nature ecology & evolution

Article

https://doi.org/10.1038/s41559-024-02387-8

Ascidian embryonic cells with properties of neural-crest cells and neuromesodermal progenitors of vertebrates

Received: 9 June 2023

Accepted: 6 March 2024

Published online: 02 April 2024

Check for updates

Tasuku Ishida & Yutaka Satou

Neural-crest cells and neuromesodermal progenitors (NMPs) are multipotent cells that are important for development of vertebrate embryos. In embryos of ascidians, which are the closest invertebrate relatives of vertebrates, several cells located at the border between the neural plate and the epidermal region have neural-crest-like properties; hence, the last common ancestor of ascidians and vertebrates may have had ancestral cells similar to neural-crest cells. However, these ascidian neural-crest-like cells do not produce cells that are commonly of mesodermal origin. Here we showed that a cell population located in the lateral region of the neural plate has properties resembling those of vertebrate neural-crest cells and NMPs. Among them, cells with *Tbx6-related* expression contribute to muscle near the tip of the tail region and cells with *Sox1/2/3* expression give rise to the nerve cord. These observations and cross-species transcriptome comparisons indicate that these cells have properties similar to those of NMPs. Meanwhile, transcription factor genes *Dlx.b*, *Zic-r.b* and *Snai*, which are reminiscent of a gene circuit in vertebrate neural-crest cells, are involved in activation of *Tbx6-related.b*. Thus, the last common ancestor of ascidians and vertebrates may have had cells with properties of neural-crest cells and NMPs and such ancestral cells may have produced cells commonly of ectodermal and mesodermal origins.

In most animal embryos, three germ layers are specified in early development and these are further specified to various cell types. However, neural-crest cells, which are formed in the neural plate border of vertebrate embryos, retain or re-activate their ability to produce cells that are commonly of mesodermal and ectodermal origins, even after gastrulation[1–3]. It is widely believed that neural-crest cells contributed to evolution of vertebrates, especially evolution of the head region[4]. Neural-crest cells give rise to various cells including the pharyngeal skeleton, smooth muscle of the aortic arches and the peripheral neurons. In other words, this cell population produces cells commonly of ectodermal origin and mesodermal origin. The gene regulatory network for differentiation of this cell population has been extensively studied and its hierarchical structure has been uncovered[5]. Specifically, a gene regulatory circuit typically containing *Dlx5/6*, *Msx1*, *Zic1*, *Tfap2* and *Pax3/7* acts at the early neural plate border and soon after it activates genes including *Snai* and *Id* in premigratory neural-crest cells, although there may be small differences among species.

Recent studies have suggested that embryos of ascidians, which belong to the subphylum Tunicata, the sister group of vertebrates, contain cells which share an evolutionary origin with vertebrate neural-crest cells[6–9]. These cells indeed differentiate into cells that include sensory neurons and pigment cells. Although homology of ascidian cells producing sensory bipolar tail neurons and vertebrate neural-crest cells is controversial[10], it is likely that the last common ancestor of vertebrates and ascidians had cells resembling vertebrate neural-crest cells. Ascidian neural-crest-like cells identified so far do

Department of Zoology, Graduate School of Science, Kyoto University, Kyoto, Japan. e-mail: yutaka@ascidian.zool.kyoto-u.ac.jp

https://doi.org/10.1038/s41559-024-02387-8

not produce cell types that are commonly of mesodermal origin, which raises the question of whether ancestral neural-crest cells of the last common ancestor of vertebrates and ascidians (ancestral Olfactores) had the ability to produce cells that were commonly not only of ectodermal origin but also of mesodermal origin.

Meanwhile, neuromesodermal progenitors (NMPs) of vertebrate embryos, which reside in the tailbud, are another cell population that has the ability to produce mesodermal (presomitic mesodermal) and ectodermal (spinal cord) cells of posterior structures in late embryos[11,12]. NMPs express *T* and *Sox2* simultaneously[13,14]. Mesodermal cells differentiating from NMPs express *Tbx6*, which represses *Sox2*, while neural cells differentiating from NMPs maintain *Sox2* expression[13,15–18]. It has not been determined whether ancestral Olfactores had NMP-like cells. If they had, did ancestral NMP-like cells produce cells that are commonly of mesodermal and ectodermal origins and how did these ancestral neural-crest cells and NMPs contribute to the body plan of ancestral Olfactores?

The central nervous system of ascidian larvae is derived from three pairs of cells of eight-cell embryos and these pairs and their descendants are called a-, b- and A-line cells[19–21]. In other words, these three lineages of cells make up the neural plate. However, the most anterior portion of the neural plate is now considered to be the structure that shares an evolutionary origin with vertebrate cranial placodes[22–27]; therefore, it may be better to call this region the anterior neural plate border. Likewise, b-line cells constitute the lateral part of the neural plate and mainly contribute to ependymal cells of the dorsal row of the nerve cord but not to neurons[19,21]. These cells are derived from a pair of cells called b6.5 at the 32-cell stage. At the 64-cell stage, its daughter cells, b7.9 and b7.10, laterally abut cells that contribute to neurons of the central nervous system (Supplementary Fig. 1). Among their daughter cells, b8.17 and b8.19 (and their descendants) also abut cells that contribute to neurons of the central nervous system, while the other daughter cells (b8.18 and b8.20) and their descendants are located more laterally at the gastrula stage (Fig. 1). The former cells (b8.17 and b8.19) have been regarded as neural plate cells[19,21] and we call them and their descendants lateral neural plate cells (LNPCs) in the present study. Peripheral sensory neurons are derived from the latter (b8.18/20-line) cells and these cells may share their evolutionary origin with vertebrate neural-crest cells (cyan-coloured cells in Fig. 1)[7,8]. In addition, another pair of neural-crest-like cells, which produce pigment cells, is identified in the neural plate (asterisks in Fig. 1)[6]. We investigated the b6.5-descendants in embryos of ascidians (*Ciona robusta* or *Ciona intestinalis* type A) to answer the question of the evolutionary origin of multipotent neural-crest cells and NMPs and found that cells of the LNPC lineage have properties of neural-crest cells and NMPs of vertebrates.

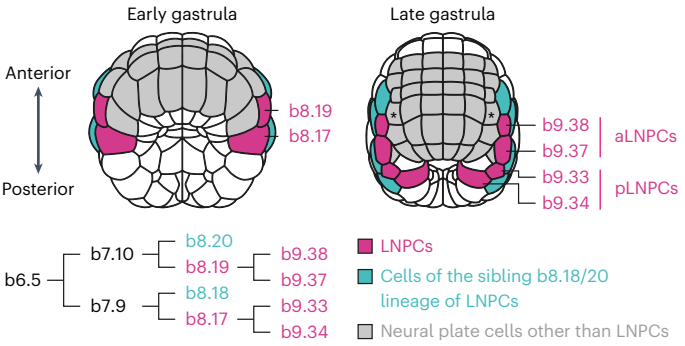


Fig. 1 | Ascidian bilaterally symmetrical embryos at the early and late gastrula stages. Neural plate cells other than b-line lineage cells are coloured grey. Magenta-coloured cells are b8.17 and b8.19 at the early gastrula stage and their descendants at the late gastrula stage, which we call LNPCs. Cyan-coloured cells are b8.18 and b8.20, which are siblings of b8.17 and b8.19, at the early gastrula stage and their descendants at the late gastrula stage. These cells produce peripheral neurons and are thought to be homologous to vertebrate neural-crest cells[7,8]. In addition, pigment cells in the brain vesicle are derived from cells marked by asterisks and these cells are also thought to share their origin with vertebrate neural-crest cells[6]. The lineage of LNPCs is shown below. These cells are derived from b6.5 cells in 32-cell embryos. Note that the most anterior part of the neural plate is thought to share its evolutionary origin with vertebrate anterior placodes[22–27].

Results

Posterior LNPCs give rise to muscle and tail-tip cells

Previous studies[19,28] have shown that b8.17 (posterior LNPCs (pLNPCs)) give rise to cells near the tip of the tail of tailbud embryos. These cells are two pairs of muscle cells and four cells of posterior parts of the nerve cord and endodermal strand. Because *Tbx6-related.b* (*Tbx6-r.b*) is required for specification of all muscle cells, including muscle cells derived from b8.17 (ref. 29), we examined *Tbx6-r.b* expression in LNPCs and found that the posterior daughter cells (b9.34) of b8.17 express *Tbx6-r.b* (Fig. 2a). Indeed, these cells expressed *Acta.a*, which encoded muscle actin (Extended Data Fig. 1). Therefore, it is highly likely that the posterior daughters give rise to muscle cells.

This observation implies that pLNPC-derived nerve-cord and endodermal-strand cells are derived from the anterior daughter of b8.17 (b9.33). *Hand-r* is expressed in cells near the tip of the tail[30–33] but it cannot be used as a specific marker because it is also expressed in other cells. Therefore, we used *Hand-r* as a marker gene to find genes expressed specifically in pLNPC-derived nerve-cord and endodermal-strand cells in single-cell RNA sequencing (scRNA-seq) data published previously[34] (Supplementary Table 1). Among these genes, we confirmed that *Hebp-r.a* is indeed expressed in putative pLNPC-derived nerve-cord and endodermal-strand cells by in situ hybridization (Fig. 2b). When we introduced a reporter gene containing the upstream regulatory region of *Hebp-r.a* with another reporter gene containing the upstream region of *Msx*, these two reporters were expressed in the same cells at the tip of the tail region, which is commonly called the tailbud (Fig. 2c). Because *Msx* is specifically expressed in b8.17, b8.18, b8.19 and b8.20 in early gastrulae[8,30,35–37] and because only b8.17 contributes to pLNPC-derived nerve-cord cells and endodermal-strand cells[19], we conclude that *Hebp-r.a* specifically marks pLNPC-derived nerve-cord cells and endodermal-strand cells. Four cells expressed *Hebp-r.a* at the early tailbud stage (Fig. 2b). This coincides with the number of b8.17 derivatives in a previous study[19]. It is likely that these pLNPC-derived nerve-cord and endodermal-strand cells divided once between the early and late tailbud stages because eight cells expressed *Hebp-r.a* reporter at the late tailbud stage (Fig. 2c). Thus, the anterior daughter (b9.33) of pLNPC (b8.17) gives rise to the most posterior parts of nerve-cord cells and endodermal-strand cells and the posterior daughter (b9.34) gives rise to muscle cells.

At the early tailbud stage, pLNPC-derived nerve-cord cells and endodermal-strand cells were initially located in the posterior part of the nerve cord (Fig. 2d). Subsequently, some of them changed their location to the ventral side and these cells were finally located in the posterior part of the endodermal strand (Fig. 2d and Extended Data Fig. 2). Using scRNA-seq data[34], we compared expression profiles of pLNPC-derived nerve-cord cells and endodermal-strand cells with those of other cell types during the larval stage and found that the transcriptome of pLNPC-derived nerve-cord cells and endodermal-strand cells was quite different from those of the other endodermal-strand and nerve-cord cells and of other tissues (Fig. 2e). This was further supported by the observation that pLNPC-derived nerve-cord cells and endodermal-strand cells did not express a pan-neural marker, *Celf3.a*[38], or an endodermal-strand marker, *Slc39a-related*[32] (Extended Data Fig. 3). Accordingly, these pLNPC-derived cells constitute a cell population distinct from nerve-cord cells or endodermal-strand cells located anteriorly.

The ascidian larval nerve cord mostly consists of four rows of cells, while the most posterior part of the nerve cord consists of a row

https://doi.org/10.1038/s41559-024-02387-8

of several cells but not of four cell rows[39]. Similarly, the endodermal strand mostly consists of two rows of cells, while the most posterior part of the endodermal strand does not and germ-line cells intervene between the anterior two rows of endodermal-strand cells and the most posterior part of the endodermal strand[39]. Judging from their locations, it is likely that these posterior nerve-cord cells and endodermal-strand cells are pLNPC-derived cells. These observations also suggest that pLNPC-derived cells may constitute a cell population distinct from nerve-cord cells or endodermal-strand cells located anteriorly (Fig. 2f). We call these pLNPC-derived cells tail-tip cells hereafter, although we do not necessarily rule out the possibility that these cells are specialized cells of the nerve cord or endodermal strand.

Regulatory factors activating *Tbx6-r.b* in LNPCs

Next, we examined how *Tbx6-r.b* is activated in b9.34. Previous studies have shown that orthologues for genes specifying neural plate border cells and neural-crest cells in vertebrate embryos are expressed in cells including LNPCs of ascidian embryos[30,36,37,40,41]. We confirmed that *Dlx.b*, *Msx*, *Snai*, *Zic-r.b*, *Tfap2-r.b*, *Pax3/7*, *Ets1/2.b*, *Lmx1* and *Id.b* were expressed in LNPCs (Extended Data Fig. 4).

Dlx.b is expressed in animal hemisphere cells, including LNPCs, in gastrulae and is important for fate decision of these cells[40]. *Msx* is expressed in LNPCs and their sibling b8.18/20-line cells[30] and is required for differentiation of sensory neurons[8,35]. While knockdown of *Msx* did not affect *Tbx6-r.b* expression, *Dlx.b* knockdown downregulated *Tbx6-r.b* in b9.34 (Fig. 3a). Note that signals for *Tbx6-r.b* expression in A-line muscle cells are seen over b9.33 cells where *Tbx6-r.b* is not expressed (Supplementary Fig. 2). *Snai* was also expressed in LNPCs (Extended Data Fig. 4) and knockdown of *Snai* downregulated *Tbx6-r.b* expression (Fig. 3a). In addition, *Zic-r.b* was downregulated in *Dlx.b* morphants (Fig. 3b) and *Snai* was downregulated in *Zic-r.b* morphants (Fig. 3c). Thus, orthologues of genes that specify fates of the neural plate border and neural-crest cells in vertebrate embryos, also regulated *Tbx6-r.b* expression in b9.34.

Tbx6-r.b potentially suppresses *Sox1/2/3* expression

While LNPCs are located in the lateral region of the neural plate and abut epidermal territory near the dorsal side of the blastopore in early gastrulae (Fig. 1), pLNPC-derived cells are found near the tip of the tail region in later embryos (Fig. 2f). The pLNPCs also contribute to muscle cells, in which *Tbx6-r.b* is expressed (Fig. 2a) and *Tbx6-r.a*, a paralogue of *Tbx6-r.b*, is expressed in b9.33 and b9.34 (Extended Data Fig. 5). In addition, anterior LNPCs (aLNPCs; b9.37 and b9.38) contribute to the dorsal row of the nerve cord (Fig. 2f)[19]. These features of LNPCs are evocative of vertebrate NMPs.

Tbx6 is expressed in NMP-derived mesodermal cells of vertebrate embryos and represses *Sox2*, which is highly expressed in NMP-derived neural cells[15,16,18]. In ascidian early embryos, *Sox1/2/3* is expressed in animal hemisphere cells including b6.5, which gives rise to LNPCs[30,40,42]. Indeed, *Sox1/2/3* was expressed in pLNPCs of middle gastrula embryos (Extended Data Fig. 6a), in which *Tbx6-r.a* and *Tbx6-r.b* are expressed (Fig. 2a and Extended Data Fig. 5). However, this *Sox1/2/3* expression became weak at the early neurula stage (Extended Data Fig. 6b). Therefore, we used a probe designed to hybridize with the first intron of *Sox1/2/3* to detect *Sox1/2/3* transcription. In normal unperturbed embryos, while nascent *Sox1/2/3* RNA was detected only in aLNPCs and in cells that contribute to the

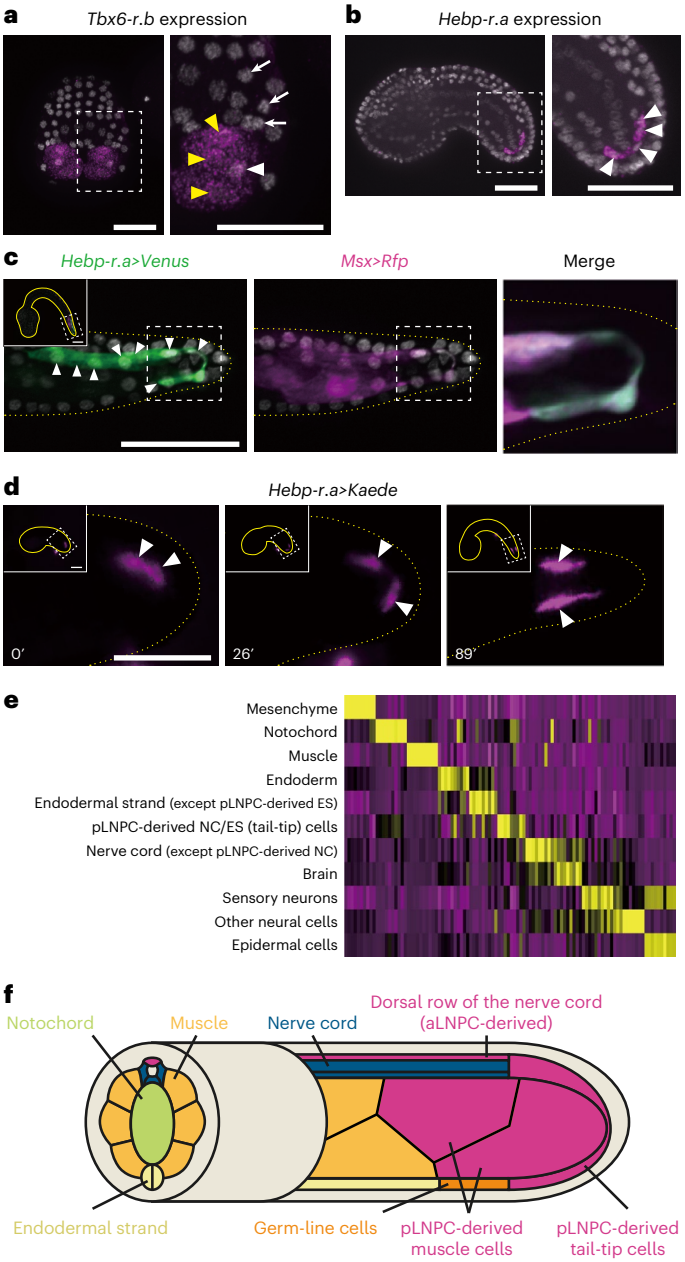


Fig. 2 | The cell lineage of b8.17 and b8.19 cells. a, *Tbx6-r.b* expression revealed by in situ hybridization in a middle gastrula embryo (magenta). A higher magnification view is shown on the right. *Tbx6-r.b* expression is observed in b9.34 (white arrowhead). Muscle cells are formed from other lineages and some of these cells also express *Tbx6-r.b* at this stage (yellow arrowheads). LNPCs other than b9.34, which do not express *Tbx6-r.b*, are indicated by white arrows. **b**, Expression of *Hebp-r.a* in a tailbud embryo (early tailbud I). Cells with *Hebp-r.a* signals are shown by magenta (arrowheads). **c**, Expression of reporter constructs which contain upstream regions of *Hebp-r.a* (green) and *Msx* (magenta) in the tip of the tail of a late tailbud embryo (late tailbud II). Overlaid photographs of signals (left) of the *Hebp-r.a* reporter and DAPI and (middle) of the *Msx* reporter and DAPI. Nuclei of eight cells labelled with *Hebp-r.a* reporter expression are shown by arrowheads in the left photograph. **d**, A tailbud embryo expressing *Hebp-r.a>Kaede* reporter was subjected to ultraviolet irradiation for photoconversion of Kaede fluorescence at the early tailbud stage (0 min; early tailbud I). One cell with photoconverted Kaede changed its location to the posterior end (26 min; middle tailbud I) and to the ventral side (89 min; late tailbud I). Note that only two cells are labelled because of mosaic incorporation of the reporter construct. Photographs at other time points are shown in Extended Data Fig. 2. **e**, A heatmap showing gene expression of various cell types. Expression levels were calculated as averages of expression levels of cells in each cell type. The top ten genes for each cell type are shown. Yellow and magenta represent high and low expression levels, respectively. Photographs in **a**–**d** are *z*-projected image stacks overlaid in pseudocolour. In **a**–**c**, nuclei are stained with DAPI (grey). Brightness and contrast of photographs in **c** and **d** were linearly adjusted. Scale bars, 50 µm. **f**, Depiction of the ascidian tail. LNPC-derived cells are coloured magenta. ES, endodermal strand; NC, nerve cord.

https://doi.org/10.1038/s41559-024-02387-8

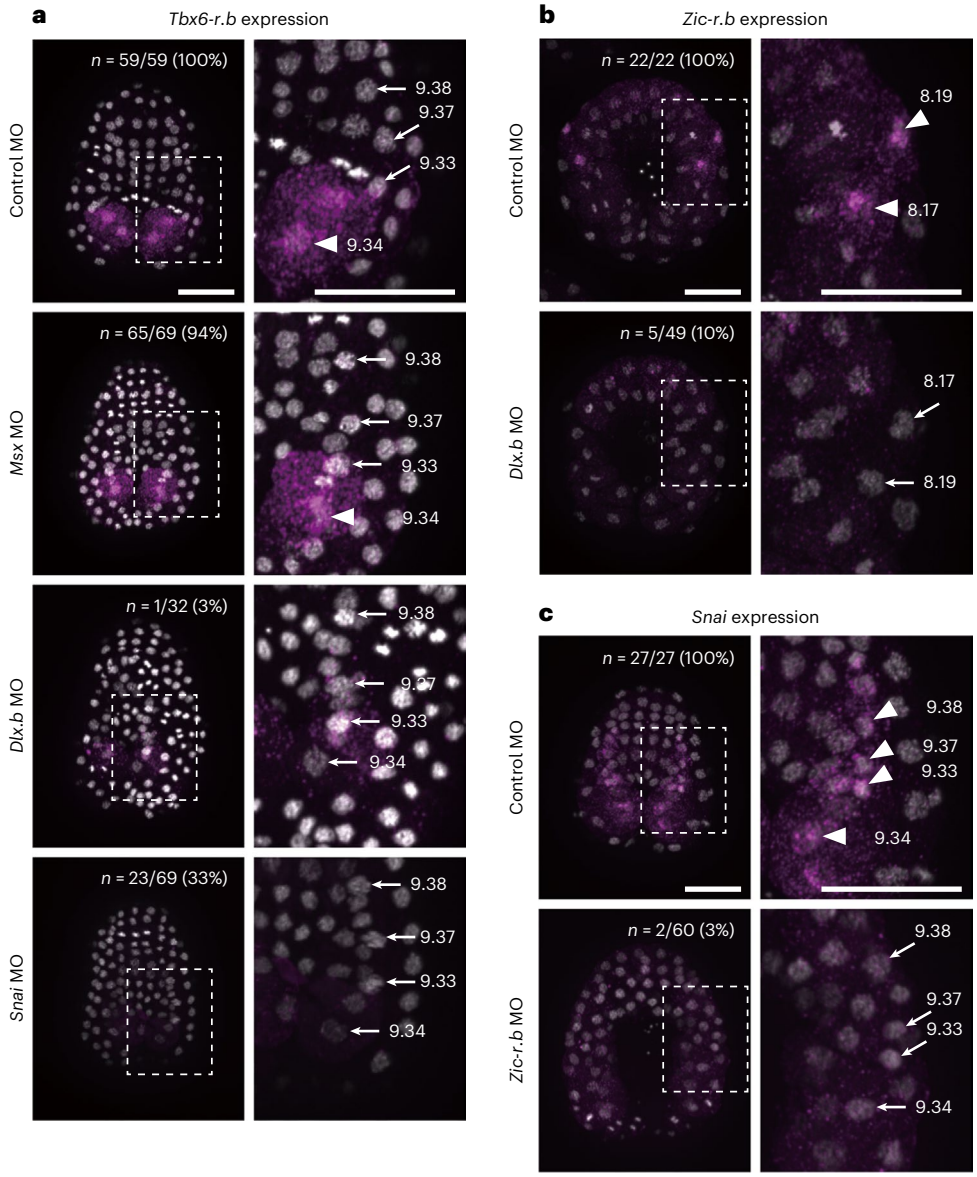


Fig. 3 | The regulatory gene circuit that activates *Tbx6-r.b* in b9.34. **a**, Expression of *Tbx6-r.b* in a control embryo and embryos injected with an MO against *Msx, Dlx.b* or *Snai*, at the early neurula stage. **b**, Expression of *Zic-r.b* in a control embryo and an embryo injected with the *Dlx.b* MO at the early gastrula stage. **c**, Expression of *Snai* in a control and an embryo injected with the *Zic-r.b* MO at the middle gastrula stage. Higher magnification views are shown on the right. LNPCs which express and do not express designated genes are indicated by arrowheads and arrows, respectively. Nuclei are stained with DAPI and are shown in grey. Photographs are *z*-projected image stacks overlaid in pseudocolour. Brightness and contrast of photographs in embryos showing *Tbx6-r.b* expression, except the *Snai* morphant embryo, were adjusted linearly. Numbers of embryos examined and embryos which expressed *Tbx6-r.b* in b9.34 (**a**), *Zic-r.b* in b8.17 and b8.19 (**b**) and *Snai* (**c**) in LNPCs are shown in the panels. In the *Msx* and *Dlx.b* morphants shown in **a**, signals for *Tbx6-r.b* expression in A-line muscle cells are seen over b9.33 cells, in which *Tbx6-r.b* is not expressed (Supplementary Fig. 2). Scale bars, 50 µm.

brain, it was not detected in pLNPCs (Extended Data Fig. 6c). When we overexpressed *Kaede* using the upstream regulatory region of *Msx* (*Msx>Kaede*) in LNPCs and their sibling b8.18/20-line cells as a control, expression of nascent *Sox1/2/3* transcripts was not changed (Fig. 4a). On the other hand, when we overexpressed *Tbx6-r.b* using the upstream regulatory region of *Msx* (*Msx>Tbx6-r.b*), nascent *Sox1/2/3* transcripts were not detected in LNPCs, whereas expression in cells that contribute to the brain, in which *Tbx6-r.b* was not overexpressed, was unaffected (Fig. 4a). *Lmx1*, which is normally expressed in aLNPCs of neurula embryos under control of *Msx*, *Nodal* and *Zic-r.b* and contributes to specification of the dorsal row of the nerve cord by activating *Pax3/7* (refs. 36,43), was also downregulated (Fig. 4b). Instead, these cells ectopically expressed a muscle marker gene, *Acta.a* (Fig. 4c). Thus, aLNPCs have the potential to become muscle cells and a key determinant for muscle fate, *Tbx6-r.b*, potentially represses *Sox1/2/3*.

We also examined *Sox1/2/3* expression in morphants of *Msx*, *Snai*, *Dlx.b* or *Zic-r.b* and found that *Dlx.b* and *Zic-r.b* but not *Msx* or *Snai*, are required for *Sox1/2/3* transcription in aLNPCs (Extended Data Fig. 7).

Similarity between ascidian LNPCs and vertebrate NMPs

The location of LNPCs in late embryos and the regulatory circuit of *Tbx6-r.b* and *Sox1/2/3* support the hypothesis that LNPCs in ascidian embryos share an evolutionary origin with vertebrate NMPs. To further confirm this hypothesis, we compared transcriptomes of LNPCs with those of zebrafish embryonic cells using scRNA-seq data published previously[34,44]. To perform a cross-species comparison, we identified orthologue groups using OrthoFinder[45] between genes of *Ciona*

https://doi.org/10.1038/s41559-024-02387-8

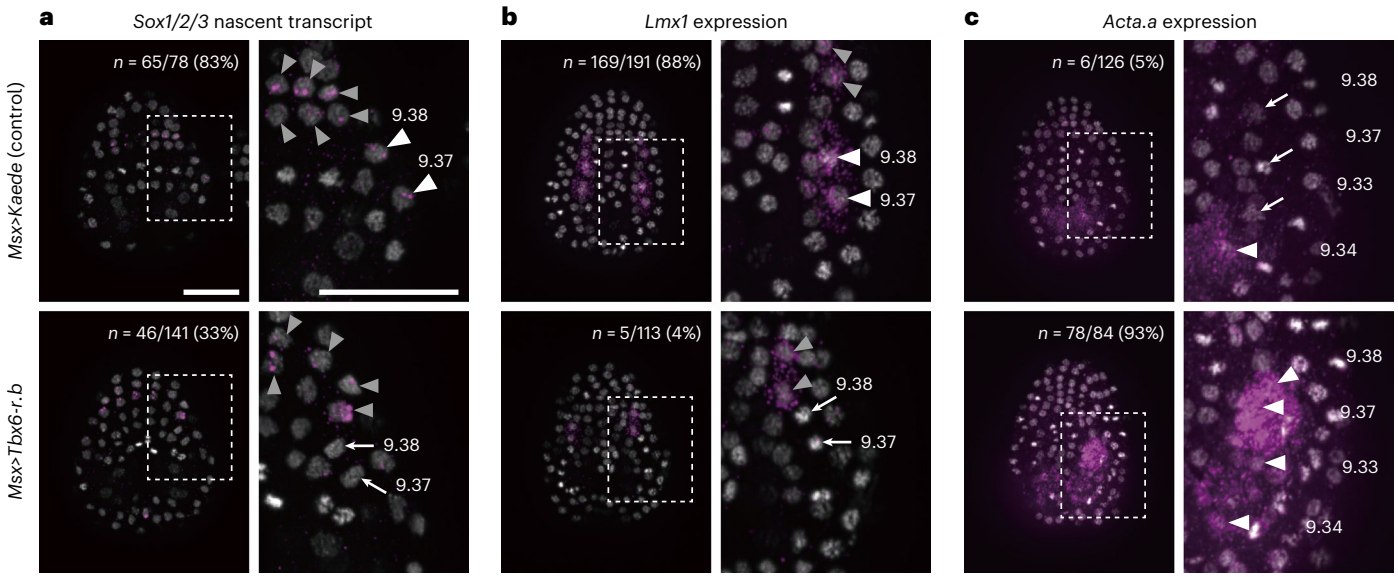


Fig. 4 | *Tbx6-r.b* can repress *Sox1/2/3* in aLNPCs and can change their fate to muscle. a, *Sox1/2/3* is transcribed in aLNPCs (b9.37 and b9.38; white arrowheads) and in cells that contribute to the central nervous system (grey arrowheads) of control middle gastrula embryos introduced with *Msx>Kaede*. On the other hand, *Sox1/2/3* is not transcribed in aLNPCs in embryos introduced with *Msx>Tbx6-r.b* (white arrows), while transcription in cells that contribute to the central nervous system is unaffected (grey arrowheads). A probe designed to hybridize with the first intron of *Sox1/2/3* was used to detect *Sox1/2/3* nascent transcripts and signals are typically seen as one or two spots in nuclei. **b**, Although *Lmx1* is expressed in aLNPCs (white arrowheads) of control early neurula embryos expressing *Msx>Kaede*, it is not expressed in these cells of embryos expressing *Msx>Tbx6-r.b* (white arrows). Note that *Lmx1* expression in cells that contribute to the central nervous system is unaffected (grey arrowheads). **c**, In control early neurula embryos introduced with *Msx>Kaede*, *Acta.a*, which encodes a muscle actin, is expressed in b9.34, while it is not expressed in the remaining LNPCs. In embryos introduced with *Msx>Tbx6-r.b*, *Acta.a* is expressed in all LNPCs. Arrowheads and arrows indicate LNPCs with and without *Acta.a* expression, respectively. Photographs are *z*-projected image stacks overlaid in pseudocolour; magenta, in situ hybridization signals; grey, nuclei stained with DAPI. Brightness and contrast of photographs were adjusted linearly. Numbers of embryos examined and embryos that expressed designated genes in LNPCs (**a**,**b**) and in b9.33/37/38 (**c**) are shown in the panels. Scale bars, 50 µm.

and zebrafish and compared ascidian LNPCs with all available cells of zebrafish embryos from 10 to 24 hours postfertilization (hpf). Ascidian notochord cells and muscle cells were also included as controls. We performed clustering of cells at three different resolutions (Supplementary Tables 2–5). At the highest resolution, all *Ciona* notochord cells and 56% of zebrafish notochord cells were found together in a single cluster and all *Ciona* muscle cells and 97% of zebrafish myotome and muscle cells were found together in three clusters (Extended Data Fig. 8). This observation indicates that our cross-species comparison successfully grouped cells according to cell types.

Similarly, at the middle resolution, all *Ciona* notochord cells and 83% of zebrafish notochord cells were found together in a single cluster and all *Ciona* muscle cells and 99% of zebrafish myotome and muscle cells were found together in a single cluster (Fig. 5a,b and Supplementary Figs. 3 and 4). At this resolution, pLNPCs (b9.33 and b9.34) and tail-tip cells (b9.33 derivatives) were found in cluster 13 (Fig. 5a,c,e). This cluster contains presomitic-mesoderm cells in the tailbud of zebrafish embryos (Fig. 5a,d,e). Zebrafish tailbud presomitic-mesoderm cells were included not only in cluster 13 but also in cluster 9 and cluster 12 (Fig. 5e). Cluster 12 mainly consists of myotome and muscle cells. Because almost all cells (99.8%) in clusters 9 and 13 constitute a single cluster at a low resolution (Extended Data Fig. 9 and Supplementary Table 2), clusters 9 and 13 are close to each other. These results indicate a close relationship between ascidian b9.33/b9.34/pLNPC-derived cells and zebrafish tailbud-presomitic-mesoderm cells.

Nearly half of b9.37 and b9.38 cells (aLNPCs) of middle gastrula embryos (42%) were found in cluster 13, which included b9.33, b9.34 (pLNPCs) and tail-tip cells, and almost all of the remaining cells (52%) were found in cluster 4 (Fig. 5a,c,e). At the early neurula stage, 86% of b9.37 and b9.38 cells were found in cluster 4 (Fig. 5a,c,e). At later stages, all b9.37/b9.38 derivatives (cells in the dorsal row of the nerve cord) were found in cluster 4 (Fig. 5a,c,e). This cluster contains spinal-cord cells of the tailbud of zebrafish embryos (Fig. 5a,d,e). In fact, among all spinal-cord cells of the tailbud of zebrafish embryos, 39% of cells were found in cluster 4 (Fig. 5e). This result indicates a close relationship between ascidian b9.37/b9.38-lineage cells and zebrafish tailbud spinal-cord cells. In other words, our analysis using single-cell transcriptome data supports the hypothesis that LNPCs have properties similar to those of zebrafish NMPs, which reside in the tailbud and contribute to the spinal cord and somites.

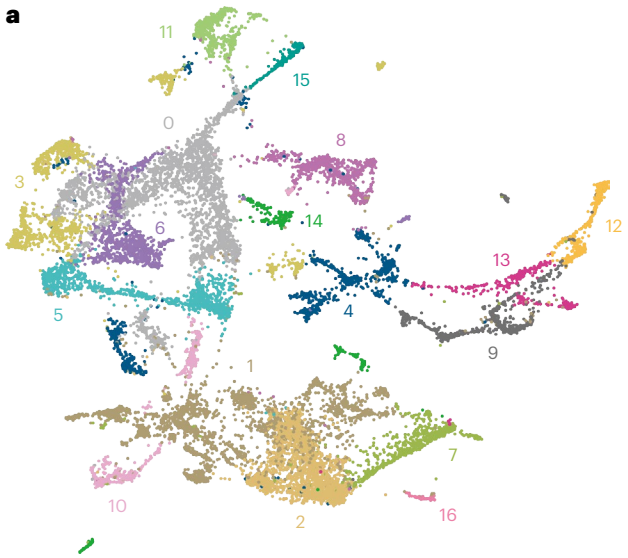
Discussion

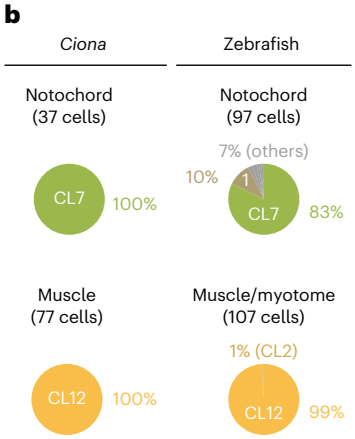
Our data show the similarity between ascidian LNPCs and vertebrate neural plate border cells which give rise to neural-crest cells. This suggests that ascidian LNPCs share an evolutionary origin with vertebrate neural-crest cells. First, both ascidian and vertebrate cell populations arise in the lateral border (or lateral side) of the neural plate. Second, most genes for transcription factors which specify the neural plate border and neural-crest cells in vertebrates are expressed in these ascidian cells. These genes include *Dlx.b*, *Msx*, *Snai*, *Zic-r.b*, *Tfap2-r.b*, *Pax3/7*, *Ets1/2.b*, *Lmx1* and *Id.b* and we showed that the gene circuit of *Dlx.b*, *Zic-r.b* and *Snai* regulates *Tbx6-r.b* in pLNPCs. In addition, in aLNPCs, *Msx* and *Zic-r.b* regulate *Lmx1* and *Lmx1* regulates *Pax3/7* (ref. 36).

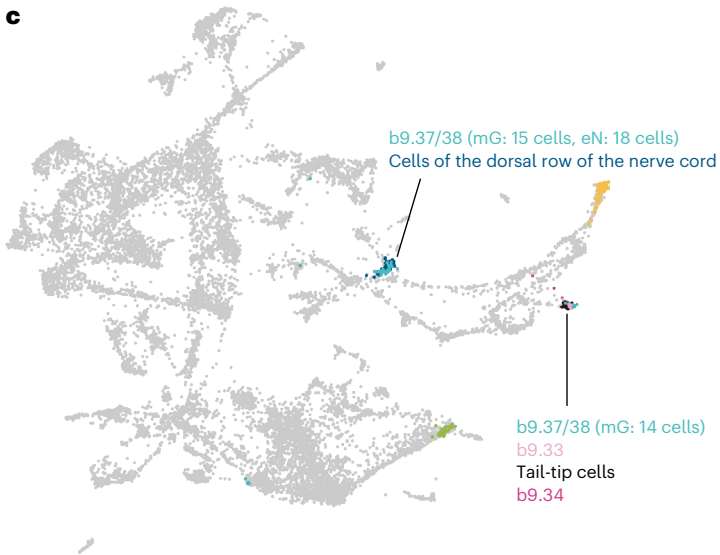
LNPCs have been regarded as part of the neural plate[19,21]. However, on the basis of our data, as well as the observation that LNPCs do not contribute to central nervous system neurons[19,21], we propose to call LNPCs lateral neural plate border cells. Because their sibling b8.18/20 lineage contributes to sensory neurons in the tail and because these cells may share an evolutionary origin with vertebrate neural-crest cells[7,8], the entire b6.5-lineage could be a cell population that shares an evolutionary origin with neural plate border cells of vertebrates.

On the other hand, our data also showed the similarity between ascidian LNPCs and vertebrate NMPs. *Sox1/2/3* is initially expressed in all animal hemisphere cells, including cells giving rise to LNPCs[30,40,42],

https://doi.org/10.1038/s41559-024-02387-8







- *Ciona* b9.37/38
- *Ciona* cells of the dorsal row of the nerve cord (b9.37/38 derivatives)
- *Ciona* b9.33
- *Ciona* tail-tip cells (b9.33 derivatives)
- *Ciona* b9.34
- *Ciona* muscle cells (including b9.34 derivatives)
- *Ciona* notochord cells
- Zebrafish cells

- Zebrafish tailbud—presomitic mesoderm
- Zebrafish tailbud—spinal cord
- Zebrafish muscle/myotome
- Zebrafish notochord
- Zebrafish other cells
- *Ciona* cells

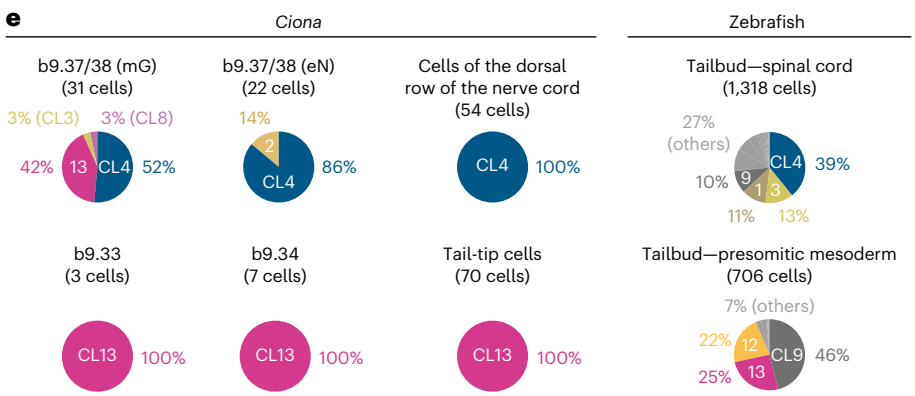


Fig. 5 | A cross-species comparison of single-cell transcriptome data indicates affinities between ascidian LNPCs and tailbud cells of zebrafish embryos. a, A UMAP plot that includes single-cell transcriptome data of 10–24 hpf embryos of zebrafish and LNPCs and their derivatives in ascidian embryos from the middle gastrula to larval stages. For controls, muscle cells and notochord cells of ascidian middle tailbud embryos are also included. Different clusters are indicated by different colours. **b**, Pie charts show that notochord and muscle cells of *Ciona* and zebrafish are largely grouped into the same clusters. **c**, *Ciona* LNPCs, muscle cells and notochord cells are coloured in the same UMAP as that shown in **a**. **d**, Zebrafish tailbud cells, muscle cells, notochord cells are coloured in the same UMAP as that shown in **a**. **e**, Proportions of clusters which include designated cells are shown by pie charts. mG, middle gastrula; eN, early neurula.

and our data indicate that it is downregulated in pLNPCs. In vertebrate NMPs, cells with strong *Tbx6* expression contribute to mesodermal cells and cells with strong *Sox2* expression contribute to the posterior part of the neural tube[13,16–18]. In ascidian embryos, b9.34 expresses *Tbx6-r.b* and gives rise to muscle cells. Another pLNPC, b9.33, expresses *Tbx6-r.a* but produces tail-tip cells. On the other hand, aLNPCs continue to express *Sox1/2/3* and give rise to the nerve cord. The ascidian larval tail nerve cord does not contain neurons[46] and it is debatable which part of the vertebrate central nervous system corresponds to the ascidian nerve cord[23,28,36,47,48]. Nevertheless, cells near the tip of the tail with *Sox2* (or its orthologue) expression contribute to the posterior part of the central nervous system in both ascidian and vertebrate embryos. In addition, in both embryos, *Tbx6* (or its orthologue) negatively regulates *Sox2* (or its orthologue) and promotes mesodermal fate. These shared properties indicate a common evolutionary origin of vertebrate NMPs and ascidian LNPCs, as previously discussed in part[28]. Consistently, our cross-species comparison of zebrafish and *Ciona* scRNA-seq data showed similarity between ascidian LNPCs and vertebrate NMPs. Intriguingly, anterior cells in the vegetal hemisphere of ascidian embryos also give rise to ectodermal (nerve cord) and mesodermal (notochord) cells[19], although it is not clear whether these cells represent ancestral states.

While NMPs are characterized by co-expression of *Sox2* and *T* in vertebrate embryos[13], *T* is not expressed in LNPCs or in any cells with *Sox1/2/3* expression in ascidian embryos[49]. *T* is expressed near the blastopore in many animals, including deuterostomes and protostomes and is used as a pan-mesodermal marker in vertebrate embryos[50]. In ascidian embryos, this conserved expression is lost and *T* is expressed exclusively in the notochord lineage[49]. Therefore, it is possible that ancestral cells of ascidian LNPCs and vertebrate NMPs expressed *T* and that this expression was lost in the ascidian lineage, although it is equally possible that this *T* expression was acquired in the vertebrate lineage. Nevertheless, data in the present study strongly indicate that the evolutionary origin of NMPs dates back to a common ancestor of ascidians and vertebrates (ancestral Olfactores).

Thus, LNPCs have properties of vertebrate neural-crest cells and NMPs. This suggests that the ancestral Olfactores may have had a cell population resembling ascidian LNPCs and that such ancestral cells may have evolved into neural-crest cells and NMPs in the vertebrate lineage. Meanwhile, because cells of the LNPC-sibling b8.18/20 lineage are thought to be homologues of neural-crest cells[7,8], it is possible that the ancestral Olfactores also had additional neural-crest-like cells which did not have NMP-like properties.

While ascidian neural-crest-like cells identified in previous studies produce sensory neurons and pigment cells[6–8], LNPCs give rise not only to cells that are normally of ectodermal origin but also cells that are normally of mesodermal origin. Contrary to the previous hypothesis[51–54], this observation indicates that the evolutionary origin of multipotency of vertebrate neural-crest cells may date back to the ancestral Olfactores, although our results are not mutually exclusive to the hypothesis[55] that various mesodermal developmental programmes have been co-opted under gene regulatory networks for neural-crest cells and NMPs in the vertebrate lineage after the spilt of vertebrates and tunicates.

Several studies have shown a close relationship between trunk neural-crest cells and NMPs using human pluripotent cultured cells[56–58]. In addition, trunk neural-crest cells are thought to have emerged earlier than other neural-crest cells[59,60]. Therefore, ancestral cells of neural-crest cells may have had properties of NMPs and ancient Olfactores may have had such cells. *Ciona* LNPCs, which have properties of neural-crest cells and NMPs, may represent such an ancestral state and multipotency of these two cell populations may share a common evolutionary origin.

Methods

Animals and gene identifiers

Adult specimens of *C. intestinalis* (type A; also called *C. robusta*) were obtained from the National BioResource Project for *Ciona*. Complementary DNA clones were obtained from our EST clone collection[31]. Identifiers for genes examined in this study are as follows: *Tbx6-r.b*, KY21.Chr11.465/466/467; *Hebp-r.a*, KY21.Chr10.215; *Hand-r*, KY21.Chr1.1701; *Tbx6-r.a*, KY21.Chr11.458; *Msx*, KY21.Chr2.1031; *Dlx.b*, KY21.Chr7.361; *Snai*, KY21.Chr3.1356; *Zic-r.b*, KY21.Chr6.26/27/28/29/30/31; *Sox1/2/3*, KY21.Chr1.254; *Lmx1*, KY21.Chr9.606; *Acta.a*, KY21.Chr1.1745; *Tfap2-r.b*, KY21.Chr7.1145; *Pax3/7*, KY21.Chr10.288; *Ets1/2.b*, KY21.Chr10.346; *Id.b*, KY21.Chr7.1129; *Celf3.a*, KY21.Chr6.58; and *Slc39a-related*, KY21.Chr4.1089. Identifiers for the latest KY21 set[61] are shown.

Functional assays

Sequences of MOs (morpholino anitsense oligonucleotides) that block translation, were as follows: *Dlx.b*, 5′-TCGGAGATTCAACGACGCTTGACAT-3′; *Zic-r.b*, 5′-GATCAACCATTACATTAGAATACAT-3′; *Snai*, 5′-GTCATGATGTAATCACAGTAATATA-3′; and *Msx*, 5′-ATTCGTTTACTGTCATTTTTAATTT-3′. These MOs were used and evaluated in previous studies[8,25,36,40,43,62,63]. They were microinjected under a microscope. For overexpression, coding sequences of *Tbx6-r.b* were cloned into the downstream region of the upstream regulatory region of *Msx* (chr2:6132658–6133440). Overexpression constructs were introduced by electroporation[64]. All functional assays were performed at least twice with different batches of embryos. Cells were identified under a microscope according to previous studies (Supplementary Fig. 5)[21,65,66].

In situ hybridization and reporter assay

Embryos were fixed in 4% paraformaldehyde in 0.1 M MOPS buffer (pH 7.5) and 0.5 M NaCl at 4 °C for over 16 h. After washing with 80% ethanol and phosphate-buffered saline containing 0.1% Tween 20 (PBST), embryos were incubated in 3% H_2O_2 for 60 min and again washed with PBST. Then, embryos were treated with 2 μg ml^{-1} of proteinase K in PBST for 30 min at 37 °C and washed with PBST. Embryos were again fixed with 4% paraformaldehyde in PBST for 1 h at room temperature, washed with PBST and immersed in hybridization buffer for at least 1 h at 50 °C. Then, hybridization buffer was replaced with fresh hybridization buffer containing a digoxigenin- and/or a fluorescein-labelled probe and embryos were incubated at 50 °C for at least 16 h. Hybridization buffer contained 50% formamide, 5× SSC, 100 μg ml^{-1} of yeast transfer RNA, 5× Denhart's solution and 0.1% Tween 20. After hybridization, embryos were washed twice at 50 °C for 15 min in 2× SSC, 50% formamide and 0.1% Tween 20. Embryos were treated for 30 min at 37 °C with 20 μg ml^{-1} of RNase A in 10 mM Tris buffer (pH 8.0), 0.5 M NaCl, 5 mM EDTA and 0.1% Tween 20. Embryos were further washed twice at 50 °C for 15 min in 0.5× SSC, 50% formamide and 0.1% Tween 20 and the washing solution was replaced with PBST. Embryos were blocked at room temperature for 60 min with 0.1% BSA in PBST and then exposed to mouse anti-digoxigenin (Roche, 11333062910; 1:100

dilution) or rabbit anti-fluorescein (Abcam, ab19491; 1:1,000 dilution) diluted in Can-Get-Signal Immunostain Immunoreaction enhancer solution A (Toyobo, NKB-501) at 4 °C overnight. After washing with PBST, embryos were incubated with HRP-conjugated secondary antibody (Thermo Fisher, B40961 and B40922; 1:1 dilution). Embryos were washed with PBST and TNT (100 mM Tris buffer (pH 7.5), 150 mM NaCl and 0.1% Tween 20). Then, embryos were stained with Alexa Fluor 488 or 555 tyramide reagent (Thermo Fisher, B40953 and B40955). For two-colour in situ hybridization, two probes were simultaneously added to the hybridization solution and the secondary-antibody reaction and detection with tyramide reagents were repeated twice. Before observation with a microscope, embryos were treated with TrueVIEW Autofluorescence Quenching Kit (Vector Laboratories, SP-8400-15).

The upstream region that is contained in reporter of *Hebp-r.a>Venus* or *Hebp-r.a>Kaede* is chr10:1456230–1457052. The upstream region of *Msx>RFP* that promotes expression in LNPCs and their sibling lineage of cells[67] is chr2:6132658–6133440. These reporter constructs were introduced by electroporation[64] or microinjection. Expression of fluorescent proteins was examined under fluorescence microscope.

Immunostaining

For immunostaining of GFP and RFP, embryos were fixed with 3.7% formaldehyde in 0.1 M MOPS buffer (pH 7.5) and 0.5 M NaCl for 5 min at room temperature. After washing with PBSTr (phosphate-buffered saline containing 0.2% Triton-X-100) and PBSTT (phosphate-buffered saline containing 0.2% Triton-X-100 and 0.4% Tween 20), embryos were incubated in blocking buffer (10% BSA in PBSTr) for 30 min and then with rabbit and mouse antibodies against GFP (1:250, Thermo Fisher, A6455) and RFP (1:250, Medical & Biological Laboratories, M1553) for 6 h at 4 °C. Then, embryos were washed with PBSTr and incubated with secondary antibodies (Alexa Fluor 488-conjugated anti-rabbit, 1:400, Thermo Fisher, A21206; Alexa Fluor 555-conjugated anti-mouse, 1:400, Thermo Fisher, A31570) for 6 h at 4 °C. Embryos were washed with PBSTr and mounted on slide glass using Vectashield mounting medium with DAPI (Vector Laboratories, H-1200).

Data analysis

To compare gene expression profiles among *Ciona* larval tissues, we used single-cell transcriptome data for *Ciona* larvae which were published previously[34] (SRA accession nos. SRR9051005 to SRR9051007). Data were mapped to the latest genome assembly (HT version)[68] and the latest version of the gene model set (KY21 version)[61] using Cell Ranger software (v.6.12, 10x Genomics). Then, mesenchymal cells, notochord cells, muscle cells, endodermal cells, endodermal-strand cells, tail-tip cells, nerve-cord cells, brain cells, sensory neurons, other neural cells and epidermal cells were annotated using Loupe Browser (10x Genomics) (Supplementary Table 6). For cell identification during the larval stage, we used *Hebp-r.a*, *Hand-r*, KY21.Chr8.1029 for tail-tip cells, KY21.Chr1.1899 and KY21.Chr1.1498 for mesenchyme[69,70], KY21.Chr6.610 (*Noto1*)[71] for notochord, KY21.Chr 11.773 (*Myh.f*) and *Acta.a* for muscle[72], KY21.Chr1.2012 (*Epi1*) and KY21.Chr7.872 (*Epib*) for epidermal cells[73], KY21.Chr9.123 (*Gnrh2*) and KY21.Chr10.869 for nerve cord[70,74], KY21.Chr11.800 (*Rlbp1*) for brain[75] and KY21.Chr2.456 (*Pou4*) for sensory neurons[30]. Other neural cells were identified by expression of KY21.Chr11.1137 and KY21.Chr2.49 (ref. 34), KY21.Chr6.690 (*Pax2/5/8.a*)[36] or KY21.Chr2.1203 (*Syt*)[76]. Endodermal cells were identified by expression of KY21.Chr6.222 (*Alp*) and KY21.Chr14.805 (ref. 32) and endodermal-strand cells were identified as cells expressing KY21.Chr14.805, without expression of *Alp*[32,77]. A heatmap containing the top 10 genes for each of 11 tissues was generated with Seurat[78] (v.4.1.3) as follows. Expression matrices of three larval samples were loaded into three Seurat objects. We normalized each dataset using the 'NormalizeData' function with default parameters and calculated 800 highly variable features using the Seurat 'FindVariableFeatures' function with default parameters on the basis of plots obtained with the 'VariableFeaturesPlot' function. A list of three Seurat objects was used for selecting features that are variable across these datasets using the 'SelectIntegrationFeatures' function with default parameters. Then, the 'FindIntegrationAnchors' function was used to identify anchors for data integration with default parameters. Using these anchors, the above datasets were integrated using the 'IntegrateData' function with default parameters except 'k.weight = 10'. The assay mode was set to 'integrated' with the 'DefaultAssay' function and data were scaled with the 'ScaleData' function. We chose only cells that we annotated as mesenchymal cells, notochord cells, muscle cells, endodermal cells, endodermal-strand cells, tail-tip cells, nerve-cord cells, brain cells, sensory neurons, other neural cells or epidermal cells, as mentioned above. Then, we tried to find genes differentially expressed in each of these cell types using the 'FindAllMarkers' function with default parameters except 'only.pos = TRUE' and 'min.pct = 0.25'. Then we picked the top ten genes for each of these cell types. After calculating averages among cells with the 'AverageExpression' function, we created a heatmap with 'DoHeatmap' function.

For a cross-species comparison between *Ciona* transcriptomes and zebrafish transcriptomes, single-cell transcriptome data for *Ciona* middle gastrula embryos to larvae, which were published previously[34], were used (SRA accession numbers: SRR9050988, SRR9050989, SRR9050992, SRR9050997, SRR9050998, SRR9051003 and SRR9051007). Data were mapped to the latest genome assembly (HT version)[68] and the latest version of the gene model set (KY21 version)[61] using Cell Ranger software (v.6.12. 10x Genomics). For zebrafish 10–24 h embryos, mapping data to gene models were downloaded from the GEO database (we used CSV files of UMI-filtered counts under accession numbers, GSM3067192 to GSM3067195; GSM3067192_10hpf.csv.gz, GSM3067193_14hpf.csv.gz, GSM3067194_18hpf.csv.gz and GSM3067195_24hpf.csv.gz)[44].

Annotations of zebrafish cells published previously[44] were used. We annotated *Ciona* cells using Loupe Browser (10x Genomics). Tail-tip cells, muscle and notochord cells in middle tailbud embryos were identified using the same marker genes that were used in larvae. At the middle gastrula and early neurula stages, we identified b9.34 as cells expressing *Msx*, *Tbx6-r.b* and *Snai* and b9.33 as cells expressing *Msx*, *Snai* and *Hox12* but not *Pax3/7* or *Tbx6-r.b*. These cells were further curated manually and are listed in Supplementary Table 7.

For comparisons with zebrafish transcriptomes, we first compared proteomes of *Ciona* (KY21 version[61]) and zebrafish (Refseq[79] proteome file named GCF_000002035.6_GRCz11_protein.faa) using OrthoFinder[45]. We took ten protein groups from the *Ciona* proteome to validate grouping by OrthoFinder and found that comparisons between *Ciona* and zebrafish proteomes, which yielded 5,422 groups, were not satisfactory (Supplementary Table 8). According to the manual of Orthofinder, we included proteomes of *Strongylocentrotus purpuratus* (sea urchin) and *Drosophila melanogaster* as controls to obtain better resolution. Proteome files we used were: 'HT.KY21Gene.protein.2.fasta.zip' for *Ciona*, which includes a proteome set derived from the latest KY21 gene model set[61], downloaded from the Ghost database[31]; 'GRCz11_protein.faa' for the zebrafish proteome, downloaded from the NCBI Refseq database[79]; 'GCF_000002235.5_Spur_5.0_translated_cds.faa.gz' was downloaded for the sea urchin proteome from the NCBI Refseq database[79]; and 'Drosophila_melanogaster.BDGP6.32.pep.all.fa.gz' for the fly proteome, downloaded from the Ensembl database (release 108)[80]. From these data, OrthoFinder yielded 6,379 orthologue groups which contained both ascidian and zebrafish proteins (Supplementary Tables 9 and 10). We confirmed that proteins in the above control ten groups are resolved better (Supplementary Table 8). When several genes for one species were included in one orthologue group, all read counts in that group were summed.

Data were analysed with Seurat[78] (v.4.1.3). After converting all gene names to orthologue-group IDs, zebrafish data were loaded into a single Seurat[78] object (listed in Supplementary Table 2; 13,951 cells).

We normalized the zebrafish data using the NormalizeData function with default parameters and calculated 800 highly variable features using the Seurat FindVariableFeatures function with default parameters on the basis of a plot obtained with the VariableFeaturesPlot function. *Ciona* expression matrices were individually loaded into seven Seurat objects because these were not normalized against one another (middle gastrula, 2,296 cells; early neurula, 1,916 cells; late neurula, 6,177 cells; middle tailbud II replica 1, 5,233 cells; middle tailbud II replica 2, 4,329 cells; late tailbud II, 5,277 cells; and larva, 6,459 cells). On the basis of scatter plots created with the FeatureScatter function, we chose cells that have unique counts within the ranges shown in Supplementary Table 11 using the subset function (middle gastrula, 1,727 cells; early neurula, 1,309 cells; late neurula, 1,600 cells; middle tailbud II replica 1, 2,254 cells; middle tailbud II replica 2, 1,853 cells; late tailbud II, 2,221 cells; and larva, 4,007 cells). We normalized the data using the NomalizeData function with default parameters and calculated 300 highly variable features similarly (this value was determined on the basis of plots obtained with the VariableFeaturesPlot function). A list of seven *Ciona* Seurat objects and one zebrafish Seurat object was used for selecting features that are variable across these datasets using the SelectIntegrationFeatures function with default parameters. Then, the FindIntegrationAnchors function was used to identify anchors for data integration with default parameters. Using these anchors, the above eight datasets were integrated using the IntegrateData function with default parameters except 'k.weight = 10'. From the integrated object, we removed irrelevant *Ciona* cells (cells other than b9.33, b9.34, b9.37, b9.38, muscle, notochord and tail-tip cells; the remaining 301 cells are listed in Supplementary Table 7). Note that we did not remove any zebrafish cells at this stage. The assay mode was set to 'integrated' with the DefaultAssay function. After scaling the data with the ScaleData function, a principal component analysis was performed with the RunPCA function with default parameters (npc = 50). To cluster cells, we used the FindNeighbors function with the parameter 'dims = 1:50' and the FindCluster function with the parameter 'resolution = 10' (high resolution), 'resolution = 0.25' (middle resolution) and 'resolution = 0.2' (low resolution). Data were visualized with the RunUMAP function with default parameters.

All matrices used for the transcriptome comparison between *Ciona* and zebrafish and R-markdowns for analyses using Seurat are available from Zenodo (https://doi.org/10.5281/zenodo.10682771)[81].

Reporting summary

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

Data availability

All data generated during this study are included in this article and Zenodo (https://doi.org/10.5281/zenodo.10682771)[81].

Code availability

No custom-made programmes were used in the present study. R-markdowns for analyses using Seurat are available from Zenodo (https://doi.org/10.5281/zenodo.10682771)[81].

References

1. Tang, W. Y. & Bronner, M. E. Neural crest lineage analysis: from past to future trajectory. *Development* **147**, dev193193 (2020).
2. Zalc, A. et al. Reactivation of the pluripotency program precedes formation of the cranial neural crest. *Science* **371**, eabb4776 (2021).
3. Buitrago-Delgado, E., Nordin, K., Rao, A., Geary, L. & LaBonne, C. Shared regulatory programs suggest retention of blastula-stage potential in neural crest cells. *Science* **348**, 1332–1335 (2015).
4. Gans, C. & Northcutt, R. G. Neural crest and the origin of vertebrates—a new head. *Science* **220**, 268–273 (1983).
5. Simoes-Costa, M. & Bronner, M. E. Establishing neural crest identity: a gene regulatory recipe. *Development* **142**, 242–257 (2015).
6. 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).
7. Stolfi, A., Ryan, K., Meinertzhagen, I. A. & Christiaen, L. Migratory neuronal progenitors arise from the neural plate borders in tunicates. *Nature* **527**, 371–374 (2015).
8. Waki, K., Imai, K. S. & Satou, Y. Genetic pathways for differentiation of the peripheral nervous system in ascidians. *Nat. Commun.* **6**, 8719 (2015).
9. Horie, R. et al. Shared evolutionary origin of vertebrate neural crest and cranial placodes. *Nature* **560**, 228–232 (2018).
10. Papadogiannis, V. et al. *Hmx* gene conservation identifies the origin of vertebrate cranial ganglia. *Nature* **605**, 701–705 (2022).
11. Tzouanacou, E., Wegener, A., Wymeersch, F. J., Wilson, V. & Nicolas, J. F. Redefining the progression of lineage segregations during mammalian embryogenesis by clonal analysis. *Dev. Cell* **17**, 365–376 (2009).
12. Martin, B. L. & Kimelman, D. Canonical Wnt signaling dynamically controls multiple stem cell fate decisions during vertebrate body formation. *Dev. Cell* **22**, 223–232 (2012).
13. Henrique, D., Abranches, E., Verrier, L. & Storey, K. G. Neuromesodermal progenitors and the making of the spinal cord. *Development* **142**, 2864–2875 (2015).
14. Olivera-Martinez, I., Harada, H., Halley, P. A. & Storey, K. G. Loss of FGF-dependent mesoderm identity and rise of endogenous retinoid signalling determine cessation of body axis elongation. *PLoS Biol.* **10**, e1001415 (2012).
15. Gouti, M., Metzis, V. & Briscoe, J. The route to spinal cord cell types: a tale of signals and switches. *Trends Genet.* **31**, 282–289 (2015).
16. Takemoto, T. et al. Tbx6-dependent Sox2 regulation determines neural or mesodermal fate in axial stem cells. *Nature* **470**, 394–398 (2011).
17. Goto, H., Kimmey, S. C., Row, R. H., Matus, D. Q. & Martin, B. L. FGF and canonical Wnt signaling cooperate to induce paraxial mesoderm from tailbud neuromesodermal progenitors through regulation of a two-step epithelial to mesenchymal transition. *Development* **144**, 1412–1424 (2017).
18. Martin, B. L. & Steventon, B. A fishy tail: insights into the cell and molecular biology of neuromesodermal cells from zebrafish embryos. *Dev. Biol.* **487**, 67–73 (2022).
19. 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).
20. Nicol, D. & Meinertzhagen, I. A. Development of the central nervous system of the larva of the ascidian, *Ciona intestinalis* L. I. The early lineages of the neural plate. *Dev. Biol.* **130**, 721–736 (1988).
21. Nicol, D. & Meinertzhagen, I. A. Development of the central nervous system of the larva of the ascidian, *Ciona intestinalis* L. II. Neural plate morphogenesis and cell lineages during neurulation. *Dev. Biol.* **130**, 737–766 (1988).
22. Mazet, F. et al. Molecular evidence from *Ciona intestinalis* for the evolutionary origin of vertebrate sensory placodes. *Dev. Biol.* **282**, 494–508 (2005).
23. Wada, H., Saiga, H., Satoh, N. & Holland, P. W. Tripartite organization of the ancestral chordate brain and the antiquity of placodes: insights from ascidian *Pax-2/5/8*, *Hox* and *Otx* genes. *Development* **125**, 1113–1122 (1998).
24. Abitua, P. B. et al. The pre-vertebrate origins of neurogenic placodes. *Nature* **524**, 462–465 (2015).

25. Ikeda, T., Matsuoka, T. & Satou, Y. A time delay gene circuit is required for palp formation in the ascidian embryo. *Development* **140**, 4703–4708 (2013).
26. Liu, B. & Satou, Y. Foxg specifies sensory neurons in the anterior neural plate border of the ascidian embryo. *Nat. Commun.* **10**, 4911 (2019).
27. Wagner, E. & Levine, M. FGF signaling establishes the anterior border of the *Ciona* neural tube. *Development* **139**, 2351–2359 (2012).
28. Hudson, C. & Yasuo, H. Neuromesodermal lineage contribution to CNS development in invertebrate and vertebrate chordates. *Genes* **12**, 592 (2021).
29. 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).
30. Imai, K. S., Hino, K., Yagi, K., Satoh, N. & Satou, Y. Gene expression profiles of transcription factors and signaling molecules in the ascidian embryo: towards a comprehensive understanding of gene networks. *Development* **131**, 4047–4058 (2004).
31. Satou, Y., Kawashima, T., Shoguchi, E., Nakayama, A. & Satoh, N. An integrated database of the ascidian, *Ciona intestinalis*: towards functional genomics. *Zool. Sci.* **22**, 837–843 (2005).
32. Satou, Y. et al. Gene expression profiles in *Ciona intestinalis* tailbud embryos. *Development* **128**, 2893–2904 (2001).
33. Satou, Y., Imai, K. & Satoh, N. The ascidian Mesp gene specifies heart precursor cells. *Development* **131**, 2533–2541 (2004).
34. Cao, C. et al. Comprehensive single-cell transcriptome lineages of a proto-vertebrate. *Nature* **571**, 349–354 (2019).
35. Roure, A. & Darras, S. *Msxb* is a core component of the genetic circuitry specifying the dorsal and ventral neurogenic midlines in the ascidian embryo. *Dev. Biol.* **409**, 277–287 (2016).
36. Imai, K. S., Stolfi, A., Levine, M. & Satou, Y. Gene regulatory networks underlying the compartmentalization of the *Ciona* central nervous system. *Development* **136**, 285–293 (2009).
37. Aniello, F. et al. Identification and developmental expression of *Ci-msxb*: a novel homologue of *Drosophila msh* gene in *Ciona intestinalis*. *Mech. Dev.* **88**, 123–126 (1999).
38. Yagi, K. & Makabe, K. W. Isolation of an early neural maker gene abundantly expressed in the nervous system of the ascidian, *Halocynthia roretzi*. *Dev. Genes Evol.* **211**, 49–53 (2001).
39. Nakamura, M. J., Terai, J., Okubo, R., Hotta, K. & Oka, K. Three-dimensional anatomy of the *Ciona intestinalis* tailbud embryo at single-cell resolution. *Dev. Biol.* **372**, 274–284 (2012).
40. Imai, K. S., Hikawa, H., Kobayashi, K. & Satou, Y. *Tfap2* and *Sox1/2/3* cooperatively specify ectodermal fates in ascidian embryos. *Development* **144**, 33–37 (2017).
41. Irvine, S. Q., Cangiano, M. C., Millette, B. J. & Gutter, E. S. Non-overlapping expression patterns of the clustered *Dll-A/B* genes in the ascidian *Ciona intestinalis*. *J. Exp. Zool. B* **308**, 428–441 (2007).
42. Miya, T. & Nishida, H. Expression pattern and transcriptional control of *SoxB1* in embryos of the ascidian *Halocynthia roretzi*. *Zool. Sci.* **20**, 59–67 (2003).
43. Imai, K. S., Levine, M., Satoh, N. & Satou, Y. Regulatory blueprint for a chordate embryo. *Science* **312**, 1183–1187 (2006).
44. Wagner, D. E. et al. Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science* **360**, 981–987 (2018).
45. Emms, D. M. & Kelly, S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **16**, 157 (2015).
46. Nicol, D. & Meinertzhagen, I. A. Cell counts and maps in the larval central nervous system of the ascidian *Ciona intestinalis* (L.). *J. Comp. Neurol.* **309**, 415–429 (1991).
47. Dufour, H. D. et al. Precraniate origin of cranial motoneurons. *Proc. Natl Acad. Sci. USA* **103**, 8727–8732 (2006).
48. Takahashi, T. & Holland, P. W. Amphioxus and ascidian *Dmbx* homeobox genes give clues to the vertebrate origins of midbrain development. *Development* **131**, 3285–3294 (2004).
49. Yasuo, H. & Satoh, N. Function of vertebrate *T* gene. *Nature* **364**, 582–583 (1993).
50. Satoh, N., Tagawa, K. & Takahashi, H. How was the notochord born? *Evol. Dev.* **14**, 56–75 (2012).
51. Schock, E. N., York, J. R. & LaBonne, C. The developmental and evolutionary origins of cellular pluripotency in the vertebrate neural crest. *Semin. Cell Dev. Biol.* **138**, 36–44 (2023).
52. York, J. R. & McCauley, D. W. The origin and evolution of vertebrate neural crest cells. *Open Biol.* **10**, 190285 (2020).
53. Martik, M. L. & Bronner, M. E. Riding the crest to get a head: neural crest evolution in vertebrates. *Nat. Rev. Neurosci.* **22**, 616–626 (2021).
54. Rothstein, M. & Simoes-Costa, M. On the evolutionary origins and regionalization of the neural crest. *Semin. Cell Dev. Biol.* **138**, 28–35 (2023).
55. Green, S. A., Simoes-Costa, M. & Bronner, M. E. Evolution of vertebrates as viewed from the crest. *Nature* **520**, 474–482 (2015).
56. Leung, A. W. et al. WNT/beta-catenin signaling mediates human neural crest induction via a pre-neural border intermediate. *Development* **143**, 398–410 (2016).
57. Frith, T. J. et al. Human axial progenitors generate trunk neural crest cells *in vitro*. *eLife* **7**, e35786 (2018).
58. Gomez, G. A. et al. WNT/β-catenin modulates the axial identity of embryonic stem cell-derived human neural crest. *Development* **146**, dev175604 (2019).
59. Martik, M. L. et al. Evolution of the new head by gradual acquisition of neural crest regulatory circuits. *Nature* **574**, 675–678 (2019).
60. Rothstein, M., Bhattacharya, D. & Simoes-Costa, M. The molecular basis of neural crest axial identity. *Dev. Biol.* **444**, S170–S180 (2018).
61. Satou, Y. et al. A manually curated gene model set for an ascidian, *Ciona robusta* (*Ciona intestinalis* Type A). *Zool. Sci.* **39**, 253–260 (2022).
62. Ikeda, T. & Satou, Y. Differential temporal control of *Foxa.a* and *Zic-r.b* specifies brain versus notochord fate in the ascidian embryo. *Development* **144**, 38–43 (2017).
63. Tokuoka, M., Kobayashi, K. & Satou, Y. Distinct regulation of *Snail* in two muscle lineages of the ascidian embryo achieves temporal coordination of muscle development. *Development* **145**, dev163915 (2018).
64. Corbo, J. C., Levine, M. & Zeller, R. W. Characterization of a notochord-specific enhancer from the Brachyury promoter region of the ascidian, *Ciona intestinalis*. *Development* **124**, 589–602 (1997).
65. Leggio, B. et al. MorphoNet: an interactive online morphological browser to explore complex multi-scale data. *Nat. Commun.* **10**, 2812 (2019).
66. Nishida, H. Cell-division pattern during gastrulation of the ascidian, *Halocynthia roretzi*. *Dev. Growth Differ.* **28**, 191–201 (1986).
67. Roure, A., Lemaire, P. & Darras, S. An otx/nodal regulatory signature for posterior neural development in ascidians. *PLoS Genet.* **10**, e1004548 (2014).
68. Satou, Y. et al. A nearly complete genome of *Ciona intestinalis* type A (*C. robusta*) reveals the contribution of inversion to chromosomal evolution in the genus *Ciona*. *Genome Biol. Evol.* **11**, 3144–3157 (2019).
69. 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).

70. Kusakabe, T. et al. Gene expression profiles in tadpole larvae of *Ciona intestinalis*. *Dev. Biol.* **242**, 188–203 (2002).
71. Hotta, K. et al. Characterization of Brachyury-downstream notochord genes in the *Ciona intestinalis* embryo. *Dev. Biol.* **224**, 69–80 (2000).
72. Chiba, S. et al. A genomewide survey of developmentally relevant genes in *Ciona intestinalis*. IX. Genes for muscle structural proteins. *Dev. Genes Evol.* **213**, 291–302 (2003).
73. Chiba, S., Satou, Y., Nishikata, T. & Satoh, N. Isolation and characterization of cDNA clones for epidermis-specific and muscle-specific genes in *Ciona savignyi* embryos. *Zool. Sci.* **15**, 239–246 (1998).
74. Kusakabe, T. G. et al. A conserved non-reproductive GnRH system in chordates. *PLoS ONE* **7**, e41955 (2012).
75. Kobayashi, K., Maeda, K., Tokuoka, M., Mochizuki, A. & Satou, Y. Controlling cell fate specification system by key genes determined from network structure. *iScience* **4**, 281–293 (2018).
76. Katsuyama, Y. et al. Regulation of synaptotagmin gene expression during ascidian embryogenesis. *Dev. Biol.* **244**, 293–304 (2002).
77. Imai, K., Takada, N., Satoh, N. & Satou, Y. β-catenin mediates the specification of endoderm cells in ascidian embryos. *Development* **127**, 3009–3020 (2000).
78. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587 (2021).
79. O'Leary, N. A. et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion and functional annotation. *Nucleic Acids Res.* **44**, D733–D745 (2016).
80. Martin, F. J. et al. Ensembl 2023. *Nucleic Acids Res.* **51**, D933–D941 (2023).
81. Ishida, T. & Satou, Y. Ascidian embryonic cells with properties of neural-crest cells and neuromesodermal progenitors of vertebrates. *Zenodo* https://zenodo.org/records/10682771 (2024).

Acknowledgements

We thank R. Masuda, S. Tokuhiro, C. Imaizumi (Kyoto University), M. Yoshida (University of Tokyo) and other members working under the National BioResource Project for *Ciona* (MEXT, Japan) at Kyoto University and the University of Tokyo for providing experimental animals. This research was supported by grants from the Japan Society for the Promotion of Science under the grant nos. 21H02486 and 21H05239 to Y.S. The manuscript was edited by a technical editor, S. D. Aird.

Author contributions

T.I. and Y.S. conceived and designed the experiments. T.I. performed experiments. T.I. and Y.S. analysed the data. Y.S. wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

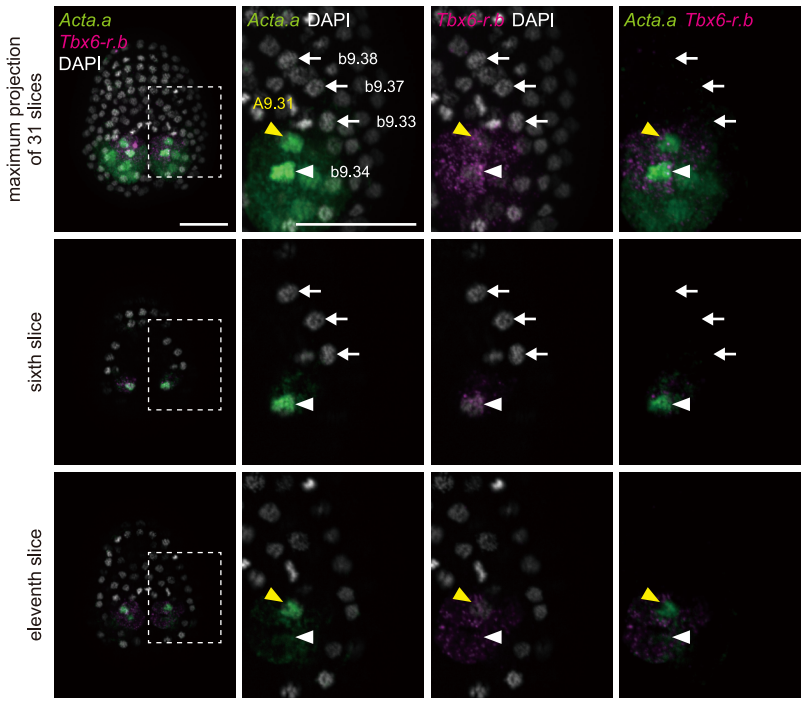
Extended data is available for this paper at https://doi.org/10.1038/s41559-024-02387-8.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41559-024-02387-8.

Correspondence and requests for materials should be addressed to Yutaka Satou.

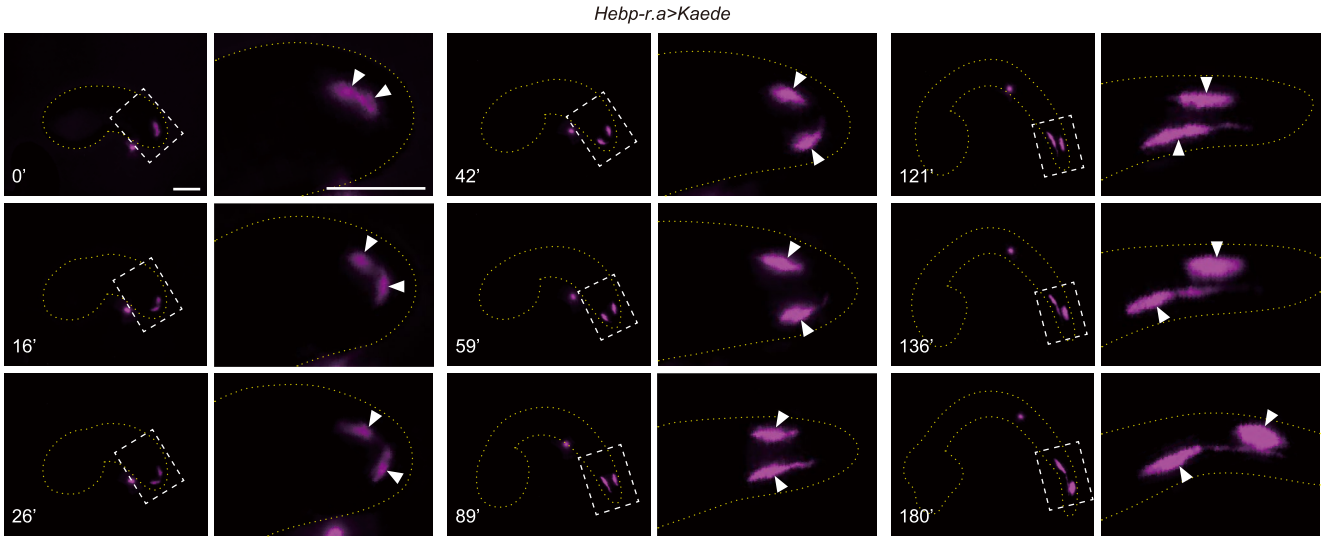
Peer review information *Nature Ecology & Evolution* thanks Joshua York and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

https://doi.org/10.1038/s41559-024-02387-8



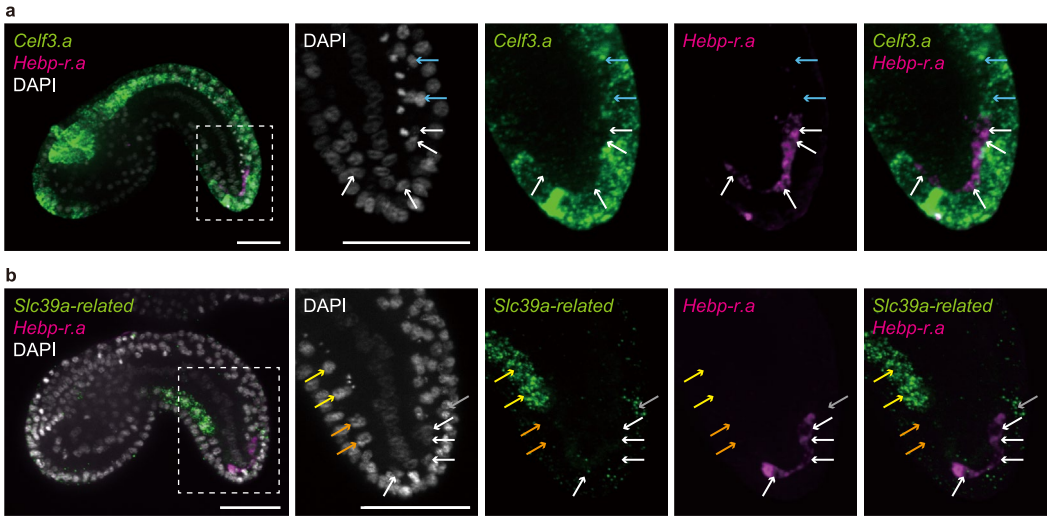
Extended Data Fig. 1 | *Acta.a*, which encodes a muscle actin, is expressed in b9.34 at the late gastrula stage. *Acta.a* expression was examined by *in situ* hybridization (green). *Tbx6-r.b* expression was also examined (magenta). *Z*-projected image stacks overlaid in pseudocolor are shown in the top row. The sixth optical slice more clearly shows that *Acta.a* and *Tbx6-r.b* are expressed in b9.34, but not in the other LNPCs. The eleventh optical slice shows expression of *Acta.a* and *Tbx6-r.b* in A9.31. Brightness and contrast of photographs were linearly adjusted. Scale bars, 50 µm.

https://doi.org/10.1038/s41559-024-02387-8



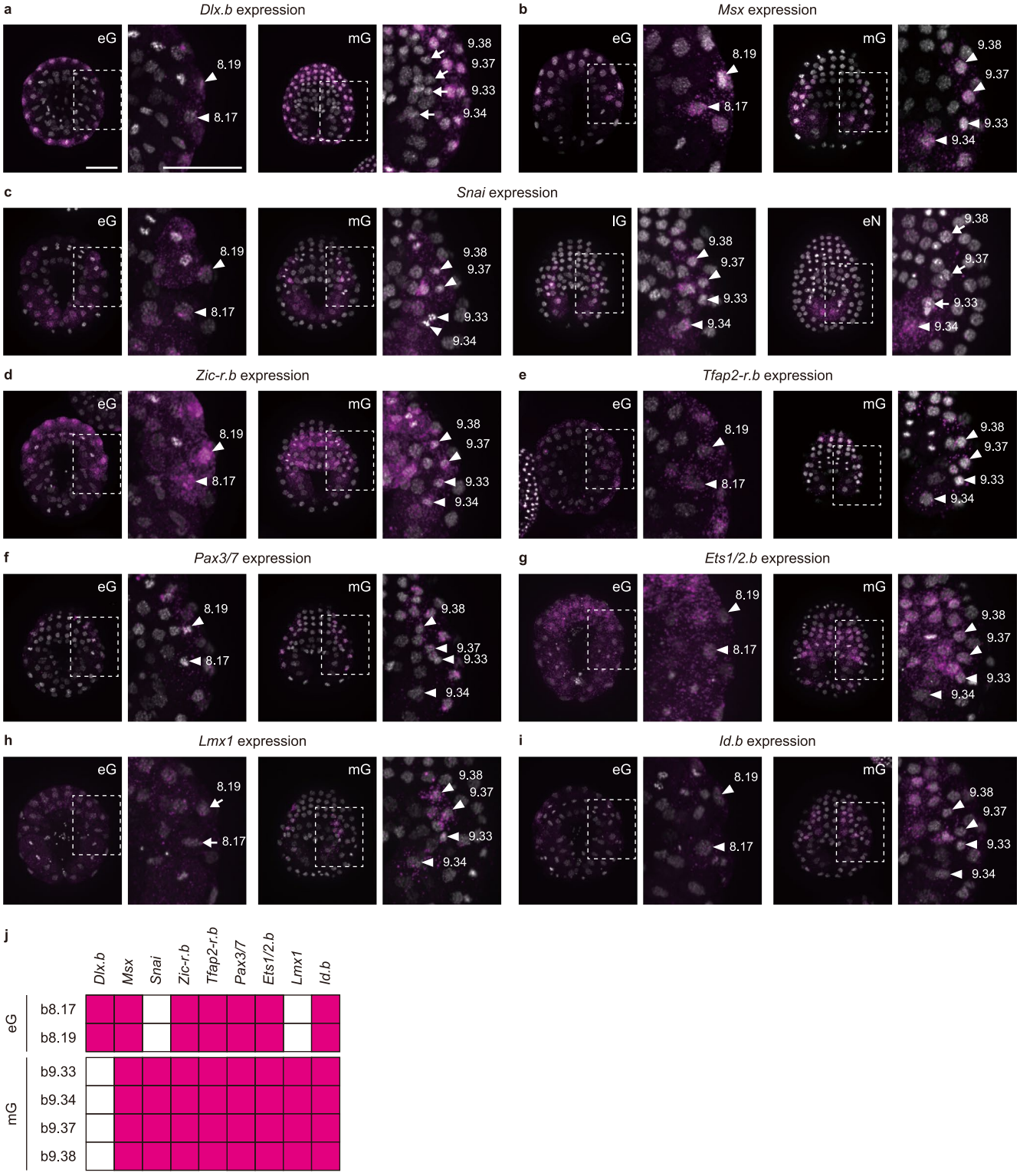
Extended Data Fig. 2 | Some pLNPC-derived cells change location from the dorsal to the ventral side between early and late tailbud stages. A tailbud embryo expressing *Hebp-r.a>Kaede* reporter was UV-irradiated for photoconversion of Kaede fluorescence at the early tailbud stage (0 min; early tailbud I). One cell with photoconverted Kaede changed its location to the ventral side. Note that only two tail-tip cells are labelled because of mosaic incorporation of the reporter construct. The first, third and sixth photographs are the same as the photograph shown in Fig. 2d. Photographs are *z*-projected image stacks overlaid in pseudocolor. Brightness and contrast of photographs were linearly adjusted. The dorsal side is up and the ventral side is down. Scale bar, 50 µm.

https://doi.org/10.1038/s41559-024-02387-8



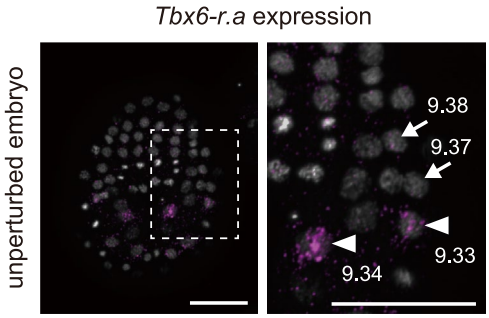
Extended Data Fig. 3 | Expression of a pan-neural marker, *Celf3.a,* and an endodermal-strand marker, *Slc39a-related*, in tailbud embryos. **a**,**b**, Expression of *Celf3.a* (**a**) and *Slc39a-related* (**b**) was examined by *in situ* hybridization and photographs are *z*-projected image stacks overlaid in pseudocolor (*Celf3.a* and *Slc39a-related*, green; *Hebp-r.a*, magenta). Tail-tip cells, which express *Hebp-r.a*, do not express *Celf3.a* or *Slc39a-related*. Brightness and contrast of photographs were linearly adjusted. Tail-tip cells are indicated by white arrows. Nerve-cord cells are indicated by cyan arrows and endodermal-strand cells are indicated by yellow arrows. There are two putative germ-line cells between anterior endodermal-strand cells and tail-tip cells (orange arrows). Grey arrows indicate epidermal cells. Nuclei are stained with DAPI (grey). Scale bar, 50 µm.

https://doi.org/10.1038/s41559-024-02387-8



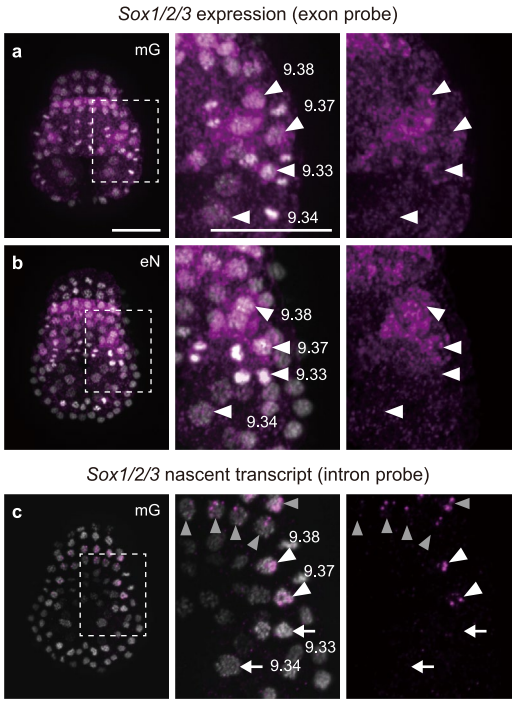
Extended Data Fig. 4 | *Dlx.b*, *Msx*, *Snai*, *Zic-r.b*, *Tfap2-r.b*, *Pax3/7*, *Ets1/2.b*, *Lmx1* and *Id.b* are expressed in LNPCs. (**a**–**i**) Expression was examined by *in situ* hybridization and photographs are *z*-projected image stacks and overlaid in pseudocolor (magenta). Nuclei were stained with DAPI (grey). Higher magnification views are shown on the right. Arrowheads indicate gene expression and arrows indicate the absence of expression. Developmental stages are shown in photographs; eG, early gastrula; mG, middle gastrula; eN, early neurula. Scale bar, 50 µm. (**j**) A summary of gene expression in LNPCs at the early and middle gastrula stages.

https://doi.org/10.1038/s41559-024-02387-8



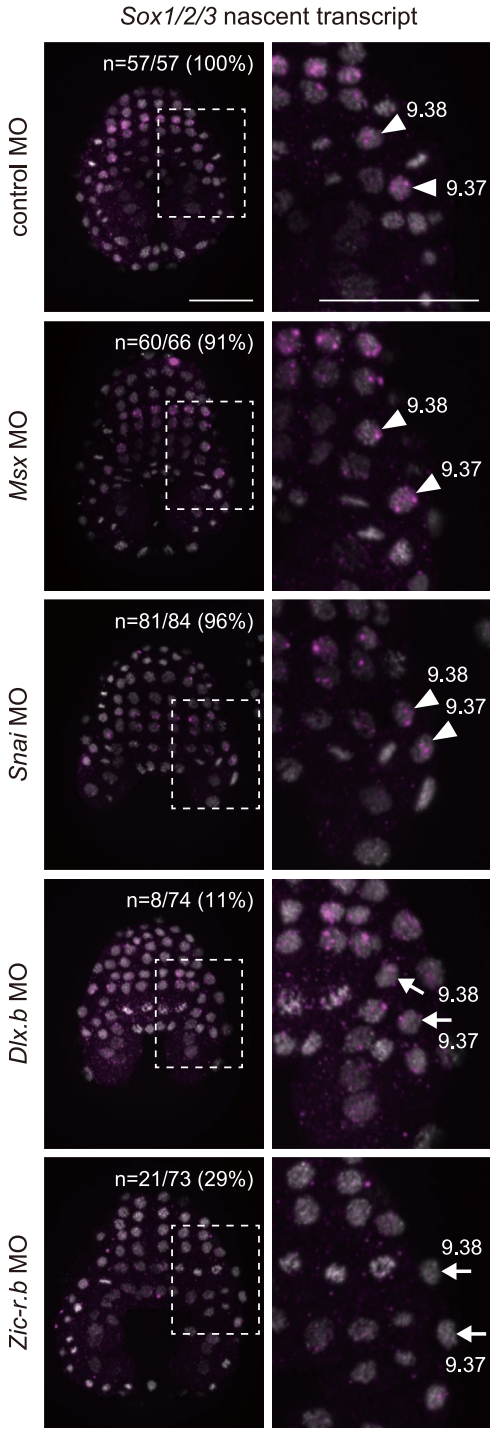
Extended Data Fig. 5 | *Tbx6-r.a* expression in a middle gastrula embryo. *Tbx6-r.a* is expressed in pLNPCs (b9.33 and b9.34) and indicated by white arrowheads. The remaining LNPCs (aLNPCs) are shown by arrows. Photographs are *z*-projected image stacks overlaid in pseudocolor. Nuclei are stained with DAPI (grey) and a higher magnification view is shown on the right. Brightness and contrast of photographs were linearly adjusted. Scale bars, 50 μm.

https://doi.org/10.1038/s41559-024-02387-8



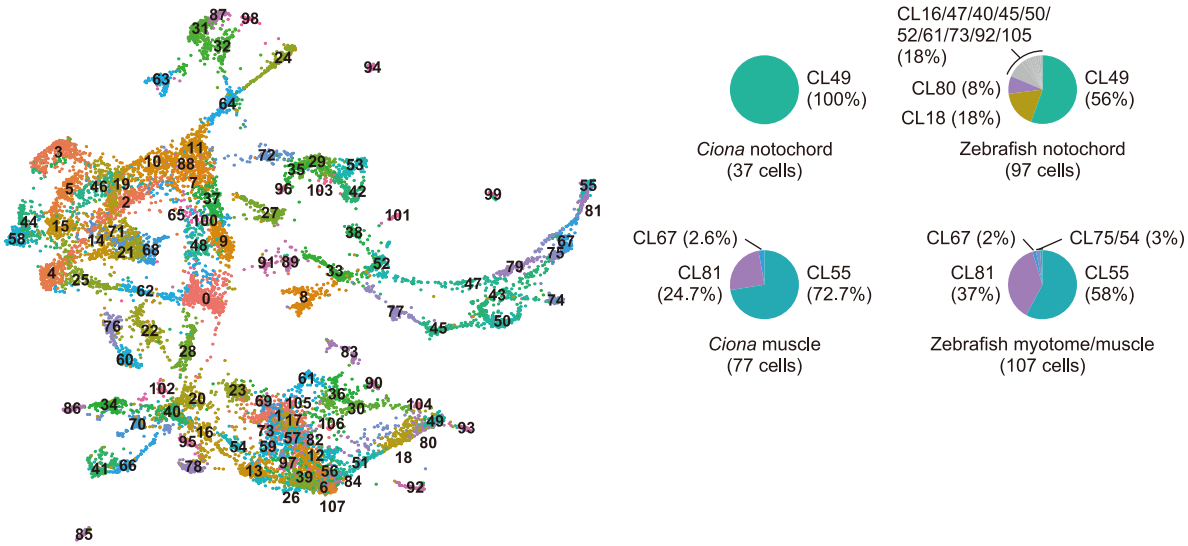
Extended Data Fig. 6 | Expression of *Sox1/2/3* in unperturbed middle gastrula embryos. (**a**, **b**) *Sox1/2/3* mRNA is detected in cells including LNPCs (white arrowheads) of normal middle gastrula gastrula (mG) and early neurula (eN) embryos. Note that signals in pLNPCs (b9.33 and b9.34) are weak at the early neurula stage. (**c**) *Sox1/2/3* nascent transcripts, which were examined with an intron probe, are not seen in pLNPCs (white arrows), which indicates that *Sox1/2/3* is not transcribed in pLNPCs. Note that *Sox1/2/3* is expressed in cells that contribute to the central nervous system (grey arrowheads) and aLNPCs (white arrowheads). Photographs are *z*-projected image stacks overlaid in pseudocolor; magenta, *in situ* hybridization signals; grey, nuclei stained with DAPI. Brightness and contrast of photographs were linearly adjusted. Scale bar, 50 µm.

https://doi.org/10.1038/s41559-024-02387-8



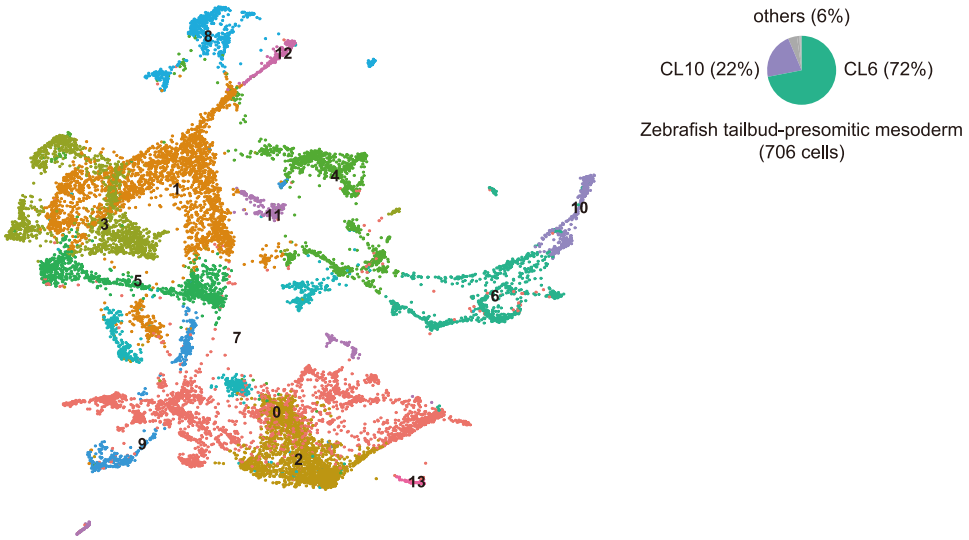
Extended Data Fig. 7 | Expression of *Sox1/2/3* in morphant embryos of *Msx*, *Snai*, *Dlx.b*, or *Zic-r.b* at the middle gastrula stage. Expression of *Sox1/2/3* nascent transcripts was examined by *in situ* hybridization using a probe designed to hybridize with the first intron of *Sox1/2/3*. Higher magnification views are shown on the right. LNPCs that express or do not express designated genes are shown by arrowheads and arrows, respectively. Nuclei are stained with DAPI and are shown in grey. Photographs are *z*-projected image stacks overlaid in pseudocolor. Brightness and contrast of photographs were linearly adjusted. Numbers of embryos examined and embryos that expressed *Sox1/2/3* nascent transcripts in b9.37 and b9.38 are shown in the panels. Scale bars, 50 µm.

https://doi.org/10.1038/s41559-024-02387-8



Extended Data Fig. 8 | High-resolution clustering of single-cell transcriptome data of *Ciona* and zebrafish. High-resolution clustering results are mapped on the same UMAP plot that is shown in Fig. 5. Different clusters are indicated by different colours. Pie charts show that notochord and muscle cells of *Ciona* and zebrafish are largely grouped into the same clusters.

https://doi.org/10.1038/s41559-024-02387-8



Extended Data Fig. 9 | Low-resolution clustering of single-cell transcriptome data of *Ciona* and zebrafish. Low-resolution clustering results are mapped on the same UMAP plot that is shown in Fig. 5. Different clusters are indicated by different colours. The pie chart shows that zebrafish cells annotated 'tailbud presomitic mesoderm' are mostly in clusters 6 and 10.

nature portfolio

Corresponding author(s): Yutaka Satou

Last updated by author(s): Jan 6, 2024

Reporting Summary

Nature Portfolio 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 Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

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

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

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Wget (version 1.19.5, GNU) and fastq-dump (version 3.0.5, NCBI) was used for downloading sequence data.
Data analysis	Cell Ranger (version 6.12, 10xGenomics), Orthofinder (version 2.5.4; Emms & Kelly 2015, Genome Biology, doi:10.1186/s13059-015-0721-2), Seurat (version 4.1.3; Hao et al., Cell 2021, 10.1016/j.cell.2021.04.048), Loupe Browser (version 6.0.0).

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 and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All data generated during this study are included in this published article and Zenodo (doi: 10.5281/zenodo.10682771). Ciona single-cell transcriptome data (SRA accession numbers: SRR9051005 to SRR9051007) and zebrafish single cell transcriptome data (GEO accession numbers: GSM3067192 to GSM3067195) were downloaded from the SRA and GEO databases.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	NA
Reporting on race, ethnicity, or other socially relevant groupings	NA
Population characteristics	NA
Recruitment	NA
Ethics oversight	NA

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

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	No statistical sample size calculation was performed. For in situ hybridization using unperturbed embryos, we used at least 20 embryos. The in situ hybridization protocol for Ciona has been well established and produces highly reproducible signals. No or little variations were observed among individuals. No comparisons among samples were expected. For in situ hybridization using morphant embryos or embryos with overexpression, we repeated experiments at least twice using different batches of embryos, and analyzed at least 20 embryos in total (numbers are shown in Figures). On the basis of protocols established in many studies using Ciona embryos, we expected that we could obtained reproducible results with 20 or more embryos (indeed reproducible results were obtained), and therefore we did not intend to perform statistical tests for these experiments. For lineage tracing experiments using reporter genes, we repeated experiments at least twice using different batches of embryos, and intended to obtain at least 20 embryos with reporter expression for each experiment. On the basis of protocols established in many studies using Ciona embryos, we expected that this sample size was sufficient for reproducible results, and we did not intend to perform statistical tests for these experiments. For embryos shown in Supplementary Figure 5, we observed at least 20 embryos for each stage. No or little variations were observed among individuals. Our purpose of this experiment was to confirm cell identities in embryos we used in the present study. Thus, the sample size is sufficient for obtaining reproducible results.
Data exclusions	No data were excluded
Replication	All experiments were repeated at least twice with different batches of embryos. Numbers of embryos examined and embryos with specific phenotypes are shown in Figures.
Randomization	Eggs were randomly selected for injection/electroporation, and resultant embryos were all analyzed. Unperturbed embryos used for in situ hybridization were randomly selected from embryos with normal morphology.
Blinding	Experiments were not blinded, because nothing was expected to vary depending on subjective judgment; expression of genes and reporters are binary (expressed or not expressed).

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 and archaeology
☐	☒ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

rabbit antibody against GFP (Thermo Fisher, #A6455); mouse antibody against RFP (Medical & biological laboratories, #M1553) ; mouse anti-digoxigenin antibody (Roche, #11333062910); rabbit anti-fluorescein antibody (Abcam, ab19491); goat HRP-conjugated anti-rabbit antibody (Thermo Fisher, #B40922); goat HRP-conjugated anti-mouse antibody (Thermo Fisher, #B40961); donkey alexa Fluor 488-conjugated anti-rabbit antibody (Thermo Fisher, #A21206); donkey alexa Fluor 555-conjugated anti-mouse antibody (Thermo Fisher, #A31570)

Validation

rabbit antibody against GFP (Thermo Fisher, #A6455): The manufacturer web site (https://www.thermofisher.com/antibody/product/GFP-Antibody-Polyclonal/A-6455) indicates that the anti-GFP antibody specifically recognizes GFP and has been used in more than 700 papers. We also used this antibody in our previous studies using Ciona (doi: 10.1242/dev.173104; 10.1038/s41467-019-12839-6; 10.1016/j.ydbio.2021.12.012).

mouse antibody against RFP (Medical & biological laboratories, #M1553) : The manufacturer web site (https://ruo.mbl.co.jp/bio/dtl/A/index.html?pcd=M155-3) indicates that the anti-RFP antibody specifically recognizes RFP and has been used in at least seven papers.

mouse anti-digoxigenin antibody (Roche, #11333062910): The manufacturer web site (https://www.sigmaaldrich.com/JP/ja/product/roche/11333062910) indicates that this antibody specifically recognizes digoxigenin.

rabbit anti-fluorescein antibody (Abcam, ab19491): The manufacturer web site (https://www.abcam.com/en-kr/products/primary-antibodies/fluorescein-antibody-ab19491) indicates that this antibody was validated by ELISA and western blotting and has been used in at least 14 papers.

goat HRP-conjugated anti-rabbit antibody (Thermo Fisher, #B40922): The user guide found in the manufacturer web site (https://www.thermofisher.com/order/catalog/product/B40922) indicates that this antibody specifically recognizes rabbit immunoglobulins.

goat HRP-conjugated anti-mouse antibody (Thermo Fisher, #B40961): The user guide found in the manufacturer web site (https://www.thermofisher.com/order/catalog/product/B40961) indicates that this antibody specifically recognizes mouse immunoglobulins.

donkey alexa Fluor 488-conjugated anti-rabbit antibody (Thermo Fisher, #A21206):The manufacturer web site (https://www.thermofisher.com/antibody/product/Donkey-anti-Rabbit-IgG-H-L-Highly-Cross-Adsorbed-Secondary-Antibody-Polyclonal/A-21206) indicates that this antibody specifically recognizes rabbit immunoglobulins and has been used in more than 6400 papers.

donkey alexa Fluor 555-conjugated anti-mouse antibody (Thermo Fisher, #A31570): The manufacturer web site (https://www.thermofisher.com/antibody/product/Donkey-anti-Mouse-IgG-H-L-Highly-Cross-Adsorbed-Secondary-Antibody-Polyclonal/A-31570) indicates that this antibody specifically recognizes mouse immunoglobulins and has been used in more than 1300 papers.

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	Adult specimens of Ciona intestinalis (type A; also called Ciona robusta; age 1-2 months; all animals are hermaphrodites) were obtained from the National BioResource Project for Ciona in Japan.
Wild animals	We do not use wild-caught animals.
Reporting on sex	NA
Field-collected samples	No field-collected samples were used in the present study.
Ethics oversight	Ciona intestinalis is excluded from legislation regulating scientific research on animals in Japan. Although there is no scientific evidence that Ciona intestinalis can experience pain, discomfort or stress, we made our best efforts to minimize potential harm that Ciona individuals might experience when we obtained eggs and sperm from them.

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