Supplementary figure and table legends

a

CD4+ | IE | LP | mLN
Tomato−
Tomato+

IE: 90.6, 8.3
LP: 81.7, 16.7
mLN: 90.6, 8.41

Tomato−
IE: 46.3, 0.47 / 31.5, 21.8
LP: 3.63, 0.45 / 23.9, 72.6
mLN: 1.14e-3, 0.09 / 0.69, 99.2

Tomato+
IE: 7.47, 0.86 / 55.1, 36.4
LP: 0.74, 0.77 / 25, 73.3
mLN: 0, 0.17 / 1.44, 98.3

CD8α−BV605 / ThPOK-GFP / Tomato-PE

b

CD4+ | IE | LP | mLN
IE: 61.3, 37.6
LP: 71.1, 27.6
mLN: 55.6, 42.9

Tomato−
IE: 40.1, 0.05 / 41.5, 18.4
LP: 0, 0 / 63.2, 36.7
mLN: 0, 0 / 2.4, 97.5

Tomato+
IE: 5.34, 0.02 / 72.6, 22.1
LP: 0.04, 0 / 65.9, 34.1
mLN: 0, 0 / 10.1, 89.9

CD8α−BV605 / ThPOK-GFP

c

Genes — IE, LP, mLN
RNA molecules — IE, LP, mLN
mtDNA — IE, LP, mLN

d

Cluster (0–24)
Mean expression per group
Percent of cells in group (100, 80, 60, 40, 20)

Genes: Lef1, Junb, Furin, Ikzf2, Gzmb, Ccl5, Capg, Gzma, Cd160, Rgs1, Rps19, Sell, Igfbp4, Izumo1r, Sostdc1, Id2, Pdcd4, Il17a, Il22, Stmn1, Hmgb2, Actb, Rps19, S100a4, Mif, Eif5a, Il10, Ccl20, Ccl4, Ifit1, Cd74

e Motility Signature
UMAP_2 / UMAP_1

f Sell, Klf2, Ccr7 — Expression Level — IE, LP, mLN

g tdTomato
UMAP_2 / UMAP_1

h Th1 Signature
UMAP_2 / UMAP_1

i Cd8a
UMAP_2 / UMAP_1

j Cluster: 18 5 14 12 8 13 22 6 4

IEL: Gzmb, Cd8a, Gzma, Ccl5, Cd7, Cd244, Nkg7, Itgae, Cd160, Lag3
Motility: Actn2, Vim, Actb, Actg1, Ppp1r16b
TCR-Pathway: Cd3e, Egr2, Nr4a2, Trac, Dusp1
Treg: Ikzf2, Tnfrsf18, Shisa5

IE_Treg2, IE_Treg1, IE_PreIEL3, IE_PreIEL2, IE_PreIEL1, IE_CD4IEL4, IE_CD4IEL3, IE_CD4IEL2, IE_CD4IEL1

k

8_IE_PreIEL1 | 12_IE_PreIEL2 | 14_IE_PreIEL3

Lgals1, Tmsb10, Gzma, Gzmb, Ccl5, Id2, Cd160, Ikzf2, Actb, Srrm2, Car2, Cd7, Capg

Normalized Z-score (2, 1, 0, −1, −2)

Supplementary Figure 1. Related to Figure 1.

(**a-k**) iFoxp3^{Tom}ThPOK^{GFP} mice were treated with tamoxifen for 10 weeks, and Tomato$^-$ and Tomato$^+$ CD4$^+$ T cells from mesenteric lymph nodes (mLN), lamina propria (LP) and intestinal epithelium (IE) were sorted for scRNA-Seq using 10X Genomics platform. Sorted Tomato$^-$ (blue gates) and Tomato$^+$ (red gates) cells were pooled in a 2:1 ratio per tissue, resulting in 3 separate libraries. (**a, b**) CD4$^+$ T cells from mLN, LP and IE before sorting (**a**) and after sorting (**b**). (**c**) Number of sequenced genes (left) and RNA molecules (middle) per cluster and percent of mitochondrial DNA (right) per library. (**d**) Top expressed genes per UMAP cluster. Circle size represents proportion of cells per cluster expressing the indicated gene and color represents expression level. (**e**) Expression levels of the motility signature among all sequenced cells. (**f**) Expression levels of *Sell* (top)*, Klf2* (middle)*, and Ccr7* (bottom) of cells from indicated tissues. (**g**) *tdTomato* gene expression in 6,668 sequenced cells. (**h-i**) Expression levels of Th1 signature (**h**) and *Cd8a* (**i**) among all sequenced cells. (**j**) Expression levels of indicated genes among IE clusters. (**k**) Expression heatmap of selected genes in pre-IEL clusters.
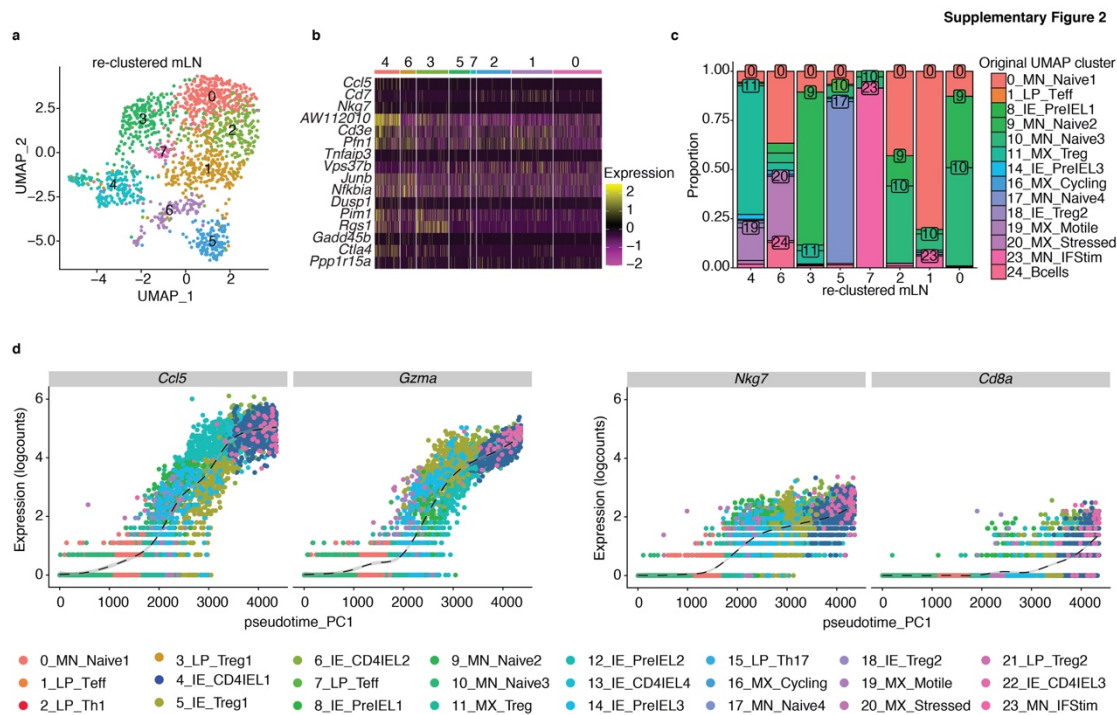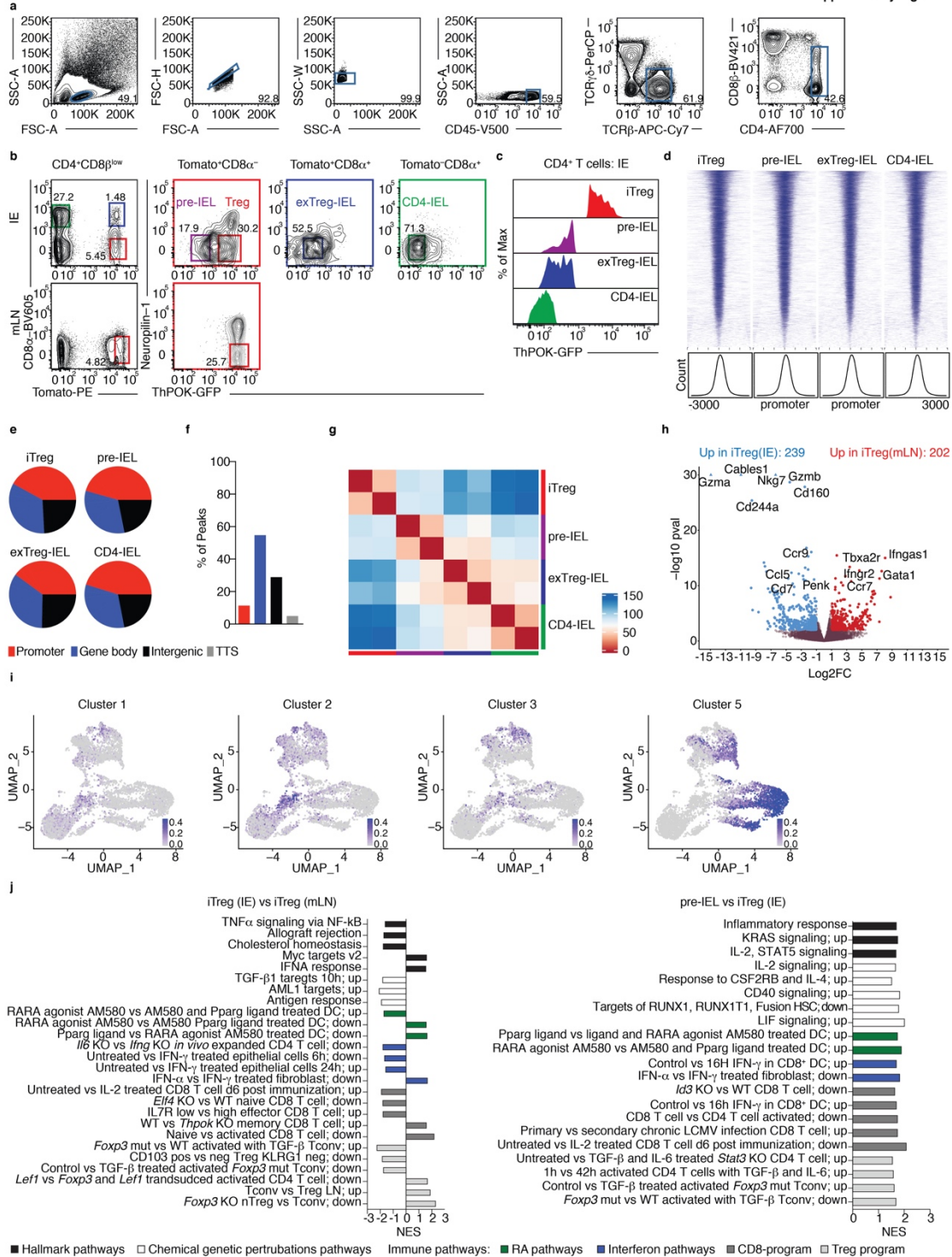
Supplementary Figure 2. Related to Figure 2.

(**a-d**) *Foxp3*[eGFP-Cre-ERT2]x*Rosa26*[lsl-tdTomato]x*Zbtb7b*[GFP](iFoxp3[Tom]ThPOK[GFP]) mice were treated with tamoxifen for 10 weeks, Tomato[−] and Tomato[+] CD4[+] T cells from mesenteric lymph nodes (mLN), lamina propria (LP) and intestinal epithelium (IE) were sorted for scRNA-Seq using 10X Genomics platform. Sorted Tomato[−] and Tomato[+] cells were pooled in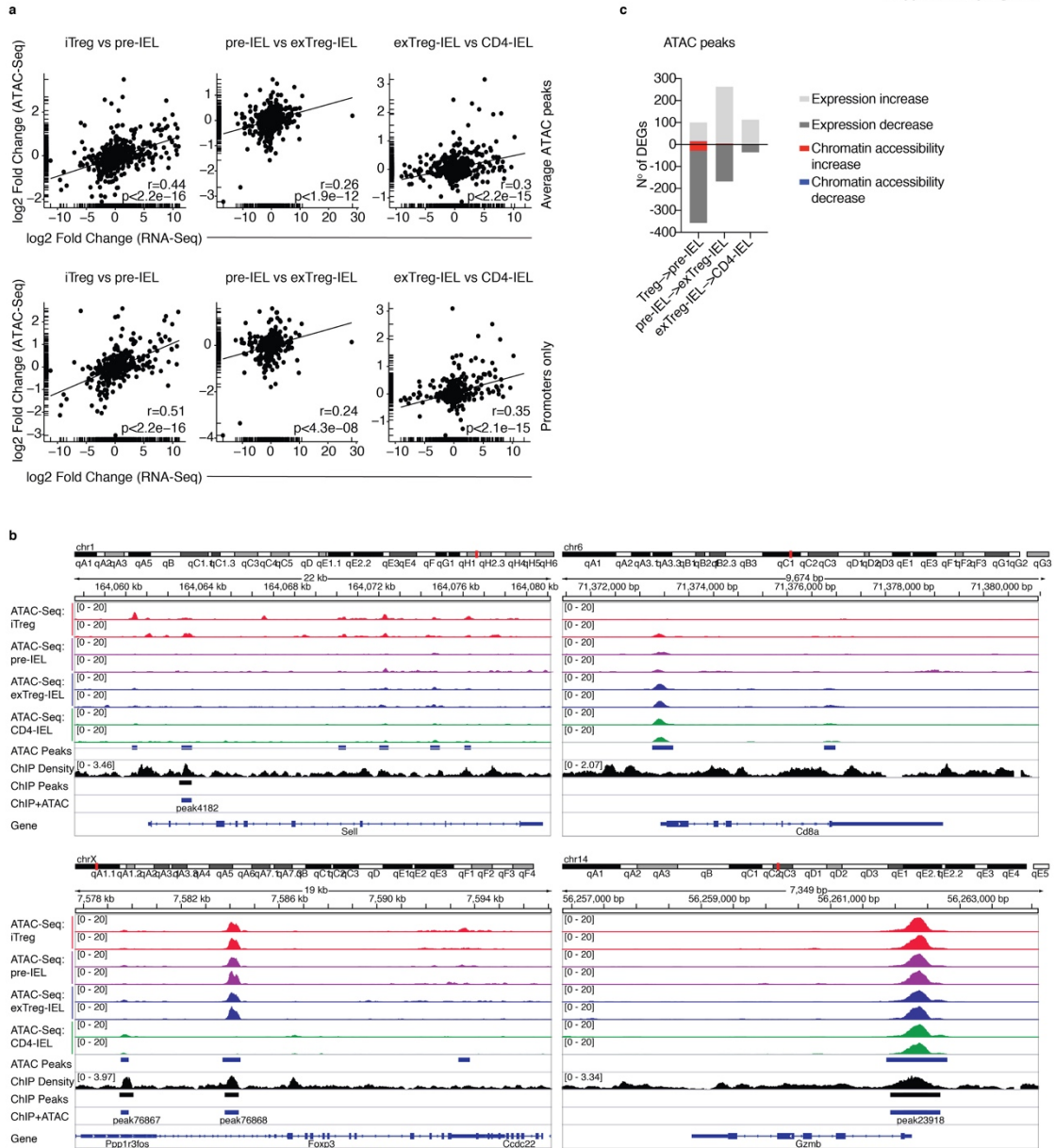 a 2:1 ratio per tissue, resulting in 3 separate libraries. (**a-c**) mLN-derived cells were re-clustered into 8 new clusters (0-7). (**a**) Re-clustered mLN cells with number indicating the new clusters, consistent throughout the figure. (**b**) Expression levels of top differentially expressed genes among re-clustered cells. (**c**) Proportion of original UMAP clusters per re-clustered mLN clusters (X-axis), as indicated. Top 3 original UMAP cluster contained within re-clustered mLN indicated in a box. (**d**) Scatter plot of gene expression ranked based on pseudotime principle component 1 (PC1).

a

SSC-A / FSC-A — 49.1

FSC-H / FSC-A — 92.8

SSC-W / SSC-A — 99.9

SSC-A / CD45-V500 — 59.5

TCRγδ-PerCP / TCRβ-APC-Cy7 — 61.9

CD8β-BV421 / CD4-AF700 — 72.6

b

CD4⁺CD8β^low — IE: 27.2, 1.48, 5.45

Tomato⁺CD8α⁻ — pre-IEL, Treg: 17.9, 30.2

Tomato⁺CD8α⁺ — exTreg-IEL: 52.5

Tomato⁻CD8α⁺ — CD4-IEL: 71.3

mLN — CD8α-BV605 / Tomato-PE: 4.82

Neuropilin-1 / ThPOK-GFP: 25.7

c CD4⁺ T cells: IE

% of Max — ThPOK-GFP: iTreg, pre-IEL, exTreg-IEL, CD4-IEL

d iTreg, pre-IEL, exTreg-IEL, CD4-IEL

Count — -3000, promoter, promoter, 3000

e iTreg, pre-IEL, exTreg-IEL, CD4-IEL

Promoter, Gene body, Intergenic, TTS

f % of Peaks

g iTreg, pre-IEL, exTreg-IEL, CD4-IEL — 150, 100, 50, 0

h Up in iTreg(IE): 239 — Up in iTreg(mLN): 202

Gzma, Cables1, Nkg7, Gzmb, Cd160, Cd244a, Ccr9, Tbxa2r, Ifngas1, Ccl5, Cd7, Penk, Ifngr2, Gata1, Ccr7

-log10 pval / Log2FC

i Cluster 1, Cluster 2, Cluster 3, Cluster 5

UMAP_2 / UMAP_1 — 0.4, 0.2, 0.0

j

iTreg (IE) vs iTreg (mLN)

- TNFα signaling via NF-kB
- Allograft rejection
- Cholesterol homeostasis
- Myc targets v2
- IFNA response
- TGF-β1 targets 10h; up
- AML1 targets; up
- Antigen response
- RARA agonist AM580 vs AM580 and Pparg ligand treated DC; up
- RARA agonist AM580 vs AM580 Pparg ligand treated DC; down
- Pparg ligand vs RARA agonist AM580 treated DC; down
- Il6 KO vs Ifng KO in vivo expanded CD4 T cell; down
- Untreated vs IFN-γ treated epithelial cells 6h; down
- Untreated vs IFN-γ treated epithelial cells 24h; up
- IFN-α vs IFN-γ treated fibroblast; down
- Untreated vs IL-2 treated CD8 T cell d6 post immunization; up
- Elf4 KO vs WT naive CD8 T cell; down
- IL7R low vs high effector CD8 T cell; up
- WT vs Thpok KO memory CD8 T cell; up
- Naive vs activated CD8 T cell; down
- Foxp3 mut vs WT activated with TGF-β Tconv; down
- CD103 pos vs neg Treg KLRG1 neg; down
- Control vs TGF-β treated activated Foxp3 mut Tconv; down
- Lef1 vs Foxp3 and Lef1 transduced activated CD4 T cell; down
- Tconv vs Treg LN; up
- Foxp3 KO nTreg vs Tconv; down

NES: -3 -2 -1 0 1 2 3

pre-IEL vs iTreg (IE)

- Inflammatory response
- KRAS signaling; up
- IL-2, STAT5 signaling
- IL-2 signaling; up
- Response to CSF2RB and IL-4; up
- CD40 signaling; up
- Targets of RUNX1, RUNX1T1, Fusion HSC; down
- LIF signaling; up
- Pparg ligand vs ligand and RARA agonist AM580 treated DC; up
- RARA agonist AM580 vs AM580 and Pparg ligand treated DC; up
- Control vs 16H IFN-γ in CD8⁺ DC; up
- IFN-α vs IFN-γ treated fibroblast; down
- Id3 KO vs WT CD8 T cell; down
- Control vs 16h IFN-γ in CD8⁺ DC; up
- CD8 T cell vs CD4 T cell activated; down
- Primary vs secondary chronic LCMV infection CD8 T cell; up
- Untreated vs IL-2 treated CD8 T cell d6 post immunization; down
- Untreated vs TGF-β and IL-6 treated Stat3 KO CD4 T cell; up
- 1h vs 42h activated CD4 T cells with TGF-β and IL-6; up
- Control vs TGF-β treated activated Foxp3 mut Tconv; up
- Foxp3 mut vs WT activated with TGF-β Tconv; down

NES: 0 1 2 3

Hallmark pathways ■ Chemical genetic pertrubations pathways □ Immune pathways: RA pathways ■ Interferon pathways ■ CD8-program ■ Treg program □

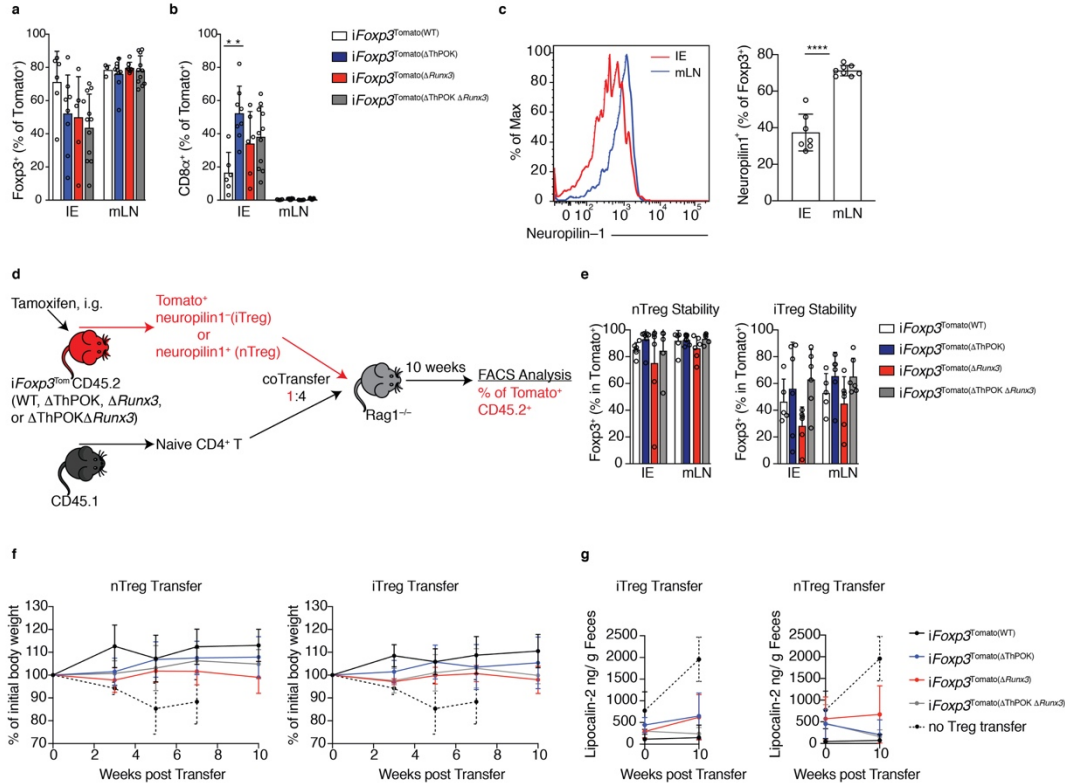**Supplementary Figure 3**. Related to Figure 3.

(**a-j**) iFoxp3^{Tom}ThPOK^{GFP} mice were treated with tamoxifen for 10 weeks and induced Tregs (iTreg; CD4$^+$Tomato$^+$GFP^{High}neuropilin-1$^-$CD8α^-), pre-IELs (CD4$^+$Tomato$^+$ GFP^{Low}CD8α^-), exTreg-IELs (CD4$^+$Tomato$^+$GFP^{Low}CD8α^+), and CD4-IELs (CD4$^+$ Tomato$^-$GFP^{Low}CD8α^+) were sorted in bulk from the IE. Assay for transposase-accessible chromatin (ATAC) or RNA libraries were prepared followed by sequencing of indicated populations. iTregs were also sorted from the mLN for RNA-seq. (**a**) Gating strategy of CD4$^+$ T cells in sequential order. (**b**) Sorting strategy of the indicated populations. CD8α and Tomato expression among CD4$^+$ CD8β^{low} T cells (left) from IE (top) and mLN (bottom). ThPOK and neuropilin-1 expression among Tomato$^+$CD8α^- (red), Tomato$^+$CD8α^+ (purple) and Tomato$^-$CD8α^+ (green) cells from IE (top) and mLN (bottom). (**c**) Levels of ThPOK expression in each sorted population in the IE. (**d**) Mapped accessible chromatin regions relative to promoters of all genes in ATAC-Seq data (top) and their relative read counts per genomic regions (bottom). (**e**) ATAC-Seq peak annotations as indicated per cell type. (**f**) Percent of total differentially accessible chromatin regions as follows: 5'UTR and promoters (Promoter; red), 3'UTR with exons and introns (Gene Body; blue), transcriptional termination site (TTS; gray) and intergenic (black). (**g**) Euclidean distance correlation of chromatin accessibility profiles of all samples. (**h**) Volcano representation of differentially expressed genes between IE iTregs (higher expression in blue) and mLN iTregs (higher expression in red), performed by Wald pairwise comparison test, p$_{adj}$<0.05 values were considered significant. (**i**) Treg signature from clusters 1- 3 and IEL signature from cluster 5 of the bulk RNA-seq heatmap (Figure 3d) overlaid onto the scRNA-Seq UMAP from figure 1. (**j**) Curated list of gene set enrichment analysis of Hallmark pathways (black), chemical genetic perturbations pathways (white), and immune pathways (Retinoic acid (RA) pathways; green, Interferon pathways; blue, CD8-program; gray, and Treg program; silver). Significant differentially accessible chromatin regions p<0.01 and significant differentially expressed genes p<0.05 in RNA-Seq. Each sample for ATAC-Seq consisted of 5,000-40,000 cells from 6 or 9 pooled mice, n=2 samples. Each sample for RNA-Seq consisted of 300-800 cells per mouse, n=2-3 mice.

a

iTreg vs pre-IEL

pre-IEL vs exTreg-IEL

exTreg-IEL vs CD4-IEL

log2 Fold Change (ATAC-Seq)

log2 Fold Change (RNA-Seq)

r=0.44
p<2.2e−16

r=0.26
p<1.9e−12

r=0.3
p<2.2e−15

Average ATAC peaks

iTreg vs pre-IEL

pre-IEL vs exTreg-IEL

exTreg-IEL vs CD4-IEL

log2 Fold Change (ATAC-Seq)

log2 Fold Change (RNA-Seq)

r=0.51
p<2.2e−16

r=0.24
p<4.3e−08

r=0.35
p<2.1e−15

Promoters only

c

ATAC peaks

N° of DEGs

- Expression increase
- Expression decrease
- Chromatin accessibility increase
- Chromatin accessibility decrease

Treg→pre-IEL
pre-IEL→exTreg-IEL
exTreg-IEL→CD4-IEL

b

chr1

chr6

ATAC-Seq: iTreg
ATAC-Seq: pre-IEL
ATAC-Seq: exTreg-IEL
ATAC-Seq: CD4-IEL
ATAC Peaks
ChIP Density
ChIP Peaks
ChIP+ATAC
Gene

peak4182   Sell

Cd8a

chrX

chr14

peak76867 peak76868
Ppp1r3fos Foxp3 Ccdc22

peak23918
Gzmb

Supplementary Figure 4. Related to Figure 4.

(**a-c**) iFoxp3^{Tom}ThPOK^{GFP} mice were treated with tamoxifen for 10 weeks and induced Tregs (iTreg; CD4$^+$Tomato$^+$GFP^{High}neuropilin-1$^-$CD8α^-), pre-IELs (CD4$^+$Tomato$^+$ GFP^{Low}CD8α^-), exTreg-IELs (CD4$^+$Tomato$^+$GFP^{Low}CD8α^+), and CD4-IELs (CD4$^+$ Tomato$^-$GFP^{Low}CD8α^+) were sorted in bulk from the IE. Assay for transposase-accessible chromatin (ATAC) or RNA libraries were prepared followed by sequencing of indicated populations. (**a**) Pearson's correlation (r) of log2 fold changes of RNA-Seq vs ATAC-Seq of total ATAC peaks averaged (top) or at promoters only (bottom) of indicated cell types. P value (p) for r coefficient as indicated. (**b**) ATAC-Seq peaks of indicated populations and ThPOK chromatin immunoprecipitation followed by sequencing (ChIP-Seq) of *in vivo* differentiated Foxp3$^+$ splenic Tregs displayed on the integrative genomics viewer (IGV) in select regions as indicated. Overlap regions between ChIP- and ATAC-Seq indicated on bottom. (**c**) Numbers of differentially expressed genes (DEGs) (top) (silver for increase, gray for decrease) in between cell types in sequential progression as performed by Wald pairwise test as indicated, with differentially accessible chromatin regions (DACR) within those genes (red for increase, blue for decrease in accessibility). Significant DACR p<0.01 and significant DEG p<0.05 in RNA-Seq. Each sample for ATAC-Seq consisted of 5,000-40,000 cells from 6 or 9 pooled mice, n=2 samples. Each sample for RNA-Seq consisted of 300-800 cells per mouse, n=2-3 mice.

Supplementary Figure 5. Related to Figure 5.

(**a**, **b**) Flow cytometry analysis of CD45$^+$TCRβ^+CD4$^+$CD8b^{low} Tomato$^+$ cells in the IE and mLN of

Zbtb7b$^{fl/+}$x*Runx3*$^{fl/+}$x*Rosa26*$^{ls|tdTomato}$x*Foxp3*^{CreER} (iFoxp3), iFoxp3x*Runx3*$^{fl/+}$ (iFoxp3$^{(\Delta ThPOK)}$),

iFoxp3x*Zbtb7b*$^{fl/+}$x*Runx3*$^{fl/fl}$ (iFoxp3$^{(\Delta Runx3)}$), iFoxp3x*Zbtb7b*$^{fl/fl}$x*Runx3*$^{fl/fl}$ (iFoxp3$^{(\Delta ThPOK \Delta Runx3)}$) mice after 10

weeks of tamoxifen treatment. (**a**) Frequency of total Foxp3$^+$ cells among Tomato$^+$ CD4$^+$ T cells. (**b**)

Frequency of total CD8α^+ cells among Tomato$^+$ CD4$^+$ T cells. (**c**) Neuropilin-1 expression among Foxp3$^+$

Tregs in the intraepithelial compartment (IE) and mesenteric lymph nodes (mLN) of WT mice. Histogram

(left) and frequency (right). (**d-g**) nTreg (neuropilin-1$^-$) or iTregs (neuropilin-1$^+$) were sorted from spleens

and mLNs of CD45.2 *Zbtb7b*$^{fl/+}$x*Runx3*$^{fl/+}$x*Rosa26*$^{ls|tdTomato}$x*Foxp3*^{CreER} (iFoxp3), iFoxp3x*Runx3*$^{fl/+}$

(iFoxp3$^{(\Delta ThPOK)}$), iFoxp3x*Zbtb7b*$^{fl/+}$x*Runx3*$^{fl/fl}$ (iFoxp3$^{(\Delta Runx3)}$), iFoxp3x*Zbtb7b*$^{fl/fl}$x*Runx3*$^{fl/fl}$

(iFoxp3$^{(\Delta ThPOK \Delta Runx3)}$) mice after tamoxifen administration and co-transferred with CD45.1 naïve CD4$^+$ T

cells to *Rag1*$^{-/-}$ hosts. CD45.2$^+$TCRβ^+CD4$^+$CD8β^-Tomato+ lymphocytes from the IE and mLN were

analyzed 10 weeks after transfer. (**d**) Experimental layout. (**e**) Frequencies of total intracellular Foxp3 cells after nTreg (left) or iTreg (right) transfer. (**f-g**) Body weight of *Rag1*[-/-] recipients after nTreg (left) or iTreg (right) transfers (**f**) and levels of fecal lipocalin-*2* before and after per cell type transferred (**g**). Dashed lines represent colitis control transfer of naïve CD45.1[+]CD4[+] T cells only. Data are expressed as mean +/- SEM of individual mice (n=5-11 per genotype, 3 separate experiments). *p<0.05, **p<0.01, ***p<0.001 [one-way ANOVA and Bonferonni test (**a,b,e-g**) and student's t test (**c**)].
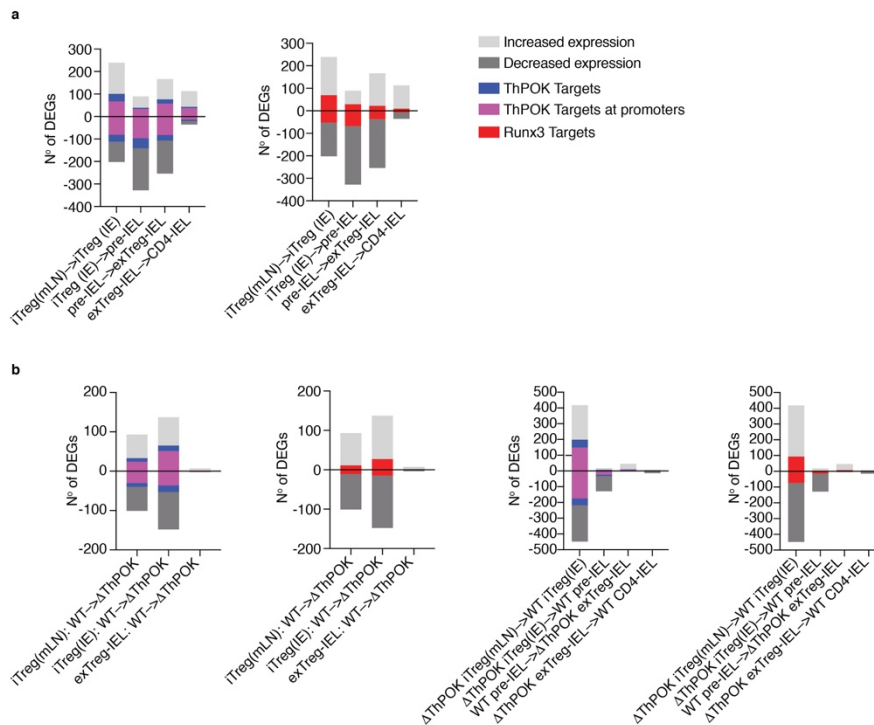
Supplementary Figure 6. Related to Figure 5.

(**a, b**) Induced Tregs (iTregs; Tomato$^+$CD8α^-neuropilin-1$^-$) and Treg-derived CD4-IELs (exTreg-IELs; Tomato$^+$CD8α^+) were sorted from *Zbtb7b*$^{fl/fl}$x*Runx3*$^{fl/+}$x*Rosa26*$^{ls|tdTomato}$ x*Foxp3*^{CreER} (iFoxp3$^{(\Delta ThPOK)}$) mice after 10 weeks of tamoxifen administration followed by RNA-Sequencing from IE or mLN. Numbers of differentially expressed genes (DEGs) (silver for increase, gray for decrease) with total putative ThPOK targets or ThPOK targets at promoters (blue and purple, respectively; left) or putative Runx3 targets (red, right) at indicated comparisons between WT cells only (**a**) and WT and ΔThPOK cells (**b**).

Supplementary Table 1. 10X genes per cluster, related to Figure 1.

Top significantly differentially expressed genes per cluster as determined by the Wilcoxon rank sum test. Columns indicate p values (p_val), average log fold change (avg_logFC), principle components 1 and 2 (pct.1and pct.2, respectively), adjusted p value (p_val_adj), cluster number and gene.

Supplementary Table 2. Levels of chromatin accessibility of WT CD4$^+$ T cells, related to Figure 3.

Differentially accessible chromatin regions per cluster, as determined by the likelihood ratio test. Columns indicate peak ID (IDs), cluster number, chromosome location, peak annotation, and correlated gene name, in order of genes visualized in heatmap of Figure 3a. n=2 per genotype.

Supplementary Table 3. Differentially accessible chromatin between WT CD4$^+$ T cells, related to Figure 3.

List of differentially accessible chromatin regions between indicated WT epithelial CD4$^+$ T cell populations, as assessed by the Wald pairwise test. Columns indicate peak ID (IDs), log fold change of accessibility (log2FoldChange), adjusted p value (padj), annotation and correlated gene name. n=2 per genotype, padj$_{sig}$<0.01.

Supplementary Table 4. Differentially expressed genes between WT CD4$^+$ T cells, related to Figure 3.

List of differentially expressed genes between indicated WT epithelial (IE) or mesenteric lymph node (mLN) CD4$^+$ T cell populations, as assessed by the Wald pairwise test. Columns genes (target_id), log2 fold change of expression (b) and adjusted p value (qval). n=2-3 per genotype, qval$_{sig}$<0.05.

Supplementary Table 5. Levels of gene expression of WT CD4$^+$ T cells, related to Figure 3.

Differentially expressed genes per cluster, as determined by the likelihood ratio test. Columns indicate gene (GeneID) and cluster number, in order of genes visualized in heatmap of Figure 3d. n=2-3 per genotype.

Supplementary Table 6. ThPOK binding regions in splenic Tregs, related to Figure 4.

Significant ThPOK binding sites in *in* vivo expanded splenic Tregs. Columns indicate binding region annotation, region location (chromosome, and start and end nucleotide position of region on indicated chromosome), peak score of regions, annotation of region, distance to transcriptional start site (TSS), and correlated gene name. Peak scores > 20 were considered to be significant.

Supplementary Table 7. Differentially accessible regions that overlap with ThPOK ChIP-Seq regions and are present near differentially expressed genes, related to Figure 4. List of ATAC-Seq peak regions that overlap with ThPOK ChIP-Seq peaks. The likelihood ratio test with a minimum adjusted-pvalue of 0.05 were used to detect 814 significant regions, which 38 were also present near differentially expressed genes.

Supplementary Table 8. Differentially expressed genes between WT and ∆ThPOK CD4[+] T cells, related to Figure 5.

List of differentially expressed genes between indicated WT and ∆ThPOK epithelial (IE) or mesenteric lymph node (mLN) CD4[+] T cell populations, as assessed by the Wald pairwise test. Columns genes (target_id), log2 fold change of expression (b) and adjusted p-value (qval). n=3 per genotype, $qval_{sig}$<0.05.

Supplementary Table 9. Levels of gene expression of WT and ∆ThPOK Tregs from the mesenteric lymph nodes and intestinal epithelium, related to Figure 5.

Differentially expressed genes per cluster, as determined by the likelihood ratio test. Columns indicate gene (GeneID) and cluster number, in order of genes visualized in heatmap of Figure 5g. n=3 per genotype.

Supplementary Table 10. Differentially accessible chromatin between WT and ∆ThPOK CD4[+] T cells, related to Figure 6.

List of differentially accessible chromatin regions between indicated WT and ΔThPOK epithelial CD4$^+$ T cell populations, as assessed by the Wald pairwise test. Columns indicate peak ID (IDs), log fold change of accessibility (log2FoldChange), adjusted p value (padj), annotation and correlated gene name. n=2 per genotype, padj$_{sig}$<0.01.