Article

Pre-T cell receptor self-MHC sampling restricts thymocyte dedifferentiation

https://doi.org/10.1038/s41586-022-05555-7

Received: 3 September 2021

Accepted: 11 November 2022

Published online: 21 November 2022

Check for updates

Jonathan S. Duke-Cohan[1,2,3]✉, Aoi Akitsu[1,2,3], Robert J. Mallis[1,2,4], Cameron M. Messier[5], Patrick H. Lizotte[5], Jon C. Aster[6], Wonmuk Hwang[7,8,9], Matthew J. Lang[10,11] & Ellis L. Reinherz[1,2,3]✉

Programming T cells to distinguish self from non-self is a vital, multi-step process that occurs in the thymus[1–4]. Signalling through the pre-T cell receptor (preTCR), a CD3-associated heterodimer comprising an invariant pTα chain and a clone-specific β chain, is a critical early checkpoint in thymocyte development within the αβ T cell lineage[5,6]. PreTCRs arrayed on $CD4^-CD8^-$ double-negative thymocytes ligate peptides bound to major histocompatibility complex molecules (pMHC) on thymic stroma, similar to αβ T cell receptors that appear on $CD4^+CD8^+$ double-positive thymocytes, but via a different molecular docking strategy[7–10]. Here we show the consequences of these distinct interactions for thymocyte progression using synchronized fetal thymic progenitor cultures that differ in the presence or absence of pMHC on support stroma, and single-cell transcriptomes at key thymocyte developmental transitions. Although major histocompatibility complex (MHC)-negative stroma fosters αβ T cell differentiation, the absence of preTCR–pMHC interactions leads to deviant thymocyte transcriptional programming associated with dedifferentiation. Highly proliferative double-negative and double-positive thymocyte subsets emerge, with antecedent characteristics of T cell lymphoblastic and myeloid malignancies. Compensatory upregulation of diverse MHC class Ib proteins in *B2m*/*H2-Ab1* MHC-knockout mice partially safeguards in vivo thymocyte progression, although disseminated double-positive thymic tumours may develop with ageing. Thus, as well as promoting β chain repertoire broadening for subsequent αβ T cell receptor utilization, preTCR–pMHC interactions limit cellular plasticity to facilitate normal thymocyte differentiation and proliferation that, if absent, introduce developmental vulnerabilities.

The αβ T cell repertoire consists of many millions to billions of T cells uniquely expressing diverse surface T cell receptors (TCRs) in a clonal manner[11–13]. These cells mediate the precise recognition and elimination of aberrant host cells displaying 'foreign' surface pMHC ligands as a result of infection or cellular transformation. The repertoire of clonotypic TCRαβ receptors and their predecessor preTCRs is generated in the thymus of jawed vertebrates during fetal, neonatal and juvenile life[6]. Thymic progenitors originating from the bone marrow (and fetal liver in utero) proliferate during the early $CD4^-CD8^-$ double-negative (DN1 and DN2) stages and, under the influence of Notch at DN2, commit to the T cell lineage[1] (Fig. 1a). Progression to the $CD44^-CD25^+CD28^{lo}$ DN3a compartment leads to further αβ T cell lineage commitment with recombination-activating genes 1 and 2 (*Rag1* and *Rag2*), supporting TCRβ locus rearrangements that produce a recombined β chain that is expressed as a disulfide-linked heterodimer with the invariant pTα subunit[14]. In turn, pTα–β associates with CD3 signalling subunits. Upon preTCR signalling at the β-selection checkpoint, the DN3b population ($CD44^-CD25^+CD28^{hi}$) undergoes a critical programme change to suppress Notch signalling, downregulate transcription of *Rag1*, *Rag2* and *Ptcra*, increase cell cycling, and mediate allelic exclusion at the TCRβ locus, enforcing expression of only one TCRβ chain per cell[6]. In turn, those thymocytes transition to DN4 ($CD44^-CD25^-$) and then immature CD8 single-positive (ISP) compartments[15]. Upon further progression to the $CD4^+CD8^+$ double-positive (DP) stage, Rag genes are upregulated for a second time, enabling recombination and transcription at the TCRα locus and thereafter expression of the TCRαβ heterodimer[16–18]. To refine the αβ T cell repertoire, both positive and negative selection events ensue at this DP stage in the thymic cortex and continue into the maturing single-positive ($CD4^+CD8^-$ or $CD4^-CD8^+$) medullary compartment, followed by their later export as peripheral T cells[19].

[1]Laboratory of Immunobiology, Dana-Farber Cancer Institute, Boston, MA, USA. [2]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA. [3]Department of Medicine, Harvard Medical School, Boston, MA, USA. [4]Department of Dermatology, Harvard Medical School, Boston, MA, USA. [5]Belfer Center for Applied Cancer Science, Dana-Farber Cancer Institute, Boston, MA, USA. [6]Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA. [7]Department of Biomedical Engineering, Texas A&M University, College Station, TX, USA. [8]Department of Materials Science and Engineering, Texas A&M University, College Station, TX, USA. [9]Department of Physics and Astronomy, Texas A&M University, College Station, TX, USA. [10]Department of Chemical and Biomolecular Engineering, Vanderbilt University, Nashville, TN, USA. [11]Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, TN, USA. ✉e-mail: jonathan_duke-cohan@dfci.harvard.edu; ellis_reinherz@dfci.harvard.edu

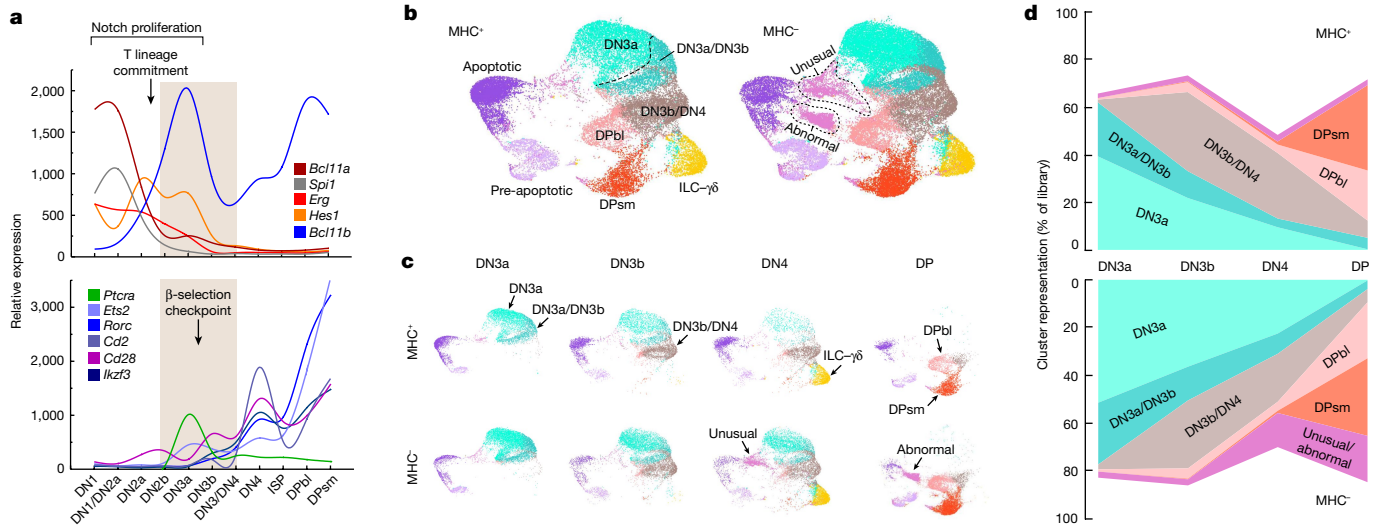


Fig. 1 | Developmental trajectories for thymocyte-like development on MHC$^+$ or MHC$^-$ supporting stroma. a, Schematic depicting representative gene transcript levels during key thymocyte developmental transitions (based on array data from the Immunological Genome Project[26]). Top, early DN1–DN3 proliferation is driven by thymocyte Notch signalling, represented here by the *Erg* and *Hes1* transcripts. Myeloid development is suppressed by downregulation of *Spi1* (encoding PU.1) during the DN2a to DN2b transition and T lineage commitment following the *Bcl11a* to *Bcl11b* switch. Bottom, following entry to the DN3 stage, the preTCR with invariant pTα (pTCRα) is expressed. PreTCR signalling downregulates Notch-driven proliferation, inhibits TCR β locus recombination, downregulates *Ptcra* and upregulates the indicated transcripts. **b**, UMAP projection of *k*-means clustering ($k = 10$) for DN3a, DN3b, DN4 and DP libraries for cells developing on either MHC$^+$ or MHC$^-$ stroma. All libraries are projected into the same space to enable direct comparison. The process for assignment of labels to each cluster is defined in the text. **c**, Cluster developmental trajectories of individual libraries and the relationship of individual clusters to phenotypically characterized thymocyte subsets. Projection of the individual FACS-sorted libraries (labelled along the top) into the primary space allows initial assignment of clusters expressing distinct transcriptomes. **d**, MHC$^-$ thymocyte-like cells progress developmentally by phenotype from DN3a to the DP stage but with altered distribution compared with MHC$^+$ cells. Focusing on the TCRαβ lineage, ILC–γδ cells, pre-apoptotic and apoptotic cells are excluded. The proportion of each defined developmental cluster is depicted for each library. $P < 2.5 \times 10^{-7}$ for the difference between MHC$^+$ and MHC$^-$ cells and stage distributions (Chi-square statistic).

Previous investigations have suggested that PreTCR signalling is independent of ligand recognition at the DN3 stage[20–23]. First, ablation of the TCR β chain variable domain that forms part of the interaction surface with pMHC on the TCRαβ receptor did not affect development through the DN3a to DN3b checkpoint. Second, a preTCR missing the extracellular domains of both the β chain and the pTα chain could drive development to the DP compartment. Third, in MHCI$^-$MHCII$^-$ double-knockout mice, thymocyte progression was unimpaired through the DN3 stage to the DP stage with respect to both cell numbers and phenotypes.

Recent structural and biophysical data, however, reveal direct interactions between preTCRs and pMHC ligands that utilize a horizontal binding mode that is compatible with facile mechanosensing[8,10,24]. Functional assays demonstrate both restricted proliferation and repertoire development in the absence of stromal MHCI and MHCII molecules[7,8]. Together these findings necessitate re-examination of the earlier results.

Early T lineage differentiation

Using an in vitro model of thymocyte differentiation, we seeded haematopoietic stem cells (HSC) from fetal liver of wild-type C57Bl/6 mice onto OP9-DL4 MHCI (MHC$^+$) stromal support cells or the same cells rendered MHCI-negative by CRISPR–Cas9 targeting of *B2m* and *Tap2* genes[7] (MHC$^-$). Both stromata lack endogenous MHCII expression. Extensive use of this model demonstrates synchronized expansion and development to the ISP and DP stage within the day 8 to day 13 window, thus recapitulating embryonic development[25].

To examine the TCR β chain selection checkpoint at the DN3a to DN3b transition, we seeded 3.2×10^4 fetal liver-derived HSC onto MHC$^+$ or MHC$^-$ stroma and quantified the developing thymocyte-like cells at day 9 (6.225×10^7 cells on MHC$^+$ stroma and 3.375×10^7 on MHC$^-$ stroma). For brevity, we refer to cells generated on MHC$^+$ and MHC$^-$ stroma with the prefix MHC$^+$ or MHC$^-$, respectively. Cells were sorted by FACS into DN3a, DN3b, DN4 and DP populations (Extended Data Fig. 1) and processed for single-cell RNA sequencing analysis (scRNA-seq) using the 10X Genomics Chromium system, simultaneously preparing from each cell bar-coded TCR α and β chain clonotype transcripts linked to that cell during downstream gene expression analysis. To reduce the dimensionality of the transcriptome information, all libraries were aggregated and projected into a single uniform manifold approximation and projection (UMAP) plane to enable direct comparison of clusters and inferred trajectory analysis incident to the FACS sorting by phenotype (Fig. 1b and Supplementary Files 2 and 3). To objectively delineate the relation of each cluster to thymocyte developmental stage, the dominant markers of normal transition from the DN3a stage to the mature DP small cell (DPsm) stage were extracted as reference arrays (Extended Data Fig. 2) from the Immunological Genome Project (https://www.immgen.org/) αβ T cell lineage database[26] and applied to each cluster, yielding a transcriptome reference trajectory that matched with relative cluster representation in each stage-specific library (Fig. 1c). Within the DN4 libraries, a population with a γδ T cell-like and innate lymphoid cell-like (ILC–γδ) transcriptome signature partitions owing to a lack of CD44 and CD25 expression (Extended Data Fig. 3a), pointing to the developmental fidelity of this in vitro system.

The pro-apoptotic (Extended Data Fig. 3b) and apoptotic populations (transcripts of predominantly mitochondrial origin) were retained as topological markers and to highlight the possibility, given the absence of thymic reticuloendothelial cells removing damaged cells, that in this assay apoptosis may be an important process even before negative selection events occurring at DP stages and beyond. The DN3a/DN3b cluster (Fig. 1c) shows early upregulation of *Ikzf3* and *Cd28*, markers of preTCR signalling (Fig. 1a and Extended Data Fig. 2a) and bridges the DN3a and DN3b libraries. Similarly, the DN3b/DN4 cluster is represented in the DN3b, DN4 and DP libraries, indicating the increased

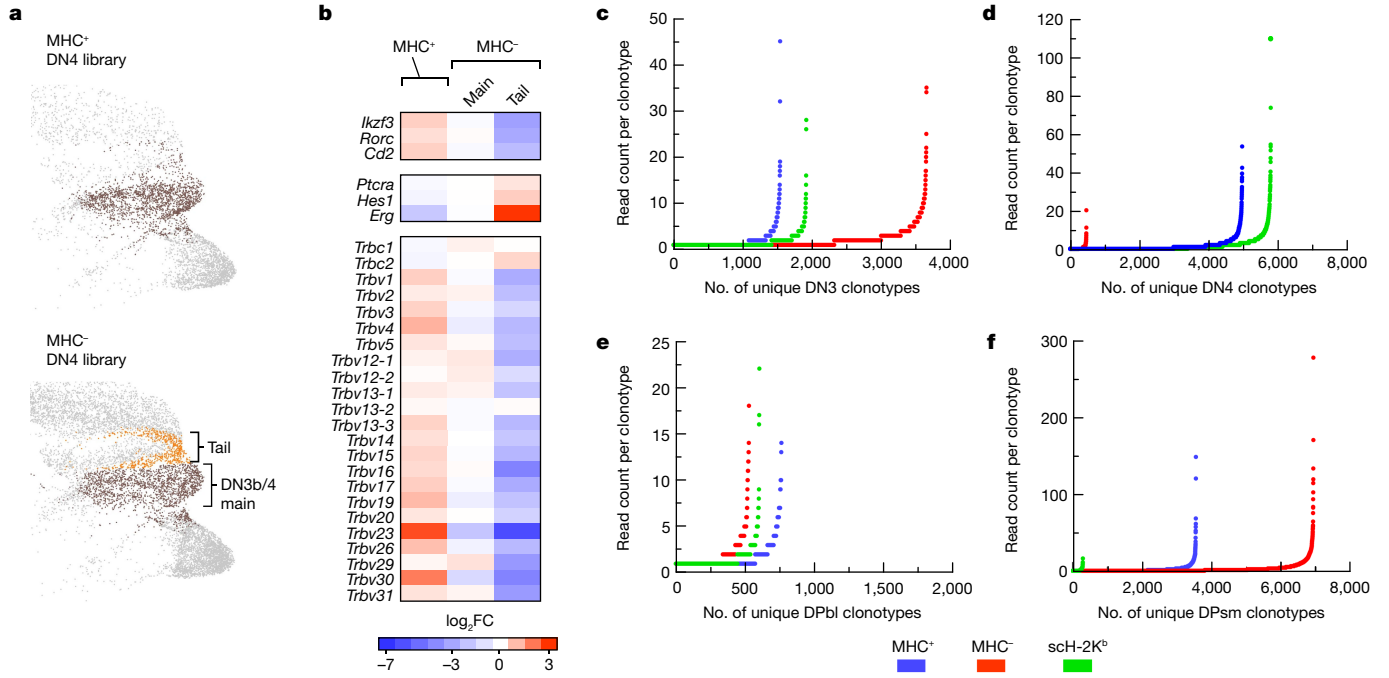


Fig. 2 | Uncoupling of the transcriptome and TCR β chain repertoire from phenotype in thymocyte-like cells developing on MHC⁻ stroma. a, The DN3b/DN4 cluster in the MHC⁻ DN4 library harbours a population with characteristics of cells not having passed through the preTCR signalling checkpoint. The MHC⁻ DN3b/DN4 cluster in the DN4 library is split into two subclusters, one corresponding to the DN3b/DN4 cluster in the MHC⁺ DN4 library (main, brown) and one corresponding to a set poorly represented in the MHC⁺ DN4 library (tail, orange). **b**, Transcript expression in MHC⁺ and MHC⁻ DN3b/DN4 cells. Transcripts that are well expressed and that mark the transition to DN3b/DN4 cells (*Ikzf3*, *Rorc* and *Cd2*; Extended Data Fig. 2b) remain low in the MHC⁻ tail subcluster, transcripts that are expected to be downregulated (Fig. 1a) remain high, and robust TCR β chain upregulation is not observed (Extended Data Fig. 2b). FC, fold change. **c**–**f**, Stage-specific analysis of β chain clonotype representation per 10,000 cells in DN3 (**c**), DN4 (**d**), DPbl (**e**) and DPsm (**f**) clusters from day 9 MHC⁺, MHC⁻ and scH-2K^b OP9-DL4 thymocyte-like development cultures. Representative of 6 experiments examining MHC⁺ ($n = 5$), MHC⁻ ($n = 6$) and scH-2K^b ($n = 3$) stromal cells.

resolution over phenotype provided by the transcriptional signature (Fig. 1c and Extended Data Fig. 2b). The immature DP blast cell (DPbl) population segregates away from the mature DPsm population on the basis of the strong representation of cell cycle-related transcripts (Extended Data Fig. 2d).

Absence of MHC affects preTCR signalling

Having established the cluster signature trajectory in normal developmental progression, we examined development in the MHC⁻ state (Fig. 1b,c). The MHC⁺ trajectory for the DN3a cluster shows a clear diminution, with progression from the DN3a to the DN4 libraries (Fig. 1c, top row and Fig. 1d). In the MHC⁻ state, there is significantly less progression—more than 36% of the phenotypically DN3b cells and more than 22% of the phenotypically DN4 cells retain a DN3a-like transcriptome, in contrast to 22% and less than 10%, respectively, in the MHC⁺ conditions (Fig. 1c,d). Nonetheless, there is phenotypic developmental progression in the absence of potential pMHC ligand binding to the preTCR. The DN3a to DN3b transition is marked by a new transcriptional programme characterized by the upregulation of *Ikzf3*, *Rorc*, *Cd2* and *Cd28* and the downregulation of *Hes1*, *Erg* and *Ptcra* (Fig. 1a and Extended Data Fig. 2b). We applied this gene panel to a subset of the DN3b/DN4 cluster that is more strongly represented in the MHC⁻ DN4 library than in the control condition, highlighted as a 'tail' moving back into the DN3a/DN3b cluster (Figs. 1b and 2a). Splitting the MHC⁻ DN4 cluster into two subclusters—one representing the main region overlapping in position with the MHC⁺ DN3b/DN4 cluster and the other representing the tail—showed clear differences. The tail, despite being phenotypically DN4, did not exhibit upregulated *Ikzf3*, *Rorc* or *Cd2* or downregulated *Hes1* and *Erg* as observed in the MHC⁺ cluster and, further, did not show robustly upregulated Trbv transcription (Fig. 2b and Extended Data Fig. 2b). Collectively, these observations are consistent with a differentiation trajectory that bypasses the β selection checkpoint. The main MHC⁻ DN3b/DN4 cluster shows an intermediate expression between the MHC⁺ DN3b/DN4 cluster and the tail, suggesting that elements of the aberrant transcriptional regulation observed in the tail subcluster extend to the main subcluster.

Reduced DN4 β clonotypic diversity

We further examined appropriate developmental regulation of Trbv gene transcription and repertoire diversity at the DN4 stage by β chain clonotype analysis of the developing MHC⁺ and MHC⁻ subpopulations using targeted RNA sequencing (RNA-seq). Wild-type HSC were seeded onto MHC⁺ OP9-DL4 stromal cells, MHC⁻ OP9-DL4 stromal cells or the same MHC⁻ cells transfected to re-express MHC class I as a single chain comprising VSV8 peptide, β_2 microglobulin (β2m) and H-2K^b (scH-2K^b). The scH-2K^b derivative expresses multiple copies of a single pMHC, thus maintaining the potential for the horizontal binding mode to the preTCR but presenting a homogenous peptide—RGYVYQGL—derived from amino acids 52–59 of vesicular stomatitis virus nucleoprotein[7]. After nine days, cell proliferation was uniformly higher on the MHC⁺ stromal cells than on the MHC⁻ or scH-2K^b support stroma (Extended Data Fig. 4a,b). Cells from each support stroma culture were sorted into phenotypically defined DN3, DN4, DPbl and DPsm populations (Supplementary File 3) and Trbv clonotypes of 10^4 cells for each stage and condition were identified by targeted RNA-seq.

TCRβ clonotype diversity is high at the DN3 stage for cells developing on all variants of the OP9-DL4 support stroma used here (Fig. 2c). The DN4 compartment reveals a consistently contracted repertoire

diversity only on the MHC$^-$ support stroma (Fig. 2d and Extended Data Fig. 4c). Up to 70% of the clonotypes developing in the MHC$^-$ DN4 population were found to make up less than 7.5% of MHC$^+$ and scH-2K^b populations (Extended Data Fig. 4d), suggesting that these clonotypes may represent a restricted population of clonotypes responding to non-classical MHC or MHC-unrelated structures on the stromal surface. Conversely, around 92% of the clonotypes expressed on cells developing on the MHC$^+$ and scH-2K^b stroma, which enable preTCR–pMHC interaction, were absent in the MHC$^-$ cultures. The limited MHC$^-$ clonotype repertoire was not a consequence of restricted cell proliferation, since the clonotype diversity of DN4 cells developing on scH-2K^b stroma was as rich as that of the cells developing in the MHC$^+$ condition (Fig. 2d and Extended Data Fig. 4c), despite similar cell representation of all 3 DN4 cell populations (10^4 cells analysed per sample). The characteristics of cells developing on MHC$^-$ stroma or scH-2K^b stroma both diverged from those on the MHC$^+$ stroma during the DP stage (Fig. 2e,f and Extended Data Fig. 4c). Cells developing on scH-2K^b stroma revealed a contraction of β-repertoire diversity, probably linked to limited positive selection afforded by a single peptide (that is, VSV8) on scH-2K^b stroma. Of note, the N15β clonotype with known specificity for VSV8 peptide presented by H-2K^b appears in the top 20 DPsm clonotypes developing on the scH-2K^b stroma (Extended Data Table 1). Conversely, the MHC$^-$ developing cells recovered diversity at the DPbl and DPsm stages, often exceeding the diversity of cells on the MHC$^+$ stroma (Fig. 2 and Extended Data Fig. 4c) and indicating aberrant β chain transcriptional regulation when MHC-dependent preTCR signalling was circumvented. Continued Notch stimulation in the absence of preTCR signalling has already been demonstrated to permit differentiation through to the DP stages[27].

Origin of β diversity in the MHCIa$^-$ system

The development of TCR clonotypes in the MHC$^-$ condition implies that thymocytes can develop and bypass the preTCR checkpoint in the absence of MHC, either via a ligandless mode or using non-classical MHCI and MHCII molecules or additional ligands. We compiled a panel of non-classical MHCI (MHCIb) (Extended Data Table 2)[28] and—following full transcriptomic analysis of the OP9 MHC$^+$ and OP9 MHC$^-$ stromal cells (Extended Data Fig. 5a–c)—examined the expression of non-β2m-dependent MHC. Loss of CD1d surface expression, which is dependent on β2m, was used as a functional validation marker of the CRISPR–Cas9 knockout in addition to loss of MHCI (Extended Data Fig. 5d), thus supporting our focus on non-β2m-dependent MHC. Transcriptome analysis identified only *Raet1d* and *Raet1e* as being expressed at the transcriptome level with detectable surface protein expression, but with no difference between MHC$^+$ OP9-DL4 and the MHC$^-$ OP9-DL4 variant (Extended Data Fig. 5e). Consequently, the origin of the 'background' clonotypes comprising the repertoire at the DN4 and subsequent stages in the MHC$^-$ condition—also found as a minor fraction of the total repertoires in the MHC$^+$ and scH-2K^b conditions (Extended Data Fig. 4d)—is uncertain but may involve non-MHC ligands or non-classical MHCIb ligands independent of β2m or the peptide-loading complex for cell surface expression.

Unusual DN4 cells develop in the absence of MHC

To further address the diminution in β chain representation at DN4 in the MHC$^-$ condition, examination of the scRNA-seq clustering is informative. Although the partitioning of the innate lymphoid cell (ILC)-like and γδ T-like cells within the DN4 represents a population low in β chain representation (Fig. 1c and Extended Data Fig. 3a), this is not the source of the difference, as the representation of this cluster is similar between the MHC$^+$ and MHC$^-$ conditions. Aside from the ILC-like and γδ T-like cells, the DN3b/DN4 cluster is the only other cluster with significant representation in the thymocyte developmental path in the MHC$^+$ DN4 library. These cells exhibit a robust upregulation of β chain transcript (264.6 ± 74.3-fold increase, median = 88.4; $P < 0.0001$) on transitioning from the DN3a/DN3b cluster (Extended Data Fig. 2b). By contrast, in the MHC$^-$ DN4 library, in addition to the DN3b/DN4 population, there remains a high representation of phenotypically defined DN4 cells with a DN3a-like transcriptome as well as a distinct population that is barely observed in the MHC$^+$ condition ('unusual'; Fig. 1c,d). As described above, compared with the MHC$^+$ DN4 library, in the MHC$^-$ DN4 library Trbv gene expression in the DN3b/DN4 main population trends toward suppression (Fig. 2b)—the DN3b/DN4 tail exhibits a significant suppression (7.65-fold down against MHC$^+$ DN3b/DN4, $P < 0.0002$; 4.48-fold down against the MHC$^-$ DN3b/DN4 main cluster, $P < 0.0025$), as do the DN3a-like cells (5.38-fold down against MHC$^-$ DN3b/DN4, $P < 0.0001$) and the MHC$^-$ unusual DN4 cluster (Fig. 3a and Extended Data Fig. 5f). The aggregate effect of all these phenomena may contribute to the low DN4 Trbv clonotype representation in the MHC$^-$ condition (Fig. 2d and Extended Data Fig. 4c).

The MHC$^-$ unusual cluster (1,776 cells, 14.3% of all DN4 cells), which is minimally represented in the MHC$^+$ DN4 library (205 cells, 2.79% of all DN4 cells), displays a complex transcriptome. Unlike the DN3b/DN4 cells expected in the DN4 library, the unusual cluster cells have not consolidated the robust expression of β chains (Fig. 3a and Extended Data Fig. 5f). Nevertheless, 83.6% of the cells in the DN4 unusual cluster express *Trbc1* and/or *Trbc2* transcripts and of these, 70.7% express *Lck* and/or *Ptcra*, confirming the T lineage origin of a large fraction of the cells (Supplementary File 2). Moreover, there is maintained expression of progenitor driver genes (*Kit*, *Lyl1*, *Ezh1* and *Id2*) as well as *Spi1*—encoding PU.1, which operates at the critical decision checkpoint determining myeloid or T cell lineage specification. These observations are consistent with the cells not having passed through the preTCR checkpoint, as is the maintained expression of early lineage and γδ T cell-linked developmental transcripts such as *Fcer1g*, *Icos* and *Il18rap* (Extended Data Fig. 3a). The high representation of these unusual cells is not part of normal ILC or γδ T cell development, otherwise they would also appear in the MHC$^+$ DN4 library that harbours a similar ILC–γδ T cell cluster. Furthermore, the MHC$^-$ unusual cells are in a cycling state with high histone transcript expression and high expression of *AY036118* (Fig. 3a,b), a long noncoding (XR_877120.4) on Chr17 implicated in regulation of thymocyte proliferation possibly mediated by telomeric association[29,30]. The volcano plot identifies transcripts of high fold change and probability averaged across the whole cluster—thus, significance may be driven by a well-represented subset of cells rather than the complete cluster population. Examining transcripts that are regulated in the same direction in most cells within a cluster in addition to *AY036118*, histone genes—represented here by *Hist1h1d* (also known as *H1f3*) and *Lars2*, which encodes mitochondrial leucyl-tRNA synthetase 2, a marker of high metabolic activity—stand out[31] (Fig. 3b).

Unexpectedly, this analysis led to identification of irregularities in *Rag1* and *Rag2* transcription in the DN4 unusual population, where these transcripts are minimally expressed (Fig. 3b). *Rag1* is well expressed in MHC$^+$ and MHC$^-$ DN3b/DN4 clusters. *Rag2* is expressed well in the MHC$^+$ DN3b/DN4 cluster, whereas expression in the MHC$^-$ DN3b/DN4 is similar to that in the DN4 unusual cluster. These findings illuminate possible differential regulation of *Rag1* and *Rag2* transcripts and show that the MHC$^-$ DN3b/DN4 cells are already experiencing transcriptional aberrations despite appearing phenotypically identical to MHC$^+$ DN3b/DN4 cells. Reduction of RAG1–RAG2 heterodimeric protein activity in the DN4 unusual population owing to regulation of *Rag2* transcripts might contribute to the loss of diversity in the β chain repertoire at this stage (Fig. 2d).

Further refinement of the properties of the MHC$^-$ DN4 unusual cluster are revealed by single-cell β clonotype analysis. Examination of the MHC$^+$ DN4 library for the top 20 clonotypes based on cellular representation (Extended Data Table 3a) shows that the majority localize to the DN3b/DN4 cluster as expected, given appropriate preTCR signalling, with minimal tracking to other clusters (Fig. 3c). Similar analysis of the MHC$^-$ DN4

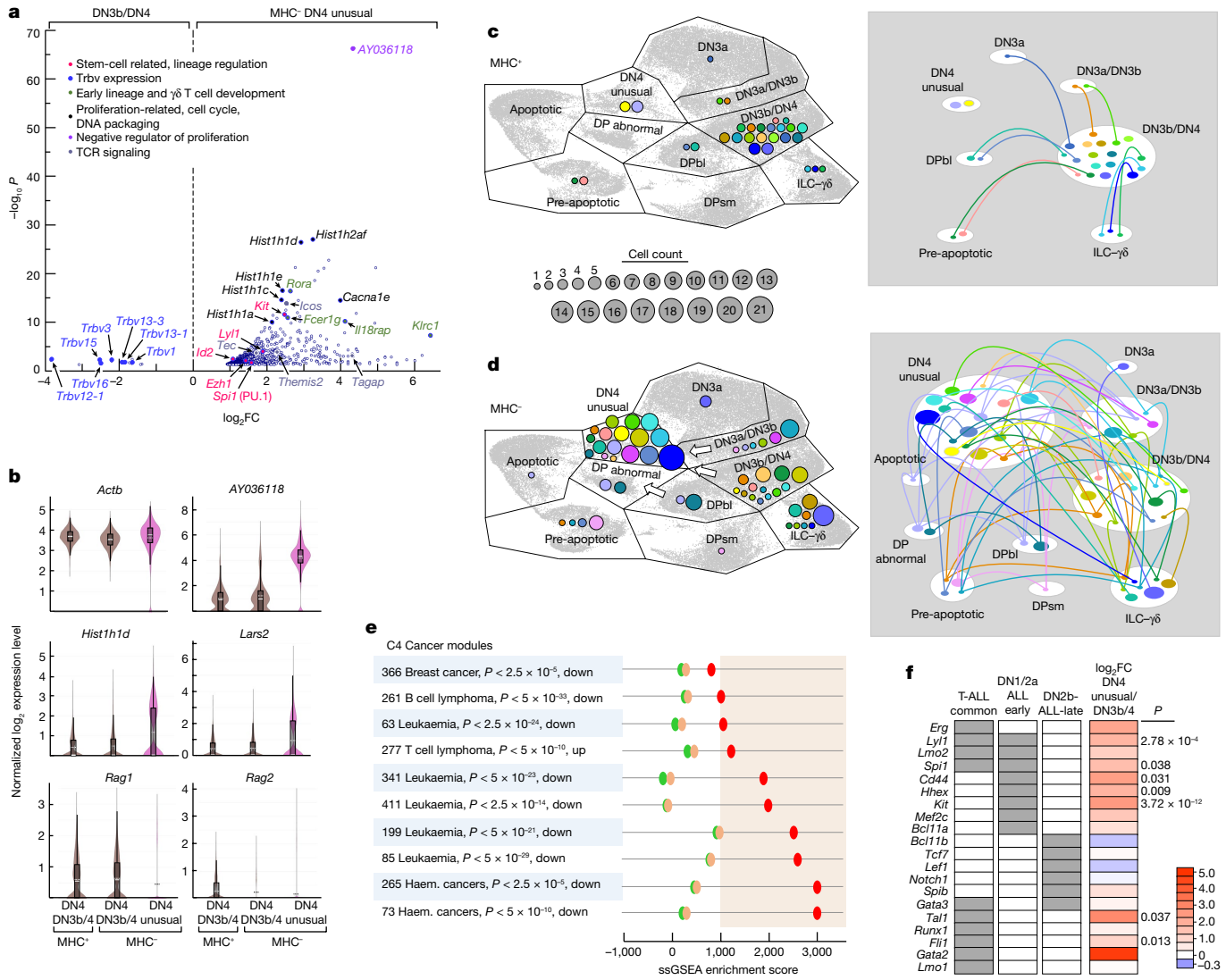


Fig. 3 | Single-cell transcriptomics of the MHC⁻ DN4 unusual cluster reveal complex proliferative and lineage abnormalities. a, Volcano plot of genes with significantly different transcript levels ($P < 0.05$) between the DN4 unusual population and the DN3b/DN4 cluster in the same MHC⁻ DN4 library. Upregulated DN4 unusual transcripts are shown to the right of zero on the *x*-axis, and downregulated transcripts are shown to the left. Inset, the functional significance of highlighted transcripts. Fourteen other Trbv gene transcripts were downregulated but did not meet the $P < 0.05$ threshold. **b**, Violin plots depicting selected log-normalized expression levels in the MHC⁺ DN4 library DN3b/DN4 cluster (DN3b/4) (2,021 cells), the MHC⁻ DN4 library DN3b/DN4 cluster (2,493 cells) and the MHC⁻ DN4 unusual cluster (1,776 cells). In box plots, the box delineates the first to third quartiles, the dotted line represents the mean, the solid line represents the median, the whisker above the box extends to the third quartile plus 1.5 times the interquartile range and the whisker below the box extends to the first quartile minus 1.5 times the interquartile range. Cluster colours are as depicted in Figs. 1 and 2. *P* values as in **a**. **c**, Cluster distribution of the top 20 DN4 clonotypes developing on MHC⁺ stroma (left) with the expected developmental trajectory to the DN3b/DN4 cluster (right). Each colour represents a unique clonotype and the circle diameter is proportional to the number of cells expressing that specific β chain. Right, clonotype tracking from cluster to cluster, with cluster and track colouring concordant with the scheme on the left; colour and clonotype specification is unique to this panel. **d**, Distribution of the top 20 DN4 clonotypes developing on MHC⁻, stroma similar to the depiction in **c**. Right, clonotype tracking from cluster to cluster. The cell count scale is as in **c**; colour and clonotype specification is unique to this panel. **e**, ssGSEA scores for MHC⁺ DN3b/DN4 (green), MHC⁻ DN3b/DN4 (light orange) and MHC⁻ DN4 unusual (red) clusters compared with gene subset modules exhibiting co-ordinated up- or downregulation in the indicated cancers (MSigDb C4). Haem., haematological. **f**, Comparison of the DN4 unusual cluster developing on MHC⁻ stroma with the DN3b/DN4 cluster developing on DN4 MHC⁺ stroma (positive control) for transcripts reported as dysregulated in human T-ALL. Shaded boxes identify regulatory transcripts that are overexpressed in the indicated forms of human T-ALL (columns 1–3). The heat map in column 4 depicts the $\log_2$ fold difference for the DN4 unusual cluster compared with the MHC⁻ DN3b/DN4 cluster and *P* value (right). DN1/2a, DN1/DN2a.

library exposes a starkly different distribution, with the majority of highly represented β clonotypes mapping to the DN4 unusual cluster (Fig. 3d). For 14 out of the 17 clonotypes represented in the DN4 unusual cluster, we can identify related cells bearing the same clonotype in the DN3a/DN3b and DN3b/DN4 clusters (Fig. 3d, right). Consequently, we propose that in the absence of pMHC, some cells may differentiate from DN3a to DN4 but deviate from the normal transcriptome trajectory to map to the unusual cluster. Cell representation of the top 20 clonotypes in the DN4 libraries, normalizing for differences in initial library size, shows 3.25 ± 0.55 cells for each MHC⁺ clonotype (only 2 out of 20 found in the unusual cluster) compared with 6.56 ± 2.07 cells for each MHC⁻ clonotype (17 out of 20 in the DN4 unusual cluster, $P < 0.0001$). This confirms the increased proliferation implied by the transcriptome signature of the MHC⁻ developing cells in this unusual cluster.

Eight of the top twenty clonotypes are found in the ILC–γδ T cluster and five of these are shared with the MHC$^-$ DN4 unusual cluster, implying that T lineage developmental options may remain open without delivery of appropriate preTCR–pMHC-dependent regulatory signals. Of interest is the observation that cells expressing the same unique β clonotype—particularly MHC$^-$ cells—tend to group closely together within the UMAP cluster, implying conservation of the transcriptional signature, even for occasional clonotypes split between clusters (Extended Data Fig. 6).

The transcript signature of the DN4 unusual population, with upregulation of early progenitor proliferative genes and of *Spi1* controlling the myeloid–T lineage decision point at the DN2a–DN2b transition, connotes a dedifferentiation of the DN4 cells in the absence of appropriate preTCR signalling. To examine the possibility that this uncommon transition may generate a transcriptional landscape consistent with aberrant transformation potential, we performed single-sample gene set expression analysis (ssGSEA) against cancer modules followed by more refined comparisons with clinically defined T cell acute lymphoblastic leukaemia (T-ALL) gene sets. By ssGSEA analysis, the DN4 unusual cluster shows a strong score (>1,000) against 9 out of the top 10 modules defined by maximal score difference from the DN3b/DN4 clusters of both the MHC$^+$ and MHC$^-$ DN4 libraries (Fig. 3e). Leukaemia and/or lymphoma transcriptomes show significant co-ordinated regulation with all nine of these gene set modules. By contrast, the DN3b/DN4 clusters of both MHC$^+$ and MHC$^-$ DN4 libraries tracked together and showed weaker association or even inverse correlation. Further refinement of this analysis compared expression in the DN4 unusual population with published transcript panels defining T-ALL focusing on early T cell precursor acute lymphoblastic leukaemia (ETP-ALL), a subset of T-ALL with poor prognosis in humans that is believed to develop from early thymic progenitors immigrating from the bone marrow[32,33]. The selected transcripts were grouped as being common to T-ALL generally, representing DN1/DN2a ETP-ALL prior to committing to the T lineage ('early') or DN2b ETP-ALL after commitment to the T lineage ('late'). Transcript representation within the MHC$^-$ DN4 unusual cluster subsequently was compared with that in MHC$^+$ DN3b/DN4 cells following the expected developmental trajectory (Fig. 3f). Seven out of ten transcripts representing the common panel trended towards upregulation, whereas three showed no change. Except for the weakly upregulated *Spib*, none of the transcripts in the DN2b ETP-ALL late panel were upregulated. By contrast, five out of the eight selected genes in the DN1/DN2a ETP-ALL early panel were significantly upregulated, and the remaining genes trended upwards, consistent with the cells dedifferentiating from a DN4 state back towards the early progenitor state. As no significant differences were noted in CDR3 length or hydropathy between DN4 β-chain variable domain (Vβ) clonotypes developing on MHC$^+$ versus MHC$^-$ stroma (ref. [8] and Supplementary File 3), the abnormal transcriptome probably emanates from a lack of preTCR ligation by MHC and not from aberrant preTCR sequences per se.

The abnormal DP subset and dedifferentiation

The DN4 unusual cluster forms one section of a bipartite UMAP cluster that also includes a unique population found only in the MHC$^-$ DP thymocyte-like library, leading to its classification here as 'abnormal' (Fig. 1b,c). This DP population projects away from the DN4 component owing to the expression of *Cd4*, *Cd8a* and *Cd8b1* (Extended Data Fig. 5g), but maps to the same DN4 cluster projection owing to the strong expression of *AY036118*, histones, early progenitor-related transcripts, *Spi1* driving non-T lineage commitment in the early double-negative (DN) stages and markers not strongly expressed in the MHC$^+$ developing DPbl or DPsm clusters (Fig. 4a). Of note, the most significantly upregulated transcripts in this DP abnormal population are transcripts that define the myeloid lineage: *Mpo* (myeloperoxidase), *Prtn3* (proteinase 3), *Ctsg* (cathepsin G) and *Elane* (neutrophil elastase)—their expression is specific to this cluster without expression in any of the DN3a to DPsm clusters representing the expected developmental trajectory or in the MHC$^-$ DN4 unusual cluster (Extended Data Fig. 5h). Selecting transcripts that are upregulated throughout the DP abnormal cluster confirms the signature *AY036118* profile, cell cycling and DNA packaging using *Hist1h1d* as representative of a broad spectrum of histones, *Plac8* as an oncogenic driver, and the key myeloid markers *Prtn3* and *Mpo* (Fig. 4b). *SPI1*, *LYL1*, *LMO2* and *MEFC2* are dominant components of a panel defining human ETP-ALL[34] and the mouse homologues are upregulated in the DP abnormal cluster, where *Lyl1* and *Spi1* are upregulated above the level seen in the DN4 unusual cluster (Figs. 3a and 4a). Supporting an origin from the αβ T cell lineage, the *Spi1*$^+$ cells in the MHC$^-$ DP cluster express Vβ region transcripts (Fig. 4g) with fully recombined clonotypic TCR β chains in more than 37% of those cells (Extended Data Fig. 5i).

To investigate the possibility that the DP abnormal cluster arose from aberrant expansion of one HSC in the progenitor pool, we determined the fraction of Y chromosome-positive cells in each cluster. The results do not support stochastic growth independent of stromal cell MHC expression (Supplementary File 6). Further strengthening the proposal that the DP abnormal cells are following a path that deviates from the wild-type pathway, 80% of the *Mpo*$^+$ cells and 84% of the *Mpo*$^+$*Spi1*$^+$ cells co-express *Lck* and/or *Cd3e* (Extended Data Fig. 5j). Confirmation that the expression of myeloid transcripts occurred in T lineage-committed cells and was not due to contaminating myeloid progenitors was derived independently via a separate set of experiments examining DP cells developing from purified DN3a and DN4 cells in vitro (Extended Data Fig. 5k–m and Supplementary File 4). We detected significantly higher levels of *Mpo* and *Spi1* transcripts in MHC$^-$ DP cells sorted by fluorescence-activated cell sorting using quantitative PCR with reverse transcription (RT–qPCR) (Extended Data Fig. 5l,m).

Both natural killer (NK) T cells and mucosal-associated invariant T (MAIT) cells develop from the DP population, but—based on two orthogonal findings in our data—there is no evidence that these cells develop via an alternative path to canonical αβ T cells in the absence of pMHC ligation. First, their respective transcriptional signatures do not map to the DP abnormal population (*Rorc*, *Tbx21* and *Gata3* for NK T cells; *Zbtb16*, *Drosha* and *Il18* for MAIT[35], although *Mr1* is twofold upregulated). Second, all three TCR β chains restricted to mouse NK T cells (*Trbv1*, Trbv13 alleles and *Trbv29*) are downregulated in the DP abnormal population, whereas those β chains restricted to mouse MAIT cells are downregulated (Trbv13 alleles) or unchanged[35] (*Trbv19*) (Fig. 4a).

As observed for the DN4 libraries, the distribution of the top 20 clonotypes by cell representation was markedly different between the MHC$^+$ DP library and the MHC$^-$ DP library (Extended Data Table 3b). The average cell representation of each clonotype was higher in the MHC$^-$ condition (Fig. 4c,d), and tracking showed that this difference was retained in the DPsm and pre-apoptotic clusters. Representatives of both the DP abnormal and of the DPsm clusters are found in the DPbl population, but there is minimal overlap between the DPsm and DP abnormal cells, implying that the cluster destiny is specified at the DPbl stage. Given the proliferative transcript signature, the early progenitor profile and the presence of myeloid markers, we examined the DP abnormal population by ssGSEA for evidence of entry into a state conducive to future myeloid dysplasia or leukaemia development (Fig. 4e). MSigDb C4 cancer module 489 (https://www.gsea-msigdb.org/gsea/msigdb/) generated the highest differential score, a cell profile that is strongly associated with leukaemias, including T-ALL and acute myeloid leukaemia (AML). The signature panel for regulatory gene abnormalities in CD34$^+$ leukaemic stem cells isolated from patients with AML overlaps completely with that for T-ALL[36] (Figs. 3f and 4f). An CD34$^+$ leukaemic stem cell panel[37] from patients with AML (LSC17) was used to assess any potential equivalence of the DP abnormal cells with transformed AML leukaemic stem cells (Fig. 4f). The DP abnormal cells expressed seven out of ten signature transcripts in the T-ALL/AML common panel

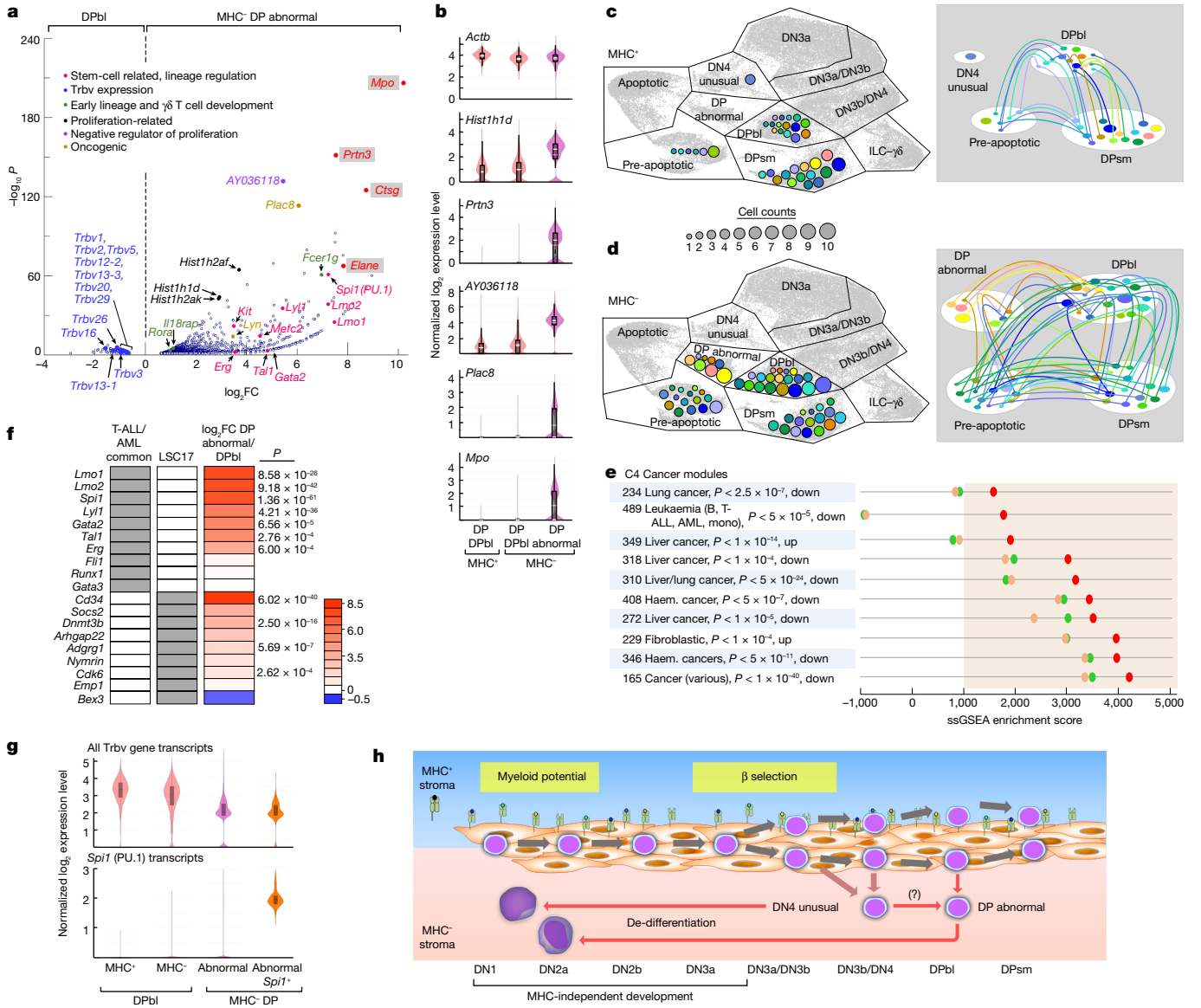


Fig. 4 | Single-cell transcriptomics of the MHC⁻ DP abnormal cluster reveal dedifferentiation and reprogramming to include a myeloid programme. **a**, Volcano plot of genes with significantly different transcript levels ($P < 0.05$) between the DP abnormal population and the DPbl cluster in the same MHC⁻ DP library. **b**, Violin plots depicting log-normalized levels of selected highly expressed genes in the MHC⁺ DP library DPbl cluster (1,212 cells), the MHC⁻ DP library DPbl cluster (3,044 cells) and the MHC⁻ DP abnormal cluster (2,512 cells). Cluster colours are as depicted in Figs. 1 and 2. *P* values are as in **a**. **c**, Cluster distribution of the top 20 DP clonotypes developing on MHC⁺ stroma (left) together with clonotype tracking (right). Cluster and track colouring on the right is concordant with the scheme on the left; colour and clonotype specification is unique to this panel. **d**, DP cluster distribution of the top 20 clonotypes developing on MHC⁻ stroma with right panel depicting clonotype tracking as in **c**. Colour and clonotype specification is unique to this panel. **e**, ssGSEA scores for MHC⁺ DPbl (green), MHC⁻ DPbl (light orange) and MHC⁻ DP abnormal (red) cells, analysing correlation with gene set modules with co-ordinated gene regulation in the indicated tumours (MSigDb C4). **f**, Comparison of the MHC⁻ DP abnormal cluster with the DP cluster developing on MHC⁺ stroma for transcripts reported as dysregulated in human AML (T-ALL/AML common) or as overexpressed in CD34⁺ leukaemic stem cells from patients with AML (LSC17). **g**, Co-expression of *Spi1* and Trbv transcripts in *Spi1*⁺ MHC⁻ DP abnormal cells. Cell numbers for MHC⁺ DPbl, MHC⁻ DPbl and MHC⁻ DP abnormal as in **b**, and 221 DP abnormal *Spi1*⁺ cells. **h**, Schematic of the proposed model for dedifferentiation of thymocytes developing in the MHC⁻ condition (pink area). In box plots, the box delineates the first to third quartiles, the dotted line represents the mean, the solid line represents the median, the whisker above the box extends to the third quartile plus 1.5 times the interquartile range and the whisker below the box extends to the first quartile minus 1.5 times the interquartile range.

at levels significantly higher than developing DPbl cells. The minimal change in *Gata3* and *Runx1* may indicate the ongoing T cell lineage programme in both subpopulations. Comparison with representatives of the human LSC17 panel found significant upregulation of four out of nine markers with a further three trending upward. *Cd34*, the canonical haematopoietic stem cell marker, was the most profoundly upregulated (293-fold). The high expression of *Cd34* coupled with persistence of *Erg* (Fig. 1a) in the MHC⁻ DP abnormal population, which was absent in the MHC⁺ libraries, points to an earlier progenitor environment in the absence of pMHC-driven preTCR signalling.

preTCR signalling in *B2m*/*H2-Ab1* dKO mice

We next examined gene expression in the thymus of MHC⁺ B6 mice and mice on the same background deficient in both *B2m* and for *H2-Ab1* (*B2m*/*H2-Ab1* dKO), previously created to abrogate expression of MHCI

and MHCII[21] (MHC$^-$). We tested whether the phenomena observed in vitro were recapitulated in vivo. We observed a significant increase in the number of DN3a cells in the MHC$^-$ thymi, a difference that extended less significantly through the DN3b to DN4 stages (Extended Data Fig. 7a). Examining gene expression for cells transitioning from the DN3a to the ISP stage (Supplementary File 4), we observed strong downregulation of *B2m* and moderate downregulation of *H2-Ab1* in H2-negative thymocytes (Extended Data Fig. 7b). However, we found no difference between the MHC$^+$ and MHC$^-$ thymocytes with respect to the transcript changes occurring in the DN3a to ISP transition depicted in Fig. 1a (Extended Data Fig. 7c). Moreover, there was no reduction in Trbv transcription at the DN4 stage in the MHC$^-$ thymocytes (Extended Data Fig. 7d). By contrast, the upregulation of anti-apoptotic *Bcl2a1* family transcripts—a defined hallmark of preTCR signalling—was clearly observed in the MHC$^+$ but not in the MHC$^-$ DN4 thymocytes (Extended Data Fig. 7e), whereas expression of canonical BCL2 pathway genes was similar in both[38]. The upregulation of Trav transcripts that are dependent on preTCR signalling was significantly greater in the MHC$^+$ mice than in MHC$^-$ mice[39] ($P = 2 \times 10^{-8}$; Extended Data Fig. 7f). Of note, the *Pim1* proto-oncogene associated with fetal haematopoiesis that is overexpressed in myeloid and lymphoid leukaemias[40] was one of the most highly expressed transcripts in the MHC$^-$ libraries, but was barely detected in the MHC$^+$ libraries (Extended Data Fig. 7e). Analysis of complete β chain repertoires for the entire thymus representation of cells from DN3a to ISP was uninformative; for each library more than 98.9% of the clonotypes were represented by 3 or fewer unique molecular identifiers (UMIs), leading to such high repertoire diversity scores that no significant differences were observed between libraries.

MHCIb upregulation in *B2m*/*H2-Ab1* dKO mice

Notably, in all *B2m*/*H2-Ab1* dKO libraries, both *H2-T3* (also known as TL) and *H2-T22* were markedly upregulated over those in B6 MHC$^+$ libraries (Extended Data Fig. 7b). By contrast, OP9-DL4 *H2-T22* expression was similar between the MHC$^+$ and MHC$^-$ cells and *H2-T3* was undetectable in either of the isogenic stroma (Extended Data Table 2). *H2-Q10* and *H2-T-ps* (which is now believed to be protein coding; NCBI Gene ID: 667803) were also notably upregulated in the MHC$^-$ libraries.

The enhanced transcription of *H2-T22*, *H2-T3*, *H2-Q10* and *H2-T-ps* genes implies that adaptation in vivo maintains functional β selection by upregulating non-classical minor MHCIb products, thereby compensating for loss of classical MHCIa alleles. This phenomenon is not operative in the OP9 cultures. Our mouse studies are consistent with the apparent normal phenotypic thymocyte development in *B2m*/*H2-Ab1* dKO mice observed previously[21], and underscore the complexity of vital in vivo biological signalling, including mechanisms to override pathway blockade via compensatory adaptation. Nonetheless, preTCR signalling is not entirely normal in this model, as evidenced by lack of upregulation of Trav and *Bcl2a1*, in agreement with the suggestion that the narrow width of the MHCIb α1α2-presenting platform relative to that of MHCIa might attenuate preTCR signalling[8].

Malignancy despite MHCIb compensation

Given the leukaemia and/or lymphoma-like transcriptome signatures of the DN4 unusual and DP abnormal clusters, we monitored the health of a cohort of mice for signs of cancer development as they aged. Of seven mice studied at 15 months of age, approximately half the lifespan of a B6 mouse, one developed significant weight loss, scruffy coat appearance, failure to thrive and was euthanized. The gross pathology depicted in Extended Data Fig. 7g reveals a massive thymus, spleen and lymph node, as well as hepatomegaly. FACS staining of organ-derived cell suspensions from this mouse revealed aberrant cell populations, including a CD4^{dull}CD8^{dull} DP subset and a CD8^{dull} single-positive subset in thymus, as well as a DP population in the spleen, presumably of thymic origin (Extended Data Fig. 7h). Haematoxylin and eosin staining of fixed tissue sections showed the thymus and spleen to be effaced by tumour, obliterating normal landmarks (Extended Data Fig. 7i). The liver disclosed tumour cells within distended sinusoids consistent with haematogenous spread. The bone marrow was also extensively replaced by acute lymphoid blast-like cells with a large nuclear-to-cytoplasmic ratio. Immunohistochemistry analysis showed that the tumour cells were positive for CD8 and the early T lineage marker TdT (Extended Data Fig. 7j), which, together with the distribution of tumour cells, is consistent with a thymic origin. The myeloid marker elastase was not detected, but the tumour was positive for activated NOTCH1 (indicated by NOTCH1 intracellular domain (NICD1) antibody) (Extended Data Fig. 7j).

preTCR–pMHC interactions safeguard development

Our in vitro study revealed that preTCR–pMHC interactions sculpt the transcriptome of DN3 and later-stage thymocytes, in addition to supporting β clonotype diversity in the αβ T cell lineage. Three irregular UMAP clusters were uncovered in culture on the MHC$^-$ stroma. The first, an aberrant DN3b/DN4 transitional population, lacked evidence of preTCR signalling but maintained Notch signalling and manifest a broad decrease in Trbv transcripts. The second, a DN4 unusual population, minimally present in the MHC$^+$ population, abnormally upregulated genes involved in earlier stages of thymic renewal (*Lyl1*, *Kit*, *Id2*, *Dtx1* and *Bcl11a*), T cell co-stimulatory function (*Icos*), adhesion function (*Itgb3*) and cytokine receptor genes involved in inflammation (*Il18r* and *Il23r*). The third, an entirely anomalous cluster, DP abnormal, expressed *Cd4*, *Cd8a* and *Cd8b1*, with a subset simultaneously expressing multiple myeloid genes (*Mpo*, *Prtn3*, *Ctsg*, *Elane*, *Hdc* and *Cst7*). Both highly proliferating DN4 unusual cells and DP abnormal cells expressed *AY036118*, which is implicated in the control of thymocyte proliferation[30].

Such aberrations of developmental programmes at DN and DP thymocyte stages are noteworthy given that human T-ALL represents aggressive malignancies of these same phenotypic subpopulations, including a subset of DN early T cell precursors[32,41]. Key human genetic abnormalities include instabilities resulting in rearrangements and/or deletions of *TCRB*, *TCRA* and *TCRD* loci, genes linked to cell cycle growth control (*Cdkn2a* or *Cdkn2b*) and mutations associated with hyperactive Notch signalling[42,43]. The latter are present in at least 50% of cases, often with additional mutations of transcription factors and signalling pathways[44].

In vivo over-expression of transcription factors (TAL1 and TLX1) in mouse also results in T-ALL[45,46], with acceleration of disease mediated by additional mutations such as those involving *Bcl11b* or *Notch1*[47,48]. Thus the activated NOTCH1 found in the MHC$^-$ dKO-derived tumour (Extended Data Fig. 7j) is not unexpected. Notably, disruption of competition between 'new' bone marrow-derived immigrants and 'existing' DN3a thymic resident progenitors in mice leads to aberrant self-renewal of the latter culminating in T-ALL that is reminiscent of human T-ALL in virtually all respects, replete with their development of activating *Notch1* mutations[49]. These DN3a thymic self-renewal progenitors give rise to TCRβ-deficient DP thymocytes with a high frequency of non-productive β gene rearrangements expressing *Notch1* and *Ptcra* transcripts, consistent with ongoing Notch signalling[50].

The expression of myeloid genes by the DP abnormal cluster and the transcriptional signatures shared with AML and ETP-ALL stem cells[36,37] suggest that a subset of myeloid malignancies may arise from the DP compartment after further transformation, particularly in light of the clinical entity of mixed phenotype acute leukaemia expressing both lymphoid and myeloid malignant markers simultaneously[34,51]. Thus, rather than singularly arising from early T cell precursors, a thymic genesis of certain haematopoietic malignancies could involve dedifferentiation from later stages of development, including DP thymocytes

or could even involve transdifferentiation to other lineages. Dedifferentiation is a normal process whereby cells progress in a retrograde manner from a more-differentiated state to a less-differentiated state as a safeguard against progenitor loss[52]. Although such phenomena have been induced by chemical or genetic means in the haematopoietic system[53–55], here we demonstrate that a lack of appropriate signalling during development leads to reprogramming.

Our UMAP projection localizes the abnormal cluster cells between the expected developmental path and the apoptotic cluster. Although detected within a synchronized window of differentiation in vitro, rapid in vivo removal of apoptotic cells by phagocytes in the thymus would obscure the destiny of the unusual and abnormal cluster cells. Their elimination also might be a factor contributing to the discordance between in vitro and in vivo results, in addition to compensatory MHC class Ib expression in vivo that maintains orderly developmental progression. Nevertheless, evolution of a thymic leukaemia despite upregulation of MHCIb in dKO mice underscores the vulnerability that arises during thymic development in the absence of fully normal preTCR triggering. In the small cohort of ageing dKO mice studied here, the frequency of early T cell malignancy was 14%. In mice lacking the Vβ domain of the preTCR—the singular pMHC binding site on the receptor—tumour penetrance was 80% over a similar timeframe[56]. Although additional dKO cohorts will need to be studied to generate robust statistics and tumour genetic features, this differential might be a result of DN3 thymocytes in dKO mice interacting with MHC class Ib ligands via a normal preTCR, in contrast to mice whose preTCRs lack the Vβ domain, and are thus incapable of binding to MHC molecules.

We postulate that self-pMHC reactivity triggers preTCRs on thymocytes during β selection, attendant downregulation of Notch signalling, modulation of cell–cell adhesion, migration and metabolism. Recent studies have demonstrated the formation of an immunological synapse between DN3a thymocytes and stroma, thereby creating a preTCR platform around β-selection to integrate cues involving Notch ligand, CXCR4 ligand and pMHC on thymic stroma, probably involving asymmetric cell division, supporting further differentiation[57,58]. A cellular niche of this type could serve as a pivotal nexus within the developmental circuit to terminate cellular plasticity and foster orderly downstream development. This circuit can malfunction however, if preTCR–pMHC ligation does not occur owing to the absence of functional ligands, a disruption of signalling pathways, or dysregulated entry to and/or exit of progenitors from their developmental niche. Similarly, the generation of intra-thymic AML can be understood as a possible consequence of early developmental plasticity and thymic niche anomalies.

TCR gene rearrangement processes necessary for T lineage repertoire formation bracket the β-selection that supports clonal expansion and repertoire diversification, thereby creating a further vulnerability for tumorigenesis. Somatic TCR repertoire formation, which enables protective adaptive immunity, comes at this potential cost. Our findings emphasize that although thymocyte progression per se can occur in the absence of classical MHC ligand-dependent preTCR function, those self-pMHC interactions are essential for normal development and to mitigate aberrant dedifferentiation. The in vivo upregulation of non-classical MHCIb in the *B2m*/*H2-Ab1* dKO mice preserves some but not all of the features of ligand-dependent preTCR function underscoring this biology, but also revealing the risk of thymic tumour evolution in the absence of classical MHC molecules.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-022-05555-7.

1. Hosokawa, H. & Rothenberg, E. V. How transcription factors drive choice of the T cell fate. *Nat. Rev. Immunol.* **21**, 162–176 (2021).
2. Koch, U. et al. Delta-like 4 is the essential, nonredundant ligand for Notch1 during thymic T cell lineage commitment. *J. Exp. Med.* **205**, 2515–2523 (2008).
3. Rodewald, H. R., Ogawa, M., Haller, C., Waskow, C. & DiSanto, J. P. Pro-thymocyte expansion by c-kit and the common cytokine receptor γ chain is essential for repertoire formation. *Immunity* **6**, 265–272 (1997).
4. Shortman, K., Egerton, M., Spangrude, G. J. & Scollay, R. The generation and fate of thymocytes. *Semin. Immunol.* **2**, 3–12 (1990).
5. Kreslavsky, T. et al. β-Selection-induced proliferation is required for αβ T cell differentiation. *Immunity* **37**, 840–853 (2012).
6. von Boehmer, H. The thymus in immunity and in malignancy. *Cancer Immunol. Res.* **2**, 592–597 (2014).
7. Das, D. K. et al. Pre-T cell receptors (Pre-TCRs) leverage Vβ complementarity determining regions (CDRs) and hydrophobic patch in mechanosensing thymic self-ligands. *J. Biol. Chem.* **291**, 25292–25305 (2016).
8. Li, X. et al. Pre-T cell receptors topologically sample self-ligands during thymocyte β-selection. *Science* **371**, 181–185 (2021).
9. Mallis, R. J., Arthanari, H., Lang, M. J., Reinherz, E. L. & Wagner, G. NMR-directed design of pre-TCRβ and pMHC molecules implies a distinct geometry for pre-TCR relative to αβTCR recognition of pMHC. *J. Biol. Chem.* **293**, 754–766 (2018).
10. Mallis, R. J. et al. Pre-TCR ligand binding impacts thymocyte development before αβTCR expression. *Proc. Natl Acad. Sci. USA* **112**, 8373–8378 (2015).
11. Davis, M. M. & Bjorkman, P. J. T-cell antigen receptor genes and T-cell recognition. *Nature* **334**, 395–402 (1988).
12. Rudolph, M. G., Stanfield, R. L. & Wilson, I. A. How TCRs bind MHCs, peptides, and coreceptors. *Annu. Rev. Immunol.* **24**, 419–466 (2006).
13. Wang, J. H. & Reinherz, E. L. The structural basis of αβ T-lineage immune recognition: TCR docking topologies, mechanotransduction, and co-receptor function. *Immunol. Rev.* **250**, 102–119 (2012).
14. Saint-Ruf, C. et al. Analysis and expression of a cloned pre-T cell receptor gene. *Science* **266**, 1208–1212 (1994).
15. Xiong, J., Armato, M. A. & Yankee, T. M. Immature single-positive $CD8^+$ thymocytes represent the transition from Notch-dependent to Notch-independent T-cell development. *Int. Immunol.* **23**, 55–64 (2011).
16. Petrie, H. T. et al. Multiple rearrangements in T cell receptor α chain genes maximize the production of useful thymocytes. *J. Exp. Med.* **178**, 615–622 (1993).
17. Shinkai, Y. et al. Restoration of T cell development in RAG-2-deficient mice by functional TCR transgenes. *Science* **259**, 822–825 (1993).
18. Wilson, A., Held, W. & MacDonald, H. R. Two waves of recombinase gene expression in developing thymocytes. *J. Exp. Med.* **179**, 1355–1360 (1994).
19. Klein, L., Kyewski, B., Allen, P. M. & Hogquist, K. A. Positive and negative selection of the T cell repertoire: what thymocytes see (and don't see). *Nat. Rev. Immunol.* **14**, 377–391 (2014).
20. Fehling, H. J., Krotkova, A., Saint-Ruf, C. & von Boehmer, H. Crucial role of the pre-T-cell receptor α gene in development of αβ but not γδ T cells. *Nature* **375**, 795–798 (1995).
21. Grusby, M. J. et al. Mice lacking major histocompatibility complex class I and class II molecules. *Proc. Natl Acad. Sci. USA* **90**, 3913–3917 (1993).
22. Irving, B. A., Alt, F. W. & Killeen, N. Thymocyte development in the absence of pre-T cell receptor extracellular immunoglobulin domains. *Science* **280**, 905–908 (1998).
23. Koller, B. H., Marrack, P., Kappler, J. W. & Smithies, O. Normal development of mice deficient in β2M, MHC class I proteins, and $CD8^+$ T cells. *Science* **248**, 1227–1230 (1990).
24. Mizsei, R. et al. A general chemical crosslinking strategy for structural analyses of weakly interacting proteins applied to preTCR–pMHC complexes. *J. Biol. Chem.* **296**, 100255 (2021).
25. Xiao, S. Y., Li, Y. & Chen, W. F. Kinetics of thymocyte developmental process in fetal and neonatal mice. *Cell Res.* **13**, 265–273 (2003).
26. Mingueneau, M. et al. The transcriptional landscape of αβ T cell differentiation. *Nat. Immunol.* **14**, 619–632 (2013).
27. Allman, D. et al. Separation of Notch1 promoted lineage commitment and expansion/transformation in developing T cells. *J. Exp. Med.* **194**, 99–106 (2001).
28. Forman, J. & Fischer Lindahl, K. Listing, location, binding motifs, and expression of nonclassical class i and related genes and molecules. *Curr. Protoc. Immunol.* **49**, A.1M.1–A.1M.13 (2002).
29. Fujita, T., Yuno, M., Okuzaki, D., Ohki, R. & Fujii, H. Identification of non-coding RNAs associated with telomeres using a combination of enChIP and RNA sequencing. *PLoS ONE* **10**, e0123387 (2015).
30. Lin, Y. W. & Aplan, P. D. Gene expression profiling of precursor T-cell lymphoblastic leukemia/lymphoma identifies oncogenic pathways that are potential therapeutic targets. *Leukemia* **21**, 1276–1284 (2007).
31. Li, R. & Guan, M. X. Human mitochondrial leucyl-tRNA synthetase corrects mitochondrial dysfunctions due to the $tRNA^{Leu(UUR)}$ A3243G mutation, associated with mitochondrial encephalomyopathy, lactic acidosis, and stroke-like symptoms and diabetes. *Mol. Cell. Biol.* **30**, 2147–2154 (2010).
32. Coustan-Smith, E. et al. Early T-cell precursor leukaemia: a subtype of very high-risk acute lymphoblastic leukaemia. *Lancet Oncol.* **10**, 147–156 (2009).
33. Vadillo, E., Dorantes-Acosta, E., Pelayo, R. & Schnoor, M. T cell acute lymphoblastic leukemia (T-ALL): New insights into the cellular origins and infiltration mechanisms common and unique among hematologic malignancies. *Blood Rev.* **32**, 36–51 (2018).
34. Dai, Y.-T. et al. Transcriptome-wide subtyping of pediatric and adult T cell acute lymphoblastic leukemia in an international study of 707 cases. *Proc. Natl Acad. Sci. USA* **119**, e2120787119 (2022).
35. Pellicci, D. G., Koay, H. F. & Berzins, S. P. Thymic development of unconventional T cells: how NKT cells, MAIT cells and γδ T cells emerge. *Nat. Rev. Immunol.* **20**, 756–770 (2020).
36. Thoms, J. A. I. et al. Disruption of a GATA2, TAL1, ERG regulatory circuit promotes erythroid transition in healthy and leukemic stem cells. *Blood* **138**, 1441–1455 (2021).

37. Ng, S. W. et al. A 17-gene stemness score for rapid determination of risk in acute leukaemia. *Nature* **540**, 433–437 (2016).
38. Mandal, M. et al. The BCL2A1 gene as a pre-T cell receptor-induced regulator of thymocyte survival. *J. Exp. Med.* **201**, 603–614 (2005).
39. Koyasu, S. et al. Pre-TCR signaling components trigger transcriptional activation of a rearranged TCR α gene locus and silencing of the pre-TCR α locus: implications for intrathymic differentiation. *Int. Immunol.* **9**, 1475–1480 (1997).
40. Amson, R. et al. The human protooncogene product p33pim is expressed during fetal hematopoiesis and in diverse leukemias. *Proc. Natl Acad. Sci. USA* **86**, 8857–8861 (1989).
41. Reinherz, E. L., Kung, P. C., Goldstein, G., Levey, R. H. & Schlossman, S. F. Discrete stages of human intrathymic differentiation: analysis of normal thymocytes and leukemic lymphoblasts of T-cell lineage. *Proc. Natl Acad. Sci. USA* **77**, 1588–1592 (1980).
42. Van Vlierberghe, P. & Ferrando, A. The molecular basis of T cell acute lymphoblastic leukemia. *J. Clin. Invest.* **122**, 3398–3406 (2012).
43. Girardi, T., Vicente, C., Cools, J. & De Keersmaecker, K. The genetics and molecular biology of T-ALL. *Blood* **129**, 1113–1123 (2017).
44. Zhang, J. et al. The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. *Nature* **481**, 157–163 (2012).
45. Condorelli, G. L. et al. T-cell-directed TAL-1 expression induces T-cell malignancies in transgenic mice. *Cancer Res.* **56**, 5113–5119 (1996).
46. Kelliher, M. A., Seldin, D. C. & Leder, P. Tal-1 induces T cell acute lymphoblastic leukemia accelerated by casein kinase IIα. *EMBO J.* **15**, 5160–5166 (1996).
47. De Keersmaecker, K. et al. The TLX1 oncogene drives aneuploidy in T cell transformation. *Nat. Med.* **16**, 1321–1327 (2010).
48. Rakowski, L. A., Lehotzky, E. A. & Chiang, M. Y. Transient responses to NOTCH and TLX1/HOX11 inhibition in T-cell acute lymphoblastic leukemia/lymphoma. *PLoS ONE* **6**, e16761 (2011).
49. Martins, V. C. et al. Cell competition is a tumour suppressor mechanism in the thymus. *Nature* **509**, 465–470 (2014).
50. Paiva, R. A. et al. Self-renewal of double-negative 3 early thymocytes enables thymus autonomy but compromises the β-selection checkpoint. *Cell Rep.* **35**, 108967 (2021).
51. Khan, M., Siddiqi, R. & Naqvi, K. An update on classification, genetics, and clinical approach to mixed phenotype acute leukemia (MPAL). *Ann. Hematol.* **97**, 945–953 (2018).
52. Kai, T. & Spradling, A. Differentiating germ cells can revert into functional stem cells in Drosophila melanogaster ovaries. *Nature* **428**, 564–569 (2004).
53. Cobaleda, C., Jochum, W. & Busslinger, M. Conversion of mature B cells into T cells by dedifferentiation to uncommitted progenitors. *Nature* **449**, 473–477 (2007).
54. Laiosa, C. V., Stadtfeld, M., Xie, H., de Andres-Aguayo, L. & Graf, T. Reprogramming of committed T cell progenitors to macrophages and dendritic cells by C/EBPα and PU.1 transcription factors. *Immunity* **25**, 731–744 (2006).
55. Riddell, J. et al. Reprogramming committed murine blood cells to induced hematopoietic stem cells with defined factors. *Cell* **157**, 549–564 (2014).
56. Jacobs, H. et al. Oncogenic potential of a pre-T cell receptor lacking the TCRβ variable domain. *Oncogene* **12**, 2089–2099 (1996).
57. Charnley, M., Ludford-Menting, M., Pham, K. & Russell, S. M. A new role for Notch in the control of polarity and asymmetric cell division of developing T cells. *J. Cell Sci.* **133**, jcs235358 (2019).
58. Mohtashami, M. et al. Direct comparison of Dll1- and Dll4-mediated Notch activation levels shows differential lymphomyeloid lineage commitment outcomes. *J. Immunol.* **185**, 867–876 (2010).

Methods

Mice

Six-week-old C57Bl/6 (B6) and B6.129-*H2-Ab1*^{tm1Gru} *B2m*^{tm1Jae} N17 (MHC$^-$) mice[21] were purchased from Taconic Farms and housed at the Dana-Farber Cancer Institute Animal Facility, accredited by the Association for Assessment and Accreditation of Laboratory Animal Care (AAALAC). All maintenance, breeding and experimental procedures were approved under Dana-Farber Cancer Institute Institutional Animal Care and Use Committee (IACUC) protocols 03-138 and 04-113. Euthanasia was by CO_2 inhalation followed by cervical dislocation. Following removal from the uterus, embryonic day (E)14.5 fetuses were euthanized by decapitation with surgical scissors. Where appropriate, no gender preference was expressed for experimental animal use.

Reagents

The OP9-DL4 parental (MHC$^+$) cell line, and the MHC$^-$ and scH-2K^b variants, were developed and used as described previously (all tested mycoplasma negative)[7,58]. Anti-mouse CD44-APC/Cy7 (clone IM7) and anti-mouse CD117-APC (c-Kit; clone 2B8) were obtained from BD Biosciences. Anti-mouse CD24 and anti-mouse CD24-FITC (clone M1/69), anti-mouse CD3e-BV605 (clone 145-2C11), anti-mouse CD4-Pacific Blue and CD4-BV711 (clone RM4-5), anti-mouse CD8a-PerCP/Cy5.5 (clone 53-6.7), anti-mouse CD8b.2-PE (clone 53-5.8), anti-mouse CD8b-PerCP/Cy5.5 (clone YTS156.7.7), anti-mouse CD11b-biotin (clone M1/70), anti-mouse CD11c-biotin (clone N418), anti-mouse CD19-biotin (clone 6D5), anti-mouse CD28-PE (clone E18), anti-mouse CD45-BV605 and anti-mouse CD45-APC (clone 30-F11), anti-mouse CD45R/B220-BV421 (clone RA3-6B2), anti-mouse NK1.1-biotin (clone PK136), anti-mouse Gr-1-biotin (clone RB6-8C5), anti-mouse Ter119-biotin (clone TER-119), anti-mouse TCRγδ-biotin (clone GL3), streptavidin-BV421, and Zombie Aqua were obtained from Biolegend. Anti-mouse Ly-6A/E (Sca1)-FITC (clone D7) and anti-mouse CD25-PE/Cy7 (clone PC61.5) were obtained from eBioscience.

The following antibodies from Cell Signaling Technology were used for immunohistochemical detection: anti-mouse CD8 (clone D4W2Z), anti-mouse neutrophil elastase (clone E8U3X) and anti-mouse NOTCH1 intracellular domain (clone D3B8). Anti-mouse TdT (clone EPR2976Y) was obtained from Abcam.

Analysis of B6 thymocyte-like development in vitro

Isolation of wild-type HSC followed the procedure described previously[7]. In brief, fetal liver cells from 30 E14.5 B6 embryos from 3 dams were depleted of B cells using anti-CD24 and complement lysis (Cedarlane) followed by staining with anti-CD4-Pacific Blue, anti-CD8-PE, anti-ScaI-FITC and anti-CD117(c-Kit)-APC. CD4$^-$CD8$^-$ (lineage negative (lin$^-$)) ScaI$^+$ c-Kit$^+$ cells (HSC) were isolated by a Becton-Dickinson FACS Aria II cell sorter. For the T cell repertoire analysis from pooled FACS-sorted thymocyte-like cells, 2,000 HSC were seeded onto 70–90% confluent layers of wild-type OP9-DL4 cells (MHC$^+$), MHC-negative OP9-DL4 cells (MHC$^-$) or scH-2K^b cells in six-well plates (that is, six independent cultures) in α-MEM without nucleosides + 15% fetal calf serum (OP9 media), Hepes (10 mM), and gentamycin supplemented with Flt3 (5 ng ml^{-1}; R&D) and IL-7 (1 ng ml^{-1}; Peprotech). For the scRNA-seq experiments, 30,000 similarly prepared HSC were seeded onto MHC$^+$ or MHC$^-$ stromal cells under the same conditions, increasing the replicates to 10× 10 cm dishes per OP9 variant. After growth for nine days, cells were isolated from the cultures and counted prior to FACS separation for enrichment by surface antigen phenotype for cells at different stages of thymocyte-like differentiation.

Cell sorting, library preparation and data processing for scRNA-seq and TCR V(D)J repertoire characterization

For scRNA-seq analysis, cells were stained with a cocktail consisting of Zombie Aqua for gating of non-viable cells, anti-CD45-APC for gating of haematopoietic cells, with biotinylated anti-CD11b, anti-CD11c, anti-NK1.1, anti-mouse TCRγδ, anti-Gr-1, anti-Ter119, and anti-CD19 followed by streptavidin-BV421 for gating of non-T lineage cells, and of anti-CD4-BV711, anti-CD8α-PerCP/Cy5.5, anti-CD44-APC/Cy7, anti-CD25-PE/Cy7 and anti-CD28-PE for gating and collection of DN3a, DN3b, DN4 and DP thymocyte-like cells on a FACS Aria II cell sorter (Extended Data Fig. 1). Note that residual ILC–γδ T cells in the DN4 subset represent cells with a ILC precursor (*Id2*, *Zbtb16*), ILC2 (*Gata3*, *Rora*), γδ T cell-like transcriptome but with no or low surface TCR expression. For each condition (MHC$^+$ or MHC$^-$), 50,000 DN3a, DN3b, DN4 and DP cells were collected by FACS for application to a 10X Chromium controller (10X Genomics) and recovery of 8,932 ± 920 (mean ± s.e.m.; $n = 8$) processed cells for gene expression (5′ GEX) and TCR V(D)J sequence library construction. Recovery for each MHC$^+$ library was as follows—DN3a: 6,970 cells, DN3b: 7,711 cells, DN4: 7,337 cells, and DP: 5,747. Similarly, recovery for each MHC$^-$ library was as follows—DN3a: 9,453 cells, DN3b: 8,776 cells, DN4: 12,454 cells, and DP: 13,011. Barcoding and 5′ library construction using v1.0 chemistry was performed following the manufacturer's protocol. Targeted mouse TCR recovery utilized the Chromium Single Cell V(D)J Enrichment Kit for mouse T cells. All libraries were single i7-indexed using the Chromium i7 Multiplex kit. Following isolation and clean-up of library DNA, integrity was assessed using an Agilent Bioanalyzer and quantification by Qubit analysis (Invitrogen). All libraries were adjusted to ~50 ng μl^{-1}, where peak fragment size (including Illumina adapters) for the gene expression (5′ GEX) libraries averaged 473 bp and ranged from 300–740 bp for the 5′ TCR libraries representing ongoing recombination products in the developing thymocyte libraries. Sequencing (150 PE) was performed on HiSeq 3000 utilizing 4 lanes where two 5′ GEX libraries (2 × 40% of reads) and two TCR libraries (2 × 10% of reads) were sequenced per lane.

Following conversion of the bcl sequencing files to fastq format, the 5′ GEX sequencing results were pipelined to Cellranger 3.1.0 using the GRCm38.p6/mm10 mouse genome as reference and the TCR files were pipelined to Cellranger V(D)J 3.1.0 using vdj_GRCm38_alts_ensembl-3.1.0.gz-3.1.0 as reference, all using default parameters (https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/algorithms/overview#secondary-analysis) as described in the following overview. Gene expression data from all libraries were aggregated by Cellranger to generate a UMAP of all libraries projected into the same 2D space. For aggregation, the count output files for each Chromium controller well were processed using the aggr command to produce a single feature-barcode matrix containing all the data. Since barcodes may overlap between libraries, a well suffix is added to each barcode-nucleotide sequence to hardcode well origin. Before merging, depth normalization is performed to subsample reads for each library to equalize the number of reads confidently mapped to the transcriptome. Prior to principal component analysis, the UMI counts were normalized towards the median across all cells by multiplying each cell's UMI count by a scaling factor of the median UMI count across all the cells divided by the UMI count for the cell. The matrix is log-transformed then centred and scaled per gene such that the mean is 0 and the s.d. is 1 prior to clustering. Consequently, all data used for differential expression is log-normalized and a pseudocount of 1 was added to both the numerator and denominator of the mean expression. For differential expression analysis and associated *P* values, Cellranger utilizes an implementation of the exact negative binomial test (https://bioconductor.org/packages/release/bioc/html/sSeq.html). For a cluster or selected cell subset within a cluster, $\log_2$-fold change was either tested against the mean expression for all other cells (global analysis) or against a selected cluster or subset (local analysis) using the Loupe browser 4.2.0 together with the Loupe V(D)J browser 3.0.0 (10X Genomics) for integration of TCR clonotype parameters. The log-normalized data for specific clusters were piped into ssGSEA (https://www.genepattern.org/modules/docs/ssGSEAProjection/4#)

for analysis of co-ordinated gene regulation in gene set modules associated with up- or downregulation in defined cancers (MSigDb C4; http://www.gsea-msigdb.org/gsea/msigdb/collections.jsp). ssGSEA scores greater than 1,000 are considered as strongly associated, equal to 0 as showing no correlation with module genes, and less than 0 as inversely correlated.

Bulk population TCR repertoire protocol and data processing for cells developing in vitro

For the bulk population β repertoire analyses of thymocyte-like cells developing on the MHC^+ and MHC^- stromata, respectively, cells were stained with anti-CD45-APC, anti-CD4-Pacific Blue, anti-CD8-PE, anti-CD25-PE/Cy7 and anti-CD44-APC/Cy7 for simultaneous collection of DN3, DN4, DPbl and DPsm thymocytes on a FACS Aria II cell sorter (Supplementary File 4). Contaminating OP9 cells expressed GFP permitting their exclusion while selection for CD45 expression ensured only haematopoietic cells were used for subset delineation. Cells were gated as $CD4^-CD8^-$ (DN) and $CD4^+CD8^+$ (DP) from which 10,000 cells each of DN3 ($CD25^+CD44^-$) cells, DN4 ($CD25^-CD44^-$) cells, DPbl (blast cells, in cell cycle; $CD4^+CD8^+$ high forward scatter) and DPsm (small, more mature cells; $CD4^+CD8^+$ low forward scatter), were collected. For each population, the cells were immediately deposited into 2× TCL lysis buffer (Qiagen) supplemented with 2-mercaptoethanol (1%) on ice, snap-frozen by immersion in dry-ice-methanol and stored at −80 °C until processed for RNA extraction and β chain repertoire analysis.

Total RNA was extracted from each sample of 10^4 cells using the PicoPure column purification system (Applied Biosystems). Subsequently, the procedure was precisely as described[59]. In brief, using a 3′ TCRβ constant region (TRBC) universal primer, 1st strand cDNA was synthesized from the starting RNA and a universal 'switch' primer ligated to the 5′ ends. Nested/extended PCR amplification through the universal ends yielded unbiased amplification of transcripts containing the complete V(D)J region and a 5′ segment of the TRBC. In the second PCR, pentanucleotide barcodes were introduced to tag each library with unique barcodes at both 5′ and 3′ ends. Following quality control using the Agilent 2100 Bioanalyzer and Illumina adapter addition, samples were sequenced (150 PE) on the MiSeq platform. Library sequences were deconvoluted from the fastx sequence output files using the barcode splitter module of the FASTX toolkit (http://hannonlab.cshl.edu/fastx_toolkit/index.html). The deconvoluted library sequences were aligned to Vβ regions in the GRCm38.p6/mm10 mouse genome followed by clone assembly and CDR3 extraction using the MiXCR suite running under Java[60]. Output provided V, D, J and Cβ usage, CDR3 nucleotide and amino acid sequence, sequence quality, and relative representation by read count. The VDJtools analytical package was used to track and compare clonotypes within the libraries[61].

Comparative transcriptomes of MHC^+ and MHC^- stromal cells

Total RNA was extracted from OP9-DL4 $MHCI^+$ parental cells and from OP9-DL4 MHC^- cells and processed commercially for standard RNA-seq and data processing (Novogene). In brief, following alignment to the GRCm38.p6/mm10 mouse genome using STAR 2.7.3a (https://github.com/alexdobin/STAR/releases), gene expression was quantified as fragments per kilobase of transcript sequence per million base pairs sequenced (FPKM) followed by differential gene expression determined by DESeq2 (https://bioconductor.org/packages/release/bioc/html/DESeq2.html) yielding expression level, $\log_2$ fold difference between the two OP9-DL4 variants, a P value and an adjusted P value (P_{adj}) using the Benjamini–Hochberg correction to control the false discovery rate.

Detection of myeloid transcripts in DP cells developing in vitro

Isolation of wild-type HSC followed the procedure described previously[7]. Thirty thousand $CD4^-CD8^-lin^-ScaI^+c\text{-}Kit^+$ cells (HSC) were isolated by FACS and placed onto 3×10^5 MHC^+ OP9-DL4 cells seeded 1 day previously in OP9 media. After growth for 7 days, cells were isolated from the culture, counted, and then stained with a cocktail of biotinylated anti-CD11b, anti-CD11c, anti-NK1.1, anti-Gr-1, anti-Ter119, anti-CD19 and anti-TCRγδ. Cells were subsequently stained with streptavidin-BV421, anti-CD45-BV605, anti-CD4-BV711, anti-CD8β-PerCP/Cy5.5, anti-CD25-PE/Cy7 anti-CD44-APC/Cy7, anti-CD28-PE and Zombie Aqua. Viable DN3a and DN4 thymocytes were collected by gating initially for Zombie $Aqua^-GFP^-CD45^+$ cells, with subsequent gating for the $CD4^-CD8^-$ DN cells and sorting of 1,000 DN3a ($CD25^+CD44^-CD28^-$) cells, and 1,000 DN4 ($CD25^-CD44^-CD28^+$). Those cells were placed on 50,000 MHC^+ or MHC^- OP9-DL4 cells seeded 1 day previously. After culture for 12 days, cells were isolated from the culture, counted, and stained with anti-B220-BV421, anti-CD45-BV605, anti-CD4-BV711, anti-CD8β-PerCP/Cy5.5 and Zombie Aqua. $CD4^+CD8^+$ DP cells were sorted into 2× TCL lysis buffer (Qiagen) supplemented with 2-mercaptoethanol (1%) on ice and stored at −80 °C. Total RNA was prepared (RNeasy Micro; Qiagen) and RT–qPCR was performed on an Applied Biosystems 7500 Fast using a SYBR green Superscript/Taq master mix (Power SYBR Green RNA-to-CT 1-Step Kit; ThermoFisher) and the following PrimerBank-validated primers[62]: mouse *Mpo*-F: AGTTGTGCTGAGCTGTATGGA; mouse *Mpo*-R: CGGCTGCTTGAAGTAAAACAGG; mouse *Spi1*-F: ATGTTACAGGCGTGCAAAATGG; mouse *Spi1*-R: TGATCGCTATGGCTTTCTCCA; mouse *Actb*-F: GGCTGTATTCCCCTCCATCG; mouse *Actb*-R: CCAGTTGGTAACAATGCCATGT. Data were processed using the LinRegPCR package[63].

Gene expression and total β clonotype analysis of thymus DN3 to ISP cells

The thymus from each of 3 B6 and 3 MHC^- mice (all males aged 3 weeks) was isolated and the cells were dispersed into RPMI-1640 medium treating each thymus as an individual sample. The cells were incubated with anti-CD4 (clone L3T4) covalently linked to microbeads (Miltenyi Biotec) used at a ratio of 100 µl beads per 10^8 cells then incubated for 10 min on ice. The thymocyte–microbead mixtures were applied to replicate LS MACS columns in a MidiMACS separator and unbound cells collected as CD4-depleted populations removing DP thymocytes and CD4SP thymocytes. The CD4-depleted populations were then sorted to remove the non-T lineage cells as described above (viable, non-T $lin^- < 0.05\%$ $CD4^+$). Following gating on DN cells ($CD4^-CD8^-$), cells were gated further on the DN3/DN4 population ($CD44^-$), and then into three further gates of $CD25^{hi}CD28^{lo/int}$ (DN3a), $CD25^{int}CD28^{hi}$ (DN3b), and $CD25^{lo}CD28^{hi}$ (DN4) as outlined in Supplementary File 4. Following gating on the $CD8^+$ cells in the viable, non-T lin^- population, the cells were further gated on the $CD24^{hi}CD3^-$ cells (ISP) and isolated populations collected into TCL lysis buffer as described above. For each mouse this procedure yielded the complete representation of all phenotypically defined DN3a, DN3b, DN4, and ISP thymocytes. Total RNA for each population was prepared using the RNAqueous-4PCR protocol (Applied Biosystems, Life Technologies). From the isolated total RNA, 200 ng was removed for NGS library preparation (SMART-Seq v4 Ultra Low Input RNA, Takara), Illumina adapter addition, and sequencing (PE150, Novaseq platform, ~40×10^6 reads per sample) for gene expression analysis (Medgenome). Read count data were normalized using DESeq2. The aligned reads were used for estimating expression of the genes using cufflinks v2.2.1. The expression values are reported in FPKM (Fragments per kilobase per million) units for each gene. The remaining RNA was used for total population Trbv repertoire determination following the protocol of Mamedov et al.[59]. with minor differences to the procedure described above. To reduce errors introduced by PCR amplification as well as estimate individual RNA contributing to a particular clonotype, the Switch primer incorporated a region with a UMI motif of 12 nucleotides within which were interspersed several deoxyuridine nucleotides subsequently treated after cDNA synthesis with uracyl deglycosylase to prevent participation of the Switch primer in the downstream PCR reactions. The individual barcoded DN3a, DN3b, DN4 and ISP libraries for each animal were pooled and Illumina adapters added

to generate one total thymus library of these stages for each animal. Following sequencing (PE150, Novaseq platform, Medgenome), library deconvolution, assembly, alignment and UMI processing was handled by the MIGEC package[64] to determine β clonotype repertoire based on UMI rather than total reads. The output was then pipelined directly to the VDJtools package[61] as described above. Repertoire diversity was assessed using CalcDiversityStats module of the VDJtools package based on the D50 and diversity index (DI)[65] that yields a value in the 0–1 range where 1 = maximal diversity.

Histological and immunohistochemical analyses

Tissue samples were fixed in formalin then paraffin-embedded using standard procedures. Sections of 5 µm were prepared, attached to slides, sections including bone were demineralized, and the samples processed for haematoxylin and eosin staining on the automated Histocore Spectra ST platform (Leica).

Immunohistochemistry was performed on the Bond III automated staining platform (Leica) using the Biosystems Refine Detection Kit (Leica). Staining utilized the following antibodies at the indicated dilution: anti-CD8 (1:200) with citrate antigen retrieval; anti-TdT (1:100) with EDTA antigen retrieval; anti-neutrophil elastase (1:200) with EDTA antigen retrieval; and anti-NOTCH1 intracellular domain (NICD1; 1:50) with EDTA antigen retrieval.

Y chromosome fractional analysis of transcriptionally defined clusters

To address the possibility that the well-represented DN4 unusual population (1,776 cells, 14.3% of DN4 library) and DP abnormal population (2,512 cells, 19.3% of DP library) in the MHC^- condition represent the clonal development of a single or limited number of aberrant progenitor cell(s) during the 9-day culture, skewing of the initial HSC male-to-female cell ratio was assessed by examining the XY cell fraction in each cluster. Given that the initial seeding of 30,000 fetal liver progenitors/initial culture plate originated from a common pool, the ratio of male (XY) cells to female (XX) cells should be maintained across all MHC^+ and MHC^- cultures through to the isolation of phenotypically defined subsets and transcriptionally defined subsets (clusters) within. Accordingly, a skewed XY/XX ratio within a cluster may indicate non-uniform clonal expansion.

Chromosome Y transcripts likely to be expressed were determined as described in the Supplementary File 6 (and refs. [66,67]) and the list screened against all libraries to generate a panel of four transcripts (*Ddx3y*, *Eif2s3y*, *Kdm5d* and *Uty*) found to be consistently expressed and detectable across all libraries and clusters. The representation of these transcripts was then used to define presence of a Y chromosome. The reverse procedure utilizing transcripts with increased representation in XX cells was not feasible as none of the identified, skewed, transcripts were detected at levels high enough or specifically enough to characterize a cell as definitively XX. Consequently, results were expressed as fraction of cells within a cluster characterized as XY. To exclude observed skewing being a result of apoptotic or other processes occurring in a specific cluster independent of supporting stroma, a panel of genes matched to expression of the Y transcript panel was determined. This latter panel of autosomal gene transcripts (*Cdk8*, *Slc25a5*, *Pank1* and *Dffb*) provided an internal control for cluster-specific skewing within the stroma-specific libraries unrelated to aberrant clonal development (Supplementary File 6).

Statistics

For all gene expression results, *P* represents the adjusted *P* value (P_{adj}) where $P < 0.05$ was the threshold for significance. Standard parametric statistics followed by Student's *t*-test and two-tailed probabilities were used for all group comparisons. For comparison of small lists of transcripts representing gene expression levels (for example, Trbv alleles), paired *t*-test or the Chi-square test (utilizing MHC^+ as 'expected') with two-tailed probability were used. No sample size calculations were undertaken, no randomization of samples was performed and there was no blinding of samples.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All sequence files have been deposited at NCBI Gene Expression Omnibus (GEO) under accession GSE186049. Source data are provided with this paper.

59. Mamedov, I. Z. et al. Preparing unbiased T-cell receptor and antibody cDNA libraries for the deep next generation sequencing profiling. *Front. Immunol.* **4**, 456 (2013).
60. Bolotin, D. A. et al. MiXCR: software for comprehensive adaptive immunity profiling. *Nat. Methods* **12**, 380–381 (2015).
61. Shugay, M. et al. VDJtools: unifying post-analysis of T cell receptor repertoires. *PLoS Comput. Biol.* **11**, e1004503 (2015).
62. Wang, X., Spandidos, A., Wang, H. & Seed, B. PrimerBank: a PCR primer database for quantitative gene expression analysis, 2012 update. *Nucleic Acids Res.* **40**, D1144–D1149 (2012).
63. Ruijter, J. M. et al. Evaluation of qPCR curve analysis methods for reliable biomarker discovery: bias, resolution, precision, and implications. *Methods* **59**, 32–46 (2013).
64. Shugay, M. et al. Towards error-free profiling of immune repertoires. *Nat. Methods* **11**, 653–655 (2014).
65. Han, F. F. et al. Profiling the pattern of human TRB/IGH-CDR3 repertoire in liver transplantation patients via high-throughput sequencing analysis. *Scand. J. Immunol.* **92**, e12912 (2020).
66. Stevant, I. et al. Dissecting cell lineage specification and sex fate determination in gonadal somatic cells using single cell transcriptomics. *Cell Rep.* **26**, 3272–3283.e3 (2019).
67. Godfrey, A. K. et al. Quantitative analysis of Y-chromosome gene expression across 36 human tissues. *Genome Res.* **30**, 860–873 (2020).

Acknowledgements This research was supported by NIH NIAID grant AI136301. C.M.M. and P.H.L. were supported additionally by the Expect Miracles Foundation and the Robert and Renée Belfer Foundation. We thank J.-h. Wang for scientific discussion and insight, D. A. Barbie and C. P. Paweletz for facilitating the scRNA-seq analyses, the Dana-Farber/Harvard Cancer Center Specialized Histopathology Core (NIH NCI grant P30 CA006516-57) and S. Moskovitz for graphic design of figures.

Author contributions Conceptualization: J.S.D.-C., A.A., R.J.M., W.H., M.J.L. and E.L.R. Methodology: J.S.D.-C., A.A., R.J.M. and E.L.R. Investigation: J.S.D.-C, A.A., C.M.M., P.H.L. and J.C.A. Writing, original draft: J.S.D.-C. and E.L.R. Writing, review and editing: J.S.D.-C., R.J.M., A.A., W.H., M.J.L. and E.L.R. Funding acquisition: M.J.L., E.L.R. and J.C.A. Supervision: J.S.D.-C. and E.L.R.

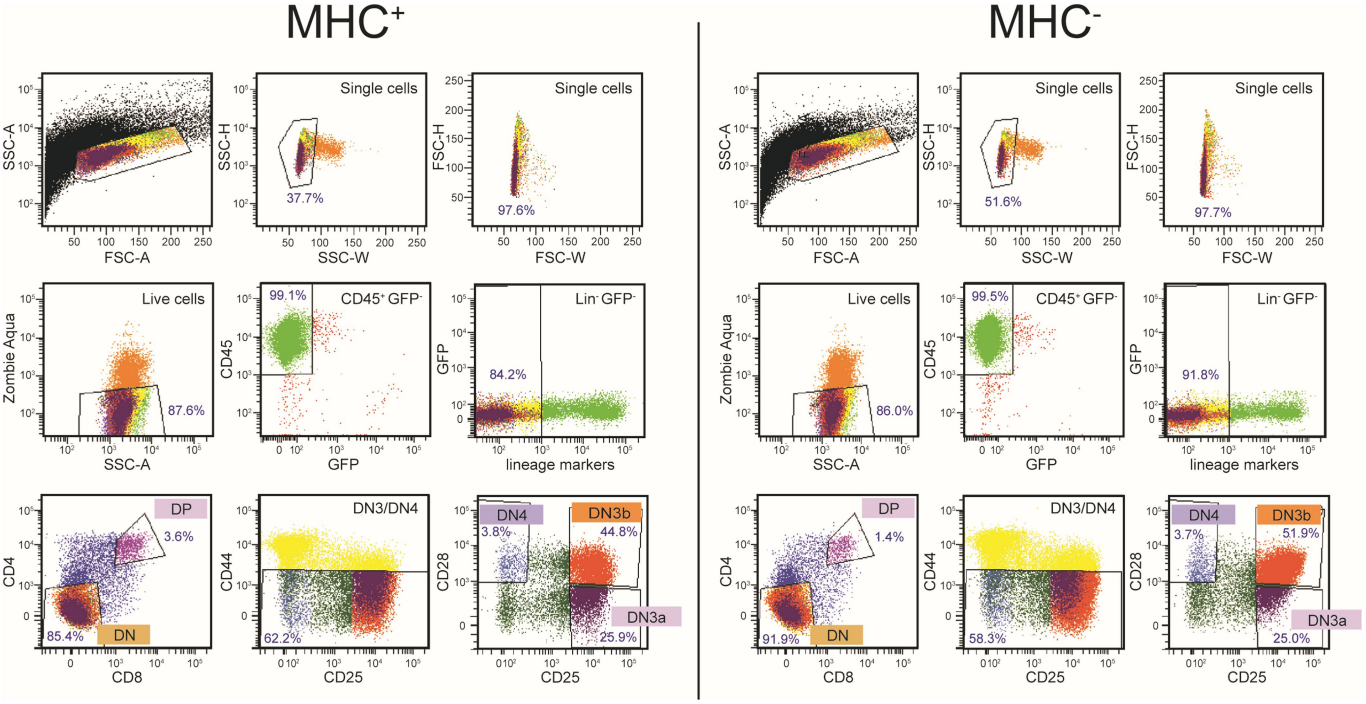
Competing interests The authors declare no competing interests.

Additional information

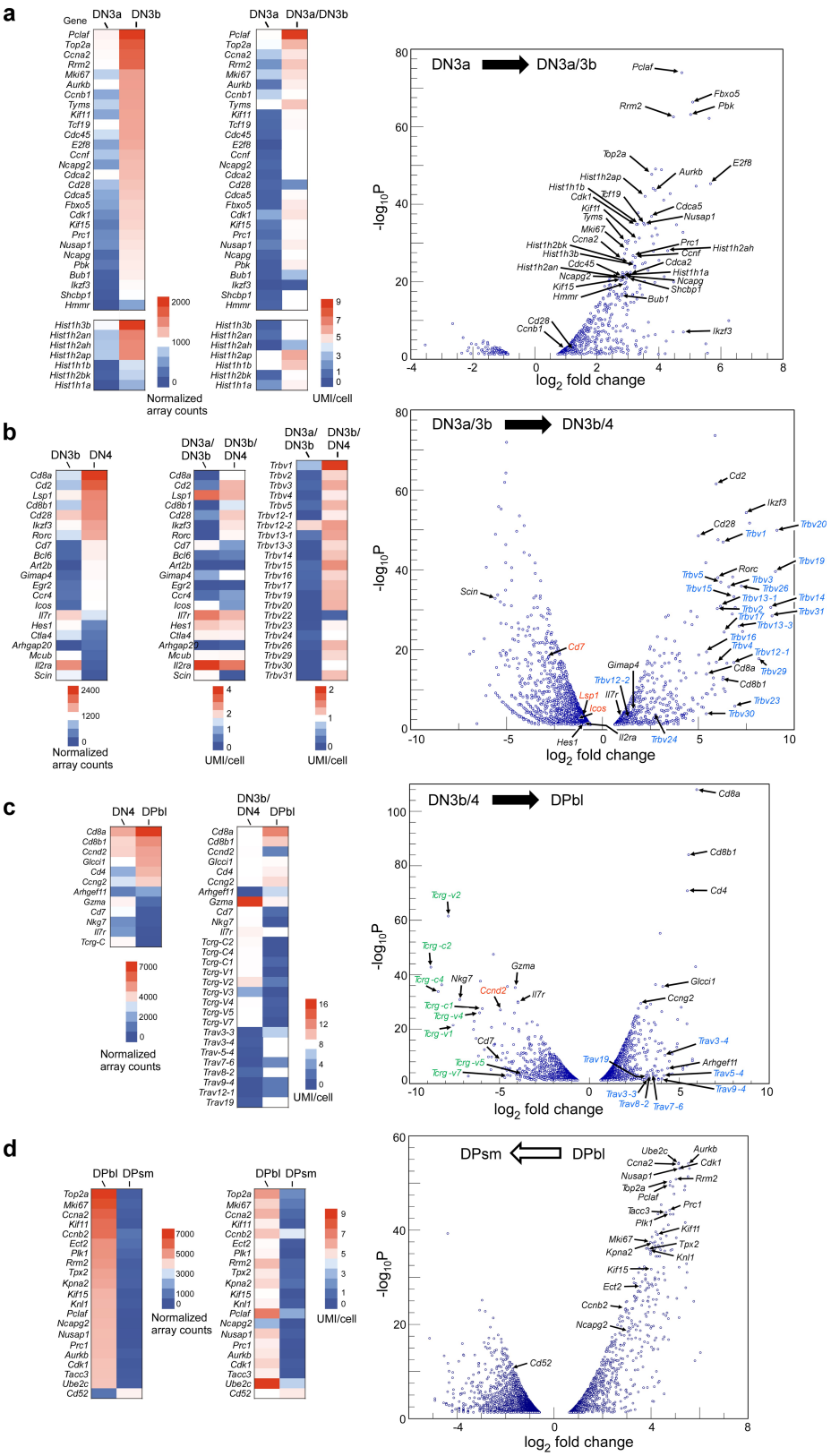
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-022-05555-7.

Correspondence and requests for materials should be addressed to Jonathan S. Duke-Cohan or Ellis L. Reinherz.

Peer review information *Nature* thanks Charles Mullighan, Hans-Reimer Rodewald and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

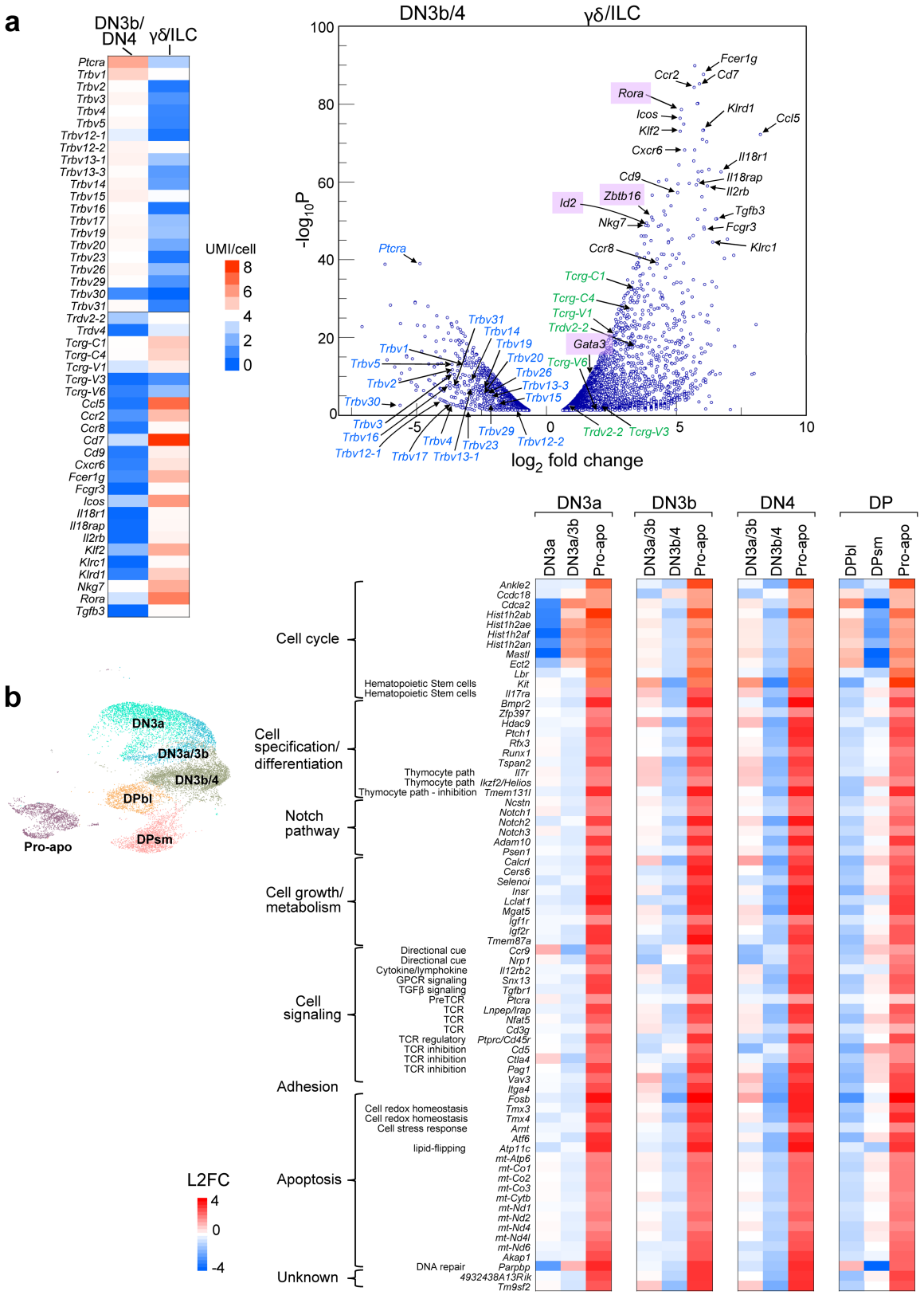


Extended Data Fig. 1 | Schematic for FACS isolation of thymocyte subsets (DN3a, DN3b, DN4, DP) for 10X scRNA-Seq and single cell TCR α and β chain clonotype sequencing. Sorted cells were isolated as DN3a cells ($CD25^+CD44^-CD28^-$), DN3b cells ($CD25^+CD44^-CD28^+$), DN4 ($CD25^-CD44^-CD28^+$) cells, and DP ($CD4^+CD8^+$) cells.

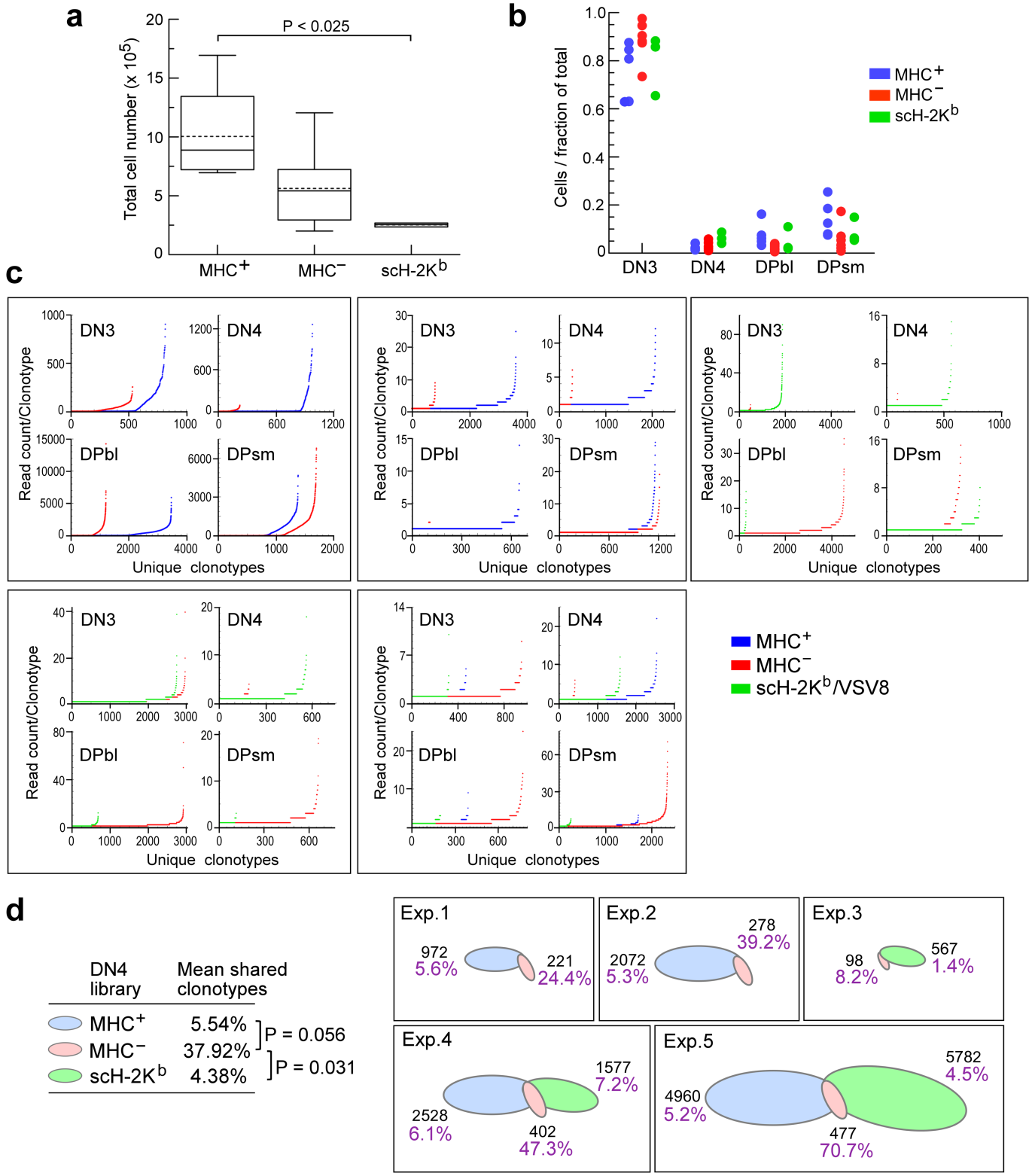


Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Cluster delineation of DN3a to DPsm cell transitions. For each transition, data from the Immune Genome Project (IGP) microarray and RNA-Seq data was used to construct a panel representing genes with the highest fold-change between phenotypically defined stages of thymocyte differentiation. The gene panel was then used to query the MHC^+ thymocyte clusters identified by UMAP projection. Combination of library phenotype together with good fit to the interrogating gene panel permitted identification of cluster relationships and developmental trajectories. **a**. Delineation of early post-β selection checkpoint DN3a/3b thymocytes from pre-β selection checkpoint DN3a thymocytes by differential gene expression. The left-hand heatmap depicts a panel selected by comparison of DN3b thymocyte gene expression from the IGP with DN3a cell expression. The same genes were examined for expression in the clusters defined as DN3a and DN3a/3b in Fig. 1B (right-hand heatmap). The volcano plot depicts the $\log_2$-fold increase of expression in the DN3a/3b population over DN3a for the expected normal developmental trajectory (x-axis). Note that for all volcano plots reported here, only the significantly changed transcripts are depicted ($P_{adj} < 0.05$; y-axis). **b**. Delineation of late post-β selection checkpoint DN3b/4 thymocytes from early pre-β selection checkpoint DN3a/3b thymocytes by differential gene expression. The heatmap on the far left depicts a panel selected by comparison of DN4 thymocyte gene expression from the IGP with DN3b cell expression (neither DN3a/3b nor DN3b/4 transitional states are explicitly defined in the IGP database). Transcripts in red were predicted from IGP data to be upregulated in the DN3b to DN4 transition but are downregulated for the conditions reported here. **c**. Delineation of late post-β selection checkpoint DN3b/4 thymocytes from DPbl thymocytes by differential gene expression. The DPbl cluster was extracted from the DP library and delineated from the more mature DPsm population by transcriptome signature as described below. **d**. Delineation of mature DPsm thymocytes from cycling DPbl thymocytes by differential gene expression. The heatmap on the far left depicts a panel selected by comparison of DPsm thymocyte gene expression from the IGP with DPbl cell. Note that during the DPbl to DPsm transition, significant cell cycling transcripts were downregulated thus significantly upregulated transcripts in the volcano plot represent the DPbl cells.

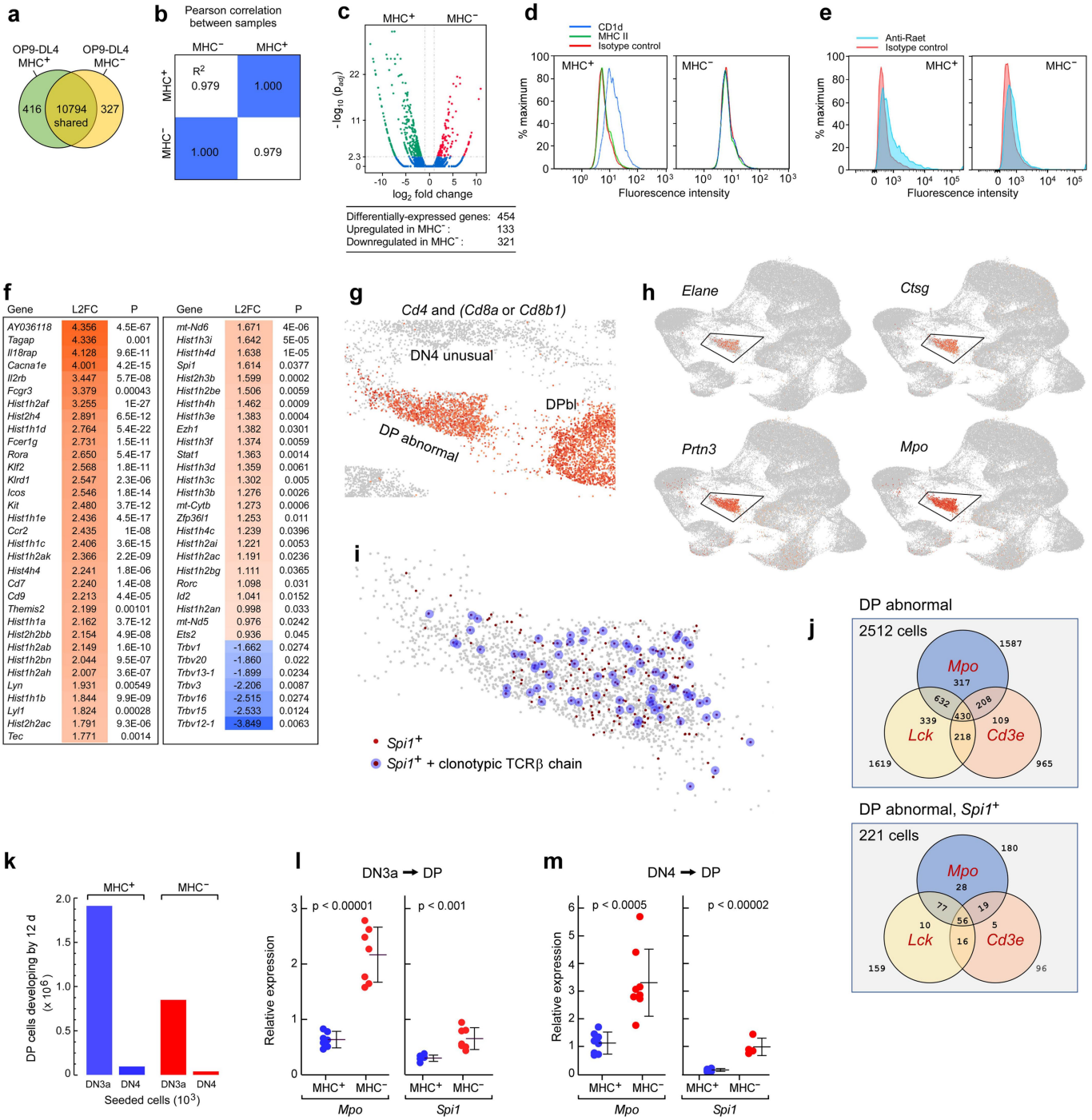


Extended Data Fig. 3 | Delineating the ILC-γ/δ TCR thymocyte cluster and pro-apoptotic cluster from the main α/β TCR lineage pathway. **a**. Distinguishing ILC-γ/δ-like cells from DN3b/4 in the DN4 libraries by gene expression. The heatmap on the left shows a manually curated panel of gene transcripts selected by likely high representation in either DN3b/4 or ILC-γ/δ-like cells. Log2 Fold-change (L2FC) and P_{adj} in the DN4 libraries for differential expression between the DN3b/4 clusters and ILC-γ/δ-like clusters are shown in the volcano plot to the right with transcripts associated with ILC development are highlighted in light purple (*Id2*, *Zbtb16*, *Gata3*, *Rora*). TCR γ and δ transcripts are highlighted in green, and *Trbv* transcripts highlighted in blue. **b**. Gene expression profile of the pro-apoptotic cluster. The dominant pro-apoptotic cluster upregulated gene expression changes are similar between all the MHC^{+} libraries on comparison with the 2 dominant clusters within each of these libraries. All log2-fold changes (L2FC) are relative only to the 3 clusters listed in each heatmap (*i.e.* local) and not to the average across all clusters in that library.

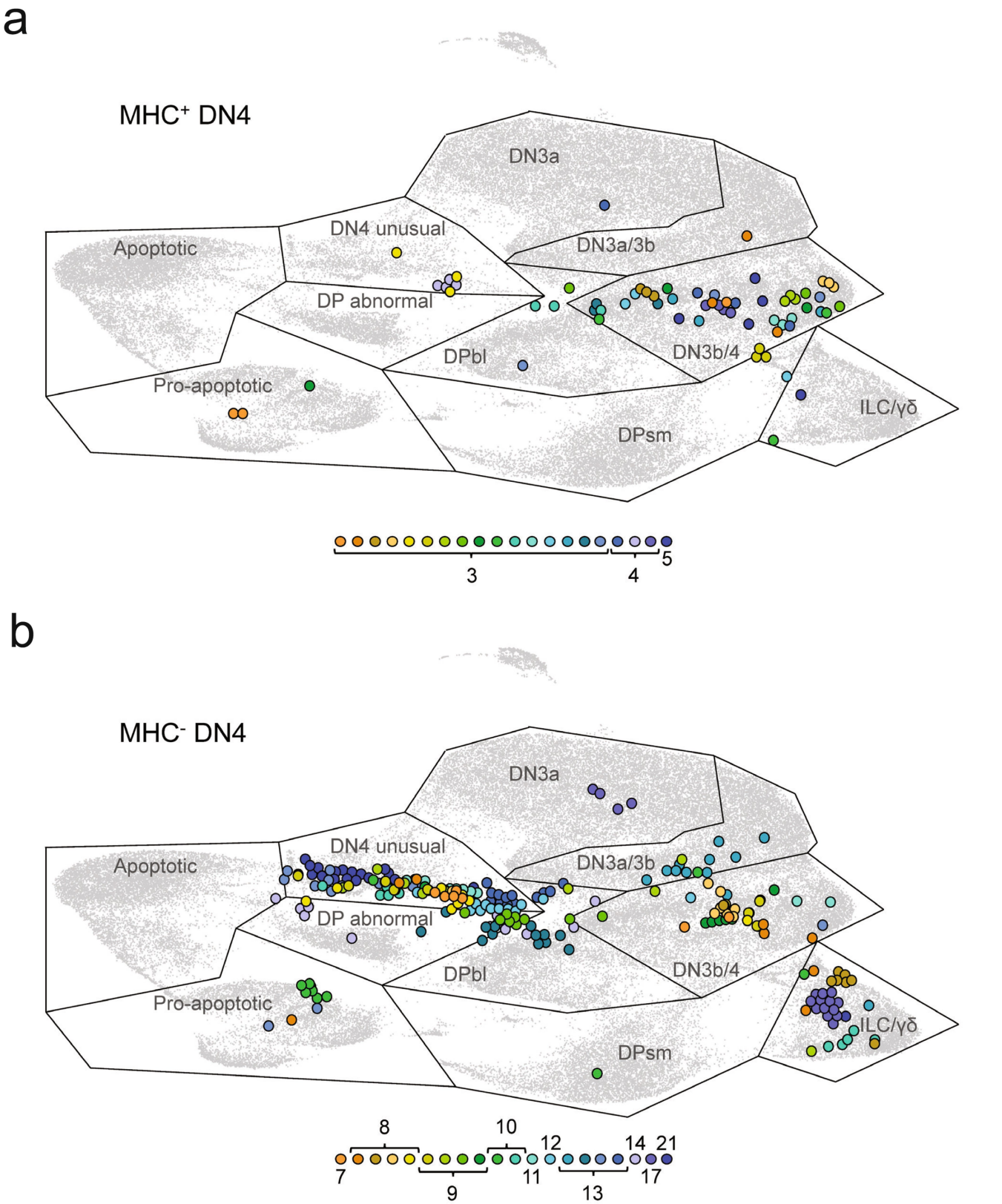


Extended Data Fig. 4 | See next page for caption.

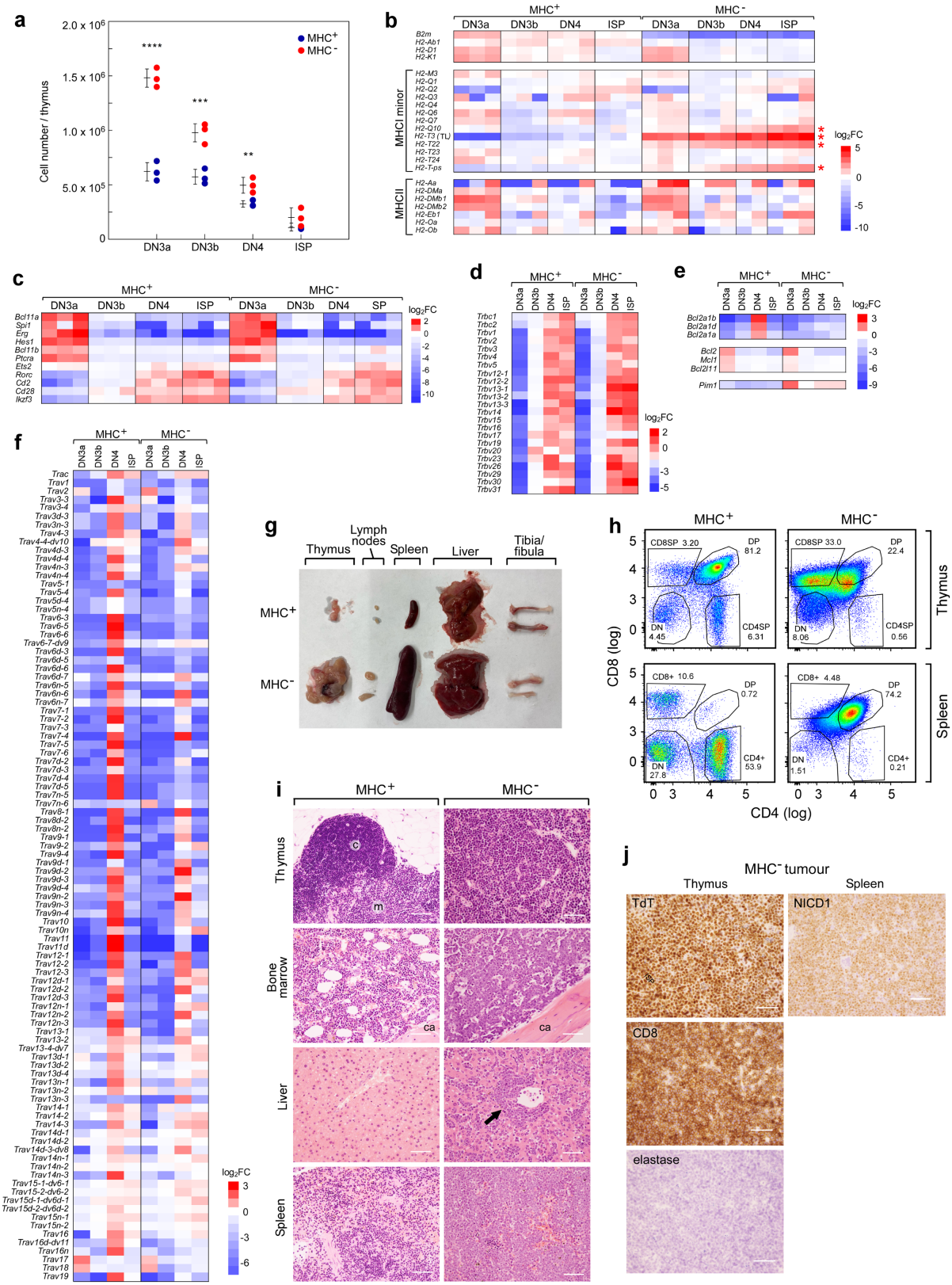
Extended Data Fig. 4 | Development and TCR repertoire analyses for cells growing on MHC$^+$, MHC$^-$ and scH-2K^b stromal support cells. a. Total cell recoveries after 9d development from 2,000 seeded HSC (Representative of 6 experiments examining MHC$^+$ (n = 5), MHC$^-$ (n = 6), and scH-2K^b (n = 3)). For all box plots, the box bounds the 1st to 3rd quartiles; where visible, the dotted line within represents mean, and the solid line represents median. Whiskers above and below (maximum and minimum) are defined as (quartile 3 + 1.5 * interquartile range) and (quartile 1 – 1.5 * interquartile range), respectively. P (= 0.0204) determined by two-tailed t test. **b**. Apparent thymocyte developmental stage representation as fraction of total cells for cultures represented in panel a. **c**. Stage-specific analysis of β chain clonotype representation/10,000 cells in d9 MHC$^+$, MHC$^-$, and scH-2K^b OP9-DL4 development cultures. Representation of data from replicate experiments of data in Fig. 2c–f. **d**. TCR β chain clonotype diversity at DN4 on MHC$^+$, MHC$^-$, and scH-2K^b stroma. The total number of TCR β chain clonotypes (black) recovered from 10^4 cells of each DN4 population isolated after growth for 9d on the varying OP9-DL4 stroma is represented by an ellipse of area in direct proportion to unique clonotype count (5 independent experiments). Percentage shared clonotypes of the total for each condition (MHC$^+$ in blue, MHC$^-$ in pink, and scH-2K^b in green) is depicted. Note that the area of overlap only approximates degree of sharing to maintain consistent orientation of the ellipses for presentation. The overlap of MHC$^+$ and scH-2K^b for experiments 4 and 5 is <1% and too small to represent in this format. Statistics and P calculated from two-tailed t test presented on left.



Extended Data Fig. 5 | Transcriptome and selected phenotype comparison of MHC$^+$ and MHC$^-$ OP9-DL4 cells and select gene expression profiles for the DN4 unusual and DPbl abnormal populations. a. Comparison of MHC$^+$ and MHC$^-$ OP9-DL4 stromal cells for transcriptome and phenotypic differences. 93.6% of transcripts detected shared by MHC$^+$ and MHC$^-$ stroma. **b**. Correlation between cell transcriptomes. Square of two-tailed Pearson correlation coefficient ($R^2 = 0.958$) ideally greater than 0.92 under optimal experimental conditions. **c**. Differential gene expression is <4% of all transcripts detected. **d**. Loss of CD1d surface expression in *B2m/Tap2* KO MHC$^-$ OP9-DL4 and confirmation of lack of MHC Class II expression in MHC$^+$ and MHC$^-$ OP9-DL4. **e**. *Raet* expression in MHC$^+$ and MHC$^-$ OP9-DL4. **f**. Select transcripts significantly differentially expressed between the MHC$^-$ DN3b/4 cluster and the DN4 "unusual" cluster. Heatmap depicts $\log_2$-fold change (L2FC) of the DN4 "unusual" cluster relative to the DN3b/4 cluster. Actual L2FC values are listed within the heatmap. **g**. Co-expression of *Cd4* transcript with *Cd8a* and/or *Cd8b1* transcripts in an overlay of the MHC$^-$ libraries focussed on the DN4 unusual, DPbl, and DP abnormal clusters. **h**. Characteristic myeloid gene transcript expression maps to the MHC$^-$ DP abnormal cluster. **i**. Full-length clonotypic TCR β chain transcript expression in 82 of 221 *Spi1*$^+$ cells (37.1%) in the MHC$^-$ DP abnormal cluster. **j**. *Mpo*-expressing cells in the DP abnormal cluster and the *Mpo*$^+$*Spi1*$^+$ subset co-express T lineage *Lck* and/or *Cd3e*. **k**. DP cell yields after 12 d for DN3a and DN4 cells seeded onto MHC+ or MHC- stromal cells. **l**. Relative expression by qRT-PCR (normalised to *Actb* = 1000) of *Mpo* and *Spi1* in DP cells developing from DN3a cells seeded 12 d earlier onto MHC$^+$ or MHC$^-$ stromal cells (Cells pooled from 3 separate cultures; n = 7 qRT-PCR replicates; *Mpo*: P < 0.00001, *Spi1*: P = 0.000655). **m**. Relative expression (normalised to *Actb* = 1000) of *Mpo* and *Spi1* in DP cells developing from DN4 cells seeded 12 d earlier onto MHC$^+$ or MHC$^-$ stromal cells (Cells pooled from 3 separate cultures; *Mpo*: n = 8 qRT-PCR replicates; *Spi1*: n = 4 qRT-PCR replicates); for **l**, **m**: mean ± s.d.; P from two-tailed t-test; representative of 2 independent experiments; *Mpo*: P = 0.000013, *Spi1*: P = 0.000233).



Extended Data Fig. 6 | Highly proliferating clonotypic progeny cluster together by transcriptional signature. a. MHC^{+} DN4 20 most highly represented clonotypes by cell number. **b**. MHC^{-} DN4 20 most highly represented clonotypes by cell number. The identical MHC^{+} and MHC^{-} DN4 clonotypic cells to those presented in Fig. 3d and Extended Data Table 3a are shown in their mapped positions in the UMAP projection. Each clonotype is represented for each panel in a unique colour with cell number indicated in key. Note that colours are not directly related to those used in Fig. 3d.



Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Transcriptome comparison of DN and ISP thymocyte subsets from MHC$^+$ and MHC$^-$ mice. a. Thymocyte subset cell recoveries from thymi of MHC$^+$ and MHC$^-$ mice. Mean ± s.d. shown; 3 mice/group; ** p = 0.0169; *** p < 0.0039; **** p < 0.0003 determined by 2-tailed t-test. **b**. Log_2-fold change in expression from global population mean for the MHC$^-$ knocked out genes (*B2m*, *H2-Ab1*), classical and minor MHCI genes, and MHCII genes. Note that for each thymocyte subset there are 3 replicates except for the MHC$^-$ DN4 cells for which there are duplicates. Asterisks highlight transcripts that are upregulated across all MHC$^-$ libraries on comparison with MHC$^+$ *Q10* ($p = 7 \times 10^{-5}$), *H2-T3 (TL)* ($p = 3 \times 10^{-7}$), *H2-T22* ($p = 1 \times 10^{-7}$) and *H2-T-ps* ($p = 4 \times 10^{-5}$). P calculated using two-tailed Chi-square test. **c**. Log_2-fold change in expression of all development stage marker genes depicted in Fig. 1a. **d**. Log_2-fold change in TCR Vβ chain segment (*Trbv*) expression. Mean depicted of triplicates for all libraries except for duplicates for MHC$^-$ DN4 samples. **e**. Log_2-fold change in *Bcl2a1* family transcripts (upper panel), canonical *Bcl2* transcripts (middle panel), and *Pim1* protooncogene (lower panel). Mean values presented. **f**. Log_2-fold change in TCR Vα chain segment (*Trav*) expression. Mean depicted of triplicates for all libraries except for duplicates for MHC$^-$ DN4 samples. **g**. Display of haematopoietic/immune organs from an MHC$^-$ dKO mouse with massive thymic growth at 15 months and from age-matched MHC$^+$ control. **h**. FACS analysis of single cell thymic and splenocyte suspensions stained for CD4 and CD8. Numbers next to gates indicate % of cells in that gate. **i**. Haematoxylin and eosin staining of representative organs from an age-matched MHC$^+$ wt B6 mouse and an MHC$^-$ dKO mouse with leukaemic growth. Thymic cortex indicated by 'c', and thymic medulla by 'm'. Cancellous bone indicated by 'ca'. Arrow indicates leukaemic cell accumulation adjacent and around a hepatic vein. **j**. Immunohistochemistry of tumour cells in dKO thymus for TdT (immature thymocytes), CD8 (T lineage) and neutrophil elastase (myeloid lineage) and in dKO spleen metastatic focus for the intracellular domain of Notch 1 (NICD1). **i, j**: For each tissue and condition, the complete section was examined down to the cellular level and the image presented (~1% of each total section) is representative of that complete section. White bar in all images represents 100 μm.

Extended Data Table 1 | Top 20 DPsm TCR β clonotypes developing on scH-2K^b stroma at d9

I.D.	Count	*Trbv*	*Trbd*	*Trbj*	*Trbc*	sequence	notes
01.	12	*Trbv15*	*Trbd1/Trbd2*	*Trbj2-3*	*Trbc2*	CASSFGTTSAETLYF	
02.	9	*Trbv3*	*Trbd1*	*Trbj2-3*	*Trbc2*	CASSRDRGAETLYF	L12 encoded by CTG
03.	1	*Trbv3*	*Trbd1*	*Trbj2-3*	*Trbc2*	CASSRDRGAETLYF	L12 encoded by TTG
04.	7	*Trbv13-2*	*Trbd2*	*Trbj2-7*	*Trbc2*	CASGVGLGGEQYF	
05.	6	*Trbv1*	*Trbd2*	*Trbj2-1*	*Trbc2*	CTCSAGTGNYAEQFF	
06.	6	*Trbv13-1*	*Trbd1*	*Trbj1-4*	*Trbc2*	CASSDGTGNERLFF	
07.	6	*Trbv13-2*	*Trbd1*	*Trbj2-3*	*Trbc2*	CASGDRTTSAETLYF	
08.	6	*Trbv17*	*Trbd2*	*Trbj2-3*	*Trbc2*	CASRPGLGGAETLYF	
09.	6	*Trbv29*	*Trbd2*	*Trbj2-3*	*Trbc2*	CASSLGWGAETLYF	
10.	6	*Trbv3*	*Trbd1*	*Trbj1-3*	*Trbc2*	CASSWTNSGNTLYF	
11.	6	*Trbv4*	*Trbd1*	*Trbj1-1*	*Trbc2*	CASSRQGAEVFF	A8 encoded by GCA
12.	1	*Trbv4*	*Trbd1*	*Trbj1-1*	*Trbc2*	CASSRQGAEVFF	A8 encoded by GCG
13.	5	*Trbv1*	*Trbd1/Trbd2*	*Trbj2-3*	*Trbc2*	CTCSADWGGAETLYF	
14.	5	*Trbv12-1*	*Trbd2*	*Trbj2-7*	*Trbc2*	CASSLRWGDEQYF	S3 encoded by AGC
15.	1	*Trbv12-1*	*Trbd2*	*Trbj2-7*	*Trbc2*	CASSLRWGDEQYF	S3 encoded by AGT
16.	5	*Trbv13-3*	*Trbd1*	*Trbj1-4*	*Trbc2*	CASRPGQGERLFF	
17.	5	*Trbv15*	*Trbd1*	*Trbj1-1*	*Trbc2*	CASSLRGTEVFF	
18.	5	*Trbv15*	*Trbd1*	*Trbj1-5*	*Trbc2*	CASSTGAPLF	
19.	5	*Trbv19*	*Trbd2*	*Trbj2-3*	*Trbc2*	CASSIWGGSAETLYF	
20.	5	*Trbv2*	*Trbd1*	*Trbj2-7*	*Trbc2*	CASSQGQGEQYF	

For 3 pairs of CDR3, the pair members are separate clonotypes developing from unique nucleotide sequences. The Count parameter represents the number of high quality, validated reads for a unique clonotype. Clone I.D.'s 14 and 15 encode a sequence already established as the CDR3 of the N15 TCR β chain known to recognize VSV8 peptide presented by H-2K^b both as a mature TCR including the N15 α chain and as a component of the preTCR[10].

Extended Data Table 2 | Non-classical MHC class I expression in MHC$^+$ and MHC$^-$ OP9-DL4

			OP9-DL4 (tpm)					OP9-DL4 (tpm)	
Gene	β2m	Ps	MHC$^+$	MHC$^-$	Gene	β2m	Ps	MHC$^+$	MHC$^-$
H2-Q1	+		33	32	*H2-M2*	+		0	0
H2-Q2	+		0	0	*H2-M3*	+		90	74
H2-Q3	+		0	0	*H2-M4-ps*	+	+	0	0
H2-Q4	+		86	2	*H2-M5*	+		41	46
H2-Q5	+		0	0	*H2-M6-ps*	+	+	4	2
H2-Q6	+		0	0	*H2-M7-ps*	+	+	0	41
H2-Q7	+		1	0	*H2-M8-ps*	+	+	0	0
H2-Q8	+		0	0	*H2-M9*	+		0	0
H2-Q9	+		0	0	*H2-M10.1*	+		0	0
H2-Q10	+		5	4	*H2-M10.2*	+		0	0
H2-T1	+		0	0	*H2-M10.3*	+		0	0
H2-T2	+		0	0	*H2-M10.4*	+		0	0
H2-T3 (TL)	+		0	0	*H2-M10.5*	+		0	0
H2-T4	+		0	0	*H2-M10.6*	+		0	0
H2-T5	+		0	0	*Cd1d1*	+		112	67
H2-T6	+		0	0	*Cd1d2*	+		0	0
H2-T7	+		0	0	*Raet1a*	no		0	0
H2-T8	+		0	0	*Raet1b*	no		0	0
H2-T9	+		0	0	*Raet1c*	no		0	0
H2-T10	+		51	13	***Raet1d***	**no**		**109**	**309**
H2-T-ps	(+)	(+)	0	0	***Raet1e***	**no**		**1128**	**1100**
H2-T12	+		0	0	*Fcgrt*	+		1867	1093
H2-T13	+		0	0	*Hfe*	+		954	1560
H2-T14	+		0	0	*Azgp1*	no		0	0
H2-T15	+		0	0	*Mr1*	+		798	571
H2-T16	+		0	0					
H2-T17	+		0	0					
H2-T18	+		0	0					
H2-T19	+		0	0	Other significant transcripts				
H2-T20	+		0	0					
H2-T21	+		0	0					
H2-T22	+		1161	753	*B2m*	mutated		9589	2003
H2-T23 (Qa-1^b)	+		404	315	*Tap2*	mutated		1196	630
H2-T24	+		21	47	*Dll4*	transduced		33378	55021
H2-Bl	+		0	0	*H2-K1*	+		327	95
H2-M1	+		0	0	*H2-D1*	+		573	481

Sixty-one non-classical MHC are listed[65]. Dependence on β2m indicated by '+' or 'no'. Genes labeled as pseudogene (Ps) may result in transcripts, initially classified as non-coding but subsequently found to be protein coding, as in the instances of *H2-Q5*, *H2-Q10, H2-T1, H2-T4, H2-T12, H2-T13, H2-T14, H2-M10.4, H2-M10.6*. *H2-T-Ps* has been provisionally redefined as protein coding. Genes in bold font are not β2m-dependent and have transcript levels above zero measured as transcripts per million (tpm). Small panel at bottom right presents expression data for the CRISPR/Cas9 targets *B2m* and *Tap2* deleted in the MHC$^-$ variant, Delta-like ligand 4, and major MHCI alleles.

Extended Data Table 3 | Well-represented clonotypes in MHC$^+$ and MHC$^-$ libraries

a

MHC$^+$ DN4

frequency	proportion	CDR3 amino acid	CDR3 length	*Trbv*
5	0.00197	CASSQDRANTEVFF	14	*Trbv5*
4	0.00158	CASSWDNYAEQFF	13	*Trbv10*
4	0.00158	CASSQGQGANTEVFF	15	*Trbv2*
4	0.00158	**CGARGAEVFF**	10	*Trbv20*
3	0.00118	CASGDAGQGGGAETLYF	17	*Trbv12-2/13-2*
3	0.00118	CASRDRSSYEQYF	13	*Trbv13-1*
3	0.00118	CASSDRGVSNERLFF	15	*Trbv13-3*
3	0.00118	CASSPGLGGREQYF	14	*Trbv15*
3	0.00118	CASSIHSGNTLYF	13	*Trbv19*
3	0.00118	**CASSPGQGADTGQLYF**	16	*Trbv19*
3	0.00118	CASSQDRAETLYF	13	*Trbv2*
3	0.00118	CASSQGHQNTLYF	13	*Trbv2*
3	0.00118	CASSQDRGEQYF	12	*Trbv2*
3	0.00118	CGARDTNTEVFF	12	*Trbv20*
3	0.00118	CGARDTNSDYTF	12	*Trbv20*
3	0.00118	CGARTGGYEQYF	12	*Trbv20*
3	0.00118	CGARDRGREQYF	12	*Trbv20*
3	0.00118	CASSGTYEQYF	11	*Trbv26*
3	0.00118	CASSLGQGANERLFF	15	*Trbv3*
3	0.00118	CASSLADWGDTQYF	14	*Trbv3*

MHC$^-$ DN4

frequency	proportion	CDR3 amino acid	CDR3 length	*Trbv*
21	0.00745	**CTCSADWGGANQDTQYF**	17	*Trbv1*
17	0.00603	CTCSPGLGGEQYF	13	*Trbv1*
15	0.00532	**CASRDSQNTLYF**	12	*Trbv19*
14	0.00497	CASGDETGVSYEQYF	15	*Trbv12-2*
13	0.00461	CASGDSTGGDQDTQYF	16	*Trbv12-2*
13	0.00461	**CASSFPASQNTLYF**	14	*Trbv14*
13	0.00461	**CASSQDWLNQDTQYF**	15	*Trbv5*
12	0.00426	**CASSLSRGEQYF**	12	*Trbv29*
11	0.00390	**CASRRTGGAGAEQFF**	15	*Trbv26*
10	0.00354	**CASSYNSGNTLYF**	13	*Trbv10*
10	0.00354	**CASSLEDNYAEQFF**	14	*Trbv29*
9	0.00319	**CASSEGATEVFF**	12	*Trbv13-3*
9	0.00319	**CASSLRENTLYF**	12	*Trbv12-2*
9	0.00319	**CASGEGRDFQDTQYF**	15	*Trbv12-2*
9	0.00319	**CASGEQYF**	8	*Trbv12-2*
9	0.00319	**CASSQEWGVQDTQYF**	15	*Trbv5*
8	0.00283	**CASSDGTGASAETLYF**	16	*Trbv13-1*
8	0.00283	**CASSGTYEQYF**	11	*Trbv13-1*
8	0.00283	**CAWSSGTGGYEQYF**	14	*Trbv31*
8	0.00283	**CASSPDRGPEVFF**	13	*Trbv5*

b

MHC$^+$ DP

frequency	proportion	CDR3 amino acid	CDR3 length	*Trbv*
5	0.00197	CASSLQGANSDYTF	14	*Trbv12-2*
4	0.00158	CTCSAAGTGPNERLFF	16	*Trbv1*
4	0.00158	CASSRQGANTEVFF	14	*Trbv10*
4	0.00158	CGARDTNTEVFF	12	*Trbv20*
3	0.00118	CASGEDTNSDYTF	13	*Trbv12-2/13-2*
3	0.00118	CASGDDSQNTLYF	13	*Trbv12-2/13-2*
3	0.00118	CASSLRGGGGAETLYF	16	*Trbv15*
3	0.00118	CASSQDWGGYEQYF	14	*Trbv2*
3	0.00118	CASSSGTGDNQAPLF	15	*Trbv29*
3	0.00118	CASSLGDSNERLFF	14	*Trbv3*
3	0.00118	CAWSLTGQLYF	11	*Trbv31*
3	0.00118	CASSQDNTEVFF	12	*Trbv5*
3	0.00118	CASSLDSSYEQYF	13	*Trbv12-2*
3	0.00118	CASSLGTGEDTQYF	14	*Trbv16*
3	0.00118	CASSQEDRDTEVFF	12	*Trbv2*
3	0.00118	CASSQDEQYF	10	*Trbv2*
3	0.00118	CGARDTGSDYTF	12	*Trbv20*
3	0.00118	CASSRDWGYEQYF	13	*Trbv3*
3	0.00118	CASSQDRGQNTLYF	14	*Trbv5*
3	0.00118	CASSQDSSYEQYF	13	*Trbv5*

MHC$^-$ DP

frequency	proportion	CDR3 amino acid	CDR3 length	*Trbv*
14	0.00152	CASGDAGGTGQLYF	14	*Trbv12-2/13-2*
14	0.00152	CASSRDRGQDTQYF	14	*Trbv17*
13	0.00141	**CASSQQGANTEVFF**	14	*Trbv5*
12	0.00131	CASSRRANTEVFF	13	*Trbv12-2*
11	0.00120	**CASSLGTGGEQYF**	13	*Trbv15*
10	0.00109	**CASGDNSPLYF**	11	*Trbv12-2/13-2*
10	0.00109	**CGARGHTEVFF**	11	*Trbv20*
10	0.00109	CASSRDTNTEVFF	13	*Trbv3*
9	0.00098	CASSLDRGRQNTLYF	15	*Trbv12-2*
9	0.00098	CASSDGTANTEVFF	14	*Trbv13-3*
9	0.00098	**CGARDRANTEVFF**	13	*Trbv20*
8	0.00087	CTCSAQANTEVFF	13	*Trbv1*
8	0.00087	CTCSADQGAETLYF	14	*Trbv1*
8	0.00087	CASSLDRDRGAEQFF	15	*Trbv16*
8	0.00087	CASSLDWGGAEQFF	14	*Trbv16*
8	0.00087	**CASRQGAGQLYF**	12	*Trbv19*
8	0.00087	CASSQDTNSDYTF	13	*Trbv2*
8	0.00087	CASSQQGAEVFF	12	*Trbv5*
7	0.00076	CTCSADQDTQYF	12	*Trbv1*
7	0.00076	**CASSLDSQNTLYF**	13	*Trbv14*

a. Top 20 clonotypes proliferating in the DN4 libraries of cells developing on MHC$^+$ and MHC$^-$ stroma. Frequency refers to cell number expressing the identical clonotype. Clonotypes present in the DN4 unusual cluster are highlighted in bold. Note that the *Trbv12-2/13-2* transcript is not a mix of *Trbv12-2* and *Trbv13-2* but rather the result of an independent recombination event between the 5′ end of *Trbv12-2* and the 3′ end of *Trbv13-2*, an event recently and frequently detected in 10X TCR single cell repertoire analyses. **b**. Top 20 clonotypes proliferating in the DP libraries of cells developing on MHC$^+$ and MHC$^-$ stroma. Clonotypes present in the DP abnormal cluster are highlighted in bold.

nature portfolio

Corresponding author(s): Jonathan S. Duke-Cohan, Ellis L. Reinherz

Last updated by author(s): Dec 6, 2022

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Software was used to collect sequencing data direct from the HiSeq,MiSeq and Novaseq machines but this is code proprietary for Illumina and intrinsic to the sequencing machines.
Data analysis	Sequencing data output from was converted from .bcl format to .fastq format and demultiplexed using the "bcl2fastq v2.0" module directly at the Novogene sequencing facility and Medgenome sequencing facility and subsequently processed by these commercial entities to yield normalized gene expression values either as transcripts per million reads (TPM) or fragments per kilobase transcript per million mapped reads (FPKM). For the scRNA-seq analysis, all .fastq for the gene expression files were pipelined through Cellranger 3.1.0 (10X Genomics) for alignment to the mouse GRCm38.p6/mm10 reference genome and UMAP projection using default settings for Cellranger. For the targeted scRNA-Seq clonotype analysis, all .fastq files were pipelined through Cellranger V(D)J 3.1.0 for alignment to the mouse V(D)J : vdj_GRCm38_alts_ensembl-3.1.0.gz-3.1.0 reference genome using default settings for Cellranger V(D)J. All gene expression data was analysed using the Loupe browser 4.2.0 (10X Genomics) and all targeted TCR data analysed using the Loupe v(D)J browser 3.0.0 (10X Genomics). Target TCR β chain data from the bulk RNA analyses was demultiplexed from the .fastq sequence files using the barcode splitter module of the FASTX toolkit (http://hannonlab.cshl.edu/fastx_toolkit/index.html) using default parameters. The demutiplexed library sequences were aligned to TCR Vβ regions in the GRCm38.p6/mm10 mouse genome followed by clone assembly and CDR3 extraction using the MiXCR suite running under Java (https://milaboratories.com/software). For UMI-tagged clonotypes, demultiplexing, alignment and assembly were accomplished using the MIGEC package (https://github.com/mikessh/migec). Further analysis for clonotype representation and tracking utilised the VDJtools package also running under Java (https://github.com/mikessh/vdjtools). All Flow cytometry and FACS analysis was acquired using FACSDiva v8.0 (BD Biosciences) and analysed using FlowJo v10 (BD). All qRT-PCR analysis was performed using the LinRegPCR 2021.2 (https://medischebiologie.nl/files/) All statistical analysis (t- test, Chi-square test) was performed using Graphpad web portal (no version number; https://www.graphpad.com/quickcalcs/)

Gene set enrichment analysis accessed through the ssGSEA module v10.1.x of GenePattern.org (https://gsea-msigdb.github.io/ssGSEA-gpmodule/v10/index.html)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All sequence files are deposited in NCBI Gene Expression Omnibus (GEO) accession GSE186049, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE186049
Immune Genome Project, https://www.immgen.org/
MSigDB, https://www.gsea-msigdb.org/gsea/msigdb/
GRCm38.p6/mm10 mouse, https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF_000001635.26/

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	For every sample, limits were placed by pragmatic laboratory parameters such as cell yield from in vitro cultures, FACS sorting time and sequencing capacity. For every sample, scRNA-Seq or bulk RNA samples, 10,000 cells was chosen as sample size meeting the technical capacity standards imposed while maintaining a robust statistical number acceptable in the field. No sample size calculations were performed. In contrast, for the animal experiments, for TCR repertoire analysis of early developing thymocytes, the whole thymus of each appropriate mouse was used for preparing a TCR beta chain-targeted library. In this instance, thymi from 3 wild-type, and 3 MHCI, MHCII double-knockout, animals were used to generate 6 independent libraries representing 2 conditions with 3 biological replicates for RNA-Seq. This arrangement is standard in the field.
Data exclusions	No data was excluded. In order to maintain the pro-apoptotic and apoptotic populations, no mitochondrial transcript gating was performed. Gene expression analyses, however, were only based on data that met the Padj <0.05 threshold for differential expression.
Replication	For the bulk RNA TCR beta clonotype analyses, samples consisted of 10,000 cells from each thymocyte subpopulation, grown on each stromal substrate, and repeated 6 times. Although scRNA-Seq by its very nature has internal replicates, the experiment here is n = 1 although 8 independent libraries were generated. For selected genes of particular interest, the developmental set-up was repeated but analysis was performed by qRT-PCR rather than RNA-Seq to successfully validate initial results.
Randomization	Although randomization was not introduced deliberately in the experimental system, it exists ab initio as gender of embryos providing foetal liver could not be determined visually and cells in aliquots of 2,000 cells of multiple embryo origin were seeded onto MHC+ control or MHC- experimental stromal support cells.
Blinding	For both scRNA-Seq and bulk RNA-Seq, data analysis was a pipeline of initial sequencing files through a series of software modules leading to final output of gene expression values and presence of TCR beta chain clonotypes - a process not formally blind since it was known which sample was being processed but practically blind as the 10X Genomic analytical software leaves no room for intervention prior to output. No deliberate blinding was introduced to any experimental procedure.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

Antibody; Clone; Label; Manufacturer; Cat #; Research Resource Identifier (RRID; https://www.rrids.org);
url with validated results

---

anti-mouse CD3e; clone 145-2C11; BV605; Biolegend; 100351; AB_2565842
https://www.biolegend.com/en-us/products/brilliant-violet-605-anti-mouse-cd3epsilon-antibody-11974
anti-mouse CD4; RM4-5; Pacific Blue; BioLegend; 100531; AB_493374;
https://www.biolegend.com/en-us/products/pacific-blue-anti-mouse-cd4-antibody-2855
anti-mouse CD8b.2; 53-5.8; PE; BioLegend; 140408; AB_10644002;
https://www.biolegend.com/en-us/products/pe-anti-mouse-cd8b-2-antibody-6871
anti-mouse Ly-6A/E (Sca-1); D7; FITC; eBioscience; 11-5981-81; AB_465332;
https://www.thermofisher.com/antibody/product/Ly-6A-E-Antibody-clone-D7-Monoclonal/A14731
anti-mouse CD117; 2B8; APC; BD Biosciences; 553356; AB_398536;
https://www.bdbiosciences.com/en-us/products/reagents/flow-cytometry-reagents/research-reagents/single-color-antibodies-ruo/apc-rat-anti-mouse-cd117.553356
anti-mouse CD45; 30-F11; APC; BioLegend; 103112; AB_312977;
https://www.biolegend.com/en-us/products/apc-anti-mouse-cd45-antibody-97
anti-mouse CD25; PC61.5; PE-Cy7; eBioscience; 25-0251-82; AB_468733;
https://www.thermofisher.com/antibody/product/CD25-Antibody-clone-PC61-5-Monoclonal/25-0251-82
anti-mouse CD44; IM7; APC-Cy7; BD Biosciences; 560568; AB_1727481;
https://www.bdbiosciences.com/en-us/products/reagents/flow-cytometry-reagents/research-reagents/single-color-antibodies-ruo/apc-cy-7-rat-anti-mouse-cd44.560568
anti-mouse CD4; RM4-5; BV711; BioLegend; 100549; AB_11219396;
https://www.biolegend.com/en-us/products/brilliant-violet-711-anti-mouse-cd4-antibody-7925
anti-mouse CD8a; 53-6.7; PerCP/Cy5.5; BioLegend; 100734; AB_2075238;
https://www.biolegend.com/en-us/products/percp-cyanine5-5-anti-mouse-cd8a-antibody-4255
anti-mouse CD28; E18; PE; BioLegend; 122010; AB_604078;
https://www.biolegend.com/en-us/products/pe-anti-mouse-cd28-antibody-3778
anti-mouse TCRg/d; GL3; Biotin; BioLegend; 118103; AB_313827;
https://www.biolegend.com/en-us/products/biotin-anti-mouse-tcr-gamma-delta-antibody-2419
anti-mouse CD11c; N418; Biotin; BioLegend; 117303; AB_313772;
https://www.biolegend.com/en-us/products/biotin-anti-mouse-cd11c-antibody-1814
anti-mouse CD11b; M1/70; Biotin; BioLegend; 101203; AB_312786;
https://www.biolegend.com/en-us/products/biotin-anti-mouse-human-cd11b-antibody-346
anti-mouse NK1.1; PK136; Biotin; BioLegend; 108703; AB_313390;
https://www.biolegend.com/en-us/products/biotin-anti-mouse-nk-1-1-antibody-428
anti-mouse Gr-1; RB6-8C5; Biotin; BioLegend; 108403; AB_313368;
https://www.biolegend.com/en-us/products/biotin-anti-mouse-ly-6g-ly-6c-gr-1-antibody-457
anti-mouse Ter-119; TER-119; Biotin; BioLegend; 116203; AB_313704;
https://www.biolegend.com/en-us/products/biotin-anti-mouse-ter-119-erythroid-cells-antibody-1864
anti-mouse CD19; 6D5; Biotin; BioLegend; 115503; AB_313638;
https://www.biolegend.com/en-us/products/biotin-anti-mouse-cd19-antibody-1527
anti-mouse CD24; M1/69; FITC; BioLegend; 101805; AB_312838;
https://www.biolegend.com/en-us/products/fitc-anti-mouse-cd24-antibody-341
anti-mouse Elastase; E8U3X; unlabeled; Cell Signaling Technology; 90120S; none;
https://www.cellsignal.com/products/primary-antibodies/neutrophil-elastase-e8u3x-rabbit-mab/90120?_=1666979524930&Ntt=90120S&tahead=true
anti-mouse Notch1 ICD; D3B8; unlabeled; Cell Signaling Technology; 4147T; AB_2153348;
https://www.cellsignal.com/products/primary-antibodies/cleaved-notch1-val1744-d3b8-rabbit-mab/4147?site-search-type=Products&N=4294956287&Ntt=4147t&fromPage=plp&_requestid=6095913
anti-mouse CD8a; D4W2Z; unlabeled; Cell Signaling Technology; 98941; AB_2756376;
https://www.cellsignal.com/products/primary-antibodies/cd8a-d4w2z-xp-rabbit-mab-mouse-specific/98941?_=1666979569545&Ntt=98941&tahead=true
anti-mouse TdT; EPR2976Y; unlabeled; Abcam; ab232378;AB_2094447;
https://www.abcam.com/tdt-antibody-epr2976y-bsa-and-azide-free-ab232378.html

Validation

All antibodies used have RRID numbers provided above and validated by the Antibody Registry. Listed URL links to validated results

Validation	for each antibody

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	OP9-DL4 parent: Dr. J. Zuniga-Pflucker, Sunnybrook Research Institute, Toronto, Canada. Derivative OP-DL4 cell lines: developed in-house. See: Das, D. K. et al. Pre-T Cell Receptors (Pre-TCRs) Leverage Vbeta Complementarity Determining Regions (CDRs) and Hydrophobic Patch in Mechanosensing Thymic Self-ligands. J Biol Chem 291, 25292-25305, doi:10.1074/jbc.M116.752865 (2016).
Authentication	All cell lines used are the parental cell lines (OP9-DL4, MHC+), B2m/Tap2 knockout (KO; OP9-DL4 MHC-) or MHCI retransfected versions of the KO cell line (OP9-DL4 scH-2Kb). Parental OP9-DL4, and all clones derived from the parent OP9-DL4 are clonal and authenticated by FACS analysis for presence or absence of MHC class I, MHC II, H-2Kb, CD1d and mouse b2m. The OP9-DL4 and OP9-DL4 MHC- have also been validated by transcriptome analysis.
Mycoplasma contamination	Tested bi-annually using the Lonza quick mycoplasma identification kit (luminescence assay) and have never tested positive.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used in this study

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	Mouse, C57Bl/6, adult females breeding with adult males were used to establish pregnant dams - embryos were isolated at e14.5 for isolation of haematopoietic stem cells from foetal liver. Embyros were not sexed and male and female embryos were used interchangeably. For sequencing from whole thymi for gene expression and targeted TCR beta clonotype analysis, C57Bl/6 mice and mice on the same background bearing a double knockout of the B2m gene and the H2-Ab1 gene (B6.129-H2-Ab1tm1Gru B2mtm1Jae N1721), all 3 weeks of age, were obtained from Taconic Biosciences.
Wild animals	This study did not involve wild animals
Field-collected samples	This study did not involve samples collected from the field
Ethics oversight	Ethical oversight of all animal use is by the Animal Research Facility and the Animal Care and Use Committee (ACUC) of the Dana-Farber Cancer Institute.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Early DN to DP thymocytes developing from haematopoietic stem cells after 9 days culture in an in vitro support system.
Instrument	Becton-Dickinson FACS Aria II cell sorter
Software	Instrument software is Becton-Dickinson FACSDiva v8.0 and post-sort analysis performed by FlowJo v10 (Becton-Dickinson)
Cell population abundance	Representative post-sort abundance profiles for the sorting of thymocyte subsets for bulk TCR beta chain repertoire analysis, and for scRNA-Seq of defined thymocyte subset, are presented in the Extended Data section.
Gating strategy	For all experiments, strategy was similar. The total cell population consisting of mainly suspension thymocytes and some released stromal support cells was stained as described in the Methods section. Initial size discrimination on a FSC-A/SSC-A plot allows discrimination of the thymocyte population from most of the much larger stromal cells. Further gating (SSC-W/SSC-H then FSC-W/FSC-H) isolates singlet cells. Gating on SSC-A/Zombie Aqua negative cells isolates the viable cell population. Gating then on the GFP-negative/CD45-positive population isolates viable, single, haematopoetic cells away from the remaining stromal cells that express GFP in a bicistronic cassette with Delta-like ligand 4 (DL4). All the alpha/beta T lineage cells should fall in the BV421-negative gate, simultaneously confirming the loss of all GFP signal in the gated cells. This population, now considered to be uniform alpha/beta T lineage thymocytes, is now stringently gated on CD8 and CD4 to give

a DN (CD4-CD8-) population and a DP (CD4+CD8+) population. Gating the DN population on the CD44-negative cells, and then gating on CD25 and CD28 yields DN3a (CD44-CD25+CD28-), DN3b (CD44-CD25+CD28+), and DN4 (CD44-CD25-CD28+) thymocytes, respectively. Sorting the the CD8+ cells from the lin- CD4- population (containing DN and CD8 cells), these cells were then gated on CD24hi CD3elo to yield the immature CD8 single positive thymocytes (ISP).

In addition to Extended data Fig.1, a further 11 fully annotated gating results are presented in Supplementary File 4.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.