



# Discovery of an unconventional lamprey lymphocyte lineage highlights divergent features in vertebrate adaptive immune system evolution

Received: 9 December 2023

Accepted: 16 August 2024

Published online: 03 September 2024

Check for updates

Yingyi Huang<sup>1,2,10</sup>, Xiang Liu<sup>1,2,3,10</sup>, Shuo Li<sup>1,2</sup>, Chen Li<sup>1,2</sup>, Hong-Yan Wang<sup>1,2</sup>, Qun Liu<sup>4,5,6</sup>, Jian-Yang Chen<sup>4,5</sup>, Yingying Zhang<sup>4,5</sup>, Yanan Li<sup>1,2</sup>, Xianghui Zhang<sup>1,2</sup>, Qian Wang<sup>1,2</sup>, Kaiqiang Liu<sup>1,2</sup>, Yu-Yan Liu<sup>1,2</sup>, Yue Pang<sup>7</sup>, Shanshan Liu<sup>8,9</sup>, Guangyi Fan<sup>1,2,4,5,8</sup> & Changwei Shao<sup>1,2</sup>

Lymphocyte receptors independently evolved in both jawed and jawless vertebrates with similar adaptive immune responses. However, the diversity of functional subtypes and molecular architecture in jawless vertebrate lymphocytes, comparable to jawed species, is not well defined. Here, we profile the gills, intestines, and blood of the lamprey, *Lampetra morii*, with single-cell RNA sequencing, using a full-length transcriptome as a reference. Our findings reveal higher tissue-specific heterogeneity among T-like cells in contrast to B-like cells. Notably, we identify a unique T-like cell subtype expressing a homolog of the nonlymphoid hematopoietic growth factor receptor, *MPL-like* (*MPL-L*). These *MPL-L*<sup>+</sup> T-like cells exhibit features distinct from T cells of jawed vertebrates, particularly in their elevated expression of hematopoietic genes. We further discovered that *MPL-L*<sup>+</sup> VLRA<sup>+</sup> T-like cells are widely present in the typhlosole, gill, liver, kidney, and skin of lamprey and they proliferate in response to both a T cell mitogen and recombinant human thrombopoietin. These findings provide new insights into the adaptive immune response in jawless vertebrates, shedding new light on the evolution of adaptive immunity.

The evolution of adaptive immunity in jawed vertebrates has been regarded as the immunological 'big bang', which was thought to be initiated by the horizontal gene transfer of an ancient RAG transposon<sup>1,2</sup>. Recent evidence suggests that the evolution of adaptive immunity is not unique to jawed vertebrates. Indeed, the discovery of a second adaptive immune system in jawless vertebrates (lampreys and hagfishes) has altered the view regarding the origin of adaptive

immunity<sup>3</sup>. Jawless vertebrates evolved different mechanisms of lymphocyte receptor diversification from those of jawed vertebrates. Jawed vertebrates rely on somatic rearrangement of the variable (V), diversity (D), and joining (J) gene segments to generate highly diverse repertoires of B and T cell receptors (BCRs and TCRs). By contrast, jawless vertebrates generate diverse repertoires of variable lymphocyte receptors (VLR) A, B, and C through somatic rearrangement of

<sup>1</sup>State Key Laboratory of Mariculture Biobreeding and Sustainable Goods, Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Qingdao, Shandong, China. <sup>2</sup>Laboratory for Marine Fisheries Science and Food Production Processes, Qingdao Marine Science and Technology Center, Qingdao, Shandong, China. <sup>3</sup>Qingdao-Europe Advanced Institute for Life Sciences, BGI Research, Qingdao, China. <sup>4</sup>BGI Research, Qingdao, China. <sup>5</sup>Qingdao Key Laboratory of Marine Genomics, BGI Research, Qingdao, China. <sup>6</sup>Department of Biology, University of Copenhagen, Copenhagen, Denmark. <sup>7</sup>College of Life Sciences, Liaoning Normal University, Dalian, China. <sup>8</sup>BGI Research, Shenzhen, China. <sup>9</sup>MGI Tech, Shenzhen, China. <sup>10</sup>These authors contributed equally: Yingyi Huang, Xiang Liu. e-mail: [shao@ysfri.ac.cn](mailto:shao@ysfri.ac.cn)

LRR cassettes that flank the gene<sup>4</sup>. Although the nature of lymphocyte receptors is different between jawless and jawed vertebrates, certain features of the two lineages are surprisingly similar<sup>5</sup>. Diverse *VLRB* genes encode secretory antigen-specific agglutinins and are only expressed by B-like lymphocytes, which are equivalent to B cells in jawed vertebrates<sup>6</sup>. The *VLRA* and *VLRC* genes are expressed in two discrepant T-like lymphocyte lineages, which are similar to the  $\alpha\beta$ -T cells and  $\gamma\delta$ -T cells in jawed vertebrates<sup>7</sup>. In contrast to the well-charted landscape of adaptive immunity in jawed vertebrates, our understanding of the lymphocyte-like cell properties in jawless vertebrates is still preliminary.

The thymus is an essential organ in jawed vertebrates where T cell development occurs, encompassing the rearrangement of germline-encoded gene segments to express antigen-specific TCRs<sup>8,9</sup>. The thymic development process, which involves the negative and positive selection of T cells, ensures their ability to discern self from non-self without inducing autoimmunity<sup>10</sup>. The notion that the thymus was absent in jawless vertebrates was invalidated by Boehm et al.<sup>11</sup>. The discrete thymus-like structure, termed thymoid, was found in the gill basket of lamprey larvae, accompanied by the concomitant expression of cytosine deaminase 1 (*CDA1*), *VLRA*, and lamprey ortholog of *FOXN1*. Parallel to the thymus in jawed vertebrates, the thymoids provide a tissue microenvironment for *VLRA* assembly in *VLRA*<sup>+</sup> lymphocytes and selection for the expression of functional *VLRA* genes. Investigations on lamprey lymphocyte-like cells are mainly focused on morphological appearance and the expression of related genes, guided by the plentiful context of jawed vertebrates. However, the heterogeneity of lymphocyte-like cell subtypes and the gene programs that underpin their pleiotropic functions remain poorly understood. It is a challenge to distinguish the potential functional subpopulations of lymphocyte-like cells in the absence of relevant molecular markers.

Single-cell RNA sequencing (scRNA-seq) provides an effective method for analyzing cell lineage development and gene expression patterns, as well as determining undefined cell lineages. While the deficiencies of genomics resources hinder the application of scRNA-seq in species without any reference genome, full-length transcriptome sequencing has been proven to be an alternative strategy with equivalent sensitivity to whole genome alignments<sup>12</sup>.

Here, we combine scRNA-seq and full-length transcriptome sequencing to generate a global transcriptomic profile of the various cell populations present in the blood, gills, and intestines collected from *Lampræta morii* (*L. morii*), a lamprey that inhabits Northeast China and belongs to the class Cyclostomata and order Petromyzontiformes with the typical adaptive immunity features of jawless vertebrates. We identify several lymphocyte-like cell subtypes and defined an unconventional T-like cell subtype (MPL-L+ T-like cells) characterized by the specific expression of the thrombopoietin receptor-like (*MPL-L*) gene. We further use several computational strategies combined with immunofluorescence staining and flow cytometry analysis to determine the transcriptome, spatial localization, immunoreactivity, and evolutionary divergence of MPL-L+ T-like cells. Altogether, this work provides insight into adaptive immunity in jawless vertebrates, highlights the contribution of MPL-L+ T-like cells to immune regulation in jawless vertebrates, and further reveals the divergence of adaptive immunity between jawed and jawless vertebrates.

## Results

### A multi-immune-tissue cell type atlas of *L. morii* larvae

The adaptive immune system (AIS) in jawless and jawed vertebrates is underpinned by distinct lymphoid receptor systems (Fig. 1a). To better characterize the transcriptional landscape of the immune system of *L. morii* larvae, which is derived from ammocoete stage and whose genome has not yet been sequenced, the full-length transcriptome of *L. morii* was utilized for further scRNA-seq data analysis (Fig. 1b and Supplementary Fig. 1). Our results proved that the gene information

provided by the full-length transcriptome was sufficient to support the subsequent analysis (Supplementary data 1-3).

To study the diversity of cell types in *L. morii* immune tissues, we sampled single cells from three tissue types: gill, blood, and intestine (including typhlosole), and prepared six gill libraries, three blood libraries, and five intestinal libraries (Fig. 1b). The gills contain thymus-like structures at the tips of the gill filaments and the neighboring secondary lamellae, where developing T-like cells are enriched. Our intestinal samples include typhlosole, an immune/lymphopoietic tissue that is highly active in ammocoete larvae and serves as the site of hematopoiesis and immune cell development in lamprey. Blood circulates throughout the body and is involved in transporting immune cells and molecules, thus playing a crucial role in systemic immunity. After quality filtering, we retained 25,978 gill cells, 16,483 blood cells, and 23,204 intestine cells for further analysis, with medians of 590, 489, and 873 transcripts/cell, respectively (Fig. 1c). We then applied the Seurat R package to group the cells into transcriptional clusters, which we used as the fundamental data for downstream analyses (Fig. 1c). We distinguished erythrocytes and immune cell populations based on the marker genes *ALAS1* and *CD45* (Fig. 1d). According to the expression of the *VLRA*, *VLRC*, and *VLRB* genes, we made a distinction between T-like cells and B-like cells. Notably, the expression of *VLRA* and *VLRC* genes was almost identical, but both differed from *VLRB*, which was consistent with previous studies<sup>7</sup> (Fig. 1d). We first annotated the cell atlas of gills, focusing on the identification of myeloid cell populations by the previous literature and the analysis of differentially expressed genes (DEGs). The myeloid cells included monocytes (*MMP19*, *MMP13*, etc.), macrophages (*CTSL*, *SIGLEC1*, etc.), and neutrophils/dendritic cell like (DC-like) (*CSF3R*, *CXCL1*, etc.) (Supplementary Fig. 2a). Erythrocytes and some somatic cell populations were also found, including erythrocytes (*ALAS1*), epithelial cells (*EPCAM/TACSTD2*, *KRT18*, etc.), fibroblasts (*COL8A1*, *FBLN2*, etc.), and vascular smooth muscle (VSM) cells (*ACTG2*, *MYLK*, etc.) (Supplementary Fig. 2a). Notably, we identified a specific cell population characterized by the expression of multiple interferon-stimulated genes (ISGs), including *MX1*, *TRIM25/29/80*, *TRIM65*, etc. This expression pattern indicates a unique response to interferon signaling within these ISGs-expressing cells. We also identified a hematopoietic population (*TAL1*, *ITGB3*, *PMP22*, etc.) (Supplementary Fig. 2a). The above marker genes were also used for the subsequent annotation process of the intestinal and blood atlases (Supplementary Fig. 2b, c).

### Cell heterogeneity across tissues

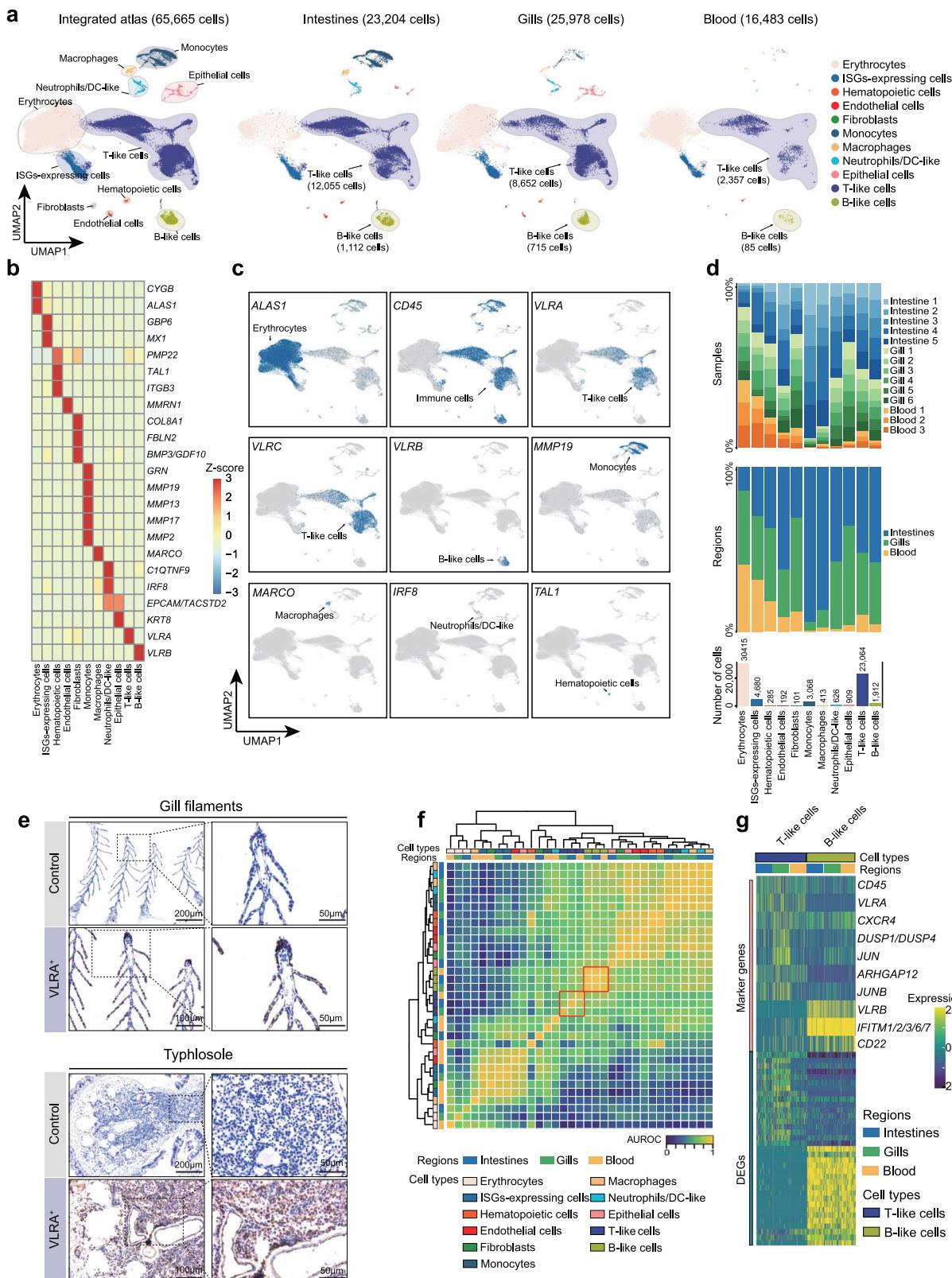
To better compare the differences in cell types among gills, blood, and intestines, we integrated and clustered the cell types and visualized the maps (Supplementary Fig. 3a, b). These populations include erythrocytes, ISGs-expressing cells, hematopoietic cells, endothelial cells, fibroblasts, monocytes, macrophages, neutrophils/DC-like, epithelial cells, T-like cells, and B-like cells (Fig. 2a–c, Supplementary Fig. 3c, and Supplementary data 11). By analyzing the cell proportion, we found that the proportions of erythrocytes and lymphoid cells were higher than the proportions of other cell types in the three tissues. Notably, the intestines exhibited a higher concentration of innate immune cells relative to the gills and blood (Fig. 2d). Immunohistochemistry with an anti-VLRA pAb was used to identify *VLRA*<sup>+</sup> cells in the gill filaments and typhlosoles. The results showed that numerous brown positive signals were observed in all investigated tissues (Fig. 2e and Supplementary Fig. 4).

We further performed a MetaNeighbor analysis to investigate the cell type relationships among the three tissues. ISGs-expressing cells, hematopoietic cells, and erythrocytes showed similar characteristics within the gills and intestines, exhibiting identical expression patterns across these three distinct tissues (Fig. 2f and Supplementary Fig. 5a). It was also found that ISGs-expressing cells and hematopoietic cells shared conserved marker genes of erythrocytes by analyzing gene



**Fig. 1 | Transcriptional characterization of the gill, intestinal, and blood atlases in *L. morii* larvae.** **a** Jawless vertebrates phylogenetic position and their unique adaptive immune system. **b** Schematic diagram of the experimental design for the full-length transcriptome and scRNA-sequencing. **c** UMAP of the three tissue types in *L. morii* larvae, including gills, intestines, and blood. The different colors in the picture represent different cell types. Violin plots in the middle show the gene

number of each cell type. The bottom bar plot shows the distribution of each cell type in each sample. The different colors in the picture indicate the different samples. The color of the abscissa represents different cell categories. Dendritic cell like, DC-like; vascular smooth muscle cell, VSM. **d** Feature plots of the differential expression of *ALAS1*, *CD45*, *VLRA*, *VLRC*, and *VLRB* in the atlases. The lamprey gene symbols in this study are indicated by mouse orthologue names.



expression patterns, which suggests that it may be one of the characteristics of lamprey blood system (Supplementary Fig. 5b). The genes related to innate immune and hematopoietic functions (*MYH9*, *GRN*, *C1QTNF9*, *CXCR4*, *CD45*, etc.) were upregulated in intestinal epithelial cells, whereas signature genes related to the development and function of thymic epithelial cells and gill respiration, such as *CXCL9*, *KRT8*, *KRT18*, and *ATP1B2*, were upregulated in gill epithelial cells

(Supplementary Fig. 5c). The heterogeneity of these cell types may be associated with the specific immune microenvironment across the tissue of lamprey. Notably, a pronounced cross-tissue similarity was observed in B-like cells but not shown in T-like cells (Fig. 2f). We further conducted an in-depth analysis of the marker genes and DEGs in T-like and B-like cells across the three tissues. Our findings indicate that the gene expression profiles of B-like cells were remarkably consistent

**Fig. 2 | Integrative transcriptome profiling of gill, intestinal, and blood atlases.** a UMAP visualization of cell type distribution after integration of the gill, blood, and intestinal atlases. The color annotation indicates each cell type annotated. **b** Heatmap of selected marker genes across various cell types. **c** Feature plot of selected marker genes across various cell types. **d** The top bar plot shows the distribution of all cell types in each sample. The color annotation indicates various samples. The bar plot in the middle shows the distribution of all cell types in three tissues. The bottom bar plot shows the cell numbers of each cell type. **e** Immunohistochemistry examinations of the gill filaments and typhlosoles.

Numerous VLRA<sup>+</sup> cells from the gill filaments and typhlosoles were well recognized by anti-VLRA pAb with experiments performed at least three times. **f** Heatmap showing the similarity of multiple cell types in three organs/tissues. The color of the outermost rectangles indicates various cell types. The color of rectangles on the second floor indicates gill, blood, or intestine. The brighter color in the picture, the higher AUROC value is. The AUROC value is proportional to the cell correlation. T-like cells and B-like cells are indicated in the red box. **g** Heatmap of the marker genes and top 20 DEGs in T-like cells and B-like cells across three tissues.

across all three tissues. However, the expression profiles of certain T-like cell markers and DEGs revealed significant heterogeneity, which could be attributed to the tissue-specific differentiation of T-like cell subtypes (Fig. 2g and Supplementary Fig. 5d).

### Discovery and characterization of distinct lymphocyte-like subtypes

The emergences of VLRs are thought to be the origin of the adaptive immune system in jawless vertebrates<sup>13</sup>. However, there is still a lack of systematic research to elucidate the cellular heterogeneity of lymphocyte populations in jawless vertebrates. To explore *L. morii* lymphoid cell biology in greater detail, we reintegrated and reanalyzed T-like cells and B-like cells (Fig. 3a and Supplementary Fig. 6a). Based on the differential gene expression profiles, five T-like cell subtypes, and one B-like cell population were identified, including (1) MPL-L+ T-like cells (*MPL-L*, *ITGB3*, and *FHOD3*); (2) T-like cells subtype 1 (*CXCR1/2/3*, *JUN*, and *IL16/PDZD2*); (3) T-like cells subtype 2 (*THEMIS1/2/3*, *CSF1R/KIT*, and *VSIG4*); (4) cycling T-like cells (*TOP2A/B*, *AURKA/B*, and *PCNA*); (5) T-like cells subtype 3 (*ZEB1/2*, *CTDSP1*, and *PRDX2*); and (6) B-like cells (*CD22*, *IFITM1/2/3/6/7*, and *BLNK*) (Fig. 3b, c). By analyzing the signature genes, we found that most T-like cell subtypes highly expressed genes involved in functions similar to those of conventional T cells in jawed vertebrates. For example, T-like cell subtype 1 expressed signatures involved in T-cell chemotaxis, T-like cell subtype 2 expressed genes related to T-cell activation, and cycling T-like cells highly expressed genes related to the cell cycle. T-like cell subtype 3 expressed genes related to cellular homeostasis (Fig. 3d and Supplementary Fig. 6b).

The signature genes of MPL-L+ T-like cells were related to hematopoiesis (*MPL-L*, *GM14681/NPM1/3*, *KDELRI/MYH9/10/11/14*, etc.) (Fig. 3d and Supplementary Fig. 6b). *MPL-L* gene highly and widely expressed in the MPL-L+ T-like cells and co-expressed with *VLRA* (Fig. 3e). Interestingly, *MPL* is a nonlymphoid hematopoietic growth factor receptor in jawed vertebrates that mainly binds with its ligand thrombopoietin (TPO) to participate in the self-renew of hematopoietic stem cells (HSCs) and the maturation of megakaryocytes-derived platelets, but not in the development and activation of lymphocytes<sup>14–16</sup>. Simultaneously *MPL* is capable activate a series of signal transduction pathways, including the JAK/STAT phosphorylation pathway and the RAS pathway<sup>14,17,18</sup>. However, current studies in mouse show that introducing *Mpl* gene into T cells through transgenic technology can enhance T-cell self-renewal and antitumor ability based on the TPO-MPL signal transduction pathways that overlap with the common signaling pathway of costimulatory molecules and cytokine production during T-cell activation (Fig. 3f)<sup>19</sup>.

To assess the functional conservation of the *MPL-L* gene in jawless vertebrates compared to its counterparts in jawed vertebrates, we analyzed the function domain of the *MPL-L* gene. Our findings indicated that the *MPL* homolog in lamprey also shared the conserved extracellular function domain of EpoR\_lig-bind in jawed vertebrates (Supplementary Fig. 7). We further compared the distribution differences among T-like cell subtypes and found that T-like cell subtype 1, T-like cell subtype 2, cycling T-like cells, and T-like cell subtype 3 mainly exist in the gills, whereas MPL-L+ T-like cells are the major T-like

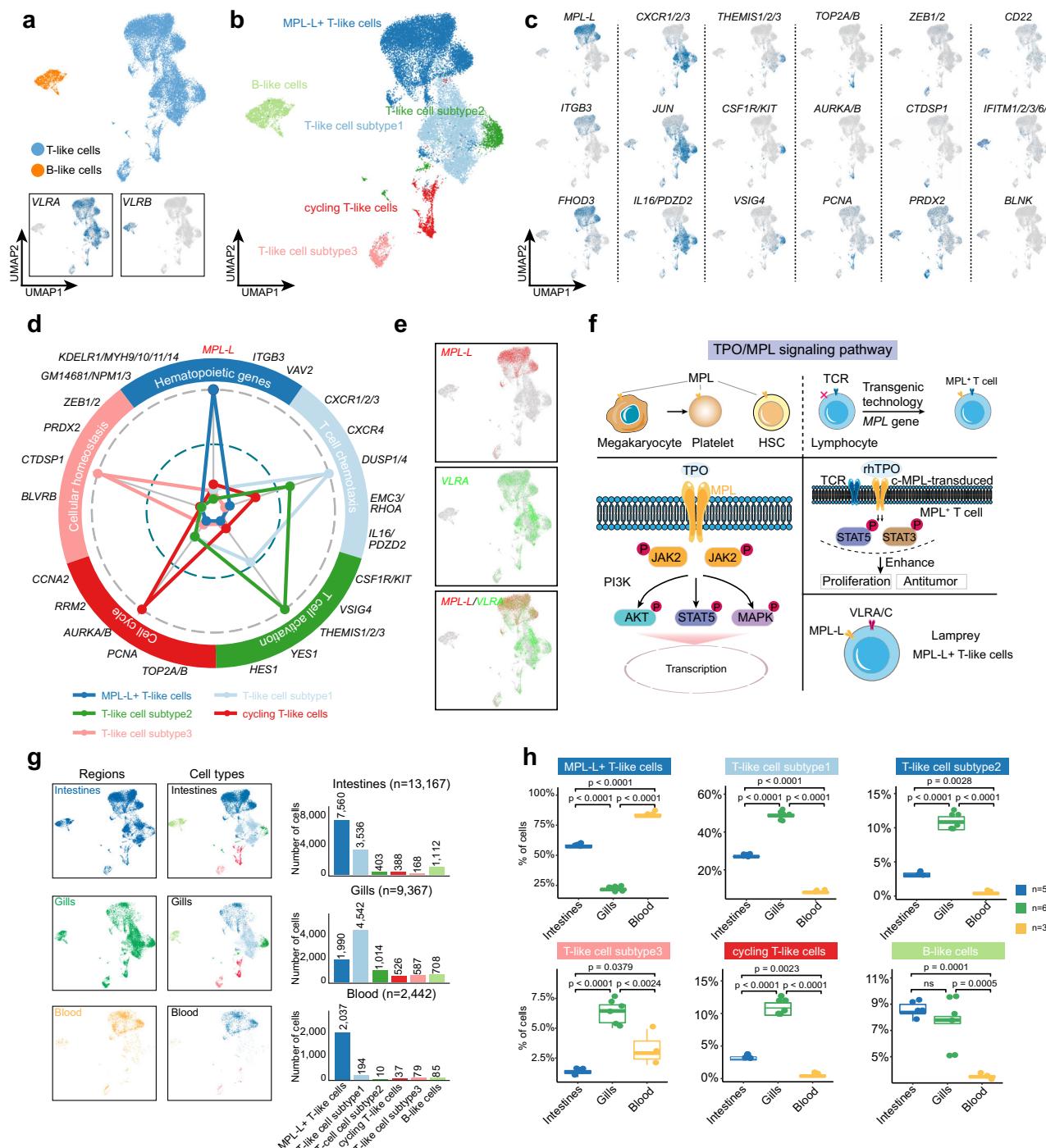
cell subtype in the intestines and blood (Fig. 3g, h). Based on the observed transcriptional heterogeneity and distinctive distribution patterns, we suggest that MPL-L+ T-like cells constitute an unconventional population within the lymphoid cell lineage of lamprey.

### MPL-L+ T-like cells represent a unique aspect of the adaptive immune system in jawless vertebrates distinguished from jawed counterparts

The adaptive immune response in jawless vertebrates is similar to that of jawed vertebrates<sup>5</sup>. However, our research has uncovered a divergence in gene expression within MPL-L+ T-like cells when contrasted with the signature genes of traditional T cells in jawed vertebrates. To explore the cellular connections between the immunological profiling of these two groups of vertebrates, we compared *L. morii* immune cells with available single-cell data from zebrafish and mouse separately, two representatives of jawed vertebrates<sup>20</sup>.

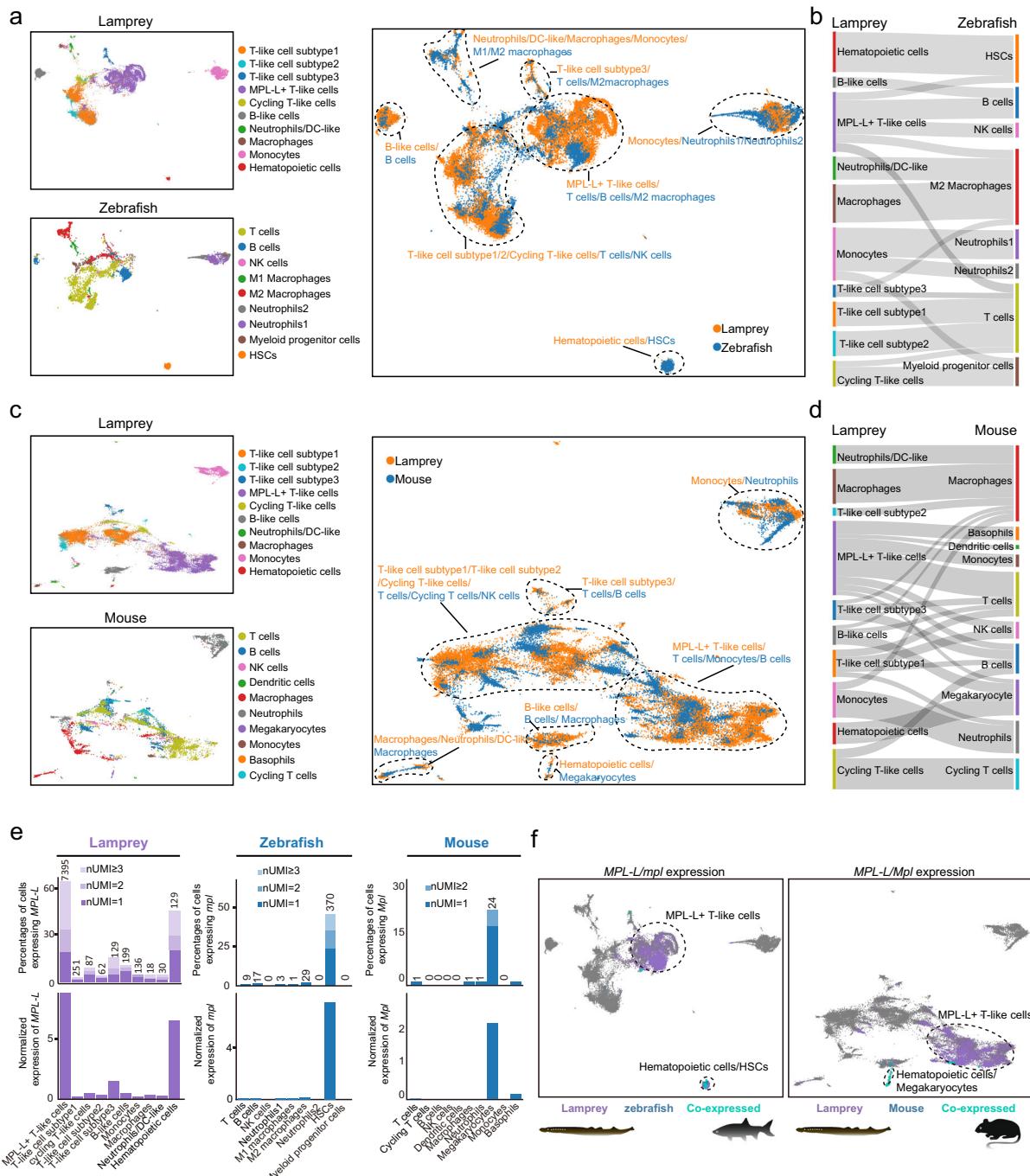
The SAMap analysis results showed broad agreement in the innate immune cells and hematopoietic cells/HSCs/megakaryocytes between lamprey and both zebrafish and mouse. B-like cells showed weaker similarity with B cells in mouse, in contrast to the pronounced alignment with B cells observed in zebrafish (Fig. 4a–d). Notably, in the lamprey T-like cell lineage cross-species mapping, the result showed complex relationships (Fig. 4b, d). To delve deeper into the cross-species mapping of T-like cell lineages, we conducted a functional enrichment analysis on the gene pairs that were significantly enriched in their homology between lampreys and either zebrafish or mouse. Our findings indicate that the enriched gene pairs within lamprey T-like cell subtype 1 exhibit functions that are consistent with T cell differentiation processes observed in zebrafish and mouse (Supplementary Fig. 8a). Whereas enriched gene pairs in MPL-L+ T-like cells/T cells homologous cell type pair shared RNA processing and ribosome biogenesis between lamprey and zebrafish/mouse (Supplementary Fig. 8b). We also observed that the hematopoietic cells of lamprey and HSCs of zebrafish displayed enrichment of gene pairs associated with the terms of erythrocyte development (Supplementary Fig. 8c). Similarly, when comparing the hematopoietic cells of lamprey with the megakaryocytes of mouse, we found an enrichment of gene pairs related to the process of blood coagulation (Supplementary Fig. 8c). Furthermore, we dissected the distribution and expression of the *mpl/Mpl/MPL-L* gene within zebrafish/mouse and lamprey. It was shown that the vast majority of MPL-L+ T-like cells and hematopoietic cells express the *MPL-L* gene in lamprey, whereas only HSCs/megakaryocytes expressed this gene in zebrafish/mouse, respectively (Fig. 4e, f). In addition, we also detected the expression pattern of *MPL-L* gene in the immune cell atlas of white-spotted bamboo shark, Chinese tongue sole, and long-tailed macaque, and no MPL-L+ T cells were identified in these species (Supplementary Fig. 8d).

To verify whether MPL-L+ T-like cells exist only in lamprey, we performed *in situ* hybridization and colocalization analysis on several organs (liver, head kidney, thymus, and gill) of zebrafish. The results showed that *mpl* was not colocalized with T cell marker *lck* (Supplementary Fig. 8e). We also prepared an anti-Lm-MPL-L polyclonal antibody and examined the colocalization between the MPL-L and VLRA in lamprey by immunofluorescence (Supplementary Fig. 9a, b).



**Fig. 3 | Transcriptional characterization and distribution of lymphoid cell subpopulations.** **a** UMAP visualization shows the distribution of T-like cells and B-like cells after reintegration. **b** UMAP visualization shows the lymphoid cell subpopulation distribution after annotation. **c** Feature plot of selected marker genes across various lymphoid cell subpopulations. **d** Radar chart shows the genes selected in the enrichment pathway of different lymphoid cell subpopulations. The farther each dot is from the center position, the higher the expression value is. **e** Feature plot of the *MPL-L* and *VLRA* gene co-expression. **f** Schematic diagram depicting the involvement of the TPO/MPL signaling pathway in the biological processes of jawed vertebrates and the expression of the *MPL-L* gene in lamprey

T-like cells. **g** UMAP shows the lymphoid cell subpopulation distribution in the three tissue types. The bar plot shows the number of cells in each lymphoid cell subpopulation in three tissue types. **h** The proportion of each lymphoid cell subpopulation in the three tissue types. The dot indicates multiple tissue samples. The box plot displays the minima, maxima, central tendency, quartile ranges, and whiskers of the data. The data presented are based on a minimum of three biological replicates per organ type: intestines ( $n = 5$ ), gills ( $n = 6$ ), and blood ( $n = 3$ ). Statistical significance was determined using one-way ANOVA with a one-sided test approach and Tukey's adjustment was made for multiple comparisons.



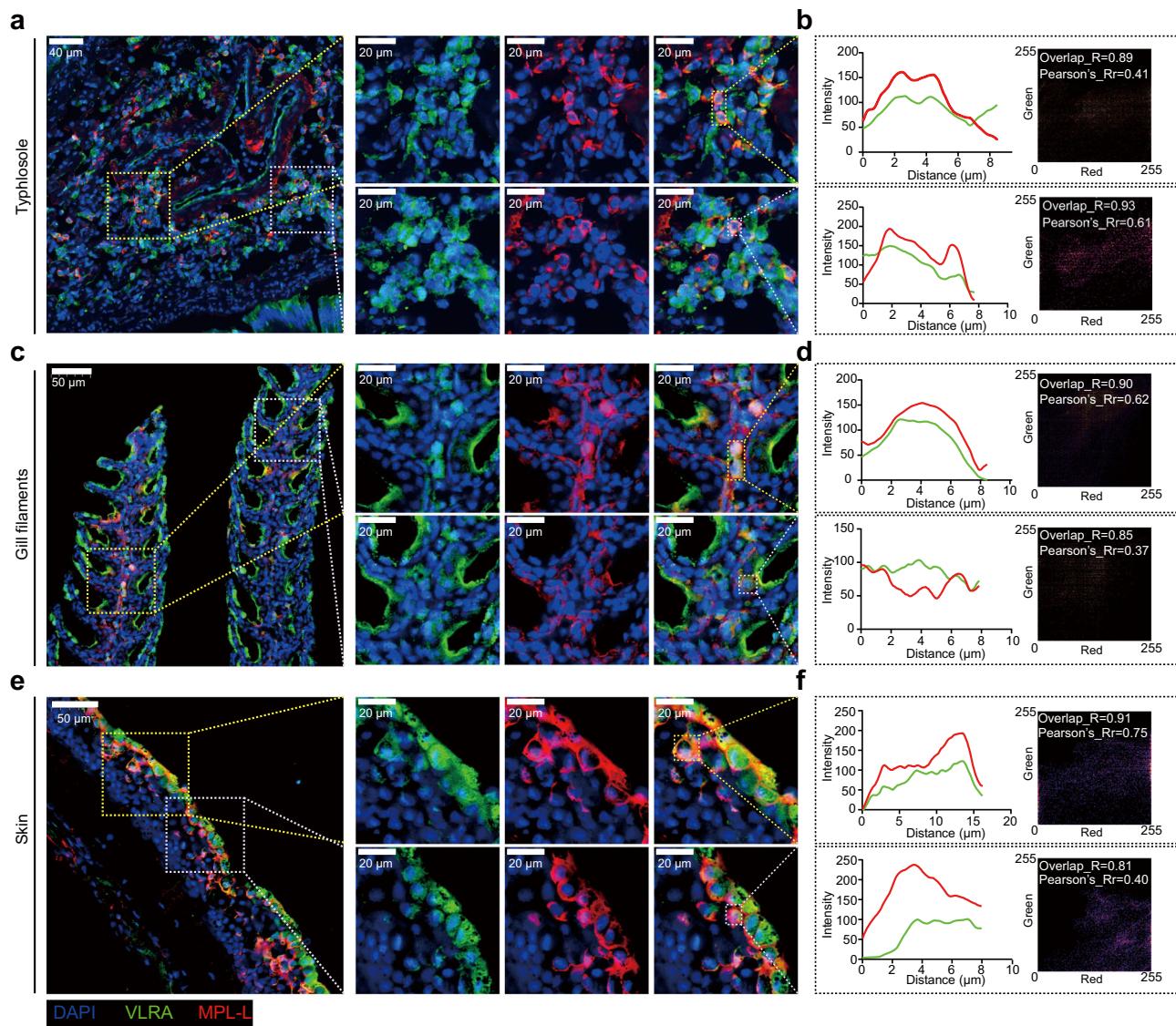
**Fig. 4 | Cross-species analysis between *L. morii* and zebrafish/mouse.** **a** Left: UMAP of lamprey and zebrafish immune cells; right: UMAP of the combined lamprey (orange) and zebrafish (blue) manifolds, with cell types circled. **b** Sankey plot summarizing the lamprey - zebrafish cell types mappings. Ignore edges with alignment scores < 0.1. **c** Left: UMAP of lamprey and mouse immune cells; right: UMAP of the combined lamprey (orange) and mouse (blue) manifolds, with cell types circled. **d** Sankey plot summarizing the lamprey - mouse cell types mappings.

Ignore edges with alignment scores < 0.1. **e** Bar plot displays the proportion of cells expressing *MPL-L*/*mpl* and the normalized expression of the *MPL-L*/*mpl*/*Mpl* gene in various cell types of *L. morii*, zebrafish, and mouse. **f** Expression of the *MPL-L*/*mpl*/*Mpl* homologous gene pairs between *L. morii* (purple) and zebrafish/mouse (blue) linked by SAMap is shown on the combined UMAP projection. Cells that highly express the *MPL-L* gene are highlighted with circles. Cells mapped together are labeled cyan.

We observed that *MPL-L* and VLRA were colocalized in the typhlosoles, gill filament, skin, kidney, and liver (Fig. 5 and Supplementary Fig. 9c). Based on these results, we suggest that the inherent expression of *MPL-L* in lamprey lymphocyte-like cells constitutes a remarkable and intriguing revelation, indicating a potential activation signal for lamprey T-like cells and the evolutionary divergence of lymphocyte lineage across vertebrates.

### ***MPL-L*<sup>+</sup> VLRA<sup>+</sup> cells responses to the phytohemagglutinin-L (PHA-L) and recombinant human TPO (rhTPO)**

Immunofluorescence colocalization analysis confirmed the distribution of *MPL-L*<sup>+</sup> VLRA<sup>+</sup> cells in almost all of the immune tissues. To explore the possibility of lamprey *MPL-L* combining with human TPO, we performed molecular docking predictions to explore the interactions between human TPO and lamprey *MPL-L*. Utilizing



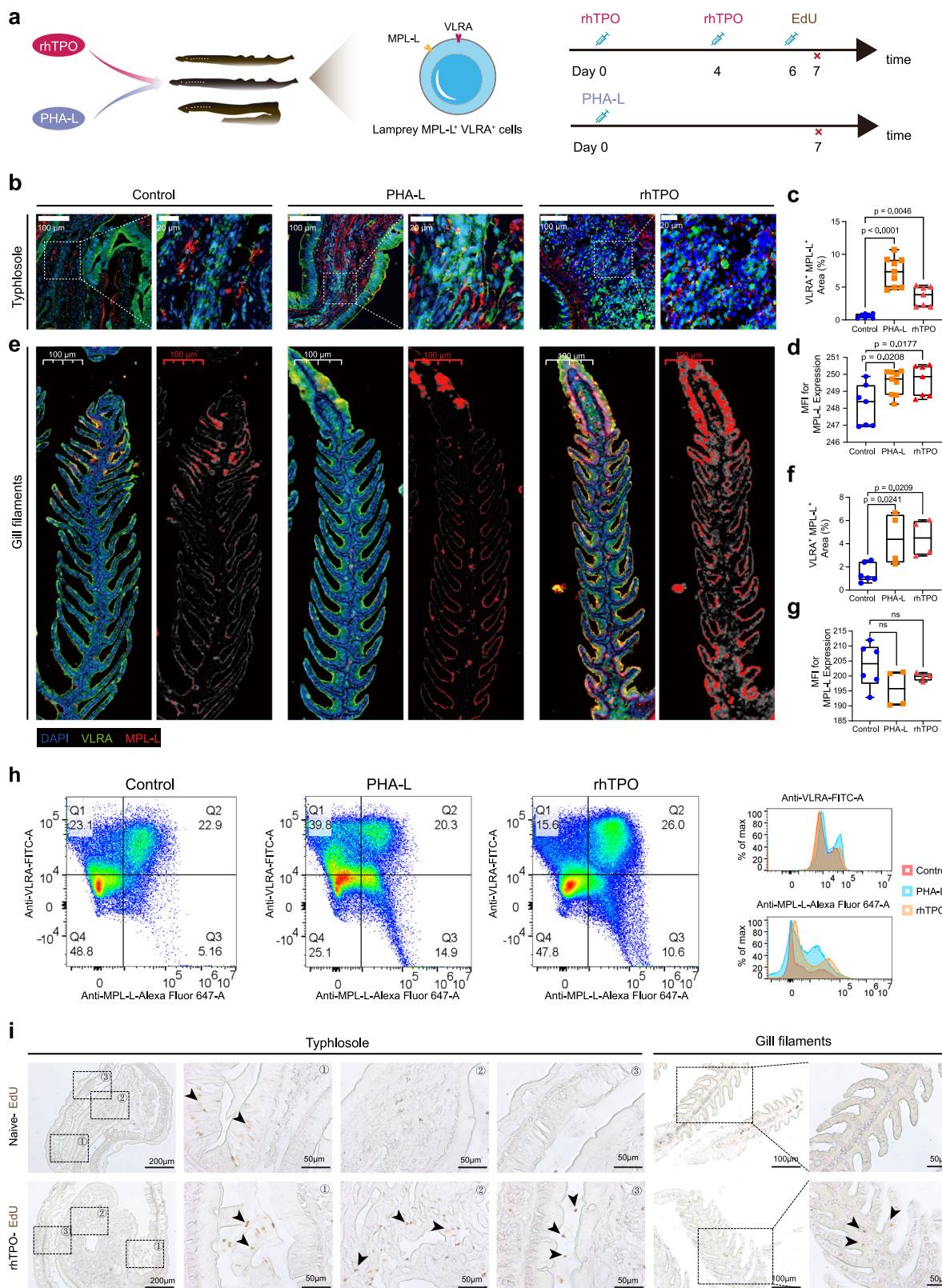
**Fig. 5 | Double-stained immunofluorescence examination of MPL-L<sup>+</sup> VLRA<sup>+</sup> cells.** a Double-stained immunofluorescence examination of MPL-L and VLRA in MPL-L<sup>+</sup> VLRA<sup>+</sup> cells in the typhlosole with experiments performed at least three times. Green fluorescence is associated with anti-VLRA pAb, and red fluorescence is associated with anti-MPL-L pAb. b Left: The intensity trajectory line charts are drawn by selecting two immunofluorescence colocalization regions in the typhlosole. The green line indicates anti-VLRA pAb and the red line represents anti-MPL-L pAb. Right: Scatterplot displays correlations of anti-VLRA pAb with anti-MPL-L pAb colocalization in the typhlosole. c Double-stained immunofluorescence examination of MPL-L and VLRA in MPL-L<sup>+</sup> VLRA<sup>+</sup> cells in the gill filaments with

experiments performed at least three times. d Left: The intensity trajectory line charts are drawn by selecting two immunofluorescence colocalization regions in the gill filaments; right: Scatterplot displays correlations of anti-VLRA pAb with anti-MPL-L pAb colocalization in the gill filaments. e Double-stained immunofluorescence examination of MPL-L and VLRA in MPL-L<sup>+</sup> VLRA<sup>+</sup> cells in the skin with experiments performed at least three times. f Left: The intensity trajectory line charts are drawn by selecting two immunofluorescence colocalization regions in the skin. Right: Scatterplot displays correlations of anti-VLRA pAb with anti-MPL-L pAb colocalization in the skin.

AlphaFold2, we performed a composite structure prediction for the MPL-L of *L. morii* and human TPO. The model-1 with the rank “0” prediction was selected. The resulting model PDB files were visualized using PyMOL, enabling us to analyze and annotate the amino acid binding sites between the ligand and receptor. Upon examination of model-1, colored according to the predicted local distance difference test (pLDDT) scale, we observed that the main body of the human TPO/lamprey MPL-L complex exhibited high confidence. Notably, several residues within human TPO were positioned at the interface of the complex, and hydrogen bonds were observed between residues of human TPO and lamprey MPL-L (Supplementary Fig. 10a). These observations suggest a potential binding interaction between human TPO and lamprey MPL-L. We also identified a potential MPL ligand, nuclear distribution C (NuDC), in lampreys,

which is also found in humans where it is believed to support megakaryocyte maturation alongside TPO. The lamprey NuDC exhibits extensive expression in the lamprey atlas and shows significant conservation in protein domains compared to the human protein (Supplementary Fig. 10b, c). Despite determining the binding sites of the lamprey NuDC/MPL-L complex, the confidence of structural model is low (Supplementary Fig. 10d).

To further verify whether MPL-L<sup>+</sup> VLRA<sup>+</sup> cells have T-cell antigen response characteristics and are responsive to human TPO, we performed the intraperitoneal injection of PHA-L and rhTPO into ammocoete larvae (Fig. 6a). The immunofluorescence assay results showed that stimulation with PHA-L and rhTPO induced an increase in the area of colocalization of VLRA and MPL-L and facilitated the expression of MPL-L in typhlosole (Fig. 6b-d). The results also showed that the



proportion of MPL-L<sup>+</sup> VLRA<sup>+</sup> cells increased in the gill filaments after injection (Fig. 6e, f). However, we have not observed a significant change in the expression levels of MPL-L within the MPL-L + VLRA+ cell population in the gill filaments (Fig. 6g). Additionally, flow cytometric analysis of cells stained with anti-VLRA and anti-MPL-L also showed that the proportion of MPL-L<sup>+</sup> VLRA<sup>+</sup> cells was increased after rhTPO injection (Fig. 6h and Supplementary Fig. 10e). Besides, we noted that

PHA-L immunostimulation promoted the proliferation of VLRA<sup>+</sup> cells and MPL<sup>+</sup> cells (Fig. 6h and Supplementary Fig. 10e). This stimulus may promote the proliferation of cells at earlier or later stages of the differentiation process, with these cells expressing MPL-L or VLRA individually. The 5-Ethynyl-2'-deoxyuridine (EdU) injection also showed that there was an increase in cell proliferation in the typhlosole and gill filament after rhTPO injection (Fig. 6i).

**Fig. 6 | Stimulation by the recombinant human TPO (rhTPO) and phytohemagglutinin-L (PHA-L) leads to the proliferation of MPL-L<sup>+</sup> VLRA<sup>+</sup> cells.** **a** Time points for injection with rhTPO, PHA-L, and the 5-Ethynyl-2'-deoxyuridine (EdU). **b** Representative sections revealing the number and location of MPL-L<sup>+</sup> VLRA<sup>+</sup> cells in the typhlosole after injection, as determined by double-stained immunofluorescence examination. Green fluorescence is associated with anti-VLRA pAb, and red fluorescence is associated with anti-MPL-L pAb. **c, d** Box plot shows the proportion of MPL-L<sup>+</sup> VLRA<sup>+</sup> cells and the mean fluorescence intensity (MFI) of MPL-L within the typhlosoles across varying experimental conditions. The data presented are based on a minimum of three biological replicates per experimental sample: Control ( $n = 7$ ), PHA-L ( $n = 9$ ), and rhTPO ( $n = 7$ ). Statistical significance was determined using one-way ANOVA with a one-sided test approach and Tukey's adjustment was made for multiple comparisons. **e** Representative sections

revealing the number and location of MPL-L<sup>+</sup> VLRA<sup>+</sup> cells in the gill filaments after injection, as determined by double-stained immunofluorescence examination. **f-g** Box plot shows the proportion of MPL-L<sup>+</sup> VLRA<sup>+</sup> cells and the MFI of MPL-L within the gill filaments across varying experimental conditions. The data presented are based on a minimum of three biological replicates per experimental sample: Control ( $n = 6$ ), PHA-L ( $n = 4$ ), and rhTPO ( $n = 4$ ). Statistical significance was determined using one-way ANOVA with a one-sided test approach and Tukey's adjustment was made for multiple comparisons. **h** Flow cytometric analysis of MPL-L<sup>+</sup> VLRA<sup>+</sup> cells before and after PHA-L and rhTPO immunostimulation. **i** Representative sections reveal the number and location of proliferating cells (EdU, brown) in the typhlosoles and gill filaments after the injection of rhTPO with experiments performed at least three times. Arrows indicate the characteristic proliferating cells.

## Distinctive molecular profiles and developmental trajectory of MPL-L<sup>+</sup> T-like cells

Having characterized *L. morii* T-like cell subclusters, it was found that MPL-L<sup>+</sup> T-like cells showed differences in gene expression and tissue distribution when compared to other T-like cell subclusters. We next explored the dynamic cell transitions in T-like cell subpopulations by inferring the state trajectories using Monocle2. Using the trajectory information, this trajectory was determined to start with T-like cells subtype 3, and finally, most MPL-L<sup>+</sup> T-like cells reached a different terminal state compared to the terminal states of the other three T-like cell subpopulations (Fig. 7a, b). Simultaneously, we calculated the proportion of cells expressing *VLRA* genes ( $n_{UMI} > 0$ ) and the average expression of *VLRA* genes in each T-like cell subpopulation. It was observed that the proportion of T-like cells subtype 3 expressing *VLRA* genes was more modest, and the average expression of *VLRA* genes in T-like cell subtype 3 ranked the lowest of all T-like cell subpopulations. The expression level of *VLRA* genes in T-like cells subtype 2 was slightly greater than that in T-like cells subtype 3, whereas T-like cells subtype 1 and MPL-L<sup>+</sup> T-like cells were primarily the terminal states (Fig. 7c, d).

We further delineated the heterogeneity of MPL-L<sup>+</sup> T-like cells in three tissues. MPL-L<sup>+</sup> T-like cell populations in the intestines, blood, and gills were distinguished by DEGs. For example, MPL-L<sup>+</sup> T-like cells in gills highly expressed genes such as *JUN*, *IERSL*, *KLF7*, and *ATF3/JDP2*, while MPL-L<sup>+</sup> T-like cells in blood expressed genes such as *ZEB1/2*, *RICTOR*, *ITGA4/9*, *RPS5*, and *HSPA9*. Intestinal MPL-L<sup>+</sup> T-like cells expressed genes such as *GRN*, *FASN*, *TLN1/2*, and *GRTP1* (Supplementary Fig. 11). To further confirm the heterogeneity of MPL-L<sup>+</sup> T-like cells across three tissues, we also utilized Monocle 2 trajectory analysis to determine the transition state of MPL-L<sup>+</sup> T-like cells within these tissues. The results showed that these cells were in a unique differentiation state (Fig. 7e). Combined with the trajectory transition information and the proportion of cells expressing *VLRA* genes ( $n_{UMI} > 0$ ), we extrapolated that MPL-L<sup>+</sup> T-like cells exhibited transcriptional and developmental diversity in their respective tissues (Fig. 7f). During the transformation of MPL-L<sup>+</sup> T-like cells on the gills, blood, and intestines, three tissue-specific gene expression patterns were formed (Fig. 7g). For example, gill MPL-L<sup>+</sup> T-like cells highly expressed genes involved in T-like cell activation, such as *FOSL2*, *JUN*, *ATF3/JDP2*, *SOX4*, and *ID2* (Fig. 7g, h). It was reported that *FOSL2* is highly expressed in early-stage T cells and plays an important role in T-cell activation, proliferation, differentiation, and transformation in higher vertebrates<sup>21</sup>. Additionally, *CYGB*, *ZEB1/2*, *ZBTB39*, and *SELL* were expressed in blood MPL-L<sup>+</sup> T-like cells. Compared with those in gills and blood, the MPL-L<sup>+</sup> T-like cells in the intestine highly expressed innate immune-related genes, such as *GRN*, *MMP19*, and *LYZL1*. The T-cell co-stimulation molecule *CD63* and the signature transcription factor of innate lymphocyte lineages *ZBTB16* were also expressed in intestinal MPL-L<sup>+</sup> T-like cells (Fig. 7g, h).

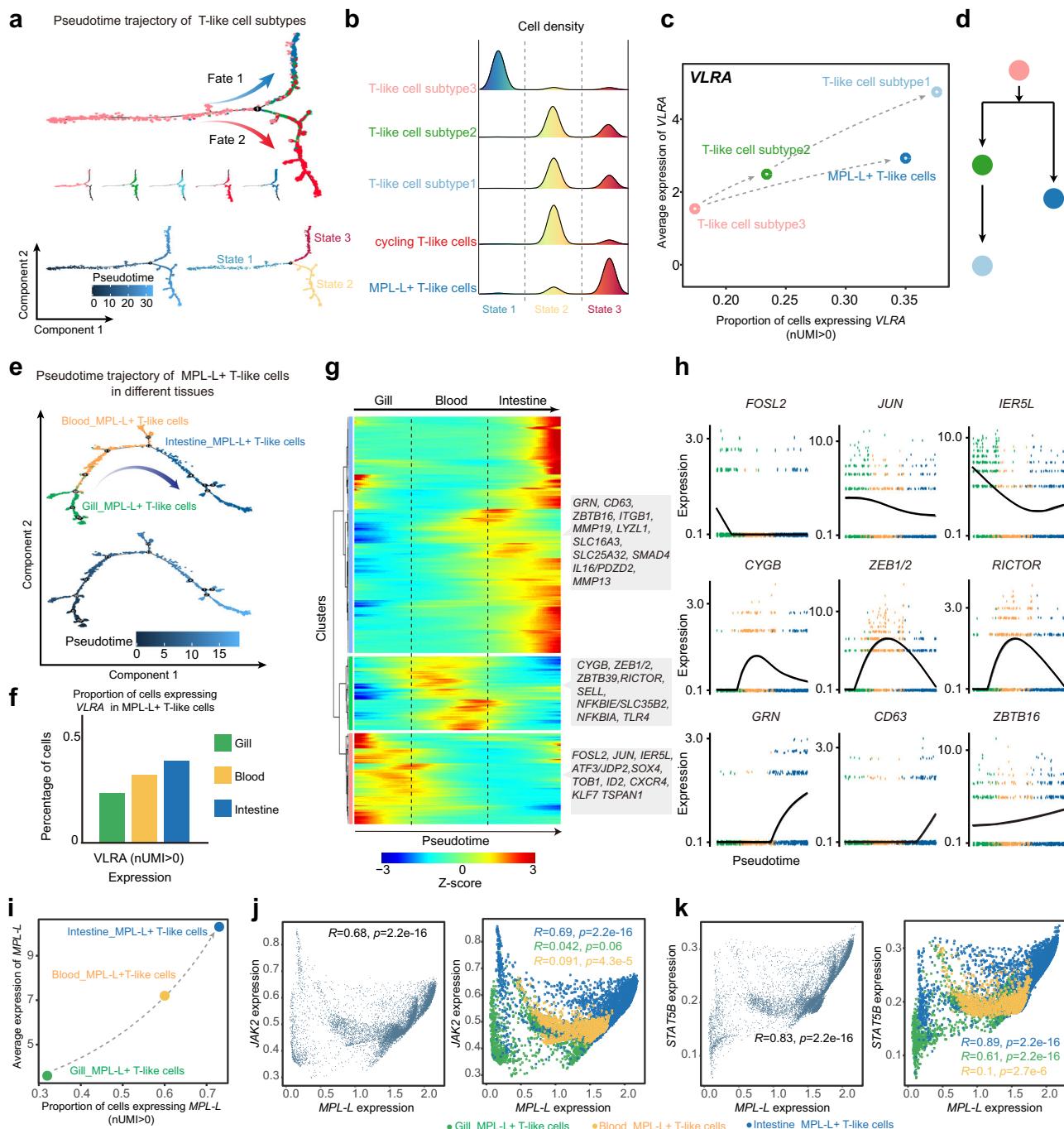
To further investigate the potential role of the *MPL-L* gene in MPL-L<sup>+</sup> T-like cells of different tissues and its regulatory mechanism, we determined the proportion of cells expressing the *MPL-L* gene and the

average expression of this gene in three MPL-L<sup>+</sup> T-like cell subpopulations. We found that the *MPL-L* gene experienced a considerable upward trend following the trajectory of MPL-L<sup>+</sup> T-like cells in the three tissues (Fig. 7i). Moreover, we carried out expression correlation analysis with MAGIC to confirm that *MPL-L* expression was positively associated with its downstream genes *JAK2* and *STATSB* in MPL-L<sup>+</sup> T-like cells (Fig. 7j, k)<sup>17,22,23</sup>. Notably, the positive correlation observed in intestinal MPL-L<sup>+</sup> T-like cells was more pronounced than in their gill and blood counterparts (Fig. 7j, k). These findings suggest a potential MPL-L-based signal transduction system within MPL-L<sup>+</sup> T-like cells, which may orchestrate the development or activation of intestinal MPL-L<sup>+</sup> T-like cells.

## Discussion

In this work, we combined full-length and single-cell sequencing to generate a comprehensive single-cell atlas of the immune tissues of *L. morii*, with a particular focus on T-like cell populations. This cell atlas advances the exploration of adaptive immunity in jawless vertebrates. Our in-depth analysis of T-like cell subtypes revealed the presence of MPL-L<sup>+</sup> T-like cells, a unique subtype characterized by the inherent expression of *MPL-L* gene and hematopoietic genes. Furthermore, our work offers information involving the transcriptomic profile, cross-species parallels, tissue-specific distribution, developmental trajectories, and specific antigen reactions to define the identity and function of the MPL-L<sup>+</sup> T-like cells.

While it has been established that the adaptive immune systems of jawless and jawed vertebrates share similar lymphocyte lineage differentiation characteristics and some molecular markers, the lymphoid cell subtypes are obviously different<sup>5</sup>. Our analysis indicated that T-like cells exhibit high heterogeneity among the three tissues based on the different gene expression patterns, which suggested that compared to B-like cells, T-like cells have more flexible and diverse cellular states and functions. Intriguingly, B-like cells did not display diverse cell subtypes such as T-like cells, and the gene expression patterns of B-like cells in three tissues were extremely conserved, which means that lamprey plasma cells were either not detected or they do not have a specific transcript signature such as those in jawed vertebrates. Notably, our research is based on the comparison of the constant regions of the *VLR* gene to identify *VLRA/B/C* genes. A potential limitation of this approach is the inability to accurately distinguish between germline *VLR* genes and functional *VLR* receptor sequences that have undergone rearrangement and assembly. This is due to the high similarity between germline genes and functional receptor sequences in the constant region. However, germline *VLR* gene transcripts also exhibit cell-specific expression patterns. For example, both germline *VLRB* genes and assembled *VLRB* genes are specifically expressed in *VLRB* cells<sup>7</sup>. This suggests that although we cannot precisely distinguish between germline genes and functional receptor sequences through constant region alignment, we can use this information to identify specific lymphoid cell types. In our study, we did observe expression patterns of *VLRA/B/C* genes in specific cell



**Fig. 7 | Analysis of MPL-L+ T-like cell transition states.** **a** Developmental pseudotime trajectory of five T-like cell subpopulations. 2D pseudotime plot showing the dynamics of five T-like cell subpopulations (upper panel). The color indicates the differentiation degree of the five cell types (lower left panel). The deeper color indicates the early state. The color of the bottom right figure shows three branches of those states (lower right panel). **b** The ridge plot shows the distribution of cell density by five T-like cell subpopulations during the transition (divided into 3 states), along with the pseudotime. **c** The figure shows the change in VLRA gene expression in five T-like cell subpopulations over pseudotime with cell proportion expressing VLRA on the x-axis and its average expression on the y-axis. **d** The diagram shows the cell transitions of T-like cell subpopulations along with the pseudotime. **e** Developmental pseudotime trajectory of MPL-L+ T-like cells across gills, blood, and intestines with the top figure color-coded by tissue and the bottom by differentiation level. **f** Bar plot displays the proportion of cells that expressed the VLRA gene (nUMI > 0) in MPL-L+ T-like cells of gills, blood, and intestines.

**g** Heatmap of gene expression arranged along the pseudotime trajectory. **h** Pseudotime plots of selected genes. **i** The figure shows the change in *MPL-L* gene expression in MPL-L+ T-like cells in three tissue types over pseudotime with cell proportion expressing *MPL-L* on the x-axis and its average expression on the y-axis. **j** Scatterplot on the left displays correlations of *MPL-L* expression with *JAK2* expression in MPL-L+ T-like cells. The scatterplot on the right displays correlations of *MPL-L* expression with *JAK2* expression in MPL-L+ T-like cells across gills, blood, and intestines. For correlation analysis, gene expression data were imputed by MAGIC. Utilizing Spearman's rank correlation coefficient to measure the strength and direction of the association between the two variables. The statistical test used was two-sided. **k** Scatterplot on the left displays correlations of *MPL-L* expression with *STAT5B* expression in MPL-L+ T-like cells. The scatterplot on the left displays correlations of *MPL-L* expression with *STAT5B* expression in MPL-L+ T-like cells across gills, blood, and intestines.

types, and we also found that the expression patterns of *VLRA* and *VLRC* genes were very similar, which is consistent with the reported expression of *VLRA/VLRC* transcripts in both VLRA and VLRC cells<sup>7</sup>. Therefore, we believe that the *VLRA/B/C* gene expression patterns observed through scRNA-seq data reflect the actual expression of these genes in cells, despite our inability to precisely distinguish between germline genes and functional receptor genes.

T-like cells showed several functional specialization subgroups, including T-like cell subtype 1, cycling T-like cells, T-like cell subtype 2, T-like cell subtype 3, and MPL-L+ T-like cells. T-like cell subtype 2 specifically expressed homologs of the regulatory T-cell selection molecule *THEMIS1/2/3* and the stem cell growth factor receptor *CSF1R/KIT*, which indicated that T-like cell subtype 2 are probably present in the early stage of development<sup>24,25</sup>. Cycling T-like cells expressed *TOP2A/B*, *PCNA*, *RRM2*, and other homologs that play a key role in T-cell division, indicating that cycling T-like cells are involved in the cell cycle<sup>26–28</sup>. *CXCR1/2/3* homolog is a chemokine receptor that is highly expressed on effector T cells in jawed vertebrates and plays an important role in T-cell transport and immune effects<sup>29</sup>. Our results showed that the *CXCR1/2/3* homolog and *VLRA* gene are co-expressed in T-like cell subtype 1, indicating that T-like cell subtype 1 are present in the late period of development and exert an adaptive immune effect through *CXCR1/2/3* receptor signaling. Interestingly, the above T-like cell subpopulations are mainly concentrated in the gills, which further confirms the claim that thymus-like structures presented in lamprey are similar to that of jawed thymus, both housing a variety of T cell subsets at different developmental phases.

A striking outcome of our study is the identification of MPL-L+ T-like cells, a special group of T-like cells that chiefly exist in the blood and intestines. They do not express the conventional T-cell genes of jawed vertebrates but highly express genes related to hematopoietic function (*MPL-L*, *GM14681/NPM1/3*, *KDEL1/MYH9/10/11/14*, *ITGB3*, *VAV2*)<sup>30,31</sup> and many ribosome-related genes (*RPS16*, *RPL7A*, *RPS25*, etc.)<sup>32</sup>. We noticed that *MPL-L* expression in a lymphoid lineage is remarkably specific. MPL is involved in the self-renewal, expansion, and maintenance of HSCs and the production and maturation of megakaryocyte-derived platelets but is not expressed in lymphocytes<sup>14–16</sup>. The combination of the hematopoietic growth factor receptor MPL and its ligand TPO activates signal transduction pathways, such as JAK2/STAT, PI3K/Akt, and Raf-1/MAPK, which overlap with the common signaling pathway of costimulatory molecules and cytokine production during T-cell activation<sup>33,34</sup>. It has been reported that the transfer of the *Mpl* gene into mouse T cells can be used as a costimulatory signal and cytokine signal for T cells to exert immune effects, which can significantly enhance persistence and antitumor ability<sup>19</sup>. Our research analyzed the conserved protein domains of MPL-L in *L. morii* and jawed vertebrates. The MPL-L protein domains in lampreys are consistent with those in other jawless and jawed species, with both EpOR\_lig-bind and FN3. We found that the EpOR\_lig-bind domain presented in the prolactin receptor of amphioxus, and also contains Ig3 and FN3 domains. The presence of the EpOR domain in amphioxus suggests that this domain was likely present in the common ancestor of extant chordates. However, in the prolactin receptor and thrombopoietin receptor of vertebrates, the Ig3 domain is not included. Previous studies have suggested that the first genome duplication event occurred during the evolution of chordates into vertebrates. Therefore, we speculate that lampreys may have gained two copies of the prolactin receptor gene during the whole genome duplication (WGD) event, and subsequently lost Ig3. It is also possible that Ig3 was lost at the same time when two gene copies were obtained during evolution. After undergoing duplication, deletion, and recombination events, lampreys gradually adapted to functional evolution to produce two genes with different functions, the prolactin receptor and the thrombopoietin receptor. Notably, the distribution of EpOR\_lig-bind and FN3 domains in all species of thrombopoietin receptors is

similar and different from that in prolactin receptors. Although *L. morii* exhibits conservation in the MPL protein domain akin to that of jawed vertebrates, the absence of experimental data regarding the functional characterization of MPL and the specific ligand identification within *L. morii* leads us to cautiously designate it as MPL-like. More functional studies will be needed in the future to clarify its functional role and its relationship to known MPL proteins.

MPL-L and VLRA protein were co-expressed in the gill filaments, typhlosoles, kidney, liver, and skin, indicating that *MPL-L* molecules are a natural feature in the T-like cells of lamprey. The abundant and specific expression of *MPL-L* molecules in T-like cells suggests their critical function. Lampreys injected with PHA-L exhibited MPL-L+ VLRA+ cell expansion and proliferation in vivo, suggesting that MPL-L+ VLRA+ cells can respond to specific antigens of T cells. Strikingly, stimulation with rhTPO, a classical ligand of MPL, also increased the number of MPL-L+ VLRA+ cells in the gill filaments and typhlosole. Although our current data is not sufficient to conclusively demonstrate that the response of MPL-L+ VLRA+ cells depends on the direct interaction between rhTPO and lamprey MPL-L, our molecular docking model offers a preliminary framework for the potential interactions between rhTPO and lamprey MPL-L. However, these models are not adequate as strong evidence of a biologically significant association. This conclusion is influenced by several factors, including the relatively low homology (approximately 25%) between the lamprey *MPL-L* gene and its human counterpart, which may limit the applicability of human models to lamprey research. Moreover, we must acknowledge the inherent limitations in predicting in vivo interactions based on in silico predictions. Additionally, the potential cross-reactivity of rhTPO with structurally similar lamprey MPL-L paralogs raises the possibility of indirect effects on T-cell-like cells. This factor further complicates the interpretation of our findings and suggests that the observed responses may not be entirely attributed to a single direct interaction. Future research should include functional assays and potentially the development of native ligands for lamprey MPL-L to gain a deeper understanding of the direct and indirect impacts on immune cell populations.

Our data suggest that MPL-L+ T-like cells in gills, blood, and intestines exist in different developmental states. Among them, MPL-L+ T-like cells in the gills mainly upregulate the homologs of TFs, such as *FOSL2*, *JUN*, *ATF3/JDP2*, *SOX4*, and *ID2*, which were found to be highly differentially expressed in early T cells and coordinated to regulate the activation and differentiation of T cells<sup>35–38</sup>. MPL-L+ T-like cells in the blood are in the intermediate differentiation state according to the pseudotime analysis and highly expressed homologs with the naive T-cell characteristic gene *SELL*<sup>39</sup>. As a costimulatory molecule of T-cell activation in jawed vertebrates, *CD63* is generally highly expressed in the process of effector T-cell activation<sup>40</sup>. Our results showed that intestinal MPL-L+ T-like cells upregulated the homolog of the costimulatory molecule *CD63* and high expression of many innate immune effector molecules (*GRN*, *ZBTB16*, *MMP19*, *LYZL1*, *MMP13*, etc.) was observed, suggesting that intestinal MPL-L+ T-like cells may perform immune functions by interacting with innate immune effector molecules<sup>41,42</sup>. We infer from the results of bioinformatics analysis that MPL-L+ T-like cells show distinct differentiation states among three tissues and may develop in the thymoid of lampreys, similar to the developmental trajectory of T cells in jawed vertebrates. However, the conclusion regarding the early development of lymphocyte lineages, especially MPL-L+ T-like cells, in jawless vertebrates, still requires further analysis and exploration. A potential strategy for studying the lineage of MPL-L+ T-like cells is to analyze scRNA-seq data from different developmental stages of lamprey larvae, perhaps coupled to spatial omics approaches. However, it should be recognized that the difficulty in extracting larval tissue directly affects the use of single-cell sequencing technology to determine their developmental trajectory. Besides omics analyses, a potentially promising means to trace the

proliferation, differentiation, and migration of MPL-L+ T-like cells in lampreys is the use of in vivo imaging technology to be explored.

Considering the rapid evolution of immune receptors and signaling mechanisms poses challenges to the reliability of many methods, we believe that SAMap is a suitable tool for our purposes. We chose SAMap primarily due to its remarkable capabilities and adaptability in handling intricate biological data. Based on the self-assembly manifold algorithm, SAMap enables identification of the shared expression program of homologous cell types of distant phylogenetic species and can reliably compare species with long evolutionary distances<sup>43</sup>. SAMap has been applied to analyze homologous cell types among humans, mice, and zebrafish, revealing the conservation of immune cells in vertebrates<sup>44</sup>. In addition, the SAMap analysis results showed that there is a good similarity between the innate immune cells and hematopoietic cells/HSCs/megakaryocytes of agnatha and gnathostomata vertebrates, which is consistent with the cognition that hematopoiesis and innate immunity are relatively conserved. This indicates that SAMap also has a good effect on the cross-species analysis of the immune system.

The key features of the adaptive immune systems in jawless and jawed vertebrates are surprisingly similar, including T/B lymphocyte lineages and their antigen response. Our analysis further corroborates the preservation of lymphocyte subtype diversity across these groups. Nonetheless, T-like cells in lamprey exhibit a distinct classification and functional profile compared to T cells in jawed vertebrates, notably lacking the distinct lineage markers of CD4<sup>+</sup> and CD8<sup>+</sup> T cells. Utilizing the expression signatures of T-cell markers and genes associated with hematopoiesis, we have delineated two categories: conventional T-like cells and novel MPL-L+ T-like cells. The latter demonstrates a clear divergence from other T-like cell subtypes in jawless vertebrates and may constitute a significant departure from the T-cell repertoire of jawed vertebrates. This distinction could be indicative of a fundamental divergence in lymphocyte lineages between these two groups apart from their unique antigen receptors. The distinctive gene expression profile of MPL-L+ T-like cells, alongside the hypothesized role of MPL signaling in T-cell function, not only deepens the understanding of the adaptive immune system of jawless vertebrates but also provides new insight into the evolution of adaptive immunity. The discovery of MPL-L+ T-like cells may improve the development of innovative treatment strategies for faulty immune functions.

However, there are still many unresolved issues that need to be further investigated. We did not find the *TPO* gene in the full-length transcriptome of the *L. morii* and other agnathan vertebrate genomes. Analysis of the TPO domain revealed that this gene is present in cartilaginous fish. The absence of TPO in agnathans may be attributed to the poor quality of the agnathan species genome or because it emerged during the second whole-genome duplication process. Interestingly, we found another ligand of MPL, NuDC, in the lamprey. In humans, NuDC functions as a cytokine by specifically binding to MPL, and it is hypothesized to work in concert with TPO to promote megakaryocyte maturation<sup>45,46</sup>. Our research revealed that lamprey NuDC is expressed extensively across a majority of cells, exhibiting remarkable conservation in its protein domain when compared to human NuDC protein domains. Furthermore, we determined the structure of the lamprey NuDC/MPL-L complex and pinpointed the binding sites secured by hydrogen bonds. Nevertheless, the confidence level in the lamprey NuDC/MPL-L complex structure was found to be low. One possible explanation for this low confidence level could be the inherent structural differences between human and lamprey MPL-L receptors. Such differences could alter the way NuDC binds to MPL-L, affecting the stability and conformation of the resulting complex. Given these considerations, it is reasonable to speculate that the binding of NuDC to MPL-L in lampreys may differ from that observed

in humans. While our research has provided valuable insights into the structural conservation of NuDC across species, further investigations are warranted to elucidate the precise nature of NuDC/MPL-L interactions in lampreys and to determine whether they indeed work in lamprey MPL-L lineage as hypothesized. Additionally, although T-like cells are similar to T cells, proving the similarities still requires more precise molecular study. Further investigation is needed to determine the functional role of MPL-L+ T-like cells in jawless vertebrates, similar to the assistance role of CD4<sup>+</sup> T cells or the cytotoxic effect of CD8<sup>+</sup> T cells in jawed vertebrates. The downstream signal of MPL-L in lampreys has not been verified, and whether it can serve as an important signal for VLRA cell activation, similar to *Mpl*-transduced T cells in mice, remains unknown.

## Methods

### Biological samples and the ethical use of animals

Lamprey larvae (*L. morii*) in the ammocoete stage were purchased in Dandong, Liaoning Province, China, in May and then cultured in the laboratory with purified water for several days for subsequent experiments. The breeding of ammocoete larvae in the laboratory was conducted using a static water culture system, aerated by an air pump. The water temperature for the culture of ammocoete larvae was kept at  $19 \pm 2$  °C. The dissolved oxygen levels were maintained at no less than 5.0 mg/L, with an overall pH range of  $7.0 \pm 0.2$ , ensuring optimal water quality conditions. Larvae were anesthetized with 0.2 g/L MS-222 (Ethyl 3-aminobenzoate methanesulfonate) before the experiment. The experimental procedures using lamprey larvae were approved by the Animal Care and Use Committee at the Chinese Academy of Fishery Sciences (YSFRI-2023037), and all experimental procedures were performed in accordance with the Guidelines for the Care and Use of Laboratory Animals at the Chinese Academy of Fishery Sciences.

### Full-length RNA-seq

For Iso-seq, four tissue types were collected, including the liver, blood, intestine, and gill. To obtain more comprehensive gene information, mixed lamprey visceral tissues were also collected. Only high-quality RNA samples with RIN  $\geq 8.5$  were used in the subsequent analysis. cDNA was prepared following the manufacturer's guidelines and converted into the Iso-seq library by a Template Prep Kit (PacBio). Thereafter, the Pacific Bioscience Sequel II platform was used for sequencing. The initial data were processed using the SMRT analysis Iso-Seq (<https://github.com/PacificBiosciences/IsoSeq>) pipeline and produced 555.23 Gb of subreads (Supplementary data 1), which were classified into 13.22 Gb of nonchimeric circular consensus (CCS) reads, and the average CCS length was 3167.80 bp (Supplementary data 2). A total of 520.16 Mb (Supplementary data 3) of high-quality consensus isoforms were produced with the Iso-Seq pipeline. The redundancy of high-quality consensus isoforms was removed by cd-hit-est with parameter "-c 0.99 -G 0 -aL 0.90 -AL 100 -aS 0.99 -AS 30". Finally, 29,267 nonredundant high-quality transcripts were predicted to have open reading frames (ORFs) with TransDecoder (v5.5.0; <https://github.com/TransDecoder/TransDecoder>) and annotated by searching against public protein databases, including eggNOG (<http://eggnog5.embl.de/#/app/home>), Non-Redundant Protein Database (NR; <https://ftp.ncbi.nlm.nih.gov/blast/db/>), and SwissProt (<https://uniprot.org>) (Supplementary data 4). To further enhance the accuracy of annotation, we also searched the corresponding peptides of the genes used in the paper against Pfam (<https://pfam.xfam.org>)<sup>47</sup> using HMMER<sup>48</sup> and compared the related function domain with eight species, which included zebrafish (*Danio rerio*), medaka (*Oryzias latipes*), western clawed frog (*Xenopus tropicalis*), red junglefowl (*Gallus gallus*), American alligator (*Alligator mississippiensis*), Indochinese rhesus macaque (*Macaca mulatta*), mouse (*Mus musculus*), and human (*Homo sapiens*) (Supplementary data 5).

## Orthology assignment and gene nomenclature

Because *L. morii* has no published genome yet, we used the third-generation full-length transcriptome data as a reference for our study. However, the full-length transcriptome data cannot reflect how these genes are related to agnathans and gnathostomes. To establish the orthology relationships among all annotated genes and infer duplication events, homology information for the set of annotated genes was retrieved from several agnathan vertebrates and two gnathostome species: brown hagfish (*Eptatretus atami*), inshore hagfish (*Eptatretus burgeri*), pouched lamprey (*Geotria australis*), southern lamprey (*Mordacia mordax*), sea lamprey (*Petromyzon marinus*), Pacific lamprey (*Entosphenus tridentatus*), arctic lamprey (*Lethenteron camtschatcicum*), zebrafish and mouse. We aimed to analyze the accuracy and evolutionary relevance of homologous genes in agnathans and gnathostomes based on the differences in homologs among different species along the evolutionary axis. The specific methods are as below: firstly, we downloaded the corresponding protein sequences of annotated genes and genomic information of the above vertebrates from NCBI. Then applied the Miniprot (<https://hub.njuu.cf/lh3/miniprot>) annotation pipeline to map the protein sequence to whole genomes. To strictly control the standard of homologous genes, we believe that the homologs with a gene length match a degree of more than 80%. In the final step, we counted the number of homologous genes at different gene loci (Supplementary data 6).

Due to most of the lamprey genes lack a clear and consistent nomenclature. We considered the specific gene names of jaw vertebrates as references. The complexity of gene matching is further compounded by the multiple rounds of whole genome duplication (WGD) events that have occurred in teleosts. In contrast to zebrafish, which have undergone numerous teleost-specific genome duplications (TGD), and other cartilaginous fish, whose gene functions may be of lower quality or incomplete, both humans and mice boast highly accurate and comprehensive genome databases. This makes it a logical choice to align the relevant lamprey genes with their human or mouse counterparts. To give *L. morii* more meaningful names and reflect the homologs among agnathans and gnathostomes in our map, we refer to the method of the brain atlas of sea lamprey<sup>49</sup>. Specifically, we have employed the OrthoFinder pipeline to infer orthology relationships between the sea lamprey (utilized in the brain atlas) and the mouse. Additionally, we have conducted a Blastp analysis between *L. morii* and the sea lamprey, adhering to strict criteria of identity > 30% and evalue > 1e-5. We used sea lamprey as a vector to speculate the evolutionary association of homologous genes between *L. morii* and mouse. In cases where multiple mouse genes correspond to a single lamprey gene (a one-to-many relationship), we have included all orthologue names, delineated by a slash (Supplementary data 7).

## Isolation of cells from the gills, blood, and intestines of *L. morii* ammocoete larvae

**Preparation of intestinal (including typhlosole) and gill cell suspensions.** For every library, the intestinal and gill tissues of *L. morii* ammocoete larvae were dissected, placed on sterile Petri dishes, and washed twice with 3 mL DMEM. Then, the intestines and gills were treated under the microscope, and the tissue was cut up with a blade on the culture dish to form tissue homogenate. Then, the intestinal and gill tissue homogenate was transferred into a 5 mL centrifuge tube containing 2-4 mL of enzyme mixture and mixed upside down to prevent the tissue blocks from adhering to each other. The mixed solution was placed in a 30 °C water bath for enzymatic digestion for 10 minutes and shaken upside down 5-10 times per minute to ensure that the enzyme contacted all tissue homogenates. After 10 min of digestion, a pipette was used to draw 10 µL of the cell suspension from the centrifuge tube onto the slide and observe it under the microscope. If the cells were not separated from the tissue, the digestion time was extended. Therefore, we removed 10 µL every minute and

observed it under the microscope until the tissue block disappeared. We placed a 40 µm cell strainer in a new culture dish, and 1 mL DMEM was pipetted onto the cell sieve to moisten it. Then, the digested cell suspension was filtered through a cell strainer with a wide-mouth sterile pipette tip, and the filtrate was transferred into a 15 mL centrifuge tube. Then, 7 mL of DMEM was added to the cell strainer, and the filtrate was transferred to the above 15 mL centrifuge tube. The cell suspension was centrifuged at a speed of 300 × g for 3 min at 4 °C. After centrifugation, the cell precipitate was visible at the bottom of the centrifuge tube. After removing the supernatant, we added 7 mL of PBS to resuspend the cells. The cells were centrifuged at 500 × g for 3 min. We collected the cell precipitate and repeated the wash 2 times. Then, we removed the supernatant after the last centrifugation and added 2 mL of 0.1% BSA-PBS mixed solution to suspend the cells. Then, we transferred the cell suspension to 2 mL of the low-adsorption centrifuge tube and placed it on ice. Nine microliters of cell suspension were pipetted into a 50 µL PCR tube, and 1 µL of 0.4% trypan blue solution was added. The cell concentration was 1000-2000 cells per 1 µL. Since the proportion of living cells was more than 80%, we repeated the experiment twice to calculate the average value of the three tests to obtain the final cell concentration.

**Preparation of blood cell suspensions.** We added 1 mL of anticoagulant into a 5 mL centrifuge tube and severed the tail to obtain the blood. The isolated blood and anticoagulant were mixed in a volume of 1:9 and collected into the above centrifuge tube. Then, we moistened the 40 µM filter screen with PBS, filtered the cell suspension, and collected the filtered cell suspension into a 50 mL tube. The cells were centrifuged at a speed of 400 × g for 5 min at 4 °C. After centrifugation, there was obvious red sediment at the bottom. Then, we removed the supernatant and added another 15 mL of PBS for resuspension. Then, we took a new 50 mL centrifuge tube and added 10 mL 31% Percoll at the bottom. In addition, we slowly added 10 mL of 54% Percoll and 15 mL of resuspension solution to the top layer. The cells were centrifuged at a speed of 400 × g for 40 min at 4 °C. After centrifugation, the leukocyte layer was collected in the solution between 31% and 54% into a new 50 mL centrifuge tube, and 20 mL of PBS was added for cleaning. The cells were centrifuged at a speed of 400 × g for 10 min at 4 °C. Then, we removed the supernatant and repeated this process twice. Finally, we used 500 µL of PBS to resuspend the cells, stained them with 0.4% trypan blue, and adjusted the cell concentration to 700-1200 cells per 1 µL.

## ScRNA-seq library preparation and sequencing

We constructed libraries of gills, intestines, and blood of *L. morii* ammocoete larvae using the DNBelab C4 series Single Cell RNA Library Preparation Kit (MGI Tech Co., Ltd.) based on droplet microfluidics technology. Cells were meticulously prepared into droplets using the DNBelab C4 portable single-cell system provided by MGI Tech Co., Ltd., specifically designed for cell lysis and mRNA capture. The single-cell microdroplets were meticulously recovered through an emulsion-breaking recovery system. Subsequently, mRNA captured by magnetic beads was transcribed into complementary DNA (cDNA) and amplified through 16 cycles of Polymerase Chain Reaction (PCR) to enrich the cDNA. The cDNA products were then subjected to a series of processes to prepare single-stranded DNA libraries, including shearing, end repair, ligation, 12 cycles of PCR, denaturation, circularization, and digestion. Finally, 10 ng of the product was obtained for sequencing analysis on the MGISEQ 2000 platform.

## ScRNA-seq data processing

The raw data obtained by scRNA-seq on the MGISEQ 2000 platform were filtered and demultiplexed using PISA (version 1.10.2) (<https://github.com/shiquan/PISA>)<sup>50</sup>. Reads were aligned to the reference genome using STAR (version 2.7.9a) (<https://github.com/alexdobin/>

STAR<sup>51</sup> and were sorted using Sambamba (version 0.7.0) (<https://github.com/biod/sambamba/releases/tag/v0.7.0>). The cell versus gene UMI count matrix was generated using PISA.

### Cell clustering and identification of cell types

For every sequencing library, unsupervised clustering was performed using Seurat (version 4.2.0) ([https://satijalab.org/seurat/pbmc3k\\_tutorial.html](https://satijalab.org/seurat/pbmc3k_tutorial.html))<sup>52</sup> after filtering, and the databases from different sequencing libraries underwent normalizing and scaling. Variable genes were determined using Seurat “FindVariableGenes” function with default parameters (selection.method = “vst”, nfeatures = 2000). For the gill, intestine, and blood atlases, we used the IntegrateData function and parameter dims (30, 30, 20) to integrate the data. Clusters were identified via the FindClusters function (resolution = 0.3, 0.5, 0.5) implemented in Seurat and subsequently visualized using the RunUMAP function (reduction = “pca”). The harmony package was used to integrate the three organs and tissues with the RunHarmony function. Clusters were identified via the FindClusters function (resolution = 0.5) implemented in Seurat and subsequently visualized using the RunUMAP function (reduction = “pca”). To identify the DEGs of different clusters, the FindAllMarkers function was performed with the following options (min.pct = 0.25, logfc.threshold = 0.25, test.use = “wilcox”). Statistical significance was determined using the Wilcoxon test with a two-sided test approach and the Bonferroni method was applied to adjust the p-values (Supplementary data 8-12).

### Homology analysis of *MPL-L* gene in vertebrates

The NR Database was obtained from NCBI and subsequently partitioned into two distinct categories: vertebrates and invertebrates. The HMM file for the EpoR\_lig-bind domain family (PF09067) of *MPL-L* genes was retrieved from the Pfam database<sup>47</sup>. Following this, the hmmsearch functionality of HMMER (version 3.4) (<https://github.com/EddyRivasLab/hmmer>)<sup>48</sup> was utilized to compare against both vertebrate and invertebrate subsets of the NR database, with an expectation value set at 1e-5. Finally, representative species from aquatic animals, amphibians, reptiles, birds, and humans were individually selected from the alignment results. Their sequence domain profiles were then examined using the CD-Search tool on NCBI and compared with the *MPL-L* domain of *L. morii* (Supplementary data 13).

### Correlation analysis

Cell type correlation analysis was performed using MetaNeighbor with the default parameters (<https://github.com/gillislab/MetaNeighbor/tree/master>). Gene correlation analysis was performed using MAGIC with the default parameters<sup>53</sup>.

### Cell trajectory analysis

Single-cell pseudotime trajectory analysis was conducted with the Monocle2 package (version 2.14.0) (<http://cole-trapnell-lab.github.io/monocle-release/docs/>)<sup>54</sup>. The starting subject CellDataSet class was created with an unnormalized expression cell-gene matrix. The clustering DEGs were identified by performing the differentialGeneTest function. Subsequently, the top 1000 genes were filtered by ordering the qval parameter to use the genes to define cell progress. The DDTTree contained in the reduceDimension function was used for data dimension reduction. The orderCells function was used to sort the cells in pseudotime. The trajectory was visualized using the plot\_cell\_trajectory function. The gene expression trajectories were visualized using plot\_pseudotime\_heatmap and plot\_genes\_in\_pseudotime functions.

### SAMap analysis

SAMap<sup>43</sup> (version 0.3.0) was performed to analyze lamprey-zebrafish and lamprey-mouse similarities and differences in immune cell types. The zebrafish immune data was from published data of the liver,

spleen, and head kidney. We took the immune cells of the thymus, fetal intestine, peripheral blood, and megakaryocyte cells of bone marrow from the published whole organ atlas of mouse<sup>20</sup>. Based on the self-assembly manifold (SAM) algorithm, SAMap enables identify the shared expression program of homologous cell types of distant phylogenetic species and can reliably compare species with long evolutionary distances. We first performed reciprocal BlastP hits between proteomes to construct a gene-gene bipartite graph between lamprey and zebrafish/mice and further screened homologous genes with a percentage of identical matches threshold of 30. Then using the “get\_mapping\_scores” function to project the two species datasets into a joint, lower-dimensional manifold representation, and the mapping strength between two atlas cell types was measured by alignment score. The “gpf.find\_all” function was used to retrieve the enriched gene pairs in the aligned cell type pairs, and the alignment score threshold was 0.2 (Supplementary data 14, 15). Statistical significance was determined using the Wilcoxon test with a two-sided test approach and the Benjamini-hochberg method was applied to adjust the p-values. The comparison score threshold of homologous cell type pairs was 0.1.

### Generation of anti-VLRA and anti-MPL-L polyclonal antibodies

Anti-VLRA (1.13 mg/mL) and anti-MPL-L (2.75 mg/mL) polyclonal antibodies were generated in rabbits by ABclonal (Wuhan, China). In brief, the fragment of those genes was cloned and inserted into the expression vector pET-28a. Their SUMO-tagged fusion antigens were synthesized from the bacteria *Escherichia coli*. The purified recombinant proteins were used to immunize rabbits. The anti-VLRA and anti-MPL-L antibodies were affinity purified. The titer of the antibodies was determined with an enzyme-linked immunosorbent assay (ELISA). The specificity of the polyclonal antibodies (1 µg/mL) was confirmed by Western blotting using rLm-VLRA and rLm-MPL-L.

### Tissue processing

For frozen sections, the fresh gill, intestinal, skin, kidney, and liver tissues were inserted into the tissue fixative for at least 24 h for subsequent experiments. Then, those tissues were removed from the fixative and trimmed with a scalpel. For dehydration, the trimmed tissues were put into a 15% sucrose solution at 4 °C until dehydration was achieved and then transferred into a 30% sucrose solution at 4 °C. For the OCT embedding process, the dehydrated tissues were removed, and filter paper was used to dry the tissue surface. Then, they were placed into the OCT embedding agent on the embedding table. Subsequently, the samples were placed on the quick-freezing stage of the frozen microtome and then sliced after fully embedded in the OCT agent. Finally, the embedded tissue block was sliced at 8-10 µM. Then, the samples were placed on clean slides and stored in a -20 °C refrigerator.

### HE staining

For dewaxing and hydration, the frozen sections were removed from the -20 °C refrigerator to thaw to room temperature, fixed with tissue fixative for 15 min, and rinsed with water. Then, the sections were placed into xylene I for 20 min, xylene II for 20 min, anhydrous ethanol I for 5 min, anhydrous ethanol II for 5 min, and 75% alcohol for 5 min and then washed with water successively. For hematoxylin staining, the sections were stained with hematoxylin staining solution for 5 min and then washed with water. After staining with the hematoxylin differentiation solution, the sections were washed with water. Finally, the samples were returned to the Scotts Tap Water bluing agent solution and rinsed with running water. For eosin staining, the sections were dehydrated with 85% alcohol for 5 min and 95% alcohol for 5 min and then stained with eosin staining solution for 5 min sequentially. For final dehydration and sealing, we placed the frozen sections into 100% ethanol I for 5 min, 100% ethanol II for 5 min, 100% ethanol III for 5 min,

xylene I for 5 min, and xylene II for 5 min and sealed them with neutral gum. Sections were observed by microscopy, and images were obtained for analysis.

### Immunohistochemistry

For analysis of T-like cell distribution in *L. morii* larvae, gill, and intestine sections were removed from the -20 °C refrigerator to room temperature and then washed with PBS thrice for 5 min. Then, the sections were incubated with critic acid buffer (pH 6.0) in a 95 °C water bath for 5 min and washed with PBS thrice for 5 min. Then, an endogenous catalase inhibitor was applied to the sections for 10 min, and the sections were washed with PBS thrice for 5 min. This step was followed by 30 min of preincubation with 10% normal bovine serum albumin (BSA) to block nonspecific binding sites. The sections and anti-VLRA (1:200) were placed in a humidified chamber and incubated overnight at 4 °C. Subsequently, the sections were washed with PBS three times for 10 min and then incubated with secondary antibodies for 1 h, followed by washing with PBS three times for 5 min. Afterward, streptomycin antibiotic protein catalase was added to the sections for 30 min, and then the sections were washed with PBS three times for 10 min. Finally, the sections were mounted with DAPI solution and hematoxylin staining solution for staining. Immunohistochemical images were obtained using a ZEISS Axio Examiner.

### Fluorescence in situ hybridization

To make paraffin sections, the zebrafish (2 months post-fertilization) was fixed in the fixing agent for at least 24 h. For dehydration and paraffin embedding, samples were placed in 75% ethanol for 4 h, 85% ethanol for 2 h, 90% alcohol for 2 h, 95% ethanol for 1 h, 100% ethanol I for 30 min, 100% ethanol II for 30 min, ethanol benzene for 5–10 min, xylene II for 5–10 min, 65 °C melting paraffin for 30 min. The oligonucleotide probe sequences are shown in Supplementary Data 16. The fluorescent probe was synthesized by Servicebio (China). Fluorescence in situ hybridization was performed using the RNASweAMI-FISH Kit (Servicebio, China).

### Immunofluorescence homologous double label staining experiment

Immunofluorescence staining was performed using the Tyramide signal amplification (TSA) Fluorescence Double Staining Kit (Servicebio, China). For analysis of anti-MPL-L and anti-VLRA colocalization in *L. morii* larvae, the gill, intestine, skin, kidney, and liver sections were removed from the -20 °C refrigerator to thaw to room temperature and then washed with PBS thrice for 5 min. Then, the sections were fixed in the fixation solution for 30 min and incubated with critical acid buffer (pH 6.0) in a 95 °C water bath for 5 min, followed by washing with PBS thrice for 5 min. Then, an endogenous catalase inhibitor was applied to the sections for 10 min, and the sections were washed with PBS thrice for 5 min. This step was followed by 30 min of preincubation with 10% normal bovine serum albumin (BSA) to block nonspecific binding sites. The sections and anti-VLRA (1:400) were placed in a humidified chamber and incubated overnight at 4 °C. Then, the sections were incubated with the corresponding HRP-labeled secondary antibodies for 50 min, followed by washing with PBS three times for 5 min. Subsequently, the first dye TSA-FITC was added dropwise onto the sections, incubated for 10 min, and washed with PBS three times for 5 min after incubation. Afterward, the sections were incubated with the critical acid buffer (pH 6.0) in a 95 °C water bath for 15 min. Then, the sections and anti-MPL-L (1:400) were placed in a humidified chamber and incubated overnight at 4 °C followed by incubation with the corresponding HRP-labeled secondary antibodies for 50 min. Subsequently, the second dye TSA-CY3 was added dropwise onto the sections for 10 min and washed with PBS thrice for 5 min after incubation. Finally, the sections were mounted with DAPI solution for staining, and a spontaneous fluorescence quenching agent was added

for 20 min. Immunofluorescence images were acquired with a microscope slide scanner (Pannoramic MIDI). The fluorescein-conjugated secondary antibodies used in this study were goat anti-rabbit IgG HRP-FITC (1:500) and goat anti-rabbit IgG HRP-CY3 (1:500). The intensity of positive cell areas was quantified using ImageJ software<sup>55</sup>.

### Protein-Protein Docking

Using AlphaFold2 (<https://github.com/google-deepmind/alphafold>), the composite structure prediction of the human TPO/*L. morii* MPL-L and the *L. morii* NuDC/MPL-L were conducted. The model with the rank “0” prediction was selected. The generated model PDB files were visualized using PyMOL (<http://www.pymol.org/pymol>) and the amino acid binding sites between the ligand and receptor were analyzed and annotated.

### Intraperitoneal injection experiment

Briefly, lamprey larvae were injected intraperitoneally with rhTPO (100 µL, 50 ng/mL) on Day 0 and Day 4, injected with EdU (5 µg) in 60 µL 0.67 X PBS on Day 6 and returned to aquaculture for 24 h before sacrifice. PHA-L (25 µg) in 0.67 X PBS was administered as a single intraperitoneal injection, and the lamprey larvae were sacrificed on Day 7 after injection. The control group was injected with an equal volume of 0.67 X PBS.

### Flow cytometry

Cells isolated from PBS-, PHA-L-, and rhTPO-stimulated *L. morii* larvae were plated in tubes and fixed for 20 min in a 4% paraformaldehyde solution in PBS at room temperature. Then, the cells were permeabilized with 0.1% Triton X-100 at room temperature for 20 mins. The cells were blocked with 10% BSA for 1 h, incubated with FITC-conjugated anti-VLRA polyclonal (1:200) and Alexa Fluor 647-conjugated anti-MPL-L polyclonal antibodies (1:200) in PBS overnight at 4 °C, and then washed three times with PBS. The cells were suspended in PBS and analyzed on a Cytek® Northern Lights™ instrument (Cytek Biosciences, Shanghai, China). Data analysis was performed using FlowJo software.

### Cell proliferation experiment

EdU staining was performed using the BeyoClick™ EdU Cell Proliferation Kit with DAB (Beyotime). Briefly, frozen sections were fixed with 4% paraformaldehyde for 20 min and washed with PBS three times for 5 min. Then, the sections were permeabilized with PBS containing 0.5% Triton X-100 for 15 min and washed with PBS three times for 5 min. Subsequently, endogenous catalase inhibitor was applied to the sections for 10 min, and the sections were washed with PBS three times for 5 min. Then, the sections were incubated with an EdU reaction cocktail for 30 min in the dark. After being washed in PBS three times for 5 min, sections were incubated in streptavidin-HRP solution for 30 min. Sections were washed several times in PBS and mounted with DAB for staining. Images were obtained using a ZEISS Axio Examiner.

### Reporting summary

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

### Data availability

Sequencing data and matrix data generated in this study have been deposited in the CNGB Sequence Archive (CNSA) of China National GeneBank DataBase (CNGBdb) under accession code [CNP0004792](#). The full-length transcriptome data are available in the Figshare (<https://doi.org/10.6084/m9.figshare.26036284>). The single-cell and full-length transcriptome raw data generated in this study have also been deposited in the Genome Sequence Archive (GSA) database<sup>56,57</sup> under accession code [CRA017568](#). The scRNA-seq data of zebrafish,

white-spotted bamboo shark (*Chiloscyllium plagiosum*), and Chinese tongue sole (*Cynoglossus semilaevis*) used in this study are available in the CNGBdb under accession code [CNP0002712<sup>ss</sup>](#). The published scRNA-seq data of mouse used in this study are available in the mouse cell atlas (GSE108097)<sup>20</sup>, which data obtained on the site: <https://figshare.com/s/865e694ad06d5857db4b?file=10756798>. The published scRNA-seq data of long-tailed macaque (*Macaca fascicularis*) used in this study are available in CNGBdb under accession code [CNP0001469](#), which processed data obtained on the site . Source data are provided with this paper.

## References

- Bernstein, R. M., Schluter, S. F., Bernstein, H. & Marchalonis, J. J. Primordial emergence of the recombination activating gene 1 (RAG1): sequence of the complete shark gene indicates homology to microbial integrases. *Proc. Natl Acad. Sci. USA* **93**, 9454–9459 (1996).
- Agrawal, A., Eastman, Q. M. & Schatz, D. G. Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. *Nature* **394**, 744–751 (1998).
- Mayer, W. E. et al. Isolation and characterization of lymphocyte-like cells from a lamprey. *Proc. Natl Acad. Sci. USA* **99**, 14350–14355 (2002).
- Alder, M. N. et al. Diversity and function of adaptive immune receptors in a jawless vertebrate. *Science* **310**, 1970–1973 (2005).
- Boehm, T. et al. Evolution of alternative adaptive immune systems in vertebrates. *Annu. Rev. Immunol.* **36**, 19–42 (2018).
- Pancer, Z. et al. Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey. *Nature* **430**, 174–180 (2004).
- Hirano, M. et al. Evolutionary implications of a third lymphocyte lineage in lampreys. *Nature* **501**, 435–438 (2013).
- Lynch, H. E. et al. Thymic involution and immune reconstitution. *Trends Immunol.* **30**, 366–373 (2009).
- Rubin, S. A. et al. Single-cell analyses reveal early thymic progenitors and pre-B cells in zebrafish. *J. Exp. Med.* **219**, <https://doi.org/10.1084/jem.20220038> (2022).
- Ashby, K. M. & Hogquist, K. A. A guide to thymic selection of T cells. *Nat. Rev. Immunol.* **24**, 103–117 (2024).
- Bajoghli, B. et al. A thymus candidate in lampreys. *Nature* **470**, 90–94 (2011).
- Huang, L. et al. Full-length transcriptome: a reliable alternative for single-cell RNA-Seq analysis in the spleen of teleost without reference genome. *Front. Immunol.* **12**, 737332 (2021).
- Boehm, T. et al. VLR-based adaptive immunity. *Annu. Rev. Immunol.* **30**, 203–220 (2012).
- Kaushansky, K. et al. Promotion of megakaryocyte progenitor expansion and differentiation by the c-Mpl ligand thrombopoietin. *Nature* **369**, 568–571 (1994).
- Fox, N., Priestley, G., Papayannopoulou, T. & Kaushansky, K. Thrombopoietin expands hematopoietic stem cells after transplantation. *J. Clin. Invest.* **110**, 389–394 (2002).
- Qian, H. et al. Critical role of thrombopoietin in maintaining adult quiescent hematopoietic stem cells. *Cell Stem Cell* **1**, 671–684 (2007).
- Tefferi, A. JAK and MPL mutations in myeloid malignancies. *Leuk. Lymphoma* **49**, 388–397 (2008).
- Matsumura, I. et al. Involvement of prolonged ras activation in thrombopoietin-induced megakaryocytic differentiation of a human factor-dependent hematopoietic cell line. *Mol. Cell Biol.* **18**, 4282–4290 (1998).
- Nishimura, C. D. et al. c-MPL provides tumor-targeted T-cell receptor-transgenic T cells with costimulation and cytokine signals. *Blood* **130**, 2739–2749 (2017).
- Han, X. et al. Mapping the Mouse Cell Atlas by Microwell-Seq. *Cell* **172**, 1091–1107.e1017 (2018).
- Jain, J., McCaffrey, P. G., Valge-Archer, V. E. & Rao, A. Nuclear factor of activated T cells contains Fos and Jun. *Nature* **356**, 801–804 (1992).
- Wingelhofer, B. et al. Implications of STAT3 and STAT5 signaling on gene regulation and chromatin remodeling in hematopoietic cancer. *Leukemia* **32**, 1713–1726 (2018).
- Wang, Z. & Bunting, K. D. STAT5 in hematopoietic stem cell biology and transplantation. *Jakstat* **2**, e27159 (2013).
- Lesourne, R. et al. Themis, a T cell-specific protein important for late thymocyte development. *Nat. Immunol.* **10**, 840–847 (2009).
- Yasuda, A. et al. The stem cell factor/c-kit receptor pathway enhances proliferation and invasion of pancreatic cancer cells. *Mol. Cancer* **5**, 46 (2006).
- Lee, M. S., Hanspers, K., Barker, C. S., Korn, A. P. & McCune, J. M. Gene expression profiles during human CD4+ T cell differentiation. *Int Immunol.* **16**, 1109–1124 (2004).
- Mauriello, J. A. Jr. et al. The role of proliferating cell nuclear antigen (PCNA) in differentiating idiopathic orbital inflammatory disease and lymphoid proliferations. *Ophthalmic Plast. Reconstr. Surg.* **13**, 26–30 (1997).
- Nabetani, A., Koujin, T., Tsutsumi, C., Haraguchi, T. & Hiraoka, Y. A conserved protein, Nuf2, is implicated in connecting the centromere to the spindle during chromosome segregation: a link between the kinetochore function and the spindle checkpoint. *Chromosoma* **110**, 322–334 (2001).
- Groom, J. R. & Luster, A. D. CXCR3 in T cell function. *Exp. Cell Res.* **317**, 620–631 (2011).
- Ito, K., Sportoletti, P., Clohessy, J. G., Silvia, G. & Pandolfi, P. P. The role of Nucleophosmin in hematopoietic stem cells and the pathogenesis of Myelodysplastic Syndrome. *Blood* **116**, 95–95 (2010).
- Bonnefoy-Bérard, N. et al. Vav: function and regulation in hematopoietic cell signaling. *Stem Cells* **14**, 250–268 (1996).
- Yoshihama, M. et al. The human ribosomal protein genes: sequencing and comparative analysis of 73 genes. *Genome Res.* **12**, 379–390 (2002).
- Hitchcock, I. S. & Kaushansky, K. Thrombopoietin from beginning to end. *Br. J. Haematol.* **165**, 259–268 (2014).
- Grozovsky, R. et al. The Ashwell-Morell receptor regulates hepatic thrombopoietin production via JAK2-STAT3 signaling. *Nat. Med.* **21**, 47–54 (2015).
- Meixner, A., Karreth, F., Kenner, L. & Wagner, E. F. JunD regulates lymphocyte proliferation and T helper cell cytokine expression. *EMBO J.* **23**, 1325–1335 (2004).
- Jadhav, K. & Zhang, Y. Activating transcription factor 3 in immune response and metabolic regulation. *Liver Res.* **1**, 96–102 (2017).
- Kuwahara, M. et al. The transcription factor Sox4 is a downstream target of signaling by the cytokine TGF-β and suppresses T(H)2 differentiation. *Nat. Immunol.* **13**, 778–786 (2012).
- Yang, C. Y. et al. The transcriptional regulators Id2 and Id3 control the formation of distinct memory CD8+ T cell subsets. *Nat. Immunol.* **12**, 1221–1229 (2011).
- Szabo, P. A. et al. Single-cell transcriptomics of human T cells reveals tissue and activation signatures in health and disease. *Nat. Commun.* **10**, 4706 (2019).
- Pfistershamer, K. et al. CD63 as an activation-linked T cell costimulatory element. *J. Immunol.* **173**, 6000–6008 (2004).
- Zhang, S., Laouar, A., Denzin, L. K. & Sant’Angelo, D. B. Zbtb16 (PLZF) is stably suppressed and not inducible in non-innate T cells via T cell receptor-mediated signaling. *Sci. Rep.* **5**, 12113 (2015).
- Brauer, R. et al. MMP-19 deficiency causes aggravation of colitis due to defects in innate immune cell function. *Mucosal. Immunol.* **9**, 974–985 (2016).

43. Tarashansky, A. J. et al. Mapping single-cell atlases throughout Metazoa unravels cell type evolution. *Elife* **10**, <https://doi.org/10.7554/elife.66747> (2021).
44. Li, J. et al. Deep learning of cross-species single-cell landscapes identifies conserved regulatory programs underlying cell types. *Nat. Genet.* **54**, 1711–1720 (2022).
45. Pan, R. M. et al. A microtubule associated protein (hNUDC) binds to the extracellular domain of thrombopoietin receptor (Mpl). *J. Cell Biochem.* **96**, 741–750 (2005).
46. Tang, Y. S., Zhang, Y. P. & Xu, P. hNUDC promotes the cell proliferation and differentiation in a leukemic cell line via activation of the thrombopoietin receptor (Mpl). *Leukemia* **22**, 1018–1025 (2008).
47. Mistry, J. et al. Pfam: The protein families database in 2021. *Nucleic Acids Res.* **49**, D412–d419 (2021).
48. Eddy, S. R. Accelerated Profile HMM Searches. *PLoS Comput. Biol.* **7**, e1002195 (2011).
49. Lamanna, F. et al. A lamprey neural cell type atlas illuminates the origins of the vertebrate brain. *Nat. Ecol. Evol.* **7**, 1714–1728 (2023).
50. Shi, Q., Liu, S., Kristiansen, K. & Liu, L. The FASTQ+ format and PISA. *Bioinformatics* **38**, 4639–4642 (2022).
51. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
52. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e3529 (2021).
53. van Dijk, D. et al. Recovering gene interactions from single-cell data using data diffusion. *Cell* **174**, 716–729.e727 (2018).
54. Qiu, X. et al. Single-cell mRNA quantification and differential analysis with Census. *Nat. Methods* **14**, 309–315 (2017).
55. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).
56. Chen, T. et al. The genome sequence archive family: toward explosive data growth and diverse data types. *Genomics Proteom. Bioinforma.* **19**, 578–583 (2021).
57. Database Resources of the National Genomics Data Center, China National Center for Bioinformation in 2022. *Nucleic Acids Res.* **50**, D27–d38, (2022).
58. Wang, H.-Y. et al. Single-cell RNA sequencing illuminates the ontogeny, conservation and diversification of cartilaginous and bony fish lymphocytes. *Nat. Commun.* <https://doi.org/10.1038/s41467-024-51761-4> (2024).

## Acknowledgements

We thank Dr. Thomas Boehm and Dr. Gaoqun Zhang (Max Planck Institute of Immunobiology and Epigenetics, Freiburg) for their comments on the manuscript. This work was supported by the Laoshan Laboratory (LSKJ202203802), the National Key Research and Development Program of China (2022YFD2400100), the Taishan Scholar Project Fund of Shandong of China, the National Ten-Thousands Talents Special Support Program, the AoShan Talents Cultivation Program Supported by Qingdao National Laboratory for Marine Science and Technology (2017ASTCP-ES06), the Central Public-interest Scientific Institution Basal Research Fund, CAFS (2020TD19) and the Central Public-interest

Scientific Institution Basal Research Fund, YSFRI, CAFS (20603022022002).

## Author contributions

Y.H. and X.L. designed the project and the experiments. Y.H. and X.L. performed single-cell sequencing data analysis. S.L. and Y.L. performed full-length transcriptome data analysis. C.L. and S.L. performed homology analysis. X.L., Y.H., and Y.Z. performed all experiments. X.L., Y.H., X.Z., and Y.P. collected the samples. C.S., Q.L., H-Y.W., J-Y.C., Q.W., Y-Y.L., S.L., G.F., and K.L. provided essential reagents and suggestions. Y.H. and X.L. wrote the manuscript. C.S. supervised the work.

## 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/s41467-024-51763-2>.

**Correspondence** and requests for materials should be addressed to Changwei Shao.

**Peer review information** *Nature Communications* thanks Jonathan Rast and Kun Wang 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) 2024