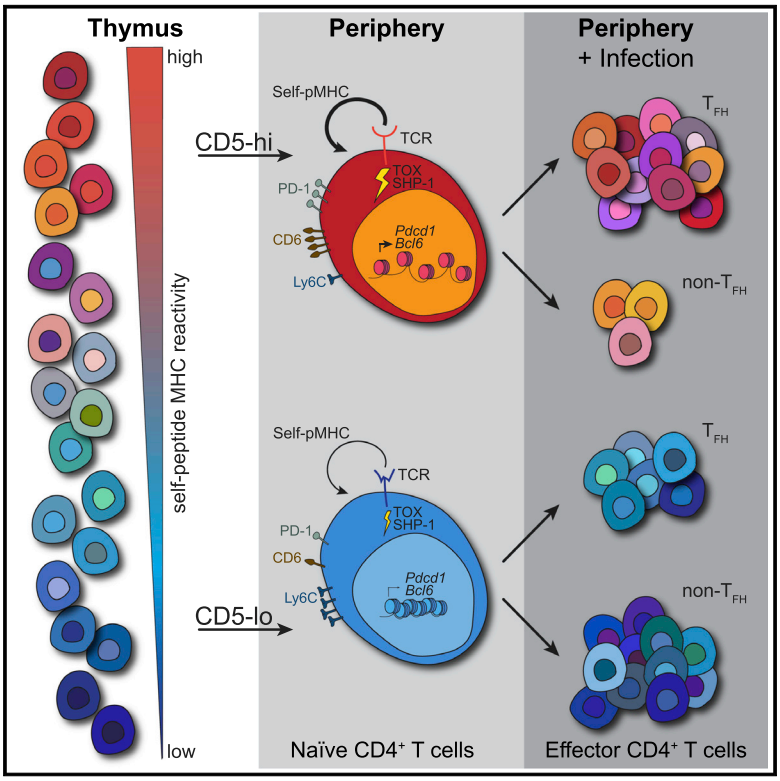
Article

Cell Reports

Pre-existing chromatin accessibility and gene expression differences among naive CD4$^+$ T cells influence effector potential

Graphical abstract



Authors

Dakota Rogers, Aditi Sood, HanChen Wang, ..., Heather J. Melichar, Johannes Textor, Judith N. Mandl

Correspondence

judith.mandl@mcgill.ca

In brief

CD4$^+$ T cell effector responses are remarkably diverse. Rogers et al. reveal that, even prior to foreign antigen encounter, naive CD4$^+$ T cell heterogeneity is only partially maintained by peripheral self-peptide MHC interactions. Thymic development pre-wires chromatin and transcriptional networks that contribute to biased effector responses.

Highlights

- Differences among naive CD4$^+$ T cells are driven by TCR genes and chromatin modifiers
- Pre-existing gene expression differences are maintained post-activation
- CD5^{hi} naive CD4$^+$ T cells have a greater propensity to become T$_{FH}$ cells
- Distinct chromatin accessibility landscapes are established during thymic development

Rogers et al., 2021, Cell Reports *37*, 110064
November 30, 2021
https://doi.org/10.1016/j.celrep.2021.110064

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

Pre-existing chromatin accessibility and gene expression differences among naive CD4$^+$ T cells influence effector potential

Dakota Rogers,[1,2] Aditi Sood,[3,4] HanChen Wang,[1,2] Jasper J.P. van Beek,[5] Thomas J. Rademaker,[6] Patricio Artusa,[1,2] Caitlin Schneider,[2,7] Connie Shen,[2,7] Dylan C. Wong,[2,7] Aanya Bhagrath,[1,2] Marie-Ève Lebel,[3] Stephanie A. Condotta,[7,13] Martin J. Richer,[7,13] Andrew J. Martins,[8] John S. Tsang,[8] Luis B. Barreiro,[9] Paul François,[6] David Langlais,[2,10,11] Heather J. Melichar,[3,12] Johannes Textor,[5] and Judith N. Mandl[1,2,7,14,*]

[1]Department of Physiology, McGill University, Montreal, QC, Canada
[2]McGill University Research Centre on Complex Traits, Montreal, QC, Canada
[3]Immunology-Oncology Unit, Maisonneuve-Rosemont Hospital Research Center, Montreal, QC, Canada
[4]Department of Microbiology, Immunology, and Infectious Disease, Université de Montréal, Montreal, QC, Canada
[5]Department of Tumor Immunology, Radboud Institute for Molecular Life Sciences, Nijmegen, the Netherlands
[6]Department of Physics, McGill University, Montreal, QC, Canada
[7]Department of Microbiology and Immunology, McGill University, Montreal, QC, Canada
[8]Multiscale Systems Biology Section, Laboratory of Immune System Biology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, USA
[9]Department of Medicine, Genetic Section, University of Chicago, Chicago, IL, USA
[10]Department of Human Genetics, McGill University, Montreal, QC, Canada
[11]McGill University Genome Centre, Montreal, QC, Canada
[12]Department of Medicine, Université de Montréal, Montréal, QC, Canada
[13]Present address: Department of Microbiology & Immunology, Indiana University School of Medicine, Indianapolis, IN, USA
[14]Lead contact
*Correspondence: judith.mandl@mcgill.ca
https://doi.org/10.1016/j.celrep.2021.110064

SUMMARY

CD4$^+$ T cells have a remarkable potential to differentiate into diverse effector lineages following activation. Here, we probe the heterogeneity present among naive CD4$^+$ T cells before encountering their cognate antigen to ask whether their effector potential is modulated by pre-existing transcriptional and chromatin landscape differences. Single-cell RNA sequencing shows that key drivers of variability are genes involved in T cell receptor (TCR) signaling. Using CD5 expression as a readout of the strength of tonic TCR interactions with self-peptide MHC, and sorting on the ends of this self-reactivity spectrum, we find that pre-existing transcriptional differences among naive CD4$^+$ T cells impact follicular helper T (T_{FH}) cell versus non-T_{FH} effector lineage choice. Moreover, our data implicate TCR signal strength during thymic development in establishing differences in naive CD4$^+$ T cell chromatin landscapes that ultimately shape their effector potential.

INTRODUCTION

Heterogeneity is a fundamental property of cellular systems (Altschuler and Wu, 2010; Mayer et al., 2016). Even clonally derived cell populations exhibit variations in gene expression that impact cell fate decisions (Carter and Zhao, 2021; Chang et al., 2008). Single-cell studies of immune cells have begun to reveal the diversity present in populations thought to be homogeneous and emphasized its role in the immune response (Villani et al., 2017; Xie et al., 2020). Such heterogeneity is perhaps nowhere as intimately tied to cellular function as it is in adaptive immune cells. T cell populations comprise a breadth of T cell receptors (TCRs), with individual cells expressing unique TCRs generated by somatic recombination (Schatz and Ji, 2011). CD4$^+$ T cells, in particular, possess a remarkable capacity for diversification into distinct effector lineages following activation that are defined by the cytokines they make and immune cells they act on. Indeed, CD4$^+$ T cell effector fate is critical in tailoring an immune response to the specific pathogen encountered (Zhou et al., 2009; Zhu et al., 2010).

CD4$^+$ T cell differentiation relies on dynamic metabolic and transcriptional changes (Almeida et al., 2016; Rodriguez et al., 2015). An early CD4$^+$ T cell fate decision is between effector subsets (T helper [Th] 1, Th2, Th9, and Th17) and follicular helper T (T_{FH}) cells. T_{FH} cells are essential to germinal center (GC) formation and generation of high-affinity plasma and memory B cells. T_{FH} development requires TCR engagement, expression of the lineage-defining transcription factor Bcl6 with inhibition of Blimp1, and upregulation of surface proteins, such as PD-1, CXCR5, and ICOS, enabling interaction with B cells (Crotty, 2019; Ruterbusch et al., 2020). Remarkably, a single CD4$^+$ T cell clone can expand into both T_{FH} and non-T_{FH} cells

(Becattini et al., 2015; Tubo et al., 2013). CD4$^+$ T cell clones are not equal, however. The propensity of a single CD4$^+$ T cell to differentiate into one helper subset over another varies between cells expressing distinct TCRs (Cho et al., 2017; Khatun et al., 2021; Tubo et al., 2013). Whether this clonal variability in cell fate is determined entirely by TCR signal strength upon antigen encounter or whether naive T cells are pre-wired with specific effector biases remains incompletely understood.

It is increasingly appreciated that naive CD4$^+$ T cells already differ prior to antigen stimulation. Despite their quiescence, naive CD4$^+$ T cells remain ready to rapidly respond to antigen (Chapman et al., 2020; Stefanová et al., 2002; Wolf et al., 2020). This is mediated, at least in part, by sub-threshold interactions with self-peptide presented on MHC (self-pMHC) that provide tonic survival signals (Vrisekoop et al., 2014). Several markers provide readouts of self-pMHC signal strength. Ly6C distinguishes naive CD4$^+$ T cells with high (Ly6C$^-$) from low (Ly6C$^+$) self-pMHC reactivity (Guichard et al., 2017; Martin et al., 2013). The expression levels of two other markers, CD5, a surface protein and negative regulator of TCR signaling, and the orphan nuclear receptor Nur77 (*Nr4a1*), positively correlate with sub-threshold TCR signal strength (Azzam et al., 1998; Moran et al., 2011). Measures of CD5 expression revealed that self-reactivity of naive T cells spans a wide spectrum, whereby the upper and lower bounds are likely set in the thymus by positive and negative selection (Fulton et al., 2015; Mandl et al., 2013). Even within a monoclonal TCR transgenic (Tg) population, cells may vary with regard to self-pMHC reactivity (Zinzow-Kramer et al., 2019). Importantly, not only does self-pMHC reactivity impact competition between T cells for homeostatic signals (Vrisekoop et al., 2017), it also influences their responses to antigen. CD5^{hi} naive CD4$^+$ T cells express higher basal levels of NF-κB, phosphorylated TCRζ, and ERK; make more IL-2 post-activation; preferentially differentiate into regulatory T cells (Tregs); predominate in acute infections; and contribute disproportionately to the memory T cell pool (Fulton et al., 2015; Henderson et al., 2015; Mandl et al., 2013; Matson et al., 2020; Persaud et al., 2014). Similarly, Ly6C$^-$ cells preferentially differentiate into Tregs and Th17 cells (Martin et al., 2013). In contrast, CD5^{lo} naive CD4$^+$ T cells produce more IFNγ upon activation (Sood et al., 2019). However, it remains unclear whether T cells are pre-programmed entirely by tonic peripheral TCR signals or whether early self-ligand encounters in the thymus play a role.

Here, we adopted an unbiased systems approach, combining single-cell RNA sequencing (scRNA-seq) with bulk RNA-seq and assay for transposase-accessible chromatin using sequencing (ATAC-seq) to comprehensively investigate the drivers of heterogeneity among naive CD4$^+$ T cells. We report that individual cell-level biases in the expression of modulators of TCR signal strength and in T_{FH} versus non-T_{FH} effector lineage choice are driven, at least in part, by pre-existing transcriptional and epigenetic (chromatin accessibility) differences between cells with low (CD5^{lo}) versus high (CD5^{hi}) self-pMHC reactivity. Unexpectedly, our data reveal that many of the differences in gene expression among naive CD4$^+$ T cells do not require continuous signaling through the TCR via self-pMHC interactions but are likely a result of variable signal strengths obtained during development.

RESULTS

TCR signaling-induced gene expression differences are drivers of variability among individual naive CD4$^+$ T cells

To define transcriptional differences among naive CD4$^+$ T cells with single-cell resolution, we performed scRNA-seq on 1,152 cells sorted from the spleen, of which 697 passed quality control (Table S1). For each individual T cell, we measured CD5 protein expression during sorting given prior studies implicating CD5 as a key readout of diversity among naive T cells (Fulton et al., 2015; Mandl et al., 2013; Persaud et al., 2014). We verified that *Cd5* transcript counts correlated with CD5 protein (Figure S1A), consistent with prior evidence for CD5 transcriptional regulation (Arman et al., 2004; Tung et al., 2001). In total, we detected 14,040 genes and an average of 1,389 genes per cell. Among the top 5% most variable genes, after accounting for technical variation, non-detection drop-outs, and removing mitochondrial genes (Table S2) (Lun et al., 2016), we detected genes important for trafficking between blood and lymphoid organs, including *Cd69*, *S1pr1*, *Sell*, *Klf2*, *Itag4*, *Tln1*, and *Foxo1*; genes involved in TCR signaling, such as *Ptpn6*, *Folr4*, *Il7r*, *Cd4*, *Jun*, *Il2rg*, *Thy1*, *Lck*, *Klf6*, *Bcl2*, *Nr4a1*, and also *Cd5*; and chromatin modifiers, including *Dnmt1*, *Hdac4*, *Sirt1*, and *Smc4* (Figure 1A). Gene Ontology (GO) enrichment analyses of the most variable genes showed an enrichment of GO terms associated with TCR signaling, including αβ T cell activation and T cell selection (Figure 1B). UMAP analysis of the 2,000 most variable genes did not reveal any detectable sub-clustering of naive cells by CD5 expression (Figure S1B). However, principal-component analyses (PCAs) based on the expression of 55 genes in the GO terms involved in T cell activation from Figure 1B (Table S3) showed that the transcriptional state of CD5^{lo} naive CD4$^+$ T cells differed from that of CD5^{hi} naive CD4$^+$ T cells, albeit with considerable overlap between the two populations (Figure 1C). Overall, our data highlighted the diversity between individual naive CD4$^+$ T cells that was detectable at the transcript level and identified expression differences in genes involved in TCR signaling as key drivers of cellular variability.

CD5 expression reveals the existence of distinct gene-expression profiles and chromatin landscapes in naive CD4$^+$ T cells

To characterize the transcriptional heterogeneity of naive CD4$^+$ T cells in greater depth, we performed bulk RNA-seq, focusing on the top and bottom 15% of the self-reactivity spectrum defined by CD5 expression. We used FoxP3^{GFP+} reporter mice to exclude Tregs, including them as an "outgroup" in our initial analyses (Figures 2A, 2B, S2A, and S2B) because Tregs express high levels of CD5 (Ordoñez-Rueda et al., 2009). We confirmed that sorted naive T cells did not contain Tregs (Figures S2C and S2D). Given the detection of chromatin modifiers in our scRNA-seq (Figure 1A), we also performed ATAC-seq to investigate the open-chromatin landscape of CD5^{lo} and CD5^{hi} naive CD4$^+$ T cells. PCAs highlighted that at the transcriptional level, CD5^{lo} and CD5^{hi} cells clustered into distinct populations

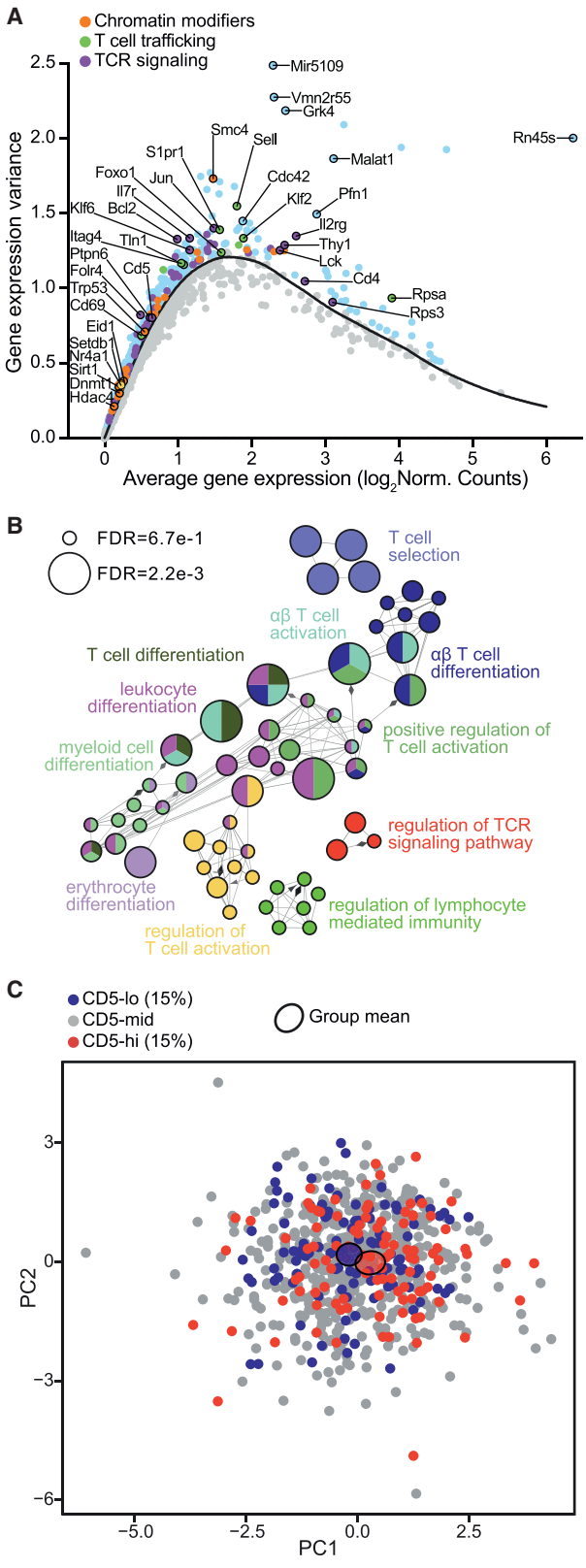


Figure 1. Cellular variability among naive CD4⁺ T cells at the single-cell level is driven by TCR signaling gene expression

(A) Plot of top 5% most variable genes (blue) across individual sorted naive CD4⁺ T cells. Select genes involved in TCR signaling (purple), T cell lymphoid organ trafficking (green), and chromatin modification (orange) are labeled.

(B) Immune system process GO enrichment analyses of top 5% variable genes from (A). Circles correspond to unique GO groups with related groups coded in the same color. Circle size reflects enrichment significance (FDR cut-offs are shown).

(C) PCA projection of scRNA-seq profiles based on genes that were in the top 5% most variable genes and also in GO terms involved in T cell activation from (B). The surface protein expression of CD5 is overlaid with the 15% $CD5^{lo}$ (blue) and $CD5^{hi}$ (red) naive $CD4^+$ T cells. Shaded colored circles represent group means. Each data point represents a single cell.

See also Figure S1; Tables S1, S2, and S3.

(Figure 2C). The $CD5^{lo}$ and $CD5^{hi}$ cells were distinguishable in PC2 and separated from Tregs in PC1 (Figure 2C). The $CD5^{lo}$ and $CD5^{hi}$ segregation was similarly reflected by their chromatin accessibility profiles (Figure 2D). Overall, $CD5^{lo}$ and $CD5^{hi}$ cells had comparable numbers and genomic annotation of accessible peaks (Figures S2E and S2F). Importantly, we identified a total of 1,006 differentially expressed genes (DEGs) at false discovery rate (FDR) $\leq$ 0.01 between $CD5^{lo}$ and $CD5^{hi}$ cells, of which 90% were found among detected differentially accessible regions (DARs) (Figure 2E). Indeed, there was significantly greater correspondence between gene expression and chromatin accessibility among DEGs than for a random gene set (Figures 2F and S2G). Plotting the fold changes (FCs) between $CD5^{lo}$ and $CD5^{hi}$ naive $CD4^+$ T cells from the bulk RNA-seq versus the ATAC-seq emphasized the dataset concordance (Figure S2H). Among the DEGs and DARs identified was *Cd5* itself, which had a ~6.7-fold greater transcript expression in $CD5^{hi}$ cells and was less accessible in $CD5^{lo}$ cells (Figures 2G and 2H). Consistent with Nur77 expression reflecting tonic signal strength, *Nr4a1* was among the DEGs (Figure 2G), and as previously described (Moran et al., 2011), *Nr4a1* expression was greater in Tregs than in $CD5^{hi}$ naive $CD4^+$ T cells. Like *Cd5*, the *Nr4a1* locus was more open in $CD5^{hi}$ cells (Figure 2H). Together, our findings revealed considerable differences both at the transcriptional and the chromatin level between $CD5^{lo}$ and $CD5^{hi}$ naive $CD4^+$ T cells.

Differences in expression of transcriptional regulators and chromatin modifiers may contribute to functional differences among $CD5^{lo}$ and $CD5^{hi}$ naive $CD4^+$ T cells

We next examined the DEGs identified between $CD5^{lo}$ and $CD5^{hi}$ naive $CD4^+$ T cell populations. More DEGs were upregulated in $CD5^{hi}$ cells and more transcripts were expressed at a FC $\geq$ 2 compared with $CD5^{lo}$ cells (Figure 3A; Table S4). Importantly, we found strong concordance between the bulk and scRNA-seq datasets with regard to specific genes. For instance, *Cd5*, *Folr4*, *Cd6*, *Nr4a1, Tox*, *Ptpn6*, and *Tcf25* were more highly expressed in $CD5^{hi}$ cells, while *Ly6c1* and *Dntt* were higher among $CD5^{lo}$ cells (Figure 3B). Indeed, as was described in CD5-sorted naive $CD8^+$ T cells (Fulton et al., 2015) and in human naive $CD4^+$ T cells (Sood et al., 2021), one of the top DEGs was *Dntt* (encoding the DNA polymerase terminal deoxynucleotidyl transferase [TdT], which inserts non-templated nucleotides during V(D)J

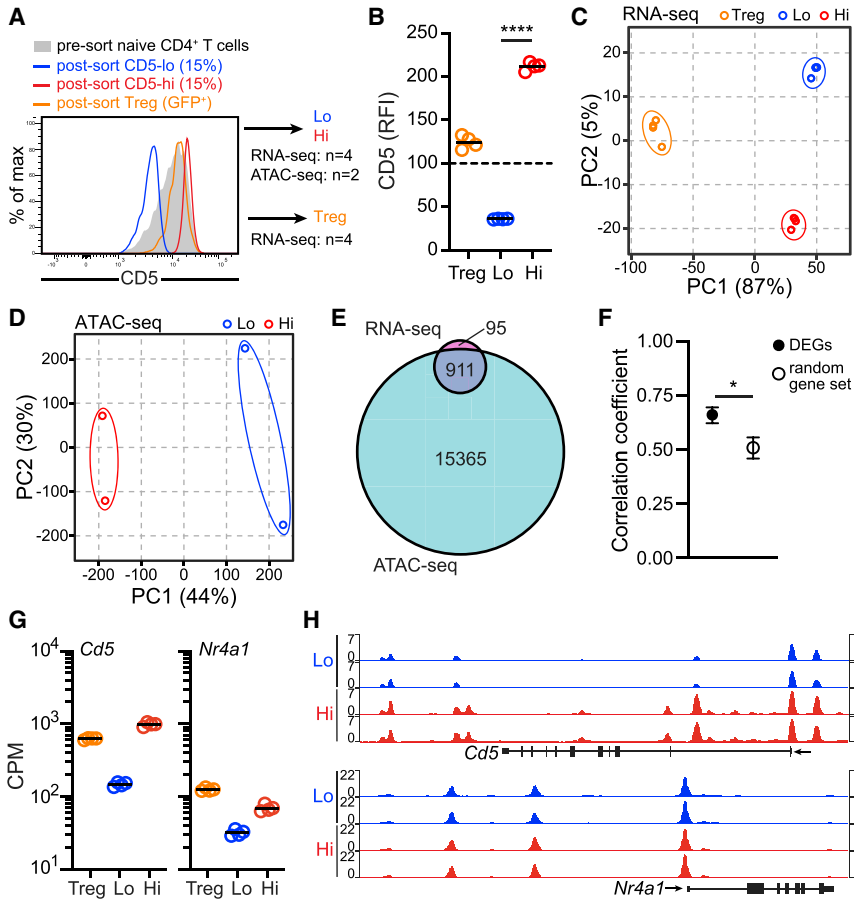


Figure 2. Sorting on CD5 expression reveals that naive CD4$^+$ T cells encompass distinct gene-expression profiles and chromatin landscapes

(A) Sorted CD5^{lo} and CD5^{hi} naive CD4$^+$ T cells, and Tregs for RNA- and ATAC-seq.

(B) Relative fluorescence intensity (RFI) of surface CD5 expression on sorted populations from (A), relative to pre-sorted naive CD4$^+$ T cells. Lines represent group means (n = 4).

(C and D) PCA of populations in (A) for RNA-seq (n = 4) (C) and ATAC-seq (n = 2) (D). Percents in parentheses indicate the variation described by each PC.

(E) Venn diagram illustrating overlap in number of DEGs (FDR $\leq$ 0.01) identified by RNA-seq and DARs identified by ATAC-seq in CD5^{lo} versus CD5^{hi} cells.

(F) Correlation between CD5^{hi} versus CD5^{lo} FC differences in identified DEGs and corresponding DARs as compared with the correlation with a random gene set.

(G) *Cd5* and *Nr4a1* mRNA expression from bulk RNA-seq (counts per million [CPM]). All group comparisons are significant at FDR $\leq$ 0.01. Lines represent group means (n = 4).

(H) ATAC-seq signal profiles across *Cd5* and *Nr4a1* gene loci from two biological replicates of sorted samples.

Statistics: paired t test (B), significant DEGs had FDR $\leq$ 0.01 (E and G), correlation coefficient with 95% confidence intervals (F). *p < 0.05, ****p < 0.0001. See also Figure S2.

TCR rearrangement), at a 14× greater abundance in CD5^{lo} cells. Given the greater *Dntt* expression in CD5^{lo} cells in both CD4$^+$ and CD8$^+$ T cells, we asked whether other DEGs were shared by naive CD8$^+$ T cells sorted on CD5. Interestingly, in line with the narrower CD5 distribution in CD8$^+$ T cells (Mandl et al., 2013), far fewer DEGs with a FC cut-off $\geq$ 2 were identified in CD8$^+$ T cells than in our CD4$^+$ dataset, and there was little overlap (Figure S3A). Among overlapping DEGs, all but 2 of the 24 DEGs showed expression concordance in CD4$^+$ and CD8$^+$ T cells (Figure S3B). Unlike CD8$^+$ T cells, CD5^{hi} naive CD4$^+$ T cells were smaller and expressed less CD44 compared with CD5^{lo} cells (Figure S3C). Overall, there were few parallels between naive CD4$^+$ and CD8$^+$ T cells sorted on CD5 expression with regard to the specific DEGs.

To investigate patterns in DEGs increased among CD5^{lo} or CD5^{hi} naive CD4$^+$ T cells, we performed GO analyses. In CD5^{lo} cells, we found an enrichment for tumor-mediated immunity and regulation of IFN-γ responses (Figure 3C), consistent with work showing that CD5^{lo} CD4$^+$ T cells produce more IFNγ than CD5^{hi} cells upon activation (Sood et al., 2019). In contrast, CD5^{hi} cells were enriched for gene networks involved in leukocyte activation, regulation of signaling, and cell migration (Figure 3C). In line with this, gene set enrichment analysis (GSEA) indicated that genes associated with CD4$^+$ T cell activation and effector responses (Gottschalk et al., 2012; Hale et al., 2013) were overrepresented in CD5^{hi} cells (Figure S3D). Of note, genes involved in the active maintenance of a quiescent state among naive T cells (Chapman et al., 2020; ElTanbouly et al., 2020; Hamilton and Jameson, 2012; Yusuf and Fruman, 2003), were detected in both CD5^{lo} and CD5^{hi} cells, but not significantly different (Tables S4 and S5). Moreover, no cytokines were among the DEGs, except *Il16*, which is constitutively expressed in naive CD4$^+$ T cells (Ren et al., 2005) and was slightly increased in CD5^{lo} cells (Tables S4 and S5). Indeed, the chromatin loci for effector cytokines, such as *Ifng*, *Il5*, *Il17a*, and *Il10*, were closed; only the transcription start site (TSS) for *Il2* and a few peaks surrounding the TSS for *Il21* had marginally greater accessibility in CD5^{hi} cells (Figure S3E). These data indicated that although CD5^{hi} cells were enriched for gene networks associated with activation, both gene expression and accessible chromatin regions were largely consistent with an equally quiescent and non-differentiated cell state among naive CD4$^+$ T cells with different self-reactivities.

Quiescence exit occurs when naive T cells are activated by antigen, after which they undergo chromatin remodeling, transcriptional changes, and ultimately effector differentiation. To understand whether naive CD4$^+$ T cells were poised to respond differently to activation as a function of our identified transcriptome and chromatin accessibility differences, we first investigated whether they differed in expression of

CellPress
OPEN ACCESS

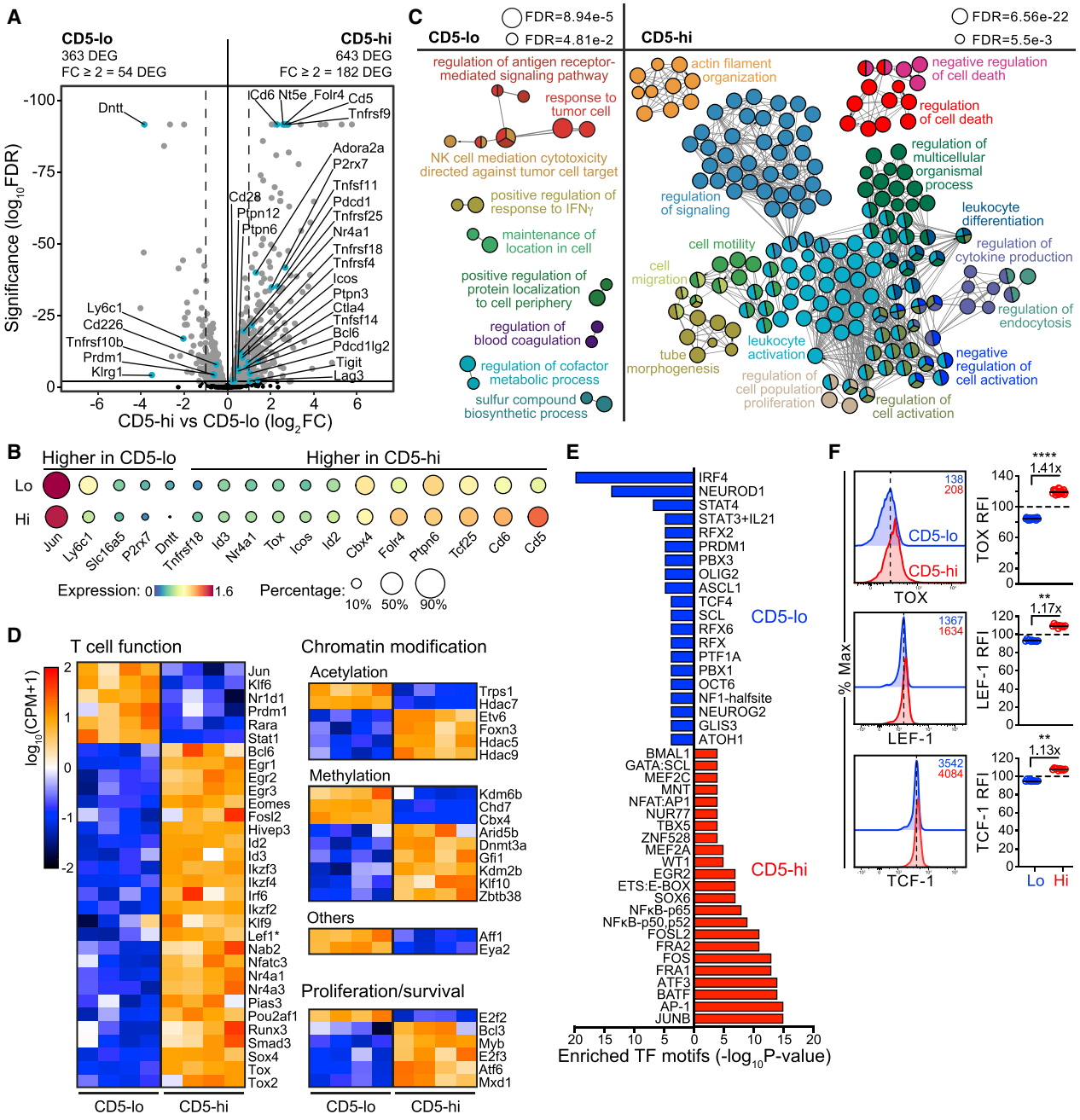


Figure 3. Transcriptional diversity among naive CD4⁺ T cells suggests functional differences through transcriptional regulator activity and chromatin modifiers

(A) Volcano plot of DEGs identified by RNA-seq comparison of $CD5^{lo}$ and $CD5^{hi}$ naive $CD4^+$ T cells. Positive FC values indicate increased expression in $CD5^{hi}$ naive $CD4^+$ T cells. Significant DEGs (FDR ≤ 0.01) are indicated in gray, and a subset of DEGs are labeled (blue). Dotted lines are drawn at FC = 2.

(B) scRNA-seq gene expression level (color) and frequency of cells with detectable expression (circle size) of select DEGs.

(C) GO enrichment analysis for genes upregulated in $CD5^{lo}$ (left) or $CD5^{hi}$ (right) cells. Circles correspond to unique GO groups; related groups are coded in the same color. Circle size reflects enrichment significance (FDR cut-offs shown).

(D) Heatmap of all differentially expressed TRs between $CD5^{lo}$ and $CD5^{hi}$ naive $CD4^+$ T cells (n = 4), grouped by function.

(E) Significant ($p \leq 10^{-4}$) unique enriched TF motifs in DARs from ATAC-seq for $CD5^{lo}$ (blue) and $CD5^{hi}$ (red) cells.

(F) Protein expression of TOX, LEF-1, and TCF-1 in $CD5^{lo}$ and $CD5^{hi}$ cells (RFI is relative to total naive $CD4^+$ T cells). Representative flow cytometry histograms are shown (with mean fluorescence intensity [MFI] in top right). Data are summarized from 2–5 independent experiments. Dotted lines in histograms denote $CD5^{lo}$ modes; data points in graphs represent individual mice (n = 10–20); lines denote group means, and average fold differences are indicated.

Statistics: all TRs had FDR ≤ 0.01 except *Lef1* (*, FDR = 0.012) (D), Wilcoxon matched-pairs signed rank test (F). **p < 0.01, ****p < 0.0001. See also Figure S3; Tables S4 and S5.

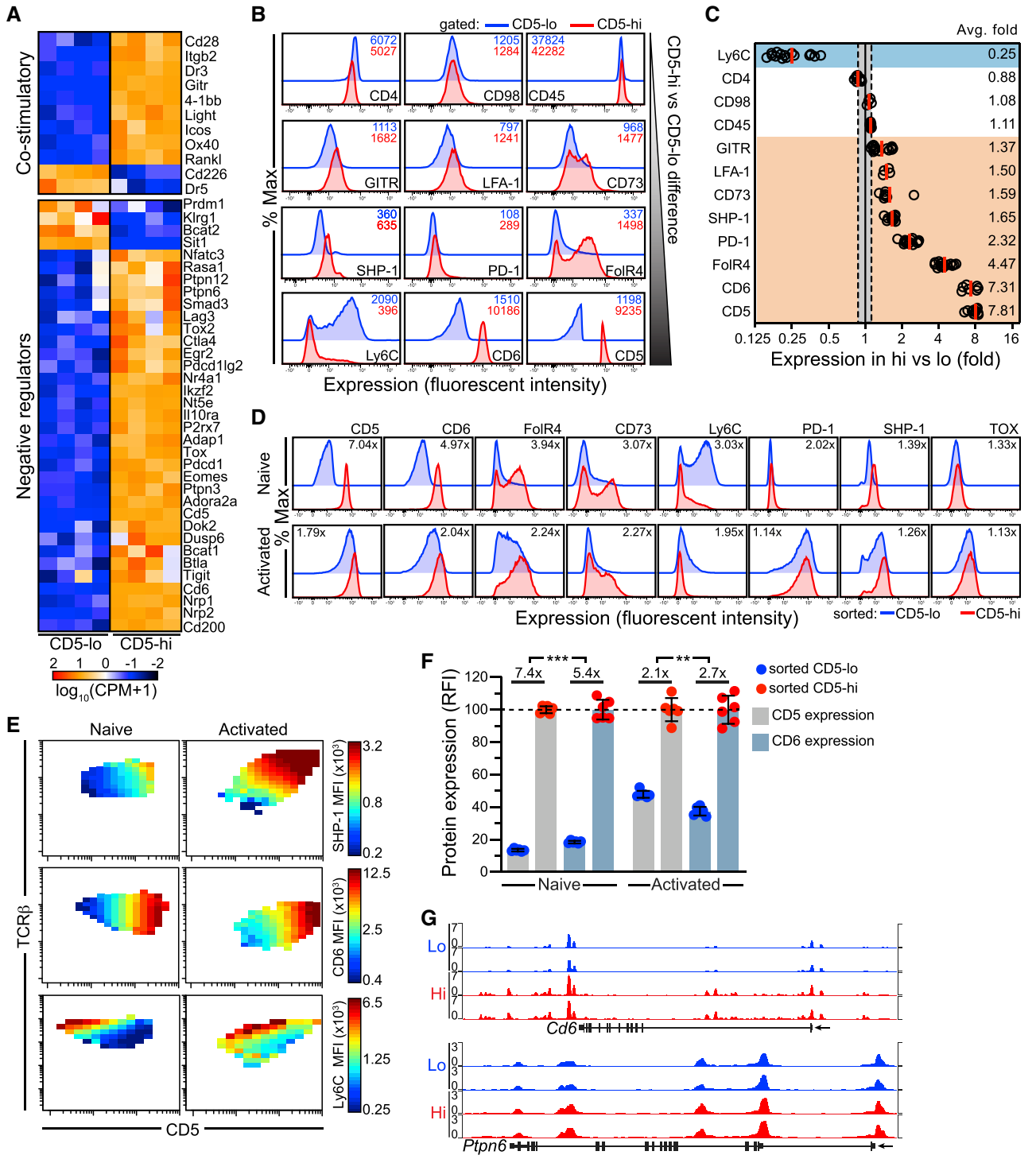


Figure 4. Pre-existing transcriptional and protein differences among naive CD4⁺ T cells are maintained post-activation

(A) Heatmap of curated list of DEGs involved in positive and negative regulation of T cell activation and TCR signaling (n = 4).

(B) Representative histograms of protein expression (measured by flow cytometry) of select DEGs identified from bulk RNA-seq analyses comparing $CD5^{lo}$ and $CD5^{hi}$ naive $CD4^{+}$ T cells. Numbers in top right of histograms indicate MFI in $CD5^{lo}$ (blue) and $CD5^{hi}$ (red) cells.

(C) Summary of fold expression differences of proteins measured in (B) that were significantly different between $CD5^{lo}$ and $CD5^{hi}$ groups from 2–5 independent experiments. Non-significant DEGs (CD4, CD98, and CD45) were used as controls to establish a cut-off for biological significance. Greater expression in $CD5^{lo}$ (blue shading) and greater expression in $CD5^{hi}$ (orange shading). Each data point is from an individual mouse (n = 4–15); red lines denote group means.

(legend continued on next page)

transcriptional regulators (TRs). We used a predefined list of 1,680 known or putative TRs (Mingueneau et al., 2013) and detected 31 TRs upregulated in CD5^{hi} cells involved in T cell proliferation or survival (*Atf6*, *Myb*, and *Bcl3*) and T cell activation or differentiation (*Egr1*, *Egr2*, *Egr3*, *Nfatc3*, *Ikfz3*, *Ikzf4*, *Tox*, *Tox2*, *Nr4a1*, *Nr4a3*, *Klf9*, *Lef1*, *Bcl6*, *Eomes*, *Irf6*, *Id2*, and *Id3*) (Figure 3D). We also detected 21 TRs that mediate chromatin modifications, such as acetylation (*Hdac5*, *Hdac9*, and *Etv6*) and methylation (*Dmnt3a*, *Klf10*, *Kdm2b*, and *Gfi1*) (Figure 3D), as well as other TRs (Figure S3F). Twenty TRs were enriched in CD5^{lo} cells, including repressors (*Hdac7*, *Nr1d1*, *Prdm1*, *Rara*, and *Trps1*) (Figures 3D and S3F).

We next asked whether there were TR binding motifs enriched among chromatin peaks that were unique to either CD5^{lo} or CD5^{hi} cells. Indeed, CD5^{hi} cells were enriched in binding motifs for TR networks downstream of TCR activation, including AP-1 and JNK transcription factors, FOS, FOSL2, AP-1, and JUNB, as well as NUR77, NFAT, and NF-κB (Figures 3E and S3G). Conversely, CD5^{lo} cells were enriched in binding motifs for IRF-4 and PRDM1 (BLIMP-1), both of which promote non-T$_{FH}$ effector differentiation (Johnston et al., 2009; Krishnamoorthy et al., 2017). Notably, although not represented in TR binding motif analysis, TOX and TOX2 were among identified upregulated TRs in CD5^{hi} cells, which function downstream of the TCR through NFAT signaling (Khan et al., 2019; Scott et al., 2019; Seo et al., 2019) and promote T$_{FH}$ differentiation by increasing Bcl6 expression through enhanced chromatin accessibility of TCF-1 and LEF-1 bound regions of the *Bcl6* locus (Xu et al., 2019). Thus, we asked whether TOX, LEF-1, and TCF-1 expression was greater at the protein level in CD5^{hi} cells. Although the detected differences were small (1.3- to 1.4-fold), they were robust across mice (Figure 3F), were not observed in unstained controls (Figure S3H), and corresponded to increased chromatin accessibility in the loci for *Tox*, *Tox2*, *Lef1*, and *Tcf7* in CD5^{hi} cells (Figure S3I).

Together, our data suggested the possibility that a network of TRs and unique chromatin accessibility profiles results in differences in cell states among CD5^{lo} and CD5^{hi} naive CD4^{+} T cells, impacting their function and/or differentiation upon activation, particularly with regard to the early T$_{FH}$ versus non-T$_{FH}$ bifurcation.

Pre-existing expression differences in regulators of TCR signaling among naive CD4^{+} T cells are maintained after activation

To investigate the cell signaling and lymphocyte activation signatures in CD5^{hi} CD4^{+} T cells, we curated a list of DEGs involved in regulating T cell activation. CD5^{hi} cells had increased expression of genes involved in co-stimulation, such as *Icos*, *Rankl*, *Itgb2*, and *Gitr* (Figure 4A). More predominantly, CD5^{hi} cells had a higher expression of genes involved in the negative regulation of TCR signaling or T cell activation, including *Cd6*, *Nt5e* (CD73), *Ptpn6* (SHP-1), *Ctla4*, *Pdcd1*, *Btla*, *IL10ra*, *P2rx7*, *Nrp1*, *Nrp2*, *Cd200*, and *Adora2a* (Figure 4A). Given the role of negative regulators in T cell exhaustion during chronic antigen stimulation, and perhaps indicative of a greater frequency and/or strength of self-pMHC signals obtained by CD5^{hi} T cells, GSEA identified an enrichment in exhaustion-associated genes among CD5^{hi} cells (Figure S4A). Corroborating our RNA-seq, we observed expression differences between CD5^{lo} and CD5^{hi} cells at the protein level for a subset of the regulators of TCR signaling, including GITR, LFA-1, CD6, FolR4, PD-1, and CD73 (Figures 4B and 4C). Of note, FCs in CD6, Ly6C, FolR4, and CD73 between CD5^{lo} and CD5^{hi} cells were detectable in neonates and remained stable with age (Figures S4B and S4C). Interestingly, signaling through LFA-1 has been shown to promote Bcl6 expression and be required for T$_{FH}$ differentiation (Meli et al., 2016), and some of the other genes modulating the TCR signal were also shown to play a role in T$_{FH}$ differentiation, including *Tox*, *Tox2*, *FolR4*, and *Icos*, all of which are expressed at greater levels in CD5^{hi} cells (Figure 4A). In addition, we confirmed greater protein expression in CD5^{hi} cells of SHP-1, a negative regulator of TCR signaling that modulates T cell antigen sensitivity (Feinerman et al., 2008; Stefanová et al., 2003), associates with CD5, CD6 (Blaize et al., 2020; Gonçalves et al., 2018), and other negative regulators of TCR signaling, such as PD-1, BTLA, and CTLA-4 (Lorenz, 2009) (Figures 4B and 4C). Further, CD5^{lo} cells expressed higher levels of Ly6C. The non-significant DEGs CD4, CD98 (LAT1), and CD45 were used as controls to establish a cut-off for biologically significant protein FCs (Figures 4B and 4C). Thus, at least for the subset of tested DEGs, transcriptional differences among naive CD4^{+} T cells correlated with differences in protein expression.

We next asked whether the differential expression of TCR signal regulators was maintained following TCR stimulation. We found that sorted CD5^{lo} and CD5^{hi} naive CD4^{+} T cells stimulated with anti-CD3/CD28, which bypasses individual TCR affinity for their agonist peptides, retained expression differences 24 h after stimulation, albeit for some proteins at reduced FCs (Figures 4D and S4D). In unsorted CD4^{+} T cells, SHP-1 and CD6 expression increased with greater CD5 levels, and this relationship was maintained following activation (Figure 4E). In contrast, Ly6C had a bimodal distribution, with most cells becoming Ly6C negative after activation (Figures 4D and 4E). In addition, the transcription factor TOX, which promotes expression of exhaustion-associated genes (Scott et al., 2019;

(D) Representative histograms of protein expression measured by flow cytometry in sorted 15% CD5^{lo} and CD5^{hi} naive CD4^{+} T cells pre-activation and 24 h post-activation with anti-CD3/CD28. Numbers in top of histograms indicate fold difference between CD5^{lo} (blue) and CD5^{hi} (red) populations.

(E) Single-cell flow cytometry analysis for SHP-1, CD6, and Ly6C expression among naive CD4^{+} T cells pre-activation and 24 h post-activation. Color scale represents MFI of proteins of interest in bins of at least 10 cells across the distribution of CD5 (x axis) and TCRβ (y axis).

(F) Naive CD4^{+} T cells were sorted into CD5^{lo} and CD5^{hi}, and CD5 and CD6 protein expression was measured pre- and post-activation (anti-CD3/CD28). RFI normalized to CD5 or CD6 expression in the CD5^{hi} population. Data are summarized from 2 independent experiments; each data point represents 4–5 pooled mice (n = 6); error bars represent mean ± SD.

(G) ATAC-seq signal profiles across *Ptpn6* and *Cd6* gene loci shown from 2 biological replicates of sorted samples.

Statistics: all genes had FDR ≤ 0.01 (A), paired t test (F). **p < 0.01, ***p < 0.001. See also Figure S4.

Seo et al., 2019), followed a similar expression pattern as SHP-1 across the full spectrum of CD5, but its expression range was reduced post-activation (Figure S4E). Mutual information analysis confirmed that CD5 was a strong predictor of both CD6 and SHP-1 expression pre- and post-activation, whereas TCRβ became predictive only post-activation for SHP-1 and was a poor predictor of CD6 (Figures S4F and S4G).

Cd5 and *Cd6* are gene homologs located on the same chromosome and have functional similarities (Lecomte et al., 1996; Padilla et al., 2000). Our data indicated a tight correlation between CD5 and CD6 levels, with CD5 expression being a robust predictor of CD6 expression both pre- and post-activation (Figures 4E and S4G). Interestingly, the FC difference in CD6 expression on sorted $CD5^{lo}$ and $CD5^{hi}$ $CD4^+$ T cells remained slightly greater post-activation than CD5, suggesting that CD6 might more reliably read out self-pMHC reactivity after activation (Figure 4F). The maintenance of differences among naive $CD4^+$ T cells in the expression of regulators of TCR signaling even following a strong TCR stimulus raised the possibility that these might be regulated through epigenetic modifications rather than modulated entirely by TCR signal strength. Indeed, akin to the differences in chromatin accessibility of the *Cd5* locus (Figure 2H), the *Cd6* and *Ptpn6* loci were more accessible in $CD5^{hi}$ cells (Figure 4G). In summary, we corroborated expression differences between $CD5^{lo}$ and $CD5^{hi}$ naive $CD4^+$ T cells at the protein level for genes important in tuning TCR signal strength and showed that these differences were maintained even after strong TCR stimulation.

$CD5^{hi}$ cells have a greater propensity to develop into T_{FH} cells than $CD5^{lo}$ naive $CD4^+$ T cells

We next investigated whether the observed transcriptional and chromatin accessibility differences alter T cell differentiation. GSEA of our RNA-seq indicated that $CD5^{hi}$ cells were enriched for genes expressed in natural Tregs (Figure S5A). To determine whether $CD5^{hi}$ cells were more likely to differentiate into Tregs *in vivo* than $CD5^{lo}$ cells, we adoptively transferred Treg-depleted $CD4^+$ T cells into $TCR\beta^{-/-}$ mice in a model of inflammatory bowel disease (Martin et al., 2004; Powrie et al., 1993). Consistent with prior *in vitro* Treg polarization experiments (Henderson et al., 2015) and studies of $Ly6C^-$ naive $CD4^+$ T cells (Martin et al., 2013), $TCR\beta^{-/-}$ mice given $CD5^{hi}$ cells subsequently had an increased survival time likely because of increased frequency of Tregs, compared with mice that received $CD5^{lo}$ cells (Figures S5B and S5C).

Interestingly, GSEA also showed that $CD5^{hi}$ cells were enriched for T_{FH} gene signatures (Figure 5A). Across all replicates, $CD5^{hi}$ cells expressed higher levels of *Pdcd1*, *Cxcr5*, and *Bcl6*, while $CD5^{lo}$ cells expressed higher levels of the T_{FH} repressor *Prdm1* (Blimp-1) (Figure 5B). Moreover, the *Pdcd1*, *Cxcr5*, and *Bcl6* loci were more accessible in $CD5^{hi}$ cells (Figure 5C). These data suggested the possibility of a pre-existing disposition among $CD5^{hi}$ cells to become T_{FH} cells relative to their $CD5^{lo}$ counterparts. In support of this hypothesis, $CD5^{hi}$ naive $CD4^+$ T cells were shown to produce more IL-2 than $CD5^{lo}$ cells post-stimulation (Persaud et al., 2014), and data have highlighted the importance of early IL-2 production in T_{FH} lineage choice, with IL-2 producers becoming T_{FH} cells and paracrine IL-2 signaling reinforcing non-T_{FH} lineage commitment of $CD4^+$ T cells obtaining weaker TCR signals (Ballesteros-Tato et al., 2012; DiToro et al., 2018).

To directly assess whether $CD5^{hi}$ naive $CD4^+$ T cells gave rise to a greater proportion of T_{FH} cells *in vivo* than $CD5^{lo}$ cells, we first sorted 15% $CD5^{lo}$ and $CD5^{hi}$ naive $CD4^+$ T cells and adoptively transferred $\sim 1 \times 10^7$ cells of each into separate recipients that were infected with LCMV. An estimated eight LCMV-GP66-specific $CD4^+$ T cells are present per 10^6 naive $CD4^+$ T cells (Jenkins and Moon, 2012; Nelson et al., 2015), and so it was not surprising that the proportion of activated cells postinfection (day 8) was highly stochastic among the transferred cells. However, as in our *in vitro* assays, the CD5 expression difference between transferred $CD5^{lo}$ and $CD5^{hi}$ $CD4^+$ T cells was maintained following activation (Figures 5D and 5E). As an alternative approach, we asked whether we would be better able to assess differences in T_{FH} differentiation potential by transferring sorted populations into $TCR\beta^{-/-}$ mice. Given the lack of competitor T cells in these recipients, we first investigated whether this would impact T_{FH} lineage choice in $CD4^+$ T cells with a fixed TCR. We performed adoptive transfers of LCMV-specific TCR Tg SMARTA $CD4^+$ T cells into both wild-type (WT) and $TCR\beta^{-/-}$ recipients, infected them 1 day later with LCMV, and then assessed the response 8 days postinfection (Figure S5D). Overall, the clonal expansion of SMARTA TCR Tg cells was greater in the $TCR\beta^{-/-}$ mice (Figure S5E). Interestingly, we found that when SMARTA TCR Tg cells were transferred into $TCR\beta^{-/-}$ recipients, CD5 surface expression was increased compared with cells transferred into WT mice (Figure S5F), suggesting that SMARTA Tg cells were receiving stronger TCR signals upon antigen encounter in the absence of other competitor T cells, as has been previously shown in lymphopenia-induced expansion (Vrisekoop et al., 2017). Importantly, the percent of T_{FH} cells among transferred SMARTA TCR Tg cells was halved in the infected $TCR\beta^{-/-}$ compared with WT recipients (Figure S5G). Given the role of IL-2 in T_{FH} differentiation and the modulation of IL-2 production by TCR signal strength (DiToro et al., 2018), we postulated that the decreased T_{FH} differentiation of SMARTA TCR Tg cells in $TCR\beta^{-/-}$ mice was a result of greater IL-2 production mediated by greater TCR signaling. Indeed, we observed that SMARTA TCR Tg cells produced more IL-2 and had greater CD25 (IL-2Rα) surface expression in the infected $TCR\beta^{-/-}$ mice (Figures S5H and S5I). Due to the impact of the lack of other T cells in the $TCR\beta^{-/-}$ mice on T_{FH} frequency post-infection, we concluded that this experimental approach would not accurately identify T_{FH} potential differences between $CD5^{lo}$ and $CD5^{hi}$ polyclonal naive $CD4^+$ T cells.

Instead, based on the observation that CD5 expression remained detectably different after transfer, we quantified T_{FH} differentiation among activated endogenous $CD5^{lo}$ and $CD5^{hi}$ $CD4^+$ T cells post-LCMV infection. We found ~2-fold increased T_{FH} differentiation within the 15% $CD5^{hi}$ activated $CD4^+$ T cells compared with $CD5^{lo}$ cells (Figures 5F, 5G, and S5J). Moreover, among $CD5^{hi}$ cells, T_{FH} ($PD\text{-}1^{hi}$ $CXCR5^+$) and $PD\text{-}1^{hi}$ $CXCR5^-$ cells expressed greater levels of PD-1 than their $CD5^{lo}$ counterparts (Figure S5K). T_{FH} cells were also overrepresented among $CD6^{hi}$ $CD4^+$ T cells and had increased PD-1 expression over their $CD6^{lo}$ counterparts (Figures 5F, 5G, S5J, and S5K). In line with an increased T_{FH} population from $CD5^{hi}$ cells, these T_{FH} cells

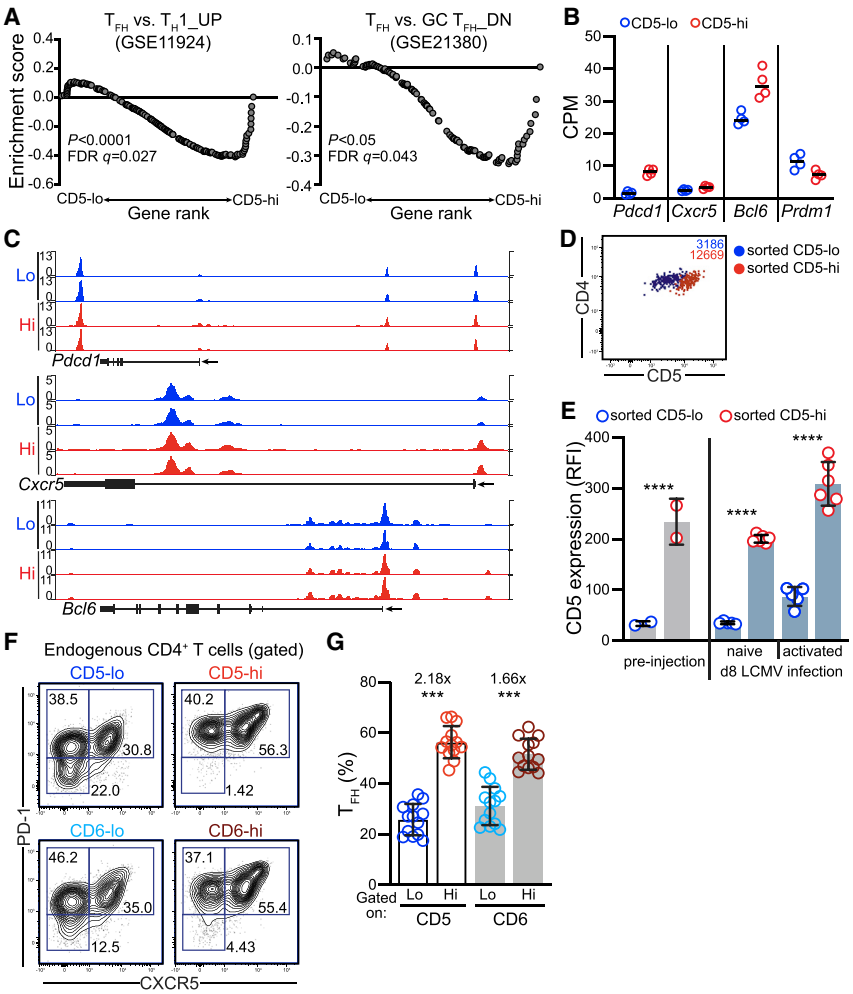


Figure 5. CD5^{hi} naive CD4^{+} T cells are enriched for T_{FH}-associated genes and have a greater T_{FH} differentiation potential upon infection than CD5^{lo} cells

(A) GSEA of T_{FH} signatures enriched in CD5^{hi} naive CD4^{+} T cells.

(B) *Pdcd1*, *Cxcr5*, *Bcl6*, and *Prdm1* mRNA expression from bulk RNA-seq for sorted CD5^{lo} and CD5^{hi} naive CD4^{+} T cell populations. Lines indicate group means (n = 4).

(C) ATAC-seq signal profiles across *Pdcd1*, *Cxcr5*, and *Bcl6* gene loci from 2 biological replicates of sorted CD5^{lo} and CD5^{hi} naive CD4^{+} T cell samples.

(D and E) 15% CD5^{lo} and CD5^{hi} naive CD4^{+} T cells were sorted and adoptively transferred into recipients and infected 1 day later with LCMV. Representative flow cytometry plot of activated (CD44^{hi}) transferred cells with numbers in top right indicating CD5 MFI of sorted CD5^{lo} (blue) and CD5^{hi} (red) cells (D), and summary of CD5 RFI (relative to endogenous total naive CD4^{+} T cells) of sorted CD5^{lo} and CD5^{hi} naive or activated CD4^{+} T cells (E). Data are summarized from 2 independent experiments (D and E); each data point is from an individual recipient mouse (n = 2–6); error bars represent mean ± SD.

(F and G) Activated (CD44^{hi}) CD4^{+} T cells isolated at day 8 after LCMV infection were gated on the top and bottom 15% CD5- or CD6-expressing cells and the percent T_{FH} was determined. Representative flow cytometry plots are shown, with numbers indicating percent cells within each gate (F); data are summarized from 2 independent experiments; each data point is from an individual mouse (n = 13); error bars represent mean ± SD (G).

Statistics: all genes had FDR ≤ 0.01 except *Cxcr5* (FDR = 0.044) (B), one-way ANOVA with Tukey's multiple comparison (E), or Wilcoxon matched-pairs signed rank test (G). ***p < 0.001, ****p < 0.0001. See also Figure S5.

expressed higher TOX, LEF-1, and TCF-1 (Figure S5L). Thus, our data suggested that T_{FH} and non-T_{FH} cell fate decision was altered by pre-existing differences present in naive CD4^{+} T cells prior to foreign antigen encounter.

Removing naive CD4^{+} T cells from continuous self-pMHC interactions reveals gene expression differences that are independent of post-thymic self-ligand recognition

Continuous tonic self-pMHC signals obtained by naive T cells in the periphery facilitate antigen recognition by maintaining partial TCRζ-chain phosphorylation and through polarization of the TCR and its signaling components (Stefanová et al., 2002). Interrupting signals from self-pMHC interactions for only 30 min leads to a loss of antigen sensitivity (Stefanová et al., 2002). It is unknown whether described differences between naive CD4^{+} T cells of low and high self-reactivity are similarly dependent on subthreshold tonic self-signals. An alternative hypothesis is that naive T cells are pre-wired by TCR signals obtained during selection in the thymus, and that these differences are then epigenetically imprinted. Given our findings that some modulators of TCR signal strength remained distinct post-activation, we examined which DEGs between CD5^{lo} and CD5^{hi} cells were dependent on continuous self-pMHC interactions and which might be a result of thymic imprinting.

To investigate this, we sorted 15% CD5^{lo} and CD5^{hi} naive CD4^{+} T cells and cultured them *ex vivo* in the absence of self-pMHC for 22 h with IL-7 and performed bulk RNA-seq. After resting, sorted cells segregated into clusters distinct from their freshly isolated counterparts along PC1, while differences between CD5^{lo} and CD5^{hi} cells were preserved in PC2 (Figure 6A). As described previously, levels of CD5 protein and transcript rapidly decreased upon resting (Figure 6B) (Mandl et al., 2012; Smith et al., 2001). Interestingly, however, CD5 mRNA and protein expression remained distinct in rested CD5^{lo} compared with CD5^{hi} cells (Figure 6C). Therefore, although CD5 expression levels are maintained by peripheral self-interactions, our data suggested that the retention of CD5 expression differences between CD5^{lo} and CD5^{hi} cells were independent of tonic self-signals.

To examine whether other genes in our identified CD5^{lo} versus CD5^{hi} DEGs followed a similar pattern to *Cd5*, we compared expression levels in fresh and rested cells and designed classification criteria to subset genes into two groups: DEG-ND, genes where expression differences between CD5-sorted populations

were not dependent on tonic self-pMHC interactions; and DEG-D, genes where differences were lost upon resting (i.e., expression differences were dependent on tonic self-pMHC interactions) (Figure 6D). A gene was classified as being DEG-ND if it was a DEG (FDR $\leq$ 0.01) in the fresh comparison and remained a DEG (FDR $\leq$ 0.01) after resting. Conversely, to be classified as DEG-D, the gene was a DEG (FDR $\leq$ 0.01) in the fresh comparison but became non-significant (FDR $\geq$ 0.3) after resting. With these criteria, 513 DEGs were DEG-ND and 212 DEG-D, while 281 DEGs did not fall into either group (Figure 6D; Tables S4 and S5). Plotting the $CD5^{lo}$ versus $CD5^{hi}$ FC of the fresh versus the rested confirmed that DEG-ND were unaffected by the absence of self-pMHC, while DEG-D converged to zero after resting (Figure 6E). Thus, transcriptional heterogeneity among the naive $CD4^+$ T cell population was explained by both differences that depended on continuous self-pMHC interactions and that were independent of tonic self-pMHC signals. Interestingly, most of the negative regulators of TCR signaling and *Dntt* were DEG-ND, consistent with their expression level being set during thymic development (Figure 6E). In addition, more than half of the T_{FH}-associated genes were DEG-ND, including *Icos*, *Itgb2*, *P2rx7*, and *Adora2a*, whereas 20% of T_{FH} associated genes were DEG-D, including *Tox2* and *Dusp6*. Of note, the TRs *Jun*, *Egr1*, *Egr3*, *Ikzf3*, and *Nfatc3* were classified as DEG-D, while *Stat1*, *Eomes*, *Ikzf2*, *Id2*, *Sox4*, and *Nr4a1* were DEG-ND, indicating that even at the level of regulation of TRs some expression differences rely on continuous self-ligand interactions, whereas others do not (Figure 6E). To verify whether the maintenance of differences among transcripts following resting was also observed at the protein level, we chose a subset involved in TCR signaling (CD6, CD73, FolR4, Ly6C, PD-1, and SHP-1) from the DEG-ND group and measured expression in sorted $CD5^{lo}$ versus $CD5^{hi}$ cells rested for 5 days to give protein levels time to turn over. We observed no loss in cell viability during this period (Figure S6A), and although protein levels did change upon resting, the differences between $CD5^{lo}$ and $CD5^{hi}$ cells were retained (Figures 6F and S6B).

Our classification of genes into two distinct groups based on their reliance on tonic self-pMHC interactions suggested that at least a subset of the described transcriptionally wired heterogeneity among naive $CD4^+$ T cells may reflect epigenetic differences established in the thymus. To test this hypothesis, we probed our ATAC-seq dataset to ask whether there were detectable differences in chromatin states between the genes in the two groups. We predicted that loci of DEG-ND genes would show a greater difference in accessible regions between $CD5^{lo}$ and $CD5^{hi}$ cells than would the loci of DEG-D genes. Indeed, on average, the fold difference in chromatin accessibility was greater in the DEG-ND compared with the DEG-D group (Figures 6G and 6H). In line with this, gene expression differences between $CD5^{lo}$ and $CD5^{hi}$ cells were greater in the DEG-ND set than the DEG-D set (Figure 6I). Collectively, these data suggested that there were two sources of heterogeneity among naive $CD4^+$ T cells that impacted their responsiveness to foreign antigen and their effector differentiation: differences that required continuous self-pMHC interactions in the periphery, and differences that did not require self-pMHC interactions that may be imprinted in the thymus.

DISCUSSION

$CD4^+$ T cells play a critical role in orchestrating an immune response, with early fate decisions between T_{FH} and non-T_{FH} effector subset lineage decisions thought to be primarily determined by TCR engagement with cognate pMHC during priming (Ruterbusch et al., 2020). Here, we showed that there are transcriptional and open-chromatin differences between $CD4^+$ T cells that are present prior to their activation, maintained after activation in the short term, and impact TCR signal strength and early lineage choice upon antigen encounter. At the single-cell level, our data highlight that heterogeneity among naive $CD4^+$ T cells is mediated by many interacting genes. Importantly, although we show that CD5 expression alone does not establish discrete clusters among naive $CD4^+$ T cells, other work has defined subpopulations that change with infectious challenge (ElTanbouly et al., 2020; Meli et al., 2020). Here, we identified expression in modulators of TCR signaling, chromatin modifiers, and steady-state T cell trafficking genes as key drivers of between-cell variability. Our data suggest that CD5 expression is a better predictor of cellular behavior at the population level than at the single-cell level. Population averages of sorted polyclonal naive $CD4^+$ T cells are therefore not always mirrored when studying the behavior of a few specific TCR clonotypes, explaining some of the contradictory results with regard to T_{FH}-lineage differentiation biases described based on specific pairs of TCR Tg clones (Bartleson et al., 2020; Persaud et al., 2014).

Our findings expand on previous work implicating self-pMHC reactivity in the effector potential of $CD4^+$ T cells (Henderson et al., 2015; Martin et al., 2013; Sood et al., 2019). Our work implicates self-reactivity in early T_{FH} cell lineage bifurcation. Previous work showed that T cells obtaining stronger TCR signals during activation become IL-2-producing cells and signal in *trans* to IL-2 non-producing cells to reinforce non-T_{FH} effector differentiation (Ballesteros-Tato et al., 2012; DiToro et al., 2018). We found that increasing the strength of self-pMHC signals obtained by SMARTA TCR Tg $CD4^+$ T cells transferred into T cell-deficient mice led to a decrease in T_{FH} differentiation as a result of enhanced IL-2 production. Thus, removing competitor T cells, by modulating both IL-2 and TCR signal strength, led to the opposite outcome with regard to T_{FH} differentiation predicted by TCR signal strength alone, in line with recent observations that $Nur77^{lo}$ $CD4^+$ T cells adoptively transferred into $TCR\alpha^{-/-}$ gave rise to a greater T_{FH} frequency than transferred $Nur77^{hi}$ cells (Bartleson et al., 2020). A recent study has corroborated the use of CD5 as a marker for the self-ligand reactivity in human T cells (Sood et al., 2021), but it will be important to investigate whether the relationships described between self-reactivity and effector potential hold.

Although we implicate thymically imprinted epigenetic differences in the transcriptional heterogeneity among naive $CD4^+$ T cells, we did not address here whether such signals are impacted at different times during development. It is increasingly appreciated that T cell development is a layered process, with neonatally derived T cells being distinct from adult-derived T cells with regard to responsiveness to stimulation and effector potential (Rudd, 2020). Moreover, in both mice and humans, CD5 expression on naive $CD4^+$ T cells is lower in adults than in neonates (Dong et al., 2017; Mandl et al., 2013). Although we

CellPress
OPEN ACCESS

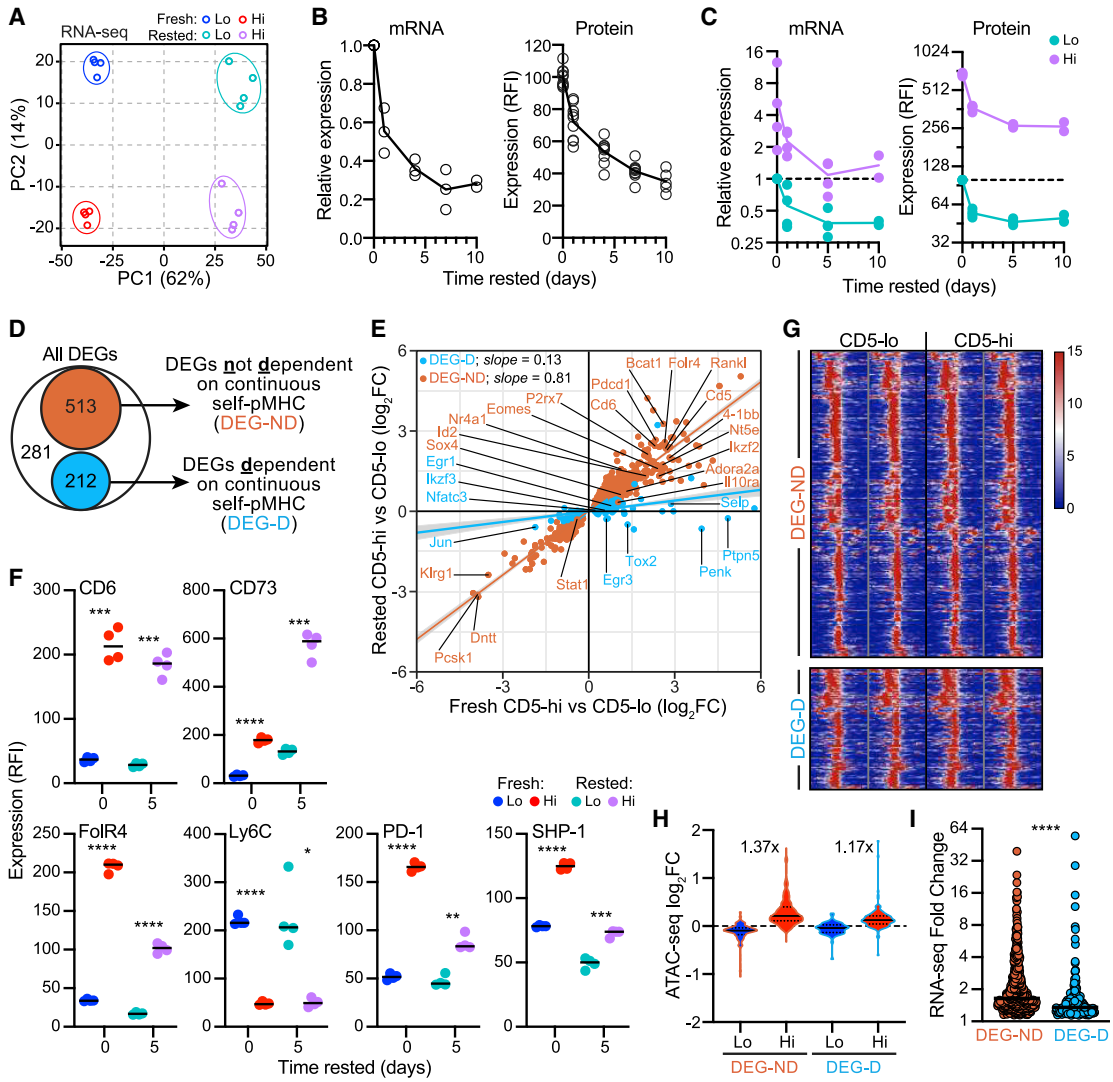


Figure 6. Withdrawal of naive CD4$^+$ T cells from self-pMHC identifies transcriptional and chromatin differences between CD5^{lo} and CD5^{hi} cells that do not rely on continuous self-pMHC interactions

(A) PCA of RNA-seq data from fresh and rested CD5^{lo} and CD5^{hi} naive CD4$^+$ T cells (n = 4).

(B) CD5 mRNA (normalized to *Gapdh*) and protein expression, relative to day 0, measured for naive CD4$^+$ T cells rested in culture in the presence of IL-7. Data are summarized from 2 independent experiments; each data point is from an individual mouse (n = 3–9).

(C) CD5 mRNA (normalized to *Tbp*) and protein expression determined after resting as in (B) of sorted 15% CD5^{lo} and CD5^{hi} naive CD4$^+$ T cells. Data are from 1 experiment; each data point represents 9–10 pooled mice. Days 0, 1, and 5 (n = 4) and day 10 (n = 2). There is a significant effect between CD5^{lo} and CD5^{hi} in both mRNA ($p < 0.05$) and protein ($p < 0.0001$) groups post-resting.

(D) Venn diagram dividing DEGs identified by RNA-seq into two groups: DEG-ND (orange) were differentially expressed (FDR $\leq$ 0.01) in the fresh CD5^{hi} versus CD5^{lo} comparison and remained differentially expressed (FDR $\leq$ 0.01) in the rested CD5^{hi} versus CD5^{lo} comparison; DEG-D (blue) were differentially expressed (FDR $\leq$ 0.01) in the fresh CD5^{hi} versus CD5^{lo} comparison but were not differentially expressed (FDR $\geq$ 0.3) in the rested comparison.

(E) FC expression of DEG-ND and DEG-D identified in (D), in both the fresh and rested samples from bulk RNA-seq. Lines indicate best fits; slopes are shown.

(F) Naive CD4$^+$ T cells were sorted into CD5^{lo} and CD5^{hi} cells and rested in culture in the presence of IL-7. RFI (normalized to total naive CD4$^+$ T cells) from genes identified from (D). Data are from 1 experiment; each data point represents 3 mice pooled (n = 4); lines denote group means.

(G) ATAC-seq chromatin accessibility heatmap for open chromatin regions among DEG-ND and DEG-D groups.

(H) Summary of data in (G) with fold differences between CD5^{lo} and CD5^{hi} open chromatin peaks indicated on the graph.

(I) Gene expression FCs from RNA-seq between CD5^{lo} and CD5^{hi} cells in DEG-ND and DEG-D groups identified in (D). Lines denote group means.

Statistics: two-way ANOVA with Sidak's multiple comparison (C), paired t test (F), or Mann-Whitney U test (I). *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$. See also Figure S6; Tables S4 and S5.

showed that differences between CD5^{lo} and CD5^{hi} naive CD4^{+} T cells remained stable between neonates and older mice, it will be important to determine whether T cell developmental origins (fetal versus adult) are a contributing factor in their diversity.

Whether differences among naive CD4^{+} T cells can ultimately be related back to features of their specific TCRs remains an open question. It is intriguing that in CD5^{lo} naive CD4^{+} T cells, one of the top DEGs was *Dntt* (encoding TdT), as was also observed in mouse naive CD8^{+} T cells and human naive CD4^{+} T cells (Fulton et al., 2015; Sood et al., 2021). TdT is responsible for adding n-nucleotides during V(D)J recombination and thus diversifying the TCR repertoire (Benedict et al., 2000; Cabaniols et al., 2001). It has been proposed that CD5^{hi} T cells have a greater proportion of germline TCRs (lacking n-nucleotides) and were evolutionarily optimized to strongly bind to pMHC (Vrisekoop et al., 2014). It is possible that differences in *Dntt* expression play a role in dictating TCR sequence length and self-pMHC reactivity. Ultimately, patterns in TCR sequences may exist that enable some prediction of self-pMHC reactivity and, therefore, the differentiation potential of individual T cell clones.

Together, our data shed light on which pre-existing transcriptome-level differences among naive CD4^{+} T cells are accessible to interventions targeting self-pMHC peripheral interactions, compared with others that would require modulation at the chromatin level, which may aid in optimizing protocols for enhancing desirable T cell functions in clinical settings, such as in adoptive cell therapies (Alspach et al., 2019; Borst et al., 2018).

Limitations of the study

Our data imply that naive CD4^{+} T cell heterogeneity is partly thymically imprinted and retained independent of interactions with self-pMHC in the periphery. Thus, our work reconciles prior studies that have described specific heterogeneous traits among naive CD4^{+} T cells, not all of which were dependent on tonic TCR signals (Mandl et al., 2013; Matson et al., 2020; Persaud et al., 2014; Stefanová et al., 2002). One limitation of our work is that not all DEGs identified could be classified into DEG-ND or DEG-D groups and, given the lack of commercially available antibodies for all genes of interest, only some of our classified genes were validated at the protein level. Further, future work will be needed to more clearly define the link between expression differences in chromatin modifiers, epigenetic changes in individual T cells, and the interactions with self-pMHC made in the thymus.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Mice
  - Infections
- METHOD DETAILS
  - Lymphocyte isolation, resting, activation, and restimulation
  - Flow cytometry
  - Cell sorts
  - Adoptive cell transfers
  - Bulk RNA sequencing
  - Single cell RNA sequencing
  - Single-cell RNA-seq data analysis
  - ATACseq library preparation, sequencing, and visualization
  - RNA extraction and quantitative real-time PCR
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Heatmaps
  - Gene set enrichment analysis
  - Gene ontology pathway analysis
  - PCA
  - Mutual Information
  - ScatterSlice analysis
  - Statistical analyses

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2021.110064.

ACKNOWLEDGMENTS

We would like to thank G. Perreault, P. D'Arcy, and the animal facility staff at McGill for their excellent care of our animal colony. We are grateful to R. Germain (NIH), T. Baldwin (University of Alberta), N. Singh (University of Maryland), N. Vrisekoop (Utrecht University), and Mandl lab members for critical feedback and comments on the manuscript. We thank C. Stegen and J. Leconte at the Cell Vision Core Facility and the core facility at the Institut de Recherches Cliniques de Montréal. We thank S. Lesage (Université de Montréal) for sharing RAG2^{GFP} mice. Computational analyses were enabled in part by support provided by Calcul Québec and Compute Canada. D.R. was supported by a Frederick Banting and Charles Best Canada Graduate Doctoral Award (CIHR) and a Tomlinson Doctoral Fellowship (McGill). A.S. was funded by a Cole Foundation Postdoctoral Fellowship. J.J.P.v.B. was supported by a NOW Vici grant (016.140.655). D.L. and H.J.M. are junior 1 and junior 2 scholars of the FRQS, respectively. J.N.M. is a Canada Research Chair for Immune Cell Dynamics. This research was supported by a NSERC Discovery Grant (2016-03808) and McGill start-up fund to J.N.M., and a Cancer Research Society (22344) and CIHR (PJT-168862) grant jointly to J.N.M. and H.J.M.

AUTHOR CONTRIBUTIONS

J.N.M. conceived and supervised the project and research, with input from H.J.M. and J.T. D.R. designed and performed most experiments with input from C. Schneider, C Shen, D.C.W., P.A., and A.B. Bulk RNA-seq was performed by J.N.M., with help from A.J.M. and J.S.T., and D.R. analyzed the data. D.R., A.S., and M.-E.L. generated the ATAC-seq dataset with input from H.J.M., and data were analyzed by D.R. and H.W. The scRNA-seq data were generated and analyzed by J.J.P.v.B. and J.T. with input from D.R. Mutual information was determined by T.J.R. and P.F. Critical reagents or intellectual input were provided by S.A.C., M.J.R., L.B.B., P.F., and D.L. The manuscript was written by J.N.M. and D.R., with critical input from H.J.M and J.T. and feedback from all authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

We worked to ensure sex balance in the selection of non-human subjects. One or more of the authors of this paper self-identifies as an underrepresented ethnic minority in science. One or more of the authors of this paper self-identifies as a member of the LGBTQ+ community. While citing references scientifically relevant for this work, we also actively worked to promote gender balance in our reference list.

Received: July 6, 2021
Revised: September 26, 2021
Accepted: November 5, 2021
Published: November 30, 2021

REFERENCES

Almeida, L., Lochner, M., Berod, L., and Sparwasser, T. (2016). Metabolic pathways in T cell activation and lineage differentiation. Semin. Immunol. *28*, 514–524.

Alspach, E., Lussier, D.M., Miceli, A.P., Kizhvatov, I., DuPage, M., Luoma, A.M., Meng, W., Lichti, C.F., Esaulova, E., Vomund, A.N., et al. (2019). MHC-II neoantigens shape tumour immunity and response to immunotherapy. Nature *574*, 696–701.

Altschuler, S.J., and Wu, L.F. (2010). Cellular heterogeneity: do differences make a difference? Cell *141*, 559–563.

Arman, M., Calvo, J., Trojanowska, M.E., Cockerill, P.N., Santana, M., López-Cabrera, M., Vives, J., and Lozano, F. (2004). Transcriptional regulation of human CD5: important role of Ets transcription factors in CD5 expression in T cells. J. Immunol. *172*, 7519–7529.

Artegiani, B., Lyubimova, A., Muraro, M., van Es, J.H., van Oudenaarden, A., and Clevers, H. (2017). A Single-Cell RNA Sequencing Study Reveals Cellular and Molecular Dynamics of the Hippocampal Neurogenic Niche. Cell Rep. *21*, 3271–3284.

Azzam, H.S., Grinberg, A., Lui, K., Shen, H., Shores, E.W., and Love, P.E. (1998). CD5 expression is developmentally regulated by T cell receptor (TCR) signals and TCR avidity. J. Exp. Med. *188*, 2301–2311.

Ballesteros-Tato, A., León, B., Graf, B.A., Moquin, A., Adams, P.S., Lund, F.E., and Randall, T.D. (2012). Interleukin-2 inhibits germinal center formation by limiting T follicular helper cell differentiation. Immunity *36*, 847–856.

Bartleson, J.M., Viehmann Milam, A.A., Donermeyer, D.L., Horvath, S., Xia, Y., Egawa, T., and Allen, P.M. (2020). Strength of tonic T cell receptor signaling instructs T follicular helper cell-fate decisions. Nat. Immunol. *21*, 1384–1396.

Becattini, S., Latorre, D., Mele, F., Foglierini, M., De Gregorio, C., Cassotta, A., Fernandez, B., Kelderman, S., Schumacher, T.N., Corti, D., et al. (2015). T cell immunity. Functional heterogeneity of human memory $CD4^+$ T cell clones primed by pathogens or vaccines. Science *347*, 400–406.

Benedict, C.L., Gilfillan, S., Thai, T.H., and Kearney, J.F. (2000). Terminal deoxynucleotidyl transferase and repertoire development. Immunol. Rev. *175*, 150–157.

Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., Fridman, W.H., Pagès, F., Trajanoski, Z., and Galon, J. (2009). ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics *25*, 1091–1093.

Blaize, G., Daniels-Treffandier, H., Aloulou, M., Rouquié, N., Yang, C., Marcellin, M., Gador, M., Benamar, M., Ducatez, M., Song, K.D., et al. (2020). CD5 signalosome coordinates antagonist TCR signals to control the generation of Treg cells induced by foreign antigens. Proc. Natl. Acad. Sci. USA *117*, 12969–12979.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.

Bolstad, B. (2020). preprocessCore: A collection of pre-processing functions. R package version 1.50.0. https://github.com/bmbolstad/preprocessCore.

Borst, J., Ahrends, T., Bąbała, N., Melief, C.J.M., and Kastenmüller, W. (2018). $CD4^+$ T cell help in cancer immunology and immunotherapy. Nat. Rev. Immunol. *18*, 635–647.

Cabaniols, J.P., Fazilleau, N., Casrouge, A., Kourilsky, P., and Kanellopoulos, J.M. (2001). Most alpha/beta T cell receptor diversity is due to terminal deoxynucleotidyl transferase. J. Exp. Med. *194*, 1385–1390.

Carter, B., and Zhao, K. (2021). The epigenetic basis of cellular heterogeneity. Nat. Rev. Genet. *22*, 235–250.

Chan, T.E., Stumpf, M.P.H., and Babtie, A.C. (2017). Gene Regulatory Network Inference from Single-Cell Data Using Multivariate Information Measures. Cell Syst. *5*, 251–267.e3.

Chang, H.H., Hemberg, M., Barahona, M., Ingber, D.E., and Huang, S. (2008). Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. Nature *453*, 544–547.

Chapman, N.M., Boothby, M.R., and Chi, H. (2020). Metabolic coordination of T cell quiescence and activation. Nat. Rev. Immunol. *20*, 55–70.

Cho, Y.L., Flossdorf, M., Kretschmer, L., Höfer, T., Busch, D.H., and Buchholz, V.R. (2017). TCR Signal Quality Modulates Fate Decisions of Single $CD4^+$ T Cells in a Probabilistic Manner. Cell Rep. *20*, 806–818.

Cotari, J.W., Voisinne, G., Dar, O.E., Karabacak, V., and Altan-Bonnet, G. (2013). Cell-to-cell variability analysis dissects the plasticity of signaling of common γ chain cytokines in T cells. Sci. Signal. *6*, ra17.

Crotty, S. (2019). T Follicular Helper Cell Biology: A Decade of Discovery and Diseases. Immunity *50*, 1132–1148.

DiToro, D., Winstead, C.J., Pham, D., Witte, S., Andargachew, R., Singer, J.R., Wilson, C.G., Zindl, C.L., Luther, R.J., Silberger, D.J., et al. (2018). Differential IL-2 expression defines developmental fates of follicular versus nonfollicular helper T cells. Science *361*, eaao2933.

Dong, M., Artusa, P., Kelly, S.A., Fournier, M., Baldwin, T.A., Mandl, J.N., and Melichar, H.J. (2017). Alterations in the Thymic Selection Threshold Skew the Self-Reactivity of the TCR Repertoire in Neonates. J. Immunol. *199*, 965–973.

ElTanbouly, M.A., Zhao, Y., Nowak, E., Li, J., Schaafsma, E., Le Mercier, I., Ceeraz, S., Lines, J.L., Peng, C., Carriere, C., et al. (2020). VISTA is a checkpoint regulator for naïve T cell quiescence and peripheral tolerance. Science *367*, eaay0524.

Feinerman, O., Veiga, J., Dorfman, J.R., Germain, R.N., and Altan-Bonnet, G. (2008). Variability and robustness in T cell activation from regulated heterogeneity in protein levels. Science *321*, 1081–1084.

Fulton, R.B., Hamilton, S.E., Xing, Y., Best, J.A., Goldrath, A.W., Hogquist, K.A., and Jameson, S.C. (2015). The TCR's sensitivity to self peptide-MHC dictates the ability of naive CD8(+) T cells to respond to foreign antigens. Nat. Immunol. *16*, 107–117.

Gonçalves, C.M., Henriques, S.N., Santos, R.F., and Carmo, A.M. (2018). CD6, a Rheostat-Type Signalosome That Tunes T Cell Activation. Front. Immunol. *9*, 2994.

Gottschalk, R.A., Hathorn, M.M., Beuneu, H., Corse, E., Dustin, M.L., Altan-Bonnet, G., and Allison, J.P. (2012). Distinct influences of peptide-MHC quality and quantity on in vivo T-cell responses. Proc. Natl. Acad. Sci. USA *109*, 881–886.

Grün, D., Kester, L., and van Oudenaarden, A. (2014). Validation of noise models for single-cell transcriptomics. Nat. Methods *11*, 637–640.

Guichard, V., Bonilla, N., Durand, A., Audemard-Verger, A., Guilbert, T., Martin, B., Lucas, B., and Auffray, C. (2017). Calcium-mediated shaping of naive CD4 T-cell phenotype and function. eLife *6*, e27215.

Hale, J.S., Youngblood, B., Latner, D.R., Mohammed, A.U., Ye, L., Akondy, R.S., Wu, T., Iyer, S.S., and Ahmed, R. (2013). Distinct memory CD4+ T cells with commitment to T follicular helper- and T helper 1-cell lineages are generated after acute viral infection. Immunity *38*, 805–817.

Hamilton, S.E., and Jameson, S.C. (2012). CD8 T cell quiescence revisited. Trends Immunol. *33*, 224–230.

Hashimshony, T., Senderovich, N., Avital, G., Klochendler, A., de Leeuw, Y., Anavy, L., Gennert, D., Li, S., Livak, K.J., Rozenblatt-Rosen, O., et al. (2016).

CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. Genome Biol. *17*, 77.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589.

Henderson, J.G., Opejin, A., Jones, A., Gross, C., and Hawiger, D. (2015). CD5 instructs extrathymic regulatory T cell development in response to self and tolerizing antigens. Immunity *42*, 471–483.

Howe, E.A., Sinha, R., Schlauch, D., and Quackenbush, J. (2011). RNA-Seq analysis in MeV. Bioinformatics *27*, 3209–3210.

Jenkins, M.K., and Moon, J.J. (2012). The role of naive T cell precursor frequency and recruitment in dictating immune response magnitude. J. Immunol. *188*, 4135–4140.

Johnston, R.J., Poholek, A.C., DiToro, D., Yusuf, I., Eto, D., Barnett, B., Dent, A.L., Craft, J., and Crotty, S. (2009). Bcl6 and Blimp-1 are reciprocal and antagonistic regulators of T follicular helper cell differentiation. Science *325*, 1006–1010.

Khan, O., Giles, J.R., McDonald, S., Manne, S., Ngiow, S.F., Patel, K.P., Werner, M.T., Huang, A.C., Alexander, K.A., Wu, J.E., et al. (2019). TOX transcriptionally and epigenetically programs $CD8^+$ T cell exhaustion. Nature *571*, 211–218.

Khatun, A., Kasmani, M.Y., Zander, R., Schauder, D.M., Snook, J.P., Shen, J., Wu, X., Burns, R., Chen, Y.G., Lin, C.W., et al. (2021). Single-cell lineage mapping of a diverse virus-specific naive CD4 T cell repertoire. J. Exp. Med. *218*, e20200650.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. *14*, R36.

Krishnamoorthy, V., Kannanganat, S., Maienschein-Cline, M., Cook, S.L., Chen, J., Bahroos, N., Sievert, E., Corse, E., Chong, A., and Sciammas, R. (2017). The IRF4 Gene Regulatory Module Functions as a Read-Write Integrator to Dynamically Coordinate T Helper Cell Fate. Immunity *47*, 481–497.e7.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25.

Lecomte, O., Bock, J.B., Birren, B.W., Vollrath, D., and Parnes, J.R. (1996). Molecular linkage of the mouse CD5 and CD6 genes. Immunogenetics *44*, 385–390.

Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics *26*, 589–595.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930.

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. *1*, 417–425.

Lorenz, U. (2009). SHP-1 and SHP-2 in T cells: two phosphatases functioning at many levels. Immunol. Rev. *228*, 342–359.

Lun, A.T., McCarthy, D.J., and Marioni, J.C. (2016). A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. F1000Res. *5*, 2122.

Mandl, J.N., Liou, R., Klauschen, F., Vrisekoop, N., Monteiro, J.P., Yates, A.J., Huang, A.Y., and Germain, R.N. (2012). Quantification of lymph node transit times reveals differences in antigen surveillance strategies of naive CD4+ and CD8+ T cells. Proc. Natl. Acad. Sci. USA *109*, 18036–18041.

Mandl, J.N., Monteiro, J.P., Vrisekoop, N., and Germain, R.N. (2013). T cell-positive selection uses self-ligand binding strength to optimize repertoire recognition of foreign antigens. Immunity *38*, 263–274.

Martin, B., Banz, A., Bienvenu, B., Cordier, C., Dautigny, N., Bécourt, C., and Lucas, B. (2004). Suppression of CD4+ T lymphocyte effector functions by CD4+CD25+ cells in vivo. J. Immunol. *172*, 3391–3398.

Martin, B., Auffray, C., Delpoux, A., Pommier, A., Durand, A., Charvet, C., Yakonowsky, P., de Boysson, H., Bonilla, N., Audemard, A., et al. (2013). Highly self-reactive naive CD4 T cells are prone to differentiate into regulatory T cells. Nat. Commun. *4*, 2209.

Matson, C.A., Choi, S., Livak, F., Zhao, B., Mitra, A., Love, P.E., and Singh, N.J. (2020). CD5 dynamically calibrates basal NF-κB signaling in T cells during thymic development and peripheral activation. Proc. Natl. Acad. Sci. USA *117*, 14342–14353.

Mayer, A., Mora, T., Rivoire, O., and Walczak, A.M. (2016). Diversity of immune strategies explained by adaptation to pathogen statistics. Proc. Natl. Acad. Sci. USA *113*, 8630–8635.

McCarthy, D.J., Campbell, K.R., Lun, A.T., and Wills, Q.F. (2017). Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. Bioinformatics *33*, 1179–1186.

McInnes, L., Healy, J., and Melville, J. (2020). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv. https://arxiv.org/abs/1802.03426.

Meli, A.P., Fontés, G., Avery, D.T., Leddon, S.A., Tam, M., Elliot, M., Ballesteros-Tato, A., Miller, J., Stevenson, M.M., Fowell, D.J., et al. (2016). The Integrin LFA-1 Controls T Follicular Helper Cell Generation and Maintenance. Immunity *45*, 831–846.

Meli, A.P., Wang, Y., de Kouchkovsky, D.A., Kong, Y., Basu, M.K., Ghosh, S., and Rothlin, C.V. (2020). IL-4-induced hysteresis in naïve T cell activation. bioRxiv. https://doi.org/10.1101/2020.08.31.275842.

Melville, J. (2019). uwot: The Uniform Manifold Approximation and Projection (UMAP) Method for Dimensionality Reduction. R package version 0.1.3.. https://CRAN.R-project.org/package=uwot.

Mingueneau, M., Kreslavsky, T., Gray, D., Heng, T., Cruse, R., Ericson, J., Bendall, S., Spitzer, M.H., Nolan, G.P., Kobayashi, K., et al.; Immunological Genome Consortium (2013). The transcriptional landscape of αβ T cell differentiation. Nat. Immunol. *14*, 619–632.

Mombaerts, P., Clarke, A.R., Rudnicki, M.A., Iacomini, J., Itohara, S., Lafaille, J.J., Wang, L., Ichikawa, Y., Jaenisch, R., Hooper, M.L., et al. (1992). Mutations in T-cell antigen receptor genes alpha and beta block thymocyte development at different stages. Nature *360*, 225–231.

Moran, A.E., Holzapfel, K.L., Xing, Y., Cunningham, N.R., Maltzman, J.S., Punt, J., and Hogquist, K.A. (2011). T cell receptor signal strength in Treg and iNKT cell development demonstrated by a novel fluorescent reporter mouse. J. Exp. Med. *208*, 1279–1289.

Muraro, M.J., Dharmadhikari, G., Grün, D., Groen, N., Dielen, T., Jansen, E., van Gurp, L., Engelse, M.A., Carlotti, F., de Koning, E.J., and van Oudenaarden, A. (2016). A Single-Cell Transcriptome Atlas of the Human Pancreas. Cell Syst. *3*, 385–394.e3.

Nelson, R.W., Beisang, D., Tubo, N.J., Dileepan, T., Wiesner, D.L., Nielsen, K., Wüthrich, M., Klein, B.S., Kotov, D.I., Spanier, J.A., et al. (2015). T cell receptor cross-reactivity between similar foreign and self peptides influences naive cell population size and autoimmunity. Immunity *42*, 95–107.

Ordoñez-Rueda, D., Lozano, F., Sarukhan, A., Raman, C., Garcia-Zepeda, E.A., and Soldevila, G. (2009). Increased numbers of thymic and peripheral CD4+ CD25+Foxp3+ cells in the absence of CD5 signaling. Eur. J. Immunol. *39*, 2233–2247.

Oukka, M. (2007). Interplay between pathogenic Th17 and regulatory T cells. Ann. Rheum. Dis. *66* (*Suppl 3*), iii87–iii90.

Oxenius, A., Bachmann, M.F., Zinkernagel, R.M., and Hengartner, H. (1998). Virus-specific MHC-class II-restricted TCR-transgenic mice: effects on humoral and cellular immune responses after viral infection. Eur. J. Immunol. *28*, 390–400.

Padilla, O., Calvo, J., Vilà, J.M., Arman, M., Gimferrer, I., Places, L., Arias, M.T., Pujana, M.A., Vives, J., and Lozano, F. (2000). Genomic organization of the human CD5 gene. Immunogenetics *51*, 993–1001.

Persaud, S.P., Parker, C.R., Lo, W.L., Weber, K.S., and Allen, P.M. (2014). Intrinsic CD4+ T cell sensitivity and response to a pathogen are set and

sustained by avidity for thymic and peripheral complexes of self peptide and MHC. Nat. Immunol. *15*, 266–274.

Powrie, F., Leach, M.W., Mauze, S., Caddle, L.B., and Coffman, R.L. (1993). Phenotypically distinct subsets of CD4+ T cells induce or protect from chronic intestinal inflammation in C. B-17 scid mice. Int. Immunol. *5*, 1461–1471.

Raivo, K. (2018). pheatmap: Pretty Heatmaps. R package version 1.0.12.. https://rdrr.io/cran/pheatmap/.

Ren, F., Zhan, X., Martens, G., Lee, J., Center, D., Hanson, S.K., and Kornfeld, H. (2005). Pro-IL-16 regulation in activated murine CD4+ lymphocytes. J. Immunol. *174*, 2738–2745.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Rodriguez, R.M., Lopez-Larrea, C., and Suarez-Alvarez, B. (2015). Epigenetic dynamics during CD4(+) T cells lineage commitment. Int. J. Biochem. Cell Biol. *67*, 75–85.

Rudd, B.D. (2020). Neonatal T Cells: A Reinterpretation. Annu. Rev. Immunol. *38*, 229–247.

Ruterbusch, M., Pruner, K.B., Shehata, L., and Pepper, M. (2020). In Vivo $CD4^+$ T Cell Differentiation and Function: Revisiting the Th1/Th2 Paradigm. Annu. Rev. Immunol. *38*, 705–725.

Salmon-Divon, M., Dvinge, H., Tammoja, K., and Bertone, P. (2010). PeakAnalyzer: genome-wide annotation of chromatin binding and modification loci. BMC Bioinformatics *11*, 415.

Schatz, D.G., and Ji, Y. (2011). Recombination centres and the orchestration of V(D)J recombination. Nat. Rev. Immunol. *11*, 251–263.

Scott, A.C., Dündar, F., Zumbo, P., Chandran, S.S., Klebanoff, C.A., Shakiba, M., Trivedi, P., Menocal, L., Appleby, H., Camara, S., et al. (2019). TOX is a critical regulator of tumour-specific T cell differentiation. Nature *571*, 270–274.

Seo, H., Chen, J., González-Avalos, E., Samaniego-Castruita, D., Das, A., Wang, Y.H., López-Moyado, I.F., Georges, R.O., Zhang, W., Onodera, A., et al. (2019). TOX and TOX2 transcription factors cooperate with NR4A transcription factors to impose $CD8^+$ T cell exhaustion. Proc. Natl. Acad. Sci. USA *116*, 12410–12415.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. *13*, 2498–2504.

Smith, K., Seddon, B., Purbhoo, M.A., Zamoyska, R., Fisher, A.G., and Merkenschlager, M. (2001). Sensory adaptation in naive peripheral CD4 T cells. J. Exp. Med. *194*, 1253–1261.

Sood, A., Lebel, M.E., Fournier, M., Rogers, D., Mandl, J.N., and Melichar, H.J. (2019). Differential interferon-gamma production potential among naïve $CD4^+$ T cells exists prior to antigen encounter. Immunol. Cell Biol. *97*, 931–940.

Sood, A., Lebel, M.E., Dong, M., Fournier, M., Vobecky, S.J., Haddad, É., Delisle, J.S., Mandl, J.N., Vrisekoop, N., and Melichar, H.J. (2021). CD5 levels define functionally heterogeneous populations of naïve human $CD4^+$ T cells. Eur. J. Immunol. *51*, 1365–1376.

Stefanová, I., Dorfman, J.R., and Germain, R.N. (2002). Self-recognition promotes the foreign antigen sensitivity of naive T lymphocytes. Nature *420*, 429–434.

Stefanová, I., Hemmer, B., Vergelli, M., Martin, R., Biddison, W.E., and Germain, R.N. (2003). TCR ligand discrimination is enforced by competing ERK positive and SHP-1 negative feedback pathways. Nat. Immunol. *4*, 248–254.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545–15550.

Thorvaldsdóttir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief. Bioinform. *14*, 178–192.

Tubo, N.J., Pagán, A.J., Taylor, J.J., Nelson, R.W., Linehan, J.L., Ertelt, J.M., Huseby, E.S., Way, S.S., and Jenkins, M.K. (2013). Single naive CD4+ T cells from a diverse repertoire produce different effector cell types during infection. Cell *153*, 785–796.

Tung, J.W., Kunnavatana, S.S., Herzenberg, L.A., and Herzenberg, L.A. (2001). The regulation of CD5 expression in murine T cells. BMC Mol. Biol. *2*, 5.

Villani, A.C., Satija, R., Reynolds, G., Sarkizova, S., Shekhar, K., Fletcher, J., Griesbeck, M., Butler, A., Zheng, S., Lazo, S., et al. (2017). Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. Science *356*, eaah4573.

Vrisekoop, N., Monteiro, J.P., Mandl, J.N., and Germain, R.N. (2014). Revisiting thymic positive selection and the mature T cell repertoire for antigen. Immunity *41*, 181–190.

Vrisekoop, N., Artusa, P., Monteiro, J.P., and Mandl, J.N. (2017). Weakly self-reactive T-cell clones can homeostatically expand when present at low numbers. Eur. J. Immunol. *47*, 68–73.

Wherry, E.J., Blattman, J.N., Murali-Krishna, K., van der Most, R., and Ahmed, R. (2003). Viral persistence alters CD8 T-cell immunodominance and tissue distribution and results in distinct stages of functional impairment. J. Virol. *77*, 4911–4927.

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis, Second Edition (Springer).

Wolf, T., Jin, W., Zoppi, G., Vogel, I.A., Akhmedov, M., Bleck, C.K.E., Beltraminelli, T., Rieckmann, J.C., Ramirez, N.J., Benevento, M., et al. (2020). Dynamics in protein translation sustaining T cell preparedness. Nat. Immunol. *21*, 927–937.

Xie, X., Shi, Q., Wu, P., Zhang, X., Kambara, H., Su, J., Yu, H., Park, S.Y., Guo, R., Ren, Q., et al. (2020). Single-cell transcriptome profiling reveals neutrophil heterogeneity in homeostasis and infection. Nat. Immunol. *21*, 1119–1133.

Xu, W., Zhao, X., Wang, X., Feng, H., Gou, M., Jin, W., Wang, X., Liu, X., and Dong, C. (2019). The Transcription Factor Tox2 Drives T Follicular Helper Cell Development via Regulating Chromatin Accessibility. Immunity *51*, 826–839.e5.

Yu, W., Nagaoka, H., Jankovic, M., Misulovin, Z., Suh, H., Rolink, A., Melchers, F., Meffre, E., and Nussenzweig, M.C. (1999). Continued RAG expression in late stages of B cell development and no apparent re-induction after immunization. Nature *400*, 682–687.

Yusuf, I., and Fruman, D.A. (2003). Regulation of quiescence in lymphocytes. Trends Immunol. *24*, 380–386.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.

Zhou, L., Chong, M.M., and Littman, D.R. (2009). Plasticity of CD4+ T cell lineage differentiation. Immunity *30*, 646–655.

Zhu, J., Yamane, H., and Paul, W.E. (2010). Differentiation of effector CD4 T cell populations (*). Annu. Rev. Immunol. *28*, 445–489.

Zinzow-Kramer, W.M., Weiss, A., and Au-Yeung, B.B. (2019). Adaptation by naïve $CD4^+$ T cells to self-antigen-dependent TCR signaling induces functional heterogeneity and tolerance. Proc. Natl. Acad. Sci. USA *116*, 15160–15169.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-mouse TCRβ (H57-597)	BioLegend	Cat #109243;; RRID: AB_2629564
Anti-mouse CD4 (RM4.5)	BioLegend	Cat #100552; RRID: AB_2563053
Anti-mouse CD8a (53-6.7)	BioLegend	Cat #100742; RRID: AB_2563056
Anti-mouse CD5 (53-7.3)	BioLegend	Cat #100627; RRID: AB_2563930
Anti-mouse Foxp3 (FJK-16 s)	Life Technologies	Cat #11-5773-82; RRID: AB_465243
Anti-mouse CD44 (IM7)	Life Technologies	Cat #45-0441-82; RRID: AB_925746
Anti-mouse CD62L (MEL-14)	BioLegend	Cat #104440; RRID: AB_2629685
Anti-mouse CD25 (PC61.5)	Life Technologies	Cat #48-0251-82; RRID: AB_10671550
Anti-mouse CD45 (30F11)	Life Technologies	Cat #17-0451-83; RRID: AB_469393
Anti-mouse CD98 (RL388)	Life Technologies	Cat #12-0981-81; RRID: AB_465792
Anti-mouse GITR (DTA-1)	Life Technologies	Cat #25-5874-80; RRID: AB_10548516
Anti-mouse LFA-1 (H155-78)	BioLegend	Cat #141007; RRID: AB_10694861
Anti-mouse CD73 (TY/11.8)	BioLegend	Cat #127210; RRID: AB_11218786
Anti-mouse PD-1 (29F.1A12)	BioLegend	Cat #135216; RRID: AB_10689635
Anti-mouse FolR4 (eBio12A5)	BioLegend	Cat #125009; RRID: AB_1134201
Anti-mouse Ly6C (HK1.4)	BioLegend	Cat #128012; RRID: AB_1659241
Anti-mouse CD6 (OX-129)	BioLegend	Cat #146404; RRID: AB_2562753
Anti-mouse CXCR5 (SPRCL5)	Life Technologies	Cat #13-7185-82; RRID: AB_2572800
Anti-mouse CD45.1 (A20)	Life Technologies	Cat #12-0453-82; RRID: AB_465675
Anti-mouse CD45.2 (104)	BioLegend	Cat #109806; RRID: AB_313443
Anti-mouse TOX (TXRX10)	Life Technologies	Cat #12-6502-82; RRID: AB_10855034
Anti-mouse CD69 (H1.2F3)	Life Technologies	Cat #25-0691-82; RRID: AB_469637
Anti-mouse B220 (RA3-6B2)	BioLegend	Cat #103241; RRID: AB_11204069
Anti-mouse F4/80 (BM8)	Life Technologies	Cat #48-4801-82; RRID: AB_1548747
Anti-mouse Ly6G (1A8)	BioLegend	Cat #127641; RRID: AB_2565881
Anti-mouse CD11b (M1/70)	BioLegend	Cat #101259; RRID: AB_2566568
Anti-mouse CD11c (N418)	Life Technologies	Cat #48-0114-82; RRID: AB_1548654
Anti-mouse NK1.1 (PK136)	BioLegend	Cat #108722; RRID: AB_2132712
Anti-mouse CD19 (eBio1D3)	Life Technologies	Cat #48-0193-82; RRID: AB_2734905
Anti-mouse IL-2 (JES6-5H4)	Life Technologies	Cat #12-7021-82; RRID: AB_466150
LEF-1 (C12A5) Rabbit mAb	Cell Signaling Technology	Cat #2230S; RRID: AB_823558
TCF-1 (C46C7) Rabbit mAb	Cell Signaling Technology	Cat #2206S; RRID: AB_2199300
SHP-1 (C14H6) Rabbit mAb	Cell Signaling Technology	Cat #3759S; RRID: AB_2173694
Goat anti-Rabbit IgG (H+L) Cross-Absorbed Secondary Antibody, Alexa Fluor 488	Life Technologies	Cat #A-11008; RRID: AB_143165
Donkey anti-Rabbit IgG (H+L) Highly Cross-Absorbed Secondary Antibody, Alexa Fluor 647	Life Technologies	Cat #A-31573; RRID: AB_2536183
Ultra-LEAF Purified anti-mouse CD3e	BioLegend	Cat #100340; RRID: AB_11149115
Ultra-LEAF Purified anti-mouse CD28	BioLegend	Cat #102116; RRID: AB_11147170
TruStain FcX (anti-mouse CD16/32)	BioLegend	Cat #101320; RRID: AB_1574975
Bacterial and virus strains		
LCMV Armstrong	Dr. John Harty Lab	University of Iowa, Iowa City, USA

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
RPMI 1640	Wisent	Cat #350-030-CL
FBS	Wisent	Cat #080-450
L-glutamine (200mM)	Wisent	Cat #609-065-EL
Penicillin/streptomycin (5000IU)	Wisent	Cat #450-201-EL
2-Mercaptoethanol (50mM)	Life Technologies	Cat #21985023
Non-essential amino acids (100X)	Wisent	Cat #321-011-EL
Sodium pyruvate (100mM)	Wisent	Cat #600-110-EL
HEPES (1M)	Wisent	Cat #330-050-EL
0.5M EDTA (pH 8.0)	Life Technologies	Cat #AM9262
IL-7	BioLegend	Cat #577804
Zombie UV Fixable Viability Kit	BioLegend	Cat #423108
Fixable Viability Dye eFluor 780	Life Technologies	Cat #65-0865-18
Fixable Viability Dye eFluor 506	Life Technologies	Cat #65-0866-18
PBS (1X)	Wisent	Cat #311-010-CL
Permeabilization Buffer (10X)	Life Technologies	Cat #00-8333-56
TRIzol	ThermoFisher Scientific	Cat #15596026
Brefeldin A solution (1000X)	BioLegend	Cat #420601
Monensin solution (1000X)	Life Technologies	Cat #00-4505-51
ACK lysis buffer	Life Technologies	Cat #A1049201
FoxP3/Transcription Factor Fixation/ Permeabilization Concentrate and Diluent	Life Technologies	Cat #00-5521-00
Trypan Blue Solution, 0.4%	ThermoFisher Scientific	Cat# 15250061
Critical commercial assays		
Transcription factor phospho buffer set	BD Biosciences	Cat #563239
EasySep Mouse CD4 T cell Isolation Kit	Stemcell	Cat #19852
EasySep Mouse T Cell Isolation Kit	Stemcell	Cat #19851
Quick-RNA miniprep kit	Zymo Research	Cat #R1054
Purelink RNA Mini Kit	Life Technologies	Cat #12183018A
High-Capacity cDNA reverse transcription kit	Life Technologies	Cat #4368814
Deposited data		
GEO SuperSeries (bulk ATAC-seq, bulk RNA-seq, and scRNA-seq)	This paper	GEO: GSE185677
Bulk ATAC-seq	This paper	GEO: GSE185674
Bulk RNA-seq	This paper	GEO: GSE185675
scRNA-seq	This paper	GEO: GSE185676
scRNA-seq source code	Zenodo, this paper	https://doi.org/10.5281/zenodo.5570794
Experimental models: Organisms/strains		
C57BL/6 mice	Jackson Laboratories	Cat #000664; RRID: IMSR_JAX:000664
B6 CD45.1 mice	Jackson Laboratories	Cat #002014; RRID: IMSR_JAX:002014
Foxp3^{GFP} transgenic mice	(Oukka, 2007)	N/A
SMARTA TCR transgenic mice	(Oxenius et al., 1998)	N/A
TCR$\beta^{-/-}$ mice	(Mombaerts et al., 1992)	N/A
RAG2^{GFP} transgenic mice	(Yu et al., 1999)	N/A
Oligonucleotides		
Cd5 (Mm00432417_m1) FAM	Life Technologies	Cat #4331182
Foxp3 (Mm00475162_m1) FAM	Life Technologies	Cat #4331182
Gapdh (Mm99999915_g1) FAM	Life Technologies	Cat #4351370

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Tbp (Mm01277042_m1) VIC	Life Technologies	Cat #4448489
Software and algorithms		
GraphPad Prism	GraphPad	https://www.graphpad.com; RRID: SCR_002798
FlowJo	BD Biosciences	https://www.flowjo.com; RRID:SCR_008520
Integrative Genomics Viewer 2.9.4	(Thorvaldsdóttir et al., 2013)	https://software.broadinstitute.org/software/igv/; RRID: SCR_011793
MeV	(Howe et al., 2011)	https://mev.tm4.org/#/about; RRID: SCR_001915
GSEA	(Subramanian et al., 2005)	https://www.gsea-msigdb.org/gsea/index.jsp; RRID: SCR_003199
Cytoscape 3.8.1	(Shannon et al., 2003)	https://cytoscape.org; RRID: SCR_003032
ClueGO 2.5.4	(Bindea et al., 2009)	http://www.ici.upmc.fr/cluego/; RRID: SCR_005748
ScatterSlice	(Cotari et al., 2013)	N/A
HOMER	(Heinz et al., 2010)	http://homer.ucsd.edu/homer; RRID: SCR_010881
CASAVA 1.8.2	Illumina	https://www.illumina.com; RRID: SCR_001802
TopHat 2.0.11	(Kim et al., 2013)	http://ccb.jhu.edu/software/tophat/index.shtml; RRID: SCR_013035
featureCounts 1.4.5 in SubRead package	(Liao et al., 2014)	N/A
EdgeR	(Robinson et al., 2010)	http://bioconductor.org/packages/release/bioc/html/edgeR.html; RRID: SCR_012802
BWA	(Li and Durbin, 2010)	http://bio-bwa.sourceforge.net/; RRID: SCR_010910
scater	(McCarthy et al., 2017)	https://bioconductor.org/packages/release/bioc/html/scater.html; RRID: SCR_015954
scran	(Lun et al., 2016)	https://bioconductor.org/packages/release/bioc/html/scran.html; RRID: SCR_016944
uwot	(Melville, 2019)	N/A
umap	(McInnes et al., 2020)	https://github.com/lmcinnes/umap; RRID: SCR_018217
FastQC 0.10.1	Babraham Bioinformatics	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/; RRID: SCR_014583
Trimmomatic 0.33	(Bolger et al., 2014)	http://www.usadellab.org/cms/index.php?page=trimmomatic; RRID: SCR_011848
Bowtie 1.0.0	(Langmead et al., 2009)	http://bowtie-bio.sourceforge.net/index.shtml; RRID: SCR_005476
MACS1.4.1	(Zhang et al., 2008)	N/A
PeakAnalyzer	(Salmon-Divon et al., 2010)	http://www.bioinformatics.org/peakanalyzer/wiki/;; RRID: SCR_001194
pheatmap	(Raivo, 2018)	https://www.rdocumentation.org/packages/pheatmap/versions/0.2/topics/pheatmap; RRID: SCR_016418
ggplot2	(Wickham, 2016)	https://cran.r-project.org/web/packages/ggplot2/index.html; RRID: SCR_014601
Other		
UltraComp eBeads compensation beads	Life Technologies	Cat #01-2222-42

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Judith Mandl (judith.mandl@mcgill.ca).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database and are publicly available as of the date of publication. Accession numbers are listed in the Key resources table.
- All original code has been deposited at Zenodo and is publicly available as of the date of publication. DOIs are listed in the Key resources table.
- Additional information required to reanalyze the data reported in this paper are available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Mice

C57BL/6 mice, CD45.1^{+}, FoxP3^{GFP+} Tg (Oukka, 2007), SMARTA TCR Tg (Oxenius et al., 1998), and TCR$\beta^{-/-}$ (Mombaerts et al., 1992) were purchased from Jackson Laboratories (Bar Harbor, ME) and bred in-house. RAG2^{GFP} Tg mice (Yu et al., 1999) were obtained from Sylvie Lesage (Université de Montréal). All mice were on a B6 background, both male and female mice were used, and were 6-12 weeks of age. Animal housing, care and research were in accordance with the Guide for the Care and Use of Laboratory Animals and all procedures performed were approved by the McGill University, Maisonneuve-Rosemont Hospital Research Center, Radboud University, or NIAID Animal Care Committee.

Infections

LCMV Armstrong was propagated from a strain provided by Dr. J. Harty (University of Iowa). Mice were infected with 2x10^{5} plaque forming units (PFU) by intra-peritoneal injection (Wherry et al., 2003). Cellular responses were assessed 8 days post-infection.

METHOD DETAILS

Lymphocyte isolation, resting, activation, and restimulation

Spleen and peripheral lymph nodes (inguinal, axillary, brachial, superficial cervical, and mesenteric) were passed through a 70 μm filter with 1% RPMI (1% FBS, 1% L-glutamine, and 1% pen/strep). ACK lysis buffer (Life Technologies) was added for 3 minutes and samples were refiltered and resuspended in 1% RPMI. Cell counts were determined by diluting a single-cell suspension 1:10 in Trypan Blue (ThermoFisher Scientific) and manually counting live single cells (Trypan Blue-negative) on a hemocytometer. For experiments where naive CD4^{+} T cells were rested in culture, cells (either total or sorted, as specified) were kept in complete RPMI (10% FBS, 1% L-glutamine, 1% pen/strep, 1% HEPES buffer, 1% Sodium Pyruvate, 1% non-essential Amino Acids, and 0.1% 2-mercapto-ethanol 1000X solution) supplemented with IL-7 (10 ng/mL, Biolegend). To activate T cells, sorted cells or total splenocytes were cultured in complete RPMI in 96-well plates coated with anti-CD3 and anti-CD28 (Invitrogen; both at 3 μg/mL). Restimulation of splenocytes for cytokine production was performed with anti-CD3 and anti-CD28 (Invitrogen; both at 2 μg/mL) in 96-well plates with brefeldin A and monensin (Invitrogen, both diluted 500X) for 5 hours at 37°C.

Flow cytometry

Samples were incubated in Fixable Viability Dye (AF780 or eF506, Life Technologies; Zombie UV, BioLegend) diluted in PBS for 20 minutes at 4°C. Extracellular antibodies were diluted in FACS buffer (2% FBS and 5mM EDTA in PBS) with Fc Block (Life Technologies) and incubated for 30 minutes at 4°C. Samples requiring intracellular staining were subsequently incubated in FoxP3 Transcription Factor Fixation/Permeabilization Concentrate and Diluent (Life Technologies) for 30 minutes at 4°C. Intracellular antibodies were diluted in permeabilization wash buffer and incubated for 30-60 minutes at 4°C. Directly conjugated antibodies used were as follows: TCRβ (H57-597), CD4 (RM4.5), CD8a (53-6.7), CD5 (53-7.3), Foxp3 (FJK-16 s), CD44 (IM7), CD62L (MEL-14), CD25 (PC61.5), CD45 (30F11), CD98 (RL388), GITR (DTA-1), LFA-1 (H155-78), CD73 (TY/11.8), PD-1 (29F.1A12), FolR4 (eBio12A5), Ly6C (HK1.4), CD6 (OX-129), CXCR5 (SPRCL5), CD45.1 (A20), CD45.2 (104), TOX (TXRX10), CD69 (H1.2F3), B220 (RA3-6B2), F4/80 (BM8), Ly6G (1A8), CD11b (M1/70), CD11c (N418), NK1.1 (PK136), CD19 (eBio1D3), and IL-2 (JES6-5H4). Primary unconjugated antibodies used were: LEF-1 (C12A5), TCF-1 (C46C7), SHP-1 (C14H6). Secondary conjugated antibodies used were either Goat anti-Rabbit IgG (H+L) Alexa Fluor 488 or Donkey anti-Rabbit IgG (H+L) Alexa Fluor 647. For samples assessed for SHP-1 expression, cells were fixed with 1X TFP Fix/Perm Buffer for 50 minutes at 4°C, then incubated in Perm Buffer III (BD Biosciences) for 20 minutes on ice. Fc Block, surface, and intracellular antibodies were diluted in 1X TFP Perm/Wash Buffer and incubated for 50 minutes at 4°C, and secondary antibody diluted in 1X TFP Perm/Wash Buffer was added for an additional 50 minutes at 4°C. For all flow cytometry experiments, cells were acquired using an LSRFortessa (BD Bioscience) and analyzed with FlowJo software (BD Bioscience).

Cell sorts

Lymphocytes from B6 or CD45.1$^+$ congenic mice were isolated in single cell suspension as described. Samples for bulk RNA-seq, ATAC-seq, *in vivo* or *in vitro* assays were pooled from spleen and lymph nodes (inguinal, axillary, brachial, mesenteric, and cervical) from 2-10 mice. Samples for scRNA-seq were from a spleen from a single mouse. Total isolated cells or cells magnetically enriched for CD4$^+$ or total T cells (Stemcell EasySep mouse total T cell or CD4$^+$ T cell enrichment kits) were then incubated in fixable viability dye, and subsequently stained with surface antibodies for 1 hour at 4°C. Naive CD4$^+$ T cell were sorted on CD5 expression (top and bottom 15%) for bulk analyses; single naive CD4$^+$ T cells were sorted into 384-well plates for subsequent scRNA-seq. Naive CD4$^+$ T cells were sorted on singlets, live, dump-negative (RNA-seq and ATAC-seq), TCRβ$^+$ (bulk- and scRNA-seq), CD4$^+$, CD8$^-$, CD25$^-$ (scRNA-seq and *in vivo* and *in vitro* assays) or FoxP3^{GFP-} (RNA-seq), CD44^{lo}, CD62L^{hi}, and 15%–25% CD5^{lo} and CD5^{hi} (RNA-seq, ATAC-seq, and *in vivo* and *in vitro* assays). Dump channel included B220, CD11b, CD11c, F4/80, Ly6G, NK1.1, and CD69 for RNA-seq; the ATAC-seq dump channel also included CD19 and CD25 (ATAC-seq). Sorts were performed on either a FACS Aria Fusion, Aria III, or Aria II SORP (BD Bioscience). All cell populations were sorted to > 90% purity for bulk populations.

Adoptive cell transfers

For all adoptive cell transfer experiments, donors and recipients were sex-matched.

TCRβ$^{-/-}$ adoptive transfer

25% CD5^{lo} and CD5^{hi} cells were sorted from CD45.2$^+$ mice and 1x10^6 cells were adoptively transferred by i.v. injection into CD45.1$^+$ TCRβ$^{-/-}$ recipients and mice were followed for 91 days.

LCMV infection

15-18 CD45.1$^+$ or CD45.2$^+$ mice were used as donors to obtain a total of 12-20x10^6 cells from each of 15% CD5^{lo} and 15% CD5^{hi} cells sorted as detailed above. 6-10x10^6 sorted donor cells were adoptively transferred by i.v. injection into CD45.2$^+$ or CD45.1$^+$ recipients (n = 2 per group), respectively. One day post-transfer, mice were infected with LCMV as described. Cells were isolated from the spleens and peripheral lymph nodes of recipient mice 8 days post-infection.

SMARTA transgenic T cell adoptive transfer

1x10^4 CD45.2$^+$ SMARTA CD4$^+$ T cells were adoptively transferred by i.v. injection into CD45.1 recipients. One day post-transfer, mice were infected with LCMV as described. Cells were isolated from the spleens of recipient mice 8 days post-infection.

Bulk RNA sequencing

1x10^6 cells from four independent samples, each with cells pooled from 2 mice, were sorted as described and CD5^{lo} and CD5^{hi} naive CD4$^+$ T cells were either directly added to 500 μL TRIzol (ThermoFisher Scientific) or rested in complete RPMI supplemented with IL-7 for 22 hr first. RNA was purified using RNA miniprep kit (Zymo Research) according to manufacturers' recommendations. 500ng of purified RNA was used to prepare RNA-seq libraries using TruSeq mRNA library preparation kit v2 (Illumina). Libraries were sequenced on an Illumina HiSeq 2000 using v3 chemistry and 50 cycle paired end reads. Illumina bcl files were converted to FASTQ using CASAVA1.8.2 and mapped to the UCSC mm9 *Mus musculus* genome annotation using Tophat 2.0.11 (Kim et al., 2013). Reads overlapping exons were counted using featureCounts version 1.4.5 from the SubRead package (Liao et al., 2014), with a minimum read mapping quality score of 10. Normalized read counts differential gene expression analysis was performed with EdgeR (Robinson et al., 2010). In order for a gene to be included in the matrix a minimum CPM value of 5 in at least 3 of the 4 replicates was required. The p values were corrected using the Benjamini-Hochberg method and an FDR threshold of $\leq$ 0.01 was considered significant. No FC threshold was set unless stated otherwise.

Single cell RNA sequencing

Each well within a 384-well plate contained CEL-Seq2 primers covered by mineral oil. Primers consisted of a 24bp polyT stretch, a 6bp random molecular barcode (UMI), a cell-specific barcode, the 5′ Illumina TruSeq small RNA kit adaptor and a T7 promoter. After sorting, the plates were frozen at −80°C until further use. Single cell RNA-seq library preparation and sequencing was performed by Single Cell Discoveries (Utrecht, Netherlands) (Artegiani et al., 2017). Libraries were prepared following the SORT-seq protocol (Muraro et al., 2016), which consists of an automated and improved version of the CEL-Seq2 protocol (Hashimshony et al., 2016). Briefly, cells were first lysed for 5 minutes at 65°C, and reverse transcription and second strand mixes were dispensed by the Nanodrop II liquid handling platform (GC Biotech). Single cell double stranded cDNAs were pooled together and *in vitro* transcribed for linear amplification. Illumina sequencing libraries were prepared using the TruSeq small RNA primers (Illumina) and sequenced paired-end at 75 bp read length the Illumina NextSeq. Paired-end reads from Illumina sequencing were aligned to the mouse transcriptome genome by BWA (Li and Durbin, 2010). Read 1 contained the barcode information and was used for assigning reads to correct cells and libraries, while read 2 was mapped to gene models. Reads that mapped equally well to multiple locations were discarded. Read counts were first corrected for UMI barcode by removing duplicate reads that had identical combinations of library, cell-specific, and molecular barcodes and were mapped to the same gene. For each cell barcode the number of UMIs for every transcript was counted, and transcript counts were then adjusted to the expected number of molecules based on counts, 256 possible UMI's and poissonian counting statistics (Grün et al., 2014). A unique feature of this protocol is the combination of both flow cytometry staining and RNA sequencing; this allowed for the simultaneous tracking of select protein expression and gene expression on single cells.

Single-cell RNA-seq data analysis

Raw read counts were first subjected to quality control. We identified two blocks of wells with fewer than 500 non-spike-in reads. To exclude these and similar low-content wells, we applied a UMAP clustering on all wells (including the spike-in reads) and excluded the cluster of cells that were mainly composed by the low-read wells. After quality control, 697 wells out of 1152 were kept in the analysis. R packages 'scater' (McCarthy et al., 2017) and 'scran' (Lun et al., 2016) were used for further processing, the spike-in reads were removed and expression values were normalized to library size and normalized log expression values, and gene variance were determined as previously described (Lun et al., 2016). Mitochondrial genes were excluded before modeling gene variance. The processed data was then plotted or subjected to UMAP clustering (McInnes et al., 2020), performed with the R package 'uwot' (Melville, 2019) using the cosine distance and a neighborhood size of 30. For defining CD5 low, mid, and high cells the CD5 MFI was logged by the flow cytometer when we sorted the cells into individual wells.

ATACseq library preparation, sequencing, and visualization

Two independent biological replicates of $CD5^{lo}$ and $CD5^{hi}$ naive $CD4^+$ T cells were sorted as described, counted and $1x10^5$ nuclei pelleted. ATAC-seq libraries were prepared from the fresh nuclei pellets using Illumina Tagment DNA TDE1 Kit by the Institut de recherches cliniques de Montréal (Montreal, Canada). Briefly, paired-end 42bp sequencing reads were generated by Illumina sequencing (using a NovaSeq6000). The quality of the sequenced reads was checked using FastQC tool v0.10.1 (Babraham Bioinformatics), and low-quality bases removed using Trimmomatic v.0.33 (Bolger et al., 2014). The trimmed reads were mapped to the mouse UCSC mm9 genome using Bowtie 1.0.0 (Langmead et al., 2009), in paired-end mode with–best parameter. Peak calling was performed using MACS1.4.1 (Zhang et al., 2008) with p values $< 10^{-7}$. Subpeaks were identified using PeakAnalyzer (Salmon-Divon et al., 2010), with parameters: *valley* = 0.5 and *cutoff* = 5 counts per million (cpm). Normalized sequenced read density profiles (bigwig) were generated using makeUCSCfile from Homer package (Heinz et al., 2010), normalizing the total number of reads in each sample to 10^6, and visualized on Integrative Genomics Viewer (IGV) (Thorvaldsdóttir et al., 2013). When visualizing ATAC-seq signal profiles for individual genes, group scaling was performed. Peaks identified in the biological replicates were pooled using mergePeaks from Homer package, merging peak summits within 50bp to each other. Read densities around the peak summits were retrieved using annotatePeaks from Homer package and quantiles normalized for FC comparison between $CD5^{lo}$ and $CD5^{hi}$ replicates (Bolstad, 2020). Transcription factor binding motif enrichment analysis was performed using Homer package on unique peaks found only in $CD5^{lo}$ or $CD5^{hi}$ replicates with a *P*-value $< 10^{-4}$. Hierarchical clustering of the peaks near the DEG-ND and DEG-D gene sets were performed using Pearson correlation with complete linkage method.

RNA extraction and quantitative real-time PCR

To assess *Cd5* and *Foxp3* expression in total and sorted $CD5^{lo}$ and $CD5^{hi}$ naive $CD4^+$ T cells or Tregs, lymphocytes from B6 or $CD45.1^+$ congenic mice were isolated in single cell suspension as described. Lymphocytes were then magnetically enriched for total T cells or $CD4^+$ T cells (Stemcell EasySep mouse total T cell or $CD4^+$ T cell enrichment kits) and sorted as described. RNA was extracted using Purelink RNA Mini Kit (Life Technologies) and cDNA converted using High-Capacity cDNA Reverse Transcription Kit (Life Technologies) according to manufacturers' recommendations. RT-qPCR analysis was performed with TaqMan Gene Expression Master Mix (Life Technologies) and TaqMan Gene Expression Assay (Life Technologies; FAM, *Cd5*, Mm00432417_m1; FAM, *Foxp3*, Mm00475162_m1). Housekeeping genes *Gapdh* (Life Technologies, FAM, Mm99999915_g1) or *Tbp* (Life Technologies, VIC, Mm01277042_m1) were used.

QUANTIFICATION AND STATISTICAL ANALYSIS

Heatmaps

For RNA-seq these were created by either showing individual replicates or average expression within replicates. Log_{10}(CPM+1) were visualized using the pheatmap package in R-Project on a color scale of black-blue-white-orange-red (Raivo, 2018). For ATAC-seq, heatmaps were created using annotatePeaks from Homer package, taking read densities ± 2.5kb with bin size of 50bp for the highest peak summit near each gene TSS. Images were generated using MeV tool with blue-white-red scale (Howe et al., 2011).

Gene set enrichment analysis

GSEAs were performed as previously described (Subramanian et al., 2005) using gene sets defined by the Molecular Signatures Database (Liberzon et al., 2015) or otherwise described.

Gene ontology pathway analysis

Enrichment of GO terms in naive $CD4^+$ T cells was performed using ClueGO (version 2.5.4) (Bindea et al., 2009) on Cytoscape (version 3.8.1) (Shannon et al., 2003). The following parameters were used when running ClueGO on the top 5% most variable genes from the scRNA-seq: Min GO Level = 4, Max GO Level = 6, Minimum Number of Genes associated to GO term = 6, and Minimum Percentage of Genes associated to GO term = 6. The following parameters were used when running ClueGO on bulk RNA-seq DEGs: Min GO Level = 3; Max GO Level = 4. For $CD5^{lo}$ cells: Minimum Number of Genes associated to GO term = 3; Minimum Percentage of Genes associated to GO term = 5. For $CD5^{hi}$ cells: Minimum Number of Genes associated to GO term = 20; Minimum Percentage of Genes

associated to GO term = 10. Enrichment p values were based on a hypergeometric test and Benjamini-Hochberg method used for multiple testing correction. For bulk RNA-seq only pathways with $p \leq 0.05$ were considered significant.

PCA

PCA plots were built using filtered $\log_2$CPM (RNA-seq) or $\log_2$-transformed read densities around peak summits (ATAC-seq) using ggplot2 package in R-Project (Wickham, 2016).

Mutual Information

Mutual Information (MI) is a robust, non-parametric measure of the statistical relationship between observables with distinct advantages over simple correlation measures (Chan et al., 2017). MI is computed as:

$$MI = \sum_{x,y} p(x,y) \log_2 \left(\frac{p(x,y)}{p(x)p(y)} \right)$$

(Joint) probability distributions are obtained by binning the data into 96 geometrically spaced bins over the full MFI range (10° – 10^6) for TCRβ, CD5, CD6, and SHP-1.

ScatterSlice analysis

Scale values corresponding to single $CD4^+$ T cells with expression of TCRβ, Ly6C, CD5, CD6, TOX, and SHP-1 from flow cytometry data were identified and exported as csv files for analysis in the R-Project ScatterSlice (Cotari et al., 2013). Cells were divided into defined bins (15x15 matrix with a minimum of 15 cells per bin) and within each bin, the average MFI of SHP-1, Ly6C, TOX, or CD6 was projected in false-color onto a plot of TCRβ versus CD5 expression.

Statistical analyses

Group comparisons were performed using Prism V9 (GraphPad). Unless specified, data are presented as mean ± standard deviations (SD) with each data point representing an individual mouse. The cut-off for significance considered was $p<0.05$ for all analyses unless otherwise stated. Information about specific statistical tests used for each experiment are listed in the figure legends.