Nature Communications

https://doi.org/10.1038/s41467-026-72760-7

Article in Press

A single-cell atlas revealing cellular heterogeneity across healthy and diseased human thymus

Received: 5 June 2023

Accepted: 17 April 2026

Cite this article as: Direder, M., Wielscher, M., Salek, M. *et al.* A single-cell atlas revealing cellular heterogeneity across healthy and diseased human thymus. *Nat Commun* (2026). https://doi.org/10.1038/s41467-026-72760-7

Martin Direder, Matthias Wielscher, Melanie Salek, Maria Laggner, Dragan Copic, Katharina Klas, Daniel Bormann, Bahar Golabi, Hannes Kühtreiber, Marie-Therese Lingitz, Leonhard Müllauer, Ana-Iris Schiefer, Wolfgang Weninger, Clemens Aigner, Hendrik Jan Ankersmit, Michael Mildner & Bernhard Moser

A single-cell atlas revealing cellular heterogeneity across healthy and diseased human thymus

Martin Direder [1,2,3], Matthias Wielscher [4], Melanie Salek [4], Maria Laggner [1,2], Dragan Copic [1,2,5], Katharina Klas [1,2], Daniel Bormann [1,2,6], Bahar Golabi [4], Hannes Kühtreiber [4], Marie-Therese Lingitz [1,2,7], Leonhard Müllauer [8], Ana-Iris Schiefer [8], Wolfgang Weninger [4], Clemens Aigner [9], Hendrik Jan Ankersmit [1,2,9,#,*], Michael Mildner [4,#,*], Bernhard Moser [9,#]

[1]Laboratory for Cardiac and Thoracic Diagnosis, Regeneration and Applied Immunology, Department of Thoracic Surgery, Medical University of Vienna, Vienna, Austria;
[2]Aposcience AG (FN 308089y), Dresdner Straße 87/A21, Vienna, Austria;
[3]Department of Orthopedics and Trauma-Surgery, Medical University of Vienna, Vienna, Austria;
[4]Department of Dermatology, Medical University of Vienna, Vienna, Austria;
[5]Department of Nephrology, Medical University of Vienna, Vienna, Austria;
[6]Department of Neurology, Medical University of Vienna, Vienna, Austria;
[7]Division of General Anesthesia and Intensive Care Medicine, Department of Anesthesia, Critical Care and Pain Medicine, Medical University of Vienna, Vienna, Austria
[8]Department of Pathology, Medical University Vienna, Vienna, Austria;
[9]Department of Thoracic Surgery, Medical University of Vienna, Vienna, Austria;

running title: The transcriptional landscape of pathologic thymus entities

 # equal contributing and joint supervising authors:
These authors contributed equally and jointly supervised this work: Hendrik Jan Ankersmit, Michael Mildner, Bernhard Moser

***corresponding authors:**
Hendrik Jan Ankersmit, MD
Laboratory for Cardiac and Thoracic Diagnosis, Regeneration and Applied Immunology, Department of Thoracic Surgery, Medical University of Vienna
Waehringer Guertel 18-20, 1090 Vienna, Austria
ORCID: 0000-0002-8761-3517

e-mail: hendrik.ankersmit@meduniwien.ac.at
phone: +43-1-40400-69790 or 67770
fax: +40-1-40400-67820

Michael Mildner, PhD
Department of Dermatology, Medical University of Vienna
Lazarettgasse 14, 1090 Vienna, Austria
ORCID: 0000-0002-6892-925X
e-mail: michael.mildner@meduniwien.ac.at
phone: +43-1-40400-73507
fax: +43-1-40400-73590

Abstract

The human thymus plays a key role in the development of the adaptive immune system. Its development and pathologic aberrations with missing involution occupy the scientific world for years. Here, we present a comprehensive single-cell RNA sequencing (scRNA-seq) analysis of 453,727 cells across 53 datasets derived from healthy prenatal, pediatric, and adult thymic tissues, as well as six pathological conditions, including thymic hyperplasia and thymic epithelial tumors (types A, AB, B, C, and micronodular thymoma). We created a high-resolution cellular atlas revealing disease-specific cellular populations and transcriptional programs, particularly within fibroblast subsets and thymic epithelial cells. Comparative analysis uncovers distinct intercellular communication patterns and identifies transcriptional alterations associated with thymic pathology. Integration with published bulk RNA-seq datasets supports the robustness and translational relevance of our findings. This study provides a foundational resource for understanding the cellular and molecular landscape of thymic development and disease, offering avenues for diagnostic and therapeutic innovations.

Keywords: prenatal thymus, pediatric thymus, adult thymus, thymoma, thymic carcinoma, thymic hyperplasia, single cell RNA sequencing

Introduction

The thymus is a primary lymphoid organ essential for the development and education of the adaptive immune system during early life [1]. Its function is orchestrated through the intricate interplay of diverse cell populations, particularly thymic epithelial cells (TECs), which guide the selection and maturation of self-tolerant, functional thymocytes (TC) [2]. Thymic activity peaks during childhood and gradually declines with age, undergoing physiological involution whereby up to 50% of its mass is replaced by adipose tissue by the age of 40 [3]. Despite its regression in adulthood, the thymus can persist abnormally, known as thymic hyperplasia (TH), a benign condition characterized by organ enlargement without severe aberration. Alternatively, the thymus can undergo neoplastic transformation into thymic epithelial tumors (TETs), including thymomas and thymic carcinomas. Thymic neuroendocrine tumors (TNETs) define a biologically distinct disease entity. TH is defined as an increase in thymic size and weight while maintaining a benign microscopic architecture [4]. The adipose tissue in TH can be either sparse between cell-rich lobules or abundant between small lobules [5,6]. In contrast, thymomas, although rare, are clinically relevant due to their association with autoimmune disorders and exhibit the lowest mutational burden among adult human cancers, with favorable outcomes following complete surgical resection [7,8].

Thymomas are histologically classified into type A (TET-A), AB (TET-AB), B (TET-B), metaplastic type C/carcinoma (TET-C), and micronodular thymoma with lymphoid stroma (TET-MNT) [9]. Recent molecular studies suggest that these subtypes may represent biologically distinct disease entities rather than a linear histopathological continuum [8]. TET-A contain bland, spindle-shaped or oval tumor cells, few or no immature T cells and usually exhibit low tumor stages [10]. TET-AB are described as epithelial thymus tumors with widely varying lymphocyte-poor (TET-A) and lymphocyte-rich (TET-B like) components [11]. B-type thymomas (TET-B) are further divided based on their lymphocyte content and epithelial morphology in type B1 (TET-

B1), B2 (TET-B2) or B3 (TET-B3). TET-B1 resembles the neonatal thymus with abundant immature T cells and sparse epithelial components, whereas TET-B3 is predominantly composed of epithelial cells with minimal lymphocytic infiltration and lacks cortical or medullary differentiation [12]. TET-MNT presents as biphasic lesions with spindle-cell nodules embedded in a lymphoid-rich, epithelium-free stroma, while TET-C (58% thymic squamous carcinoma) exhibits lobular architecture with islands of polygonal tumor cells, focal keratinization, and intercellular bridges [11,13].

The postnatal thymus itself is encapsulated and divided into lobules by connective tissue septa. Its parenchyma primarily comprises developing TC supported by a stromal network of TECs, fibroblasts (FB), endothelial cells (EC), and other accessory cell types that coordinate TC maturation and central tolerance induction [14].

The advent of single-cell RNA sequencing (scRNA-seq) has revolutionized the analysis of complex tissues by enabling the high-resolution profiling of cellular heterogeneity, gene expression dynamics, and intercellular communication [15]. With regard to the thymus and TETs, scRNA-seq has already advanced our understanding of thymic organogenesis, TC differentiation, and the molecular underpinnings of thymoma subtypes, particularly those associated with myasthenia gravis (MG) [16-23]. Furthermore, integrated multi-omics approaches, such as that by Radovich et al., have characterized the genomic landscape of TETs, highlighting *GTF2I* mutations and their subtype-specific distribution [8].

In the present study, we aimed to expand this basis by applying scRNA-seq to systematically characterize the cellular composition and transcriptional programs of benign thymic tissue, TH, and TETs, including thymic carcinoma. Our goal was to elucidate common and unique cellular features across thymic conditions and to identify potentially pathogenic cell populations and signalling networks. The datasets generated here are publicly available and offer a valuable

resource for future integrative analyses, contributing to a deeper understanding of thymoma biology and facilitating the discovery of diagnostic and therapeutic targets.

Results

Uncovering the cellular heterogeneity in normal and diseased human thymus by scRNA-seq.

Recent scRNA-seq studies analysing the cellular microenvironment of the thymus and its pathological alterations have provided critical insights over the past years [16,20,22,23]. Given that these studies focus on the same organ and that advancements in scRNA-seq technology have significantly increased the number of analyzable cells, we consolidated published datasets from independent research groups. To expand the published datasets, especially for TH and TET-B, we performed additional scRNA-seq experiments. This integrated dataset encompasses transcriptional profiles of prenatal, pediatric, and adult thymus, as well as various TET subtypes, including type A, AB, B, C, and MNT, to construct a comprehensive representation of the thymic cellular landscape (Figure 1a).

In total, fifty-three datasets from five independent research groups were analyzed in a unified and extensive manner. This combination included eighteen prenatal thymus samples, eight pediatric thymus samples, five adult thymus samples, six TH samples, and samples representing thymoma subtypes: one TET-A, four TET-AB, seven TET-B (including B1-B3 but also intermediate types), three TET-C (thymic carcinoma, all squamous cell carcinomas), and one TET_MNT (donor information detailed in Supplementary Table 1). Due to challenges in histologically distinguishing the TET-B subtypes in all samples, these were collectively analyzed as a merged group. Histopathological validation of all tissue samples was conducted by expert pathologists (Supplementary Figure 1), and representative hematoxylin and eosin (H&E) stainings of all thymic conditions are provided (Figure 1b).

Following rigorous preprocessing and quality control, a total of 453,727 high-quality single-cell transcriptomes were selected for analysis. Unbiased clustering identified nine major cell clusters (Figure 1c). Previously published cluster markers (Bautista et al., Xin et al., Yasumizu et al.) were used to annotate cell types, facilitating cross-validation with published cell populations

from other studies (Supplementary Figure 2a) [16,22,23]. Annotation accuracy was further confirmed by cross-referencing marker genes within our generated clusters (Supplementary Figure 2b, Supplementary Data 1). We successfully identified T cells (TC), B cells (BC), thymic epithelial cells (TEC), dendritic cells (DC), macrophages/monocytes (MAC/Mono), endothelial cells (EC), fibroblasts (FB), and vascular smooth muscle cells/pericytes (VSMC/Peri) in varying proportions depending on the tissue condition (Figure 1c).

The relative cellular composition per tissue type and research group is summarized in Figure 1d. Differences in tissue processing protocols, including differential cellular enrichment, across the contributing studies results in divergent proportional representations of major cell clusters (Supplementary Table 2).

To address discrepancies in cellular proportions arising from differential tissue processing in TETs, bulk RNA-seq deconvolution was performed across thymoma subtypes [24]. This analysis demonstrated a prominent presence of TECs in TET-A, co-dominance of T cells and TECs in TET-AB, and a progressive decrease in T cells alongside increased TEC prevalence from TET-B1 to B3. Other cell types remained minor constituents in these groups, whereas TET-C and TET-MNT exhibited markedly higher proportions of additional cell types (Figure 1e, Supplementary Data 2).

Compared to bulk RNA sequencing, scRNA-seq enables detailed comparison of specific cell types across distinct thymic conditions. To gain an initial overview of cellular heterogeneity in pathologic thymic conditions, we performed a cell type-specific differential expression analysis to identify the most significantly up- and downregulated genes in each pathological thymic entity relative to the healthy adult thymus. Analysis of differentially expressed genes (DEGs) across all cell types revealed numerous significant transcriptional alterations. Notably, TECs exhibited the most pronounced changes in almost all pathological conditions. Additionally, Mac/Mono and FB showed substantial dysregulation compared to healthy adult thymus (Supplementary Figure 3).

Among the pathologies, TET-C and TET-MNT displayed the highest degree of cell type-specific alterations (Supplementary Data 3). GO Term analysis uncovered functional shifts within TECs of each tissue in comparison with the healthy adult thymus. They especially show an upregulation of genes associated with cellular transition and development in the pathologic tissues. (Supplementary Data 4)

Together, this integrative scRNA-seq analysis represents the most comprehensive cellular characterization of the human thymus and its pathological states to date, providing a refined and unified cellular atlas that highlights key cellular composition shifts across neoplastic transformation. Direct comparison of gene expression in the identified cell types of the various thymus conditions provides direct insight into the individual cell stages present and their dominant transcriptional situations.

Integrated analysis reveals distinct thymic epithelial subtypes associated with TET.

Previous studies have provided valuable but heterogeneous characterizations of the cellular landscapes in both healthy and pathological thymic conditions [16,20,22,23]. However, variations in cluster nomenclature and analytic depth limited unified interpretation. To address this, we performed a comprehensive re-analysis of all major cell clusters across all datasets to establish a consistent transcriptional framework.

Given the proposed pivotal role of TECs in TETs [2], we first focused on in-depth subclustering of the TEC population, identifying ten distinct TEC subclusters (Figure 2a). Utilizing module scores derived from top DEGs reported in prior TEC subtype characterizations as well as marker genes established by Yayon et al., we assigned specific TEC subtypes within the current dataset (Figure 2b) [16,22,23,25]. This juxtaposition of previous annotations also points out certain analogies between the referenced studies. One cluster was annotated as cortical TEC (cTEC) and one as

corticomedullary TEC-like (mcTEC), consistent with previous genetic annotations [16,22,25]. Medullary TECs segregated into two distinct clusters corresponding to mTEC_hi and mTEC_lo. While mTEC_hi included keratinocyte-like TECs, mTEC_lo interestingly also expressed marker genes characteristic for immature TECs (Figure 2b) [23,25]. Strikingly, one TEC cluster was predominantly restricted to pathological thymic conditions (TH, TET-A, TET-AB, TET-B, and TET-MNT) and was barely noticeable in healthy tissue (denoted as TET_TEC, Figure 2a). This cluster shared transcriptional features with KRT14+ mTEC-like and mcTEC cells described by Xin et al., both of which are linked to thymic epithelial progenitor characteristics, and also showed a slight upregulation of typical cTEC marker as *LY75* and *CCL25* [23]. However, TET_TECs showed no strong expression of marker genes for regular mcTECs identified in healthy thymus, as mentioned by Yayon et al. (Figure 2b)[25]. Differential expression analysis revealed 228 upregulated genes (fold change >2) including *MAOA*, *RORA*, *GMDS*, *ERBB4*, *IGFBP7*, and *SCPEP1* (Figure 2c, Supplementary Data 5). Gene Ontology (GO) enrichment implicated TET_TECs in developmental and morphogenic processes such as cell morphogenesis, skeletal system development, muscle structure formation, regulation of growth but also regeneration and stem cell development (Figure 2d, Supplementary Data 6). Immunofluorescence (IF) of MAOA confirmed elevated protein levels in TET-A, TET-AB, TET-MNT, as well as TET-B1 and TET-B2, compared to the healthy postnatal thymus (Figure 3a, Supplementary Figure 4). Two carcinoma-specific TEC clusters (Carc_TEC_1, Carc_TEC_2) dominated the epithelial subcluster in TET_C samples. Notably, both clusters shared characteristics with the CHI3L1_mTEC subtype described by Xin et al. (Figure 2a & b) [23]. Carc_TEC_1 was characterized by a strong expression of *DUSP2*, *CCL20* and *SOX18* while Carc_TEC_2 showed high level of *CD177*, *PLAU* and *CA9*. Their transcriptional profiles indicate potential pathological roles in epithelial cell differentiation and cell migration (Figure 3b & c, Supplementary Data 6). Functionally, cTEC, mcTEC and mTEC clusters exhibited characteristics linked to T cell development, including regulation of T cell receptor signaling,

interferon responses, and immune modulation (Supplementary Figure 5, Supplementary Data 5, 6). The adult thymus persisted as expected primarily of mTEC_lo followed by mTEC_hi and almost no cTECs. TEC subtypes previously described by Bautista et al., Xin et al. and Yayon et al., including neuronendocrine and ciliated assembling cluster (TEC_neuro_ciliated), muscle-like myoid TECs (TEC_myo_1, TEC_myo_2), and tuft cells (TEC_tuft) (Figure 2b, Supplementary Figure 5, Supplementary Data 5, 6) [22,23,25]. GO and marker analysis supported the correct annotation. While TEC_tuft appeared to be present in several conditions, TEC_myo_1, as one of the main cell types in TET_C, showed a strong involvement in tissue morphogenesis.

This analysis provides evidence that a significant number of TECs in diseased thymuses exhibit a pathology-specific transcription pattern, suggesting strong involvement in morphogenetic processes. These cells show moderate similarity to bipotent TEC precursor cells in healthy thymuses but nevertheless differ significantly from them in terms of gene expression.

Integrated analysis reveals distinct fibroblast subtypes associated with TET.

Given pronounced FB alterations in TET pathology, we further dissected the detected FBs, identifying ten distinct FB subclusters (Figure 4a, Supplementary Data 7). Initial classification attempts based on Yasumizu et al. were inconclusive (Supplementary Figure 6a), prompting integration of additional fibroblast markers from Yayon et al. [25]. This enabled identification of Perilobular-, Interlobular-, and medullary fibroblasts (Supplementary Figure 6b). However, a predominant FB subcluster almost exclusively present in all TET types (TET-FB) was identified. This fibroblast group exhibited elevated expression of *TWIST1*, *ITGA11*, and *SDC1* (Figure 4b, Supplementary Data 7). TET-FB seemed to be especially associated with extracellular matrix organization and embryonic morphogenesis (Figure 4c, Supplementary Data 8). Pathway analysis underscored robust TGF-β signaling activity within this subtype (Figure 4d).

Immunofluorescence dual-staining for SDC1 and CD90 (THY1) confirmed increased presence of spindle-shaped fibroblasts in TET-A, TET-AB, TET-MNT, TET-B1, TET-B2, TET-B3, and TET-C (Figures 4e & f, Supplementary Figure 7). Two distinct cluster of perilobular fibroblasts, especially present in prenatal thymus, were identified. One positive for proliferation marker *TOP2A* (PeriloFB-Prolif) and another one positive for *GRM7* and *ADAMTSL1* with supposed involvement in tissue morphogenesis (PeriloFB) (Figure 4a, Supplementary Figure 8, Supplementary Data 8). One cluster of interlobular FBs (InterloFB), characterized by expression of genes involved in vasculature development, was only present in the adult thymus. The second Interlobular fibroblast cluster, annotated according to its expression of *COL9A3* (InterloFB_COL9A3), was especially present in prenatal thymus but also in TET_A, _AB and _B implicating functions in muscle structure development (Figure 4a, Supplementary Figure 8, Supplementary Data 8). Strikingly, one mixed cell cluster consisting of medullary, interlobular and perilobular FBs (Perilo_Interlo_medFB) built the major amount of FBs in the pediatric thymus (Figure 4a), characterized by high expression of *ECM2*, *ANGPTL1* and *MYOC* (Supplementary Figure 8, Supplementary Data 8) Two further medullary clusters were identified (medFB-MHCIIh, medFB-CCL2) present only in healthy and benign adult thymic conditions, such as TH and TET_MNT. (Supplementary Figure 8, Supplementary Data 8). One fibroblast cluster almost exclusively present in fetal tissue (Supplementary Figure 6c) showed a strong expression of *POSTN* and exhibited a gene signature associated with embryonic organ morphogenesis (Figure 4a, Supplementary Figure 8, Supplementary Data 8). Additionally, one FB type, expected to be present in fetal thymic tissue (fetal-FB-SFRP1) showed cellular presence also in TH with the potential ability to regulate stem cell differentiation (Supplementary Figure 8, Supplementary Data 8).

Overall, we show that TETs contain a specific type of fibroblast that appears to be particularly involved in forced matrix formation and morphogenic functions. Furthermore, the cellular state in

TH is examined in more detail, which, in addition to healthy adult thymus, shows particular parallels with prenatal thymus.

Since important findings regarding TECs and FBs were made during the creation of the ´thymus cell atlas in healthy and diseased conditions´, it was decided to address the subtype analyses of the remaining cell types in the supplements in order to maintain the readability of the study. In Supplementary Note1, we provide the in-depth subcluster analysis of all remaining celltypes including Figures and Tables depicting subtype annotation and distribution, DEGs and corresponding GO-terms, drawing a more complete image of the cellular spectrum (Note S1, Supplementary Figures 9-19, Supplementary Data 9 - 22).

In-depth analysis uncovers cell type-specific changes of interactions and pathway regulations in normal and pathologic thymus.

ScRNAseq enables the prediction of potential cell-cell interactions and regulatory pathways within tissues. To further explore pathway activity, we applied decoupleR for pathway inference analysis. Averaged pathway activity scores across conditions indicated increased TGF-β signaling in all pathological thymic entities compared to the healthy adult thymus, with particularly pronounced activity in EC, FB, and VSMC/Peri (Figure 5a). Statistical analyses demonstrated significant upregulation, particularly in TET-AB, TET-B and TH populations (Figure 5b).

To delineate potential variations in intercellular communication between pathological thymic states and the healthy thymus, we applied CellChat with its multiple dataset comparison functionality. Our results indicated that the postnatal thymus exhibits a significantly greater

number of cell-cell interactions than the prenatal thymus. Interestingly, more interactions were observed in less malignant conditions (Figure 5c).

Mapping the cellular origins of these interactions revealed that T cells of the healthy adult thymus were subject to particularly strong incoming signalling. In TH, TECs demonstrated the highest incoming interaction strength. Notably, Mac/Mono and FB of TETs exhibited prominent cell communication potential (Supplementary Figure 20a). Direct comparisons between each pathological condition and the healthy adult thymus showed increased interactions particularly involving FB, ECs, and VSMC/Peri (Supplementary Figure 20b). The strongest enhancement of TEC-mediated interactions was observed in TH. While TET-A, TET-AB, TET-C and TET-MNT exhibited increased TEC communication with FB, VSMC/Peri, and EC, TET-B did not show marked changes in TEC intercellular signaling.

Screening for upregulated signaling pathways (Supplementary Figure 21) revealed a significant increase of APP-CD74 signaling among various cell types in TET-A (Figure 5d). Immunofluorescence staining for APP and CD74 confirmed these findings, demonstrating strong co-localization of these proteins specifically in TET-A samples (Figure 5e, Supplementary Figure 22, Supplementary Data 23). Examination of previously identified cellular subtypes in TET-A revealed the strongest upregulation of APP in InterloFB-COL9A3 followed by TET-FB, TET-TEC and distinct EC types, while pDCs and the distinct macrophages showed the strongest expression of *CD74* (Figure 5f).

In summary, we reveal a comparison of the significantly different pathway regulations in healthy thymus at different stages of development as well as in TH and the various TETs. Furthermore, we were able to identify increased cell-cell interaction via APP-CD74 in TET-A using transcriptional analysis. In addition to the FBs referred to here as InterloFB-COL9A3, the TET-specific FBs and TECs described above also appear to be involved in this interaction.

Bulkseq data with known *GTF2I* mutation disclose specific gene expression pattern

Radovich et al. uncovered the increased occurrence of *GTF2I* mutations, specifically at a single codon (L424H), in humans developing TET-A and TET-AB in comparison with all distinct TET-types by exome sequencing [8]. To uncover potential *GTF2I*-mutation-affected genes, we performed bulkseq analysis using data published by Radovich et al. (Figure 6a). A general bulk-seq comparison of all *GTF2I* positive TETs with the remaining ones uncovered 109 up- and 308 down-regulated genes (Figure 6a, Supplementary Data 24). GO term analysis of these two gene lists go along with the features reported by Radovich et al., including participation in WNT signalling pathway, but also show involvement of *GTF2I*-affected genes in matrix shaping processes and cell differentiation (Figure 6b & c, Supplementary Data 25) [8]. To uncover potential *GTF2I* mutations in our scRNAseq data, we formed module scores of the up and down regulated genes and checked their expression in the scRNAseq donors. Strikingly, the *GTF2I*-Up regulated gene pattern was positive in TET-A (Figure 6d).

In summary, the officially published Bulkseq and mutation analyses were combined with the scRNAseq analyses performed here to identify and verify the presence of up- and downregulated transcriptional patterns in *GTF2I*-mutated TETs in the tissues included here. Interestingly, the gene pattern was specifically confirmed under the benign TET-A condition, which is consistent with the published data. Through this analysis, a potential *GTF2I* mutation in our scRNAseq data cannot be proven, however the combination of up-regulated genes is suggested as promising indicator for a benign pathologic state.

Discussion

Previous studies have begun to elucidate the cellular heterogeneity of the healthy and diseased thymus using scRNA-seq [16,20,22,23]. Here, we integrated these published datasets with generated data from TETs and included the additional condition “thymic hyperplasia” (TH). This comprehensive approach allowed us to construct a unified cellular landscape of both healthy and pathological thymus, revealing conserved cellular patterns across studies, transcriptional distinctions between healthy and diseased tissues, and refining the classification of thymic cell types.

Unlike conventional approaches relying on single marker genes [16,20,22,23], our inclusion of module scores allowed more precise cell type assignments, capturing nuanced heterogeneity beyond individual markers. Despite the limitations of scRNAseq data due to methodological cellular enrichment of various groups, a combination of scRNAseq and bulk deconvolution made it possible to create a highly reliable picture of the cellular composition of different TETs. Consistent with literature, TET-A samples contained over 60% TECs, underscoring their central role in pathology, while TCs were present at levels comparable to endothelial cells [10]. TET-B1 was dominated by TCs, with their abundance declining across TET-B2 and B3 as TECs increased [11,26]. TET-AB exhibited an intermediate profile with both TECs and TCs, aligning with its histology [11]. TET-C and TET-MNT, largely TEC-driven, also showed increased presence of BC, DC, FB, MAC/Mono, and EC, reflecting their histologic and transcriptomic similarity [11,13].

Our TH samples, processed without cellular enrichment, revealed a native-like cellular composition with over 80% TCs, consistent with previous reports [4]. Although scRNA-seq quantification must be interpreted cautiously due to tissue dissociation biases [27], our integrated dataset provides a hint of the relative cell proportions. However, for a clear confirmation of the cellular composition of the tissue based on the scRNAseq data, future studies with extensive multiplex immunohistochemistry analysis using celltype specific markers would be necessary.

TECs remain pivotal in driving TET subtypes, but cell-cell communication analyses revealed a more complex picture. Robust TEC signaling was mainly observed in TH, with only TET-A and TET-MNT showing increased TEC interactions versus healthy thymus. Non-TEC populations exhibited subtype-specific increased communication (VSMC/Peri and FB in TET-A, TET-MNT, TET-C) highlighting the tumor microenvironment's multifaceted contribution. Although the total number of ligand-receptor pairs did not differ markedly, qualitative differences emerged. Notably, APP-CD74 interaction was upregulated in TET-A, a pathway linked to tumor progression, Alzheimer's disease, and fibrosis [28-32]. Subtype analysis shows a strong expression of APP especially in InterloFB-COL9A3, but also in TET-TEC, TET-FB and arterial, venous and capillary ECs. A significant upregulation of *CD74* was detected in pDCs and distinct macrophage subtypes, suggesting an particular interaction of these cellular subtypes. The interaction of ECs and MACs is particularly likely in this environment, as it is known to be induced by an increase in TGF-ß [30]. This aligns with heightened fibrotic signaling and TGF-β pathway activation, particularly in TET-A, but also in the remaining pathological conditions, echoing Yu et al. [33]. TGF-β is a well-established promoter of tumor growth, fibrosis, and angiogenesis [34]. These data emphasize that both TECs and other cell types contribute to aberrant TET tissue architecture and identifying key ligand-receptor axes might serve as potential therapeutic targets. As such, Bintrafusp alfa, a bifunctional inhibitor of TGF-β and PD-L1, exemplifies a promising treatment under investigation for TETs [35].

A direct comparison of healthy and diseased tissues at the cell-type level provided initial insights into cellular alterations, further refined by deeper analyses. We identified a distinct TEC population, termed TET_TEC, predominantly present in TH and all TET subtypes except metaplastic Type-C. This population showed strong transcriptional overlap with $KRT14^{+}$ and medullary-cortical-like TECs described by Xin et al. [23], both linked to thymic epithelial progenitor cells (TEPCs) [23]. TEPCs serve as common progenitors for cortical and medullary TECs during

fetal and early postnatal thymus development [36,37]. In comparison, an independent mcTEC cluster was identified in healthy thymus. When comparing these results with those of healthy thymus at different stages published by Yayon et al., it is striking that there is only weak agreement between mcTEC marker genes such as *DLK2* or *CCL2* [25]. Only *IGFBP6* and *CCN2* show weak positivity. This comparison, in conjunction with the preliminary data from Xin et al., suggests that TET-TECs are in a progenitor-like state, which differs from mcTECs in healthy thymus based on the expression of specific genes. Conversely, only a few of the cluster-defining genes of TET-TECs, such as *APBB2*, *TGM2*, *GPC4*, and *RBMS3*, appear in healthy mcTECs. Especially *TGM2* and *GPC4* are known to support the development and growth of organs [38,39].The TET_TEC cluster is marked by a unique gene signature including *GMDS*, *EYA1*, *ZBTB20*, *GAS6*, *IGFBP7*, and *MAOA*, genes associated with various cancers [40-49]. In particular, *ZBTB20* promotes proliferation, migration, and apoptosis resistance [45], while *KIAA1217*, *RBMS3*, and *SCPEP1* are implicated in epithelial-to-mesenchymal transition (EMT) [50,51], with *RBMS3* additionally inhibiting WNT signaling [52]. *MAOA* and *GAS6* are potential cancer therapy targets [46-48]. Conversely, *TBC1D5*, upregulated here, acts as a tumor suppressor potentially limiting metaplastic progression in less aggressive forms of TETs [53]. The gene ontology terms related to cell morphogenesis and development suggest notable plasticity and influence on thymic architecture. While these data position TET_TECs as likely pathogenic drivers across multiple TET subtypes, functional studies are needed to confirm their role.

Rare TEC subtypes previously described by Bautista et al. (neuroendocrine, ciliated, myoid, tuft/ionocyte, keratinocyte-like TECs) and by Xin et al. ($CHGA^+$ and $MYOG^+$ TECs) were primarily connected in one common mixed cluster [23]. This supports the transcriptional equivalence of $MYOG^+$ muscle-like TECs and myoid TECs, suggesting shared identity or close lineage relationships. The core transcriptional program of this cluster resembles conventional cortical and medullary TECs, indicating fundamental similarity despite phenotypic diversity. The

Carc_TEC clusters provide an initial molecular snapshot of thymic squamous carcinoma microenvironment, which accounts for 58% of type C thymomas. Although these clusters show fundamentally differences in their gene expression, both clusters appear to have a strong impact on epithelial cell differentiation. The expression of *CCL20* in CARC-TEC-1 and *MUC4* and *MUC1* in Carc-TEC-2 go in line with published literature discussing these factors as potential biomarkers for thymic carcinoma [54-56]. More comprehensive insights require additional data from rarer thymic carcinomas such as hyalinizing clear cell and sebaceous carcinomas [9,13].

Looking at the results of the FBs as a whole, it is striking that medFB-CCL2 accounts for a large proportion of the healthy adult thymus and in TET-MNT, whereas the medFB-MHCIIh are found particularly in adult healthy thymus. According to Yayon et al., medullary FBs are found primarily in the postnatal state in thymic tissue, which is also reflected in the results shown here, while peri- and interlobular FBs are primarily present in the prenatal thymus. In contrast to TET_TEC, the TET_FB cluster is found across all TET subtypes but absent in healthy postnatal thymus and TH. These fibroblasts exhibit strong signatures of extracellular matrix remodeling and TGF-β signalling. While fibroblasts contribute to thymic organogenesis and TEC development under physiological conditions [57,58], they can facilitate tumor progression by promoting epithelial transformation and proliferation [59-61]. We observed increased expression of *SDC1* in TET_FB, corroborated at the protein level by immunofluorescence. Although elevated stromal SDC1 often correlates with poor prognosis in cancers [62], its higher expression in benign TETs may indicate a favorable outcome, suggesting SDC1 as a potential therapeutic target. Antibodies targeting SDC1, such as indatuxumab ravtansine and VIS832, currently tested in clinical trials for multiple myeloma, may have potential in TETs [63-66].

TET_FB cells upregulate fibrosis- and EMT-associated genes including *FN1*, *ITGA11*, *MMP11*, *CTHRC1*, *TWIST1*, and *SFRP4* [67-73]. MMP11 supports invasion and migration, while CTHRC1 promotes ECM deposition and tumor progression. SFRP4, a WNT inhibitor, shows cancer-type

dependent expression and impacts ECM formation, especially in prostate cancer [72]. The upregulation of *HTRA1*, a modulator of TGF-β signaling and fibroblast transdifferentiation into cancer-associated fibroblasts (CAFs), further implicates TET_FB cells in tumor stroma remodeling [74,75]. Collectively, these data highlight TET_FBs as important contributors to EMT, fibrosis, and tumor microenvironment remodeling in TETs. Given the limited current focus on CAFs in thymomas, deeper exploration of their roles could be relevant for potential treatment options in the future.

This study examines TH using scRNAseq. As a benign phenomenon, TH is primarily characterized by the presence of an oversized thymus that continues to function normally in adulthood. It is a condition easy to treat surgically, but nevertheless represents a stressful condition for many patients. Here, it was shown that TECs with numerous cell-cell interactions appear to be the most active cell-type. Furthermore, the TGF-ß pathway is significantly upregulated in EC, similar to more serious thymus diseases. At the cellular level, there are strong similarities to healthy adult thymus (FBs), prenatal thymus (PCs), and TETs (TECs). However, special fibroblasts which are primarily found in fetal tissue (fetal-FB-SFRP1) were also detected in TH. These FBs mainly express factors associated with stem cell differentiation and response to tumor necrosis factor. Top regulated genes of this celltype like *DLX5*, *SFRP1* and *HES1* [76] are mentioned to be involved in embryogenesis and organ development[77-79]. These findings suggest TECs and FBs as key drivers in TH.

In summary, our integrated single-cell analysis provides a comprehensive and refined cellular atlas of the healthy and pathological thymus, encompassing multiple TET types and TH. We highlight the pivotal role of TEC, particularly the TET_TEC population, in driving tumor pathogenesis, alongside a dynamic interplay with diverse stromal and immune cell types that shape the tumor microenvironment. Our findings emphasize the complexity of cell-cell communication and the significance of fibrotic and immunoregulatory pathways, such as TGF-β

signaling, in thymic tumor progression. Importantly, the identification of specific molecular signatures and ligand-receptor interactions opens promising avenues for targeted therapies, including agents like Bintrafusp alfa and anti-SDC1 antibodies. This work not only advances our understanding of thymic tumor biology but also lays a foundation for developing precision medicine approaches tailored to the distinct cellular landscapes of thymic diseases.

Methods

Study approval/ Ethics statement

The study was conducted according to the guidelines of the Declaration of Helsinki. The use of resected tissue was approved by the ethics committee of the Medical University of Vienna (vote 2167/2020).

Tissue acquisition

Human tissue samples of 6 TH, 4 TET-B (Type B1-3) and 2 TET-C were obtained from patients, who underwent extended thymectomy (Supplementary Table 1). Surplus tissue not required for pathologic evaluation was processed. Patients who had received any previous chemotherapy or radiation treatment were excluded from the study. Upon surgical resection, the tissue was examined by a pathologist for clinical and/or research purposes. The included patients were aged between 23 and 77 years.

Generation of single-cell suspensions

Samples were rinsed with sterile Dulbecco`s phosphate-buffered saline (PBS, without Ca^{2+} and Mg^{2+}, Gibco, Thermo Fisher Scientific, Waltham, MA, USA), mechanically minced and enzymatically dissociated using MACS Miltenyi Tumor Dissociation Kit, human (Miltenyi Biotec, Bergisch Gladbach, Germany) in accordance with the manufacturer´s instructions. Sample-Enzyme mix were placed in gentleMACS C-tubes (Miltenyi) and digested on the gentleMACS OctoDissociator (Miltenyi) with the selected gentleMACS program 37C_h_TDK_2. The resulting cell suspension was sequentially passed through 100 and 40 µm cell strainers and washed twice with 0.04% bovine serum albumin (BSA, Sigma Aldrich, St. Louis, MO, USA) in PBS. Cellular viability and concentration were determined by LUNA-FL™ Dual Fluorescence Cell Counter (Logos Biosystems, Anyang-si, Gyeonggi-do, South Korea) using the Acridine Orange/Propidium Iodide (AO/PI) Cell Viability Kit (Logos Biosystems). Suspensions with a cell

viability <80% were processed with Dead Cell Removal Kit (Miltenyi). The final cell concentration was adjusted to 0.7-1.2*10^6 cells/ml. All isolation procedures were performed within 4 hours after surgery and samples were kept on ice for transport.

Single cell RNA sequencing

Gel bead-in-emulsion (GEM) preparation, cDNA amplification and library preparation were performed with Chromium Next GEM Single Cell 5´ Kit v2 and Dual Index Kit TT Set A (all 10x Genomics, Pleasanton, CA, USA). Viable cells were loaded on Chromium Next GEM Chips type K (10X Genomics), GEMs were prepared using the Chromium controller (10X Genomics). The Biomedical Sequencing Core Facility of the Center for Molecular Medicine (CeMM, Vienna, Austria) was commissioned for RNA-sequencing, demultiplexing and counting. All samples were sequenced paired-end by Illumina HiSeq 75PE, or Illumina NovaSeq SP, 50PE (all Illumina, San Diego, CA, USA). Raw sequencing files were demultiplexed and aligned to the human reference genome (GrCh38) and counted using the Cellranger pipelines (Cellranger v5.0.1, 10x Genomics).

Data acquisition

Transcriptional datasets of distinct TET types (A, AB, B3, C, MNT) published by Xin et al. (2022) were downloaded from the Genome Sequence Archive in National Genomics Data Center (HRA002334) [https://ngdc.cncb.ac.cn/gsa-human/browse/HRA002334] [23]. Single cell RNA sequencing from Yasumizu et al. (2022), were downloaded from the JGA (JGAS000482) [https://ddbj.nig.ac.jp/resource/jga-study/JGAS000482] [16]. Transcriptomic data from Bautista et al. (2021) were downloaded from the Gene Expression Omnibus (GEO) database (GSE147520) [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE147520] [22]. Datasets from Park et al. (2020) were downloaded from ArrayExpress (E-MTAB-8581) [https://www.ebi.ac.uk/biostudies/ArrayExpress/studies/E-MTAB-8581?query=E-MTAB-8581] [20].

Bulkseq data were downloaded from GDC Data Portal (TCGA-THYM) [https://portal.gdc.cancer.gov/projects/TCGA-THYM] [8].

Pre-processing, quality control and sketch integration

Single cell RNA sequencing data were analyzed using R (v4.3.1, The R Foundation, Vienna, Austria), R-studio and Seurat (Seurat v5.0.2, Satija Lab) [80]. Unique molecular identifier (UMI) count matrices were generated for each individual dataset. Features of all datasets were screened and doublets removed. The feature designation was harmonized and only features detected in all datasets were kept for further analysis. Potential ambient RNA contamination was removed using DecontX [81]. Synchronized datasets were converted to Seurat Objects. Erythrocytes, identified by their expression of hemoglobin subunit beta [HBB >=5], were removed from all datasets. Data normalization and variable feature selection was performed. Cells with unique feature counts above 3000 or below 200 as well as cells with more than 7% mitochondrial counts were excluded from downstream analysis. A representative set of 5000 cells from each dataset was then identified applying a leverage score computation. The selected cells were then preprocessed using first the “FindVariableFeatures”, “ScaleData” commands followed by a principal component analysis (PCA) dimensionality reduction by default. Due to the complexity of the study design with data sets from different working groups, each of which does not contain all cell types due to enrichment, several tools were tested as integration methods. RPCA and Harmony achieved the best results with no significant differences in cluster distribution between each other. Since all data sets originate from the same platform (10x), not all cell fractions are contained in the included data sets, and due to the extensive amount of data, the analysis was ultimately performed using RPCA. The RPCA Integration method with 30 dimensions and 20 k.anchors was subsequently utilized for integration UMAP-Plots split by tissue, coloured according to the batch are depicted in Supplementary Figure 23.

"FindNeighbors", "FindClusters" and "RunUMAP" were further applied for graph-based clustering of the representative cell selection. Complete datasets were then loaded in individually and all cells were integrated in the previously determined integrated low-dimensional space. Clusters were projected on the full datasets applying the "ProjectData" command.

Cluster annotation using Modul score calculation

After the described integration process, the identified clusters were annotated applying Seurats "AddModuleScore" function. The top differentially expressed gene sets of the main clusters published by Bautista et al. [22], Yasumizu et al. [16] and Xin et al. [23] were used and classification of the here identified cluster was performed with all results. Manual inspection/annotation of remaining doublets has been performed on main and subset level using canonical markergenes, with a particular focus on TC and TEC clusters. Since stromal-stromal doublets have not been screened, potential intermediate stromal doublets might still remain in the dataset. Although HBB-positive erythrocytes were previously excluded, a cluster consisting of Red blood cells has been identified. The mentioned analytical impurities were excluded from further downstream analysis. Clustermarker lists were obtained using the "FindAllMarkers" feature (min.pct = 0.25, logfc.threshold = 0.25, test.use= "wilcox").

Cell type specific subclustering

Subset analyses of all clusters were performed based on the raw data of the cell cluster. Variable feature selection was performed followed by data scaling and PCA. "RunUMAP", "FindNeighbors" and "FindClusters" were performed for UMAP dimensionality reduction and Louvain clustering. Thymic epithelial cells and T-cells were annotated according to cell type-specific module scores created from Top differentially regulated expressed genes published by Xin et al. [23] and Yasumizu et al. and by marker genes published by Yayon et al. [16,25]. B-cells were classified according to specific marker genes published by Stewart et al. [82]. Dendritic cell

subtypes were assigned based on the markers published by Yayon and Yasumizu et al. [16,25]. Macrophage/monocyte subtypes were identified using module scores published by Yayon et al. and Yasumizu et al. [16] in combination with well-known marker genes [25,83]. Identification of subtypes of 1) plasma cells, 2) endothelial cells and 3) vascular smooth muscle cells according to marker genes published by Yayon et al. [25]. The relative composition of fibroblasts and thymus epithelial cells was determined by calculating the average proportions of cellular percentages within each individual donor. Donors with fewer than 50 cells and a composition consisting of only one clustertype were excluded.

Differential gene expression analysis

Seurat´s “FindAllMarkers” and “FindMarkers” commands were used to identify genes with a significant differential expression between distinct cell groups and to uncover cluster-specific marker genes. Only genes detected in a minimum fraction of 0.25 in either of the compared cell populations with a minimum of 0.25-fold difference (log-scale) between the distinct groups were included in the clustermarker computation. Specific differential gene expression analysis was performed with default settings.

Enrichment analysis

Gene Ontology (GO) enrichment analysis was performed using Metascape with a minimum enrichment score of 2 and a p-value cutoff of 0.05 [84]. Genes resulting from differential gene expression analysis with an average fold change ≥ 2 were included to obtain information on biological function.

Cell-cell communication inference and pathway analysis

The CellChat package was used to reveal potential cell-cell interactions within the distinct tissue types [85]. For comparative analysis, condition-specific CellChat objects were merged with the CellChat object from healthy thymus, including prenatal, pediatric and adult datasets. The R package decoupleR was used according to the developer vignettes to identify potential pathway

alterations and transcription factor activities [86]. The average pathway activities determined from all donors are shown. Statistical analyses was also performed using the detected values of each individual donor.

Extraction of *GTF2I* expression pattern from Bulkseq data

TCGA-Thym bulk RNA seq data published by Radovich et al. [8] and the donor- specific clinical information were downloaded using the R package TCGAbiolink. The genetic pattern specific for the *GTF2I* mutation was identified after adaptation of the code published by Yasumizu et al. using the same packages as described [16]. Up- and downregulated genes in GTF2I-positive donors compared to wildtype donors with a log2 fold change higher than 2 and an adjusted p-value <0.01 were used for downstream analysis. The “final diagnosis” described in the donor-specific clinical information was used as diagnosis. Since literature considers TET types as autonomous pathologic process without an explicit relation, except for TET-Bs, we first aimed to identify a type-specific gene pattern, comparing exclusively TET-ABs with and without *GTF2I* mutation. However, this analysis revealed no differentially gene regulation with a fold change of 2 or higher. Therefore, the entire range of samples was analyzed. The resulting gene lists were converted to NCBI symbols and potential duplicates were removed by default. Only genes also present in the Seurat object were kept for the next steps. Module Scores were created for up- and for downregulated genes and the distinct expression was visualized as Dotplots.

Bulkseq deconvolution

Deconvolution of bulk RNA-seq data was performed applying the “granulator” package maintained by Pfister et al. [87]. Counts of scRNAseq data were transformed into transcript per Million (tpm) followed by calculation of average tpm per gene and cluster. Deconvolution was then performed according the vignette published by Kuettel et al. using the linear mixing model

(https://bioconductor.org/packages/release/bioc/vignettes/granulator/inst/doc/granulator.html, accessed on 04/12/2024) [24].

Data analysis software

The following R packages were used for analyses: granulator_1.10.0, decontX_1.0.0, DT_0.32, TCGAmutations_0.4.0, data.table_1.15.2, maftools_2.18.0, scales_1.3.0, lubridate_1.9.3, forcats_1.0.0, stringr_1.5.1, purrr_1.0.2, readr_2.1.5, tidyverse_2.0.0, pals_1.8, vsn_3.70.0, plyr_1.8.9, EnhancedVolcano_1.20.0, biomaRt_2.58.2, DESeq2_1.42.1, TCGAbiolinks_2.30.0, ComplexHeatmap_2.18.0, RColorBrewer_1.1-3, enrichR_3.2, celda_1.18.2, SingleCellExperiment_1.24.0, SummarizedExperiment_1.32.0, GenomicRanges_1.54.1, GenomeInfoDb_1.38.7, IRanges_2.36.0, S4Vectors_0.40.2, MatrixGenerics_1.14.0, matrixStats_1.2.0, pheatmap_1.0.12, tibble_3.2.1, decoupleR_2.8.0, readxl_1.4.3, rstatix_0.7.2.999, Matrix.utils_0.9.7, Matrix_1.6-5, limma_3.58.1, ggsignif_0.6.4, clustree_0.5.1, ggraph_2.2.1, tidyr_1.3.1, CellChat_1.1.3, Biobase_2.62.0, BiocGenerics_0.48.1, igraph_2.0.3, ggrepel_0.9.5, SeuratWrappers_0.3.0, sctransform_0.4.1, patchwork_1.2.0, xlsx_0.6.5, ggplot2_3.5.0, magrittr_2.0.3, dplyr_1.1.4, SeuratObject_5.0.1, sp_2.1-3, viridis_0.6.5, viridisLite_0.4.2, Polychrome_1.5.1

All analyses were conducted using the cited packages, according to publicly available instructions, provided by the developers. No novel data analysis tools have been generated in this study.

Immunofluorescence

After deparaffinization and antigen unmasking using 10x R-Universal Epitope Recovery Buffer (Aptum Biologics Ltd, Southampton, UK) sections were incubated with primary antibody diluted according to Supplementary Table 3 in 2% PBS/BSA at 4°C over night. The anti-SDC1 antibody was directly labeled using the FlexAble CoraLite® Plus 488 Antibody Labeling Kit (Proteintech,

Manchester, UK; Cat. No. KFA001) following the manufacturer's instructions. Sections were washed in PBS three times for 5 min and incubated with secondary antibody in PBS with 2% BSA, 1% goat serum and DAPI (1:1000) (ThermoFisher Scientific, cat#: 62248) diluted according to Supplementary Table 3 for 30 min at room temperature. Sections were washed three times in PBS, twice with dH_2O, and mounted using mounting medium (Dako/Agilent Technologies, Santa Clara, CA, USA; Cat. No. S3023). Images were acquired on a Zeiss LSM 700 laser scanning confocal microscope (Carl Zeiss Microscopy) with the ZEN 2012 SP1 (black edition) software (Version 8.1.0.484). All images of a specific staining were captured using strictly identical microscopic settings with standardized exposure. During subsequent image processing, the intensity of DAPI was optimized linear to improve the visualization of cell nuclei using ImageJ v1.53c [88].

Hematoxylin and eosin (HE-stainings)

HE staining was performed on 2 µm sections of 7.5% neutral formalin-fixed, paraffin- embedded tissues according to routine protocols.

Study overview

The schematic visualisation of the study setting was created using Adobe Illustrator CS6 v16.0.3 (Adobe Inc., San Jose, California, USA).

Statistics & Reproducibility

Specifically generated and publicly available scRNAseq and bulkseq datasets were analyzed for the study. No statistical method was used to predetermine the sample size for scRNAseq or Bulkseq analyses. These are analyses of rare diseases and therefore no data were excluded. During the scRNAseq analyses, poor-quality cell data were eliminated from further investigation.

The biological N for validating immunofluorescence was extrapolated from previous work, with similar research questions and study designs. The experiments were not randomized. The investigators were not blinded. Normal distribution within a group was tested by Shapiro-Wilk test. Groups with normally distributed data were compared by one-way ANOVA with Tukey post hoc test. Groups without normal distribution were compared by Kruskal-Wallis test, followed by Dunn's post-hoc comparisons. All statistical analyses were performed in GraphPad Prism v8.0.1 (GraphPad Software, San Diego, USA). Clustermarker genes of all main cell-cluster were identified by testing for differential expression of a cellcluster against all other cells using the Wilcoxon rank-sum test implemented in Seurat (FindAllMarkers). Statistical significance of GO-term analyses was calculated using a hypergeometric test with Benjamini–Hochberg multiple testing correction implemented in metascape The statistical significance of ligand–receptor interactions was determined using a permutation test implemented in the Cellchat package. P-values for bioinformatical data were marked in figure using asterisks indicating *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$. Exact p-values of GO-term and Cellchat analyses can be found in the supplementary information.

Informed Consent Statement

All donors, who were registered by our research group, provided their written informed consent and agreed to the publication of the anonymized patient data listed (such as rare diagnosis, sex or age). Participation in the study was voluntary, and there was no compensation for expenses. Since this analysis concerns rare diseases, the sex of the donors was not taken into account due to the small number of samples. Information regarding donor consent for the datasets that have already been published can be found in the relevant studies.

Data Availability Statement

The ScRNAseq data generated in this study have been deposited in the NCBI´s Gene Expression Omnibus (GEO) database under GEP series accession number GSE228033 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE228033]. All the data used to produce the figures in this study are available in the Supplementary and Source Data files, which accompany this paper. Publicly available datasets were downloaded from the following sources: Xin et al., 2022 (HRA002334) [https://ngdc.cncb.ac.cn/gsa-human/browse/HRA002334] [23]; Yasumizu et al., 2022 (JGAS000482) [https://ddbj.nig.ac.jp/resource/jga-study/JGAS000482] [16]; Bautista et al., 2021 (GSE147520) [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE147520] [22]; Park et al., 2020 (E-MTAB-8581) [https://www.ebi.ac.uk/biostudies/ArrayExpress/studies/E-MTAB-8581?query=E-MTAB-8581] [20]; Radovich et al. (2018) (TCGA-THYM [https://portal.gdc.cancer.gov/projects/TCGA-THYM] [8]

Code Availability Statement

No original custom code was developed within this work. Only previously published and open-access software was used following developer's instructions, cited and described in detail under the relevant method subsections and supplementary notes. The R-code applied in this study is accessible under the following link: [https://github.com/Mwielscher/scRNAseq/tree/main/thymus] [89]. A README file has been included to the link for general guidance.

References

1 Kyewski, B. & Derbinski, J. Self-representation in the thymus: an extended view. *Nat Rev Immunol* **4**, 688-698, doi:10.1038/nri1436 (2004).

2 Hogquist, K. A., Baldwin, T. A. & Jameson, S. C. Central tolerance: learning self-control in the thymus. *Nat Rev Immunol* **5**, 772-782, doi:10.1038/nri1707 (2005).

3 Francis, I. R., Glazer, G. M., Bookstein, F. L. & Gross, B. H. The thymus: reexamination of age-related changes in size and shape. *AJR Am J Roentgenol* **145**, 249-254, doi:10.2214/ajr.145.2.249 (1985).

4 Ricci, C. *et al.* True thymic hyperplasia: a clinicopathological study. *Ann Thorac Surg* **47**, 741-745, doi:10.1016/0003-4975(89)90131-8 (1989).

5 Suster, D., Ronen, N., Pierce, D. C. & Suster, S. Thymic Parenchymal Hyperplasia. *Mod Pathol* **36**, 100207, doi:10.1016/j.modpat.2023.100207 (2023).

6 Marx, A., Ströbel, P. & Weis, C.-A. The pathology of the thymus in myasthenia gravis. *Mediastinum* **2** (2018).

7 Detterbeck, F. C. Evaluation and treatment of stage I and II thymoma. *J Thorac Oncol* **5**, S318-322, doi:10.1097/JTO.0b013e3181f20dab (2010).

8 Radovich, M. *et al.* The Integrated Genomic Landscape of Thymic Epithelial Tumors. *Cancer Cell* **33**, 244-258.e210, doi:10.1016/j.ccell.2018.01.003 (2018).

9 Marx, A. *et al.* The 2021 WHO Classification of Tumors of the Thymus and Mediastinum: What Is New in Thymic Epithelial, Germ Cell, and Mesenchymal Tumors? *J Thorac Oncol* **17**, 200-213, doi:10.1016/j.jtho.2021.10.010 (2022).

10 Marx, A. *et al.* The 2015 World Health Organization Classification of Tumors of the Thymus: Continuity and Changes. *J Thorac Oncol* **10**, 1383-1395, doi:10.1097/jto.0000000000000654 (2015).

11 Travis, W. D. *et al.* The 2015 World Health Organization Classification of Lung Tumors: Impact of Genetic, Clinical and Radiologic Advances Since the 2004 Classification. *J Thorac Oncol* **10**, 1243-1260, doi:10.1097/jto.0000000000000630 (2015).

12 Kim, H. K. *et al.* Type B thymoma: is prognosis predicted only by World Health Organization classification? *J Thorac Cardiovasc Surg* **139**, 1431-1435.e1431, doi:10.1016/j.jtcvs.2009.10.024 (2010).

13 Weissferdt, A. & Moran, C. A. Thymic carcinoma, part 1: a clinicopathologic and immunohistochemical study of 65 cases. *Am J Clin Pathol* **138**, 103-114, doi:10.1309/ajcp88fztwanlrcb (2012).

14 von Gaudecker, B. Functional histology of the human thymus. *Anat Embryol (Berl)* **183**, 1-15, doi:10.1007/bf00185830 (1991).

15 Jovic, D. *et al.* Single-cell RNA sequencing technologies and applications: A brief overview. *Clin Transl Med* **12**, e694, doi:10.1002/ctm2.694 (2022).

16 Yasumizu, Y. *et al.* Myasthenia gravis-specific aberrant neuromuscular gene expression by medullary thymic epithelial cells in thymoma. *Nat Commun* **13**, 4230, doi:10.1038/s41467-022-31951-8 (2022).

17 Zeng, Y. *et al.* Single-Cell RNA Sequencing Resolves Spatiotemporal Development of Pre-thymic Lymphoid Progenitors and Thymus Organogenesis in Human Embryos. *Immunity* **51**, 930-948.e936, doi:10.1016/j.immuni.2019.09.008 (2019).

18 Kernfeld, E. M. *et al.* A Single-Cell Transcriptomic Atlas of Thymus Organogenesis Resolves Cell Types and Developmental Maturation. *Immunity* **48**, 1258-1270.e1256, doi:10.1016/j.immuni.2018.04.015 (2018).

19 Lee, M. *et al.* Single-cell RNA sequencing identifies shared differentiation paths of mouse thymic innate T cells. *Nat Commun* **11**, 4367, doi:10.1038/s41467-020-18155-8 (2020).

20 Park, J. E. *et al.* A cell atlas of human thymic development defines T cell repertoire formation. *Science* **367**, doi:10.1126/science.aay3224 (2020).

21 Le, J. *et al.* Single-Cell RNA-Seq Mapping of Human Thymopoiesis Reveals Lineage Specification Trajectories and a Commitment Spectrum in T Cell Development. *Immunity* **52**, 1105-1118.e1109, doi:10.1016/j.immuni.2020.05.010 (2020).

22 Bautista, J. L. *et al.* Single-cell transcriptional profiling of human thymic stroma uncovers novel cellular heterogeneity in the thymic medulla. *Nat Commun* **12**, 1096, doi:10.1038/s41467-021-21346-6 (2021).

23 Xin, Z. *et al.* The immune landscape of human thymic epithelial tumors. *Nat Commun* **13**, 5463, doi:10.1038/s41467-022-33170-7 (2022).

24 Hunt, G. J., Freytag, S., Bahlo, M. & Gagnon-Bartsch, J. A. dtangle: accurate and robust cell type deconvolution. *Bioinformatics* **35**, 2093-2099, doi:10.1093/bioinformatics/bty926 (2019).

25 Yayon, N. *et al.* A spatial human thymus cell atlas mapped to a continuous tissue axis. *Nature* **635**, 708-718, doi:10.1038/s41586-024-07944-6 (2024).

26 Ströbel, P. *et al.* Corticomedullary differentiation and maturational arrest in thymomas. *Histopathology* **64**, 557-566, doi:10.1111/his.12279 (2014).

27 Denisenko, E. *et al.* Systematic assessment of tissue dissociation and storage biases in single-cell and single-nucleus RNA-seq workflows. *Genome Biol* **21**, 130, doi:10.1186/s13059-020-02048-6 (2020).

28 Ma, C. *et al.* Therapeutic modulation of APP-CD74 axis can activate phagocytosis of TAMs in GBM. *Biochim Biophys Acta Mol Basis Dis* **1870**, 167449, doi:10.1016/j.bbadis.2024.167449 (2024).

29 Matsuda, S., Matsuda, Y. & D'Adamio, L. CD74 interacts with APP and suppresses the production of Abeta. *Mol Neurodegener* **4**, 41, doi:10.1186/1750-1326-4-41 (2009).

30 Liu, B. *et al.* APP-CD74 axis mediates endothelial cell-macrophage communication to promote kidney injury and fibrosis. *Front Pharmacol* **15**, 1437113, doi:10.3389/fphar.2024.1437113 (2024).

31 Chen, G. *et al.* Single-cell transcriptomic analysis reveals that the APP-CD74 axis promotes immunosuppression and progression of testicular tumors. *J Pathol* **264**, 250-269, doi:10.1002/path.6343 (2024).

32 Li, R. Q., Yan, L., Zhang, L., Zhao, Y. & Lian, J. CD74 as a prognostic and M1 macrophage infiltration marker in a comprehensive pan-cancer analysis. *Sci Rep* **14**, 8125, doi:10.1038/s41598-024-58899-7 (2024).

33 Yu, L., Ke, J., Du, X., Yu, Z. & Gao, D. Genetic characterization of thymoma. *Sci Rep* **9**, 2369, doi:10.1038/s41598-019-38878-z (2019).

34 Gulley, J. L. *et al.* Dual inhibition of TGF-β and PD-L1: a novel approach to cancer treatment. *Mol Oncol* **16**, 2117-2134, doi:10.1002/1878-0261.13146 (2022).

35 Yang, Y., Yu, Y., Fan, Y. & Li, H. Evolving treatment landscape in thymic epithelial tumors: From mechanism to therapy. *Biochim Biophys Acta Rev Cancer* **1879**, 189145, doi:10.1016/j.bbcan.2024.189145 (2024).

36 Rossi, S. W., Jenkinson, W. E., Anderson, G. & Jenkinson, E. J. Clonal analysis reveals a common progenitor for thymic cortical and medullary epithelium. *Nature* **441**, 988-991, doi:10.1038/nature04813 (2006).

37 Bleul, C. C. *et al.* Formation of a functional thymus initiated by a postnatal epithelial progenitor cell. *Nature* **441**, 992-996, doi:10.1038/nature04850 (2006).

38 Lai, T. S., Davies, C. & Greenberg, C. Tissue Transglutaminase Enhances Fibrin-Dependent Angiogenesis and Extracellular Matrix Formation During Tissue Repair by Altering Gene Expression and Is Inhibited by Aspirin. *Blood* **114**, 3055-3055, doi:10.1182/blood.V114.22.3055.3055 (2009).

39 Filmus, J. The function of glypicans in the mammalian embryo. *Am J Physiol Cell Physiol* **322**, C694-c698, doi:10.1152/ajpcell.00045.2022 (2022).

40 Wei, X. *et al.* GMDS knockdown impairs cell proliferation and survival in human lung adenocarcinoma. *BMC Cancer* **18**, 600, doi:10.1186/s12885-018-4524-1 (2018).

41 Li, C. M. *et al.* Gene expression in Wilms' tumor mimics the earliest committed stage in the metanephric mesenchymal-epithelial transition. *Am J Pathol* **160**, 2181-2190, doi:10.1016/s0002-9440(10)61166-2 (2002).

42 Wu, K. *et al.* EYA1 phosphatase function is essential to drive breast cancer cell proliferation through cyclin D1. *Cancer Res* **73**, 4488-4499, doi:10.1158/0008-5472.Can-12-4078 (2013).

43 Nikpour, P., Emadi-Baygi, M., Emadi-Andani, E. & Rahmati, S. EYA1 expression in gastric carcinoma and its association with clinicopathological characteristics: a pilot study. *Med Oncol* **31**, 955, doi:10.1007/s12032-014-0955-y (2014).

44 Wang, Q. F. *et al.* MLL fusion proteins preferentially regulate a subset of wild-type MLL target genes in the leukemic genome. *Blood* **117**, 6895-6905, doi:10.1182/blood-2010-12-324699 (2011).

45 Stoyanov, D., Stoyanov, G. S., Ivanov, M. N., Spasov, R. H. & Tonchev, A. B. Transcription Factor Zbtb20 as a Regulator of Malignancy and Its Practical Applications. *Int J Mol Sci* **24**, doi:10.3390/ijms241813763 (2023).

46 Han, H. *et al.* Monoamine oxidase A (MAOA): A promising target for prostate cancer therapy. *Cancer Lett* **563**, 216188, doi:10.1016/j.canlet.2023.216188 (2023).

47 Wu, J. B. *et al.* MAOA-Dependent Activation of Shh-IL6-RANKL Signaling Network Promotes Prostate Cancer Metastasis by Engaging Tumor-Stromal Cell Interactions. *Cancer Cell* **31**, 368-382, doi:10.1016/j.ccell.2017.02.003 (2017).

48 Wu, G. *et al.* Targeting Gas6/TAM in cancer cells and tumor microenvironment. *Mol Cancer* **17**, 20, doi:10.1186/s12943-018-0769-1 (2018).

49 Li, D. *et al.* Cancer-associated fibroblast-secreted IGFBP7 promotes gastric cancer by enhancing tumor associated macrophage infiltration via FGF2/FGFR1/PI3K/AKT axis. *Cell Death Discov* **9**, 17, doi:10.1038/s41420-023-01336-x (2023).

50 Wang, Y. *et al.* KIAA1217 Promotes Epithelial-Mesenchymal Transition and Hepatocellular Carcinoma Metastasis by Interacting with and Activating STAT3. *Int J Mol Sci* **23**, doi:10.3390/ijms23010104 (2021).

51 Paul, I. *et al.* Parallelized multidimensional analytic framework applied to mammary epithelial cells uncovers regulatory principles in EMT. *Nat Commun* **14**, 688, doi:10.1038/s41467-023-36122-x (2023).

52 Górnicki, T. *et al.* Role of RBMS3 Novel Potential Regulator of the EMT Phenomenon in Physiological and Pathological Processes. *Int J Mol Sci* **23**, doi:10.3390/ijms231810875 (2022).

53 Huang, Y. *et al.* TBC1D5 reverses the capability of HIF-2α in tumor progression and lipid metabolism in clear cell renal cell carcinoma by regulating the autophagy. *J Transl Med* **22**, 212, doi:10.1186/s12967-024-05015-y (2024).

54 Liu, X. *et al.* Abnormal Cellular Populations Shape Thymic Epithelial Tumor Heterogeneity and Anti-Tumor by Blocking Metabolic Interactions in Organoids. *Adv Sci (Weinh)* **11**, e2406653, doi:10.1002/advs.202406653 (2024).

55 Kaira, K. *et al.* MUC1 expression in thymic epithelial tumors: MUC1 may be useful marker as differential diagnosis between type B3 thymoma and thymic carcinoma. *Virchows Arch* **458**, 615-620, doi:10.1007/s00428-011-1041-x (2011).

56 Pardini, E. *et al.* Somatic mutations of thymic epithelial tumors with myasthenia gravis. *Front Oncol* **13**, 1224491, doi:10.3389/fonc.2023.1224491 (2023).

57 Kadouri, N., Nevo, S., Goldfarb, Y. & Abramson, J. Thymic epithelial cell heterogeneity: TEC by TEC. *Nat Rev Immunol* **20**, 239-253, doi:10.1038/s41577-019-0238-0 (2020).

58 Nitta, T., Ota, A., Iguchi, T., Muro, R. & Takayanagi, H. The fibroblast: An emerging key player in thymic T cell selection. *Immunol Rev* **302**, 68-85, doi:10.1111/imr.12985 (2021).

59 Olumi, A. F. *et al.* Carcinoma-associated fibroblasts direct tumor progression of initiated human prostatic epithelium. *Cancer Res* **59**, 5002-5011, doi:10.1186/bcr138 (1999).

60 Hayward, S. W. *et al.* Malignant transformation in a nontumorigenic human prostatic epithelial cell line. *Cancer Res* **61**, 8135-8142 (2001).

61 Barcellos-Hoff, M. H. & Ravani, S. A. Irradiated mammary gland stroma promotes the expression of tumorigenic potential by unirradiated epithelial cells. *Cancer Res* **60**, 1254-1260 (2000).

62 Handra-Luca, A. Syndecan-1 in the Tumor Microenvironment. *Adv Exp Med Biol* **1272**, 39-53, doi:10.1007/978-3-030-48457-6_3 (2020).

63 Iftikhar, A. *et al.* Investigational Monoclonal Antibodies in the Treatment of Multiple Myeloma: A Systematic Review of Agents under Clinical Development. *Antibodies (Basel)* **8**, doi:10.3390/antib8020034 (2019).

64 Schönfeld, K. *et al.* Indatuximab ravtansine (BT062) combination treatment in multiple myeloma: pre-clinical studies. *J Hematol Oncol* **10**, 13, doi:10.1186/s13045-016-0380-0 (2017).

65 Musto, P. & La Rocca, F. Monoclonal antibodies in relapsed/refractory myeloma: updated evidence from clinical trials, real-life studies, and meta-analyses. *Expert Rev Hematol* **13**, 331-349, doi:10.1080/17474086.2020.1740084 (2020).

66 Yu, T. *et al.* VIS832, a novel CD138-targeting monoclonal antibody, potently induces killing of human multiple myeloma and further synergizes with IMiDs or bortezomib in vitro and in vivo. *Blood Cancer J* **10**, 110, doi:10.1038/s41408-020-00378-z (2020).

67 Bansal, R. *et al.* Integrin alpha 11 in the regulation of the myofibroblast phenotype: implications for fibrotic diseases. *Exp Mol Med* **49**, e396, doi:10.1038/emm.2017.213 (2017).

68 Liu, X. *et al.* Regulation of FN1 degradation by the p62/SQSTM1-dependent autophagy-lysosome pathway in HNSCC. *Int J Oral Sci* **12**, 34, doi:10.1038/s41368-020-00101-5 (2020).

69 Zhuang, Y., Li, X., Zhan, P., Pi, G. & Wen, G. MMP11 promotes the proliferation and progression of breast cancer through stabilizing Smad2 protein. *Oncol Rep* **45**, doi:10.3892/or.2021.7967 (2021).

70 Ma, B., Ran, R., Liao, H. Y. & Zhang, H. H. The paradoxical role of matrix metalloproteinase-11 in cancer. *Biomed Pharmacother* **141**, 111899, doi:10.1016/j.biopha.2021.111899 (2021).

71 Pyagay, P. *et al.* Collagen triple helix repeat containing 1, a novel secreted protein in injured and diseased arteries, inhibits collagen expression and promotes cell migration. *Circ Res* **96**, 261-268, doi:10.1161/01.Res.0000154262.07264.12 (2005).

72 Andersen, M. K. *et al.* Spatial transcriptomics reveals strong association between SFRP4 and extracellular matrix remodeling in prostate cancer. *Commun Biol* **7**, 1462, doi:10.1038/s42003-024-07161-x (2024).

73 Yu, X. *et al.* Molecular mechanisms of TWIST1-regulated transcription in EMT and cancer metastasis. *EMBO Rep* **24**, e56902, doi:10.15252/embr.202356902 (2023).

74 Wu, H., Ma, S., Xiang, M. & Tong, S. HTRA1 promotes transdifferentiation of normal fibroblasts to cancer-associated fibroblasts through activation of the NF-κB/bFGF signaling pathway in

gastric cancer. *Biochem Biophys Res Commun* **514**, 933-939, doi:10.1016/j.bbrc.2019.05.076 (2019).

75 Rosochowicz, M. A., Kulcenty, K. & Suchorska, W. M. Exploring the Role of HtrA Family Genes in Cancer: A Systematic Review. *Mol Diagn Ther* **28**, 347-377, doi:10.1007/s40291-024-00712-2 (2024).

76 Akimoto, M. *et al.* Hes1 regulates formations of the hypophyseal pars tuberalis and the hypothalamus. *Cell Tissue Res* **340**, 509-521, doi:10.1007/s00441-010-0951-2 (2010).

77 Hao, D. *et al.* A bioactive material with dual integrin-targeting ligands regulates specific endogenous cell adhesion and promotes vascularized bone regeneration in adult and fetal bone defects. *Bioact Mater* **20**, 179-193, doi:10.1016/j.bioactmat.2022.05.027 (2023).

78 Yang, L. *et al.* An early phase of embryonic Dlx5 expression defines the rostral boundary of the neural plate. *J Neurosci* **18**, 8322-8330, doi:10.1523/jneurosci.18-20-08322.1998 (1998).

79 Trevant, B. *et al.* Expression of secreted frizzled related protein 1, a Wnt antagonist, in brain, kidney, and skeleton is dispensable for normal embryonic development. *J Cell Physiol* **217**, 113-126, doi:10.1002/jcp.21482 (2008).

80 Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e3529, doi:10.1016/j.cell.2021.04.048 (2021).

81 Yang, S. *et al.* Decontamination of ambient RNA in single-cell RNA-seq with DecontX. *Genome Biol* **21**, 57, doi:10.1186/s13059-020-1950-6 (2020).

82 Stewart, A. *et al.* Single-Cell Transcriptomic Analyses Define Distinct Peripheral B Cell Subsets and Discrete Development Pathways. *Front Immunol* **12**, 602539, doi:10.3389/fimmu.2021.602539 (2021).

83 Direder, M. *et al.* Schwann cells contribute to keloid formation. *Matrix Biol*, doi:10.1016/j.matbio.2022.03.001 (2022).

84 Zhou, Y. *et al.* Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* **10**, 1523, doi:10.1038/s41467-019-09234-6 (2019).

85 Jin, S. *et al.* Inference and analysis of cell-cell communication using CellChat. *Nat Commun* **12**, 1088, doi:10.1038/s41467-021-21246-9 (2021).

86 Badia, I. M. P. *et al.* decoupleR: ensemble of computational methods to infer biological activities from omics data. *Bioinform Adv* **2**, vbac016, doi:10.1093/bioadv/vbac016 (2022).

87 Pfister, E. K., V.; Ferrero, E. granulator: Rapid benchmarking of methods for in silico deconvolution in bulk RNA-seq data. (2025).

88 Rueden, C. T. *et al.* ImageJ2: ImageJ for the next generation of scientific image data. *BMC Bioinformatics* **18**, 529, doi:10.1186/s12859-017-1934-z (2017).

89 Direder, M. *et al.* A single-cell atlas revealing cellular heterogeneity across healthy and diseased human thymus. Github, doi:10.5281/zenodo.19474046 (2026).

Acknowledgments

We would like to thank Dr. HP Haselsteiner and the CRISCAR Familienstiftung for their support and belief in the Medical University/Aposcience AG public private partnership. The authors additionally acknowledge the core facilities of the Medical University of Vienna, a member of Vienna Life Science Instruments. Some of the computational results presented were achieved using the Vienna Scientific Cluster (VSC).

Funding

H.J.A. discloses support for the research and publication of this work from the FFG Grant "APOSEC" (852748 and 862068; 2015-2019), the Vienna Business Agency "APOSEC to clinic" (ID 2343727, 2018-2020) and the Aposcience AG. M.M. discloses support for the research and publication of this work from the Federal Ministry Women, Science and Research, Republic of Austria and the Austrian Science Fund (Grant-DOI: 10.55776/PAT8996524).

Author Contributions Statement

Conceptualization, M.D., H.J.A., M.M. and B.M.; methodology, M.D., M.W., M.S., M.L., D.C., K.K., D.B., B.G., H.K., M.T.L., L.M., A.I.S., W.W., C.A. and M.M.; software, M.D., M.W., D.C., and K.K.; validation, M.D. and B.M.; formal analysis, M.D.; investigation, M.D., M.W., M.M and B.M.; resources, H.J.A., M.M.; data curation, M.D., D.B., K.K.; writing original draft preparation, M.D., M.L.; writing, review and editing, M.D., M.W., M.S., M.L., D.C., K.K., D.B., H.K., M.T.L., C.A., M.M., B.M.; visualization, M.D.; supervision, H.J.A., M.M., B.M.; project administration, M.D. and B.M.; funding acquisition, H.J.A., M.M.; equal contribution, H.J.A., M.M., B.M.; All authors checked and agreed to the final version of the manuscript.

Competing Interests Statement

All authors declare no conflict of interest.

Figure legends

Figure 1: Single cell RNA sequencing uncovers cellular composition of healthy and pathologic thymic conditions

(a) Schematic illustration of the study setting (b) Hematoxylin/Eosin (HE) staining of prenatal thymus (prenatal Thym), postnatal thymus (postnatal Thym), thymic hyperplasia (TH), thymoma type-A (TET-A), -AB (TET-AB), -B1 (TET-B1), -B2 (TET-B2), -B3 (TET-B3), -micronodular (TET-MNT), -C/ squamous carcinoma (TET-C). Scale bars: 500 µm, coloured frames symbolize condition affiliation- healthy thymus, TH or TET; One representative micrograph of n = 3 donors per condition is shown. (c) UMAP-Plot revealing common primary cell cluster split in healthy thymus, TH and TET, coloured frames symbolize condition affiliation- healthy thymus, TH or TET, Cluster characterisation as TC: thymocytes; BC: B-cells; PC: plasma cells; TEC: thymic epithelial cells; DC: dendritic cells; MAC/Mono: macrophages/monocytes; EC: endothelial cells; FB: fibroblasts; VSMC/Peri: vascular smooth muscle cells/pericytes; coloured frames symbolize condition affiliation- healthy thymus, TH or TET; (d) Bar plot indicates relative amounts of cell types within TET-A, -AB, -B1, -B2, -B3, -C, -MNT based on bulk RNA-seq deconvolution. , coloured frames symbolize condition affiliation- healthy thymus, TH or TET; (e) Bar plot indicates relative amounts of cell types exclusive for research-group and condition based on scRNAseq. Attached Bar plot shows total number of included cells per entity. Source data are provided within Supplementary data.

Figure 2: Integrated analysis reveals distinct thymic epithelial subtypes associated with TETs

(a) Combined plot including from left to right - UMAP-Plot showing identified TEC subtypes - Bar plot indicating relative amounts of cell types within all conditions – Bar plot displaying total number of detected cells per condition – Heatmap uncovering cluster individuality by DEG; Abbreviations: thymic hyperplasia (TH), thymoma type-A (TET-A), -AB (TET-AB), -B (TET-B), -micronodular (TET-MNT), -C/ squamous carcinoma (TET-C); (b) Dot plots depicting matching cluster annotation applying module scores based on published clustermarker and marker genes; (c) Dot plot highlighting top clusterdefining genes of TET-TEC; Dot color shows level of average gene expression, Dot size reveals percentage of cells expressing the gene; (d) GO-Term analysis of TET-TEC. Bar length depicts statistical significance. Pathway and functional enrichment analyses were performed using Metascape, statistical significance was calculated using a hypergeometric test with Benjamini–Hochberg multiple testing correction. Source data are provided within Supplementary data.

Figure 3: Immunostaining reveals increase of MAOA in distinct TETs

(a) Representative immunofluorescence of postnatal thymus (postnatal Thym), thymic hyperplasia (TH), thymoma type-A (TET-A), -AB (TET-AB), -B (TET-B), -micronodular (TET-MNT), -C/ squamous carcinoma (TET-C) for MAOA. Scale bar: 100 µm. One representative micrograph of n = 3 donors per condition is shown. (b) Combined plot including - Heatmap highlighting top 10 clusterdefining genes of Carc_TEC_1; color shows level of average gene expression, - GO-Term analysis of Carc_TEC_1. Statistical significance is depicted by bar length. (c) Combined plot including - Heatmap highlighting top 10 clusterdefining genes of Carc_TEC_2; color shows level of average gene expression, - GO-Term analysis of Carc_TEC_2. Statistical significance is depicted by bar length. Source data are provided within Supplementary data.

Figure 4: Integrated analysis reveals distinct fibroblast subtypes associated with TETs

(a) Combined plot including from left to right - UMAP-Plot showing identified FB subtypes - Bar plot indicating relative amounts of cell types within all conditions – Bar plot displaying total number of detected cells per condition – Heatmap uncovering cluster individuality by identified DEG expression; Abbreviations: thymic hyperplasia (TH), thymoma type-A (TET-A), -AB (TET-AB), -B (TET-B), -micronodular (TET-MNT), -C/ squamous carcinoma (TET-C); (b) Dot plot highlighting top clusterdefining genes of TET-FB; Dot color shows level of average gene expression, Dot size reveals percentage of cells expressing the gene; (c) GO-Term analysis of TET-FB. Pathway and functional enrichment analyses were performed using Metascape, where statistical significance was calculated using a hypergeometric test with Benjamini–Hochberg multiple testing correction. Statistical significance is depicted by bar length; (d) Heatmap of predicted pathway using PROGENy scores calculated over single-cell RNA-seq data for each FB subtype. (e) Feature plot verifying increased THY1 expression in TET-FB. Color intensity shows average gene expression. (f) Representative immunofluorescence of postnatal thymus (postnatal Thym), thymic hyperplasia (TH), thymoma type-A (TET-A), -AB (TET-AB), -B (TET-B), -micronodular (TET-MNT), -C/ squamous carcinoma (TET-C) for SDC1 and CD90. Scale bar: 100 µm. Arrows and dashed rectangles indicate double-positive cells. Continuous rectangels show magnification of double-positive cells. One representative micrograph of n = 3 donors per condition is shown. Source data are provided within Supplementary data.

Figure 5: Pathology-specific changes in cellular pathway regulation and cell-cell interactions of normal and pathologic thymus

(a) Heatmap of predicted pathway signaling for prenatal thymus, pediatric thymus, adult thymus, thymic hyperplasia (TH), thymoma type-A (TET-A), -AB (TET-AB), -B (TET-B), -micronodular (TET-MNT), -C/ squamous carcinoma (TET-C) using PROGENy scores calculated over single-cell RNA-seq data for each cell type. coloured frames symbolize condition affiliation- healthy thymus, TH or TET; (b) Comparison of TGFB pathway activities detected in all donors by tissue. Lines and error bars indicate mean and standard deviation. Individual counts, for each donor are depicted as dots. Gray background indicates conditions with less than 3 donors (TET-A and TET-MNT), which excluded from the statistical analyses but are shown in the figure for the sake of comparability. Pathway activities identified in n = VSMC: 4 adult thymus, 4 pediatric thymus, 18 prenatal thymus, 4 TH, 4 TET-AB, 7 TET-B, 3 TET-C; FB: 4 adult thymus, 4 pediatric thymus, 18 prenatal thymus, 5 TH, 4 TET-AB, 6 TET-B, 3 TET-C; EC: 4 adult thymus, 4 pediatric thymus, 17 prenatal thymus, 4 TH, 4 TET-AB, 7 TET-B, 3 TET-C - donors, *p* values derived in FB and VSMC from one-way ANOVA tests and Tukey post hoc tests, in EC from Kruskal–Wallis Test and Dunn's post-hoc comparisons. Exact p-values are provided directly in the figure. coloured borderlines symbolize condition affiliation- healthy thymus, TH or TET; (c) Barplot depicting absolute number of detected interactions by each condition. (d) Dot plot revealing high communication probability of APP-CD74 in TET-A. Red squares indicate cell-cell interactions identified in TET_A. Font color denotes thymus condition. Dot color shows communication probability. Cell–cell communication probabilities were inferred using CellChat, and the statistical significance of ligand–receptor interactions was determined using a permutation test. (e) Representative immunofluorescence of postnatal thymus (postnatal Thym), thymic hyperplasia (TH), thymoma type-A (TET-A), -AB (TET-AB), -B (TET-B), -micronodular (TET-MNT), -C/ squamous carcinoma (TET-C) for APP and CD74. Scale bar: 100 µm. One representative micrograph of n = 3 donors per condition is shown. (f) Heatmap highlighting APP and CD74 expression in identified TET-A cellsubtypes; color shows level of average gene expression; Source data are provided within Supplementary data and Source Data file.

Figure 6: Bulk RNAseq analysis reveals potential *GTF2I* mutation-associated transcriptional pattern.

(a) PCA plots based on bulk RNAseq analysis of distinct TET types. (b) Heatmap of significantly up- and down regulated genes comparing *GTF2I*-mutation positive vs. negative TETs by bulk RNAseq. Color represents the Z-score of normalized expression by DESeq2. Final diagnosis and mutation status are shown on top of the heatmap; (c-d) GO-Term analysis of up- and down regulated genes comparing *GTF2I*-mutation positive vs. negative TETs by bulk RNAseq. Pathway and functional enrichment analyses were performed using Metascape, where statistical significance was calculated using a hypergeometric test with Benjamini–Hochberg multiple testing correction. Statistical significance is depicted by bar length; (e) Dot plots showing matching upregulation in TET-A and TET-MNT of module scores based on published GTF2I up- and down regulated genes, Dot color shows level of average gene expression, Dot size reveals percentage of cells expressing the score; Abbreviations: thymic hyperplasia (TH), thymoma type-A (TET-A), -AB (TET-AB), -B (TET-B), -micronodular (TET-MNT), -C/ squamous carcinoma (TET-C); Source data are provided within Supplementary data.

Editor's Summary

The human thymus is essential for building the adaptive immune system, yet its development and disease-related changes remain incompletely understood. Here, the authors show through large-scale single-cell RNA sequencing that distinct cell populations and signaling programs define normal development and multiple thymic diseases.

Peer Review Information: *Nature Communications* thanks Sarah Teichmann, who co-reviewed with Veronika Kedlian, Giuseppe Giaccone and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Figure 1

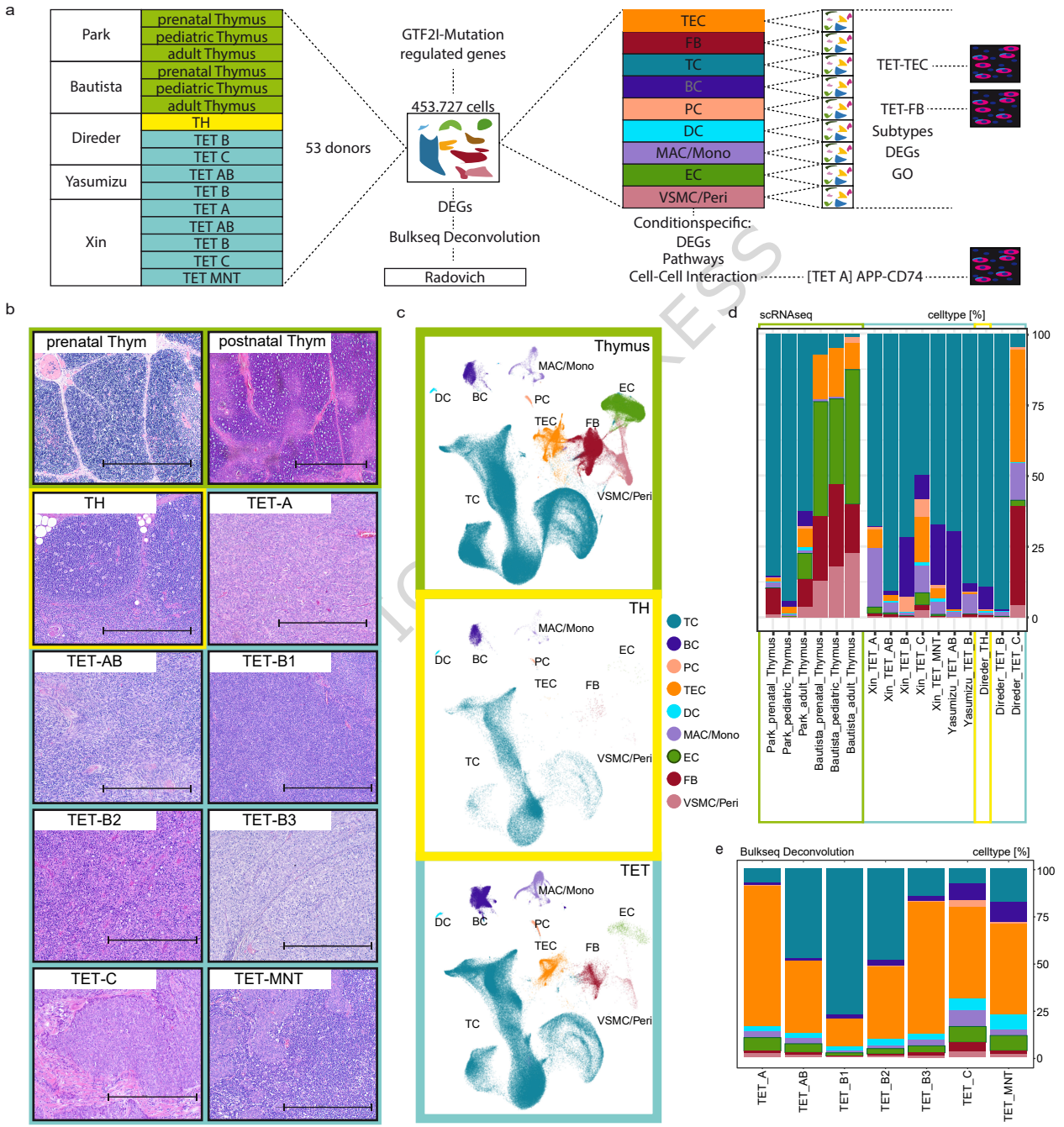


Figure 2

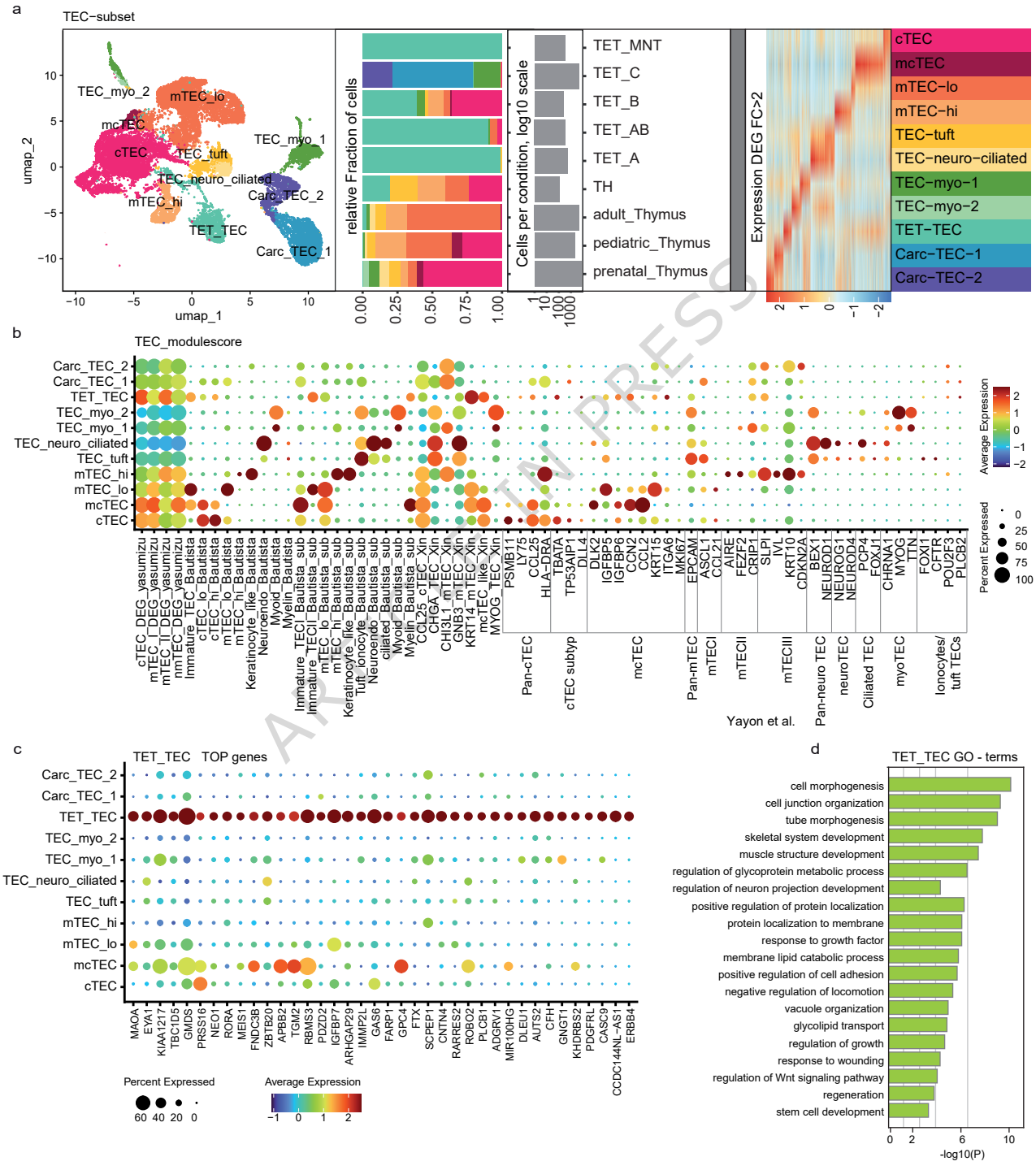


Figure 3

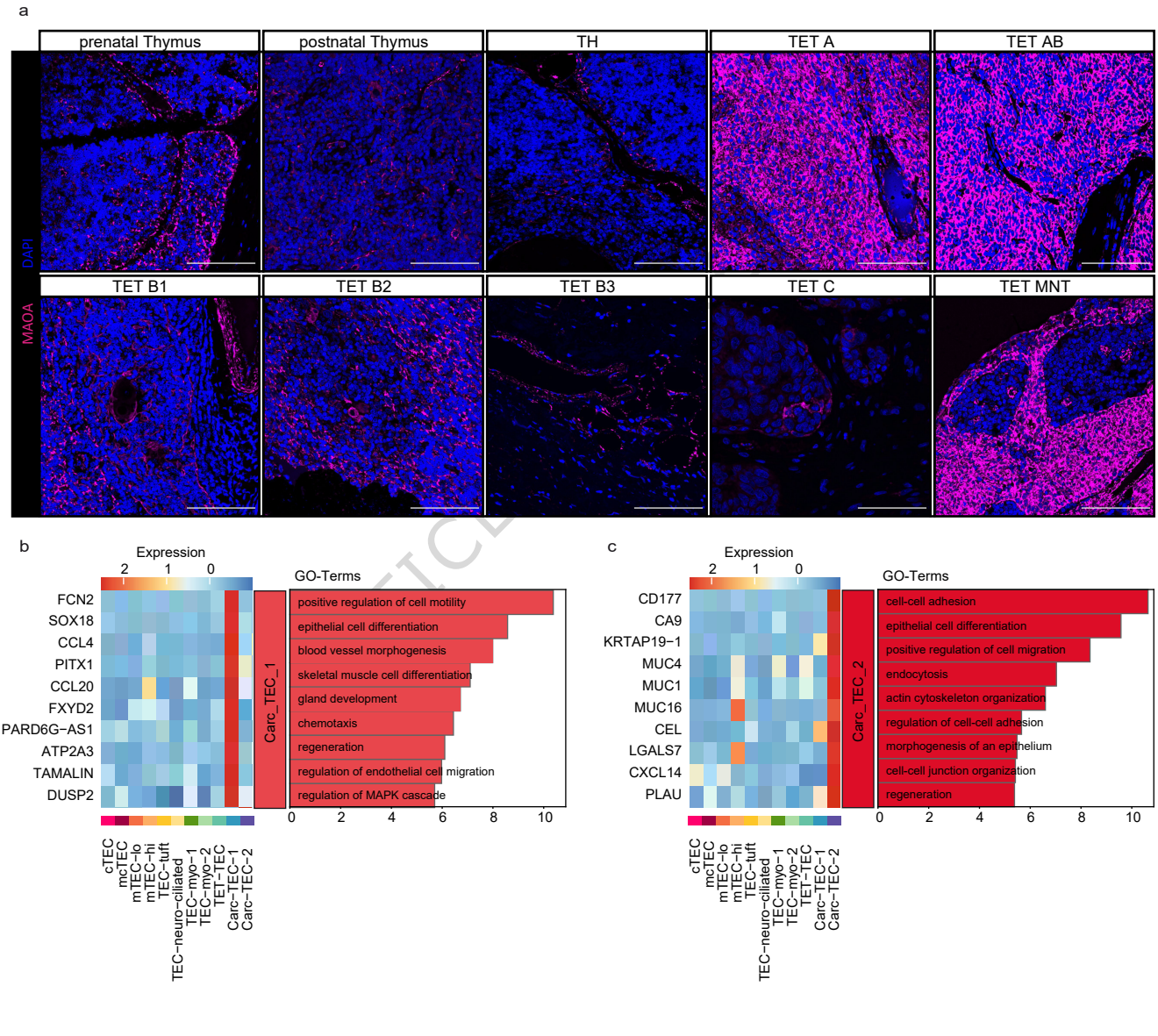

a
prenatal Thymus
postnatal Thymus
TH
TET A
TET AB
DAPI
TET B1
TET B2
TET B3
TET C
TET MNT
MAOA
b
Expression
2
1
0
GO-Terms
FCN2
SOX18
CCL4
PITX1
CCL20
FXYD2
PARD6G−AS1
ATP2A3
TAMALIN
DUSP2
Carc_TEC_1
positive regulation of cell motility
epithelial cell differentiation
blood vessel morphogenesis
skeletal muscle cell differentiation
gland development
chemotaxis
regeneration
regulation of endothelial cell migration
regulation of MAPK cascade
0
2
4
6
8
10
cTEC
mcTEC
mTEC−lo
mTEC−hi
TEC−tuft
TEC−neuro−ciliated
TEC−myo−1
TEC−myo−2
TET−TEC
Carc−TEC−1
Carc−TEC−2
c
Expression
2
1
0
GO-Terms
CD177
CA9
KRTAP19−1
MUC4
MUC1
MUC16
CEL
LGALS7
CXCL14
PLAU
Carc_TEC_2
cell-cell adhesion
epithelial cell differentiation
positive regulation of cell migration
endocytosis
actin cytoskeleton organization
regulation of cell-cell adhesion
morphogenesis of an epithelium
cell-cell junction organization
regeneration
0
2
4
6
8
10
cTEC
mcTEC
mTEC−lo
mTEC−hi
TEC−tuft
TEC−neuro−ciliated
TEC−myo−1
TEC−myo−2
TET−TEC
Carc−TEC−1
Carc−TEC−2

Figure 4

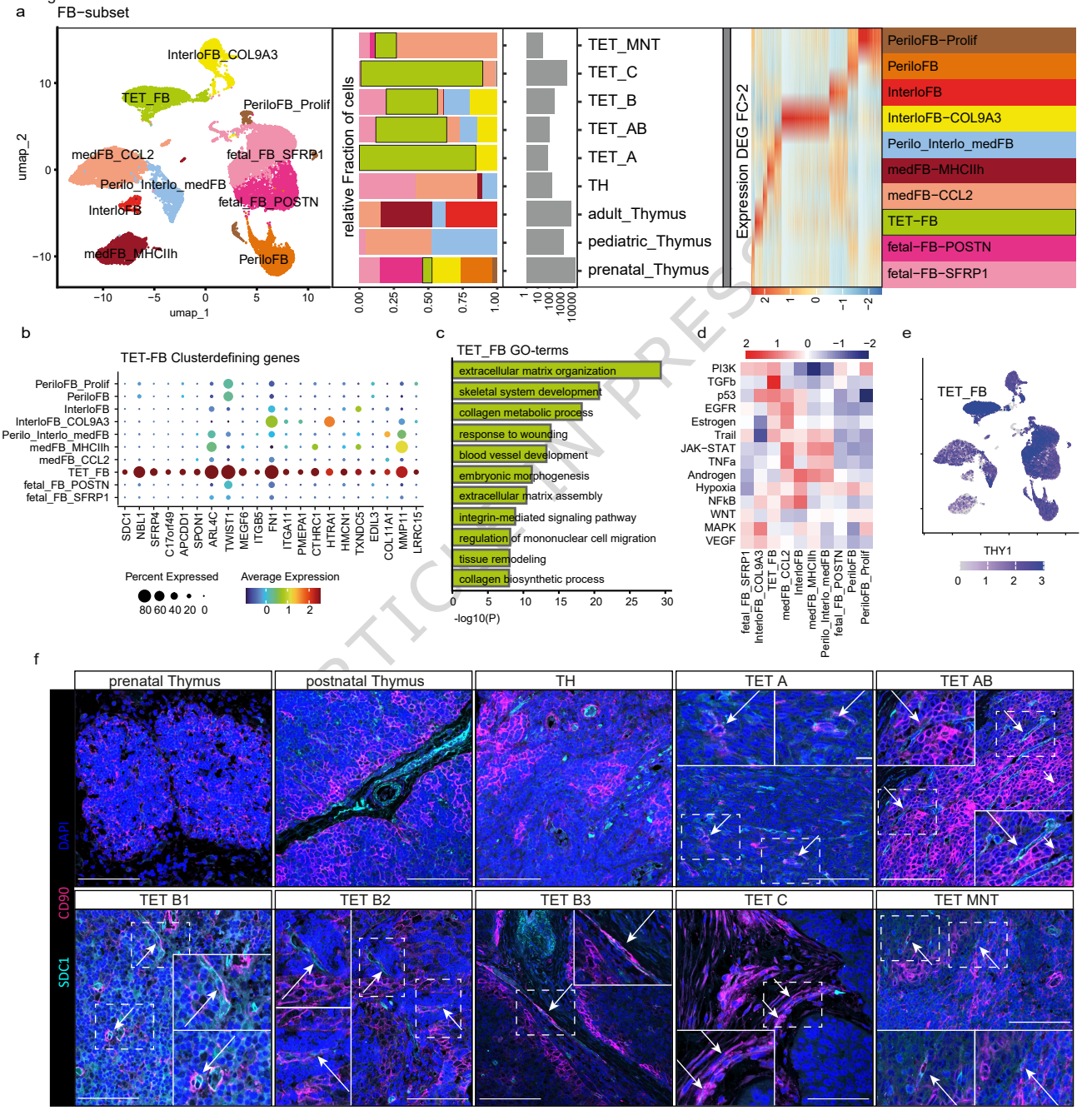


Figure 5

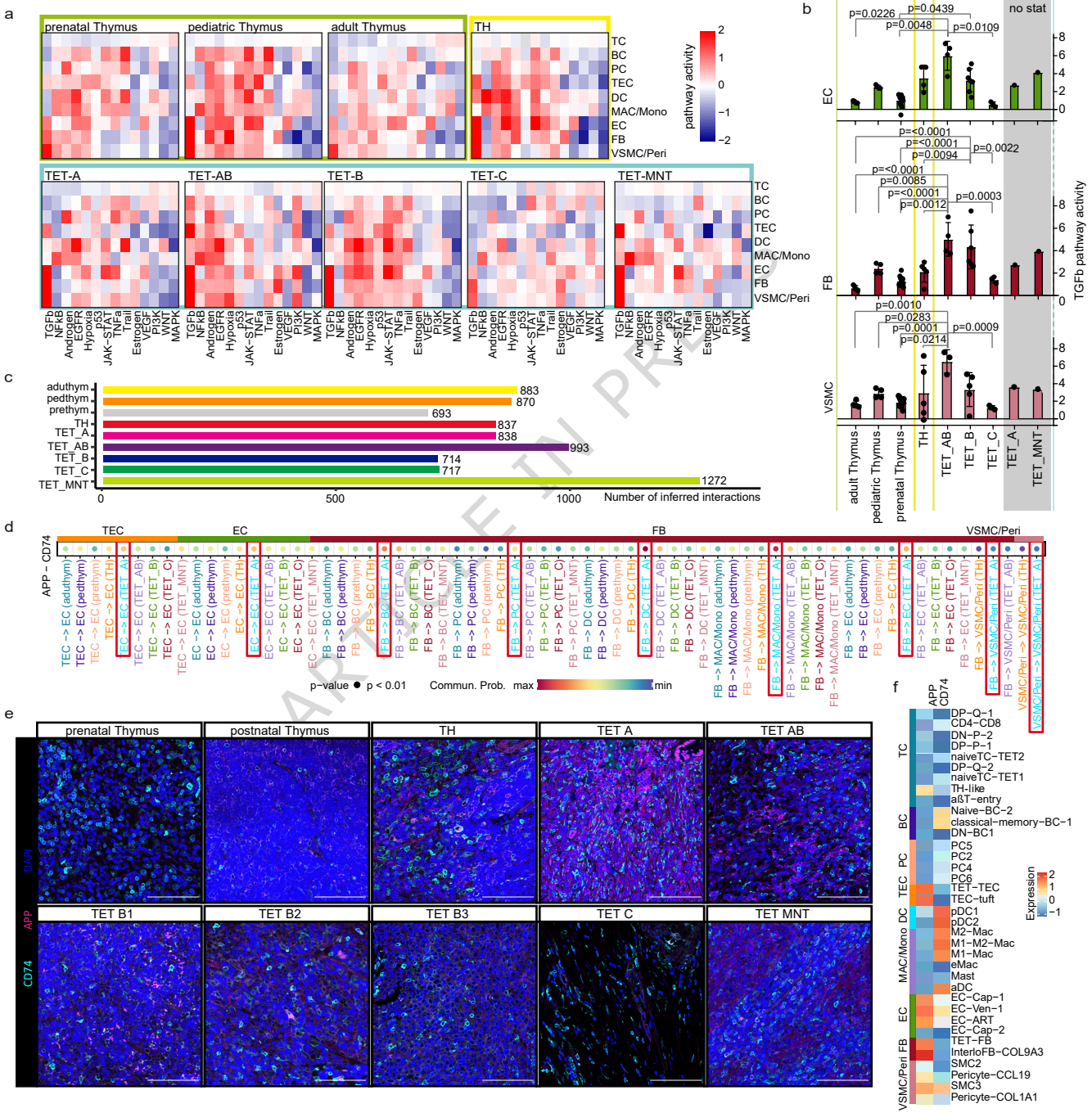
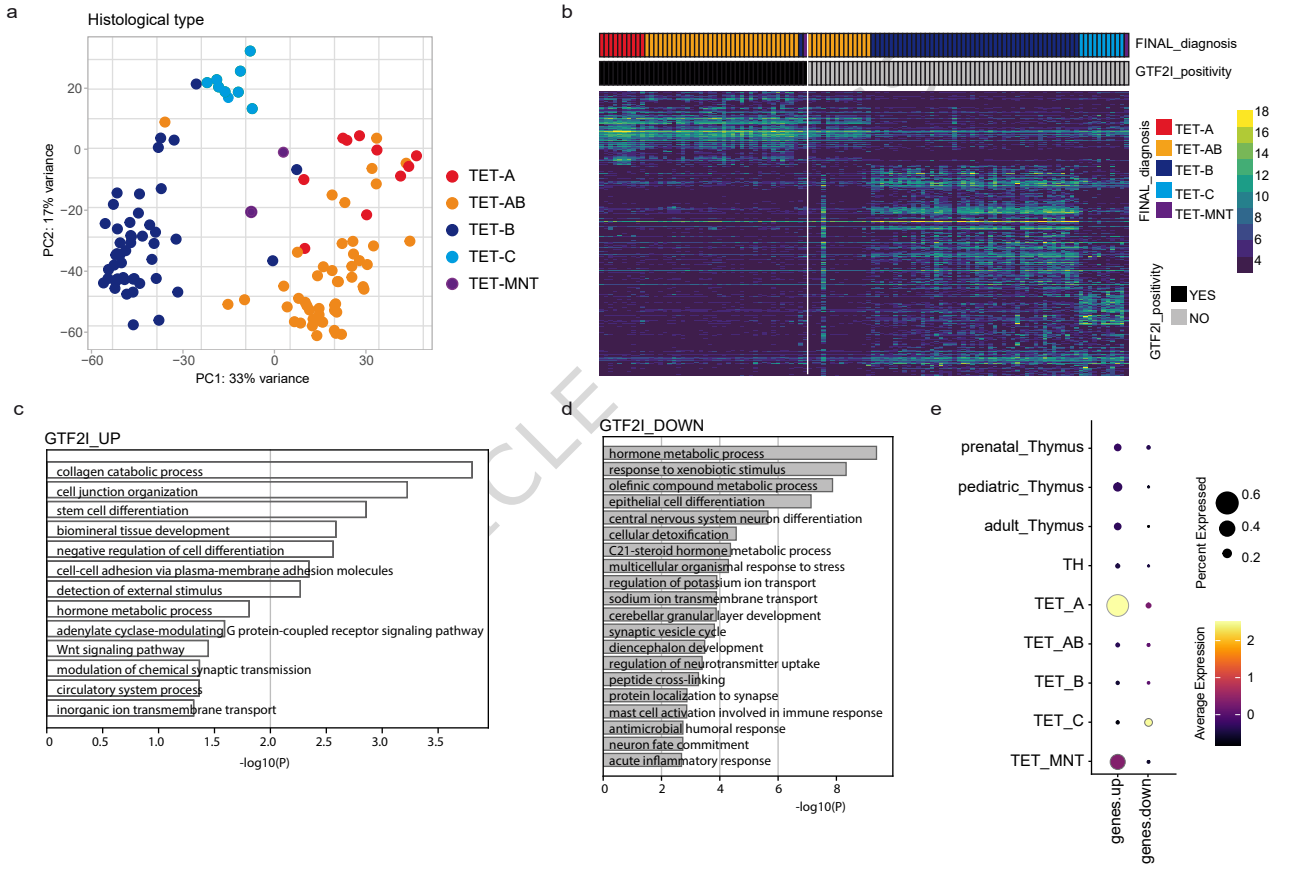

a
prenatal Thymus
pediatric Thymus
adult Thymus
TH
TET-A
TET-AB
TET-B
TET-C
TET-MNT
TC
BC
PC
TEC
DC
MAC/Mono
EC
FB
VSMC/Peri
pathway activity
TGFb NFkB Androgen EGFR Hypoxia p53 JAK-STAT TNFa Trail Estrogen VEGF PI3K WNT MAPK
b
no stat
EC
FB
VSMC
TGFb pathway activity
p=0.0226
p=0.0439
p=0.0048
p=0.0109
p=<0.0001
p=<0.0001
p=0.0094
p=0.0022
p=<0.0001
p=0.0085
p=<0.0001
p=0.0003
p=0.0012
p=0.0010
p=0.0283
p=0.0001
p=0.0009
p=0.0214
adult Thymus
pediatric Thymus
prenatal Thymus
TH
TET_AB
TET_B
TET_C
TET_A
TET_MNT
c
adythym
pedthym
prethym
TH
TET_A
TET_AB
TET_B
TET_C
TET_MNT
883
870
693
837
838
993
714
717
1272
Number of inferred interactions
d
TEC
EC
FB
VSMC/Peri
APP – CD74
p-value ● p < 0.01
Commun. Prob. max min
e
prenatal Thymus
postnatal Thymus
TH
TET A
TET AB
TET B1
TET B2
TET B3
TET C
TET MNT
DAPI
APP
CD74
f
APP
CD74
DP-Q-1
CD4-CD8
DN-P-2
DP-P-1
naiveTC-TET2
DP-Q-2
naiveTC-TET1
TH-like
aβT-entry
Naive-BC-2
classical-memory-BC-1
DN-BC1
PC5
PC2
PC4
PC6
TET-TEC
TEC-tuft
pDC1
pDC2
M2-Mac
M1-M2-Mac
M1-Mac
eMac
Mast
aDC
EC-Cap-1
EC-Ven-1
EC-ART
EC-Cap-2
TET-FB
InterloFB-COL9A3
SMC2
Pericyte-CCL19
SMC3
Pericyte-COL1A1
Expression

Figure 6


a
Histological type
PC2: 17% variance
PC1: 33% variance
TET-A
TET-AB
TET-B
TET-C
TET-MNT
b
FINAL_diagnosis
GTF2I_positivity
TET-A
TET-AB
TET-B
TET-C
TET-MNT
YES
NO
c
GTF2I_UP
collagen catabolic process
cell junction organization
stem cell differentiation
biomineral tissue development
negative regulation of cell differentiation
cell-cell adhesion via plasma-membrane adhesion molecules
detection of external stimulus
hormone metabolic process
adenylate cyclase-modulating G protein-coupled receptor signaling pathway
Wnt signaling pathway
modulation of chemical synaptic transmission
circulatory system process
inorganic ion transmembrane transport
-log10(P)
d
GTF2I_DOWN
hormone metabolic process
response to xenobiotic stimulus
olefinic compound metabolic process
epithelial cell differentiation
central nervous system neuron differentiation
cellular detoxification
C21-steroid hormone metabolic process
multicellular organismal response to stress
regulation of potassium ion transport
sodium ion transmembrane transport
cerebellar granular layer development
synaptic vesicle cycle
diencephalon development
regulation of neurotransmitter uptake
peptide cross-linking
protein localization to synapse
mast cell activation involved in immune response
antimicrobial humoral response
neuron fate commitment
acute inflammatory response
-log10(P)
e
prenatal_Thymus
pediatric_Thymus
adult_Thymus
TH
TET_A
TET_AB
TET_B
TET_C
TET_MNT
genes.up
genes.down
Percent Expressed
Average Expression