

# SCIENTIFIC REPORTS



OPEN

## SOXE neofunctionalization and elaboration of the neural crest during chordate evolution

Andrew Tai<sup>1,\*</sup>, Martin Cheung<sup>1,\*</sup>, Yong-Heng Huang<sup>2</sup>, Ralf Jauch<sup>2</sup>, Marianne E. Bronner<sup>3</sup> & Kathryn S. E. Cheah<sup>1</sup>

Received: 20 June 2016

Accepted: 15 September 2016

Published: 13 October 2016

During chordate evolution, two genome-wide duplications facilitated acquisition of vertebrate traits, including emergence of neural crest cells (NCCs), in which neofunctionalization of the duplicated genes are thought to have facilitated development of craniofacial structures and the peripheral nervous system. How these duplicated genes evolve and acquire the ability to specify NC and their derivatives are largely unknown. Vertebrate *SoxE* paralogues, most notably *Sox9/10*, are essential for NC induction, delamination and lineage specification. In contrast, the basal chordate, amphioxus, has a single *SoxE* gene and lacks NC-like cells. Here, we test the hypothesis that duplication and divergence of an ancestral *SoxE* gene may have facilitated elaboration of NC lineages. By using an *in vivo* expression assay to compare effects of *AmphiSoxE* and vertebrate *Sox9* on NC development, we demonstrate that all SOXE proteins possess similar DNA binding and homodimerization properties and can induce NCCs. However, *AmphiSOXE* is less efficient than *SOX9* in transactivation activity and in the ability to preferentially promote glial over neuronal fate, a difference that lies within the combined properties of amino terminal and transactivation domains. We propose that acquisition of *AmphiSoxE* expression in the neural plate border led to NCC emergence while duplication and divergence produced advantageous mutations in vertebrate homologues, promoting elaboration of NC traits.

Urochordates (Tunicates), Cephalochordates (Lancelets) and Vertebrates constitute the three extant groups within the chordate lineage. Although they share the fundamental chordate body plan with a dorsal hollow nerve cord, notochord and paired gill arches, vertebrates uniquely possess neural crest cells (NCCs). The neural crest is a multipotent stem cell population that arises from the dorsal region of the neural tube, undergo an epithelial-mesenchymal transition and migrate to distant locations to differentiate into multiple, diverse derivatives, including melanocytes, peripheral ganglia and craniofacial cartilage. Acquisition of the NC is thought to have facilitated a shift in the chordate lineage from filter-feeders to active predators<sup>1</sup>. Deciphering the molecular underpinnings of how NC is evolved as a multipotent migratory population is essential for understanding the evolutionary origin of vertebrates.

The previous discoveries of migratory pigment cells<sup>2,3</sup> and neuronal progenitors with a subset of neural crest-like properties<sup>4</sup> within the neural plate border together with phylogenetic analysis place tunicates as the sister group of vertebrates<sup>5</sup>. In contrast, amphioxus completely lacks NC-like cells<sup>6</sup>. Tunicates have small genomes with loss of several developmental genes whereas the genome of amphioxus has retained synteny with vertebrate genome<sup>7,8</sup>. Thus amphioxus represents an excellent extant surrogate to the chordate ancestor for exploring conservation and divergence of evolutionary mechanism in NC development.

All vertebrates studied to date contain at least two to three paralogs of *SoxE* transcription factors family (*Sox8*, *Sox9* and *Sox10*), which are crucial for NC development. The relative timing of their expression onset and function during NC development differ depending on the species<sup>9–15</sup>. In birds and mammals, *Sox9* is the first *SoxE* gene to be expressed in the prospective NCCs and is rapidly downregulated as NC migrates further, whereas

<sup>1</sup>School of Biomedical Sciences, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China.

<sup>2</sup>Genome Regulation Laboratory, Drug Discovery Pipeline, Key Laboratory of Regenerative Biology, Guangdong Provincial Key Laboratory of Stem Cell and Regenerative Medicine, South China Institute for Stem Cell Biology and Regenerative Medicine, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou, Guangdong, 510530, China. <sup>3</sup>Division of Biology 139-74, California Institute of Technology, Pasadena, USA. \*These authors contributed equally to this work. Correspondence and requests for materials should be addressed to K.S.E.C. (email: kathycheah@hku.hk)

*Sox10* expression is activated in the premigratory NC domain and maintained in migrating NCCs<sup>13</sup>. In developing chick embryos, SOX9 induces *Sox10* expression and both are sufficient to transform neural tube progenitors into NC identity. In addition, SOX9 cooperates with SNAIL2 to induce ectopic NCCs with epithelial-mesenchymal transition (EMT) properties<sup>16,17</sup>. Prolonged *Sox9* expression biased NC to form glial cells but not neurons<sup>13</sup>. However, NCCs are still specified in mouse embryos with germ-line specific deletion of *Sox9* (genetically null) but undergo apoptosis resulting in defects in the formation of trunk NC-derivatives<sup>16</sup>. In contrast, *Sox8* expression precedes that of *Sox9* and *Sox10* in NC progenitors of *Xenopus* embryos and is required for the proper onset of NC specification<sup>18</sup>. However, in mammals, SOX8 appears to be dispensable for NC development<sup>19</sup>, whereas SOX9 is essential for chondrogenic differentiation<sup>13,16,20</sup>, and SOX10 for specification of glial, melanocyte and enteric lineages<sup>21,22</sup>. These studies highlight species-specific differences in the relative importance of SOXE proteins during NC development. In addition, the chondrogenic expression and function of vertebrate *Sox9* is conserved in three *SoxE* paralogs in lampreys (jawless vertebrates)<sup>9</sup> and probably in the basal chordate amphioxus, which has a single *SoxE* gene (*AmphiSoxE*) with expression in oral skeleton<sup>23</sup>, suggesting an ancestral role of *SoxE* gene in chondrogenesis. Although previous studies demonstrated a large degree of functional similarity between a *SoxE* ortholog from *Drosophila melanogaster*, *Sox100B* and *Sox10* in mouse NC development<sup>24</sup>, whether *SoxE* gene from amphioxus, the closest living relatives of vertebrates also retains function of and/or differs from vertebrate *Sox9* to regulate NC ontogeny is not clear. Molecular analyses show that the neural-plate border region of amphioxus has many transcription factors common to vertebrates but lacks most “neural crest specifier” genes, most notably *AmphiSoxE*. The exception is *AmphiSNAIL* which is expressed in the dorsolateral portion of the amphioxus neural tube, though no neural crest cells arise from this domain<sup>25</sup>. This implies that the redeployment of *AmphiSoxE* gene and/or other ancestral NC-specifiers in the border region may have facilitated emergence of NCCs. Consistent with this, recent studies suggest that such co-option events could have helped to mediate acquisition of NC cis-regulatory sequences in the *AmphiSoxE* gene<sup>23</sup>. In addition, genome wide duplication results in increasing the number of *SoxE* paralogues and their functional diversity in vertebrates. However, it remains to be determined whether the changes in the coding sequence of *SoxE* genes contribute to their neofunctionalization in NC lineage diversification during the chordate evolution. To address whether duplication and divergence of an ancestral *SoxE* gene may have led to novel functions that facilitated emergence of the NC and its traits in vertebrates, we compared the activity of AmphiSOXE with its vertebrate homologue by analyzing their effects on chicken neural crest development as an assay system.

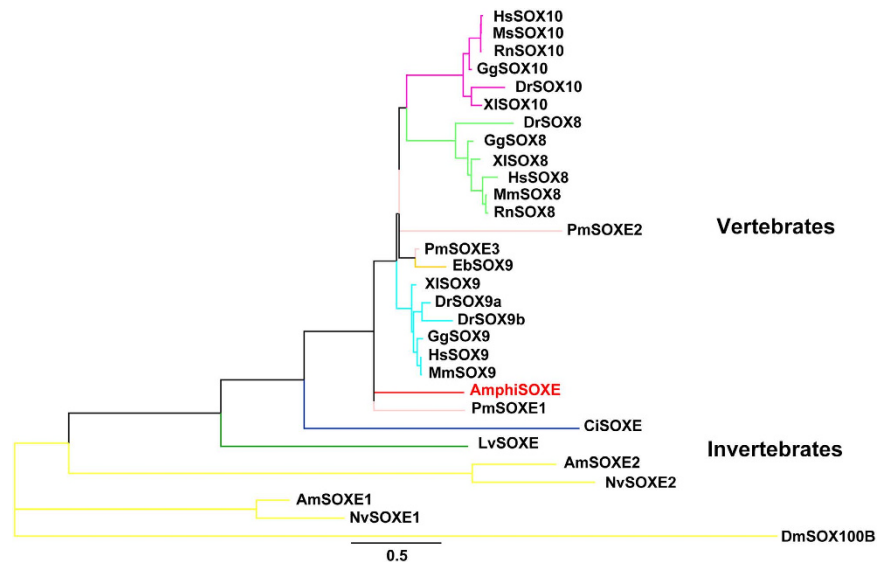
## Results

**Phylogenetic analyses reveal the intermediate state of amphioxus SOXE during evolution.** In contrast to vertebrate *SoxE*, *AmphiSoxE* is not expressed in the neural plate border region<sup>26</sup>. Phylogenetic analysis places AmphiSOXE between invertebrates and vertebrates, suggesting its intermediate state during evolution (Fig. 1). Comparative sequence analysis reveals AmphiSOXE shares only 39.1% amino acid identity with Human SOX9 (Supplemental Fig. 1), with highest conservation in the dimerization (71.4%) and HMG (96.2%) domains but only 22% identity in the transactivation domain.

### Amphioxus SOXE and SOX9 show similar DNA binding and cooperative homodimerization.

Next we asked whether there are detectable changes on the level of SOXE-DNA association that may have contributed to the evolution of new biochemical activities within the SOXE subgroup. AmphiSOXE shares a highly conserved amino acid sequence in both the high-mobility group (HMG) box and the DNA-dependent dimerization (DIM) domain with the mammalian SOXE proteins SOX8, SOX9 and SOX10 (Fig. 2a). The AmphiSOXE HMG box encodes two unique amino acids within the HMG box and a further four amino acids show variability within the SOXE subgroup (Fig. 2a). However, as these residues map to protein interfaces remote from the DNA they are unlikely to directly modulate DNA recognition (Fig. 2b). The N-terminal region upstream of the AmphiSOXE HMG box encodes the 40 amino acid DIM domain characteristic for the SOXE subgroup<sup>27</sup>. The DIM is evolutionarily less conserved than the HMG box and several amino acid variants within the AmphiSOXE-DIM map to positions that were previously reported to influence cooperative homodimerization on palindromic DNA elements<sup>28</sup>. Moreover, SOXE dimerization has recently been suggested to be mediated by DIM:HMG rather than by DIM:DIM interactions<sup>29</sup> (Fig. 2b). The same study demonstrated that SOXE factors tolerate a flexible half-site spacing and dimeric complexes were observed on all tested elements with spacers ranging from 1 to 10 bp (ACAATG(n<sub>1-10</sub>)CATTGT). However, the cooperativity factor and thus the efficiency of the SOXE homodimerization was found to be highest when the half-spacing was 3, 4 or 5 bp<sup>29</sup>. We therefore compared DNA dependent dimerization of AmphiSOXE to mammalian SOX9 with quantitative electrophoretic mobility shift assays (EMSAs) using two palindromic SOXE binding sites with 4 or 5 bp spacers between the half-sites (Fig. 3a). In the absence of the DIM, the SOX9-HMG as well as the AmphiSOXE-HMG form additive dimers as indicated by a cooperativity factor  $\omega$  of ~1 (Fig. 3b,d). In contrast, the SOX9-NHMG and AmphiSOXE-NHMG constructs possessing both DIM domain and HMG box dimerize substantially more effectively with strong positive cooperativity on CD-Rap and Zero DNA elements (Fig. 3c,d). However, both SOX9 and AmphiSOXE bound to and dimerized on DNA in an indistinguishable fashion indicating that they exhibit similar capacity for cooperative binding to the same DNA sequence, despite differences in amino acids within DIM and HMG regions.

**Amphioxus SOXE is capable of inducing neural crest-like cells.** To compare the activity of AmphiSOXE with its vertebrate homologue, we used an electroporation assay to introduce *AmphiSoxE* or chick *Sox9* cDNA in the bicistronic pCIG nuclear-EGFP expression vector into the caudal neural tube of stage HH10–11 chick embryos prior to neural crest emigration and assessed subsequent effects on NC development (Fig. 4a). After one day (stage HH15), NCCs are actively migrating and by 2 days (HH19) have begun to differentiate



**Figure 1. Molecular Phylogenetic analysis of SOXE proteins by Maximum Likelihood method.** The evolutionary history was inferred by using the Maximum Likelihood method based on the Dayhoff matrix based model<sup>53</sup>. The tree with the highest log likelihood ( $-8351.5037$ ) is shown. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.5064)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 30 amino acid sequences. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 273 positions in the final dataset. Evolutionary analyses were conducted in MEGA7<sup>54,55</sup>. The trees are exported to Newick files, which is further modified by the figtree without changing the evolutionary distance. *Homo sapiens* (Hs), *Mus musculus* (Mm), *Rattus norvegicus* (Rn), *Gallus gallus* (Gg), *Xenopus laevis* (Xl), *Danio rerio* (Dr), *Petromyzon marinus* (Pm), *Eptatretus burgeri* (Eb), *Branchiostoma lanceolatum* (Amphi), *Ciona intestinalis* (Ci), *Lytechinus variegatus* (Lv), *Drosophila melanogaster* (Dm), *Apis mellifera* (Am), *Nasonia vitripennis* (Nv).

into neurons within the ganglion core, whereas bipotential neural and glial precursors remain in the dorsal root ganglia (DRG) periphery<sup>16</sup>. Consistent with the possibility that recruitment of SOXE to the neural plate border may have facilitated emergence of NC, AmphiSOXE promoted NC specification and emigration from the dorsal neural tube similar to SOX9. By 24 hours post-transfection (hpt), like Sox9, AmphiSOXE also induced ectopic expression of markers characteristic of NC identity, including *Sox10*, *FoxD3*, *Wnt3a*, and HNK1, while repressing the neural marker SOX2 (Fig. 4b–i; Supplemental Fig. 2a–f). In contrast, neither ectopic expression of these markers nor alteration of their endogenous levels was observed in the untransfected side of the neural tube or embryos treated with pCIG control vector (Supplemental Figs 2 and 3). As with SOX9<sup>13</sup>, transfected cells delaminated from lateral regions of the neural tube, accompanied by breakdown of LAMININ in the basal lamina (Fig. 5c,d,g,h) and reduction of N-CADHERIN expression in the apical region (Fig. 5a,b,e,f). Like co-expression of *Sox9* and *Snail2*<sup>16</sup>, combined electroporation of *AmphiSoxE* and *Snail2* resulted in ectopic NC emigration due to loss of apical-basal polarity in the transfected cells (Fig. 5i–p).

### Overexpression of Amphioxus SOXE drives both dorsally and ventrally migrating cells toward DRG lineages.

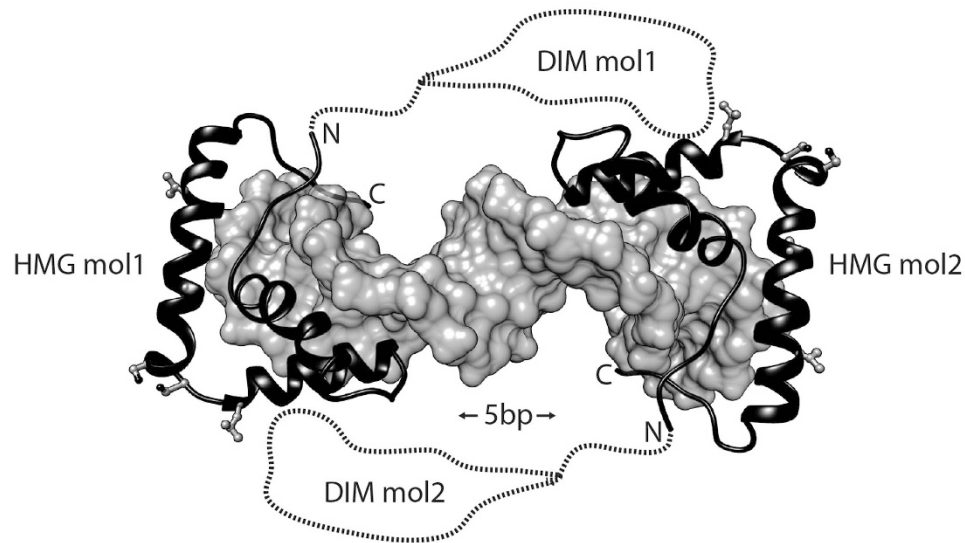
Despite these similarities at early stages, marked differences were noted between AmphiSOXE, SOX9 and control electroporations at 48hpt. Whereas control electroporations result in GFP<sup>+</sup> Schwann cell precursors expressing P0 along the ventral roots, cells overexpressing AmphiSOXE or SOX9 still colonized to the ventral root but failed to differentiate into Schwann cells, suggesting that prolonged expression of AmphiSOXE or Sox9 inhibited NC to differentiate into Schwann cell lineage (Supplemental Fig. 4a–c). In addition, we observed premature migration of GFP<sup>+</sup> cells along the dorsolateral pigment pathway, which normally opens for melanoblast migration only one day later (HH21) in control embryos (Fig. 5q–t). These dorsolaterally located cells fail to express the melanocyte marker MeEM, instead expressing SOX2 (Fig. 5q–s; Fig. 6g,j, n = 10/10), characteristic of sensory neuron specification of NC<sup>30,31</sup>, suggesting a cell fate switch from melanocytic to DRG lineages.

Despite these functional similarities, we also noted marked differences in effects mediated by AmphiSOXE versus SOX9. Similar to GFP control, AmphiSOXE overexpressing embryos had GFP<sup>+</sup> cells expressing early markers of neuronal differentiation HuC/D or ISLET1/2<sup>32</sup> and SOX2, within the core of the DRG that differentiated into neurons and glial (Fig. 6b–g; Supplemental Fig. 5 a–d'; n = 8/8). By contrast, GFP<sup>+</sup> cells in SOX9-electroporated embryos did not express neuronal markers in the DRG core, instead localizing to the periphery where glial cells differentiate<sup>30</sup> (Fig. 6h–j; Supplemental Fig. 5e,f'; n = 12/12). This difference in lineage

a

Amphi_SOXE	MS.....ETETPVEVK.....EETLESLIKREEDTDEESTSGSEDSGESDR.....KRHKFFSQTRAVSNVLKGYDWLVEMFVRWG..SSKSKPHVK	87
SOX8	ML.....DMSEARAQFP.....CSFSQTSSMHSVEDSDSAPPFACSEGLRAGCGGRCDTAAEADERFFAQIRAVSNVLKGYDWLVEMFVRWGCGGTLKPKPHVK	100
SOX9	MNLLDFPMKMTDECEKGLSGAPSPMSDESAQSECF.SGSGSETENTRECENTIFPKGEF...CLKRSEEDKFFVQIRAVSNVLKGYDWLVEMFVRWG..SSKSKPHVK	106
SOX10	MAEQ...LISFVELSFVGSSEPRCLSPGSAFSLGFDGCGGGSGLRASPCGELKRVKK..EQDGEADDDKFFVQIRAVSNVLKGYDWLVEMFVRWG..ASKSKPHVK	105
Consensus	m s HMG fp ir avs vl gydw lvemfcvr g k kchvk	
Amphi_SOXE	RFMNAFMWVCAARRKLADQYPHLHNAELSKTLGKLWRMLNEDKRPFFEEAERLRQHKKDHPYKYCPRRRNSKQCNQSGSDEAGSEASPTISANTIFKALQAESPTGSE	199
SOX8	RFMNAFMWVCAARRKLADQYPHLHNAELSKTLGKLWRMLNEDKRPFFEEAERLRQHKKDHPYKYCPRRRNSKQCNQSGSDEAGSEASPTISANTIFKALQAESPTGSE	207
SOX9	RFMNAFMWVCAARRKLADQYPHLHNAELSKTLGKLWRMLNEDKRPFFEEAERLRQHKKDHPYKYCPRRRNSKQCNQSGSDEAGSEASPTISANTIFKALQAESPTGSE	216
SOX10	RFMNAFMWVCAARRKLADQYPHLHNAELSKTLGKLWRMLNEDKRPFFEEAERLRQHKKDHPYKYCPRRRNSKQCNQSGSDEAGSEASPTISANTIFKALQAESPTGSE	215
Consensus	rfcmnafmwvcaarrkladqyphlhnaelsktlgklwrmlnedkrpffeeaerlrqhkhdhpykycprrrnskqcnqsgsdeagseasptisanti fkalqaesptgse	
Amphi_SOXE	PHSEELKGPSPHDGSYGVTPSSCAPPTPTTPKQDQGMALKADGMKRCSTNLTAIHFDGPHHHHHHFCGHGHNICESNVDIQLFL..LVMSMESEFVBEEDCYLLEN	309
SOX8	HHHS.....DHHGCT...HGPPTPTTPK...IDLHCASN.....GSKCELRLFG...RRIVDSGR..QIDFESNVDISELSEVSNMDFVBEEDCYLLEN	291
SOX9	GMSEVHS...PGEHSGCS...CGPPTPTTPK...IDVQAG.....KVDLKRFG...RPLAEGGRQFIDFRIVDIGELSSDVISNIEFDVBEEDCYLLEN	301
SOX10	HPPEGSFMSDGNPEHPSGCS...HGPPTPTTPK...DELQSG.....KADPKRFG...RSLGEGGK..FIDFCGNVDIGEISHEWSSIMETFDVBEEDCYLLEN	304
Consensus	g pptpcttpk t g idf vdi v s fdv e dyp l n	
Amphi_SOXE	GHFASAS..GHHFSPPTTSYQMSSSATT...TVISS.....SSWAKQNTSFRDNGE.....QLRPFVMEQEHLPFPPFPYTT...	379
SOX8	GHFALPTEFSQATASGSYCGASYSHSCATG...IGASEVW.....AHKCAFASASATEACP.....LRFQIKTEQLSESHYNDQSHGSP	368
SOX9	GHFGVEATHGCVITYTGSYGISSTAPTFAIAGHVMMSKQAPPPFPQPCAPQCAFQPCQCAFQCCQAHILTLTSSSEFGCSCHTHIKTEQLSPSHYSECCCHS.	412
SOX10	GHFGHVG...SYSAAQYGLCSALAVASGHSA...WISKE.....EGVALETVSPFGVDKAKA.....CVKTTETTCFQGEPPHYTLQPS...	374
Consensus	gh q	
Amphi_SOXE	....EHPEASSNYPCQSSYCHSPPEFCYTDYPPPAHSECCFYSFHP...SSSTPPNYNAEPESLPTVAG.....AP..STWHP..SYTQIARE	463
SOX8	GRACYGSYACASVITAASATAASSFASACC..DYT...ILQASNYSPYPG.YPPSLY..QYFHFSSRRFYASPLIN...GLSMPEAHSASSNWCQPVYTLITRP	464
SOX9	....EQIISYSPFNLFHYSYSPETITRCH.DYA...DHONGGSYSHAAGCCSCLYSTFTYNAQREMVTPIALTSGVPSIEQTHSE..QHWEQEVYTLITRP	507
SOX10	....TSQIAYTSLSLHHYGSAFPISRFCH.DYS...DHCPGSGYCHAG.CASGLYSAFSYGHSORPLYTAISDF...SPSGFSHSE..THWEQEVYTLITRP	466
Consensus	q dy y y r p w yt l rp	

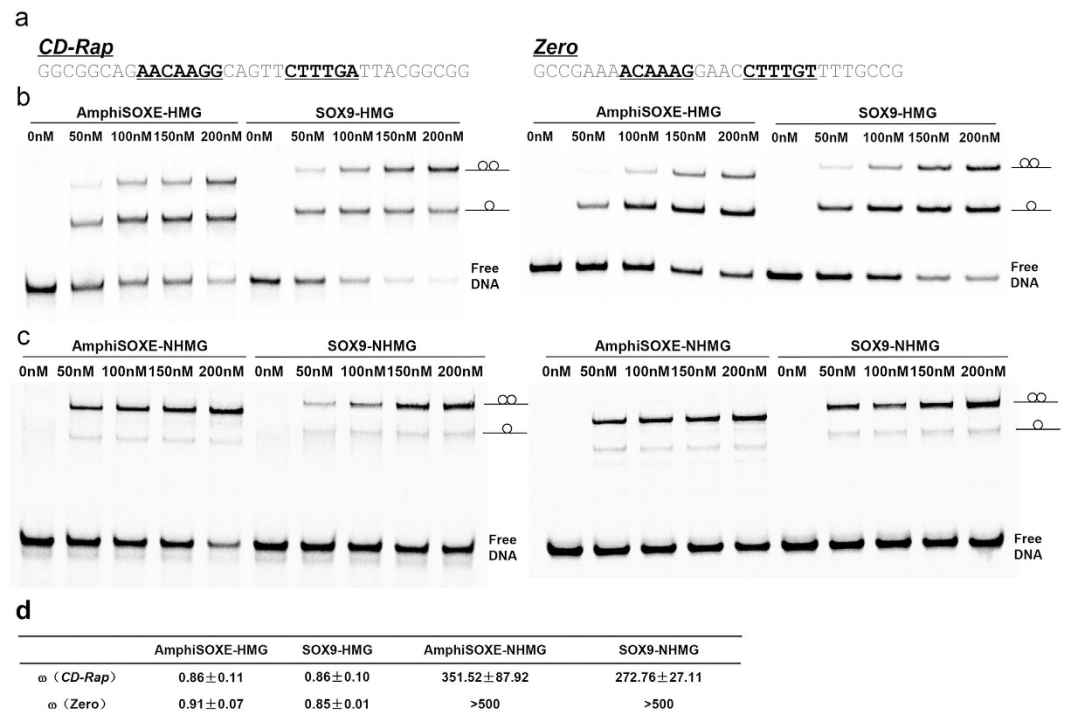
b



**Figure 2. HMG Structure and sequence alignment of SOXE proteins.** (a) Sequence alignment of SOXE proteins. Invariant residues are shaded black, conserved residues gray and variable residues white. The DIM and the HMG are boxed. The arrows indicate the sites with potential effects on dimerization some of which were interrogated experimentally<sup>28</sup>. (b) Structural model of a SOXE dimer on a palindromic DNA sequence with 5bp spacer. Models were prepared as described in Palasingam *et al.*<sup>56</sup> using structural coordinates from protein data bank entry 3f27 as template (<http://www.sciencedirect.com/science/article/pii/S0022283609003635>). DNA is shown as gray surface and the HMG boxes of two juxtaposed SOXE molecules as black cartoon. Residues varying amongst HMG boxes of SOXE proteins are shown as ball-and-stick. The DIM domains of unknown structure are schematically depicted with dotted lines.

specification was insensitive to dosage (Supplemental Fig. 6;  $n = 3/3$ ). Thus, vertebrate SOX9 appears to be functionally divergent from AmphiSOXE in influencing glial versus neuronal differentiation.

**Divergence of SOXE N- and C-terminal domains enabled glial versus neuronal differentiation bias.** To examine which protein domains may be responsible for this differential activity, we generated chimeric

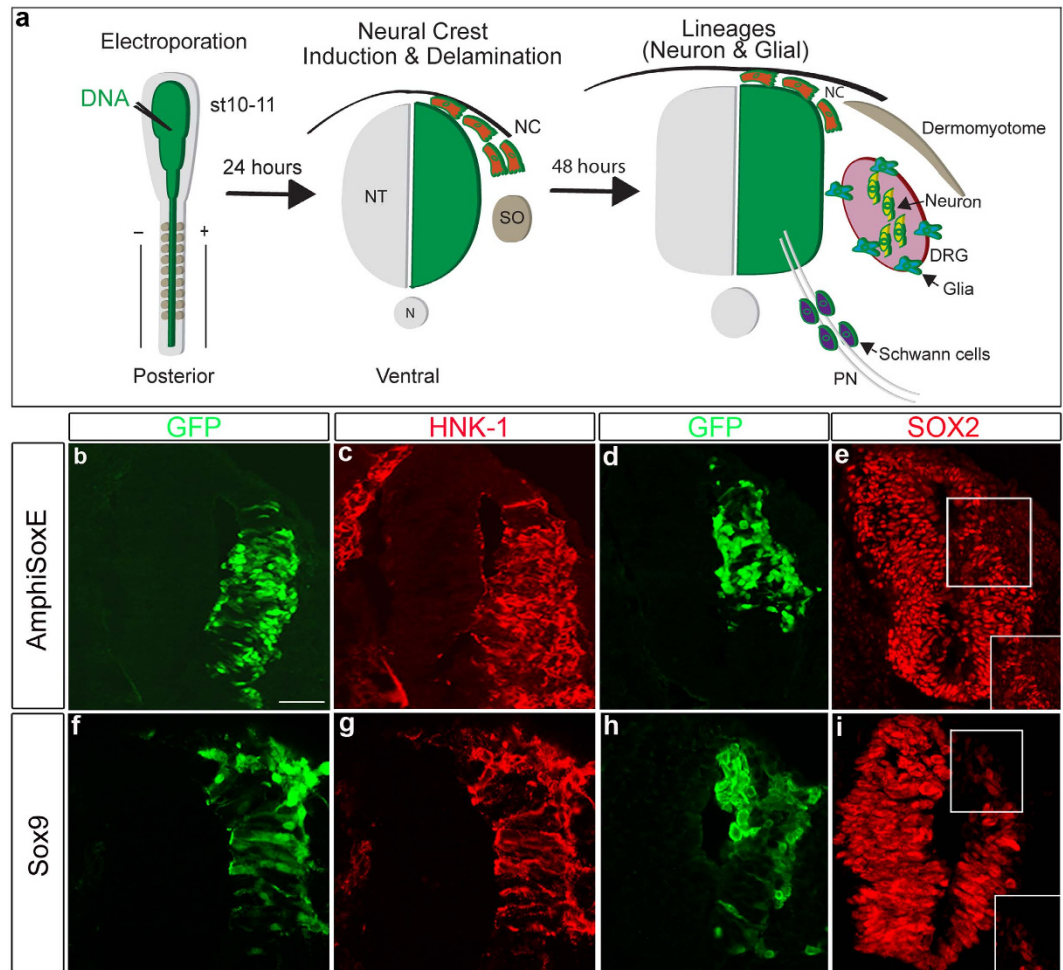


**Figure 3. AmphiSOXE and SOX9 show no difference in DNA binding and homodimerization.** (a) *CD-Rap* (5 bp spacer)<sup>57</sup> and 4 bp spacer elements used for dimerization EMSA test, the black and bold characters show the core SOX binding sites. (b,c) Gel images showing AmphiSOXE and SOX9 binding to the two DNA elements with HMG box constructs (b) and NHMG constructs (c). (d) Cooperativity factors for homodimerization were estimated for the indicated proteins and *CD-Rap* and *Zero* DNA elements. Values are only calculated when the fractional contribution of each of the 3 bands is at least 5%.

proteins between the N-terminus, HMG-box and C-terminal transactivation domains of SOX9 and AmphiSOXE and compared their activities (Fig. 6a). N-, HMG or C-terminal domains were designated 9 when derived from SOX9 or E when derived from AmphiSOXE. All the chimeric proteins can induce HNK1 (Supplemental Fig. 7). Each combination (9-E-9; E-9-9; 9-9-E) behaved like SOX9, resulting in differentiation into glial cells but not neurons (Fig. 6k–m, q–s, t–v; Supplemental Fig. 5g,h', k,l'; m–o'). However, NC expressing chimeric protein containing the SOX9 HMG box with both the N- and C- domains of AmphiSOXE (E-9-E), differentiated into both neurons and glial (Fig. 6n–p; Supplemental Fig. 5i,j'). These results suggest that N- and C-terminal domains of AmphiSOXE but not the HMG box are responsible for the different influence on neuronal and glial differentiation.

**Transactivation activity of AmphiSOXE is weak compared with SOX9.** Accordingly, we hypothesized that alterations in the ancestral SOXE N-terminus and transactivation domain may have resulted in differences in transactivation activity that could affect lineage decisions. To test this, we compared the transactivation activity of AmphiSOXE with SOX9 *in vivo* using *Sox2*- and *Sox10*-regulatory regions to drive luciferase (*luc*) reporters known to be active in the DRG<sup>33,34</sup>. Consistent with the fact that *Sox2* expression is initiated in SOX9<sup>+</sup> emigrating NCCs<sup>35</sup> and levels of SOX2 expression are known to influence neuronal differentiation<sup>31</sup>, we found that SOX9 stimulated a marked increase of *Sox2-luc* reporter activity compared to AmphiSOXE, SOXE-9-E and vector control (Fig. 7a). These data suggest that AmphiSOXE is a weaker activator of *Sox2* transcription, while SOX9 may inhibit neuronal differentiation by activating high levels of *Sox2* expression (*Sox2*<sup>High</sup>). To test this, we inhibited SOX2 function by overexpressing a *Sox2-EnR* construct<sup>36</sup> and *Sox9* together and examined the effect on neuronal differentiation in the DRG. By 48hpt, the majority of GFP<sup>+</sup> cells were localized in the periphery of the DRG rather than the core. Notably DRG size was unchanged and comparable numbers of HuC/D<sup>+</sup> cells formed in both transfected and control sides (Fig. 7b). These results suggest that blocking SOX2 function in SOX9 overexpressing cells can overcome the differentiation bias in the DRG. In contrast to differences in activation of *Sox2*, we observed similarities in activation of the *Sox10-luc* reporter by AmphiSOXE, SOX9 and SOXE-9-E (Fig. 7a), consistent with their similar potency in directing GFP<sup>+</sup> cells to differentiate into glial.

The basal jawless vertebrate, lamprey (*Petromyzom marinus*) (Fig. 1), has migrating NCCs and most of the NC derivatives. By ectopically expressing the lamprey *Sox9* ortholog, *SoxE3* (Supplemental Fig. 1), we found *SoxE3* functions similarly to SOX9, inhibiting sensory neuronal formation without affecting glial differentiation (Fig. 7c). *SoxE3* also transactivated *Sox10-luc* reporter activity to a similar degree to AmphiSOXE, SOX9 and SOXE-9-E, whereas *SoxE3* activated the *Sox2-luc* reporter at levels less than for SOX9 but higher than for



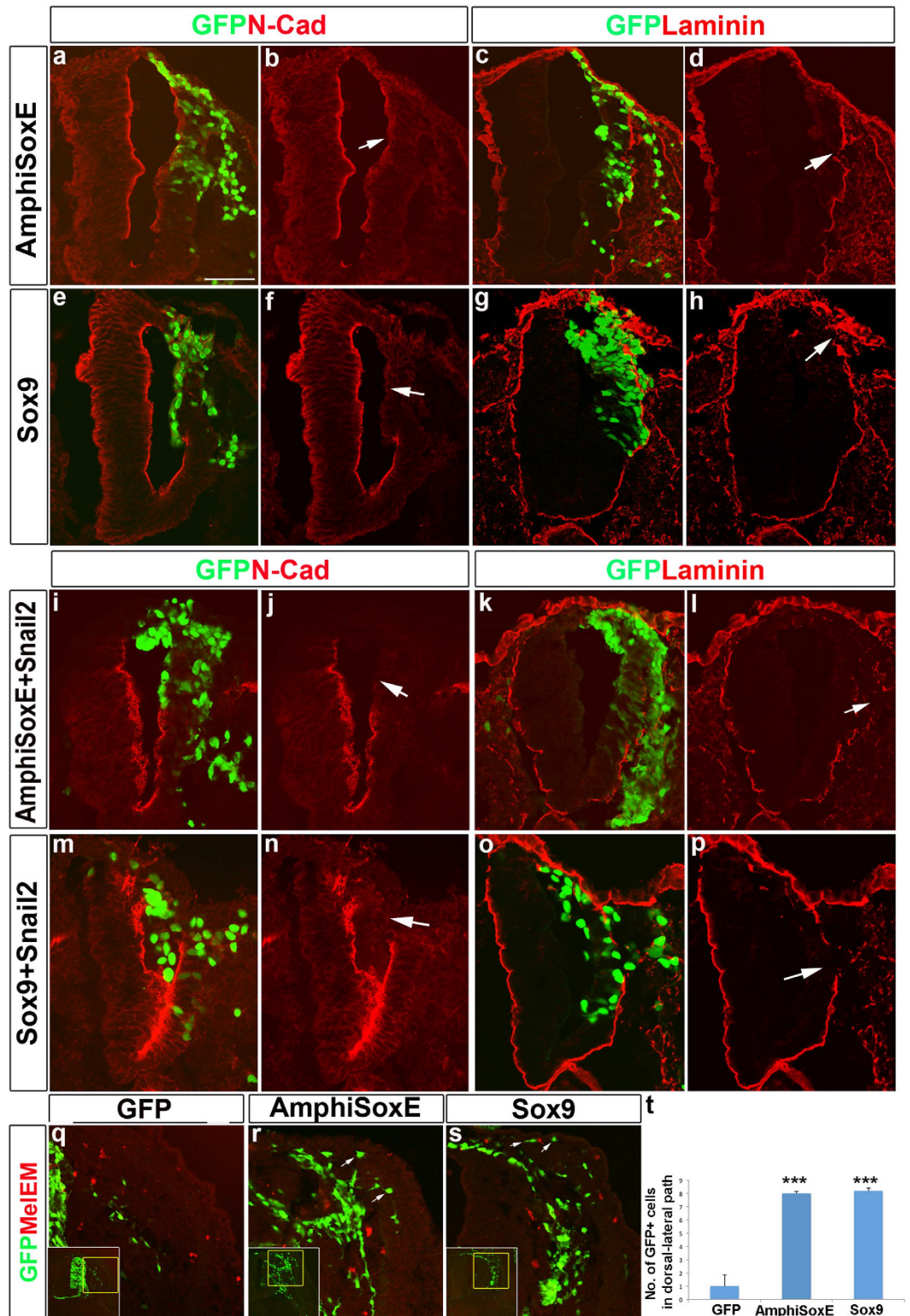
**Figure 4. Forced expression of AmphiSOXE induces neural crest-like cells in chick neural tube.** (a) Schematic representation of the experimental strategy. (b,f) Transverse sections of the neural tube electroporated with *AmphiSoxE/EGFP* or *Sox9/EGFP* analysed at 24 hours post-transfection. (c,g) Overexpression of *AmphiSoxE* or *Sox9* induces ectopic HNK-1 expression. (e,i) Insets show a magnified view of the neural tube region with SOX2 negative cells shown by the white box. The neuronal progenitor marker SOX2 is repressed in (d,e) *AmphiSOXE* or (h,i) *SOX9* expressing cells in a cell-autonomous manner. Scale bar: 50  $\mu\text{m}$ .

*AmphiSOXE* or *SOXE-9-E* (Fig. 7a). These data suggest that *SOXE3* functions similar to *SOX9* in directing NC cells to glial rather than neuron differentiation in the DRG.

Lastly, we examined the other two paralogs of vertebrate *SOXE*, *SOX8* and *SOX10*, which also possessed NC-inducing activity when overexpressed in chick neural tube<sup>13,37</sup>. Similar to *Sox9*, overexpression of *SOX8* or *SOX10* inhibited neuronal fate without affecting glial cell formation (Supplemental Fig. 8a–l) and exhibited more efficient in transactivating *Sox2-luc* reporter activity than *AmphiSOXE* (Supplemental Fig. 8m), whereas the activity in transactivating *Sox10-luc* reporter is similar among all tested *SOXE* proteins (Supplemental Fig. 8n). These findings suggest that divergent function in lineage determination has been fixed and retained in the *SoxE* duplicate paralogs following the split of jawed vertebrates from their ancestors.

## Discussion

NC formation is a complex multi-step process<sup>38</sup>, regulated by a hierarchical NC gene regulatory network (NC-GRN) in vertebrates. Studies in lamprey reveal extensive functional conservation of the NC-GRN to the base of vertebrates<sup>39</sup>. In the basal chordate, amphioxus, homologues of vertebrate NC specifier genes are present in the genome but are not expressed at the neural plate border region, with the exception of *AmphiSnail*<sup>26</sup>. Recent data suggest that acquisition of *SoxE* expression in NCCs may have been driven by evolution of new *cis*-regulatory sequences<sup>23</sup>. In addition to regulatory changes, our data suggest that protein changes between *AmphiSOXE* and vertebrate *SOXE* may be critical. While *AmphiSOXE*, *SOX9* and even the *Drosophila* ortholog *SOX100B* can induce NC traits at the expense of neural cell fate<sup>24</sup>, there are clear differences with respect to lineage specification. The domains largely responsible for this differential activity appear to be located in the N- and C-terminus, consistent with the finding that *AmphiSOXE* is a weaker transactivator of *Sox2* when compared with *SOX9* (Fig. 8). Previous studies showed that *Sox2* expression is initiated in early migrating NCCs, maintained in the periphery of the DRG and later restricted to the satellite glial<sup>40</sup>. Overexpression of *SOX2* in trunk NC culture inhibited both



**Figure 5.** Co-expression of *AmphiSoxE* and *Snail2* causes disruption of apical-basal polarity of neuroepithelial cells. (a,b,e,f) Transfection of *AmphiSoxE* or *Sox9* alone: Apical N-CAD expression is reduced as indicated by white arrows. (c,d,g,h) Most of the transfected cells are retained in the neural tube except for the dorsal and medial regions where basal LAMININ expression is reduced (white arrows). (i,j,k,l) Cotransfection of *AmphiSoxE* and *Snail2* assayed at 24 hours post-transfection. (i,j) The apical N-CADHERIN expression is lost (white arrow). (k,l) Many cells have delaminated from the neural tube and basal LAMININ expression is lost (white arrow). (m–p) A similar observation is seen in the neural tube transfected with *Sox9* and *Snail2*, with loss of N-CAD and LAMININ expression (white arrows). Scale bar: 100  $\mu$ m. Transverse section of the neural tube electroporated with EGFP (q), *AmphiSoxE/EGFP* (r) and *Sox9/EGFP* analyzed at 48 hours post-transfection. Early onset of migration onto the dorsolateral migratory pathway is observed in *AmphiSoxE/EGFP*<sup>+</sup> (r) or *Sox9/*

EGFP<sup>+</sup> cells(s) as indicated by the white arrows, but not observed for control EGFP<sup>+</sup> cells (q). These dorsolateral migrating cells do not express the melanocyte marker, MeLEM (r,s). Insets show a low magnified view of panels q-s with yellow boxes indicating the high magnification of the region where MeLEM immunofluorescence is shown in each panel. (t) Quantification of the number of EGFP<sup>+</sup>, *AmphiSoxE*/EGFP<sup>+</sup> and *Sox9*/EGFP<sup>+</sup> cells migrating via the dorsolateral route. nt, neural tube. \*\*\*p < 0.001 as compared to EGFP control. Scale bar: 20 μm.

neuronal and glial differentiation<sup>40</sup>, whereas downregulation of *Sox2* by shRNA in migrating NCCs or *Wnt1-Cre* mediated knockout of *Sox2* in mouse NC lineage resulted in reduced sensory neurons formation in the DRG<sup>31</sup>. Since SOX2 expression and function was genetically manipulated in uncommitted NC progenitors, the findings from these studies indicate an essential requirement for optimal SOX2 dosage in maintaining bipotent state of NC progenitor in which either excessive or below the optimal level of SOX2 could lead to defects in lineage differentiation. Once NCCs reach the periphery of the DRG, SOX2 and SOX10 are co-expressed in bipotential neural progenitors, which subsequently segregate into SOX2<sup>low</sup>/SOX10<sup>-</sup> neurons and SOX2<sup>high</sup>/SOX10<sup>+</sup> glial cells<sup>30,31</sup>. This suggests that downregulation of *Sox2* is required for sensory neuron differentiation consistent with its role in the developing chick spinal cord<sup>36</sup> while SOX2<sup>high</sup> may promote satellite glial differentiation similar to its function in the oligodendrocyte lineage<sup>41</sup>. In agreement with this, promotion of higher *Sox2* expression by SOX9 compared to *AmphiSOXE*, correlated with a bias towards glial rather than neuronal differentiation, suggesting that proper levels of SOX2 are required for acquisition of neuronal fate within the DRG. This is further supported by the ability of SOX2-EnR protein to abolish the effect of SOX9 over-expression. Competition between the ectopic SOX2-EnR repressor and the induced endogenous SOX2 for transcriptional targets could lower effective SOX2 transactivation levels to permit neuronal differentiation. Therefore, overexpression of *Sox9* exhibited two distinct effects on *Sox2* expression at two different stages of NC development, induction (repression of *Sox2*) and differentiation into glial cells (activation of *Sox2*) likely through association with stage specific cofactors, which remain to be identified.

As lamprey SOXE3 behaved like SOX9 in promoting glial over neuronal fate, we speculate that the ability of all SOXE proteins to direct NC differentiation into glial may have been acquired in basal vertebrates. In contrast, although another NC specifier, vertebrate *FoxD3*, was able to induce ectopic NCCs when overexpressed in chick neural tube, neither Amphioxus *FoxD* (*AmphiFoxD3*) nor lamprey *FoxD* paralogs exhibited this activity. Domain mapping studies further revealed that N-terminal sequence is critical for the NCC differentiation-inducing activity of *FoxD3*<sup>42</sup>. Altogether, these results suggest that co-option of *SoxE* into NC-GRN occurred before the genome duplication, whereas among the five known vertebrate *FoxD* paralogs, only *FoxD3* is expressed in the NC and might have been recruited slightly later than other NC specifiers, after the genome duplication. In addition, duplicated paralogs of *SoxE* and *FoxD* family acquire neofunctionalization properties at the level of NC lineage determination and specification, respectively via changes of specific protein sequences. The reasons underlying the distinct onset of neofunctionalization in both NC specifiers are not clear. Since the C-terminal motif of *FoxD3* essential for the differentiation of dorsal mesodermal in *Xenopus* embryos is conserved in *AmphiFoxD*<sup>42,43</sup>, it is tempting to speculate that ancestral *FoxD* might be important in mesoderm development, resulting in late onset of acquiring the novel ability to induce NCCs.

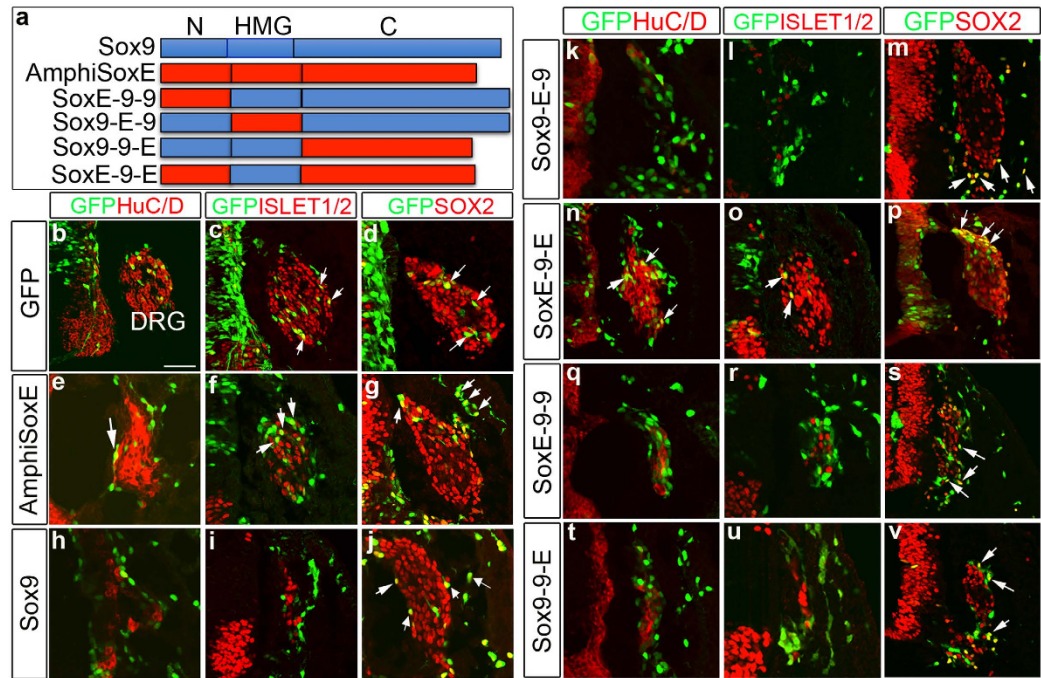
Our data also imply that acquisition of a stronger transactivation potential in vertebrate SOX9 may have been advantageous during evolution of transcriptional regulation to ensure activation of SOX9 target genes at appropriate levels for tissue development. Consistent with the important contribution of evolving a strongly transactivating SOXE factor for the neural crest in vertebrates, it is noted that heterozygous mutations in the transactivation domain of SOX9 may lead to differential severity of Campomelic Dysplasia because of hypomorphic effects<sup>44,45</sup>. It is also notable that some mutations in the transactivation domain of SOX10 can lead to milder Waardenberg syndrome phenotypes than those that cause loss of function<sup>46,47</sup>. Our findings indicate that a weaker transactivation activity leads to loss of differential differentiation fate. Thus, we speculate that the two genome-wide duplications in the vertebrate lineage allowed duplication and divergence of *SoxE* genes, conferring additional later functions that may be essential for controlling proper migration and cell lineage segregation. Altogether, our results suggest that acquisition of SOXE to the neural plate border (through a regulatory change) accompanied by duplication and divergence of its protein domains (via advantageous mutations) that strengthened its transactivation ability, allowed elaboration of NC derivatives into peripheral ganglia and perhaps also the craniofacial skeleton (Fig. 8).

## Materials and Methods

All experimental protocols were approved by the Committee on the Use of Live Animals in Teaching and Research (CULATR) in the University of Hong Kong. All the methods were carried out in accordance with the guidelines approved by the Committee on the Use of Live Animals in Teaching and Research (CULATR) in the University of Hong Kong.

**Protein production.** The SOX9-HMG and SOX9-NHMG proteins were prepared using a bacterial pET-vector based expression system as described<sup>29</sup>. *AmphiSOXE*-HMG and the *AmphiSOXE*-NHMG were PCR amplified using primers with overhangs for attB sites and an N-terminal tobacco etch virus (TEV) protease cleavage site and cloned into a pDEST-hisMBP expression plasmid<sup>48</sup> using the Gateway BP and LR cloning system (Invitrogen). The expression plasmids were transferred into the BL21(DE3)pLySs *E. coli* strain and the proteins were expressed and purified using established methods<sup>29</sup>.



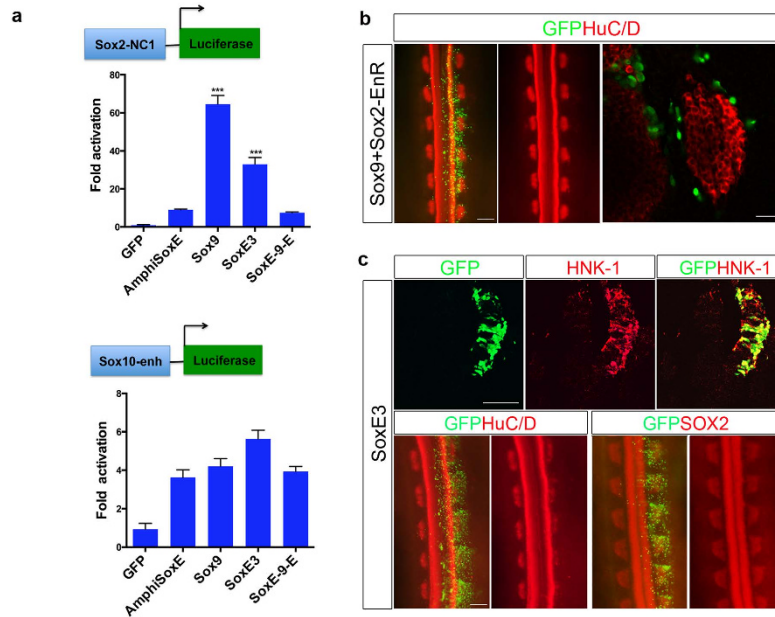


**Figure 6. Prolonged expression of *AmphiSoxE* does not affect NC-derived lineages specification at trunk level.** (a) Schematic diagram of the constructs used for the analysis. Transverse sections of the neural tube electroporated with EGFP (b–d), *AmphiSoxE/EGFP* (e–g), *Sox9/EGFP* (h–j), *Sox9-E-9/EGFP* (k–m), *Sox9-E-9/EGFP* (n–p), *Sox9-9-9/EGFP* (q–s) and *Sox9-9-E/EGFP* (t–v) and analysed after 48 hours. (b–d) Control EGFP<sup>+</sup> cells migrate to dorsal root ganglia (DRG) and express markers for neuron, HuC/D and ISLET1/2 and glia, SOX2. *AmphiSoxE*<sup>+</sup> (e,f) and *SoxE-9-E*<sup>+</sup> cells (n,o) express HuC/D and ISLET1/2 in DRG whereas expression of these neuronal markers are lost in embryos transfected with *Sox9* (h,i), *Sox9-E-9* (k,l), *Sox9-9-9* (q,r) and *Sox9-9-E* (t,u). SOX2 expression in DRG is unaffected by ectopic expression of each construct (g,j,m,p,s,v). The white arrows indicate the transfected cells co-expressing either ISLET1/2 or SOX2. Scale bar: 50 μm.

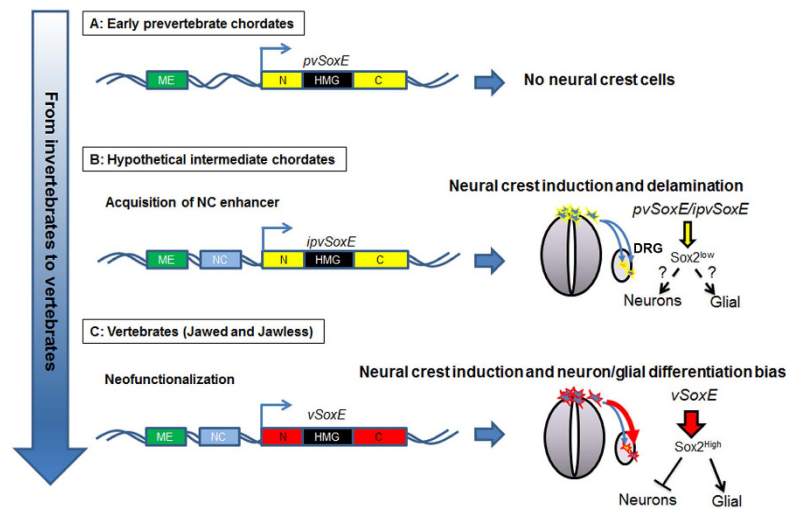
**EMSA.** EMSAs were carried out as reported recently<sup>29</sup>. In brief, dsDNA probes were prepared by combining forward strands with 5' cy5 label and unlabeled reverse strands (Life Technologies) in 1X annealing buffer (20 mM Tris–HCl, pH 8.0; 50 mM MgCl<sub>2</sub>; 50 mM KCl) followed by heating to 95 °C for 5 min and cooling to 4 °C at 1 °C/min. For each EMSA reaction, 50 nM dsDNA probes were mixed with varying concentrations of protein in 1X EMSA buffer (10 mM Tris–HCl pH 8.0, 0.1 mg/ml bovine serum albumin, 50 μM ZnCl<sub>2</sub>, 100 mM KCl, 10% (v/v) glycerol, 0.1% (v/v) Igepal CA630 and 2 mM beta-mercaptoethanol). Samples were loaded onto 12% 1X Tris-glycine (TG, 25 mM Tris–HCl pH 8.0, 192 mM glycine) native PAGE gels after incubating at 4 °C in the dark for 4 h. The gel were run at 200 V for 40 min in 1X TG buffer in the cold room and imaged with a Typhoon FLA-7000 PhosphorImager (FUJIFILM). The intensities were detected with the Image Quant software (GE Healthcare) and cooperativity factors were calculated using previously reported equations<sup>49</sup>.

**In ovo electroporation and expression vectors.** Fertilized white Leghorn eggs were obtained from Tin Hang Technology Co. Limited and incubated in a humidified incubator at 38 °C. Embryos were staged as described previously<sup>50</sup>. *In ovo* electroporation was carried out as described<sup>13</sup>. Expression constructs were generated in the pCAGGS vector for *mouse Sox8*, *Sox9*, *Sox10*, *AmphiSoxE* and *LampreySoxE3*. Chimeric *SoxE-9-9*, *Sox9-E-9*, *Sox9-9-E* and *SoxE-9-E* constructs were generated by DNA recombineering using DY380 bacteria<sup>51</sup>. For electroporation, each *Sox* construct was co-injected with an EGFP expression construct into the lumen of Hamburger Hamilton stage 10–12 neural tubes. Electroporation was carried out using a BTX electroporator delivering five 50-ms pulses of 30V across the neural tube. Transfected embryos were incubated for 24 and 48 h before processing. *Sox2-Luc* reporter was generated by cloning the *Sox2*-NC1 enhancer<sup>33</sup> into the pGL3-promoter vector (Promega). *Sox10-Luc* reporter (pE1B-C-F/*Sox10*-MCS4) was purchased in Addgen (#20243)<sup>34</sup>. *Sox2-EnR* was a gift of Prof. Jonas Muhr<sup>36</sup>.

**In situ hybridisation and immunohistochemistry.** Transfected embryos were fixed for an hour at 4 °C in 4% paraformaldehyde in 0.1 M phosphate buffer (PB), cryoprotected with 30% sucrose in PB and cryo-sectioned. *In situ* hybridization on cryosections was performed as described<sup>13</sup>, using probes for chick, *Sox10*<sup>52</sup>, *Wnt3a* (a gift from N. Itasaki) and *FoxD3* chick expressed sequence tag clone. Immunohistochemical staining was performed using the antibodies against: green fluorescent protein (GFP) (Molecular Probes), Laminin (Sigma), HNK-1 (Becton Dickinson), HuC/D (Invitrogen), SOX2 (Neuromics), *Islet1/2* (DSHB), P0 (DSHB) and MeEM



**Figure 7. AmphiSOXE is less effective than vertebrate SOX9 in directing *Sox2* transcription.** (a) In ovo *Sox2-NC1*- and *Sox10* enhancer (*Sox10-enh*) driven luciferase reporter assays: pCAGGS, *AmphiSoxE* (SOXE), *Sox9*, *SoxE3* and *SoxE-9-E* were cotransfected into the neural tube with a *Sox2-NC1*- or *Sox10*-enhancer (enh)-driven luciferase reporter and a Renilla control plasmid in neural tube. The relative luciferase activity is compared with pCAGGS control. AmphiSOXE or SOXE-9-E protein weakly transactivates the *Sox2-NC1* luciferase reporter whereas SOX9 and SOXE3 yields exceptionally high and moderate activation of the reporter respectively. By contrast, AmphiSOXE, SOX9 or SOXE-9-E exhibits similar transactivation capacities on the *Sox10-enh* luciferase reporter. (b) Immunofluorescence on embryos transfected with *Sox9 + Sox2-EnR* showing that the transfected cells migrate to the periphery of DRG without affecting HuC/D expression compared to the untransfected side (n = 6). Scale Bar: 50 μm. (c) Overexpression of *SoxE3* results in ectopic HNK-1 expression and marked reduction of HuC/D expression in the DRG as compared with the untransfected side, while SOX2 expression remains altered. \*\*\*p < 0.001 as compared to AmphiSOXE (SOXE).



**Figure 8. Linking neofunctionalization of SoxE genes to vertebrate neural crest traits.** (A) An ancient prevertebrate chordate (pv-chordate) *pvSoxE* gene is expressed in mesoderm/endoderm under the control of a mesoderm-endodermal enhancer (ME), but lacks a regulatory element(s) that directs expression in the neural plate border and is a weak transactivator. As a consequence the pv-chordate lacks neural crest cells. (B) In a hypothetical intermediate pv-chordate/vertebrate (*ipv*), *ipvSoxE* has acquired a neural crest (NC) enhancer, which directs expression in the neural plate border resulting in neural crest formation (blue). However, the *ipv* SOXE remains as a weak transactivator able to induce low levels of *Sox2* expression without affecting specification of neuronal and glial lineages specification in the DRG. (C) Vertebrate *Sox9* evolves through alteration of protein sequences in *pvSOXE* N- and C- terminal domains, which confer with high transactivation capacity to activate high *Sox2* expression resulting in glial fate and suppression of neuronal formation in the DRG.

(DSHB). Immunofluorescence images were photographed using a Zeiss LSM700 confocal microscope in the Faculty Core Facility, Li Ka Shing Faculty of Medicine, the University of Hong Kong.

**Luciferase assay.** *Sox2-NC1* or *Sox10-MCS4* enhancer driven luciferase reporters were mixed with Renilla and the relevant *Sox* expression construct and electroporated into the lumen of the chick neural tube. The trunk parts of embryos were collected 1–2 days later, lysed and assayed following the Promega's manual. Statistical analysis was performed using a Student's *t*-test.

## References

- Green, S. A., Simoes-Costa, M. & Bronner, M. E. Evolution of vertebrates as viewed from the crest. *Nature* **520**, 474–482, doi: 10.1038/nature14436 (2015).
- Jeffery, W. R., Strickler, A. G. & Yamamoto, Y. Migratory neural crest-like cells form body pigmentation in a urochordate embryo. *Nature* **431**, 696–699, doi: 10.1038/nature02975 (2004).
- Abitua, P. B., Wagner, E., Navarrete, I. A. & Levine, M. Identification of a rudimentary neural crest in a non-vertebrate chordate. *Nature* **492**, 104–107, doi: 10.1038/nature11589 (2012).
- 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).
- Delsuc, F., Brinkmann, H., Chourrout, D. & Philippe, H. Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature* **439**, 965–968, doi: 10.1038/nature04336 (2006).
- Holland, L. Z., Laudet, V. & Schubert, M. The chordate amphioxus: an emerging model organism for developmental biology. *Cell Mol Life Sci* **61**, 2290–2308, doi: 10.1007/s00018-004-4075-2 (2004).
- Dehal, P. *et al.* The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. *Science* **298**, 2157–2167, doi: 10.1126/science.1080049 (2002).
- Putnam, N. H. *et al.* The amphioxus genome and the evolution of the chordate karyotype. *Nature* **453**, 1064–1071, doi: 10.1038/nature06967 (2008).
- McCauley, D. W. & Bronner-Fraser, M. Importance of SoxE in neural crest development and the evolution of the pharynx. *Nature* **441**, 750–752, doi: 10.1038/nature04691 (2006).
- McCauley, D. W. & Bronner-Fraser, M. Conservation and divergence of BMP2/4 genes in the lamprey: expression and phylogenetic analysis suggest a single ancestral vertebrate gene. *Evolution & development* **6**, 411–422, doi: 10.1111/j.1525-142X.2004.04054.x (2004).
- Yan, Y. L. *et al.* A pair of Sox: distinct and overlapping functions of zebrafish *sox9* co-orthologs in craniofacial and pectoral fin development. *Development* **132**, 1069–1083, doi: 10.1242/dev.01674 (2005).
- Kluver, N., Kondo, M., Herpin, A., Mitani, H. & Schartl, M. Divergent expression patterns of Sox9 duplicates in teleosts indicate a lineage specific subfunctionalization. *Development genes and evolution* **215**, 297–305, doi: 10.1007/s00427-005-0477-x (2005).
- Cheung, M. & Briscoe, J. Neural crest development is regulated by the transcription factor Sox9. *Development* **130**, 5681–5693, doi: 10.1242/dev.00808 (2003).
- Sakai, D., Suzuki, T., Osumi, N. & Wakamatsu, Y. Cooperative action of Sox9, Snail2 and PKA signaling in early neural crest development. *Development* **133**, 1323–1333, doi: 10.1242/dev.02297 (2006).
- Kellerer, S. *et al.* Replacement of the Sox10 transcription factor by Sox8 reveals incomplete functional equivalence. *Development* **133**, 2875–2886, doi: 10.1242/dev.02477 (2006).
- Cheung, M. *et al.* The transcriptional control of trunk neural crest induction, survival, and delamination. *Developmental cell* **8**, 179–192, doi: 10.1016/j.devcel.2004.12.010 (2005).
- Liu, J. A. *et al.* Phosphorylation of Sox9 is required for neural crest delamination and is regulated downstream of BMP and canonical Wnt signaling. *Proc Natl Acad Sci USA* **110**, 2882–2887, doi: 10.1073/pnas.1211747110 (2013).
- O'Donnell, M., Hong, C. S., Huang, X., Delnicki, R. J. & Saint-Jeannet, J. P. Functional analysis of Sox8 during neural crest development in *Xenopus*. *Development* **133**, 3817–3826, doi: 10.1242/dev.02558 (2006).
- Sock, E., Schmidt, K., Hermanns-Borgmeyer, I., Bosl, M. R. & Wegner, M. Idiopathic weight reduction in mice deficient in the high-mobility-group transcription factor Sox8. *Molecular and cellular biology* **21**, 6951–6959, doi: 10.1128/MCB.21.20.6951-6959.2001 (2001).
- Mori-Akiyama, Y., Akiyama, H., Rowitch, D. H. & de Crombrughe, B. Sox9 is required for determination of the chondrogenic cell lineage in the cranial neural crest. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 9360–9365, doi: 10.1073/pnas.1631288100 (2003).
- Sonnenberg-Riethmacher, E. *et al.* Development and degeneration of dorsal root ganglia in the absence of the HMG-domain transcription factor Sox10. *Mechanisms of development* **109**, 253–265 (2001).
- Southard-Smith, E. M., Kos, L. & Pavan, W. J. Sox10 mutation disrupts neural crest development in Dom Hirschsprung mouse model. *Nature genetics* **18**, 60–64, doi: 10.1038/ng0198-60 (1998).
- Jandzik, D. *et al.* Evolution of the new vertebrate head by co-option of an ancient chordate skeletal tissue. *Nature* **518**, 534–537, doi: 10.1038/nature14000 (2015).
- Cossais, F. *et al.* Replacement of mouse Sox10 by the *Drosophila* ortholog Sox100B provides evidence for co-option of SoxE proteins into vertebrate-specific gene-regulatory networks through altered expression. *Developmental biology* **341**, 267–281, doi: 10.1016/j.ydbio.2010.01.038 (2010).
- Yu, J. K. The evolutionary origin of the vertebrate neural crest and its developmental gene regulatory network—insights from amphioxus. *Zoology* **113**, 1–9, doi: 10.1016/j.zool.2009.06.001 (2010).
- Yu, J. K., Meulemans, D., McKeown, S. J. & Bronner-Fraser, M. Insights from the amphioxus genome on the origin of vertebrate neural crest. *Genome research* **18**, 1127–1132, doi: 10.1101/gr.076208.108 (2008).
- Peirano, R. I. & Wegner, M. The glial transcription factor Sox10 binds to DNA both as monomer and dimer with different functional consequences. *Nucleic acids research* **28**, 3047–3055 (2000).
- Schlierf, B., Ludwig, A., Klenovsek, K. & Wegner, M. Cooperative binding of Sox10 to DNA: requirements and consequences. *Nucleic acids research* **30**, 5509–5516 (2002).
- Huang, Y. H., Jankowski, A., Cheah, K. S., Prabhakar, S. & Jauch, R. SOXE transcription factors form selective dimers on non-compact DNA motifs through multifaceted interactions between dimerization and high-mobility group domains. *Scientific reports* **5**, 10398, doi: 10.1038/srep10398 (2015).
- Wakamatsu, Y., Maynard, T. M. & Weston, J. A. Fate determination of neural crest cells by NOTCH-mediated lateral inhibition and asymmetrical cell division during gangliogenesis. *Development* **127**, 2811–2821 (2000).
- Cimadamore, F. *et al.* Human ESC-derived neural crest model reveals a key role for SOX2 in sensory neurogenesis. *Cell stem cell* **8**, 538–551, doi: 10.1016/j.stem.2011.03.011 (2011).
- Cui, S. & Goldstein, R. S. Early markers of neuronal differentiation in DRG: islet-1 expression precedes that of Hu. *Brain research. Developmental brain research* **121**, 209–212 (2000).

33. Uchikawa, M., Ishida, Y., Takemoto, T., Kamachi, Y. & Kondoh, H. Functional analysis of chicken Sox2 enhancers highlights an array of diverse regulatory elements that are conserved in mammals. *Developmental cell* **4**, 509–519 (2003).
34. Antonellis, A. *et al.* Identification of neural crest and glial enhancers at the mouse Sox10 locus through transgenesis in zebrafish. *PLoS genetics* **4**, e1000174, doi: 10.1371/journal.pgen.1000174 (2008).
35. Adameyko, I. *et al.* Sox2 and Mitf cross-regulatory interactions consolidate progenitor and melanocyte lineages in the cranial neural crest. *Development* **139**, 397–410, doi: 10.1242/dev.065581 (2012).
36. Bylund, M., Andersson, E., Novitsch, B. G. & Muhr, J. Vertebrate neurogenesis is counteracted by Sox1-3 activity. *Nature neuroscience* **6**, 1162–1168, doi: 10.1038/nn1131 (2003).
37. McKeown, S. J., Lee, V. M., Bronner-Fraser, M., Newgreen, D. F. & Farlie, P. G. Sox10 overexpression induces neural crest-like cells from all dorsoventral levels of the neural tube but inhibits differentiation. *Dev Dyn* **233**, 430–444, doi: 10.1002/dvdy.20341 (2005).
38. Bronner, M. E. & LeDouarin, N. M. Development and evolution of the neural crest: an overview. *Developmental biology* **366**, 2–9, doi: 10.1016/j.ydbio.2011.12.042 (2012).
39. Sauka-Spengler, T., Meulemans, D., Jones, M. & Bronner-Fraser, M. Ancient evolutionary origin of the neural crest gene regulatory network. *Developmental cell* **13**, 405–420, doi: 10.1016/j.devcel.2007.08.005 (2007).
40. Wakamatsu, Y., Endo, Y., Osumi, N. & Weston, J. A. Multiple roles of Sox2, an HMG-box transcription factor in avian neural crest development. *Dev Dyn* **229**, 74–86, doi: 10.1002/dvdy.10498 (2004).
41. Hoffmann, S. A. *et al.* Stem cell factor Sox2 and its close relative Sox3 have differentiation functions in oligodendrocytes. *Development* **141**, 39–50, doi: 10.1242/dev.098418 (2014).
42. Ono, H., Kozmik, Z., Yu, J. K. & Wada, H. A novel N-terminal motif is responsible for the evolution of neural crest-specific gene-regulatory activity in vertebrate FoxD3. *Dev Biol* **385**, 396–404, doi: 10.1016/j.ydbio.2013.11.010 (2014).
43. Yaklichkin, S., Steiner, A. B., Lu, Q. & Kessler, D. S. FoxD3 and Grg4 physically interact to repress transcription and induce mesoderm in *Xenopus*. *J Biol Chem* **282**, 2548–2557, doi: 10.1074/jbc.M607412200 (2007).
44. Bi, W., Deng, J. M., Zhang, Z., Behringer, R. R. & de Crombrughe, B. Sox9 is required for cartilage formation. *Nat Genet* **22**, 85–89, doi: 10.1038/8792 (1999).
45. Pop, R., Zaragoza, M. V., Gaudette, M., Dohrmann, U. & Scherer, G. A homozygous nonsense mutation in SOX9 in the dominant disorder campomelic dysplasia: a case of mitotic gene conversion. *Human genetics* **117**, 43–53, doi: 10.1007/s00439-005-1295-y (2005).
46. Chaoui, A. *et al.* Identification and functional analysis of SOX10 missense mutations in different subtypes of Waardenburg syndrome. *Human mutation* **32**, 1436–1449, doi: 10.1002/humu.21583 (2011).
47. Jung, H. J., Jin, S. A., Choi, S. J., Lee, S. C. & Yun, S. J. A de novo SOX10 mutation in a patient with Waardenburg syndrome type IV. *Journal of the American Academy of Dermatology* **68**, e177–e178, doi: 10.1016/j.jaad.2012.10.021 (2013).
48. Nallamsetty, S. & Waugh, D. S. A generic protocol for the expression and purification of recombinant proteins in *Escherichia coli* using a combinatorial His6-maltose binding protein fusion tag. *Nature protocols* **2**, 383–391, doi: 10.1038/nprot.2007.50 (2007).
49. BabuRajendran, N. *et al.* Structure of Smad1 MH1/DNA complex reveals distinctive rearrangements of BMP and TGF-beta effectors. *Nucleic acids research* **38**, 3477–3488, doi: 10.1093/nar/gkq046 (2010).
50. Hamburger, V. & Hamilton, H. L. A series of normal stages in the development of the chick embryo. *Journal of morphology* **88**, 49–92 (1951).
51. Zhang, Y., Buchholz, F., Muyrers, J. P. & Stewart, A. F. A new logic for DNA engineering using recombination in *Escherichia coli*. *Nature genetics* **20**, 123–128, doi: 10.1038/2417 (1998).
52. Cheng, Y., Cheung, M., Abu-Elmagd, M. M., Orme, A. & Scotting, P. J. Chick sox10, a transcription factor expressed in both early neural crest cells and central nervous system. *Brain research. Developmental brain research* **121**, 233–241 (2000).
53. Schwarz, R. & Dayhoff, M. *Matrices for detecting distant relationships*. (National Biomedical Research Foundation, 1979).
54. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular biology and evolution* **33**, 1870–1874, doi: 10.1093/molbev/msw054 (2016).
55. M., N. & S., K. *Molecular Evolution and Phylogenetics*. (Oxford University Press, New York, 2000).
56. Palasingam, P. *et al.* The Structure of Sox17 Bound to DNA Reveals a Conserved Bending Topology but Selective Protein Interaction Platforms. *J. Mol. Bio.* **388**, 619–630 (2003).
57. Sock, E. *et al.* Loss of DNA-dependent dimerization of the transcription factor SOX9 as a cause for campomelic dysplasia. *Human molecular genetics* **12**, 1439–1447 (2003).

## Acknowledgements

This work was supported by grants from the Research Grants Council and University Grants Council of Hong Kong (7337/01M), (AoE/M-04/04), (T12-708/12-N) and (T12C-714/14-R) to K.S.E.C., GRF\_17110715 to M.C. and NS086907 to M.E.B. We thank Robin Lovell-Badge and Patrick Tam for helpful discussion, and Hisato Kondoh and Masanori Uchikawa for providing the Sox2-NC1 construct. We thank Yogesh Srivastava (GIBH) for help with structural modeling and May Cheung for technical assistance. R.J. is supported by a 2013 MOST China-EU Science and Technology Cooperation Program (grant number 2013DFE33080), by the National Natural Science Foundation of China (grant number 31471238), a 100 talent award of the Chinese Academy of Sciences and a Science and Technology Planning Project of Guangdong Province, China (2014B030301058).

## Author Contributions

A.T. and M.C. designed and performed experiments, analyzed data, and wrote the manuscript. Y.-H.H. designed and performed experiments. R.J. designed experiments, analyzed data and wrote the manuscript. M.E.B. and K.S.E.C. designed experiments, analyzed data and wrote the manuscript.

## Additional Information

**Accession codes:** Accession numbers for the sequences used in this study are as follows: *Rattus norvegicus* (Sox8: NP\_001100458; Sox10: NP\_); *Mus musculus* (Sox8: NP\_035577; Sox9: NM\_035578; Sox10: NP\_035567); *Gallus gallus* (Sox8: NP\_990062; Sox9: NP\_989612; Sox10: NP\_990123); *Homo sapiens* (Sox8: NM\_055402; Sox9: NP\_000337; Sox10: NP\_990123); *Xenopus laevis* (Sox8: NP\_001083964; Sox9: NP\_001084276; Sox10: NP\_001082358); *Danio rerio* (Sox8: NP\_001020636; Sox9a: NP\_571718; Sox9b: AAH67133; Sox10: NP\_571950); *Petromyzon marinus* (SoxE1: AAW34332; SoxE2: ABC58684; SoxE3: ABC58685); *Eptatretus burgeri* (Sox9: BAG11536); *Ciona intestinalis* (SoxE: CAD58841); *Lytechinus variegatus* (SoxE: ABY40629); *Drosophila melanogaster* (Sox100B: NP\_651839); *Apis mellifera* (SoxE1: XP\_001122993; SoxE2: XP\_001122996) and *Nasonia vitripennis* (SoxE1: XP\_001604913; SoxE2: XP\_008213434).

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

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

**How to cite this article:** Tai, A. *et al.* SOXE neofunctionalization and elaboration of the neural crest during chordate evolution. *Sci. Rep.* **6**, 34964; doi: 10.1038/srep34964 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016