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

Running Title: Foetal αβ TCR repertoire

Authors: Jasmine Rowell[1], Ching-In Lau[1], Susan Ross[1], Diana C. Yanez[1], Oscar A. Peña[2], Benny Chain[3] and Tessa Crompton[1]

Affiliations of authors:

[1]UCL Great Ormond Street Institute of Child Health

30 Guilford Street

London WC1N 1EH, UK

[2] School of Biochemistry

University of Bristol

Bristol, United Kingdom

[3] Division of Infection and Immunity

University College London

London, United Kingdom

Address correspondence to Tessa Crompton, Phone: +44 (0)207 905 2893; Fax: +44 (0)207 813 8494; email: t.crompton@ucl.ac.uk

Key words: T-cell development, embryo, T cell receptor (TCR), Positive selection, MHC-restriction, TCR sequencing, TCR repertoire

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 in

the 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 TCRβ expansion with a less diverse TCRβ 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 TCRβ and TCRα 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 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.

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).

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 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 VJα 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 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 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 ($f(x) = kx^{-\alpha}$) 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.

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.

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.

Data Availability

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

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., SHUGAY, M., MERZLYAK, E. M., STAROVEROV, D. B., PUTINTSEVA, E. V., TURCHANINOVA, M. A., MAMEDOV, I. Z., POGORELYY, M. V., BOLOTIN, D. A., IZRAELSON, M., DAVYDOV, A. N., EGOROV, E. S., KASATSKAYA, S. A., REBRIKOV, D. V., LUKYANOV, S. & CHUDAKOV, D. M. 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., DE MASSY, M. R., ISMAIL, M., READING, J. L., UDDIN, I., WOOLSTON, A., HATIPOGLU, E., OAKES, T., ROSENTHAL, R., PEACOCK, T., RONEL, T., NOURSADEGHI, M., TURATI, V., FURNESS, A. J. S., GEORGIOU, A., WONG, Y. N. S., BEN AISSA, A., SUNDERLAND, M. W., JAMAL-HANJANI, M., VEERIAH, S., BIRKBAK, N. J., WILSON, G. A., HILEY, C. T., GHORANI, E., GUERRA-ASSUNÇÃO, J. A., HERRERO, J., ENVER, T., HADRUP, S. R., HACKSHAW, A., PEGGS, K. S., MCGRANAHAN, N., SWANTON, C., SWANTON, C., JAMAL-HANJANI, M., GEORGIOU, A., SUNDERLAND, M. W., READING, J. L., PEGGS, K. S., GHORANI, E., DE MASSY, M. R., HATIPOGLU, E., QUEZADA, S. A., CHAIN, B., MCGRANAHAN, N., HACKSHAW, A., HILEY, C. T., VEERIAH, S., ROSENTHAL, R., WILSON, G. A., HERRERO, J., BIRKBAK, N. J., NGAI, Y., SHARP, A., RODRIGUES, C., PRESSEY, O., SMITH, S., GOWER, N., DHANDA, H., WATKINS, T. B. K., ESCUDERO, M., STEWART, A., ROWAN, A., GOLDMAN, J., VAN LOO, P., STONE, R. K., DENNER, T., NYE, E., WARD, S., LIM, E., BOEING, S., GRECO, M., BAKIR, M. A., LITCHFIELD, K., NICOD, J., PUTTICK, C., ENFIELD, K., COLLIVER, E., CAMPBELL, B., ABBOSH, C., WU, Y., SKRZYPSKI, M., HYNDS, R. E., MARAFIOTI, T., HARTLEY, J. A., GORMAN, P., LOWE, H. L., ENSELL, L., SPANSWICK, V., KARAMANI, A., MOORE, D., BISWAS, D., RAZAQ, M., BECK, S., HUEBNER, A., DIETZEN, M., NACEUR-LOMBARDELLI, C., AKTHER, M. A., ZHAI, H., KANNU, N., MANZANO, E., 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. 2023a. ggpubr: 'ggplot2' Based Publication Ready Plots.

KASSAMBARA, A. 2023b. rstatix: Pipe-Friendly Framework for Basic Statistical Tests. R package version 0.7.2 ed.

KASSAMBARA A, M. F. 2020. factoextra: Extract and Visualize the Results of Multivariate Data Analyses. R package version 1.0.7 ed.

KOLDE, R. 2019. pheatmap: Pretty Heatmaps. R package version 1.0.12 ed.

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**,** e21627.

MONTECINO-RODRIGUEZ, E. & DORSHKIND, K. 2023. The layered development of mouse B and T Cells. *Immunol Rev,* 315**,** 79-88.

OAKES, T., HEATHER, J. M., BEST, K., BYNG-MADDICK, R., HUSOVSKY, C., ISMAIL, M., JOSHI, K., MAXWELL, G., NOURSADEGHI, M., RIDDELL, N., RUEHL, T., TURNER, C. T., UDDIN, I. & CHAIN, B. 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**,** 1267.

OKSANEN, J., BLANCHET, F. G., FRIENDLY, M., KINDT, R., LEGENDRE, P., MCGLINN, D., MINCHIN, P. R., O'HARA, R. B., SIMPSON, G. L., SOLYMOS, P., STEVENS, M. H. H., SZOECS, E. & WAGNER, H. 2022. vegan: Community Ecology Package. R package version 2.6-4 ed.

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., BOTTING, R. A., DOMINGUEZ CONDE, C., POPESCU, D. M., LAVAERT, M., KUNZ, D. J., GOH, I., STEPHENSON, E., RAGAZZINI, R., TUCK, E., WILBREY-CLARK, A., ROBERTS, K., KEDLIAN, V. R., FERDINAND, J. R., HE, X., WEBB, S., MAUNDER, D., VANDAMME, N., MAHBUBANI, K. T., POLANSKI, K., MAMANOVA, L., BOLT, L., CROSSLAND, D., DE RITA, F., FULLER, A., FILBY, A., REYNOLDS, G., DIXON, D., SAEB-PARSY, K., LISGO, S., HENDERSON, D., VENTO-TORMO, R., BAYRAKTAR, O. A., BARKER, R. A., MEYER, K. B., SAEYS, Y., BONFANTI, P., BEHJATI, S., CLATWORTHY, M. R., TAGHON, T., HANIFFA, M. & TEICHMANN, S. A. 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., ELHANATI, Y., MARCOU, Q., SYCHEVA, A. L., KOMECH, E. A., NAZAROV, V. I., BRITANOVA, O. V., CHUDAKOV, D. M., MAMEDOV, I. Z., LEBEDEV, Y. B., MORA, T. & WALCZAK, A. M. 2017. Persisting fetal

clonotypes influence the structure and overlap of adult human T cell receptor repertoires. *PLoS Comput Biol,* 13**,** e1005572.

RACKAITYTE, E. & HALKIAS, J. 2020. Mechanisms of Fetal T Cell Tolerance and Immune Regulation. *Front Immunol,* 11**,** 588.

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., JR., 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., PATEL, R. K., REYNALDI, A., GRENIER, J. K., WANG, J., WATSON, N. B., NZINGHA, K., YEE MON, K. J., PENG, S. A., GRIMSON, A., DAVENPORT, M. P. & RUDD, B. D. 2018. Developmental Origin Governs CD8(+) T Cell Fate Decisions during Infection. *Cell,* 174**,** 117-130 e14.

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., POLANSKI, K., DANN, E., LINDEBOOM, R. G. H., VILARRASA-BLASI, R., VENTO-TORMO, R., HANIFFA, M., MEYER, K. B., DRATVA, L. M., TUONG, Z. K., CLATWORTHY, M. R. & TEICHMANN, S. A. 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, A. M., BRYAN J, CHANG W, MCGOWAN LD, FRANÇOIS R, GROLEMUND G, HAYES A, HENRY L, HESTER J, KUHN M, PEDERSEN TL, MILLER E, BACHE SM, MÜLLER K, OOMS J, ROBINSON D, SEIDEL DP, SPINU V, TAKAHASHI K, VAUGHAN D, WILKE C, WOO K, YUTANI H 2019. Welcome to the tidyverse. *Journal of Open Source Software,* 4**,** 1686.

WONG, W. K., LEEM, J. & DEANE, C. M. 2019. Comparative Analysis of the CDR Loops of Antigen Receptors. *Front Immunol,* 10**,** 2454.

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. R package version 0.2-13 ed.

Figure Legends

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 (>99th 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.

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.

Figure 5. Decreased TCRβ 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: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

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.

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.

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.

Supplementary Figure Legends

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 (>99th 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: TCRα 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.

Supplementary Tables show antibodies used for sorting adult (Table 1) and foetal (Table 2) thymus and primers used (Table 3).

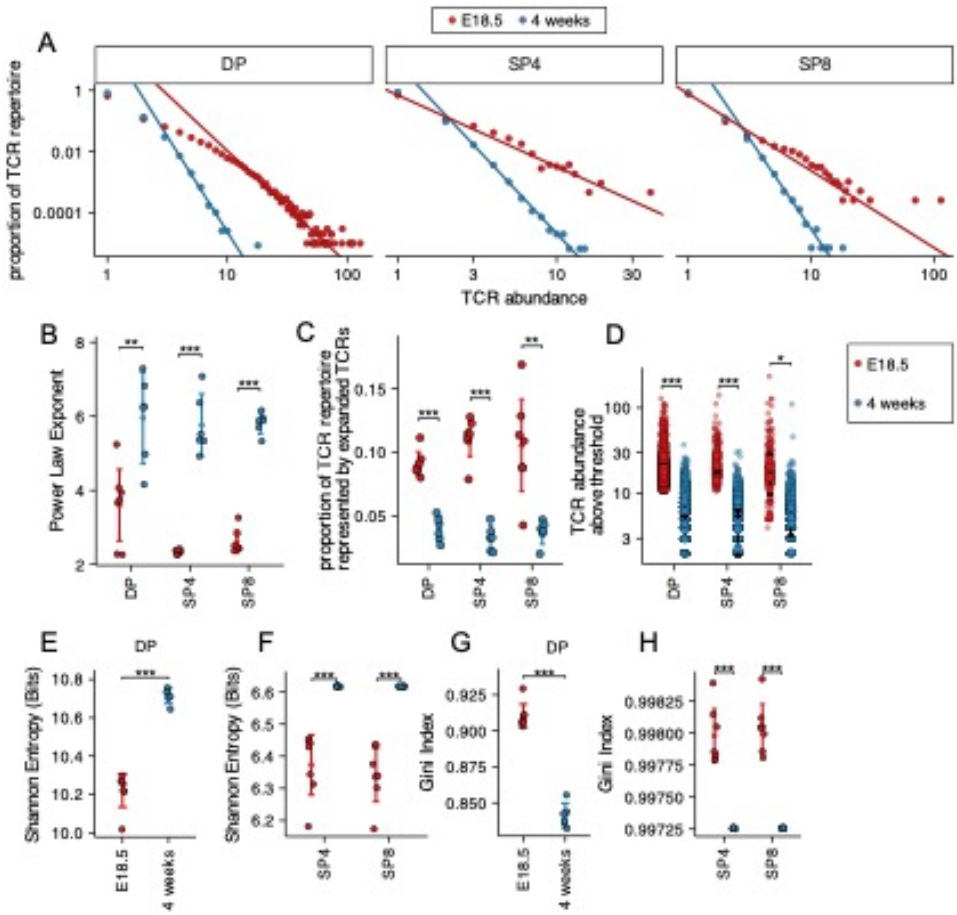

E18.5
4 weeks
A
DP
SP4
SP8
proportion of TCR repertoire
TCR abundance
B
Power Law Exponent
C
proportion of TCR repertoire represented by expanded TCRs
D
TCR abundance above threshold
E
Shannon Entropy (Bits)
F
G
Gini Index
H

Figure 1

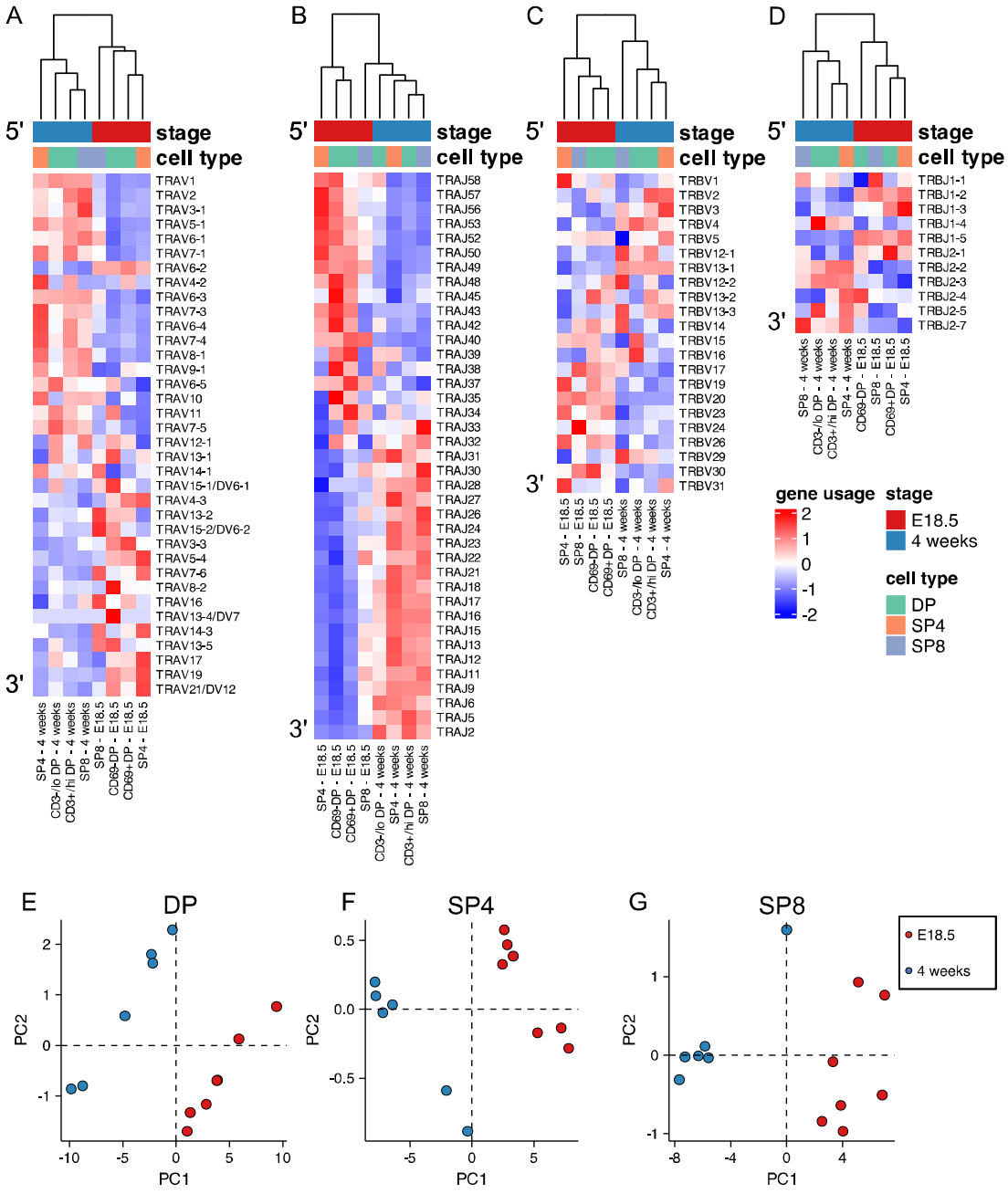

A
B
C
D
5'
3'
stage
cell type
TRAV1
TRAV2
TRAV3-1
TRAV5-1
TRAV6-1
TRAV7-1
TRAV6-2
TRAV4-2
TRAV6-3
TRAV7-3
TRAV6-4
TRAV7-4
TRAV8-1
TRAV9-1
TRAV6-5
TRAV10
TRAV11
TRAV7-5
TRAV12-1
TRAV13-1
TRAV14-1
TRAV15-1/DV6-1
TRAV4-3
TRAV13-2
TRAV15-2/DV6-2
TRAV3-3
TRAV5-4
TRAV7-6
TRAV8-2
TRAV16
TRAV13-4/DV7
TRAV14-3
TRAV13-5
TRAV17
TRAV19
TRAV21/DV12
SP4 - 4 weeks
CD3-/lo DP - 4 weeks
CD3+/hi DP - 4 weeks
SP8 - 4 weeks
SP8 - E18.5
CD69-DP - E18.5
CD69+DP - E18.5
SP4 - E18.5
TRAJ58
TRAJ57
TRAJ56
TRAJ53
TRAJ52
TRAJ50
TRAJ49
TRAJ48
TRAJ45
TRAJ43
TRAJ42
TRAJ40
TRAJ39
TRAJ38
TRAJ37
TRAJ35
TRAJ34
TRAJ33
TRAJ32
TRAJ31
TRAJ30
TRAJ28
TRAJ27
TRAJ26
TRAJ24
TRAJ23
TRAJ22
TRAJ21
TRAJ18
TRAJ17
TRAJ16
TRAJ15
TRAJ13
TRAJ12
TRAJ11
TRAJ9
TRAJ6
TRAJ5
TRAJ2
SP4 - E18.5
CD69-DP - E18.5
CD69+DP - E18.5
SP8 - E18.5
CD3-/lo DP - 4 weeks
SP4 - 4 weeks
CD3+/hi DP - 4 weeks
SP8 - 4 weeks
TRBV1
TRBV2
TRBV3
TRBV4
TRBV5
TRBV12-1
TRBV13-1
TRBV12-2
TRBV13-2
TRBV13-3
TRBV14
TRBV15
TRBV16
TRBV17
TRBV19
TRBV20
TRBV23
TRBV24
TRBV26
TRBV29
TRBV30
TRBV31
SP4 - E18.5
SP8 - E18.5
CD69-DP - E18.5
CD69+DP - E18.5
SP8 - 4 weeks
CD3-/lo DP - 4 weeks
CD3+/hi DP - 4 weeks
SP4 - 4 weeks
TRBJ1-1
TRBJ1-2
TRBJ1-3
TRBJ1-4
TRBJ1-5
TRBJ2-1
TRBJ2-2
TRBJ2-3
TRBJ2-4
TRBJ2-5
TRBJ2-7
SP8 - 4 weeks
CD3-/lo DP - 4 weeks
CD3+/hi DP - 4 weeks
SP4 - 4 weeks
CD69-DP - E18.5
SP8 - E18.5
CD69+DP - E18.5
SP4 - E18.5
gene usage
2
1
0
-1
-2
stage
E18.5
4 weeks
cell type
DP
SP4
SP8
E
DP
PC2
PC1
F
SP4
PC2
PC1
G
SP8
PC2
PC1
E18.5
4 weeks

Figure 2

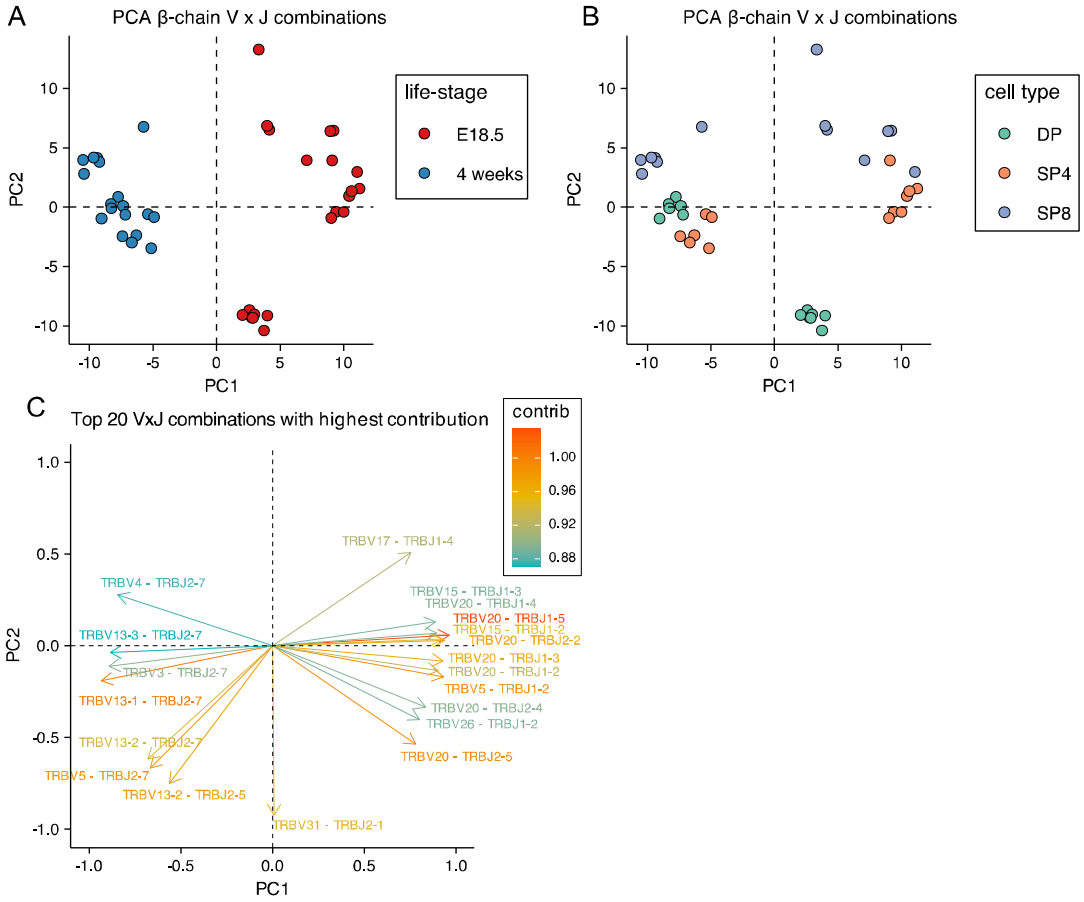

A
PCA β-chain V x J combinations
life-stage
E18.5
4 weeks
PC1
PC2
B
PCA β-chain V x J combinations
cell type
DP
SP4
SP8
PC1
PC2
C
Top 20 VxJ combinations with highest contribution
contrib
1.00
0.96
0.92
0.88
TRBV17 - TRBJ1-4
TRBV4 - TRBJ2-7
TRBV15 - TRBJ1-3
TRBV20 - TRBJ1-4
TRBV20 - TRBJ1-5
TRBV15 - TRBJ1-2
TRBV20 - TRBJ2-2
TRBV13-3 - TRBJ2-7
TRBV20 - TRBJ1-3
TRBV3 - TRBJ2-7
TRBV20 - TRBJ1-2
TRBV5 - TRBJ1-2
TRBV13-1 - TRBJ2-7
TRBV20 - TRBJ2-4
TRBV26 - TRBJ1-2
TRBV13-2 - TRBJ2-7
TRBV20 - TRBJ2-5
TRBV5 - TRBJ2-7
TRBV13-2 - TRBJ2-5
TRBV31 - TRBJ2-1
PC1
PC2

Figure 3

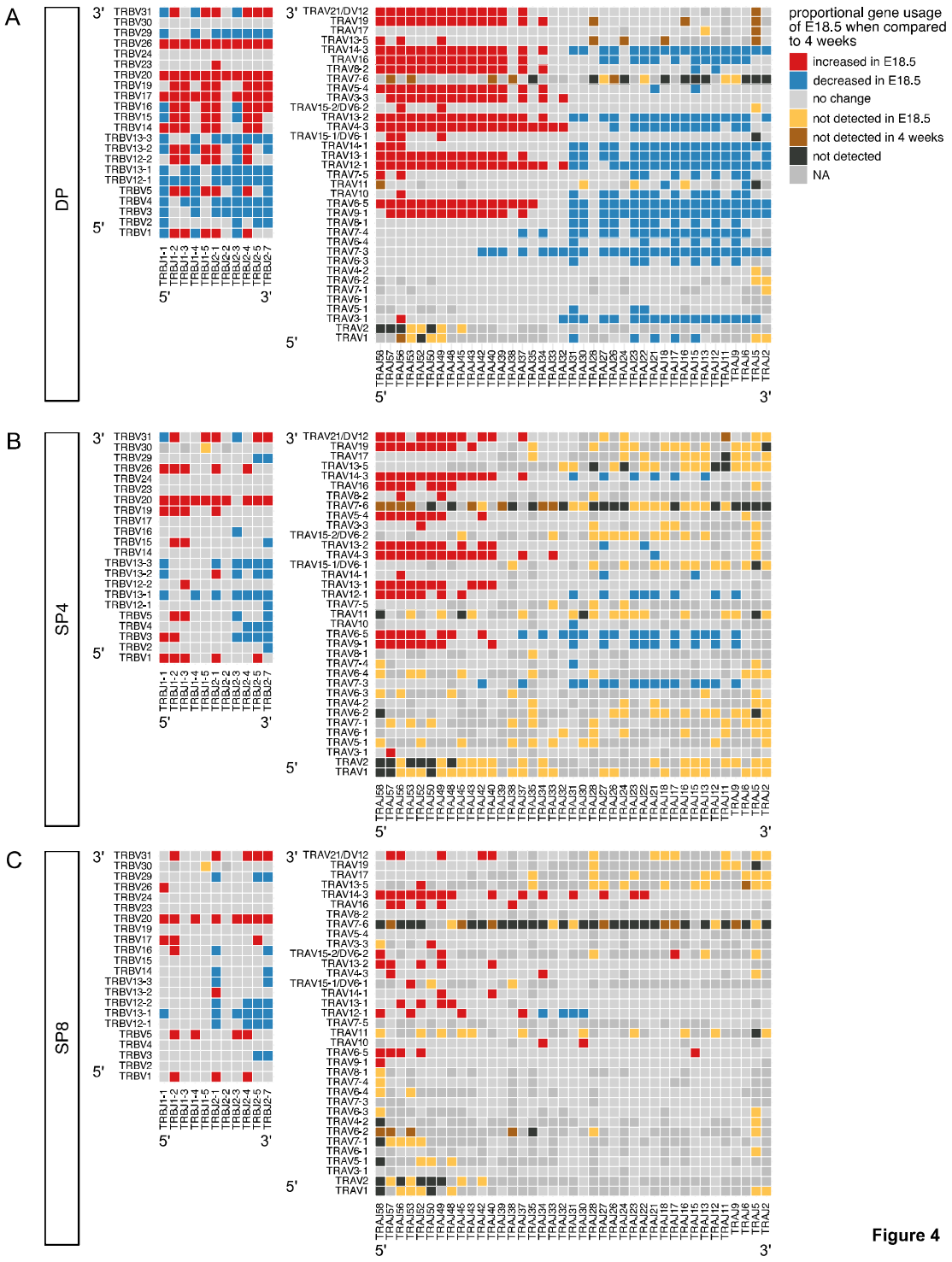

A
DP
B
SP4
C
SP8
proportional gene usage of E18.5 when compared to 4 weeks
increased in E18.5
decreased in E18.5
no change
not detected in E18.5
not detected in 4 weeks
not detected
NA
3'
5'

Figure 4

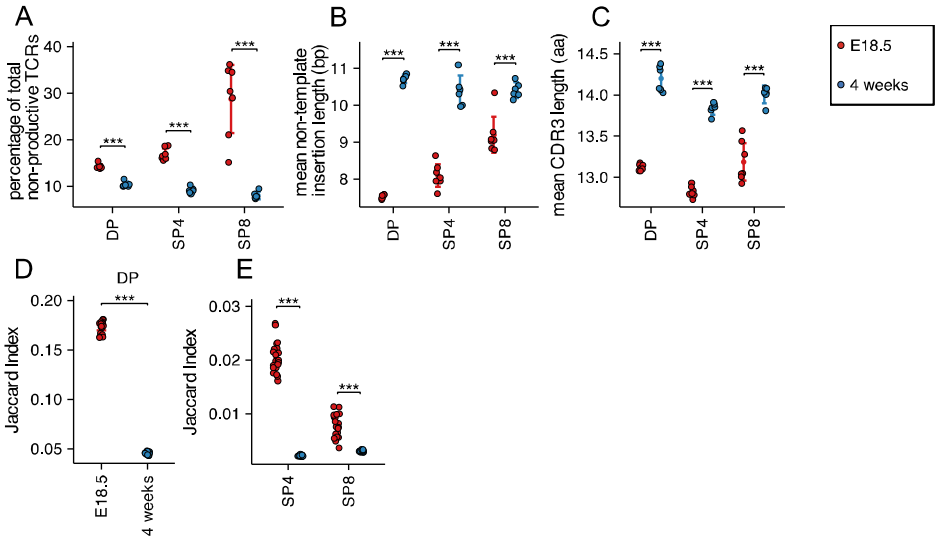

A
percentage of total non-productive TCRs
40
30
20
10
***
DP
SP4
SP8
B
mean non-template insertion length (bp)
11
10
9
8
C
mean CDR3 length (aa)
14.5
14.0
13.5
13.0
E18.5
4 weeks
D
DP
Jaccard Index
0.20
0.15
0.10
0.05
E18.5
4 weeks
E
0.03
0.02
0.01

Figure 5

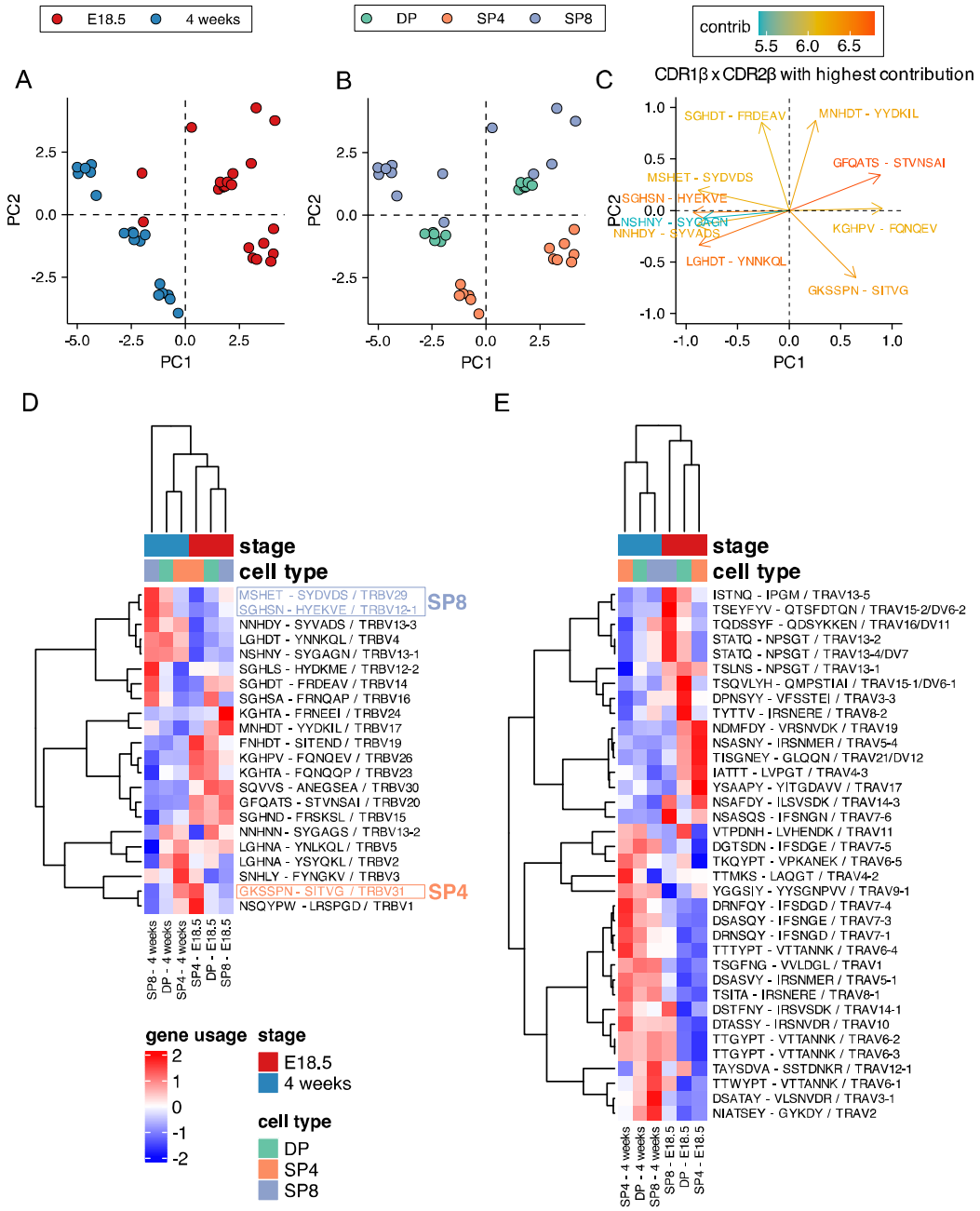

E18.5
4 weeks
DP
SP4
SP8
contrib
5.5 6.0 6.5
A
B
C
CDR1β x CDR2β with highest contribution
PC1
PC2
SGHDT - FRDEAV
MNHDT - YYDKIL
GFQATS - STVNSAI
MSHET - SYDVDS
SGHSN - HYEKVE
NSHNY - SYGAGN
NNHDY - SYVADS
KGHPV - FQNQEV
LGHDT - YNNKQL
GKSSPN - SITVG
D
E
stage
cell type
MSHET - SYDVDS / TRBV29
SGHSN - HYEKVE / TRBV12-1
SP8
NNHDY - SYVADS / TRBV13-3
LGHDT - YNNKQL / TRBV4
NSHNY - SYGAGN / TRBV13-1
SGHLS - HYDKME / TRBV12-2
SGHDT - FRDEAV / TRBV14
SGHSA - FRNQAP / TRBV16
KGHTA - FRNEEI / TRBV24
MNHDT - YYDKIL / TRBV17
FNHDT - SITEND / TRBV19
KGHPV - FQNQEV / TRBV26
KGHTA - FQNQQP / TRBV23
SQVVS - ANEGSEA / TRBV30
GFQATS - STVNSAI / TRBV20
SGHND - FRSKSL / TRBV15
NNHNN - SYGAGS / TRBV13-2
LGHNA - YNLKQL / TRBV5
LGHNA - YSYQKL / TRBV2
SNHLY - FYNGKV / TRBV3
GKSSPN - SITVG / TRBV31
SP4
NSQYPW - LRSPGD / TRBV1
SP8 - 4 weeks
DP - 4 weeks
SP4 - 4 weeks
SP4 - E18.5
DP - E18.5
SP8 - E18.5
gene usage
2
1
0
-1
-2
stage
E18.5
4 weeks
cell type
DP
SP4
SP8
ISTNQ - IPGM / TRAV13-5
TSEYFYV - QTSFDTQN / TRAV15-2/DV6-2
TQDSSYF - QDSYKKEN / TRAV16/DV11
STATQ - NPSGT / TRAV13-2
STATQ - NPSGT / TRAV13-4/DV7
TSLNS - NPSGT / TRAV13-1
TSQVLYH - QMPSTIAI / TRAV15-1/DV6-1
DPNSYY - VFSSTEI / TRAV3-3
TYTTV - IRSNERE / TRAV8-2
NDMFDY - VRSNVDK / TRAV19
NSASNY - IRSNMER / TRAV5-4
TISGNEY - GLQQN / TRAV21/DV12
IATTT - LVPGT / TRAV4-3
YSAAPY - YITGDAVV / TRAV17
NSAFDY - ILSVSDK / TRAV14-3
NSASQS - IFSNGN / TRAV7-6
VTPDNH - LVHENDK / TRAV11
DGTSDN - IFSDGE / TRAV7-5
TKQYPT - VPKANEK / TRAV6-5
TTMKS - LAQGT / TRAV4-2
YGGSIY - YYSGNPVV / TRAV9-1
DRNFQY - IFSDGD / TRAV7-4
DSASQY - IFSNGE / TRAV7-3
DRNSQY - IFSNGD / TRAV7-1
TTTYPT - VTTANNK / TRAV6-4
TSGFNG - VVLDGL / TRAV1
DSASVY - IRSNMER / TRAV5-1
TSITA - IRSNERE / TRAV8-1
DSTFNY - IRSVSDK / TRAV14-1
DTASSY - IRSNVDR / TRAV10
TTGYPT - VTTANNK / TRAV6-2
TTGYPT - VTTANNK / TRAV6-3
TAYSDVA - SSTDNKR / TRAV12-1
TTWYPT - VTTANNK / TRAV6-1
DSATAY - VLSNVDR / TRAV3-1
NIATSEY - GYKDY / TRAV2
SP4 - 4 weeks
DP - 4 weeks
SP8 - 4 weeks
SP8 - E18.5
DP - E18.5
SP4 - E18.5

Figure 6

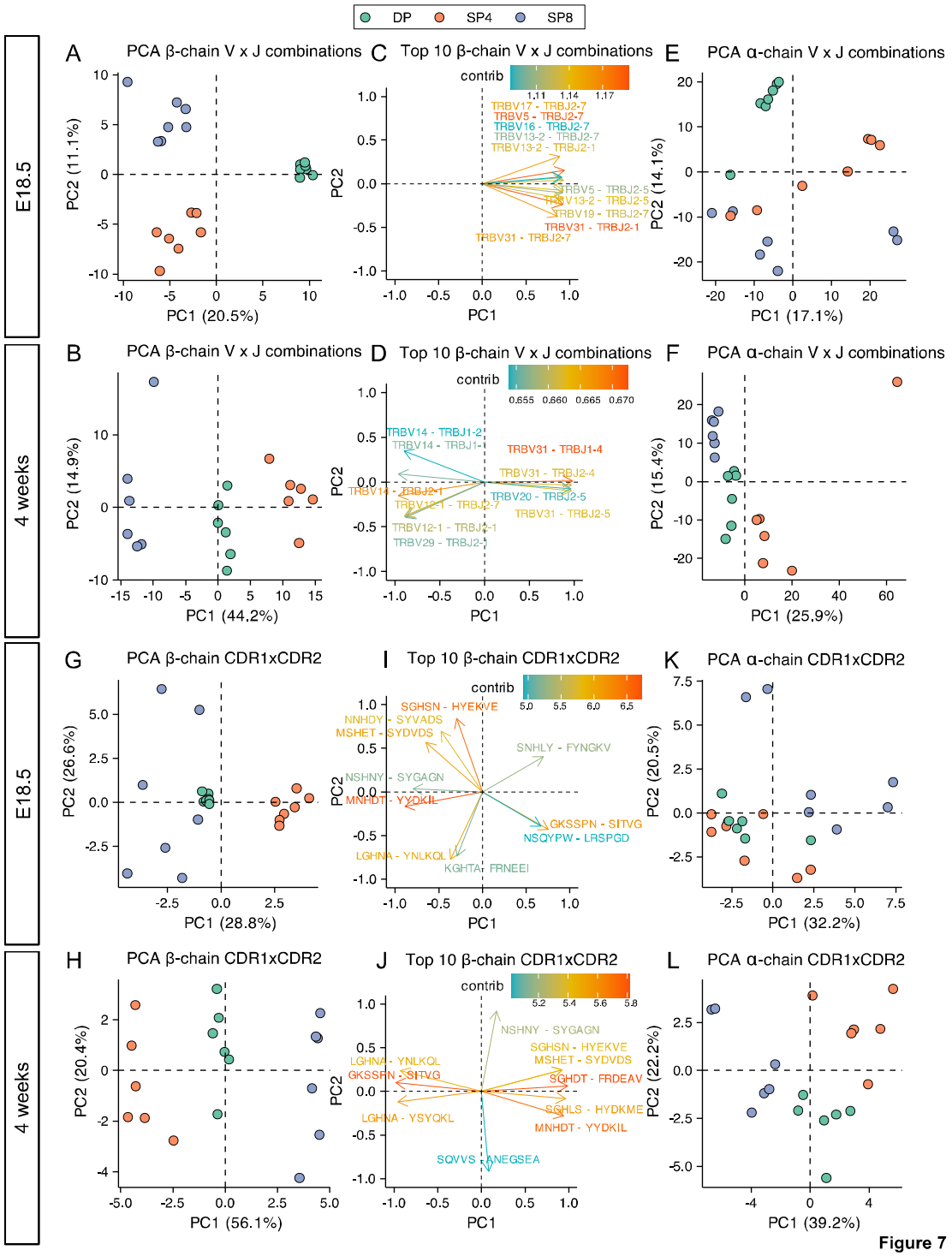

DP
SP4
SP8
E18.5
4 weeks
E18.5
4 weeks
A
PCA β-chain V x J combinations
PC1 (20.5%)
PC2 (11.1%)
B
PCA β-chain V x J combinations
PC1 (44.2%)
PC2 (14.9%)
C
Top 10 β-chain V x J combinations
contrib
TRBV17 - TRBJ2-7
TRBV5 - TRBJ2-7
TRBV16 - TRBJ2-7
TRBV13-2 - TRBJ2-7
TRBV13-2 - TRBJ2-1
TRBV5 - TRBJ2-5
TRBV13-2 - TRBJ2-5
TRBV19 - TRBJ2-7
TRBV31 - TRBJ2-1
TRBV31 - TRBJ2-7
D
Top 10 β-chain V x J combinations
TRBV14 - TRBJ1-2
TRBV14 - TRBJ1-1
TRBV31 - TRBJ1-4
TRBV31 - TRBJ2-4
TRBV20 - TRBJ2-5
TRBV31 - TRBJ2-5
TRBV12-1 - TRBJ2-1
TRBV29 - TRBJ2-1
E
PCA α-chain V x J combinations
PC1 (17.1%)
PC2 (14.1%)
F
PCA α-chain V x J combinations
PC1 (25.9%)
PC2 (15.4%)
G
PCA β-chain CDR1xCDR2
PC1 (28.8%)
PC2 (26.6%)
H
PCA β-chain CDR1xCDR2
PC1 (56.1%)
PC2 (20.4%)
I
Top 10 β-chain CDR1xCDR2
SGHSN - HYEKVE
NNHDY - SYVADS
MSHET - SYDVDS
SNHLY - FYNGKV
NSHNY - SYGAGN
MNHDT - YYDKIL
GKSSPN - SITVG
NSQYPW - LRSPGD
LGHNA - YNLKQL
KGHTA - FRNEEI
J
Top 10 β-chain CDR1xCDR2
NSHNY - SYGAGN
SGHSN - HYEKVE
MSHET - SYDVDS
LGHNA - YNLKQL
GKSSPN - SITVG
SGHDT - FRDEAV
SGHLS - HYDKME
LGHNA - YSYQKL
MNHDT - YYDKIL
SQVVS - ANEGSEA
K
PCA α-chain CDR1xCDR2
PC1 (32.2%)
PC2 (20.5%)
L
PCA α-chain CDR1xCDR2
PC1 (39.2%)
PC2 (22.2%)

Figure 7

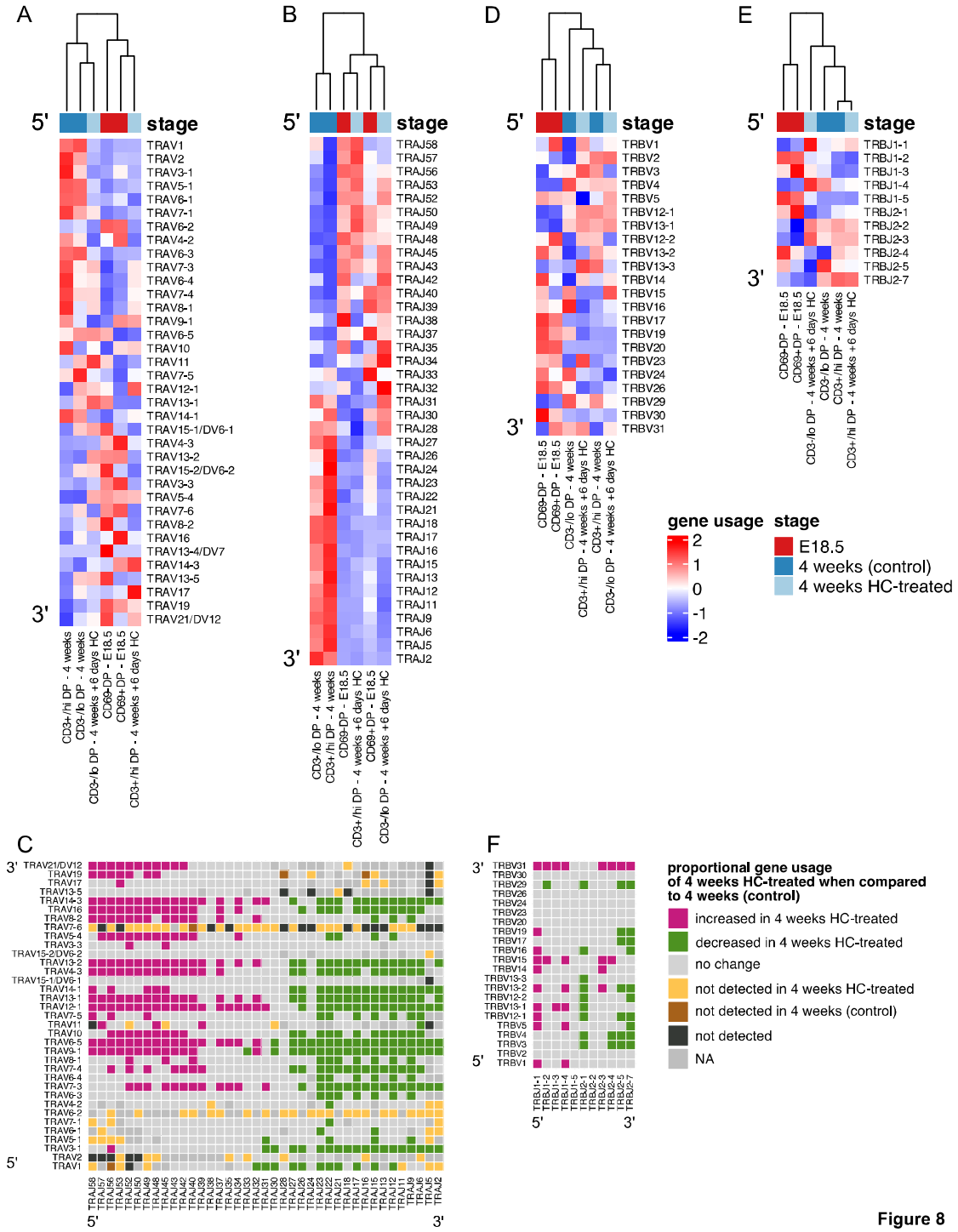
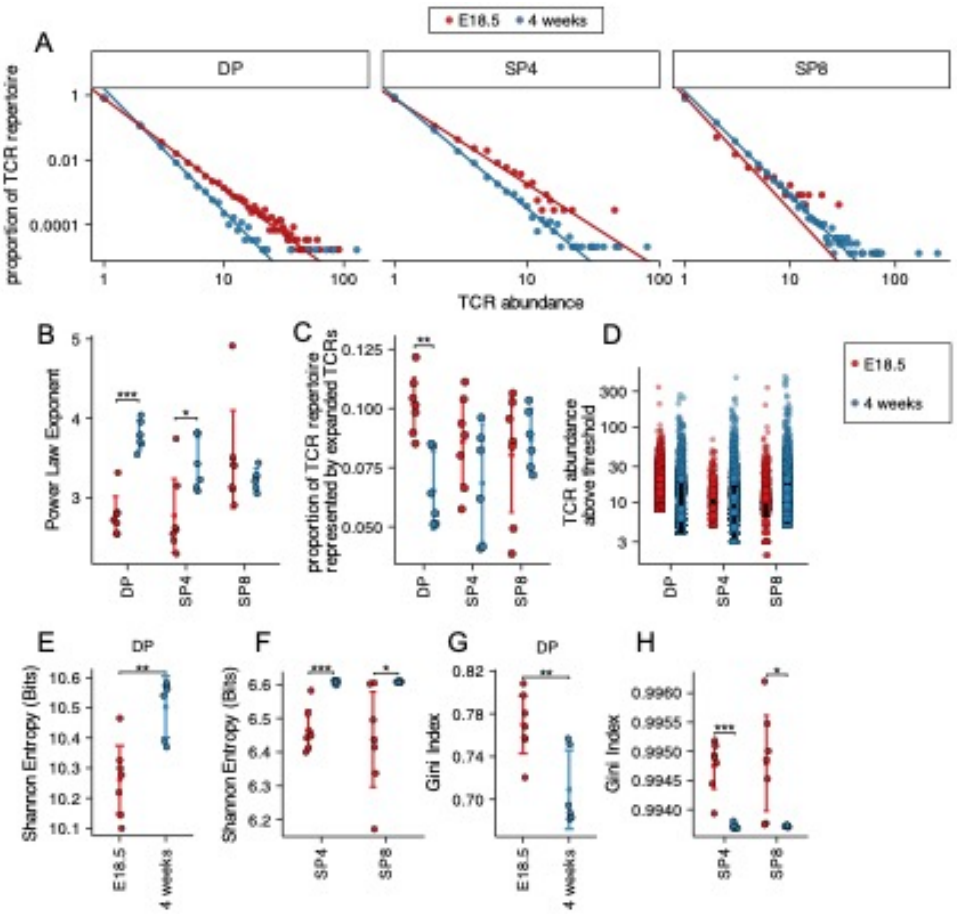
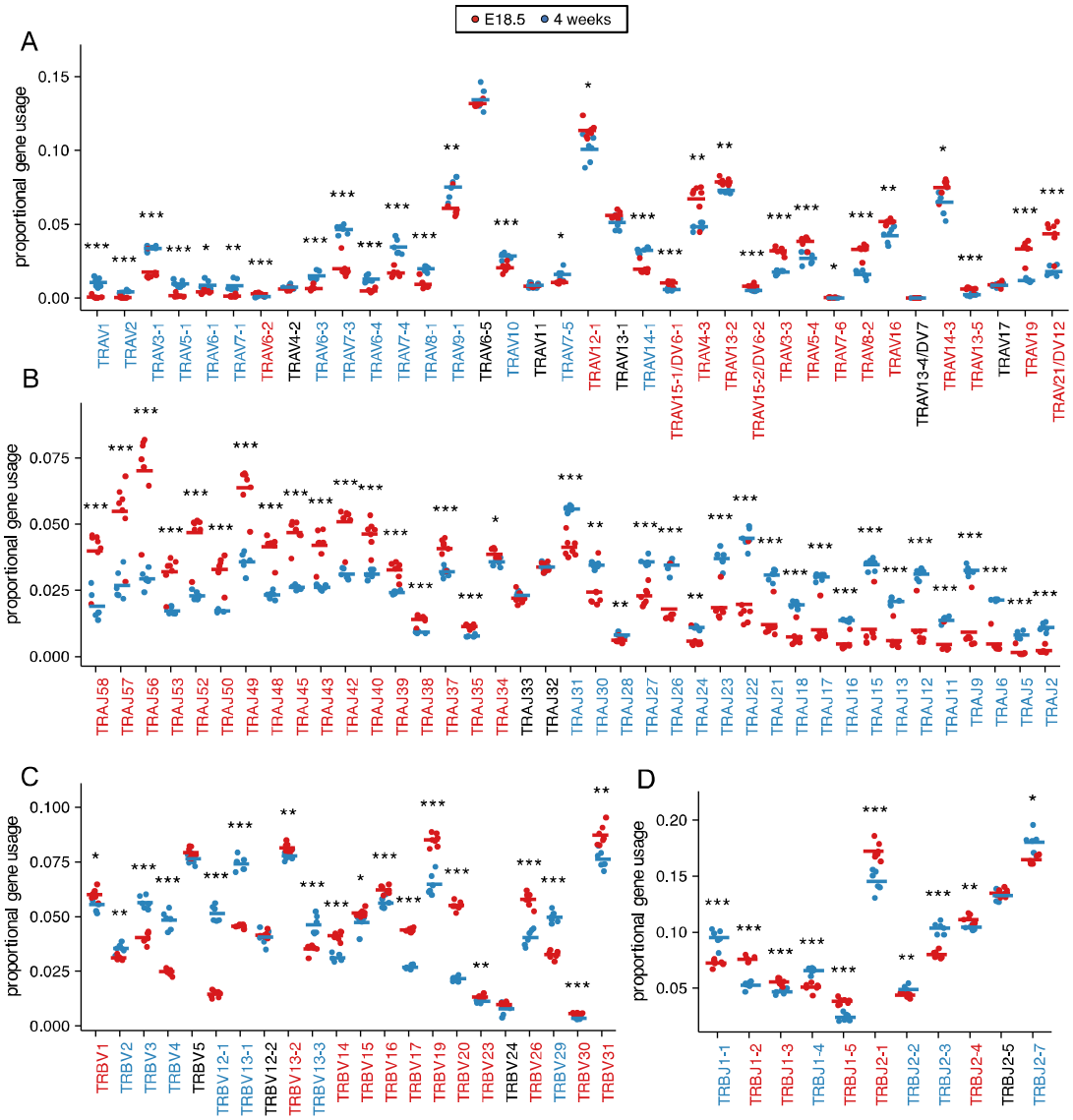

A
B
D
E
C
F
5'
3'
stage
TRAV1
TRAV2
TRAV3-1
TRAV5-1
TRAV6-1
TRAV7-1
TRAV6-2
TRAV4-2
TRAV6-3
TRAV7-3
TRAV6-4
TRAV7-4
TRAV8-1
TRAV9-1
TRAV6-5
TRAV10
TRAV11
TRAV7-5
TRAV12-1
TRAV13-1
TRAV14-1
TRAV15-1/DV6-1
TRAV4-3
TRAV13-2
TRAV15-2/DV6-2
TRAV3-3
TRAV5-4
TRAV7-6
TRAV8-2
TRAV16
TRAV13-4/DV7
TRAV14-3
TRAV13-5
TRAV17
TRAV19
TRAV21/DV12
CD3+/hi DP - 4 weeks
CD3-/lo DP - 4 weeks
CD3-/lo DP - 4 weeks +6 days HC
CD69-DP - E18.5
CD69+DP - E18.5
CD3+/hi DP - 4 weeks +6 days HC
TRAJ58
TRAJ57
TRAJ56
TRAJ53
TRAJ52
TRAJ50
TRAJ49
TRAJ48
TRAJ45
TRAJ43
TRAJ42
TRAJ40
TRAJ39
TRAJ38
TRAJ37
TRAJ35
TRAJ34
TRAJ33
TRAJ32
TRAJ31
TRAJ30
TRAJ28
TRAJ27
TRAJ26
TRAJ24
TRAJ23
TRAJ22
TRAJ21
TRAJ18
TRAJ17
TRAJ16
TRAJ15
TRAJ13
TRAJ12
TRAJ11
TRAJ9
TRAJ6
TRAJ5
TRAJ2
TRBV1
TRBV2
TRBV3
TRBV4
TRBV5
TRBV12-1
TRBV13-1
TRBV12-2
TRBV13-2
TRBV13-3
TRBV14
TRBV15
TRBV16
TRBV17
TRBV19
TRBV20
TRBV23
TRBV24
TRBV26
TRBV29
TRBV30
TRBV31
TRBJ1-1
TRBJ1-2
TRBJ1-3
TRBJ1-4
TRBJ1-5
TRBJ2-1
TRBJ2-2
TRBJ2-3
TRBJ2-4
TRBJ2-5
TRBJ2-7
gene usage
2
1
0
-1
-2
stage
E18.5
4 weeks (control)
4 weeks HC-treated
proportional gene usage
of 4 weeks HC-treated when compared
to 4 weeks (control)
increased in 4 weeks HC-treated
decreased in 4 weeks HC-treated
no change
not detected in 4 weeks HC-treated
not detected in 4 weeks (control)
not detected
NA

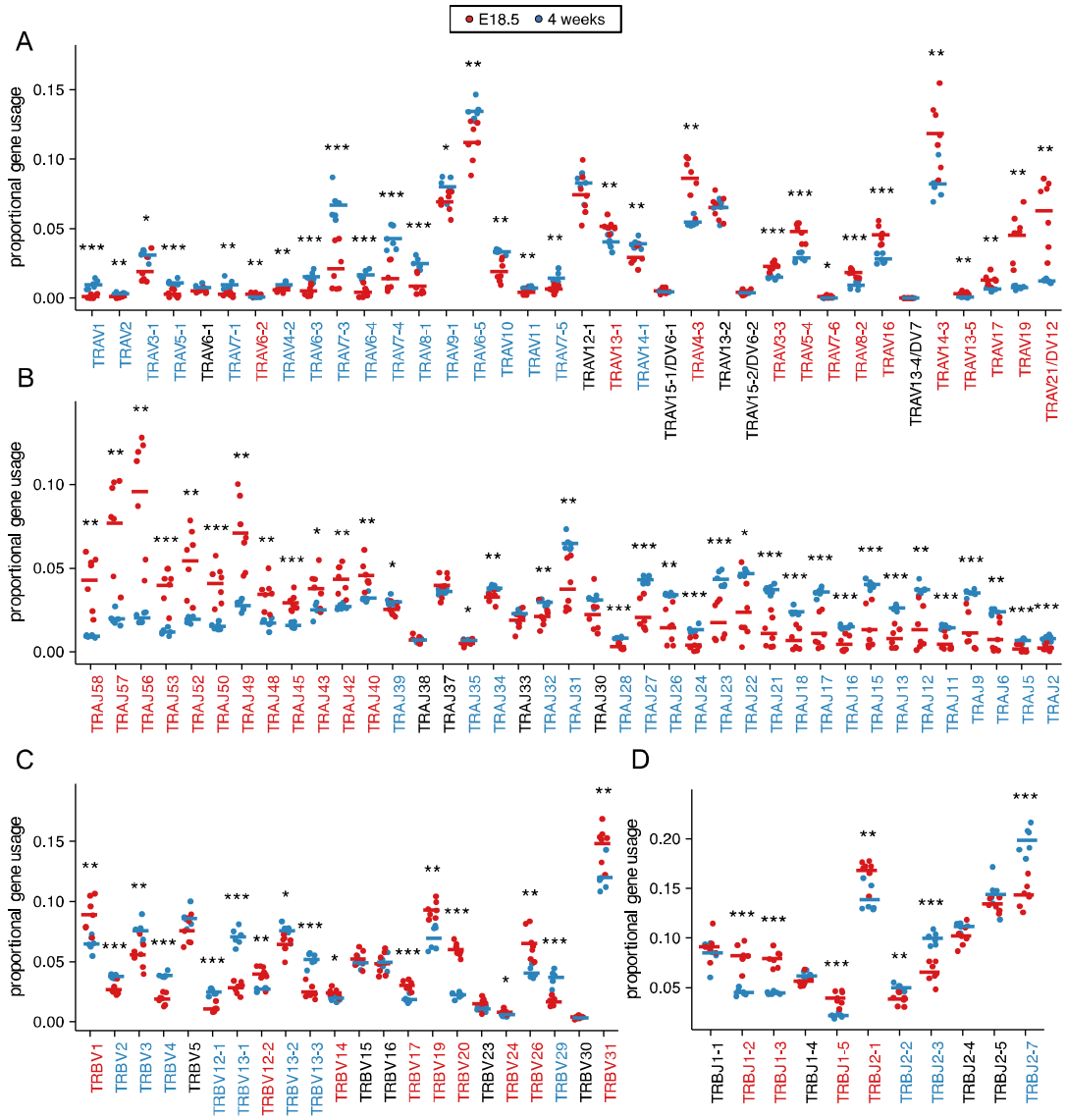
Figure 8



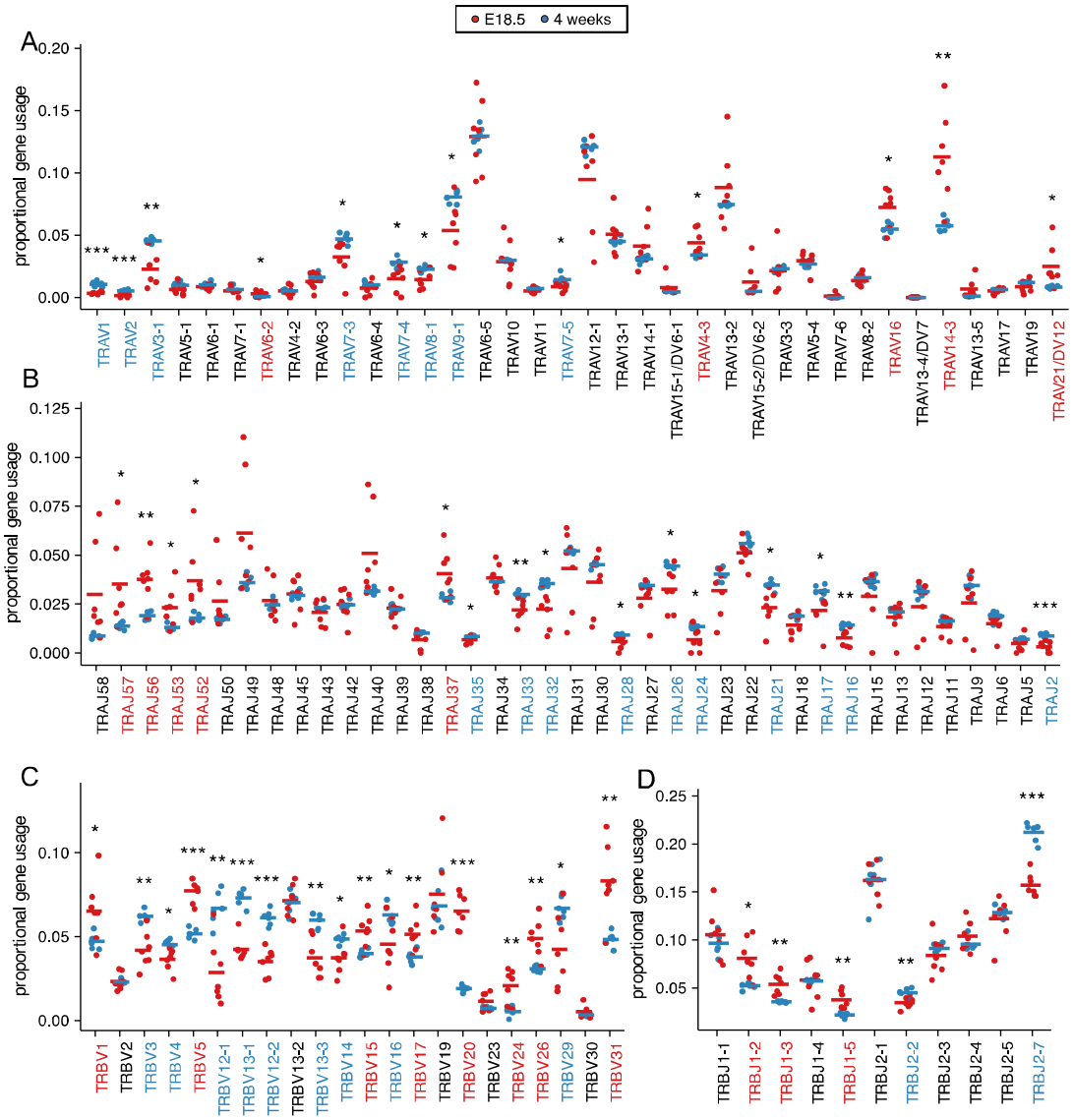
Supplementary Figure 1



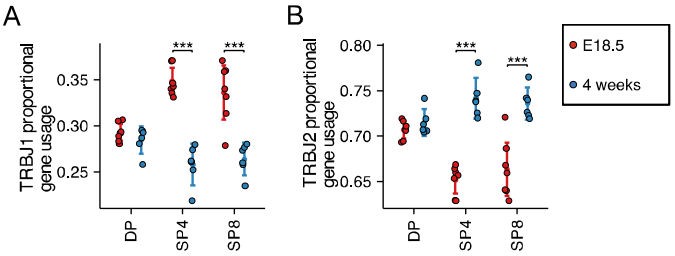
Supplementary Figure 2



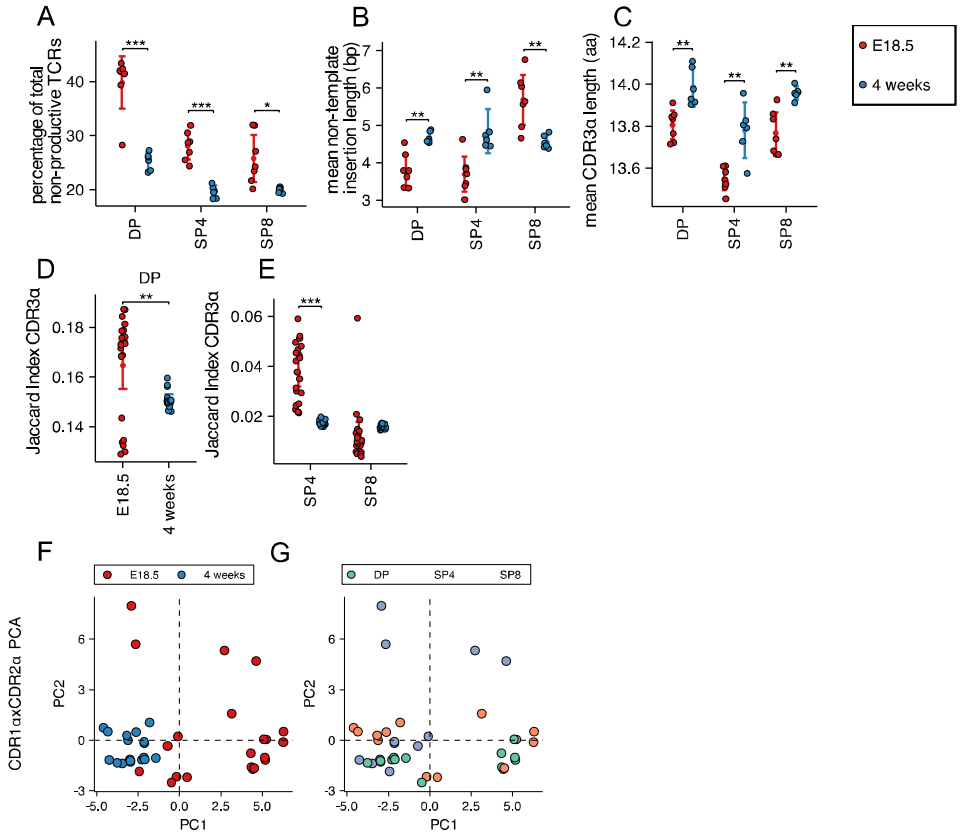
Supplementary Figure 3



Supplementary Figure 4

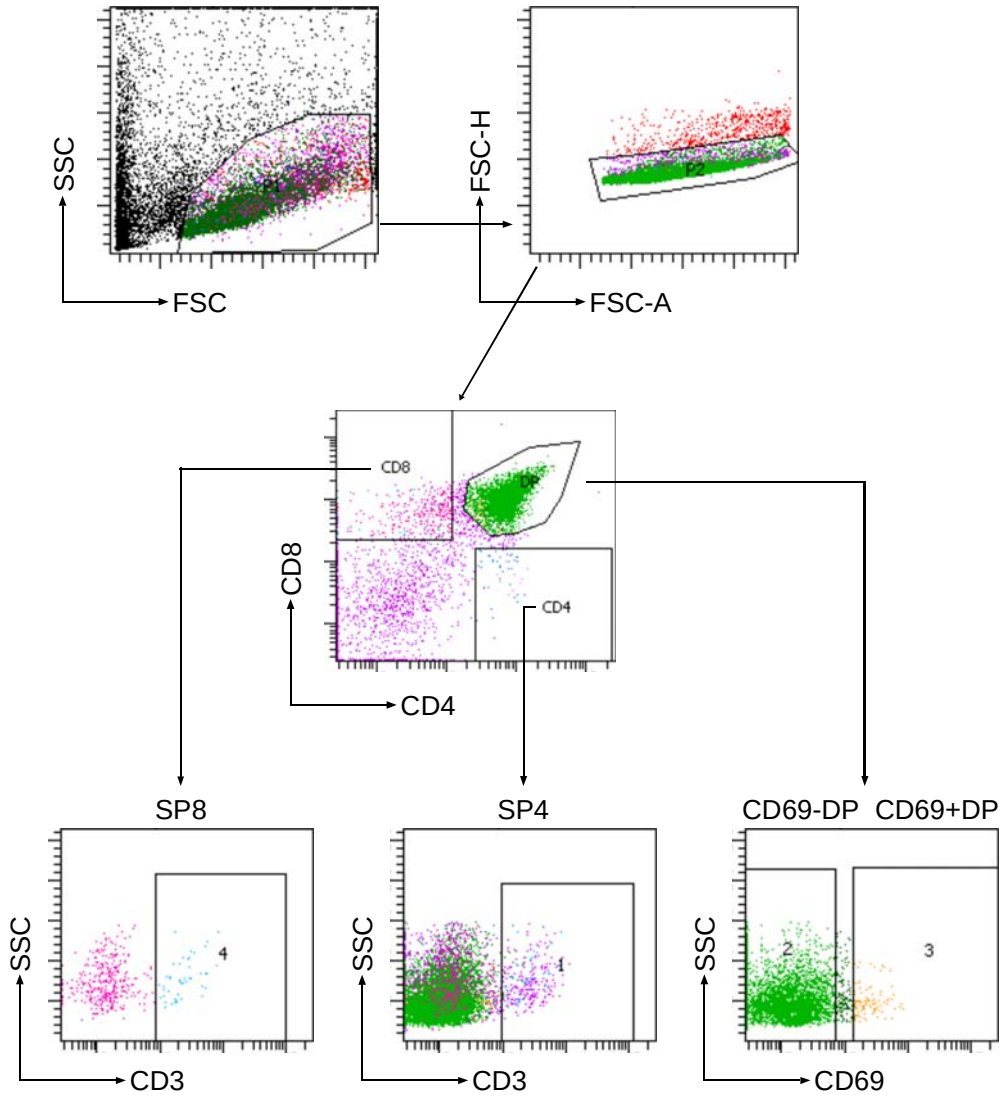


Supplementary Figure 5



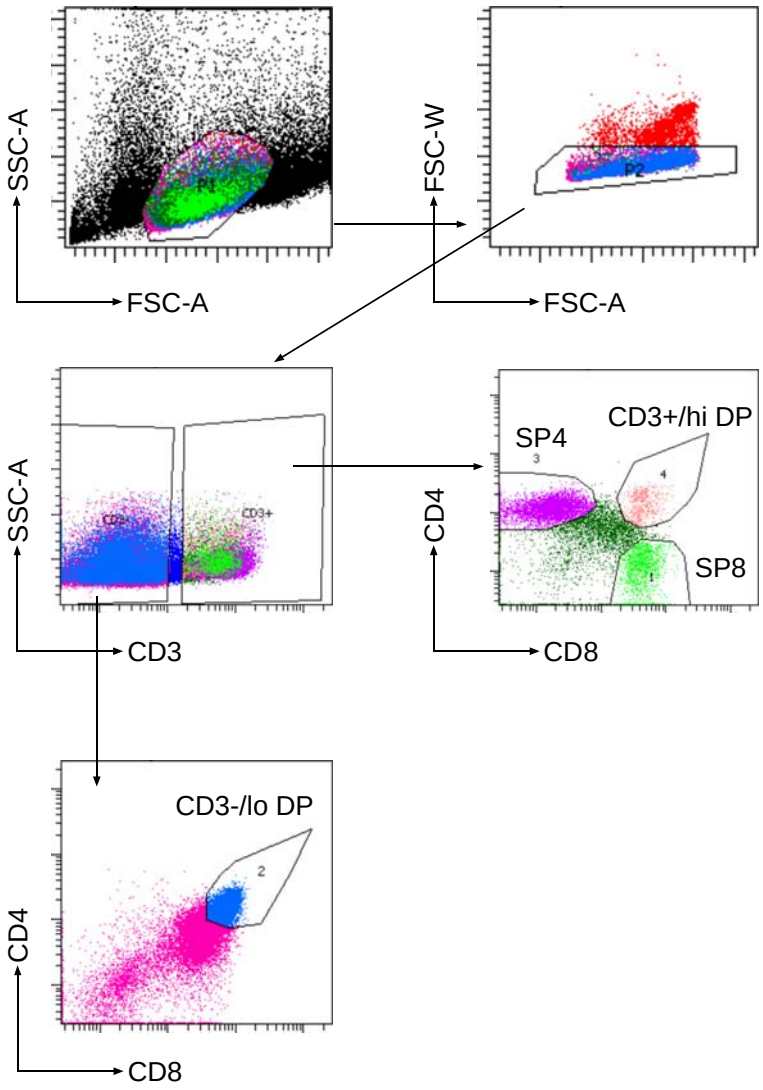
Supplementary Figure 6

E18.5 cell sorting layout



Supplementary Figure 7

4 weeks cell sorting layout



Supplementary Figure 8