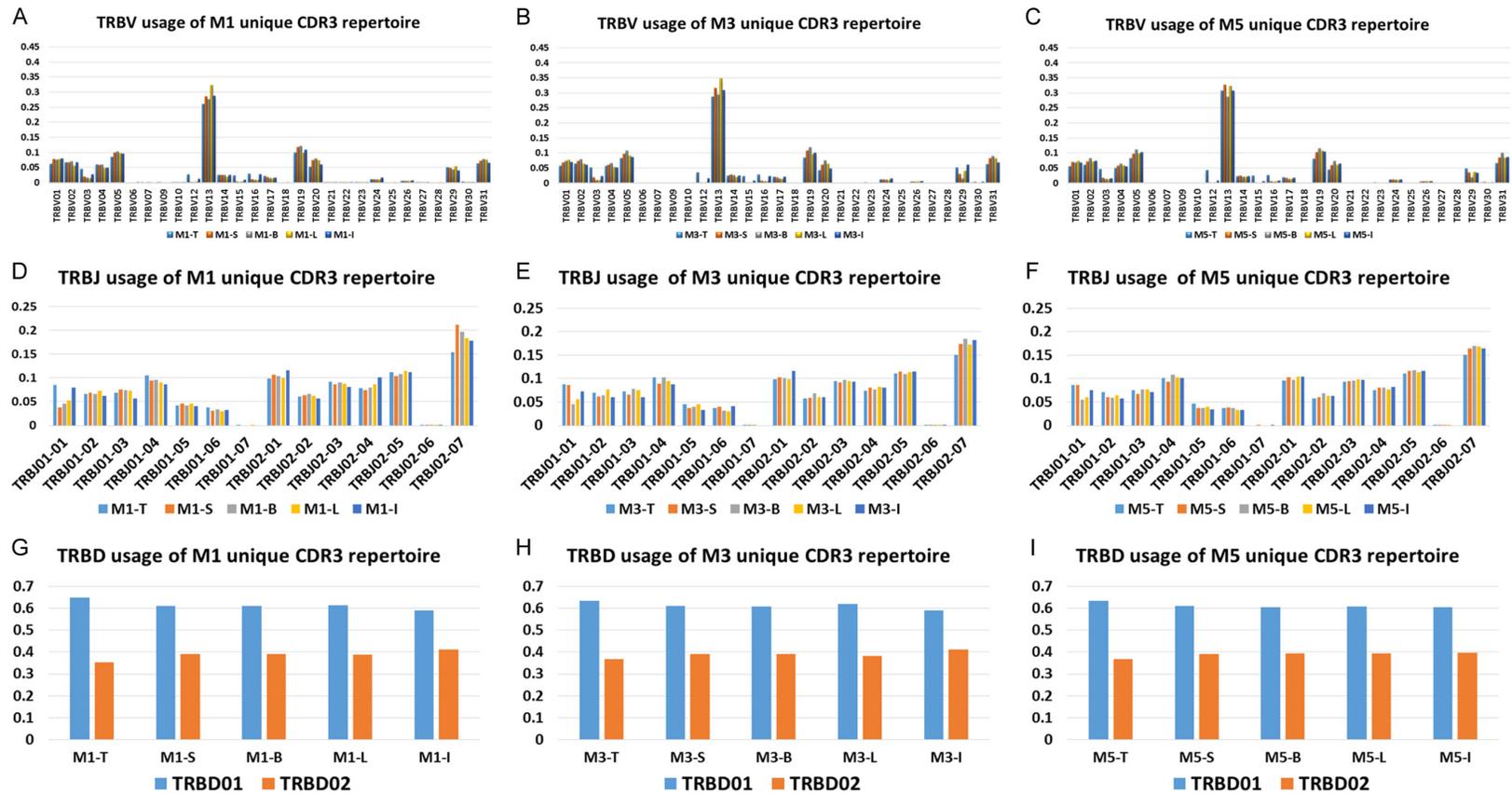


Figure 9. The TRBV, TRBJ, TRBD usage of unique CDR3 repertoires (gene sequences) in the thymus, spleen, blood, liver, and small intestine of M1, M3 and M5 separately (A-I). Note: (1) The X axis was the family name of TRBV, TRBJ or TRBD. (2) The Y axis was the percent of TRBV, TRBJ or TRBD usage in unique CDR3 repertoires.

Analysis of TCR β chain in BALB/c mice

data of certain tissues from each mouse, as well as tissue comparisons between M1, M3 and M5, are depicted in [Figure S4](#) and [Table S1](#).

Discussion

T cell characteristics, including the development, differentiation, and proliferation in the thymus, spleen and lymph nodes, have been elucidated by the clonal selection theory of Burnet. However, distinct T cell attributes such as origin, differentiation, cloning and oligo-clonal proliferation have been less well-studied in non-lymphoid tissues (i.e., skin, liver and small intestine). Differences and relationships among distinct T cells in non-lymphoid tissues have not been extensively studied, nor have the differences and relationships between lymphoid tissues and non-lymphoid or circulatory tissues.

The composition and characteristics of TCR CDR3 repertoires was used to study T-cell response and tolerance. Previously, capillary electrophoretic analysis was used for CDR3 spectratyping [22, 23].

The first direct estimation of the diversity of the human alpha beta T-cell receptor was carried out in 1999. Arstila's group calculated the number of $V\beta 18$ - $J\beta 1.4$ sequences and $V\beta 16$ - $J\beta 2.2$ sequences and then estimated the diversity of whole β chains using the frequency of $V\beta 18$ - $J\beta 1.4$ and $V\beta 16$ - $J\beta 2.2$ sequences [24]. In 2000, Casrouge and colleagues reported the TCR repertoire in the mouse spleen. The authors performed extensive sequencing of immunoscope CDR3 size peaks in TRBV10-TRBJ1.2 combinations. They then evaluated the size of the TCR β -chain pool in the TRAV2 subset and finally obtained a minimal value for the $\alpha\beta$ repertoire size [25].

In 2009, high-throughput sequencing technology was used to study the TCR CDR3 repertoire [17, 26]. Since that time, the composition of TCR CDR3 repertoires in different physiological and pathological conditions have been analysed using HTS technology and immune bioinformatics approaches [27-31]. Here, we sequenced the CDR3 repertoire in five tissues (thymus, spleen, blood, liver, and small intestine) at three different ages (one, three and five months) of BALB/c mice and analysed the

homogeneity and heterogeneity among the different tissues at different ages.

We compared and analysed the composition and characteristics of diversity, clonality, overlapping, CDR3 aa length, CDR3 aa usage, insertion or deletion of v/j/d5/d3, and TRBV/TRBJ/TRBD usage of the total T-cell CDR3 repertoires in a single BALB/c mouse at one-, three- and five months of age separately. The 1/DS and unique sequence percent of the TCR CDR3 repertoire analysis indicated that lymphoid tissue (thymus and spleen) had more diversity of distribution when compared to the blood. The diversity of the CDR3 repertoire in the thymus of the one-month-old mouse was greater than in the older mice, and these data agree with the diversity of CDR3 repertoire data from MRL/lpr in the thymus at all mouse ages, according to a previous study [27]. The diversity of CDR3 repertoires in the spleen, blood, and liver decreased with increased age. Thymic function has been confirmed to be mature in one-month-old mice and then gradually diminishes with age. The diversity of T cells in the peripheral circulation also decreases over time. In 2004, Messaoudi I et al. showed that the presence of T-cell clonal expansion reduced $CD8^+$, but not $CD4^+$ T-cells, in aged mice, and they demonstrated that age-related T-cell clonal expansion functionally impairs the efficacy of antiviral $CD8^+$ T-cell immunity in an antigen-specific manner [32]. In 2012, we used CDR3 spectratyping (Immunoscope) techniques to analyse the TRBV-TRBJ1-1 CDR3 repertoire in the thymus, spleen, liver, small intestine, and blood of a one-month-old BALB/c mouse [33]. In the current study, our results show for the first time the murine T-cell diversity across different ages using the composition of the CDR3 repertoire sequence. The data show that diversity of the intestinal CDR3 repertoire in the one-month-old mouse was greater than in the three-month-old mouse but lower than in the five-month-old mouse. This may be explained by an increased diversity of the intestinal flora in the five-month-old mouse, and more work is required to confirm this. In this experiment, the total high-throughput sequencing results for in-frame CDR3 in the 1-, 3-, and 5-month-old mice was 219899, 280340, and 433708, respectively. The total number of sequenced CDR3 sequences was less than in the thymus, spleen, blood, and liver tissues. This may have caused incon-

Analysis of TCR β chain in BALB/c mice

sistency in the analysis of the results. Recently, Chara W proposed a solution to the limitations of HTS TCR diversity analyses [34]. In subsequent in-depth studies, further quality control measures will be carried out based on sample cell numbers and sequencing depth.

CDR3 aa repertoire clonal expansion frequency analysis in all tissues of the different mouse ages indicated that the LEC (<0.001%) CDR3 aa distributions in the thymus and spleen were greater than in the liver and intestine, and the HEC distributions in these tissues (>0.01%) showed the opposite pattern. However, both the LEC and HEC CDR3 aa distributions in the blood fell in between the results of the other tissues. The distribution of the CDR3 aa repertoire clonal frequency in the lymphoid tissues may explain the diversity of the systemic immune response. HECs of the CDR3 aa repertoires in non-lymphoid tissues were associated with a distinct local immune response and the presence of high memory T cells in non-lymphoid tissues. MECs (0.01-0.001%) of the CDR3 aa repertoires in the liver of all mice had the greatest distribution compared to the other tissues in all mice. Thus, a complicated immune response mechanism exists in the liver. More work is required to understand the distribution of MECs in the liver and the greater distribution of HECs in the small intestine.

The thymus of the one-month-old mouse had the highest frequency of unique CDR3 aa overlap sequences of all tissues when compared to the three- and five-month-old BALB/c mice. The top-15 high-frequency CDR3 aa overlap sequences (in the thymus of the one-month mouse) were found at a low frequency in the other tissues. This suggests that the young thymus is a major source of public CDR3 in all tissues under physiological conditions. CDR3 aa overlap sequences had the highest frequency in the three- and five-month-old mice spleens, and the top-15 high-frequency CDR3 aa overlap sequences (in the spleen of three- and five-month-old mice) also had a high frequency in other tissues (Figure S5). This suggests that the spleen is a major resource of public CDR3 in the blood, liver, and small intestine of older mice under physiological conditions. These experimental data agree with the development, distribution and migration of T cells in the thymus as mice age. A statistical analysis across all ages of each mouse tissue for unique CDR3

aa overlap sequences with regards to the total unique CDR3 sequences indicated the greatest amount of overlap in the thymus and spleen (which had the greatest amount of public T cells compared to other tissues). There was a lower amount of overlap in the blood and liver (which had the greatest amount of public T cells compared to the other tissues), with the least amount found in the small intestine (which had the least amount of public T cells).

The CDR3 repertoire length distribution in all tissues and ages of the BALB/c mice showed a normal distribution of 14 aa as the median length, except in the small intestine of the one-month-old mouse, which had a 12 aa median, but how this occurs is uncertain. The amino acid usage in all tissues of all mice was consistent. However, K, M, H, and I usage was abnormally low in the whole TCR β chain CDR3 repertoires, suggesting that K, M, H, and I do not contribute to the response effect of CDR3 and the corresponding antigen.

The usage of TRBV, TRBD, and TRBJ in unique and total CDR3 repertoires did not show significant differences among the thymus, spleen, blood, liver, or small intestine of the one-, three-, and five-month-old mice. There are several notable common characteristics of TRBV, TRBD, and TRBJ usage in the BALB/c mouse CDR3 repertoires-the top-five major TRBV families with the highest frequency of usage are TRBV1, TRBV5, TRBV13, TRBV19, and TRBV31, among which the frequency of the TRBV13 family usage was significantly greater than for other family usages. The TRBJ01-7 (ORF) and TRBJ02-6 (P) families had low frequency usage in the studied tissues, and the TRBJ02-7 family had high-frequency usage. The TRBD01 family was used more than the TRBD02 family in the studied tissues. Characteristics of family usages may be attributed to the composition and structure of the TCR gene during BALB/c animal evolution, which may have contributed to gene rearrangements. Additionally, these gene sequences are associated with effects of the TCR β chain CDR3 antigen.

The insertion or deletion of a unique CDR3 repertoire in the thymus, blood, spleen and liver were different than in the small intestine, and this may be associated with the small CDR3 repertoire of the small intestine in our data. There were no differences in the V and d3 dele-

tions in any tissues among the mice, suggesting that the total number of 3'-end deletions in V and D genes was less than in the TCR gene rearrangement. More research is required to explore the insertion and deletion distribution among the different tissues and ages of BALB/c mice.

Conclusions

In summary, the diversity, clonality and other characteristics of the total T-cell TCR CDR3 repertoire can be identified and compared among tissues using high-throughput sequencing technology and immune bioinformatics approaches. The results from this study suggested that there were significant differences in the composition and characteristics of the total T-cell TCR β chain CDR3 repertoire among lymphoid tissues (thymus and spleen), blood (circulating T cells), and non-lymphoid tissues (liver and small intestine) in one-, three-, and five-month-old BALB/c mice. However, the composition and characteristics of the CDR3 repertoire in the thymus was similar to that in the spleen, and the blood repertoire was similar to the liver; only the small intestine had a unique composition. These characteristics may be associated with the development, maturation, migration, proliferation and effect of distinct T cells in the different tissues. However, this experiment only used a single mouse per condition (one, three, and five months of age), and we did not sort the T cell subgroups from the thymus, spleen, blood, liver and small intestine; therefore, the homogeneity and heterogeneity of the total T-cell TCR β chain CDR3 repertoires lack the cellular composition of the different individual organs. Recently, Paul-Gydéon Ritvo et al. used HTS technology to analyse the TCR CDR3 repertoire of T_H follicular helper (T_{fh}) and T follicular regulatory (T_{fr}) cells and found that the T_{fr} TCR repertoire is more similar to that of Tregs than to that of T_{fh} or T_{eff} cells, suggesting the Treg cell origin of T_{fr} cells [35]. Maceiras AR et al. analysed the T_H CDR3 repertoire, indicating that GC T_{fh} and T_{fr} pools are generated from distinct TCR repertoires, with T_{fh} cells expressing antigen-responsive TCRs to promote antibody responses, and T_{fr} cells expressing potentially autoreactive TCRs to suppress autoimmunity [36]. This requires further study of the TCR β chain CDR3 repertoires of Treg, Th17, T_{fh} or other T-cell subgroups in different mouse tissues and ages.

Acknowledgements

We thank Adaptive Biotechnologies Corporation (Seattle, WA, US) for help with mouse TCR beta chain CDR3 repertoire sequencing and analysis. The work was supported by grants from the National Prophase Project on Basic Research of China (973pre-Program, 2008CB-517310) and the National Natural Science Foundation of China (81660269 & 81441048).

Disclosure of conflict of interest

None.

Address correspondence to: Xinsheng Yao, Department of Immunology, Zunyi Medical University, Zunyi 563003, China. Tel: 0851-28642716; E-mail: immunology01@126.com

References

- [1] Masopust D, Vezys V, Marzo AL, Lefrançois L. Preferential localization of effector memory cells in nonlymphoid tissue. *Science* 2001; 291: 2413-2417.
- [2] Masopust D, Vezys V, Wherry EJ, Barber DL, Ahmed R. Cutting edge: gut microenvironment promotes differentiation of a unique memory CD8 T cell population. *J Immunol* 2006; 176: 2079-83.
- [3] Casey KA, Fraser KA, Schenkel JM, Moran A, Abt MC, Beura LK, Lucas PJ, Artis D, Wherry EJ, Hogquist K, Vezys V, Masopust D. Antigen-independent differentiation and maintenance of effector-like resident memory T cells in tissues. *J Immunol* 2012; 188: 4866-75.
- [4] Hovhannisyan Z, Treatman J, Dan RL and Mayer L. Characterization of interleukin-17-producing regulatory T cells in inflamed intestinal mucosa from patients with inflammatory bowel diseases. *Gastroenterology* 2011; 140: 957-965.
- [5] Xiaodong J, Clark RA, Luzheng L, Wagers AJ, Fuhlbrigge RC and Kupper TS. Skin infection generates non-migratory memory CD8+ T(RM) cells providing global skin immunity. *Nature* 2012; 483: 227-231.
- [6] Clark RA, Watanabe R, Teague JE, Schlapbach C, Tawa MC, Adams N, Dorosario AA, Chaney KS, Cutler CS, Leboeuf NR, Carter JB, Fisher DC, Kupper TS. Skin effector memory T cells do not recirculate and provide immune protection in alemtuzumab-treated CTCL patients. *Sci Transl Med* 2012; 4: 117ra7.
- [7] Rahul P, James C, George M, Richards WG, Clark RA and Kupper TS. Resident memory T cells (T(RM)) are abundant in human lung: di-

Analysis of TCR β chain in BALB/c mice

- iversity, function, and antigen specificity. *PLoS One* 2011; 6: e16245.
- [8] Teijaro JR, Turner D, Pham Q, Wherry EJ, Lefrançois L and Farber DL. Cutting edge: tissue-retentive lung memory CD4 T cells mediate optimal protection to respiratory virus infection. *J Immunol* 2011; 187: 5510-4.
- [9] Anna C, Michal S, Rempala GA, Simarjot Singh P, Mcindoe RA, Denning TL, Lynn B, Piotr K, Pawel K and Leszek I. Thymus-derived regulatory T cells contribute to tolerance to commensal microbiota. *Nature* 2013; 497: 258-262.
- [10] Emerson RO, Sherwood AM, Rieder MJ, Jamie G, Williamson DW, Carlson CS, Drescher CW, Muneesh T, Bielas JH and Robins HS. High-throughput sequencing of T-cell receptors reveals a homogeneous repertoire of tumour-infiltrating lymphocytes in ovarian cancer. *J Pathol* 2013; 231: 433-440.
- [11] Shifrut E, Baruch K, Gal H, Ndifon W, Deczkowska A, Schwartz M and Friedman N. CD4+ T cell-receptor repertoire diversity is compromised in the spleen but not in the bone marrow of aged mice due to private and sporadic clonal expansions. *Front Immunol* 2013; 4: 379.
- [12] Sathaliyawala T, Kubota M, Yudanin N, Turner D, Camp P, Thome JC, Bickham K, Lerner H, Goldstein M and Sykes M. Distribution and compartmentalization of human circulating and tissue-resident memory T cell subsets. *Immunity* 2013; 38: 187-197.
- [13] Bergot AS, Chaara W, Ruggiero E, Mariotti-Ferandiz E, Dulauroy S, Schmidt M, von Kalle C, Six A and Klatzmann D. TCR sequences and tissue distribution discriminate the subsets of naïve and activated/memory Treg cells in mice. *Eur J Immunol* 2015; 45: 1524-34.
- [14] Carlson CS, Emerson RO, Sherwood AM, Desmarais C, Chung MW, Parsons JM, Steen MS, Lamadriherrmannsfeldt MA, Williamson DW and Livingston RJ. Using synthetic templates to design an unbiased multiplex PCR assay. *Nat Commun* 2013; 4: 2680.
- [15] Marrero I, Hamm DE and Davies JD. High-throughput sequencing of islet-infiltrating memory CD4(+) T cells reveals a similar pattern of TCR V β usage in prediabetic and diabetic NOD mice. *PLoS One* 2013; 8: e76546.
- [16] Nunes-Alves C, Booty MG, Carpenter SM, Rothchild AC, Martin CJ, Desjardins D, Steblenko K, Kløverpris HN, Madansein R, Ramsuran D, Leslie A, Correia-Neves M, Behar SM. Human and murine clonal CD8+ T cell expansions arise during tuberculosis because of TCR selection. *PLoS Pathog* 2015; 11: e1004849.
- [17] Robins HS, Campregher PV, Srivastava SK, Wacher A, Turtle CJ, Kahsai O, Riddell SR, Warren EH, Carlson CS. Comprehensive assessment of T-cell receptor beta-chain diversity in alphabeta T cells. *Blood* 2009; 114: 4099-4107.
- [18] Toivonen R, Arstila TP and Hanninen A. Islet-associated T-cell receptor-beta CDR sequence repertoire in prediabetic NOD mice reveals antigen-driven T-cell expansion and shared usage of VbetaJbeta TCR chains. *Mol Immunol* 2015; 64: 127-135.
- [19] van Heijst JW, Ceberio I, Lipuma LB, Samilo DW, Wasilewski GD, Gonzales AM, Nieves JL, van den Brink MR, Perales MA, Pamer EG. Quantitative assessment of T cell repertoire recovery after hematopoietic stem cell transplantation. *Nat Med* 2013; 19: 372-7.
- [20] Lefranc MP. 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; 2011: 633-642.
- [21] Yousfi Monod M, Giudicelli V, Chaume D, Lefranc MP. IMGT/JunctionAnalysis: the first tool for the analysis of the immunoglobulin and T cell receptor complex V-J and V-D-J JUNCTIONS. *Bioinformatics* 2004; 20: 379-385.
- [22] Gorski J, Yassai M, Zhu X, Kissela B, Keever C, Flomenberg N. Circulating T cell repertoire complexity in normal individuals and bone marrow recipients analyzed by CDR3 size spectratyping. Correlation with immune status. *J Immunol* 1994; 152: 5109-19.
- [23] Yao XS, Diao Y, Sun WB, Luo JM, Qin M and Tang XY. Analysis of the CDR3 length repertoire and the diversity of TCR alpha chain in human peripheral blood T lymphocytes. *Cell Mol Immunol* 2007; 4: 215-20.
- [24] Arstila TP, Casrouge A, Baron V, Even J, Kanellopoulos J and Kourilsky P. A direct estimate of the human alphabeta T cell receptor diversity. *Science* 1999; 286: 958-961.
- [25] Casrouge A, Beaudoin E, Dalle S, Pannetier C, Kanellopoulos J and Kourilsky P. Size estimate of the $\alpha\beta$ TCR repertoire of naive mouse splenocytes. *J Immunol* 2000; 164: 5782-7.
- [26] Freeman JD, Warren RL, Webb JR, Nelson BH, Holt RA. Profiling the T-cell receptor beta-chain repertoire by massively parallel sequencing. *Genome Res* 2009; 19: 1817-24.
- [27] Emerson R, Sherwood A, Desmarais C, Malhotra S, Phippard D and Robins H. Estimating the ratio of CD4+ to CD8+ T cells using high-throughput sequence data. *J Immunol Methods* 2013; 391: 14-21.
- [28] Fang H, Rui Y, Liu X, Daigo Y, Yew PY, Tanikawa C, Matsuda K, Imoto S, Miyano S and Nakamura Y. Quantitative T cell repertoire analysis by deep cDNA sequencing of T cell receptor α and β chains using next-generation sequencing (NGS). *Oncoimmunology* 2015; 3: e968467.

Analysis of TCR β chain in BALB/c mice

- [29] Li Z, Long M, ChunMei L, Bin S, Jiang Y, Rui M, Qingqing M and XinSheng Y. Composition and variation analysis of TCR β -chain CDR3 repertoire in the thymus and spleen of MRL/lpr mouse at different ages. *Immunogenetics* 2015; 67: 25-37.
- [30] Long M, Yang L, Shi B, He X, Peng A, Li Y, Teng Z, Sun S, Rui M and 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.
- [31] Six A, Mariotti-Ferrandiz ME, Chaara W, Magadan S, Pham HP, Lefranc MP, Mora T, Thomas-Vaslin V, Walczak AM and Boudinot P. The past, present, and future of immune repertoire biology - the rise of next-generation repertoire analysis. *Front Immunol* 2013; 4: 413.
- [32] Messaoudi I, Lemaoult J, Guevara-Patino JA, Metzner BM, Nikolich-Zugich J. Age-related CD8 T cell clonal expansions constrict CD8 T cell repertoire and have the potential to impair immune defense. *J Exp Med* 2004; 200: 1347-58.
- [33] Xiaoyan HE, Wang P and Rui MA. Analysis of the TRBV-TRBJ1-1 CDR3 repertoire in the thymus, spleen, liver, small intestine, and blood of one-month-old Balb/c mouse. *Immunological Journal* 2012; 10.
- [34] Chaara W, Gonzalez-Tort A, Florez LM, Klatzmann D, Mariotti-Ferrandiz E and Six A. Rep-Seq data representativeness and robustness assessment by shannon entropy. *Front Immunol* 2018; 9: 1038.
- [35] Ritvo PG, Chaara W, El Soufi K, Bonnet B, Six A, Mariotti-Ferrandiz E and Klatzmann D. High-resolution repertoire analysis of Tfr and Tfh cells reveals unexpectedly high diversities indicating a bystander activation of follicular T cells. *BioRxiv* 2017; 231977.
- [36] Maceiras AR, Almeida SCP, Mariotti-Ferrandiz E, Chaara W, Jebbawi F, Six A, Hori S, Klatzmann D, Faro J, Graca L. T follicular helper and T follicular regulatory cells have different TCR specificity. *Nat Commun* 2017; 8: 15067.

Analysis of TCR β chain in BALB/c mice

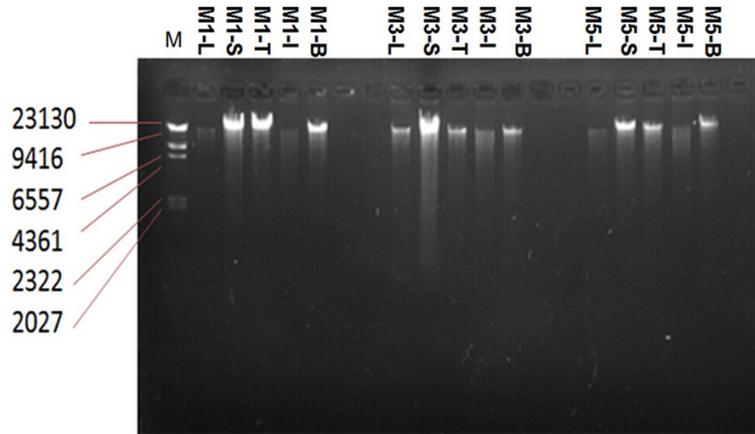


Figure S1. Agarose gel electrophoresis of genomic DNA extracted thymus, spleen, blood, liver, and small intestine of all ages of mice. Note: M1: mouse at 1 month old; M3: mouse at 3 months old ; M5: mouse at 5 months old; T: thymus tissue; S: spleen tissue; B: blood; L: liver tissue; I: small intestine.

Analysis of TCR β chain in BALB/c mice

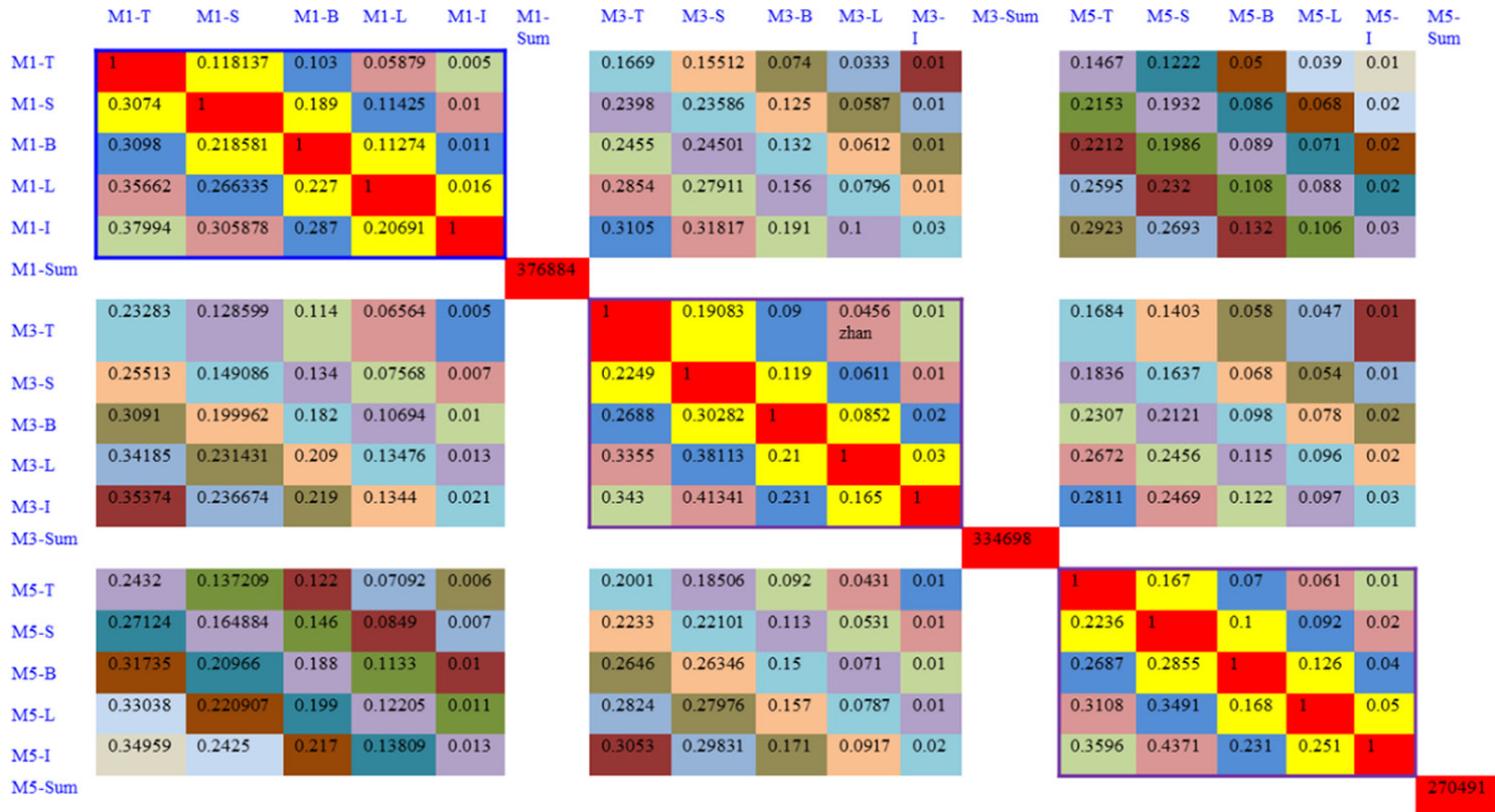


Figure S2. Percent of unique CDR3 amino acid sequence overlap of total unique CDR3 amino acid sequences in thymus, spleen, blood, liver, and small intestine of all ages of mice.

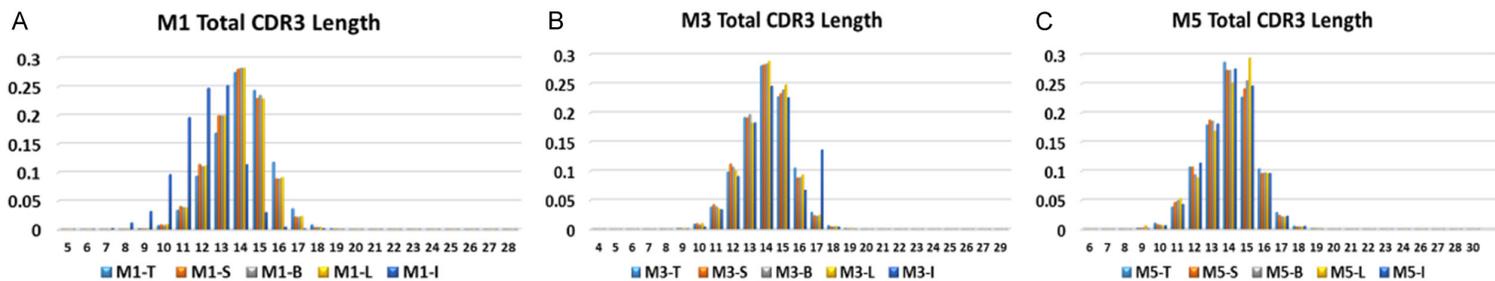


Figure S3. Total CDR3 repertoire length distribution in thymus, spleen, blood, liver, and small intestine for all mouse ages.

Analysis of TCR β chain in BALB/c mice

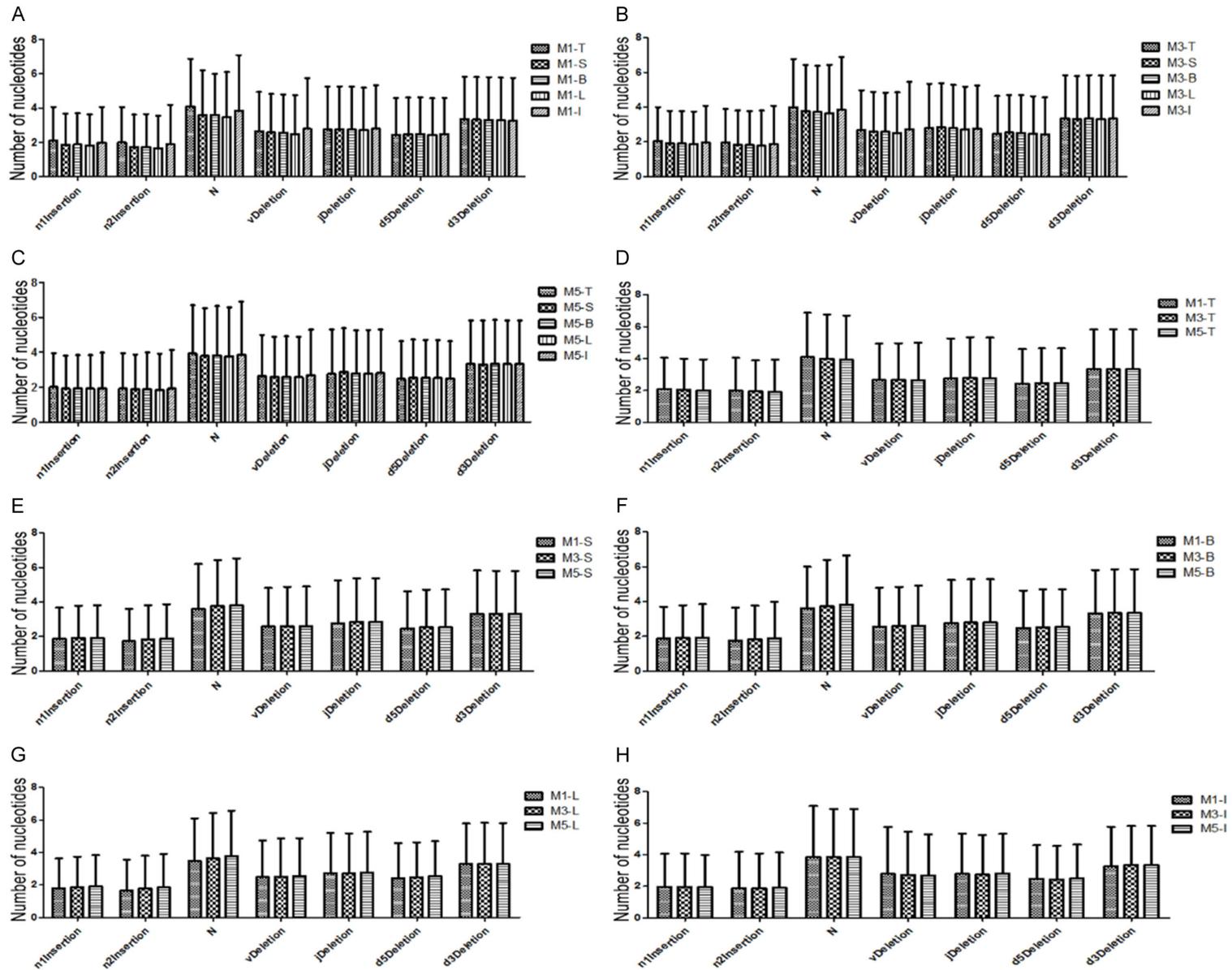


Figure S4. Statistical analysis of n1, n2 and both insertions, V, J, D5 and D3 deletions in thymus, spleen, blood, liver, and small intestine of all ages of mice.

Analysis of TCR β chain in BALB/c mice

Table S1. *p*-value pairwise comparisons of insertions and deletions among thymus, spleen, blood, liver, and small intestine of all ages of mice

	M1T-M1S	M1T-M1B	M1T-M1L	M1T-M1I	M3T-M3S	M3T-M3B	M3T-M3L	M3T-M3I	M5T-M5S	M5T-M5B	M5T-M5L	M5T-M5I
Row Factor	<i>P</i> value											
n1 Insertion	P<0.001	P<0.001	P<0.001	P<0.01	P<0.001	P<0.001	P<0.001	P>0.05	P<0.001	P<0.001	P<0.001	P<0.05
n2 Insertion	P<0.001	P<0.001	P<0.001	P<0.01	P<0.001	P<0.001	P<0.001	P>0.05	P<0.001	P>0.05	P<0.001	P>0.05
N	P<0.001	P>0.05										
v Deletion	P<0.001	P<0.001	P<0.001	P<0.01	P<0.001	P<0.001	P<0.001	P>0.05	P<0.001	P<0.001	P<0.001	P>0.05
j Deletion	P<0.01	P>0.05	P>0.05	P>0.05	P<0.001	P>0.05	P<0.001	P>0.05	P<0.001	P>0.05	P>0.05	P>0.05
d5 Deletion	P<0.001	P<0.001	P>0.05	P>0.05	P<0.001	P<0.001	P>0.05	P>0.05	P<0.001	P<0.001	P<0.001	P>0.05
d3 Deletion	P<0.05	P<0.001	P<0.001	P>0.05	P<0.001	P>0.05	P>0.05	P>0.05	P<0.01	P>0.05	P>0.05	P>0.05
	M1S-M1B	M1S-M1L	M1S-M1I		M3S-M3B	M3S-M3L	M3S-M3I		M5S-M5B	M5S-M5L	M5S-M5I	
Row Factor	<i>P</i> value	<i>P</i> value	<i>P</i> value		<i>P</i> value	<i>P</i> value	<i>P</i> value		<i>P</i> value	<i>P</i> value	<i>P</i> value	
n1 Insertion	P>0.05	P<0.001	P<0.01		P>0.05	P<0.01	P>0.05		P>0.05	P>0.05	P>0.05	
n2 Insertion	P>0.05	P<0.001	P<0.001		P>0.05	P<0.001	P>0.05		P>0.05	P>0.05	P>0.05	
N	P<0.05	P<0.001	P<0.001		P<0.001	P<0.001	P<0.05		P>0.05	P>0.05	P>0.05	
v Deletion	P>0.05	P<0.001	P<0.001		P>0.05	P<0.001	P<0.001		P>0.05	P>0.05	P<0.05	
j Deletion	P>0.05	P<0.001	P>0.05		P<0.001	P<0.001	P<0.01		P<0.001	P<0.001	P>0.05	
d5 Deletion	P>0.05	P>0.05	P>0.05		P>0.05	P<0.001	P<0.05		P>0.05	P>0.05	P>0.05	
d3 Deletion	P>0.05	P<0.05	P>0.05		P<0.001	P>0.05	P>0.05		P<0.01	P>0.05	P>0.05	
	M1B-M1L	M1B-M1I			M3B-M3L	M3B-M3I			M5B-M5L	M5B-M5I		
Row Factor	<i>P</i> value	<i>P</i> value			<i>P</i> value	<i>P</i> value			<i>P</i> value	<i>P</i> value		
n1 Insertion	P<0.001	P<0.05			P>0.05	P>0.05			P>0.05	P>0.05		
n2 Insertion	P<0.001	P<0.001			P<0.05	P>0.05			P>0.05	P>0.05		
N	P<0.001	P<0.001			P<0.001	P<0.001			P<0.05	P>0.05		
v Deletion	P<0.001	P<0.001			P<0.001	P<0.001			P>0.05	P<0.05		
j Deletion	P<0.05	P>0.05			P<0.001	P>0.05			P>0.05	P>0.05		
d5 Deletion	P<0.05	P>0.05			P<0.05	P>0.05			P>0.05	P>0.05		
d3 Deletion	P>0.05	P>0.05			P>0.05	P>0.05			P>0.05	P>0.05		
	M1L-M1I				M3L-M3I				M5L-M5I			
Row Factor	<i>P</i> value				<i>P</i> value				<i>P</i> value			
n1 Insertion	P<0.001				P<0.05				P>0.05			
n2 Insertion	P<0.001				P<0.05				P>0.05			
N	P<0.001				P<0.001				P<0.05			
v Deletion	P<0.001				P<0.001				P<0.01			
j Deletion	P>0.05				P>0.05				P>0.05			
d5 Deletion	P>0.05				P>0.05				P>0.05			
d3 Deletion	P>0.05				P>0.05				P>0.05			

Analysis of TCR β chain in BALB/c mice

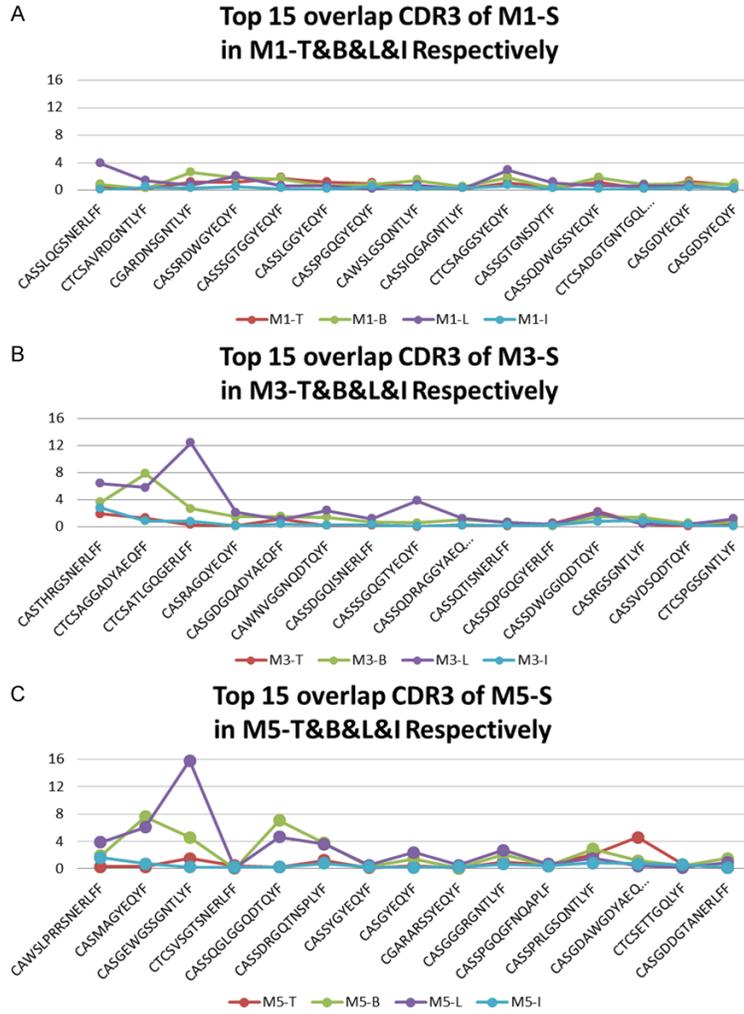


Figure S5. The distribution of the top 15 high-frequency CDR3 aa overlap sequences (spleen) in the thymus, blood, liver and small intestine of M1, M3 and M5 separately.