CellPress

Supplemental figures

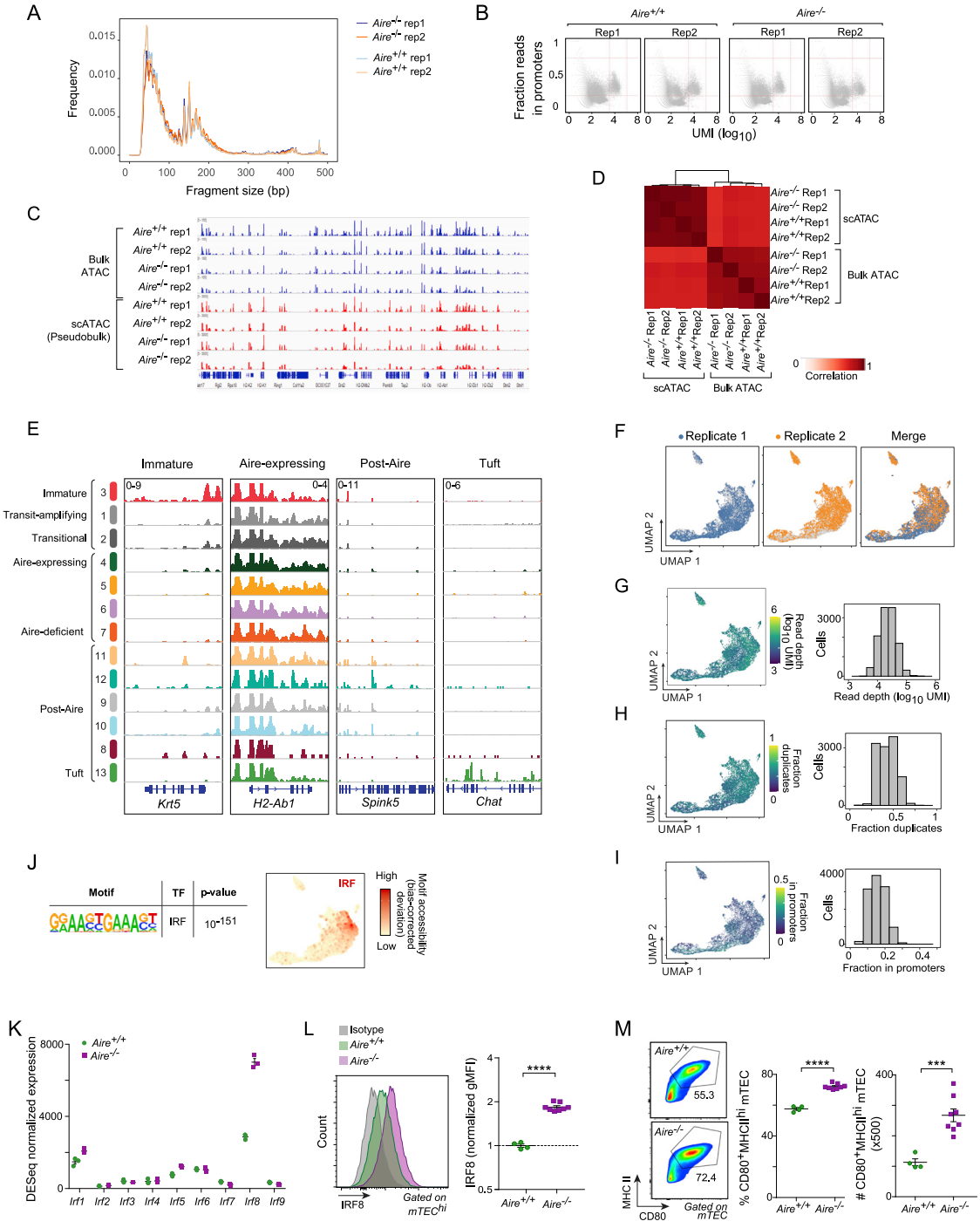

Figure S1. Quality control and additional analysis of scATAC-seq data from mTECs, related to Figure 1

(A) Fragment size distribution for scATAC-seq libraries generated from mTEC^{hi} from the thymi of *Aire*$^{+/+}$ and *Aire*$^{-/-}$ mice.

(B) Fraction of reads in promoters versus number of UMIs per cell barcode for each scATAC-seq library. Each dot is one cell barcode, and dashed red lines indicate the criteria for filtering high-quality cells.

(C and D) Genome tracks (C) and inter-sample Pearson correlation (D) comparing chromatin-accessibility signal in bulk ATAC-seq of mTEC^{hi} (Bansal et al., 2021) versus pseudobulk scATAC-seq of mTEC^{hi} (this study).

(legend continued on next page)

(E) Additional chromatin-accessibility tracks for mTEC clusters at marker genes. Signal is in CPM.
(F) Distribution of first (left), second (middle), and merged (right) scATAC-seq replicates, independent of genotype, in UMAP space.
(G–I) UMAP (left) and histogram (right) of read depth (G), fraction of duplicated reads (H), and fraction of reads in promoters (I).
(J) Cluster-level (left) and single-cell (right) TF-motif-enrichment analysis of scATAC-seq data for IRF family motifs.
(K) Normalized expression of IRF family member transcripts in *Aire*$^{+/+}$ (n = 3) and *Aire*$^{-/-}$ (n = 3) mTEC^{hi}, assayed by bulk RNA-seq. Each dot is one biological replicate, and bars show mean ± SEM.
(L) Representative flow plot (left) and summarized data (right) of IRF8 expression in *Aire*$^{+/+}$ and *Aire*$^{-/-}$ mTEC^{hi}, assayed by flow cytometry. Data are normalized to mean IRF8 geometric mean fluorescence intensity (gMFI) in mTEC^{hi} from *Aire*$^{+/+}$ mice.
(M) Representative flow plots (left) and summarized data (right) of the fraction and number of CD80^{+}MHCII^{hi} mTECs in the thymi of *Aire*$^{+/+}$ and *Aire*$^{-/-}$ mice. For (L) and (M), each dot is one mouse, n = 4 for *Aire*$^{+/+}$ and n = 8 for *Aire*$^{-/-}$, data were pooled from two independent experiments, bars show mean ± SEM, and p values were calculated by unpaired, two-sided Student's t tests.

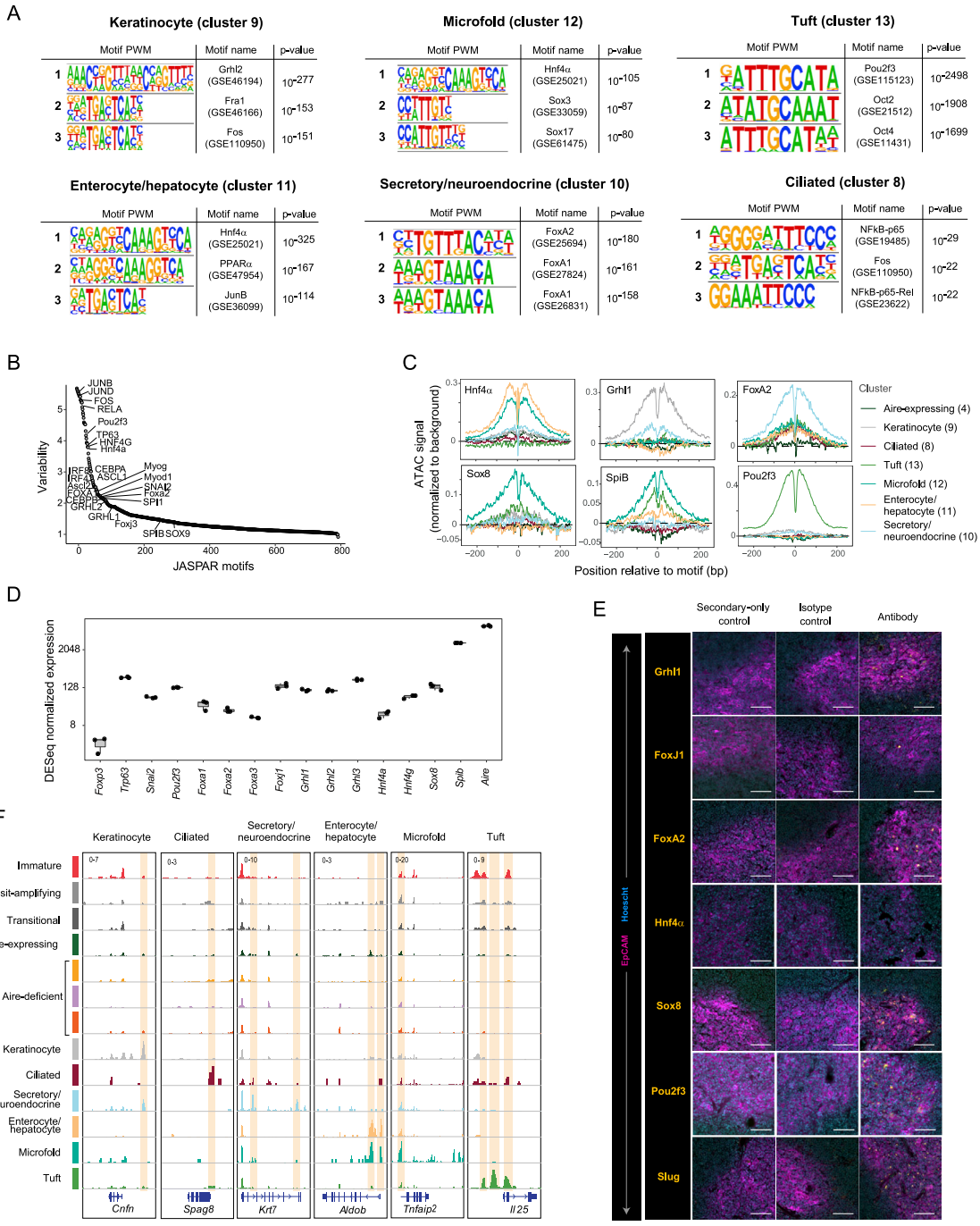

Figure S2. Additional analysis of TF-motif accessibility in mTECs, related to Figure 2

(A) Top motif enrichments for each post-Aire mTEC cluster, as detected by cluster-level TF-motif-enrichment analysis. Note that the secretory/neuroendocrine and ciliated clusters were originally called as one cluster; the joint enrichment p value for the combined cluster is shown in Figure 2.

(B) Variability of TF motifs across individual mTECs, ranked from most to least variable. Some key TFs are labeled.

(C) Cluster-level TF footprinting of scATAC-seq data. Signal is in arbitrary units and represents bias-normalized Tn5 insertions over background at TF footprints within the accessible peaks of each cluster.

(D) Expression of transcripts encoding the indicated TFs in mTEC^{hi} (n = 3), assayed by bulk RNA-seq. Each dot is one biological replicate.

(E) Immunofluorescence demonstrating the specificity of each of the anti-TF antibodies used in Figure 2. Scale bars, 200 μm.

(F) Additional chromatin-accessibility tracks for each of the mTEC clusters identified by scATAC-seq at the indicated loci. Signal is in CPM.

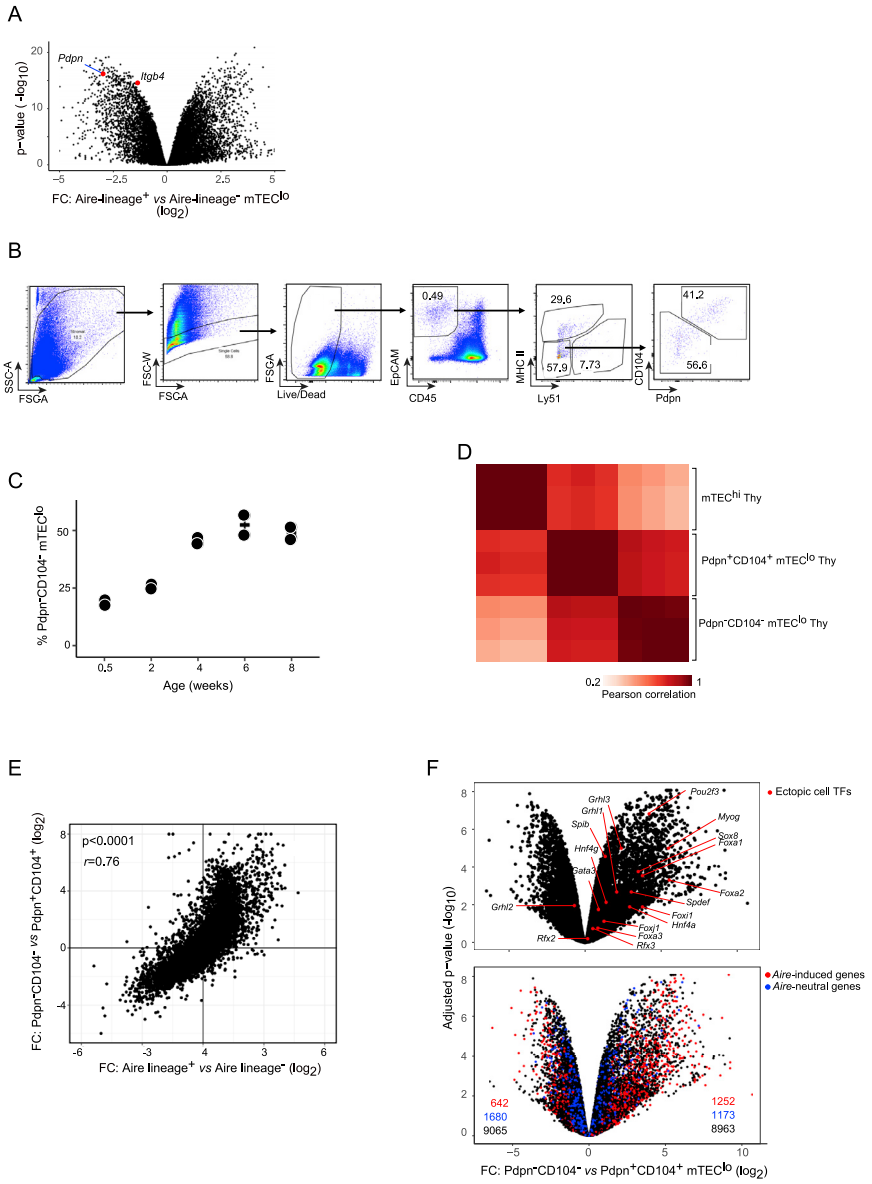

Figure S3. Additional analysis of bulk RNA-seq of post-Aire mTECs, related to Figure 3

(A) Volcano plot of bulk RNA-seq of Aire-lineage-positive mTEC^{lo} (n = 4) versus Aire-lineage-negative mTEC^{lo} (n = 4). Data were reanalyzed from Miller et al. (2018).

(B) Cytofluorometric gating scheme for analysis and isolation of Pdpn^{-}CD104^{-} mTEC^{lo}.

(C) Abundance of Pdpn^{-}CD104^{-} mTEC^{lo} at indicated mouse ages (n = 2 for each). Each dot is one mouse, and bars show mean ± SEM.

(D) Pearson correlation of bulk RNA-seq replicates of Pdpn^{+}CD104^{+} mTEC^{lo}, mTEC^{hi}, and Pdpn^{-}CD104^{-} mTEC^{lo} (n = 3 for each).

(E) FC/FC plot of Pdpn^{-}CD104^{-} mTEC^{lo} versus Pdpn^{+}CD104^{+} mTEC^{lo} compared with Aire-lineage-positive mTEC^{lo} versus Aire-lineage-negative mTEC^{lo}. *r*, Pearson correlation.

(F) Volcano plots of bulk RNA-seq of Pdpn^{-}CD104^{-} mTEC^{lo} versus Pdpn^{+}CD104^{+} mTEC^{lo}. Lineage-defining TFs identified by scATAC-seq (top) and Aire-induced and Aire-neutral genes (bottom) are highlighted. p values were BH-corrected.

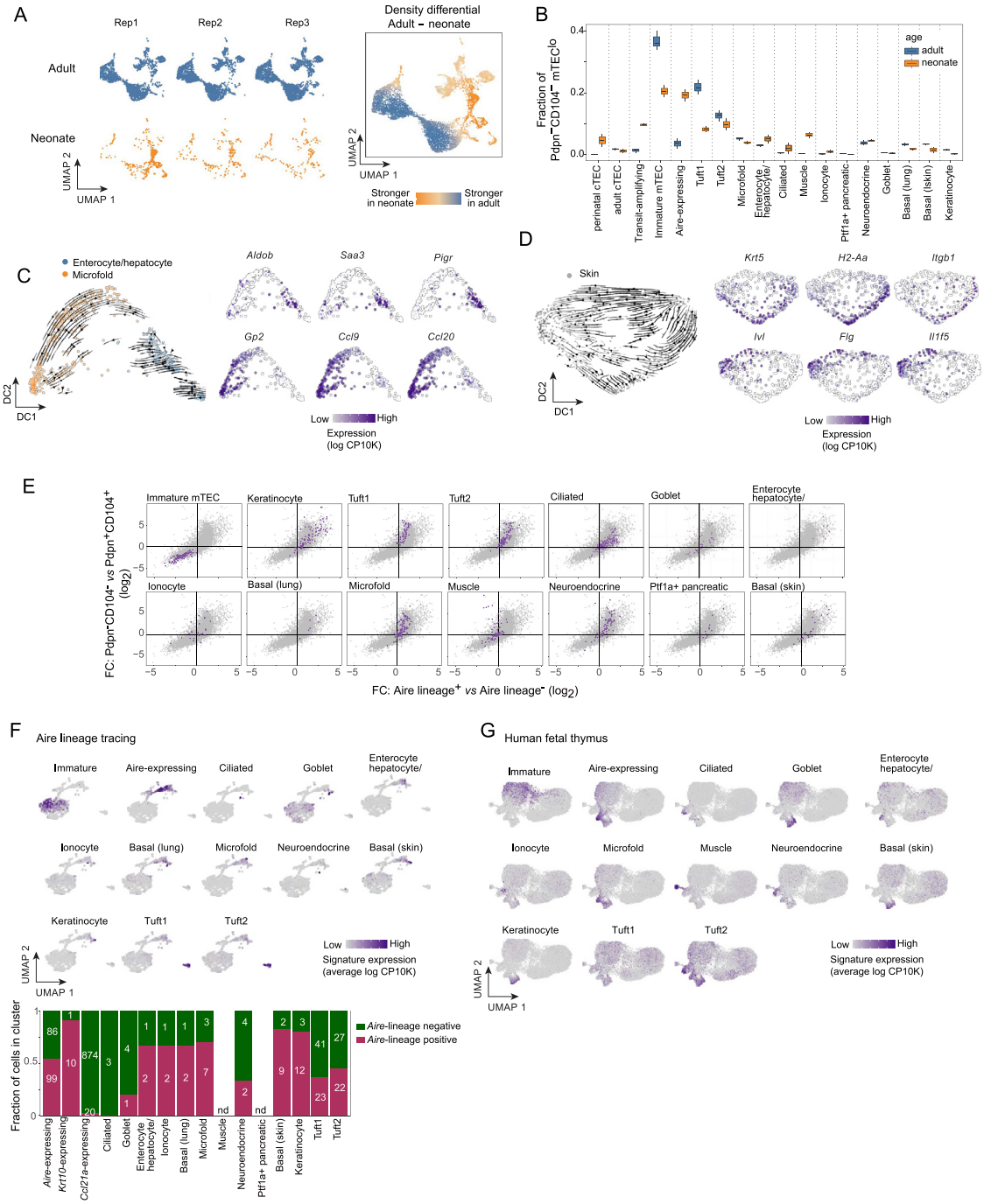

Figure S4. Additional analysis of scRNA-seq of post-Aire mTECs, related to Figure 3

(A) Per-replicate UMAPs of scRNA-seq data from perinatal and adult mTECs (left) and differential density UMAP comparing the distribution of perinatal and adult mTECs (right).

(B) Abundance of indicated mTEC subtypes relative to total Pdpn^{-}CD104^{-} mTEC^{lo} in perinatal (n = 2) and adult (n = 3) mice.

(C and D) RNA velocity analysis of enterocyte/hepatocyte and microfold mTECs (C) and basal skin and keratinocyte mTECs (D), showing velocity arrows representing the predicted trajectory of cell differentiation (left) and expression of marker genes (right). DC, diffusion component.

(E) FC/FC plots comparing Pdpn^{-}CD104^{-} mTEC^{lo} and Aire-lineage-positive mTEC^{lo} as in Figure S3E, with mimetic-cell gene signatures derived from scRNA-seq data overlaid in purple.

(F) UMAPs (top) of scRNA-seq data from Aire-lineage-traced mTECs, highlighting the expression of the indicated mimetic-cell gene signatures, and fraction of each mimetic cell subtype (bottom) that was Aire-lineage-positive or -negative.

(legend continued on next page)

(G) UMAPs of scRNA-seq data from human fetal thymus, with the expression of the indicated orthologue-converted mimetic-cell gene signatures overlaid. For (A) and (B), note that neonatal replicate #1 was excluded from differential density UMAP calculation and fraction quantification due to a slight difference in gating during sorting.

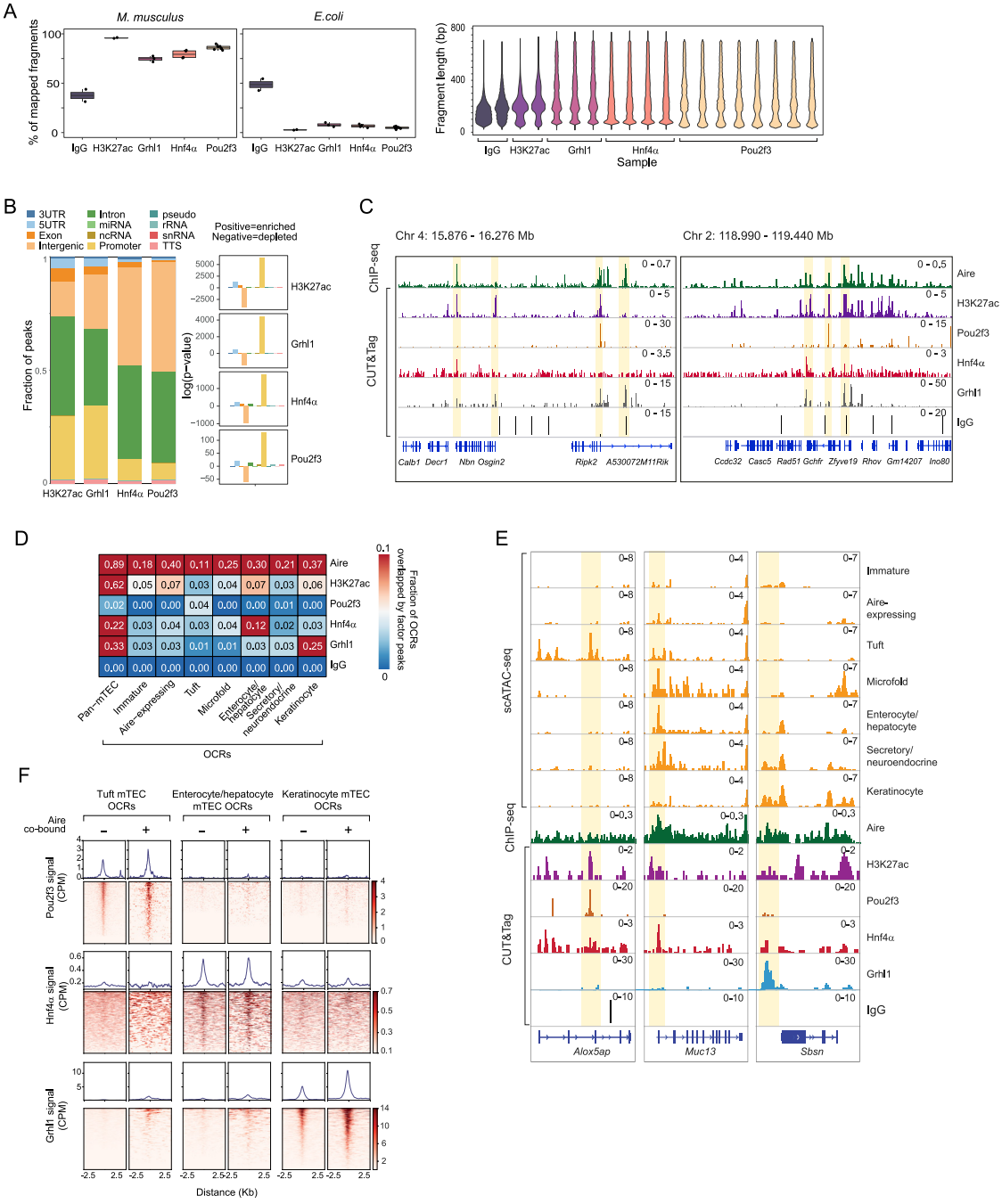

Figure S5. Additional analysis of CUT&Tag of mimetic-cell TFs, related to Figure 4

(A) Quality-control metrics for CUT&Tag samples, including the percent of fragments mapped to the *M. musculus* (left) and *E. coli* (middle) genomes, and the fragment-length distribution for each replicate (right). For fragment mapping plots, each dot is one replicate.

(B) Fractional representation (left) and enrichment (right) of indicated genomic features within binding peaks of the indicated factors. p values were calculated in HOMER.

(C) Genome tracks showing wide views of binding of the indicated factors.

(D) Heatmap of the fraction of each indicated set of OCRs that overlap binding peaks for the indicated factors.

(E) Additional genome tracks displaying chromatin accessibility in mTEC subtypes (yellow) and binding of the indicated factors at the indicated loci.

(F) Profile plots and heatmaps of Pou2f3, Hnf4α, and Grhl1 CUT&Tag signal at the indicated mimetic-cell-specific OCRs, stratified by Aire co-binding. For profile plots, mean signal is shown. For heatmaps, each row is one OCR. In (C), (E), and (F), signal is in CPM and was merged from two (Aire, H3K27ac, IgG), three (Grhl1), four (Hnf4α), or eight (Pou2f3) independent replicates.

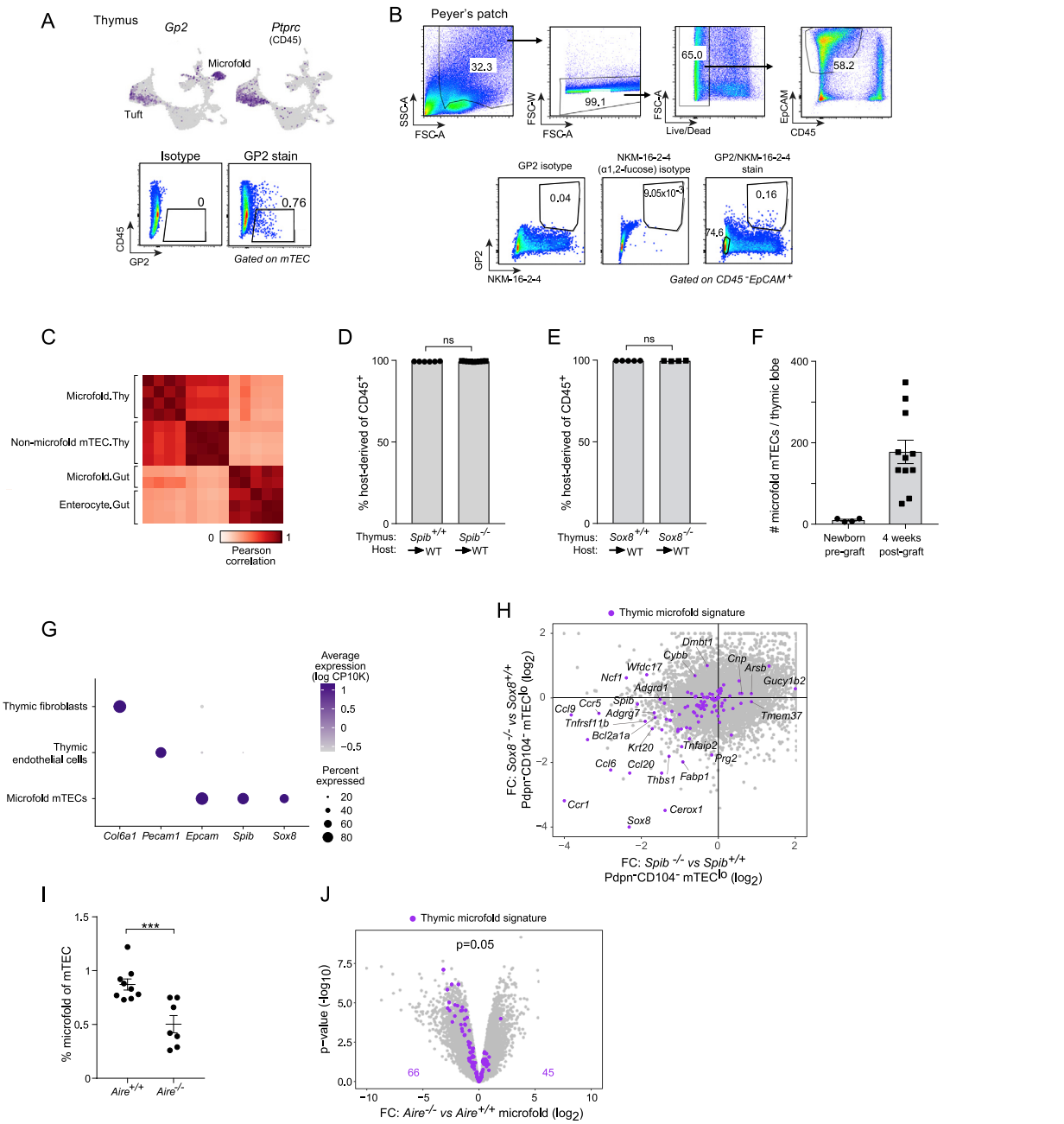

Figure S6. Additional characterization of microfold mTECs, related to Figure 5

(A) UMAPs of scRNA-seq data from mTECs (top) showing expression of *Gp2* and *Ptprc* (encoding CD45), and flow gating using these markers to identify microfold mTECs (bottom).

(B) Flow gating of Peyer's patches to isolate gut M cells.

(C) Pearson correlation matrix of bulk RNA-seq replicates from non-microfold mTECs, microfold mTECs, gut M cells, and enterocytes.

(D and E) Fraction of host-derived (CD45.1$^+$) cells among total hematopoietic cells in thymi from (D) *Spib*$^{+/+}$ and *Spib*$^{-/-}$ or (E) *Sox8*$^{+/+}$ and *Sox8*$^{-/-}$ donor mice 4 weeks after grafting into WT hosts.

(F) Number of microfold mTECs in individual thymic lobes from newborn pre-graft and 4-week post-graft WT thymi. Each dot is one thymic lobe, and data were pooled from 5 independent experiments.

(G) Expression of the indicated transcripts in thymic fibroblasts, thymic endothelial cells, and microfold mTECs, assayed by scRNA-seq.

(H) FC/FC plot comparing gene expression in *Sox8*$^{-/-}$ and *Spib*$^{-/-}$ Pdpn$^-$CD104$^-$ mTEC^{lo} versus WT controls.

(I) Fraction of microfold mTECs in thymi from *Aire*$^{+/+}$ and *Aire*$^{-/-}$ mice. Each dot is one mouse, and data were pooled from three independent experiments. p value was calculated by unpaired, two-sided Student's t test.

(J) Volcano plot of bulk RNA-seq of purified GP2$^+$CD45^{neg} mTECs from *Aire*$^{-/-}$ versus *Aire*$^{+/+}$ mice. Signature p value was calculated by chi-square test. In (H) and (J), the microfold signature is highlighted in purple. For (D)–(F) and (I), bars show mean ± SEM.

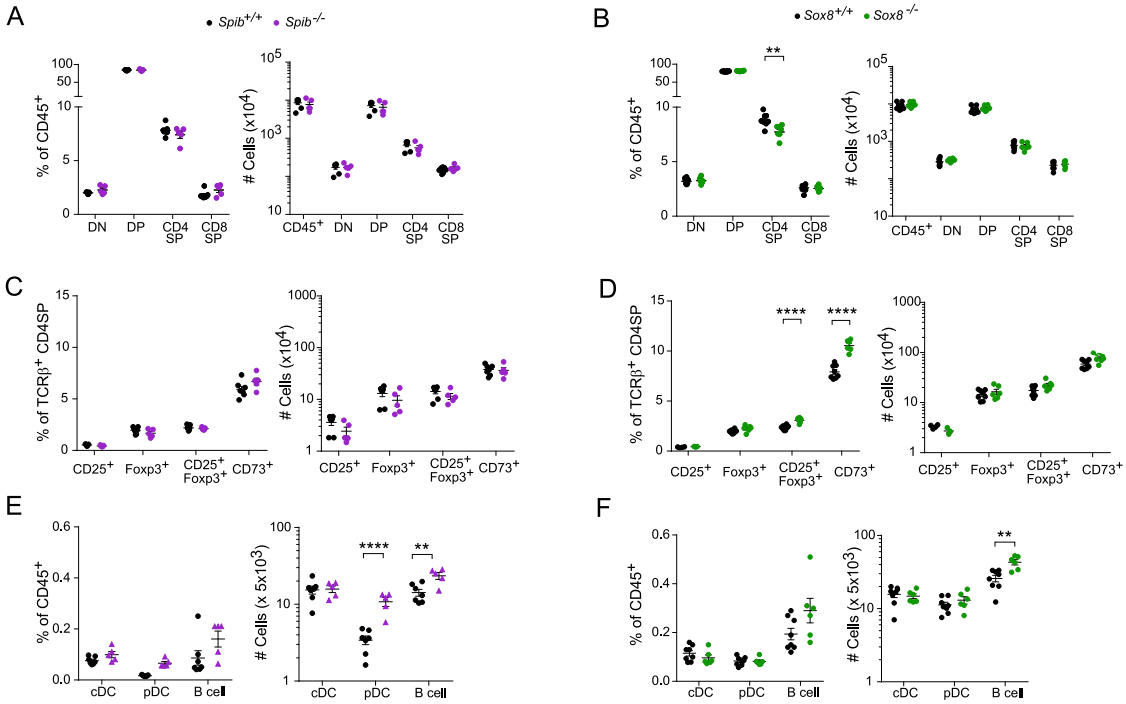

Figure S7. Hematopoietic cells in *Spib*$^{-/-}$ and *Sox8*$^{-/-}$ thymi, related to Figure 5

Fraction (left) and number (right) of (A and B) T cell, (C and D) CD4 single-positive T cell, and (E and F) B cell (CD19^{+}B220^{+}) and dendritic cell subsets in thymi from *Spib*$^{+/+}$ and *Spib*$^{-/-}$ mice (A, C, and E) and from *Sox8*$^{+/+}$ and *Sox8*$^{-/-}$ mice (B, D, and F). DN, CD4/CD8 double negative; DP, CD4/CD8 double positive; SP, single positive; cDC, classical dendritic cell (CD11c^{+}B220^{-}); pDC, plasmacytoid dendritic cell (CD11c^{+}B220^{+}). Each dot is one mouse, bars show mean ± SEM, p values were calculated by unpaired, two-sided Student's t tests, and for both genotypes, data are pooled from two independent experiments.