Single-cell RNA-sequencing methods are revolutionizing our understanding of how cells are specified to become definitive tissues during development\(^1\). 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 vertebrates\(^6\), the ascidian Ciona 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 embryogenesis\(^8\) 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 neurons\(^8\). The resulting high-resolution transcriptome trajectories for different tissues 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.

Comprehensive 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 synаптоме of swimming tadpoles and tracing the evolutionary origin of cell types such as the vertebrate telencephalon.

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-sequencing methods are revolutionizing our understanding of how cells are specified to become definitive tissues during development\(^1\). 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 vertebrates\(^6\), the ascidian Ciona 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 embryogenesis\(^8\) 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 neurons\(^8\). The resulting high-resolution transcriptome trajectories for different tissues 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.

**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).

\[^{1}\] Lewis-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. Present address: The Yunnan Key Laboratory of Primate Translational Medicine, Kunming University of Science and Technology, Kunming, China. 4These authors contributed equally: Chen Cao, Laurence A. Lemaire. *e-mail: msl2@princeton.edu; chenk@lpbr.cn

---

**Comprehensive single-cell transcriptome lineages of a proto-vertebrate**

Chen Cao\(^1\)\(^5\), Laurence A. Lemaire\(^1\)\(^5\), Wei Wang\(^2\), Peter H. Yoon\(^3\), Yoolim A. Choi\(^3\), Lance R. Parsons\(^1\), John C. Matese\(^1\), Wei Wang\(^1\), Michael Levine\(^1\)\(^5\) & Kai Chen\(^1\)\(^5\)

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.
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 late determinants (Extended Data Fig. 3, Supplementary Table 2); for example, \textit{Jrx-B} is specifically expressed in a lineage (anterior) epidermis and \textit{Not} is expressed in the b-lineage (posterior).

**Reconstructing cell lineages**

The alignment of transcriptome profiles of individual cell types at each of the 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 notochord (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), 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, f). This is a considerably higher level of resolution than that obtained by conventional experimental studies.

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 neuropore. 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 chordates. 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 tail, 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 *Zic1* and *Notch*), 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 *Tbx6*, *Tbx8* and *Tbx16* (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 (*Mlr1/Mlr2/Myl5*) and tropomyosin (*Tpm1/2/3*) (Extended Data Fig. 5d, Supplementary Table 2). None of these genes is expressed in the primary notochord. 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 neurons, 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 cells. 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 tadpoles, and these were found to express galanin and two of its receptors (*Galr1* and *Galr2*) (Supplementary Table 2). Galanin has previously been implicated in neuro-regeneration and axogenesis. 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. 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 average level of expression. **a**ATENs, anterior apical trunk epithelial neurons; **ANB**, anterior neural boundary; **aSV**, anterior apical trunk epidermal neurons; **ANB**, anterior neural boundary; **aSV**, anterior apical trunk epidermal neurons; **MHB**, midbrain–hindbrain boundary; **pATENs**, posterior apical trunk epithelial neurons; **RTENs**, rostral trunk epithelial neurons; **SV**, sensory vesicle. Letters in parentheses denote lineages.
of vertebrates\textsuperscript{25,26}—was unambiguously identified on their 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 serotonin, substance P, and an uncharacterized neuropeptide \textsuperscript{28} (Extended Data Fig. 7b). Upon overexpression of Prop in the anterior neural plate, supernumerary S39aa\textsuperscript{+} cells (H2B–mCherry (H2B–mCh), red) are observed (bottom, Dmrt1 > Prop).

**Transcriptome and synaptome integration**

The recently reported Giona synaptome identified a single Eminens neuron (Em2) as a key regulator of decussating neurons\textsuperscript{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 Eminens neurons, including morphology and location\textsuperscript{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 single-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\textsuperscript{2,33} (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 FoxH1-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 frogs. The decussating neurons integrate a variety of sensory information to trigger a fast escape reflex. As predicted by previous studies, 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 vertebrates. Galanin promotes survival of dorsal root ganglia neurons during development and after injury. It is possible that galanin serves as a trophic 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 synaptoptome studies, 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. 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 neuromere, 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 cells, 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. 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.

We propose that the vertebrate telencephalon arose through the incorporation of non-neural ectoderm in anterior regions of the 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 palps 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 telencephalon. Current single-cell studies encompasses a broad spectrum of cell types and systems, 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.
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. Delucu, F., Brinkmann, H., Courrout, 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. *Development* **133**, 138–149 (2010).
8. Suzuki, M. M. & Satoh, N. Genes expressed in the amphioxus notochord revealed by EST analysis. *Dev. Biol.*** **224**, 168–177 (2000).
9. Ryan, K. et al. The peripheral nervous system of the larval brain. *Ciona intestinalis* (L) highlights sidedness in the brain of a chordate sibling. *eLife* **5**, e16962 (2016).
10. Drom, 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).
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. Ishida, H. Cell lineage analysis in ascidian embryos by intracellular injection of a tracer enzyme. Ill. 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. Veerman, 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. Stemml, D. L. Structure and function of the notochord: an essential organ for chordate development. *Development* **132**, 2503–2512 (2005).
16. Yagi, K., Satou, T. & Satoh, N. A zinc finger transcription factor, ZicL, is a direct activator of Brachury in the notochord specification of *Ciona intestinalis*. *Development* **131**, 1279–1288 (2004).
17. Hudson, C. & Yasuo, H. A signaling relay involving Nodal and Delta ligands acts during secondary notochord induction in *Ciona* embryos. *Development* **133**, 2555–2564 (2006).
18. Yagi, K., Takatori, N., Satou, Y. & Satoh, N. C-tbx6b and C-tbx6c are key mediators of the maternal effect gene C-mach1 in muscle cell differentiation in *Ciona intestinalis* embryos. *Dev. Biol.*** **282**, 539–549 (2005).
19. Takashashi, H. et al. Brachury downstream notochord differentiation in the ascidian embryo. *Science* **332**, 1297–1302 (2011).
20. Horie, T. et al. Regulatory circuitry for dopaminergic neurons in a prot vertebrate identified by whole-embryo single-cell transcriptomics. *Genes Dev.*** **32**, 1297–1302 (2018).
21. Stolfi, A. et al. Circuit homology between decussating pathways in the *Ciona* larval CNS and the vertebrate startle-response pathway. *Curr. Biol.*** **27**, 721–728 (2017).
22. Korn, H. & Faber, D. S. The Mauthner cell half a century later: a neurobiological model for decision-making? *Neuron* **47**, 13–28 (2005).
23. Stolfi, A. & Levine, M. Neuronal subtype specification in the spinal cord of a protovertebrate. *Development* **138**, 995–1004 (2011).
24. Hamada, M. et al. Expression of neuropeptide- and hormone-encoding genes in the *Ciona intestinalis* larval brain. *Dev. Biol.*** **352**, 202–214 (2011).
25. 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).
26. Ryan, K. et al. The CNS connectome of a tadpole larva of *Ciona intestinalis*: I. Neuronal neurons of the ascidian larval nervous system in *Ciona intestinalis*: I. Central nervous system. *J. Comp. Neurol.*** **501**, 316–334 (2007).
27. 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).
28. 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).
29. 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).
30. Abitua, P. B. et al. The pre-vertebrate origins of neurogenic placodes. *Nature*** **524**, 462–465 (2015).
31. 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).
32. Stolfi, A. et al. Early chordate origins of the vertebrate second heart field. *Science*** **329**, 565–568 (2010).
33. Horie, R. et al. Shared evolutionary origin of vertebrate neural crest and cranial placodes. *Nature*** **560**, 228–232 (2018).
34. Zeng, F. et al. Papillae revisited and the nature of the adhesive secreting cillocytes. *Dev. Biol.*** **448**, 183–198 (2019).
35. Hiebert, J. M. & Fishell, G. The genetics of early telencephalon patterning: some assembly required. *Nat. Rev. Neurosci.*** **9**, 678–685 (2008).
36. Zembrzycki, 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).
37. 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).
38. 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).
39. Zhang, 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
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 further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis. We further normalized the read counts of each cell by Seurat methods47, and the normalized read counts were log-transformed for further analysis.
transition probabilities between cells with the destiny package. 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 = 40, 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× higher than in the sibling branch, and the genes were 1.25× 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 and selected highly expressed genes for specific cell types at the final time point that had a fold change (expressed in log2) above one between groups were all used in investigating the putative direct interaction. We used cluster-buster 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 database. 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 v2.3.2. Only genes with an average fold change (expressed log-arithmically) > 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 DESeq2: (1) FDR (false discovery rate) adjusted P value below 0.05; and (2) absolute fold change (expressed in log2) 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 heatmap2.10. For Fig. 3e, genes with an average fold change (expressed log-arithmically) > 1.5 are shown. Both Fig. 3e and Extended Data Fig. 5d were plotted with heatmap. 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 monochrome 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.

Twist and c-tail regulatory sequences have previously been described. They were cloned in pCESA expression vector upstream of the reporter gene GFPCAAX (CAAX is the palmitoylation motif to target a protein to the membrane), H2B-mApple, H2B-YFP, mNeonGreen-PH (nG-PH), mCherryCAAX and H2B-mCherry using NotI and Ascl restriction enzymes (NEB). The expression vector with H2B-mApple reporter construct was obtained by inserting mApple26 (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)27 using NotI and FseI (NEB) into a pCESA expression vector and then replacing the GFP coding sequence with mNeonGreen28 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 Ascl 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 fop (pCESA-fop-GFP) using Ascl and XbaI restriction enzymes (NEB). The regulatory sequences of NP (KH.C11.631) were PCR-amplified and cloned into pCESA-fop-GFPCAAX. After PCR amplification from the prop > GFPCAAX (primers in Supplementary Table 7), Prop 900 bp, Prop 700 bp and Prop 300 bp were cloned into pCESA-GFPCAAX vector using Ascl and XbaI. Prop mutations in FoxH1-a binding site of the Prop 260 bp regulatory sequences were obtained by plasmid PCR of Prop 260 bp fop-GFPCAAX (primers in Supplementary Table 7).

Gαδ2 regulatory sequences specifically active in bipolar tail neurons were amplified from Gab2-H2B-mCherry (primers in Supplementary Table 7) and cloned into pCESA-fop-mCherryCAAX using Ascl and XbaI restriction sites. The 2.2 kb Znt3 regulatory sequences were PCR-amplified (primers in Supplementary Table 7) and then assembled into pSP-Kacle expression vector using NEBuilder. Ptf1a was obtained by PCR-amplifying an expression vector that contains the full-length Ptf1a regulatory sequences29 (primers in Supplementary Table 7). The PCR product was self-recombined using NEBuilder. Ptf1a was then subcloned upstream of mCherryCAAX in the pCESA expression vector.

A LacZ expression vector under the control of Dmrt1 (Dmrt1 > LacZ) has previously been described. 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).

Cliona electroporation and imaging. After fertilization, one-cell-stage embryos were electroporated using 20 to 100 µg of each expression construct as previously described. The embryos were raised at 16°C, 18°C or 21°C in ASW and fixed at the desired stage following a previously described protocol. 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 Atempo 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 ASW30, and the imaging was performed at 18°C from the latTI to latTHI stage. The images were assembled using Fiji 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 χ2 test followed by the post hoc test for pairwise comparison, Fisher’s test with Bonferroni adjustment was also performed with Tidyverse. 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 the African killifish N. furzeri was performed according to the guidelines of the Stowers Institute for Medical Research.

A 4-kb Cliona 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. F₂ 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 F₂ embryos showed GFP expression in the forebrain. For the second transgenic line, 8 out of 25 F₂ embryos showed GFP expression in the forebrain. Finally, 10 out 37 F₂ 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 predict determine 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.

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., K.C. and M.L. wrote the manuscript. C.C., L.A.L., K.C. and M.L. contributed to interpretation of the results, and C.C., L.A.L., K.C. and M.L. wrote the supplementary information 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) 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; midTI, 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 transcription 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 shown on seven reconstructed transcriptome trajectories of endoderm (n = 14,162 cells). Scale bars, 50 μm.
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 Casq1/2 fag>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 α-(a), β- (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>GFPCAAX (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>GFPCAAX (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 \( Prop>GFPCAAX \) 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 \( \mu \)m.
Extended Data Fig. 9  |  Manipulation of Eminens gene regulatory network.  

**a.** Pseudotemporal expression profiles of regulatory genes and signalling components in Eminens neurons. A representative embryo is shown for the different fusion genes (GFPCAAX, 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.  

**b.** Diagram of the Prop regulatory sequences with their length indicated on the left. A 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. χ² test with four degrees of freedom was performed (P < 2.2 × 10⁻¹⁶), followed by two-sided Fisher’s exact test with Bonferroni adjustment for multiple comparisons. P values: 900 bp versus 700 bp, 1.05 × 10⁻⁷; 900 bp versus 300 bp, 3.47 × 10⁻¹⁵; 900 bp versus 260 bp, 2.36 × 10⁻¹⁴; 900 bp versus 260-bp FoxH-a mut, 1.81 × 10⁻¹⁰; 700 bp versus 300 bp, 0.011; 700 bp versus 260 bp, 0.36; 700 bp versus 260-bp FoxH-a mut, 0.088; 300 bp versus 260 bp, 5.59 × 10⁻⁶; 300 bp versus 260-bp FoxH-a mut, 0.69; 260 bp versus 260-bp FoxH-a mut, 1.27 × 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.** Numbers of 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.  

Numbers of embryos: 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 × 10⁻⁹. Scale bars, 20 μm.
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 μm.
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.

- 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
- 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

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 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.
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 UMIs 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 gastula 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 |
|-----|-----------------------|
| ☒  | Antibodies            |
| ☒  | Eukaryotic cell lines |
| ☒  | Palaeontology         |
| ☒  | Animals and other organisms |
| ☒  | Human research participants |
| ☒  | Clinical data         |

Methods

| n/a | Involved in the study |
|-----|-----------------------|
| ☒  | ChIP-seq              |
| ☒  | Flow cytometry        |
| ☒  | 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 killfish 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.