**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 D. Solomon 2, Priscila F. Slepicka 1, Kelsea M. Hubka 1, Hanh Dan Nguyen 1, Michael G. Chavez 3, Christine Y. Yeh 4, Doo Kyung Kim 3, Vittorio Sebastiano 5, Kenneth I Weinberg 1, Virginia Winn 5, Casey A. Gifford 6, Veronika Kedlian 7, Jong-Eun Park 7, Georg A. Hollander 8, Purvesh Khatri 9, Sarah A. Teichmann 7, Andrew J. Gentles 4, 10*#, and Katja G. Weinacht 1*#

1 Department of Pediatrics, Division of Stem Cell Transplantation and Regenerative Medicine, Stanford University, Stanford CA, USA
2 Department of Pediatrics, Division of Immunology, Stanford University, 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, Stanford CA, USA
6 Department of Pediatrics, Division of Cardiology, Stanford University, Stanford CA, USA
7 Wellcome Sanger Institute, Cambridge, UK
8 Department of Pediatrics and Institute of Developmental and Regenerative Medicine, University of Oxford, Oxford, United Kingdom
9 Institute for Immunity, Transplantation, and Infection, Stanford University, Stanford, CA, USA.
10 Department of Pathology, Stanford School of Medicine, Stanford, CA, USA

*These authors contributed equally to this work
#Corresponding authors

**ABSTRACT**

The thymus is essential in directing T cell immunity and central tolerance, yet the signals that drive thymic ontogeny beyond the anterior foregut remain incompletely understood. Here we elucidate key pathways instructing commitment and specialization of the human thymic epithelial stroma through complementary single cell transcriptomic approaches. First, we infer inductive cues that contribute to the initiation of thymic organogenesis by comparing gene regulatory networks expressed in human fetal epithelial cells of the thymus and other anterior foregut-derived organs. Second, we define lineage trajectories within the thymic epithelial compartment by evaluating thymus samples sequentially across embryonic, fetal, and early postnatal stages of development. Responses to types I and II interferon were the predominant distinguishing feature between epithelial cells of the thymus and other anterior foregut-derived organs. Distinct expression patterns for the transcription factors *SOX2* and *TP63*, as well as key retinoic acid metabolizing enzymes were also observed. Lineage trajectory analysis revealed preferential engagement of the *NFκB* gene regulatory network in cortical thymic epithelial cells (TECs) while expression of the interferon response network predominated in the medullary trajectory. Prenatally, medullary TECs showed a pre-AIRE signature while post-AIRE maturation occurred predominantly after birth. However, unique clusters of atypical TECs lacking antigen presenting molecules and resembling peripheral muscle and neuroendocrine tissues were already present midgestation. Our study highlights a distinctive role for interferon signaling during thymic morphogenesis, differential regulatory

programs directing cortical and medullary specialization, and alternate patterns of peripheral tissue antigen expression arising at different stages of development.

## INTRODUCTION

The thymic epithelium comprises a highly specialized set of cells that attract lymphoid progenitors, promote their proliferation and maturation into thymocytes, and facilitate the selection of a diverse, self-tolerant T cell receptor repertoire. As such, the thymus is central to the integrity of the adaptive immune system. Thymic function declines in the second decade of life and is susceptible to a wide range of pathogenic insults[1,2]. Any endogenous or acquired compromise of thymic functions leads to immunodeficiency. Reconstituting thymic function through *in vitro*-generated induced pluripotent stem cell (iPSC)-derived thymic tissues holds great therapeutic promise. The signals that induce anterior foregut endoderm (AFE) from iPSCs have long been discovered and protocols to *in vitro* differentiate epithelial cells of other AFE-derived organs such as lung and esophagus have been successfully established[3-8]. Recent studies of murine and human thymus using single cell transcriptomics have greatly expanded our understanding of thymic epithelial cell (TEC) diversity[9-17], yet the signals that drive TEC fate beyond the AFE stage remain incompletely understood. Elucidating these signals is essential to derive TECs from iPSCs *in vitro* and enable regenerative cell therapies for clinical applications.

The anterior foregut gives rise to the ventrally located respiratory system and the dorsally located esophagus[18]. The anterior-most segment of the foregut develops into the pharyngeal apparatus. Essential to the development of the mammalian pharyngeal apparatus is the out-pocketing of the pharyngeal endoderm to form the pharyngeal pouches[19]. The thymic primordium, also referred to as thymus anlage, forms from the ventro-caudal region of the third pharyngeal pouch while the dorsal aspect gives rise to the parathyroid (Figure 1A). In human embryonic samples, the third pharyngeal pouch transcriptional regulators TBX1, HOXA3, PAX1 and PAX9 are detected as early as the beginning of post conception week (pcw) 6 and preceded the expression of the master thymic transcription factor FOXN1 mid-week 6[20]. Lymphoid progenitors enter the thymus anlage as early as pcw 7[12,20] and are indispensable for epithelial cell maturation and expression of functional markers, such as PSMB11 and AIRE[21-23]. Murine studies have implicated two developmentally separate steps in thymic organogenesis. Emergence of the earliest discernable thymic primordium is instructed by signals from the surrounding mesenchyme. In the second phase, bone marrow-derived lymphoid precursors migrate into the developing thymus and reciprocal signals between cells of hematopoietic and thymic epithelial origin promote thymic organogenesis and thymopoiesis, i.e. the development of thymocytes into mature T cells, respectively[2,24-26]. Branching morphogenesis of the airway and cortico-medullary separation in the thymus exhibit notable parallels. In both settings, a homogenous epithelial cell population begins to polarize and display distinct apical-basal features. In the thymus, centrally located epithelial cells gain Cldn3/4 expression and go on to form the medulla. However, unlike in other AFE-derived organs, a thymic lumen does not form[27,28]. Important insights into the developmental programs that instruct thymic organogenesis have been derived from mice with biallelic loss of function mutations in *Foxn1* which result in an athymic and hairless (nude) phenotype[29,30]. Thymic development in nude mice arrests after Cldn3/4 are expressed but before colonization with lymphoid precursors occurs[28,31]. These findings align with multiple studies indicating Foxn1 is required cell-autonomously for lympho-attraction and maturation of the thymic epithelium but not for initiation of thymic patterning[24,31-34]. The thymic rudiment of *Foxn1*$^{null}$ mice displays nonfunctional cystic features strongly resembling respiratory epithelium in morphology and gene expression[28,33-35]. Notably, NOD-scid IL2R$^{null}$ (NSG) mice with genetically impaired immune functions of hematopoietic origin leading to a lack of thymocytes also display distorted thymic architecture with lack of cortico-medullary demarcation and cystic features[36]. Although NSG mice do not have genetic defects in thymic epithelial development, they exhibit permanently impaired thymic organ function leading to poor T cell reconstitution even if healthy hematopoietic stem cells are transplanted later in life[36,37]. This suggests Foxn1 is necessary but not sufficient to fully suppress respiratory cell fate and promote

functional organogenesis of the developing thymus. However, the corresponding gene regulatory networks governing the lymphoid-driven maturation of the thymic epithelium have not yet been characterized.

Here we identify potential drivers underlying thymic epithelial differentiation by comparing epithelial cells from human fetal thymus, bronchus, lung, esophagus and parathyroid using single cell transcriptomics. In addition, we characterize developmental programs specific to thymic epithelial subsets by analyzing human samples from embryonic, fetal, and early postnatal stages. Our studies yield translatable insights critical to advance *in vitro* differentiation approaches into functional thymic epithelium and promote cortical and medullary specialization.

## RESULTS

### Human fetal anterior foregut-derived organs at single cell resolution

Single cell RNA-sequencing (scRNA-seq) was performed on a total of 17 samples from 15 biological specimens. For the comparison of anterior foregut-derived organs, we sampled tissues from human fetal thymus (pcw 16, 17 and 18), airway/bronchus and lung parenchyma (pcw 16, 17 and 18), as well as esophagus (pcw 16 and 17). To examine lineage trajectories and diversification of the thymic epithelium, we studied very early human thymus samples (pcw 6, 7), in comparison to mid-gestation (pcw 16, 17, 18) and neonatal (age 3 months) tissues (Figure 1A). The earliest thymus samples were used unmanipulated due to very low numbers of thymocytes at this stage of development. For all later stages, thymic samples were depleted of CD45+ thymocytes using magnetic beads. All tissues were enriched for epithelial cells using EPCAM positive selection with magnetic beads (Figure 1B), except for two thymic samples which were processed by FACS sorting. All sample characteristics are detailed in Supplementary Table S1.

Preprocessing and integration of scRNA-seq samples yielded a dataset of 52,162 single cell transcriptomes with a mean of 3522 unique genes per cell, and an average of 14,839 total counts per cell. For the comparison of AFG-derived organs, human thymus, bronchus, lung, and esophagus samples (pcw 16, 17 and 18) 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. Cells within unsupervised clusters were annotated in concordance with nomenclature employed in previously published organ-specific human single cell datasets for thymus[12], respiratory system[38,39] and esophagus[40], respectively.

Within the thymus, we identified a total of 18 epithelial and non-epithelial cell populations (Figure 1C). Epithelial cells further subclustered into canonical[10,12] *CCL25, PSMB11, KRT8*-high cortical thymic epithelial cells (cTEC) and *CCL19*, *KRT5/14*-high medullary thymic epithelial cells (mTECI). A population of TECs with dim expression of cTEC and mTEC markers and uniquely high expression of *CTGF* and *DLK2* previously termed mcTEC[12], was the third most abundant population. Classical mTECII were characterized by high transcript levels of *AIRE* and *CLDN4*. The medullary TEC subpopulations mTECIII and mTECIV were defined by expression of *KRT1* as well as *FOXI1* and *POU2F,* respectively, which correspond to the post-AIRE keratinocyte-like mTECs and Tuft cells in the mouse[10,11,41,42]. The recently discovered muscle and neuroendocrine-like populations TECneuro and TECmyo, identified by *NEUROD1*, *CHGA* as well as *MYOD1* and *CHRNA1* expression, respectively, were also evident[12]. Nonepithelial clusters included double-negative (DN) and double-positive (DP) thymocytes, along with small numbers of single positive T cells (*PTPRC*, *CD3D*, *CD4, CD8A, CD8B*), macrophages (*S100A9, RETN*), dendritic cells (*LAMP3*), vascular smooth muscle cells (*PDGFRB, ACTA2*), aldehyde dehydrogenase-positive fibroblasts, previously described as thymic fibroblasts 1[12] (*PDGFRA, ALDH1A2*), fibronectin-high fibroblasts, previously described as thymic fibroblasts 2[37] (*FBN1*), vascular endothelial cells (*PECAM1, CDH5, AQP1*), lymphatic endothelial cells (*LYVE1, PROX1*) and neuronal cells (*NRXN1, MPZ*) (Figure 1C top left, Figure 1D, Supplementary Figure S1A).

Tissues of the bronchus clustered into airway epithelial basal cells (*KRT15, IL33, DLK2*), club cells (*SCGB1A1, BPIFA1*), goblet cells (*MUC5B, BPIFB1*), submucosal gland cells (*LTF*), secretory progenitor cells (*SFTPB, SCGB3A2*), multiciliated cells (*FOXJ1, PIFO*) 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 (Figure 1C top right, Figure 1D, Supplementary Figure S1A).

Lung tissues expectedly shared many of the same cell types surrounding the airways albeit in reverse ratios. Bud tip progenitor cells (*SFTPC, SOX9*) and bud tip progenitor adjacent cells (*AGER, SFTPB*) were the most abundant populations, followed by multiciliated cells (*FOXJ1, PIFO*), secretory progenitors (*SFTPB, SCGB3A2*) and very small numbers of epithelial basal cells (*KRT15, IL33, DKL2*). 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 (Figure 1C bottom left, Figure 1D, Supplementary Figure S1A).

The esophageal tissue included epithelial basal cells (*KRT15, DLK2, CLDN4*), suprabasal cells (*KRT13, KRT4, CLDN4*), multiciliated cells (*FOXJ1, PIFO*), gastric epithelial cells (*TFF1, MUC5B*) and non-epithelial cells including immune cells (T cells, macrophages), smooth muscle cells, vascular smooth muscle cells, fibroblasts, endothelial cells and a cluster of neuronal cells (Figure 1C bottom right, Figure 1D, Supplementary Figure S1A).

After integration of the individual data sets from thymus, bronchus, lung and esophagus we verified successful batch correction (Figure 1E, left). Subsequent clustering was defined based on shared tissue identity, regardless of organ of origin, separating epithelial cells (red) from mesenchymal cells (purple), immune cells (blue), endothelial cells (green), myoepithelial cells (pink) and neuronal cells (gray) (Figure 1E, middle; Supplementary Figure S1B). When compared across tissue types, top differentially expressed genes within the epithelial cell population included the epithelial cell adhesion molecule *EPCAM*, the transcription factors *ELF3* and *KLF5*, a select group of keratins (*KRT8, 17, 18, 19*), E-cadherin (*CDH1*), desmoplakin (*DSP*), the calcium binding and cytoskeletal-associated proteins *S100A11* and *S100A14*[43], the tight junction associated protein *CLDN4*, and *LMO7*, a negative regulator of transforming growth factor β (TGFβ) that has been implicated in maintaining cell fate and counteracting epithelial to mesenchymal transition[44]. cTEC, mcTEC and mTECI clustered together within the epithelial super-cluster (Figure 1E, right).

**Inflammatory response network activation distinguishes thymic epithelium from other anterior foregut-derived epithelia**

To better characterize the unique features of TECs among epithelial cells of other anterior foregut-derived organs, we performed subclustering of epithelial cells from all tissues (Figure 2A, Supplementary Figure S2A). Epithelial cells derived from the same organ system and those with shared function, e.g. multiciliated cells of bronchus and esophagus, demonstrated a closer transcriptional correlation (Figure 2B). Top differentially expressed genes shown for select epithelial cell subsets included known markers of key effector functions, such as lympho-attractant and peptide processing properties in cTEC (*CCL25, TBATA, PRESS16, CTSV),* and *s*urfactant production for bud tip progenitor cells of the lung (*SFTA3, SFTPB, SFTPC)*. Notably, mTECI also showed uniquely high expression of multiple interferon-inducible genes including *IFI6, IFI27, IFIT3, IFITM3, LTB, MX1* and *ISG15* (Supplementary Figure S2B).

Development is driven by the coordinated initiation of transcription factor (TF)-induced gene regulatory networks. To gain insight into the diverse gene regulatory networks underlying the development of different anterior foregut-derived epithelial cell subsets, we applied the single cell regulatory network inference and

clustering algorithm SCENIC[45, 46]. Sets of genes that are co-expressed with TFs were identified using GRNBoost2[47]. Processed modules representing a TF and its validated targets across all annotated motifs are referred to as regulons. Top differentially expressed regulons across all epithelial cells with highest expression in the thymus included the known regulators of early thymic patterning *PAX9* and *FOXG1*[20], and a broad selection of regulators of inflammatory response pathways. Out of 23 regulons with differentially high expression in the thymus, more than half represented interferon response factors (*IRF1, IRF7, IRF8, IRF9)*, signal transducer and activator of transcription factors (*STAT5A, STAT3, STAT1*), nuclear factor-κB (NF-κB) components (*NFKB1, NFKB2, RELB)*, as well as additional modulators of inflammatory response signaling such as *BHLHE40*[48] and *RFX5*, a modifier of MHCII expression[49] (Figure 2C). Of note, due to lack of consensus target sequence annotation, *FOXN1* is not included in the SCENIC algorithm. Uniform Manifold Approximation and Projection for Dimension Reduction (UMAP) visualization of the collective epithelial cell-specific TF regulon network distinctly separated the thymic epithelium from epithelial cells of the three other organ systems (Figure 2D). To identify specific subnetworks[50,51] within the complexity of the inferred co-activating TF global network, we employed the Leiden algorithm[52]. This process delineated 10 distinct subnetworks (A-J) (Figure 2E). Modular expression of the regulons defining these subnetworks demonstrated organ and cell type specificity for select subnetworks (A-D) while other subnetworks were broadly shared between organ systems (G, H, I) (Figure 2G). Subnetworks A and B were unique to the thymus with subnetwork A defining all primary canonical TEC populations cTEC, mcTEC and mTECI, and mTECII. Subnetwork B was specific to TECmyo. Subnetwork A linked the interferon response factor regulons *IRF1, IRF3, IRF7, IRF8, IRF9* with the STAT family regulons *STAT5A, STAT3, STAT2* and *STAT1,* the NF-κB family subunit regulons *NFKB1, NFKB2* and *RELB,* and the NF-κB modulators *SPIB*[53] and *HIVEP1*[54], further extending the collection of previously identified thymus-specific inflammatory response regulons. Interferons (IFN) are chemokines that induce cell intrinsic programs mediating a wide range of biological responses through the JAK-STAT pathway[55-57]. Type I IFNs include IFNα and IFNβ; type II IFNs include IFNγ. Some target genes are shared between both types, while others are specific, e.g. *IRF1* is preferentially induced by IFNγ[55,58]. IFN-activated signaling pathways can lead to NF-κB activation and, in turn, NF-κB can modulate the cellular response to IFNs[59]. Interestingly, the developmental transcription factors *FOXG1, PAX9,* and *ASCL1* were also part of the thymus-specific interferon response-centric subnetwork A (Figure 2H). Together, these findings raise the possibility that interferon response signaling during fetal development promotes thymic epithelial cell differentiation relative to other anterior foregut derived epithelia.

**Interferon response signatures in thymic epithelial cells increase during development**

To complement our gene regulatory network analysis by an orthogonal approach that includes close and distant interaction partners united by a common biological process, we performed Gene Set Enrichment Analysis (GSEA). We maximized discrimination between the primary parenchymal organ-defining epithelial cell types by subclustering cTEC, mTECI, bronchial epithelial basal cells, esophageal epithelial basal cells, esophageal suprabasal cells as well as bud tip progenitor cells of the lung (Figure 3A) and identified differentially expressed gene sets (Figure S3A). Network visualization of positively enriched gene sets across these epithelial cell subsets revealed three distinct clusters. An mTECI-predominant cluster was highly enriched for genes involved in IFNαβ and IFNγ signaling pathways, antigen presentation and TCR signaling. In a cTEC-predominant cluster, genes involved in TCR signaling, proteasome peptide processing and IFNγ signaling were enriched. The final cluster, composed of both bud tip progenitor and esophageal basal epithelial cells, demonstrated enrichment of pathways related to ribosomal function, highlighting the importance of protein synthesis in both cell types (Figure 3B).

In mTECI, genes with greatest differential expression within the enriched gene sets included previously identified downstream effectors of type I and II IFNs (*IRF1, IRF2, IRF3, IRF4, IRF7, IRF8, IRF9, IFI6, ISG15, IFI27, BST2, IFITM1, IFITM3*, *MX1* and *MX2)* as well as IFN signal transduction mediators (STAT1 and

STAT2). In cTEC, genes implicated in IFN response and signal transduction such as *IRF1, IRF8, STAT1, JAK1* and *TYK2*, as well as MHC class I and II alleles implicated in antigen presentation exhibited highest differential expression. The IFNγ receptor genes *IFNGR1* and *IFNGR2* showed increased differential expression in both mTECI and cTEC subsets (Supplementary Figure S3B).

To establish a quantifiable measure of IFN response across epithelial subsets, we integrated the expression of genes within type I and II IFN response gene sets into corresponding expression modules. Both modules were most highly expressed in mTECI, closely followed by cTEC. Conversely, type I and II IFN modules were dimly expressed in epithelial subpopulations of bronchial, lung, esophageal, and parathyroid tissues (Figure 3C, D). Type I and II IFN module expression increased with advancing gestational age in both mTECI and cTEC (Figure 3E). This trajectory parallels the expansion of thymocytes within the thymus during embryonic and fetal stages of development.

We sought to investigate the relative contributions of hematopoietic and stromal cells to the production of interferon in the thymus. Mature, activated T cells are widely recognized as producers of IFN, yet less is known about the role and capacity for IFN production in human thymocytes. Recent single cell transcriptomic studies focusing on human T lineage trajectories from CD34-positive thymus seeding hematopoietic progenitor cells found that *IRF7* and *IRF8* expression defined early T-lineage specified precursors[60,61]. Because *IFNG, IFNA1, IFNA2* and *IFNB1* single cell transcripts largely evaded detection across subsets of hematopoietic, epithelial and non-epithelial stromal origin in our data set (Supplementary Figure S3B), we performed bulk RNA-sequencing on magnetic bead-enriched fractions of CD45 positive thymocytes, TECs, bronchial epithelial cells (BECs) and lung epithelial cells (LECs). Thymocytes expressed significantly higher levels of *IFNG* than epithelial cells from thymus, bronchus or lung indicating them as the primary source of interferon within the thymus (Figure 3F). Together, these finding suggest IFN release plays a role in the lymphoid-driven maturation of the thymic epithelium.

***SOX2 and TP63 show divergent expression patterns in thymic epithelial cells***

Parallels between branching morphogenesis of the airway and patterning of thymic cortex and medulla have been proposed[27,28]. SOX2 and TP63 are master regulators of branching morphogenesis. During development of the respiratory system, the cycling expression of SOX2 and TP63 promotes forking of the airway. In the mouse, *Sox2* is highly expressed in non-branching regions and absent from branching regions, suggesting that downregulation of Sox2 is necessary for airway epithelium to respond to fork-inducing signals[62]. *Sox2* expression increases after the branching process and directly induces the expression of *Tp63*, reinstituting basal epithelial cell fate[63]. In our dataset, *SOX2* and *TP63* were highly co-expressed in bronchial and esophageal basal epithelial cells and gradually decreased as differentiation progressed towards the parenchymal regions of the lung (Figure 4A). Key foregut-defining transcription factors including *FOXA1, IRX3, IRX5*[4,64] were also highly expressed in epithelial cells of bronchus and esophagus at fetal stages but showed minimal expression in all TECs, paralleling the expression of *SOX2*. In contrast, TECs retained strong expression of the third pharyngeal pouch markers *PAX1, SIX1, FOXG1* and *EYA1*. Regulon expression of *SOX2* and *TP63* across all anterior foregut-derived epithelial cells consolidated these findings on a target gene level (Figure 4B). We further validated our data using immunohistochemistry (IHC) which confirmed concordant protein expression of SOX2 and P63 in fetal sections of human thymus, bronchus, lung and esophagus. P63 protein was strongly expressed in human fetal thymus sections, while SOX2 protein was absent in the thymus compared to all other tested tissues (Figure 4C). Although, we extensively looked for potential binding sites for *FOXN1* and any of the IFN response TFs upstream of *SOX2* and *TP63*, a direct transcriptional link between these transcriptional programs could not be established based on annotated target sequences.

**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[65-67]. A dose-dependent excess of exogenous retinoic acid intake during pregnancy is associated with morphogenic defects in thymus and parathyroid[68,69]. 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 toxicity[70,71]. These findings suggest that an intricate balance of RA signaling is essential for thymic development.
To investigate possible differences in RA metabolism as a source for divergent differentiation between anterior foregut-derived epithelial cells, we integrated a curated set of developmentally relevant RA genes[72,73] into expression modules for (I) RA synthesis, (II) RA nuclear receptor activity and (III) RA degradation (Supplementary Figure S4A). There were no major differences in RA synthetic capacity or RA nuclear activity across all analyzed anterior foregut-derived epithelial cell types. However, cTEC had a significantly higher expression of the RA degradation module compared to all other epithelial cells, including PTECs of gestationally matched samples. (Supplementary Figure S4B, C). RA degradation in cTECs was stable throughout development while RA degradation in mTECI sharply declined mid gestation (Supplementary Figure S4D). As RA production and degradation can occur autocrine or paracrine in non-target tissues[66], we investigated the contribution of other cell types within the thymic microenvironment to RA metabolism. The highest expression of RA synthesizing enzymes was found in both *ALDH1A2*-expressing (Fib-1) and *FBN1*-expressing (Fib-2) fibroblasts (Supplementary Figure S4E, F). Endothelial cells and lymphatic endothelial cells of the thymic microenvironment showed highest expression of RA nuclear activity, while cTEC again showed highest expression of the RA degradation module. These data suggest that the non-epithelial thymic stroma contributes to RA synthesis while cTECs most tightly control local RA levels.

**The thymic microenvironment provides distinct signals to diverse TEC subsets**
Receptor ligand interactions between the thymic epithelial and lymphoid compartment are essential in instructing thymic organogenesis[12,74]. 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[15]. To investigate the role the microenvironment plays during human thymic development, we analyzed the transcriptomic data using CellPhoneDB v2.0[75] and inferred signaling between TECs and non-epithelial cells (Supplementary Figure S4G).

The strongest receptor-ligand interaction inferred among TEC subtypes was that between *CD74* and the cytokine *MIF*, expressed by thymocytes, fibroblasts, and endothelial cells. CD74 is the invariant chain for Major Histocompatibility Complex (MHC)II and various roles as an immunomodulatory co-receptor have recently emerged[76,77]. Other interactions of interest included canonical and non-canonical NOTCH signaling. Historically, investigations of thymic NOTCH signaling largely focused on activating canonical Delta Like Ligands (DLLs) on TECs and their role in T cell maturation[78-89]. Our data suggest that endothelial cells and fibroblasts are also major contributors of *DLL1* and *DLL4* signals. Notably, the inhibitory non-canonical NOTCH ligand *DLK1* is expressed on both endothelial cells and mcTECs and interactions with cTECs via the *NOTCH2* receptor are inferred. Non-canonical inhibitory DLKs have been implicated in the inhibition of differentiation into adipocytes in other epithelial cells[90-93]. In the context of adipose transformation of the thymic stroma, these findings open new avenues of investigation into thymic epithelial maintenance. Thymic fibroblasts appeared as the main source of *FGF7* and *FGF10*. cTEC were the primary source of *BMP7*. *WNT5A* was expressed in fibroblasts and lymphatic endothelial cells and interactions with a wide range of canonical and non-canonical WNT receptors on cTECs (*ROR1*, *ROR2*, *FZD3*, *FZD5*, *FZD6* and *PTPRK)* were inferred, consistent with recent murine studies[7]. As an indicator of validity, we also inferred interactions between well-known receptor-ligand pairs[94], including thymocyte *CCR9* and stromal *CCL25*, as well as thymocyte *CCR7* and stromal *CCL19*.

**The cellular composition of the thymus is dynamic during development**

To examine lineage trajectories and epithelial cell diversification throughout thymic development, we subclustered TECs from samples of subsequent gestational ages (pcw 6, 7, 16, 17, 18 and 3 months post birth) and verified successful batch correction (Figure 5A). tSNE visualization of this data set revealed a distinct subset of TECs in spatial proximity to mcTEC, not previously identified in our earlier analysis (*Figure 1*). The newly identified population clustered independently due to high differential expression of 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 5B, Supplemental Figure S5A). Accordingly, we designated this population cycling TECs (cyTEC). The intertypical mcTEC population showed high differential expression for the Early Growth Response Family TFs *EGR1*, *EGR2* and *EGR3* and the C-C motif chemokine ligand *CCL2*. To date, only *CTGF* and *DLK2* have been described as mcTEC defining genes[12]. The previously noted small population of PTECs (*PTH, GCM2, CASR, TBX1, GATA3*) present in one of the early thymus samples was also included in the dataset. All other TEC population were characterized as before.
Cycling TECs were predominantly present in the very early sample and decreased in frequency with increasing gestational ages, revealing why they had previously gone unnoticed (Figure 5C). After birth, cycling TECs were only marginally present. The fraction of mcTEC remained constant throughout development until 3 months post birth. cTEC were the most abundant TEC population until mid-gestation after which their frequency decreased. Starting mid gestation mTECI were the most abundant TEC population and remained predominant until 3 months post birth. mTECII showed a relative increase in number over time but remained infrequent relative to their frequency during murine development[16,95] . mTECIII/IV largely appeared after birth.

By overall PCA-based gene expression correlation, cycling TECs demonstrated greater similarity to mcTEC than to any other population, while mcTECs and cTECs showed greatest correlation overall (Figure 5D, Supplementary Figure S5B. The findings were upheld after cycling genes were removed from the analysis.

Histone acetyltransferases and histone deacetylases are important regulators of chromatin remodeling and have been shown to alter IFNα- and IFNγ-inducible transcription[96,97]. Furthermore, Hdac3 has recently been described as master regulator of murine mTEC development[98]. To assess if histone deacetylases are differentially expressed in human TECs, we examined *HDAC1-8* in our dataset. *HDAC4* was the most differentially expressed histone deacetylase between cTECs and mTECIs (Figure 5E). *HDAC4* peaked mid-gestation and was exclusively expressed in cTECs after birth (Figure 5F). Cycling TECs showed highest expression of *HDAC1* which remained stable during gestation (Supplementary Figure S5C). mcTEC showed a similar HDAC expression profile as cycling TECs, but HDAC expression was confined to a smaller fraction of cells. The most highly expressed HDACs in mTECIII/IV and TECmyo/neuros were *HDAC5* and *HDAC2,* respectively. To corroborate preferential expression of *HDAC4* in cTEC, we performed IHC on human fetal thymus tissue. HDAC4 colocalized with the cTEC marker PSMB11 (Figure 5G), supporting our transcriptional findings.
Overall, the differential expression of specific *HDAC* subtypes correlated with cell identity. Gene expression signature, correlation analysis, and HDAC pattern raise the possibility that cycling TEC are an early proliferating progenitor pool most closely resembling mcTEC.

**Distinct gene regulatory networks direct cortical and medullary lineage specification**
To dissect the complex developmental sequences that give rise to cortical and medullary thymic epithelial lineages, we employed tSpace, an algorithm based on the concept of "trajectory space", in which high dimensional data are reduced by defining cells through nearest neighbor relative distance[119]. The tSpace projection rendered an mcTEC centric view of the thymus with mTECI and cTEC populations flanking mcTEC (Figure 6A). We used Slingshot to infer lineage trajectories along pseudotime[120]. To focus on fundamental lineage branchpoints, we restricted the initial analysis to cycling TEC, mcTEC, cTEC and mTECI. Slingshot

inferred two distinct lineage trajectories, originating at cycling TEC, bifurcating at mcTEC and ending at mTECI (T1) and cTECs (T2) (Figure 6B). We then expanded the analysis to include mTECII which extended the original trajectory T1 (Supplementary Figure S6A). Although the lineage tracing was set to originate at cycling TEC, the trajectories did not branch until the mcTEC cluster, suggesting mcTEC as the immediate biological precursor of cortical and medullary lineages.
Combining cellular barcoding and scRNA-seq for *in vivo* lineage tracing, a recent study identified distinct prenatal and postnatal thymic epithelial progenitor populations in the murine thymus[104]. When projecting the modular expression of the human homolog genes defining these progenitor populations onto tSpace, the prenatal progenitor signature most closely overlapped with mcTEC while the postnatal progenitor signature shifted along the medullary lineage trajectory (Supplementary Figure S6B). These observations suggest that a prenatal progenitor population with mcTEC signature gives rise to cortical and medullary lineages in both humans and mice.

To explore transcriptional drivers directing lineage bifurcation, we interrogated mTECI and cTEC specific gene regulatory networks using GRNBoost2[45], an algorithm for regulatory network inference based on gene co-expression. The GRNBoost2 adjacency matrix was filtered to isolate TFs that target cTEC and mTECI defining markers. Two discrete networks, one with mTECI predominance and one with cTEC predominance were linked by abundant central nodes identifying shared regulators. *IRF7, IRF8, IRF9, STAT1, BHLHE40*, *ASCL1, ANXA1, FOSB* and *SOX4* were primary drivers of medullary fate while *NFATC1, IRX6, HMGB1, HMGB2, SOD1, ETV1* and *CYB5R1* were differential regulators of cTEC provenance. Key TFs shared between medullary and cortical networks included *PAX1, FOXN1, IRF1, NFκB1, SPIB, ASCL2, SREBF1, PRDX5, TAGLN2, JUNB* and *ODC1* (Figure 6C). Notwithstanding, central TFs exhibited a lineage bias based on the amount of lineage-specific target genes. These lineage biases were also validated transcriptionally (Supplementary Figure S6C). Interestingly, *NFκB1,* which has been previously associated with medullary fate[40,121-123], displayed a cTEC bias (Figure 6D).
To infer dynamic changes in gene expression along lineage trajectories, we used tradeSeq, a gene additive model based on negative binomial distribution that allows inference of both, within-lineage and between-lineage differential expression[124]. Differential gene expression was determined along defined pseudotime intervals (knots). Knots 1 and 2, located in cycling TEC and mcTEC, overlapped between trajectories. TradeSeq was run on the top 2000 most variably expressed genes. We used the earlyDE test to determine overall lineage differences between knots 2 to 5, represented as rolling average expression of the 60 genes with the highest Wald statistic score (pvalue≈0) (Figure 6E). Top differentially expressed genes along the medullary lineage again included IFN response elements *IFITM1, IFIT1, IFITM3, IFI6, ISG15, TMEN173* , encoding for the stimulator of interferon response (STING) protein[125], *SOD2*, encoding for superoxide dismutase 2 as well as *CCDC3*, a recently described TP63 target[126]. Genes with highest differential expression in the cortical trajectory included the transcription factors *NFκB1* and *ASCL2*. Notably, interrogating lineage segments earlier after the mcTEC branch point (knots 2 to 4), polarized some of the previously shared regulators. E.g. *SPIB* now demonstrated differentially high expression in mTECs while *FOXN1* and *PAX1* were increased in cTEC (Supplementary Figure S6D).

We also interrogated differentially expressed TF regulons along pseudotime (Figure 6F). Due to limitations of TF motif annotation within SCENIC[43], targets inferred to be activated by the known thymic transcription factors *FOXN1* and *PAX1* were absent. Differentially expressed regulons in cycling TECs (*E2F1, E2F2, E2F7, E2F8*) and mcTEC (*EGR2, EGR23*) represent within lineage differential expression reflecting the overlap of knots 1 and 2. The extended medullary trajectory demonstrated differentially high expression of previously characterized inflammatory response regulators (*IRFs*, *STATs*, etc) except for *IRF6* and *STAT6* regulons which were higher along T2. The previously emerging *NFκB1* bias in the cortical lineage, as well as the dichotomy between *ASCL1* (T1) and *ASCL2* (T2) were also corroborated on a regulon level. *SREBF2* was higher along T1

while *SREBF1* was higher along T2. Interestingly, many *SOX* regulons demonstrated higher expression levels along the medullary lineage, specifically *SOX2* which gained expression while *TP63* decreased. High *AIRE* regulon expression marked the mTECII stage along pseudotime. However, *CREB3L1, RELB, SPIB, SPIC* and *POU2F3* regulons which identified AIRE as a target, exhibited slow and steady gain in regulon expression starting early along T1. MHC class I and II expression along extended trajectory 1 demonstrated consistent transcriptional gain along the mTECI stage followed by a marked increase in mTECII (Figure 6G).

**Patterns of thymic epithelial cell identity and peripheral tissue antigen expression**

Central tolerance is induced in the thymus through negative selection of autoreactive T cells with high affinity to self-antigens presented on TECs[127,128]. Negative selection must "anticipate" that T cells will encounter peripheral tissue antigens (PTA) once released from the thymus. This was believed to occur by cellular mimicry of medullary TECs resembling peripheral tissues[129]. The concept was largely abandoned with discovery of the TF Auto Immune Regulator (AIRE) in a subset of medullary TECs[130]. AIRE's function is unique in that it triggers PTA expression independent of target-sequence specificity by leveraging chromatin plasticity[131]. FEZF2 has been described as an additional regulator of PTA expression[132], although its mechanistic underpinnings remain illdefined[133]. AIRE and FEZF2 together are thought to account for approximately 60% of PTA expression[134]. This suggests the existence of alternate mechanisms inducing PTA expression. To investigate signatures of PTA expression and their relationship to *AIRE* and *FEZF2* during development, we isolate the medullary subpopulations mTECII, mTECIII/IV, TECmyo and TECneuro. mTECII and mTECIII/IV separated distinctly from the TECmyo and TECneuro population (Figure 7A). Unsupervised clustering revealed the six populations, immature *AIRE*$^{low}$ mTECII (*CCL19*), cycling mTECII (*TOP2A*), mature *AIRE*$^{high}$ mTECII, post-*AIRE* mTECIII/IV (*SPINK5*), muscle-like TEC (*ACTA1*) and neuroendocrine-like TEC (*BEX1*) (Figure 7B, C). mature *AIRE*$^{high}$ mTECII and post-*AIRE* mTECIII/IV populations significantly increased in frequency throughout the observed time interval while cycling mTECII and muscle and neuroendocrine-like TECs were stable in frequency beyond midgestation (Figure 7D, E). Maturation and specification of medullary TECs from mTECI to mature *AIRE*$^{high}$ mTECII stages were marked by decreasing *CCL19* and *TP63* expression while *CLDN3* and *CLDN4* steadily increased. *AIRE*$^{high}$ mTECII cells exclusively co-expressed both *AIRE* and *FEZF2*. mTECIII/IV were high in *SPINK5* and *IVL* indicating post-*AIRE* provenance. Characteristic peripheral tissue defining signatures in post-*AIRE* TECs included *GRHL1*, *GRHL3* and *KRTDAP*, denoting a keratinocyte-like state while *POU2F3* indicated the emergence of Tuft-like cells. Neuroendocrine-like TECs high in *BEX1*, *NEUROD1*, *NRX1* and muscle-like TECs high in *MYOG1*, *ACTA1*, and *TTN1* exhibited minimal keratinization and absent *FOXN1* expression but were high in *EPCAM*, supporting their epithelial nature (Figure 7F). To distinguish global cell fate from isolated, possibly promiscuous gene expression, we probed for tentative TF-target interactions using SCENIC. Muscle and neuroendocrine-like TECs showed full expression of muscle-specific (*MYOD1, MYOG1, MEF2A, MEF2C, TEAD4, MYF6*) or neuroendocrine-specific (*NEUROD1, NEUROD2, SOX4, SOX11, POU4F1, NPDC1*) regulons (Figure 7G). In contrast, *AIRE*-positive and post-AIRE medullary TECs did not express comprehensive sets of peripheral tissue defining regulons. Interestingly, this correlated inversely with the expression of antigen presenting molecules on the cell surface. MHC class I and II expression was highest in mature mTECII whereas muscle and neuroendocrine-like TECs expressed low levels of MHC class I and no MHC class II (Figure 7H). These findings point to divergent patterns of AIRE-dependent and AIRE-independent peripheral tissue antigen expression in discrete thymic epithelial cell subsets and suggest that maturation and diversity of the medullary TEC compartment extends beyond birth.

**DISCUSSION**

Progress in genomic analysis methods have redefined our concept of thymic epithelial architecture. The goal of this study was to understand the dynamic signals that drive thymic fate during development. Comparing human fetal epithelial cells from different AFG-derived organs, we provide compelling evidence that type I and II IFNs exert inductive signals distinctive to the thymic epithelial stroma. We further show that the pervasive effects of

IFN extend differentially into the thymic epithelial compartment, reflected in characteristic *IRF* and *NFκB* signatures specific to cortical and medullary lineages, respectively. Previous studies pointed to an activation of the TNF signaling axis, mediated by CD40 and RANK, during medullary TEC development[135,136]. These signals were shown to stimulate the NFκB pathway and proved essential for various aspects of mTEC fate and function[121-123]. Our data points to a central role for type I and II IFNs during thymic morphogenesis with IRFs, STATs and NFκB emerging as downstream mediators that fine-tune its effects along cortical and medullary trajectories. How exactly the global effects of type I and II IFNs are differentially mediated in cTEC and mTEC remains the subject of ongoing investigations. Our dataset opens up multiple novel avenues of inquiry. First, the RA and IFN signaling axes intersect, with RA augmenting the effect of IFN[137]. IFNγ has, in turn, been shown to selectively increase the expression of *IRF1* and *STAT1*[138-140]. Our data indicates that RA is preferentially degraded in cTEC, which could explain the relatively lower average expression of *IFN* target genes in the cortex compared to the medulla. Cross signaling also occurs between IFN and the MIF:CD74 complex[75,81,82]. Contrary to other cytokines that are locally induced, MIF is constitutively expressed, suggesting that its effect is regulated on the receptor level[75-77]. CD74 acts as a transcriptional modulator via the ERG1/2 and NFκB pathways[75,78,79] and promotes the production of proinflammatory molecules, including IFNγ, TNF, IL-1β, IL-2, IL-6, IL-8[75,81,82]. Previous studies suggested that CD74 positively influences mTEC cellularity in the mouse[80]. We posit that differential expression patterns in genes altering the IFN response, including RA, MIF and other cytokines and their receptors , help orchestrate the spatially distinct thymic niches that ultimately give rise to a diversity of unique TEC subsets.

This framework, however, does not preclude that differences in patterning underlying the thymic cortex and medulla precede the recruitment of interferon-releasing hematopoietic elements into the thymus. To this end, our data on the divergent expression of *SOX2, FOXA1, IRX3, IRX5* and *TP63* is intriguing and consistent with the historical hypothesis that a program overriding classical basal epithelial cell fate must be initiated for the thymus to mature[22,23]. When exactly *SOX2* begins to decrease during thymic morphogenesis is not directly accessible in human samples. In the mouse, *Sox2* is expressed in the PPEIII[141,142]. One might therefore speculate that *SOX2* and *TP63* remain co-expressed while pharyngeal endoderm cells maintain apical-basal polarity, and that changes in the *SOX2-TP63* expression pattern occur as the thymus anlage expands into three dimensional space.

Our study also yields novel insights into the proliferative dynamics within the human thymic epithelial compartment. We denote a pool of highly proliferative progenitors designated cycling TEC early during embryonic development that recedes by mid gestation. Correlation studies and patterns of HDAC expression indicate similarities between cycling TEC and mcTEC. The latter are, in turn, is not merely a population of intertypical TEC with biphenotypic cortical and medullary features. Rather, mcTEC are defined by dim expression of canonical cTEC or mTEC markers such as PSMB11, PRSS16, TBATA or CCL19 and may be more accurately designated non-mcTEC. Instead, they show uniquely high expression in *CCL2*, *MAFF* and the previously recognized non-canonical notch ligand *DLK2*. Trajectory analysis demarks mcTEC as the bona fide progenitor population from which the developmental bifurcation towards cortical and medullary lineages originates. mcTEC show lower MHCII expression and keratinization than their lineage-committed cortical and medullary counterparts supporting their immature developmental state. Their fraction within the entirely of the TEC pool remains constant during embryonic, fetal and early postnatal stages of development.

We specifically leveraged our data set to understand what distinguishes the regulatory networks underlying cTEC and mTEC commitment. Beyond differences in IFN processing resulting in an *NFκB1* bias in the cortex and *IRF-STAT* bias in the medulla, we noted a seesaw in regulon expression of *SREBF1* versus *SREBF2, SOX2* versus *TP63,* and *ASCL1* versus *ASCL2*. ASCLs are achaete-scute family basic helix-loop-helix transcription

factors. In the context of ectoderm development, *ASCL1* is necessary and sufficient to induce neuronal lineage progenitors[51,52]. *ASCL1* expression has previously been identified in human TEC subsets, albeit fate mapping experiments in *Ascl1*-reporter mice have not yielded definitive conclusions on the role of *Ascl1* during thymic epithelial lineage commitment and specification[5].
Also of interest, *CCL19* $^{high}$ mTECI comprise by far the largest population of medullary TECs during development. Classical AIRE-positive mTECII constitute a surprisingly small fraction of TECs, albeit with a rising trajectory after birth. Keratinocyte-like mTECIII and Tuft cell-like mTECIV display a classic post-AIRE signature and largely emerge postnatally. Muscle- and neuroendocrine-like TECs, on the other hand, are stably present starting midgestation, show reduced keratinization and lack expression of post-AIRE markers, *FOXN1* and antigen presenting molecules. High expression of lineage defining regulons in muscle- and neuroendocrine-like TECs mimics native muscle and neuroendocrine tissue signatures. Nonetheless, high levels of EPCAM support their epithelial origin. Fundamental differences in cell identity and surface architecture, specifically MCH class II expression, between classical post-AIRE mTEC and muscle- and neuroendocrine-like TECs point to distinct mechanisms for PTA expression that emerge in parallel at different stages during development and may represent alternate means of tolerance induction.

A recent study in Aire-sufficient and Aire-deficient mice found a distinct motif enrichment for members of lineage defining TF families, such as *FoxA, FoxJ1, Sox8, Hnf4* in small numbers of mTECs with high expression of PTA signatures[40] (Supplementary Figure 7A). These were not present in our developmentally early human data set. It seems likely that the AIRE-dependent diversification of the thymic medulla has not yet reached its full expression *in utero* and that more lineage defining TFs will appear over time. From a clinical perspective, these data should make us carefully consider the indications for iatrogenic removal of large parts of thymus shortly after birth.

Between the extended maturation of the thymic medulla during development and the involution of the thymus in the second decade of life, the window of thymic peak performance is relatively small. In addition, genetic defects and a myriad of iatrogenic and biological insults to the thymus, such as irradiation, medication, infection, and graft-versus-host disease can compromise its structural integrity and function. This underscores the urgent need for therapies to regenerate thymic tissues. Our study dissecting the transcriptional framework underlying commitment and specialization of the human thymic epithelial stroma during development reveals novel regulators of thymic development and offers actionable insights can be translated to advance assays for the in vitro generation of thymic epithelial replacement therapies from pluripotent stem cells.

## 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 6-7 samples was obtained in collaboration with the laboratory of Prof. 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),$10^8$ cells/column, according to the manufacturer's instructions. Single cell suspensions of all tissues, (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 manufacturer's instructions. Quality control for thymocyte depletion and epithelial cell enrichment for thymic single cell suspensions was performed by flow cytomery on Beckman FACSAria II with CD45-APC, CD3-FITC, CD205-PE, EpCAM-APC, and Live Dead-Near InfraRed-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-Bead in Emulsions (GEMs) were generated from single cell suspensions using the 3′ V3 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 sample indexing as per manufacturer's protocol. Libraries were quantified using Bioanalyzer (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. After genome alignment and filtering for mitochondrial content and number of genes/reads per cell, the data was normalized with regularized negative Binomial regression using Seurat V.4.0[99]. Cells from different samples were then combined using integration anchors[100].

**Differential gene expression**

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

**SCENIC**

The single cell regulatory network inference and clustering algorithm, SCENIC was used to identify transcription factor regulons as previously descripbed[45, 46]. In brief, sets of genes that are co-expressed with TFs were identified using GRNBoost2[47]. Then, each potential TF-target interaction was subjected to *cis*-regulatory motif analysis to detect putative direct-binding targets. Interactions with significant motif enrichment of the respective upstream TF were retained while indirect targets lacking motif support were removed.

**Gene set analysis**

Gene sets were obtained from the Molecular Signature Database (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 set 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 were then reduced using uniform manifold approximation and projection (UMAP). We then fit principal 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 psueudotime 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 transcriptional expression values or Regulon AUC values were used to construct an expression matrix along 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,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), log1p normalized expression matrices were filtered to remove low quality genes according to an average expression threshold value across all cells of 0.01. The filtered expression matrices were converted into a loom file and passed into GRNBoost2 to evaluate co-expression based network interference. The resultant adjacency matrices were passed into the RcisTarget step for regulon prediction. Predicted regulons and an aggregate of their targets from different motifs were 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 Seurat's centered log ratio transformation. The data were subject to a standard Seurat clustering pipeline consisting of FindVariableFeatures(), ScaleData(), followed by RunPCA(). UMAP dimensionality reduction was carried out using the first 20 PCs for the pcw 16,17,18 samples and the first 18 PCs for the pcw 6w sample. 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 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**
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). Only interactions with a GRNBoost2 importance score >20 were retained (440/7363 TF-TF interactions). 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 in ggnetwork using the Fruchterman-Reingold layout algorithm.

**Trajectory Heatmap Construction**
Scaled expression values and SCENIC Regulon AUC values 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.

**GRNBoost2 Network Construction**
GRNBoost2 output of all cells within the vertical analysis was filtered using an "importance" threshold > 2 - 376,017 interactions out of 6,342,692 were retained (Python/3.9.0, Pandas/1.5.2). Seurat's FindMarkers(test.use = "wilcox") was used to identify differentially expressed genes (DEGs) between cTEC and mTEC I clusters. If a transcription factor had >= 20 targets that intersected with the top 50 DEGs of either cluster, an edge was drawn between the TF and all tentative targets. Transcription factors which met the threshold for both clusters were included. This filtered edge list was then passed to igraph ggplot2/3.4.0 and ggnetwork/0.5.10 for graphical network representation. Nodes were force directed in ggnetwork using the Fruchterman-Reingold layout algorithm.

**SCENIC GRNBoost2/RcisTarget Refined Network Construction**
An adjacency matrix between all TF-target interactions was constructed using the SCENIC RcisTarget output of mTECII,mTECIII/IV,TEC myo/neuro clusters (Python/3.9.0, Pandas/1.5.2). Seurat's FindMarkers(test.use = "wilcox") was used to identify differentially expressed genes (DEGs) between Muscle-like and Neuroendocrine-like clusters. If a transcription factor had >= 15 targets that intersected with the top 50 DEGs of either cluster, an edge was drawn between the TF and all tentative targets. No transcription factors which met the threshold for both clusters were found. The filtered edge list was then passed into igraph ggplot2/3.4.0 and ggnetwork/0.5.10 for graphical network representation. Nodes were force directed in ggnetwork using the Fruchterman-Reingold layout algorithm.

**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 table 1 for reference) diluted in PBS with 1% DS in a humid chamber for 16 hours at 4˚C. Slides were 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

## AUTHOR INFORMATION

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

## AUTHOR CONTRIBUTIONS

A.M., A.G., and K.G.W. interpreted the data and wrote the manuscript. The experiments were designed, performed and analyzed by A.M., B.S., A.G., and K.G.W.. P.F.S, K.M.H., D.K.K., H.D.N., 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. C. G., V.S. and P.K. reviewed design, bioinformatic analysis and edited the manuscript.

## ACKNOWLEDGEMENTS

This work was supported by the California Institute of Regenerative Medicine, the Emerson Collective, the Baxter Foundation. Specifically, 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, and Priscila F. Slepicka for the graphical illustrations.

## REFERENCES

1. Kadouri N, Nevo S, Goldfarb Y, Abramson J. Thymic epithelial cell heterogeneity: TEC by TEC. *Nat Rev Immunol.* 2020;20(4):239-253.
2. Abramson J, Anderson G. Thymic Epithelial Cells. *Annu Rev Immunol.* 2017;35:85-118.
3. 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.
4. 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.
5. 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.
6. 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.
7. 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.
8. 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.
9. 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.
10. 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.
11. 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.

12. 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).
13. 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.
14. 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.
15. 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.
16. Baran-Gale J, Morgan MD, Maio S, et al. Ageing compromises mouse thymus function and remodels epithelial cell differentiation. *Elife.* 2020;9.
17. Aad G, Abajyan T, Abbott B, et al. Measurement of top quark polarization in top-antitop events from proton-proton collisions at radicals=7 TeV using the ATLAS detector. *Phys Rev Lett.* 2013;111(23):232002.
18. Zorn AM, Wells JM. Vertebrate endoderm development and organ formation. *Annu Rev Cell Dev Biol.* 2009;25:221-251.
19. Graham A, Richardson J. Developmental and evolutionary origins of the pharyngeal apparatus. *Evodevo.* 2012;3(1):24.
20. Farley AM, Morris LX, Vroegindeweij E, et al. Dynamics of thymus organogenesis and colonization in early human development. *Development.* 2013;140(9):2015-2026.
21. 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.
22. 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.
23. 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.
24. Nehls M, Kyewski B, Messerle M, et al. Two genetically separable steps in the differentiation of thymic epithelium. *Science.* 1996;272(5263):886-889.
25. Anderson G, Jenkinson EJ. Lymphostromal interactions in thymic development and function. *Nat Rev Immunol.* 2001;1(1):31-40.
26. 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.
27. 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.
28. 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.
29. Corbeaux T, Hess I, Swann JB, Kanzler B, Haas-Assenbaum A, Boehm T. Thymopoiesis in mice depends on a Foxn1-positive thymic epithelial cell lineage. *Proc Natl Acad Sci U S A.* 2010;107(38):16613-16618.
30. Flanagan SP. 'Nude', a new hairless gene with pleiotropic effects in the mouse. *Genet Res.* 1966;8(3):295-309.
31. 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.
32. 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.
33. 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.

34. Dooley J, Erickson M, Roelink H, Farr AG. Nude thymic rudiment lacking functional foxn1 resembles respiratory epithelium. *Dev Dyn.* 2005;233(4):1605-1612.
35. Cordier AC. Ultrastructure of the cilia of thymic cysts in "nude" mice. *Anat Rec.* 1975;181(2):227-249.
36. Shultz LD, Lyons BL, Burzenski LM, et al. Human lymphoid and myeloid cell development in NOD/LtSz-scid IL2R gamma null mice engrafted with mobilized human hemopoietic stem cells. *J Immunol.* 2005;174(10):6477-6489.
37. Zeleniak A, Wiegand C, Liu W, et al. De novo construction of T cell compartment in humanized mice engrafted with iPSC-derived thymus organoids. *Nature methods.* 2022;19(10):1306-1319.
38. 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.
39. 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.
40. 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.
41. 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.
42. Michelson DA, Hase K, Kaisho T, Benoist C, Mathis D. Thymic epithelial cells co-opt lineage-defining transcription factors to eliminate autoreactive T cells. *Cell.* 2022;185(14):2542-2558 e2518.
43. Sakaguchi M, Huh NH. S100A11, a dual growth regulator of epidermal keratinocytes. *Amino Acids.* 2011;41(4):797-807.
44. 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.
45. Aibar S, Gonzalez-Blas CB, Moerman T, et al. SCENIC: single-cell regulatory network inference and clustering. *Nature methods.* 2017;14(11):1083-1086.
46. Van de Sande B, Flerin C, Davie K, et al. A scalable SCENIC workflow for single-cell gene regulatory network analysis. *Nat Protoc.* 2020;15(7):2247-2276.
47. Moerman T, Aibar Santos S, Bravo Gonzalez-Blas C, et al. GRNBoost2 and Arboreto: efficient and scalable inference of gene regulatory networks. *Bioinformatics.* 2019;35(12):2159-2161.
48. Cook ME, Jarjour NN, Lin CC, Edelson BT. Transcription Factor Bhlhe40 in Immunity and Autoimmunity. *Trends Immunol.* 2020;41(11):1023-1036.
49. DeSandro AM, Nagarajan UM, Boss JM. Associations and interactions between bare lymphocyte syndrome factors. *Mol Cell Biol.* 2000;20(17):6587-6599.
50. Silva FN, Albeshri A, Thayananthan V, Alhalabi W, Fortunato S. Robustness modularity in complex networks. *Phys Rev E.* 2022;105(5-1):054308.
51. Tandon A, Albeshri A, Thayananthan V, Alhalabi W, Radicchi F, Fortunato S. Community detection in networks using graph embeddings. *Phys Rev E.* 2021;103(2-1):022316.
52. Traag VA, Waltman L, van Eck NJ. From Louvain to Leiden: guaranteeing well-connected communities. *Sci Rep.* 2019;9(1):5233.
53. 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.
54. 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.
55. Pestka S, Krause CD, Walter MR. Interferons, interferon-like cytokines, and their receptors. *Immunol Rev.* 2004;202:8-32.
56. 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.

57. 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.
58. Platanias LC. Mechanisms of type-I- and type-II-interferon-mediated signalling. *Nat Rev Immunol.* 2005;5(5):375-386.
59. Pfeffer LM. The role of nuclear factor kappaB in the interferon response. *J Interferon Cytokine Res.* 2011;31(7):553-559.
60. 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.
61. 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.
62. 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.
63. 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.
64. 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.
65. 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.
66. Duester G. Retinoic acid synthesis and signaling during early organogenesis. *Cell.* 2008;134(6):921-931.
67. 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.
68. Lammer EJ, Chen DT, Hoar RM, et al. Retinoic acid embryopathy. *N Engl J Med.* 1985;313(14):837-841.
69. 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.
70. 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.
71. Pavan M, Ruiz VF, Silva FA, et al. ALDH1A2 (RALDH2) genetic variation in human congenital heart disease. *BMC Med Genet.* 2009;10:113.
72. Rhinn M, Dolle P. Retinoic acid signalling during development. *Development.* 2012;139(5):843-858.
73. Kam RK, Deng Y, Chen Y, Zhao H. Retinoic acid synthesis and functions in early embryonic development. *Cell Biosci.* 2012;2(1):11.
74. Han J, Zuniga-Pflucker JC. A 2020 View of Thymus Stromal Cells in T Cell Development. *J Immunol.* 2021;206(2):249-256.
75. 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.
76. 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.
77. 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.
78. 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.
79. Robey E. Regulation of T cell fate by Notch. *Annu Rev Immunol.* 1999;17:283-295.
80. Deftos ML, Bevan MJ. Notch signaling in T cell development. *Curr Opin Immunol.* 2000;12(2):166-172.

81. 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.
82. 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.
83. 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.
84. 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.
85. Deftos ML, He YW, Ojala EW, Bevan MJ. Correlating notch signaling with thymocyte maturation. *Immunity.* 1998;9(6):777-786.
86. 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.
87. 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.
88. 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.
89. 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.
90. 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.
91. 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.
92. Smas CM, Sul HS. Pref-1, a protein containing EGF-like repeats, inhibits adipocyte differentiation. *Cell.* 1993;73(4):725-734.
93. 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.
94. 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.
95. Wells KL, Miller CN, Gschwind AR, et al. Combined transient ablation and single-cell RNA-sequencing reveals the development of medullary thymic epithelial cells. *Elife.* 2020;9.
96. 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.
97. 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.
98. Goldfarb Y, Kadouri N, Levi B, et al. HDAC3 Is a Master Regulator of mTEC Development. *Cell Rep.* 2016;15(3):651-665.
99. 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.
100. Stuart T, Butler A, Hoffman P, et al. Comprehensive Integration of Single-Cell Data. *Cell.* 2019;177(7):1888-1902 e1821.
101. 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.
102. Komiya Y, Habas R. Wnt signal transduction pathways. *Organogenesis.* 2008;4(2):68-75.

103. 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.
104. Nusser A, Sagar, Swann JB, et al. Developmental dynamics of two bipotent thymic epithelial progenitor types. *Nature.* 2022;606(7912):165-171.
105. 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.
106. 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.
107. 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.
108. Simon T, Bromberg JS. Regulation of the Immune System by Laminins. *Trends in Immunology.* 2017;38(11):858-871.
109. 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.
110. <1020.full.pdf>.
111. 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.
112. Bleul CC, Boehm T. Chemokines define distinct microenvironments in the developing thymus. *Eur J Immunol.* 2000;30(12):3371-3379.
113. Cedile O, Lobner M, Toft-Hansen H, et al. Thymic CCL2 influences induction of T-cell tolerance. *J Autoimmun.* 2014;55:73-85.
8. Baran-Gale J, Morgan MD, Maio S, et al. Ageing compromises mouse thymus function and remodels epithelial cell differentiation. *Elife.* 2020;9.
9. Wells KL, Miller CN, Gschwind AR, et al. Combined transient ablation and single-cell RNA-sequencing reveals the development of medullary thymic epithelial cells. *Elife.* 2020;9.

114. Goldfarb Y, Kadouri N, Levi B, et al. HDAC3 Is a Master Regulator of mTEC Development. *Cell Rep.* 2016;15(3):651-665.
115. 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.
116. 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.
117. 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.
118. 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.
119. Dermadi D, Bscheider M, Bjegovic K, et al. Exploration of Cell Development Pathways through High-Dimensional Single Cell Analysis in Trajectory Space. *iScience.* 2020;23(2):100842.
120. Street K, Risso D, Fletcher RB, et al. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics.* 2018;19(1):477.
121. 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.
122. 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.

123. 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.
124. Van den Berge K, Roux de Bezieux H, Street K, et al. Trajectory-based differential expression analysis for single-cell sequencing data. *Nat Commun.* 2020;11(1):1201.
125. Barber GN. STING: infection, inflammation and cancer. *Nat Rev Immunol.* 2015;15(12):760-770.
126. Liao W, Liu H, Zhang Y, et al. Ccdc3: A New P63 Target Involved in Regulation Of Liver Lipid Metabolism. *Sci Rep.* 2017;7(1):9020.
127. Stritesky GL, Xing Y, Erickson JR, et al. Murine thymic selection quantified using a unique method to capture deleted T cells. *Proc Natl Acad Sci U S A.* 2013;110(12):4679-4684.
128. Klein L, Kyewski B, Allen PM, Hogquist KA. Positive and negative selection of the T cell repertoire: what thymocytes see (and don't see). *Nat Rev Immunol.* 2014;14(6):377-391.
129. Farr AG, Rudensky A. Medullary thymic epithelium: a mosaic of epithelial "self"? *J Exp Med.* 1998;188(1):1-4.
130. Anderson MS, Venanzi ES, Klein L, et al. Projection of an immunological self shadow within the thymus by the aire protein. *Science.* 2002;298(5597):1395-1401.
131. Kaiser C, Bradu A, Gamble N, Caldwell JA, Koh AS. AIRE in context: Leveraging chromatin plasticity to trigger ectopic gene expression. *Immunol Rev.* 2022;305(1):59-76.
132. Takaba H, Morishita Y, Tomofuji Y, et al. Fezf2 Orchestrates a Thymic Program of Self-Antigen Expression for Immune Tolerance. *Cell.* 2015;163(4):975-987.
133. Qi Y, Zhang R, Lu Y, Zou X, Yang W. Aire and Fezf2, two regulators in medullary thymic epithelial cells, control autoimmune diseases by regulating TSAs: Partner or complementer? *Frontiers in immunology.* 2022;13:948259.
134. Rode I, Rodewald HR. Transcription factor hijacking in the name of tolerance. *Cell.* 2022;185(14):2398-2400.
135. 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.
136. 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.
137. 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.
138. 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.
139. 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.
140. 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.
141. Rizzoti K, Lovell-Badge R. SOX3 activity during pharyngeal segmentation is required for craniofacial morphogenesis. *Development.* 2007;134(19):3437-3448.
142. 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.

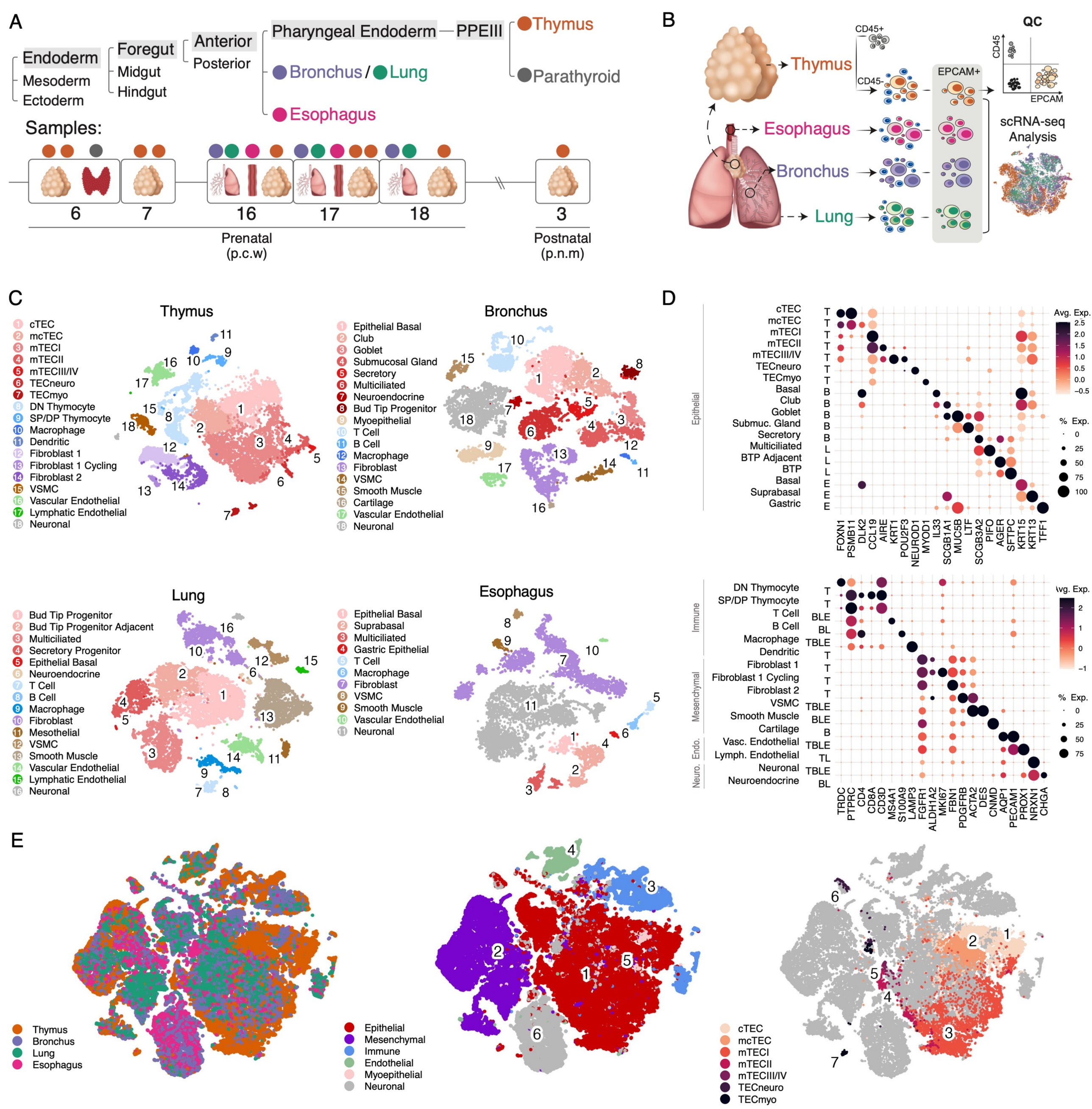

A
Endoderm
Mesoderm
Ectoderm
Foregut
Midgut
Hindgut
Anterior
Posterior
Pharyngeal Endoderm
PPEIII
Thymus
Parathyroid
Bronchus / Lung
Esophagus
Samples:
6
7
16
17
18
3
Prenatal (p.c.w)
Postnatal (p.n.m)
B
Thymus
Esophagus
Bronchus
Lung
CD45+
CD45-
EPCAM+
QC
CD45
EPCAM
scRNA-seq Analysis
C
Thymus
1 cTEC
2 mcTEC
3 mTECI
4 mTECII
5 mTECIII/IV
6 TECneuro
7 TECmyo
8 DN Thymocyte
9 SP/DP Thymocyte
10 Macrophage
11 Dendritic
12 Fibroblast 1
13 Fibroblast 1 Cycling
14 Fibroblast 2
15 VSMC
16 Vascular Endothelial
17 Lymphatic Endothelial
18 Neuronal
Bronchus
1 Epithelial Basal
2 Club
3 Goblet
4 Submucosal Gland
5 Secretory
6 Multiciliated
7 Neuroendocrine
8 Bud Tip Progenitor
9 Myoepithelial
10 T Cell
11 B Cell
12 Macrophage
13 Fibroblast
14 VSMC
15 Smooth Muscle
16 Cartilage
17 Vascular Endothelial
18 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 Vascular Endothelial
15 Lymphatic Endothelial
16 Neuronal
Esophagus
1 Epithelial Basal
2 Suprabasal
3 Multiciliated
4 Gastric Epithelial
5 T Cell
6 Macrophage
7 Fibroblast
8 VSMC
9 Smooth Muscle
10 Vascular Endothelial
11 Neuronal
D
Epithelial
cTEC
mcTEC
mTECI
mTECII
mTECIII/IV
TECneuro
TECmyo
Basal
Club
Goblet
Submuc. Gland
Secretory
Multiciliated
BTP Adjacent
BTP
Basal
Suprabasal
Gastric
FOXN1 PSMB11 DLK2 CCL19 AIRE KRT1 POU2F3 NEUROD1 MYOD1 IL33 SCGB1A1 MUC5B LTF SCGB3A2 PIFO AGER SFTPC KRT15 KRT13 TFF1
Avg. Exp.
% Exp.
Immune
DN Thymocyte
SP/DP Thymocyte
T Cell
B Cell
Macrophage
Dendritic
Mesenchymal
Fibroblast 1
Fibroblast 1 Cycling
Fibroblast 2
VSMC
Smooth Muscle
Cartilage
Endo.
Vasc. Endothelial
Lymph. Endothelial
Neuro.
Neuronal
Neuroendocrine
TRDC PTPRC CD4 CD8A CD3D MS4A1 S100A9 LAMP3 FGFR1 ALDH1A2 MKI67 FBN1 PDGFRB ACTA2 DES CNMD AQP1 PECAM1 PROX1 NRXN1 CHGA
E
Thymus
Bronchus
Lung
Esophagus
Epithelial
Mesenchymal
Immune
Endothelial
Myoepithelial
Neuronal
cTEC
mcTEC
mTECI
mTECII
mTECIII/IV
TECneuro
TECmyo

**Figure 1. Cellular composition of human fetal anterior foregut-derived tissues.** (A) Schematic of ontogeny for thymus, bronchus, lung, esophagus and parathyroid, arising from the endodermal germ layer (top). Summary of gestational stage/age and tissue origin of samples. Colors denote thymus (orange), parathyroid (grey), bronchus (blue), lung (green) and esophagus (pink) (bottom). (B) Workflow of tissue processing for single cell transcriptomic profiling. (C) tSNE visualization of the cellular composition of human fetal thymus, bronchus, lung and esophagus colored by cell type. DN: double negative thymocyte, SP/DP: double positive/single positive thymocyte, VSMC: vascular smooth muscle cell. (D) Dot plot showing select, highly expressed cell type defining marker genes used for cell type annotation. Here and onwards, color represents average expression of marker genes in each cell type. Dot size represents percentage of cells in each cell type expressing the selected marker gene. BTP: bud tip progenitor, Submuc: submucosal, Vasc: vascular, Lymph: lymphatic, Endo: endo, Neuro: neuronal (E) tSNE visualization of integrated cellular composition of anterior foregut tissues colored by organ of origin (left); tissue type (middle); epithelial cell subpopulations of the thymus (right).

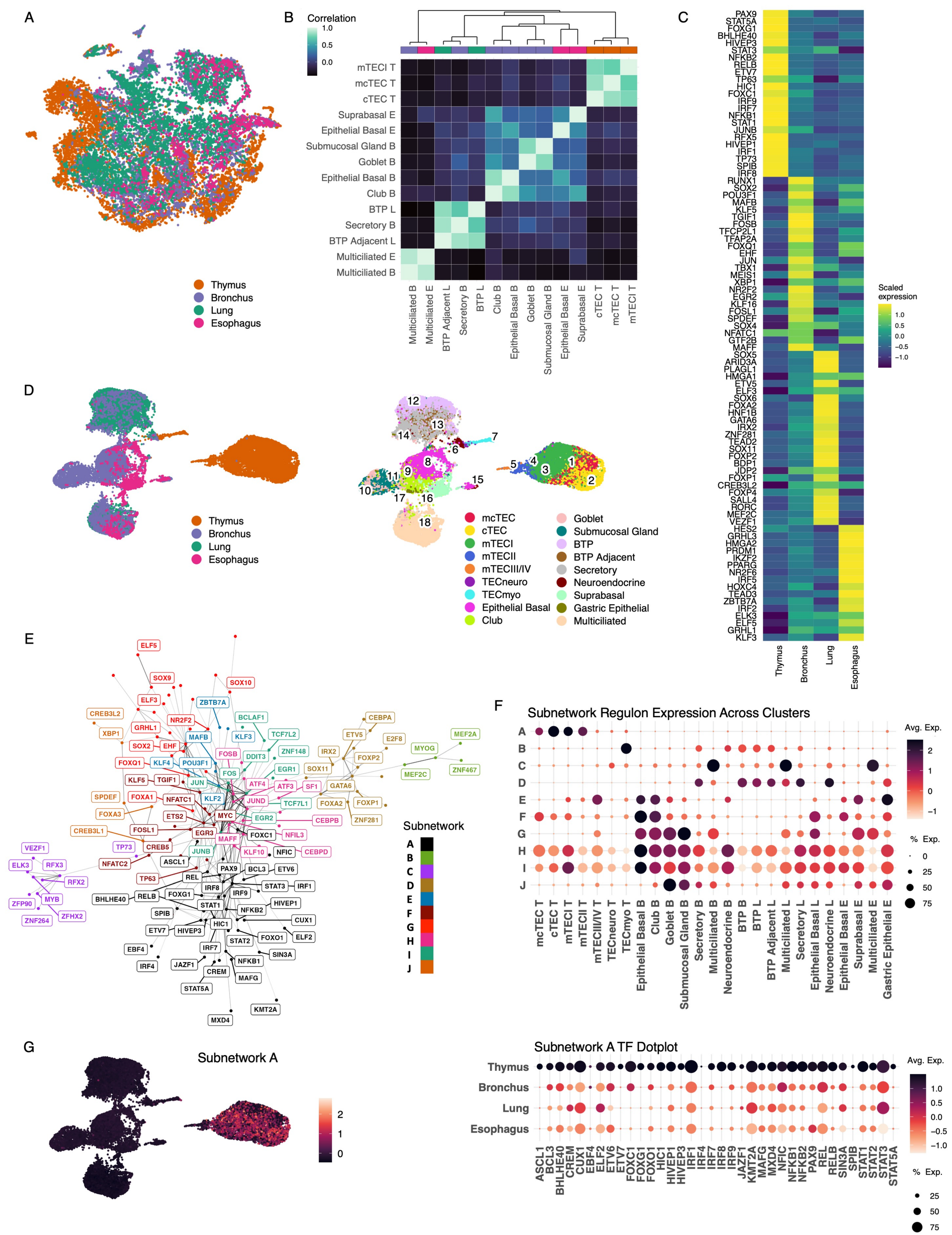

A
Thymus
Bronchus
Lung
Esophagus
B
Correlation
1.0
0.5
0.0
mTECI T
mcTEC T
cTEC T
Suprabasal E
Epithelial Basal E
Submucosal Gland B
Goblet B
Epithelial Basal B
Club B
BTP L
Secretory B
BTP Adjacent L
Multiciliated E
Multiciliated B
C
Scaled expression
1.0
0.5
0.0
−0.5
−1.0
Thymus
Bronchus
Lung
Esophagus
D
Thymus
Bronchus
Lung
Esophagus
mcTEC
cTEC
mTECI
mTECII
mTECIII/IV
TECneuro
TECmyo
Epithelial Basal
Club
Goblet
Submucosal Gland
BTP
BTP Adjacent
Secretory
Neuroendocrine
Suprabasal
Gastric Epithelial
Multiciliated
E
F
Subnetwork Regulon Expression Across Clusters
Avg. Exp.
% Exp.
Subnetwork
G
Subnetwork A
Subnetwork A TF Dotplot
Avg. Exp.
% Exp.

**Figure 2. Gene regulatory landscape of AFG-derived epithelial cells reveals unique inflammatory signature in the thymic epithelium.** (A) tSNE visualization of epithelial cells from all anterior foregut tissues colored by tissue of origin. (B) Correlation matrix plot comparing anterior foregut tissues. T= Thymus, B=Bronchus, L=Lung, B=Bronchus, BTP= Bud Tip Progenitor Cells. (C) Heatmap of differentially expressed SCENIC regulons across AFG-derived tissues. (D) SCENIC UMAP of all AFG-derived epithelial cells colored by tissue of origin (left), and cell type (right). (E) Network visualization of prominent epithelial AFG transcription factors. Transcription factor subnetworks are defined by Leiden algorithm-based subnetwork identification. (F) Dot plot representation of subnetwork-specific expression across AFG- derived epithelial cell subsets. (G) UMAP of average subnetwork A expression (left) and corresponding dot plot of all subnetwork A genes across epithelial cells from AFG derived organs (right).

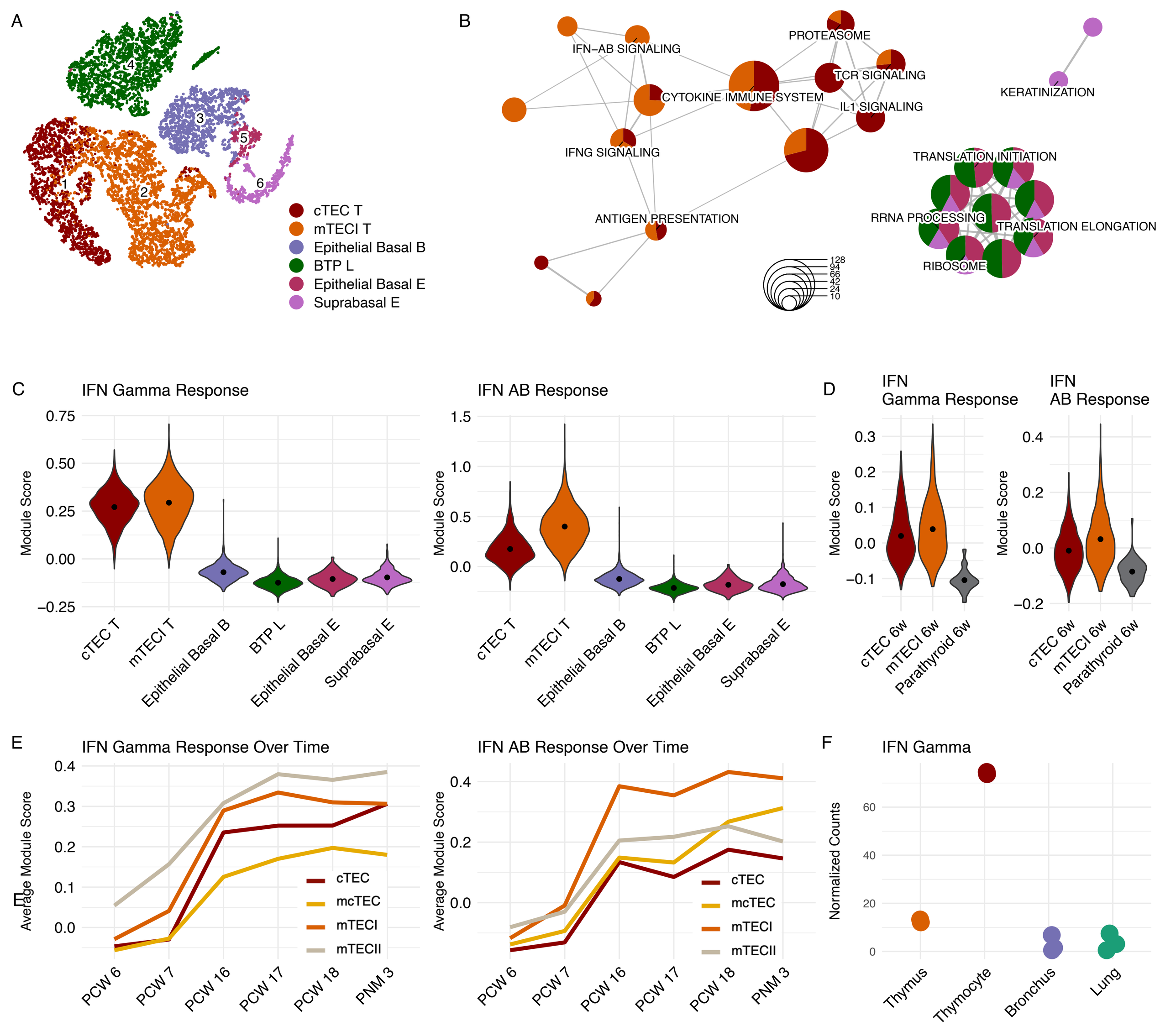

A
1
2
3
4
5
6
cTEC T
mTECI T
Epithelial Basal B
BTP L
Epithelial Basal E
Suprabasal E
B
IFN–AB SIGNALING
PROTEASOME
TCR SIGNALING
CYTOKINE IMMUNE SYSTEM
IL1 SIGNALING
KERATINIZATION
IFNG SIGNALING
TRANSLATION INITIATION
RRNA PROCESSING
TRANSLATION ELONGATION
ANTIGEN PRESENTATION
RIBOSOME
128
94
66
42
24
10
C
IFN Gamma Response
IFN AB Response
Module Score
cTEC T
mTECI T
Epithelial Basal B
BTP L
Epithelial Basal E
Suprabasal E
D
IFN Gamma Response
IFN AB Response
Module Score
cTEC 6w
mTECI 6w
Parathyroid 6w
E
IFN Gamma Response Over Time
IFN AB Response Over Time
Average Module Score
cTEC
mcTEC
mTECI
mTECII
PCW 6
PCW 7
PCW 16
PCW 17
PCW 18
PNM 3
F
IFN Gamma
Normalized Counts
Thymus
Thymocyte
Bronchus
Lung

**Figure 3. Enrichment in Interferon type 1 and 2 signaling distinguish cortical and medullary TECs from other AFG-derived epithelial cell types.**
(A) tSNE visualization of select tissue-defining epithelial cell populations of thymus, bronchus, lung and esophagus. (B) Enrichment map of differentially expressed gene sets across select tissue-defining epithelial cell populations. Each node represents a specific gene set. Edges between nodes denote a significant number of shared genes between nodes. Node size reflects the number of genes within each gene set. Small isolated nodes were removed. (C) Violin plot representation of Reactome-curated IFN type I and II gene module expression in select AFG-derived epithelial cell populations. Dot within violin plot indicates mean module score. (D) Violin plot representation of Reactome-curated IFN type I and II gene module expression in embryonic (PCW 6) thymic and parathyroid epithelial cells. (E) Average module expression of Reactome-curated IFN type 1 and II across embryonic, fetal, and postnatal time points in select thymic epithelial subtypes. (F) Bulk RNAseq expression of IFN Gamma in thymocytes and select AFG organs.

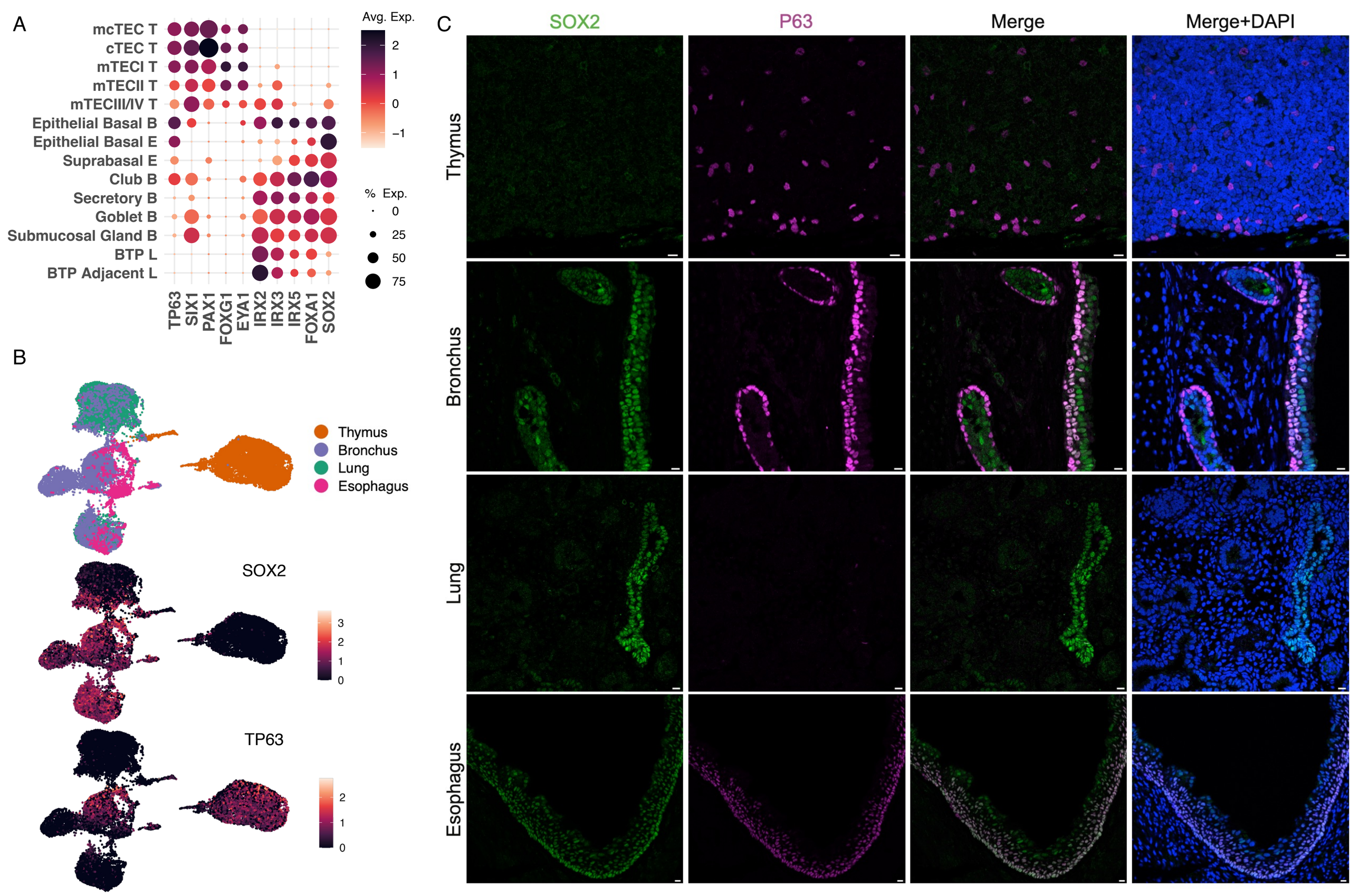

A
mcTEC T
cTEC T
mTECI T
mTECII T
mTECIII/IV T
Epithelial Basal B
Epithelial Basal E
Suprabasal E
Club B
Secretory B
Goblet B
Submucosal Gland B
BTP L
BTP Adjacent L
TP63
SIX1
PAX1
FOXG1
EYA1
IRX2
IRX3
IRX5
FOXA1
SOX2
Avg. Exp.
% Exp.
B
Thymus
Bronchus
Lung
Esophagus
SOX2
TP63
C
SOX2
P63
Merge
Merge+DAPI
Thymus
Bronchus
Lung
Esophagus

**Figure 4. Unique *SOX2*/*TP63* expression characterizes thymic epithelium from other AFG derived epithelia**

(A) Dotplot of curated PPEIII-defining transcription factors within select AFG-derived epithelial cell populations. (B) UMAP visualization based on SCENIC regulon expression in all AFG-derived epithelial cells colored by tissue of origin (top). *SOX2* (middle) and *TP63* (bottom) regulons are highlighted. (C) Representative immunohistochemistry-stained sections of human fetal thymus, bronchus, lung, and esophagus demonstrating SOX2 (green) and P63 (magenta) protein expression. Counterstain with DAPI (blue). Scale bars: 50um.

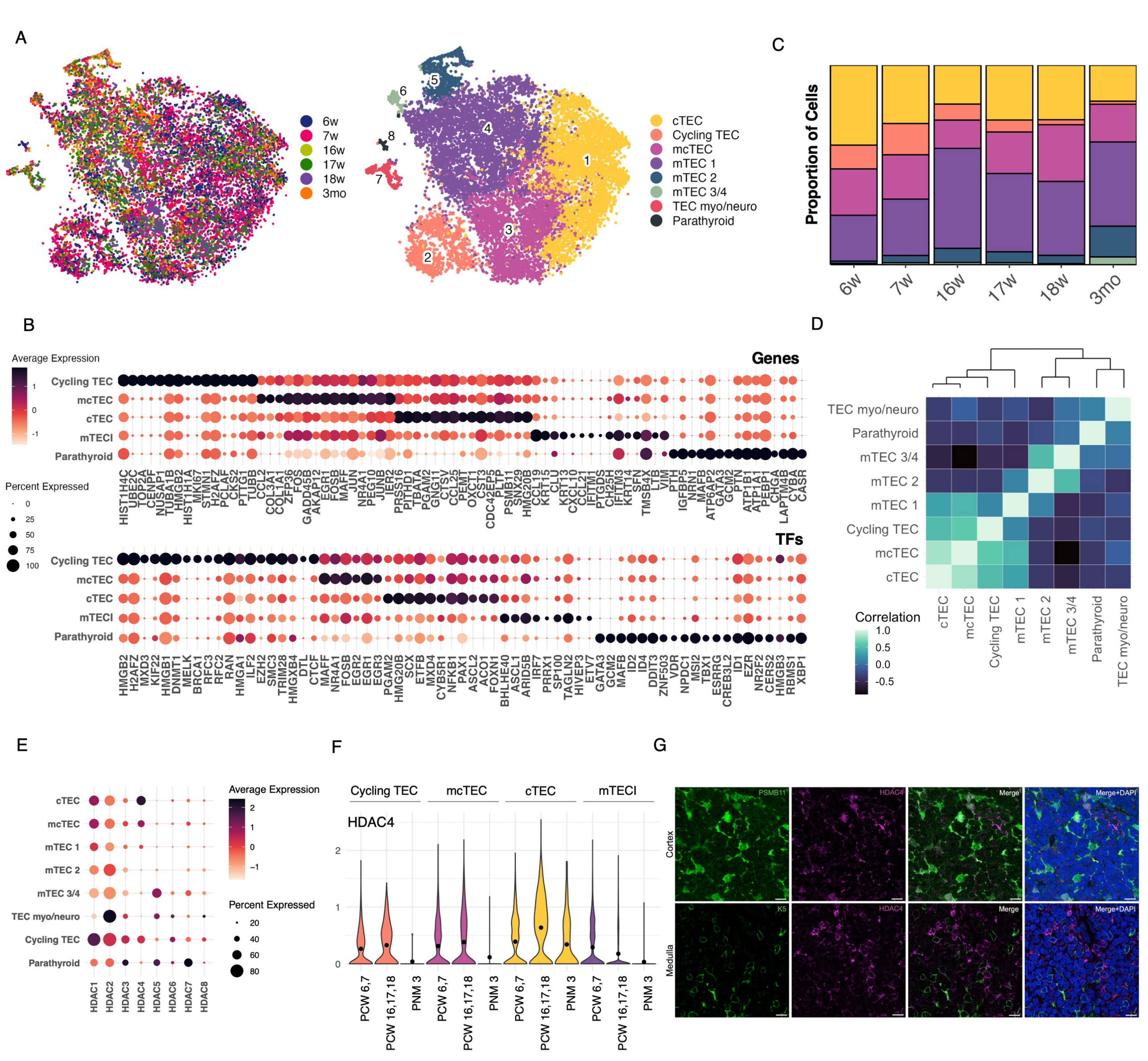

A
6w
7w
16w
17w
18w
3mo
cTEC
Cycling TEC
mcTEC
mTEC 1
mTEC 2
mTEC 3/4
TEC myo/neuro
Parathyroid
B
Average Expression
Genes
TFs
Percent Expressed
C
Proportion of Cells
D
Correlation
E
F
G
HDAC4
PCW 6,7
PCW 16,17,18
PNM 3
Cortex
Medulla
PSMB11
K5
Merge
Merge+DAPI

**Figure 5. Cellular composition of the human thymus during gestation and early prenatal stages.**
**(A)** tSNE visualization of thymic epithelial cells across all sampled developmental time points, colored by gestational/postnatal age (left) or cell type (right). (B) Dot plot representation of differentially expressed genes (top) and transcription factors (bottom) in select thymic and parathyroid epithelial cell types. Color representative of average expression level. Dot size representative of percentage of cells in cell type expressing selected marker gene. (C) Relative frequency of select thymic epithelial cell types across increasing developmental time points. (D) Correlation matrix plot comparing thymic epithelial subpopulations. (E) Dot plot representation of HDAC gene expression across 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) Violin plot representation of HDAC4 expression in key thymic epithelial cell types at different time points. Dot within violin plot represents mean module score. (G) Representative immunohistochemistry image of human fetal thymus section stained for PSMB11 (green) and HDAC4 (magenta). Counterstain with DAPI (Blue). C: Cortex; M: Medulla. Scale bars: 50um.

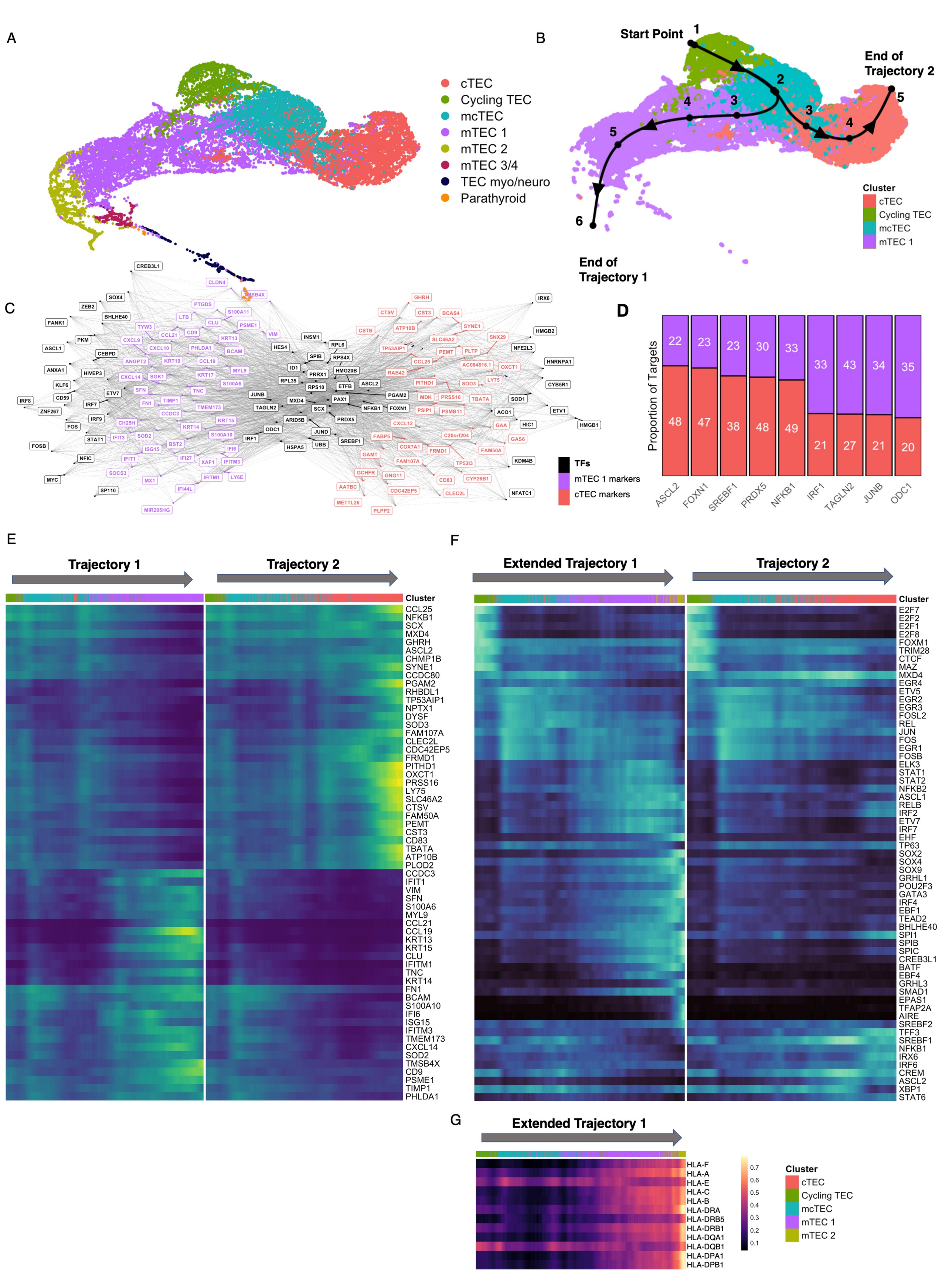

A
cTEC
Cycling TEC
mcTEC
mTEC 1
mTEC 2
mTEC 3/4
TEC myo/neuro
Parathyroid
B
Start Point
End of Trajectory 2
End of Trajectory 1
Cluster
cTEC
Cycling TEC
mcTEC
mTEC 1
C
TFs
mTEC 1 markers
cTEC markers
D
Proportion of Targets
ASCL2
FOXN1
SREBF1
PRDX5
NFKB1
IRF1
TAGLN2
JUNB
ODC1
E
Trajectory 1
Trajectory 2
Cluster
F
Extended Trajectory 1
Trajectory 2
Cluster
G
Extended Trajectory 1
Cluster
cTEC
Cycling TEC
mcTEC
mTEC 1
mTEC 2

**Figure 6. Gene regulatory networks directing cortical and medullary lineage specification.**
(A) tSpace UMAP of all thymic epithelial cells across advancing gestational and postnatal ages. (B) Slingshot-identified cortical and medullary lineage trajectories plotted onto tSpace UMAP. mTEC2,mTEC 3/4, TEC myo/neuro, and parathyroid clusters are not included. Pseudotime intervals, identified by TradeSeq (knots) are represented as dots. (C) Gene regulatory network visualization of interactions between transcription factors and cTEC/mTECI defining markers. (D) Barplot visualization of cluster specific bias in select central nodes from Figure 6C. Labels in each bar represent the number of identified transcription factor targets that intersect with the 50 defining markers for each cluster. (E) Heatmaps of the rolling average of top differentially expressed genes between knots 2-5 along trajectories 1 and 2. Rolling average and quantile scaling separate for each trajectory. (F) Heatmap representation of the rolling average of SCENIC regulon AUC scores along extended trajectories 1 and 2 for select regulons. Rolling average and quantile scaling separate for each trajectory. (G) Heatmaps of the rolling average of select HLA genes along trajectories 1 and 2.

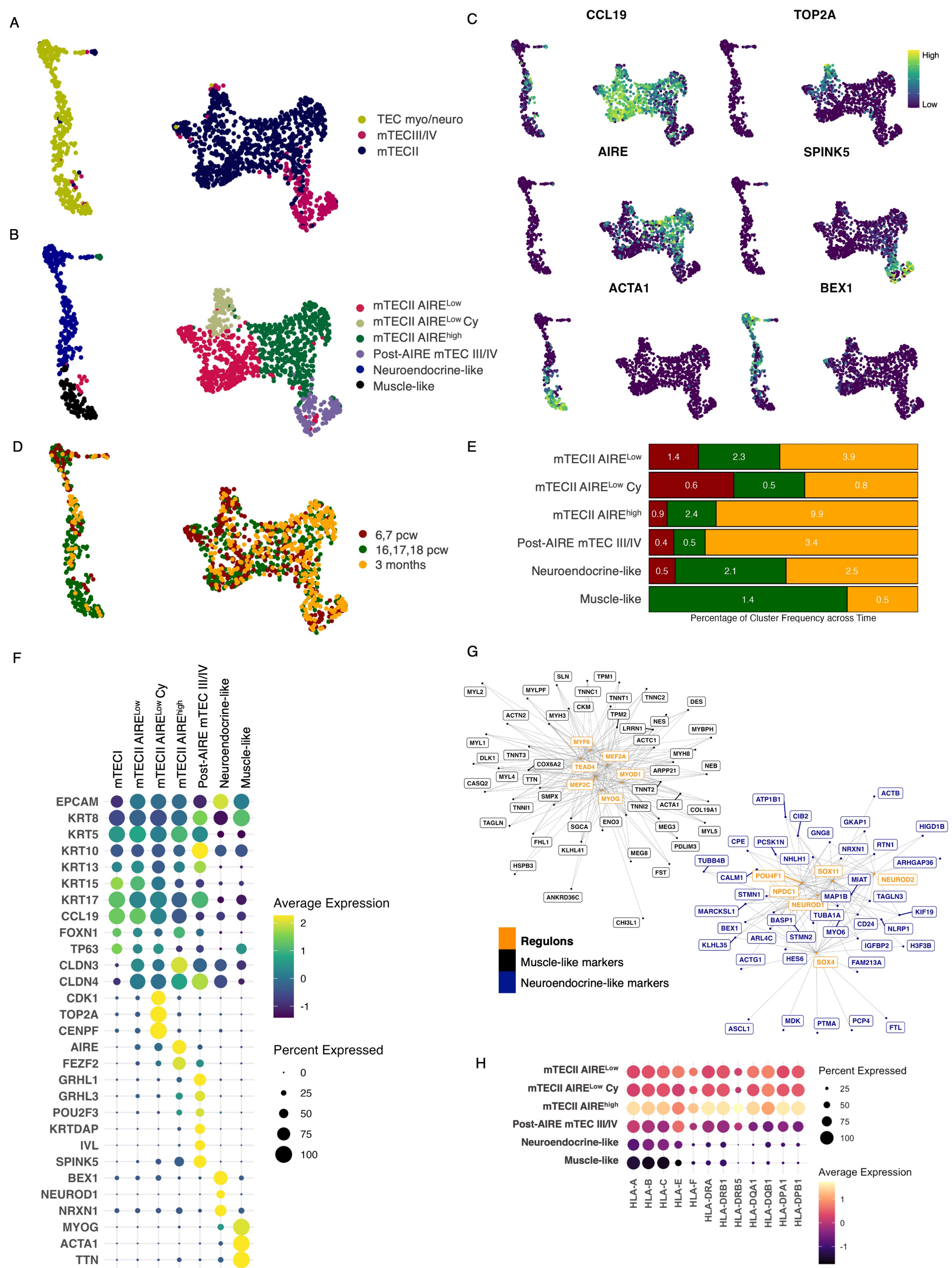
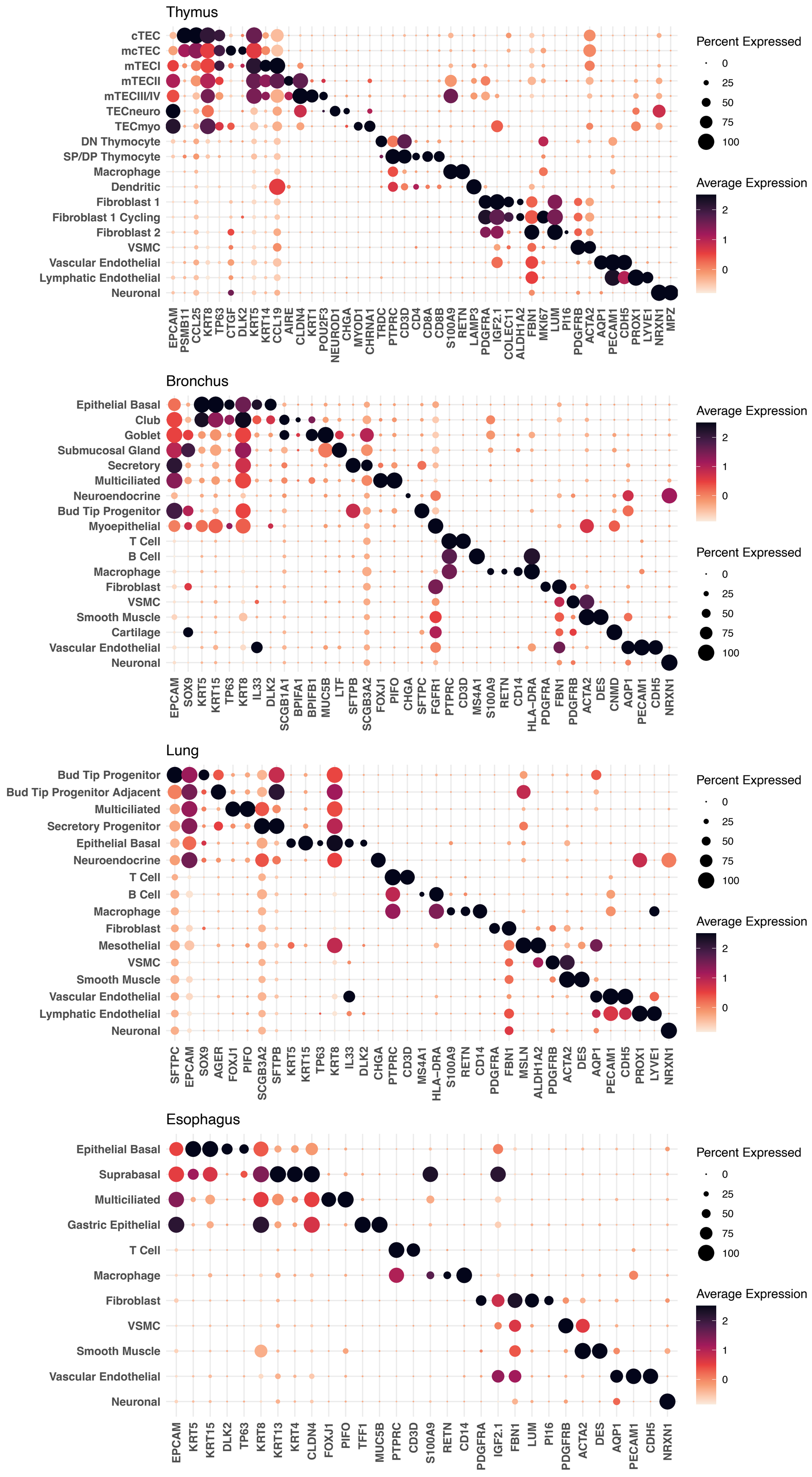
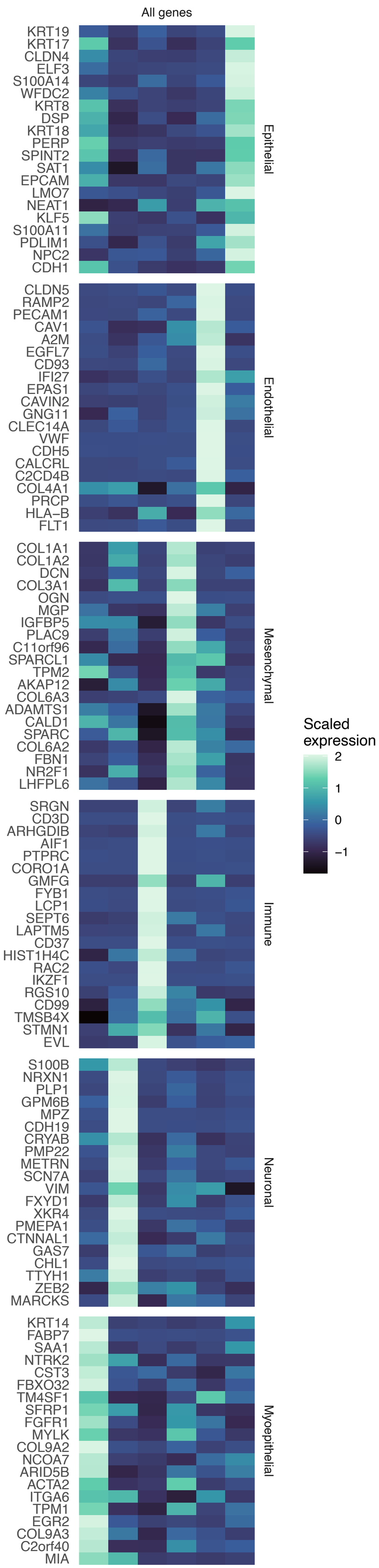

A
TEC myo/neuro
mTECIII/IV
mTECII
B
mTECII AIRELow
mTECII AIRELow Cy
mTECII AIREhigh
Post-AIRE mTEC III/IV
Neuroendocrine-like
Muscle-like
C
CCL19
TOP2A
High
Low
AIRE
SPINK5
ACTA1
BEX1
D
6,7 pcw
16,17,18 pcw
3 months
E
mTECII AIRELow
mTECII AIRELow Cy
mTECII AIREhigh
Post-AIRE mTEC III/IV
Neuroendocrine-like
Muscle-like
1.4 2.3 3.9
0.6 0.5 0.8
0.9 2.4 9.9
0.4 0.5 3.4
0.5 2.1 2.5
1.4 0.5
Percentage of Cluster Frequency across Time
F
mTECI
mTECII AIRELow
mTECII AIRELow Cy
mTECII AIREhigh
Post-AIRE mTEC III/IV
Neuroendocrine-like
Muscle-like
EPCAM
KRT8
KRT5
KRT10
KRT13
KRT15
KRT17
CCL19
FOXN1
TP63
CLDN3
CLDN4
CDK1
TOP2A
CENPF
AIRE
FEZF2
GRHL1
GRHL3
POU2F3
KRTDAP
IVL
SPINK5
BEX1
NEUROD1
NRXN1
MYOG
ACTA1
TTN
Average Expression
2
1
0
-1
Percent Expressed
0
25
50
75
100
G
Regulons
Muscle-like markers
Neuroendocrine-like markers
MYF6
MEF2A
TEAD4
MYOD1
MEF2C
MYOG
SOX11
NPDC1
NEUROD1
POU4F1
SOX4
H
mTECII AIRELow
mTECII AIRELow Cy
mTECII AIREhigh
Post-AIRE mTEC III/IV
Neuroendocrine-like
Muscle-like
HLA-A
HLA-B
HLA-C
HLA-E
HLA-F
HLA-DRA
HLA-DRB1
HLA-DRB5
HLA-DQA1
HLA-DQB1
HLA-DPA1
HLA-DPB1
Percent Expressed
25
50
75
100
Average Expression
1
0
-1

**Figure 7. Patterns of medullary thymic epithelial cell maturation and peripheral tissue antigen expression**

(A) UMAP visualization of mTECII, mTECIII/IV, and TECmyo/neuro across increasing gestational and postnatal ages. Cells colored according to cell type. (B) Unsupervised clustering of UMAP from Figure 7A. (C) UMAP visualization of expression of cluster-defining marker genes. (D) UMAP from (7A) colored according to developmental age. (E) Barplots depicting percent of frequency of clusters across developmental time. (F) Dot plot representation of select transcription factors and key marker expression across select thymic cell types. Color representative of average expression level. Dot size representative of percentage of cells in cell type expressing selected marker gene. (G) SCENIC-based network visualization of interactions between transcription factors and target genes in TECmyo/neuro. (H) Dot plot representation of select HLA gene expression across mature medullary thymic epithelial cell types. Color representative of average expression level. Dot size representative of percentage of cells in cell type expressing selected marker gene.


Thymus
cTEC
mcTEC
mTECI
mTECII
mTECIII/IV
TECneuro
TECmyo
DN Thymocyte
SP/DP Thymocyte
Macrophage
Dendritic
Fibroblast 1
Fibroblast 1 Cycling
Fibroblast 2
VSMC
Vascular Endothelial
Lymphatic Endothelial
Neuronal
EPCAM PSMB11 CCL25 KRT8 TP63 CTGF DLK2 KRT5 KRT14 CCL19 AIRE CLDN4 KRT1 POU2F3 NEUROD1 CHGA MYOD1 CHRNA1 TRDC PTPRC CD3D CD4 CD8A CD8B S100A9 RETN LAMP3 PDGFRA IGF2.1 COLEC11 ALDH1A2 FBN1 MKI67 LUM PI16 PDGFRB ACTA2 AQP1 PECAM1 CDH5 PROX1 LYVE1 NRXN1 MPZ
Percent Expressed
0
25
50
75
100
Average Expression
2
1
0
Bronchus
Epithelial Basal
Club
Goblet
Submucosal Gland
Secretory
Multiciliated
Neuroendocrine
Bud Tip Progenitor
Myoepithelial
T Cell
B Cell
Macrophage
Fibroblast
VSMC
Smooth Muscle
Cartilage
Vascular Endothelial
Neuronal
EPCAM SOX9 KRT5 KRT15 TP63 KRT8 IL33 DLK2 SCGB1A1 BPIFA1 BPIFB1 MUC5B LTF SFTPB SCGB3A2 FOXJ1 PIFO CHGA SFTPC FGFR1 PTPRC CD3D MS4A1 S100A9 RETN CD14 HLA−DRA PDGFRA FBN1 PDGFRB ACTA2 DES CNMD AQP1 PECAM1 CDH5 NRXN1
Average Expression
2
1
0
Percent Expressed
0
25
50
75
100
Lung
Bud Tip Progenitor
Bud Tip Progenitor Adjacent
Multiciliated
Secretory Progenitor
Epithelial Basal
Neuroendocrine
T Cell
B Cell
Macrophage
Fibroblast
Mesothelial
VSMC
Smooth Muscle
Vascular Endothelial
Lymphatic Endothelial
Neuronal
SFTPC EPCAM SOX9 AGER FOXJ1 PIFO SCGB3A2 SFTPB KRT5 KRT15 TP63 KRT8 IL33 DLK2 CHGA PTPRC CD3D MS4A1 HLA−DRA S100A9 RETN CD14 PDGFRA FBN1 MSLN ALDH1A2 PDGFRB ACTA2 DES AQP1 PECAM1 CDH5 PROX1 LYVE1 NRXN1
Percent Expressed
0
25
50
75
100
Average Expression
2
1
0
Esophagus
Epithelial Basal
Suprabasal
Multiciliated
Gastric Epithelial
T Cell
Macrophage
Fibroblast
VSMC
Smooth Muscle
Vascular Endothelial
Neuronal
EPCAM KRT5 KRT15 DLK2 TP63 KRT8 KRT13 KRT4 CLDN4 FOXJ1 PIFO TFF1 MUC5B PTPRC CD3D S100A9 RETN CD14 PDGFRA IGF2.1 FBN1 LUM PI16 PDGFRB ACTA2 DES AQP1 PECAM1 CDH5 NRXN1
Percent Expressed
0
25
50
75
100
Average Expression
2
1
0

**Supplementary Figure S1A**
Dot plot showing marker gene expression for annotated cell types. DN: double negative T cells, SP/DP: double positive/single positive thymocyte, VSMCs: vascular smooth muscle cells.


All genes
KRT19
KRT17
CLDN4
ELF3
S100A14
WFDC2
KRT8
DSP
KRT18
PERP
SPINT2
SAT1
EPCAM
LMO7
NEAT1
KLF5
S100A11
PDLIM1
NPC2
CDH1
Epithelial
CLDN5
RAMP2
PECAM1
CAV1
A2M
EGFL7
CD93
IFI27
EPAS1
CAVIN2
GNG11
CLEC14A
VWF
CDH5
CALCRL
C2CD4B
COL4A1
PRCP
HLA−B
FLT1
Endothelial
COL1A1
COL1A2
DCN
COL3A1
OGN
MGP
IGFBP5
PLAC9
C11orf96
SPARCL1
TPM2
AKAP12
COL6A3
ADAMTS1
CALD1
SPARC
COL6A2
FBN1
NR2F1
LHFPL6
Mesenchymal
SRGN
CD3D
ARHGDIB
AIF1
PTPRC
CORO1A
GMFG
FYB1
LCP1
SEPT6
LAPTM5
CD37
HIST1H4C
RAC2
IKZF1
RGS10
CD99
TMSB4X
STMN1
EVL
Immune
S100B
NRXN1
PLP1
GPM6B
MPZ
CDH19
CRYAB
PMP22
METRN
SCN7A
VIM
FXYD1
XKR4
PMEPA1
CTNNAL1
GAS7
CHL1
TTYH1
ZEB2
MARCKS
Neuronal
KRT14
FABP7
SAA1
NTRK2
CST3
FBXO32
TM4SF1
SFRP1
FGFR1
MYLK
COL9A2
NCOA7
ARID5B
ACTA2
ITGA6
TPM1
EGR2
COL9A3
C2orf40
MIA
Myoepithelial
Scaled expression
2
1
0
−1

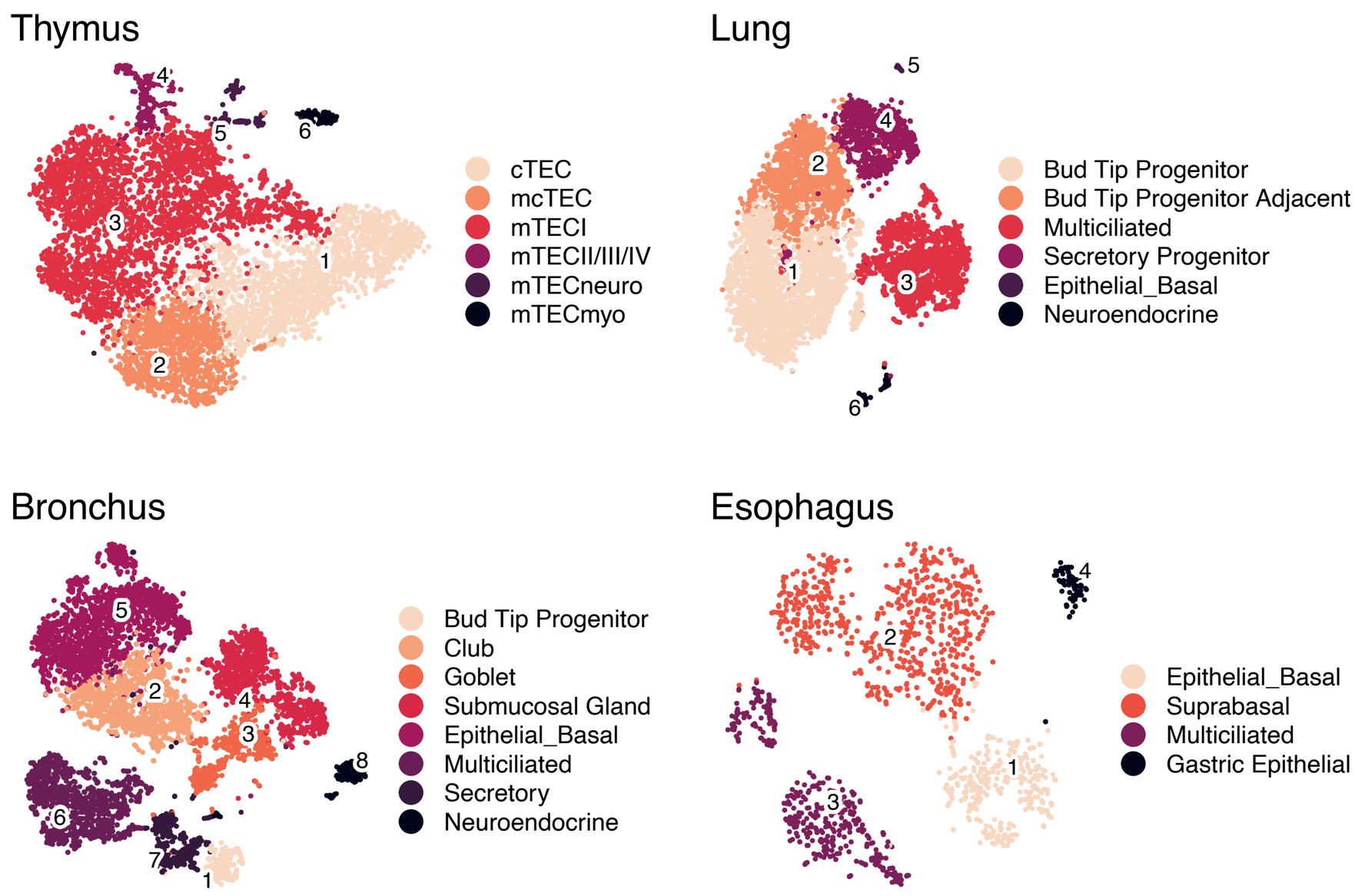
**Supplementary Figure S1B**
Heatmap of top differentially expressed genes for all major tissue types across cells from human fetal thymus, bronchus, lung and esophagus.

Supplementary Table S1. metadata for sequenced samples

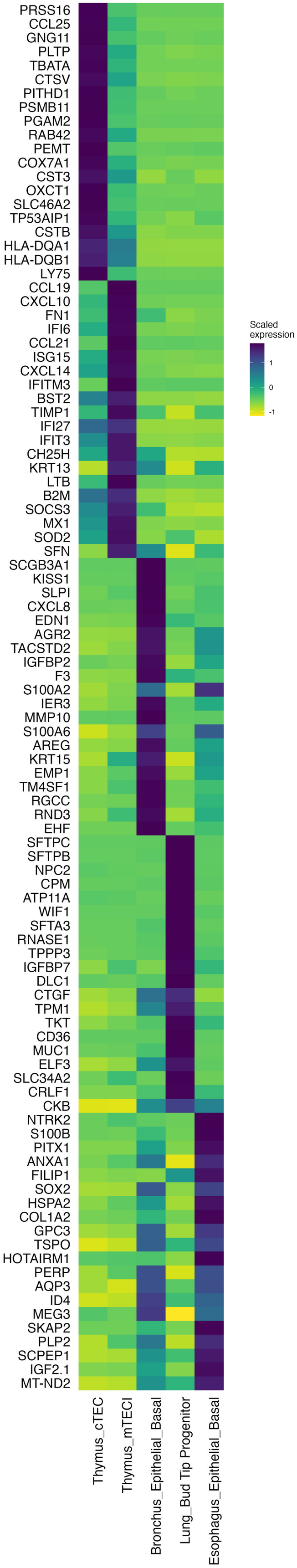
| Sample ID | Organ | Age | NumCells | MeanCountsPerCell | MeanGenesPerCell | MedianMitoPercentage | Gender | Tissue processing |
|---|---|---|---|---|---|---|---|---|
| C40_TH_TOT_2 | Thymus | Post Conception Week 7 | 7180 | 11139 | 3054.5 | 1.97 | Female | All cells |
| C40_TH_TOT_1 | Thymus | Post Conception Week 7 | 6841 | 11293.8 | 3098.1 | 1.97 | Female | All cells |
| C41_TH_TOT_2 | Thymus | Post Conception Week 6 | 5789 | 14190.5 | 3544.9 | 2.61 | Male | All cells |
| C41_TH_TOT_1 | Thymus | Post Conception Week 6 | 5620 | 14244.3 | 3547.6 | 2.52 | Male | All cells |
| KGW1_TH_EPCAM_45 | Thymus | Post Conception Week 16 | 3046 | 21747 | 4135 | 7.29 | Male | EPCAM positive selection, CD45 depletion |
| F83_TH_EPCAM | Thymus | Post Conception Week 17 | 3678 | 9575.2 | 2634.2 | 2.55 | Female | EPCAM FACS sorted |
| KGW2_TH_EPCAM_45 | Thymus | Post Conception Week 17 | 7634 | 11754 | 2993 | 6.81 | Female | EPCAM positive selection, CD45 depletion |
| KGW3_TH_EPCAM_45 | Thymus | Post Conception Week 18 | 8648 | 9349 | 2721 | 5.73 | Male | EPCAM positive selection, CD45 depletion |
| T07_TH_EPCAM_S3 | Thymus | 3 Months Postnatal | 3149 | 11355.6 | 2819.6 | 3.86 | Male | EPCAM FACS sorted |
| KGW4_B_EPCAM | Bronchus | Post Conception Week 16 | 5920 | 15651 | 3525 | 7.69 | Male | EPCAM positive selection |
| KGW5_B_EPCAM | Bronchus | Post Conception Week 17 | 4376 | 16449 | 3188 | 8.56 | Female | EPCAM positive selection |
| KGW6_B_EPCAM | Bronchus | Post Conception Week 18 | 7414 | 15387 | 2979 | 9.43 | Male | EPCAM positive selection |
| KGW7_L_EPCAM | Lung | Post Conception Week 16 | 3726 | 20366 | 3862 | 7.95 | Male | EPCAM positive selection |
| KGW8_L_EPCAM | Lung | Post Conception Week 17 | 7075 | 11074 | 2810 | 8.11 | Female | EPCAM positive selection |
| KGW9_L_EPCAM | Lung | Post Conception Week 18 | 7208 | 12965 | 3254 | 7.12 | Male | EPCAM positive selection |
| KGW10_E_EPCAM | Esophagus | Post Conception Week 16 | 4603 | 16004 | 3513 | 7 | Male | EPCAM positive selection |
| KGW11_E_EPCAM | Esophagus | Post Conception Week 17 | 6464 | 11431 | 2723 | 8.48 | Female | EPCAM positive selection |

**Table S1. Table_S1.xlsx (separate file)**
**Excel file containing metadata for sequenced samples.**

[Sample ID] TH: Thymus, B: Bronchus, L: Lung, E: Esophagus


Thymus
4
5
6
3
1
2
cTEC
mcTEC
mTECI
mTECII/III/IV
mTECneuro
mTECmyo
Lung
5
4
2
1
3
6
Bud Tip Progenitor
Bud Tip Progenitor Adjacent
Multiciliated
Secretory Progenitor
Epithelial_Basal
Neuroendocrine
Bronchus
5
2
4
3
8
6
7
1
Bud Tip Progenitor
Club
Goblet
Submucosal Gland
Epithelial_Basal
Multiciliated
Secretory
Neuroendocrine
Esophagus
4
2
1
3
Epithelial_Basal
Suprabasal
Multiciliated
Gastric Epithelial

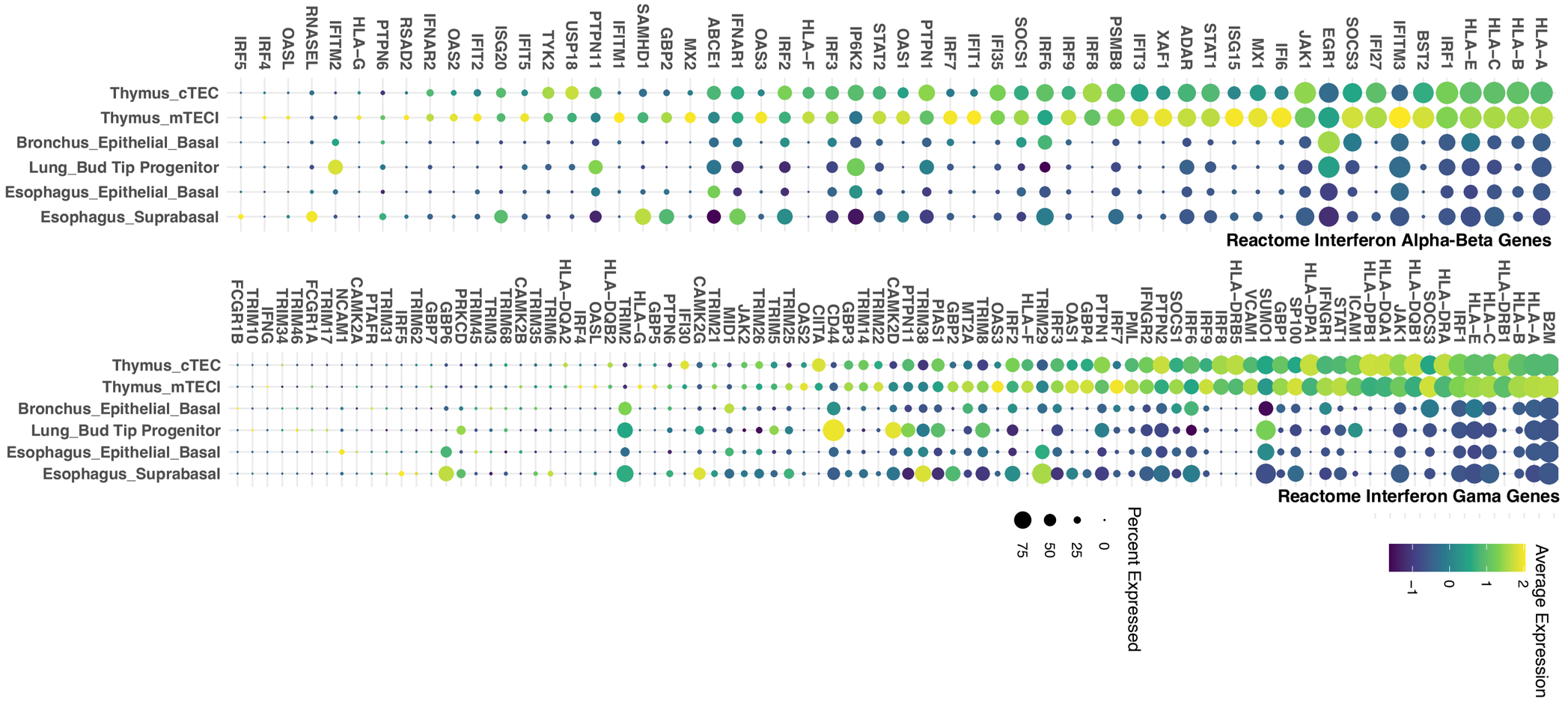
**Supplementary Figure S2A.**
tSNE visualization of epithelial cells colored by cell type from individual anterior foregut tissues, and tSNE visualization of epithelial cells from all anterior foregut tissues colored by tissue of origin (right panel).


PRSS16
CCL25
GNG11
PLTP
TBATA
CTSV
PITHD1
PSMB11
PGAM2
RAB42
PEMT
COX7A1
CST3
OXCT1
SLC46A2
TP53AIP1
CSTB
HLA-DQA1
HLA-DQB1
LY75
CCL19
CXCL10
FN1
IFI6
CCL21
ISG15
CXCL14
IFITM3
BST2
TIMP1
IFI27
IFIT3
CH25H
KRT13
LTB
B2M
SOCS3
MX1
SOD2
SFN
SCGB3A1
KISS1
SLPI
CXCL8
EDN1
AGR2
TACSTD2
IGFBP2
F3
S100A2
IER3
MMP10
S100A6
AREG
KRT15
EMP1
TM4SF1
RGCC
RND3
EHF
SFTPC
SFTPB
NPC2
CPM
ATP11A
WIF1
SFTA3
RNASE1
TPPP3
IGFBP7
DLC1
CTGF
TPM1
TKT
CD36
MUC1
ELF3
SLC34A2
CRLF1
CKB
NTRK2
S100B
PITX1
ANXA1
FILIP1
SOX2
HSPA2
COL1A2
GPC3
TSPO
HOTAIRM1
PERP
AQP3
ID4
MEG3
SKAP2
PLP2
SCPEP1
IGF2.1
MT-ND2
Thymus_cTEC
Thymus_mTECI
Bronchus_Epithelial_Basal
Lung_Bud Tip Progenitor
Esophagus_Epithelial_Basal
Scaled expression
1
0
-1

**Supplementary Figure S2B.**
Heatmap of differentially expressed genes across key AFG-derived epithelial cell subsets.

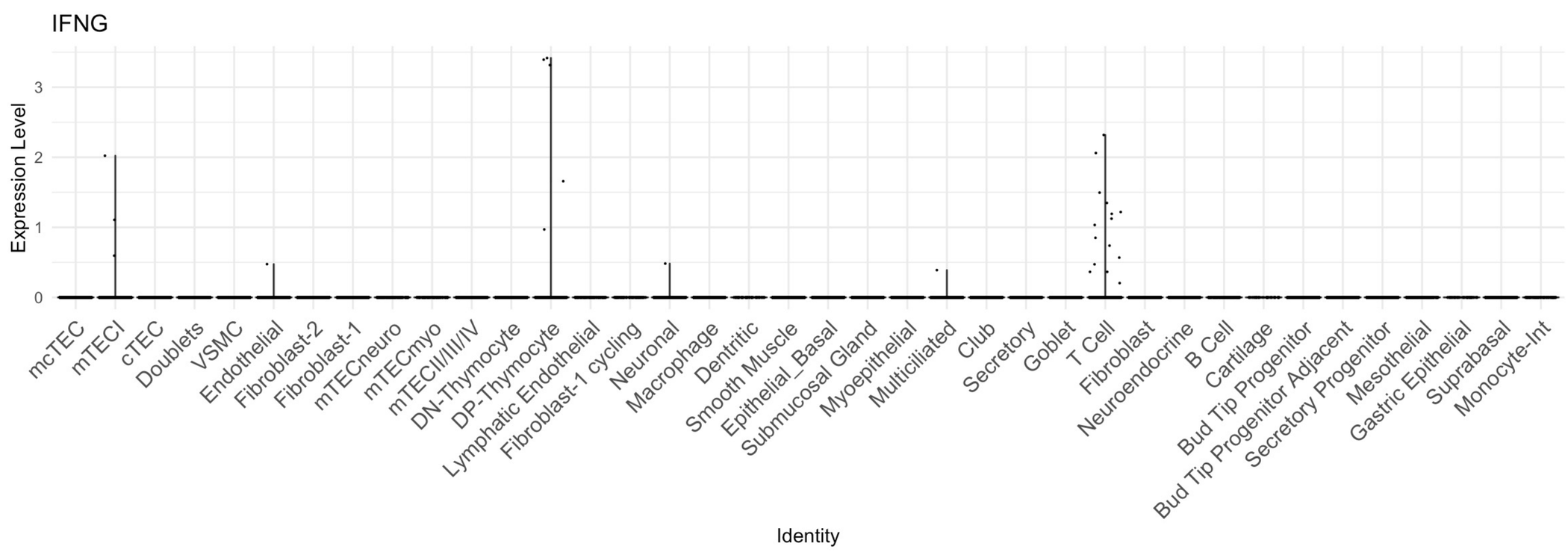
REACTOME_CELL_CELL_JUNCTION_ORGANIZATION
REACTOME_O_LINKED_GLYCOSYLATION_OF_MUCINS
REACTOME_KERATINIZATION
REACTOME_FORMATION_OF_THE_CORNIFIED_ENVELOPE
REACTOME_POST_TRANSLATIONAL_MODIFICATION_SYNTHESIS_OF_GPI_ANCHORED_PROTEINS
REACTOME_THE_CITRIC_ACID_TCA_CYCLE_AND_RESPIRATORY_ELECTRON_TRANSPORT
KEGG_OXIDATIVE_PHOSPHORYLATION
REACTOME_RESPIRATORY_ELECTRON_TRANSPORT
REACTOME_RESPIRATORY_ELECTRON_TRANSPORT_ATP_SYNTHESIS_BY_CHEMIOSMOTIC_COUPLING_AND_HEAT_PRODUCTION_BY_UNCOUPLING_PROTEINS
REACTOME_COMPLEX_I_BIOGENESIS
REACTOME_RECRUITMENT_OF_NUMA_TO_MITOTIC_CENTROSOMES
REACTOME_DISEASES_OF_METABOLISM
REACTOME_INFLUENZA_INFECTION
REACTOME_CELLULAR_RESPONSE_TO_STARVATION
REACTOME_RRNA_PROCESSING
REACTOME_TIGHT_JUNCTION_INTERACTIONS
REACTOME_TRANSLATION
REACTOME_NONSENSE_MEDIATED_DECAY_NMD
REACTOME_EUKARYOTIC_TRANSLATION_INITIATION
REACTOME_RESPONSE_OF_EIF2AK4_GCN2_TO_AMINO_ACID_DEFICIENCY
REACTOME_ACTIVATION_OF_THE_MRNA_UPON_BINDING_OF_THE_CAP_BINDING_COMPLEX_AND_EIFS_AND_SUBSEQUENT_BINDING_TO_43S
REACTOME_SELENOAMINO_ACID_METABOLISM
KEGG_RIBOSOME
REACTOME_EUKARYOTIC_TRANSLATION_ELONGATION
REACTOME_SRP_DEPENDENT_COTRANSLATIONAL_PROTEIN_TARGETING_TO_MEMBRANE
REACTOME_SURFACTANT_METABOLISM
REACTOME_SIGNALING_BY_RECEPTOR_TYROSINE_KINASES
REACTOME_SIGNALING_BY_NTRKS
REACTOME_CELLULAR_SENESCENCE
KEGG_PATHOGENIC_ESCHERICHIA_COLI_INFECTION
REACTOME_CARDIAC_CONDUCTION
REACTOME_NUCLEAR_EVENTS_KINASE_AND_TRANSCRIPTION_FACTOR_ACTIVATION
KEGG_HYPERTROPHIC_CARDIOMYOPATHY_HCM
KEGG_DILATED_CARDIOMYOPATHY
REACTOME_NGF_STIMULATED_TRANSCRIPTION
REACTOME_ANTIVIRAL_MECHANISM_BY_IFN_STIMULATED_GENES
REACTOME_ANTIGEN_PRESENTATION_FOLDING_ASSEMBLY_AND_PEPTIDE_LOADING_OF_CLASS_I_MHC
REACTOME_CHEMOKINE_RECEPTORS_BIND_CHEMOKINES
REACTOME_INTERFERON_ALPHA_BETA_SIGNALING
REACTOME_INTERFERON_SIGNALING
REACTOME_INTERFERON_GAMMA_SIGNALING
KEGG_ANTIGEN_PROCESSING_AND_PRESENTATION
REACTOME_TCR_SIGNALING
REACTOME_CROSS_PRESENTATION_OF_SOLUBLE_EXOGENOUS_ANTIGENS_ENDOSOMES
KEGG_PROTEASOME
KEGG_NATURAL_KILLER_CELL_MEDIATED_CYTOTOXICITY
KEGG_LEISHMANIA_INFECTION
REACTOME_COSTIMULATION_BY_THE_CD28_FAMILY
KEGG_TYPE_I_DIABETES_MELLITUS
KEGG_GRAFT_VERSUS_HOST_DISEASE
KEGG_ALLOGRAFT_REJECTION
KEGG_AUTOIMMUNE_THYROID_DISEASE
KEGG_SYSTEMIC_LUPUS_ERYTHEMATOSUS
KEGG_INTESTINAL_IMMUNE_NETWORK_FOR_IGA_PRODUCTION
KEGG_ASTHMA
REACTOME_GENERATION_OF_SECOND_MESSENGER_MOLECULES
cTEC T
mTECI T
Epithelial Basal B
BTP L
Epithelial Basal E
Suprabasal E
−log10(p.adjust)
2
4
6
8
enrichmentScore
0.75
0.50
0.25
0.00

**Supplementary Figure S3A.**
Dot plots of significant differential gene sets across select anterior foregut epithelial cell types.

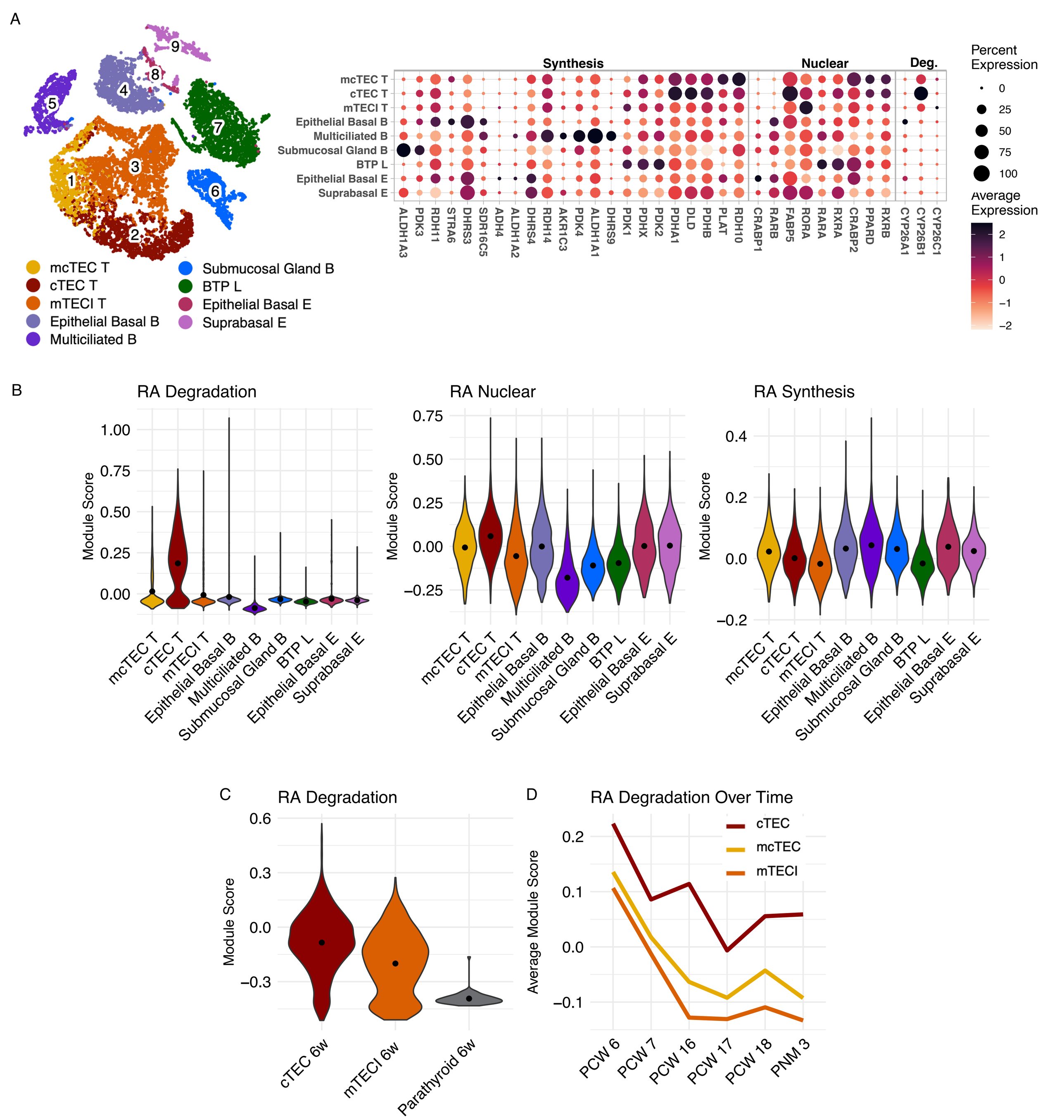

HLA-A
HLA-B
HLA-C
HLA-E
IRF1
BST2
IFITM3
IFI27
SOCS3
EGR1
JAK1
IFI6
MX1
ISG15
STAT1
ADAR
XAF1
IFIT3
PSMB8
IRF8
IRF9
IRF6
SOCS1
IFI35
IFIT1
IRF7
PTPN1
OAS1
STAT2
IP6K2
IRF3
HLA-F
IRF2
OAS3
IFNAR1
ABCE1
MX2
GBP2
SAMHD1
IFITM1
PTPN11
USP18
TYK2
IFIT5
ISG20
IFIT2
OAS2
IFNAR2
RSAD2
PTPN6
HLA-G
IFITM2
RNASEL
OASL
IRF4
IRF5
Thymus_cTEC
Thymus_mTECI
Bronchus_Epithelial_Basal
Lung_Bud Tip Progenitor
Esophagus_Epithelial_Basal
Esophagus_Suprabasal
Reactome Interferon Alpha-Beta 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
Reactome Interferon Gama Genes
Average Expression
2
1
0
−1
Percent Expressed
0
25
50
75

**Supplementary Figure S3B.**
Dot plot representations of Reactome-curated 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.

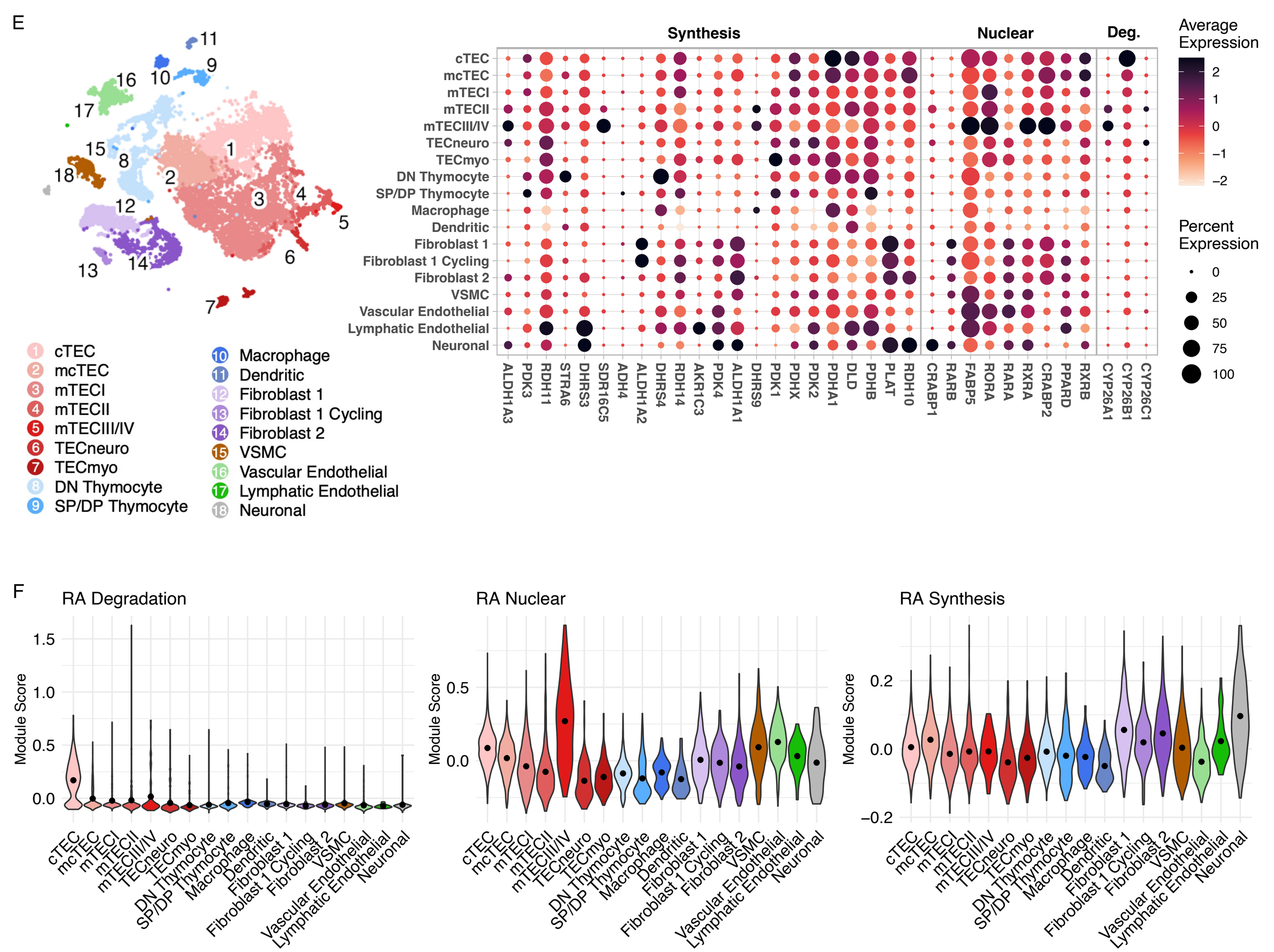
**Expression of *IFNA1, IFNA2, IFNB1* not detected**



**Supplementary Figure S3C.**
Violin plot representation of select interferon relevant gene expression across all AFG cell types.


A
mcTEC T
cTEC T
mTECI T
Epithelial Basal B
Multiciliated B
Submucosal Gland B
BTP L
Epithelial Basal E
Suprabasal E
Synthesis
Nuclear
Deg.
ALDH1A3
PDK3
RDH11
STRA6
DHRS3
SDR16C5
ADH4
ALDH1A2
DHRS4
RDH14
AKR1C3
PDK4
ALDH1A1
DHRS9
PDK1
PDHX
PDK2
PDHA1
DLD
PDHB
PLAT
RDH10
CRABP1
RARB
FABP5
RORA
RARA
RXRA
CRABP2
PPARD
RXRB
CYP26A1
CYP26B1
CYP26C1
Percent Expression
Average Expression
B
RA Degradation
RA Nuclear
RA Synthesis
Module Score
C
RA Degradation
cTEC 6w
mTECI 6w
Parathyroid 6w
D
RA Degradation Over Time
Average Module Score
cTEC
mcTEC
mTECI
PCW 6
PCW 7
PCW 16
PCW 17
PCW 18
PNM 3

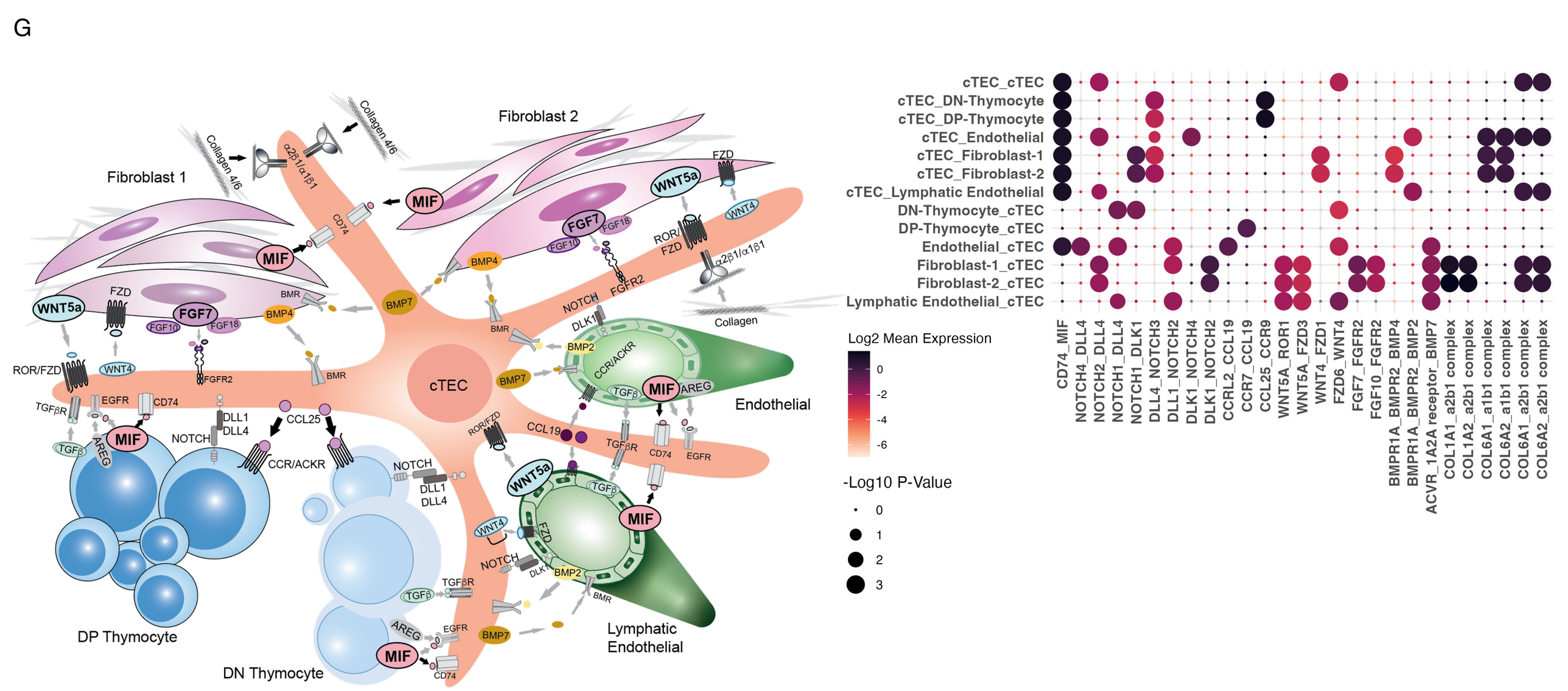
**Supplement S4 A-D**
(A) Left panel: tSNE visualization of select AFG-derived epithelial cells. Right panel: Dotplot of curated set of genes involved in retinoic acid synthesis, nuclear signaling, and degradation across select anterior foregut epithelial cell types. (B) Violin plot representation of composite expression of retinoic acid metabolism gene sets from Figure S4A across select AFG-derived epithelial cell types. (C) Violin plot representation of composite expression of retinoic acid metabolism gene sets across select thymic and parathyroid epithelial cells at pcw 6. (D) Average module expression of curated retinoic acid metabolism gene sets across embryonic, fetal, and postnatal time points in select thymic epithelial subtypes.


E
Synthesis
Nuclear
Deg.
Average Expression
Percent Expression
cTEC
mcTEC
mTECI
mTECII
mTECIII/IV
TECneuro
TECmyo
DN Thymocyte
SP/DP Thymocyte
Macrophage
Dendritic
Fibroblast 1
Fibroblast 1 Cycling
Fibroblast 2
VSMC
Vascular Endothelial
Lymphatic Endothelial
Neuronal
ALDH1A3 PDK3 RDH11 STRA6 DHRS3 SDR16C5 ADH4 ALDH1A2 DHRS4 RDH14 AKR1C3 PDK4 ALDH1A1 DHRS9 PDK1 PDHX PDK2 PDHA1 DLD PDHB PLAT RDH10 CRABP1 RARB FABP5 RORA RARA RXRA CRABP2 PPARD RXRB CYP26A1 CYP26B1 CYP26C1
F
RA Degradation
RA Nuclear
RA Synthesis
Module Score

**Supplement S4 E,F**

(E) Left panel: tSNE visualization of thymic cells. Right panel: Dotplot of curated set of genes involved in retinoic acid synthesis, nuclear signaling, and degradation across all thymic cell types. (B) Violin plot representation of composite expression of retinoic acid metabolism gene sets from Figure S4A across all thymic cell types.

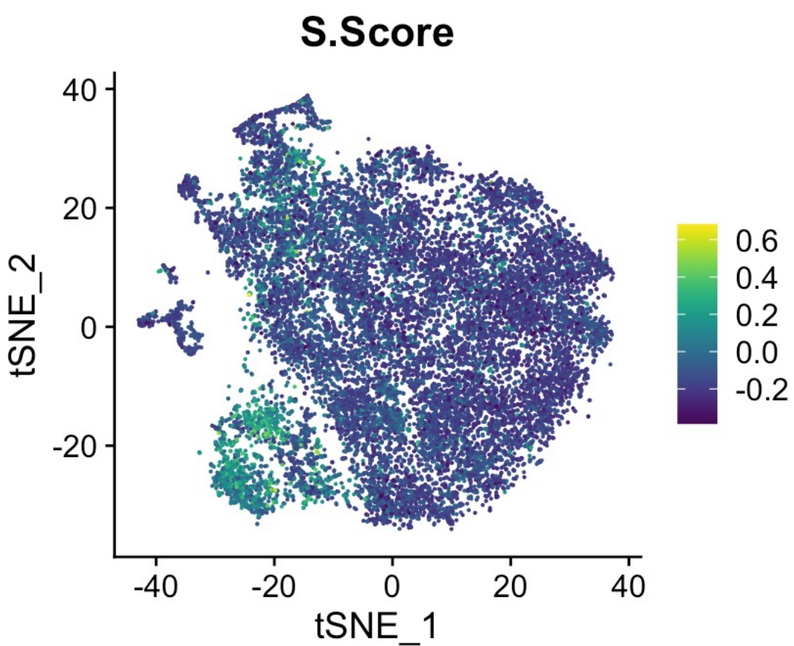
G

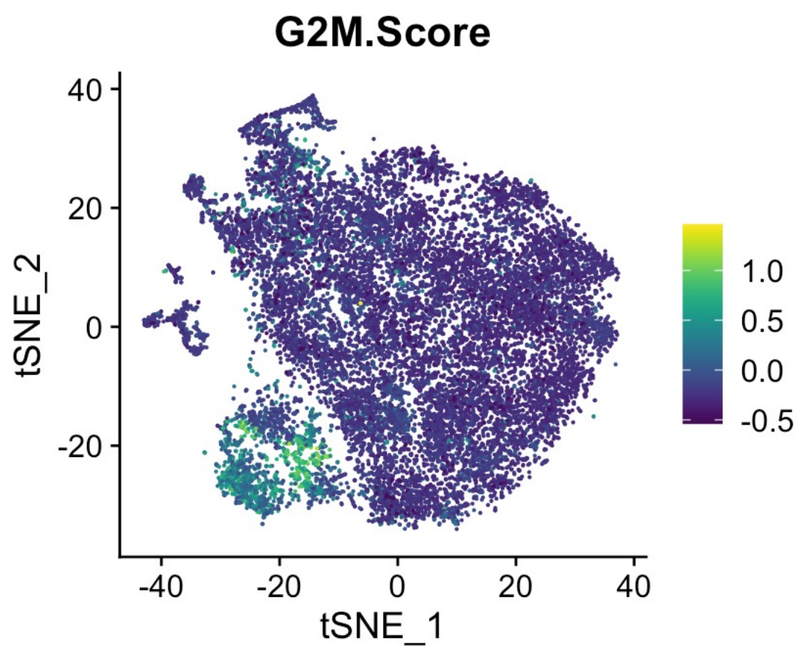

Fibroblast 1
Fibroblast 2
Collagen 4/6
Collagen 4/6
α2β1/α1β1
CD74
MIF
MIF
WNT5a
FZD
FGF7
FGF10
FGF18
BMP4
BMR
BMP7
ROR/FZD
WNT4
FGFR2
TGFβR
EGFR
CD74
TGFβ
AREG
MIF
NOTCH
DLL1
DLL4
CCL25
CCR/ACKR
cTEC
NOTCH
DLK1
BMP2
CCR/ACKR
TGFβ
MIF
AREG
Endothelial
CCL19
TGFβR
CD74
EGFR
WNT5a
TGFβ
MIF
WNT4
FZD
NOTCH
DLK1
BMP2
BMR
TGFβ
TGFβR
AREG
EGFR
BMP7
MIF
CD74
DP Thymocyte
DN Thymocyte
Lymphatic
Endothelial
Collagen
α2β1/α1β1
ROR/
FZD
WNT5a
FGF7
FGF10
FGF18
FGFR2
WNT4
FZD
BMP4
BMR
BMP7
NOTCH
DLK1
BMP2
CCR/ACKR
TGFβ
MIF
AREG
Endothelial
CCL19
TGFβR
CD74
EGFR
WNT5a
TGFβ
MIF
WNT4
FZD
NOTCH
DLK1
BMP2
BMR
TGFβ
TGFβR
AREG
EGFR
BMP7
MIF
CD74
DP Thymocyte
DN Thymocyte
Lymphatic
Endothelial


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
CD74_MIF
NOTCH4_DLL4
NOTCH2_DLL4
NOTCH1_DLL4
NOTCH1_DLK1
DLL4_NOTCH3
DLL1_NOTCH2
DLK1_NOTCH4
DLK1_NOTCH2
CCRL2_CCL19
CCR7_CCL19
CCL25_CCR9
WNT5A_ROR1
WNT5A_FZD3
WNT4_FZD1
FZD6_WNT4
FGF7_FGFR2
FGF10_FGFR2
BMPR1A_BMPR2_BMP4
BMPR1A_BMPR2_BMP2
ACVR_1A2A receptor_BMP7
COL1A1_a2b1 complex
COL1A2_a2b1 complex
COL6A1_a1b1 complex
COL6A2_a1b1 complex
COL6A1_a2b1 complex
COL6A2_a2b1 complex
Log2 Mean Expression
0
-2
-4
-6
-Log10 P-Value
0
1
2
3

**Supplement S4 G**

(G) Left panel: Graphical representation of key receptor ligand interactions within the thymic microenvironment. Right panel: Dot plot visualization of select CellPhone DB receptor-ligand interactions across select thymic cell types.

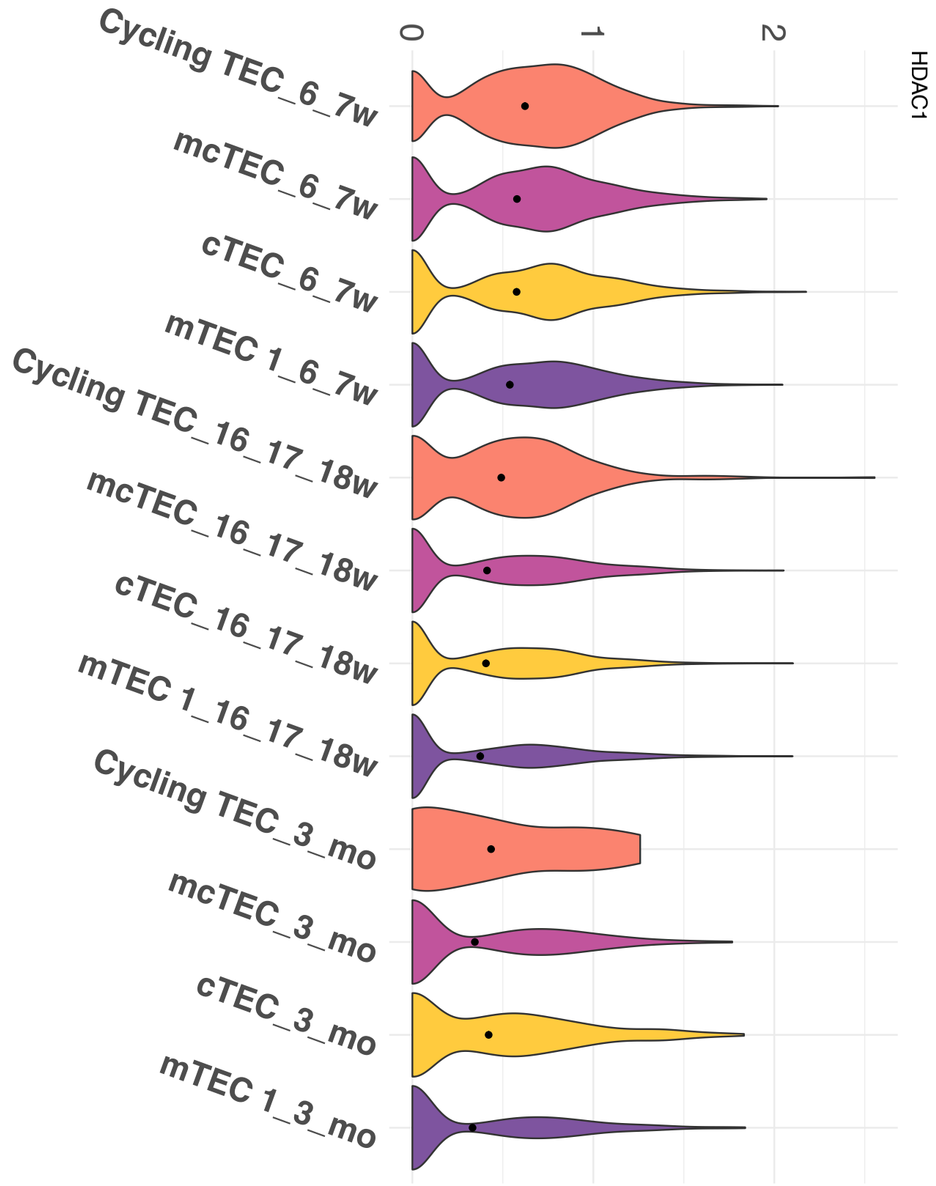

S.Score
40
20
0
-20
tSNE_2
-40
-20
0
20
40
tSNE_1
0.6
0.4
0.2
0.0
-0.2


G2M.Score
40
20
0
-20
tSNE_2
-40
-20
0
20
40
tSNE_1
1.0
0.5
0.0
-0.5

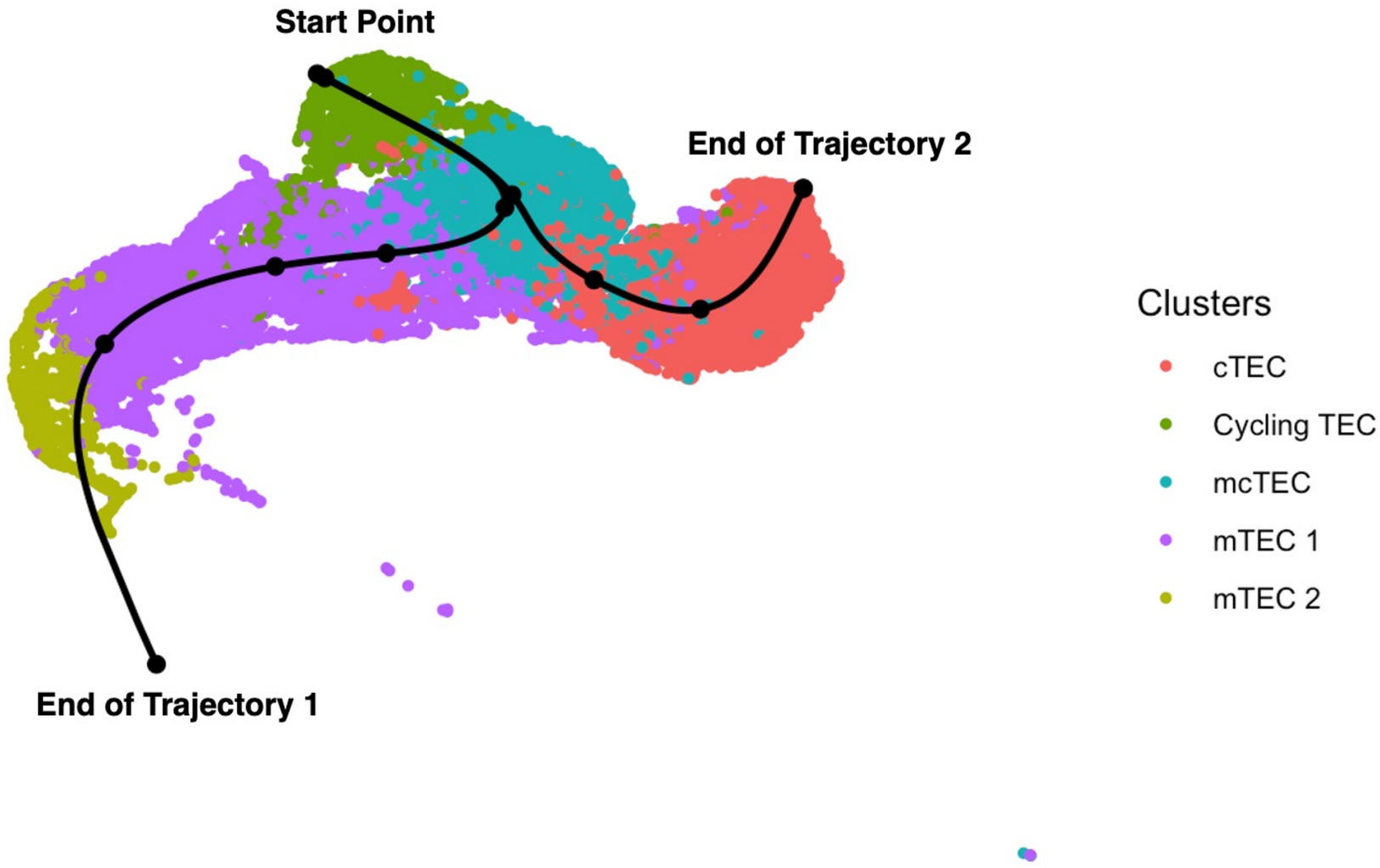
**Supplementary Figure S5A.**
tSNE visualization of thymic epithelial cells across all sampled developmental time points colored Seurat's G2M and S phase modular score expression.

| | cTEC | Cycling TEC | mcTEC | mTEC 1 | mTEC 2 | mTEC 3/4 | Parathyroid | TEC myo/neuro |
|---|---|---|---|---|---|---|---|---|
| **cTEC** | 1.0000000 | 0.4911876 | 0.8140999 | 0.3195151 | -0.3966713 | -0.574202417 | -0.38886332 | -0.356109098 |
| **Cycling TEC** | 0.4911876 | 1.0000000 | 0.5923346 | 0.1347333 | -0.3968367 | -0.483725579 | -0.27806839 | -0.353732234 |
| **mcTEC** | 0.8140999 | 0.5923346 | 1.0000000 | 0.4610582 | -0.3994865 | -0.883557995 | -0.39598630 | -0.113158977 |
| **mTEC 1** | 0.3195151 | 0.1347333 | 0.4610582 | 1.0000000 | 0.1161085 | -0.509068566 | -0.43564113 | -0.218665874 |
| **mTEC 2** | -0.3966713 | -0.3968367 | -0.3994865 | 0.1161085 | 1.0000000 | 0.481200775 | -0.25174182 | -0.420507793 |
| **mTEC 3/4** | -0.5742024 | -0.4837256 | -0.8835580 | -0.5090686 | 0.4812008 | 1.000000000 | 0.01899081 | 0.001222167 |
| **Parathyroid** | -0.3888633 | -0.2780684 | -0.3959863 | -0.4356411 | -0.2517418 | 0.018990806 | 1.00000000 | 0.113643786 |
| **TEC myo/neuro** | -0.3561091 | -0.3537322 | -0.1131590 | -0.2186659 | -0.4205078 | 0.001222167 | 0.11364379 | 1.000000000 |

**Supplementary Figure S5B.**
Data frame of values constituting the heatmap found in (5D).

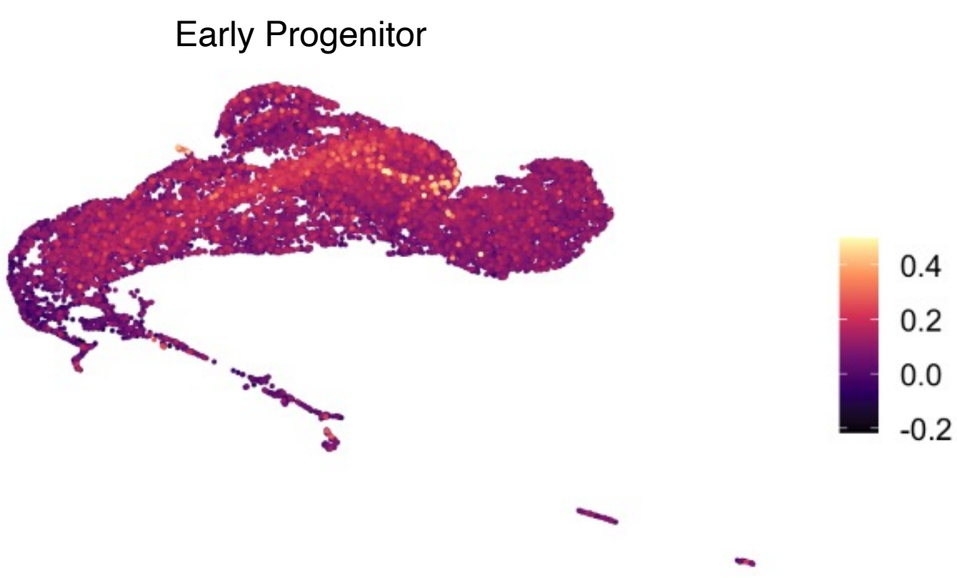

HDAC1
0
1
2
Cycling TEC_6_7w
mcTEC_6_7w
cTEC_6_7w
mTEC 1_6_7w
Cycling TEC_16_17_18w
mcTEC_16_17_18w
cTEC_16_17_18w
mTEC 1_16_17_18w
Cycling TEC_3_mo
mcTEC_3_mo
cTEC_3_mo
mTEC 1_3_mo

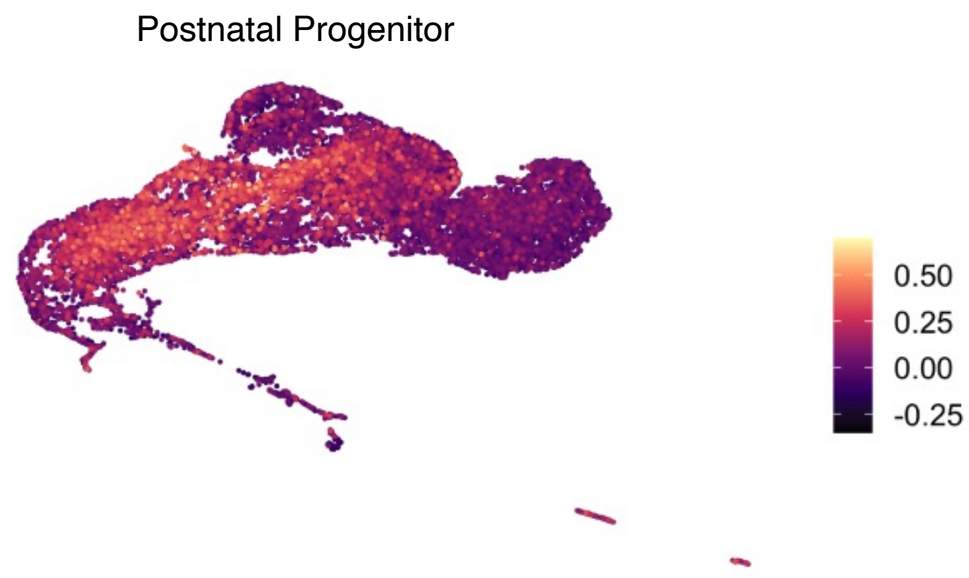
**Supplementary Figure S5C.**
Violin plot representation of HDAC1 expression in key thymic epithelial cell types at different time points. Dot within violin plot represents mean module score.

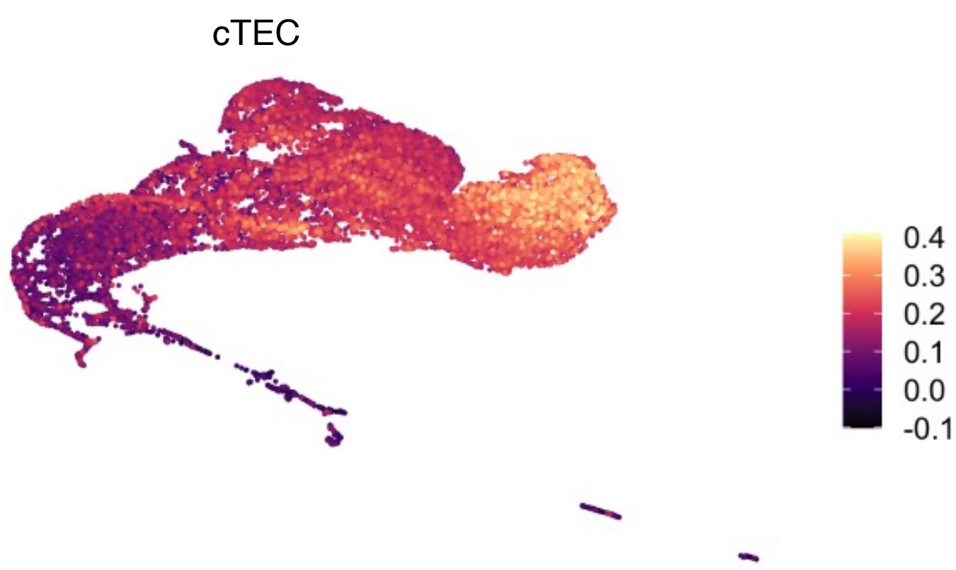

Start Point
End of Trajectory 2
End of Trajectory 1
Clusters
cTEC
Cycling TEC
mcTEC
mTEC 1
mTEC 2

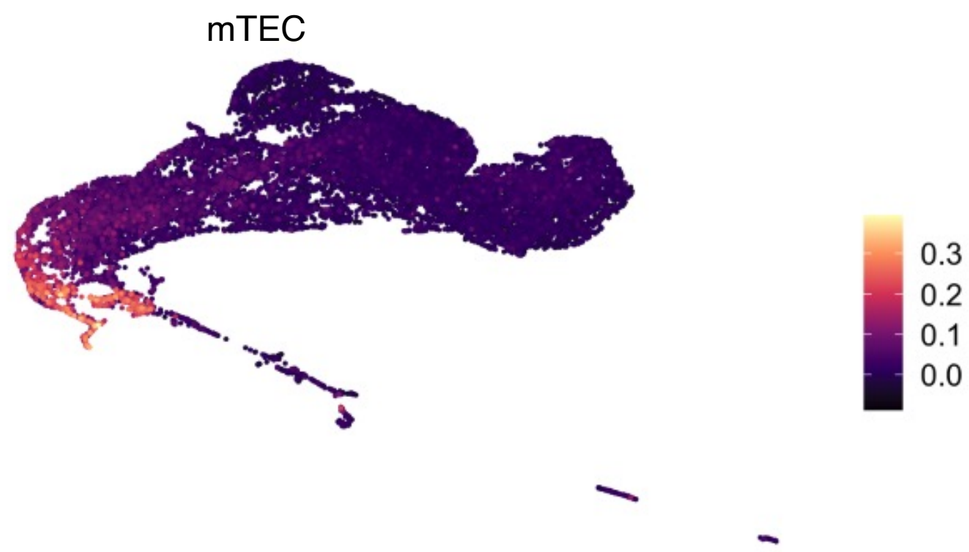
**Supplementary Figure S6A.**
tSpace UMAP of thymic epithelial cells across all gestational ages (excluding mTEC 3/4, TEC myo/neuro, and Parathyroid) colored by cell type. (B) Slingshot identified trajectories 1 and 2 are plotted on the tSpace UMAP from (6A). Pseudotime knots used for TradeSeq analysis are also visualized.

cTEC
Cycling TEC
mcTEC
mTEC 1
mTEC 2
mTEC 3/4
TEC myo/neuro
Parathyroid

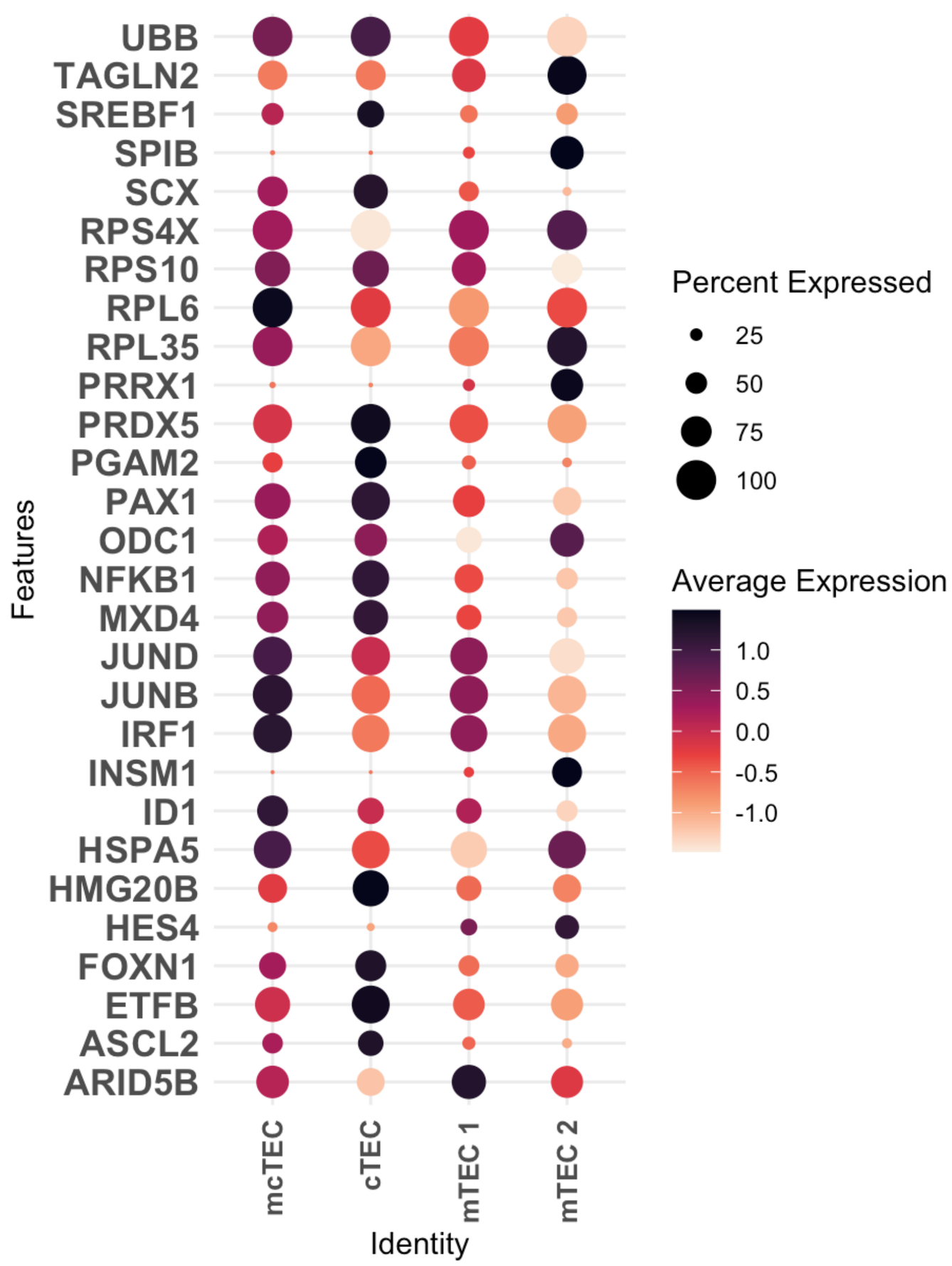

Early Progenitor
0.4
0.2
0.0
-0.2


Postnatal Progenitor
0.50
0.25
0.00
-0.25


cTEC
0.4
0.3
0.2
0.1
0.0
-0.1

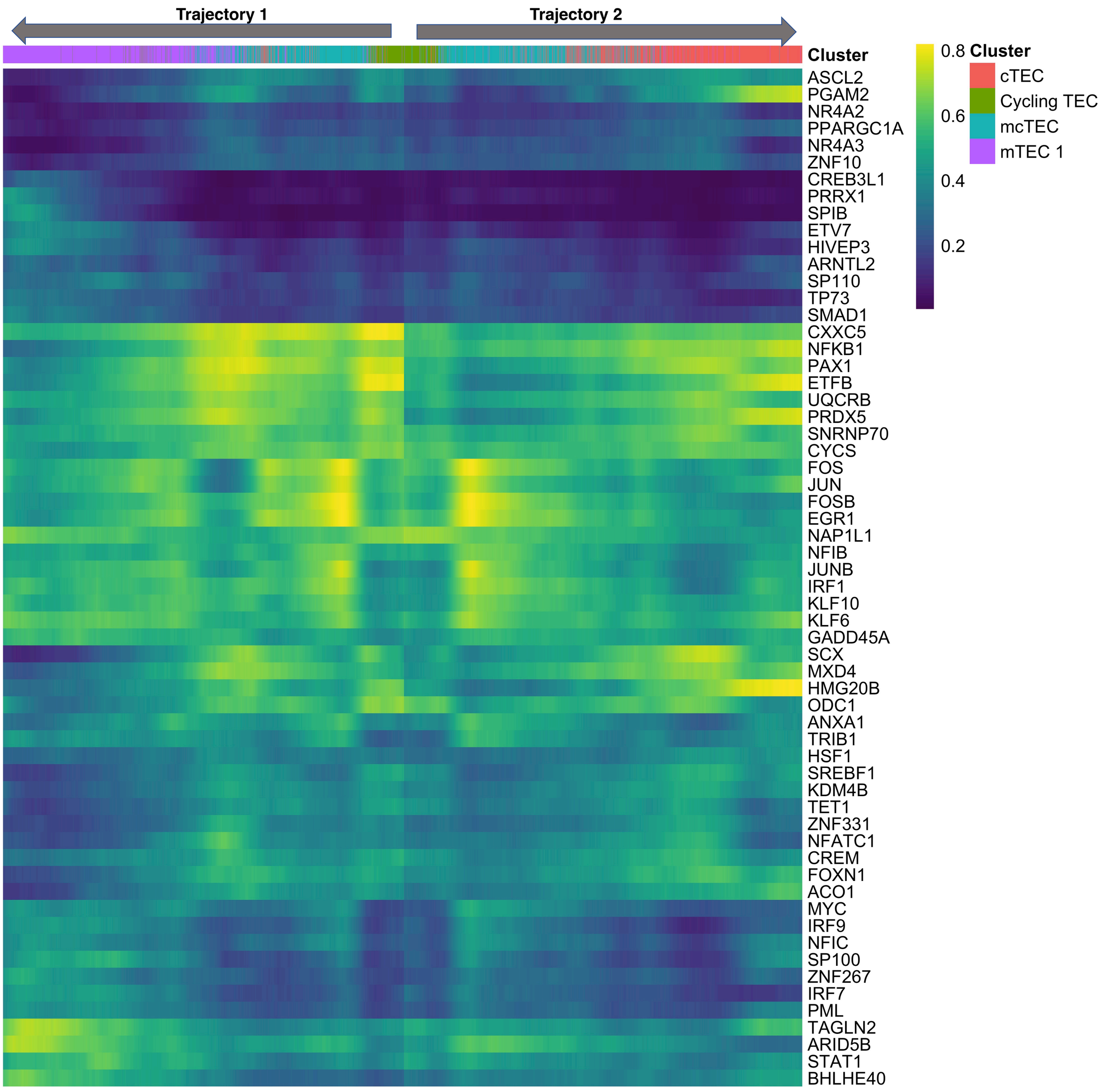

mTEC
0.3
0.2
0.1
0.0

**Supplementary Figure S6B.**
tSpace UMAP visualization from Figure 6A of thymic epithelial cells across all sampled developmental time points, overlaid with score of gene expression modules composed of gene sets reflecting different thymic epithelial cell populations described by Nusser et al. 2022[104].


UBB
TAGLN2
SREBF1
SPIB
SCX
RPS4X
RPS10
RPL6
RPL35
PRRX1
PRDX5
PGAM2
PAX1
ODC1
NFKB1
MXD4
JUND
JUNB
IRF1
INSM1
ID1
HSPA5
HMG20B
HES4
FOXN1
ETFB
ASCL2
ARID5B
Features
mcTEC
cTEC
mTEC 1
mTEC 2
Identity
Percent Expressed
25
50
75
100
Average Expression
1.0
0.5
0.0
-0.5
-1.0

**Supplementary Figure S6C.**
Dot plot representation of central TF expression from figure (6C) across key TEC clusters. Color representative of average expression level. Dot size representative of percentage of cells in cell type expressing selected marker gene.

Supplementary Figure S6D



**Supplementary Figure S6D.**
Heatmaps of the rolling average of top differentially expressed TFs between knots 2-4 along trajectories 1 and 2. Rolling average and quantile scaling separate for each trajectory.