

Spatial cartography of human thymus enables the geopositioning of lineage transcription factors in rare mimetic thymic epithelial cells

Received: 25 June 2024

Accepted: 12 January 2026

Published online: 10 March 2026

 Check for updates

Uma S. Kamaraj^{1,14}, Ying Chen^{1,14}, Junjie Lei¹, Pradeep Gautam¹, Pongsatorn Horcharoensuk^{1,2}, Czaryna K. M. Clemente³, Katja G. Weinacht⁴, Nicholas R. J. Gascoigne^{5,6,7,8}, Jinmiao Chen^{9,10}, Ching Kit Chen^{3,11}, Qingfeng Chen^{1,9}, Qi-Jing Li^{1,9}, Lai Guan Ng¹² & Yui-Han Loh^{1,2,8,13} ✉

The thymus is the primary site for T cell maturation. While transcriptional profiling of human thymus has been reported, a high-resolution spatial atlas is needed. Here we use Stereo-seq spatial transcriptomics to generate a spatial atlas of the human fetal (13, 14, 17 or 18 weeks post-conception) and pediatric (7 weeks, and 2, 5 or 6 years old) thymus. The architecture of the thymus comprises regions such as the outer cortex, inner medulla, and septa, and contains multiple cell types, including thymic epithelial cells (TEC), thymocytes, dendritic cells, macrophages, and B cells. Utilising this spatial transcriptomics and proteomics information, we further describe lineage-defining transcription factors (TF) that govern molecular signatures of rare mimetic TEC regulation. Our study thus establishes a high-resolution spatial atlas of the human fetal and pediatric thymus to uncover distinct architectural features and TFs regulating these rare cell types, and serves as a resource for further studies.

The maturation and selection of T cells are mediated by the thymus, which is crucial for the development of adaptive immunity and central tolerance¹. T cell progenitors from the fetal liver or bone marrow interact with different thymic microenvironments, giving rise to distinct T cells². Histologically, the human thymus can be divided into the

cortex, medulla, and septa³. In the cortex, cortical thymic epithelial cells (cTECs) support T lineage commitment and subsequent positive selection. Surviving thymocytes then enter the medulla, where a diverse pool of antigen-presenting cells (APCs) is present, including medullary thymic epithelial cells (mTECs) and dendritic cells (DCs).

¹Institute of Molecular and Cell Biology (IMCB), Agency for Science, Technology and Research (A*STAR), Singapore, Republic of Singapore. ²Department of Physiology, NUS Yong Loo Lin School of Medicine, Singapore, Singapore. ³Department of Pediatrics, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore. ⁴Department of Pediatrics, Division of Stem Cell Transplantation and Regenerative Medicine, Stanford University, Stanford, CA, USA. ⁵Immunology Translational Research Programme, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore.

⁶Department of Microbiology and Immunology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore. ⁷Cancer Translational Research Programme, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore. ⁸NUS Graduate School's Integrative Sciences and Engineering Programme, National University of Singapore, Singapore, Singapore. ⁹Singapore Immunology Network (SigN), Agency for Science, Technology and Research (A*STAR), Singapore, Singapore. ¹⁰Immunology Translational Research Program, Department of Microbiology and Immunology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore. ¹¹Division of Cardiology, Department of Paediatrics, Khoo Teck Puat-National University Children's Medical Institute, National University Health System, Singapore, Singapore. ¹²Shanghai Immune Therapy Institute, School of Medicine, Shanghai Jiao Tong University, Renji Hospital, Shanghai, China. ¹³Department of Biological Sciences, National University of Singapore, Singapore, Singapore. ¹⁴These authors contributed equally: Uma S. Kamaraj, Ying Chen. ✉e-mail: Jonathan_LOH@a-star.edu.sg

Those medullary APCs, which express a wide range of peripheral tissue antigens (PTAs), facilitate negative selection to eliminate overreactive thymocytes. Interestingly, studies have identified TECs expressing extra-thymic lineage transcription factors (TFs) and proposed that these TEC mimetics (i.e. to mimic peripheral counterparts) play an important role in central tolerance via compartmentalised self-antigen expression⁴.

The existence of many different thymic cell types and their complex interactions remains a considerable challenge for understanding T cell development. To date, a fine dissection of various thymic cell types has been performed by different studies^{5–8} using scRNA-seq transcriptome profiling and recently using spatial data^{9,10}. These efforts have generated an overview of thymic cellular composition. Due to the comparatively lower resolution of spatial data, these studies have performed the projection of scRNA-seq data onto the spatial tissues to infer the location of the cell types^{11,12}. However, a high-resolution spatial atlas is needed for the discovery of rare cell types directly on spatial tissue using cell type marker genes to understand the nuances of T cell maturation and specification processes.

In this study, we employed Stereo-seq¹³ spatial transcriptomics to create a high-resolution spatial atlas of the human fetal and paediatric thymi. Stereo-seq is a high-resolution spatial transcriptomics technology that captures genome-wide gene expression directly from intact tissue sections using dense DNA nanoball arrays and in situ RNA capture. Complementing this approach, we use MACSima¹⁴ to profile selected protein markers at cellular resolutions as multiplexed immunofluorescence imaging. Further, we employ Stereo-CITE, which integrates spatial multi-omics profiling of transcriptomics and proteomics from the same tissue. Together, these complementary platforms provide a comprehensive spatial omics framework for studying thymic architecture, cell-type locations, and cellular communication.

Using spatial profiling has enabled us to segregate known histological regions via the direct expression of marker genes of different cell types. Furthermore, we resolved each region into sub-regional niches and examined the cellular composition and active signalling pathways within these niches. Recently, various efforts have been taken to profile the different types of rare TEC mimetics^{15–17}. There is also a concept that mTECs utilise lineage-defining transcription factors (TFs) to promote their differentiation into diverse sub-cell types with distinct molecular signatures, for instance, different stages of muscle mimetic TECs¹⁸. Here, we employ high-resolution spatial profiling to map rare mimetic TEC cell types and uncover the core transcriptional features.

In this study, we generate a high-resolution spatial atlas of the human fetal and paediatric thymi using Stereo-seq spatial profiling. This atlas resolves canonical histological regions into fine-grained cellular niches and reveals the spatial organisation of different known and rare thymic cell types. These findings provide a detailed map of thymic microenvironments and their signalling landscapes. We systematically identify TFs expressed by mimetic TECs using spatial transcriptomics and proteomics, characterising their potential tissue antigen-presenting function. More broadly, our human thymus atlas enables deeper investigation of T-cell maturation dynamics and serves as a resource for future studies on thymic development, central tolerance and immune dysregulation.

Results

High-resolution spatial transcriptomic atlas of human thymus

To build a spatial transcriptomic atlas for the human fetal and paediatric thymi, we collected thymus from the early fetal (F) stages of development, including 13 weeks (w), 14w, 17w and 18w post conception. We also collected paediatric (P) thymic samples comprising the ages of 7w, 2 years (y), 5 y and 6 y old (Fig. 1a and Supplementary Fig. 1a). The thymus comprises two lobes, and each lobe comprises

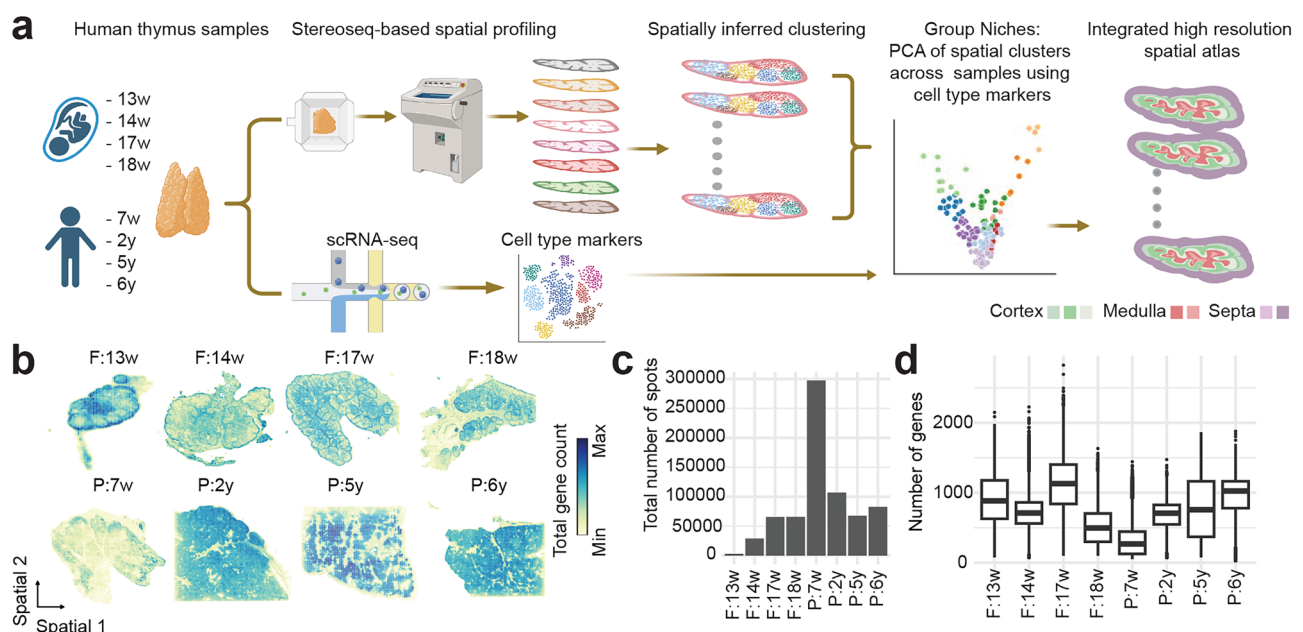


Fig. 1 | High-resolution spatial transcriptomics atlas of human fetal and paediatric thymus. **a** Schematic of the overall workflow for profiling the human thymus in this study. The thymus tissue samples were collected from human fetal (13w, 14w, 17w, 18w) and paediatric (7w, 2 y, 5 y, 6 y) ages. Spatial transcriptome and single-cell profiling of thymus tissue sections were performed. Spatially inferred clusters were determined based on the transcription profile of the spatial spots. The thymus sections across developmental ages were integrated using PCA on the pseudobulk gene expression profile of the spatial clusters, and then K-means

clustering was used to determine the groups. The major thymus domains were defined by the expression of key marker genes. Created in BioRender. Chen, Y. (2026) <https://BioRender.com/wx66ksu>. **b** Spatial visualisation of the number of genes detected with Stereo-seq spatial transcriptomics at bin 50 spot resolution or 25 µm diameter spot across fetal and paediatric tissues. **c** Total number of spots that were sequenced in the thymus tissue section across ages. **d** The number of unique genes detected across the spots in **c** of the tissue section at the developmental ages. F fetal, P paediatric.

sub-structures called lobules³. Histologically, the thymus can be further divided briefly into cortex and medulla (Supplementary Fig. 1b). We spatially and transcriptionally profiled one lobe of the human thymus using Stereo-seq and scRNA-seq techniques (Fig. 1a, b and Supplementary Data 1). We optimised our tissue processing and analytical pipeline using mouse thymus samples from 1- and 3-month-old mice (Supplementary Fig. 1c, d). The Stereo-seq methodology is based on DNA nanoball technology, and in this study, we use bin 50, where the spatial spots are 25 µm in diameter. The number of spots represents the lobe size across the different fetal ages, where 13w has 2868 spots, 14w has 28,800 spots, 17w has 65,438 spots, and 18w has 65,730 spots (Supplementary Data 1 and Fig. 1c). In addition, we also observed that the number of lobule structures increases, indicating the growing complexity of the thymus organ during fetal development. The paediatric thymus samples are usually 5 cm or longer in length (Supplementary Fig. 1a). Therefore, we spatially profiled a portion of the thymic lobe. For paediatric samples ages 2 y, 5 y and 6 y, the 1 × 1 cm chip was used. The number of spatial spots sequenced for 2 y was 106,933, 5 y was 67,617 and 6 y was 82,716 (Fig. 1c). For the paediatric 7w sample, we used a larger chip (2 × 3 cm) to capture the spatial delineation of a larger thymic area with 297,572 spots (Fig. 1b, c). We noticed that using a bigger chip for larger tissues enables better spatial profiling as it facilitates the capture of intact thymus morphology.

Next, we determined the transcriptional depth of the samples and found an average of around 900 genes detected in 13w, 700 in 14w, 1100 in 17w and 500 in 18w across the fetal tissues and around 300 genes detected in 7w, 700 in 2 y, 800 in 5 y and 900 in 6 y in paediatric samples (Fig. 1d). However, across all these samples, the sequencing depth enabled the detection of marker genes for various thymic cell types (Fig. 2a and Supplementary Fig. 1e). Thymocyte markers such as *CD4* and *CD8A* were detected throughout the tissue, as expected. Tissue compartmentalisation was revealed by comparing *PRSSI6* (a cTEC marker) and *AIRE* (a mTEC marker). *PRSSI6* was localised mainly in the outer cortical compartment, while *AIRE* was primarily detected in the inner medullary region. The *CD34* expression pattern was concentrated in the septa regions, which mark the precursors of hematopoietic stem cells and early thymocytes, such as double-negative T cells¹⁹. Besides, different immunological markers, including *MS4A1*, *CD83* and *CD14*, were detected, suggesting the presence of other thymic cell types such as B cells, DCs and macrophages²⁰. Indeed, co-immunostaining of *CD14* and *CD68* identified the presence of macrophages (Supplementary Fig. 1f). Moreover, as revealed by *COL1A2*, thymic fibroblasts displayed a preferential localisation to septa and outlined areas of the thymic lobe, suggesting a distinct compartmentalisation function. Overall, this revealed the capabilities of spatial transcriptomics to capture diverse thymic cell types and their distinct spatial distribution patterns within the thymus.

To comprehensively delineate thymic cellular composition, we performed scRNA-seq of thymic tissue and integrated it with publicly available fetal and paediatric data⁵. This integrated scRNA-seq comprises fetal ages 10w to 18w and paediatric ages 7w to 13 y (Supplementary Fig. 1g–j). The abundant cell types, such as double-positive T cells (DPs), macrophages and fibroblasts, are represented across different studies. In contrast, rare cell types such as mTEC-III are detected in a higher proportion in EPCAM+ sorted scRNA-seq public samples (Supplementary Fig. 1k). Hence, to ensure the representation of most of the cell types in the thymus, we used the integrated scRNA-seq data for further analysis. The different cell types identified in the scRNA-seq data expressing known marker genes^{5,19} are shown in Fig. 2b. These annotated cell types provide a comprehensive reference for resolving spatial thymic composition and organisation.

Thymus architecture and geospatial positioning of cell types

In human thymus, the major anatomic regions (cortex and medulla) can be distinguished via the differential cellular density using H&E

staining (Supplementary Fig. 1b). To define similar regional compartments on our spatial atlas, we performed spatially constrained clustering based on the spatial transcriptome in fetal and paediatric samples (Supplementary Fig. 2a). These spatial clusters take into account the neighbourhood of the spatial spots and the gene expression profile. Next, to integrate samples across different ages, we generated a pseudo-bulk transcriptional profile for each cluster using a curated set of genes representing the major cell types found in the thymus (see Methods). The resulting pseudo-bulk profiles were then batch-corrected across samples, followed by PCA transformation and K-means clustering, which stratified into 10 transcriptionally and spatially distinct groups (Supplementary Fig. 2b, c). The groups that expressed *AIRE* were referred to as the medulla, *PRSSI6* as the cortex, *COL1A2*, *VWF* and *CD34* as septa and the group with the least overall gene expression as low quality (LQ) (Fig. 3a and Supplementary Fig. 2d). The expression of marker genes in the groups across the sample ages indicates the medulla of fetal and paediatric thymic samples systematically separated into different groups as Medulla I and II (Fig. 3b). We found that the LQ region exhibits higher expression of *KRT1*, a marker of mTEC III. This suggests that the LQ region may partially capture medullary regions enriched for mTEC III-like cell states.

We further subdivided the cortex, medulla and septa groups into three concentric niches using the continuous tissue axis method⁹ (Fig. 3c). The proportion of these niches shows an overall more than 45% cortex, 10% medulla, 25% septa and 10% LQ regions (Supplementary Fig. 2e). Since the spatial data is of higher resolution, we used two statistically based methods instead of a single-cell projection-based approach to determine the cellular composition. Firstly, we computed a cell-type-specificity score to determine the cell type present within each spatial spot. The score comprised known cell-type marker genes from Fig. 2b and the top ten differentially expressed genes derived from integrated thymus scRNA-seq data. The cell-type-specificity scores of the cell types within each niche across samples are shown in Fig. 3d. The cortical niches showed a higher enrichment for cortical-associated cell types such as cTECs and DPs (Fig. 3d and Supplementary Fig. 2f). The genes highly expressed in cTECs (*PRSSI6*, and *PSMB11*) and DP (*SMPD3*) show a cortical spatial signature (Fig. 3e, g and Supplementary Fig. 2g). Meanwhile, the medulla niches show higher enrichment for mTECs, dendritic cells (DCs), B cells and single-positive T cells (SPs) (Fig. 3d). mTEC II marked by *AIRE*, DCs by *CD14* and B cells by *MS4A1* shows a medullary spatial signature (Fig. 3f, h and Supplementary Fig. 2h). Further, to validate the cell types in spatial locations, we profiled the protein abundance in thymus tissue using MACSima, and these regions were segregated similarly into niches for comparison (Fig. 3i). In line with the enrichment of cell types by gene expression (Fig. 3d), the proteins show a higher abundance of proliferating T cells in the cortex, immune cells such as B cells and macrophages and TECs in the medulla, and fibroblast and endothelial cells in the septa (Fig. 3j and Supplementary Fig. 2i). Our findings highlight distinct cellular compositions within the thymic niches, further substantiated by both gene expression and protein abundance analyses, demonstrating a spatial organisation that reflects the functional diversity across the cortex, medulla, and septa regions.

Signalling pathways active during positive selection of T cells

Given the dependence of thymocytes on interactions with thymic APCs for maturation, we next sought to decipher the signalling pathways in the microenvironment within different thymic niches. The high-resolution thymus atlas enables the direct detection and comparison of signalling pathways across the developmental stages (Fig. 4a and Supplementary Fig. 3a). Across the niches, prominent temporal changes were seen in pathways such as MK and IGF in cortex niche I in fetal samples. This might be due to the profiling of the whole thymus lobe in

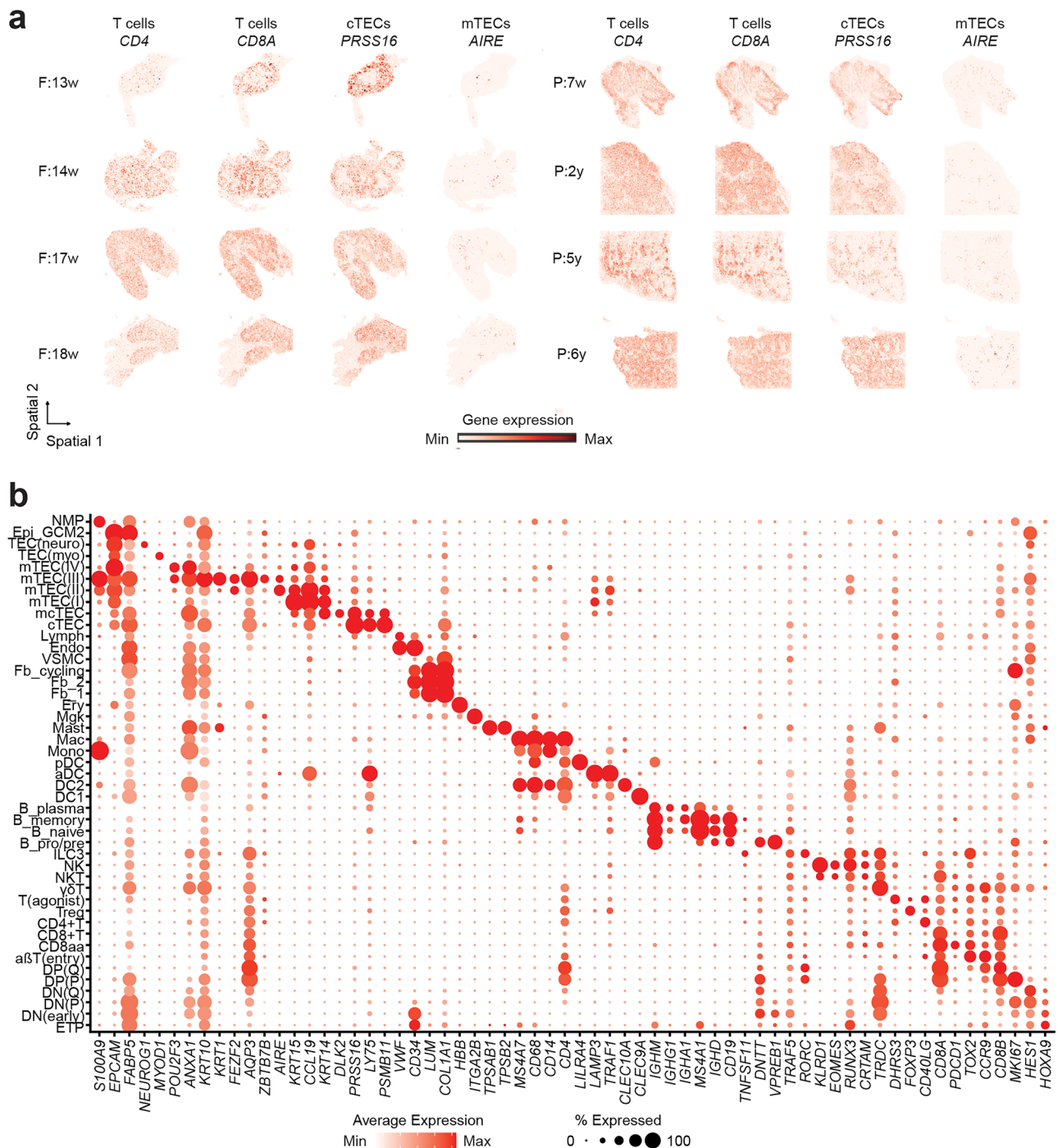


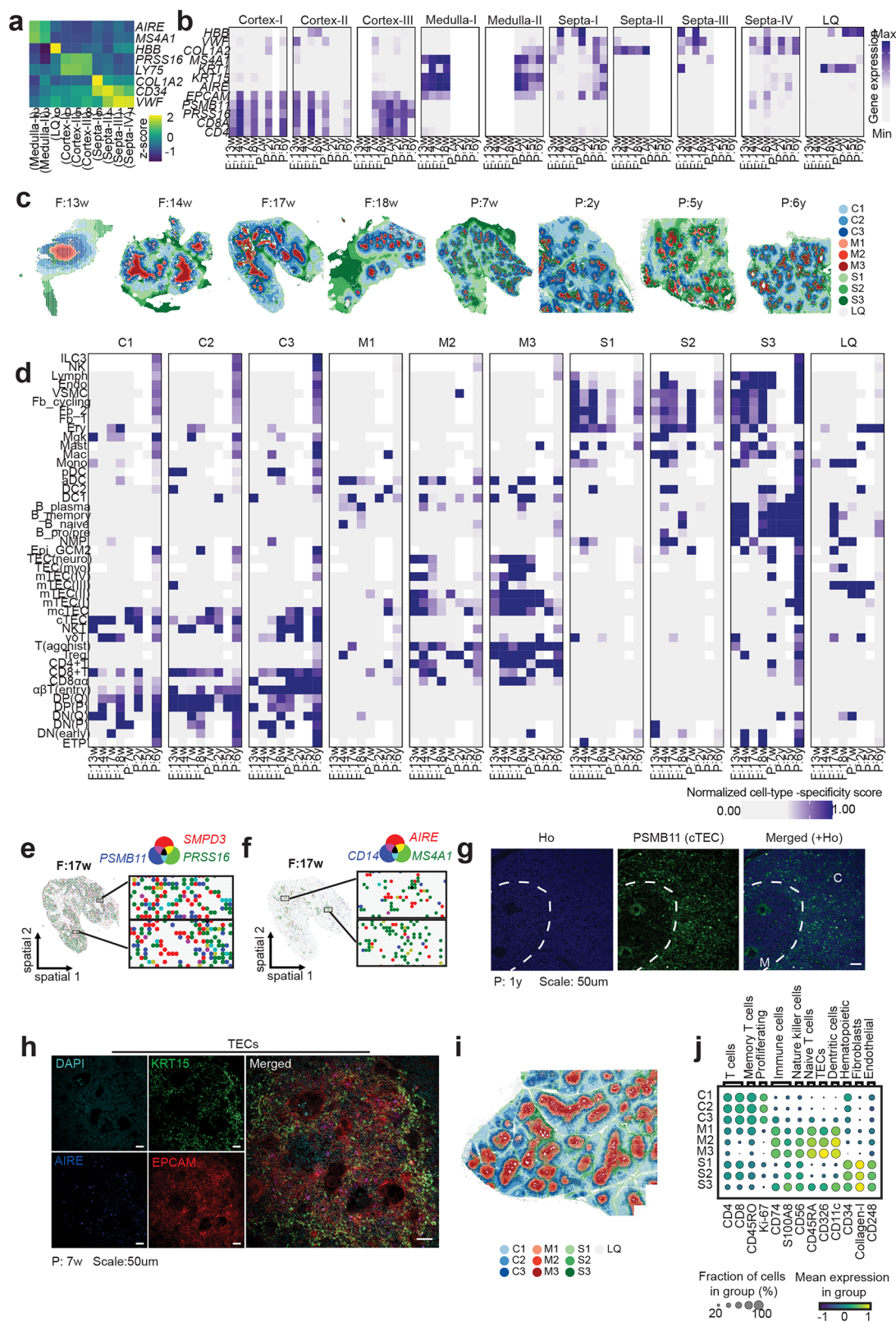
Fig. 2 | Expression profile of marker genes in the spatial and single-cell transcriptomics atlas. a The spatial gene expression of known markers for thymocytes (*CD4* and *CD8A*) and TECs (*PRSS16* and *AIRE*). **b** Dotplot shows the

expression of marker genes for all the cell types present in the thymus based on scRNA-seq data. F fetal, P paediatric.

fetal samples, which contains the capsule intact, compared to the paediatric samples, where a small portion of the lobe was sequenced. Spatial expression of *MDK* (ligand) and *LRP1* (receptor) for the MK pathway revealed a pattern that closely resembled that of *COL1A1* (Fig. 4c and Supplementary Fig. 3c). *MDK* is expressed by cTECs and fibroblasts, whereas *LRP1* is expressed by fibroblasts (Fig. 4b). Consistently, previous studies have shown the positive effects of MDK in stimulating fibroblast growth²¹, further implying the role of the MDK pathway in supporting cortical fibroblast maintenance. Similarly, the IGF pathway involved fibroblasts and DN(Q) (Fig. 4b and

Supplementary Fig. 3b). Together, these findings highlight niche-specific signalling interactions that support fibroblast function.

The major signalling pathway enriched in cortical regions is CXCL (Fig. 4a). The expression of *CXCL12* (ligand) and *CXCR4* (receptor) in the CXCL pathway was more prevalent in cortical regions (Fig. 4d, e and Supplementary Fig. 3d, e). While scRNA-seq showed expression of *CXCL12* by some fibroblasts, co-immunostaining of fibroblastic marker Col-I and CXCL12 also revealed their co-localisation or at least close proximity. This suggests that thymic fibroblasts might be one of the sources for CXCL12. (Fig. 4b, f and Supplementary Fig. 3f). In



consensus with recent studies^{22–24}, *CXCL12* demonstrated a higher enrichment in TEC populations (Fig. 4b), whereas *CXCR4* was detected preferentially in thymocytes (Fig. 4b). These observations were further supported by co-immunostaining of respective TEC and T cell markers with *CXCL12* or *CXCR4* (Fig. 4g, h and Supplementary Fig. 3g, h), consistent with the role of *CXCL12* in thymic migration stimulation²⁵ (Fig. 4d and Supplementary Fig. 3d). Besides, *IL16* (ligand) and *CD4*

(receptor) from the *IL16* pathway show enrichment in cortex and medulla (Fig. 4a and Supplementary Fig. 3b), aligned with the reported roles of the *IL16* pathway in T cell development^{26–28}. This elucidates the intricate signalling pathways within thymic niches, highlighting the critical role of signalling pathways in mediating thymocyte maturation and migration, thereby underscoring the importance of the thymic microenvironment in T cell development.

Fig. 3 | Geopositioning of cell types in thymic niches across age. **a** The heatmap shows the normalised gene expression of key markers to define the thymus domains from the integration of spatially aware transcriptome-based groups combining all ages. Domains were defined as medulla (*AIRE* and *MS4A1*), cortex (*PRSS16* and *LY75*) and septa (*COL1A2*, *CD34* and *VWF*). The colour indicates the normalised gene expression as a z-score. **b** The heatmap shows the gene expression of marker genes across thymus age within each integrated group identified. **c** The spatial domains are concentrically divided into three niches from the outer to inner axis using the TissueTag method. **d** The heatmap illustrates the normalised cell type enrichment based on the cell-type-specificity score. This score is derived from the expression of cell-specific genes within spatial spots, which were identified through differential scRNA-seq analysis. **e** The co-localisation of DP T cells (*SMPD3*) with cTECs (*PRSS16* and *PSMB11*) in the cortex is shown using marker genes in red, blue and green, and the spots with overlapping expression of genes are shown in the

respective overlapping colours. **f** The co-localisation of mTECs (*AIRE*), macrophages (*CD14*) and B cells (*MS4A1*) in the medulla is shown using marker genes in red, blue and green, and the spots with overlapping expression of genes are shown in the respective overlapping colours. **g** Immunostaining of paediatric thymic slide (1y) with PSMB11 (cTEC marker) revealed the cortical compartment. C cortex, M medulla. Scale bar: 50 μ m. A representative region is shown from three independent regions. **h** Immunostaining of different TEC markers (*AIRE*, *KRT15* and *EPCAM*) in the paediatric thymic slide (7w) revealed localisation of different mTEC subtypes (mTEC I: *KRT15*+*EPCAM*+; mTEC II: *AIRE*+*EPCAM*+) in the medulla. Scale bar: 50 μ m. A representative region is shown from four independent regions. **i** MACSima spatial profiling of 22 proteins was used to define the spatial domains, and these domains were concentrically divided into three niches from outer to inner axis in P:7w thymus. **j** The dot plot shows the cell types based on the abundance of proteins in the niches of the MACSima spatial profiled sample. F fetal, P paediatric.

Signalling pathways active during negative selection of T cells

Similarly, we determined the major signalling pathways within the medulla niches across the samples from different developmental ages and found enrichment for pathways such as CCL, MIF, IL16, GALECTIN and TGFB (Fig. 5a). The primary function of the medulla has been the establishment of self-tolerance, which requires multiple components for efficient antigen presentation. For instance, *CD74* is involved in the MIF pathway and is a component of the MHC-II antigen presentation process²⁹. Consistently, it shows an expression pattern specific to medullary regions (Fig. 5c and Supplementary Fig. 3b, i). *MIF* is expressed in mTECs, and *CD44* is expressed by SPs, DCs and B cells (Fig. 5b). The spatial expression map of *CD74*, *CD44* and *MS4A1* (B cell) showed spots with both co-expression of *CD74* and *MS4A1* and co-expression of *CD44* and *MS4A1*, supporting the B cell's antigen presentation function in the medulla. This suggests a central role of MIF signalling in mediating antigen presentation within the medulla.

The CCL signalling pathway is known to regulate the migration of selected thymocytes³⁰, which could help establish medullary self-tolerance. One of the receptors on thymocytes mediating this process is *CCR7*, which interacts with *CCL19* and *CCL21*^{31,32}. Notably, the expression patterns of *CCL19* and *CCR7* were highly restricted to medullary regions (Fig. 5d, e and Supplementary Fig. 3j, k). Besides, *CCL19* was more enriched in mTEC populations (Fig. 5f and Supplementary Fig. 3l), while *CCR7* was comparatively found in SPs, B cells and activated DCs (aDCs) (Fig. 5b, g and Supplementary Fig. 3m). The co-localisation of *CCL19* and *CCR7* with another medullary APC marker *IL3RA* for DCs revealed multiple spots with co-expression among them (Fig. 5d and Supplementary Fig. 3j). In line with the negative selection of SP T cells leading to apoptosis, the medulla is enriched for the Galectin pathway, which is involved in cell death signalling. Additionally, galectins are also known to affect T cell activation and differentiation via APC modulation³³. In summary, the analysis of the medulla niches reveals significant enrichment of key signalling pathways such as CCL, MIF, IL16, GALECTIN, and TGFB, which are crucial for promoting self-tolerance and efficient antigen presentation.

TFs regulating the cell states of mimetic TECs

The mimetic TECs refer to a group of TECs with distinct expressions of TFs that are usually from non-thymic lineages⁴. A recent study implied that these unconventional mTEC subsets could regulate diverse processes and functionally mimic endocrine, microfold or myocyte cell types¹⁵. The mTECs have the potential to differentiate into various molecularly distinct cells, including different stages of muscle differentiation¹⁸. However, a comprehensive spatial understanding of these cell states and the TFs that regulate mimetic TECs is still lacking. To systematically identify the TFs implicated in mimetic TECs, we utilised our high-resolution spatial atlas and scRNA-seq data (Fig. 6a). Since mTECs are epithelial cells located in the medulla, we focused on spatial spots of bin 50 or 25 μ m resolution that expressed *EPCAM*, a

well-known epithelial marker. We then screened for all known human TFs that were spatially present within *EPCAM*+ spots and expressed in *EPCAM*-sorted TECs within the scRNA-seq datasets^{5,17}. To determine the TFs responsible for the molecular variation and cell states of mimetic TECs, we employed a majority voting method to assign the TFs identified in the *EPCAM* spots to the cortex, medulla, or septa (Fig. 6b and Supplementary Fig. 4a–e). It is important to note that mimetic TECs are relatively rare and were detected in 2674 spots across all samples, representing 0.4% of tissue sections sequenced.

A previous study estimated the location of a few known mimetic TECs, such as mTEC III keratinocytes, myoTEC, neuroTEC, ciliated TECs and tuft cells to be in the medulla and near Hassall's corpuscles (HCs) based on scRNA-seq cell type projection onto spatial data⁹. However, the lower spatial resolution and the limited area of tissue sequenced hindered the unbiased identification of mimeTFs (Supplementary Fig. 4f–i). For instance, in a previous study, though the myoTECs are projected onto the spatial data, the key TF *MYOG* in myoTECs is not detected in conjunction with *EPCAM* due to lower spatial resolution (Supplementary Fig. 4i). In this study, we aim to address this limitation by utilising our high-resolution thymus atlas to identify TFs that are involved in mimetic TEC regulation systematically. We identified 70 mimeTFs specific to the medullary region, and the percentage of *EPCAM*+ spots expressing these mimeTFs across the samples displayed rare occurrences (Fig. 6b, Supplementary Fig. 4e and Supplementary Data 3). Similarly, we also identified TFs specific to the cortex and septa; however, these TFs showed relatively less abundance compared to the medulla (Supplementary Fig. 4a, b). Notably, the mimeTFs did not consistently exhibit differences based on fetal or paediatric origins. Key mTEC markers, such as *AIRE* and *FEZF2*, known for their role in regulating mimetic TECs, were detected^{34,35}. We catalogued the known TEC mimetics based on the expression of lineage-specific TFs and genes found in previous studies^{15–17} (Supplementary Data 4). We uncovered several well-studied TFs in mimetic TECs, including *ASCL1*, *INSM1*, *NEUROD1*, *MYOG*, *POU4F1*, *SOX2*, *SPIB* and *JUN* (Supplementary Fig. 4j). Additionally, our approach revealed 62 TFs that may influence mimetic TEC cell states, suggesting extensive regulatory complexity within these rare cell populations.

To validate the detection capabilities for mimeTFs using our approach on spatial transcriptomics, we conducted a Stereo-CITE spatial profiling to capture both the transcriptome and proteome (including *EPCAM*) simultaneously. We further analysed the data at a resolution of 10 μ m or bin 20 (instead of 25 μ m), which better aligns with the diameter of epithelial cells. We identified the cortex, medulla, and septa niches for the CITE-spatial sample, and these regions are comparable to those found in other transcriptionally sequenced samples (Fig. 6c). The spatial plot indicates that *EPCAM* gene expression is sparse, whereas the corresponding protein profile reveals a concentration of *EPCAM* protein (CD326) in the medulla region as expected (Fig. 6d). To determine whether the same mimeTFs are

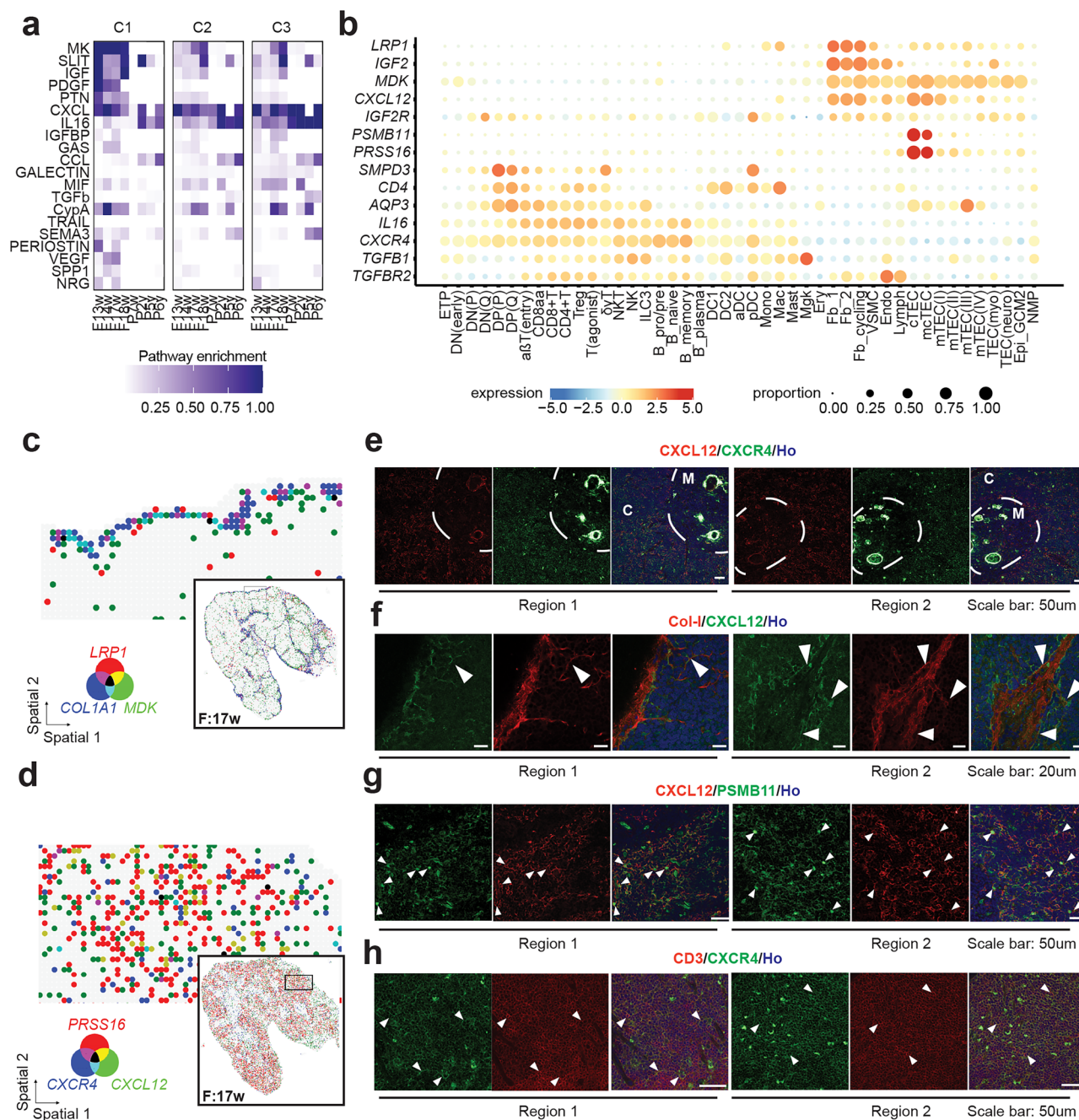


Fig. 4 | Signalling pathways active during positive selection of T cells. a The heatmap shows the normalised enrichment of the signalling pathways that are active within the cortical niches across the thymus ages. **b** The dot plot shows the gene expression of selected ligands and receptors in the cell types found in scRNA-seq data. **c** Spatial co-localisation of the *LRP1* and *MDK* and the cell type involved, like fibroblasts, is marked by *COL1A1* in the fetal 17w thymus. **d** The spatial co-localisation of *CXCL12*, *CXCR4*, and the cell type involved, like cTECs, is marked by *PRSS16*. **e** Immunostaining of human thymic slides (P:1y) revealed cortical enrichment of *CXCL12/CXCR4* pathway. Co-staining of *CXCL12* and *CXCR4* showed preferential expression in cortex rather than in medulla. Two representative regions are shown from four independent regions. **f** Immunostaining of human thymic

slides (P:7w) revealed co-expression of *Col-I* (fibroblast marker) and *CXCL12*, especially at the capsule layer, as indicated by white arrows, thus supporting cortical fibroblasts as another source for *CXCL12* in cortex. Two representative regions were shown from three independent regions. **g** Immunostaining of human thymic slides (P:1yr) showed co-expression between *PSMB11* (cTEC marker) and *CXCL12*, as indicated by white arrows, thus supporting cTECs as a source for *CXCL12* in cortex. Two representative regions are shown from three independent regions. **h** Immunostaining of human thymic slides (P:1 yr) of T cell marker *CD3* with *CXCR4* showed their co-expression, supporting T cells as a source for *CXCR4* in cortex. Two representative regions are shown from three independent regions. F fetal, P paediatric.

observed at 10 μ m resolution, we focused on the abundant EPCAM protein spots in the medulla and almost all the predicted mimeTFs were detected (Fig. 6b). Further co-immunostaining of mimeTFs with the pan-TEC marker EPCAM shows that TECs express these mimeTFs. Previously reported mimeTFs involved in microfoldTEC, myoTEC and

neuroTEC, such as SPIB, MYOG and SOX2, can be detected (Fig. 7a). Notably, other mimeTFs identified in our analysis, including BHLHE40, DLX5, EGR3, and NFKB2 are expressed along with the TEC marker (Fig. 7b and Supplementary Fig. 5a), suggesting their involvement in mimetic TEC cell states. Overall, this study demonstrates a robust

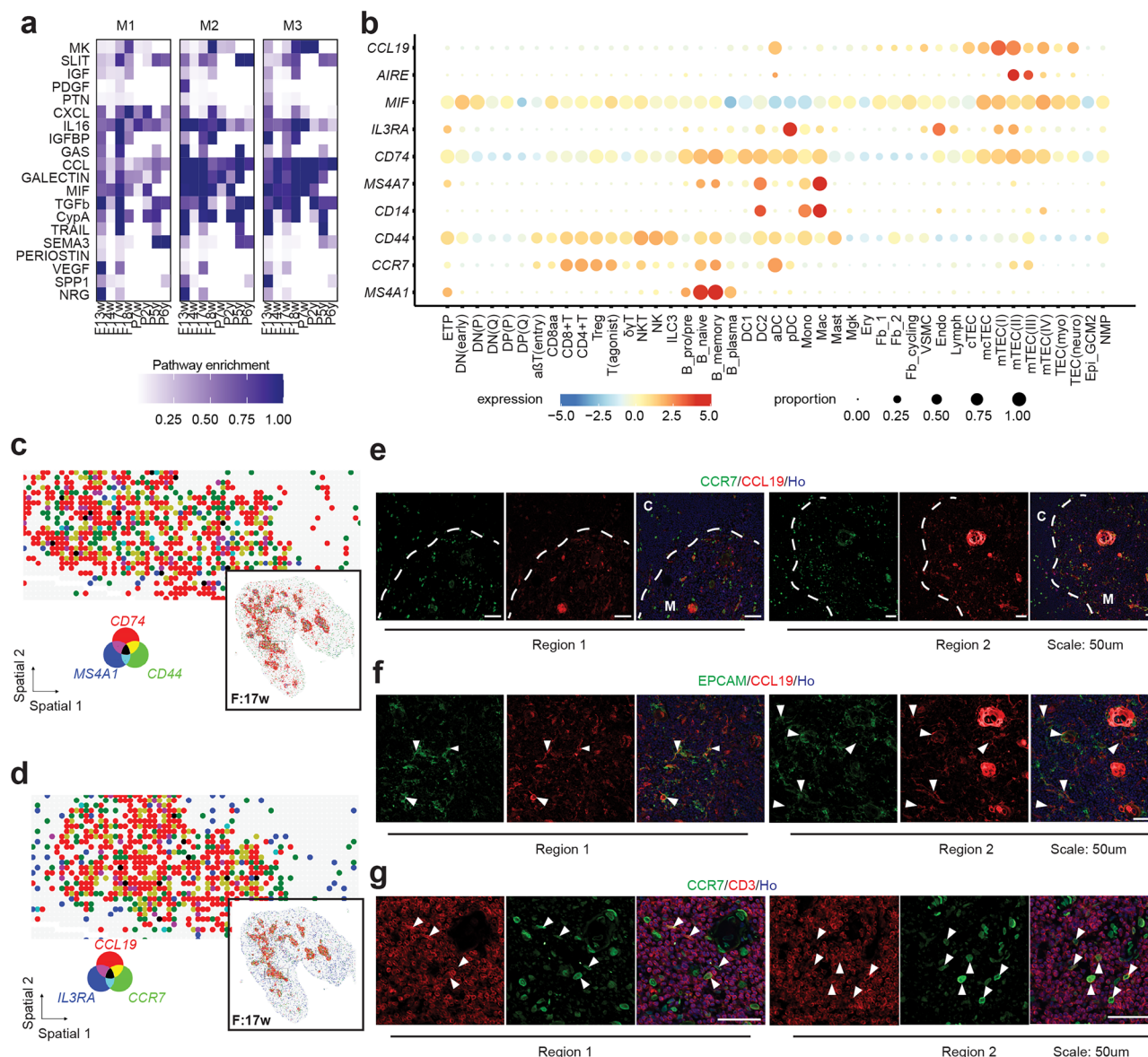


Fig. 5 | Signalling pathways active during negative selection of T cells. a The heatmap shows the normalised enrichment of the signalling pathways that are active within the medullary niches across the thymus ages. **b** The dot plot shows the gene expression of selected ligands and receptors in the cell types found in scRNA-seq data. **c** Spatial co-localisation of the *CD74*, *CD44*, and the cell type involved, like B cell, is marked by *MS4A1*. **d** Spatial visualisation of the significantly expressed ligand (*CCL19*) and receptor (*CCR7*) of the CCL pathway and *IL3RA* (aDC). **e** Immunostaining of human thymic slides (P: 2 m) revealed medullary enrichment of CCR7/CCL19 pathway. Two representative regions are shown from five

independent regions. **f** Immunostaining of human thymic slides (P: 2 m) showed co-expression between EPCAM (TEC marker) and CCL19 in medulla, as indicated by white arrows, thus supporting mTECs as a source for CCL19 in cortex. Two representative regions are shown from four independent regions. **g** Immunostaining of human thymic slides (P: 2 m) revealed co-expression of CD3 (T cell marker) and CCR7, as indicated by white arrows, thus supporting T cells as a source for CCR7 in medulla. Two representative regions are shown from three independent regions. F fetal, P paediatric.

approach for resolving mimetic TECs spatially and to dissect their regulatory programmes.

Abundance and geopositioning of mimeTFs expressing cells

The mimetic TECs are rare cell types in the thymus, as indicated by the low percentage of spots expressing these mimeTFs (Fig. 6b). These TFs are generally also expressed at low levels, and spatial profiling is more susceptible to missing low-expressing genes. Thus, we analysed the number of EPCAM-sorted TEC cells that expressed these TFs in scRNA-seq data¹⁷ (Fig. 8a). TFs such as *JUN*, *HES6*, *MEIS2*, *TSC22D1* and *STAT3* are present in more than 80% of the cells and could be pan-mTEC TFs. Some known mimetic TFs such as *INSML1*, *NEUROD1*, *POU4F1* and *SOX2* are expressed in more than 50% of the cells, whereas TFs such as *AIRE*,

FEZF2, *MYOG*, and *SPIB* are rare. Likewise, the novel mimeTFs *BHLHE40*, *DLX5*, *EGR3* and *NFKB2* are rare. Overall, these findings reinforce the rarity of lineage-defining mimeTFs and illustrate the technical and biological challenges of capturing such low-frequency TEC states.

One limitation of our thymus spatial atlas is that the HC region cannot be precisely pinpointed, as these areas are typically defined by H&E staining. We observed that some of the HC-like structures present within the medulla were assigned to LQ regions, mainly because very few genes are expressed within the HC (Fig. 3c). Additionally, the LQ region exhibited a high expression of *KRT1*, which is a mTEC-III marker (Fig. 3b, d). A previous study suggests that mimetic TECs and mTEC-III are closer to HCs⁹. Therefore, we focused on the *KRT1* spots in our

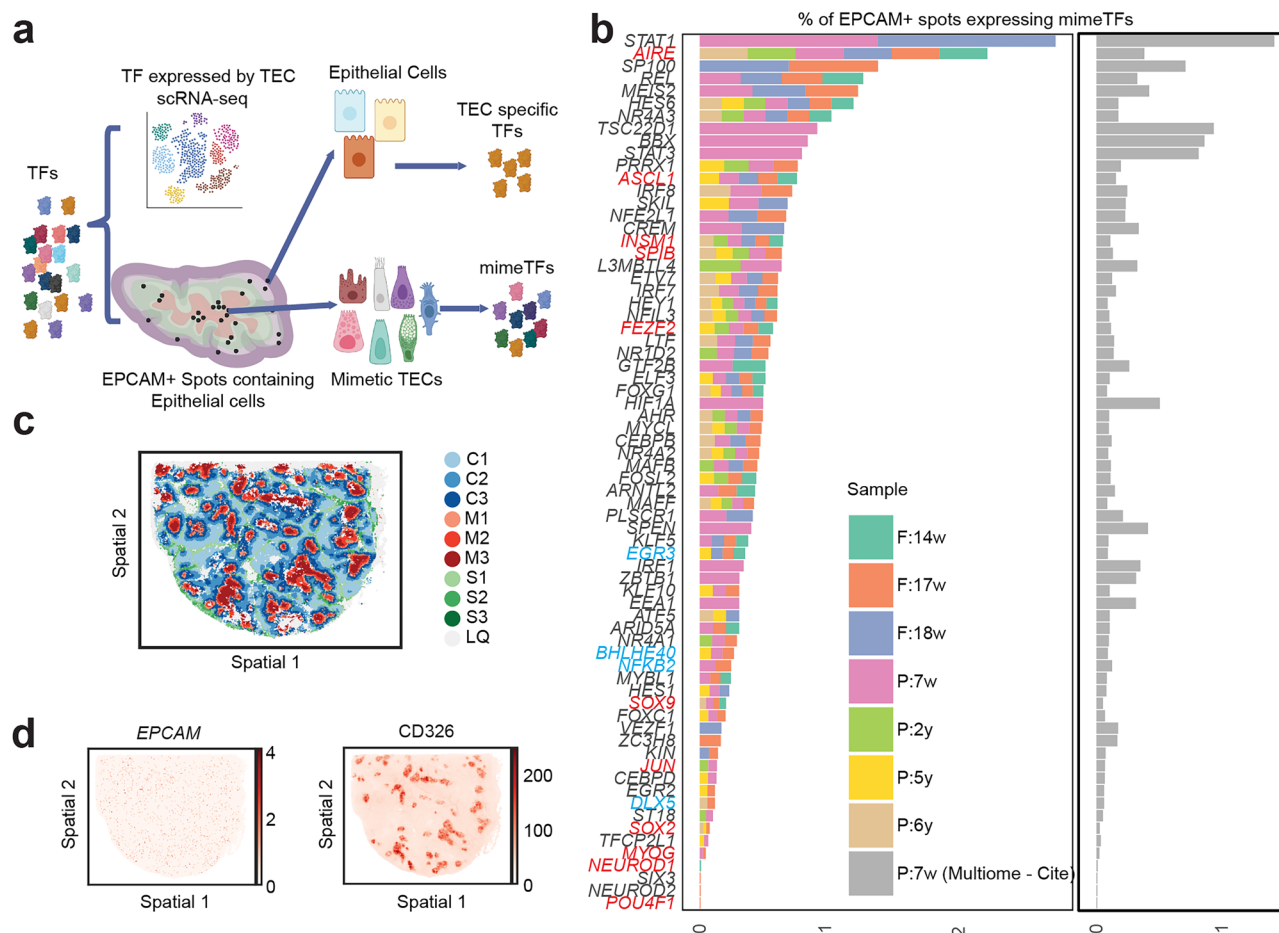


Fig. 6 | Identification of the TFs implicated in rare mimetic TECs. **a** The schematic illustration depicts the identification of mimetic TFs (mimeTFs) expressed by mimetic TECs. The TEC spots on the spatial section at a bin 50 (~25 μm) resolution were determined using the expression of *EPCAM*. The TFs expressed within the spatial TEC spots were determined and then screened for the respective TF's expression in scRNA-seq data of TECs. A majority voting strategy was used to stratify the identified TFs into the domains, and the TFs associated with the medulla were called mimeTFs. Created in BioRender. Chen, Y. (2026) <https://BioRender.com/4jzktkg>. **b** The stacked bar chart shows the percentage of spots expressing mimeTFs compared to all EPCAM spots within each sample. The colours in the bar

chart indicate the sample. The grey colour represents the TFs identified from Stereo-CITE at bin 20, whereas the rest of the colours represent the samples that detected TF using Stereo-seq at bin 50. The labels highlighted in red are the TFs known in mTECs and mimetic TECs, and the TFs highlighted in blue are the mimeTFs found and verified in this study. **c** The niches are defined for the spatial multi-omics (Stereo-CITE) profiling of P:7w thymus, which included the information on abundance of 31 proteins coupled with whole transcriptomics at bin 20 (~10 μm) resolution. **d** The spatial distribution of *EPCAM* gene expression and the corresponding protein (CD326) abundance of the gene *EPCAM* in spatial multi-omics. F fetal, P paediatric.

samples and found an enrichment of *KRT1*-high cells near the HC, along with some cells forming a boundary around it (Fig. 8b and Supplementary Fig. 5b). Next, we calculated the relative percentage of spots in each niche that expressed mimeTFs within the *KRT1*-high region. When comparing the outer medulla to the innermost medulla (M3), we found that M3 had the highest proportion of spots expressing mimeTFs (Fig. 8c and Supplementary Fig. 5c). The *KRT1*-high region showed the highest number of spots with mimeTFs expression (Fig. 8c). The ranking of mimeTFs based on the frequency of occurrence within the *KRT1*-high region is shown in Fig. 8d. Notably, *AIRE* and *FEZF2*, known for tissue-restricted antigen expression regulation^{34,35} showed the highest specificity in the *KRT1*-high region, followed by other mimeTFs such as *ASCL1*, *INSML*, *DLX5* and *MYOG*. These results therefore pinpoint mimeTF-expressing TECs to discrete *KRT1*-high regions, establishing their geopositioning within specialised medullary microenvironments and HCs.

In addition to TFs, several genes are linked to mimetic TECs, such as *KRT10* in keratinocyte-TEC, *SPINK5* and *LY6D* in corneo-TEC, *DES* and *MYL1* in muscle-TEC^{15–17} (Supplementary Data 4). For example, most mimeTF spots express *KRT10*, *TTN*, and *UBB4B*, while some genes

are more specific; for instance, *KRT7* is associated with *SOX9*, and the *SPINK5* gene is expressed in spots containing mimeTFs such as *SOX2*, *IRF1*, *NEUROD2*, and *SPEN* at varying degrees (Supplementary Fig. 5d). The *NEUROD2* mimeTF-containing spots show high expression of *KRT10*, *GNG8*, *SFTPD*, *IL25*, and *SPINK5*. This suggests that these mTECs have the potential to differentiate into various molecularly distinct cell types based on the expression of mimeTFs.

Subsequently, to examine the physical proximity of mimeTFs expressed by the mimetic TECs, we computed the co-occurrence of mimeTFs (Fig. 8e). *AIRE* and *FEZF2*, known for tissue-restricted antigen expression regulation^{34,35}, were primarily co-localised with each other and formed a hub across all samples. Besides, a higher degree of co-localisation was observed among the mimeTFs *INSML*, *HES6*, and *ASCL1*. Furthermore, we analysed the spatial neighbours of mimeTFs within the same spot, five spots and 10 spots (Supplementary Data S5). The network illustrates that mimetic TECs expressing mimeTFs are typically located in close proximity to one another. Likewise, co-immunostaining of *MYOG* and *AIRE* supported their physical proximity, with 6.5% of *MYOG* and *AIRE* co-occurring at the same spots, 58.1% within five spots, and 74.2% within 10 spots,

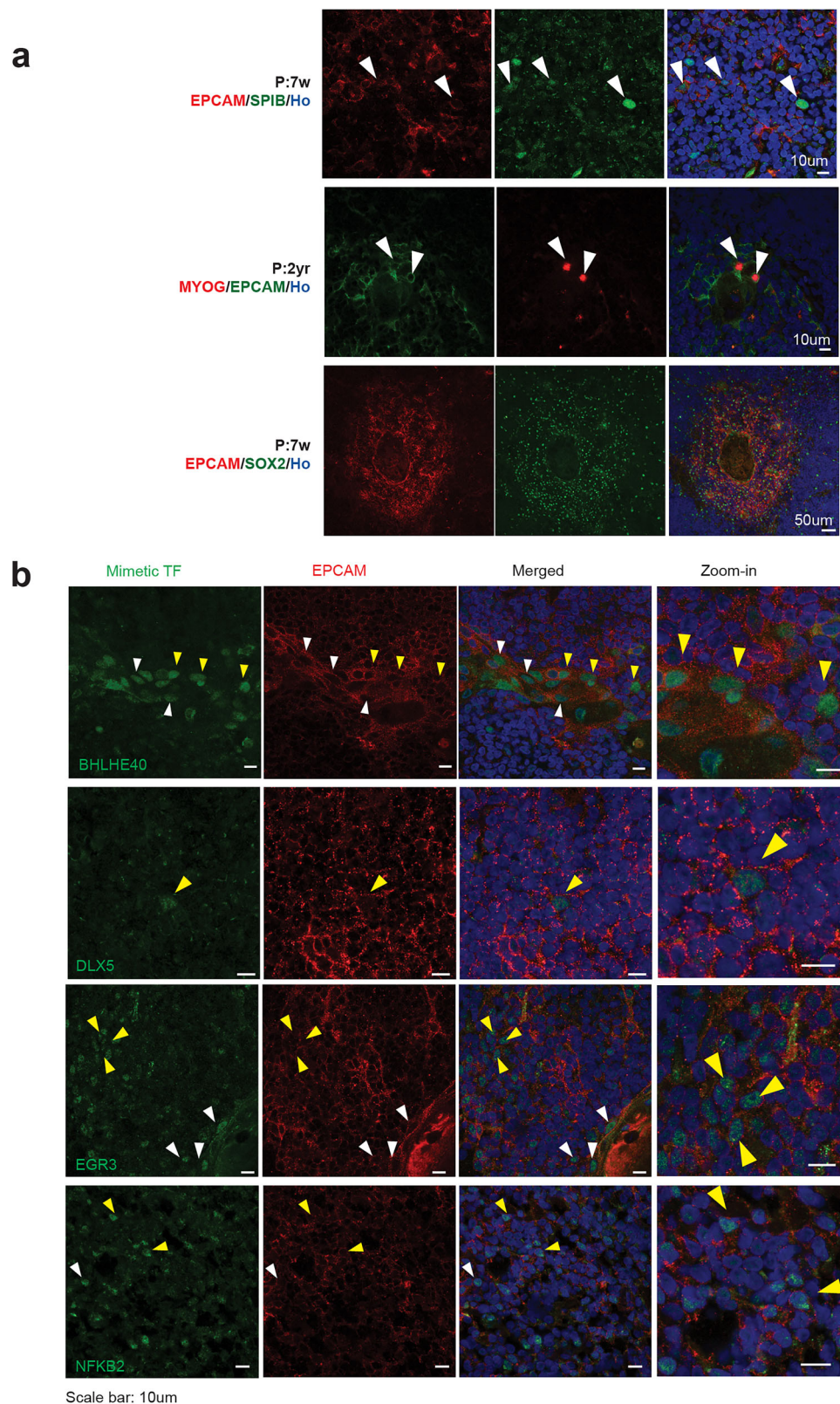
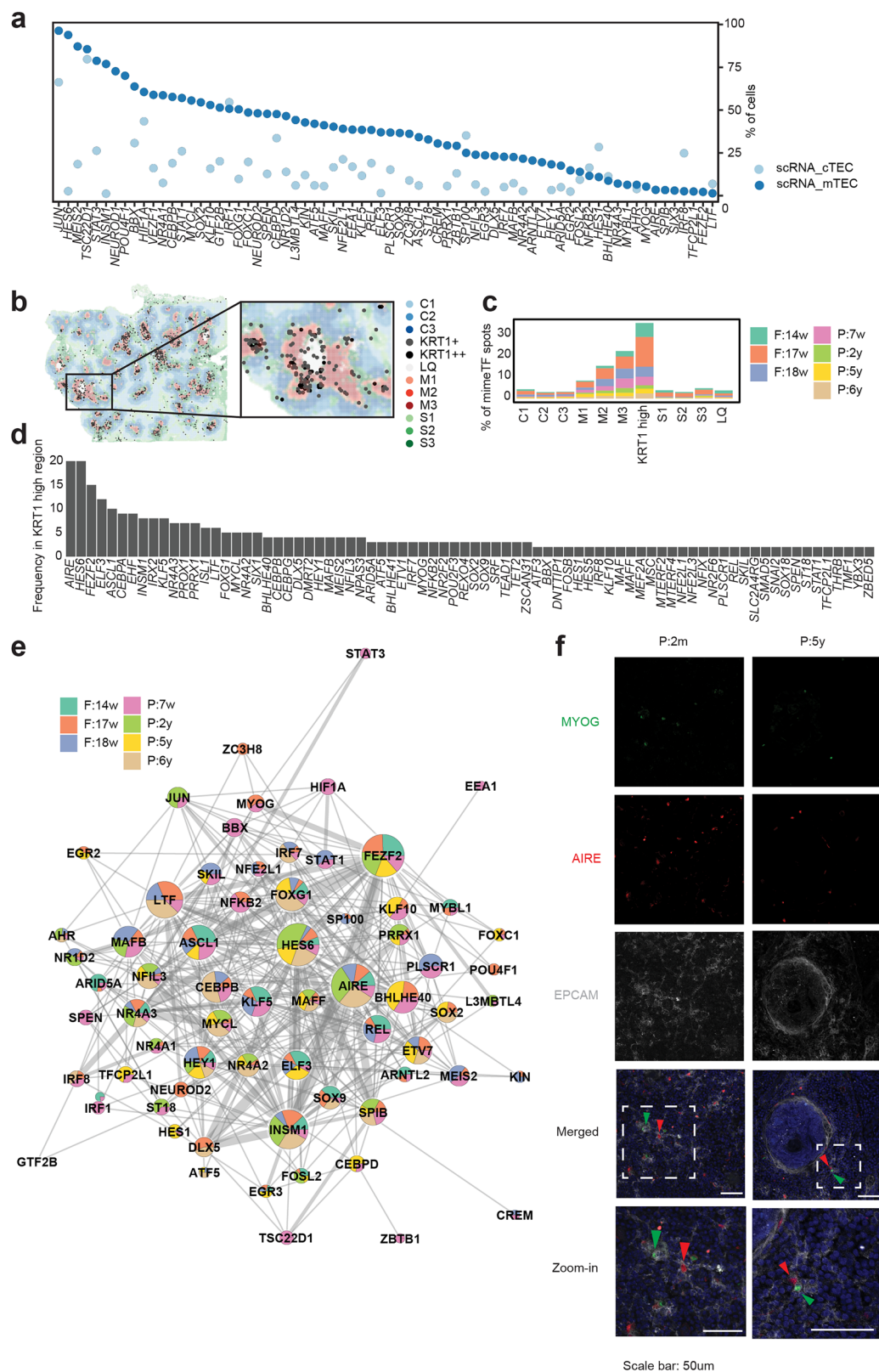


Fig. 7 | Immunostaining of the TFs expressed by rare mimetic TECs. a Co-immunostaining of reported mimetic TFs with pan-TEC marker EPCAM in the human thymic slides. Co-staining of mimetic TF SPIB or MYOG with EPCAM revealed their rarity in number, while co-staining of mimetic TF SOX2 with EPCAM showed a high number of SOX2+ TECs. Rare mimetic TF+ TECs are indicated by white arrows. Nuclei were stained with Hoechst. A representative region is shown

from at least three independent regions. **b** Co-immunostaining of mimetic TFs and EPCAM in human thymic slides of 5 y sample. Examples of cells co-expressing mimetic TFs and EPCAM were marked with arrows. Further zoom-in of selected cells was marked by yellow arrows. A representative region is shown from two to six independent regions. Nuclei were stained with Hoechst.



and significantly co-localised by Fisher's exact test with p values of 0.009, $<2.2 \times 10^{-16}$ and $<2.2 \times 10^{-16}$, respectively (Fig. 8f and Supplementary Fig. 5e, f). This shows a similar trend to that observed for the MYOG and AIRE spatial transcriptional co-localisation. Together, these multilayered analyses reveal that mimetic TECs are rare but spatially organised into tightly clustered, HC-proximal microenvironments, where their coordinated mimeTF expression and close physical

proximity suggest a specialised regulatory hub for self-antigen presentation and central tolerance.

Cell type mimicking function by mimeTFs expressing TECs

To better understand the role of mimeTFs in regulating cell type mimicking function, we conducted a Gene Ontology (GO) biological processes enrichment analysis on the differentially expressed genes in

Fig. 8 | Abundance of mimeTFs and enrichment in KRT1+ spots. **a** The points in the graph indicate the percentage of TEC cells that have a normalised expression of the identified mimeTFs greater than one in the scRNA-seq dataset comprising EPCAM+ sorted cTECs and mTECs. **b** Spatial visualisation of KRT1 spots in the medulla and LQ regions. KRT1 expression greater than one is referred to as KRT1+, and greater than 10 as KRT1++. **c** The relative percentage of spots containing mimeTFs within each niche and KRT1 high region, which includes KRT1+ and KRT1++ spots. **d** The identified mimeTFs are ranked based on the frequency of occurrence within the KRT1-expressing spots. **e** The network shows the colocalisation of

mimeTFs that occur in the same spots. The node size represents the degree or number of connections, and the pie chart displays the relative abundance of mimeTFs in each sample. The thickness of the edges shows the relative number of spots that express the connecting mimeTFs together. **f** Immunostaining of AIRE+ and MYOG+ TECs in human thymic slides. Co-staining of mimetic TF MYOG and AIRE, with pan-TEC marker EPCAM revealed the instances where TECs expressing these mimetic TFs were close (within 50 µm). Nuclei were stained with DAPI. A representative region is shown from four independent regions. F fetal, P paediatric.

spots that express mimeTFs (Supplementary Data 6). This revealed the presence of many non-thymic cell fate/lineage-related terms, which could infer the origin of their counterpart cell type. For instance, the GO term enrichment for known mimeTFs such as *INSML* correlated with pancreatic B cell differentiation. This finding aligns with observed autoimmune phenotypes in the murine pancreas following an *Insm1* conditional knockout in TECs³⁶. Additionally, *ASCL1* was associated with terms related to neuronal development, *MYOG* was connected to skeletal muscle development, and *POU4F1* indicated roles in bone and nerve development. *SOX2* was found to be involved in endocrine, eye, and pituitary gland development, while *BHLHE40* was implicated in the nervous system, *DLX5* in bone, skeletal and nervous system, and *EGR3* in endothelial and muscle development. Subsequently, we aimed to map the lineage-specific TF that controls the mimetic TECs for antigen presentation for the lineage. For this, we assessed the knockout effects of these TFs using known zebrafish knockout phenotypes from the ZFIN database³⁷ and human diseases from the Disease Ontology database³⁸ (Supplementary Fig. 6a and Supplementary Data 7). The mimeTFs exhibited similar functions based on the involved organs, with major organ groups including the nervous system, brain, cardiac system, and eye. For instance, similar to the GO analysis, *ASCL1*, *DLX5* and *BHLHE40* were also associated with the nervous system. These findings highlight the potential roles of these factors in presenting antigens by mimicking other cell types for T cell maturation.

Next, we resolved the cell types located in close proximity to the mimeTF-expressing TECs. As expected, immune cell types such as B cells and DCs, mTECs, and SP thymocytes showed higher enrichment (Supplementary Fig. 6b, c). Notably, the mimetic TEC subtypes myo and neuro exhibited some of the highest enrichment across all spots expressing mimeTFs (Supplementary Fig. 6c). Since TFs are known to regulate tissue-specific expression, in this study, we aim to investigate whether these mimeTFs influence the genes responsible for mimicking other cell types. Specifically, we sought to determine if spatial data can be used to directly quantify genes that are not specific to the thymus, as well as antigen-presenting genes. For instance, Supplementary Data 8 presents the TFs alongside their target genes, which include known tissue-specific and antigen-presenting genes. The tissue-specific genes listed are not related to the thymus, suggesting a cell-mimicking function regulated by these mimeTFs. Next, the TF regulon for each mimeTF was computed, and for some mimeTFs, such as *DLX5*, *SOX9* and *MEIS2*, the top TF regulon identified was the corresponding TFs (Supplementary Fig. 6d). For these, we explored the expression of tissue-specific genes from the Human Protein Atlas³⁹ within the TF regulon target genes and found a varying number of tissue-specific genes regulated. For instance, *MEIS2* regulates *FBXO32*, which is specifically expressed in skeletal muscle (Supplementary Data 8). Some of the target genes were also involved in the antigen presentation process, including the MHC class I and II. In line with the enrichment of APC cells in the mimetic spots, we also observed varying expression of HLA variants, DC, and NK cell receptors (Supplementary Fig. 6e). The associations with tissue-specific GO terms, disease phenotypes, and regulon targets indicate that mimeTFs drive cell type mimicking functions that broaden the self-antigen landscape presented to developing T cells. Overall, this study demonstrates the ability of high-

resolution spatial omics to geolocation, characterise and investigate rare cell types transcriptionally primed to support central tolerance within the thymus.

Discussion

Our study presents a comprehensive, high-resolution spatial atlas of the fetal and paediatric human thymus, featuring a fine dissection of cellular compositions for each regional niche. The cartographic nature of our integrative spatial map enables us not only to directly locate the previously reported cell types via gene marker visualisation but also to resolve the spatial distances among subtypes of each thymic cell type. Our study has demonstrated the geo-position of thymic niches, paving the way for further examination of sub-regional functions during thymus development. Moreover, using the defined cell-type-specificity scoring metrics, we can directly assign the geo-position of the cell type in the thymus tissue.

Recent advances in both single-cell and spatial transcriptomics have enhanced our understanding of human thymic development, enabling the definition of the structural domains (i.e., cortex, medulla, and septa) despite the compact nature of thymic tissue. The application of spatial Stereo-seq technology serves as one approach to decipher the regional complexity of the human thymus. Compared to a previous study^{9,10,12}, our high-resolution atlas enables capturing cell types via direct detection of gene markers. Known histological regions, including cortex, medulla and septa, are identified via spatial profiling of regional markers, such as *PRSS16* (cortex) and *AIRE* (medulla). Each thymic cell type comprises multiple subtypes, usually differentiated by a few gene markers. Our high-resolution map allows for a direct comparison of their regional distribution.

The study examined the signalling pathways active during T-cell selection in thymic niches. Since the fetal samples could capture the capsule intact, pathways such as MK and IGF were prominent, indicating interactions between cTECs and fibroblasts. While the CXCL pathway was key to the migration of DP thymocytes in cortex. In contrast, the thymic medulla showed enrichment in pathways like CCL, MIF, and TGFB, which are crucial for establishing self-tolerance via antigen presentation. *MIF* was associated with MHC-II presentation, while *CCL19* and *CCR7* were essential for thymocyte migration and maturation, as discussed previously^{9,10}. This highlights the intricate signalling networks within thymic niches that are essential for T cell maturation and self-tolerance.

Recent studies have profiled many different types of TEC mimetics at the single-cell level^{15–18}. However, a systematic method for unbiased detection of TFs involved in TEC mimetics using spatial omics and their spatially interacting information remains less examined. In our study, we employed our high-resolution thymus atlas and multi-omic spatial data, including proteomic and transcriptomic information, to identify TFs that regulate the mimetic TEC. We found that around 0.4% of cells spatially expressed mimeTFs associated with mimetic TECs. Several well-studied TFs in mimetic TECs, including *ASCL1*, *INSML*, *NEUROD1*, *MYOG*, *POU4F1*, *SOX2*, *SP1* and *JUN* were identified. We observe a significant increase in the presence of mimeTFs in the KRT1-high region located near and surrounding the HCs. We estimate the functionality of these novel mimeTFs, such as

BHLHE40 to be implicated in the nervous system, *DLX5* in the nervous system and bone, and *EGR3* in endothelial and muscle-related cell type presentation. Further, we establish that the mimeTFs are spatially co-localised with *AIRE*, *FEZF2*, *HES6* and *INSML1*, forming a hub. Our study has a unique advantage in determining the spatial geopositioning of the mimeTFs with their interaction partners and estimated antigen presentation from different tissue origins.

Nevertheless, this study also presents some limitations. Owing to the technical challenges, a single-cell resolution of the spatial atlas remains lacking. Currently, larger cell sizes like fibroblasts can be near the single-cell level at 50 bin or 25 μm spots, but thymocytes are cell types that are many magnitudes smaller. Here, we focus on thymic epithelial cells in identifying mimetic TECs, which are approximately 10 μm . Hence, we try to overcome this by defining the spots at bin 20 or 10 μm resolution as close as epithelial cell size in multi-omic transcriptome and proteome data. However, the thymus is a highly compact tissue, making it more challenging to identify the borders of individual cells within it. Therefore, a higher-resolution atlas with cell border imaging could allow for better single-cell visualisation of cell-cell interactions among the subtypes of thymic cells. In addition, higher spatial resolution would enable detailed mimeTF-based TF-gene regulon identification. Further, the spatial aspect can be integrated with multimodal data, including scRNA-seq^{40–42}, chromatin accessibility⁴³, histone profiling⁴⁴, DNA modification⁴⁵ and RNA modification⁴⁶ to capture the overall cellular dynamics in the thymus.

In summary, our study has established an integrated high-resolution spatial atlas for the human fetal and paediatric thymi, enabling fine dissection of thymic niches for major histological structures. We leverage the spatial profiling to identify TFs implicated in rare mimetic TECs. This spatial atlas enables the identification of rare cell types, as the mimeTFs are detected in less than 0.4% of the tissue section and spatially co-localised to form a hub that matures T cells via antigen presentation. This atlas could serve as a resource for future studies dissecting thymic microenvironments, central tolerance and immune dysregulation.

Method

Sample collection

Mouse thymic samples were harvested from euthanised C57BL/6, specifically 1-month-old and 3-month-old (IACUC: 221691). All work involving animals was conducted following the ethical guidelines of A*STAR. The housing facility was the Biological Resource Centre, A*STAR, which was a pathogen-free facility. The housing temperature was maintained at 21 $\pm$ 1 degrees, 45–70% humidity and a 12/12 dark/light cycle.

Human thymic samples were collected after obtaining consent from patients, with approval from the SingHealth Centralised Institutional Review Board and National Healthcare Group Health Domain Specific Review Board (CIRB Ref: 2013/837/D; NHG DSRB Ref: 2023/00049 – “Understanding Human Thymic Ageing and Its Mechanisms”). Fetal thymic samples were obtained from aborted fetuses of indicated weeks of gestation (i.e. 13w, 14w, 17w and 18w). Informed consents were obtained from all women for the donation of their fetal tissues for research. Paediatric thymic samples (i.e., 7w, 2 y, 5 y, and 6 years) were collected from patients undergoing cardiac surgery after obtaining consent from them or their legal guardians (prior to surgery). Freshly harvested thymus samples from donors were kept and transported on cold DMEM media (Gibco, 11965092).

Sample embedding and freezing

Prior to embedding, each thymic tissue was wiped dry with tissues and Kimwipe to remove unwanted surface liquid. Subsequently, the tissue was dissected and trimmed to smaller sizes comparable to that of a cryomold (Peel-A-Way® 22 mm square, 20 mm deep, 27110 for paediatric samples; SAKURA Finetek Cryomold Intermediate,

15 $\times$ 15 $\times$ 5 mm, 4566 for fetal samples). Trimmed tissues were embedded in OCT (Tissue-Tek O.C.T. Compound, SAKURA, 4583) with adjustment to desired orientation. After that, the cryomold with tissues embedded in OCT was placed in a mixture of dry ice and ethanol for freezing. Upon a complete change in OCT from transparent to fully white in appearance, the freezing was finished. The OCT-tissue blocks were then stored on dry ice and subsequently at -80°C for long-term storage.

For all subsequent spatial profiling, frozen human thymic tissues were sectioned with the same cryostat (LEICA CM3050S).

Stereo-seq spatial transcriptomics profiling

The samples were processed for Stereo-seq according to the protocol described in study¹³ as well as manufacturer's instructions. For fetal thymic samples (13 week, 14 week, 17 week, 18 week), Stereo-seq was performed with standard procedure. Briefly, the tissues were sectioned at 10 μm , and mounted to a 1 $\times$ 1 cm Stereo-seq chip (T chip, 11OCT114). Mounted tissue on chip was dried at 37 degrees, followed by fixation with cold methanol at -20°C . Nuclei staining was performed with Qubit ssDNA reagent (Invitrogen, Q10212A) as recommended and imaged at $\times 10$ magnification (Olympus, FV3000RS Upright). Subsequently, the tissue section was permeabilized (37 degrees) and then reverse transcribed (42 degrees) for at least 3 h. Permeabilization duration was pre-determined with permeabilization test kits (P chip kit, 11ISP118). Tissues were then removed (55 degrees, 10 to 20 min), and cDNA was released overnight (55 degrees, between 3 to 16 h). Released cDNA was collected and purified following manufacturer's instructions with SPRI beads (SPRIselect, B23318). Raw cDNA was amplified further and purified again with SPRI beads. A total of 20 ng amplified cDNA was input for library processing. Libraries were prepared based on manufacturer's instructions (Stereo-seq Library Preparation Kit, 11IKL114). Processed libraries were sequenced with a MGI DNASEQ-Tx sequencer.

In-house modifications were tested for paediatric thymic samples (2 years, 5 years, 6 years) to improve gene capture. Two key modifications were implemented: (1) mounted tissues were dried at -20°C with desiccant for around 30 min (instead of 37 degrees) prior to fixation to reduce RNA degradation; (2) the overall workflow was changed to Stereo-CITE protocol without antibody added, to accommodate the change of fixative from methanol to 4% paraformaldehyde (PFA) (Boster, AR1068). In brief, tissue sections were mounted to a 1 $\times$ 1 cm Stereo-seq chip slide (10 μm for 2 years old sample, 5 μm for 6 years old sample, T slide, 21OCT114), followed by low temperature drying at -20°C . Mounted tissue was subsequently fixed with 4% PFA (Boster, AR1068) at room temperature (RT) for 10 min, followed by blocking at RT with pre-prepared blocking buffer as recommended by the manufacturer (Horse Serum, Thermo Fisher, 26050070; Goat Serum, Thermo Fisher, 16210064; Triton X-100 Solution, Sigma Aldrich, 93443-100 ML; Blocking reagent from Stereo-CITE seq's protein assist kit; RNase Inhibitor, Thermo Fisher; Salmon Sperm DNA sheared, Invitrogen, AM9680) for 65 min. No FcR blocking reagent was added as no antibody would be added later. Nuclei staining was then performed with DAPI (Thermo Fisher, 62248) and imaged at $\times 10$ magnification (Olympus, FV3000RS Upright). De-crosslinking was conducted at 70 degrees for 15 min. Upon cooling down to room temperature, the section was permeabilized at 37 degrees with duration pre-determined using Permeabilization test kits (21ISP118), followed by reverse transcription for at least 3 h at 42 degrees. The downstream steps were similar to those for fetal sample processing, including tissue removal, cDNA release, cDNA purification and amplification, as well as library processing.

Stereo-CITE seq spatial transcriptomics profiling

Stereo-CITE seq was performed only for paediatric 7 week sample, following manufacturer's instructions. The tissue was sectioned at

10 µm and mounted to a 1 × 1 cm Stero-seq T chip slide (210CT114). The section was dried at 37 degrees, followed by 4% PFA fixation. The remaining steps were the same as described above, except for blocking step and imaging step. Blocking was performed at room temperature with FcR blocking reagent (Biolegend, 422301) added for 20 min. A cocktail of 31 primary antibodies was prepared and added to the tissue section (45 min at room temperature). Secondary antibody incubation was performed at room temperature for 15 min. DAPI staining was performed after secondary antibody incubation, at room temperature for 2 min. Nuclei and protein imaging were then performed. The primary antibody list is shown in Supplementary Table 1.

MACSima spatial profiling

MACSima spatial profiling was performed by Singapore Immunology Network (SiGN) as a contracted service. Frozen human thymic tissue was sectioned at 7 µm, and mounted onto a SuperFrost Plus slide (Fisher Scientific, FIS12-550-15). The slide was transported on dry ice to SiGN for downstream MACSima processing, which followed the guidelines described by the earlier study¹⁴. Briefly, the mounted section underwent fixation and permeabilization (in blocking buffer) prior to cyclic antibody staining. A total of 22 primary antibodies were used (see below), with conjugation to either fluorescein isothiocyanate (FITC), phycoerythrin (PE) or allophycocyanin (APC). Each cycle of antibody staining was followed by immediate imaging, before the start of the next antibody staining cycle. The entire process was automated using the MACSima instrument. The primary antibody list is shown in Supplementary Table 2.

Hematoxylin and eosin (H&E) stain of thymic slides

The OCT blocks of both murine and human thymic tissues were sectioned using a cryostat to 10 µm per section for H&E staining. All H&E staining of thymic slides was performed with the H&E Staining Kit (Hematoxylin and Eosin) (Abcam, ab245880) according to the manufacturer's protocol. Briefly, the thymic section was fixed with cold methanol at −20 degrees for 30 min. Subsequently, the fixed slides were stained with Hematoxylin, Mayer's solution at room temperature for 5 min, followed by bluing reagent for 10–15 s. Then, section was stained with Eosin Y Solution for 3 min. The stained tissue was mounted with DPX Mountant for histology (Sigma, 06522-100 ML). In between each staining step, sections were washed with distilled water or absolute ethanol as instructed in the manufacturer's protocol. All H&E slides were imaged with the Zeiss AxioObserver7 HCS at ×10 magnification.

Immunostaining of thymic slides

Human thymic slides (10 µm per section) were fixed with 4% paraformaldehyde at room temperature for 10 min, followed by one to two times washing with 1X Phosphate buffered saline (PBS, Gibco, 10010031). Fixed sections were subsequently permeabilized with 0.25% Triton-X (Sigma) at room temperature for 10 min, followed by washing one to two times with 1X PBS. Blocking was performed with 7% Fetal bovine serum (FBS, heat-inactivated, Gibco, 16140071) or 1% Bovine Serum Albumin (BSA, Capricorn Scientific, BSA-1S) at room temperature for at least 1 h. Sections were then incubated with primary antibodies followed by secondary antibodies. Prior to mounting with a mounting medium (VectorShield, H-1000-10), sections were stained with Hoechst for nuclei (1000x dilution with blocking buffer, Invitrogen, H3570) at room temperature for 10 min if the mounting medium was without DAPI. In between each incubation, 6 times of 1xPBS washes were performed. Primary antibody incubation was performed at 4 degrees for 16 h and secondary antibody incubation was done at room temperature in the dark for 1–2 h. All antibodies were diluted in respective blocking reagents. Primary antibodies (diluted to 1 µg/mL for all) used are in Supplementary Table 3. Secondary antibodies (1000x diluted) used were Goat anti-Rabbit Alexa Fluor 488 (Life

Technologies, A11008), Goat anti-Mouse Alexa Fluor 555 (Life Technologies, A21422), Donkey anti-Mouse Alexa Fluor Plus 488 (Thermo Fisher, A32766), Donkey anti-Rabbit 594 (Thermo Fisher, A21207), and Donkey anti-Goat 647 (Thermo Fisher, A32814). All slides were imaged with the Olympus FV3000RS Upright Microscope or the Zeiss LSM700inv Microscope.

Quantification of co-localisation using immunofluorescence

Quantification analysis on the immunofluorescence slides was performed using open-source software, Fiji⁴⁷, which is a distribution of ImageJ. A custom ImageJ macro was written to perform the analysis in Fiji. Due to the rare nature of myo-TECs, images were taken on regions with at least 1 myo-TEC present and with this ImageJ macro, those images can be batch processed. This macro uses Bio-Formats⁴⁸ plugins to open the microscope images and nuclei on Hoechst channel are segmented using the StarDist plugin. The segmented nuclei are then added as regions of interest (ROIs) to ImageJ ROI manager. Nuclei ROIs in the ROI manager that do not have EPCAM signal or are too small are then removed. The remaining nuclei ROIs are the TEC nuclei.

MYOG staining and AIRE staining were then used to identify which TEC nuclei ROIs in ROI manager are myo-TEC and mTECII, respectively. Staining control slides (i.e. slides stained with only secondary antibodies) were used to set the minimal intensity thresholds to reduce false-positive nuclei. For each positive myo-TEC nuclei, the macro can check if there is any mTECII nucleus within a circle of diameter either 50, 125 or 250 µm was labelled to identify the existence of any mTECII nucleus. A total of 23 images across six patients were included for quantification analysis.

At a defined spatial diameter (50, 125 or 250 µm), EPCAM⁺ spots were classified based on MYOG and AIRE expression into four groups: MYOG⁺/AIRE⁺ double-positive, MYOG⁺ only, AIRE⁺ only, and double-negative. Counts for each group were extracted, and the conditional proportion of MYOG⁺ cells co-localised with AIRE was computed. A 2 × 2 contingency table was then constructed to assess the statistical association between MYOG and AIRE expression using Fisher's exact test.

Sample dissociation to single cells

All fresh paediatric samples were dissected and trimmed prior to tissue embedding and freezing in OCT as described above. Small tissue pieces were randomly selected and collected for tissue dissociation, yielding single-cell suspensions. In brief, fresh thymic samples were minced into small pieces, followed by enzymatic dissociation with dissociation solution. Dissociation solution was prepared by mixing Collagenase IV (Gibco, 17104019) at 2U/µL and DNase I (STEMCELL Technologies, 07900) at 10 µg/mL. Minced tissues were incubated with dissociation solution at 37 degrees for 30 min, with agitation using pipettes every 5 min. At the end of incubation, dissociated tissues were passed through a 70 µm cell strainer followed by a 40 µm cell strainer to create a single-cell suspension. Subsequently, red blood cells were lysed with ACK lysis buffer (Gibco, A1049201) via incubating at room temperature for 10 min. Cells were resuspended at 0.5% BSA for cell counting with an automated cell counter. Cell viability was measured with trypan blue (Gibco, 15250061). For long-term storage, cells were frozen in freezing media (DMEM/10%FBS supplemented with 10% DMSO) in liquid nitrogen.

snRNA-seq and scRNA-seq profiling

Due to limited recruitment of fetal samples, no extra fetal samples were collected for dissociation to yield fresh single-cell suspensions. Hence, snRNA-seq was performed instead. Five to ten sections of frozen thymic tissues (at 10 µm thickness) were used to extract nuclei via commercial nuclei isolation kits (10x Genomics). Extracted nuclei were fixed and subsequently subjected to fixed RNA profiling kits (10x Genomics). All steps were performed following the manufacturer's instructions.

For paediatric samples with dissociated single cells, frozen vials were thawed, followed by dead cell removal (Dead Cell Removal Kit, Miltenyi, 130-090-101). Viable cells were recovered and subsequently subjected to MACS sorting prior to scRNA-seq. MACS sorting was performed according to the manufacturer's protocols (Miltenyi) with the following microbeads: CD3 (Miltenyi, 130-097-043), EPCAM (Miltenyi, 130-061-101). Briefly, total thymic cells were incubated with CD3 MACS beads at 4 °C for 15 min, followed by magnetic separation with MS columns (Miltenyi, 130-042-201) or LS columns (Miltenyi, 130-042-401). Flow-through containing CD3-ve populations were then collected and incubated with EPCAM MACS beads at 4 degrees for 30 min. Upon magnetic separation, both populations were collected. MACS-sorted CD3+, CD3-/EPCAM+ and CD3-/EPCAM- populations were subjected to scRNA-seq with 3' scRNA-seq kits (10x Genomics, 1000127, 1000268, 1000190). All sample processing and subsequent library construction were performed according to the manufacturer's instructions.

sc/snRNA-seq data processing

Cellranger v7.1 was utilised to map the sequencing data from 10x scRNA-seq and fixed RNA profiling snRNA-seq to the respective mm10 v32 mouse and GRCh38 v32 human reference genomes. The count data was processed for downstream analysis using Seurat v4.3⁴⁹. The cells were then filtered out if the cells had less than 500 genes or greater than 10% mitochondrial reads. Additionally, cells with UMI counts less than 0.01% quantile and greater than 95% quantile across all cells were removed. The genes were filtered out if they were detected in less than 10 cells. Only genes with log10 of average UMI count across all cells greater than -2.5, genes with at least 2 UMIs and genes detected in at least 10 cells were used for downstream analysis. To identify and eliminate doublets, DoubletFinder v2.0.3³⁰ was used. The gene count was normalised using SCTransform⁵¹ and dimension reduction was performed using PCA and UMAP.

sn/scRNA-seq integration with public data

The sc/snRNA-seq data from this study were integrated with a publicly available thymus scRNA-seq dataset. First, the cells from the public dataset were grouped by sample. Then, the samples from both the public dataset and this study were integrated. Integration anchors were identified using the 'FindIntegrationAnchors' function with RPCA reduction, incorporating the first 50 PCs. The integrated data was scaled, and a UMAP projection was computed. There were 290,401 cells representing the diverse cell types present in the thymus. The cell types were annotated by projecting the cells in this study to the reference public dataset³ cell types in Seurat. The differentially expressed genes were identified using the FindAllMarkers function with MAST differential expression testing from Seurat, with a minimum proportion of cells of 0.25 and a minimum log fold change of 0.25.

Spatial transcriptomics data processing

We performed spatial transcriptomics on thymus tissues using Stereo-seq¹³. The SAW v6¹³ pipeline was used to process the spatial image and transcriptome to obtain the gene expression profile for each DNB spot. The sequencing reads of mouse and human thymus were aligned to mm10 v32 mouse and GRCh38 v32 reference genomes, respectively. The downstream analysis was performed at a bin 50 resolution, corresponding to a 25 µm diameter spot in the tissue. The analysis was performed using StereoPy v1.0¹³ and Scanpy v1.9.6⁵². The low-quality spots are based on the number of genes expressed in each spot. For bin 50, spots expressing more than 100 genes were selected for samples with an average of unique genes detected greater than 500. For samples with fewer than 500 unique genes, spots that expressed more than 75 genes were selected. The spots with mitochondria reads greater than 10% were removed. The spots were then normalised using SCTransform normalisation. Spatial neighbours were identified using

the top 30 PCs. Then, spatially constrained Leiden clustering¹³ was performed to identify the transcriptionally and spatially similar spots with default parameters. The spatial neighbourhood enrichment analysis was performed using Squidpy v1.3.0⁵³. The identified Leiden clusters were then visualised as edge-weighted networks based on the neighbourhood z-score between clusters.

Integration of spatial clusters across samples

First, for each sample, we identified the Leiden clusters through spatial inference. To obtain the gene signature of these clusters, we utilised decoupler v1.8.0⁵⁴ to sum the raw counts, creating pseudobulk gene expression profiles for the clusters across all samples. We used a set of genes representing the major cell types in the thymus as features: *LY75*, *PRSSI6*, *PSMB11*, *FABP5*, *RUNX3*, *AQP3*, *ZBTB7B*, *FOXP3*, *TOX2*, *LAMP3*, *IGHM*, *DNTT*, *MKI67*, *TRAF1*, *MS4A1*, *TRAF5*, *CD8A*, *CLEC10A*, *CD14*, *CD68*, *MS4A7*, *TPSAB1*, *TPSB2*, *KRT15*, *EPCAM*, *AIRE*, *FEZF2*, *KRT1*, *KRT10*, *VWF*, *PECAM1*, *CDH5*, *HBB*, *CD34*, *COL1A1*, *COL1A2*, *LUM* and *ACTA2*. The pseudobulk transcriptional profiles of these features were then batch-corrected using pyComBat v0.3.3⁵⁵. Following this, we conducted PCA analysis on the batch-corrected pseudobulk cluster profiles and used the top five principal components to fit the model. The K-means algorithm was applied to cluster the pseudobulk profiles into ten groups. These ten groups were then assigned to specific regions: the cortex, medulla, and septa. The assignments were based on the expression of key marker genes: *AIRE* and *MS4A1* for the medulla, *PRSSI6* and *LY75* for the cortex, and *COL1A2*, *CD34*, and *VWF* for the septa. The group exhibiting the lowest overall gene expression was classified as low quality. For comparison between different technologies, such as transcriptome and proteome analysis, we utilised TissueTag v0.2.2⁹, which marks a common continuous axis to subdivide the cortex, medulla, and septa. The cortex was further divided into three niches, from the outer cortex I to the innermost cortex III, by splitting the region into equal parts containing a similar number of cells. Similarly, the medulla and septa were separated into three niches.

Spatial cell type annotation

The cell-type deconvolution of spatial bin 50 spots was carried out using RCTD v2.2.1⁵⁶. RCTD used a supervised learning approach to decompose the integrated scRNA-seq data and predict the cell types present in the spatial spots. RCTD was run in full mode across all samples. We set the parameter of a maximum of 3 cells within a 25 µm spot based on the assumption that an inactive T cell is approximately 8 µm, an epithelial cell is about 12 µm, and a fibroblast is around 15 µm in size. The number of cell types assigned to each cluster or niche was scaled to a range of 0 to 1 based on the total number of spots within each niche. Cell types with values greater than 0.5 were retained and visualised using a heatmap.

To calculate the cell-type-specificity score, we combined known cell-type marker genes with the top ten differentially expressed genes identified from integrated fetal and paediatric scRNA-seq data using the FindMarkers function in Seurat. The known cell-type marker genes are detailed in Fig. 2b. A complete list of the cell type-specific genes utilised in this study for calculating the specificity score can be found in Supplementary Data 2. These cell type-specific genes were used to compute the expression score via the gene score function from Scanpy. Spots with cell type-specific scores equal to or greater than 0.75 were assigned to the corresponding cell type. The number of cell types assigned to each cluster or niche was scaled to a range of 0 to 1 based on the total number of spots within each niche. Cell types with values greater than 0.5 were retained and visualised using a heatmap.

Three gene co-expression

The enrichment for each of the genes in the three gene combination was calculated using Scanpy score genes, and spots that were greater than a threshold of 0.1 were used to select spots expressing the gene.

The colours used were red, blue and green to mark the spots expressing only one gene and the respective combination of colours was used to indicate the presence of 2 or 3 gene expressions. To quantitatively show the spots co-expressing two or more genes, we visualised a heatmap with a number showing the exact number of spots and the colour gradient normalised across domains within each sample.

MACSima spatial image analysis

The cyclic Macsima run for 22 proteins stated above was performed in a paediatric sample of age 7 weeks. The image datasets for each protein were imported into the MACS iQ View Software v1.3.1 for analysis. Segmentation and manual contrast adjustments were applied to the images, and the mean fluorescent intensities for each cell were calculated against the background. The protein abundance values for each spot, based on the *x* and *y* axes, were exported for further analysis using Seurat. The data were log-normalised, and PCA was performed. The cells were clustered into niches based on a continuous tissue axis, and the protein abundance within each niche was visualised using a dot plot.

Signalling pathways

For the spatial transcriptomic data used in the study, cell-cell communication patterns within the cortical and medulla niches at bin 50 resolution were computed using CellChat v2.1.2⁵⁷. The spots were spatially constrained to 200 µm in diameter for the ligand-receptor pair interaction calculation. The communication probability was calculated using type as “truncatedMean”, trim parameter at 0.05 and scale distance at 0.1 for all the samples. The communication pathways were filtered based on the presence in at least 10 spots. The signalling role score across all the samples was obtained and normalised to a 0 to 1 scale within each sample. The overall signalling pathways with scores combining the incoming and outgoing signals were then visualised using a ggplot2 v3.5.1 tile graph.

Identification of mimeTFs in spatial transcriptome

Since mimetic TECs refer to a group of TECs with distinct expressions of TFs that are usually from non-thymic lineages, we used an *in silico* approach to screen for all possible TFs expressed by these TECs. The pipeline is in Python using scanpy v1.9.6, pandas v2.2.3, and matplotlib v3.7.5.

Step I: Identifying and extracting the TEC spots. The mimetics are TECs, and from the scRNA-seq data, we see that *EPCAM* is only expressed by TECs in the thymus. We used *EPCAM*, a pan-TEC marker, to mark the spots that contained TECs. The spots that expressed *EPCAM* greater than one were selected.

Step II: Selecting the TF that are expressed in the TEC spots. The mimetic TECs are known to express lineage TFs; hence, 1639 known human TFs were obtained from the previously published human transcription factor database⁵⁸. Next, within the *EPCAM*+ spots, the TFs expressed with normalised expression levels greater than 10 and in at least three spots on the tissue were selected.

Step III: Determination of TF expression that is specific to TECs. Next, the ZNF TFs within the filtered TFs were removed as the biological function for these TFs is complex to interpret. The TFs are expressed in a spot that may also contain other cell types along with TECs. To ensure the TFs are being expressed by the TECs, we used publicly available scRNA-seq data on *EPCAM*+ sorted TECs^{5,17}. The TFs were further sub-selected based on whether the TFs had a normalised expression level of at least five in TECs from scRNA-seq data.

Step IV: Spatial-specific TFs were determined. Based on the distribution of spots, the TFs were assigned to the cortex, medulla, and

septa domains using the majority voting approach. Seventy TFs were assigned to the medulla, 138 to the cortex, and 46 to the septa. The TFs were present in 2674 spots across all samples, which accounts for 0.37% of all the tissue spots profiled. The TF assigned to medulla were referred to as mimeTFs and was present in 783 spots across all ages.

Spatial multi-omics data processing

We performed spatial transcriptomics and proteomics on thymus tissue using Stereo-CITE. The SAW v7.1.2 pipeline processed the spatial image and transcriptome to obtain the gene expression profile for each DNB spot. The sequencing reads of the human thymus were aligned to the GRCh38 v32 reference genome. The cite-spatial data was analysed at a bin 20 or 10 µm resolution to be closer to the epithelial cell type. The transcriptome data were log-normalised using Scanpy functions. The spot of the proteome data for each protein was normalised using SpatialGlue v1.1.5⁵⁹.

Identification of mimeTFs in spatial multiome

The Stereo-CITE spatial transcriptomics and proteomics were carried out on the P:7w sample simultaneously on the same tissue section, resulting in 635,400 spots. The analysis was performed at a bin 20 resolution, and the domains such as cortex, medulla and septa were defined using marker genes *PRSSI6* and *AIRE* in transcriptome data. The niches were defined using the tissue continuous axis, similar to the other transcriptome-only data. The proteome data spots greater than the mean distribution of CD326 (*EPCAM* protein) were selected to be *EPCAM*+ protein spots. Then the corresponding *EPCAM*+ transcriptome spots were used to identify mimeTFs. A similar approach as described in the previous section was used to identify mimeTFs. The only exception was that the gene expression requirement of the TF in *EPCAM*+ spots was reduced to 5 instead of 10 to accommodate the lower gene expression profile at bin 20 resolution. The mimeTFs were present in 17459 spots in medulla, which is 2.7% of the tissue.

Defining *KRT1* high regions

Since *KRT1* expression was seen in the medulla and LQ regions, the determination of *KRT1*-high spots was restricted to these regions. The spots with *KRT1* expression greater than zero were grouped as *KRT1*+ and with greater than 10 as *KRT1*++ regions. Collectively, these spots are called *KRT1* high regions. The relative frequency of occurrence of mimeTFs within the *KRT1* high spots was determined.

Differential GO function

The spots expressing mimeTFs were grouped, and differential expression analysis was carried out for GO Biological Process₂₀₂₃ using drug2cell v0.1.0⁶⁰. The top 10 differential gene ontology terms were identified, and development-related terms were highlighted.

Co-localisation of mimeTFs

The co-localisation of expression on mimeTFs with other cell types was computed based on the cell-type-specificity score. Additionally, the co-localisation of different mimeTFs in the same spot was calculated based on the number of spots expressing mimeTFs. The co-localisation network was visualised using Networkx v3.1⁶¹ and Cytoscape v3.10.3⁶². The nearest neighbours with five bins-50 spots or 100 µm, and 10 spots or 200 µm were calculated using Euclidean distance between the *x-y* coordinates.

TF knockout phenotypes and disease phenotypes

The developmental knockout effect was found for each TF identified using the reported experimental phenotypes from the ZFIN database³⁷, which comprises zebrafish knockout phenotypes. mimeTFs associated with human diseases were obtained from the Disease Ontology database³⁸. Affected organ labels were assigned manually based on the affected organ in the zebrafish knockout phenotype or human disease.

Public Visium data analysis

The publicly available spatial Visium data was obtained from this study⁹. A similar procedure of computing the mimeTFs from spatial spots was employed. The spots that expressed *EPCAM* greater than one and were present in the medulla region were selected. Since the Visium data had lower resolution, the number of genes detected was comparatively lower. We reduced the gene expression cutoff from 10 to 5 for selecting TFs in *EPCAM*⁺ spots. Then TF, which were expressed in at least three spots, were selected as mimeTFs.

TF and target gene analysis

The TF-based gene-regulatory network for the spatial mimeTF spots was determined by using pySCENIC v0.12.1⁶³ and cistargetDB with the TFs and targets in the human genome. The target genes were screened for tissue-specific genes from the Human Protein Atlas³⁹ and antigen processing and presentation from the KEGG pathway.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The scRNA-seq and spatial transcriptomics data of the mouse and human thymi were visualised using a website developed using Shiny-Cell v2.1.0⁶⁴. The raw spatial and single-cell transcriptomic data are available in the GEO⁶⁵ repository under <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1045362> and <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1213311>. The analysed data generated in this study are available at the Zenodo repository <https://doi.org/10.5281/zenodo.12595241>. Public scRNA-seq data^{5,17} were obtained from the Zenodo repository (<https://doi.org/10.5281/zenodo.5500511>) and GEO: GSE220830. The publicly available spatial Visium data⁹ were obtained from <https://cellxgene.cziscience.com/collections/fc19ae6c-d7c1-4dce-b703-62c5d52061b4>. All data included in the Supplementary Information are available from the authors. The raw numbers for charts and graphs are available in the Source Data file whenever possible. Source data are provided with this paper.

Code availability

The custom scripts developed and used in this study are made available on GitHub and the Zenodo repository as Jupiter notebooks at <https://github.com/UmaSangumathi/mimeTFs.git> (<https://doi.org/10.5281/zenodo.17851510>).

References

- Xing, Y. & Hogquist, K. A. T-cell tolerance: central and peripheral. *Cold Spring Harb. Perspect. Biol.* **4**, a006957 (2012).
- Kumar, B. V., Connors, T. & Farber, D. L. Human T cell development, localization, and function throughout life. *Immunity* **48**, 202 (2018).
- Pearse, G. Normal structure, function and histology of the thymus. *Toxicol. Pathol.* <https://doi.org/10.1080/01926230600865549> (2006).
- Michelson, D. A. & Mathis, D. Thymic mimetic cells: tolerogenic masqueraders. *Trends Immunol.* **43**, 782–791 (2022).
- Park, J.-E. et al. A cell atlas of human thymic development defines T cell repertoire formation. *Science* **367**, eaay3224 (2020).
- Kernfeld, E. M. et al. A single-cell transcriptomic atlas of thymus organogenesis resolves cell types and developmental maturation. *Immunity* **48**, 1258–1270.e6 (2018).
- Lavaert, M. et al. Integrated scRNA-Seq identifies human postnatal thymus seeding progenitors and regulatory dynamics of differentiating immature thymocytes. *Immunity* **52**, 1088–1104.e6 (2020).
- Han, Y. et al. scRNA-seq profiling of neonatal and adult thymus-derived CD4⁺ T cells by a T cell origin-time tracing model. *J. Mol. Cell Biol.* **14**, mjac072 (2023).
- Yayon, N. et al. A spatial human thymus cell atlas mapped to a continuous tissue axis. *Nature* **635**, 708–718 (2024).
- Li, Y. et al. Unraveling the spatial organization and development of human thymocytes through integration of spatial transcriptomics and single-cell multi-omics profiling. *Nat. Commun.* **15**, 7784 (2024).
- Heimli, M. et al. Multimodal human thymic profiling reveals trajectories and cellular milieu for T agonist selection. *Front. Immunol.* **13**, 1092028 (2023).
- Suo, C. et al. Mapping the developing human immune system across organs. *Science* **376**, eabo0510 (2022).
- Chen, A. et al. Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell* **185**, 1777–1792.e21 (2022).
- Kinkhabwala, A. et al. MACSima imaging cyclic staining (MICS) technology reveals combinatorial target pairs for CAR T cell treatment of solid tumors. *Sci. Rep.* **12**, 1–16 (2022).
- Givony, T. et al. Thymic mimetic cells function beyond self-tolerance. *Nature* **622**, 164–172 (2023).
- Michelson, D. A., Hase, K., Kaisho, T., Benoist, C. & Mathis, D. Thymic epithelial cells co-opt lineage-defining transcription factors to eliminate autoreactive T cells. *Cell* **185**, 2542–2558.e18 (2022).
- Ragazzini, R. et al. Defining the identity and the niches of epithelial stem cells with highly pleiotropic multilineage potency in the human thymus. *Dev. Cell.* <https://doi.org/10.1016/j.devcel.2023.08.017> (2023).
- Huisman, B. D. et al. Cross-species analyses of thymic mimetic cells reveal evolutionarily ancient origins and both conserved and species-specific elements. *Immunity* **58**, 108–123.e7 (2025).
- Cordes, M. et al. Single-cell immune profiling reveals thymus-seeding populations, T cell commitment, and multilineage development in the human thymus. *Sci. Immunol.* **7**, eade0182 (2022).
- Li, Z. et al. CD83: activation marker for antigen presenting cells and its therapeutic potential. *Front. Immunol.* **10**, 1312 (2019).
- Muramatsu, H. & Muramatsu, T. Purification of recombinant midkine and examination of its biological activities: functional comparison of new heparin binding factors. *Biochem. Biophys. Res. Commun.* **177**, 652–658 (1991).
- Alawam, A. S., Anderson, G. & Lucas, B. Generation and regeneration of thymic epithelial cells. *Front. Immunol.* **11**, 535698 (2020).
- Lucas, B. et al. Progressive changes in CXCR4 expression that define thymocyte positive selection are dispensable for both innate and conventional $\alpha\beta$ T-cell development. *Sci. Rep.* **7**, 1–11 (2017).
- Tramont, P. C. et al. CXCR4 acts as a costimulator during thymic β selection. *Nat. Immunol.* **11**, 162 (2010).
- Dzhagalov, I. & Phee, H. How to find your way through the thymus: a practical guide for aspiring T cells. *Cell. Mol. Life Sci.* **69**, 663–682 (2012).
- Lynch et al. Cutting edge: IL-16/CD4 preferentially induces Th1 cell migration: requirement of CCR5 1. *J. Immunol.* **171**, 4965–4968 (2003).
- Cruikshank, W. W. et al. Molecular and functional analysis of a lymphocyte chemoattractant factor: association of biologic function with CD4 expression. *Proc. Natl. Acad. Sci. USA* **91**, 5109–5113 (1994).
- Liu, Y. et al. Identification of a CD4 domain required for interleukin-16 binding and lymphocyte activation. *J. Biol. Chem.* **274**, 23387–23395 (1999).
- Schröder, B. The multifaceted roles of the invariant chain CD74-More than just a chaperone. *Biochim. Biophys. Acta* **1863**, 1269–1281 (2016).
- Love, P. E. & Bhandoola, A. Signal integration and cross-talk during thymocyte migration and emigration. *Nat. Rev. Immunol.* **11**, 469 (2011).

31. Kozai, M. et al. Essential role of CCL21 in establishment of central self-tolerance in T cells. *J. Exp. Med.* **214**, 1925 (2017).
32. Kwan, J. & Killeen, N. CCR7 directs the migration of thymocytes into the thymic medulla. *J. Immunol.* **172**, 3999–4007 (2004).
33. Liu, F.-T. & Stowell, S. R. The role of galectins in immunity and infection. *Nat. Rev. Immunol.* **23**, 479–494 (2023).
34. Nishijima, H. et al. Aire controls heterogeneity of medullary thymic epithelial cells for the expression of self-antigens. *J. Immunol.* **208**, 303–320 (2022).
35. Takaba, H. et al. Fezf2 orchestrates a thymic program of self-antigen expression for immune tolerance. *Cell* **163**, 975–987 (2015).
36. Tao, W. et al. Insm1 regulates mTEC development and immune tolerance. *Cell. Mol. Immunol.* **20**, 1472–1486 (2023).
37. Bradford, Y. M. et al. Zebrafish information network, the knowledgebase for Danio rerio research. *Genetics* **220**, iyac016 (2022).
38. Schriml, L. M. et al. Human Disease Ontology 2018 update: classification, content and workflow expansion. *Nucleic Acids Res.* **47**, D955–D962 (2019).
39. Karlsson, M. et al. A single-cell type transcriptomics map of human tissues. *Sci. Adv.* **7**, eabh2169 (2021).
40. Gautam, P. et al. Multi-species single-cell transcriptomic analysis of ocular compartment regulons. *Nat. Commun.* **12**, 5675 (2021).
41. Ng, M. S. F. et al. Deterministic reprogramming of neutrophils within tumors. *Science* **383**, eadf6493 (2024).
42. Xing, Q. R. et al. Unraveling heterogeneity in transcriptome and its regulation through single-cell multi-omics technologies. *Front. Genet.* **11**, 662 (2020).
43. Xing, Q. R. et al. Parallel bimodal single-cell sequencing of transcriptome and chromatin accessibility. *Genome Res.* **30**, 1027–1039 (2020).
44. Warriar, T. et al. SETDB1 acts as a topological accessory to Cohesin via an H3K9me3-independent, genomic shunt for regulating cell fates. *Nucleic Acids Res.* **50**, 7326–7349 (2022).
45. Viswanathan, R. et al. DARE SOME enables concurrent profiling of multiple DNA modifications with restriction enzymes in single cells and cell-free DNA. *Sci. Adv.* **9**, eadi0197 (2023).
46. Hamashima, K. et al. Single-nucleus multiomic mapping of mA methylomes and transcriptomes in native populations of cells with sn-m6A-CT. *Mol. Cell.* <https://doi.org/10.1016/j.molcel.2023.08.010> (2023).
47. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
48. Linkert, M. et al. Metadata matters: access to image data in the real world. *J. Cell Biol.* **189**, 777–782 (2010).
49. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).
50. McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst.* **8**, 329–337.e4 (2019).
51. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019).
52. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
53. Palla, G. et al. Squidpy: a scalable framework for spatial omics analysis. *Nat. Methods* **19**, 171–178 (2022).
54. Badia-I-Mompel, P. et al. decoupleR: ensemble of computational methods to infer biological activities from omics data. *Bioinform. Adv.* **2**, vbac016 (2022).
55. Behdenna, A. et al. pyComBat, a Python tool for batch effects correction in high-throughput molecular data using empirical Bayes methods. *BMC Bioinforma.* **24**, 459 (2023).
56. Cable, D. M. et al. Robust decomposition of cell type mixtures in spatial transcriptomics. *Nat. Biotechnol.* **40**, 517–526 (2022).
57. Jin, S. et al. Inference and analysis of cell-cell communication using CellChat. *Nat. Commun.* **12**, 1088 (2021).
58. Lambert, S. A. et al. The human transcription factors. *Cell* **172**, 650–665 (2018).
59. Long, Y. et al. Deciphering spatial domains from spatial multi-omics with SpatialGlue. *Nat. Methods* **21**, 1658–1667 (2024).
60. Kanemaru, K. et al. Spatially resolved multiomics of human cardiac niches. *Nature* **619**, 801–810 (2023).
61. Hagberg, A. A., Schult, D. A. & Swart, P. J. Exploring network structure, dynamics, and function using NetworkX. In *Proc. 7th Python in Science Conference* (eds Varoquaux, G. et al.) 11–15 (Pasadena, CA, 2008).
62. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
63. Van de Sande, B. et al. A scalable SCENIC workflow for single-cell gene regulatory network analysis. *Nat. Protoc.* **15**, 2247–2276 (2020).
64. Ouyang, J. F., Kamaraj, U. S., Cao, E. Y. & Rackham, O. J. L. ShinyCell: simple and sharable visualization of single-cell gene expression data. *Bioinformatics* **37**, 3374–3376 (2021).
65. Barrett, T. et al. NCBI GEO: archive for functional genomics data sets-update. *Nucleic Acids Res.* **41**, D991–D995 (2013).

Acknowledgements

The National Research Foundation, Singapore (NRF) Investigatorship award NRFI2018-02 (Y.-H.L.); National Medical Research Council NMRC/OFIRG21nov-0088 (Y.-H.L.); Singapore Food Story (SFS) R&D Programme W22W3D0007 (Y.-H.L.); A*STAR Biomedical Research Council, Central Research Fund, Use-Inspired Basic Research—CRF UIBR (Y.-H.L.); Competitive Research Programme—CRP NRF-CRP29-2022-0005 (Y.-H.L.); A*STAR Industry Alignment Fund—Prepositioning Programme IAF-PP: H23J2a0095 (Y.-H.L.), EVANTICA IAF-PP: H23J2a0097 (Y.-H.L.). The Singapore Ministry of Education grant MOE-000112 (N.R.J.G.).

Author contributions

Y.-H.L. conceptualised, designed and supervised this study. U.S.K. conceptualised the computational pipeline and carried out analysis. Y.C. conceptualised, designed and carried out the experimental data generation and analysis. J.L. carried out the computational analysis. P.G., P.H. and C.K.C. generated experimental data. K.W., N.R.J.G., J.C., C.K.M.C., Q.C., Q.L. and L.G.N. analysed the data. U.S.K., C.Y. and Y.-H.L. wrote the manuscript with feedback from all authors. All authors approved and contributed to the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-026-68596-w>.

Correspondence and requests for materials should be addressed to Yuin-Han Loh.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2026