Published in final edited form as:

Nature. 2019 July ; 571(7765): 349–354. doi:10.1038/s41586-019-1385-y.

Comprehensive single cell transcriptome lineages of a proto-vertebrate

Chen Cao[1,5], **Laurence A. Lemaire**[1,5], **Wei Wang**[3], **Peter H. Yoon**[2], **Yoolim A. Choi**[2], **Lance R. Parsons**[1], **John C. Matese**[1], **Wei Wang**[1], **Michael Levine**[1,2,*], **Kai Chen**[1,4,*]

[1]Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ, USA

[2]Department of Molecular Biology, Princeton University, Princeton, NJ, USA

[3]Stowers Institute for Medical Research, Kansas City, MO, USA

[4]Current address: The Yunnan Key Laboratory of Primate Biomedical Research, Institute of Primate Translational Medicine, Kunming University of Science and Technology, Kunming, YN, China

[5]These authors contributed equally

Abstract

Ascidian embryos highlight the importance of cell lineages in animal development. As simple proto-vertebrates they also provide insights into the evolutionary origins of novel cell types, such as cranial placodes and neural crest. To build upon these efforts we have determined single cell transcriptomes for more than 90,000 cells spanning the entirety of *Ciona intestinalis* development, from the onset of gastrulation to swimming tadpoles. This represents an average of over 12-fold coverage for every cell at every stage of development, owing to the small cell numbers of ascidian embryos. Single cell transcriptome trajectories were used to construct “virtual” cell lineage maps and provisional gene networks for nearly 40 different neuronal subtypes comprising the larval nervous system. We summarize several applications of these datasets, including annotating the synaptome of swimming tadpoles and tracing the evolutionary origin of novel cell types such as the vertebrate telencephalon.

Single cell RNA sequencing (scRNA-seq) methods are revolutionizing our understanding of how cells are specified to become definitive tissues during development[1–5]. These studies

*Correspondence should be addressed to M.L. (msl2@princeton.edu) and K.C. (chenk@lpbr.cn).

Author contributions

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

The authors declare no competing interests.

Data availability

Raw sequencing data and gene expression matrix are available in 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.

permit the elucidation of “virtual lineages” for select tissues, and also provide detailed expression profiles for interesting cell types such as pluripotent progenitor cells. However, a limitation of the earlier studies is the incomplete coverage of vertebrate embryos due to the large cell numbers.

As the closest living relatives of vertebrates[6], the ascidian, *Ciona intestinalis*, serves a critical role in understanding comparable, but far more complex, developmental and physiological processes in vertebrates. Contrary to vertebrate embryos, ascidian embryos are quite simple. Gastrulating embryos are composed of just 100-200 cells, while swimming tadpoles contain ~2500 cells. Due to these small cell numbers, it is possible to obtain comprehensive coverage of every cell type during development, including rare neuronal subtypes.

Here, we extend the regulatory “blueprint” spanning the early phases of embryogenesis[7] by profiling the transcriptomes of individual cells in sequentially staged *Ciona* embryos, from gastrulation at the 110-cell stage to neurula and larval stage. Reconstructed temporal expression profiles illuminate the specification and differentiation of individual cell types. Nearly 40 neuronal subtypes were identified, even though the central nervous system is composed of just 177 neurons[8]. The resulting high-resolution transcriptome trajectories, regulatory cascades and provisional gene networks provide a variety of insights, including the evolution of novel cell types such as the telencephalon of vertebrates.

CELL FATE SPECIFICATION

Synchronized embryos from 10 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 2 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 to swimming tadpoles when all larval cell types, tissues and organs are formed (Fig. 1b). In total, we profiled 90,579 cells, corresponding 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 ~12K UMIs (Unique Molecular Identifiers), thereby enabling the recovery of rare populations such as germ cells (~0.1% in swimming tadpoles).

t-distributed Stochastic Neighbor Embedding (tSNE) projections of the transcriptomes at all 10 developmental stages identified coherent clusters of individual tissues, including heart, tail muscles, endoderm, notochord, germ cells, epidermis, nervous system, and mesenchyme (Extended Data Fig. 2a–I). Several tissues, such as the nervous system and mesenchyme, exhibit a progressive increase in cell complexity during development (Extended Data Fig. 2c–I), resulting 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 as compared with divergent cell types at the same time points. This is particularly evident for the developing germ line since it is transcriptionally quiescent during the time frame of the analysis[9].

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

Classical cell lineage studies suggest that all of the major tissues of the ascidian tadpole are specified prior to gastrulation, at the 110-cell stage (Fig. 1c)[12]. Most of the internal organs are derived from vegetal lineages, including the notochord, endoderm, tail muscles, heart, germ cells, and regions of the nervous system. In 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, including the notochord, 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. The expression profiles of individual cell types identified known and new potential fate determinants (Extended Data Fig. 3; Supplementary Table 2). For example, *Irx-B* is specifically expressed in a-line (anterior) epidermis, while *Not* is expressed in the b-line (posterior).

RECONSTRUCTING CELL LINEAGES

The alignment of transcriptome profiles of individual cell types at sequential stages of development permitted the reconstruction of “virtual” lineage maps (Fig. 1e; Methods). In total, 60 different cell types were identified in swimming tadpoles, and their corresponding virtual lineages 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, but also provide new insights into the specification and differentiation of individual cell types. For example, the transcriptome profiles accurately capture muscle and heart lineages (Extended Data Fig. 4a, b), as well as the primary (A8) and secondary (B8) lineages of the notochord (Extended Data Fig. 5)[12]. The mesenchyme has been shown to be derived from three separate lineages (A7.6, B7.7, and B8.5)[11], and our analyses suggest they segregate to produce nine different cell types (Extended Data Fig. 4c, d). Similarly, head and trunk endoderm produce seven distinct 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 CNS were found to share a common lineage with the pro-anterior sensory vesicle (pro-aSV), the anterior-most terminus of the neural tube that fuses with the stomodeum to form the neuropore[14]. Both derivatives share a common origin with palp sensory cells (PSCs), which arise from the non-neural proto-placodal territory located immediately anterior of the neural tube (Fig. 1e). This observation

is consistent with a specific model for the evolution of the vertebrate telencephalon, as discussed below.

TRANSITIONAL PROPERTIES OF NOTOCHORD

The notochord is a derivative of the mesoderm, and a defining innovation of chordates[15]. However, it exhibits distinctive properties in cephalochordates and vertebrates. Cephalochordates such as *Amphioxus* contain a muscular notochord that helps power movements of the tail[16], while 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 different subclusters throughout development (Extended Data Fig. 5a). By constructing single cell trajectories, it was possible to identify cell signaling 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], it was possible 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, c, d* (Extended Data Fig. 3)[19]. They precede expression of muscle identity genes such as Calsequestrin (*Casq1/2*), myosin (*Mlra/Mlrv/Myl5*), and tropomyosin (*Tpm1/2/3*) (Extended Data Fig. 5d; Supplementary Table 2). None of these genes are expressed in the primary notochord[20]. Moreover, the 5' regulatory regions of these genes contain clusters of Tbx6 binding motifs (Supplementary Table 3), suggesting 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 primary, notochord. These distinctive developmental programs suggest that *Ciona* possesses both properties of the notochords seen in cephalochordates and vertebrates.

IDENTIFICATION OF INDIVIDUAL NEURONS

The central nervous system of swimming tadpoles is composed of only 177 neurons[8], thereby permitting 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 ~7-fold 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). They map in 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 different neural subtypes (Fig. 2; Extended Data Fig. 6e; Supplementary Table 2). For example, coronet cells are the only dopaminergic neurons in the *Ciona* CNS. They express high levels of *Ptf1a* and *Meis*, which are sufficient to reprogram the CNS into supernumerary coronet cells[21]. It is possible that other

combinations of cell-specific transcription factors specify additional neural subtypes (e.g., *Bsh*, *Lhx2/9*, and *Aristaless* in the anterior sensory vesicle).

The high coverage of individual transcriptomes permitted the identification of scarce neuronal subtypes (Extended Data Fig. 7). For example, there are only two pairs of BTNs (bipolar tail neurons) in swimming tadpoles[22], and these were found to express galanin and two of its receptors (*Gal1r* and *Gal2r*). Galanin has been implicated in neuro-regeneration and axogenesis[23,24]. A reporter gene containing *Galr2* regulatory sequences mediates restricted expression in the BTNs (Extended Data Fig. 7a). Similarly, a pair of decussating neurons (ddNs), which play a central role in the startle response of vertebrates[25,26], was unambiguously identified based on their selective expression of the homeobox gene Dmbx27 (Extended Data Fig. 7b).

Additional specific neuronal subtypes were identified on the basis of restricted expression of select marker genes. Ci-*VP*$^{+}$ pSV neurons express several secreted neuropeptides (Supplementary Table 5), including vasopressin/oxytocin and an uncharacterized neuropeptide, *NP*28. *NP* expression is restricted to two small groups of neurons located in the posterior-most regions of the sensory vesicle (Extended Data Fig. 7c, e–f). A pair of Eminens neurons (Ems) was identified by expression of a reporter gene containing *Prop* 5' regulatory sequences (Extended Data Fig. 7d). Altogether, these studies document the feasibility of identifying the transcriptome trajectories and virtual lineages of individual defined neurons, including rare subtypes such as Ems.

TRANSCRIPTOME AND SYNAPTOME INTEGRATION

The recently reported *Ciona* synaptome identified a single Eminens neuron, Em2, as a key regulator of the ddNs[8,25]. The pair of Ems was identified in our datasets on the basis of expressing GABAergic marker genes and *Prop* (Extended Data Fig. 8a, b). Moreover, reporter genes containing *Prop* regulatory sequences are selectively expressed in a pair of neurons that display all the properties of Ems, including morphology and location (Extended Data Fig. 7d)[22,25,29–31]. The transcriptome trajectories of Ems neurons suggest that they arise from the a-lineage (Fig. 2), even though they are located in posterior regions of the sensory vesicle. This apparent discrepancy was resolved by live cell imaging. We found that Ems undergoes long-range migration from forebrain to posterior regions of the sensory vesicle (Fig. 3a; Supplementary Video 1). These movements correlate with the expression of a variety of genes implicated in migration and axogenesis, including *Nav2* and *Trim9* (Fig. 3b)[32,33].

Regulatory cascades of cell signaling components and transcription factors permitted the formulation of a provisional gene regulatory network for the Ems 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 involved in neuronal function, including neuropeptide receptors (*Oxyr*, *Glpr*, *Galr2*), zinc neuromodulation (e.g. *Znt3* and *S39aa*), and GABAergic markers (e.g., *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 one of the predicted upstream regulators of *Prop*, *FoxH-a*, caused a significant 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 Ems and ectopic activation of downstream reporter genes (e.g., *S39aa*) (Fig. 3d; Extended Data Fig. 9d, e).

We next sought to leverage this information to gain insights into the nature of the neuronal interactions underlying the startle response (Fig. 3e). A centerpiece of the startle circuit is the pair of ddNs, which correspond to the Mauthner neurons in the brain stem of fish and frogs[26]. They integrate a variety of sensory information to trigger a fast escape reflex. As predicted by previous studies[25,31], interactions between Em2 and the ddNs (Fig. 3f) are probably inhibitory since Ems express GABAergic markers such as *vGat* and *Gad* (Extended Data Fig. 6), while the ddNs express GABA receptors (Fig. 3e; Supplementary Table 2). The ddNs also express glutamate receptors (Fig. 3e), suggesting response 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 Ems and ddNs express receptors for galanin, which is expressed in the BTNs (Fig. 3g, Extended Data Fig. 8h). The BTNs have been likened to the dorsal root ganglia derivatives of the neural crest in vertebrates[22]. Galanin promotes survival of dorsal root ganglia neurons during development and after injury. It is possible that it serves as a tropic factor for Em2 since the BTNs directly interact with the cell body of this neuron (Fig. 3g). Moreover, modulation of Em2 by additional neuropeptides is suggested by its expression of a vasopressin/oxytocin receptor. As shown above, Ci-*VP*$^{+}$ cells express a number of secreted neuropeptides including vasopressin/oxytocin (Extended Data Fig. 7e). They are in close proximity with Em2 (Extended Data Fig. 7f).

Altogether, the transcriptome datasets provide substantive annotations of the neuronal circuits described by recent synaptome studies[8,29]. They suggest both targeted growth and feedback inhibition of the startle response by BTNs, and implicate neuropeptides such as galanin and oxytocin/vasopressin as potential modulators of the circuit, in addition to canonical neurotransmitters.

EVOLUTION OF NOVEL CELL TYPES

Previous evo-devo studies suggest that *Ciona* possesses rudiments of key vertebrate innovations such as neural crest, cranial placodes and cardio-pharyngeal mesoderm[22,34–37]. However, the evolutionary origin of the telencephalon, arising from the anterior-most regions of the forebrain, remains uncertain. The telencephalon contains the olfactory bulb and is also the source of higher order brain functions such as the neocortex of humans. Forebrain regions of the *Ciona* CNS 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 (PSCs) and the pro-aSV (Extended Data Figs. 10–12; Methods). The PSCs--axial columnar

cells[38]--express a cascade of cell signaling components and regulatory genes, including *FoxC*, *Dlx*, *FoxG*, *Isl* and *SP8* (Extended Data Fig. 12a, c; Supplementary Table 2). A similar regulatory cascade has been implicated in the specification of the telencephalon in vertebrates[39,40].

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

We propose that the vertebrate telencephalon arose by 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 larvae and transgenic killifish 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 placodal derivatives such as the lens of the eyes (Extended Data Fig. 12e). These observations are consistent with the incorporation of proto-placodal gene regulatory modules (e.g., 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 different cell types (including nearly 40 neuronal subtypes) comprising the *Ciona* tadpole. These datasets significantly extend classical lineage maps and regulatory blueprints, and provide a rich 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 controlling behavior. They also provide new insights into the evolutionary transition of invertebrates, including the dual properties of the *Ciona* notochord and the expansion of the vertebrate forebrain. This is an exciting era of single cell studies that encompasses a broad spectrum of cell types and systems[1–4,43]. The resulting databases provide new opportunities to trace the evolutionary origins of every cell, tissue and organ in the human body.

METHODS

Ciona handling, collection and dissociation of embryos

The adults of *Ciona intestinalis* were purchased from M-Rep, San Diego, CA. The eggs and sperm was obtained as described in Christiaen et al, 2009[44]. Sperm was added to the eggs for 10 minutes (min). Then the fertilized eggs were washed with filtered sea water twice. Except in the case of the larva stage, the same animal provide 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 Hotta et al 2007[45]. At least 2 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 Ca^{2+}-free artificial sea water (Ca^{2+}-free ASW, 10 mM KCl, 40 mM $MgCl_2$, 15 mM $MgSO_4$, 435 mM NaCl, 2.5 mM $NaHCO_3$, 7 mM Tris base, 13 mM Tris-HCl). Embryos were immediately dissociated with 0.5 to 1% trypsin in Ca^{2+}-free ASW with 5 mM EGTA (ASW-EGTA) for 2 min (gastrula and neurula stages) to 10 min (tailbud stages). Embryos were pipetted 5 min on ice to complete dissociation of individual cells. Then, the digestion was inhibited with 0.2% BSA in Ca^{2+}-free ASW or quenched by 20% FBS. Cells were collected by centrifugation at 4°C at 500 g for 2 to 5 min and then resuspended in ice-cold Ca^{2+}-free ASW containing 0.5% BSA. For the swimming larva stage, the embryos were either homogenized (H100, Waverly) and dissociated using 1% trypsin or dissociated with 1% trypsin, 1mg/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 previously described.

Single cell barcoding, on-chip and off-chip technical replicates, library preparation and sequencing

Cell concentration of each sample was checked by TC20™ Automated Cell Counter to ensure within 1000-2000 cell/μl. Single cell suspensions were loaded onto The Chromium Controller (10x Genomics, CA). To assess technical variations between replicates, on-chip and off-chip experiments were performed. The on-chip experiment consisted to load in two different lanes of cells from LTBII embryos on the same chip, obtained by fertilizing eggs with the sperm from the same animal. In the off-chip experiment, dissociated cells from LTBI embryos obtained with the same fertilization strategy, were loaded on the same lane on two different chips and processed separately. For all the samples, cells were lysed, cDNAs were barcoded and amplified with Chromium Single Cell 3' Library and Gel Bead Kit v2 (10x Genomics, CA) following the instructions of the manufacturer. Illumina sequencing libraries were prepared from the cDNA samples using the Nextera DNA library prep kit (Illumina, CA). All the libraries were sequenced on Illumina HiSeq 2500 Rapid flowcells (Illumina Inc., CA) with paired-end 26nt + 125nt reads following standard Illumina protocols.

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. The reference sequence was obtained from the Ghost database[46].

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 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. We further normalized the read counts of each cell by Seurat methods[47], 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 1000 highest standard deviations were obtained as highly variable genes. After significant principal components (PCs) were identified, a graph-based clustering approach was used for partitioning the cellular distance matrix into clusters. Cell distance was visualized by t-Distributed Stochastic Neighbor Embedding (t-SNE) method in reduced 2D 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 the samples were kept for canonical correlation analysis (CCA). Top 50 canonical correlation vectors (CCs) were calculated, and each dataset was projected into the maximally correlated subspaces. According to the relationship between the number of CCs and the percentage of the variance explained, 1-18 CCs were used for subspace alignment[47] and dimensional reduction. Graph-based clustering was performed in the lower-dimensional space, and an approximate nearest neighbor search was performed. We employed 10 random starts for clustering and selected the result with highest modularity. A modified Fast Fourier Transform-accelerated Interpolation-based t-SNE[48] was used in the visualization of our dataset. Max iteration times was set to 2000 and the perplexity parameter was set to 30.

For t-SNE visualization of cells in each tissue type from different developmental stages, we employed distinct strategy. As we focused on identification of subpopulation cells within the same tissue type across time, whereas CCA and subspace alignment focused on shared similarities between samples, we instead regressed out the effects produced by the UMI counts, experiments batch and sample identities with negative binomial regression modelling before dimensional reduction and clustering. Similar with approaches mentioned earlier, after significant PCs were identified (1-20 PCs were used for subclustering of cell 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 as 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_insitu) database 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 their gene expression pattern to make sure there was no ambiguous expression of known markers, which might indicate the cell doublets. We also consulted with experts in ascidians

community to validate our findings. These novel cell types were named with their specifically expressed genes, e.g. *TTI1*$^{+}$ cells in mesenchyme. If cells' position, morphology are verified by our reporter assays, with transcriptomic characteristics, we compare them to published paper with morphological information (e.g., the synaptome paper) and identify the cell types (e.g. Eminens neurons, Ems).

Cell state mapping across time points

As mentioned in the main text, most of the major tissues of ascidian tadpole are first specified prior to gastrulation, which is our start time point of sampling. 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[1]. In brief, between every two adjacent time points, all the cells were embedded to the PCA space of the late time point (1–50 PCs were used). For each cluster identified in the late time point, each cell in the cluster and its nearest 5 neighbor cells in the previous time point were calculated. The voting results for all the cells in each cluster of the late time point were aggregated, and the percentage of "ancestor cells" in each cluster of the early time point was calculated. In the case that one cluster had more than one ancestor cluster, the cluster in early time point with a percentage number at least twice higher than the other clusters was considered as the winning ancestor.

For highly differentiated tissue types, to better capture all the sub-populations, we sub-clustered the cells from each stages and did the ancestor voting process between sub-clusters across time points. For mesenchyme cells, most winning ancestors were unambiguous assigned, with > 90% winning share on average. For epidermis and nervous system systems, as some sensory neurons differentiate from the epidermis, we sub-clustered the epidermis and nervous systems cells together. We deleted the sub-clusters if they were assigned to multiple ancestor clusters and there was no winning ancestor (no winning cluster met our criteria: percentage number 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 Ems, we employed monocle 2[49] to construct the single cell trajectory for cleanly defined lineage with known markers. Highly variable genes with q value less than 0.01 were selected across time points, 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 value less than $1e^{-100}$ and $1e^{-20}$ were shown in the pseudotemporal expression pattern of primary and secondary notochord in Extended Data Fig. 5, respectively.

For tissues that harbored more complexity during development, such as the mesenchyme and nervous system, which had 9 and 41 identified sub-clusters at larva stage, respectively, we employed a simulated diffusion-based computational reconstruction method, URD[2], 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 10 developmental stages as described earlier, we calculated transition probabilities between cells with the destiny package[50]. We next assigned 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 last stage we covered. The walk simulated through cells based on the transition probabilities and the transitions were only allowed to 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 same cells were visited from each tip.

For cells in the mesenchyme, we optimized the number of nearest neighbors (knn) and set it to 250, and the width of the Gaussian used to transform cell-cell distances into transition probabilities (sigma) was set to 6. 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. For cells in the nervous system, to avoid ambiguities in reconstructing gene expression lineages, we excluded or combined those cell clusters that 1) were not well defined or determined during neurogenesis based on prior knowledge; 2) could not be resolved by diffusion components, such as very small population of cells *e.g.* ddNs; 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 was recovered followed the criteria set in the URD package: cells in the segment were compared in a pairwise manner with their 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 1.5x higher than in the sibling branch, and the genes were 1.25x better than a random classifier for the population. Then an impulse model was fit 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 expression pattern that was not fitted with the impulse model were arranged at the bottom of the cascade. In the heatmap, 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, signaling pathway genes recovered in each developmental trajectory cascade, and representative highly variable function genes of specific cell types of the last time point with log2 fold change above 1 between groups were used in investigating the

putative direct interaction. We employed cluster-buster[51] to find clusters of pre-specified motifs in 5kbp upstream of the TSS of each gene. The parameters was set as follows: g = 1, m = 0, c = 0, score >= 6. The position frequency matrix (PFM) was downloaded from JASPAR 2018 database[52]. Genes with no PFM recorded in JASPAR was not considered in constructing the regulatory network. The regulatory network were plotted with Biotapestry. Each line between every two genes represents putative direct interaction as the binding motif of regulatory gene was identified in the motif clusters region of the target gene.

Heatmap

Heatmaps in Extended Data Fig. 3 was plotted with DoHeatmap function of Seurat v2.3.2. Only genes with average fold change (log) >0.3 were shown. For Extended Data Fig. 5d, differentially expressed genes (DEGs) between different primary notochord and secondary notochord were identified by the following criteria using the DESeq2[53]: 1) FDR adjusted p value below 0.05; 2) absolute log2 fold change between groups were larger than 1.5. The mean expression level of each gene within one developmental stage was calculated, and the genes scaled expression was clustered based on the Euclidean distance with pheatmap 1.0.10. For Fig. 3e, genes with average fold change (log) > 1.5 were shown. Both Fig. 3e and Extended Data Fig. 5d were plotted with pheatmap. The pseudo-temporal expression heatmaps in Extended Data Fig. 5b–c and Extended Data Fig. 9a, and the expression dynamics in Fig. 3b was plotted with monocle2.

Molecular Cloning

The KH number of all the genes mentioned in the manuscript as well as their official gene name if not used can be found in the Supplementary Table 6.

Dmbx, Dmrt1, Gad, Prop, *Twist* and *vGat*, regulatory sequences have been previously described[27,31,34,54,55]. They were cloned in *pCESA* expression vector upstream of the reporter genes *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 AscI restriction enzymes (NEB, England). The expression vector with *H2B:mApple* reporter construct was obtained by inserting *mApple*56 (primers in Supplementary Table 7) into the expression vectors *pCESA* containing *H2B* using NEBuilder *(NEB, England)*. The expression vector containing *nG:PH* reporter gene was obtained by first inserting *GFP:PH* (courtesy of T. Meyer)[57] using NotI and FseI *(NEB, England)* into a *pCESA* expression vector and then replacing the *GFP* coding sequence by *nG58*by recombination using NEBuilder (primers in Supplementary Table 7).

Asic1b, Calm, *Fgf13*, *Galr2, S39aa, S39aa 2.2kb*, 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 containing *GFP* downstream of the minimal promoter of *fog* (*pCESA-fog>GFP*) using AscI and XbaI restriction enzymes (NEB, England). The regulatory sequences of *NP* (*KH.C11.631*) were PCR amplified and cloned into *pCESA-fog>GFPCAAX.*

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

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

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

Galr2-BTN was amplified from *Galr2>H2BmCherry* (primers in Supplementary Table 7) and cloned into *pCESA-fog>mCherryCAAX* using AscI and XbaI restriction sites.

Tll1, *Hlx, 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 containing the full length *Ptf1a* regulatory sequences[21] (primers in Supplementary Table 7). The PCR product was self-recombined using NEBuilder. *Ptf1a* was then subcloned upstream of *mCherryCAAX* in *pCESA* expression vector.

LacZ expression vector under the control of *Dmrt1* (*Dmrt1>LacZ*) has been previously described[54]. *Prop* coding sequence was amplified from mid tailbud embryo cDNA and cloned downstream of *Dmrt1* regulatory sequences using NotI and FseI restriction enzymes (NEB, England).

Ciona electroporation and imaging

After fertilization, one cell-stage embryos were electroporated using 20 to 100 μg of each expression construct as described in Corbo et al, 1997[10].

The embryo were rise at 16, 18 or 21°C in ASW; fixed at the desired stage following the protocol described in Wagner and Levine 2012[54]. The embryo were washed several times with 0.05% BSA in PBS before being mount 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 were performed in duplicate or triplicate (see related figure legends). Between 18 and 610 embryos were recovered per condition. No specific randomization strategy was performed except the assignment of the fertilized eggs to the different conditions.

Live imaging was performed on a home-build two-photon microscope system. Embryos were anesthetized with 16mg/ml MS-222 in ASW (Sigma-Aldrich). They were placed in

microwells casted in 1% agarose in ASW[59] and the imaging was performed at 18 degree from LTBII to LTBIII stage. The images were assembled using Fiji[60] and the final rendering obtained with Imaris (Bitplane, Switzerland).

Statistical analysis of the functional assays

For the statistical tests, the embryos with the same electroporated plasmids were pooled over the different experiments. Mann Withney test was performed with the package Tidyverse of R software, the Chi-square test followed by the post-hoc tests for pairwise comparison, Fisher test with Bonferroni adjustment was also performed with Tidyverse[61].

Fish husbandry, generation of transgenic animal and imaging

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

A 4kb *Ciona FoxG* regulatory sequences were cloned into pDest-Tol2-miniP-GFP-Cryaa-Venus transgenic vector through Gibson assembly. In order 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 the Yamamoto embryo solution (17 mM NaCl, 2.7 mM KCl, 2.5 mM $CaCl_2$, 0.02 mM NaHCO3, pH 7.3) at 28 °C for two 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 show GFP expression in the forebrain. For the second transgenic line, 8 out of 25 F1 embryos show GFP expression in the forebrain. Finally 10 out 37 F1 embryos of the third transgenic line had GFP expression in the forebrain. No particular blinding or randomization strategy was implemented.

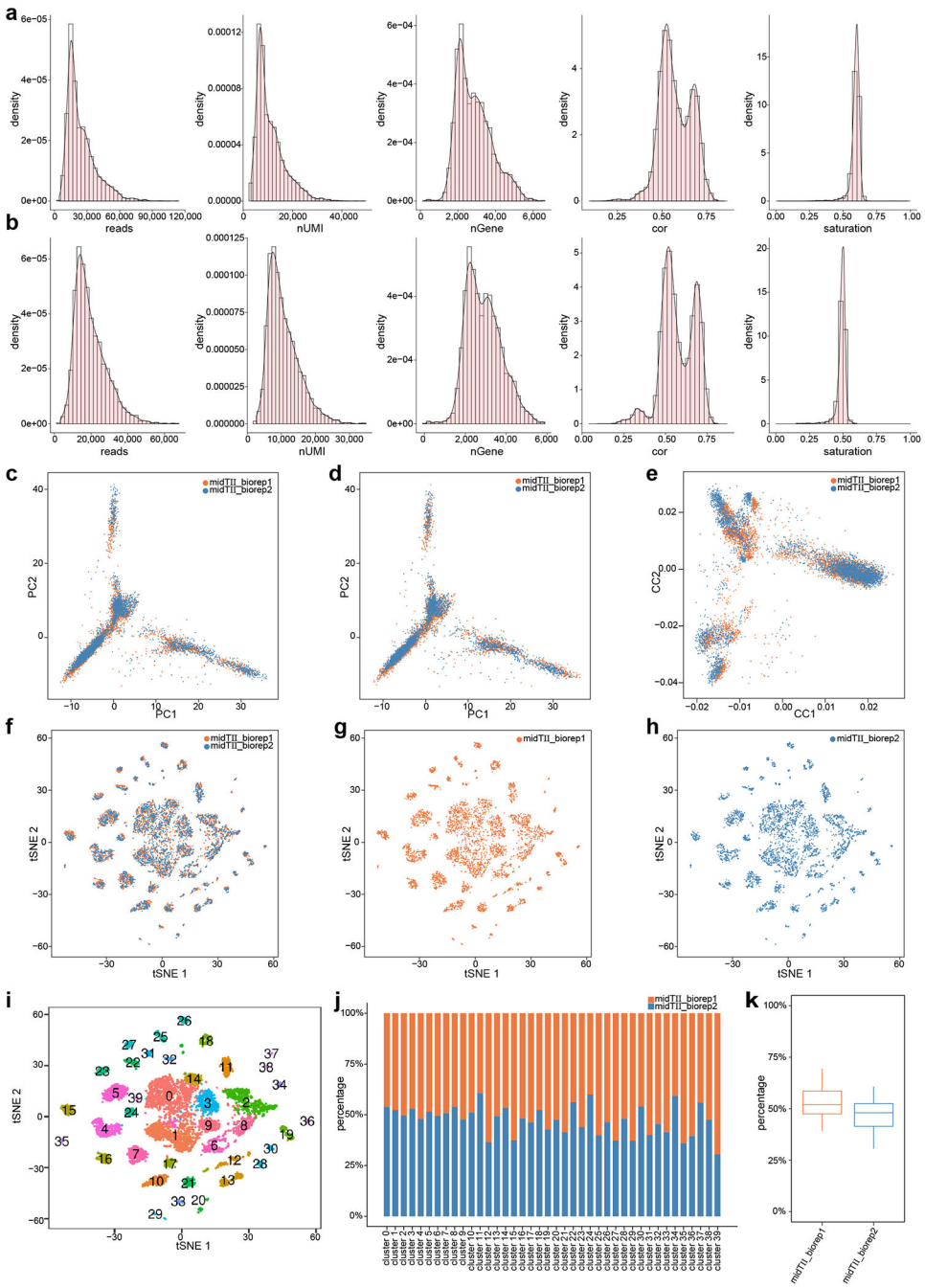
Killifish embryos were removed manually from the chorion before imaging. The juvenile fish were anesthetized in 150 mg/l MS-222 for 5min at room temperature. Images were taken with Ultraview R2 spinning disk confocal microscope.

Blinding

For the single cell experiments, since 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.

Supplementary Information is available in the online version of the article.

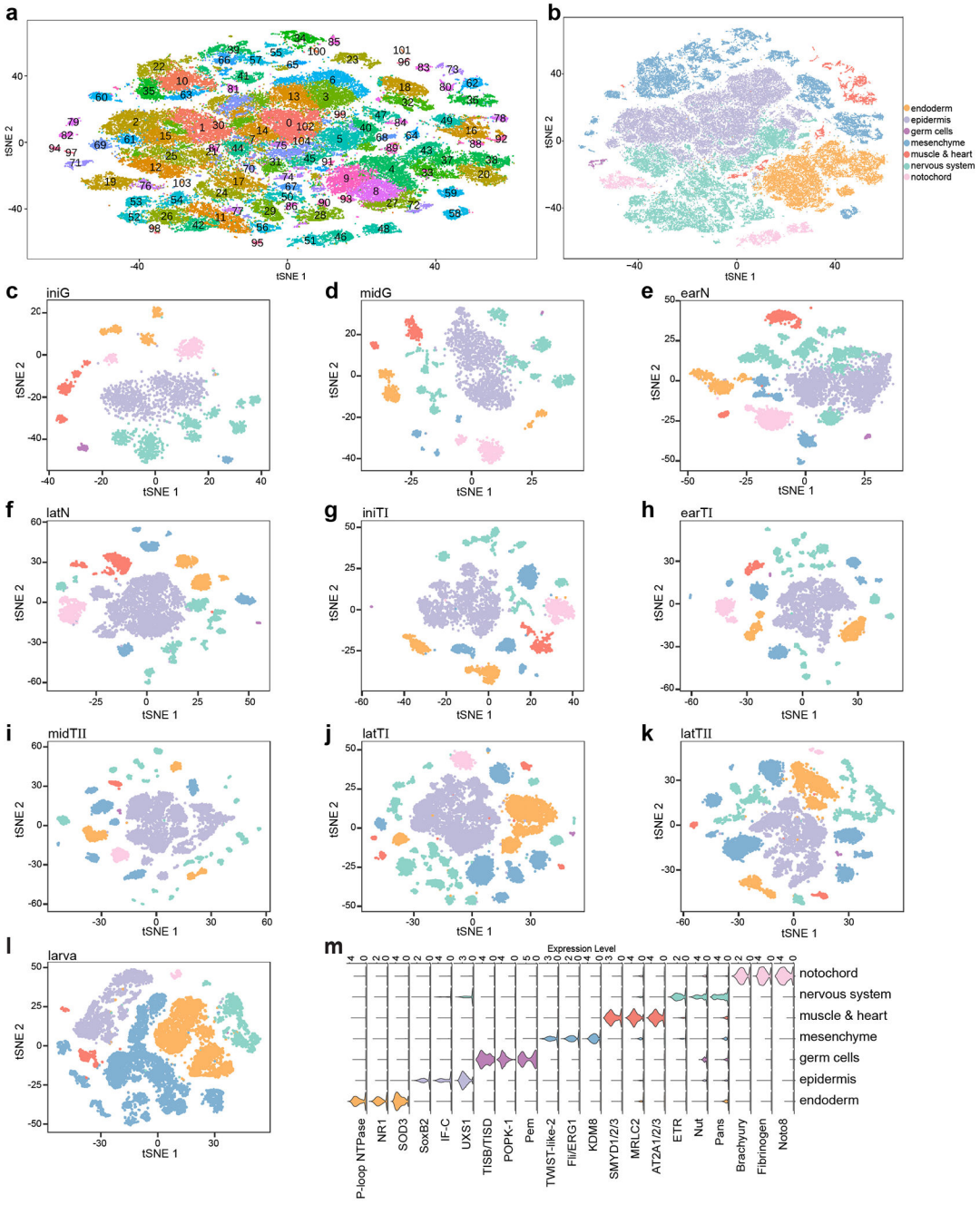
Extended Data



Extended Data Fig. 1. Data quality and biological replicates from middle tailbud stage.

(a, b) Distribution plot of reads number, UMIs, gene number, correlation coefficient (spearman) and saturation level per cell from middle tailbud (**a**, midTII_biorep1; **b**, midTII_biorep2) **(c)** First 2 pcs were plotted for cells regressed by UMIs (midTII_biorep1: n = 4,929 cells; midTII_biorep2: n = 4,062 cells). **(d)** First 2 pcs were plotted for cells regressed by both UMIs and batches. **(e)** First 2 canonical correlation vectors were plotted after cca alignment. Merged **(f)** and split **(g, h)** tSNE clustering for the biological replicates. **(i)** tSNE plot of cca aligned samples of biological replicates (n = 8,991 cells). The number

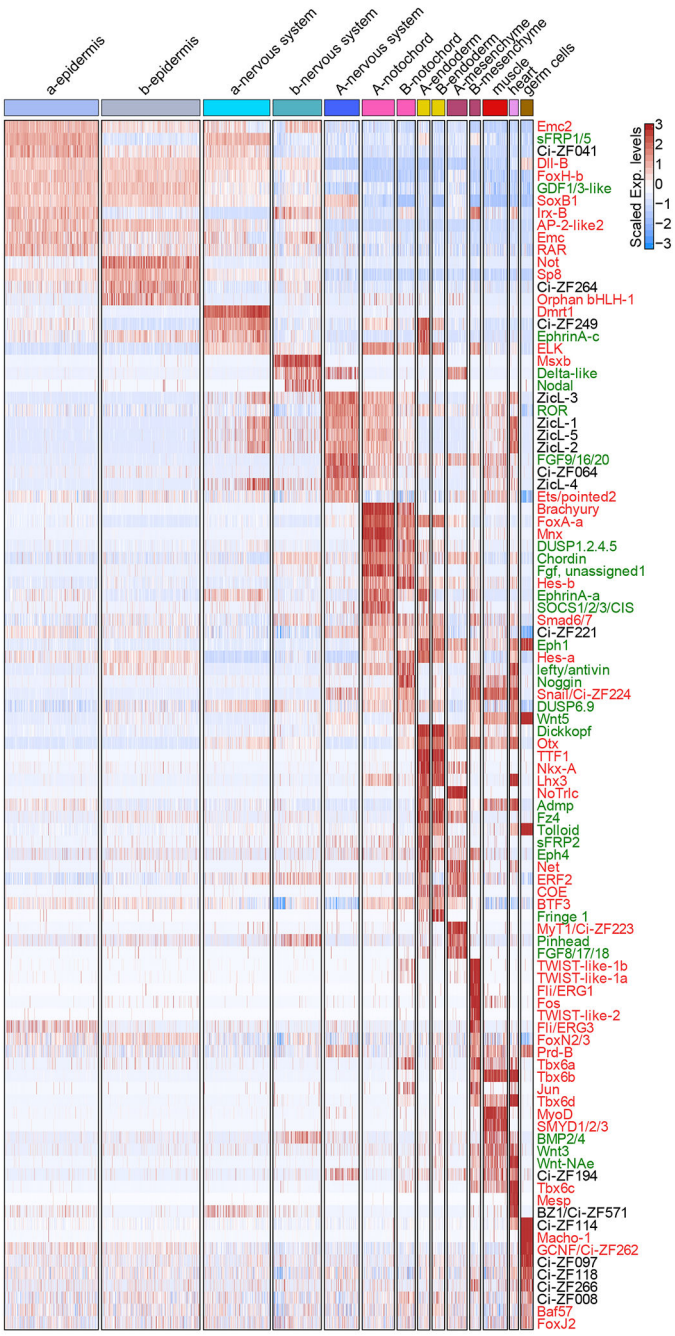
indicates different clusters. **(j)** The percentage of cells between replicates within the same cluster shown in (i). **(k)** Boxplot of percentage of cells in each cluster (n = 40 clusters) between replicates. The lower, middle and upper hinges correspond to the first and third quartiles (the 25th and 75th percentiles), and the middle hinge corresponds to median.



Extended Data Fig. 2. tSNE projections of 10 stages scRNA-Seq data.

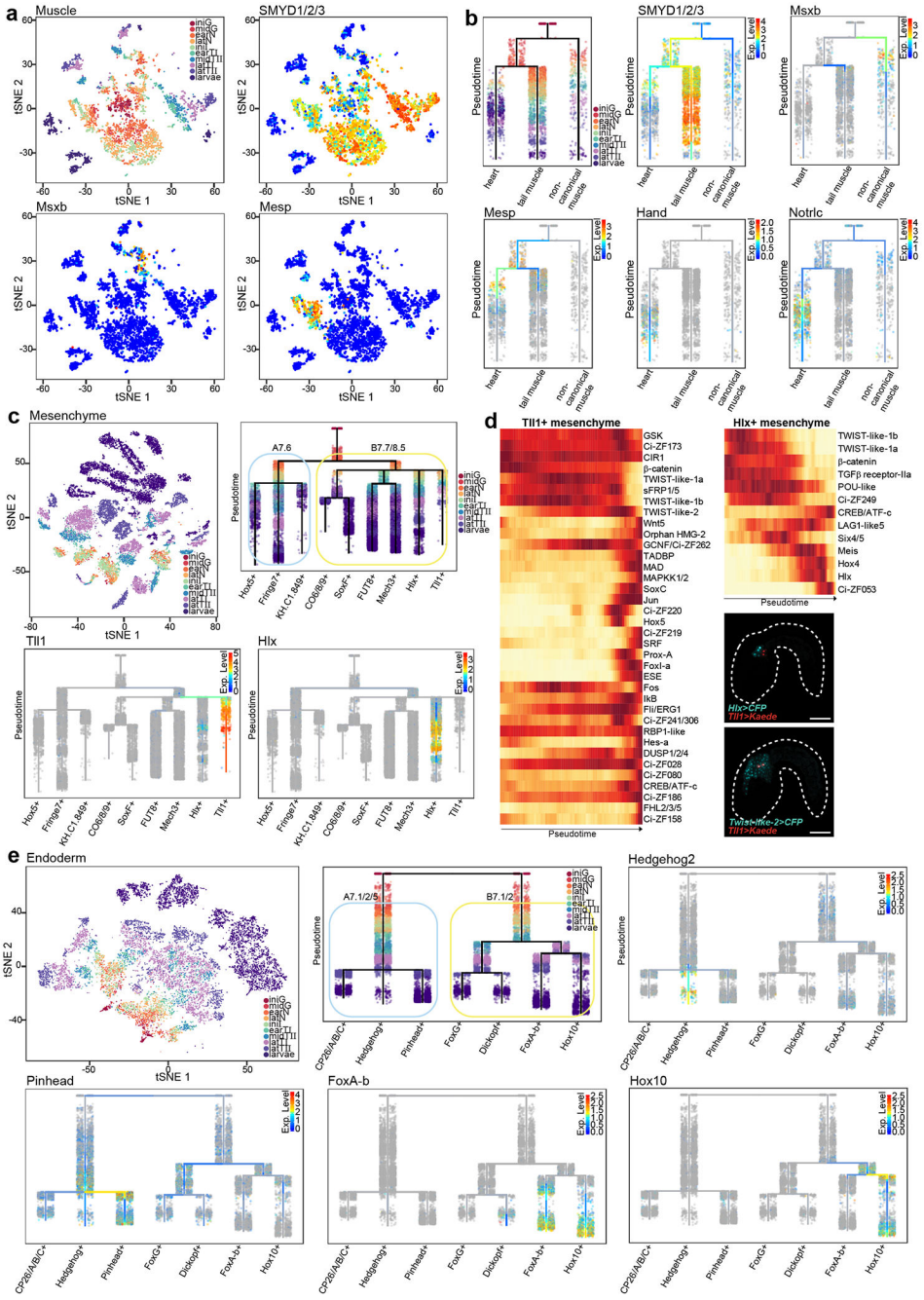
(a) t-SNE plot of the entire dataset (n = 90,579 cells). Cells were colored and labelled by clusters. Differentially expressed genes in each cluster can be found in Supplementary Table 2. **(b)** t-SNE plot of all the cells colored according to tissue type. **(c-l)** tSNE projections of cells colored by tissue types in different developmental stages (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, larva: n = 15,791 cells). Color 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 & heart: n = 3,691 cells, nervous system: n = 22,198 cells, notochord: n = 4,053 cells). Color code is the same as in (b).



Extended Data Fig. 3. Cell type specification at the onset of gastrulation.

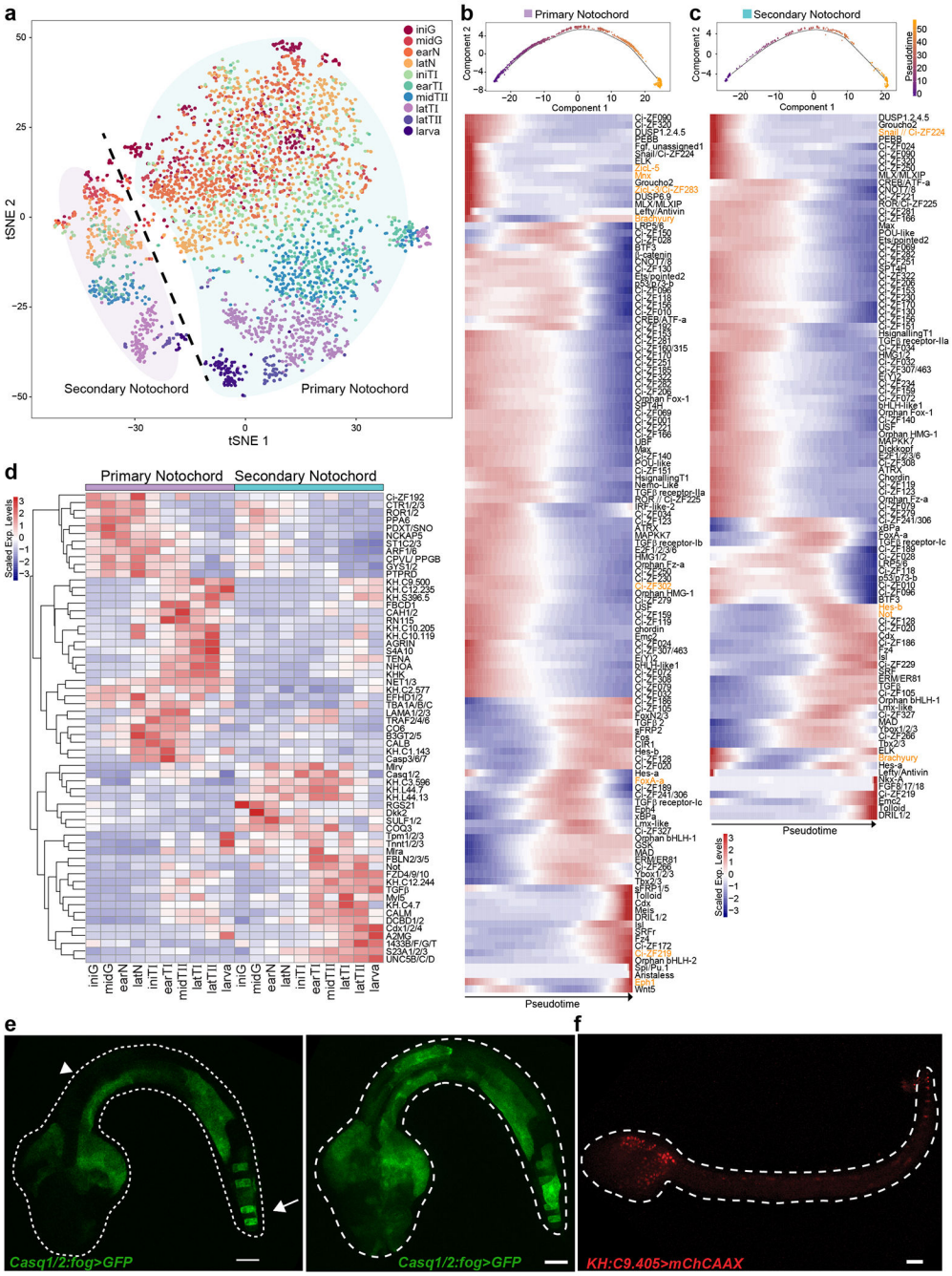
Heatmap showing scaled expression of differentially expressed genes encoding transcription factors (red) and cell signaling components (green). Many new marker genes were identified for each tissue.



Extended Data Fig. 4. Reconstructed transcriptional trajectories of muscle, mesenchyme and endoderm.

(a) tSNE 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)** tSNE 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 signaling pathway genes along pseudotime in *Tll1*$^{+}$ and *Hlx*$^{+}$

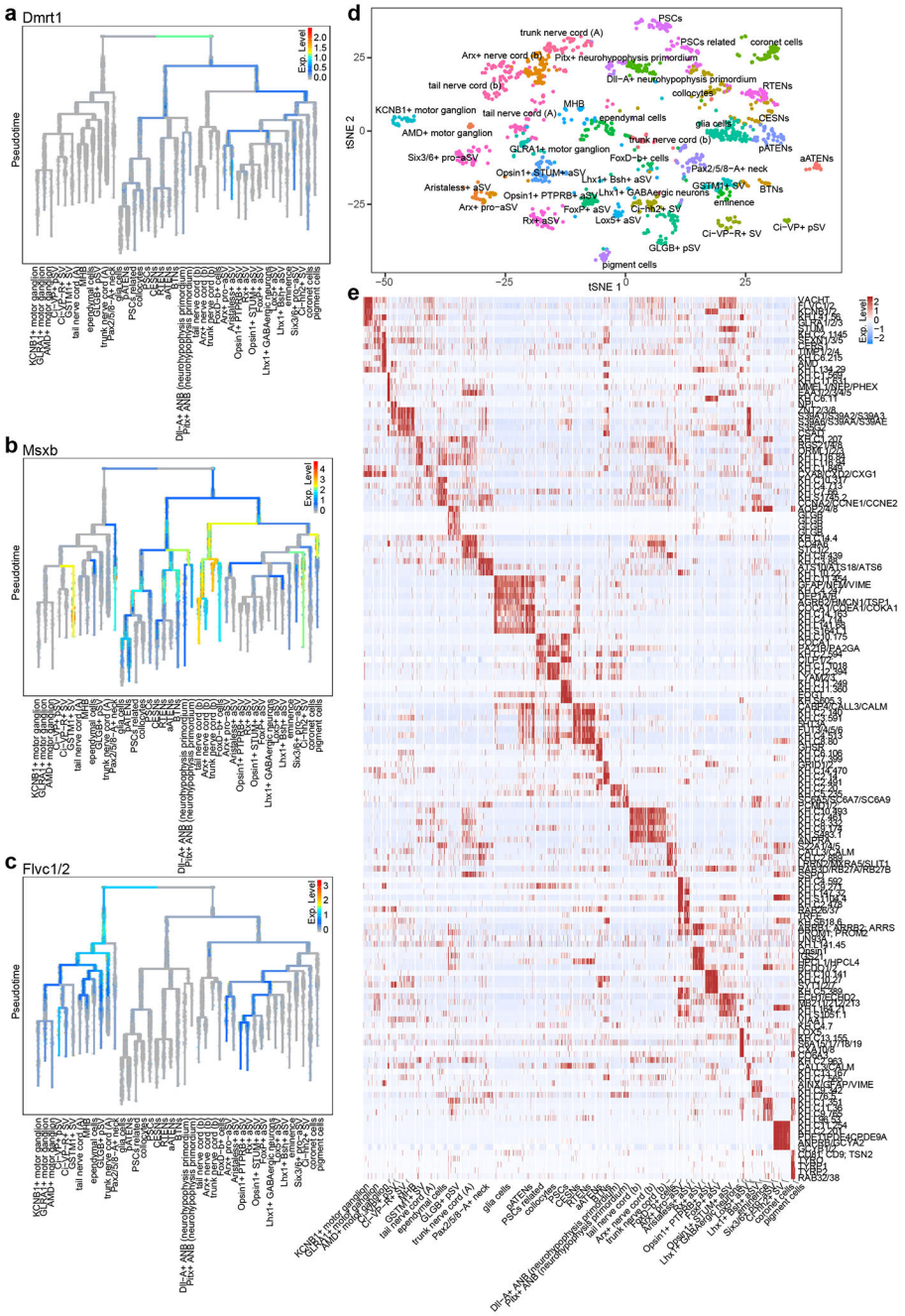
mesenchyme. Lower-right panel shows mid tailbud embryos expressing a Twist-like-2 (cyan), a mesenchymal marker, and *Tll1* (red) reporter gene, and an *Hlx* (cyan) and *Tll1* (red) reporter gene (n= 3 electroporations). **(e)** tSNE projection and expression patterns of representative marker genes shown on reconstructed 7 transcriptome trajectories of endoderm (n = 14,162 cells). Scale bar: 50 μm.



Extended Data Fig. 5. Transcriptome profiles of Ciona notochord cells during development.

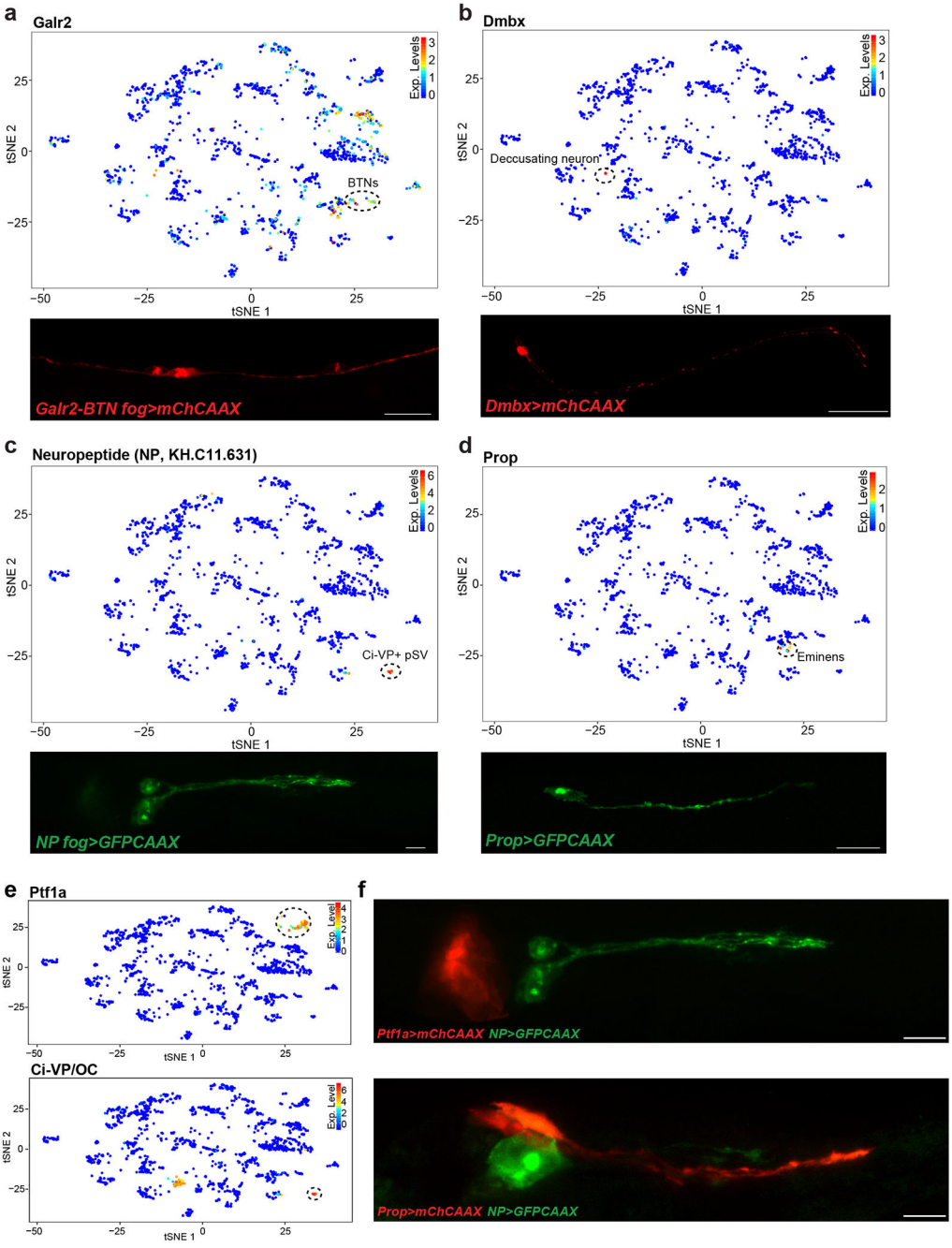
(a) t-SNE plot of notochord cells. Cells were colored by developmental stage (key shown in upper right). The dash line shows the separation between the primary (n = 3,123 cells) and secondary lineages (n = 627 cells). **(b, c)** The single cell transcriptome trajectory (top) and pseudo-temporal gene expression profiles (bottom) of the primary notochord and the secondary notochord. Cells were ordered along the trajectory across pseudotime. Only significant expressed genes (likelihood ratio test) with q-value less than $1e^{-100}$ (primary) and $1e^{-20}$ (secondary) are shown. Selected transcription factors and signaling molecules are

labeled in orange. **(d)** Heatmap of differentially expressed genes between the primary and secondary notochord. Genes were clustered by Euclidean distance. **(e)** Expression of a *Casq1/2>GFP* reporter gene in late tailbud stage embryo (left panel, one optical plane; right panel, maximum intensity projection, n = 3 electroporations). GFP (green) was present in the muscle and in the secondary notochord (arrow), and no expression was observed in the primary notochord (arrowhead).**(f)** Expression of a *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 electroporations). Scale bar: 20 μm.



Extended Data Fig. 6. Neural cells.

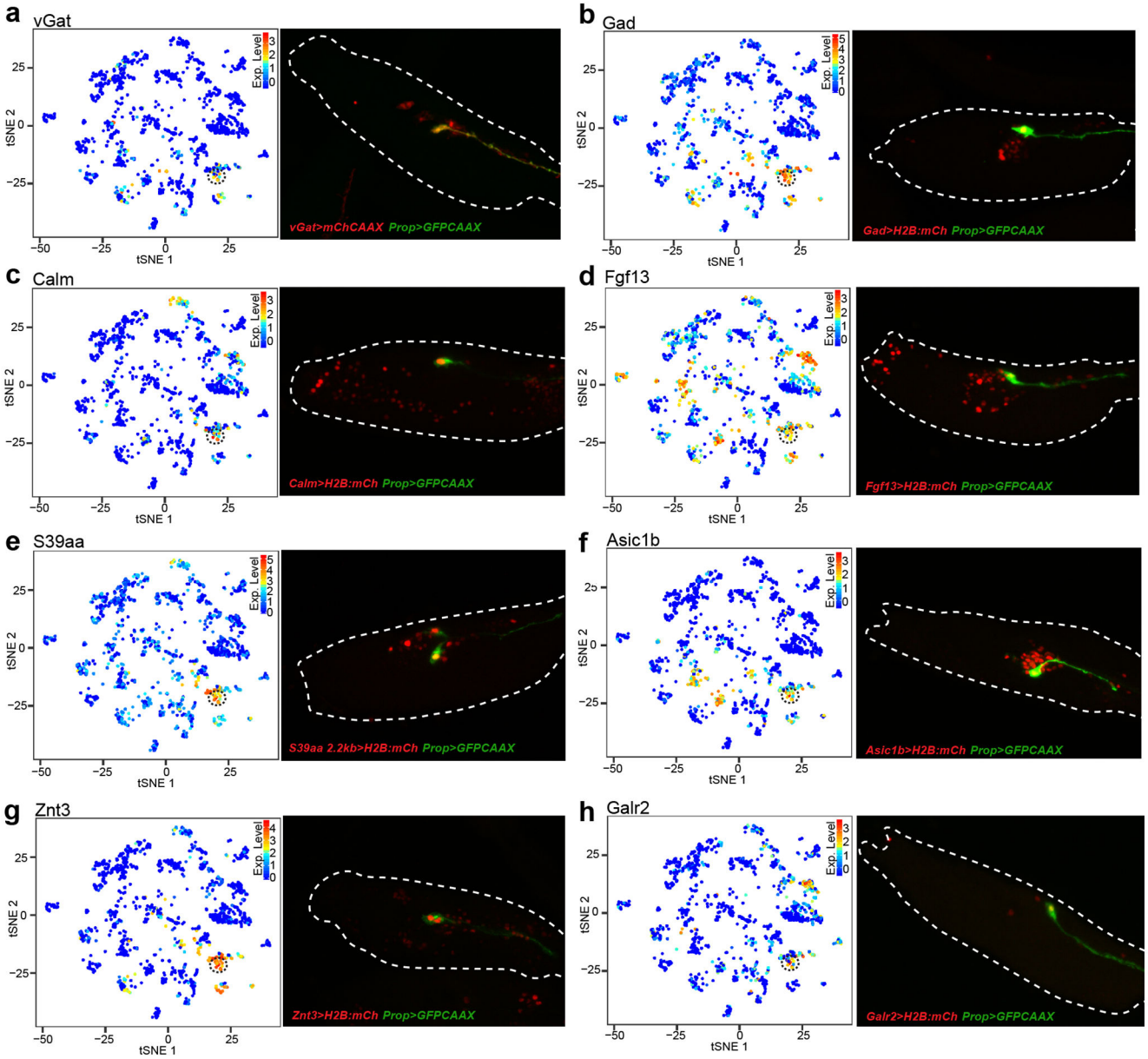
Expression patterns of representative marker genes for a- **(a)**, b- **(b)** and A-lineage **(c)** were shown in reconstructed transcriptome trajectories of neural cells spanning 10 developmental stages. **(d)** t-SNE plot of neural cells recovered from larva stage (n = 1,704 cells). Identified cell types are labelled. **(e)** Heatmap of top 5 differentially expressed genes (no transcription factors) for each type of neural cells in larva stage.



Extended Data Fig. 7. The identification of scarce neuronal subtypes in larva stage.

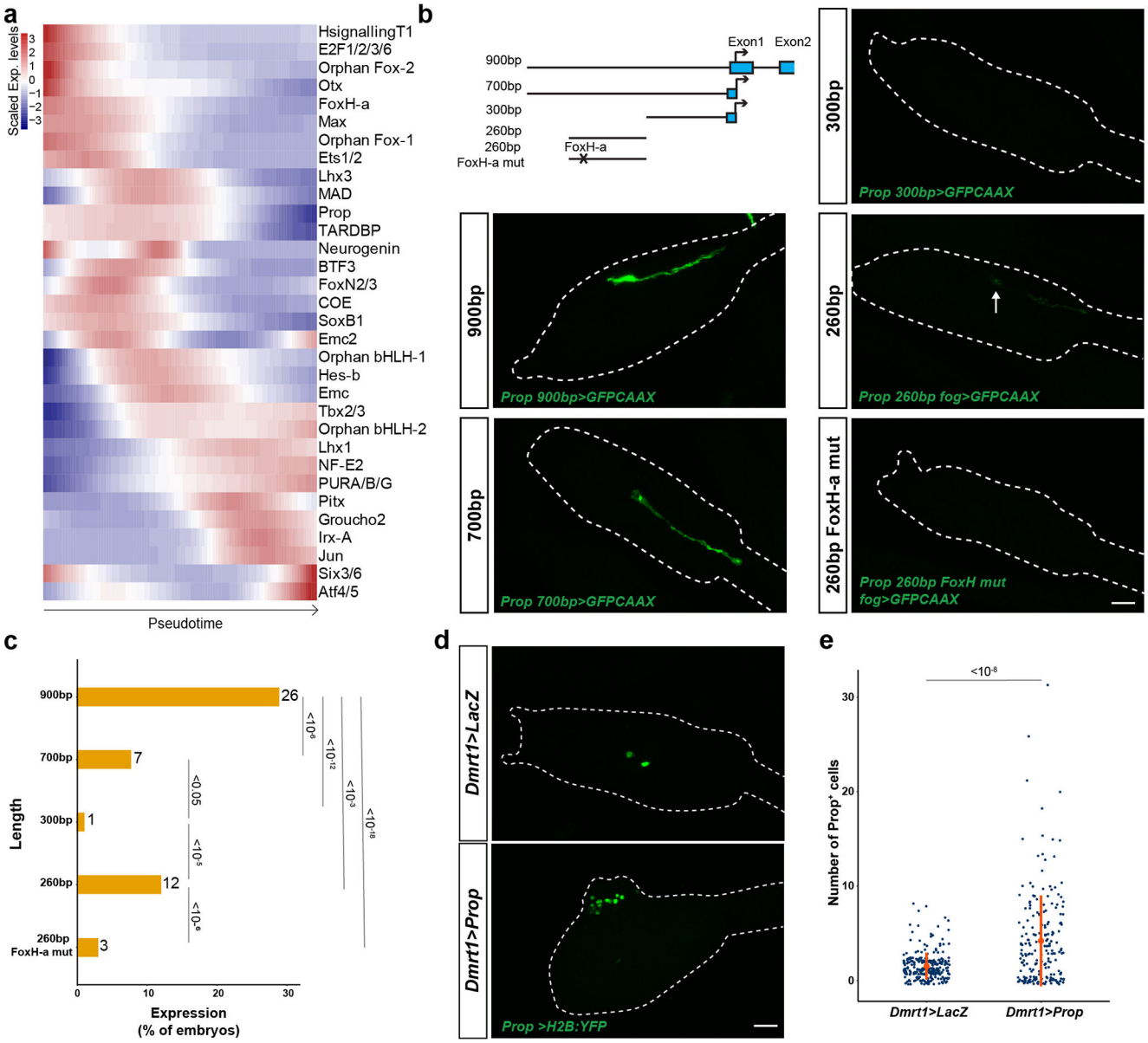
(a) Distribution of cells expressing *Galr2* in t-SNE plot. Cells within dashed circle show *Galr2* expression in BTNs (n = 26 cells). Reporter assay with a BTN-minimal enhancer for *Galr2* shows its specific activity in the BTNs (n = 3 electroporations). **(b)** Distribution of cells expressing *Dmbx* in t-SNE plot. Cells within dashed circle show *Dmbx* expression in the decussating neurons (ddNs) (n = 4 cells). The 5' regulatory sequences of *Dmbx* is active in the ddNs (red, n = 3 electroporations). **(c)** Distribution of cells expressing *NP (KH:C11.631)* in t-SNE plot. Cells within dashed circle show *NP* expression in Ci-VP^{+} pSV

(n = 11 cells). Reporter assay for *NP (*green*)* shows its specific expression in neurons in the posterior sensory vesicle (n = 3 electroporations). **(d)** Distribution of cells expressing *Prop* in t-SNE plot. Cells within dashed circle show *Prop* expression in Eminens (n = 17 cells). Expression of *Prop* reporter gene is specific to the Eminens neurons (green) (n = 3 electroporations). **(e)** t-SNE plot of the larval nervous system for *Ptf1a* (top) and *VP/OC* (bottom). The dotted circle corresponds to coronet cells (top panel, n = 72 cells) and the *ci-VP*$^{+}$ pSV cluster (bottom panels, n = 11 cells). **(f)** Expression of the reporter *Ptf1a>mChCAAX* (red) for the coronet cells and *NP>GFPCAAX* (green) for the ci-VP+ pSV showed both cell populations do not contact each other but are in close vicinity (top panel, n = 3 electroporations). Expression of the reporter *Prop>mChCAAX* (red) for the Eminens neurons and *NP>GFPCAAX* (green) for the ci-VP+ pSV. The *NP*$^{+}$ cells are also in proximity of the Eminens neurons (bottom panel, n = 2 electroporations). Scale bar: 10 μm.



Extended Data Fig. 8. Neuronal subtypes and expression of marker genes for Eminens neurons

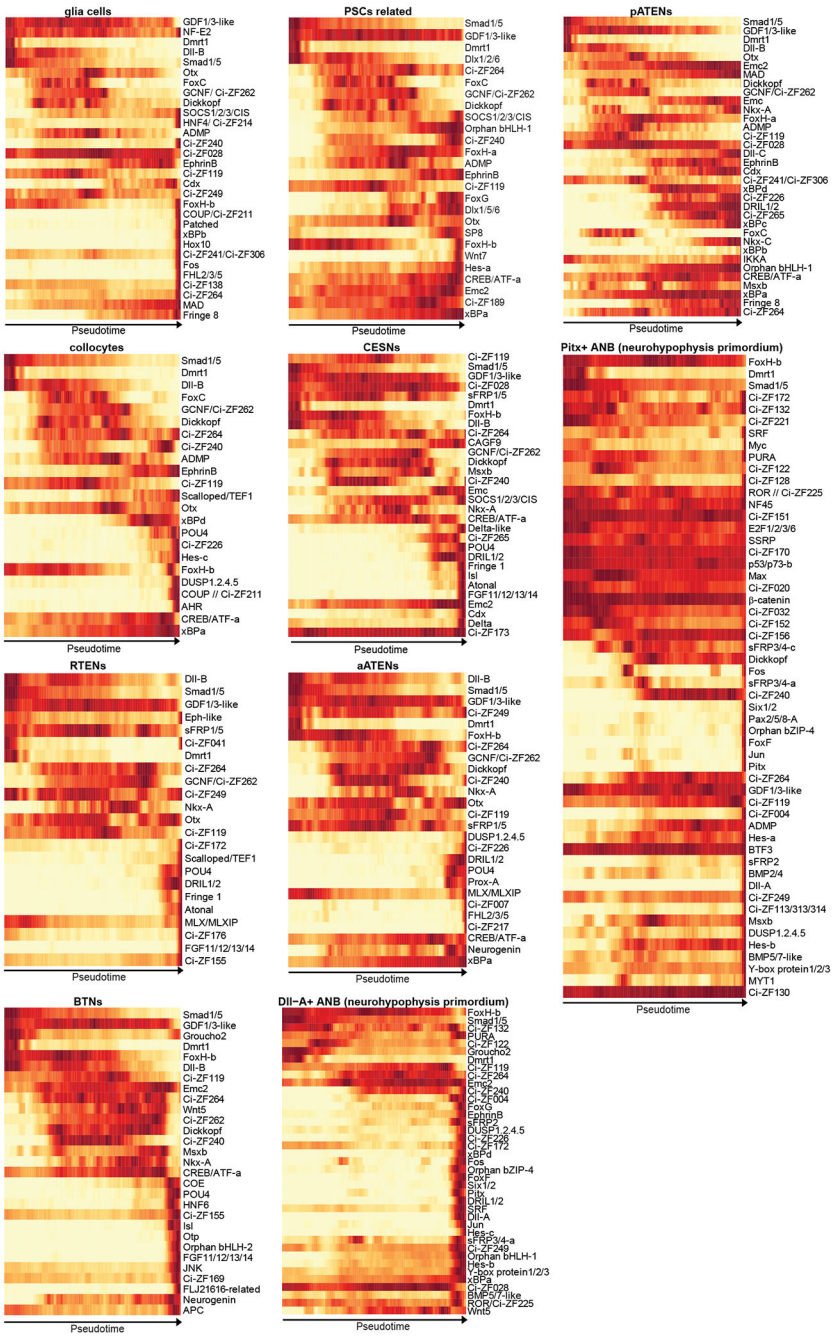
(a-h) Expression levels of 8 marker genes in larval nervous system shown in tSNE plot (left panel, n = 1,704 cells), and their corresponding reporter assays (*H2B:mCh*, red) with a *Prop>GFP* reporter (*GFPCAAX*, green, right panel, n = 2 electroporations for *Gad*, *S39aa 2.2kb*, *Znt3,* and *Asic1b* and n = 3 electroporations for *vGat*, *Calm*, *Fgf13*, and *Galr2*). The dashed circle in the tSNE plots identifies the Eminens neurons. Scale bar: 20 μm.



Extended Data Fig. 9. Manipulation of Eminens gene regulatory network.

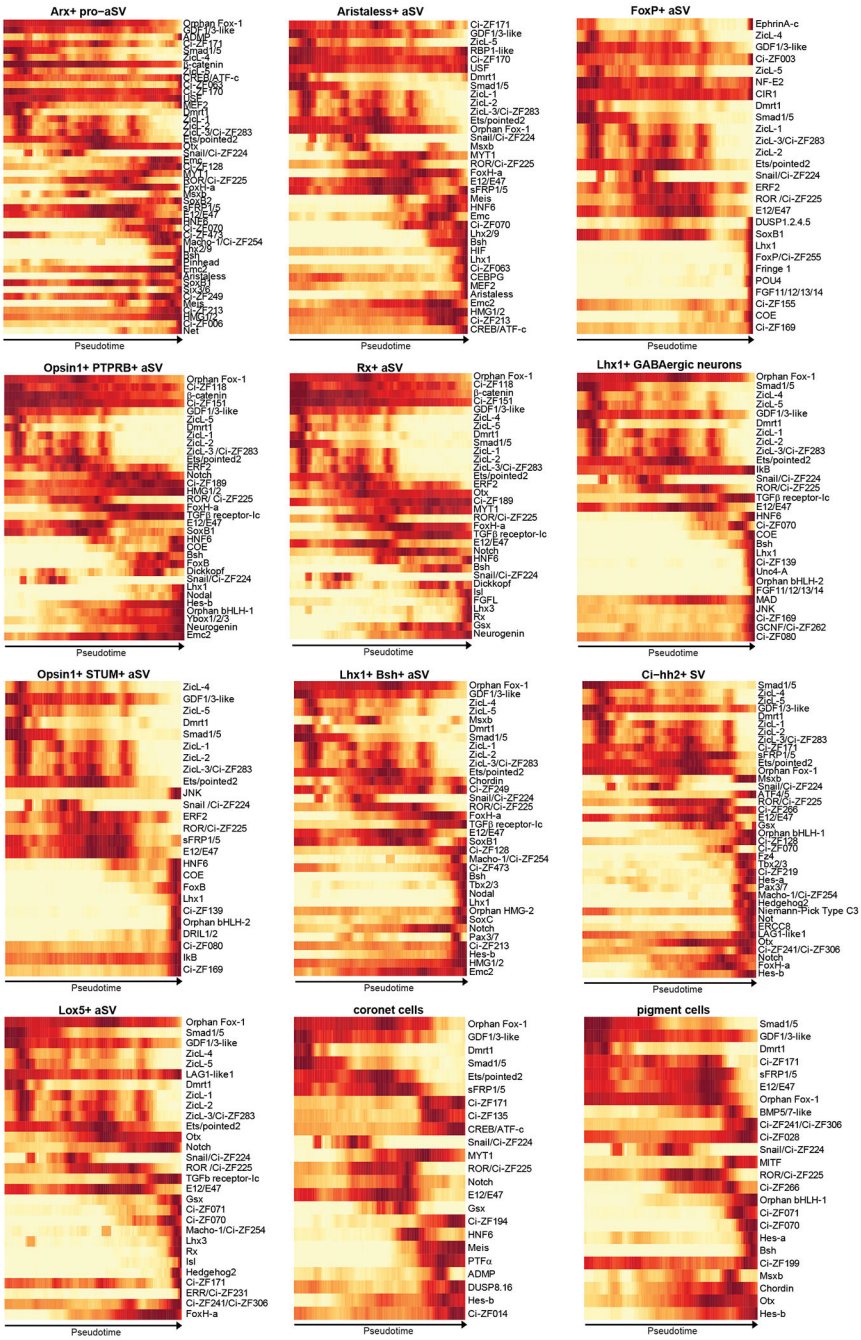
(a) Pseudo-temporal expression profiles of regulatory genes and signaling components in Eminens. **(b)** Diagram of the *Prop* regulatory sequences with their length indicated on the left. Representative embryo is shown for the different fusion genes (GFPCAAX, green). The minimal *Prop* enhancer has weak expression in Eminens (arrow). When FoxH-a binding site is mutated, it is even less active (260 FoxH-a mut). **(c)** Bar plot of the percentage of the embryos expressing GFP shown in panel (b). Numbers on the right of the column correspond to the percentage of GFP^+ embryos. Chi-square test with 4 degree of freedom was performed (p-value $<2.2x10^{-16}$), followed by Fisher exact test, two-sided, with Bonferroni adjustment for multiple comparison (p-value: 900bp *vs* 700bp: $1.05x10^{-7}$; 900bp *vs* 300bp: $3.47x10^{-13}$; 900bp *vs* 260bp: $2.36x10^{-4}$; 900bp *vs* 260bp FoxH-a mut:

$1.81x10^{-19}$; 700bp *vs* 300bp: 0.011; 700bp *vs* 260bp: 0.36; 700bp *vs* 260bp FoxH mut: 0.088; 300bp *vs* 260bp: $5.59x10^{-6}$; 300bp *vs* 260bp FoxH mut: 0.69; 260bp *vs* 260bp FoxH-a mut: $1.27x10^{-7}$; number of embryos: 900bp n = 207, 700bp n = 300, 300bp n = 160 pooled over two electroporations, 260bp n = 440, 260bp FoxH mut n = 750 pooled over three electroporations). **(d)** Overexpression of *Prop* with *Dmrt1* regulatory sequences causes supernumerary $Prop^+$ cells (bottom panel) compared to control embryo expressing *LacZ (*top panel*).* The Eminens cells are detected with a 2kb *Prop* reporter gene (H2B:YFP, green). **(e)** Quantification of $Prop^+$ cells from the electroporations in panel (d). (*Dmrt1>LacZ* n = 269 embryos; *Dmrt1>Prop* n = 210 embryos, pooled over three electroporation). The orange dots indicate the mean and the bars, the standard deviation. *Dmrt1>LacZ*: 1.5 cells+/−1.4 *Dmrt1>Prop*: 4.2 cells+/−4.8. Mann Withney test, p-value: $3.65x10^{-9}$. Scale bar: 20 μm.

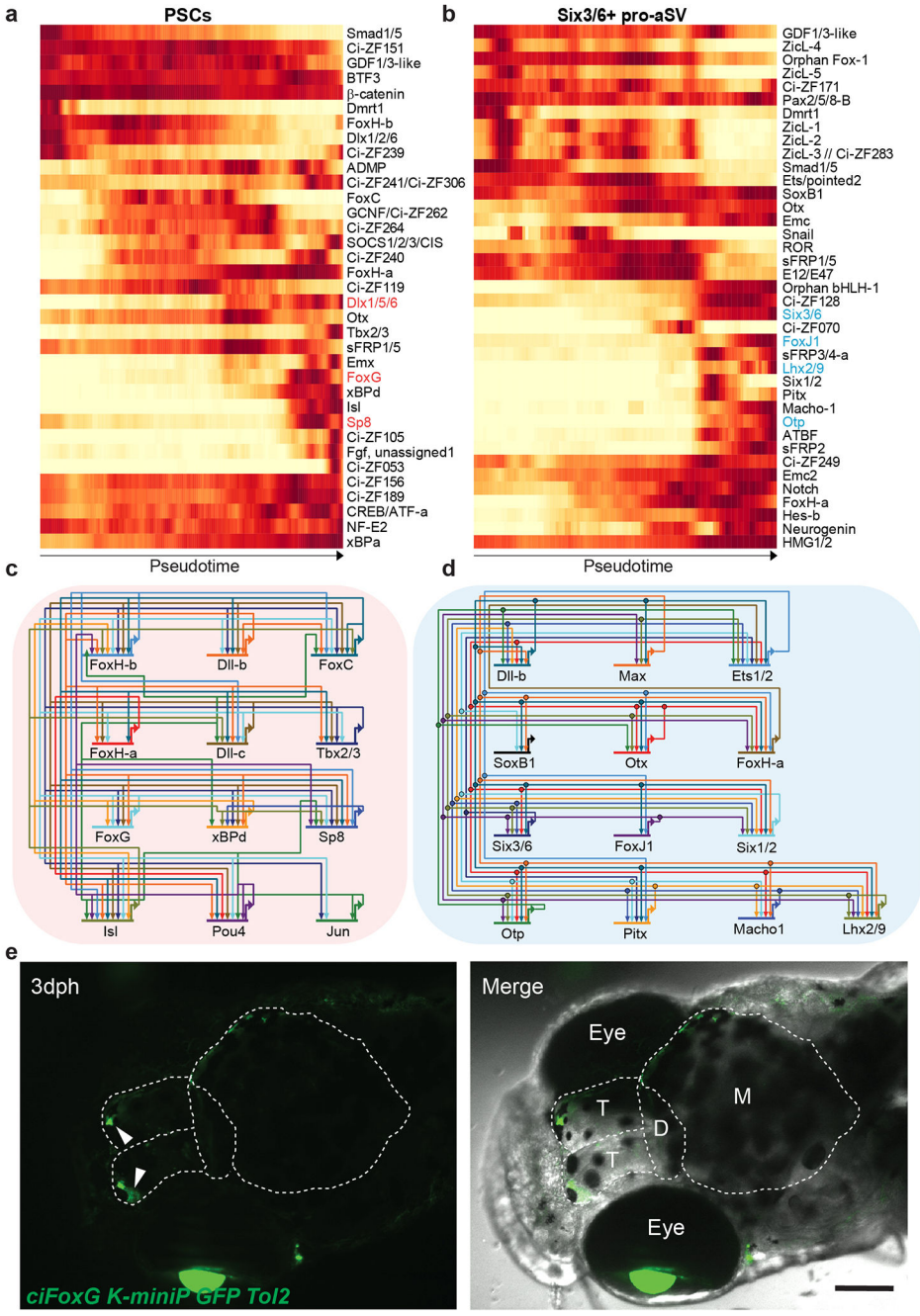


Extended Data Fig. 10. Pseudotemporal gene expression cascade of PNS.

Representative transcription factors and signaling pathway genes along pseudotime in reconstructed developmental trajectories of peripheral nervous system are shown.



Extended Data Fig. 11. Pseudotemporal gene expression cascade of a-lineage CNS.
Representative transcription factors and signaling pathway genes along pseudotime in reconstructed developmental trajectories of a-lineage central nervous system are shown. aSV is an abbreviation for anterior sensory vesicle.



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

(a) Gene expression cascade of regulatory genes and signaling components of palp sensory cells (ACCs). Genes implicated in the development of the vertebrate telencephalon were labeled in red. **(b)** Gene expression cascade of regulatory genes and signaling components in the anterior-most regions of the sensory vesicle (*Six3/6*$^{+}$ pro-aSV). Genes implicated in vertebrate telencephalon development are labelled in blue. The putative regulatory interactions among transcription factors from the cascade of PSCs **(c)** and *Six3/6*$^{+}$ pro-aSV **(d)** along 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 (left panel: green channel; right panel: merged image of bright field and green channel images). (n = three independent transgenic lines, see methods) T: telencephalon; D: diencephalon; M: midbrain. Scale bar: 400 μm.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank J.B. Wiggins, J.M. Miller and the Genomics Core Facility for technical support of 10x Chromium platform; IT/Bioinformatics staff at the Lewis-Sigler Institute for Integrative Genomics (LSI) for sequence alignment pipeline development; E. G. Gatzogiannis, director of the LSI Imaging Core Facility, for building the 2photon microscope and help on imaging; the Shvartsman lab in LSI for Imaris access, A. Sánchez Alvarado at the Stowers Institute for Medical Research (SIMR) for providing lab support and resources. We also thank Levine Lab members for helpful discussions and especially N. Treen to suggest mNeonGreen as fluorescent reporter.This study was supported by a grant from the NIH (NS076542) to M.L. W.W. (SIMR) was funded by Stowers Institute.

References

1. Briggs JA et al. The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. Science 360, (2018).
2. Farrell JA et al. Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. Science 360, (2018).
3. Wagner DE et al. Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. Science 360, 981–987, (2018). [PubMed: 29700229]
4. Pijuan-Sala B et al. A single-cell molecular map of mouse gastrulation and early organogenesis. Nature 566, 490–495, (2019). [PubMed: 30787436]
5. Cao J et al. The single-cell transcriptional landscape of mammalian organogenesis. Nature 566, 496–502, (2019). [PubMed: 30787437]
6. Delsuc F, Brinkmann H, Chourrout D & Philippe H Tunicates and not cephalochordates are the closest living relatives of vertebrates. Nature 439, 965–968, (2006). [PubMed: 16495997]
7. Imai KS, Levine M, Satoh N & Satou Y Regulatory blueprint for a chordate embryo. Science 312, 1183–1187, (2006). [PubMed: 16728634]
8. Ryan K, Lu Z & Meinertzhagen IA The CNS connectome of a tadpole larva of Ciona intestinalis (L.) highlights sidedness in the brain of a chordate sibling. Elife 5, (2016).
9. Prodon 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). [PubMed: 17366574]
10. Corbo JC, Levine M & Zeller RW Characterization of a notochord-specific enhancer from the Brachyury promoter region of the ascidian, Ciona intestinalis. Development 124, 589–602, (1997). [PubMed: 9043074]
11. Tokuoka M, Imai KS, 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). [PubMed: 15355799]
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). [PubMed: 3582738]
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). [PubMed: 23813578]
14. Veeman MT, Newman-Smith E, El-Nachef D & Smith WC 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). [PubMed: 20438724]

15. Stemple DL Structure and function of the notochord: an essential organ for chordate development. Development 132, 2503–2512, (2005). [PubMed: 15890825]
16. Suzuki MM & Satoh N Genes expressed in the amphioxus notochord revealed by EST analysis. Dev Biol 224, 168–177, (2000). [PubMed: 10926757]
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). [PubMed: 14993185]
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). [PubMed: 16835438]
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). [PubMed: 15950616]
20. Takahashi H et al. Brachyury downstream notochord differentiation in the ascidian embryo. Genes Dev 13, 1519–1523, (1999). [PubMed: 10385620]
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). [PubMed: 30228204]
22. Stolfi A, Ryan K, Meinertzhagen IA & Christiaen L Migratory neuronal progenitors arise from the neural plate borders in tunicates. Nature 527, 371–374, (2015). [PubMed: 26524532]
23. Shi TJ 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). [PubMed: 16487144]
24. Holmes FE et al. Targeted disruption of the galanin gene reduces the number of sensory neurons and their regenerative capacity. Proc Natl Acad Sci U S A 97, 11563–11568, (2000). [PubMed: 11016970]
25. Ryan K, Lu Z & Meinertzhagen IA Circuit Homology between Decussating Pathways in the Ciona Larval CNS and the Vertebrate Startle-Response Pathway. Curr Biol 27, 721–728, (2017). [PubMed: 28216318]
26. Korn H & Faber DS The Mauthner cell half a century later: a neurobiological model for decision-making? Neuron 47, 13–28, (2005). [PubMed: 15996545]
27. Stolfi A & Levine M Neuronal subtype specification in the spinal cord of a protovertebrate. Development 138, 995–1004, (2011). [PubMed: 21303852]
28. Hamada M et al. Expression of neuropeptide- and hormone-encoding genes in the Ciona intestinalis larval brain. Dev Biol 352, 202–214, (2011). [PubMed: 21237141]
29. Ryan K, Lu Z & Meinertzhagen IA The peripheral nervous system of the ascidian tadpole larva: Types of neurons and their synaptic networks. J Comp Neurol 526, 583–608, (2018). [PubMed: 29124768]
30. Imai JH & Meinertzhagen IA Neurons of the ascidian larval nervous system in Ciona intestinalis: I. Central nervous system. J Comp Neurol 501, 316–334, (2007). [PubMed: 17245701]
31. Takamura K, Minamida N & Okabe S Neural map of the larval central nervous system in the ascidian Ciona intestinalis. Zoolog Sci 27, 191–203, (2010). [PubMed: 20141424]
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). [PubMed: 8410186]
33. Winkle CC 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). [PubMed: 27147649]
34. Abitua PB et al. The pre-vertebrate origins of neurogenic placodes. Nature 524, 462–465, (2015). [PubMed: 26258298]
35. Abitua PB, Wagner E, Navarrete IA & Levine M Identification of a rudimentary neural crest in a non-vertebrate chordate. Nature 492, 104–107, (2012). [PubMed: 23135395]
36. Stolfi A et al. Early chordate origins of the vertebrate second heart field. Science 329, 565–568, (2010). [PubMed: 20671188]

37. Horie R et al. Shared evolutionary origin of vertebrate neural crest and cranial placodes. Nature 560, 228–232, (2018). [PubMed: 30069052]
38. Zeng F et al. Papillae revisited and the nature of the adhesive secreting collocytes. Dev Biol, (2018).
39. Hebert JM & Fishell G The genetics of early telencephalon patterning: some assembly required. Nat Rev Neurosci 9, 678–685, (2008). [PubMed: 19143049]
40. 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). [PubMed: 17470284]
41. Jacquet BV 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). [PubMed: 21697387]
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). [PubMed: 22736245]
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). [PubMed: 29539641]
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). [PubMed: 17557317]
46. Satou Y, Kawashima T, Shoguchi E, Nakayama A & Satoh N An integrated database of the ascidian, Ciona intestinalis: towards functional genomics. Zoolog Sci 22, 837–843, (2005). [PubMed: 16141696]
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). [PubMed: 29608179]
48. Linderman GC, Rachh M, Hoskins JG, Steinerberger S & Kluger Y Fast interpolation-based t-SNE for improved visualization of single-cell RNA-seq data. Nat Methods 16, 243–245, (2019). [PubMed: 30742040]
49. Qiu X et al. Reversed graph embedding resolves complex single-cell trajectories. Nat Methods 14, 979–982, (2017). [PubMed: 28825705]
50. Haghverdi L, Buettner F & Theis FJ Diffusion maps for high-dimensional single-cell analysis of differentiation data. Bioinformatics 31, 2989–2998, (2015). [PubMed: 26002886]
51. Frith MC, Li MC & Weng Z Cluster-Buster: Finding dense clusters of motifs in DNA sequences. Nucleic Acids Res 31, 3666–3668, (2003). [PubMed: 12824389]
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). [PubMed: 29161433]
53. Love MI, Huber W & Anders S Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15, 550, (2014). [PubMed: 25516281]
54. Wagner E & Levine M FGF signaling establishes the anterior border of the Ciona neural tube. Development 139, 2351–2359, (2012). [PubMed: 22627287]
55. Yoshida R et al. Identification of neuron-specific promoters in Ciona intestinalis. Genesis 39, 130–140, (2004). [PubMed: 15170699]
56. Shaner NC et al. Improving the photostability of bright monomeric orange and red fluorescent proteins. Nat Methods 5, 545–551, (2008). [PubMed: 18454154]
57. Stauffer TP, Ahn S & Meyer T Receptor-induced transient reduction in plasma membrane PtdIns(4,5)P2 concentration monitored in living cells. Curr Biol 8, 343–346, (1998). [PubMed: 9512420]
58. Shaner NC et al. A bright monomeric green fluorescent protein derived from Branchiostoma lanceolatum. Nat Methods 10, 407–409, (2013). [PubMed: 23524392]

59. Gregory C & Veeman M 3D-printed microwell arrays for Ciona microinjection and timelapse imaging. PLoS One 8, e82307, (2013). [PubMed: 24324769]

60. Schindelin J et al. Fiji: an open-source platform for biological-image analysis. Nat Methods 9, 676–682, (2012). [PubMed: 22743772]

61. R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing, 2013).

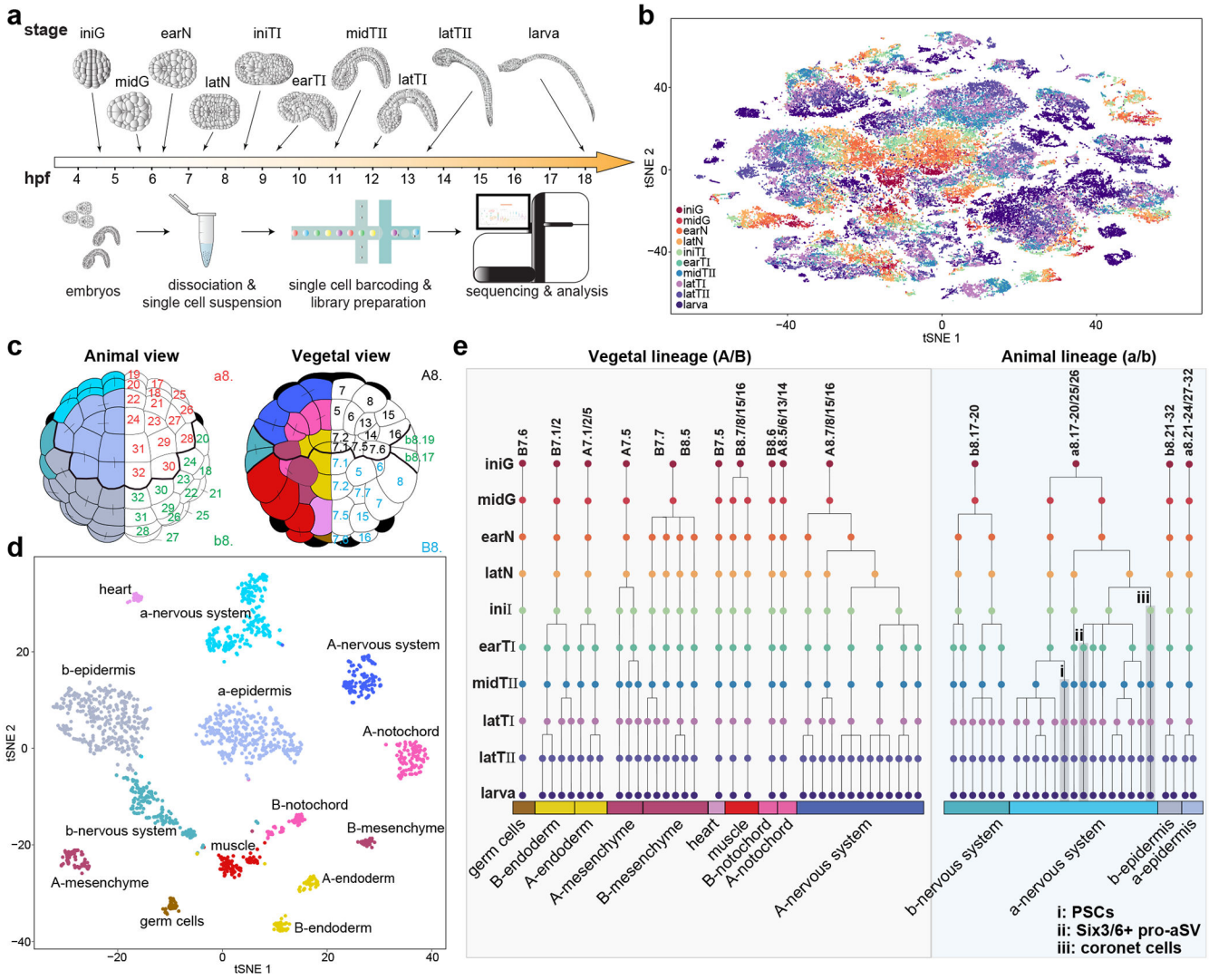


Fig. 1. Overview of scRNA-Seq assays and cell type specification at the onset of gastrulation.

(a) Staged embryos were collected from 10 different developmental stages, beginning with the initial gastrula stage (iniG), middle gastrula (midG), early neurula (earN), late neurula (latN), initial tailbud I (iniTI), early tailbud I (earTI), middle tailbud II (midTII), late tailbud I (latTI), late tailbud II (latTII), and larva (larva) (from iniG to latTII: n = 2 biological replicates per stage, larva stage: n = 3 biological replicates). **(b)** t-SNE plot of the entire dataset (n = 90,579 cells). Cells color-coded according to developmental stage (key in lower left). **(c)** Schematics of animal (left) and vegetal (right) blastomeres of a *Ciona* embryo at the initial gastrula stage. Tissue types were color-coded (left) and named according to Conklin's nomenclature (right). Bold lines indicate the boundaries between the a-, b-, A-, and B-lineage blastomeres. **(d)** t-SNE plot of transcriptomes from single cells at the initial gastrula stage (n = 1,731 cells) using the color-coding scheme shown in (c). Each of the major tissues maps within a separate cluster. **(e)** Virtual lineage trees were reconstructed using transcriptome profiles from sequential developmental stages. The points in the tree represent inferred developmental transitions from initial gastrula to larva. Only unambiguous

alignments are shown (Methods). Branches labeled in shadow represent PSCs (i), Six3/6+-pro-aSV (ii) and coronet cells (iii) lineages, respectively.

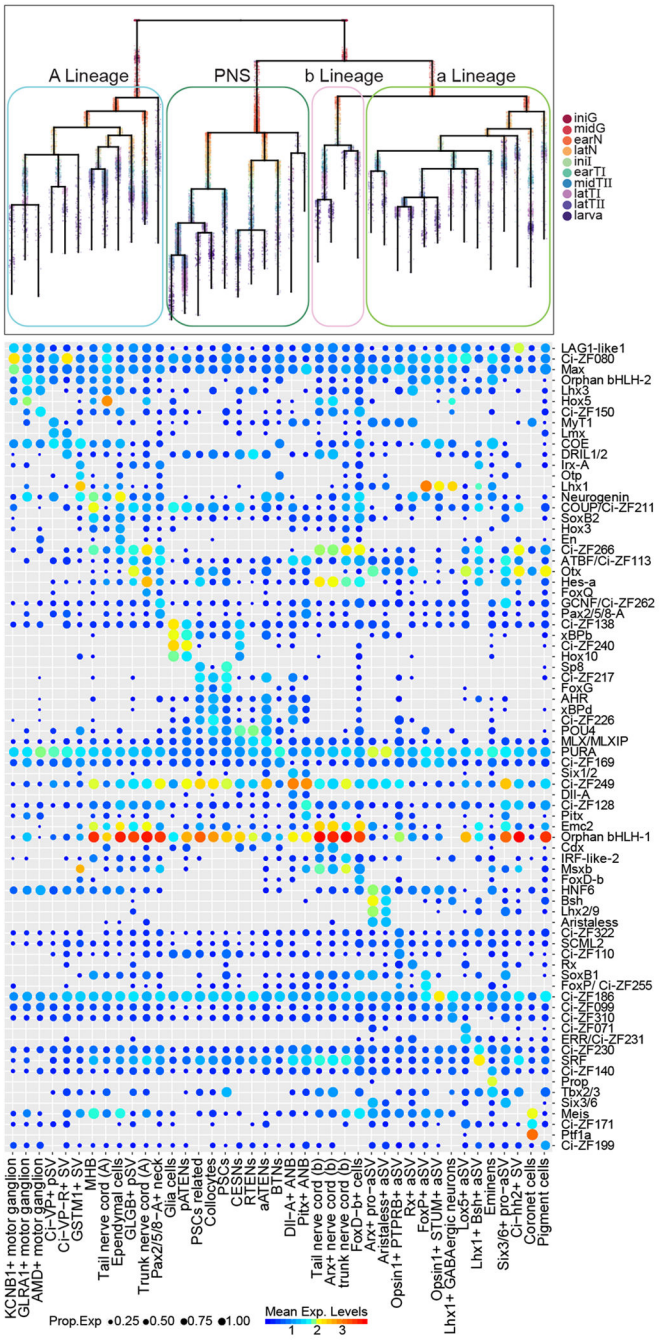


Fig. 2. Transcriptome trajectories for defined individual neurons.

Reconstructed expression lineage for the entire nervous system. Cells were colored by developmental stage, and the a-, b- and A- lineage branches of the central nervous system and peripheral nervous system are identified (top). Cells were ordered by pseudotime along each trajectory. Dotplot of the top 3 highly expressed regulatory genes in each neural subcluster at larva stage (bottom). Dot size represents the percentage of cells expressing the transcription factor, and the dot color shows the averaged expression level. MHB: midbrain-hindbrain boundary, pATENs: posterior apical trunk epidermal neurons, CESNs: caudal

epidermal sensory neurons, RTENs: rostral trunk epidermal neurons, aATENs: anterior apical trunk epidermal neurons, ANB: anterior neural boundary.

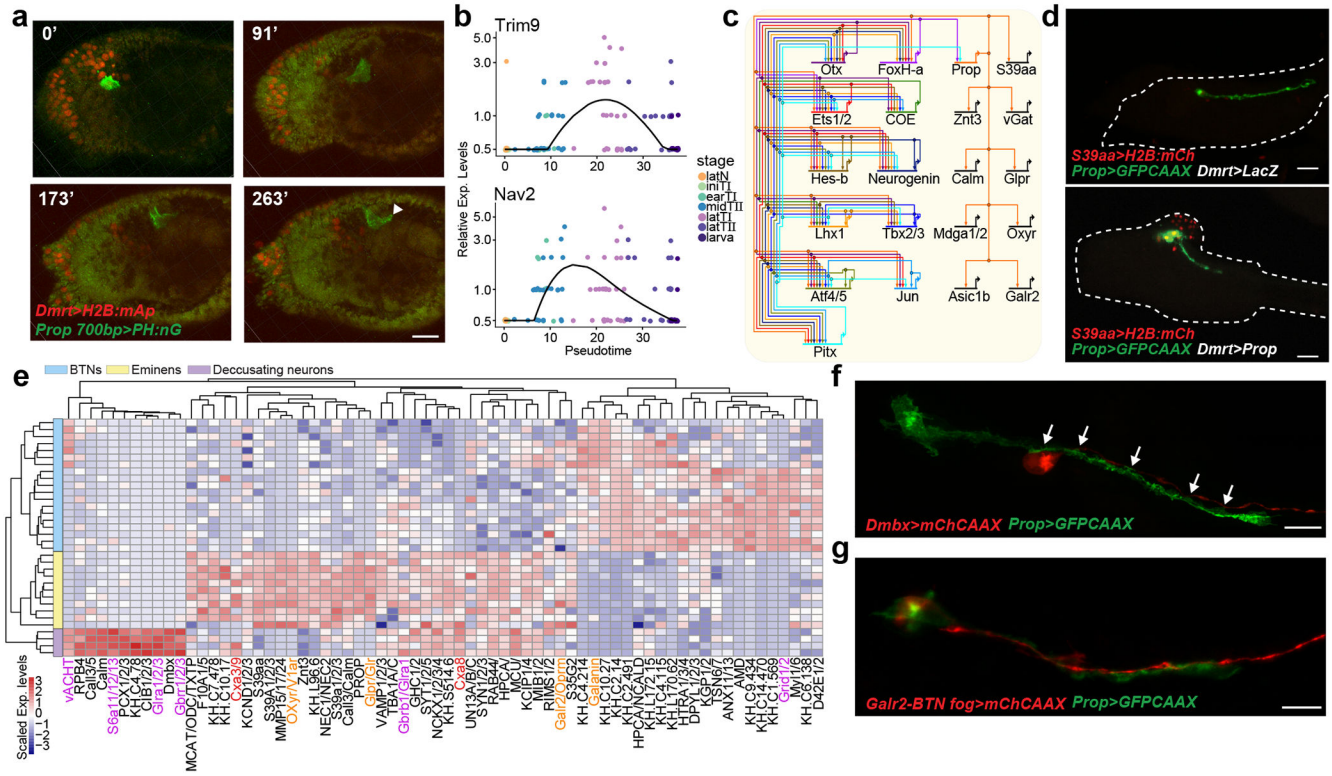


Fig.3. Integration of transcriptome maps and synaptome neuronal circuits.

(a) Snap shots of a time-lapse video of Eminens cell migration (from LTBI to LTBIII, n = 2 embryos). The embryo expressed H2B:mApple under the regulatory sequences of *Dmrt1* (red), and PH:nG driven by *Prop* regulatory sequences (green). The Eminens migrates from the anterior side toward the posterior of the sensory vesicle before axogenesis (arrowhead). Numbers in upper left indicate elapsing time (minute). **(b)** Pseudo-temporal expression of 2 migration related genes from early neurula to larva stage. **(c)** A provisional gene regulatory network of Eminens was shown based on the regulatory cascade (Extendend Data Fig. 9a). Only representative regulatory genes are shown (Methods). **(d)** Upon overexpression of *Prop* in the anterior neural plate, supernumerary *S39aa*$^{+}$ cells (H2B:mCh, red) are observed (bottom, *Dmrt1>Prop)* compared with control embryos *(*top*, Dmrt>LacZ).* The embryos also expressed a membrane GFP as *Prop* reporter gene (GFPCAAX, green, n= 3 electroporations for both conditions). **(e)** Heatmap of differentially expressed genes between Ems, ddNs, and Ems, BTNs. Genes involved in neurotransmission were colored in magenta, connexin genes in red, neuropeptide-associated genes in orange. **(f)** Em2 has multiple contacts (arrows) with the ddNs, as observed by double labeling with *Prop* and *Dmbx* reporter genes (green and red, respectively). (n= 3 electroporations). **(g)** Double labeling with *Prop* and BTN-specific-*Galr2* reporter genes (green and red, respectively) shows extensive contacts of BTN axon with Ems (n= 2 electroporations). Scale bar: (a) and (d): 20 μm (f) and (g): 10 μm.

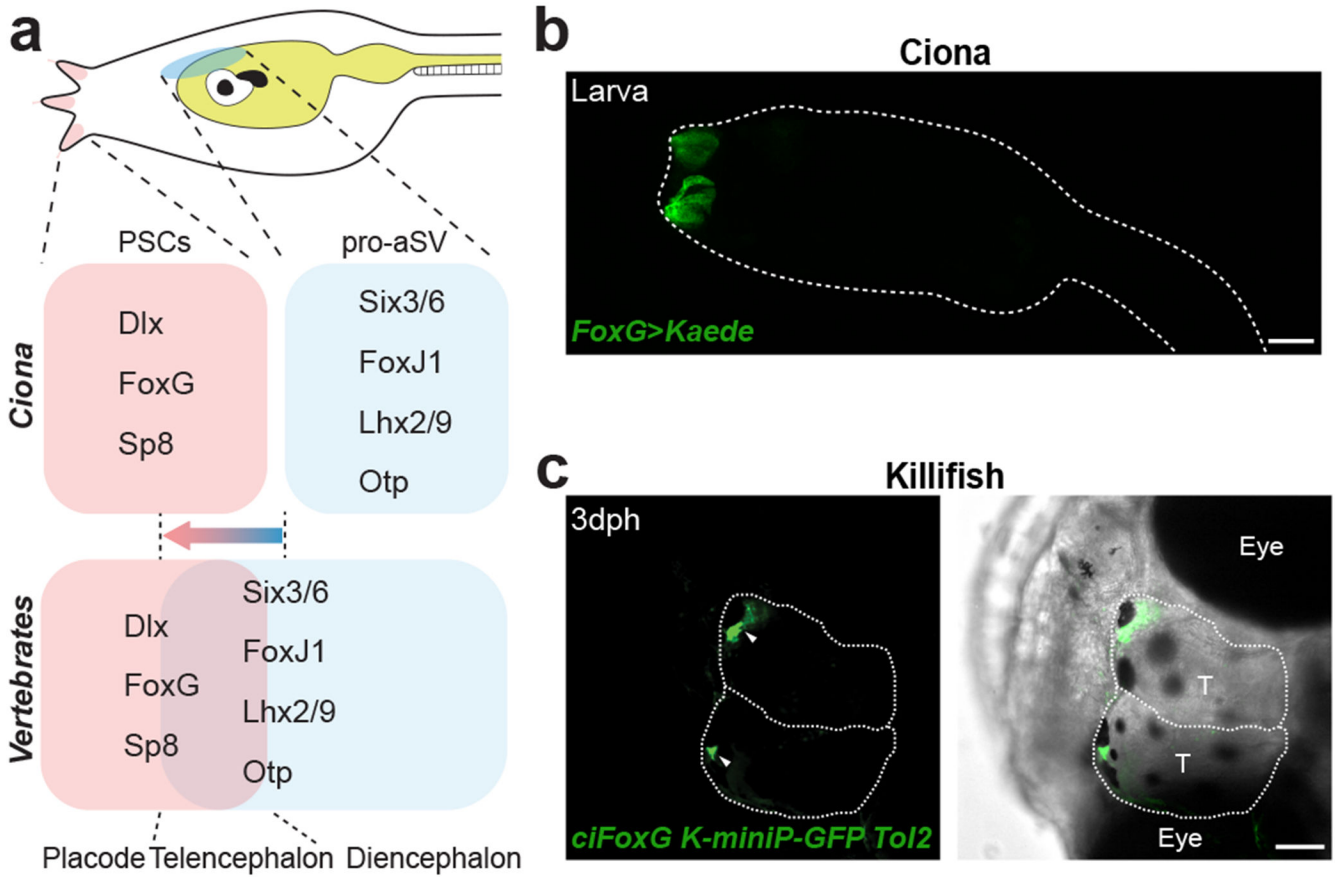


Fig. 4. Model for the evolution of the telencephalon.

(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* palps, including *Dlx*, *FoxG*, *Sp8*, and in pro-aSV, 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 electroporations). **(c)** In killifish, GFP driven by the *Ciona FoxG* regulatory sequences and a zebrafish minimal enhancer is expressed in a subset of cells in the olfactory bulb of the telencephalon (arrowheads, left panel) (n = 3 independent transgenic lines, see methods). T, telencephalon. Scale bar (b): 20 μm and (c):250 μm.