Hmx gene conservation identifies the evolutionary origin of vertebrate cranial ganglia

Vasileios Papdogiannis[1], Hugo J. Parker[2], Alessandro Pennati[1], Cedric Patthey[1,3], Marianne E. Bronner[4] and Sebastian M. Shimeld[1]*

[1] Department of Zoology, University of Oxford, 11a Mansfield Road, Oxford OX1 3SZ.
[2] Stowers Institute for Medical Research, Kansas City, MO 64110, USA
[3] Present address: Department of Radiosciences, Umeå University, 901 85 Umeå, Sweden.
[4] Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125, USA.

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

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) which develop 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 able 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 confirm BTNs as CSG homologs. Our analysis also identifies derived evolutionary changes, including a genetic basis for secondary simplicity in *Ciona* and unique regulatory complexity in vertebrates.

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 and some neurons of the trigeminal ganglion. 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 encodes homeodomain transcription factors. 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), as well as 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. To gain insight into roles of Hmx genes in invertebrate chordates, 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 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, and delaminate and migrate before connecting epidermal sensory cells to the CNS[3,16]. Previously, they have been likened to neural crest[2,3].

Vertebrate Hmx genes are necessary for the correct development of CSG[11,17-19]. To test whether *Ciona Hmx* functions in BTN development, we used the *Ciona epiB* promoter[20] to drive *Hmx* expression broadly in the embryonic epidermis, with transgenic embryos subsequently prepared for transcriptome analysis. Comparison of experimental to control embryos (Fig. 2A) revealed up- and down-regulated genes that were compared to cell-type expression profiles extracted from single-cell sequence data[2]. The results showed that *Ciona Hmx* upregulated the BTN differentiation program (Fig. 2B), concomitantly suppressing expression of epidermal genes, and of genes associated with other cells that develop from this lineage (Fig. 2B). This suggests that *Hmx* is sufficient to drive the BTN transcriptional program in *Ciona* embryonic epidermis.

The expression domains and developmental roles of Hmx genes in vertebrates and *Ciona* may reflect either shared ancestry, or derived traits in either lineage. Discriminating between these possibilities requires comparison of the genetic program(s) upstream of Hmx expression in each lineage. To address this we mapped and tested Hmx regulatory elements in *Ciona* and lamprey. Comparison of tunicate genomes[21] showed 2kb of sequence 5' to the *Hmx* transcription start site to be conserved between *Ciona* species (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* (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).

Hmx gene evolution in vertebrates is more complex. Jawed vertebrate Hmx genes are located in two paralagous 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 ancestor a single Hmx gene duplicated in tandem to yield a two-gene cluster. In jawless vertebrates a second tandem duplication occurred, yielding the three-gene cluster state found in lamprey and hagfish. In jawed vertebrates the two-gene cluster duplicated as a block to form the paralagous two-gene clusters *Hmx3-Hmx2* and *Hmx1-SOHo* (Fig. 3E). Analysis of the broader chromosomal regions in which these genes lie has shown this occurred as part of the two rounds of genome duplication (2R) that occurred in early vertebrate evolution[22].

Comparative genomics across all jawed vertebrates 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 (Figure S1), exceptionally large for ancient vertebrate CNEs[23,24]. Unusually, they are also homologous to each other, sharing a core of around 500bp of highly conserved sequence between paralogous CNEs (Fig. 3C. S1). We searched for these CNEs in lamprey and hagfish, identifying one 5' of *HmxA*, and a second between *HmxB* and *HmxC* (Fig. 3B; Fig. S1). Molecular phylogenetic analysis confirmed their orthology to jawed vertebrate uCNE and dCNE respectively (Fig. 3D). These data show the uCNE and dCNE originally evolved by tandem duplication of one ancestral CNE (udCNE: Fig. 3C, E). Parsimoniously this happened at the same time as the ancestral Hmx gene was duplicated, predating the separation of jawless and jawed vertebrate lineages over 500 million years ago[25].

To test the functions of uCNE and dCNE we generated lamprey embryos transgenic for reporter constructs (Fig. 4A)[26]. 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 infer this ancestral element would also have had the capacity to drive gene expression into CNS and CSG (Fig. 4H).

The 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, when we tested the activity of lamprey CNEs in transgenic *Ciona* (Fig. 4H) we found dCNE was able to recapitulate *Ciona Hmx* expression in BTNs and CNS (Fig. 4I-L). BTN expression was specific and robust (Fig. 4J, L). CNS expression extended along the majority of the neural tube (Fig. 4J), 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-L with Fig. 4B, C and Fig. 1I-M).

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 (AP) patterning network into a posterior 'proto-neural crest' domain and anterior 'proto-placode' domain[2]. However, Hmx constitutively marks vertebrate placodes and CSG, and not neural crest: while Hmx expression in DRG and ciliary ganglia (both neural crest derivatives) has been reported in the mouse [7,27,28], this is not seen in other vertebrates including *Xenopus*, zebrafish, medaka and lamprey[6,8,9,13] (Supplementary Figure S4). Our expression data point to BTNs as homologous to placodal CSG neurons and not to neural crest. This is reinforced by finding that a vertebrate Hmx CNE robustly drives expression of a reporter in *Ciona* BTNs and Lamprey CSG neurons, implying the two cell types share similar *trans*-regulatory environments. 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[2,29]. Both, however, are homologous to placode-derived CSG, and not neural crest.

3

We also note evolution of Hmx in tunicates and vertebrates parallels derived aspects of neural evolution and includes highly unusual features. Tunicates are secondarily-reduced[30], and we see direct 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 in the stem vertebrate. That both CNEs have been subsequently conserved and maintained in tandem over the remainder of vertebrate evolution appears to be unique. These Hmx CNEs are also unusually large, with other identified jawed vertebrate CNEs conserved with lamprey typically much smaller [24]. While the functional implications of these unusual features are unknown, such conservation speaks to extreme evolutionary constraint. We speculate this reflects 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 the instructive role of Hmx in directing the development of an ancient cell type involved in environmental sensing.

References

1   Northcutt, R. G. & Gans, C. The genesis of neural crest and epidermal placodes: a reinterpretation of vertebrate origins. *Q Rev Biol* **58**, 1-28 (1983).

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).

3   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).

4   Shimeld, S. M. & Holland, P. W. Vertebrate innovations. *Proc Natl Acad Sci U S A* **97**, 4449-4452, 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).

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).

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).

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).

9   Kelly, L. E. & El-Hodiri, H. M. 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).

10   Kiernan, A. E., Nunes, F., Wu, D. K. & Fekete, D. M. 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).

11   Quina, L. A., Tempest, L., Hsu, Y. W., Cox, T. C. & Turner, E. E. Hmx1 is required for the normal development of somatosensory neurons in the geniculate ganglion. *Dev Biol* **365**, 152-163 (2012).

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).

13   Bayramov, A. V., Martynova, N. Y., Eroshkin, F. M., Ermakova, G. V. & Zaraisky, A. G. 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).

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).

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).

16   Ryan, K., Lu, Z. & Meinertzhagen, I. A. 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, E. K., Baron, S., Van de Water, T. & Lufkin, T. Hmx2 homeobox gene control of murine vestibular morphogenesis. *Development* **128**, 5017-5029 (2001).

18   Wang, W., Grimmer, J. F., Van De Water, T. R. & 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).

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).

20   Joyce Tang, W., 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   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).

22   Wotton, K. R. *et al.* Conservation of gene linkage in dispersed vertebrate NK homeobox clusters. *Dev Genes Evol* **219**, 481-496, doi:10.1007/s00427-009-0311-y (2009).

23   Doglio, L. *et al.* Parallel evolution of chordate cis-regulatory code for development. *PLoS Genet* **9**, e1003904, doi:10.1371/journal.pgen.1003904 (2013).

24   McEwen, G. K. *et al.* Early evolution of conserved regulatory sequences associated with development in vertebrates. *PLoS Genet* **5**, e1000762, doi:10.1371/journal.pgen.1000762 (2009).

25   Shimeld, S. M. & Donoghue, P. C. Evolutionary crossroads in developmental biology: cyclostomes (lamprey and hagfish). *Development* **139**, 2091-2099, doi:10.1242/dev.074716 (2012).

26   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).

27   Bober, E., Baum, C., Braun, T. & Arnold, H. H. 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).

28   Munroe, R. J. *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).

29   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).

30   Holland, L. Z. Tunicates. *Curr Biol* **26**, R146-152, doi:10.1016/j.cub.2015.12.024 (2016).

Figures and legends

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. Tg, trigeminal. CNS, central nervous system. X indicates lateral line ganglia have been lost by these species. The olfactory is not shown as Hmx has not been reported from this placode. Full analysis underlying this figure in Supplementary Figure S4.

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. **F.** Hindbrain and spinal cord are strongly marked after stage 25. **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 compared to the neural marker *Hu/ELAV*. Arrows mark some of the peripheral neurons expressing *Hu/ELAV*.

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.
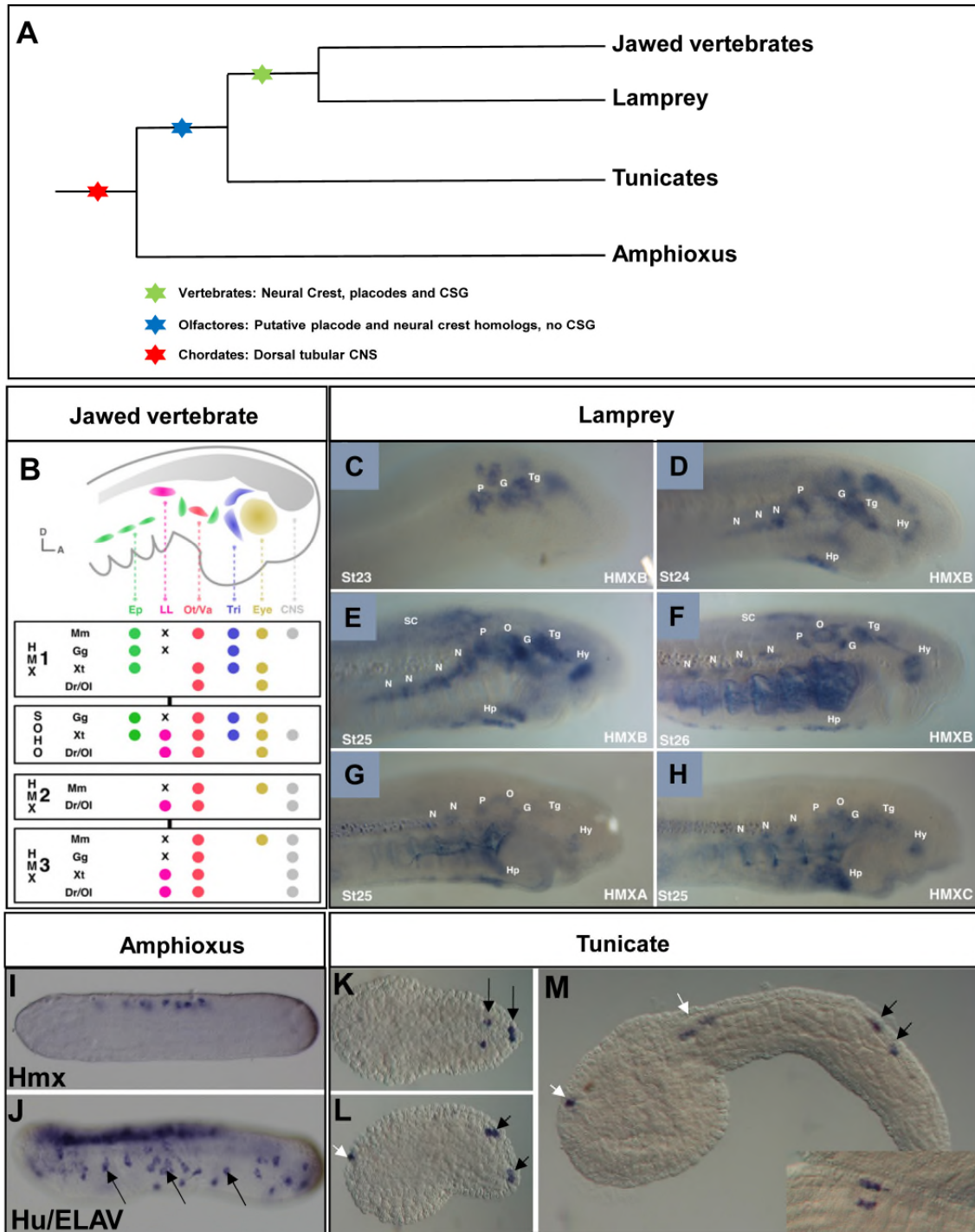
6

Figure 1

7

Figure 2. *Hmx* function and regulation in *Ciona*

A. Schematic of the experimental strategy used to drive *Ciona Hmx* overexpression.

B. Volcano plot of 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], and the inset table shows these quantified for BTN, palp and epidermal (EPI) cell types.

C. Sequence conservation between *C. robusta* and *Ciona savignyi* identifies an approximately 2Kb CNE 5' to the first *Hmx* exon.

D. Schematic of the experimental strategy for characterising *Hmx* CNE activity in *Ciona.*

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'. No CNS stain was observed at this stage. F and G show BTN and CNS staining respectively in early larvae. Numbers indicate the number of embryos showing staining in the indicated structures, out of the total number of surviving embryos. Full embryo counts are in Supplementary Figure S5.
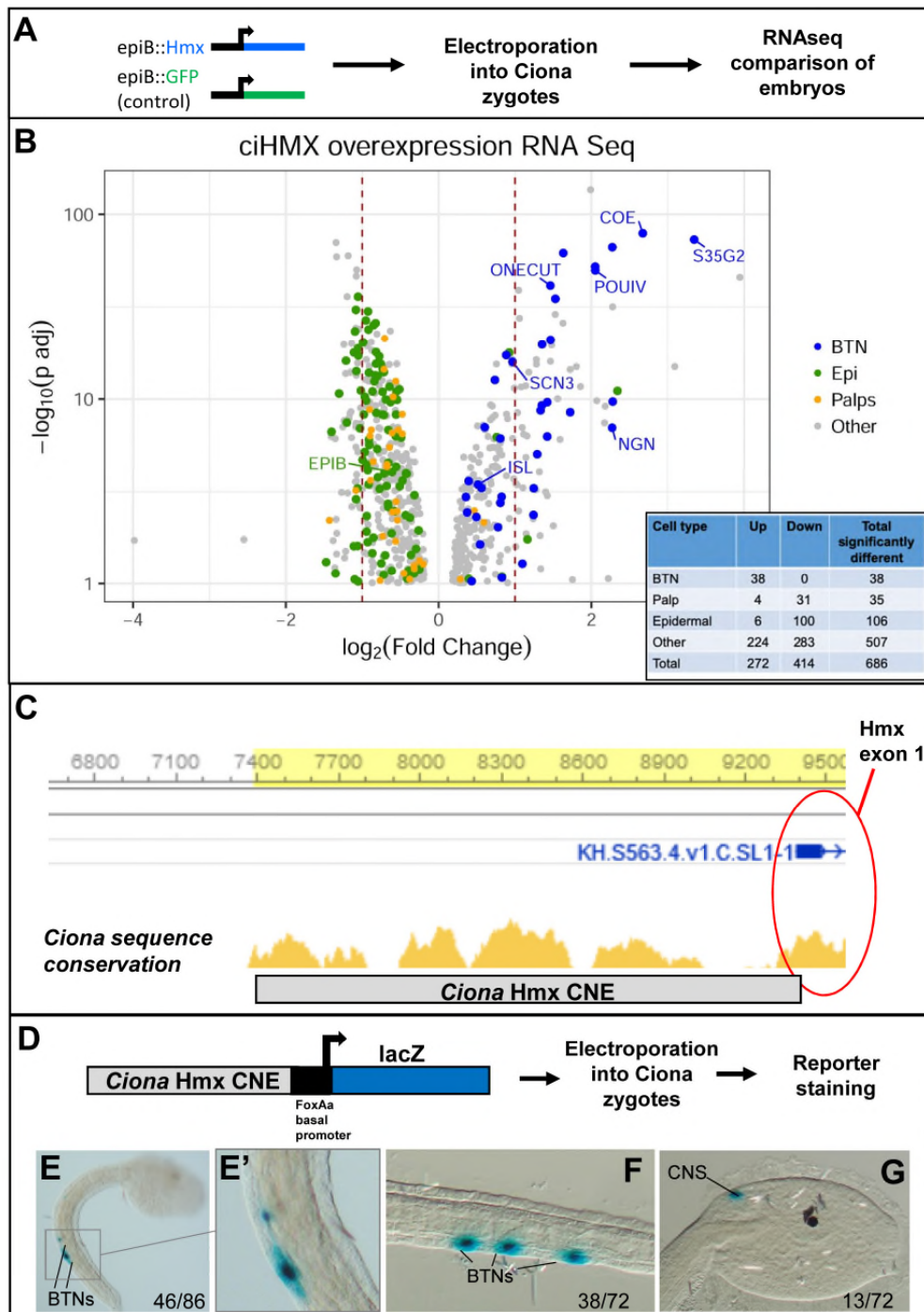
8

Figure 2

Figure 3. Vertebrate Hmx locus evolution and CNE identification

A. Phylogeny of vertebrate HMX proteins. Values indicate percentage bootstrap support.

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 Figure 1.

D. uCNE and dCNE sequences form monophyletic groups in molecular phylogenetic analysis.

E. A 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.
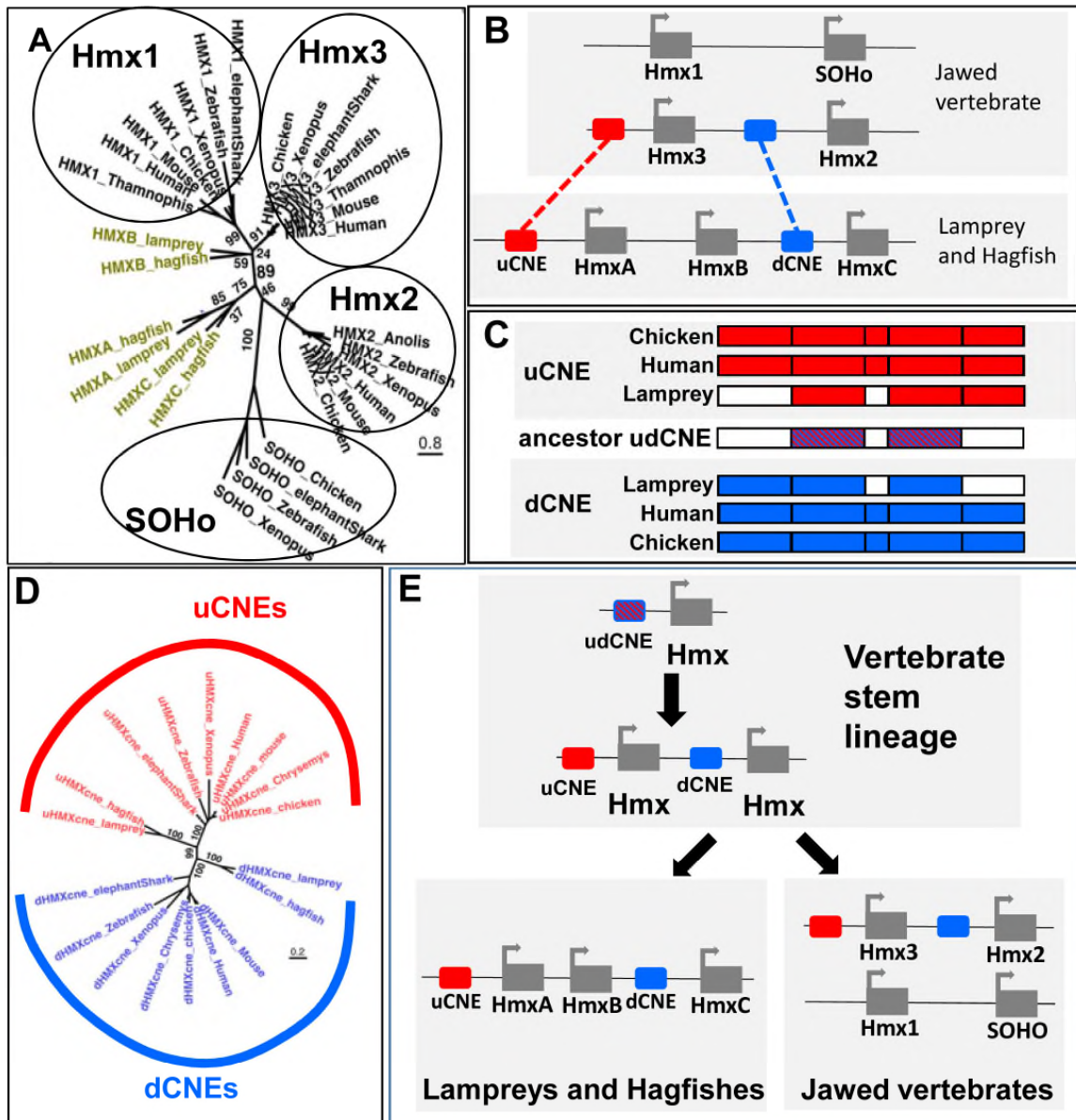
10

Figure 3

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 Supplementary Figure S3.

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. Summary of lamprey Hmx CNE activity.

I-L. I shows a schematic of the experimental method used to examine lamprey CNE activity in *Ciona*. Below are transgenic 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 are in Supplementary Figure S6.

Figure 4
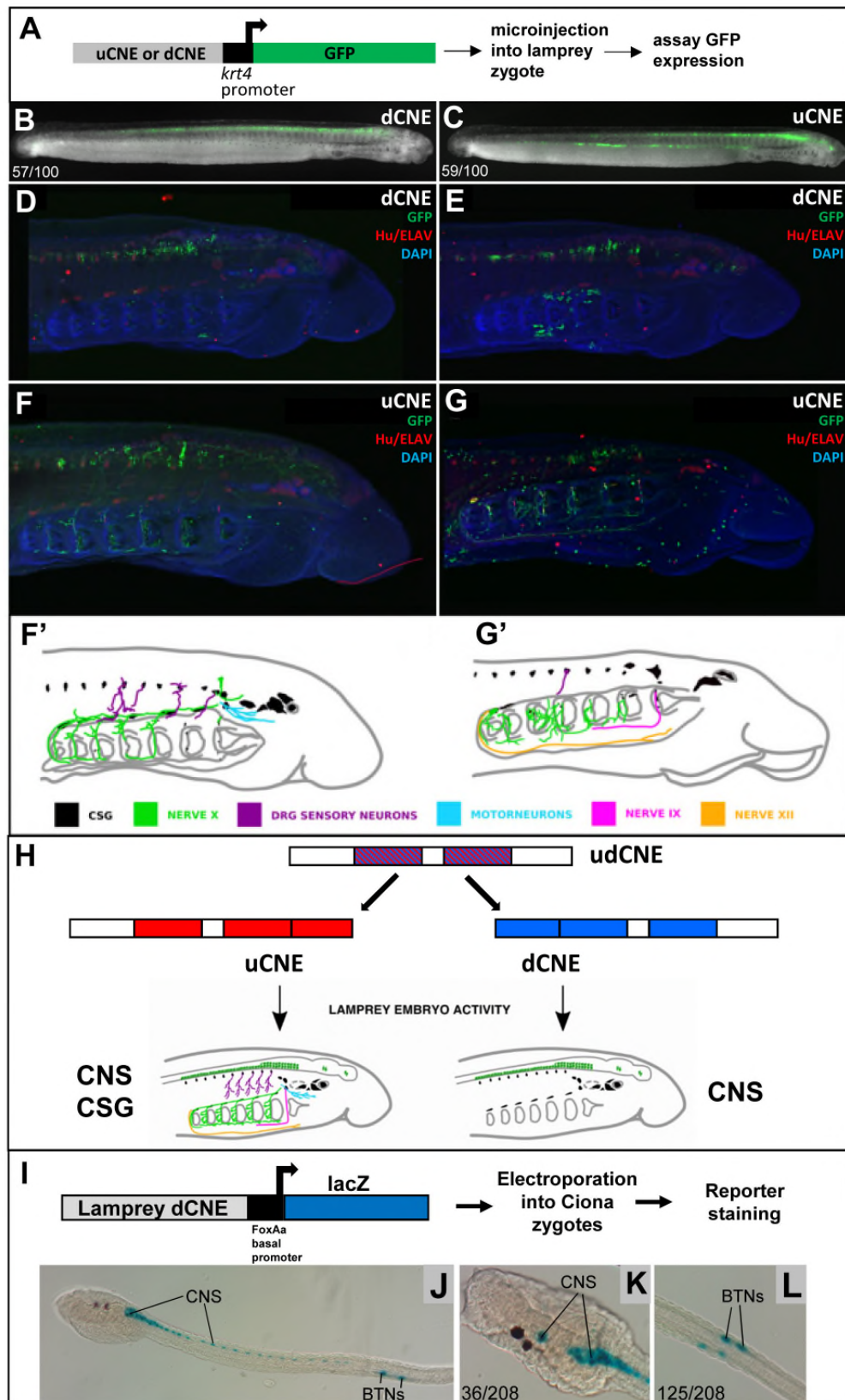
Materials and Methods

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 [1], genome assemblies for *Petromyzon marinus* [2,3], a *P. marinus* transcriptome assembly built in-house from Illumina GAII data available on SRA, the *Lethenteron camtschaticum* genome assembly [4], 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 (Figure S3, supplementary data). 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 [5]. 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. Alignments and molecular phylogenetic analyses were undertaken using MUSCLE [6] and RAxML [7] 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 Hmx/Nkx5* locus was already annotated [8], 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 [9], to identify the full open reading frame. This was amplified, using the primers shown below, for the overexpression experiment. *Branchiostoma lanceolatum Hmx* was identified by searching the genome [10]. 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
L. planeri uCNE	GCAGCAAGCGATGTATGCAA	CGAGTACGCGAAGAGGTGG
L. planeri dCNE	CATCATCAGCGCGCGTATTT	TTCCTGCACACAACTCTCGC
C. intestinalis Hmx probe	AGATCTGGTTTCAGAACCGACG	GGGATTCCCCTCTTTGCGTA
Ciona CNE	CACTATAGAACTCGACGTCCAACTAAATCGTCGAAG	CTGACCATGGGGATCCTTTTCGGCGGTTCGTTCTC
Lamprey uCNE into pCES	CACTATAGAACTCGAGCAGCAAGCGATGTATGC	GAGTCGACCTGCAGGCGAGTACGCGAAGAGGTGG
Lamprey dCNE into pCES	CACTATAGAACTCGACATCATCAGCGCGCGTATTTAAG	GAGTCGACCTGCAGGTTCCTGCACACAACTCTCGC
B. lanceolatum Hmx	CTCCAACAACCCGTTCTCGA	TGGGATAAGGGCGGGTTTTC
B. lanceolatum Hu/Elav	GGTGGAGTCCTGCAAGCTGA	GAATGGGCCAAAGAGCTGCC

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 [11]. Adult *Ciona intestinalis* were collected from Northney Marina, UK, and maintained in a circulating sea water aquarium at 14°C under constant light. 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 [12]. Adult *Branchiostoma lanceolatum* were collected near Banyuls-sur-Mer, France and spawning was induced by heat stimulation [13,14]. 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 [15].

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 [16]. Lamprey transient transgenesis was performed in *P. marinus* embryos as previously described [16,17]. 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 [18]. 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 (Figure S2), 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 ganglia. 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: Figure S2, supplementary data) and did not label the same cells as seen with Hmx uCNE.

Ciona transgenics and sequence analysis

The plasmid vector containing the *epiB* promoter driving *GFP* (*epiB::GFP*) was kindly provided by Bob Zeller [19]. 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* 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 C*. intestinalis* zygotes as previously described [20]. 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. Embryos were grown to the tailbud stage when GFP was visible, identifying transgenic embryos which were then manually selected and processed for RNA extraction. Three full biological replicates were performed on embryo batches derived from different fertilisations, with each biological replicate combining RNA from at least 50 individual 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 [21] to trim off adapters and Sickle [22] to trim low quality reads. Remaining reads were then mapped to the *Ciona robusta* genome (KH2012) using STAR [23]. Differential expression analysis was carried out using the DESeq2 R package [24], 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 *C. intestinalis* single cell sequencing were extracted from the supplementary files of published literature [25], and cross-correlated with the up and down regulated gene lists to provide the annotation of data shown in Figure 2B.

3

To test CNE activity in *Ciona*, we cloned the 2kb 5' to *C. intestinalis Hmx*, lamprey uCNE or lamprey dCNE into the reporter vector pCES [26] using the primers shown in Table 1. Constructs were electroporated into *C. intestinalis* zygotes as above, and embryos stained for β-galactosidase activity as described [27].

References for materials and methods

1   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).

2   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).

3   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).

4   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).

5   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).

6   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).

7   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).

8   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).

9   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).

10   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).

11   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).

12   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).

13   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).

14   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).

15   Holland, P. W. Wholemount in situ hybridization to amphioxus embryos. *Methods Mol Biol* **97**, 641-644, doi:10.1385/1-59259-270-8:641 (1999).

16   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).

17   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).

18   Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**, 676-682, doi:10.1038/nmeth.2019 (2012).

19   Joyce Tang, W., 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).

20   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).

21   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).

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

23   Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).

24   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).

25   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).

26   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).

27   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).