EVOLUTIONARY BIOLOGY

Spatially resolved single-cell atlas of ascidian endostyle provides insight into the origin of vertebrate pharyngeal organs

An Jiang[1]†, Kai Han[2]†, Jiankai Wei[1,3]†, Xiaoshan Su[2]†, Rui Wang[2], Wei Zhang[1], Xiawei Liu[2], Jinghan Qiao[1], Penghui Liu[1], Qun Liu[2], Jin Zhang[1], Nannan Zhang[2], Yonghang Ge[1], Yuan Zhuang[1], Haiyan Yu[1], Shi Wang[1,3], Kai Chen[4], Wange Lu[5], Xun Xu[6], Huanming Yang[6], Guangyi Fan[2,6,7]*, Bo Dong[1,3,8]*

The pharyngeal endoderm, an innovation of deuterostome ancestors, contributes to pharyngeal development by influencing the patterning and differentiation of pharyngeal structures in vertebrates; however, the evolutionary origin of the pharyngeal organs in vertebrates is largely unknown. The endostyle, a distinct pharyngeal organ exclusively present in basal chordates, represents a good model for understanding pharyngeal organ origins. Using Stereo-seq and single-cell RNA sequencing, we constructed aspatially resolved single-cell atlas for the endostyle of the ascidian *Styela clava*. We determined the cell composition of the hemolymphoid region, which illuminates a mixed ancestral structure for the blood and lymphoid system. In addition, we discovered a cluster of hair cell–like cells in zone 3, which has transcriptomic similarity with the hair cells of the vertebrate acoustico-lateralis system. These findings reshape our understanding of the pharynx of the basal chordate and provide insights into the evolutionary origin of multiplexed pharyngeal organs.

INTRODUCTION

The development of pharyngeal organs, including pharyngeal slits and its associated apparatus, represents a remarkable morphological innovation in deuterostome ancestors (*1–3*). This has been documented in stem echinoderms (*4*) and stem vertebrates (*5*). The most primitive form of the pharynx in extant deuterostomes has been described in hemichordates as simple pharyngeal endoderm outpockets arising from the foregut (*6*). The pharyngeal endoderm in vertebrates exhibits a common developmental pattern and regulatory network that closely resemble those of their ancestral species (*2, 7*). It also plays an important role in the development of the vertebrate pharynx (*8, 9*). The complexity of the vertebrate pharynx increases with the incorporation of the cranial paraxial mesoderm and neural crest–derived mesenchyme. This gives rise to a number of accessory organs within the pharyngeal arches. In addition to the thyroid gland, which develops from the pharyngeal endoderm, various specialized organs originate from the pharynx primordium, including the parathyroid gland, thymus, and auditory organ (*10*).

The endostyle is a unique pharyngeal organ found exclusively in nonvertebrate chordates and ammocoetes, and represents a key point in the evolutionary hierarchy bridging vertebrates and nonchordate deuterostomes. The endostyle takes the form of a groove-like channel that extends longitudinally along the ventral body wall. Both lateral sides of the endostyle, consisting of regional ciliated epithelial cells specialized in mucus secretion and histological support (*11*), play a role in facilitating the filter-feeding process (*12*). Because of its endodermal origin (*13*), the endostyle expresses regional-specific endoderm markers, showing tissue-specific expression patterns (*14–16*). The endostyle is found in three categories, including cephalochordates, urochordates, and ammocoetes, and has been recognized as a thyroid-equivalent organ because of the presence of cellular components capable of concentrating iodine (*17, 18*) and the capacity to synthesize thyroid hormones (*14, 19–21*). Recent studies have uncovered various functions for the endostyle and suggest that it may have a common evolutionary origin with vertebrate pharyngeal organs. For example, in colonial ascidians, the sinus region of the endostyle contains both hematopoietic stem cells (*22*) and other somatic stem cells (*23*). Peptidergic neurons have also been found within the sinus regions of the endostyle (*24*). Moreover, the endostyle and its encompassing sinus region exhibit immune functions as evidenced by immunohistochemical labeling in response to exposure to pathogen-related biochemical compounds (*25, 26*). Together, the endostyle contains diverse cell types with distinct properties. These cell types may share common ancestors with functional-equivalent cell populations in vertebrates.

The basic steps for investigating complex organ evolution include defining and comparing cell types within organs (*27*). Using single-cell transcriptomics, it is feasible to define the cell profile of an organ at single-cell resolution (*28–30*). Nevertheless, when analyzing organs characterized by specialized tissue arrangement patterns, such as the endostyle, it becomes critical to capture transcriptomic data at specified locations. To thoroughly investigate the cell composition of the endostyle and its role in the evolution of pharyngeal organs, we used a spatial transcriptomic technique known as Stereo-seq (*31*) in conjunction with a high-throughput, single-cell transcriptomic

method provided by 10x Genomics (*32*). Using this approach, we constructed a location-addressable cell transcriptional profile for transverse sections of the endostyle in the urochordate ascidian, *Styela clava*, in which the genome and transcription factors have been well characterized (*33*, *34*). We produced a spatially resolved atlas of the endostyle and defined 23 cell clusters by integrating three bio-replicons of single-cell RNA-sequencing (scRNA-seq) and six Stereo-seq sections. We used electron microscopy to visualize the ultracellular structures of the endostyle, performed in situ hybridization to verify marker gene expression, and conducted a comprehensive cross-species analysis of the various cell types. On the basis of these data, we identified a region containing hemolymphoid cells as a component of the endostyle. In addition, we defined cell types with specialized functions within the dense tissue region of the endostyle, such as a cell population that expresses markers for mechanically sensitive hair cells and neurons. Our research effectively revealed the cellular composition of the endostyle, a pharyngeal organ with exceptional importance in pharyngeal evolution, as well as provided insights into cellular foundations for complex organ evolution in vertebrates.

RESULTS

Pharynx-related diverse cellular composition of the ascidian endostyle

The pharynx is possessed by animals including hemichordates, basal chordates, and vertebrates (Fig. 1A). The hemichordates exhibit the simplest pharyngeal structure, which is characterized as pharyngeal slits originating from out-pockets of the endoderm. In vertebrates, besides the feeding and respiration tunnel, multiple derivative organs with versatile functions have developed from the pharyngeal arch, such as the auditory organ, thyroid, tonsil, and thymus. In the evolutionary continuum bridging invertebrates and vertebrates, the endostyle, a distinct pharyngeal organ, is a crucial component of the pharynx.

The transverse section of the endostyle consists of various regional components. A bilaterally symmetric groove is flanked by nine ciliated epithelial regions on the lateral sides, which are surrounded by a sinus region adjacent to the ventral blood vessel (VBV) (Fig. 1B). Although the gene expression pattern along the endostyle longitudinal axis showed a considerable homogeneity, several *Hox* genes were exceptions (*35*). To identify the cell components and potential vertebrate homologs in the endostyle, we established a spatially resolved atlas at single-cell resolution (Fig. 1B and fig. S1A). The scRNA-seq dataset, which consisted of 10,017 valid cells, was processed using the Seurat workflow. The cell compositions of the scRNA-seq dataset were preliminarily defined (fig. S1B). Multiple cell types, including immune cells, secretory epithelial cells, and blood cells, were discovered with cell cluster–specific markers (fig. S1C and data S1). For Stereo-seq, the expression profile, detected by spatially barcoded DNA nano-ball (DNB), was aligned with a single-stranded DNA dye-stained optical image captured in library construction (fig. S1D). Two cell segregation strategies, including square bin division and cell segregation, were applied to the dense and sparse tissue regions, respectively, based on different cell distribution characteristics. Using these methods, we segregated the DNB spots into cell units, which represent a compromised reflection of the real cell morphology (fig. S1E). Cell unit boundaries segregating the DNB spots on silicon chips were constructed as the backbone of

the atlas (fig. S1F). After cell segregation, 18,371 cell units were obtained in six Stereo-seq sections, and cell type annotation was performed (fig. S1G and data S2).

Next, we integrated the dataset from three biological replicates of scRNA-seq and Stereo-seq from six sections of 10-μm thickness (Fig. 1C and fig. S2) and annotated the cell types de novo based on cluster-specific gene expression (fig. S3A and data S3). The annotation was then projected onto the spatial landscape to create the cell atlas (Fig. 1B).

In the dense tissue regions of the atlas, regional cell types formed a clear bilaterally symmetric pattern, consistent with previous studies labeling with regional markers (*15*). The surrounding sinus region, filled with immune and blood cell clusters, is defined as the hemolymphoid region (HLR). To determine the batch effect of two technical replicates, we compared the cell numbers of each cell clusters between the scRNA-seq and spatial transcriptome in the atlas. The results indicated a general homogenous distribution of cell numbers among the clusters (fig. S3B). By performing integration process using Seurat, we observed that cells from scRNA-seq and Stereo-seq datasets are well co-embedded within the same uniform manifold approximation and projection (UMAP) space (fig. S3C). Cell annotations from each individual dataset were also projected onto the integrated space, showing that cells of the same cell type clumped together and cell groups with similar functional profiles were nearly located (fig. S3D), further supporting the similarity between scRNA-seq and spatial transcriptome datasets.

To discover possible biological processes associated with endostyle at the expression level, we selected highly variable genes in the integrated dataset (fig. S3E) and performed a functional enrichment analysis. Gene ontology (GO) enrichment analysis revealed functions that are highly variable in the endostyle, including neuron transduction (axon guidance and glial cell differentiation), immune process (adaptive immune system), muscle (muscle filament sliding), and pathways related to microtubule cytoskeleton or supramolecular fiber organization (fig. S3F). Additionally, we identified 13 gene modules with notable spatial autocorrelation in the Stereo-seq dataset using hotspot (Fig. 1D and fig. S4A) (*36*). These 13 gene modules were further classified into five categories based on their spatial distribution similarity (fig. S4B). We detected a greater variety of modules and genes in the sparse tissue region than in the dense tissue region, which may indicate the functional diversity of the sparse region (fig. S4C). Module score heatmaps revealed regional enrichment of potential functions, such as response to hormone function in the dorsal parts of the dense tissue region, regulation of the microtubule-based process in the ventral parts of the dense tissue region, and erythrocyte differentiation function in the HLR. These terms are consistent with the corresponding biological characteristics of the regions, including hormone metabolism, mucus net for filter feeding (*37*), and surrounding blood tissue (*22*) (Fig. 1E).

To identify potential conservation between the endostyle and vertebrate tissues, we reanalyzed the published single-cell dataset from zebrafish (*Danio rerio*) (*30*, *38–40*). The results indicated that the embryonic lineages of zebrafish, including the pharyngeal endoderm, paraxial mesoderm, neural crest, and pharyngeal arch, exhibited high similarity in expression compared with the endostyle of *S. clava* (Fig. 1F), suggesting substantial similarity between the ectoderm-derived tissues from zebrafish and the endostyle. Moreover, a comparison between the endostyle and adult zebrafish tissues revealed high transcriptomic similarity between the endostyle and the thymus,
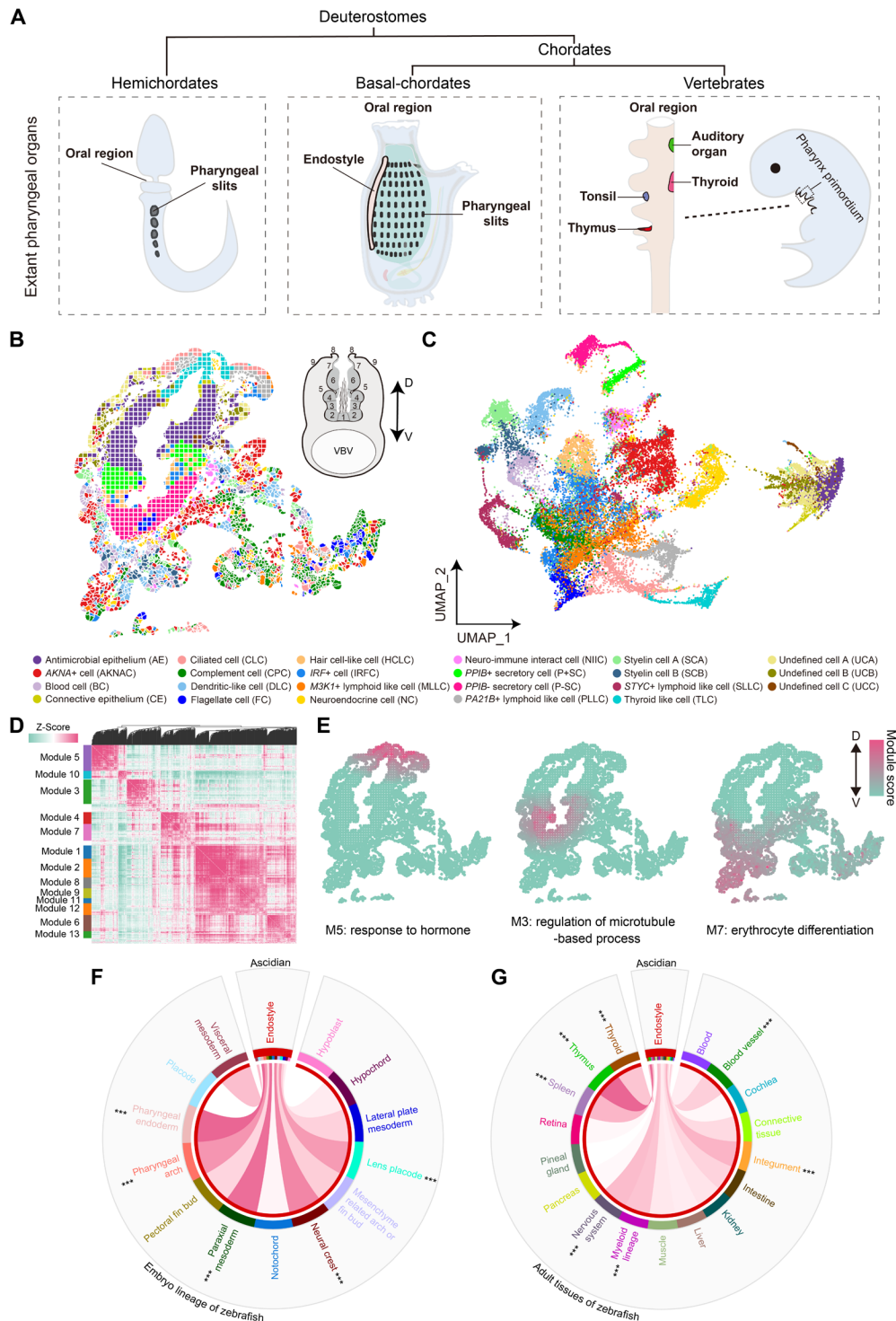
Fig. 1. Cell atlas construction for the endostyle of the adult basal chordate, *S. clava*. (**A**) Extant pharyngeal organs and apparatus in three deuterostomes categories, including hemichordates, basal chordates, and vertebrates. Left: Schematic of the pharyngeal organ, pharyngeal slits in hemichordates. Middle: Schematic of the pharyngeal organ in basal chordates, represented by an ascidian, including the pharyngeal slits (on the pharyngeal sac) and the endostyle. Right: Schematic of the pharyngeal primordium in vertebrates. Pharyngeal organs (tonsil, thymus, auditory organ, and thyroid) have primordium in the pharynx arch. (**B**) Spatial atlas showing the distribution of cell units on the transverse section colored by cell annotation. The upper right schematic shows the transverse structure of the endostyle, consisting of bilaterally symmetric zones 1 to 9 with a ciliated groove in the center, surrounded by the sinus region. The subendostylar vessel is located on the ventral side. D, dorsal; V, ventral. (**C**) UMAP of all 28,388 scRNA-seq and Stereo-seq cell units, colored with cell annotation. (**D**) Heatmap showing gene modules based on spatial autocorrelation. (**E**) Spatial visualization of selected modules in the heatmap of the gene modules. Three modules labeled with the enriched gene function. (**F** and **G**) Circos plots showing transcriptomic similarity between the endostyle and the embryo lineage (F) or adult tissue (G) of zebrafish (*D. rerio*). Here, *** indicates the top 40% similarity.

thyroid, myeloid lineage, and nervous system from zebrafish (Fig. 1G), suggesting a similarity between the endostyle and immune/blood system–related tissues in zebrafish at the transcriptome level.

The HLR provides a niche harboring blood and immune cells
To identify the cell composition of the HLR in the spatially resolved single-cell atlas, we selected cell clusters dominated by cell units located in the sparse region as shown in the Stereo-seq results (fig. S5A). All scRNA-seq and Stereo-seq cell units within the sparse regions of these clusters were included and visualized for further analysis (Fig. 2A). Cluster-specific markers supported the cell type annotations, with most clusters being related to immune/blood function (Fig. 2B). The expression of markers, complement C3 (*C3*) and hemocyanin 2 (*HCY2*), were verified by in situ hybridization (fig. S5B).

Morphological observation of the HLR was conducted using confocal and transmission electron microscopy. Two adjacent tissue components, the HLR and VBV, were observed in addition to the dense tissue region in the transverse section (Fig. 2C). The HLR predominantly contained a highly developed extracellular matrix, forming cave-like structures within the region. A large number of cells were scattered throughout the extracellular matrix space (fig. S5C). We systematically compared the morphology of cells in the HLR with that of blood cells reported in *Ciona robusta* (*41*). Most of the blood cell types reported in *C. robusta* were also found in the HLR of the endostyle (fig. S5D).

To verify the divergence of cell clusters in the HLR and peripheral blood system, in situ hybridization was used to assess the expression of *complement C3*, a marker gene for the complement cell cluster in the HLR. We calculated the ratios of cells expressing *complement C3* in the HLR to those in the peripheral blood tissues (Fig. 2D). There is a notably larger ratio of *C3*+ cells in the HLR compared to that in the peripheral blood. This finding indicates an enrichment of *C3*+ cells within the HLR relative to the peripheral blood, suggesting enhanced immune activity within the HLR (fig. S5, E and F).

Next, we conducted a pseudotime analysis on cell clusters within the HLR. Cells annotated as "stem cell" within the single-cell dataset were assigned as the root of the pseudotime trajectory, representing cells with distinct expression of proliferation markers (fig. S1C). The results indicated that the complement cell cluster and *IRF*+ cell (IRFC) cluster had the lowest pseudotime value, indicating a low differentiation status. Conversely, clusters including the dendritic-like cell, blood cell, and *AKNA*+ cell corresponded to high pseudotime values and were associated with a high differentiation status (Fig. 2E and fig. S5G). Additionally, RNA velocity analysis using dynamo also revealed a similar transition trend as the above results. The general trend of the streamlines exhibited a pattern starting from the complement cell cluster and the IRFC cluster (fig. S5H). Furthermore, streamlines of RNA velocity were projected onto the spatial landscape and showed a trend starting from the ventral part and ending at the dorsal part (Fig. 2F).

To infer the potential stemness of the blood/immune lineages, we visualized the expression of categories of stemness marker genes (Fig. 2G). The IRFC cluster showed intense expression of genes associated with hematopoietic stem cell differentiation and hematopoietic progenitor cell differentiation. Cell populations, including *M3K1*+ lymphoid like, *STYC*+ lymphoid like, and *PA21B*+ lymphoid like cell clusters, exhibited enriched expression of genes in the lymphocyte differentiation category. These findings suggest that cell clusters with lower pseudotime values or immature status have higher stem or proliferation capacity. Additionally, the results indicate a nonuniformity of cell distribution. Cells at the beginning of the developmental trajectory tended to be located in the ventral region of the HLR, while cells at the dorsal part were more likely to be found at the end of the developmental trajectory.

On the basis of the morphological and expressional evidence, we hypothesized that the HLR may serve as a potential hub for the generation of blood and immune cells, potentially acting as a supplementary source to the peripheral blood. Consequently, we performed scRNA-seq on the HLR tissue and blood from *S. clava* (fig. S6A). The resulting datasets from the two samples were merged, resulting in the identification of 12 cell clusters (Fig. 2H and data S4). Gene expression scoring revealed that cell clusters C2, C9, and C11 uniquely expressed hematopoietic stem cell–related pathways, along with myeloid/lymphoid gene pathways. Furthermore, cells from the HLR demonstrated more intense expression patterns in gene pathways, including hematopoietic progenitor cell differentiation, myeloid leukocyte/lymphoid progenitor cell differentiation, and repressive genes related to differentiation (fig. S6B). To further demonstrate the differential characteristics among C2, C9, and C11, a pseudotime trajectory analysis revealed divergent differential statuses of cells from different clusters and samples (Fig. 2I and fig. S6C). Moreover, the distinct expression of transcription factors was observed in C11, including *early growth response factor 1* (*EGR1*; known for its role in maintaining hematopoietic stem cell homeostasis in vertebrates) (*42*), *catenin beta 1* (*CTNNB1*; known for its role in expanding the hematopoietic stem cell pool in vertebrate through overexpression) (*43*), *claudin 7* (*CLDN7*), and goosecoid homeobox 2 (*GSC2*; associated with embryo head and pharynx organization) (fig. S6D) (*44*). Consequently, cells in C11 were considered to be in the stem developmental stage. The C11 cell cluster, located at one end of the trajectory, predominantly consisted of cells originating from the HLR (Fig. 2I and fig. S6, E and F). These in silico findings primarily support the differential trajectory from a cluster of the HLR cells that potentially contribute to the blood system, which further supports our hypothesis.

These results suggest that the HLR is a potential hub for the generation and maintenance of blood and immune cells. It may represent a primordial hemolymphoid system that shares homology with the blood and immune systems of vertebrate species. Furthermore, the HLR may act as a stem cell center, contributing to the production of functional cells within the immune and blood systems, potentially playing a crucial role in replenishing the blood supply (Fig. 2J).

Hair cell–like cells have transcriptomic similarity with hair cells in vertebrates
Hair cell–like cells (HCLCs) were identified in zone 3 of the dense tissue region spatially mixed with the IRFCs (Fig. 3A). The HCLCs exhibited expression of several markers, including *protein tyrosine phosphatase receptor type Q* (*PTPRQ*), *usherin* (*USH2A*), *whirlin* (*WHRN*), and *adhesion G protein–coupled receptor V1* (*ADGRV1*), which are important factors in maintaining functionally active hair bundles and stereocilia in hair cells of the vertebrate (Fig. 3B and fig. S7A) (*45–49*). Additionally, neuronal markers such as *neuronal cell adhesion molecule* (*NRCAM*), *semaphorin-1A* (*SEM1A*), and *solute carrier family 26 member 4* (*SLC26A4*) were detected. *NRCAM* is essential for the development of spiral ganglion neurites (*50*). *SEM1A* plays an important role in growth cone guidance (*51*). *SLC26A4* is an electroneutral sodium-independent transporter for chloride and
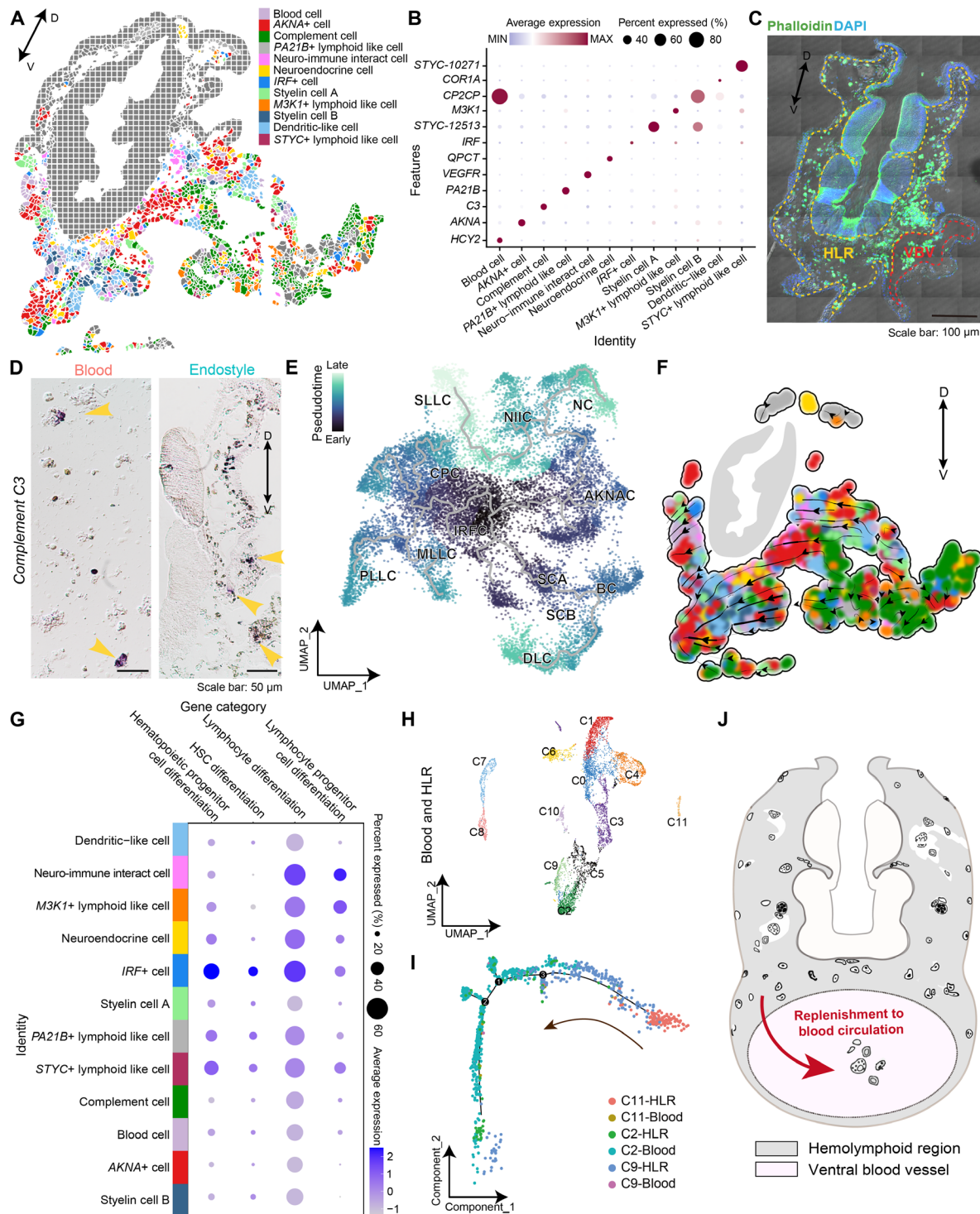
Fig. 2. The HLR provides a lymphoid-like niche for blood/immune cell lineages. (**A**) Spatial visualization of the cell clusters in the HLR. (**B**) Bubble plot showing the expression of marker genes in the indicated cell types. (**C**) Structure and location of the HLR based on confocal microscopy. Cells in the HLR were scattered. (**D**) In situ hybridization showing the expression pattern of *complement C3*. Left: Results on the peripheral blood cell. Right: Results on the endostyle cryosection. Yellow arrows point at signals. (**E**) Pseudotime trajectory analysis showing trajectories among cell clusters in the HLR based on the UMAP. Cells are colored with pseudotime values on the UMAP. (**F**) RNA velocity streamline plots showing the predicted trajectory of the cell cluster transition in the HLR of the endostyle. Cells are plotted based on their location in the spatial atlas. (**G**) Bubble plot showing the expression intensity of the HLR cell clusters for the gene categories. (**H**) UMAP showing the merged scRNA-seq from the blood and HLR samples. Cells were clustered into 12 clusters. Cell identities were denoted with cluster IDs starting with C. (**I**) Cell trajectory constructed with cells from C2, C9, and C11. Cells on the trajectory were colored by six types separated by cell clusters and sample origin, blood, or the HLR. (**J**) Model of the immune and blood system in the HLR.
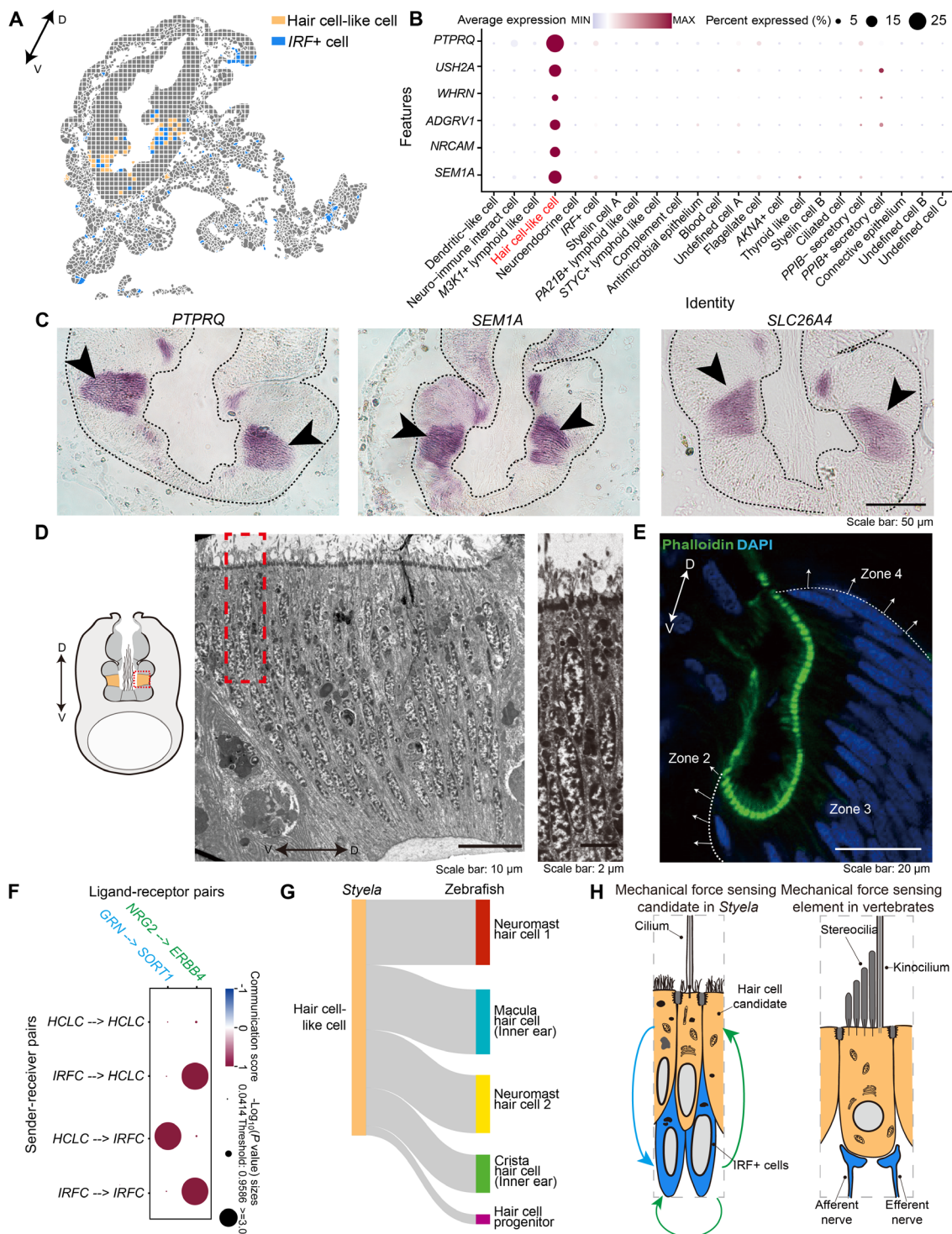
Fig. 3. The HCLC cluster defined in zone 3 has transcriptomic similarity with the acoustico-lateralis system. (**A**) Spatial visualization of the HCLC cluster and IRFC clusters. (**B**) Bubble plot showing the expression of marker genes in the indicated cell types. (**C**) In situ hybridization showing the expression of *PTPRQ*, *SEM1A*, and *SLC26A4*. Black arrows indicate the signals. (**D**) Left: Schematic for the location of zone 3 in the endostyle transverse section (orange region). Middle: Detailed visualization of the subcellular structure in the region of the red rectangle in the left figure. Right: A more detailed visualization of the red rectangle region in the middle figure. (**E**) Confocal microscopy showing the zone 3 region on phalloidin (green)–and DAPI (blue)–stained cryosections. Zone 2 to 3 and zone 3 to 4 tissue boundaries are indicated by white dashed lines. (**F**) Bubble plot showing communication scores of ligand-receptor pairs between sender and receiver cell populations. (**G**) Cross-species comparison between the HCLC cluster and cell populations of the acoustic-lateralis system in zebrafish. The thickness of the connections indicates the correlation value of the connected cell clusters. (**H**) Left: Schematic for hypothetical mechanical force-sensing candidate in the endostyle. The blue arrow and green arrows correspond to the ligand-receptor relationship, *GRN* to *SORT1* and *NRG2* to *ERBB4*, respectively. Right: Schematic of the mechanical force sensing element in vertebrates.

iodide, associated with deafness and Pendred syndrome (*52*). The expression of *PTPRQ*, *SEM1A*, and *SLC26A4* was verified with in situ hybridization and showed unique signals in zone 3 (Fig. 3C and fig. S7B). Distinct IRFC markers compared to the HCLC cluster, including *gelsolin* (*GSN*), *actinin alpha 2* (*ACTN2*), *intermediated filament* (*IF*), *plastin 3* (*PLS3*), and *annexin A7* (*AXNA7*), are important cytoskeletal components, suggesting potential correlations with morphology architecture and mechanical support (fig. S7C).

To describe the morphological features and relationship between the HCLC and IRFC populations, we performed visualization by confocal and transmission electron microscopy. Transmission electron microscopy results exhibited that the cells in zone 3 manifested a slender, shuttle-like shape, closely interleaving with one another (Fig. 3D and fig. S7, D and E). Notably, long F-actin cilia were distributed on the apical surface, whereas densely and tightly arranged cell nuclei were distributed underneath (Fig. 3E). The HCLC population with specialized apical cilia were connected tightly to basal cells, which were thought to be IRFC. The association between two cell populations was supported by a ligand-receptor analysis, which revealed interactions from *neuregulin 2* (*NRG2*) to *Erb-B2 receptor tyrosine kinase 4* (*ERBB4*) and *granulin* (*GRN*) to *sortilin1* (*SORT1*). Both interactions are important cell-cell interactions in the nervous system (Fig. 3F and fig. S7F) (*53*, *54*). To determine the evolutionary relationship of the HCLC in zone 3 with organs/tissues in vertebrates, we performed a cross-species comparison analysis. Spearman correlation between HCLC cells and zebrafish tissues was calculated based on the average expression of transcription factors. The results showed that the HCLC cluster had a higher similarity to the zebrafish cochlea compared to other tissues (fig. S7G). Consequently, we conducted a cross-species comparison between the HCLC cluster to hair cell–related tissues in zebrafish (*39*). The results indicated that the HCLC cluster in zone 3 exhibits a high similarity with hair cell populations and hair cell progenitors of the acoustico-lateralis system in zebrafish (Fig. 3G).

On the basis of the gene expression and morphological results, we hypothesized a model with potential mechanical force-sensing capacity. In contrast with vertebrate mechanical force sensing elements consisting of hair cells and synapses, the HCLC and IRFC populations in the *Styela* endostyle form tightly interlaced layers. These cell clusters are distributed from the apical ciliated surface to the basal membrane. This arrangement may facilitate the transmission capacity of nerve electrical signals through tight junctions in the absence of a synaptic structure (Fig. 3H).

The thyroid-equivalent element has cells of different maturation status

In the spatial atlas, a unique cluster of thyroid-like cells was observed in zone 7 of the endostyle (Fig. 4A). This cell cluster specifically expressed key markers associated with thyroid hormone synthesis, including *thyroid peroxidase* (*TPO*) (*14*) and *dual oxidase 1* (*DUOX1*) (*55*). Additionally, we detected the expression of *Intelectin 1* (*ITLN1*) (*56*, *57*) and *milk fat globule membrane* (*MFGM*) (*58*). These findings suggest a potential immune capacity and a role in maintaining epithelial tissue homeostasis for this group of cells (Fig. 4B and fig. S8A). The expression of an important marker, *TPO*, was verified by in situ hybridization, which is consistent with findings from a previous study (*14*) (Fig. 4C and fig. S8B). Ultrastructural observations revealed a single layer of nonciliated cuboidal epithelial cells (Fig. 4D). To verify the function of thyroid-like cells, we estimated the module score of

pathways related to the thyroid hormone, serving as an important indicator of thyroid activity. The expression scores for the functional pathways of the thyroid hormone generation and metabolic processes were mapped onto the spatial landscape, revealing intense signals in zone 7 (Fig. 4E and fig. S8C).

The thyroid-like cell cluster is located in the thyroid homologous region of the ascidian endostyle, which has been reported to be zone 7 (*14*). To verify that the transcriptional data obtained by Stereo-seq and scRNA-seq could support the reported homologous relationship between the endostyle regions and the thyroid gland (*19*), we conducted a cross-species comparison between the thyroid-like cell cluster of the endostyle and developmental lineages/organs of zebrafish. The result indicated that the thyroid-like cell cluster exhibited the highest similarity with the thyroid, liver, and pharyngeal endoderm (Fig. 4F). On the basis of the results of marker expression, in situ hybridization verification, and cross-species comparison, we defined the thyroid-like cell cluster as a cell population with thyroid-equivalent features in the endostyle, which further confirmed the reliability of the cell atlas.

To explore the potential developmental status of the thyroid-like cell cluster, a developmental trajectory assignment was performed. We constructed a cell trajectory for the pharyngeal endoderm and the adult thyroid of zebrafish and identified important genes for the trajectory of thyroid gland development. These genes were enriched in functional pathways, including the extracellular matrix–receptor interaction (fig. S8, D to F). The expression of vital thyroid developmental gene homologs within the thyroid-like cell cluster of the endostyle was visualized (Fig. 4G). Pseudotime values were assigned to the thyroid-like cells based on the value of the most highly expressed gene. The frequency distribution revealed the different maturation status of the cells in the thyroid-like cell cluster (Fig. 4H). In addition, we projected pseudotime values onto the cell atlas of the region (Fig. 4I). The pseudotime distribution was calculated based on a curve fit with cell units from a bilateral-symmetric region. Analysis of the evenly segmented curve revealed enriched and highly differentiated cells in the dorsal region of the thyroid-like cell cluster (Fig. 4J). These results suggest a renewal trend or a differentiation trajectory of the cluster, from ventral to dorsal.

Cell populations of other dense tissue regions revealed the endostyle transcriptome diversity

In addition to the zones described above, we conducted a detailed analysis of the cellular components and gene expression in the dense tissue regions (fig. S9, A to C). Zones 1 and 8 were characterized by flagellate cell and ciliated cell clusters, both of which exhibited similar expression profiles associated with the ciliated cell structures. Zones 2 and 4 displayed a similar morphology and gene expression pattern, and the two cell clusters were defined as *peptidylprolyl isomerase B* (*PPIB*)− and *PPIB*+ secretory cells. The cell components in zone 6 were defined as the antimicrobial epithelium based on the *intelectin 2* (*ITLN2*) expression, whereas information on zone 5 was still lacking because of its small area as a connective tissue.

Combined with the cellular components of the HLR (Fig. 2A) and previous studies in colonial ascidians (*22*, *23*), the regional cellular components of the endostyle were revealed. The cell populations defined in this atlas not only are consistent with the previous understanding of the endostyle regions (*15*) but also provide insight into the evolution of pharyngeal organs.
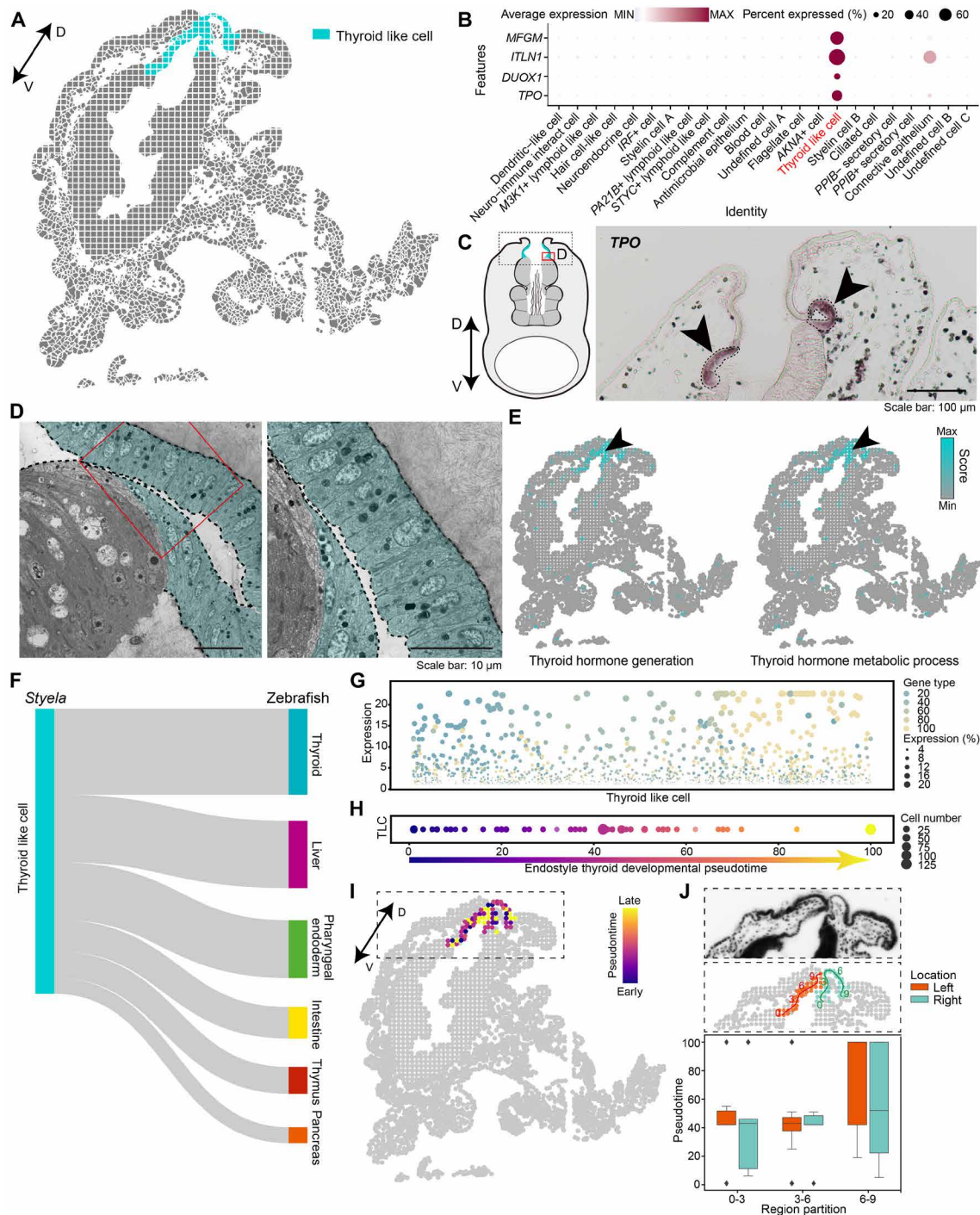
Fig. 4. Homologous cell cluster to the thyroid gland in the endostyle cell atlas. (**A**) Spatial visualization of the thyroid-like cell cluster. (**B**) Bubble plot showing the expression of marker genes for the thyroid-like cell cluster. (**C**) Left: Schematic showing the location of the thyroid-like cell in the transverse section. Right: In situ hybridization showing the expression pattern of *TPO*. Black arrows indicate the signals. The field is the left black rectangular region. (**D**) Left: Transmission electron microscope showing ultracellular structures of cells in the thyroid-like cell cluster (blue masked). Right: Detailed visualization of the red rectangle region. (**E**) Heatmap showing the spatial expression of functional pathways. (**F**) Cross-species comparison between the thyroid-like cell cluster in the endostyle and tissue/organs in zebrafish. The thickness of the connections indicates the correlation value of the connected cell clusters. (**G**) Expression of genes with trajectory information (fig. S8, D and E). Each bubble represents the expression level of a specific gene in a cell. Bubbles are colored with the trajectory location of the gene. The size of bubbles indicates the expression level. Cells in the thyroid-like cell cluster were ranked on the horizontal axis based on the trajectory type of the top expressed genes. (**H**) Distribution of cells on the pseudotime trajectory defined by the trajectory type of their top expressed genes. The bubble size indicates the number of cells for a given pseudotime status. (**I**) Visualization of the trajectory status. (**J**) Top: Single-stranded DNA image for the rectangular region in (I). Middle: Track line showing the epithelium extending from the ventral region to the dorsal region, divided into nine segments evenly. The left and right zone 7 were labeled with different colors. Bottom: Bar plot showing the statistics for the pseudotime values for cells on nine segments.

DISCUSSION

The evolution and origination of complex organs have long been an intriguing topic in evo-devo ever since Charles Darwin emphasized the importance of embryology for evolution (*59*). With the advent of multi-omics techniques such as single-cell and spatial transcriptomics, the identification of cell type and cell-to-cell interactions from an evo-devo perspective has advanced rapidly, thus providing opportunities to study the evolution of organs in animals of evolutionary importance (*27*). Using cutting-edge Stereo-seq and scRNA-seq, we constructed the first spatially resolved cell atlas at single-cell resolution for the endostyle, an important pharyngeal organ in basal chordates. This study provides the first spatial resolved cell atlas for a marine animal. This is important for resolving the cell composition of organs with complex structural organization and tissue arrangement patterns compared with previous single-cell datasets on marine animal samples (*28*, *60–62*).

Marine organisms exhibit remarkable variability among different species. The sample preparation method for spatial transcriptomics was optimized based on previous results obtained using fixed and embedded methods for in situ hybridization and single-molecule RNA fluorescence in situ hybridization (*63–65*). To ensure the reproducibility of the atlas, we combined three replicates of scRNA-seq and six sections of Stereo-seq. These data exhibited similar expression patterns, except for the statistical differences between the two techniques (fig. S3, B to D). We analyzed data from replicates and technical batches not only separately but also in combination, which supported the reliability of our conclusion. To improve the alignment efficiency of the transcriptomic data, we adjusted the gene structural annotation and replenished the alternative splicing information based on a previous publication (*33*), which largely improved the mapping results of the endostyle datasets. The updated version of the gene structural annotation was released on EDomics, a multi-omics database for animal evo-devo (*66*). scRNA-seq and Stereo-seq identified 18,685 and 11,912 genes, respectively, compared with the bulk RNA-sequencing results (11,083 genes from bulk RNA-sequencing) (*35*) (fig. S3, G and H, and data S5), which, to some extent, exhibited differences among the techniques. Although spatial transcriptomics still has limited capture efficiency compared with scRNA-seq, it has irreplaceable advantages, including providing detailed spatial information and exhibiting less bias in the capture of rare cell types and signatures. Our results provided a detailed reference for the de novo construction of a cell atlas in marine species with limited previous research.

We performed a comprehensive comparison between the endostyle and zebrafish tissues. The results indicated a high expression similarity and potential homology. The endostyle exhibited a high transcriptome similarity to the myeloid lineage, which is consistent with the hematopoietic niche in the endostyle of colonial ascidians (*22*). On the basis of the results, we fully characterized the enclosed region surrounded by the epithelia extending from zone 9 and other dense tissue regions, which were previously defined as a blood sinus region. This region, which was defined as the HLR in this research, was characterized by the presence of blood and immune cell populations, potential developmental lineages, and highly developed cave-like extracellular matrix structures. Our findings represent a remarkable advance over previous discoveries of the hematopoietic niche (*22*) and stem cell characteristics (*23*). The highly developed extracellular matrix may provide a favorable microenvironment for cell proliferation and lineage development, although further experimental verification is needed. An

HLR-like region is likely to be present in a common ancestor of vertebrates and tunicates. This region may evolve into the HLR region in tunicates and the blood and lymphoid system, including organs like the thymus and tonsil organs in vertebrates (Fig. 5, A and B). However, the mixed blood and immune system in the HLR is different from the highly developed erythroid and myeloid cell lineage derived from vertebrate hematopoietic stem cells. A more precise and robust comparison remains to be investigated.

In the comprehensive cross-species comparison, besides the expression similarity between the ascidian endostyle and the zebrafish thyroid or pharyngeal endoderm, cell types in the endostyle also exhibited high transcriptome similarity to the placode and cochlea, which are ectoderm-origin tissue/organs. We assessed the expression of a set of markers with functions such as maintenance of hair cell architecture and neuron signaling. These marker genes are homologous to those in hair cells of the vertebrate acoustic-lateralis system. Morphological and in silico evidence supported the existence of the HCLC cluster in zone 3. Recent studies on the endostyle of both *Oikopleura* (*67*) and *Ciona* (*68*) revealed that, even when the Nkx homeobox-1 (Nkx2.1) gene was deleted, the supporting regions still developed normally. This result suggests that the supporting components in the endostyle are established under the control of transcription factors other than *Nkx2.1*. These data, together with our results, suggest the potential existence of a non-endoderm component in the endostyle, although the endostyle has long been believed to originate from the ventral pharyngeal endoderm (*11*, *13*). While the hair cells of the acoustic-lateralis system are derived from ectodermal placodes, which are thought to be homologous to the siphon in ascidians (*69*), we currently lack detailed evidence for the developmental contribution of the placode or ectoderm to the pharyngeal region in basal chordates. Further studies are needed to address these unresolved issues, including tracing the spatially resolved developmental lineages that contribute to the formation of the adult endostyle and determining the function of the cell populations identified in this study.

The evolutionary origins of an important endocrine endoderm gland, the thyroid organ, can be traced back to cell clusters in the endostyle, which are believed to concentrate iodine and secrete thyroid hormone. In the ascidian endostyle, zone 7 has been confirmed as the thyroid equivalent region (*14*). Here, we identified a cell cluster dominating zone 7 and defined it as the thyroid-like cell cluster based on multiple pieces of evidence. When mapping the pseudotime trajectory with the developmental trajectory of the thyroid gland in zebrafish, we observed a diverse developmental status within the thyroid-like cell cluster. Cells with different developing status coexist in the region, and statistical analysis revealed a transitional maturation tendency from the ventral to the dorsal region. This suggests that the thyroid-like cell cluster may be capable of replenishing cells. This feature of the cluster in ascidians contrasts with the limited renewal capacity of the adult vertebrates' thyroid gland (*70*) and might provide a model for thyroid research.

Here, aiming to elucidate the evolution of multi-germ layered complexed organs in the vertebrate pharynx, we generated a high-quality dataset for a pharyngeal organ that exclusively exists in basal chordates. We found that the endostyle features a diverse cellular composition and exhibits potential functional similarities with organs in vertebrate pharynx. The HLR, which contains immune/blood cell lineage and developmental trajectory, may share common ancestry with vertebrate immune organs such as the tonsil and thymus. Furthermore, cell populations in zone 3 are likely homolog candidates to
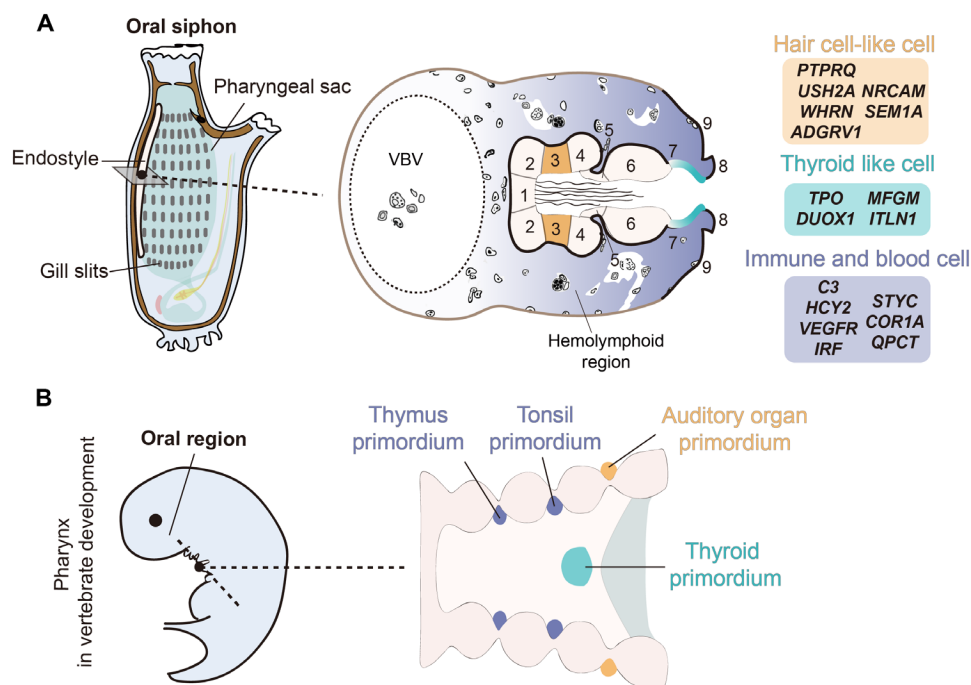
Fig. 5. Comparison of the pharynx organ between the endostyle in ascidian and developing pharynx in vertebrates. (**A**) Left: Schematic showing the location of the endostyle in the adult ascidian. Right: Transverse section of the endostyle showing the diverse cellular composition of the endostyle, including the immune and blood cells, HCLC, and the thyroid-like cell cluster, which correspond to the major cell populations for the HLR, zone 3, and zone 7 in the schematic. The gradient color in three regions shows their distinct developmental status within the tissue. Distinct marker genes are labeled. (**B**) Left: Schematic showing the morphology location for the developing pharynx in vertebrates. Right: Coronal plane of the developing pharynx showing the multiple organ primordium, including the thyroid, auditory organ, tonsil, and thymus.

the auditory/lateralis system in vertebrates. We also found that the thyroid-like cell cluster corresponds to the vertebrate thyroid gland and showed the coexistence of cells with different maturation states. These findings support the idea that the endostyle is a key organ in pharyngeal evolution and that the endostyle may have function similarities to those of vertebrate pharyngeal organs. The involvement of multiple cell components in the pharynx of the basal chordate implicates a cellular basis for the emergence of multiple organs in vertebrates and provides insight into the evolutionary footprint of pharyngeal architecture.

MATERIALS AND METHODS
Animals and sample collection
Adults *S. clava* were purchased from the Xunshan Company in Weihai, China, and temporarily preserved in a circulating aquarium system within the laboratory at 18°C, equipped with air pumps. The endostyle of healthy individuals was gently dissected using disinfected instruments as the previous description (*35*). All guidelines for animal experiments were approved by the Institutional Animal Care and Use Committee of Ocean University of China (OUC-IACUC), with approval numbers 2020-0032-0517 and 2023-0032-0039.

Single-cell isolation and scRNA-seq library construction for the endostyle, HLR, and blood
Tissues from five adult individuals were used for single-cell isolation and library construction. Among them, three endostyles from different animals were used separately as biological replicates for the endostyle library, one endostyle tissue for the HLR library, and one intact heart tissue for the blood library.

To obtain a cell suspension of the whole endostyle, fresh endostyle was transferred into clean vials and immediately dissociated with 0.2% trypsin in a solvent of Ca^{2+}-free artificial seawater with 5 mM EGTA. The tissue was consistently scissored and pipetted for 15 to 20 min on ice to ensure complete dissociation. To obtain a cell suspension from the HLR, we conducted a two-step digestion. First, fresh endostyle tissue was cut into segments, transferred into a clean vial, and immediately dissociated using 0.2% trypsin and 0.2% collagenase in the solvent mentioned above. The tissue was pipetted (without scissored) for 40 to 50 min on ice to completely remove the dense tissue, retaining the intact sparse tissue. Then, the sparse tissue was transferred to Ca^{2+}-free artificial seawater, rinsed, and dissociated using 2.5% trypsin in the same solvent, followed by pipetting on ice for 5 to 10 min for complete dissociation. Digestion was inhibited in all samples using 0.2% bovine serum albumin (BSA) in Ca^{2+}-free artificial seawater. The suspension was filtered into a vial with a 35-μm cell filter cap. Cells were collected via centrifugation at 4°C, 500*g* for 2 to 5 min and resuspended in ice-cold Ca^{2+}-free artificial seawater containing 0.1% BSA. To obtain fresh blood cells, the intact heart region of an adult individual was carefully exposed. The beating heart (typically observed to beat a few times per minute in *S. clava*) was pierced with a syringe, and the blood was immediately drawn and transferred to glass tubes.

For the whole endostyle single-cell library, single-cell gel beads in the emulsion were generated using a Chromium Controller instrument (10x Genomics). Sequencing libraries were prepared using

Chromium Single-Cell 3′ Reagent Kits (10x Genomics) according to the manufacturer's instructions. After performing cleanup using a SPRIselect Reagent Kit, the libraries were constructed through the following steps: fragmentation, end repair, A-tailing, SPRIselect cleanup, adapter ligation, SPRIselect cleanup, sample index polymerase chain reaction (PCR), and SPRIselect size selection. The libraries were subsequently sequenced using the Illumina NovaSeq system.

For the HLR and blood single-cell libraries, cell number and viability were estimated using a fluorescence cell analyzer (Countstar Rigel S2) with AO/PI reagent. Libraries were prepared using SeekOne MM Single Cell 3′ Library Preparation Kit (SeekGene, Cat# SO01V3.1). An appropriate number of cells were loaded into the flow channel of the SeekOne MM chip and allowed cells to settle by gravity in microwells. Unsettled cells were removed, and cell barcoded magnetic beads (CBBs) were pipetted into the flow channel and also allowed to settle in microwells under a magnetic field. Excess CBBs were rinsed out, and cells in the MM chip were lysed to release and capture RNA. Reverse transcription was performed at 37°C for 30 min. Barcoded cDNA on the CBBs was hybridized with a random primer, which had read 2 SeqPrimer sequences at the 5′ end and could extend to form the second strand with a 3′ end cell barcode. The resulting second-strand DNA was denatured off the CBBs, purified, and amplified in a PCR. The cDNA product was cleaned and added to an adapter and sample index. Finally, the indexed sequencing libraries were cleaned with SPRI beads, quantified, and sequenced on the Illumina NovaSeq.

Data analysis for scRNA-seq for the endostyle, HLR, and blood

For the scRNA-seq results obtained from the 10x Genomics Chromium platform, we processed the three endostyle samples separately. BCL files were converted to FASTQ format using bcl2fastq v1.8.4 from Illumina. Cell Ranger v3.0.2 was used to generate gene-barcode matrices for each sample, using the mkref and count functions with default parameters. Previously published genome sequences (*33*) were used as a reference. Next, Seurat objects were created using Seurat 4.0 (*71*) based on the gene expression matrix, followed by several preprocessing steps, including filtering, integration, and clustering. Initial filtering was based on gene expression levels, with nCounts ranging from 0 to 5000 and nFeature ranging from 300 to 1200. Cells with high mitochondrial expression percentages, defined as the mitochondrial gene counts divided by the total cell count, were excluded using specific thresholds (30, 45, and 35 for three samples, respectively). To identify and eliminate potential doublets, we normalized the data using the SCTransform function, with parameters set to 2000 variable features and the qpoisson method. We processed the data using DoubletFinder v3.0 (*72*), using parameters of 0.075 for the assumed doublet formation rate and 15 for principal component. We used SoupX (*73*) to remove ambient RNA that may have come from the buffer system. Subsequently, each high-quality dataset underwent another SCTransform normalization and scaling under a reshaped dimensionality. Finally, the three processed single-cell datasets were integrated using the IntegrateData function to remove batch effects and co-embed them into the same space. The integrated dataset was subjected to principal components analysis (PCA) and further clustered with a resolution of 0.3 based on the precalculated shared nearest neighbors. Marker genes for each cell cluster were identified using the FindAllMarkers function.

For scRNA-seq data from the HLR and blood samples, which were sequenced using the SeekOne platform, we processed the paired-end FASTQ format files using the seeksoultools rna module (http://seeksoul.seekgene.com/en/v1.2.0/index.html) to generate gene-barcode matrices. The FASTQ files of HLR were down-sampled by SEQTK with seed 100 and a ratio of 0.2 to ensure balanced library depth for the two samples before the processing with the seeksoultools. We used Seurat 4.0 (*71*) for cell cluster identification, following the same steps as previously mentioned. For the HLR sample, filtering thresholds were set at nCounts ranging from 100 to 3000, nFeature ranging from 50 to 1000, and the lowest mitochondrial expression percent of 40. For the blood sample, filtering thresholds were set as nCounts ranging from 0 to 1000, nFeature ranging from 10 to 500, and a lowest mitochondrial expression percent of 50. The filtered data from the HLR and blood were processed with SCTransform separately and then merged using the Seurat merge function. Clustering was performed on the merged object with a resolution of 0.3. Subsequently, the FindMarkers function was executed to identify marker genes for C2, C9, and C11 clusters. To investigate the potential transitional relationship between the two samples, we constructed a cell trajectory using Monocle 2 (*74*) and focused on cell clusters C2, C9, and C11. Highly expressed genes (log_avg2FC ≥ 0.25) for these clusters were estimated using the differentialGeneTest function to identify genes that change along pseudotime.

Stereo-seq library preparation and sequencing

For Stereo-seq cryosection, an endostyle sample from a single adult animal was isolated and immediately immersed in methane precooled to −20°C for 5 min. The rapidly fixed tissue was snap-frozen in Tissue-Tek OCT (4583, Sakura, Torrance, CA) with liquid nitrogen prechilled isopentane. It was then transferred to a −80°C refrigerator for storage before cryosection. Before sectioning, the OCT-embedded tissue was placed in a −20°C freezing microtome for 30 min for equilibration. Cryosections were collected at 10-μm intervals along the transverse plane using a Leica CM1950 cryostat.

Stereo-seq experiments were performed as previously described by Chen *et al.* (*31*). Briefly, the Stereo-seq chip was washed with NF-H_2O supplemented with ribonuclease (RNase) inhibitor (0.05 U/μl) (NEB, M0314L). Six sections from the anterior half of the endostyle were mounted onto a single Stereo-seq chip with a 3-min incubation on a thermocycler adaptor at 37°C. Subsequently, sections were fixed using methanol and incubated at −20°C for 40 min. Sections on the chip were then stained with a nucleic acid dye (Thermo Fisher Scientific, Q10212) for single-stranded DNA imaging and scanned using the Motic Custom PA53 FS6 microscope system to capture the full-structure images (20× objective).

Tissue sections were permeabilized in 0.1% pepsin (Sigma-Aldrich, P7000) with a 0.01 M HCl buffer and incubated at 37°C for 6 min. The RNA released and captured on the Stereo-seq chip was reverse-transcribed overnight at 42°C using SuperScript II Reverse Transcription (RT) Mix. Sections were washed twice with a Wash Buffer and digested with Tissue Removal buffer (10 mM tris-HCl, 25 mM EDTA, 100 mM NaCl, 0.5% SDS) at 37°C for 30 min. The remaining RT products were collected and amplified using KAPA HiFi Hotstart Ready Mix (Roche, KK2602) with 0.8 μM cDNA-PCR primer. Sequencing libraries were prepared with PCR products following the steps that included concentration quantification, DNA fragmentation, PCR amplification, and purification. The purified PCR products were used to generate DNBs and were finally sequenced on the

MGI DNBSEQ-T1 sequencer [35 base pairs (bp) for Read1, 100 bp for Read2].

Stereo-seq data processing

Spatially resolved RNA-seq data generated by Stereo-seq were preprocessed to generate gene expression matrices for subsequent analyses. Sequenced reads were aligned to their spatial locations and the corresponding gene structures. Briefly, forward reads containing spatial barcodes were first aligned to coordinate identifiers (CIDs in the mask file) with a 1-bp mismatch tolerance. The remaining spatially valid data were then cleaned by filtering read pairs if the reverse reads providing cDNA sequences contained any sequencing adapters. The reverse reads were further aligned to the reference genome, and only mapped reads with a MAPQ greater than 10 were annotated using handleBam (https://github.com/BGIResearch/handleBam, code version: 1.0.0, accession date: 13 April 2022). During this annotation process, mapped reads that overlapped with more than 50% of the exon region were classified as exon transcripts, while others were annotated as intron or intergenic by checking whether they were located in the intron region or not. The final spatial gene expression profile matrix was generated using the exonic reads.

Cell segmentation and DNB assignment

To obtain the final spatial gene expression matrices, we performed data correspondence and cell assignment for DNBs. We performed object segmentation based on the nuclei in the single-stranded DNA staining image and then overlaid the outlines of these objects on the corresponding aligned gene expression spatial images.

To do this, we first performed image registration to align the single-stranded DNA images with the coordinates in the gene expression matrices. The spatial gene expression matrices were then transformed into two-dimensional spatial images by color-coding the unique molecular identifier counts of each DNB. Simultaneously, the corresponding single-stranded DNA images underwent a series of pre-processing steps, including background noise filtering and local contrast enhancement. These images were then transformed to their true physical size by replacing the pixel unit with the distance of the DNBs (715 nm). Subsequently, the pre-processed image datasets were aligned with the TrakEM2 algorithm by Fiji, where tissue morphology was manually identified and used for feature matching. The single-stranded DNA images were transformed to match the spatial coordinates of gene expression based on the affine matrix.

Next, we detected and segmented the cell objects in the transformed single-stranded DNA images using CellProfiler software (75). Briefly, nucleus locations were determined using the global Otsu thresholding strategy without manual annotation. The Mutex Watershed algorithm was used to automatically determine whether the pixel belonged to a specific nucleus, resulting in the creation of cell boundaries. These boundaries of segmented cells had preset object diameters ranging from 6 to 40 pixels. The RNA data associated with each DNB location were then assigned to individual cell objects based on whether they fell within the segmented cell boundaries. For the regions where dense nuclei were located and the boundary could not be detected, we manually masked these regions and retained bin15 partitions, each approximately 10 μm in width and height, instead of cell segmentation. Using the aligned coordinates of cells and gene expression, the gene expression matrices were further aggregated into putative cells ready for downstream analysis.

Stereo-seq data clustering

We filtered the aggregated cell-level spatial RNA data, retaining cells with a minimum of 25 gene types and a maximum of 7500. The filtered matrices were then normalized using the SCTransform function in Seurat 4.0. Data from different sections were then integrated using the IntegrateData function. PCA was performed with default parameters, and the top 30 principal components were extracted for unsupervised clustering with a resolution parameter set to 0.8. Spatial information was not considered for the clustering algorithm and we projected the resulting Seurat clusters into their spatial coordinates based on the cell-segmented locations.

Data integration of scRNA-seq and Stereo-seq

Three biological replicons from scRNA-seq and six Stereo-seq sections were integrated using the Seurat IntegrateData function, using the SCT normalization method and including 2000 variable features. The integrated data were processed using PCA and clustered with a resolution of 0.8. The resulting clusters were visualized using UMAP.

Cell type annotation

For the clustering results generated from the single-cell dataset, the Stereo-seq dataset, and the integrated dataset of single-cell and Stereo-seq, we identified a list of differentially expressed genes (DEGs) using the FindAllMarkers function in Seurat with default parameters, separately. For each of the datasets, we extracted the top 30 genes with greater fold change and significant *P* values and further queried their function in the Swiss-Prot database (https://www.ebi.ac.uk/uniprot/index).

Cell types or cluster identifiers were manually annotated based on the cluster-enriched DEGs for the single-cell dataset and the Stereo-seq dataset separately. To compare the similarities and differences between the scRNA-seq data and the spatial data, we projected the annotations from the two datasets into the same integrated UMAP dimension and manually verified their consistency. The final comprehensive annotations were determined based on the integrated clustering results, taking into account the previous single-cell and Stereo-seq annotation process.

Identification of spatially correlated modules

Gene modules were categorized using the autocorrelation algorithm, Hotspot, based on spatial location (36). The spatial expression matrix was normalized by the total unique molecular identifier count of each putative cell after filtering genes that expressed in less than three cells. A *K*-nearest neighbor (KNN) graph of genes was constructed with a parameter of 300 neighbors in spatial space. Genes with significant spatial autocorrelation [false discovery rate (FDR) < 0.05] were retained for the calculation of local correlations. Gene modules were identified using the create_modules function with a minimum gene threshold of 20 and an FDR threshold of 0.05.

Reuse of zebrafish single-cell transcriptomic dataset

To investigate the similarity between the endostyle cell types and vertebrate tissues, we collected scRNA-seq data of zebrafish from Farnsworth *et al.* (30), Qian *et al.* (39), and Gillotay *et al.* (40) as reference datasets. Before comparing the endostyle and zebrafish tissues, we preprocess the zebrafish datasets to unify the annotations and eliminate batch effects between different samples. These three zebrafish developmental and tissue-specific single-cell expression objects were

individually normalized and then integrated using Seurat 4.0. PCA was performed on the integrated assay, and the top 30 principal components were selected for clustering. The integrated assay was then reclustered at a resolution of 0.8 to generate uniform clustering labels, ensuring that cells with similar statistics or identities received the same label, even if they came from different data sources. Annotations were manually updated based on the original cell types and our integrated clustering.

Similarity inference of *S. clava* endostyle and zebrafish tissue/lineage

To infer the conservation of vertebrate tissue and the endostyle, we used previously reported Kullback-Leibler divergence methods (*76*) to calculate the similarities of endostyle to each zebrafish tissue. We performed separate analyses for adult tissues and developing tissues/lineages. The reason for this is that we considered that a mixture of gene signatures from tissues/lineages at different developmental status (potentially overlapped/similar expression profile in adjacent development status) may lead to a failure in the comparison process.

We identified homologous genes between *S. clava* and zebrafish by performing TBLASTX, which aligns the coding sequences of *S. clava* with those of zebrafish. Gene pairs with reciprocal best matches in both forward and reverse alignments were retained as true aligned hits. For one-to-many homologous gene pairs, we determined knee points based on the identity, bit score, and *e* values of the top 5 hits to each *S. clava* gene. Matches greater than the maximum identity values of the knee points were considered homologous hits.

SAMap analysis

Blast matches with an *e* value less than 10^{-6} and an identity value greater than 30% were kept as confident hits and were subsequently used in the SAMap comparison (*77*). Before comparison, we aggregated every 10 cells of the same cell type into pseudo-meta cells to emphasize the cell type–specific features. For the comparison between specific endostyle cell types and zebrafish tissues, we subset the endostyle thyroid-like cell cluster and zebrafish pharyngeal endoderm–related groups, and the endostyle HCLCs and zebrafish hair cell–related groups to apply SAMap analysis.

Cell trajectory analysis of the HLR cells

We used Monocle 3 (*78*) to construct a pseudotime trajectory for the HLR cells, which we assumed to have a hierarchy lineage based on the expressed signatures. Only cell types that were primarily located in the HLR were retained and used to construct the construction of trajectory, while cells located in the dense tissue region were filtered out. The subset data from each section were re-processed using Seurat 4.0 with SCTransform function and further integrated with 2000 integration features. PCA was performed on the top 30 principal components. The processed Seurat object was then converted to a Monocle3 object using as.cell_data_set function, and it was clustered using the UMAP dimensional reduction calculated from Seurat. We used methods including learn_graph and order_cells to construct the trajectory. The root of trajectory was assigned as the cell type closest to "stem cell" from the scRNA-seq dataset.

In addition, Dynamo (*79*) analysis was performed following the instructions at https://dynamo-release.readthedocs.io/. The unspliced and spliced raw count matrices were calculated using the Velocity package, based on the annotation file and bam file generated by SAW. These matrices were further processed using the recipe_monocle function in Dynamo. The top 3000 highly expressed genes were selected for dimension reduction using UMAP with default parameters. RNA velocity vectors were estimated on the normalized matrix and projected into the spatial visualization.

Cell-cell interaction analysis

We used ligand-receptor pairs from CellChatDB (*80*) to evaluate cell-cell interactions between each pair of cell types. To identify potential ligand and receptor genes in *S. clava*, we used the SonicParanoid (*81*) software with default parameters to identify orthologous relationships between *S. clava* and the human, which is the source of known ligand-receptor database. All orthologous genes, including one-to-one and one-to-many, were used as possible matches to further infer the ligand-receptor relationship in *S. clava*. The interaction scores of each ligand and receptor gene were calculated using the cell2cell (*82*) package, following the instructions at https://earmingol.github.io/cell2cell (code version: 0.6.4, accession date: 15 February 2023). Target cell types, specifically HCLCs and IRFCs, were selected based on manual determination of their spatial proximity.

Cross-species trajectory projection

Inspired by the evolutionary homology between the thyroid-like cell cluster in the endostyle and the thyroid organ in vertebrates, we performed similarity calculations between the endostyle thyroid-like cell cluster and tissues from different developmental processes of vertebrate thyroid. To investigate the evolutionary and developmental status of endostyle thyroid-like cells, we evaluated the similarity between thyroid-like cells and cells from different trajectory points in thyroid development and projected the thyroid-like cells onto a pseudo-evolutionary trajectory. We first constructed the pseudo-time trajectory of the pharyngeal endoderm (developmental origin of the thyroid) and the thyroid. Monocle 2 (*74*) was used to trace the developmental footprint of the zebrafish thyroid. Highly expressed genes (log_avg2FC ≥ 0.25) in pharyngeal endoderm and thyroid were estimated using the differentialGeneTest function to find genes that change along the pseudotime. By identifying the pseudotime point with maximum expression, these trajectory-related genes were then classified into different clusters and were treated as representing developmental stages. Finally, we ordered the endostyle thyroid-like cells based on the pseudo-time value of their most highly expressed trajectory-related genes and inferred the order as the pseudo-trajectory of endostyle thyroid-like cells.

To assess the spatial preference of endostyle thyroid-like cell pseudo-trajectory, we compared pseudotime values across different partitions along the spatial region. First, we manually designated two regions within the thyroid-like cell cluster located on either side of the bilateral-sympathetic region. We then used curve fitting to represent the spatial distribution of cells. The fitted curve was divided into three subregions, and each thyroid-like cell region was projected into these subregions based on the nearest distances.

In situ hybridization on the endostyle and peripheral blood

The fresh endostyle sample was fixed with 4% paraformaldehyde (PFA) in artificial water at 4°C overnight. Sucrose descent was performed with a 30% sucrose in phosphate-buffered saline (PBS). The dried tissue was embedded in OCT compound (Sakura 4583)

and snap-frozen in liquid nitrogen. Cryosections for in situ hybridization were prepared using Leica CM3050, and the sections were mounted onto adhesive slides. Peripheral blood from *S. clava* was collected from the non-endostyle pharynx blanket tissue and fixed with 4% PFA. Blood cell droplets were washed with PBS and then dried on adhesive slides. Digoxigenin (DIG)–RNA probes were designed (data S6) and synthesized by in vitro transcription using SP6 RNA polymerase and T7 RNA polymerase (Thermo Fisher Scientific, EP0131, EP0111). The in situ hybridization procedure was performed in a humidified black box to prevent the drying of the slides. Regions of tissue on the slides were outlined with hydrophobic paint. The tissue was digested with proteinase K (10 µg/ml) for 30 min at 37°C and washed with PBST (0.1% Tween 20 in PBS). The tissue was then refixed with 4% PFA in PBS solution for 30 min at room temperature and washed with PBST. The tissue was then incubated with pre-hybridization buffer [50% formamide, 5× saline sodium citrate buffer (SSC), 0.1% Tween 20] for 10 min at room temperature. The tissue was then incubated in hybridization buffer [50% formamide, 5× SSC, 0.1% Tween 20, 5× Denhardt's solution, yeast tRNA (100 µg/ml), and heparin (50 µg/ml)] at the hybridization temperature for 4 hours. A fresh hybridization buffer containing probes at a concentration of 0.5 ng/µl was added, and the tissue was incubated at the hybridization temperature for 20 hours. The tissue was washed four times with a solution (50% formamide, 5× SSC, 0.1% Tween 20) and four times with PBST. The tissue was then incubated in 1% blocking buffer (Roche) with Dig antibody at 4°C overnight. Finally, the tissue was incubated with nitro blue tetrazolium–bromochloroindolyl phosphate for chromogenic detection.

Cytoskeleton and nuclear staining of transverse sections

Cryosections were prepared as described above. The slide was incubated with 0.1% PBST (Triton X-100) for 20 min at room temperature. The slide was then washed three times with PBS solution. Phalloidin 488 (Alexa, A12379) was diluted at 1:100 with PBS, and the slide was stained in the absence of light at room temperature for 30 min. The slides were then sealed with VECTASHIELD Antifade Mounting Medium with DAPI (4′,6-diamidino-2-phenylindole) (H-1200-10) and resin. The prepared slides were visualized using a ZEISS LSM 900 confocal microscope.

Sample preparation and transmission electron microscopy observation

Fresh endostyle was collected and immersed into the fixation buffer (2.5% glutaraldehyde in artificial seawater) and stored at 4°C. The semi-thin section and ultrathin sections were prepared by the Biomedical Center of Qingdao Medical College, Qingdao University. The sample was observed by transmission electron microscope JEM-1200EX.

Statistical analysis

All the statistical analyses were performed using R (version 3.6). Student's *t* test and Spearman's rank correlation analysis were used in this study. The P value <0.05 was considered as a significant difference exists. $*0.01 < P < 0.05$; $**0.001 < P < 0.01$; $***P < 0.001$.

Supplementary Materials

This PDF file includes:
Figs. S1 to S9
Legends for data S1 to S6

Other Supplementary Material for this manuscript includes the following:
Data S1 to S6

REFERENCES AND NOTES

1. H. Gee, *Across the Bridge: Understanding the Origin of the Vertebrates* (The University of Chicago Press, 2018).
2. J. A. Gillis, J. H. Fritzenwanker, C. J. Lowe, A stem-deuterostome origin of the vertebrate pharyngeal transcriptional network. *Proc. Biol. Sci.* **279**, 237–246 (2012).
3. C. J. Lowe, D. N. Clarke, D. M. Medeiros, D. S. Rokhsar, J. Gerhart, The deuterostome context of chordate origins. *Nature* **520**, 456–465 (2015).
4. S. Clausen, A. B. Smith, Palaeoanatomy and biological affinities of a Cambrian deuterostome (Stylophora). *Nature* **438**, 351–354 (2005).
5. Q. Tian, F. Zhao, H. Zeng, M. Zhu, B. Jiang, Ultrastructure reveals ancestral vertebrate pharyngeal skeleton in yunnanozoans. *Science* **377**, 218–222 (2022).
6. F. Pardos, Fine structure and function of pharynx cilia in *Glossobalanus minutus* Kowalewsky (Enteropneusta). *Acta Zool.* **69**, 1–12 (1988).
7. E. Veitch, J. Begbie, T. F. Schilling, M. M. Smith, A. Graham, Pharyngeal arch patterning in the absence of neural crest. *Curr. Biol.* **9**, 1481–1484 (1999).
8. A. Graham, J. Richardson, Developmental and evolutionary origins of the pharyngeal apparatus. *EvoDevo* **3**, 24 (2012).
9. T. Piotrowski, C. Nusslein-Volhard, The endoderm plays an important role in patterning the segmented pharyngeal region in zebrafish (*Danio rerio*). *Dev. Biol.* **225**, 339–356 (2000).
10. N. Anthwal, H. Thompson, The development of the mammalian outer and middle ear. *J. Anat.* **228**, 217–232 (2016).
11. R. Olsson, Endostyles and endostylar secretions: A comparative histochemical study. *Acta Zool.* **44**, 299–328 (1963).
12. J. K. Petersen, Ascidian suspension feeding. *J. Exp. Mar. Biol. Ecol.* **342**, 127–137 (2007).
13. T. Hirano, H. Nishida, Developmental fates of larval tissues after metamorphosis in the ascidian, *Halocynthia roretzi*. II. Origin of endodermal tissues of the juvenile. *Dev. Genes Evol.* **210**, 55–63 (2000).
14. M. Ogasawara, R. Di Lauro, N. Satoh, Ascidian homologs of mammalian thyroid peroxidase genes are expressed in the thyroid-equivalent region of the endostyle. *J. Exp. Zool.* **285**, 158–169 (1999).
15. J. Hiruta, F. Mazet, K. Yasui, P. Zhang, M. Ogasawara, Comparative expression analysis of transcription factor genes in the endostyle of invertebrate chordates. *Dev. Dyn.* **233**, 1031–1037 (2005).
16. A. Sasaki, Y. Miyamoto, Y. Satou, N. Satoh, M. Ogasawara, Novel endostyle-specific genes in the ascidian *Ciona intestinalis*. *Zoolog. Sci.* **20**, 1025–1030 (2003).
17. E. J. Barrington, A. Thorpe, The identification of monoiodotyrosine, diiodotyrosine and thyroxine in extracts of the endostyle of the ascidian, *Ciona Intestinalis L. Proc. R. Soc. B Biol. Sci.* **163**, 136–149 (1965).
18. E. J. W. Barrington, A. Thorpe, An autoradiographic study of the binding of iodine125 in the endostyle and pharynx of the ascidian, *Ciona intestinalis L. Gen. Comp. Endocrinol.* **5**, 373–385 (1965).
19. M. Nilsson, H. Fagman, Development of the thyroid gland. *Development* **144**, 2123–2140 (2017).
20. B. Kluge, N. Renault, K. B. Rohr, Anatomical and molecular reinvestigation of lamprey endostyle development provides new insight into thyroid gland evolution. *Dev. Genes Evol.* **215**, 32–40 (2005).
21. M. Ogasawara, Overlapping expression of amphioxus homologs of the thyroid transcription factor-1 gene and thyroid peroxidase gene in the endostyle: Insight into evolution of the thyroid gland. *Dev. Genes Evol.* **210**, 231–242 (2000).
22. B. Rosental, M. Kowarsky, J. Seita, D. M. Corey, K. J. Ishizuka, K. J. Palmeri, S. Y. Chen, R. Sinha, J. Okamoto, G. Mantalas, L. Manni, T. Raveh, D. N. Clarke, J. M. Tsai, A. M. Newman, N. F. Neff, G. P. Nolan, S. R. Quake, I. L. Weissman, A. Voskoboynik, Complex mammalian-like haematopoietic system found in a colonial chordate. *Nature* **564**, 425–429 (2018).
23. A. Voskoboynik, Y. Soen, Y. Rinkevich, A. Rosner, H. Ueno, R. Reshef, K. J. Ishizuka, K. J. Palmeri, E. Moiseeva, B. Rinkevich, I. L. Weissman, Identification of the endostyle as a stem cell niche in a colonial chordate. *Cell Stem Cell* **3**, 456–464 (2008).
24. T. Osugi, Y. Sasakura, H. Satake, The ventral peptidergic system of the adult ascidian *Ciona robusta* (*Ciona intestinalis* Type A) insights from a transgenic animal model. *Sci. Rep.* **10**, 1892 (2020).
25. D. Parrinello, M. A. Sanfratello, A. Vizzini, N. Parrinello, M. Cammarata, Ciona intestinalis galectin (CiLgals-a and CiLgals-b) genes are differentially expressed in endostyle zones and challenged by LPS. *Fish Shellfish Immunol.* **42**, 171–176 (2015).
26. A. Vizzini, D. Parrinello, M. A. Sanfratello, M. R. Trapani, V. Mangano, N. Parrinello, M. Cammarata, Upregulated transcription of phenoloxidase genes in the pharynx and endostyle of *Ciona intestinalis* in response to LPS. *J. Invertebr. Pathol.* **126**, 6–11 (2015).

27. D. Arendt, J. M. Musser, C. V. H. Baker, A. Bergman, C. Cepko, D. H. Erwin, M. Pavlicev, G. Schlosser, S. Widder, M. D. Laubichler, G. P. Wagner, The origin and evolution of cell types. *Nat. Rev. Genet.* **17**, 744–757 (2016).

28. C. Cao, L. A. Lemaire, W. Wang, P. H. Yoon, Y. A. Choi, L. R. Parsons, J. C. Matese, W. Wang, M. Levine, K. Chen, Comprehensive single-cell transcriptome lineages of a proto-vertebrate. *Nature* **571**, 349–354 (2019).

29. E. Khrameeva, I. Kurochkin, D. Han, P. Guijarro, S. Kanton, M. Santel, Z. Qian, S. Rong, P. Mazin, M. Sabirov, M. Bulat, O. Efimova, A. Tkachev, S. Guo, C. C. Sherwood, J. G. Camp, S. Paabo, B. Treutlein, P. Khaitovich, Single-cell-resolution transcriptome map of human, chimpanzee, bonobo, and macaque brains. *Genome Res.* **30**, 776–789 (2020).

30. D. R. Farnsworth, L. M. Saunders, A. C. Miller, A single-cell transcriptome atlas for zebrafish development. *Dev. Biol.* **459**, 100–108 (2020).

31. A. Chen, S. Liao, M. Cheng, K. Ma, L. Wu, Y. Lai, X. Qiu, J. Yang, J. Xu, S. Hao, X. Wang, H. Lu, X. Chen, X. Liu, X. Huang, Z. Li, Y. Hong, Y. Jiang, J. Peng, S. Liu, M. Shen, C. Liu, Q. Li, Y. Yuan, X. Wei, H. Zheng, W. Feng, Z. Wang, Y. Liu, Z. Wang, Y. Yang, H. Xiang, L. Han, B. Qin, P. Guo, G. Lai, P. Munoz-Canoves, P. H. Maxwell, J. P. Thiery, Q. F. Wu, F. Zhao, B. Chen, M. Li, X. Dai, S. Wang, H. Kuang, J. Hui, L. Wang, J. F. Fei, O. Wang, X. Wei, H. Lu, B. Wang, S. Liu, Y. Gu, M. Ni, W. Zhang, F. Mu, Y. Yin, H. Yang, M. Lisby, R. J. Cornall, J. Mulder, M. Uhlen, M. A. Esteban, Y. Li, L. Liu, X. Xu, J. Wang, Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell* **185**, 1777–1792.e21 (2022).

32. E. Z. Macosko, A. Basu, R. Satija, J. Nemesh, K. Shekhar, M. Goldman, I. Tirosh, A. R. Bialas, N. Kamitaki, E. M. Martersteck, J. J. Trombetta, D. A. Weitz, J. R. Sanes, A. K. Shalek, A. Regev, S. A. McCarroll, Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).

33. J. Wei, J. Zhang, Q. Lu, P. Ren, X. Guo, J. Wang, X. Li, Y. Chang, S. Duan, S. Wang, H. Yu, X. Zhang, X. Yang, H. Gao, B. Dong, Genomic basis of environmental adaptation in the leathery sea squirt (*Styela clava*). *Mol. Ecol. Resour.* **20**, 1414–1431 (2020).

34. J. Zhang, J. Wei, H. Yu, B. Dong, Genome-wide identification, comparison, and expression analysis of transcription factors in ascidian *Styela clava*. *Int. J. Mol. Sci.* **22**, 4317 (2021).

35. A. Jiang, W. Zhang, J. Wei, P. Liu, B. Dong, Transcriptional analysis of the endostyle reveals pharyngeal organ functions in ascidian. *Biology* **12**, 245 (2023).

36. D. DeTomaso, N. Yosef, Hotspot identifies informative gene modules across modalities of single-cell genomics. *Cell Syst.* **12**, 446–456.e9 (2021).

37. M. C. Holley, Cell shape, spatial patterns of cilia, and mucus-net construction in the ascidian endostyle. *Tissue Cell* **18**, 667–684 (1986).

38. M. Tambalo, R. Mitter, D. G. Wilkinson, A single cell transcriptome atlas of the developing zebrafish hindbrain. *Development* **147**, dev184143 (2020).

39. F. Qian, G. Wei, Y. Gao, X. Wang, J. Gong, C. Guo, X. Wang, X. Zhang, J. Zhao, C. Wang, M. Xu, Y. Hu, G. Yin, J. Kang, R. Chai, G. Xie, D. Liu, Single-cell RNA-sequencing of zebrafish hair cells reveals novel genes potentially involved in hearing loss. *Cell. Mol. Life Sci.* **79**, 385 (2022).

40. P. Gillotay, M. Shankar, B. Haerlingen, E. Sema Elif, M. Pozo-Morales, I. Garteizgogeascoa, S. Reinhardt, A. Kränkel, J. Bläsche, A. Petzold, N. Ninov, G. Kesavan, C. Lange, M. Brand, A. Lefort, F. Libert, V. Detours, S. Costagliola, S. Sumeet Pal, Single-cell transcriptome analysis reveals thyrocyte diversity in the zebrafish thyroid gland. *EMBO Rep.* **21**, e50612 (2020).

41. V. Longo, D. Parrinello, A. Longo, M. G. Parisi, N. Parrinello, P. Colombo, M. Cammarata, The conservation and diversity of ascidian cells and molecules involved in the inflammatory reaction: The *Ciona robusta* model. *Fish Shellfish Immunol.* **119**, 384–396 (2021).

42. I. M. Min, G. Pietramaggiori, F. S. Kim, E. Passegue, K. E. Stevenson, A. J. Wagers, The transcription factor EGR1 controls both the proliferation and localization of hematopoietic stem cells. *Cell Stem Cell* **2**, 380–391 (2008).

43. T. Reya, A. W. Duncan, L. Ailles, J. Domen, D. C. Scherer, K. Willert, L. Hintz, R. Nusse, I. L. Weissman, A role for Wnt signalling in self-renewal of haematopoietic stem cells. *Nature* **423**, 409–414 (2003).

44. E. M. De Roberts, M. Blum, C. Niehrs, H. Steinbeisser, Goosecoid and the organizer. *Dev. Suppl.* **116**, 167–171 (1992).

45. W. Zhang, Z. Tang, S. Fan, D. Yao, Z. Zhang, C. Guan, W. Deng, Y. Ying, Protein tyrosine phosphatase receptor-type Q: Structure, activity, and implications in human disease. *Protein Pept. Lett.* **29**, 567–573 (2022).

46. A. Castiglione, C. Möller, Usher Syndrome. *Audiol. Res.* **12**, 42–65 (2022).

47. N. Michalski, C. Petit, Genes involved in the development and physiology of both the peripheral and central auditory systems. *Annu. Rev. Neurosci.* **42**, 67–86 (2019).

48. P. D. Mathur, J. Yang, Usher syndrome and non-syndromic deafness: Functions of different whirlin isoforms in the cochlea, vestibular organs, and retina. *Hear. Res.* **375**, 14–24 (2019).

49. P. Zhou, H. Meng, X. Liang, X. Lei, J. Zhang, W. Bian, N. He, Z. Lin, X. Song, W. Zhu, B. Hu, B. Li, L. Yan, B. Tang, T. Su, H. Liu, Y. Mao, Q. Zhai, Y. Yi, ADGRV1 variants in febrile seizures/epilepsy with antecedent febrile seizures and their associations with audio-visual abnormalities. *Front. Mol. Neurosci.* **15**, 864074 (2022).

50. Y. Brand, M. Sung, K. Pak, E. Chavez, E. Wei, V. Radojevic, D. Bodmer, A. F. Ryan, Neural cell adhesion molecule NrCAM is expressed in the mammalian inner ear and modulates spiral ganglion neurite outgrowth in an in vitro alternate choice assay. *J. Mol. Neurosci.* **55**, 836–844 (2015).

51. P. Cafferty, L. Yu, H. Long, Y. Rao, Semaphorin-1a functions as a guidance receptor in the Drosophila visual system. *J. Neurosci.* **26**, 3999–4003 (2006).

52. J. L. Wémeau, P. Kopp, Pendred syndrome. *Best Prac. Res. Clin. Endocrinol. Metab.* **31**, 213–224 (2017).

53. D. Vullhorst, R. M. Mitchell, C. Keating, S. Roychowdhury, I. Karavanova, J. H. Tao-Cheng, A. Buonanno, A negative feedback loop controls NMDA receptor function in cortical interneurons via neuregulin 2/ErbB4 signalling. *Nat. Commun.* **6**, 7222 (2015).

54. F. Hu, T. Padukkavidana, C. B. Vægter, O. A. Brady, Y. Zheng, I. R. Mackenzie, H. H. Feldman, A. Nykjaer, S. M. Strittmatter, Sortilin-mediated endocytosis determines levels of the frontotemporal dementia protein, progranulin. *Neuron* **68**, 654–667 (2010).

55. J. Hiruta, F. Mazet, M. Ogasawara, Restricted expression of NADPH oxidase/peroxidase gene (Duox) in zone VII of the ascidian endostyle. *Cell Tissue Res.* **326**, 835–841 (2006).

56. L. Chen, J. Yan, J. Shi, W. Sun, Z. Chen, J. Yu, J. Qi, Y. Du, H. Zhang, L. Feng, Zebrafish intelectin 1 (zITLN1) plays a role in the innate immune response. *Fish Shellfish Immunol.* **83**, 96–103 (2018).

57. L. Chen, J. Li, G. Yang, A comparative review of intelectins. *Scand. J. Immunol.* **92**, e12882 (2020).

58. X. Guo, X. Dou, B. Dong, Identification and functional characterization of lactadherin, an agglutinating glycoprotein from the chordate Styela clava. *In Vitro Cell. Dev. Biol.* **55**, 405–415 (2019).

59. B. K. Hall, Evolutionary developmental biology (evo-devo): Past, present, and future. *Evol. Educ. Outreach* **5**, 184–193 (2012).

60. P. Ma, X. Liu, Z. Xu, H. Liu, X. Ding, Z. Huang, C. Shi, L. Liang, L. Xu, X. Li, G. Li, Y. He, Z. Ding, C. Chai, H. Wang, J. Qiu, J. Zhu, X. Wang, P. Ding, S. Zhou, Y. Yuan, W. Wu, C. Wan, Y. Yan, Y. Zhou, Q. J. Zhou, G. D. Wang, Q. Zhang, X. Xu, G. Li, S. Zhang, B. Mao, D. Chen, Joint profiling of gene expression and chromatin accessibility during amphioxus development at single-cell resolution. *Cell Rep.* **39**, 110979 (2022).

61. W. Wang, X. Niu, T. Stuart, E. Jullian, W. M. Mauck III, R. G. Kelly, R. Satija, L. Christiaen, A single-cell transcriptional roadmap for cardiopharyngeal fate diversification. *Nat. Cell Biol.* **21**, 674–686 (2019).

62. S. Sharma, W. Wang, A. Stolfi, Single-cell transcriptome profiling of the Ciona larval brain. *Dev. Biol.* **448**, 226–236 (2019).

63. E. M. Erkenbrack, J. C. Croce, E. Miranda, S. Gautam, M. Martinez-Bartolome, S. Yaguchi, R. C. Range, Whole mount in situ hybridization techniques for analysis of the spatial distribution of mRNAs in sea urchin embryos and early larvae. *Methods Cell Biol.* **151**, 177–196 (2019).

64. Y. Hasegawa, J. L. Mark Welch, B. J. Rossetti, G. G. Borisy, Preservation of three-dimensional spatial structure in the gut microbiome. *PLOS ONE* **12**, e0188257 (2017).

65. N. Redmayne, S. L. Chavez, Optimizing tissue preservation for high-resolution confocal imaging of single-molecule RNA-FISH. *Curr. Protoc. Mol. Biol.* **129**, e107 (2019).

66. J. Wei, P. Liu, F. Liu, A. Jiang, J. Qiao, Z. Pu, B. Wang, J. Zhang, D. Jia, Y. Li, S. Wang, B. Dong, EDomics: A comprehensive and comparative multi-omics database for animal evo-devo. *Nucleic Acids Res.* **51**, D913–D923 (2023).

67. T. A. Onuma, R. Nakanishi, Y. Sasakura, M. Ogasawara, Nkx2-1 and FoxE regionalize glandular (mucus-producing) and thyroid-equivalent traits in the endostyle of the chordate *Oikopleura dioica*. *Dev. Biol.* **477**, 219–231 (2021).

68. M. Yamagishi, T. Huang, A. Hozumi, T. A. Onuma, Y. Sasakura, M. Ogasawara, Differentiation of endostyle cells by Nkx2-1 and FoxE in the ascidian *Ciona intestinalis* type A: Insights into shared gene regulation in glandular- and thyroid-equivalent elements of the chordate endostyle. *Cell Tissue Res.* **390**, 189–205 (2022).

69. A. Graham, S. M. Shimeld, The origin and evolution of the ectodermal placodes. *J. Anat.* **222**, 32–40 (2013).

70. T. Ozaki, T. Matsubara, D. Seo, M. Okamoto, K. Nagashima, Y. Sasaki, S. Hayase, T. Murata, X. H. Liao, J. Hanson, J. Rodriguez-Canales, S. S. Thorgeirsson, K. Kakudo, S. Refetoff, S. Kimura, Thyroid regeneration: Characterization of clear cells after partial thyroidectomy. *Endocrinology* **153**, 2514–2525 (2012).

71. Y. Hao, S. Hao, E. Andersen-Nissen, W. M. Mauck III, S. Zheng, A. Butler, M. J. Lee, A. J. Wilk, C. Darby, M. Zager, P. Hoffman, M. Stoeckius, E. Papalexi, E. P. Mimitou, J. Jain, A. Srivastava, T. Stuart, L. M. Fleming, B. Yeung, A. J. Rogers, J. M. McElrath, C. A. Blish, R. Gottardo, P. Smibert, R. Satija, Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).

72. C. S. McGinnis, L. M. Murrow, Z. J. Gartner, DoubletFinder: Doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst.* **8**, 329–337.e4 (2019).

73. M. D. Young, S. Behjati, SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. *Gigascience* **9**, giaa151 (2020).

74. X. Qiu, A. Hill, J. Packer, D. Lin, Y. A. Ma, C. Trapnell, Single-cell mRNA quantification and differential analysis with Census. *Nat. Methods* **14**, 309–315 (2017).

75. D. R. Stirling, M. J. Swain-Bowden, A. M. Lucas, A. E. Carpenter, B. A. Cimini, A. Goodman, CellProfiler 4: Improvements in speed, utility and usability. *BMC Bioinformatics* **22**, 433 (2021).

76. S. Levy, A. Elek, X. Grau-Bové, S. Menéndez-Bravo, M. Iglesias, A. Tanay, T. Mass, A. Sebé-Pedrós, A stony coral cell atlas illuminates the molecular and cellular basis of coral symbiosis, calcification, and immunity. *Cell* **184**, 2973–2987.e18 (2021).

77. A. J. Tarashansky, J. M. Musser, M. Khariton, P. Li, D. Arendt, S. R. Quake, B. Wang, Mapping single-cell atlases throughout Metazoa unravels cell type evolution. *eLife* **10**, e66747 (2021).

78. J. Cao, M. Spielmann, X. Qiu, X. Huang, D. M. Ibrahim, A. J. Hill, F. Zhang, S. Mundlos, L. Christiansen, F. J. Steemers, C. Trapnell, J. Shendure, The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019).

79. X. Qiu, Y. Zhang, J. D. Martin-Rufino, C. Weng, S. Hosseinzadeh, D. Yang, A. N. Pogson, M. Y. Hein, K. H. J. Min, L. Wang, E. I. Grody, M. J. Shurtleff, R. Yuan, S. Xu, Y. Ma, J. M. Replogle, E. S. Lander, S. Darmanis, I. Bahar, V. G. Sankaran, J. Xing, J. S. Weissman, Mapping transcriptomic vector fields of single cells. *Cell* **185**, 690–711.e45 (2022).

80. S. Jin, C. F. Guerrero-Juarez, L. Zhang, I. Chang, R. Ramos, C. H. Kuan, P. Myung, M. V. Plikus, Q. Nie, Inference and analysis of cell-cell communication using CellChat. *Nat. Commun.* **12**, 1088 (2021).

81. S. Cosentino, W. Iwasaki, SonicParanoid: Fast, accurate and easy orthology inference. *Bioinformatics* **35**, 149–151 (2019).

82. E. Armingol, A. Ghaddar, C. J. Joshi, H. Baghdassarian, I. Shamie, J. Chan, H. L. Her, S. Berhanu, A. Dar, F. Rodriguez-Armstrong, O. Yang, E. J. O'Rourke, N. E. Lewis, Inferring a spatial code of cell-cell interactions across a whole animal body. *PLOS Comput. Biol.* **18**, e1010715 (2022).