UNIVERSITY OF OXFORD

Reverse Engineering Gene Regulatory Circuitries in Neural Crest Development and Evolution

A thesis submitted for the degree of
D.Phil. in Medical Sciences

at
St Johns's College

Author: Ivan L. Candido-Ferreira
Supervisor: Professor Tatjana Sauka-Spengler

November 2021

Abstract

Reverse Engineering Gene Regulatory Circuitries in Neural Crest Development and Evolution

Ivan Candido-Ferreira, University of Oxford, St John's College

D.Phil. in Medical Sciences, November 2021

The neural crest (NC) is an emblematic population of embryonic, stem-like cells characterised by their unique multipotency, migratory behaviour and broad developmental potential. Acquisition of these cells represents one major evolutionary transition in the tree of life, endowing vertebrates with their unique body plan.

A gene regulatory network (GRN) has been proposed to orchestrate NC development. To interrogate the NC-GRN at unprecedented resolution in a genome-wide fashion, we employed single cell transcriptomics and epigenomics at multiple stages of avian NC development. Development of a new pipeline allowed us to reverse engineer the NC-GRN in an unbiased manner. This approach enabled us to identify the *cis*-regulatory architecture embedding the logic of NC ontogeny. We identified regulatory circuitries involved in the pre-patterning of the early NC. Interrogation of the combinatorial *cis*-regulatory codes encoded in these enhancers revealed opposing circuitries controlling NC fate decisions.

We next compared the transcriptional heterogeneity of NC cells at single cell resolution across three-species: mouse, chicken and zebrafish. Our results indicate conservation of transcriptional states despite extensive divergence of genomic sequences across these species. Moreover, we also pioneered single-cell transcriptomics in the marine lamprey, enabling us to identify novel genes likely involved in lamprey NC development.

Finally, we interrogated the *cis*-regulatory landscape of zebrafish and lamprey NC cells, revealing conserved properties between chicken, zebrafish and lampreys, such as motif content, enhancer-promoter communications and dynamics of chromatin accessibility during development. We identified a group of pleiotropic enhancers involved in NC-dependent neurogenesis. We propose these elements have played a pivotal role in the early elaboration of multipotent NC in the stem vertebrate lineage via pleiotropic changes in ancestral enhancer elements. Taken together, this thesis provides important insights into the *cis*-regulatory basis of NC development and evolution.

Contents

List of Figures

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university.

This dissertation is my own work, and I am fully responsible (1) for all computational analyses of RNA-seq, ATAC-seq, ChIP-seq, scRNA-seq and scATAC-seq data, (2) development of the pipeline that enabled the integration of *cis*-regulatory sequences with scRNA-seq which was used to reverse engineer developmental regulatory circuitries, (3) enhancer dissection *in silico*, (4) motif content identification, (5) differential accessibility analysis and K-means clustering, (6) development of the analysis pipeline for combinatorial enrichment of motifs within enhancer elements, (7) Sox10 ChIP-seq optimisations, experiments and construction of universal recipient cloning vectors for biotin ChIP-seq, (8) cloning and *in vivo* validation of *Sox5*, *Fzd3* and *Chd7* chicken enhancers, screening and confocal imaging, (9) generation of zebrafish ATAC-seq profiles, (10) multimodal single cell analysis and integration of single cell transcriptomes across species.

Exception is made in the text and in the following:

- Avian NC RNA-seq, ATAC-seq, Capture-C, H3K27Ac ChIP-seq and scRNA-seq were conducted in collaboration with Dr. Ruth Williams.
- Confocal imaging figures 2.1, 2.12, 2.14, 2.16, 2.33, 2.34, 3.1, 3.27 and 7.11 resulted from collaborations with Dr. Ruth Williams, Dr. Martyna Lukoseviciute and Dr. Irving Ling.

- Weighted Gene Co-Expression Network Analysis (WGCNA) analysis was conducted in collaboration with Dr. Daria Gavriouchkina.
- Sox10 ChIP-seq was performed in collaboration with Dr. Ruth Williams.
- Dissection of the avian *CHD7* locus was performed in collaboration with Prof. Tatjana Sauka-Spengler, Dr. Ruth Williams, Dr. Gunes Taylor and Dr. Irving Ling.
- Bilateral electroporations for CRISPR/Cas9-mediated knockout and RNA-seq experiments were performed by Dr. Ruth Williams.
- Capture-C experiments were designed and performed by Dr. Ruth Williams, and Dr. Jelena Telenius and prof. Jim Hughes contributed with Capture-C data analysis.
- Analysis of the SmartSeq2 scRNA-seq dataset and construction of the interactive ShinyApp was conducted in collaboration with Dr. Emmanouela Repapi.
- Zebrafish NC scRNA-seq experiments were performed in collaboration with Dr. Martyna Lukoseviciute.
- Lamprey scRNA-seq, ATAC-seq and cross-species enhancer reporter assays were performed in collaboration with Dr. Ruth Williams and Dr. Dorit Hockman.

Ivan Candido-Ferreira, November 2021

List of Abbreviations

3C - Chromatin Conformation Capture

3D - Three Dimensions

ATAC-seq - Assay for Tranposase Accessible Chromatin followed by sequencing

BAFs - BRG1/BRM associated factors

BMP - Bone Morphogenetic Proteins

Cas9 - CRISPR-associated protein 9

CBEs - chromatin boundary elements

CBP/p300 - cAMP-regulated-enhancer (CRE)-binding protein (CREB)-binding protein

ChIP-seq - Chromatin Immuno-Precipitation followed by sequencing

cl - cluster

CRISPR - Clustered Regularly Interspaced Short Palindromic Repeats

CTCF - CCCTC-binding factor

CUT&RUN - Cleavage Under Targets and Release Using Nuclease

DA - Differential Accessibility

di - diencephalon

DP - Double positive population of EdnrB+ and FoxD3+ vagal NC

E2 - EdnrB+ population of vagal NC

ECM - extracellular matrix

EMT - Epithelial-to-Mesenchymal Transition

Enh - Enhancer

FACS - fluorescence-activated cell sorting

FGF - Fibroblast Growth Factor

FNP - frontonasal proemience

G2M - G2-phase checkpoint

GO - gene ontology

GRN - gene regulatory network

gRNAs - guide Rnas

hb - hindbrain

HH - Hamburger Hamilton

mb - midbrain

mes - mesencephalon

MHB - midbrain-hindbrain boundary

NC - neural crest

np - nodose placode

NPB - Neural Plate Border

ot - otic placode

PA1-6 - pharyngeal arches 1-6

Pc - peak centre

PNS - Peripheral Nervous System

PRC1/2 - Polycomb Repressive Complex 1/2

r1-r8 - rhombomeres 1 to 8

RNA-seq - RNA sequencing

SAMAP - self-assembling manifolds across species

SE - super-enhancer

spc - spinal cord

ss - somite stage

STARR-seq - self-transcribing active regulatory region followed by sequencing

TADs - Topologically Associating Domains

TEs - transposable elements

TF - transcription factor

TFBS - transcription factor binding sites

tNC - trunk NC

UAS/Gal4 - GAL4/upstream activating sequence system

UTR - untranslated region

WGCNA - Weighted Gene Co-expression Network Analysis

WNT - Wingless/Integrated

1 Introduction

1.1 Regulatory control of NC ontogeny

The incredible biodiversity of shapes and forms has always captivated humanity. Such staggering diversity is illustrated by the multitude of unique body plans and pigmentation patterns in chordates, the group comprised of filter-feeding invertebrates such as amphioxus and sea squirts, as well as vertebrates [1]. The remarkable complexity and diverse cellular make up of metazoans evolved through the accurate regulation of when, where and to what extent each gene is expressed during embryonic development, giving rise to cell- and tissue-specific transcriptional programmes [2–5].

Precision and robustness of such mechanisms regulating the emergence of cell identities are encoded in genomes in the form of gene regulatory networks (GRNs). GRNs engage signalling pathways, morphogens, transcription factors (TFs) and effector molecules that form the *trans* regulatory environment, as well as DNA responsive elements collectively known as enhancers, constituting a *cis*-regulatory network dynamically remodelled and bound by chromatin regulators, TFs and nuclear receptors during development [4–8]. The interplay of such *cis* and *trans* regulatory components, together with the biophysical sculpting of the developing embryo, hierarchically establishes body plans and direct complex patterns of cell-type specific gene expression during metazoan embryogenesis [4–6].

In this Chapter, I discuss the role of genomic regulatory mechanisms in development and evolution that shape the identity of the neural crest (NC), an iconic population of multipotent progenitors with remarkable migratory ability in the vertebrate embryo. NC cells evolved during the radiation of vertebrates and uniquely contribute to the vertebrate ”new head” [9, 10]. NC is an attractive and experimentally tractable system that can advance our understanding of the role of developmental GRNs in evolutionary transitions, such as the emergence of novel cell-types and unique body plans. In fact, the fantastic craniofacial adaptations of birds, which captivated Darwin and many generations of biologists, are derived directly from adaptive diversifications of the NC (reviewed in [11]). However, the genomic mechanisms underlying such diversity remain poorly understood.

Finally, I take a comparative approach within a phylogenetic framework, to first outline findings from molecular embryology, genomics and comparative morphology in jawed vertebrates, which revealed conserved processes regulating NC ontogeny, and then briefly discuss how these processes are implicated in human disease. I then compare these findings to functional and genomics studies in the closest living relatives of jawed vertebrates, lampreys, hagfish, urochordates, cephalochordates and hemichordates,respectively, which in turn revealed incremental changes underlying the appearance of NC. I discuss how mechanisms of regulatory divergence in *cis* and *trans* have facilitated the evolutionary origins and diversification of the NC, and outline exciting novel functional genomic and computational approaches that are poised to advance our understanding of NC role in developmental and evolutionary dynamics.

1.2 Brief history of the NC

Discovered more than 150 years ago by Wilhelm His in the chicken embryo, the NC was initially named ”zwischenstrang”, German for ”intermediate cord” [12]. Due to the observation that these cells are located above the neural tube (at the ”crest”), extensively migrate and give contribute to the formation of spinal ganglia and cranial nerves, this transient cell population was later named neural crest [13]. Pioneer studies conducted by Julia Platt in the salamander embryo in the late XIX century suggested a controversial hypothesis: NC cells also contribute to cranial skeleton and cartilage [14]. However, the scientific community readily dismissed her results and ideas, given the dominant "germ layer doctrine" that proposed mesoderm as the single tissue of origin of all skeletal elements [15].

Such controversy was only resolved beyond doubt decades later by the elegant fate mapping experiments conducted by Nicole Le Douarin and her team [15–23]. Nicole Le Douarin used heterospecific graft experiments in which quail NC was transplanted to stage-matched chicken host embryos (quail-chick chimeras). The morphologically and immunochemically distinct quail nuclei in derived tissues of embryos or juvenile chickens confirmed and greatly extended our knowledge of the unique contribution of the cephalic NC to cartilage and skeletal elements of the head [19–21, 23]. More recent lineage tracing experiments [24], as well as unbiased multi-omic approaches [25], are consistent with the idea that early NC cells are multipotent [26, 27], and broadened our understanding of the NC contribution to a plethora of derivatives [15, 17, 19–22, 26–30].

Tight control of such complex developmental trajectories involves the cross-talk of multiple regulatory elements, including morphogens and their antagonists, as well as intrinsic factors expressed in the NC, collectively forming the NC *trans*-regulatory environment [31–33]. Such *trans* components interact and can be dynamically controlled by developmental *cis*-regulatory elements such as enhancers, which encode the genomic logic of developmental programmes [8].

1.2.1 NC induction and specification

The NC is a transient, stem cell-like migratory embryonic population of vertebrates uniquely contributing to vertebrate body plan [15, 19, 23, 29, 34–36]. NC induction takes place within the neural plate border at the end of gastrulation, following instruction from extracellular signals such as Wnt1/Wnt4, Fgf10/Fgf3, Bmp2/3/4 and Smad3, emanating from both the environing ectoderm and underlying mesoderm [19, 31–33]. In fact, intermediate levels of BMP specify the neural plate border (NPB) in a manner similar to the classical "French Flag" model (Figure 1.1a), [37]. NPB cells will then give rise to NC cells and cranial placodes [31, 37, 38]. However, whether NPB cells are pre-patterned towards NC versus placodal fates remain a matter of debate [39–41]. These upstream signalling pathways induce the activation of NPB specification factors such as Gbx2, Zic, Pax3/7, Prdm1 and Msx1/2, which in turn instruct the activation of NC and placodal specifiers. NC specification genes include TFAP2a/c, FoxD3, Id, Snai1/2 and Sox9/10, and are reciprocally repressed by neural plate and pro-placodal factors, such as Sox2/3 and Six1/2, respectively (Figure 1.1e). NC cells then undergo epithelial-to-mesenchymal transitions (EMT), acquiring the ability to extensively migrate and colonise distal sites of the embryo [31–33, 38, 42, 43].

Classical embryology approaches and lineage tracing experiments have identified the extensive contribution of the NC to sensory, adrenergic and cholinergic neurons, melanocytes, cartilage and skeletal elements of the head, smooth and striated muscles of the heart, Schwann cells, enteric and sympathetic ganglia, among many others [19, 20, 23, 24, 26, 27, 45, 46]. Because of their unique contribution to evolutionary novelties of vertebrates such as jaws and sensory neurons from the peripheral nervous system (PNS), the NC is thought to have played a central role in the transition from a filter feeding to a predatory lifestyle in the common ancestor of vertebrates [9, 18, 47]. Remarkably, some of these derivative cell types also originate from mesoderm, such as chondroblasts and osteoblasts, suggesting that mesodermal-like GRN subcircuits involved in mesenchymal cell differentiation have been partially co-opted into the NC GRN [47, 48]. Henceforth, NC-derived

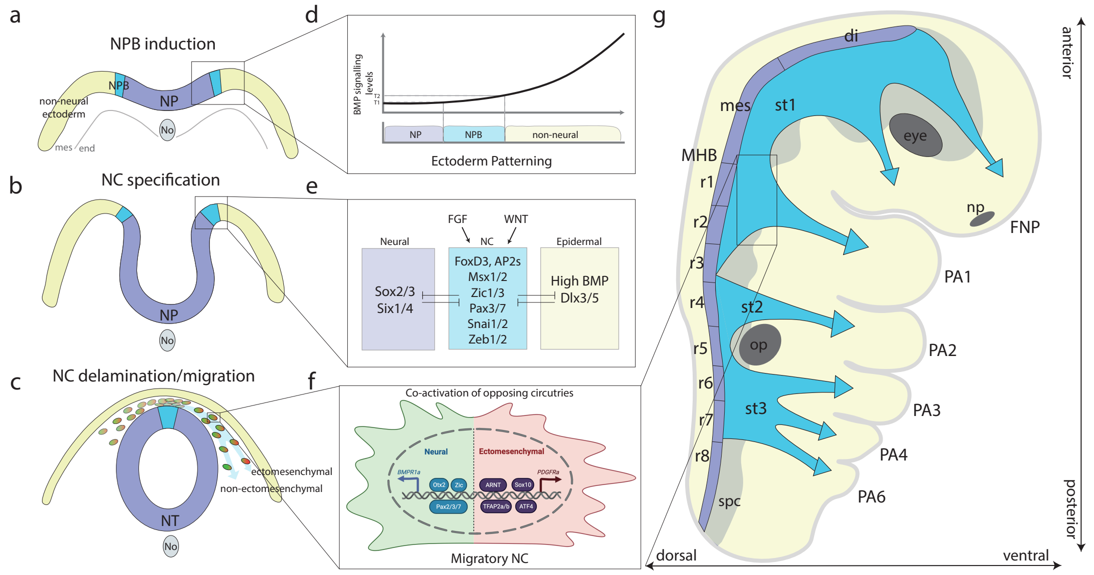

Figure 1.1 Regulation of NC formation. a) Neural plate border (NPB) induction, in between the neural plate (NP) and non-neural ectoderm is patterned early by varying levels of BMP signaling, in a manner consistent with the ”French flag” model. (b) NC specification is regulated by the combinatorial action of transcription factors (TFs). (**a-c**) Early NC ontogeny, from induction to specification and delamination. Gene regulatory mechanisms underlying these processes are showed next to each panel. (**d**) Neural plate border (NPB) induction, in between the neural plate (NP) and non-neural ectoderm is patterned early by varying levels of BMP signaling, in a manner consistent with the â€œFrench flagâ€ model. (**e**) Mutual repression between presumptive neural plate and NC territories establish the sharp boundaries. (**f**) Activation of opposing circuits establish early heterogeneity in delaminating NC. (**g**) NC migration along stereotypical streams in a generalised gnasthosome embryo. di - diencephalon; mes - mesencephalon, MHB - midbrain-hindbrain boundary; PA1-6 - pharyngeal arches 1-6; FNP - frontonasal proemience; np - nodose placode; op - optic placode; ; r1-r8- rhombomeres 1 to 8; spc - spinal cord. Adapted from [44].

mesenchyme will be termed ectomesenchyme as a means of distinguishing from mesenchymal cell-types derived from the mesoderm.

An impressive body of literature details the epistatic relationships between regulators involved in NC development in jawed vertebrates revealed by bottom-up approaches, which enabled proposing a hierarchical and modular model depicting the NC GRN from candidate regulators (Figure 1.2), (extensively reviewed in: [31–33, 49]). This model outlines the key circuits controlling NC induction, specification, multipotency, proliferation, migration and downstream cell fate decisions. However, unbiased, high-throughput, top-down systems-level interrogation of the NC has only been made possible in the last decade by technological advances enabling high throughput transcriptomic and epigenomic profiling, thus allowing for the identification of *cis* and *trans* regulators of NC development genome-wide in an unbiased fashion. These studies identified an extensive cohort of NC-associated chromatin modifiers, such as ATP-dependent BRG1/BRM associated factors (BAFs, also known as SWI/SNF chromatin remodelling complex in mammals), including CHD7, NC specifiers such as Snai1/2, Zeb1/2, Twist1/2, ZFHX4, Myc, nuclear receptors (such as NR2F1/NR2F2), lineage-determining factors (e.g., Olig2, Ascl1, Mitf) and effector molecules involved in migration, colonisation and adoption of developmental fates, such as EdnrA/B, Cxcr4, Ret and PdgfrA [50–61].

1.2.2 Schrodinger's crest

A major controversy arising from NC lineage tracing studies pertains to whether the NC comprises a population of homogeneous multipotent precursors or lineage-restricted oligopotent precursors (reviewed in [22]). *In vivo*, NC cells faithfully follow stereotypical patterns of differentiation, suggesting they are pre-patterned before colonising distal sites of the embryo. The first cells to delaminate from the neural tube follow a dorsal to ventral order during their colonisation; for example, the first cells delaminating from the trunk/sacral region gives rise to sympathetic ganglia, followed by Schwann cells, sensory ganglia and finally melanocytes [45, 46]. A core regulatory circuit involving the TFs FoxD3, Sox9 and Snai2 in the avian NC specifies restriction in lineage potential [63]. However, under experimental conditions, partial ablation of NC leads to the compensation from neighbouring cells, which are able to deflect their migratory routes, proliferate and fill in for the missing cells, demonstrating the remarkable ability of the NC to compute complex cellular decisions and acquire plasticity during development [29, 64]. Interestingly, the identification of melanocytes differentiating from Schwann cells precursors il-

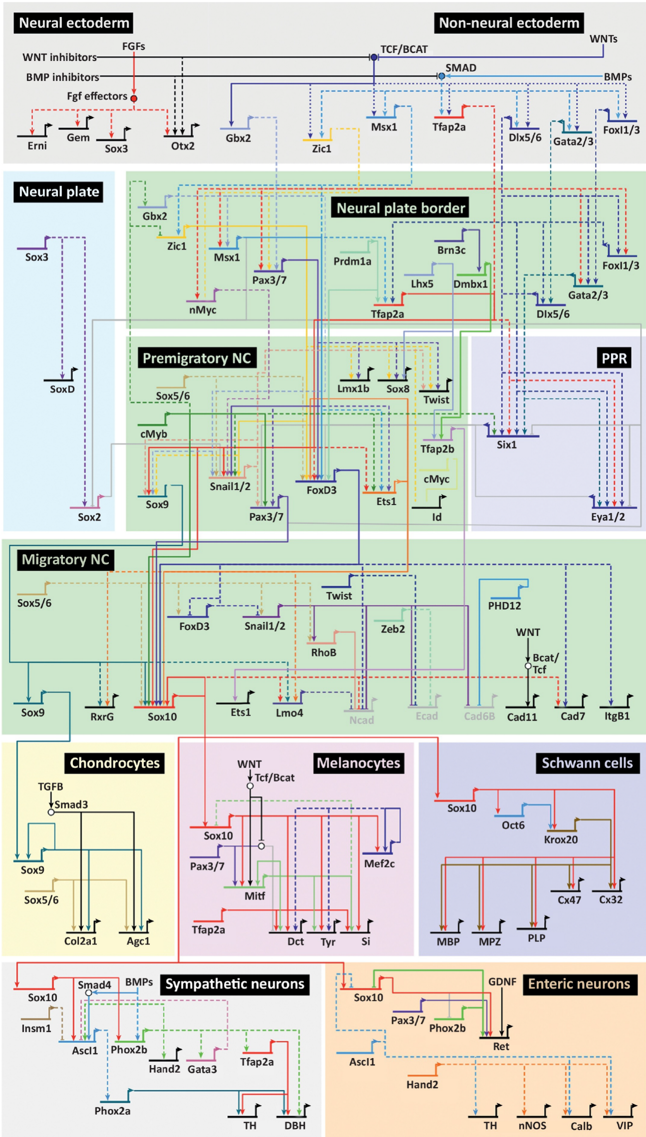

Figure 1.2 The provisional gene regulatory network orchestrating NC development. Gene regulatory interactions have been implied via *cis*-regulatory analysis of candidate genes (see text). This model, illustrated in Biotapestry format, highlights the topology and the central TFs regulating NC formation. Top in the GRN hierarchy is displayed the signalling molecules and their downstream targets (e.g., *Otx2*, *Gbx2*, *Zic1*, *Msx1* and *Tfap2a*) patterning the ectoderm into neural and non-neural domains. Upon establishment of the neural plate border, which express *Brn3c*, *Lhx5* and *Dmbx1*, this territory will later give rise to pro-placodal regions (PPR) and the pre-migratory, FoxD3+ NC. After EMT and delamination, Sox10+ NC cells become migratory. These cells then colonise distant sites within the embryo and differentiate into a plethora of derivatives, illustrated here by several derivatives: chondrocytes (Sox9+), melanocytes (Mitf+), Schwann cells (Sox10/Oct6+), sympathetic neurons (Ascl1/Phox2B+) and enteric neurons (Hand2/Ret+). Adapted from [32, 62]

lustrates the ability of some glial cells to retain NC-like oligopotency [65]. Such "Schrodinger's cat"-like paradoxical behaviour of the NC, representing both multipotent and pre-patterned precursors, is a matter of intensely debated within the literature (*e.g.*, [22, 24]).

Single cell transcriptomics and population-based epigenomic approaches have recently showed heterogeneities in the early zebrafish [60], *Xenopus* [66], avian [58, 59] and murine NC [25, 67] at the onset of delamination and during early migration. These studies suggest that subpopulations of transcriptionally- and epigenomically-biased NC pools harbour information of their future fates *in vivo*. Indeed, single cell profiling of the murine and avian NC at single cell resolution resolved the early developmental trajectories at the cranial and trunk axial levels, indicating combinations of TFs driving developmental fates [25, 58, 59]. For example, co-expression of Phox2b and Ascl1 intrinsically biases migratory cells towards neuronal/glial fates [25, 58].

Importantly, NC specification progresses in an anterior-to-posterior (rostro-caudal) fashion within the embryo. However, not all axial levels are capable of producing the same derivatives: for example, the cranial NC of jawed vertebrates is uniquely endowed with the potential to contribute to ectomesenchymal derivatives, such as cartilage and bone within the head, whereas vagal NC contributes neurons and glia of the enteric nervous system [10, 26, 27, 68, 69]. Interestingly, NC from all axial levels can give rise to melanocytes [17, 18, 22], suggesting that the transcriptional programme controlling pigment cell differentiation is evolutionary ancient and deeply rooted within the NC GRN, possibly predating axial regionalisation of the NC [47]. Moreover, derivatives from trunk NC are largely depleted of ectomesenchymal potential, and in heterotopic grafting experiments, the majority of NC trunk neuronal derivatives can be contributed by the NC cells grafted from other axial levels [19, 20, 22, 27]. Additionally, identification of a territory within the trunk region, but not the cephalic, homologous to presumptive NC in invertebrate chordates [70] suggests that the ancestral NC may have originated first as a compartment within the trunk neural plate border, which later invaded the cranial region. Therefore, the ground state of NC cells may be reflected by NC cells within the trunk level, with additional axial-regionalisation evolving within vertebrates [10, 47].

The axial-specific heterogeneity of NC cells is tightly controlled by expression of different combinations of TFs - markedly homeobox genes. The homeodomain proteins Dmbx1 and Otx2 instruct the early specification of cranial NC identity at the midbrain level [10, 55, 62, 69]. The posteriorising factor Gbx2, a repressor of *Otx2* involved in the midbrain-hindbrain boundary formation, neural plate caudalisation and suppression of placode specification [71], also instructs the formation of a trunk-like identity in post-otic NC [67]. Moreover, Cdx1/2 acting upstream of *Pax3/7*, *Msx1/2* and *FoxD3* determines trunk NC [72, 73]. Collinear expression of *Hox* genes patterns the post-otic NC, with anterior Hox factors (e.g., HoxA2, HoxB1/B2) patterning hindbrain NC, and subsequent acting Hoxes instruct the identity of the vagal (e.g., HoxB3, HoxA4) and trunk (e.g., HoxB5) NC [74, 75]. In fact, Hox TFs have a well-established role in repressing ectomesenchymal fates, as indicated by gain-of-function experiments in Hox depleted regions of the cranial NC [76]. Such spatiotemporal regulation of axial-specific transcriptional programmes highlights principles of GRN modularity within the NC GRN and indicates how such distinctive modules restrict cell fate potential. For example, ectopic delivery of a cranial-specific circuit (consisting of Sox8, TFAP2b and Ets1) to the trunk NC instructed partial reprogramming of the avian trunk NC, resulting in ectopic chondroblast differentiation [56].

The unique ability of the cranial NC of jawed vertebrates to contribute to craniofacial ectomesenchyme has been linked to the coupling of the cranial NC GRN to the core pluripotency transcriptional network, whereby cranial NC was hypothesised to either acquire mesoderm-like potential *de novo* or retain such potency from early gastrulation [66, 67, 77]. Consistent with these hypotheses, depletion of VENTX/NANOG and POU5/OCT4 in *Xenopus* NC disrupts ectomesenchymal development but does not disrupt NC-derived neural- and pigmented cell-type derivatives [78]. Moreover, single cell profiling of presumptive murine cranial NC revealed that these cells transiently reactivate core Yamanaka pluripotency factors (Nanog, Oct4 and Sox2) [67]. Thus, an *in vivo* programming event seems to redirect cranial NC towards multipotency, endowing these cells with their unique ability to generate mesoderm-like ectomesenchyme and pattern the vertebrate head, thus suggesting multipotency is acquired *de novo* during development, possibly through the evolution of novel distal-acting *cis*-regulatory elements [47].

An elegant study conducted on murine NC has provided further evidence that cranial NC cells encode such developmental plasticity at the pre-migratory stage through modifications at the chromatin level [79]. Conditional knockout of methyltransferase *ezh2* (*enhancer of zeste homolog 2*), a core component of the Polycomb Repressive Complex 2 (PRC2), in the cranial region of the murine NC, which is devoid of Hox genes expression, mimics the phenotype of HoxA2, HoxA3 and HoxB4 overexpression in these cephalic subpopulations [76]. These mutants display inhibition of almost all development of NC-derived craniofacial bone and cartilage [79]. In wild type embryos, distinctive sets of positional gene promoters (e.g., *Hoxa2*, *Dlx5*/*Dlx6*, or *Hand2*) and their *cis*-regulatory elements display similar levels of accessibility, a requirement for enhancer activity. Yet these genes are expressed at different levels in different positions of the cranial region, and thus patterned, transcriptionally heterogeneous NC progenitor subpopulations will give rise to the frontonasal, maxillary and mandibular processes. How are these genes expressed heterogeneously, when the chromatin topology shows a homogeneous permissive landscape? These loci are marked by bivalent histone modifications; chromatin at their promoters is trimethylated on lysine 27 of histone H3 (H3K27me3), (mark associated to transcriptional repression deposited by PRC2), but they are also, at the same time, marked by modifications associated to transcriptional activation, such as histone H3 trimethylation of lysine 4 (H3K4me3) [80]. These elements and several other gene promoters are maintained in such a poised state during cephalic NC migration, resulting in rapid onset of different sets of positional genes at the post-migratory stage, likely due to the erasure of repressive H3K27me3 mark in combination with the rapid activation of transcription. These results suggest that differential deposition of chromatin bivalency at distinct positional genes precedes NC axial regionalisation. Maintenance of chromatin accessibility together with chromatin poising for transcriptional activation in these positional genes allow for their rapid deployment upon erasure of repressive states instructed by changes in the *trans* environment during early migration. Therefore, differential chromatin dynamics offers an attractive mechanism reconciling the paradoxical observation of fate-restriction in the NC, yet retaining a plastic, multipotent behaviour.

Consistent with these findings, a forward genetic screen in zebrafish identified an *h3f3a* mutant [81] displaying a ”sucker” phenotype [82], reminiscent of jawless vertebrates. Such mutants lack all cranial NC derivatives, including head pigmentation and jaws, but do not display defects in the trunk, or major abnormalities in embryonic development outside of the cranial NC. *H3f3a* encodes one of the five histone 3 variants H3.3 in the zebrafish genome, suggesting a surprisingly unique role for this H3.3 histone variant in poising cranial NC towards ectomesenchymal fates, further suggesting a central role of differential chromatin regulation in determining NC ectomesenchymal potential and axial regionalisation.

Moreover, genome-wide chromatin profiling of avian and murine NC enabled the identification of distinctive sets of enhancers controlling loci associated to NC multipotency and fate decisions [58, 59, 67]. These enhancers display differential spatio-temporal accessibility and activity, motif content, posttranslational histone modifications and association to target promoters predictive of the future transcriptional state (Figure 1.3), thus defining or at least biasing developmental fates earlier than previously thought. For example, a subset of cranial-specific enhancers is associated to ectomesenchymal fates, whereas enhancers shared between different axial levels are associated to neuronal/glial fates [58, 59], thus suggesting the existence of a *cis*-regulatory architecture pre-patterning lineage decisions of the NC.

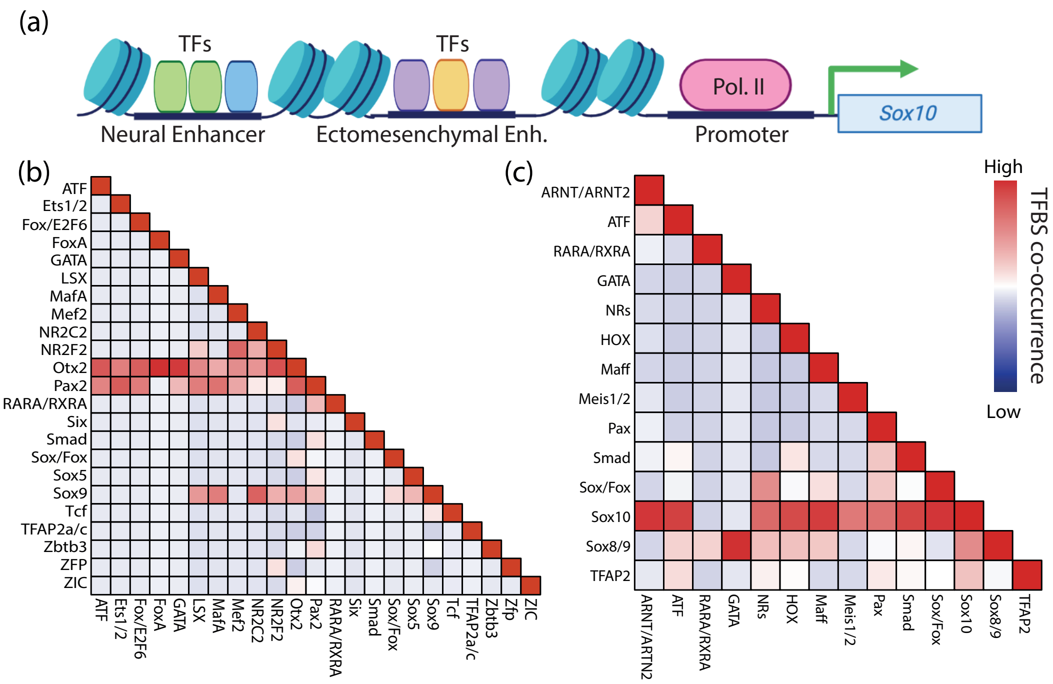

Figure 1.3 Combinatorial regulation of cranial NC development by distinct *cis*-regulatory architectures. (a) Enhancers regulating expression of NC regulators, such as *Sox10*, display signatures of distinct developmental fates (*e.g.*, neural and ectomesenchymal). One signature, for example, is the putative co-binding of distinct combinations of TFs and lineage-determining factors. TFs depicted in green and blue represent factors involved in neural differentiation; TFs depicted in purple and orange represent regulators involved in ectomesenchymal fates, such as skeletogenesis. (b) Co-occurring TF motif binding sites enriched in these putative neural and (c) ectomesenchymal enhancers in the avian NC, reflecting the distinct *cis*-regulatory logic involved in the early control of NC developmental fate. TFBS, transcription factor binding sites. Based on data from Williams, Candido-Ferreira, *et al* [59].

Heterogeneity of the chromatin landscape in subpopulations residing at the same axial level (for example, cranial or vagal) points towards cell-autonomous epigenomic remodelling as a mechanism underlying early poising of developmental fates. Indeed, in our recent study of the avian cranial NC [59], two major groups of enhancers were identified (Figure 1.3a). The first group represents NC-specific enhancers displaying binding motif combinations of canonical NC factors, such as Sox10 and TFAP2. However, these enhancers are also remarkably enriched for binding sites of hypoxia-inducible factor HIF-1b, also known as Aryl Hydrocarbon Receptor Nuclear Translocator (ARNT, [83]), (Figure 1.3c), in agreement with previous functional studies suggesting the role of hypoxia-inducible genes in NC cells proliferation, migration and differentiation into ectomesenchyme [84–86]. These enhancers also regulate loci involved in ectomesenchymal fates, such as *PdgfrA*, suggesting activity of these enhancers is indeed associated with chondrogenesis and skeletogenesis. However, enhancer-mediated hypoxia-responsiveness is likely not limited to cranial-specific processes [86].

Another group of enhancers is decompacted and established early, as indicated by their accessibility in presumptive NC, and likely maintained in Otx2+ NC [59]. These enhancers are also shared with other cranial cell-types such as neurogenic placodes and neural plate, possibly indicating that an early, neuroepithelial NC state is regulated by a pleiotropic *cis*-regulatory logic dependent on Otx2, divergent from that of ectomesenchymal enhancers (Figure 1.3b) and may represent regulatory elements controlling the transcriptional network of neurogenic NC at the midbrain level [59]. Interestingly, we note that the genomic arrangement of such elements reflects the developmental order of NC differentiation. For instance, within the *Sox10* enhancer cluster located upstream of the gene promoter, the neural enhancers, located in the distal-most region of the Sox10 TAD, are the earliest to be activated followed by subsequently acting ectomesenchymal enhancers. It is therefore plausible that the genomic order of enhancers within a *cis*-regulatory cluster may reflect their order of activation, reminiscent of the organisation of *Hox* gene clusters [87, 88].

1.2.3 NC self-renewal, EMT and migration

The size of the NC migratory population is regulated prior to NC delamination from the neural tube [89]. The oncogenic TF cMyc, regarded as a late NC specifier [32, 38], seems to control NC self-renewal during the pre-migratory stage [32, 33]. However, migratory NC seems to have limited ability to proliferate [90], further indicating that such ability is transitory and limited to a niche of pre-migratory cells. However, the regulatory circuitry underlying key properties of multipotent cells, such as self-renewal and a sustained capacity to proliferate, remains elusive in the NC.

Nevertheless, a recent study suggests that cMyc controls NC self-renewal but not its multipotency or proliferating capacities in chicken embryos [89]. Limited self-renewal potential caused a reduction in the size of the migratory progenitor pool, as indicated by morpholino knockdowns. Importantly, it was proposed that cMyc acts not by binding to its canonical E-Box sites through heterodimerisation with the basic helix-loop-helix leucine zipper TF, Max, but rather via a non-canonical interaction with Miz1 *in vivo*. Although such switch from the activator heterodimer cMyc/Max to the repressor complex cMyc/Miz appears to underlie NC self-renewal capacity *in vivo*, the authors reported conflicting results *in vitro* [89]. Therefore, further studies probing cMyc binding genome-wide *in vivo* are needed to comprehensively reveal its targets and investigate to what extent dynamic recruitment of co-binding partners, such as Max and Miz factors, regulates different circuitries.

Once NC is specified and the size of the migratory progenitor pool is established, NC cells show a notable ability to delaminate by undergoing EMT. Firstly, NC cells reshape their cellular properties, which allow for their delamination from the neural tube, acquiring mesenchymal properties and migrating to distant sites in the developing embryo [108]. Remodelling of the extracellular matrix (ECM) involves a cadherin switch, in which pre-migratory NC cells expressing type I cadherins (for example, Cadherins 1-4) start expressing type II cadherins (Cadherin 6/7/10/11) during migration [91]. The onset of EMT is controlled by Zeb1/2, Snail1/2, FoxD3, mediating the repression of type I cadherin genes, whereas transcriptional activators such as Pax3/7, Sox9/10, Ets1 and Twist1/2 induce the acquisition of the mesenchymal-state, including the acquisition of type II cadherins and several ECM modifiers, such as ADAM10/13 metalloproteases [92–95].

In the cranial regions, NC assembles into three main migrating streams, the trigeminal, hyoid and branchial (Figure 1.1g)[44]. These streams are thought to be pre-patterned by inhibitory signals, such as semaphorins secreted by the cranial placodes, rhombomeres and surrounding mesenchyme [34, 42–44]. However, recent computational modelling not taking into account any inhibitory signals suggested that the biomechanical interactions between the NC and placodes are sufficient to cause NC streams to emerge. It is therefore possible that streaming may result primarily from the mechanical interactions between NC and placodes, aand that further inhibitory signals may serve to canalise such streams, increasing the robustness of NC stereotypical migration [96]. In addition to semaphorins, NC migration in streams seems to be also regulated by Ephrin-Eph signalling, plexins, neuropilins, and Slit-Robo interactions (reviewed in [44]). Moreover, during migration, endothelin receptors A and B (EdnrA/B) seem to play a predominant role. A large population of migratory NC cells expresses those receptors, and are guided by the environing cells expressing cognate ligands, such as endothelin-3 [97]. Such central role in guiding the migration and colonisation of the NC by EdnrA/B is further supported by their function in the development of the NC-derived enteric nervous system [58].

Such extraordinary ability to sense multiple developmental cues, navigate the developing embryo and perform logical operations illustrate the unique role of membrane receptors in the tight control of NC differentiation. Again, expression of such receptors and signalling molecules is controlled by combinations of lineage-determining TFs, often in an axial-specific manner [10, 32, 33, 55, 62, 69]. For example, Sox10 regulates multitude of loci involved in the migratory state (e.g., *EdnrB*, *PdgfrA*, among many others), and acts in combination with other activator TFs, such as TFAP2a/b, establishing feed-forward loops that increase robustness of such transcriptional states [58, 59].

1.2.4 NC terminal differentiation

Finally, following extensive collective migration throughout the embryo guided by environing cues, NC cells settle in different distal sites and deploy distinct gene batteries that regulate their terminal differentiation. This process is mediated through the action of multiple lineage-determining TFs, which co-bind hundreds or thousands of downstream targets loci to effectuate differentiation into a plethora of NC derivatives. Interestingly, TFs are often reiteratively used to drive different developmental lineages, such as Sox10 regulating both pigment cells differentiation together with MITF, and in Olig1/2-mediated differentiation of oligodendrocytes [10, 25, 47, 56, 62, 68, 69].

1.3 Distal control of development: enhancers

Although previous reconstructions of metazoan GRNs afforded novel insights into development [98–101], a major criticism is their depiction of genomes as a constant *cis* environment, with the dynamics of TF-mediated operations being the only active force driving development. However, the chromatin is a highly compacted structure, composed of nucleosomes as the basic unit. Such condensed package in the form of chromatin harbours the whole genome, and thus efficient gene regulatory mechanisms evolved to access, identify, unwrap, read, decode and archive genomic information [102]. A major source of developmental information is encoded in regulatory regions often located distally from gene bodies, collectively known as enhancers [8, 103]. This suggests that the *cis*-regulatory environment is highly active and differentially deployed [104]. It is thus crucial to understand the chromatin dynamics to decipher the genomic basis of development. Below, we briefly discuss enhancer properties and their dynamics in development, and in turn, how this knowledge is informing and shaping our current understanding of NC development.

1.3.1 Mapping enhancers genome-wide

Active enhancers are found in nucleasome-depleted regions, also known as accessible chromatin, and are often flanked by nucleosomes displaying ”permissive” histone post-translational modifications, such as H3K4me1 and H3K27ac, deposited by lysine methyltransferases and acetylases (e.g., CBP/p300), respectively [8, 103, 105–107]. However, these *cis*-regulatory elements can be virtually located anywhere in the genome, establishing long-range interactions with target promoters [108, 109], including interchromosomal communications [110]. Additionally, enhancers can also be found in inactive or poised states [111]. In fact, poised enhancers are important for efficient cellular differentiation, largely due to their fast deployment in which the repressive H3K27me3 mark is replaced by H3K27ac [112], as previously discussed [79].

Mapping enhancers genome-wide is therefore challenging [103]. The main methods for enhancer identification rely on properties correlated to their function, such as nucleasome depletion (MNase-seq), enrichment in accessible chromatin (ATAC-seq, DNAse-seq), as well as chromatin immunopreciptation followed by sequencing (ChIP-seq) for histone marks associated/correlated with their function, such as H3K27ac, as well as targeting TF-bound chromatin [8, 103, 104, 113, 114].

Unfortunately, these methods require on millions of cells, and have largely being used to profile cells cultured *in vitro*. Novel approaches, such as CUT&RUN and CUT&TAG, which use an antibody-fused to MNase or the Tn5 transposase, respectively, are capable of generating high-resolution maps of TF binding in ultra low input samples (ranging from 1 to 100.000/cells per experiment)[115]. These methods therefore are enabling the interrogation of cells difficult to culture, including the NC [116, 117], which can be extracted directly from the their endogenous context in the developing embryo. Moreover, our lab has pioneered the development of tissue-specific biotin ChIP-seq for low input material profiling (100.000 cells/experiment) in avian and zebrafish embryos [58, 60]. Our approach relies on Avi-tagging proteins of interest, such as TFs, which can be biotinylated *in vivo* and precipitated with streptavidin beads, resulting in high-resolution maps of TF binding *in vivo*.

Nevertheless, these approaches indirectly map enhancers by profiling chromatin regions which are associated, but no exclusive to, enhancers. To circumvent this limitation, a novel approach named self-transcribing active regulatory region followed by sequencing (STARR-seq) was developed [118]. This method capitalises on the fact that enhancers can modulate transcription in an position and orientation-independent manner [119]. Elegantly, Stark and colleagues [118, 120] designed this approach by creating libraries of putative enhancers cloned downstream of a minimal promoter, followed by polyA-tail. This way, this library can be transfected in a cell-type of interest and then transcripts isolated and quantified. In this method each enhancer functions as a self-reporter, which can be transcribed, sequenced and mapped back to a reference genome, thus yielding quantitative maps of enhancer activity genome-wide [118]. We believe that adapting STARR-seq to lower cell numbers or single cells is poised to transform this technology into an even more powerful method for the quantitative identification of enhancers *in vivo*.

1.3.2 Enhancer clusters and enhancer 'redundancy'

Although individual enhancers are key to transcriptional activation, that is, switching gene expression on and off, the regulatory landscape of a single locus often involves several enhancers [8, 98, 101, 103, 108, 121], suggesting that enhancers of seemingly redundant activities may have additional function. Indeed, redundant enhancers are involved in transcriptional precision, that is, the increase in the robustness of gene expression, as well as the precise regulation of sharp boundaries of transcription patterns [122–125], such as those involved in neural crest patterning [59] and hindbrain segmentation [74]. In this context, robustness is defined here as the persistence of a pattern of gene expression, resisting changes caused by genetic or environmental perturbations. Therefore, overlapping enhancer activities may decrease fluctuations in expression patterns, which can be caused by environmental, phenotypic and genotypic differences, thus fostering robustness in gene expression [124–126].

Genome-wide identification of cell-type specific enhancers revealed clusters of enhancers located in lineage-determining loci, such as *Myc* in embryonic stem cells [121]. A controversial hypothesis suggested that such clusters of enhancers display synergistic properties; in other words, their combined effect on transcription is higher than the sum of the individual effect of each enhancer, thus establishing "super-enhancers"[121, 127, 128]. However, dissection of "super-enhancers" suggests additive or sub-additive [129, 130], rather than non-linear, higher order synergistic effects, arguing against a hierarchical "super-enhancer" hypothesis. Therefore, little evidence supports the existence of synergistic properties between spatially clustered enhancers [130]. Consistent with this, our recent study on the *cis*-regulatory landscapes of early avian NC [59] revealed a self-interacting domain containing a ”super-enhancer” upstream of the NC factor *Sox10*. This regulatory domain features multiple accessible enhancers, is highly enriched for Brd4 binding and H3K27Ac modification, and seems to be conserved in mammals [131] and zebrafish (this thesis). The modular structure of this *Sox10* ”super-enhancer” suggests the presence of 6-7 enhancers, which function in an additive or sub-additive manner and display complementary domains of activity [59].

Some enhancers are perceived as redundant because their depletion results in normal development under optimal conditions (*e.g.*, [129]). However, their importance to transcriptional precision becomes obvious under stressful developmental conditions, as, for instance, enhancer knockouts may reduce viability [125]. Therefore, ”redundant” enhancers seem to be key to canalising developmental fates [125]. Additionally, if two ”redundant” enhancers are weak, they might cooperatively or synergistically enhance transcription by displaying additive effects, as long-range enhancer-promoter interactions are stochastic [132]. In contrast, if two strong enhancers are communicating with the same promoter, their combined activation might be less than the sum of their individual effects owing to their competing behaviours [132], further arguing against a ”super-enhancer” hypothesis[130]. Ironically, if spatial clusters of enhancers indeed function in synergy, this necessarily means that the majority of individual enhancer elements within a ”super-enhancer’ are intrinsically weak.

The general architecture of enhancer is constituted by multiple motif-containing TFs binding sites (TFBS) [8, 103]. These inputs are represented by general pleiotropic factors acting downstream of signalling pathways such as nuclear receptors, which act in combination with cell-type- or tissue-specific restricted TFs (e.g., [117]). For example, NC enhancers often employ inputs from WNT, FGF and

BMP signalling together with cell-type restricted TF, Sox10 [54, 58, 59, 117]. The sharp boundaries of gene expression are tuned by TFs repressing such enhancers. In fact, overlapping binding sites for activators and repressors are often found within a *cis*-regulatory environment, suggesting that TF binding competition is a important parameter for tissue specificity [126]. Because of the combinatorial nature of inputs binding to enhancers, a long-debated *cis*-regulatory code has been proposed to exist with properties similar to that of the genetic code, such as redundancy, combinatorics, modularity and to some extent universality [133].

Enhancer activity is primarily dictated by TFBS syntax and binding of a critical number of TFs [8]. It has been recently demonstrated that low-affinity TFBS are essential for tissue-specificity [134–136]. Indeed, conversion of low-affinity to high-affinity or optimal spacing between TFBS cause ectopic gene expression outside of the endogenous territory [137]. This is probably due to the increased occupancy of pleiotropic regulators in different cellular contexts. Furthermore, *de novo* identification of endogenous low-affinity TFBS uncovered hidden enhancers surprising strength that control notochord development in tunicates, suggesting that low-affinity binding sites are a key mechanism for transcriptional precision [138]. Similarly, low-affinity binding sites were also discovered in enhancers regulating NC development [117]. Therefore, a key parameter of enhancer strength and tissue-specificity seems to be dictated by a TFBS affinity. The fact that low-affinity TFBS are so often found within enhancers led to the suggestion that developmental enhancers are in fact ”suboptimised”, suggesting why deciphering a hypothetical *cis*-regulatory code is challenging: the inherent nature of enhancers employ rules favouring a degenerate syntax. Indeed, a deep-learning algorithm trained on *Drosophila* enhancers identified by STARR-seq [139] has suggested that such principles of motif syntax and enhancer organisation are present genome-wide and across developmental contexts. This fact enabled the authors to design synthetic enhancers with tunable levels of activity *in vitro*, providing further support to the existence of a general regulatory syntax.

However, because enhancers often employ such low-affinity, suboptimal binding sites, individual developmental enhancers seem to be generally weak [135, 137, 138]. However, as suggested by cloning weak enhancers in tandem, their clustering was found to experimentally increase transcriptional activity without causing ectopic activation in different tissues. These results led the authors to propose that enhancer clustering is evolutionarily favoured within a locus as a manner to increase the robust expression without trade-offs to precision [137]; furthermore, these results may explain the underlying mechanism of why multiple enhancers are often found close to each other within ”super enhancers”.

From a mechanistic point of view, enhancer strength is related to the efficiency of transcripts production. Because transcription is a discontinuous event consisting of a series of bursts [140], strong enhancers can either modulate the production of more bursts (higher frequency) and/or production of more transcripts (higher amplitude) in a shorter period of time. In an elegant study, Fukaya *et al* [141] have used a live-imaging approach to visualise single transcripts in developing *Drosophila* embryos. Their results suggest that strong and weak enhancers produce transcriptional bursts of similar amplitudes and duration. However, strong enhancers produce bursts more often than the weak ones. Interestingly, a single enhancer can simultaneously communicate with two promoters producing transcriptional bursts of equivalent frequencies, amplitudes and duration, suggesting that modulation of enhancer-promoter communication relies less on direct enhancer-promoter contacts, but rather on a chromatin environment such as that of self-interacting/liquid-liquid phase separated domain [141], where a single enhancer can modulate different promoters simultaneously.

1.3.3 Combinatorial regulation by transcription factors and co-factors

Recent studies provide exciting evidence that TF combinations can be learnt from *cis*-regulatory sequences to which TFs bind [133, 142], indeed suggesting the existence of a combinatorial *cis*-regulatory code. Therefore, the minimal 'regulatory cocktails' required for cellular identities is likely to be deciphered from transcriptional enhancers. Nevertheless, a very limited number of TFs also act as "master" regulators, which we define here as factors that display the capacity of programming cellular identity on its own, such as the master regulator of eye morphogenesis *eyeless* in *Drosophila* [143] (also called *Pax6* in vertebrates). However, unifying features of such "master" regulators, such as pioneering activity, have not been identified yet.

However, little is known about the functional constraints on TFs combinatorics. Are there subsets of TFs that can (or cannot) interact with each other, or all TFs promiscuously interact with others? Are these interactions direct or mediated by co-factors? Are there intrinsic properties of TFs such as structural features that constrain or enhance their combinatorial action?

GRN rewiring by gains of direct TF inputs suggests, to a certain extent, that transcriptional regulators can promiscuously interact. Indeed, transcriptional synergy between the yeast Gal4 protein, the rat glucocorticoid receptor (rGR) and the human activator ATF [8, 144] indicate that interactions between TFs do not necessarily involve direct protein-protein contacts. Therefore, combinatorial binding of a critical number of TFs to enhancers might be sufficient for the recruitment of transcriptional machineries such as CBP/p300, which are able to mediate indirect synergy between them. Direct interaction of the TAZ2 domain in CBP/p300 to several TFs [145] suggests a structural basis for the promiscuous interaction among a wide variety of transcriptional regulators. However, once a regulatory output resulting from a combination of TFs is evolutionarily fixed in a *cis*-regulatory context, it is likely that such combination becomes less flexible to changes [146], such as those evident in enhaceosomes [147–149]. Hence, the complexity of metazoan GRNs might emerge from the plasticity of TF combinatorics, whereas the regulation of conserved developmental fates might result from evolutionarily fixed combinations of driver TFs.

Dissection of the *Drosophila evenskipped* (*eve*) stripe 2 enhancer, which contain TFBS for the weak activators bicoid and huntchback (hb) and the repressors kruppel and giant [150] using *trans* and *cis* complementation assays has shown that both TFs and TFBS required for enhancer activity can be replaced [150]. For instance, the same regulatory output can result from the conversion of a few low affinity TFBS to a single optimal TFBS and even from the replacement of a developmental TF (e.g., hb) by a generic activator (Sp1), arguing in favour of a view where TFs can promiscuously interact and result in equivalent regulatory outputs in a developmental context [147]. However, a recent comprehensive *trans* complementation study comparing the regulatory output of 474 transcriptional regulators recruited by the UAS/Gal4 system in *Drosophila* S2 cells has shown that at least 100 TFs can replace a TF in an specific *cis*-regulatory context [151]. Further comparison of 24 enhancer contexts showed that 42% of all 474 TFs could recapitulate enhancer activity in at least one enhancer context, whereas the vast majority of the remaining TFs acted as repressors [151], further suggesting the importance of TF-mediated repression for transcription precision. Importantly, transcriptional regulators and co-factors could be clustered in fifteen groups, consistent with their regulatory function, sequence features and functional role (that is, their intrinsic properties), providing evidence of equivalency within TFs belonging to each cluster (and often within) [151]. Therefore, although enhancer activity seems to be dictated to some extent by intrinsic properties of bound TFs, there seems to exist a large degree of flexibility to bound factors. Such plasticity might increase the robustness of GRNs by rapid TF binding turnover, and may have obscured the existence of a *cis*-regulatory code.

1.3.4 Topologically associating domains

Finally, it is important to note that long-range enhancer-promoter communications are achieved through genome folding. In the nucleus, the genome is folded in three-dimensional (3D), higher-order chromatin structures, with genomic regions containing interacting loci and regions compartmentalised into so-called topologically associating domains (TADs). Genome compartmentalisation contributes to the fine-tuning of genome functions, and spatial positioning and folding of genomic regions contribute to the control of gene expression [88, 152, 153].

This 3D compartmentalisation is achieved via insulator proteins at the TAD boundaries, such as CTCF binding in mammals [154]. It has been suggested that chromatin compartments are built by loop extrusion, in which chromatin is progressively extruded through a cohesin ring, until it encouters two CTCF molecules bound with convergent orientations [154], functioning as anchors to the extruded loop, thus creating a stable conformation leading to the formation of a TAD [155, 156]. Moreover, cohesin loading and cohesin release protein complexes, SCC2/SCC4 and WAPL, respectively, were shown to be important for controlling size of an extruded loop, by regulating the residence time of cohesin on DNA [157].

However, it is not clear yet to what extent the endogenous chromatin environment, including epigenetic histone modifications, chromatin compartmentalisation and sub compartmentalisation into TADs by insulator elements regulate transcription. For instance, depletion of insulation by knocking down cohesin reduces the sharp boundaries of TADs, but it is not sufficient to disrupt the overall endogenous compartments[158]. In addition to TADs, specialised macromolecular complexes mediate enhancer-promoter communication, transcription initiation and specificity, such as TFIID, E, F, H and large co-activators like Mediator (reviewed in ref. [108]).

Although after the depletion of TADs enhancer-promoter interactions are still maintained, loss of insulator elements also create ectopic communications between promoters located within the vicinity of developmental enhancers [159, 160], likely causing developmental abnormalities such as brachydactyly, F-syndrome and polydactyly.

1.4 Deep conservation of the NC transcriptional programme

Agnathans, also known as jawless vertebrates, represent the closest relatives to jawed vertebrates. They thus occupy a strategic phylogenetic position for studying NC regulators that may reflect ancestral states. Two extant lineages of agnathans, lampreys and hagfishes have been suggested to form a monophyletic group known as cyclostomes [161, 162]. Morphological, embryological, molecular and genomic studies of these animals have offered unique insights into the early evolution of the NC, as mechanisms shared between gnathostomes and agnathans may have been conserved for more than 500 million years (MY). Lamprey neural crest possibly reflects the ground state of this cell population and may help resolve species-specific differences in NC formation identified within vertebrates [162, 163]. Therefore, dissecting lamprey NC development using genomic approaches is essential to the study the evolution of the NC, and may provide further evidence for a conserved NC *cis*-regulatory code, at least within vertebrates [164].

Most developmental studies have concentrated on the marine lamprey *Petromyzon marinus* [163, 165–167], the Japanese marine lamprey *Lampetra japonica* [168] and the brook lamprey *Lampetra planeri* [169, 170], with some genomic resources also available for the Arctic lamprey *Lethenteron camtschaticum* [171, 172]. However, molecular embryology approaches and functional studies in cyclostomes are challenging not only because of the relative difficulty to obtain and manipulate embryos, but also due to programmed developmental genomic rearrangements that take place during embryogenesis, in which large portions of the germline genome are eliminated from somatic tissues [167]. Therefore, the recent sequencing of *P. marinus* genomes [166, 167] have been instrumental for comparative genomics in lampreys [169].

Although the common ancestor of vertebrates underwent one round of whole genome duplication [173], cyclostomes, or at least lampreys, have also undergone a second round, which contributed to the evolution of six *Hox* clusters [172], whereas gnathostomes also underwent an independent round of whole genome duplication [173], resulting in four Hox clusters in the stem lineage of gnathostomes. These differences have obscured the phylogenetic relationships between lamprey and jawed vertebrates homologous genes. Nevertheless, conserved factors have been identified in the lamprey NC, thus indicating the existence of a core set of ancestral regulators present before lineage-specific expansion/contraction of vertebrate genomes.

Similarly to jawed vertebrates, BMP, Wnt and FGF signalling also pattern lamprey ectoderm into the non-neural, neural and NPB domains [163]. Lamprey NC induction and specification involves the core transcriptional circuit conserved to jawed vertebrates involving TFAP2, FoxD-A, SnaiA, Sip1 (also known as Zeb2 or ZFHX1B), Pax3/7, MsxA, Id, n-Myc and ZicA [163, 174–177]. Three SoxE orthologues have been identified in lamprey, SoxE1, SoxE2 and SoxE3, displaying functional similarities to Sox8, Sox10 and Sox9, respectively [178]. Migratory NC also express conserved effector genes involved in NC lineage commitments, such as neurogenins and POU domain TF Brn3s involved in PNS development [10, 163]. Furthermore, lamprey cranial NC at midbrain level has also been shown to express an orthologue of Otx1/2, and migrate towards the mandibular process despite the absence of jaws, suggesting a regulatory logic and developmental fate of the midbrain NC predating the evolution of gnathostomes [179]. Equally, post-otic NC cells express collinear combinations of Hox TFs [180], which pattern the hindbrain and trunk NC, but seem to be depleted of canonical vagal NC, consistent with their apparent lack of a canonical enteric nervous system [181]. Lineage tracing studies in lamprey have identified cranial NC migration within the trigeminal, hyoid and branchial streams similarly to jawed vertebrates, and have shown that NC cells at this axial level also contribute to melanophores, sensory neurons, glia and chondrocytes [165, 182]. Lamprey NC may also give rise to other connective tissues [169]. For example, NC migration into branchial arches [162, 183] is consistent with a putative role in NC oropharyngeal skeleton contributions in agnathans.

However, orthologues of the gnathostome NC specifiers Ets1 and Twist1/2 - important regulators of ectomesenchymal development - are depleted in cranial NC specification of the sea lamprey, *P. marinus* [10, 163], indicating some level of plasticity in the evolution of the NC transcriptional programme. In the sea lamprey NC, Ets1 and Twist1/2 are only expressed during late migration and colonisation of the branchial arches [10]. Cross-species comparison of lamprey and gnathostomes suggest Ets1 and Twist1/2 were co-opted, likely through heterochronic shifts, during the evolution of ectomesenchymal NC in the stem lineage of vertebrates. This raises an attractive hypothesis: cranial NC regulators may have been acquired secondarily, and thus trunk NC may represent a simpler transcriptional network, possibly reflecting the ground state of the ancestral NC [10, 47, 163]. However, another hypothesis is that Ets1 and Twist1/2 were secondarily lost in a species- or lineage-specific manner. Thus, investigating their expression in other cyclostomes is necessary to test the hypothesis of gradual acquisition of ectomesenchymal specifiers into the NC GRN. For example, NC EMT in chicken and *Xenopus*, but not in mice, requires Snai1/2, suggesting mice may have evolved divergent mechanisms compensating for Snai1/2 depletion in the NC [184].

Transcriptomic and epigenomic profiling of dissected tissue enriched for NC in the marine lamprey *P. marinus* at different stages of embryonic development, further validated by *in situ* hybridisation analyses, identified gene co-expression modules underlying conserved ontogenetic processes derived from the NC [169]. These approaches extended the putative GRN underlying cartilage and connective tissue development, as well as implicated novel TFs in the specification of lamprey ectomesenchyme, such as Fli1, Zfhx4 and signalling molecules (Tead1/2). Therefore, a complex transcriptional network seems to be highly conserved across vertebrates. Consistent with this hypothesis, enhancers regulating the expression of *TFAP2b* and *SoxE1* in *P. marinus*, containing combinations of Hoxb2/b3/a2, Sox10 and Tfap2a binding motifs, which also displayed orthologous activity when tested in cross-species enhancer reporter assays in chicken and zebrafish embryos, further highlight the existence of a combinatorial logic functionally conserved across vertebrates.

Nevertheless, important distinctions between agnathan and gnathostome NC ontogeny have also been identified. For instance, cyclostomes lack sympathetic chain ganglia. Consistently, NC progenitors giving rise to sympathetic neurons in jawed vertebrates express Phox2a/b, Ascl1 and Hand, but are not co-expressed in lamprey NC [163, 181]. As they lack enteric nervous system, their autonomic control in the gut derives from spinal neurons, similar to amphioxus [181]. Moreover, lamprey peripheral neurons are not myelinated, although they are covered by glial ensheathments [185] likely derived from NC. Furthermore, gnathostome NC EMT involves extracellular matrix (ECM) remodelling mediated by Snai1/2, including a dynamic change of type I epithelial cadherins (E-Cadherin and N-Cadherin) in premigratory NC by mesenchymal, type II cadherins (such as Cdh6, Cdh7 and Cdh11) during delamination from the neural tube and migration. In lampreys, however, only single copies of Type I and II cadherins have been identified [167].

Surprisingly, lamprey NC undergoes EMT without a differential shift in these cadherins, keeping cadherin Type II expression in pre-migratory and early migratory NC [186].

Moreover, previous claims suggested NC evolved its ability to contribute ectomesenchyme through co-option of mesodermal transcriptional networks, based for example on the expression of Snai1/2 [48]. However, a recent study suggesting Snai1/2 role in NC has been co-opted from neurogenic circuits challenges this long-favoured assumption [186].

Until recently, however, conclusive evidence linking *de novo* evolution of genes, expansion/contraction of paralogous regulators and genome duplication to the diversification of the NC was lacking. An elegant functional investigation on the roles of the endothelin pathway in the marine lamprey, *P. marinus*, paved the way for such a link. Endothelin ligands (Edns) and their receptors (EdnRs) are unique to vertebrates. By disrupting their expression in *Xenopus* and lamprey embryos, Square *et al* [97] revealed that endothelin signalling is not required for NC specification and differentiation. This suggests that endothelin pathway co-opted into the NC transcriptional programmed occurred before whole-genome duplication events, in a vertebrate ancestor displaying multipotent NC. Nevertheless, differences in the paralogous components of the endothelin pathway deployed in agnathan and gnathosome NC further suggests that whole-genome duplications underlied the diversification of the NC in a step-wise manner, as indicated by different requirements of endothelin signalling in NC subopopulations. For example, NC-derived PNS seem to be largely Edn-independent, whereas melanocytes seem to have integrated Edn-signalling before the whole-genome duplication. Ectomesenchymal differentiation, however, display differential requirements in lampreys and gnathosomes, as indicated by the cooperation of EdnrA and EndrB in mucocartilage development in lampreys, and EdnrA-dependent development of the craniofacial elements in jawed vertebrates. We posit that these results also favour our view, in which genomic regulatory elements deeply conserved in vertebrates pre-pattern NC into neurogenic, melanocytic and skeletogenic progenitors, thus pre-parsing NC developmental potential. Therefore, the *EdnrA/B* loci represent exciting genomic substrates for evolutionary studies.

Studies of hagfish development, however, have been hampered for a long time by the inability to collect embryos [187]. The recent culture of hagfish embryos developed by Kuratani and colleagues (reviewed in [187]) thus represent a major breakthrough in evolutionary developmental biology affording unique insights to which studies in lampreys can be compared to another agnathan [188, 189]. Identification of Sox9+ NC highlights SoxE expression as a highly conserved feature in all vertebrates studied to date [188]. We thus expect that future studies of the hagfish embryos, combined with genomics, will continue to illuminate the early origins of vertebrates.

Taken together, these studies indicate conserved regulators shared within jawless and jawed vertebrates, thus establishing hypothetical requirements for the early appearance of the early, migratory, multipotent NC at the dawn of vertebrates. However, they also indicate a certain level of plasticity within the gene-regulatory interactions underlying NC ontogeny.

1.5 Pre-vertebrate origins of the NC

Although the NC represents a synapormophy of vertebrates, understanding whether the NC appeared abruptly in the common ancestor of vertebrates (*i.e.*, as an evolutionary innovation) or through gradual changes from homologous cell-type(s) (that is, as an evolutionary elaboration) remained a mystery for several decades [163]. Comparative studies in the closest living relatives of vertebrates - invertebrate chordates - have thus been essential to illuminate the early origins of the NC. Here, we highlight how such studies yielded insights into the incremental incorporations of ancestral circuits within the NC GRN through gene co-option, remodelling of an ancestral transcriptional network, and the coupling of GRN subcircuits controlling cellular differentiation of particularly informative cell-types homologous to NC derivatives, such as pigmented cells, dorsal root ganglia-like neurons, and chondrocytes.

1.6 Genome-scale rewiring of an ancestral chordate GRN

Urochordates, commonly known as sea squirts, are a diverse group of filter-feeding, metamorphic, marine invertebrates, representing the unlikely closest invertebrate relatives to us, vertebrates [1]. Within urochordates, a common group of sessile filter feeders, the ascidians, are remarkable by displaying an adult body plan highly divergent to that of vertebrates. However, after gastrulation, these animals develop into free-swimming tadpoles, displaying a simple cephalic region with light-detecting organs, pigmentation, and an active tail controlling navigation, highly similar to that of vertebrates.

Elegant lineage-tracing experiments on *Ecteinascidia turbinata*, a group of tunicates displaying gigantic tadpoles, provided the first elegant evidence suggesting that the pigment cells of urochordates derive from migratory, MITF-driven progenitors from the trunk lateral plate ectoderm, in a manner similar to that of melanocytes derived from the NC [190]. In *Ciona*, similarly to the results from *E. turbinata*, pigmented cell-types have also been identified displaying similarities to NC-derived melanocytes; however, such results yielded a controversy.

Pigment cells derived from the cephalic region (the ”a” lineage) or the trunk lateral plate ectoderm (the ”A” lineage) contribute to MITF-expressing, light sensing, pigmented cells types contributing to the ocellus and the otolith [191, 192]. In vertebrates, however, light-sensing pigment cells are derived from otic placodes. In the trunk region, migratory border cells from the ”A” contribute to melanocytes within the body. Are both progenitor lineages, the cephalic ”a” and the trunk ”A”, orthologous to the NC? One hypothesis suggested that such differences could point to axial-regionalisation predating the origin of NC; that is, the NC was already pre-parsed into cephalic and trunk regions before the early evolution of vertebrates; or both lineages were co-opted into the NC transcriptional network regardless of its axial regionalisation [192].

The identification of a transcriptional NC-like domain of the NPB expressing Msxb and Pax3/7 within the *Ciona* trunk region, as well as a pro-placodal-like region expressing Dmrta, FoxC and Six1/2 in the cephalic domain, indicate that the early, melanocytic NC was in fact restricted to the trunk region (”A” progenitor cells), whereas pigment cells derived from cephalic regions share an evolutionary origin to neurogenic placodes [70]. Later invasion of this NC-like domain into the head region could have endowed the cephalic NC with its unique ability to pattern the vertebrate ”new head”, facilitating the transition from a filter feeding to predatory life-style. Indeed, overexpression of the canonical EMT driver Twist in pigment cell progenitors (although experiments were conducted in the cephalic ”’a9.49” lineage) resulted in the extended migration of such cells away from their usual sites of colonisation [191]. These results strongly support the idea that co-option of EMT subcircuits (such as Twist1/2, Lmo3/4, Snai1/2, ZFHX4 and Zeb1/2), or more generally of additional transcriptional modules, endowed NC with its ability to migrate extensively and colonise distal sites within the embryo.

Interestingly, the sea squirt *Ciona intestinalis* tadpole displays a neuronal cell-type responsible for the sensory processing of the tail region. These cells present a remarkable similarity in the transcriptional network and ontogeny to sensory neurons in the dorsal root ganglia (DRG) of vertebrates, derived from the trunk NC. These DRG-like neurons, known as bipolar tail neurons (BTNs), are derived from the lateral ectoderm, display HMX expression [65] similar to that of DRG neurons [193], and migrate away from the NPB to give rise to the sensory nervous system of the *Ciona* tadpole, reminiscent of neurogenic NC. Therefore, it is likely that NPB cells with migratory ability and contributing to the peripheral sensory nervous system predated the origin of the NC, being co-opted and patterned into the NC GRN.

Our most distant invertebrate chordate relatives - cephalochordates, as known as amphioxus - display a simple free-swimming body plan, highly similar to the ammocoete larvae of lampreys. Therefore, these animals, represented by species such as *Branchiostoma lanceolatum* and *Branchiostoma floridae*, may display a body plan likely conserved to that of the last common ancestor of vertebrates [1]. Therefore, comparisons to amphioxus can provide important insights into the ancestral state of the common ancestor of chordates.

Extensive analysis of the spatial distribution of orthologues of NC regulators in the amphioxus embryo revealed conserved expression of Pax3/7, Msx and Zic within the neural plate border [175, 194]. However, NC regulators, such as FoxD3, Twist, Id, TFAP2 and SoxE are excluded from this domain [174]. Therefore, it is possible that the early acquisition of NC-like inducing NPB cells in the common ancestor of vertebrates and urochordates involved the co-option of TFAP2, SoxE, FoxD3, Twist and Id via changes in their *cis*-regulatory architectures [195].

Moreover, amphioxus also displays chondrogenesis controlled by a transcriptional module driven by SoxE, which is derived from the mesoderm. Therefore, the cartilage to which NC gives rise to is not an evolutionary novelty; the novelty here is an ectoderm-derived cell-type (the NC) being uniquely able to contribute to cartilage [48]. An elegant study conducted by Jandzik and colleagues [48], in which the whole amphioxus SoxE regulatory region controlling expression of a reporter was inserted into the zebrafish genome resulted in reporter expression in a manner which recapitulated SoxE-expressing domains in the amphioxus embryo, but reporter activity was not found in NC. Further analyses indicated that chondrogenic NC rely on a *cis*-regulatory architecture dependent on TFAP2-mediated regulation of SoxE in vertebrates. Therefore, acquisition of TFAP2-driven SoxE regulation via changes in *cis* provides an elegant mechanism explaining the co-option event of the chondrogenic module into the NC GRN, likely resulting in NC-derived craniofacial cartilage in the vertebrate head. Consistent with this idea, double-knock out of the NC-specifiers FoxD3 and TFAP2 in zebrafish completely ablates NC ontogeny, suggesting TFAP2 is indeed a central anchor of the NC transcriptional network together with FoxD3 [195].

Surprisingly, a recent comparison of whole-embryo, single cell transcriptomes at high-resolution across diverse species suggests that cellular differentiation modules coupled to different germ layers switch to other germ layers more frequently than previously thought [196]. Therefore, transcriptional module co-options into different germ layers via changes in *cis* may represent a relatively common mechanism shaping the evolution of metazoan body plans. Thus, the co-option event that resulted in evolution of chondrogenic NC may offer a paradigmatic example of such elusive mechanism.

Surprisingly, the identification of a domain of Bmp3/4, Dlx, Pax3/7 and SoxE expression in the dorsal neural chord of an acorn worm, the hemichordate *Balanoglossus simodensis* [197], points to the potential deep conservation of a transcriptional program patterning the neural plate and neural plate border, likely pre-dating chordates (Figure 1.4a). This would necessarily imply that cephalochordates have secondarily lost these domains of expression. However, the last common ancestor of hemichordates and chordates has diverged during the Cambrian explosion, and such similarities could also represent patterns of gene expression evolved in parallel (that is, homoplasies). Additionally, the homology between the chordate neural tube and dorsal neural chord of acorn worms is still debatable. Therefore, further studies involving extensive taxon sampling within hemichordates are required to test such equally parsimonious hypotheses.

Finally, the anatomy of cephalochordates also suggests that pharyngeal gills and somites extensively co-exist in a manner similar to the domain where vagal NC is specified in gnastosomes. This provides further support that a posterior, trunk-like environment, and not a cephalic region, reflects the ancestral state of the NC [198]. Indeed, loss-of-function of the trunk NC specifier Cdx in the ascidian *Halocynthia roretzi* resulted in complete loss of body pigmentation but not of pigmentation in the ocellus and otoliths [199]. Altogether, these studies strongly support the idea that a MITF-driven, posterior pigment cell progenitor, and not cephalic, was integrated into the ancestral, melanocytic NC transcriptional network.

In summary, these studies indicate that the invention of multipotent, migratory NC within vertebrates was likely incremental, involving extensive changes in the *cis*-regulatory architecture of NC-associated loci, such as *Twist, Id, TFAP2, FoxD3* and *SoxE*. The uneven distribution of such regulators within the phylogeny of chordates thus allows us to schematically polarise the acquisition of individual transcriptional modules driving a directed, step-wise evolution of the ancestral NC (Figure 1.4b).

1.6.1 *Trans*-regulatory changes: insights from FoxD3

Although the studies outlined above suggest that the adaptive elaboration of the ancestral NC involved genome-scale rewiring events relying on extensive remodelling of enhancer-promoter networks, extensive innovations in *trans* have also been proposed to the *de novo* evolution of NC traits.

One attractive example illustrating *de novo* changes facilitating the evolutionary appearance of the NC is elaboration of the transcriptional activation mechanism

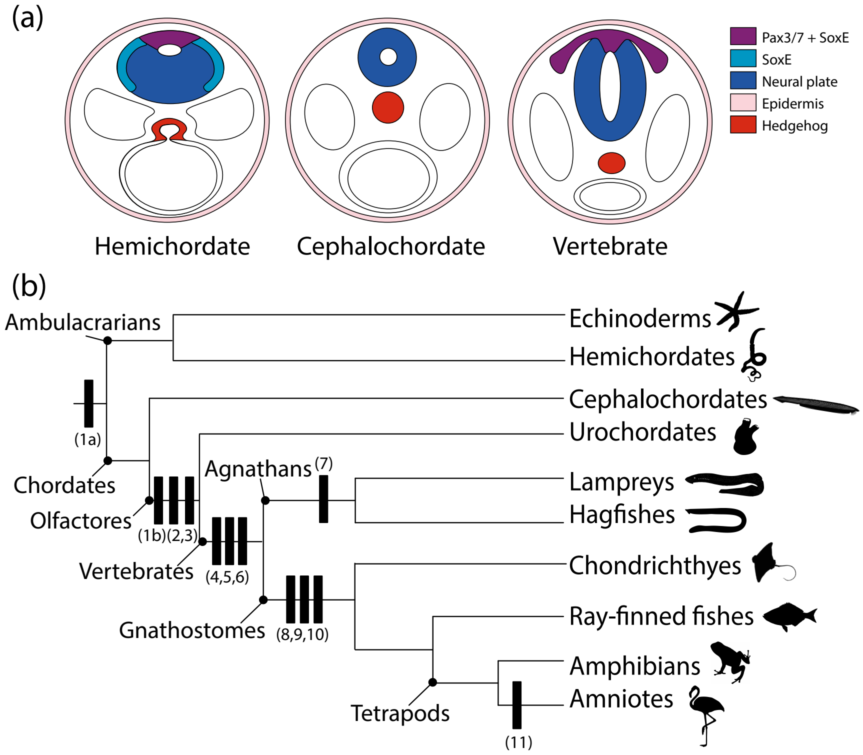

Figure 1.4 Early origins and transcriptional modularity of the NC (**a**) Schematic comparison of the NC markers Pax3/7 and SoxE expression in pre-metamorphic enteropneust hemichordate larvae, and cephalochordate and vertebrate embryos. Cross-section through *B. simodensis* collar cord, and neural plate of neurula amphioxus and vertebrate embryos, suggesting potential homology between the vertebrate neural plate border (NPB) and the dorsal aspect of *B. simodensis* collar chord. (**b**) Phylogenetic relationships of selected extant deuterostomes highlighting key evolutionary transitions associated to NC evolution. Hypothetical character changes within branches represent: (**1**) Origin of SoxE+ and Pax3/7+ expressing NPB-like cells: under the pre-chordate scenario (the "enteropneust theory"), NPB-like cells appeared in the common ancestor of chordates and hemichordates (**1a**). Alternatively, NPB-like cells evolved in the common ancestor of vertebrates and urochordates (**1b**); (2) NPB-like cells compartmentalised into proto-placodal and proto-NC territories expressing Six1/2 and Msx1/2, respectively; (**3**) co-option of core NC specifiers (FoxD, SoxE, TFAP2, Snail) downstream of Msx1/2, coupled with a Mitf-driven circuit delineating pigment cell lineage; (**4**) first round (1R) of genome duplication (autotetraploidization); (**5**) canonical NC, giving rise to melanocytes and neurons and glia in the PNS; (**6**) ectomesenchymal NC; (**7**) independent genome duplication(s) in agnathans; (**8**) second genome duplication (2R) in jawed vertebrates; (**9**) Cranial NC-derived jaws and odontoblasts; (**10**) vagal NC; (**11**) NC regionalised into cranial, vagal, trunk and sacral domains.

in FoxD3 protein. Fork head box (Fox) proteins, also known for their winged helix structure, contain a highly conserved forkhead (FKH) DNA-binding domain (Figure 1.5). FKH is composed of three alfa-helices, three beta-sheets and two "wing" regions, which is strikingly similar to the three-dimensional structure of the linker histone H1 [200]. Such resemblance to histones is functionally linked to FKH binding and ability to remodel compacted chromatin [60, 201]. It has been proposed that Fox factors pioneering activity depends on the interaction with the core histones H3 and H4, through displacement of linker histone H1s in compacted chromatin, thus establishing accessible chromatin permissive to the binding of other TFs [200]. In jawed vertebrates, FoxD3 has been identified as a central regulator of NC development, functioning as a bimodal player: as a pioneer factor, binding to nucleosomal DNA and decommissioning enhancers (Figure 1.5c), and as a canonical transcriptional repressor, likely acting via a C-terminal domain interacting with additional transcriptional repressors (Figure 1.5d) [60].

Interestingly, phylogenetic and functional comparison of vertebrate FoxD3s to non-vertebrate chordates and other vertebrate FoxDs revealed the presence of an N-terminal segment necessary for FoxD3 to activate transcription [202], likely underlying the pioneering activity of FoxD3, but not necessary for repression. This evidence suggests that FoxD factors play an ancestral role as transcriptional repressors, binding to accessible chromatin similarly to other Fox factors (Figure 1.5a-b). The novel N-terminal domain, however, may represent an innovation in vertebrates, which endowed FoxD3 with the property to remodel chromatin landscapes and activate distal *cis*-regulatory networks during embryonic development.

SoxE proteins, represented in jawed vertebrates by Sox8, Sox9 and Sox10, are transcriptional activators containing the High Mobility Group (HMG) box domain capable of binding to accessible chromatin [203]. Replacement of Sox9 by the Amphioxus SoxE suggests that AmphiSoxE possess similar properties to Sox9 in avian NC, although with a reduced transactivation activity, and preferentially promoting a glial over neuronal fate [204]. Similar results were found in the replacement of the mouse Sox10 by the SoxE orthologue from *Drosophila*, Sox100B [205]. Therefore, a *cis*-regulatory co-option event may have been sufficient to integrate SoxEs into the

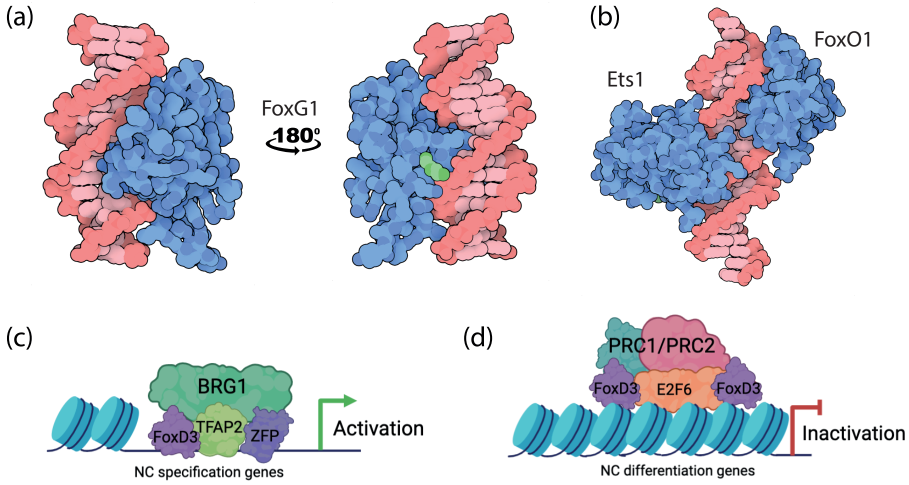

Figure 1.5 Structural properties underlying trans-regulatory evolution of a central NC specifier, FoxD3. (**a**) A 3D cartoon model of FKH-dependent FoxG1 (PDB id: 7CBY) binding to the major groove and (**b**) FoxO1 co-biding with Ets1 (PDB id: 4LG0). Cartoon visualisations were created with Illustrate. (**c**) Co-binding of FoxD3 to accessible chromatin with other negative regulators, such as E2F6, may underlie FoxD3 ancestral activity as a transcriptional repressor. (**d**) The novel vertebrate-specific N-terminal domain of FoxD3 may have contributed to the acquisition of pioneering activity, enabling FoxD3 to bind compacted chromatin, likely in combinations with other transcriptional activators, such as TFAP2.

NC, but some level of innovations/optimisations in *trans* has taken place during the evolution of SoxEs integrated to the NC transcriptional programme.

However, the high conservation of protein-coding genes orchestrating NC development, coupled with the identification of changes in the spatiotemporal pattern of gene expression suggests that innovations in trans were less common than *cis*-regulatory divergence during the evolution of the NC. We therefore posit that enhancer evolution was central in facilitating the early appearance of NC, followed by changes in *trans*, which together diversified the NC GRN and functioned as substrates for adaptive morphological adaptations.

1.6.2 *Cis*-regulatory divergence underlying the elaboration of the NC

Although the aforementioned functional analyses convincingly suggest enhancer evolution as a means of rewiring entire transcriptional networks, the mechanistic basis by which *cis*-regulatory divergence happened during the emergence of the NC remains elusive. Nevertheless, comparative regulatory genomics across different species (in the NC and other systems) have revealed important mechanisms by which rapid changes can be achieved, playing roles in the diversification of novel cell-types, morphologies and speciations, briefly discussed below.

Exaptation of transposable elements (TEs) as a driver of craniofacial diversity

The craniofacial diversity of vertebrates is one of the most attractive examples of adaptive radiations in the tree of life. An iconic example is illustrated by the craniofacial adaptations of Galapagos finches that captivated Darwin, however little is known about the genomic mechanisms underlying such adaptations. These morphological traits are largely derived from the NC, and therefore functional investigations of the NC have been essential to advance our understanding of the drivers of morphological diversity in vertebrates.

Using an *in vitro* differentiation protocol, Prescott *et al* [206] were able to investigate in detail the NC cells giving rise to the craniofacial features in human and chimps (reviewed in [104]). Their study revealed that differential colonisation and exaptation of transposable elements (TEs), including endogenous retroviruses (ERVs) have played a pivotal role in the diversification of such morphologies. TEs are selfish genomic elements that can rapidly expand within genomes and represent a large proportion of the human genome [206]. These elements commonly contain suboptimal binding sites for endogenous TFs, or transposase-derived DNA-binding proteins that can rapidly evolve into TFs [8]. Enhancers enriched for ERV1, ERVL-MaLR, ERVK, and L1 elements were associated to species-specific biased *cis*-regulatory elements, suggesting their differential exaptation in a species-specific manner have underlined the evolution of chimp and human facial features. Moreover, the NC loci associated to such species-specific biases include components of the endothelin pathway, Pax3/7, BMP4 and the BMP receptor, BMPER. Therefore, the adaptation of TEs harbouring suboptimal TF-binding sites into regulatory elements may represent a common evolutionary mechanism underlying the rapid diversification of morphological traits in general and the craniofacial diversity in particular.

However, species like Galapagos finches feature genomes with constrained TE composition. It will thus be interesting to include such species, as well as vertebrates representing different taxa, in further studies to identify other genomic drivers of NC-dependent morphological adaptations.

Heterochronic shifts in NC development through changes in TF binding

A recent investigation of ray-finned fishes revealed a high plasticity of NC cells in acquiring heterochronic shifts [207]. Although NC ontogeny has been proposed to follow a conserved anterior to posterior sequence across vertebrates, the NC in bichir, sterlet, gar, and pike embryos display several divergences from this model. Firstly, cephalic NC migration starts prematurely when compared to other vertebrates, at early stages of neurulation and before neural tube closure. Strikingly, however, the migration of the post-otic (hyoid) stream occurs prior to the more anterior, midbrain stream. This has been associated to several species-specific morphologies: for example, key hyoid structures such as external gills (bichir), a large opercular flap (gar) and formation of cartilage (pike) develop earlier in these lineages. The inclusion of such diverse taxa in this study thus illustrates widespread differences in the timing and order of NC ontogeny.

How have these heterochronic shifts evolved? A study in marsupials might provide some potential clues. Neonates of marsupials are altricial at birth and develop under a shorter gestation period. Opossum (*Monodelphis domestica*) develop their jaw primordia significantly earlier than mouse or quail embryos, likely due to the fact that neonates need to feed significantly earlier than their mammal and avian counterparts. Skeletogenic NC is differentially enriched for the SoxE regulator, Sox9. *Cis*-regulatory analysis of the opossum Sox9 revealed a cMyb-responsive enhancer (E3) in marsupials, displaying conserved activity in avian embryos [208]. The orthologous enhancer in avian embryos, however, lack a marsupial sequence required for cMyb-dependent activity. Moreover, the "marsupialised" enhancer displays activity before the endogenous activation of avian *Sox9*. Collectively, these results suggest that heterochronic shifts in the marsupial NC are dependent on the *cis*-regulatory rewiring of Sox9 and cMyb. One of such changes was the "marsupialisation" in *cis* of *Sox9* E3 enhancer with the acquisition of a cMyb-responsive sequence.

These results tempt us to hypothesise that genome-wide changes in *cis*, through the vast acquisition of cMyb-responsiveness, could have endowed the marsupial NC with the ability to initiate skeletogenic programmes earlier than NC in mammalian and avian embryos [208]. Similar mechanisms, dependent on the acquisition/loss of TF binding has the potential to explain the differences in the development of bichir, pike, gar and sterlet NC. However, an alternative explanation could relate to species-specific differences in the biophysical sculpting of these embryos. These hypotheses also need to be taken into account in further studies attempting to decipher the molecular and morphogenetic mechanisms contributing to such diversity.

De novo **evolution of enhancers**

Another mechanism of evolution in *cis* is the *de novo* acquisition of enhancers from inactive, ancestral sequences. Although this mechanism has been long illustrated in several systems, to what extent enhancers are lost and acquired remains an essential question in genomics and developmental biology. A quantitative study across 11 million years of *Drosophila* evolution using STARR-seq in a constant *trans*-regulatory environment and in a genome-wide fashion revealed hundreds of enhancers evolving *de novo*, albeit the majority was maintained through stabilising selection [120]. Importantly, newly acquired enhancers commonly evolve on "neutral" sequences without apparent linkage to adaptive selection. Novel enhancers seem to compensate the loss of conserved enhancers, likely stabilising total strength of enhancer clusters per loci. Moreover, acquired enhancers, compared to their inactive orthologous sequences, display TF binding similar to those enhancers within a locus they might be compensating for, suggesting that selection on TF binding, rather than on sequence itself, might drive evolution of cell-type or tissue-specific gene regulation [120]. Advancing our understanding of enhancer evolution in the context of NC will thus require similar efforts, at micro- and macro-evolutionary scales.

Changes in enhancer-promoter contacts

In metazoans, a major determinant of enhancer communication to their target promoter is a class of *cis*-regulatory elements known as insulators. Insulators are also called chromatin boundary elements (CBEs) as they are situated at the edges of TADs [209]. CBEs play an essential role in altering gene expression at loci within the TAD, acting both as barriers that shield against the spreading of the repressive chromatin and as enhancer-blockers that prevent the communication between distant elements or promoters in other TADs that can influence gene expression. [152, 209]. In vertebrates, boundary elements display a conserved regulatory grammar, and are responsive to binding of sequence-specific DNA-binding protein CTCF (CCCTC-binding factor) and the multiprotein cohesin complex, which display an ancient origin within bilaterians [209].

Therefore, an attractive mechanism of evolution in *cis* relates to the extensive rewiring of enhancer-promoter contacts via genomic inversions, duplication and fusions, changing the conformation of TADs during evolutionary time [210]. Such mechanism has been essential for the evolution of Hox cluster regulation in vertebrates. However, to what extent this is a general mechanism in different loci and cell-types remain challenging to study owing to the technological requirements and costs associated to studying 3D-conformation of the chromatin.

Interestingly, this mechanism of evolution in *cis* (rearrangement of TADs and rewiring of enhancer-promoter contacts) was revealed from a study on the intersexuality of the female iberian mole, *Talpa occidentalis* [211]. In this species, females display a masculinised organ, the ovotestis. The development of ovotestes in female moles was linked to several genomic changes, including extensive rearrangement of TADs. However, the regulatory expansion within a TAD was associated to increased production of androgen. Within this syntenic region, the androgen-producing enzyme, CYP17A1, was triplicated and contrary to expectations, expression of the three CYP17A1 paralogs was highly unequal, with one CYP17A1 copy dominating the transcription within the locus. In fact, the two new copies, CYP17A1-2 and CYP17A1-3, together contribute less than 5% of CYP17A1 transcripts. Further investigation of the *cis*-regulatory architecture of this intriguing locus identified the duplication of CYP17A1 enhancers followed by their fusion. The super-active new enhancer from the Iberian mole was further tested in a murine model, revealing a 3- and 2-fold increase in CYP17A1 expression in females and males, respectively, thus convincingly suggesting that the genomic expansion of CYP17A1 locus contributed to the masculinisation of female moles not via increased CYP17A1 expression from multiple loci, but via evolution of a novel potent enhancer resulting in transcriptional boost from a single CYP17A1 locus. Importantly, these results warrant caution when assuming protein-coding gene duplication as sole drivers of morphological diversification, as often hypothesised in the context of NC evolution.

1.7 Conclusions

Comparative transcriptomic and epigenomic approaches across different taxa distributed within phylogenetically informative lineages are thus poised to advance and refine our understanding of NC development and evolution. In fact, the development of multiple single-cell ”omics” approaches, including single cell RNA-seq, has illuminated the evolution of cell-type specific transcriptional networks across metazoans. Nevertheless, their wide-adoption for comparing evolutionarily distant species has been hampered, among other reasons, by the availability of computational algorithms that can integrate such highly multi-dimensional datasets within the same dimensionally reduced space, owing in part to the limitation imposed by the number of informative ”anchoring” orthologs shared between such transcriptomes, as well as the non-linear relationships represented by the divergence of transcriptional programmes [196].

The rapid development of novel computational algorithms (e.g., [196]), have been essential for the robust analysis of single cell transcriptomes within and between metazoan lineages, thus greatly enriching the available toolkit for such investigations. We expect that adoption of similar approaches targeted at functionally profiling the NC and putative homologous cell-types from species as diverse as vertebrates, tunicates, cephalochordates and hemichordates will provide unparalleled insights into the transcriptomic evolution of the NC.

Comparing epigenomes and regulomes, however, is exceptionally difficult owing to multiple genomic rearrangements, fusions, deletions, neutral drifts and insertions of TEs within genomes, which hamper the identification of orthologous non-coding regions across large evolutionary scales. The majority of such approaches aiming at comparing epigenomes have thus focused on relatively small evolutionary distances [212] or within taxa with constrained genomic evolution [120]. Nevertheless, the technological advancements over the last couple of years, including the development of robust epigenome profiling techniques in population of cells or single-cells of a wide array of species, such as STARR-seq, ChIP-seq, ATAC-seq, and more recently Cut&Run and Cut&Tag, are enabling the identification of distal-acting *cis*-regulatory elements which can be identified within syntenic regions, and thus their relative distribution, motif content and nucleotide distribution, as well as content of inserted TEs, can be used to identify putative orthologous or divergent enhancers across different species at a genome-scale resolution (Figure 1.6a). Such experimental approaches will be essential to afford insights into the modes of evolution of *cis*-regulatory elements (Figure 1.6b). We propose that such approaches will ultimately enable the identification of the genomic substrates underlying the dawn of the NC, as well as enabling the identification of non-coding alleles under differential selective pressures. However, the development of novel computational approaches, pipelines and algorithms, based, for example, on deep-learning, reference-free [213] and/or highly-accurate, whole-genome alignments are still necessary for our progress into understanding the *cis*-regulatory basis of adaptive evolution of the NC at a genome-wide, or at least genome-scale, resolution.

Such technological and computational developments are fundamental for us to move from anecdotal illustrations of NC evolution to a systematic understanding of the evolutionary trajectories during major transitions, such as the dawn of NC. This view favoured by us reflects the influence that technological advancements had on moving us from traditional molecular embryology to developmental biology coupled with genomics, which refined, extended and outlined novel rules within a systems-level understanding of fate-acquisition, lineage-specification, hierarchical deployment of genomic regulatory modules, and beyond [5, 32, 100, 214].

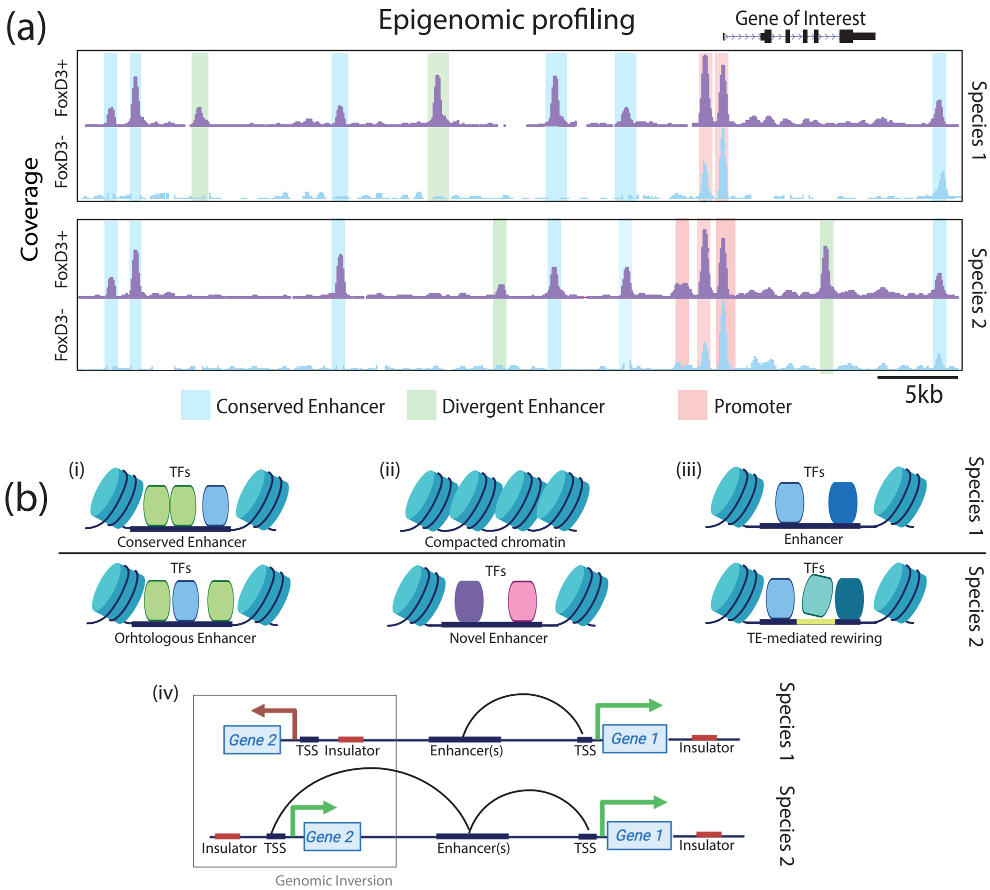

Figure 1.6 Cross-species integrative analysis of epigenomes as a manner to resolve cis-regulatory divergence during NC evolution. (**a**) Epigenomic profiling (e.g., ChIP-seq, ATAC-seq, STARR-seq, CUT&TAG) of a population of cells isolated based on a marker of interest (for example, a FoxD3+ expressing population) in the two species, or reconstructed pseudo-bulk single cell epigenomes (e.g., scATAC-seq) of orthologous cell-types enables for the identification of cis-regulatory elements (conserved and divergent) in syntenic regions. In this illustrative example, six distal-activing regions are highly conserved, whereas other regions with varying levels of conservation have diverged, including the acquisition of a novel intronic enhancer in Species 2. (**b**) Analysis of such orthologous or diverging cis-regulatory sequences across species have identified four main modes of cis-regulatory evolution: (**i**) conservation of TF binding, with varying levels of nucleotide and/or motif turn-over; (**ii**) de novo evolution of novel, functional enhancers from ancestral non-functional sequences; (**iii**) insertion of transposable-elements (TE)-containing functional, yet suboptimal, TF binding sites, thus rewiring previously existing regulatory interactions; and (**iv**) genome rearrangements, such as inversions, duplications, insertions, recombinations or deletions, generating novel enhancer-promoter interactions (represented by loops) within an insulated domain. This novel interaction switches on (green) a gene that is transcriptionally inactive (red) in the sister species.

In the context of evolutionary developmental biology, one attractive hypothesis that could benefit from such technological advancements is, for example, the idea that repeated evolution of morphological traits is coupled to parallel changes in orthologous non-coding regions [215–217]. This idea has long been hypothesised and discussed in the context of relevant NC-related loci, such as *EdnrB* and *Ret* [216]. Kingsley and colleagues have more recently been able to narrow down such loci of repeated evolution to genomic regions harbouring fragile conformations of chromatin [218] in the context of adaptations in armoured and non-armoured stickleback fishes. Such chromatin fragility causes frequent double-stranded DNA breaks, and their repair might accelerate nucleotide turnover. Similar environmental pressures in independent populations might constrain the pool of fixed, adaptive alleles within such non-coding regions, thus illustrating a mechanism that might be able to explain how convergent adaptations directly regulated by enhancers might evolve repeatedly and rapidly. However, such hypotheses still lack support from genome-wide (or at least, genome-scale) studies and across multiple taxa.

Another data-driven attractive idea that warrants further investigations of the frequency and scale of developmentally controlled transcriptional modules switching in between germ layers/developmental contexts, such as mesoderm-derived cartilage in invertebrates, in contrast to mesoderm- and NC-derived chondrogenesis in vertebrates. These events can now be investigated at high-resolution via the integration of whole-embryo, developmental single-cell transcriptomes across species [196]. These experiments will be interesting to advance our understanding of the evolutionary dynamics of such transitions, including their frequency, and illuminate the driving forces (*e.g.*, neutral drifts, directional selection, etc.) shaping embryonic development at macro- and micro-evolutionary scales.

Finally, we posit that adopting of such genomic and functional approaches in non-reference organisms will provide novel insights into the evolution of the NC. Increased taxon sampling is poised to illuminate the origins and adaptive diversifications and will further our understanding of the role of NC in the morphological evolution of vertebrates.

In conclusion, the current systems-level investigations of a relatively limited number of vertebrate taxa have suggested that NC functions as ”a transcriptional platform”, which integrates diverse developmental modules [177], such as EMT, migration, embryonic colonisation of distal sites, environmental-guidance, as well as a multitude of cellular differentiation programmes. Remarkable flexibility to integrate (and lose) such modules via regulatory changes demonstrate the utility of the NC in advancing our understanding onto the nature of emergence and diversification of novel cell types. The exciting technological advancements we witnessed in such short years aimed at interrogating whole-genomes and their regulatory architectures in small populations, or where possible, in single-cells, are thus waiting to be adopted in NC cells isolated directly from their endogenous context within the live embryo from diverse taxa - many of which remain underrepresented, such as sharks, rays, chimaeras, ray-finned fishes, lungfishes, amphibians, snakes, lizards, avians, marsupials, among many others.

We expect the development of many creative approaches to investigate the diversity of cell-types shaping life forms in general, and the vertebrate NC, in particular, will continue to benefit from the ever-expanding toolkit of these exciting genomic and computational approaches.

1.8 Thesis outline

The main aim of this Thesis is to reverse engineer the GRN governing avian NC development at two stages of development: specification at the pre-migratory stage, and early migratory. Secondly, I aim at comparing this GRN across species as a manner to decipher how GRN evolution has contributed to the early appearance and diversification of the NC within the radiation of vertebrates. In Chapter 1, I review the literature detailing the regulatory mechanisms controlling NC development, particularly focusing on the transcriptional mechanisms that have been useful to reverse engineer a provisional NC GRN. I then undertake a comparative approach, discussing evidence in the literature that reveals homology between the NC of vertebrates and embryonic cell-types from invertebrate chordates, as well as GRN mechanisms likely contributing to the early origins of the NC. I also discuss the *cis* and *trans* regulatory mechanisms that have shaped the diversification of the NC. Chapter 2 aims at detailing the first step in the process of reverse-engineering the first genome-wide GRN of the avian NC, describing the pipeline that I developed to reconstruct gene regulatory interactions at high-resolution through the integration of multimodal datasets. My results indicate the existence of two circuitries regulating NC progenitor identities, which likely contribute to ectomesechymal development and PNS neurogenesis. In Chapter 3, I first aim at examining NC evolution at single-cell resolution of the avian, zebrafish and murine NC using a novel strategy [196] to integrate single cell transcriptomic datasets within the same lower dimensional space. I then use the approach I developed to reverse engineer GRNs in Chapter 2 to investigate NC development in two species: zebrafish and lampreys. Epigenomic comparisons of chicken and zebrafish NC reveals pleiotropic *cis*-regulatory elements controlling early NC development, providing important evidence of the role of enhancer pleiotropy in the early origins of the NC. Moreover, I present data suggesting that the acquisition of novel enhancers may have contributed to NC capacity to produce ectomesechyme, as well as in further diversifying the NC within vertebrates, such as additional axial regionalisation within amniotes. Finally, Chapter 4 provide concluding remarks and future directions of this work. Materials and methods used in this work are detailed in Chapter 5.

2

Reverse engineering the avian NC gene regulatory circuitry *in vivo*

In this Chapter, I developed a computational pipeline to reverse engineer the NC GRN in a genome-wide fashion, by integrating epigenomic and transcriptomic profiles of developing avian NC cells isolated directly from their endogenous context. We have conducted an unbiased, systems-level functional study of the NC regulatory landscape *in vivo*, which enabled us to identify the dynamic remodelling of *cis*-regulatory landscapes and the spatio-temporal dynamics of enhancer deployment during NC development at unprecedented resolution. Assigning NC enhancers to downstream target promoters using a high-resolution targeted chromatin conformation capture (3C) based method (Capture-C) [219] and positional association to expressed genes, together with the definition of upstream TF inputs using scRNA-seq and motif sequence analysis has enabled comprehensive reverse engineering of the GRN underlying early NC ontogeny. *De novo* identification of the core transcriptional networks underlying NC ontogeny revealed a combinatorial *cis*-regulatory code comprised of Sox9, Otx2, NR2F2 and Pax2/3/7 orchestrating NC neurogenesis, whereas the canonical regulators Sox10, TFAP2, as well as novel regulators Arnt2 and ATF2, regulated ectomesenchymal fates. Through CRISPR/Cas9-mediated knockout of these factors, and their causal link to downstream *cis*-regulatory responsive elements, we revealed direct feed-forward loops controlled by the heterotypic binding of these TFs, as well as the cross-negative regulation of these two identities, which function as core logical features of the NC

GRN. Taken together, the ability of our approach to define and interrogate the GRN underlying a system as complex as the NC demonstrates its unique power to dissect gene regulatory circuits *in vivo*, with broad implications for vertebrate GRN discovery and study. Integration of our data sets allowed us to assemble a comprehensive GRN underlying early NC development and provide an interactive resource for exploring new regulatory hierarchies.

2.1 The *trans*-regulatory environment of developing cranial NC

We used the cranial NC-specific *FoxD3* enhancer NC1 [220] driving Citrine (Figure 2.1) to isolate avian NC cells via fluorescence-activated cell sorting (FACS) to collect Citrine-positive NC and Citrine-negative non-NC cells. We collected samples from developing chicken embryos at two stages: premigratory, 5-6ss (ss, somite-stage, HH9; approximately 36h of incubation [221]) and early-migratory, 8-10ss (HH10; approximately 42h of incubation [221]). At the pre-migratory stage, these cells maintain a neuroepithelial structure within the dorsal neural tube, forming a niche of multipotent, self-renewing cells; whereas at 8-10ss they undergo EMT, delaminate from the neural tube and start migration. Henceforth, we designated these cells as pre-migratory and early migratory NC, respectively.

We conducted RNA-seq on NC cells (Citrine+) at both stages and their corresponding non-NC (Citrine-) cells, followed by analysis of differential gene expression, which revealed approximately 400-500 genes per stage enriched in NC cells (Figure 2.2). NC TFs (Msx1, Pax7, TFAP2A/B, Sox9/10) and cell adhesion molecules (Cdh6, Cdh13, Tmem132c/d) were enriched at 5-6ss, as were signalling components (Wnt3a/8b, Bmp5, Bmper), (Figure 2.2a). At 8-10ss, NC TFs were maintained and other TFs were added to the programme, such as Tle1, Sox5/8, TFAP2C, Pax3 (Figure 2.2b). Factors involved in cell migration (Ednrb, Cxcr4), extracellular matrix (ECM) remodelling (Adamts1/20/l1), and factors re-

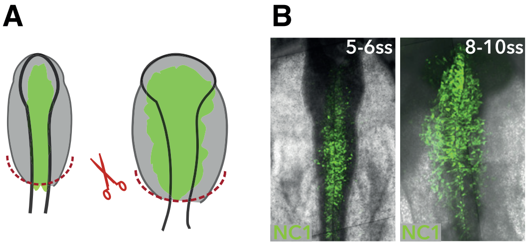

Figure 2.1 The FoxD3 enhancer NC1. (a) Embryos electroporated with the vector containing FoxD3 NC1::Citrine at gastrulation stages (HH4+) are raised until early NC formation. Schematic representation of NC cells labeled by Citrine (green) and Citrine-non-NC cells (gray) in early chicken embryos. Red dashed line indicates dissected cranial regions. (**b**) Confocal images of embryos electroporated with FoxD3 NC1::Citrine reporters at the pre-migratory (5-6ss) and early migratory (8-10ss) stages. Image adapted from [59]. Electroporation, FACS and downstream assays (RNA-seq, ATAC-seq and ChIP-seq, except for Sox10 Biotin ChIP-seq) reported in this Chapter were performed by Dr. Ruth Williams.

flecting various NC derivative lineages during NC differentiation (Col9a3, Enc1, Ltk, RXRG) were all enriched later (Figure 2.3).

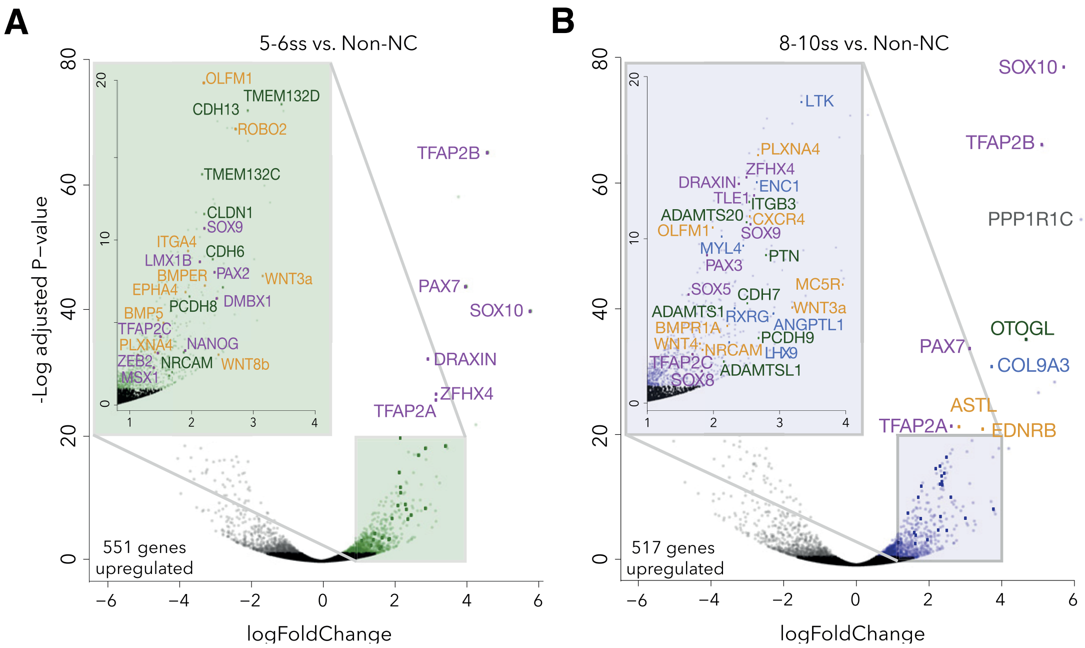

Figure 2.2 Volcano plots showing enriched genes in the chicken NC. Genes enriched (LogFoldChange >1, average of the normalized count values (base mean) >50, -Log(adjusted P-value) >2) in NC at 5-6ss (**a**) and 8-10ss (**b**) compared to corresponding non-NC cells; magenta, transcription factors; green, cell-surface molecules; yellow, signalling molecules; blue, differentiation genes. Differential expression was determined using DESeq2 with a negative binomial model, p-values calculated using Wald test, corrected for multiple testing using the Benjamin-Hochberg method (padj).

Consistent with gene ontology (GO) analysis (Figure 2.3), these results suggest that at 5-6ss, the pre-migratory NC is engaged in stem cell development and differentiation processes, whereas at 8-10ss they were starting differentiation programmes (*e.g.*, cartilage and connective tissue, myelination and skeletal muscle development). Taken together, we observed a general shift from cell-cell to cell-ECM interactions as NC cells undergo EMT and initiate delamination from the neural tube. Genes enriched in control non-NC reflect predominantly extrinsic signalling components such as BMP pathway or development of other embryonic structures, consistent with their presence in other cranial tissues surrounding the NC (Figure 2.3).

Given the wealth of transcriptomes at multiple stages of NC development and their cognate non-NC controls, we used Weighted Gene Co-expression Network Analysis (WGCNA) [222, 223] to further cluster the expressed genes into distinct subgroups (Figure 2.4). We annotated known NC factors, which allowed us to assign these subgroups to previously proposed NC GRN modules governing critical stages of early NC ontogeny, with further subdivision into early (Clusters-i and -ii) and (Cluster iii) NC specification modules. NC specifiers (Snail2, Sox8) were identified in cluster-i, along with other early NC regulators, such as Msx1 and Noelin1. Cluster-ii was characterised by NC factors Sox5 and Chd7, as well as type II cadherins (Cdh6, Cdh10) representing the progression of EMT. NC TFs (Sox10, Pax7, TFAP2A) were recovered in Cluster-iii and were maintained from 5ss onwards. NC migration was represented by the presence of ECM remodelling factors (Adamts20, AdamtsL1, Col9a1/a3) in Cluster-iv. Signalling components involved for early NC formation were present in Clusters-v and viii-x, where they were predominantly expressed in non-NC cells consistent with dynamic signalling interplay of NC cells with environing tissues (Appendix 7.1). Hence, the majority of genes depleted in NC and enriched in non-NC controls, as per our differential expression

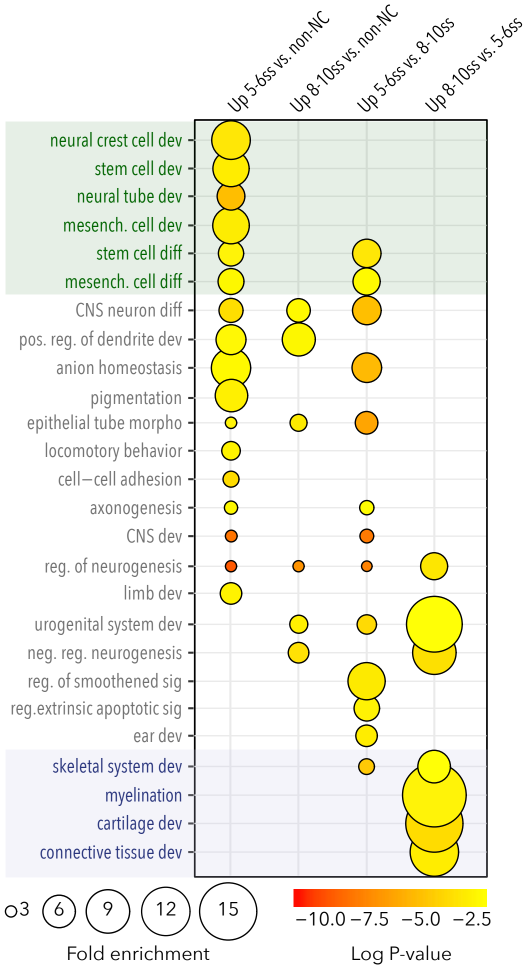

Figure 2.3 GO terms associated with NC-enriched genes. Fold enrichment was obtained using statistical overrepresentation test, p-values calculated with binomial distributions and Bonferroni correction for multiple hypothesis testing.

analysis, were recovered in Cluster-v (Figure 2.4). Such cluster assignment reflecting co-expression with known NC genes provides an excellent framework from which the role of previously uncharacterised and novel factors can be inferred. For instance, we identified Zfhx4 in Cluster-iii as co-expressed with canonical NC specifiers Pax7, Sox10 and TFAP2A/B, while Lmo4, Mef2A/C and Map4k4, together with non-canonical Wnt factors (Wnt7A/B), Plk2 and protein phosphatase 1 regulatory inhibitor subunits (PPP1Rs), were recovered in Clusters-iv and -vii, the modules harbouring downstream effectors specific to late migration/differentiation into distinct NC lineages (Figure 2.4).

Taken together, these results suggest that our WGCNA clustering parsed NC gene expression dynamics and in return, revealed factors previously undescribed in NC development. Thus, our clustering approach provides insights into the modularity of co-expressed genes, likely indicating co-regulation. Importantly, these co-expressed cluster of genes reflect specific circuitries regulating NC development, such as NC specification (Clusters-i) and EMT (Cluster-ii), expanding the depth and complexity of the NC GRN.

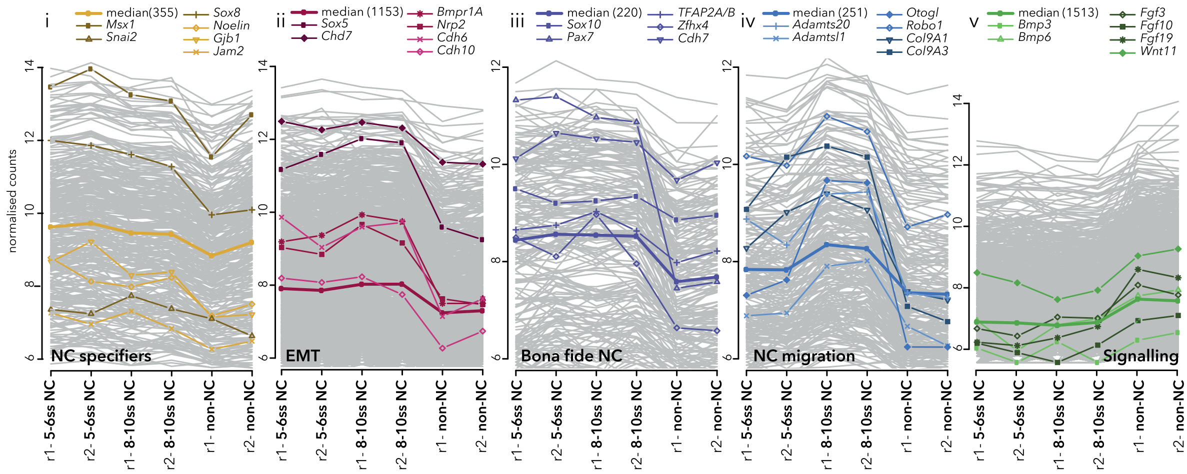

Figure 2.4 Co-expression clusters of highly correlated genes identified by WGCNA. Representative genes are shown. Two replicates per stage for NC and non-NC cells are shown on the x axis. Contributed by Dr. Daria Gavriouchkina.

2.2 Cell-type specific enhancer activities can be predicted from chromatin accessibility dynamics

We next sought to identify regulatory features controlling NC programmes. To this end, we performed ATAC-seq [224] to generate high-resolution maps of chromatin accessibility, a hallmark of *cis*-regulatory elements. In addition to NC and non-NC cells at 5-6ss and 8-10ss, ATAC-seq was also performed on dissociated cells from dissected epiblast regions at HH4 and somites from 10ss embryos. These represent excellent positive control samples such that NC could be compared to other cellular contexts. Furthermore, they can also provide insights into *cis*-regulatory elements that may be maintained from early gastrulation (*i.e.*, epiblast cells) and shared with mesoderm (*e.g.*, somites). Linear regression analysis showed a high reproducibility of biological replicates (Appendix 7.2). Exploratory analysis allowed us to notice dynamic changes in chromatin accessibility across the samples analysed. For example, in the vicinity of the NC specifier *Snail2* (Figure 2.5a) we identified a number of differentially accessible elements, open specifically in NC cells (Figure 2.5a-b), and hypothesised these could represent putative NC enhancers. To determine physical interactions between such distal regulatory elements and their putative target genes we performed Next-Generation Capture-C [219]. Using dorsal neural tubes dissected from 6ss embryos and peripheral blood cells from HH36 embryos as controls we determined the topologically-associating domain (TAD) of *Snai2* in both tissues (Figure 2.5c). We found a broad NC-specific TAD spanning 700kb, located predominantly downstream from the promoter and encompassing a number of open chromatin elements.

Using *in vivo* reporter assays we identified five active enhancers within the *Snai2* TAD, two distal (enh-332 and enh-334) and three proximal (enh-13, enh-239, enh-242), (Figure 2.5). Enh-13 (Figure 2.6d-e) was active in the cranial neural tube including NC at 5ss becoming confined to NC cells by 9ss. Enhancers-334, 242 and 239 were active in both NC and non-NC at 6-10ss (Figure 2.6a-c'), consistent with these regions being accessible in both NC and non-NC cells. A region specifically open in the somites was found to have somite-specific activity (enh-241, Appendix 7.4). Overall this exploratory data analysis suggested that chromatin accessibility dynamics across multiple cell types and developmental stages reflects tissue-specific enhancer activity.

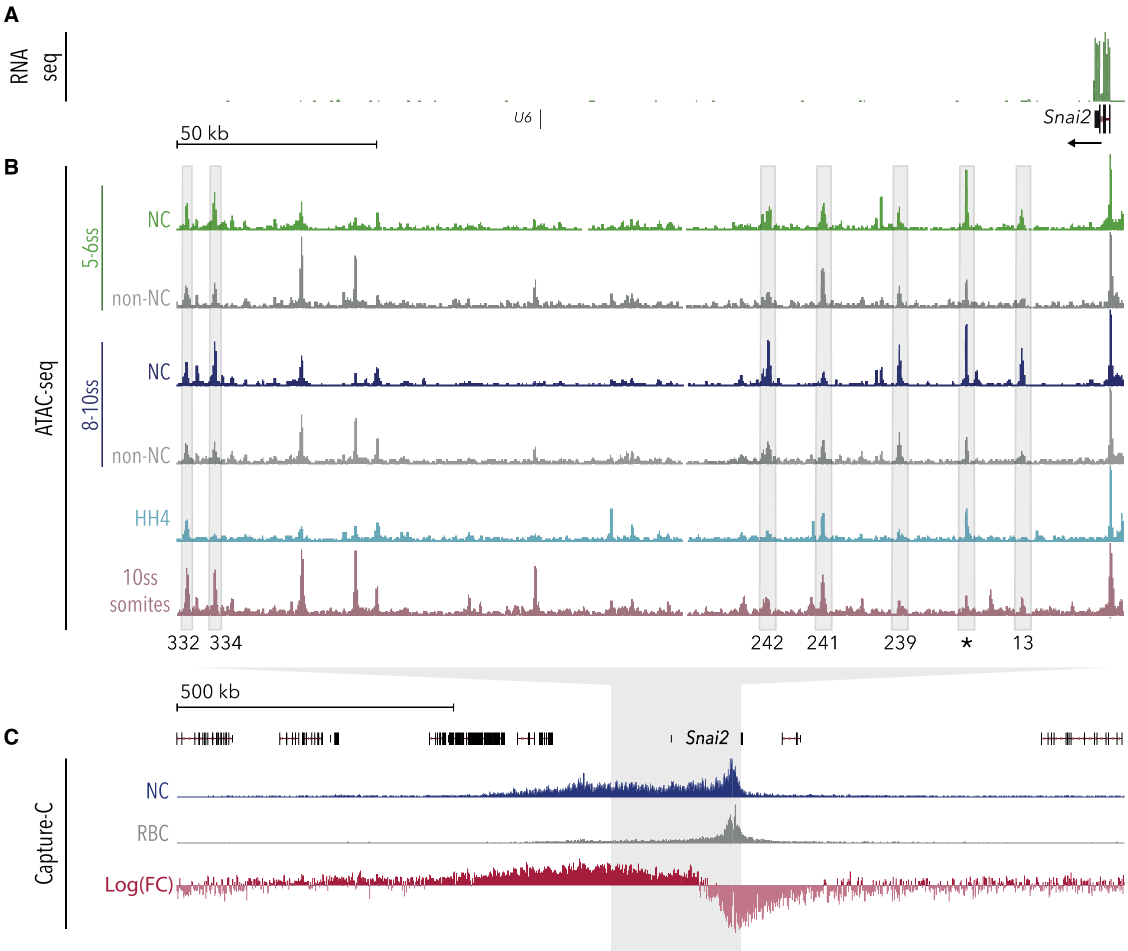

Figure 2.5 Profiling Chromatin Accessibility Dynamics during Early NC Development. Genome browser views of (**A**) RNA-seq data (5-6ss; pre-migratory NC) and (**B**) ATAC-seq profiles in NC and non-NC cells upstream of the *Snai2* locus. 5-6ss NC ATAC is shown in green and 8-10ss in blue; corresponding non-NC samples are shown in gray. ATAC data from HH4 and somite tissue are shown in light blue and pink, respectively. Vertical grey bars indicate putative enhancers tested *in vivo*. Asterisk (*) highlights a putative enhancer that did not drive fluorescent reporter expression when tested *in vivo* (**C**) Capture-C tracks showing the TAD at the Snai2 locus in NC (blue track) and control red blood cells (RBC, gray track). Differential interactions were determined using DESeq2, with statistical support calculated by Wald tests and corrected for multiple testing using the Benjamin-Hochberg method. Ratio of LogFoldChange (logFC, red track) values and their standard errors were calculated with Wald statistics in DEseq2. Capture-C data analysis highlighted in **c** was contributed by Dr. Jelena Telenius and prof. Jim Hughes; electroporations and genomic assays were performed by Dr. Ruth Williams.

2.3 Dynamic deployment of enhancers during NC ontogeny

We next sought to interrogate the regulatory information encoded in enhancer elements. We first performed *k*-means linear clustering of our ATAC-seq datasets

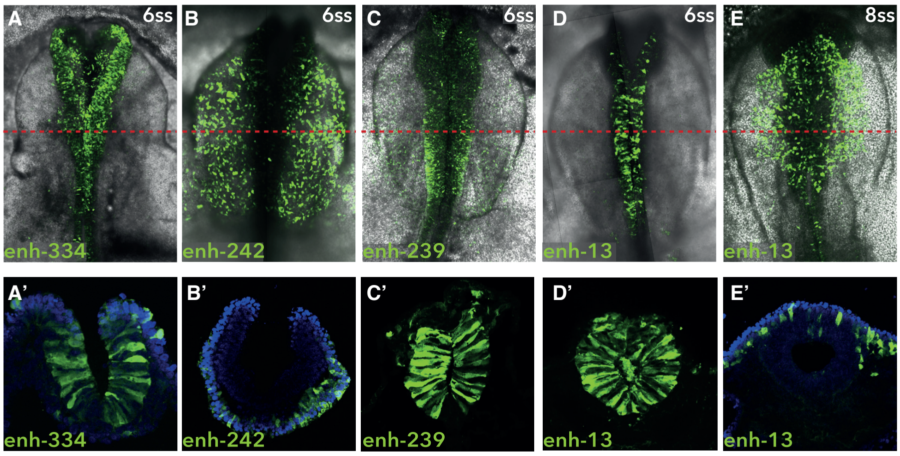

Figure 2.6 *In vivo* enhancer reporter assays for Snai2 enhancers. (A-E) Enhancer driven *in vivo* reporter (Citrine) expression of tested enhancers and their respective transverse sections; red line indicates approximate location of section. Blue, DAPI. Enhancer names correspond to the annotations in Figure 2.5. Contributed by Dr. Ruth Williams.

using the seqMINER platform [225], which assigned all open chromatin peaks genome-wide into patterns of chromatin accessibility. We focused on the top seven clusters based on their density and patterns of accessibility in NC cells (Figure 2.7a). Broad signal elements constitutively accessible across all samples were recovered in *k*-Cl-4 (Spearman correlation coefficient r=0.94), (Figure 2.7a-b); 20% of which represented promoter regions (Figure 2.8), suggesting that the remaining 80% of elements may represent other classes of *cis*-regulatory elements or non-annotated promoters. Elements specifically accessible in NC cells were found in *k*-Cl-3 (r=0.74, $p<2.2 \times 10^{-16}$), (Figure 2.7), which we assigned as the canonical NC enhancer group. Accessibility in NC cells also characterised the *k*-Cl-1 elements (r=0.74, $p<2.2 \times 10^{-16}$), (Figure 2.7b).; however, these were also open in Citrine- non-NC cells, indicating possible activity in other cell-types. Chromatin accessibility was also seen in pre-migratory NC and in the early (HH4) epiblast (*k*-Cl-6), consistent with previous reports suggesting that the NC regulatory landscape is established as early as during gastrulation [77, 226]. *K*-Cl-6 accessibility was reduced in NC cells at 8-10ss, suggesting these elements might be involved in the early establishment of NC programmes and are later decommissioned. Elements recovered in *k*-Clusters were predominantly located in intergenic or intronic regions (Figure 2.8), consistent with their *cis*-regulatory activity. These results thus establish the dynamic *cis*-regulatory landscapes across different stages and cell-types likely underlying NC ontogeny.

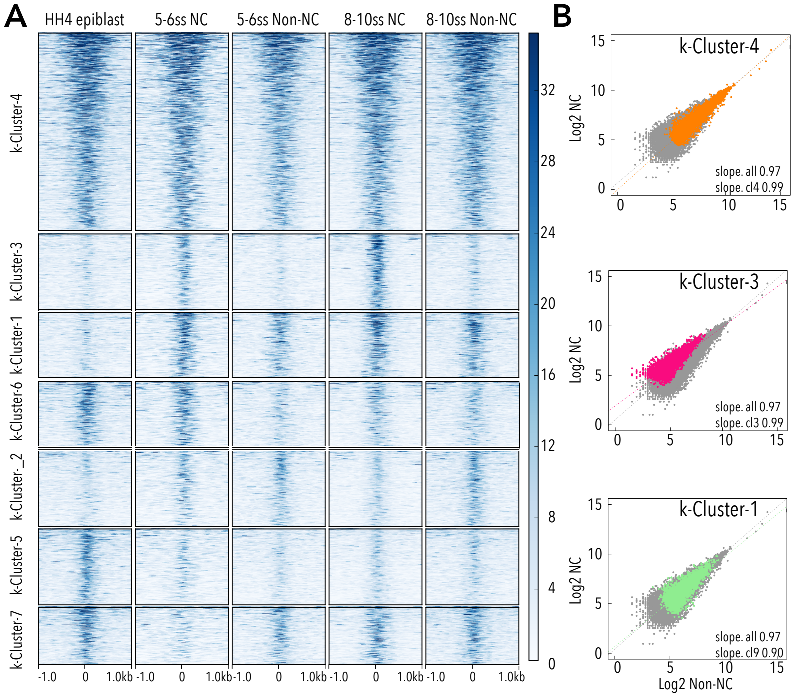

Figure 2.7 *K*-Means Clustering and Differential Analysis of Chromatin Accessibility during Early NC Development. (a) Read-count-normalized ATAC-seq samples obtained from NC and non-NC control cells at HH4, 5-6ss, and 8-10ss were reiteratively assigned into cluster of elements (k=10) showing similar patterns of chromatin accessibility dynamics genome-wide using k-means algorithm. Only the top 7 clusters are shown. (b) Log-scaled scatter plots of normalised ATAC-seq counts of selected k-means clusters, quantifying differential accessibility by calculating slopes and Spearman correlation coefficients in NC and control (non-NC) cells.

In parallel, we employed a stringent statistical analysis of differential accessibility (DA) in purified NC cells compared to neighbouring non-NC cells using the DEseq2 algorithm, part of the DiffBind package. We identified 1605 statistically significant NC-specific elements (from now on referred to as DA elements) at 5-6ss and 5734 at 8-10ss (FDR<0.1, Log2FoldChange>1), (Figure 2.9), suggesting the progressive

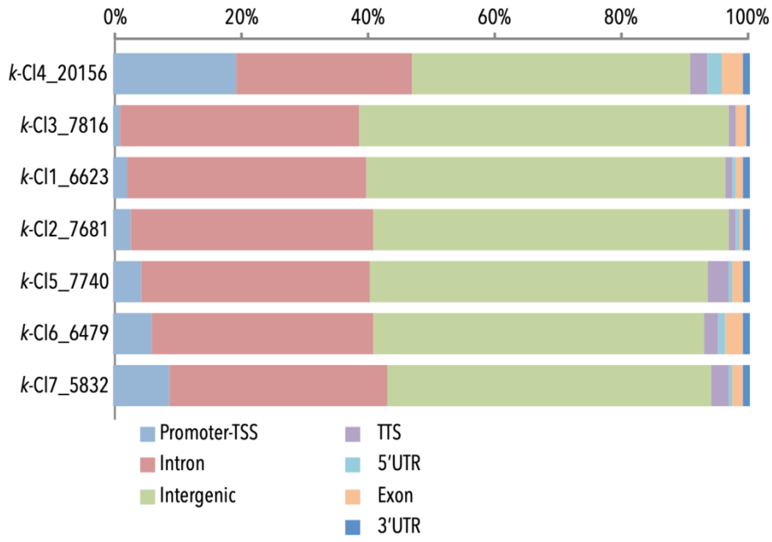

Figure 2.8 Stacked bar plot showing genomic annotation of *k*-Cluster elements. The number of elements in each k-Cluster is also shown.

establishment of the canonical NC *cis*-regulatory landscape.

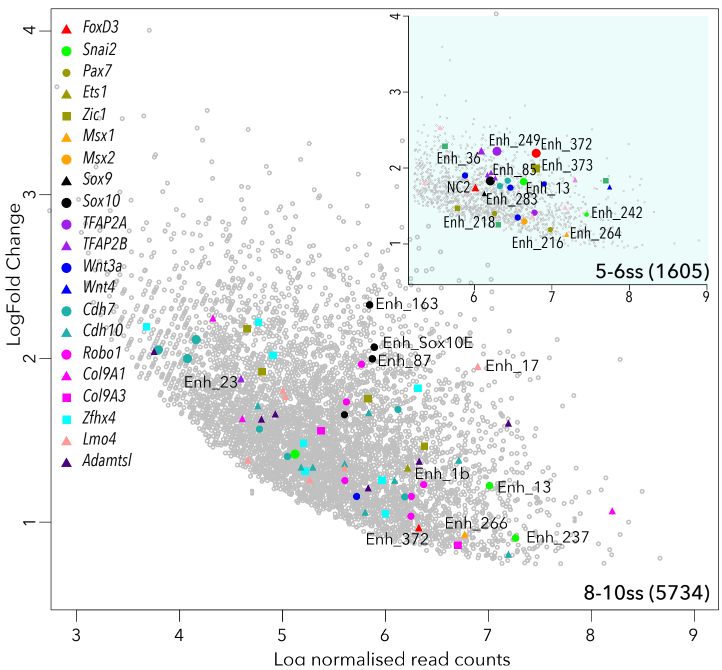

Figure 2.9 MA plot depicting differentially accessible elements. DA elements using a negative binomial distribution model (FDR < 0.1, Fold enrichment > 1), A selection of putative enhancers to key NC genes are color coded; numbered enhancers have been tested *in vivo*.

DA elements were enriched at key NC loci (Figure 2.9) and differentially overlapped individual k-Clusters (Figure 2.10). Consistent with their early opening, DA elements at 5-6ss were mainly shared with k-Cl-1 (27%), k-Cl-6 (14%), and to a lesser extent with k-Cl-3 (7%). DA elements at 8-10ss predominantly overlapped NC k-Cl-3 elements (47%), and to a lesser degree k-Cl-1 (12%) and k-Cl-6 (6%), (Figure 2.10).

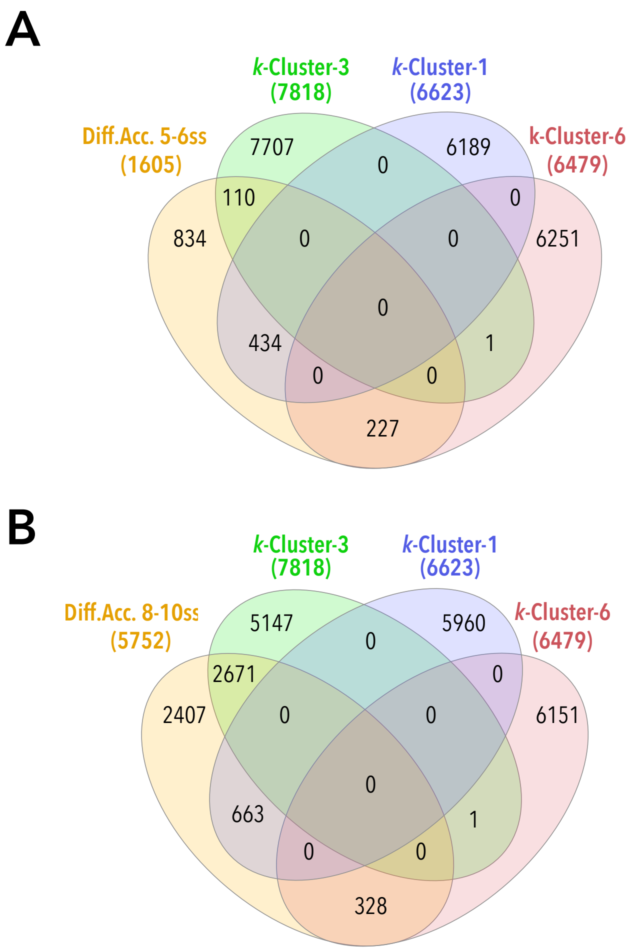

Figure 2.10 Venn diagrams showing overlap of k-Cluster and DA elements Overlaps are shown for (a) DA elements at 5-6ss (b) and 8-10ss.

Assignment of *cis*-regulatory elements from k-Cl-3 and k-Cl-1 to the nearest expressed gene indicated that key NC genes were controlled by multiple regulatory elements. We further distinguished genes predominantly regulated either by specific k-Clusters to those with shared regulation from two or more k-Clusters. For example, canonical NC factors *Sox10*, *EdnrB* and *Myc*, along with lesser-characterised NC genes such as *Lrrc4c* and *Epha5* [227, 228] were enriched for k-Cl-3 elements (Figures 2.11, 2.12a). In contrast, genes involved in neural development such as *Sox11*, *Grid2*, *MycN* and cell migration/guidance *Fgfrl1*, *Wnt5a* and *Sema3a* were largely enriched for k-Cl-1 elements (Figure 2.12b). However, other NC genes *TFAP2B* and *Snai2*, and EMT factors *Zeb2* and *Lmo4* were driven by a combination of k-Cl-3 and k-Cl-1 elements (Figure 2.12a), suggesting a complex interplay between cell-types, developmental stages and *cis*-regulatory dynamics. Collectively these results suggest that distinct clusters of enhancers may differentially regulate different NC lineages, suggesting some level of modularity to enhancer activities.

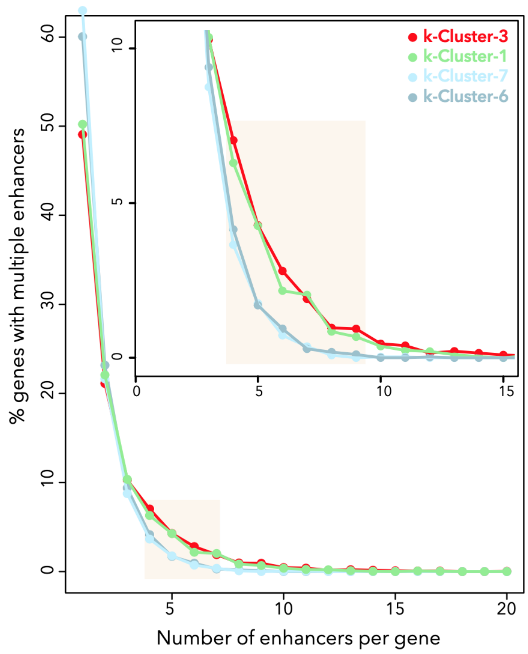

Figure 2.11 Percentage of genes with multiple associated enhancers from k-clusters.

Dr. Ruth Williams extensively tested the genome-wide enhancer predictions by assessing the spatiotemporal activity of putative enhancers *in vivo* (Appendix

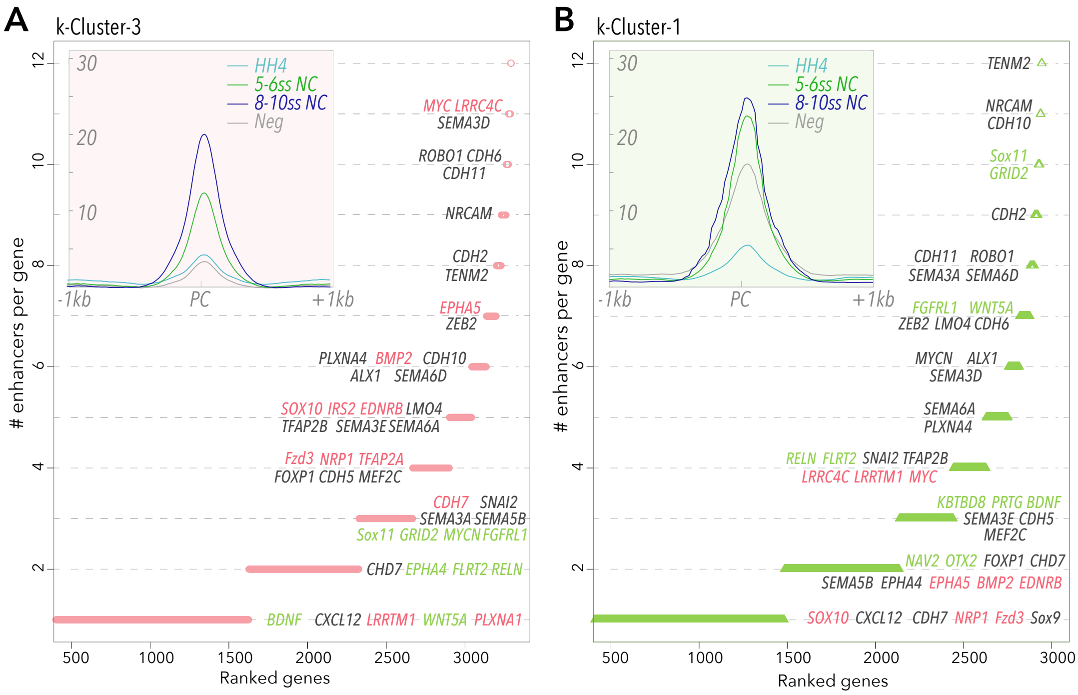

Figure 2.12 Gene rankings by number of associated *k*-Cl-3 and *k*-Cl-1 elements. (a) Genes associated to *k*-Cl-3 and (b) to *k*-Cl-1. Inserts show mean merged accessibility profiles per cluster.

7.4). Moreover, we determined the TADs by Capture-C and annotated putative enhancers of *Sox10*, *Lmo4* and *Zfhx4*. TADs varied in size and distribution. For example, the *Lmo4* TAD spread 500kb downstream and 200kb upstream of the promoter (Figure 2.13a), whereas the *Zfhx4* TAD spanned 700kb predominantly upstream of the coding region (Figure 2.13b). Importantly, the *k*-Cl-3 enh-17 element specific to NC cells displayed activity in early delaminating and migratory NC, whereas *k*-Cl-9 enh-15 was active in non-NC cells of the ectoderm (Figure 2.13a). similarly, we identified two enhancers, the *k*-Cl-9 enh-5 and the *k*-Cl-4 enh-316 displaying activity specifically in delaminating NC cells. These results validate our integrative framework for identification of cell-type- and stage-specific NC enhancers and provide a resource for the study of NC gene regulation *in vivo*.

Taken together, our results suggested that differential chromatin accessibility analysis and *k*-means clustering captured complementary aspects of the NC *cis*-

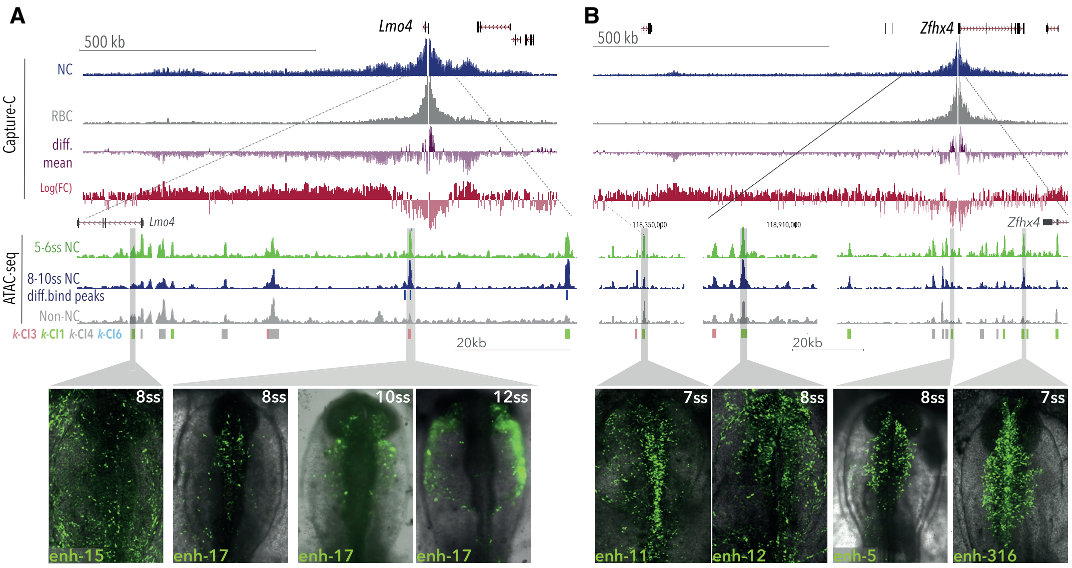

Figure 2.13 Genome browser snapshots of *Lmo4* and *Zfhx4* *cis*-regulatory landscapes. (a) *Lmo4* and (b) *Zfhx4* loci are shown along with Capture-C tracks (blue, NC cells; gray, RBC; purple, differential mean between experimental and control samples; red, LogFoldChange values calculated by DESeq2. P-values were calculated with Wald test and Benjamin-Hochberg correction. Statistically significant chromatin interactions identified by Capture-C are depicted by positive values in the LogFoldChange (red) track. ATAC-seq (green, 5-6ss; blue, 8-10ss; gray, negative non-NC) tracks; annotation of *k*-means elements are shown. *In vivo* activity of selected enhancers, boxed in gray, is shown at the bottom. *Zfhx4* and *Lmo4* featured NC-specific upstream regulatory apparatus. *Zfhx4* expression is solely controlled by *k*-Cl-1 enhancers, with NC-specific activity (enh-5 and enh-316) as well as non-NC regulation (enh-11 and enh-12). *Lmo4* NC-specific expression is governed by *k*-Cl-3 enh-17 and non-NC activity regulated by *k*-Cl-1 enh-15. Capture-C data analysis was contributed by Dr. Jelena Telenius and prof. Jim Hughes; electroporations, genomic assays and confocal imaging were contributed by Dr. Ruth Williams.

regulatory landscape. On one hand, *k*-Cl-3 display activity of NC enhancers, showing motif content, enhancer-promoter contacts and *in vivo* enhancer activity in a manner largely specific to NC cells. On the other hand, although some *k*-Cl-1 elements display enhancer activity specifically to NC cells, we uncovered several instances in which they are active in non-NC cells, thus suggesting these elements are shared with different cell types.

2.3.1 An enhancer cluster regulates *Sox10*

The observed spatial organisation of dynamically accessible regions, as well as the increased ratio of enhancers per gene at developmental loci (Figure 2.12) suggested that enhancers form clusters of elements within TADs, similar to ”super-enhancers” (SEs) [121, 129]. Importantly, SE-like regions function as core regulatory circuits within GRNs, specifying the cell-type identity and their dysregulation can impair cellular programming and be involved in the emergence of neoplasias [121, 129, 229]. H3K27Ac ChIP-seq in NC cells showed high enrichment for this histone modification, considered a hallmark of SE-like regions [128]. ROSE (Rank Ordering of Super Enhancers) algorithm [127, 128] on H3K27Ac ChIP-seq profiles identified 1379 SEs (Figure 2.14a) at the delaminating NC stage (8-10ss), and 1288 SEs at the pre-migratory NC stage (5-6ss). Highly ranked SE-like clusters were associated to canonical NC regulators loci, including *Sox10*, *Fzd3*, *TFAP2A*, *TFAP2B*, *Chd7*, *Sox5* and *Snai2* (Figure 2.14a) at 8-10ss, but not at the earlier pre-migratory stage (data not shown), suggesting remodelling of the NC *cis*-regulatory landscape at delaminating/actively migrating stage, when the majority of enhancers are deployed. Moreover, we inferred occupancy of the Mediator complex by ChIP-seq for the transcriptional regulator Brd4, suggesting that approximately 30% of NC elements displaying differential accessibility at 8-10ss had SE-like characteristics (Figure 2.14b).

Given the pivotal role of *Sox10* in NC ontogeny [32, 230, 231] and its high SE ranking (Figure 2.14a), we sought to thoroughly survey the *Sox10* locus at the epigenomic level. We identified a cluster of open elements spanning a 20kb region, 2kb upstream from the promoter, within the defined *Sox10* TAD (Figure 2.15a). Each of these elements were present in our DA sets and the majority were recovered in *k*-Cl-3, with the exception of the two most distal peaks (enh-99 and enh-89), belonging to *k*-Cl-1 and *k*-Cl-6, respectively (Figure 2.15a). *In vivo* en-

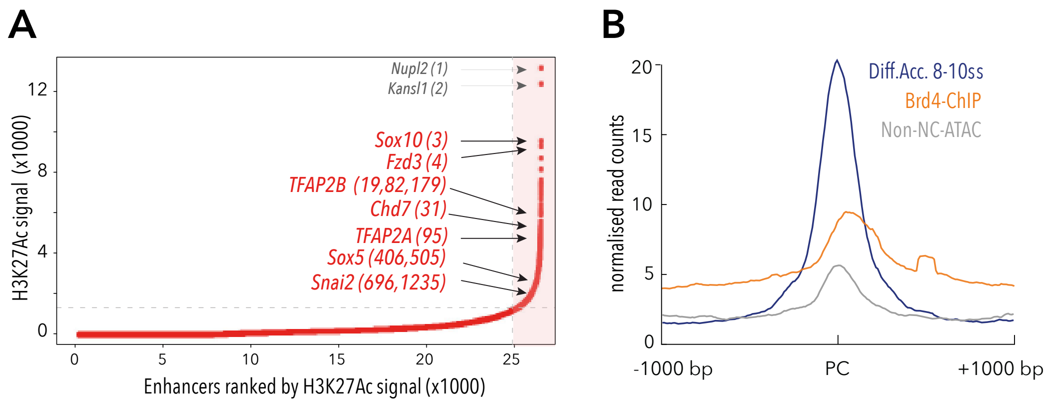

Figure 2.14 Super-enhancer-like Clusters Regulate Key NC Genes. (a) Super-enhancers ranked by H3K27ac signal in 8-10ss NC, using the ROSE algorithm; top-ranked NC genes are annotated. (b) Mean merged profile of DA elements at 8-10ss and control non-NC elements occupied by Brd4.

hancer reporter assays showed these elements have diverse spatiotemporal patterns of activity in NC cells (Figure 2.15b-g), collectively reconstituting the endogenous *Sox10* expression pattern. Enh-99 identified in *k*-Cl-1 and co-occupied by Brd4, is the earliest acting *Sox10* enhancer (detected from 5ss) that likely onsets the *Sox10* expression (Figure 2.15b). Interestingly, we found that a specific region within the enh-99 (Enh-99a) predominantly drove specific reporter activity in the trunk NC (Appendix 7.4). Enh-99 activity was followed by enhancers enh-87, -85, and -84 which onset at 7ss (Figure 2.15c-g) and continued throughout early NC migration, with enh-84 activity being particularly prominent in migrating NC (8-10ss), consistent with the occupancy of these elements by Brd4 (with the exception of enh-85) and their assignment to the *k*-Cl-3. Surprisingly, enh-89, a *k*-Cl-6 element accessible from the epiblast stage, was active at 8ss, predominantly at the hindbrain level, suggesting this enhancer might be primed early in development for later deployment in the NC. Conversely, enh-87, a *k*-Cl-3 element, was active in a complimentary domain (anterior cranial region, figure 2.15c), suggesting that enh-89 and enh-87 collectively controlled cranial *Sox10* expression in a manner redundant to enh-99, possibly acting as shadow enhancers. All five new

enhancers displayed statistically significant interactions with the *Sox10* promoter (Figure 2.15a). The most proximal open element, enh-82, which did not display any *in vivo* enhancer activity at the stages tested, possibly represents a *cis*-regulatory repressor, a late-acting enhancer or an insulator. Interestingly, we could recover both previously described *Sox10* enhancers (cranial-specific, 10E2; trunk and placodes, 10E1 [230]) in our DA analysis, suggesting chromatin accessibility at these enhancers is not differentially regulated at the cranial versus trunk levels. However, only the cranial-specific enhancer 10E2 displayed statistically significant interactions captured from the promoter by Capture-C.
(Figure 2.15h)

Collectively, our results suggests that additive or sub-additive, combinatorial action of enhancers might be critical for sustained *Sox10* expression during NC ontogeny. It will thus be interesting to test different combinations of enhancers within the same construct in order to further test their additivity/synergy in the future. Moreover, our results also provided a locus specific example of the global dynamics of NC transcriptional regulation at a key SE, by providing evidence that: (1) *k*-Cl-1 elements may be required to initiate NC-specific programmes, as in the case of enh-99, and were shared with neural progenitor cells (such as trunk neural tube cells, enh-99a); and (2) multiple *k*-Cl-3 elements (enhs-84, 85, 87 and 10e2) represent cell-type restricted, NC enhancers essential for transcriptional robustness within the NC-GRN.

In fact, the spatial organisation of the *Sox10* SE, with the early acting enhancers located at the 3' most region (enh-99) towards later acting, proximally located enhancers (*e.g,* enh-84) are consitent with the idea that within a enhancer cluster, enhancers are activated in a collinear fashion.

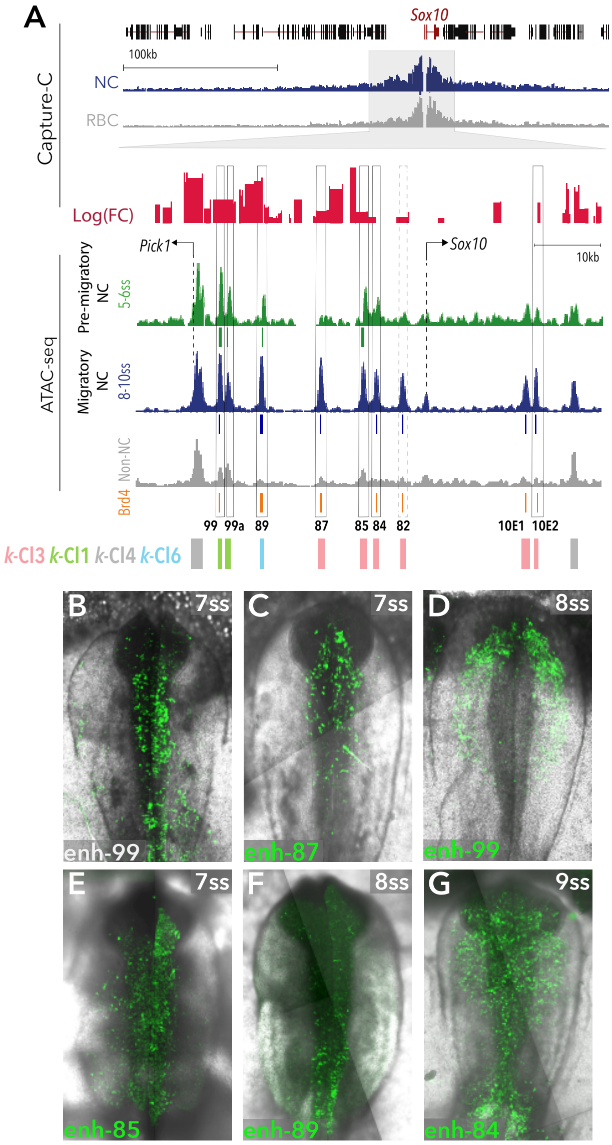

Figure 2.15 Dissection of the *Sox10* super-enhancer-like cluster.(a) Capture-C genome browser tracks at the *Sox10* locus. Blue, NC cells; gray, RBC; red, log-transformed adjusted P-value, representing ratio of LogFoldChange values and their standard errors, determined using DESeq2. P-values were calculated with Wald test and Benjamin-Hochberg correction and ATAC data (green, 5-6ss; blue, 8-10ss; gray, non-NC). DA elements at 5-6ss (green) and 8-10ss (blue), Brd4-bound peaks (orange), and *k*-cluster elements are indicated. Selected putative enhancers are boxed and their *in vivo* activity is shown. Confocal imaging of active enhancers: (b) distal k-Cl-1 enh-99 at 7ss, (c) enh-87 at 7ss, (d) enh-99 at 8ss, (e) enh-85 at 7ss, (f) k-Cl-6 enh-89 at 8ss, and (g) enh-84 at 9ss. Capture-C data analysis was contributed by Dr. Jelena Telenius and prof. Jim Hughes; electroporations, genomic assays and confocal imaging were contributed by Dr. Ruth Williams.

2.3.2 An enhancer cluster regulates the chromatin remodeller *CHD7*

Surprisingly, we found that a high-ranked locus in our SE analysis (Figure 2.14a) is associated to the chromodomain helicase DNA-binding domain member *CHD7*, an ATP-dependent chromatin remodeller homologous to the *Drosophila* trithorax-group protein Kismet previously shown to co-bind with Sox9 and Sox10 [51]. Although CHD7 is also associated to PBAF (polybromo- and BRG1-associated factor-containing complex, also known as SWI/SNF complexes in mammals) and remodels chromatin accessibility in multiple cellular contexts [51], this result led us to investigate whether this chromatin remodeller displays NC-specific regulation *in vivo*.

We first determined the *CHD7* TAD by Capture-C, revealing a large NC-specific TAD, spanning approximately 1.1Mb: 85Kb upstream and 1Mb downstream from the Chd7 promoter, encompassing the entire gene body (Figure 2.16a). ATAC-seq profiles of NC cells at 5-6ss and 8-10ss revealed multiple putative *cis*-regulatory elements with characteristics of developmental regulation, including *k*-Cl-3-like accessibility (Figure 2.16b).

A cohort of candidate elements, to which we proposed the names enhancers A, C, F, H and T (Figure 2.16b) were successfully cloned and tested *in vivo*. Enhancer reporter assays validated specific activity in NC cells (Figure 2.17), in a manner similar to the uncovered enhancers from the canonical NC regulator *Sox10* (Figure 2.15b), in addition to non-NC cells, such as the hindbrain. Enh-A was specifically active in the migratory cranial NC at the midbrain levels; enh-C displayed specific activity in the neural tube, markedly at post-midbrain levels; enh-F was active in the neural tube and ectoderm; enh-H and -T displayed activity in migratory NC in a manner similar to enh-A (Figure 2.17).

We next dissected the motif content of these *cis*-regulatory elements. A motif database including canonical NC factors, such as TFAP2-family members, SoxE,

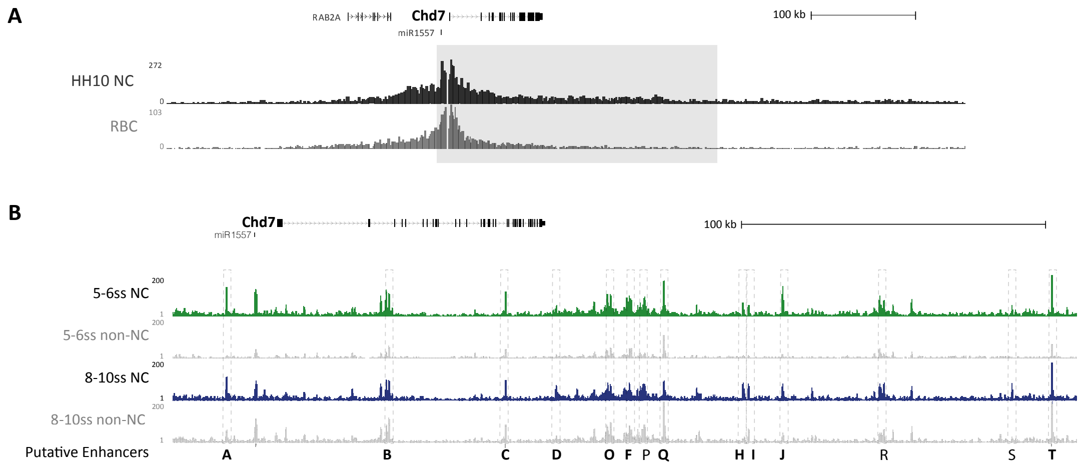

Figure 2.16 *Chd7* enhancer cluster. (a) Capture-C genome browser tracks at the *CHD7* locus (dark gray, HH10 NC cells; light gray, RBC). Statiscally significant interactions, identified by the ratio of LogFoldChange values and their standard errors, were determined by DESeq2, and p-values calculated with Wald test and Benjamin-Hochberg correction. Boxed region is the genomic region showed below. (b) ATAC-seq data (green, 5-6ss; blue, 8-10ss; gray, non-NC). Selected putative enhancers (enhancers A-T) are boxed. Electroporations and genomic assays were performed by Dr. Ruth Williams.

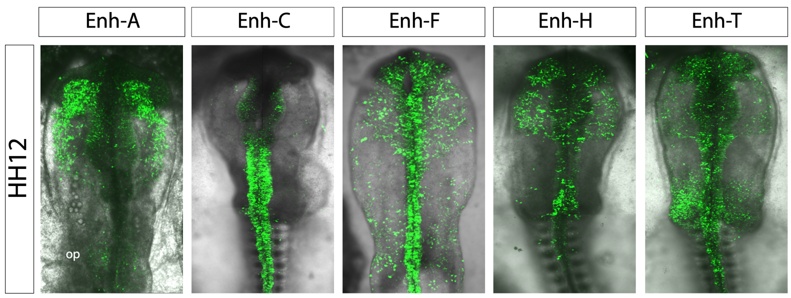

Figure 2.17 NC-specific *cis*-regulatory control of the chromatin remodeller *Chd7 in vivo* Enhancer reporter assays of putative Chd7 regulatory elements driving Citrine expression. Enh-A, Enh-C, Enh-F, Enh-H and Enh-T are shown in embryos at HH12. Confocal images were contributed by Dr. Ruth Williams.

Ets1, ZIC and SMAD, among others, was screened on these elements and enrichment was calculated by log-transformed Odds scores (Figure 2.18a). Interestingly, we found motif enrichment of neural tube associated regulators onto enh-C, such as Zic2 and RXRA/B, consistent with enh-C activity in the hindbrain. This result indeed suggests this element is driven by these TFs. The unique combination of TF binding sites onto each of these elements support NC-specific control of *Chd7*, while also indicating multiple mechanisms for *Chd7* regulation within the same cell population.

Importantly, we uncovered unique contributions of Sox10 inputs to *Chd7* enhancers A, C, F, P, Q, H, R and S, supported by the *in vivo* activity of enh-A and enh-C (Figure 2.17) in a manner that is consistent with Sox10 regulation. To further validate this hypothesis, we developed a toolkit for *in vivo* biotinylation of target proteins, which enable us to optimise and develop Biotin-Streptavidin based Chromatin Immuno-Precipitation (biotin ChIP-seq) in the avian model (Figure 2.18b). Specifically, we designed universal recipient vectors that enable C-terminal and N-terminal Avi-tagging of a protein-of-interest. We also incorporated an enhancer cloning site such that the Avi-tagged-TF can be tissue-specifically expressed under the control of its cognate enhancer and made these constructs available (Addgene #110204, #110205). Our approach enable us to isolate profile TF binding directly in their endogenous tissue, at different stages and in a genome-wide fashion.

To assess Sox10 binding to *Chd7* enhancers, we cloned the *Sox10* Enh-99, driving a C-terminally tagged Sox10. Following co-electroporation with ubiquitously expressed biotin ligase, BirA, (Addgene #127781) into HH4 embryos, the Avi-tagged-TF was biotinylated *in vivo*, nuclei were isolated, chromatin was sonicated, and TF-bound regions were pulled down by Streptavidin beads, allowing us to build Sox10 ChIP libraries and sequence (Figure 2.18c). Our results revealed that all active *Chd7* enhancers showed some binding by Sox10, compared to the control, providing strong support to the idea that Sox10 is a key player in Chd7 regulation (Figure 2.18c). In fact, NC-specific enhancers, Enh-A and Enh-T showed particularly high Sox10 binding, consistent with our motif analysis.

Next, we compared the *Chd7* locus between the chicken and human genomes. To this end, we examined epigenomic data obtained from *in vitro* derived human cranial NC cells [50, 206]. Specifically, we used ChIP data for histone modifications associated with active enhancers (H3K4me1, H3K27ac), promoters (H3K4me3) and repressed chromatin (H3K27me3), as well as ATAC-seq data. The majority of avian *Chd7* enhancers were detected by LiftOver, with the exception of Enh-A and Enh-T which we predicted based on synteny with other elements. We found a number of avian *Chd7* enhancers were marked by histone modifications indicative of enhancer elements (Figure 2.18d). In fact, putative human enh-A region was marked by H3K4me1, suggesting this is a poised element in the human NC (Figure 2.18). Taken together, our results underscores the importance of Sox10-driven combinatorial regulation of *Chd7*, and indicate this might be a conserved feature in the human genome.

Finally, we functionally perturbed the upstream TFs controlling enhancer activity. We employed a CRISPR/Cas9-mediated knockout approach [232], in which bilateral assays were performed (n=10 in each experiment). Using gRNAs previously described [232], we electroporated the left side of each embryo with experimental gRNAs, whereas the right side was electroporated with scrambled guides serving as positive controls, which allowed us to assess enhancer activity within the same embryos. Both sides received Cas9-RFP. We focused our functional assessments on the developmental enhancers: Enh-A, Enh-T and Enh-C. (Figure 2.19)

Enhancer activity, motif content, and Sox10 binding suggests Enh-A is driven by canonical NC regulators, including but not limited to Sox9, Sox10 and Ets1. We thus co-electroporated gRNAs targeting these factors in combination with Enh-A driving Citrine, thus allowing us to directly measure enhancer pertubation by confocal imaging of Citrine. Our results indeed reveal a reduction in enhancer activity in the experimental side compared to the control side, providing support to enh-A activity being controled by canonical NC regulators (Figure 2.19). We also tested if Sox2/3 factors were important for Enh-A activity, however combined knock-out of these factors had no effect on Enh-A driven Citrine expression (data not shown). These results thus suggest that enh-A activity in migratory NC is

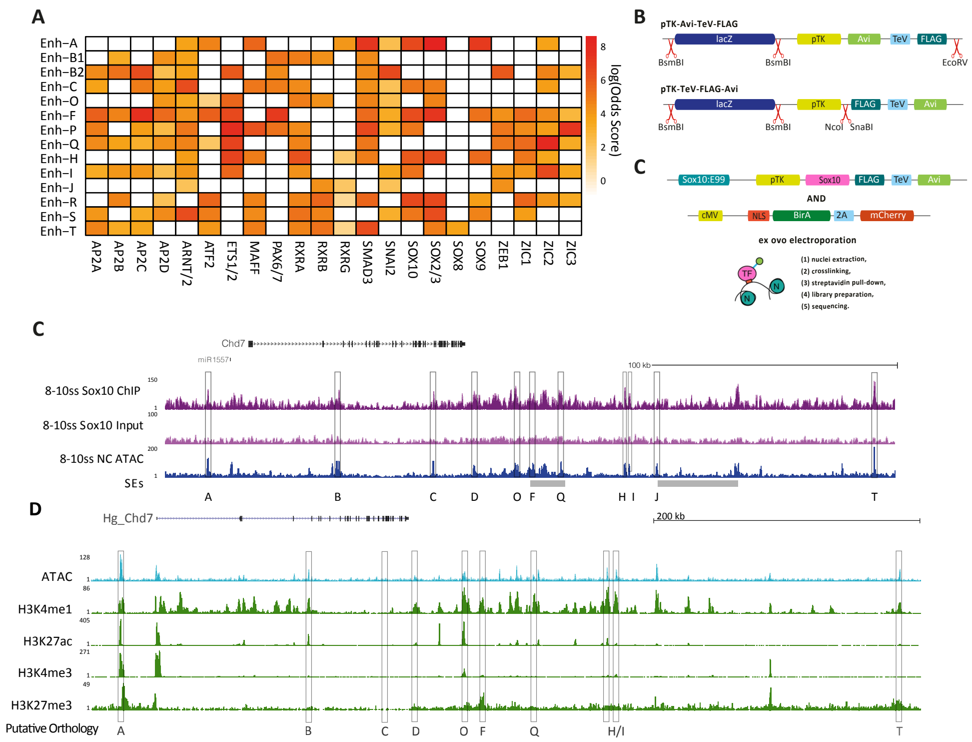

Figure 2.18 Sox10 regulates *Chd7* enhancers. (**a**) Motif composition analysis of avian *Chd7* enhancers. The heatmap shows log(Odds Scores) of screened motifs, predictive of TF binding. (**b**) Biotin ChIP-seq toolkit for *in vivo* biotinylation of TFs of interest. We designed two universal recipient vectors, which allow for N- and C-terminal Avi-tagging of a protein of interest. Vectors also containing a enhancer-cloning site. (**c**) Sox10 biotin ChIP-seq workflow. (**d**) Sox10 binding to *Chd7* enhancers. Magenta track shows normalised genome-browser tracks of the Sox10 biotin ChIP-seq from samples collected at 8-10ss; light-purple represents the input; blue, 8-10ss ATAC-seq. Grey bars indicate the location of Chd7 super-enhancer-like clusters. (**e**) Genome-browser tracks of ChIP-seq data for histone modifications associated to *cis*-regulatory in an *in vitro* model of human NC. Boxed are putative orthologous enhancers between human-chicken genomes. One biological replicate of Sox10 biotin ChiP-seq was contributed by Dr. Ruth Williams.

specifically regulated by canonical NC regulators.

Similarly, we targeted Pax7 and Sox9 to functionally validate enh-C, which is active in the hindbrain. Consistent with our motif composition analysis, this functional perturbation caused a a reduction in enhancer activity on the experimental side, indeed suggesting Pax7/Sox9 control of *Chd7* enh-C (Figure 2.20). Finally, we tested if a combination of Sox3, Sox10 and TFAP2b regulated enh-T as indicated by motif content analysis. Again, our functional experiments support our *in*

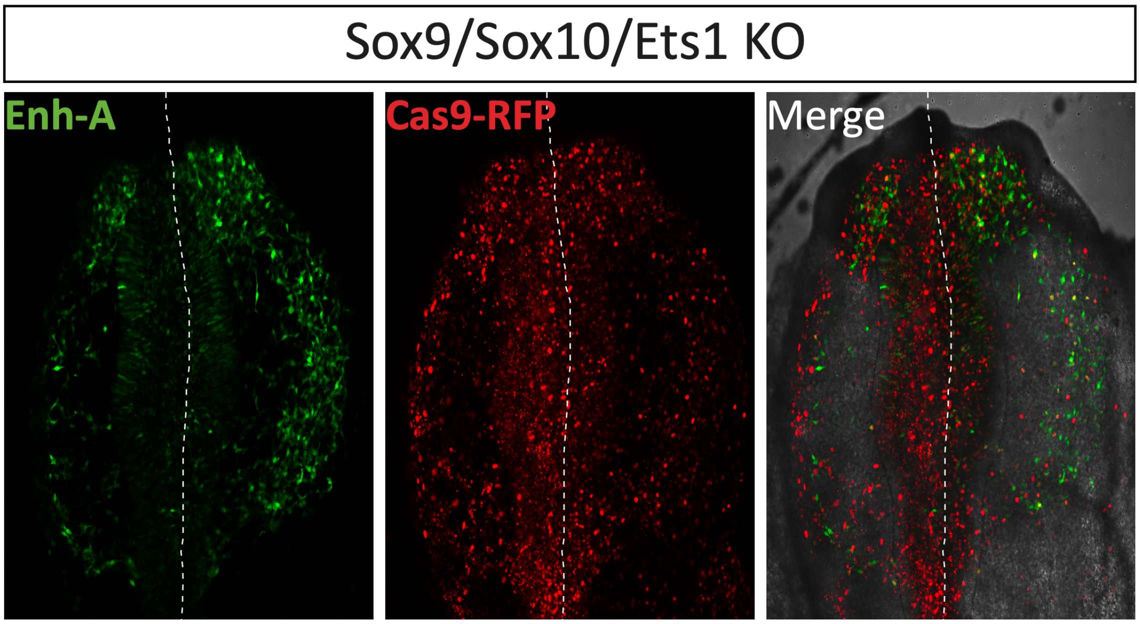

Figure 2.19 Functional perturbation of upstream transcription factors regulating *Chd7* enh-A disrupts enhancer activity *in vivo*. Embryos at gastrulation stages (HH4+) are electroporated on the experimental (left) side with sgRNAs targeting selected factors (Sox9, Sox10, Ets1). The control (right) side of the embryo is electroporated with scrambled sgRNAs. Both sides were electroporated with Cas9-RFP and enhancer reporter construct driving Citrine. Electroporation efficiency on both sides can be directly assessed by imaging Cas9-RFP. Embryos are incubated until the pre-migratory stages of NC development (HH8-10) and then analysed by confocal imaging. Bilateral electroporations were performed by prof. Tatjana Sauka-Spengler; confocal imaging was contributed by Prof. Tatjana Sauka-Spengler and Dr. Ruth Williams.

silico predictions, as enh-T activity was reduced on the experimental side (Figure 2.21).

Taken together, these results reveal a novel mechanism by which the chromatin remodeller *Chd7* is specifically regulated in the NC. We provide evidence of combinatorial regulation at the *cis* and *trans* levels, validate Sox10 binding, and provide functional validation to the combinatorial regulation of three key developmental enhancers, which are likely conserved in humans. Our results thus have important implications to clinical research associating Chd7 to human developmental abnormalities.

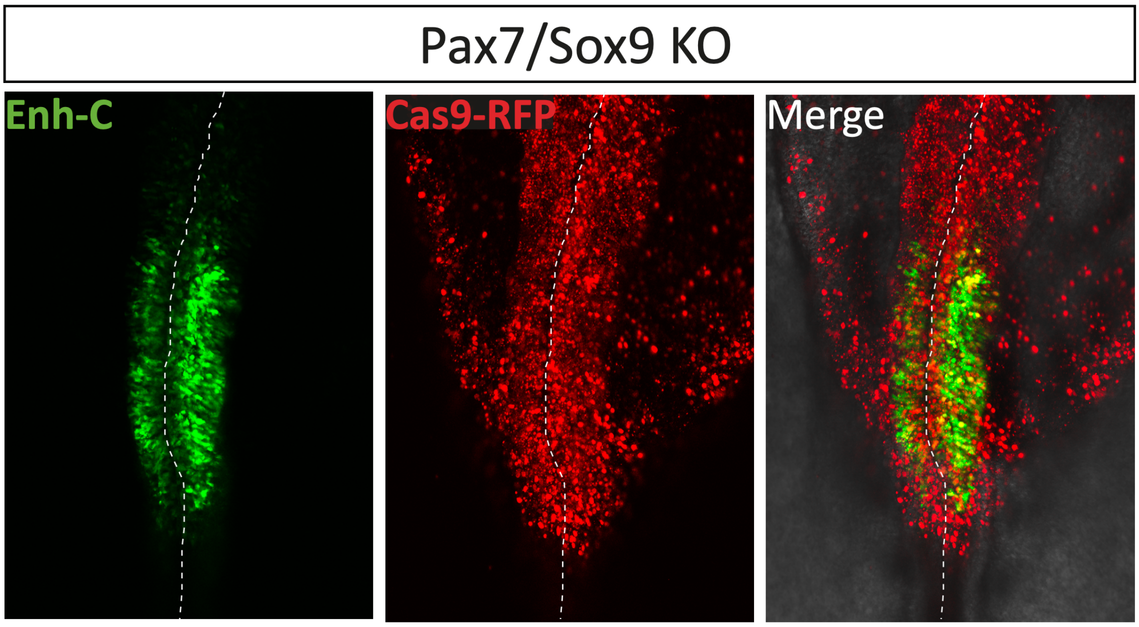

Figure 2.20 Functional perturbation of upstream transcription factors disrupts *Chd7* enh-C activity *in vivo*. Embryos at gastrulation stages (HH4+) are electroporated on the experimental (left) side with sgRNAs targeting selected factors (Pax7 and Sox9). The control (right) side of the embryo is electroporated with scrambled sgRNAs. Both sides were electroporated with Cas9-RFP and enhancer reporter construct driving Citrine. Electroporation efficiency on both sides can be directly assessed by imaging Cas9-RFP. Embryos are incubated until the pre-migratory stages of NC development (HH8-10) and then analysed by confocal imaging. Bilateral electroporations were performed by prof. Tatjana Sauka-Spengler; confocal imaging was contributed by Prof. Tatjana Sauka-Spengler and Dr. Ruth Williams.

2.4 *Cis*-regulatory architecture controlling early NC lineage decisions

2.4.1 NC heterogeneity at the epigenomic level

The observed *cis*-regulatory dynamics (Figures 2.5, 2.7 and 2.9) provides evidence of early heterogeneity at the epigenomic level, consistent with previous reports suggesting regulatory mechanisms that limit NC potential and thus determine NC lineage decisions, with the implications for the proposed multipotency of the NC [22, 63]. Consistently, GO term enrichment analysis of genes associated to each group of clustered elements revealed that pathways associated with canoni-

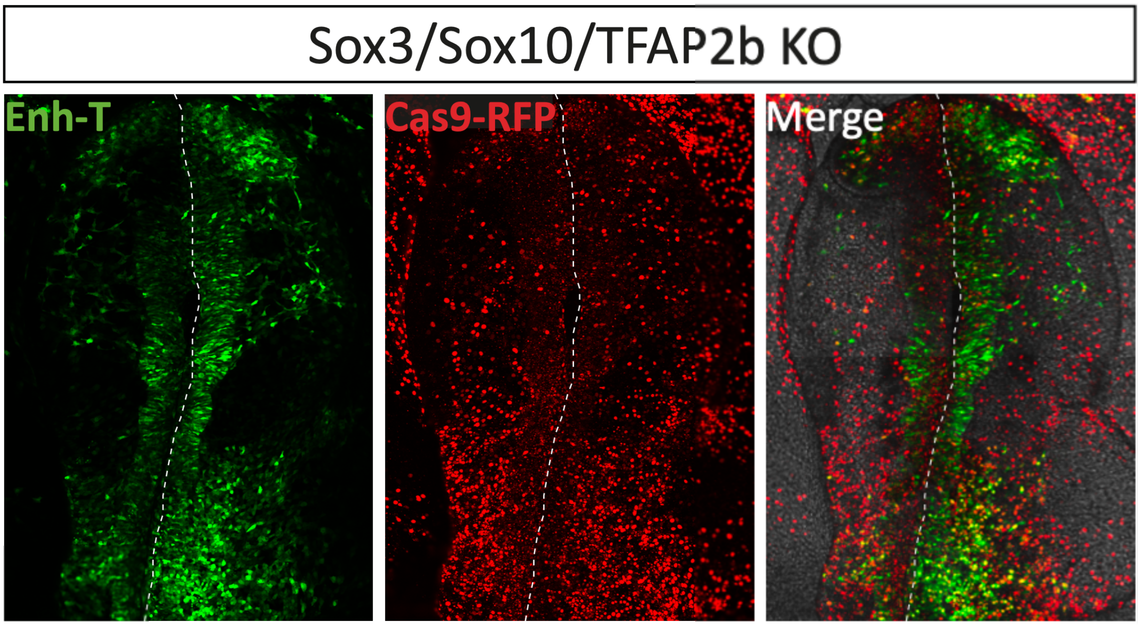

Figure 2.21 Functional perturbation of upstream transcription factors disrupts *Chd7* enhancer T activity *in vivo*. Embryos at gastrulation stages (HH4+) are electroporated on the experimental (left) side with sgRNAs targeting selected factors (Sox3, Sox10 and TFAP2b). The control (right) side of the embryo is electroporated with scrambled sgRNAs. Both sides were electroporated with Cas9-RFP and enhancer reporter construct driving Citrine. Electroporation efficiency on both sides can be directly assessed by imaging Cas9-RFP. Embryos are incubated until the pre-migratory stages of NC development (HH8-10) and then analysed by confocal imaging. Bilateral electroporations were performed by prof. Tatjana Sauka-Spengler; confocal imaging was contributed by Prof. Tatjana Sauka-Spengler and Dr. Ruth Williams.

cal NC development and differentiation programmes ($p<0.01$, Binomial test with Bonferroni correction) were exclusively associated with *k*-Cl-3 and DA elements (Figure 2.22a, red box). *k*-Cl-3 and DA enhancers only became accessible from pre-migratory NC stages (Figure 2.22), providing evidence that they mediated NC programmes. Conversely, programmes associated with neurogenesis are shared (and hierarchically regulated) by the successive action of *k*-Cl-6, *k*-Cl-1 and *k*-Cl-3 elements (Figure 2.22, green box), revealing an early regulatory split between ectomesenchymal progenitors and neurogenic progenitors at the *cis*-regulatory level. In addition, DA elements at 5-6ss suggested the existence of a stem cell-like niche within, premigratory NC, likely maintained at least until 8-10ss (Figure 2.22, red box).

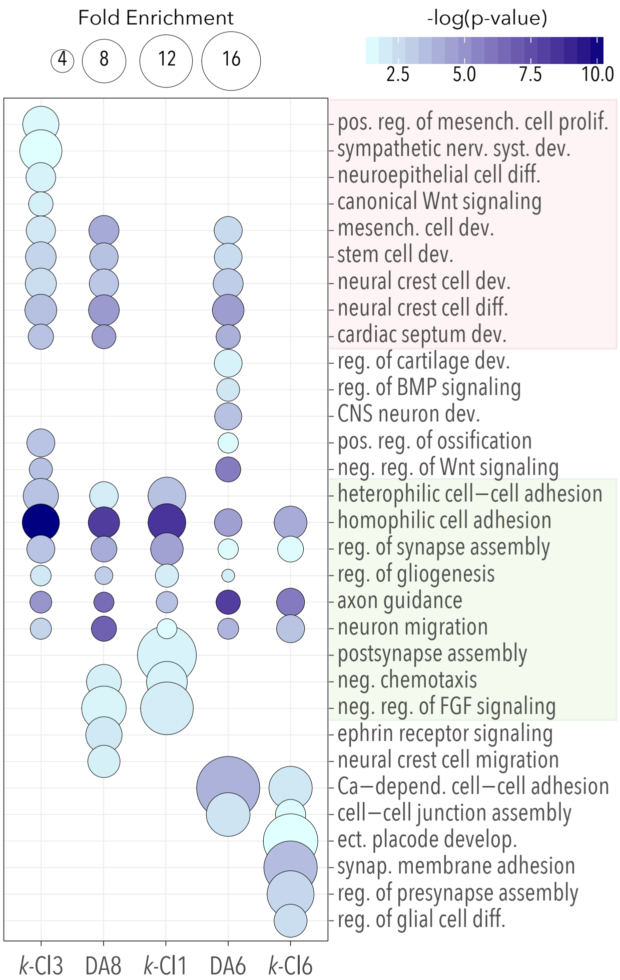

Figure 2.22 Gene ontology (GO) terms associated to *k*-means clusters and DA elements. Fold enrichments were obtained using statistical overrepresentation test, p-values calculated with binomial distributions and Bonferroni correction for multiple hypothesis testing. *K*-Cl-3 and DA elements exclusively reflect canonical NC and ectomesenchymal cell development, and specific NC differentiation programs (sympathetic nervous system), while sharing roles in homophilic cell adhesion, gliogenesis, and axonogenesis with k-Cl-6 and k-Cl-1 (**$p < 0.01$, Binomial test with Bonferroni correction, fold change >4). Only late-acting elements (*k*-Cl-3, *k*-Cl-1, and DA at 8-10ss) correlate to heterophilic cell-cell adhesion, while *k*-Cl-6 elements play a role in early regulation of neuronal NC lineages, ectodermal placode formation, and cell-cell adhesion.

In addition to sharing NC terms with *k*-Cl-3 and *k*-Cl-1, DA regulatory elements yielded terms indicative of the specific NC developmental processes taking place at stages at which the cells were characterised (early multipotent, premigratory NC at 5-6ss, and early migratory NC at 8-10ss). Analysis also highlighted the future state of NC cells relative to the sample stage. For instance, terms including cartilage development, regulation of ossification and CNS neuron development, more reflective of later NC behaviour, were over-represented at 5-6ss, whereas terms such as gliogenesis and neuron migration were enriched at 8-10ss, suggesting that the *cis*-regulatory landscapes are primed earlier than the onset of corresponding gene expression and observable phenotype. Interestingly, early-opening *k*-Cl-6 elements (Figure 2.22) function in the upstream regulation of neuronal differentiation processes (pre-synapse assembly, synaptic membrane adhesion and gliogenesis), suggesting that at the regulatory level, neurogenic NC cells begin to be primed very early in development. This was in line with the finding that all *k*-Cl-1 elements that associate with neural differentiation appear to be fully accessible by 5-6ss (Figure 2.7).

We next investigated the upstream TF codes driving the observed functional heterogeneity between enhancer groups. Firstly, using the Homer algorithm [233] we identified enriched *de novo* TF binding motifs within our clustered elements with high statistical significance ($p<10^{-12}$, Binomial test), (Figure 2.23). We detected *trans*-regulatory inputs encrypted in *cis*-elements, including binding signatures of early acting factors, canonical NC factors, and inputs specifically correlated to either neural or mesenchymal lineages (Figure 2.23). A canonical NC signature highly enriched in *k*-Cl-3 and late DA elements (Figure 2.23, red box), providing genome-wide evidence of the pivotal role of TFAP2s, Sox10 and Fox factors in the NC *cis*-regulatory landscapes, suggested a role for feed-forward loops in the global regulation of NC. Moreover, we also found factors suggestive of ectomesenchymal fates, such as cardiac specifier Pax3, myoblast regulator MyoD, and mesenchymal stem regulator NFAT; some of which are known to act downstream of Sox10, TFAP2 and FoxD3 [32, 33]. Amongst other *k*-Cl-3 motifs, we also found subgroups of exclusive inputs (e.g., Sox3) and those shared with *k*-Cl-6 (Tead1/2 and

Meis1/2) and *k*-Cl-1 (i.e., ATF, Maf, retinoic acid receptors, Zic, Zfp and GATA), possibly reflecting the role for some of these factors in stem cell-like and neural lineages, respectively.

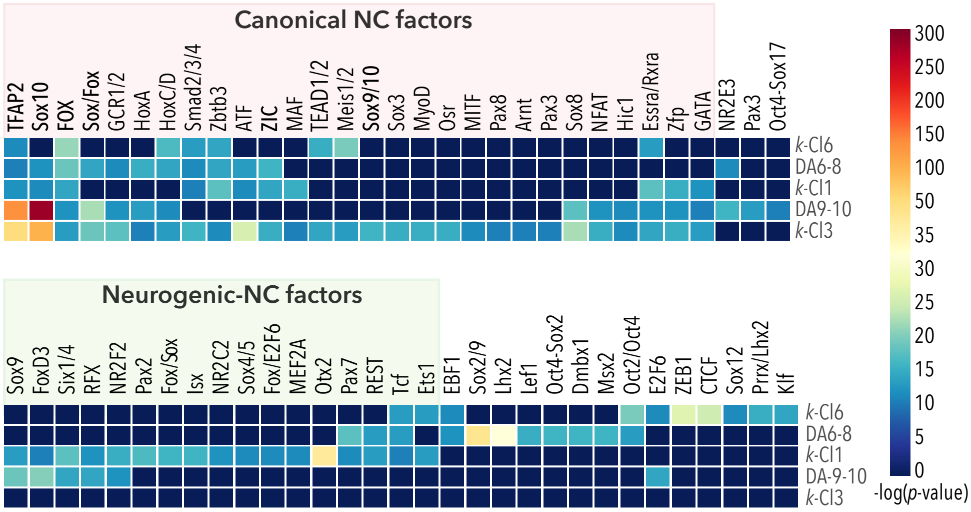

Figure 2.23 *De novo* TF binding motifs enriched in *k*-means clusters and DA elements. *De novo* motifs were identified using Homer. Binominal p-testing was used to determine motifs with p < 1 x10e^{-11}.

The *k*-Cl-1 motif signature (Figure 2.23, green box) was consistent with the putative role of these elements in neural lineages. For instance, high enrichment in Sox9 and FoxD3 motifs is in line with the postulated role for these factors in driving the switch from neural to melanocytic lineages in early trunk NC [63]. Intriguingly, we also found significant inputs from Otx2 and Pax2, factors known to play canonical roles in the early neural specification and neural differentiation [234], respectively, whose placement within the NC-GRN has been long hypothesised [235]. Moreover, NC nuclear receptors NR2C2 and NR2F2 [50] were also found to drive *k*-Cl-1 elements, as well as SoxC/SoxD factors (Sox4/5), consistent with the role of *k*-Cl-1 in the early onset of NC programmes. We also uncovered a *k*-Cl-6 specific niche of upstream inputs including self-renewal drivers Oct2/4 and Klf, EMT regulator Zeb1. Finally, a signature of pre-migratory NC associated TFs (e.g., Sox2/9, Msx2, Lef1/Tcf and Pax7) was associated with DA elements at 5-6ss, whereas DA elements at 8-10ss showed the highest enrichment of factors associated with migratory NC (Sox10, Sox8, TFAP2s) (Figure 2.23), indicating the role of a Sox8/10-TFAP2-dependent feed-forward loop on the global regulation of migratory NC. Remarkably, these results allow us to conclude that our DA elements indeed capture the temporal dynamics of pre- and migratory NC, whereas *k*-Cl-3, *k*-Cl-6 and *k*-Cl-1 elements revealed a regulatory interconnection of canonical and non-canonical NC programmes to different cell-types and lineages, including naïve stem cells, neural and ectomesenchymal progenitors. Moreover, enrichment of binding sites for factors not expressed at stages or cell-types studied here (such as anterior Hox genes and nuclear receptors responsive to retinoic acid (RA) signalling), suggest that these enhancers may be involved in the later patterning of the NC through response to external cues. Such diversity of canonical and novel, lesser characterised TF inputs reflects the broad potential of NC cells and suggests further regulatory refinement is required to direct NC fates.

Interestingly, although *k*-Cluster motifs were largely associated to their postulated role as promoters, we also uncovered a *de novo* motif containing a combination of canonical FKH binding sites with the PRC2-associated TF, E2F6 (Figure 2.24). Consistent with previous findings from our laboratory [60], this finding suggest some *k*-cluster 4 elements might function as *cis*-regulatory repressors. Such repressive sites might be responsive to FoxD3 or other Fox TFs, which might co-bind with E2F6 and recruit PRC1/PRC2 in a cell type-specific manner, and induce transcription repression of developmental loci.

2.4.2 Resolving NC heterogeneity at single-cell resolution

We next inquired whether this heterogeneous regulatory logic was reflected at the transcriptional level. To this end, we performed single-cell RNA-seq using a Smart-Seq2 strategy that resulted in 137 deep NC single cell transcriptomes within the

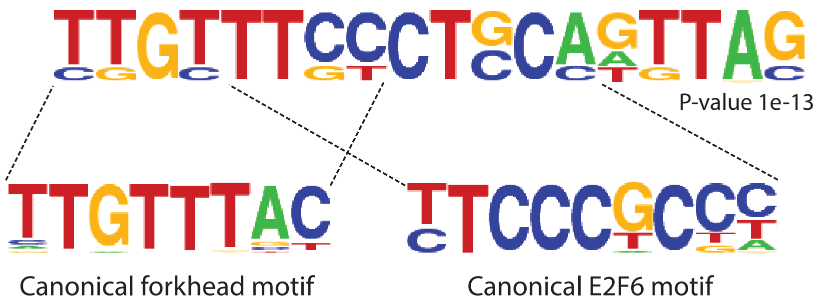

Figure 2.24 A *de novo* Fox/E2F6 motif suggests a subset of k-Cluster 4 elements might function as *cis*-regulatory repressors.

narrow dynamic window (6-7ss) that captured the transcriptional dynamics of NC before and during early delamination (Figure 2.25). Principle component analysis (PCA) revealed three distinct populations: Sc-cluster-1, -2 and -3 (Figure 2.25). Sc-cluster-1 and -3 presented clear NC identity, expressing hallmark NC genes (Pax7, Sox9, Snai2, Sox10, TFAP2A/B, Ets1), but segregated according to developmental stage and were uniquely characterised by a number of early-stage specific within Sc-cluster-1 (Brn3, Alx1, Cldn1 and Sfrp2,) and later-stage specific, ectomesenchymal-related NC genes within Sc-cluster-3 (Col9A3, Elk3, Ltk, ItgA9 and Cxcr4). In contrast, Sc-cluster-2 contained cells collected at both 6ss and 7ss that in addition to sharing some of the NC factors with Sc-cluster-1 and -3, also uniquely expressed genes indicative of a neural phenotype, such as Sox3 and Cdh2, including a number of key neural genes exclusively driven only by *k*-Cl-1 elements (Vimentin, Nbeal1, Pax6, etc) or in concert with *k*-Cl-6 (Otx2, Gli2, MylK, Nlgn1, and retinoic acid signalling players Nav2, Cyp1b). Interestingly, while initiating neural differentiation programmes, Sc-cluster-2 cells also expressed a number of NPB specification factors (Sall1, Zic1, Msx2, Meis1/2, Frzb, etc.) involved in the maintenance of NC programmes, that were mainly driven by *k*-Cl-6 elements. Strikingly, Sc-cluster-2 cells did not express majority of canonical, ectomesenchymal NC markers (Figure 2.25). In summary, single-cell transcriptional analyses strongly support an early split of the developing NC into neurogenic and multipotent NC progenitors.

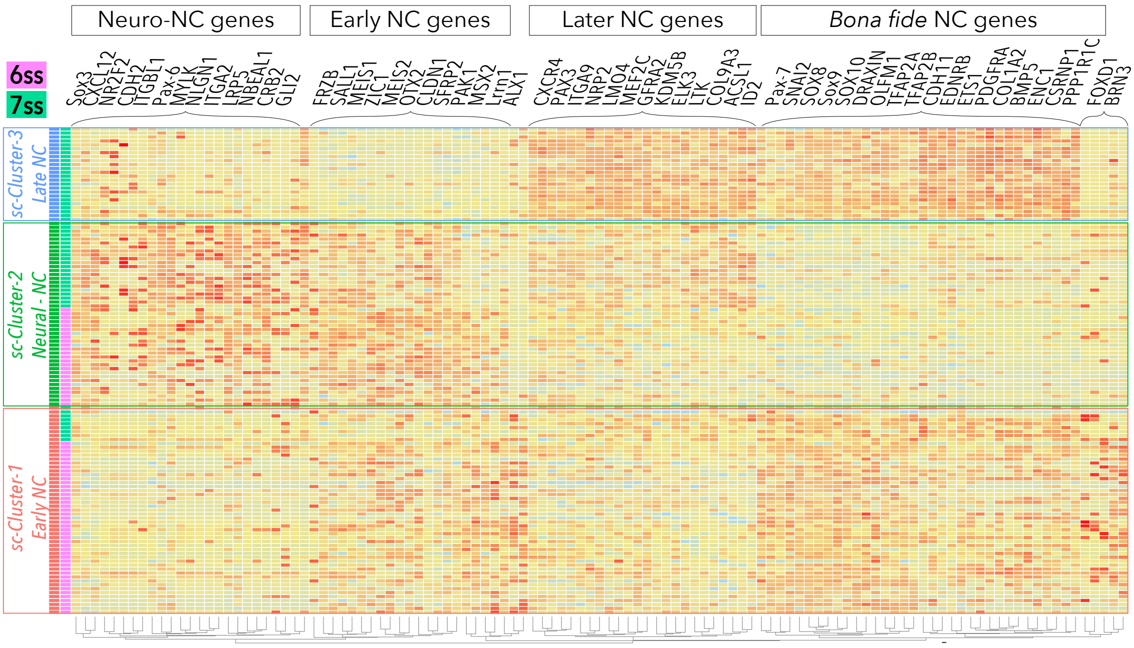

Figure 2.25 scRNA-seq heatmap visualizing hierarchical clustering of single NC cells at 6-7ss (total 124 cells, 74 6ss, and 63 7ss), sc-Cluster-1 (early NC, in red), sc-Cluster-3 (late NC, in blue), sc-Cluster-2 (neuronal-NC, in green). Top 50 differentially expressed genes are shown. PAGODA analysis was performed by Dr. Emmanouela Repapi.

We next determined single-cell clusters statistical linkage to the uncovered epigenomic k-clusters by hypergeometric P-tests. Pre-migratory NC Sc-cluster-1 only showed stastically significant links to the k-Cl-3 elements (*p=0.02), whereas NC sc-cluster-3 is significantly associated to both canonical k-Cl-3 and the early neurogenic k-Cl-1 elements (****p=2.23x10^{-6} and **p=0.003, respectively; two-tailed hypergeometric P-test), (Figure 2.26. Strikingly, the singular neurogenic NC cluster Sc-cluster-2 seems to be controlled only by early k-Cl-6 and k-Cl-1 enhancers (*p=0.018 and ***p=4.0x10^{-4}, respectively), (Figure 2.26). Such a strong correlation between k-Cluster enhancers and single-cell gene expression clusters indicate the role of early *cis*-regulatory heterogeneity in establishing diverse NC progenitor subpopulations.

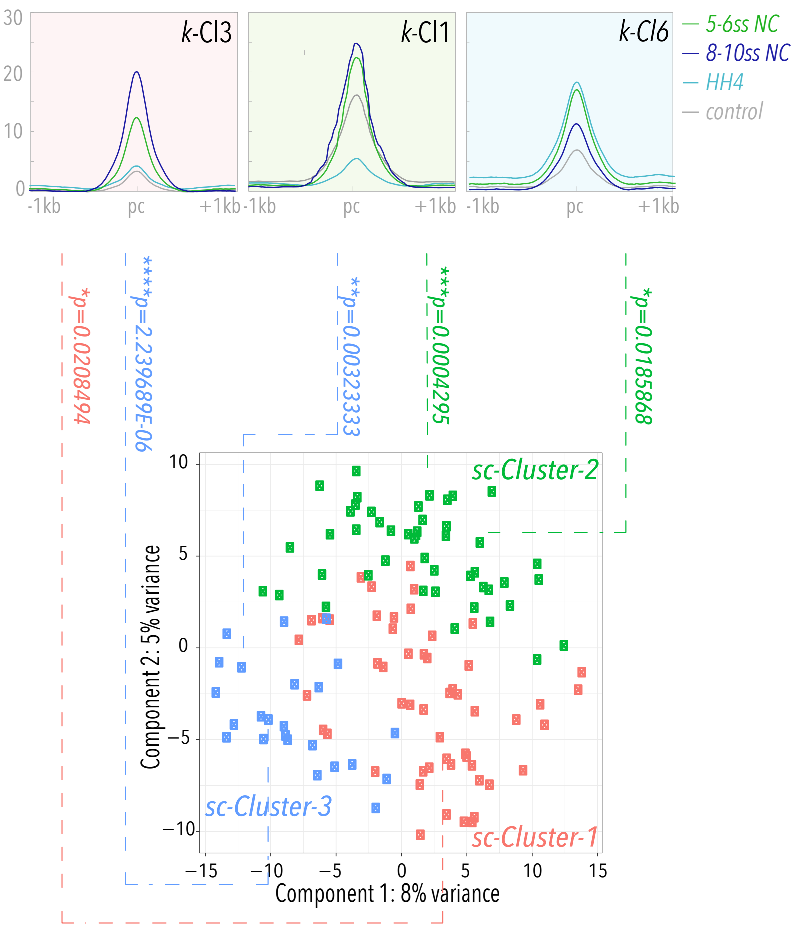

Figure 2.26 PCA of top 100 genes from scRNA-seq and their statistical association to *k*-means clusters. P-values reflecting statistical significance of the single cell RNA-seq and *k*-means cluster associations were calculated using two-tailed hypergeometric P-tests.

2.5 A combinatorial *cis*-regulatory code for NC fate restriction

The establishment of developmental regulatory programmes requires combinatorial TF activity [142, 236]. We therefore examined putative TF co-binding patterns at NC enhancers as a means of identifying the core transcriptional networks underlying these distinctive identities. Screening all 2-way combinations of enriched *de novo* TF motifs in our enhancer clusters allowed us to detect a cohort of significantly enriched co-occurring TFs (*$p<0.05$, two-sided Chi-squared test with Bonferroni correction), (Figures 2.27, 2.28, 2.29).

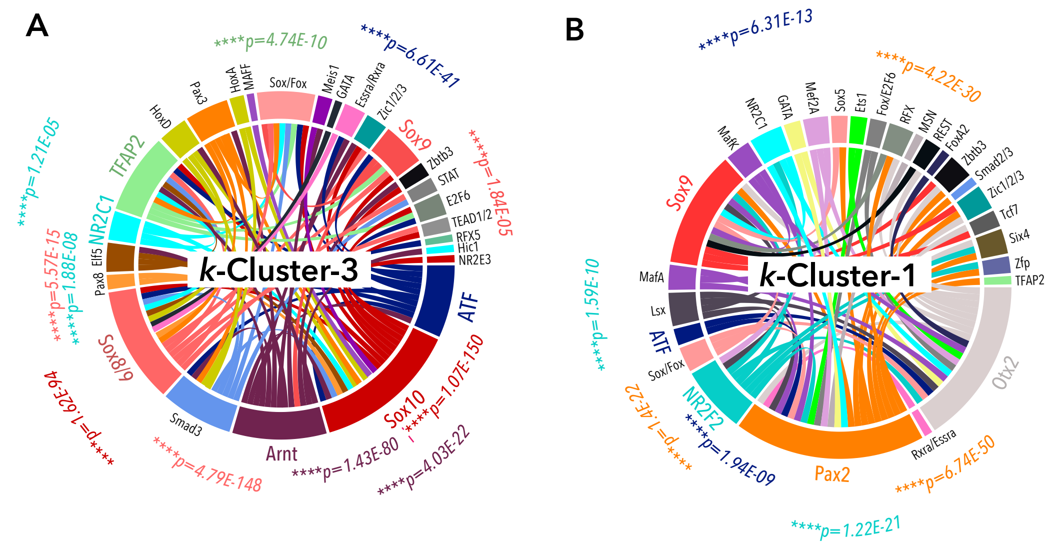

Figure 2.27 TF co-binding relationships predicted for k-Cl-3 (A) and k-Cl-1 (B) elements. P-values were calculated using two-tailed Chi-squared test, with Bonferroni correction for multiple hypothesis.

In *k*-Cl-3, we found a striking enrichment of canonical interactions [32] specific to NC (Figures 2.27a, 2.29), such as Sox10-TFAP2 (p=1.6x10^{-94}), Sox10-Sox9 (p=5.8x10-13), Sox10-Sox8 (p=4.8x10-148), Sox10-Sox/Fox (p=7.7x10^{-85}), as well as novel ones, such as ATF-Sox10 (p=1.07x10^{-150}) and Arnt/Arnt2-Sox10 (p=1.4x10^{-80}), suggesting a dynamic interplay of TF combinations in the regulation of NC-specific identity. Moreover, analysis of DA at 5-6ss and 8-10ss enabled the further identification of temporally restricted, NC-specific combinations controlling pre-migratory (e.g. Lhx-Sox2, Fox-Sox2, Dmbx1-Lhx, p $<$ 2.7x10^{-28}), (Figure 2.28a) and early migrating (i.e., Sox10-TFAP2s, Smad3-Sox/Fox, Sox10-Smad3, TFAP2-Sox/Fox, p$<$ 2.3x10^{-32}), (Figures 2.28B, 2.29) identities, respectively [32]. *k*-Cl-1, however, showed an enrichment in non-canonical NC TF interactions (Figures 2.27b, 2.29), such as co-binding of Otx2-Pax2 (p=6.7x10^{-50}), Otx2-Sox9 (p=4.2x10^{-30}), Pax2-Sox9 (p=8.5x10^{-24}), NR2F2-Otx2 (p=1.4x10^{-22}), Otx2-Zic1/2/3 (p= 2.9x10^{-11}) and NR2F2-Sox9 (p=1.6x10^{-10}) amongst others, highlighting known and integrating the previously unknown core transcriptional circuitry underlying NC-derived neural progenitors into our current view of the NC-GRN [32, 33].Importantly, these results place Otx2 and Pax2 alongside the canonical NC regulators Sox9, Zic1-3, Smad2/3 and NR2F2 [32, 50], as putative core regulators of the NC-GRN, displaying previously unappreciated connectivity to downstream, neural effector genes.

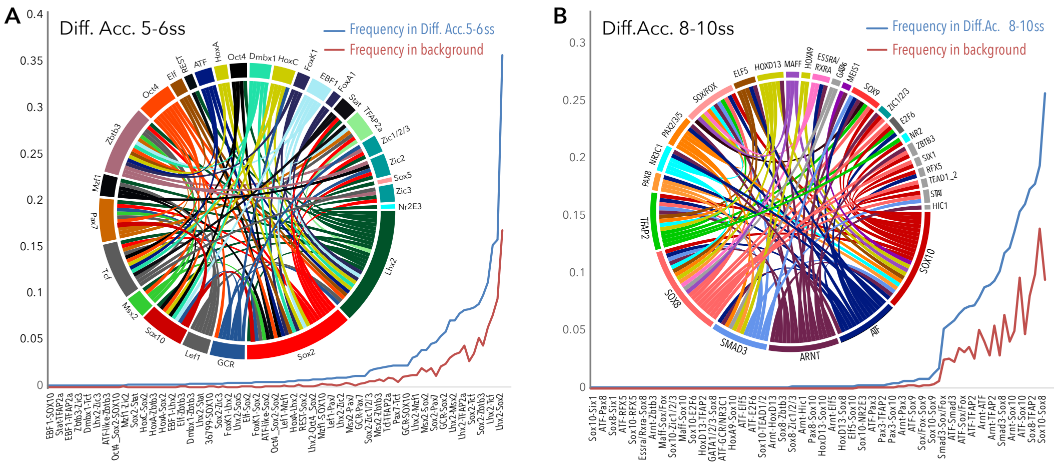

Figure 2.28 TF co-binding relationships predicted for DA elements at 5-6ss (A) and 8-10ss (B). P-values were calculated using two-tailed Chi-squared test, with Bonferroni correction for multiple hypothesis. Motif co-occurrences are ranked below, with the most frequent combinations to the right.

To functionally probe the core transcriptional circuits governing NC heterogeneous identities (*i.e.*, ectomesenchymal versus neurogenic NC), we used a CRISPR-Cas9 toolkit [232] to endogenously knockout combinations of upstream co-acting factors. Using *in vivo* bilateral co-electroporation assay, we delivered the target gRNAs with wildtype Cas9 to the left and the control gRNA with Cas9 to the right side of the embryo and performed RNA-seq on left and right dissected dorsal neural tubes at 7ss, to assess the global effect of the perturbation on the NC-GRN (Figure 2.30). This approach represents an excellent strategy to test the contribution of these factors to the proposed gene-regulatory interactions, as each side of the same embryo can be used for the control and experimental groups, thus allowing us to rule out variations that would have been otherwise caused by

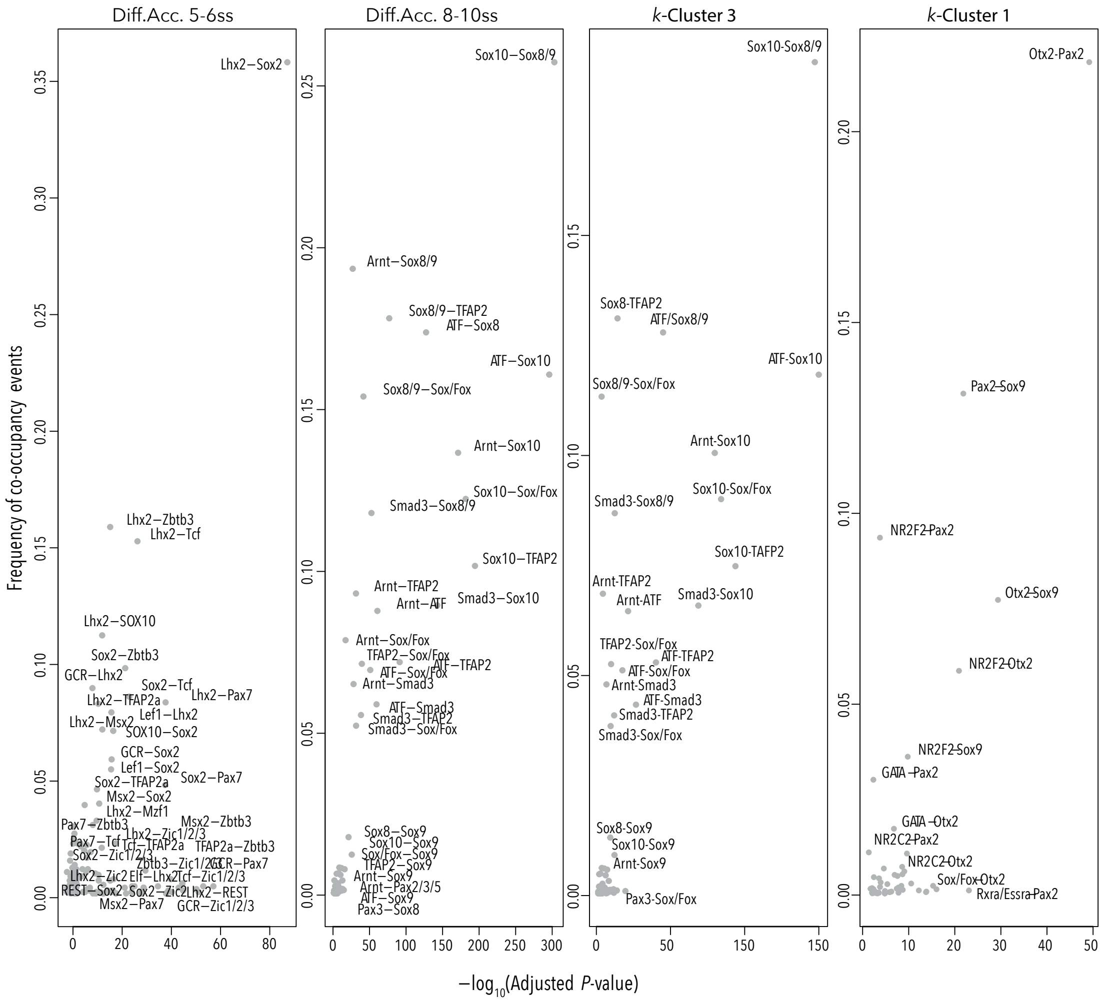

Figure 2.29 Scatter plot showing the frequency of TF co-occupancy events and statistical significance in DA elements at 5-6ss and 8-10ss, and *k*-Cluster-3 and *k*-Cluster-1. P-values were calculated using two-tailed Chi-squared test, with Bonferroni correction for multiple hypothesis.

embryo-to-embryo differences, such as genetics, developmental heterochrony often seen between embryos, batch effects, among others.

On the one hand, we targeted the factors constituting the *k*-Cl-3 minimal transcriptional core (Sox10, Sox9, TFAP2B, ATF2, Arnt2 and NR2C1), which we have proposed to underlie NC canonical/mesenchymal progenitor identities (Figures 2.27a, 2.29). This strategy caused specific downregulation of 241 and upregulation of 62 genes (FDR<0.1, Log2Foldchange>1). Downregulated genes included key

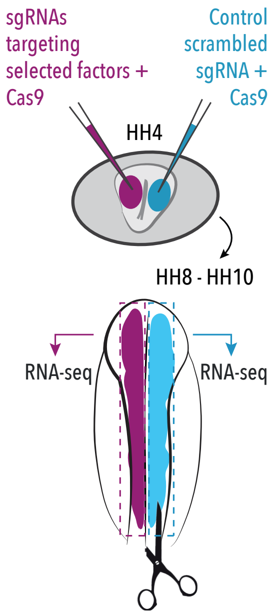

Figure 2.30 Schematic of bilateral electroporation assay for CRISPR-mediated GRN perturbation experiment. Embryos at gastrulation stages (HH4+) are injected on the experimental side with sgRNAs targeting selected factors. The control side of the embryo is injected with scrambled sgRNAs. Both sides were electroporated with Cas9. Embryos are incubated until the pre-migratory stages of NC development (HH8-10), and dorsal neural tubes were dissected, separating the control side from the experimental one. These samples were then analysed by RNA-seq. Bilateral electroporations and RNA-seq were performed by Dr. Ruth Williams.

premigratory NC regulators such as Msx1 and Draxin, suggesting these genes are directly co-regulated by Sox10, Sox9 and TFAP2b. These factors also display positive cross-regulation between them, further reinforcing the role of feedback loops in this transcriptional network. Furthermore, downregulation of a number of downstream effector genes (PlxnA4, Pdgfra, Cxcr7, Col9A1/A3, Pcdh18 and Cdh13), (Figure 2.31b), some of which were previously described as functional components of NC programme with roles in cell migration, indicated a general disruption of effector gene batteries. Conversely, the upregulated genes were involved in neural programmes (FoxG1, Hesx1, Lhx2/9, Six3/6, ALDH1A1), consistent with the GO terms associated to them ($p<0.05$, Binomial test with Bonferroni correction), (Figure 2.32a).

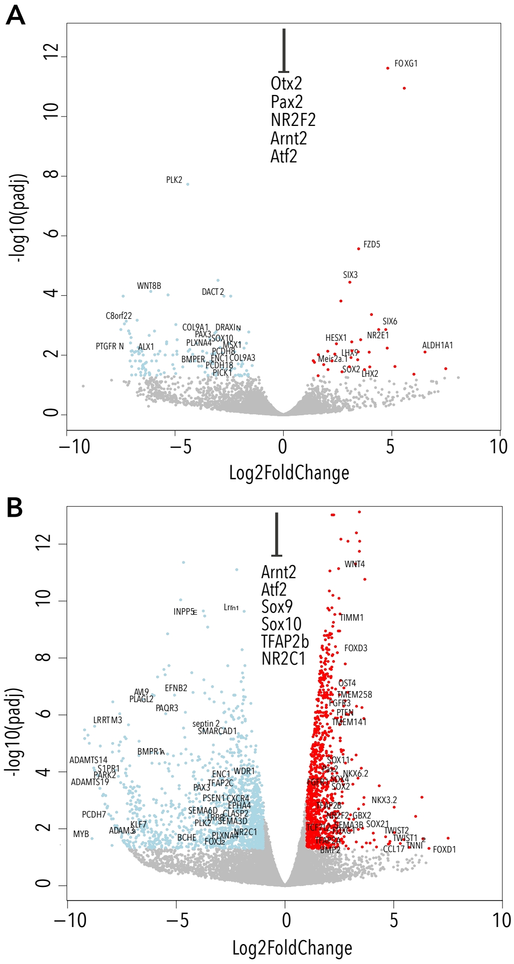

Figure 2.31 Volcano plots showing mis-regulated genes following CRISPR/Cas9-mediated knockout of core TFs associated with *k*-Cl-3 (A) and *k*-Cl-1 (B), respectively. Differential expression was analysed by DEseq2, comparing gene expression from three biological replicates, bilaterally electroporated on the left (experimental) and right (control) sides with a CRISPR/Cas9 toolkit. Dissected dorsal neural tubes from each side were individually subjected to RNA-seq and used as biological replicates in each experiment. Factors highlighted in (a) represent the knocked-down k-Cl-1 core regulators: Otx2, Pax2, NR2F2, Arnt2 and Atf2; in (b), the highlighted k-Cl-3 core regulators Sox9, Sox10, NR2C1, TFAP2b and Atf2 were knocked-down. Red, up-regulated genes; blue, down-regulated genes.

In a separate experiment, we sought to perturb the factors constituting the *k*-Cl-1 minimal transcriptional core (Pax2, Otx2, NR2F2, ATF2, and Arnt2), (Figure 2.31a), which we proposed to play a critical role within the core transcriptional network driving the establishment of a neurogenic NC identity (Figures 2.27, 2.29). GO term analysis of these genes (*p<0.05, Binomial test with Bonferroni correction) revealed a complex perturbation of a large number of general processes, including catabolism and glycerolipid metabolism, but also downregulation of neural genes (BCHE, BMPR1a, Lrrn1, PCDH7, Sema6, Cxcr4, PTRM, RARb, TFAP2C, Pax3) involved in cell migration, neurogenesis, cell adhesion and neural differentiation (Figures 2.32b). Interestingly, key TFs and nuclear receptors governing ectomesenchymal programmes were modestly but significantly upregulated (TFAP2A/B, Twist1/2, FoxD1/3, Wnt4, FGFR1/3, Bmp2), alongside other neural-related pathways dependent on NKX6.2, Sox2/11/21, PTEN and Sema3b, known to govern neural differentiation (Figures, 2.31a, 2.32b).

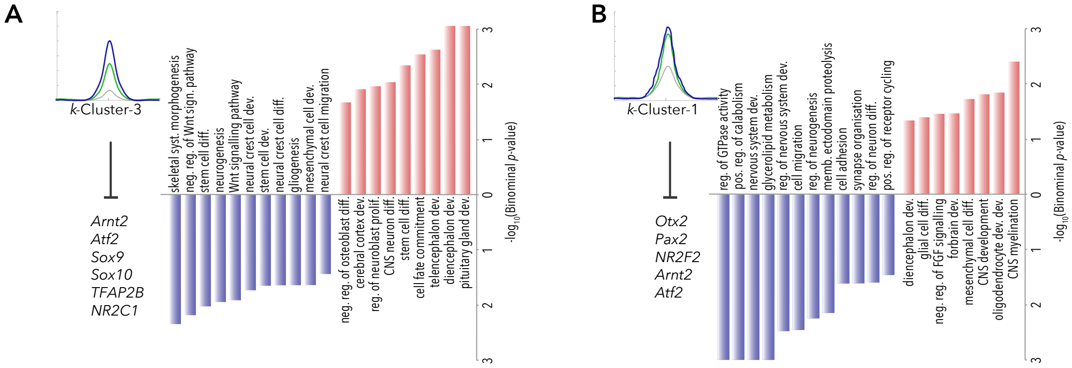

Figure 2.32 GO terms associated with mis-regulated genes following knockout of indicated TFs associated with *k*-Cl-3 (A) and *k*-Cl-1 (B). Enriched GO terms were obtained using statistical overrepresentation test, p-values calculated with binomial distributions and Bonferroni correction for multiple hypothesis testing.

Our results indicated that while perturbing the core transcriptional network of *k*-Cl-3 caused the direct suppression of NC-specific ectomesenchymal progenitor programmes with concurrent upregulation of neural signature genes, the disruption of *k*-Cl-1 regulatory core appeared to have an opposite effect on the global transcriptional landscape. Thus, our knockout strategy caused a dysregulation of the NC-specific networks balancing the shift of neural-to-ectomesenchymal progenitor states, suggesting that the uncovered NC transcriptional cores, mediated by *k*-Cl-3 and *k*-Cl-1 regulatory elements balanced the shift between NC-derived

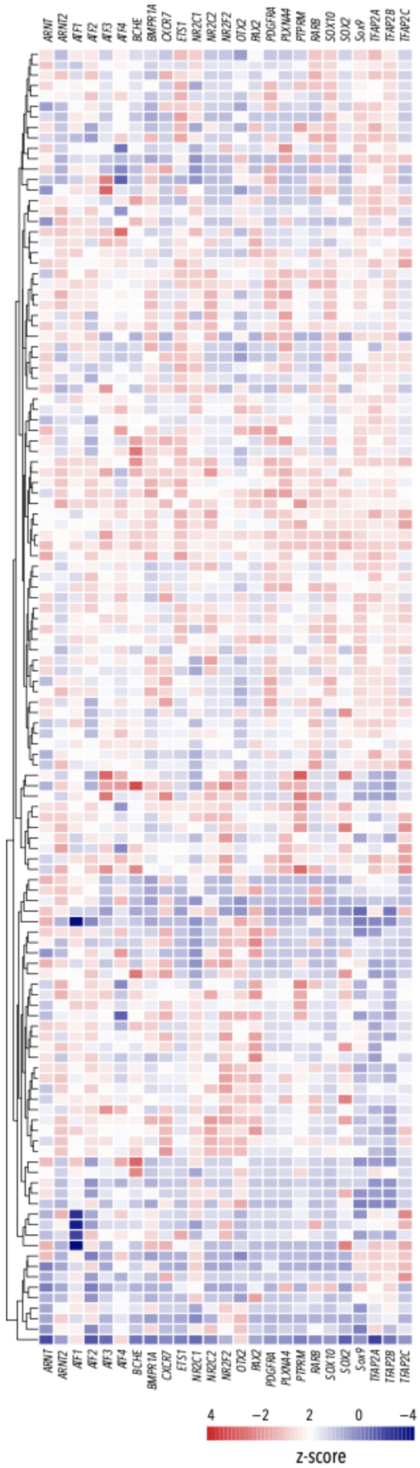

Figure 2.33 Heatmap showing single-cell co- expression of targeted TFs and selected genes, which were dysregulated following CRISPR/Cas9-mediated knockout of core TFs.

ectomesenchymal and neural lineages, likely through mutual cross-negative regulation.

2.6 Integrating regulatory information into GRN circuitry

State-of-the-art approaches to reverse engineer global vertebrate GRNs have either used the empirical reconstruction of GRN hierarchies by screening motif binding sites around promoters, together with high-resolution single cell profiles [237], or mathematical modelling of best-fit topologies of epistatic interactions and developmental trajectories determined at the single cell transcriptomic level [238]. I have build on these strategies by benefiting from unbiased, high signal-to-noise, low input epigenomes and high-resolution single-cell transcriptomes of NC cells. I unravelled the co-binding dynamics of TFs directly regulating the NC programmes, by performing a comprehensive screen to identify all known high-resolution vertebrate motifs ($p<0.0001$, Binomial test) present in our enhancers (*k*-Cl-3, *k*-Cl-1, and DA elements) as a means of identifying the full ensemble of TF inputs. I further filtered motif occurrences by integrating TF co-expression data from the single-cell profiles (Appendix 2.25, 7.5, ShinyApp). I next produced comprehensive, genome-wide reconstructions of the *k*-Cl-3 and *k*-Cl-1-dependent GRNs (Appendix figures 7.8, 7.9, respectively), allowing us to identify the hierarchical position of *k*-Cl-3 (Sox10, Sox9, TFAP2B, ATF2, Arnt2 and NR2C1) and *k*-Cl-1 (Otx2, Pax2, NR2F2, Atf2 and Arn2) minimal core targeted factors within their respective NC-GRNs, as well to predict their downstream targets. Genome-wide reconstructions were also performed using DA *cis*-regulatory elements at the pre-migratory stage (Appendix 7.6) and migratory stage (Appendix 7.7).

The extracted GRN circuits that control representative genes downregulated upon *k*-Cl-3 TF core knockout (Figure 2.34a), feature contributions from multiple enhancers within these loci varying between six (*PlxnA4*) to two (*Pdgfra*) per gene

(Figure 2.34c), further highlighting the importance of enhancer redundancy for the robustness of transcriptional networks. These enhancers primarily mediate heterotypic binding of the core TFs, TFAP2B and Sox9/10 as well as additional inputs from other co-binding partners, including Arnt2, NR2C2, MafF, retinoid acid receptors, Zeb1 and Zic3. Some of these inputs form the *k*-Cl-3 core circuits (Figure 2.27, 2.29) and are themselves regulated by other targeted factors (NR2C1 and Arnt2, for instance) again highlighting feed-forward loops as core features of the NC-GRN (Figure 7.8). Moreover, selected genes (*PlxnA4*, *Msx1*, *PdgfrA*) affected by the knockout of *k*-Cl-3 TFs were assessed by *in situ* hybridization chain reaction (HCR) along with an unaffacted gene (*FoxD3*), which was used as an internal control, thus validating our RNA-seq results (Figure 2.34b).

The extracted GRN circuits controlling representative genes downregulated (Figure 2.35a) upon *k*-Cl-1 TF core knockout identify the direct inputs of disrupted TFs, such as Otx2-Lhx2 and Smad3-Pax2, into BCHE, BMPER1 and TFAP2C, and Otx2-Zic1/3 into RXRb, with additional inputs (NR2F2, Arnt2, Meis1/2, Sox8-10) contributing to individual enhancers (Figure 2.35c). Again, we selected genes affected by the knockout (*PlxnA4*, *Cxcr4*, *Bmpr1A*) and assessed their disruption by *in situ* HCR (Figure 2.35b) along with the same unaffacted gene (*FoxD3*), which was used as an internal control, thus validating our RNA-seq results. Taken together, these results suggested that the core transcriptional network captured by *k*-Cl-1 elements acts upstream of RA signalling, through Otx2-dependent activation of RXRb, which may dynamically control the response of these cells to gradients of RA, a critical process in the patterning of neuronal lineages [238]. Taken together, our results corroborate the idea that *k*-Cl-1 elements orchestrate the early establishment of the early neural progenitor state shared between NC and non-NC cells.

Generating sub-circuits governing a small number of genes allows representation

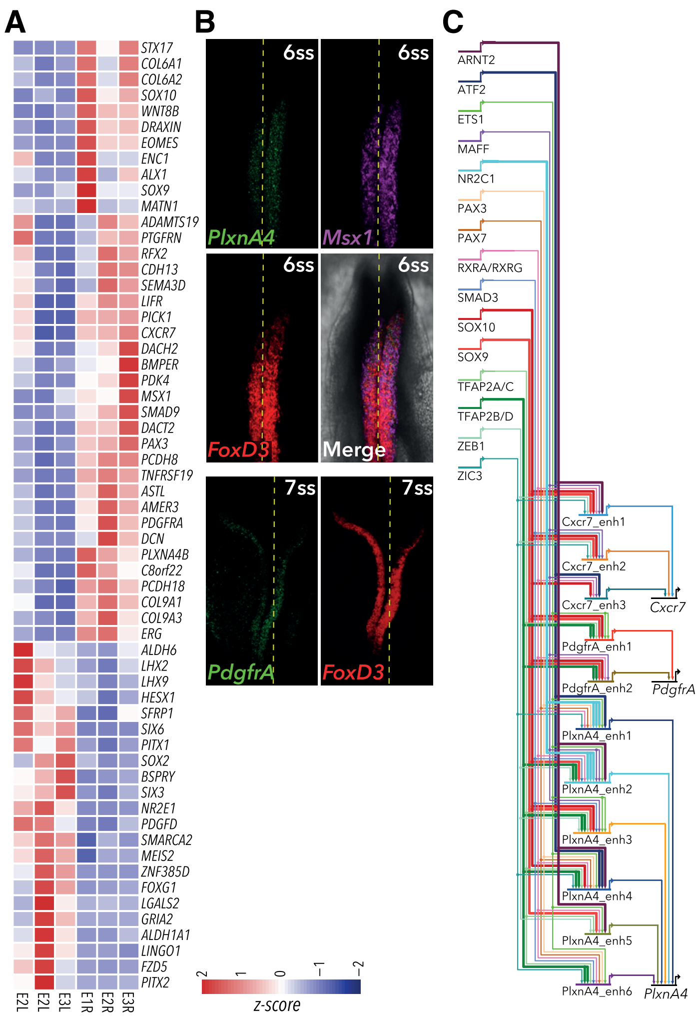

Figure 2.34 Heatmap comparing gene expression in left (experimental) and right (control) dissected dorsal neural tubes from three representative embryos where k-Cl-3 core TFs were targeted. (a) Differential expression was determined using DESeq2 with a negative binomial model, p-values calculated using Wald test, corrected for multiple testing using the Benjamin-Hochberg method (padj<0.1, Log2Foldchange >1), (b) Expression patterns of perturbed genes (*PlxA4*, *Msx1*, *PdgfrA*) following knockout of k-Cl-3 core TFs on the left side of the embryo, obtained by HCR. Control side, right. FoxD3 was used as non-pertubated, upstream factor which does not depend on the regulation of the targeted factors. (c) Sub-circuitry of gene regulatory interactions governing Cxcr7, PlxnA4, and PdgfrA, all downregulated in perturbation of *k*-Cl-3 core. HCR and confocal images were contributed by Dr. Ruth Williams.

of regulatory dynamics and demonstrates that patterns of co-regulation can be resolved for functionally related genes. Moreover, combinatorial patterns of TF inputs suggest the existence of regulatory modules within the broader NC-GRN, possibly related to the previously identified co-expression gene networks (Figure

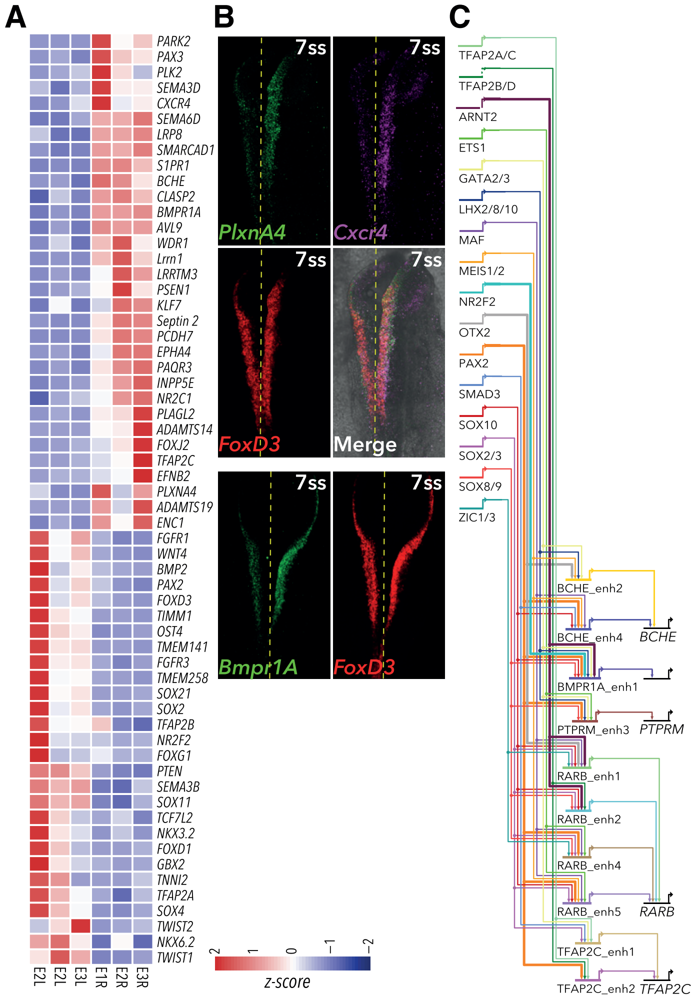

Figure 2.35 Heatmap comparing gene expression in left (experimental) and right (control) dissected dorsal neural tubes from three representative embryos where k-Cl-1 core TFs were targeted. (a) Differential expression was determined using DESeq2 with a negative binomial model, p-values calculated using Wald test, corrected for multiple testing using the Benjamin-Hochberg method (padj<0.1, Log2Foldchange >1), (b) Expression patterns of perturbed genes (*PlxnA4*, *Cxcr4*, *Bmpr1a*) following knockout of *k*-Cl-1 core TFs on the left side of the embryo, obtained by HCR. Again, *FoxD3* was used as a internal control, displaying no pertubation in the left (experimental) versus right (control) sides of the embryos. (c) Sub-circuitry of gene regulatory interactions governing Bche, Bmpr1, Ptprm, RarB, and TFAP2C, all downregulated in perturbation of *k*-Cl-1 core. HCR and confocal images were contributed by Dr. Ruth Williams.

2.4). As such circuitry requires cellular co-expression of all molecules involved, we used our single-cell profiles to assess co-expression of upstream TF inputs and downstream genes in order to validate our networks further (Figure 2.33). This approach validated novel regulatory relationships and elucidated the core components of the NC-GRN, highlighting the importance of enhancer regulation by combinatorial TF binding for normal NC development.

2.7 Database resource

Finally, we have assembled our data into a user-friendly, interactive Shiny App (Figure 7.5). This provides a searchable interface whereby individual or multiple genes can be queried for (1) gene expression; (2) WGCNA cluster association; (3) regulatory elements and retrieval of upstream regulatory inputs; (4) k-Cluster dependent transcriptional networks visualisation; and (5) visualisation of single-cell co-expression profiles. This unique tool offers valuable resource for the NC and wider GRN fields, as it allows exploration of the wealth of novel regulatory sequences and interactions reported here. We expect our resource to contribute to end-to-end investigations of the full complement of NC genes. In-depth analysis of the NC-GRN circuits of interest that are easily retrievable from our databases is poised to provide critical insight into precise gene regulatory events underlying NC ontogeny.

2.8 Discussion

Gene regulatory networks control key regulatory events that ensure accurate development. Over many years of investigation, a wealth of knowledge about NC formation has been accumulated through bottom-up, candidate-gene approach studies, providing essential insights into developmental processes governing NC ontogeny [32, 33, 57, 235]. We provided a top-down, unbiased approach to interrogate NC gene regulatory mechanisms genome-wide, allowing us to reverse engineer a comprehensive representation of the NC-GRN, encompassing a a genome-scale identification of NC enhancers and their upstream TF inputs. Specifically, we uncovered *cis*-regulatory interactions driving different NC lineages including a number of novel factors; ”super-enhancer”-like regions regulating NC regulators; and single-cell signatures transcriptomic heterogeneity, and their correlated epigenomic signatures, within the developing NC. To our knowledge, this represents the first high-resolution, genome-wide reconstruction of a vertebrate GRN *in vivo* integrating distal-regulatory elements as the core computational units, highlighting the potential of our method to decode complex biological networks underlying ontogeny.

2.8.1 Gene expression signatures of NC

Our transcriptional analysis has yielded an comprehensive view of the factors influencing NC formation, significantly increasing the resolution of NC-GRN [31, 33]. For instance, in addition to assigning known and novel regulators to the canonical modules of premigratory (TFAP2A/B, Msx1, Zeb2, Sox9, Pax7, Dmbx1 etc.) and migrating (Sox5, Sox10, Sox8, Lhx9, Pax3, Wnt4, Wnt3a, Ednrb, Robo1, Cxcr4, etc.) NC, we have also observed a clearly defined shift from homophilic cell-to-cell communication based on cadherins and protocadherins (Cdh6/13, Pcdh8) to heterophilic cell-ECM communication (Ednrb, Cxcr4, Plxna4, Adamts1/20), demonstrating comprehesinve remodelling of the gene regulatory network as NC cells undergo EMT and commence migration.

Moreover, characterisation of co-expression clusters has expanded the previously proposed modular network structure to refine NC factors into more precise categories (early NC specification, onset of EMT, (late) NC specification, active NC migration, extracellular signalling linked to NC specification, etc.), enabling us to include and classify novel regulators within the NC-GRN. For instance, the zinc-finger homeobox protein 4 (Zfhx4), which was recently shown to regulate multipotency, self-renewal and migration of therapy-resistant glioblastomas (NC-derived brain cancer) [239], was recovered in the late NC specification module containing core NC specifiers (Sox10, TFAP2A, TFAP2B, Pax7, etc.). The LIM adaptor protein, Lmo4, previously described as an essential regulator of NC induction [240, 241] EMT [240] was recovered in the GRN module featuring myocyte-specific enhancer factors MEF2A/C, and late non-canonical Wnt factors (Wnt7A/B), required in NC for proper craniofacial development and during migration, respectively [242, 243]. Both Zfhx4 and Lmo4 featured NC-specific upstream regulatory apparatus. In particular, Zfhx4 expression was solely controlled by *k*-Cl-1 enhancers, showing NC specific activity (enh-5 and enh-316) as well as non-NC regulation (enh-11 and enh-12). Similarly, NC-specific expression of Lmo4 was governed by *k*-Cl-3 enh-17, whereas non-NC activity was controlled by *k*-Cl-1 enh-15.

2.8.2 NC *cis*-regulatory architecture

Investigation of NC regulation at the epigenomic level has uncovered dynamic changes in chromatin accessibility of distal regulatory elements employed in the NC and other embryonic cell-types at different stages of NC ontogeny. Our global analysis of chromatin dynamics has revealed: (1) an NC-independent class of ubiquitously accessible elements, enriched at promoters (*k*-Cl-4); (2) a class of elements specifically open only in the NC, at either premigratory and/or migrating stages (*k*-Cl-3); (3) a class of elements dynamically open in both NC and neuroepithelial cells (*k*-Cl-1); and (4) a class of elements accessible in naïve epiblast and premigratory NC cells that become inaccessible by migrating stages (*k*-Cl-6). Dissection of the specific enhancer clusters by TF binding and co-binding motif enrichment analysis, target gene assignment by positional information and high-resolution targeted chromosome conformation capture analysis of enhancer-promoter contacts, and functional classification of assigned target genes has indicated that three key clusters (*k*-Cl-3, *k*-Cl-6 and *k*-Cl-1) embed the regulatory logic of NC development.

The *k*-Cl-3 elements were mainly associated with the NC loci, such as TFAP2A/B and Sox10, whereas *k*-Cl-6 and *k*-Cl-1 elements were respectively enriched at loci related to stem cells and neural progenitors, such as Oct4 and Otx2. We propose that *k*-Cl-6 elements initiate NC specification, likely through *cis*-regulatory pleiotropy, for example, reusing enhancer elements from epiblast cells in the premigratory NC. Establishment of an early NC programme is captured by *k*-Cl-1 elements, consistent with the NC origin at the dorsal neuroepithelium. Therefore, we hypothesise that *k*-Cl-1 might represent NC and non-NC neural progenitors, to which the early, premigratory NC share regulatory circuits. In contrast, *k*-Cl-3 elements may represent spatially restricted, NC. DiffBind clusters at 5-6ss and 8-10ss capture the temporal dynamics of canonical premigratory and early delaminating NC, revealing a global picture of NC spatiotemporal heterogeneity at the epigenomic level. Taken together, our data suggest that the regulatory logic of NC development is hierarchically embedded in the dynamic *cis*-regulatory landscapes. This approach has thus allowed us to not only comprehensively profile active NC enhancers but also to dissect the regulatory inputs and their downstream targets, from which the NC-GRN could be reconstructed in a genome-wide fashion.

2.8.3 Pre-patterning of the early NC

By analysing circuitries mediated by distinct clusters of NC regulatory elements (*k*-Cl-3 versus *k*-Cl-1), we have revealed the heterogeneity of *cis*-regulatory landscapes within cranial NC cells. Furthermore, we clearly link this regulatory dichotomy to the heterogeneity uncovered at the transcriptional level using single-cell profiling. Our results show that identified single-cell NC clusters featured either canonical NC or neural-NC identity, and associated to the corresponding *cis*-regulatory clusters with high-statistical support, thus linking their *cis*-regulatory and intrinsic transcriptional dynamics.

2.8.4 Combinatorial regulatory codes for specification of NC fates

We have identified unique combinatorial *cis*-regulatory codes underlying either canonical NC identity featuring NC interactions Sox10-Sox8/9, Sox10-TFAP2, as well as co-activities with critical new players ATF and Arnt) or neural-NC identity enriched for non-canonical interactions including Otx2-Pax2, Otx2-NR2F2, Pax2-Sox9 and Otx2-Sox9. Our results build on the previously proposed minimal network underlying the switch from neural to mesenchymal fates, involving Sox9 and FoxD3 [63], suggesting a previously unappreciated involvement of Otx2, Pax2 and NR2F2 in the NC-GRN. By performing CRISPR/Cas9 knockout of top regulators of both programmes *in vivo*, we showed that corresponding transcriptional networks mediated the activation of canonical NC regulators and direct repression of neural pathways. The activation of an entire *cis*-regulatory landscape specific to canonical NC and mesenchymal progenitors, represented by *k*-Cl-3 elements, appears vulnerable to perturbation, as the disruption of the corresponding core transcriptional network leads to a robust change towards a neural progenitor state. However, the rewiring of the early neural *cis*-regulatory landscape, represented by a large number of strongly interconnected *k*-Cl-1 elements robustly coupled by auto-regulatory and feed-forward loops and active in both NC and neuroepithelial progenitors, appears more difficult to achieve. Such a strong, early relationship between NC and non-NC neuroepithelial programmes may to some extent explain the NC-derived neural progenitors' resilience to reprogramming [56]. Nonetheless, the modest yet significant effect in shifting from neural to mesenchymal fates (as revealed by upregulation of Twist1/2 and TFAP2A/B, among others), primarily dictated by Otx2 and Pax2 disruption, indeed suggests that these two core networks mutually negatively regulate one another and thus dynamically balance neural and canonical/mesenchymal NC identities. In order to test this hypothesis in the future, it will be necessary to identify negative regulators repressing one identity at the expense of the other, and directly demonstrate such putative repressors directly bind downstream targets promoting alternative fates. A similar approach was used in our lab, while investigating the bimodal role of foxd3 in zebrafish NC, where foxd3 was shown to function directly as a negative regulator of neurogenic NC binding downstream genes promoting neurogenesis, while at the same promoting NC multipotency [60].

The involvement of Orthodenticle homeobox 2 (Otx2) gene within the NC-GRN has long been hypothesised [235], given the craniofacial abnormalities that accompanied heterozygous Otx2+/- mice Because Otx2 deficiency causes atrophies and deletions, Otx2 has thus been proposed to act as a gap gene in vertebrates, in contrast to its function both in homeotic and gap-like transformations in *Drosophila* [234, 235]. Our unbiased, systems-level approach, has enabled the positioning of this factor within the NC-GRN, by identifying its upstream regulators (for example, Sox9/10, Sox2/3, Pax2, Msx1/2, Arnt/Arnt2, E2F6, Lhx, RXRg, Pax2/8 and self-regulation) as well as numerous downstream targets (including *Adamts19, Cdh6/9/10/13, Col4a2/6, Ednrb, Epha6/7, Fgfr3, Pcdh8/9/15/19, Pdfrl, PlxnA4, Robo1, Sema3D/5b/6A/6D, Smad3, Sox11, Zfhx4* and *Zic1* among many others). Similarly, the non-canonical role of Pax2 and NR2F2 in controlling neural fate restriction in NC cells illustrates the power of our approach to resolve novel hierarchies and genome-wide circuits.

2.8.5 A genome-wide view of the NC-GRN

Our integrated approach, using the intermediary of active NC-specific enhancers, has enabled the assembly of comprehensive NC-GRNs, linking predicted upstream regulatory factors to their downstream target genes (Appendix figures 7.6, 7.7, 7.8, 7.9). Our approach has revealed new hierarchies and modular relationships as well as recapitulated previously known interactions, highlighting its utility to interrogate the complex regulatory logic controlling early NC ontogeny. Moreover, the fact that at least 57% of the genes downregulated upon specific network core perturbations were directly regulated by at least one *k*-Cl-3 or *k*-Cl-1 element as per our analysis confirms the causal links within the proposed NC-GRN circuits. With the recent and on-going development of novel, low input methods to profile TF binding *in vivo* [115], we expect to soon be able to circumvent the key limitation of our approach, which is the reliance on *cis*-regulatory dissection through the use of known, high-resolution motifs. We also assembled our data into a interactive ShinyApp (Appendix Figure 7.5), which facilitate access and browsing of these multimodal datasets by the scientific community, freely available at: https://livedataoxford.shinyapps.io/Chick_NC_GRN-TSS-Lab/.

Taken together, our data provide whole-genome framework for understanding the early regulatory events underlying essential NC cell functions. We expect that our resources will be useful for the interrogation of the circuitries underlying essential processes during NC ontogeny, such as maintenance of stemness and multipotency, as well as broader cellular processes and behaviours such as EMT, migration, differentiation and fate restriction.

2.8.6 Limitations of our study

Importantly, this study heavily relies on the identification of positive interactions driven by enhancers and their activators. Negative interactions driven by *cis*-regulatory repressors, repressive TFs and their co-factors were not taken into account, which represents an important limitation of our approach. I anticipate that the cross-balance between *k*-Cl-3 and *k*-Cl-1 networks may be regulated by putative, cross-negative interactions that await to be uncovered.

2.8.7 Summary

This chapter of my doctoral thesis presents evidence that the *cis*-regulatory architecture encodes information pre-patterning the early NC into neurogenic progenitors *versus* ectomesechymal progenitors. We identified an extensive cohort of enhancer elements, which were dissected and integrated to transcriptomic profiles at single cell resolution, allowing us to reverse engineer the avian NC-GRN. Interrogation of the combinatorial *cis*-regulatory code controlling neurogenic enhancers (*e.g.*, *k*-Cl-1) versus enhancers active in, ectomesechymal progenitors (*e.g.*, *k*-Cl-3) allowed us to identify a opposing gene circuits that control NC fate decisions.

3

The *cis*-regulatory architecture underlying the early origins of the NC

In this Chapter, we took on a comparative approach across different vertebrate species to investigate the regulatory basis of NC evolution. We first generated novel transcriptomic datasets at single cell resolution from the avian and zebrafish models to characterise transcriptional dynamics during early NC development, and re-utilised a dataset recently published [67] enabling us to include the murine NC into our comparative analyses. We also conducted a preliminary investigation of the lamprey NC at single cell resolution, revealing wide-spread deployment of Zeb-like paralogues likely sustaining lamprey EMT. Next, we used a computational approach that enables us to "integrate" such single cell profiles obtained in different vertebrate species within the same low dimensional space; co-embedding of subpopulations of NC cells from different species within similar clusters provides strong statistical support for a deep conservation of the transcriptional programmes underlying early NC heterogeneity. To deepen our investigation, we conducted co-expression analysis and K-means clustering of bulk RNA-seq and ATAC-seq, respectively, spanning early NC induction and specification in zebrafish using an experimental design that parallels our investigation in the avian embryo. This allowed us to directly compare our results between the avian and zebrafish NC throughout their ontogeny.

On one hand, our results suggest that, at the *trans* level, NC programmes are highly conserved between zebrafish and amniotes, albeit deploying different paralogues, consistent with the idea that NC cells evolved before a whole-genome duplication event in the stem vertebrate. On the other hand, our results indicate that early neurogenic progenitors were co-opted into the NC programme via enhancer pleiotropy. Indeed, pleiotropic enhancers are shared between neurogenic NC and neuronal cell-types, such as hindbrain cells and cranial placodes, suggesting a common regulatory architecture between neural cell-types. Surprisingly, we find that enhancers involved in ectomesenchymal development are restricted to the NC. Contrary to previous hypothesis [48, 77], these ectomesenchymal circuits are not shared with mesodermal cell-types, such as pre-somitic mesoderm, and have likely evolved *de novo*. Our results, thus, have important implications for understanding NC origins and diversification at the dawn of vertebrates.

3.1 Experimental strategy to profile early NC cells across vertebrates

NC evolution represents a major transition in the radiation of vertebrates [47]. Proposed to be an evolutionary elaboration of vertebrates [191], NC uniquely contribute to the vertebrate body plan likely underlying the transition from a filter feeding to predatory life-style [34]. NC are thought to have contributed to the sensory ability of vertebrates to navigate their environment and to the acquisition of adaptations that enable them to camouflage in their habitat. Nevertheless, our understanding of NC evolution still remains sparse owing to a limited number of vertebrate taxa used as models, and the challenge to study NC cells *in vivo* due to their transient nature and highly migratory behaviour. Here, we circumvent some of these challenges using a combination of transgenic models that specifically label NC cells in living vertebrate embryos. We then interrogate these NC cells, collected directly from their endogenous environment, with single cell RNA-seq (scRNA-seq) and ATAC-seq allowed us to study NC gene regulation across 540 million years of vertebrate evolution.

To interrogate the NC cells in a unbiased manner, we first conducted scRNA-seq in the avian, zebrafish and lamprey embryos. Although we previously characterised early avian NC cells using the Smart-seq2 approach (Figure 2.25)[59], this technology displays an inherent lower throughput. We circumvented this limitation using the microfluidics-based approach from the 10X Genomics Chromium platform. Dr. Ruth Williams electroporated avian embryos with *FoxD3* NC1.M3 enhancer [220] driving *Citrine*, and cranial NC cells were collected from embryos at 7-8ss for scRNA-seq (Figure 3.1a). After library preparation and sequencing, scRNA-seq profiles were t compared and annotated in accordance to [59, 67].

Conversely, we used the zebrafish *FoxD3* gene trap line *Gt(foxd3-Citrine)*^{ct110a} [244], in which one allele of *FoxD3* is fused to *Citrine*, thus allowing us to sort FoxD3-labelled embryonic cells at desired stages (Figure 3.1b) for a representative teleostei. For single cell profiling using scRNA-seq and scATAC-seq, we collected embryos with FoxD3-expressing Citrine-labelled NC cells at 6-7 somite stage (ss), performed by Dr. Martyna Lukoseviucute. Bulk-RNA-seq profiles from [60] were also reanalysed. These datasets span the stages covering the early NC induction through NC specification. We also performed high-resolution bulk-ATAC-seq profiling at similar stages, in a manner that allowed us to directly compare the epigenomic landscapes to their transcriptional states. Moreover, this experimental design mimicked closely our approach used to reverse engineer the avian NC, thus allowing us to directly compare regulatory architecture of the avian and zebrafish NC.

Finally, we used Tahara stage 21 (T21) embryos [245] from the marine lamprey, *P. marinus* (Figure 3.1c), to profile for the first time lamprey NC cells at single cell resolution. Dr. Ruth Williams and Dr. Dorit Hockman optimised marine lamprey dissociation. In the absence of fluorescent markers for direct profiling of the cell-types of interest, Dr. Dorit Hockman dissected dorsal neural tubes, containing a region enriched for NC cells. Furthermore, Dr. Dorit Hockman also conducted ATAC-seq profiling of lamprey NC cells [169], suggesting the existence of enhancer elements with similar properties (*i.e.*, development dynamics of chromatin accessibility, functional activity and motif content) to that of avian and zebrafish NC.

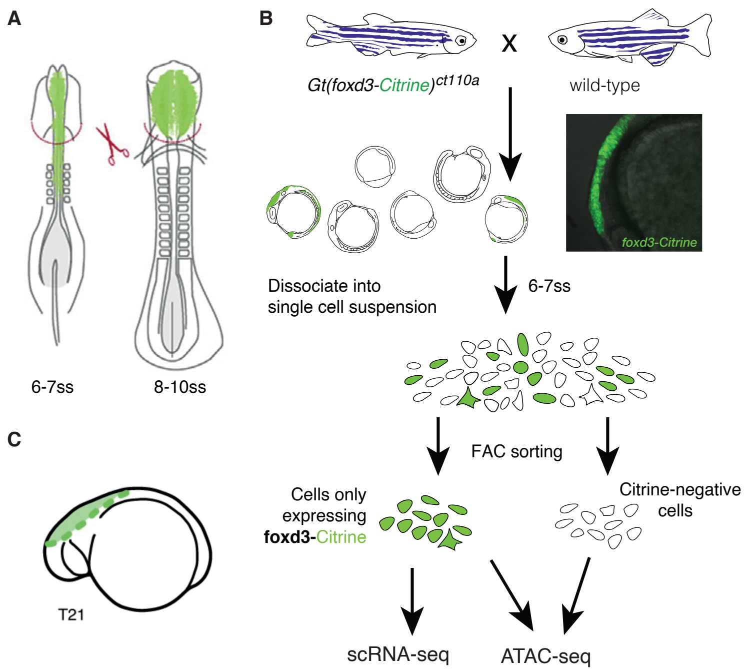

Figure 3.1 Schematic diagram of the experimental strategy to profile NC across species.. (a) Chicken NC cells labelled by *FoxD3* NC1.M3::Citrine at 7ss were FAC sorted for single cell transcriptomic profiling. (b) *FoxD3* gene trap line Gt(foxd3-Citrine)^{ct110a} was outcrossed, fluorescent embryos were sorted and grown to desired stages, and cells were dissociated and FAC-sorted for downstream analysis. (c) Dorsal neural tubes were dissected from marine lamprey embryos at T21 stages, a tissue enriched for pre-migratory NC. Figures adapted from [59, 60, 169].

3.1.1 The amniote NC at single cell resolution

First, we characterised the avian NC labeled with Citrine driven by the *FoxD3* NC1.M3 enhancer (Figure 2.1). Importantly, this mutated enhancer displays stronger activity than the endogenous NC1 enhancer, due to the mutated FoxD3 repressor binding sites [59, 220], allowing us to collect both the pre-migratory and delaminating and early migrating cranial NC cells. Our approach resulted in 1509 high-resolution profiles, which were processed with standard scRNA-seq work flow (see Methods), involving quality control, dimensional reduction, calculation of UMAP (Uniform Manifold Approximation and Projection) embeddings and clustering (see Methods). We identified eight Leiden clusters (Figure 3.2a), which were annotated using comparisons to our previous results (Figure 3.1) and further comparison with the murine NC (Figure 3.2b), [67]. Visualisation of the expression of predictive canonical regulators, such as *Ets1*, *Draxin*, *TFAP2a*, *Sox10*, *Sox9*, *Sox8*, *Msx1*, *Wnt4*, in addition to cell-cycle markers, *Cdk1* and *Kpna2*, among others (Figure 3.1c), provided support to the identification of our single cell clusters.

Marker genes (Figure 3.3) and relative levels of expression (Figure 3.4) further extended our analyses. We named different subpopulations of NC cells: **migratory NC** based on the high levels of expression of *Lmo4* (Figure 3.4), as well as canonical NC factors, such as *Ets1* and *Sox10* (Figure 3.2); **delaminating NC** based on co-expression of *LdhB* and *Draxin*; **non-neural NC** characterised by a niche of *EPCAM* and *FABP3*-expressing cells; and **anterior precursors** (APs), with overlapping identity to **early neurogenic NC**, represented by double positive Six3/Otx2-expressing cells. **Pre-migratory cells** were largely identified by positional signals located in the Neural Plate Border (NPB) and dorsal neural plate, such as *Wnt4* and *Msx1*; **multipotent NC** expressed pluripotency markers, such as *Cdk1*, as well as G2M genes such as *Tuba1b*, in addition to the anteriorising homeobox gene *Hesx1*, whereas **KPNA2+ positive cells** displayed expression of similar G2M markers but also relatively high levels of this karyopherin-a paralogue. Finally, **ectomesenchyme progenitors**/undifferentiated ectomesenchyme cells were defined by high levels of *Pitx2*, *Six1* and *Prrx1*.

Taken together, our analysis of the avian NC is consistent with the idea that the

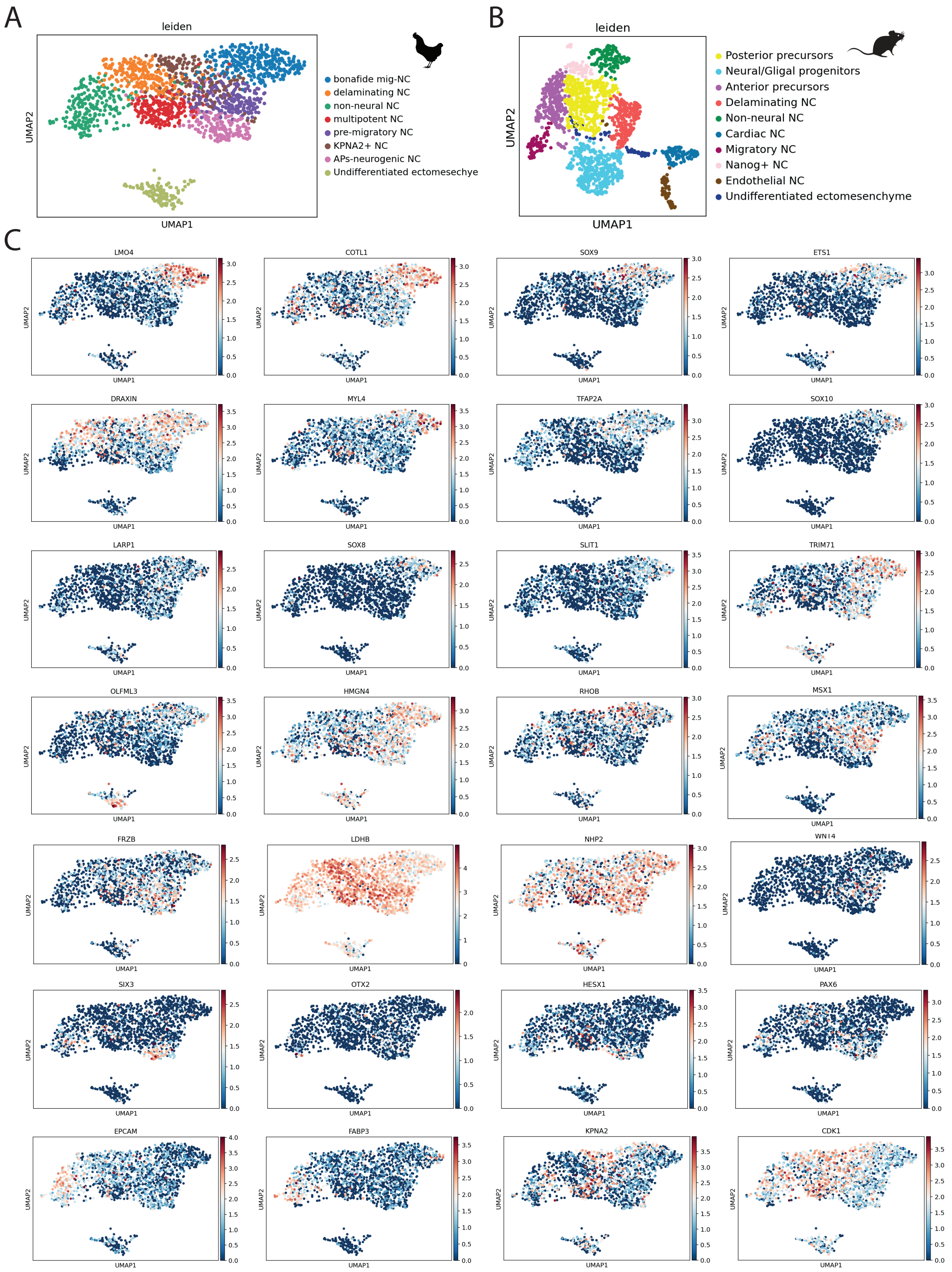

Figure 3.2 The avian and murine NC at single cell resolution. (a) UMAP embedding of the chicken NC coloured by cluster annotations. (b) UMAP embedding of the murine NC coloured by cluster annotations. (c) UMAP embedding of the chicken NC coloured by gene expression of multiple marker genes.

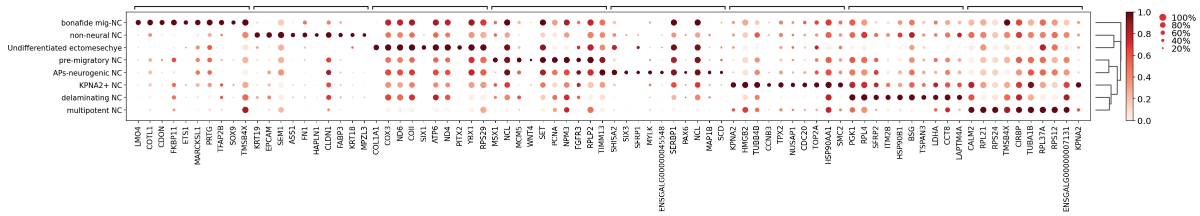

Figure 3.3 Dotplot of the top10 markers of each cluster identified in the avian NC.

early NC is at least transitionally heterogeneous at the transcriptional level. We identified early segregation of the NC subpopulations arising at the midbrain level. Co-expression of markers with genes previously unrelated to NC development, including for example *Tetraspanin3*, *Cotl1*, *Trim7l*, and *Olfml3*, indicate these genes extend our analyses and reveal novel factors that may be implicated in early NC development. Moreover, unsupervised identification of *Lmo4* as the main marker of migratory NC illustrates the importance of this *Snai1/2* co-factor to avian NC EMT.

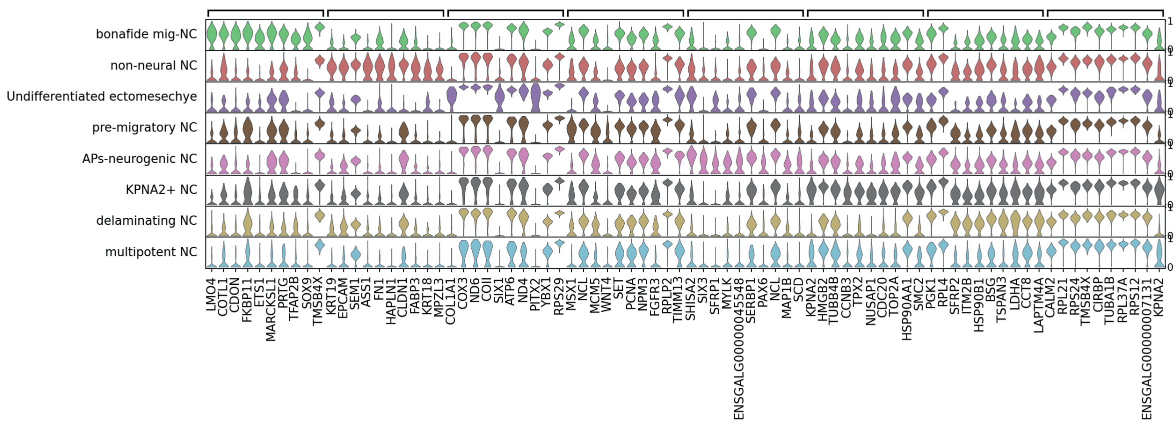

Figure 3.4 Violin plots for top10 marker genes of each cluster identified in the avian NC.

Next, we revisited the recently published murine NC dataset [67] spanning the ontogeny of NC cells from early induction to early differentiation. Importantly, murine NC were collected from *Wnt1::Cre; Rosa26*$^{TdTomato/+}$ m mouse line at the 4ss (early NC induction), 6ss (pre-migratory NC), 8ss (NC specification), and

10ss (NC migration and early differentiation) embryos dissected at the midbrain-hindbrain and FAC-sorted. We analysed the cells using a similar computational approach used for the avian NC (see Methods), and our results recapitulated previous features [67]. 10 clusters were identified and annotated according to the original manuscript (Figure 3.2)b).

Again, marker genes (Figure 3.5) and relative levels of expression (Figure 3.6) allowed us to identify: (1) **Early Wnt1 precursors** based on the expression of low levels of *Sox2*, *Oct4*, *Nanog* and *Otx2*; (2) **anterior precursors**, displaying similarities to our cluster of **neurogenic progenitors** in the avian NC (Figure 3.4) were enriched for *Sox2*/*Otx2* double positive cells, as well as *Rax*; (3) **posterior precursors** representing post-otic NC cells, identified based on the caudalising homeobox *Gbx2*, as well as *En1*; (4) **non-neural NC** cells, identified based on the expression of *Epcam*, as well as *Cdh1*; (5) **delaminating NC** displayed similarities to our migratory NC cluster, including the cells co-expressing of *Ets1* and *Sox10* and also enriched for *Snai1*, *Mef2c*, *Erbb3*, *Camk2b* and *Dlx2*; (5) committed **neural/glial progenitors** displayiing high levels of *Camk2b* and *Dlx2*; (6) **cardiac NC** expressing *FoxC2* and Pitx2; a putative **endothelial progenitor** cluster (7) containing cells expressing *Fli1* and *Sox7*; (8) **undifferentiated ectomesenchyme/ectomesenchyme progenitors**, identified based on the expression of *Prrx1*; (9) undefined Wnt1+ cells expressing *Psap* and *Gpx3*; and finally (10) **nanog+ cells**, a small niche of multipotent precursors displaying expression of Yamanaka factors (*Nanog*, *Sox2* and *Oct4*).

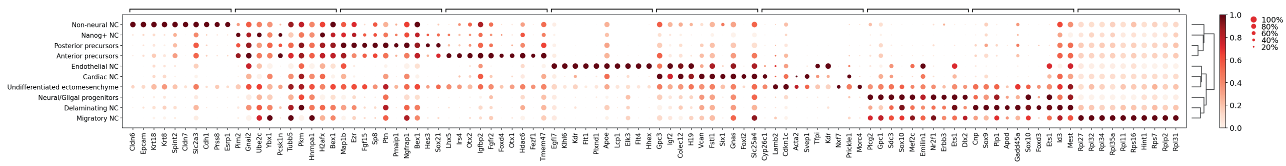

Figure 3.5 Dotplot of the top10 markers of each cluster identified in the murine NC.

Taken together, these results establish the transcriptional landscape of NC cells in two amniote species, the chicken and mouse. Collectively, we provide evidence for transcriptional heterogeneity in the NC of these two species. Expression of orthologous markers in subpopulations of NC cells indicate that such heterogeneity likely predates the radiation of amniotes.

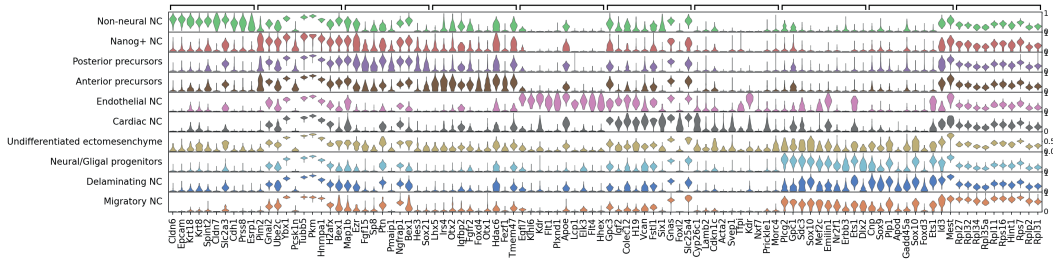

Figure 3.6 Violin plots for top10 marker genes of each cluster identified in the murine NC.

3.1.2 The zebrafish NC at single cell resolution

Next, we used *Gt(foxd3-Citrine)*^{ct110a} embryos to profile foxd3+ cells from a species that has diverged before the radiation of amniotes, the zebrafish. After quality control, dimensional reduction, UMAP projection, embedding and clustering, we identified two main population of *foxd3* expressing cells: a neuromesodermal progenitor (NMP)-like population, thoroughly investigated by Dr. Martyna Lukoseviciute [246], and a population of foxd3$^{\text{high}}$ cells, corresponding to cranial and vagal NC. We eliminated NMP-like cells for downstream analysis, and focused on the latter group (see Methods). Importantly, this marked difference to our avian and murine datasets arises from the use of *Gt(foxd3-Citrine)*^{ct110a} transgenic line, which tags *foxd3* expressing cells at all axial levels. Foxd3$^{\text{high}}$ NC cells were clustered into nine groups (Figure 3.7), and differential gene expression (Figures 3.8 and 3.9) was analysed to identify cluster identity.

Analyses of marker genes (Figures 3.8 and 3.9) as well as their relative levels (Figure 3.10), revealed the presence of: (1) a **NC** subpopulation, expressing

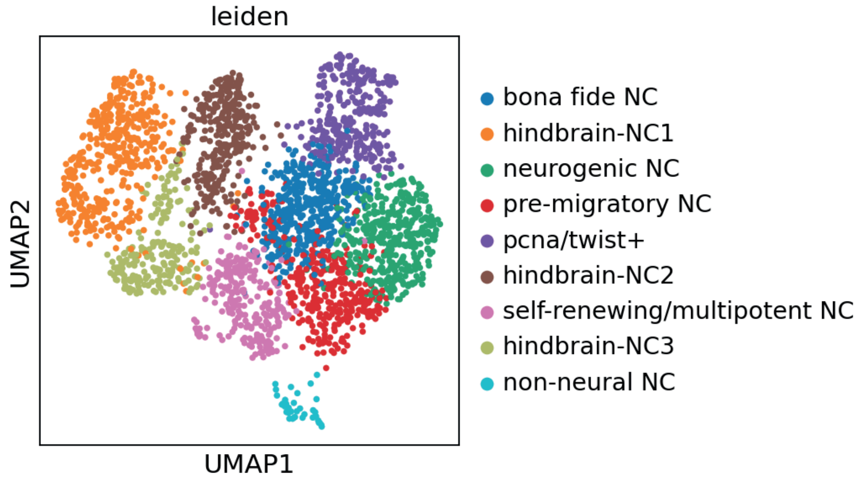

Figure 3.7 The zebrafish NC at single cell resolution. UMAP embedding of the zebrafish NC colored by annotated clusters.

twist1b, *col15a1b*, *nrp2a*, *pdgfra*, *alcama*, *cldn5a*; (2) **neurogenic NC**, expressing *pax3a*$^{\text{high}}$, *fgfr4*$^{\text{low}}$, *alcama*, *otx1*, *msi1*, wnt4; (3) **pre-migratory NC**, expressing *pax3a*$^{\text{low}}$, *fgfr4*$^{\text{high}}$, *otx2b*, *msx1b*, *zic3*, *her11*, *wnt8b*; (4) **non-neural NC**, expressing *epcam*, *apoeb*, *alx1* and *actb1*; (5-7) **hindbrain NC** clusters, expressing *mafba*, *crabp2a*, *egr2b*, *crestin*, *hoxb2a*, *hoxb3a*, *irx7*, *meis1b* and *meis3*; (8) **self-renewing/multipotent NC** cells, expressing *apoeb*, *actb1*; and finally (9) ***pcna/twist1b*-expressing cells**, which display a similar profile to NC cells, but also express differential levels of *rrm2*, *cdc6*, *asf1ba*, *kpna2*.

Interestingly, these subpopulations display similar properties to those identified in the chicken and murine NC, including the expression of orthologues, such as *wnt4*, *fgfr4*, *alx1* and *epcam*.

However, some zebrafish subpopulations seem to have "substituted" the reciprocal 1:1 orthologue, as the expression of homologous genes is better explained by paralogues. For example, zebrafish **pre-migratory NC** expressed *msx1b*, in contrast to *Msx2* in chicken and mouse; conversely, we found that putative neurogenic NC express *otx1*, instead of the canonical *Otx2* in amniotes.

These results led us to suggest that although the transcriptional landscape seems

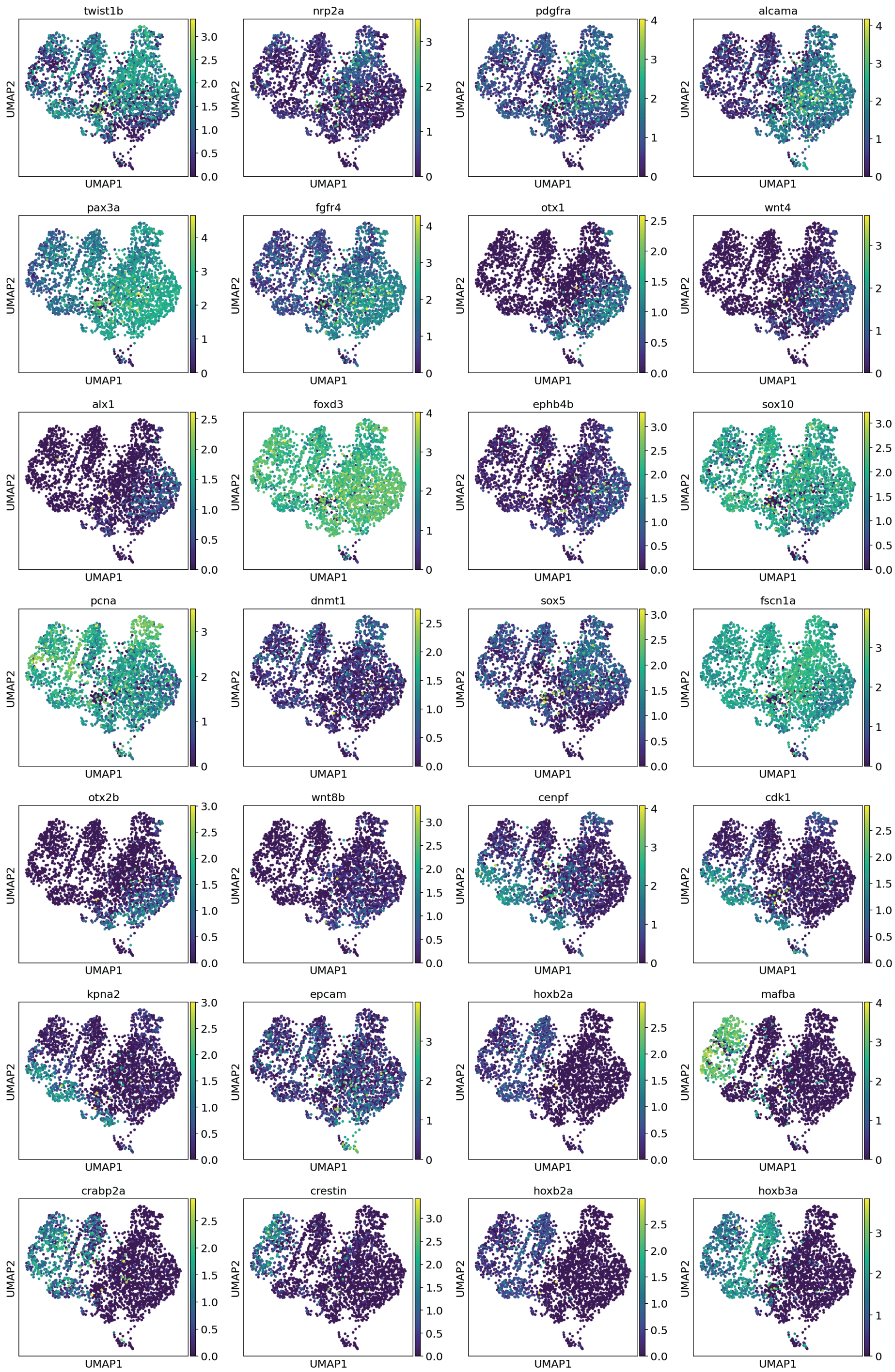

Figure 3.8 UMAP embedding of the zebrafish NC coloured by expression of marker genes.

to be largely conserved across the three species (zebrafish/chick/mouse), in several instances paralogous gene-pair "substitutions" seems to have taken place during

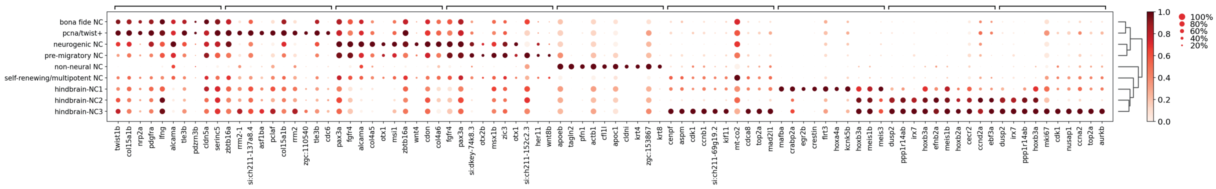

Figure 3.9 Dotplot of top10 markers for each cluster identified in the zebrafish NC.

the evolution of the NC. This observation challenges the common assumption to attribute orthologous functions to the best 1:1 orthologous genes, and may instead suggest some flexibility in the deployment of paralogous genes.

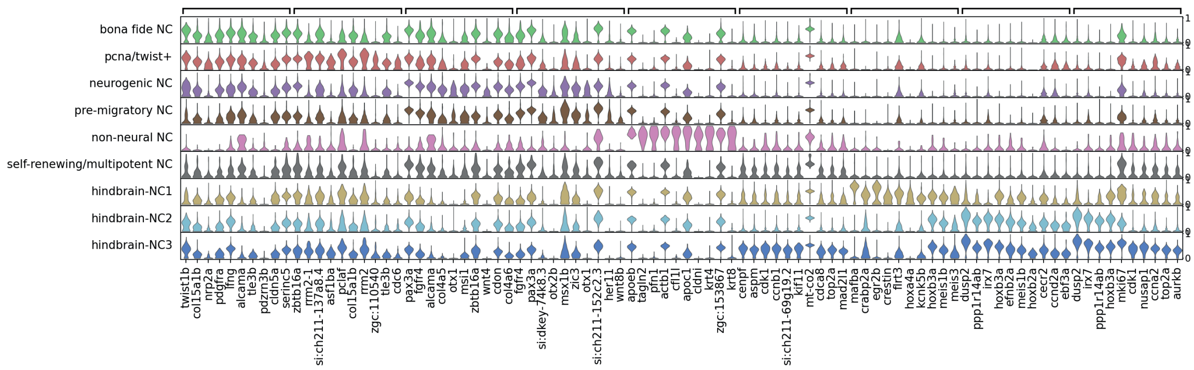

Figure 3.10 Violin plots for top10 marker genes for each cluster identified in the zebrafish NC.

Moreover, these results are also consistent with the idea that the NC evolved before the whole genome duplication event in the stem vertebrate ancestor [177]. A parsimonious explanation for the differential use of paralogous genes in these species is the acquisition of duplicated genes/genomes followed by their neo/subfunctionalisation, with one or the other paralogue being retained in the NC, leading to the appearance of paralogous "substitutions". An alternative explanation is the actual substitution of 1:1 orthologues by their paralogues via compensatory evolution.

3.1.3 Single cell analysis of NC across species

Next, I employed a novel algorithm, self-assembling manifolds mapped across species (SAMap) [196], which enables the alignment of scRNA-seq across large evolutionary distances within the same manifold. Firstly, SAMap uses a homology graph, which in this case is represented by best reciprocal hits (BRHs) identified by blast. In other words, all transcripts were translated and sequence similarities were analysed for every translated sequence between genomes (but not within the same genome). BRHs between two or more species are thus considered orthologues if they have diverged according to the species tree. This homology graph, representing sequences ranked according to their sequence similarities, is then used during the first iteration of SAMAP. In the first iteration, BRHs are used as anchors between each single-cell dataset, thus allowing them to align within the same lower dimensional space. However, in several instances, the BRH are not expressed, and therefore their sequence similarity may not reflect functional conservation. To circumvent that, SAMap then reiteratively assign each "anchoring" gene to the consecutive best hit, thus enabling for the identification of homologous genes that may not represent the most conserved ones at the sequence level, but rather those that share similar levels of expression between species. This strategy thus circumvents the constraints imposed by sequence orthology, which are bypassed via the analysis of expression similarities between homologous genes between mapped cells, sometimes representing genes that have diverged during large evolutionary distances [196]. Relative contributions of homologous genes are then assessed through multiple iterations until the manifolds robustly converge.

We first performed a pairwise analysis between the avian, murine and zebrafish NC profiles (Figure 3.11). Interestingly, chicken to mouse and zebrafish to mouse (Figure 3.11a-d) displayed strong species-specific bias, whereas chicken to zebrafish integration was largely overlapping (Figure 3.11e-f). This result suggested a strong technology-bias in the mouse dataset, which resulted from Smart-seq2 methodology used to profile murine NC cells. However, we cannot rule out species-specific effects with the current data; in order to test that, we will use a Pax3:Cre mouse line and profile Pax3+ NC cells with 10X Chromium, as a manner to rule out technology-biases.

Nevertheless, our pair-wise assessment indicate that the transcriptional programmes underlying NC heterogeneity are highly conserved. Mapping scores (Figure 3.11, b,d,f) between NC clusters of each species indicate high overlap of the transcriptional states between these species. Interestingly, the highest scoring pair is illustrated by the cluster of epcam+ **non-neural NC**, suggesting high divergence of this transcriptional programme compared to the others.

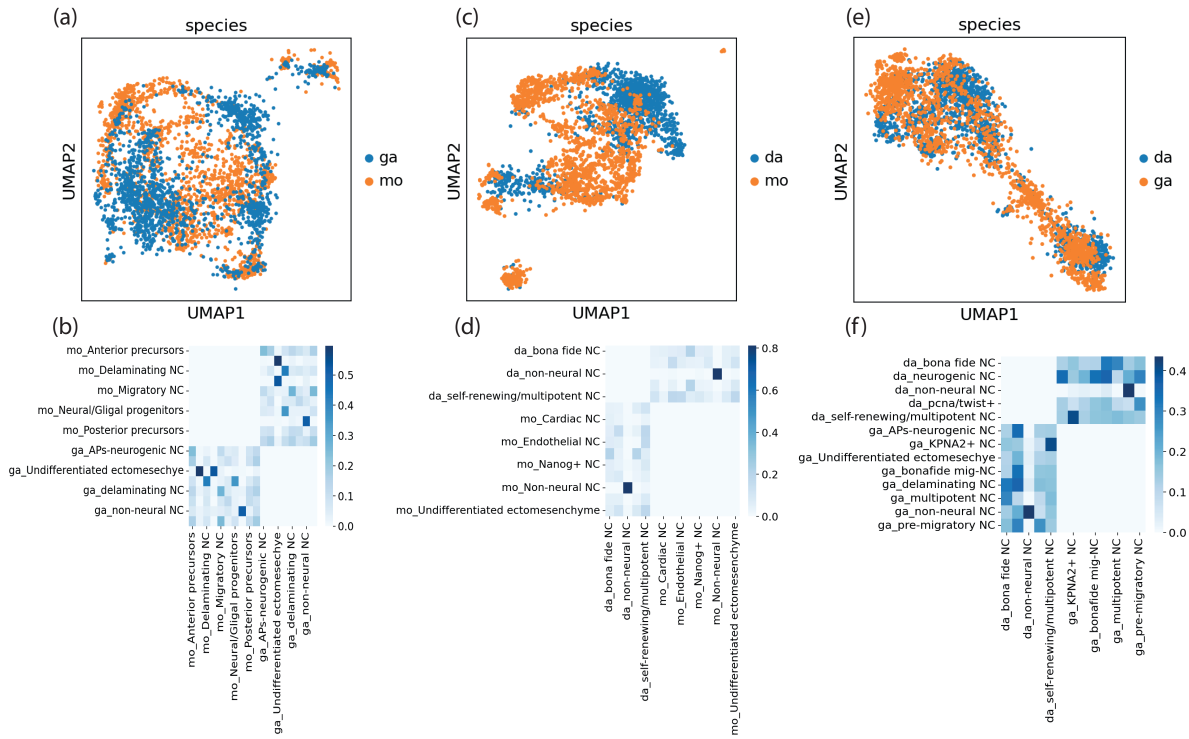

Figure 3.11 Pairiwse SAMap alignment of single cell profiles across species. (a-b) mouse chicken, (c-d) mouse zebrafish, (e-f) chicken zebrafish. Top panels show UMAP embeddings coloured by species. Bottom panels show a matrix of mapping scores between clusters from each species. mo, *Mus musculus*, ga, *Gallus gallus*, da, *Danio rerio*.

We further assessed the interrelationships of NC clusters across species using the mapping scores as a manner to identify reciprocal matchings. Sankey plots for chicken to mouse, zebrafish to mouse and zebrafish to chicken (Figure 3.12a-c) revealed higher inter-connectivity between more developmentally differentiated clusters (e.g., undifferentiated ectomesenchyme, cardiac NC and non-neural NC) versus biased but not committed lineage progenitors, such as neurogenic NC and ectomesechymal NC. This is further illustrated by avian multipotent NC precursors, which displayed connectivity to multiple murine clusters, including nanog+, anterior (otx2+) and posterior (gbx2+) precursors and neuro/glial progenitors.

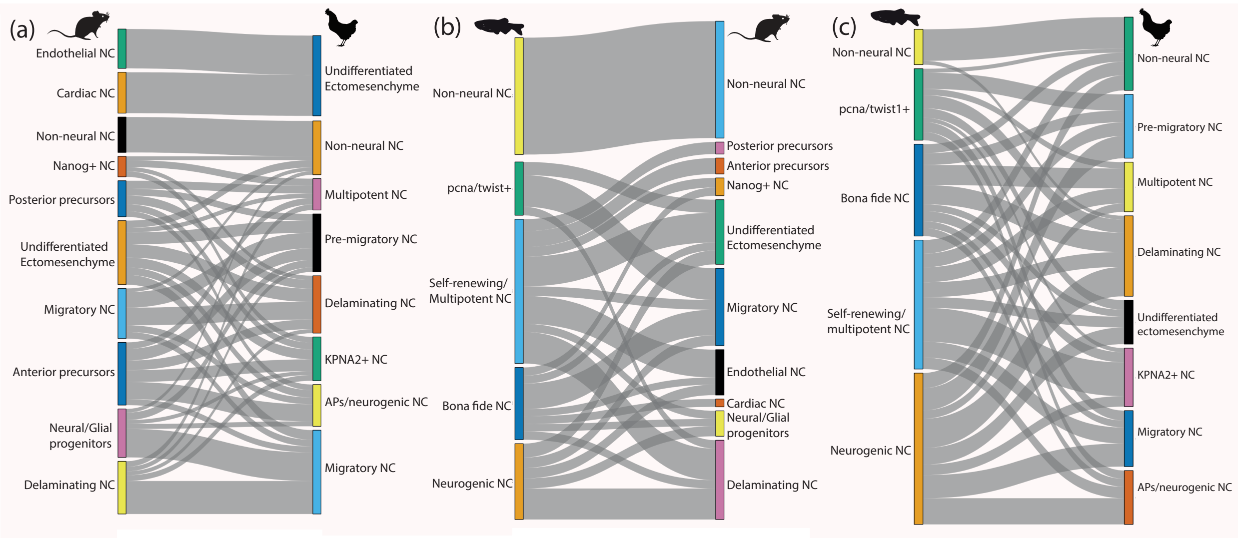

Figure 3.12 Sankey plot mapping of pairwise integration between mouse, chicken and zebrafish NC cells. (a) SAMap alignment of mouse and chicken, (b) mouse and zebrafish, and (c) chicken and zebrafish datasets. mo, *Mus musculus*, ga, *Gallus gallus*, da, *Danio rerio*.

To further illustrate high conservation of the transcriptional states across these species, we identified the gene-pairs underlying such mapping scores. As expected, chicken-to-mouse comparison revealed similar deployment of 1:1 orthologous genes. For example, cross-mapping of *bona fide* NC is sustained by similar levels of *Sox9*, *Sox10*, *Ets1* and *Id2*. Nevertheless, even at this shorter evolutionary distance, instances of paralogous "substitutions" could be identified, such as MYC expression the chicken *bona fide* NC in contrast to Mycn in mouse. Similarly, the cross-mapping of more committed lineages, such as undifferentiated ectomesenchyme from chicken to cardiac NC from mice is sustained by expression of *Six1*; but chicken expresses *Tbx22* in contrast to *Tbx1* in mice. Another example is revealed in the cross-mapping of chicken neurogenic NC to anterior precursors from mice. To a large extent, the mapping is sustained by 1:1 orthologues of *Shisa2*, *Sfrp1*, *Hesx1*, *Frzb* and *Cdh2*; but FGF receptors displayed paralogous "substitutions": chicken NC deployed *Fgfr1/3*, whereas mouse deployed *Fgfr2*.

Such instances of paralogous "substitutions" are clearer in the comparisons of our ammniote datasets to zebrafish. *fzd7a-Fzd2 col15a1b-Col3a1*, *col15a1b-Col23a1*, *col15a1b-Col2a1*. However, they also display 1:1 orthologues: *tspan7-Tspan7*, *fzd7a-Fzd2*, and *degs1-Degs1*. Similarly, zebrafish *bona fide* NC mapping to endothelial NC is supported by *itm2ba-Itm2a*, *cers2b-Cers2*, but also the paralogues *etv5b-Etv2*. High-conservation of Migratory NC states is illustrated by *snai1b-Snai1*, *mycn-Mycn* and *sox10-Sox10*; neurogenic NC/Delaminating NC by *sox9b-Sox9* and *tcf12-Tcf12*; and non-neural NC by *epcam-Epcam* and *krt8-Krt7*.

The comparison between zebrafish and chicken displayed similar properties. For example, *bona fide* NC mapped to delaminating NC was supported by *itm2ba-ITM2B* and *tspan7-TSPAN3*; *bona fide* NC mapping to multipotent NC by *ppp1caa-PPP2CA* and *id2a*-ID2; *bona fide* NC to pre-migratory NC by *msx3-MSX1*; and neurogenic NC by *fgfr4-FGFR1/3*.

We next assessed the 3-way species alignment within the same manifold (Figure 3.13). Our results indicate similarities to the 2-way alignments, including extensive mouse-biased clusters despite the more close evolutionary relantionship between chicken and mouse (Figure 3.13a). This strongly supports the idea that whereas SAMap successfully aligns transcriptomes at large evolutionary distances, overcoming the biases imposed by different batches, technologies and species remains challenging.

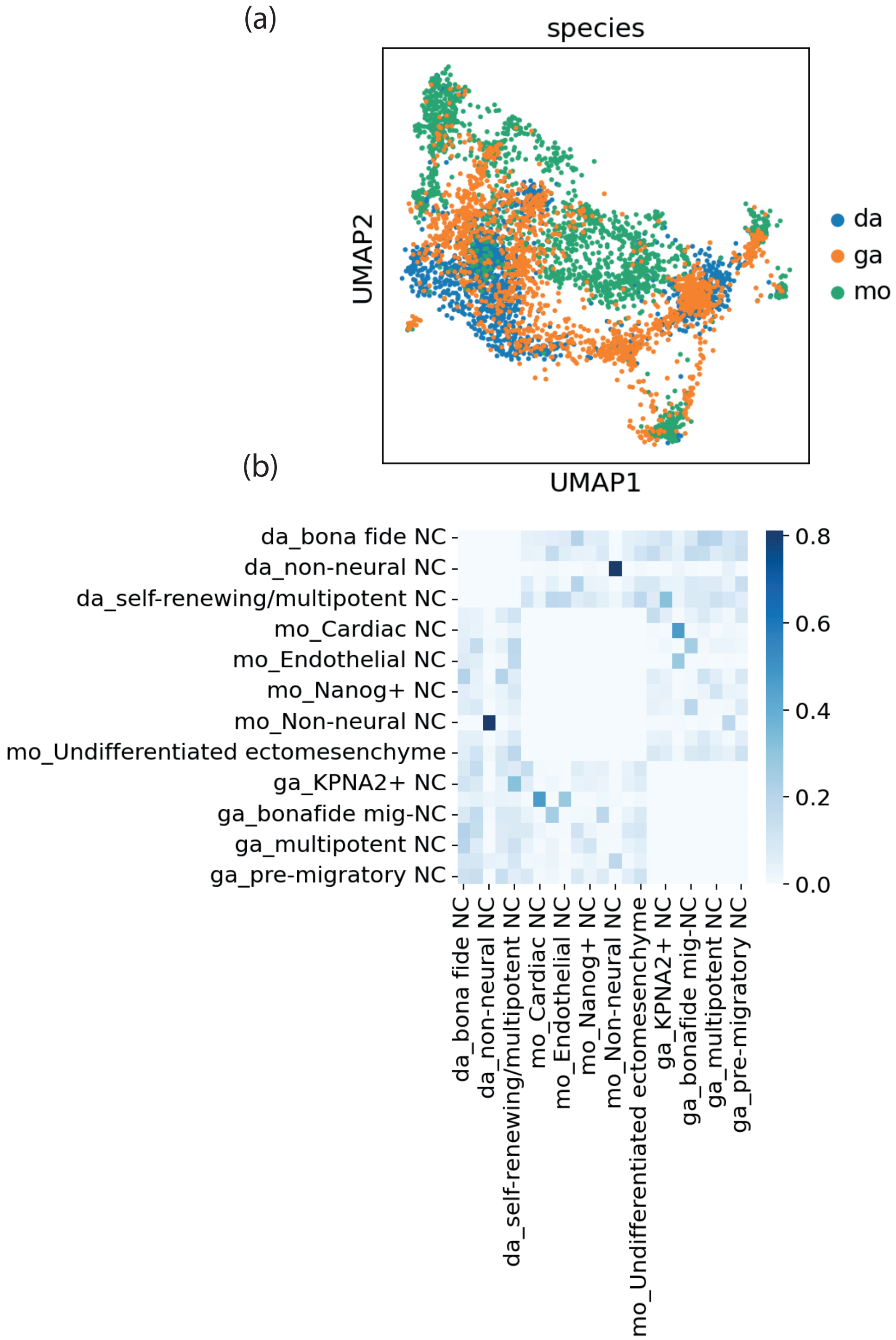

Figure 3.13 Three-way SAMap alignment of single cell profiles across species. (a) UMAP embedding coloured by species. (b) Heatmap of mapping scores between clusters from each species. mo, *Mus musculus*, ga, *Gallus gallus*, da, *Danio rerio*

Nevertheless, we could recapitulate similar properties identified in the higher-resolved 2-way species comparisons. Mapping scores (Figure 3.13b) between clusters, and the interrelationship of the multiple clusters (Figure 3.14) supports high conservation of trans environment across such large evolutionary distances, spanning the divergence of bony fishes.

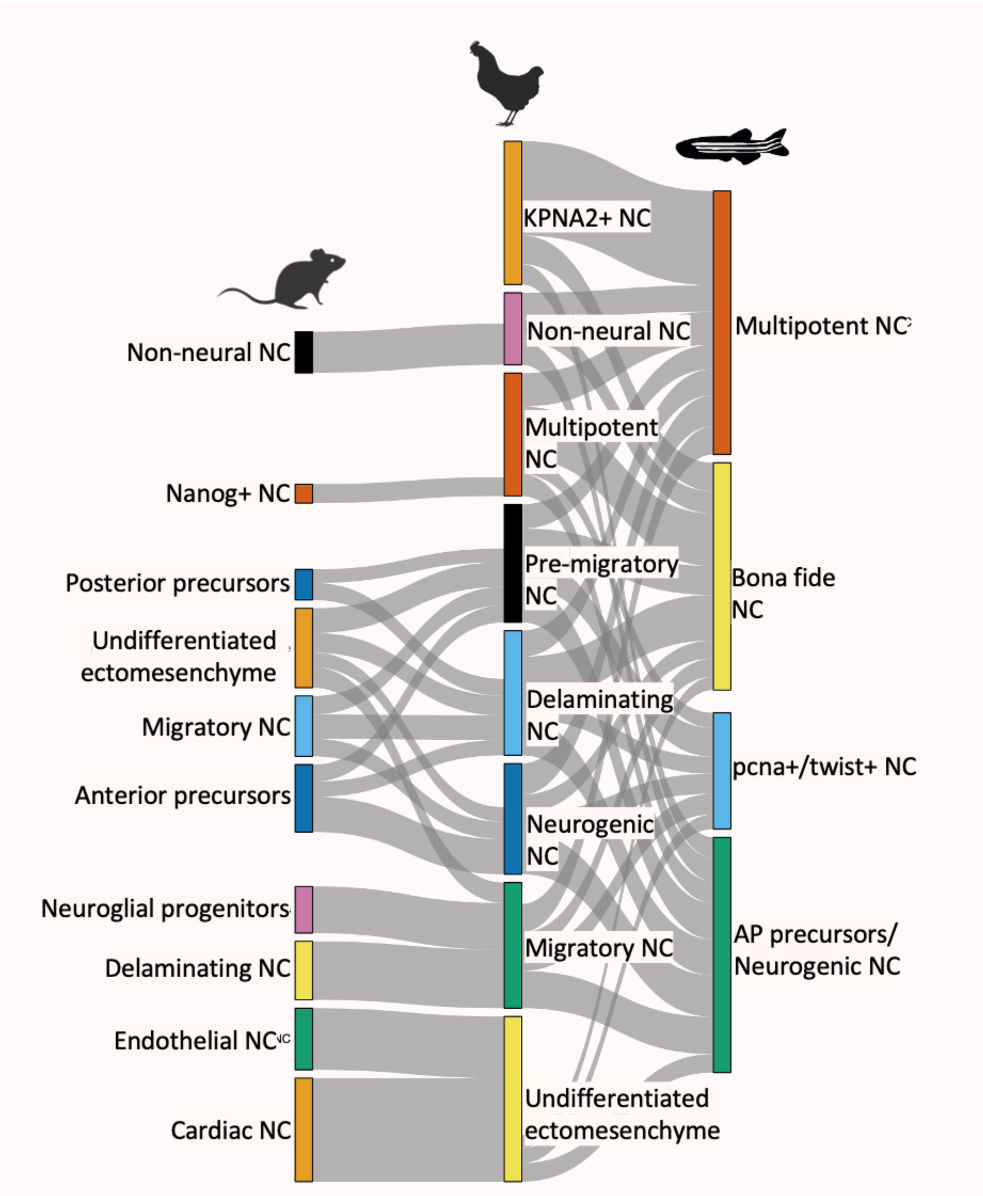

Figure 3.14 Sankey plot mapping mouse, chicken and zebrafish NC cells.

Importantly, our results indicate that similar *trans* factors are deployed in NC development, and reveals instances of "paralogous substitutions", suggesting that the *cis*-regulatory architecture controlling paralogous genes may have either been retained before duplication events or evolved *de novo*.

Moreover, our cross-species mapping illustrates the challenge of comparing developmental progenitors across large evolutionary distances. The fact that several genes are reiteratively used during cell fate decisions may cause confounding alignments between distinct cell clusters. In contrast, lineage committed cells, such as our non-neural NC clusters, are more easily mapped due to the deployment of differential markers. I propose that higher throughput datasets with deeper coverage might improve such alignments.

3.1.4 The lamprey NC at single cell resolution

Excitingly, we performed the first investigation, to our knowledge, of the marine lamprey NC at single cells resolution. After surgical collection of dorsal neural tubes from T21 *P. marinus* embryos (Figure 3.1c), cells were dissociated and scRNA-seq was performed using the 10X Genomics Chromium platform, followed by library preparation and sequencing (see Methods). Although our quality control revealed the majority of cells were indeed stressed (see Methods), likely owing to difficulty of collecting such elusive cells from this challenging research model, we were able to identify 700 high-resolution profiles of cell-types associated to mesendoderm, non-neural epithelium, cranial mesenchyme, neural tube and putative cranial NC cells (CNCCs), (Figure 3.15a).

Indeed, we identified a cluster of 100 cells expressing *bona fide* NC regulators, such as *MSX2*, *Id2*, *id3*, *Lin7b*, *Mycn/Mycn.1*, *Chd7*, *FoxD3*, *Sox9.1/Sox9*, *SOX10*, *SMAD6*, *Tfap2b*, which we proposed to represent lamprey NC. Moreover, these cells also expressed the chromatin remodeller *Chd7* and the multipotency marker *Lin7b*. We also identified three members of the zinc finger homeobox family expressed in putative lamprey NC cells, *Zeb1.1*, *Zeb2*, *Zeb2.1*, as well as *znf593*. Moreover, we identified several genes enriched in the lamprey NC, such as *SHROOM2*, which encodes a protein involved in cell contractility, *GEMIN2/5* (encoding Survival Of Motor Neuron Protein Interacting Protein 2 and 5). Identification of expression of a novel annotated transcript, PMZ-0041112, further illustrates expansion of the lamprey NC toolkit in a manner likely divergent from jawed vertebrates.

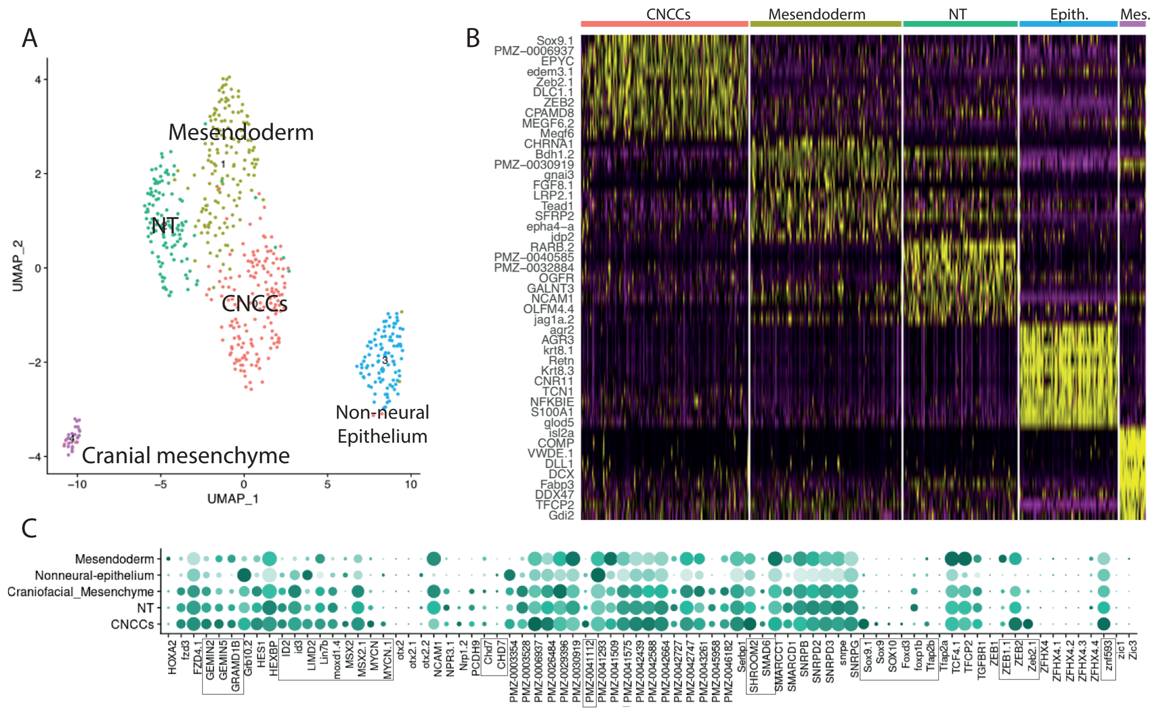

Figure 3.15 Profiling the marine lamprey NC at single cell resolution. (a) Annotated UMAP embedding of dorsal neural tubes dissected from T21 marine lampreys. (b) Heatmap showing hierarchical clustering of top marker genes for each cluster. (c) Dotplot of marker genes. Boxed are genes displaying differential enrichment in the lamprey NC.

Although further work in needed to improve the dissociation step and increase the throughput and depth of our single cell libraries, these results illustrate divergence of the transcriptional network of lamprey NC cells to that of zebrafish, avian and murine NC. Nevertheless, our dataset provides an excellent database of predicted gene expression catalogues, which can guide the functional dissection and experimental validation in the lamprey NC *in vivo*. In our future work, we will aim to extensively collect tissue enriched for pre-migratory and migratory lamprey NC and perform single nuclei RNA-seq, as manner to circumvent the clear drop-out ratio we have experienced using 10X scRNA-seq.

3.2 Transcriptional and epigenomic dynamics during zebrafish NC ontogeny

Although our scRNA-seq profiling approach provides insights at high-resolution to the zebrafish NC, it does not cover the multiple steps involved in the induction, specification, delamination and migration of NC, encompassing the early NC ontogeny. We thus used a wealth of transcriptomes at multiple stages of NC development and their cognate non-NC controls previously published from our laboratory [60], again using the foxd3 gene trap line. This dataset was collected at 75% epiboly, 1-3ss, 5-7ss and 14-18ss (Figure 3.16), representing late gastrulation stage, early NC induction, NC specification and migration, respectively. Importantly, this experimental strategy also parallels our approach used to profile the avian NC, with embryonic stages largely representing the ones we used in the previous Chapter.

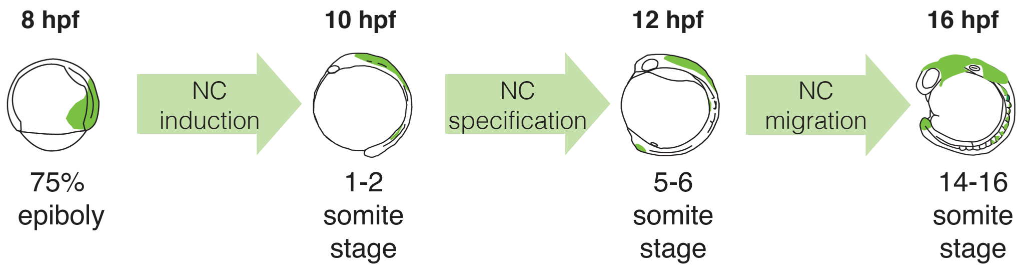

Figure 3.16 Zebrafish embryo stages examined in this study. 75% epiboly, a gastrulation stage during which embryonic shield and hypoblast are formed. 1-3ss and 5-6ss, induced and specified pre-migratory neural crest (NC), respectively. 14-16ss, migratory NC. Foxd3 expression is labelled in green; hpf - hours post fertilization. Adapted from [60].

Again, we used Weighted Gene Co-expression Network Analysis (WGCNA) [222, 223] to cluster the expressed genes into distinct subgroups (Figure 3.17).

Annotation of known zebrafish TFs enabled us to assign these subgroups to previously proposed NC GRN modules governing critical stages of early NC ontogeny. Early NC clusters (Figure 3.17a-b) deployed a transcriptional programme consisting of known canonical regulators, such as *pax3a*, *irx2a*, *twist2*, *snai1b*, *tcf7*, *neurod2*, *LHX3/4*, *pou4f3*; a *bona fide* NC cluster (Figure 3.17e) displayed expression of transcriptional activators involved in maintaining multipotency and activating EMT (*sox5*, *sox9b*, *sox10*, *sox8b*, *mafk*). We also recovered a NC induction module (Figure 3.17d) marked by expression of *tfap2a*, *tfap2e*, *lhx8b*, *pou6f1*, *prdm1a*, factors acting upstream of *foxd3/sox10*; a module involved in early NC regionalisation (Figure 3.17c) displaying expression of *mef2d*, *hoxd3a*, *hox13a*, *gli3*, *pax7b*, *gbx2*). Two clusters involved in migratory NC were also identified (Figure 3.17f, i). The first display factors involved in ectomesenchyme differentiation (*sox4a*, *irx4a*, *tcf20*, *rxraa*, *alx4b*, *ets1*); whereas the second displays factors involved in vagal NC identity (*hoxb3a*, *hoxa2b*, *dlx1a*, *id2a*, *alx1*, *alx4a*, *prr1a*, *foxn1*, *myod1*, *nr2f2*). We also recovered a module shared with gastrulation stages (Figure 3.17g) displaying expression of *tcf3a*, *ppp1r113bb*, *hesx1*; remaining clusters displayed enrichment in non-NC (Figure 3.17h; Appendix 7.10).

Again, we propose that such cluster assignment reflecting co-expression with known NC genes provides an excellent framework for which the role of previously uncharacterised and novel factors can be inferred. Taken together, these results suggest that our WGCNA clustering parsed NC gene expression dynamics and in return, revealed factors not previously described in NC development, thus expanding the depth and complexity of the zebrafish NC GRN.

We next sought to interrogate the *cis*-regulatory landscapes controlling the deployment of these complex NC programmes. Again, FAC sorted foxd3+ cells were collected 3-4ss, 7-10ss and 14-18ss, representing early NC induction, NC specification and migration, respectively. For the earliest stage we used the 75% epiboly

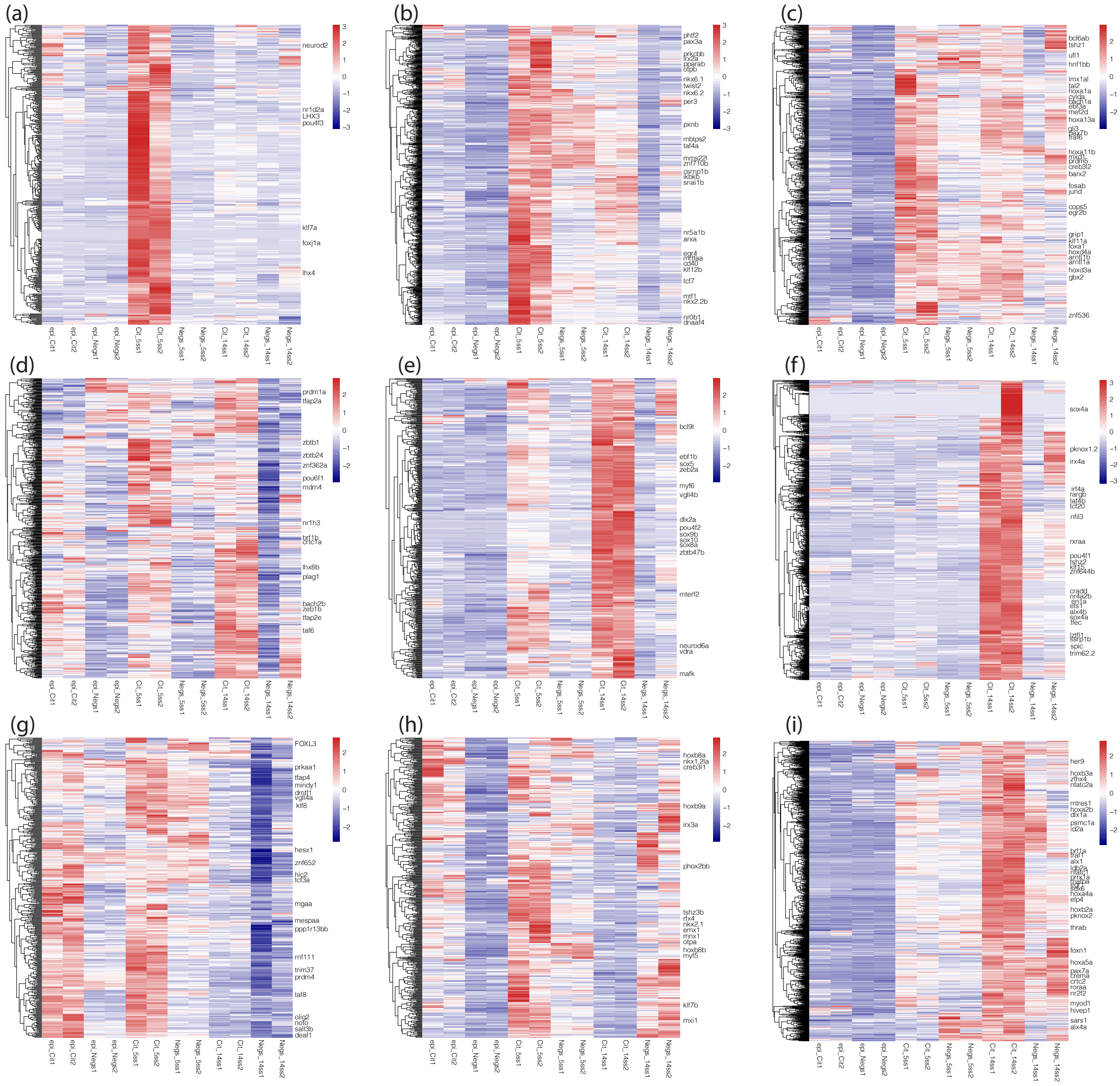

Figure 3.17 WGCNA analysis reveals co-expressed TFs during the zebrafish NC ontogeny.(**a-b**) Early, pre-migratory NC clusters; (**c**) NC regionalisation module; (**d**) NC induction module; (**e**) *bona fide* NC module; (**f**) migratory NC module; (**g**) a module shared with gastrulation stages, and (**h**) module enriched for genes expressed in non-NC cells. Two replicates per stage for NC and non-NC cells are shown on the x axis.

data previously reported in [60]. Our results recapitulate similar *cis*-regulatory dynamics that we previously reported for avian NC. For example, at the *snai1a* locus (Figure 3.18), a paralogue of *Snai2*, we identified regions of accessible chromatin displaying different levels of accessibility; however, most peaks were largely shared with non-NC cells. In fact, a proximal peak upstream of the *snai1a* promoter specifically accessible in NC cells suggests this might be a enh-13-like *snai1a* NC enhancer. This results highlight the high-resolution of our ATAC-seq profiles from NC cells.

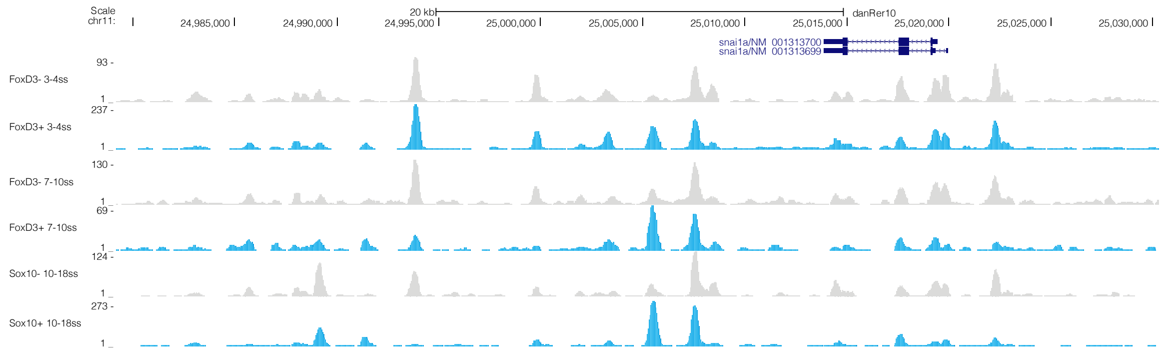

Figure 3.18 Profiling chromatin dynamics during early NC ontogeny in zebrafish. Genome browser tracks showing the *sna1a* locus. In grey, non-NC cells. Blue, NC cells at different stages of development.

To identify patterns of chromatin accessibility dynamics in a manner consistent with our findings in the chicken embryo, we next employed k-means clustering. We used samples at stages largely corresponding to the ones from the avian NC: 75% epiboly stage matching the HH4 sample; 3-4ss in zebrafish matching 5-6ss to chicken (pre-migratory sample); and 7-10ss in zebrafish, matching the 8-10ss stage in the chicken, representing NC cells at early delamination stages. We also included the stage-matched cognate non-NC controls. K-means clustering resulted in seven clusters (Figure 3.19a). They displayed chromatin dynamics similar to the ones that were recovered in chicken, and therefore we annotated them in a manner consistent with their cognates in chicken. k-Cl-4 display broad signal, accessible at all stages and cell-types, with around 25% of them overlapping annotated promoters. k-Cl-1 and k-Cl-1-like displayed patterns of accessibility enriched in NC and non-NC cells, indicative of pleiotropic roles in early development. k-Cl-3 displayed accessibility specifically in NC cells, whereas k-Cl-5 and k-Cl-6 displayed activity shared with 75% epiboly stages. We next analysed the *sox10* locus 3.19b), which identified putative enhancer elements in a manner consistent with their spatial organisation of the chicken *Sox10* "super-enhancer". This locus is enriched for *k*-Cl-3 (n=4) and *k*-Cl-1/*k*-Cl-1-like (n=3) elements, similarly to our previous results in chicken.

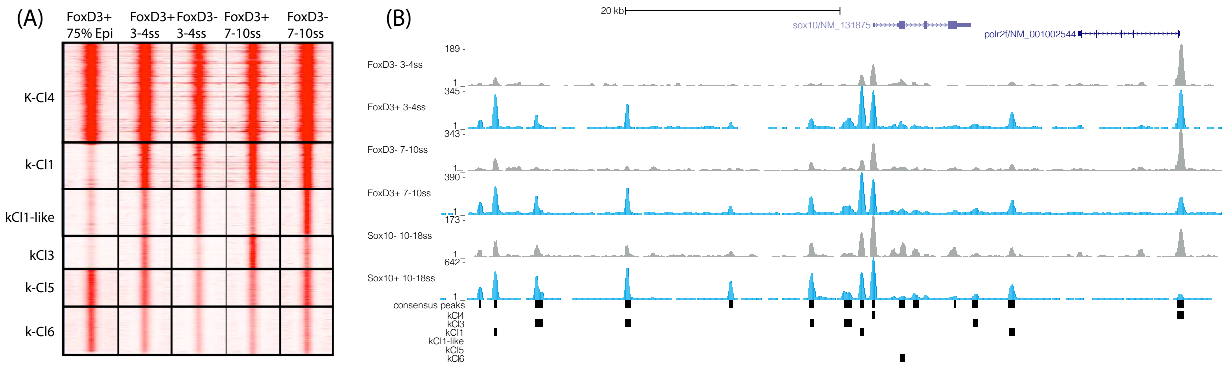

Figure 3.19 *K*-means clustering of chromatin accessibility during early zebrafish NC development. (**a**) Read-count-normalized ATAC-seq samples obtained from NC and non-NC control cells were reiteratively assigned into cluster of elements (k=10) showing similar patterns of chromatin accessibility dynamics genome-wide using *k*-means algorithm. Only the top 7 clusters are shown. (**b**) The zebrafish *sox10* locus; putative enhancers belonging to each cluster are annotated.

We next used LiftOver tool from the UCSC Genome Browser Interface to identify cross-matching regions in between the chicken and zebrafish genomes, restricting our predictions to syntenic regions. This approach enable us, to some extent, to transfer the enhancer annotations from the chicken genome into zebrafish, thus allowing us to predict orthologous enhancers (Figures 3.20, 3.22). For the *sox10* locus (Figure 3.20a-b), we could identify orthology and similar spatial organisation in between both species, suggesting functional conservation of the *sox10* *cis*-regulatory landscape across such large evolutionary distance.

Interestingly, the two downstream enhancers of *Sox10*, Sox10E1 and Sox10E2, which display differential axial-specific activities (i.e., cranial *versus* trunk, respectively [230, 231]) map to a single region in the zebrafish genome (Figure 3.21). This result led us to hypothesise that gene regulatory elaborations via duplications or gains of novel enhancers could have underlined additional axial regionalisation of the amniote NC, warranting further investigation in the future.

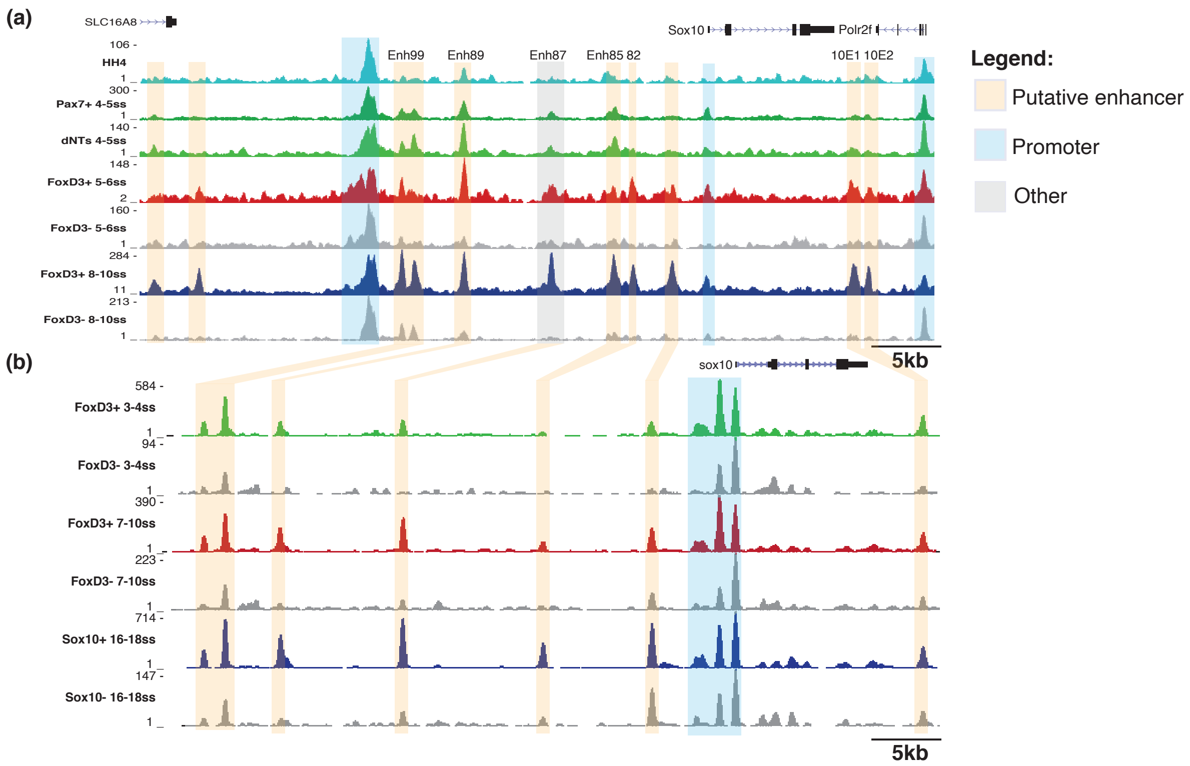

Figure 3.20 LiftOver of enhancer genomic coordinates in between the chicken and zebrafish genomes. Genome browser tracks showing ATAC-seq profiles at different stages of NC development at (**a**) the chicken *Sox10* and (**b**) zebrafish *sox10*. Putative enhancers matching between genomes are highlighted. Genome browser tracks in green, red and dark blue are ATAC-seq data from NC cells at pre-migratory and migratory stages; light blue, ATAC-seq from gastrulation stages embryos; grey, ATAC-seq data from non-NC FAC sorted cells.

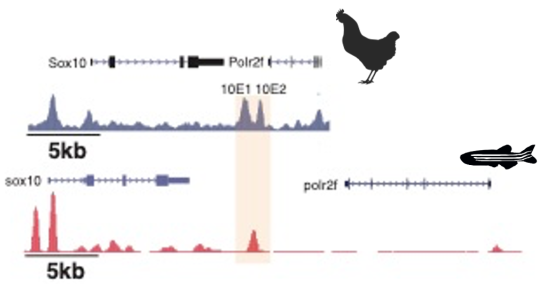

Figure 3.21 Alignment of two enhancers downstream of chicken *Sox10* map to a single putative *cis*-regulatory element in the zebrafish genome wthin a syntenic region. Genome browser tracks showing NC ATAC-seq profiles at the chicken *Sox10* and zebrafish *sox10*. The highlight indicate the mapping of two annotated enhancer elements in the chicken genome (10E1 and 10E2) to a single putative accessible element in zebrafish. Blue, chicken NC ATAC-seq; red, zebrafish NC ATAC-seq.

Interestingly, annotation of enhancer elements within the *Sox5* locus (Figure 3.22a-b) suggest a large expansion enhancers in the chicken *Sox5*, as indicated by the more compact organisation of the *sox5* locus in zebrafish. This results are consistent with the idea that enhancer turn over is common between species [120, 212]. Nevertheless, two clusters of super-enhancer-like elements (one intronic and one intergenic positioned upstream) seem to be conserved between the two species.

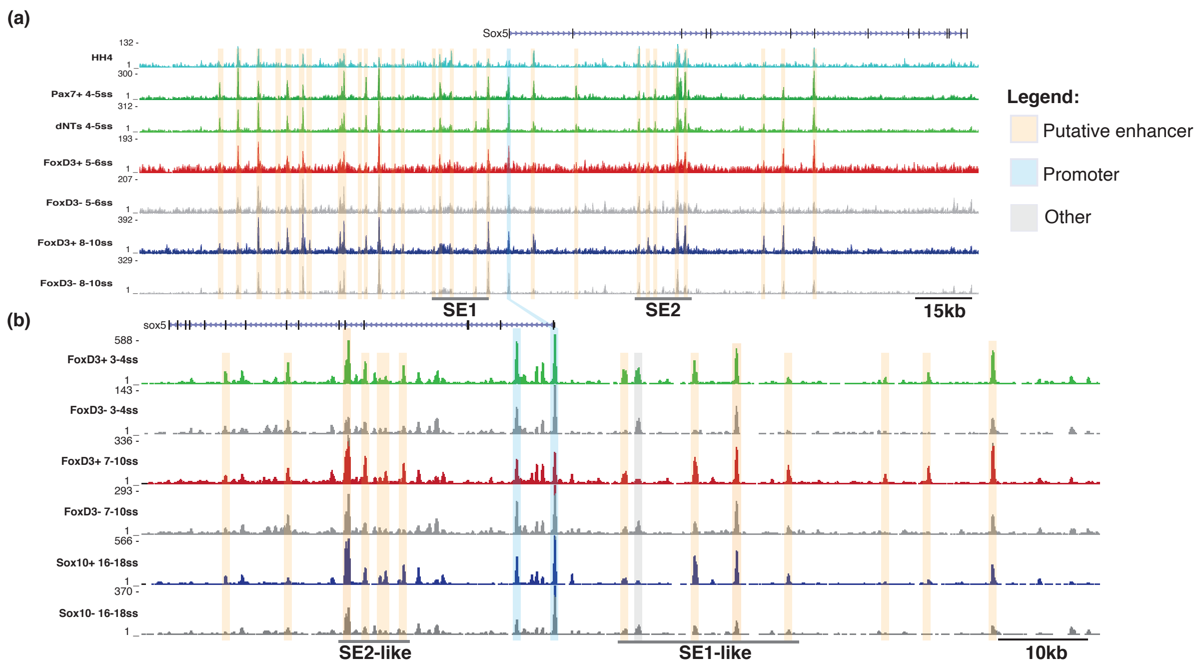

Figure 3.22 LiftOver of enhancer genomic coordinates in between the chicken and zebrafish genomes at the *Sox5* locus. Genome browser tracks showing ATAC-seq profiles at different stages of NC development at (**a**) the chicken *Sox5* and (**b**) zebrafish *sox5*. Putative enhancers matching between genomes are highlighted. Genome browser tracks in green, red and dark blue are ATAC-seq data from NC cells at pre-migratory and migratory stages; light blue, ATAC-seq from gastrulation stages embryos; grey, ATAC-seq data from non-NC FAC sorted cells. SE, super-enhancer-like regions.

Finally, we conducted motif enrichment analysis on our zebrafish *k*-Clusters (Figure 3.23). Our results indicate enrichment of motifs associated to NC factors in *k*-Cl-1 and *k*-Cl-3, such as Sox10, Sox9, TFAP2a. Interestingly, k-cl-1 again displayed motifs reminiscent of a role in neurogenesis, such as Otx2, Sox2 and Sox3.

Promoter-associated TFBS were identified in *k*-Cl-4 (e.g., NFY, YY1), whereas *k*-Cl-5 and *k*-Cl-6 displayed enrichment of upstream factors associated with gastrulation, such as Otx2, Sox2, Nanog and FoxA2.

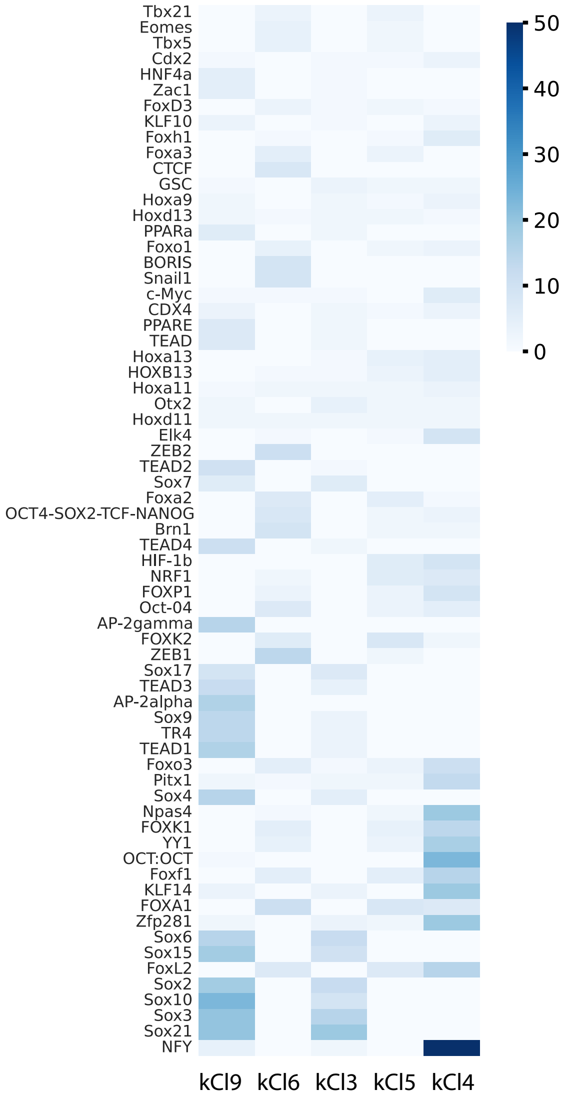

Figure 3.23 TF binding motifs enriched in zebrafish *k*-means clusters elements. The scale shows -log10 transformed P-values, calculated with binomial distributions.

Taken together, our results indicate the validity of our approach to identify enhancers across species and reveal a large cohort of enhancer elements displayed similar dynamics to the ones characterised in the chicken NC.

3.3 Epigenomic dynamics during lamprey NC ontogeny

To assess the NC *cis*-regulatory landscape of the marine lamprey, Dr. Dorit Hockman optimised and performed bulk ATAC-seq in dorsal neural tubes of the marine lamprey *P. marinus* at two stages of NC evolution: early specification (T20) and early delamination (T21) [169]. Dr. Dorit Hockman also dissected whole heads at T23, which were also used for epigenomic profiling, representing tissue enriched for migratory NC (Figure 3.24a). We mapped ATAC-seq profiles to the marine lamprey germline genome assembly [167]. To reduce background noise, we focused our analyses on peaks associated with NC GRN genes, by annotating peaks to promoters, as well as intergenic and intronic clusters, which were then restricted to peaks that were associated with genes expressed at T21. We further restricted our analysis to intergenic and intronic regions as a manner to identify distal *cis*-regulatory elements.

K-means clustering of the ATAC-seq signal over consensus peaksets (8998 intergenic; 17,908 intronic), performed by Dr. Dorit Hockman, revealed dynamics of chromatin accessibility over the course of development (Figure 3.24). Similar to our previous findings in the chicken and zebrafish NC, we found clusters indicative of *cis*-regulatory activity. kCl-1 displayed broad signal and accessibility across all developmental stages assessed here, suggesting a large proportion of these elements may function as alternative promoters. In particular, two large clusters (intergenic cluster 5; intronic cluster 4) displayed increased accessibility at T21 and T23 compared with T20 (Figure 3.24d). These clusters contained elements associated with known NC GRN TFs, SoxE1 (intergenic cluster 5) and Tfap2B (intronic cluster 4), suggesting they might contain enhancer elements associated to lamprey NC EMT.

We next used TF-binding site motif analysis to further interrogate the ATAC-seq k-means clusters. Similar to our recent ATAC-seq analyses in zebrafish and chicken NC, several intergenic clusters, including cluster 5, were enriched for Sox,

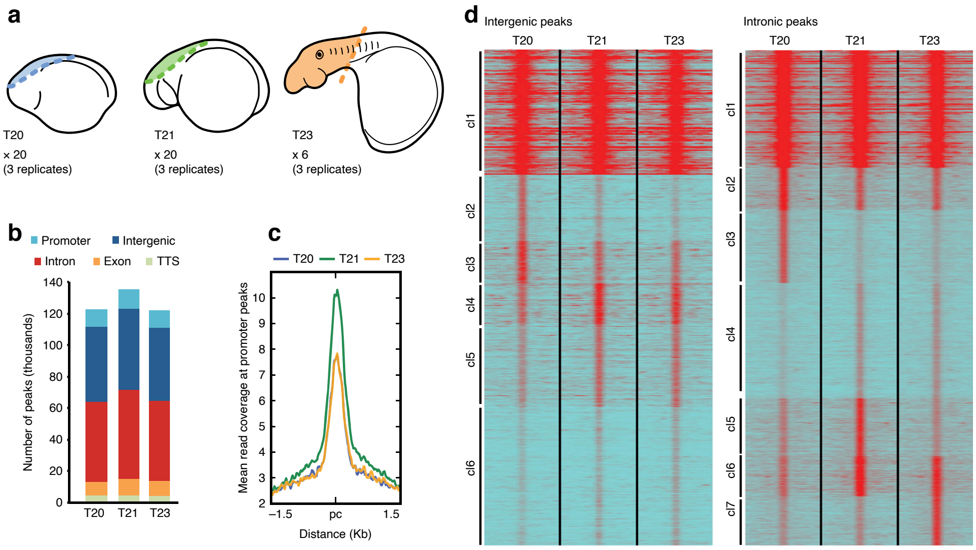

Figure 3.24 Profiling of chromatin dynamics in the marine lamprey developing NC. (a). Schematics indicating the region of dorsal neural tube or head dissected from T20, T21 and T23 lamprey embryos for ATAC-seq and the number of biologically independent samples analysed. (b) Functional annotation of our ATAC-seq peaksets for all stages. (c) Mean ATAC-seq read coverage map at each stage, showing higher read coverage at T21. (d) Heatmaps depicting *k*-means linear enrichment clustering of ATAC-seq reads at all stages across consensus intergenic and intronic peaksets. Boxes indicated the large "EMT" clusters that show enriched signal at T21 and T23. cl, cluster; Pc, peak centre. Contributed by Dr. Dorit Hockman and available at [169].

Tfap2 and Tead motifs (Figure 3.25a). GRHL2, Pax, Ets1 and Fox motifs were also enriched. Intronic cluster 4 displayed similar enrichment for Sox and Pax motifs, as well as Fox sites, whereas GRHL2 and REST-NRSF motifs showed the highest enrichment in this cluster (Figure 3.25b). The presence of motifs for key NC specification TFs (Sox, Fox, Tfap, Ets1, Pax families), as well those that play a role in EMT (GRHL236), suggested that these clusters harbour *cis*-regulatory elements that provide connections between NC GRN players. Enrichment of CTCF-binding sites in all intergenic clusters further suggests these peaks may represent putative chromatin boundary elements.

Moreover, *k*-Cluster 1 corresponding to intergenic and intronic peaksets had a distinct TF-binding site profile from the other clusters, displaying enrichment of factors associated to promoter activity (Figure 3.25c). This result further indicated that these elements may display promoter-like activity.

Finally, we used two-way combinatorial analysis to test for evidence of TF co-binding at putative cis-regulatory elements. We focused on intergenic cluster 5 and selected 18 TFs that displayed enriched motifs and enriched gene expression at T21. The motifs of six TFs (Smad2, Sox10, TFAP2a, Pit1, Bcl-6 and Pbx3) were significantly enriched in pair-wise analyses ($p < 0.05$, two-tailed Chi-squared test), (Figure 3.25d). Interestingly, we found Smad2 sites were enriched in combination with all the other five motifs. This suggests that the combinatorial activity of TGFb-signalling with canonical NC specification factors at enhancers is a conserved property of the NC GRN.

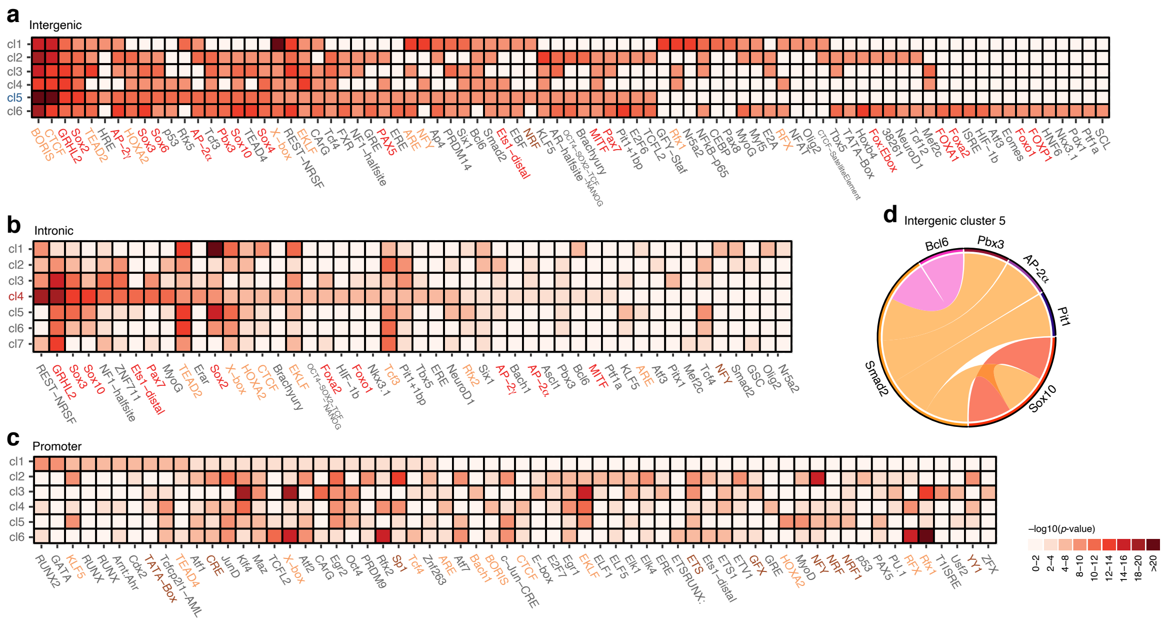

Figure 3.25 Motif enrichment in lamprey NC *k*-means clusters. (a) intergenic (b), intronic and (c) promoter *k*-means clusters. Canonical regulator motifs of NC development are highlighted in red. Similar motifs shared between intergenic and intronic cluster 1 and promoter clusters are highlighted in orange. Canonical promoter motifs are highlighted in brown. (d) TFs that were significantly enriched in pair-wise co-binding analyses conducted on intergenic *k*-means cluster 5.

Collectively, these results indicate that a similar *cis*-regulatory architecture is deployed during the ontogeny of the early NC in lampreys. We identified broadly accessible elements, likely representing housekeeping and developmental promoters, boundary elements and other *cis*-regulatory elements. Additionally, motif content and differential chromatin accessibility dynamics suggests that lampreys employ a similar *cis*-regulatory architecture to that of other vertebrates, relying on developmentally-regulated enhancer elements, that may display a conserved *cis*-regulatory code. Therefore, similar evolutionary pressures might shape the *cis*-regulatory evolution of lampreys.

3.3.1 *In vivo* assessment of *cis*-regulatory activity

Our results indicate similar dynamics, motif content, and enhancer-promoter communications of developmental enhancers across large evolutionary distances, suggesting functional constraints on the evolution of NC gene regulation. To validate enhancer activity across species, we first performed an extensive screen in three loci displaying Sox10 binding, enrichment of H3K27ac, and enrichment for putative pleiotropic enhancers belonging to *k*-Cl-1 and *k*-cl-4, as well as NC-specific *k*-Cl-3 elements in the chicken genome. Although we have previously characterised enhancer activities *in vivo*, our analyses were limited to the early steps of NC specification and delamination (6-10ss), largely precluding the identification of later-acting enhancers, as well as enhancers that display pleiotropic activities. We thus extensively analysed enhancer activities spanning early NC specification (HH7, 5-6ss) to colonisation (HH16-18).

We first analysed the Sox5 locus, a large domain spanning 400kb. We identified and cloned 12 putative enhancers upstream of *Sox5* and 9 intronic elements (Figure 3.26a). However, only two enhancers displayed enhancer activity in the developing NC: the *k*-Cl-4 element Sox5E8 and the *k*-Cl-9 element Sox5E9 (Figure 3.26a), both showing strong Sox10 binding. Sox5E8 was active in late NC cells colonising the craniofacial ectomesenchyme surrounding the eyes, as well as the trunk NC giving rise to DRG. Similarly, Sox5E9 displayed activity in the cranial region, but also in the vagal NC, which give rise to the enteric nervous sytem and contributes to cardiac development. Other elements, such as Sox5E27 and Sox5E33, displayed activity in the midbrain and tail bud, respectively.

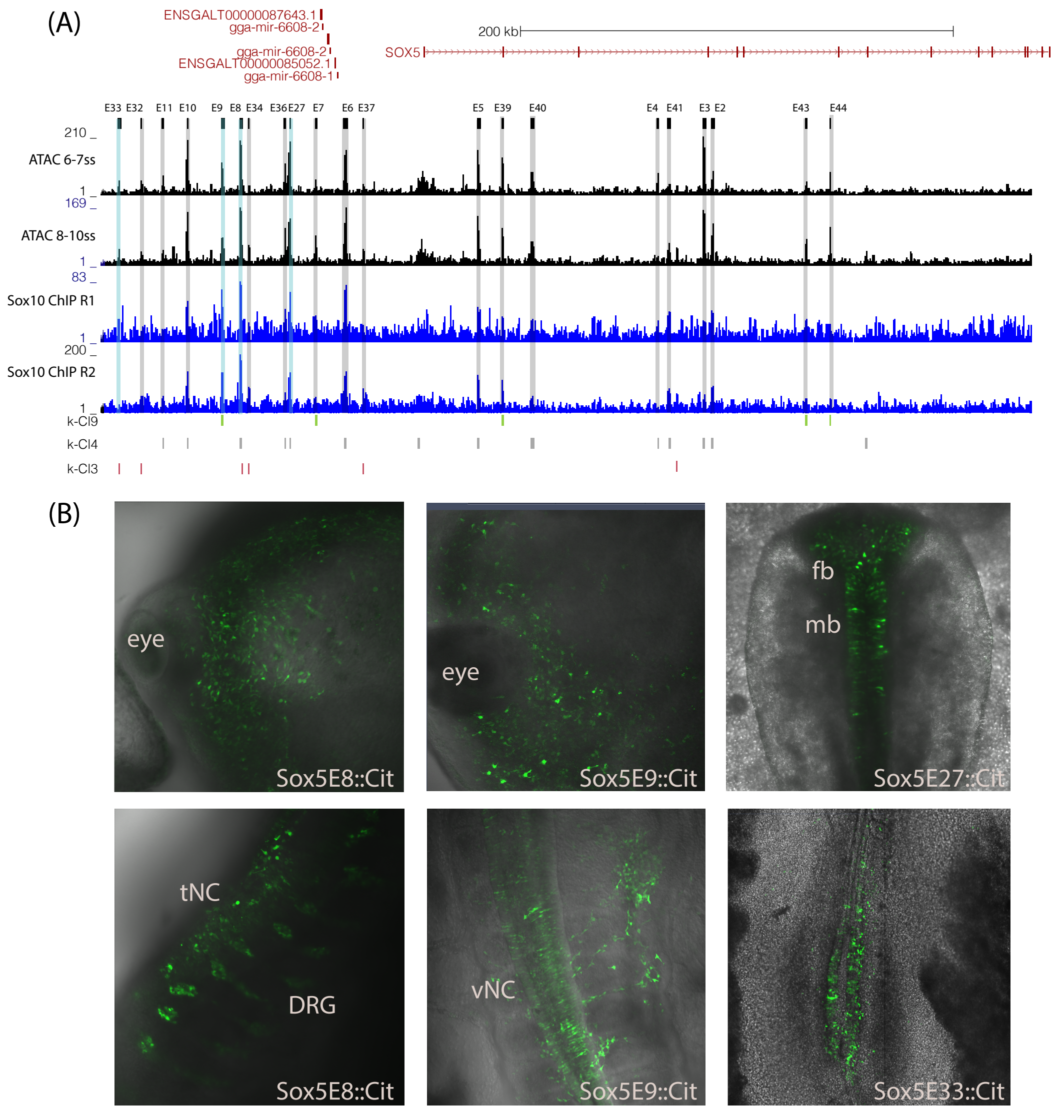

Figure 3.26 Dissection of the Sox5 locus. (a) Genome browser tracks of the Sox5 locus. In black, ATAC-seq of NC cells at 6-7ss and 8-10ss, respectively. In blue, Sox10 ChIP-seq profiles. R1, replicate one; R2, replicate 2. (b) *In vivo* validation of *cis*-regulatory elements in chicken embryos.

Next, we screened the Fzd3 locus, encompassing a 150kb genomic region (Figure

3.27a). 11 elements were identified and cloned; however, the majority displayed pleiotropic activity in non-NC cells, such as Fzd3E6, active in the hindbrain and anterior midbrain (Figure 3.27b). A single NC element, Fzd3E10, was identified, displaying activity in the cranial NC at migratory stages (Figure 3.27c).

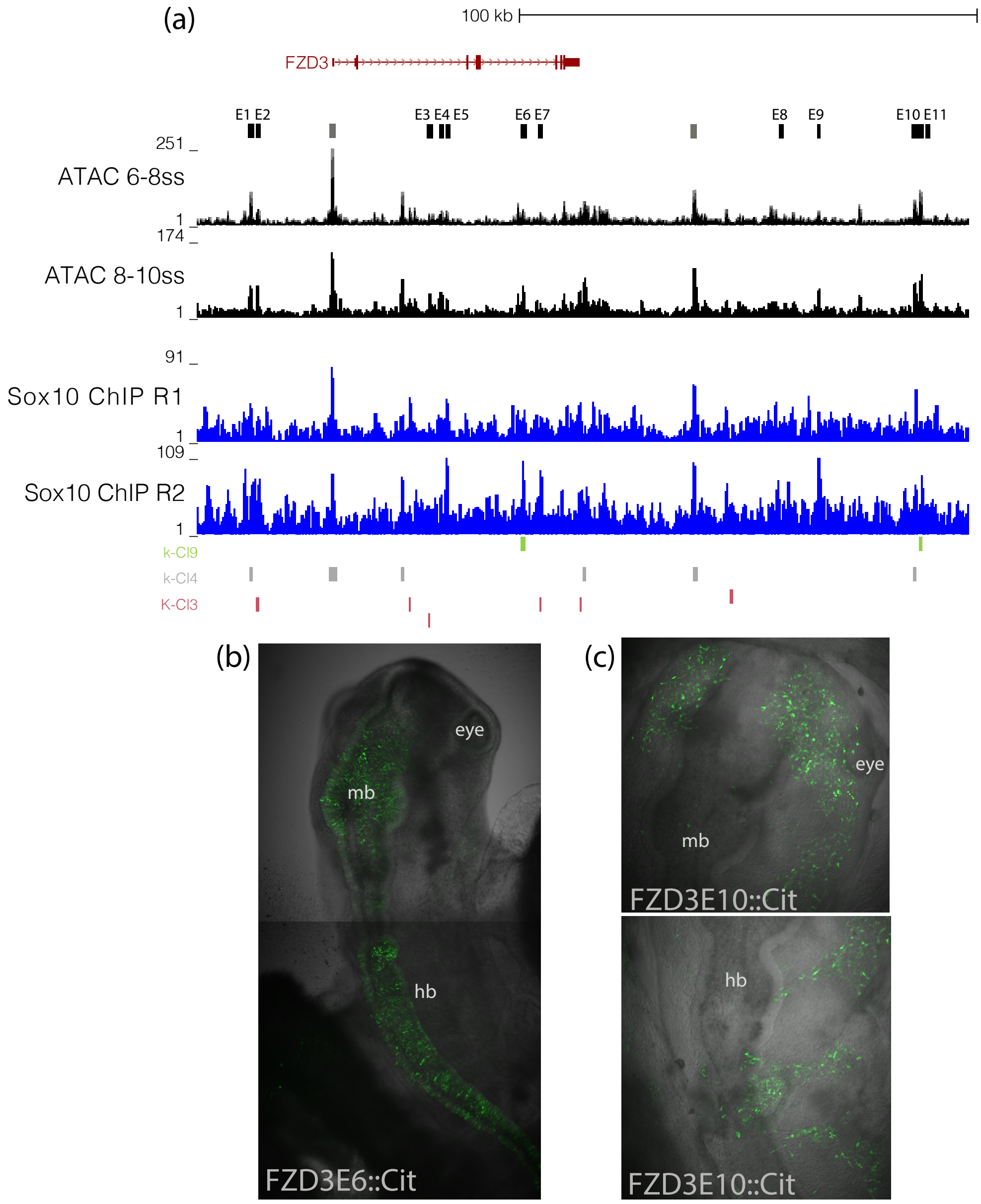

Figure 3.27 Dissection of the *Fzd3* locus.(a) Genome browser tracks of the Fzd3 locus. In black, ATAC-seq of NC cells at 6-7ss and 8-10ss, respectively. In blue, Sox10 ChIP-seq profiles. R1, replicate one; R2, replicate 2. (b) *In vivo* validation of *cis*-regulatory elements in chicken embryos.

Finally, we identified a *k*-Cl-3, Sox10-bound enhancer located 200kb upstream of *Chd7*, at the 3' UTR of the *Carbonic Anhydrase 8* (*CA8*) locus (Figure 3.28a), a gene involved in cerebellar ataxia (OMIM 114815). In contrast to the previous NC elements, this enhancer is active at pharyngula stages (Figure 3.28b) in trunk NC cells, but is also shared with neurogenic placodes, hindbrain and midbrain cells (Figure 3.28c), thus displaying pleiotropic activity in the developing nervous systems.

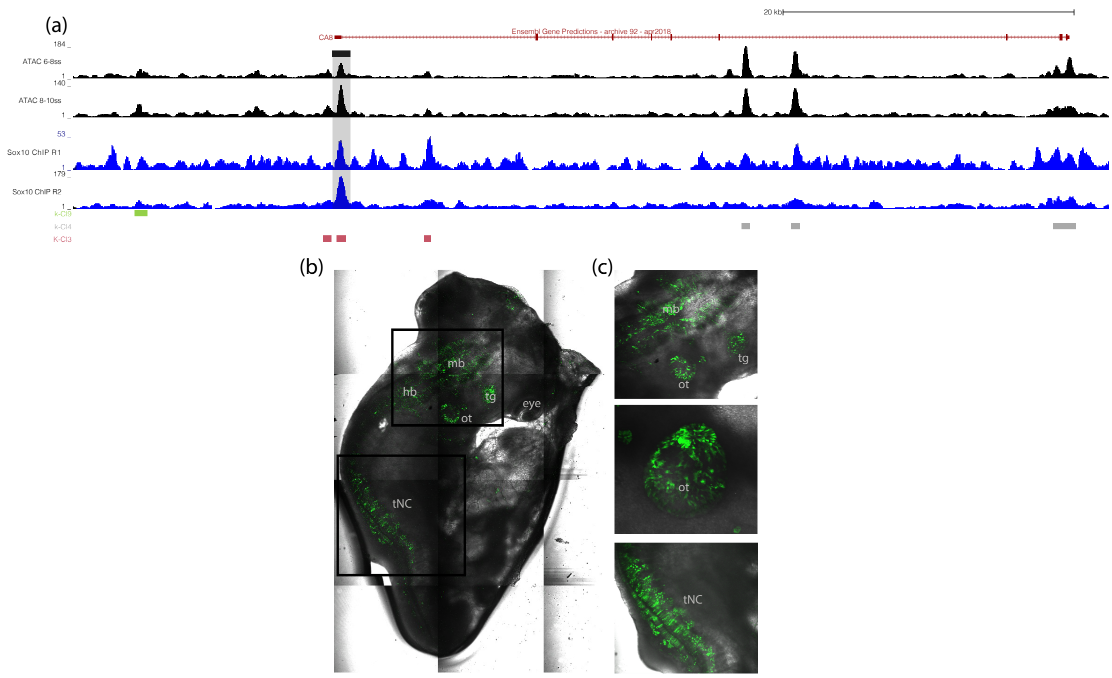

Figure 3.28 Dissection of the *CA8* locus.a) Genome browser tracks of the Fzd3 locus. In black, ATAC-seq of NC cells at 6-7ss and 8-10ss, respectively. In blue, Sox10 ChIP-seq profiles. R1, replicate one; R2, replicate 2. (b-c) *In vivo* validation of *cis*-regulatory elements in chicken embryos. hb, hindbrain; mb-midbrain; ot - otic placode; tNC - trunk NC.

Similarly, a previous enhancer screen in our laboratory, aimed at identifying *cis*-regulatory regions involved in enteric nervous system development [58] identified two Sox10-bound elements in loci involved in neurogenesis: *Ret* and *Ascl-1*. Interestingly, these elements also displayed activity at pharyngula stages, and were shared with rhombomeres and the trigeminal ganglia (Figure 3.29a-b).

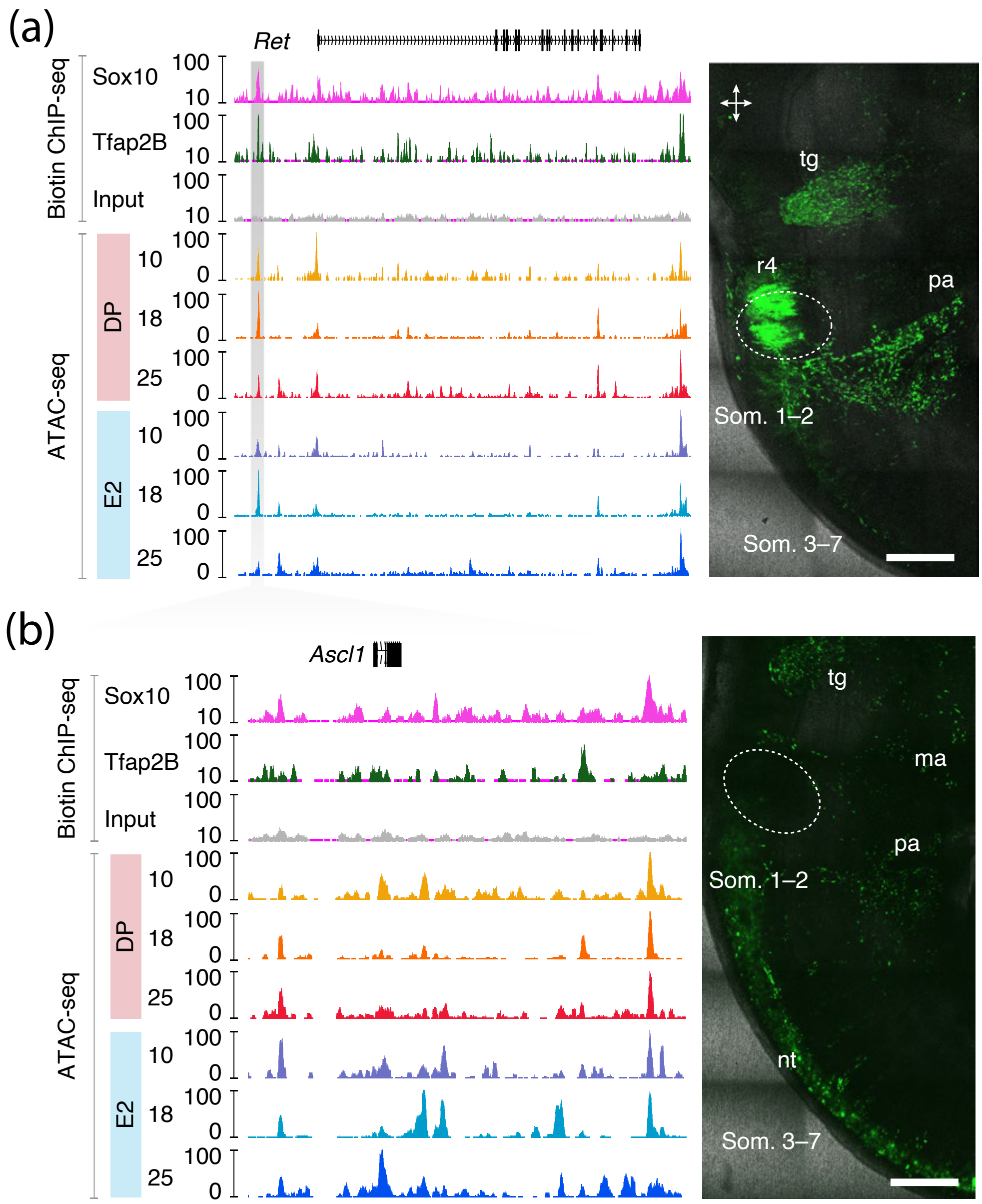

Figure 3.29 The *cis*-regulatory landscape of *Ret* and *Ascl-1*. (**a**) The *Ret* locus, showing ChIP-seq profiles and developmental ATAC-seq profiles at HH10, HH18 and HH25. Boxed is the enhancer element validated *in vivo*. (**b**) The *Ascl-1* locus, showing ChIP-seq profiles and developmental ATAC-seq profiles at HH10, HH18 and HH25. Boxed is the enhancer element validated *in vivo* DP, double-positive EdnrB+, FoxD+ vagal NC cells. Adapted from [58].

Surprisingly, despite extensive screening of several NC regulators (Figure 3.24, Figure 3.25, Figure 7.4), we found NC enhancers involved in EMT and ectomesenchyme development to be specifically active in the NC. This contrasts with previous proposals, suggesting NC role in ectomesenchyme development evolved through the co-option of mesoderm-associated circuitries. In fact, enhancer pleiotropy was only identified in NC enhancers associated to neurogenic roles, as illustrated by CA8E1, RetE1 and Ascl1E1. This suggests that the early neurogenic NC evolved through pleiotropic changes in ancestral chordate enhancers, which were already involved in regulating the development of nervous systems.

Finally, to test for functional conservation, we employed cross-species enhancer reporter assays. We tested three chicken enhancers, the pleiotropic enhancer Sox10E99 and Sox10E99a, and the NC-restricted Sox10E84 in lamprey embryos (Appendix Figure 7.11). Excitingly, we could recapitulate enhancer activity driven by pleiotropic chicken enhancer elements in the lamprey embryo. Sox10E99a showed activity in the pre-migratory NC, whereas Sox10E99 displayed activity in the migrating NC. However, the NC-restricted enhancer Sox10E84 did not display any activity. These results are consistent with the idea that enhancer pleiotropy might functionally constrain enhancer evolution, thus explaining the remarkable conservation of enhancer activities at such large evolutionary scales. We posit that NC-restricted enhancers might have evolved more recently, and thus display weaker yet species-specific differences related to the diversification of the NC within vertebrates.

3.4 Discussion

This chapter of my DPhil thesis has revealed high conservation of the transcriptional programmes deployed during NC ontogeny at unprecedented resolution. To demonstrate that, we have employed three scRNA-seq datasets: the murine Wnt1+ NC (Figure 3.2b), previously reported by [67], and the foxd3-expressing avian (Figure 3.2a) and zebrafish NC (Figures 3.7, 3.8, 3.9, 3.10). We then used a novel approach, implemented in SAMap [196], which enables the alignment of multiple transcriptomes across different species within the same lower dimensional space. It improves on previous approaches which restricted the query to 1:1 orthologues by taking into account the evolutionary divergence of homologous gene pairs displaying similar levels of expression, and reiteratively assigning clusters to their best-matching counterpart. Using this approach, we were able to identify several subpopulations of NC cells displaying conserved transcriptional states across these three species; however, the fact that these cells represent transcriptionally-biased, but not lineage commitment populations, preclude 1:1 matches across these species. This was revealed, for instance, by the sharing expressed markers between the mouse nanog+ cluster and the multipotent NC, delaminating NC and KPNA2+ NC clusters in chicken (Figures 3.11 and 3.12). Interestingly, this contrasts with the more terminally differentiated non-neural NC clusters, which could be identified across these species.

Moreover, we pioneered single cell transcriptomics in the marine lamprey embryo (Figure 3.15). Using dorsal neural tubes dissected at T21, we enriched our samples for NC, which revealed at high resolution the transcriptional state of the early lamprey NC. In the near future, we plan to built upon this approach to further refine and expand our analyses, thus allowing us to compare single cell transcriptomes of lampreys with the remaining species.

In parallel, we comprehensively interrogated the *cis*-regulatory landscapes of zebrafish (Figure 3.19) and lampreys (Figure 3.24) using ATAC-seq. Our results identify clusters of elements displaying differential dynamics of accessibility (Figure 3.19a) and motif content (Figure 3.23), predictive of their role in NC development. Consistent with our results from Chapter 2, we identified NC-restricted elements displaying TF biding motifs of canonical NC regulators, such as Sox10, Sox9 and TFAP2a, as well as pleiotropic enhancers shared with non-NC cells. Indeed, *in vivo* validation of these enhancers indicate pleiotropy of NC enhancers in neurodevelopment. However, in contrast to previous proposals [47, 48] and to our surprise, NC enhancers involved in EMT and ectomesenchyme development were not shared with mesoderm cell-types, arguing against a hypothesis in which the unique ability of NC to contribute to the craniofacial ectomesenchyme evolved through the co-option of mesoderm circuits.

3.4.1 Did enhancer pleiotropy mediate the early elaboration of the NC?

An intriguing hypothesis arising from this study is the early elaboration of the NC via pleiotropic changes of ancestral chordate enhancers. Indeed, previous studies have identified homologous cell-types of NC derivatives in invertebrates chordates [190, 191], including bipolar-tail neurons [193], which form the peripheral nervous system of tadpole tunicates. Moreover, it has been previously proposed that a trunk-like state may represent the ancestral NC identity [10], which later invaded the cranial region [70]. This early NC, restricted to the trunk, likely evolved in combination with cranial placodes. Therefore, our observation of enhancers active in neurogenic NC in combination with hindbrain and neurogenic placodes (Figures 3.28 and 3.29) is consistent with the idea that these enhancers might reflect a highly pleiotropic, constrained *cis*-regulatory logic reminiscent of the elaboration of the NC.

Such early appearance of the NC has been a matter of intense debate. Evolution of a new cell type, as complex as the NC, requires concerted evolutionary changes genome-wide, and thus such genomic correlates of NC origins have been challenging to identify. Thus, enhancer pleiotropy might represent an attractive hypothesis mechanistically explaining the early evolution of the NC. For example, standing genetic variation on enhancers involved in peripheral nervous system development in populations of the stem vertebrate ancestor could have been selected and rapidly fixed, thus causing genome-wide directional selection underlying the appearance of novel roles to neurogenic progenitors arising at the trunk lateral plate border. Subfuntionalisation of this cell type might have given rise to NC. Finally, later expansion of genes and/or genomes within the stem vertebrate lineage might have facilitated the evolution of novel developmental roles for the NC, such as in ectomesenchyme development.

Consistent with the hourglass model for vertebrate embryonic development [247], pleiotropic enhancers were active during the pharyngula stage, representing a phase of increased functional constraint in development (*i.e.*, the middle of the hourglass, Figure 3.30). Therefore, enhancers acting before or after this constrained developmental phase may evolve at faster rates, and thus underlie evolutionary innovations within vertebrates.

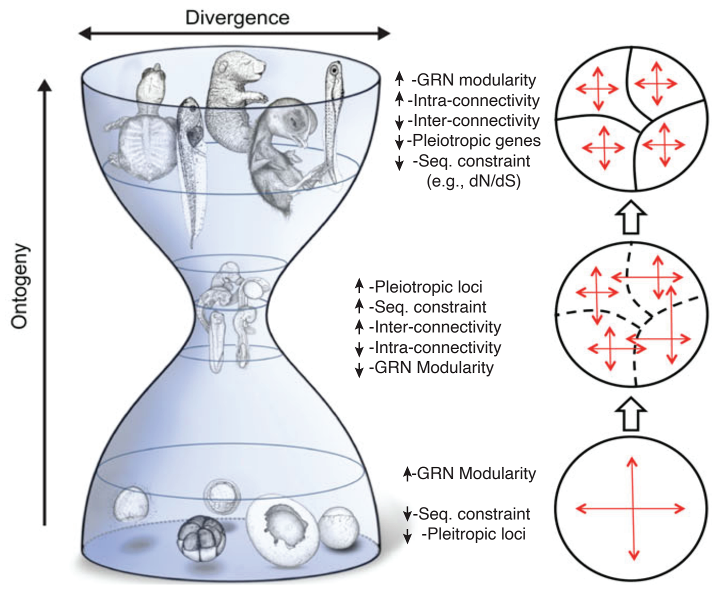

Figure 3.30 The developmental hourglass model. The model predicts that mid-embryonic stages (the pharyngula stage) represent the period of highest conservation owing to increased pleiotropy, functional constraint and inter-connectivity of GRNs. Adapted from [247]

.

3.4.2 Cell-type restricted enhancers and the diversification of the NC

Although cell type-restricted enhancers seem to be less constrained, and thus displaying higher rate of motif turnover [120], TF biding [212] and weaker activities [132], another intriguing hypothesis stemming from our cross-species comparisons is the possibility that NC-restricted enhancers may underlie evolutionary innovations within vertebrates. For example, comparison of the *Sox10* locus between zebrafish and chicken (Figures 3.20, 3.21) suggests the acquisition of putative enhancers within amniotes or the loss of *cis*-regulatory elements in zebrafish. It is possible that orthologous zebrafish enhancers reflects conserved elements likely reflecting their ancestral role. One intriguing possibility is that putatively duplicated enhancers in the stem amniote ancestor may have played a role in additional axial regionalisation of the NC, parsing axial identities into cranial, vagal, cardiac, trunk and sacral within amniotes, *versus* cranial and trunk in anamniotes [10]. This hypothesis will be interesting to test, and may provide a framework for the identification of enhancers underlying species-specific differences in the vertebrate NC.

3.5 Summary

This final experimental chapter of my doctoral thesis presents evidence that two main classes of *cis*-regulatory elements display important differences in their evolutionary dynamics. The first group, represented by pleiotropic enhancers, regulate neurogenic NC. *In vivo* validation of these elements indicate they are largely shared with diverse neurogenic tissues, including hindbrain neuronal cells and neurogenic placodes. These results indicate that the early neurogenic NC was elaborated via pleiotropic changes in ancestral enhancer elements. In contrast, NC-restricted enhancers are restricted to the cranial region and display specific activity ectomesenchymal progenitor NC cells. Surprisingly, we did not find enhancer elements shared with other mesenchymal derivatives, such as those derived from the mesoderm. These results suggest that NC contribution to the ectomesenchyme evolved *de novo*, likely through the acquisition of novel enhancers. Our results thus have important implications to the understanding of how novel cell types evolve.

4 Conclusions and Future Directions

The dynamics of embryonic and post-embryonic development has puzzled humanity since our beginning. Abya Yalan indigenous rituals and paintings, classical Greek philosophers, Celtic ceremonies, Chinese medicines, Tibetan Buddhist bardos, to name but a few of Eastern and Western approaches and knowledges, illustrate the numerous reports from humanity attempting to reduce the complexity of the steps bridging fertilisation to death in a manner that we might be able to grasp, albeit through different lenses and angles. For developmental biologists, this complexity represents our everyday work and passion. Driven by curiosity, an equally complex question has challenged generations of developmental biologists: how can one complex organism, displaying such intricate symmetries and asymmetries, sculpted body plans and stunning patterns of gene expression, faithfully emerge from a single cell?

I attempted to understand such complex question of how cell fate decisions are computed in live vertebrate embryos using NC cells as a model system. Using a regulatory-centric approach based on gene regulatory theory, I interrogated NC development at multiple scales (i.e., at the single cell- and population levels; in *cis* and in *trans*), which enabled myself to: (1) uncover early heterogeneity of NC cells in the avian and zebrafish embryo at the transcriptomic and epigenomic level; (2) decode the *cis*-regulatory architectures underlying such heterogeneity; (3) identify a combinatorial *cis*-regulatory code encoding developmental information that directs (or at least bias) NC cells towards neurogenic *versus* ectomesenchymal fates; (4) reverse engineer the gene regulatory hierarchies directing NC cells towards these fates; and finally (5) identify two classes of *cis*-regulatory elements - pleiotropic enhancers and NC-restricted enhancers - displaying differential dynamics in development and evolution, revealing how such cell fates were incorporated into NC development during the radiation of vertebrates.

My DPhil provided insights onto the early NC development of three vertebrate species, chicken, zebrafish and lamprey, representing lineages that span >540 million years of vertebrate evolution. I developed a computational pipeline to reverse engineer the NC genomic regulatory network, which was successfully applied to the chicken NC. My work combined single cell transcriptomic profiling to identify co-expressed developmental genes, which are then integrated to their cognate epigenomic landscapes in a genome-wide manner. This state-of-the-art approach allowed us to decode at high-resolution the gene regulatory hierarchies controlling NC development. However, two main limitations remain: firstly, our approach is based on the identification of TF binding sites that were largely characterised in human and mouse genomes. Divergences in TF binding in chicken and zebrafish may cause important noise to my predictions. Secondly, our approach relies on the parallel profiling of cells for transcriptomics and epigenomics. Novel technological approaches enabling the simultaneous profiling of single cells provide an elegant advancement to improve my approach to reverse engineer GRNs, allowing the same cell to be interrogated at the *cis* and *trans* level.

My future work thus need to tackle these important limitations. Adoption of a novel exciting approach, CUT&RUN [115], a method for low input (1-100.000 target cells) high-resolution profiling of TF binding, is starting to enable our lab and others to map TF binding and identify motifs *de novo* cells directly isolated from their endogenous environment. Nevertheless, future work should also aim at expanding the vertebrate models used to interrogate cell-types homologous to the NC, and may further benefit from the inclusion of invertebrate chordates. Several taxa at key phylogenetic positions representing important transitions in the radiation of vertebrate remain poorly studied: lampreys and hagfishes, sharks, skates, rays and chimaeras, bony fishes, lung fishes, amphibians (including caecilins), and multiple amniotes, such as lizards, snakes, avians, marsupials and mammals, among many others. Important limitations have been identified with the species we used here as models [184], and we therefore need to be cautious to do not generalise our findings to >600 million years of vertebrate evolution. Moreover, adoption of single-nuclei ATAC-seq and CUT&RUN methods in embryos from these lineages, together with targeted transcriptomic profiling of fixed or frozen tissue [248] might further enable us to advance our understanding of the regulatory mechanisms shaping the diversity of chordates.

We also expect that novel deep learning approaches, such as deep-STARR [139], might provide exciting new paths to circumvent the challenge of identifying orthologous *cis*-regulatory elements at large evolutionary distances. Deep learning might provide a way forward for us to identify informative features predictive of enhancer activity shared across several experimentally-tractable species, and transfer-learning can be applied to predict similar *cis*-regulatory elements in diverse taxa for which no functional data can be obtained. Furthermore, deep learning approaches may be used to identify disease-associated *cis*-regulatory elements; in combination with the evolutionary constraints we have identified here, this may provide key insights to rank enhancers related to human neurocristopathies for clinical research prioritisation.

Importantly, the reverse engineered circuitries presented here represent high-resolution topologies describing TF-driven gene batteries regulating neurogenic *versus* ectomesenchymal progenitor states of NC cells. However, further work exploiting these topologies, using combination of mathematical models (i.e., based in differential models and/or graph theory) are still needed for us test our ability to predict normal gene expression dynamics *in silico*. Such computational modelling of our circuitries will be important to also transfer the knowledge I developed here in a manner that it can be useful to identify developmental nodes and linkages likely disrupted in human diseases.

Excitingly, my final experimental chapter revealed two divergent *cis*-regulatory architectures constraining the evolutionary dynamics of the NC. On one hand, highly-constrained, densely-packed with motif content, pleiotropic enhancers were shown to regulate neurogenic NC. I have thus proposed that the early NC might have evolved through pleiotropic changes of ancestral chordate enhancers, which incorporated neurogenesis into the early, melanocytic NC. In contrast, functionally weak NC-restricted enhancer elements, which likely evolved *de novo*, may have mediated the evolution of novel ectomesenchyme-derived morphologies in the stem vertebrate ancestor.

My future work would benefit from the use synthetic libraries of enhancers to identify motif changes involved in the acquisition of such novel developmental fates. Interestingly, recent work using a library of synthetic enhancers was used to identify important syntactical rules controlling enhancer activity in *Drosophila*. In my future work, it would be fantastic to integrate deep learning approaches to identify predictive features of enhancer activity in the NC, and use them to design novel enhancers with desired properties, as a manner to provide proof-of-principle to the evolutionary scenario proposed here (i.e., synthetic enhancers involved in NC-mediated neurogenesis versus ectomesenchymal development).

Finally, another important limitation worth of citing of my DPhil is lack of integration with the biophysical regulation of the NC. My results suggest that the unique ability of NC progenitor cells to direct daughter cells to one fate or another in a sterotypical manner following reproducible patterns is certainly encoded in the genome. However, coarse-grained computational modelling of embryonic vertebrate development has already provided important insights suggesting that biomechanical interactions might be enough for some of the complex operations the NC operate [96]. Therefore, I envision future systems-biology approaches requiring the integration of biophysics and genomics to computationally model embryonic development in a manner that will allow for the faithful reconstruction of developmental dynamics. Computational modelling of development will likely enable us to identify biomechanic forces acting in a concerted manner with gene regulatory networks; such intense computational process can bring us to a time where whole-organism development from a single-cell can be recapitulated at high-resolution *in silico*. Such multi-lab work to decode the forces underlying metazoan development might prove most useful for us to advance our understanding of developmental forces shaping evolution, as well as to decipher human developmental diseases.

These are exciting times to study development.

5 Materials and Methods

5.0.1 Chick embryos

Fertilised Bovans Brown chicken eggs (Henry Stewart Co.) were incubated at 37°C with approximately 40% humidity. Embryos were staged according to (Hamburger and Hamilton, 1951). All experiments were performed on chicken embryos younger than 12 days of development, and as such were not regulated by the Animals (Scientific Procedures) Act 1986.

5.0.2 Embryo preparation and *ex ovo* electroporation

Hamburger and Hamilton (HH4) embryos were captured using the filter paper based 'easy-culture', method. Eggs were opened after desired incubation period, albumin was removed and embryos were removed from yolk using punctured filter paper and kept in Ringers solution until electroporation, as previously described [55, 230, 232]. Plasmid(s) were injected across the whole the epiblast and electroporated: 5 pulses of 5V, 50ms on, 100ms off. We used the bilateral electroporation technique for perturbation experiments, whereby control and experimental regents are delivered to opposite sides of the primitive streak, providing ideal internal, stage matched controls for each experiment. Embryos were cultured on albumin at 37°C/5% CO2 overnight to the desired stage.

5.0.3 Zebrafish husbandry

Animals were handled in accordance to procedures authorised by the UK Home Office in accordance with UK law (Animals [Scientific Procedures] Act 1986) and the recommendations in the Guide for the Care and Use of Laboratory Animals. All vertebrate animal work was performed at the facilities of Oxford University Biomedical Services. Adult fish were maintained as described previously [249]. Embryos were staged as described previously [250]. In brief, embryos were staged using a dissecting stereo-microscope. 75% epiboly stage was identified by observing a distinctively thicker dorsal side and visible epiblast, hypoblast and evacuation zone. 1-2ss - by observing first/second segment furrow. 5-6ss - counting 5-6 somites, apparent optical and Kupffer's vesicles and prominent polster. 14-16ss - counting 14-16 somites, observing otic placode, v-shaped trunk somites.

5.0.4 Lamprey embryos

Adult sea lamprey (*Petromyzon marinus*) were supplied by the US Fish and Wildlife Service and Department of the Interior. Embryos obtained by *in vitro* fertilisation, were grown to the desired stage in compliance with California Institute of Technology Institutional Animal Care and Use Committee protocol #1436. Prior to dissection, embryos were dechorionated in 0.1x Marc's Modified Ringers buffer (MMR) in a dish lined with 1% agarose. For dissection, embryos were moved into a shallow well in the agarose. T18, T20 and T21 DNTs including premigratory, early-delaminating and/or late-delaminating NC cells were dissected from the head using an eye-lash knife. T20 and T23 heads were dissected using forceps.

5.0.5 Cell dissociation and FAC-sorting

Following electroporation with the NC specific Foxd3 enhancer NC1 (Simoes-Costa et al., 2012), driving a Citrine reporter (2ug/ul), we used fluorescence activated cell sorting (FACS) to collect Citrine-positive NC and Citrine-negative control cells at two distinct stages of cranial NC development (5-6ss and 8-10ss). Dissected cranial regions from electroporated embryos were dissociated with dispase (1.5mg/ml in DMEM/10mM Hepes pH 7.5) at 37°C for 15 minutes with intermittent pipetting to achieve a single cell suspension and with 0.05% Trypsin at 37°C for a final 3 minutes dissociation step. The reaction was stopped and cells were re-suspended in an excess of Hanks (1X HBSS, 0.25% BSA, 10mM Hepes pH8) buffer. Cells were centrifuged at 500g for 10min, re-suspended in Hanks buffer, passed through a 40-um cell strainers, and centrifuged at 750g for 10min. Pelleted cells were re-suspended in 500ul Hanks buffer. Citrine positive cells were sorted and collected using BD FACS-Aria Fusion. We collected 300 and 600 NC cells per embryo at 5-6ss and 8-10ss, respectively. FACS isolated cells were used for RNA-seq and epigenome profiling experiments (bulk and single-cell).

5.0.6 Bulk RNA extraction, library preparation and sequencing

FAC-sorted cells were centrifuged at 2000g for 5mins at 4°C, washed with PBS, and resuspended in lysis buffer and stored at -80°C. RNA was extracted using Ambion RNAqueous Micro Total RNA isolation kit (Cat. #AM1931, ThermoFisher Scientific), Turbo-DNAse treatment was performed for 1 hour to ensure removal of plasmid. RNA integrity was checked using Bioanalyser, and only samples with RIN>7 were used. Sequencing libraries were prepared using Takara SMARTerïf' low input RNA kit and sequenced using 100bp paired-end reads on the Illumina Hiseq2000 platform. Minimum of three biological replicas for each stage were used for analysis; however, only duplicates were used for data visualisation.

5.0.7 Single cell RNA preparation, library preparation and sequencing

Individual NC cells were collected by FACS. For SmartSeq2 cDNA was generated and sequencing libraries were prepared as previously described (Picelli et al., 2014). Briefly, mRNAs were primed with oligo-dT and reverse transcribed using an LNA-containing template switching oligo. Libraries were generated from amplified cDNA by Tagmentation with Tn5. Libraries were sequenced using 50bp single end reads for 96 cells. A 4 x107 dilution of ERCC spike-in control was used. For 10X Genomics based single-cell RNA-seq the Chromium Single Cell 3' Library and Gel Bead Kit v2 (Cat. #120267) was used following manufacturers protocol. Briefly, NC cells were collected by FACS into 2ul Hanks buffer, approximately 8000 cells were loaded onto the Chromium. Libraries were quantified using Qubit and Kappa qPCR and 4nM 10X scRNA-seq library was sequenced on a NextSeq500 platform (Illumina) using high output v2.5 150-cycle kit in 26 x 8 x 0 x 98 mode.

5.0.8 ATAC, library preparation and sequencing

ATAC was performed using a protocol adapted from (Buenrostro et al., 2013). FAC-sorted cells were lysed (10mM Tris-HCl, pH7.4, 10mM NaCl, 3mM MgCl2, 0.1% Igepal) and tagmented using Illumina Nextera DNA kit (FC-121-1030) for 30mins at 37°C. Tagmented DNA was amplified using NEB Next High-Fidelity 2X PCR Master Mix (Cat. #M0543S) for 11 cycles. Tagmentation efficiency was assessed using Agilent Tapestation.

Single-cell ATAC was performed using the 10X Genomics Chromium platform. Nuclei were prepared using demonstrated protocol (CG000169 | Rev A) from 10X Genomics. Briefly, pelleted cells (500g, 5mins, 4°C) were washed with PBS supplemented with BSA (0.04%) and resuspended in nuclei lysis buffer (10mM TrisHCL-pH7.4, 10mM NaCl, 3mM MgCl2, 0.1% Tween-20, 0.1% Noidet P40 substitute, 0.01% Digitonin, and 1% BSA), lysis efficiency was determined using Countess

II FL automated cell counter. Tagmentation was conducted on bulk nuclei and libraries were generated using the Chromium Chip E Single Cell ATAC Kit and Chromium Single Cell ATAC Library Gel Bead Kit (PN-1000111) before loading (approximately 5000 nuclei) into Chromium and following supplier's protocol (CG000168). ATAC-seq libraries were sequenced using paired-end 40bp reads on the Illumina NextSeq500 platform. Three biological replicates were obtained for each stage. 4nM 10X scATAC-seq library was sequenced on a NextSeq500 platform (Illumina) using high output v2.5 75-cycle kit in 36 x 8 x 16 x 36 mode. 5% of PhiX control was spiked in and 1.8pM loaded onto the flow cell according to manufacturer's instructions.

5.0.9 H3K27Ac ChIP, library preparation and sequencing

FAC-sorted cells were cross-linked with 1% formaldehyde for 9 mins. Fixation was quenched with 125mM of glycine for 5min. Cross-linker was washed out by 3x pellet washes with 1xPBS ((supplemented with 1X PI (protease inhibitors)), 1mM DTT and 0.2mM PMSF) centrifuging at 2000g for 4mins at 4°C. Pellets were resuspended in isotonic nuclei extraction buffer (NEB: 0.5% NP40, 0.25% Triton-X, 10mM Tris-HCl-pH 7.5, 3mM CaCl2, 0.25M sucrose, 1mM DTT, 0.2mM PMSF, 1X PI). Cell nuclei were expulsed with 20 strokes using pestle B in a glass homogeniser, pelleted and washed with 1xPBS (with 1X PI, 1mM DTT and 0.2mM PMSF). Nuclei were lysed in SDS lysis buffer (0.7% SDS, 10mM EDTA, 50mM Tris-HCl (pH 7.5), 1x PI). Cross-linked chromatin was sonicated at 11A 10x (10s on 30s off) followed by 6A 8x (30s on 30s off) using a probe sonicator, to achieve 300-800bp fragments. Pre-blocked Protein A Dynabeads were incubated with antibody (H3K27ac, Abcam Ab4729, Brd4, Bethyl Laboratories A301-985A). Sonicated DNA-protein complexes were applied to beads overnight at 4°C. IgG antibody (Millipore 12-370) was used as control and an input sample was taken. Samples were washed 8x with

RIPA buffer (50mM Hepes-KOH (pH 8.0), 500mM LiCl, 1mM EDTA, 1% NP40, 0.7% Na-Deoxycholate, 1x PI) and 1x NaCl TE wash (1x TE, 50mM NaCl) at 4°C. Chromatin was eluted from the beads with SDS ChIP elution buffer (50mM Tris-HCl (pH 7.5), 10mM EDTA, 1% SDS). Cross-linking was reversed overnight at 65°C at 1300 rpm. Cellular RNA was digested with RNaseA (0.2mg/ml) at 37°C for 1 hour, and cellular proteins were removed with Proteinase-K (0.4mg/ml) at 55°C for 1 hour. Samples were purified by standard phenol-chloroform extraction and ethanol precipitation. ChIP-ed DNA was amplified using published protocol (Adli and Bernstein, 2011) for small cell number ChIP, whereby Seqeunase was used to prime ChIP DNA with custom oligos containing PCR primers and restriction enzyme (BciVI) sites allowing amplification and subsequent digestion yielding library prep ready fragmented DNA. Libraries were prepared using NEBNext Ultra DNATM library prep kit following manufacturers protocol. ChIP libraries were sequenced using 50bp paired-end reads using Illumina Hiseq2500 platform. Three biological replicates for each stage and antibody were obtained.

5.0.10 Capture-C

Capture-C was performed according to the published protocol (Davies et al., 2016), with a few minor adaptions. To collect the material, dorsal neural tubes were dissected from 6-7ss embryos and dissociated in nuclei extraction buffer (NEB, as above), 20 strokes with pestle A in a glass homogeniser. Cells were cross-linked with 2% formaldehyde for 8min. Fixation was quenched with cold 1M glycine. Pelleted cells (1500rpm, 4°C, 5min) were washed with 1xPBS, re-pelleted and resuspended in lysis buffer (10mM Tris-HCl (pH8), 10mM NaCl, 0.2% Igepal-NP40, 1x cOmplete protease inhibitor (PI)), and left on ice for 20min before pelleting. Supernatant was removed and pellets were stored at -80°C. On the day of the Capture, cross-linked cells were re-suspended in water on ice and homogenised with pestle B, 45 strokes, and centrifuged at 14,000 rpm for 5min. Pelleted cells were re-suspended in water, restriction enzyme buffer and SDS (0.25% final) and incubated at 37°C with 1400 rpm shaking for 1 hour. Triton-X-100 (8.25% final) was added and samples were incubated again with shaking for 1 hour. 3x 125uL aliquots of restriction enzyme (DpnII, NEB High-concentration 50,000 U/ml), were added several hours apart, such that samples were digested for 16-24 hours. Restriction enzyme was inactivated by 20min incubation qt 65°C. To ligate fragmented DNA, water, ligation buffer and 240uL T4 ligase (high-concentration 30U/ul; Cat. #EL0013, ThermoFisher Scientific) was added and samples were incubated overnight at 16°C with 1400rpm mixing. Samples were uncrosslinked overnight at 65°C in presence of 50ug of Proteinase-K and then treated with 300uL RNaseA (Cat. #1119915, Roche) for 30min at 37°C. DNA was extracted using standard Phenol:Chloroform method and ethanol precipitation. Thus, obtained 3C libraries were then sonicated to 200bp, using Covaris S220 with the following settings: duty cycle 10%, intensity 5, cycles per burst 200, 6 cycles each of 60s and set mode. Sonicated samples were cleaned using Ampure XP beads, quantified by Qubit (ThermoFisher Scientific) and checked using 2200 Agilent Tapestation System. Libraries were prepared using NEBNext Ultra library prep kit and NEBNext Multiplex oligos for Illumina according to the manufacturer's protocol with the exception of using Herculase enhanced DNA polymerase (Agilent). Libraries were cleaned up with Ampure XP beads and again assessed using Tapestation and Qubit.

Libraries were prepared for hybridisation by adding 5ug Chicken Hybloc competitor DNA (Applied Genetics Laboratories) to 1.5-2ug 3C library, 1nmol Nimblegen-HE universal oligo, and 1nmol Nimblegen-HE index oligo (specific to the indeces used in library preparation) (Nimblegen SeqCap EZ hybridisation and wash kit, Cat. #05634261001, Roche). The assembled reaction was dried by vacuum centrifuge at 50°C. Hybridisation buffer and component were added and the DNA was carefully reconstituted. Target specific biotinylated oligos were preheated to 47°C in 0.2ml PCR tubes. 3C libraries and blocking oligos were denatured at 95°C for 10min before adding to the preheated biotinylated oligos. The mixture was then incubated at 47°C for 64-72 hours. Hybridised samples were washed and M-270 Streptavidin Dynabeads (Cat. #65305, ThermoFisher Scientific) were prepared according to Nimblegen SeqCap EZ hybridisation and wash kit protocol. Hybridised samples we applied to streptavidin beads and incubated at 47°C for 45min with 600rpm mixing. Washing steps were performed as per kit's instructions. Amplification of captured multiplex DNA sample was done using the V2 Nimblegen SeqCap EZ accessory kit (Cat. #07145594001, Roche), followed by an Ampure XP bead clean up. As recommended (Davies et al., 2016), to maximise the capture, we repeated the hybridisation reaction (24 hours), as well as subsequent washes and amplification steps. Finally, captured libraries were sequenced using 150bp paired-end reads on the Illumina Miseq platform (300-cycle MiSeq Reagent Kit v2, Cat. #MS-102-2002, Illumina). A full version protocol is available at http://www.tsslab.co.uk/resources.

5.0.11 Generating Nanotag reporter vectors

pTK Citrine/Cerulean/mCherry reporter vectors [220] were modified to generate an 'easy-insert' integration site enabling highly efficient cloning of putative enhancers using rapid simultaneous ligation/digestion reaction in a single tube. To this end, a lacZ expression cassette flanked with restriction sites for the type IIs restriction endonuclease BsmBI was cloned upstream of theof the minimal Thymidine Kinase (TK) promoter. To enable multiplexing and simultaneous detection of reporter activity from different enhancers, we further modified pTK reporter vectors to include unique barcoded DNA-tags ('nanotags') (Nam and Davidson, 2012) downstream of fluorescent reporters, within their 3'UTR regions and upstream of their polyadenylation signal. We thus generated 48 new Nanotag reporter vectors (16 each for pTK-Citrine, pTK-Cerulean and pTK-mCherry). pTK Nanotag vector collection is available from Addgene (https://www.addgene.org/Tatjana-Sauka-Spengler/).

5.0.12 Enhancer cloning and testing

Multiplexed, high-throughput enhancer cloning pipeline is abridged below, the full technical protocol is available at http://www.tsslab.co.uk/resources. Putative enhancers were amplified from chick gDNA using primers containing specific sequence tails (see supplementary excel file) to facilitate subsequent cloning into nanotagged reporter vector using a modified GoldenGate (Engler et al., 2009) protocol. Briefly, gel-purified amplicons were combined with modified pTK nanotag reporter vector with T4 DNA ligase and BsmBI restriction enzyme and subjected to a cycling reaction that allows simultaneous BsmBI digestion and T4-mediated ligation of amplicon into the reporter vector. Endotoxin-free plasmid preparations (E.Z.N.A. Endo Free Plasmid Mini Kit II, Cat. #D6950-02, Omega Bio-Tek or Qiagen endo-free maxi prep kit, Cat. #12362, Qiagen) were prepared for electroporation. Putative enhancers in reporter plasmids were pooled at 0.2ug/ul each, 10-12 plasmids including positive, NC1, and negative (short-oligos, table S2) controls, and electroporated into the entire epiblast of the early chick gastrula embryos (HH4) [220, 232]. Embryos were allowed to develop to desired stages using ex ovo culture in thin albumin. Cranial regions were dissected and extracted RNA was hybridised with Nanostring reporter code-set oligos designed to detect nanotag transcripts. Nanostring assay was conducted as manufacturers protocol and absolute transcript counts were recorded. Enhancers with a count >50 were electroporated individually at 2ug/ul and fluorescent reporter activity was imaged throughout early NC developmental stages. An extended set of 'electroporation-ready' novel positive NC enhancers cloned into the fluorescent reporters is available from Addgene (https://www.addgene.org/Tatjana-Sauka-Spengler/). Imaging analysis Embryos were imaged on an Olympus MVX10 stereomicroscope with 2.5X objective using Axio Vision 4.8 software. Zeiss 780 Upright confocal microscope was used for imaging at high cellular resolution

5.0.13 Cryosectioning and immunostaining

Embryos selected for cryosectioning were fixed in 4% PFA (paraformaldehyde) for 1 hour at room temperature (RT) or at 4°C overnight, then washed 3x 10min in 1x PBS. Embryos were cryoprotected in 15% sucrose/ PBS overnight at 4°C, followed by overnight incubation in 7.5% gelatine/15% sucrose /PBS at 37°C. Following 4-hour incubation in 20% gelatine/PBS embryos were embedded in 20% gelatine/PBS, snap frozen in LN2 and stored at -80°C. Cryosectioning was performed at 10um thickness. Prior to immunocytochemistry, excess gelatine was removed from the slides by a brief rinse in pre-warmed PBS (37°C). Sections were rinsed 3x 5min in PBT (2% DMSO, 0.5% Triton X-100 in PBS), blocked in 10% donkey serum in PBT (block solution) for 1 hour at RT, and incubated overnight at 4°C with primary antibody (1:200 dilution of rabbit anti-GFP in block solution, Cat. #TP401, Torrey Pines Biolabs). Sections were then washed in PBT at RT 5x 10min, followed by incubation with secondary antibody (AlexaFluor-488-conjugated donkey anti-rabbit IgG diluted at 1:1000 in PBT, Cat. #A21206, ThermoFisher Scientific) for 2 hours at RT. Sections were washed 6-8x 10min in PBT at RT then overnight at 4°C and mounted using Vectashield with DAPI (Cat. #H-1200, Vector Laboratories).

5.0.14 *In vivo* CRISPR-mediated perturbation assays

sgRNAs were cloned into a chicken pU6-3 vector using a BsmBI mediated approach as previously described (Williams et al., 2018) and used in conjunction with Cas9-RFP or Cas9-2A-Citrine (Addgene #92358) (Williams et al., 2018) to knockout transcription factors. Following bilateral electroporation, left and right dorsal neural tubes were dissected separately; RNA was extracted using Ambion RNAqueous Micro Total RNA isolation kit (Cat. #AM1931, ThermoFisher Scientific). For TF knockout experiments RNA-seq libraries were prepared using SMART-Seqâ„¢ v4 Ultraâ„¢ Low Input RNA Kit for Sequencing (Cat. #634889, Takara Bio Clontech) and sequenced using 40bp paired-end reads on Illumina NextSeq500. 6 individual embryos and associated RNA-sew libraries were analysed.

5.0.15 Quantification and statistical analysis

Statistical details of experiments can be found in the figure legends and text, including p-values and FDR cutoffs. Statistical analyses were performed in Microsoft Excel or R, unless differently stated.

5.0.16 NGS data processing

Sequencing files from each sequencing lane were de-multiplexed and the resulting files were merged. Reads were trimmed for quality using sickle (Joshi, 2011) (v.1.33) quality control was performed using fastQC (v.0.10.1) and deeptools (Ramirez et al., 2016) (v.2.2.2). Reads were mapped to the chicken genome galGal5, lamprey petMar3, and zebrafish danRer10 assemblies using bowtie2 (Langmead and Salzberg, 2012) (v.1.0.0), except for RNA-seq files, which were mapped using RNA-STAR (2.4.2a) (Dobin et al., 2013). Sequence alignment/map files were compressed to the binary version (BAM) for downstream analysis. Differential expression analysis was carried out using in DESeq2 (Love et al., 2014) (v.1.14.1).

Peaks were called for each ATAC-seq sample using MACS2 (v.2.0.10) (Feng et al., 2012) and ATAC-seq consensus peaks were generated using bedtools (Langmead et al., 2009) (v. 2.25.0) intersect command. For RNA- and ATAC-seq reproducibility analyses, BAM files were first normalized for sequencing depth using samtools (Li et al., 2009) (v.1.3) and count tables were retrieved using Subread (Liao et al., 2013) (v.1.4.5) featureCounts (Liao et al., 2014) tool from an Ensembl gene list in GTF format (for RNA-seq) and from a consensus set of 99,583 accessible regions (for ATAC-seq), respectively. Count tables were analysed in R (v.3.2.3) and principal component analysis (PCA) was performed using the DEseq2 (v1.4.5) package in the R (v.3.2.1) environment. A custom Perl script was used to generate smoothened genome browser tracks in BigWig format for ATAC-seq data visualisation.

5.0.17 WGCNA

Weighted correlation network analysis (WGCNA) [222] was performed on normalised gene count tables generated by DESeq2, according to the pipeline detailed in [222, 223].

5.0.18 Single-Cell Analysis

SmartSeq2 scRNA-seq Reads were mapped using RNA-STAR (2.4.2a) (Dobin et al., 2013) to the chicken genome Galgal4 assembly and the ERCC spike-in controls using the default parameters. Subread package with default parameters. Reads that aligned to more than one locus as well as ambiguous fragments were excluded from all further analysis. To remove cells with low quality sequencing, cells with a) less than 100,000 sequenced reads or b) less than 50% uniquely mapped reads were excluded from any further analysis. Further filtering was done based on the distributions of a) the number of expressed genes per cell, b) the proportion of reads on mitochondrial genes or c) the proportion of reads on ERCC spike-ins, requiring that the cells are within 3 MADs (Median Absolute Deviation) for each distribution (McCarthy et al., 2017). In the remaining 124 cells, 10,291 genes were detected using the following criteria: a) they had any reads in more than 5 cells and b) had an average number of counts above 1. The cell-based factors from the deconvolved pool-based size factors were used for the normalisation of the gene counts, as described in (Lun et al., 2016), using pool sizes of 10, 20, 40 and 60 cells. The clustering was done with the pagoda2 package (https://github.com/hms-dbmi/pagoda2) using the infomap and the walktrap community methods. There was a 96% concordance between the two methods and the infomap method was used to obtain the final clusters. An analysis for differentially expressed genes was performed using the SCDE package (Kharchenko et al., 2014) to find cluster specific gene markers. For each cluster a one-versus-one approach was applied between all pairs of clusters using the default parameters.

10X Chromium scRNA-seq Single-cell RNA-seq raw base call files were demultiplexed using Cellranger (v3.0.0) (Zheng et al., 2017) and mapped to the galGal6, danRer11, petMar3 and mm10 transcriptomes, and compiled using the Cellranger "mkref" function. 3,108 estimated number of cells were obtained with mean reads per cell of 152,318, median genes per cell of 1,501, median UMI counts per cell of 4,968 and total genes detected was 15,511. 97.7% of reads had valid barcodes with a Q30 of 97.5% and 83.6% of the reads mapped confidently to the chicken genome. Downstream analysis was carried out using Seurat package (v3.0.0)(Butler et al., 2018) in R. Matrices were filtered to remove barcodes with fewer than 200 genes, more than 3000 genes expressed and carried a high percentage of UMIs from mitochondrial features (greater than 10%) (Figure S5A). Expression values for the total UMI counts per cell were normalised and permutation test carried to determine the significant principal components and dimensions in the data before performing linear dimensional reduction (resolution = 0.45, dims = 1:15). Seven cell clusters were obtained and visualised on a t-SNE plot. After identifying cell markers, we removed two clusters that contained mostly housekeeping or cycling genes and perform further subclustering of the remaining 5 clusters with a total of 1509 cells (resolution = 0.45, dims = 1:20).

5.0.19 *K*-means clustering of ATAC data.

Clustering of downsampled ATAC-seq datasets was performed with the seqMINER platform (v.1.3.4) (Ye et al., 2011) on a consensus set of 99,583 accessible regions using the k-means enrichment linear clustering normalisation algorithm set to k=10 and a window of +/-1.0 kb from the peak centre. Heatmaps and merged profiles were plotted using the Deeptools (v.2.2.2) package (Ramirez et al., 2016). K-Cl specific count tables were then used for correlation and linear regression analyses and visualization in R (v.3.2.3).

5.0.20 Differential accessibility.

Differential accessibility analysis was carried out in R (v.3.2.1) using the DiffBind package (v1.10.2). Differential accessibility across samples was calculated using a negative binomial distribution model implemented in DEseq2 (v1.4.5). A stringent threshold (FDR<0.1, Fold enrichment>1) was used to define high-confidence differentially accessible ATAC-seq peaks by comparing the enrichment of accessible regions in NC cells at two stages (5-6ss and 8-10ss) over non-NC cells. Normalised read counts were clustered and visualised using pheatmap (v. 1.0.10) (Kolde, 2018) with default settings.

5.0.21 Next-Generation Capture-C analysis

NG Capture-C was performed as described previously (Davies et al., 2016) on dissected dorsal neural tube tissue and chicken erythrocytes. Samples were indexed for multiplexing and co-capture of promoters using biotinylated 120-mers (Integrated DNA technologies, IDT), designed with the CapSequm webtool (http://apps.molbiol.ox.ac.uk/CaptureC/cgi-bin/CapSequm.cgi) (Hughes et al., 2014) and pooled to a final concentration of 2.9nM (Table S1). Captured material was pooled and sequenced using 150bp paired-end reads on the Illumina Miseq platform (300-cycle MiSeq Reagent Kit v2, Cat. #MS-102-2002, Illumina). Reads were mapped using Capture-C scripts (https://github.com/Hughes-Genome-Group/CCseqBasicF/releases), and analysed as previously described (Hay et al., 2016) using intra-house R-scripts and DESeq2.

5.0.22 Functional annotation of CREs

A custom Python (v.2.7.5) script was used to assign peaks to nearest expressed genes. Ensembl identifiers were converted to gene names using the Chicken Ensembl Gene ID converter (https://www.biotools.fr/). Genes associated to two or more enhancers were used for functional analysis and GRN assembly. High-resolution TADs were determined as described in (Davies et al., 2016). The Panther package (Mi et al., 2017) (v.11) was used to calculate statistical overrepresentation tests using default settings. P-values were calculated with binomial distributions with Bonferroni correction for multiple hypothesis testing.

5.0.23 Motif analysis

De novo motif discovery was performed using Homer (Heinz et al., 2010) (v.4.7) findMotifsGenome.pl script, using clusters of ATAC-seq peaks identified by differential accessibility analysis and k-means clustering. Motifs of 8-22bp in length were searched in +/-250bp windows from the peak centre and a constant background containing a set of 99,583 peaks was used. Non-redundant matrices with $p < 10^{-11}$ (Binomial P-test) were retained. De novo motifs were annotated using Homer (v.4.7), STAMP (Mahony and Benos, 2007), TOMTOM (Gupta et al., 2007) and manual comparisons. Log10 transformed P-values were clustered and visualised using pheatmap (v. 1.0.10) in R (v. 3.2.1) with default settings.

5.0.24 Combinatorial binding

All possible pairs of cluster-specific de novo motif combinations were computed in R (v. 3.2.1). Positive controls were defined by using a set of 99,583 consensus peaks subtracted to each cluster of elements using bedtools (v. 2.25.0) intersect command. Homer (v.4.7) annotatePeaks.pl script was used to screen all de novo motifs occurring in windows of +/-250bp from the peak centre. Motif occurrences were converted to a matrix of motif presence (1) or absence (0) in R (v.3.2.3), and a custom Python3 script using the Pandas package was used to calculate motif co-occurrences. Combinations enriched at $\alpha = 5\%$ (two-tailed Chi-squared test) with Bonferroni correction for multiple hypothesis (m) testing were retained for P-values $<\alpha/m$ (Noble, 2009). Motif combination frequencies in differentially accessible and in k-Clusters groups of elements versus that of positive controls were used as a metric of combinatorial enrichment. Co-binding networks were plotted using the Circlize package in R (v.3.2.3).

5.0.25 Identification and ranking of super-enhancer-like clusters

Super-enhancers were identified using the ROSE (Ranked Order of Super Enhancers) pipeline (http://younglab.wi.mit.edu/super-enhancer-code.html). This algorithm groups (stitches) identified individual enhancers positioned within a defined distance into putative super-enhancers (SEs) and ranks them by their input-subtracted H3K27Acetylation signal. We used the default options with 12.5 kb as the maximum grouping (stitching) distance, the reads per million normalization for H3K27Ac ChIP signal from 5-6ss and 8-10ss FAC-sorted NC and a promoter exclusion zone of 1,000 or 2,500 basepairs. The enhancer clusters were ranked by cumulative normalised H3K27Ac signal (H3K27ac signal versus enhancer rank) and super-enhancers were identified as clusters lying beyond the identified inflection point on the ranking curve (Whyte et al., 2013). Super-enhancers cluster were associated to the NC enhancers identified in this study and assigned to the closest expressed gene(s). The identification of groups of NC enhancers enriched in Brd4 occupancy on was achieved by k-means clustering (Seqminer platform) (Ye et al., 2011) of the Brd4 ChIP datasets using 8-10ss DiffBind enhancers as a reference. GRN assembly

A comprehensive database collection of high-resolution TF binding models from vertebrates (HOCOMOCO v.11) (Kulakovskiy et al., 2018) was downloaded in Homer (v.4.7) format. Position weight matrixes (PWMs) with a stringent Binomial 3-value < 0.0001 were selected, resulting in a list of 769 PWMs. These TF binding models were first filtered based on PWM redundancy and were further eliminated by applying a minimum expression level threshold (1>FPKM) to their cognate TFs, resulting in a list of 169 motifs. scRNA-seq was used to define TFs co-expressed in clusters of single cells. This list was further refined by eliminating ubiquitous TFs and by combining TF paralogs (e.g. TFAP2A and C; TFAP2B and D etc. based on high similarity of their binding motifs). This resulted in a final list of 49 PWMs being retained. Homer (v.4.7) annotatePeaks.pl script was used to screen these PWMs in all differentially accessible and in k-Clusters groups of elements, allowing the inference of the upstream (TF) inputs putatively binding to such elements. Matrices containing the list of CREs and upstream TF inputs were merged with another matrix containing the CREs assignments to their target promoters, which were determined either by Capture-C or by positional information from expressed genes. These elements were grouped based in the differential accessibility analysis and k-means clustering, which facilitated the retaining of the GRN spatiotemporal dynamics. The combination of such matrixes linked the upstream (TF) inputs and the downstream outputs (target genes), resolving GRN hierarchies and modules in a genome-wide fashion, at single-cell resolution and with epigenomic details depicted. GRN hierarchical relationships were assembled and used for network analysis and visualization in BioTapestry (Longabaugh et al., 2005) (v. 7.1.1). Sub-networks for specific genes or groups of genes were retrieved from the matrixes containing all hierarchies and can be interactively surveyed in the ShinyApp.

Data and software availability The beta version of the ShinyApp allowing access to the data and network construction can be downloaded from https://github.com/tsslab/chick-NC-GRN. The data and images associated with the ShinyApp can be downloaded from https://Figure share.com/articles/chick-NC-GRN-RData/6953294 and https://Figure share.com/articles/chick-NC-GRN-images/6953306. The raw and processed data generated in this study have also been submitted to GEO (GEO record GSE121527).

6 Acknowledgments

First of all, I would like to thank my supervisor, Tatjana, for believing in me. She taught me the joy of being outside of my comfort zone, showing the importance of embracing the unknown and nurturing an intellectually-stimulating environment. My doctoral work would not have been the same without her unorthodox mentoring style.

I have also been privileged to work alongside a number of brilliant colleagues and collaborators over these years: Ruth Williams, Martyna Lukoseviciute, Dorit Hockman, Filipa Simões, Irving Ling, Daria Gavriouchkina, Mara Artibani, Van Chong-Morrison, Lucy, Sarah, Amy, Michael, Gunes, Oana, Andrew, Hanifi, Zhyuan, Evgeny, Daniel, Jasmina, Seda and my undergraduate students, Ethan and Ramona. Working with you has been a pleasure.

I am also grateful to my co-supervisor, Marella, who always offered her kindness when I most needed. I am also thankful to my thesis committee, Jim, Tudor, Tom, Aziz, as well as St. John's College staff, Zoltan, Caroline and Joyce.

My DPhil work would not be possible without the excellent WIMM Core facilities. I am thankful to Kevin for helping with FACS, Pablo and Cris (Wolfson Imaging facility), Neil Ashley from the Single-Cell facility for providing support, as well as the BMS staff for zebrafish husbandry.

This work would not be complete without mentioning Seb, Chico, Yayoi, Lili Barbieri, Pablo Carravilla, Narain, Erdinc, Falk, Pablo Infate Amate, Laura Molloy, Natsuno, Adir, Bia, Sergio, Femi, Panqueca, Porpeta, Cida and Fatima. On the other side of the ocean, I was lucky to always have the support from Daniel, Marina, Cacicão André, Tai, Vandy, Feto, Naty, Laurex, Vits, Lala, Caito, Ma, Ka, Cabelo and Bizu.

I did not find words to thank mom, dad, muchacha, gordinha and Gabi for their unconditional love and support.

This work is dedicated to the memory of Maheli Jizar Lavander.

7 Appendix

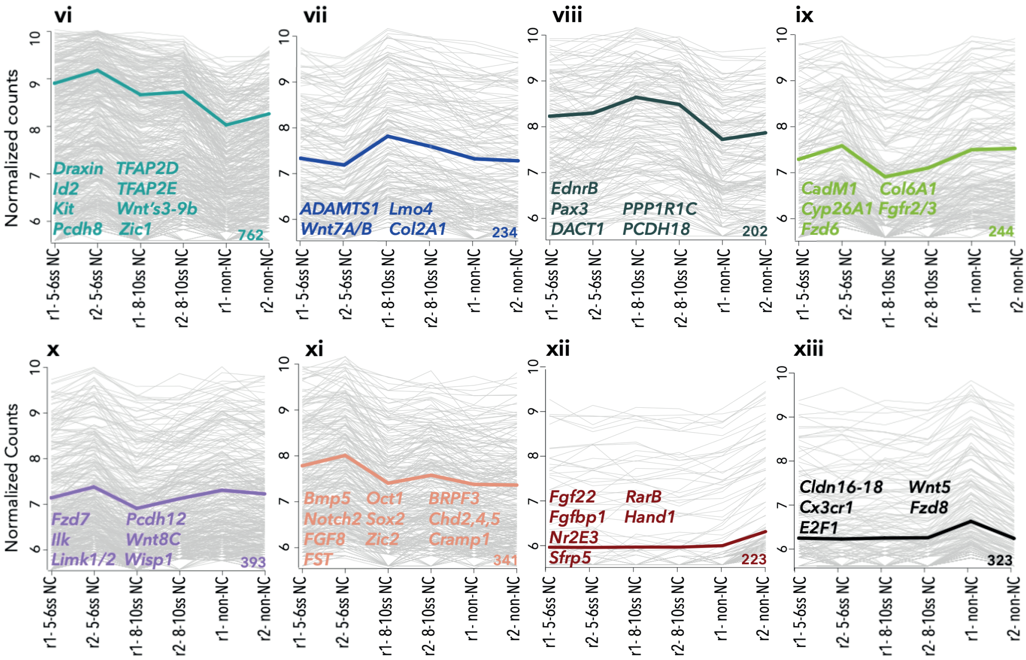

Figure 7.1 Appendix 1 - Clusters (vi-xiii) of highly correlated genes identified by WGCNA.

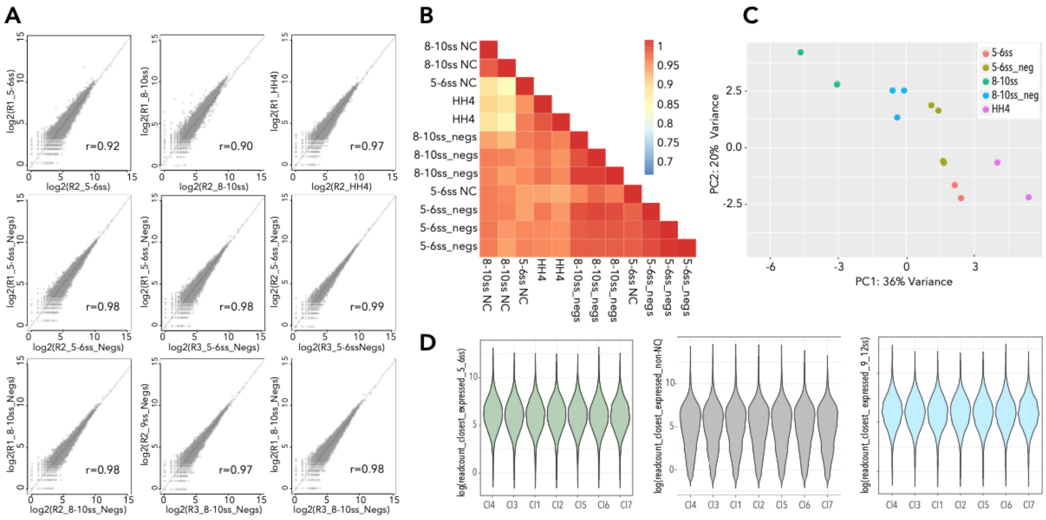

Figure 7.2 Appendix 2 - ATAC-seq quality control. (A) Scatter plots showing correlation of ATAC-seq replicas, r= Pearson correlation co-efficient. (B) Matrix presenting the correlation coefficients to all possible pairwise comparisons of replicates/samples. (C) PCA comparing NC and non-NC cells at both stages and HH4 ATAC-seq samples. (D) Violin plots showing correlation between k-Cluster elements and gene expression levels.

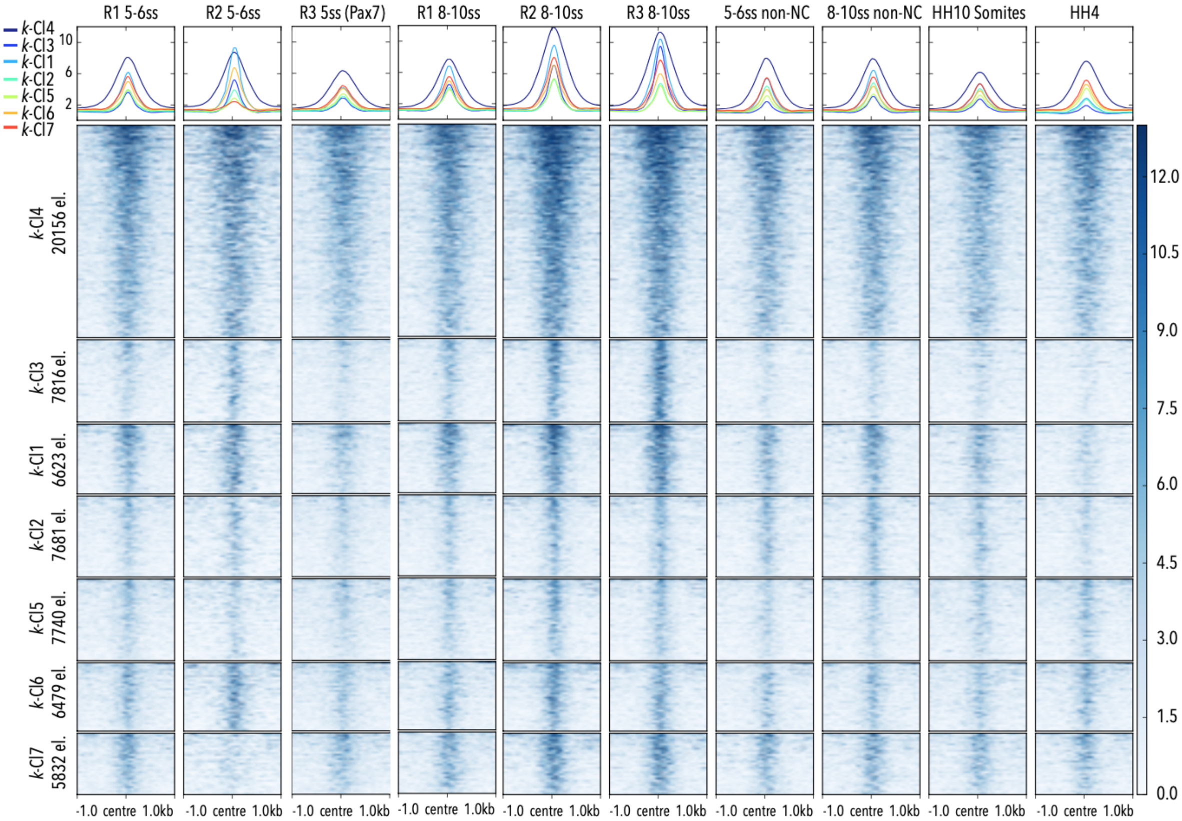

Figure 7.3 Appendix 3 - heatmap and merged profiles depicting *k*-means linear enrichment clustering of ATAC signal across all samples/stages analysed. *Pax7* sample is NC cells isolated using the Pax7 enh-195.

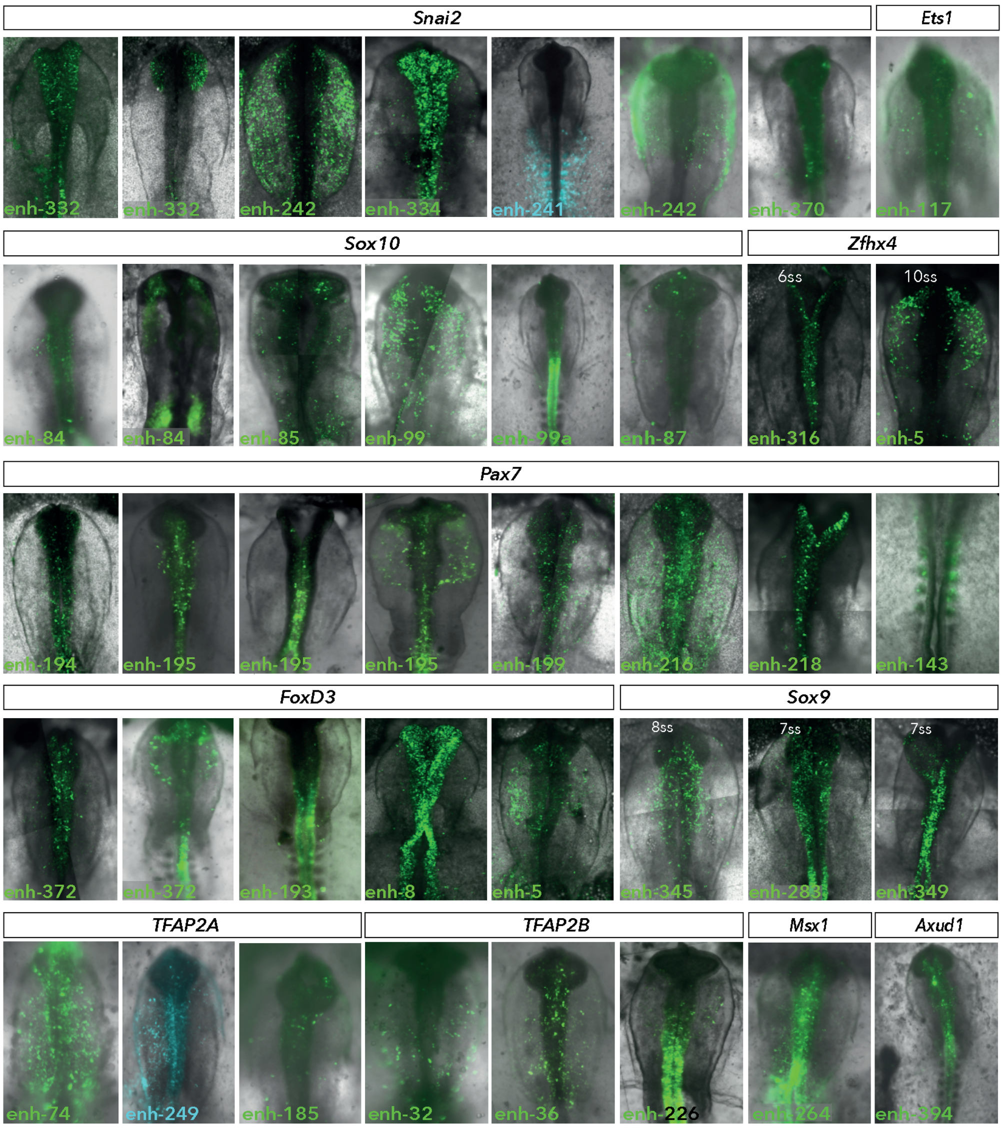

Figure 7.4 Appendix 4- *In vivo* activity of selected enhancers. Contributed by Dr. Ruth Williams.

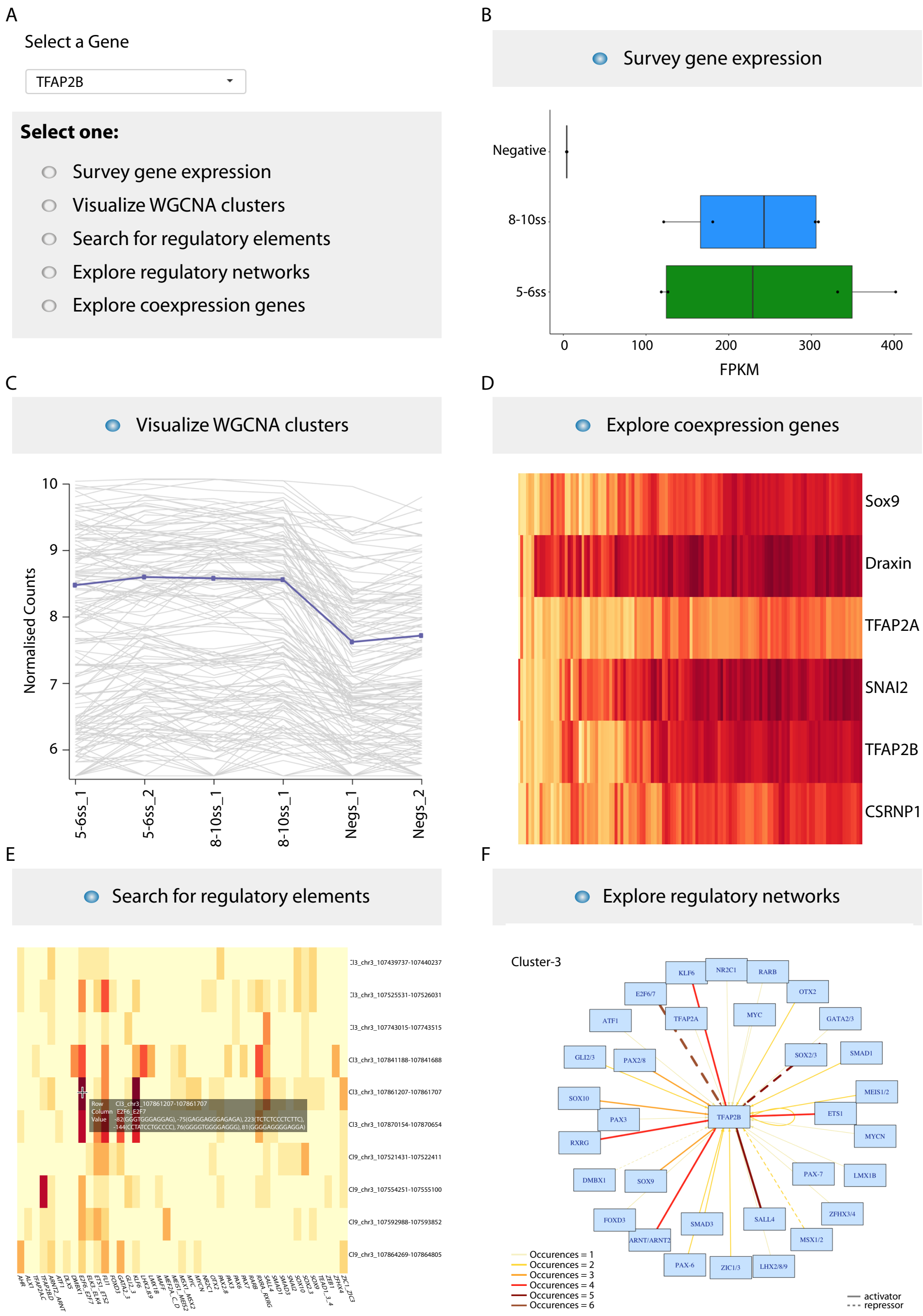

Figure 7.5 Shiny App Tool for Exploration of NC Gene Expression and Regulation. (A) Searchable interface for gene of interest (GOI). (B) Gene expression tab allows to explore levels (shown in FPKMs) as well as dynamics and differential gene expression stats across samples analyzed. (C) WGCNA cluster tab shows assignment to co-expression clusters. (D) Gene co-expression at the single-cell level is visualized in a searchable heatmap; features adjustable cell percentage overlap and multiple gene-name searches. (E) Regulatory tab enables exploration of assigned CREs (genome co-ordinates in galgal4) and inherent TF motifs (exact location within element provided). (F) Network tab enables visualization of individual regulatory circuits assembled based on either k-Cl3 and k-Cl1 enhancers and their respective TF inputs. TF occurrence frequency and activating/repressing interactions are shown.

Figure 7.6 Appendix 6- NC Gene Regulatory Circuitry Depicting TF Inputs, Inferred from Known Vertebrate TF Binding Models, into DA Elements at 5-6ss and Their Downstream Target Genes.

Figure 7.7 Appendix 7 - NC Gene Regulatory Circuitry Depicting TF Inputs, Inferred from Known Vertebrate TF Binding Models, into DA Elements at 8-10ss and Their Downstream Target Genes.

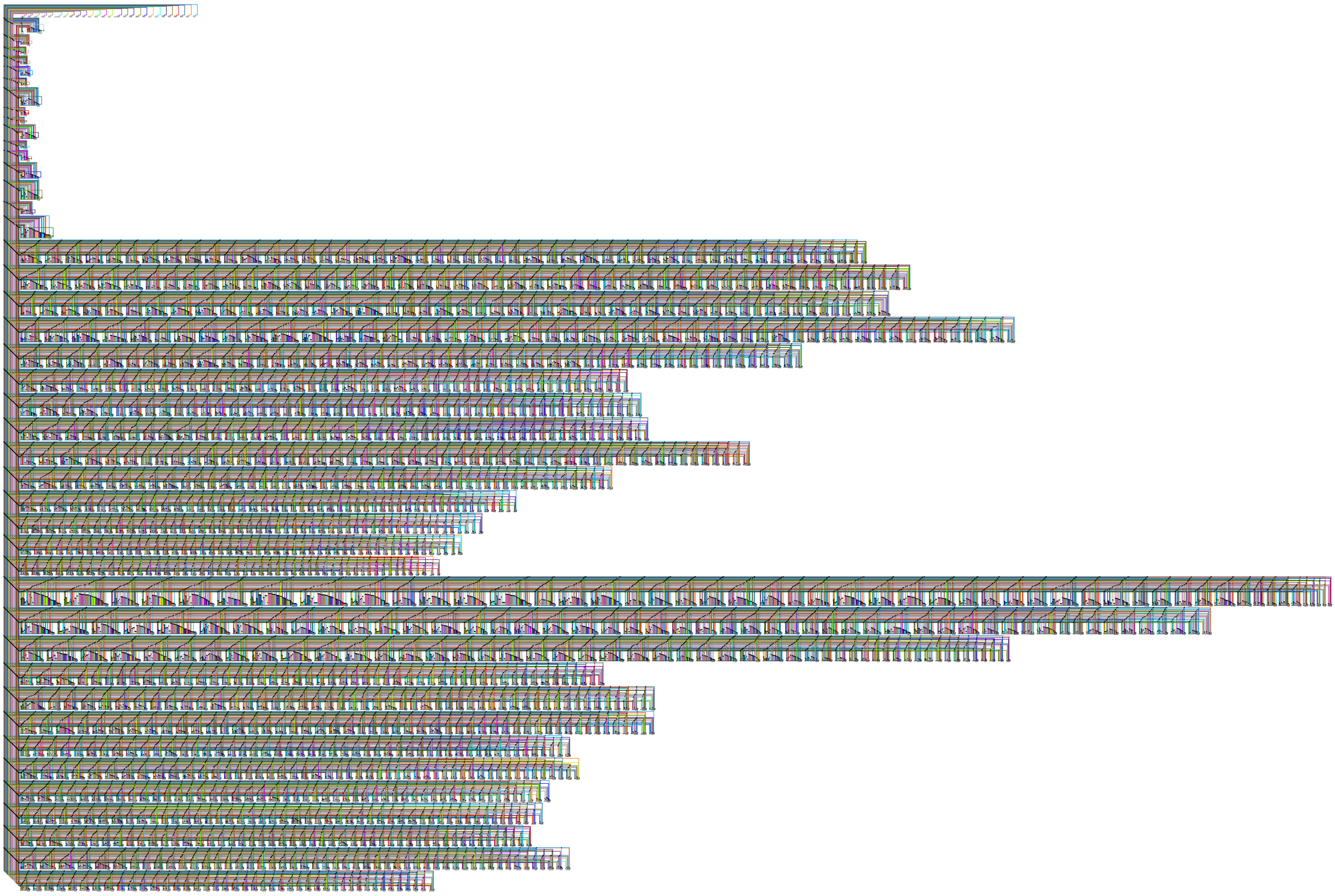

Figure 7.8 Appendix 8- NC Gene Regulatory Circuitry Depicting TF Inputs, Inferred from Known Vertebrate TF Binding Models, into *k*-Cluster 1 elements and Their Downstream Target Genes.

Figure 7.9 Appendix 8- NC Gene Regulatory Circuitry Depicting TF Inputs, Inferred from Known Vertebrate TF Binding Models, into *k*-Cluster 3 elements and Their Downstream Target Genes.

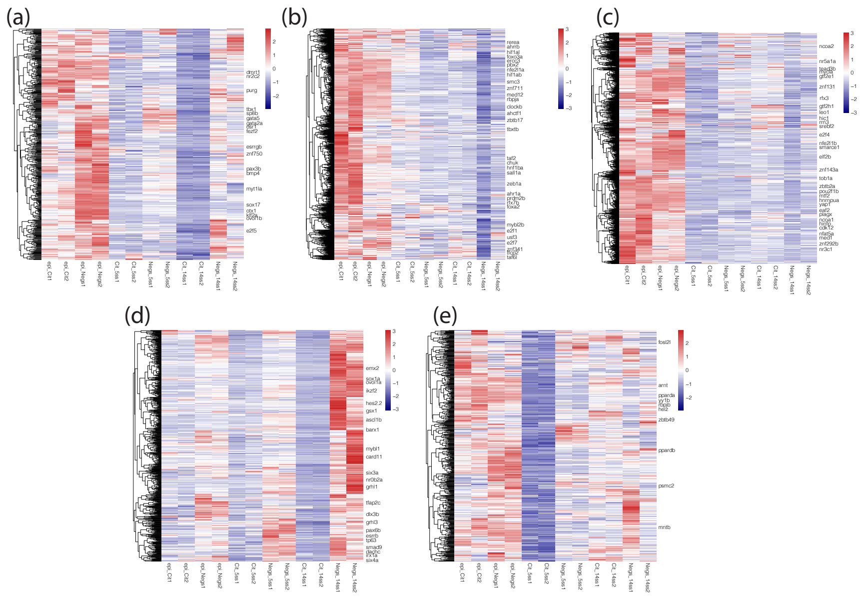

Figure 7.10 Appendix 10 - Additional WGCNA clusters identified during zebrafish NC development. Two replicates per stage for NC and non-NC cells are shown on the x axis.

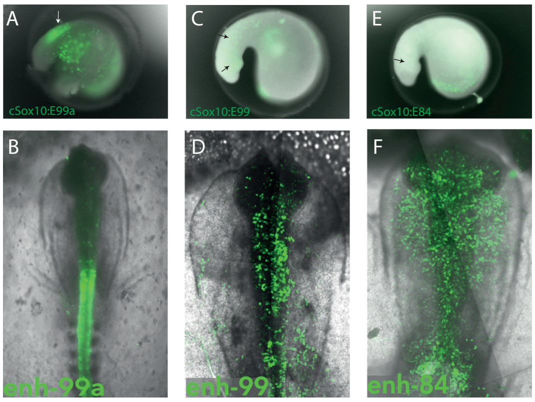

Figure 7.11 Appendix 10- Cross-species enhancer reporter assays. Chicken *Sox10* tested in marine lamprey embryos. (a) Sox10 E in T21 lamprey. The arrow indicate activity in the pre-migratory cranial NC. (b) Sox10 E in T21 lamprey. The arrow indicate activity in the migratory cranial NC. (c) Sox10 E in T21 lamprey. The arrow indicate the lack activity in NC. (d-f) Chicken embryos at HH9-10 displaying activity of the same enhancers tested in lampreys.

Bibliography

[1] N. Satoh, D. Rokhsar, and T. Nishikawa. Chordate evolution and the three-phylum system. *Proceedings of the Royal Society B: Biological Sciences*, 281(1794):20141729, 2014.

[2] D. Arendt, J. M. Musser, C. V. H. Baker, A. Bergman, C. Cepko, D. H. Erwin, M. Pavlicev, G. Schlosser, S. Widder, M. D. Laubichler, and G. P. Wagner. The origin and evolution of cell types. *Nature Reviews Genetics*, 17(12):744–757, 2016.

[3] D. Arendt. The evolution of cell types in animals: emerging principles from molecular studies. *Nature Reviews Genetics*, 9(11):868–882, 2008.

[4] D. H. Erwin and E. H. Davidson. The evolution of hierarchical gene regulatory networks. *Nature Reviews Genetics*, 10(2):141–148, 2009.

[5] M. Levine and R. Tjian. Transcription regulation and animal diversity. *Nature*, 424(6945):147–151, 2003.

[6] M. Lagha, J. P. Bothma, and M. Levine. Mechanisms of transcriptional precision in animal development. *Trends in Genetics*, 28(8):409–416, 2012.

[7] M. S. Halfon. Silencers, enhancers, and the multifunctional regulatory genome. *Trends in Genetics*, 36(3):149–151, 2020.

[8] F. Spitz and E. E. M. Furlong. Transcription factors: from enhancer binding to developmental control. *Nature Reviews Genetics*, 13(9):613–626, 2012.

[9] C. Gans and R. G. Northcutt. Neural crest and the origin of vertebrates: A new head. *Science*, 220(4594):268–273, 1983.

[10] M. L. Martik, S. Gandhi, B. R. Uy, J. A. Gillis, S. A. Green, M. Simoes-Costa, and M. E. Bronner. Evolution of the new head by gradual acquisition of neural crest regulatory circuits. *Nature*, 574(7780):675–678, 2019.

[11] R. A. Schneider. Neural crest and the origin of species-specific pattern. *Genesis*, 56(6-7):e23219, 2018.

[12] W. His. *Untersuchungen uber die erste Anlage des Wirbelthierleibes: Die erste Entwickelung des Huhnchens im Ei*. His, Wilhelm, 1831-1904 Untersuchungen uber die erste Anlage des Wirbelthierleibes. Leipzig,, 1868.

[13] A. M. Marshall. Memoirs: The development of the cranial nerves in the chick. *Quarterly Journal of Microscopical Science*, s2-18(69):10–40, 1878.

[14] J. B. Platt. Ectodermic origin of the cartilages of the head. *Anat. Anz.*, (7431):506–509, 1883.

[15] N. M. L. Douarin and E. Dupin. The “beginnings” of the neural crest. *Developmental Biology*, 444:S3–s13, 2018.

[16] C. S. Le Lievre and N. M. Le Douarin. Mesenchymal derivatives of the neural crest: analysis of chimaeric quail and chick embryos. *J Embryol Exp Morphol*, 34(1):125–54, 1975.

[17] E. Dupin and N. M. L. Douarin. Development of melanocyte precursors from the vertebrate neural crest. *Oncogene*, 22(20):3016–3023, 2003.

[18] N. M. L. Douarin and E. Dupin. The neural crest in vertebrate evolution. *Current Opinion in Genetics &amp Development*, 22(4):381–389, 2012.

[19] M. E. Bronner and N. M. LeDouarin. Development and evolution of the neural crest: An overview. *Developmental Biology*, 366(1):2–9, 2012.

[20] H. C. Etchevers, E. Dupin, and N. M. L. Douarin. The diverse neural crest: from embryology to human pathology. *Development*, 146(5), 2019.

[21] N. M. L. Douarin. Piecing together the vertebrate skull. *Development*, 139(23):4293–4296, 2012.

[22] E. Dupin, G. W. Calloni, J. M. Coelho-Aguiar, and N. M. L. Douarin. The issue of the multipotency of the neural crest cells. *Developmental Biology*, 444:S47–s59, 2018.

[23] G. Couly, P. Coltey, and N. L. Douarin. The triple origin of skull in higher vertebrates: a study in quail-chick chimeras. *Development*, 117(2):409–429, 1993.

[24] A. Baggiolini, S. Varum, J. M. Mateos, D. Bettosini, N. John, M. Bonalli, U. Ziegler, L. Dimou, H. Clevers, R. Furrer, and L. Sommer. Premigratory and migratory neural crest cells are multipotent in vivo. *Cell Stem Cell*, 16(3):314–322, 2015.

[25] R. Soldatov, M. Kaucka, M. E. Kastriti, J. Petersen, T. Chontorotzea, L. Englmaier, N. Akkuratova, Y. Yang, M. Häring, V. Dyachuk, C. Bock, M. Farlik, M. L. Piacentino, F. Boismoreau, M. M. Hilscher, C. Yokota, X. Qian, M. Nilsson, M. E. Bronner, L. Croci, W.-Y. Hsiao, D. A. Guertin, J.-F. Brunet, G. G. Consalez, P. Ernfors, K. Fried, P. V. Kharchenko, and I. Adameyko. Spatiotemporal structure of cell fate decisions in murine neural crest. *Science*, 364(6444), 2019.

[26] M. Bronner-Fraser and S. E. Fraser. Cell lineage analysis reveals multipotency of some avian neural crest cells. *Nature*, 335(6186):161–164, 1988.

[27] M. Bronner-Fraser and S. Fraser. Developmental potential of avian trunk neural crest cells in situ. *Neuron*, 3(6):755–766, 1989.

[28] W. Tang and M. E. Bronner. Neural crest lineage analysis: from past to future trajectory. *Development*, 147(20), 2020.

[29] G. Couly, A. Grapin-Botton, P. Coltey, and N. L. Douarin. The regeneration of the cephalic neural crest, a problem revisited: the regenerating cells originate from the contralateral or from the anterior and posterior neural fold. *Development*, 122(11):3393–3407, 1996.

[30] S. Fraser and M. Bronner-Fraser. Migrating neural crest cells in the trunk of the avian embryo are multipotent. *Development*, 112(4):913–920, 1991.

[31] P. Betancur, M. Bronner-Fraser, and T. Sauka-Spengler. Assembling neural crest regulatory circuits into a gene regulatory network. *Annual Review of Cell and Developmental Biology*, 26(1):581–603, 2010.

[32] T. Sauka-Spengler and M. Bronner-Fraser. A gene regulatory network orchestrates neural crest formation. *Nature Reviews Molecular Cell Biology*, 9(7):557–568, 2008.

[33] M. Simões-Costa and M. E. Bronner. Establishing neural crest identity: a gene regulatory recipe. *Development*, 142(2):242–257, 2015.

[34] C. V. Baker. The evolution and elaboration of vertebrate neural crest cells. *Current Opinion in Genetics &amp Development*, 18(6):536–543, 2008.

[35] N. M. Young, D. Hu, A. J. Lainoff, F. J. Smith, R. Diaz, A. S. Tucker, P. A. Trainor, R. A. Schneider, B. Hallgrímsson, and R. S. Marcucio. Embryonic bauplans and the developmental origins of facial diversity and constraint. *Development*, 141(5):1059–1063, 2014.

[36] J. R. York and D. W. McCauley. The origin and evolution of vertebrate neural crest cells. *Open Biology*, 10(1):190285, 2020.

[37] C. LaBonne and M. Bronner-Fraser. Neural crest induction in xenopus: evidence for a two-signal model. *Development*, 125(13):2403–2414, 1998.

[38] M. S. Prasad, T. Sauka-Spengler, and C. LaBonne. Induction of the neural crest state: Control of stem cell attributes by gene regulatory, post-transcriptional and epigenetic interactions. *Developmental Biology*, 366(1):10–21, 2012.

[39] J. Stundl, P. Y. Bertucci, A. Lauri, D. Arendt, and M. E. Bronner. Chapter six - evolution of new cell types at the lateral neural border. *Current Topics in Developmental Biology*, 141:173–205, 2021.

[40] S. Krispin, E. Nitzan, and C. Kalcheim. The dorsal neural tube: a dynamic setting for cell fate decisions. *Dev Neurobiol*, 70(12):796–812, 2010.

[41] S. E. M. E. B. Daniela Roellig, Johanna Tan-Cabugao. Dynamic transcriptional signature and cell fate analysis reveals plasticity of individual neural plate border cell. *eLife*, 6:e21620, 2017.

[42] B. Steventon, R. Mayor, and A. Streit. Neural crest and placode interaction during the development of the cranial sensory system. *Developmental Biology*, 389(1):28–38, 2014.

[43] E. Theveneau and R. Mayor. Neural crest delamination and migration: From epithelium-to-mesenchyme transition to collective cell migration. *Developmental Biology*, 366(1):34–54, 2012.

[44] M. Minoux and F. M. Rijli. Molecular mechanisms of cranial neural crest cell migration and patterning in craniofacial development. *Development*, 137(16):2605–2621, 2010.

[45] C. L. Lievre, G. Schweizer, C. Ziller, and N. L. Douarin. Restrictions of developmental capabilities in neural crest cell derivatives as tested by in vivo transplantation experiments. *Developmental Biology*, 77(2):362–378, 1980.

[46] S. Krispin, E. Nitzan, Y. Kassem, and C. Kalcheim. Evidence for a dynamic spatiotemporal fate map and early fate restrictions of premigratory avian neural crest. *Development*, 137(4):585–595, 2010.

[47] S. A. Green, M. Simoes-Costa, and M. E. Bronner. Evolution of vertebrates as viewed from the crest. *Nature*, 520(7548):474–482, 2015.

[48] D. Jandzik, A. T. Garnett, T. A. Square, M. V. Cattell, J.-K. Yu, and D. M. Medeiros. Evolution of the new vertebrate head by co-option of an ancient chordate skeletal tissue. *Nature*, 518(7540):534–537, 2014.

[49] A. S. Hovland, M. Rothstein, and M. Simoes-Costa. Network architecture and regulatory logic in neural crest development. *WIREs Systems Biology and Medicine*, 12(2), 2019.

[50] A. Rada-Iglesias, R. Bajpai, S. Prescott, S. A. Brugmann, T. Swigut, and J. Wysocka. Epigenomic annotation of enhancers predicts transcriptional regulators of human neural crest. *Cell Stem Cell*, 11(5):633–648, 2012.

[51] R. Bajpai, D. A. Chen, A. Rada-Iglesias, J. Zhang, Y. Xiong, J. Helms, C.-P. Chang, Y. Zhao, T. Swigut, and J. Wysocka. CHD7 cooperates with PBAF to control multipotent neural crest formation. *Nature*, 463(7283):958–962, 2010.

[52] L. S. Gammill and M. Bronner-Fraser. Neural crest specification: migrating into genomics. *Nature Reviews Neuroscience*, 4(10):795–805, 2003.

[53] L. S. Gammill and M. Bronner-Fraser. Genomic analysis of neural crest induction. *Development*, 129(24):5731–5741, 2002.

[54] M. Simões-Costa, M. Stone, and M. E. Bronner. Axud1 integrates wnt signaling and transcriptional inputs to drive neural crest formation. *Developmental Cell*, 34(5):544–554, 2015.

[55] M. Simões-Costa, J. Tan-Cabugao, I. Antoshechkin, T. Sauka-Spengler, and M. E. Bronner. Transcriptome analysis reveals novel players in the cranial neural crest gene regulatory network. *Genome Research*, 24(2):281–290, 2014.

[56] M. Simoes-Costa and M. E. Bronner. Reprogramming of avian neural crest axial identity and cell fate. *Science*, 352(6293):1570–1573, 2016.

[57] M. Simões-Costa and M. E. Bronner. Insights into neural crest development and evolution from genomic analysis. *Genome Research*, 23(7):1069–1080, 2013.

[58] I. T. C. Ling and T. Sauka-Spengler. Early chromatin shaping predetermines multipotent vagal neural crest into neural, neuronal and mesenchymal lineages. *Nature Cell Biology*, 21(12):1504–1517, 2019.

[59] R. M. Williams, I. Candido-Ferreira, E. Repapi, D. Gavriouchkina, U. Senanayake, I. T. Ling, J. Telenius, S. Taylor, J. Hughes, and T. Sauka-Spengler. Reconstruction of the global neural crest gene regulatory network in vivo. *Developmental Cell*, 51(2):255–276.e7, 2019.

[60] M. Lukoseviciute, D. Gavriouchkina, R. M. Williams, T. Hochgreb-Hagele, U. Senanayake, V. Chong-Morrison, S. Thongjuea, E. Repapi, A. Mead, and T. Sauka-Spengler. From pioneer to repressor: Bimodal foxd3 activity dynamically remodels neural crest regulatory landscape in vivo. *Developmental Cell*, 47(5):608–628.e6, 2018.

[61] L. A. Trinh, V. Chong-Morrison, D. Gavriouchkina, T. Hochgreb-Hägele, U. Senanayake, S. E. Fraser, and T. Sauka-Spengler. Biotagging of specific

cell populations in zebrafish reveals gene regulatory logic encoded in the nuclear transcriptome. *Cell Reports*, 19(2):425–440, 2017.

[62] M. L. Martik and M. E. Bronner. Regulatory logic underlying diversification of the neural crest. *Trends in Genetics*, 33(10):715–727, 2017.

[63] E. Nitzan, S. Krispin, E. R. Pfaltzgraff, A. Klar, P. A. Labosky, and C. Kalcheim. A dynamic code of dorsal neural tube genes regulates the segregation between neurogenic and melanogenic neural crest cells. *Development*, 140(11):2269–2279, 2013.

[64] P. M. Kulesa, C. M. Bailey, J. C. Kasemeier-Kulesa, and R. McLennan. Cranial neural crest migration: New rules for an old road. *Developmental Biology*, 344(2):543–554, 2010.

[65] I. Adameyko, F. Lallemend, J. B. Aquino, J. A. Pereira, P. Topilko, T. Müller, N. Fritz, A. Beljajeva, M. Mochii, I. Liste, D. Usoskin, U. Suter, C. Birchmeier, and P. Ernfors. Schwann cell precursors from nerve innervation are a cellular origin of melanocytes in skin. *Cell*, 139(2):366–379, 2009.

[66] J. A. Briggs, C. Weinreb, D. E. Wagner, S. Megason, L. Peshkin, M. W. Kirschner, and A. M. Klein. The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. *Science*, 360(6392):eaar5780, 2018.

[67] A. Zalc, R. Sinha, G. S. Gulati, D. J. Wesche, P. Daszczuk, T. Swigut, I. L. Weissman, and J. Wysocka. Reactivation of the pluripotency program precedes formation of the cranial neural crest. *Science*, 371(6529), 2021.

[68] M. Rocha, A. Beiriger, E. E. Kushkowski, T. Miyashita, N. Singh, V. Venkataraman, and V. E. Prince. From head to tail: regionalization of the neural crest. *Development*, 147(20), 2020.

[69] M. Rothstein, D. Bhattacharya, and M. Simoes-Costa. The molecular basis of neural crest axial identity. *Developmental Biology*, 444:S170–s180, 2018.

[70] R. Horie, A. Hazbun, K. Chen, C. Cao, M. Levine, and T. Horie. Shared evolutionary origin of vertebrate neural crest and cranial placodes. *Nature*, 560(7717):228–232, 2018.

[71] B. Li, S. Kuriyama, M. Moreno, and R. Mayor. The posteriorizing gene *Gbx2* is a direct target of Wnt signalling and the earliest factor in neural crest induction. *Development*, 136(19):3267–3278, 2009.

[72] R. Neijts, S. Amin, C. van Rooijen, and J. Deschamps. Cdx is crucial for the timing mechanism driving colinear hox activation and defines a trunk segment in the hox cluster topology. *Dev Biol*, 422(2):146–154, 2017.

[73] J. G. Savory, N. Bouchard, V. Pierre, F. M. Rijli, Y. De Repentigny, R. Kothary, and D. Lohnes. Cdx2 regulation of posterior development through non-hox targets. *Development*, 136(24):4099–110, 2009.

[74] A. Gavalas, M. Studer, A. Lumsden, F. M. Rijli, R. Krumlauf, and P. Chambon. Hoxa1 and hoxb1 synergize in patterning the hindbrain, cranial nerves and second pharyngeal arch. *Development*, 125(6):1123–36, 1998.

[75] A. Gavalas, P. Trainor, L. Ariza-McNaughton, and R. Krumlauf. Synergy between hoxa1 and hoxb1: the relationship between arch patterning and the generation of cranial neural crest. *Development*, 128(15):3017–27, 2001.

[76] S. Creuzet, G. Couly, C. Vincent, and N. M. Le Douarin. Negative effect of hox gene expression on the development of the neural crest-derived facial skeleton. *Development*, 129(18):4301–4313, 2002.

[77] E. Buitrago-Delgado, K. Nordin, A. Rao, L. Geary, and C. LaBonne. Shared regulatory programs suggest retention of blastula-stage potential in neural crest cells. *Science*, 348(6241):1332–1335, 2015.

[78] P. Scerbo and A. H. Monsoro-Burq. The vertebrate-specific VENTX/NANOG gene empowers neural crest with ectomesenchyme potential. *Science Advances*, 6(18), 2020.

[79] M. Minoux, S. Holwerda, A. Vitobello, T. Kitazawa, H. Kohler, M. B. Stadler, and F. M. Rijli. Gene bivalency at polycomb domains regulates cranial neural crest positional identity. *Science*, 355(6332), 2017.

[80] P. Voigt, W.-W. Tee, and D. Reinberg. A double take on bivalent promoters. *Genes & Development*, 27(12):1318–1338, 2013.

[81] S. G. Cox, H. Kim, A. T. Garnett, D. M. Medeiros, W. An, and J. G. Crump. An essential role of variant histone h3.3 for ectomesenchyme potential of the cranial neural crest. *PLoS Genetics*, 8(9):e1002938, 2012.

[82] C. Miller, T. Schilling, K. Lee, J. Parker, and C. Kimmel. sucker encodes a zebrafish endothelin-1 required for ventral pharyngeal arch development. *Development*, 127(17):3815–3828, 2000.

[83] M. Mandl and R. Depping. Hypoxia-inducible aryl hydrocarbon receptor nuclear translocator (ARNT) (HIF-1): Is it a rare exception? *Molecular Medicine*, 20(1):215–220, 2014.

[84] V. Compernolle. Cardia bifida, defective heart development and abnormal neural crest migration in embryos lacking hypoxia-inducible factor-1. *Cardiovascular Research*, 60(3):569–579, 2003.

[85] E. H. Barriga, P. H. Maxwell, A. E. Reyes, and R. Mayor. The hypoxia factor hif-1 controls neural crest chemotaxis and epithelial to mesenchymal transition. *Journal of Cell Biology*, 201(5):759–776, 2013.

[86] C. U. Niklasson, E. Fredlund, E. Monni, J. M. Lindvall, Z. Kokaia, E. U. Hammarlund, M. E. Bronner, and S. Mohlin. Hypoxia inducible factor-2 importance for migration, proliferation, and self-renewal of trunk neural crest cells. *Developmental Dynamics*, 250(2):191–236, 2020.

[87] E. Rodriguez Carballo, L. Lopez-Delisle, N. Yakushiji-Kaminatsui, A. Ullate-Agote, and D. Duboule. Impact of genome architecture on the functional activation and repression of hox regulatory landscapes. *BMC Biol*, 17(1):55, 2019.

[88] E. Rodríguez-Carballo, L. Lopez-Delisle, A. Willemin, L. Beccari, S. Gitto, B. Mascrez, and D. Duboule. Chromatin topology and the timing of enhancer function at the iHoxD/i locus. *Proceedings of the National Academy of Sciences*, 117(49):31231–31241, 2020.

[89] L. Kerosuo and M. E. Bronner. cMyc regulates the size of the premigratory neural crest stem cell pool. *Cell Reports*, 17(10):2648–2659, 2016.

[90] D. A. Ridenour, R. McLennan, J. M. Teddy, C. L. Semerad, J. S. Haug, and P. M. Kulesa. The neural crest cell cycle is related to phases of migration in the head. *Development*, 141(5):1095–1103, 2014.

[91] A. Dady, C. Blavet, and J. L. Duband. Timing and kinetics of e- to n-cadherin switch during neurulation in the avian embryo. *Dev Dyn*, 241(8):1333–49, 2012.

[92] M. Cheung, M. C. Chaboissier, A. Mynett, E. Hirst, A. Schedl, and

J. Briscoe. The transcriptional control of trunk neural crest induction, survival, and delamination. *Dev Cell*, 8(2):179–92, 2005.

[93] S. J. McKeown, A. S. Wallace, and R. B. Anderson. Expression and function of cell adhesion molecules during neural crest migration. *Dev Biol*, 373(2):244–57, 2013.

[94] A. Rinon, A. Molchadsky, E. Nathan, G. Yovel, V. Rotter, R. Sarig, and E. Tzahor. p53 coordinates cranial neural crest cell growth and epithelial-mesenchymal transition/delamination processes. *Development*, 138(9):1827–38, 2011.

[95] E. Théveneau, J.-L. Duband, and M. Altabef. Ets-1 confers cranial features on neural crest delamination. *PLoS ONE*, 2(11):e1142, 2007.

[96] A. Szabó, E. Theveneau, M. Turan, and R. Mayor. Neural crest streaming as an emergent property of tissue interactions during morphogenesis. *PLOS Computational Biology*, 15(4):e1007002, 2019.

[97] T. A. Square, D. Jandzik, J. L. Massey, M. Romášek, H. P. Stein, A. W. Hansen, A. Purkayastha, M. V. Cattell, and D. M. Medeiros. Evolution of the endothelin pathway drove neural crest cell diversification. *Nature*, 585(7826):563–568, 2020.

[98] M. Arnone and E. Davidson. The hardwiring of development: organization and function of genomic regulatory systems. *Development*, 124(10):1851–1864, 1997.

[99] I. S. Peter and E. H. Davidson. Modularity and design principles in the sea urchin embryo gene regulatory network. *FEBS Letters*, 583(24):3948–3958, 2009.

[100] E. H. Davidson, J. P. Rast, P. Oliveri, A. Ransick, C. Calestani, C.-H. Yuh, T. Minokawa, G. Amore, V. Hinman, C. Arenas-Mena, O. Otim, C. T. Brown, C. B. Livi, P. Y. Lee, R. Revilla, A. G. Rust, Z. jun Pan, M. J. Schilstra, P. J. C. Clarke, M. I. Arnone, L. Rowen, R. A. Cameron, D. R. McClay, L. Hood, and H. Bolouri. A genomic regulatory network for development. *Science*, 295(5560):1669–1678, 2002.

[101] M. Levine and E. H. Davidson. Gene regulatory networks for development. 102(14):4936–4942, 2005.

[102] P. B. Talbert, M. P. Meers, and S. Henikoff. Old cogs, new tricks: the evolution of gene expression in a chromatin context. *Nature Reviews Genetics*, 20(5):283–297, 2019.

[103] D. Shlyueva, G. Stampfel, and A. Stark. Transcriptional enhancers: from properties to genome-wide predictions. *Nature Reviews Genetics*, 15(4):272–286, 2014.

[104] H. K. Long, S. L. Prescott, and J. Wysocka. Ever-changing landscapes: Transcriptional enhancers in development and evolution. *Cell*, 167(5):1170–1187, 2016.

[105] N. D. Heintzman, R. K. Stuart, G. Hon, Y. Fu, C. W. Ching, R. D. Hawkins, L. O. Barrera, S. Van Calcar, C. Qu, K. A. Ching, W. Wang, Z. Weng, R. D. Green, G. E. Crawford, and B. Ren. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet*, 39(3):311–8, 2007.

[106] A. Eberharter and P. B. Becker. Histone acetylation: a switch between repressive and permissive chromatin. second in review series on chromatin dynamics. *EMBO Rep*, 3(3):224–9, 2002.

[107] N. T. Crump and T. A. Milne. Why are so many mll lysine methyltransferases required for normal mammalian development? *Cell Mol Life Sci*, 76(15):2885–2898, 2019.

[108] M. Levine, C. Cattoglio, and R. Tjian. Looping back to leap forward: Transcription enters a new era. *Cell*, 157(1):13–25, 2014.

[109] W. de Laat and D. Duboule. Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature*, 502(7472):499–506, 2013.

[110] B. Lim, T. Heist, M. Levine, and T. Fukaya. Visualization of transvection in living drosophila embryos. *Molecular Cell*, 70(2):287–296.e6, 2018.

[111] R. Margueron and D. Reinberg. The polycomb complex prc2 and its mark in life. *Nature*, 469(7330):343–9, 2011.

[112] M. P. Creyghton, A. W. Cheng, G. G. Welstead, T. Kooistra, B. W. Carey, E. J. Steine, J. Hanna, M. A. Lodato, G. M. Frampton, P. A. Sharp, L. A. Boyer, R. A. Young, and R. Jaenisch. Histone h3k27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A*, 107(50):21931–6, 2010.

[113] A. Visel, E. M. Rubin, and L. A. Pennacchio. Genomic views of distant-acting enhancers. *Nature*, 461(7261):199–205, 2009.

[114] D. Schmidt, M. D. Wilson, B. Ballester, P. C. Schwalie, G. D. Brown, A. Marshall, C. Kutter, S. Watt, C. P. Martinez-Jimenez, S. Mackay, I. Talianidis, P. Flicek, and D. T. Odom. Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. 328:6, 2010.

[115] P. J. Skene and S. Henikoff. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife*, 6, 2017.

[116] M. Rothstein and M. Simoes-Costa. Heterodimerization of TFAP2 pioneer factors drives epigenomic remodeling during neural crest specification. *Genome Research*, 30(1):35–48, 2019.

[117] A. P. Azambuja and M. Simoes-Costa. The connectome of neural crest enhancers reveals regulatory features of signaling systems. *Developmental Cell*, 56(9):1268–1282.e6, 2021.

[118] C. D. Arnold, D. Gerlach, C. Stelzer, L. M. Boryń, M. Rath, and A. Stark. Genome-wide quantitative enhancer activity maps identified by starr-seq. *Science*, 339(6123):1074–1077, 2013.

[119] J. Banerji, S. Rusconi, and W. Schaffner. Expression of a beta-globin gene is enhanced by remote sv40 dna sequences. *Cell*, 27(2 Pt 1):299–308, 1981.

[120] C. D. Arnold, D. Gerlach, D. Spies, J. A. Matts, Y. A. Sytnikova, M. Pagani, N. C. Lau, and A. Stark. Quantitative genome-wide enhancer activity maps for five drosophila species show functional enhancer conservation and turnover during cis-regulatory evolution. *Nature Genetics*, 46(7):685–692, 2014.

[121] D. Hnisz, B. J. Abraham, T. I. Lee, A. Lau, V. Saint-André, A. A. Sigova, H. A. Hoke, and R. A. Young. Super-enhancers in the control of cell identity and disease. *Cell*, 155(4):934–947, 2013.

[122] Z. Wunderlich, M. D. Bragdon, B. J. Vincent, J. A. White, J. Estrada, and A. H. DePace. Krüppel expression levels are maintained through compensatory evolution of shadow enhancers. *Cell Reports*, 12(11):1740–1747, 2015.

[123] J.-W. Hong, D. A. Hendrix, and M. S. Levine. Shadow enhancers as a source of evolutionary novelty. *Science*, 321(5894):1314–1314, 2008.

[124] E. Cannavò, P. Khoueiry, D. A. Garfield, P. Geeleher, T. Zichner, E. H. Gustafson, L. Ciglar, J. O. Korbel, and E. E. Furlong. Shadow enhancers are pervasive features of developmental regulatory networks. *Current Biology*, 26(1):38–51, 2016.

[125] M. W. Perry, A. N. Boettiger, J. P. Bothma, and M. Levine. Shadow enhancers foster robustness of drosophila gastrulation. *Current Biology*, 20(17):1562–1567, 2010.

[126] J. Crocker, G. R. Ilsley, and D. L. Stern. Quantitatively predictable control of drosophila transcriptional enhancers in vivo with engineered transcription factors. *Nature genetics*, 48(3):292:298, 2016.

[127] W. A. Whyte, D. A. Orlando, D. Hnisz, B. J. Abraham, C. Y. Lin, M. H. Kagey, P. B. Rahl, T. I. Lee, and R. A. Young. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell*, 153(2):307–319, 2013.

[128] J. Lovén, H. A. Hoke, C. Y. Lin, A. Lau, D. A. Orlando, C. R. Vakoc, J. E. Bradner, T. I. Lee, and R. A. Young. Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell*, 153(2):320–334, 2013.

[129] D. Hay, J. R. Hughes, C. Babbs, J. O. J. Davies, B. J. Graham, L. L. P. Hanssen, M. T. Kassouf, A. M. Oudelaar, J. A. Sharpe, M. C. Suciu, J. Telenius, R. Williams, C. Rode, P.-S. Li, L. A. Pennacchio, J. A. Sloane-Stanley, H. Ayyub, S. Butler, T. Sauka-Spengler, R. J. Gibbons, A. J. H. Smith, W. G. Wood, and D. R. Higgs. Genetic dissection of the -globin super-enhancer in vivo. *Nature Genetics*, 48(8):895–903, 2016.

[130] N. Dukler, B. Gulko, Y.-F. Huang, and A. Siepel. Is a super-enhancer greater than the sum of its parts? *Nature Genetics*, 49(1):2–3, 2017.

[131] T. Werner, A. Hammer, M. Wahlbuhl, M. R. Bösl, and M. Wegner. Multiple conserved regulatory elements with overlapping functions determine sox10 expression in mouse embryogenesis. *Nucleic Acids Research*, 35(19):6526–6538, 2007.

[132] J. P. Bothma, H. G. Garcia, S. Ng, M. W. Perry, T. Gregor, and M. Levine. Enhancer additivity and non-additivity are determined by enhancer strength in the drosophila embryo. *eLife*, 4, 2015.

[133] J. O. Yáñez-Cuna, E. Z. Kvon, and A. Stark. Deciphering the transcriptional cis-regulatory code. *Trends in Genetics*, 29(1):11–22, 2013.

[134] A. I. Ramos and S. Barolo. Low-affinity transcription factor binding sites shape morphogen responses and enhancer evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368(1632):20130018, 2013.

[135] J. Crocker, N. Abe, L. Rinaldi, A. P. McGregor, N. Frankel, S. Wang, A. Alsawadi, P. Valenti, S. Plaza, F. Payre, R. S. Mann, and D. L. Stern. Low affinity binding site clusters confer hox specificity and regulatory robustness. *Cell*, 160(1-2):191–203, 2015.

[136] J. Crocker, E. P.-B. Noon, and D. L. Stern. The soft touch. In *Current Topics in Developmental Biology*, pages 455–469. 2016.

[137] E. K. Farley, K. M. Olson, W. Zhang, A. J. Brandt, D. S. Rokhsar, and M. S. Levine. Suboptimization of developmental enhancers. *Science*, 350(6258):325–328, 2015.

[138] E. K. Farley, K. M. Olson, W. Zhang, D. S. Rokhsar, and M. S. Levine. Syntax compensates for poor binding sites to encode tissue specificity of developmental enhancers. *Proceedings of the National Academy of Sciences*, 113(23):6508–6513, 2016.

[139] B. P. de Almeida, F. Reiter, M. Pagani, and A. Stark. Deepstarr predicts enhancer activity from dna sequence and enables the de novo design of enhancers. *bioRxiv*, 2021.

[140] D. M. Suter, N. Molina, D. Gatfield, K. Schneider, U. Schibler, and F. Naef. Mammalian genes are transcribed with widely different bursting kinetics. *Science*, 332(6028):472–474, 2011.

[141] T. Fukaya, B. Lim, and M. Levine. Enhancer control of transcriptional bursting. *Cell*, 166(2):358–368, 2016.

[142] R. P. Zinzen, C. Girardot, J. Gagneur, M. Braun, and E. E. M. Furlong. Combinatorial binding predicts spatio-temporal cis-regulatory activity. *Nature*, 462(7269):65–70, 2009.

[143] J. Zhu, S. Palliyil, C. Ran, and J. P. Kumar. *Drosophila* Pax6 promotes development of the entire eye-antennal disc, thereby ensuring proper adult head formation. *Proceedings of the National Academy of Sciences*, 114(23):5846–5853, 2017.

[144] Y.-S. Lin, M. Carey, M. Ptashne, and M. R. Green. How different eukaryotic transcriptional activators can cooperate promiscuously. *Nature*, 345(6273):359–361, 1990.

[145] J. He, J. Ye, Y. Cai, C. Riquelme, J. O. Liu, X. Liu, A. Han, and L. Chen. Structure of p300 bound to MEF2 on DNA reveals a mechanism of enhanceosome assembly. *Nucleic Acids Research*, 39(10):4464–4474, 2011.

[146] C. D. Brown, D. S. Johnson, and A. Sidow. Functional architecture and evolution of transcriptional elements that drive gene coexpression. *Science*, 317(5844):1557–1560, 2007.

[147] M. Carey. The enhanceosome and transcriptional synergy. *Cell*, 92(1):5–8, 1998.

[148] D. Thanos and T. Maniatis. Virus induction of human IFN gene expression requires the assembly of an enhanceosome. *Cell*, 83(7):1091–1100, 1995.

[149] M. Levine. Transcriptional enhancers in animal development and evolution. *Current Biology*, 20(17):R754–r763, 2010.

[150] D. Arnosti, S. Barolo, M. Levine, and S. Small. The eve stripe 2 enhancer employs multiple modes of transcriptional synergy. *Development*, 122(1):205–214, 1996.

[151] G. Stampfel, T. Kazmar, O. Frank, S. Wienerroither, F. Reiter, and A. Stark. Transcriptional regulators form diverse groups with context-dependent regulatory functions. *Nature*, 528(7580):147–151, 2015.

[152] O. Symmons, L. Pan, S. Remeseiro, T. Aktas, F. Klein, W. Huber, and F. Spitz. The shh topological domain facilitates the action of remote enhancers by reducing the effects of genomic distances. *Developmental Cell*, 39(5):529–543, 2016.

[153] A. S. Weintraub, C. H. Li, A. V. Zamudio, A. A. Sigova, N. M. Hannett, D. S. Day, B. J. Abraham, M. A. Cohen, B. Nabet, D. L. Buckley, Y. E. Guo, D. Hnisz, R. Jaenisch, J. E. Bradner, N. S. Gray, and R. A. Young. YY1 is a structural regulator of enhancer-promoter loops. *Cell*, 171(7):1573–1588.e28, 2017.

[154] E. de Wit, E. S. Vos, S. J. Holwerda, C. Valdes-Quezada, M. J. Verstegen, H. Teunissen, E. Splinter, P. J. Wijchers, P. H. Krijger, and W. de Laat. CTCF binding polarity determines chromatin looping. *Molecular Cell*, 60(4):676–684, 2015.

[155] A. L. Sanborn, S. S. Rao, S. C. Huang, N. C. Durand, M. H. Huntley, A. I. Jewett, I. D. Bochkov, D. Chinnappan, A. Cutkosky, J. Li, K. P. Geeting, A. Gnirke, A. Melnikov, D. McKenna, E. K. Stamenova, E. S. Lander, and E. L. Aiden. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc Natl Acad Sci U S A*, 112(47):E6456–65, 2015.

[156] G. Fudenberg, M. Imakaev, C. Lu, A. Goloborodko, N. Abdennur, and L. A. Mirny. Formation of chromosomal domains by loop extrusion. *Cell Rep*, 15(9):2038–49, 2016.

[157] J. H. I. Haarhuis, R. H. van der Weide, V. A. Blomen, J. O. Yanez-Cuna, M. Amendola, M. S. van Ruiten, P. H. L. Krijger, H. Teunissen, R. H. Medema, B. van Steensel, T. R. Brummelkamp, E. de Wit, and B. D. Rowland. The cohesin release factor wapl restricts chromatin loop extension. *Cell*, 169(4):693–707.e14, 2017.

[158] V. C. Seitan, A. J. Faure, Y. Zhan, R. P. McCord, B. R. Lajoie, E. Ing-Simmons, B. Lenhard, L. Giorgetti, E. Heard, A. G. Fisher, P. Flicek, J. Dekker, and M. Merkenschlager. Cohesin-based chromatin interactions enable regulated gene expression within preexisting architectural compartments. *Genome Research*, 23(12):2066–2077, 2013.

[159] M. Franke, D. M. Ibrahim, G. Andrey, W. Schwarzer, V. Heinrich, R. Schöpflin, K. Kraft, R. Kempfer, I. Jerković, W.-L. Chan, M. Spielmann, B. Timmermann, L. Wittler, I. Kurth, P. Cambiaso, O. Zuffardi, G. Houge, L. Lambie, F. Brancati, A. Pombo, M. Vingron, F. Spitz, and S. Mundlos. Formation of new chromatin domains determines pathogenicity of genomic duplications. *Cell*, 538(7624):265–269, 2016.

[160] D. G. Lupiáñez, K. Kraft, V. Heinrich, P. Krawitz, F. Brancati, E. Klopocki,

D. Horn, H. Kayserili, J. M. Opitz, R. Laxova, F. Santos-Simarro, B. Gilbert-Dussardier, L. Wittler, M. Borschiwer, S. A. Haas, M. Osterwalder, M. Franke, B. Timmermann, J. Hecht, M. Spielmann, A. Visel, and S. Mundlos. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell*, 161(5):1012–1025, 2015.

[161] S. M. Shimeld and P. C. J. Donoghue. Evolutionary crossroads in developmental biology: cyclostomes (lamprey and hagfish). *Development*, 139(12):2091–2099, 2012.

[162] S. A. Green and M. E. Bronner. The lamprey: A jawless vertebrate model system for examining origin of the neural crest and other vertebrate traits. *Differentiation*, 87(1-2):44–51, 2014.

[163] T. Sauka-Spengler, D. Meulemans, M. Jones, and M. Bronner-Fraser. Ancient evolutionary origin of the neural crest gene regulatory network. *Developmental Cell*, 13(3):405–420, 2007.

[164] L. Doglio, D. K. Goode, M. C. Pelleri, S. Pauls, F. Frabetti, S. M. Shimeld, T. Vavouri, and G. Elgar. Parallel evolution of chordate cis-regulatory code for development. *PLoS Genetics*, 9(11):e1003904, 2013.

[165] M. S. Modrell, D. Hockman, B. Uy, D. Buckley, T. Sauka-Spengler, M. E. Bronner, and C. V. Baker. A fate-map for cranial sensory ganglia in the sea lamprey. *Developmental Biology*, 385(2):405–416, 2014.

[166] J. J. Smith, S. Kuraku, C. Holt, T. Sauka-Spengler, N. Jiang, M. S. Campbell, M. D. Yandell, T. Manousaki, A. Meyer, O. E. Bloom, J. R. Morgan, J. D. Buxbaum, R. Sachidanandam, C. Sims, A. S. Garruss, M. Cook, R. Krumlauf, L. M. Wiedemann, S. A. Sower, W. A. Decatur, J. A. Hall,

C. T. Amemiya, N. R. Saha, K. M. Buckley, J. P. Rast, S. Das, M. Hirano, N. McCurley, P. Guo, N. Rohner, C. J. Tabin, P. Piccinelli, G. Elgar, M. Ruffier, B. L. Aken, S. M. J. Searle, M. Muffato, M. Pignatelli, J. Herrero, M. Jones, C. T. Brown, Y.-W. Chung-Davidson, K. G. Nanlohy, S. V. Libants, C.-Y. Yeh, D. W. McCauley, J. A. Langeland, Z. Pancer, B. Fritzsch, P. J. de Jong, B. Zhu, L. L. Fulton, B. Theising, P. Flicek, M. E. Bronner, W. C. Warren, S. W. Clifton, R. K. Wilson, and W. Li. Sequencing of the sea lamprey (petromyzon marinus) genome provides insights into vertebrate evolution. *Nat Genet*, 45(4):415–421, 2013.

[167] J. J. Smith, N. Timoshevskaya, C. Ye, C. Holt, M. C. Keinath, H. J. Parker, M. E. Cook, J. E. Hess, S. R. Narum, F. Lamanna, H. Kaessmann, V. A. Timoshevskiy, C. K. M. Waterbury, C. Saraceno, L. M. Wiedemann, S. M. C. Robb, C. Baker, E. E. Eichler, D. Hockman, T. Sauka-Spengler, M. Yandell, R. Krumlauf, G. Elgar, and C. T. Amemiya. The sea lamprey germline genome provides insights into programmed genome rearrangement and vertebrate evolution. *Nature Genetics*, 50(2):270–277, 2018.

[168] N. Horigome, M. Myojin, T. Ueki, S. Hirano, S. Aizawa, and S. Kuratani. Development of cephalic neural crest cells in embryos of lampetra japonica, with special reference to the evolution of the jaw. *Developmental Biology*, 207(2):287–308, 1999.

[169] D. Hockman, V. Chong-Morrison, S. A. Green, D. Gavriouchkina, I. Candido-Ferreira, I. T. C. Ling, R. M. Williams, C. T. Amemiya, J. J. Smith, M. E. Bronner, and T. Sauka-Spengler. A genome-wide assessment of the ancestral neural crest gene regulatory network. *Nature Communications*, 10(1), 2019.

[170] V. Papdogiannis, H. J. Parker, A. Pennati, C. Patthey, M. E. Bronner, and

S. M. Shimeld. Hmx gene conservation identifies the evolutionary origin of vertebrate cranial ganglia, 2020.

[171] S. Kuratani. Cephalic neural crest cells and the evolution of craniofacial structures in vertebrates: morphological and embryological significance of the premandibular–mandibular boundary. *Zoology*, 108(1):13–25, 2005.

[172] T. K. Mehta, V. Ravi, S. Yamasaki, A. P. Lee, M. M. Lian, B.-H. Tay, S. Tohari, S. Yanai, A. Tay, S. Brenner, and B. Venkatesh. Evidence for at least six hox clusters in the japanese lamprey (lethenteron japonicum). *Proceedings of the National Academy of Sciences*, 110(40):16044–16049, 2013.

[173] O. Simakov, F. Marlétaz, J.-X. Yue, B. O'Connell, J. Jenkins, A. Brandt, R. Calef, C.-H. Tung, T.-K. Huang, J. Schmutz, N. Satoh, J.-K. Yu, N. H. Putnam, R. E. Green, and D. S. Rokhsar. Deeply conserved synteny resolves early events in vertebrate evolution. *Nature Ecology &amp Evolution*, 4(6):820–830, 2020.

[174] D. Meulemans, D. McCauley, and M. Bronner-Fraser. Id expression in amphioxus and lamprey highlights the role of gene cooption during neural crest evolution. *Developmental Biology*, 264(2):430–442, 2003.

[175] D. Meulemans and M. Bronner-Fraser. Gene-regulatory interactions in neural crest evolution and development. *Developmental Cell*, 7(3):291–299, 2004.

[176] D. Meulemans and M. Bronner-Fraser. Amphioxus and lamprey AP-2 genes: implications for neural crest evolution and migration patterns. *Development*, 129(21):4953–4962, 2002.

[177] D. M. Medeiros. The evolution of the neural crest: new perspectives from lamprey and invertebrate neural crest-like cells. *Wiley Interdisciplinary Reviews: Developmental Biology*, 2(1):1–15, 2012.

[178] B. R. Uy, M. Simoes-Costa, T. Sauka-Spengler, and M. E. Bronner. Expression of sox family genes in early lamprey development. *The International Journal of Developmental Biology*, 56(5):377–383, 2012.

[179] J. M. Tomsa and J. A. Langeland. OtxExpression during lamprey embryogenesis provides insights into the evolution of the vertebrate head and jaw. *Developmental Biology*, 207(1):26–37, 1999.

[180] H. J. Parker, B. D. Kumar, S. A. Green, K. D. Prummel, C. Hess, C. K. Kaufman, C. Mosimann, L. M. Wiedemann, M. E. Bronner, and R. Krumlauf. A hox-TALE regulatory circuit for neural crest patterning is conserved across vertebrates. *Nat Commun*, 10(1), 2019.

[181] S. A. Green, B. R. Uy, and M. E. Bronner. Ancient evolutionary origin of vertebrate enteric neurons from trunk-derived neural crest. *Nature*, 544(7648):88–91, 2017.

[182] D. W. McCauley and M. Bronner-Fraser. Neural crest contributions to the lamprey head. *Development*, 130(11):2317–2327, 2003.

[183] S. A. Green and M. E. Bronner. Gene duplications and the early evolution of neural crest development. *Seminars in Cell &amp Developmental Biology*, 24(2):95–100, 2013.

[184] E. H. Barriga, P. A. Trainor, M. Bronner, and R. Mayor. Animal models for studying neural crest development: is the mouse different? *Development*, 142(9):1555–1560, 2015.

[185] M.-T. Weil, S. Heibeck, M. Töpperwien, S. tom Dieck, T. Ruhwedel, T. Salditt, M. C. Rodicio, J. R. Morgan, K.-A. Nave, W. Möbius, and H. B. Werner. Axonal ensheathment in the nervous system of lamprey: Implica-

tions for the evolution of myelinating glia. *The Journal of Neuroscience*, 38(29):6586–6596, 2018.

[186] J. R. York, T. Yuan, K. Zehnder, and D. W. McCauley. Lamprey neural crest migration is snail-dependent and occurs without a differential shift in cadherin expression. *Developmental Biology*, 428(1):176–187, 2017.

[187] K. G. Ota and S. Kuratani. The history of scientific endeavors towards understanding hagfish embryology. *Zoological Science*, 23(5):403–418, 2006.

[188] K. G. Ota, S. Kuraku, and S. Kuratani. Hagfish embryology with reference to the evolution of the neural crest. *Nature*, 446(7136):672–675, 2007.

[189] Y. Oisi, K. G. Ota, S. Kuraku, S. Fujimoto, and S. Kuratani. Craniofacial development of hagfishes and the evolution of vertebrates. *Nature*, 493(7431):175–180, 2012.

[190] W. R. Jeffery, A. G. Strickler, and Y. Yamamoto. Migratory neural crest-like cells form body pigmentation in a urochordate embryo. *Nature*, 431(7009):696–699, 2004.

[191] P. B. Abitua, E. Wagner, I. A. Navarrete, and M. Levine. Identification of a rudimentary neural crest in a non-vertebrate chordate. *Nature*, 492(7427):104–107, 2012.

[192] W. R. Jeffery, T. Chiba, F. R. Krajka, C. Deyts, N. Satoh, and J.-S. Joly. Trunk lateral cells are neural crest-like cells in the ascidian ciona intestinalis: Insights into the ancestry and evolution of the neural crest. *Developmental Biology*, 324(1):152–160, 2008.

[193] A. Stolfi, K. Ryan, I. A. Meinertzhagen, and L. Christiaen. Migratory neuronal progenitors arise from the neural plate borders in tunicates. *Nature*, 527(7578):371–374, 2015.

[194] D. Meulemans and M. Bronner-Fraser. Central role of gene cooption in neural crest evolution. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 304b(4):298–303, 2005.

[195] E. V. Otterloo, W. Li, A. Garnett, M. Cattell, D. M. Medeiros, and R. A. Cornell. Novel tfap2-mediated control ofisoxE/iexpression facilitated the evolutionary emergence of the neural crest. *Development*, 139(4):720–730, 2012.

[196] A. J. Tarashansky, J. M. Musser, M. Khariton, P. Li, D. Arendt, S. R. Quake, and B. Wang. Mapping single-cell atlases throughout metazoa unravels cell type evolution. *eLife*, 10, 2021.

[197] N. Miyamoto and H. Wada. Hemichordate neurulation and the origin of the neural tube. *Nature Communications*, 4(1), 2013.

[198] B. K. Hall and J. A. Gillis. Incremental evolution of the neural crest, neural crest cells and neural crest-derived skeletal tissues. *Journal of Anatomy*, 222(1):19–31, 2012.

[199] I. Yajima, K. Endo, S. Sato, R. Toyoda, H. Wada, S. Shibahara, T. Numakunai, K. Ikeo, T. Gojobori, C. R. Goding, and H. Yamamoto. Cloning and functional analysis of ascidian mitf in vivo: insights into the origin of vertebrate pigment cells. *Mechanisms of Development*, 120(12):1489–1504, 2003.

[200] S. Hannenhalli and K. H. Kaestner. The evolution of fox genes and their role in development and disease. *Nat Rev Genet*, 10(4):233–240, 2009.

[201] L. A. Cirillo, F. R. Lin, I. Cuesta, D. Friedman, M. Jarnik, and K. S. Zaret. Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Molecular Cell*, 9(2):279–289, 2002.

[202] H. Ono, Z. Kozmik, J.-K. Yu, and H. Wada. A novel n-terminal motif is responsible for the evolution of neural crest-specific gene-regulatory activity in vertebrate FoxD3. *Developmental Biology*, 385(2):396–404, 2014.

[203] Y. Kamachi and H. Kondoh. Sox proteins: regulators of cell fate specification and differentiation. *Development*, 140(20):4129–4144, 2013.

[204] A. Tai, M. Cheung, Y.-H. Huang, R. Jauch, M. E. Bronner, and K. S. E. Cheah. SOXE neofunctionalization and elaboration of the neural crest during chordate evolution. *Scientific Reports*, 6(1), 2016.

[205] F. Cossais, E. Sock, J. Hornig, S. Schreiner, S. Kellerer, M. R. Bösl, S. Russell, and M. Wegner. 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(1):267–281, 2010.

[206] S. L. Prescott, R. Srinivasan, M. C. Marchetto, I. Grishina, I. Narvaiza, L. Selleri, F. H. Gage, T. Swigut, and J. Wysocka. Enhancer divergence and cis-regulatory evolution in the human and chimp neural crest. *Cell*, 163(1):68–83, 2015.

[207] J. Stundl, A. Pospisilova, T. Matějková, M. Psenicka, M. E. Bronner, and R. Cerny. Migratory patterns and evolutionary plasticity of cranial neural crest cells in ray-finned fishes. *Developmental Biology*, 467(1-2):14–29, 2020.

[208] Y. Wakamatsu and K. Suzuki. Sequence alteration in the enhancer contributes to the heterochronic sox9 expression in marsupial cranial neural crest. *Developmental Biology*, 456(1):31–39, 2019.

[209] D. Hnisz, D. S. Day, and R. A. Young. Insulated neighborhoods: Structural

and functional units of mammalian gene control. *Cell*, 167(5):1188–1200, 2016.

[210] S. Kittelmann, A. D. Buffry, F. A. Franke, I. Almudi, M. Yoth, G. Sabaris, J. P. Couso, M. D. S. Nunes, N. Frankel, J. L. Gómez-Skarmeta, J. Pueyo-Marques, S. Arif, and A. P. McGregor. Gene regulatory network architecture in different developmental contexts influences the genetic basis of morphological evolution, 2017.

[211] F. M. Real, S. A. Haas, P. Franchini, P. Xiong, O. Simakov, H. Kuhl, R. Schöpflin, D. Heller, M.-H. Moeinzadeh, V. Heinrich, T. Krannich, A. Bressin, M. F. Hartmann, S. A. Wudy, D. K. N. Dechmann, A. Hurtado, F. J. Barrionuevo, M. Schindler, I. Harabula, M. Osterwalder, M. Hiller, L. Wittler, A. Visel, B. Timmermann, A. Meyer, M. Vingron, R. Jiménez, S. Mundlos, and D. G. Lupiáñez. The mole genome reveals regulatory rearrangements associated with adaptive intersexuality. *Science*, 370(6513):208–214, 2020.

[212] D. Villar, C. Berthelot, S. Aldridge, T. F. Rayner, M. Lukk, M. Pignatelli, T. J. Park, R. Deaville, J. T. Erichsen, A. J. Jasinska, J. M. Turner, M. F. Bertelsen, E. P. Murchison, P. Flicek, and D. T. Odom. Enhancer evolution across 20 mammalian species. *Cell*, 160(3):554–566, 2015.

[213] C. Berthelot, D. Villar, J. E. Horvath, D. T. Odom, and P. Flicek. Complexity and conservation of regulatory landscapes underlie evolutionary resilience of mammalian gene expression. *Nature Ecology &amp Evolution*, 2(1):152–163, 2017.

[214] D. H. Erwin. The origin of animal body plans: a view from fossil evidence and the regulatory genome. *Development*, 147(4), 2020.

[215] B. Prud'homme, N. Gompel, A. Rokas, V. A. Kassner, T. M. Williams, S.-D. Yeh, J. R. True, and S. B. Carroll. Repeated morphological evolution through cis-regulatory changes in a pleiotropic gene. *Nature*, 440(7087):1050–1053, 2006.

[216] D. L. Stern and V. Orgogozo. Is genetic evolution predictable? *Science*, 323(5915):746–751, 2009.

[217] D. L. Stern and V. Orgogozo. The Loci Of Evolution: How Predictable Is Genetic Evolution? *Evolution*, 62(9):2155–2177, 2008.

[218] K. T. Xie, G. Wang, A. C. Thompson, J. I. Wucherpfennig, T. E. Reimchen, A. D. C. MacColl, D. Schluter, M. A. Bell, K. M. Vasquez, and D. M. Kingsley. DNA fragility in the parallel evolution of pelvic reduction in stickleback fish. *Science*, 363(6422):81–84, 2019.

[219] J. O. J. Davies, J. M. Telenius, S. J. McGowan, N. A. Roberts, S. Taylor, D. R. Higgs, and J. R. Hughes. Multiplexed analysis of chromosome conformation at vastly improved sensitivity. *Nat Methods*, 13(1):74–80, 2015.

[220] M. S. Simoes-Costa, S. J. McKeown, J. Tan-Cabugao, T. Sauka-Spengler, and M. E. Bronner. Dynamic and differential regulation of stem cell factor foxd3 in the neural crest is encrypted in the genome. *PLoS Genet*, 8(12):e1003142, 2012.

[221] V. Hamburger and H. L. Hamilton. A series of normal stages in the development of the chick embryo. *Dev Dyn*, 195(4):231–272, 1992.

[222] P. Langfelder and S. Horvath. WGCNA: an r package for weighted correlation network analysis. *BMC Bioinformatics*, 9(1), 2008.

[223] B. Z. Peter Langfelder and S. Horvath. Defining clusters from a hierarchical cluster tree: the dynamic tree cut package for r. *Bioinformatics*, 24(5):719–720, 2007.

[224] J. D. Buenrostro, P. G. Giresi, L. C. Zaba, H. Y. Chang, and W. J. Greenleaf. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, dna-binding proteins and nucleosome position. *Nat Methods*, 10(12):1213–8, 2013.

[225] X. Zhan and D. J. Liu. SEQMINER: An r-package to facilitate the functional interpretation of sequence-based associations. *Genet Epidemiol* ., 39(8):619–623, 2015.

[226] M. L. Basch, M. Bronner-Fraser, and M. I. Garcia-Castro. Specification of the neural crest occurs during gastrulation and requires pax7. *Nature*, 441(7090):218–22, 2006.

[227] C. Murko, F. M. Vieceli, and M. Bronner. Transcriptome dataset of trunk neural crest cells migrating along the ventral pathway of chick embryos. *Data Brief*, 21:2547–2553, 2018.

[228] V. Robinson, A. Smith, A. M. Flenniken, and D. G. Wilkinson. Roles of eph receptors and ephrins in neural crest pathfinding. *Cell Tissue Res.*, 290(2):265–274, 1997.

[229] V. Boeva, C. Louis-Brennetot, A. Peltier, S. Durand, C. Pierre-Eugène, V. Raynal, H. C. Etchevers, S. Thomas, A. Lermine, E. Daudigeos-Dubus, B. Geoerger, M. F. Orth, T. G. P. Grünewald, E. Diaz, B. Ducos, D. Surdez, A. M. Carcaboso, I. Medvedeva, T. Deller, V. Combaret, E. Lapouble, G. Pierron, S. Grossetête-Lalami, S. Baulande, G. Schleiermacher, E. Barillot, H. Rohrer, O. Delattre, and I. Janoueix-Lerosey. Heterogeneity of neurob-

lastoma cell identity defined by transcriptional circuitries. *Nature Genetics*, 49(9):1408–1413, 2017.

[230] P. Betancur, M. Bronner-Fraser, and T. Sauka-Spengler. Genomic code for isox10/i activation reveals a key regulatory enhancer for cranial neural crest. *Proceedings of the National Academy of Sciences*, 107(8):3570–3575, 2010.

[231] P. Betancur, T. Sauka-Spengler, and M. Bronner. A sox10 enhancer element common to the otic placode and neural crest is activated by tissue-specific paralogs. *Development*, 138(17):3689–3698, 2011.

[232] R. M. Williams, U. Senanayake, M. Artibani, G. Taylor, D. Wells, A. A. Ahmed, and T. Sauka-Spengler. Genome and epigenome engineering CRISPR toolkit for iin vivo/i modulation of icis/i-regulatory interactions and gene expression in the chicken embryo. *Development*, 2018.

[233] S. Heinz, C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre, H. Singh, and C. K. Glass. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and b cell identities. *Mol Cell.*, 38(4):576–589, 2010.

[234] R. Finkelstein and N. Perrimon. The molecular genetics of head development in *Drosophila melanogaster*. 112(4):899–912, 1991.

[235] N. M. LeDouarin and C. Kalcheim. The neural crest. *(Second Edition), Cambridge University Press*, Vol.36, 1999.

[236] K. M. Lelli, M. Slattery, and R. S. Mann. Disentangling the many layers of eukaryotic transcriptional regulation. *Annual Review of Genetics*, 46(1):43–68, 2012.

[237] S. Aibar, C. B. González-Blas, T. Moerman, V. A. Huynh-Thu, H. Imrichova, G. Hulselmans, F. Rambow, J.-C. Marine, P. Geurts, J. Aerts, J. van den

Oord, Z. K. Atak, J. Wouters, and S. Aerts. SCENIC: single-cell regulatory network inference and clustering. *Nat Methods.*, 14(11):1083–1086, 2017.

[238] M. Gouti, J. Delile, D. Stamataki, F. J. Wymeersch, Y. Huang, J. Kleinjung, V. Wilson, and J. Briscoe. A gene regulatory network balances neural and mesoderm specification during vertebrate trunk development. *Dev Cell*, 41(3):243–261.e7, 2017.

[239] Y. Chudnovsky, D. Kim, S. Zheng, W. A. Whyte, M. Bansal, M.-A. Bray, S. Gopal, M. A. Theisen, S. Bilodeau, P. Thiru, J. Muffat, O. H. Yilmaz, M. Mitalipova, K. Woolard, J. Lee, R. Nishimura, N. Sakata, H. A. Fine, A. E. Carpenter, S. J. Silver, R. G. Verhaak, A. Califano, R. A. Young, K. L. Ligon, I. K. Mellinghoff, D. E. Root, D. M. Sabatini, W. C. Hahn, and M. G. Chheda. ZFHX4 interacts with the NuRD core member CHD4 and regulates the glioblastoma tumor-initiating cell state. *Cell Rep.*, 6(2):313–324, 2014.

[240] T. Ferronha, M. A. Rabadan, E. Gil-Guinon, G. L. Dreau, C. de Torres, and E. Marti. LMO4 is an essential cofactor in the snail2-mediated epithelial-to-mesenchymal transition of neuroblastoma and neural crest cells. *Journal of Neuroscience*, 33(7):2773–2783, 2013.

[241] S. D. Ochoa, S. Salvador, and C. LaBonne. The LIM adaptor protein LMO4 is an essential regulator of neural crest development. *Dev Biol.*, 361(2):313–325, 2012.

[242] J. D. Calisto, C. Araya, L. Marchant, C. F. Riaz, and R. Mayor. Essential role of non-canonical wnt signalling in neural crest migration. *Development*, 132(11):2587–2597, 2005.

[243] M. P. Verzi, P. Agarwal, C. Brown, D. J. McCulley, J. J. Schwarz, and

B. L. Black. The transcription factor MEF2c is required for craniofacial development. *Dev Cell.*, 12(4):645–652, 2007.

[244] T. Hochgreb-Hagele and M. E. Bronner. A novel foxd3 gene trap line reveals neural crest precursor movement and a role for foxd3 in their specification. *Dev Biol*, 374(1):1–11, 2013.

[245] N. Tahara, H. Kawakami, K. Q. Chen, A. Anderson, M. Yamashita Peterson, W. Gong, P. Shah, S. Hayashi, R. Nishinakamura, Y. Nakagawa, D. J. Garry, and Y. Kawakami. Sall4 regulates neuromesodermal progenitors and their descendants during body elongation in mouse embryos. *Development*, 146(14), 2019.

[246] M. Lukoseviciute, S. Mayes, and T. Sauka-Spengler. Neuromesodermal progenitor origin of trunk neural crest in vivo. *bioRxiv*, 2021.

[247] N. Irie and S. Kuratani. The developmental hourglass model: a predictor of the basic body plan? *Development*, 141(24):4649–4655, 2014.

[248] R. Amamoto, M. D. Garcia, E. R. West, J. Choi, S. W. Lapan, E. A. Lane, N. Perrimon, and C. L. Cepko. Probe-seq enables transcriptional profiling of specific cell types from heterogeneous tissue by RNA-based isolation. *eLife*, 8, 2019.

[249] M. Westerfield. *The zebrafish book : a guide for the laboratory use of zebrafish (Danio rerio). 4th ed.* Eugene, Or., 2000.

[250] C. B. Kimmel, W. W. Ballard, S. R. Kimmel, B. Ullmann, and T. F. Schilling. Stages of embryonic development of the zebrafish. *Dev Dyn*, 203(3):253–310, 1995.