Materials and Methods

Ciona species used in experiments

Ciona intestinalis (Linnaeus 1767) commonly used in experiments has been recently proposed to consist of two populations[1]. One, described in some publications as *Ciona intestinalis* (Type B), has been proposed by some authors as being the original species first described by Linnaeus. The other, known as *Ciona intestinalis* (Type A), has also been described as *Ciona robusta*. All experiments in this paper were conducted with *Ciona intestinalis* (Type B). Genome data used in this study primarily derive from *Ciona intestinalis* (Type A) (*Ciona robusta*). This is indicated in the methods below each time *Ciona* is referred to.

Hmx gene identification, cloning and sequence analyses

We searched multiple sources of lamprey and hagfish sequence data for potential Hmx genes. For lamprey this constituted *Lampetra planeri* transcriptome data [2], genome assemblies for *Petromyzon marinus* [3,4], a *P. marinus* transcriptome assembly built in-house from Illumina GAII data available on SRA, the *Lethenteron camtschaticum* genome assembly [5], and an L. *camtschaticum* transcriptome assembly kindly provided by Juan Pascual-Anaya. For hagfish we searched the *Eptatretus burgeri* genome Eburgeri_3.2 genome assembly. In each dataset we identified the three genes as described in the main manuscript. In *P. marinus* (genome version Pmar_germline 1.0/petMar3) these are located on scaffold_00015 represented by gene models PMZ_0020818-RA, PMZ_0048148 and PMZ_0028877-RA. An additional *P. marinus* scaffold, scaffold_00813, also contained a gene model (PMZ_0038761-RA) with an Hmx type homeobox. However when we examined the sequence of this locus it was found to have >99.5% identity to part of the Hmx locus from scaffold_00015 (Supplementary File 7). We concluded it is either a very recent duplication of sequence from scaffold_00015, or an artefact of the genome assembly process, and have not considered it further.

To identify CNEs we first compared jawed vertebrate loci using the Conserved Non-coDing Orthologous Regions (CONDOR) database [6]. This identified a small number of elements surrounding the *Hmx3-Hmx2* locus that were conserved across jawed vertebrates. We extended this to lamprey and hagfish, using sequence similarity searches to search specifically for these conserved elements in these lineages. In addition, we also carried out extensive comparison of the lamprey and hagfish locus (extending 500kb upstream *HmxA* and downstream *HmxC*) to jawed vertebrate *Hmx3-Hmx2* and *Hmx1-SOHo*, which did not reveal additional non-coding elements shared between the two lineages. Alignments and molecular phylogenetic analyses were undertaken using MUSCLE [7] and RAxML [8] using the Maximum Likelihood method. 1000 bootstrap replicates were used to assess node confidence.

Lamprey *Hmx* genes were cloned from *L. planeri* cDNA, and uCNE and dCNE sequences from *L. planeri* genomic DNA, using the primers shown in the Table 1 below. The *Ciona intestinalis* (Type A) *Hmx/Nkx5* locus was already annotated [9], though the gene model was incomplete. Since no ESTs mapped to this gene, no clones were available in arrayed plasmid libraries. We hence first cloned a fragment of the gene using the primers shown in the table below, and used this for in situ hybridisation. We then used homology to *Ciona savignyi*, coupled with RNAseq data mapped on ANISEED [10], to identify the full open reading frame. This was amplified, in two sections, one 5' and one 3' using the primers shown in Table 1 and cloned into the vector using the Cold Fusion system (System

Biosciences). *Branchiostoma lanceolatum Hmx* was identified by searching the genome [11]. The in situ probe was cloned by PCR from 24-36 hours post fertilisation larvae cDNA.

All clones were verified by sequencing, and new cloned sequences have been deposited in Genbank accessions MN264670- MN264672.

Table 1: PCR primers used in this study

Gene/region	5′ primer	3′ primer
L. planeri HmxA probe	ACGTCGTCCGGGTTTGATAA	CGGAAGACGACAACGAAGGA
L. planeri HmxB probe	CCGCGCTGGCATTTCTTATT	GTAGCGCTTCATGTCGAACG
L. planeri HmxC probe	GACGCAGGTGAAGATCTGGT	GGATGGACAAAAAGCGGCAG
Amphioxus Hmx probe	CTCCAACAACCCGTTCTCGA	TGGGATAAGGGCGGGTTTTC
L. planeri uCNE	GCAGCAAGCGATGTATGCAA	CGAGTACGCGAAGAGGTGG
L. planeri dCNE	CATCATCAGCGCGCGTATTT	TTCCTGCACACAACTCTCGC
C. intestinalis Hmx probe	AGATCTGGTTTCAGAACCGACG	GGGATTCCCCTCTTTGCGTA
Ciona full ORF 5′ section	TAAAATAGTAAAATGGTACCTATGACGTCACTGTGCCAATTG	TTCCCCTTCTGACGTAGGGA
Ciona full ORF 3′ section	TCCCTACGTCAGAAGGGGAAG	ACCGGCGCTCAGCTGGAATTATGATTGTCTCACACCACGGAA
Ciona CNE	CACTATAGAACTCGACGTCCAACTAAATCGTCGAAG	CTGACCATGGGGATCCTTTTCGGCGGTTCGTTCTC
Lamprey uCNE into pCES	CACTATAGAACTCGAGCAGCAAGCGATGTATGC	GAGTCGACCTGCAGGCGAGTACGCGAAGAGGTGG
Lamprey dCNE into pCES	CACTATAGAACTCGACATCATCAGCGCGCGTATTTAAG	GAGTCGACCTGCAGGTTCCTGCACACAACTCTCGC
Ngn FwB	ATTGGACAGTTACGGCTTAG	
Ngn RevB	GACCATGCATTCGGTTTCGTTC	
Hmx_overlapATG FwA	TTGTCTTAATAATTATAACTATGAC	
Hmx ovelapEND Rev	GAATATATATGTTGTCTATGATTGT	

Embryos and In situ hybridisation

Naturally spawned *Lampetra planeri* embryos were collected from a shallow stream in the New Forest, UK, under a Permission granted by Forestry England. They were cultured in filtered river water at 16°C and processed for in situ hybridisation as previously described [12]. Adult *Ciona intestinalis* (Type B) were collected from Northney Marina, UK, and maintained in a circulating sea water aquarium at 14°C under constant light. For the CRISPR experiments, Adult *Ciona intestinalis* (Type B) were collected and shipped from the Roscoff marine station (France) and maintained in aquaria at 18°C. Gametes were liberated by dissection, fertilised in vitro and embryos allowed to grow to the desired stage before fixation and storage. Methods for fixation, storage and in situ hybridisation were as previously described [13]. Adult *Branchiostoma lanceolatum* were collected near Banyuls-sur-Mer, France and spawning was induced by heat stimulation [14,15]. Embryos were grown for 36hr at 19°C in natural sea water. Fixation was performed for 2hr on ice in 4% PFA in MOPS buffer containing 0.1M MOPS, 1mM EGTA, 2mM $MgSO_4$ and 500mM NaCl. In situ hybridization was performed as previously described [16].

Lamprey transgenics, imaging and controls

Lamprey uCNE and dCNE sequences from *L. planeri* were amplified by PCR (primers in Table 1) and cloned into the HLC vector with a zebrafish *krt4* minimal promoter [17]. Lamprey

transient transgenesis was performed in *P. marinus* embryos as previously described [17,18]. Briefly, injection mixes consisting of 20ng μl^{-1} reporter plasmid, 1x CutSmart buffer (NEB), and 0.5U μl^{-1} I-SceI enzyme (NEB) in water were incubated at 37°C for 30 minutes and then micro-injected at a volume of approximately 2nl per embryo into lamprey embryos at the one-cell stage. Embryos were then raised and screened for GFP reporter expression using a Zeiss SteREO Discovery V12 microscope. Transient transgenic reporter assays may generate mosaicism in reporter expression patterns, with variation in levels and domains between embryos. 100 embryos were screened for each construct at two stages (25 and 27)

Representative GFP-expressing embryos were first imaged live to record GFP fluorescence, using a Zeiss SteREO Discovery V12 microscope and a Zeiss Axiocam MRm camera with AxioVision Rel 4.6 software. Embryos were then fixed in 4% paraformaldehyde and stained with a Chicken Polyclonal Anti-GFP antibody (Abcam AB13970), and a Mouse HuC/HuD Monoclonal Antibody (Invitrogen 16A11). These were visualised with Goat Anti-Chicken IgY H&L Alexa Fluor® 488 (Abcam AB150169) and anti-Mouse alexa594 (Abcam AB150116). Before imaging, embryos were counterstained with DAPI. Embryos were viewed on an Olympus FV1000 Confocal microscope. Reconstructions and analysis were carried out using FIJI- imageJ v.1.52g [19]. Z stacks and 3D projects of confocal data were built using maximum intensity projection.

Confocal microscopy was able to reveal GFP expressing cells not possible to image in live embryos. Since lamprey transgenesis is a relatively new technique and previous studies have not assessed levels of background at this resolution, we analysed reporter activity in embryos injected with the plasmid vector HLC (Extended data 10), focusing on ganglia and CNS expression that might overlap with endogenous *Hmx* staining and confound interpretation. This revealed GFP expression in skin cells, head muscle and branchiomeric muscle. We also saw occasional expression in CNS and CSG. CNS expression was clearly distinct from that observed with *Hmx* enhancers and did not overlap with *Hmx* gene expression. Ganglia expression was infrequent (22% of embryos analysed: Extended data 10) and did not label the same cells as seen with Hmx uCNE.

Ciona Hmx overexpression and sequence analysis

The plasmid vector containing the *epiB* promoter driving *GFP* (*epiB>GFP*) was kindly provided by Bob Zeller [20]. The full *Hmx* open reading frame was amplified by PCR and cloned downstream of the *epiB* promoter, replacing GFP and creating *epiB::Hmx. Ciona intestinalis* (Type B) Hmx was amplified in two sections, a 5' region using the primers TAAAATAGTAAAATGGTACCTATGACGTCACTGTGCCAATTG and TTCCCCTTCTGACGTAGGGA, and a 3' section using the primers TCCCTACGTCAGAAGGGGAAG and ACCGGCGCTCAGCTGGAATTATGATTGTCTCACACCACGGAA. This resulted in two fragments: each had a homologous arm overlapping the other fragment and a homologous arm overlapping one end of the vector digested with KpnI and EcoRI. The Cold Fusion system (System Biosciences) was used to insert these into the vector via recombination, fusing the 5' end of the resulting full *ciHMX* ORF with the 3' end of the *epiB* promoter. Integrity of the resulting construct was confirmed with sequencing.

Constructs were electroporated into *Ciona intestinalis* (Type B) zygotes as previously described [21]. We first confirmed that these constructs drove their respective transgenes into the epidermis as expected, using GFP live imaging and *Hmx* in situ hybridisation respectively. We then electroporated parallel batches with either *epiB>GFP* only (control) or *epiB>GFP* and *epiB>Hmx* (*Hmx* overexpression) constructs. As each electroporation results

in 100s of growing embryos, some of which are transgenic and some of which are not, embryos were grown to the tailbud stage when GFP was visible, allowing us to identify transgenic embryos. At this stage the epidermis makes up approximately 50% of the total cells of the embryo [22]. Transgenic embryos were then manually selected and processed for RNA extraction. Three full biological replicates were performed on embryo batches derived from different fertilisations. Each biological replicate combined RNA from at least 50 individual embryos.

In summary, each of these 6 samples (three experimental, three control) derives from a minimum of 50 pooled embryos. Each embryo was confirmed as transgenic by GFP expression, and in each embryo the *epiB* promoter was driving expression of the transgene(s) into the epidermis, an ectodermal tissue comprising a substantial proportion of the overall embryo. The epidermis shares germ layer origins with the neural cells that express *Hmx*, but does not itself express *Hmx* in wild-type embryos. All six RNA samples were sequenced by Illumina HiSeq4000 following polyA selection, yielding approximately 28 million paired end 75bp reads per sample.

For differential gene expression analysis, fastQC was used to assess sequencing quality, Trimmomatic [23] to trim off adapters and Sickle [24] to trim low quality reads. Remaining reads were then mapped to the *Ciona intestinalis* (Type A) (*Ciona robusta*) genome (KH2012) using STAR [25]. Differential expression analysis was carried out using the DESeq2 R package [26], using an adjusted p value threshold of 0.01. Finally, a minimum FPKM threshold of 2 was also applied to exclude very lowly expressed transcripts. This yielded a list of genes significantly (adjusted p<0.01) up or downregulated in the *Hmx* overexpression treatment compared to the control. Gene lists deriving from *Ciona intestinalis* (Type A) (*Ciona robusta*) single cell sequencing were extracted from the supplementary files of published literature[27,28], and cross-correlated with the up and down regulated gene lists to provide the annotation of data shown in Figure 2B-D. Full gene lists are in supplementary files 1 and 2.

CNE analysis in transgenic *Ciona*

To test CNE activity in *Ciona*, we cloned the 2kb 5' to *Ciona intestinalis* (Type B) *Hmx*, lamprey uCNE or lamprey dCNE into the reporter vector pCES [29] using the primers shown in Table 1. Constructs were electroporated into *Ciona intestinalis* (Type B) zygotes as above, and embryos stained for β-galactosidase activity as described [30].

CRISPR-Cas9 knockout of *Ciona Hmx* and *Ngn*

Dechorionated eggs were fertilized and electroporated as described[31]. The plasmid vectors, *Ngn>Unc-76:GFP*, *Asic>Unc-76:GFP*, *Fog>H2B:mCherry* and *Fog>Cas9* were kindly provided by Alberto Stolfi[32]. Single guide RNAs (10 for *Hmx* and 4 for *Ngn*) were cloned into the *U6>sgRNA(F+E)* vector (provided by Addgene) as previously established[33] and an unspecific control sgRNA (CTTTGCTACGATCTACATT)[33] was used in every experimental replicate. sgRNA specificities were validated in pairs by electroporation, PCR amplification of targeted regions and Sanger sequencing (Microsynth, Basel, Switzerland). Electroporation mixes were as follows:

<u>Hmx and Ngn CRISPR:</u>

Fog>H2B:mCherry, 10 µg

Fog>Cas9, 30 µg

U6>Hmx sgRNAs (30 µg each) or U6>Ngn sgRNAs (30µg each) or U6>Control sgRNA, 60µg

Ngn>Unc-76:GFP or Asic>Unc-76:GFP or Hmx CNE(2K-E1)>LacZ, 70µg

<u>Hmx overexpression</u>

Fog>H2B:mCherry, 10 µg
epiB>Hmx, 25 µg
Ngn>Unc-76:GFP or Asic>Unc-76:GFP, 70 µg

Primer sequences, target regions and sequencing results for best gRNA pairs producing the phenotypes presented in the manuscript are shown in Extended data 3, and the sequences of these guide RNAs are:.

```
sgHmx6 (rev):          GTGACGTAGACAGGGAACGG CGG
sgHmx10 (rev):         GCAGGGGGCCATGGGAAATG GGG
sgNgn P2 new (rev):    GACGTAACAAAGCATAGCCG CGG
sgNgn 2 new (rev):     ATGCATGCCGGGCCCGCCGT CGG
```

An antibody anti-β galactosidase (Promega) was used to stain *Hmx CNE (2K-E1)>LacZ* positive cells. Samples were mounted in Vectashield and images were obtained using a Leica DM5000 B microscopy.

Transcription Factor Binding Site Prediction
Transcription factor binding sites were searched using vertebrate JASPAR database profiles. Sites predicted to be bound with a probability above or equal to 0.7 were kept. For vertebrate CNEs, predicted sites were compared across different species (same sequences used for phylogenetic tree in Fig. 4D - Human, mouse, chicken, painted turtle, zebrafish, elephant shark, African clawed frog and lamprey), retaining only sites that were conserved at least in lamprey and five other species. For *Ciona* CNE, predicted sites were compared between *Ciona intestinalis* (Type A) (*Ciona robusta*) and *C. savignyi*, retaining only sites conserved between the two species.

References for materials and methods

1 Brunetti, R. *et al.* Morphological evidence that the molecularly determined Ciona intestinalis type A and type B are different species: Ciona robusta and Ciona intestinalis. *J Zool Syst Evol Res* **53**, 186-193, doi:10.1111/jzs.12101 (2015).

2 Lara-Ramirez, R., Poncelet, G., Patthey, C. & Shimeld, S. M. The structure, splicing, synteny and expression of lamprey COE genes and the evolution of the COE gene family in chordates. *Dev Genes Evol* **227**, 319-338, doi:10.1007/s00427-017-0591-6 (2017).

3 Smith, J. J. *et al.* Sequencing of the sea lamprey (Petromyzon marinus) genome provides insights into vertebrate evolution. *Nat Genet* **45**, 415-421, 421e411-412, doi:10.1038/ng.2568 (2013).

4 Smith, J. J. *et al.* The sea lamprey germline genome provides insights into programmed genome rearrangement and vertebrate evolution. *Nat Genet* **50**, 270-277, doi:10.1038/s41588-017-0036-1 (2018).

5 Mehta, T. K. *et al.* Evidence for at least six Hox clusters in the Japanese lamprey (Lethenteron japonicum). *Proc Natl Acad Sci U S A* **110**, 16044-16049, doi:10.1073/pnas.1315760110 (2013).

6 Woolfe, A. *et al.* CONDOR: a database resource of developmentally associated conserved non-coding elements. *BMC Dev Biol* **7**, 100, doi:10.1186/1471-213X-7-100 (2007).

7 Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792-1797, doi:10.1093/nar/gkh340 (2004).

8 Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313, doi:10.1093/bioinformatics/btu033 (2014).

9 Wada, S. *et al.* A genomewide survey of developmentally relevant genes in Ciona intestinalis. II. Genes for homeobox transcription factors. *Dev Genes Evol* **213**, 222-234, doi:10.1007/s00427-003-0321-0 (2003).
10 Brozovic, M. *et al.* ANISEED 2017: extending the integrated ascidian database to the exploration and evolutionary comparison of genome-scale datasets. *Nucleic Acids Res* **46**, D718-D725, doi:10.1093/nar/gkx1108 (2018).
11 Marletaz, F. *et al.* Amphioxus functional genomics and the origins of vertebrate gene regulation. *Nature* **564**, 64-70, doi:10.1038/s41586-018-0734-6 (2018).
12 Lara-Ramirez, R., Patthey, C. & Shimeld, S. M. Characterization of two neurogenin genes from the brook lamprey lampetra planeri and their expression in the lamprey nervous system. *Dev Dyn* **244**, 1096-1108, doi:10.1002/dvdy.24273 (2015).
13 Boorman, C. J. & Shimeld, S. M. Pitx homeobox genes in Ciona and amphioxus show left-right asymmetry is a conserved chordate character and define the ascidian adenohypophysis. *Evol Dev* **4**, 354-365 (2002).
14 Fuentes, M. *et al.* Insights into spawning behavior and development of the European amphioxus (Branchiostoma lanceolatum). *J Exp Zool B Mol Dev Evol* **308**, 484-493, doi:10.1002/jez.b.21179 (2007).
15 Fuentes, M. *et al.* Preliminary observations on the spawning conditions of the European amphioxus (Branchiostoma lanceolatum) in captivity. *J Exp Zool B Mol Dev Evol* **302**, 384-391, doi:10.1002/jez.b.20025 (2004).
16 Holland, P. W. H. Wholemount in situ hybridization to amphioxus embryos. *Methods Mol Biol* **97**, 641-644, doi:10.1385/1-59259-270-8:641 (1999).
17 Parker, H. J., Bronner, M. E. & Krumlauf, R. A Hox regulatory network of hindbrain segmentation is conserved to the base of vertebrates. *Nature* **514**, 490-493, doi:10.1038/nature13723 (2014).
18 Parker, H. J., Sauka-Spengler, T., Bronner, M. & Elgar, G. A reporter assay in lamprey embryos reveals both functional conservation and elaboration of vertebrate enhancers. *PLoS One* **9**, e85492, doi:10.1371/journal.pone.0085492 (2014).
19 Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**, 676-682, doi:10.1038/nmeth.2019 (2012).
20 Tang, W. J., Chen, J. S. & Zeller, R. W. Transcriptional regulation of the peripheral nervous system in Ciona intestinalis. *Dev Biol* **378**, 183-193, doi:10.1016/j.ydbio.2013.03.016 (2013).
21 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).
22 Nakamura, M. J., Terai, J., Okubo, R., Hotta, K. & Oka, K. Three-dimensional anatomy of the Ciona intestinalis tailbud embryo at single-cell resolution. *Developmental Biology* **372**, 274-284, doi:10.1016/j.ydbio.2012.09.007 (2012).
23 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120, doi:10.1093/bioinformatics/btu170 (2014).
24 Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33). (https://github.com/najoshi/sickle, 2011).
25 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).
26 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).
27 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).
28 Sharma, S., Wang, W. & Stolfi, A. Single-cell transcriptome profiling of the Ciona larval brain. *Dev Biol* **448**, 226-236, doi:10.1016/j.ydbio.2018.09.023 (2019).

29 Harafuji, N., Keys, D. N. & Levine, M. Genome-wide identification of tissue-specific enhancers in the Ciona tadpole. *Proc Natl Acad Sci U S A* **99**, 6802-6805, doi:10.1073/pnas.052024999 (2002).
30 Chen, W. C. *et al.* Dissection of a Ciona regulatory element reveals complexity of cross-species enhancer activity. *Developmental Biology* **390**, 261-272, doi:10.1016/j.ydbio.2014.03.013 (2014).
31 Kari, W., Zeng, F., Zitzelsberger, L., Will, J. & Rothbacher, U. Embryo Microinjection and Electroporation in the Chordate Ciona intestinalis. *J Vis Exp*, doi:10.3791/54313 (2016).
32 Kim, K. *et al.* Regulation of Neurogenesis by FGF Signaling and Neurogenin in the Invertebrate Chordate Ciona. *Front Cell Dev Biol* **8**, 477, doi:10.3389/fcell.2020.00477 (2020).
33 Stolfi, A., Gandhi, S., Salek, F. & Christiaen, L. Tissue-specific genome editing in Ciona embryos by CRISPR/Cas9. *Development* **141**, 4115-4120, doi:10.1242/dev.114488 (2014).