Data in Brief 35 (2021) 106751

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

# Characterization of human T cell receptor repertoire data in eight thymus samples and four related blood samples

Nelli Heikkilä[a,b,*], Iivari Kleino[a], Reetta Vanhanen[a,b], Dawit A. Yohannes[a,c], Ilkka P. Mattila[d], Jari Saramäki[e], T. Petteri Arstila[a,b,*]

[a] Research Programs Unit, Translational Immunology, University of Helsinki, Haartmaninkatu 3, 00290 Helsinki, Finland
[b] Medicum, Department of Bacteriology and Immunology, University of Helsinki, Haartmaninkatu 3, 00290 Helsinki, Finland
[c] Department of Medical and Clinical Genetics, University of Helsinki, Haartmaninkatu 8, 00290 Helsinki, Finland
[d] Department of Pediatric Cardiac and Transplantation Surgery, Hospital for Children and Adolescents, Helsinki University Central Hospital, Stenbäckinkatu 9, 00290 Helsinki, Finland
[e] Department of Computer Science, Aalto University, Konemiehentie 2, 02150 Espoo, Finland

ARTICLE INFO

*Article history:*
Received 8 September 2020
Revised 22 December 2020
Accepted 11 January 2021
Available online 20 January 2021

*Keywords:*
TCR repertoire
CDR3
Public immune responses
thymus

ABSTRACT

T cell receptor (TCR) is a heterodimer consisting of TCR$\alpha$ and TCR$\beta$ chains that are generated by somatic recombination of multiple gene segments. Nascent TCR repertoire undergoes thymic selections where non-functional and potentially autoreactive receptors are removed. During the last years, the development of high-throughput sequencing technology has allowed a large scale assessment of TCR repertoire and multiple analysis tools are now also available.

In our recent manuscript, *Human thymic T cell repertoire is imprinted with strong convergence to shared sequences* [1], we show highly overlapping thymic TCR repertoires in unrelated individuals. In the current Data in Brief article, we provide a more detailed characterization of the basic features of these thymic and related peripheral blood TCR repertoires. The thymus samples were collected from eight infants undergoing corrective cardiac surgery, two of whom were monozygous

DOI of original article: 10.1016/j.molimm.2020.09.003
* Corresponding authors.
*E-mail addresses:* nelli.heikkila@helsinki.fi (N. Heikkilä), petteri.arstila@helsinki.fi (T.P. Arstila).

https://doi.org/10.1016/j.dib.2021.106751

twins [2]. In parallel with the surgery, a small aliquot of peripheral blood was drawn from four of the donors. Genomic DNA was extracted from mechanically released thymocytes and circulating leukocytes. The sequencing of TCR$\alpha$ and TCR$\beta$ repertoires was performed at ImmunoSEQ platform (Adaptive Biotechnologies). The obtained repertoire data were analysed applying relevant features from immunoSEQ® 3.0 Analyzer (Adaptive Biotechnologies) and a freely available VDJTools software package for programming language R [3]. The current data analysis displays the basic features of the sequenced repertoires including observed TCR diversity, various descriptive TCR diversity measures, and V and J gene usage. In addition, multiple methods to calculate repertoire overlap between two individuals are applied. The raw sequence data provide a large database of reference TCRs in healthy individuals at an early developmental stage. The data can be exploited to improve existing computational models on TCR repertoire behaviour as well as in the generation of new models.



## Specifications Table

| | |
|---|---|
| Subject | Immunology |
| Specific subject area | T cell antigen receptor (TCR) alpha chain and beta chain diversity and characteristics in thymus and in peripheral blood |
| Type of data | Table: Sample description by immunoSEQ and VDJTools softwares (Table 1), repertoire diversity metrics (Table 2), resampled repertoire diversity metrics (Table 3), repertoire overlap measures (Table 4).<br>Graph: V gene usage heatmap (Figure 1), J gene usage heatmap (Figure 2), rarefaction plots (Figure 3), clustering of overlap analyses (Figure 4). |
| How data were acquired | TCRAD and TCRB sequencing was performed at ImmunoSEQ platform (Adaptive Biotechnologies). TCR analysis was performed using immunoSEQ® 3.0 Analyzer (Adaptive Biotechnologies) and VDJTools software [3]. |
| Data format | Raw<br>Analysed |
| Parameters for data collection | Thymus samples were obtained from eight immunologically healthy infants undergoing open cardiac surgery for congenital heart defects. A small aliquot of blood (0.5–1 mL) was drawn from four subjects during the operation. The study was approved by the Pediatric Ethical Committee of the Helsinki University Hospital (HUS/747/2019) and a written informed consent was obtained from the parents. |
| Description of data collection | Thymocytes were extracted mechanically from tissue resects. Blood samples were treated with ACK lysis buffer (Thermo Fisher Scientific) to remove erythrocytes. DNA was extracted from 10–30 million thymocytes and from all available PBMCs. TCRAD and TCRB sequencing was performed as previously described [4] from a standardized quantity of genomic DNA using ImmunoSEQ assay (Adaptive Biotechnologies), which exploits a multiplex PCR system spanning the V(D)J region at a length that is sufficient to identify V and J genes and cover unique CDR3 regions. |
| Data source location | Institution: University of Helsinki<br>City/Town/Region: Helsinki<br>Country: Finland |
| Data accessibility | Repository name: The European Nucleotide Archive (ENA) at EMBL-EBI<br>Data identification number: PRJEB41936<br>Direct URL to data: https://www.ebi.ac.uk/ena/browser/view/PRJEB41936 |

(*continued on next page*)

| | |
|---|---|
| Related research article | Heikkilä Nelli, Vanhanen Reetta, Yohannes Dawit A., Kleino Iivari, Mattila Ilkka P., Saramäki Jari, Arstila T. Petteri<br>Human thymic T cell repertoire is imprinted with strong convergence to shared sequences<br>Molecular Immunology, Volume 127, November 2020, Pages 112–123<br>doi.org/10.1016/j.molimm.2020.09.003 |

## Value of the Data

- These data consist of a unique collection of over 62 million T cell receptor (TCR) sequences obtained directly from human thymus. It is a large scale resource of human TCR$\alpha$ and TCR$\beta$ repertoires at an early developmental stage before clonal selections by peripheral antigens and devoid of medical or immunological interventions.
- The data are useful for those who wish to compare TCR repertoires from healthy thymus and from individuals affected by immunological diseases or other medical conditions. The large scale thymic repertoire data can also benefit computational experiments which have been typically limited to peripheral blood TCR data.
- These data can be directly exploited to improve existing computational models on TCR repertoire generation as well as in the generation of new models. These data can also guide design of human TCR sequencing experiments and serve as a reference database for new experiments.

## 1. Data Description

All TCRAD and TCRB sequences obtained from eight thymus (donors A-D and donors 1–4) and four related blood samples (donors 1–4) have been deposited in the European Nucleotide Archive (ENA) at EMBL-EBI under accession number PRJEB41936 (https://www.ebi.ac.uk/ena/browser/view/PRJEB41936). In addition, the sequences are available at immuneACCESS® repository in the form of immunoSEQ™ output format and can also be downloaded as raw FASTA files (https://clients.adaptivebiotech.com/pub/heikkila-2020-mi). On average, we obtained 4.1 million unique TCR$\alpha$ and 810 000 unique TCR$\beta$ clonotypes from each thymus. From blood samples we obtained on average 150 000 and 84 000 unique TCR$\alpha$ and TCR$\beta$ sequences, respectively. An overview of sequence diversities, total counts and sequence productivity (in-frame vs. non-coding sequences) was generated both by immunoSEQ™ and VDJTools softwares and is displayed together with donor details in Table 1. Two of the donors (A and B) were monozygous twins and the influence of genetics in the repertoire has been analysed previously [2]. The V and J gene usage has been shown to be biased in the peripheral blood but also already in the thymus [5–7]. The gene segment usage in the current samples is also biased (Figs. 1 & 2).

The TCR diversity has been previously assessed both in the peripheral blood and in the thymus and multiple diversity metrics are available [4,8–10]. The diversity estimates for the current samples were calculated using VDJTools software with default settings. To estimate the lower bound of total species richness, VDJTools provide unmodified Chao1, extrapolated Chao (chaoE) and Efron-Thisted estimates while the repertoire diversity is depicted with Shannon's index and inverse Simpson's index (Table 2). The species richness and repertoire diversity indexes are also calculated for datasets down-sampled to the size of the smallest dataset to facilitate the comparison of samples with different sequencing depths (Table 3). Furthermore, a rarefaction curve based on the relationship between the sample diversity and the sample size was plotted for TCR$\alpha$ and TCR$\beta$ with extrapolation to the size of the largest sample (Fig. 3).

Despite the high potential diversity of TCR repertoires, a surprisingly high fraction of the repertoire is shared between individuals [1]. Here, we calculated various overlap measures with VDJTools: Pearson correlation, relative overlap measure [rationale explained in 11], Jaccard index

**Table 1**
Description of the sequenced samples.

| TCRAD | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Sample id | Age (days) | Sex | Tissue | immunoSEQ™: count | VDJTools: count | immunoSEQ™: diversity | VDJTools: diversity | immunoSEQ™: non-coding diversity | VDJTools: non-coding diversity | immunoSEQ™: non-coding frequency | VDJTools: non-coding frequency |
| thymus A | 243 | M | thymus | 11 838 086 | 11 838 086 | 6 907 422 | 6 763 870 | 4 719 902 | 2 090 241 | 68.33% | 31.10% |
| thymus B | 244 | M | thymus | 12 849 473 | 12 849 473 | 7 578 104 | 7 419 245 | 5 179 754 | 2 307 719 | 68.35% | 31.22% |
| thymus C | 225 | F | thymus | 8 359 283 | 8 359 283 | 5 347 824 | 5 259 057 | 3 663 398 | 1 752 208 | 68.50% | 33.98% |
| thymus D | 126 | M | thymus | 11 063 464 | 11 063 464 | 6 743 495 | 6 610 182 | 4 617 533 | 2 007 905 | 68.47% | 30.29% |
| thymus 1 | 7 | M | thymus | 3 179 774 | 3 179 774 | 2 089 557 | 1 984 292 | 1 447 726 | 601 486 | 69.28% | 30.01% |
| thymus 2 | 52 | M | thymus | 1 747 487 | 1 747 487 | 1 262 845 | 1 198 677 | 883 536 | 385 749 | 69.96% | 32.05% |
| thymus 3 | 107 | M | thymus | 2 158 043 | 2 158 043 | 1 289 728 | 1 230 436 | 902 934 | 398 227 | 70.01% | 32.24% |
| thymus 4 | 156 | F | thymus | 1 848 851 | 1 848 851 | 1 419 013 | 1 345 927 | 997 764 | 441 964 | 70.31% | 32.71% |
| blood 1 | 7 | M | blood | 154 682 | 154 682 | 138 159 | 130 307 | 86 201 | 34 147 | 62.39% | 26.02% |
| blood 2 | 52 | M | blood | 123 523 | 123 523 | 109 171 | 103 142 | 65 096 | 26 413 | 59.63% | 25.12% |
| blood 3 | 107 | M | blood | 245 126 | 245 126 | 180 100 | 170 333 | 110 571 | 45 852 | 61.39% | 24.94% |
| blood 4 | 156 | F | blood | 199 326 | 199 326 | 167 266 | 157 728 | 104 846 | 45 404 | 62.68% | 27.89% |
| TCRB | | | | | | | | | | |
| Sample id | Age (days) | Sex | Tissue | immunoSEQ™: count | VDJTools: count | immunoSEQ™: diversity | VDJTools: diversity | immunoSEQ™: non-coding diversity | VDJTools: non-coding diversity | immunoSEQ™: non-coding frequency | VDJTools: non-coding frequency |
| thymus A | 243 | M | thymus | 1 647 656 | 1 647 656 | 1 254 760 | 1 245 029 | 288 199 | 108 933 | 22.97% | 8.46% |
| thymus B | 244 | M | thymus | 1 783 878 | 1 783 878 | 1 540 161 | 1 526 694 | 363 551 | 138 386 | 23.60% | 8.89% |
| thymus C | 225 | F | thymus | 1 850 299 | 1 850 299 | 1 568 528 | 1 551 603 | 248 898 | 93 724 | 15.87% | 5.89% |
| thymus D | 126 | M | thymus | 1 726 796 | 1 726 796 | 1 462 150 | 1 449 881 | 279 672 | 106 019 | 19.13% | 7.15% |
| thymus 1 | 7 | M | thymus | 237 063 | 237 063 | 223 725 | 222 925 | 53 389 | 19 585 | 23.86% | 8.76% |
| thymus 2 | 52 | M | thymus | 182 356 | 182 356 | 173 368 | 172 746 | 35 779 | 14 443 | 20.64% | 8.28% |
| thymus 3 | 107 | M | thymus | 142 903 | 142 903 | 138 544 | 137 920 | 31 385 | 12 183 | 22.65% | 8.75% |
| thymus 4 | 156 | F | thymus | 128 228 | 128 228 | 122 195 | 121 483 | 25 475 | 10 129 | 20.85% | 8.24% |
| blood 1 | 7 | M | blood | 82 418 | 82 418 | 77 868 | 77 281 | 21 203 | 7 462 | 27.23% | 9.51% |
| blood 2 | 52 | M | blood | 73 945 | 73 945 | 69 875 | 69 404 | 17 566 | 6 783 | 25.14% | 9.81% |
| blood 3 | 107 | M | blood | 134 110 | 134 110 | 104 236 | 103 551 | 26 162 | 9 831 | 25.10% | 8.73% |
| blood 4 | 156 | F | blood | 88 901 | 88 901 | 82 550 | 81 935 | 20 852 | 8 151 | 25.26% | 10.19% |

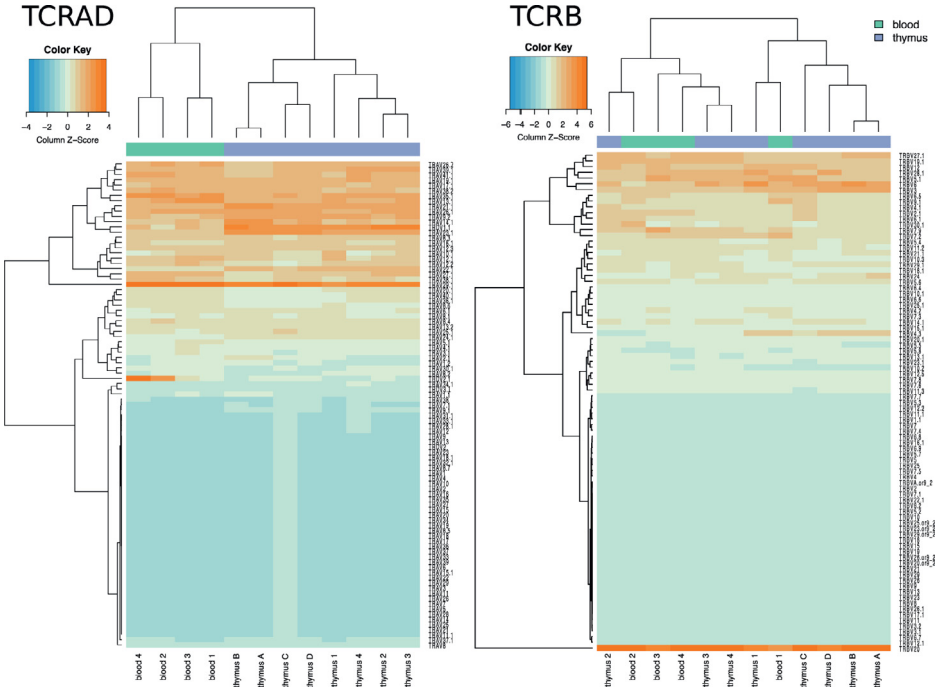


**Fig. 1.** The V gene usage in TCRAD and TCRB repertoires. Z-scores indicate the relative frequency of each segment. Dendrograms show clustering of the samples and the gene segments.

and Morisita-Horn index (Table 4). The calculations were performed on the entire repertoire and exact matching of V gene, J gene and the CDR3 region was required. The clustering of different samples with multidimensional scaling is depicted for Jaccard index (Fig. 4).

## 2. Experimental Design, Materials and Methods

Thymus samples were obtained from eight immunologically healthy infants undergoing corrective cardiac surgery for congenital heart defects. The study was approved by the Pediatric Ethical Committee of the Helsinki University Hospital (HUS/747/2019). A written informed consent was obtained from the parents. Thymocytes were extracted mechanically from tissue resects and stored as pellets of 10–30 million thymocytes in −70 °C. From four donors a small aliquot of 0.5–1 mL peripheral blood was drawn during the surgery. To remove erythrocytes, the blood samples were treated with ACK lysis buffer (Thermo Fisher Scientific, USA) according to manufacturer's orders and the obtained leukocytes were stored as pellets in −70 °C. Genomic DNA was extracted from frozen pellets with QIASymphony™ (Qiagen, Germany) according to manufacturer's orders. TCRAD and TCRB regions were sequenced from a standardized quantity of quality-controlled genomic DNA using ImmunoSEQ™ assay (Adaptive Biotechnologies). The assay uses a multiplex PCR system spanning the TCRAD VJ and TCRB VDJ regions at a length that is sufficient to cover unique CDR3 regions and to identify V and J genes. Amplicon sequencing was performed on Illumina platform. TCRAD and TCRB definitions were based on IMGT database (www.imgt.org). Primer bias and sequencing errors were corrected as previously described [4].

For each sequenced sample the ImmunoSEQ™ assay outputs a file of unique nucleotide sequences covering V and J genes and the CDR3 region, the count and frequency of each sequence, the CDR3 region length, and whether the sequence is in-frame, out-of-frame or contains a

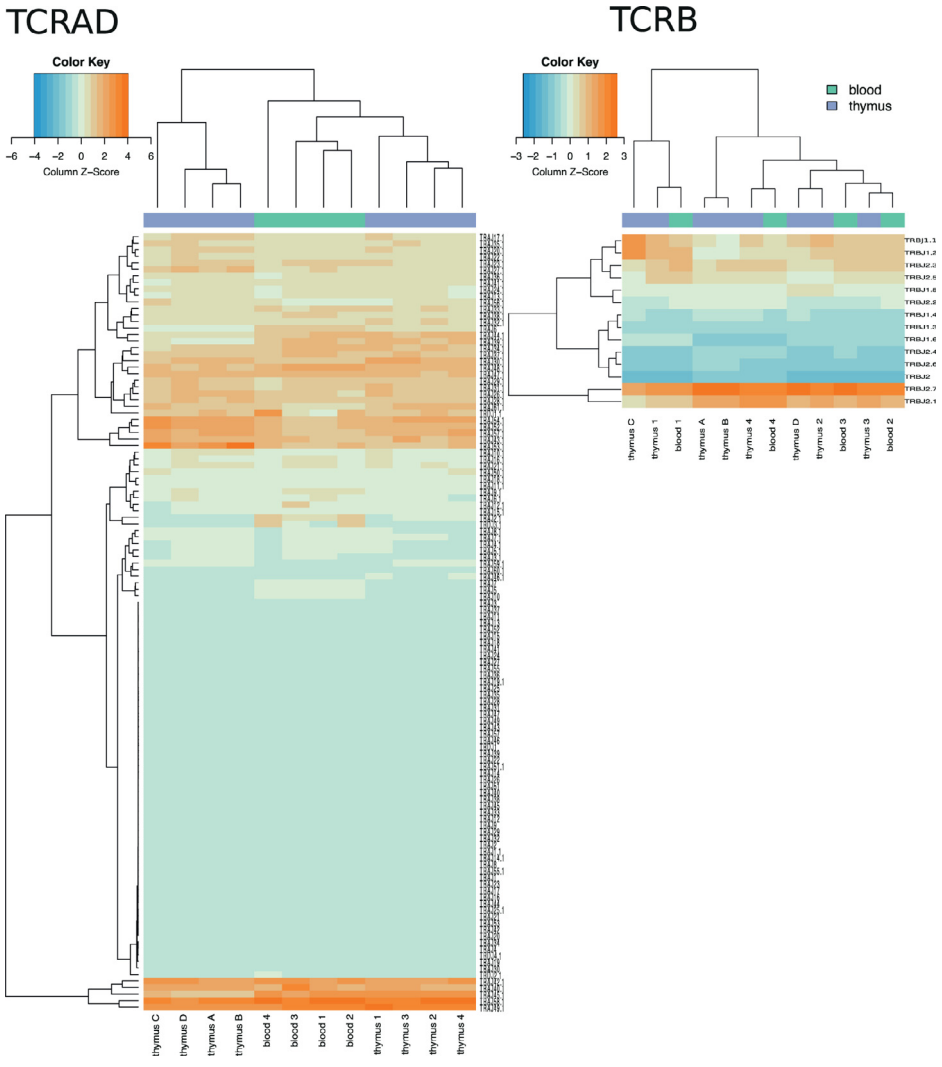


**Fig. 2.** The J gene usage in TCRAD and TCRB repertoires. Z-scores indicate the relative frequency of each segment. Dendrograms show clustering of the samples and the gene segments.

premature STOP codon. For in-frame and 'has stop' sequences the nucleotide sequence is converted to CDR3 amino acid sequence and * symbol indicates the STOP codon. In addition, the V gene, D gene and J gene names, the number of non-templated nucleotide insertions and the locations of insertions in V and J gene segments are provided. The raw FASTA files are also available but not directly used in the present analysis.

In the current article we applied TCR analysis tools form two platforms: immunoSEQ$^{TM}$ ANALYZER 3.0 run on Adaptive Biotech website (adaptivebiotech.com/products-services/immunoseq/immunoseq-analyzer/) and a java based non-commercial software package VDJTools [3]. From immunoSEQ$^{TM}$ we adapted "Sample Overview" to calculate the sample diversity and counts. VDJTools readily accepts the basic immunoSEQ$^{TM}$ output format and converts it to a VDJTools output file. From VDJTools we used "CalcBasicStats" command to calculate the sample diversity and counts, "CalcSegmentUsage" command to produce V and J gene usage heatmaps,

**Table 2**
Diversity estimates.

TCRAD

| Sample id | Age (days) | Sex | Tissue | Observed counts | Observed diversity | Chao1 (mean±std) | Extrapolated ChaoE (mean±std) | Efron-Thisted (mean±std) | Shannon's index (mean) | Inversed Simpson's index (mean) |
|---|---|---|---|---|---|---|---|---|---|---|
| thymus A | 243 | M | thymus | 11 838 086 | 6 763 870 | 16 826 997 ± 13 658 | 7 160 320 ± 2 017 | 46 864 937 ± 2 469 178 | 4 617 696 | 1 633 144 |
| thymus B | 244 | M | thymus | 12 849 473 | 7 419 245 | 18 828 064 ± 14 772 | 7 419 245 ± 2 120 | 41 786 084 ± 2 534 757 | 5 058 759 | 1 796 531 |
| thymus C | 225 | F | thymus | 8 359 283 | 5 259 057 | 14 602 322 ± 14 425 | 7 124 951 ± 2 068 | 45 140 688 ± 3 293 359 | 3 861 431 | 1 603 852 |
| thymus D | 126 | M | thymus | 11 063 464 | 6 610 182 | 16 698 624 ± 13 750 | 7 342 590 ± 2 021 | 23 898 564 ± 1 279 838 | 4 677 030 | 1 764 774 |
| thymus 1 | 7 | M | thymus | 3 179 774 | 1 984 292 | 4 754 478 ± 6 900 | 4 141 693 ± 3 605 | 11 705 484 ± 692 435 | 1 588 328 | 1 011 926 |
| thymus 2 | 52 | M | thymus | 1 747 487 | 1 198 677 | 3 089 405 ± 5 992 | 2 989 346 ± 4 808 | 7 000 872 ± 416 869 | 1 017 426 | 771 547 |
| thymus 3 | 107 | M | thymus | 2 158 043 | 1 230 436 | 2 468 027 ± 3 953 | 2 410 929 ± 3 240 | 9 121 480 ± 635 237 | 985 201 | 708 503 |
| thymus 4 | 156 | F | thymus | 1 848 851 | 1 345 927 | 3 812 418 ± 7 380 | 3 606 280 ± 5 422 | 6 826 960 ± 368 256 | 1 166 396 | 885 107 |
| blood 1 | 7 | M | blood | 154 682 | 130 307 | 540 969 ± 4 198 | 540 969 ± 4 123 | 827 694 ± 56 037 | 121 367 | 108 361 |
| blood 2 | 52 | M | blood | 123 523 | 103 142 | 430 223 ± 3 766 | 430 223 ± 3 700 | 681 737 ± 35 608 | 94 556 | 76 513 |
| blood 3 | 107 | M | blood | 245 126 | 170 333 | 522 505 ± 3 003 | 522 505 ± 2 923 | 885 943 ± 47 301 | 114 452 | 5 901 |
| blood 4 | 156 | F | blood | 199 326 | 157 728 | 56 686 ± 3 696 | 566 868 ± 3 615 | 1 270 020 ± 90 170 | 137 613 | 62 127 |

TCRB

| Sample id | Age (days) | Sex | Tissue | Observed counts | Observed diversity | Chao1 (mean±std) | Extrapolated ChaoE (mean±std) | Efron-Thisted (mean±std) | Shannon's index (mean) | Inversed Simpson's index (mean) |
|---|---|---|---|---|---|---|---|---|---|---|
| thymus A | 243 | M | thymus | 1 647 656 | 1 245 029 | 3 463 693 ± 6 808 | 1,360,137 ± 899 | 5 568 945 ± 281 571 | 1 115 157 | 949 669 |
| thymus B | 244 | M | thymus | 1 783 878 | 1 526 694 | 7 198 338 ± 17 565 | 1,576,012 ± 1 097 | 11 613 883 ± 828 672 | 1 426 418 | 1 275 231 |
| thymus C | 225 | F | thymus | 1 850 299 | 1 551 603 | 6 175 838 ± 13 550 | 1 551 603 ± 1 077 | 11 065 398 ± 661 911 | 1 446 717 | 1 302 914 |
| thymus D | 126 | M | thymus | 1 726 796 | 1 449 881 | 6 208 706 ± 14 737 | 1 537 939 ± 1 056 | 9 910 366 ± 569 512 | 1 343 436 | 1 185 806 |
| thymus 1 | 7 | M | thymus | 237 063 | 222 925 | 2 056 755 ± 18 593 | 1 215 349 ± 4 192 | 2 736 169 ± 206 873 | 217 609 | 209 474 |
| thymus 2 | 52 | M | thymus | 182 356 | 172 746 | 1 737 599 ± 18 637 | 1 136 517 ± 5 332 | 2 180 279 ± 113 043 | 169 209 | 163 907 |
| thymus 3 | 107 | M | thymus | 142 903 | 137 920 | 2 039 501 ± 29 774 | 1 215 583 ± 6 936 | 1 782 350 ± 140 139 | 136 060 | 133 205 |
| thymus 4 | 156 | F | thymus | 128 228 | 121 483 | 1 269 967 ± 16 587 | 971 553 ± 6 988 | 1 569 435 ± 90 703 | 118 934 | 115 045 |
| blood 1 | 7 | M | blood | 82 418 | 77 281 | 713 105 ± 10 949 | 658 522 ± 7 882 | 865 983 ± 57 876 | 75 069 | 70 167 |
| blood 2 | 52 | M | blood | 73 945 | 69 404 | 704 739 ± 12 003 | 651 718 ± 8 696 | 796 697 ± 63 815 | 67 046 | 59 899 |
| blood 3 | 107 | M | blood | 134 110 | 103 551 | 1 029 542 ± 14 292 | 787 853 ± 6 017 | 1 089 644 ± 87 860 | 47 612 | 1 278 |
| blood 4 | 156 | F | blood | 88 901 | 81 935 | 790 042 ± 12 109 | 707 833 ± 7 921 | 898 294 ± 62 024 | 77 810 | 64 418 |

**Table 3**
Resampled diversity estimates.

TCRAD

| Sample id | Age (days) | Sex | Tissue | Observed counts | Observed diversity | Resampled Chao1 (mean±std) | Resampled Efron-Thisted (mean±std) | Resampled Shannon's index (mean) | Resampled inversed Simpson's index (mean) |
|---|---|---|---|---|---|---|---|---|---|
| thymus A | 243 | M | thymus | 11 838 086 | 6 763 870 | 3 439 802 ± 48 656 | 1 928 183 ± 109 031 | 119 114 | 116 022 |
| thymus B | 244 | M | thymus | 12 849 473 | 7 419 245 | 3 717 184 ± 117 571 | 1 934 212 ± 165 908 | 119 461 | 116 728 |
| thymus C | 225 | F | thymus | 8 359 283 | 5 259 057 | 3 519 847 ± 71 955 | 2 023 854 ± 156 215 | 119 278 | 116 379 |
| thymus D | 126 | M | thymus | 11 063 464 | 6 610 182 | 3 799 437 ± 74 891 | 2 113 509 ± 195 602 | 119 460 | 116 538 |
| thymus 1 | 7 | M | thymus | 3 179 774 | 1 984 292 | 2 131 831 ± 30 009 | 1 796 761 ± 112 417 | 117 634 | 114 042 |
| thymus 2 | 52 | M | thymus | 1 747 487 | 1 198 677 | 1 703 786 ± 14 235 | 1 686 940 ± 56 182 | 116 726 | 113 381 |
| thymus 3 | 107 | M | thymus | 2 158 043 | 1 230 436 | 1 302 763 ± 10 861 | 1 458 706 ± 13 181 | 114 778 | 110 665 |
| thymus 4 | 156 | F | thymus | 1 848 851 | 1 345 927 | 2 227 657 ± 27 749 | 1 956 797 ± 46 144 | 118 117 | 115 234 |
| blood 1 | 7 | M | blood | 154 682 | 130 307 | 515 756 ± 2 407 | 850 317 ± 49 399 | 101 160 | 92 134 |
| blood 2 | 52 | M | blood | 123 523 | 103 142 | 430 223 ± 0 | 681 737 ± 0 | 94 556 | 76 513 |
| blood 3 | 107 | M | blood | 245 126 | 170 333 | 421 347 ± 1 715 | 760 367 ± 58 240 | 71 403 | 5 814 |
| blood 4 | 156 | F | blood | 199 326 | 157 728 | 506 377 ± 1 192 | 824 158 ± 38 655 | 95 097 | 51 910 |

TCRB

| Sample id | Age (days) | Sex | Tissue | Observed counts | Observed diversity | Resampled Chao1 (mean±std) | Resampled Efron-Thisted (mean±std) | Resampled Shannon's index (mean) | Resampled inversed Simpson's index (mean) |
|---|---|---|---|---|---|---|---|---|---|
| thymus A | 243 | M | thymus | 1 647 656 | 1 245 029 | 2 402 723 ± 69 516 | 1 320 715 ± 8 691 | 72 359 | 71 675 |
| thymus B | 244 | M | thymus | 1 783 878 | 1 526 694 | 4 788 673 ± 151 470 | 1 437 422 ± 2 372 | 73 122 | 72 751 |
| thymus C | 225 | F | thymus | 1 850 299 | 1 551 603 | 4 582 751 ± 147 345 | 1 423 151 ± 2 423 | 73 119 | 72 760 |
| thymus D | 126 | M | thymus | 1 726 796 | 1 449 881 | 4 058 264 ± 145 518 | 1 413 346 ± 5 795 | 72 976 | 72 540 |
| thymus 1 | 7 | M | thymus | 237 063 | 222 925 | 1 912 950 ± 45 644 | 1 275 516 ± 15 408 | 71 922 | 71 039 |
| thymus 2 | 52 | M | thymus | 182 356 | 172 746 | 1 663 403 ± 52 735 | 1 238 325 ± 16 294 | 71 651 | 70 673 |
| thymus 3 | 107 | M | thymus | 142 903 | 137 920 | 1 974 903 ± 20 177 | 1 263 684 ± 15 635 | 72 045 | 71 233 |
| thymus 4 | 156 | F | thymus | 128 228 | 121 483 | 1 206 007 ± 23 326 | 1 135 274 ± 12 074 | 70 702 | 69 306 |
| blood 1 | 7 | M | blood | 82 418 | 77 281 | 709 395 ± 4 431 | 816 448 ± 47 909 | 67 932 | 63 859 |
| blood 2 | 52 | M | blood | 73 945 | 69 404 | 704 739 ± 0 | 796 697 ± 0 | 67 046 | 59 899 |
| blood 3 | 107 | M | blood | 134 110 | 103 551 | 895 010 ± 8 921 | 798 491 ± 78 565 | 30 118 | 1 268 |
| blood 4 | 156 | F | blood | 88 901 | 81 935 | 756 652 ± 5 086 | 741 337 ± 57 954 | 65 811 | 56 040 |

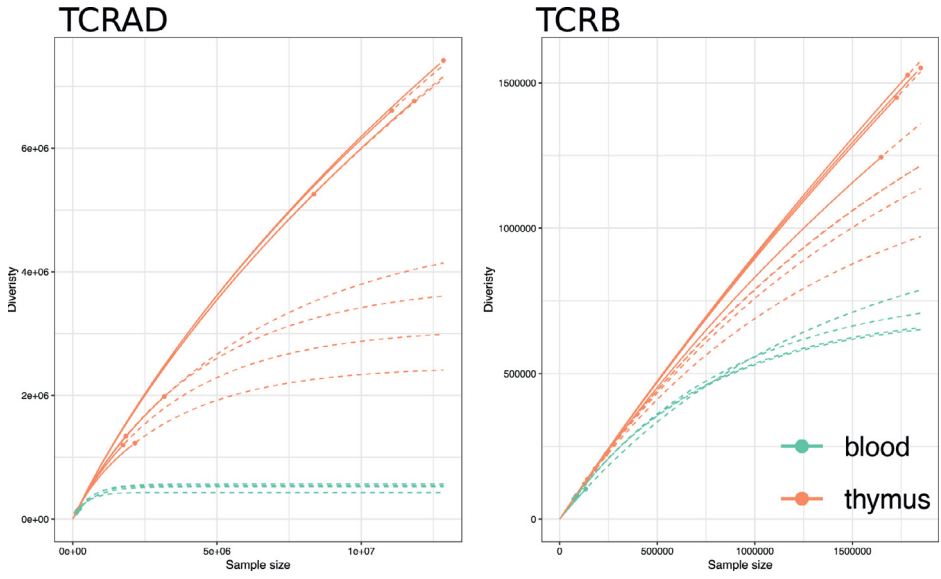


**Fig. 3.** The rarefaction curves of TCRAD and TCRB diversities for each sample. The thymic samples are marked in red and peripheral samples in blue. The dots indicate observed diversity and counts, solid lines interpolated and dashed lines extrapolated values.

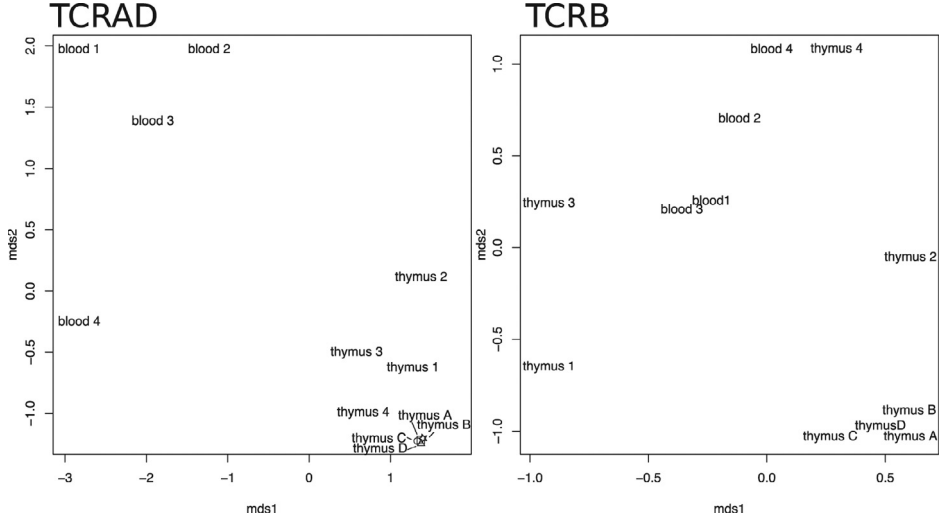


**Fig. 4.** Clustering of the samples on multi-dimensional scaling according to pairwise repertoire overlap comparison with Jaccard index.

"CalcDiversityStats" and "RarefactionPlot" commands with default settings to calculate and visualise diversity estimations, and finally "CalcPairwiseDistances" command to calculate the sequence overlap between two samples. For sequence overlap we selected the setting "strict",

**Table 4**
Overlap measures.

TCRAD

| Sample 1 id | Sample 2 id | Diversity sample 1 | Diversity sample 2 | Number of overlapping clonotypes | Counts sample 1 | Counts sample 2 | Counts of overlapping clonotypes in sample 1 | Counts of overlapping clonotypes in sample 2 | Pearson correlation | Relative overlap measure | Jaccard index | Morisita-Horn index |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| thymus A | thymus B | 6 763 870 | 7 419 245 | 866 679 | 11 838 086 | 12 849 473 | 3 401 700 | 3 572 924 | 0,881 | 1,73E-08 | 0,0651 | 0,904 |
| thymus A | thymus C | 6 763 870 | 5 259 057 | 630 350 | 11 838 086 | 8 359 283 | 2 823 352 | 2 076 828 | 0,844 | 1,77E-08 | 0,0553 | 0,876 |
| thymus A | thymus D | 6 763 870 | 6 610 182 | 773 916 | 11 838 086 | 11 063 464 | 3 177 831 | 2 897 003 | 0,853 | 1,73E-08 | 0,0614 | 0,882 |
| thymus A | thymus 1 | 6 763 870 | 1 984 292 | 287 646 | 11 838 086 | 3 179 774 | 1 679 632 | 658 811 | 0,655 | 2,14E-08 | 0,0340 | 0,727 |
| thymus A | thymus 2 | 6 763 870 | 1 198 677 | 177 662 | 11 838 086 | 1 747 487 | 1 213 379 | 332 560 | 0,600 | 2,19E-08 | 0,0228 | 0,673 |
| thymus A | thymus 3 | 6 763 870 | 1 230 436 | 189 717 | 11 838 086 | 2 158 043 | 1 277 978 | 442 972 | 0,604 | 2,28E-08 | 0,0243 | 0,675 |
| thymus A | thymus 4 | 6 763 870 | 1 345 927 | 203 596 | 11 838 086 | 1 848 851 | 1 353 803 | 361 818 | 0,633 | 2,24E-08 | 0,0258 | 0,702 |
| thymus A | blood 1 | 6 763 870 | 130 307 | 29 044 | 11 838 086 | 154 682 | 316 005 | 37 176 | 0,251 | 3,30E-08 | 0,0042 | 0,218 |
| thymus A | blood 2 | 6 763 870 | 103 142 | 20 600 | 11 838 086 | 123 523 | 223 783 | 26 054 | 0,148 | 2,95E-08 | 0,0030 | 0,151 |
| thymus A | blood 3 | 6 763 870 | 170 333 | 34 574 | 11 838 086 | 245 126 | 347 204 | 57 652 | 0,014 | 3,00E-08 | 0,0050 | 0,004 |
| thymus A | blood 4 | 6 763 870 | 157 728 | 28 303 | 11 838 086 | 199 326 | 300 479 | 38 946 | 0,042 | 2,65E-08 | 0,0041 | 0,048 |
| thymus B | thymus C | 7 419 245 | 5 259 057 | 657 775 | 12 849 473 | 8 359 283 | 3 031 445 | 2 129 578 | 0,837 | 1,69E-08 | 0,0547 | 0,870 |
| thymus B | thymus D | 7 419 245 | 6 610 182 | 808 433 | 12 849 473 | 11 063 464 | 3 416 024 | 2 969 838 | 0,851 | 1,65E-08 | 0,0611 | 0,882 |
| thymus B | thymus 1 | 7 419 245 | 1 984 292 | 297 894 | 12 849 473 | 3 179 774 | 1 791 741 | 676 485 | 0,646 | 2,02E-08 | 0,0327 | 0,719 |
| thymus B | thymus 2 | 7 419 245 | 1 198 677 | 184 222 | 12 849 473 | 1 747 487 | 1 292 592 | 342 539 | 0,586 | 2,07E-08 | 0,0218 | 0,663 |
| thymus B | thymus 3 | 7 419 245 | 1 230 436 | 196 494 | 12 849 473 | 2 158 043 | 1 360 763 | 455 236 | 0,588 | 2,15E-08 | 0,0232 | 0,661 |
| thymus B | thymus 4 | 7 419 245 | 1 345 927 | 210 731 | 12 849 473 | 1 848 851 | 1 438 108 | 371 861 | 0,619 | 2,11E-08 | 0,0246 | 0,693 |
| thymus B | blood 1 | 7 419 245 | 130 307 | 29 717 | 12 849 473 | 154 682 | 330 377 | 37 934 | 0,246 | 3,07E-08 | 0,0040 | 0,206 |
| thymus B | blood 2 | 7 419 245 | 103 142 | 21 239 | 12 849 473 | 123 523 | 236 641 | 26 746 | 0,151 | 2,78E-08 | 0,0028 | 0,145 |
| thymus B | blood 3 | 7 419 245 | 170 333 | 35 705 | 12 849 473 | 245 126 | 366 684 | 57 206 | 0,022 | 2,83E-08 | 0,0047 | 0,008 |
| thymus B | blood 4 | 7 419 245 | 157 728 | 29 322 | 12 849 473 | 199 326 | 316 077 | 40 215 | 0,041 | 2,51E-08 | 0,0039 | 0,046 |
| thymus C | thymus D | 5 259 057 | 6 610 182 | 596 367 | 8 359 283 | 11 063 464 | 2 003 982 | 2 481 328 | 0,827 | 1,72E-08 | 0,0529 | 0,863 |
| thymus C | thymus 1 | 5 259 057 | 1 984 292 | 231 423 | 8 359 283 | 3 179 774 | 1 073 774 | 557 955 | 0,639 | 2,22E-08 | 0,0330 | 0,722 |
| thymus C | thymus 2 | 5 259 057 | 1 198 677 | 147 145 | 8 359 283 | 1 747 487 | 794 176 | 283 569 | 0,598 | 2,33E-08 | 0,0233 | 0,686 |
| thymus C | thymus 3 | 5 259 057 | 1 230 436 | 160 296 | 8 359 283 | 2 158 043 | 853 824 | 388 406 | 0,614 | 2,48E-08 | 0,0253 | 0,695 |
| thymus C | thymus 4 | 5 259 057 | 1 345 927 | 169 687 | 8 359 283 | 1 848 851 | 897 840 | 311 017 | 0,640 | 2,40E-08 | 0,0264 | 0,719 |
| thymus C | blood 1 | 5 259 057 | 130 307 | 24 337 | 8 359 283 | 154 682 | 206 500 | 31 520 | 0,238 | 3,55E-08 | 0,0045 | 0,229 |
| thymus C | blood 2 | 5 259 057 | 103 142 | 17 448 | 8 359 283 | 123 523 | 150 315 | 22 268 | 0,132 | 3,22E-08 | 0,0033 | 0,154 |
| thymus C | blood 3 | 5 259 057 | 170 333 | 29 699 | 8 359 283 | 245 126 | 233 789 | 46 909 | 0,072 | 3,32E-08 | 0,0055 | 0,070 |
| thymus C | blood 4 | 5 259 057 | 157 728 | 24 276 | 8 359 283 | 199 326 | 204 214 | 33 909 | 0,047 | 2,93E-08 | 0,0045 | 0,049 |
| thymus D | thymus 1 | 6 610 182 | 1 984 292 | 279 803 | 11 063 464 | 3 179 774 | 1 506 996 | 648 316 | 0,684 | 2,13E-08 | 0,0337 | 0,748 |
| thymus D | thymus 2 | 6 610 182 | 1 198 677 | 174 157 | 11 063 464 | 1 747 487 | 1 089 914 | 327 307 | 0,621 | 2,20E-08 | 0,0228 | 0,689 |
| thymus D | thymus 3 | 6 610 182 | 1 230 436 | 182 717 | 11 063 464 | 2 158 043 | 1 128 178 | 431 118 | 0,623 | 2,25E-08 | 0,0239 | 0,686 |
| thymus D | thymus 4 | 6 610 182 | 1 345 927 | 196 751 | 11 063 464 | 1 848 851 | 1 200 784 | 351 610 | 0,655 | 2,21E-08 | 0,0254 | 0,719 |
| thymus D | blood 1 | 6 610 182 | 130 307 | 28 731 | 11 063 464 | 154 682 | 287 455 | 36 894 | 0,268 | 3,34E-08 | 0,0043 | 0,217 |
| thymus D | blood 2 | 6 610 182 | 103 142 | 20 581 | 11 063 464 | 123 523 | 211 785 | 26 154 | 0,169 | 3,02E-08 | 0,0031 | 0,156 |
| thymus D | blood 3 | 6 610 182 | 170 333 | 33 661 | 11 063 464 | 245 126 | 313 960 | 54 297 | 0,023 | 2,99E-08 | 0,0050 | 0,007 |
| thymus D | blood 4 | 6 610 182 | 157 728 | 27 545 | 11 063 464 | 199 326 | 269 325 | 38 363 | 0,067 | 2,64E-08 | 0,0041 | 0,052 |
| thymus 1 | thymus 2 | 1 984 292 | 1 198 677 | 84 060 | 3 179 774 | 1 747 487 | 252 989 | 176 256 | 0,645 | 3,53E-08 | 0,0271 | 0,749 |
| thymus 1 | thymus 3 | 1 984 292 | 1 230 436 | 86 181 | 3 179 774 | 2 158 043 | 256 879 | 230 215 | 0,652 | 3,53E-08 | 0,0275 | 0,752 |
| thymus 1 | thymus 4 | 1 984 292 | 1 345 927 | 91 687 | 3 179 774 | 1 848 851 | 270 795 | 184 905 | 0,657 | 3,43E-08 | 0,0283 | 0,758 |
| thymus 1 | blood 1 | 1 984 292 | 130 307 | 16 192 | 3 179 774 | 154 682 | 70 223 | 21 396 | 0,341 | 6,26E-08 | 0,0077 | 0,310 |
| thymus 1 | blood 2 | 1 984 292 | 103 142 | 11 184 | 3 179 774 | 123 523 | 50 623 | 14 471 | 0,245 | 5,46E-08 | 0,0054 | 0,248 |
| thymus 1 | blood 3 | 1 984 292 | 170 333 | 17 726 | 3 179 774 | 245 126 | 72 795 | 30 442 | 0,026 | 5,24E-08 | 0,0083 | 0,008 |
| thymus 1 | blood 4 | 1 984 292 | 157 728 | 14 345 | 3 179 774 | 199 326 | 60 045 | 19 682 | 0,137 | 4,58E-08 | 0,0067 | 0,215 |
| thymus 2 | thymus 3 | 1 198 677 | 1 230 436 | 57 752 | 1 747 487 | 2 158 043 | 128 317 | 166 052 | 0,625 | 3,92E-08 | 0,0244 | 0,762 |
| thymus 2 | thymus 4 | 1 198 677 | 1 345 927 | 62 995 | 1 747 487 | 1 848 851 | 138 699 | 136 036 | 0,641 | 3,90E-08 | 0,0254 | 0,779 |
| thymus 2 | blood 1 | 1 198 677 | 130 307 | 10 777 | 1 747 487 | 154 682 | 31 281 | 14 595 | 0,317 | 6,90E-08 | 0,0082 | 0,333 |
| thymus 2 | blood 2 | 1 198 677 | 103 142 | 8 069 | 1 747 487 | 123 523 | 23 923 | 10 565 | 0,229 | 6,53E-08 | 0,0062 | 0,274 |
| thymus 2 | blood 3 | 1 198 677 | 170 333 | 11 853 | 1 747 487 | 245 126 | 32 976 | 18 846 | 0,253 | 5,81E-08 | 0,0087 | 0,310 |
| thymus 2 | blood 4 | 1 198 677 | 157 728 | 10 266 | 1 747 487 | 199 326 | 29 087 | 14 063 | 0,260 | 5,43E-08 | 0,0076 | 0,374 |
| thymus 3 | thymus 4 | 1 230 436 | 1 345 927 | 66 420 | 2 158 043 | 1 848 851 | 186 971 | 141 677 | 0,630 | 4,01E-08 | 0,0265 | 0,761 |
| thymus 3 | blood 1 | 1 230 436 | 130 307 | 11 134 | 2 158 043 | 154 682 | 42 595 | 15 018 | 0,320 | 6,94E-08 | 0,0082 | 0,356 |
| thymus 3 | blood 2 | 1 230 436 | 103 142 | 7 922 | 2 158 043 | 123 523 | 31 210 | 10 341 | 0,217 | 6,24E-08 | 0,0060 | 0,289 |
| thymus 3 | blood 3 | 1 230 436 | 170 333 | 13 436 | 2 158 043 | 245 126 | 48 717 | 27 106 | 0,013 | 6,41E-08 | 0,0097 | 0,004 |
| thymus 3 | blood 4 | 1 230 436 | 157 728 | 10 754 | 2 158 043 | 199 326 | 39 869 | 14 765 | 0,168 | 5,54E-08 | 0,0078 | 0,309 |
| thymus 4 | blood 1 | 1 345 927 | 130 307 | 11 729 | 1 848 851 | 154 682 | 32 338 | 15 669 | 0,296 | 6,69E-08 | 0,0080 | 0,308 |
| thymus 4 | blood 2 | 1 345 927 | 103 142 | 8 364 | 1 848 851 | 123 523 | 23 731 | 10 893 | 0,182 | 6,02E-08 | 0,0058 | 0,233 |
| thymus 4 | blood 3 | 1 345 927 | 170 333 | 13 389 | 1 848 851 | 245 126 | 36 254 | 21 997 | 0,075 | 5,84E-08 | 0,0089 | 0,064 |
| thymus 4 | blood 4 | 1 345 927 | 157 728 | 11 776 | 1 848 851 | 199 326 | 32 772 | 16 401 | 0,061 | 5,55E-08 | 0,0079 | 0,081 |
| blood 1 | blood 2 | 130 307 | 103 142 | 2 210 | 154 682 | 123 523 | 3 440 | 3 269 | 0,329 | 1,64E-07 | 0,0096 | 0,689 |
| blood 1 | blood 3 | 130 307 | 170 333 | 3 425 | 154 682 | 245 126 | 5 091 | 6 453 | 0,355 | 1,54E-07 | 0,0115 | 0,663 |
| blood 1 | blood 4 | 130 307 | 157 728 | 2 625 | 154 682 | 199 326 | 3 937 | 4 343 | 0,078 | 1,28E-07 | 0,0092 | 0,217 |
| blood 2 | blood 3 | 103 142 | 170 333 | 2 564 | 123 523 | 245 126 | 3 985 | 5 130 | 0,268 | 1,46E-07 | 0,0095 | 0,478 |
| blood 2 | blood 4 | 103 142 | 157 728 | 2 037 | 123 523 | 199 326 | 3 335 | 4 881 | 0,614 | 1,25E-07 | 0,0079 | 0,408 |
| blood 3 | blood 4 | 170 333 | 157 728 | 3 028 | 245 126 | 199 326 | 6 058 | 6 352 | 0,141 | 1,13E-07 | 0,0093 | 0,132 |

(*continued on next page*)

which requires matching CDR3 nucleotide regions as well as matching V genes and J genes. For visualisation of Jaccard index overlap values we used “ClusterSamples” tool that provides a multi-dimensional scaling plot created with isoMDS() function of MASS package for R.

## Ethics Statement

The study was approved by the Pediatric Ethical Committee of the Helsinki University Hospital (HUS/747/2019) and a written informed consent was obtained from the parents.

**Table 4** (*continued*)

TCRB

| Sample 1 id | Sample 2 id | Diversity sample 1 | Diversity sample 2 | Number of overlapping clonotypes | Counts sample 1 | Counts sample 2 | Counts of overlapping clonotypes in sample 1 | Counts of overlapping clonotypes in sample 2 | Pearson correlation | Relative overlap measure | Jaccard index | Morisita-Horn index |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| thymus A | thymus B | 1 245 029 | 1 526 694 | 3 244 | 1 647 656 | 1 783 878 | 5 295 | 5 045 | 0,326 | 1,70E-09 | 0,00117 | 0,715 |
| thymus A | thymus C | 1 245 029 | 1 551 603 | 2 545 | 1 647 656 | 1 850 299 | 4 108 | 3 863 | 0,266 | 1,30E-09 | 0,00091 | 0,737 |
| thymus A | thymus D | 1 245 029 | 1 449 881 | 2 939 | 1 647 656 | 1 726 796 | 4 752 | 4 385 | 0,277 | 1,60E-09 | 0,00109 | 0,741 |
| thymus A | thymus 1 | 1 245 029 | 222 925 | 529 | 1 647 656 | 237 063 | 965 | 627 | 0,240 | 1,90E-09 | 0,00036 | 0,370 |
| thymus A | thymus 2 | 1 245 029 | 172 746 | 301 | 1 647 656 | 182 356 | 509 | 345 | 0,204 | 1,40E-09 | 0,00021 | 0,293 |
| thymus A | thymus 3 | 1 245 029 | 137 920 | 252 | 1 647 656 | 142 903 | 431 | 278 | 0,026 | 1,50E-09 | 0,00018 | 0,230 |
| thymus A | thymus 4 | 1 245 029 | 121 483 | 221 | 1 647 656 | 128 228 | 373 | 249 | 0,063 | 1,50E-09 | 0,00016 | 0,208 |
| thymus A | blood 1 | 1 245 029 | 77 281 | 555 | 1 647 656 | 82 418 | 929 | 694 | 0,246 | 5,80E-09 | 0,00042 | 0,110 |
| thymus A | blood 2 | 1 245 029 | 69 404 | 361 | 1 647 656 | 73 945 | 632 | 433 | -0,013 | 4,20E-09 | 0,00027 | 0,096 |
| thymus A | blood 3 | 1 245 029 | 103 551 | 550 | 1 647 656 | 134 110 | 1 023 | 651 | 0,109 | 4,30E-09 | 0,00041 | 0,216 |
| thymus A | blood 4 | 1 245 029 | 81 935 | 299 | 1 647 656 | 88 901 | 545 | 346 | 0,015 | 2,90E-09 | 0,00023 | 0,138 |
| thymus B | thymus C | 1 526 694 | 1 551 603 | 2 919 | 1 783 878 | 1 850 299 | 4 371 | 4 373 | 0,280 | 1,20E-09 | 0,00095 | 0,740 |
| thymus B | thymus D | 1 526 694 | 1 449 881 | 3 357 | 1 783 878 | 1 726 796 | 5 097 | 4 974 | 0,345 | 1,50E-09 | 0,00113 | 0,774 |
| thymus B | thymus 1 | 1 526 694 | 222 925 | 597 | 1 783 878 | 237 063 | 997 | 700 | 0,339 | 1,80E-09 | 0,00034 | 0,335 |
| thymus B | thymus 2 | 1 526 694 | 172 746 | 340 | 1 783 878 | 182 356 | 562 | 375 | 0,134 | 1,30E-09 | 0,00020 | 0,282 |
| thymus B | thymus 3 | 1 526 694 | 137 920 | 319 | 1 783 878 | 142 903 | 541 | 354 | 0,042 | 1,50E-09 | 0,00019 | 0,213 |
| thymus B | thymus 4 | 1 526 694 | 121 483 | 290 | 1 783 878 | 128 228 | 497 | 325 | 0,085 | 1,60E-09 | 0,00018 | 0,198 |
| thymus B | blood 1 | 1 526 694 | 77 281 | 615 | 1 783 878 | 82 418 | 1 035 | 762 | 0,181 | 5,20E-09 | 0,00038 | 0,100 |
| thymus B | blood 2 | 1 526 694 | 69 404 | 401 | 1 783 878 | 73 945 | 679 | 458 | 0,119 | 3,80E-09 | 0,00025 | 0,112 |
| thymus B | blood 3 | 1 526 694 | 103 551 | 623 | 1 783 878 | 134 110 | 1 056 | 769 | 0,118 | 3,90E-09 | 0,00038 | 0,163 |
| thymus B | blood 4 | 1 526 694 | 81 935 | 349 | 1 783 878 | 88 901 | 578 | 401 | 0,052 | 2,80E-09 | 0,00022 | 0,124 |
| thymus C | thymus D | 1 551 603 | 1 449 881 | 3 300 | 1 850 299 | 1 726 796 | 5 015 | 4 900 | 0,244 | 1,50E-09 | 0,00110 | 0,750 |
| thymus C | thymus 1 | 1 551 603 | 222 925 | 671 | 1 850 299 | 237 063 | 1 110 | 803 | 0,251 | 1,90E-09 | 0,00038 | 0,307 |
| thymus C | thymus 2 | 1 551 603 | 172 746 | 387 | 1 850 299 | 182 356 | 674 | 452 | 0,204 | 1,40E-09 | 0,00022 | 0,266 |
| thymus C | thymus 3 | 1 551 603 | 137 920 | 280 | 1 850 299 | 142 903 | 471 | 314 | 0,061 | 1,30E-09 | 0,00017 | 0,199 |
| thymus C | thymus 4 | 1 551 603 | 121 483 | 269 | 1 850 299 | 128 228 | 439 | 307 | 0,060 | 1,40E-09 | 0,00016 | 0,174 |
| thymus C | blood 1 | 1 551 603 | 77 281 | 752 | 1 850 299 | 82 418 | 1 283 | 983 | 0,253 | 6,30E-09 | 0,00046 | 0,094 |
| thymus C | blood 2 | 1 551 603 | 69 404 | 468 | 1 850 299 | 73 945 | 779 | 559 | 0,057 | 4,30E-09 | 0,00029 | 0,096 |
| thymus C | blood 3 | 1 551 603 | 103 551 | 702 | 1 850 299 | 134 110 | 1 252 | 887 | 0,192 | 4,40E-09 | 0,00042 | 0,160 |
| thymus C | blood 4 | 1 551 603 | 81 935 | 330 | 1 850 299 | 88 901 | 585 | 387 | 0,090 | 2,60E-09 | 0,00020 | 0,130 |
| thymus D | thymus 1 | 1 449 881 | 222 925 | 656 | 1 726 796 | 237 063 | 1 048 | 786 | 0,279 | 2,00E-09 | 0,00039 | 0,315 |
| thymus D | thymus 2 | 1 449 881 | 172 746 | 386 | 1 726 796 | 182 356 | 632 | 445 | 0,120 | 1,50E-09 | 0,00024 | 0,269 |
| thymus D | thymus 3 | 1 449 881 | 137 920 | 316 | 1 726 796 | 142 903 | 524 | 360 | 0,015 | 1,60E-09 | 0,00020 | 0,206 |
| thymus D | thymus 4 | 1 449 881 | 121 483 | 276 | 1 726 796 | 128 228 | 412 | 311 | 0,036 | 1,60E-09 | 0,00018 | 0,176 |
| thymus D | blood 1 | 1 449 881 | 77 281 | 711 | 1 726 796 | 82 418 | 1 173 | 935 | 0,072 | 6,30E-09 | 0,00047 | 0,088 |
| thymus D | blood 2 | 1 449 881 | 69 404 | 412 | 1 726 796 | 73 945 | 692 | 466 | 0,126 | 4,10E-09 | 0,00027 | 0,117 |
| thymus D | blood 3 | 1 449 881 | 103 551 | 718 | 1 726 796 | 134 110 | 1 199 | 897 | 0,120 | 4,80E-09 | 0,00046 | 0,152 |
| thymus D | blood 4 | 1 449 881 | 81 935 | 386 | 1 726 796 | 88 901 | 600 | 472 | 0,055 | 3,20E-09 | 0,00025 | 0,094 |
| thymus 1 | thymus 2 | 222 925 | 172 746 | 68 | 237 063 | 182 356 | 92 | 82 | 0,165 | 1,80E-09 | 0,00017 | 0,748 |
| thymus 1 | thymus 3 | 222 925 | 137 920 | 75 | 237 063 | 142 903 | 98 | 86 | 0,059 | 2,40E-09 | 0,00021 | 0,749 |
| thymus 1 | thymus 4 | 222 925 | 121 483 | 47 | 237 063 | 128 228 | 55 | 55 | -0,048 | 1,70E-09 | 0,00014 | 0,750 |
| thymus 1 | blood 1 | 222 925 | 77 281 | 301 | 237 063 | 82 418 | 371 | 464 | 0,097 | 1,75E-08 | 0,00100 | 0,292 |
| thymus 1 | blood 2 | 222 925 | 69 404 | 99 | 237 063 | 73 945 | 118 | 122 | 0,153 | 6,40E-09 | 0,00034 | 0,463 |
| thymus 1 | blood 3 | 222 925 | 103 551 | 192 | 237 063 | 134 110 | 243 | 236 | 0,284 | 8,30E-09 | 0,00059 | 0,741 |
| thymus 1 | blood 4 | 222 925 | 81 935 | 63 | 237 063 | 88 901 | 85 | 78 | 0,032 | 3,40E-09 | 0,00021 | 0,488 |
| thymus 2 | thymus 3 | 172 746 | 137 920 | 32 | 182 356 | 142 903 | 40 | 40 | 0,000 | 1,30E-09 | 0,00010 | 0,848 |
| thymus 2 | thymus 4 | 172 746 | 121 483 | 35 | 182 356 | 128 228 | 40 | 42 | 0,000 | 1,70E-09 | 0,00012 | 0,836 |
| thymus 2 | blood 1 | 172 746 | 77 281 | 92 | 182 356 | 82 418 | 104 | 112 | 0,074 | 6,90E-09 | 0,00037 | 0,590 |
| thymus 2 | blood 2 | 172 746 | 69 404 | 77 | 182 356 | 73 945 | 88 | 102 | 0,051 | 6,40E-09 | 0,00032 | 0,405 |
| thymus 2 | blood 3 | 172 746 | 103 551 | 99 | 182 356 | 134 110 | 114 | 3 222 | -0,049 | 5,50E-09 | 0,00036 | 0,001 |
| thymus 2 | blood 4 | 172 746 | 81 935 | 42 | 182 356 | 88 901 | 49 | 48 | 0,183 | 3,00E-09 | 0,00016 | 0,741 |
| thymus 3 | thymus 4 | 137 920 | 121 483 | 29 | 142 903 | 128 228 | 33 | 30 | -0,060 | 1,70E-09 | 0,00011 | 0,916 |
| thymus 3 | blood 1 | 137 920 | 77 281 | 76 | 142 903 | 82 418 | 89 | 104 | 0,214 | 7,10E-09 | 0,00035 | 0,591 |
| thymus 3 | blood 2 | 137 920 | 69 404 | 46 | 142 903 | 73 945 | 49 | 49 | -0,041 | 4,80E-09 | 0,00022 | 0,745 |
| thymus 3 | blood 3 | 137 920 | 103 551 | 126 | 142 903 | 134 110 | 153 | 5 923 | -0,042 | 8,80E-09 | 0,00052 | 0,001 |
| thymus 3 | blood 4 | 137 920 | 81 935 | 40 | 142 903 | 88 901 | 47 | 43 | -0,062 | 3,50E-09 | 0,00018 | 0,774 |
| thymus 4 | blood 1 | 121 483 | 77 281 | 44 | 128 228 | 82 418 | 50 | 53 | 0,065 | 4,70E-09 | 0,00022 | 0,672 |
| thymus 4 | blood 2 | 121 483 | 69 404 | 46 | 128 228 | 73 945 | 49 | 64 | -0,009 | 5,50E-09 | 0,00024 | 0,341 |
| thymus 4 | blood 3 | 121 483 | 103 551 | 53 | 128 228 | 134 110 | 66 | 71 | 0,181 | 4,20E-09 | 0,00024 | 0,772 |
| thymus 4 | blood 4 | 121 483 | 81 935 | 81 | 128 228 | 88 901 | 92 | 137 | 0,112 | 8,10E-09 | 0,00040 | 0,369 |
| blood 1 | blood 2 | 77 281 | 69 404 | 195 | 82 418 | 73 945 | 272 | 240 | 0,045 | 3,64E-08 | 0,00133 | 0,736 |
| blood 1 | blood 3 | 77 281 | 103 551 | 328 | 82 418 | 134 110 | 496 | 448 | 0,303 | 4,10E-08 | 0,00182 | 0,656 |
| blood 1 | blood 4 | 77 281 | 81 935 | 126 | 82 418 | 88 901 | 159 | 156 | 0,219 | 1,99E-08 | 0,00079 | 0,825 |
| blood 2 | blood 3 | 69 404 | 103 551 | 173 | 73 945 | 134 110 | 219 | 218 | 0,108 | 2,41E-08 | 0,00100 | 0,665 |
| blood 2 | blood 4 | 69 404 | 81 935 | 87 | 73 945 | 88 901 | 114 | 111 | 0,032 | 1,53E-08 | 0,00058 | 0,737 |
| blood 3 | blood 4 | 103 551 | 81 935 | 130 | 134 110 | 88 901 | 175 | 158 | 0,102 | 1,53E-08 | 0,00070 | 0,704 |

## CRediT Author Statement

NH and TPA conceptualised the study and wrote the original manuscript. NH and RV collected and prepared the samples. IK, DAY and JS implemented the software usage. IPM provided the study material. All authors reviewed and accepted the manuscript.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have, or could be perceived to have, influenced the work reported in this article.

## Acknowledgements

We thank Tamás Bazsinka for technical assistance. This study has been funded by University of Helsinki Research Foundation, Biomedicum Helsinki Foundation, Emil Aaltonen Foundation, The Finnish Medical Foundation and the MD PhD program of the University of Helsinki.

## References

[1] N. Heikkilä, R. Vanhanen, D.A. Yohannes, I. Kleino, I.P. Mattila, J. Saramäki, et al., Human thymic T cell repertoire is imprinted with strong convergence to shared sequences, Mol. Immunol. 127 (2020) 112–123.
[2] N. Heikkilä, R. Vanhanen, D.A. Yohannes, P. Saavalainen, S. Meri, T.S. Jokiranta, et al., Identifying the inheritable component of human thymic T cell repertoire generation in monozygous twins, Eur. J. Immunol. (2019), doi:10.1002/eji.201948404.
[3] M. Shugay, D.V. Bagaev, M.A. Turchaninova, D.A. Bolotin, O.V. Britanova, E.V. Putintseva, et al., VDJtools: unifying post-analysis of T Cell receptor repertoires, PLoS Comput. Biol. 11 (2015) e1004503.
[4] R. Vanhanen, N. Heikkilä, K. Aggarwal, D. Hamm, H. Tarkkila, T. Pätilä, et al., T cell receptor diversity in the human thymus, Mol. Immunol. 76 (2016) 116–122.
[5] E. Quirós Roldan, A. Sottini, A. Bettinardi, A. Albertini, L. Imberti, D. Primi, Different TCRBV genes generate biased patterns of V-D-J diversity in human T cells, Immunogenetics 41 (1995) 91–100.
[6] I.V. Zvyagin, M.V. Pogorelyy, M.E. Ivanova, E.A. Komech, M. Shugay, D.A. Bolotin, et al., Distinctive properties of identical twins' TCR repertoires revealed by high-throughput sequencing, Proc. Natl. Acad. Sci. U. S. A. 111 (2014) 5980–5985.
[7] J.-.E. Park, R.A. Botting, C. Domínguez Conde, D-M Popescu, M. Lavaert, D.J. Kunz, et al., A cell atlas of human thymic development defines T cell repertoire formation, Science 367 (2020), doi:10.1126/science.aay3224.
[8] H.S. Robins, P.V. Campregher, S.K. Srivastava, A. Wacher, C.J. Turtle, O. Kahsai, et al., Comprehensive assessment of T-cell receptor beta-chain diversity in alphabeta T cells, Blood 114 (2009) 4099–4107.
[9] D.J. Laydon, C.R. Bangham, B. Asquith, Estimating T-cell repertoire diversity: limitations of classical estimators and a new approach, Philos. Trans. R. Soc. Lond. B Biol. Sci. 370 (2015), doi:10.1098/rstb.2014.0291.
[10] T. Mora, A.M. Walczak, How many different clonotypes do immune repertoires contain? Curr. Opin. Syst. Biol. 18 (2019) 104–110.
[11] M. Shugay, D.A. Bolotin, E.V. Putintseva, M.V. Pogorelyy, I.Z. Mamedov, D.M. Chudakov, Huge overlap of individual TCR beta repertoires, Front Immunol. 4 (2013) 466.