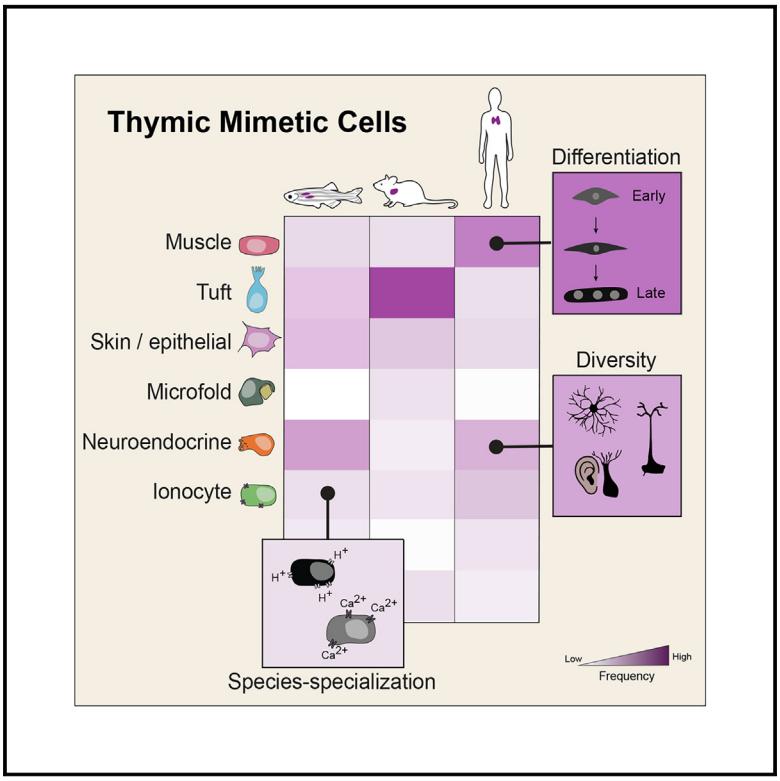
Article

Immunity

Cross-species analyses of thymic mimetic cells reveal evolutionarily ancient origins and both conserved and species-specific elements

Graphical abstract



Highlights

- Human thymic mimetic cells mirror a diversity of peripheral cell types
- Human mimetic cells include a massive expansion of muscle mTECs vis-à-vis mice
- Imaging reveals striated muscle mTECs and groupings akin to neuromuscular junctions
- Zebrafish mimetic cells include conserved and fish-specific subtypes

Authors

Brooke D. Huisman, Daniel A. Michelson, Sara A. Rubin, ..., Christophe Benoist, Leonard Zon, Diane Mathis

Correspondence

dm@hms.harvard.edu

In brief

Thymic mimetic cells are chimeras between medullary-thymic-epithelial cells (mTECs) and peripheral cell types. Huisman, Michelson, et al. perform high-resolution profiling of human mimetics, revealing a diversification of nerve and ionocyte mTECs and a massive expansion of muscle mTECs vis-à-vis mice. Comparison of human, mouse, and zebrafish mimetics reveals species-specific and species-conserved subtypes.

Huisman et al., 2025, Immunity *58*, 108–123
January 14, 2025
https://doi.org/10.1016/j.immuni.2024.11.025

CellPress

CellPress

Immunity

Article

Cross-species analyses of thymic mimetic cells reveal evolutionarily ancient origins and both conserved and species-specific elements

Brooke D. Huisman,[1,9] Daniel A. Michelson,[1,2,3,9] Sara A. Rubin,[2,3,4,5] Katherine Kohlsaat,[6] Wilson Gomarga,[4,5] Yuan Fang,[1] Ji Myung Lee,[6] Pedro del Nido,[6,7] Meena Nathan,[6,7] Christophe Benoist,[1] Leonard Zon,[4,5,8] and Diane Mathis[1,10,*]

[1]Department of Immunology, Harvard Medical School, Boston, MA, USA
[2]Harvard-MIT Program in Health Sciences and Technology, Harvard Medical School, Boston, MA, USA
[3]PhD Program in Immunology, Harvard Medical School, Boston, MA, USA
[4]Stem Cell Program and Division of Hematology/Oncology, Boston Children's Hospital and Dana-Farber Cancer Institute, Boston, MA, USA
[5]Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA, USA
[6]Department of Cardiac Surgery, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA
[7]Department of Surgery, Harvard Medical School, Boston, MA, USA
[8]Howard Hughes Medical Institute and Boston Children's Hospital, Boston, MA, USA
[9]These authors contributed equally
[10]Lead contact
*Correspondence: dm@hms.harvard.edu
https://doi.org/10.1016/j.immuni.2024.11.025

SUMMARY

Thymic mimetic cells are molecular hybrids between medullary-thymic-epithelial cells (mTECs) and diverse peripheral cell types. They are involved in eliminating autoreactive T cells and can perform supplementary functions reflective of their peripheral-cell counterparts. Current knowledge about mimetic cells derives largely from mouse models. To provide the high resolution that proved revelatory for mice, we performed single-cell RNA sequencing on purified mimetic-cell compartments from human pediatric donors. The single-cell profiles of individual donors were surprisingly similar, with diversification of neuroendocrine subtypes and expansion of the muscle subtype relative to mice. Informatic and imaging studies on the muscle-mTEC population highlighted a maturation trajectory suggestive of skeletal-muscle differentiation, some striated structures, and occasional cellular groupings reminiscent of neuromuscular junctions. We also profiled thymic mimetic cells from zebrafish. Integration of data from the three species identified species-specific adaptations but substantial interspecies conservation, highlighting the evolutionarily ancient nature of mimetic mTECs. Our findings provide a landscape view of human mimetic cells, with anticipated relevance in autoimmunity.

INTRODUCTION

Discrimination between self and non-self is a fundamental requirement of the immune system. Failure to identify and target non-self results in infection or cancer, while failure to identify and protect self can result in autoimmunity. To limit self-reactivity, the T cell repertoire is pruned in the thymus. Maturing T cells are screened against a broad representation of self-antigens, including peripheral-tissue antigens (PTAs) expressed by medullary-thymic-epithelial cells (mTECs), resulting in the deletion of self-reactive T cells or their diversion into the regulatory T (Treg)-cell lineage (reviewed in Abramson and Anderson[1]). One mechanism by which thymic-epithelial cells express PTA genes is through quasi-random, Aire-mediated augmentation of transcription.[2,3] A second mechanism has only more recently been appreciated: subsets of mTECs, termed "thymic mimetic cells," mimic the transcriptional programs of particular extra-thymic cell types.[4–8] Cells of this nature, seemingly "misplaced" in the thymus, were reported as early as the mid-1800s, with the observation of cornified epithelial cells on thymic sections.[9] These and other ectopic cell types underlay one of the early explanations for the widespread expression of PTAs in the thymus.[10] Single-cell RNA sequencing (scRNA-seq) data reawakened interest in misplaced thymic stromal cells—first, thymic tuft cells and later mTEC subtypes resembling neurons, ciliated cells, muscle, ionocytes, or keratinocytes.[11–16] These cells were conceptually unified and collectively categorized as thymic mimetic cells following the discovery that murine mTECs co-opt lineage-defining transcription factors (TFs) to drive their differentiation.[4] Mimetic mTECs coordinate tolerance induction

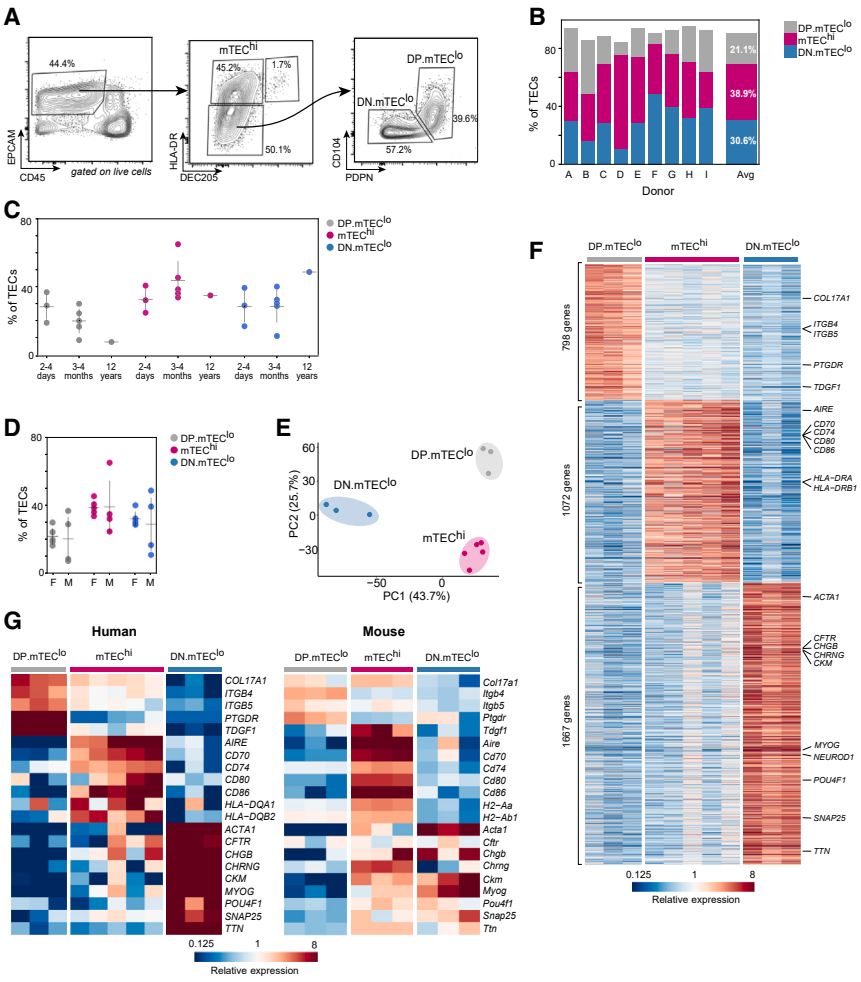


Figure 1. The human thymus contains a mimetic-cell compartment analogous to mouse

(A) Representative gating of human thymic-epithelial cells into mTEC^{hi} and mTEC^{lo} populations and subdivision of mTEC^{lo} cells into CD104 and PDPN double positive (DP.mTEC^{lo}) or double negative (DN.mTEC^{lo}) populations.

(B–D) Relative abundance of the mTEC populations among human TECs by flow cytometry, (B) by donor, with average indicated, or stratified by (C) age and (D) sex. (C and D) Mean and standard deviation are shown in black. There was no significant difference in the frequency of each compartment between 2- to 4-day-old and 3- to 4-month-old donors or between male and female donors by two-sided Wilcoxon rank sum test without multiple hypothesis testing correction and a p value cutoff of 0.05.

(E) Principal-component analysis of human population-level RNA-seq data on differential genes.

(F) Heatmap of differentially expressed genes (FC $\geq$ 2 and FDR < 0.05) across human thymus compartments in population RNA-seq data.

(G) Heatmap of selected genes in the human or mouse thymus, by population RNA-seq.

See also Figure S1 and Table S1.

RESULTS

Humans and mice harbor analogous post-Aire mTEC compartments

To investigate mTECs from the human thymus, we examined cells from de-identified pediatric thymi removed during the normal course of corrective cardiothoracic surgery. The nine donors ranged in age from 2 days to 12 years (Table S1). All nine presented with a cardiac defect, and donors A and B also had genetic syndromes (Down and Noonan syndrome, respectively).

In wild-type mice, three major mTEC compartments can be discerned by flow cytometry[4,19]: mTEC^{hi}, expressing high levels of major histocompatibility complex class II (MHC class II) molecules and encompassing the Aire-expressing cells; DP.mTEC^{lo}, which are pre-Aire-expressing cells that are Pdpn^{+}CD104^{+} MHC class II^{lo}; and DN.mTEC^{lo}, which are post-Aire-expressing cells with low levels of MHC class II molecules and no PDPN or CD104. To determine whether these compartments exist in humans and whether human post-Aire mTECs are similarly enriched with mimetic cells, we examined the analogous compartments in human thymi, enriching for TECs via bead-based CD45^{+}-cell depletion or EpCAM^{+}-cell enrichment (Figure 1A; Figure S1A). The thymi from all nine donors hosted all three major mTEC compartments at levels similar to those found in mice[4]; in particular, post-Aire mTECs comprised an average of around 30% of total TECs in both species (Figure 1B). The various donors showed no significant age- or sex-related differences in the frequencies of the three compartments (Figures 1C and 1D).

As an initial characterization of these compartments, we sorted cells from 3 to 5 donors and performed population-level

to the PTAs of the peripheral cells they mimic and can also perform unanticipated functions.[4,7]

Studies on human thymic mimetic cells have not advanced as far as those on mice. There have been variable reports of tuft, muscle, neuroendocrine, ionocyte, ciliated, myelin-expressing, and/or mechanosensory thymic mimetic cells.[13,14,17,18] But none of these reports focused on the post-Aire-expressing mTEC compartment wherein mimetic cells reside, rather assaying whole thymus, unfractionated mTECs, or Aire-stage mTECs, resulting in inconsistent and relatively shallow mimetic-mTEC coverage. In-depth studies of the landscape and phenotype of human thymic mimetic cells thus remain lacking, a gap made particularly acute by their anticipated relevance to autoimmune diseases.

Here, we characterized thymic mimetic cells from multiple human pediatric thymi, focusing on the post-Aire mTEC compartment. We observed a diversity of subtypes that both overlapped with and diverged from those of mice, including a striking expansion of human muscle mTECs. Additionally, we identified a plethora of thymic mimetic-cell subtypes in zebrafish, indicating that mimetic mTECs are evolutionarily ancient. Cross-species analyses among humans, mice, and zebrafish revealed many mimetic-cell subtypes that were conserved across all three species and others that reflected species-specific physiology.

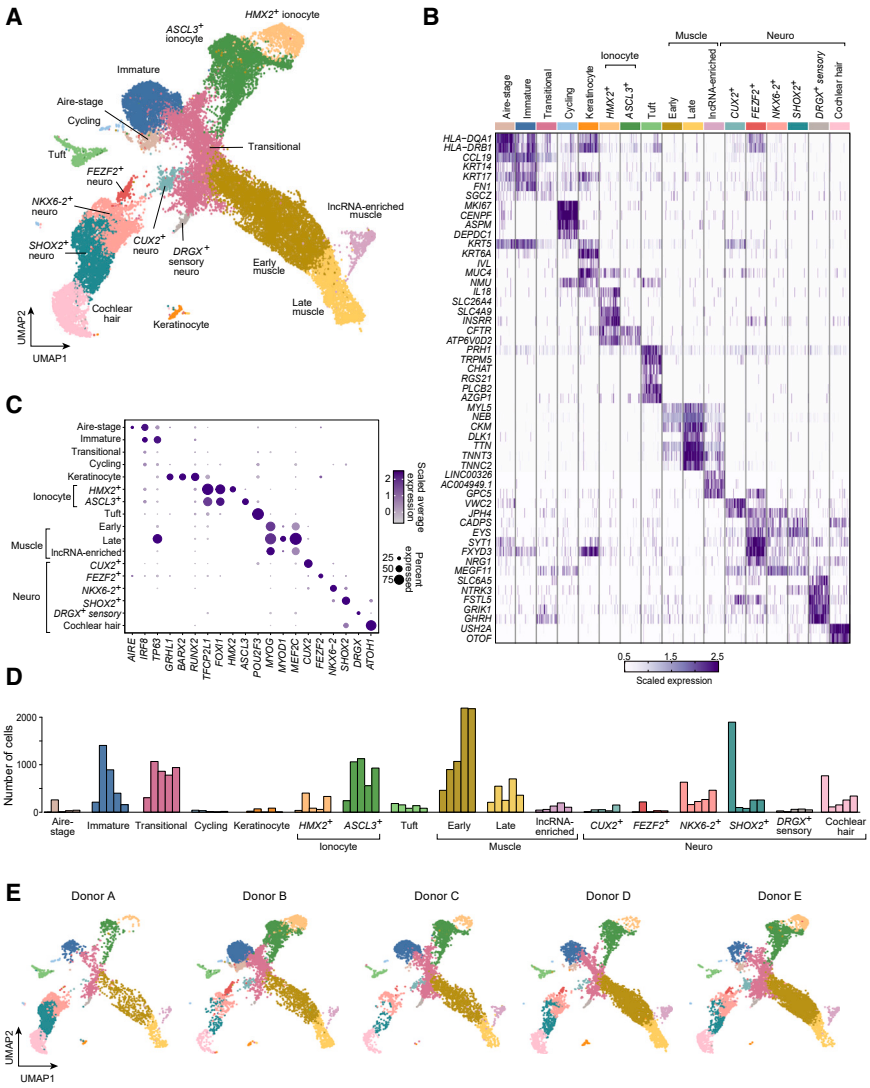


Figure 2. scRNA-seq reveals a diverse repertoire of mimetic cells in the human thymus, including large populations of muscle, neuroendocrine, and ionocyte mimetic cells

(A) UMAP of scRNA-seq of human mimetic cells.
(B) Heatmap of select genes. For each cluster, up to 100 randomly sampled cells are shown.
(C) Expression of genes encoding relevant TFs.
(D) Donor composition of clusters (in order left to right, donors A–E)
(E) UMAP, as in (A), split by donor.
See also Figures S1 and S4 and Table S1.

RNA-seq. According to principal-component analysis (PCA), samples from the same compartments clustered together across the donors (Figure 1E). Clustering of differentially expressed transcripts (fold change [FC] $\geq$ 2; false discovery rate [FDR] < 0.05) revealed distinct programs of gene expression in each of the three compartments (Figure 1F). The DP.mTEC^{lo} population resembled the corresponding pre-Aire compartment of mice in its enrichment for collagen- and integrin-encoding transcripts.[4] The mTEC^{hi} population preferentially expressed the *AIRE* gene, consistent with its Aire-stage designation in mice, as well as transcripts encoding human leukocyte antigen (HLA) class II molecules (e.g., *HLA-DRA1* and *HLA-DRB1*), associated antigen-processing molecules (e.g., *CD74*), and costimulatory ligands (e.g., *CD70*, *CD80*, and *CD86*). DN.mTEC^{lo} expressed transcripts encoding lineage-defining TFs such as *MYOG* and *NEUROD1*, as well as other PTA transcripts such as *CKM*, *TTN*, and *CFTR*, consistent with a post-Aire, mimetic-cell-enriched compartment like that of mice. Indeed, homologous marker genes showed similar expression patterns in human and murine mTECs, although PTAs such as *CFTR*, *CHRNG*, and *TTN* seemed to be more strongly enriched in the DN.mTEC^{lo} compartment of humans (Figure 1G).

In sum, human mTECs can be separated into populations analogous to the pre-Aire, Aire-stage, and mimetic-cell-enriched post-Aire mTEC compartments of mice based on HLA-DR, PDPN, and CD104 expression. There is overall conservation of these populations' transcriptional programs between mice and humans, but with some species-specific nuances.

Diverse mimetic-cell subtypes populate the human thymus

We more deeply characterized human mimetic cells by performing scRNA-seq analysis of DN.mTEC^{lo} cells from five donors (A–E). After imposing quality-control filters, 29,310 mTECs (5,097–6,609 per donor) were retained. After cell clustering and dimensionality reduction, we observed a heterogeneous mimetic-cell compartment reflecting diverse peripheral cell and tissue types: muscle (three subtypes), neuroendocrine (six subtypes), tuft, keratinocyte, ionocyte (two subtypes), and transitional mTECs (Figure 2A). Gene-clustering and marker-gene analyses validated the cell-cluster distinctions, verifying that the various subtypes expressed TFs and genes representative of their corresponding peripheral tissues (Figures 2B and 2C). Each of these clusters is discussed in detail in subsequent sections. We also captured small numbers of immature, cycling, and Aire-stage mTECs, likely representing sorting contamination.

Muscle, ionocyte, and neuroendocrine mimetic cells were the most abundant clusters in the scRNA-seq data, whereas keratinocyte and tuft mimetic cells were less abundant (Figure 2D). The relative cluster frequencies were well-conserved across donors and did not vary substantially by sex or across the ages surveyed (Figure 2E; Figures S1B–S1E). The only donor with a substantially altered frequency distribution was donor A, who possessed a relative enrichment of neuroendocrine mTECs and a relative paucity of muscle and ionocyte mTECs. This was also the only donor who had Down syndrome.

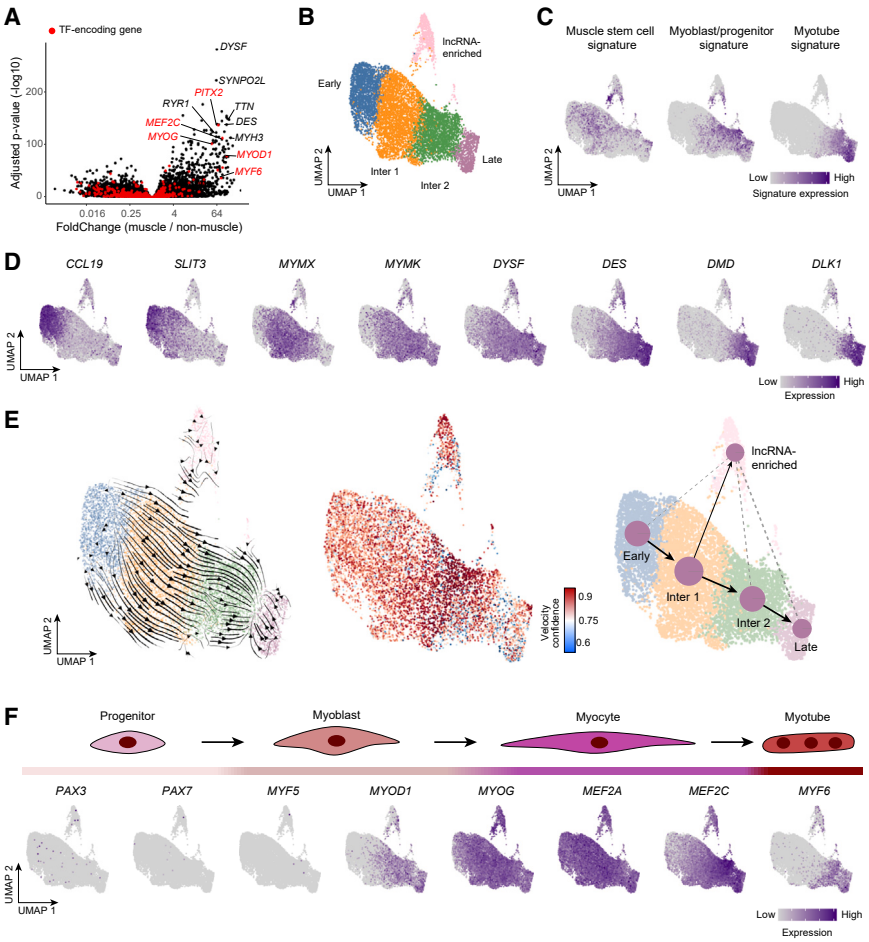


Figure 3. Muscle mimetic cells in the human thymus reflect a differentiation trajectory

(A) Differentially expressed genes in muscle clusters versus other mimetic-cell clusters, by pseudo-bulk analysis. TFs (as curated in Lambert et al.[24]) are in red. Select genes relevant in muscle are labeled.

(B) Human muscle mimetic-cell clusters, reclustered, with outlier clusters (non-muscle) removed.

(C) Signatures from a skeletal-muscle dataset from McKellar et al.[22] overlaid on the muscle mimetic-cell UMAP.

(D) Marker genes enriched in early, intermediate, and late muscle mimetic-cell stages.

(E) RNA velocity using scvelo, with velocities shown as streamlines and velocity confidence, and partition-based graph abstraction (PAGA) representation.

(F) Schematic of muscle development and expression of genes encoding TFs known to be involved in peripheral skeletal-muscle differentiation.

See also Figure S2.

The muscle mimetic-cell population is greatly expanded in the human thymus

Arguably the most striking difference in the mimetic-cell compartment of humans and mice was the approximately 25-fold more frequent occurrence of muscle mTECs in humans (36% versus 1.5%).[4] Human muscle mimetics expressed transcripts encoding myogenic TFs (*MYOG* and *MEF2C*) as well as key structural constituents of skeletal muscle (*TTN* and *DMD*). Thymic mimetics resembling both early and late skeletal-muscle cells were observed, the latter exhibiting enhanced expression of muscle markers such as *DLK1*, *TTN*, *CKM*, and *TNNT3* (Figure 2B). We were surprised to see a cluster of muscle mTECs distinguished by its enrichment in long non-coding RNAs (lncRNAs), strongly conserved across all five donors (Figures 2A and 2B; Figure S2A). Despite our best efforts, we were unable to define marker genes that could offer insight into this cluster's function beyond its strong enrichment for numerous lncRNAs, such as *LINC00326* and *AC004949.1*. Nor could we find a literature analog in peripheral muscle, though a few studies have reported lncRNAs generally to be important in skeletal muscle.[20] Whether this cluster is a mimetic-cell-related phenomenon, reflects some as-yet-undiscovered muscle biology, or is some type of experimental anomaly remains to be seen.

We were particularly intrigued by the expansion of muscle mimetic cells because of their potential relevance to the autoimmune disorder, myasthenia gravis (MG), in which autoantibodies target the neuromuscular junction and cause muscle weakness, sometimes so severe as to be fatal. MG is strongly associated with the presence of thymic tumors, and thymectomy often results in a marked clinical improvement in myasthenic symptoms.[21] Thus, given their enrichment in humans and potential disease relevance, we first focused on muscle mimetic cells for deeper analysis.

As an initial step, we compared muscle mimetic cells with all of the other mimetic-cell subtypes. According to pseudo-bulk analysis, the muscle mimetics showed increased expression of genes encoding canonical muscle TFs (including *MYOG*, *MYOD1*, *MEF2C*, and *MYF6*) (Figure 3A). Reclustering them in isolation yielded five subclusters, composed of 9,060 total cells (Figure 3B; Figure S2B). Overlaying published transcript signatures from peripheral muscle types[22] suggested a differentiation trajectory (Figure 3C): a muscle stem-cell signature was enriched in the "early" and "intermediate-1" clusters, and cells in the intermediate-1 and intermediate-2 clusters were enriched for myoblast and muscle-progenitor transcripts, and myotube signature transcripts were increased in the "late" cluster. This differentiation trajectory was also reflected in the transcription of individual marker genes (Figure 3D). For example, early muscle mimetic cells transcribed *CCL19*, which is preferentially expressed in immature mTECs,[23] as well as *SLIT3*, and the intermediate muscle-mTEC clusters expressed elevated levels of the genes encoding myomixer (MYMX), myomaker (MYMK), and dysferlin (DYSF), while the late cluster showed enriched expression of genes specifying desmin (DES) and dystrophin (DMD), whose products are found in peripheral myofibers. RNA

velocity analysis, which focuses on the ratio of spliced to unspliced mRNA, also predicted a differentiation trajectory from the early to late muscle-mTEC clusters (Figure 3E). Integration of the lncRNA-enriched cluster into this trajectory was less apparent but, most confidently, it seemed to derive from the intermediate-1 cluster.

We next investigated how transcripts encoding TFs progressed along the muscle-mTEC putative differentiation trajectory, a progression that has been well characterized in skeletal-muscle differentiation (reviewed by Imbriano and Molinari[25]). Peripheral-muscle differentiation trajectories usually cite PAX3 and PAX7 expression in muscle progenitors; expression of the pioneer myogenic determinants, MYF5 and MYOD1, in myoblasts; expression of MYOG, which is required for terminal muscle-cell differentiation, alongside MEF2A and MEF2C, in myocytes; and MYF6 in mature myotubes.[25,26] Transcripts encoding PAX3 and PAX7 were very sparsely expressed in muscle mTECs, without a discernible gradient along the muscle differentiation trajectory, suggesting that muscle mTECs might not rely on these factors to initiate muscle differentiation programs (Figure 3F). Instead, early muscle mimetics exhibited persistently high expression of *MYOG* and *MEF2A* transcripts (Figure 3F), both of which are only transiently expressed at an intermediate phase of peripheral-muscle differentiation.[22,26,27] *MYOD1* transcription, which in the periphery precedes that of *MYOG*, was evident only at the intermediate stage of muscle-mTEC differentiation, after *MYOG* transcription had already been turned on. Finally, intermediate-2 and late muscle mTECs showed increased expression of the genes encoding MEF2C and MYF6, which are also associated with late muscle differentiation in the periphery. Despite their divergent path(s) to the differentiated muscle fate, late muscle mTECs nonetheless expressed an array of transcripts encoding mature-muscle markers, highlighting the central role of TFs, rather than strict ontogenic recapitulation, in driving mimetic-cell gene-expression programs.

We also wondered to what extent the various muscle-mTEC subtypes reflected different peripheral-muscle types. Therefore, we examined the expression of myosin-chain transcripts, which have specific patterns of expression in peripheral slow- and fast-twitch (type I and type II) skeletal muscle as well as in atrial and ventricular cardiac muscle. When present, these transcripts were most highly expressed in the late muscle-mTEC cluster (Figure S2C), without distinct skeletal or cardiac gene-expression patterns. Similarly, we examined expression of both a signature of transcripts enriched in cardiomyocytes[28] and a cardiac-type troponin (*TNNI3*) (Figure S2D), both of which were enriched in late mTECs. Thus, late muscle mimetic cells appeared to be a broad representation of mature-muscle types.

In brief, then, muscle mimetic cells are strikingly enriched in human pediatric vis-à-vis mouse thymi. They manifest a clear differentiation trajectory, one that relies on some of the same TFs as, but perhaps follows a different route from, what is characteristic of peripheral muscle. An enigmatic lncRNA-enriched muscle subset may reflect as-yet undiscovered muscle biology.

Muscle mimetic cells are histologically abundant, and some of them may form neuromuscular junctions

To confirm the abundance and nature of the muscle mimetic cells revealed by scRNA-seq analysis, we performed immunofluorescence staining of thymus sections using antibodies against various muscle markers. DES$^+$ cells were abundant in, and specific to, the thymus medulla (Figure 4A), with many of them exhibiting elongated morphology resembling that of peripheral myocytes. DES staining overlapped with that of myogenin in some nuclei (Figure 4B, white arrows), but there were also myogenin$^+$ DES$^-$ muscle mimetics (Figure 4B, blue arrow). DES and myosin-heavy-chain stainings also colocalized, occurring at a frequency of approximately 40 per mm^2 in the medulla (Figure 4C). Similarly, the DYSF and myosin-heavy-chain stainings overlapped (Figure 4D), although DYSF$^+$ cells were about 7-fold more abundant than the cells expressing myosin heavy chain. That DYSF expression has also been observed in thymic endothelial and vascular smooth muscle cells in the thymus[13] may explain its greater abundance. Strikingly, some muscle mTECs had striated morphology that mirrored the cellular microstructure of mature skeletal or cardiac muscle (Figure 4E).

Hassall's corpuscles are a prominent structure in the human thymus. Alongside keratin10 ringing Hassall's corpuscles, as previously observed,[29] we also found DMD$^+$ cells (Figure S3A). This morphology was not recapitulated with the other muscle markers, although some DES$^+$ cells were located at the edge of Hassall's corpuscles (Figures 4A and 4B). Since the genes encoding DES and DMD are expressed in late muscle mimetic cells (Figure 3D), the DES$^+$ or DMD$^+$ cells associated with Hassall's corpuscles may represent a subset of mature-muscle mimetic cells, although the significance of this localization is unclear.

We were curious whether muscle mimetic cells formed neuromuscular junctions in a manner analogous to that of skeletal muscle, especially given that such junctions appear to be autoimmunizing agents in MG. We stained thymus sections with antibodies against myosin heavy chain and the neuroendocrine markers SOX2, S100, and CADPS. SOX2 and S100 staining revealed adjacent muscle and neuroendocrine cells. Approximately half of the myosin-heavy-chain$^+$ cells showed overlapping staining of CADPS, which is involved in secretory vesicle exocytosis (Figures 5A–5C). To determine whether these cells formed potentially functional neuromuscular junctions, we stained additional thymus sections with fluorescently labeled alpha-bungarotoxin, a neurotoxin that binds to the nicotinic acetylcholine receptor (AChR)[30] (Figure S3B). Alpha-bungarotoxin staining was not detected at the interface between neuroendocrine and muscle mimetic cells according to most muscle markers, suggesting that they generally did not form cholinergic synapses. However, alpha-bungarotoxin did colocalize with DYSF$^+$ cells (Figure 5D). Recalling that tuft mimetic cells expressed *CHAT* (Figure 2B), which encodes the enzyme that catalyzes the synthesis of acetylcholine, we wondered whether tuft mTECs might provide acetylcholinergic stimulus to DYSF$^+$ cells. Of note, tuft-muscle interactions have been observed in peripheral tissues (reviewed by O'Leary et al.[31]). Co-staining for DYSF, CHAT, and alpha-bungarotoxin uncovered multiple, though rare, instances of tuft-muscle interfaces with intervening alpha-bungarotoxin (at about 1 in 800 DYSF$^+$ cells) (Figure 5E). Alpha-Bungarotoxin staining also colocalized with an alternative tuft marker, TRPM5, indicating that we were specifically labeling tuft mimetics in the thymus (Figure S3C). About 30% and 10% of tuft mTECs and DYSF$^+$ cells, respectively, had associated

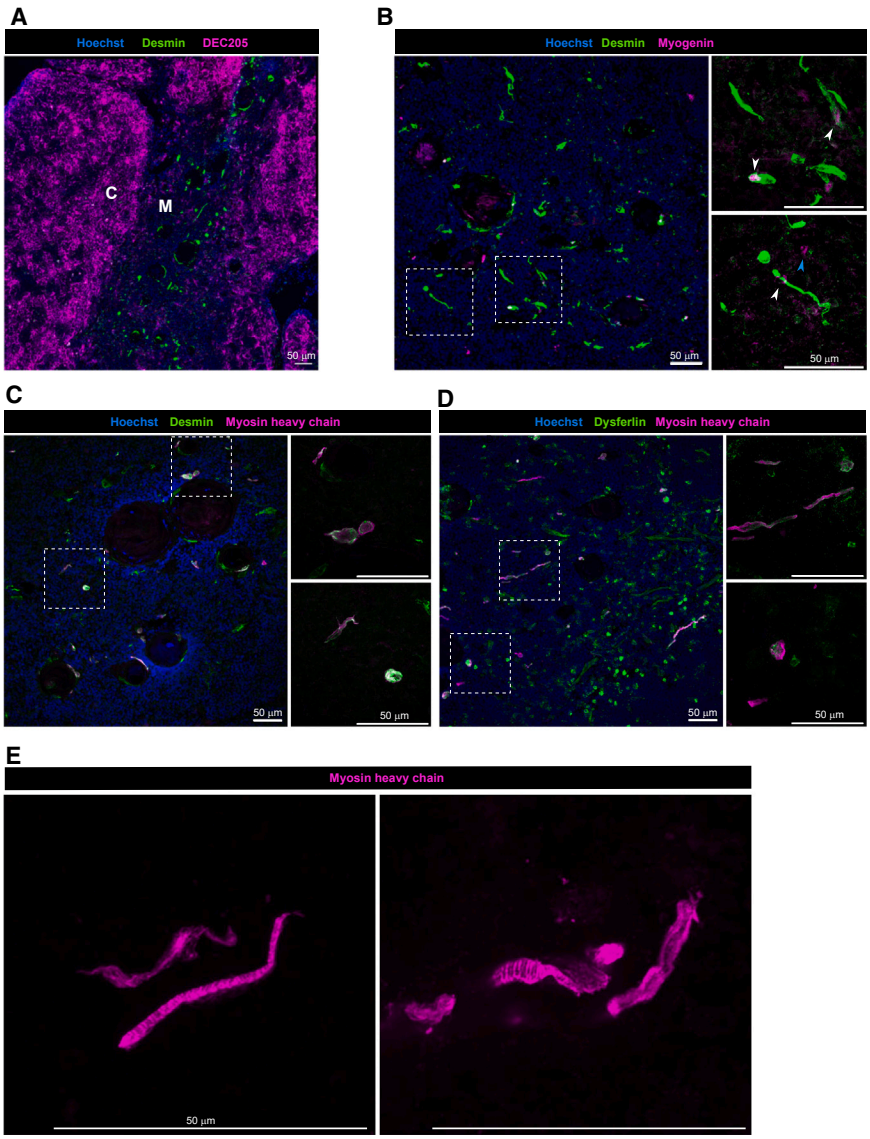


Figure 4. Histological examination of the human thymus reveals abundant muscle mimetic cells

(A–E) Sections from 4-month-old female with tetralogy of Fallot and pulmonary stenosis. Hoechst nuclear stain included in low-power images in blue. (B) White arrows highlight myogenin$^+$ desmin$^+$ cells, and blue arrows highlight myogenin$^+$ desmin$^-$ cells.

See also Figure S3.

alpha-bungarotoxin staining, while approximately 8% of tuft mTECs overlapped with DYSF$^+$ cells.

Thus, immunohistology confirmed the abundance and heterogeneity of human muscle mimetic cells. It also revealed a frequent association of muscle and neuroendocrine mimetics and an intriguing, though occasional, colocalization of tuft mimetic cells, DYSF$^+$ cells, and alpha-bungarotoxin, raising the possibility that tuft and at least some muscle mimetics may interface in the thymus in an acetylcholine-driven manner.

Neuroendocrine mTECs mimic a multiplicity of cell types

scRNA-seq of the human mimetic-cell compartment also revealed a striking expansion and diversification of neuroendocrine mimetic cells, subdivided into six subtypes, delineated by their expression of genes encoding the TFs ATOH1, NKX6-2, SHOX2, FEZF2, CUX2, and DRGX (Figures 2A–2C; subclustered in Figures S4A and S4B). *ATOH1*-expressing mimetic cells resembled cochlear-hair cells, which enable the inner ear to detect sound via mechanotransduction.[32] In addition to *ATOH1*, these mimetics expressed genes specifying the TFs BARHL1, POU4F3, PAX2, and LHX3 (Figure S4B), which comprise a known TF network in inner-ear hair cells and are important for their differentiation and maintenance,[33–38] underscoring the biological logic of PTA expression in this mimetic-cell subtype. In addition to these TFs, cochlear-hair mimetic cells expressed the genes *OTOF* and *USH2A* (Figure 2B), both of which have been implicated in human otic diseases: mutations in *OTOF* cause non-syndromic deafness,[39] while mutations in *USH2A* underlie Usher syndrome, which is accompanied by hearing loss.[40] *ATOH1* expression has been reported in the human thymus, by cells described as either mechanosensory[17] or ciliated[14] mTECs. While the mimetic-cell repertoire of mice includes *FoxJ1*$^+$ ciliated mTECs, these cells lack *Atoh1* expression as well as the neuroendocrine features of *ATOH1*$^+$ cochlear-hair mimetic cells.[4]

There was also a population of interneuron mimetic cells, expressing transcripts specifying the TF NKX6-2 (Figure 2C), which is characteristic of peripheral cortical interneurons and their ventral telencephalon progenitors.[41] In addition, *SIM1* and *NKX2-2* transcripts (Figure S4B) were found in this mimetic-cell subtype, reflective of analogous spinal-cord interneurons.[42] Interestingly, the *NKX6-2*$^+$ neuro cluster also contained a subset of cells that expressed the gene encoding MYT1L (Figure S4B), which is associated with neurodevelopmental disorders (reviewed in Chen et al.[43]). *NKX6-2* transcription has been reported in the thymus in "mTEC-neuroendocrine-early" cells,[17] although a distinct mTEC subtype characterized by its expression has not previously been described in humans or mice.

Another neuroendocrine mimetic-cell subtype expressed the gene encoding the TF SHOX2 (Figure 2C), transcribed as a gradient overlapping *ATOH1*$^+$ and *NKX6-2*$^+$ mimetic cells (Figure S4B). In the periphery, SHOX2 is required for the differentiation of subsets of motor neurons.[44] A small subset of *SHOX2*$^+$ mimetic cells also specifically expressed the gene encoding

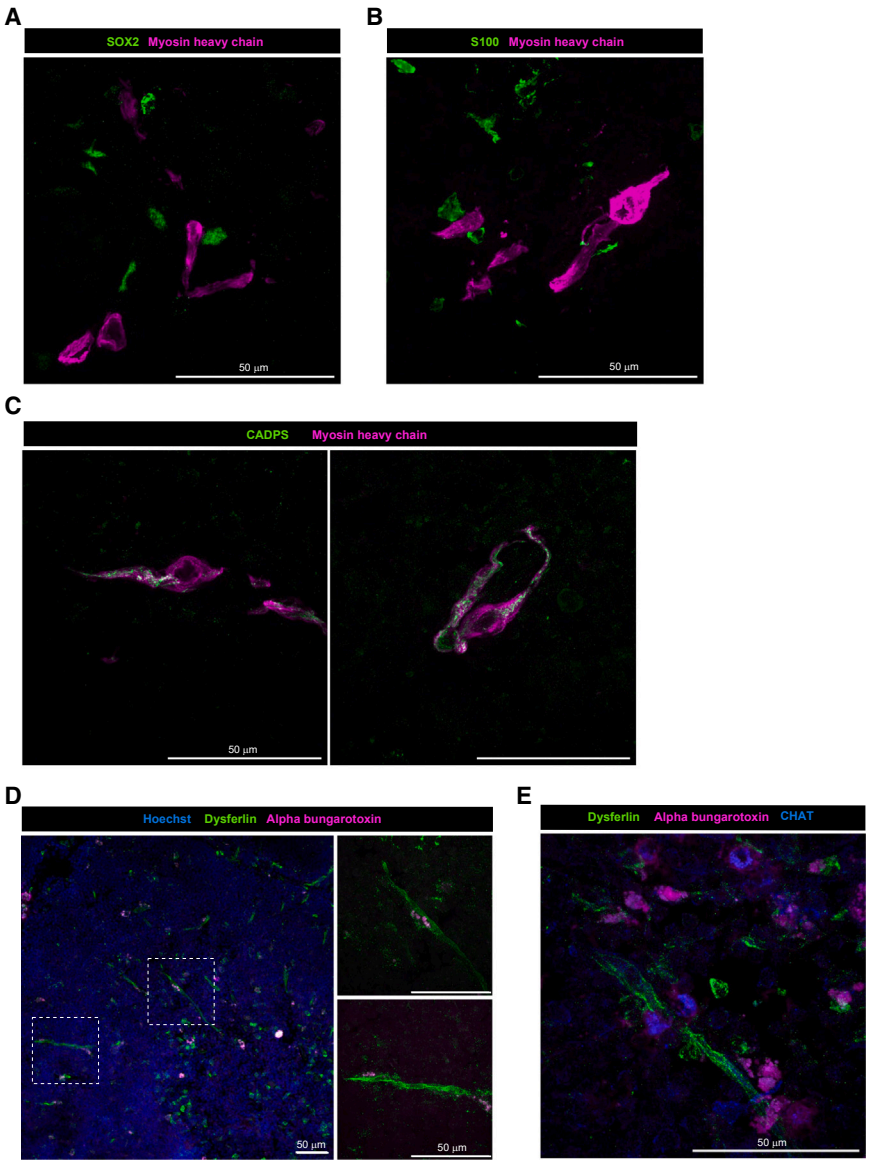


Figure 5. Mimetic cells in the human thymus form cellular groupings reminiscent of neuromuscular junctions

(A–E) Sections from 4-month-old female with tetralogy of Fallot and pulmonary stenosis. Hoechst nuclear stain included in low-power images in blue.

See also Figure S3.

the TF PHOX2B (Figure S4B), which is required for differentiation of visceral motor neurons,[45,46] suggesting that *SHOX2*$^+$ mimetic cells are analogous to motor neurons and that the *PHOX2B*$^+$ subset, in particular, is analogous to visceral motor neurons.

Other clusters of neuroendocrine mimetic cells transcribed genes encoding the TFs FEZF2 and CUX2 (Figure 2C; Figure S4B), thereby being analogous to lineages of cortical projection neurons in the brain, i.e., *FEZF2*$^+$ or *CUX2*$^+$ radial glial cells engender cortical projection neurons,[47,48] and *FEZF2*$^+$ radial glial cells also give rise to other glia populations.[48] Fezf2 has been reported to drive PTA expression in the murine thymus, though the underlying mechanism remains unclear.[49] A specific *FEZF2*$^+$ cluster of neuro-mimetic cells could potentially explain the thymic transcriptional changes and autoimmunity associated with *Fezf2* deletion in mice.[49]

Lastly, we identified a population of mimetic cells expressing the gene encoding the TF DRGX, as well as the *NTRK3* gene (Figures 2B and 2C; Figure S4B). These two loci are transcribed in the dorsal root ganglia and primary sensory neurons,[50,51] identifying this mimetic-mTEC subtype as analogous to sensory neurons. *DRGX*$^+$ neuro-mimetic cells also expressed the genes specifying the glutamate receptor subunit, GRIK1; the glycine neurotransmitter transporter, SLC6A5; and the hypothalamic hormone, GHRH (Figure 2B). *DRGX*$^+$ sensory populations have not to our knowledge previously been described in mouse or human thymi.

In sum, humans display a diverse collection of neuroendocrine mimetic-cell subtypes, collectively comprising a markedly larger fraction of the mimetic-mTEC compartment than that observed in mice.[4] Human neuroendocrine mimetics, characterized by expression of transcripts specifying TFs such as ATOH1, SHOX2, NKX6-2, DRGX, CUX2, and FEZF2, mimic sensory neurons, motor neurons, interneurons, and neurons of the brain.

Ionocyte mimetic cells resemble multiple peripheral ionocyte types

Also prominent among the human mimetic-cell subtypes were *FOXI1*-expressing ionocyte mimetics. Ionocytes are ion-secreting cells with tissue-specialized functions[52]—for example, intercalated cells of the kidney control urine pH,[53] and pulmonary ionocytes regulate airway fluid and mucus properties.[54,55] They express the gene encoding CFTR, the anion channel for which loss-of-function mutations cause cystic fibrosis,[54–56] and are dependent on the lineage-defining TF, FOXI1.[57] Peripheral ionocyte subtypes also share transcripts specifying the TF, TFCP2L1, and the V-ATPase subunits.[52,54,57–59] Mirroring their peripheral counterparts, thymic ionocyte mTECs expressed *CFTR*, *FOXI1*, *TFCP2L1*, and V-ATPase-subunit-encoding genes such as *ATP6V0D2* (Figures 2B and 2C). The population of human ionocyte mimetic cells was not homogeneous, encompassing two distinct subtypes. One transcribed the gene encoding the TF ASCL3 (Figures 2A–2C)[52,60]; the other specifically expressed the gene specifying the TF HMX2, as well as *SLC26A4*, *SLC4A9*, and *INSRR* transcripts (Figures 2B and 2C; Figure S4C). This set of transcripts identified this subtype as analogous to type B intercalated cells in the kidney, which regulate

bicarbonate secretion in the urine.[53,57,58] These cells also expressed *IL18* (Figure 2B), encoding a proinflammatory mediator thought to be important during urinary tract infections.[61]

Additional mimetic-cell subtypes diversify the thymic PTA repertoire

While the muscle, neuroendocrine, and ionocyte mimetic-cell subtypes were abundant in the human thymus, several other subtypes, namely tuft and keratinocyte mTECs, were present at lower frequencies (Figures 2A and 2D). Tuft mTECs mimic peripheral tuft cells, chemosensory first-responders found in the intestine, trachea, and taste buds.[62] Though peripheral tuft cells were only recently reported to be key controllers of type-2 immunity, via support of type-2 innate lymphoid cells (ILC2s),[63–65] tuft mTECs were the first mouse thymus mimetics to be deeply characterized.[11,12,66,67] In agreement with these reports, we observed human tuft mimetics to express the gene encoding the lineage-defining TF, *POU2F3*, as well as transcripts specifying CHAT and the taste transduction molecules, TRPM5 and PLCB2 (Figures 2B and 2C; Figure S4D). However, human tuft mTECs did not specifically express *DCLK1* transcripts (Figure S4D), consistent with their absence from other human tuft-cell populations but in contrast to the situation in mice, where DCLK1 is a reliable marker of tuft cells.[68]

Keratinocyte mTECs mimic peripheral keratinocytes of the epidermis.[69] In humans, keratinocyte mimetic cells expressed the gene encoding the lineage-defining TF, GRHL1, and were enriched for keratin-gene transcripts, including *KRT5* and *KRT6A* (Figure 2B). Keratin-rich cells are frequently found in Hassall's corpuscles.[18] Though Hassall's corpuscles are abundant in the human thymus according to histologic analyses, there was a relative paucity of keratinocyte mimetics by scRNA-seq analysis. We speculate that they may be more resistant to release by our digestion protocol.

Finally, we annotated a subset of mTECs as transitional mimetic cells, given their occupancy of uniform manifold approximation and projection (UMAP) space intermediate between multiple mimetic-cell subtypes. They were characterized by low-level expression of genes encoding multiple lineage-defining TFs and PTAs (e.g., MYOG and POU2F3), as well as residual expression of genes encoding some immature mTEC markers (e.g., CCL19) (Figures 2B and 2C). These cells may represent an intermediate state between immature and/or Aire-expressing mTECs and differentiated mimetic-cell types.

Zebrafish also have thymic mimetic cells—Some conserved, some reflecting unique fish physiology

We also investigated the evolutionary conservation of thymic mimetic cells in zebrafish, which diverged from humans approximately 450 million years ago, after the evolution of vertebrate adaptive immunity (Figure 6A).[70,71] Sparse thymic mimetic cells, including neural and tuft TECs, were previously observed in the course of profiling zebrafish lymphocyte differentiation.[72] To enrich for thymic-epithelial cells, we performed scRNA-seq on dissociated thymus tissue depleted of T and B cells on the basis of transgenic reporters (given the paucity of antibody reagents for zebrafish) (Figures S5A and S5B; Table S2). These data were analyzed alongside published zebrafish thymus scRNA-seq datasets of sorted live cells.[72] After excluding hematopoietic cells, we annotated mimetic-cell subtypes based on gene expression and computationally predicted label transfer using the recently described pre-trained scGPT model,[73] fine-tuned on our human mimetic-cell scRNA-seq data (Figure S6). We identified a diversity of mimetic-cell subtypes enriched for *epcam* expression (Figures 6B and 6C; Figure S5C): muscle, ionocyte (two subtypes), tuft (three subtypes), neuroendocrine (five subtypes), periderm, ciliated, metaphocyte, macrophage-like, structural, and ear non-sensory mTECs.

Zebrafish ionocyte mimetics were of two subtypes, which reflected specialization in zebrafish physiology necessary for maintaining homeostasis and survival in a freshwater aquatic environment. One subtype expressed *atp6v1aa*, *slc9a3.2*, and *ca15a* (Figures 6C and 6D), identifying it as analogous to H^+-ATPase-rich (HR) ionocytes, which are responsible for acid-base regulation and sodium uptake, crucial functions due to low environmental ion concentrations and variable water pH in the zebrafish habitat.[75,76] The other subtype expressed *trpv6* and *slc8a1b* (Figures 6C and 6D), suggesting that it mimics Na^+-K^+-ATPase-rich (NaR) ionocytes, which play a role in calcium uptake, important for maintaining physiological calcium ion concentrations in the face of environmental fluctuations.[77,78] Zebrafish also had two additional primary ionocyte subtypes: Na^+-Cl^- cotransporter-expressing (NCC) and K^+-secreting (KS).[76,79] The KS ionocyte marker gene *kcnj1a.1* was expressed alongside HR marker genes, while the NCC marker gene, *slc12a10.2*, was expressed in only a subset of these cells (Figure 6D).

The zebrafish thymus contained three tuft mimetic subtypes (Figures 6B, 6C, and 6E). Immune-type tuft cells transcribed *alox5a*, which encodes a molecule involved in proinflammatory arachidonic acid metabolism, as well as *il17rc* transcripts, encoding a member of the interleukin (IL)-17 cytokine-signaling axis. Gut-type tuft cells expressed transcripts specifying the aldehyde dehydrogenase enzyme (i.e., *aldh1a3*) and *vil1*, which is frequently used as a marker of tuft cells. Neurosensory-type tuft cells expressed *gng13a* and *trpm5*, encoding molecules involved in the sensory perception of taste.

As in humans, the zebrafish thymus contained a wealth of neuroendocrine-mTEC subtypes, some fish-specific. Ear sensory-hair mimetics expressed the marker genes *atoh1a*, *otofb*, and *foxj1a*, as also observed in humans (Figures 2B, 2C, 6B, and 6C). A different, related zebrafish subtype also expressed *atoh1a* transcripts, in addition to those encoding the TFs Sox2, Prox1a, and Drgx (Figure S5D). These markers suggested this cluster to be analogous to the neuromast, which detects water displacement via mechanosensory hair cells on the surface of fish.[80,81] Two major types of olfactory sensory neurons, microvillous and ciliated, are found in zebrafish,[82] and thymic mimetic-cell subtypes analogous to each of them were identifiable (Figures 6B and 6C). Mimetics of ciliated olfactory sensory neurons expressed olfactory receptor activity genes (such as *or132-5* and *or111-11*), *ompb*, and *foxj1b*, while mimetics of microvillous olfactory sensory neurons transcribed *trpc2b* and genes encoding the V2R-type receptor, including *olfcd3* (Figure 6C; Figure S5D). Yet another neuroendocrine mimetic-cell subtype in zebrafish expressed *neurod1*, *neuroregulin 3b* (*nrg3b*), and *lhx9* transcripts (Figure 6C; Figure S5D).

The zebrafish thymus contained three structural mimetic-cell subtypes. One expressed *sox10*, *bricd5*, *stm*, *otomp*, and *otog*

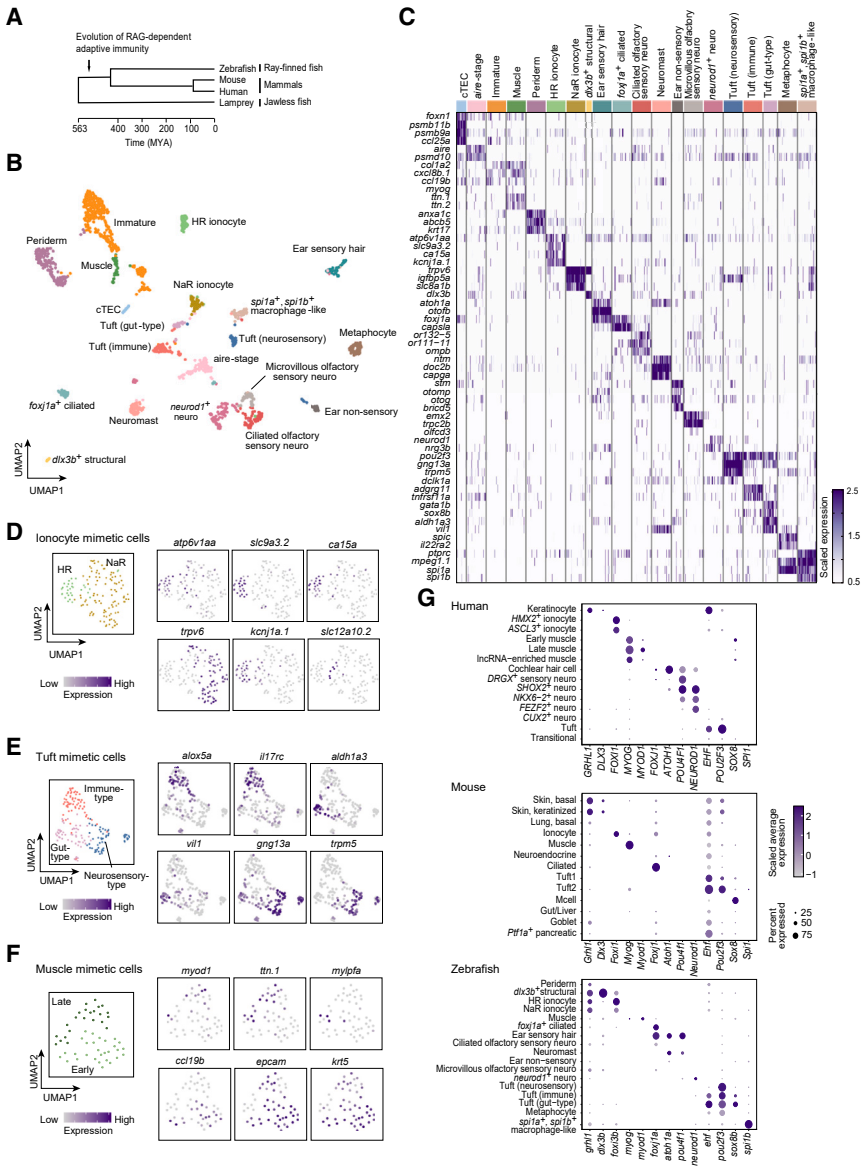


Figure 6. The zebrafish mimetic-cell compartment contains subtypes that overlap with humans and mice, as well as species-specific subtypes reflecting unique fish physiology

(A) Evolutionary tree, showing relationships between lamprey, zebrafish, humans, and mice. Adapted from visualization generated by TimeTree.[74]

(B) UMAP of zebrafish scRNA-seq.

(C) Heatmap of select genes. For each cluster, up to 50 randomly sampled cells are shown.

(D) Subclustered zebrafish ionocyte mimetic cells and expression of select relevant genes.

(E) Subclustered zebrafish tuft mimetic cells and expression of select relevant genes.

(F) Subclustered zebrafish muscle mimetic cells and expression of select relevant genes.

(G) Expression of conserved transcription factors across mimetic cells in three species. See also Figures S5 and S6 and Tables S2 and S3.

transcripts (Figure 6C; Figure S5D), suggesting it to be analogous to ear non-sensory cells, a structural cell type forming the otic capsule.[83] A second structural subtype was reminiscent of the periderm, a developmental epidermal layer that expresses, for example, *krt17, anxa1c*, and *grhl3* (Figure 6C; Figure S5D). A third transcribed the gene encoding the TF dlx3b, which is found in peripheral structural tissues such as tooth epithelium,[84] though this cluster was composed of only fourteen cells and so was less well defined (Figures 6B and 6C; Figure S5D).

We discovered two interesting *epcam*$^{+}$ populations in the zebrafish thymus. One expressed *mpeg1.1*, *ptprc*, *il22ra2*, and an unusual group of genes encoding TFs (e.g., *pou2f3*, *spic*, and *prox1a*) (Figures 6B and 6C; Figure S5D). This collection of marker genes evoked metaphocytes, ectoderm-derived, macrophage-like cells occupying zebrafish barrier tissues.[85,86] Though not known to be resident in the thymus, these cells may represent bona fide metaphocytes or, instead, reflect a metaphocyte-like mimetic-cell population. The other population transcribed *spi1a* and *spi1b* as well as genes encoding conventional macrophage markers, such as *fcer1g* and *mpeg1.1* (Figures 6B and 6C; Figure S5D). Though these cells were enriched for *epcam* expression, we cannot entirely rule out that they were thymus-localized macrophages.

Finally, like that of mice and humans, the zebrafish thymus contained muscle mimetic cells, though, as in mice, they comprised a markedly smaller fraction of the mimetic-cell compartment than did those in humans (Figure 6B). The zebrafish muscle cluster was most abundant in sample 2, which was treated with liberase, suggesting that enzymatic digestion may be necessary to efficiently release zebrafish muscle mTECs (Figure S5B; Table S2). The zebrafish muscle mimetics expressed genes encoding classical muscle TFs (e.g., *myog* and *myod1*) as well as genes specifying mature-muscle markers (e.g., *ttn.1* and *mylpfa*) (Figures 6C and 6F). These cells adjoined immature mTECs in the UMAP, with an expression gradient of immature marker genes such as *ccl19b* and *ccl25a*, suggesting that they may represent a trajectory of muscle mimetic-cell differentiation states akin to those in humans (Figures 6C and 6F). These cells are enriched for the thymic-epithelial genes *epcam* and *krt5* (Figure 6F), suggesting that they are bona fide mimetic cells rather than contaminants from thymus-adjacent skeletal muscle. Nevertheless, to validate the identity of zebrafish muscle mimetic cells, we crossed fish carrying a fluorescent mCherry reporter under the dictates of the *mylpfa* promoter with fish expressing GFP driven by the *lck* promoter, the latter to highlight the thymus organ. By confocal imaging of 2-week-old fish, some mCherry-expressing cells were clearly within the thymus

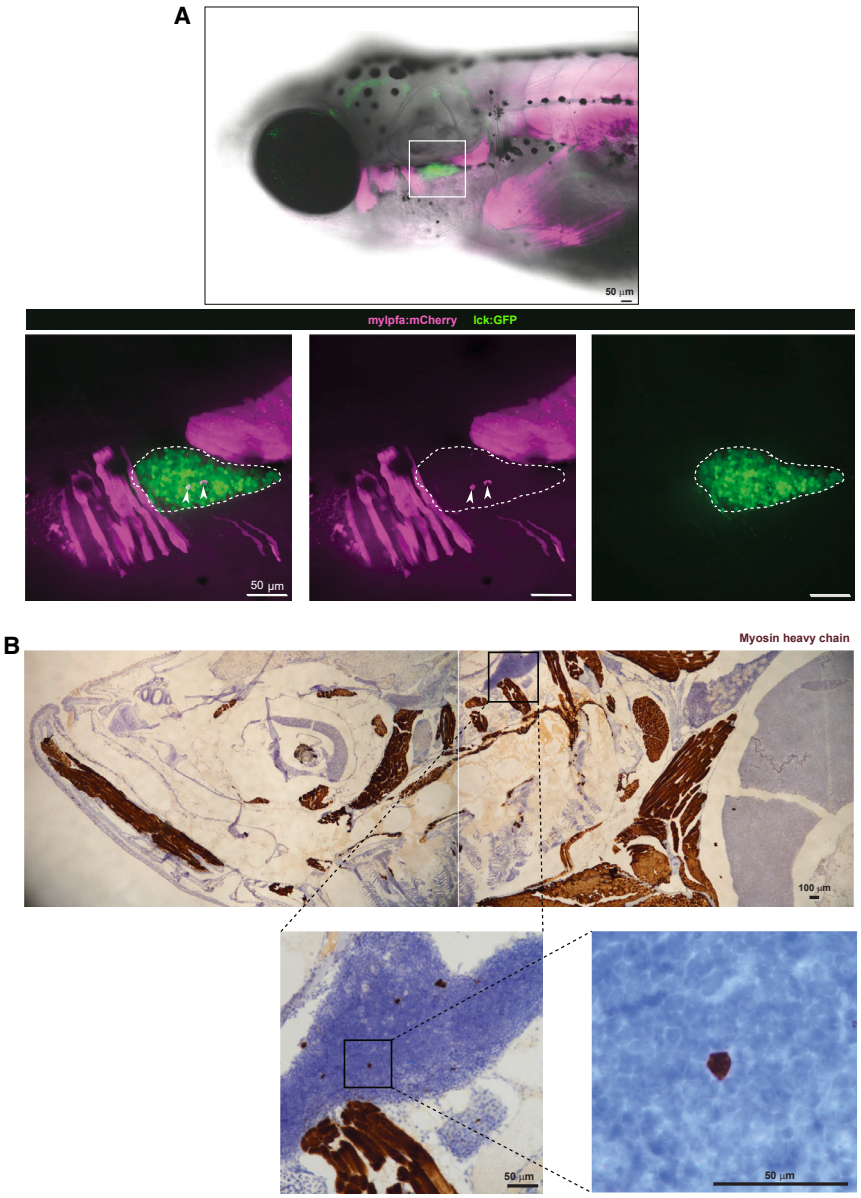


Figure 7. The zebrafish thymus contains muscle mimetic cells

(A) Imaging of the thymus of a 2-week-old zebrafish bearing the transgenes *mylpfa:mCherry* and *lck:GFP*. High-magnification images are maximum intensity projections, and merged and individual color images are included. See also Video S1.

(B) Immunohistochemistry staining with myosin heavy-chain-binding antibody on the thymus of a 3-month-old zebrafish, counterstained with hematoxylin.

See also Video S1.

(Figure 7A; Video S1), suggesting the presence of muscle mimetic cells. These results were confirmed by immunohistopathology using an antibody that binds to myosin heavy chains, and cells expressing myosin heavy chain could be observed in the thymus of a 3-month-old fish (Figure 7B). More muscle mTECs were observed in the 3-month-old compared with the 2-week-old fish, consistent with mimetic cells accumulating with age and populating the fish thymus through adulthood (Figures 7A and 7B).

Mimetic cells can be both species-shared and species-specific

Despite millions of years of divergent evolution (Figure 6A), zebrafish, humans, and mice all possessed thymic mimetic cells, with a remarkable overall conservation of mimetic-cell subtypes (summarized in Table S3). Namely, all three species had tuft, ionocyte, muscle, neuroendocrine, skin, and ciliated mimetics. However, species-relevant specialization of these subtypes was also evident, such as in the human and zebrafish ionocyte subtypes and in the wealth of human and zebrafish neuroendocrine subtypes (Table S3). Zebrafish and mice both hosted mimetic-cell types that were not observed in other species. The mouse-specific subtypes—microfold, gut/liver, goblet, and *Ptf1a*+ pancreatic mTECs—resembled peripheral intestinal and digestive cells, while the subtypes specific to zebrafish mimicked peripheral structural or supporting cells (Table S3).

Moreover, mimetic-cell subtypes expressed largely orthologous lineage-defining TFs across the three species, including Grhl1, Foxi1, Myod1, and Pou2f3 (Figure 6G). But some distinctions in TF gene expression were also evident—for example, several neuroendocrine mimetic-cell subtypes strongly expressed *NEUROD1* in humans, while its ortholog was only weakly transcribed in zebrafish and was not expressed in mice (Figure 6G). Similarly, *Sox8* transcripts mark microfold cells in the mouse, but they were also expressed in zebrafish tuft mimetics (Figure 6G). Such conservation of TF expression supports the centrality of lineage-defining TFs in mimetic-cell differentiation.

DISCUSSION

To provide a high-resolution view of human thymic mimetic cells, we performed scRNA-seq analysis on the flow-cytometrically sorted post-Aire mTEC compartments of multiple pediatric donors, identifying a diversity of mimetic-cell subtypes, many of which had counterparts in the mouse. Yet there were striking differences in the representation of certain subtypes in the two species, most notably a strong enrichment of muscle mTECs, diversification of neuroendocrine and ionocyte mTECs, and impoverishment in tuft mTECs in humans. Relatedly, characterization of the zebrafish mimetic-cell repertoire and its

comparison with that of mice and humans revealed both species-conserved and species-specific mimetic-cell subtypes. For instance, the zebrafish thymus contains aquatic-adapted ionocyte and water-mechanosensing neuromast mimetics.

Several features of the three species' mimetic-mTEC profiles raise questions about the processes underlying mimetic-cell differentiation and regulation. One somewhat surprising facet is that there was no apparent correlation between the abundance or physiological importance of a particular mimetic-cell subtype and that of its peripheral counterpart. Examples include the dominance of tuft mTECs in mice and the mere existence of cochlear-hair mTECs in humans.

Another curiosity was the divergent representation of some of the subtypes in the different species, the most pronounced example being the impressive expansion of the muscle-mTEC population in humans. What factors, then, regulate the abundance of a mimetic-cell subtype? We first acknowledge that aspects of the post-Aire-mTEC isolation procedure (e.g., thymus dissociation) could potentially affect mimetic-cell representation in the scRNA-seq data. However, since the processing conditions were analogous for humans and mice,[4] it seems unlikely that the cell-extraction efficiency is the major contributing factor in differences between these two species. Another potential experimental factor might be that different stages of the life cycle were being compared, e.g., adult mice versus young humans. But it is important to keep in mind that thymus development is much more advanced at birth in humans than in mice. Indeed, we saw little evidence of age-related changes in children ranging from 2 days to 4 months of age, consistent with the very similar fractional representation (within the entire pool) of the various mimetic-cell subtypes detected in perinatal versus adult humans in another study.[14]

More likely, one or more biological factors probably come into play. One such factor might be that there has been evolutionary selection for mTEC subtypes that mimic peripheral cells not well covered by other tolerance mechanisms (although it is difficult to envisage how this kind of selection could operate). Another factor might be selection for mTEC subtypes that perform a new function in the thymus, in reflection of the activities of their peripheral counterparts. We already know, for example, that tuft mTECs regulate thymic generation of invariant natural killer (NK)-T cells and ILC2s[11,12,66,87] and that microfold mTECs influence the thymic immunoglobulin A (IgA) repertoire.[4,7] It is also possible that the molecular mechanics of mimetic-mTEC differentiation are an important factor. Accumulation of each mimetic-cell subtype depends on the induction of one or more genes encoding a lineage-defining TF, and several features of these genes could influence their propensity for transcriptional induction: enhancer:promoter distance, the abundance of Z-DNA in the promoter region,[88] other DNA-sequence features, and chromatin architecture. Lastly, it was recently shown that different lineage-promoting TFs have varying propensities, in and of themselves, to drive cell differentiation to a late or terminal stage.[89]

The mechanics underlying induced transcription of the genes encoding lineage-defining and cooperating TFs probably also explain some of the particularities of the eye-catching muscle-mTEC conglomerate. Several computational approaches argued for a differentiation trajectory reminiscent of that followed during skeletal-muscle differentiation. Yet transcription of some of the genes encoding key TFs (e.g., *PAX3/7* and *MYOD1*) did not seem to adhere to the classical trajectory even though late-stage muscle mTECs ended up having many transcriptional and morphological features (even striations!) in common with skeletal muscle. Also unanticipated was the diversity of mature-muscle markers that could be detected, including different types of skeletal and cardiac muscle. These findings highlight the laxity of the muscle differentiation process in the thymic context, raising the question of whether the mTEC component of the muscle-mTEC transcriptional chimera might be having an influence on muscle-related transcriptional programs. Alternatively, cell-intrinsic or -extrinsic factors normally present in muscle but absent or reduced in the thymus might more faithfully drive the classical muscle differentiation pathway or select for cells that are on track to the appropriate terminal differentiation state.

Certain features of human mimetic cells raise questions about their function and implications in disease. First, the only donor with a substantially different distribution of mimetic-cell subtypes was donor A, who possessed a relative enrichment of neuroendocrine mTECs and a relative paucity of muscle and ionocyte mTECs. This donor had Down syndrome (trisomy 21), which has been associated with changes in thymus morphology, altered AIRE expression, and autoimmunity.[90,91] Of note, AIRE is encoded on chromosome 21, leading to the hypothesis that Down syndrome may alter central tolerance via perturbed AIRE expression.[92] The changed mimetic-cell distributions of donor A suggest altered central tolerance in Down syndrome, although caution must be exercised in drawing conclusions from a single case.

Second, we observed overlap of some neuroendocrine and muscle mimetic cells, as well as striated muscle-mTEC substructures. We are intrigued by the possibility that muscle mTECs could exert a specialized function within the thymus or could exhibit behavior mirroring peripheral muscle, such as contraction. Indeed, spontaneous contraction of myoid cells in thymic stromal cultures was observed in the 1970–80s,[93–95] although the exact origin of these cells remains in question. We also observed rare colocalization of tuft mimetic cells, $DYSF^{+}$ cells, and alpha-bungarotoxin in the human thymus, raising the possibility that mimetic cells can engage in acetylcholine-mediated interfaces, mimicking peripheral neuromuscular junctions. This finding is of special interest because autoantibodies recognizing the AChR occur in around 80% of individuals diagnosed with MG.[21] In addition to these autoantibodies, individuals with MG often exhibit B cell infiltration and ectopic germinal centers in the thymus.[96,97] Indeed, it has been proposed that the thymus is the site of autoreactive B cell maturation and serves as a reservoir for them,[98,99] although the mechanism for such thymic maturation is unknown. We hypothesize that the acetylcholine-associated cellular interaction observed here may supply the autoimmunizing agent in MG. However, this notion remains in the realm of speculation until experimental validation can be provided. Establishing a mouse model lacking the AchR specifically in TECs is one long-term approach, assuming that mice will also show muscle-mTEC:AchR:tuft-mTEC colocalization analogous to that in humans. Characterizing the mimetic-cell compartment in MG patients is another approach, perhaps providing insights into the repertoire of,

and contributions of, mimetic cells in both this disease and autoimmunity more broadly.

In brief, our characterization of the human mimetic-cell compartment of pediatric donors provides a high-resolution look at the mimetic-mTEC repertoire of the human thymus, laying the foundation for future studies on mimetic cells in diseased states. The potential significance of such efforts is underlined by the importance of murine ciliated, muscle, entero-hepato, tuft, and endocrine mTECs in tolerization to neo- or endogenous antigens.[4,7,8,11]

Limitations of the study

We chose to sort the human post-Aire MHC class II^{lo} mTEC population for this study because the strong mimetic-cell enrichment afforded by employing this approach with mice had revealed mTEC subtypes that were missing or too rare to study in previous analyses. However, one might question whether the strong mouse:human differences in distribution we observed actually reflected differential expression of sorting markers in the two species. We do not think this is the case, however. Focused re-analysis of two recent human scRNA-seq studies wherein the investigators surveyed the entire mTEC compartment[14,17] revealed that—just as in our study—muscle and neuroendocrine mTECs were dominant mimetic-cell subtypes; keratinocyte and tuft mTECs were among the least abundant subtypes; and microfold mTECs were not found. Thus, the differential mTEC subtype distributions we reported cannot be attributed simply to our sorting protocol.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Diane Mathis (dm@hms.harvard.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

scRNA-seq and population RNA-seq datasets generated in this study are available at the Gene Expression Omnibus (GEO): human population RNA-seq (GEO: GSE261700), human scRNA-seq (GEO: GSE262749), and zebrafish scRNA-seq, including reanalyzed data from Rubin et al.[72] (GEO: GSE259418). Published datasets utilized in this study: mouse population RNA-seq data from DP.mTEC^{lo}, mTEC^{hi}, and DN.mTEC^{lo} were previously published[4] (GEO: GSE194253). Mouse scRNA-seq data were previously published[4] (GEO: GSE194252).
Muscle signatures were generated from published sequencing data: data from McKellar et al. was downloaded from Dryad (https://doi.org/10.5061/dryad.t4b8gtj34),[22] and data from Feng et al.[28] was downloaded from GEO (GEO: GSE193346).
Scripts from major analyses are available on GitHub: https://github.com/huismanb/human-mimetic-cells.

ACKNOWLEDGMENTS

We thank I. Magill, J. Lee, and the Broad Genomics platform for RNA-seq; the Harvard Medical School (HMS) and Boston Children's Hospital (BCH) flow cytometry cores; the HMS MicRoN core; the Brigham and Women's Hospital Pathology Cores; BCH Aquatics Research Facilities staff for their assistance with fish maintenance; A. Ludin Tal for *Tg(lck:mCherry)* zebrafish; the Langenau Lab for *Tg(lck:GFP; mylpfa:mCherry)* zebrafish; D. Mallah, L. Yang, and S. Yang for assistance with data processing; Dr. C. Laplace for graphics; Dr. A. Ortiz-Lopez for experimental help; A. Corbin for contributions to experimental design and technical expertise; and Dr. Sasha Opotowsky for helpful discussions. This work was supported by NIH grants R01DK060027 and R01AI088204 (to D.M.), R01HL144780 (to L.Z.), T32GM144273 (for D.A.M.), T32GM007753 (for S.A.R.), and F30HL152628 (to S.A.R.); a Damon Runyon Cancer Research Fellowship (DRG-2519-24 to B.D.H.); and an anonymous donation.

AUTHOR CONTRIBUTIONS

Conceptualization: B.D.H., D.A.M., and D.M.; conducted experiments: B.D.H., D.A.M., S.A.R., and W.G.; performed data analysis: B.D.H., D.A.M., and Y.F.; supervision: C.B., L.Z., and D.M.; provision of human samples: K.K., J.M.L., P.d.N., and M.N.; wrote manuscript: B.D.H., D.A.M., and D.M.; edited manuscript: B.D.H., D.A.M., S.A.R., M.N., C.B., L.Z., and D.M.; funding acquisition: D.M. and L.Z.

DECLARATION OF INTERESTS

D.M. is a cofounder and is on the scientific advisory board of Zag Therapeutics, Inc.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Human samples
  - Zebrafish
- METHOD DETAILS
  - Human thymus processing
  - Human thymus flow cytometry
  - Human population-level RNA-seq
  - Human scRNA-seq library preparation
  - Human scRNA-seq analysis
  - Human thymus histology
  - Zebrafish tissue collection and cell sorting
  - Zebrafish scRNA-seq library preparation
  - Zebrafish scRNA-seq analysis
  - Imaging of transgenic zebrafish
  - Zebrafish immunohistochemistry
  - Evolutionary tree visualization
  - Fine-tuning scGPT on human mTEC datasets
  - Applying fine-tuned scGPT model
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.immuni.2024.11.025.

Received: May 22, 2024
Revised: August 19, 2024
Accepted: November 27, 2024
Published: December 27, 2024

REFERENCES

1. Abramson, J., and Anderson, G. (2017). Thymic epithelial cells. Annu. Rev. Immunol. *35*, 85–118. https://doi.org/10.1146/annurev-immunol-051116-052320.
2. Anderson, M.S., Venanzi, E.S., Klein, L., Chen, Z., Berzins, S.P., Turley, S.J., von Boehmer, H., Bronson, R., Dierich, A., Benoist, C., et al. (2002). Projection of an immunological self shadow within the thymus

by the aire protein. Science *298*, 1395–1401. https://doi.org/10.1126/science.1075958.

3. Anderson, M.S., and Su, M.A. (2016). AIRE expands: new roles in immune tolerance and beyond. Nat. Rev. Immunol. *16*, 247–258. https://doi.org/10.1038/nri.2016.9.
4. Michelson, D.A., Hase, K., Kaisho, T., Benoist, C., and Mathis, D. (2022). Thymic epithelial cells co-opt lineage-defining transcription factors to eliminate autoreactive T cells. Cell *185*, 2542–2558.e18. https://doi.org/10.1016/j.cell.2022.05.018.
5. Michelson, D.A., and Mathis, D. (2022). Thymic mimetic cells: tolerogenic masqueraders. Trends Immunol. *43*, 782–791. https://doi.org/10.1016/j.it.2022.07.010.
6. Michelson, D.A., and Mathis, D. (2024). Thymic mimetic cells: ontogeny as immunology. Annu. Rev. Cell Dev. Biol. *40*, 283–300. https://doi.org/10.1146/annurev-cellbio-112122-023316.
7. Givony, T., Leshkowitz, D., Del Castillo, D., Nevo, S., Kadouri, N., Dassa, B., Gruper, Y., Khalaila, R., Ben-Nun, O., Gome, T., et al. (2023). Thymic mimetic cells function beyond self-tolerance. Nature *622*, 164–172. https://doi.org/10.1038/s41586-023-06512-8.
8. Michelson, D.A., Zuo, C., Verzi, M., Benoist, C., and Mathis, D. (2023). Hnf4 activates mimetic-cell enhancers to recapitulate gut and liver development within the thymus. J. Exp. Med. *220*, e20230461. https://doi.org/10.1084/jem.20230461.
9. Hassall, A.H. (1846). The Microscopic Anatomy of the Human Body: in Health and Disease (S. Highley).
10. Farr, A.G., and Rudensky, A. (1998). Medullary thymic epithelium: a mosaic of epithelial "self"? J. Exp. Med. *188*, 1–4. https://doi.org/10.1084/jem.188.1.1.
11. Miller, C.N., Proekt, I., von Moltke, J., Wells, K.L., Rajpurkar, A.R., Wang, H., Rattay, K., Khan, I.S., Metzger, T.C., Pollack, J.L., et al. (2018). Thymic tuft cells promote an IL-4-enriched medulla and shape thymocyte development. Nature *559*, 627–631. https://doi.org/10.1038/s41586-018-0345-2.
12. Bornstein, C., Nevo, S., Giladi, A., Kadouri, N., Pouzolles, M., Gerbe, F., David, E., Machado, A., Chuprin, A., Tóth, B., et al. (2018). Single-cell mapping of the thymic stroma identifies IL-25-producing tuft epithelial cells. Nature *559*, 622–626. https://doi.org/10.1038/s41586-018-0346-1.
13. Park, J.E., Botting, R.A., Domínguez Conde, C.C., Popescu, D.M., Lavaert, M., Kunz, D.J., Goh, I., Stephenson, E., Ragazzini, R., Tuck, E., et al. (2020). A cell atlas of human thymic development defines T cell repertoire formation. Science *367*, eaay3224. https://doi.org/10.1126/science.aay3224.
14. Bautista, J.L., Cramer, N.T., Miller, C.N., Chavez, J., Berrios, D.I., Byrnes, L.E., Germino, J., Ntranos, V., Sneddon, J.B., Burt, T.D., et al. (2021). Single-cell transcriptional profiling of human thymic stroma uncovers novel cellular heterogeneity in the thymic medulla. Nat. Commun. *12*, 1096. https://doi.org/10.1038/s41467-021-21346-6.
15. Dhalla, F., Baran-Gale, J., Maio, S., Chappell, L., Holländer, G.A., and Ponting, C.P. (2020). Biologically indeterminate yet ordered promiscuous gene expression in single medullary thymic epithelial cells. EMBO J. *39*, e101828. https://doi.org/10.15252/embj.2019101828.
16. Baran-Gale, J., Morgan, M.D., Maio, S., Dhalla, F., Calvo-Asensio, I., Deadman, M.E., Handel, A.E., Maynard, A., Chen, S., Green, F., et al. (2020). Ageing compromises mouse thymus function and remodels epithelial cell differentiation. eLife *9*, e56221. https://doi.org/10.7554/eLife.56221.
17. Ragazzini, R., Boeing, S., Zanieri, L., Green, M., D'Agostino, G., Bartolovic, K., Agua-Doce, A., Greco, M., Watson, S.A., Batsivari, A., et al. (2023). Defining the identity and the niches of epithelial stem cells with highly pleiotropic multilineage potency in the human thymus. Dev. Cell *58*, 2428–2446.e9. https://doi.org/10.1016/j.devcel.2023.08.017.
18. Yayon, N., Kedlian, V.R., Boehme, L., Suo, C., Wachter, B., Beuschel, R.T., Amsalem, O., Polanski, K., Koplev, S., Tuck, E., et al. (2023). A spatial human thymus cell atlas mapped to a continuous tissue axis. Nature *635*, 708–718.
19. Metzger, T.C., Khan, I.S., Gardner, J.M., Mouchess, M.L., Johannes, K.P., Krawisz, A.K., Skrzypczynska, K.M., and Anderson, M.S. (2013). Lineage tracing and cell ablation identify a post-Aire-expressing thymic epithelial cell population. Cell Rep. *5*, 166–179. https://doi.org/10.1016/j.celrep.2013.08.038.
20. Hitachi, K., Honda, M., and Tsuchida, K. (2022). The functional role of long non-coding RNA in myogenesis and skeletal muscle atrophy. Cells *11*, 2291. https://doi.org/10.3390/cells11152291.
21. Gilhus, N.E., Tzartos, S., Evoli, A., Palace, J., Burns, T.M., and Verschuuren, J.J.G.M. (2019). Myasthenia gravis. Nat. Rev. Dis. Primers *5*, 30. https://doi.org/10.1038/s41572-019-0079-y.
22. McKellar, D.W., Walter, L.D., Song, L.T., Mantri, M., Wang, M.F.Z., De Vlaminck, I., and Cosgrove, B.D. (2021). Large-scale integration of single-cell transcriptomic data captures transitional progenitor states in mouse skeletal muscle regeneration. Commun. Biol. *4*, 1280. https://doi.org/10.1038/s42003-021-02810-x.
23. Wells, K.L., Miller, C.N., Gschwind, A.R., Wei, W., Phipps, J.D., Anderson, M.S., and Steinmetz, L.M. (2020). Combined transient ablation and single-cell RNA-sequencing reveals the development of medullary thymic epithelial cells. eLife *9*, e60188. https://doi.org/10.7554/eLife.60188.
24. Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T.R., and Weirauch, M.T. (2018). The human transcription factors. Cell *172*, 650–665. https://doi.org/10.1016/j.cell.2018.01.029.
25. Imbriano, C., and Molinari, S. (2018). Alternative splicing of transcription factors genes in muscle physiology and pathology. Genes (Basel) *9*, 107. https://doi.org/10.3390/genes9020107.
26. Lazure, F., Blackburn, D.M., Corchado, A.H., Sahinyan, K., Karam, N., Sharanek, A., Nguyen, D., Lepper, C., Najafabadi, H.S., Perkins, T.J., et al. (2020). Myf6/MRF4 is a myogenic niche regulator required for the maintenance of the muscle stem cell pool. EMBO Rep. *21*, e49499. https://doi.org/10.15252/embr.201949499.
27. Andrés, V., and Walsh, K. (1996). Myogenin expression, cell cycle withdrawal, and phenotypic differentiation are temporally separable events that precede cell fusion upon myogenesis. J. Cell Biol. *132*, 657–666. https://doi.org/10.1083/jcb.132.4.657.
28. Feng, W., Bais, A., He, H., Rios, C., Jiang, S., Xu, J., Chang, C., Kostka, D., and Li, G. (2022). Single-cell transcriptomic analysis identifies murine heart molecular features at embryonic and neonatal stages. Nat. Commun. *13*, 7960. https://doi.org/10.1038/s41467-022-35691-7.
29. Hale, L.P., and Markert, M.L. (2004). Corticosteroids regulate epithelial cell differentiation and Hassall body formation in the human thymus. J. Immunol. *172*, 617–624. https://doi.org/10.4049/jimmunol.172.1.617.
30. Greene, L.A., Sytkowski, A.J., Vogel, Z., and Nirenberg, M.W. (1973). α-Bungarotoxin used as a probe for acetylcholine receptors of cultured neurones. Nature *243*, 163–166. https://doi.org/10.1038/243163a0.
31. O'Leary, C.E., Schneider, C., and Locksley, R.M. (2019). Tuft cells-systemically dispersed sensory epithelia integrating immune and neural circuitry. Annu. Rev. Immunol. *37*, 47–72. https://doi.org/10.1146/annurev-immunol-042718-041505.
32. Atkinson, P.J., Huarcaya Najarro, E., Sayyid, Z.N., and Cheng, A.G. (2015). Sensory hair cell development and regeneration: similarities and differences. Development *142*, 1561–1571. https://doi.org/10.1242/dev.114926.
33. Burton, Q., Cole, L.K., Mulheisen, M., Chang, W., and Wu, D.K. (2004). The role of Pax2 in mouse inner ear development. Dev. Biol. *272*, 161–175. https://doi.org/10.1016/j.ydbio.2004.04.024.
34. Kopecky, B., and Fritzsch, B. (2011). Regeneration of hair cells: making sense of all the noise. Pharmaceuticals (Basel) *4*, 848–879. https://doi.org/10.3390/ph4060848.

35. Hertzano, R., Dror, A.A., Montcouquiol, M., Ahmed, Z.M., Ellsworth, B., Camper, S., Friedman, T.B., Kelley, M.W., and Avraham, K.B. (2007). Lhx3, a LIM domain transcription factor, is regulated by Pou4f3 in the auditory but not in the vestibular system. Eur. J. Neurosci. *25*, 999–1005. https://doi.org/10.1111/j.1460-9568.2007.05332.x.

36. Hertzano, R., Montcouquiol, M., Rashi-Elkeles, S., Elkon, R., Yücel, R., Frankel, W.N., Rechavi, G., Möröy, T., Friedman, T.B., Kelley, M.W., et al. (2004). Transcription profiling of inner ears from *Pou4f3*$^{ddl/ddl}$ identifies Gfi1 as a target of the Pou4f3 deafness gene. Hum. Mol. Genet. *13*, 2143–2153. https://doi.org/10.1093/hmg/ddh218.

37. Hou, K., Jiang, H., Karim, M.R., Zhong, C., Xu, Z., Liu, L., Guan, M., Shao, J., and Huang, X. (2019). A critical E-box in *Barhl1* 3′ enhancer is essential for auditory hair cell differentiation. Cells *8*, 458. https://doi.org/10.3390/cells8050458.

38. Masuda, M., Dulon, D., Pak, K., Mullen, L.M., Li, Y., Erkman, L., and Ryan, A.F. (2011). Regulation of POU4F3 gene expression in hair cells by 5′ DNA in mice. Neuroscience *197*, 48–64. https://doi.org/10.1016/j.neuroscience.2011.09.033.

39. Yasunaga, S., Grati, M., Cohen-Salmon, M., El-Amraoui, A., Mustapha, M., Salem, N., El-Zir, E., Loiselet, J., and Petit, C. (1999). A mutation in OTOF, encoding otoferlin, a FER-1-like protein, causes DFNB9, a non-syndromic form of deafness. Nat. Genet. *21*, 363–369. https://doi.org/10.1038/7693.

40. Zaw, K., Carvalho, L.S., Aung-Htut, M.T., Fletcher, S., Wilton, S.D., Chen, F.K., and McLenachan, S. (2022). Pathogenesis and treatment of Usher syndrome type IIA. Asia Pac. J. Ophthalmol. (Phila) *11*, 369–379. https://doi.org/10.1097/APO.0000000000000546.

41. Sousa, V.H., Miyoshi, G., Hjerling-Leffler, J., Karayannis, T., and Fishell, G. (2009). Characterization of Nkx6-2-derived neocortical interneuron lineages. Cereb. Cortex *19*, i1–i10. https://doi.org/10.1093/cercor/bhp038.

42. Briscoe, J., Sussel, L., Serup, P., Hartigan-O'Connor, D., Jessell, T.M., Rubenstein, J.L., and Ericson, J. (1999). Homeobox gene Nkx2.2 and specification of neuronal identity by graded Sonic hedgehog signalling. Nature *398*, 622–627. https://doi.org/10.1038/19315.

43. Chen, J., Yen, A., Florian, C.P., and Dougherty, J.D. (2022). MYT1L in the making: emerging insights on functions of a neurodevelopmental disorder gene. Transl. Psychiatry *12*, 292. https://doi.org/10.1038/s41398-022-02058-x.

44. Rosin, J.M., Kurrasch, D.M., and Cobb, J. (2015). Shox2 is required for the proper development of the facial motor nucleus and the establishment of the facial nerves. BMC Neurosci. *16*, 39. https://doi.org/10.1186/s12868-015-0176-0.

45. Dubreuil, V., Hirsch, M.R., Jouve, C., Brunet, J.F., and Goridis, C. (2002). The role of Phox2b in synchronizing pan-neuronal and type-specific aspects of neurogenesis. Development *129*, 5241–5253. https://doi.org/10.1242/dev.129.22.5241.

46. Pattyn, A., Hirsch, M., Goridis, C., and Brunet, J.F. (2000). Control of hindbrain motor neuron differentiation by the homeobox gene Phox2b. Development *127*, 1349–1358. https://doi.org/10.1242/dev.127.7.1349.

47. Franco, S.J., Gil-Sanz, C., Martinez-Garay, I., Espinosa, A., Harkins-Perry, S.R., Ramos, C., and Müller, U. (2012). Fate-restricted neural progenitors in the mammalian cerebral cortex. Science *337*, 746–749. https://doi.org/10.1126/science.1223616.

48. Guo, C., Eckler, M.J., McKenna, W.L., McKinsey, G.L., Rubenstein, J.L.R., and Chen, B. (2013). Fezf2 expression identifies a multipotent progenitor for neocortical projection neurons, astrocytes, and oligodendrocytes. Neuron *80*, 1167–1174. https://doi.org/10.1016/j.neuron.2013.09.037.

49. Takaba, H., Morishita, Y., Tomofuji, Y., Danks, L., Nitta, T., Komatsu, N., Kodama, T., and Takayanagi, H. (2015). Fezf2 orchestrates a thymic program of self-antigen expression for immune tolerance. Cell *163*, 975–987. https://doi.org/10.1016/j.cell.2015.10.013.

50. Saito, T., Greenwood, A., Sun, Q., and Anderson, D.J. (1995). Identification by differential RT-PCR of a novel paired homeodomain protein specifically expressed in sensory neurons and a subset of their CNS targets. Mol. Cell. Neurosci. *6*, 280–292. https://doi.org/10.1006/mcne.1995.1022.

51. Nguyen, M.Q., von Buchholtz, L.J., Reker, A.N., Ryba, N.J., and Davidson, S. (2021). Single-nucleus transcriptomic analysis of human dorsal root ganglion neurons. eLife *10*, e71752. https://doi.org/10.7554/eLife.71752.

52. Pou Casellas, C., Pleguezuelos-Manzano, C., Rookmaaker, M.B., Verhaar, M.C., and Clevers, H. (2023). Transcriptomic profile comparison reveals conservation of ionocytes across multiple organs. Sci. Rep. *13*, 3516. https://doi.org/10.1038/s41598-023-30603-1.

53. Purkerson, J.M., Tsuruoka, S., Suter, D.Z., Nakamori, A., and Schwartz, G.J. (2010). Adaptation to metabolic acidosis and its recovery are associated with changes in anion exchanger distribution and expression in the cortical collecting duct. Kidney Int. *78*, 993–1005. https://doi.org/10.1038/ki.2010.195.

54. Montoro, D.T., Haber, A.L., Biton, M., Vinarsky, V., Lin, B., Birket, S.E., Yuan, F., Chen, S., Leung, H.M., Villoria, J., et al. (2018). A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. Nature *560*, 319–324. https://doi.org/10.1038/s41586-018-0393-7.

55. Plasschaert, L.W., Žilionis, R., Choo-Wing, R., Savova, V., Knehr, J., Roma, G., Klein, A.M., and Jaffe, A.B. (2018). A single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte. Nature *560*, 377–381. https://doi.org/10.1038/s41586-018-0394-6.

56. Riordan, J.R., Rommens, J.M., Kerem, B., Alon, N., Rozmahel, R., Grzelczak, Z., Zielenski, J., Lok, S., Plavsic, N., and Chou, J.L. (1989). Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. Science *245*, 1066–1073. https://doi.org/10.1126/science.2475911.

57. Shah, V.S., Chivukula, R.R., Lin, B., Waghray, A., and Rajagopal, J. (2022). Cystic fibrosis and the cells of the airway epithelium: what are Ionocytes and what do they do? Annu. Rev. Pathol. *17*, 23–46. https://doi.org/10.1146/annurev-pathol-042420-094031.

58. Chen, L., Lee, J.W., Chou, C.L., Nair, A.V., Battistone, M.A., Păunescu, T.G., Merkulova, M., Breton, S., Verlander, J.W., Wall, S.M., et al. (2017). Transcriptomes of major renal collecting duct cell types in mouse identified by single-cell RNA-seq. Proc. Natl. Acad. Sci. USA *114*, E9989–E9998. https://doi.org/10.1073/pnas.1710964114.

59. Werth, M., Schmidt-Ott, K.M., Leete, T., Qiu, A., Hinze, C., Viltard, M., Paragas, N., Shawber, C.J., Yu, W., Lee, P., et al. (2017). Transcription factor *TFCP2L1* patterns cells in the mouse kidney collecting ducts. eLife *6*, e24265. https://doi.org/10.7554/eLife.24265.

60. Mauduit, O., Aure, M.H., Delcroix, V., Basova, L., Srivastava, A., Umazume, T., Mays, J.W., Bellusci, S., Tucker, A.S., Hajihosseini, M.K., et al. (2022). A mesenchymal to epithelial switch in Fgf10 expression specifies an evolutionary-conserved population of ionocytes in salivary glands. Cell Rep. *39*, 110663. https://doi.org/10.1016/j.celrep.2022.110663.

61. Gauer, S., Sichler, O., Obermüller, N., Holzmann, Y., Kiss, E., Sobkowiak, E., Pfeilschifter, J., Geiger, H., Mühl, H., and Hauser, I.A. (2007). IL-18 is expressed in the intercalated cell of human kidney. Kidney Int. *72*, 1081–1087. https://doi.org/10.1038/sj.ki.5002473.

62. Du, Y., Gao, H., He, C., Xin, S., Wang, B., Zhang, S., Gong, F., Yu, X., Pan, L., Sun, F., et al. (2022). An update on the biological characteristics and functions of tuft cells in the gut. Front. Cell Dev. Biol. *10*, 1102978. https://doi.org/10.3389/fcell.2022.1102978.

63. Gerbe, F., Sidot, E., Smyth, D.J., Ohmoto, M., Matsumoto, I., Dardalhon, V., Cesses, P., Garnier, L., Pouzolles, M., Brulin, B., et al. (2016). Intestinal epithelial tuft cells initiate type 2 mucosal immunity to helminth parasites. Nature *529*, 226–230. https://doi.org/10.1038/nature16527.

64. Howitt, M.R., Lavoie, S., Michaud, M., Blum, A.M., Tran, S.V., Weinstock, J.V., Gallini, C.A., Redding, K., Margolskee, R.F., Osborne, L.C., et al. (2016). Tuft cells, taste-chemosensory cells, orchestrate parasite type 2 immunity in the gut. Science *351*, 1329–1333. https://doi.org/10.1126/science.aaf1648.

65. Schneider, C., O'Leary, C.E., and Locksley, R.M. (2019). Regulation of immune responses by tuft cells. Nat. Rev. Immunol. *19*, 584–593. https://doi.org/10.1038/s41577-019-0176-x.

66. Lucas, B., White, A.J., Cosway, E.J., Parnell, S.M., James, K.D., Jones, N.D., Ohigashi, I., Takahama, Y., Jenkinson, W.E., and Anderson, G. (2020). Diversity in medullary thymic epithelial cells controls the activity and availability of iNKT cells. Nat. Commun. *11*, 2198. https://doi.org/10.1038/s41467-020-16041-x.

67. Panneck, A.R., Rafiq, A., Schütz, B., Soultanova, A., Deckmann, K., Chubanov, V., Gudermann, T., Weihe, E., Krasteva-Christ, G., Grau, V., et al. (2014). Cholinergic epithelial cell with chemosensory traits in murine thymic medulla. Cell Tissue Res. *358*, 737–748. https://doi.org/10.1007/s00441-014-2002-x.

68. Strine, M.S., and Wilen, C.B. (2022). Tuft cells are key mediators of interkingdom interactions at mucosal barrier surfaces. PLoS Pathog. *18*, e1010318. https://doi.org/10.1371/journal.ppat.1010318.

69. Edqvist, P.H.D., Fagerberg, L., Hallström, B.M., Danielsson, A., Edlund, K., Uhlén, M., and Pontén, F. (2015). Expression of human skin-specific genes defined by transcriptomics and antibody-based profiling. J. Histochem. Cytochem. *63*, 129–141. https://doi.org/10.1369/0022155414562646.

70. Kumar, S., and Hedges, S.B. (1998). A molecular timescale for vertebrate evolution. Nature *392*, 917–920. https://doi.org/10.1038/31927.

71. Woods, I.G., Kelly, P.D., Chu, F., Ngo-Hazelett, P., Yan, Y.L., Huang, H., Postlethwait, J.H., and Talbot, W.S. (2000). A comparative map of the zebrafish genome. Genome Res. *10*, 1903–1914. https://doi.org/10.1101/gr.10.12.1903.

72. Rubin, S.A., Baron, C.S., Pessoa Rodrigues, C., Duran, M., Corbin, A.F., Yang, S.P., Trapnell, C., and Zon, L.I. (2022). Single-cell analyses reveal early thymic progenitors and pre-B cells in zebrafish. J. Exp. Med. *219*, e20220038. https://doi.org/10.1084/jem.20220038.

73. Cui, H., Wang, C., Maan, H., Pang, K., Luo, F., Duan, N., and Wang, B. (2024). scGPT: toward building a foundation model for single-cell multi-omics using generative AI. Nat. Methods *21*, 1470–1480. https://doi.org/10.1038/s41592-024-02201-0.

74. Kumar, S., Suleski, M., Craig, J.M., Kasprowicz, A.E., Sanderford, M., Li, M., Stecher, G., and Hedges, S.B. (2022). TimeTree 5: an expanded resource for species divergence times. Mol. Biol. Evol. *39*, msac174. https://doi.org/10.1093/molbev/msac174.

75. Lin, L.Y., Horng, J.L., Kunkel, J.G., and Hwang, P.P. (2006). Proton pump-rich cell secretes acid in skin of zebrafish larvae. Am. J. Physiol. Cell Physiol. *290*, C371–C378. https://doi.org/10.1152/ajpcell.00281.2005.

76. Hwang, P.P., and Chou, M.Y. (2013). Zebrafish as an animal model to study ion homeostasis. Pflugers Arch. *465*, 1233–1247. https://doi.org/10.1007/s00424-013-1269-1.

77. Pan, T.C., Liao, B.K., Huang, C.J., Lin, L.Y., and Hwang, P.P. (2005). Epithelial Ca^{2+} channel expression and Ca^{2+} uptake in developing zebrafish. Am. J. Physiol. Regul. Integr. Comp. Physiol. *289*, R1202–R1211. https://doi.org/10.1152/ajpregu.00816.2004.

78. Hwang, P.P. (2009). Ion uptake and acid secretion in zebrafish (Danio rerio). J. Exp. Biol. *212*, 1745–1752. https://doi.org/10.1242/jeb.026054.

79. Chang, W.J., and Hwang, P.P. (2011). Development of zebrafish epidermis. Birth Defects Res. C Embryo Today *93*, 205–214. https://doi.org/10.1002/bdrc.20215.

80. Pistocchi, A., Feijóo, C.G., Cabrera, P., Villablanca, E.J., Allende, M.L., and Cotelli, F. (2009). The zebrafish Prospero homolog prox1 is required for mechanosensory hair cell differentiation and functionality in the lateral line. BMC Dev. Biol. *9*, 58. https://doi.org/10.1186/1471-213X-9-58.

81. Hernández, P.P., Olivari, F.A., Sarrazin, A.F., Sandoval, P.C., and Allende, M.L. (2007). Regeneration in zebrafish lateral line neuromasts: expression of the neural progenitor cell marker sox2 and proliferation-dependent and-independent mechanisms of hair cell renewal. Dev. Neurobiol. *67*, 637–654. https://doi.org/10.1002/dneu.20386.

82. Sato, Y., Miyasaka, N., and Yoshihara, Y. (2005). Mutually exclusive glomerular innervation by two distinct types of olfactory sensory neurons revealed in transgenic zebrafish. J. Neurosci. *25*, 4889–4897. https://doi.org/10.1523/JNEUROSCI.0679-05.2005.

83. Shi, T., Beaulieu, M.O., Saunders, L.M., Fabian, P., Trapnell, C., Segil, N., Crump, J.G., and Raible, D.W. (2023). Single-cell transcriptomic profiling of the zebrafish inner ear reveals molecularly distinct hair cell and supporting cell subtypes. eLife *12*, e82978. https://doi.org/10.7554/eLife.82978.

84. Pang, L., Zhang, Z., Shen, Y., Cheng, Z., Gao, X., Zhang, B., Wang, X., and Tian, H. (2020). Mutant dlx3b disturbs normal tooth mineralization and bone formation in zebrafish. PeerJ *8*, e8515. https://doi.org/10.7717/peerj.8515.

85. Zhou, Q., Zhao, C., Yang, Z., Qu, R., Li, Y., Fan, Y., Tang, J., Xie, T., and Wen, Z. (2023). Cross-organ single-cell transcriptome profiling reveals macrophage and dendritic cell heterogeneity in zebrafish. Cell Rep. *42*, 112793. https://doi.org/10.1016/j.celrep.2023.112793.

86. Lin, X., Zhou, Q., Zhao, C., Lin, G., Xu, J., and Wen, Z. (2019). An ectoderm-derived myeloid-like cell population functions as antigen transporters for langerhans cells in zebrafish epidermis. Dev. Cell *49*, 605–617.e5. https://doi.org/10.1016/j.devcel.2019.03.028.

87. Nevo, S., Frenkel, N., Kadouri, N., Gome, T., Rosenthal, N., Givony, T., Avin, A., Peligero Cruz, C.C., Kedmi, M., Lindzen, M., et al. (2024). Tuft cells and fibroblasts promote thymus regeneration through ILC2-mediated type 2 immune response. Sci. Immunol. *9*, eabq6930. https://doi.org/10.1126/sciimmunol.abq6930.

88. Fang, Y., Bansal, K., Mostafavi, S., Benoist, C., and Mathis, D. (2024). AIRE relies on Z-DNA to flag gene targets for thymic T cell tolerization. Nature *628*, 400–407. https://doi.org/10.1038/s41586-024-07169-7.

89. Joung, J., Ma, S., Tay, T., Geiger-Schuller, K.R., Kirchgatterer, P.C., Verdine, V.K., Guo, B., Arias-Garcia, M.A., Allen, W.E., Singh, A., et al. (2023). A transcription factor atlas of directed differentiation. Cell *186*, 209–229.e26. https://doi.org/10.1016/j.cell.2022.11.026.

90. Bull, M.J. (2020). Down Syndrome. N. Engl. J. Med. *382*, 2344–2352. https://doi.org/10.1056/NEJMra1706537.

91. Marcovecchio, G.E., Bortolomai, I., Ferrua, F., Fontana, E., Imberti, L., Conforti, E., Amodio, D., Bergante, S., Macchiarulo, G., D'Oria, V., et al. (2019). Thymic epithelium abnormalities in DiGeorge and Down syndrome patients contribute to dysregulation in T cell development. Front. Immunol. *10*, 447. https://doi.org/10.3389/fimmu.2019.00447.

92. Giménez-Barcons, M., Casteràs, A., Armengol, M.P., Porta, E., Correa, P.A., Marín, A., Pujol-Borrell, R., and Colobran, R. (2014). Autoimmune predisposition in Down syndrome may result from a partial central tolerance failure due to insufficient intrathymic expression of AIRE and peripheral antigens. J. Immunol. *193*, 3872–3879. https://doi.org/10.4049/jimmunol.1400223.

93. Itoh, T. (1983). Establishment of a myoid cell clone from rat thymus. Cell Tissue Res. *231*, 39–47. https://doi.org/10.1007/BF00215772.

94. Wekerle, H., Cohen, I.R., and Feldman, M. (1973). Thymus reticulum cell cultures confer T cell properties on spleen cells from thymus-deprived animals. Eur. J. Immunol. *3*, 745–748. https://doi.org/10.1002/eji.1830031202.

95. Wekerle, T.H., Paterson, B., Ketelsen, U., and Feldman, M. (1975). Striated muscle fibres differentiate in monolayer cultures of adult thymus reticulum. Nature *256*, 493–494. https://doi.org/10.1038/256493a0.

96. Berrih-Aknin, S., and Le Panse, R. (2014). Myasthenia gravis: a comprehensive review of immune dysregulation and etiological mechanisms. J. Autoimmun. *52*, 90–100. https://doi.org/10.1016/j.jaut.2013.12.011.

97. Lefeuvre, C.M., Payet, C.A., Fayet, O.M., Maillard, S., Truffault, F., Bondet, V., Duffy, D., de Montpreville, V., Ghigna, M.R., Fadel, E., et al. (2020). Risk factors associated with myasthenia gravis in thymoma patients: the potential role of thymic germinal centers. J. Autoimmun. *106*, 102337. https://doi.org/10.1016/j.jaut.2019.102337.

98. Okumura, M., Ohta, M., Takeuchi, Y., Shiono, H., Inoue, M., Fukuhara, K., Kadota, Y., Miyoshi, S., Fujii, Y., and Matsuda, H. (2003). The immunologic role of thymectomy in the treatment of myasthenia gravis: implication of

thymus-associated B-lymphocyte subset in reduction of the anti-acetylcholine receptor antibody titer. J. Thorac. Cardiovasc. Surg. *126*, 1922–1928. https://doi.org/10.1016/s0022-5223(03)00938-3.

99. Jiang, R., Hoehn, K.B., Lee, C.S., Pham, M.C., Homer, R.J., Detterbeck, F.C., Aban, I., Jacobson, L., Vincent, A., Nowak, R.J., et al. (2020). Thymus-derived B cell clones persist in the circulation after thymectomy in myasthenia gravis. Proc. Natl. Acad. Sci. USA *117*, 30649–30660. https://doi.org/10.1073/pnas.2007206117.

100. Liu, X., Li, Y.S., Shinton, S.A., Rhodes, J., Tang, L., Feng, H., Jette, C.A., Look, A.T., Hayakawa, K., and Hardy, R.R. (2017). Zebrafish B cell development without a pre-B cell stage, revealed by CD79 fluorescence reporter transgenes. J. Immunol. *199*, 1706–1715. https://doi.org/10.4049/jimmunol.1700552.

101. White, R.M., Sessa, A., Burke, C., Bowman, T., LeBlanc, J., Ceol, C., Bourque, C., Dovey, M., Goessling, W., Burns, C.E., et al. (2008). Transparent adult zebrafish as a tool for in vivo transplantation analysis. Cell Stem Cell *2*, 183–189. https://doi.org/10.1016/j.stem.2007.11.002.

102. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.

103. Liao, Y., Smyth, G.K., and Shi, W. (2013). The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Res. *41*, e108. https://doi.org/10.1093/nar/gkt214.

104. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550. https://doi.org/10.1186/s13059-014-0550-8.

105. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., III, Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573–3587.e29. https://doi.org/10.1016/j.cell.2021.04.048.

106. McGinnis, C.S., Murrow, L.M., and Gartner, Z.J. (2019). DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. Cell Syst. *8*, 329–337.e4. https://doi.org/10.1016/j.cels.2019.03.003.

107. Melsted, P., Booeshaghi, A.S., Liu, L., Gao, F., Lu, L., Min, K.H.J., da Veiga Beltrame, E., Hjörleifsson, K.E., Gehring, J., and Pachter, L. (2021). Modular, efficient and constant-memory single-cell RNA-seq preprocessing. Nat. Biotechnol. *39*, 813–818. https://doi.org/10.1038/s41587-021-00870-2.

108. Bergen, V., Lange, M., Peidli, S., Wolf, F.A., and Theis, F.J. (2020). Generalizing RNA velocity to transient cell states through dynamical modeling. Nat. Biotechnol. *38*, 1408–1414. https://doi.org/10.1038/s41587-020-0591-3.

109. Lange, M., Bergen, V., Klein, M., Setty, M., Reuter, B., Bakhti, M., Lickert, H., Ansari, M., Schniering, J., Schiller, H.B., et al. (2022). CellRank for directed single-cell fate mapping. Nat. Methods *19*, 159–170. https://doi.org/10.1038/s41592-021-01346-6.

110. Stacklies, W., Redestig, H., Scholz, M., Walther, D., and Selbig, J. (2007). pcaMethods–a bioconductor package providing PCA methods for incomplete data. Bioinformatics *23*, 1164–1167. https://doi.org/10.1093/bioinformatics/btm069.

111. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods *9*, 676–682. https://doi.org/10.1038/nmeth.2019.

112. Volders, P.J., Anckaert, J., Verheggen, K., Nuytens, J., Martens, L., Mestdagh, P., and Vandesompele, J. (2019). LNCipedia 5: towards a reference set of human long non-coding RNAs. Nucleic Acids Res. *47*, D135–D139. https://doi.org/10.1093/nar/gky1031.

113. Frankish, A., Diekhans, M., Jungreis, I., Lagarde, J., Loveland, J.E., Mudge, J.M., Sisu, C., Wright, J.C., Armstrong, J., Barnes, I., et al. (2021). GENCODE 2021. Nucleic Acids Res. *49*, D916–D923. https://doi.org/10.1093/nar/gkaa1087.

114. La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastriti, M.E., Lönnerberg, P., Furlan, A., et al. (2018). RNA velocity of single cells. Nature *560*, 494–498. https://doi.org/10.1038/s41586-018-0414-6.

115. Wolf, F.A., Hamey, F.K., Plass, M., Solana, J., Dahlin, J.S., Göttgens, B., Rajewsky, N., Simon, L., and Theis, F.J. (2019). PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. Genome Biol. *20*, 59. https://doi.org/10.1186/s13059-019-1663-x.

116. Zheng, G.X.Y., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J., et al. (2017). Massively parallel digital transcriptional profiling of single cells. Nat. Commun. *8*, 14049. https://doi.org/10.1038/ncomms14049.

117. Song, Y., Miao, Z., Brazma, A., and Papatheodorou, I. (2023). Benchmarking strategies for cross-species integration of single-cell RNA sequencing data. Nat. Commun. *14*, 6495. https://doi.org/10.1038/s41467-023-41855-w.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse monoclonal anti-CD45, Brilliant Violet 605	BioLegend	Cat#304041; RRID: AB_2562105
Mouse monoclonal anti-EpCAM, APC-Cy7	BioLegend	Cat#324245; RRID: AB_2783193
Mouse monoclonal anti-HLA-DR, APC	BioLegend	Cat#307609; RRID: AB_314687
Mouse monoclonal anti-CD205, PE	BioLegend	Cat#342203; RRID: AB_1626209
Rat monoclonal anti-PDPN, PE-Cy7	BioLegend	Cat#337013; RRID: AB_2563367
Mouse monoclonal anti-CD104, FITC	BioLegend	Cat#327805; RRID: AB_893216
Rabbit polyclonal anti-Desmin	abcam	Cat#ab15200; RRID: AB_301744
Mouse monoclonal anti-Myogenin, PE	BD Biosciences	Cat#563120; RRID: AB_2738016
Mouse monoclonal anti-Myosin heavy chain (clone MF20)	R&D Systems	Cat#MAB4470; RRID: AB_1293549
Rabbit polyclonal anti-Dysferlin	Novus	Cat#NBP1-84696-25ul; RRID: AB_11039027
Goat polyclonal anti-ChAT	R&D	Cat#AF3447-SP; RRID: AB_2079603
Rabbit polyclonal anti-CADPS	Novus	Cat#NBP1-77323; RRID: AB_11028350
Rabbit monoclonal anti-SOX2	Cell Signaling Technology	Cat#23064S; RRID: AB_2714146
Rabbit polyclonal anti-S100	Abcam	Cat#ab868; RRID: AB_306716
Mouse monoclonal anti-Cytokeratin 10	Abcam	Cat#ab212546
Rabbit polyclonal anti-Dystrophin	Abcam	Cat#ab15277; RRID: AB_301813
Rabbit polyclonal anti-TRPM5	Proteintech	Cat#18027-1-AP; RRID: AB_2287825
Donkey polyclonal anti-rabbit IgG, FITC	Jackson ImmunoResearch	Cat#711-096-152; RRID: AB_2340597
Donkey polyclonal anti-rabbit IgG, Cy3	Jackson ImmunoResearch	Cat#711-166-152; RRID: AB_2313568
Donkey polyclonal anti-mouse IgG, Alexa Fluor 488	Invitrogen	Cat#A21202; RRID: AB_141607
Donkey polyclonal anti-rabbit IgG, Alexa Fluor 647	Jackson ImmunoResearch	Cat#711-606-152; RRID: AB_2340625
Donkey polyclonal anti-goat IgG, Alexa Fluor 647	Jackson ImmunoResearch	Cat#705-605-147; RRID: AB_2340437
Donkey polyclonal anti-mouse IgG, Cy5	Jackson ImmunoResearch	Cat#715-175-151;RRID: AB_2340820
Mouse monoclonal anti-Myosin 4 antibody (clone MF20)	ThermoFisher	Cat#14-6503-82; RRID: AB_2572894
Biological samples		
Human thymus tissue	Boston Children's Hospital	N/A
Chemicals, peptides, and recombinant proteins		
Collagenase, from Clostridium histolyticum	Sigma	Cat#6885
Collagenase/Dispase	RocheMillipore	Cat#10269638001
Deoxyribonuclease I from bovine pancreas	Sigma	Cat#DN25
Hoechst 33342	ThermoFisher	Cat#H3570
LIVE/DEAD Fixable Violet Cell Stain kit	Invitrogen	Cat#L34963
Normal donkey serum	Jackson ImmunoResearch	Cat#017-000-121
TCL RNA lysis buffer	Qiagen	Cat#1031576
Tween-20	Fisher Scientific	BP337-500
Alpha-bungarotoxin-CF488	VWR	Cat#76221-710
Prolong Diamond Antifade Mountant	Invitrogen	Cat#P36970
Liberase TM, medium Thermolysin concentration	Roche	Cat#5401119001
10% formalin	VWR	Cat#16004-121
Epitope antigen retrieval solution	Leica	Cat#AR9961
Leica Biosystems Refine Detection Kit	Leica	Cat#DS9800
HistoreCore SPECTRA CV mounting medium	Leica	Cat#3801733

(*Continued on next page*)

Continued		
REAGENT or RESOURCE	SOURCE	IDENTIFIER
Dewax deparaffinization solution	Leica	Cat#AR9222
Critical commercial assays		
Anti-EpCAM microbeads	Miltenyi	Cat#130-061-101
Anti-CD45 microbeads	Miltenyi	Cat#130-045-801
MACS LS columns	Miltenyi	Cat#130-042-401
Chromium Next Gem Single Cell 3' kit v3.1	10X Genomics	Cat#PN-1000128
Deposited data		
Raw and analyzed human population RNA-seq	This paper	GEO: GSE261700
Raw and analyzed human scRNA-seq	This paper	GEO: GSE262749
Raw and analyzed zebrafish scRNA-seq	This paper and Rubin et al.[72]	GEO: GSE259418
Mouse population RNA-seq	Michelson et al.[4]	GEO: GSE194253
Mouse scRNA-seq	Michelson et al.[4]	GEO: GSE194252
Skeletal muscle scRNA-seq	McKellar et al.[22]	https://doi.org/10.5061/dryad.t4b8gtj34
Cardiac muscle scRNA-seq	Feng et al.[28]	GEO: GSE193346
Experimental models: Organisms/strains		
Zebrafish: *Tg(lck:mCherry)*	Aya Ludin Tal, Leonard Zon Lab	N/A
Zebrafish: *Tg(cd79a:eGFP)*	Liu et al.[100]	N/A
Zebrafish: *Tg(lck:GFP; mylpfa:mCherry)*	David Langenau Lab	N/A
Zebrafish: Casper	White et al.[101]	N/A
Software and algorithms		
STAR v2.7.3a	Dobin et al.[102]	https://github.com/alexdobin/STAR/
Subread v2.0.0	Liao et al.[103]	https://subread.sourceforge.net/
DESeq2 v1.36.0	Love et al.[104]	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
Cell Ranger v7.0.0	10X Genomics	N/A
Seurat v4.3.0	Hao et al.[105]	https://satijalab.org/seurat/
DoubletFinder v2.0.3	McGinnis et al.[106]	https://github.com/chris-mcginnis-ucsf/DoubletFinder
kallisto-bustools v0.27.3	Melsted et al.[107]	https://www.kallistobus.tools/
scVelo v0.2.5	Bergen et al.[108]	https://scvelo.readthedocs.io/en/stable/
CellRank v1.5.1	Lange et al.[109]	https://cellrank.readthedocs.io/en/latest/#
FlowJo v10.8.2	BD Biosciences	N/A
Python 3.7.3	Python Software Foundation	https://www.python.org/
R v4.2.2	R Core Team	https://www.r-project.org/
matplotlib v3.5.3	The Matplotlib development team	https://matplotlib.org/
numpy v1.21.6	NumPy team	https://numpy.org/
scipy v1.4.1	SciPy development team	https://scipy.org/
pybiomart v0.2.0	N/A	https://anaconda.org/bioconda/pybiomart
pcaMethods v1.88.0	Stacklies et al.[110]	https://www.bioconductor.org/packages/release/bioc/html/pcaMethods.html
TimeTree5	Kumar et al.[74]	https://timetree.org/
Pre-trained scGPT model	Cui et al.[73]	https://drive.google.com/drive/folders/1oWh_-zRdhtoGQ2Fw24HP41FgLoomVo-y.
Torch v1.13.0	Linux Foundation	https://pytorch.org/
NIS Nikon Elements AR v5	Nikon	N/A
Fiji (ImageJ2) v2.9.0	Schindelin et al.[111]	https://imagej.net/

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Human samples

Human samples were discarded, deidentified specimens from donors undergoing pediatric cardiothoracic corrective surgery at Boston Children's Hospital (BCH). All procedures were approved by the BCH institutional review board (protocol number IRB-P00041973), and written consent was obtained for use of deidentified discarded tissue as part of the consent for surgery. Samples were from male and female donors, and donors ranged in age from two days to twelve years.

Zebrafish

Zebrafish were maintained in accordance with Boston Children's Hospital Institutional Animal Care and Use Committee protocols and in line with Animal Resources at Children's Hospital guidelines (protocol 20-10-4254R). Strains used were: *Tg(lck:mCherry; cd79a:eGFP)*, *Tg(lck:GFP; mylpfa:mCherry)*, and Casper. Zebrafish ranged from two to fourteen weeks post-fertilization. Both male and female zebrafish were included; the sex of the zebrafish is indicated unless the fish was too young/small for sex determination.

METHOD DETAILS

Human thymus processing

Approximately 500 mg of non-frozen thymic tissue was placed in 5 mL of DMEM on ice. Tissue was vigorously minced with scissors to release thymocytes. The sample was rested on ice for 1 minute for thymic pieces to settle, and the lymphocyte-rich supernatant was removed. Thymic fragments were digested with 10 mL collagenase digest buffer (basic media (DMEM, 25 mM HEPES, 2% FCS (Gemini Bio)) plus 0.5mg/mL collagenase II (Sigma), 0.1mg/mL dNaseI (Millipore)) at 37°C for 20 minutes, with agitation, centrifuged then digested with collagenase/dispase buffer (basic media plus 0.5mg/mL collagenase/dispase (Millipore), 0.1mg/mL dNaseI) at 37°C for 20 minutes. EDTA was added to 10 mM, and the sample was centrifuged and resuspended in FACS buffer (PBS, 2% FCS) and filtered through a 100 μm filter. $EpCAM^+$ cells were enriched (Donors A, B, and E) using magnetic anti-EpCAM beads and LS columns or $CD45^+$ cells were depleted (Donors C and D) using magnetic anti-CD45 beads and LS columns (all Miltenyi). Cells were rinsed on the column with MACS buffer (PBS, 2% FCS, 1mM EDTA). EpCAM-enriched cells were eluted in 5 mL of MACS buffer.

Human thymus flow cytometry

Cells were stained with CD45 Brilliant Violet 605 (Biolegend 304041; Clone HI30; RRID AB_2562105), EpCAM APC-Cy7 (Biolegend 324245; Clone 9C4; RRID AB_2783193), HLA-DR APC (Biolegend 307609; Clone L243; RRID AB_314687), CD205 PE (Biolegend 342203; Clone HD30; RRID AB_1626209), PDPN PE-Cy7 (Biolegend 337013; Clone NC-08; RRID AB_2563367), CD104 FITC (Biolegend 327805; Clone 58XB4; RRID AB_893216) antibodies, each at 1:50 dilutions in FACS buffer, and dead cells were labeled with DAPI. Samples from Donors G-I were stained with the LIVE/DEAD Fixable Violet Cell Stain kit (Invitrogen L34963), rather than DAPI, at 1:1000 for 10 min, washed with FACS buffer, and were additionally fixed in 1% formaldehyde (Thermo Scientific 28906) in PBS for 10 min at room temperature, quenched with equal volume of room temperature 0.25 M glycine in PBS, and washed twice with 0.04% bovine serum albumin in PBS. Cells were analyzed and sorted on a 4-laser FACSAria (BD) instrument, gated as in Figure S1A. Flow cytometry data were analyzed in FlowJo 10.8.2.

Summary visualizations of flow cytometry-defined compartment composition were generated using Python 3.7.3 and matplotlib (version 3.5.3). Error bar mean and standard deviation calculations were performed using the numpy package (version 1.21.6). Statistical testing was performed using the scipy package (version 1.4.1) and the two-sided Wilcoxon rank sum test with no correction for multiple hypothesis testing and a cutoff for significance of $p < 0.05$.

Human population-level RNA-seq

Cell populations were sorted twice by flow cytometry for purity, sorting 1,000 cells directly into 5 μL TCL buffer with 1% 2-mercaptoethanol in the final sort for cell lysis. Cell libraries were prepared using Smart-seq2 and sequenced by the Broad Genomics Platform, according to the ImmGen ultra-low-input sequencing protocol (immgen.org).

Data were preprocessed according to the ImmGen preprocessing protocols (immgen.org). Reads were aligned to the GRCh38 human genome with STAR and gene counts quantified with featureCounts (Subread package). The standard ImmGen quality control filters were applied to ensure sample exclusion if a sample had <8,000 genes with more than ten reads, high hematopoietic-cell transcript contamination, housekeeping gene median transcript integrity number below 45, or poor replicate correlation.

Raw read count tables were normalized with the DESeq2 median of ratios methods, the standard ImmGen approach, and utilized for plotting heatmaps and PCA analysis. Downstream analyses were performed in R (version 4.2.2). Genes were excluded if their normalized expression did not exceed 20 in at least two samples.

Differential expression analysis was performed to select genes for heatmap visualization and PCA analysis. Differentially expressed genes were identified for each compartment vs the remaining compartments using DESeq2 applied to count tables. In these calculations, low expression genes were excluded if they did not exceed 20 counts in at least two samples. Differentially expressed

genes were selected for heatmap inclusion using a cutoff of fold change (FC) $\geq$ 2 and false discovery rate (FDR) < 0.05 with Benjamini-Hochberg correction. Differentially expressed genes were selected for PCA analysis using the above procedure with |FC| $\geq$ 2 and FDR < 0.05.

PCA was performed with the pcaMethods package, using method 'svd', scaled using unit variance ('uv') and a mean-centered matrix ('center=TRUE'). To plot heatmaps, normalized data tables were $\log_2$ transformed and genes were row-centered by subtraction of row means.

Human scRNA-seq library preparation

10,000 - 20,000 human PDPN$^-$CD104$^-$ mTEC^{lo} cells were sorted into an FCS-coated eppendorf collection tube containing basic media and 10% FCS. Cells were encapsulated and processed using the Chromium Next Gem Single Cell 3' Dual Index kit (v3.1 chemistry) from 10x Genomics as per manufacturer's instructions. Samples were sequenced on a NovaSeq6000 (Broad Institute) for a minimum of 80 million reads per sample (Read 1: 28 cycles, i5 index: 10 cycles, i7 index: 10 cycles, Read 2 index: 90 cycles).

Human scRNA-seq analysis

Sequencing reads were demultiplexed and aligned to the GRCh38 human reference genome and output as transcript-by-cell matrices using Cell Ranger (version 7.0). Downstream single cell RNA-sequencing analysis was performed in R using Seurat. Filtered feature matrices were imported as Seurat objects, with a minimum feature filter of 3 cells ('min.cells') and raw count matrices were merged across donors without formal batch correction ('merge'). Cells with $\geq$12.5% of reads mapped to the mitochondrial genome were excluded. Similarly, cells with $\leq$1000 genes or $\geq$7000 genes were excluded. Data were scaled and normalized ('NormalizeData'), in which gene counts for each cell were divided by total counts for the corresponding cell and multiplied by scale-factor 10,000 and natural-log transformed (log1p). These data were used in gene feature plots ('FeaturePlot'). Variable features (2,000) were selected using the 'vst' selection method and variable features were scaled and centered ('ScaleData') and utilized for principal component analysis (PCA) dimensionality reduction ('RunPCA'). Cells were clustered, followed by non-linear dimensional reduction and visualization in UMAP space using the first 30 dimensions (using Seurat functions 'FindNeighbors', 'FindClusters', and 'RunUMAP'). Doublet prediction was run using the DoubletFinder package,[106] using an expected doublet rate of 4% (based on 10x Genomics expected multiplet rates) ('doubletFinder_v3' function with pN = 0.25, pK = 0.09, PCs = 1:10) and run separately on each donor. Contaminating fibroblasts and T cells were eliminated alongside a cluster enriched for predicted doublets. After eliminating clusters, the normalization, scaling, and dimensionality reduction procedures were repeated. Cells were clustered in a semi-supervised fashion by unbiased Louvain clustering at high resolution (resolution = 2) to capture fine-grained heterogeneity among mimetic cells, followed by consolidation of biologically homologous clusters. To clarify subtype and transitional mimetic cell boundaries (i.e. the ionocyte/transitional cell cluster boundary), we subset cells, subclustered as described above (with Louvain clustering at resolution = 1), and integrated subclusters into the larger dataset in a supervised manner.

Dot plot visualizations were generated with the Seurat function 'DotPlot' with defaults including scaling, and minimum threshold of expression in 5% of cells for the smallest dot. Scaled data were visualized in heatmaps using the Seurat function 'DoHeatmap', with a minimum display value of 0.5. Muscle-specific analysis was performed by subsetting the three muscle clusters and reclustering, as described above; small subclusters enriched for ionocyte and intermediate genes were excluded, and cells were reclustered. For the muscle-specific heatmap, the Seurat function 'FindAllMarkers' was used to identify cluster-specific genes, expressed in at least 25% of cells (in either of two populations being compared) and at least an average of 1.2-fold change difference between the two groups. Neuroendocrine-specific visualizations were generated by subsetting the six neuroendocrine clusters and reclustering; small subclusters enriched for intermediate genes were excluded and cells were revisualized, with the original cluster identities labeled.

Module scores for muscle types from published data were calculated using the Seurat 'AddModuleScore' function, which calculates average expression levels of signature genes against aggregate expression-matched control genes. Marker genes were selected using the Seurat function 'FindAllMarkers' to identify upregulated genes, expressed in at least 10% of cells (in either of two populations being compared) and at least an average two-fold change difference between the two groups. Mouse gene names were converted to their human equivalents in Python using the pybiomart package (version 0.2.0) by converting via gene stable IDs. If multiple human genes existed for one mouse gene, all were included. If no human equivalent existed, that gene was skipped. The signature for myotubes was generated by combining signature genes from clusters of type iIx and type iIb myonuclei.

The three muscle clusters from the scRNA-seq data were grouped and the non-muscle mimetic cell clusters were grouped for pseudo-bulk analysis. Gene counts were summed across cells for each donor in muscle or non-muscle groups, to make 10 groups (5 donors, muscle or non-muscle groups). DESeq2 was utilized for differential expression analysis. TF-encoding genes were identified based on compiled TFs from Lambert et al.[24] An analogous procedure was followed for pseudo-bulk analysis of lncRNA-enriched muscle mTECs compared to the remaining mimetic cells. LncRNA-encoding genes were identified based on compiled genes from LNCipedia v5.2[112] and GENCODE v46.[113]

For RNA velocity analysis, FASTQ files were preprocessed using kallisto-bustools, with prebuilt human Linnarsson index and lamanno workflow option, to map spliced and unspliced transcripts.[107,114] RNA velocity was calculated and plotted using scVelo (version 0.2.5) in dynamical mode.[108] Genes were filtered for the top 20,000 genes and required minimum shared counts of 20 for a gene. Cluster assignments and UMAP coordinates were directly carried over from Seurat and velocities were plotted as streamlines on the embedding, alongside computed velocity confidence. CellRank (version 1.5.1)[109] and scVelo were utilized for calculating terminal and initial states and plotting a partition-based graph abstraction (PAGA) representation.[115]

Human thymus histology

Thymus tissue was sectioned sagittally through the central part of a thymic lobe (to approximately 5 mm). Tissue was rinsed in PBS then incubated in 4% paraformaldehyde in PBS overnight at 4°C. Tissue was then rinsed three times in excess PBS and transferred to a sucrose gradient in PBS of layered 5%, 15%, and 30% solutions and allowed to sink overnight at 4°C. Thymus tissue was patted dry and embedded in optimal cutting temperature media (OCT) (Sakura Tissue-Tek), frozen on dry ice, and stored at -80°C. OCT embedded tissue was cryostat sectioned to 8 μm. At the start of staining, sections were permeabilized in PBS with 0.05% Tween-20 (Fisher BP337-500) (PBS-T) for 10 minutes. Tissue was blocked for 1 hr at room temperature with 5% normal donkey serum (Jackson ImmunoResearch 017-000-121), stained with primary stain in PBS-T at 1:50-1:100 dilution for 1 hr at room temperature, and rinsed three times quickly with PBS-T then three times for 10 minutes each, with agitation. Tissue was stained with secondary antibody at 1:1000 dilution in PBS-T for 1 hr at room temperature, and again washed three times quickly in PBS-T then three times for 20 minutes each, with agitation. Tissue was stained with Hoechst nuclear counterstain at 1:500 for 10 minutes at room temperature, rinsed in PBS-T and mounted with ProLong Diamond Antifade Mountant (ThermoFisher P36970) and coverslip (No. 1 ½ optically clear glass).

Tissue sections were stained with the following antibodies: anti-Desmin (ab15200), anti-DEC205-PE (Biolegend 342203), anti-Myogenin-PE (BD Pharmingen 563120), anti-Myosin heavy chain (R&D MAB4470), anti-Dysferlin (Novus NBP1-84696-25ul), anti-ChAT (R&D AF3447-SP), anti-CADPS (Novus NBP1-77323), anti-SOX2 (Cell Signaling Technology 23064S), anti-S100 (Abcam ab868), anti-Cytokeratin 10 (Abcam 212546), anti-Dystrophin (Abcam 15277), anti-TRPM5 (Proteintech 18027-1-AP), donkey anti-rabbit FITC (Jackson ImmunoResearch 711-096-152), donkey anti-rabbit Cy3 (Jackson ImmunoResearch 711-166-152), donkey anti-mouse AF488 (Invitrogen A21202), donkey anti-rabbit AF647 (Jackson ImmunoResearch 711-606-152), donkey anti-goat AF647 (Jackson ImmunoResearch 705-605-147), donkey anti-mouse Cy5 (Jackson ImmunoResearch 715-175-151), and staining reagent Alpha-bungarotoxin-CF488 (VWR 76221-710).

Thymic sections were imaged by spinning-disk confocal microscopy using Nikon Elements acquisition software, a Nikon Ti inverted microscope with W1 Yokogawa Spinning disk with 50 μm pinhole disk, and Plan Apo 20x air or Plan Apo 100x oil objective. Images with multiple Z-planes were processed as the maximum intensity projection. Images were processed in Fiji (ImageJ2).

Zebrafish tissue collection and cell sorting

Ten adult zebrafish (13 weeks post-fertilization (wpf)) were analyzed by scRNA-seq in this study. These ten zebrafish were *lck:mCherry*; *cd79a:eGFP* double transgenics, generated by crossing *Tg(lck:mCherry)* (gift from Aya Ludin Tal, Zon Laboratory, Harvard University) and *Tg(cd79a:eGFP)* zebrafish.[100] This pair of transgenic markers allowed for the detection and visualization of T and B cells. The sex of these ten fish could not be confidently determined at the time of analysis due to their small size and young age. An additional adult Tübingen (TU) wildtype zebrafish was used as a control for fluorescence-activated cell sorting (FACS) gating.

lck:mCherry; *cd79a:eGFP* transgenic zebrafish were euthanized by rapid chilling and confirmed deceased prior to thymi dissection. Thymus dissections were performed as previously described.[72] Briefly, each zebrafish was positioned on a Styrofoam surface and appropriately braced under a fluorescent dissecting scope. Each thymus was clearly visualized following the removal of the operculum and dissected into 0.5% BSA in HBSS without Ca^{2+} and Mg^{2+}. Five samples were liberase treated by adding a 2.5 mg/mL Liberase TM (medium Thermolysin concentration, Roche) stock solution at 1:60 dilution to each sample for a final total collagenase concentration of 42 μg/mL. These samples were incubated at 37°C on a heat block at 800 rpm for 20 minutes. To stop the enzymatic digestion, pure fetal calf serum (FCS) was added to a final concentration of 10%, and the samples were filtered through a 50 μm disposable filter (CellTrics), centrifuged at 400 g x 5 min, and resuspended in 0.5% BSA in HBSS without Ca^{2+} and Mg^{2+}. Five untreated samples were mechanically dissociated by pipetting prior to filtering, centrifugation, and resuspension.

Sytox blue was used to determine cell viability and was added to the samples immediately before FACS. Live, non-autofluorescent, GFP-negative, mCherry-negative cells were sorted (85 μm nozzle) on a BD FACSAriaII from each of the 5 samples per condition into 300 μL of 0.5% BSA in DPBS and kept on ice following the sort. This sorting strategy depleted T and B cells from the sample, allowing for the relative enrichment of thymic epithelial cells. In total, approximately 18,000 cells were sorted for the untreated condition and 33,000 cells were sorted for the liberase-treated condition.

Zebrafish scRNA-seq library preparation

Sorted cells were pelleted by centrifugation at 400 g for 5 min. Following Trypan blue staining, they were resuspended to final concentrations of 400 and 700 cells/μL for the untreated and liberase-treated samples, respectively. Transcriptome profiling was performed using the Chromium Next GEM Single Cell v3.1 Reagent Kits (10x Genomics) with a targeted recovery of 4,000 cells per lane. Two lanes were run for the liberase-treated condition, whereas one lane was run for the untreated condition. The final transcriptome libraries were pooled with 4 additional libraries, each targeting 6,000 cells. These pooled libraries, targeting a total of 36,000 cells, were sequenced at the Harvard University Bauer Core on an Illumina NovaSeq 6000 using an S4 flow cell (Read 1: 28 cycles, i7 index: 8 cycles, i5 index: 0 cycles, Read 2: 91 cycles).

Zebrafish scRNA-seq analysis

Demultiplexed FASTQ files were processed using 10x Genomics Cell Ranger 7.0.0.[116] Alignment was performed to a custom genome consisting of zebrafish Ensembl GRCz11 and five additional genes (*EGFP*, *mCherry*, *dsRed*, *TagBFP*, and *BRAFv600e*). Previously published zebrafish thymus data[72] were reprocessed and included in our analysis. Specifically, the demultiplexed FASTQ files

from four technical replicates of cells derived from the thymi of 21 juvenile *Tg(lck:eGFP)* zebrafish (4 wpf), two adult *Tg(lck:eGFP)* zebrafish thymi, and two adult GESTALT zebrafish thymi were included. Reprocessing was performed to include reads mapping to intronic regions to increase sensitivity and for consistency with mouse and human data processing (Technical Note - Interpreting Single Cell Gene Expression Data With and Without Intronic Reads, Document Number CG000554 (Rev A), 10x Genomics, (2022, June 21)).

Zebrafish scRNA-seq data were analyzed as described for human data above, with several modifications, described here. Cells with $\geq$10% of reads mapped to the mitochondrial genome or cells with $\leq$300 genes were excluded. Doublet prediction was run on separately encapsulated samples, with an expected doublet rate of 1%. Immunocytes and non-TECs, including cells expressing canonical T cell, B cell, NK cell, or myeloid markers, were iteratively excluded.

Ionocyte-specific analysis was performed by subsetting HR and NaR ionocyte mimetic cell clusters and reclustering (resolution = 0.5). Labels were reassigned and homologous clusters combined. Similarly, neurosensory, immune, and gut-type tuft cells were subset and reclustered (resolution = 1). For muscle-specific analysis, the muscle mimetic cell cluster was subset and reclustered (resolution = 1, ndims = 50).

Imaging of transgenic zebrafish

Larvae (2 wpf) from *Tg(lck:GFP; mylpfa:mCherry)* transgenic zebrafish were used in this study. Confocal microscopy was performed using a Yokogawa CSU-X1 spinning disk mounted on an inverted Nikon Eclipse Ti microscope equipped with dual Andor iXon EMCCD cameras. Animals were only included for imaging and analysis if the expression of all transgenes could be identified. All images were acquired using NIS-Elements (Nikon). Whole fish were mounted in 0.8% LMP agarose with tricaine (0.16 mg/mL) in glass bottom 6-well plates and covered with E3 media containing tricaine (0.16 mg/mL). GFP and mCherry expression were detected using 488-nm and 561-nm lasers, respectively.

Zebrafish immunohistochemistry

Wild-type Casper fish were fixed in 10% formalin (VWR cat. 16004-121) and tissues were dehydrated through a series of graded ethanol: 70% (x1), 95% (x3), 100% (x3), followed by xylene (x3), and liquid paraffin. After processing, tissues were embedded in paraffin. Tissues were sectioned at 5 μm and mounted onto Superfrost Plus Slides (Fisherbrand 12-550-15). Tissues were stained with Harris' hematoxylin solution for 6 minutes at 60-70°C and then rinsed in tap water. Next, 0.25% acetic acid and 70% ethanol in water were used to differentiate the tissue for 30 seconds, and tissues were then rinsed with water. Tissues were blued in lithium carbonate solution for 30 seconds and rinsed with water. Staining was performed with eosin Y ethanol solution for 30 seconds followed by dehydration and covered with a coverslip. Immunohistochemistry was performed on the Leica Bond III automated staining platform using the Leica Biosystems Refine Detection Kit (Leica; DS9800). FFPE tissue sections were baked for 30 minutes at 60°C and deparaffinized (Leica AR9222) prior to staining. Primary antibodies were incubated for 30 minutes, visualized via DAB, and counterstained with hematoxylin (Leica DS9800). The slides were rehydrated in graded alcohol and coverslipped with HistoreCore Spectra CV mounting medium (Leica 3801733) added. Myosin 4 antibody (Thermo Fisher 14-6503-82; clone MF20) was run at a 1:100 dilution with a 30-minute citrate antigen retrieval (Leica ER1 AR9961).

Evolutionary tree visualization

Evolutionary tree schematic was adapted from visualization generated by TimeTree5 (https://timetree.org/).[74]

Fine-tuning scGPT on human mTEC datasets

The pre-trained scGPT model was downloaded from:
https://drive.google.com/drive/folders/1oWh_-zRdhtoGQ2Fw24HP41FgLoomVo-y. Our human mimetic cell scRNA-seq datasets were used to finetune the pre-trained scGPT model. Specifically, we excluded non-mimetic cell subtypes, and 19,558 cells from Donors A, B, C, and D were used as the reference dataset for fine-tuning, and 6,226 cells from Donor E were used as the query dataset to evaluate the performance of the fine-tuned scGPT. The fine-tuning task was to predict the mimetic cell subtype (identified using Seurat) for each cell based on its transcriptome profiled by scRNA-seq. The fine-tuning model was implemented in Torch (v1.13.0), and was trained on a single NVIDIA A100 GPU card for around ~50 hours. We used a batch size of 8, and fine-tuned the model for 15 epochs until convergence and then returned to the model that had achieved the smallest validation loss. The learning rate of the model was 1e-5, which was selected by comparing the performances of multiple learning rates. To evaluate the performance of the fine-tuned scGPT model, we used the model to predict mimetic cell subtypes for cells in the query dataset (i.e. test set, Donor E). The fine-tuned scGPT model achieved an accuracy of 0.886 on the query dataset. We also generated a confusion matrix to visualize its performance on individual mimetic cell subtypes.

Applying fine-tuned scGPT model

To use the scGPT model, which was pre-trained and fine-tuned on human cells, for mouse and zebrafish datasets, we converted the mouse and zebrafish gene names to human gene names. We adapted a pipeline from Song et al.[117] for one-to-one gene conversion. In brief, we identified orthologous genes using pybiomart (version 0.2.0) and, where multiple orthologs existed, selected the top ortholog based on orthology confidence (primarily) and gene order conservation (secondarily), and removed the genes that were not conserved between mouse/zebrafish and human. The converted scRNA-seq datasets of mouse and zebrafish mimetic cells then underwent routine processing using Seurat, including quality control (min.cells = 3, min.features = 200), normalization, scaling,

PCA and clustering. After converting gene names and cleaning cells, 4,341 mouse mimetic cells x 15,130 genes, and 1,104 zebrafish mimetic cells x 8,699 genes were input to the fine-tuned scGPT model for prediction. For each mouse/zebrafish mimetic cell, the fine-tuned scGPT model predicted to which human mimetic cell subtype it is most similar based on its transcriptome profiled by scRNA-seq. For a given cell, if the predictive probability was smaller than 0.5 for every human mimetic cell subtype label, the mouse/zebrafish cell would be labeled as "novel" to indicate that it might be a mimetic cell subtype that was specific to mouse/zebrafish and not identified in human. To visualize predicted labels on mouse and zebrafish UMAPs, the scGPT predictions were imported into R and assigned as a feature in the Seurat object metadata. Mouse/zebrafish mimetic cells (excluding Aire-stage, immature mTECs, transit amplifying mTECs, and cTECs) were revisualized and predicted labels overlaid. Original labels are also shown for reference.

QUANTIFICATION AND STATISTICAL ANALYSIS

Significance was assessed for flow cytometry data by two-sided Wilcoxon rank sum test with no correction for multiple hypothesis testing and a cutoff for significance of $p < 0.05$, as described in the corresponding figure legend. Definitions of center and dispersion are also indicated in the figure legend. *p*-values for population RNA-seq or pseudo-bulk differential expression analysis were calculated in DESeq2 using a Wald test with Benjamini-Hochberg correction for multiple comparisons. Statistical analyses were performed using R or Python.