HHS Public Access

Author manuscript

Nature. Author manuscript; available in PMC 2023 May 26.

Published in final edited form as:

Nature. 2022 May ; 605(7911): 701–705. doi:10.1038/s41586-022-04742-w.

Hmx gene conservation identifies the origin of vertebrate cranial ganglia

Vasileios Papadogiannis[1,2], **Alessandro Pennati**[1,3], **Hugo J. Parker**[4], **Ute Rothbächer**[3], **Cedric Patthey**[5], **Marianne E. Bronner**[6], **Sebastian M. Shimeld**[1,*]

[1]Department of Zoology, University of Oxford, 11a Mansfield Road, Oxford OX1 3SZ, UK.

[2]Institute of Marine Biology, Biotechnology and Aquaculture, Hellenic Centre for Marine Research, 71500 Gournes, Crete, Greece.

[3]Institute of Zoology and Center of Molecular Biosciences, University of Innsbruck, Technikerstrasse 25, Innsbruck, 6020 Austria.

[4]Stowers Institute for Medical Research, Kansas City, MO 64110, USA.

[5]Department of Radiation Sciences, Umeå University, 901 85 Umeå, Sweden.

[6]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125, USA.

Abstract

[*]Correspondence to sebastian.shimeld@zoo.ox.ac.uk.

Author contributions

V.P., C.P and S.M.S. conceived the study. V. P. conducted lamprey gene expression analysis, CNE identification, analysis of lamprey reporter gene experiments, *Ciona Hmx* expression analysis, *Ciona Hmx* overexpression analysis and RNAseq and the molecular phylogenetic analyses. A.P. conducted *Ciona* CNE identification and reporter gene experiments, tests of lamprey CNE activity in *Ciona*, CRISPR and overexpression reporter analyses, and analysis of *Ciona Hmx* and *Ngn* gene expression. C. P. conducted the amphioxus in situ hybridisation and participated in RNAseq data analysis. H.J.P. conducted the lamprey reporter construct injections and analysis. U.R., M.E.B. and S.M.S. supervised the work. All authors contributing to drafting and editing the manuscript.

Ethics Statement

Lampetra planeri experiments approved by Department of Zoology, University of Oxford Animal Welfare and Ethical Review Board. *Petromyzon marinus* were maintained under the parameters set in accordance with the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health, with protocols approved by the Institutional Animal Care and Use Committees of the California Institute of Technology (lamprey, Protocol #1436–17)

Additional information

Correspondence regarding reprints, permissions and requests for materials should be sent to Sebastian Shimeld (sebastian.shimeld@zoo.ox.ac.uk).

Statistics and Reproducibility

For image data shown in Figure 1 and Extended Data Figure 4: Lamprey images representative of at least 5 embryos for each stage, experiment repeated 3 times. *Ciona* images representative of at least 50 embryos and repeated 3 times (2 times for experiments in Extended Data Figure 4). Amphioxus images representative of two replicates of 20–30 embryos each. Transgenesis experiments on C. intestinalis presented in Figure 2b,d,h were replicated in duplicate, screening a minimum of 50 total embryos for reporter construct activity, while performing gDNA extraction, PCR (presented in Extended Data Figure 3) and Sanger sequencing from pooled embryos (>100) from each of the two replicate batches to confirm CRISPR editing success. Experiments presented in Figure 2 e,e',f,g and Figure 4i,j,k were repeated 4 times, screening at least 50 embryos each time. Lamprey transgenesis (shown in Figure 4b,c and Extended Data Figure 7) was carried out twice, injecting over 500 embryos each time and screening 100 embryos each time. Immunostaining of lamprey transgenic embryos was repeated in triplicate, with a minimum of 10 embryos stained each time.

Competing interests

We have no competing interests to declare.

The evolutionary origin of vertebrates included innovations in sensory processing associated with the acquisition of a predatory lifestyle[1]. Vertebrates perceive external stimuli through sensory systems serviced by cranial sensory ganglia (CSG), whose neurons predominantly arise from cranial placodes; however, understanding the evolutionary origin of placodes and CSGs is hampered by the gulf between living lineages and difficulty in assigning homology between cell types and structures. Here we use the ***Hmx*** gene family to address this question. We show ***Hmx*** is a constitutive component of vertebrate CSG development and that ***Hmx*** in the tunicate ***Ciona*** is necessary and sufficient to drive the differentiation program of Bipolar Tail Neurons (BTNs), cells previously thought neural crest homologs[2,3]. Using ***Ciona*** and lamprey transgenesis we demonstrate that a unique, tandemly duplicated enhancer pair regulated ***Hmx*** in the stem-vertebrate lineage. Strikingly, we also show robust vertebrate ***Hmx*** enhancer function in ***Ciona***, demonstrating that deep conservation of the upstream regulatory network spans the evolutionary origin of vertebrates. These experiments demonstrate regulatory and functional conservation between ***Ciona*** and vertebrate ***Hmx***, and point to BTNs as CSG homologs.

CSG, including the trigeminal, vestibuloacoustic and epibranchial ganglia, relay information from sensory cells to the brain. CSG neurons derive from two sources: cranial placodes provide neurons that delaminate from the cranial ectoderm, and cranial neural crest cells migrate into ganglia providing all of the glia plus some neurons of the trigeminal ganglia. The evolution of neural crest, placodes and CSG form part of the influential ‘New Head Hypothesis’ which posits that these and other innovations underlie the transformation of an ancestral chordate filter feeder into the ancestral vertebrate-type predator[1]. Our molecular and genetic understanding of this transformation has been limited, however, by the substantial anatomical gulf between vertebrates and their nearest living relatives, amphioxus and tunicates, which lack most or all of these characters (Fig. 1a)[4].

To address this gap, we focused on the *Hmx* gene family, which encode homeodomain transcription factors (TF). We previously used transcriptomics to find markers for placode-derived CSG neurons, identifying *Hmx3* as one such gene[5]. Jawed vertebrates have 4 *Hmx* family genes named *Hmx1, Hmx2, Hmx3* and *SOHo*[6], with expression in mouse, chicken, *Xenopus* and zebrafish primarily confined to the central nervous system (CNS) and cranial placodes and the CSG they form (Fig. 1b)[6–13]. Lampreys are members of the earliest-diverging living vertebrate lineage and offer insight into the basal vertebrate state. We identified three *Hmx* genes in lamprey, *HmxA*, *HmxB* and *HmxC*, all expressed in the CNS, cranial placodes and CSG (Fig. 1c–h). Lamprey *Hmx* genes were not expressed in the olfactory placode or in other parts of the peripheral nervous system (PNS) such as the dorsal root ganglia (DRG). This suggests that expression of vertebrate *Hmx* genes in the CNS, posterior placodes and their descendent CSG reflects the ancestral condition. We next examined the expression of the single *Hmx* genes found in amphioxus[14] and tunicates[15]. Amphioxus *Hmx* was expressed in the CNS but not in the PNS (Fig. 1i,j). In embryos of the tunicate *Ciona intestinalis*, *Hmx* was expressed in the CNS and in a subpopulation of PNS cells, the BTNs (Fig. 1k–m). BTNs are born lateral to the neural plate, then delaminate and migrate before connecting epidermal sensory cells to the CNS[3,16]. Previously, they have been likened to neural crest[2,3].

Hmx regulates *Ciona* BTN development

Vertebrate *Hmx* genes are necessary for correct development of CSG[11,17–19]. To test whether *Ciona Hmx* imparts a BTN phenotype in other cells, we used the *Ciona epiB* promoter[20] to drive *Hmx* expression (*epiB>Hmx*) broadly in the embryonic epidermis and compared these embryos to control embryos using transcriptomics (Fig. 2a). Comparison of up- and downregulated genes to cell-type expression profiles extracted from single-cell sequencing data[2,21] showed *Ciona Hmx* upregulated the BTN differentiation program (Fig. 2a, Table 1, Supplementary File 1). We also confirmed this effect experimentally, using BTN-specific GFP reporters driven by enhancers from the *Ngn* and *Asic* genes (Extended Data Fig. 2a), both of which were ectopically expressed after *Hmx* overexpression by co-electroporation with *epiB>Hmx* (Fig. 2b). *Hmx* overexpression also suppressed expression of epidermal genes and genes associated with palp sensory cells, an anterior sensory cell population that does not express *Hmx* (Fig. 2a, Table 1). This suggests that *Hmx* is sufficient to drive the BTN transcriptional program in *Ciona* embryonic epidermis. BTN neurogenesis is also driven by *Neurogenin* (*Ngn*), in turn activated by *FGF9*[22]. *Ngn* was strongly upregulated after *Hmx* overexpression (4.8x vs control, padj=1.02E-07: Supplementary File 1), prompting us to compare the downstream targets of *Hmx* with those previously characterised for *Ngn*[22]. We found a large overlap in the downstream network (87/291 genes), including 36 of 41 BTN markers controlled by *Hmx* (Fig. 2c, Supplementary File 2). Strikingly, *Hmx* was also upregulated by *Ngn* overexpression, suggesting a circuit built on a feedback loop between *Hmx* and *Ngn*. *POU IV* may also be involved in this circuit, as it is upregulated by *epiB>Hmx* (Fig. 2a) and upregulates *Ngn*[23].

To dissect the role of *Hmx* in BTN fate determination, we investigated both downstream and upstream interactions within the BTN network. First, to test whether *Hmx* was also necessary for the activation of the BTN network we performed CRISPR-Cas9 knockout of *Hmx* and characterized embryos using BTN-specific GFP reporter expression driven by enhancers from the *Ngn* and *Asic* genes (Fig. 2d; Extended Data Fig. 2b; Extended Data Fig. 3a–d). BTN reporter signal was lost in a significant majority of embryos compared to controls (1.45E-07 and 8.33E-04, chi square test, *Ngn* and *Asic* respectively: Extended Data Table 1). Furthermore, comparison of tunicate genomes[24] showed 2kb of sequence 5’ to the *Hmx* transcription start site to be conserved between *Ciona* species (Extended Data Fig. 2c). We hypothesized this was a conserved non-coding element (CNE), with sequence conservation reflecting evolutionary constraint deriving from a role in gene regulation. We tested this in transgenic *Ciona* (Extended Data Fig. 2d) and found the 2kb fragment able to drive robust and specific reporter expression in BTNs and in part of the CNS, recapitulating aspects of endogenous *Hmx* expression (Fig. 2e–g). To further understand BTN network activation, we carried out CRISPR-Cas9 knockout of *Ngn* (Fig. 2h; Extended Data Fig. 2b, Extended Data Fig. 3e–h) and found loss of *Hmx* CNE activity in co-electroporated embryos compared to controls (1.50E-02, chi squared: Extended Data Table 2). We also examined gene expression in early developmental stages, revealing *Ngn* expression precedes *Hmx* (Extended Data Fig. 4). *FGF9*, which activates *Ngn*[22], was downregulated by *Hmx* overexpression (Supplementary File 1). Integrating these data gives a model for BTN specification in which FGF9 kick-starts the circuit via *Ngn,* which initiates an *Hmx-Ngn*

feedback loop that upregulates other BTN genes, plus represses *FGF9* and epidermal and palp genes (Fig. 2i).

Evolution of *Hmx* gene architecture

The expression domains and developmental roles of *Hmx* genes in vertebrates and *Ciona* may reflect shared ancestry, or derived traits in one or both lineages. Discriminating between these requires comparison of the genetic program upstream of *Hmx* in each lineage. To address this, we mapped and tested *Hmx* regulatory elements in lamprey. Jawed vertebrate *Hmx* genes are located in two paralogous two-gene clusters[6]. We found that the jawless vertebrate *Hmx* genes are in a single three-gene cluster in both lamprey and hagfish genomes (Fig. 3a,b). Sequence comparison shows these genomic arrangements evolved by gene duplication (Fig. 3e). In the vertebrate stem lineage, a single Hmx gene duplicated in tandem to yield a two-gene cluster. In jawless vertebrates a second tandem duplication yielded the three-gene cluster state found in lamprey and hagfish. While all three lamprey *Hmx* genes have an identical homeobox, *HmxA* and *HmxC* share additional conserved sequence outside the homeobox and group together in molecular phylogenetic analysis (Fig. 3a: Extended Data Fig. 5a; Extended Data Fig. 5b). This shows *HmxA* and *HmxC* were separated by a duplication within the lamprey/hagfish lineage. In jawed vertebrates the two-gene cluster duplicated as a block to form the paralogous two-gene clusters *Hmx3-Hmx2* and *Hmx1-SOHo* (Fig. 3e).

Genome comparisons across jawed vertebrate *Hmx* loci revealed two CNEs, both associated with the *Hmx3-Hmx2* locus. One (uCNE) lies 5' of *Hmx3*, the other (dCNE) lies between *Hmx3* and *Hmx2*, 5' of *Hmx2* (Fig. 3b). Both are over 1kb long (Supplementary Files 3–5), exceptionally large for ancient vertebrate CNEs[25,26]. Unusually, they are also homologous to each other, sharing a conserved core of around 500bp (Fig. 3c: Supplementary Files 3–5). We searched for these CNEs in lamprey and hagfish, identifying one 5' of *HmxA*, and a second between *HmxB* and *HmxC* (Fig. 3b: Supplementary Files 3–5). We also compared the lamprey and hagfish *Hmx* locus to the jawed vertebrate *Hmx1-SOHo* locus, but this did not reveal any shared conserved elements. Molecular phylogenetic analysis confirmed the orthology of the lamprey/hagfish CNEs to jawed vertebrate uCNE and dCNE respectively (Fig. 3d). These data show uCNE and dCNE originally evolved by tandem duplication of one ancestral CNE (udCNE: Fig. 3c,e) in the stem vertebrate lineage, over 500 million years ago[27]. Parsimony suggests this happened at the same time as the ancestral *Hmx* gene was duplicated. Only the *Hmx3-Hmx2* locus has retained its association with this CNE, so can be regarded as ancestral in this context. In jawed vertebrates *Hmx1* and *SOHo* do not have uCNE or dCNE but retain aspects of ancestral expression (Extended Data Fig. 1) including broadly in CSG, showing enhancers that are lineage-specific or have not been conserved at the sequence level are also part of the overall regulation of vertebrate *Hmx* genes.

Conservation of *Hmx* gene regulation

To test the functions of uCNE and dCNE we generated lamprey embryos transgenic for reporter constructs (Fig. 4a)[28]. Both lamprey CNEs drove reporter expression in the CNS in a pattern similar to endogenous *Hmx* gene expression (Fig. 4b–g). Confocal imaging showed

uCNE was also able to drive reporter expression into CSG derivatives including the facial, glossopharyngeal and vagus nerves, as well as into some structures which do not derive from CSG or express *Hmx* (Fig. 4f–g'). These data confirm that the CNEs are regulatory elements which capture aspects of the spatial expression of lamprey *Hmx* genes. Since they evolved by duplication and divergence from udCNE, we deduce this ancestral element would also have had the capacity to drive gene expression into CNS and CSG (Extended Data Fig. 6).

This deduced ancestral vertebrate regulation of *Hmx* has similarity to the regulation of *Hmx* in living *Ciona,* with both including a large, proximal CNE driving expression into both PNS and CNS. This raises the possibility that regulation in the two lineages is homologous, in turn predicting BTNs and CSG will share the regulatory environment needed to activate *Hmx* expression. While the *Ciona* and vertebrate regulatory elements do not show sequence conservation, TF binding site prediction revealed a large number of common candidate upstream regulators (35: Supplementary File 6). 16 were shared with both vertebrate CNEs, 16 with dCNE only, and 3 with uCNE only. When we tested the activity of lamprey CNEs in transgenic *Ciona* (Fig. 4h) we found uCNE was not functional but dCNE was able to recapitulate *Ciona Hmx* expression in BTNs and CNS (Fig. 4i–k). BTN expression was specific and robust (Fig. 4i–k). CNS expression extended along the majority of the neural tube (Fig. 4i), encompassing cells in the anterior *Ciona* CNS that express *Hmx*, but resembling the more extensive CNS expression of *Hmx* genes and reporters in amphioxus and vertebrates (compare Fig. 4i–k with Fig. 4b–e and Fig. 1i–m).

Discussion

It has been previously suggested that BTNs are homologs of the neural crest[3], and this has been elaborated into a gene regulatory model in which the *Ciona* neural plate border divides under an anterior-posterior patterning network into a posterior 'proto-neural crest' domain and anterior 'proto-placode' domain[2]. However, the *Hmx* family constitutively marks vertebrate placodes and CSG and is necessary for proper CSG neuron development[11,17–19]. In contrast, while *Hmx* expression has been reported later in development in mouse and chicken DRG, *Hmx* function is not required for the specification of their neural crest derived neurons[11]. Furthermore, *Hmx* expression is not observed in neural crest in vertebrates including *Xenopus*, zebrafish, medaka and lamprey[6,8,9,13] (Fig. 1c–h: Extended Data Fig. 1). This identifies CSG and not neural crest as the shared site of vertebrate *Hmx* expression. Expression in some neural crest derivatives in some vertebrates therefore reflects derived evolutionary co-option of *Hmx*, possibly to help form specific sensory neurons deriving from a different cellular source following the transit of neural crest precursors through a reacquired pluripotent state[29,30]. Our data hence point to BTNs as homologous to CSG neurons. This is reinforced by finding that a vertebrate *Hmx* CNE drives expression of a reporter in *Ciona* BTNs. Furthermore, this matches the embryonic origin of both BTNs and placodes as lateral to the neural plate. *Ciona* also produces *Hmx*-negative PNS cells from the anterior neural plate border. These cells express markers of vertebrate chemosensory and GnRH neurons[2], which in vertebrates develop from the olfactory placode. Our data are hence in keeping with proposals that the ancestral neural plate border had two domains yielding PNS cells[2,31]. Both, however, are homologous to placode-derived neurons. Both may also lie within the *Otx*-expressing anterior region in *Ciona*[32].

The evolution of *Hmx* in tunicates and vertebrates parallels derived aspects of neural evolution and includes several unusual features. Tunicates are secondarily-reduced[33], and we see evidence for how this evolved in the broad activity of vertebrate dCNE in the *Ciona* CNS. This shows the gene regulatory environment necessary for broad ancestral *Hmx* expression has persisted in *Ciona*, but that *Ciona Hmx* expression has become restricted to the anterior CNS by changes in its *cis*-regulation. Following divergence of the tunicate and vertebrate lineages, *Hmx* evolved by tandem duplication of gene and CNE. That both CNEs have been conserved and maintained in tandem over the remainder of vertebrate evolution appears to be unique. These *Hmx* CNEs are also unusually large at around 1kb in length, with similarly-aged vertebrate CNEs much smaller[26]. While the functional implications of these unusual features are unknown, such conservation speaks to extreme evolutionary constraint. We speculate this stems from a requirement for robustness in interacting with the ancient *trans* regulatory environment that evolved in the stem lineage of the tunicates and vertebrates, perhaps reflecting an instructive role for *Hmx* in directing the development of ancient cell types including ones involved in environmental sensing.

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[35]. 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.

Vertebrate *Hmx* gene expression survey

To develop an overview of vertebrate *Hmx* gene expression by cluster and paralogue group we extracted descriptions of relevant expression from published literature [5–10,12,13,18,36–45]. The outcome of this analysis is shown in Extended Data Figure 1.

Hmx gene identification, cloning and sequence analyses

We searched multiple sources of lamprey and hagfish sequence data for potential *Hmx* genes using the BLAST+ suite (2.7.1). For lamprey this constituted *Lampetra planeri* transcriptome data [46], genome assemblies for *Petromyzon marinus* [47,48], a *P. marinus* transcriptome assembly built in-house from Illumina GAII data available on SRA, the *Lethenteron camtschaticum* genome assembly [49], 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 (http://condor.fugu.biology.qmul.ac.uk/, now part of the UCNE database - https://ccg.epfl.ch/UCNEbase/external_search.php) [50]. 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 (3.8.31) [51] and RAxML (8.2.12) [52] using the Maximum Likelihood method. 1000 bootstrap replicates were used to assess node confidence. Accession numbers and sequences used in molecular phylogenetic analyses are in Supplementary File 8 and Supplementary File 9.

Lamprey *Hmx* genes were cloned from *L. planeri* cDNA, and uCNE and dCNE sequences from *L. planeri* genomic DNA, using the primers shown in Supplementary Table 1. The *Ciona intestinalis* (Type A) *Hmx/Nkx5* locus was already annotated [15], 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 Supplementary Table 1, and used this for in situ hybridisation. We then used homology to *Ciona savignyi*, coupled with RNAseq data mapped on ANISEED [24], to identify the full open reading frame. This was amplified, in two sections, one 5’ and one 3’ using the primers shown in Supplementary Table 1 and cloned into the vector using the Cold Fusion system (System Biosciences). *Branchiostoma lanceolatum Hmx* was identified by searching the genome [53]. 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.

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 [54]. 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 [55]. We did not examine gene expression in adult animals. Adult

Branchiostoma lanceolatum were collected near Banyuls-sur-Mer, France and spawning was induced by heat stimulation [56,57]. 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 [58].

Lamprey transgenics, imaging and controls

Lamprey uCNE and dCNE sequences from *L. planeri* were amplified by PCR (primers in Supplementary Table 1) and cloned into the HLC vector with a zebrafish *krt4* minimal promoter [28]. Lamprey transient transgenesis was performed in *P. marinus* embryos as previously described [28,59]. 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) at 1:1000, and a Mouse HuC/HuD Monoclonal Antibody (Invitrogen 16A11) at 1:500. These were visualised with Goat anti-Chicken IgY H&L Alexa Fluor® 488 (Abcam AB150169) at 1:1000 and anti-Mouse alexa594 (Abcam AB150116) at 1:1000. 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 [60]. 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 7), 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 Fig. 7) 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 Robert 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 [61]. 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 [62]. 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 (0.11.7) was used to assess sequencing quality, Trimmomatic (0.31) [63] to trim off adapters and Sickle [64] to trim low quality reads. Remaining reads were then mapped to the *Ciona intestinalis* (Type A) (*Ciona robusta*) genome (KH2012) using STAR (2.7.0c) [65]. Differential expression analysis was carried out using the DESeq2 (1.34.0) R package [34], 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[2,21], and cross-correlated with the up and down regulated gene lists to provide the annotation of data shown in Figure 2a,c. 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 [66] using the primers shown in Supplementary Table 1. Constructs were electroporated into *Ciona intestinalis* (Type B) zygotes as above, and embryos stained for β-galactosidase activity as described [67].

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

Dechorionated eggs were fertilized and electroporated as described[68]. The plasmid vectors, *Ngn>Unc-76:GFP*, *Asic>Unc-76:GFP*, *Fog>H2B:mCherry* and *Fog>Cas9* were kindly provided by Alberto Stolfi[22]. *Ngn>Unc-76:GFP* and *Asic>Unc-76:GFP* recapitulate *Ngn* and *Asic* expression respectively in the BTNs[22]. 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[69] and an unspecific control sgRNA (CTTTGCTACGATCTACATT)[69] 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:

Hmx and Ngn CRISPR:

Fog>H2B:mCherry, 10 μg

Fog>Cas9, 30 μg

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

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

Hmx overexpression

Fog>H2B:mCherry, 10 μg

epiB>Hmx, 25 μg

Ngn>Unc-76:GFP or Asic>Unc-76:GFP, 70 μg

Primer design, target regions and sequencing results for best gRNA pairs producing the phenotypes presented in the manuscript are shown in Extended Data Fig. 3, and the sequences of these guide RNAs (which were designed specifically to *Ciona intestinalis* (Type B)) are:.

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

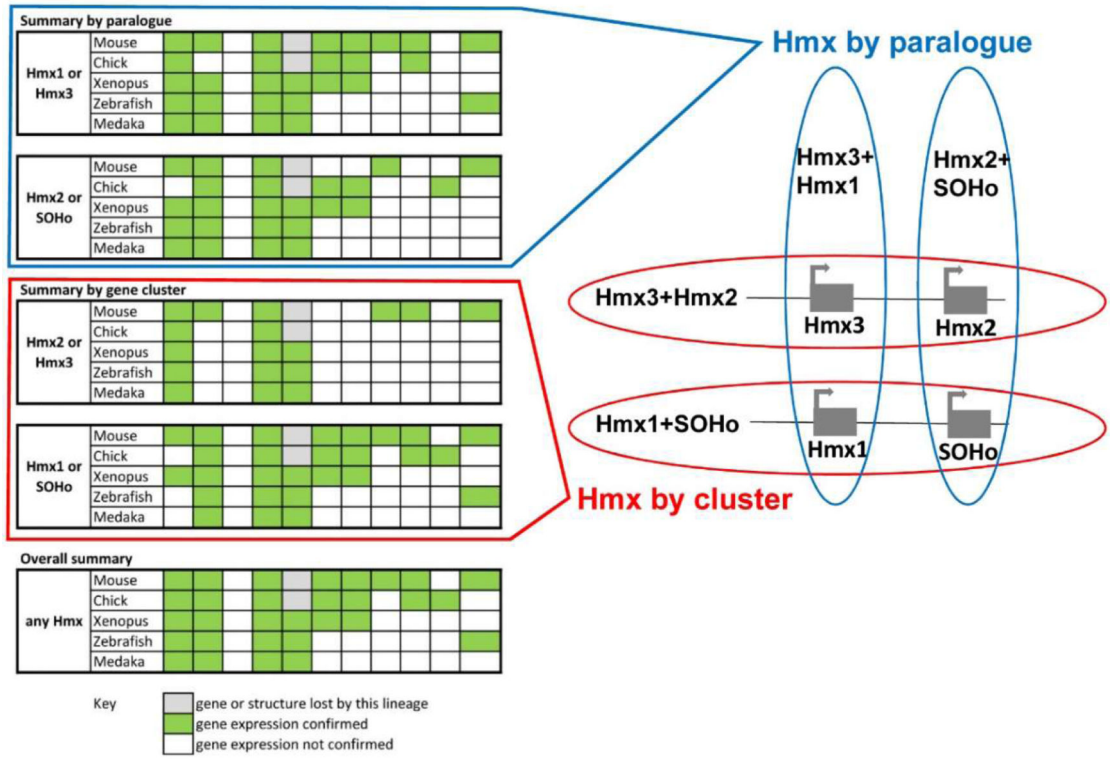
An antibody against anti-β galactosidase (Promega Z3781) at 1:1000 was used to stain *Hmx CNE (2K-E1)>LacZ* positive cells. Samples were mounted in Vectashield and images were obtained using Leica DM5000 B microscopy.

Transcription Factor Binding Site Prediction

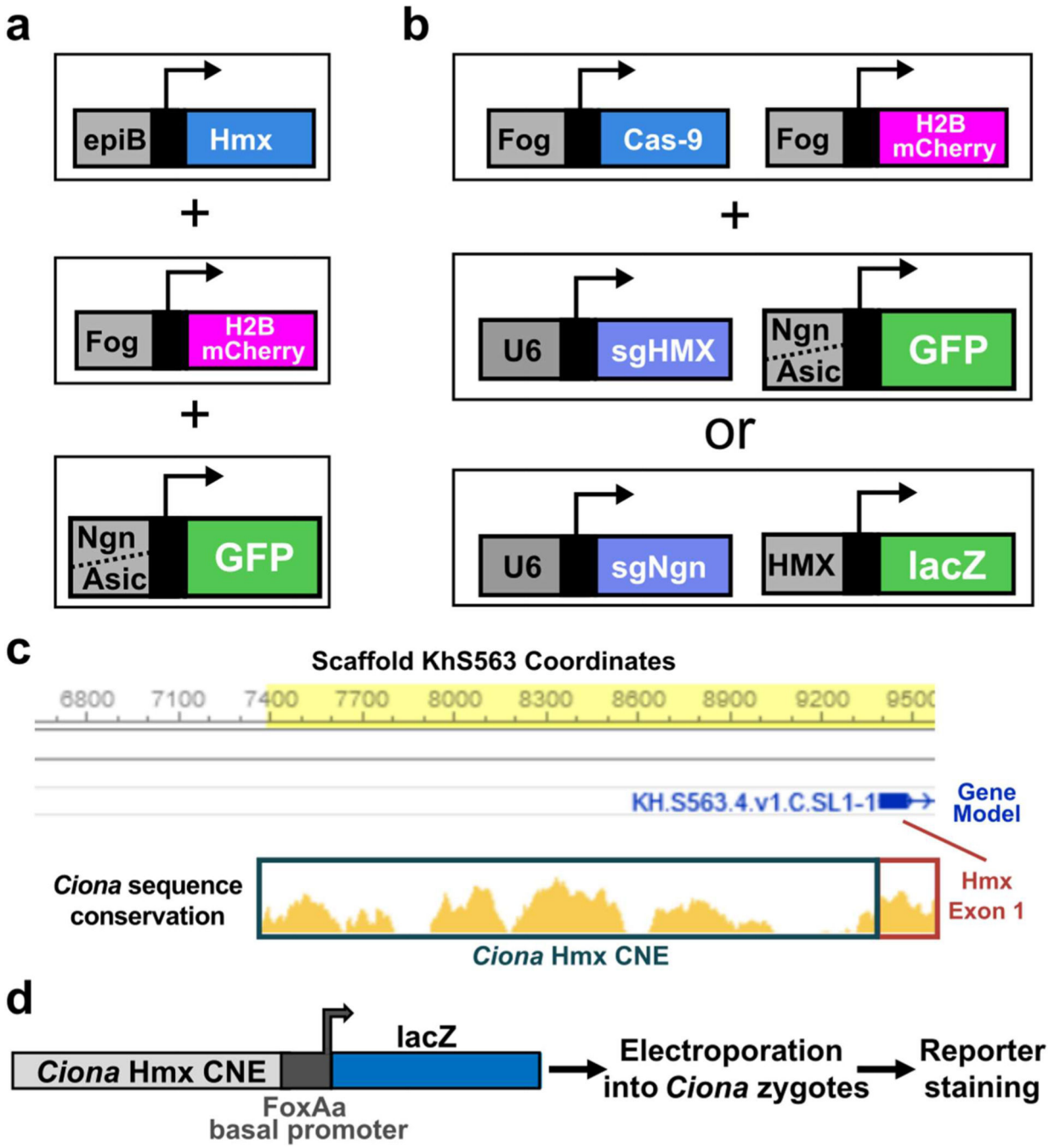
Transcription factor binding sites were searched using vertebrate JASPAR database (https://jaspar.genereg.net/) profiles. Sites predicted to be bound with a probability above or equal to 0.7 were kept. For vertebrate CNEs, predicted sites were compared across different species (same sequences used for phylogenetic tree in Fig. 3d - 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.

Extended Data

		CNS		CSG/Placodal					Other PNS			Other	
Gene	Species	Brain/spinal cord	Eye	Olfactory	Otic/VA	Lateral line	Trigeminal	Epibranchial	Ciliary ganglion	DRG	Sympathetic ganglia	Branchial arch	References
Hmx1	Mouse	■	■		■		■	■		■		■	Wang et al 2000, Quina et al 2012, Munroe et al 2009
	Chick						■	■		■			Adameyko et al 2009, Patthey et al 2016
	Xenopus		■		■	■	■	■					Kelly and El-Hodiri 2016
	Zebrafish		■		■							■	Hartwell et al 2019, Feng and Xu 2010, Boisset and Schorderet 2016
	Medaka		■										Adamska et al 2001
Hmx2	Mouse	■	■		■				■			■	Wang et al 2000
	Chick												
	Xenopus												
	Zebrafish	■			■	■							Feng and Xu 2010, Liu et al 2015
	Medaka	■			■	■							Adamska et al 2001
Hmx3	Mouse	■	■		■				■	■		■	Wang et al 2000, Bober et al 1994
	Chick	■			■								Patthey et al 2016
	Xenopus	■			■	■							Bayranov et al 2004
	Zebrafish	■			■	■							Feng and Xu 2010, Adamska et al 2000
	Medaka	■			■	■							Adamska et al 2001
SOHo	Mouse												
	Chick		■								■		Kiernan et al 1997, Takahashi et al 2003, Apostolova et al 2007
	Xenopus	■	■		■	■	■	■					Kelly and El-Hodiri 2016
	Zebrafish		■		■	■							Feng and Xu 2010
	Medaka		■		■	■							Adamska et al 2001

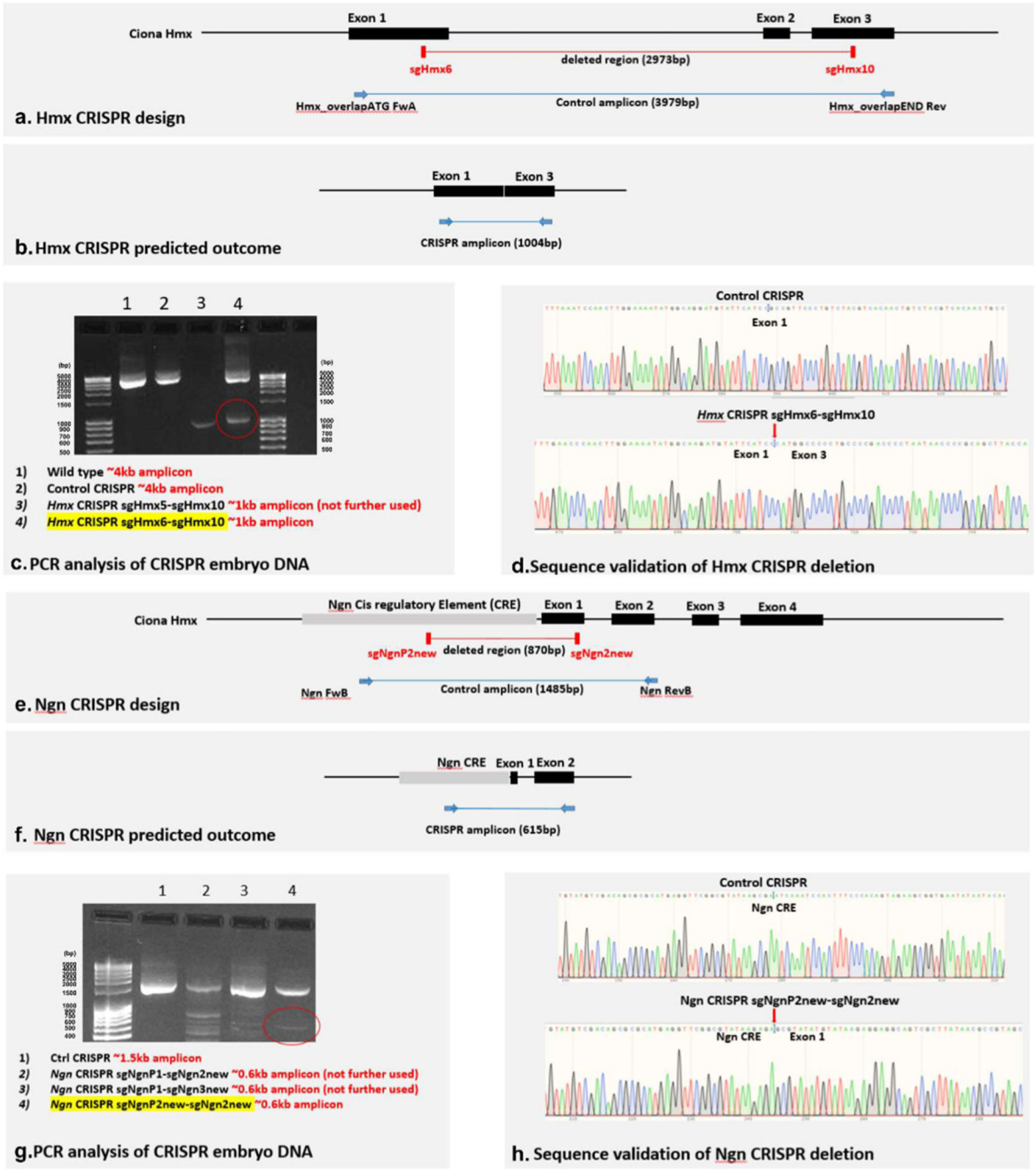


Extended Data Fig. 1. Expression of jawed vertebrate *Hmx* genes in neural derivatives.
The summaries show expression by gene cluster, by genome duplication paralogue (as in the associated diagram), or overall.



Extended Data Fig. 2. Schematics of experimental strategies for reporter assays and *Ciona Hmx* CNE identification.

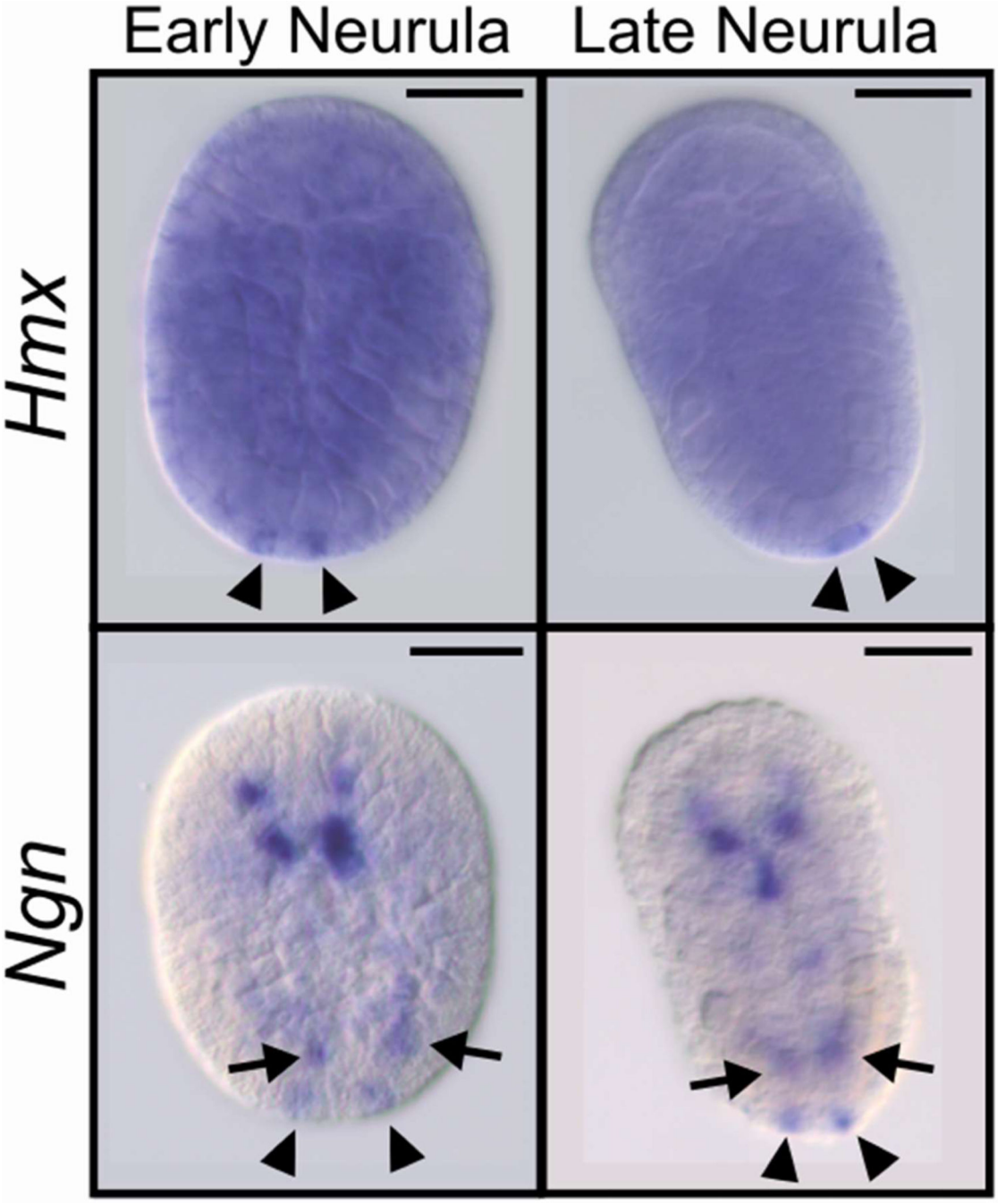
a. *Hmx* overexpression in *Ciona*. **b.** *Hmx* or *Ngn* CRISPR Cas9 knockout in *Ciona*. **c.** *Ciona Hmx* CNE identification. Approximately 2Kbp 5' to the first *Hmx* exon in *Ciona intestinalis* (Type A) (*Ciona robusta*) scaffold KhS563 is shown, with conservation to the *Hmx* locus in C*iona savignyi* shown below. **d.** *Ciona Hmx* CNE analysis in *Ciona*.



Extended Data Fig. 3. CRISPR-Cas9 knockout of *Ciona Hmx* and *Ciona Ngn*.

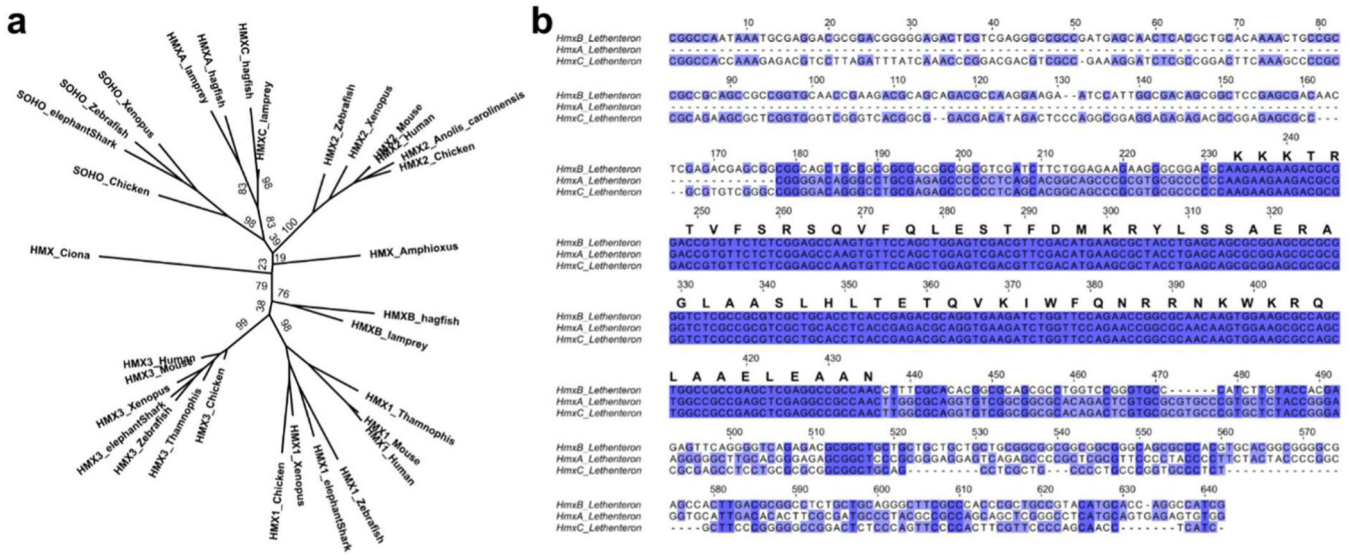
a. Placement of sgRNA guides for *Hmx* CRISPR knockout and primers used for validation, relative to gene structure. Guide and primer sequences in Methods and Supplementary Table 1. **b**. Predicted engineered outcome of *Hmx* CRISPR knockout. **c**. PCR amplification of *Ciona intestinalis* (Type B) genomic DNA from wild type, CRISPR control and *Hmx* CRISPR embryo DNA (as well as from additional sgRNAs that were tested but not used in further experiments). The guide used in further experiments is marked in yellow. Sizes of bands in the DNA ladder (100bp DNA-Ladder, extended: Carl Roth) are given in base

pairs (bp). **d**. Sequencing of amplified bands with sequence identity matching the predicted outcomes in (**b**). **e**. Placement of sgRNA guides for *Ngn* CRISPR knockout and primers used for validation, relative to gene structure. Guide and primer sequences in Methods and Supplementary Table 1. **f**. Predicted engineered outcome of *Ngn* CRISPR knockout. **g**. PCR amplification of *Ciona intestinalis* (Type B) genomic DNA from wild type, CRISPR control and *Ngn* CRISPR embryo DNA (as well as from additional sgRNAs that were tested but not used in further experiments). The guide used in further experiments is marked in yellow. Sizes of bands in the DNA ladder (100bp DNA-Ladder, extended: Carl Roth) are given in base pairs (bp). **h**. Sequencing of amplified bands with sequence identity matching the predicted outcomes in (**f**).



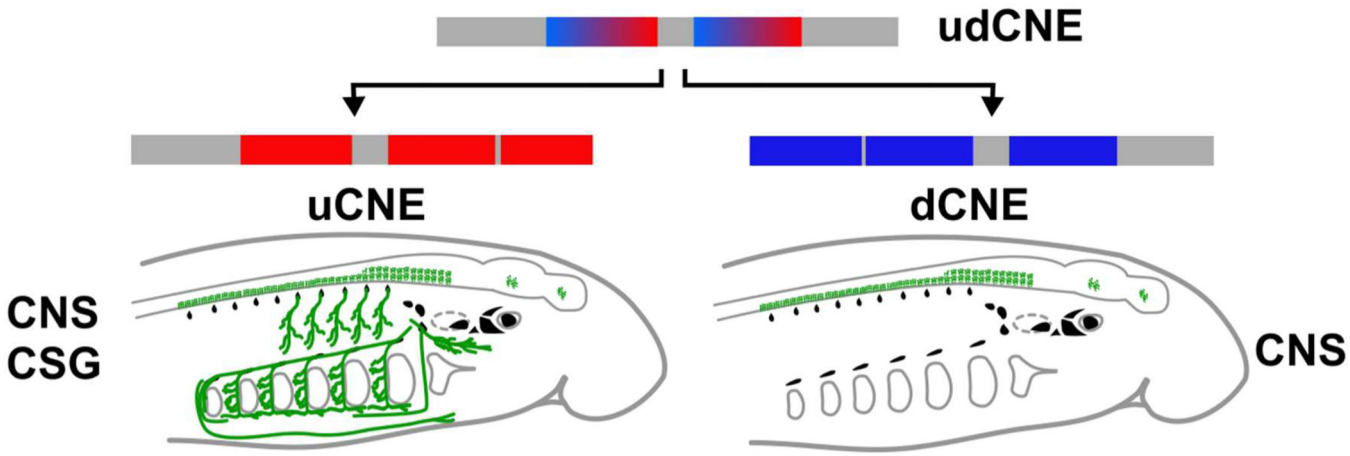
Extended Data Fig. 4. Early developmental expression of *Hmx* and *Ngn* in *C. intestinalis* (Type B).

Gene expression was analysed by whole mount *in situ* hybridisation. Only posterior BTNs (arrowheads) are marked by faint *Hmx* expression during neurula stages, while *Ngn* is expressed in posterior BTNs (arrowheads) and anterior BTNs (arrows) and the CNS. Scale bars 100μM.



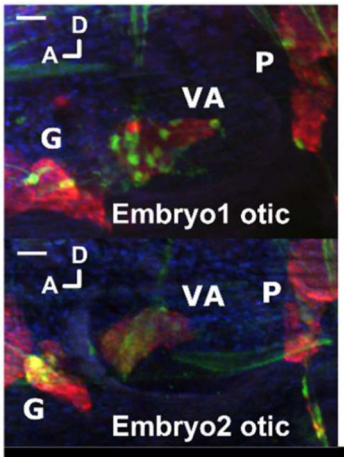
Extended Data Fig. 5. Molecular phylogenetic analysis of chordate Hmx sequences and alignment of lamprey Hmx sequences.

a. This phylogenetic analysis includes Hmx sequences from amphioxus and *Ciona*. The analysis was conducted using the Maximum Likelihood method and numbers indicate percentage node support out of 1000 bootstraps. **b.** Lamprey *HmxA*, *HmxB* and *HmxC* nucleotide sequence alignment. The translation shows the identical homeodomain amino-acid sequence encoded by all three genes. *HmxA* and *HmxC* share additional nucleotide sequence identity before and after the homeodomain encoding sequence. Nucleotide sequences are from the lamprey *Lethenteron camtschaticum.*



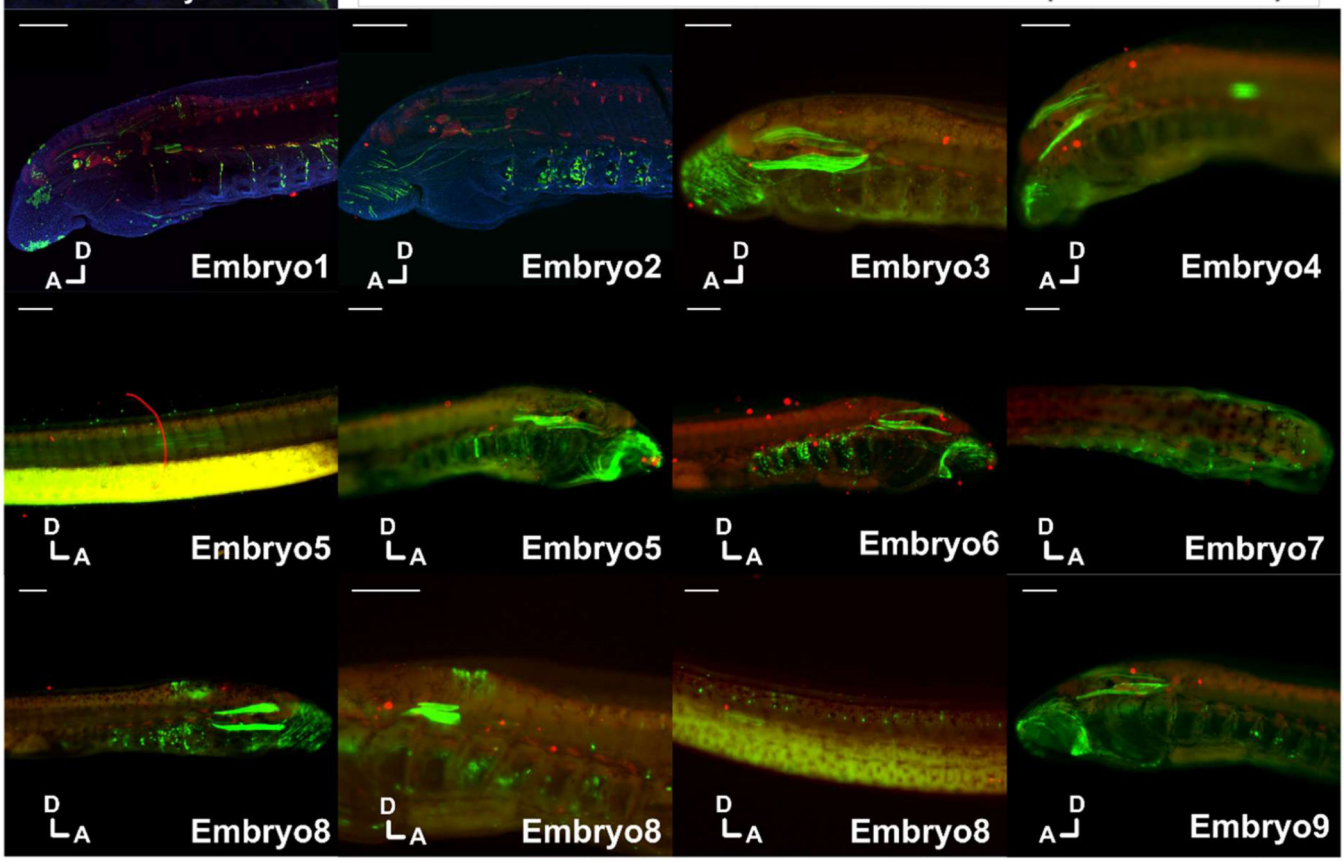
Extended Data Fig. 6. Model of evolution of vertebrate *Hmx* uCNE and dCNE from an ancestral udCNE.

CNE activity is shown in green on the embryo diagrams in the Central Nervous System (CNS) and Cranial Sensory Ganglia (CSG).



	CSG	Brain*	Spinal Cord	Skin	Face Muscle	Branchial Arch Capillary muscle
embryo 1	■	■	■	■		
embryo 2	■	■		■	■	■
embryo 3		■		■	■	■
embryo 4				■	■	
embryo 5			■	■	■	■
embryo 6				■	■	■
embryo 7				■		
embryo 8		■	■	■	■	■
embryo 9				■	■	■
TOTAL	2	4	3	9	7	6
%	22%	44%	33%	100%	78%	67%

*Circuit between Dorsal Midbrain and Dorsal Posterior Hindbrain (3rd-4th branchial arch)



Extended Data Fig. 7. Assessment of background deriving from the vector used to generate lamprey transgenics.

Embryos were injected with vector only (which includes the zebrafish *krt4* minimal promoter and reporter gene but no cloned enhancer), allowed to develop then fixed and labelled for DNA (DAPI, blue), GFP (green) and Hu/ELAV (red) before analysis by confocal microscopy. Each embryo was scored for expression in multiple tissues, as shown in the table at the top right of the picture, with a focus on tissues overlapping with *Hmx* expression. D and A indicate dorsal and anterior orientations for each image. CSG, cranial sensory ganglia. G, geniculate ganglion. VA, vestibuloacoustic ganglion. P, petrosal ganglion. Spinal cord expression was confined to isolated cells and distinct from

the consistent column of expression generated by *Hmx* enhancers (see Figure 4, main text). Brain expression appeared in the dorsal hindbrain and midbrain and was also distinct from *Hmx* and *Hmx* enhancer expression. In two embryos (1 and 2 below: also in high magnification in top left focused on the otic area) close examination revealed scattered cells around some cranial ganglia, including a few co-expressing Hu/ELAV. These differed from those labelled with *Hmx* enhancers in that GFP staining did not penetrate into axons. Scale bars 100μM except for the high magnification views of the otic region where they are 10μM.

Extended Data Table 1.

Quantification of BTN signal in control and *Hmx* CRISPR knockout embryos using the *Ngn* or *Asic* markers. Images in Fig. 2d.

Readout	**CRISPR control**		***Hmx* CRISPR 6–10**		**Chi squared (control vs. *Hmx* CRISPR**
	Embryo count	**% with marked BTN**	**Embryo count**	**% with marked BTN**	
Foq>H2B:mCherry	61	85%	52	29%	1.45E-07
BTNs *Ngn*	52		15		
Fog>H2B:mCherry	43	84%	50	34%	8.33E-04
BTNs *Asic*	36		17		

Extended Data Table 2.

Quantification of *Ciona Hmx* CNE activity in control and *Ngn* CRISPR knockout embryos. Images in Fig. 2h.

Readout	**CRISPR control**		***Ngn* CRISPR 6–10**		**Chi squared (control vs. *Ngn* CRISPR**
	Embryo count	**% with marked BTN**	**Embryo count**	**% with marked BTN**	
Fog>H2B:mCherry	50	98%	49	65%	1.50E-02
BTNs: *Hmx CNE (2K-E1)>lacZ*	49		32		

Extended Data Table 3

Full count data for *Ciona* Hmx −2kb transgenic reporter analysis (Figure 2e–g) and for dCNE reporter analysis in *Ciona* transgenic embryos (Figure 4h).

Full count data for *Ciona Hmx* −2kb transgenic reporter analysis (Figure 2E–G)	**Number of embryos**
Tailbud stage	
Total surviving	86
of which	
BTN only stained	46
No stain	40
Larval stage	
Total surviving	72

Full count data for *Ciona Hmx* −2kb transgenic reporter analysis (Figure 2E–G)	**Number of embryos**
of which	
BTN only stained	25
BTN and CNS stained	13
No stain	34
Full count data for dCNE reporter analysis in *Ciona* transgenic embryos (Figure 4H)	
Tailbud stage	
Total surviving	74
of which	
BTN only stained	23
BTNs and CNS stained	29
CNS only stained	0
No stain	22
Larval stage	
Total surviving	208
of which	
BTN only stained	91
BTNs and CNS stained	34

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We are grateful to Prof. Alberto Stolfi and Prof. Robert Zeller for sharing plasmids used in the *Ciona* CRISPR and overexpression studies, respectively. We thank Dr Hector Escriva for hosting C.P. and access to his amphioxus facility. We thank Stephen Green for lamprey husbandry assistance. V.P. was supported by a Natural Motion scholarship. V.P. and S.M.S. acknowledge the Elizabeth Hannah Jenkinson fund for financial support. V.P. also thanks Dr Tereza Manousaki and Dr Costas Tsigenopoulos for their support while based in HCMR. A.P. was supported by the Accademia Nazionale dei Lincei while working in Oxford and by the H2020 Marie Sklodowska-Curie COFUND ARDRE to U.R. while working in Innsbruck. C.P. was supported by an EMBO Long Term Fellowship while working in Oxford. M.E.B. acknowledges support from award R35NS111564 from the NIH. H.J.P. was supported by funds from the Stowers Institute (grant #1001).

Data Availability

Cloned *Hmx* gene sequences have been deposited in Genbank accessions MN264670-MN264672. RNAseq data have been deposited in SRA accession GSE141046. Original data underlying Figures 4b–c of this manuscript can be accessed from the Stowers Original Data Repository at [http://odr.stowers.org/websimr/].

References

1. Northcutt RG & Gans C. The genesis of neural crest and epidermal placodes: a reinterpretation of vertebrate origins. Q Rev Biol 58, 1–28 (1983). [PubMed: 6346380]

2. 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). [PubMed: 30069052]

3. Stolfi A, Ryan K, Meinertzhagen IA & Christiaen L. Migratory neuronal progenitors arise from the neural plate borders in tunicates. Nature 527, 371–374, doi:10.1038/nature15758 (2015). [PubMed: 26524532]

4. Shimeld SM & Holland PW Vertebrate innovations. Proc Natl Acad Sci U S A 97, 44494452, doi:10.1073/pnas.97.9.4449 (2000).

5. Patthey C. et al. Identification of molecular signatures specific for distinct cranial sensory ganglia in the developing chick. Neural Dev 11, 3, doi:10.1186/s13064-016-0057-y (2016). [PubMed: 26819088]

6. Adamska M. et al. Five Nkx5 genes show differential expression patterns in anlagen of sensory organs in medaka: insight into the evolution of the gene family. Dev Genes Evol 211, 338–349 (2001). [PubMed: 11466530]

7. Wang W, Lo P, Frasch M. & Lufkin T. Hmx: an evolutionary conserved homeobox gene family expressed in the developing nervous system in mice and Drosophila. Mech Dev 99, 123–137 (2000). [PubMed: 11091080]

8. Feng Y. & Xu Q. Pivotal role of hmx2 and hmx3 in zebrafish inner ear and lateral line development. Dev Biol 339, 507–518, doi:10.1016/j.ydbio.2009.12.028 (2010). [PubMed: 20043901]

9. Kelly LE & El-Hodiri HM Xenopus laevis Nkx5.3 and sensory organ homeobox (SOHo) are expressed in developing sensory organs and ganglia of the head and anterior trunk. Dev Genes Evol 226, 423–428, doi:10.1007/s00427-016-0555-2 (2016). [PubMed: 27392729]

10. Kiernan AE, Nunes F, Wu DK & Fekete DM The expression domain of two related homeobox genes defines a compartment in the chicken inner ear that may be involved in semicircular canal formation. Dev Biol 191, 215–229, doi:10.1006/dbio.1997.8716 (1997). [PubMed: 9398436]

11. Quina LA, Tempest L, Hsu YW, Cox TC & Turner EE Hmx1 is required for the normal development of somatosensory neurons in the geniculate ganglion. Dev Biol 365, 152–163 (2012). [PubMed: 22586713]

12. Takahashi H, Shintani T, Sakuta H. & Noda M. CBF1 controls the retinotectal topographical map along the anteroposterior axis through multiple mechanisms. Development 130, 5203–5215, doi:10.1242/dev.00724 (2003). [PubMed: 12954716]

13. Bayramov AV, Martynova NY, Eroshkin FM, Ermakova GV & Zaraisky AG The homeodomain-containing transcription factor X-nkx-5.1 inhibits expression of the homeobox gene Xanf-1 during the Xenopus laevis forebrain development. Mech Dev 121, 1425–1441, doi:10.1016/j.mod.2004.08.002 (2004). [PubMed: 15511636]

14. Takatori N. et al. Comprehensive survey and classification of homeobox genes in the genome of amphioxus, Branchiostoma floridae. Dev Genes Evol 218, 579–590, doi:10.1007/s00427-008-0245-9 (2008). [PubMed: 18797923]

15. 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). [PubMed: 12736825]

16. Ryan K, Lu Z. & Meinertzhagen IA The CNS connectome of a tadpole larva of Ciona intestinalis (L.) highlights sidedness in the brain of a chordate sibling. Elife 5, doi:10.7554/eLife.16962 (2016).

17. Wang W, Chan EK, Baron S, Van de Water T. & Lufkin T. Hmx2 homeobox gene control of murine vestibular morphogenesis. Development 128, 5017–5029 (2001). [PubMed: 11748138]

18. Wang W, Grimmer JF, Van De Water TR & Lufkin T. Hmx2 and Hmx3 homeobox genes direct development of the murine inner ear and hypothalamus and can be functionally replaced by Drosophila Hmx. Dev Cell 7, 439–453, doi:10.1016/j.devcel.2004.06.016 (2004). [PubMed: 15363417]

19. Wang W, Van De Water T. & Lufkin T. Inner ear and maternal reproductive defects in mice lacking the Hmx3 homeobox gene. Development 125, 621–634 (1998). [PubMed: 9435283]

20. Tang WJ, Chen JS & Zeller RW Transcriptional regulation of the peripheral nervous system in Ciona intestinalis. Dev Biol 378, 183–193, doi:10.1016/j.ydbio.2013.03.016 (2013). [PubMed: 23545329]

21. 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). [PubMed: 30392840]
22. 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). [PubMed: 32656209]
23. Chacha PP et al. Neuronal identities derived by misexpression of the POU IV sensory determinant in a protovertebrate. Proc Natl Acad Sci U S A 119, doi:10.1073/pnas.2118817119 (2022).
24. 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). [PubMed: 29149270]
25. Doglio L. et al. Parallel evolution of chordate cis-regulatory code for development. PLoS Genet 9, e1003904, doi:10.1371/journal.pgen.1003904 (2013). [PubMed: 24282393]
26. McEwen GK et al. Early evolution of conserved regulatory sequences associated with development in vertebrates. PLoS Genet 5, e1000762, doi:10.1371/journal.pgen.1000762 (2009). [PubMed: 20011110]
27. Shimeld SM & Donoghue PC Evolutionary crossroads in developmental biology: cyclostomes (lamprey and hagfish). Development 139, 2091–2099, doi:10.1242/dev.074716 (2012). [PubMed: 22619386]
28. Parker HJ, Bronner ME & 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). [PubMed: 25219855]
29. Scerbo P. & Monsoro-Burq AH The vertebrate-specific VENTX/NANOG gene empowers neural crest with ectomesenchyme potential. Sci Adv 6, eaaz1469, doi:10.1126/sciadv.aaz1469 (2020).
30. Zalc A. et al. Reactivation of the pluripotency program precedes formation of the cranial neural crest. Science 371, doi:10.1126/science.abb4776 (2021).
31. Mazet F. et al. Molecular evidence from Ciona intestinalis for the evolutionary origin of vertebrate sensory placodes. Dev Biol 282, 494–508, doi:10.1016/j.ydbio.2005.02.021 (2005). [PubMed: 15950613]
32. Roure A, Lemaire P. & Darras S. An otx/nodal regulatory signature for posterior neural development in ascidians. PLoS Genet 10, e1004548, doi:10.1371/journal.pgen.1004548 (2014). [PubMed: 25121599]
33. Holland LZ Tunicates. Curr Biol 26, R146–152, doi:10.1016/j.cub.2015.12.024 (2016). [PubMed: 26906481]
34. Love MI, 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). [PubMed: 25516281]
35. 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).
36. Adameyko I. et al. Schwann cell precursors from nerve innervation are a cellular origin of melanocytes in skin. Cell 139, 366–379, doi:10.1016/j.cell.2009.07.049 (2009). [PubMed: 19837037]
37. Adamska M. et al. Inner ear and lateral line expression of a zebrafish Nkx5–1 gene and its downregulation in the ears of FGF8 mutant, ace. Mech Dev 97, 161–165, doi:10.1016/s0925-4773(00)00414-7 (2000). [PubMed: 11025218]
38. Apostolova G. et al. Neurotransmitter phenotype-specific expression changes in developing sympathetic neurons. Mol Cell Neurosci 35, 397–408, doi:10.1016/j.mcn.2007.03.014 (2007). [PubMed: 17513123]
39. Bober E, Baum C, Braun T. & Arnold HH A novel NK-related mouse homeobox gene: expression in central and peripheral nervous structures during embryonic development. Dev Biol 162, 288–303, doi:10.1006/dbio.1994.1086 (1994). [PubMed: 7510254]
40. Boisset G. & Schorderet DF Zebrafish hmx1 promotes retinogenesis. Exp Eye Res 105, 34–42, doi:10.1016/j.exer.2012.10.002 (2012). [PubMed: 23068565]

41. Herbrand H. et al. Two regulatory genes, cNkx5–1 and cPax2, show different responses to local signals during otic placode and vesicle formation in the chick embryo. Development 125, 645–654 (1998). [PubMed: 9435285]
42. Munroe RJ et al. Mouse H6 Homeobox 1 (Hmx1) mutations cause cranial abnormalities and reduced body mass. BMC Dev Biol 9, 27, doi:10.1186/1471-213X-9-27 (2009). [PubMed: 19379485]
43. Quina LA et al. Deletion of a conserved regulatory element required for Hmx1 expression in craniofacial mesenchyme in the dumbo rat: a newly identified cause of congenital ear malformation. Dis Model Mech 5, 812–822, doi:10.1242/dmm.009910 (2012). [PubMed: 22736458]
44. Hartwell RD et al. Anteroposterior patterning of the zebrafish ear through Fgf- and Hh-dependent regulation of hmx3a expression. PLoS Genet 15, e1008051, doi:10.1371/journal.pgen.1008051 (2019). [PubMed: 31022185]
45. Liu J. et al. Evolutionarily conserved regulation of hypocretin neuron specification by Lhx9. Development 142, 1113–1124, doi:10.1242/dev.117424 (2015). [PubMed: 25725064]
46. Lara-Ramirez R, Poncelet G, Patthey C. & Shimeld SM 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). [PubMed: 28871438]
47. Smith JJ 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). [PubMed: 23435085]
48. Smith JJ 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). [PubMed: 29358652]
49. Mehta TK 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). [PubMed: 24043829]
50. 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). [PubMed: 17760977]
51. Edgar RC MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32, 1792–1797, doi:10.1093/nar/gkh340 (2004). [PubMed: 15034147]
52. 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). [PubMed: 24451623]
53. 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). [PubMed: 30464347]
54. Lara-Ramirez R, Patthey C. & Shimeld SM 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). [PubMed: 25809594]
55. Boorman CJ & Shimeld SM 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). [PubMed: 12356265]
56. 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). [PubMed: 17520703]
57. 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). [PubMed: 15287102]
58. Holland PWH Wholemount in situ hybridization to amphioxus embryos. Methods Mol Biol 97, 641–644, doi:10.1385/1-59259-270-8:641 (1999). [PubMed: 10443398]
59. Parker HJ, 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). [PubMed: 24416417]

60. Schindelin J. et al. Fiji: an open-source platform for biological-image analysis. Nat Methods 9, 676–682, doi:10.1038/nmeth.2019 (2012). [PubMed: 22743772]

61. Corbo JC, Levine M. & Zeller RW Characterization of a notochord-specific enhancer from the Brachyury promoter region of the ascidian, Ciona intestinalis. Development 124, 589–602 (1997). [PubMed: 9043074]

62. Nakamura MJ, 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). [PubMed: 23022659]

63. Bolger AM, Lohse M. & Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120, doi:10.1093/bioinformatics/btu170 (2014). [PubMed: 24695404]

64. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33). (https://github.com/najoshi/sickle, 2011).

65. Dobin A. et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21, doi:10.1093/bioinformatics/bts635 (2013). [PubMed: 23104886]

66. Harafuji N, Keys DN & 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). [PubMed: 12011440]

67. Chen WC 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). [PubMed: 24680932]

68. 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).

69. 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). [PubMed: 25336740]

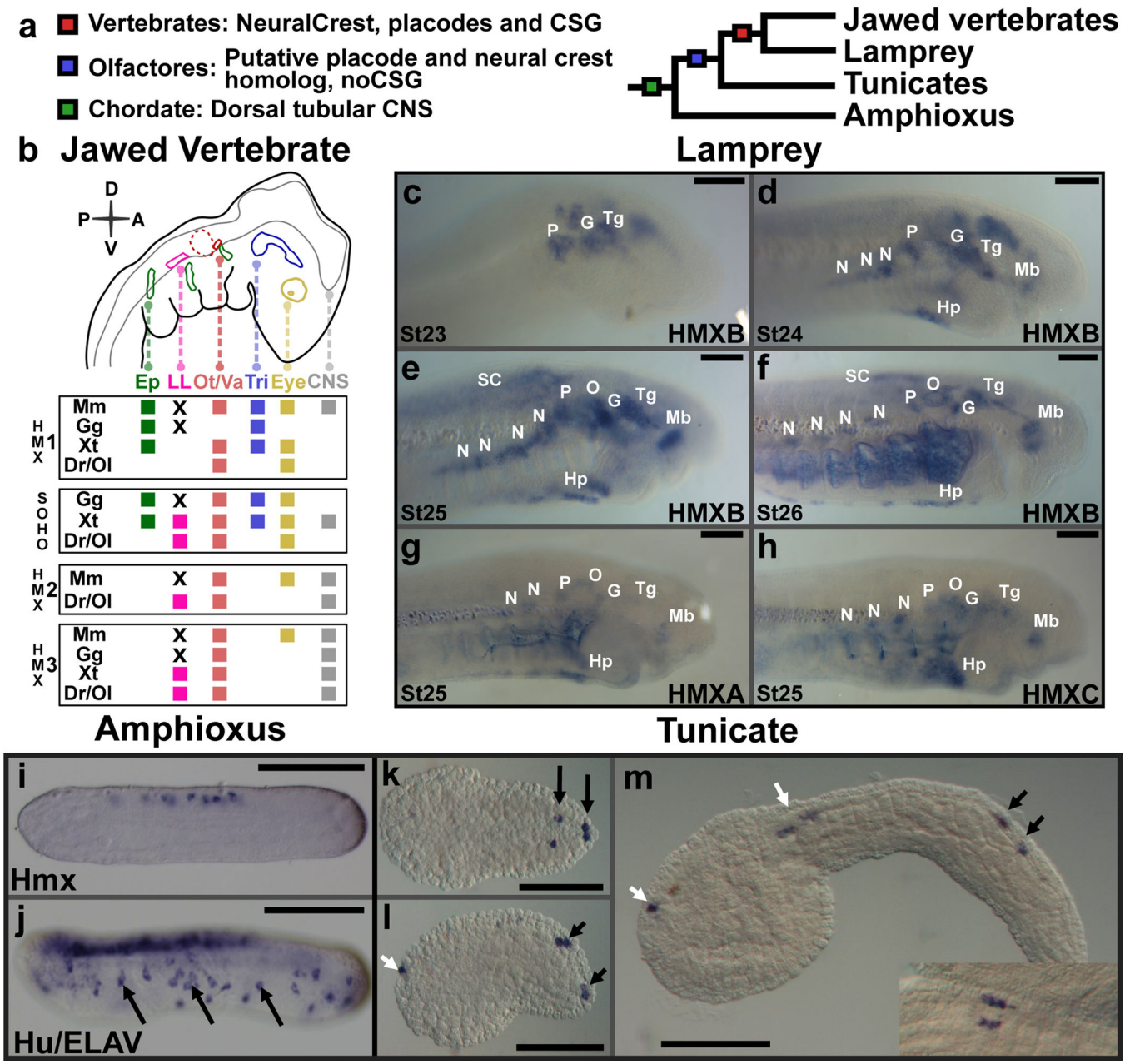


Figure 1. *Hmx* expression in chordates.

a. Phylogeny of the chordates showing the evolutionary ancestry of key characters. **b.** Schematic depiction of *Hmx* expression in jawed vertebrate cranial placodes/ganglia. Species shown are: Mm, *Mus musculus*. Gg, *Gallus gallus*. Xt-*Xenopus tropicalis*. Dr, *Danio rerio*. Ol, *Oryzias latipes.* Where species are missing it means either the gene has not been analysed (*Hmx2* in Gg and Xt) or the gene has been lost (*SOHo* in Mm). Other abbreviations: Ep, epibranchial. LL, lateral line. Ot/Va, otic/vestibuloacoustic. Tri, trigeminal. CNS, central nervous system. X indicates lateral line ganglia have been lost by these species. The olfactory is not shown as *Hmx* expression has not been reported from this placode. Full analysis underlying this figure in Extended Data Fig. 1. **c-h.** Expression

of lamprey *Hmx* genes. **c.** *HmxB* expression starts in cranial sensory ganglia at stage 23. **d.** Hypothalamus expression in also seen by stage 24. An expression domain below the branchial basket corresponds to the position of the hypobranchial ganglia. **e, f.** Hindbrain and spinal cord express *HmxB* at stages 25 and 26: branchial arch stain in (**f**) is an artefact of antibody trapping. **g, h**. Expression of lamprey *HmxA* and *HmxC* is identical to *HmxB*. Abbreviations: N, nodose. P, petrosal. G, geniculate. O, otic. Tg, trigeminal. SC, spinal cord. Hp, hypobranchial. Hy, hypothalamus. **i, j.** *Hmx* expression in amphioxus (**i**) compared to the neural marker *Hu/ELAV* (**j**). Arrows mark some of the peripheral neurons expressing *Hu/ELAV*. *Hmx* expression is confined to the CNS. **k-m.** *Hmx* expression In *Ciona*. Black arrows identify BTNs, white arrows expression in the CNS. The inset on (**m**) shows a dorsal view with BTNs lying parallel to the CNS. All scale bars 100μM. Lamprey images representative of at least 5 embryos for each stage, amphioxus and *Ciona* images representative of at least 10 embryos.

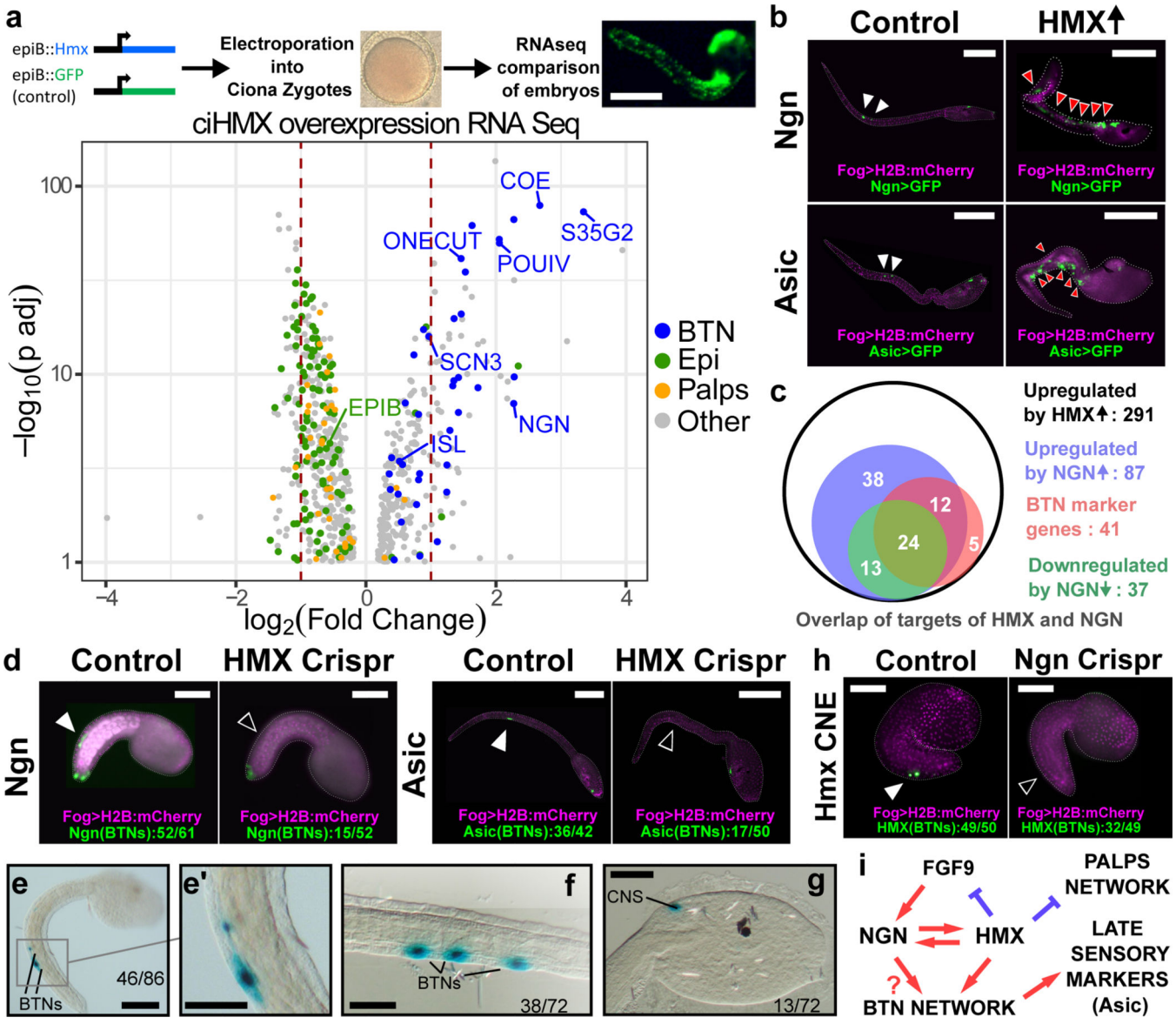


Figure 2. *Hmx* regulation and downstream target genes in *Ciona*.

a. Above is a schematic of the strategy used to drive *Ciona Hmx* overexpression. Below the volcano plot shows genes up or down regulated after *Ciona Hmx* overexpression. Selected genes are named. Colour coding reflects genes identified as cell type expressed in single cell sequencing data[2], according to the code shown. Data analysed using negative binomial generalized linear models the DESeq2[34], P values adjusted (p adj) for multiple testing. Table 1 shows precise numbers, underlying data in Supplementary File 1.**b**. Expression of *Ngn>GFP* and *Asic>GFP* constructs in control embryos (white arrowheads) and in cells ectopically expressing *Hmx* driven by *epiB>Hmx* (red arrowheads). **c.** Overlap of upregulated *Hmx* target genes and genes differentially expressed after *Ngn* overexpression and knockdown[22]. Subset of genes upregulated by *Hmx* overexpression that were also upregulated by *Ngn* overexpression or downregulated by *Ngn* knockdown and BTN expressed genes from single cell data[2] are shown. Full data in Supplementary File 2. **d**. BTN

marker activity after *Hmx* CRISPR Cas9 knockout. *Fog>H2B:mCherry* marks successful uptake of the electroporation mix. BTN signal for both early (*Ngn*) and late (*Asic*) markers (closed arrowheads) was lost after *Hmx* knockout (open arrowheads). Numbers in Extended Data Table 1. **e-g**. *Ciona Hmx* CNE activity in *Ciona* embryos, visualised by lacZ staining. (**e**) shows a tailbud stage embryo with stain in BTNs, seen in close up in (**e'**). (**f**,**g**) show BTN and CNS staining respectively in early larvae. Numbers indicate number of embryos showing staining in the indicated structures, out of total surviving embryos. Full embryo counts in Extended Data Table 3. **h**. *Ciona Hmx* CNE activity in controls (closed arrowhead) and its loss after *Ngn* CRISPR Cas9 knockout (open arrowhead). *Fog>H2B:mCherry* marks successful uptake of the electroporation mix. Numbers in Extended Data Table 2. **i**. BTN specification network model. Scale bars 100μM except **e-g** (50μM).

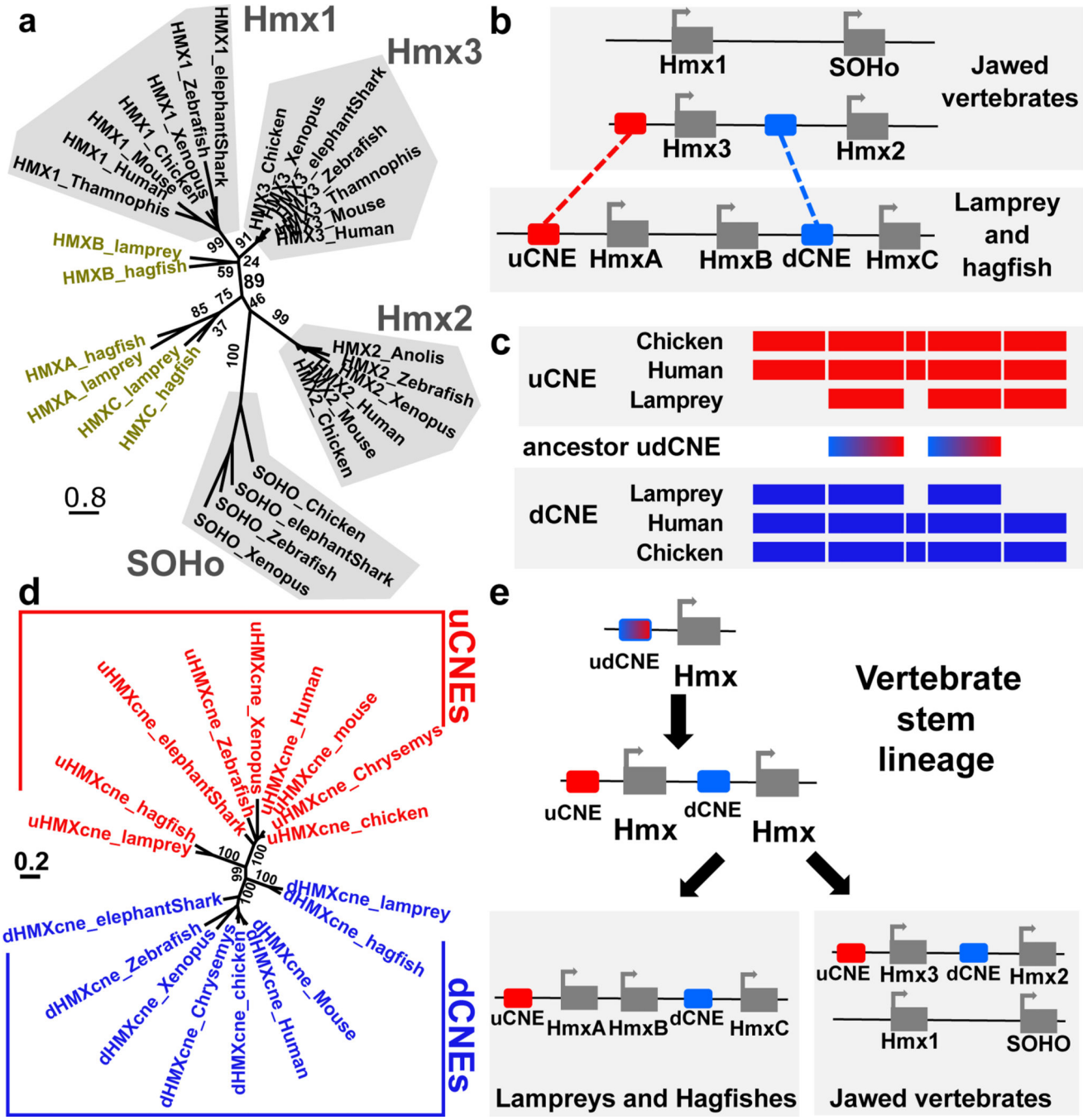


Figure 3. Vertebrate *Hmx* locus evolution and CNE identification

a. Phylogeny of vertebrate Hmx proteins. Tree constructed using the Maximum Likelihood method, values indicate percentage bootstrap support and scale indicates number of substitutions per site. **b.** Comparative mapping of jawed and jawless vertebrate *Hmx* loci identifies two CNEs. **c.** Sequence similarity between uCNE and dCNE sequences in lamprey, human and chicken show they derived from an ancestral pre-duplication CNE. Coloured boxes indicate regions of sequence similarity amongst uCNE sequences (red), dCNE sequences (blue) and between uCNE and dCNE sequences (dashed red and

blue). Sequence alignments in Supplementary Files 3–5. **d.** uCNE and dCNE sequences form monophyletic groups in molecular phylogenetic analysis. Tree constructed using the Maximum Likelihood method, values indicate percentage bootstrap support and scale indicates number of substitutions per site. **e.** Model for the evolution of *Hmx* loci. The current arrangements in vertebrates evolved from a single ancestral cluster with both uCNE and dCNE elements present, which itself evolved from a one-gene state with a single ancestral udCNE.

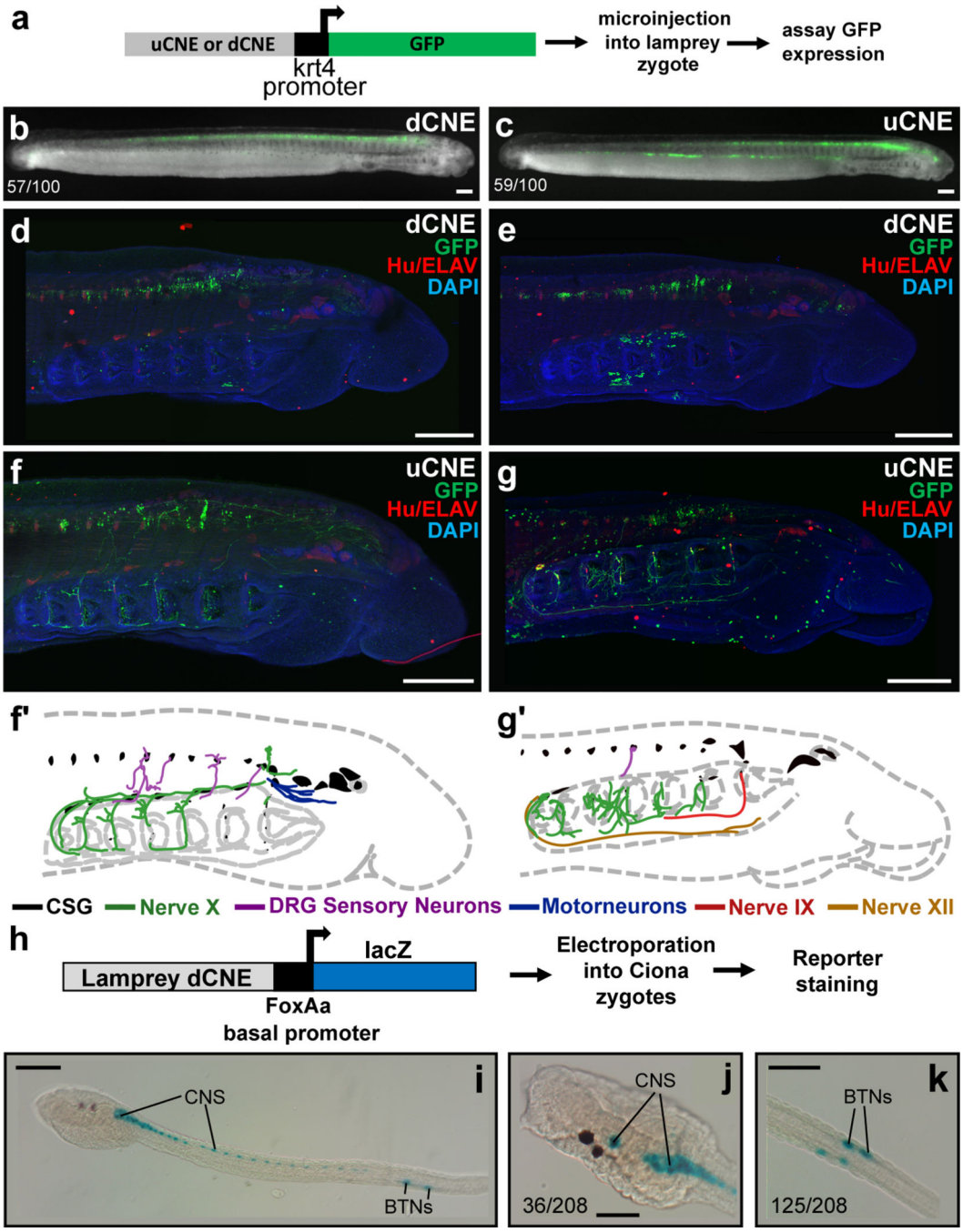


Figure 4. Lamprey *Hmx* CNE activity in transgenic lamprey and *Ciona* embryos

a. Experimental strategy for detecting reporter activity in lamprey embryos. **b, c.** Representative embryos showing *dCNE>GFP* and *uCNE>GFP* activity in the CNS. Numbers show the number of times CNS expression was seen out of the number of embryos screened. Analysis of vector-only controls is shown in Extended Data Fig. 7. **d, e.** Confocal reconstructions of lamprey embryos transgenic for *dCNE>GFP*, showing activity (green) in forebrain, midbrain, hindbrain and spinal cord. **f, g.** Confocal reconstructions of lamprey embryos transgenic for *uCNE>GFP*, showing activity in the CNS, and in peripheral nerves.

In (**d-g**) ganglia and other neurons are stained red using an antibody to Hu/ELAV. **f', g'** Schematic tracing of embryos shown in (**f**) and (**g**) respectively, with nerve CNE reporter activity traced and colour coded. **h.** Schematic of the experimental method used to examine lamprey CNE activity in *Ciona*. **i-k**. Transgenic *Ciona* larvae stained for lamprey dCNE reporter activity. Numbers indicate the number of times expression in the cells identified was seen, out of total surviving larvae. Full embryo counts in Extended Data Table 3. All scale bars 100μM.

Table 1.

Differentially expressed genes from Fig. 2a, quantified for BTN, palp and epidermal cell types.

Cell Type	Number upregulated	Number downregulated	Total significantly different
BTN	38	0	38
Palp	4	31	35
Epidermal	6	100	106
Other	224	283	507
Total	272	414	686