TITLE

Single cell transcriptomics of human prenatal anterior foregut-derived organs identifies distinct developmental signatures directing commitment and specialization of the thymic epithelial stroma

AUTHORS

Abdulvasey Mohammed[1,*], Benjamin Solomon[2,*], Priscila F. Slepicka[1], Kelsea M. Hubka[1], Hanh Dan Nguyen[1], Michael G. Chavez[3], Christine Y. Yeh[4], Virginia D. Winn[5], Casey A. Gifford[6], Purvesh Khatri[7,8*#], Andrew Gentles[9*#], and Katja G. Weinacht[1*#]

1. Department of Pediatrics/Division of Hematology, Oncology, Stem Cell Transplantation, and Regenerative Medicine, Stanford University School of Medicine, Stanford, CA, USA
2. Department of Pediatrics, Stanford University School of Medicine, Stanford, CA, USA
3. Department of Bioengineering, Stanford University, Stanford, CA, USA
4. Department of Biomedical Data Science, Stanford University, Stanford, CA, USA
5. Department of Obstetrics and Gynecology, Stanford University School of Medicine, Stanford.CA, USA
6. Department of Pediatrics/Division of Cardiology, Stanford University School of Medicine, Stanford, CA, USA
7. Institute for Immunity, Transplantation and Infection, Stanford University School of Medicine, Stanford, CA, USA
8. Center for Biomedical Informatics Research, Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA
9. Department of Pathology, of Medicine (Quantitative Sciences Unit), and Biomedical Data Sciences. Stanford University School of Medicine, Stanford, CA, USA

*These authors contributed equally to this work
#Corresponding authors

Correspondence should be sent to Katja G. Weinacht (kgw1@stanford.edu) and Andrew Gentles (andrewg@stanford.edu).

ABSTRACT
The thymus is the primary lymphoid organ that instructs adaptive T cell immunity and central tolerance. Within the thymus, the thymic epithelium comprises a diverse and highly specialized set of cells that promote T cell maturation, proliferation, and selection of a diverse and self-tolerant T cell receptor repertoire. As such, the function of the thymic epithelium is central to the integrity of the immune system in health and disease. Accordingly, regenerating thymic function for therapeutic purposes holds great promise. However, the gene regulatory networks that drive thymic epithelial cell ontogeny and maintenance during human embryonic development remain incompletely understood. Elucidating the complex interplay between epithelial cell intrinsic and environmental signals that contribute to thymic morphogenesis is essential for our quest to derive thymic epithelial cells from pluripotent stem cells to reconstitute thymic function for clinical purposes.

To deduce the signals instructing the development and specialization of the human thymic epithelial stroma, we have used a dual-trajectory single cell transcriptomic approach: (A) To delineate the trajectory from anterior foregut towards thymic fate and dissect alternate lineage bifurcations, we have compared transcriptional signatures of epithelial cells from human fetal thymus to epithelial cells derived from the other major anterior foregut-derived organs, i.e., bronchus, lung, and esophagus as well as the parathyroid. (B) To define the proliferative dynamics that give rise to developmental hierarchies within specialized thymic epithelial compartments, we have compared human thymus samples sequentially during embryonic, fetal, and early postnatal stages of organogenesis.

Distinct gene regulatory networks delineate cortical, and medullary thymic epithelial cells, conducting airway basal cells, respiratory bud-tip progenitor cells and esophageal basal cells. Activation of the NFKB/Interferon response axis, it the prevailing distinguishing feature that sets the thymic epithelium apart from all other anterior foregut derived organs. This correlates with highest interferon gamma protein expression by thymocytes in the thymic medulla. Unique expression patterns for the transcription factors SOX2 and TP63, and the relative expression of retinoic acid producing, and metabolizing enzymes are also observed. Within the thymic epithelial compartment, a cycling progenitor pool gives rise to three distinctive developmental trajectories. Cortical TECs and medullary TECs show distinctive lineage trajectories originating from a cluster of predominantly cycling TECs as early embryonic stages. Compared to medullary TECs, cortical TECs show increased Histone Deacetylase 4 expression paired with decreased levels of histone acetylation. Our studies trace thymus specific gene regulatory networks within the larger framework of epithelial cells from all major anterior foregut derived organs with the goal to provide insights into actionable pathways that help advance directed differentiation approaches for translational applications.

INTRODUCTION
Recent studies of murine and human thymus using single cell transcriptomics have dramatically transformed our understanding of thymic epithelial cell diversity[1-6]; yet the signals that drive thymic epithelial cell fate during human foregut development remain incompletely understood. The anterior part of the mammalian foregut gives rise to the ventrally located respiratory system and the dorsally located esophagus which extends into the stomach[7]. The anterior-most segment of the foregut develops into the pharyngeal apparatus. Essential to mammalian pharyngeal development is the out-pocketing of the endoderm to form the pharyngeal pouches[8].

The thymus is derived from the ventro-caudal aspect of the third pharyngeal pouch (PPEIII) (Figure 1A). Two developmentally separate steps have been implicated in thymic organogenesis in murine studies. Initially, the development of the Thymic Anlage from the PPEIII is instructed by signals from the surrounding mesenchyme. In the second phase, bone marrow-derived lymphoid precursors migrate to the thymic primordium. Reciprocal signals between cells of hematopoietic and thymic epithelial origin promote thymic organogenesis and T cell maturation, respectively[9-12]. The earliest detection of the PPEIII transcriptional regulators TBX1, HOXA3, PAX1 and PAX9 in human embryonic samples was demonstrated at the beginning of week 6 (CS16) and preceded the expression of master thymic transcription factor FOXN1 mid-week 6 (CS 16-17)[13]. These findings are concordant with murine data in which Foxn1 is required cell-autonomously for maturation of the thymic epithelium but not for initiation of thymus organogenesis[9,14,15]. Lymphoid progenitors enter the human thymic anlage as early as week 7[4,13] and are indispensable for epithelial cell maturation and expression of functional markers, such as PSMB11 or AIRE[16-18]. Parallels between branching morphogenesis of the airway and the initiation of cortical and medullary separation in the thymus have been proposed. In both settings, a homogenous epithelial cell population begins to polarize and display distinct apical-basal features. Epithelial cells in the center of the thymic anlage gain CLDN3/4 expression and proceed to form the medulla, however, a lumen does not form[19,20].

Important insights into the developmental programs that instruct thymic organogenesis were derived from murine knock out studies in *Foxn1* which resulted in an athymic and hairless (nude) phenotype[21]. Interestingly, the thymic development of nude mice arrests after Cldn3/4 is expressed but before colonization with lymphoid precursors occurs[15,20]. The thymic rudiment of *Foxn1* knock-out mice displays nonfunctional cystic features strongly resembling respiratory epithelium in both, morphology, and gene expression[9,22-24]. It has therefore been suggested that Foxn1 induces thymic epithelial cell commitment at the expense of respiratory epithelial cell fate as the default pathway by inhibiting branching morphogenesis [20,22,23].

Patterning of the esophagus, the other major anterior foregut-derived organ, occurs in tight coordination with that of the respiratory system[25] and any deviations from the default program threaten appropriate separation between both organ systems[26].

The signals that drive anterior foregut development from pluripotent cells along with those that induce lung and esophageal commitment from anterior foregut endoderm have been successfully established[27-32]. The developmental programs that define patterning of the third pharyngeal pouch and thymus from the anterior foregut endoderm, however, remain largely elusive. We sought to close this knowledge gap by dissecting the lineage branch points between human fetal thymus and other organs that develop in close temporo-spatial proximity from the anterior foregut, i.e., bronchus, lung, esophagus and parathyroid, using single cell transcriptomics. Next, we defined how thymic epithelial cell development progresses and specializes once thymic commitment has occurred by comparing human thymus samples sequentially during embryonic, fetal, and early postnatal stages of organogenesis.
Our studies build upon recent advances defining mouse and human thymic microenvironment and architecture using single cell transcriptomics[1-6,33]. These studies demonstrated an abundance of novel and unique thymic cell subpopulations which had evaded discrimination by conventional analytic metrics. Our work specifically focuses on understanding differential developmental trajectories between closely related anterior foregut derived epithelial cell populations with a view towards advancing directed differentiation and tissue regeneration approaches.

RESULTS

Human fetal anterior foregut derived epithelial cells at single cell resolution

We analyzed a total of 16 samples from 14 biological specimens from human embryonic, fetal and postnatal tissues estimated to range from post conception week (pcw) 7 to 3 months post birth. For the comparison of the major anterior foregut derived organs, we sampled tissues from human fetal thymus (pcw 16, 17 and 18), airway/bronchus, lung parenchyma (pcw 16, 17 and 18) and esophagus (pcw 16 and 17). To examine the longitudinal progression and diversification of the thymic epithelium, we studied very early human thymic samples (pcw 7 and 8), in comparison to mid-gestation (pcw 16, 17, 18) and neonatal (age 3 months) tissue (Figure 1A, Supplementary Table S1*). EPCAM positive selection with magnetic beads was used for all but the earliest thymic samples to enrich for cells of epithelial origin while preserving fractions of all other subpopulations of the microenvironment. In addition, thymic samples were partially depleted of CD45+ thymocytes (Figure 1B). The very early (pcw 7, 8) thymus samples were used unmanipulated due to very small numbers of thymocytes at this stage of development.

After genome alignment, and filtering for mitochondrial content and number of genes/reads per cell, data were normalized using regularized negative Binomial regression in Seurat[34]. Cells from different samples were then combined using integration anchors[35], resulting in a combined dataset of 52,162 single cell transcriptomes with a mean of 3522 unique genes per cell, and mean of 14,839 total counts per cell. For the comparison of anterior foregut derived organs, post conception week 16, 17 and 18 samples of human thymus, bronchus, lung, and esophagus were organized by performing PCA dimensionality reduction on the 3000 most variable genes followed by unsupervised Louvain clustering on the 50 top principal components and visualized using t-SNE. Samples were annotated in concordance with nomenclature employed in previously published organ-specific human single cell datasets for thymus[4], respiratory system[36,37] and esophagus[38], respectively.

Within the thymus, we identified a total of 17 epithelial and non-epithelial cell populations (Figure 1C). Epithelial cells further subdivided into canonical *CCL25, PSMB11, KRT8-* high cortical thymic epithelial cells (cTEC) and *CCL19*^{high}, *KRT5/14*^{high} medullary thymic epithelial cells (mTECI). The previously described population of TECs with overlapping cTEC and mTEC features termed mcTECs[4] with uniquely high expression of *CTGF* and *DLK2*, was the third most abundant population visualized. Classical mTECII were characterized by high transcript levels of *AIRE* and *CLDN4* (Figure 1D). The recently described medullary TEC subpopulations mTECIII and mTECIV were also visualized. These cells are characterized by expression of *KRT1* and *POUF3* respectively, which correspond to the corneocyte-like post-AIRE mTECs and Tufft cells in the mouse[2,3,39]. TECmyo and TECneuro were defined by *NEUROD1*, *CHGA*, *CHGB*, *MYOD1* and *CHRNA1* expression, respectively (Figure 1D). Nonepithelial clusters included double negative and double positive thymocytes, as well as small numbers of single positive T cells (*PTPRC*, *CD3D*, *CD4, CD8A, CD8B*), macrophages (*S100A9, RETN*), dendritic cells (*LAMP3*), vascular smooth muscle cells (*PDGFRB, RGS5*), aldehyde-dehydrogenase positive fibroblasts, also described as thymic fibroblasts 1[4] (*PDGFRA, ALDH1A2*), fibronectin-high fibroblasts also described as thymic fibroblasts 2[4] (*FBN1*), vascular endothelial cells (*PECAM1, CDH5, AQP1*), lymphatic endothelial cells (*LYVE1, PROX1*) and neuronal cells (*NRXN1, MPZ*).

Tissues of the bronchus clustered into conducting airway epithelial basal cells (*KRT5, KRT15, TP63*), club cells (*SCGB1A1, BPIFA1*), multi-ciliated cells (*FOXJ1, PIFO*) and submucosal gland cells (*LTF*), along with goblet cells (*MUC5B, BPIFB1*), secretory progenitor cells (*SFTPB, SCGB3A2*) and bud tip

progenitor cells from the adjacent respiratory parenchyma (*SFTPC, SOX9*). In addition, the airway microenvironment was comprised of immune cells (B cells, T cells), myoepithelial cells, smooth muscle, vascular smooth muscle, fibroblasts, cartilage, endothelial cells, neuroendocrine and neuronal cells.

Lung tissues expectedly shared many of the same cell types surrounding the airways albeit in reverse ratios. Bud tip progenitors (*SFTPC, SOX9*) and bud tip progenitor adjacent cells (*AGER, SFTPB*) were the most abundant, followed by multiciliate cells (*FOXJ1, PIFO*), secretory progenitors (*SFTPB, SCGB3A2*) and very small numbers of epithelial basal cells (*KRT5, KRT15, TP63*). Cell types not derived from the foregut included neuroendocrine cells, immune cells (B cells, T cells, macrophages), fibroblasts, mesothelial cells, vascular smooth muscle cells, smooth muscle cells, endothelial cells, lymphatic endothelial cells and neuronal cells.

The esophageal tissue included esophageal basal cells (*KRT5, KRT15, TP63*), suprabasal cells (*KRT13, KRT4, CLDN4*), multiciliate cells (*FOXJ1, PIFO*), gastric epithelial cells (*TFF1, MUC5B*) and non-foregut-derived cell types including immune cells (T cells, macrophages), smooth muscle cells, vascular smooth muscle cells, two different types of fibroblasts, endothelial cells and a sizable cluster of neuronal cells from the vagal nerve that was accidentally included during sample dissection (Figure 1C, D, E; Supplementary Figure 1).

Integration of the individual data sets from thymus, bronchus, lung and esophagus followed by batch correction confirmed representation of all four tissues within all major cell types (Figure 1F, top). The clustering was defined by shared cellular identity regardless of organ system of origin, sharply separating epithelial cells (red) from mesenchymal cells (purple), immune cells (light blue), endothelial cells (green), myoepithelial cells (light pink) and neuronal cells (gray) (Figure 1F, middle). Thymic epithelial subpopulations displayed distinctive clustering within the epithelial cell population (Figure 1F), bottom. Top differentially expressed genes expressed broadly across all anterior foregut-derived epithelial cells included a select group of keratins (*KRT8, 17, 18, 19*), e-cadherin (*CDH1*), desmoplakin (*DSM*), the calcium binding proteins *S100A11* and *S100A14,* that can interact with a variety of cytoskeletal structures[40], the tight junction associated protein claudin 4 (*CLDN4*), the transcription factors *ELF3* and *KLF5*, and interestingly, *LMO7*, a negative regulator of transforming growth factor (TGF) that has been implicated in maintaining cell fate and counteracting fibrosis[41](Figure 1G)

Thymus-specific transcriptional factor regulons show unique interferon response network absent in other foregut-derived epithelial cells

To better understand the differences between anterior foregut derived epithelial cells, we computationally isolated and reintegrated the epithelial clusters only (Figure 2A). Expectedly, epithelial cells derived from the same organ system displayed closer dendrogram relationships than those of different organs, as did cells of shared function (i.e., multiciliate cells of bronchus and esophagus) (Figure 2B). Top differentially expressed genes (shown for select epithelial cell subsets) validated known key effector functions, such as lympho-attractant and peptide processing properties in cTECs (*CCL25, TBATA, PRESS16, CTSV),* and *s*urfactant production for bud tip progenitor cells of the lung (*SFTA3, SFTPB, SFTPC),* respectively. Notably, mcTECs showed differentially high expression for the chemokine *CCL2* and the hematopoietic stem cell maker *THY1* (also known as CD90) while mTECIs showed a differentially high expression for various interferon-inducible genes including *IFI6, IFIT3, IFITM3, LTB, MX1*and *ISG15* (Figure 2C).

To interrogate the function of transcription factors (TFs) and shed light on the distinctive developmental programs that underlie lineage fate decisions emerging from the anterior foregut, we implemented the single cell regulatory network inference and clustering (SCENIC) algorithm[42]. This process combines gene correlations with binding motif analysis within individual cells to infer regulons that correspond with co-expressed target genes. This is particularly critical in the thymus to distinguish low abundance TFs with important developmental functions from those that are promiscuously expressed genes. Top differentially expressed TF regulons across anterior foregut derived epithelial cells with highest expression in the thymus included a remarkable concentration of TF regulons that are were linked to fundamental inflammatory response pathways. Out of the top 23 differentially expression TF regulons, more than half were key regulators of interferon (IFN) responses (*IRF1, IRF7, IRF8, IRF9)*, Signal Transducer and Activator of Transcription (*STAT5A, STAT3, STAT1*), Nuclear Factor- - *NFKB1, NFKB2, RELB)*, and other inflammatory signaling pathways*,* including *RFX5*, a modulator of MHCII expression[43], and *BHLE40*, which is emerging as a key regulator of immunity during infection, autoimmunity, and inflammatory conditions[44](Figure 2D). To interrogate the relationship of individual TF regulons within the global TF network and within different organ systems, we identified the 250 most variable expressed regulons and clustered them based on coactivation using the Leiden clustering algorithm (Figure 2E). Uniform Manifold Approximation and Projection for Dimension Reduction (UMAP) visualization of the TF regulon network in its entirety not only showed a characteristic demarcation of individual regulons based on tissue of origin but distinctly separated the thymus in contrast to the three other organ systems (Figure 2F). Further unsupervised clustering revealed 9 subnetworks with dedicated organ and cell type specificity (Figure 2G). There were two thymus specific subnetworks (A and B). Network A largely defined the primary canonical TEC populations cTEC, mcTEC and mTECI, as well as the smaller medullary population mTECII, III and IV while subnetwork B was TECmyo specific.

Subnetwork A was the primary driver behind the previously noted thymus-specific high expression of inflammatory response pathway regulons, connecting the interferon response factor regulons *IRF1, IRF7, IRF8, IRF9* with an even larger number of STAT family regulons (*STAT5A, STAT3, STAT2 and STAT1),* NF- family subunit regulons *NFKB1, NFKB2* and *RELB,* and the NF- *SPIB*[45]*, HIVEP1*[46]. Interferon (IFNs) are a family of cytokines known to induce cell intrinsic programs with antiviral, immunomodulatory and antiproliferative properties[47]. There are two main classes of interferon: type I, also known as IFN and IFN , and type II, also known as IFN [47]. Upon receptor binding, signal transduction occurs through the JAK-STAT pathway. Both types of IFN can activate common and distinct STAT complexes, which regulate the transcription of target genes[48]. It is well established that IFNs induce the expression of hundreds of genes, which mediate various biological responses[49]. Some genes are regulated by both classes of IFNs, while others are specific such as the IFN-regulatory factor 1 (*IRF1*) gene which is preferentially induced by IFN [50] . In addition, IFN-activated signaling pathways can lead to NF- , in turn, NF- can modulate the cellular response to IFNs[51]. These findings raise the possibility that interferon (IFN) release is the inciting event behind the observed gene regulator network activation.

Of note, *ASCL1* was also part of the thymus-specific interferon response centric subnetwork. ASCL1 is an achaete-scute family basic helix-loop-helix transcription factor. In the context of the ectoderm development, *ASCL1* is necessary and sufficient to induce neuronal lineage progenitors[52]. When expressed in already committed neural progenitor cells, *Ascl1* promotes cell cycle exit and neuronal maturation[53]. *ASCL1* expression has previously been identified in human TEC subsets, albeit fate mapping experiments in *Ascl1*-reporter mice yield definitive conclusions on the role of *Ascl1* during thymic epithelial lineage commitment and specification[5].

Medullary thymic epithelial cells demonstrate distinctive interferon type 1 and 2 response signature.

To corroborate these findings using the SCENIC algorithm by orthogonal means and infer signaling pathways that could direct fundamental lineage commitment within the anterior foregut, we performed Gene Set Enrichment Analysis (GSEA) restricted to the primary, organ system-defining epithelial subsets cTECs, mTECI, bronchial epithelial basal cells, esophageal epithelial basal cell, esophageal suprabasal cells and bud tip progenitor cells of the lung (Figure 3A). Network visualization of positively enriched gene sets across select epithelial cell subsets revealed three distinct clusters of nodes, representing a cTEC-predominant thymic epithelial cluster, an mTECI-predominant thymic epithelial cluster, and a combined cluster of esophageal (basal and suprabasal) and bud tip progenitor cell identity in which nodes were divided. Clusters of cTEC and mTECI predominance shared abundant connections between nodes. Basal cells of bronchial origin were spread apart with minimal spatial concentration (Figure 3B). Gene sets with highest enrichment score in cTECs related to TCR signaling, protein processing by the proteasome, antigen cross presentation, NF- activation, IFN and interleukin (IL)-1 signaling. In mTECIs, both IFN and IFN signaling as well as antigen presentation were the leading gene sets with highest overall enrichment score. Shared gene sets between bud tip progenitor and esophageal basal epithelial cells related to ribosomal function, highlighting the importance protein synthesis plays in these cells. Esophageal suprabasal cells were characterized by genes defining keratinization (Figure 3C).

To investigate shared developmental drivers between cTEC and mTECI subsets that may be concealed by the dominant effect of highly expressed individual genes or gene subsets, we removed gene sets related to proteasome structure and function from the analysis. Even after removal of the proteasome gene sets, the IFN and IFN signaling modules remained uniquely enriched in the mTECI compartment. Leading edge genes with the greatest abundance in expression in the mTECI subset included direct effectors of interferon type I and II signaling including *IRF1, IRF2, IRF3, IRF4, IRF7, IRF8, IRF9, IFI6, ISG15, IFI27, BST2, IFITM1, IFITM3*, *MX1* and *MX2* as well as known interferon signal transduction mediators STAT1 and STAT2. In cTECs, genes implicated in interferon response and signal transduction such as *IRF1, IRF8, STAT1, JAK1* and *TYK2* were also highly expressed (Figure 3D). The IFN receptor genes *IFGR1* and *IFGR2* showed increased expression in both mTECI and cTEC subsets. Leading edge genes in the cTEC subset were largely comprised of MHC class I and II alleles implicated in antigen presentation.

While activated T cells are widely recognized as producers of interferon, less is known about the role interferons play during human thymopoiesis. Tonic type I IFN and NF- signaling during late-stage T cell development have been described in the murine thymus[54]. Expression of *IFNA1, IFNA2, IFNB1, IFNG as well as IFNAR1, IFNAR2* transcripts largely evaded the level of detection across subsets of hematopoietic, epithelial and non-epithelial stromal origin (Supplementary Figure S2). This is a recognized challenge in the field as *IFN* mRNA expression on a single cell level is extremely low. Interestingly, recent integrated single cell transcriptomic studies focusing on T lineage trajectories of CD34-positive thymus seeding hematopoietic progenitor cells found that *IRF7* and *IRF8* expression defined early thymus seeding progenitors and T-lineage specified precursors [55,56].

To establish a unified measure of IFN response across our select group of organ system defining epithelial subsets, we integrated Reactome curated type I and type II IFN specific gene sets into type I and type II IFN-response modules. The mTECI subpopulation demonstrated highest integrated scores

from both, the IFN type I and II response models, respectively. cTECs also showed significantly increased IFN type I and II response scores compared all other select epithelial subsets. IFN response scores in epithelial subpopulations of bronchial, esophageal and lung origin were extremely low, consistent with virtually absent expression of any of the included genes (Figure 3E, right panel). Interferon type I and II response scores in both TECs and mTECI subpopulations increased with advanced gestational age. (Figure 3E, left panel) This trajectory parallels the expansion of T cell progenitors within the thymic epithelial stroma over time.

To determine from which cell within the thymic microenvironment interferon production originates, we visualized IFN protein using immunohistochemistry (IHC) staining on gestational week 16 human fetal tissue slices. IHC staining demonstrated widespread intracellular IFN protein across most cells within the thymic microenvironment, with slight but noticeable predominance of the IFN protein signal in the thymic medulla (Figure 3F, panel 1). Although most cells within the thymic microenvironment at this stage in development are T cell precursors (Supplementary Figure S3), intracellular IFN protein was also clearly detected in thymic epithelial cells (Figure 3F, panel 2, 3, 4). These findings could either be consistent with endogenous *IFNG* expression by thymic epithelial cells or by IFN protein uptake through *IFNR1* receptor-mediated endocytosis and nuclear import via NPI-1 [57-59]. IFN protein was not visualized in fetal bronchus, lung and esophagus, corroborating our transcriptional findings.

SOX2 and TP63 show divergent expression pattern in thymic epithelial cells

Although parallels between branching morphogenesis of the airway and the initiation of cortical and medullary separation of the thymus have been proposed[19,20], the question stands unanswered, and any implicated gene regulatory networks remain elusive. SOX2 and TP63 are the master regulators of branching morphogenesis. During development of the respiratory system, the cycling expression of SOX2 and TP63 promotes forking of the airways. In the mouse *Sox2* is highly expressed in non-branching and absent from branching regions, suggesting that downregulation of Sox2 is necessary for airway epithelium to respond to fork-inducing signals[60]. *Sox2* expression was shown to increase again after the branching process and directly induced the expression of *Tp63*, reinstituting basal epithelial cell fate[61]. In order to unveil more subtle TF factor patterns that could have been concealed by the dominant interferon response signature present in thymic epithelial cells, we specifically interrogated *SOX2* and *TP63* along with a select set on curated TFs, known to regulate cell fate decisions in the anterior foregut and pharyngeal apparatus across a broad subset anterior foregut derived epithelial cell subsets. *SOX2* and *TP63* were highly co-expressed in basal epithelial cells of bronchus and esophagus and gradually decreased towards the parenchymal regions of the lung. All epithelial cells of the thymus, however, were marked by strong expression of *TP63* while *SOX2* was absent. Interestingly, the key foregut-defining transcription factors *FOXA1, SOX2, IRX3, IRX5*[28,62] remained highly expressed in epithelial cells of bronchial and esophageal origin and maintained moderate expression in the lung but were entirely absent in all epithelial cells of thymic origin. Thymic epithelial cells, in turn, demonstrated persistently strong expression of the third pharyngeal pouch markers *PAX1, PAX9, EYA1,* and *SIX1* (Figure 4A and B). To corroborate the divergent expression of *SOX2* and *TP63* observed in select thymic epithelial cell subsets, we visualized the *SOX2* and *TP63* regulons using the previously introduced SCENIC algorithm[42]. Expression of the *SOX2* regulon in its entirely was absent in thymic epithelial cells while the *TP63* regulon showed strong expression cells of thymic, bronchial, and esophageal origin (Figure 4C). Concordant protein expression of SOX2 and P63 in fetal sections of human thymus, bronchus, lung and esophagus was visualized by IHC (Figure 4D). These observations are compatible with previously put forth hypotheses[20,22,23] that the

development of the Thymic Anlage is linked to an overriding transcriptional program the abrogates branching morphogenesis in basal epithelial cells of the anterior foregut.

Next, we interrogated Claudin (CLDN) family genes which encode tight junction proteins that are known to play a critical role in apical basal polarity and tube formation[63,64]. Claudin-3 and 4 are specifically implicated in epithelial to mesenchymal transition[65] and the gain for Claudin-3 and 4 is classically considered a hallmark of medullary thymic epithelial commitment[19,20]. *CLDN3* and *CLDN4* were widely expressed in epithelial cells of bronchus, lung and esophagus while *CLDN5*, a gene characteristic for endothelial cells was uniquely high in the mTECII population (we are showing mTECII/III/IV) (Figure 4E). Interestingly, patients with 22q11.2 Deletion Syndrome are haploinsufficient for *CLDN5*[66].

FN1, VIM and *VCAM1*, also genes associated with epithelial to mesenchymal transition[67] were among the top differentially expressed genes in mTECI. Although the contribution of laminins to the thymic extracellular matrix has frequently been described[68-71], collagens *COL4A1, COL6A1, COL6A2, COL17A1*, integrin alpha subunit V (*ITGAV*) and Galectin 1 (*LGALS1*) predominated in our human thymic samples while unique laminins and other extracellular matrix proteins, e.g. perlecan (*HSPG2*), were selectively high in basal cells of bronchus and esophagus as well as bud tip progenitor adjacent cells of the lung. The distinctive extracellular environment of epithelial cells is matched by a characteristic intracellular scaffold with keratins predominating the cytoskeleton while functioning as main mediators for adhesion, migration, stress response and metabolic reactions[72]. Keratins (*KRT*) 8, 5, 14 and 17 predominated in cTECs and mTECs. Interestingly, mcTECs overall show a less keratinized phenotype possibly indicative of a developmentally less mature cell state. Bronchial basal cells share many of the keratins with TECs and show a uniquely high expression of *KRT15* while esophageal basal and suprabasal cells were specifically high in *KRT4*, *6B* and *13*. Bud tip progenitor and bud tip progenitor adjacent cells of the lung use *EPCAM* as the prevailing cell cytoskeletal element, consistent with their more pliable tissue architecture (Figure 4E).

To investigate the thymus to attract immune cells, and specifically to understand how this differs from other anterior foregut derived organs, we queried the expression of a wide range of chemokine receptors and ligands in our data set. Chemokine ligands are important chemo-attractants that actively recruit immune cells with cell type- and tissue-specific tropism. We found high expression levels of *CXCL14* in medullary TECs. CXCL14 is a natural inhibitor of the CXCL12:CXCR4 signaling axis[73]. The CXCL12:CXCR4 axis is used to attract CXCR4 positive thymocytes to the thymic cortex[74]. Expressing an inhibitor of this pathway in locations devoid of CXCR4 positive thymocytes raises the possibility that CXCL14 contributes to the spatial demarcation of immune cells within the thymic epithelial architecture. Interestingly, we also found a previously not recognized abundance of the C-C motif chemokine ligand *CCL2* in mcTECs. CCL2 has been shown to play a role in attracting macrophages and dendritic cells as well as in inducing tolerance in T cells[75,76]. Our data suggest that *CCL2* is a novel gene characteristic for mcTECs, for which connective tissue growth factor (*CTGF*) and the non-canonical inhibitor of NOTCH signaling *DKL2* have thus far been the only uniquely described genes[4]. Because CCL2 is secreted extracellularly, its visualization does not aid in spatially localizing the distribution mcTECs within the thymus (Supplementary Figure S4). We also observed the well know cell type specific expression of chemokine ligands *CXCL12* and *CCL25* in cTECs and *CCL19* and *CCL21* mTECIs, respectively[77] (Figure 4E).

Cortical TECs tightly regulate paracrine sources of retinoic acid

Retinoic acid (RA) signaling plays an important role in vertebrate morphogenesis, including the patterning of the pharyngeal pouches[78-80]. A dose-dependent excess of exogenous retinoic acid intake during pregnancy, e.g., in the form of vitamin A is associated with defects in thymus and parathyroid along with various other malformations[81,82]. Interestingly, children with mutations in RA metabolizing enzymes (*STRA6*, *ALDH1A2*) resulting in RA deficiency present with the same spectrum of congenital defects associated with RA toxicicy[83,84]. These findings suggest that an intricate balance of retinoic acid signaling is essential for thymic development. To understand the differential impact that RA plays in the development of anterior foregut-derived epithelial cells, we developed an RA metabolism gene regulatory module that integrates a curated set developmentally relevant RA metabolism genes[85,86] separated into RA synthesis, RA homeostasis and nuclear receptor interactions and RA degradation (Figure 5A).

Out of our data set encompassing all RA relevant anterior foregut derived epithelial subsets, basal epithelial cell of bronchial and esophageal origin scored highest in RA synthetic capacity within an overall narrow range across all queried cell types. cTECs, bronchial basal epithelial cells and esophageal basal epithelial cells displayed the highest RA nuclear activity score, once again within a relatively narrow range. Strikingly, cTECs had the highest RA degradation score (Figure 5B). As RA is frequently degraded by non-target tissues[79], the cell-intrinsic expression of RA metabolizing enzymes raises the possibility that RA levels in cTECs are under uniquely tight control. Retinoic acid signaling can be mediated through autocrine or paracrine sources of RA[79]. As RA synthesis scores were low across all types on epithelial cells including cTECs, we investigated exogenous sources of RA in the thymic microenvironment using the same functional RA metabolism gene regulatory module as above. Highest expression of RA synthesizing enzymes was found in both types of thymic fibroblasts, including aldehyde-dehydrogenase (*ALDH1A2*) positive fibroblasts also known as thymic fibroblasts 1[4], and fibronectin-high (*FBN1*) fibroblasts also known as thymic fibroblasts 2[4]. These data suggest that TECs, specifically cTECs tightly regulate paracrine sources of RA from the microenvironment. Interestingly, endothelial cells and lymphatic endothelial cells of the thymic microenvironment showed highest scores of nuclear activities while cTECs once again showed highest scores of RA degradation not only within the epithelial but also across the thymic microenvironmental data subset (Figure 5C).

The thymic microenvironment provides distinct signals to diverse TEC subsets

Critical receptor ligand interactions between the thymic epithelial and lymphoid compartment are well described[4,87]. In addition, a recent report highlighted the contribution of the non-epithelial non-lymphoid compartment to thymic development in murine models of health and disease[33]. To investigate the unique role the microenvironment plays in human thymic development, we used Cellphone DB v2.0. This database integrates the heteromeric architecture of receptor-ligand pairs to infer possible cell-cell communication based on matched gene expression levels in an unbiased fashion[88]. First, we assessed locally induced signaling between cortical TECs and primary non-epithelial subclusters within the thymic microenvironment, i.e., DN-Thymocytes, DP-Thymocytes, Endothelial Cells, Lymphatic Endothelial Cells, as well as *ALDH1A1* and *FBN1*-expressing Fibroblasts (Fib-1 and Fib-2). Only interactions with highest significance (p-value <0.01) are discussed (Figure 6).

The pivotal role of Wnt5a in thymic development was originally ascribed to its cell-autonomous expression within the thymic epithelium[89]. Recent murine studies, however, have suggested that *Wnt5a* displays a distinctive expression pattern in the non-epithelial thymic stroma[33]. Our human data showed that *WNT5a* stands out as the prevailing WNT ligand, which is broadly expressed in the non-epithelial microenvironment by thymic fibroblasts and endothelial cells. Cortical TECs express its

cognate receptors *ROR1*, *ROR2*, *FZD3*, *FZD5*, *FZD6* and *PTPRK*, that elicit a wide array of canonical, non-canonical WNT and WNT/Ca++ pathways signaling responses[90-92], respectively. The predominant cell autonomously expressed WNT ligand expressed by human cTECs is *WNT4*.

Elegant work in a transgenic *FGF7* mouse model has recently highlighted how early FGF7 begins to impact the development of two different bipotent thymic progenitor cell populations[93]. Based on the highly significant expression of cognate ligand pairs, our data suggest that human prenatal thymic fibroblasts are the primary source of *FGF7*, leading to FGFR2 stimulation on cTECs. Thymic fibroblasts are also the sole source of *FGF10* and *FGF18*. *BMP2* and *BMP4* is expressed by fibroblasts, endothelial and lymphatic endothelial cells and can activate a variety of BMP and Activin receptors on cTECs. In return, cTECs are the primary source of *BMP7* to promote Activin signaling in the non-epithelial stroma. Thymocytes and endothelial cells express the secreted TGF 1 and the EGF-receptor ligands *HBEGF* and *AREG* which can drive proliferation and differentiation via the MAPK pathway in cTECs. A strong synergistic effect of TGF 1 and EGF-ligands in intestinal epithelial cells has been proposed[94]. CellPhoneDB identified integrin complex A1B1 as the prevailing integrin expressed by cTECs, capable of interacting with a diverse set of collagens produced by thymic fibroblasts and endothelial cells. The thymus is also capable of producing a wide variety of collagens which interact almost exclusively with integrin complex A1B1 in the thymic fibroblasts as compared to both A1B1 and A2B1 in endothelial cells. A crucial role for laminins in murine thymic development and function has frequently been highlighted[68-71]. Within the human thymic microenvironment, fibroblasts, VSMCs, endothelial cells and cTECs also express an array of laminins albeit to a lesser extent than the other aforementioned extracellular matrix components, corroborating our gene expression data above (Supplementary Figure S5). This data offers actionable insights for the development of a clinically translatable extracellular thymic matrix.

Previously known receptor-ligand pairs which were used to validate the CellPhoneDB algorithm included CCL25 on cTECs interacting with CCR9 on immature T cells, which is believed to contribute to T cell homing and development[95]. In addition, we identified cTEC chemokine ligand *CCL19* interactions with the atypical chemokine receptor like 2 (*CCRL2*) on endothelial cells. In contrast to the known canonical CCL19:CCR7 interaction between cTECs and thymocytes, binding to CCRL2 does not result in cellular activation. Instead CCRL2 is constitutively recycled via clathrin-coated pits. Thus CCRL2 may have a role in modulating chemokine-triggered immune responses by controlling CCL19 densitiy[96].

Next, we extended the analysis to include all TEC subtypes. TECs exchange canonical and non-canonical NOTCH signals with the microenvironment[4,38,97]. Historically, investigations of thymic NOTCH signaling focused on activating canonical Delta Like Ligands (DLLs) on TECs that are indispensable for T cell maturation[98-109]. Our data suggest that endothelial cells are major contributors of DLL1 and DLL4 signals that lead to NOTCH activation in immature T cells via NOTCH3. Interestingly, the inhibitory NOTCH ligand Delta Like Non-Canonical Notch Ligand 1 (*DLK1*) is expressed on both endothelial cells and mcTECs and predominant interactions with cTECs via the NOTCH2 receptor are inferred. *DLK2* but not *DLK1* has recently been described as a gene uniquely expressed by mcTECs[4]. Both non-canonical inhibitory DLKs have been implicated in the inhibition of differentiation into adipocytes in other epithelial cells[110-113]. In the context of adipose transformation of the thymic stroma, these findings open new avenues of investigation into the role of mcTECs in thymic epithelial maintenance.

Within our dataset, the strongest, most significant receptor-ligand interactions present across all TEC subtypes were inferred between *CD74* (cTEC, mcTEC, mTECI/II/III/IV) and the Macrophage Migration

Inhibiting Factor (*MIF*) expressed by all major cell types of the microenvironment (endothelial, mesenchymal, neuronal, immune). MIF was originally described as an atypical lymphokine implicated in delayed hypersensitivity reactions. Interestingly, contrary to other cytokines that are locally induced, MIF is constitutively expressed by many tissues, raising the possibility that its effect is regulated on the receptor level. Proteins with receptor function for MIF had not been identified until many years later and include CD74, CXCR2 and CXCR4[114-116]. CD74 was first described as the invariant chain for Major Histocompatibility Complex II (MHCII). In this context, CD74 acts as a chaperone and is indispensable for antigen presentation[117]. Accordingly, it is highly expressed on thymic epithelial cells. Over time, MCHII-independent roles for CD74 as an important immunomodulatory co-receptor have emerged[118,119]. Upon MIF binding, CD74 releases its cytosolic intracellular domain (CD74 ICD), which goes on to act as a transcriptional modulator promoting proliferative, proinflammatory and B pathways as well as direct DNA-binding interactions[114,117,118].

Interactions between different thymic epithelial cell subsets were queried. The most frequent inferred level of significance.

Lineage tracing identifies common progenitor population for cortical and medullary thymic epithelial cells

To examine lineage diversification within the thymic epithelium, we focused our analysis on thymic epithelial from different gestational ages, integrating pcw 6, 7, 16, 17, 18 and a neonatal (age 3 months) sample. In addition to the canonical and other known thymic epithelial cell populations (cTEC, mcTEC, mTECI, mTECII, mTECIII/IV, TECmyo/neuro), we defined a sizable population of TEC that was visualized near mcTECs and demonstrated the closest dendrogram relationship (Figure 8A). This population, however, clustered independently according to genes implicated in cell cycle, DNA synthesis and proliferation (*TOP2A, CENPF, NUSAP1, TUBAB1, TUBB, MKI67, CKS2*) and had the highest G2M/S phase score (Figure 8B). Accordingly, we designated this population cycling TECs. In addition, we identified a small population of parathyroid epithelial cells (*PTH, GCM2, CASR, TBX1, GATA3*). With increasing gestational and postal age, cycling TECs and cTECs showed a decrease in relative abundance compared the other subpopulations, while mTECIs and mTECIIs showed a relative increase over time (Figure 8C). Hdac3 has been described as master regulator of murine mTEC development[120]. Histone acetyltransferase and histone deacetylase activity are important regulators of chromatin remodeling which - -inducible transcription[121,122]. We therefore queried the expression of the most abundant histone deacetylases (*HDAC1-8*) expressed in the thymic epithelium. Interestingly, in our human data set, the histone deacetylases most differentially expressed between cTEC and mTECI subpopulations was *HDAC4. HDAC4* demonstrated significantly increased expression in cTECs which became even more specific with increasing age.

To corroborate preferential expression of HDAC4 in cTECs, we performed IHC on human fetal thymus sections. HDAC4 was coexpressed with PSMB11 in the thymic cortex while Keratin 5 positive mTECs colocalized with HDAC4 to a lesser degree. In addition, we evaluated human fetal thymus sections for global differences in histone acetylation between cortex and medulla. H4K16ac, a target of multiple HDACs including HDAC1, HDAC2, and HDAC4[123,124] demonstrated increased signal intensity in the thymic medulla, however, the signal originated largely from non-epithelial cells preventing the exclusive assessment of acetylation status in mTECs (Figure 8D; Supplementary Figure S6). Other epithelial subpopulations also demonstrated preferential expression of specific

HDAC subtypes. Cycling TECs showed highest expressing of *HDAC1*, TECmyo/neuro of *HDAC2*, mTECIII/IV in *HDAC5*, suggesting differences in chromatin states match differences in cell identify.

Discussion:

Recent progress in genomic and epigenomic analysis methods have redefined our understanding of thymic epithelial diversity and architecture with unprecedented granularity. The signals that drive epithelial cell fate decisions are now starting to emerge. Our comparison of human fetal epithelial cells from the major anterior foregut derived organs provides compelling evidence for the inductive effect interferon type 1 and 2 exert during commitment and specification of the thymus. As interferon gene expression is known to show extremely low abundance in transcriptomic studies, we explored the origin of interferon production by alternate means. We demonstrate intracellular IFN protein in T cell precursors and thymic epithelial cells. Although both, endogenous expression of *IFNG* by thymic epithelial cells or uptake of exogenously produced IFN by thymic epithelial cells are plausible explanations, we favor the latter. In the context of thymic development, developing T lineage committed precursors are most likely the main producers of IFN . IFN and IFN can be produced by many cell types including T cells, although dendritic cells and macrophages are classically regarded as primary source of type I interferon. Differential expression patterns in genes altering the interferon response, including cytokine receptors, their ligands and RA in any of the cells of the thymic microenvironment may help orchestrating the intricate signals that define the fate of varied epithelial subpopulations in spatially distinct microclimates. On the other hand, the fact that e.g., cTECs and mTECs seem to respond differentially to the inductive effect of interferon aligns with the notion that the initial patterning of thymic cortex and medulla indeed precedes the recruitment of hematopoietic elements into the thymus.

To this end, the data on the divergent expression of SOX2 and TP63 is intriguing. When exactly SOX2 begins to decrease during human thymic epithelial development evades our means of analysis in human samples. In the mouse, Sox2 is expressed in the third pharyngeal pouch[126,127]. It is conceivable that SOX2 and TP63 remain co-expressed while pharyngeal endoderm cells abide by the law of apical- basal polarity and attach to a basement membrane, and that changes in the SOX2-TP63 expression pattern occur as the cells of the Thymic Anlage proliferate and expand in 3D. This is in line with the historical hypothesis that a program overriding classical basal epithelial cell fate must be initiated for the Thymic Anlage to form.

Previous studies pointed to a role for the TNF signaling axis mediated via CD40 and RANK in medullary TEC development[128,129]. These signals were shown to act through activation of the pathway[130]. In turn, disruption of effectors downstream of the pathway abrogated essential aspects of mTEC development and function[131,132]. The overarching programs that govern cortical TEC commitment and differentiation are yet to be defined. Our data points to a more global role interferon type I and II play in thymic development while IRFs and emerge as downstream mediators that fine tune the interferon response along the development of the mTEC and cTEC lineages. How exactly the global effects of type I and type II are differentially mediated in cTEC and mTEC remains subject of active investigation. Multiped findings within our dataset open novel avenues of inquiry into the differential response to interferons in spatially different compartments of the thymus. There is extensive crosstalk between retinoic and interferon signaling[133] and interferon has been shown to increase the expression of IRF1 and STAT1[134-136]. Our data indicate that RA is preferentially degraded in cTECs which could explain the lower IFN modules scores observed in our analysis. The same goes for extensive cross signaling interactions between interferon and MIF:CD74. A growing

body of evidence suggests that MIF upregulates TLR4 expression and promotes the production of - -2, IL-6, IL-8[114,137,138]. This raises the possibility that MIF:CD74 interactions contribute to modulating the TEC subpopulation specific interferon-response signature highlighted above. Previous studies suggested that CD74 positively influences mTEC cellularity in the mouse[119]. Our data offers a larger framework in which the MIF:CD74 axis is part of the modulatory circuits that fine tune interferon and responses.

While our data shed light on the role of interferons in the developmental onset of the thymus, it begs the question what role interferons play in the life cycle of the thymus overall. Interferon induces a strong antiproliferative gene expression patterns. In the hematopoietic stem cell niche, interferons have been intimately linked to hematopoietic stem cell ageing and exhaustion. Like others before us, we evaluated for makers of epithelial stem cell-like fate like, the leucine-rich repeat (LRR)-containing G-protein-coupled receptor 4/5/6 (Lgr4/5/6) gene family[139], the hemidesmosome (*ITGA6*)[140], *KRT15*[140,141], *BCL-2*[142] without solid evidence for a long lived thymic stem cell. Is it by coincidence that the human thymus involutes early in life and postnatal thymic stem cell have yet to be identified? Is the tale of the thymus the story of an organ whose rise and fall is intimately linked to IFN?

MATERIALS AND METHODS

Tissue Acquisition

Human fetal tissues from thymus, bronchus, lung and esophagus from 7 to 18 post conception week (pcw) specimens were obtained under Stanford School of Medicine IRB-approved protocols from the Department of Obstetrics and Gynecology at Stanford Hospital. Samples were derived from deidentified fetal tissue donations with written informed consent after elective termination of pregnancy. The study does not qualify as human subjects research, as defined by the Stanford School of Medicine IRB.

Single cell transcriptomic data from pcw 7-8 samples was obtained in collaboration with the laboratory of Sarah Teichmann, from the MRC/Wellcome Trust-funded Human Developmental Biology Resources (HDBR, http://www.hdbr.org).

Tissue Digestion

Human fetal thymus and lung samples (pcw 16 -18) were mechanically dissected into <2 mm^3 pieces using surgical scissors. Human fetal bronchus and esophagus samples (pcw 16 -18) were opened longitudinally and mechanically scraped with scalpel to harvest the inner epithelial layer of bronchus and esophagus, respectively. All tissue samples were enzymatically digested with DNAse (10 mg/ml) and Liberase (2.5 mg/ml, Sigma Aldrich) per 10 g tissue per tube using a tissue dissociator (gentleMACS, C tube, program m_spleen_02 (Miltenyi), 1-3 runs depending on tissue). Tissue fragments were transferred to a 50ml tube and incubate at 37 C for 20 min. Cells were pelleted at 110g, RT for 2 min and resuspended in 10 ml PBS. For persisting fragments, the procedure was repeated 1-2 times to yield single cells. Cells were resuspended in MACS buffer (PBS, 1 %BSA, 2mM EDTA), filtered through a 100um cell strainer and counted.

Epithelial Cell Enrichment

For fetal thymic tissue (16-18 pcw) only, cell suspension was depleted of thymocytes and other immune cells using Human CD45 MicroBeads (Miltenyi) and LD depletion columns (Miltenyi),10^8

(CD45-depleted thymus, bronchus, lung, and esophagus) were enriched for epithelial cells using Human CD326 (EpCAM) Microbeads and LS positive selection columns (Miltenyi) according to the n and epithelial cell enrichment for

thymic single cell suspensions was performed by flow cytometry on Beckman FACSAria II with CD45-APC, CD3-FITC, CD205-PE, EpCAM-APC, and L/D-NearIR-APC-Cy7 prior to MACS purification. Due to epitope blockage from magnetic bead antibodies, additional markers, i.e., CD7-AF647 (thymocytes) and CD105-PerCP/Cy5.5 (thymic mesenchymal cells) were used to assess post enrichment purity.

scRNAseq Library Preparation and Sequencing
Gel-
chemistry and Chromium microfluidics system (10X Genomics). After reverse transcription, barcoded cDNA was extracted from the GEMs by Post-GEM RT-cleanup and amplified for 12 cycles. Amplified cDNA was subjected to 0.6x SPRI beads cleanup (Beckman, B23318). 25% of the amplified cDNA was subjected to enzymatic fragmentation, end-repair, A tailing, adaptor ligation and 10X specific

(Agilent) analysis. Libraries were pooled and sequenced on an NovaSeq 6000 instrument (Illumina) using the recommended sequencing read lengths of 28 bp (Read 1), 8 bp (i7 Index Read), and 91 bp (Read 2).

scRNA-seq data processing
Single cell data were aligned to human genome library and downstream processing was performed using Seurat V4.0.

Differential gene expression
Differential gene expression between single-cell clusters was determined using the FindAllMarkers function within Seurat.

Gene set analysis
Gene sets were obtained from MSigDB curated C2 dataset[58]. We then performed gene set enrichment analysis (GSEA) using these sets and differential gene expression between annotated single-cell clusters as determined in Seurat. In order to visualize biological themes, created a network representation of positively enriched gene sets expression across single-cell clusters. In these networks, gene sets represent nodes and edges are drawn between nodes based on the proportion of shared genes between them. Each node is then colored proportionally by the relative enrichment of that gene set in each single-cell cluster. GSEA was performed using clusterProfiler/4.5.1.902 and gene set network visualization was performed using enrichplot/1.17.0.992[59].

Module scores
We determined gene module scores by applying the AddModuleScore function within Seurat to gene lists of interest. In determining module scores from gene lists determined by Nusser et al. mouse gene IDs were converted to corresponding human gene IDs using biomaRt/2.46.3. Genes without corresponding human homologs were excluded.

Trajectory and pseudotime analysis
We utilized tSpace/0.1.0 [60], to better represent continuous relationships between single-cell expression profiles. In brief, tSpace determines cellular distances based on the shortest path between cellular k-nearest neighbor (KNN) subgraphs using the Wanderlust algorithm. These distances with tSpace were then reduced using uniform manifold approximation and projection (UMAP). We then fit principle curves through dimensionally reduced tSpace data using slingshot/1.8.0 and SingleCellExperiment/1.12.0. Cycling TECs were selected as the starting-point cluster, while end-

point clusters were not specified. We then extracted principal curve coordinates and cellular pseudotime values based on the resulting principal curves.

Modeling gene expression along pseudotime

We utilized tradeSeq/1.4.0[61] to build a negative binomial generalized additive model of single-cell gene expression along pseudotime. This allowed for statistical testing of both within-trajectory and between-trajectory gene expression differences as well as smoothing of gene expression along pseudotime for the purpose of visualization. earlyDETest was used to identify the key differential genes between lineages 1 and 2 across knots 2-5. The genes were then filtered using a p-value = 0 and sorted according to the highest Wald Statistic value. The top 60 genes were then used for heatmap construction.

Trajectory Heatmap Construction

Scaled expression values (horizontal analysis) or Regulon AUC values (vertical analysis) were used to construct an expression matrix across pseudotime. The input matrix was then quantile scaled using scale_quantile(outlier_cutoff = 0.1) SCORPIUS/1.0.8.
The rolling average of this matrix was then constructed using rollapply(width=700,FUN = mean,fill=NA,align="right") zoo/1.8-11 and passed into pheatmap/1.0.12 for construction.

Gene regulatory network inference

Prior to pySCENIC/0.12.0 input (Python/3.9.0), the log1p normalized expression matrices were filtered to remove low quality genes according to an average expression threshold across all cells of .01. The filtered expression matrices were then converted into a loom file and passed into GRNBoost2 to evaluate co-expression-based network interference. The resultant adjacency matrices were then passed into the RcisTarget step for regulon prediction. Predicted regulons and an aggregate of their targets from different motifs were then used to determine cellular enrichment for each regulon using AUCell.
Motif database used: hg38__refseqr80__10kb_up_and_down_tss.mc9nr.genes_vs_motifs.rankings
TF list used: allTFs_hg38.txt
Motif2TF annotations: motifs-v9-nr.hgnc-m0.001-o0.0.tbl

Regulon AUCell Output Seurat Integration

The AUCell matrix was passed into Seurat/4.3 and normalized across features using a centered log ratio transformation. The data then underwent the typical Seurat clustering pipeline consisting of FindVariableFeatures(), ScaleData(), followed by RunPCA(). UMAP dimensionality reduction was carried out using the first 20 PCs for the vertical analysis and the first 8 PCs for the vertical analysis. Previously defined cell identities were used to verify SCENIC based clustering. Differential regulons were identified using the FindAllMarkers() Seurat function with a log2fc.threshold = 0.01. Differentially identified regulons were then cross referenced for differential expression in the original transcriptional data. Module scores for composite regulon expression were calculated using AddModuleScore(parameters: nbin = 24, ctrl = 12) to account for input matrix size.

Differential Regulon Network Construction

The top 250 most variable regul
horizontal analysis were identified. RcisTarget refined regulon and target interactions across all motifs were aggregated and an edge list was constructed by drawing an edge if a variable feature was the target of a different variable feature (Python Pandas/1.5.2 used). The resultant edge list was then passed into igraph/1.3.5 and simplified using the igraph::simplify() function and clustered using igraph::cluster_leiden(). ggnetwork/0.5.10 and ggplot2/3.4.0 were used for the final network visualization. Nodes were force directed using igraph::layout_with_fr().

Trajectory Heatmap Construction
Scaled expression values (horizontal analysis) or Regulon AUC values (vertical analysis) were used to construct an expression matrix across pseudotime. The input matrix was then quantile scaled using scale_quantile(outlier_cutoff = 0.1) SCORPIUS/1.0.8.
The rolling average of this matrix was then constructed using rollapply(width=700,FUN = mean,fill=NA,align="right") zoo/1.8-11 and passed into pheatmap/1.0.12 for construction.

Ligand-Receptor Interaction Analysis

Ligand-Receptor interaction analysis was performed utilizing the command line statistical_analysis module within the CellPhoneDB package. Specific cell subtypes (e.g. mTECI, mcTEC, mTECII/III/IV, DP-Thymocytes, DN-Thymocytes etc) were selected for focused analyses. Interactions were filtered for pairwise interactions with a p-value < 0.001. Select interactions were then visualized in a dot plot using ggplot2/3.40.

Immunohistochemistry and confocal imaging
Paraffin-embedded human fetal thymus tissues were deparaffinized in xylene and rehydrated with ethanol solutions of decreasing concentrations until a final wash step with milliQ water. Antigen retrieval was performed with IHC-Tek™ Epitope Retrieval Solution (#IW-1100-1L, IHC World) using steam. Tissue slices were permeabilized with permeabilization solution (PBS, 0.2%TritonX-100) for 30 minutes at room temperature, blocked with PBS containing 10% of Donkey Serum (DS) and 1% Bovine Serum Albumin for 1 hour at room temperature, and incubated with primary antibodies (see

washed with PBS and incubated with fluorescently-labeled secondary antibody for 1 hour at room temperature. After subsequent washes with PBS, tissue autofluorescence was quenched with Vector TrueVIEW Autofluorescence Quenching Kit (# SP-8400, Vector Labs). Nuclei were stained with 5 µg/ml of DAPI in PBS for 10 minutes at room temperature. Slides were mounted with VectaMount. Tissue visualization and image collection was performed using a Leica Stellaris 5 Confocal microscope and LAS X software (Leica Microsystems). All images were acquired using 40X (HC PL APO CS2, NA 1.30) or 63X (HC PL APO CS2, NA 1.40) oil-immersion objectives.

Table 1. List of Antibodies

Antibody	Dilution	Source	Identifier
Mouse anti-human p63	1:100	Abcam	ab735
Mouse anti-human CCL2/MCP1	1:100	Novus biologicals	NBP2-22115SS
Mouse anti-human Histone deacetylase 4 (HDAC4)	1:50	Santa Cruz	sc-46671
Mouse anti-rabbit Pan-Keratin	1:100	Santa Cruz	Sc-81714
Rabbit anti-human PSMB11	1:50	Abcam	Ab170832
Rabbit anti-human Histone H4 acetyl K16 (H4K16ac)	1:100	Abcam	ab109463

Rabbit anti-human Interferon gamma	1:100	Abcam	ab25101
Rabbit anti-human Keratin 5	1:100	Biolegend	905501
Rabbit anti-human SOX2	1:500	Abcam	ab97959
Donkey anti-rabbit Alexa Fluor 488	1:500	Abcam	ab150073
Donkey anti-mouse Alexa Fluor 647	1:500	Abcam	ab150107

REFERENCES

1. Kernfeld EM, Genga RMJ, Neherin K, Magaletta ME, Xu P, Maehr R. A Single-Cell Transcriptomic Atlas of Thymus Organogenesis Resolves Cell Types and Developmental Maturation. *Immunity.* 2018;48(6):1258-1270 e1256.
2. Bornstein C, Nevo S, Giladi A, et al. Single-cell mapping of the thymic stroma identifies IL-25-producing tuft epithelial cells. *Nature.* 2018;559(7715):622-626.
3. Miller CN, Proekt I, von Moltke J, et al. Thymic tuft cells promote an IL-4-enriched medulla and shape thymocyte development. *Nature.* 2018;559(7715):627-631.
4. Park JE, Botting RA, Dominguez Conde C, et al. A cell atlas of human thymic development defines T cell repertoire formation. *Science.* 2020;367(6480).
5. Bautista JL, Cramer NT, Miller CN, et al. Single-cell transcriptional profiling of human thymic stroma uncovers novel cellular heterogeneity in the thymic medulla. *Nat Commun.* 2021;12(1):1096.
6. Zeng Y, Liu C, Gong Y, et al. Single-Cell RNA Sequencing Resolves Spatiotemporal Development of Pre-thymic Lymphoid Progenitors and Thymus Organogenesis in Human Embryos. *Immunity.* 2019;51(5):930-948 e936.
7. Zorn AM, Wells JM. Vertebrate endoderm development and organ formation. *Annu Rev Cell Dev Biol.* 2009;25:221-251.
8. Graham A, Richardson J. Developmental and evolutionary origins of the pharyngeal apparatus. *Evodevo.* 2012;3(1):24.
9. Nehls M, Kyewski B, Messerle M, et al. Two genetically separable steps in the differentiation of thymic epithelium. *Science.* 1996;272(5263):886-889.
10. Abramson J, Anderson G. Thymic Epithelial Cells. *Annu Rev Immunol.* 2017;35:85-118.
11. Anderson G, Jenkinson EJ. Lymphostromal interactions in thymic development and function. *Nat Rev Immunol.* 2001;1(1):31-40.
12. Zuklys S, Handel A, Zhanybekova S, et al. Foxn1 regulates key target genes essential for T cell development in postnatal thymic epithelial cells. *Nat Immunol.* 2016;17(10):1206-1215.
13. Farley AM, Morris LX, Vroegindeweij E, et al. Dynamics of thymus organogenesis and colonization in early human development. *Development.* 2013;140(9):2015-2026.
14. Blackburn CC, Augustine CL, Li R, et al. The nu gene acts cell-autonomously and is required for differentiation of thymic epithelial progenitors. *Proc Natl Acad Sci U S A.* 1996;93(12):5742-5746.
15. Nowell CS, Bredenkamp N, Tetelin S, et al. Foxn1 regulates lineage progression in cortical and medullary thymic epithelial cells but is dispensable for medullary sublineage divergence. *PLoS Genet.* 2011;7(11):e1002348.
16. Shakib S, Desanti GE, Jenkinson WE, Parnell SM, Jenkinson EJ, Anderson G. Checkpoints in the development of thymic cortical epithelial cells. *J Immunol.* 2009;182(1):130-137.

17. White AJ, Withers DR, Parnell SM, et al. Sequential phases in the development of Aire-expressing medullary thymic epithelial cells involve distinct cellular input. *Eur J Immunol.* 2008;38(4):942-947.
18. Klug DB, Carter C, Gimenez-Conti IB, Richie ER. Cutting edge: thymocyte-independent and thymocyte-dependent phases of epithelial patterning in the fetal thymus. *J Immunol.* 2002;169(6):2842-2845.
19. Munoz JJ, Cejalvo T, Tobajas E, Fanlo L, Cortes A, Zapata AG. 3D immunofluorescence analysis of early thymic morphogenesis and medulla development. *Histol Histopathol.* 2015;30(5):589-599.
20. Munoz JJ, Tobajas E, Juara S, Montero S, Zapata AG. FoxN1 mediates thymic cortex-medulla differentiation through modifying a developmental pattern based on epithelial tubulogenesis. *Histochem Cell Biol.* 2019;152(6):397-413.
21. Flanagan SP. 'Nude', a new hairless gene with pleiotropic effects in the mouse. *Genet Res.* 1966;8(3):295-309.
22. Dooley J, Erickson M, Roelink H, Farr AG. Nude thymic rudiment lacking functional foxn1 resembles respiratory epithelium. *Dev Dyn.* 2005;233(4):1605-1612.
23. Dooley J, Erickson M, Farr AG. An organized medullary epithelial structure in the normal thymus expresses molecules of respiratory epithelium and resembles the epithelial thymic rudiment of nude mice. *J Immunol.* 2005;175(7):4331-4337.
24. Cordier AC. Ultrastructure of the cilia of thymic cysts in "nude" mice. *Anat Rec.* 1975;181(2):227-249.
25. Edwards NA, Shacham-Silverberg V, Weitz L, et al. Developmental basis of trachea-esophageal birth defects. *Dev Biol.* 2021;477:85-97.
26. Domyan ET, Ferretti E, Throckmorton K, Mishina Y, Nicolis SK, Sun X. Signaling through BMP receptors promotes respiratory identity in the foregut via repression of Sox2. *Development.* 2011;138(5):971-981.
27. Kishimoto K, Iwasawa K, Sorel A, et al. Directed differentiation of human pluripotent stem cells into diverse organ-specific mesenchyme of the digestive and respiratory systems. *Nat Protoc.* 2022.
28. Green MD, Chen A, Nostro MC, et al. Generation of anterior foregut endoderm from human embryonic and induced pluripotent stem cells. *Nat Biotechnol.* 2011;29(3):267-272.
29. Miller AJ, Dye BR, Ferrer-Torres D, et al. Generation of lung organoids from human pluripotent stem cells in vitro. *Nat Protoc.* 2019;14(2):518-540.
30. Trisno SL, Philo KED, McCracken KW, et al. Esophageal Organoids from Human Pluripotent Stem Cells Delineate Sox2 Functions during Esophageal Specification. *Cell Stem Cell.* 2018;23(4):501-515 e507.
31. Zhang Y, Yang Y, Jiang M, et al. 3D Modeling of Esophageal Development using Human PSC-Derived Basal Progenitors Reveals a Critical Role for Notch Signaling. *Cell Stem Cell.* 2018;23(4):516-529 e515.
32. Huang SX, Green MD, de Carvalho AT, et al. The in vitro generation of lung and airway progenitor cells from human pluripotent stem cells. *Nat Protoc.* 2015;10(3):413-425.
33. Handel AE, Cheuk S, Dhalla F, et al. Developmental dynamics of the neural crest-mesenchymal axis in creating the thymic microenvironment. *Sci Adv.* 2022;8(19):eabm9844.
34. Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* 2019;20(1):296.
35. Stuart T, Butler A, Hoffman P, et al. Comprehensive Integration of Single-Cell Data. *Cell.* 2019;177(7):1888-1902 e1821.

36. Miller AJ, Yu Q, Czerwinski M, et al. In Vitro and In Vivo Development of the Human Airway at Single-Cell Resolution. *Dev Cell.* 2020;54(6):818.
37. Travaglini KJ, Nabhan AN, Penland L, et al. A molecular cell atlas of the human lung from single-cell RNA sequencing. *Nature.* 2020;587(7835):619-625.
38. Busslinger GA, Weusten BLA, Bogte A, Begthel H, Brosens LAA, Clevers H. Human gastrointestinal epithelia of the esophagus, stomach, and duodenum resolved at single-cell resolution. *Cell Rep.* 2021;34(10):108819.
39. Metzger TC, Khan IS, Gardner JM, et al. Lineage tracing and cell ablation identify a post-Aire-expressing thymic epithelial cell population. *Cell Rep.* 2013;5(1):166-179.
40. Sakaguchi M, Huh NH. S100A11, a dual growth regulator of epidermal keratinocytes. *Amino Acids.* 2011;41(4):797-807.
41. Xie Y, Ostriker AC, Jin Y, et al. LMO7 Is a Negative Feedback Regulator of Transforming Growth Factor beta Signaling and Fibrosis. *Circulation.* 2019;139(5):679-693.
42. Aibar S, Gonzalez-Blas CB, Moerman T, et al. SCENIC: single-cell regulatory network inference and clustering. *Nature methods.* 2017;14(11):1083-1086.
43. DeSandro AM, Nagarajan UM, Boss JM. Associations and interactions between bare lymphocyte syndrome factors. *Mol Cell Biol.* 2000;20(17):6587-6599.
44. Cook ME, Jarjour NN, Lin CC, Edelson BT. Transcription Factor Bhlhe40 in Immunity and Autoimmunity. *Trends Immunol.* 2020;41(11):1023-1036.
45. Zhao X, Li L, Yuan S, Zhang Q, Jiang X, Luo T. SPIB acts as a tumor suppressor by activating the NFkB and JNK signaling pathways through MAP4K1 in colorectal cancer cells. *Cell Signal.* 2021;88:110148.
46. Matsumoto H, Scicluna BP, Jim KK, et al. HIVEP1 Is a Negative Regulator of NF-kappaB That Inhibits Systemic Inflammation in Sepsis. *Frontiers in immunology.* 2021;12:744358.
47. Pestka S, Krause CD, Walter MR. Interferons, interferon-like cytokines, and their receptors. *Immunol Rev.* 2004;202:8-32.
48. Darnell JE, Jr., Kerr IM, Stark GR. Jak-STAT pathways and transcriptional activation in response to IFNs and other extracellular signaling proteins. *Science.* 1994;264(5164):1415-1421.
49. Der SD, Zhou A, Williams BR, Silverman RH. Identification of genes differentially regulated by interferon alpha, beta, or gamma using oligonucleotide arrays. *Proc Natl Acad Sci U S A.* 1998;95(26):15623-15628.
50. Platanias LC. Mechanisms of type-I- and type-II-interferon-mediated signalling. *Nat Rev Immunol.* 2005;5(5):375-386.
51. Pfeffer LM. The role of nuclear factor kappaB in the interferon response. *J Interferon Cytokine Res.* 2011;31(7):553-559.
52. Bertrand N, Castro DS, Guillemot F. Proneural genes and the specification of neural cell types. *Nat Rev Neurosci.* 2002;3(7):517-530.
53. Castro DS, Martynoga B, Parras C, et al. A novel function of the proneural factor Ascl1 in progenitor proliferation identified by genome-wide characterization of its targets. *Genes Dev.* 2011;25(9):930-945.
54. Xing Y, Wang X, Jameson SC, Hogquist KA. Late stages of T cell maturation in the thymus involve NF-kappaB and tonic type I interferon signaling. *Nat Immunol.* 2016;17(5):565-573.
55. Lavaert M, Liang KL, Vandamme N, et al. Integrated scRNA-Seq Identifies Human Postnatal Thymus Seeding Progenitors and Regulatory Dynamics of Differentiating Immature Thymocytes. *Immunity.* 2020;52(6):1088-1104 e1086.
56. Le J, Park JE, Ha VL, et al. Single-Cell RNA-Seq Mapping of Human Thymopoiesis Reveals Lineage Specification Trajectories and a Commitment Spectrum in T Cell Development. *Immunity.* 2020;52(6):1105-1118 e1109.

57. Larkin J, 3rd, Johnson HM, Subramaniam PS. Differential nuclear localization of the IFNGR-1 and IFNGR-2 subunits of the IFN-gamma receptor complex following activation by IFN-gamma. *J Interferon Cytokine Res.* 2000;20(6):565-576.
58. Subramaniam PS, Johnson HM. Lipid microdomains are required sites for the selective endocytosis and nuclear translocation of IFN-gamma, its receptor chain IFN-gamma receptor-1, and the phosphorylation and nuclear translocation of STAT1alpha. *J Immunol.* 2002;169(4):1959-1969.
59. Subramaniam PS, Green MM, Larkin J, 3rd, Torres BA, Johnson HM. Nuclear translocation of IFN-gamma is an intrinsic requirement for its biologic activity and can be driven by a heterologous nuclear localization sequence. *J Interferon Cytokine Res.* 2001;21(11):951-959.
60. Gontan C, de Munck A, Vermeij M, Grosveld F, Tibboel D, Rottier R. Sox2 is important for two crucial processes in lung development: branching morphogenesis and epithelial cell differentiation. *Dev Biol.* 2008;317(1):296-309.
61. Ochieng JK, Schilders K, Kool H, et al. Sox2 regulates the emergence of lung basal cells by directly activating the transcription of Trp63. *Am J Respir Cell Mol Biol.* 2014;51(2):311-322.
62. Loh KM, Ang LT, Zhang J, et al. Efficient endoderm induction from human pluripotent stem cells by logically directing signals controlling lineage bifurcations. *Cell Stem Cell.* 2014;14(2):237-252.
63. Hewitt KJ, Agarwal R, Morin PJ. The claudin gene family: expression in normal and neoplastic tissues. *BMC Cancer.* 2006;6:186.
64. Baumholtz AI, Gupta IR, Ryan AK. Claudins in morphogenesis: Forming an epithelial tube. *Tissue Barriers.* 2017;5(4):e1361899.
65. Lin X, Shang X, Manorek G, Howell SB. Regulation of the Epithelial-Mesenchymal Transition by Claudin-3 and Claudin-4. *PLoS One.* 2013;8(6):e67496.
66. McDonald-McGinn DM, Sullivan KE, Marino B, et al. 22q11.2 deletion syndrome. *Nat Rev Dis Primers.* 2015;1:15071.
67. Kalluri R, Weinberg RA. The basics of epithelial-mesenchymal transition. *The Journal of clinical investigation.* 2009;119(6):1420-1428.
68. Berrih S, Savino W, Cohen S. Extracellular matrix of the human thymus: immunofluorescence studies on frozen sections and cultured epithelial cells. *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society.* 1985;33(7):655-664.
69. Ocampo JSP, Brito JMd, Corrêa-de-Santana E, Borojevic R, Villa-Verde DMS, Savino W. Laminin-211 controls thymocyte thymic epithelial cell interactions. *Cellular Immunology.* 2008;254(1):1-9.
70. Savino W, Mendes-da-Cruz DA, Golbert DC, Riederer I, Cotta-de-Almeida V. Laminin-Mediated Interactions in Thymocyte Migration and Development. *Frontiers in immunology.* 2015;6:579.
71. Simon T, Bromberg JS. Regulation of the Immune System by Laminins. *Trends in Immunology.* 2017;38(11):858-871.
72. Windoffer R, Beil M, Magin TM, Leube RE. Cytoskeleton in motion: the dynamics of keratin intermediate filaments in epithelia. *J Cell Biol.* 2011;194(5):669-678.
73. Tanegashima K, Suzuki K, Nakayama Y, et al. CXCL14 is a natural inhibitor of the CXCL12-CXCR4 signaling axis. *FEBS Lett.* 2013;587(12):1731-1735.
74. Lancaster JN, Li Y, Ehrlich LIR. Chemokine-Mediated Choreography of Thymocyte Development and Selection. *Trends Immunol.* 2018;39(2):86-98.
75. Bleul CC, Boehm T. Chemokines define distinct microenvironments in the developing thymus. *Eur J Immunol.* 2000;30(12):3371-3379.
76. Cedile O, Lobner M, Toft-Hansen H, et al. Thymic CCL2 influences induction of T-cell tolerance. *J Autoimmun.* 2014;55:73-85.
77. <1020.full.pdf>.

78. Cunningham TJ, Duester G. Mechanisms of retinoic acid signalling and its roles in organ and limb development. *Nat Rev Mol Cell Biol.* 2015;16(2):110-123.
79. Duester G. Retinoic acid synthesis and signaling during early organogenesis. *Cell.* 2008;134(6):921-931.
80. Kopinke D, Sasine J, Swift J, Stephens WZ, Piotrowski T. Retinoic acid is required for endodermal pouch morphogenesis and not for pharyngeal endoderm specification. *Dev Dyn.* 2006;235(10):2695-2709.
81. Lammer EJ, Chen DT, Hoar RM, et al. Retinoic acid embryopathy. *N Engl J Med.* 1985;313(14):837-841.
82. Rothman KJ, Moore LL, Singer MR, Nguyen US, Mannino S, Milunsky A. Teratogenicity of high vitamin A intake. *N Engl J Med.* 1995;333(21):1369-1373.
83. Pasutto F, Sticht H, Hammersen G, et al. Mutations in STRA6 cause a broad spectrum of malformations including anophthalmia, congenital heart defects, diaphragmatic hernia, alveolar capillary dysplasia, lung hypoplasia, and mental retardation. *Am J Hum Genet.* 2007;80(3):550-560.
84. Pavan M, Ruiz VF, Silva FA, et al. ALDH1A2 (RALDH2) genetic variation in human congenital heart disease. *BMC Med Genet.* 2009;10:113.
85. Rhinn M, Dolle P. Retinoic acid signalling during development. *Development.* 2012;139(5):843-858.
86. Kam RK, Deng Y, Chen Y, Zhao H. Retinoic acid synthesis and functions in early embryonic development. *Cell Biosci.* 2012;2(1):11.
87. Han J, Zuniga-Pflucker JC. A 2020 View of Thymus Stromal Cells in T Cell Development. *J Immunol.* 2021;206(2):249-256.
88. Efremova M, Vento-Tormo M, Teichmann SA, Vento-Tormo R. CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat Protoc.* 2020;15(4):1484-1506.
89. Liang H, Coles AH, Zhu Z, et al. Noncanonical Wnt signaling promotes apoptosis in thymocyte development. *J Exp Med.* 2007;204(13):3077-3084.
90. Endo M, Kamizaki K, Minami Y. The Ror-Family Receptors in Development, Tissue Regeneration and Age-Related Disease. *Front Cell Dev Biol.* 2022;10:891763.
91. Komiya Y, Habas R. Wnt signal transduction pathways. *Organogenesis.* 2008;4(2):68-75.
92. Scheibner K, Bakhti M, Bastidas-Ponce A, Lickert H. Wnt signaling: implications in endoderm development and pancreas organogenesis. *Curr Opin Cell Biol.* 2019;61:48-55.
93. Nusser A, Sagar, Swann JB, et al. Developmental dynamics of two bipotent thymic epithelial progenitor types. *Nature.* 2022;606(7912):165-171.
94. Uttamsingh S, Bao X, Nguyen KT, et al. Synergistic effect between EGF and TGF-beta1 in inducing oncogenic properties of intestinal epithelial cells. *Oncogene.* 2008;27(18):2626-2634.
95. Wu X, Sun M, Yang Z, et al. The Roles of CCR9/CCL25 in Inflammation and Inflammation-Associated Diseases. *Front Cell Dev Biol.* 2021;9:686548.
96. Leick M, Catusse J, Follo M, et al. CCL19 is a specific ligand of the constitutively recycling atypical human chemokine receptor CRAM-B. *Immunology.* 2010;129(4):536-546.
97. Liu D, Kousa AI, O'Neill KE, et al. Canonical Notch signaling controls the early thymic epithelial progenitor cell state and emergence of the medullary epithelial lineage in fetal thymus development. *Development.* 2020;147(12).
98. Hozumi K, Mailhos C, Negishi N, et al. Delta-like 4 is indispensable in thymic environment specific for T cell development. *J Exp Med.* 2008;205(11):2507-2513.
99. Robey E. Regulation of T cell fate by Notch. *Annu Rev Immunol.* 1999;17:283-295.

100. Deftos ML, Bevan MJ. Notch signaling in T cell development. *Curr Opin Immunol.* 2000;12(2):166-172.
101. Radtke F, Wilson A, Stark G, et al. Deficient T cell fate specification in mice with an induced inactivation of Notch1. *Immunity.* 1999;10(5):547-558.
102. Pui JC, Allman D, Xu L, et al. Notch1 expression in early lymphopoiesis influences B versus T lineage determination. *Immunity.* 1999;11(3):299-308.
103. Felli MP, Maroder M, Mitsiadis TA, et al. Expression pattern of notch1, 2 and 3 and Jagged1 and 2 in lymphoid and stromal thymus components: distinct ligand-receptor interactions in intrathymic T cell development. *Int Immunol.* 1999;11(7):1017-1025.
104. Washburn T, Schweighoffer E, Gridley T, et al. Notch activity influences the alphabeta versus gammadelta T cell lineage decision. *Cell.* 1997;88(6):833-843.
105. Deftos ML, He YW, Ojala EW, Bevan MJ. Correlating notch signaling with thymocyte maturation. *Immunity.* 1998;9(6):777-786.
106. Jehn BM, Bielke W, Pear WS, Osborne BA. Cutting edge: protective effects of notch-1 on TCR-induced apoptosis. *J Immunol.* 1999;162(2):635-638.
107. Robey E, Chang D, Itano A, et al. An activated form of Notch influences the choice between CD4 and CD8 T cell lineages. *Cell.* 1996;87(3):483-492.
108. Izon DJ, Punt JA, Xu L, et al. Notch1 regulates maturation of CD4+ and CD8+ thymocytes by modulating TCR signal strength. *Immunity.* 2001;14(3):253-264.
109. Jimenez E, Vicente A, Sacedon R, et al. Distinct mechanisms contribute to generate and change the CD4:CD8 cell ratio during thymus development: a role for the Notch ligand, Jagged1. *J Immunol.* 2001;166(10):5898-5908.
110. Nueda ML, Baladron V, Garcia-Ramirez JJ, et al. The novel gene EGFL9/Dlk2, highly homologous to Dlk1, functions as a modulator of adipogenesis. *J Mol Biol.* 2007;367(5):1270-1280.
111. Bauer SR, Ruiz-Hidalgo MJ, Rudikoff EK, Goldstein J, Laborda J. Modulated expression of the epidermal growth factor-like homeotic protein dlk influences stromal-cell-pre-B-cell interactions, stromal cell adipogenesis, and pre-B-cell interleukin-7 requirements. *Mol Cell Biol.* 1998;18(9):5247-5255.
112. Smas CM, Sul HS. Pref-1, a protein containing EGF-like repeats, inhibits adipocyte differentiation. *Cell.* 1993;73(4):725-734.
113. Garces C, Ruiz-Hidalgo MJ, Font de Mora J, et al. Notch-1 controls the expression of fatty acid-activated transcription factors and is required for adipogenesis. *J Biol Chem.* 1997;272(47):29729-29734.
114. Calandra T, Roger T. Macrophage migration inhibitory factor: a regulator of innate immunity. *Nat Rev Immunol.* 2003;3(10):791-800.
115. Roger T, David J, Glauser MP, Calandra T. MIF regulates innate immune responses through modulation of Toll-like receptor 4. *Nature.* 2001;414(6866):920-924.
116. Bernhagen J, Krohn R, Lue H, et al. MIF is a noncognate ligand of CXC chemokine receptors in inflammatory and atherogenic cell recruitment. *Nature Medicine.* 2007;13(5):587-596.
117. Starlets D, Gore Y, Binsky I, et al. Cell-surface CD74 initiates a signaling cascade leading to cell proliferation and survival. *Blood.* 2006;107(12):4807-4816.
118. Gil-Yarom N, Radomir L, Sever L, et al. CD74 is a novel transcription regulator. *Proc Natl Acad Sci U S A.* 2017;114(3):562-567.
119. Wang HX, Zhang Q, Zhang J, et al. CD74 regulates cellularity and maturation of medullary thymic epithelial cells partially by activating the canonical NF-kappaB signaling pathway. *FASEB J.* 2021;35(5):e21535.

120. Goldfarb Y, Kadouri N, Levi B, et al. HDAC3 Is a Master Regulator of mTEC Development. *Cell Rep.* 2016;15(3):651-665.
121. Zhang JJ, Zhao Y, Chait BT, et al. Ser727-dependent recruitment of MCM5 by Stat1alpha in IFN-gamma-induced transcriptional activation. *EMBO J.* 1998;17(23):6963-6971.
122. DaFonseca CJ, Shu F, Zhang JJ. Identification of two residues in MCM5 critical for the assembly of MCM complexes and Stat1-mediated transcription activation in response to IFN-gamma. *Proc Natl Acad Sci U S A.* 2001;98(6):3034-3039.
123. Liu D, Wu D, Zhao L, et al. Arsenic Trioxide Reduces Global Histone H4 Acetylation at Lysine 16 through Direct Binding to Histone Acetyltransferase hMOF in Human Cells. *PLoS One.* 2015;10(10):e0141014.
124. Bhaskara S, Jacques V, Rusche JR, Olson EN, Cairns BR, Chandrasekharan MB. Histone deacetylases 1 and 2 maintain S-phase chromatin and DNA replication fork progression. *Epigenetics Chromatin.* 2013;6(1):27.
125. White AJ, Parnell SM, Handel A, et al. Diversity in Cortical Thymic Epithelial Cells Occurs through Loss of a Foxn1-Dependent Gene Signature Driven by Stage-Specific Thymocyte Cross-Talk. *J Immunol.* 2022.
126. Rizzoti K, Lovell-Badge R. SOX3 activity during pharyngeal segmentation is required for craniofacial morphogenesis. *Development.* 2007;134(19):3437-3448.
127. Wei Q, Condie BG. A focused in situ hybridization screen identifies candidate transcriptional regulators of thymic epithelial cell development and function. *PLoS One.* 2011;6(11):e26795.
128. Akiyama T, Shimo Y, Yanai H, et al. The tumor necrosis factor family receptors RANK and CD40 cooperatively establish the thymic medullary microenvironment and self-tolerance. *Immunity.* 2008;29(3):423-437.
129. Boehm T, Scheu S, Pfeffer K, Bleul CC. Thymic medullary epithelial cell differentiation, thymocyte emigration, and the control of autoimmunity require lympho-epithelial cross talk via LTbetaR. *J Exp Med.* 2003;198(5):757-769.
130. Burkly L, Hession C, Ogata L, et al. Expression of relB is required for the development of thymic medulla and dendritic cells. *Nature.* 1995;373(6514):531-536.
131. Akiyama T, Maeda S, Yamane S, et al. Dependence of self-tolerance on TRAF6-directed development of thymic stroma. *Science.* 2005;308(5719):248-251.
132. Kinoshita D, Hirota F, Kaisho T, et al. Essential role of IkappaB kinase alpha in thymic organogenesis required for the establishment of self-tolerance. *J Immunol.* 2006;176(7):3995-4002.
133. Chelbi-Alix MK, Pelicano L. Retinoic acid and interferon signaling cross talk in normal and RA-resistant APL cells. *Leukemia.* 1999;13(8):1167-1174.
134. Pelicano L, Li F, Schindler C, Chelbi-Alix MK. Retinoic acid enhances the expression of interferon-induced proteins: evidence for multiple mechanisms of action. *Oncogene.* 1997;15(19):2349-2359.
135. Matikainen S, Ronni T, Hurme M, Pine R, Julkunen I. Retinoic acid activates interferon regulatory factor-1 gene expression in myeloid cells. *Blood.* 1996;88(1):114-123.
136. Kolla V, Lindner DJ, Xiao W, Borden EC, Kalvakolanu DV. Modulation of interferon (IFN)-inducible gene expression by retinoic acid. Up-regulation of STAT1 protein in IFN-unresponsive cells. *J Biol Chem.* 1996;271(18):10508-10514.
137. Calandra T, Bernhagen J, Mitchell RA, Bucala R. The macrophage is an important and previously unrecognized source of macrophage migration inhibitory factor. *J Exp Med.* 1994;179(6):1895-1902.
138. Bacher M, Metz CN, Calandra T, et al. An essential regulatory role for macrophage migration inhibitory factor in T-cell activation. *Proc Natl Acad Sci U S A.* 1996;93(15):7849-7854.

139. Barker N, Tan S, Clevers H. Lgr proteins in epithelial stem cell biology. *Development.* 2013;140(12):2484-2494.
140. Tumbar T, Guasch G, Greco V, et al. Defining the epithelial stem cell niche in skin. *Science.* 2004;303(5656):359-363.
141. Cotsarelis G. Epithelial stem cells: a folliculocentric view. *J Invest Dermatol.* 2006;126(7):1459-1468.
142. Potten CS, Booth C, Pritchard DM. The intestinal epithelial stem cell: the mucosal governor. *Int J Exp Pathol.* 1997;78(4):219-243.

AUTHOR INFORMATION

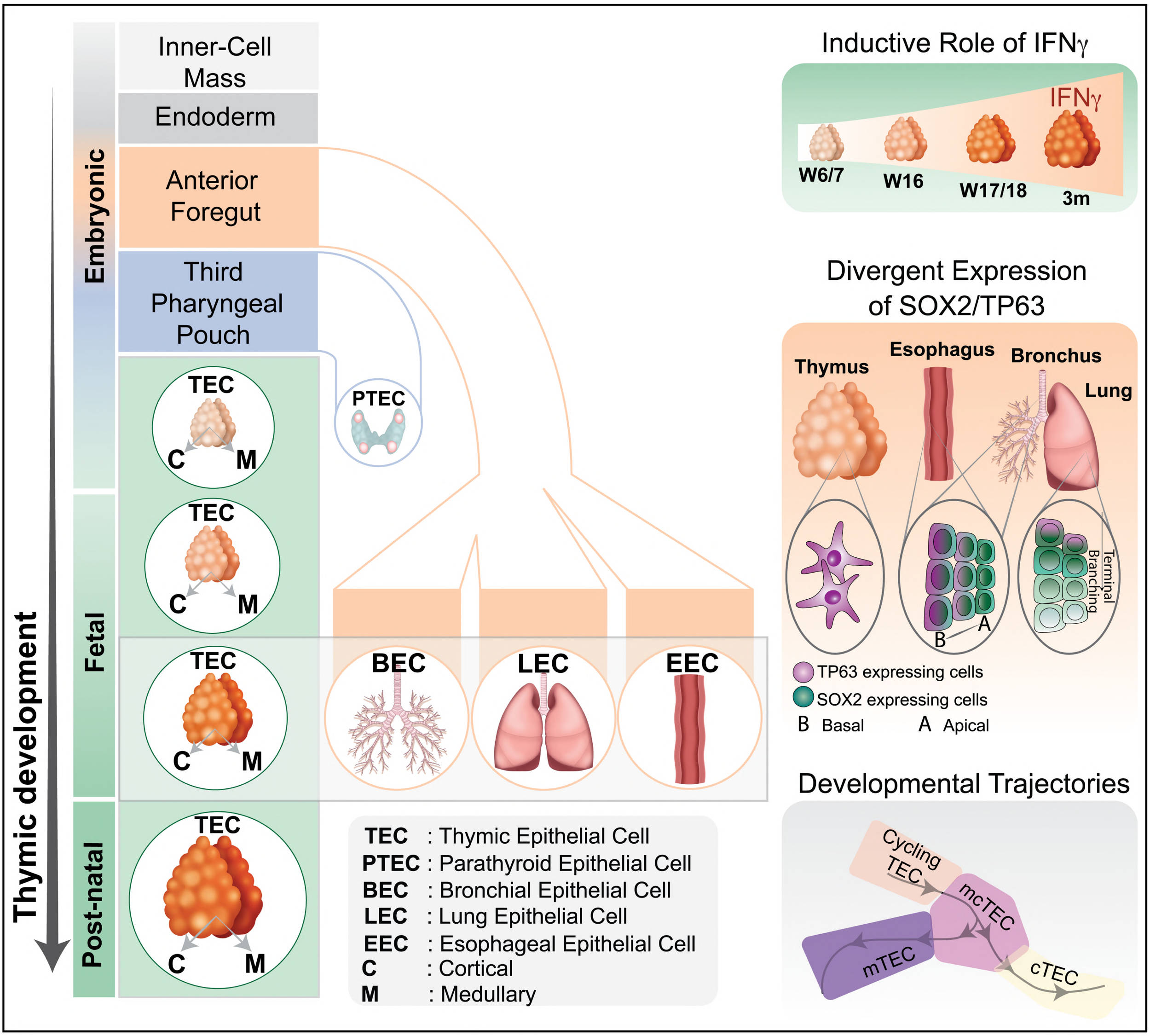
These authors contributed equally: Abdulvasey Mohammed and Ben Solomon
These authors contributed equally; correspondence should be addressed to: Andrew Gentles and Katja G. Weinacht

CONTRIBUTIONS

A.M., B.S., P.F.S, A.G., and K.G.W. interpreted the data and wrote the manuscript. The experiments were designed, performed and analyzed by B.S., A.M., A.G., and K.G.W.. P.F.S, K.M.H., M.C. and C.Y.Y. performed additional experiments and bioinformatic data processing. V.W., S.T., J.P. and V.K. provided samples and analytical input. G.A.H. and P.K. reviewed design, bioinformatic analysis and edited the manuscript.

ACKNOWLEDGEMENTS

We thank the anonymous donors of the Stanford 22q11 Deletions Syndrome Program for their generous support of our work. We thank all Stanford 22q11 Deletion Syndrome Program directors and advisory board (Maria Grazia Roncarolo, Irving Weissman, Gay Crooks, Antonio Baldini) and consortium members (Ken Weinberg, Elizabeth Illingworth, Rosa Bacchetta, and Andrea Cipriani), for insightful input and discussions. scRNA library preparation and sequencing by Hui Gai and MedGenome, Inc.. We thank Riley Suhar, PhD, for her contribution in some of the graphical illustrations.



ABSTRACT FIGURE

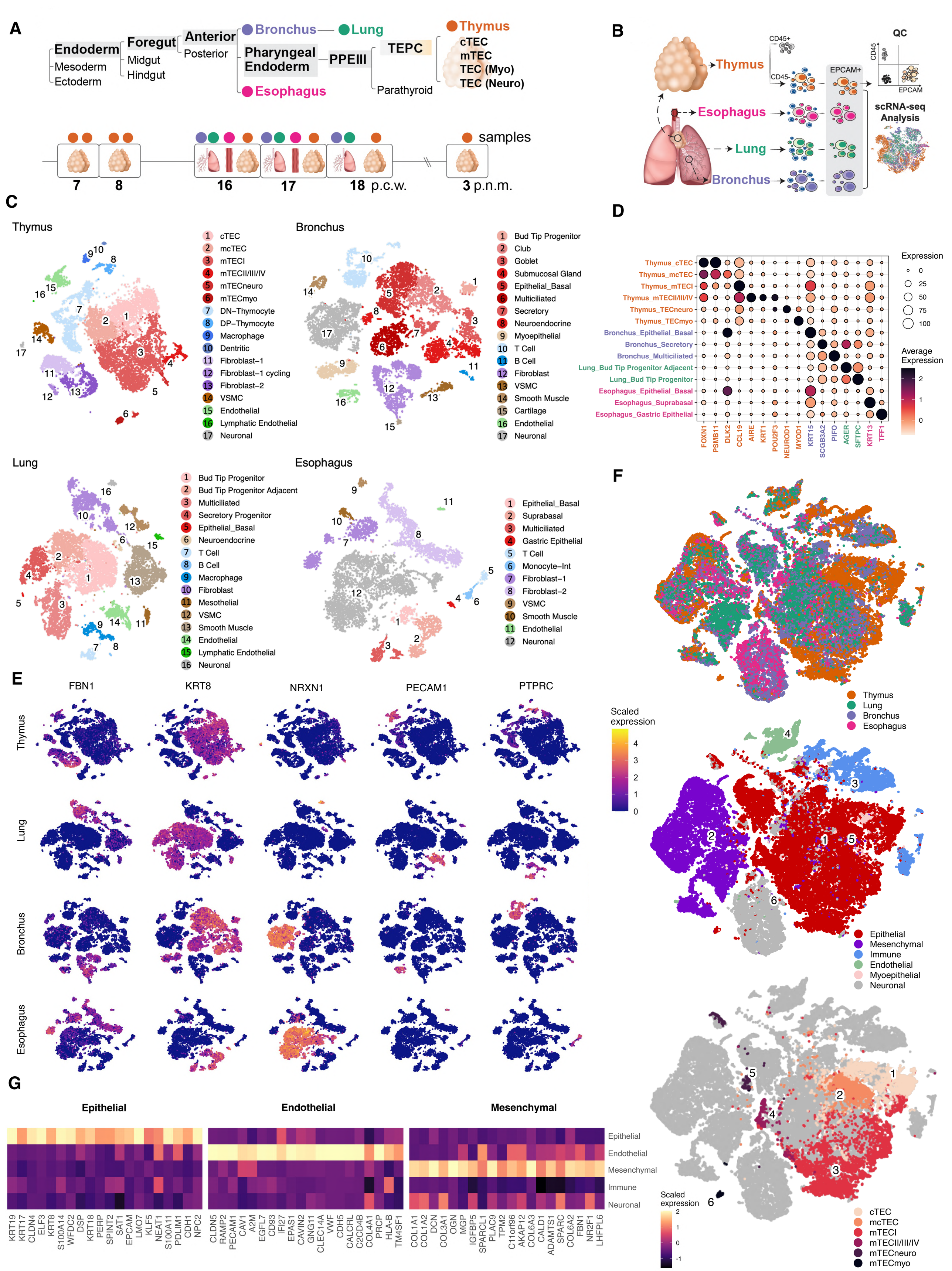

A
Endoderm
Mesoderm
Ectoderm
Foregut
Midgut
Hindgut
Anterior
Posterior
Bronchus
Lung
Pharyngeal Endoderm
PPEIII
TEPC
Parathyroid
Esophagus
Thymus
cTEC
mTEC
TEC (Myo)
TEC (Neuro)
samples
7
8
16
17
18 p.c.w.
3 p.n.m.
B
Thymus
CD45+
CD45-
EPCAM+
QC
CD45
EPCAM
Esophagus
Lung
Bronchus
scRNA-seq Analysis
C
Thymus
1 cTEC
2 mcTEC
3 mTECI
4 mTECII/III/IV
5 mTECneuro
6 mTECmyo
7 DN–Thymocyte
8 DP–Thymocyte
9 Macrophage
10 Dentritic
11 Fibroblast–1
12 Fibroblast–1 cycling
13 Fibroblast–2
14 VSMC
15 Endothelial
16 Lymphatic Endothelial
17 Neuronal
Bronchus
1 Bud Tip Progenitor
2 Club
3 Goblet
4 Submucosal Gland
5 Epithelial_Basal
6 Multiciliated
7 Secretory
8 Neuroendocrine
9 Myoepithelial
10 T Cell
11 B Cell
12 Fibroblast
13 VSMC
14 Smooth Muscle
15 Cartilage
16 Endothelial
17 Neuronal
Lung
1 Bud Tip Progenitor
2 Bud Tip Progenitor Adjacent
3 Multiciliated
4 Secretory Progenitor
5 Epithelial_Basal
6 Neuroendocrine
7 T Cell
8 B Cell
9 Macrophage
10 Fibroblast
11 Mesothelial
12 VSMC
13 Smooth Muscle
14 Endothelial
15 Lymphatic Endothelial
16 Neuronal
Esophagus
1 Epithelial_Basal
2 Suprabasal
3 Multiciliated
4 Gastric Epithelial
5 T Cell
6 Monocyte–Int
7 Fibroblast–1
8 Fibroblast–2
9 VSMC
10 Smooth Muscle
11 Endothelial
12 Neuronal
D
Thymus_cTEC
Thymus_mcTEC
Thymus_mTECI
Thymus_mTECII/III/IV
Thymus_TECneuro
Thymus_TECmyo
Bronchus_Epithelial_Basal
Bronchus_Secretory
Bronchus_Multiciliated
Lung_Bud Tip Progenitor Adjacent
Lung_Bud Tip Progenitor
Esophagus_Epithelial_Basal
Esophagus_Suprabasal
Esophagus_Gastric Epithelial
FOXN1
PSMB11
DLK2
CCL19
AIRE
KRT1
POU2F3
NEUROD1
MYOD1
KRT15
SCGB3A2
PIFO
AGER
SFTPC
KRT13
TFF1
Expression
0
25
50
75
100
Average Expression
2
1
0
E
FBN1
KRT8
NRXN1
PECAM1
PTPRC
Thymus
Lung
Bronchus
Esophagus
Scaled expression
4
3
2
1
0
F
Thymus
Lung
Bronchus
Esophagus
Epithelial
Mesenchymal
Immune
Endothelial
Myoepithelial
Neuronal
Scaled expression
2
1
0
-1
cTEC
mcTEC
mTECI
mTECII/III/IV
mTECneuro
mTECmyo
G
Epithelial
Endothelial
Mesenchymal
KRT19
KRT17
CLDN4
ELF3
KRT8
S100A14
WFDC2
DSP
KRT18
PERP
SPINT2
SAT1
EPCAM
LMO7
KLF5
NEAT1
S100A11
PDLIM1
CDH1
NPC2
CLDN5
RAMP2
PECAM1
CAV1
A2M
EGFL7
CD93
IFI27
EPAS1
CAVIN2
GNG11
CLEC14A
VWF
CDH5
CALCRL
C2CD4B
COL4A1
PRCP
HLA-B
TM4SF1
COL1A1
COL1A2
DCN
COL3A1
OGN
MGP
IGFBP5
SPARCL1
PLAC9
TPM2
C11orf96
AKAP12
COL6A3
CALD1
ADAMTS1
SPARC
COL6A2
FBN1
NR2F1
LHFPL6
Epithelial
Endothelial
Mesenchymal
Immune
Neuronal

Figure 1. Dissecting and integrating the gene expression profile of fetal derived anterior foregut tissues using scRNAseq.

(A) Sequential embryonic development showing the foregut development into the pharyngeal apparatus focusing on the thymic ontogeny, including the bipotent third (III) pharyngeal pouch endoderm (PPEIII), ventralization of the PPE (PPEv), thymic epithelial progenitor cell development (TEPC), and the thymus (medullary thymic epithelial cells, mTEC; cortical TEC, cTEC; neuroTEC; myoTEC). (B) Scheme of sequential cell digestion and cell separation of embryonic Thymus, Esophagus, Bronchus and Lungs for quality control check and single cell RNA sequencing library preparation. (C) tSNE visualization of epithelial cells colored by cell type from individual anterior foregut tissues (D) Dot plot showing top expressing genes within selected anterior foregut tissue epithelial cell type populations. Color representative of average expression level. Dot size representative of percentage of cells in cell type expressing selected marker gene. (E) tSNE visualization of cells from individual anterior foregut tissues overlaid with expression pattern of genes representative of specific cell classes. (F) tSNE visualization of cells colored by cell type from individual anterior foregut tissues. (G) Heatmap of top differentially expressed genes in specific cell types across all AFG-derived organs.

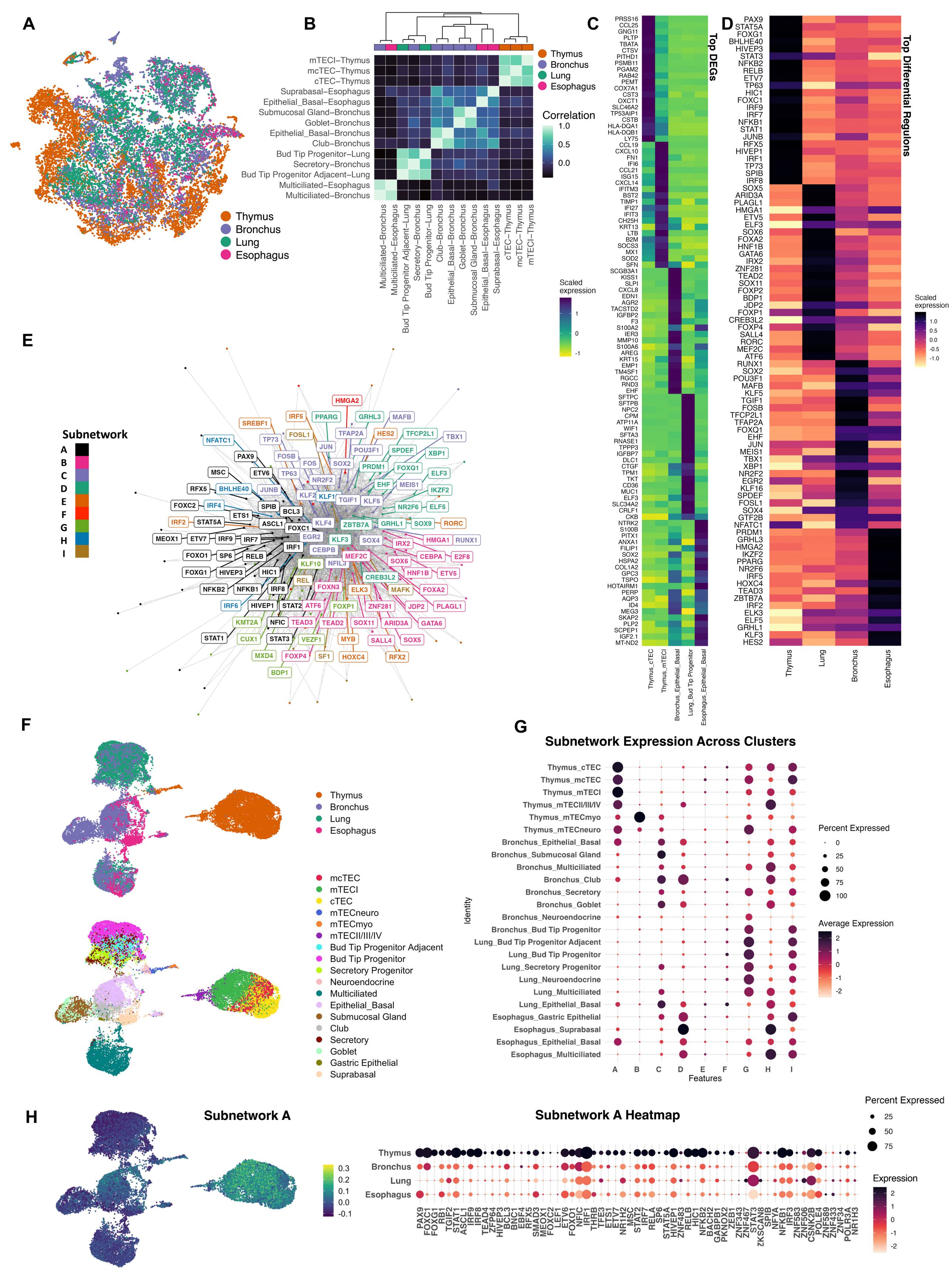

A
Thymus
Bronchus
Lung
Esophagus
B
mTECI–Thymus
mcTEC–Thymus
cTEC–Thymus
Suprabasal–Esophagus
Epithelial_Basal–Esophagus
Submucosal Gland–Bronchus
Goblet–Bronchus
Epithelial_Basal–Bronchus
Club–Bronchus
Bud Tip Progenitor–Lung
Secretory–Bronchus
Bud Tip Progenitor Adjacent–Lung
Multiciliated–Esophagus
Multiciliated–Bronchus
Correlation
1.0
0.5
0.0
Scaled expression
C
Top DEGs
D
Top Differential Regulons
Scaled expression
Thymus_cTEC
Thymus_mTECI
Bronchus_Epithelial_Basal
Lung_Bud Tip Progenitor
Esophagus_Epithelial_Basal
Thymus
Lung
Bronchus
Esophagus
E
Subnetwork
A
B
C
D
E
F
G
H
I
F
Thymus
Bronchus
Lung
Esophagus
mcTEC
mTECI
cTEC
mTECneuro
mTECmyo
mTECII/III/IV
Bud Tip Progenitor Adjacent
Bud Tip Progenitor
Secretory Progenitor
Neuroendocrine
Multiciliated
Epithelial_Basal
Submucosal Gland
Club
Secretory
Goblet
Gastric Epithelial
Suprabasal
G
Subnetwork Expression Across Clusters
Thymus_cTEC
Thymus_mcTEC
Thymus_mTECI
Thymus_mTECII/III/IV
Thymus_mTECmyo
Thymus_mTECneuro
Bronchus_Epithelial_Basal
Bronchus_Submucosal Gland
Bronchus_Multiciliated
Bronchus_Club
Bronchus_Secretory
Bronchus_Goblet
Bronchus_Neuroendocrine
Bronchus_Bud Tip Progenitor
Lung_Bud Tip Progenitor Adjacent
Lung_Bud Tip Progenitor
Lung_Secretory Progenitor
Lung_Neuroendocrine
Lung_Multiciliated
Lung_Epithelial_Basal
Esophagus_Gastric Epithelial
Esophagus_Suprabasal
Esophagus_Epithelial_Basal
Esophagus_Multiciliated
Identity
Features
Percent Expressed
Average Expression
H
Subnetwork A
Subnetwork A Heatmap
Thymus
Bronchus
Lung
Esophagus
Percent Expressed
Expression

Figure 2. Differential and regulatory analysis of all AFG-derived epithelial cells reveals unique thymic inflammatory signature. (A) tSNE visualization of epithelial cells from all anterior foregut tissues colored by tissue of origin. (B) Correlation matrix plot comparing anterior foregut tissues. (C) Heatmap of differentially expressed genes across key AFG-derived epithelial cell subsets. (D) Heatmap of differentially expressed SCENIC regulons across AFG derived tissues. (E) Network visualization of transcriptional factors derived from SCENIC specified regulons. Transcription factors are classified according to leiden based subnetwork identification. (F) UMAP visualization of SCENIC regulon AUC scores of all epithelial cells from AFG-derived tissue classified by tissue and AFG epithelial cell types. (G) Dot plot representation of subnetwork specific composite regulon expression across AFG epithelial cell types. Color representative of average expression level. Dot size representative of percentage of cells in cell type expressing selected marker gene. (H) UMAP visualization of Subnetwork A regulon expression accompanied by a dot plot of Subnetwork A specific TF transcriptional expression across AFG derived tissues. Color representative of average expression level within dot plot whereas size is representative of percentage of cells in cell type expressing selected marker gene.

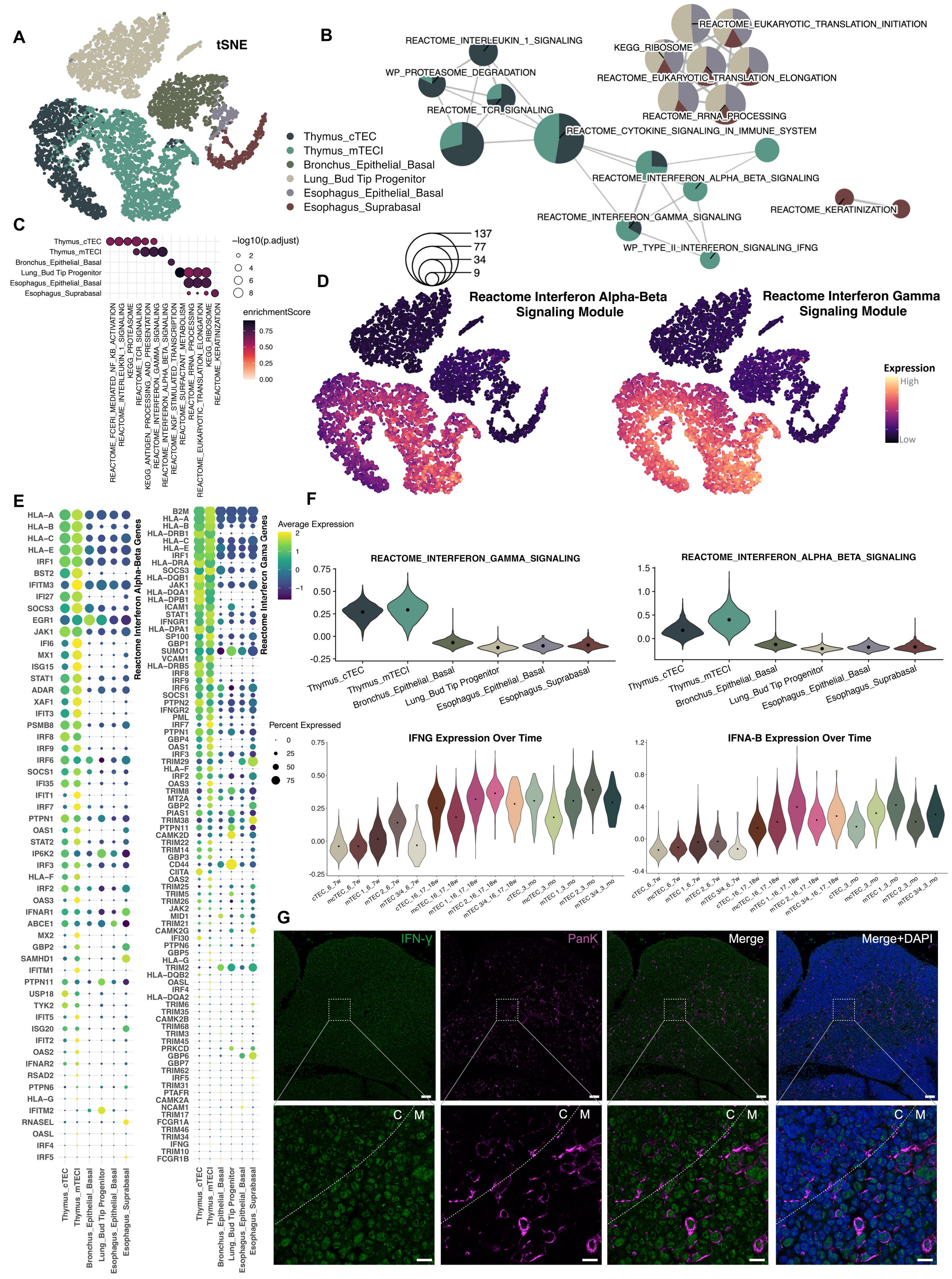

A
tSNE
Thymus_cTEC
Thymus_mTECI
Bronchus_Epithelial_Basal
Lung_Bud Tip Progenitor
Esophagus_Epithelial_Basal
Esophagus_Suprabasal
B
REACTOME_INTERLEUKIN_1_SIGNALING
WP_PROTEASOME_DEGRADATION
REACTOME_TCR_SIGNALING
REACTOME_EUKARYOTIC_TRANSLATION_INITIATION
KEGG_RIBOSOME
REACTOME_EUKARYOTIC_TRANSLATION_ELONGATION
REACTOME_RRNA_PROCESSING
REACTOME_CYTOKINE_SIGNALING_IN_IMMUNE_SYSTEM
REACTOME_INTERFERON_ALPHA_BETA_SIGNALING
REACTOME_KERATINIZATION
REACTOME_INTERFERON_GAMMA_SIGNALING
WP_TYPE_II_INTERFERON_SIGNALING_IFNG
137
77
34
9
C
Thymus_cTEC
Thymus_mTECI
Bronchus_Epithelial_Basal
Lung_Bud Tip Progenitor
Esophagus_Epithelial_Basal
Esophagus_Suprabasal
–log10(p.adjust)
2
4
6
8
enrichmentScore
0.75
0.50
0.25
0.00
REACTOME_FCERI_MEDIATED_NF_KB_ACTIVATION
REACTOME_INTERLEUKIN_1_SIGNALING
KEGG_PROTEASOME
REACTOME_TCR_SIGNALING
KEGG_ANTIGEN_PROCESSING_AND_PRESENTATION
REACTOME_INTERFERON_GAMMA_SIGNALING
REACTOME_INTERFERON_ALPHA_BETA_SIGNALING
REACTOME_NGF_STIMULATED_TRANSCRIPTION
REACTOME_SURFACTANT_METABOLISM
REACTOME_RRNA_PROCESSING
REACTOME_EUKARYOTIC_TRANSLATION_ELONGATION
KEGG_RIBOSOME
REACTOME_KERATINIZATION
D
Reactome Interferon Alpha-Beta Signaling Module
Reactome Interferon Gamma Signaling Module
Expression
High
Low
E
Reactome Interferon Alpha-Beta Genes
HLA-A
HLA-B
HLA-C
HLA-E
IRF1
BST2
IFITM3
IFI27
SOCS3
EGR1
JAK1
IFI6
MX1
ISG15
STAT1
ADAR
XAF1
IFIT3
PSMB8
IRF8
IRF9
IRF6
SOCS1
IFI35
IFIT1
IRF7
PTPN1
OAS1
STAT2
IP6K2
IRF3
HLA-F
IRF2
OAS3
IFNAR1
ABCE1
MX2
GBP2
SAMHD1
IFITM1
PTPN11
USP18
TYK2
IFIT5
ISG20
IFIT2
OAS2
IFNAR2
RSAD2
PTPN6
HLA-G
IFITM2
RNASEL
OASL
IRF4
IRF5
Reactome Interferon Gama Genes
B2M
HLA-A
HLA-B
HLA-DRB1
HLA-C
HLA-E
IRF1
HLA-DRA
SOCS3
HLA-DQB1
JAK1
HLA-DQA1
HLA-DPB1
ICAM1
STAT1
IFNGR1
HLA-DPA1
SP100
GBP1
SUMO1
VCAM1
HLA-DRB5
IRF8
IRF9
IRF6
SOCS1
PTPN2
IFNGR2
PML
IRF7
PTPN1
GBP4
OAS1
IRF3
TRIM29
HLA-F
IRF2
OAS3
TRIM8
MT2A
GBP2
PIAS1
TRIM38
PTPN11
CAMK2D
TRIM22
TRIM14
GBP3
CD44
CIITA
OAS2
TRIM25
TRIM5
TRIM26
JAK2
MID1
TRIM21
CAMK2G
IFI30
PTPN6
GBP5
HLA-G
TRIM2
HLA-DQB2
OASL
IRF4
HLA-DQA2
TRIM6
TRIM35
CAMK2B
TRIM68
TRIM3
TRIM45
PRKCD
GBP6
GBP7
TRIM62
IRF5
TRIM31
PTAFR
CAMK2A
NCAM1
TRIM17
FCGR1A
TRIM46
TRIM34
IFNG
TRIM10
FCGR1B
Thymus_cTEC
Thymus_mTECI
Bronchus_Epithelial_Basal
Lung_Bud Tip Progenitor
Esophagus_Epithelial_Basal
Esophagus_Suprabasal
Average Expression
2
1
0
−1
Percent Expressed
0
25
50
75
F
REACTOME_INTERFERON_GAMMA_SIGNALING
REACTOME_INTERFERON_ALPHA_BETA_SIGNALING
IFNG Expression Over Time
IFNA-B Expression Over Time
cTEC_6_7w
mcTEC_6_7w
mTEC 1_6_7w
mTEC 2_6_7w
mTEC 3/4_6_7w
cTEC_16_17_18w
mcTEC_16_17_18w
mTEC 1_16_17_18w
mTEC 2_16_17_18w
mTEC 3/4_16_17_18w
cTEC_3_mo
mcTEC_3_mo
mTEC 1_3_mo
mTEC 2_3_mo
mTEC 3/4_3_mo
G
IFN-γ
PanK
Merge
Merge+DAPI
C
M

Figure 3. Interferon type 1 and 2 signaling pathways are enriched in cortical and medullary TECs compared to other AFG-derived epithelial cell types. (A) tSNE visualization of key AFG-derived epithelial cell type populations. (B) Enrichment map of differentially expressed genes clustered according to shared genes across gene subsets of key AFG-derived epithelial cell type populations. Node size reflects the number of genes within the gene set. (C) Dot plots of gene set enrichment across selected anterior foregut epithelial cell types. Only selected gene sets are included. Color representative of gene set enrichment score. Dot size representative of percentage of -value. (D) tSNE visualization of Reactome IFN gene set module expression of key AFG-derived epithelial cell type populations. (E) Dot plot representations of Reactome IFN gene set expression of key AFG-derived epithelial cell type populations. Color representative of average expression level. Dot size representative of percentage of cells in cell type expressing selected marker gene. (F) Violin lot representation of Reactome IFN gene set module expression of key AFG-derived epithelial cell type populations accompanied by expression over different time points in key thymic epithelial cell types. (G) Representative immunofluorescence confocal microscopy image of a human fetal thymus stained for IFN (green), PanKeratin (PanK; magenta), and DAPI (blue). C: Cortex; M: Medulla. Scale bars: 50um and 10um for full scaled and zoomed in images, respectively.

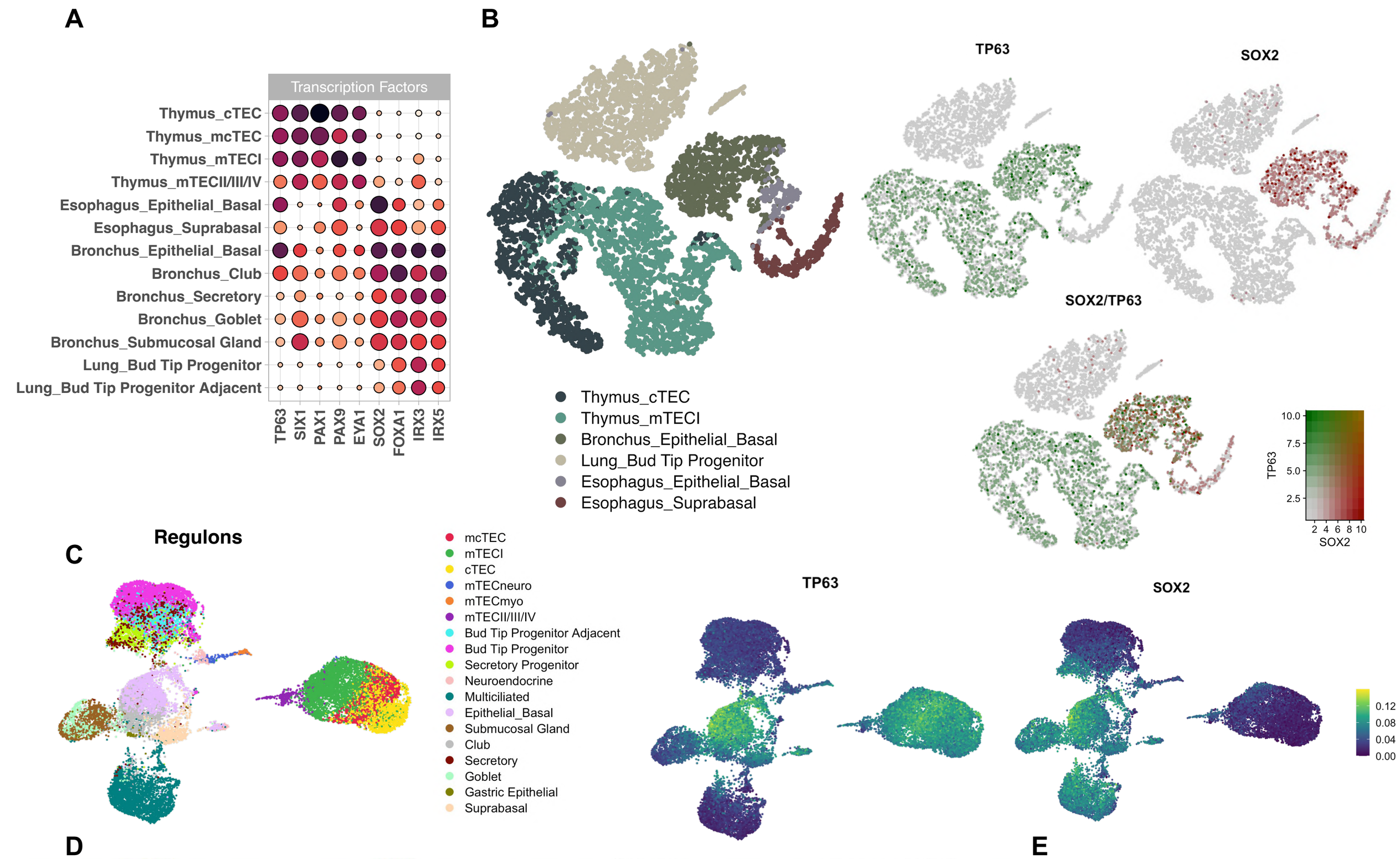

A
Transcription Factors
Thymus_cTEC
Thymus_mcTEC
Thymus_mTECI
Thymus_mTECII/III/IV
Esophagus_Epithelial_Basal
Esophagus_Suprabasal
Bronchus_Epithelial_Basal
Bronchus_Club
Bronchus_Secretory
Bronchus_Goblet
Bronchus_Submucosal Gland
Lung_Bud Tip Progenitor
Lung_Bud Tip Progenitor Adjacent
TP63
SIX1
PAX1
PAX9
EYA1
SOX2
FOXA1
IRX3
IRX5
B
TP63
SOX2
SOX2/TP63
Thymus_cTEC
Thymus_mTECI
Bronchus_Epithelial_Basal
Lung_Bud Tip Progenitor
Esophagus_Epithelial_Basal
Esophagus_Suprabasal
TP63
SOX2
C
Regulons
mcTEC
mTECI
cTEC
mTECneuro
mTECmyo
mTECII/III/IV
Bud Tip Progenitor Adjacent
Bud Tip Progenitor
Secretory Progenitor
Neuroendocrine
Multiciliated
Epithelial_Basal
Submucosal Gland
Club
Secretory
Goblet
Gastric Epithelial
Suprabasal
TP63
SOX2
D
SOX2
P63
Merge
Merge+DAPI
Thymus
Bronchus
Lung
Esophagus
E
Chemokines
CXCL14
CCL21
CCL19
CCL2
CXCL12
CCL25
Epithelial Keratins
EPCAM
KRT13
KRT4
KRT6B
KRT15
KRT17
KRT14
KRT5
KRT8
ECM
HSPG2
LAMC1
LAMA5
LAMA3
LAMA4
LAMA2
COL1A2
COL3A1
COL7A1
LAMB3
LAMC2
LGALS1
COL17A1
ITGAV
COL6A2
COL6A1
COL4A1
VCAM1
VIM
FN1
Claudins
CLDN5
CLDN4
CLDN3
Thymus_cTEC
Thymus_mcTEC
Thymus_mTECI
Thymus_mTECII/III/IV
Esophagus_Epithelial_Basal
Esophagus_Suprabasal
Bronchus_Epithelial_Basal
Bronchus_Club
Bronchus_Secretory
Bronchus_Goblet
Bronchus_Submucosal Gland
Lung_Bud Tip Progenitor
Lung_Bud Tip Progenitor Adjacent
Average Expression
Percent Expression

Figure 4. Differential distribution of TP63 and SOX2 expression across AFG-derived epithelial cells. (A) Dot plot showing expression of genes from a selected gene set of transcription factors within selected anterior foregut tissue epithelial cell type populations. Color representative of average expression level. Dot size representative of percentage of cells in cell type expressing selected marker gene. (B) tSNE visualization of SOX2 and TP63 expression of key AFG-derived epithelial cells described in (4B). (C) SCENIC UMAP visualization of SOX2 and TP63 regulon expression of all AFG epithelial cells described in (2F). (D) Representative immunofluorescence confocal microscopy image of a human fetal thymus, bronchus, lung, and esophagus stained for SOX2 (green), P63 (magenta), and DAPI (blue). Scale bars: 50um. (E) Dot plot showing expression of genes from selected gene sets (claudins, ECM, epithelial keratins, chemokines) within selected anterior foregut tissue epithelial cell type populations. Color representative of average expression level. Dot size representative of percentage of cells in cell type expressing selected marker gene.

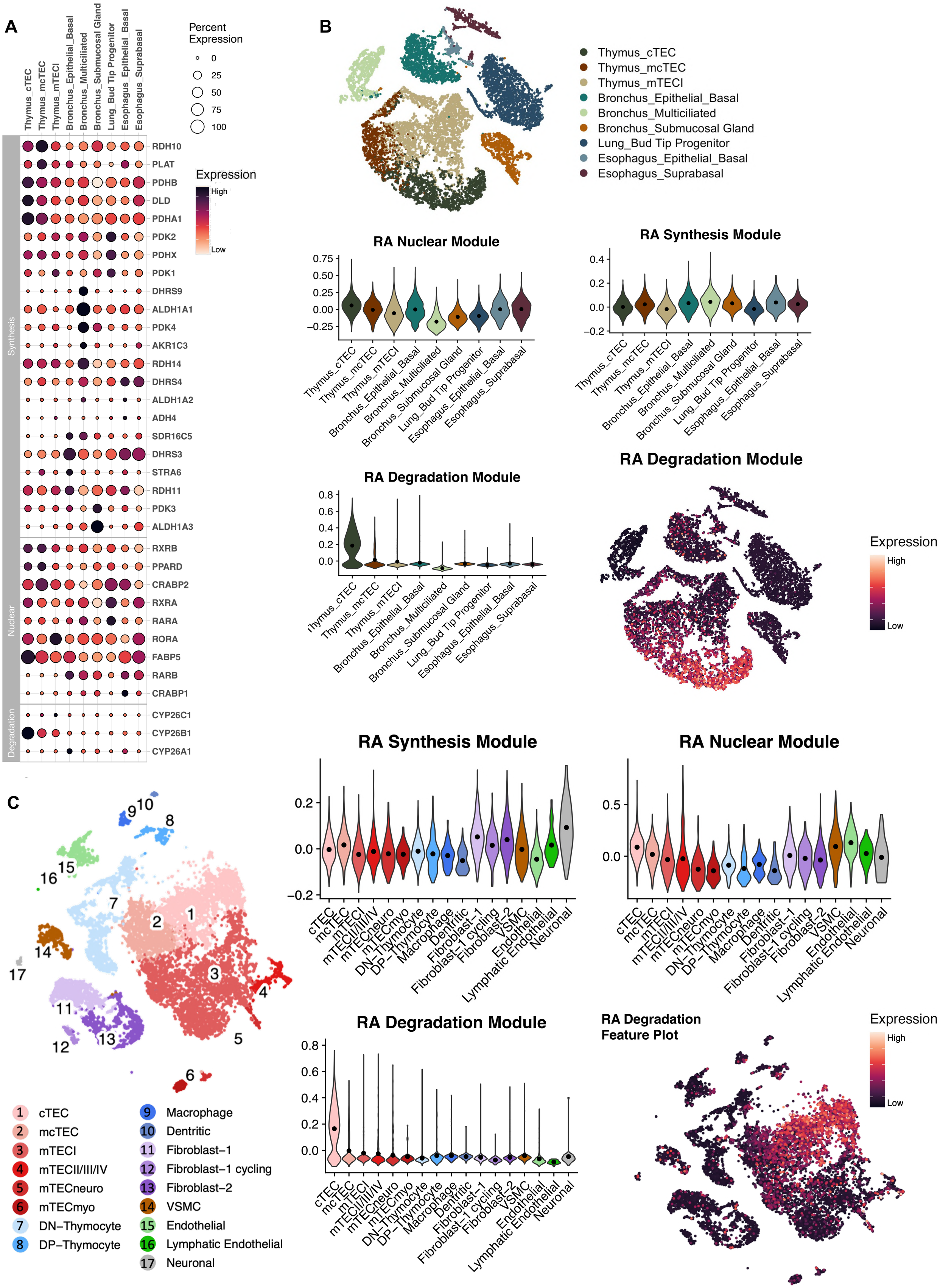

A
Thymus_cTEC
Thymus_mcTEC
Thymus_mTECI
Bronchus_Epithelial_Basal
Bronchus_Multiciliated
Bronchus_Submucosal Gland
Lung_Bud Tip Progenitor
Esophagus_Epithelial_Basal
Esophagus_Suprabasal
Percent Expression
0
25
50
75
100
Expression
High
Low
Synthesis
RDH10
PLAT
PDHB
DLD
PDHA1
PDK2
PDHX
PDK1
DHRS9
ALDH1A1
PDK4
AKR1C3
RDH14
DHRS4
ALDH1A2
ADH4
SDR16C5
DHRS3
STRA6
RDH11
PDK3
ALDH1A3
Nuclear
RXRB
PPARD
CRABP2
RXRA
RARA
RORA
FABP5
RARB
CRABP1
Degradation
CYP26C1
CYP26B1
CYP26A1
B
Thymus_cTEC
Thymus_mcTEC
Thymus_mTECI
Bronchus_Epithelial_Basal
Bronchus_Multiciliated
Bronchus_Submucosal Gland
Lung_Bud Tip Progenitor
Esophagus_Epithelial_Basal
Esophagus_Suprabasal
RA Nuclear Module
RA Synthesis Module
RA Degradation Module
RA Degradation Module
Expression
High
Low
C
1 cTEC
2 mcTEC
3 mTECI
4 mTECII/III/IV
5 mTECneuro
6 mTECmyo
7 DN−Thymocyte
8 DP−Thymocyte
9 Macrophage
10 Dentritic
11 Fibroblast−1
12 Fibroblast−1 cycling
13 Fibroblast−2
14 VSMC
15 Endothelial
16 Lymphatic Endothelial
17 Neuronal
RA Synthesis Module
RA Nuclear Module
RA Degradation Module
RA Degradation Feature Plot
Expression
High
Low

Figure 5. Regulation of retinoic acid signaling across AFG-derived is differential, specially within cortical thymic epithelial cells. (A) Dot plots of gene expression for curated genes involved in retinoic acid synthesis, nuclear signaling, and metabolism across select anterior foregut epithelial cell types. Color representative of average expression level. Dot size representative of percentage of cells in cell type expressing selected marker gene. (B) tSNE visualization of key AFG-derived epithelial cells accompanied by Violin plot representation of composite expression of curated genes from (4A) across identified cell types. Dot in Violin Plot represents mean module score. tSNE visualization of composite RA degradation module expression. Color representative of average expression level. (C) tSNE visualization of thymus cells colored by cell type as in Figure 1 focusing on RA modules. tSNE visualization of thymus cells overlaid with RA degradation module score for each cell as described in (4B). Color representative of average expression level.

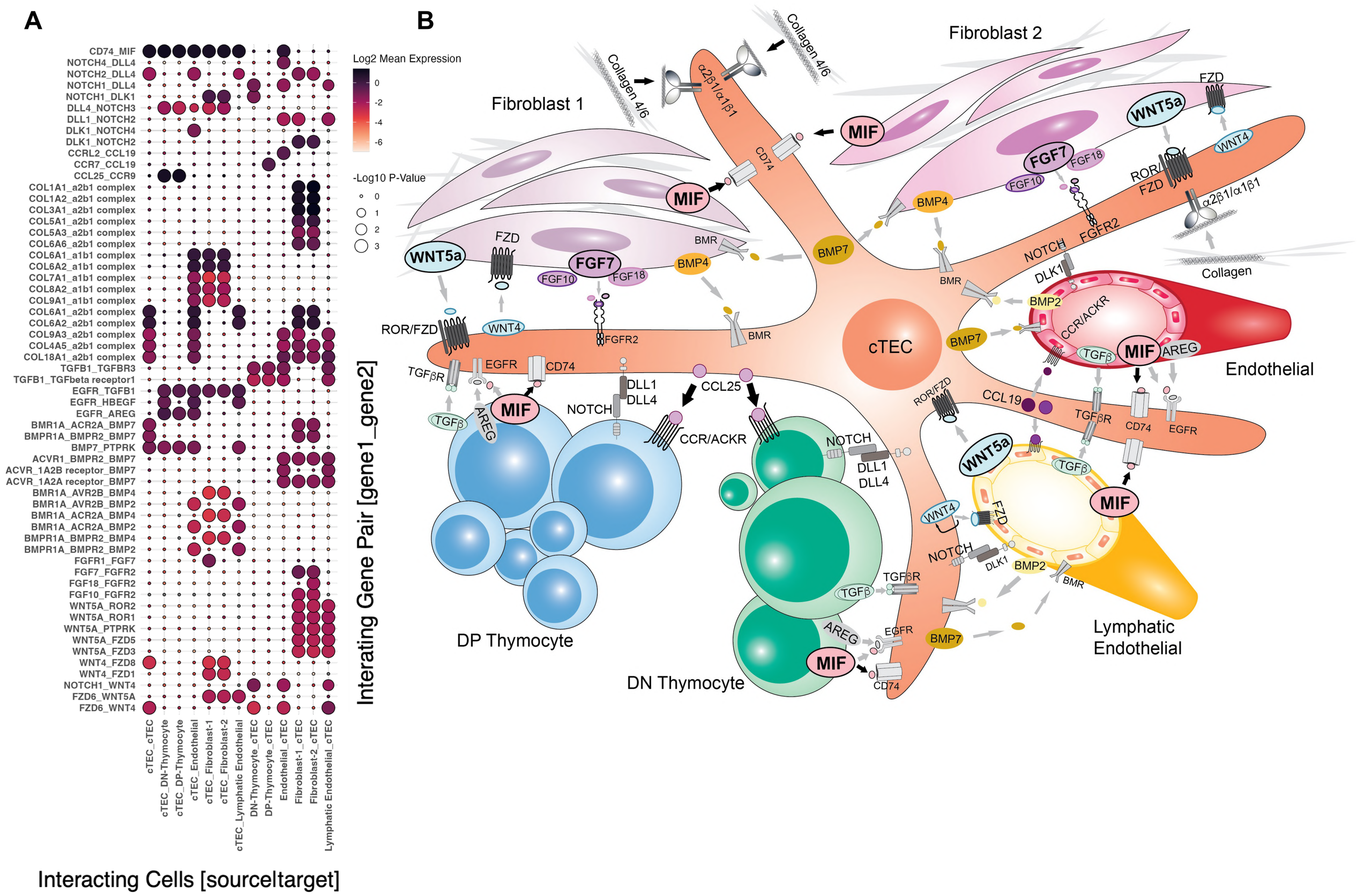

A
B
Log2 Mean Expression
0
-2
-4
-6
-Log10 P-Value
0
1
2
3
CD74_MIF
NOTCH4_DLL4
NOTCH2_DLL4
NOTCH1_DLL4
NOTCH1_DLK1
DLL4_NOTCH3
DLL1_NOTCH2
DLK1_NOTCH4
DLK1_NOTCH2
CCRL2_CCL19
CCR7_CCL19
CCL25_CCR9
COL1A1_a2b1 complex
COL1A2_a2b1 complex
COL3A1_a2b1 complex
COL5A1_a2b1 complex
COL5A3_a2b1 complex
COL6A6_a2b1 complex
COL6A1_a1b1 complex
COL6A2_a1b1 complex
COL7A1_a1b1 complex
COL8A2_a1b1 complex
COL9A1_a1b1 complex
COL6A1_a2b1 complex
COL6A2_a2b1 complex
COL9A3_a2b1 complex
COL4A5_a2b1 complex
COL18A1_a2b1 complex
TGFB1_TGFBR3
TGFB1_TGFbeta receptor1
EGFR_TGFB1
EGFR_HBEGF
EGFR_AREG
BMR1A_ACR2A_BMP7
BMPR1A_BMPR2_BMP7
BMP7_PTPRK
ACVR1_BMPR2_BMP7
ACVR_1A2B receptor_BMP7
ACVR_1A2A receptor_BMP7
BMR1A_AVR2B_BMP4
BMR1A_AVR2B_BMP2
BMR1A_ACR2A_BMP4
BMR1A_ACR2A_BMP2
BMPR1A_BMPR2_BMP4
BMPR1A_BMPR2_BMP2
FGFR1_FGF7
FGF7_FGFR2
FGF18_FGFR2
FGF10_FGFR2
WNT5A_ROR2
WNT5A_ROR1
WNT5A_PTPRK
WNT5A_FZD5
WNT5A_FZD3
WNT4_FZD8
WNT4_FZD1
NOTCH1_WNT4
FZD6_WNT5A
FZD6_WNT4
cTEC_cTEC
cTEC_DN-Thymocyte
cTEC_DP-Thymocyte
cTEC_Endothelial
cTEC_Fibroblast-1
cTEC_Fibroblast-2
cTEC_Lymphatic Endothelial
DN-Thymocyte_cTEC
DP-Thymocyte_cTEC
Endothelial_cTEC
Fibroblast-1_cTEC
Fibroblast-2_cTEC
Lymphatic Endothelial_cTEC
Interating Gene Pair [gene1_gene2]
Interacting Cells [sourceltarget]
Fibroblast 1
Fibroblast 2
Collagen 4/6
Collagen 4/6
α2β1/α1β1
CD74
MIF
MIF
WNT5a
FZD
FGF7
FGF10
FGF18
BMR
BMP4
BMP7
ROR/FZD
WNT4
FGFR2
TGFβR
EGFR
CD74
TGFβ
AREG
MIF
NOTCH
DLL1
DLL4
CCL25
CCR/ACKR
cTEC
DP Thymocyte
DN Thymocyte
ROR/
FZD
WNT5a
WNT4
α2β1/α1β1
NOTCH
FGFR2
DLK1
Collagen
BMP2
CCR/ACKR
TGFβ
MIF
AREG
Endothelial
CCL19
TGFβR
CD74
EGFR
ROR/FZD
WNT5a
TGFβ
MIF
WNT4
FZD
NOTCH
DLK1
BMP2
BMR
TGFβR
TGFβ
AREG
EGFR
BMP7
MIF
CD74
Lymphatic
Endothelial

Figure 6. Identification of critical receptor-ligand interactions within the thymic microenvironment. (A) Dot plot visualization of select CellPhoneDB cell-cell interactions across select thymic cell types. Dot size corresponds with -Log10 P-Value while dot color corresponds with average expression. (B) Graphical representation of key interactions identified in (6A). Black arrows represent higher expression levels of the interactors.

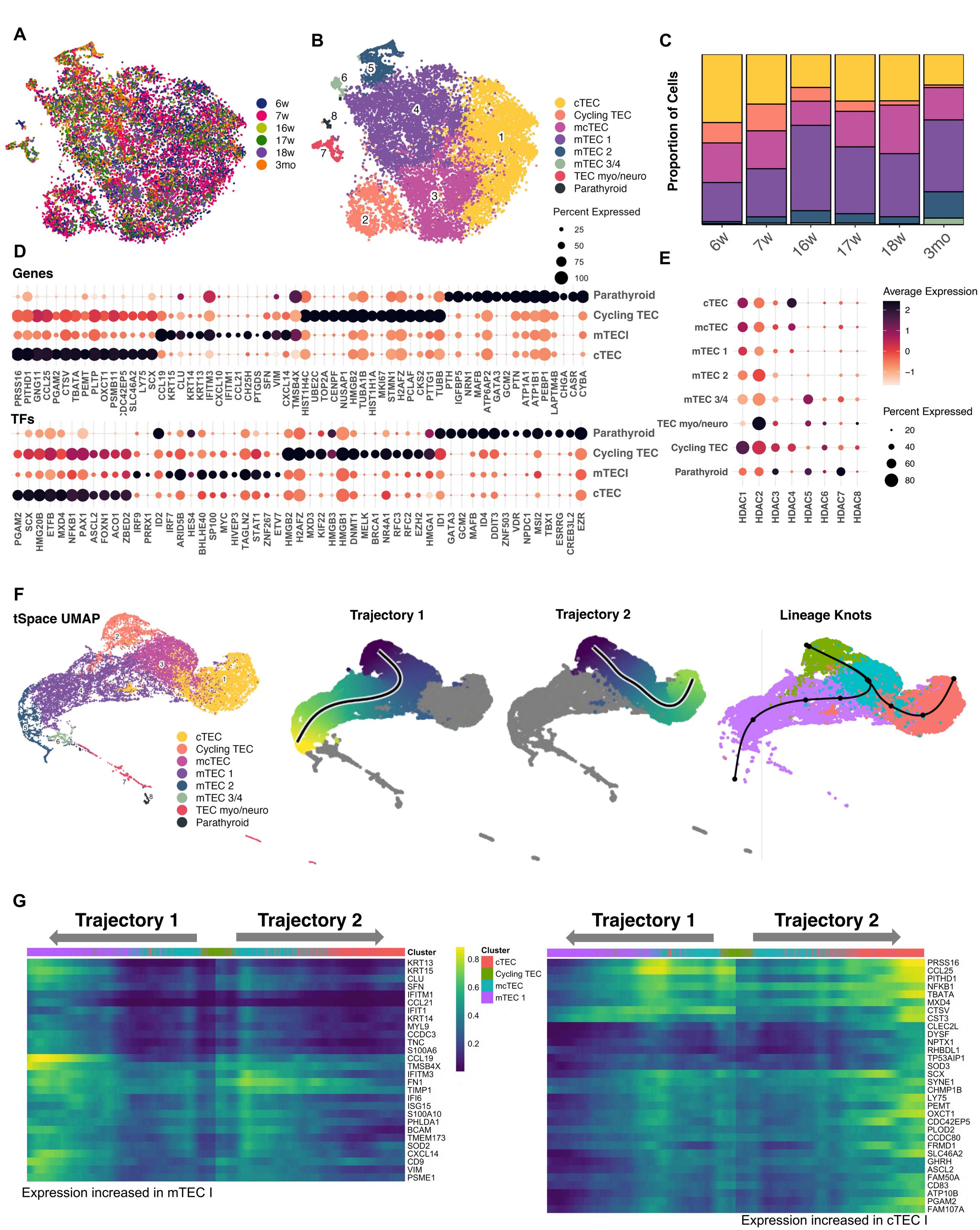

A
6w
7w
16w
17w
18w
3mo
B
cTEC
Cycling TEC
mcTEC
mTEC 1
mTEC 2
mTEC 3/4
TEC myo/neuro
Parathyroid
Percent Expressed
25
50
75
100
C
Proportion of Cells
6w
7w
16w
17w
18w
3mo
D
Genes
TFs
Parathyroid
Cycling TEC
mTECI
cTEC
E
Average Expression
Percent Expressed
HDAC1
HDAC2
HDAC3
HDAC4
HDAC5
HDAC6
HDAC7
HDAC8
F
tSpace UMAP
Trajectory 1
Trajectory 2
Lineage Knots
G
Trajectory 1
Trajectory 2
Cluster
Expression increased in mTEC I
Expression increased in cTEC I

Figure 7. Identification of key genes and lineages within vertical thymic development. tSNE visualization of thymic epithelial cells across all sampled developmental time points, colored by cell types (A) or by developmental time point (B). (C) Relative proportion of select thymic epithelial cell types at each developmental time point. (D) Dot plot representation of all differential gene and transcription factor expression of Parathyroid and key thymic epithelial cell types. Color representative of average expression level. Dot size representative of percentage of cells in cell type expressing selected marker gene. (E) Dot plot representation of all identified HDAC gene expression of all thymic epithelial cell types. Color representative of average expression level. Dot size representative of percentage of cells in cell type expressing selected marker gene. (F) tSpace UMAP of all vertical cells colored by cell type. Slingshot identified trajectories 1 and 2 are plotted on the tSpace UMAP. Pseudotime knots used for TradeSeq analysis are also visualized. (G) Heatmaps of the rolling average of differentially expressed genes between knots 2-5 along trajectories 1 and 2. Heatmaps are differentiated by upregulation along the trajectory of thymic cell type cTEC and mTEC 1.

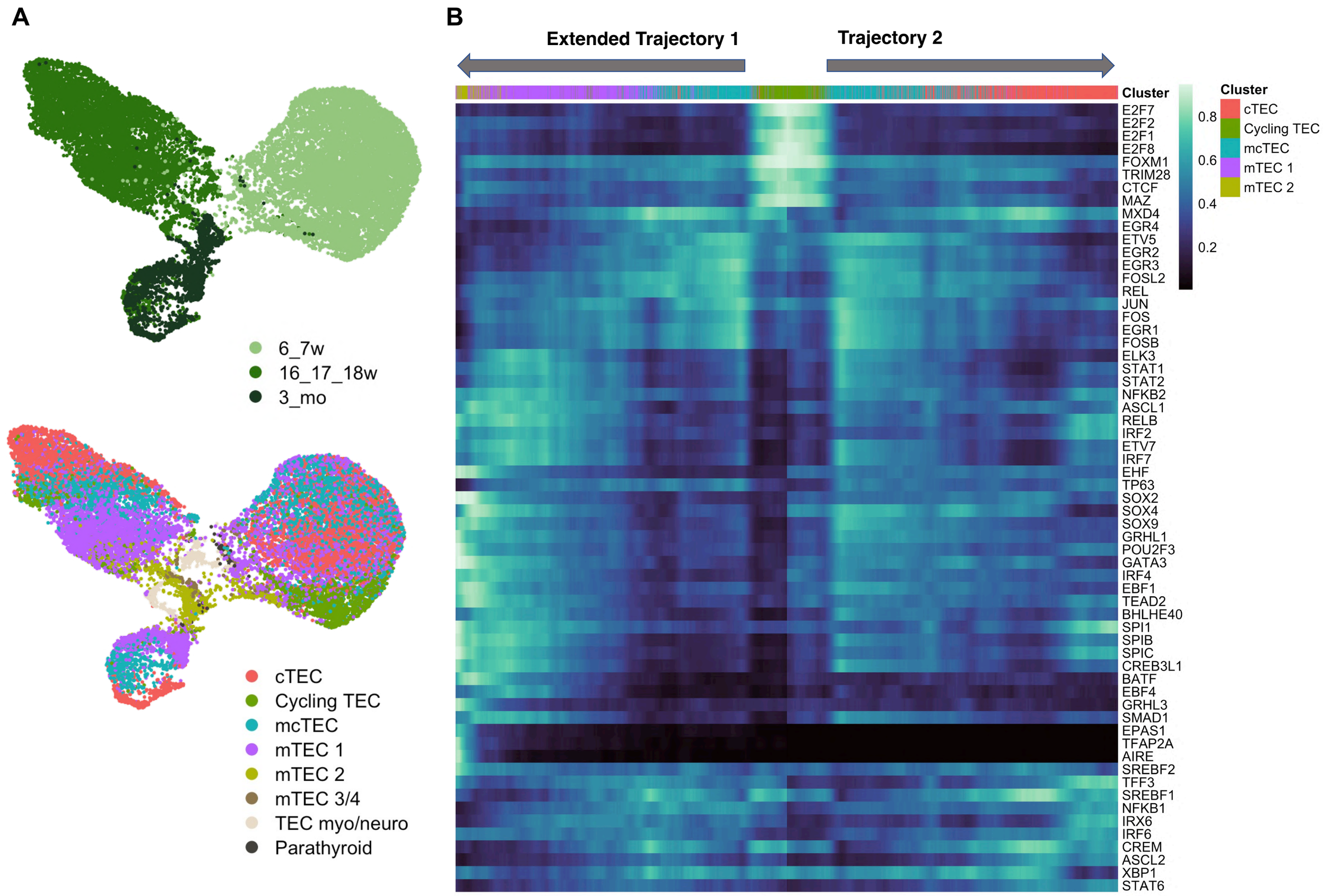
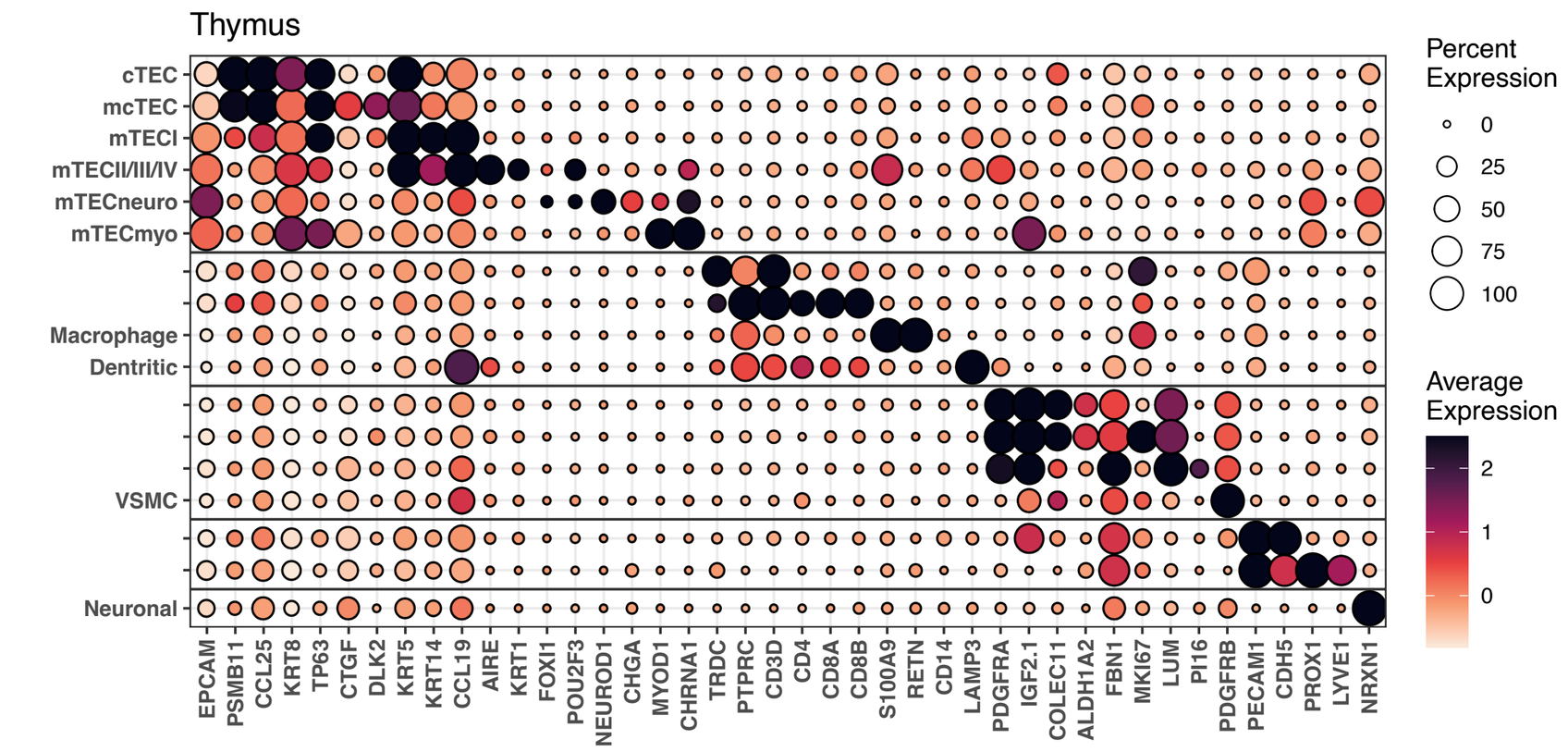
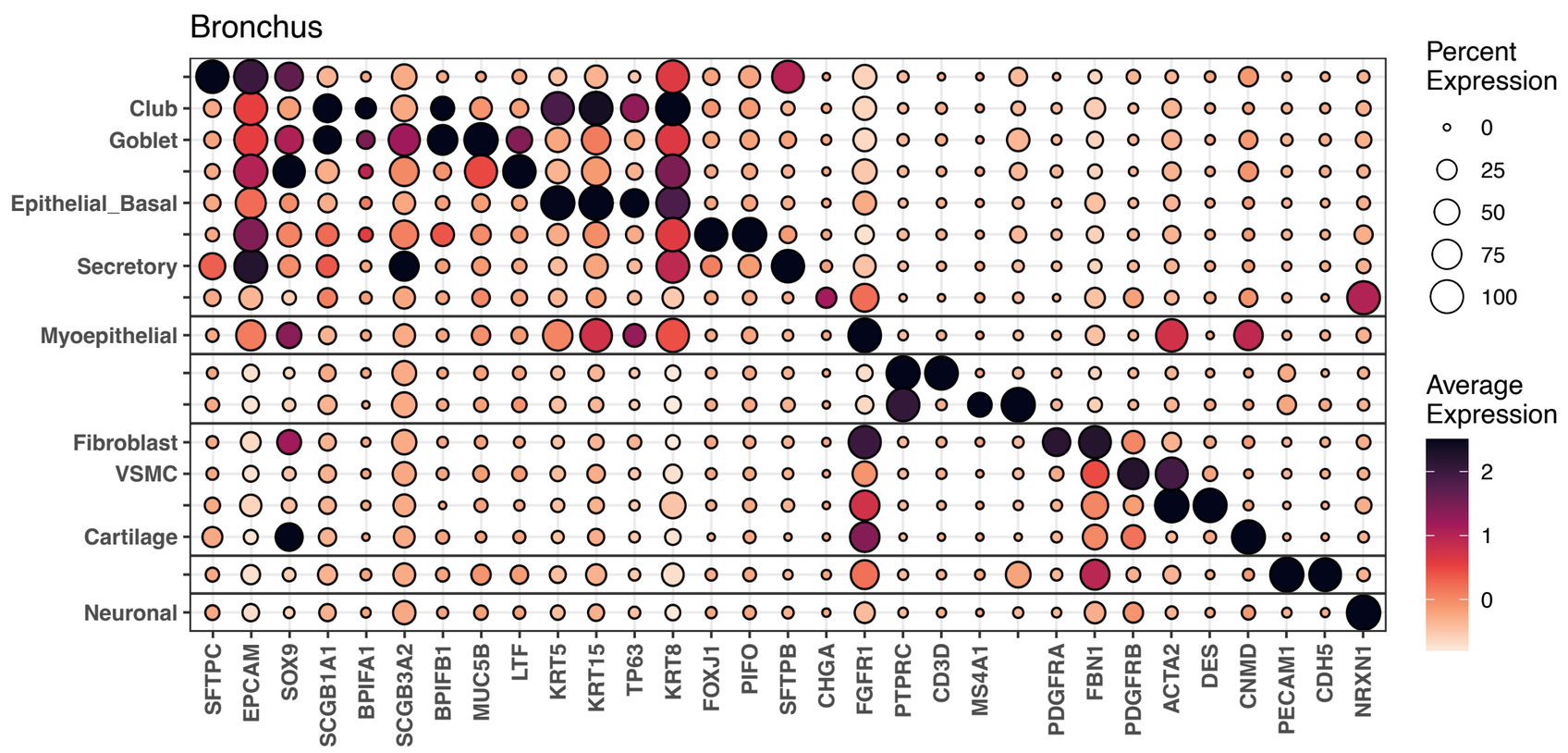
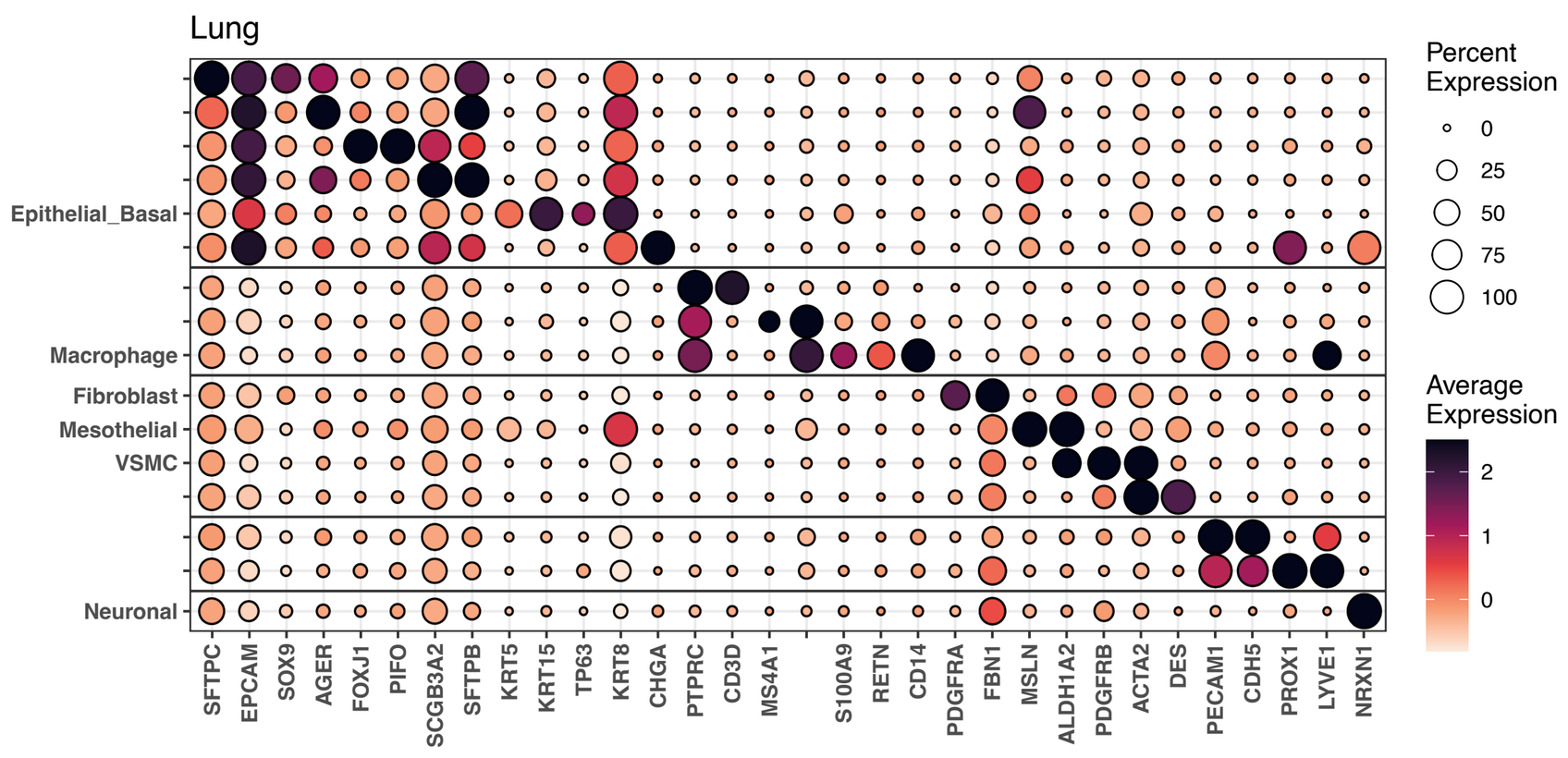

A
6_7w
16_17_18w
3_mo
cTEC
Cycling TEC
mcTEC
mTEC 1
mTEC 2
mTEC 3/4
TEC myo/neuro
Parathyroid
B
Extended Trajectory 1
Trajectory 2
Cluster
E2F7
E2F2
E2F1
E2F8
FOXM1
TRIM28
CTCF
MAZ
MXD4
EGR4
ETV5
EGR2
EGR3
FOSL2
REL
JUN
FOS
EGR1
FOSB
ELK3
STAT1
STAT2
NFKB2
ASCL1
RELB
IRF2
ETV7
IRF7
EHF
TP63
SOX2
SOX4
SOX9
GRHL1
POU2F3
GATA3
IRF4
EBF1
TEAD2
BHLHE40
SPI1
SPIB
SPIC
CREB3L1
BATF
EBF4
GRHL3
SMAD1
EPAS1
TFAP2A
AIRE
SREBF2
TFF3
SREBF1
NFKB1
IRX6
IRF6
CREM
ASCL2
XBP1
STAT6
0.8
0.6
0.4
0.2
Cluster
cTEC
Cycling TEC
mcTEC
mTEC 1
mTEC 2

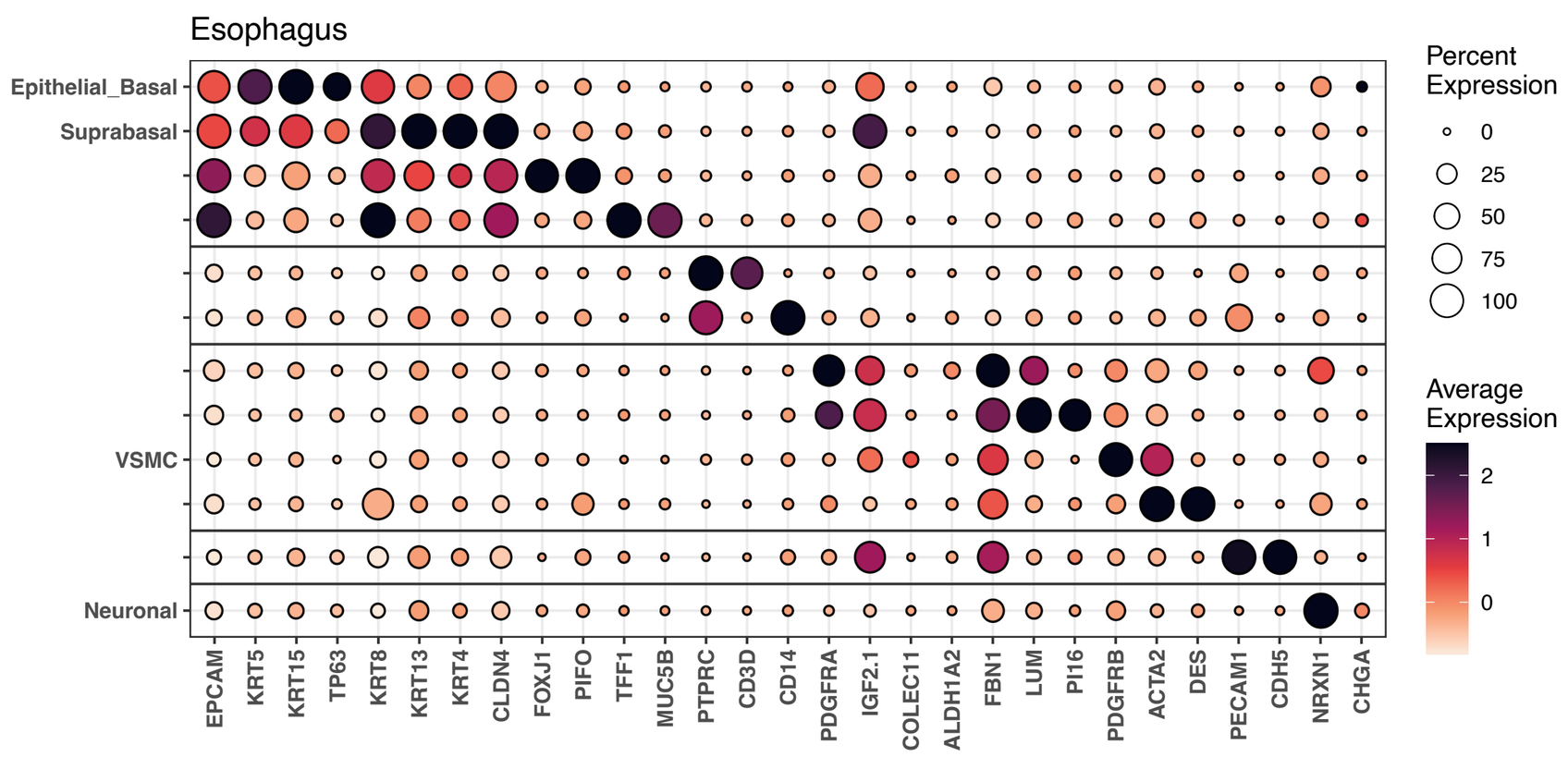
Figure 8. Thymic epithelial lineage progression through a SCENIC lens. (A) UMAP visualization of SCENIC regulon AUC scores of all thymic epithelial cells from AFG-derived tissue classified by age and cell type. (B) Heatmap representation of the rolling average of SCENIC regulon AUC scores across both trajectories 1 and 2 of select regulons.

Supplementary Figure S1



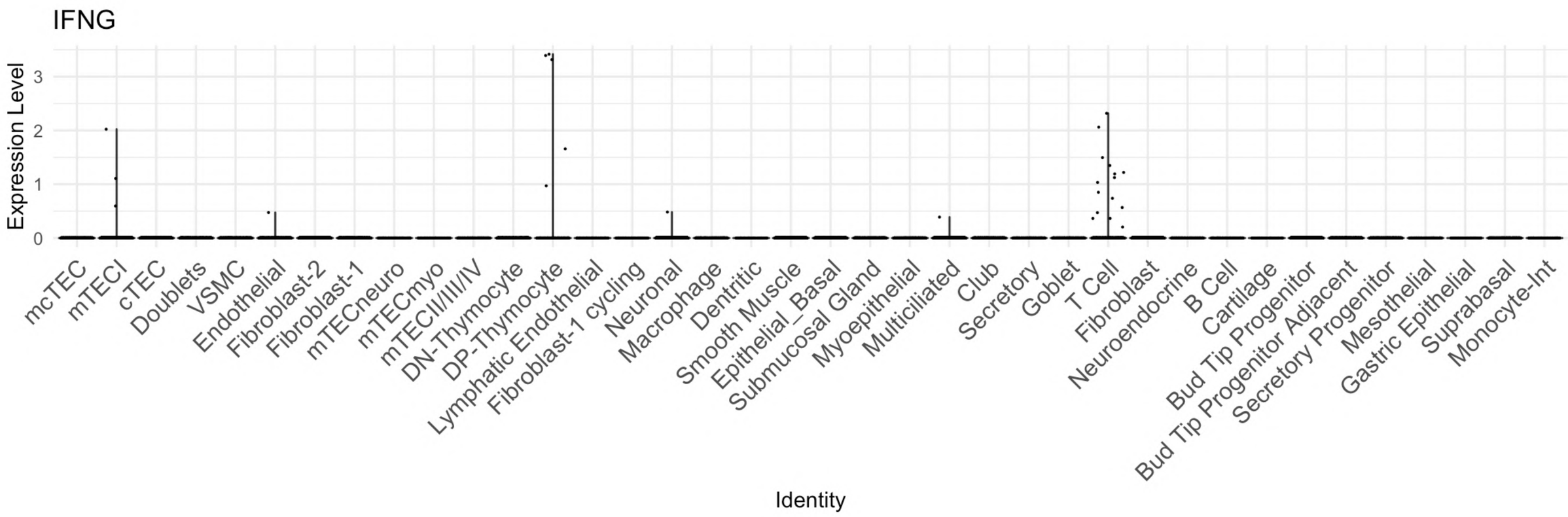






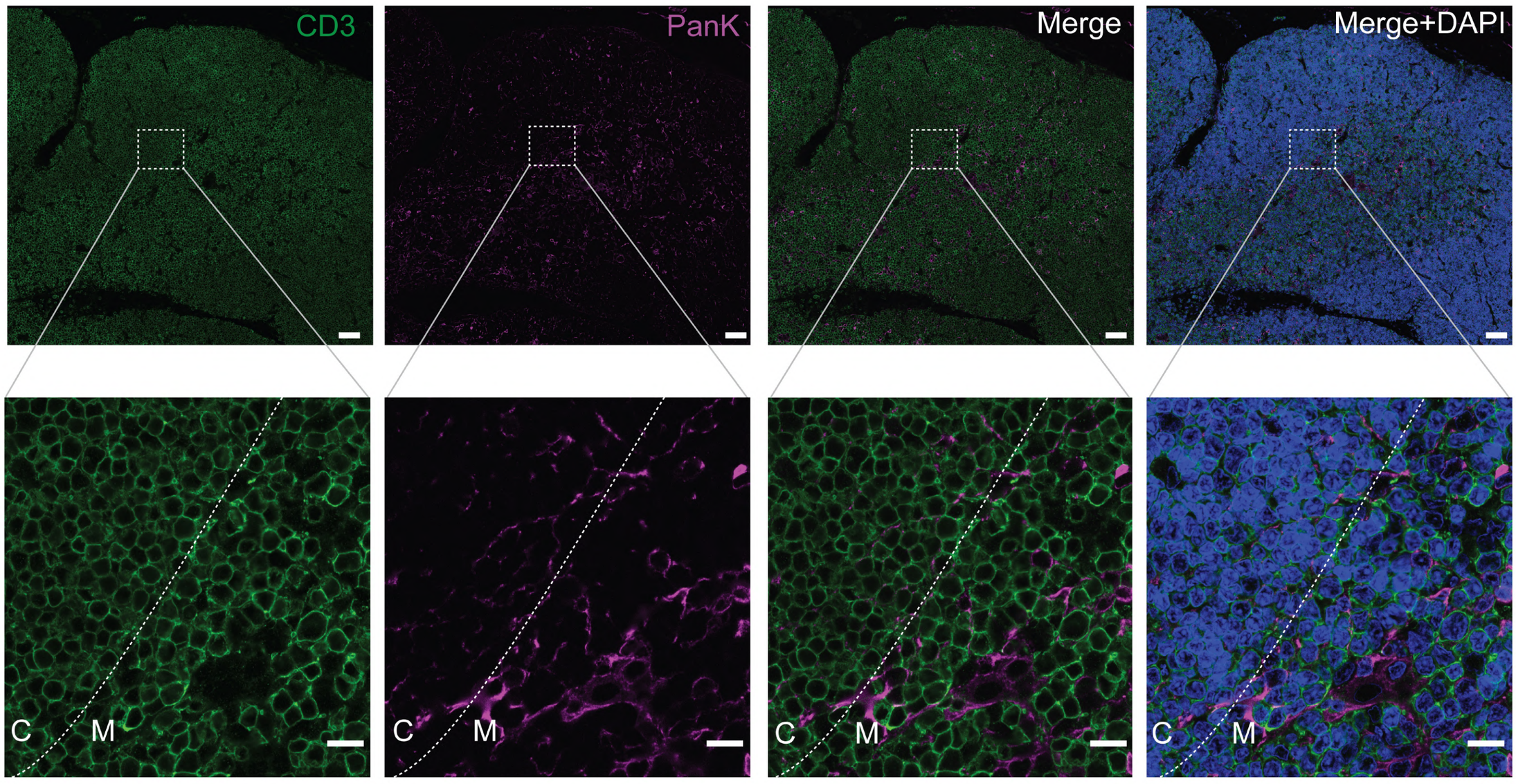
Supplementary Figure S1. Dot plot showing representative marker gene expression of anterior foregut cell types across individual tissues. Color representative of average expression level. Dot size representative of percentage of cells in cell type expressing selected marker gene.

Supplementary Figure S2



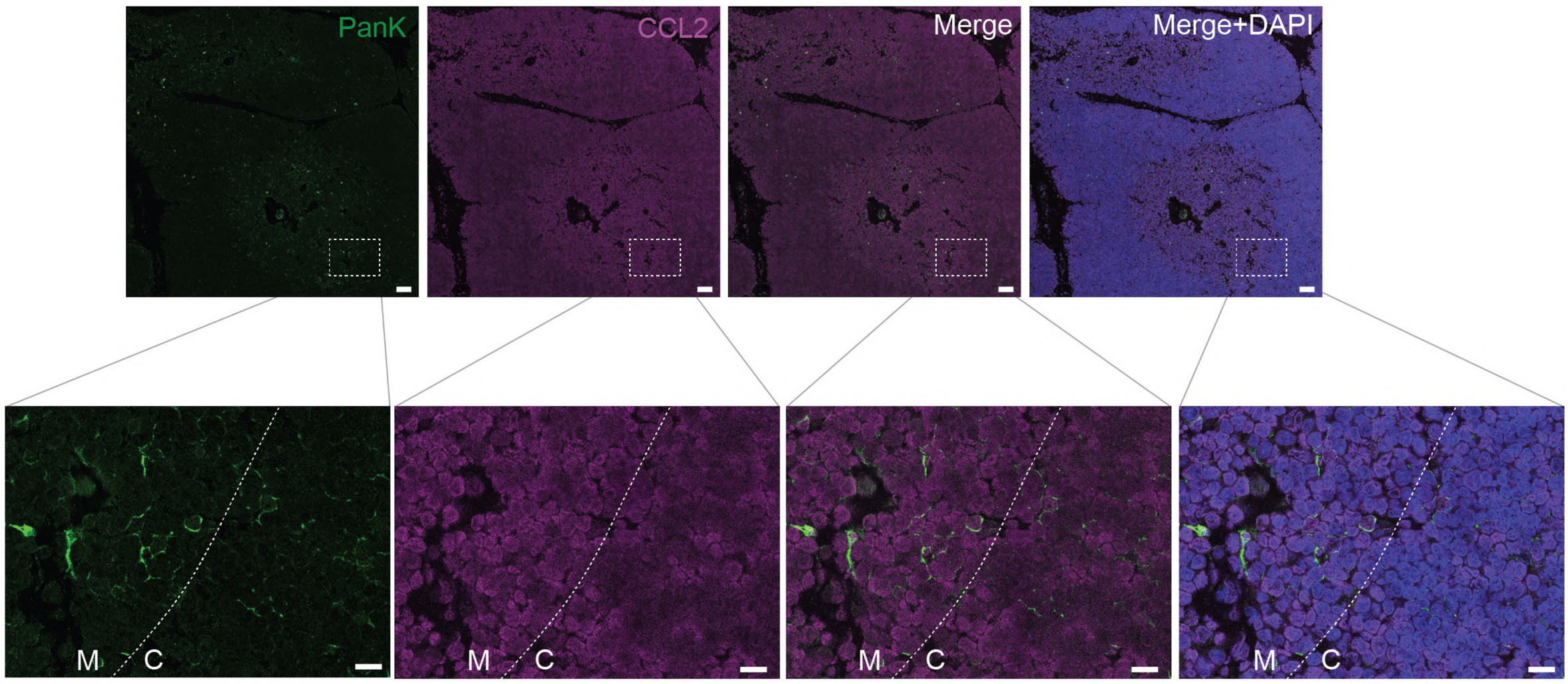
Supplementary Figure S2. Violin plot representation of Reactome IFN Gamma module expression across all identified cell types

Supplementary Figure S3



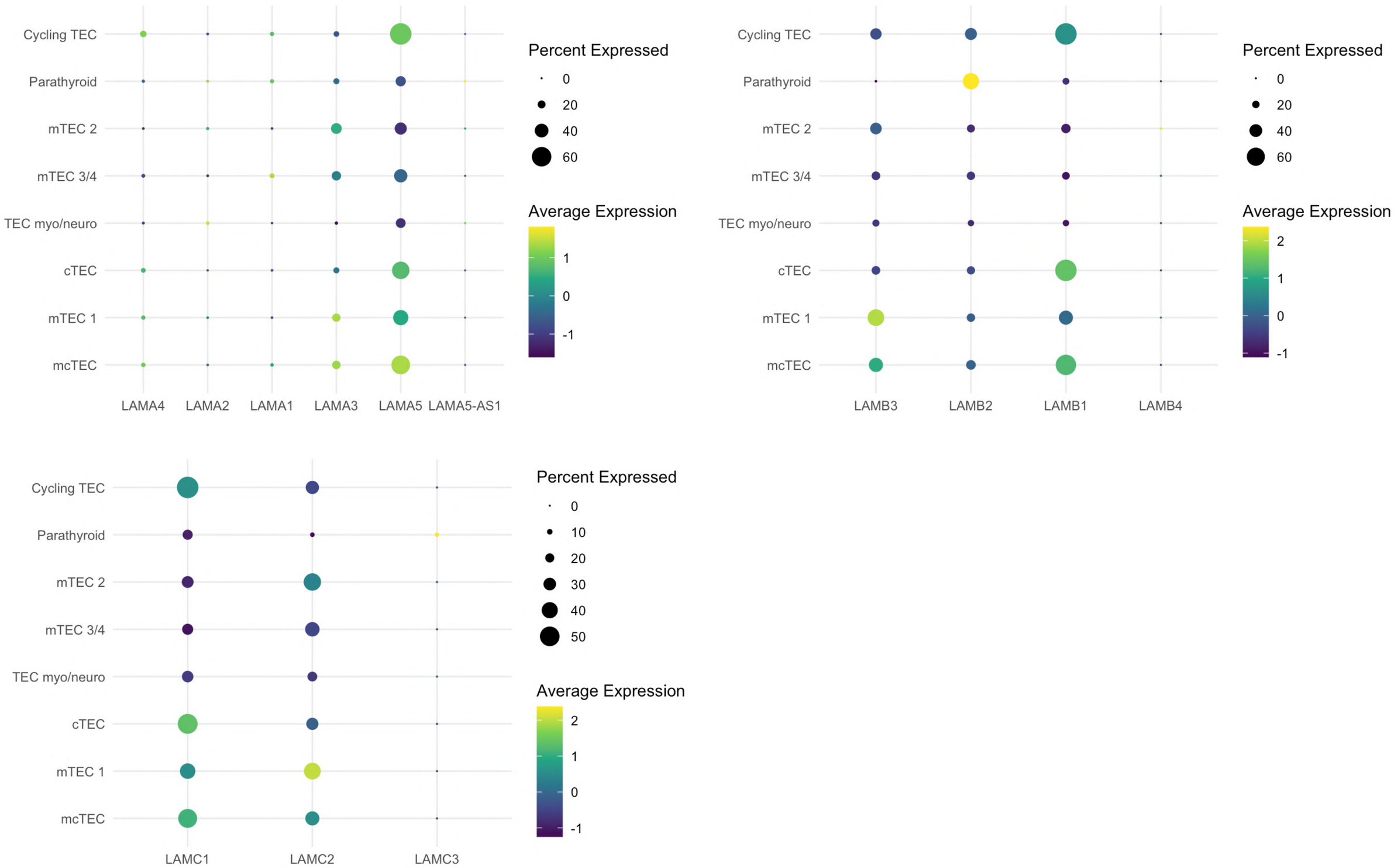
Supplementary Figure S3. Representative immunofluorescence confocal microscopy image of a 16-week human fetal thymus stained for CD3 (green), PanK (magenta) and DAPI (Blue). C: Cortex; M: Medulla. Scale bars: 50um and 10um for full scaled and zoomed in images, respectively.

Supplementary Figure S4



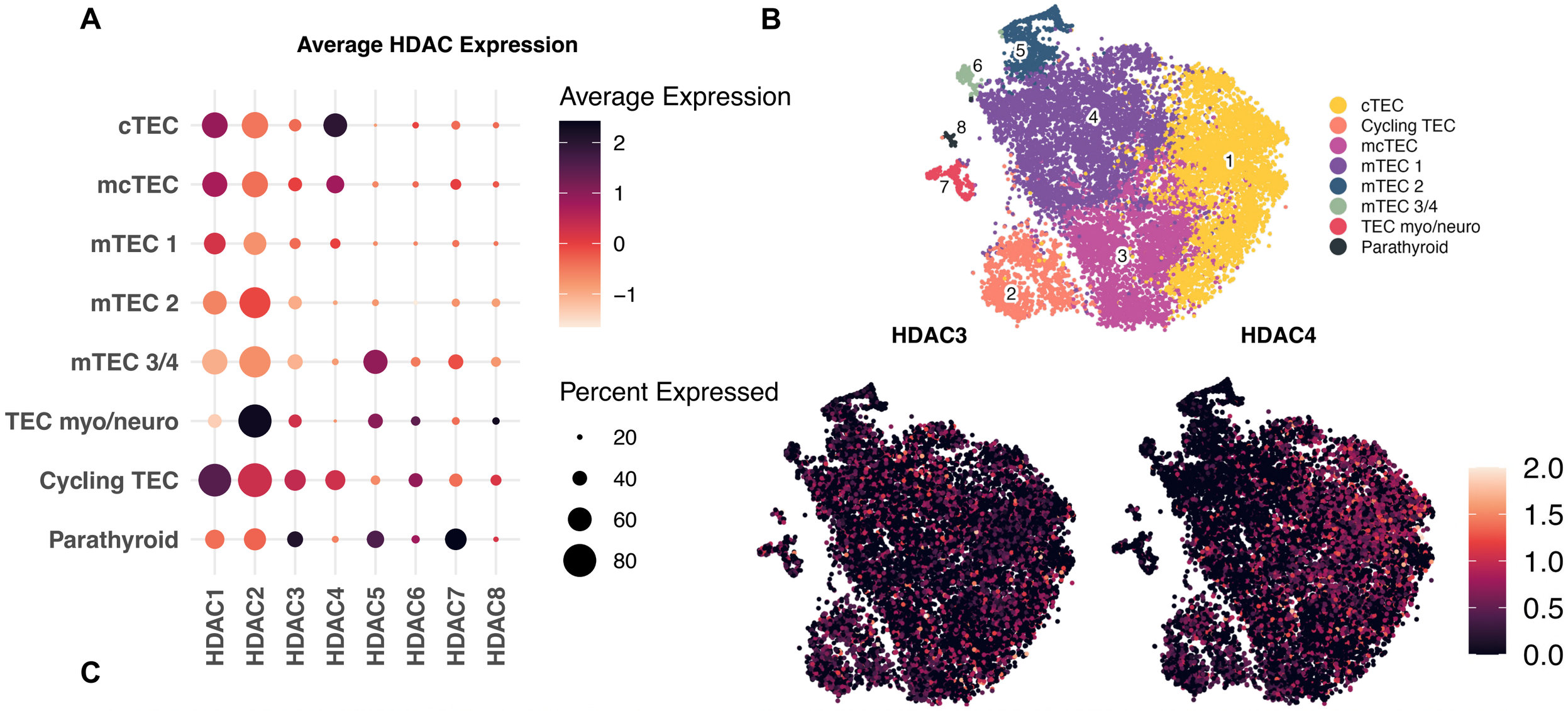
Supplementary Figure S4. CCL2 protein levels in thymic cortex and medulla. Representative immunofluorescence confocal microscopy image of a 16-week human fetal thymus stained for PanK (green), CCL2 (magenta) and DAPI (Blue). C: Cortex; M: Medulla. Scale bars: 50um and 10um for full scaled and zoomed in images, respectively.

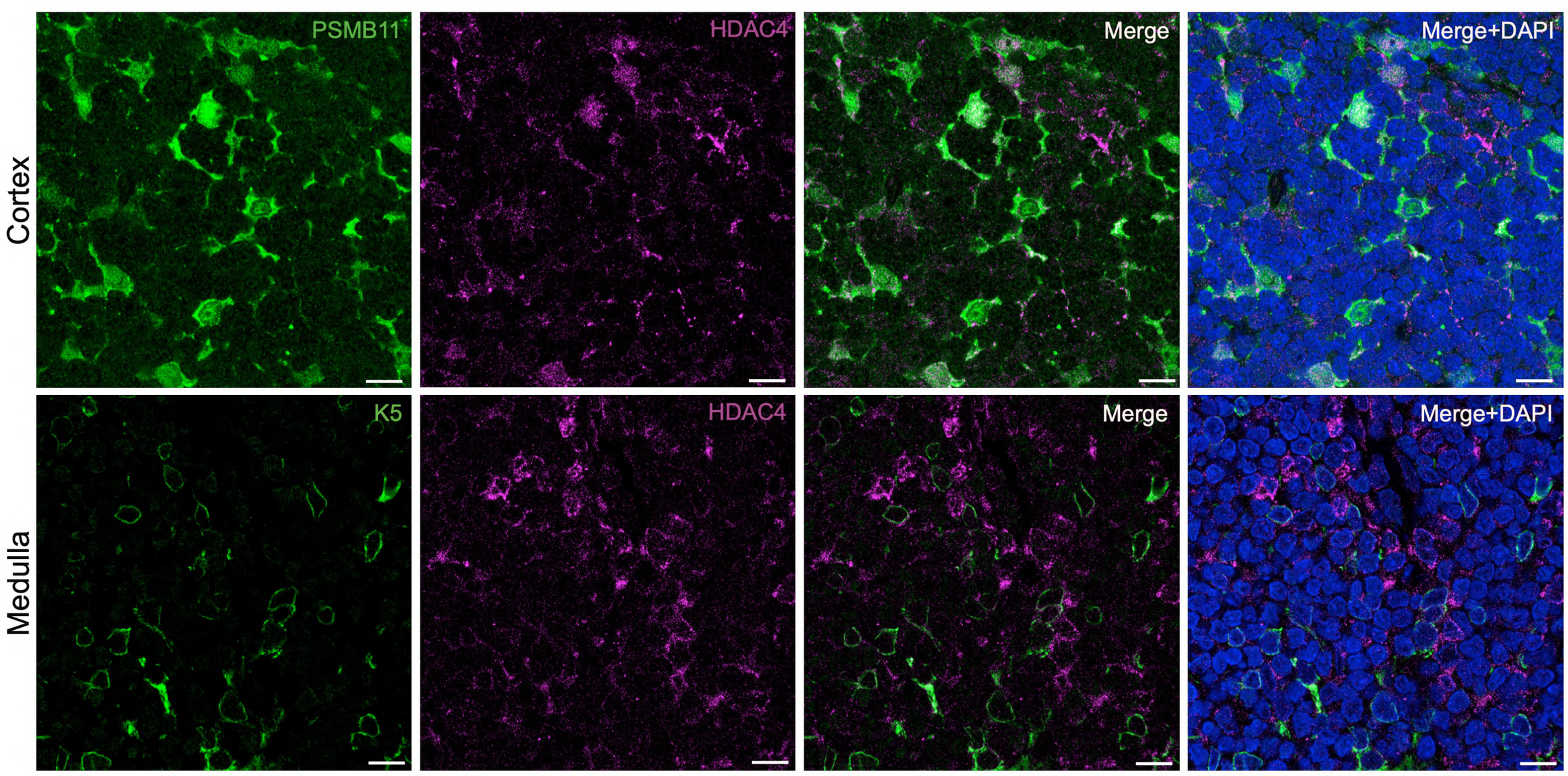
Supplementary Figure S5

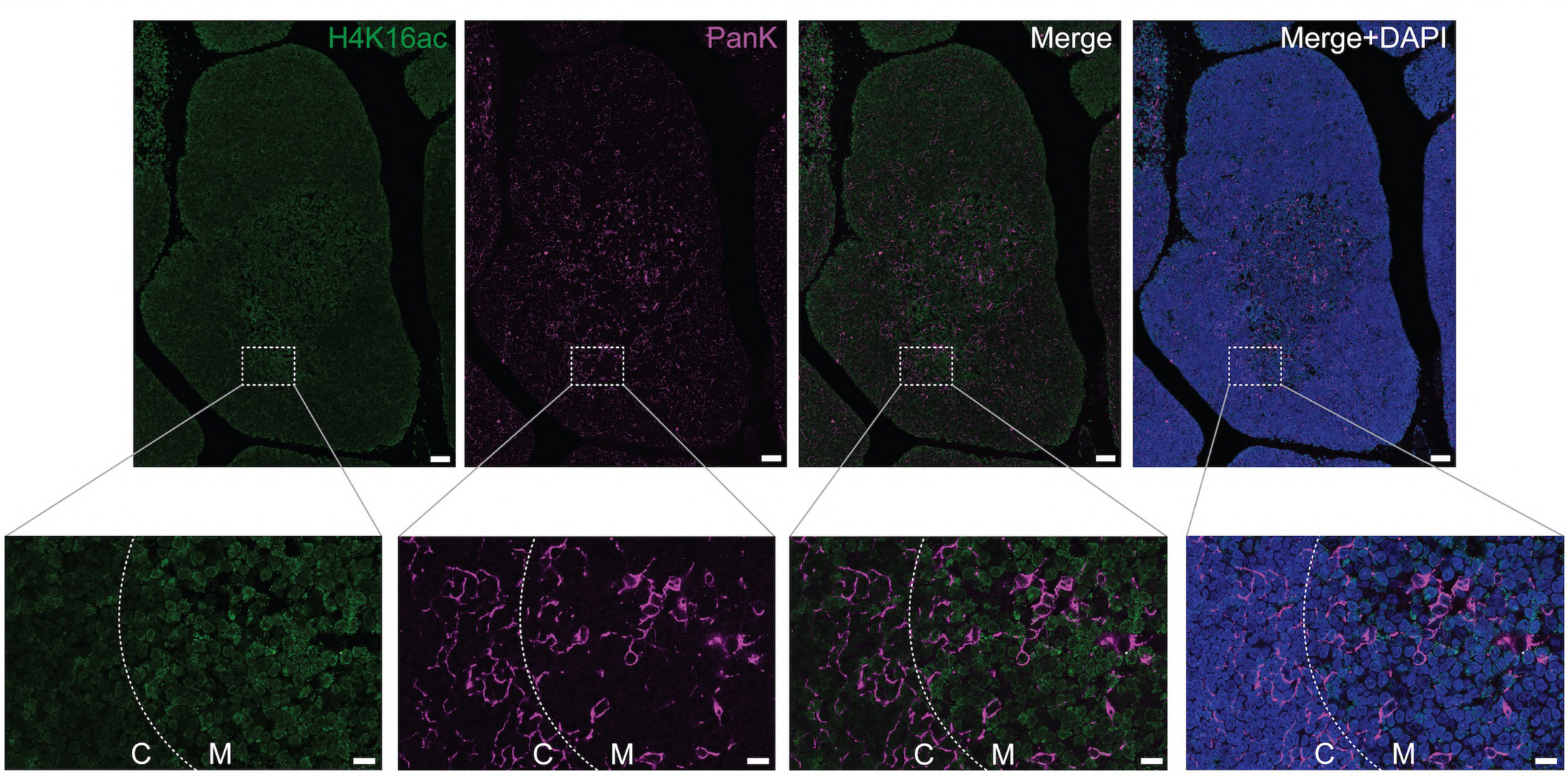


Supplementary Figure S5. Dot plot representation of all identified Lamnin gene expression across all thymic epithelial cell types. Color representative of average expression level. Dot size representative of percentage of cells in cell type expressing selected marker gene.

Supplementary Figure S6







Supplementary Figure S6. (A) Dot plot representation of all identified HDAC gene expression of all thymic epithelial cell types. Color representative of average expression level. Dot size representative of percentage of cells in cell type expressing selected marker gene. Same as (7E) (B) tSNE visualization of thymic epithelial cells across all sampled developmental time points colored by cell types. t-SNE visualization of thymus cells overlaid with HDAC3 and HDAC4 expression. (C) Representative immunofluorescence confocal microscopy image of a human fetal thymus stained for PSMB11 (green), HDAC4 (magenta) and DAPI (Blue). C: Cortex; M: Medulla. Scale bars: 50um. (D) Representative immunofluorescence confocal microscopy image of a human fetal thymus stained for H4K16ac (green), PanK (magenta) and DAPI (Blue). C: Cortex; M: Medulla. Scale bars: 50um and 10um for full scaled and zoomed in images, respectively.

Supplementary Table S1

Supplementary Table S1. metadata for sequenced samples								
Sample ID	Organ	Age	NumCells	MeanCounts	MeanGenesP	MedianMitol	Gender	Tissue processing
C40_TH_TOT_2	Thymus	PCW7	7180	11139	3054.5	1.97	Female	All cells
C40_TH_TOT_1	Thymus	PCW7	6841	11293.8	3098.1	1.97	Female	All cells
C41_TH_TOT_2	Thymus	PCW7 to 8	5789	14190.5	3544.9	2.61	Male	All cells
C41_TH_TOT_1	Thymus	PCW7 to 8	5620	14244.3	3547.6	2.52	Male	All cells
KGW1 (previously HGSC5)	Thymus	PCW16	3046	21747	4135	7.29	Male	EPCAM positive selection, CD45 depletion
F83_TH_EPCAM	Thymus	PCW17	3678	9575.2	2634.2	2.55	Female	EPCAM FACS sorted
KGW2 (previously HGSC1)	Thymus	PCW17	7634	11754	2993	6.81	Female	EPCAM positive selection, CD45 depletion
KGW3 (previously HG1)	Thymus	PCW18	8648	9349	2721	5.73	Male	EPCAM positive selection, CD45 depletion
T07_TH_EPCAM_S3	Thymus	3 months	3149	11355.6	2819.6	3.86	Male	EPCAM positive selection
KGW4 (previously HGSC6)	Bronchus	PCW16	5920	15651	3525	7.69	Male	EPCAM positive selection
KGW5 (previously HGSC2)	Bronchus	PCW17	4376	16449	3188	8.56	Female	EPCAM positive selection
KGW6 (previously HG2)	Bronchus	PCW18	7414	15387	2979	9.43	Male	EPCAM positive selection
KGW7 (previously HGSC7)	Lung	PCW16	3726	20366	3862	7.95	Male	EPCAM positive selection
KGW8 (previously HGSC3)	Lung	PCW17	7075	11074	2810	8.11	Female	EPCAM positive selection
KGW9 (previously HG3)	Lung	PCW18	7208	12965	3254	7.12	Male	EPCAM positive selection
KGW10 (previously HGSC8)	Esophagus	PCW16	4603	16004	3513	7	Male	EPCAM positive selection
KGW11 (previously HGSC4)	Esophagus	PCW17	6464	11431	2723	8.48	Female	EPCAM positive selection