

Comprehensive single-cell transcriptome lineages of a proto-vertebrate

Chen Cao^{1,5}, Laurence A. Lemire^{1,5}, Wei Wang², Peter H. Yoon³, Yoolim A. Choi³, Lance R. Parsons¹, John C. Matese¹, Wei Wang¹, Michael Levine^{1,3*} & Kai Chen^{1,4*}

Ascidian embryos highlight the importance of cell lineages in animal development. As simple proto-vertebrates, they also provide insights into the evolutionary origins of cell types such as cranial placodes and neural crest cells. Here we have determined single-cell transcriptomes for more than 90,000 cells that span the entirety of development—from the onset of gastrulation to swimming tadpoles—in *Ciona intestinalis*. Owing to the small numbers of cells in ascidian embryos, this represents an average of over 12-fold coverage for every cell at every stage of development. We used single-cell transcriptome trajectories to construct virtual cell-lineage maps and provisional gene networks for 41 neural subtypes that comprise the larval nervous system. We summarize several applications of these datasets, including annotating the synaptome of swimming tadpoles and tracing the evolutionary origin of cell types such as the vertebrate telencephalon.

Single-cell RNA-seq methods are revolutionizing our understanding of how cells are specified to become definitive tissues during development^{1–5}. These studies allow the elucidation of virtual lineages for select tissues, and provide detailed expression profiles for cell types such as pluripotent progenitor cells. However, a limitation of previous studies has been the incomplete coverage of vertebrate embryos, owing to the large numbers of cells present in these embryos.

As one of the closest living relatives of vertebrates⁶, the ascidian *C. intestinalis* serves a critical role in understanding developmental and physiological processes that are comparable to—but far less complex than—those of vertebrates. In comparison to vertebrate embryos, ascidian embryos are simple: gastrulating embryos are composed of only 100–200 cells, and swimming tadpoles contain about 2,500 cells. Owing to these small numbers of cells, it is possible to obtain comprehensive coverage of every cell type during development, including rare neuronal subtypes.

Here we extend insights into the regulatory ‘blueprint’ that spans the early phases of embryogenesis⁷ by profiling the transcriptomes of individual cells in sequentially staged *Ciona* embryos, from gastrulation at the 110-cell stage to the neurula and larval stages. Reconstructed temporal expression profiles reveal the specification and differentiation of individual cell types. Nearly 40 subtypes of neurons were identified, even though the central nervous system of the *Ciona* larva is composed of only 177 neurons⁸. The resulting high-resolution transcriptome trajectories, regulatory cascades and provisional gene networks provide insights into the evolution of novel cell types in vertebrates, including those of the telencephalon.

Specification of cell fate

Synchronized embryos from ten different stages of development were rapidly dissociated in RNase-free calcium-free synthetic seawater, and individual cells were processed in the 10x Genomics Chromium system with at least two biological replicates for each developmental stage (Fig. 1a, Extended Data Fig. 1, Supplementary Table 1, Methods). The staged embryos span all of the hallmark processes of development, beginning with gastrulation and culminating in swimming tadpoles (at which point all larval cell types, tissues and organs are formed)

(Fig. 1b). In total, we profiled 90,579 cells, which corresponds to an average of over 12-fold coverage for every cell across each of the sampled stages (Supplementary Table 1). Individual cells were sequenced to an average depth of about 12,000 unique molecular identifiers, which enabled the recovery of rare populations such as germ cells (which constitute about 0.1% of cells in swimming tadpoles).

t-distributed stochastic neighbour-embedding (*t*-SNE) projections of the transcriptomes at all ten stages of development identified coherent clusters of individual tissues, including heart, tail muscles, endoderm, notochord, germ cells, epidermis, nervous system and mesenchyme (Extended Data Fig. 2a–l). Several tissues—such as the nervous system and mesenchyme—exhibit a progressive increase in cell complexity during development (Extended Data Fig. 2c–l), which results in the appearance of more cell clusters at later stages of embryogenesis. We also found that most individual tissues displayed less variation in their transcriptome profiles during development, when compared with divergent cell types at the same time points. This is particularly evident for the developing germ line, because it is transcriptionally quiescent during the time frame of our analysis⁹.

The specific and stable expression of cell-specific marker genes (Extended Data Fig. 2m, Supplementary Table 2), such as *Brachyury* for the developing notochord and *Twist-like-2* for the mesenchyme^{10,11}, facilitated the reconstruction of temporal profiles for different tissues. This study also identified a variety of genes, including *Kdm8* (a histone H3K36me2 demethylase expressed in mesenchyme lineages), as tissue-specific markers.

Classical cell-lineage studies suggest that all of the major tissues of the ascidian tadpole are specified before gastrulation, at the 110-cell stage¹² (Fig. 1c). Most of the internal organs—including the notochord, endoderm, tail muscles, heart, germ cells, and regions of the nervous system—are derived from vegetal lineages. By contrast, animal blastomeres give rise to ectodermal derivatives, including epidermis and associated sensory neurons, and regions of the nervous system. Each of these cell types was identified as a discrete cluster in the *t*-SNE projections of dissociated 110-cell embryos (Fig. 1d).

Several tissues are already seen to segregate into distinct anterior and posterior clusters by the 110-cell stage, including the notochord,

¹Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ, USA. ²Stowers Institute for Medical Research, Kansas City, MO, USA. ³Department of Molecular Biology, Princeton University, Princeton, NJ, USA. ⁴Present address: The Yunnan Key Laboratory of Primate Biomedical Research, Institute of Primate Translational Medicine, Kunming University of Science and Technology, Kunming, China. ⁵These authors contributed equally: Chen Cao, Laurence A. Lemire. *e-mail: msl2@princeton.edu; chenk@lpbr.cn

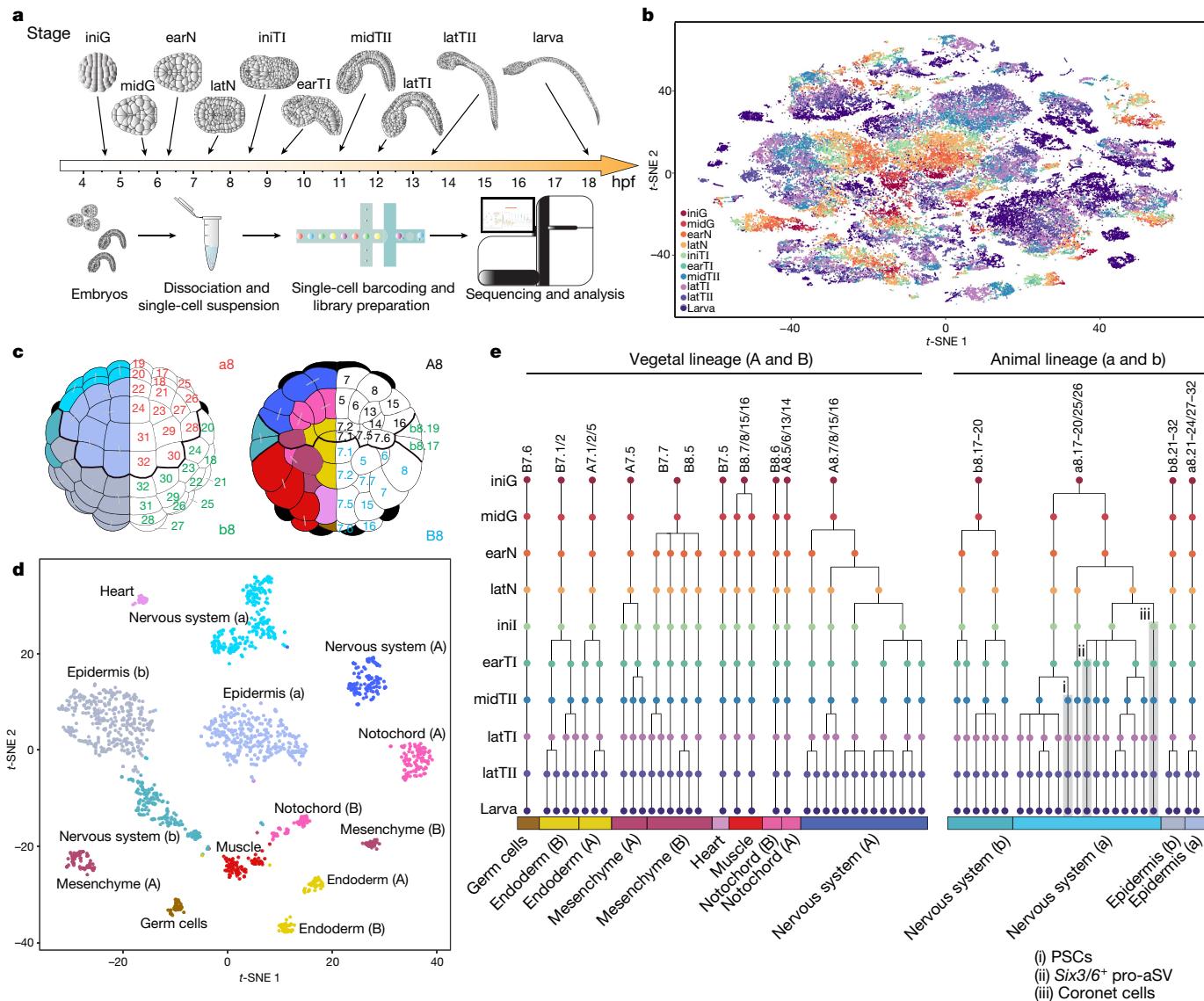


Fig. 1 | Overview of single-cell RNA-seq assays and cell-type specification at the onset of gastrulation. **a**, Staged embryos were collected from ten developmental stages: the initial gastrula (iniG), mid-gastrula (midG), early neurula (earN), late neurula (latN), initial tailbud I (iniTI), early tailbud I (earTI), mid-tailbud II (midTI), late tailbud I (latTI), late tailbud II (latTII) and larval (larva) stages. $n = 2$ biological replicates per stage (iniG to latTII stages); $n = 3$ biological replicates (larval stage). hpf, hours post-fertilization. **b**, t-SNE plot of the entire dataset ($n = 90,579$ cells). Cells are colour-coded according to developmental stage. **c**, Schematics of animal (left) and vegetal (right) blastomeres of a *Ciona* embryo at the initial gastrula stage. Tissue types

endoderm and lateral plate sensory cells. These observations validate and extend classical evidence for the specification of all major larval tissues at the 110-cell stage. Our expression profiles of individual cell types revealed previously known and newly identified potential fate determinants (Extended Data Fig. 3, Supplementary Table 2): for example, *Irx-B* is specifically expressed in a-lineage (anterior) epidermis and *Not* is expressed in the b-lineage (posterior).

Reconstructing cell lineages

The alignment of transcriptome profiles of individual cell types at sequential stages of development enabled the reconstruction of virtual lineage maps (Fig. 1e, Methods). In total, 60 cell types were identified in swimming tadpoles, and the corresponding virtual lineages of these cell types could be traced to blastomeres at the 110-cell stage (the time of fate restriction). The reconstructed lineages are in close

agreement with known lineage information, and provide insights into the specification and differentiation of individual cell types. For example, the transcriptome profiles accurately capture the muscle and heart lineages (Extended Data Fig. 4a, b), as well as the primary (from A8 blastomeres) and secondary (from B8 blastomeres) lineages of the notochord¹² (Extended Data Fig. 5). The mesenchyme has previously been shown to be derived from three separate lineages (from A7.6, B7.7 and B8.5 blastomeres)¹¹, and our analyses suggest they segregate to produce nine cell types (Extended Data Fig. 4c, d). Similarly, the head and trunk endoderm produce seven cell types (Extended Data Fig. 4e). This is a considerably higher level of resolution than that obtained by conventional experimental studies^{12,13}.

The transcriptome maps also capture more-nuanced lineage information. For example, dopaminergic neurons (coronet cells) of the central nervous system were found to share a common lineage with the

pro-anterior sensory vesicle, the anterior-most terminus of the neural tube that fuses with the stomodeum to form the neuropore¹⁴. Both derivatives share a common origin with palp sensory cells, which arise from the non-neural proto-placodal territory located immediately anterior of the neural tube (Fig. 1e)—this is consistent with the model for the evolution of the vertebrate telencephalon discussed in ‘Evolution of cell types’ below.

Transitional properties of the notochord

The notochord is a derivative of the mesoderm, and is a defining innovation of chordates¹⁵. However, the notochord exhibits distinctive properties in cephalochordates and vertebrates. Cephalochordates such as *Amphioxus* contain a muscular notochord that helps to power movements of the tail¹⁶, whereas the vertebrate notochord is non-muscular and provides structural support for derivatives of the paraxial mesoderm. The *Ciona* notochord appears to contain a mixture of both properties.

The primary (A-lineage) and secondary (B-lineage) notochord cells are clearly resolved into subclusters throughout development (Extended Data Fig. 5a). By constructing single-cell trajectories, it was possible to identify cell signalling and regulatory genes in each lineage (Extended Data Fig. 5b, c). In addition to the identification of genes that are known to be differentially expressed in the two lineages (such as *ZicL* and *Notch*)^{17,18}, we were able to identify distinctive regulatory strategies for the two lineages (Extended Data Fig. 5b, c). For example, *Otx* and *Not* are specifically expressed in the secondary notochord, along with the muscle determinants *Tbx6a*, *Tbx6c* and *Tbx6d*¹⁹ (Extended Data Fig. 3). They precede expression of muscle identity genes such as calsequestrin (*Casq1/2*; *solidi* in *Ciona* gene symbols separate multiple vertebrate homologues (as *Ciona* has not undergone genome duplication)), myosin (*Mla/MlrV/Myl5*) and tropomyosin (*Tpm1/2/3*) (Extended Data Fig. 5d, Supplementary Table 2). None of these genes is expressed in the primary notochord²⁰. Moreover, the 5' regulatory regions of these genes contain clusters of *Tbx6* binding motifs (Supplementary Table 3), which suggests their direct regulation by muscle determinants. Gene reporter assays verified restricted expression of *Casq1/2* and *KH.C9.405* (Supplementary Table 2) in the secondary notochord and tail muscles (Extended Data Fig. 5e). It therefore appears that a muscle differentiation program is purposefully deployed in the secondary, but not the primary, notochord. These developmental programs suggest that *Ciona* possesses properties of both the notochords seen in cephalochordates and those of vertebrates.

Identification of individual neurons

The central nervous system of swimming tadpoles is composed of only 177 neurons⁸, which allows for the reconstruction of detailed transcriptome trajectories for individual neurons (Methods). We profiled 22,198 neural cells derived from the a-, b- and A-lineages (Extended Data Fig. 6a–c) across all 10 stages of development. This represents an average of about sevenfold coverage for every cell type (Supplementary Table 4). A total of 41 neural derivatives were identified in swimming tadpoles (Fig. 2, Extended Data Fig. 6d). These cells map to different regions of the nervous system, including the sensory vesicle, motor ganglion, nerve cord, peripheral sensory cells and associated interneurons. Distinctive combinations of regulatory genes were identified in the neural subtypes (Fig. 2, Extended Data Fig. 6e, Supplementary Table 2). For example, coronet cells are the only dopaminergic neurons in the *Ciona* central nervous system. Coronet cells express high levels of *Ptf1a* and *Meis*, which are sufficient to reprogram the central nervous system into supernumerary coronet cells²¹. It is possible that other combinations of cell-specific transcription factors specify additional neural subtypes (for example, *Bsh*, *Lhx2/9* and *Aristaless* in the anterior sensory vesicle).

The high coverage of individual transcriptomes enabled the identification of rare neuronal subtypes (Extended Data Fig. 7). For example, there are only two pairs of bipolar tail neurons in swimming tadpoles²², and these were found to express galanin and two of its receptors (*Galr1* and *Galr2*) (Supplementary Table 2). Galanin has previously been

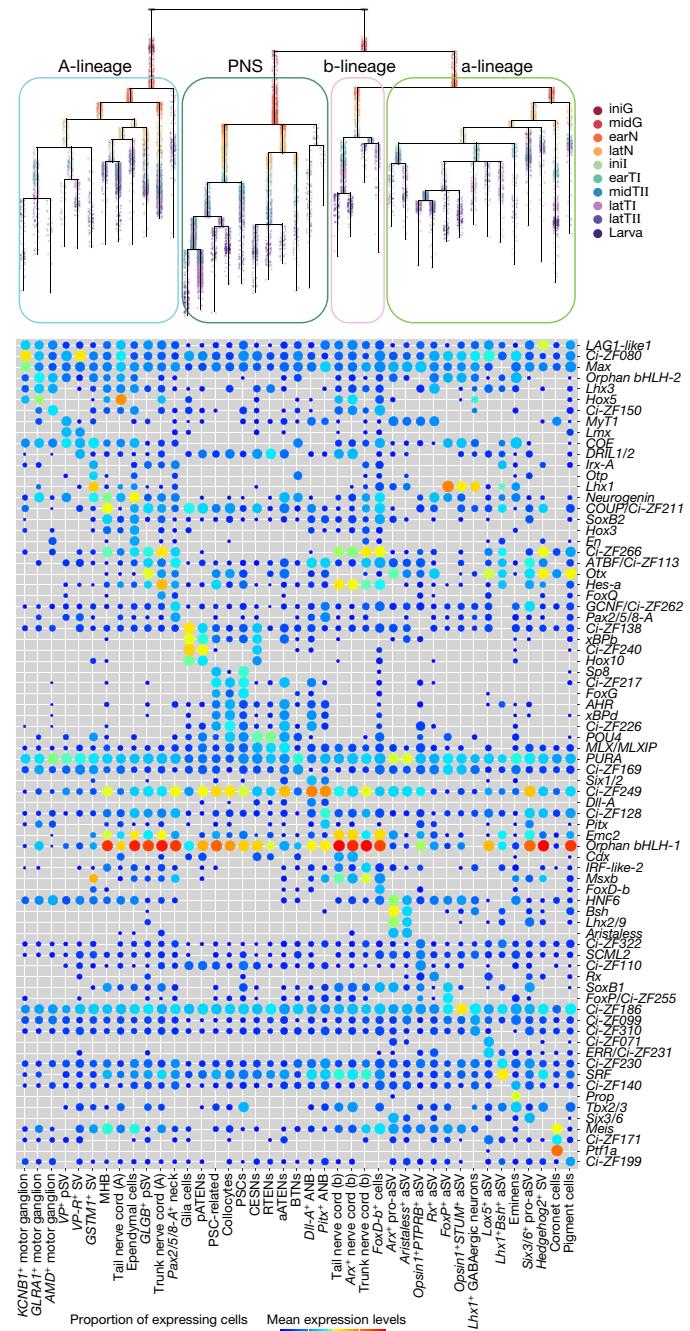


Fig. 2 | Transcriptome trajectories for defined individual neurons. Reconstructed expression lineage for the entire nervous system. Top, cells are coloured by developmental stage, and the a-lineage, b-lineage and A-lineage branches of the central nervous system and peripheral nervous system (PNS) are identified. Cells are ordered by pseudotime along each trajectory. Bottom, dot plot of the top three most-highly expressed regulatory genes in each neural subcluster at the larval stage. Dot size represents the percentage of cells that express the transcription factor, and the dot colour shows the averaged level of expression. aATENs, anterior apical trunk epidermal neurons; ANB, anterior neural boundary; aSV, anterior sensory vesicle; BTNs, bipolar tail neurons; CESNs, caudal epidermal sensory neurons; MHB, midbrain–hindbrain boundary; pATENs, posterior apical trunk epidermal neurons; RTENs, rostral trunk epidermal neurons; SV, sensory vesicle. Letters in parentheses denote lineages.

implicated in neuro-regeneration and axogenesis^{23,24}. A reporter gene that contains *Galr2* regulatory sequences mediates restricted expression in the bipolar tail neurons (Extended Data Fig. 7a). Similarly, a pair of decussating neurons—which have a central role in the startle response

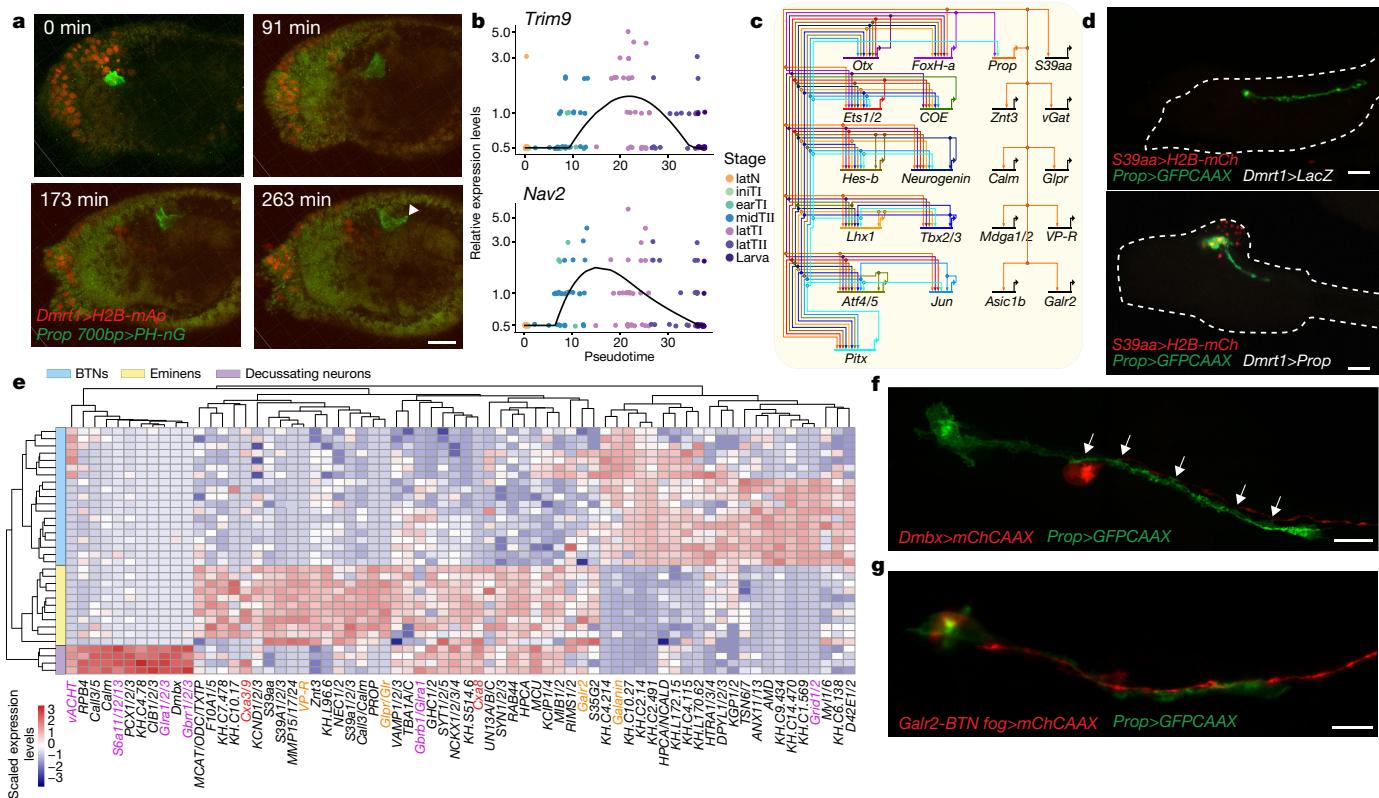


Fig. 3 | Integration of transcriptome maps and synaptome neuronal circuits. **a**, Snapshots of a time-lapse video of Eminens cell migration (from latTI to latTIII, $n = 2$ embryos). The embryo expressed H2B-mApple (H2B-mAp) under the regulatory sequences of *Dmrt1* (red), and PH-mNeonGreen (PH-nG) driven by *Prop* regulatory sequences (green). The Eminens cell migrates from the anterior side towards the posterior of the sensory vesicle before axogenesis (arrowhead). Numbers in top left indicate elapsed time in minutes. **b**, Pseudo-temporal expression of two migration-related genes from the early neurula to the larval stage. **c**, A provisional gene regulatory network of Eminens cells is shown, based on the regulatory cascade (Extended Data Fig. 9a). Only representative regulatory genes are shown (Methods). **d**, Upon overexpression of *Prop* in the anterior neural plate, supernumerary S39aa⁺ cells (H2B-mCherry (H2B-mCh), red) are observed (bottom, *Dmrt1>Prop*)

of vertebrates^{25,26}—was unambiguously identified on the basis of their selective expression of the homeobox gene *Dmbx*²⁷ (Extended Data Fig. 7b).

Additional specific neuronal subtypes were identified on the basis of restricted expression of select marker genes. Posterior sensory vesicle neurons that are positive for the vasopressin/oxytocin (VP) gene express several neuropeptides (Supplementary Table 5), including VP and an uncharacterized neuropeptide²⁸ (*NP*; see Supplementary Table 6). *NP* expression is restricted to a small group of neurons located in the posterior-most regions of the sensory vesicle (Extended Data Fig. 7c, e–f). A pair of Eminens neurons was identified by expression of a reporter gene containing *Prop* 5' regulatory sequences (Extended Data Fig. 7d). These studies document the feasibility of identifying the transcriptome trajectories and virtual lineages of individual defined neurons, including rare subtypes such as Eminens neurons.

Transcriptome and synaptome integration

The recently reported *Ciona* synaptome identified a single Eminens neuron (Em2) as a key regulator of decussating neurons^{8,25}. The pair of Eminens neurons was identified in our datasets on the basis of their expression of marker genes of GABAergic (γ -aminobutyric-acid-releasing) neurons and *Prop* (Extended Data Fig. 8a, b). Moreover, reporter genes that contain *Prop* regulatory sequences are selectively expressed in a pair of neurons that display all of the properties of

compared with control embryos (top, *Dmrt1>LacZ*). The embryos also expressed a membrane GFP as a *Prop* reporter gene (GFP-CAAX, green, $n = 3$ electroporation experiments for both conditions). **e**, Heat map of differentially expressed genes between Eminens neurons, decussating neurons and bipolar tail neurons. Genes involved in neurotransmission are coloured in magenta, connexin genes in red and neuropeptide-associated genes in orange. **f**, Em2 has multiple contacts (arrows) with the decussating neurons, as observed by double labelling with *Prop* and *Dmbx* reporter genes (green and red, respectively). $n = 3$ electroporation experiments. **g**, Double labelling with *Prop* and bipolar-tail-neuron-specific *Galr2* reporter genes (green and red, respectively) shows extensive contacts of bipolar tail neuron axon with Eminens neurons. $n = 2$ electroporation experiments. Scale bars, 20 μm (a, d), 10 μm (f, g).

Eminens neurons, including morphology and location^{25,29–31} (Extended Data Fig. 7d). Transcriptome trajectories of Eminens neurons suggest that they arise from the a-lineage (Fig. 2), even though they are located in the posterior regions of the sensory vesicle. This apparent discrepancy was resolved by live-cell imaging. We found that Eminens neurons undergo long-range migration from the forebrain to posterior regions of the sensory vesicle (Fig. 3a, Supplementary Video 1). These movements correlate with the expression of a variety of genes that are implicated in migration and axogenesis, including *Nav2* and *Trim9*^{32,33} (Fig. 3b).

Regulatory cascades of cell-signalling components and transcription factors enabled the formulation of a provisional gene regulatory network for the Eminens neurons (Fig. 3c, Extended Data Fig. 9a). The lynchpin of this network is *Prop*, a homeobox gene that appears to regulate a variety of genes that are involved in neuronal function; these include neuropeptide receptors (VP-R, *Glpr* and *Galr2*), zinc neuromodulation (*Znt3* and *S39aa*) and GABAergic markers (*vGat*) (Fig. 3c, Extended Data Fig. 8). Support for this network was obtained by manipulating a minimal *Prop* enhancer. Point mutations in the binding site for *FoxH-a*, one of the predicted upstream regulators of *Prop*, caused a significant (Fisher's exact test, $P = 1.27 \times 10^{-7}$) reduction in the expression of the minimal *Prop* reporter gene (Extended Data Fig. 9b, c). More importantly, overexpression of *Prop* in anterior regions of the sensory vesicle (via a *Dmrt1>Prop* fusion gene) resulted in the

formation of supernumerary Eminens neurons and ectopic activation of downstream reporter genes (for example, *S39aa*) (Fig. 3d, Extended Data Fig. 9d, e).

We next sought to leverage this information to gain insights into the neuronal interactions that underlie the startle response (Fig. 3e). A centrepiece of the startle circuit is the pair of decussating neurons, which correspond to the Mauthner neurons in the brain stem of fish and frogs²⁶. The decussating neurons integrate a variety of sensory information to trigger a fast escape reflex. As predicted by previous studies^{25,31}, interactions between Em2 and the decussating neurons (Fig. 3f) are probably inhibitory, as Eminens neurons express GABAergic markers such as *vGat* and *Gad* (Extended Data Fig. 8a, b) whereas the decussating neurons express GABA receptors (Fig. 3e, Supplementary Table 2). The decussating neurons also express glutamate receptors (Supplementary Table 2), which suggests that they respond to tonic glutamate signals.

The transcriptome datasets further raise the possibility that the startle circuit may be modulated by secreted neuropeptides (Supplementary Table 5). Both Eminens and decussating neurons express receptors for galanin, which is expressed in the bipolar tail neurons (Fig. 3e, g, Extended Data Fig. 8h). The bipolar tail neurons have previously been likened to the dorsal root ganglia derivatives of the neural crest in vertebrates²². Galanin promotes survival of dorsal root ganglia neurons during development and after injury. It is possible that galanin serves as a tropic factor for Em2, because the bipolar tail neurons directly interact with the cell body of this neuron (Fig. 3g). Moreover, modulation of Em2 by additional neuropeptides is suggested by the fact that Em2 expresses a VP receptor. As shown above, *VP⁺* cells express genes for a number of secreted neuropeptides—including *VP* and *NP* (Extended Data Fig. 7e). The *VP⁺* cells are in close proximity with Em2 (Extended Data Fig. 7f).

Our transcriptome datasets provide substantive annotations of the neuronal circuits that have been described by recent synaptome studies^{8,29}, suggest both targeted growth and feedback inhibition of the startle response by bipolar tail neurons, and implicate neuropeptides (such as galanin and vasopressin/oxytocin) as potential modulators of the circuit, in addition to canonical neurotransmitters.

Evolution of cell types

Previous studies suggest that *Ciona* possesses the rudiments of key vertebrate innovations, including the neural crest, cranial placodes and the cardio-pharyngeal mesoderm^{22,34–37}. However, the evolutionary origin of the telencephalon, which arises from the anterior-most regions of the forebrain, remains uncertain. The telencephalon contains the olfactory bulb and regions that control higher-order brain functions, such as the neocortex of humans. Forebrain regions of the *Ciona* central nervous system give rise to dopaminergic coronet cells and neuropore, but lack telencephalon derivatives such as the olfactory bulb.

To explore the origins of the telencephalon, we examined the gene-regulatory cascades for derivatives of the anterior-most regions of the neural plate, particularly palp sensory cells and the pro-anterior sensory vesicle (Extended Data Figs. 10–12, Methods). The palp sensory cells, also known as axial columnar cells³⁸, express a cascade of cell-signalling components and regulatory genes, including *FoxC*, *Dlx*, *FoxG*, *Isl* and *SP8* (Extended Data Fig. 12a, c, Supplementary Table 2). A similar regulatory cascade has previously been implicated in the specification of the telencephalon in vertebrates^{39,40}.

We also determined transcriptome trajectories for the pro-anterior sensory vesicle (the anterior-most regions of the neural tube), located adjacent to the proto-placodal territory that forms the palp sensory cells. The pro-anterior sensory vesicle first expresses anterior determinants (for example, *Otx*), followed by cell-specification genes such as *FoxJ1*, *Six1/2*, *Six3/6*, *Lhx2/9*, *Pitx* and *Otp* (Extended Data Fig. 12b, d, Supplementary Table 2). Many of these genes have also previously been implicated in the development of forebrain derivatives, including regions of the telencephalon^{41,42}.

We propose that the vertebrate telencephalon arose through the incorporation of non-neural ectoderm in anterior regions of the

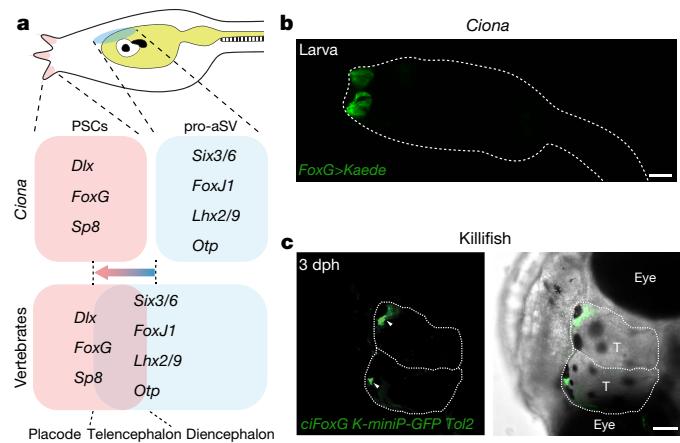


Fig. 4 | Model for the evolution of the telencephalon. **a**, Proposed model of the evolution of the vertebrate telencephalon. The telencephalon arose from the incorporation of anterior placodal gene-regulatory module into forebrain regions of the neural tube. Key regulatory components in *Ciona* palp sensory cells (including *Dlx*, *FoxG* and *Sp8*) and in the pro-anterior sensory vesicle (including *Six3/6*, *FoxJ1*, *Lhx2/9* and *Otp*) are conserved in the vertebrate telencephalon. **b**, The *FoxG* reporter gene (Kaede, green) exhibits restricted expression in palp sensory cells but not anterior regions of the sensory vesicle of *Ciona* larvae ($n = 2$ electroporation experiments). **c**, In killifish, GFP driven by the *Ciona FoxG* regulatory sequences and a zebrafish minimal promoter is expressed in a subset of cells in the olfactory bulb of the telencephalon (arrowheads, left). $n = 3$ independent transgenic lines (Methods). T, telencephalon. dph, days post-hatching. Scale bars, 20 μm (b), 250 μm (c).

neural tube (Fig. 4a). To test this model, we examined the expression of a *Ciona FoxG* reporter gene in *Ciona* larvae and transgenic killifish (*Nothobranchius furzeri*) embryos (Fig. 4b, c, Methods). This reporter is expressed in palp cells of *Ciona* embryos (Fig. 4b). It also mediates expression in subsets of cells in the olfactory bulb of the killifish telencephalon (Fig. 4c), as well as in placodal derivatives such as the lens of the eye (Extended Data Fig. 12e). These observations are consistent with the incorporation of proto-placodal gene-regulatory modules (for example, axial columnar cells) into an expanded forebrain of vertebrates.

In summary, we have presented comprehensive transcriptome trajectories, regulatory cascades and provisional gene networks for over 60 cell types (including nearly 40 neuronal subtypes) that comprise the *Ciona* tadpole. These datasets substantially extend classical lineage maps and regulatory blueprints, and provide a source of information for reconstructing the contributions of individual cells, lineages and tissues to critical morphogenetic processes, such as gastrulation, neurulation, notochord intercalation, tail elongation, compartmentalization of the gut and nervous system, and the formation of complex neuronal circuits that control behaviour. Our datasets also provide insights into the evolutionary transition between invertebrates and vertebrates, including the dual properties of the *Ciona* notochord and the expansion of the vertebrate forebrain. Current single-cell studies encompass a broad spectrum of cell types and systems^{1–5,43}, offering unprecedented opportunities to trace the evolutionary origins of every cell, tissue and organ in the human body.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests are available at <https://doi.org/10.1038/s41586-019-1385-y>.

Received: 27 August 2018; Accepted: 10 June 2019;

Published online 10 July 2019.

- Briggs, J. A. et al. The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. *Science* **360**, eaar5780 (2018).
- Farrell, J. A. et al. Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. *Science* **360**, eaar3131 (2018).

3. Wagner, D. E. et al. Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science* **360**, 981–987 (2018).
4. Pijuan-Sala, B. et al. A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature* **566**, 490–495 (2019).
5. Cao, J. et al. The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019).
6. Delsuc, F., Brinkmann, H., Chourrut, D. & Philippe, H. Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature* **439**, 965–968 (2006).
7. Imai, K. S., Levine, M., Satoh, N. & Satou, Y. Regulatory blueprint for a chordate embryo. *Science* **312**, 1183–1187 (2006).
8. Ryan, K., Lu, Z. & Meinertzhagen, I. A. The CNS connectome of a tadpole larva of *Ciona intestinalis* (L.) highlights sidedness in the brain of a chordate sibling. *eLife* **5**, e16962 (2016).
9. Prodöhl, F., Yamada, L., Shirae-Kurabayashi, M., Nakamura, Y. & Sasakura, Y. Postplasmic/PEM RNAs: a class of localized maternal mRNAs with multiple roles in cell polarity and development in ascidian embryos. *Dev. Dyn.* **236**, 1698–1715 (2007).
10. Corbo, J. C., Levine, M. & Zeller, R. W. Characterization of a notochord-specific enhancer from the *Brachyury* promoter region of the ascidian, *Ciona intestinalis*. *Development* **124**, 589–602 (1997).
11. Tokuoka, M., Imai, K. S., Satou, Y. & Satoh, N. Three distinct lineages of mesenchymal cells in *Ciona intestinalis* embryos demonstrated by specific gene expression. *Dev. Biol.* **274**, 211–224 (2004).
12. Nishida, H. Cell lineage analysis in ascidian embryos by intracellular injection of a tracer enzyme. III. Up to the tissue restricted stage. *Dev. Biol.* **121**, 526–541 (1987).
13. Nakazawa, K. et al. Formation of the digestive tract in *Ciona intestinalis* includes two distinct morphogenic processes between its anterior and posterior parts. *Dev. Dyn.* **242**, 1172–1183 (2013).
14. Veeman, M. T., Newman-Smith, E., El-Nachef, D. & Smith, W. C. The ascidian mouth opening is derived from the anterior neuropore: reassessing the mouth/neural tube relationship in chordate evolution. *Dev. Biol.* **344**, 138–149 (2010).
15. Stemple, D. L. Structure and function of the notochord: an essential organ for chordate development. *Development* **132**, 2503–2512 (2005).
16. Suzuki, M. M. & Satoh, N. Genes expressed in the amphioxus notochord revealed by EST analysis. *Dev. Biol.* **224**, 168–177 (2000).
17. Yagi, K., Satou, Y. & Satoh, N. A zinc finger transcription factor, ZicL, is a direct activator of *Brachyury* in the notochord specification of *Ciona intestinalis*. *Development* **131**, 1279–1288 (2004).
18. Hudson, C. & Yasuo, H. A signalling relay involving Nodal and Delta ligands acts during secondary notochord induction in *Ciona* embryos. *Development* **133**, 2855–2864 (2006).
19. Yagi, K., Takatori, N., Satou, Y. & Satoh, N. *Ci-Tbx6b* and *Ci-Tbx6c* are key mediators of the maternal effect gene *Ci-macho1* in muscle cell differentiation in *Ciona intestinalis* embryos. *Dev. Biol.* **282**, 535–549 (2005).
20. Takahashi, H. et al. *Brachyury* downstream notochord differentiation in the ascidian embryo. *Genes Dev.* **13**, 1519–1523 (1999).
21. Horie, T. et al. Regulatory cocktail for dopaminergic neurons in a protovertebrate identified by whole-embryo single-cell transcriptomics. *Genes Dev.* **32**, 1297–1302 (2018).
22. Stolfi, A., Ryan, K., Meinertzhagen, I. A. & Christiaen, L. Migratory neuronal progenitors arise from the neural plate borders in tunicates. *Nature* **527**, 371–374 (2015).
23. Shi, T. J. et al. Sensory neuronal phenotype in galanin receptor 2 knockout mice: focus on dorsal root ganglion neurone development and pain behaviour. *Eur. J. Neurosci.* **23**, 627–636 (2006).
24. Holmes, F. E. et al. Targeted disruption of the galanin gene reduces the number of sensory neurons and their regenerative capacity. *Proc. Natl. Acad. Sci. USA* **97**, 11563–11568 (2000).
25. Ryan, K., Lu, Z. & Meinertzhagen, I. A. Circuit homology between decussating pathways in the *Ciona* larval CNS and the vertebrate startle-response pathway. *Curr. Biol.* **27**, 721–728 (2017).
26. Korn, H. & Faber, D. S. The Mauthner cell half a century later: a neurobiological model for decision-making? *Neuron* **47**, 13–28 (2005).
27. Stolfi, A. & Levine, M. Neuronal subtype specification in the spinal cord of a prototypical vertebrate. *Development* **138**, 995–1004 (2011).
28. Hamada, M. et al. Expression of neuropeptide- and hormone-encoding genes in the *Ciona intestinalis* larval brain. *Dev. Biol.* **352**, 202–214 (2011).
29. Ryan, K., Lu, Z. & Meinertzhagen, I. A. The peripheral nervous system of the ascidian tadpole larva: types of neurons and their synaptic networks. *J. Comp. Neurol.* **526**, 583–608 (2018).
30. Imai, J. H. & Meinertzhagen, I. A. Neurons of the ascidian larval nervous system in *Ciona intestinalis*: I. Central nervous system. *J. Comp. Neurol.* **501**, 316–334 (2007).
31. Takamura, K., Minamida, N. & Okabe, S. Neural map of the larval central nervous system in the ascidian *Ciona intestinalis*. *Zool. Sci.* **27**, 191–203 (2010).
32. Hekimi, S. & Kershaw, D. Axonal guidance defects in a *Caenorhabditis elegans* mutant reveal cell-extrinsic determinants of neuronal morphology. *J. Neurosci.* **13**, 4254–4271 (1993).
33. Winkle, C. C. et al. *Trim9* deletion alters the morphogenesis of developing and adult-born hippocampal neurons and impairs spatial learning and memory. *J. Neurosci.* **36**, 4940–4958 (2016).
34. Abitua, P. B. et al. The pre-vertebrate origins of neurogenic placodes. *Nature* **524**, 462–465 (2015).
35. Abitua, P. B., Wagner, E., Navarrete, I. A. & Levine, M. Identification of a rudimentary neural crest in a non-vertebrate chordate. *Nature* **492**, 104–107 (2012).
36. Stolfi, A. et al. Early chordate origins of the vertebrate second heart field. *Science* **329**, 565–568 (2010).
37. Horie, R. et al. Shared evolutionary origin of vertebrate neural crest and cranial placodes. *Nature* **560**, 228–232 (2018).
38. Zeng, F. et al. Papillae revisited and the nature of the adhesive secreting cell types. *Dev. Biol.* **448**, 183–198 (2019).
39. Hébert, J. M. & Fishell, G. The genetics of early telencephalon patterning: some assembly required. *Nat. Rev. Neurosci.* **9**, 678–685 (2008).
40. Zembrzuski, A., Griesel, G., Stoykova, A. & Mansouri, A. Genetic interplay between the transcription factors Sp8 and Emx2 in the patterning of the forebrain. *Neural Dev.* **2**, 8 (2007).
41. Jacquet, B. V. et al. Specification of a Foxj1-dependent lineage in the forebrain is required for embryonic-to-postnatal transition of neurogenesis in the olfactory bulb. *J. Neurosci.* **31**, 9368–9382 (2011).
42. Carlin, D. et al. Six3 cooperates with Hedgehog signaling to specify ventral telencephalon by promoting early expression of Foxg1a and repressing Wnt signaling. *Development* **139**, 2614–2624 (2012).
43. Zhong, S. et al. A single-cell RNA-seq survey of the developmental landscape of the human prefrontal cortex. *Nature* **555**, 524–528 (2018).

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

© The Author(s), under exclusive licence to Springer Nature Limited 2019

METHODS

Ciona handling, collection and dissociation of embryos. The adults of *C. intestinalis* were purchased from M-Rep. The eggs and sperm were obtained as previously described⁴⁴. Sperm was added to the eggs for 10 min. Then, the fertilized eggs were washed twice with filtered sea water. Except in the case of the larval stage, the same animal provided both the eggs and the sperm to lower the polymorphism rate for downstream analysis. Embryos were raised to different stages at 18 °C according to a previously described method⁴⁵. At least two biological replicates from each developmental stage were collected (Supplementary Table 1). For each sample, 100 to 500 morphologically normal embryos were randomly picked and transferred into tubes pre-coated with 5% BSA in Ca²⁺-free artificial sea water (Ca²⁺-free ASW, 10 mM KCl, 40 mM MgCl₂, 15 mM MgSO₄, 435 mM NaCl, 2.5 mM NaHCO₃, 7 mM Tris base, 13 mM Tris-HCl). Embryos were immediately dissociated with 0.5 to 1% trypsin in Ca²⁺-free ASW with 5 mM EGTA (ASW EGTA) for 2 min (gastrula and neurula stages) to 10 min (tailbud stages). Embryos were pipetted for 5 min on ice to complete the dissociation of individual cells. Then, the digestion was inhibited with 0.2% BSA in Ca²⁺-free ASW or quenched by 20% FBS. Cells were collected by centrifugation at 4 °C at 500g for 2 to 5 min and then resuspended in ice-cold Ca²⁺-free ASW containing 0.5% BSA. For the swimming larval stage, the embryos were either homogenized (H100, Waverly) and dissociated using 1% trypsin or dissociated with 1% trypsin, 1 mg/ml collagenase, 0.5% pronase and 0.5 mg/ml cellulase in ASW EGTA. Once dissociated, the enzymes were inhibited by 20% FBS and 2 mg/ml glycine. The cells were washed and resuspended as described above. **Single-cell barcoding, on-chip and off-chip technical replicates, library preparation and sequencing.** The cell concentration of each sample was checked by TC20 Automated Cell Counter to ensure it was within 1,000–2,000 cells per micro-litre. Single-cell suspensions were loaded onto The Chromium Controller (10x Genomics). To assess technical variations between replicates, on-chip and off-chip experiments were performed. The on-chip experiment consisted of loading two lanes of cells from latTII embryos on the same chip, obtained by fertilizing eggs with sperm from the same animal. In the off-chip experiment, dissociated cells from latTI embryos obtained with the same fertilization strategy were loaded on the same lane on two different chips and processed separately. For all of the samples, cells were lysed, cDNAs were barcoded and amplified with Chromium Single Cell 3' Library and Gel Bead Kit v2 (10x Genomics) following the instructions of the manufacturer. Illumina sequencing libraries were prepared from the cDNA samples using the Nextera DNA library prep kit (Illumina). All of the libraries were sequenced on Illumina HiSeq 2500 Rapid flowcells (Illumina) with paired-end 26 nucleotides (nt) + 125 nt reads following standard Illumina protocols.

Raw sequencing reads were filtered by Illumina HiSeq Control Software and only pass-filtered reads were used for further analysis. Samples were run on both lanes of a HiSeq 2500 Rapid Run mode flow cell instrument. Base calling was performed by Illumina RTA version 1.18.64.0. BCL files were then converted to FASTQ format using bcl2fastq version 1.8.4 (Illumina). Reads that aligned to phix (using Bowtie version 1.1.1) were removed, as were reads that failed Illumina's default chastity filter. We then combined the FASTQ files from each lane and separated the samples using the barcode sequences allowing one mismatch (using barcode_splitter version 0.18.2). Using 10x CellRanger version 2.0.1, the count pipeline was run with default settings on the FASTQ files to generate gene-barcode matrices for each sample. The reference sequence was obtained from the Ghost database⁴⁶.

Data quality control and visualization. To remove signals from putative empty droplet or degraded RNA, low-quality transcriptomes were filtered for each time course sample as follows: (1) we discarded cells with less than 1,000 expressed genes; (2) cells with unique molecular identifiers (UMIs) of five s.d. above the mean were not included in our analyses (Supplementary Table 1); (3) we considered only genes that were expressed in at least three cells in each dataset. In total, 90,579 cells were kept for subsequent analysis. We further normalized the read counts of each cell by Seurat methods⁴⁷, and the normalized read counts were log-transformed for downstream analyses and visualizations. For dimensional reduction, the relative expression measurement of each gene was used to remove unwanted variation. Genes with the top 1,000 highest standard deviations were obtained as highly variable genes. After significant (jackstraw procedure) principal components were identified, a graph-based clustering approach was used for partitioning the cellular distance matrix into clusters. Cell distance was visualized by t-SNE method in reduced two-dimensional space.

For t-SNE visualization of the whole dataset (as shown in Fig. 1b, Extended Data Fig. 2a, b), the shared highly variable genes (500 genes) among all of the samples were kept for canonical correlation analysis. The top 50 canonical correlation vectors were calculated, and each dataset was projected into the maximally correlated subspaces. According to the relationship between the number of canonical correlation vectors and the percentage of variance explained, 1–18 canonical correlation vectors were used for subspace alignment⁴⁷ and dimensional reduction. Graph-based clustering was performed in the lower-dimensional space,

and an approximate nearest neighbour search was performed. We used ten random starts for clustering and selected the result with highest modularity. A modified fast Fourier transform-accelerated interpolation-based t-SNE⁴⁸ was used in the visualization of our dataset. Maximum iteration times were set to 2,000 and the perplexity parameter was set to 30.

For t-SNE visualization of cells in each tissue type from different developmental stages, we regressed out the effects produced by the UMI counts, experiment batch and sample identities with negative binomial regression modelling before dimensional reduction and clustering. Similar to approaches mentioned above, after significant (jackstraw procedure) principal components were identified (1–20 principal components were used for the subclustering of cells in each tissue type), graph-based clustering was performed and a modified fast Fourier transform-accelerated interpolation-based t-SNE was used in the visualization.

Annotation of cell clusters. We annotated the clusters for assigning clustering results to tissue types. For the clustering applied to t-SNE coordinates of the whole dataset, three steps were followed to refine the annotation results. (1) The expression pattern of top marker genes and regulatory genes were compared between clusters. Clusters with similar expression pattern of key regulatory genes and known markers were considered to be the same tissue type. (2) We carefully compared the annotation results with the in situ images recorded in the Ghost (<http://ghost.zool.kyoto-u.ac.jp>) and Aniseed (https://www.aniseed.cnrs.fr/aniseed/experiment/find_in situ) databases, or published papers, to validate the annotation results. (3) For putative newly discovered cell types in clusters with poorly annotated marker genes, we carefully checked the gene-expression pattern to make sure there was no ambiguous expression of known markers, which might indicate cell doublets. We also consulted with experts in ascidian research to validate our findings. The newly identified cell types were named according to their specifically expressed genes—for example, *Tll1*⁺ cells in the mesenchyme. If the position and morphology of the cells were verified by our reporter assays, we compared them to published papers with morphological information (for example, the previously published synaptome⁸) and identified the cell types (for example, Eminens neurons).

Cell-state mapping across time points. Most of the major tissues of the ascidian tadpole are first specified before gastrulation, which is our starting time point for sampling. To capture the developmental transitions that stem from different blastomeres at the 110-cell stage, we performed 'ancestor voting' between clusters across time, as previously described¹. In brief, between every two adjacent time points, all of the cells were embedded into the principal component analysis space of the later time point (1–50 principal components were used). For each cluster identified in the later time point, each cell in the cluster and its nearest five neighbouring cells in the previous time point were calculated based on the similarity between transcriptomes. The voting results for all of the cells in each cluster of the later time point were aggregated, and the percentage of 'ancestor cells' in each cluster of the earlier time point was calculated. In cases in which a cluster had more than one ancestor cluster, the cluster in the earlier time point was considered to be the winning ancestor if it had a percentage number that was at least two times that of the other clusters.

To better capture all of the subpopulations for highly differentiated tissue types, we subclustered the cells from each of the stages and performed the ancestor voting process between subclusters across time points. For mesenchyme cells, most of the winning ancestors were unambiguously assigned, with a >90% winning share on average. For cells from the epidermis and nervous system systems (as some sensory neurons differentiate from the epidermis), we subclustered the epidermis and nervous system cells together. We deleted the subclusters if they were assigned to multiple ancestor clusters and there was no winning ancestor (no winning cluster met our criteria (that is, a percentage number that was at least twice of the other clusters)) to make sure the intermediate state or immature neuron types did not affect the cell-state mapping results.

Single-cell trajectory construction. For notochord and Eminens neurons, we used monocle 2⁴⁹ to construct the single-cell trajectory for each cleanly defined lineage with known markers. Highly variable genes with $q < 0.01$ were selected across time points, a discriminative dimensionality reduction (DDRTree) was performed with regression on the UMI counts to eliminate unwanted variation introduced by sequencing depth between samples, and cells were ordered along the trajectory according to their pseudotime value. A subset of significant genes with $q < 1 \times 10^{-100}$ and $q < 1 \times 10^{-20}$ are shown in the pseudotemporal expression pattern of primary and secondary notochord, respectively, in Extended Data Fig. 5.

For tissues that contained more complexity during development, such as the mesenchyme and nervous system (which had 9 and 41 identified subclusters at the larval stage, respectively), we used a simulated diffusion-based computational reconstruction method (URD²) for acquiring the transcriptional trajectories during embryogenesis. In brief, after differentially expressed genes were picked and dimensional reduction was performed for cells in specific tissues from ten developmental stages as described in 'Data quality control and visualization', we calculated

transition probabilities between cells with the destiny package⁵⁰. We next assigned to the cells a pseudotime value with a probabilistic breadth-first graph search using the transition probabilities. To find the developmental trajectories, we performed biased random walks that started from a random cell in each refined cluster of the final stage (that is, the larval stage) that we covered. The walk was simulated through cells on the basis of the transition probabilities, and the transitions were only allowed for cells with younger or similar pseudotimes to make sure the trajectory between the root (cells from the earliest time point) and the tip (cells from the last time point) was found. Then, the biased random walk was processed into visitation frequencies. The URD tree structure was built by aggregating trajectories when the same cells were visited from each tip.

For cells in the mesenchyme, we optimized the number of nearest neighbours (k -nearest neighbour) and set it to 250, and the width of the Gaussian used to transform cell–cell distances into transition probabilities (sigma) was set to six. We also modified parameters for constructing the URD tree as follows: divergence.method = “preference”, cells.per.pseudotime.bin = 75, bins.per.pseudotime.window = 10, p.thresh = 0.01. To avoid ambiguities in reconstructing gene-expression lineages for cells in the nervous system, we excluded or combined those cell clusters that (1) were not well-defined or determined during neurogenesis on the basis of prior knowledge; (2) could not be resolved by diffusion components (such as very small population of cells; for example, decussating neurons); and (3) exhibited intermixing in the diffusion maps. The parameters were set as follows: divergence.method = “preference”, cells.per.pseudotime.bin = 28, bins.per.pseudotime.window = 4, p.thresh = 0.025, minimum.visits = 40. For endodermal cells, the parameters were set as follows: divergence.method = “preference”, cells.per.pseudotime.bin = 65, bins.per.pseudotime.window = 10, p.thresh = 0.01, minimum.visits = 20. For muscle cells, the parameters were set as follows: divergence.method = “preference”, cells.per.pseudotime.bin = 20, bins.per.pseudotime.window = 10, p.thresh = 0.01, minimum.visits = 20.

Gene-expression cascade. The genes included in the cascade of each trajectory were recovered following the criteria set in the URD package: cells in the segment were compared in a pairwise manner with cells from each of that segment’s siblings and children, and differentially expressed genes were kept if they were expressed in more than 10% of the population, their mean expression level was $1.5 \times$ higher than in the sibling branch, and the genes were $1.25 \times$ better classifiers than a random classifier for the population. Then, an impulse model was fitted to the expression of each gene recovered in the cascade for determining the ‘on and off’ timing of expression, and the genes were ordered by the ‘on-time’ in the cascade. Genes with an expression pattern that was not fitted with the impulse model were arranged at the bottom of the cascade. In the heat map, cells were ordered with the progression of pseudotime using a moving window, and the scaled mean expression within each pseudotime moving-window was plotted.

Regulatory network. Regulatory genes, signalling pathway genes recovered in each developmental trajectory cascade and selected highly expressed genes for specific cell types at the final time point that had a fold change (expressed in \log_2) above one between groups were all used in investigating the putative direct interaction. We used cluster-buster⁵¹ to find clusters of pre-specified motifs 2 kb upstream of the transcription start site of each gene. The parameters were set as follows: $g = 1$, $m = 0$, $c = 0$, score ≥ 6 . The position frequency matrix was downloaded from the JASPAR 2018 database⁵². Genes with no position frequency matrix recorded in JASPAR was not considered in constructing the regulatory network. The regulatory network was plotted with BiotaPestry. Each line between every two genes represents a putative direct interaction, as the binding motif of the regulatory gene was identified in the motif-cluster region of the target gene.

Heat maps. Heat maps in Extended Data Fig. 3 were plotted with the DoHeatmap function of Seurat v.2.3.2. Only genes with an average fold change (expressed logarithmically) > 0.3 are shown. For Extended Data Fig. 5d, differentially expressed genes between primary notochord and secondary notochord were identified by the following criteria using DESeq2⁵³: (1) FDR (false discovery rate) adjusted P value below 0.05; and (2) absolute fold change (expressed in \log_2) between groups was larger than 1.5. The mean expression level of each gene within one developmental stage was calculated, and the scaled expression of the genes was on the basis of the Euclidean distance using pheatmap 1.0.10. For Fig. 3e, genes with an average fold change (expressed logarithmically) > 1.5 are shown. Both Fig. 3e and Extended Data Fig. 5d were plotted with pheatmap. The pseudotemporal expression heat maps in Extended Data Fig. 5b, c and Extended Data Fig. 9a, and the expression dynamics in Fig. 3b, were plotted using monocle 2.

Molecular cloning. The KH number of all of the genes mentioned in the manuscript as well as other names that are commonly used in the *Ciona* field can be found in the Supplementary Table 6.

Dmbx, *Dmrt1*, *Gad*, *Prop*, *Twist* and *vGat* regulatory sequences have previously been described^{27,31,34,54,55}. They were cloned in *pCESA* expression vector upstream of the reporter genes *GFP-CAAX* (CAAX is the palmitoylation motif to target a protein to the membrane), *H2B-mApple*, *H2B-YFP*, *mNeonGreen-PH* (*nG-PH*),

mCherry-CAAX and *H2B-mCherry* using NotI and AscI restriction enzymes (NEB). The expression vector with *H2B-mApple* reporter construct was obtained by inserting *mApple*⁵⁶ (primers in Supplementary Table 7) into the *pCESA* expression vector that contains *H2B*, using NEBuilder (NEB). The expression vector that contains the *nG-PH* reporter gene was obtained by first inserting *GFP-PH* (courtesy of T. Meyer)⁵⁷ using NotI and FseI (NEB) into a *pCESA* expression vector and then replacing the *GFP* coding sequence with *mNeonGreen*⁵⁸ by recombination using NEBuilder (primers in Supplementary Table 7).

Asic1b, *Calm*, *Fgf13*, *Galr2*, *S39aa*, *S39aa* 2.2 kb and *Znt3* regulatory sequences were PCR-amplified (primers in Supplementary Table 7) from genomic DNA and cloned into *pCESA-H2B:mCherry* using AscI and NotI restriction enzymes.

After PCR amplification (primers in Supplementary Table 7) *Casq1/2* regulatory sequences were cloned into an expression vector that contains *GFP* downstream of the minimal promoter of *fog* (*pCESA-fog>GFP*) using AscI and XbaI restriction enzymes (NEB). The regulatory sequences of *NP* (*KH.C11.631*) were PCR-amplified and cloned into *pCESA-fog>GFP-CAAX*.

After PCR amplification from the *Prop>GFP-CAAX* (primers in Supplementary Table 7), *Prop* 900 bp, *Prop* 700 bp and *Prop* 300 bp were cloned into *pCESA-GFP-CAAX* vector using AscI and NotI.

For live imaging, *Prop* 700 bp was also cloned upstream of *PH-nG*. The reporter gene was obtained by NEBuilder assembly. First the *PH* domain, *GFP* and the degradation signal of *Hes-b* (*PH-GFP* primers), which was obtained from *Ciona* cDNA, were assembled into a *pCESA* expression vector using NEB builder. Then, *GFP* and the degradation signal coding sequences were replaced by *mNeonGreen* and a shorter degradation sequence using NEBuilder assembly (*PH-nG* primers). Finally a second degradation signal was inserted before the stop codon using NEBuilder assembly (deg primers, primers in Supplementary Table 7).

Prop 260 bp was cloned into *pCESA-fog GFP-CAAX* vector using AscI and XbaI. Point mutations in FoxH-a binding site of the *Prop* 260 bp regulatory sequences were obtained by plasmid PCR of *Prop* 260 bp *fog>GFP-CAAX* (primers in Supplementary Table 7).

Galr2 regulatory sequences specifically active in bipolar tail neurons were amplified from *Galr2>H2B-mCherry* (primers in Supplementary Table 7) and cloned into *pCESA-fog>mCherry-CAAX* using AscI and XbaI restriction sites.

Tll1, *Hlx* and *FoxG* regulatory sequences were PCR-amplified (primers in Supplementary Table 7) and then assembled into *pSP-Kaede* expression vector using NEBuilder. *Ptf1a* was obtained by PCR-amplifying an expression vector that contains the full-length *Ptf1a* regulatory sequences²¹ (primers in Supplementary Table 7). The PCR product was self-recombined using NEBuilder. *Ptf1a* was then subcloned upstream of *mCherry-CAAX* in the *pCESA* expression vector.

A *LacZ* expression vector under the control of *Dmrt1* (*Dmrt1>LacZ*) has previously been described⁵⁴. The *Prop* coding sequence was amplified from mid-tailbud embryo cDNA and cloned downstream of *Dmrt1* regulatory sequences using NotI and FseI restriction enzymes (NEB).

Ciona electroporation and imaging. After fertilization, one-cell-stage embryos were electroporated using 20 to 100 μ g of each expression construct as previously described¹⁰.

The embryos were raised at 16 °C, 18 °C or 21 °C in ASW and fixed at the desired stage following a previously described protocol⁵⁴. The embryos were washed several times with 0.05% BSA in PBS before being mounted using FluorSave Reagent (Millipore). Images were acquired with a Zeiss 880 confocal microscope with or without the Airyscan module, and a wide-field Zeiss Axio Observer Z1/7 combined to the Apotome 2.0 module.

All electroporation was performed in duplicate or triplicate. Between 18 and 610 embryos were recovered per condition. No specific randomization strategy was performed, except for the assignment of the fertilized eggs to the different conditions.

Live imaging was performed using a two-photon microscope system built in-house. Embryos were anaesthetized with 16 mg/ml MS-222 in ASW (Sigma-Aldrich). They were placed in microwells cast in 1% agarose in ASW⁵⁹, and the imaging was performed at 18 °C from the latTI to latTIII stage. The images were assembled using Fiji⁶⁰ and the final rendering obtained with Imaris (Bitplane).

Statistical analysis of the functional assays. For the statistical tests, the embryos with the same electroporated plasmids were pooled over the different experiments. Mann–Whitney *U*-test was performed with the package Tidyverse of R software, the χ^2 test followed by the post hoc test for pairwise comparison, Fisher’s test with Bonferroni adjustment was also performed with Tidyverse⁶¹.

Fish husbandry, generation of transgenic fish and imaging. All experiments with the African killifish *N. furzeri* were performed using the GRZ strain. All of the fish were housed at 27 °C in a facility overseen by the Stowers Institute for Medical Research (SIMR) Institutional Animal Care and Use Committee. Work with fish was performed according to the guidelines of the Stowers Institute for Medical Research.

A 4-kb *Ciona FoxG* regulatory sequence was cloned into *pDest-Tol2-miniP-GFP-Cryaa-Venus* transgenic vector through Gibson assembly. To generate the

transgenic killifish, 15–20 pg DNA was co-injected with 30 pg transposase mRNA into one-cell-stage *N. furzeri* embryos and the injected embryos were maintained in Yamamoto embryo solution (17 mM NaCl, 2.7 mM KCl, 2.5 mM CaCl₂, 0.02 mM NaHCO₃, pH 7.3) at 28 °C for 2 weeks before hatching. F₀ founders were crossed with wild-type GRZ fish and three independent lines were established for gene expression studies. No genotyping was performed to detect the transgene. However, for the first transgenic line, 15 out of 46 F₁ embryos showed GFP expression in the forebrain. For the second transgenic line, 8 out of 25 F₁ embryos showed GFP expression in the forebrain. Finally, 10 out of 37 F₁ embryos of the third transgenic line had GFP expression in the forebrain. No particular randomization strategy was implemented.

Killifish embryos were removed manually from the chorion before imaging. The juvenile fish were anaesthetized in 150 mg/l MS-222 for 5 min at room temperature. Images were taken with Ultraview R2 spinning disk confocal microscope. **Estimation of sample size, blinding and randomization.** No statistical methods were used to predetermine sample size. For the single-cell experiments, because the embryo collection and the subsequent data analysis was performed by different researchers, the investigators were blinded to group allocation. For the functional assays, no particular blinding strategy was adopted. As stated in the specific sections above, the assignment of the *Ciona* embryos to the different conditions were randomized. Otherwise, no particular randomization strategy was used.

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

Data availability

Raw sequencing data and the gene-expression matrix are available in the Gene Expression Omnibus (GEO) under accession number GSE131155. Our data can be explored at https://portals.broadinstitute.org/single_cell/study/SCP454/comprehensive-single-cell-transcriptome-lineages-of-a-proto-vertebrate. All other data are available from the corresponding authors on reasonable request.

44. Christiaen, L., Wagner, E., Shi, W. & Levine, M. Isolation of sea squirt (*Ciona*) gametes, fertilization, dechorionation, and development. *Cold Spring Harb. Protoc.* **2009**, pdb.prot5344, (2009).
45. Hotta, K. et al. A web-based interactive developmental table for the ascidian *Ciona intestinalis*, including 3D real-image embryo reconstructions: I. From fertilized egg to hatching larva. *Dev. Dyn.* **236**, 1790–1805 (2007).
46. Satou, Y., Kawashima, T., Shoguchi, E., Nakayama, A. & Satoh, N. An integrated database of the ascidian, *Ciona intestinalis*: towards functional genomics. *Zool. Sci.* **22**, 837–843 (2005).
47. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
48. Linderman, G. C., Rachh, M., Hoskins, J. G., Steinerberger, S. & Kluger, Y. Fast interpolation-based t-SNE for improved visualization of single-cell RNA-seq data. *Nat. Methods* **16**, 243–245 (2019).
49. Qiu, X. et al. Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* **14**, 979–982 (2017).
50. Haghverdi, L., Buettner, F. & Theis, F. J. Diffusion maps for high-dimensional single-cell analysis of differentiation data. *Bioinformatics* **31**, 2989–2998 (2015).
51. Frith, M. C., Li, M. C. & Weng, Z. Cluster-Buster: finding dense clusters of motifs in DNA sequences. *Nucleic Acids Res.* **31**, 3666–3668 (2003).
52. Khan, A. et al. JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res.* **46**, D1284 (2018).
53. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
54. Wagner, E. & Levine, M. FGF signaling establishes the anterior border of the *Ciona* neural tube. *Development* **139**, 2351–2359 (2012).
55. Yoshida, R. et al. Identification of neuron-specific promoters in *Ciona intestinalis*. *Genesis* **39**, 130–140 (2004).
56. Shaner, N. C. et al. Improving the photostability of bright monomeric orange and red fluorescent proteins. *Nat. Methods* **5**, 545–551 (2008).
57. Stauffer, T. P., Ahn, S. & Meyer, T. Receptor-induced transient reduction in plasma membrane PtIns(4,5)P2 concentration monitored in living cells. *Curr. Biol.* **8**, 343–346 (1998).
58. Shaner, N. C. et al. A bright monomeric green fluorescent protein derived from *Branchiostoma lanceolatum*. *Nat. Methods* **10**, 407–409 (2013).
59. Gregory, C. & Veeman, M. 3D-printed microwell arrays for *Ciona* microinjection and timelapse imaging. *PLoS ONE* **8**, e82307 (2013).
60. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
61. R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2013).

Acknowledgements We thank J. B. Wiggins, J. M. Miller and the Genomics Core Facility for technical support of 10x Chromium platform; IT and Bioinformatics staff at the Lewis-Sigler Institute for Integrative Genomics (LSI) for development of the sequence alignment pipeline; E. G. Gatzogiannis (director of the LSI Imaging Core Facility) for building the two-photon microscope and help with imaging; the Shvartsman laboratory in LSI for Imaris access; A. Sánchez Alvarado at SIMR for providing laboratory support and resources; and members of the Levine Laboratory for helpful discussions, and especially N. Treen for suggesting mNeonGreen as fluorescent reporter. This study was supported by a grant from the NIH (NS076542) to M.L. W.W. (SIMR) was funded by the Stowers Institute.

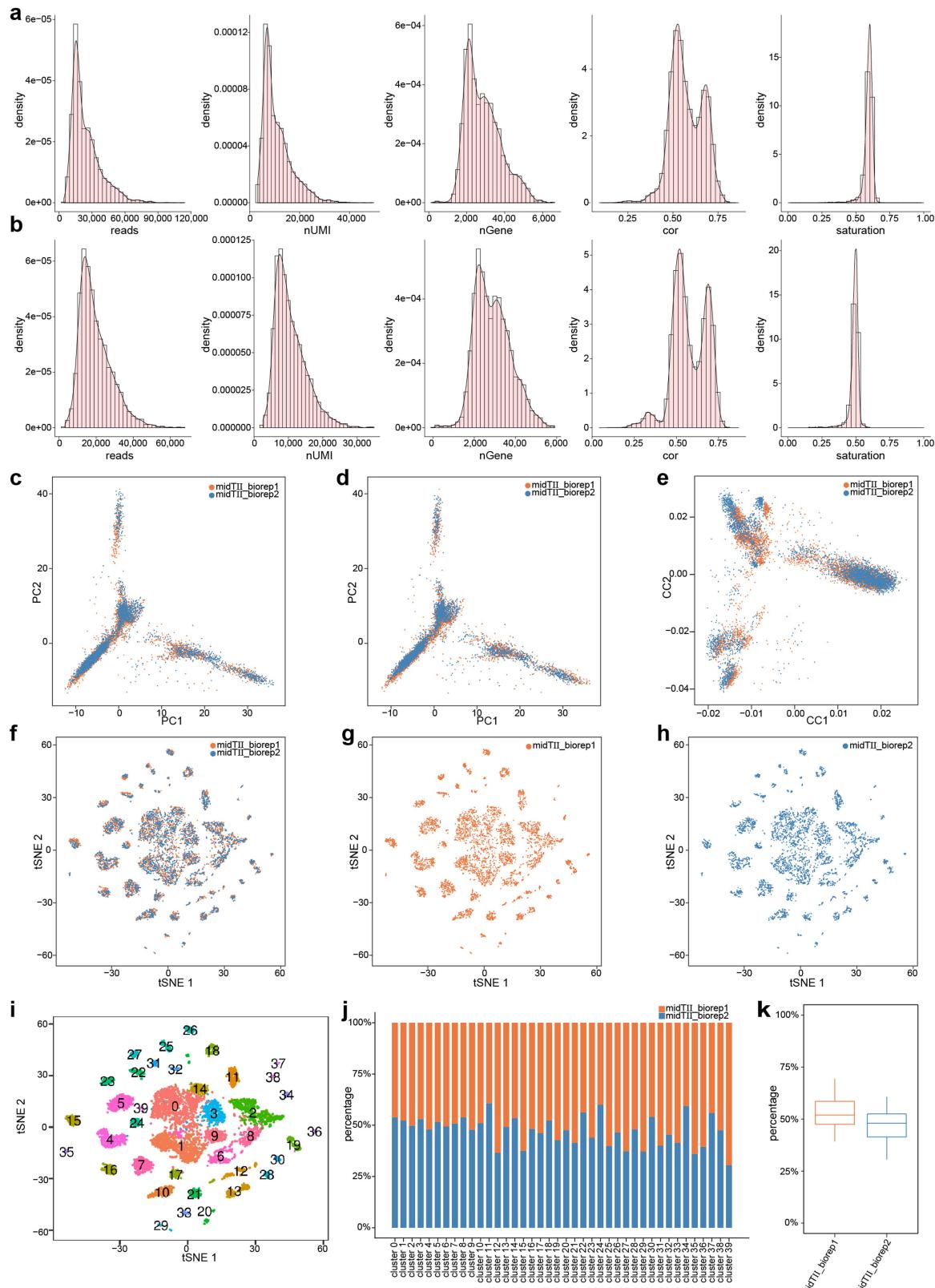
Author contributions K.C. and M.L. conceived the project. K.C., M.L., C.C., L.A.L. and W.W. (SIMR) designed the experiments. L.A.L., P.H.Y., Y.A.C. and K.C. performed *Ciona* experiments, W.W. (SIMR) performed killifish experiments, C.C. performed computational data analysis. L.R.P., J.C.M. and W.W. (LSI) set up the single-cell RNA-sequencing pipeline. M.L. supervised the project. All authors contributed to interpretation of the results, and C.C., L.A.L., K.C. and M.L. wrote the manuscript.

Competing interests The authors declare no competing interests.

Additional information

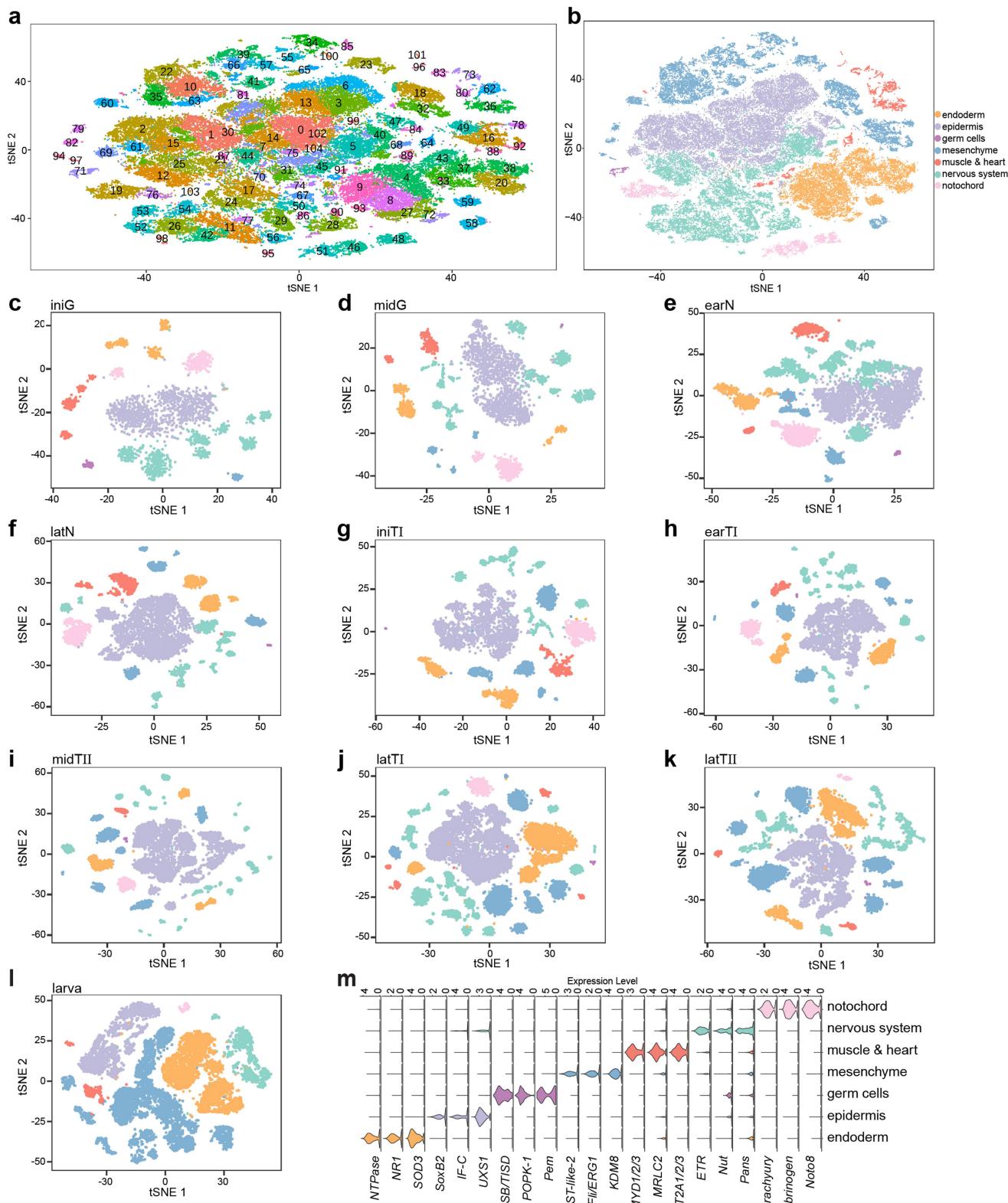
Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-019-1385-y>.

Correspondence and requests for materials should be addressed to M.L. or K.C. **Reprints and permissions information** is available at <http://www.nature.com/reprints>.



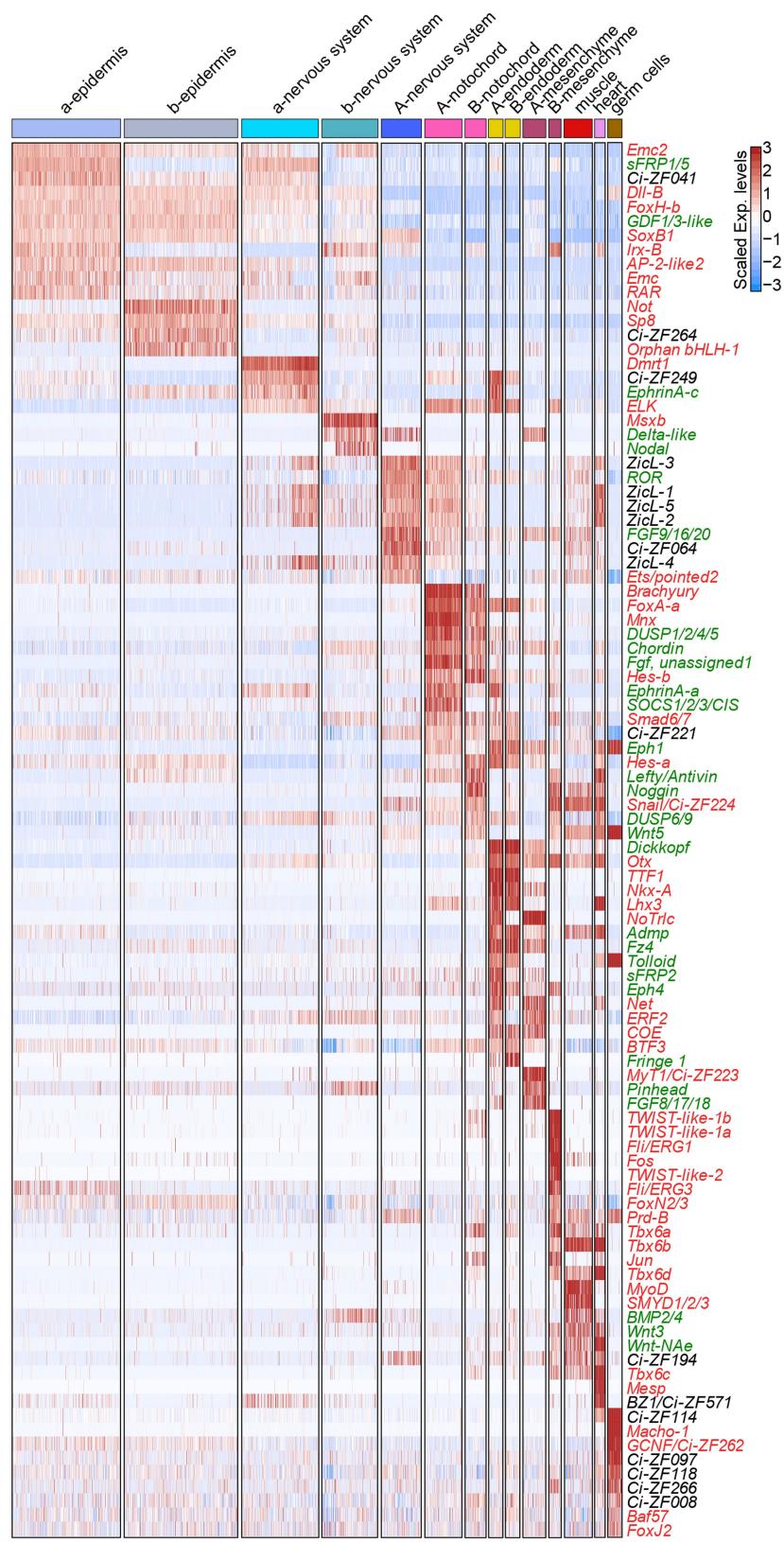
Extended Data Fig. 1 | Data quality and biological replicates from mid-tailbud stage. **a, b**, Distribution plot of reads numbers, UMIs, gene numbers, correlation coefficient (Spearman) and saturation level per cell from mid-tailbud (**a**, midTII.biorep1; **b**, midTII.biorep2). **c**, The first two principal components were plotted for cells regressed by UMIs (midTII.biorep1, $n = 4,929$ cells; midTII.biorep2, $n = 4,062$ cells). **d**, The first two principal components were plotted for cells regressed by both UMIs and batches. **e**, The first two canonical correlation vectors were plotted after alignment by canonical correlation analysis. **f–h**, Merged (**f**)–(**h**) t-SNE clustering for the biological replicates. **i**, t-SNE plot of canonical-correlation-analysis-aligned samples of biological replicates ($n = 8,991$ cells). The numbers indicate different clusters. **j**, The percentage of cells between replicates within the same cluster (clusters shown in **i**). **k**, Box plot of the percentage of cells in each cluster ($n = 40$ clusters) between replicates. The lower, middle and upper hinges correspond to the first and third quartiles (the 25th and 75th percentiles), and the middle hinge corresponds to the median.

and split (**g, h**) t-SNE clustering for the biological replicates. **i**, t-SNE plot of canonical-correlation-analysis-aligned samples of biological replicates ($n = 8,991$ cells). The numbers indicate different clusters. **j**, The percentage of cells between replicates within the same cluster (clusters shown in **i**). **k**, Box plot of the percentage of cells in each cluster ($n = 40$ clusters) between replicates. The lower, middle and upper hinges correspond to the first and third quartiles (the 25th and 75th percentiles), and the middle hinge corresponds to the median.



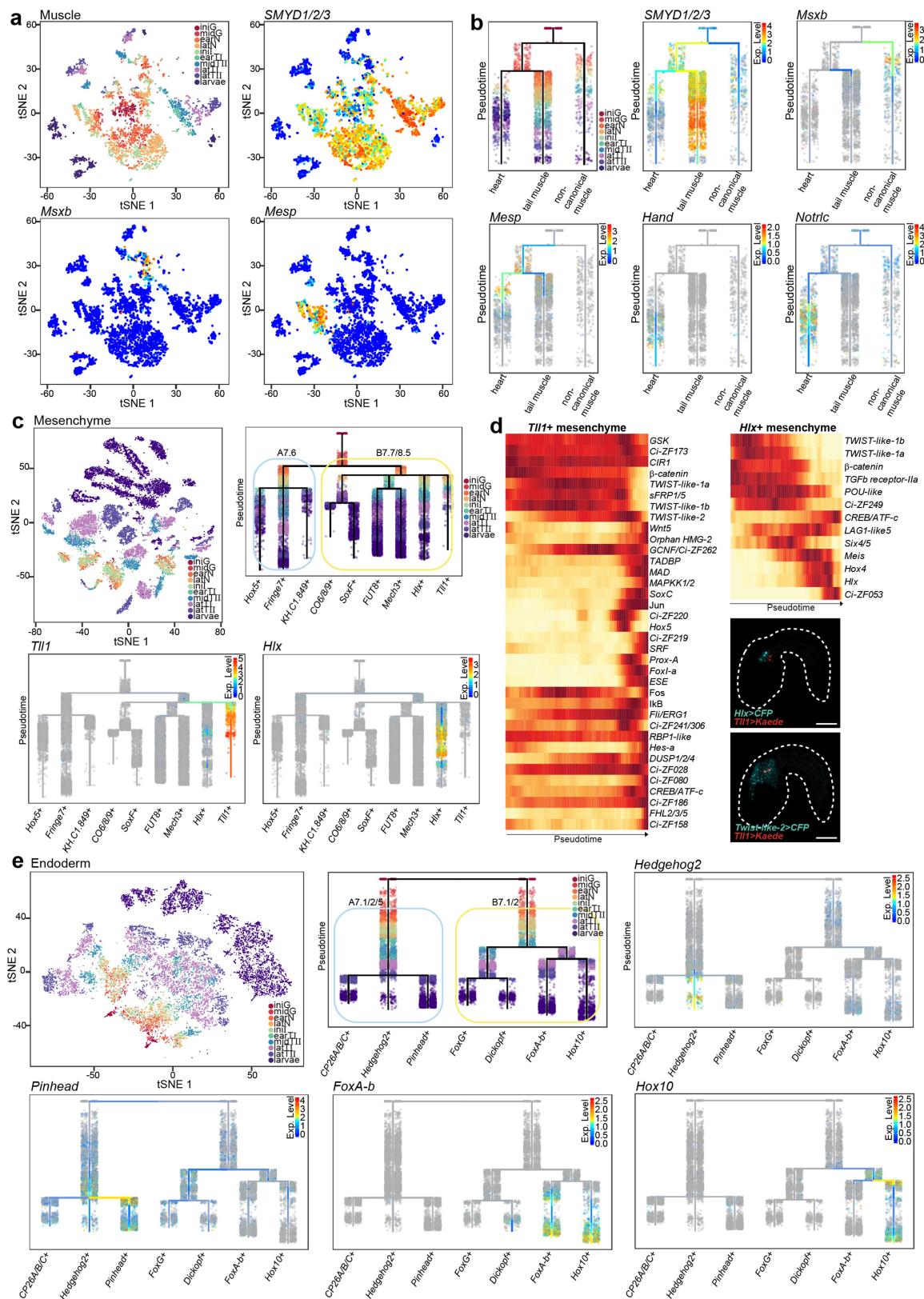
Extended Data Fig. 2 | t-SNE projections of ten stages from single-cell RNA-sequencing data. **a**, t-SNE plot of the entire dataset ($n = 90,579$ cells). Cells are coloured and labelled by clusters. Differentially expressed genes in each cluster can be found in Supplementary Table 2. **b**, t-SNE plot of all of the cells, coloured according to tissue type. **c–l**, t-SNE projections of cells, coloured by tissue types at different stages of development (iniG, $n = 2,863$ cells; midG, $n = 3,384$ cells; earN, $n = 7,154$ cells; latN, $n = 8,449$ cells; iniTI,

$n = 5,668$ cells; earTI, $n = 7,109$ cells; midTII, $n = 8,991$ cells; latTI, $n = 18,535$ cells; latTII, $n = 12,635$ cells; and larva, $n = 15,791$ cells). The colour code is the same as in b. m, Violin plots illustrating expression levels of representative marker genes per cell per tissue type (endoderm, $n = 14,162$ cells; epidermis, $n = 26,936$ cells; germ cells, $n = 396$ cells; mesenchyme, $n = 19,143$ cells; muscle and heart, $n = 3,691$ cells; nervous system, $n = 22,198$ cells; and notochord, $n = 4,053$ cells). Colour code is the same as in b.



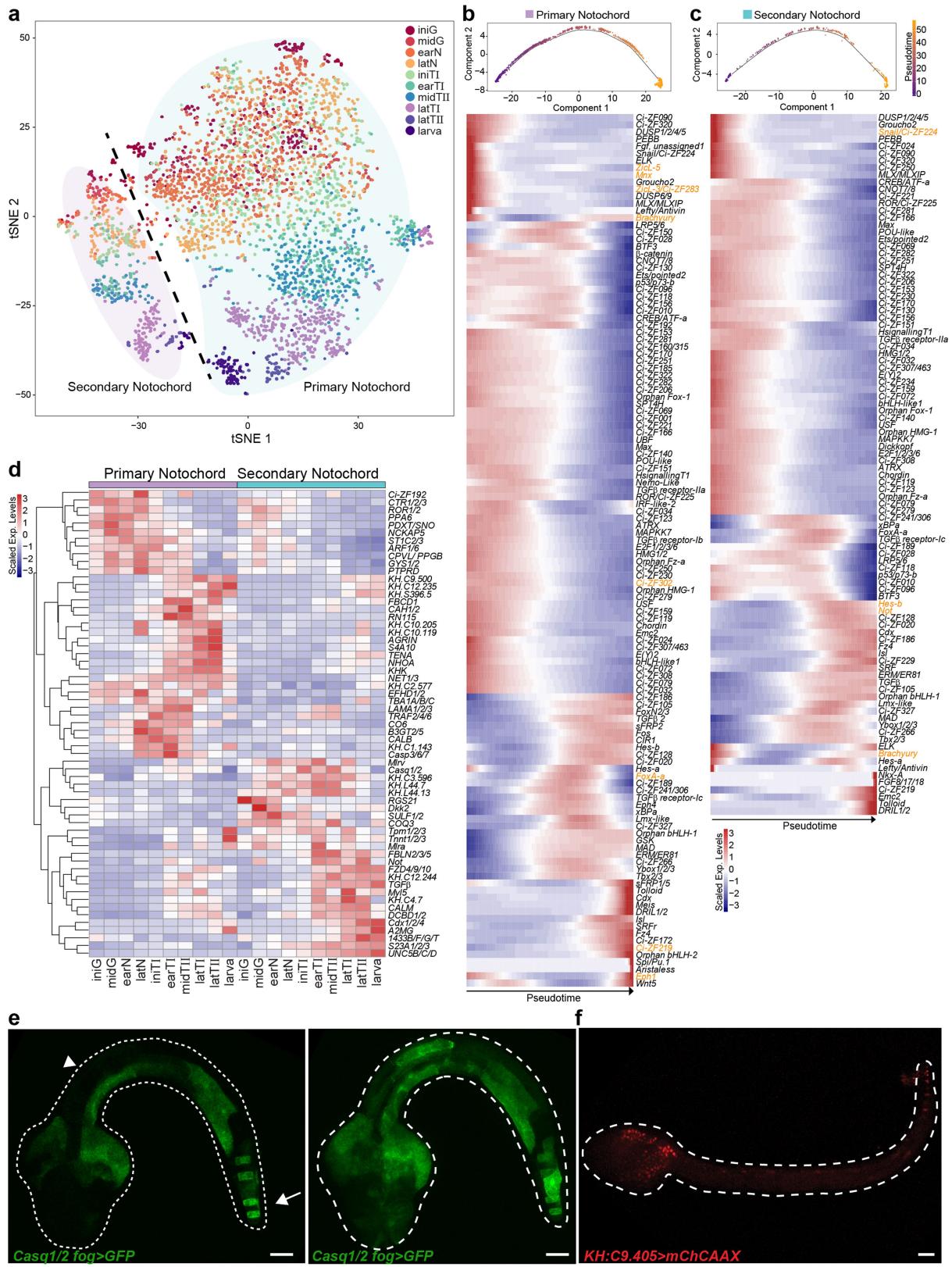
Extended Data Fig. 3 | Specification of cell types at the onset of gastrulation. The heat map shows the scaled expression of differentially expressed genes that encode transcription factors (red) and cell-signalling

components (green). Many marker genes were newly identified for each tissue.



Extended Data Fig. 4 | Reconstructed transcriptional trajectories of muscle, mesenchyme and endoderm. **a**, *t*-SNE projection and expression patterns of representative marker genes of tail muscle, non-canonical muscle and heart ($n = 3,691$ cells). **b**, Reconstructed transcriptome trajectories and expression patterns of representative marker genes in muscle. **c**, *t*-SNE projection and expression patterns of representative marker genes shown on reconstructed transcriptome trajectories of mesenchyme ($n = 19,143$ cells). **d**, Cascade of representative transcriptional trajectories for *Tll1*⁺ mesenchyme and *Hlx*⁺ mesenchyme. **e**, *t*-SNE projection and expression patterns of representative marker genes shown on seven reconstructed transcriptome trajectories of endoderm ($n = 14,162$ cells). Scale bars, 50 μ m.

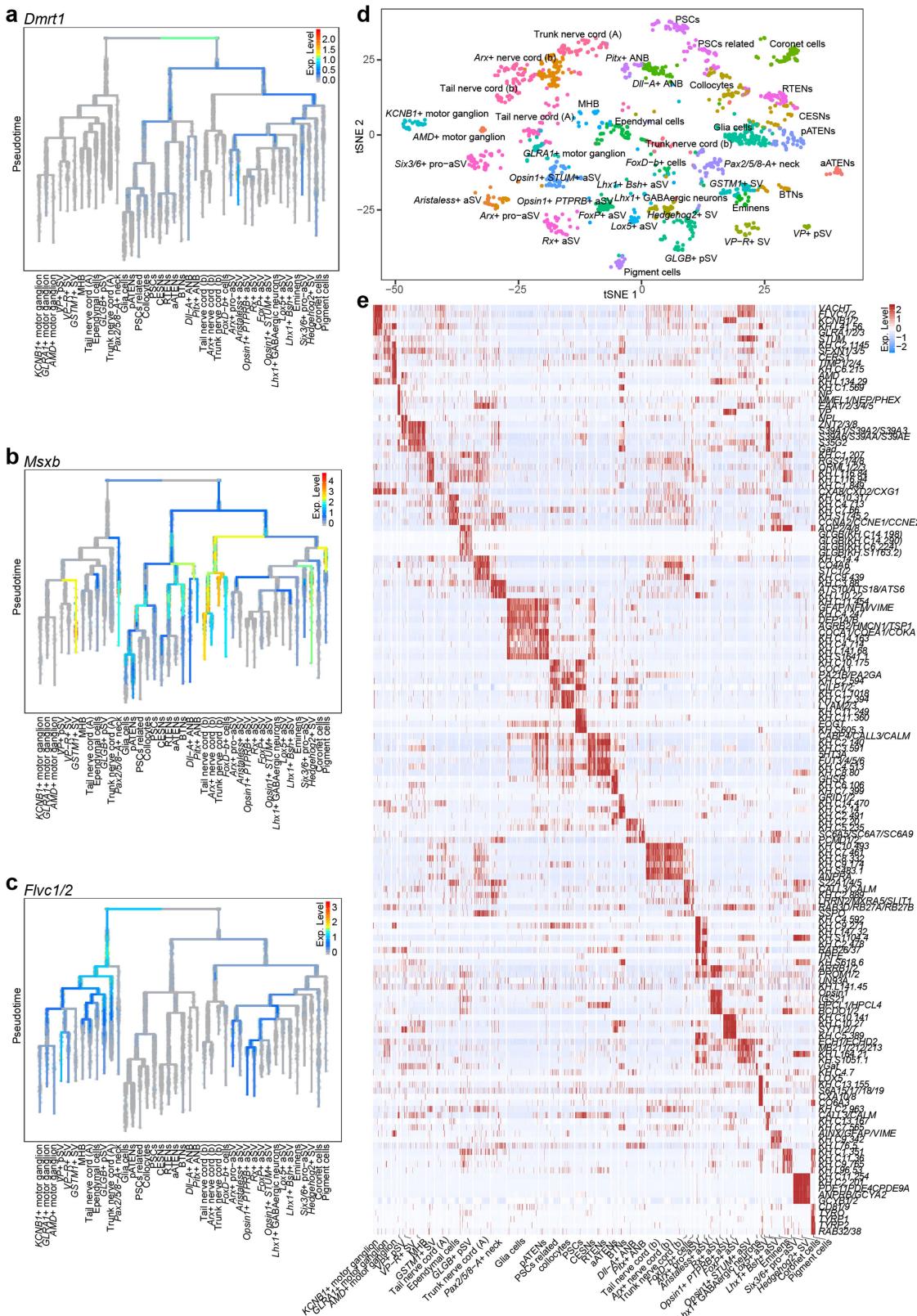
factors and signalling pathway genes along pseudotime in *Tll1*⁺ and *Hlx*⁺ mesenchyme. Mid-tailbud embryos that express *Twist-like-2* (cyan), a mesenchymal marker, and *Tll1* (red) reporter gene (top), and an *Hlx* (cyan) and *Tll1* (red) reporter gene (bottom, $n = 3$ electroporation experiments). **e**, *t*-SNE projection and expression patterns of representative marker genes along pseudotime in *Tll1*⁺ and *Hlx*⁺ mesenchyme. Mid-tailbud embryos that express *Twist-like-2* (cyan), a mesenchymal marker, and *Tll1* (red) reporter gene (top), and an *Hlx* (cyan) and *Tll1* (red) reporter gene (bottom, $n = 3$ electroporation experiments). **f**, *t*-SNE projection and expression patterns of representative marker genes shown on seven reconstructed transcriptome trajectories of endoderm ($n = 14,162$ cells). Scale bars, 50 μ m.



Extended Data Fig. 5 | See next page for caption.

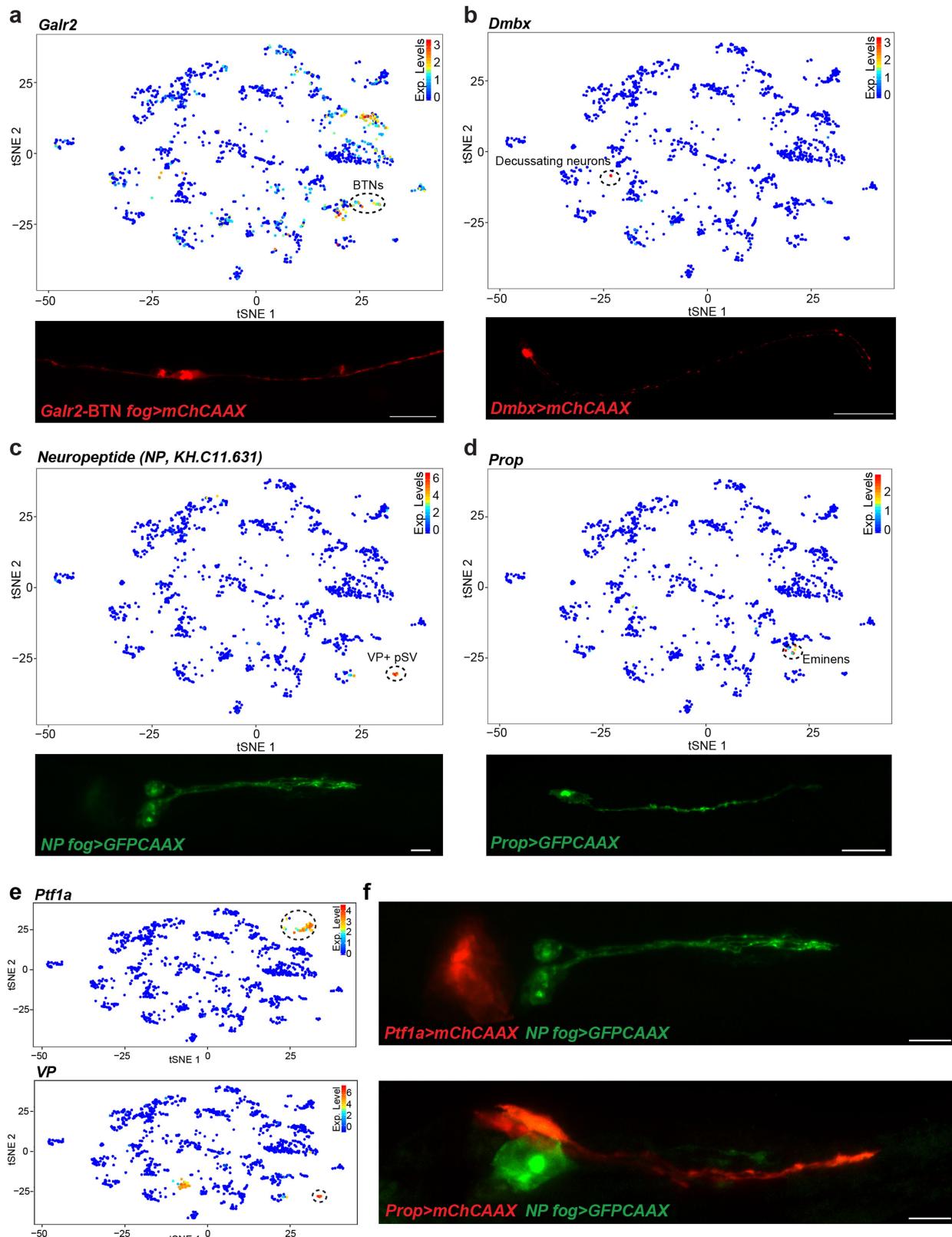
Extended Data Fig. 5 | Transcriptome profiles of *Ciona* notochord cells during development. **a**, t-SNE plot of notochord cells. Cells are coloured by developmental stage. The dashed line shows the separation between the primary ($n = 3,123$ cells) and secondary lineages ($n = 627$ cells). **b, c**, The single-cell transcriptome trajectory (top) and pseudotemporal gene-expression profiles (bottom) of the primary notochord and the secondary notochord. Cells were ordered along the trajectory across pseudotime. Only significantly expressed genes (likelihood ratio test) with $q < 1 \times 10^{-100}$ (primary notochord) and $q < 1 \times 10^{-20}$ (secondary notochord) are shown. Selected transcription factors and signalling molecules are labelled in orange. **d**, Heat map of differentially expressed

genes between the primary and secondary notochord. Genes are clustered by Euclidean distance. **e**, Expression of a *Casq1/2fog>GFP* reporter gene in a late-tailbud-stage embryo (left, one optical plane; right, maximum intensity projection). $n = 3$ electroporation experiments. GFP (green) was present in the muscle and in the secondary notochord (arrow), but no expression was observed in the primary notochord (arrowhead). **f**, Expression of *KH.C9.405>mChCAAX* reporter gene in late tailbud II stage embryo. *mChCAAX* (red) was present in the secondary notochord but not the primary notochord. $n = 3$ electroporation experiments. Scale bars, 20 μm .



Extended Data Fig. 6 | Neural cells. **a–c**, Expression patterns of representative marker genes for the a- (**a**), b- (**b**) and A-lineages (**c**) are shown in reconstructed transcriptome trajectories of neural cells that span ten developmental stages. **d**, t-SNE plot of neural cells recovered from the

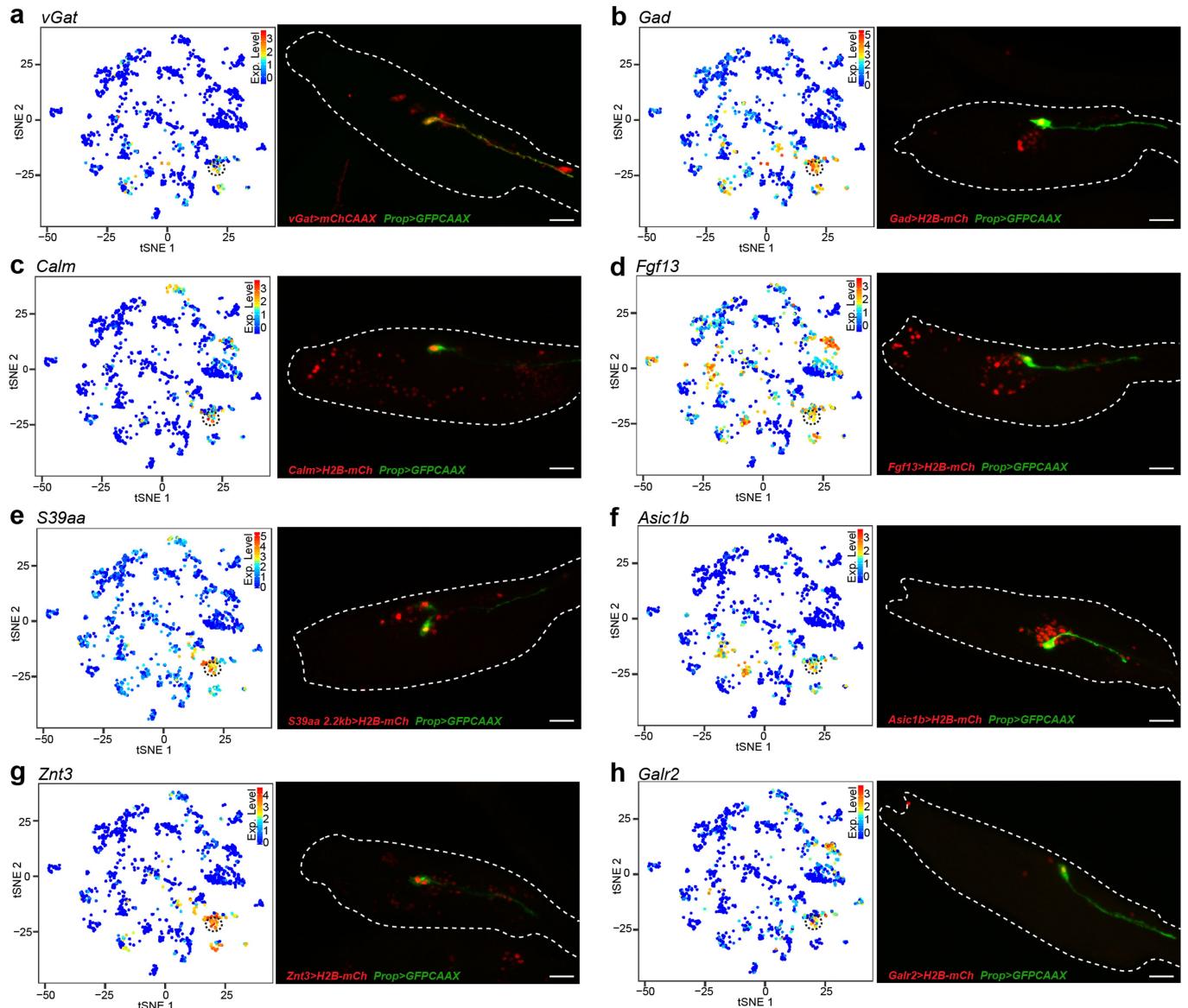
larval stage ($n = 1,704$ cells). Identified cell types are labelled. e, Heat map of the top-five differentially expressed genes (not including those encoding transcription factors) for each type of neural cell in the larval stage.



Extended Data Fig. 7 | See next page for caption.

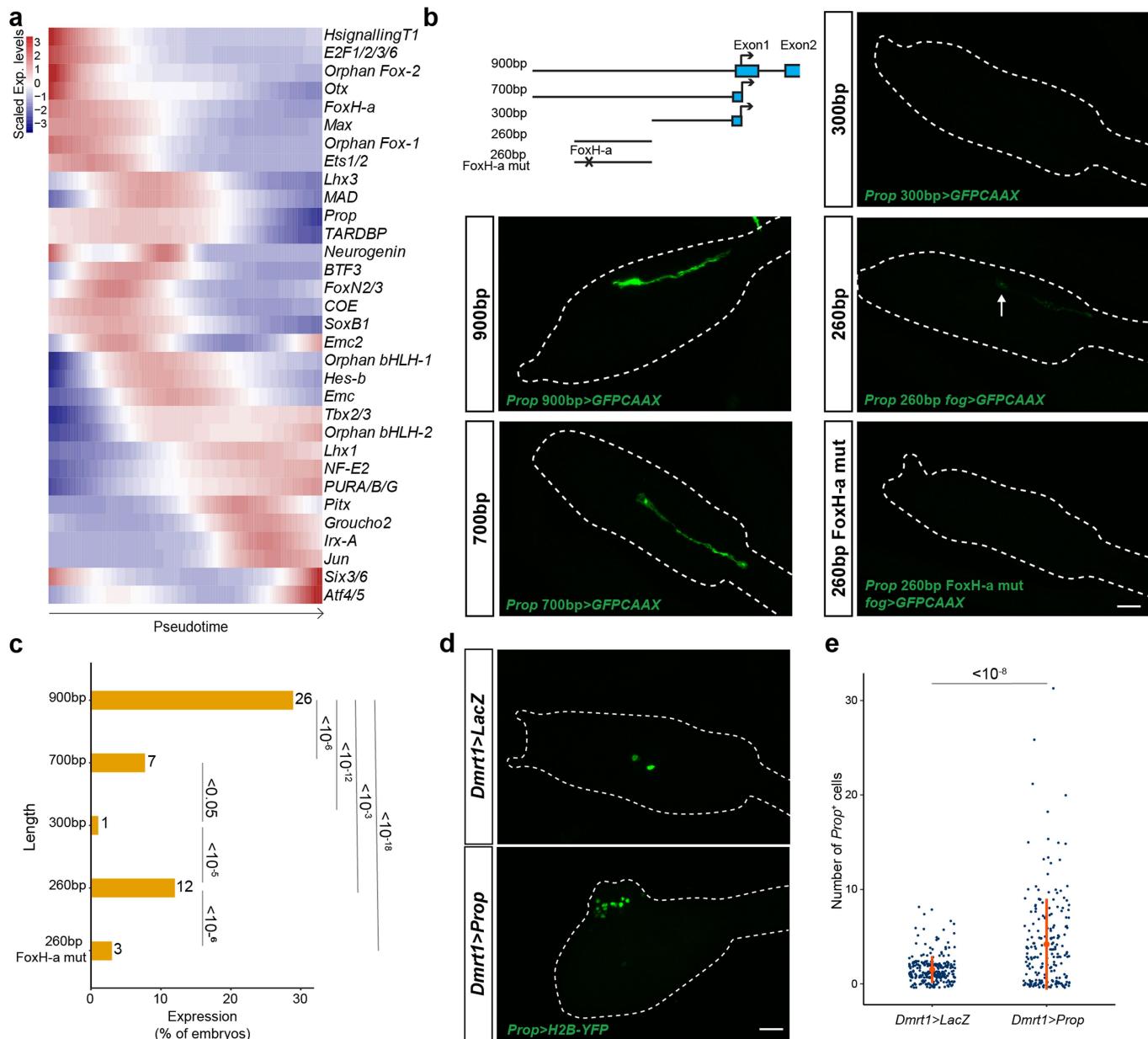
Extended Data Fig. 7 | The identification of rare neuronal subtypes in the larval stage. **a**, Distribution of cells that express *Galr2* in the t-SNE plot. Cells within the dashed circle show *Galr2* expression in bipolar tail neurons ($n = 26$ cells). Reporter assay with a bipolar-tail-neuron minimal enhancer for *Galr2* shows the specific activity of *Galr2* in bipolar tail neurons ($n = 3$ electroporation experiments). **b**, Distribution of cells that express *Dmbx* in the t-SNE plot. Cells within the dashed circle show *Dmbx* expression in decussating neurons ($n = 4$ cells). The 5' regulatory sequences of *Dmbx* are active in decussating neurons (red, $n = 3$ electroporation experiments). **c**, Distribution of cells that express *NP* in the t-SNE plot. Cells within the dashed circle show *NP* expression in VP^+ posterior sensory vesicle ($n = 11$ cells). Reporter assay for *NP* (green) shows the specific expression of *NP* in neurons in the posterior sensory vesicle ($n = 3$ electroporation experiments). **d**, Distribution of cells that express *Prop* in the t-SNE plot. Cells within the dashed circle

show *Prop* expression in Eminens neurons ($n = 17$ cells). Expression of the *Prop* reporter gene is specific to Eminens neurons (green) ($n = 3$ electroporation experiments). **e**, t-SNE plot of the larval nervous system showing cells that express *Ptf1a* (top) and VP (bottom). The dotted circle corresponds to coronet cells (top, $n = 72$ cells) and VP^+ posterior sensory vesicle cluster (bottom, $n = 11$ cells). **f**, Expression of the reporter *Ptf1a>mChCAAX* (red) for coronet cells and *NP>GFPCAAAX* (green) for VP^+ posterior sensory vesicle shows that these cell populations do not contact each other, but are in close vicinity (top, $n = 3$ electroporation experiments; the GFP channel is shown in c). Expression of the reporter *Prop>mChCAAX* (red) for Eminens neurons and *NP>GFPCAAAX* (green) for VP^+ posterior sensory vesicle. NP^+ cells are also in proximity to Eminens neurons (bottom, $n = 2$ electroporation experiments). Scale bars, 10 μm .



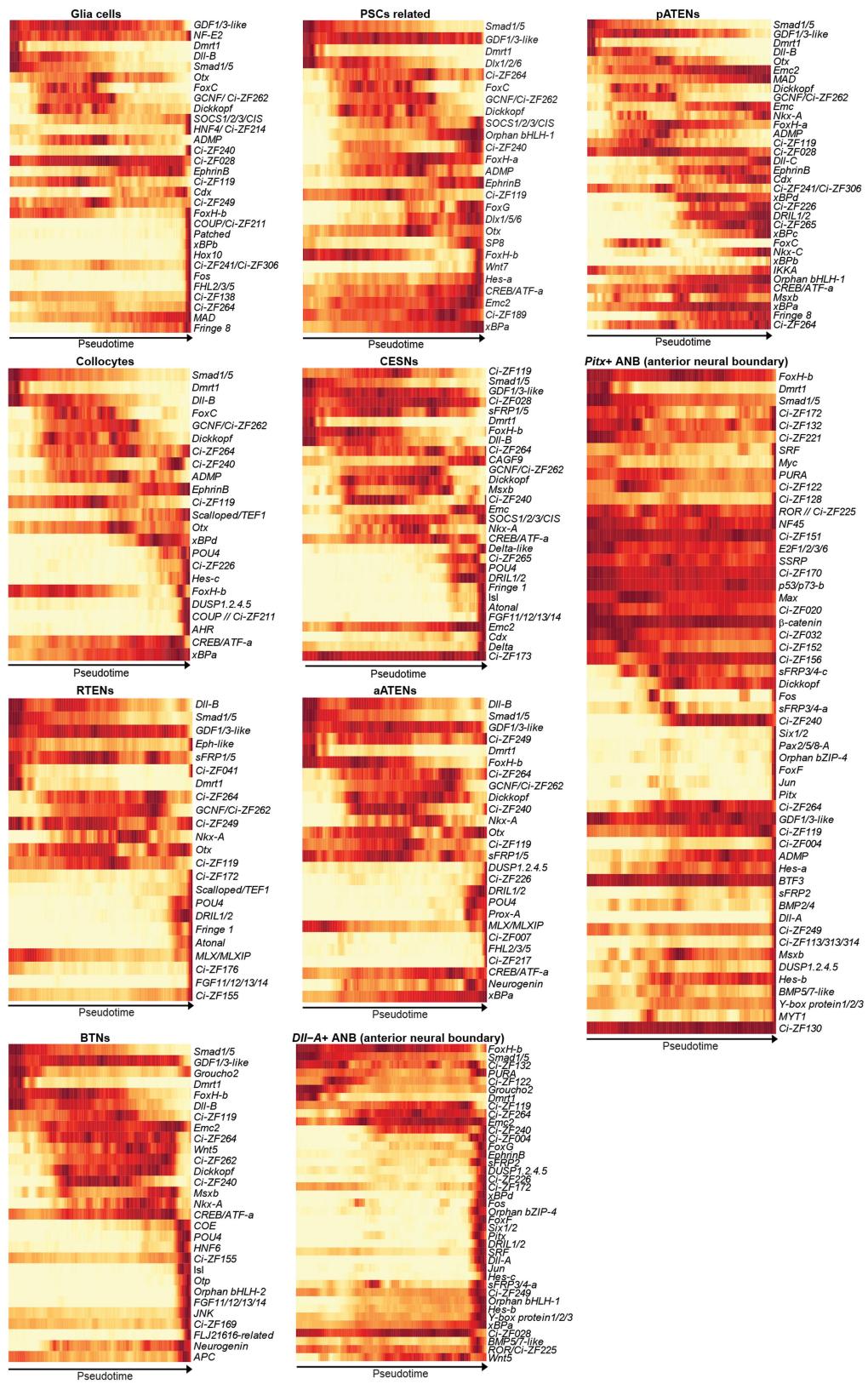
Extended Data Fig. 8 | Expression of marker genes for Eminens neurons. **a–h**, Expression levels of eight marker genes in the larval nervous system, shown in *t*-SNE plots (left, $n = 1,704$ cells), and their corresponding reporter assays (*mChCAAX* for *vGat* and *H2B-mCherry* for the other genes, red) with a *Prop>GFP CAAX* reporter (green, right).

$n = 2$ electroporation experiments for *Gad*, *S39aa* 2.2 kb, *Znt3* and *Asic1b*; $n = 3$ electroporation experiments for *vGat*, *Calm*, *Fgf13* and *Galr2*. The dashed circle in the *t*-SNE plots identifies Eminens neurons. Scale bars, 20 μm .

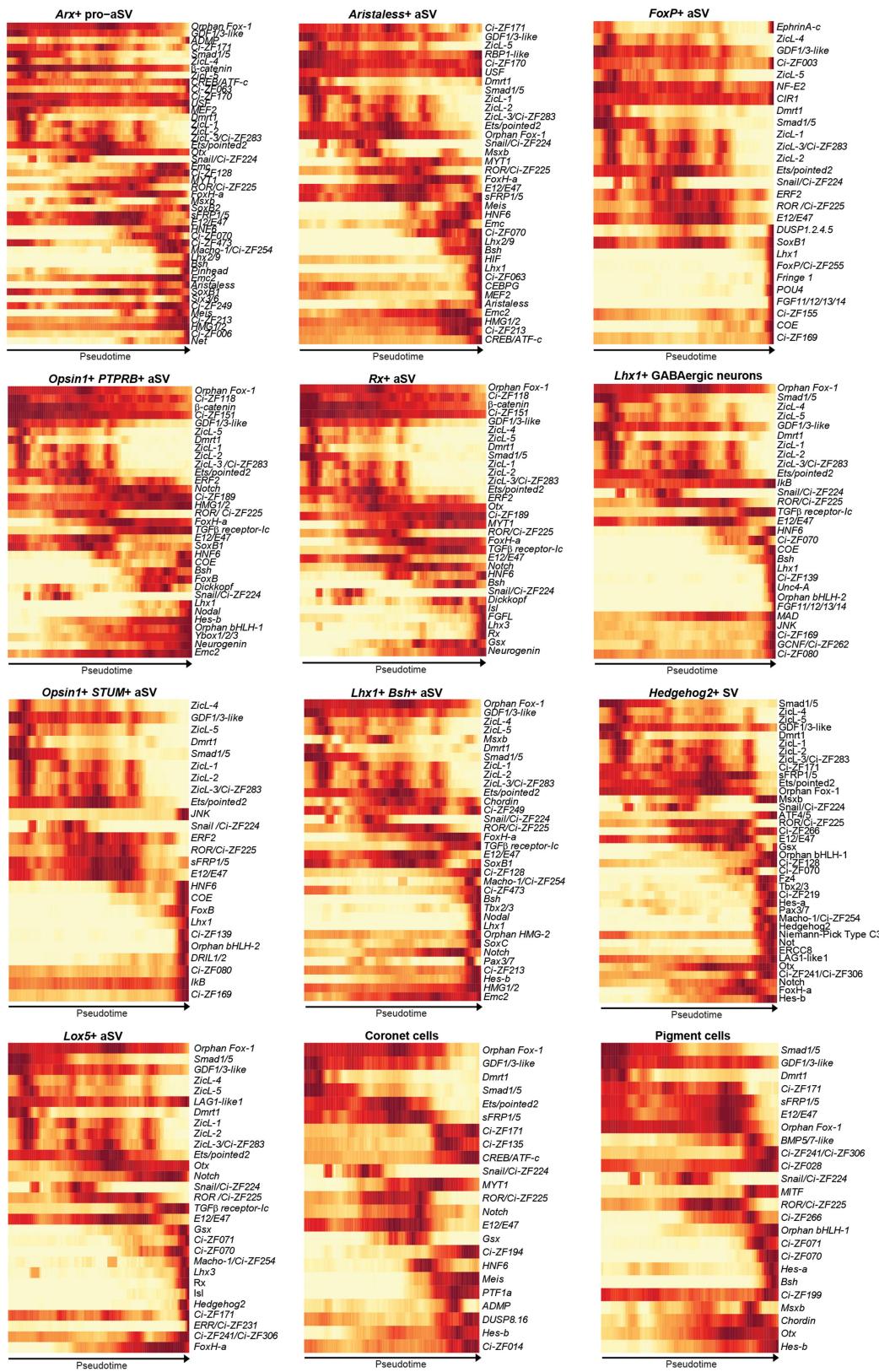


Extended Data Fig. 9 | Manipulation of Eminens gene regulatory network. **a**, Pseudotemporal expression profiles of regulatory genes and signalling components in Eminens neurons. **b**, Diagram of the *Prop* regulatory sequences with their length indicated on the left. A representative embryo is shown for the different fusion genes (GFP-CAAX, green). The minimal *Prop* enhancer has weak expression in Eminens neurons (arrow). When the binding site for FoxH-a was mutated (260 bp FoxH-a mut), these regulatory sequences show even less activity. **c**, Bar plot of the percentage of the embryos that express GFP shown in **b**. Numbers on the right of the column correspond to the percentage of GFP+ embryos. χ^2 test with four degrees of freedom was performed ($P < 2.2 \times 10^{-16}$), followed by two-sided Fisher's exact test with Bonferroni adjustment for multiple comparisons. P values: 900 bp versus 700 bp, $P = 1.05 \times 10^{-7}$; 900 bp versus 300 bp, $P = 3.47 \times 10^{-13}$; 900 bp versus 260 bp, $P = 2.36 \times 10^{-4}$; 900 bp versus 260-bp FoxH-a mut, $P = 1.81 \times 10^{-19}$; 700 bp versus 300 bp, $P = 0.011$; 700 bp versus 260 bp,

$P = 0.36$; 700 bp versus 260-bp FoxH-a mut, $P = 0.088$; 300 bp versus 260 bp, $P = 5.59 \times 10^{-6}$; 300 bp versus 260-bp FoxH-a mut, $P = 0.69$; 260 bp versus 260-bp FoxH-a mut, $P = 1.27 \times 10^{-7}$. Numbers of embryos: 900 bp, $n = 207$; 700 bp, $n = 300$; 300 bp, $n = 160$, all pooled over 2 electroporation experiments; 260 bp, $n = 440$, 260-bp FoxH-a mut, $n = 750$, all pooled over 3 electroporation experiments. **d**, Overexpression of *Prop* using *Dmrt1* regulatory sequences causes supernumerary *Prop*+ cells (bottom panel) compared to control embryos expressing *LacZ* (top). The 2-kb *Prop* reporter gene shows specific expression in Eminens neurons (H2B-YFP, green). The images show representative embryos for both conditions. **e**, Quantification of *Prop*+ cells from the experiments in **d**. *Dmrt1>LacZ*, $n = 269$ embryos; *Dmrt1>Prop*, $n = 210$ embryos, pooled over 3 electroporation experiments. The orange dots indicate the mean and the bars indicate the s.d. *Dmrt1>LacZ*, 1.5 ± 1.4 cells; *Dmrt1>Prop*, 4.2 ± 4.8 cells. Mann-Whitney *U*-test, $P = 3.65 \times 10^{-9}$. Scale bars, 20 μ m.

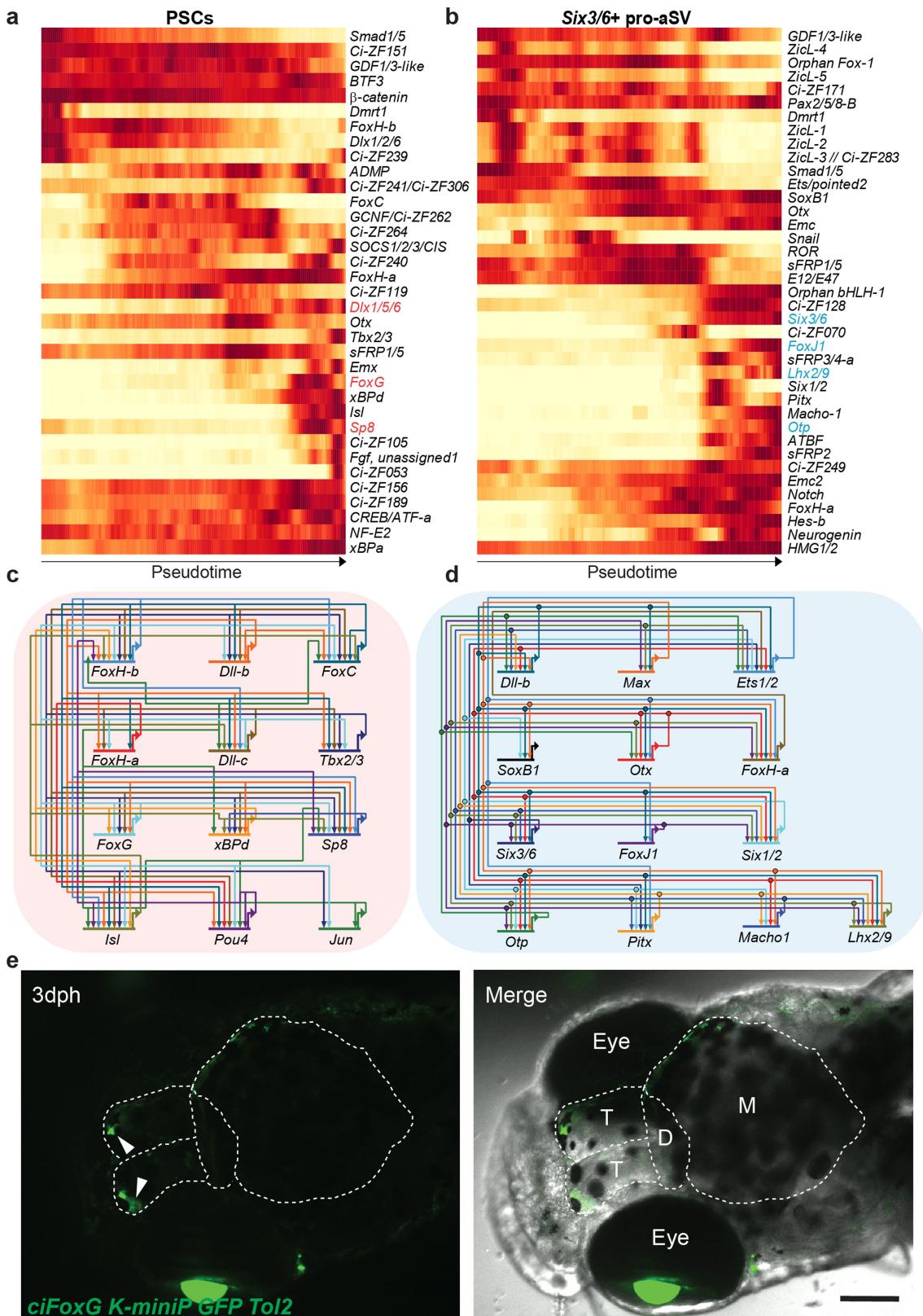


Extended Data Fig. 10 | Pseudotemporal gene-expression cascade of the peripheral nervous system. Representative transcription factors and signalling pathway genes along pseudotime in the reconstructed developmental trajectories of the peripheral nervous system are shown.



Extended Data Fig. 11 | Pseudotemporal gene-expression cascade of the central nervous system of a-lineage. Representative transcription factors and signalling pathway genes along pseudotime in reconstructed

developmental trajectories of the central nervous system of a-lineage are shown.



Extended Data Fig. 12 | Model for the evolution of the telencephalon.

a, Gene-expression cascade of regulatory genes and signalling components of palp sensory cells (also known as axial columnar cells). Genes implicated in the development of the vertebrate telencephalon are labelled in red. **b**, Gene-expression cascade of regulatory genes and signalling components in the anterior-most regions of the sensory vesicle (*Six3/6*⁺ pro-anterior sensory vesicle). Genes implicated in vertebrate telencephalon development are labelled in blue. **c, d**, The putative

regulatory interactions among transcription factors from the cascade of palp sensory cells (**c**) and *Six3/6*⁺ pro-anterior sensory vesicle (**d**) along their developmental trajectories. **e**, The *FoxG* reporter gene with *Ciona* enhancer sequence exhibits restricted expression in a subset of cells in the olfactory bulb of the killifish telencephalon (arrowheads) and in the eye lens (left, GFP channel; right, merged image of bright-field and GFP channel images). $n = 3$ independent transgenic lines (Methods). D, diencephalon; M, midbrain; T, telencephalon. Scale bar, 400 μ m.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Raw sequencing reads were filtered by Illumina HiSeq Control Software and only pass-filter reads were used for further analysis. Samples were run on both lanes of a HiSeq 2500 Rapid Run mode flow cell. Base calling was performed by Illumina RTA version 1.18.64.0. BCL files were then converted to FASTQ format using bcl2fastq version 1.8.4 (Illumina). Reads that aligned to phix (using Bowtie version 1.1.1) were removed as well as reads that failed Illumina's default chastity filter. We then combined the FASTQ files from each lane and separated the samples using the barcode sequences allowing 1 mismatch (using barcode_splitter version 0.18.2). Using 10x CellRanger version 2.0.1, the count pipeline was run with default settings on the FASTQ files to generate gene–barcode matrices for each sample.

Data analysis

For dimensional reduction, clustering and t-SNE visualization, Seurat v2.3.4 was applied with an implement of a modified Fast Fourier Transform-accelerated Interpolation-based t-SNE method.
In order to capture the developmental transitions stemming from different blastomeres at 110 cell stage, we performed “ancestor voting” between clusters across time as described in Briggs, J. A. et al. 2018. For notochord and Eminens cells, we employed monocle 2 to construct the single cell trajectory. For tissues that harbored more complexity during development, such as the mesenchyme and nervous system, we employed a simulated diffusion-based computational reconstruction method, URD, for acquiring the transcriptional trajectories during embryogenesis. Cluster-buster was used to find clusters of pre-specified motifs in 2kb upstream of the TSS of each gene.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Raw sequencing data and gene expression matrix are available in Gene Expression Omnibus (GEO) under accession number GSE131155

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	For the sequencing, 100 to 500 embryos per samples were collected for following dissociation and scRNA-Seq. For the functional analysis and reporter assay, 18 to 610 embryos per sample were used. For the killifish reporter assay, between 25 and 46 embryos per lines were analyzed.
Data exclusions	To remove signals from putative empty droplet or degraded RNA, low-quality transcriptomes were filtered for each time course sample, as follows: 1) we discarded cells with less than 1000 expressed genes; 2) Cells with UMLs exhibiting five SDs above the mean were not included in our analyses (Supplementary Table 1); 3) we only consider genes that were expressed in at least 3 cells in each dataset. In total, 90,579 cells were kept for the following analysis.
Replication	We performed 2 biological replicates for stages from initial gastrula to late tail bud II, and 3 replicates for swimming tadpole stage. All the reporter assays were performed at least twice. Two replicates were done for the minimal Prop enhancer assay. Three replicates were done for the overexpression assay as well as for the mutation assay on Prop minimal enhancer. Three different lines of transgenic killifish were generated for the reporter assay.
Randomization	Embryos in experiment were randomized and collected for dissociation before cell indexing on 10X Genomics Chromium system.
Blinding	Investigators were blinded to group allocation during data collection and analysis: embryo collection and scRNA-Seq data analysis were performed by two different researchers. For functional assays, no particular blinding strategy was adopted. Experiments were performed by one researcher.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Animals and other organisms

Policy information about [studies involving animals; ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals

Experiment of African killifish *N. furzeri* were performed using the GRZ strain

Wild animals

Ciona intestinalis were purchased from M-REP, San Diego, California, which collected them in San Diego area.

Field-collected samples

OR

Ethics oversight

Ciona intestinalis are non-vertebrates; Work with killifish was performed according to guidelines of the Stowers Institute for Medical Research.

Note that full information on the approval of the study protocol must also be provided in the manuscript.