Ascidian cells with traits of neural crest and tailbud cells of vertebrates

Tasuku Ishida and Yutaka Satou*

Department of Zoology, Graduate School of Science, Kyoto University,
Sakyo, Kyoto, 606-8502, Japan

Correspondence to Yutaka Satou (yutaka@ascidian.zool.kyoto-u.ac.jp)

Summary

Neural crest cells (NCCs) 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 NCC-like properties; hence, their last common ancestor may have had primitive cells like NCCs[1-4]. However, these ascidian NCC-like cells do not produce cells that are commonly of mesodermal origin. Here, we showed that a cell population located in the lateral to posterior border of the neural plate has properties resembling those of vertebrate NCCs and NMPs. Among them, cells with *Tbx6-related* expression contribute to muscle near the tailbud, and cells with *Sox1/2/3* expression give rise to the nerve cord. These observations and cross-species transcriptome comparison indicates that these cells have properties like those of NMPs. Meanwhile, transcription factor genes *Dlx.b*, *Zic-r.b*, and *Snai,* which are reminiscent of a gene circuit in vertebrate NCCs, are involved in activation of *Tbx6-related.b*. Thus, the last common ancestor of ascidians and vertebrates may have had cells with hybrid properties of NMPs and NCCs, and such ancestral cells may have produced cells commonly of ectodermal and mesodermal origins.

Introduction

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 (NCCs), which are formed in the neural plate border (NPB) of vertebrate embryos, retain or re-activate their ability to produce cells that are commonly of mesodermal and ectodermal origins, even after gastrulation[5-7]. It is widely believed that NCCs contributed to evolution of vertebrates, especially evolution of the head region[8].

Recent studies have suggested that embryos of ascidians, which belong to the subphylum Tunicata, the sister group of vertebrates, contain cells that share an evolutionary origin with vertebrate neural crest cells[1-4]. These cells indeed differentiate into cells that include sensory neurons and pigment cells. Although homology of cells producing bipolar tail neurons to NCCs is controversial[9], it is likely that the last common ancestor of vertebrates and ascidians had primitive neural crest cells. Ascidian NCC-like cells identified so far do not produce cell types that are commonly of mesodermal origin, which raises the question of whether primitive NCCs 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 are another cell population that maintains the ability to produce mesodermal and ectodermal cells of posterior structures in late embryos, and reside in the tailbud[10,11]. It has not been determined whether ancestral Olfactores had NMP-like cells. If it had, did primitive NMP-like cells produce cells that are commonly of mesodermal and ectodermal origins and how did these primitive NCCs and NMPs contribute together to constructing the body plan of ancestral Olfactores?

Cells in the posterior border of the neural plate give rise to muscle cells and tail-tip cells near the tailbud.

Sister cells called b7.9 and b7.10 of embryos of ascidians (*Ciona robusta*, or *Ciona intestinalis* type A) laterally abut cells that contribute to the central nervous system (CNS) (Figure S1). Among their daughter cells, b8.17 and b8.19 also abut cells that contribute to the CNS, while the remaining cells (b8.18 and b8.20) are located more laterally (Figure 1a). We call the former cells (b8.17 and b8.19) and their descendants “proximal neural plate border cells” (pNPBCs) and the latter cells “distal neural plate border cells” (dNPBCs), hereafter. Anterior pNPBCs (b8.19) give rise to the dorsal row of the nerve cord[12]. Posterior pNPBCs (b8.17) give rise to two pairs of muscle cells in the tip of the tail, and cells located posteriorly to the nerve cord and endodermal strand[12,13]. *Tbx6-related.b* (*Tbx6-r.b*) is required for specification of all muscle cells, including muscle cells derived from b8.17[14]. The posterior daughter cells of b8.17 (b9.34) indeed express *Tbx6-r.b* (Figure 1b). Therefore, it is highly likely that the posterior daughters give rise to muscle cells.

This observation implies that cells located posteriorly to the nerve cord and endodermal strand

66 [we call these cells "tail-tip cells" (TTCs), hereafter] are derived from the anterior daughter of b8.17
67 (b9.33). For confirmation, we first identified genes expressed specifically in putative TTCs. *Hand-r* is
68 expressed in putative TTCs[15-18], but it cannot be used as a specific marker, because it is also expressed in
69 other cells. Therefore, we used *Hand-r* as a marker gene to find genes expressed specifically in TTCs in
70 single-cell RNA-sequencing (scRNA-seq) data published previously[19] (Table S1). Among these genes, we
71 confirmed that *Hebp-r.a* is indeed expressed in putative TTCs by *in situ* hybridization (Figure 1c). When
72 we introduced a reporter gene containing the upstream regulatory region of *Hebp-r.a* with another reporter
73 gene containing the upstream region of *Msx*, these two reporters were expressed in the same cells at the
74 tailbud (Figure 1d). Because *Msx* is specifically expressed in b8.17, b8.18, b8.19, and b8.20 in early
75 gastrulae[3,15,20,21], and because only b8.17 contributes to TTCs[12], we conclude that *Hebp-r.a* specifically
76 marks TTCs. Thus, the anterior daughter (b9.33) of b8.17 gives rise to TTCs and the posterior daughter
77 (b9.34) gives rise to muscle cells.

78 TTCs were initially located posterior to the nerve cord in early tailbud embryos (Figure 1e).
79 Subsequently, some TTCs changed their location to the ventral side, and these cells were finally located
80 posteriorly to the endodermal strand (Figure 1e; Figure S2). Consequently, TTCs were located posteriorly
81 to the nerve cord and the endodermal strand. Therefore, these cells have been regarded as nerve cord cells
82 and endodermal strand cells[12]. Using scRNA-seq data[19], we compared their expression profiles with those
83 of other cell types during the larval stage, and found that the transcriptome of TTCs was quite different
84 from those of the endodermal strand and nerve cord cells and of other tissues (Figure 1f). Therefore, it is
85 likely that TTCs are distinct from the endodermal strand or nerve cord. This was further supported by the
86 observation that TTCs did not express a pan-neural marker, *Celf3.a*[22], or an endodermal-strand marker,
87 *Slc39a-related*[17] (Figure S3). ScRNA-seq data also indicated that *Tbx6-r.a*, a paralog of *Tbx6-r.b*, is
88 expressed in the TTC lineage (Figure 1g).

89

90 **_Tbx6-r.b_ is activated by a set of regulatory factors similar to a set that specifies the NPB and NCCs in**
91 **vertebrate embryos**

92 Next, we examined how *Tbx6-r.b* is activated in b9.34. Previous studies have shown that
93 orthologs for genes specifying NPB cells and NCCs in vertebrate embryos are expressed in cells including
94 pNPBCs of ascidian embryos[15,21,23-25]. We confirmed that *Dlx.b, Msx*, *Snai*, *Zic-r.b*, *Tfap2-r.b*, *Pax3/7*,
95 *Ets1/2.b*, *Lmx1*, and *Id.b* were indeed expressed in pNPBCs (Figure S4).

96 *Dlx.b* is expressed in animal hemisphere cells, including pNPBCs, in gastrulae, and is important
97 for fate decision of these cells[23]. *Msx* is expressed in pNPBCs and dNPBC (b8.17, b8.18, b8.19, and
98 b8.20)[15], and is required for differentiation of sensory neurons[3,20]. While knockdown of *Msx* did not affect
99 *Tbx6-r.b* expression, *Dlx.b* knockdown downregulated *Tbx6-r.b* in b9.34 (Figure 2a; Figure S5). *Snai* was
100 also expressed in pNPBCs (see Figure 2c), and knockdown of *Snai* downregulated *Tbx6-r.b* expression
101 (Figure 2a). In addition, *Zic-r.b* was downregulated in *Dlx.b* morphants (Figure 2b), and *Snai* was down-

regulated in *Zic-r.b* morphants (Figure 2c). Thus, orthologs of genes that specify fates of the NPB and NCCs in vertebrate embryos, also regulated *Tbx6-r.b* expression in b9.34.

Tbx6-r.b potentially suppresses _Sox1/2/3_ expression

While pNPBCs are located at the boundary between the neural plate and epidermal territory near the dorsal side of the blastopore in early gastrulae (see Figure 1a), posterior pNPBCs are found near the tailbud in later embryos (see Figure 1c, d). Posterior pNPBCs also contribute to muscle cells, in which *Tbx6-r.b* is expressed, as we indicated above. In addition, anterior pNPBCs contribute to the nerve cord. These features of pNPBCs 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[26-28]. Because *Tbx6-r.a* and *Tbx6-r.b* are expressed in pNBPCs (b9.33 and b9.34) of ascidian embryos (see Figure 1b, g), we next examined transcription of *Sox1/2/3* by *in situ* hybridization. We used a probe designed to hybridize with the first intron of *Sox1/2/3*, because *Sox1/2/3* is expressed in all animal hemisphere cells of early embryos[15,23,29] and because residual *Sox1/2/3* mRNA potentially prevents precise analysis. Nascent *Sox1/2/3* RNA was detected in anterior pNPBCs (b9.37 and b9.38) and in cells that contribute to the CNS (Figure S6). When we overexpressed *Kaede* using the upstream regulatory region of *Msx* in pNPBCs and dNPBCs (*Msx>Kaede*) as a control, expression of nascent *Sox1/2/3* transcripts was not changed (Figure 3a). On the other hand, when we overexpressed *Tbx6-r.b* using the upstream regulatory region of *Msx* in pNPBCs and dNPBCs (*Msx>Tbx6-r.b*), nascent *Sox1/2/3* transcripts were not detected in pNPBCs, whereas expression in cells that contribute to the CNS was unaffected (Figure 3a). *Lmx1*, which is normally expressed in anterior pNPBCs under control of *Msx*, *Nodal* and *Zic-r.b*, and contributes to specification of dorsal nerve cord cells by activating *Pax3/7*[21,30], was also downregulated (Figure 3b). Instead, these cells ectopically expressed a muscle marker gene, *Acta.a* (Figure 3c). Thus, anterior pNPBCs have the potential to become muscle cells, and a key determinant for muscle fate, *Tbx6-r.b*, negatively regulates *Sox1/2/3*.

Single-cell transcriptome data support similarity of ascidian pNPBCs and vertebrate NMPs

The location of oNPBCs in late embryos and the regulatory circuit of *Tbx6-r.b* and *Sox1/2/3* support the hypothesis that pNPBCs in ascidian embryos share an evolutionary origin with vertebrate NMPs. To further confirm this hypothesis, we compared transcriptomes of pNPBCs with those of zebrafish embryonic cells using scRNA-seq data published previously[19,31]. To perform a cross-species comparison, we identified ortholog groups using OrthoFinder[32] between genes of *Ciona* and zebrafish, and compared ascidian pNPBCs with all available cells of zebrafish embryos from 10 to 24 hpf. Ascidian muscle cells and notochord cells were also included as controls, and these control cells were found in the close vicinity of muscle cells and notochord cells of zebrafish embryos in a UMAP plot, respectively (Figure 4; Figure S7), as expected. Anterior pNPBCs (b9.37 and b9.38) and posterior pNPBCs (b9.33 and b9.34) were found in

the close vicinity of cells annotated as spinal cord and pre-somitic mesoderm in the tailbud of zebrafish embryos, respectively. This observation supports the hypothesis that pNPBCs have properties similar to those of zebrafish NMPs, which reside in the tailbud and contribute to the spinal cord and somites.

141 **Discussion**

142 Our data show the similarity between ascidian pNPBCs and vertebrate NPB cells that give rise to
143 NCCs. This suggests that ascidian pNPBCs share an evolutionary origin with vertebrate NCCs. The first
144 point is that both ascidian and vertebrate cell populations arise in the lateral border of the neural plate.
145 Second, most genes for transcription factors that specify the NPB and NCCs in vertebrates are expressed in
146 these ascidian cells. These genes include *Dlx.b, Msx*, *Snai*, *Zic-r.b*, *Tfap2-r.b*, *Pax3/7*, *Ets1/2.b*, *Lmx1*, and
147 *Id.b*, and we showed that the gene circuit of *Dlx.b*, *Zic-r.b,* and *Snai* indeed regulates *Tbx6-r.b* expression in
148 posterior pNPBCs. In addition, in anterior pNPBCs, *Msx* and *Zic-r.b* regulate *Lmx1*, and *Lmx1* regulates
149 *Pax3/7*[21].

150 On the other hand, our data also showed the similarity between ascidian pNPBCs and vertebrate
151 NMPs. In vertebrate embryos, NMPs are characterized by co-expression of *Sox2* and *Brachyury*[33]. In
152 ascidian embryos, *Brachyury* is expressed exclusively in the notochord lineage[34] and never in cells with
153 *Sox1/2/3* expression. On the other hand, *Sox1/2/3* is initially expressed in all animal hemisphere cells
154 including cells giving rise to pNPBCs[15,23,29], and our data indicate that it is downregulated in posterior
155 pNPBCs.

156 Meanwhile, in vertebrate NMPs, cells with strong *Tbx6* expression contribute to mesodermal
157 cells and cells with strong *Sox2* expression contribute to the posterior part of the neural tube[26,28,33,35]. In
158 ascidian embryos, b9.34 expresses *Tbx6-r.b* and gives rise to muscle cells. Another posterior pNPBC,
159 b9.33, also expresses *Tbx6-r.a*, but produces cells of unknown identity. Anterior pNPBCs cells continue to
160 express *Sox1/2/3* and give rise to the nerve cord. The ascidian larval tail nerve cord does not contain
161 neurons[36], and it is debatable which part of the vertebrate CNS corresponds to the ascidian nerve cord[21,37-
162 39]. Nevertheless, tailbud cells with *Sox2* (or its ortholog) expression contribute to the posterior part of the
163 CNS in both ascidian and vertebrate embryos. In addition, in both embryos, *Tbx6* (or its ortholog)
164 negatively regulates *Sox2* (or its ortholog). These shared properties indicate a common evolutionary origin
165 of vertebrate NMPs and ascidian pNPBCs. Consistently, our cross-species comparison of zebrafish and
166 *Ciona* scRNA-seq data showed similarity between ascidian pNPBCs and vertebrate NMPs.

167 Thus, pNPBCs have properties of vertebrate NCCs and NMPs. This suggests that the ancestral
168 Olfactores may have had a cell population resembling ascidian pNPBCs, and that such ancestral cells may
169 have evolved into NCCs and NMPs in the vertebrate lineage. Nevertheless, because some dNPBCs are
170 thought to be homologs of NCCs[2,3], it is possible that the ancestral Olfactores also had additional NCC-like
171 cells that did not have NMP-like properties.

172 While ascidian NCC-like cells derived from dNPBCs and another lineage produce sensory
173 neurons and pigment cells[1-3], pNPBCs give rise not only to cells that are normally of ectodermal origin, but
174 also cells that are normally of mesodermal origin. Contrary to the previous hypothesis[40-43], this observation
175 indicates that the evolutionary origin of multipotency of vertebrate NCCs may date back to the ancestral
176 Olfactores.

Several studies have shown a close relationship between trunk NCCs and NMPs using human pluripotent cultured cells[44-46]. In addition, trunk NCCs are thought to have emerged earlier than other NCCs[47,48]. Therefore, ancestral cells of NCCs may have had properties of NMPs and ancient Olfactores may have had such cells. *Ciona* pNPBCs, which have hybrid properties of NCCs and NMPs, may represent such an ancestral state, and multipotency of these two cell populations may share a common evolutionary origin.

Acknowledgements

We thank Reiji Masuda, Shinichi Tokuhiro, Chikako Imaizumi (Kyoto University), Manabu Yoshida (University of Tokyo), and other members working under the National BioResource Project for *Ciona* (MEXT, Japan) in 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 numbers 21H02486 and 21H05239 to YS**.**

Author contribution

TI performed experiments. YS drafted the original manuscript. TI and YS reviewed the manuscript draft, revised it, and approved the final version.

Data availability

All data generated or analyzed during this study are included in this published article.

Competing interests

The authors declare no competing interests.

References

1 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, doi:10.1038/nature11589 (2012).

2 Stolfi, A., Ryan, K., Meinertzhagen, I. A. & Christiaen, L. Migratory neuronal progenitors arise from the neural plate borders in tunicates. *Nature* **527**, 371-374, doi:10.1038/nature15758 (2015).

3 Waki, K., Imai, K. S. & Satou, Y. Genetic pathways for differentiation of the peripheral nervous system in ascidians. *Nat Commun* **6**, 8719, doi:10.1038/ncomms9719 (2015).

4 Horie, R. *et al.* Shared evolutionary origin of vertebrate neural crest and cranial placodes. *Nature* **560**, 228-232, doi:10.1038/s41586-018-0385-7 (2018).

5 Tang, W. Y. & Bronner, M. E. Neural crest lineage analysis: from past to future trajectory. *Development* **147**, doi:10.1242/dev.193193 (2020).

6 Zalc, A. *et al.* Reactivation of the pluripotency program precedes formation of the cranial neural crest. *Science* **371**, eabb4776ro, doi:10.1126/science.abb4776 (2021).

7 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, doi:10.1126/science.aaa3655 (2015).

8 Gans, C. & Northcutt, R. G. Neural Crest and the Origin of Vertebrates - a New Head. *Science* **220**, 268-273, doi:10.1126/science.220.4594.268 (1983).

9 Papadogiannis, V. *et al.* Hmx gene conservation identifies the origin of vertebrate cranial ganglia. *Nature* **605**, 701-705, doi:10.1038/s41586-022-04742-w (2022).

10 Tzouanacou, E., Wegener, A., Wymeersch, F. J., Wilson, V. & Nicolas, J. F. Redefining the Progression of Lineage Segregations during Mammalian Embryogenesis by Clonal Analysis. *Developmental Cell* **17**, 365-376, doi:10.1016/j.devcel.2009.08.002 (2009).

11 Martin, B. L. & Kimelman, D. Canonical Wnt Signaling Dynamically Controls Multiple Stem Cell Fate Decisions during Vertebrate Body Formation. *Developmental Cell* **22**, 223-232, doi:10.1016/j.devcel.2011.11.001 (2012).

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, doi:10.1016/0012-1606(87)90188-6 (1987).

13 Hudson, C. & Yasuo, H. Neuromesodermal Lineage Contribution to CNS Development in Invertebrate and Vertebrate Chordates. *Genes-Basel* **12**, doi:10.3390/genes12040592 (2021).

14 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, doi:10.1016/j.ydbio.2005.03.029 (2005).

15 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, doi:10.1242/dev.01270 (2004).

16 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, doi:10.2108/zsj.22.837 (2005).

17 Satou, Y. *et al.* Gene expression profiles in Ciona intestinalis tailbud embryos. *Development* **128**, 2893-2904, doi:10.1242/dev.128.15.2893 (2001).

18 Satou, Y., Imai, K. & Satoh, N. The ascidian Mesp gene specifies heart precursor cells. *Development* **131**, 2533-2541, doi:10.1242/dev.01145 (2004).

19 Cao, C. *et al.* Comprehensive single-cell transcriptome lineages of a proto-vertebrate. *Nature* **571**, 349-354, doi:10.1038/s41586-019-1385-y (2019).

20 Roure, A. & Darras, S. Msxb is a core component of the genetic circuitry specifying the dorsal and ventral neurogenic midlines in the ascidian embryo. *Developmental Biology* **409**, 277-287, doi:10.1016/j.ydbio.2015.11.009 (2016).

21 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, doi:10.1242/dev.026419 (2009).

22 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, doi:10.1007/s004270000118 (2001).

23 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, doi:10.1242/dev.142109 (2017).

24 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, doi:10.1016/S0925-4773(99)00178-1 (1999).

25 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. *Journal of experimental zoology. Part B, Molecular and developmental evolution* **308**, 428-441, doi:10.1002/jez.b.21169 (2007).

26 Takemoto, T. *et al.* Tbx6-dependent Sox2 regulation determines neural or mesodermal fate in axial stem cells. *Nature* **470**, 394-398, doi:10.1038/nature09729 (2011).

27 Gouti, M., Metzis, V. & Briscoe, J. The route to spinal cord cell types: a tale of signals and switches. *Trends in Genetics* **31**, 282-289, doi:10.1016/j.tig.2015.03.001 (2015).

28 Martin, B. L. & Steventon, B. A fishy tail: Insights into the cell and molecular biology of neuromesodermal cells from zebrafish embryos. *Developmental Biology* **487**, 67-73, doi:10.1016/j.ydbio.2022.04.010 (2022).

29 Miya, T. & Nishida, H. Expression pattern and transcriptional control of SoxB1 in embryos of the ascidian Halocynthia roretzi. *Zool Sci* **20**, 59-67, doi:10.2108/zsj.20.59 (2003).

30 Imai, K. S., Levine, M., Satoh, N. & Satou, Y. Regulatory blueprint for a chordate embryo. *Science* **312**, 1183-1187, doi:10.1126/science.1123404 (2006).

31 Wagner, D. E. *et al.* Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science* **360**, 981-987, doi:10.1126/science.aar4362 (2018).

32 Emms, D. M. & Kelly, S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology* **16**, doi:10.1186/s13059-015-0721-2 (2015).

33 Henrique, D., Abranches, E., Verrier, L. & Storey, K. G. Neuromesodermal progenitors and the making of the spinal cord. *Development* **142**, 2864-2875, doi:10.1242/dev.119768 (2015).

34 Yasuo, H. & Satoh, N. Function of vertebrate T gene. *Nature* **364**, 582-583, doi:10.1038/364582b0 (1993).

35 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, doi:10.1242/dev.143578 (2017).

36 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, doi:10.1002/cne.903090402 (1991).

37 Dufour, H. D. *et al.* Precraniate origin of cranial motoneurons. *Proc Natl Acad Sci U S A* **103**, 8727-8732, doi:10.1073/pnas.0600805103 (2006).

38 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).

39 Takahashi, T. & Holland, P. W. Amphioxus and ascidian Dmbx homeobox genes give clues to the vertebrate origins of midbrain development. *Development* **131**, 3285-3294, doi:10.1242/dev.01201 (2004).

40 Schock, E. N., York, J. R. & LaBonne, C. The developmental and evolutionary origins of cellular pluripotency in the vertebrate neural crest. *Seminars in Cell & Developmental Biology* **138**, 36-44, doi:10.1016/j.semcdb.2022.04.008 (2023).

41 York, J. R. & McCauley, D. W. The origin and evolution of vertebrate neural crest cells. *Open Biology* **10**, doi:10.1098/rsob.190285 (2020).

42 Martik, M. L. & Bronner, M. E. Riding the crest to get a head: neural crest evolution in vertebrates. *Nature Reviews Neuroscience* **22**, 616-626, doi:10.1038/s41583-021-00503-2 (2021).

43 Rothstein, M. & Simoes-Costa, M. On the evolutionary origins and regionalization of the neural crest. *Seminars in Cell & Developmental Biology* **138**, 28-35, doi:10.1016/j.semcdb.2022.06.008 (2023).

44 Leung, A. W. *et al.* WNT/beta-catenin signaling mediates human neural crest induction via a pre-neural border intermediate. *Development* **143**, 398-410, doi:10.1242/dev.130849 (2016).

45 Frith, T. J. *et al.* Human axial progenitors generate trunk neural crest cells in vitro. *Elife* **7**, doi:10.7554/eLife.35786 (2018).

46 Gomez, G. A. *et al.* WNT/beta-catenin modulates the axial identity of embryonic stem cell-derived human neural crest. *Development* **146**, doi:10.1242/dev.175604 (2019).

47 Martik, M. L. *et al.* Evolution of the new head by gradual acquisition of neural crest regulatory circuits. *Nature* **574**, 675-+, doi:10.1038/s41586-019-1691-4 (2019).

48 Rothstein, M., Bhattacharya, D. & Simoes-Costa, M. The molecular basis of neural crest axial identity. *Developmental Biology* **444**, S170-S180, doi:10.1016/j.ydbio.2018.07.026 (2018).

49 Satou, Y. *et al.* A Manually Curated Gene Model Set for an Ascidian, *Ciona robusta* (*Ciona intestinalis* Type A). *Zool Sci* **39**, 253-260, doi:10.2108/zs210102 (2022).

50 Ikeda, T., Matsuoka, T. & Satou, Y. A time delay gene circuit is required for palp formation in the ascidian

embryo. *Development* **140**, 4703-4708, doi:10.1242/dev.100339 (2013).

51 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, doi:10.1242/dev.142174 (2017).

52 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, doi:10.1242/dev.163915 (2018).

53 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, doi:10.1242/dev.124.3.589 (1997).

54 Roure, A., Lemaire, P. & Darras, S. An otx/nodal regulatory signature for posterior neural development in ascidians. *PLoS genetics* **10**, e1004548, doi:10.1371/journal.pgen.1004548 (2014).

55 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 biology and evolution* **11**, 3144-3157, doi:10.1093/gbe/evz228 (2019).

56 O'Leary, N. A. *et al.* Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research* **44**, D733-D745, doi:10.1093/nar/gkv1189 (2016).

57 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, doi:10.1016/j.ydbio.2004.07.007 (2004).

58 Kusakabe, T. *et al.* Gene expression profiles in tadpole larvae of Ciona intestinalis. *Dev Biol* **242**, 188-203, doi:10.1006/dbio.2002.0538 (2002).

59 Hotta, K. *et al.* Characterization of Brachyury-downstream notochord genes in the Ciona intestinalis embryo. *Dev Biol* **224**, 69-80, doi:10.1006/dbio.2000.9765 (2000).

60 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, doi:10.1007/s00427-003-0324-x (2003).

61 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, doi:10.2108/zsj.15.239 (1998).

62 Kusakabe, T. G. *et al.* A conserved non-reproductive GnRH system in chordates. *PLoS One* **7**, e41955, doi:10.1371/journal.pone.0041955 (2012).

63 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, doi:10.1016/j.isci.2018.05.004 (2018).

64 Katsuyama, Y. *et al.* Regulation of synaptotagmin gene expression during ascidian embryogenesis. *Developmental Biology* **244**, 293-304, doi:10.1006/dbio.2002.0584 (2002).

65 Imai, K., Takada, N., Satoh, N. & Satou, Y. β-catenin mediates the specification of endoderm cells in ascidian embryos. *Development* **127**, 3009-3020, doi:10.1242/dev.127.14.3009 (2000).

356 **Materials and Methods**

357 **Animals and gene identifiers**

358 Adult specimens of *Ciona intestinalis* (type A; also called *Ciona robusta*) were obtained from the
359 National BioResource Project for *Ciona*. cDNA clones were obtained from our EST clone collection[16].
360 Identifiers for genes examined in this study are as follows: *Tbx6-r.b*, KY21.Chr11.465/466/467; *Hebp-r.a,*
361 KY21.Chr10.215; *Hand-r*, KY21.Chr1.1701; *Tbx6-r.a*, KY21.Chr11.458; *Msx*, KY21.Chr2.1031; *Dlx.b*,
362 KY21.Chr7.361; *Snai*, KY21.Chr3.1356; *Zic-r.b*, KY21.Chr6.26/27/28/29/30/31; *Sox1/2/3*,
363 KY21.Chr1.254; *Lmx1*, KY21.Chr9.606; *Acta.a*, KY21.Chr1.1745; *Tfap2-r.b*, KY21.Chr7.1145; *Pax3/7*,
364 KY21.Chr10.288; *Ets1/2.b*, KY21.Chr10.346; *Id.b*, KY21.Chr7.1129; *Celf3.a,* KY21.Chr6.58; *Slc39a-*
365 *related*, KY21.Chr4.1089. Identifiers for the latest KY21 set[49] are shown.

366

367 **Functional assays**

368 Sequences of MOs that block translation, were as follows: *Dlx.b,* 5'-
369 TCGGAGATTCAACGACGCTTGACAT-3'; *Zic-r.b*, 5'-GATCAACCATTACATTAGAATACAT-3';
370 *Snai*, 5'-GTCATGATGTAATCACAGTAATATA-3'; *Msx*, 5'-ATTCGTTTACTGTCATTTTTAATTT-3'.
371 These MOs were used and evaluated in previous studies[3,21,23,30,50-52]. They were microinjected under a
372 microscope. For overexpression, coding sequences of *Tbx6-r.b* were cloned into the downstream region of
373 the upstream regulatory region of *Msx* (Chr2:6,132,658-6,133,440). Overexpression constructs were
374 introduced by electroporation[53]. All functional assays were performed at least twice with different batches
375 of embryos.

376

377 ***In situ* hybridization and reporter assay**

378 Embryos were fixed in 4% paraformaldehyde in 0.1 M MOPS buffer (pH 7.5) and 0.5 M
379 NaCl at 4 °C for over 16 h. After washing with 80% ethanol and phosphate-buffered saline containing
380 0.1% Tween 20 (PBST), embryos were incubated in 3% H_2O_2 for 60 min, and again washed with
381 PBST. Then, embryos were treated with 2 μg/mL proteinase K in PBST for 30 min at 37 °C, and
382 washed with PBST. Embryos were again fixed with 4% paraformaldehyde in PBST for 1 h at room
383 temperature, washed with PBST, and immersed in hybridization buffer for at least 1 h at 50 °C. Then,
384 the hybridization buffer was replaced with fresh hybridization buffer containing a digoxigenin- or a
385 fluorescein-labeled probe, and embryos were incubated at 50 °C for at least 18 h. The hybridization
386 buffer contained 50% formamide, 5×SSC, 100 μg/mL yeast tRNA, 5×Denhart's solution, and 0.1%
387 Tween 20. After hybridization, embryos were washed twice at 50 °C for 15 min in 2×SSC, 50%
388 formamide, and 0.1% Tween 20. Embryos were treated for 30 min at 37 °C with 20 μg/mL RNase A in
389 10 mM Tris buffer (pH 8.0), 0.5 M NaCl, 5 mM EDTA, and 0.1% Tween 20. Embryos were further
390 washed twice at 50 °C for 15 min in 0.5×SSC, 50% formamide, and 0.1% Tween 20, and the washing

391 solution was replaced with PBST. Embryos were blocked at room temperature for 60 min with 0.1%
392 BSA in PBST, and then exposed to mouse anti-digoxigenin antibody (Roche, #11333062910; 1/100
393 dilution) or rabbit anti-fluorescein antibody (Abcam, ab19491; 1/1000 dilution) diluted in Can-Get-
394 Signal Immunostain Immunoreaction enhancer solution A (Toyobo, NKB-501) at 4 °C overnight.
395 After washing with PBST, embryos were incubated with HRP-conjugated secondary antibody
396 (Thermo Fisher, #B40961 and #B40922). Embryos were washed with PBST and TNT [100 mM Tris
397 buffer (pH7.5), 150 mM NaCl, and 0.1% Tween 20]. Then, embryos were stained with Alexa Fluor
398 488 or 555 tyramide reagent (Thermo Fisher, #B40953 and #B40955). For two-color *in situ*
399 hybridization, two probes were simultaneously added to the hybridization solution, and the secondary-
400 antibody reaction and detection with tyramide reagents were repeated twice. Before observation with a
401 microscope, embryos were treated with TrueVIEW Autofluorescence Quenching Kit (Vector, #SP-8400-
402 15).

403 The upstream region that is contained in reporter of *Hebp-r.a>Venus* or *Hebp-r.a>Kaede* is
404 Chr10:1,456,230-1,457,052. The upstream region of *Msx>RFP* that promotes expression in dNPBCs and
405 pNPBCs[54], is Chr2:6,132,658-6,133,440. These reporter constructs were introduced by electroporation[53] or
406 microinjection. Expression of fluorescent proteins was examined under fluorescence microscope.

407

408 **Data analysis**

409 To compare gene expression profiles among *Ciona* larval tissues, we used single cell
410 transcriptome data for *Ciona* larvae that were published previously[19] (SRA accession numbers:
411 SRR9051005 to SRR9051007). Data were mapped to the latest genome assembly (HT version)[55] and the
412 latest version of the gene model set (KY21 version)[49] using Cell Ranger software (version 6.12.
413 10xGenomics). A heatmap containing the top 10 genes for each of 11 tissues was generated with Seurat
414 (version 4.1.3).

415 For a cross-species comparison between *Ciona* transcriptomes and zebrafish transcriptomes,
416 single-cell transcriptome data for *Ciona* middle gastrula embryos to larvae, which were published
417 previously[19], were used (SRA accession numbers: SRR9050986, SRR9050989, SRR9050992,
418 SRR9050997, SRR9050998, SRR9051003, and SRR9051007). Data were mapped to the latest genome
419 assembly (HT version)[55] and the latest version of the gene model set (KY21 version)[49] using Cell Ranger
420 software (version 6.12. 10xGenomics). For zebrafish 10- to 24-hr embryos, mapping data to gene models
421 were downloaded from the GEO database (accession numbers: GSM3067192 to GSM3067195)[31]. For
422 comparisons to zebrafish transcriptomes, we first compared proteomes of *Ciona* (KY21 version) and
423 zebrafish (Refseq[56] proteome file named GCF_000002035.6_GRCz11_protein.faa) using OrthoFinder[32]. As
424 controls to obtain better resolution, we included proteomes of *Strongylocentrotus purpuratus* (sea urchin)
425 and *Drosophila melanogaster*. Proteome files we used were: HT.KY21Gene.protein.2.fasta.zip for *Ciona*,
426 which includes a proteome set derived from the latest KY21 gene model set[49], downloaded from the Ghost

427 database[16]; GRCz11_protein.faa for the zebrafish proteome, downloaded from the NCBI Refseq database[56];
428 GCF_000002235.5_Spur_5.0_translated_cds.faa.gz was downloaded for the sea urchin proteome from the
429 NCBI Refseq database[56]; and Drosophila_melanogaster.BDGP6.32.pep.all.fa.gz for the vinegar fly
430 proteome, downloaded from the Ensembl database (release 108). OrthoFinder yielded 6379 ortholog
431 groups that contained both ascidian and zebrafish proteins (Table S2). When multiple genes for one species
432 were included in one ortholog group, all read counts in that group were summed. Expression matrices of
433 *Ciona* and zebrafish were integrated and analyzed with Seurat (version 4.1.3).

434 Annotations of zebrafish cells published previously[31] were used. We annotated *Ciona* cells using
435 Loupe Browser (10xGenomics). For cell identification during the larval stage, we used *Hebp-r.a*, *Hand-r*,
436 KY21.Chr8.1029 for TTCs, KY21.Chr1.1899 and KY21.Chr1.1498 for mesenchyme[57,58], KY21.Chr6.610
437 (*Noto1*)[59] for notochord, KY21.Chr 11.773 (*Myh.f*) and *Acta.a* for muscle[60], KY21.Chr1.2012 (*Epi1*) and
438 KY21.Chr7.872 (*Epib*) for epidermal cell[61], KY21.Chr9.123 (*Gnrh2*) and KY21.Chr10.869 for nerve
439 cord[58,62], KY21.Chr11.800 (*Rlbp1*) for brain[63], and KY21.Chr2.456 (*Pou4*) for sensory neurons[15]. Other
440 neural cells were identified by expression of KY21.Chr11.1137 and KY21.Chr2.49[19], KY21.Chr6.690
441 (*Pax2/5/8.a*)[21], or KY21.Chr2.1203 (*Syt*)[64]. Endodermal cells were identified by expression of
442 KY21.Chr6.222 (*Alp*) and KY21.Chr14.805[17], and endodermal strand cells were identified as cells
443 expressing KY21.Chr14.805, without expression of *Alp*[17,65]. These cells were further curated manually and
444 are listed in Table S3. For the cross-species comparison, TTCs, muscle and notochord cells in middle
445 tailbud embryos were identified using the same marker genes that were used in larvae. At the middle
446 gastrula and early neurula stages, we identified b9.34 as cells expressing *Msx*, *Tbx6-r.b*, and *Snai*, and
447 b9.33 as cells expressing *Msx*, *Snai*, and *Hox12* but not *Pax3/7* or *Tbx6-r.b*. These cells were further curated
448 manually and are listed in Table S4.

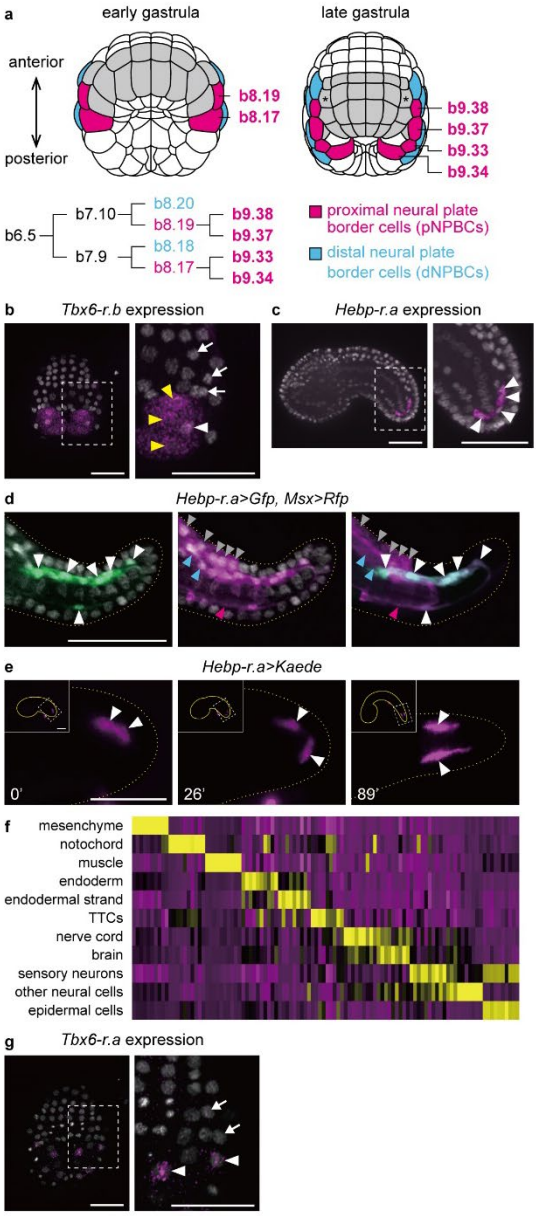
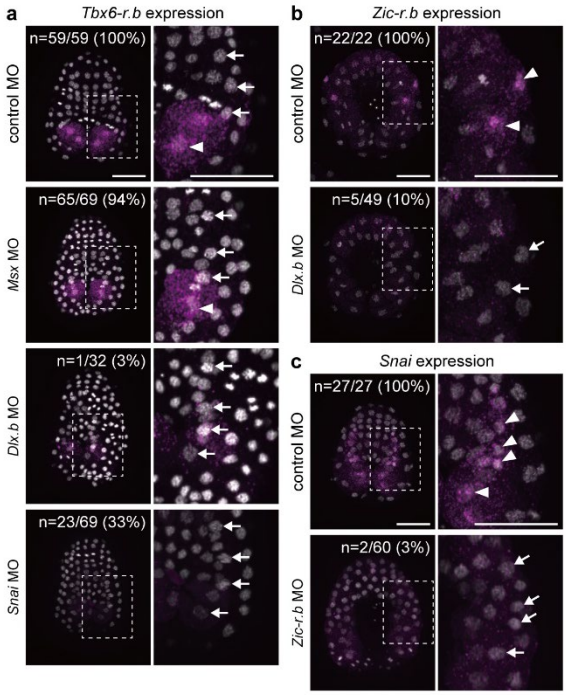


Figure 1. The cell lineage of b8.17 and b8.19 cells. (a) Ascidian bilaterally symmetrical embryos at the early and late gastrula stages. Neural plate cells, which contribute to the central nervous system, are colored gray. Magenta cells are b8.17 and b8.19 at the early gastrula stage, and their descendants at the late gastrula stage, which we call pNPBCs. Cyan cells are b8.18 and b8.20 at the early gastrula stage, and their descendants at the late gastrula stage, which we call dNPBCs. Cells that have been considered to be homologous to vertebrate NCCs[2,3] are derived from these cyan cells. In addition, pigment cells in the brain vesicle are derived from cells marked by asterisks, and these cells are also thought to share an origin with vertebrate NCCs[1]. The lineage of pNPBCs and dNPBCs is shown below. These cells are derived from b6.5 cells in 32-cell embryos. (b) *Tbx6-r.b* expression revealed by *in situ* hybridization in a middle gastrula embryo (magenta). *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). Other

pNBPCs are indicated by white arrows. (c) Expression of a TTC marker, *Hebp-r.a*, in a tailbud embryo (early tailbud I). *Hebp-r.a* signals are shown by magenta (arrowheads). (d) Expression of reporter constructs that contain upstream regions of *Hebp-r.a* (green) and *Msx* (magenta) in the tailbud region of a late tailbud embryo (late tailbud III). Overlaid photographs of signals (left) of the *Hebp-r.a* reporter and DAPI, (middle) of the *Msx* reporter and DAPI, and (right) of the *Hebp-r.a* reporter and the *Msx* reporter. The *Msx* reporter is expressed not only in TTCs, but also in muscle cells (magenta arrowheads), epidermal cells (gray arrowheads), and the dorsal row of the nerve cord (cyan arrowheads). The dorsal side is up, and the ventral side is down. (e) A tailbud embryo expressing *Hebp-r.a>Kaede* reporter was subjected to UV-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 at the middle tailbud stage (26 min; middle tailbud I), and to the ventral side at the late tailbud stage (89 min; late tailbud I). Note that only two TTCs are labelled because of mosaic incorporation of the reporter construct. Photographs at other time points are shown in Figure S2. (f) A heatmap showing gene expression of various cell types. Each vertical line represents expression levels of a single gene. Expression levels were calculated as averages of expression levels of cells in each cell type. The top 10 genes for each cell type are shown. Yellow and magenta represent high and low expression levels, respectively. (g) *Tbx6-r.a* expression in b9.33 and b9.34 (magenta), which are indicated by white arrowheads, in a late gastrula embryo. The remaining pNPBCs are shown by arrows. Photographs in b, c, d, e, and g are z-projected image stacks overlaid in pseudocolor. In b, c, d, and g, nuclei are stained with DAPI (gray), and higher magnification views are shown on the right. Brightness and contrast of photographs in d, e, g were linearly adjusted. Scale bars, 50 μm.



482

483 **Figure 2. The regulatory gene circuit that activates *Tbx6-r.b* in b9.34.** (a) Expression of *Tbx6-r.b* in a
484 control embryo and embryos injected with an MO against *Msx, Dlx.b,* or *Snai*, at the early neurula stage.
485 (b) Expression of *Zic-r.b* in a control embryo and an embryo injected with the *Dlx.b* MO at the early
486 gastrula stage. (c) Expression of *Snai* in a control and an embryo injected with the *Zic-r.b* MO at the middle
487 gastrula stage. Higher magnification views are shown on the right. pNPBCs that express and do not express
488 designated genes are shown by arrowheads and arrows, respectively. Nuclei are stained with DAPI and are
489 shown in gray. Photographs are z-projected image stacks overlaid in pseudocolor. Brightness and contrast
490 of photographs in embryos showing *Tbx6-r.b* expression, except the *Snai* morphant embryo, were linearly
491 adjusted. Numbers of embryos examined and embryos that expressed (a) *Tbx6-r.b* in b9.34, (b) *Zic-r.b* in
492 b8.17 and b8.19, and (c) *Snai* in pNPBCs are shown in the panels. In (a), signals for *Tbx6-r.b* expression in
493 A-line muscle cells are seen over b9.33 cells, in which *Tbx6-r.b* is not expressed (see Figure S5). Scale
494 bars, 50 μm.

495

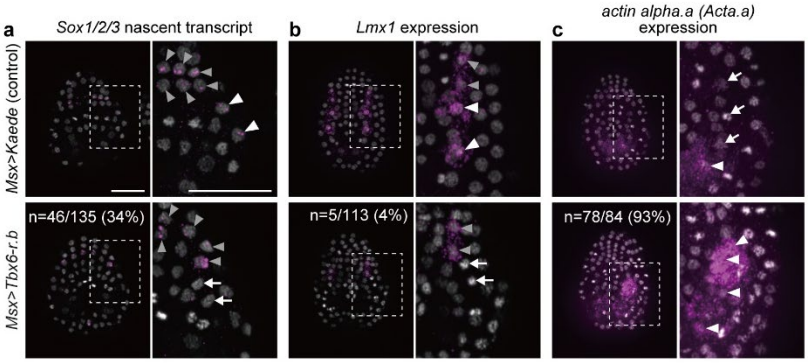


Figure 3. *Tbx6-r.b* can repress *Sox1/2/3* in anterior pNPBCs and can change their fate to muscle. (a) *Sox1/2/3* is transcribed in anterior pNPBCs (b9.37 and b9.38; white arrowheads) and in cells that contribute to the central nervous system (gray arrowheads) of control middle gastrula embryos introduced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

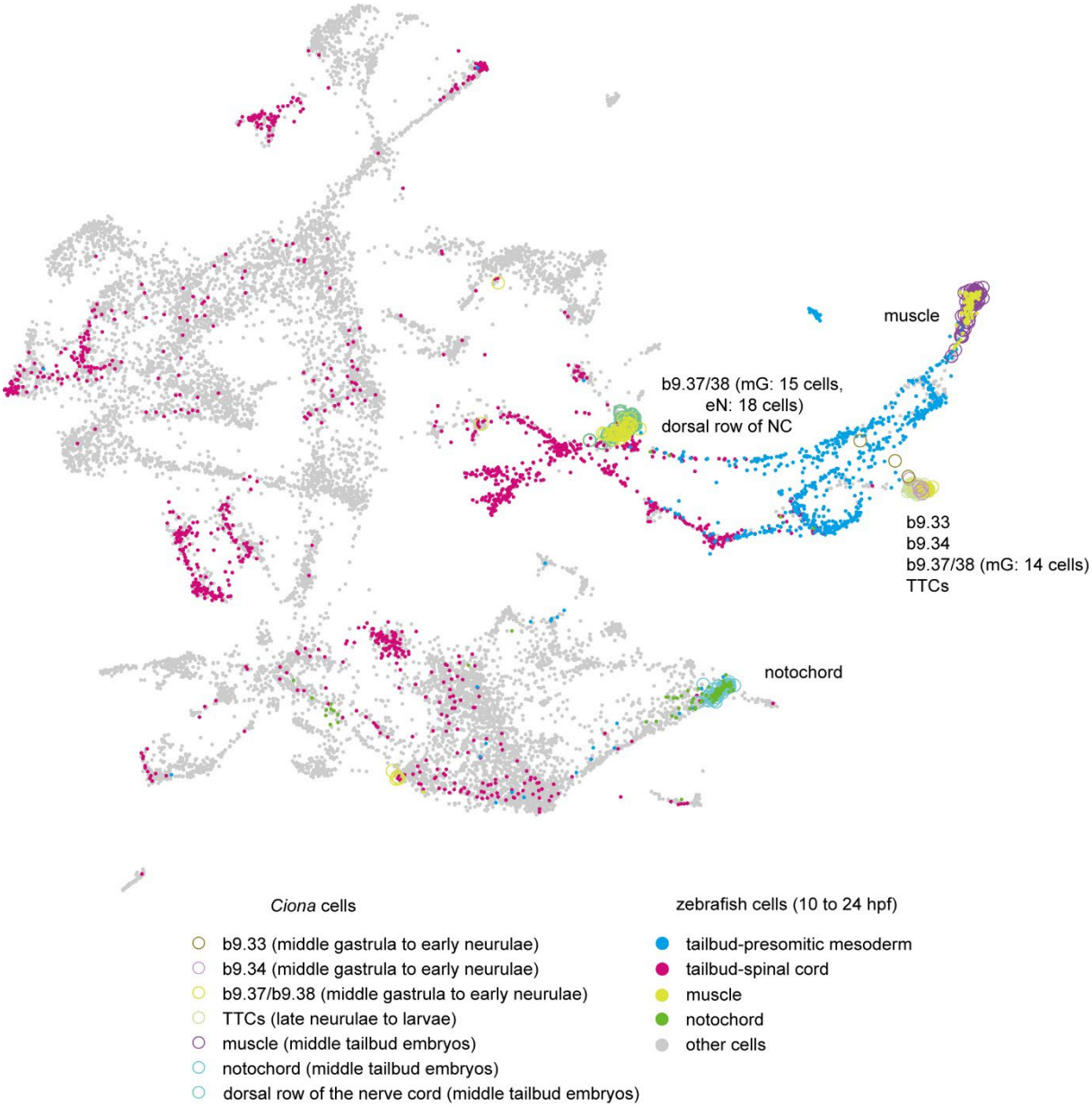


Figure 4. A cross-species comparison of single-cell transcriptome data indicates affinities of ascidian pNPBCs and tailbud cells of zebrafish embryos. A UMAP plot that includes single-cell transcriptome data of 10 to 24 hpf embryos of zebrafish, and pNPBCs and its derivatives in ascidian embryos from the middle gastrula to late tailbud stages and larvae. For positive controls, muscle cells and notochord cells of an ascidian middle tailbud embryo are also included.

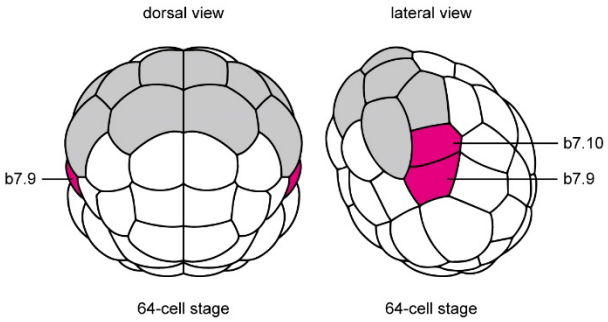


Figure S1. An ascidian 64-cell embryo. Cells, which contribute to the central nervous system, are colored with gray, and b7.9 and b7.10, which produce dNPBCs and pNPBCs, are colored magenta.

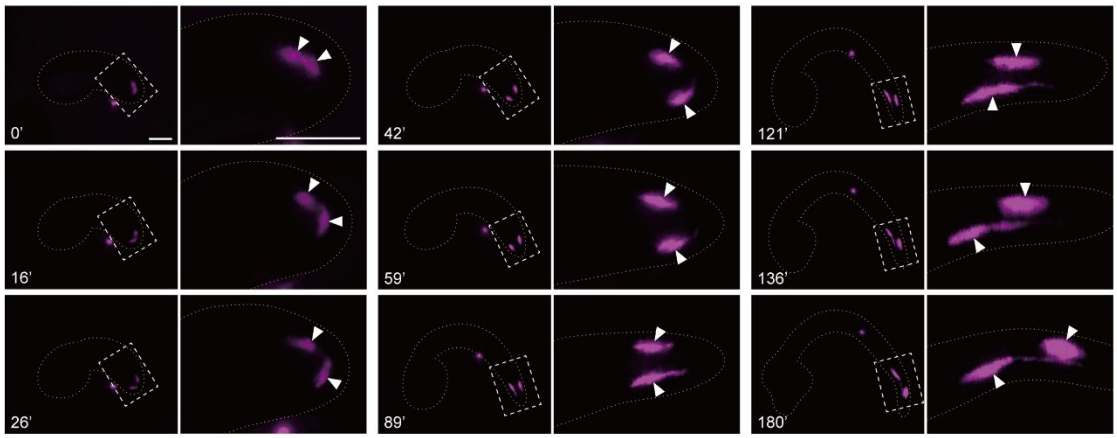
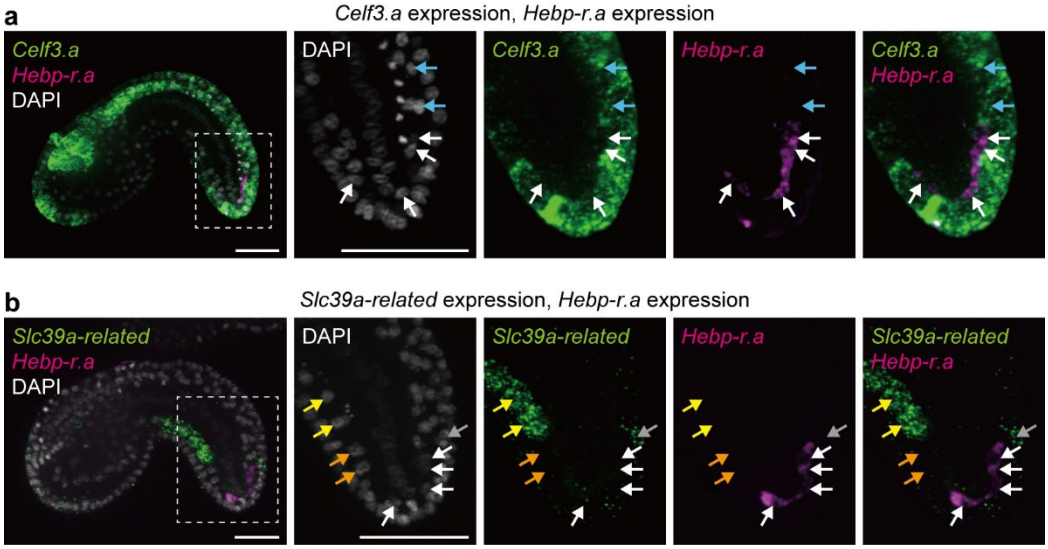


Figure S2. Some tail-tip cells change location from the dorsal to the ventral side between early and late tailbud stages. A tailbud embryo expressing *Hebp-r.a>Kaede* reporter was subjected to UV-irradiation for photoconversion of Kaede fluorescence at the early tailbud stage (0 min; early tailbud I). One cell with photoconverted Kaede changes its location to the ventral side. Note that only two TTCs are labelled because of mosaic incorporation of the reporter construct. The first, third, and sixth photographs are the same as the photograph shown in Figure 1e. 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.



535

536 **Figure S3. Expression of a pan neural marker, *Celf3.a*, and an endodermal-strand marker, *Slc39a-***
537 ***related* in tailbud embryos.** Expression was examined by *in situ* hybridization, and photographs are z-
538 projected image stacks overlaid in pseudocolor (*Celf3.a* and *Slc39a-related*, green; *Hebp-r.a*, magenta).
539 TTCs, which express *Hebp-r.a*, do not express *Celf3.a* or *Slc39a-related*. Brightness and contrast of
540 photographs were linearly adjusted. TTCs are indicated by white arrows. Nerve cord cells are indicated by
541 cyan arrows, and endodermal strand cells are indicated by yellow arrows. There are two cells between
542 TTCs and endodermal strand cells (orange arrows). From their position, these two cells are probably germ-
543 line cells. Scale bar, 50 μm.

544

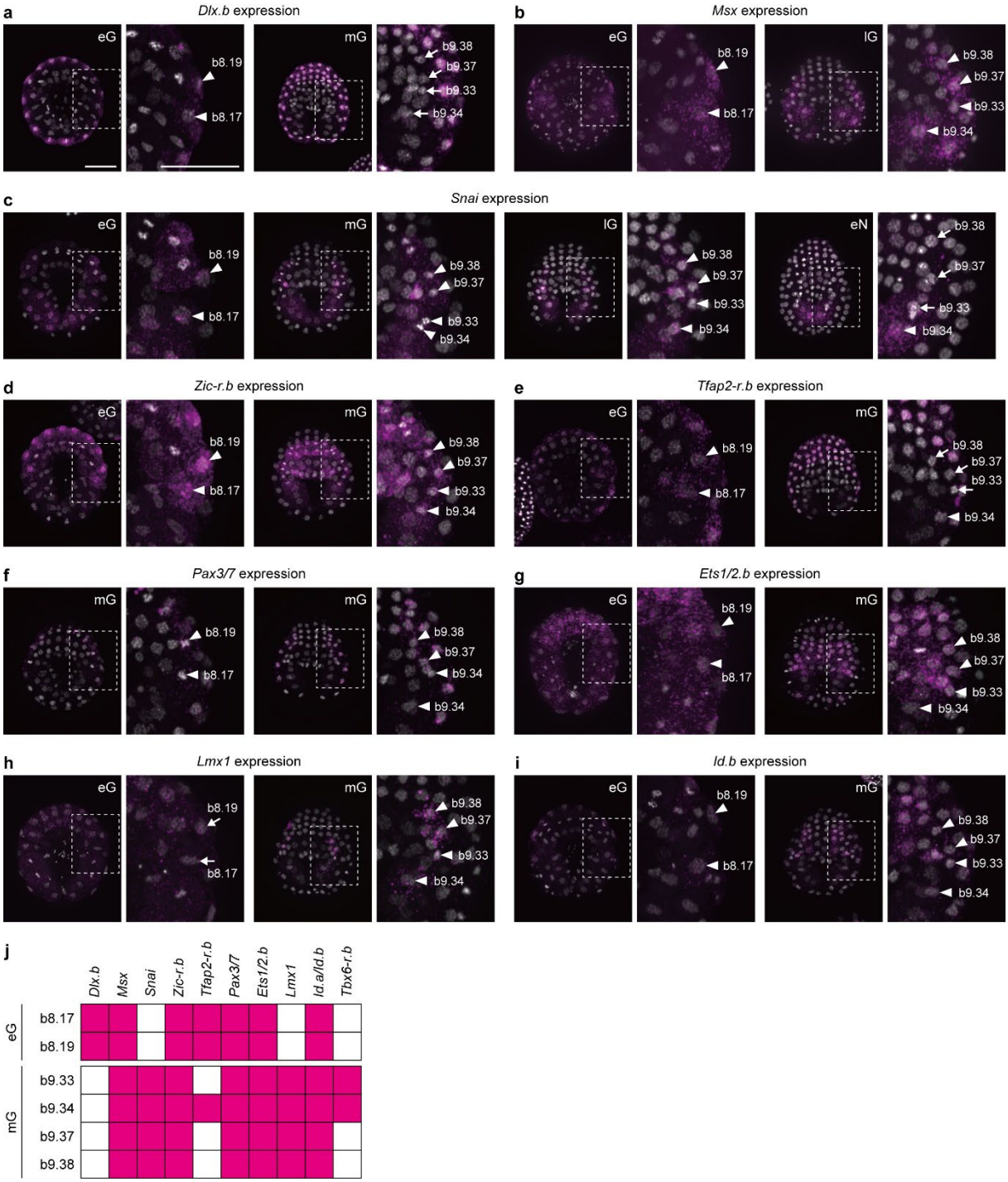
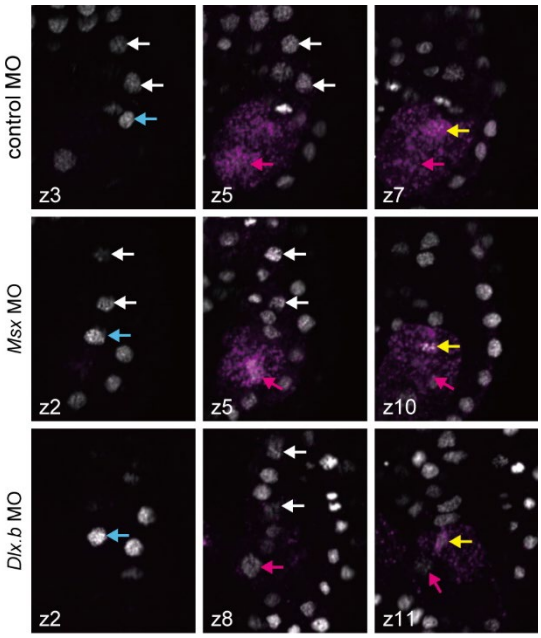


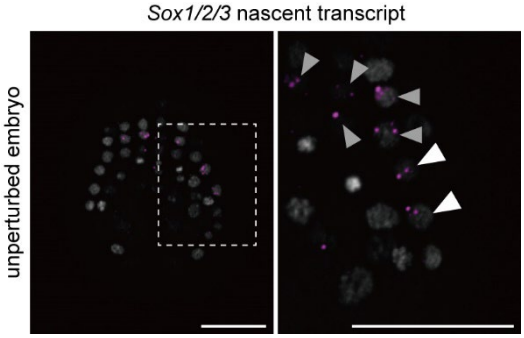
Figure S4. *Dlx.b*, *Msx*, *Snai*, *Zic-r.b*, *Tfap2-r.b*, *Pax3/7*, *Ets1/2.b*, *Lmx1*, and *Id.b* are expressed in pNPBCs. (a-i) Expression was examined by *in situ* hybridization, and photographs are z-projected image stacks and overlaid in pseudocolor (magenta). Nuclei were stained with DAPI (gray). 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; lG, late gastrula, eN, early neurula. Scale bar, 50 μm. (j) A summary of gene expression in pNPBCs at the early and middle gastrula stages.



553

554 **Figure S5. Expression of *Tbx6-r.b* in a control embryo and embryos injected with an MO against *Msx***
555 **or *Dlx.b*.** Optical slices showing *Tbx6-r.b* expression in control and embryos injected with the *Msx* MO or
556 *Dlx.b* MO. Although signals are seen around b9.33 in z-projected image stacks shown in Figure 2, this
557 signal shows expression in an A-line muscle cell (A9.31), but not in b9.33. The anterior pNPBCs (b9.37
558 and b9.38) are shown by white arrows, b9.33 by cyan arrows, b9.34 by magenta arrows, and the A-line
559 muscle cell by yellow arrows. Nuclei are stained with DAPI (gray). Brightness and contrast of photographs
560 were linearly adjusted. Photographs are overlaid in pseudocolor. Scale bar, 50 μm.

561



562

563 **Figure S6. Expression of nascent *Sox1/2/3* in unperturbed middle gastrula embryos.** *Sox1/2/3* is
564 transcribed in anterior pNPBCs (b9.37 and b9.38; white arrowheads) and in cells that contribute to the
565 central nervous system (gray arrowheads) of normal middle gastrula embryos. Photographs are z-projected
566 image stacks overlaid in pseudocolor; magenta, *in situ* hybridization signals; gray, nuclei stained with
567 DAPI. Brightness and contrast of photographs were linearly adjusted.

568

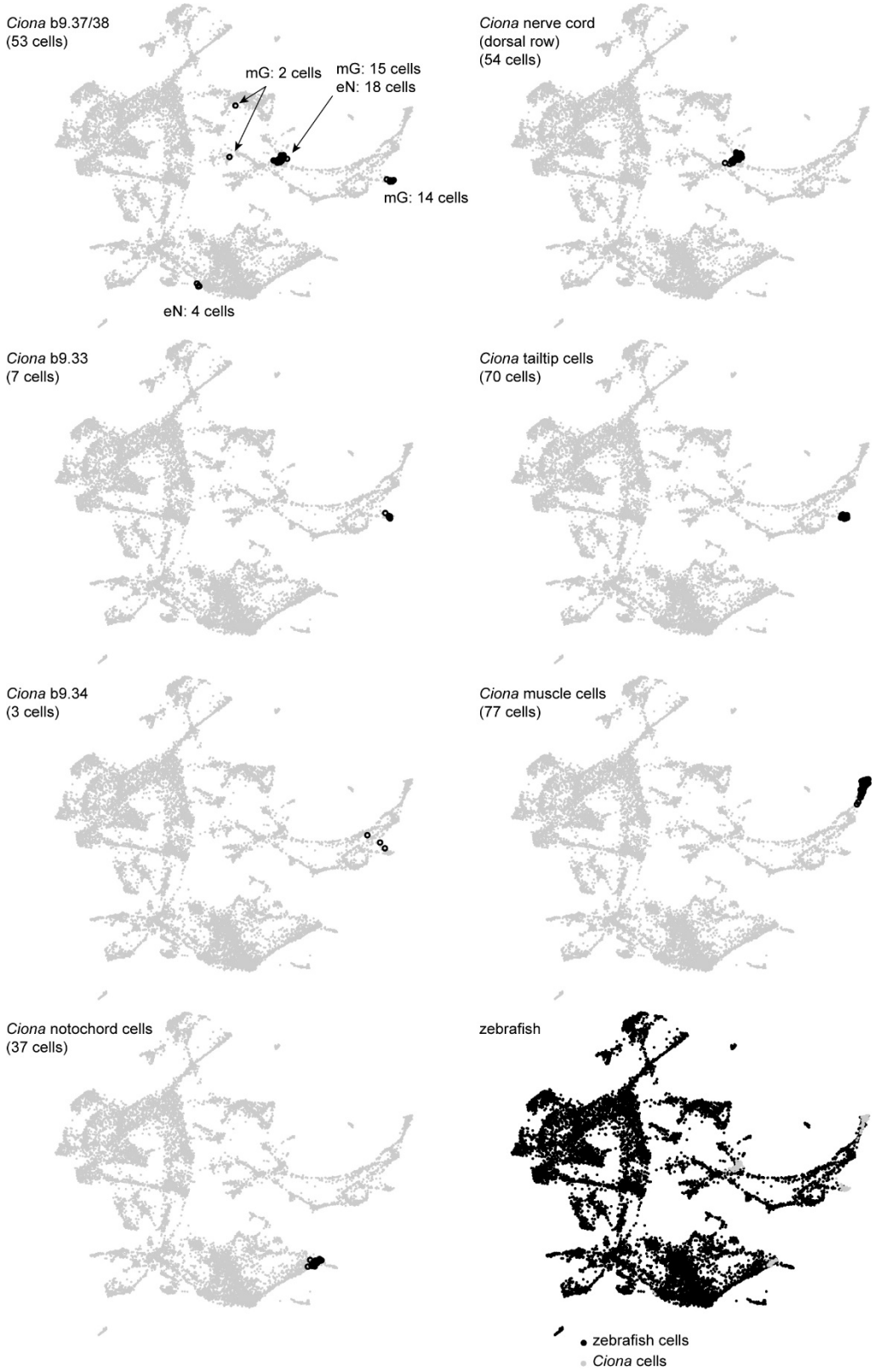


Figure S7. Locations of individual cell types in the cross-species comparison UMAP shown in Figure 4. Seven types of cell in *Ciona* embryos are individually shown by black circles. Zebrafish cells are shown in the bottom right panel in black.

Table S1. Genes abundantly expressed in TTCs.

Separate Excel File

Table S2. Ortholog groups identified by OrthoFinder

Separate Excel File

Table S3. Annotation for cells used to identify highly expressed genes in representative tissues.

Separate Excel File

Table S4. Annotation for *Ciona* cells used for cross-species analysis.

Separate Excel File