A Hox-TALE regulatory circuit for neural crest patterning is conserved across vertebrates

Hugo J. Parker1, Bony De Kumar2, Stephen A. Green2, Karin D. Prummel3, Christopher Hess3, Charles K. Kaufman4, Christian Mosimann3, Leanne M. Wiedemann1,5, Marianne E. Bronner2 & Robb Krumlauf1,6

In jawed vertebrates (gnathostomes), Hox genes play an important role in patterning head and jaw formation, but mechanisms coupling Hox genes to neural crest (NC) are unknown. Here we use cross-species regulatory comparisons between gnathostomes and lamprey, a jawless extant vertebrate, to investigate conserved ancestral mechanisms regulating Hox2 genes in NC. Gnathostome Hoxa2 and Hoxb2 NC enhancers mediate equivalent NC expression in lamprey and gnathostomes, revealing ancient conservation of Hox upstream regulatory components in NC. In characterizing a lamprey hoxα2 NC/hindbrain enhancer, we identify essential Meis, Pbx, and Hox binding sites that are functionally conserved within Hoxa2/Hoxb2 NC enhancers. This suggests that the lamprey hoxα2 enhancer retains ancestral activity and that Hoxa2/Hoxb2 NC enhancers are ancient paralogues, which diverged in hindbrain and NC activities. This identifies an ancestral mechanism for Hox2 NC regulation involving a Hox-TALE regulatory circuit, potentiated by inputs from Meis and Pbx proteins and Hox auto-/cross-regulatory interactions.
Neural crest (NC) cells are a migratory and multipotent cell type, representing a defining characteristic of vertebrates\(^1\). The cranial NC emerges from the mid-hindbrain region and contributes to cartilage and bone of the pharynx. Gnathostomes (jawed vertebrates) exhibit nested Hox gene expression domains (Hox codes) along the anteroposterior (AP) extent of both the neural tube and adjacent NC. However, little is known regarding shared versus independent regulation of Hox expression in these two tissues. In NC of the pharyngeal arches (PAs), the NC Hox code confers positional identity to each arch\(^3\). There is also evidence for an earlier role of Hox genes in regional generation of NC from the hindbrain\(^4\,\(^5\). Emergence of NC and its underlying Hox code played an important role in craniofacial evolution, giving rise to unique structures of the head and neck, including the jaws and hyoid apparatus used in feeding and respiration. Genome analyses together with gene expression and functional studies have described a common framework for a gene regulatory network (GRN) that drives NC induction, specification, and migration across vertebrates, and some regulatory components are present in non-vertebrate chordates\(^6\,\(^7\). However, Hox genes have not yet been integrated within the current formulation of the NC GRN. This is in part because the mechanisms regulating Hox expression in NC are relatively unclear compared to the current knowledge of Hox regulation in hindbrain segmentation\(^4\). Hence, despite the established functional roles of Hox genes in cranial NC\(^8\,\(^9\), whether the Hox code is coupled to this conserved NC GRN and if so, how, is unknown. It is unclear whether Hox networks that control axial patterning and generation of NC are integrated within or are working in parallel, independently from the NC GRN.

Hox2 genes are critical components of the cranial NC Hox code and provide an important avenue for addressing this question. Experimental alterations in expression of gnathostome Hox2 paralogues (Hoxa2 and Hoxb2) lead to homeotic transformations. In mice, ectopic Hoxa2 suppresses jaw formation\(^10\), while Hoxa2 mutants exhibit mirror image jaw duplication, indicating a role as a selector gene in NC AP identity\(^11\,\(^12\). Regulatory analyses in mouse, chick, and zebrafish have identified evolutionarily conserved enhancers in the Hoxa2 and the Hoxb2 genes that mediate their expression in NC and hindbrain segments (rhombomeres (r)) (Fig. 1a, b)\(^10\,\(^11\,\(^13\). Analyses of mouse Hoxa2 and Hoxb2 NC enhancers provide conflicting mechanisms for NC expression of Hox genes. A single 5′-flanking enhancer drives Hoxb2 in r4 and its NC\(^11\), supporting a model for shared regulation in these tissues (Fig. 1b). In contrast, independent exonic/intronic and 5′-flanking enhancers regulate Hoxb2 expression in r4\(^12\) versus r4-derived NC\(^14\) (PA2), suggesting the evolution of distinct regulatory mechanisms between the hindbrain and NC (Fig. 1a). Characterisation of cis-elements required for the activity of these enhancers has provided some insight into their underlying regulatory mechanisms (Fig. 1a, b). Both the Hoxb2 r4/NC and the Hoxa2 exonic/intronic r4 enhancers depend upon the combined inputs from Meis and Pbx-Hox binding sites for their activities\(^11\,\(^15\,\(^16\). However, the inputs into the independent Hoxa2 NC enhancer are largely unknown, except for a binding site for transcription factor AP2-α (Tfap2a) in the mouse enhancer that is not conserved across gnathostome species\(^10\,\(^12\). Taken together, these analyses reveal differences between Hoxa2 and Hoxb2 in the enhancers and regulatory mechanisms underlying their expression in r4 and NC. Hence, in mouse, each gene supports a different model with respect to shared versus independent regulation in the hindbrain and NC. Given that Hoxa2 and Hoxb2 are paralogous genes, this raises two questions: what is the evolutionary relationship between their NC enhancers and which of these mechanisms is ancestral?

The sea lamprey offers an opportunity to address conserved ancestral mechanisms of Hox2 NC regulation in vertebrates. The extant jawless vertebrates (cyclostomes), lamprey and hagfish, represent a sister group to gnathostomes, so comparisons between these lineages can provide insights into the evolution of gene regulatory programs in vertebrates\(^18\,\(^19\,\(^20\). Such comparisons can reveal aspects of vertebrate biology that were present in the common ancestor of extant vertebrates and which have been conserved in each lineage. These studies can also identify features that differ between each lineage, which could represent divergence from the ancestral vertebrate state in either/both lineages. While comparisons between cyclostomes and gnathostomes may reveal ancestral features of vertebrate hox2 regulation in the NC, little is known about the enhancers and regulatory factors underlying Hox NC expression in cyclostomes. Here, we employ cross-species regulatory comparisons between lamprey and gnathostomes to isolate and characterize NC enhancers and investigate ancestral regulation of Hox2 in vertebrates.

### Results

#### Expression of hox2 genes in lamprey cranial NC.

In two species of lamprey, sea lamprey (Petromyzon marinus) and Arctic lamprey (Lethenteron camtschaticum), only two Hox2 paralogues (hoxa2 and hoxδ2) have been identified, but their orthology to gnathostome Hoxa2/Hoxb2 remains unresolved (Fig. 2a)\(^21\,\(^22\). The sea lamprey has six Hox clusters, compared to the four clusters inferred in ancestral gnathostomes\(^22\). A recent reconstruction based on comparisons of gene order at the chromosomal level between vertebrate species supports a model in which the ancestor of cyclostomes and gnathostomes also had four Hox clusters\(^22\), suggesting that the two additional Hox clusters in lamprey arose from duplication events in the lamprey/cyclostome lineage. To investigate the duplication history of lamprey Hox clusters, previous work employed a chromosome-wide analysis of genomic synteny (duplicate gene retention) between lamprey Hox-bearing chromosomes\(^22\). These pairwise comparisons indicated that the chromosomes bearing the hoxa and hoxδ clusters display a significantly closer relationship to each other than to any other Hox-bearing chromosomes. Similarly, the chromosomes carrying the hoxb and hoxε clusters also show a significant enrichment for shared paralogues with each other. This suggests that these pairs of chromosomes derive from duplication events that occurred more recently than the duplication events that gave rise to the other Hox-bearing chromosomes\(^22\). Phylogenetic analyses reveal hoxb and hoxε paralogues consistently clustering in protein trees\(^21\,\(^22\), while hoxa and hoxδ paralogues do not show any clear patterns of clustering. This may be due to the limitations of phylogenetic analyses in resolving the relative timing of ancient duplication events\(^22\). Thus, the evidence from synteny analysis leads us to infer that hoxa2 and hoxδ2 genes are paralogues that arose from duplication in the lamprey/cyclostome lineage.

In exploring hox gene NC regulation in sea lamprey embryos, our time-course analysis of hox1–3 expression revealed nested domains in the pharynx from stage (st) 23, reminiscent of those in Arctic lamprey\(^24\,\(^25\) and in gnathostomes\(^4\,\(^8\) (Fig. 2b–e). hoxa2 is expressed in the NC posterior to PA1 and in the hindbrain posterior to r1, similar to gnathostome Hoxa2. In contrast, hoxδ2 is expressed in r3/r5, notochord, and in posterior pharyngeal endoderm and mesenchyme, but not at high levels in NC (Fig. 2b, c). If hoxδ clusters arose by duplication in the lamprey/cyclostome lineage, as supported by synteny analysis, this suggests that hoxδ2 expression diverged after this duplication. Given the similarity in NC expression of hoxa2 to gnathostome Hoxa2, we focused our comparative regulatory analysis on mechanisms mediating lamprey hoxa2 expression in NC.
Gnathostome Hox2 NC enhancers function in lamprey. To explore conservation of the regulatory network upstream of Hox2 in the NC, we sought to perform cross-species enhancer analysis by testing gnathostome NC enhancers for activity in lamprey embryos. As part of this approach, we wanted a way to globally monitor the NC in vivo. In zebrafish, the crestin promoter/enhancer element is a highly specific tool for monitoring NC development, as it is active from pre- to post-migratory stages across multiple axial levels, making it a good candidate for marking NC in lamprey. In lamprey transgenic reporter assays, the crestin element mediates spatiotemporal expression in NC similar to its activity in zebrafish, and its activity is sensitive to perturbation of the same transcription factor binding sites (Supplementary Figure. 1a–d; Fig. 3b; Supplementary Table 1).

Initial green fluorescent protein (GFP) expression in premigratory NC was observed at st21 and maintained as NC delaminated and migrated ventro-laterally to populate the pharynx (Supplementary Figure 1b–c). Frontal sections at st24 revealed that GFP transcripts are present in the NC-derived pharyngeal mesenchyme (Fig. 3d). These data demonstrate that NC cells can be labelled and visualized in vivo by a reporter assay in the developing lamprey and suggest that the crestin element is interpreted in lamprey by conserved upstream components of an ancestral NC GRN. This serves as a proof of concept for interspecies analysis of NC enhancers.

Next, we investigated whether upstream regulatory inputs required for Hoxa2 expression in gnathostome NC are present in lamprey using homologous Hoxa2 NC enhancers from three regions from the lamprey Hox2 locus have enhancer activity, with an r2/r4 enhancer positioned within the exons and intron (c). The Hoxa2-hoxc3 intergenic region drives reporter expression in the hindbrain and NC. However, it is not known whether this is through independent or shared NC/hindbrain enhancers, specific cis-elements have not been identified, and the relationship of this region to the gnathostome Hoxa2 and Hoxb2 enhancers is unclear.
gnathostomes (zebrafish (zf), fugu (f), mouse (m)) (Fig. 3a). In both lamprey and zebrafish embryos, all three gnathostome enhancers mediated GFP reporter expression in the developing pharynx posterior to PA1 (Fig. 3b; Supplementary Table 2; Supplementary Figure 2). Frontal sections of transient transgenic lamprey embryos confirmed that GFP transcripts were prominently expressed in NC-derived pharyngeal mesenchyme equivalent to crestin reporter expression domains and endogenous hox\(\alpha_2\) (Fig. 3d).

In gnathostomes, NC Hoxa2 expression is regulated by 5’-flanking elements (NC1–5) that partially overlap those of a separate r3/r5 enhancer (RE1-5, Krox20, Sox) (Figs. 1a, 3a).
Fig. 2 Embryonic time course showing expression of hox genes in the lamprey hindbrain and cranial neural crest (NC). a Genomic organization of Hox genes in lamprey and mouse. Boxes represent Hox genes, which are organized into paralogue groups based on their sequence. The arrow above the clusters denotes the direction of hox gene transcription. Lamprey hox genes from paralogue groups 1–3 were examined for NC expression in this study and their expression in cranial NC is denoted by green/white shading. b Lateral views of lamprey embryos from stages (st)21 to 26, showing hox gene expression domains in the developing head. Pharyngeal arches are numbered and rhombomere-specific domains (r) indicated. The arrowhead marks weak hoxb2 expression in mandibular mesoderm at st26. c Frontal sections through lamprey embryos showing hox gene expression domains within the developing pharynx. Pharyngeal arches are numbered. d Schematic of a frontal section through the lamprey st24.5 embryonic pharynx with tissue domains annotated; NC domains are shaded in blue. Scale bars: 200 μm (b); 100 μm (c). e Schematic depicting hox expression in the lamprey hindbrain and NC at st23 and st24. ec, ectoderm; en, endoderm; m, mesoderm; mo, mouth; nc, neural crest; r, rhombomere; st, stage

Of these, the NC3 element is the most highly conserved: global sequence alignment using Multi-LAGAN identified sequence conservation of NC3 extending to sharks (Fig. 3a). In previous work, two 15 bp deletions within NC3 were each found to abolish NC reporter expression in mouse. To determine whether the same cis-elements are required for NC activity of Hox2α(m) in lamprey, we generated two variants with these deletions in NC3. While hindbrain activity persisted, these reporters exhibited severely diminished NC activity in zebrafish and lamprey (Fig. 3a, c; Supplementary Figure 3a, b; Supplementary Table 3). Analogously, activity of a gnathostome Hoxb2 NC enhancer (hoxb2a( zf)) also depended upon the same regulatory sites in lamprey as in zebrafish (Supplementary Figure 3c, d). Since Hoxa2(m) and hoxb2a( zf) NC enhancers require the same cis-motifs across distant species, this suggests that an ancestral GRN upstream of Hox2 in NC patterning has been retained in lamprey and gnathostome lineages.

Conservation of binding sites in a lamprey hoxα2 NC enhancer. Gnathostome Hoxα2 and lamprey hoxα2 share similarity in their cis-regulatory architecture for rhombomeric hindbrain expression, as r2 and r4 enhancers are embedded in conserved locations (Fig. 1a, c)13,18. However, our global sequence alignment failed to reveal conservation of NC regulatory elements (NC1–5) 5′ of the lamprey hoxα2 genes (Fig. 3a). Hence, we manually searched for short sequence motifs within an enhancer (hoxα2–4 kb) in the 5′-flanking region of hoxα2 that recapitulates endogenous hoxα2 expression in the hindbrain (r3–r5), NC posterior to PA1 and somites (Fig. 4a–e; Supplementary Figure 4; Supplementary Table 2). In mouse Hoxα2, the most highly conserved elements for NC expression are NC3 and part of NC2, while those necessary for r3/r5 activity are Krox20, Sox, RE2, and RE3 sites. Surveying the lamprey enhancer, we identified short matching sequences for Krox20, Sox, and NC3 (Fig. 4f; Supplementary Figure 5). To address whether a smaller region containing these sites retains enhancer activity, we cloned a 1530 bp region encompassing the conserved sites with ~500 bp on each side (hoxα2 element A) and demonstrated that it mediates reporter expression equivalent to hoxα2–4 kb in lamprey (Fig. 4g, j). To test if the conserved sites are required for tissue-specific enhancer activity, we assayed variants of hoxα2–4 kb in which these sites were mutated (Fig. 4f, h–j). Deleting Krox20-Sox sites (ΔKrox20) resulted in the loss of r3/r5 expression, but maintenance in r4 and NC (Fig. 4f, h). In contrast, deleting the conserved sites within NC3 (ΔNC3) caused loss of all rhombomeric and NC activities, but somitic expression is retained (Fig. 4f, i–j).

Inspection of the conserved sites between the lamprey hoxα2 and gnathostome Hoxα2 enhancers revealed that they match closely to consensus transcription factor binding site motifs for factors involved in early hindbrain and NC patterning. In addition to the previously characterised Krox20 sites, we identified three short blocks of conserved sequence that correspond to consensus binding motifs for Sox, Meis, and Pbx-Hox factors (Fig. 4f). These motifs each fall within regions functionally required for enhancer activity in gnathostomes and lamprey, notably NC3 (Fig. 4a, f, h–j; Supplementary Figure 3a–b), suggesting that Meis, Pbx, and Hox factors may provide conserved and essential inputs into these enhancers. Thus, lamprey hoxα2 and gnathostome Hoxα2 appear to be regulated in the hindbrain and NC through conserved transcription factor binding sites retained during vertebrate evolution. This provides further support for an ancestral GRN upstream of Hox2 in NC patterning that has been retained in lamprey and gnathostomes.

Hoxα2 and Hoxb2 NC enhancers are divergent paralogues. The identification of conserved Meis, Pbx, and Hox binding motifs in the lamprey hoxα2 and gnathostome Hoxα2 enhancers is significant as equivalent sites are present and functionally required in the mouse Hoxb2 and zebrafish hoxb2α enhancers (Fig. 1b; Supplementary Figure 3c). To explore whether these are homologous sites and to interrogate common and diverged features of the hoxα2, Hoxα2, and Hoxb2 enhancers, we used the Krox20 sites, implicated in r3/r5 expression, as an anchor to align the enhancer sequences. This revealed striking conservation of the sequence and order of Krox20, Sox, Meis, and Pbx-Hox sites between these enhancers, with relatively low sequence conservation within the intervening regions (Fig. 5a; Supplementary Figure 6). Based on these conserved sites and relative positions, we infer that Hoxα2 and Hoxb2 NC enhancers are ancient paralogues, derived from an ancestral vertebrate Hox2 enhancer that contained these sites. These paralogous enhancers appear to have diverged in gnathostomes, such that the mouse Hoxα2 enhancer is inactive in r4 but expressed in r4-derived NC, while the Hoxb2 enhancer is active in both r4 and its NC (Fig. 5a)11. The lamprey hoxα2 NC enhancer exhibits the combined activity of both mouse Hoxα2 and Hoxb2 enhancers, suggesting that it may reflect the ancestral state. We searched upstream of lamprey hoxb2, finding conservation of Krox20 and Sox sites, but no Meis or Pbx-Hox sites (Fig. 5b). This suggests that the ancestral sites for r4/NC enhancer activity were lost upstream of hoxb2, consistent with the expression of hoxb2 in r3/r5 but not in r4 and r4-derived NC (Fig. 2b, c).

Lamprey meisc and hoxα2 are similarly expressed in NC. Identification of Meis and Pbx-Hox consensus binding motifs in the hoxα2 enhancers of lamprey and gnathostomes implies that these factors may play conserved roles in regulating vertebrate Hox2 genes in the hindbrain and NC. Meis, Prep, and Pbx are members of the TALE (Three-Amino-Acid-Loop-Extension) homeodomain family that have diverse roles in patterning tissues, including hindbrain and NC30–33. Additionally, they can act as cofactors for other transcription factors, including Hox proteins. To initially explore whether Meis factors are linked with regulation of NC in lamprey, we characterized expression of four lamprey meisc genes. We found that meisc shows early NC expression with a spatiotemporal pattern similar to hoxα2
Fig. 3  Conserved activity of gnathostome Hoxa2 neural crest (NC) enhancers in zebrafish and lamprey.  

**a** Sequence alignment of gnathostome Hoxa2-Hoxa3 and lamprey hox2-hox3 gene loci against the human locus. Conserved non-coding sequences (pink), untranslated regions (UTRs) (cyan) and coding sequences (blue) are highlighted. The relative locations of the mouse hindbrain and NC cis-elements (top) are shown. Gnathostome Hoxa2 enhancers used for cross-species reporter analysis are detailed below the alignment. Letters within parenthesis indicate species of origin of the enhancer: zf, zebrafish; f, fugu; m, mouse.  

**b** Green fluorescent protein (GFP) reporter expression in zebrafish and lamprey embryos (lateral views), mediated by wild-type (b) and mutated (c) gnathostome NC enhancers. For zebrafish, the otic vesicle is circled and GFP expression in rhombomeres (r) and pharyngeal arches (2–5) indicated. Lamprey pharyngeal arches are labelled (2–4). GFP-expressing embryos shown are representative of the expression potential of the reporter construct in each case, as inferred from screening many (typically more than 100) injected embryos. Supplementary Table 2 provides the number of embryos and details of specific expression for all constructs in lamprey. Injection statistics for the transient transgenic zebrafish embryos shown in c are given in Supplementary Table 3.  

**d** Frontal sections through the transient transgenic lamprey embryos shown in Fig. 3b, with GFP transcripts detected by in situ hybridisation, revealing expression in NC-derived mesenchyme (arrowheads) in the pharyngeal arches (numbered). Scale bars: 100 μm
Fig. 4 Characterization of a lamprey hox2 neural crest (NC)/hindbrain enhancer. a The mouse Hoxa2-Hoxa3 genomic region and its equivalent from the lamprey hox2 are shown, with Hox gene exons annotated (blue arrows). hox2 upstream regions assayed for reporter activity in this study, with or without the c-Fos minimal promoter, are shown. b–e Lateral views (b, d) and frontal sections (c, e) of st24.5 lamprey embryos, comparing endogenous expression of hox2 (b, c) to GFP reporter expression mediated by hox2 − 4 kb (d, e). Pharyngeal arches are numbered and rhombomeric expression detailed. Arrowheads point to PA2 NC expression. f Multiple sequence alignment of the Hoxa2 NC enhancer from gnathostomes with the lamprey hox2 enhancer, showing conserved sites (yellow). The positions of characterized mouse cis-elements (Krox20, Sox, RE2-3, NC3) are marked above the alignment. The enhancer schematic (a) shows the position of these elements within the assayed hox2 upstream regions, with conserved (shaded boxes) or divergent (empty boxes) cis-elements highlighted. Consensus binding motifs from the JASPAR database76 for Krox2077, Sox1178, Meis179, and Pbx-Hox80 factors are shown below the alignment, as well as sequences deleted in hox2 − 4 kb ΔKrox20 and ΔNC3 variants. The non-aligning interval between these conserved regions is ~250–400 bp and varies in length between species. Supplementary Figure 5 contains the full alignment. g–j Lateral (g, i) and dorsal (j) views of st24.5 lamprey embryos showing GFP reporter expression driven by the enhancers detailed in a. Pharyngeal arches are numbered, with expression in rhombomeres (r) and somites (s) annotated. GFP-expressing embryos shown are representative of the expression potential of the reporter construct in each case, as inferred from screening many (typically more than 100) injected embryos. Supplementary Table 2 provides the number of embryos and details of specific expression for all constructs in lamprey (Fig. 4a–d). This expression data and the conserved consensus Meis binding sites are consistent with the notion that meis genes may have played an ancestral role in regulating Hox2 in NC during vertebrate evolution.

Occupancy of TALE and Hox proteins on Hox2 enhancers. Hoxb1, Pbx, and Meis bind to sites within the mouse Hoxb2 hindbrain/NC enhancer that are required for its activity11,17. The presence of homologous sites within NC3 of the mouse Hoxa2 enhancer suggests that such factors may also regulate its activity in the hindbrain and NC. This is significant because previous characterisation of this NC enhancer did not uncover factors that bind to NC3 nor provide insight into its underlying regulatory mechanism10,35. The divergent activities of the Hoxa2 and Hoxb2 enhancers, particularly in r4, suggest that there may be differences in their interactions with upstream regulatory factors despite the presence of homologous binding sites. This led us to investigate binding properties of Meis, Pbx, and Hox proteins on these enhancers.
For insight into NC, we harnessed published genome-wide binding data for Meis, Pbx, and Hoxa2 in PA2 of mouse embryos.26,27 Focusing on the NC enhancers of Hoxa2 (NC3) and Hoxb2 (HRE), we observed similar binding profiles for these factors over each of the enhancers, consistent with their regulatory activity in NC. There is enrichment for occupancy of Meis and Pbx, in keeping with a role for these TALE factors in controlling Hox2 NC activity (Fig. 7a, b). It is interesting that Hoxa2 also binds to these enhancers in PA2, suggesting that there may be auto- and cross-regulatory inputs from Hox2 proteins into the NC Hox code. This is analogous to the important roles for auto- and cross-regulatory circuits in regulating Hox expression in other tissues, including hindbrain rhombomeres.11,13,15 Since Pbx and Meis can act as Hox cofactors, they may interact with Hoxa2 on these NC enhancers. Together with the presence of essential Meis motifs
and bipartite Pbx-Hox sites, this raises the possibility of both Hox-dependent and -independent inputs of Pbx and Meis into NC Hox expression.

With respect to the developing hindbrain, chromatin immunoprecipitation-sequencing (ChIP-seq) approaches are not feasible for individual rhombomeres due to the small number of cells and limiting amounts of embryonic material. In addition, there is no suitable anti-Hoxb1 antibody for ChIP-seq experiments. To circumvent these challenges, we generated a mouse embryonic stem (ES) cell line (KH2) carrying an inducible...
locus-specific insertion of Hoxb1 marked with Flag epitopes and used it in combination with programmed differentiation of ES cells into neural fates. This enabled us to use anti-Flag antibodies for Hoxb1 ChIP-seq and to obtain sufficient cell populations for a comparative series of genome-wide binding experiments. This cell culture system has previously been shown to exhibit global changes of gene expression that are similar to early in vivo phases of neural development, including the sequential activation of hindbrain-expressed Hox genes and their cofactors (such as Meis)\(^38,40\). We have previously applied this system to investigate the genome-wide binding properties of Hoxa1 and TALE proteins in neural cells, uncovering in vivo regulatory interactions relevant to hindbrain patterning\(^39-42\).

Using this same approach for neural cells, we found similarities and significant differences in binding patterns of Hox and TALE proteins between the paralogous Hox2 enhancers (Fig. 7a, b). The Hoxb2 (HRE) enhancer shows robust binding of Hoxb1, Pbx, and Meis, plus prominent p300 recruitment, consistent with their established in vivo role in mediating Hoxb2 expression in r4\(^11,17\). In contrast, the Hoxa2 (NC3) enhancer lacks discernable binding of Hoxb1, has reduced levels of Pbx and Meis occupancy, and displays differential binding patterns of Prep1 and Prep2. These properties, in combination with absence of p300, directly correlate with its lack of activity in r4. These differences in TALE and Hoxb1 binding and r4 activity presumably reflect sequence divergence or differences in epigenetic states between the two enhancers. Studies on Pbx-Hox protein binding have shown that small sequence variations within the canonical Pbx-Hox bipartite binding sites influence the selectivity for specific Hox proteins\(^38\). However, sequence comparisons of the Hoxb2, Hoxa2, and hoxa2 enhancers show that the core consensus Meis and Pbx-Hox sites are identical (Figs. 4f, 5a; Supplementary Figure 6). In contrast, the sequences around these sites differ considerably between enhancers: for example, sequences immediately 5' of the Meis site are shared between Hoxb2 and hoxa2 but not Hoxa2 (Supplementary Figure 6). While these differences in neighbouring sequences may have arisen by sequence drift and be functionally neutral, an intriguing alternative is that they may have functional significance in modulating binding to the conserved motifs and impacting r4 activity. Taken together, ChIP-seq data link TALE (Pbx and Meis) and Hox proteins to regulation of Hox2 genes in both the hindbrain and NC.

Discussion
Here, we have investigated the ancestral regulation of Hox2 genes in the NC and hindbrain of vertebrates, using interspecies cis-regulatory comparisons between gnathostomes and lamprey. We demonstrated that gnathostome Hoxa2 and Hoxb2 NC enhancers are capable of driving equivalent NC expression in lamprey as they do in gnathostomes and that a homologous enhancer is present upstream of lamprey hoxa2. Sequence comparisons and regulatory analysis revealed conserved Meis, Pbx, and Hox binding sites between gnathostome Hoxa2/Hoxb2 and lamprey hoxa2 NC enhancers that are required for their activity. The lamprey hoxa2 NC enhancer appears to have retained ancestral activity in both NC and hindbrain, while the paralogous Hoxa2 and Hoxb2 NC enhancers have differentially partitioned NC and hindbrain activities in the gnathostome lineage. Regulatory divergence has also occurred between lamprey Hox2 paralogues, with hoxδ2 appearing to have lost the regulatory sites for r4/NC enhancer activity. This suggests that a regulatory circuit with input from TALE and Hox proteins was an important component of the GRN for Hox2-dependent NC patterning in ancestral vertebrates that has been maintained during evolution (Fig. 8a). TALE factors are part of an ancient patterning system\(^38,43\) that may have multiple roles in coupling Hox expression to the core NC GRN. These findings raise a number of interesting points and avenues for further investigation.

At the mechanistic level, little is known with respect to shared versus independent inputs that govern axial patterning in the hindbrain and NC. This is because the mechanisms regulating Hox expression in the NC are relatively unclear compared to the current knowledge of rhombomeric Hox regulation\(^3\). Analyses of mouse Hoxa2 and Hoxb2 NC enhancers provided conflicting mechanisms for NC expression of Hox genes. Hoxa2 supported evidence for independent enhancers mediating expression in r4 and r4-derived NC, since the NC enhancer is not active in r4 and a separate intronic/exonic enhancer drives r4 expression\(^10,13\). In contrast, Hoxb2 uses common elements to control r4 and NC expression, suggesting similar or shared regulatory requirements in these tissues\(^11\). Our analyses resolve this paradox, providing evidence that the Hoxa2/Hoxb2 NC enhancers each retain conserved Meis, Pbx, and Hox binding sites, which are deployed in slightly different ways in mediating tissue-specific activities. hoxa2 is the only lamprey hox2 parologue expressed in PA2 NC (Fig. 2b, c) and we uncovered the presence of homologous functional motifs (Meis, Pbx, and Hox) in an upstream enhancer with activity in r4 and NC. Based on sequence conservation, position, and regulatory activity, we infer that this enhancer is homologous to gnathostome Hoxa2/Hoxb2 NC enhancers. Sequence comparisons suggest that lamprey hoxδ2 has diverged and is missing these NC motifs but retains Krox20 and Sox sites, consistent with its expression in r3/r5. Since the lamprey hoxa2 NC enhancer exhibits the combined activity of the Hoxa2 and Hoxb2 enhancers, we consider it likely that it reflects the ancestral state. Thus, we suggest that Hox2 was ancestrally regulated in r4 and NC by a shared enhancer through inputs by Meis, Pbx, and Hox (Fig. 8b). Alternatively, if the ancestral NC enhancer did not have the r4 activity, then gnathostome Hoxb2 and the lamprey hoxa2 NC enhancers independently evolved the ability to mediate expression in r4.

The loss of r4 activity from the Hoxa2 NC enhancer may have been mitigated by the existence of a second r4 enhancer, located in the exon/intronic region. We have previously shown that a region containing the lamprey hoxa2 exon1, intron, and exon2 was also capable of driving reporter expression in r4 in lamprey embryos\(^18\). This implies that the rhombomeric activity of this genomic region is ancestral to extant vertebrates. Hence, in lamprey hoxa2 there are two r4 enhancers, which may have partially redundant/shadow activities. It is notable that while both mediate expression in r4 in lamprey, only the upstream enhancer drives expression in NC, which indicates that there is something unique about the upstream enhancer that helps to potentiate its activity in both rhombomeres and NC.

The close proximity and partial overlap between the r3/r5 and r4/NC enhancers in hoxa2, in a manner analogous to its gnathostome counterparts (Fig. 1), suggests that some components of these cis-elements may be required for expression in both tissues. Our experiments with the lamprey hoxa2 enhancer showed that deletion of the Meis and Pbx-Hox sites not only led to the loss of r4 and NC expression but also to the loss of r3/r5 expression (Fig. 4). This implies that Meis and/or Pbx-Hox factors are also involved in regulating the expression of hoxa2 in r3/r5. In this regard, the mouse Hoxa2 NC enhancer partially overlaps elements required for r3/r5 activity: deletion of the Meis site in NC3 (ANC3_1) causes the loss of NC activity and also reduces r3/r5 expression when tested in mouse\(^10,44\) and zebrafish (Fig. 3c; Supplementary Figure 3a–b). Thus, the Meis site contributes to both r3/r5 and NC activities of the Hoxa2 enhancer. Deletion of the Pbx-Hox site (ANC3_2) removes NC expression but does not influence r3/r5 activity in mouse\(^10\), which may reflect the control
Evolution of the neural crest Hox code in vertebrates

- Non-vertebrate deuterostomes
- Cyclostomes
- Gnathostomes

Saccoglossus  Amphioxus  Ciona  Lamprey  Zebrafish  Mouse

- hox2 and hoxi2 enhancer divergence
- NC GRN
- Hox GRN for hindbrain segmentation and NC Hox code
- TALE- and Hox-dependent Hox2/hindbrain/NC enhancer
- Nested Hox expression in the neuroectoderm, coincident with Meis

Divergence of Hox2 hindbrain/NC enhancers

Putative ancestral Hox2 upstream enhancer activity

Lamprey hox2 enhancers

Mouse Hox2 enhancers

Conserved functional motifs in ancestral Hox2 upstream enhancer

No anterior NC or r4 activity

No r4 activity

No NC activity posterior to PA2

- r3/r5 activity
- NC-like cells

Fig. 8 Evolutionary model for regulation of Hox2 genes in vertebrates. A model for evolution of the neural crest (NC) Hox code, based on our data. A NC gene regulatory network (GRN) and NC Hox code evolved in ancestral vertebrates and are conserved between cyclostomes and gnathostomes. In ancestral vertebrates, Hox2 NC expression was regulated by TALE (Three-Amino-Acid-Loop-Extension) and Hox factors, through a putative ancestral enhancer with shared NC and hindbrain activities. A model for the divergence of lamprey and mouse Hox2 NC/hindbrain enhancers. Enhancer activity domains are depicted in blue in schematic dorsal views of the hindbrain and pharyngeal arches. Conserved functional motifs (Krox20, Sox, Meis, Pbx-Hox) present upstream of lamprey and mouse Hox2 genes are shown. Lamprey and mouse enhancers show divergent activities. Comparison between expression domains and conserved motifs leads us to suggest that a putative ancestral vertebrate Hox2 enhancer contained cis-elements for r3/r5 expression (Krox20, Sox) and r4/NC expression (Meis, Pbx-Hox). These scenarios assume that duplication events that gave rise to four Hox clusters in early vertebrates occurred prior to the cyclostome/gnathostome split, as the most parsimonious explanation. However, it is also possible that independent genome duplication events may have occurred in cyclostome and gnathostome lineages (see Holland and Ocampo Daza for a recent discussion).
The emergence of the NC during vertebrate evolution provides a key example of how regulatory codes coevolved with novel cell types in an animal body plan. Non-vertebrate deuterostomes, like the hemichordate *Saccoglossus* and the cephaliocordate amphioxus, lack NC but deploy nested AP domains of *Hox* expression to pattern their nervous system. This raises the intriguing possibility that the NC *Hox* code in ancestral vertebrates may have arisen independently, by evolution of new regulatory inputs into *Hox* genes. A further possibility invokes a combination of both, with shared inputs creating a *Hox* prepattern and independent inputs evolving to modulate this in a tissue-specific manner. Our investigation of ancestral *Hox* NC regulation in vertebrates sets the stage for examining the emergence of *Hox* regulation in NC during chordate evolution. This requires comparison of deuterostome development, with a focus on non-vertebrate deuterostome cell types that may be evolutionarily related to NC.

Studies in tunicates and cephaliocordates suggest that they employ similar gene regulatory programs to specify the neural plate border. Recent studies in tunicates, the vertebrate sister group (Fig. 8a), have identified certain embryonic cell populations that display some characteristics of NC cells. For example, in *Ecteinascidia turbinata*, a colonial tunicate, the trunk lateral cells originate beside the neural tube and migrate to give rise to pigmented cell types, leading to their designation as NC-like cells. Trunk lateral cells are also identifiable in *Ciona*, where they express homologues of some key genes of the vertebrate NC-GRN, including *Tfap2a* and *Twit*. However, the homology of trunk lateral cells to NC has been called into question.

*Ciona* *Hox* genes are dispersed across two chromosomes and residual spatial colinearity of expression in the nervous system has been detected for some of them. This may be a general feature of tunicates, with similar *Hox* cluster disintegration seen in species from other tunicate classes. Comparison with amphioxus, which has a single *Hox* cluster and nested colinear *Hox* expression along the AP axis of the neuroepithelium, suggests that *Tfap2a* and *Twit* are relatively divergent in terms of their *Hox* genomic content and expression. *Ciona intestinalis* *Hox2* expression has not been detected in the developing neural tube or neural plate border, but has been described in the larval ectodermal atrial primordia. No defects in larval morphogenesis were detected upon residual spatial colinearity of expression in the nervous system has been observed, leading us to speculate that these factors may have comprised an ancestral deuterostome regulatory code in ancestral vertebrates.

The homology of *Tfap2a* and *Twit* represent ancestral neural elements with a capacity for mediating NC activity in vertebrates. In summary, our finding of functionally conserved *Meis* and *Pbx-Hox* sites in lamprey and gnathostome *Hox2* NC enhancers focuses attention on the role of these factors in NC development. *Meis* and *Pbx* play important roles in patterning diverse tissues during development, including the hindbrain and NC, some of which may be independent of *Hox*. For example, mouse embryos with a conditional deletion of *Meis2* in NC display abnormalities in patterning the bones and connective tissue in PA1, where *Hox* genes are not expressed. Hence, they could also serve as cofactors for other transcription factors, or have independent roles in patterning NC. Therefore, while they have not been linked to the current NC-GRN, TALE factors (Pbx and Meis) may be important components in this network. If so, these transcription factors could be part of a mechanism that couples *Meis* genes to the NC GRN in vertebrate evolution. The conserved expression of *Meis* genes in NC from gnathostomes and lamprey is consistent with an ancestral role in NC and their roles and interactions in NC development require further study.

**Methods**

**Sequence alignment.** Global sequence alignment of *Hox* genomic loci (Fig. 3a) was performed using Multi-LAGAN, with human as the baseline sequence and conserved sequences defined by 60% conservation over 40 bp. Sequence alignments of *Hox2* enhancers (Figs. 4f, 5a, b; Supplementary Figures 3, 5, 6) were performed using AlignX in VectorNTI (Life Technologies).

**Enhancer elements.** Enhancer elements were selected from the published data or based on sequence conservation in cross-species alignments. The DNA for each element was amplified by PCR from genomic DNA or from pre-existing plasmids using KOD Hot Start Master Mix (Novagen). The following primers were used for amplification. The sequences in uppercase represent homology to genomic DNA, and adaptor sequences for cloning are in lowercase text. References are given for primary literature in which the enhancers were identified.

**cresinin-1 kb (25):**

F: 5′-ccgctgcggtaagcggctcaagcctcagc-3′
R: 5′-ggagatctcggcggaggtgac3′
cresinin-296 bp: 5′-ggggtgtggtgggtggggccc-3′
R: 5′-ggagatctcggcggaggtgac3′

**Ciona intestinalis Hox2a (21):**

F: 5′-ggggtgtggtgggtggggccc-3′
R: 5′-ggagatctcggcggaggtgac3′

**Ciona intestinalis Hox2b (21):**

F: 5′-ggggtgtggtgggtggggccc-3′
R: 5′-ggagatctcggcggaggtgac3′

**Generation of reporter constructs.** The HLC vector was created in a previous study focusing on the hindbrain. The following enhancer elements were cloned into the Hugo’s lamprey construct (HLC) reporter vector in a previous study focusing on the hindbrain.

**hox2a 296 bp: 5′-ccgctgcggtaagcggctcaagcctcagc-3′
R: 5′-ggagatctcggcggaggtgac3′

**hox2a 4 kb: 5′-ccgctgcggtaagcggctcaagcctcagc-3′
R: 5′-ggagatctcggcggaggtgac3′

**hox2b 296 bp: 5′-ccgctgcggtaagcggctcaagcctcagc-3′
R: 5′-ggagatctcggcggaggtgac3′

**hox2b 4 kb: 5′-ccgctgcggtaagcggctcaagcctcagc-3′
R: 5′-ggagatctcggcggaggtgac3′

**cresinin-1 kb (25):**

F: 5′-ccgctgcggtaagcggctcaagcctcagc-3′
R: 5′-ggagatctcggcggaggtgac3′

**cresinin-296 bp: 5′-ggggtgtggtgggtggggccc-3′
R: 5′-ggagatctcggcggaggtgac3′

**Ciona intestinalis Hox2a (21):**

F: 5′-ggggtgtggtgggtggggccc-3′
R: 5′-ggagatctcggcggaggtgac3′

**Ciona intestinalis Hox2b (21):**

F: 5′-ggggtgtggtgggtggggccc-3′
R: 5′-ggagatctcggcggaggtgac3′
and Protocol #2018-0184) and California Institute of Technology (lamprey, Pro- 

Lamprey reporter assay. The wild-type Slusarski AB zebrafish line was used for embryo micro-injection experiments using Tol2-mediated transgenesis. A standard injection mix containing 25 ng/l reporter plasmid (generated by mini-prep), 35 ng/l Tol2 transposase messenger RNA, and 0.05% phenol red was micro-injected into fertilized zebrafish embryos at the one-cell stage at a volume of 1 nl. Embryos were screened at 24–30 h post fertilization for fluorescent reporter expression using a Leica M205FA microscope. In assay reports, transfections are characterized by the number and proportion of screened embryos exhibiting the indicated gene expression. In each screen, the tissue-specific GFP expression domains were noted, along with the number and proportion of screened embryos exhibiting GFP expression in each of those domains. The empty HLC vector was used for transgenic reporter assays. For fluorescence imaging in zebrafish embryos, embryos were fixed in MEMFA and dehydrated in methanol for in situ hybridisation.

**Zebrafish and lamprey experiments.** This study was conducted in accordance with the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health and protocols were approved by the Institutional Animal Care and Use Committees of the Stowers Institute (zebrafish, RK Protocol #2015-0149 and Protocol #2018-0184) and California Institute of Technology (lamprey, Protocol #1436-17).

**Zebrafish reporter assay.** The wild-type Slusarski AB zebrafish line was used for embryo micro-injection experiments using Tol2-mediated transgenesis. A standard injection mix containing 25 ng/l reporter plasmid (generated by mini-prep), 35 ng/l Tol2 transposase messenger RNA, and 0.05% phenol red was micro-injected into fertilized zebrafish embryos at the one-cell stage at a volume of 1 nl. Embryos were screened at 24–30 h post fertilization for fluorescent reporter expression using a Leica M205FA microscope. In assay reports, transfections are characterized by the number and proportion of screened embryos exhibiting the indicated gene expression. In each screen, the tissue-specific GFP expression domains were noted, along with the number and proportion of screened embryos exhibiting GFP expression in each of those domains. The empty HLC vector was used for transgenic reporter assays. For fluorescence imaging in zebrafish embryos, embryos were fixed in MEMFA and dehydrated in methanol for in situ hybridisation.

**Lamprey reporter assay.** Lamprey transgenic transgenesis was performed using *P. marinus* embryos at the one-cell stage and I-SceI meganuclease-mediated transgenesis. Injection mixes containing 20 nl/l reporter plasmid (generated by mini-prep), 1× CutSmart buffer (NEB), and 0.5U µl−1 I-SceI enzyme (NEB) were used for lamprey transgenesis. 

**ChIP-seq.** ChIP-seq was performed according to the Upstate protocol as described with modifications. Cells were fixed with 1% formaldehyde by incubating at 37 °C for 11 min. The reaction was quenched for 30 min in a 10% (vol/vol) Serum-NEAA solution overlaid with 100 µl of a 500 mM glycine solution. Cells were sonicated for 25 min in a Bioruptor at high setting and 30 s on-off cycle. Respective antibodies attached to sepharose-A beads were used for immunoprecipitation. Sequencing of ChIP-seq libraries was performed on the Illumina HiSeq 2500, 51 bp single end. Raw reads were aligned to the UCSC mm10 mouse genome with bowtie2. Primary reads from each bam were normalized to reads per million and bigWig tracks visualized at the UCSC genome browser (https://genome.ucsc.edu/).

**Assay for transposable-accessible chromatin-sequencing.** Assay for transposase-accessible chromatin-sequencing was performed as described previously. Fifty thousand cells were counted using the sestere 2.0 cell counter (EMD Millipore). The tagmentation reaction was performed using the Nextera DNA Library Prep Kit (Illumina, FC-121-1030) and libraries indexed using the Nextera Index Kit (Illumina, FC-121-1011). Libraries were size selected by Bluepippin (Sage Science) and sequenced on Illumina HiSeq 2500 instrument. Following sequencing, Illumina Real Time Analysis v1.18.64 and CASAVA v1.8.2 were run to demultiplex reads and generate FASTQ files.

**Reporting summary.** Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

**Data availability**

The authors declare that all data supporting the findings of this study are available within the article and its supplementary information files or from the corresponding author upon reasonable request. All raw sequencing data from this study underlying Fig. 7a have been deposited in the NCBI BionProject database [https://www.ncbi.nlm.nih.gov/ bioproject] under accession code PRJNA341679 and Sequence Read Archive under accession code SRP079975 and PRJNA503882. Original data underlying this manuscript can be accessed from the Stowers Original Data Repository at [http://odc.stowers.org/webmgr/]. A reporting summary for this Article is available as a Supplementary Information file.

**Received:** 5 December 2018 **Accepted:** 26 February 2019 **Published online:** 13 March 2019

| Reference | Title |
|-----------|-------|
| 1. | Green, S. A., Simoes-Costa, M. & Bronner, M. E. | Evolution of vertebrates as viewed from the crest. *Nature*, 520, 474–482 (2015). |
| 2. | Le Douarin, N. & Kalcheim, C. | The *Neural Crest* 2nd edn (Cambridge University Press, Cambridge, 1999). |
10. Maconochie, M. et al. Regulation of regulatory networks patterning cranial neural crest. Dev. Biol. https://doi.org/10.1016/j.ydbio.2018.1003.1016 (2018).

11. McEllin, J. A., Alexander, T. B., Tumpel, S., Wiedemann, L. M. & Krumlauf, R. Ancient evolutionary origin of the neural crest gene regulatory network. Dev. Cell 13, 405–420 (2007).

12. Minoux, M. & Rijli, F. M. Molecular mechanisms of cranial neural crest cell migration and patterning in craniofacial development. Development 137, 2605–2621 (2010).

13. Parker, H. J. & Krumlauf, R. Segmental arithmetic: summing up the cranial neural crest. Trends Dev. Biol. 126, 1483–1494 (1999).

14. Parker, H. J., Pushel, I. & Krumlauf, R. Coupling the roles of Hox genes to cranial neural crest identity during melanoma initiation. Nature 390, 3990–4001 (2012).

15. Mehta, T. K. et al. Evidence for at least six Hox clusters in the Japanese lamprey. Proc. Natl Acad. Sci. USA 110, 1501–1506 (2013).

16. Minoux, M. & Rijli, F. M. Analysis of fugu hoxa2 genes provide evidence for subfunctionalization of neural crest cell and rhombomere cis-regulatory modules during vertebrate evolution. Dev. Biol. 409, 530–542 (2016).

17. Parker, H. J. & Krumlauf, R. Segmental arithmetic: summing up the Hox gene regulatory network for hindbrain development in chordates. Wiley Interdiscip. Rev. Dev. Biol. 6, https://doi.org/10.1002/wdev.2867 (2017).

18. Wada, H., Escriva, H., Zhang, S. & Laufer, V. Conserved RARE localization in amphioxus Hox clusters and implications for Hox code evolution in the vertebrate neural crest. Dev. Dyn. 235, 1522–1531 (2006).

19. Tümpel, S. et al. Expression of Hoxa9 in rhombomere 4 is regulated by a conserved cross-regulatory mechanism dependent upon Hoxb1. Dev. Biol. 312, 646–660 (2007).

20. Shm, M. H. et al. The zinc finger gene Krox-20 regulates Hox-2 (Hox2.8) during hindbrain segmentation. Cell 72, 183–196 (1993).

21. Prince, V. & Lumsden, A. Hox-a2 expression in normal and transposed rhombomeres: independent regulation in the neural tube and neural crest. Development 120, 911–923 (1994).

22. Trainor, P. A., Arita-McNaughton, L. & Krumlauf, R. Role of the isthmus and FGFs in resolving the paradox of neural crest plasticity and prepatternning. Science 295, 1288–1291 (2002).

23. Trainor, P. & Krumlauf, R. Plasticity in mouse neural crest cells reveals a new patterning role for cranial mesoderm. Nat. Cell Biol. 2, 96–102 (2000).

24. Shimeld, S. M. & Donoghue, P. C. Evolutionary crossroads in developmental biology: cyclostomes (lamprey and hagfish). Development 139, 2091–2099 (2012).

25. Mehta, T. K. et al. Evidence for at least six Hox clusters in the Japanese lamprey (Lethenteron japonicum). Proc. Natl Acad. Sci. USA 110, 16044–16049 (2013).

26. Smith, J. et al. The sea lamprey gene network provides insights into programmed genome rearrangement and vertebrate evolution. Nat. Genet. 50, 270–277 (2018).

27. Pascual-Anaya, J. et al. Hagfish and lamprey Hox genes reveal conservation of a temporal colinearity in vertebrates. Nat. Ecol. Evol. 2, 859–866 (2018).

28. Takio, Y. et al. Evolutionary biology: lamprey Hox genes and the evolution of jaws. Nature 429, 1–2 (2004).

29. Takio, Y. et al. Hox gene expression patterns in Lethenteron japonicum embryos—insights into the evolution of the vertebrate Hox code. Dev. Biol. 308, 606–620 (2007).

30. Kaufman, C. M. et al. A zebrafish melanoma model reveals emergence of neural crest identity during melanoma initiation. Science 351, aaad2197 (2016).

31. Nonchev, S. et al. Segmental expression of Hox-2 in the hindbrain is directly regulated by Krox-20. Development 20, 122, 543–554 (1996).

32. Budnoch, M. et al. LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. Genome Res. 13, 721–731 (2003).

33. Hahn, R. et al. Analysis of dynamic changes in retinoic-induced transcription and epigenetic profiles of murine Hox clusters in ES cells. Genome Res. 25, 1229–1243 (2015).

34. Medeiros, D. M. The evolution of the neural crest: new perspectives from lamprey and invertebrate neural crest-like cells. Wiley Interdiscip. Rev. Dev. Biol. 2, 1–15 (2013).

35. Yu, J. C., Meulemans, D., McKeown, S. J. & Bronner-Fraser, M. Insights from the amphioxus genome on the origin of vertebrate neural crest. Genome Res. 18, 1127–1132 (2008).

36. Denes, A. et al. Molecular architecture of annelid nerve cord supports common origin of nervous system centralization in bilateria. Cell 129, 277–288 (2007).

37. Jeffery, W. R., Strickler, A. G. & Yamamoto, Y. Migratory neural crest-like cells form body pigmentation in a urochordate embryo. Nature 431, 696–699 (2004).

38. Jeffery, W. R. et al. Trunk lateral cells are neural crest-like cells in the ascidian Ciona intestinalis: insights into the ancestry and evolution of the neural crest. Dev. Biol. 324, 152–160 (2008).

39. Hall, B. K. & Gillis, J. A. Incremental evolution of the neural crest, neural crest cells and neural crest-derived skeletal tissues. J. Anat. 222, 19–31 (2013).

40. Ikuta, T., Yoshida, N., Satoh, N. & Saiga, H. Ciona intestinalis Hox gene cluster: Its dispersed structure and residual colinear expression in development. Proc. Natl Acad. Sci. USA 101, 15118–15123 (2004).
60. Seo, H. C. et al. Hox cluster disintegration with persistent anteroposterior order of expression in *Oikopleura dioica*. *Nature* **431**, 67–71 (2004).

61. Sekigami, Y. et al. Hox gene cluster of the ascidian, *Halocynthia roretzi*, reveals multiple ancient steps of cluster disintegration during ascidian evolution. *Zool. Lett.* **5**, 17 (2017).

62. García-Fernandez, J. & Holland, P. W. H. Archetypal organisation of the amphioxus Hox gene cluster. *Nature* **370**, 563–566 (1994).

63. Mihara, T., Saiga, H., Leduc, N., Zouhir, M. & Hox genes in the larval development of the ascidian *Ciona intestinalis*. *Development* **137**, 1505–1513 (2010).

64. Lemons, D., Fritzenwanker, J. H., Gerhart, J., Lowe, C. J. & McGinnis, W. Co-ordinated patterning of appendages in early bilaterian evolution. *Dev. Biol.* **344**, 358–362 (2010).

65. Albueixch-Crespo, B. et al. Molecular regionalization of the developing amphioxus neural tube challenges major partitions of the vertebrate brain. *PLoS Biol.* **15**, e2001573 (2017).

66. Manzanares, M. et al. Conservation and elaboration of Hox gene regulation during evolution of the vertebrate head. *Nature* **408**, 854–857 (2000).

67. Fisher, S. et al. Evaluating the biological relevance of putative enhancers using Tol2 transposon-mediated transgenesis in zebrafish. *Nat. Protoc.* **1**, 1297–1305 (2006).

68. 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 (2014).

69. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **10**, 21–26 (2013).

70. Smith, K. T., Martin-Brown, S. A., Florens, L., Washburn, M. P. & Workman, J. L. Deacetylase inhibitors dissociate the histone-targeting ING2 subunit from the Sin3 complex. *Chem. Biol.* **17**, 65–74 (2010).

71. Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: a method to assay chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.* **109**, 21 29–21 29 (2015).

72. Khan, A. et al. JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res.* **46**, D1284 (2018).

73. Smith, K. T., Martin-Brown, S. A., Florens, L., Washburn, M. P. & Workman, J. L. Deacetylase inhibitors dissociate the histone-targeting ING2 subunit from the Sin3 complex. *Chem. Biol.* **17**, 65–74 (2010).

74. Langmead, B. & Saltzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

75. Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: a method to assay chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.* **109**, 21 29–21 29 (2015).

76. Khan, A. et al. JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res.* **46**, D1284 (2018).

77. Meng, X., Brodsky, M. H. & Wolfe, S. A. A bacterial one-hybrid system for determining the DNA-binding specificity of transcription factors. *Nat. Biotechnol.* **23**, 988–994 (2005).

78. Badis, G. et al. Diversity and complexity in DNA recognition by transcription factors. *Science* **324**, 1720–1723 (2009).

79. Chang, C. P. et al. Meis proteins are major in vivo DNA binding partners for wild-type but not chimeric Pax proteins. *Mol. Cell. Biol.* **17**, 5679–5687 (1997).

80. Qu, L., Wright, D. D. & Kamps, M. P. Fusion with E2A converts the Pax1 homeodomain protein into a constitutive transcriptional activator in human leukemias carrying the t(1;19) translocation. *Mol. Cell. Biol.* **14**, 3938–3948 (1994).

81. Sacerdot, C., Louis, A., Bon, C., Berthelot, C. & Roest Crollius, H. Chromosome evolution at the origin of the ancestral vertebrate genome. *Genome Biol.* **19**, 166 (2018).

82. Holland, L. Z. & Ocampo Daza, D. A new look at an old question: when did the second whole genome duplication occur in vertebrate evolution? *Genome Biol.* **19**, 209 (2018).

**Acknowledgements**

We thank Dorit Hockman, Tetsuto Miyashita, and Megan Martik for lamprey husbandry assistance, the Stowers Institute aquatics facility for zebrafish care, and Histology facility for sectioning assistance. This study was conducted in accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the NIH and protocols approved by the Institutional Animal Care and Use Committees of the Stowers Institute (Zebrafish, RK Protocol: #2015-0149), California Institute of Technology (lamprey, MEB Protocol: #1436-11), and the veterinary office of UZH and the Cantor of Zürich. K.D.P., C.H., and C.M. were supported by Science Foundation (SNSF) professorship (C.M. grant 170623), a Marie Curie Career Integration Grant from the European Commission (C.M. grant PCIG14-GA-2013-631984), the Swiss Cancer League, and the Cantor of Zürich. H.J.P., B.D.K., L.M.W., and R.K. were supported by the Stowers Institute (R.K. grant #2013-1001). S.A.G. and M.E.B. were supported by grants R01NS086907 and R01DE0197911.

**Author contributions**

H.J.P., M.E.B. and R.K. conceived this research programme. H.J.P., B.D.K., K.D.P. and C. H. conducted the experiments. S.A.G. performed lamprey husbandry. C.K.K. and C.M. developed the *crestin* transgenic zebrafish line and associated constructs. H.J.P., B.D.K., C.M., L.M.W., M.E.B. and R.K. analysed the data, discussed the ideas and interpretations, and wrote the manuscript.

**Additional information**

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-019-09197-8.

**Competing interests:** The authors declare no competing interests.

**Reprints and permissions information** is available online at http://npg.nature.com/reprintsandpermissions/

**Journal peer review information:** Nature Communications thanks David McCauley, Elena Silva and the other anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

**Publisher’s note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s) 2019