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

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

*These authors contributed equally
#Corresponding authors

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

ABSTRACT

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

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

Distinct gene regulatory patterns delineate cortical, and medullary thymic epithelial cells, conducting airway basal cells, respiratory bud-tip progenitor cells and esophageal basal cells. Specifically, the transcription factors SOX2 and TP63, that promote basal epithelial cell fate when co-expressed and drive branching morphogenesis when cycled in a coordinated fashion, show a uniquely divergent $TP63^{high}SOX2^{low}$ expression pattern in the thymic epithelium. In addition, the thymic epithelium shows distinctive interferon and NFκB-driven signature which is not present in other epithelial cells. Within the thymic epithelial compartments, a cycling progenitor pool gives rise to three distinct developmental trajectories. A dichotomy between PI3Kinase and TGFb signaling between cTECs and mTECs was observed. Our studies map gene regulatory signatures during thymic epithelial cell ontogeny with the intent to advance directed differentiation approaches for translational applications.

INTRODUCTION

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

derived from the ventro-caudal aspect of the third pharyngeal pouch (PPEIII). Two genetically separate steps have been implicated in the development of the thymus: Initially, the development of the PPEIII is instructed by signals from the surrounding mesenchyme. In the second phase of development, bone marrow-derived lymphoid precursors migrate to the thymic primordium. Reciprocal signals between cells of hematopoietic and thymic epithelial origin promote thymic organogenesis and T cell maturation, respectively[9]. The earliest detection of the PPEIII transcriptional regulators TBX1, HOXA3, PAX1 and PAX9 in human embryonic samples was demonstrated at the beginning of week 6 post-conception (CS16) and preceded the expression of the master thymic transcription factor FOXN1 mid-week 6 (CS 16-17)[10]. These findings are concordant with murine data in which Foxn1 is required cell-autonomously for maturation of the thymic epithelium but not for initiation of thymus organogenesis[9,11,12]. Lymphoid progenitors enter the human thymic anlage as early as week 7[4,10] and are indispensable for thymic epithelial maturation and expression of functional markers, such as PSMB11 or AIRE[13-15]. Parallels between branching morphogenesis of the airway and the initiation of cortical and medullary separation in the thymus have been proposed. In both setting, a homogenous epithelial cell population begins to polarize and display distinct apical-basal features. Epithelial cells in the center of the Thymic Anlage gain CLDN3/4 expression and go on to form the medulla, however, a lumen doesn't form[16,17]. Important insights into the developmental programs that instruct thymic organogenesis were derived from murine knock out studies in *Foxn1* which generate an athymic and hairless (nude) phenotype[18]. The thymic development of nude mice arrests after Cldn3/4 is expressed but before colonization with lymphoid precursors occurs[12,17]. The thymic rudiment of *Foxn1* knock-out mice displays nonfunctional cystic features strongly resembling respiratory epithelium in both, morphology, and gene expression[9,19-21]. It has therefore been suggested that Foxn1 induces thymic epithelial cell commitment at the expense of respiratory epithelial cell fate as the default pathway by inhibiting branching morphogenesis [17,19,20]. Morphogenesis of the esophagus, the other major anterior foregut-derived organ, occurs in tight coordination with the patterning of respiratory system[22] and deviations threaten appropriate separation between both organ systems[23].

While the signals that drive anterior foregut development from pluripotent cells as well as lung and esophageal commitment from anterior foregut have been successfully established[24-29], the developmental programs that define patterning of the third pharyngeal pouch and thymus from the anterior foregut endoderm remain elusive. To close this knowledge gap, we sought to investigate the signals that instruct the development and specialization of the human thymic epithelium, using a dual single cell RNA sequencing (scRNA-seq) approach: (A) To understand lineage branch points between fetal thymus to other organs that develop in close temporo-spatial proximity from the anterior foregut, we have compared the transcriptomes of human fetal bronchus, lung, esophagus as well as the parathyroid epithelial cells, in the following referred to as *horizontal* comparison. (B) To define the proliferative dynamics that give rise to developmental hierarchies within specialized thymic epithelial compartments, we have compared human thymus samples sequentially during embryonic, fetal, and early postnatal stages of organogenesis, in the following referred to as *vertical* comparison.

Our studies build on the recent advances in mouse and human thymic epithelial cell characterization using single cell transcriptomics that demonstrated an abundance of unique thymic epithelial cell subpopulations which had evaded discrimination by conventional flow cytometric means of characterization. In this work, we focus on understanding differential developmental trajectories between closely related anterior foregut derived epithelial cell populations with a view towards directed differentiation and tissue regeneration approaches.

RESULTS

Human fetal anterior foregut derived epithelial cells at single cell resolution

We analyzed a total of 14 biological specimens from human embryonic, fetal and postnatal tissues estimated to range from post conception week (pcw) 7 to 3 months post birth. For the comparison of the major anterior foregut derived organs (horizontal comparison), we sampled tissue from thymus, airway epithelium, lung parenchymal tissue (pcw 16, 17 and 18) and esophageal epithelium (pcw 16 and 17) from human fetal samples.

To dissect developmental hierarchies within the thymic epithelial lineages, we studied two very early thymic samples (estimated at or before pcw 7), in comparison to pcw 16, 17, 18 and neonatal (age 3 months) thymic tissue. EPCAM+ magnetic bead selection was used for all but the earliest thymic samples to enrich for cells of epithelial origin while preserving fractions of all other subpopulations of the microenvironment. In addition, thymic samples were partially depleted of CD45+ thymocytes (Figure 1B). The very early (pcw 7) thymus samples were used unmanipulated due to very small numbers of thymocytes at this stage of development (Supplementary Table S1).

After sequence alignment to human genome library, quality control including mitochondrial gene removal, and batch correction, the resulting data set yielded single cell transcriptomes with a median of 3000 unique genes per cell. For the horizontal comparison of anterior foregut derived organs, pcw 16, 17 and 18 samples of human thymus, bronchus, lung, and esophagus were clustered unsupervised and visualized using t-SNE. Samples were annotated in concordance with nomenclature employed in previously published organ-specific human single cell datasets for thymus[4], respiratory system[30,31] and esophagus[32], respectively. Within the thymus, we identified a total of 17 epithelial and non-epithelial cell populations. Epithelial cells further subdivided into canonical *CCL25, PSMB11, KRT8*- high cortical thymic epithelial cells (cTEC) and *CCL19* high, *KRT5/14*-high medullary thymic epithelial cells (mTECI). The previously described population of TECs with overlapping cTEC and mTEC features as well as unique expression of *CTGF* and *DLK2* termed mcTECs[4], was the third most abundant population visualized. Classical mTECII were characterized by high transcript levels of *AIRE* and *CLDN4*. The recently described medullary subpopulations mTECIII and mTECIV, characterized by expression of *KRT1* and *POUF3* respectively, which correspond to the corneocyte-like post-AIRE mTECs and Tufft cells in the mouse[2,3,33], were also visualized. TECmyo and TECneuro were defined by *NEUROD1*, *CHGA*, *CHGB*, *MYOD1* and *CHRNA1*, respectively. Nonepithelial clusters included double negative and double positive thymocytes, as well as small numbers of single positive T

cells (*PTPRC*, *CD3D*, *CD4, CD8A, CD8B*), macrophages (*S100A9, RETN*), dendritic cells (*LAMP3*), vascular smooth muscle cells (*PDGFRB, RGS5*), aldehyde-dehydrogenase positive fibroblasts, also known as thymic fibroblasts 1[4] (*PDGFRA, ALDH1A2*), fibronectin-high fibroblasts also known as thymic fibroblasts 2[4] (*FBN1*), vascular endothelial cells (*PECAM1, CDH5, AQP1*), lymphatic endothelial cells (*LYVE1, PROX1*) and neuronal cells (*NRXN1, MPZ*).

Tissues of the conducting airway clustered into bronchial epithelial basal cells (*KRT5, KRT15, TP63*), club cells (*SCGB1A1, BPIFA1*), multi-ciliated cells (*FOXJ1, PIFO*) and submucosal gland cells (*LTF*), along with goblet cells (*MUC5B, BPIFB1*), secretory progenitor cells (SFTPB, SCGB3A2) and bud tip progenitors from the adjacent respiratory parenchyma (*SFTPC, SOX9*). In addition, the airway microenvironment was comprised of immune cells (B cells, T cells), myoepithelial cells, smooth muscle, vascular smooth muscle, fibroblasts, cartilage, endothelial cells, neuroendocrine and neuronal cells.

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

The esophageal tissue included esophageal basal cells (*KRT5, KRT15, TP63*), suprabasal cells (*KRT13, KRT4, CLDN4*), multiciliate cells (*FOXJ1, PIFO*), gastric epithelial cells (*TFF1, MUC5B*) and non-foregut-derived cell types including immune cells (T cells, macrophages), smooth muscle, vascular smooth muscle, two different types of fibroblasts, endothelial cells and a sizable cluster of neuronal cells from the vagal nerve that was accidentally included during sample dissection (Figure 1C, D, E and Supplementary Figure S1).

Integration of all four tissues followed by batch correction and quality control showed an even contribution of all tissues (thymus, bronchus, lung, esophagus) to all major cell types (epithelial, endothelial, mesenchymal, immune) with a sharp clustering of cells stem from a shared developmental origin, separating epithelial cells (red) from mesenchymal cells (purple), immune cells (light blue), endothelial cells (green), myoepithelial cells (light pink) and neuronal cells (gray) (Figure 1F). Top differentially expressed genes in different tissue types that were shared across all anterior foregut derived epithelial cells included a variety of keratins (*KRT5, 8, 13, 15, 17, 18, 19*), claudins (*CLDN3, 4*), e-cadherin (*CDH1*), desmoplakin (*DSM*), the transcription factors *ELF3* and *HOPX*, the calcium binding protein *CAPS2* and interestingly the arginase *ARG2* that is known to govern mTOR signaling, autophagy and senesce in other cell types[34,35] (Figure 1G, Supplementary Figure S2).

SOX2 and TP63 show divergent expression pattern in the Thymic Epithelial Cells
To better understand the distinct developmental programs that give rise to epithelial cells of different anterior foregut origins at highest resolution, we computationally isolated epithelial cells (Figure 2A, Supplementary Figure S2), followed by reintegration and repeat clustering (Figure 2B). Cells derived from the same organ generally clustered together as did cells of shared function (i.e., multiciliate cells). The closest cluster relationship to thymic epithelial cells was shown by suprabasal and basal epithelial cells (Figure 2B).
Given the unresolved question of the role of branching morphogenesis in thymic development, we specifically interrogated the expression of the transcription factors SOX2 and TP63 that cycle *on* and *off* in a synchronized fashion to promote branching of the airways during development of the respiratory system. In the mouse *Sox2* is known to be highly expressed in non-branching and absent from branching regions, suggesting that downregulation of Sox2 is mandatory for airway epithelium to respond to branch inducing signals[36]. *Sox2* expression was shown to increase again after the branching process and directly induced the expression of *Tp63*, the master switch toward the basal cell lineage[37]. In our human data set, SOX2 and TP63 were highly co-expressed in basal epithelial cells of bronchus and esophagus and gradually decreased towards the parenchymal regions of the lung. All epithelial cells of the thymus, however, were marked by strong expression of TP63 while SOX2 was absent, both on an RNA and protein level (Figure 2C). Interestingly, the key foregut gene regulatory network transcription factors including FOXA1, SOX2, IRX3, IRX5[25,38] also remained highly expressed in epithelial cells of bronchial and esophageal origin and maintained moderate expression in the lung but were entirely absent in all epithelial cells of thymic origin which demonstrated persistently strong expression of the third pharyngeal pouch markers PAX1, PAX9, EYA1, SIX1 and FOXG1. (Figure 2D).This observation is compatible with the previously put forth hypothesis that the development of the third pharyngeal pouch is linked to an overriding transcriptional program the abrogates branching morphogenesis in basal epithelial cells of the anterior foregut.

Next, we interrogated Claudin (CLDN) family genes which encode tight junction proteins that are known to play a critical role in apical basal polarity and tube formation[39,40] (Figure 2B). Claudin-3 and 4 are specifically implicated in epithelial to mesenchymal transition[41] and the gain for Claudin-3 and 4 is classically considered a hallmark of medullary thymic epithelial commitment[16,17]. Interestingly, *CLDN3* and 4 were highly expressed in epithelial cells of bronchus, lung and esophagus while *CLDN5*, a gene missing in the commonly deleted region in 22q11 Deletion Syndrome was the prevailing *CLDN* gene in the majority of epithelial cells in thymus. *CLDN3* and *CLDN4* expression was uniquely high in the mTECII population.

FN1, VIM and VCAM1, other genes associated with epithelial to mesenchymal transition[42] were among the top differentially expressed genes in mTEC1. Although the contribution of laminins to the thymic extracellular matrix has frequently been described[43-46], collagens COL4A1, COL6A1, COL6A2, COL17A1, integrin alpha subunit V (ITGAV) and Galectin 1 predominated in the thymic samples while unique laminins and other extracellular matrix proteins, e.g. perlecan, were selectively high in basal cells

of bronchus and esophagus as well as bud tip progenitor adjacent cells of the lung. The distinctive extracellular environment is matched by a characteristic intracellular scaffold with keratins predominating the cytoskeleton while functioning as main mediators for adhesion, migration, stress response and metabolic reactions in epithelial cells[47]. Keratins (*KRT*) 8, 5, 14 and 17 predominate in cTECs and mTECs. Interestingly, mcTECs overall show a less keratinized phenotype possibly indicative of a developmentally less mature cell state. Bronchial basal cells share many of the keratins with TECs and show a uniquely high expression of *KRT* 15 while esophageal basal and suprabasal cells were specifically high in *KRT4*, *6B* and 13. Bud tip and bud tip adjacent cells of the lung use EPCAM as the prevailing cell cytoskeletal element, in line with their more pliable tissue architecture.

To investigate the unique ability thymus to attract thymocytes and other immune cells and specifically, how this differs from other anterior foregut derived organs, we queried the expression of a wide range of chemokine receptors and ligands in our data set. Chemokine ligands are important chemo-attractants that actively recruit immune cells with cell type- and tissue-specific tropism. In addition to the well know chemokine ligands *CXCL12*, *CCL25* as well as *CCL19* and *CCL21* that have been shown to display cell type specificity for cortical and medullary TECs respectively[48], we found high expression levels of *CXCL14* in medullary TECs. CXCL14 is a natural inhibitor of the CXCL12:CXCR4 signaling axis[49]. The CXCL12:CXCR4 axis is used to attract CXCR4 positive thymocytes to the thymic cortex[50]. Expressing an inhibitor of this pathway in locations devoid of CXCR4 positive thymocytes raises the possibility that CXCL14 contributes to the spatial demarcation of immune cells within the thymic epithelial architecture. Interestingly, we also found a previously not recognized abundance of the C-C motif chemokine ligand *CCL2* in mcTECs. CCL2 has been shown to play a role in attracting macrophages and dendritic cells as well as in inducing tolerance in T cells[51,52]. Our data suggest that *CCL2* is a novel uniquely expressed gene in mcTECs, for which connective tissue growth factor *CTGF* and the non-canonical inhibitor of NOTCH signaling *DKL2* have thus far been the only uniquely described genes[4]. (Figure 2B, C, D).

Medullary thymic epithelial cells demonstrate unique response to interferon type 1 and 2

To infer signaling pathways that could direct cell-type differentiation in an unbiased fashion, we performed Gene Set Enrichment Analysis (GSEA). Network visualization of positively enriched gene sets across select epithelial cells revealed three distinct clusters of nodes, representing cortical epithelial cells (cTEC), medullary epithelial cells (mTECI) and a combined cluster of esophageal basal cell (EEC) and bud tip progenitor cell (LEC) identity in which nodes were divided. Clusters of cTEC and mTEC origin shared abundant connections between nodes. Basal cells of bronchial origin (BEC) were spread apart with minimal spatial concentration (Figure 3A, and Supplementary Figure S3A). The most highly scoring gene sets in cTECs related to TCR signaling, protein processing by the proteasome, antigen cross presentation and NFκB activation. In mTECs, interferon $\alpha\beta$ and interferon γ signaling as well as antigen presentation were

the leading enriched gene sets. Shared gene sets in LECs and EECs all related to ribosomal function, highlighting the importance protein synthesis plays in these cell types (Figure 3B, top panel, and Supplementary Figure S3). To investigate developmental drivers shared by cortical and medullary thymic epithelial cell that may be overpowered due to the dominant effect of highly expressed individual genes or gene subsets, we removed gene sets related to proteasome structure and function from the analysis (Figure 3B, bottom panel, and Supplementary Figure S3). Even after removal of the proteasome gene sets, interferon $\alpha\beta$ and interferon γ module remained uniquely enriched in the mTEC compartment. Leading edge genes with the greatest abundance in the mTEC compartment were defined by direct effectors of interferon signaling including *IFI6, ISG15, IFI27, BST2, IFITM3* and *MX1*; while leading edge genes in cTEC compartment where largely MHC class 1 and 2 alleles implicated in antigen presentation.

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[53]. A dose-dependent excess of exogenous retinoic acid intake during pregnancy, e.g. in the form of vitamin A is associated with defects in thymus and parathyroid along with various other malformations[54,55]. Interestingly, children with mutations in RA metabolism (*STRA6*, *ALDH1A2*) resulting in RA deficiency present with the same spectrum of congenital defects associated with RA toxicicy[56,57]. These findings suggest that an intricate balance of retinoic acid signaling is essential for thymic development. To understand the differential impact that RA plays in the development of anterior foregut-derived epithelial cells, we developed an RA metabolism gene regulatory module that integrates a curated set developmentally relevant RA metabolism genes, separated into RA synthesis, RA homeostasis and nuclear receptor interactions and RA degradation (Figure 4A).

Out of our data set, basal epithelial cell of bronchial (BEC) and esophageal (EEC) origin scored highest in RA synthetic capacity within an overall narrow range across all queried cell types. BECs and cTECs displayed the nuclear activity score, once again within a relatively narrow range. Strikingly, cTECs had substantially the highest RA degradation score and were the only queried cell type with significant RA breakdown by any of the three RA cytochrome P450 family 26 subfamily members solely specific to retinoic acid, but not other aldehydes or alcohols (Figure 4A and B). This raises the possibility that RA levels in TEC are under uniquely tight control. Retinoic acid signaling can be mediated through autocrine or paracrine sources of RA. As RA synthesis scores were low across all types on epithelial cells including cTECs, we investigated exogenous sources of RA in the thymic microenvironment using the same functional RA metabolism models. Highest expression of RA metabolizing enzymes was found in both types of thymic fibroblasts, including aldehyde-dehydrogenase positive fibroblasts also known as thymic fibroblasts 1^4 (*ALDH1A2*), and fibronectin-high fibroblasts also known as thymic fibroblasts 2^4 (*FBN1*) (Figure 4C and D). These data suggest that TEC, specifically cortical TECs tightly regulate paracrine sources of retinoic acid. Interestingly, endothelial cells and lymphatic endothelial cells of the thymic microenvironment showed highest scores of nuclear activities while cTECs once again

showed highest scores of RA degradation not only within the epithelial but also across the thymic microenvironmental data subset.

MATERIALS AND METHODS

Tissue Acquisition

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

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

Tissue Digestion

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

Epithelial Cell Enrichment

For fetal thymic tissue (16-18 pcw) only, cell suspension was depleted of thymocytes and other immune cells using Human CD45 MicroBeads (Miltenyi) and LD depletion columns (Miltenyi),10^8 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 flowcytomery on Beckman FACSAria II with CD45-APC, CD3-FITC, CD205-PE, EpCAM-APC, and L/D-NearIR-APC-Cy7 prior to MACS purification. Due to epitope blockage from magnetic

bead antibodies, additional markers, i.e. CD7-AF647 (thymocytes) and CD105-PerCP/Cy5.5 (thymic mesenchymal cells) were used to assess post enrichment purity.

scRNAseq Library Preparation and Sequencing

Gel-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 and downstream processing was performed using Seurat V4.0.

Differential gene expression

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

Gene set analysis

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

Module scores

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

Trajectory and pseudotime analysis

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

SingleCellExperiment/1.12.0. Cycling TECs were selected as the starting-point cluster, while end-point clusters were not specified. We then extracted principal curve coordinates and cellular pseudotime values based on the resulting principal curves.

Modeling gene expression along pseudotime

We utilized tradeSeq/1.4.0[61] to build a negative binomial generalized additive model of single-cell gene expression along pseudotime. This allowed for statistical testing of both within-trajectory and between-trajectory gene expression differences as well as smoothing of gene expression along psueudotime for the purpose of visualization. Genes with statistically different expression patterns along an individual trajectory as well as those with statistically different expression between the ends of the two trajectories were determined using the Wald test. To determine transiently expressed genes, we identified those with significant variation in expression along pseudotime but with minimal difference in expression between the ends of pseudotime trajectories. We calculated a transient score based on the sum of squared Wald score for within-trajectory gene expression and the squared negative Wald square for between-trajectory end expression.

REFERENCES

1. Kernfeld EM, Genga RMJ, Neherin K, Magaletta ME, Xu P, Maehr R. A Single-Cell Transcriptomic Atlas of Thymus Organogenesis Resolves Cell Types and Developmental Maturation. *Immunity.* 2018;48(6):1258-1270 e1256.
2. Bornstein C, Nevo S, Giladi A, et al. Single-cell mapping of the thymic stroma identifies IL-25-producing tuft epithelial cells. *Nature.* 2018;559(7715):622-626.
3. Miller CN, Proekt I, von Moltke J, et al. Thymic tuft cells promote an IL-4-enriched medulla and shape thymocyte development. *Nature.* 2018;559(7715):627-631.
4. Park JE, Botting RA, Dominguez Conde C, et al. A cell atlas of human thymic development defines T cell repertoire formation. *Science.* 2020;367(6480).
5. Bautista JL, Cramer NT, Miller CN, et al. Single-cell transcriptional profiling of human thymic stroma uncovers novel cellular heterogeneity in the thymic medulla. *Nat Commun.* 2021;12(1):1096.
6. Zeng Y, Liu C, Gong Y, et al. Single-Cell RNA Sequencing Resolves Spatiotemporal Development of Pre-thymic Lymphoid Progenitors and Thymus Organogenesis in Human Embryos. *Immunity.* 2019;51(5):930-948 e936.
7. Zorn AM, Wells JM. Vertebrate endoderm development and organ formation. *Annu Rev Cell Dev Biol.* 2009;25:221-251.
8. Graham A, Richardson J. Developmental and evolutionary origins of the pharyngeal apparatus. *Evodevo.* 2012;3(1):24.
9. Nehls M, Kyewski B, Messerle M, et al. Two genetically separable steps in the differentiation of thymic epithelium. *Science.* 1996;272(5263):886-889.
10. Farley AM, Morris LX, Vroegindeweij E, et al. Dynamics of thymus organogenesis and colonization in early human development. *Development.* 2013;140(9):2015-2026.

11. 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.
12. 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.
13. 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.
14. 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.
15. 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.
16. 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.
17. 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.
18. Flanagan SP. 'Nude', a new hairless gene with pleiotropic effects in the mouse. *Genet Res.* 1966;8(3):295-309.
19. Dooley J, Erickson M, Roelink H, Farr AG. Nude thymic rudiment lacking functional foxn1 resembles respiratory epithelium. *Dev Dyn.* 2005;233(4):1605-1612.
20. 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.
21. Cordier AC. Ultrastructure of the cilia of thymic cysts in "nude" mice. *Anat Rec.* 1975;181(2):227-249.
22. Edwards NA, Shacham-Silverberg V, Weitz L, et al. Developmental basis of trachea-esophageal birth defects. *Dev Biol.* 2021;477:85-97.
23. Domyan ET, Ferretti E, Throckmorton K, Mishina Y, Nicolis SK, Sun X. Signaling through BMP receptors promotes respiratory identity in the foregut via repression of Sox2. *Development.* 2011;138(5):971-981.
24. 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.
25. 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.
26. 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.

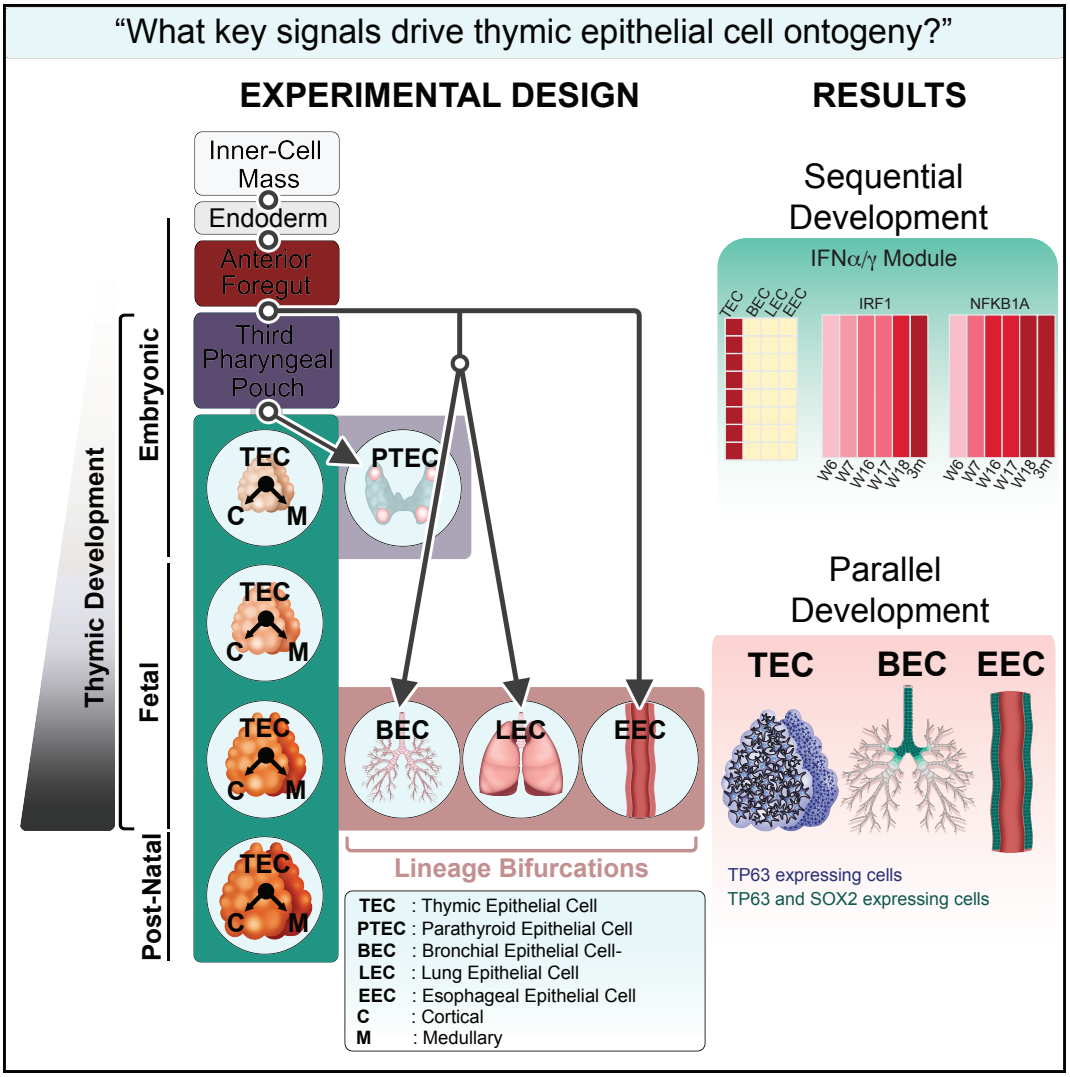
27. Trisno SL, Philo KED, McCracken KW, et al. Esophageal Organoids from Human Pluripotent Stem Cells Delineate Sox2 Functions during Esophageal Specification. *Cell Stem Cell.* 2018;23(4):501-515 e507.
28. Zhang Y, Yang Y, Jiang M, et al. 3D Modeling of Esophageal Development using Human PSC-Derived Basal Progenitors Reveals a Critical Role for Notch Signaling. *Cell Stem Cell.* 2018;23(4):516-529 e515.
29. 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.
30. 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.
31. 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.
32. 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.
33. 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.
34. Xiong Y, Yepuri G, Forbiteh M, et al. ARG2 impairs endothelial autophagy through regulation of MTOR and PRKAA/AMPK signaling in advanced atherosclerosis. *Autophagy.* 2014;10(12):2223-2238.
35. Xiong Y, Yu Y, Montani JP, Yang Z, Ming XF. Arginase-II induces vascular smooth muscle cell senescence and apoptosis through p66Shc and p53 independently of its l-arginine ureahydrolase activity: implications for atherosclerotic plaque vulnerability. *J Am Heart Assoc.* 2013;2(4):e000096.
36. 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.
37. 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.
38. 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.
39. Hewitt KJ, Agarwal R, Morin PJ. The claudin gene family: expression in normal and neoplastic tissues. *BMC Cancer.* 2006;6:186.
40. Baumholtz AI, Gupta IR, Ryan AK. Claudins in morphogenesis: Forming an epithelial tube. *Tissue Barriers.* 2017;5(4):e1361899.
41. Lin X, Shang X, Manorek G, Howell SB. Regulation of the Epithelial-Mesenchymal Transition by Claudin-3 and Claudin-4. *PLoS One.* 2013;8(6):e67496.
42. Kalluri R, Weinberg RA. The basics of epithelial-mesenchymal transition. *The Journal of clinical investigation.* 2009;119(6):1420-1428.
43. 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.
44. 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.
45. 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.
46. Simon T, Bromberg JS. Regulation of the Immune System by Laminins. *Trends in Immunology.* 2017;38(11):858-871.
47. Windoffer R, Beil M, Magin TM, Leube RE. Cytoskeleton in motion: the dynamics of keratin intermediate filaments in epithelia. *J Cell Biol.* 2011;194(5):669-678.
48. <1020.full.pdf>.
49. Tanegashima K, Suzuki K, Nakayama Y, et al. CXCL14 is a natural inhibitor of the CXCL12-CXCR4 signaling axis. *FEBS Lett.* 2013;587(12):1731-1735.
50. Lancaster JN, Li Y, Ehrlich LIR. Chemokine-Mediated Choreography of Thymocyte Development and Selection. *Trends Immunol.* 2018;39(2):86-98.
51. Bleul CC, Boehm T. Chemokines define distinct microenvironments in the developing thymus. *Eur J Immunol.* 2000;30(12):3371-3379.
52. Cedile O, Lobner M, Toft-Hansen H, et al. Thymic CCL2 influences induction of T-cell tolerance. *J Autoimmun.* 2014;55:73-85.
53. 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.
54. Lammer EJ, Chen DT, Hoar RM, et al. Retinoic acid embryopathy. *N Engl J Med.* 1985;313(14):837-841.
55. 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.
56. 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.
57. Pavan M, Ruiz VF, Silva FA, et al. ALDH1A2 (RALDH2) genetic variation in human congenital heart disease. *BMC Med Genet.* 2009;10:113.
58. Liberzon A, Birger B, Thorvaldsdóttir H. et al. The molecular signatures database hallmark gene set collection. *Cell Systems*. 2015; 1:417–425.
59. Wu T, Hu E, Xu S. et al. ClusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation*. 2021; 3:100141.
60. Dermadi D, Bschneider M, Lazarus NH, *et al.* Exploration of cell development pathways through high-dimensional single cell analysis in trajectory space. *iScience*. 2020; 23:100842.
61. Van den Berge K, Bezieux HR, Street K, et al. Trajectory-based differential expression analysis for single-cell sequencing data. *Nat Commun*. 2020; 11:1201.
62. Nusser A, Sagar, Swann JB, et al. Developmental dynamics of two bipotent thymic epithelial progenitor types. Nature. 2022; 606(7912):165-171.

AUTHOR INFORMATION
These authors contributed equally: Abdulvasey Mohammed and Priscila Ferreira Slepicka
These authors contributed equally; correspondence should be addressed to: Andrew Gentles (andrewg@stanford.edu) and Katja G. Weinacht (kgw1@stanford.edu).

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

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



ABSTRACT FIGURE

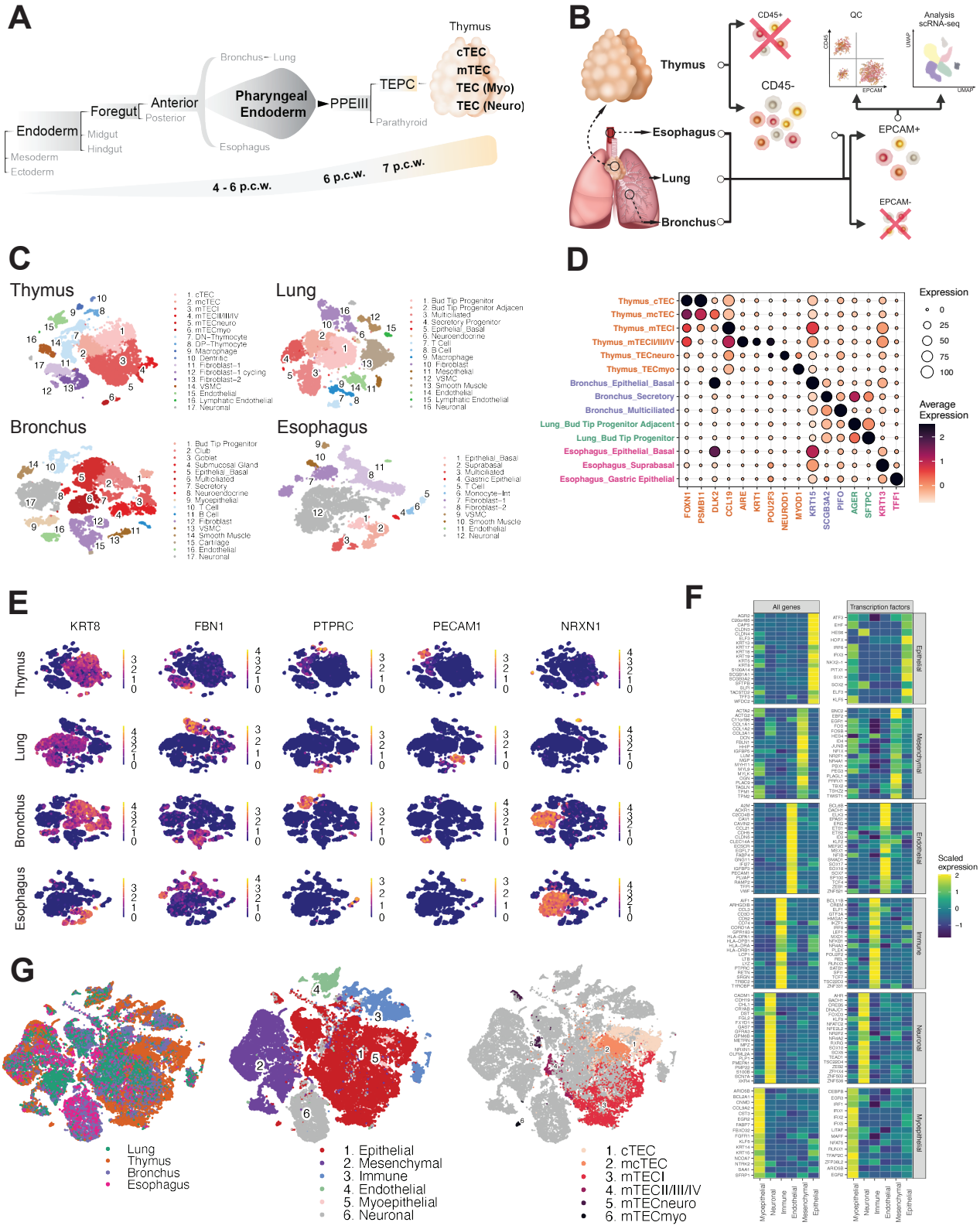

A
Endoderm
Mesoderm
Ectoderm
Foregut
Anterior
Posterior
Midgut
Hindgut
Bronchus
Lung
Pharyngeal Endoderm
Esophagus
PPEIII
TEPC
Parathyroid
Thymus
cTEC
mTEC
TEC (Myo)
TEC (Neuro)
4 - 6 p.c.w.
6 p.c.w.
7 p.c.w.
B
Thymus
Esophagus
Lung
Bronchus
CD45+
CD45-
QC
Analysis scRNA-seq
EPCAM+
EPCAM-
C
Thymus
Lung
Bronchus
Esophagus
D
Expression
Average Expression
E
KRT8
FBN1
PTPRC
PECAM1
NRXN1
Thymus
Lung
Bronchus
Esophagus
F
All genes
Transcription factors
Epithelial
Mesenchymal
Endothelial
Immune
Neuronal
Myoepithelial
Scaled expression
G
Lung
Thymus
Bronchus
Esophagus
1. Epithelial
2. Mesenchymal
3. Immune
4. Endothelial
5. Myoepithelial
6. Neuronal
1. cTEC
2. mcTEC
3. mTECI
4. mTECII/III/IV
5. mTECneuro
6. mTECmyo

Figure 1. Dissecting and integrating the gene expression profile of fetal derived anterior foregut tissues using scRNAseq. (A) Sequential embryonic development showing the foregut development into the pharyngeal apparatus focusing on the thymic ontogeny, including the bipotent third (III) pharyngeal pouch endoderm (PPEIII), ventralization of the PPE (PPEv), thymic epithelial progenitor cell development (TEPC), and the thymus (medullary thymic epithelial cells, mTEC; cortical TEC, cTEC; neuroTEC; myoTEC). (B) Scheme of sequential cell digestion and cell separation of embryonic Thymus, Esophagus, Bronchus and Lungs for quality control check and single cell RNA sequencing library preparation. (C) tSNE visualization of epithelial cells colored by cell type from individual anterior foregut tissues (D) Dot plot showing top expressing genes within selected anterior foregut tissue epithelial cell type populations. Color representative of average expression level. Dot size representative of percentage of cells in cell type expressing selected marker gene. (E)
tSNE visualization of cells from individual anterior foregut tissues overlaid with expression pattern of genes representative of specific cell classes. (F) Heatmap of top differentially expressed genes in specific cell types of different tissue. (G) tSNE visualization of cells colored by cell type from individual anterior foregut tissues.

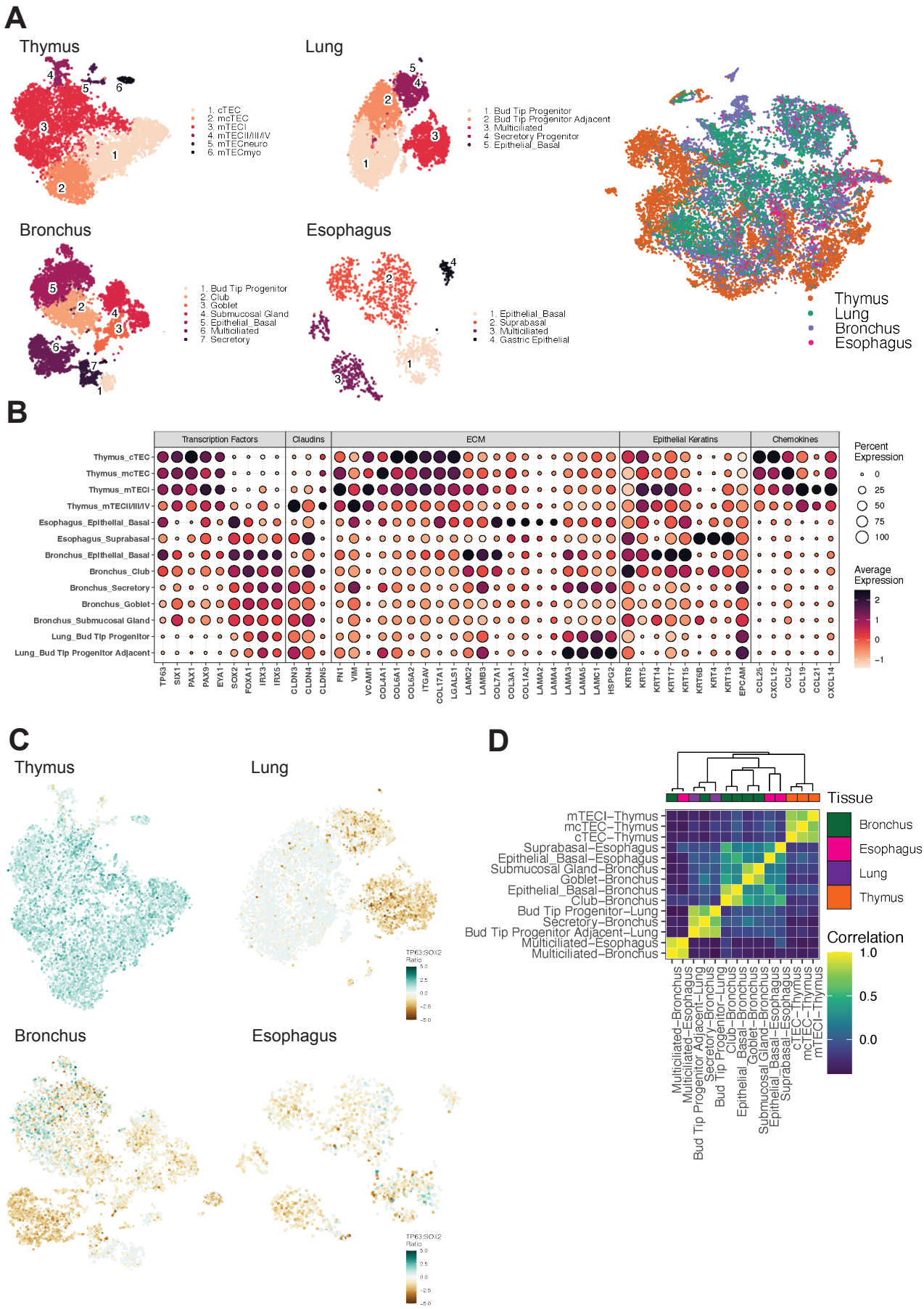

A
Thymus
1. cTEC
2. mcTEC
3. mTECI
4. mTECII/III/IV
5. mTECneuro
6. mTECmyo
Lung
1. Bud Tip Progenitor
2. Bud Tip Progenitor Adjacent
3. Multiciliated
4. Secretory Progenitor
5. Epithelial_Basal
Bronchus
1. Bud Tip Progenitor
2. Club
3. Goblet
4. Submucosal Gland
5. Epithelial_Basal
6. Multiciliated
7. Secretory
Esophagus
1. Epithelial_Basal
2. Suprabasal
3. Multiciliated
4. Gastric Epithelial
Thymus
Lung
Bronchus
Esophagus
B
Transcription Factors
Claudins
ECM
Epithelial Keratins
Chemokines
Thymus_cTEC
Thymus_mcTEC
Thymus_mTECI
Thymus_mTECII/III/IV
Esophagus_Epithelial_Basal
Esophagus_Suprabasal
Bronchus_Epithelial_Basal
Bronchus_Club
Bronchus_Secretory
Bronchus_Goblet
Bronchus_Submucosal Gland
Lung_Bud Tip Progenitor
Lung_Bud Tip Progenitor Adjacent
TP63
SIX1
PAX1
PAX9
EYA1
SOX2
FOXA1
IRX3
IRX5
CLDN3
CLDN4
CLDN5
FN1
VIM
VCAM1
COL4A1
COL6A1
COL6A2
ITGAV
COL17A1
LGALS1
LAMC2
LAMB3
COL7A1
COL3A1
COL1A2
LAMA2
LAMA4
LAMA3
LAMA5
LAMC1
HSPG2
KRT8
KRT5
KRT14
KRT17
KRT15
KRT6B
KRT4
KRT13
EPCAM
CCL25
CXCL12
CCL2
CCL19
CCL21
CXCL14
Percent Expression
0
25
50
75
100
Average Expression
2
1
0
−1
C
Thymus
Lung
Bronchus
Esophagus
TP63:SOX2 Ratio
5.0
2.5
0.0
−2.5
−5.0
D
Tissue
Bronchus
Esophagus
Lung
Thymus
Correlation
1.0
0.5
0.0
mTECI−Thymus
mcTEC−Thymus
cTEC−Thymus
Suprabasal−Esophagus
Epithelial_Basal−Esophagus
Submucosal Gland−Bronchus
Goblet−Bronchus
Epithelial_Basal−Bronchus
Club−Bronchus
Bud Tip Progenitor−Lung
Secretory−Bronchus
Bud Tip Progenitor Adjacent−Lung
Multiciliated−Esophagus
Multiciliated−Bronchus
Multiciliated−Bronchus
Multiciliated−Esophagus
Bud Tip Progenitor Adjacent−Lung
Secretory−Bronchus
Bud Tip Progenitor−Lung
Club−Bronchus
Epithelial_Basal−Bronchus
Goblet−Bronchus
Submucosal Gland−Bronchus
Epithelial_Basal−Esophagus
Suprabasal−Esophagus
cTEC−Thymus
mcTEC−Thymus
mTECI−Thymus

Figure 2. Differential distribution of *TP63* and *SOX2* expression across AFG-derived epithelial cells. (A) 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). (B) Dot plot showing expression of genes from selected gene sets (transcription factors, claudins, ECM, epithelial keratins, chemokines) within selected anterior foregut tissue epithelial cell type populations. Color representative of average expression level. Dot size representative of percentage of cells in cell type expressing selected marker gene. (C) tSNE visualization of epithelial cells from individual anterior foregut tissue overlaid with ratio of *TP63* to *SOX2* gene expression. Ratio represents the scaled difference between log-normalized counts of the two genes. (D) Correlation matrix plot comparing anterior foregut tissues.

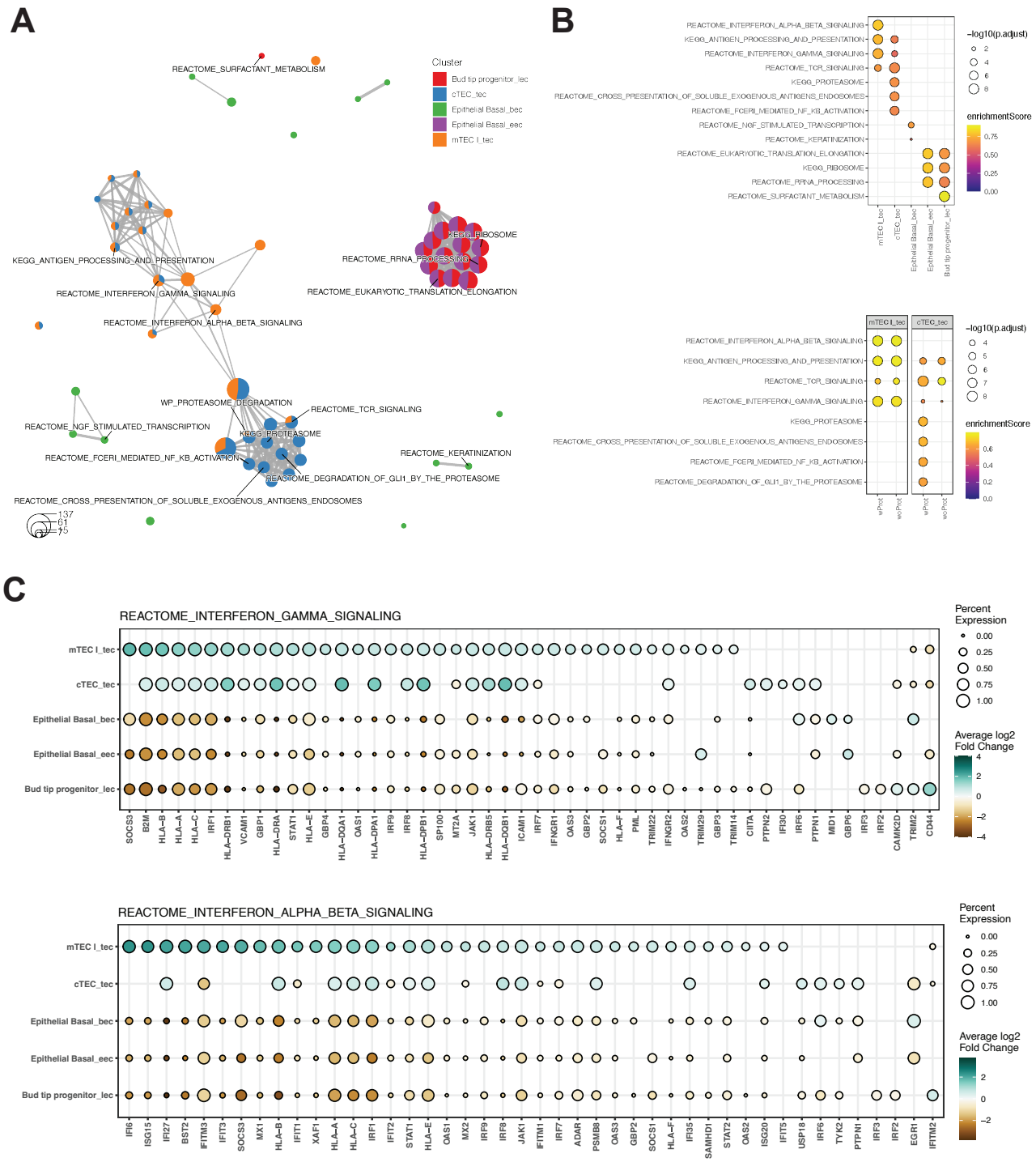

A
Cluster
Bud tip progenitor_lec
cTEC_tec
Epithelial Basal_bec
Epithelial Basal_eec
mTEC I_tec
REACTOME_SURFACTANT_METABOLISM
KEGG_ANTIGEN_PROCESSING_AND_PRESENTATION
REACTOME_INTERFERON_GAMMA_SIGNALING
REACTOME_INTERFERON_ALPHA_BETA_SIGNALING
KEGG_RIBOSOME
REACTOME_RRNA_PROCESSING
REACTOME_EUKARYOTIC_TRANSLATION_ELONGATION
WP_PROTEASOME_DEGRADATION
REACTOME_TCR_SIGNALING
REACTOME_NGF_STIMULATED_TRANSCRIPTION
KEGG_PROTEASOME
REACTOME_FCERI_MEDIATED_NF_KB_ACTIVATION
REACTOME_KERATINIZATION
REACTOME_DEGRADATION_OF_GLI1_BY_THE_PROTEASOME
REACTOME_CROSS_PRESENTATION_OF_SOLUBLE_EXOGENOUS_ANTIGENS_ENDOSOMES
137
61
75
B
REACTOME_INTERFERON_ALPHA_BETA_SIGNALING
KEGG_ANTIGEN_PROCESSING_AND_PRESENTATION
REACTOME_INTERFERON_GAMMA_SIGNALING
REACTOME_TCR_SIGNALING
KEGG_PROTEASOME
REACTOME_CROSS_PRESENTATION_OF_SOLUBLE_EXOGENOUS_ANTIGENS_ENDOSOMES
REACTOME_FCERI_MEDIATED_NF_KB_ACTIVATION
REACTOME_NGF_STIMULATED_TRANSCRIPTION
REACTOME_KERATINIZATION
REACTOME_EUKARYOTIC_TRANSLATION_ELONGATION
KEGG_RIBOSOME
REACTOME_RRNA_PROCESSING
REACTOME_SURFACTANT_METABOLISM
mTEC I_tec
cTEC_tec
Epithelial Basal_bec
Epithelial Basal_eec
Bud tip progenitor_lec
−log10(p.adjust)
enrichmentScore
REACTOME_DEGRADATION_OF_GLI1_BY_THE_PROTEASOME
wProt
woProt
C
REACTOME_INTERFERON_GAMMA_SIGNALING
Percent Expression
Average log2 Fold Change
SOCS3 B2M HLA−B HLA−A HLA−C IRF1 HLA−DRB1 VCAM1 GBP1 HLA−DRA STAT1 HLA−E GBP4 HLA−DQA1 OAS1 HLA−DPA1 IRF9 IRF8 HLA−DPB1 SP100 MT2A JAK1 HLA−DRB5 HLA−DQB1 ICAM1 IRF7 IFNGR1 OAS3 GBP2 SOCS1 HLA−F PML TRIM22 IFNGR2 OAS2 TRIM29 GBP3 TRIM14 CIITA PTPN2 IFI30 IRF6 PTPN1 MID1 GBP6 IRF3 IRF2 CAMK2D TRIM2 CD44
REACTOME_INTERFERON_ALPHA_BETA_SIGNALING
IFI6 ISG15 IFI27 BST2 IFITM3 IFIT3 SOCS3 MX1 HLA−B IFIT1 XAF1 HLA−A HLA−C IRF1 IFIT2 STAT1 HLA−E OAS1 MX2 IRF9 IRF8 JAK1 IFITM1 IRF7 ADAR PSMB8 OAS3 GBP2 SOCS1 HLA−F IFI35 SAMHD1 STAT2 OAS2 ISG20 IFIT5 USP18 IRF6 TYK2 PTPN1 IRF3 IRF2 EGR1 IFITM2

Figure 3. Interferon type 1 and 2 signaling pathway is enriched in medullary TECs compared to other AFG-derived epithelial cells. (A) Network visualization of positively enriched gene sets across select anterior foregut epithelial cells. Only selected gene sets are labeled. Nodes represent individual gene sets. Node color represents relative use of gene set genes by each cell type. Node size reflects number of genes in gene set. Edges weight reflects proportion of genes shared between gene sets with a minimum proportion of 0.2. (B) Dot plots of comparing gene set enrichment of original genes sets and gene sets with proteasomal genes removed across selected anterior foregut epithelial cell types (bottom panel). Dot plots of gene set enrichment across selected anterior foregut epithelial cell types (top panel). Only selected gene sets are included for both dot plots. Color representative of gene set enrichment score. Dot size representative of percentage of -log10 p-value. (C) Dot plot showing relative expression of genes from select gene sets across select anterior foregut epithelial cell types. Color representative of average log2 expression difference between indicated cell type and all other epithelial cell types. Dot size representative of percentage of cells in cell type expressing selected gene.

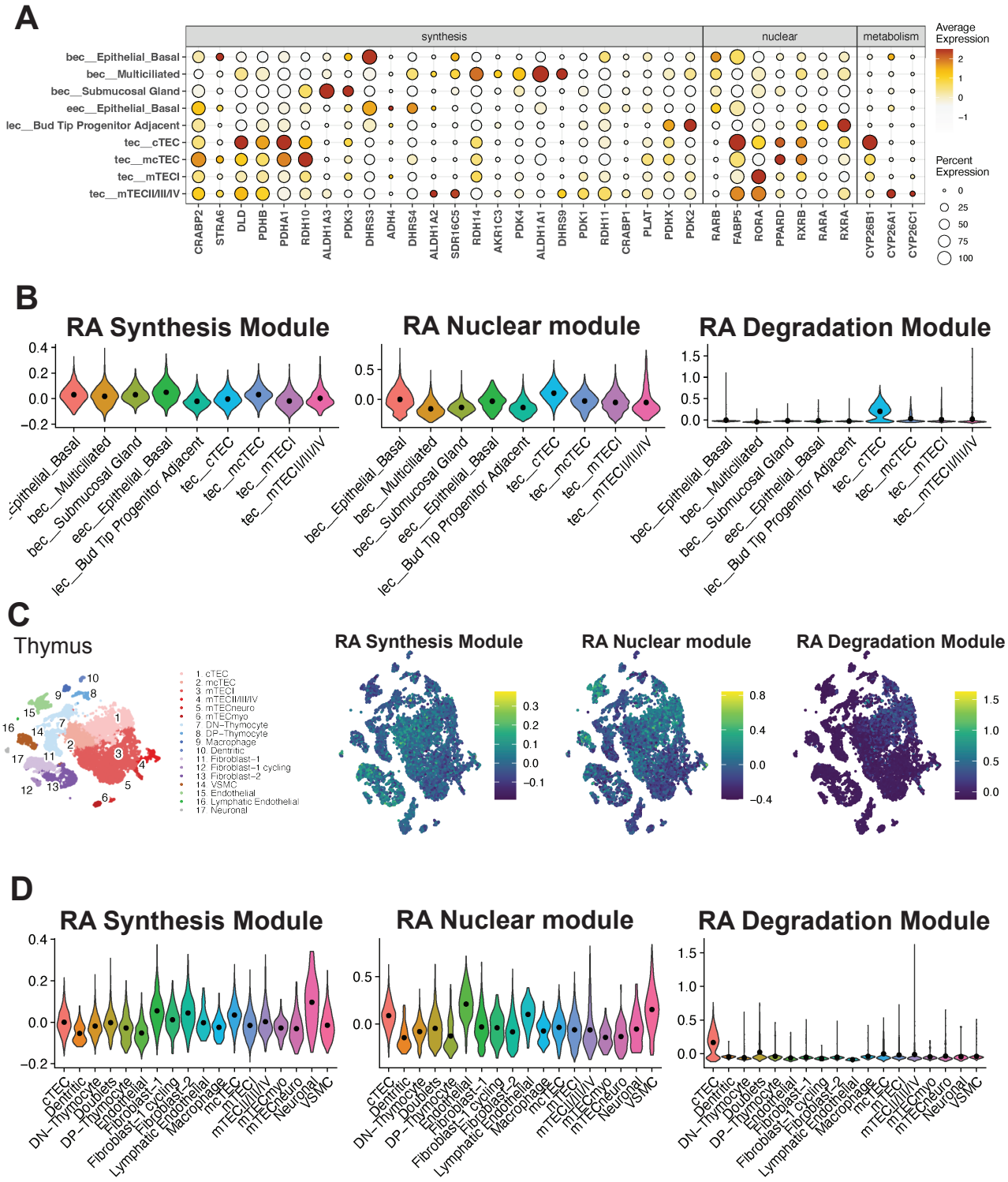

A
synthesis
nuclear
metabolism
Average Expression
Percent Expression
bec__Epithelial_Basal
bec__Multiciliated
bec__Submucosal Gland
eec__Epithelial_Basal
lec__Bud Tip Progenitor Adjacent
tec__cTEC
tec__mcTEC
tec__mTECI
tec__mTECII/III/IV
CRABP2
STRA6
DLD
PDHB
PDHA1
RDH10
ALDH1A3
PDK3
DHRS3
ADH4
DHRS4
ALDH1A2
SDR16C5
RDH14
AKR1C3
PDK4
ALDH1A1
DHRS9
PDK1
RDH11
CRABP1
PLAT
PDHX
PDK2
RARB
FABP5
RORA
PPARD
RXRB
RARA
RXRA
CYP26B1
CYP26A1
CYP26C1
B
RA Synthesis Module
RA Nuclear module
RA Degradation Module
C
Thymus
1 cTEC
2 mcTEC
3 mTECI
4 mTECII/III/IV
5 mTECneuro
6 mTECmyo
7 DN-Thymocyte
8 DP-Thymocyte
9 Macrophage
10 Dentritic
11 Fibroblast-1
12 Fibroblast-1 cycling
13 Fibroblast-2
14 VSMC
15 Endothelial
16 Lymphatic Endothelial
17 Neuronal
RA Synthesis Module
RA Nuclear module
RA Degradation Module
D
RA Synthesis Module
RA Nuclear module
RA Degradation Module
cTEC
Dentritic
DN-Thymocyte
DP-Thymocyte
Doublets
Endothelial
Fibroblast-1
Fibroblast-1 cycling
Fibroblast-2
Lymphatic Endothelial
Macrophage
mcTEC
mTECI
mTECII/III/IV
mTECmyo
mTECneuro
Neuronal
VSMC

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

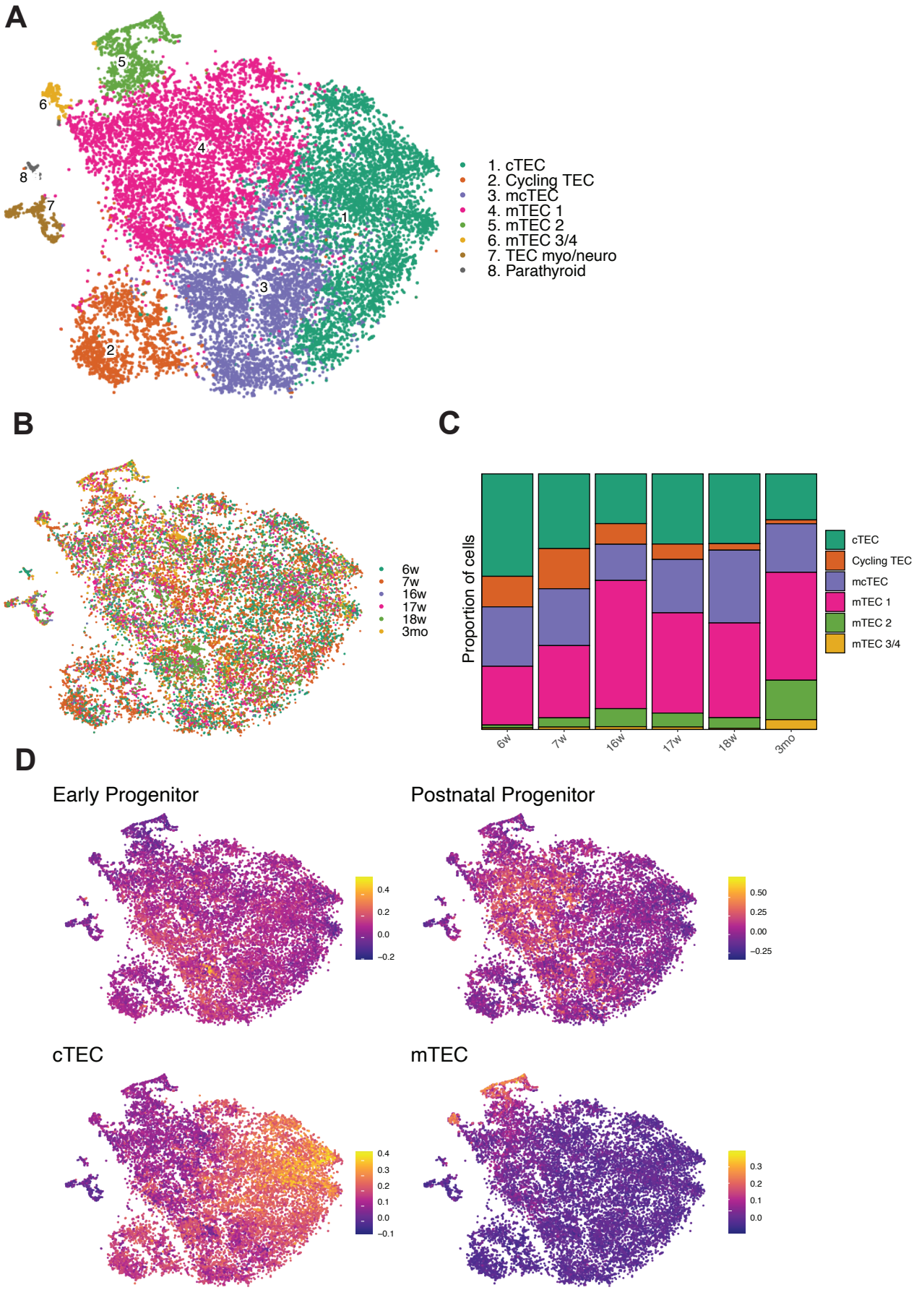

A
1. cTEC
2. Cycling TEC
3. mcTEC
4. mTEC 1
5. mTEC 2
6. mTEC 3/4
7. TEC myo/neuro
8. Parathyroid
B
6w
7w
16w
17w
18w
3mo
C
Proportion of cells
cTEC
Cycling TEC
mcTEC
mTEC 1
mTEC 2
mTEC 3/4
6w
7w
16w
17w
18w
3mo
D
Early Progenitor
Postnatal Progenitor
cTEC
mTEC

Figure 5. Vertical development of thymic epithelial cells in sequential timepoints. tSNE visualization of thymic epithelial cells across all sampled developmental time points, colored by cell types (A) or by developmental time point (B). (C) Relative proportion of select thymic epithelial cell types at each developmental time point. (D) tSNE visualization 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.

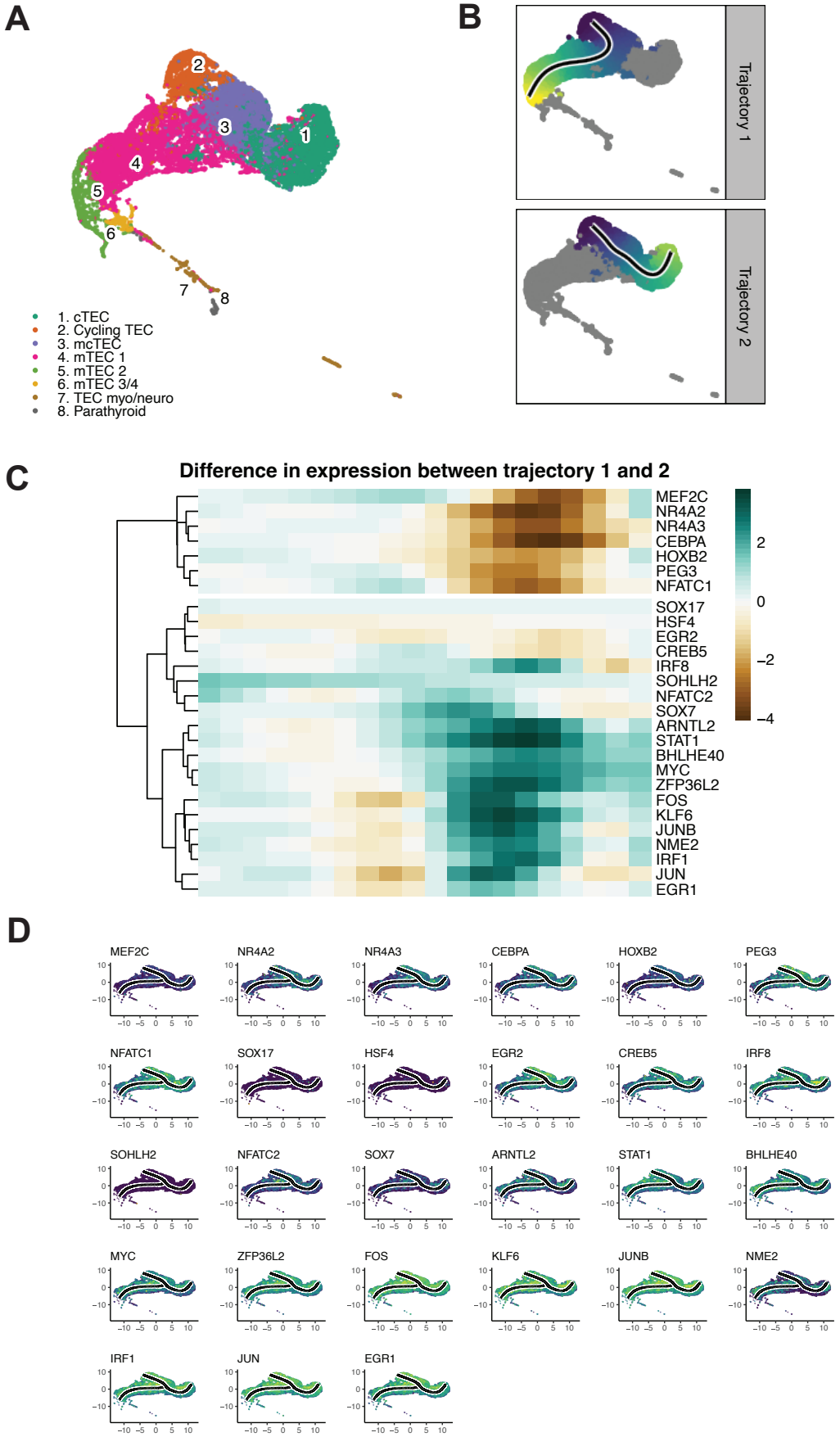
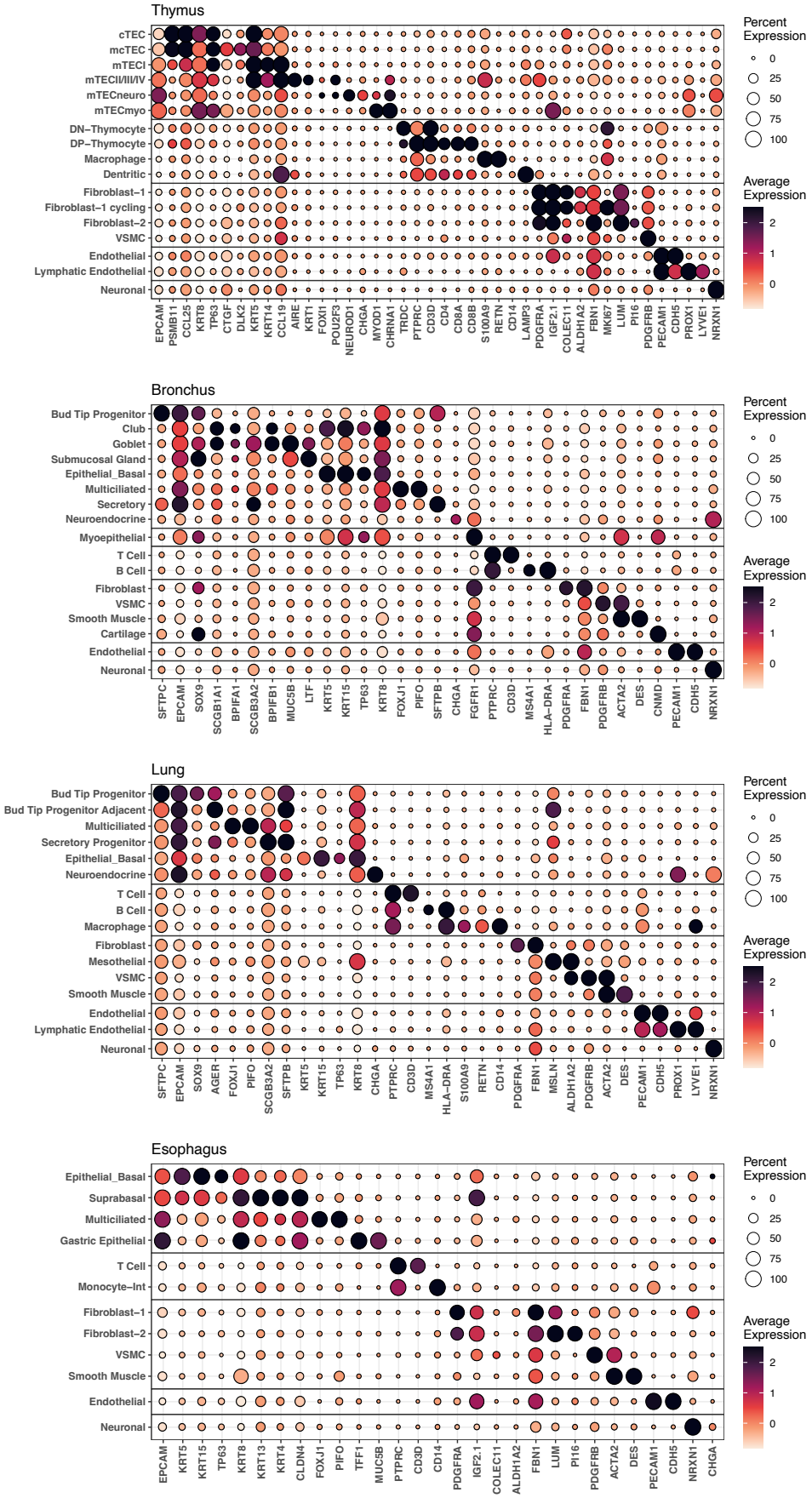

A
1. cTEC
2. Cycling TEC
3. mcTEC
4. mTEC 1
5. mTEC 2
6. mTEC 3/4
7. TEC myo/neuro
8. Parathyroid
B
Trajectory 1
Trajectory 2
C
Difference in expression between trajectory 1 and 2
MEF2C
NR4A2
NR4A3
CEBPA
HOXB2
PEG3
NFATC1
SOX17
HSF4
EGR2
CREB5
IRF8
SOHLH2
NFATC2
SOX7
ARNTL2
STAT1
BHLHE40
MYC
ZFP36L2
FOS
KLF6
JUNB
NME2
IRF1
JUN
EGR1
D

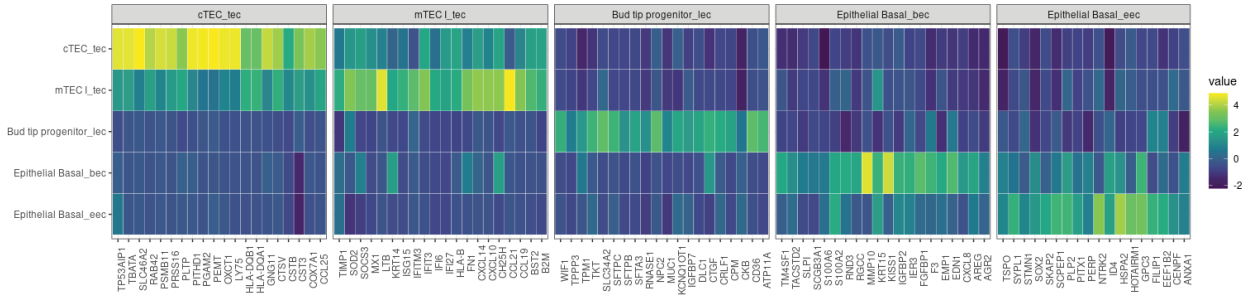
Figure 6. Development-related gene signature modulates dual trajectory of thymic epithelial cell differentiation. (A) tSpace-UMAP visualization of thymic epithelial cells across all sampled developmental time points, colored by cell types. (B) tSpace-UMAP from above overlaid with pseudotime trajectories based on underlying gene expression. (C) Heatmap of difference in select transcription factor expression between trajectories along pseudotime. Expression based on negative binomial generalized additive model (GAM) to standardize differences in trajectory lengths. Transient transcription factors selected by including only those with significant difference in expression between trajectories and an absolute expression difference between trajectories of < 1.5 at the beginning and end of pseudotime. (D) tSpace-UMAP visualization of thymic epithelial cells across all sampled developmental time points, overlaid with trajectories and expression of transiently expressed transcription factors included in above heatmap.

A


Thymus
cTEC
mcTEC
mTECI
mTECII/III/IV
mTECneuro
mTECmyo
DN−Thymocyte
DP−Thymocyte
Macrophage
Dentritic
Fibroblast−1
Fibroblast−1 cycling
Fibroblast−2
VSMC
Endothelial
Lymphatic Endothelial
Neuronal
EPCAM PSMB11 CCL25 KRT8 TP63 CTGF DLK2 KRT5 KRT14 CCL19 AIRE KRT1 FOXI1 POU2F3 NEUROD1 CHGA MYOD1 CHRNA1 TRDC PTPRC CD3D CD4 CD8A CD8B S100A9 RETN CD14 LAMP3 PDGFRA IGF2.1 COLEC11 ALDH1A2 FBN1 MKI67 LUM PI16 PDGFRB PECAM1 CDH5 PROX1 LYVE1 NRXN1
Bronchus
Bud Tip Progenitor
Club
Goblet
Submucosal Gland
Epithelial_Basal
Multiciliated
Secretory
Neuroendocrine
Myoepithelial
T Cell
B Cell
Fibroblast
VSMC
Smooth Muscle
Cartilage
Endothelial
Neuronal
SFTPC EPCAM SOX9 SCGB1A1 BPIFA1 SCGB3A2 BPIFB1 MUC5B LTF KRT5 KRT15 TP63 KRT8 FOXJ1 PIFO SFTPB CHGA FGFR1 PTPRC CD3D MS4A1 HLA−DRA PDGFRA FBN1 PDGFRB ACTA2 DES CNMD PECAM1 CDH5 NRXN1
Lung
Bud Tip Progenitor
Bud Tip Progenitor Adjacent
Multiciliated
Secretory Progenitor
Epithelial_Basal
Neuroendocrine
T Cell
B Cell
Macrophage
Fibroblast
Mesothelial
VSMC
Smooth Muscle
Endothelial
Lymphatic Endothelial
Neuronal
SFTPC EPCAM SOX9 AGER FOXJ1 PIFO SCGB3A2 SFTPB KRT5 KRT15 TP63 KRT8 CHGA PTPRC CD3D MS4A1 HLA−DRA S100A9 RETN CD14 PDGFRA FBN1 MSLN ALDH1A2 PDGFRB ACTA2 DES PECAM1 CDH5 PROX1 LYVE1 NRXN1
Esophagus
Epithelial_Basal
Suprabasal
Multiciliated
Gastric Epithelial
T Cell
Monocyte−Int
Fibroblast−1
Fibroblast−2
VSMC
Smooth Muscle
Endothelial
Neuronal
EPCAM KRT5 KRT15 TP63 KRT8 KRT13 KRT4 CLDN4 FOXJ1 PIFO TFF1 MUC5B PTPRC CD3D CD14 PDGFRA IGF2.1 COLEC11 ALDH1A2 FBN1 LUM PI16 PDGFRB ACTA2 DES PECAM1 CDH5 NRXN1 CHGA
Percent Expression
0
25
50
75
100
Average Expression
2
1
0

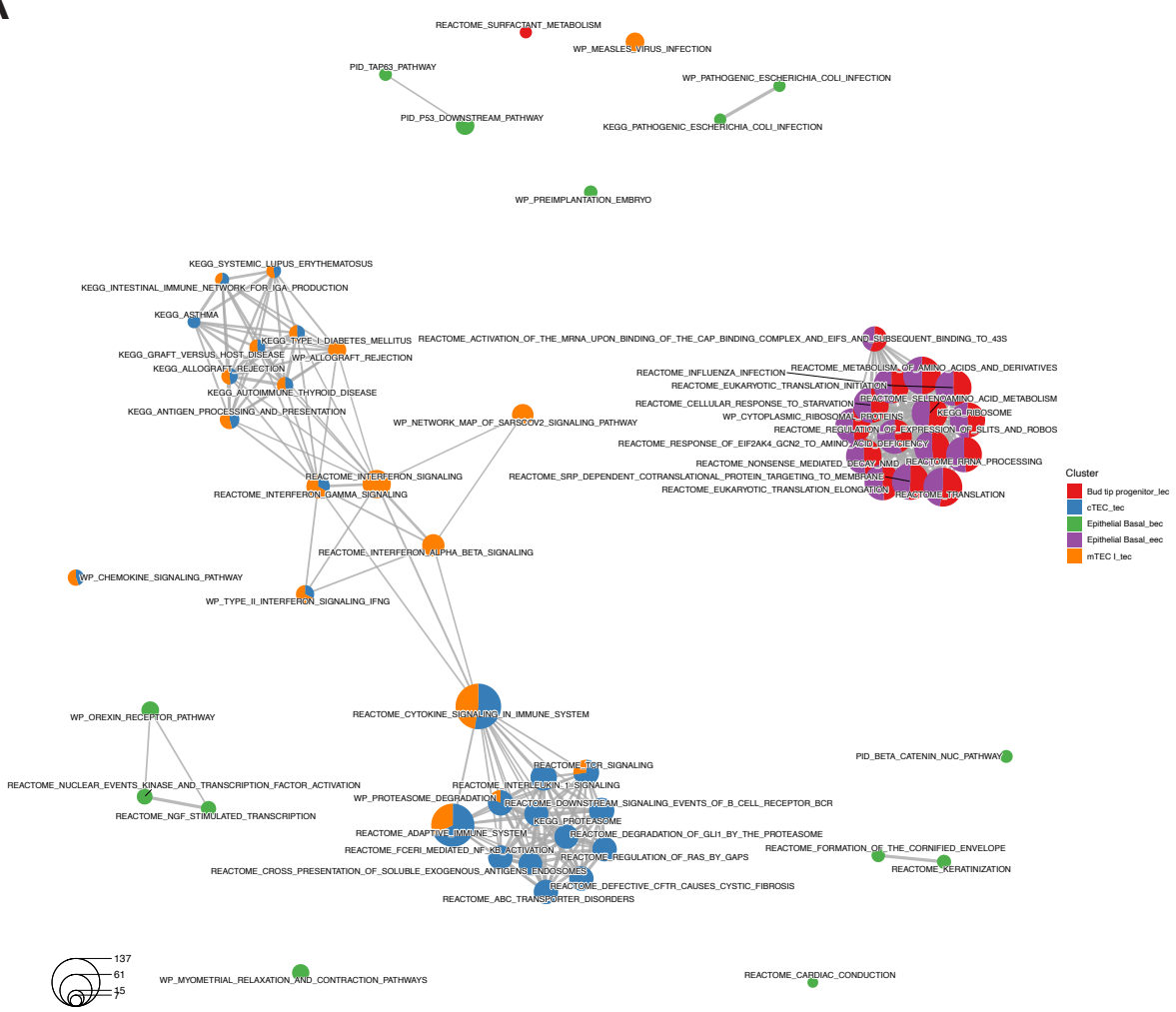
Supplementary Figure S1. Dot plot showing representative marker gene expression of anterior foregut cell types across individual tissues. Color representative of average expression level. Dot size representative of percentage of cells in cell type expressing selected marker gene.

A

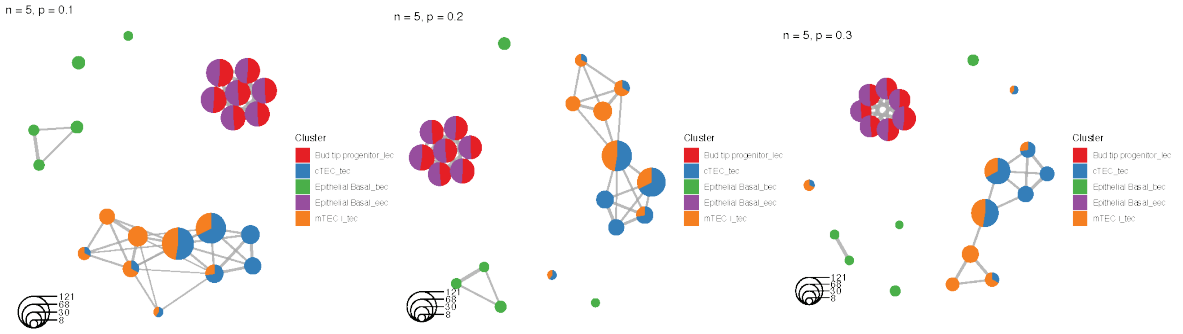

cTEC_tec
mTEC I_tec
Bud tip progenitor_lec
Epithelial Basal_bec
Epithelial Basal_eec
value
4
2
0
-2
TP53AIP1 TBATA SLC46A2 RAB42 PSMB11 PRSS16 PLTP PITHD1 PGAM2 PEMT OXCT1 LY75 HLA-DQB1 HLA-DQA1 GNG11 CTSV CSTB CST3 COX7A1 CCL25
TIMP1 SOD2 SOCS3 MX1 LTB KRT14 ISG15 IFITM3 IFIT3 IFI6 IFI27 HLA-B FN1 CXCL14 CXCL10 CH25H CCL21 CCL19 BST2 B2M
WIF1 TPPP3 TPM1 TKT SLC34A2 SFTPC SFTPB SFTA3 RNASE1 NPC2 MUC1 KCNQ1OT1 IGFBP7 DLC1 CTGF CRLF1 CPM CKB CD36 ATP11A
TM4SF1 TACSTD2 SLPI SCGB3A1 S100A6 S100A2 RND3 RGCC MMP10 KRT15 KISS1 IGFBP2 IER3 FGFBP1 F3 EMP1 EDN1 CXCL8 AREG AGR2
TSPO SYPL1 STMN1 SOX2 SKAP2 SCPEP1 PLP2 PITX1 PERP NTRK2 ID4 HSPA2 HOTAIRM1 GPC3 FILIP1 EEF1B2 CENPF ANXA1

Supplementary Figure S2. Heatmap of selected genes across thymic epithelial cells.

A

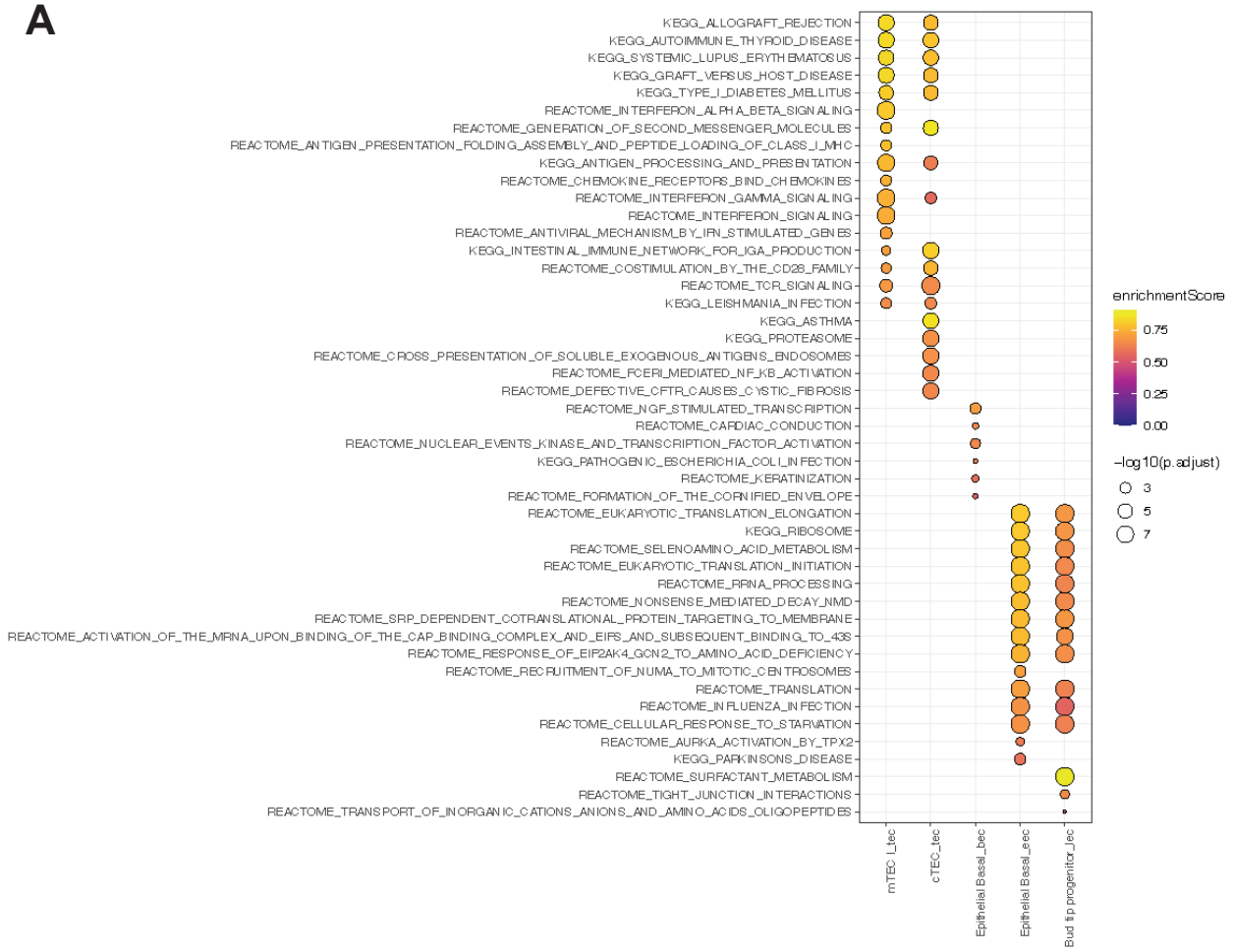

REACTOME_SURFACTANT_METABOLISM
WP_MEASLES_VIRUS_INFECTION
PID_TAP63_PATHWAY
WP_PATHOGENIC_ESCHERICHIA_COLI_INFECTION
PID_P53_DOWNSTREAM_PATHWAY
KEGG_PATHOGENIC_ESCHERICHIA_COLI_INFECTION
WP_PREIMPLANTATION_EMBRYO
KEGG_SYSTEMIC_LUPUS_ERYTHEMATOSUS
KEGG_INTESTINAL_IMMUNE_NETWORK_FOR_IGA_PRODUCTION
KEGG_ASTHMA
KEGG_TYPE_I_DIABETES_MELLITUS
REACTOME_ACTIVATION_OF_THE_MRNA_UPON_BINDING_OF_THE_CAP_BINDING_COMPLEX_AND_EIFS_AND_SUBSEQUENT_BINDING_TO_43S
KEGG_GRAFT_VERSUS_HOST_DISEASE
WP_ALLOGRAFT_REJECTION
KEGG_ALLOGRAFT_REJECTION
REACTOME_INFLUENZA_INFECTION
REACTOME_METABOLISM_OF_AMINO_ACIDS_AND_DERIVATIVES
REACTOME_EUKARYOTIC_TRANSLATION_INITIATION
KEGG_AUTOIMMUNE_THYROID_DISEASE
REACTOME_CELLULAR_RESPONSE_TO_STARVATION
REACTOME_SELENOAMINO_ACID_METABOLISM
KEGG_ANTIGEN_PROCESSING_AND_PRESENTATION
WP_CYTOPLASMIC_RIBOSOMAL_PROTEINS
KEGG_RIBOSOME
WP_NETWORK_MAP_OF_SARSCOV2_SIGNALING_PATHWAY
REACTOME_REGULATION_OF_EXPRESSION_OF_SLITS_AND_ROBOS
REACTOME_RESPONSE_OF_EIF2AK4_GCN2_TO_AMINO_ACID_DEFICIENCY
REACTOME_NONSENSE_MEDIATED_DECAY_NMD
REACTOME_RRNA_PROCESSING
REACTOME_INTERFERON_SIGNALING
REACTOME_SRP_DEPENDENT_COTRANSLATIONAL_PROTEIN_TARGETING_TO_MEMBRANE
REACTOME_INTERFERON_GAMMA_SIGNALING
REACTOME_EUKARYOTIC_TRANSLATION_ELONGATION
REACTOME_TRANSLATION
Cluster
Bud tip progenitor_lec
cTEC_tec
Epithelial Basal_bec
Epithelial Basal_eec
mTEC I_tec
REACTOME_INTERFERON_ALPHA_BETA_SIGNALING
WP_CHEMOKINE_SIGNALING_PATHWAY
WP_TYPE_II_INTERFERON_SIGNALING_IFNG
WP_OREXIN_RECEPTOR_PATHWAY
REACTOME_CYTOKINE_SIGNALING_IN_IMMUNE_SYSTEM
PID_BETA_CATENIN_NUC_PATHWAY
REACTOME_TCR_SIGNALING
REACTOME_NUCLEAR_EVENTS_KINASE_AND_TRANSCRIPTION_FACTOR_ACTIVATION
REACTOME_INTERLEUKIN_1_SIGNALING
WP_PROTEASOME_DEGRADATION
REACTOME_DOWNSTREAM_SIGNALING_EVENTS_OF_B_CELL_RECEPTOR_BCR
REACTOME_NGF_STIMULATED_TRANSCRIPTION
KEGG_PROTEASOME
REACTOME_ADAPTIVE_IMMUNE_SYSTEM
REACTOME_DEGRADATION_OF_GLI1_BY_THE_PROTEASOME
REACTOME_FCERI_MEDIATED_NF_KB_ACTIVATION
REACTOME_FORMATION_OF_THE_CORNIFIED_ENVELOPE
REACTOME_REGULATION_OF_RAS_BY_GAPS
REACTOME_KERATINIZATION
REACTOME_CROSS_PRESENTATION_OF_SOLUBLE_EXOGENOUS_ANTIGENS_ENDOSOMES
REACTOME_DEFECTIVE_CFTR_CAUSES_CYSTIC_FIBROSIS
REACTOME_ABC_TRANSPORTER_DISORDERS
137
61
45
7
WP_MYOMETRIAL_RELAXATION_AND_CONTRACTION_PATHWAYS
REACTOME_CARDIAC_CONDUCTION

B


n = 5, p = 0.1
n = 5, p = 0.2
n = 5, p = 0.3
Cluster
Bud tip progenitor_lec
cTEC_tec
Epithelial Basal_bec
Epithelial Basal_eec
mTEC I_tec
121
68
30
8

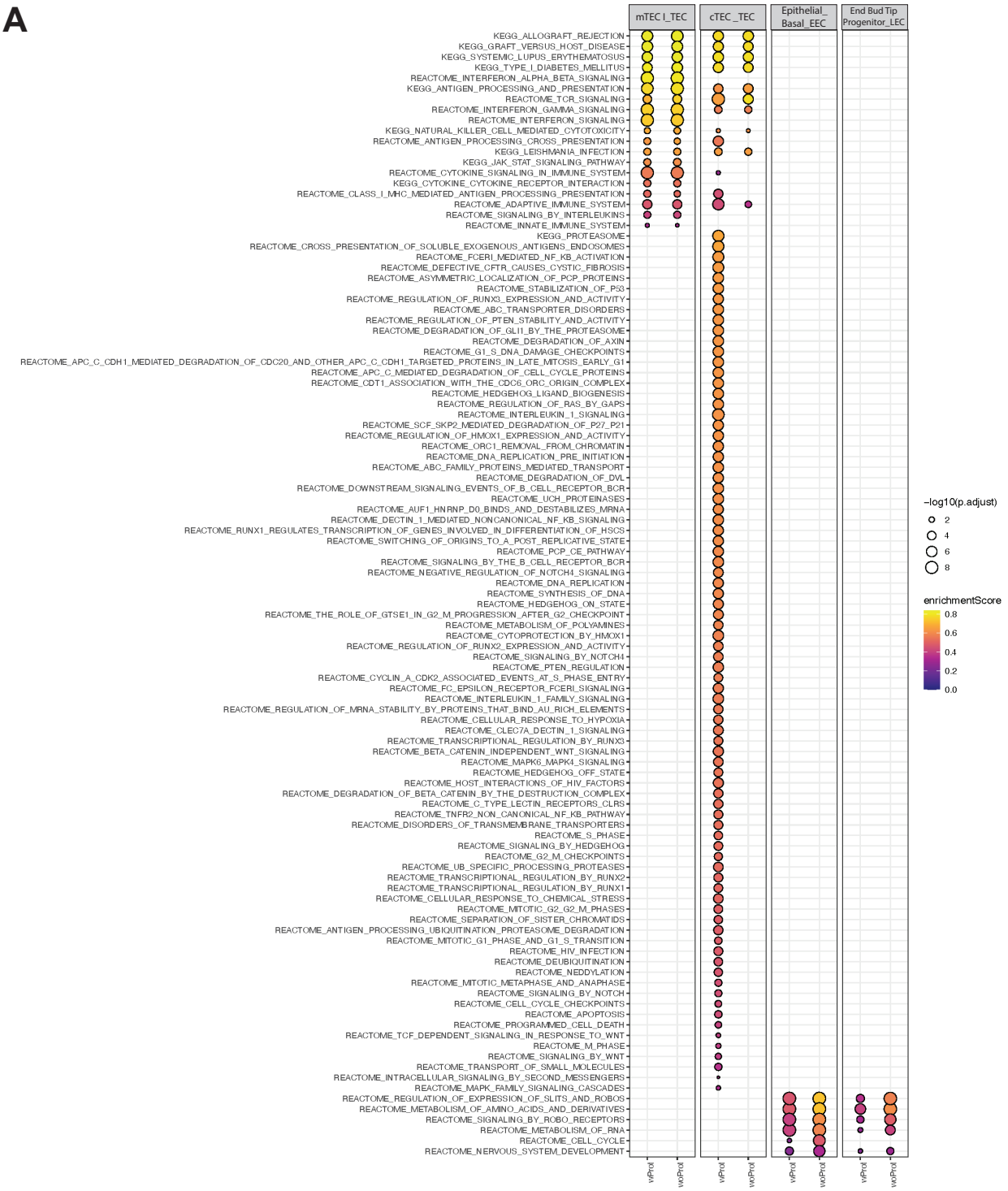
Supplementary Figure S3. (A) Network visualization of positively enriched gene sets across select anterior foregut epithelial cells. Edges weight reflects proportion of genes shared between gene sets with a minimum proportion of 0.2. All genes are represented. **(B)** Network visualization of positively enriched gene sets across select anterior foregut epithelial cells. Nodes represent individual gene sets. Node color represents relative use of gene set genes by each cell type. Node size reflects number of genes in gene set. Edges weight reflects proportion of genes shared between gene sets. N = number of pathways included for each cell type. P = minimum proportion of gene set overlap required for an edge to be drawn between them.

A

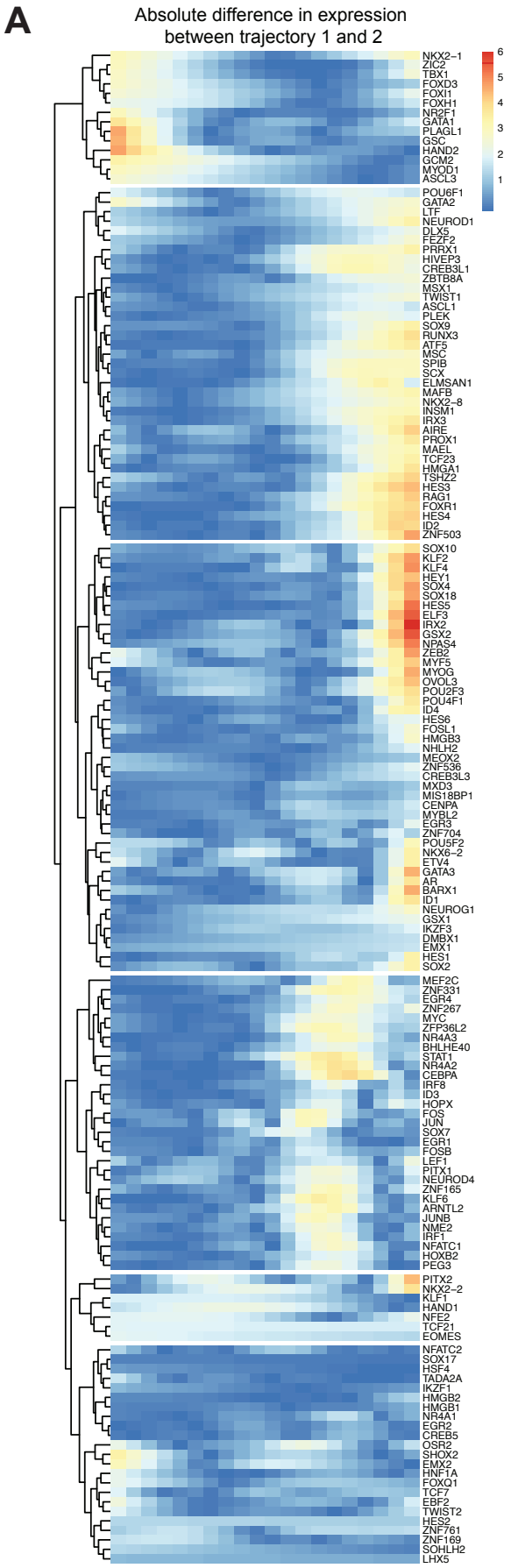

KEGG_ALLOGRAFT_REJECTION
KEGG_AUTOIMMUNE_THYROID_DISEASE
KEGG_SYSTEMIC_LUPUS_ERYTHEMATOSUS
KEGG_GRAFT_VERSUS_HOST_DISEASE
KEGG_TYPE_I_DIABETES_MELLITUS
REACTOME_INTERFERON_ALPHA_BETA_SIGNALING
REACTOME_GENERATION_OF_SECOND_MESSENGER_MOLECULES
REACTOME_ANTIGEN_PRESENTATION_FOLDING_ASSEMBLY_AND_PEPTIDE_LOADING_OF_CLASS_I_MHC
KEGG_ANTIGEN_PROCESSING_AND_PRESENTATION
REACTOME_CHEMOKINE_RECEPTORS_BIND_CHEMOKINES
REACTOME_INTERFERON_GAMMA_SIGNALING
REACTOME_INTERFERON_SIGNALING
REACTOME_ANTIVIRAL_MECHANISM_BY_IFN_STIMULATED_GENES
KEGG_INTESTINAL_IMMUNE_NETWORK_FOR_IGA_PRODUCTION
REACTOME_COSTIMULATION_BY_THE_CD28_FAMILY
REACTOME_TCR_SIGNALING
KEGG_LEISHMANIA_INFECTION
KEGG_ASTHMA
KEGG_PROTEASOME
REACTOME_CROSS_PRESENTATION_OF_SOLUBLE_EXOGENOUS_ANTIGENS_ENDOSOMES
REACTOME_FCERI_MEDIATED_NF_KB_ACTIVATION
REACTOME_DEFECTIVE_CFTR_CAUSES_CYSTIC_FIBROSIS
REACTOME_NGF_STIMULATED_TRANSCRIPTION
REACTOME_CARDIAC_CONDUCTION
REACTOME_NUCLEAR_EVENTS_KINASE_AND_TRANSCRIPTION_FACTOR_ACTIVATION
KEGG_PATHOGENIC_ESCHERICHIA_COLI_INFECTION
REACTOME_KERATINIZATION
REACTOME_FORMATION_OF_THE_CORNIFIED_ENVELOPE
REACTOME_EUKARYOTIC_TRANSLATION_ELONGATION
KEGG_RIBOSOME
REACTOME_SELENOAMINO_ACID_METABOLISM
REACTOME_EUKARYOTIC_TRANSLATION_INITIATION
REACTOME_RRNA_PROCESSING
REACTOME_NONSENSE_MEDIATED_DECAY_NMD
REACTOME_SRP_DEPENDENT_COTRANSLATIONAL_PROTEIN_TARGETING_TO_MEMBRANE
REACTOME_ACTIVATION_OF_THE_MRNA_UPON_BINDING_OF_THE_CAP_BINDING_COMPLEX_AND_EIFS_AND_SUBSEQUENT_BINDING_TO_43S
REACTOME_RESPONSE_OF_EIF2AK4_GCN2_TO_AMINO_ACID_DEFICIENCY
REACTOME_RECRUITMENT_OF_NUMA_TO_MITOTIC_CENTROSOMES
REACTOME_TRANSLATION
REACTOME_INFLUENZA_INFECTION
REACTOME_CELLULAR_RESPONSE_TO_STARVATION
REACTOME_AURKA_ACTIVATION_BY_TPX2
KEGG_PARKINSONS_DISEASE
REACTOME_SURFACTANT_METABOLISM
REACTOME_TIGHT_JUNCTION_INTERACTIONS
REACTOME_TRANSPORT_OF_INORGANIC_CATIONS_ANIONS_AND_AMINO_ACIDS_OLIGOPEPTIDES
mTEC_Lec
cTEC_lec
Epithelial Basal_bec
Epithelial Basal_eec
Bud tip progenitor_lec
enrichmentScore
0.75
0.50
0.25
0.00
−log10(p.adjust)
3
5
7

Supplementary Figure S4. Dot plots of gene set enrichment across selected anterior foregut epithelial cell types. Only selected gene sets are included. Color representative of gene set enrichment score. Dot size representative of percentage of -log10 p-value. All gene sets are included

A


mTEC I_TEC
cTEC _TEC
Epithelial_ Basal_EEC
End Bud Tip Progenitor_LEC
KEGG_ALLOGRAFT_REJECTION
KEGG_GRAFT_VERSUS_HOST_DISEASE
KEGG_SYSTEMIC_LUPUS_ERYTHEMATOSUS
KEGG_TYPE_I_DIABETES_MELLITUS
REACTOME_INTERFERON_ALPHA_BETA_SIGNALING
KEGG_ANTIGEN_PROCESSING_AND_PRESENTATION
REACTOME_TCR_SIGNALING
REACTOME_INTERFERON_GAMMA_SIGNALING
REACTOME_INTERFERON_SIGNALING
KEGG_NATURAL_KILLER_CELL_MEDIATED_CYTOTOXICITY
REACTOME_ANTIGEN_PROCESSING_CROSS_PRESENTATION
KEGG_LEISHMANIA_INFECTION
KEGG_JAK_STAT_SIGNALING_PATHWAY
REACTOME_CYTOKINE_SIGNALING_IN_IMMUNE_SYSTEM
KEGG_CYTOKINE_CYTOKINE_RECEPTOR_INTERACTION
REACTOME_CLASS_I_MHC_MEDIATED_ANTIGEN_PROCESSING_PRESENTATION
REACTOME_ADAPTIVE_IMMUNE_SYSTEM
REACTOME_SIGNALING_BY_INTERLEUKINS
REACTOME_INNATE_IMMUNE_SYSTEM
KEGG_PROTEASOME
REACTOME_CROSS_PRESENTATION_OF_SOLUBLE_EXOGENOUS_ANTIGENS_ENDOSOMES
REACTOME_FCERI_MEDIATED_NF_KB_ACTIVATION
REACTOME_DEFECTIVE_CFTR_CAUSES_CYSTIC_FIBROSIS
REACTOME_ASYMMETRIC_LOCALIZATION_OF_PCP_PROTEINS
REACTOME_STABILIZATION_OF_P53
REACTOME_REGULATION_OF_RUNX3_EXPRESSION_AND_ACTIVITY
REACTOME_ABC_TRANSPORTER_DISORDERS
REACTOME_REGULATION_OF_PTEN_STABILITY_AND_ACTIVITY
REACTOME_DEGRADATION_OF_GLI1_BY_THE_PROTEASOME
REACTOME_DEGRADATION_OF_AXIN
REACTOME_G1_S_DNA_DAMAGE_CHECKPOINTS
REACTOME_APC_C_CDH1_MEDIATED_DEGRADATION_OF_CDC20_AND_OTHER_APC_C_CDH1_TARGETED_PROTEINS_IN_LATE_MITOSIS_EARLY_G1
REACTOME_APC_C_MEDIATED_DEGRADATION_OF_CELL_CYCLE_PROTEINS
REACTOME_CDT1_ASSOCIATION_WITH_THE_CDC6_ORC_ORIGIN_COMPLEX
REACTOME_HEDGEHOG_LIGAND_BIOGENESIS
REACTOME_REGULATION_OF_RAS_BY_GAPS
REACTOME_INTERLEUKIN_1_SIGNALING
REACTOME_SCF_SKP2_MEDIATED_DEGRADATION_OF_P27_P21
REACTOME_REGULATION_OF_HMOX1_EXPRESSION_AND_ACTIVITY
REACTOME_ORC1_REMOVAL_FROM_CHROMATIN
REACTOME_DNA_REPLICATION_PRE_INITIATION
REACTOME_ABC_FAMILY_PROTEINS_MEDIATED_TRANSPORT
REACTOME_DEGRADATION_OF_DVL
REACTOME_DOWNSTREAM_SIGNALING_EVENTS_OF_B_CELL_RECEPTOR_BCR
REACTOME_UCH_PROTEINASES
REACTOME_AUF1_HNRNP_D0_BINDS_AND_DESTABILIZES_MRNA
REACTOME_DECTIN_1_MEDIATED_NONCANONICAL_NF_KB_SIGNALING
REACTOME_RUNX1_REGULATES_TRANSCRIPTION_OF_GENES_INVOLVED_IN_DIFFERENTIATION_OF_HSCS
REACTOME_SWITCHING_OF_ORIGINS_TO_A_POST_REPLICATIVE_STATE
REACTOME_PCP_CE_PATHWAY
REACTOME_SIGNALING_BY_THE_B_CELL_RECEPTOR_BCR
REACTOME_NEGATIVE_REGULATION_OF_NOTCH4_SIGNALING
REACTOME_DNA_REPLICATION
REACTOME_SYNTHESIS_OF_DNA
REACTOME_HEDGEHOG_ON_STATE
REACTOME_THE_ROLE_OF_GTSE1_IN_G2_M_PROGRESSION_AFTER_G2_CHECKPOINT
REACTOME_METABOLISM_OF_POLYAMINES
REACTOME_CYTOPROTECTION_BY_HMOX1
REACTOME_REGULATION_OF_RUNX2_EXPRESSION_AND_ACTIVITY
REACTOME_SIGNALING_BY_NOTCH4
REACTOME_PTEN_REGULATION
REACTOME_CYCLIN_A_CDK2_ASSOCIATED_EVENTS_AT_S_PHASE_ENTRY
REACTOME_FC_EPSILON_RECEPTOR_FCERI_SIGNALING
REACTOME_INTERLEUKIN_1_FAMILY_SIGNALING
REACTOME_REGULATION_OF_MRNA_STABILITY_BY_PROTEINS_THAT_BIND_AU_RICH_ELEMENTS
REACTOME_CELLULAR_RESPONSE_TO_HYPOXIA
REACTOME_CLEC7A_DECTIN_1_SIGNALING
REACTOME_TRANSCRIPTIONAL_REGULATION_BY_RUNX3
REACTOME_BETA_CATENIN_INDEPENDENT_WNT_SIGNALING
REACTOME_MAPK6_MAPK4_SIGNALING
REACTOME_HEDGEHOG_OFF_STATE
REACTOME_HOST_INTERACTIONS_OF_HIV_FACTORS
REACTOME_DEGRADATION_OF_BETA_CATENIN_BY_THE_DESTRUCTION_COMPLEX
REACTOME_C_TYPE_LECTIN_RECEPTORS_CLRS
REACTOME_TNFR2_NON_CANONICAL_NF_KB_PATHWAY
REACTOME_DISORDERS_OF_TRANSMEMBRANE_TRANSPORTERS
REACTOME_S_PHASE
REACTOME_SIGNALING_BY_HEDGEHOG
REACTOME_G2_M_CHECKPOINTS
REACTOME_UB_SPECIFIC_PROCESSING_PROTEASES
REACTOME_TRANSCRIPTIONAL_REGULATION_BY_RUNX2
REACTOME_TRANSCRIPTIONAL_REGULATION_BY_RUNX1
REACTOME_CELLULAR_RESPONSE_TO_CHEMICAL_STRESS
REACTOME_MITOTIC_G2_G2_M_PHASES
REACTOME_SEPARATION_OF_SISTER_CHROMATIDS
REACTOME_ANTIGEN_PROCESSING_UBIQUITINATION_PROTEASOME_DEGRADATION
REACTOME_MITOTIC_G1_PHASE_AND_G1_S_TRANSITION
REACTOME_HIV_INFECTION
REACTOME_DEUBIQUITINATION
REACTOME_NEDDYLATION
REACTOME_MITOTIC_METAPHASE_AND_ANAPHASE
REACTOME_SIGNALING_BY_NOTCH
REACTOME_CELL_CYCLE_CHECKPOINTS
REACTOME_APOPTOSIS
REACTOME_PROGRAMMED_CELL_DEATH
REACTOME_TCF_DEPENDENT_SIGNALING_IN_RESPONSE_TO_WNT
REACTOME_M_PHASE
REACTOME_SIGNALING_BY_WNT
REACTOME_TRANSPORT_OF_SMALL_MOLECULES
REACTOME_INTRACELLULAR_SIGNALING_BY_SECOND_MESSENGERS
REACTOME_MAPK_FAMILY_SIGNALING_CASCADES
REACTOME_REGULATION_OF_EXPRESSION_OF_SLITS_AND_ROBOS
REACTOME_METABOLISM_OF_AMINO_ACIDS_AND_DERIVATIVES
REACTOME_SIGNALING_BY_ROBO_RECEPTORS
REACTOME_METABOLISM_OF_RNA
REACTOME_CELL_CYCLE
REACTOME_NERVOUS_SYSTEM_DEVELOPMENT
vPrd
woPrd
vPrd
woPrd
vPrd
woPrd
vPrd
woPrd
−log10(p.adjust)
2
4
6
8
enrichmentScore
0.8
0.6
0.4
0.2
0.0

Supplementary Figure S5. Dot plots of comparing gene set enrichment of original genes sets and gene sets with proteasomal genes removed across selected anterior foregut epithelial cell types. Color representative of gene set enrichment score. Dot size representative of percentage of -log10 p-value. All gene sets are included.


A
Absolute difference in expression
between trajectory 1 and 2
6
5
4
3
2
1
NKX2−1
ZIC2
TBX1
FOXD3
FOXI1
FOXH1
NR2F1
GATA1
PLAGL1
GSC
HAND2
GCM2
MYOD1
ASCL3
POU6F1
GATA2
LTF
NEUROD1
DLX5
FEZF2
PRRX1
HIVEP3
CREB3L1
ZBTB8A
MSX1
TWIST1
ASCL1
PLEK
SOX9
RUNX3
ATF5
MSC
SPIB
SCX
ELMSAN1
MAFB
NKX2−8
INSM1
IRX3
AIRE
PROX1
MAEL
TCF23
HMGA1
TSHZ2
HES3
RAG1
FOXR1
HES4
ID2
ZNF503
SOX10
KLF2
KLF4
HEY1
SOX4
SOX18
HES5
ELF3
IRX2
GSX2
NPAS4
ZEB2
MYF5
MYOG
OVOL3
POU2F3
POU4F1
ID4
HES6
FOSL1
HMGB3
NHLH2
MEOX2
ZNF536
CREB3L3
MXD3
MIS18BP1
CENPA
MYBL2
EGR3
ZNF704
POU5F2
NKX6−2
ETV4
GATA3
AR
BARX1
ID1
NEUROG1
GSX1
IKZF3
DMBX1
EMX1
HES1
SOX2
MEF2C
ZNF331
EGR4
ZNF267
MYC
ZFP36L2
NR4A3
BHLHE40
STAT1
NR4A2
CEBPA
IRF8
ID3
HOPX
FOS
JUN
SOX7
EGR1
FOSB
LEF1
PITX1
NEUROD4
ZNF165
KLF6
ARNTL2
JUNB
NME2
IRF1
NFATC1
HOXB2
PEG3
PITX2
NKX2−2
KLF1
HAND1
NFE2
TCF21
EOMES
NFATC2
SOX17
HSF4
TADA2A
IKZF1
HMGB2
HMGB1
NR4A1
EGR2
CREB5
OSR2
SHOX2
EMX2
HNF1A
FOXQ1
TCF7
EBF2
TWIST2
HES2
ZNF761
ZNF169
SOHLH2
LHX5

Supplementary Figure S6. Heatmap of absolute difference in select transcription factor expression between trajectories along pseudotime. Expression based on negative binomial generalized additive model (GAM) to standardize differences in trajectory lengths. Transcription factors selected by including only those with significant difference in expression between trajectories.

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