eLife

Reviewed Preprint
v2 • November 20, 2024
Revised by authors

Reviewed Preprint
v1 • March 5, 2024

Immunology and Inflammation

Distinct T Cell Receptor (TCR) gene segment usage and MHC-restriction between foetal and adult thymus

Jasmine Rowell, Ching-In Lau, Susan Ross, Diana C Yanez, Oscar A Peña, Benny Chain, Tessa Crompton

UCL Great Ormond Street Institute of Child Health, 30 Guilford Street, London, UK • School of Biochemistry University of Bristol, Bristol, United Kingdom • Division of Infection and Immunity, University College London, London, United Kingdom

eLife Assessment

This **important** manuscript provides an extensive and **convincing** analysis of the foetal and adult TCR repertoire in the mouse thymus. A potential implication of the work is that the earliest appearing T cells during ontogeny may have properties that are fundamentally distinct from those appearing later in life. The study will be of interest to immunologists concerned with T cell development and TCR repertoires.

https://doi.org/10.7554/eLife.93493.2.sa3

Abstract

Here we sequenced rearranged TCRβ and TCRα chain sequences in CD4+CD8+ double positive (DP), CD4+CD8-single positive (SP4) and CD4-CD8+ (SP8) thymocyte populations from the foetus and young adult mouse. We found that life-stage had a greater impact on TCRβ and TCRα gene segment usage than cell-type. Foetal repertoires showed bias towards 3'TRAV and 5'TRAJ rearrangements in all populations, whereas adult repertoires used more 5'TRAV gene segments, suggesting that progressive TCRα rearrangements occur less frequently in foetal DP cells. When we synchronised young adult DP thymocyte differentiation by hydrocortisone treatment the new recovering DP thymocyte population showed more foetal-like 3'TRAV and 5'TRAJ gene segment usage. In foetus we identified less influence of MHC-restriction on α-chain and β-chain combinatorial VxJ usage and CDR1xCDR2 (V region) usage in SP compared to adult, indicating weaker impact of MHC-restriction on the foetal TCR repertoire.

The foetal TCRβ repertoire was less diverse, less evenly distributed, with fewer non-template insertions, and all foetal populations contained more clonotypic expansions than adult. The differences between the foetal and adult thymus TCR repertoires are consistent with the foetal thymus producing αβT-cells with properties and functions that are distinct from adult T-cells: their repertoire is less governed by MHC-restriction, with preference for particular gene segment usage, less diverse with more clonotypic expansions, and more closely encoded by genomic sequence.

Introduction

T-cells are essential for adaptive immunity to mount an immune response against a wide and unpredictable array of foreign pathogens. T-cells recognise pathogens through their T-cell receptor, and adaptive immunity relies upon a broad repertoire of T-cells with TCRs with different specificities to potentially unknown pathogens, so that should a new threat arise, the immune system can respond. The diverse repertoire of possible TCRs is generated from gene rearrangement of two independently generated chains, the β chain and α chain, which come together to form the TCR. This diverse pool of αβTCR is generated during T-cell development in the thymus by the RAG-dependent joining of variable (V), joining (J), and diversity (D) gene segments at the β locus, and V and J gene segments at the α locus, with additional diversity coming from non-template nucleotides (junctional diversity)(Schatz and Swanson, 2011).

The thymus is essential for both T-cell development and TCR repertoire selection, but undergoes changes in output and function across the life of a mouse(Kondo et al., 2019, Montecino-Rodriguez and Dorshkind, 2023). In this study, we compared the αβTCR repertoire generated in the foetal thymus to that of the young adult, to investigate how these life-stages influence V(D)J gene usage and TCR repertoire diversity and distribution.

The thymus and the parathyroid organs, emerge from the third pharyngeal pouch and cleft during mid gestation of the foetal mouse(Gordon and Manley, 2011). Haematopoietic progenitors migrate from the foetal liver through the blood stream to colonise the thymus from embryonic day (E) 10-12. The first CD4-CD8-double negative (DN) thymocytes are already present at E11 and proliferate over the following days, producing a larger population by E14. CD4+CD8+ double positive (DP) and CD8+ immature single positive (ISP) thymocytes are first detectable at ~E16, while mature CD4+CD8-CD3+ single positive (SP4) and CD4-CD8+CD3+ single positive (SP8) thymocytes can be detected at day E18 (Solanki et al., 2018, Solanki et al., 2020). After birth, the thymus continues to grow rapidly and reaches its peak size at 4 to 6 weeks of age (Xiao et al., 2003).

During this developmental progression, the *T cell receptor (TCR) β* and *TCRα* loci are sequentially rearranged to produce T-cells with a diverse TCR repertoire. Successful rearrangement of *TCRβ* and pre-TCR expression are required for commitment to the αβT-cell lineage and differentiation from the CD25+DN (DN3) stage to DP cell, whereas cells that successfully rearrange the γδTCR commit to the γδT-cell lineage at the CD25+DN stage. *TCRα* rearrangement and signalling through the αβTCR are essential for MHC-restriction and differentiation from DP to SP cell (positive selection) and then also for deletion of T-cells bearing self-reactive TCR (negative selection). Pre-TCR signalling is required for survival, expansion and differentiation of cells that have completed TCRβ rearrangement, to produce a pool of cells which express a single functional TCRβ chain, in which TCRα rearrangement and αβTCR repertoire selection will occur(Dutta et al., 2021) and β-selection is believed to involve immune synapse formation and MHC-interactions(Allam et al., 2021). In general, allelic exclusion of the TCRβ chain locus is believed to prevent developing thymocytes from simultaneously rearranging and expressing two TCRβ chains, whereas in DP cells the TCRα chain locus undergoes biallelic Vα-Jα rearrangement, with many possible rounds of rearrangement on each allele(Carico and Krangel, 2015, Carico et al., 2017, Genolet et al., 2012). Although the sequence of this developmental programme is the same between foetal and adult αβT-cell development, progression beyond the DN3 stage is less tightly linked to TCRβ expression and β-selection in the foetal thymus(Hager-Theodorides et al., 2007).

There are several other fundamental differences between adult and foetal thymus that may affect the generation and selection of the TCR repertoire. In contrast to adult αβT-cells which are produced in a thymus structure that is mature, during foetal T-cell development, the thymus is still

forming: the structure of the medulla and differentiation of the medullary thymic epithelial cell (mTEC) population, which express tissue restricted antigens for tolerance induction, are dependent on the haematopoietic compartment, with different requirements in foetal and adult thymus(Desanti et al., 2012). There are also differences inthe progenitor cells that seed the thymus at different life-stages(Ramond et al., 2014), which arise from either the foetal liver or postnatal bone-marrow, and enter the thymus through different locations(Montecino-Rodriguez and Dorshkind, 2023). Some foetal haematopoietic stem cells (HSC) have been shown to have increased capacity to proliferate and to differentiate preferentially into innate-like lymphocytes compared with adult cells(Beaudin et al., 2016). Additionally, CD8 T-cells derived from HSCs from the foetal liver are functionally distinct from adult CD8 T-cells(Wang et al., 2016).

The foetal TCR repertoire is known to be less diverse than the adult repertoire. TdT is absent from the foetal thymus, and is first expressed a few days after birth, so foetal repertoires contain less non template nucleotide additions(Bogue et al., 1992). This has been confirmed by next generation TCR sequencing from mouse and human foetal and adult repertoires, which showed increased diversity after birth that was attributed to the postnatal increase in non-template nucleotide additions(Sethna et al., 2017, Britanova et al., 2016, Pogorelyy et al., 2017). Additionally, a PCR approach has indicated that the foetal TCRα repertoire uses a limited range of VJ rearrangements with enrichment of proximal VJ rearrangements (Pasqual et al., 2002).

In this study we compared the TCR repertoire that is generated and selected in the foetal thymus with that from the young adult thymus. In addition to the reduction in non-template nucleotide additions seen in the embryonic repertoire, we identified biases in VJ usage and VxJ pairing of both TCRβ and TCRα repertoires. The foetal TCRβ repertoires were less diverse and less equally distributed with more clonal expansions than the young adult repertoires.

Results

Foetal thymocyte populations show clonotypic TCR β expansion with a less diverse TCR β repertoire than adult

To compare the TCRβ chain repertoire generated in the embryonic and young adult thymus we FACS-sorted DP (CD4+CD8+), SP4 (CD4+CD8-CD3+) and SP8 (CD4-CD8+CD3+) populations from E18.5 (foetal) and 4 week-old (young adult) thymus and sequenced their rearranged TCR transcripts. After RNA extraction from purified populations we used a TCR sequencing protocol and analysis pipeline which includes single-strand DNA ligation that tags each molecule of TCR α- and β-chain mRNA with a unique molecular identifier (UMI), allowing PCR bias to be corrected for in later analysis(Oakes et al., 2017, Uddin et al., 2019, Thomas et al., 2013). Firstly, to visualise the frequency distribution of the TCRβ repertoire in each population, we plotted the TCRβ abundance (number of copies of each sequence) against their proportion of the repertoire (**Fig.1A**). For each of the developmentally-defined thymocyte populations, the foetal thymus showed an increase in the proportion of abundant clones compared to adult mice in DP, SP4 and SP8 populations (**Fig.1A**). The frequency distributions of TCRβ abundances were fitted to a discrete power law, where the power law exponent corresponds to the gradient on the log-log plots (**Fig.1A**). The power law exponent was significantly lower in each foetal population compared to its adult counterpart (**Fig.1B**), demonstrating a difference in distribution of abundance of TCRβ clones between foetal and adult, consistent with an increase in clonotypic expansion of abundant clones in the foetal TCRβ repertoire. To confirm this, we calculated the proportion of the TCRβ repertoire represented by the top 1% most abundant TCRβ clones in each population. In foetal DP and SP4 and SP8 populations the proportion of the TCRβ repertoire represented by the top 1% most abundant TCRβ clones was more than twice that of the adult populations (**Fig1C**), and the mean abundance of TCRβ sequences detected in the top 1% most abundant (expanded) clones was also significantly higher (**Fig. 1D**). in the foetal SP4 population only, suggesting a difference in

equality of distribution in the top 1% most abundant (expanded) clones in the DP and SP8 populations compared to SP4 (**Fig.1D**). We therefore calculated standard indices of diversity, richness and distribution: Shannon entropy and the Gini index. Shannon entropy is derived from information theory and represents the complexity of the system, taking into account both the number of unique clonotypes and the frequencies of each clonotype, to generate a measure of richness and diversity(Shannon and Weaver, 1949). For each thymocyte population, the Shannon entropy was higher in the young adult TCRβ repertoire than in the foetal, demonstrating a richer more diverse repertoire in the adult thymocyte populations (**Fig1E-F**). The Gini index is a measure of the equality (evenness) of the distribution, and is frequently used in economics to demonstrate the distribution of a population's wealth, where a lower index represents a more equal distribution(Ceriani and Verme, 2012). The Gini index of the distribution of TCRβ clonotypes was significantly lower in each adult population compared to its foetal counterpart, confirming a more equally distributed TCRβ repertoire in the young adult thymus (**Fig1G-H**).

In parallel we analysed the repertoire diversity, richness and distribution of the TCRα repertoires (**SFig1**). The fitted power law exponent of the frequency distribution of TCRα abundances was significantly lower in the foetal DP and SP4 populations compared to their adult counterparts (**SFig1A-B**), whereas the proportion of the TCRα repertoire represented by the top 1% most abundant TCRα clones was higher in the foetal DP population than the adult DP population (**SFig1C**). However, we detected no significant differences between foetal and adult populations in the abundance of TCRα sequences detected in the top 1% most frequent (expanded) clones (**SFig1D**). The Shannon entropy was significantly lower in each foetal population compared to its adult counterparts (**SFig1E-F**), but the Gini index was significantly higher in each foetal population (SFigG-H), confirming that foetal TCRα repertoires were less rich, diverse and evenly distributed than young adult TCRα repertoires.

Summary

Foetal DP, SP4 and SP8 TCRβ and TCRα repertoires are less evenly distributed and less diverse than their adult counterparts and foetal TCRβ repertoires contain more clonal expansions.

Foetal and young adult TCR β and TCR α repertoires use distinct TRAV, TRAJ, TRBV and TRBJ gene segments

We next tested if the foetal and young adult repertoires showed differences in usage of TCR V and J gene segments. We found many differences between foetus and adult in proportional V and J segment usage within unique TCRβ and TCRα sequences for DP (**SFig2**), SP4 (**SFig3**) and SP8 (**SFig4**) populations. Preferential gene segment usage according to life-stage appeared to correlate with the chromosomal location of the gene segment, particularly for TRAV and TRBV usage within the DP population, where the adult repertoire favoured the 5' gene segments, and the foetus the 3' gene segments. To visualise these differences, we plotted heatmaps of mean proportional gene segment usage for unique TCR sequences, clustering each life-stage and thymocyte population, but showing each gene segment in chromosomal order (**Fig2A-D**). The samples clustered by life-stage first before cell-type (E18.5 populations clustered together) for both TCRα and TCRβ, suggesting that VJ gene usage is altered in the foetus compared to young adult. Both TRAV and TRAJ showed clear preference for chromosomal location, with foetal populations favouring 3' TRAV segments and 5' TRAJ segments, whereas adult populations showed enrichment of 5' TRAV and 3' TRAJ (**Fig2A-B**). To test if this bias in gene usage was present in the unselected TCRα repertoire, we separated the DP population into cells that have not yet entered positive selection ($CD3^{-/lo}$DP/$CD69^{-}$DP) and cells that have entered TCR repertoire selection ($CD3^{+/hi}$DP/$CD69^{+}$DP). Both foetal DP populations clustered together and showed similar bias towards 3' TRAV and 5' TRAJ, suggesting that the bias was independent of positive selection. In the case of the adult DP populations, however, the more immature DP population clustered separately from the other young adult populations and showed weaker enrichment 3'TRAJ than the

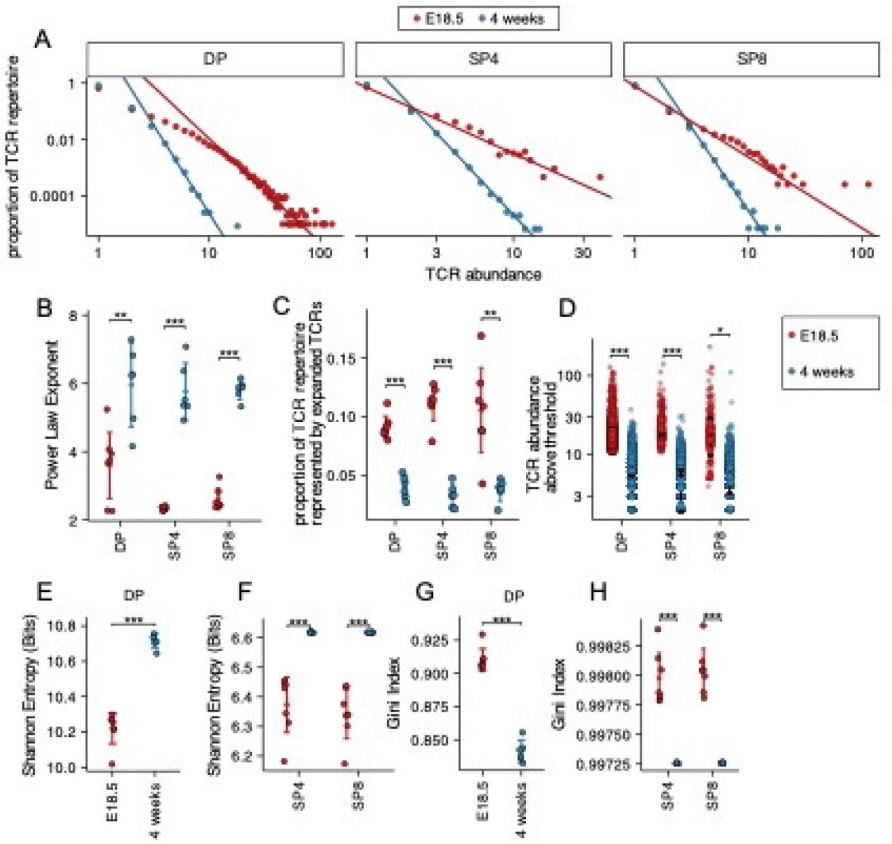


Figure 1.

Decreased diversity and increased clonality of foetal compared to young adult TCRβ repertoires

TCR β-chain repertoires were sequenced from FACS-sorted $CD4^+CD8^+$ (DP), $CD4^+CD8^-CD3^+$ (SP4) and $CD4^-CD8^+CD3^+$ (SP8) thymocyte populations from E18.5 (red circles; n=7) and 4 week-old (blue circles; n=6) C57BL/6 thymus.
(A) The frequency distribution of TCR β-chain abundance was fitted to a discrete power law ($f(k) = Ck^{-\alpha}$) by maximum likelihood (solid line). These logged plots show a representative individual mouse TCR repertoire frequency distribution for E18.5 and 4 weeks in DP, SP4 and SP8 populations. The x axis represents TCR abundance (size of clone), and the y axis represents the proportion of the repertoire. The negative of the power law exponent corresponds to the slope of the logged plot. **(B)** The power law exponents of the frequency distribution of TCR β-chain abundances for E18.5 and 4 weeks in DP, SP4 and SP8 populations, where each point represents an individual mouse/embryo. **(C)**The proportion of the total TCR β-chain repertoire accounted for by the expanded top 1% most abundant sequences (>99^{th} percentile) for E18.5 and 4 weeks for DP, SP4 and SP8 populations, where each point represents an individual mouse/embryo. **(D)** The number of β-chain sequences detected above the given frequency threshold (top 1% most abundant sequences) is shown for E18.5 (red circles) and 4 weeks (blue circles) in DP, SP4 and SP8 populations. Each small translucent point represents the abundance of any β-chain sequence detected above the threshold for all mice/embryos at that life-stage, while each larger solid point represents the mean of β-chain sequence abundance for each individual mouse/embryo. **(E)** The Shannon entropies for the β-chain TCR repertoires of E18.5 and 4 weeks in DP populations, where each point represents an individual mouse/embryo. Each repertoire was subsampled to 50000 TCRs 1000 times before calculating the Shannon entropy. **(F)** The Shannon entropies for the β-chain TCR repertoires of E18.5 and 4 weeks in SP4 and SP8 populations, where each point represents an individual mouse/embryo. Each repertoire was subsampled to 750 TCRs 1000 times before calculating the Shannon entropy. **(G)** The Gini indexes for the β-chain TCR repertoires of E18.5 and 4 weeks in DP populations, where each point represents an individual mouse/embryo. Each repertoire was subsampled to 50000 TCRs 1000 times before calculating the Gini index. **(H)** The Gini indexes for the β-chain TCR repertoires of E18.5 and 4 weeks in SP4 and SP8 populations, where each point represents an individual mouse/embryo. Each repertoire was subsampled to 750 TCRs 1000 times before calculating the Gini index.
Dotplots show mean±c.i and statistical comparisons were carried out by unpaired Student's t-test or Welch's t-test as appropriate: *** p <0.001; ** p <0.01; *p < 0.05.

populations that had undergone or were undergoing positive selection ($CD3^{+/hi}$DP, SP4 and SP8). This suggests that in the adult thymus, positive selection changes proportional TRAJ gene usage, whereas in the foetal thymus an inherent bias in VJα gene usage persists after positive selection.

Heatmaps for TRBV also broadly showed chromosomal location-bias for gene usage at the different life-stages, with foetal populations preferentially using 3'TRBV, although the pattern appeared less pronounced (**Fig2C-D**). Comparison of TRBJ cluster usage between adult and foetus showed that foetal SP TCRβ repertoires used proportionally more TRBJ gene segments from the TRBJ1 gene cluster than their adult counterparts, whereas adult SP repertoires used proportionally more TRBJ gene segments from the TRBJ2 cluster suggesting that positive selection influences TRBJ cluster usage (**SFig5** and **SFig2-4**). Therefore, to evaluate further the differences in TCRβ gene segment usage between foetal and adult repertoires, we carried out Principal Component Analysis (PCA) of the β-chain variable and joining gene counts of the unique TCRβ chains for each population (**Fig. 2E-J**). PCA separated the repertoires for DP, SP4 and SP8 by life-stage on PC1, confirming differences in TRBV and TRBJ usage between life-stage for each population.

Summary

Foetal and adult DP SP4 and SP8 thymocyte populations use distinct TCRβ and TCRα gene segments, with foetal populations showing bias towards 3' TRAV, 5' TRAJ, 3'TRBV and the TRBJ1 gene cluster.

Foetal and adult TCRβ and TCRα repertoires use distinct combinations of V and J segments

Given this difference in TRBV and TRBJ gene usage between foetal and adult, we next tested if TCRβ foetal and young adult thymocyte repertoires use different combinations of V and J gene segments. We carried out PCA using the counts of each VxJ combination from unique TCRβ sequences in foetal and adult DP, SP4 and SP8 populations (**Fig3A-C**). PCA separated by life-stage on PC1, with adult repertoires clustering tightly together (**Fig3A**). When we compared by thymocyte population, each adult population clustered together, with adult SP4 separating from adult SP8 on PC2, and DP cells scoring in between, suggesting that PC2 corresponds to MHC-restriction of the adult populations (**Fig3B**). In contrast, the foetal populations were more dispersed, and did not segregate fully by cell type on PC2 (**Fig3B**). When we examined which VxJ combinations contributed most to PC1, we found strong correlation to the proportional TRBV and TRBJ usage detected: TRBV13-1 and TRBJ2-7 both showed adult bias in the analysis of individual gene usage, and in combination also contributed strongly to the positive score on PC1; whereas TRBV20 and TRBJ1-3 both showed foetal bias in the analysis of individual gene usage and contributed strongly to the negative (foetal) score on PC1, reflecting chromosomal gene segment location (**SFig2-4** and **Fig2-D**).

We next compared usage of all possible VxJ combinations between the adult and foetal thymocyte populations for both TCRβ and TCRα repertoires, identifying all combinations that showed a significant change in proportional usage in foetus compared to young adult (**Fig4A-C**). The DP foetal population showed many increases (86, shown in red) and decreases (72, shown in blue) in proportional usage of VxJ combinations for TCRβ repertoire compared to young adult, which after selection were reduced in the SP4 and SP8 populations, to 37 increases and 34 decreases in SP4, and 26 increases and 25 decreases in SP8. In the case of the TCRα chain, the foetal DP repertoire favoured VxJ combinations from the 3' location of TRAV and 5' location of TRAJ, and this bias persisted but was less pronounced in the SP populations (**Fig.4B-C**). Concomitantly, the DP foetal population showed clear proportional decrease in usage of combinations of 5' TRAV and 3' TRAJ in comparison to young adult.

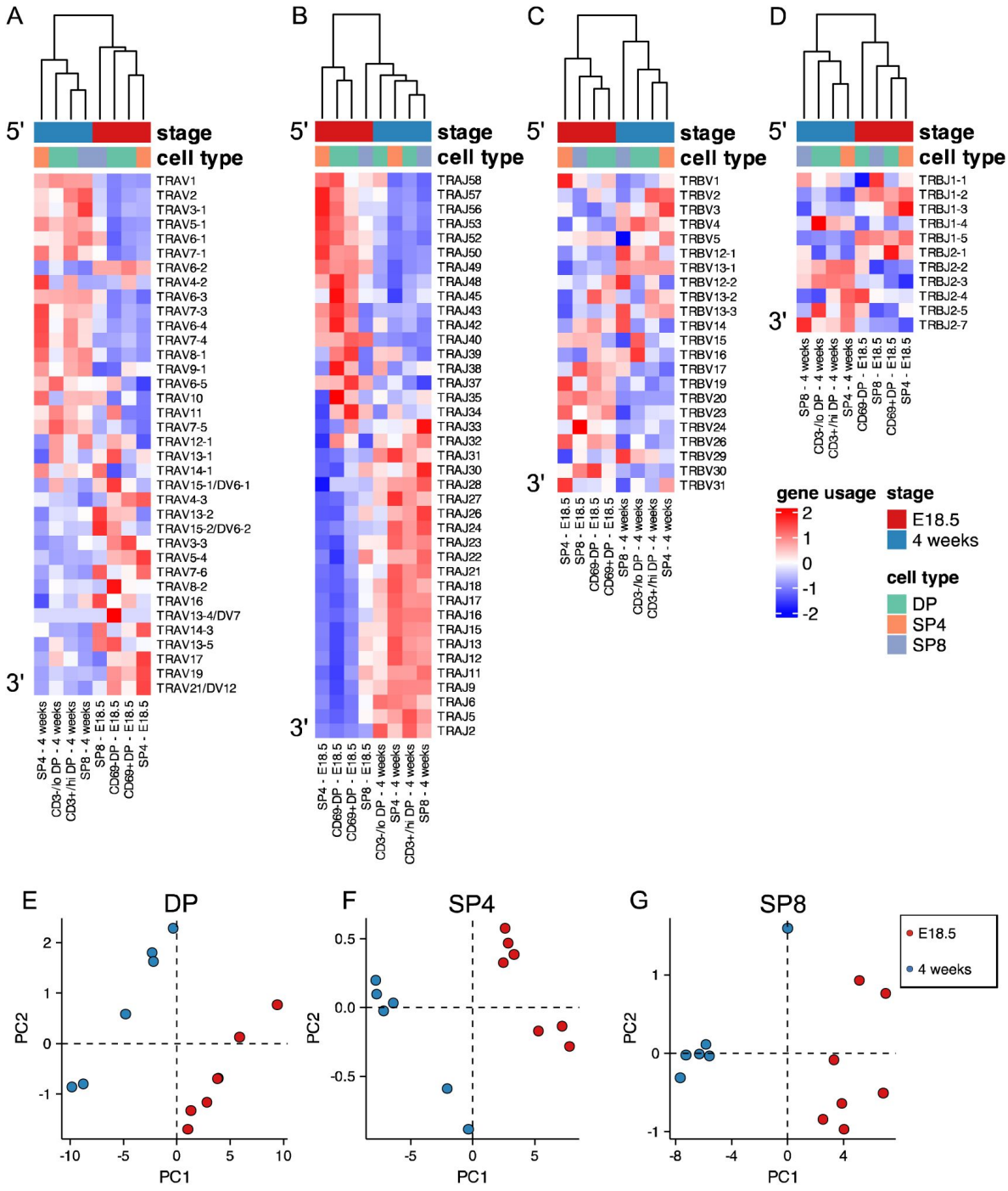


Figure 2.

Foetal and young adult TCR repertoires favour different gene segments

TCR α-chain (A-B) and β-chain (C-G) repertoires were sequenced from FACS-sorted $CD4^+CD8^-CD3^+$ (SP4) and $CD4^-CD8^+CD3^+$ (SP8) $CD4^+CD8^+CD69^-$ (CD69-DP, foetal), $CD4^+CD8^+CD69^+$ ($CD69^+$DP foetal) $CD4^+CD8^+CD3^{-/lo}$ ($CD3^{-/lo}$DP, adult), $CD4^+CD8^+CD3^{+/hi}$ ($CD3^{+/hi}$DP adult) thymocyte populations from E18.5 (red; n=7) and 4 week-old (blue; n=6) C57BL/6 thymus. DP, SP4 and SP8 populations are coloured in green, orange and purple respectively.

(A-D) Heatmaps of proportional α-chain variable (V) **(A)**, α-chain joining (J) **(B),** β-chain V**(C)** and β-chain J **(D)** gene usage of unique TCRs for E18.5 and 4 weeks in DP, SP4 and SP8 populations. Each column represents a mean of 7 embryos or 6 mice and was clustered using Euclidian distance. Genes are shown in chromosomal order (5' to 3') from top to bottom. **(E-G)** PCA biplot of β-chain V and J gene counts of unique TCRs for E18.5 and 4 weeks in DP, **(E)** SP4 **(F)** and SP8 **(G)** populations.

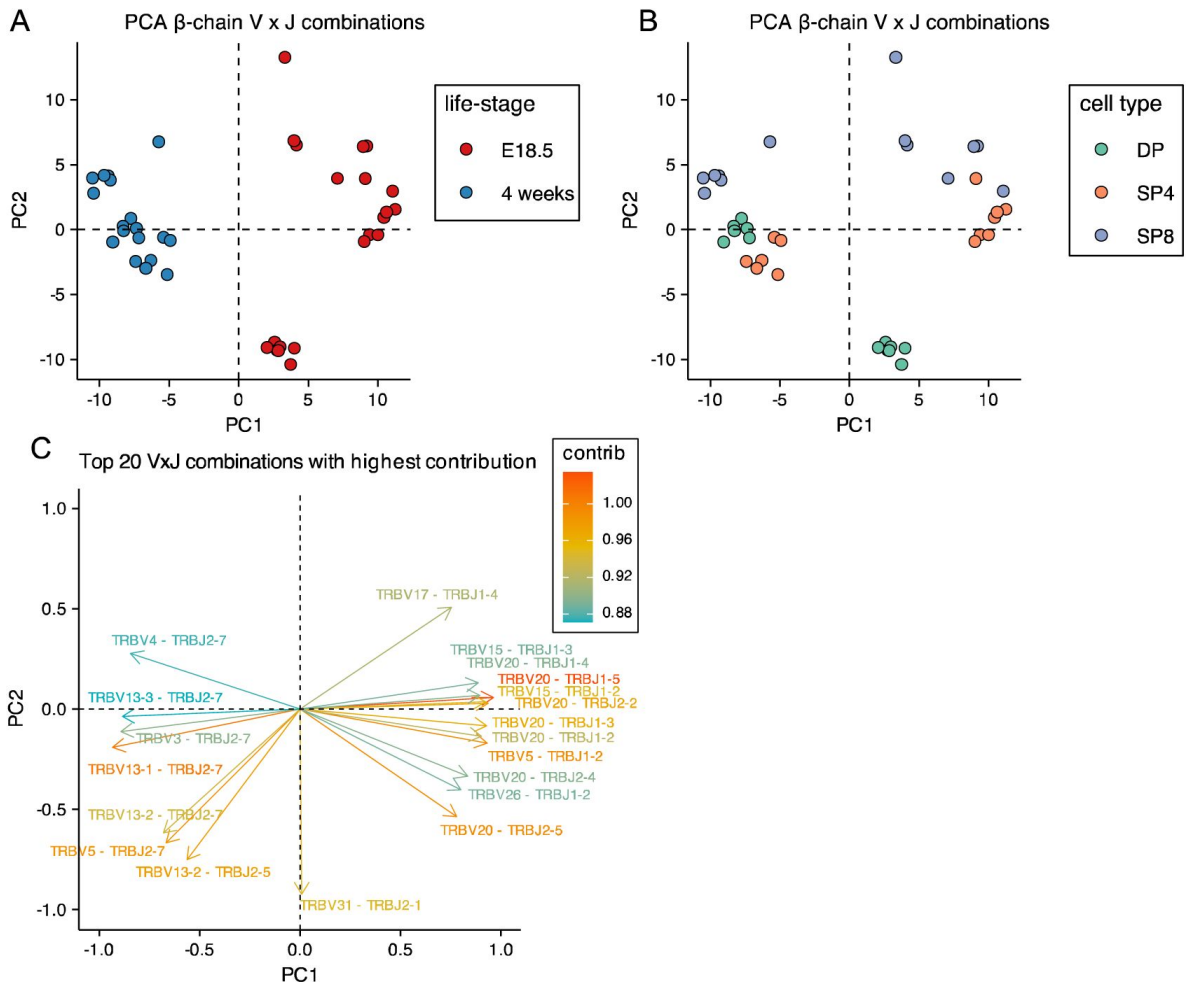


Figure 3.

Principal Component Analysis (PCA) of Variable x Joining combinations in embryonic and young adult TCRβ repertoires cluster by life-stage before cell-type

TCR β-chain repertoires were sequenced from FACS-sorted $CD4^+CD8^+$ (DP), $CD4^+CD8^-CD3^+$ (SP4) and $CD4^-CD8^+CD3^+$ (SP8) thymocyte populations from E18.5 (n=7) and 4 week-old (n=6) C57BL/6 thymus.
(A-B) PCA biplot of β-chain VxJ gene counts of unique TCRβ sequences for E18.5 and 4 weeks in DP, SP4 and SP8 populations. **(A)** E18.5 and 4 weeks are coloured in red and blue circles respectively. **(B)** DP, SP4 and SP8 populations are coloured in green, orange and purple respectively. **(C)** Top 20 highest β-chain VxJ gene combinations that contribute to PC1 and PC2 of the PCA.

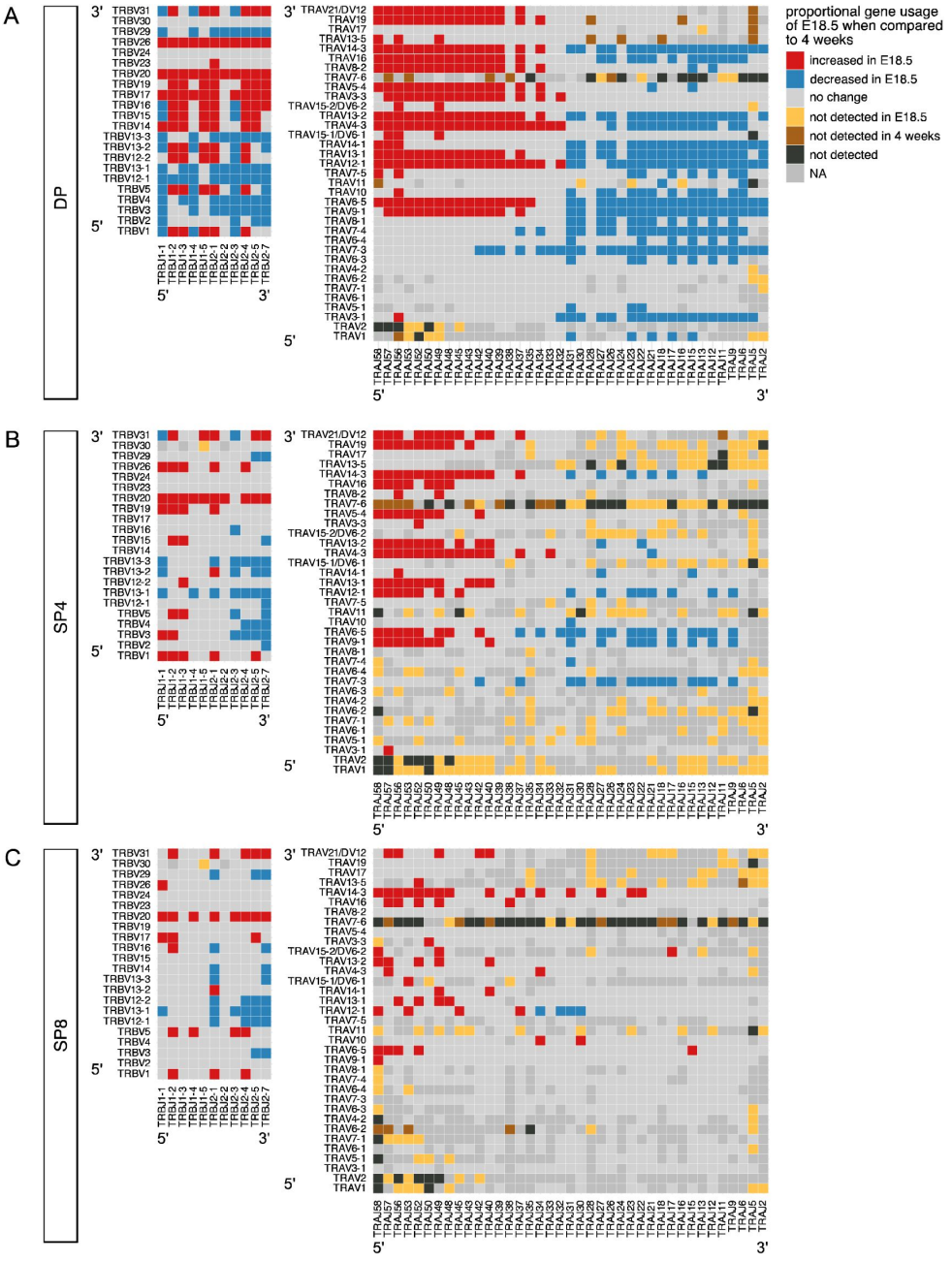


Figure 4.

Proportional bias in VxJ combinations in foetal and young adult repertoires by cell-type

TCR β-chain and α-chain repertoires were sequenced from FACS-sorted $CD4^+CD8^+$ (DP), $CD4^+CD8^-CD3^+$ (SP4) and $CD4^-CD8^+CD3^+$ (SP8) thymocyte populations from E18.5 (red; n=7) and 4 week-old (blue; n=6) C57BL/6 thymus.
(A-C) Plots show β-chain (left column) and α-chain (right column) proportional VxJ gene usage of total TCRs for E18.5 compared to 4 weeks in DP **(A)**, SP4 **(B)** and SP8 **(C)** populations. Red tiles signify increased usage of VxJ combinations in E18.5 ($p < 0.05$), while blue tiles signify decreased usage in E18.5 ($p < 0.05$) compared to 4 weeks. Light grey tiles signify no change in gene usage ($p > 0.05$), yellow tiles signify VxJ combinations not detected in E18.5, brown tiles signify VxJ combinations not detected in 4 weeks and black tiles signify VxJ combinations not detected in both E18.5 and 4 weeks. Only combinations that were detected in at least 3 mice mice/embryos per group were compared. Dark grey tiles signify combinations that were not compared. Statistical comparisons were carried out by unpaired Student's t-test followed by FDR-adjustment (5%, Benjamini-Hochberg procedure) of p values.

Summary

Foetal and adult DP, SP4 and SP8 thymocyte populations show bias in proportional combinatorial VxJ usage for both TCRα and TCRβ repertoires, with foetal TCRα repertoires enriched for proximal VJ rearrangements. PCA indicates that TCRβ combinatorial VXJ usage is less influenced by cell type (MHC-restriction) in foetal compared to adult repertoires.

Shorter CDR3 length and reduced non-template insertion length in foetal TCRβ repertoires

We next tested if the percentage of non-productive TCRβ and TCRα rearrangements was different in foetal and young adult populations. Out of frame (non-productive) rearrangements generated during TCR gene rearrangement are eliminated by nonsense-mediated decay during T-cell development(Weischenfeldt et al., 2008), although some non-productive TCR transcripts persist (Mahowald et al., 2011, Britanova et al., 2016, Camaglia et al., 2023). The proportion of non-productive rearrangements was higher in each foetal population than adult (**Fig5A** and **SFig6A**).

Furthermore, we examined non-template additions/deletions to V and J gene segments, which in the TCRβ chain include TRBD nucleotides. TRBD segments are difficult to identify as they are highly homologous and short. As previously reported (Sethna et al., 2017), the mean TCRβ non-template nucleotide insertion length was lower in the foetal DP, SP4 and SP8 populations compared to adult. To confirm this we then investigated the length of the complementary determining region 3 (CDR3) of the TCRβ chain as it is the most hypervariable region and primary site of antigen recognition. The mean CDR3 length was shorter in all three populations in foetal TCRβ repertoires compared to young adult, confirming previous studies (Sethna et al., 2017) (**Fig5B-C**).

Likewise, when we compared amino acid CDR3 sequences for the TCRα repertoires, we found lower mean length of non-template insertions (**SFig6B**) and mean α-chain CDR3 length in the foetal thymocyte populations compared to their adult counterparts (**SFig6C**).

Sharing of CDR3 sequences differs between adult and foetal TCRβ and TCRα repertoires

In order to investigate sharing between the TCRβ CDR3 repertoires within the same cell population from different embryos/mice of the same life-stage, we used the Jaccard Index of Similarity. This index measures the size of the intersection divided by the size of the union of finite sample sets(Jaccard, 1912). We calculated the Jaccard Index of Similarity for each TCRβ CDR3 repertoire from each thymocyte population, comparing different mice/embryos to determine how similar each foetal repertoire is to the repertoire of the same thymocyte population from another foetus, and likewise for young adults. For each population, the Jaccard index was significantly higher in the foetal samples compared to adult, indicating that the TCRβ CDR3 repertoires in each population from different embryos shared more sequences than the CDR3 repertoires from different individual adults (**Fig5D-E**). This increased sharing of TCRβ CDR3 sequences in the foetal repertoires is consistent with shorter CDR3 length and shorter non-template insertions, leading to a repertoire that contains more genomic sequence than adult.

In the case of the α-chain repertoires, foetal DP and SP4 repertoires shared more TCRα CDR3 sequences between individuals than adults shared between individuals (**SFig6D-E**).

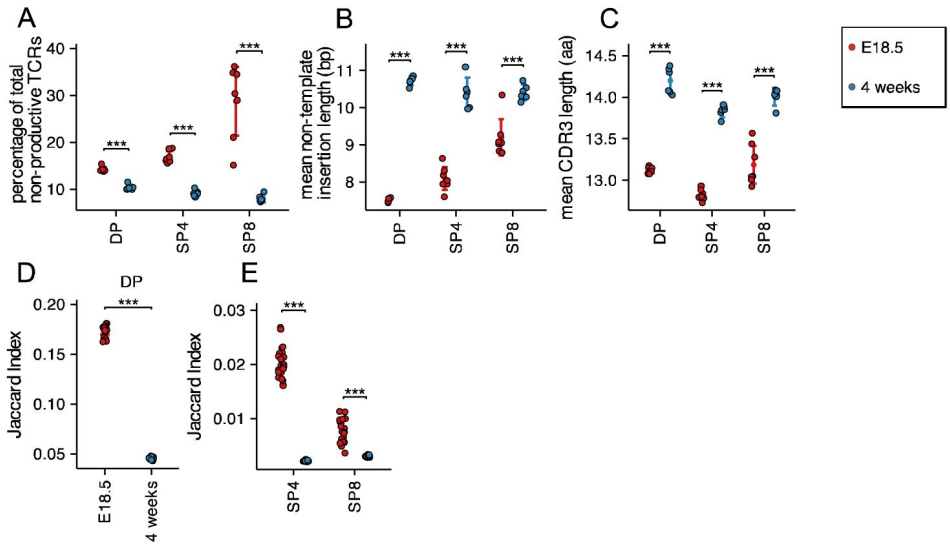


Figure 5.

Decreased TCR β non-template insertions, CDR3 length with increased CDR3 sharing, but reduced amino acid motif sharing in foetal compared to young adult

TCR β-chain repertoires were sequenced from FACS-sorted $CD4^+CD8^+$ (DP), $CD4^+CD8^-CD3^+$ (SP4) and $CD4^-CD8^+CD3^+$ (SP8) thymocyte populations from E18.5 (red circles; n=7) and 4 week-old (blue circles; n=6) C57BL/6 thymus. **(A)** Percentage of non-productive β-chain TCRs from total rearrangements for E18.5 and 4 weeks in DP, SP4 and SP8 populations. **(B)** The weighted mean of non-template insertion length for all β -chain TCRs (basepairs) for E18.5 and 4 weeks in DP, SP4 and SP8 populations. **(C)** The mean of β-chain unique predicted CDR3 length (number of amino acids) for E18.5 and 4 weeks in DP, SP4 and SP8 populations. **(D)** The intra life-stage Jaccard Index of Similarity of β-chain CDR3s for E18.5 and 4 weeks in the DP population. Each repertoire was subsampled to 30000 CDR3s 1000 times before calculating the Jaccard Index. **(E)** The intra life-stage Jaccard Index of Similarity of β-chain CDR3s for E18.5 and 4 weeks in the SP4 and SP8 population. Each repertoire was subsampled to 500 CDR3s 1000 times before calculating the Jaccard Index. Dotplots show mean±c.i and statistical comparisons were carried out by unpaired Student's t-test or Welch's t-test as appropriate: *** p <0.001; ** p <0.01; *p < 0.05.

Summary

Foetal TCRβ and TCRα CDR3 repertoires have lower mean non-template insertion length and lower mean CDR3 length than adult, and share more CDR3 sequences with one another than adult.

Different CDR1 and CDR2 usage between foetal and adult TCRβ and TCRα repertoires: less MHC-restriction in foetal repertoires

The CDR1 and CDR2 loops in the TCR are encoded in the V gene segments and typically contact the MHC's conserved α-helices. They can determine MHC restriction during repertoire selection(Wong et al., 2019), so we examined CDR1xCDR2 combinations in the foetal and adult thymocyte populations. We carried out PCA, using the frequency distributions (counts) of each β-chain CDR1xCDR2 combination (corresponding to V gene segment) as input (**Fig6A-C**). The PCA clustered the populations by life-stage and cell-type, with all adult samples falling on the negative side of PC1, and clustering tightly by cell type on PC2, so that adult SP8 repertoires were positive on PC2, and adult DP and SP4 populations negative on PC2, with adult DP positioned between adult SP clusters. In contrast the foetal SP8 population was more dispersed and failed to cluster tightly on PC2, and DP repertoires fell on the positive side of PC2. To further visualise this, we generated a heatmap of the proportional usage of β-chain CDR1xCDR2 combinations for each foetal or adult thymocyte population (**Fig. 6D**). The foetal populations clustered together, showing preference for distinct CDR1xCDR2 combinations, which reflect the bias observed in their TRBV gene usage (**Fig 2C**). For both foetal and adult samples, DP populations were positioned between SP8 and SP4, indicating a divergence in CDR1xCDR2 usage as a result of MHC restriction following positive selection. Indeed, CDR1xCDR2 combinations that contributed strongly to the positive side on PC2 (**Fig6C**), such as MSHET-SYDVDS (TRBV29) and SGHSN-HYEKVE (TRBV12-1), showed increased expression in SP8 and decreased expression in SP4 (consistent with MHCI-restriction), while combinations that contributed strongly to the negative side of PC2, such as GKSSPN-SITVG (TRBV31), showed increased expression in SP4 and decreased expression in SP8 (consistent with MHCII-restriction). TCRα CDR1 and CDR2 are also encoded in V gene segments and may also be determined by MHC-restriction (Sim et al., 1996), so we additionally investigated TCRα CDR1xCDR2 combinations (**Fig6E**). Again, samples from each life-stage grouped together, and consistent with the strong life-stage dependent influence on TRAV use, several CDR1xCDR2 combinations showed an adult-bias in usage. PCA, using the frequency distributions (counts) of each predicted α-chain CDR1xCDR2 combination as input clustered adult repertoires on the negative axis of PC1, and separated adult SP4 repertoires from adult SP8 on PC2 indicating that TCRα CDR1/CDR2 usage was influenced by MHC restriction in adult repertoire selection, consistent with other reports in mouse and human(Camaglia et al., 2023, Suo et al., 2023). In contrast, foetal repertoires were highly dispersed and failed to segregate on either PC1 or PC2 (**SFig6F-G**).

Summary

CDR1xCDR2 (V gene segment) usage is different between adult and foetal repertoires, with adult TCRβ and TCRα CDR1xCDR2 repertoires clustering more tightly by PCA and segregating according to cell type (MHC-restriction).

Distinct MHC-restriction between adult and foetal TCR repertoires

PCA of β-chain VxJ, CDR1βxCDR2β (Vβ), and CDR1αxCDR2α (Vα), separated adult but not foetal SP repertoires according to cell type, suggesting a weaker influence of MHC-restriction on the foetal single positive repertoires. However, foetal and adult TCR repertoires show bias to usage of different V and J gene segments (**Fig2 A-D**, **SFig2-3**), so it is possible that foetal repertoires would cluster by cell type if PCA was carried out using foetal repertoires alone. We therefore repeated the PCA analyses on foetal and adult samples separately (**Fig.7A-L**). For foetal β-chain VxJ combinations, DP repertoires were separated from both SP repertoires on PC1 (corresponding

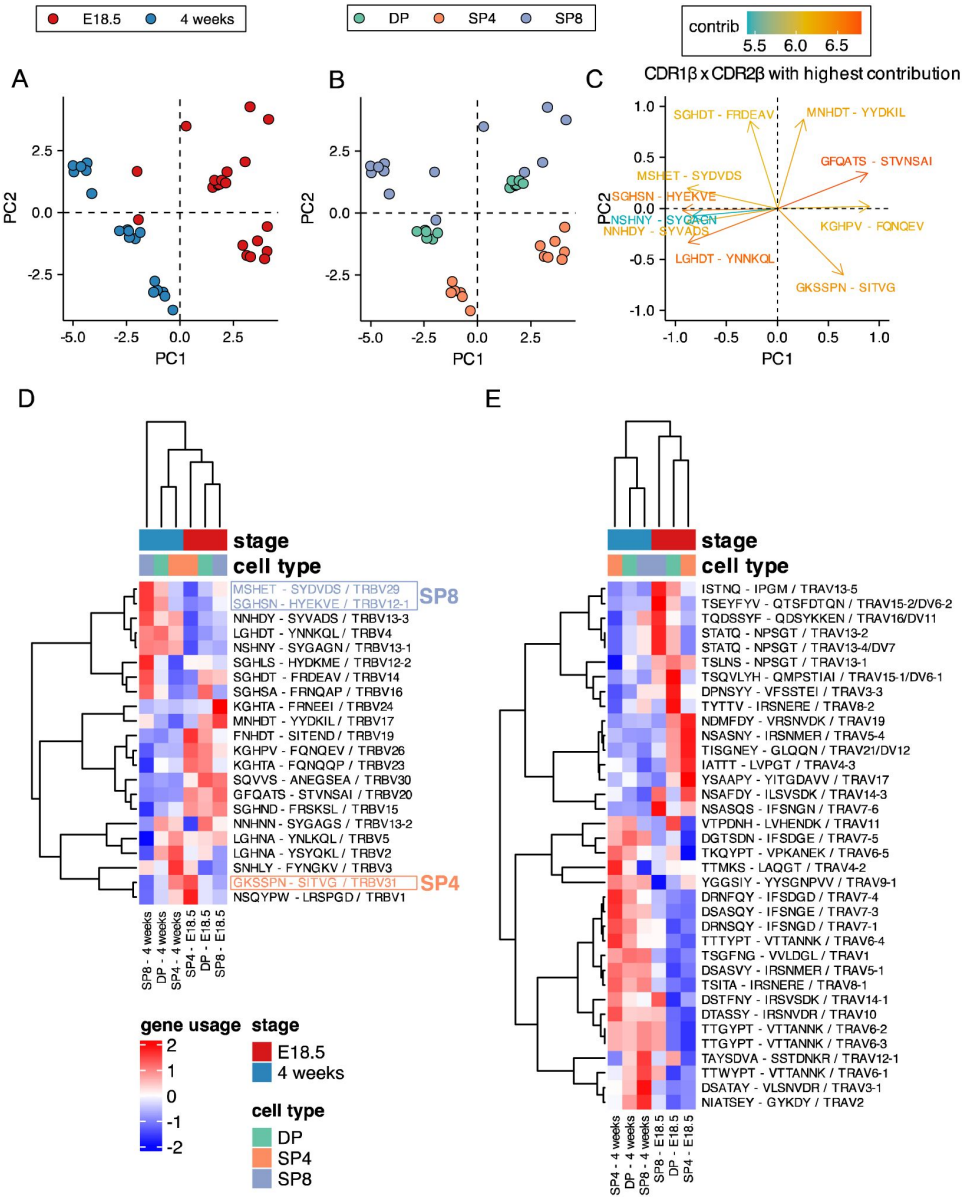


Figure 6.

Proportional combinatorial CDR1xCDR2 bias in foetal and young adult thymocyte populations indicates reduced MHC-restriction in foetal thymus

TCR β-chain and α-chain repertoires were sequenced from FACS-sorted $CD4^+CD8^+$ (DP), $CD4^+CD8^-CD3^+$ (SP4) and $CD4^-CD8^+CD3^+$ (SP8) thymocyte populations from E18.5 (n=7) and 4-week-old (n=6) C57BL/6 thymus and frequency of CDR1 and CDR2 sequences calculated for each population. E18.5 and 4-week samples are coloured red and blue respectively, while DP, SP4 and SP8 populations are coloured green, orange and purple respectively. **(A-B)** PCA biplot of β-chain CDR1xCDR2 frequency distributions for E18.5 and 4 weeks in DP, SP4 and SP8 populations. In **(A)**, E18.5 and 4-weeks samples are coloured in red and blue respectively, while in **(B)**, DP, SP4 and SP8 populations are coloured in green, orange and purple respectively. **(C)** Top 10 highest CDR1xCDR2 combinations that contribute to PC1 and PC2 of the PCA of β-chain CDR1xCDR2 frequency distributions for E18.5 and 4 weeks in DP, SP4 and SP8 population (shown in A-B). **(D)** Heatmap of proportional β-chain CDR1xCDR2 usage E18.5 and 4 weeks in DP, SP4 and SP8 populations. Each column represents a mean of 7 embryos or 6 adult mice and was clustered using Euclidian distance, while rows (CDR1 x CDR2 combinations) were clustered using Pearson correlation. **(E)** Heatmap of proportional α-chain CDR1xCDR2 usage in E18.5 and 4 week samples in DP, SP4 and SP8 populations. Each column represents a mean of 7 embryos or 6 mice and was clustered using Euclidian distance, while rows (CDR1xCDR2 combinations) were clustered using Pearson correlation.

to maturation), whereas SP4 repertoires separated from SP8 repertoires on PC2 (corresponding to MHC-restriction), which accounted for 11.1% of the variance between repertoires (**Fig.7A**). In contrast, for adult repertoires SP4 repertoires were separated from SP8 repertoires on PC1, which accounted for 44.2% of variance (**Fig. 7B**). Thus, MHC-restriction accounted for approximately 4-fold more of the variance in β-chain VxJ counts in the adult repertoires than in foetal repertoires. For the adult repertoires, the 10 β-chain VxJ combinations that contributed most to PC1 (MHC-restriction) included combinations that contributed to both the positive (SP4) and negative (SP8) side of PC1; whereas for foetal repertoires the top 10 β-chain VxJ combinations all contributed to the positive side of PC1 (DP cells) (**Fig. 7C-D**).

PCA of foetal α-chain VxJ counts failed to cluster repertoires by cell type on either PC1 or PC2 (**Fig. 7E**). In contrast, PCA of adult α-chain VxJ counts separated repertoires by SP cell type on PC1, which accounted for 25.9% of variance, with all SP4 repertoires falling on the positive side of the axis, and SP8 repertoires on the negative side (**Fig. 7F**).

We also carried out PCA of β-chain and α-chain CDR1xCDR2 (V region) frequency distributions for adult and foetal repertoires separately. For both adult and foetal repertoires, PCA separated β-chain repertoires by SP cell type (MHC-restriction) on PC1 (**Fig. 7G and H**). PC1 accounted for 56.1% of variability in the adult data, but only 28.8% of variance in foetal data, indicating a stronger impact of MHC-restriction on β-chain V region usage in adult than foetal thymocytes. Some combinations that contributed strongly to the negative (SP4) side of PC1 in the adult PCA, such as GKSSFN-SITVG (TRBV31), also contributed strongly to the foetal SP4 axis, but top contributors to SP4 in each PCA also included life-stage specific CDR1xCDR2 combinations, such as LGHNA-YSYQKL (TRBV2) and LGHNA-YNLKQL (TRBV5) for adult repertoires, and NSQYPQ-LRSPGD (TRBV1) for foetal repertoires (**Fig. 7I-J**). Several CDR1xCDR2 contributed strongly to the SP8 axis in both adult and foetal PCA, such as MNHDT-YYDKIL (TRBV17) and MSHET-SYDVDS (TRBV29), but life-stage specific combinations (KGHTA-FRNEEI (TRBV24) for foetal SP8 and SGHLS-HYDKME ((TRBV12-2) for adult SP8) also contributed strongly. Thus, for both SP4 and SP8 populations, life-stage influences which V region genes are selected.

PCA of α-chain CDR1xCDR2 (V region) counts for adult repertoires again separated SP4 repertoires from SP8 repertoires on PC1, accounting for 39.2% of variance, but failed to separate foetal repertoires by SP cell type on either PC1 or PC2 (**Fig7K-L**), confirming a stronger impact of MHC-restriction on the adult than foetal TCRα repertoire.

Summary

MHC-restriction contributes more to the variance in adult than foetal TCRβ and TCRα repertoires. This indicates that V gene region usage and VxJ combinatorial usage are more tightly governed by MHC-restriction in adult than in foetal thymus.

Young adult DP VJ α gene segment usage becomes more foetal-like after hydrocortisone treatment

We observed distinct patterns of V and J gene usage between foetus and young adult, with foetal DP repertoires showing bias towards use of 3' V and 5' J rearrangements (**Fig2**-**4**). During TCRβ gene rearrangement, these segments are closest to one another in the looping structure that enables V to DJ rearrangements (Skok et al., 2007), suggesting this might account for the foetal bias in proportional TRBVxTRBJ segment usage. In contrast to TCRβ locus rearrangement, the TCRα locus can undergo multiple rounds of rearrangement on each chromosome, sequentially moving 3' to 5' along the series of TRAV gene segments, while moving 5' to 3' along the series of TRAJ gene segments, with proximal pairs (3' TRAV segments with 5' TRAJ segments) initiating this sequence of rearrangements(Carico et al., 2017). We observed that the foetal DP repertoire showed TRAVxTRAJ bias by their chromosomal position. One possible explanation for this bias is that in

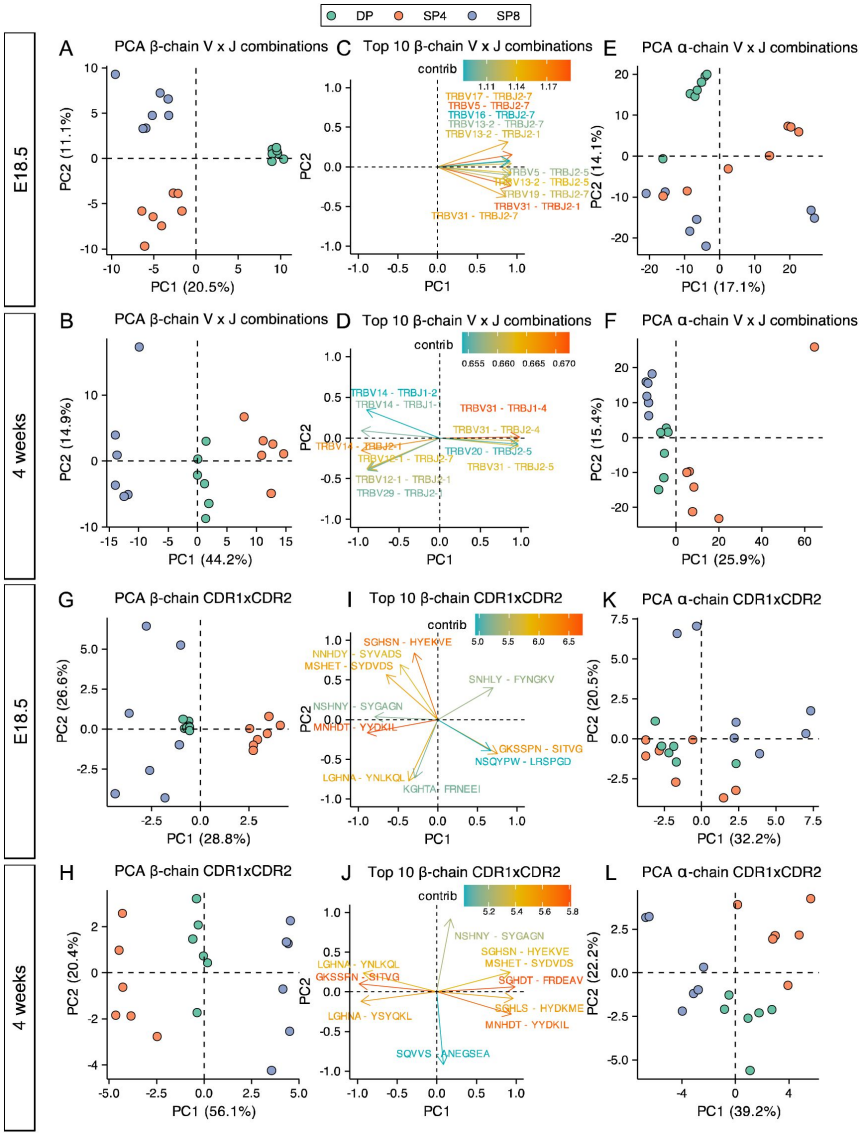


Figure 7.

PCA of VxJ and CDR1xCDR2 combinations in foetal and young adult thymocyte populations indicate reduced MHC-restriction of foetal TCR repertoires

TCR β-chain and α-chain repertoires were sequenced from FACS-sorted $CD4^+CD8^+$ (DP), $CD4^+CD8^-CD3^+$ (SP4) and $CD4^-CD8^+CD3^+$ (SP8) thymocyte populations from E18.5 (n=7) and 4 week-old (n=6) C57BL/6 thymus.
(A-B) PCA biplot of β-chain VxJ gene counts of unique TCRβ sequences for E18.5 **(A)** and 4 week **(B)** in DP, SP4 and SP8 populations (coloured in green, orange and purple respectively). The percentage of variance for each principal component is indicated on the axis. **(C)** Ten β-chain VxJ gene combinations that contribute most to PC1 and PC2 of the PCA for E18.5. **(D)** Ten β-chain VxJ gene combinations that contribute most to PC1 and PC2 of the PCA for 4 weeks. **(E-F)** PCA biplot of α-chain VxJ gene counts of unique TCRβ sequences for E18.5 **(E)** and 4 weeks **(F)** in DP, SP4 and SP8 populations (coloured in green, orange and purple respectively). The percentage of variance for each principal component is indicated on the axis. **(G-H)** PCA biplot of β-chain CDR1xCDR2 frequency distributions of unique TCRβ sequences for E18.5 **(G)** and 4 weeks **(H)** in DP, SP4 and SP8 populations (coloured in green, orange and purple respectively). The percentage of variance for each principal component is indicated on the axis. **(I)** Ten β-chain CDR1xCDR2 combinations that contribute most to PC1 and PC2 of the PCA of β-chain CDR1xCDR2 frequency distributions for E18.5. **(J)** Ten β-chain CDR1xCDR2 combinations that contribute most to PC1 and PC2 of the PCA of β-chain CDR1xCDR2 frequency distributions for 4 weeks. **(K-L)** PCA biplot of α-chain CDR1xCDR2 frequency distributions of unique TCRβ sequences for E18.5 **(K)** and 4 weeks **(L)** in DP, SP4 and SP8 populations (coloured in green, orange and purple respectively). The percentage of variance for each principal component is indicated on the axis.

the foetus progressive rounds of TCRα rearrangement are less common than in young adult, perhaps as a result of differences in the kinetics of T-cell development. Differentiation of foetal thymocytes occurs in a largely synchronized wave, whereas adult thymocyte production is unsynchronised, continuous and slower. Once the adult thymus has reached steady-state, homeostasis between thymocyte populations regulates the rate of differentiation to DP cell (Hager-Theodorides et al., 2007, Outram et al., 2009). To examine the influence of the rate of differentiation on VJ gene usage, we synchronized the differentiation of adult DP thymocytes by treating young adult mice with hydrocortisone (HC) to deplete the adult thymus of all but the most mature cells. At 2 days after treatment, the thymus was depleted of DP cells. Four days later (6 days after treatment), we FACS-sorted and TCR sequenced the replenished $CD3^{-/lo}$DP ($CD3^{-/lo}$CD4+CD8+), and $CD3^{+/hi}$DP ($CD3^{+/hi}$CD4+CD8+) populations. We plotted heatmaps of mean proportional V and J gene segment usage for unique TCR sequences in foetal, HC-treated and control young adult repertoires, clustering each life-stage and thymocyte population, but showing each gene segment in chromosomal order (**Fig8A-D**). For TRAJ usage, foetal and adult HC-treated repertoires clustered together, with untreated adult repertoires clustering separately (**Fig8A-B**). Interestingly, both foetal and HC-treated young adult populations favoured 5' TRAJ segments, in contrast to control young adult populations, that displayed bias for 3' TRAJ (**Fig8A-B**). We next compared usage of all possible VxJ combinations between control and HC-treated young adult DP populations for all TCRα sequences, identifying combinations that showed a change in proportional usage in the HC-treated young adult compared to control young adult (**Fig4A-C**). The DP HC-treated young adult population favoured VxJ combinations from the 3' location of TRAV and 5' location of TRAJ along with clear proportional decrease in usage of combinations of 5' TRAV and 3' TRAJ in comparison to young adult control (**Fig8C**). Therefore, by synchronizing young adult thymus using HC, foetal and young adult gene segment usage became more alike, with 5' TRAV and 3' TRAJ bias (**Fig4A** and **Fig8C-F**).

In the case of the β-chain, comparison of proportional usage of all possible unique TCRβ VxJ combinations between HC-treated and control young adult DP populations showed a 5' bias in TRBJ gene segment usage (**Fig8F**), with several VxJ combinations that were proportionally increased or decreased in common with the comparison between foetal and young adult combinatorial gene usage (**Fig4A** and **Fig8F**). These data indicate that when the adult young adult DP population arises quickly and synchronously its TRAV, TRAJ and TRBJ segment usage and TRAVxTRAJ combinations closely resemble the foetal DP repertoire, suggesting that the kinetics of differentiation are important in determining gene segment usage.

Summary

As the young adult thymus recovers from hydrocortisone depletion, combinatorial TCRα VxJ usage becomes more foetal-like, suggesting that the synchronicity and kinetics of foetal T-cell development influence VxJ combination.

Discussion

Here we employed a bulk population-based strategy to investigate the TCRβ and TCRα chain repertoires from developmentally defined thymocyte populations during T-cell development. Our study revealed many differences between the foetal and young adult thymus in TCR gene segment usage, repertoire diversity, and clonality by bulk sequencing separate TCRβ and TCRα repertoires. These data indicate that MHC restriction and positive selection have a weaker impact on the TCR repertoire in foetal thymus compared to adult. Overall, the differences between the foetal and adult thymus TCR repertoires are consistent with the foetal thymus producing αβT-cells with properties and functions that are distinct from adult T-cells: their repertoire is less diverse, more closely encoded by genomic sequence, less governed by MHC-restriction, and with preference for particular gene segment usage. Several recent studies have demonstrated distinct and more

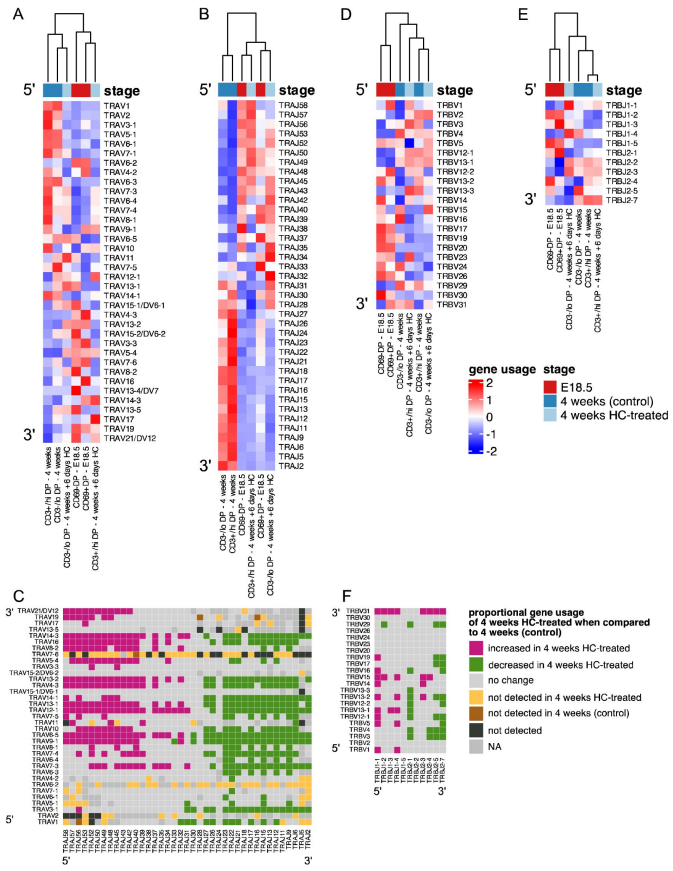


Figure 8.

In hydrocortisone-depleted adult thymus recovering DP populations use foetal-like gene segments

TCR α-chain (A-C) and β-chain (D-F) repertoires were sequenced from FACS-sorted $CD4^+CD8^+CD69^-$ (CD69-DP, foetal), $CD4^+CD8^+CD69^+$ ($CD69^+$DP foetal) $CD4^+CD8^+CD3^{-/lo}$ ($CD3^{-/lo}$DP, adult), $CD4^+CD8^+CD3^{+/hi}$ ($CD3^{+/hi}$DP adult) thymocyte populations from E18.5 (red; n=7), 4 week-old (blue; n=6), and 4 week-old treated with hydrocortisone on day 6 after treatment (light blue; n=5) C57BL/6 thymus.

(A-B) Heatmaps of proportional α-chain variable (V) **(A)** and α-chain joining (J) **(B)** gene usage of productive TCRs for E18.5, 4 weeks (control) and 4 weeks treated with hydrocortisone in DP populations. Each column represents a mean of 7 or 6 (E18.5 and 4 weeks) or 5 (4 weeks treated with hydrocortisone) embryos or mice and was clustered using Euclidian distance. Genes are shown in chromosomal order (5′ to 3′) from top to bottom. **(C)** Plot shows α-chain proportional VxJ gene usage of total TCRs for 4 weeks HC-treated compared to 4 weeks (control) in DP populations. Pink tiles signify increased usage of VxJ combinations in 4 weeks HC-treated ($p < 0.05$), while green tiles signify decreased usage in 4 weeks HC-treated ($p < 0.05$) compared to 4 weeks (control). Grey tiles signify no change in gene usage ($p > 0.05$), yellow tiles signify VxJ combinations not detected in 4 weeks HC-treated, brown tiles signify VxJ combinations not detected in 4 weeks (control) and black tiles signify VxJ combinations not detected in both 4 weeks HC-treated and 4 weeks (control). Only combinations that were detected in at least 3 mice per group were compared. Dark grey tiles signify combinations that were not compared. Statistical comparisons were carried out by unpaired Student's t-test followed by FDR-adjustment (5%, Benjamini-Hochberg procedure) of p values.

(D-E) Heatmaps of proportional β-chain V**(D)** and β-chain J **(E)** gene usage of unique TCRs for E18.5, 4 weeks (control) and 4 weeks treated with hydrocortisone in DP populations. Each column represents a mean of 7 or 6 (E18.5 and 4 weeks) or 5 (4 weeks treated with hydrocortisone) embryos or mice and was clustered using Euclidian distance. Genes are shown in chromosomal order (5′ to 3′) from top to bottom. **(F)** Plot shows β-chain proportional VxJ gene usage of total TCRs for 4 weeks HC-treated compared to 4 weeks (control) in DP populations. Pink tiles signify increased usage of VxJ combinations in 4 weeks HC-treated ($p < 0.05$), while green tiles signify decreased usage in 4 weeks HC-treated ($p < 0.05$) compared to 4 weeks (control). Grey tiles signify no significant change in gene usage ($p > 0.05$), yellow tiles signify VxJ combinations not detected in 4 weeks HC-treated, brown tiles signify VxJ combinations not detected in 4 weeks (control) and black tiles signify VxJ combinations not detected in both 4 weeks HC-treated and 4 weeks (control). Statistical comparisons were carried out by unpaired Student's t-test followed by FDR-adjustment (5%, Benjamini-Hochberg procedure) of p values.

innate-like features of foetal and neonatal T-cell populations in both humans and mice (Beaudin et al., 2016, Wang et al., 2016, Rackaityte and Halkias, 2020, Smith et al., 2018). The first wave of foetal αβT-cells that leave the thymus must provide early protection against infection in the neonatal animal but also need to be tolerant to both self and maternal MHC/antigens. It is possible that the distinct features and potentially weaker MHC-restriction of the foetal TCR repertoire have evolved to provide these two important features of foetal adaptive immunity. It will be interesting to investigate if the foetal-bias in gene segment usage confers specific properties on foetal TCRs, with respect to their binding specificity, affinity and avidity and immunity to neonatal infection (Thomas and Crawford, 2019).

In validation of our approach, we were able to quantify known differences in TCRβ repertoire diversity, CDR3 length and non-template insertion length between foetal and adult(Sethna et al., 2017). The foetal TCRβ and TCRα repertoires contained fewer non-template nucleotides than adult, indicating that they were more closely aligned to genomic sequence. Consistent with this, the foetal β-chain and α-chain CDR3 repertoires shared more sequences with one another than their young adult counterparts. Our study also showed that the foetal TCRβ repertoire had a less equal distribution in DP, SP4 and SP8 populations compared to adult, and increased clonal expansion of the most abundant clones in DP and SP4 populations. Thus, the reduction in TCRβ repertoire diversity observed in the foetal thymus could not simply be attributed to the reduction in N-nucleotide insertions. These differences in TCRβ repertoire distribution and clonal expansion may be the result of differences in the kinetics of expansion and differentiation of thymocyte subsets in the foetus, leading to greater clonal expansion of TCRβ clones following β-selection. The transition from DP to SP cell in the foetus takes just 2-3 days at most, as DP cells first appear around E16, and the mature SP populations were FACS-sorted on E18.5, whereas in the adult thymus differentiation from DN3 to SP cell may take place over a longer period of time (Ross et al., 2014, Solanki et al., 2020).

We found many differences in proportional TRBV, TRBJ, TRAV, and TRAJ gene usage between foetal and adult thymocyte populations, with foetal populations showing proportional bias towards 3' TRBV usage. We observed a particularly striking bias towards use of 3' TRAV and 5' TRAJ gene segments in the foetus. As a previous study used a PCR-based approach to identify enrichment of 3'TRAV to 5'TRAJ rearrangements in foetal BALB/c thymocytes(Pasqual et al., 2002), the bias is not dependent on mouse strain or MHC haplotype

Bias in TCRα gene segment usage by chromosomal location has been reported in human DN thymocyte populations by single cell RNAseq but was found to diminish as thymocytes mature and progressive recombination of the TCRα loci occurs(Park et al., 2020). In contrast, in the mouse foetal thymus, we found that the bias in proportional TRAV and TRAJ gene segment use by chromosomal position persisted in the positively selecting $CD69^{+}$DP and mature SP4 populations, but was not apparent in the adult $CD3^{-/lo}$DP population. Our data therefore suggest that progressive rounds of TCRα rearrangement are less common in the foetal DP population. Indeed, when we depleted the young adult thymus by HC-treatment, we found an increase in 3'TRAV and 5'TRAJ gene segment usage in the synchronized recently differentiated recovering DP population, making these repertoires more similar to the foetal DP TCRα repertoires. Thus differences between foetal and control young adult TCRαV-J rearrangements in DP cells may be the result of slower differentiation in the adult thymus, allowing more time for multiple rounds of TCRαV-J rearrangements in the DP population. Consistent with this, the foetal DP population showed a different TCRα frequency distribution than the adult DP population, with a lower power law exponent.

Comparison of proportional use of TCRβVxJ and TCRαVxJ combinations between adult and foetal populations also highlighted the 3' V gene segment preference in the foetal DP population, and the 5' V gene segment preference in the adult DP population for both β-chain and α-chain.

Interestingly, life-stage had a greater impact on TCRβ and TCRα gene segment usage than MHC-restriction in all populations.

When we carried out PCA on adult and foetal samples separately, PCA clustered β-chain VxJ combination counts and β-chain CDR1xCDR2 by SP population in both adult and foetal samples, showing the impact of MHC restriction (positive selection) on the SP repertoires. However, the percentage of variability that was attributable to MHC-restriction was greater in adult repertoires than foetal repertoires. PCA also clustered the adult α-chain V x J and CDR1xCDR2 combinations by SP cell type, confirming impact of MHC restriction on adult Vα gene usage. However, foetal populations failed to segregate by cell type in either α-chain PCA, indicating that MHC restriction/positive selection has less influence on α-chain repertoires in foetal SP populations. Thus, our data suggests that in the foetal thymus the influence of MHC-restriction/positive selection on both the TCRβ and TCRα repertoires that is weaker than in the adult thymus.

Our approach of bulk TCR sequencing from FACS-sorted thymocyte populations had the advantages of scale (allowing the sequencing of thousands of rearrangements from thousands of cells from each embryo or mouse), at relatively low cost. However, it had the disadvantage that we were unable to investigate which TCRβ rearrangements pair with which TCRα rearrangements in single cells. Clearly, in future it will be important and interesting to expand this analysis to include TCRβ/α pairing by single cell RNAseq.

Materials and methods

Mice

C57/BL6 mice were bred and maintained in specific pathogen-free conditions at University College London (UK) under UK Home Office regulations, following ethical review. Mixed sex of foetal (E18.5) and young adult (4 weeks old) mice were used. For hydrocortisone (HC) treatment, mice were injected intraperitoneally with 0.6mg/gram of body weight pure HC sodium phosphate (Sigma Aldrich, cat. no: BP188) in sterile PBS, and analysed after 6 days.

Fluorescence activated cell sorting

Each foetal or young adult thymus was disaggregated into a single cell solution using the back of a syringe on a 70-μn cell strainer. The cells were washed through the filter into falcon tubes using ice cold FACs buffer (1x AIM-V medium (research grade), AlbuMAX Supplement (Gibco, cat. no: 31035025)) to eliminate any large clumps. Cells were then counted using Accuri C6 flow cytometer and then pelleted by centrifugation for 5 minutes at 1400 rpm and supernatant was removed. The whole cell pellet from the entire organ was stained in FACS buffer with a panel of directly conjugated antibodies supplied by Biolegend (San Diego, US) or eBioscience (San Diego, US) (Supplementary Table 1 and 2) on ice for 30 minutes as described(Lau et al., 2021). To obtain $CD3^{-/lo}$DP ($CD4^+CD8^+CD3^{-/lo}$), $CD69^-$DP ($CD4^+CD8^+CD69^-$), $CD69^+$DP ($CD4^+CD8^+CD69^+$), $CD3^{+/hi}$ DP ($CD4+CD8+CD3^{+/hi}$), SP4 ($CD4^+CD8^-CD3^+$), SP8 ($CD4^-CD8^+CD3^+$) thymocyte cell suspensions were sorted using a BD FACS Aria III. CD69 expression was used to differentiate DP populations in the foetus, as anti-CD3 staining is dull in DP thymocyte populations at E18.5(Solanki et al., 2018). Only live cells were gated and collected using forward and side scatter. Gating strategies to sort the required cell populations are shown in **SFig7** (for foetal thymus) and **SFig8** (for adult thymus).

The cells were collected in 100 μL (foetus) or 400 μL (young adult) of FACS buffer and were then pelleted by centrifugation for 25 minutes at 1400 rpm at 4 °C. The supernatant was removed and 100 μL of Extraction buffer from PicoPure RNA isolation kit (Applied Biosystems, cat. no. KIT0204) was added. Finally, the samples were incubated for 30 minutes at 42 °C on a shaking heatblock and then stored at -80 °C until the RNA was ready to be extracted.

RNA extraction

RNA was extracted by PicoPure RNA isolation kit (Applied Biosystems, cat. no. KIT0204) according to manufacturer's instructions and eluted in 11 µL of elution buffer. RNA concentration was assessed using a spectrophotometer.

TCR amplification and sequencing

We used the protocol for TCR amplification and sequencing as described (Oakes et al., 2017, Uddin et al., 2019), with some minor modifications for mouse samples. See Supplementary Table 3 for the primer list with the sequences and Supplementary Table 4 for a full list of reagents.

RNA was first DNase treated. A maximum of 500 ng of RNA was mixed with 1 µL RQ1 10x Buffer (Promega, cat. no: M6101) and 1 µL RQ1 DNase (1 U/µL, Promega, cat. no: M6101) and then incubated at 37 °C for 30 minutes. 1 µL of RQ1 DNase Stop Solution (Promega, cat. no: M6101) was then added and the mix was incubated at 65 °C to stop the reaction. The RNA was then reverse transcribed by adding 1.5 µL of 10 µM TRAC2 primer, 1.5 µL of 10 µM TRBC3 primer, and 1.5 µL of 10 mM dNTPs (Promega, cat. no: U1515) and incubating at 65 °C for 5 minutes. The reaction was then rapid cooled down on ice and 6 µL of 5x FS Buffer (Invitrogen, cat. no: 18080044), 1.5 µL of 0.1 M DTT (Invitrogen, cat. no: 18080044), 1.5 µL of RNasin Ribonuclease Inhibitor (40 U/µL, Promega, cat. no: N2111) and 1.5 µL of SuperScript III RT (200 U/µL, Invitrogen, cat. no: 18080044) was added and it was incubated at 55 °C for 30 minutes, followed by 70 °C for 15 minutes. These reactions were purified by Minelute PCR purification (Qiagen, cat. no: 28004) following the manufacturer's instructions and eluted in 10.5 µL nuclease-free water.

The cDNA was then ligated to the M13 ligation primer which is composed of the Illumina SP2 primer, an 8 base spacer and a 12 base Unique Molecular Identifier (UMI). This UMI consists of two random hexamers separated by the 8 base spacer. On the 5' end the primer is phosphorylated, and on the 3' end it is blocked with a Spacer C3 moiety to prevent primer concatemerization. The ligation mix consisted of: 10 µL cDNA, 3 µL of bovine serum albumin (BSA)/hexamine cobalt chloride (HCC) mixture (1 mg/mL BSA, 10 mM HCC), 3 µL of T4 RNA ligase buffer (NEB, cat. no: M0204S), 1 µL of 10 mM ATP (NEB, cat. no: M0204S), 1 µL of 10 µM M13 Ligation primer, 2 µL of T4 RNA Ligase (10 000 U/mL, NEB, M0204S), and 10 µL of PEG 8000 50% (NEB, cat. no: M0204S). The ligation mix was then incubated for at least 18 hours up to a maximum of 23 hours at 16 °C and then heat inactivated at 65 °C for 10 minutes.

70 µL of nuclease-free water was added to the ligation products. This was then purified by adding 50 µL of AMPure XP Beads (Beckman Coulter, cat. no: A63881) to the 100 µL ligation product, with some modifications to manufacturer's instructions: 300 µL of 80% ethanol was used for the washing steps and the product was eluted in 31 µL nuclease-free water. The purified ligation product was then amplified in a PCR reaction consisting of 31 µL purified product, 10 µL of 5x HF Buffer (NEB, cat. no: M0530L), 2.5 µL of 10 µM m-alpha-RC1 primer, 2.5 µL of 10 µM m-beta-RC1 primer, 2.5 µL of 10 µM SP2-M13 primer, 1 µL of 10 mM dNTPs (Promega, cat. no: U1515), and 0.5 µL of Phusion Polymerase (NEB, cat. no:M0530L). The m-alpha-RC1 and m-beta-RC1 primers hybridize to the constant region of the α-chain and β -chain genes in the mouse respectively. The SP2-M13 primer hybridizes to the sequence introduced by the M13 ligation primer earlier during the ligation step. PCR-1 was run using the following conditions: initial denaturation at 98°C for 3 minutes, followed by 4 cycles of denaturation at 98°C for 15 seconds, annealing at 69°C for 30 seconds and extension at 72°C for 40 seconds, followed by a final extension at 72°C for 5 minutes.

The 50 µL PCR product was then purified by adding 40 µL of AMPure XP Beads (Beckman Coulter, cat. no: A63881) with some modifications to manufacturer's instructions: 200 µL of 80% ethanol was used for the washing steps. At this stage the samples were split into two: TCRα and TCRβ, each eluted in 31 µL nuclease-free water.

The purified PCR product was then amplified in a PCR reaction to add indices and Illumina adaptors P5 and P7 and sequencing primer 1 (SP1) sequence. The PCR reaction mix consisted of: 31 µL purified product, 2.5 µL of 1 µM SP1 index (mSP1-6N-I-**X**-αRC1 or mSP1-6N-I-**X**-βRC1, for TCRα and TCRβ respectively), 2.5 µL of 10 µM SP2 index (P7-L**X**), 2.5 µL of 10 µM SP1-P5 primer, 10 µL of 5x HF Buffer (NEB, cat. no: M0530L), 1 µL of 10 mM dNTPs (Promega, cat. no: U1515), and 0.5 µL of Phusion Polymerase (NEB, cat. no: M0530L). PCR-2 was run using the following conditions: initial denaturation at 98°C for 3 minutes, followed by 6 cycles of denaturation at 98°C for 15 seconds, annealing at 69°C for 30 seconds and extension at 72°C for 40 seconds, followed by a final extension at 72°C for 5 minutes. The 50 µL PCR product was then purified by adding 40 µL of AMPure XP Beads (Beckman Coulter, cat. no: A63881) with some modifications to manufacturer's instructions: 200 µL of 80% ethanol was used for the washing steps, and the product was eluted in 30 µL nuclease-free water.

The purified PCR product was then amplified in a qPCR reaction, so that the reaction can be monitored and stopped in real time when the reaction passes the threshold of 0.1 ΔRn to prevent overamplification and minimise the introduction of PCR artifacts. SYBR Green I Nucleic Acid Gel Stain, 10,000x concentrate in DMSO (ThermoFisher Scientific, cat. no: S7567) was diluted firstly 1:100 in DMSO and this stock solution was stored at -20 °C. Prior to the assembly of the PCR mix, this stock solution was diluted a further 1:50 in nuclease-free water.

The PCR reaction consisted of: 28 µL purified product, 10 µL of 5x HF Buffer (NEB, cat. no: M0530L), 5 µL of SYBR Green (1:50), 1.25 µL of 10 mM dNTPs (Promega, cat. no: U1515), 1 µL of Rox (Invitrogen, cat. no: 12223012), 2.5 µL of 10 µM P5 primer, 2.5 µL of 10 µM P7 primer and 0.5 µL of Phusion Polymerase (NEB, cat. no: M0530L). The qPCR was run using the following conditions: initial denaturation at 98°C for 3 minutes, followed by a variable number of cycles of denaturation at 98°C for 15 seconds, annealing at 69°C for 30 seconds and extension at 72°C for 40 seconds, followed by a final extension at 72°C for 5 minutes.

The reaction was stopped when the samples passed the threshold of 0.1 ΔRn. The 50 µL PCR product was then purified by adding 40 µL of AMPure XP Beads (Beckman Coulter, cat. no: A63881) with some modifications to manufacturer's instructions: 80% ethanol was used for the washing steps and the product was eluted in 30 µL nuclease-free water.

The purified products were then quantified using Qubit dsDNA high sensitivity reagents (ThermoFisher Scientific, cat. no: Q32854) and the size was confirmed with the TapeStation System HSD1000 (Agilent, cat. no: 5067-5584, cat. no: 5067-5585). The expected size is around 650 bp for a successful library preparation.

Sequencing

The concentrations acquired by the Qubit and the size determined by the TapeStation, were used to calculate the molarity of the samples in nM. The samples were then diluted to 12 nM and pooled into one 12 nM library containing samples with different indices. 84 samples were prepared per sequencing run (42 α-chain and 42 β -chain libraries). A Vacuum Concentrator Centrifuge was used to pellet the library and remove excess volume, so that 30 µL of the pooled library could then be run on the Pippin Prep System using a Pippin Gel Cassette 1.5% w/v agarose dye free 250 bp-1.5 kb DNA size range (Sage Science, cat. no: CDF1510) to size select for sequences between 350 to 750 bp, following the manufacturer's instructions.

40 µL of library eluted from the Pippin Prep was then quantified using a Qubit and TapeStation System HSD1000 as detailed above. The expected molarity was between 2 nM and 5 nM. The library was then denatured and diluted to 1.2 pM according to standard Illumina's protocols (for 4 nM libraries) and was spiked with 22% of 20 pM PhiX control (Illumina, cat. no: FC-110-3001). The 1.2 pM library was sequenced on the NextSeq using the NextSeq 500/550 Mid Output Kit v2.5 (300 Cycles) (Illumina, cat. no: 20024905) at UCL genomics.

Error correction and outputting

The NextSeq outputs files in the format named binary based call (.bcl) which were converted into FASTQ files using bcl2fastq for downstream processing using a pipeline of scripts described previously(Oakes et al., 2017, Thomas et al., 2013): *Decombinator_v3.1* (available at: *https://github.com/innate2adaptive/Decombinator/*) in Python 2.7. This pipeline identifies the errors introduced by PCR amplification using the UMI attached during the ligation step and thus produces an output of a corrected abundance of the TCR in the sample along with determining V, J and CDR3 regions for each TCR.

Downstream analysis

For further downstream analyses, *R Studio* was used. The *tidyverse* set of packages were used for data manipulation and visualisations, in particular *ggplot2*(Wickham, 2016, Wickham H, 2019). Dotplots show mean±c.i (package *ggpubr*(Kassambara, 2023a)) and statistical comparisons were carried out by unpaired Student's t-test or Welch's t-test as appropriate using the package *rstatix(*Kassambara, 2023b*)*. *Pheatmap* and *ComplexHeatmap* were used to generate Pearson correlation heatmaps(Gu et al., 2016, Kolde, 2019).

TCR abundance distribution

TCR repertoire abundance distributions are typically L-shaped, with a high number of distinct TCRs or clonotypes present only once, and a low number of hyperexpanded clones. Therefore, the TCR abundance (number of copies of each sequence) was plotted against their proportion of the repertoire in a log-log plot. This transformed distribution follows approximately a linear distribution apart from the largest clones and can be fitted to a discrete power law ($f(x) = kx^{-a}$) as seen in previous studies(Oakes et al., 2017, Joshi et al., 2019). The TCR abundance frequency was fitted to a discrete power law using maximum likelihood estimation(Bauke, 2007, Clauset et al., 2009) using the *PoweRLaw* package(Gillespie, 2015). The power law exponents (α) were then plotted and compared.

Diversity indices

Before calculating diversity indices, α and β identified TCRs or CDR3s were subsampled (rarefied) to a number lower than the smallest repertoire using the package *vegan*(Oksanen et al., 2022) in order to correct for differences in diversity due to sample size. The mean diversity indices (Shannon Entropy, Gini Index and Jaccard Index of Similarity) were then calculated from 1000 repeats of this random sampling. The Shannon Entropy was computed using the package *vegan*(Oksanen et al., 2022). Gini index was computed using *ineq* package(Zeileis, 2014). The base R function *dist* was used to calculate the Jaccard Index of Similarity.

Principle Component Analysis

Before PCA was applied, raw counts were first $\log_{10}$ transformed to account for differences in sample size, using a pseudocount of 0.01 for any counts that were not observed. Z-scores were then calculated from these values by subtracting the mean and dividing by the standard deviation. PCA was then computed using the base R function prcomp and factoextra package was used to investigate the results(Kassambara A, 2020).

VxJ combinations

Proportional VxJ gene usage of total TCRs was calculated for each sample and only VxJ combinations that were detected in all samples were compared. Statistical comparisons were carried out by unpaired Student's t-test followed by FDR-adjustment (5%, Benjamini-Hochberg procedure) of p values.

Acknowledgements

This work was funded by the MRC (MR/P000843/1, MR/S037764/1) and BBSRC (BB/T020970/1). JR was supported by a studentship from the BBSRC London Interdisciplinary Biosciences Consortium (1903458). Research at UCL Great Ormond Street Institute of Child Health is supported by the NIHR Biomedical Research Centre at Great Ormond Street Hospital and UCL. The authors declare no competing financial interests.

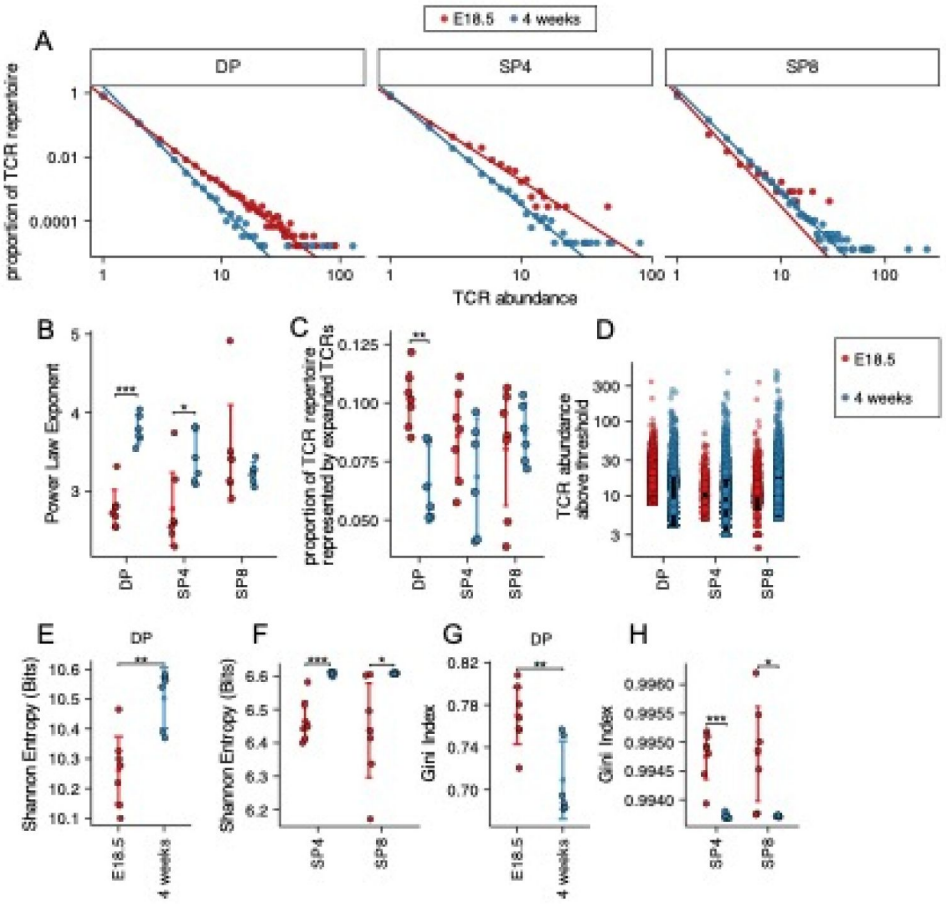
Authors' Contributions

JR, BC and TC conceived the study; JR and TC designed experiments and wrote the manuscript. JR and OAP analysed sequencing data; JR, C-IL, DY and SR performed experiments. C-IL and BC critically reviewed the manuscript.

Data availability

TCR sequencing data will be made available on UCL Research Data Repository at *https://figshare.com/s/6c02ef78423ddd898854*

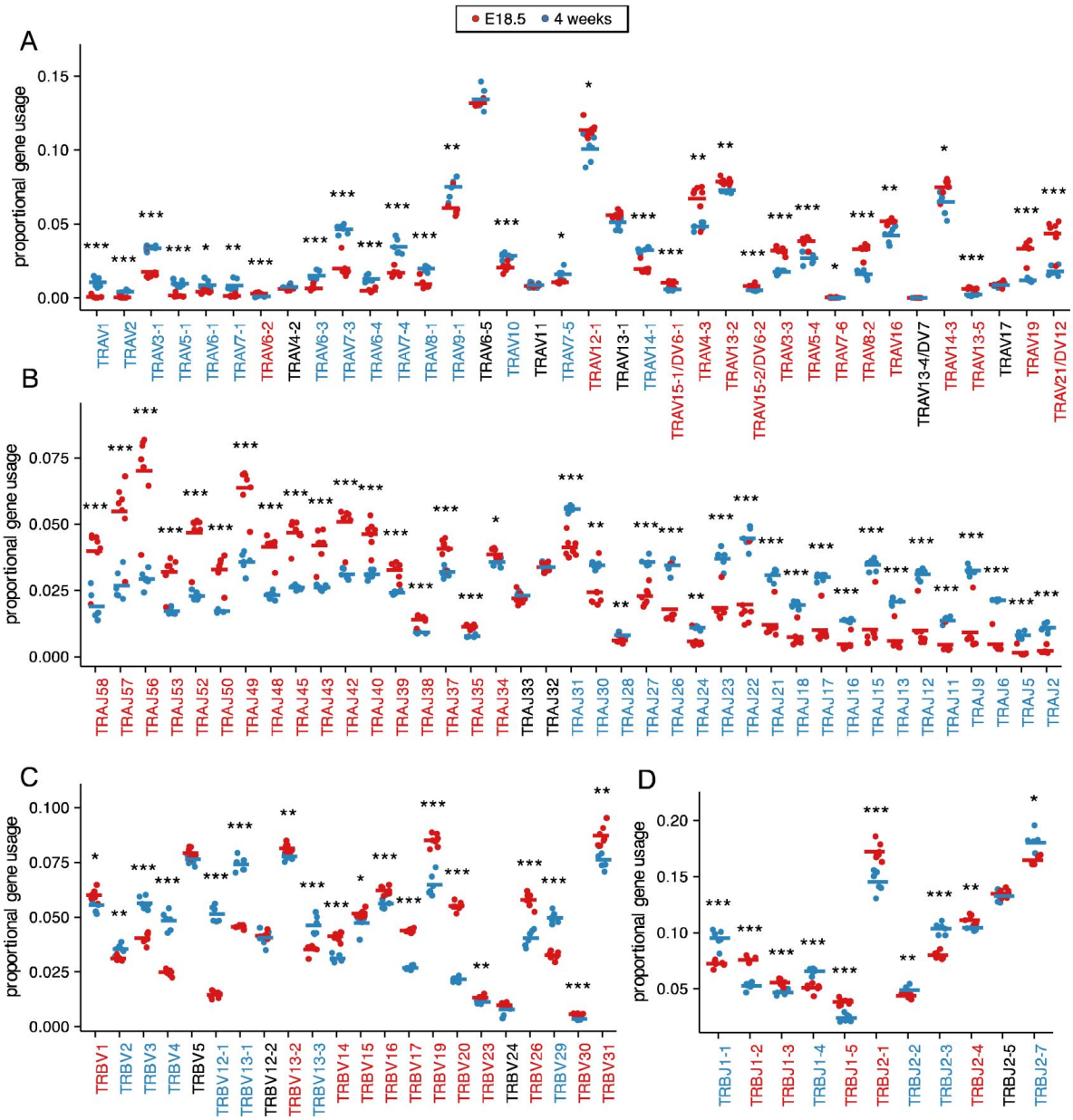
Supplementary figures



Supplementary Figure 1

Distribution and diversity of embryonic and young adult TCRα repertoire

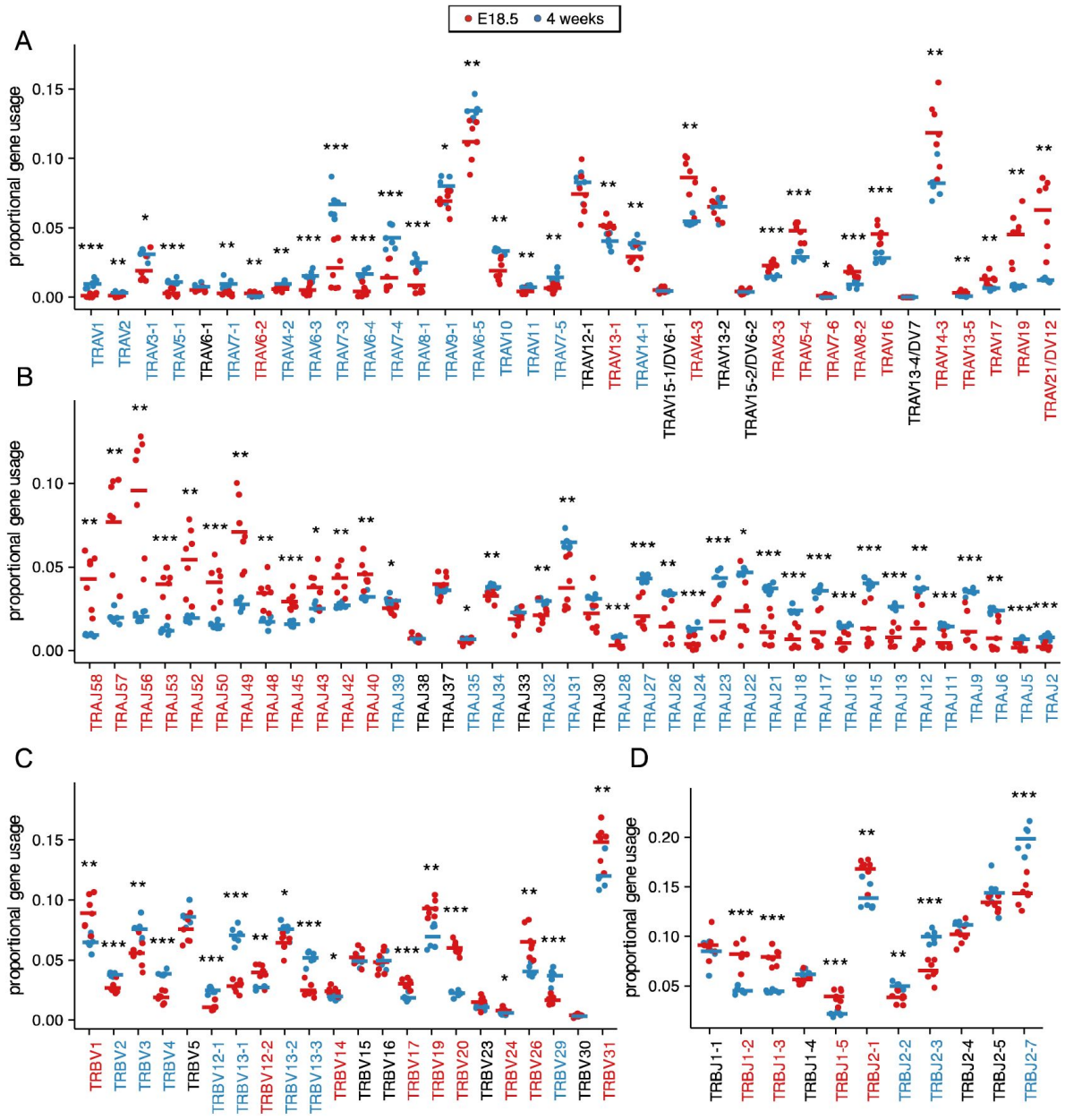
TCR α-chain repertoires were sequenced from FACS-sorted $CD4^+CD8^+$ (DP), $CD4^+CD8^-CD3^+$ (SP4) and $CD4^-CD8^+CD3^+$ (SP8) thymocyte populations from E18.5 (red circles; n=7) and 4 week-old (blue circles; n=6) C57BL/6 thymus. **(A)** The frequency distribution of TCR α-chain abundance was fitted to a discrete power law ($f(k) = Ck^{-\alpha}$) by maximum likelihood. These logged plots show a representative individual mouse TCR repertoire frequency distribution for E18.5 and 4 weeks in DP, SP4 and SP8 populations. The x axis represents TCR abundance (size of clone), and the y axis represents the proportion of the repertoire. The negative of the power law exponent corresponds to the slope of the logged plot. **(B)** The power law exponents of the frequency distribution of TCR α-chain abundances for E18.5 and 4 weeks in DP, SP4 and SP8 populations, where each point represents an individual mouse/embryo. **(C)** The proportion of the total TCR α-chain repertoire accounted for by the expanded top 1% most abundant sequences (>99^{th} percentile) for E18.5 and 4 weeks for DP, SP4 and SP8 populations, where each point represents an individual mouse/embryo. **(D)** The number of α-chain sequences detected above the given frequency threshold (top 1% most abundant sequences) is shown for E18.5 (red circles) and 4 weeks (blue circles) in DP, SP4 and SP8 populations. Each small translucent point represents the abundance of any β-chain sequence detected above the threshold for all mice/embryos at that life-stage, while each larger solid point represents the mean of α-chain sequence abundance for each individual mouse/embryo. **(E)** The Shannon entropies for the α-chain TCR repertoires of E18.5 and 4 weeks in DP populations, where each point represents an individual mouse/embryo. Each repertoire was subsampled to 50000 TCRs 1000 times before calculating the Shannon entropy. **(F)** The Shannon entropies for the α-chain TCR repertoires of E18.5 and 4 weeks in SP4 and SP8 populations, where each point represents an individual mouse/embryo. Each repertoire was subsampled to 750 TCRs 1000 times before calculating the Shannon entropy. **(G)** The Gini indexes for the α-chain TCR repertoires of E18.5 and 4 weeks in DP populations, where each point represents an individual mouse/embryo. Each repertoire was subsampled to 50000 TCRs 1000 times before calculating the Gini index. **(H)** The Gini indexes for the α-chain TCR repertoires of E18.5 and 4 weeks in SP4 and SP8 populations, where each point represents an individual mouse/embryo. Each repertoire was subsampled to 750 TCRs 1000 times before calculating the Gini index.
Dotplots show mean±c.i and statistical comparisons were carried out by unpaired Student's t-test or Welch's t-test as appropriate: *** p <0.001; ** p <0.01; *p < 0.05.



Supplementary Figure 2

Foetal and young adult DP populations have proportionally different TRAV, TRAJ, TRBV and TRBJ gene usage

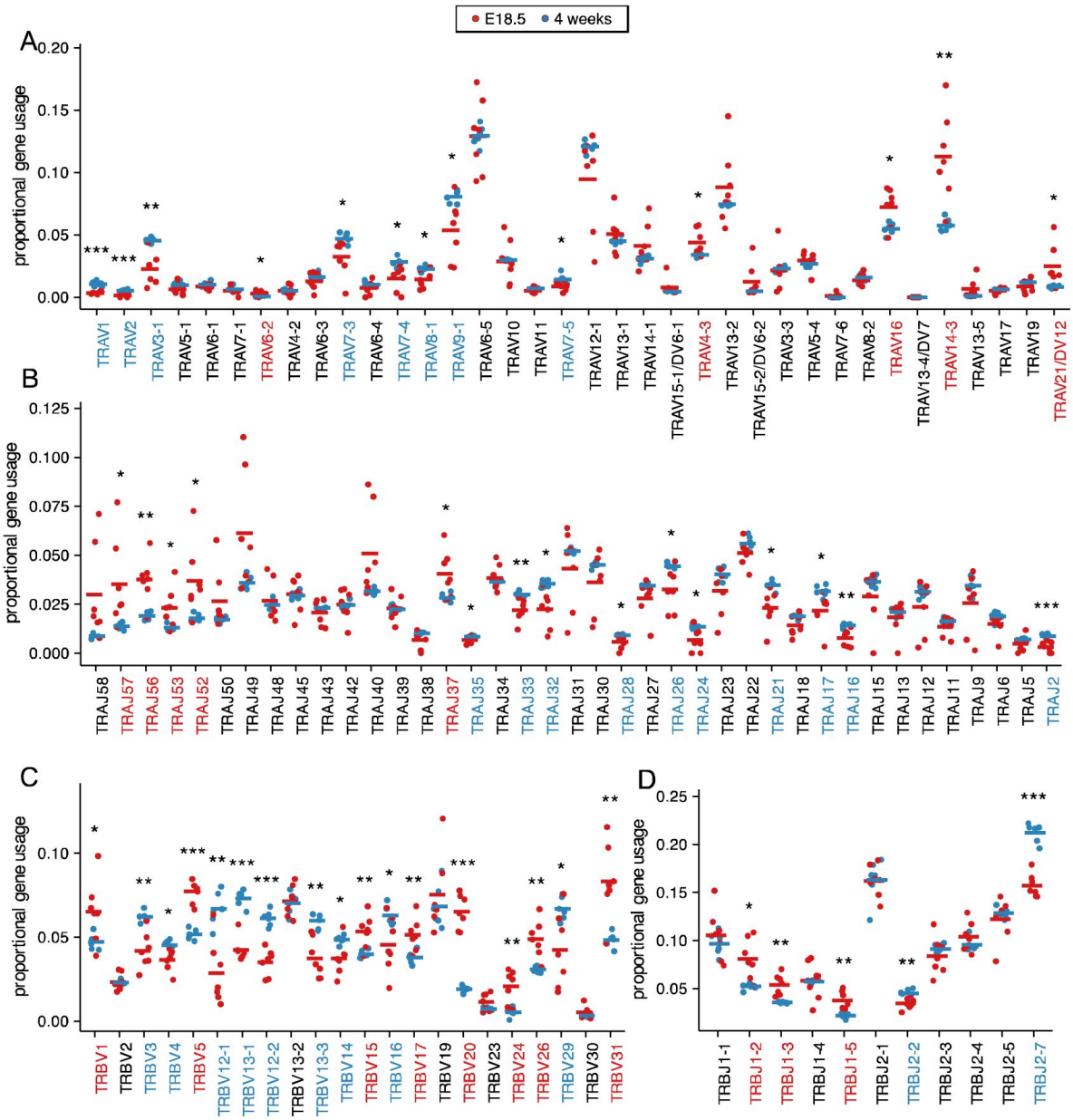
(A-D) Dotplots show proportional TRAV **(A)**, TRAJ **(B)**, TRBV **(C)**, and TRBJ **(D)** gene usage of unique TCRs for E18.5 (red circles, n=7) and 4 weeks (blue circles, n=6) in DP populations. Genes are shown in chromosomal order (5' to 3') from left to right. Dots represent individual mice (n=7 or 6) and bars show the mean. Statistical comparisons were carried out by unpaired Student's t-test or Welch's t-test as appropriate. Genes significantly increased were highlighted in red (when increased in E18.5) or blue (when increased in 4 weeks). *** p <0.001; ** p <0.01; *p < 0.05.



Supplementary Figure 3

Foetal and young adult SP4 populations have proportionally different TRAV, TRAJ, TRBV and TRBJ gene usage

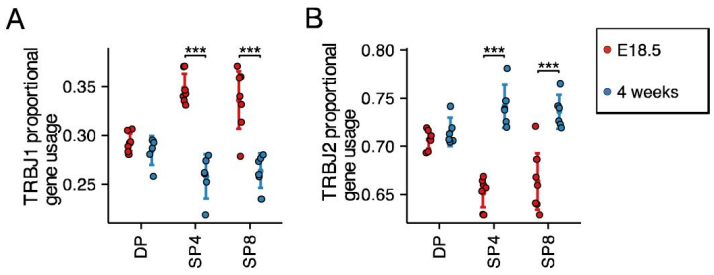
(A-D) Dotplots show proportional TRAV **(A)**, TRAJ **(B)**, TRBV **(C)**, and TRBJ **(D)** gene usage of unique TCRs for E18.5 (red circles, n=7) and 4 weeks (blue circles, n=6) in SP4 populations. Genes are shown in chromosomal order (5' to 3') from left to right. Dots represent individual mice and bars show the mean. Statistical comparisons were carried out by unpaired Student's t-test or Welch's t-test as appropriate. Genes significantly increased were highlighted in red (when increased in E18.5) or blue (when increased in 4 weeks). *** p <0.001; ** p <0.01; *p < 0.05.



Supplementary Figure 4

Foetal and young adult SP8 populations have proportionally different TRAV, TRAJ, TRBV and TRBJ gene usage

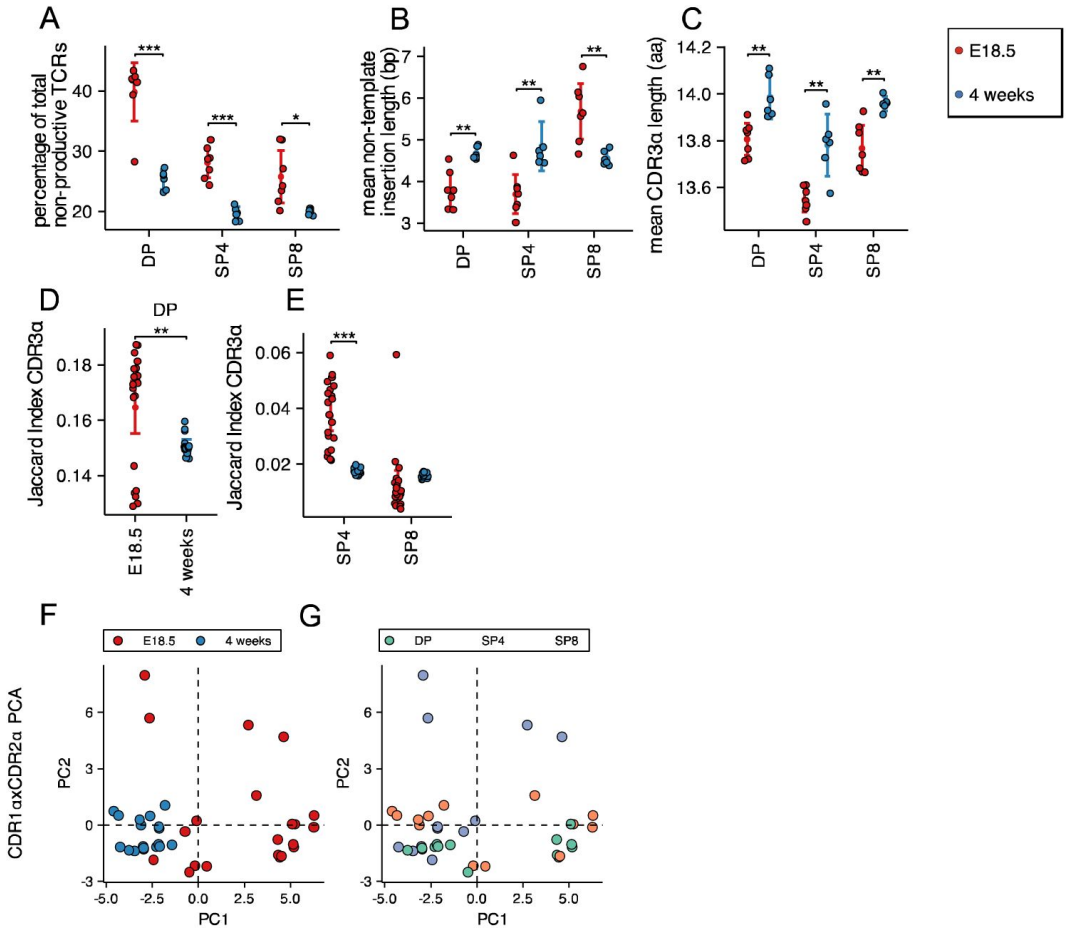
(A-D) Dotplots show proportional TRAV **(A)**, TRAJ **(B)**, TRBV **(C)**, and TRBJ **(D)** gene usage of unique TCRs for E18.5 (red circles, n=7) and 4 weeks (blue circles, n=6) in SP8 populations. Genes are shown in chromosomal order (5' to 3') from left to right. Dots represent individual mice (n=4) and bars show the mean. Statistical comparisons were carried out by unpaired Student's t-test or Welch's t-test as appropriate. Genes significantly increased were highlighted in red (when increased in E18.5) or blue (when increased in 4 weeks). *** p <0.001; ** p <0.01; *p < 0.05.



Supplementary Figure 5

Proportional TRBJ gene segment usage from TRBJ1 and TRBJ2 gene clusters

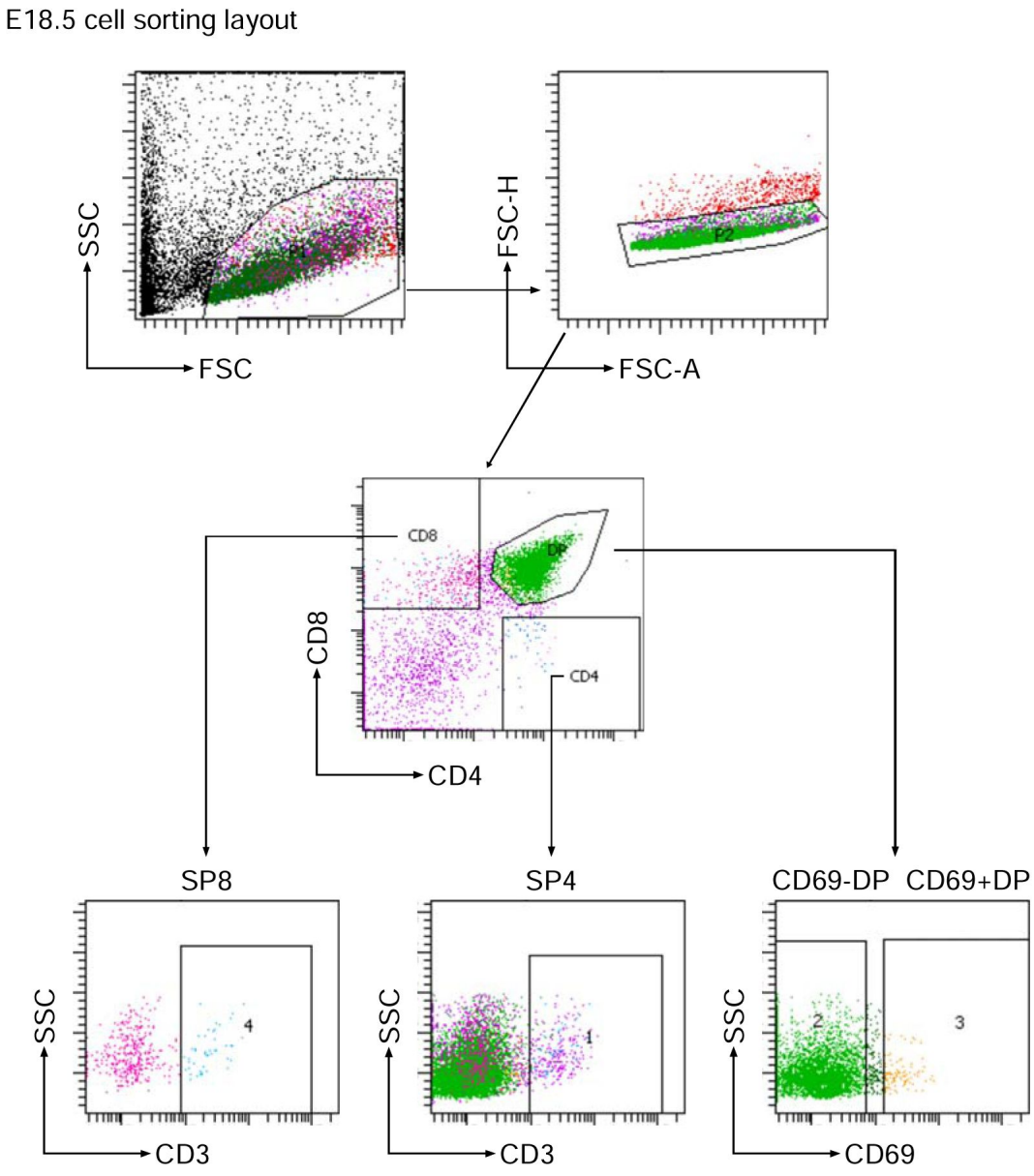
(A-B) Dotplots show proportional TRBJ gene segment usage from TRBJ1 cluster **(A)** and TRBJ2 cluster **(B)**, for E18.5 (red circles, n=7) and 4 weeks (blue circles, n=6) in DP, SP4 and SP8 populations. Statistical comparisons were carried out by unpaired Student's t-test. *** p <0.001; ** p <0.01; *p < 0.05.



Supplementary Figure 6

TCR α CDR3 non-template insertions, length, sharing and PCA of α-chain CDR1xCDR2 frequency distributions in foetal and young adult thymocyte populations

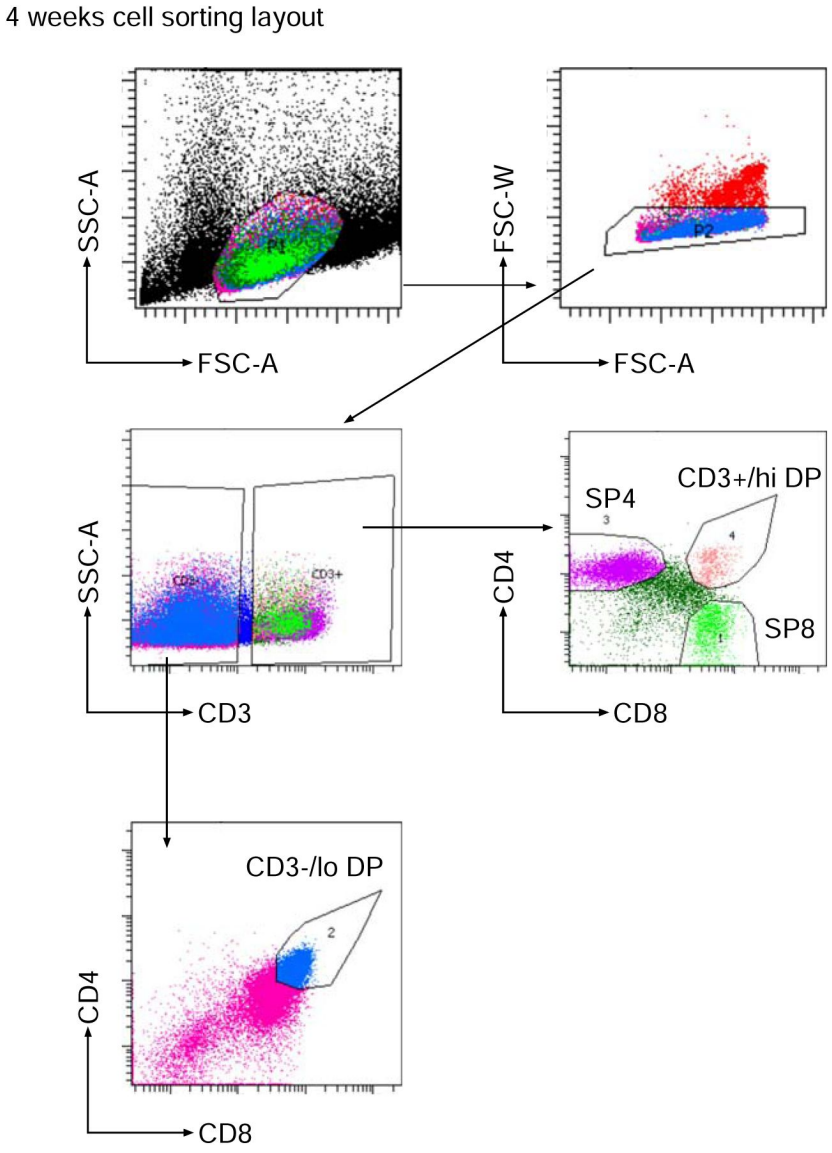
TCR α-chain repertoires were sequenced from FACS-sorted $CD4^+CD8^+$ (DP), $CD4^+CD8^-CD3^+$ (SP4) and $CD4^-CD8^+CD3^+$ (SP8) thymocyte populations from E18.5 (red circles; n=7) and 4 week-old (blue circles; n=6) C57BL/6 thymus.
(A) Percentage of non-productive α-chain TCRs from total rearrangements for E18.5 and 4 weeks in DP, SP4 and SP8 populations. **(B)** The weighted mean of non-template insertion length for all α-chain TCRs (basepairs) for E18.5 and 4 weeks in DP, SP4 and SP8 populations. **(C)** The mean of α-chain unique predicted CDR3 length (number of amino acids) for E18.5 and 4 weeks in DP, SP4 and SP8 populations. **(D)** The intra life-stage Jaccard Index of Similarity of α-chain CDR3s for E18.5 and 4 weeks in the DP population. Each repertoire was subsampled to 30000 CDR3s 1000 times before calculating the Jaccard Index. **(E)** The intra life-stage Jaccard Index of Similarity of α-chain CDR3s for E18.5 and 4 weeks in the SP4 and SP8 population. Each repertoire was subsampled to 500 CDR3s 1000 times before calculating the Jaccard Index.
(F-G) PCA biplot of α-chain CDR1xCDR2 frequency distributions for E18.5 and 4-weeks in DP, SP4 and SP8 populations. In (**F),** E18.5 and 4-weeks samples are coloured in red and blue respectively, while in (**G**), DP, SP4 and SP8 populations are coloured in green, orange and purple respectively. Dotplots show mean±c.i and statistical comparisons were carried out by unpaired Student's t-test or Welch's t-test as appropriate: *** p <0.001; ** p <0.01; *p < 0.05.



Supplementary Figure 7

Gating strategy for cell sorting for foetal thymus.

Figure shows representative FACS plots and gates for cell sorting of foetal thymus.



Supplementary Figure 8

Gating strategy for cell sorting for adult thymus.

Figure shows representative FACS plots and gates for cell sorting of adult thymus.

References

Allam A. H., Charnley M., Pham K., Russell S. M (2021) **Developing T cells form an immunological synapse for passage through the β-selection checkpoint** *J Cell Biol* **220**

Bauke H (2007) **Parameter estimation for power-law distributions by maximum likelihood methods** *The European Physical Journal B* **58**:167–173

Beaudin A. E., Boyer S. W., Perez-Cunningham J., Hernandez G. E., Derderian S. C., Jujjavarapu C., Aaserude E., Mackenzie T., Forsberg E. C (2016) **A Transient Developmental Hematopoietic Stem Cell Gives Rise to Innate-like B and T Cells** *Cell Stem Cell* **19**:768–783

Bogue M., Gilfillan S., Benoist C., Mathis D (1992) **Regulation of N-region diversity in antigen receptors through thymocyte differentiation and thymus ontogeny** *Proc Natl Acad Sci U S A* **89**:11011–5

Britanova O. V. *et al.* (2016) **Dynamics of Individual T Cell Repertoires: From Cord Blood to Centenarians** *J Immunol* **196**:5005–13

Camaglia F., Ryvkin A., Greenstein E., Reich-Zeliger S., Chain B., Mora T., Walczak A. M., Friedman N (2023) **Quantifying changes in the T cell receptor repertoire during thymic development** *Elife* **12**

Carico Z., Krangel M. S (2015) **Chromatin Dynamics and the Development of the TCRalpha and TCRdelta Repertoires** *Adv Immunol* **128**:307–61

Carico Z. M., Roy Choudhury K., Zhang B., Zhuang Y., Krangel M. S. (2017) **Tcrd Rearrangement Redirects a Processive Tcra Recombination Program to Expand the Tcra Repertoire** *Cell Rep* **19**:2157–2173

Ceriani L., Verme P (2012) **The origins of the Gini index: extracts from Variabilità e Mutabilità (1912) by Corrado Gini** *The Journal of Economic Inequality* **10**:421–443

Clauset A., Shalizi C. R., Newman M. E. J (2009) **Power-law distributions in empirical data** *SIAM Review* **51**:661–703

Desanti G. E., Cowan J. E., Baik S., Parnell S. M., White A. J., Penninger J. M., Lane P. J., Jenkinson E. J., Jenkinson W. E., Anderson G (2012) **Developmentally regulated availability of RANKL and CD40 ligand reveals distinct mechanisms of fetal and adult cross-talk in the thymus medulla** *J Immunol* **189**:5519–26

Dutta A., Zhao B., Love P. E (2021) **New insights into TCR beta-selection** *Trends Immunol* **42**:735–750

Genolet R., Stevenson B. J., Farinelli L., Osteras M., Luescher I. F (2012) **Highly diverse TCRalpha chain repertoire of pre-immune CD8(+) T cells reveals new insights in gene recombination** *EMBO J* **31**:4247–8

Gillespie C. S (2015) **Fitting Heavy Tailed Distributions: The poweRlaw Package** *Journal of Statistical Software* **64**:1–16

Gordon J., Manley N. R (2011) **Mechanisms of thymus organogenesis and morphogenesis** *Development* **138**:3865–78

Gu Z., Eils R., Schlesner M (2016) **Complex heatmaps reveal patterns and correlations in multidimensional genomic data** *Bioinformatics* **32**:2847–2849

Hager-Theodorides A. L., Rowbotham N. J., Outram S. V., Dessens J. T., Crompton T (2007) **Beta-selection: abundance of TCRbeta-/gammadelta-CD44-CD25-(DN4) cells in the foetal thymus** *Eur J Immunol* **37**:487–500

Jaccard P (1912) **The Distribution of the Flora in the Alpine Zone** *The New Phytologist* **11**:37–50

Joshi K. *et al.* (2019) **Spatial heterogeneity of the T cell receptor repertoire reflects the mutational landscape in lung cancer** *Nature Medicine* **25**:1549–1559

Kassambara A. (2023) **ggpubr: 'ggplot2' Based Publication Ready Plots** *CRAN*

Kassambara A (2023) **rstatix: Pipe-Friendly Framework for Basic Statistical Tests** *CRAN*

Kassambara A (2020) **factoextra: Extract and Visualize the Results of Multivariate Data Analyses** *CRAN*

Kolde R (2019) **pheatmap: Pretty Heatmaps** *CRAN*

Kondo K., Ohigashi I., Takahama Y (2019) **Thymus machinery for T-cell selection** *Int Immunol* **31**:119–125

Lau C. I., Rowell J., Yanez D. C., Solanki A., Ross S., Ono M., Crompton T (2021) **The pioneer transcription factors Foxa1 and Foxa2 regulate alternative RNA splicing during thymocyte positive selection** *Development* **148**

Mahowald G. K., Mahowald M. A., Moon C., Khor B., Sleckman B. P (2011) **Out-of-frame T cell receptor beta transcripts are eliminated by multiple pathways in vivo** *PLoS One* **6**

Montecino-Rodriguez E., Dorshkind K (2023) **The layered development of mouse B and T Cells** *Immunol Rev* **315**:79–88

Oakes T. *et al.* (2017) **Quantitative Characterization of the T Cell Receptor Repertoire of Naive and Memory Subsets Using an Integrated Experimental and Computational Pipeline Which Is Robust, Economical, and Versatile** *Front Immunol* **8**

Oksanen J. *et al.* (2022) **vegan: Community Ecology Package** *CRAN*

Outram S. V., Hager-Theodorides A. L., Shah D. K., Rowbotham N. J., Drakopoulou E., Ross S. E., Lanske B., Dessens J. T., Crompton T (2009) **Indian hedgehog (Ihh) both promotes and restricts thymocyte differentiation** *Blood* **113**:2217–2228

Park J. E. *et al.* (2020) **A cell atlas of human thymic development defines T cell repertoire formation** *Science* **367**

Pasqual N., Gallagher M. A., Aude-Garcia C., Loiodice M. L., Thuderoz F., Demongeot J., Ceredig R., Marche P. N. L., Jouvin-Marche E. (2002) **Quantitative and Qualitative Changes in V-J α Rearrangements During Mouse Thymocytes Differentiation : Implication For a Limited T Cell Receptor α Chain Repertoire** *Journal of Experimental Medicine* **196**:1163–1174

Pogorelyy M. V. *et al.* (2017) **Persisting fetal clonotypes influence the structure and overlap of adult human T cell receptor repertoires** *PLoS Comput Biol* **13**

Rackaityte E., Halkias J (2020) **Mechanisms of Fetal T Cell Tolerance and Immune Regulation** *Front Immunol* **11**

Ramond C., Berthault C., Burlen-Defranoux O., de Sousa A. P., Guy-Grand D., Vieira P., Pereira P., & Cumano A. (2014) **Two waves of distinct hematopoietic progenitor cells colonize the fetal thymus** *Nat Immunol* **15**:27–35

Ross J. O., Melichar H. J., Au-Yeung B. B., Herzmark P., Weiss A., Robey E. A (2014) **Distinct phases in the positive selection of CD8+ T cells distinguished by intrathymic migration and T-cell receptor signaling patterns** *Proc Natl Acad Sci U S A* **111**:E2550–8

Schatz D. G., Swanson P. C (2011) **V(D)J recombination: mechanisms of initiation** *Annu Rev Genet* **45**:167–202

Sethna Z., Elhanati Y., Dudgeon C. S., Callan C. G., Levine A. J., Mora T., Walczak A. M. (2017) **Insights into immune system development and function from mouse T-cell repertoires** *Proc Natl Acad Sci U S A* **114**:2253–2258

Shannon C. E., Weaver W (1949) **The mathematical theory of communication, Champaign** Il, Us: University of Illinois Press

Sim B. C., Zerva L., Greene M. I., Gascoigne N. R (1996) **Control of MHC restriction by TCR Valpha CDR1 and CDR2** *Science* **273**:963–6

Skok J. A., Gisler R., Novatchkova M., Farmer D., De Laat W., Busslinger M. (2007) **Reversible contraction by looping of the Tcra and Tcrb loci in rearranging thymocytes** *Nat Immunol* **8**:378–87

Smith N. L. *et al.* (2018) **Developmental Origin Governs CD8(+) T Cell Fate Decisions during Infection** *Cell* **174**:117–130

Solanki A., Yanez D. C., Lau C. I., Rowell J., Barbarulo A., Ross S., Sahni H., Crompton T (2020) **The transcriptional repressor Bcl6 promotes pre-TCR-induced thymocyte differentiation and attenuates Notch1 activation** *Development* **147**

Solanki A., Yanez D. C., Ross S., Lau C.-I., Papaioannou E., Li J., Saldaña J. I., Crompton T (2018) **Gli3 in fetal thymic epithelial cells promotes thymocyte positive selection and differentiation by repression of Shh** *Development* **145**

Suo C. *et al.* (2023) **Dandelion uses the single-cell adaptive immune receptor repertoire to explore lymphocyte developmental origins** *Nat Biotechnol*

Thomas N., Heather J., Ndifon W., Shawe-Taylor J., Chain B (2013) **Decombinator: a tool for fast, efficient gene assignment in T-cell receptor sequences using a finite state machine** *Bioinformatics* **29**:542–550

Thomas P. G., Crawford J. C (2019) **Selected before selection: A case for inherent antigen bias in the T cell receptor repertoire** *Curr Opin Syst Biol* **18**:36–43

Uddin I., Joshi K., Oakes T., Heather J. M., Swanton C., Consortium T. R., Chain B (2019) **An Economical, Quantitative, and Robust Protocol for High-Throughput T Cell Receptor Sequencing from Tumor or Blood** *Methods Mol Biol* **1884**:15–42

Wang J., Wissink E. M., Watson N. B., Smith N. L., Grimson A., Rudd B. D (2016) **Fetal and adult progenitors give rise to unique populations of CD8+ T cells** *Blood* **128**:3073–3082

Weischenfeldt J., Damgaard I., Bryder D., Theilgaard-Monch K., Thoren L. A., Nielsen F. C., Jacobsen S. E., Nerlov C., Porse B. T (2008) **NMD is essential for hematopoietic stem and progenitor cells and for eliminating by-products of programmed DNA rearrangements** *Genes Dev* **22**:1381–96

Wickham H. (2016) **ggplot2: Elegant Graphics for Data Analysis** Springer-Verlag New York

Wickham H *et al.* (2019) **Welcome to the tidyverse** *Journal of Open Source Software* **4**

Wong W. K., Leem J., Deane C. M (2019) **Comparative Analysis of the CDR Loops of Antigen Receptors** *Front Immunol* **10**

Xiao S. Y., Li Y., Chen W. F (2003) **Kinetics of thymocyte developmental process in fetal and neonatal mice** *Cell Res* **13**:265–73

Zeileis A (2014) **ineq: Measuring Inequality, Concentration, and Poverty**

Author information

Jasmine Rowell
UCL Great Ormond Street Institute of Child Health, 30 Guilford Street, London, UK
ORCID iD: 0000-0001-7040-8528

Ching-In Lau
UCL Great Ormond Street Institute of Child Health, 30 Guilford Street, London, UK

Susan Ross
UCL Great Ormond Street Institute of Child Health, 30 Guilford Street, London, UK

Diana C Yanez
UCL Great Ormond Street Institute of Child Health, 30 Guilford Street, London, UK

Oscar A Peña
School of Biochemistry University of Bristol, Bristol, United Kingdom

Benny Chain
Division of Infection and Immunity, University College London, London, United Kingdom

Tessa Crompton
UCL Great Ormond Street Institute of Child Health, 30 Guilford Street, London, UK

ORCID iD: 0000-0002-8973-4021

For correspondence: t.crompton@ucl.ac.uk

Editors

Reviewer #1 (Public review):

Summary:

The manuscript by Rowell et al aims to identify differences in TCR recombination and selection between foetal and adult thymus in mice. Authors sequenced the unpaired bulk TCR repertoire in foetal and adult mice thymi and studied both TCRB and TCRa characteristics in the double negative (DN, CD4-CD8-) and single positive (SP4 CD4+CD8- and SP8 CD4-CD8+) populations. They identified age-related differences in TCRa and TCRB segment usage, including a preferential bias toward 3'TRAV and 5' TRAJ rearrangements in foetal cells compared to adults who had a larger perveance for 5'TRAV segments. By depleting the thymocyte population in adult thymi using hydrocortisone, the authors demonstrated that the repertoire became more foetal like, they, therefore, argue that the preferential 5'TRAV rearrangements in adults may be resulting from prolonged/progressive TCRa rearrangements in the adult thymocytes. In line with previous studies, Authors demonstrate that the foetal TCR repertoire was less diverse, less evenly distributed and had fewer non-template insertions while containing more clonal expansions. In addition, the authors claim that changes in V-J usage and CDR1 and CDR2 in the DN vs SP repertoires indicated that positive selection of foetal thymocytes are less dependent on interactions with the MHC.

Strengths:

Overall, the manuscript provides an extensive analysis of the foetal and adult TCR repertoire in the thymus, resulting in new insights in T cell development in foetal and adult thymi.

Weaknesses:

Three major concerns arise:
(1) the authors have analysed TCR repertoires of only 4 foetal and 4 adult mice, considering the high spread the study may have been underpowered.

- The sample size was increased in the revised version

(2) Gating strategies are missing and

- These have now been provided in the revised version

(3) The manuscript is very technical and clearly aimed for a highly specialised audience with expertise in both thymocyte development and TCR analysis. Considering eLife is a scientific journal with a broader readership, Authors are recommended to provide schematics of the TCR rearrangements/their findings and include a summary of conclusions/implications of their findings at the end of each results section rather than waiting till the discussion. This will help the reader to interpret their findings while reading the results.

- These have now been included in the revised version

https://doi.org/10.7554/eLife.93493.2.sa2

Reviewer #2 (Public review):

Summary:

The authors comprehensively assess differences in the TCRB and TCRA repertoires in the fetal and adult mouse thymus by deep sequencing of sorted cell populations. For TCRB and TCRA they observed biased gene segment usage, less diversity, and greater repertoire sharing among individuals in fetal thymocytes. The TCRB repertoire was less evenly distributed and displayed more evidence of clonal expansions in fetal thymocytes. Both fetal and adult thymocytes demonstrated repertoire skewing in CD4 and CD8 as compared to DP thymocytes, which was attributed to MHC-I- vs MHC-II-restriction during positive selection. Effects of MHC-restriction were notably weaker in fetal thymocytes. The authors conclude that in multiple respects fetal repertoires are distinct from adult repertoires.

Strengths:

The analyses of the F18.5 and adult thymic repertoires are comprehensive with respect to the cell populations analyzed and the diversity of statistical approaches used to characterize the repertoires. Because repertoires were analyzed in pre- and post-selection thymocyte subsets, the data allowed assessment of repertoire selection at different developmental stages. Intriguing differences between fetus and adult are identified.

Weaknesses:

Some of the repertoire characteristics reported are already fairly well documented in the literature. Moreover, an unaddressed limitation of the study is that fetal thymocytes were analyzed at single time-point in their development. As a result, at least some of the conclusions about the fetal repertoire may be viewed not as general conclusions, but rather, due to the synchronous development of fetal thymocytes, as pertaining to the one day of fetal/early neonatal development assayed. Statements suggesting that (1) "progressive TCRa rearrangements occur less frequently in foetal DP cells" (Abstract), (2) "One possible explanation for this bias is that in the foetus progressive rounds of TCRa rearrangement are less common than in young adult" (Discussion), and (3) "Overall, the differences between the foetal and adult thymus TCR repertoires are consistent with the foetal thymus producing abT-cells ... with preference for particular gene segment usage" (Discussion), are oversimplified and potentially misleading.

https://doi.org/10.7554/eLife.93493.2.sa1

Author response:

The following is the authors' response to the original reviews.

> ***Public Reviews:***
>
> ***Reviewer #1 (Public Review):***
>
> *Summary:*
>
> *The manuscript by Rowell et al aims to identify differences in TCR recombination and selection between foetal and adult thymus in mice. Authors sequenced the unpaired bulk*

> *TCR repertoire in foetal and adult mice thymi and studied both TCRB and TCRα characteristics in the double positive (DP, CD4+CD8+) and single positive (SP4 CD4+CD8CD3+ and SP8 CD4-CD8+CD3+) populations. They identified age-related differences in TCRα and TCRB segment usage, including a preferential bias toward 3'TRAV and 5' TRAJ rearrangements in foetal cells compared to adults who had a larger perveance for 5'TRAV segments. By depleting the thymocyte population in adult thymi using hydrocortisone, the authors demonstrated that the repertoire became more foetal like, they therefore argue that the preferential 5'TRAV rearrangements in adults may be resulting from prolonged/progressive TCRα rearrangements in the adult thymocytes. In line with previous studies, Authors demonstrate that the foetal TCR repertoire was less diverse, less evenly distributed and had fewer non-template insertions while containing more clonal expansions. In addition, the authors claim that changes in V-J usage and CDR1 and CDR2 in the DP vs SP repertoires indicated that positive selection of foetal thymocytes are less dependent on interactions with the MHC.*
>
> *Strengths:*
>
> *Overall, the manuscript provides an extensive analysis of the foetal and adult TCR repertoire in the thymus, resulting in new insights in T cell development in foetal and adult thymi.*
>
> *Weaknesses:*
>
> *Three major concerns arise:*
>
> *(1) the authors have analysed TCR repertoires of only 4 foetal and 4 adult mice, considering the high spread the study may have been underpowered.*

Given the concerns of the reviewer we have sequenced more libraries and added more data to include repertoires from 7 embryos and 6 young adults (biological replicates from different sorts). We believe that including more replicates has indeed strengthened our study.

Our experimental approach was to sequence TCR transcripts, and in studies using RNA-sequencing of inbred mice, often only 3 individuals (biological replicates) are sequenced.

Our study sequenced from 7 foetal thymuses (generating TCRα and TCRβ repertoires from 4 FACS-sorted cell populations); 6 adult thymuses (generating TCRα and TCRβ repertoires from 4 FACS-sorted cell populations); and 5 adult thymuses from hydrocortisone-treated mice (generating TCRα and TCRβ repertoires from FACS-sorted CD3lo and CD3hi DP populations). We thus analysed 124 distinct repertoires from different populations and libraries, and many tens of thousands of unique sequences.

> *(2) Gating strategies are missing and*

We have included gating strategies for cell-sorting as SFig7 and SFig8.

> *(3) the manuscript is very technical and clearly aimed for a highly specialised audience with expertise in both thymocyte development and TCR analysis. Authors are recommended to provide schematics of the TCR rearrangements/their findings and include a summary conclusions/implications of their findings at the end of each results section rather than waiting till the discussion. This will help the reader to interpret their findings while reading the results.*

We have modified the manuscript to include a more general introductory paragraph (page 3) to introduce the reader to the topic and we have included brief summaries of the findings at the end of each result section (pages 7,9,10,12,13,15).

Reviewer #2 (Public Review):

Summary:

The authors comprehensively assess differences in the TCRB and TCRA repertoires in the fetal and adult mouse thymus by deep sequencing of sorted cell populations. For TCRB and

TCRA they observed biased gene segment usage and less diversity in fetal thymocytes. The TCRB repertoire was less evenly distributed and displayed more evidence of clonal expansions and repertoire sharing among individuals in fetal thymocytes. In both fetal and adult thymocytes they show skewing of V segment (CDR1-2) repertoires in CD4 and CD8 as compared to DP thymocytes, which they attribute to MHC-I vs MHC-II restriction during positive selection. However the authors assess these effects to be weaker in fetal thymocytes, suggesting weaker MHC-restriction. They conclude that in multiple respects fetal repertoires are distinct from and more innate-like than adult.

Strengths:

The analyses of the F18.5 and adult thymic repertoires are comprehensive with respect to the cell populations analyzed and the diversity of approaches used to characterize the repertoires. Because repertoires were analyzed in pre- and post-selection thymocyte subsets, the data offer the potential to assess repertoire selection at different developmental stages. The analysis of repertoire selection in fetal thymocytes may be unique.

Weaknesses:

(1) Problematic experimental design and some lack of familiarity with prior work have resulted in highly problematic interpretations of the data, particularly for TCRA repertoire development.

The authors note fetal but not adult thymocytes to be biased towards usage of 3' V segments and 5'J segments. It should be noted that these basic observations were made 20 years ago using PCR approaches (Pasqual et al., J.Exp.Med. 196:1163 (2002)), and even earlier by others.

We have cited this manuscript (Introduction, page 5) which used PCR of genomic DNA to investigate some TCRα VJ rearrangements in foetal and adult thymus. In contrast, our study uses next generation sequencing of transcripts to investigate all possible combinations of TCRα and TCRβ VJ combinations in different sorted thymocyte populations ex vivo. The greater sensitivity of this more modern technology has thus enabled us to detect many more TCRαVJ rearrangements than the 2002 study, and to conclude on basis of stringent statistical testing that the foetal repertoire is enriched for 3'V to 5'J combinations (Fig. 4).

The authors also note that in fetal thymus this bias persists after positive selection, and it can be reproduced in adults during recovery from hydrocortisone treatment. The authors conclude that there are fewer rounds of sequential TCRA rearrangements in the fetal thymus, perhaps due to less time spent in the DP compartment in fetus versus adult. However, the repertoire difference noted by the authors does not require such an explanation. What the authors are analyzing in the fetus is the leading edge of a synchronous wave of TCRA rearrangements, whereas what they are analyzing in adults is the unsynchronized steady state distribution. It is certainly true, as has been shown previously, that the earliest TCRA rearrangements use 3' TRAV and 5'TRAJ segments. But analysis of adult thymocytes has shown that the progression from use of 3' TRAV and 5' TRAJ to use of 5' TRAV and 3' TRAJ takes several days (Carico et al., Cell Rep. 19:2157

> *(2017)). The same kinetics, imposed on fetal development, would put development of a more complete TCRA repertoire at or shortly after birth. In fact, Pasqual showed exactly this type of progression from F18 through D1 after birth, and could reproduce the progression by placing F16 thymic lobes in FTOC. It is not appropriate to compare a single snapshot of a synchronized process in early fetal thymocytes to the unsynchronized steady state situation in adults. In fact, the authors' own data support this contention, because when they synchronize adult thymocytes by using hydroxycortisone, they can replicate the fetal distribution. Along these lines, the fact that positive selection of fetal thymocytes using 3' TRAV and 5' TRAJ segments occurs within 2 days of thymocyte entry into the DP compartment does not mean that DP development in the fetus is intrinsically rapid and restricted to 2 days. It simply means that thymocytes bearing an early rearranging TCR can be positively selected shortly after TCR expression. The expectation would be that those DP thymocytes that had not undergone early positive selection using a 3' TRAV and a 5' TRAJ would remain longer in the DP compartment and continue the progression of TCRA rearrangements, with the potential for selection several days later using more 5'TRAV and 3'TRAJ.*

We agree with this summary provided by the reviewer which corresponds closely to the points we made ourselves in the manuscript. Indeed, we discuss the synchronization and kinetics of first wave of T-cell development in Results page 13 and Discussion page 17, which was the rationale for the hydrocortisone experiment. We have also discussed findings from Carico et al 2017 in this context (see pages 13, 16, 17).

> *(2) The authors note 3' V and 5'J biases for TCRB in fetal thymocytes. The previously outlined concerns about interpreting TCRA repertoire development do not directly apply here. But it would be appropriate to note that by deep sequencing, Sethna (PNAS 114:2253 (2017)) identified skewed usage of some of the same TRBV gene segments in fetal versus adult. It should also be noted that Sethna did not detect significantly skewed usage of TRBJ segments. Regardless, one might question whether the skewed usage of TRBJ segments detected here should be characterized as relating to chromosomal location. There are two logical ways one can think about chromosomal location of TRBJ segments - one being TRBJ1 cluster vs TRBJ2 cluster, the other being 5' to 3' within each cluster. The variation reported here does not obviously fit either pattern. Is there a statistically significant difference in aggregate use of the two clusters? There is certainly no clear pattern of use 5' to 3' across each cluster.*

We have included a statistical comparison of the aggregate TRBJ use between the J1 cluster and the J2 cluster (see SFig5) and Results page 9.

> *(3) The authors show that biases in TCRA and TCRB V and J gene usage between fetal and adult thymocytes are mostly conserved between pre- and post-selection thymocytes (Fig 2). In striking contrast, TCRA and TCRB combinatorial repertoires show strong biases preselection that are largely erased in post-selection thymocytes (Fig 3). This apparent discrepancy is not addressed, but interpretation is challenging.*

I think the reviewer is referring to heatmaps for individual gene segment usage shown in Figure 2 in comparison to combinatorial usage shown in Figure 4. There is not a discrepancy in the data, but rather the differences between these two figures lie in the way in which the comparisons are made and visualised. The heatmaps in Figure 2A-D show mean proportional usage of each individual gene segment for each cell type in the two life stages, clustered by Euclidian distance. This visualisation clearly shows bias in foetal 3’ TRAV usage and 5’TRAJ usage (looking at areas of red, which have higher usage), with less pronounced enrichment for TRBV and TRBJ. The heatmaps also show differences in intensity between different cell populations in each life-stage.

In contrast, in Figure 4 the tiles show combinations with statistically significant (P<0.05) differences in mean counts for each VJ combination in each cell type between 7 foetal and 6 adult repertoires by Student's t-test, after correcting for False discovery rate (FDR) due to multiple combinations. It is the case, that there are fewer significant differences in proportional combinatorial VxJ use between foetal and adult repertoires after selection. We find this an interesting finding and have expanded our discussion of this aspect of the data (page 10). More than half of the significant differences persist after repertoire selection, and the reduction in each individual SP population, of course in part reflects the lineage divergence.

> *(4) The observation that there is a higher proportion of nonproductive TCRB rearrangements in fetal thymus compared to adult is challenging to interpret, given that the results are based upon RNA sequencing so are unlikely to reflect the ratio in genomic DNA due to processes like NMD.*

We have added two sentences to explain that transcripts of non-productive rearrangements are eliminated by nonsense-mediated decay (NMD), but some non-productive transcripts are detected in many studies of TCR repertoire sequencing, and we have cited three studies from different groups that document this (see Results, page 10-11). We have not commented on how the increase in non-productive TCR rearrangements in the foetal populations (in comparison to adult) relates to rearrangements in genomic DNA or NMD. We have likewise not commented on the possible significance or biological role of nonproductive TCR transcripts, but simply reported our findings.

> *(5) An intriguing and paradoxical finding is that fetal DP, CD4 and CD8 thymocytes all display greater sharing of TCRB CDR3 sequences among individuals than do adults (Fig 5DE), whereas DP and CD8 thymocytes are shown to display greater CDR3 amino acid triplet motif sharing in adults (with a similar trend in CD4).*

As foetal DP, CD4SP and CD8SP TCRbeta repertoires have fewer non-template insertions and lower means CDR3 length, they are expected to share more CDR3 repertoires than their adult counterparts. However, in the case of CDR3 amino acid triplet motifs (k-mers) what is being analysed is the sharing of each possible individual k-mer. If k-mers are shared more in the adult for some populations, but CDR3 repertoires are shared more in the foetus, we think it means that some k-mers appear in many different CDR3 sequences in the adult, so that they are over-represented in multiple different CDR3s (presumably due to selection processes, although we agree that this is just an assumption).

> *The authors attribute high amino acid triplet sharing to the result of selection of recurrent motifs by contact with pMHC during positive selection. But this interpretation seems highly problematic because the difference between fetal and adult thymocytes is dramatic even in unfractionated DP thymocytes, the vast majority of which have not yet undergone positive selection. How then to explain the differences in CDR3 sharing visualized by the different approaches?*

The TCRβ repertoire has been selected in the adult DP population through the process of β-selection, which is believed to involve immune synapse formation and MHC-interactions (Allam et al 2021,10.1083/jcb.201908108). We have now included this reference in the introduction to make this clear (page 4). However, we agree with the reviewer's comments that it is challenging to explain the k-mer analysis and that we have not been able to actually show that increased k-mer sharing in the adult is a direct consequence of increased positive selection: it was our interpretation of this seemingly paradoxical finding. For clarity, we have therefore removed the k-mer analyses from the manuscript.

> *(6) The authors conclude that there is less MHC restriction in fetal thymocytes, based on measures of repertoire divergence from DP to CD4 and CD8 populations (Fig. 6). But the authors point to no evidence of this in analysis of TRBV usage, either by PC or heatmap analyses (A,B,D). The argument seems to rest on PC analysis of TRAV usage (Fig S6), despite the fact that dramatic differences in the SP4 and SP8 repertoires are readily apparent in the fetal thymocyte heatmaps. The data do not appear to be robust enough to provide strong support for the authors' conclusion.*

We have written the text very carefully so as not to make the claim too strong, stating in the abstract: “In foetus we identified less influence of MHC-restriction on α-chain and β-chain combinatorial VxJ usage and CDR1xCDR2 (V region) usage in SP compared to adult, indicating weaker impact of MHC-restriction on the foetal TCR repertoire.” We are not saying that MHC-restriction does not impact VJ gene usage in foetal repertoires, but rather that it has less influence (particularly when compared to life-stage). Evidence for this comes from: [1] Heatmaps in Fig2A-D which show that all repertoires cluster first by life-stage ahead of cell type; [2] Fig3A and B: PCA of adult and foetal TCRβ VXJ combinations: All repertoires cluster by life-stage on PC1. PC2 separates adult repertoires by cell type (adult SP8 are positive on PC2 while adult SP4 are negative on PC2, and DP cells are between them) but for foetal repertoires the SP8 and SP4 are highly dispersed with some SP4 cells falling on positive side of PC2. Only foetal DP repertoires cluster tightly. [3] Fig6A-C: PCA of β−chain CDR1xCDR2 (corresponding to Vβ gene segment usage) again shows the same pattern. Adult repertoires separate by cell type on PC2, (SP8 positive on PC2, SP4 negative on PC2, with DP in between), but foetal SP8 repertoires are much more dispersed. [5] SFig6J-K: PCA of α−chain CDR1xCDR2 (Vα usage) frequency distributions: adult repertoires cluster together and are separated by cell type on PC2 (SP4 positive, SP8 negative), but foetal populations are highly dispersed and fail to cluster by cell type on either axis. [6] We have additionally added new PCA analyses to explore differences in MHC-restriction between foetal and adult SP populations. This is shown in the new Figure 7. We reasoned that in a PCA that included foetal and adult repertoires together, the foetal repertoires might not segregate by SP cell type (MHC-restriction) because of their overall bias towards particular VJ combinations, which would mean that effectively the PCA would be imposing adult MHC restriction on the foetal repertoires. We therefore carried out PCA in which we analysed the adult repertoires separately from the foetal repertoires. As expected for adult repertoires, PCA separated SP4 repertoires from SP8 repertoires on PC1 in each comparison (β-chain VxJ (Fig. 7B), α-chain VxJ (Fig. 7F), β-chain CDR1xCDR2 (V region) (Fig. 7H) and α-chain CDR1xCDR2 (V region) (Fig. 7L)). In contrast, for foetal TCRα repertoires (α-chain VxJ and α-chain CDR1xCDR2 (V region)), PCA failed to separate SP4 from SP8 repertoires on PC1 or PC2, so we did not detect impact of MHC-restriction on foetal TCRβ repertoires (Fig. 7E and K). For foetal TCRβ repertoires, PCA separated SP4 β-chain VxJ from SP8 on PC2, accounting for only 11.1% of variance (Fig. 7A) (in contrast to the 44.2% of variance accounted for by MHC-restriction in adult β-chain VxJ PCA (Fig. 7B)). Thus, in adult repertoires ~4-fold more of the variance in β-chain VxJ usage can be accounted for by MHC-restriction than in foetal repertoires. PCA of foetal β-chain CDR1xCDR2 (V region) separated SP4 from SP8 on PC1, accounting for 28.8% of variance, whereas in PCA of adult β-chain CDR1xCDR2, MHCrestriction accounted for 56.1% (>2-foldmore than in foetus). Thus, even when we considered only V-region usage alone, we detected a stronger influence of MHC-restriction on the TCRβ repertoire in adult compared to foetal thymus.

> ***Reviewer #3 (Public Review):***
>
> *Summary:*
>
> *This study provides a comparison of TCR gene segment usage between foetal and adult thymus.*

> *Strengths:*
>
> *Interesting computational analyses was performed to find interesting differences in TCR gene usage within unpaired TCRa and TCRb chains between foetal and adult thymus.*
>
> *Weaknesses:*
>
> *This study was significantly lacking insight and interpretation into what the data analysed actually means for the biology. The dataset discussed in the paper is from only two experiments. One comparing foetal and adult thymi from 4 mice per group and another which involved hydrocortisone treatment. The paper uses TCR sequencing methodology that sequences each TCR alpha and beta chains in an unpaired way, meaning that the true identity of the TCR heterodimer is lost. This also has the added problem of overestimating clonality, and underestimating diversity.*

We have discussed the limitations and benefits of our approach of sequencing TCRβ and TCRα repertoires separately in the Discussion (page 19). This approach allows the analysis of thousands of sequences from different cell types and different individuals at relatively low cost. We have made no claims in our manuscript about overall diversity or pairing, and given that each chain's gene locus rearranges at a different time point in development, we believe it is of interest to consider the repertoires individually within this context.

> *Limited detail in the methods sections also limits the ability for readers to properly interpret the dataset. What sex of mice were used? Are there any sex differences? What were the animal ethics approvals for the study?*

We have included this information in the Methods (page 19). Both sexes were used and we found no sex differences, although that was not the focus of our study. All animal experimentation in the UK is carried out under UK Home Office Regulations (following ethical review). This is included in the Methods (page 19).

> ***Recommendations for the authors:***
>
> ***Reviewer #1 (Recommendations For The Authors):***
>
> *Major points:*
>
> *- Group sizes are very small (4 foetal and 4 adult mice). Considering the spread in TCR analysis (eg fig 1 B-H, Sup figures 2-4), the study is likely underpowered as it often looks like one mouse prevents or supports a statistical difference. Authors should therefore consider increasing the group size.*

We have sequenced more libraries and included more data, from 7 foetal and 6 young adult animals (biological replicates).

> *- The authors should include a gating strategy for their sorted cells. This is essential to verify the quality of their findings.*

We have added this to the Methods and SFig7 and SFig8.

> *Authors should include a summary sentence at the end of each result section which interprets the main finding. Furthermore, the manuscript would greatly benefit from a schematic figure of their main findings, particularly with regards to the rearrangements and selection differences in foetal and adult thymi.*

We have added a summary sentence to the end of each results section.

> *- Authors should be more careful with their claim that MHC has less of an effect foetal TCR selection. Authors demonstrated that there is a difference in VJ recombination between the foetal and adult TCR repertoire, skewing the foetal TCR repertoire to certain variable and junctional segments. Since both CDR1 and CDR2 are encoded by the variable gene, this is likely to affect their ability to interact with the MHC during positive selection. Have Authors considered whether the selection process is actually a bystander effect of the differences in the rearrangement process? One way to support the authors claim is to demonstrate that mice with an alternative MHC background, have similar foetal/adult gene rearrangements but a different TCR repertoire in the SP populations.*

Time and resources have prevented us from repeating our experiments in another strain of inbred mice. However, we note that a previous PCR study that showed 3'TRAV to 5'TRAJ bias in foetal repertoires was carried out in BALB/c mice (Pasqual JEM 2002). We have added this point to the Discussion (page 17).

> *- (supplementary) tables have not been provided.*

Supplementary Tables were uploaded with the submission. STables 1 and 2 show antibodies used for cell sorts and STable 3 primers used.

> *Moderate points:*
>
> *- The loading plots in Figure 3 onward are visually strong. Authors could consider including an V and J (separate) loading plots for Figure 3 E, F and G to demonstrate preferential V and J usage.*

We have included additional loading plots in Figure 7 for the new PCA we have added (see Fig. 7C, D,I and J).

> *- "the proportion of non-productive rearrangements was higher in the foetal SP8 population than adults (Fig 5A)" Authors should explain how non-productive TCRs end up in SP populations as they need to pass positive and negative selection which both require interactions between the TCR and the MHC.*

As we used RNA sequencing in our study, we did not comment on how the increase in nonproductive TCRbeta rearrangements in the foetal populations (in comparison to adult) relates to rearrangements in genomic DNA or to nonsense-mediated decay (NMD) that is believed to down-regulate transcripts of non-productively rearranged TCR. We have not commented on the possible significance or biological role of non-productive TCR transcripts, but simply reported our findings.

> *- Authors have studied CDR3 sequential amino acid triplets (k-mers). However, CDR3 regions are longer than 3 amino acids in length, hence authors should provide 1) an overview/comparison of the identified k-mers in foetal or adult thymocytes 2) explain how different k-mers relate to each other, eg whether they are expressed in the same TCR. Have authors considered using alternative programs to identify CDR3 motifs that are based on the full CDR3amino acid sequence, eg TCRdist provides motifs and indicated which amino acids are germline encoded or inserted.*

In light of this comment from this reviewer and also comments from Reviewer 2, we have removed the comparison of k-mers from the manuscript. Please see response to point 5 of Reviewer 2.

> *- The term "innate-like" is confusing as it implies that foetal cells are not antigen specific.*
>
> *However, once in the circulation, foetal cells will respond in an antigen-specific manner.*
>
> *Hence authors should use another term.*

We have removed the term “innate-like” from the abstract and the first time we used it in the first paragraph of the Discussion. However, the second time we used the term, we are actually taking it from the manuscript we cited (Beaudin et al 2016) and in this case we left it in. We agree that foetal cells are likely to respond in an antigen-specific manner.

> *- To support their hypothesis in the discussion "However, as TCRd gene segments are nested.... so that 5' TRAV segments are not favoured" can authors confirm that there are indeed less yd T cells in the foetal repertoire?*

We have removed this section from the discussion, because although it is interesting, it is highly speculative, and the manuscript is already quite complicated to interpret.

> *Minor points:*
>
> *- The authors may find the publication by De Greef 2021 PNAS of interest to identify TRBD segments*
>
> *- Authors need to clarify that they mean CDR3-beta in the sentence "The mean predicted CDR3 length.... compared to young adult"*

We have included new data in the manuscript to show that mean CDR3 length is lower in all foetal populations of beta (Fig5C) and alpha (SFig5C) and clarified which we are referring to in the text.

> *- Authors should bring the section "During TCRb gene rearrangement, these segments.... Initiating the sequence of rearrangements" forward and include a schematic." Forward to figure 2 and provide the reader with a visual schematic of the foetal vs adult recombination events.*
>
> *- Discussion: "The first wave of foetal abT-cells that leave the thymus... tolerant to both self and maternal MHC/antigens". Have Authors considered the alternative hypothesis published by Thomas 2019 in Curr Opin System Biol that the observed bias could potentially provide better protection against childhood pathogens?*

We have indeed considered this, as stated in the first paragraph of the Discussion “The first wave of foetal αβT-cells that leave the thymus must provide early protection against infection in the neonatal animal”. We have now cited the Thomas 2019 study.

> *- Discussion: Authors should rephrase the sentence "The transition from DP to SP cell in the foetus.... From DN3 to SP cell may be slower" as it is unclear what the authors mean.*

We have rephrased this (see page 17)

> *- Discussion "TRAV and TRAJ Array" do authors mean "TRAV and TRAJ area"?*

We did indeed mean array (as in series of gene segments) but we have changed the wording for clarity (page 14).

> *- Methods, Fluorescence activated cell sorting: can authors clarify whether they stained, sorted and sequenced the full thymus and /or specify how many cells were included. Can authors also explain why foetal and adult cells were treated differently (eg the volume of master mix)?*
>
> *- Methods Fluorescence activated cell sorting authors should specify what they mean with "mastermix of either 1:50 (foetal thymus) or 1:100 (adult thymus)". Does this mean all antibodies in the foetal mastermix were 1:50 and all antibodies in the adult master mix were 1:100? If so, why were different concentrations used and why were antibodies not individually titrated before use?*

We have clarified the methods and antibodies used are listed with clones in supplementary tables.

> *Figures:*
>
> *- Several figures did not fit on the page and therefore missed the top or side*
>
> *- Figure 1A: missing a label on the Y axis*

This is visible

> *- Figure 2A-D: please indicate the 5' and 3' terminus in each graph. The cell type legend should include two separate colours for the two DP populations.*

We have added 5’ and 3’ labels. The two DP populations are clearly labelled.

> *- Figure 4: please indicate the 5' and 3' terminus in each graph.*

We have added 5’ and 3’ labels.

> *- Figure 5C: y axis should read mean CDR3B length (aa), Figure 5D and E: y axis should read Jaccard Index CDR3B, Figure 5 F and G: y axis should read Jaccard index CDR3B k-mers. Same comment for Sup Fig 5 but then CDR3a.*

We have added these labels for both Figure 5 and Supplementary Figure 6 (was SFig5 previously).

> *- Figure 6C top label should read CDR1B x CDR2B with highest contribution*

We have added this label.

> *- Figure 7: please indicate the 5' and 3' terminus in each graph.*

We have added 5’ and 3’ labels. This is now Figure 8, as we have added new analyses (new Figure 7).

> *- Supplementary Figure 1-4 are missing a colour legend next to the graphs.*

We have added the legends in.

> ***Reviewer #2 (Recommendations For The Authors):***
>
> *(1) The authors need to provide better support for the notion that the fetal thymus produces ab T cells with properties and functions that are distinct from adult T cells. There are several ways they might provide a more meaningful assessment: (1) They could analyze the fetal repertoire at multiple time points. (2) They could compare instead the steady state distributions in early postnatal and adult thymus samples. (3) They could compare the peripheral T cell repertoires in the first week of life versus adult. This last approach would allow them to draw the most impactful conclusion.*

We appreciate these suggestions. Sadly, it is beyond our budget for the current manuscript and beyond the scope of our current study that we believe provides interesting new information.

> *(2) Fig S2D shows TRBJ1-4 in black lettering meant to indicate no significant difference whereas the figure shows use of this gene segment to be elevated in adult. I believe TRBJ1-4 should be in blue lettering.*

This is now coloured correctly.

> *(3) The figure call out on p11 (Fig5I-J) should be H-I.*

This is now corrected.

> *(4) Please indicate in the main text that Jaccard analysis in Fig 5 D-E is for TCRB.*

This is now corrected.

> *(5) The analysis of usage of TCRB CDR1xCDR2 combinations in Fig6D is said to "reflect the bias observed in their TRBV gene usage (Fig 2C)". Isn't it the case that every TRBV gene presents a distinct CDR1xCDR2 combination, meaning that there is no difference between TRBV usage and TRBV CDR1xCDR2 usage? If so, please make this clearer.*

Yes, this is the case, we have made this clearer in the text.

> ***Reviewer #3 (Recommendations For The Authors):***
>
> *In general, although there is lots of interesting analyses that can be done with these large datasets, I feel as though the authors did not fully interpret the real meaning and significance of many of these results. Whilst there were some speculation on why a foetal repertoire might be different to those of adults in the discussion sections, the rationale for each individual analyses was not clearly explained. I would suggest that the rationale and a thorough explanation of each analyses be added to the results section, including a finishing sentence on what it means.*

We have added short summaries to each results section to make the points we are making clearer.

> *The authors did not mention how many cells were sorted for from each thymus for sequencing. Was the cell number normalised between each population? As this might have an influence on various downstream measurements of diversity, evenness and clonality, if there is a sampling issue.*

This is explained in the methods. We used sampling to allow comparisons between repertoires of different sizes, and this is also explained in the methods.

> *The authors should include the cell sorting profiles and example flow cytometry plots, including gating strategies and the post sort purity of each sorted population.*

We have included sorting strategies in the methods (SFig7 and SFig8).

> *I think the manuscript could also be improved if there were some basic characterisation of foetal vs. adult thymus development. How many thymocytes are in a foetal vs adult thymus at the timepoints chosen?*
>
> *I think there were some interesting findings in this paper. Given that overall, the foetal thymus appeared to be less diverse than that of the adult, one question I thought would be interesting to discuss was the overlap between the two repertoires. Is the foetal thymus simply a sub-fraction of the adult repertoire or is it totally distinct with no overlapping sequences?*

Our analyses indicate that the repertoires are actually different. This is evident in Fig4 and in PCA loading plots shown in Fig, 3C and new Fig. 7C, D, I and J.

> *I think that some of the interpretation in the results section may be a bit vague. "When we compaired by thymocyte population, each adult population clustered together, with adult SP4 separating from adult SP8 on PC2 and DP cells scoring in between, suggesting that PC2 might correspond to MHC restriction of the adult populations." - whilst I think I know what the authors mean, I do believe that this could be explained in clearer detail and more explicit. SP4 and SP8 are known to be positively selected in the thymus on distinct MHC class I and MHC class II molecules for example.*

We have tried to clarify the text describing that PCA and additionally added a new Figure (new Fig. &) to compare the influence of MHC-restriction on the TCR repertoire in foetal and adult thymus.

> *In the methods section, the age and sex of mice used were not explained at all. What was used in the experiment? Are there any sex differences?*

Age and sex of mice is given in the methods. We have not detected sex differences.

> *This is a huge omission from the manuscript. In general, I don't believe the methods section has described the analysis in sufficient detail for replication. All analysis code and data should be publicly accessible and be in a format that allows for the reader to replicate the figures in the paper upon running the code. Perhaps even allowing them to run their own TCR datasets. Overall, I think the manuscript needs some rewriting to include additional details and deeper interpretation of each individual analyses.*

Sequencing data files will be made publicly available on UCL Research Data Repository.

https://doi.org/10.7554/eLife.93493.2.sa0