Published in final edited form as:

Genesis. 2013 July ; 51(7): 457–470. doi:10.1002/dvg.22403.

Gene regulatory evolution and the origin of macroevolutionary novelties: insights from the neural crest

Eric Van Otterloo^{1,*}, Robert A. Cornell¹, Daniel Meulemans Medeiros², and Aaron T. Garnett^{2,*}

¹Department of Anatomy and Cell Biology, University of Iowa, Iowa City, Iowa

²Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, Colorado

Abstract

The appearance of novel anatomic structures during evolution is driven by changes to the networks of transcription factors, signaling pathways, and downstream effector genes controlling development. The nature of the changes to these developmental gene regulatory networks (GRNs) is poorly understood. A striking test case is the evolution of the GRN controlling development of the neural crest (NC). NC cells emerge from the neural plate border (NPB) and contribute to multiple adult structures. While all chordates have a NPB, only in vertebrates do NPB cells express all the genes constituting the neural crest GRN (NC-GRN). Interestingly, invertebrate chordates express orthologs of NC-GRN components in other tissues, revealing that during vertebrate evolution new regulatory connections emerged between transcription factors primitively expressed in the NPB and genes primitively expressed in other tissues. Such interactions could have evolved by two mechanisms. First, transcription factors primitively expressed in the NPB may have evolved new DNA and/or cofactor binding properties (protein neofunctionalization). Alternately, cis-regulatory elements driving NPB expression may have evolved near genes primitively expressed in other tissues (cis-regulatory neofunctionalization). Here we discuss how gene duplication can, in principle, promote either form of neofunctionalization. We review recent published examples of interspecies gene-swap, or regulatory-element-swap, experiments that test both models. Such experiments have yielded little evidence to support the importance of protein neofunctionalization in the emergence of the NC-GRN, but do support the importance of novel cis-regulatory elements in this process. The NC-GRN is an excellent model for the study of gene regulatory and macroevolutionary innovation.

Keywords

Development; Evolution; Transcription; Neural Crest

Corresponding author: Aaron T. Garnett, Department of Ecology and Evolutionary Biology, Campus Box 334, University of Colorado, Boulder, CO 80309, Phone # 303-735-1984, Fax # 303-492-8699.

*Equal contribution

The origin of novel gene regulatory networks

Cellular differentiation, division, and movement during development are controlled by regulatory molecules such as growth factors, growth factor receptors, signal transduction molecules, and transcription factors. The net effect of these regulatory cascades is to stimulate the expression of genes encoding the enzymes and structural molecules that effect cell-specific functions and behaviors. These collections of genes encoding regulatory molecules and differentiation effectors, and the *cis*-regulatory DNA sequences mediating their cell-type specific expression, are called gene regulatory networks (GRNs) (Davidson et al., 2003). During evolution, mutations that alter GRNs are thought to drive changes in organismal morphology. Modest changes in morphology may be caused by slight changes in the GRN. For instance, a change in a *cis*-regulatory element which stabilizes the binding of a transcription factor might result in a small increase in expression of a gene. If the gene encodes a growth factor, cell division in the developing embryo might be prolonged and the size of the adult organ increased. However, macroevolutionary changes, like the emergence of a qualitatively new structure or cell type, presumably involves the formation of novel regulatory connections between genes, or even the assembly of entirely new GRNs. Understanding how novel regulatory connections are formed is thus essential to understanding the mechanistic bases of macroevolutionary novelties.

The neural crest (NC) is an excellent model for investigating the genomic changes associated with the emergence of a major macroevolutionary novelty. The NC is a population of embryonic precursor cells that arises from the border of the nascent central nervous system and contributes to a range of adult structures including the head skeleton, paired sensory organs, peripheral nervous system, and heart. Because a bona fide NC, and many of its adult derivatives, is found in all living vertebrates but in no invertebrates, it is considered a defining feature of the vertebrate subphylum. In addition to being a clear example of an important macroevolutionary novelty, recent work has convincingly linked the evolution of NC cells to the establishment of dozens of new gene regulatory interactions between neural border-expressed transcription factors and genes primitively expressed in other germ layers (Abitua et al., 2012; Meulemans and Bronner-Fraser, 2004; Meulemans and Bronner-Fraser, 2005; Meulemans et al., 2003; Van Otterloo et al., 2012; Yu et al., 2008a). Thus, the NC and its GRN lends itself to the question; ‘how do the novel gene regulatory connections that underlie major macroevolutionary innovations emerge?’

At the core of all GRNs are the physical interactions between transcription factors and their cognate *cis*-regulatory sequences. Many of the transcription factors essential for NC development have been identified, and, like most developmental regulators, they all have invertebrate homologs. Members of these NC-specifier gene families are deployed in the mesoderm and other parts of the ectoderm of invertebrate chordates (Yu et al., 2002; Yu et al., 2008a). However, in all living vertebrates, they are expressed in the neural border and migratory NC cells (Dottori et al., 2001; Southard-Smith et al., 1998), suggesting they were recruited to the NC early in vertebrate evolution. The recruitment of mesodermal genes to the NPB may have proven advantageous since many of the genes are involved in the epithelial to mesenchymal transition, allowing the NC to delaminate from the neural tube and migrate throughout the embryo.

In the context of this conserved developmental toolkit, one way novel gene regulatory interactions could arise is via mutations in NC-GRN transcription factors that allow them to bind new DNA sequences and/or protein cofactors (illustrated in Figure 1A). In the case of the NC, this scenario is especially attractive as two rounds of large-scale genome duplication occurred in the vertebrate lineage near the time NC emerged. Such gene duplications are thought to greatly facilitate the evolution of new protein functions (*protein neofunctionalization*) as partially redundant duplicated genes are freed to acquire new biochemical properties. A non-mutually exclusive alternative is that the new connections of the NC-GRN formed as new cis-regulatory DNA sequences arose flanking NC-GRN genes (*cis-regulatory neofunctionalization*) (illustrated in Figure 1B). Such sequences could have arisen *de novo* in previously non-functional DNA sequences or via accumulation of mutations by existing cis-regulatory elements. Recent reviews have focused on the general role of genome duplication in the emergence of new structures in vertebrate evolution (Holland, 2013) and on what developmental comparisons of chordate groups can tell us about the role of genome duplication in NC evolution (Green and Bronner, 2013; Hall and Gillis, 2013). Here we review recent papers examining the way in which new regulatory interactions are established between transcription factors and their targets. We examine what these studies reveal about the relationship between genome duplication and NC evolution. Finally, we discuss the contribution of both protein neofunctionalization and cis-regulatory neofunctionalization to the evolution of novel regulatory interactions of the NC-GRN.

Gene duplication and protein neofunctionalization in neural crest evolution

Did neofunctionalization of transcription factors contribute to the assembly of the NC-GRN? More than 40 years ago Ohno famously proposed that gene duplication might facilitate protein neofunctionalization by relaxing selective pressure on the duplicated copies, allowing them to accumulate novel, potentially fortuitous mutations (Ohno, 1970). Modifications or extensions to Ohno's central idea have since been proposed. For instance, the Duplication, Degeneration, Complementation (DDC) model seeks to explain the retention of duplicated genes, given the strong prediction that function-destroying mutations are much more probable than function-modifying ones (Force et al., 1999). The DDC model posits that selective pressure to maintain the two duplicates, protecting them from those deleterious mutations, results from *subfunctionalization*, in which the functions once carried out by the ancestral single gene-copy (singleton) become split between duplicates (illustrated in Figure 2A). The subfunctionalization initially occurs at the cis-regulatory level. Degeneration of complementary regulatory elements causes loss of each duplicate in distinct subdomains of the expression domain of the ancient singleton gene (temporal or spatial). Since the function is required in each subdomain, there is now selective pressure to maintain both duplicates. Subsequently, the duplicates may diverge from one another as they become optimized for the biochemical function they perform in each particular subdomain.

The Escape from Adaptive Conflict model is another refinement of the Ohno hypothesis. It posits that a gene under optimization for two independent functions prior to duplication is constrained by the need to carry out both functions ("gene sharing") (Hittinger and Carroll, 2007; Hughes, 1994; Piatigorsky, 1990). Subsequently, upon duplication, the two copies "escape" this constraint and each can be further optimized for its particular function. Thus,

relatively freed from constraint, the duplicates are more likely to acquire a novel biochemical function (neofunctionalization) than prior to duplication (illustrated in Figure 2B).

Does gene duplication indeed result in functional specialization of the new genes? One mode of assessing this question is to directly compare changes at the amino acid sequence level of duplicated proteins, both to each other and to non-duplicates. Based on the number and type of amino acid changes it can be inferred the amount of selective pressure on a given coding sequence (either positive or purifying). Several studies have found that following gene duplication there are distinct patterns in amino acid substitutions, indicating a decrease in purifying selection on young duplicates relative to the ancestral gene (Kondrashov et al., 2002; Lynch and Conery, 2000). Interestingly, frequently one duplicate exhibits a higher evolutionary rate than the other, suggesting that the selective pressures on the duplicates are distinct and that they are being selected for different properties (Brunet et al., 2006; Conant and Wagner, 2003; Zhang et al., 2003).

While such analyses are provocative evidence for sub and/or neofunctionalization, they cannot reveal if these changes have significant consequences on protein function. Fortunately, in many metazoan model systems it is now possible to directly test the extent to which paralogs have diverged from one another by replacing one paralog with another. By replacing a duplicated gene with an ortholog from a basal species containing just one copy of the gene, one can infer whether the protein duplicate has acquired a novel function not possessed by the ancestral gene. We will review several examples of such experiments in the context of transcription factors in the NC-GRN (summarized in Table 1).

The neural-crest-inducing quality of Pax group III proteins (Pax3 and Pax7) was likely to be present in the ancestral group III ortholog

The family of *Pax* transcription factors, whose members contribute to developmental GRNs of a variety of tissues [Reviewed in (Buckingham and Relaix, 2007; Wang et al., 2008)], has been expanded by gene duplication during evolution. A single *Pax* ortholog has been identified in *Trichoplax adhaerens*, a simple metazoan (Hadrys et al., 2005), while four paralogs are present in the basal chordate amphioxus (Holland et al., 1995; Holland, 1999). In gnathostomes, there are four groups of *Pax* paralogs (groups I–IV), containing two or three members each; each group is orthologous to one of the four paralogs in amphioxus.

Pax group III members (*Pax3* and *Pax7*) are essential for early NC development and are candidates to have undergone neofunctionalization during the emergence of the NC-GRN. *Pax3* mutant mice lack many trunk-NC and vagal-NC derivatives, including the peripheral nervous system, melanocytes, and NC-derived cells populating the heart; in addition they exhibit defects in other tissues (spina bifida, exencephaly, and loss of muscle progenitor cells) (Bajolle et al., 2006; Franz, 1993; Franz and Kothary, 1993; Franz et al., 1993; Tremblay et al., 1995; Watanabe et al., 1998). In contrast, *Pax7* mutant mice display defects in cranial NC development and derivatives, such as the maxilla (Mansouri et al., 1996). *Pax7* likely also carries out some redundant roles with *Pax3* in other aspects of NC development given their overlapping expression patterns (Lacosta et al., 2005; Maczkowiak

et al., 2010; Murdoch et al., 2012). Studies in various vertebrate species indicate that Pax3 and Pax7 are direct and/or indirect regulators of known NC-GRN transcription factors, including *Tfap2*, *SoxE*, *Snail*, *FoxD3*, *c-RET* and *Mitf* (Basch et al., 2006; de Croze et al., 2011; Dottori et al., 2001; Lang and Epstein, 2003; Maczkowiak et al., 2010; Monsoro-Burg et al., 2005; Sato et al., 2005). Interestingly, in cephalochordates the single Pax group III ortholog is expressed in the neural plate border (NPB) (Yu et al., 2008b). In contrast, expression of other important transcription factors in the NC-GRN including *Tfap2*, *SoxE*, and *FoxD3* orthologs is missing in this region. Based on these observations, Pax3 and Pax7, or the Pax group III ortholog from which they were derived, may have undergone neofunctionalization in order to gain specificity for one or more of the other NC-GRN genes.

To address functional diversity among Pax paralogs, Relaix and colleagues replaced the coding sequence of Pax3, a member of subfamily III, with that of Pax7, also of subfamily III (Relaix et al., 2004). In these mice, the Pax3 gene was absent but Pax7 was expressed within the normal Pax3 expression domain. Interestingly, NC development was largely normal (Relaix et al., 2004). By contrast, when Pax3 was replaced with Pax8, a member of group II, NC was abnormal, as in non-transgenic Pax3 mutants (Hayashi et al., 2011). Muscle development, also affected in Pax3 mutants, was not restored by Pax8 expression (Hayashi et al., 2011). In this example, paralogs resulting from a more recent duplication event are functionally equivalent, while a paralog resulting from a more ancient one has a distinct function. Because both subfamily III members possess the ability to induce NC, the gene duplication event that expanded subfamily III could not have promoted the appearance of this functionality. However, an earlier duplication event might have done so. The appropriate functionality of the Pax protein necessary for it to contribute to the NC-GRN, for instance the specific DNA binding preference, may have emerged long before the NC itself. Appropriate knock-in experiments could test whether the amphioxus subfamily III Pax protein and the Trichoplax Pax ortholog possess this functionality.

Interphylum gene swaps argue against an important role for protein neofunctionalization in SoxE or Tfap2 in neural crest evolution

The Sox family of transcription factors, which is divided into multiple subfamilies (SoxA-J) based on the sequence similarity of their respective DNA-binding domains (Bowles et al., 2000), is also of interest in the evolution of the NC-GRN. The *SoxE* group is comprised of *Sox8*, *Sox9*, and *Sox10* in gnathostomes; the basal vertebrate lamprey also contains three *SoxE* paralogs (Lakiza et al., 2011; McCauley and Bronner-Fraser, 2006). Vertebrate SoxE paralogs are expressed in mesoderm and in NC. In gnathostomes, loss of *Sox10* disrupts specification or survival of early NC and *Sox10* is also required for terminal differentiation of multiple NC derivatives (Britsch et al., 2001; Carney et al., 2006; Dutton et al., 2001; Elworthy et al., 2003; Kapur, 1999; Kim et al., 2003; Reiprich et al., 2008; Southard-Smith et al., 1998). Of note, amphioxus contains just one *SoxE* paralog and it is expressed in mesoderm in early neurulae (Yu et al., 2008b), and throughout the neural tube of early larvae (Meulemans and Bronner-Fraser 2007). Therefore, one possibility is that the single SoxE gene in the vertebrate precursor was constrained by its requirement to function in the

mesoderm and differentiating central nervous system, and duplication of the *SoxE* gene facilitated acquisition of NC functions in one of the duplicates.

Did the *SoxE* genes acquire NC functionality through protein neofunctionalization subsequent to gene duplication? To address this question, Cossais and colleagues replaced mouse *Sox10* gene with a *Sox* ortholog from *Drosophila melanogaster*, *Sox100B* (Cossais et al., 2010). Transgenic mice were initially monitored for proper endogenous expression levels of *Sox100B* using real-time PCR, which showed similar levels to *Sox10*. Transgenic mice were assessed for changes in expression levels of other *SoxE* family members, as auto-regulation between *SoxE* family members has been observed (Cheung et al., 2005). No significant differences were observed for either *Sox8* or *Sox9* at the mRNA or protein levels. Interestingly, upon replacement of *Sox10* with *Sox100B*, mice showed development of multiple NC derivatives; however, most derivatives showed some mild defects as compared to WT embryos. Specification of melanocytes, as observed by *Mitf* and *Dct* expression, occurred in *Sox100B* expressing, *Sox10* deficient mice, although fewer cells were detected than in control embryos. In such mice, enteric nervous system development was largely unaffected at more rostral locations but was reduced, compared to in controls, at more caudal locations of the gut. Similarly, the sympathetic ganglia were present in *Sox100B* expressing, *Sox10* deficient mice, but were more perfectly rescued at rostral than caudal levels. Analysis of dorsal root ganglia development in transgenic mice revealed subtle defects, most notably in neuronal numbers (reduced 40%); however, satellite glia were present at only slightly reduced numbers (78% of WT numbers). Finally, Schwann cell development was relatively unaffected in *Sox100B* expressing mice. Taken together, these findings suggest that to a large degree *Sox100B* is functionally equivalent to *Sox10* in mouse NC development, and therefore protein neofunctionalization of *SoxE*, whether facilitated by gene duplication or not, did not play a major role in the inclusion of *Sox* in the NC-GRN.

Interestingly, another gene-swap experiment, carried out by the same group, revealed protein subfunctionalization among *SoxE* subfamily members. *Sox8* expression partially overlaps *Sox10* expression during development, and compound mutants (*Sox10* and *Sox8*) show more dramatic defects than single mutants in certain NC derivatives, such as the enteric nervous system (as well as non-NC derived oligodendrocytes) (Maka et al., 2005). These findings suggest that the role of an ancient *SoxE* ortholog may have been split among these duplicates (Maka et al., 2005; Stolt et al., 2004). Have the functions of the duplicates diverged? Replacement of *Sox10* with *Sox8* results in fairly normal NC development, with some interesting exceptions. The peripheral nervous system (cranial ganglia, dorsal root ganglia, sympathetic ganglia), which is virtually absent in *Sox10*-mutants, is patterned normally in embryos expressing *Sox8* from the *Sox10* locus (Kellerer et al., 2006). However, similar to embryos with *Sox100B* knocked-in to the *Sox10* locus, the enteric nervous system was normal in rostral, but completely absent in more caudal portions of the gut. Interestingly, the *Sox8* expressing, *Sox10* deficient embryos are nearly completely devoid of melanocyte precursors, melanoblasts, as detected by *Dct*, *Mitf*, and *Kit* expression, similar to *Sox10*-mutants. This contrasts to *Sox100B* knock-in embryos, where melanocytes are restored to a much greater extent than in *Sox8* knock-in embryos. These results indicate that

Sox10 has retained but Sox8 has lost some of the NC-patterning functions, specifically the ability to promote melanocyte development, that were present in the ortholog from which both were derived by duplication.

We recently explored the role genome duplication may have played in promoting novel NC-inducing function to the transcription factor activator protein 2 (*Tfap2*) family (Van Otterloo et al., 2012). Similar to previous examples, *Tfap2* is a strong candidate for such a scenario because it is essential for NC specification and was duplicated in the chordate lineage (Van Otterloo et al., 2012). Two *Tfap2* paralogs are present in the agnathan vertebrate, lamprey, while five are seen in jawed-vertebrates (*Tfap2a-e*) and only a single *Tfap2* homolog (*Bf-AP2*) has been identified in the non-vertebrate chordate, amphioxus [see (Meulemans and Bronner-Fraser, 2002; Van Otterloo et al., 2012) for details]. The importance of *Tfap2* in NC development is well established in a variety of model systems (Brewer et al., 2004; Brewer et al., 2002; Hoffman et al., 2007; Knight et al., 2003; Li and Cornell, 2007; Luo et al., 2003; O'Brien et al., 2004). We tested whether gene duplication facilitated the acquisition of a NC-inducing function (protein neofunctionalization) in vertebrate *Tfap2*. In zebrafish, simultaneous inhibition of *tfap2a* and *tfap2c*, the two paralogs expressed in the NPB, prevents induction of NC (Hoffman et al., 2007; Li and Cornell, 2007). mRNA encoding the single amphioxus *Tfap2* paralog was injected into *tfap2a/c*-depleted zebrafish embryos and NC induction was assessed. Interestingly, the amphioxus *Tfap2* protein was able to fully restore NC induction as assessed by expression of the NC-GRN component *foxd3* (Van Otterloo et al., 2012). In addition, markers of both cranial (*dlx2* expression) and trunk (melanophores) NC were restored in this context. The *AP2* ortholog from *Drosophila* was similarly competent to induce NC. However, while NC was induced in *tfap2a/c*-depleted zebrafish embryos injected with RNA encoding the *Drosophila AP2* ortholog, melanocytes and other NC derivatives were not rescued to the same extent as in embryos injected with, for instance, zebrafish *tfap2b*. These results imply that protein neofunctionalization did not imbue TFAP2 with the ability to induce NC though genome duplication may have allowed vertebrate *Tfap2* paralogs to become optimized for later functions in specific NC derivatives (Van Otterloo et al., 2012).

The differential ability of duplicates to pattern the NC is an interesting finding highlighted by these studies. This observation strongly suggests a sequential establishment of NC derivatives over evolutionary time likely attributed to optimization of protein function within a given lineage. Given their appearance in more rudimentary organisms it has been speculated that melanocytes are a more ancient NC cell-type from which other derivatives have emerged. Recruitment, or co-option, of additional transcription factors within the melanocyte lineage may have facilitated its divergence into these additional cell types (Abitua et al., 2012; Jeffery, 2006; Jeffery, 2007; Jeffery et al., 2004). Interestingly, both for analysis of SoxE (Kellerer et al., 2006) and *Tfap2* (Van Otterloo et al., 2012) melanocytes were the most sensitive to replacement, either by an ortholog (Sox8 for Sox10) or “ancient” paralog (*Drosophila AP-2* for *Tfap2a/Tfap2c*), respectively. Such findings suggest that additional protein optimization may have occurred subsequent to duplication; this optimization may have included altered protein stability, cis/trans-binding partners, and

transcriptional potency, among other properties to facilitate diversification of NC derivatives.

In sum, the findings highlighted above suggest that for three key NC-GRN transcription factors, the vertebrate gene duplications were not followed by neofunctionalization. Furthermore, for *Tfap2* and *SoxE*, invertebrate pre-duplication orthologs were shown to be capable of performing the NC-specific function of the vertebrate duplicates. Thus, for these gene families, neofunctionalization also did not occur before the vertebrate genome duplications. Additional studies will be necessary to determine if this trend of functional conservation is true for other NC-GRN transcription factors.

The contribution of novel cis-regulatory elements to the neural crest GRN

We have just reviewed that gene-swap experiments have yet to yield evidence that neofunctionalization of transcription factors contributed to the changes in gene expression just mentioned. The alternative is that the novel expression of these genes in the NPB/NC resulted from the appearance of novel cis-regulatory elements (CREs). The way in which NPB/NC CREs evolved is largely an open question. Much of what we know about CRE evolution comes from experiments in *Drosophila*. Gene regulation in the *Drosophila* embryo often involves the establishment of sharply defined gene expression domains in response to broadly-distributed regulatory factors. This is particularly true in the early syncytial blastoderm. Similarly in vertebrates, initiation of NPB gene expression involves the establishment of a sharply-defined gene expression domain in response to broadly distributed signaling molecules such as BMPs, Wnts and FGFs. Thus an examination of *Drosophila* enhancer evolution is useful in considering how vertebrate NC/NPB enhancers may have evolved.

Novel CREs can evolve in at least two ways. First, regulatory elements can appear *de novo* near a gene, via DNA mutation or transposition of a mobile element. For instance, in *Drosophila*, enhancers for the *yellow* gene appear to have evolved repeatedly via the accumulation of point mutations in previously non-regulatory DNA sequences, leading to divergent pigmentation patterns (Kalay and Wittkopp, 2010). Moreover, a recent study in teleost fish used the sequence conservation of pseudogenes to demonstrate that enhancer sequences can indeed evolve from non-regulatory sequence (Eichenlaub and Ettwiller, 2011). In another example, transposable elements in the human and zebrafish genomes have given rise to large numbers of p53 responsive enhancers (Micale et al., 2012; Wang et al., 2007).

Second, new CREs can evolve from pre-existing CREs as mutations change the number and/or spacing of transcription factor binding sites. In one example, in *Drosophila santomea* the *Neprilysin-1* gene acquired novel expression in the optic lobe through the accumulation of a small number of mutations in a pre-existing enhancer (Rebeiz et al., 2011). In another example, the activities of a family of neurogenic ectoderm enhancers were altered during lineage divergence to compensate for changes in embryo size (Crocker et al., 2008). Thus, in the early *Drosophila* embryo there is a gradient of nuclear localization of the transcription factor Dorsal (Roth et al., 1989), with highest levels in the ventral embryo. Moderate levels

of nuclear Dorsal protein specify neurogenic ectoderm in an intermediate zone of the embryo. Enhancers that drive gene expression in this tissue often contain linked binding sites for Dorsal and the E-box transcription factor Twist (Crocker et al., 2008). Mutations that alter the spacing between the two sites appear to alter the sensitivity of the enhancer to the Dorsal protein, thereby expanding or shrinking the embryonic domain of gene expression.

Evidence that incorporation of *FoxD3* and *SoxE* into the NC-GRN involved the evolution of new cis-regulatory elements

Recent regulatory-element-swap studies have begun to identify the novel CREs that drive expression of *FoxD* and *SoxE* genes in the NC. In vertebrates, *FoxD3* (Dirksen and Jamrich, 1995) and to a lesser extent *FoxD2* (Pohl and Knochel, 2002) are expressed in NC. In amphioxus, the single *FoxD* ortholog is expressed in the diencephalon, notochord and paraxial mesoderm (Yu et al., 2002). The region upstream of zebrafish *foxD3* contains an enhancer that drives gene expression in the NC (Gilmour et al., 2002). Interestingly, the region immediately upstream of amphioxus *FoxD* can direct reporter expression in the paraxial mesoderm and notochord in both amphioxus and chick; however, it cannot drive expression in the NC of chicken (Yu et al., 2008a). Recent experiments have demonstrated that, in addition to the promoter region, two distant enhancers drive *FoxD3* expression in chick NC (Simões-Costa et al., 2013). An alignment of the genomic regions surrounding chicken *FoxD3* and amphioxus *FoxD* using LAGAN (Brudno et al., 2003) indicates that neither of these distal elements are conserved at the primary sequence level in amphioxus. However, it is possible that the elements are conserved and the sequences are no longer recognizable as being homologous to the chicken enhancers or the sequences could lie in regions of the genome that have not yet been assembled properly. These studies have identified CREs that drive NC expression of *FoxD*, and have determined that it likely emerged sometime between the split of amphioxus and the vertebrates.

A similar analysis identified a CRE driving NC expression of *soxE*, and suggested that the construction of this element required accumulation of binding sites for the transcription factor Tfp2 (Van Otterloo et al., 2012). As mentioned above, vertebrate *SoxE* genes *Sox8*, *Sox9* and *Sox10* are expressed in the NPB during neurulation (Bell et al., 2000; Cheung and Briscoe, 2003; Southard-Smith et al., 1998), while amphioxus *SoxE* is expressed only in the mesoderm during this stage (Yu et al., 2008b). A CRE adjacent to zebrafish *sox10* is sufficient to drive reporter expression in the presumptive NC, and the Tfp2 binding sites are critical for this activity (Van Otterloo et al., 2012). The corresponding region upstream of amphioxus *soxE* has enhancer activity in zebrafish embryos, but it does not contain Tfp2 binding sites and is unable to drive gene expression in the zebrafish NPB (Van Otterloo et al., 2012). This suggests that the evolution of these binding sites facilitated *SoxE* expression in the NPB. It is unclear whether these binding sites are part of a NPB/NC enhancer that arose *de novo* or endowed an existing enhancer with the ability to drive gene expression in the developing NC. Deeper comparative studies may shed light on this issue.

Did genome duplication facilitate the generation of novel cis-regulatory elements driving neural crest expression?

Given that the construction of the novel NC-GRN required the generation of novel CREs, could gene duplication have facilitated such CRE neofunctionalization? In principle, recently duplicated paralogous enhancers would be under reduced selective pressure to maintain their ancient expression domains, permitting one to be modified such that it becomes active in the NPB/NC. If this were the case, we would expect that the ancestor of a NC gene was not expressed in the NPB and that after duplication one paralog, but not the other, acquired a NPB/NC expression domain while simultaneously losing an ancestral expression domain.

As *Tfap2*, *SoxE* and *FoxD* families were all duplicated concomitant with their recruitment to the NC, did any of them simultaneously lose expression in another domain? *Tfap2* was deployed in the NPB very early in vertebrate evolution (Figure 3A). Gnathostome *Tfap2a* and *Tfap2c* as well as agnathan (lamprey) *Tfap2* are expressed in the NPB and NC (Meulemans and Bronner-Fraser, 2002; Van Otterloo et al., 2012). The phylogeny and expression of these genes suggests that they evolved from a common ancestor that was once only deployed in the early epidermis and brain, similar to amphioxus *Tfap2*. This common ancestor underwent a regulatory change causing its expression in the NPB in addition to the epidermis and brain. Thus the *Tfap2a/b/c* ancestor maintained its ancestral expression domains while evolving a new NPB expression domain, suggesting that genome duplication was not a prerequisite for deployment of *Tfap2* in the NPB. Similarly, while amphioxus *SoxE* is expressed in mesoderm, all *SoxE* genes in gnathostomes (*Sox8*, *Sox9* and *Sox10*) and in two lamprey *SoxEs* (*SoxE1* and *SoxE2*) (McCauley and Bronner-Fraser, 2006) are expressed in the developing NC, suggesting the CRE driving *SoxE* expression in the NPB/NC evolved before the duplications that generated *Sox8*, *Sox9* and *Sox10* (Figure 3B). By contrast, only two of five vertebrate *Foxd* paralogs (*Foxd2*, *Foxd3*) are expressed in the NPB/NC, raising the possibility that novel expression in the NPB/NC arose after one or both of the vertebrate genome duplications (Figure 3C). However, the urochordate *Ciona intestinalis* has only one *FoxD* gene and it is expressed in the a9.49 cell lineage, which recent studies indicate deploy portions of the NC-GRN (Abitua et al., 2012; Imai et al., 2002). If the *Ciona FoxD* expression is homologous to the vertebrate *FoxD2/3* NPB/NC expression, then expression of *FoxD* in the NC probably predates the vertebrate genome duplications. In summary, none of these results support the notion that duplication of *Tfap2*, *SoxE* and *FoxD* was necessary for NC to arise.

Gene duplication as a means to alter gene dosage during neural crest patterning

While genome duplication may not have been critical for the NC-GRN to initially evolve, it may have proven useful in allowing organisms to change what they do with the NC. Genome duplication may have facilitated elaboration of NC derivatives by providing multiple copies of essentially identical proteins that could be deployed in different combinations to give different outputs. In the *Tfap2* family, for example, *tfap2a* and *tfap2c*

are both expressed in the zebrafish NPB during gastrulation, but only *tfap2a* is highly expressed in the premigratory NC during neurulation and segmentation (Hoffman et al., 2007; Li and Cornell, 2007; Van Otterloo et al., 2012). Later in development different combinations of *tfap2* paralogs are expressed in different subsets of NC derivatives. Similar nested gene expression patterns among paralogs are also observed in the *SoxE*, *Zic*, *Pax3/7* and *Dlx* families (Depew et al., 2002; Hong and Saint-Jeannet, 2005; Merzdorf, 2007; Minchin and Hughes, 2008). Differential expression of paralogs could be an easy mechanism for achieving coarse control over the levels of a protein family expressed in different cell types. These different protein levels could then be exploited to trigger different GRNs. For example, gnathostome genomes contain 6–8 paralogs from the *Dlx* gene family that likely arose from a single *dlx* gene in the ancestral chordate (Stock et al., 1996). These paralogs are now expressed in a nested pattern in the developing jaw and different *Dlx* gene combinations, allowing different parts of the jaw to receive vastly different doses of Dlx proteins (Depew et al., 2002). This could have facilitated some of the phenotypic diversity seen in the head skeletons of modern vertebrates.

Current directions

Although limited, current evidence suggests protein neofunctionalization of NC specific transcription factors was not a key event in the emergence of NC in vertebrates. However, as discussed, the studies highlighted above suggest duplicated proteins may become optimized for their roles in particular NC derivatives including specification, maintenance, and/or differentiation. As genome manipulation becomes more readily available for an extended number of species, “gene-swap” experiments can be utilized to further characterize how changes at the protein sequence level contributes to changes in NC behavior and differentiation. More refined experiments, such as “domain-swapping” between orthologs can also be implemented to dissect what key changes occurred to the protein to alter its biochemical function. In the context of transcription factors, these studies could identify protein changes important for binding to novel downstream effectors in a specific NC lineage. These studies should begin to add clarity to how protein changes have driven the evolution of NC derivatives.

An additional current challenge to the field is to determine how NPB/NC enhancers were generated during evolution. If such an enhancer emerged *de novo*, then a prediction is that the homologous region of the amphioxus genome would lack any regulatory function, whereas if it emerged instead through modification of a pre-existent enhancer, then the homologous region from amphioxus may have detectable enhancer activity. Testing these predictions is hampered by the difficulty of identifying homologous chromatin regions between distant species. In addition, homologous enhancers can be difficult to recognize, as it appears they can undergo drastic changes in sequence while maintaining the same activity (Fisher et al., 2006; Hare et al., 2008). In previous studies showing *de novo* enhancer generation, homology was established by studying closely related organisms in which low amounts of neutral DNA mutation had occurred (Kalay and Wittkopp, 2010) or by studying pseudogene sequences which had maintained their sequence similarity because of non- cis-regulatory constraints (Eichenlaub and Ettwiller, 2011). These options are unlikely to be available when assessing non-coding sequence homology between amphioxus and

vertebrates. However, there are other types of evidence that can suggest homology. For example, if an amphioxus enhancer and a vertebrate enhancer are located in the same position relative to the gene they regulate, this suggests homology, particularly if the enhancers are in a homologous intron. This type of deep conservation of enhancer position was previously observed in dorsal-ventral patterning enhancers in *Drosophila* and flour beetle (Cande et al., 2009).

Another type of evidence that would suggest enhancer modification is if the ancestral activity of an enhancer is entangled with a derived activity. For example, the ancestral *Tfap2* was expressed in the epidermis and vertebrates expanded this expression dorsally into the NPB (Meulemans and Bronner-Fraser, 2002; Van Otterloo et al., 2012). This expansion within the plane of an epithelium is reminiscent of the expansion of neurogenic ectoderm gene expression discussed above, and may occur by similar mechanisms. The isolation of a *Tfap2* enhancer with activity in both the epidermis and NPB would indicate the ancestral epidermal enhancer might have been modified to have NPB expression.

Several new technologies and reagents promise to greatly facilitate future studies of CRE evolution. Next generation sequencing technologies have hugely accelerated the rate at which CREs can be identified (Kundaje et al., 2012; Wang et al., 2012). In addition, antibodies are being developed for amphioxus NPB specifiers. These can be used with ChIP-Seq methodology, making it possible to identify differences in their regulatory targets and the targets of their vertebrate homologs. BAC transgenesis, in model vertebrate species and in tunicates and amphioxus, will be useful in detecting and comparing CREs associated with *SoxE*, *FoxD* and *Tfap2* genes in each group (Minguillon et al., 2009).

Acknowledgments

Work on the neural crest GRN in the Cornell group is funded by NSF grant IOS- 1147221. Work on the neural crest GRN in the Medeiros group is funded by NIH grant 1R03DE019804. ATG was supported by NSF grant DBI-0905991.

We are grateful to Albert Erives for comments on earlier drafts of the manuscript.

References

- Abitua PB, Wagner E, Navarrete IA, Levine M. Identification of a rudimentary neural crest in a non-vertebrate chordate. *Nature*. 2012; 492:104–7. [PubMed: 23135395]
- Bajolle F, Zaffran S, Kelly RG, Hadchouel J, Bonnet D, Brown NA, Buckingham ME. Rotation of the myocardial wall of the outflow tract is implicated in the normal positioning of the great arteries. *Circ Res*. 2006; 98:421–8. [PubMed: 16397144]
- Basch ML, Bronner-Fraser M, Garcia-Castro MI. Specification of the neural crest occurs during gastrulation and requires Pax7. *Nature*. 2006; 441:218–22. [PubMed: 16688176]
- Bell KM, Western PS, Sinclair AH. SOX8 expression during chick embryogenesis. *Mech Dev*. 2000; 94:257–60. [PubMed: 10842083]
- Bowles J, Schepers G, Koopman P. Phylogeny of the SOX family of developmental transcription factors based on sequence and structural indicators. *Dev Biol*. 2000; 227:239–55. [PubMed: 11071752]
- Brewer S, Feng W, Huang J, Sullivan S, Williams T. Wnt1-Cre-mediated deletion of AP-2alpha causes multiple neural crest-related defects. *Dev Biol*. 2004; 267:135–52. [PubMed: 14975722]

- Brewer S, Jiang X, Donaldson S, Williams T, Sucov HM. Requirement for AP-2alpha in cardiac outflow tract morphogenesis. *Mech Dev.* 2002; 110:139–49. [PubMed: 11744375]
- Britsch S, Goerich DE, Riethmacher D, Peirano RI, Rossner M, Nave KA, Birchmeier C, Wegner M. The transcription factor Sox10 is a key regulator of peripheral glial development. *Genes Dev.* 2001; 15:66–78. [PubMed: 11156606]
- Brunet FG, Roest Crollius H, Paris M, Aury JM, Gibert P, Jaillon O, Laudet V, Robinson-Rechavi M. Gene loss and evolutionary rates following whole-genome duplication in teleost fishes. *Mol Biol Evol.* 2006; 23:1808–16. [PubMed: 16809621]
- Buckingham M, Relaix F. The role of Pax genes in the development of tissues and organs: Pax3 and Pax7 regulate muscle progenitor cell functions. *Annu Rev Cell Dev Biol.* 2007; 23:645–73. [PubMed: 17506689]
- Cande J, Goltsev Y, Levine MS. Conservation of enhancer location in divergent insects. *Proc Natl Acad Sci U S A.* 2009; 106:14414–9. [PubMed: 19666595]
- Carney TJ, Dutton KA, Greenhill E, Delfino-Machin M, Dufourcq P, Blader P, Kelsh RN. A direct role for Sox10 in specification of neural crest-derived sensory neurons. *Development.* 2006; 133:4619–30. [PubMed: 17065232]
- Cheung M, Briscoe J. Neural crest development is regulated by the transcription factor Sox9. *Development.* 2003; 130:5681–93. [PubMed: 14522876]
- Cheung M, Chaboissier MC, Mynett A, Hirst E, Schedl A, Briscoe J. The transcriptional control of trunk neural crest induction, survival, and delamination. *Dev Cell.* 2005; 8:179–92. [PubMed: 15691760]
- Conant GC, Wagner A. Asymmetric sequence divergence of duplicate genes. *Genome Res.* 2003; 13:2052–8. [PubMed: 12952876]
- Cossais F, Sock E, Hornig J, Schreiner S, Kellerer S, Bosl MR, Russell S, Wegner M. 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. *Dev Biol.* 2010; 341:267–81. [PubMed: 20144603]
- Crocker J, Tamori Y, Erives A. Evolution acts on enhancer organization to fine-tune gradient threshold readouts. *PLoS Biol.* 2008; 6:e263. [PubMed: 18986212]
- Davidson EH, McClay DR, Hood L. Regulatory gene networks and the properties of the developmental process. *Proc Natl Acad Sci U S A.* 2003; 100:1475–80. [PubMed: 12578984]
- de Croze N, Maczkowiak F, Monsoro-Burq AH. Reiterative AP2a activity controls sequential steps in the neural crest gene regulatory network. *Proc Natl Acad Sci U S A.* 2011; 108:155–60. [PubMed: 21169220]
- Depew MJ, Lufkin T, Rubenstein JL. Specification of jaw subdivisions by Dlx genes. *Science.* 2002; 298:381–5. [PubMed: 12193642]
- Dirksen ML, Jamrich M. Differential expression of fork head genes during early *Xenopus* and zebrafish development. *Dev Genet.* 1995; 17:107–16. [PubMed: 7586752]
- Dottori M, Gross MK, Labosky P, Goulding M. The winged-helix transcription factor Foxd3 suppresses interneuron differentiation and promotes neural crest cell fate. *Development.* 2001; 128:4127–38. [PubMed: 11684651]
- Dutton KA, Pauliny A, Lopes SS, Elworthy S, Carney TJ, Rauch J, Geisler R, Haffter P, Kelsh RN. Zebrafish colourless encodes sox10 and specifies non-ectomesenchymal neural crest fates. *Development.* 2001; 128:4113–25. [PubMed: 11684650]
- Eichenlaub MP, Ettwiller L. De novo genesis of enhancers in vertebrates. *PLoS Biol.* 2011; 9:e1001188. [PubMed: 22069375]
- Elworthy S, Lister JA, Carney TJ, Raible DW, Kelsh RN. Transcriptional regulation of mitfa accounts for the sox10 requirement in zebrafish melanophore development. *Development.* 2003; 130:2809–18. [PubMed: 12736222]
- Fisher S, Grice EA, Vinton RM, Bessling SL, McCallion AS. Conservation of RET regulatory function from human to zebrafish without sequence similarity. *Science.* 2006; 312:276–9. [PubMed: 16556802]
- Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics.* 1999; 151:1531–45. [PubMed: 10101175]

- Franz T. The Splotch (Sp1H) and Splotch-delayed (Spd) alleles: differential phenotypic effects on neural crest and limb musculature. *Anat Embryol (Berl)*. 1993; 187:371–7. [PubMed: 8512089]
- Franz T, Kothary R. Characterization of the neural crest defect in Splotch (Sp1H) mutant mice using a lacZ transgene. *Brain Res Dev Brain Res*. 1993; 72:99–105.
- Franz T, Kothary R, Surani MA, Halata Z, Grim M. The Splotch mutation interferes with muscle development in the limbs. *Anat Embryol (Berl)*. 1993; 187:153–60. [PubMed: 8238963]
- Gilmour DT, Maischein HM, Nusslein-Volhard C. Migration and function of a glial subtype in the vertebrate peripheral nervous system. *Neuron*. 2002; 34:577–88. [PubMed: 12062041]
- Green SA, Bronner ME. Gene duplications and the early evolution of neural crest development. *Semin Cell Dev Biol*. 2013; 24:95–100. [PubMed: 23287633]
- Hadrys T, DeSalle R, Sagasser S, Fischer N, Schierwater B. The Trichoplax PaxB gene: a putative Proto-PaxA/B/C gene predating the origin of nerve and sensory cells. *Mol Biol Evol*. 2005; 22:1569–78. [PubMed: 15858210]
- Hall BK, Gillis JA. Incremental evolution of the neural crest, neural crest cells and neural crest-derived skeletal tissues. *J Anat*. 2013; 222:19–31. [PubMed: 22414251]
- Hare EE, Peterson BK, Iyer VN, Meier R, Eisen MB. Sepsid even-skipped enhancers are functionally conserved in *Drosophila* despite lack of sequence conservation. *PLoS Genet*. 2008; 4:e1000106. [PubMed: 18584029]
- Hayashi S, Rocancourt D, Buckingham M, Relaix F. Lack of in vivo functional compensation between Pax family groups II and III in rodents. *Mol Biol Evol*. 2011; 28:2787–98. [PubMed: 21512107]
- Hittinger CT, Carroll SB. Gene duplication and the adaptive evolution of a classic genetic switch. *Nature*. 2007; 449:677–81. [PubMed: 17928853]
- Hoffman TL, Javier AL, Campeau SA, Knight RD, Schilling TF. Tfap2 transcription factors in zebrafish neural crest development and ectodermal evolution. *J Exp Zool B Mol Dev Evol*. 2007; 308:679–91. [PubMed: 17724731]
- Holland LZ. Evolution of new characters after whole genome duplications: Insights from amphioxus. *Semin Cell Dev Biol*. 2013; 24:101–9. [PubMed: 23291260]
- Holland ND, Holland LZ, Kozmik Z. An amphioxus Pax gene, *AmphiPax-1*, expressed in embryonic endoderm, but not in mesoderm: implications for the evolution of class I paired box genes. *Mol Mar Biol Biotechnol*. 1995; 4:206–14. [PubMed: 7670596]
- Holland PW. The effect of gene duplication on homology. *Novartis Found Symp*. 1999; 222:226–36. discussion 236–42. [PubMed: 10332763]
- Hong CS, Saint-Jeannet JP. Sox proteins and neural crest development. *Semin Cell Dev Biol*. 2005; 16:694–703. [PubMed: 16039883]
- Hughes AL. The evolution of functionally novel proteins after gene duplication. *Proc Biol Sci*. 1994; 256:119–24. [PubMed: 8029240]
- Imai KS, Satoh N, Satou Y. An essential role of a *FoxD* gene in notochord induction in *Ciona* embryos. *Development*. 2002; 129:3441–53. [PubMed: 12091314]
- Jeffery WR. Ascidian neural crest-like cells: phylogenetic distribution, relationship to larval complexity, and pigment cell fate. *J Exp Zool B Mol Dev Evol*. 2006; 306:470–80. [PubMed: 16619245]
- Jeffery WR. Chordate ancestry of the neural crest: new insights from ascidians. *Semin Cell Dev Biol*. 2007; 18:481–91. [PubMed: 17509911]
- Jeffery WR, Strickler AG, Yamamoto Y. Migratory neural crest-like cells form body pigmentation in a urochordate embryo. *Nature*. 2004; 431:696–9. [PubMed: 15470430]
- Kalay G, Wittkopp PJ. Nomadic enhancers: tissue-specific cis-regulatory elements of yellow have divergent genomic positions among *Drosophila* species. *PLoS Genet*. 2010; 6:e1001222. [PubMed: 21151964]
- Kapur RP. Early death of neural crest cells is responsible for total enteric aganglionosis in *Sox10(Dom)/Sox10(Dom)* mouse embryos. *Pediatr Dev Pathol*. 1999; 2:559–69. [PubMed: 10508880]

- Kellerer S, Schreiner S, Stolt CC, Scholz S, Bosl MR, Wegner M. Replacement of the Sox10 transcription factor by Sox8 reveals incomplete functional equivalence. *Development*. 2006; 133:2875–86. [PubMed: 16790476]
- Kim J, Lo L, Dormand E, Anderson DJ. SOX10 maintains multipotency and inhibits neuronal differentiation of neural crest stem cells. *Neuron*. 2003; 38:17–31. [PubMed: 12691661]
- Knight RD, Nair S, Nelson SS, Afshar A, Javidan Y, Geisler R, Rauch GJ, Schilling TF. lockjaw encodes a zebrafish tfap2a required for early neural crest development. *Development*. 2003; 130:5755–68. [PubMed: 14534133]
- Kondrashov FA, Rogozin IB, Wolf YI, Koonin EV. Selection in the evolution of gene duplications. *Genome Biol*. 2002; 3:RESEARCH0008. [PubMed: 11864370]
- Kundaje A, Kyriazopoulou-Panagiotopoulou S, Libbrecht M, Smith CL, Raha D, Winters EE, Johnson SM, Snyder M, Batzoglou S, Sidow A. Ubiquitous heterogeneity and asymmetry of the chromatin environment at regulatory elements. *Genome Res*. 2012; 22:1735–47. [PubMed: 22955985]
- Lacosta AM, Muniesa P, Ruberte J, Sarasa M, Dominguez L. Novel expression patterns of Pax3/Pax7 in early trunk neural crest and its melanocyte and non-melanocyte lineages in amniote embryos. *Pigment Cell Res*. 2005; 18:243–51. [PubMed: 16029418]
- Lakiza O, Miller S, Bunce A, Lee EM, McCauley DW. SoxE gene duplication and development of the lamprey branchial skeleton: Insights into development and evolution of the neural crest. *Dev Biol*. 2011; 359:149–61. [PubMed: 21889937]
- Lang D, Epstein JA. Sox10 and Pax3 physically interact to mediate activation of a conserved c-RET enhancer. *Hum Mol Genet*. 2003; 12:937–45. [PubMed: 12668617]
- Li W, Cornell RA. Redundant activities of Tfap2a and Tfap2c are required for neural crest induction and development of other non-neural ectoderm derivatives in zebrafish embryos. *Dev Biol*. 2007; 304:338–54. [PubMed: 17258188]
- Luo T, Lee YH, Saint-Jeannet JP, Sargent TD. Induction of neural crest in *Xenopus* by transcription factor AP2alpha. *Proc Natl Acad Sci U S A*. 2003; 100:532–7. [PubMed: 12511599]
- Lynch M, Conery JS. The evolutionary fate and consequences of duplicate genes. *Science*. 2000; 290:1151–5. [PubMed: 11073452]
- Maczkowiak F, Mateos S, Wang E, Roche D, Harland R, Monsoro-Burq AH. The Pax3 and Pax7 paralogs cooperate in neural and neural crest patterning using distinct molecular mechanisms, in *Xenopus laevis* embryos. *Dev Biol*. 2010; 340:381–96. [PubMed: 20116373]
- Maka M, Stolt CC, Wegner M. Identification of Sox8 as a modifier gene in a mouse model of Hirschsprung disease reveals underlying molecular defect. *Dev Biol*. 2005; 277:155–69. [PubMed: 15572147]
- Mansouri A, Stoykova A, Torres M, Gruss P. Dysgenesis of cephalic neural crest derivatives in Pax7–/– mutant mice. *Development*. 1996; 122:831–8. [PubMed: 8631261]
- McCauley DW, Bronner-Fraser M. Importance of SoxE in neural crest development and the evolution of the pharynx. *Nature*. 2006; 441:750–2. [PubMed: 16760978]
- Merzdorf CS. Emerging roles for zic genes in early development. *Dev Dyn*. 2007; 236:922–40. [PubMed: 17330889]
- Meulemans D, Bronner-Fraser M. Amphioxus and lamprey AP-2 genes: implications for neural crest evolution and migration patterns. *Development*. 2002; 129:4953–62. [PubMed: 12397104]
- Meulemans D, Bronner-Fraser M. Gene-regulatory interactions in neural crest evolution and development. *Dev Cell*. 2004; 7:291–9. [PubMed: 15363405]
- Meulemans D, Bronner-Fraser M. Central role of gene cooption in neural crest evolution. *J Exp Zool B Mol Dev Evol*. 2005; 304:298–303. [PubMed: 15880502]
- Meulemans D, Bronner-Fraser M. Insights from amphioxus into the evolution of vertebrate cartilage. *PLoS One*. 2007; 2:e787. [PubMed: 17726517]
- Meulemans D, McCauley D, Bronner-Fraser M. Id expression in amphioxus and lamprey highlights the role of gene cooption during neural crest evolution. *Dev Biol*. 2003; 264:430–42. [PubMed: 14651928]
- Micale L, Loviglio MN, Manzoni M, Fusco C, Augello B, Migliavacca E, Cotugno G, Monti E, Borsani G, Reymond A, Merla G. A fish-specific transposable element shapes the repertoire of p53 target genes in zebrafish. *PLoS One*. 2012; 7:e46642. [PubMed: 23118857]

- Minchin JE, Hughes SM. Sequential actions of Pax3 and Pax7 drive xanthophore development in zebrafish neural crest. *Dev Biol.* 2008; 317:508–22. [PubMed: 18417109]
- Minguillon C, Gibson-Brown JJ, Logan MP. Tbx4/5 gene duplication and the origin of vertebrate paired appendages. *Proc Natl Acad Sci U S A.* 2009; 106:21726–30. [PubMed: 19995988]
- Monsoro-Burq AH, Wang E, Harland R. Msx1 and Pax3 cooperate to mediate FGF8 and WNT signals during *Xenopus* neural crest induction. *Dev Cell.* 2005; 8:167–78. [PubMed: 15691759]
- Murdoch B, DelConte C, Garcia-Castro MI. Pax7 lineage contributions to the mammalian neural crest. *PLoS One.* 2012; 7:e41089. [PubMed: 22848431]
- O'Brien EK, d'Alencon C, Bonde G, Li W, Schoenebeck J, Allende ML, Gelb BD, Yelon D, Eisen JS, Cornell RA. Transcription factor Ap-2alpha is necessary for development of embryonic melanophores, autonomic neurons and pharyngeal skeleton in zebrafish. *Dev Biol.* 2004; 265:246–61. [PubMed: 14697367]
- Ohno, S. Evolution by gene duplication. Springer-Verlag; Berlin, New York: 1970.
- Piatigorsky J. Molecular biology: recent studies on enzyme/crystallins and alpha-crystallin gene expression. *Exp Eye Res.* 1990; 50:725–8. [PubMed: 2197110]
- Pohl BS, Knochel W. Temporal and spatial expression patterns of FoxD2 during the early development of *Xenopus laevis*. *Mech Dev.* 2002; 111:181–4. [PubMed: 11804794]
- Rebeiz M, Jikomes N, Kassner VA, Carroll SB. Evolutionary origin of a novel gene expression pattern through co-option of the latent activities of existing regulatory sequences. *Proc Natl Acad Sci U S A.* 2011; 108:10036–43. [PubMed: 21593416]
- Reiprich S, Stolt CC, Schreiner S, Parlato R, Wegner M. SoxE proteins are differentially required in mouse adrenal gland development. *Mol Biol Cell.* 2008; 19:1575–86. [PubMed: 18272785]
- Relaix F, Rocancourt D, Mansouri A, Buckingham M. Divergent functions of murine Pax3 and Pax7 in limb muscle development. *Genes Dev.* 2004; 18:1088–105. [PubMed: 15132998]
- Roth S, Stein D, Nusslein-Volhard C. A gradient of nuclear localization of the dorsal protein determines dorsoventral pattern in the *Drosophila* embryo. *Cell.* 1989; 59:1189–202. [PubMed: 2688897]
- Sato T, Sasai N, Sasai Y. Neural crest determination by co-activation of Pax3 and Zic1 genes in *Xenopus* ectoderm. *Development.* 2005; 132:2355–63. [PubMed: 15843410]
- Southard-Smith EM, Kos L, Pavan WJ. Sox10 mutation disrupts neural crest development in Dom Hirschsprung mouse model. *Nat Genet.* 1998; 18:60–4. [PubMed: 9425902]
- Stock DW, Ellies DL, Zhao Z, Ekker M, Ruddle FH, Weiss KM. The evolution of the vertebrate Dlx gene family. *Proc Natl Acad Sci U S A.* 1996; 93:10858–63. [PubMed: 8855272]
- Stolt CC, Lommes P, Friedrich RP, Wegner M. Transcription factors Sox8 and Sox10 perform non-equivalent roles during oligodendrocyte development despite functional redundancy. *Development.* 2004; 131:2349–58. [PubMed: 15102707]
- Tremblay P, Kessel M, Gruss P. A transgenic neuroanatomical marker identifies cranial neural crest deficiencies associated with the Pax3 mutant *Spotch*. *Dev Biol.* 1995; 171:317–29. [PubMed: 7556916]
- Van Otterloo E, Li W, Garnett A, Cattell M, Medeiros DM, Cornell RA. Novel Tfp2-mediated control of soxE expression facilitated the evolutionary emergence of the neural crest. *Development.* 2012; 139:720–30. [PubMed: 22241841]
- Wang J, Zhuang J, Iyer S, Lin X, Whitfield TW, Greven MC, Pierce BG, Dong X, Kundaje A, Cheng Y, Rando OJ, Birney E, Myers RM, Noble WS, Snyder M, Weng Z. Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. *Genome Res.* 2012; 22:1798–812. [PubMed: 22955990]
- Wang Q, Fang WH, Krupinski J, Kumar S, Slevin M, Kumar P. Pax genes in embryogenesis and oncogenesis. *J Cell Mol Med.* 2008; 12:2281–94. [PubMed: 18627422]
- Wang T, Zeng J, Lowe CB, Sellers RG, Salama SR, Yang M, Burgess SM, Brachmann RK, Haussler D. Species-specific endogenous retroviruses shape the transcriptional network of the human tumor suppressor protein p53. *Proc Natl Acad Sci U S A.* 2007; 104:18613–8. [PubMed: 18003932]
- Watanabe Y, Duprez D, Monsoro-Burq AH, Vincent C, Le Douarin NM. Two domains in vertebral development: antagonistic regulation by SHH and BMP4 proteins. *Development.* 1998; 125:2631–9. [PubMed: 9636078]

- Yu JK, Holland ND, Holland LZ. An amphioxus winged helix/forkhead gene, AmphiFoxD: insights into vertebrate neural crest evolution. *Dev Dyn.* 2002; 225:289–97. [PubMed: 12412011]
- Yu JK, Mazet F, Chen YT, Huang SW, Jung KC, Shimeld SM. The Fox genes of Branchiostoma floridae. *Dev Genes Evol.* 2008a; 218:629–38. [PubMed: 18773219]
- Yu JK, Meulemans D, McKeown SJ, Bronner-Fraser M. Insights from the amphioxus genome on the origin of vertebrate neural crest. *Genome Res.* 2008b; 18:1127–32. [PubMed: 18562679]
- Zhang P, Gu Z, Li WH. Different evolutionary patterns between young duplicate genes in the human genome. *Genome Biol.* 2003; 4:R56. [PubMed: 12952535]

Figure 1. Illustration of protein neofunctionalization and cis-regulatory element neofunctionalization in neural crest evolution

(A) Protein neofunctionalization. A transcription factor expressed in the NPB of the pre-vertebrate ancestor (shown in black) regulates a suite of genes through a set of enhancers (an example is shown as yellow gene with a black enhancer). The NPB transcription factor is duplicated leading to two paralogous genes in vertebrates, one paralog mutates in such a way that it can now regulate a new set of genes through a new set of enhancers (the green gene can now be regulated by the black and white striped transcription factor through the white enhancer). (B) cis-regulatory element neofunctionalization. A gene in the prevertebrate ancestor (shown in yellow) is expressed in non-NPB tissues under the regulation of a set of transcription factors (shown in black) binding its regulatory DNA (shown as a black enhancer). The ancestral gene is duplicated leading to two paralogs in vertebrates. The regulatory DNA of one of the paralogs changes in such a way that it can now bind a NPB transcription factor (shown in white) leading to its derived expression in the NPB.

Figure 2. Models depicting two variations of gene retention after duplication and protein neofunctionalization

(A) Force et al. proposed the Duplication Degeneration Complementation (DDC) model of gene retention (Force et al. 1999). 1) Subsequent to gene duplication (either whole genome or localized) there is 2) degeneration of cis-regulatory enhancer elements (Enhancer 1 and Enhancer 2 in figure) between duplicates, such that their expression patterns "complement" that of the original singleton ("subfunctionalization"). 3) Isolation to this more restricted expression domain then frees the duplicated protein to be further optimized for this isolated tissue ("neofunctionalization"). (B) In the Escape from Adaptive Conflict (EAC) model, initially tested by Hittinger and Carroll, 1) the singleton, prior to gene duplication, is already carrying out multiple functions at a suboptimal level ("gene sharing") (Hittinger CT & Carroll SB, 2007). 2) Subsequently, as a result of gene duplication, this conflict of suboptimal function is resolved as the duplicates can be further optimized for their isolated function. In the context of neural crest evolution, under either scenario, gene duplication facilitates the acquisition of a novel "neural crest promoting" function.

Figure 3. Phylogenetic distribution of neural plate border/neural crest gene expression in the *Tfp2*, *FoxD* and *SoxE* gene families

The gene trees for the (A) *Tfp2* (Van Otterloo et al., 2012), (B) *SoxE* (McCauley and Bronner-Fraser, 2006; Lakiza et al., 2011), and (C) *FoxD* (Yu et al., 2002; Sauka-Spengler et al., 2007) gene families are shown and the domains in which each gene is expressed are shown below the trees. Solid rectangles indicate where NPB/NC expression is likely to have arisen in each family. The origin of NPB/NC expression in the *Tfp2* family is ambiguous since *tfap2e* is expressed in NC derivatives (Van Otterloo et al., 2012) indicating that the

Tfap2 NC/NPB expression domain may have arisen in the *Tfap2a/b/c/e* ancestor. The dashed rectangles reflect this ambiguity. The trees were constructed using the neighbor-joining method with alignments of the full-length protein sequences. **Ciona FoxD* is expressed in the anterior neural tube (Imai et al., 2002), and in the a9.49 cell lineage which is at the NPB and may represent a rudimentary NC (Abitua et al., 2012). These regions may be equivalent to the diencephalon and NC expression domains, respectively.

Table 1

Summary of gene replacement experiments

Gene	Replacement	Strategy	Results*	Reference
Pax				
<i>Pax3 (Mus musculus)</i> <i>Pax</i> subfamily III	<i>Pax7 (Mus musculus)</i> <i>Pax</i> subfamily III	KI	Functionally equivalent	Relaix et al. 2004
<i>Pax3 (Mus musculus)</i> <i>Pax</i> subfamily III	<i>Pax8 (Mus musculus)</i> <i>Pax</i> subfamily II	KI	Unable to compensate	Hayashi et al. 2011
<i>Pax3 (Mus musculus)</i> <i>Pax</i> subfamily III	<i>AmphioxusPax (Subfamily III) or Trichoplax adhaerens Pax</i>		?	
Sox				
<i>Sox10 (Mus musculus)</i> <i>SoxE</i> subfamily	<i>sox100B (Drosophila melanogaster)</i>	KI	Mostly capable of compensating – derivatives not fully rescued	Cossais et al. 2010
<i>Sox10 (Mus musculus)</i> <i>SoxE</i> subfamily	<i>Sox8 (Mus musculus)</i> <i>SoxE</i> subfamily	KI	Partially capable of compensating – No rescue of melanocytes	Kellerer et al. 2006
<i>Sox10 (Mus musculus)</i> <i>SoxE</i> subfamily	<i>SoxE(Amphioxus)</i>		?	
Tfap2				
<i>Tfap2a/Tfap2c (Danio rerio)</i>	<i>AP-2(Amphioxus)</i>	KD/OX	Functionally equivalent (for markers assessed)	Van Otterloo et al. 2012
<i>Tfap2a/Tfap2c (Danio rerio)</i>	<i>AP-2 (Drosophila melanogaster)</i>	KD/OX	Mostly capable of compensating – less efficient at restoring NC derivatives (ones assessed)	Van Otterloo et al. 2012

Abbreviations: KI, knock-in; KD, knock-down; OX, over-expression

* Rescue results in the context of neural crest development

Red text, not yet done