ARTICLE

Check for updates

Cell landscape of larval and adult *Xenopus laevis* at single-cell resolution

Yuan Liao[1,2,7], Lifeng Ma[1,7], Qile Guo [1,3,7], Weigao E[1,7], Xing Fang[1,2,7], Lei Yang[1,7], Fanwei Ruan[1], Jingjing Wang[1,4], Peijing Zhang[1,4], Zhongyi Sun[1], Haide Chen[1,4], Zhongliang Lin[1], Xueyi Wang[1], Xinru Wang[1], Huiyu Sun[1], Xiunan Fang[4], Yincong Zhou[5], Ming Chen [5], Wanhua Shen [6,8✉], Guoji Guo [1,2,3,4,8✉] & Xiaoping Han [1,2,8✉]

The rapid development of high-throughput single-cell RNA sequencing technology offers a good opportunity to dissect cell heterogeneity of animals. A large number of organism-wide single-cell atlases have been constructed for vertebrates such as *Homo sapiens*, *Macaca fascicularis*, *Mus musculus* and *Danio rerio*. However, an intermediate taxon that links mammals to vertebrates of more ancient origin is still lacking. Here, we construct the first *Xenopus* cell landscape to date, including larval and adult organs. Common cell lineage-specific transcription factors have been identified in vertebrates, including fish, amphibians and mammals. The comparison of larval and adult erythrocytes identifies stage-specific hemoglobin subtypes, as well as a common type of cluster containing both larval and adult hemoglobin, mainly at NF59. In addition, cell lineages originating from all three layers exhibits both antigen processing and presentation during metamorphosis, indicating a common regulatory mechanism during metamorphosis. Overall, our study provides a large-scale resource for research on *Xenopus* metamorphosis and adult organs.

[1] Center for Stem Cell and Regenerative Medicine, and Bone Marrow Transplantation Center of the First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou 310000, China. [2] Zhejiang Provincial Key Lab for Tissue Engineering and Regenerative Medicine, Dr. Li Dak Sum & Yip Yio Chin Center for Stem Cell and Regenerative Medicine, Hangzhou, Zhejiang 310058, China. [3] ZJU-UOE Institute, Zhejiang University, School of Medicine, Haining, Zhejiang, China. [4] Liangzhu Laboratory, Zhejiang University Medical Center, Hangzhou, Zhejiang 311121, China. [5] College of Life Sciences, Zhejiang University, Hangzhou 310003, China. [6] Zhejiang Key Laboratory of Organ Development and Regeneration, College of Life and Environmental Sciences, Hangzhou Normal University, Hangzhou, Zhejiang, China. [7] These authors contributed equally: Yuan Liao, Lifeng Ma, Qile Guo, Weigao E, Xing Fang, Lei Yang. [8] These authors jointly supervised this work: Wanhua Shen, Guoji Guo, Xiaoping Han. ✉email: shen@hznu.edu.cn; ggj@zju.edu.cn; xhan@zju.edu.cn

The rapid development of high-throughput single-cell RNA sequencing (scRNA-seq) technology offer a good opportunity to dissect cell heterogeneity of *Xenopus*, from embryogenesis[1,2] to individual organs such as cornea and kidney[3,4], and to reveal the cellular mechanism of *Xenopus* regeneration such as in limb[5,6], tail[7,8] and spinal cord[9]. A large number of organism-wide single-cell atlases have been constructed for various species, such as *Homo sapiens*[10], *Macaca fascicularis*[11], *Mus musculus*[12], *Danio rerio*[13], *Drosophila melanogaster*[14], *Ciona intestinalis*[15], *Caenorhabditis elegans*[16], *Schmidtea mediterranea*[17], and *Stylophora pistillata*[18]. From an evolutionary perspective, the amphibian is an intermediate taxon that links mammals to vertebrates of more ancient origin. A comprehensive larval and adult *Xenopus* cell landscape is helpful in all aspects of a better understanding of genetic evolution relations. For example, to achieve the transition from an aquatic lifestyle to a terrestrial one, anuran tadpoles undergo drastic morphological and physiological changes during metamorphosis, such as the appearance of limbs, loss of larval tail, remodeling of digestive organs, and decrease in regeneration ability[19]. The mammalian perinatal period shares strong similarities with amphibian metamorphosis[20], making the process of metamorphosis an excellent model from which to learn about mammalian postembryonic development. However, the molecular networks that regulate metamorphosis at single-cell resolution are still unclear.

In this work, we perform highly parallel single-cell transcriptomics with Microwell-seq to construct an initial "*Xenopus* Cell Landscape (XCL)", comprising more than 500,000 cells isolated from four *Xenopus laevis* stages, from larva to juvenile, and 17 adult tissues. XCL, the first such comprehensive resource for clawed frog research, profiles more than 100 major cell types in *Xenopus laevis*, which provides valuable scientific evidence of cross-species evolution. The XCL database website is publicly available at http://bis.zju.edu.cn/XCL/ for all researchers in the *Xenopus* community.

Results

Construction of the *Xenopus* cell landscape using Microwell-seq.
Xenopus laevis is a classical model of vertebrate embryonic development[1], regeneration potential[8], and neuronal regulation[21]. Using Microwell-seq, we constructed a comprehensive XCL including adult and larval metamorphosis stages (Fig. 1a)[12]. A detailed multi-organ transcriptome landscape covered brain, liver, lung, kidney, intestine, heart, muscle, skin, pancreas, eye, testis, ovary, bladder, stomach, bone marrow, spleen, and oviduct tissues from 1-year-old *Xenopus laevis*. The whole body of Nieuwkoop and Faber (NF) stage 48 (premetamorphosis), NF54 (prometamorphosis), NF59 (metamorphosis climax), and NF66 (end of metamorphosis) were dissociated respectively to fully analyze the process of metamorphosis. Tissues and tadpoles were prepared as single-cell suspensions (Supplementary Dataset 1).

The single-cell transcriptomics data were processed using published pipelines[22]. In total, we obtained 501, 358 qualified single cells from 17 adult tissues and tadpoles at NF48, NF54, NF59, and NF66 (Fig. 1b–d and Fig. S1a, b). The complete tissue dataset was grouped into 106 clusters and over 581 cell-type subclusters in the hierarchy (Fig. 1b, c and Fig. S1c). The numbers of cells and unique molecular identifiers (UMIs) per cell detected for each tissue are shown in Fig. 1d and Fig. S1b. The average number of genes detected per cell was ~600. The tissues with the highest and lowest average numbers of genes detected were the testis and pancreas, with ~1000 genes and 300 genes respectively (Fig. S1a). We annotated the cell types based on clustering markers referred in previous studies. The marker genes for each

cell type are listed in Fig. S1d and Supplementary Dataset 1. Multiple tissues, including bladder, eye, heart, lung, muscle, ovary, pancreas, oviduct, and stomach, contributed to the defined adult stromal cells (Cluster (C)1, C22, and C59), while several clusters (C0, C12, C24, and C105) were larva-specific stromal cells. Other clusters with significant multi-tissue contributions corresponded to epithelial cells (C8), endothelial cells (C9), and muscle cells (C43) (Fig. 1b, c).

To define the cell-type similarities and relationships, pairwise unsupervised MetaNeighbor[23] analyses were performed for these 106 clusters (Fig. 1e). The resulting circular cell-type dendrogram was organized according to major lineages rather than tissues. For example, most cell types were matched within the same category, such as stromal, muscle, and epithelial cells.

In total, our landscape provided cellular signatures for major tissues of *Xenopus laevis*, providing a valuable resource at single-cell resolution for the *Xenopus* community. The resource is publicly available at http://bis.zju.edu.cn/XCL/.

Cellular heterogeneity in diverse *Xenopus laevis* tissues.
Using t-distributed stochastic neighbor embedding (t-SNE) and differential gene expression analyses, we constructed more detailed transcriptional profiles of *Xenopus* tissues, including previously unrecognized cell heterogeneity. The brain is the most complicated tissue, and our data covered most known cell types, such as excitatory and inhibitory neurons, radial glia, astrocytes and oligodendrocytes. We also identified several endocrine cells from the pituitary gland, including gonadotroph cells, growth hormone cells, thyrotroph cells, and melanotrope cells (Fig. 2a and Supplementary Dataset 2). We further examined specific marker genes related to each cluster. Three GABAergic neurons (inhibitory neuron) subclusters (C5, C6, and C27) expressed high levels of *slc32a1* and *gad2*, whereas three excitatory neuron subclusters (C19, C28, and C32) expressed high levels of *slc17a7*. Myelinating oligodendrocytes (C2) expressed high levels of myelin-associated genes *plp1*, *mbp*, and *olig2*, and progenitor cells (C17) expressed high levels of *pdgfra* and the neurogenesis-related transcription factor *nkx2-2*[24] (Fig. 2b and Supplementary Dataset 2). Notably, a large number of endocrine cells were identified from the pituitary gland that expressed different hormone-associated genes as well as co-expression of high levels of *chga* and *chgb* (Fig. S2). For instance, gonadotroph cells (C1 and C26) expressed high levels of *fshb*, growth hormone cells (C3, C7, and C15) expressed high levels of *gh1*, prolactin progenitor cells (C9) expressed high levels of *prl.1*, and thyrotroph cells (C12) expressed high levels of *tshb*. In addition, C10, which expressed the astrocyte marker *atp1a2*, was identified as representative fibrous astrocytes for its high expression of the myelin-associated gene *pmp2*[25] (Fig. 2b).

The lung is an organ of particular interest because *Xenopus* undergoes a transition from living in an aquatic environment to living in a terrestrial environment. We annotated 17 cell-type clusters in the lung with analysis based on known markers (Fig. 2c, d and Supplementary Dataset 2). Alveolar epithelial (AE) cells (C5) were enriched in *sftpb* and *sftpc*, which are typical markers of alveolar type 2 (AT2) cells in the mouse and human lungs (Fig. S3a). The lack of alveolar type 1 (AT1) cells which are easily found in the lungs of mice and humans were missed in our data and enriched pulmonary ionocytes (C10) with high expression of *atp6v0d2* and *atp6v1b1* (Fig. 2d and Fig. S3a), which implied that the respiratory organ of *Xenopus* was intermediate between the gill enrichment of ionocytes and mammalian lung enrichment of alveolar epithelial cells. We also identified melanocytes (C15) with high expression of *pmel-like* and *tyrp1*, which were previously reported to exist in the lungs of
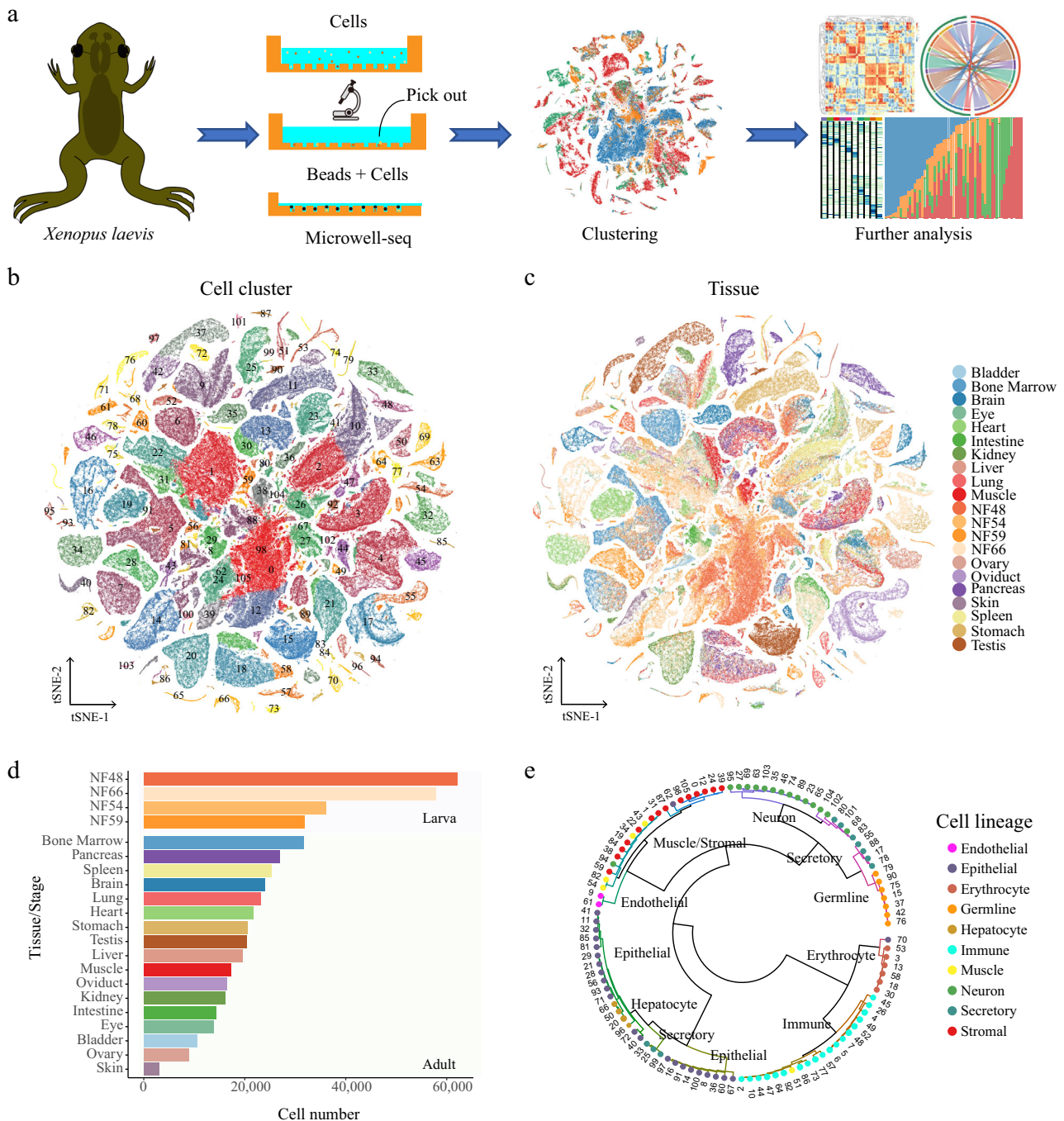
Fig. 1 Constructing an XCL using Microwell-seq. a A schematic of the basic workflow for the *Xenopus* cell landscape using the Microwell-seq platform.
b t-SNE analysis of 501, 358 single cells collected from larval and adult tissues. The 106 cell-type clusters are labeled in different colors. Cell cluster markers are listed in Supplementary Dataset 1. **c** t-SNE analysis of 501, 358 single cells collected from larval and adult tissues. Tissues/stages are labeled in different colors. **d** Cell number of each tissue/stage is detected at the XCL. Tissues/stages are labeled in different colors. **e** The 106 cell-type clusters are re-clustered into ten cell lineages and circles showing the relationships among the 106 cell types. Cell lineages are labeled in different colors.

Xenopus with unknown function[26] (Fig. 2c and Supplementary Dataset 2). To infer the evolutionary relationship between the mouse lung and zebrafish respiratory organs, we performed a cross-species clustering analysis of these tissues. To enable cross-species analysis, orthologous and paralogous genes from mouse, zebrafish and *Xenopus* were extracted[27]. As revealed in Fig. S3b, several cell types were conserved among zebrafish, *Xenopus*, and mouse. For example, the gene expression patterns of *Xenopus* pulmonary ionocytes showed strong correlations with zebrafish gill ionocytes, zebrafish swim bladder epithelial cells and mouse

AT2 cells. Interestingly, although *sftpb* and *sftpc* were highly expressed, the gene expression patterns of *Xenopus* AE cells showed strong correlations with not only mouse AT2 cells but also mouse AT1 cells, Club cells, and alveolar bipotent progenitors, indicating that *Xenopus* AE cells performed a variety of functions similar to those of mouse AT1, AT2 and Club cells in adult *Xenopus* lungs. *Xenopus* AE cells also showed strong correlations with zebrafish swim bladder epithelial cells, implying the evolutionary relationships among the *Xenopus* lung, zebrafish swim bladder, and mouse lung. These results were consistent with

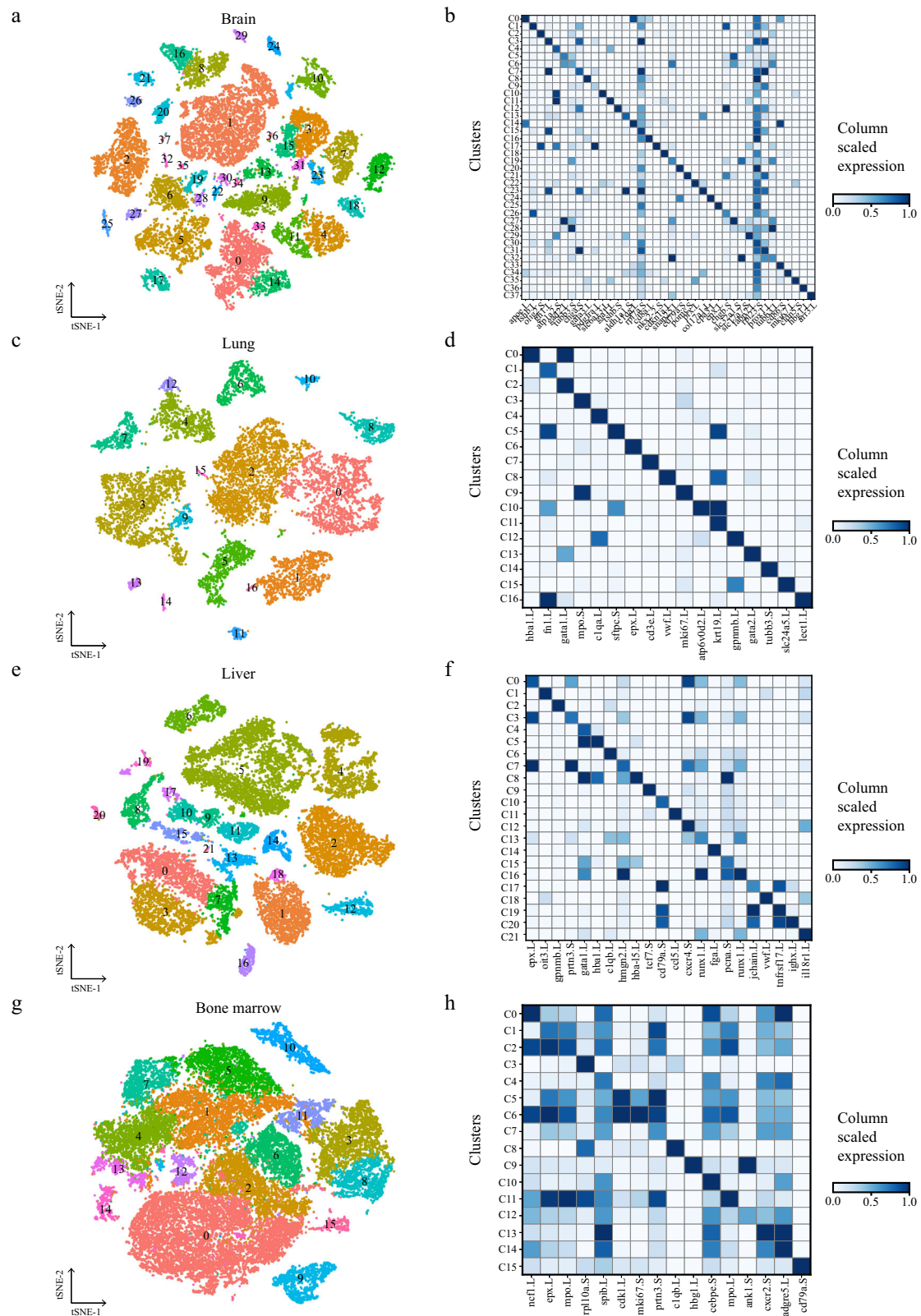**Fig. 2 Cellular heterogeneity in diverse adult *Xenopus laevis* tissues.** t-SNE map of *Xenopus* brain (**a**), lung (**c**), liver (**e**), and bone marrow (**g**) single-cell data. Cells are colored by cell-type cluster; Heatmap showing representative gene expression in each cluster of *Xenopus* brain (**b**), lung (**d**), liver (**f**), and bone marrow (**h**). The color encodes the average expression level.

a previous hypothesis that cell types rather than organs might be proposed as "evolutionary units" in evolutionary biology[28,29].

Unlike that in mammals, hematopoiesis in *Xenopus* occurs in the liver rather than bone marrow[30,31]. To further study the function of hematopoiesis at the single-cell level, we analyzed the cellular components in *Xenopus* liver and bone marrow, respectively. We identified 21 clusters in the *Xenopus* liver (Fig. 2e, f) and expectedly most were hematopoiesis-related clusters except liver parenchyma cell hepatocytes (C14). For example, hematopoietic stem and progenitor cells (HSPCs and C16) were enriched in hematopoiesis transcription factors *gata2*, *runx1*, and *tal1* (Fig. 2f and Supplementary Dataset 2). Next, we identified erythrocytes at various differentiation stages, including erythroid progenitor cells (C4), proliferating early erythrocytes (C15), proliferating middle erythrocytes (C8), and late erythrocytes (C5) based on previously reported marker genes[32] (Fig. S4a). In addition, among 16 identified clusters in *Xenopus* bone marrow, mainly composing granulocytes such as neutrophils and eosinophils (Fig. 2g, h). In contrast to adult mouse bone marrow, *Xenopus* bone marrow contained no HSPCs, but its other cellular components, such as major granulocytes, were similar (Fig. S4b). These data indicated that the liver was the primary hematopoietic organ of adult *Xenopus* while the bone marrow harbored granulocytic potential.

Further, we defined cell-type clusters and their markers in all the other tissues and summarized them in Supplementary Dataset 2 and on the XCL website (Fig. S5 and Supplementary Dataset 2).

Cell-type evolution between adult *Xenopus*, zebrafish, and mammals. Cross-species single-cell RNA analysis provides new insight into animal evolution at the single-cell level[28,33]. To identify the evolutionary process in vertebrates, updated datasets from the Human Cell Landscape (HCL), the Mouse Cell Atlas (MCA), and the Zebrafish Cell Landscape (ZCL) were collected. Details of selected homologous genes across four species were shown in Supplementary Dataset 3. To proceed with network construction, pseudocells for each cell cluster of *Xenopus laevis* were constructed and an integrated manifold with a high degree of cross-species alignment was produced by SAMap to ensure the accuracy of cell types and cell states comparison (Fig. S6). As shown in Fig. 3a, the gene expression patterns of endothelial cells showed the strongest correlations between both species compared with other lineages. Besides, stromal cells were also highly conserved across these four species. The germline of *Xenopus* showed the strongest correlation with the zebrafish germline and lower correlations with mouse germline, and few correlations with human germline, indicating that the germline of *Xenopus* was evolutionarily closer to fish than to mammals.

To investigate the genetic network that underlies cell lineage fate determination, transcription factor specificity scores were calculated for each cell type of *Xenopus laevis* (Fig. 3b). Stromal, muscle, and endothelial cells shared several TFs, such as the *notch* effector *hes1/4*, which are reported to regulate myogenesis[34,35]. *Twist1/2* and *snail*, which are related to the epithelial-to-mesenchymal transition in humans, were highly associated with stromal cells[36]. *Meox2*, a response gene for cell migration and fatty acid transport, was identified in endothelial cells[37,38]. The key regulators of erythropoiesis *gata1* and *tal1* in mice were also highly associated with erythrocytes in *Xenopus*[39,40]. For immune cells of *Xenopus*, the major hematopoietic TF *runx1* and the key regulator of T cell differentiation *runx3* were identified[41,42]. *Sox1*, reported to be implicated in *Xenopus* neurogenesis and the *Xenopus* myelination-related gene *sox10* were detected in our neuronal cells[43,44] (Fig. 3b). To confirm the conserved cell-

lineage-specific TFs in vertebrates, we collected TF data from humans, mice, and zebrafish, and further identified conserved TFs for each lineage based on homologous relationships and expression (Fig. 3c and Figs. S7, S8). Our data showed that endothelial cells shared several TFs in all lineages, such as the regulator of endothelial differentiation *erg*[45], a key gene for vascular development *sox7*[46], and the regulator of transcription of lipid transport *meox2* for endothelial cells[47] (Fig. 3c). In addition, neurons shared most TFs in all lineages. For instance, the key neural crest specifier *sox10*[48], a common regulator of the neuronal conversion gene *neurod1*[49], and the regulator of the oligodendrocyte lineage and neural cell specification *olig2* for neurons[50] (Fig. S8a). In particular, some highly conserved TFs such as the regulator of neural stem cells *tcf4*[51] and the known neural developmental regulator *lhx6*[52] were detected in four species (Fig. S8a). Notably, *olig1*, a common oligodendrocyte transcription factor in mice and humans[53], was not detected in *Xenopus* adult neurons and was only weakly expressed prior to the tadpole stage[54] (Fig. S8a).

Cell landscape of larval *Xenopus* during metamorphosis. To describe the cellular mechanisms regulating amphibian metamorphosis at the single-cell level, NF48, NF54, NF59, and NF66 tadpoles were processed using Microwell-seq and 188,020 single cells passed our quality control tests. Metamorphosis is a process regulated by thyroid hormone (TH) T3. Previous research indicates that T3 is able to induce complete metamorphosis and blockade of T3 synthesis prevents metamorphosis[55]. To ensure that the collected samples represented different stages of metamorphosis, we cross-referenced our data with the expression patterns of four previously reported T3-dependent genes[56], which indicated that our samples represented premetamorphosis, prometamorphosis, metamorphic climax, and the end of metamorphosis (Fig. 4a). The marker genes for each major cell type were listed in Fig. S9a. In our data, several major cell type categories were present in all four stages, such as the defined neurons (C13, C20, C29, C41, and C43), immune cells (C4, C11, and C31), and proximal tubule cells (C27) (Fig. 4b, c). Several digestion system clusters were specifically present in certain stages. For instance, stomach parietal cells (C22) and hepatocytes (C24) were present only in NF66 and NF54, respectively. Meanwhile, enterocytes (C12) were present in NF59 and enterocytes (C15 and C25) were present in N66; however, enterocytes (C48) were present in NF54, NF59, and NF66. Our results showed that remodeling of the intestine is the one of the most dramatic changes during metamorphosis as noted previously[57]. Enterocytes, especially at later stages, exhibited high heterogeneity due to changes in diet during metamorphosis to form the complex adult intestine[56,58].

Next, the different stage tadpole datasets were grouped into 38, 46, 48, and 56 clusters respectively (Fig. S9b–e). We analyzed the proportions of four stages in each cluster. As shown in Fig. 4d, almost no intertemporal cell types emerged, indicating reliable temporality of our data. To further understand the transformations that occur during metamorphosis, we performed another round of clustering specifically for epithelial, stromal, and immune cells (Fig. S10 and Supplementary Dataset 4). Obviously, epithelial and stromal cells exerted strong transcriptomic heterogeneity rather than immune cells. Overall, NF66 is more self-consistent in Epithelial and Stromal in batch-integrated cellular profiles. We also analyzed the relationship between larval and adult cell clusters (Figs. S11–S12 and Supplementary Dataset 5). As shown in Fig. S11b, c, the enterocytes at NF66 and adult shared transcriptional similarity, while enterocytes at NF54 and NF59 show stronger heterogeneity. The correlation of clusters and DGEs between larval and adult enterocytes were
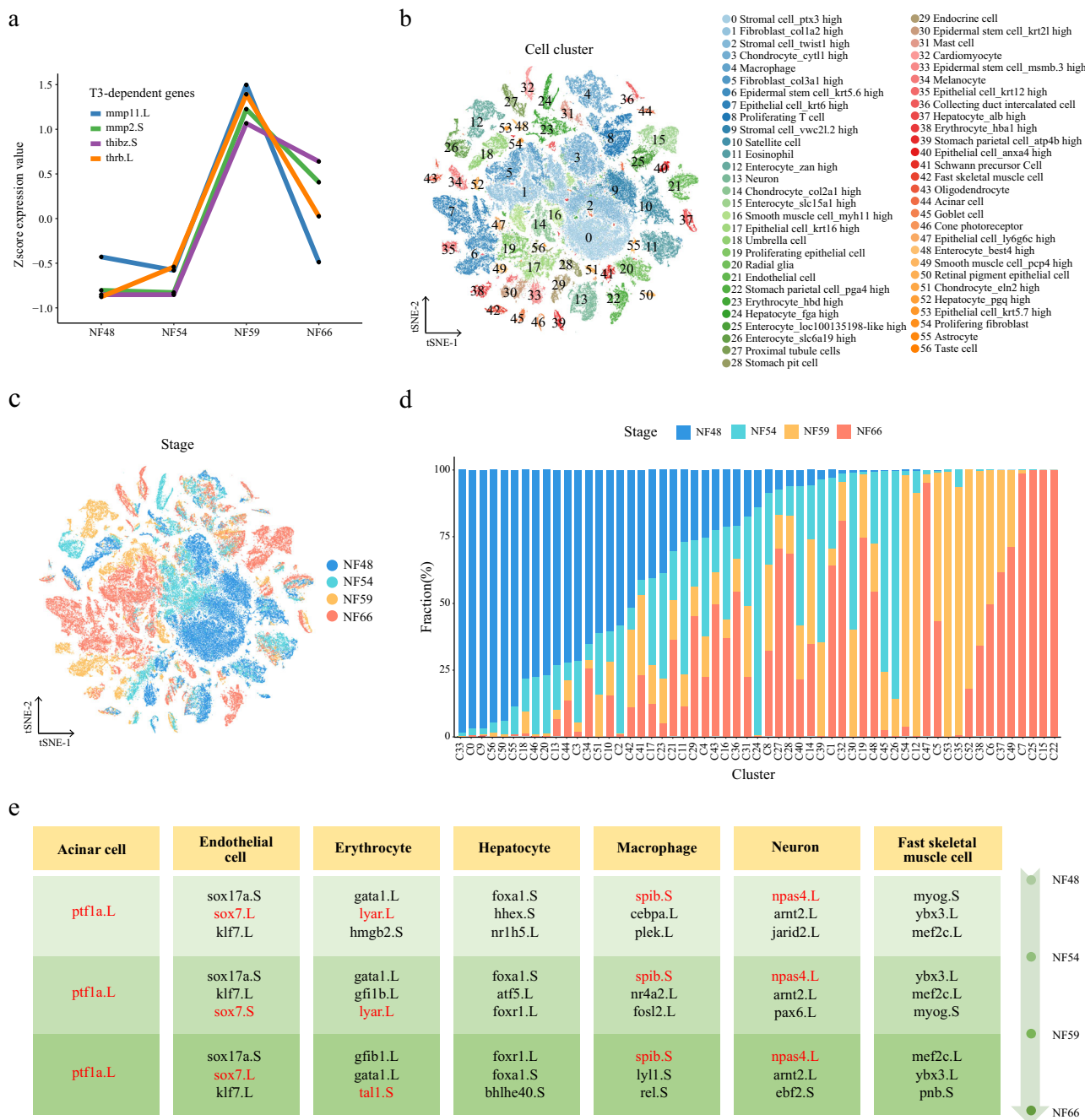
Fig. 3 Cell-type evolution between adult *Xenopus*, zebrafish, and mammals. a Circle plot showing the similarity of cell lineages in *Xenopus laevis*, zebrafish (left), mice (middle), and humans (right). Pairs of cell types with mapping scores greater than 0.1 are connected by lines. Cell lineages are labeled in different colors. **b** Heatmap showing the specific scores of TFs in *Xenopus* for each cell type. Each row represents one TF, and each column represents one cell lineage. Cell lineages are labeled in different colors. The representative TFs in each cell lineage are presented in the right panel. **c** Sankey plot showing common lineage-specific transcription factors in endothelial for four vertebrates (humans, mice, *Xenopus*, and zebrafish). The representative TFs in *Xenopus laevis* are labeled by red. Different TFs in each species are marked in different color boxes. Homologous TFs between species obtained from SAMap are connected by lines.

Fig. 4 Cell landscape of larval *Xenopus* during metamorphosis. a Expression trajectory of genes regulated by T3 during metamorphosis. **b**, t-SNE analysis of 188, 020 single cells collected from larval *Xenopus*. 57 cell-type clusters are labeled in different colors. Cell cluster markers are listed in Supplementary Dataset 1. **c** t-SNE analysis of 188, 020 single cells collected from larval *Xenopus*. Stages are labeled in different colors. **d** Stage proportion in different cell clusters. Stages are labeled in different colors. **e** dTFs are listed in different stages for cell lineages. Each row represents one stage and each column represents one cell type. The representative TFs for each cell types are labeled by red.

shown in Fig. S11c. The high correlation coefficient between C3 and C4 with the most proportion of NF54 and NF59 data suggested some NF-specific regulation, while DGEs contributing to the combined cluster from larval and adult integration indicated that enterocytes were undergoing extensive remodeling during metamorphosis. Similar to the enterocytes, most stomach parietal cells at NF66 and adults shared transcriptional similarity but low correlation with that at NF54 and NF59 (Fig. S11d–f). Hepatocytes at NF54 showed transcriptional heterogeneity but hepatocytes at NF59 displayed a high correlation with NF66 and adult hepatocytes (Fig. S12a–c). In terms of neuron cells, general

integration showed a high correlation among adult and most larval neurons, with some adult neurons showing heterogeneity mostly in C1 and C3 (Fig. S12d–f). Taken together, these data suggested that significant transcriptome changes occurred in the vast majority of adult cells compared with larval cells that underwent metamorphosis.

To identify the stage-specific TFs for each cell type, we collected TF data from all cell types at four stages and defined the stage-specific TFs as "driver TFs (dTFs)" (details in methods and Supplementary Dataset 6). As shown in Fig. 4e, several cell types, including endothelial cells and acinar cells, exhibited relatively stable

regulatory factor expression during metamorphosis. By contrary, erythrocytes dTFs underwent tremendous changes, with the exception of *gata1*. It was worth noting that *lyar*, which regulates globin gene expression during development was no longer a specific transcription factor at NF66[59], indicating a hemoglobin (Hb) transition during this stage. Meanwhile, *tal1*, a regulator of erythroid differentiation, became a specific transcription factor at NF66[60]. These changes were likely important for the larval erythrocytes to undergo Hb transition and to differentiate into adult erythrocytes. In total, our data provided a useful resource for identifying the key regulators of each cell lineage during metamorphosis.

Conversion of erythrocyte hemoglobin from larval to adult. Hb transition from larva to adult during vertebrate development is a physiologically important process for animals to adapt to environmental changes from the fetal environment to an atmospheric environment in animals. Hb switching is also TH-dependent and which kinds of globin are expressed in individual erythrocytes at different stages in *Xenopus laevis* has been controversial[61]. To investigate this transition during metamorphosis in *Xenopus*, we reanalyzed the whole larval and adult erythrocytes. As expected, adult and larval erythrocytes showed strong heterogeneity before and after metamorphosis (Fig. 5a, b). In adults and NF66, erythrocytes mainly expressed *hba1* and *hbg1*, while in early tadpoles at NF48 and NF54, erythrocytes expressed *hbg2* and *hbd* instead. Notably, during premetamorphosis at NF59, the cluster of erythrocytes coexpressing *hba1*, *hbg1*, *hbg2*, and *hbd* indicated a transient state in the development of erythrocytes (Fig. 5c). The expression of *hba1* and *hbd* confirmed that hemoglobin was coexpressed in individual erythrocytes (Fig. 5d). Although several reports have claimed that no erythrocytes contained both larval and adult globin[61,62], our data identified that there was a cluster of erythrocytes, mainly at NF59, that contained both larval and adult globin, as reported previously[63,64]. Interestingly, our data showed that few or no *mhc1a* (MHC class I) molecules were detected in larval erythrocytes before metamorphosis, consistent with a previous report (Fig. 5c), but a continuous increase in expression occurred following development until adult erythrocyte transition[65]. Our result suggested that immune cells might be involved in the removal of larval erythrocytes without *mhc1a*. To investigate the molecular pathways involved in erythrocyte development, we sub-clustered whole erythrocytes into three groups: a larval cluster (C0) of erythrocytes mainly at NF48 and NF54; a transient cluster (C6) of erythrocytes at NF59; and an adult cluster (C2), of erythrocytes mainly at NF66 and in adults. GO classification was performed to analyze the biological functions of the differentially expressed genes (DEGs) in each group based on human homologs (Fig. 5e). The larval cluster was enriched mainly in biological pathway categories associated with RNA metabolism and telomere maintenance, suggesting that transcription and telomerase activity were necessary at the early stage of erythrocytes. Erythrocyte differentiation- and hemoglobin biosynthesis-related biological pathways were enriched in transient clusters. This result implied that erythrocytes at this stage might synthesize new hemoglobin and differentiate into new cell types during the transient state. In conclusion, our data indicated that *hba1* and *hbg1* are adult-special hemoglobin in *Xenopus laevis*, while *hbg2* and *hbd* are larval-special hemoglobin, and that there is a transient cluster of cells containing both larval and adult hemoglobin only at NF59 that might be a progenitor of adult erythrocytes.

Gene expression profiles of the remodeling multisystem during metamorphosis. As shown in Fig. 4c, digestion system cells derived from the endoderm, including enterocytes, hepatocytes and stomach parietal cells showed strong heterogeneity during metamorphosis. To investigate common molecular pathways involved in tissue development during metamorphosis, we re-collected data from digestion system cells, erythrocytes derived from mesoderm and neurons derived from ectoderm to analyze. We clustered selected genes expressed (for details, see Methods) into four modules: module 1, thyroid hormone-synchronous upregulated genes; module 2, thyroid hormone-synchronous downregulated genes; module 3, continuously downregulated genes; module 4, continuously upregulated genes (Fig. 6a and Supplementary Dataset 7). The number of genes in module 1 was larger than that of module 4 for each cell lineage (Fig. 6b), indicating that upregulated genes involved in metamorphosis were mainly thyroid hormone-dependent.

Next, GO analysis was performed (based on the human homologs) to clarify the biological and molecular functions of the genes in each module (Fig. 6c–f). The entire GO term for each module could be found in Supplementary Dataset 8. The common genes in module 1 across cell types during metamorphosis were enriched in antigen processing and presentation of exogenous antigen, immune system process, cellular location, granulocyte activation, and heterocycle catabolic process. Interestingly, the shared genes enriched in antigen processing and presentation of exogenous antigen catalog were proteasome family genes, although different subtypes, such as *psmb4*, *psmb5*, and *psmc2*, are involved in the expression of MHC class I antigens and protein degradation[66]. MHC class I antigens were reported to be important surface markers of adult erythrocytes during *Xenopus* metamorphosis[67]. Our data suggested that antigen processing and presentation in NF59 was required for the formation of MHC I molecules at the end of metamorphosis and was a common regulatory mechanism in each lineage during metamorphosis. To detect the common genes shared by multiple cell types, we also generated Venn diagrams of the genes in different cell types. Genes present with a high frequency in module 1 were *zan* (freq = 5), *nme2* (freq = 4), *cwcl5* (freq = 4), *lypla2* (freq = 4), *taf13* (freq = 4), and so on.

The common genes in module 2 across cell types during metamorphosis were mainly enriched in metabolic processes such as nitrogen compound metabolic process, organic substance metabolic process, mRNA metabolic process, and cellular macromolecule metabolic process. The common genes in module 3 across cell types during metamorphosis were also mainly enriched in metabolic processes, such as cellular metabolic process, mRNA metabolic process, and oxoacid metabolic process (Fig. 6d, e). These data suggested that some cellular metabolic processes such as nitrogen compound and mRNA metabolic process were continually downregulated from the climax of metamorphosis to the end of metamorphosis. Genes with a high frequency in module 2 and 3 were also mainly translation-related genes, such as *eef1d* (freq = 4), *eif2s3* (freq = 4), and *ilf3* (freq = 4), which are involved in the elongation and translation initiation process of protein synthesis[68,69].

The common genes in module 4 across cell types during metamorphosis were enriched in cellular response to chemical or cytokine stimulus, amide biosynthetic process, immune response and translational initiation. These data suggested that upregulation of these processes at the end of metamorphosis was required for the maturation of the adult organ. Genes with a high frequency of module 4 were *s100a11* (freq = 5), *aldh9a1* (freq = 4), *anxa2* (freq = 4), *col1a1* (freq = 4), *col1a2* (freq = 4), *crip1* (freq = 4), *krt19* (freq = 4), *mhc1a* (freq = 4), *pepd* (freq = 4), and *txn* (freq = 4). Notably, these genes are involved in diverse cellular processes, highlighting important biomarkers and targets for metamorphosis in *Xenopus*.

Overall, our data demonstrated that upregulation of antigen processing and presentation and downregulation of cellular
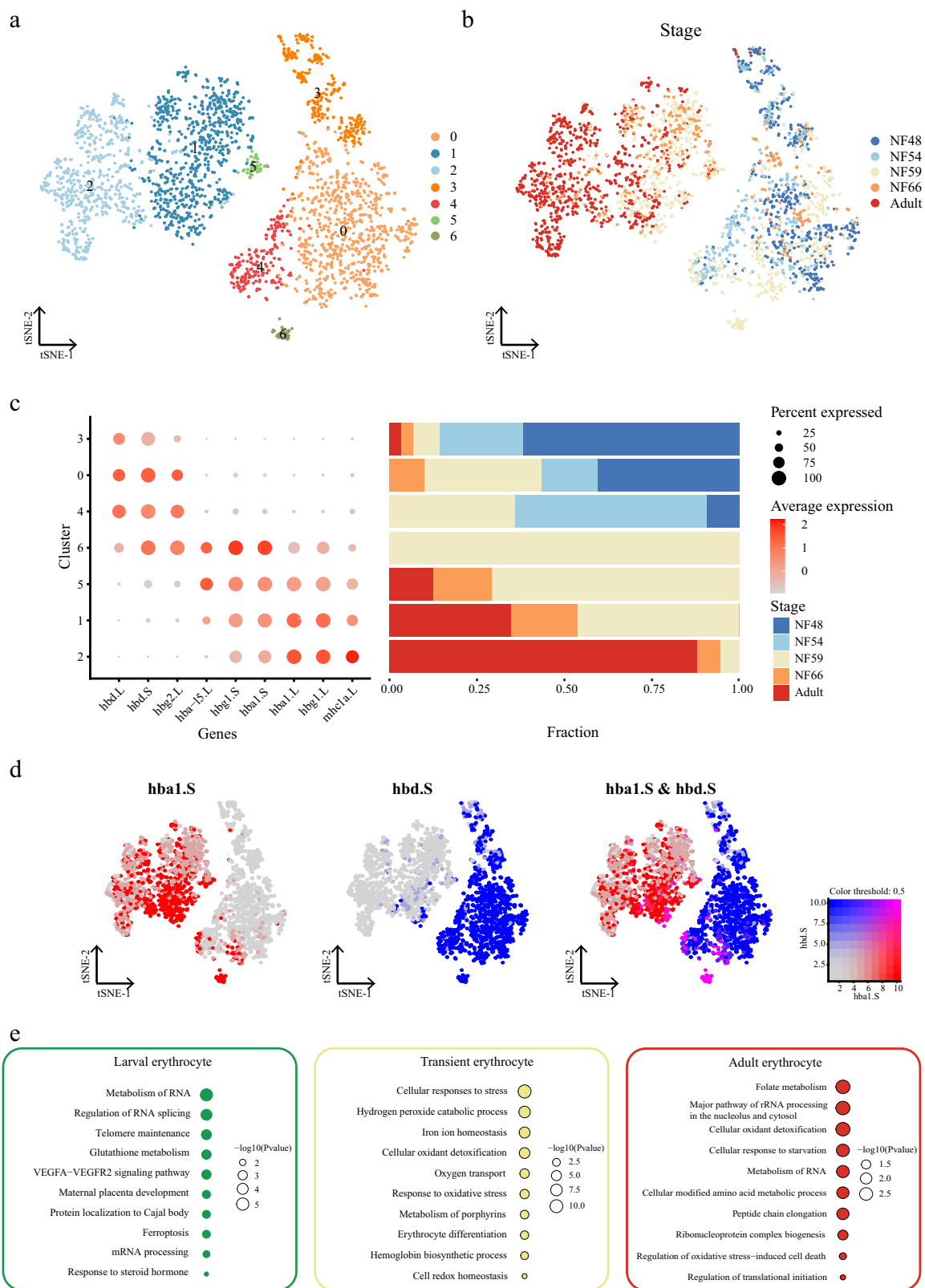
metabolic processes are common biological processes in *Xenopus* metamorphosis for major cell lineage.

Discussion

The constructed organism-wide cell atlases, such as those for *Homo sapiens*, *Mus musculus*, and *Danio rerio*, alongside that of

Xenopus laevis generated here, allowed us to systematically compare cell types across vertebrates. During the transition from an aquatic to a terrestrial lifestyle, the respiratory system of amphibians undergoes drastic changes. The evolutionary relationships among the lung, gill and swim bladder are controversial based on current evidence[70]. Our data identified that *Xenopus* AE

Fig. 5 Conversion of erythrocyte hemoglobin from larva to adult. a t-SNE analysis of erythrocytes collected from larval and adult *Xenopus*. Different clusters are labeled in different colors. Erythrocytes from adults were sampled to the same order of magnitude as in larva. Cell cluster markers are listed in Supplementary Dataset 3. **b** t-SNE analysis of erythrocytes collected from larval and adult *Xenopus*. Different stages are labeled in different colors.
c Representative gene expression in each cluster of erythrocytes and cell proportions in different stages are shown. Stages are labeled in different colors. The size of the dot encodes the percentage of cells within a cell type, and the color encodes the average expression level. **d** Feature plot in the t-SNE map of erythrocytes collected from larval and adult *Xenopus*. Cells are colored according to the expression of the indicated marker genes *hba1.S* (red), *hbd.S* (blue), and co-expression *hba1.S* and *hbd.S* (pink). **e** Representative GO terms enriched in each module for three cell types. Different cell clusters are labeled in different colors. *P* value was calculated by the hypergeometric distribution, a statistical test is one-sided, adjustments *P* values were made after *P* value is corrected by Benjamin–Hochberg multiple test.

cells showed strong correlations with zebrafish swim bladder epithelial cells and mouse AT2 cells, AT1 cells, Club cells and alveolar bipotent progenitors. Meanwhile, *Xenopus* pulmonary ionocytes showed strong relationships to zebrafish gill ionocytes, zebrafish swim bladder epithelial cells, and mouse AT2 cells. Recently, cell types rather than tissues have been proposed as "evolutionary units"[28,29]. Our results also demonstrated that the evolutionary conservation and divergence of the vertebrate respiratory system relied mainly on cell types and not tissues, as previous research proposed.

Amphibian metamorphosis is a dramatic biological phenomenon that could be controlled by the simple TH. Our data indicated that antigen processing and presentation was a common regulatory mechanism involved in cell apoptosis in lineages derived from three germ layers during metamorphosis. GO analysis of thyroid hormone-synchronous upregulated genes during metamorphosis revealed enrichment in genes associated with antigen processing and presentation of exogenous antigens. MHC class I antigens were used to distinguish larval and adult erythrocyte clusters in a previous study[67]. The absence of MHC antigens on the surface of larval erythrocytes is consistent with our results. However, larval erythrocytes were cleared from circulation after 60 days after metamorphosis, indicating a mechanism for the protection of adult erythrocytes after metamorphosis. To our surprise, other cell types, such as hepatocytes, enterocytes, neurons, and stomach parietal cells, contained both antigen processing and presentation genes. These data from ectoderm, mesoderm, and endoderm implied that antigen processing and presentation might be a shared biological pathway involving tissue remodeling during metamorphosis.

In conclusion, we constructed the first *Xenopus* cell landscape at the single-cell level, providing large-scale resource for research on *Xenopus* metamorphosis and adult organs. We performed comparative analyses among vertebrate species at single-cell resolution and revealed an evolutionary relationship between the respiratory systems of aquatic and terrestrial animals. We identified a common biological pathway involving cell apoptosis and differentiation for cell lineages from the three germ layers. These data are valuable for future research in vertebrate biology.

Methods

Animals. Stage 48–66 tadpoles (ten in NF48 and NF54 and one in NF59 and NF66) were obtained from lab-bred mature *Xenopus laevis* (from Shen's lab at Hangzhou Normal University) and reared in a 22 °C incubator with 12 h light /12 h dark cycle in 0.1×Steinberg's solution [in mM: 10 HEPES (Cat. No. H3375, Sigma), 58 NaCl (Cat. No. S5886, Sigma), 0.67 KCl (Cat. No. P5405, Sigma), 0.34 Ca(NO₃)₂ (Cat. No. 31218, Sigma), 0.83 MgSO₄ (Cat. No. M2643, Sigma), pH 7.2]. The adult *Xenopus laevis* (wide type, including three males and two females) aged from 1-year older were studied. All animal protocols in this study were approved by the Ethics Committee of the Zhejiang University Laboratory Animal Center (Lot number: 20201049 and ZJU20220195).

Fabrication of the microwell device. Firstly, a silicon plate (Suzhou chip scientific instrument Co., Ltd., Suzhou, China) containing 100,000 microwells whose diameter and depth were 28 and 35 μm respectively was manufactured. Secondly, a polydimethylsiloxane (PDMS) plate with the same number of micropillars was made using the silicon microwell plate. Lastly, the PDMS plate was used as a mold to make a disposable agarose microwell plate by pouring 5% agarose (Cat. No. BY-R0100, Baygene) solution onto the surface of the PDMS plate.

Synthesis of barcoded beads. For bead synthesis, 350 μl carboxyl magnetic beads (20–25 μm in diameter, 50 mg/ml, Suzhou Knowledge & Benefit Sphere Tech. Co., Ltd., Suzhou, China) were washed twice with 0.1 M 2-[N-morpholino] ethane-sulfonic acid (MES, Cat. No. A610341-0100, Sangon). The beads were suspended in a 635 μl of 0.1 M MES, 125 μl of 0.2 M MES, and 3.08 mg 1-Ethyl-3-(3-dimethyl aminopropyl) carbodiimide hydrochloride (EDC, Cat. No. C600433-0025, Sangon; 3.08 mg). Then 7.5 μl beads were then divided into each well of a 96-well plate. Amino-modified oligonucleotides (2.5 μl, 50 μM in 0.1 M MES, all the sequences used are listed in Supplementary Dataset 9 and were synthesized by Sangon Biotech Co., Ltd. with high-performance liquid chromatography purification) were then added to each well. Incubating it for 20 min at room temperature and a 0.5 μl mix (6 mg EDC in 100 μl of 0.1 M MES) was distributed into each well. Then repeated once. The beads were collected and washed in 1 ml of 0.1 M phosphate-buffered saline (PBS, Cat. No. DT20012, Dianrui) containing 0.02% Tween 20 after 80 min vortexing and incubation. The supernatant was carefully removed using a magnetic grate. Finally, the beads were then washed twice in 1 ml TE (pH 8.0).

In the second split-pool, the beads were divided into each well of a new 96-well plate after washing with water. A 15 μl polymerase chain reaction (PCR) mix (1× Phanta Master Mix, Cat. No. P510-01, Vazyme) and 5 μM oligonucleotides were added to each well. The oligonucleotides in each well included a sequence with reverse complementarity to linker 1, a unique barcode and a linker 2 sequence, and the PCR program was as follows: 94 °C for 5 min; five cycles of 94 °C for 15 s, 48.8 °C for 4 min, and 72 °C for 4 min; and a 4 °C hold. Notably, the beads were mixed between denaturation (95 °C) and primer annealing (48.8 °C) in each cycle. After PCR program, the beads were collected in a 95 °C water bath for 6 min and separated using a magnet, removing complementary chains, two times.

In the third split-pool, the PCR program was as follows: 94 °C for 5 min, 48.8 °C for 20 min, and 72 °C for 4 min and a 4 °C hold. The oligonucleotides used in each well included a linker 2 reverse-complementary sequence, a UMI sequence, a unique barcode, and a poly-T tail. After the PCR program, the beads were collected and suspended in 200 μl exonuclease I mix (containing 1× exonuclease I buffer and 1 U/μl exonuclease I) and incubated at 37 °C for 15 min with a rotary mixer. Then the beads were washed with 200 μl TE-TW [10 mM Tris (pH 8.0) (Cat. No. 15568-025, Invitrogen), 1 mM EDTA (Cat. No. AM9260G, Invitrogen), 0.01% Tween 20 (Cat. No. A600560-0500, Sangon), 200 μl of 10 mM Tris-HCl (pH 8.0), and 1 ml double-distilled water. Finally, the beads could be stored in 1 ml TE-TW for one month at 4 °C.

Cell preparation. Adult *Xenopus laevis* tissues were minced into pieces (~1 mm) at room temperature using sterilized scissors. The tissue pieces were then transferred to a 15 ml centrifuge tube, washed with DPBS, and suspended in a 5 ml solution containing various dissociation enzymes for different amounts of time until no tissue fragments were visible. Details for single-cell isolation from different tissues are listed in Supplementary Dataset 1. To collect dissociated cells, the centrifuge tube was at 300 × *g* for 5 min at 4 °C, removed the supernatant, and resuspended in 2 ml DPBS. Then the dissociated cells were passed through a 70 μm strainer (Cat. No. F613462-9001, Sangon). After being washed twice (centrifuged at 300 × *g* for 5 min at 4 °C), the cells were resuspended at a density of 1 × 10⁵ cells/ml in DPBS.

Tadpoles whole body were also minced into pieces (~1 mm) at room temperature using scissors. The tadpole pieces were transferred to a 15 ml centrifuge tube and dissociated in 5 ml 1x TrypLE containing collagenase I (0.25 mg/ml), collagenase II (0.25 mg/ml), collagenase IV (0.25 mg/ml) and collagenase V (0.25 mg/ml). The remaining steps were consistent with adult *Xenopus laevis* tissues mentioned above.

Cell collection and lysis. To eliminate the rate of the double cell during Microwell-seq, the cell concentration should be carefully controlled. The suitable cell concentration is about 100,000/ml (with 10% of the wells occupied by individual cells) and can be estimated using a hemocytometer. The cell suspension was pipetted onto the microwell plate and the plate was detected under a microscope to wash extra cells away. Then the bead suspension was loaded into the microwell plate which was placed on a magnet. The four sides of the microwell plate were
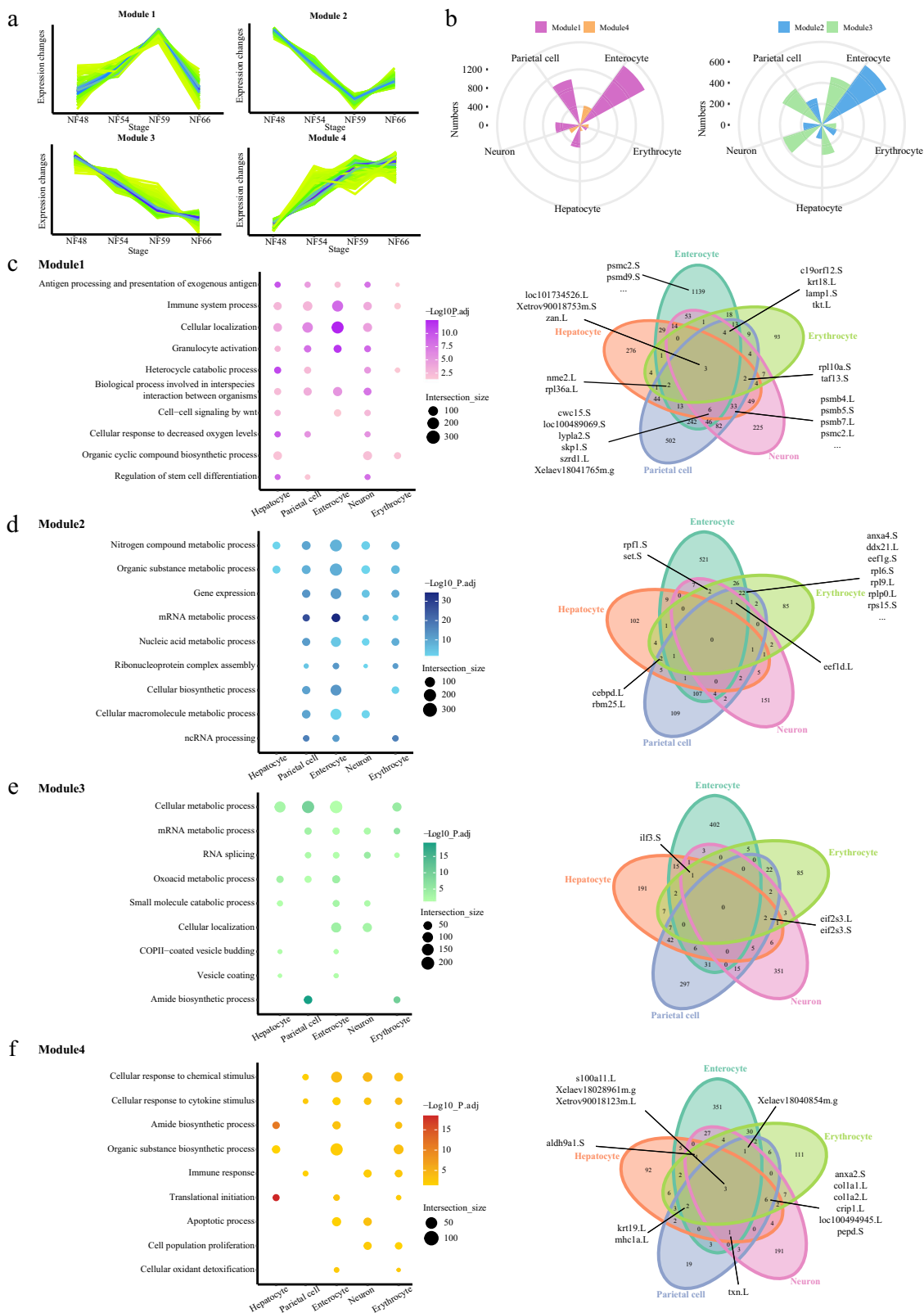
Fig. 6 Gene expression profiles of the remodeling multisystem during metamorphosis. a Schematic diagram of the four modules. **b** Rose diagrams showing the distribution of DEGs in each cell type. Different modules are labeled in different colors. **c–f** Representative GO terms enriched in module 1 (**c**), 2 (**d**), 3 (**e**), and 4 (**f**) for five cell types, respectively (left). GO terms enriched in different modules are labeled in different colors. Venn diagrams showing the numbers of shared genes in each module (right). The bold lines indicate the representative genes shared in at least three cell types. *P* value was calculated by the hypergeometric distribution, statistical test is one-sided, adjustments *P* values were made after *P* value is corrected by Benjamin–Hochberg multiple tests.

removed and excess beads were washed away slightly. Cold lysis buffer [0.1 M Tris-HCl (pH 7.5) (Cat. No. 15567-027, Invitrogen), 0.5 M LiCl (Cat. No. A100416-0025, Sangon), 1% sodium dodecyl sulfate (SDS, Cat. No. B548118-0100, Sangon), 10 mM EDTA, and 5 mM dithiothreitol (Cat. No. A620058-0005, Sangon)] was pipetted over the surface of the microwell plate and incubated for 12 min. Next, the microwell plate was turned over and the beads were collected in a 1.5 ml RNase-free tube. The collected beads were washed once with 1 ml of 6× SSC (Cat. No. AM9770, Invitrogen), followed by 500 µl of 6× SSC and 200 µl of 50 mM Tris-HCl pH 8.0.

Reverse transcription. Briefly, 20 µl RT mix [200 U SuperScript II reverse transcriptase (Cat. No. 2690 A, Takara), 1× Superscript II first-strand buffer (Cat. No. 2690 A, Takara), 20 U Murine RNase inhibitor (Cat. No. R301-2, Vazyme), 1 M betaine (Cat. No. 107-43-7, Sigma), 6 mM MgCl$_2$ (Cat. No. AM9530G, Invitrogen), 2.5 mM dithiothreitol (Cat. No. A620058-0005, Sangon), 1 mM deoxynucleoside triphosphate (dNTP, Cat. No. P031-1, Vazyme), and 1 µM TSO primer] was added to the collected beads and were incubated at 42 °C for 90 min on a rotary mixer to keep even reaction. The beads were then washed with 200 µl TE-SDS (1× TE + 0.5% SDS) to inactivate reverse transcriptase.

Exonuclease I treatment. The beads were washed with 200 µl TE-TW and 200 µl of 10 mM Tris-HCl (pH 8.0), pooled, and resuspended in 100 µl exonuclease I mix [1× exonuclease I buffer (Cat. No. B0293S, NEB), and 1 U/µl exonuclease I (Cat. No. M0293L, NEB)] and incubated at 37 °C for 30 min with mixing on a rotary mixer to remove single-stranded oligonucleotides without capturing mRNA. The beads were then washed once with TE-SDS, followed by 1 ml TE-TW and 200 µl of 10 mM Tris-HCl (pH 8.0).

cDNA amplification. The beads were divided equally into four PCR tubes with 12.5 µl PCR mix [1× HiFi HotStart Readymix (Cat. No. KK2601, Kapa Biosystems.) and 0.1 µM TSO-PCR primer]. The PCR program was as follows: 98 °C for 3 min; six cycles of 98 °C for 20 s, 65 °C for 45 s, and 72 °C for 6 min; 72 °C for 10 min; and a 4 °C hold. The beads were mixed between denaturation (95 °C) and primer annealing (65 °C) in each cycle. Next, all PCR products were pooled together and VAHTS DNA Clean Beads (Cat. No. N411-03, Vazyme) were used to purify the pooling cDNA samples. To purify the DNA sample, a 25 µl PCR mix (1× HiFi HotStart Readymix and 0.1 µM TSO-PCR primer) was added again. The PCR program was as follows: 10 to 12 cycles of 98 °C for 3 min, 98 °C for 20 s, 67 °C for 20 s, and 72 °C for 6 min; 72 °C for 10 min; and a 4 °C hold. VAHTS DNA Clean beads were then used to purify the DNA sample and get the cDNA library.

Transposase fragmentation and selective PCR. The purified cDNA libraries were fragmented using a customized transposase that carries two identical insertion sequences. The customized transposase was included in the TruePrep Homo-N7 DNA Library Prep Kit for Illumina (Cat. No. TD513, Vazyme) or TruePrep Homo-N7 DNA Library Prep Kit for MGI (Cat. No. L-N7E461L0, Vazyme). The fragmentation reaction was performed according to the instructions provided by the manufacturer. We used customed P5 primer (listed in Supplementary Dataset 9) and VAHTS RNA Adapters for Illumina (N8XX, Cat. No. TD203-1, Vazyme) or our MGI P7 primers (M8XX, listed in Supplementary Dataset 9) to specifically amplify fragments that contain the 3′ ends of transcripts. Other fragments will form self-loops, impeding their binding to PCR primers. The PCR program was as follows: 72 °C for 3 min; 98 °C for 1 min; five cycles of 98 °C for 15 s, 60 °C for 30 s, and 72 °C for 3 min; 72 °C for 5 min; and a 4 °C hold. The PCR product was purified using 0.9× VAHTS DNA Clean beads. Then, a 25 µL PCR mix (1×HiFi HotStart Readymix and 0.2 µM 2100 primer) was added to each sample. The PCR program was as follows: 95 °C for 3 min; five cycles of 98 °C for 20 s, 60 °C for 15 s, and 72 °C for 15 s; 72 °C for 3 min; and a 4 °C hold. To eliminate primer dimers and large fragments, 0.55-0.15× VAHTS DNA Clean beads were then used to purify the cDNA library. The size distribution of the products was analyzed on an Agilent 2100 bioanalyzer, and a peak in the 400 to 700 bp range was observed. Finally, the samples were subjected to sequencing on the Illumina HiSeq or MGI DNBSEQ-T7. For MGI sequencing, we applied the protocol provided by VAHTS Circularization Kit for MGI (Cat. No. NM201-01, Vazyme) to obtain single-stranded circular cDNA available for DNB (DNA Nanoball) generation. We also replaced the official R1 sequencing primers with our custom R1 sequencing primers A&B (listed in Supplementary Dataset 9) to ensure the completion of the sequencing.

Quantification and statistical analysis

Preprocessing of Microwell-seq sequence data. Microwell-seq datasets were processed using the protocols described by ref. [71]. The raw fastq-format sequencing data from a DNBSEQ-T7 were first split into i7 indexed sub-libraries using "splitBarcode [https://github.com/MGI-tech-bioinformatics/splitBarcode]". Reads from *Xenopus laevis* were aligned to the *Xenopus laevis* v9.1 genome, using STAR[72]. The DGE data matrices were obtained using the modified Dropseq tools (https://github.com/ggjlab/mca_data_analysis/tree/master/preprocessing/Dropseq_tools-1.12/) and the corresponding protocol is available at http://mccarrolllab.org/dropseq/. The DGE data containing the top 30,000 cells sorted by the total number of transcripts were obtained after preprocessing. For quality control, we filtered out cells in which fewer than 500 transcripts were detected.

Correction of RNA contamination. During scRNA-seq data analysis, we commonly assume that all RNAs are endogenous to each individual cell. It has been recognized that some contaminating nonendogenous RNAs are also present even within datasets of the highest quality[73]. While applying Microwell-seq technology, there is a probability that a certain amount of cell-free RNA from the input solution admixed with the cell in a well are also sequenced. Although the amount of cell-free RNAs is very low compared with the total RNA content, it is still important to recognize and correct this contamination. To solve this problem, we strictly removed the batch gene background. We assumed that, for each batch of experiments, cell barcodes with fewer than 500 UMIs correspond to empty beads exposed to free RNA during cell lysis, RNA capture, and washing steps. Genes with extensive-expression in all beads were considered batch genes. The batch gene background value was defined as the average gene detection level for all cell barcodes with fewer than 500 UMIs, multiplied by the median of the fold difference between the detected gene expression of a cell and the average detected gene expression for beads with fewer than 500 UMIs and then rounded to the nearest integer. We subtracted the batch gene background for each cell from the digital expression matrix before performing the cross-tissue or cross-stage comparisons. We used the background-removed matrix to perform downstream analyses.

Filter out potential doublets. scRNA-seq data are commonly influenced by technical artifacts known as doublets, in which two or more different cells receive the same barcode, resulting in an aggregated transcriptome. We used the R package DoubletFinder[74] to detect the potential doublets in each individual library. The overall doublet rate for our experiment was no more than 1.2%[12]. Approximately 5% of cells were labeled as doublets and were removed.

Cell clustering for single-cell datasets on a per-dataset basis. Seurat[75] was used as a tool for cell clustering on a per-dataset basis. The data were $\log_2$ (counts per million (CPM)/100 + 1) transformed, and the number of UMIs and the percentage of mitochondrial gene content were regressed. A total of 2000 genes were selected through the "FindVariableFeatures" function by using the "vst" method and used as inputs for initial principal component analysis (PCA) and the number of principal components (PCs) used for nonlinear dimensional reduction (t-SNE). For clustering, we set different resolution parameters between 0.5 and 2.5 in the "FindCluster" function and chose the final cluster number by distinguishing differential genes among clusters. These parameters, including the resolution and number of PCs, were adjusted on a per-dataset basis.

Batch effect evaluation and clustering analysis for single-cell datasets on all dataset bases. The gene expression matrices for all single cells were merged together and fed to Seurat for clustering analysis. Briefly, the data were first processed by the "SCTransform" function. After that, processed data from larva were subjected to the Seurat standard procedure as described above. We chose 50 PCs for PCA and computed the neighborhood graph of cells. We then used the "FindCluster" function to cluster cells with a resolution of 0.8. Processed data from adults were fed to Scanpy[76]. We chose 100 PCs for PCA and computed the neighborhood graph of cells. We then used Leiden clustering to cluster cells with a resolution of 1 and neighbor number of 15. Finally, 57 clusters for larvae and 106 clusters for adults were produced, and marker genes were defined by the Wilcoxon rank-sum test. t-SNE was applied to visualize the single-cell transcriptional profile in 2D space. The Wilcoxon rank-sum test was used by running the 'FindAllMarkers' function in Seurat and the 'rank genes groups' function in Scanpy to find DEGs in each cluster. We annotated each cell type with marker genes selected based on an extensive literature review.

Collection of orthologous genes. Human, mouse, and zebrafish orthologous pairs were obtained from Ensembl v96 by BioMarkt. *Xenopus laevis* and human orthologous pairs were obtained from "Xenbase Home [http://www.xenbase.org/common/]". In this study, we considered one-to-one orthologous and one-to-many paralogous pairs for further analysis.

Cross-species transcriptome comparison. Before transcriptome comparison, we collected human, mouse, and zebrafish cell-gene matrices and cell annotations from the previously published HCL, MCA, and ZCL datasets (https://db.cngb.org/HCL/; http://bis.zju.edu.cn/MCA; http://bis.zju.edu.cn/ZCA). We extracted the cells in the adult stage and divided the 102 human, 104 mouse, and 135 zebrafish cell clusters into 12 major cell lineages. To reduce the effects of noise and outliers, we randomly sampled 100,000 cells for each species, and we calculated the pseudocells, which was an average of every 50 cells from the same cell type, for further analysis. The gene expression data for each species underwent normalization to the sum of transcripts and multiplication by 100,000. The combined projection was visualized with the "SAMAP.scatter" function. To compare transcriptomes across species, we applied SAMap analysis between *Xenopus laevis* and humans, mice, and zebrafish. The cell type pairs with mapping scores higher than 0.1 were plotted between species by the package Circlize.

Lineage-specific transcription factor (TF) analysis. In this study, *Xenopus laevis* TFs were defined as the one-to-one orthologs of TFs found in *Xenopus tropicalis* downloaded from AnimalTFDB. We applied the same method used in previous research to judge cell-type-specific TFs for *Xenopus laevis*. We calculated the Jensen–Shannon divergence (JSD) for every TF expression status and defined the TF specificity score as $1 - \sqrt{JSD}$. The Z-score-normalized TF specificity score was used to infer the specific TFs in each lineage. With reference to lineage-specific TF data for humans, mice, and zebrafish, we identified TFs conserved in four species for each lineage.

Identification of TFs that drive cell type mapping during metamorphosis. We define g_1 and g_2 to contain TFs from stage 1 and stage 2, respectively. Let $X_{a_1 a_2}$ denote the set of all cells with cross-stage edges between cell types a_1 and a_2; JSD was used to evaluate cell-type-specific TFs for each stage such as stage1: $Y_{1_{g_1}} = 1 - \sqrt{JSD}$; We assign each TF a score: $h_g = T\left(S(Y_{1, g1})\right) \circ T\left(S(Y_{2, g2})\right)$, where $S(Z)$ standardizes vector Z to have zero mean and unit variance, and $T(Z)$ sets negative values in vector Z to zero in order to ignore weakly expressed genes. To be inclusive, the top 5 TFs according to h_g are identified as "drive TFs".

Gene expression trajectory analysis during metamorphosis. For module analysis, we used a similar method described in Kazer et al.[77]. We identified the DEGs between each stage using the "FindMarkers" function in Seurat. The parameters were logfc.threshold $= 0.1$, min.pct $= 0.1$ and min.diff.pct $= -Inf$. Then, we computed averaged expression values for each gene using the "AverageExpression" function in Seurat and grouped them into 15–30 clusters using the Mfuzz package in R with the fuzzy c-means algorithm[78].

To determine whether cluster expression varied over time, every cell was scored for the genes within the cluster using the "AddModuleScore" function in Seurat. Then, 1000 two-sided Wilcoxon rank-sum tests were performed between the distribution of scores from 150–1000 cells at the first stage and the same number of cells from each other stage. For each stage, the P values from the 1000 tests were averaged. After FDR correction, if $q < 1e-10$ for any stage, the module was considered to vary significantly in expression in that stage.

Finally, these clusters were grouped into four modules using one-sided Wilcoxon rank-sum tests: module 1 was defined as the genes with the highest expression at NF59, module 2 was defined as the genes with the lowest expression at NF59, module 3 was defined as the genes that were consistently downregulated during metamorphosis, and module 4 was defined as the genes that were consistently upregulated.

Gene ontology analysis. Gene ontology analysis was performed on *Homo sapiens* orthologs using "g:profiler [https://biit.cs.ut.ee/gprofiler/gost]", and all orthologous genes were taken as the universe. We selected the Benjamini–Hochberg FDR to compute multiple testing corrections and considered biological pathways with p values smaller than 0.05.

Statistics and reproducibility. No statistical method was used to predetermine sample size. No data were excluded from the analyses and all analyses were not randomized.

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

Data availability

Code availability

References

1. Briggs, J. A. et al. The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. *Science* https://doi.org/10.1126/science.aar5780 (2018).
2. Bright, A. R. et al. Combinatorial transcription factor activities on open chromatin induce embryonic heterogeneity in vertebrates. *EMBO J.* **40**, e104913 (2021).
3. Sonam, S., Bangru, S., Perry, K. J., Kalsotra, A. & Henry, J. J. Single-cell analyses of the corneal epithelium: Unique cell types 1 and gene expression profiles. Preprint at bioRxiv https://doi.org/10.1101/2020.08.06.240036 (2020).
4. Corkins, M. E. et al. A comparative study of cellular diversity between the Xenopus pronephric and mouse metanephric nephron. Preprint at *bioRxiv* https://doi.org/10.1101/2022.01.11.475739 (2022).
5. Lin, T.-Y. et al. Fibroblast dedifferentiation as a determinant of successful regeneration. *Dev. Cell* **56**, 1541–1551.e1546 (2021).
6. Aztekin, C. et al. Secreted inhibitors drive the loss of regeneration competence in Xenopus limbs. *Development* https://doi.org/10.1242/dev.199158 (2021).
7. Kakebeen, A. D., Chitsazan, A. D., Williams, M. C., Saunders, L. M. & Wills, A. E. Chromatin accessibility dynamics and single cell RNA-Seq reveal new regulators of regeneration in neural progenitors. *Elife* https://doi.org/10.7554/eLife.52648 (2020).
8. Aztekin, C. et al. Identification of a regeneration-organizing cell in the Xenopus tail. *Science* **364**, 653–658 (2019).
9. Pelzer, D. et al. Foxm1 regulates neural progenitor fate during spinal cord regeneration. *EMBO Rep.* https://doi.org/10.15252/embr.202050932 (2021).
10. Han, X. P. et al. Construction of a human cell landscape at single-cell level. *Nature* **581**, 303–309 (2020).
11. Qu, J. et al. A reference single-cell regulomic and transcriptomic map of cynomolgus monkeys. *Nat. Commun.* **13**, https://doi.org/10.1038/s41467-022-31770-x (2022).
12. Han, X. et al. Mapping the mouse cell atlas by microwell-seq. *Cell* **173**, 1307 (2018).
13. Spanjaard, B. et al. Simultaneous lineage tracing and cell-type identification using CRISPR-Cas9-induced genetic scars. *Nat. Biotechnol.* **36**, 469–473 (2018).
14. Li, H. et al. Fly cell atlas: a single-nucleus transcriptomic atlas of the adult fruit fly. *Science* **375**, eabk2432 (2022).
15. Cao, C. et al. Comprehensive single-cell transcriptome lineages of a proto-vertebrate. *Nature* **571**, 349–354 (2019).
16. Cao, J. et al. Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* **357**, 661–667 (2017).
17. Fincher, C. T., Wurtzel, O., de Hoog, T., Kravarik, K. M. & Reddien, P. W. Cell type transcriptome atlas for the planarian Schmidtea mediterranea. *Science* https://doi.org/10.1126/science.aaq1736 (2018).
18. Levy, S. et al. A stony coral cell atlas illuminates the molecular and cellular basis of coral symbiosis, calcification, and immunity. *Cell* **184**, 2973–2987.e2918 (2021).
19. Marshall, L. N. et al. Stage-dependent cardiac regeneration in Xenopus is regulated by thyroid hormone availability. *Proc. Natl Acad. Sci. USA* **116**, 3614–3623 (2019).
20. Tata, J. R. Gene expression during metamorphosis: an ideal model for post-embryonic development. *Bioessays* **15**, 239–248 (1993).
21. Sin, W. C., Haas, K., Ruthazer, E. S. & Cline, H. T. Dendrite growth increased by visual activity requires NMDA receptor and Rho GTPases. *Nature* **419**, 475–480 (2002).
22. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).
23. Crow, M., Paul, A., Ballouz, S., Huang, Z. J. & Gillis, J. Characterizing the replicability of cell types defined by single cell RNA-sequencing data using MetaNeighbor. *Nat. Commun.* **9**, 884 (2018).
24. Zhang, C. et al. The transcription factor NKX2-2 regulates oligodendrocyte differentiation through domain-specific interactions with transcriptional corepressors. *J. Biol. Chem.* **295**, 1879–1888 (2020).
25. Kohler, S., Winkler, U. & Hirrlinger, J. Heterogeneity of astrocytes in grey and white matter. *Neurochem. Res.* **46**, 3–14 (2021).
26. Zuasti, A., Jimenez-Cervantes, C., Garcia-Borron, J. C. & Ferrer, C. The melanogenic system of Xenopus laevis. *Arch. Histol. Cytol.* **61**, 305–316 (1998).
27. Tarashansky, A. J. et al. Mapping single-cell atlases throughout Metazoa unravels cell type evolution. *Elife* https://doi.org/10.7554/eLife.66747 (2021).
28. Wang, J. et al. Tracing cell-type evolution by cross-species comparison of cell atlases. *Cell Rep.* **34**, 108803 (2021).
29. Jiang, M. et al. Characterization of the zebrafish cell landscape at single-cell resolution. *Front. Cell Dev. Biol.* https://doi.org/10.3389/fcell.2021.743421 (2021).
30. Hadji-Azimi, I., Coosemans, V. & Canicatti, C. Atlas of adult Xenopus laevis laevis hematology. *Dev. Comp. Immunol.* **11**, 807–874 (1987).
31. Nogawa-Kosaka, N. et al. Identification of erythroid progenitors induced by erythropoietic activity in Xenopus laevis. *J. Exp. Biol.* **214**, 921–927 (2011).
32. Popescu, D. M. et al. Decoding human fetal liver haematopoiesis. *Nature* **574**, 365–371 (2019).
33. Geirsdottir, L. et al. Cross-species single-cell analysis reveals divergence of the primate microglia program. *Cell* **179**, 1609–1622.e1616 (2019).
34. Sueda, R. & Kageyama, R. Regulation of active and quiescent somatic stem cells by Notch signaling. *Dev. Growth Differ.* **62**, 59–66 (2020).

35. Stoeck, A. et al. Discovery of biomarkers predictive of GSI response in triple-negative breast cancer and adenoid cystic carcinoma. *Cancer Disco.* **4**, 1154–1167 (2014).

36. Phinney, D. G. Twist, epithelial-to-mesenchymal transition, and stem cells. *Stem Cells* **29**, 3–4 (2011).

37. Coppiello, G. et al. Meox2/Tcf15 heterodimers program the heart capillary endothelium for cardiac fatty acid uptake. *Circulation* **131**, 815–826 (2015).

38. Gohn, C. R., Blue, E. K., Sheehan, B. M., Varberg, K. M. & Haneline, L. S. Mesenchyme homeobox 2 enhances migration of endothelial colony forming cells exposed to intrauterine diabetes mellitus. *J. Cell Physiol.* **232**, 1885–1892 (2017).

39. Doty, R. T. et al. Single-cell analyses demonstrate that a heme-GATA1 feedback loop regulates red cell differentiation. *Blood* **133**, 457–469 (2019).

40. Horie, K. et al. Down-regulation of GATA1-dependent erythrocyte-related genes in the spleens of mice exposed to a space travel. *Sci. Rep.* **9**, 7654 (2019).

41. Daly, M. E. Transcription factor defects causing platelet disorders. *Blood Rev.* **31**, 1–10 (2017).

42. Milner, J. J. et al. Runx3 programs CD8(+) T cell residency in non-lymphoid tissues and tumours. *Nature* **552**, 253–257 (2017).

43. Sekizar, S. et al. Remyelination by resident oligodendrocyte precursor cells in a Xenopus laevis inducible model of demyelination. *Dev. Neurosci.* **37**, 232–242 (2015).

44. Nitta, K. R. et al. Expression of Sox1 during Xenopus early embryogenesis. *Biochem. Biophys. Res. Commun.* **351**, 287–293 (2006).

45. Baltzinger, M., Mager-Heckel, A. M. & Remy, P. Xl erg: expression pattern and overexpression during development plead for a role in endothelial cell differentiation. *Dev. Dyn.* **216**, 420–433 (1999).

46. Cermenati, S. et al. Sox18 and Sox7 play redundant roles in vascular development. *Blood* **111**, 2657–2666 (2008).

47. Yang, W. Y. et al. Coronary risk in relation to genetic variation in MEOX2 and TCF15 in a Flemish population. *BMC Genet.* **16**, 116 (2015).

48. Betancur, P., Bronner-Fraser, M. & Sauka-Spengler, T. Genomic code for Sox10 activation reveals a key regulatory enhancer for cranial neural crest. *Proc. Natl Acad. Sci. USA* **107**, 3570–3575 (2010).

49. Zhang, L. et al. Small molecules efficiently reprogram human astroglial cells into functional neurons. *Cell Stem Cell* **17**, 735–747 (2015).

50. Jakovcevski, I. & Zecevic, N. Olig transcription factors are expressed in oligodendrocyte and neuronal cells in human fetal CNS. *J. Neurosci.* **25**, 10064–10073 (2005).

51. Papes, F. et al. Transcription factor 4 loss-of-function is associated with deficits in progenitor proliferation and cortical neuron content. *Nat. Commun.* **13**, 2387 (2022).

52. Yuan, F. et al. LHX6 is essential for the migration of human pluripotent stem cell-derived GABAergic interneurons. *Protein Cell* **11**, 286–291 (2020).

53. Dai, J. X., Bercury, K. K., Ahrendsen, J. T. & Macklin, W. B. Olig1 function is required for oligodendrocyte differentiation in the mouse brain. *J. Neurosci.* **35**, 4386–4402 (2015).

54. Bronchain, O. J. et al. The olig family: phylogenetic analysis and early gene expression in Xenopus tropicalis. *Dev. Genes Evol.* **217**, 485–497 (2007).

55. Furlow, J. D. & Neff, E. S. A developmental switch induced by thyroid hormone: Xenopus laevis metamorphosis. *Trends Endocrinol. Metab.* **17**, 40–47 (2006).

56. Heimeier, R. A., Das, B., Buchholz, D. R., Fiorentino, M. & Shi, Y. B. Studies on Xenopus laevis intestine reveal biological pathways underlying vertebrate gut adaptation from embryo to adult. *Genome Biol.* **11**, R55 (2010).

57. Brown, D. D. & Cai, L. Amphibian metamorphosis. *Dev. Biol.* **306**, 20–33 (2007).

58. Schreiber, A. M., Cai, L. & Brown, D. D. Remodeling of the intestine during metamorphosis of Xenopus laevis. *Proc. Natl Acad. Sci. USA* **102**, 3720–3725 (2005).

59. Ju, J. Y. et al. Human fetal globin gene expression is regulated by LYAR. *Nucleic Acids Res.* **42**, 9740–9752 (2014).

60. Kassouf, M. T. et al. Genome-wide identification of TAL1's functional targets: Insights into its mechanisms of action in primary erythroid cells. *Genome Res.* **20**, 1064–1083 (2010).

61. Mukhi, S., Cai, L. Q. & Brown, D. D. Gene switching at Xenopus laevis metamorphosis. *Dev. Biol.* **338**, 117–126 (2010).

62. Maniatis, G. M. & Ingram, V. M. Erythropoiesis during amphibian metamorphosis: I. Site of maturation of erythrocytes in Rana catesbeiana. *J. Cell Biol.* **49**, 372–379 (1971).

63. Jurd, R. D. & Maclean, N. An immunofluorescent study of the haemoglobins in metamorphosing Xenopus laevis. *J. Embryol. Exp. Morphol.* **23**, 299–309 (1970).

64. Widmer, H. J., Hosbach, H. A. & Weber, R. Globin gene expression in Xenopus laevis: anemia induces precocious globin transition and appearance of adult erythroblasts during metamorphosis. *Dev. Biol.* **99**, 50–60 (1983).

65. Flajnik, M. F. et al. Major histocompatibility complex-encoded class I molecules are absent in immunologically competent Xenopus before metamorphosis. *J. Immunol.* **137**, 3891–3899 (1986).

66. Wong, M. L., Dong, C., Maestre-Mesa, J. & Licinio, J. Polymorphisms in inflammation-related genes are associated with susceptibility to major depression and antidepressant response. *Mol. Psychiatr.* **13**, 800–812 (2008).

67. Flajnik, M. F. & Dupasquier, L. Mhc class-I antigens as surface-markers of adult erythrocytes during the metamorphosis of xenopus. *Dev. Biol.* **128**, 198–206 (1988).

68. Flores, I. L. et al. EEF1D modulates proliferation and epithelial-mesenchymal transition in oral squamous cell carcinoma. *Clin. Sci.* **130**, 785–799 (2016).

69. Liu, Y. et al. LAIR-1 suppresses cell growth of ovarian cancer cell via the PI3K-AKT-mTOR pathway. *Aging* **12**, 16142–16154 (2020).

70. Perry, S. F. & Sander, M. Reconstructing the evolution of the respiratory apparatus in tetrapods. *Respir. Physiol. Neurobiol.* **144**, 125–139 (2004).

71. Han, X. et al. Mapping the mouse cell atlas by microwell-seq. *Cell* **172**, 1091–1107.e1017 (2018).

72. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

73. Zheng, G. X. et al. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).

74. McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst.* **8**, 329–337.e324 (2019).

75. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e1821 (2019).

76. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).

77. Kazer, S. W. et al. Integrated single-cell analysis of multicellular immune dynamics during hyperacute HIV-1 infection. *Nat. Med.* **26**, 511–518 (2020).

78. Kumar, L. & E Futschik, M. Mfuzz: a software package for soft clustering of microarray data. *Bioinformation* **2**, 5–7 (2007).

79. Liao, Y. et al. Cell landscape of larval and adult Xenopus laevis at single-cell resolution (2022).

Author contributions

The project was conceived by X.H. and G.G. Tissue digestion were performed by Y.L., L.Y., and X.F. Experiments were performed by Y.L., F.R., H.C., Z.L., Xueyi W., and Xinru W. Bioinformatics analysis was performed by L.M., W.E., Q.G., X.F., Z.S., J.W., P.Z., and H.S. Website construction was performed by M.C., Y.Z., L.M., and Q.G. Frogs were provided by W.S. and supported on animal housing and maintenance. The paper was written by X.H., G.G., and W.S. and funding was acquired by X.H., G.G., and Y.L.

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/s41467-022-31949-2.

Correspondence and requests for materials should be addressed to Wanhua Shen, Guoji Guo or Xiaoping Han.

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

Reprints and permission 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.