



UNIVERSITY OF OXFORD

Tracking Chromatin Dynamics and Bimodal foxd3 Mechanism on Neural Crest gene regulation *in vivo*

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

at
St Catherine's College

Author: Martyna Lukoseviciute
Supervisor: Professor Tatjana Sauka-Spengler

Trinity 2019

Martyna Lukoseviciute, University of Oxford, St Catherine's College

D.Phil. in Medical Sciences, Trinity 2019

Tracking Chromatin Dynamics and Bimodal *foxd3* Mechanism on Neural Crest gene regulation *in vivo*

The neural crest (NC) is a transient embryonic stem cell-like population characterised by its multipotency and broad developmental potential. The core NC transcription factor *foxd3*, known to mediate late NC lineage decisions via transcriptional repression, is also pervasively expressed in the pre-migratory NC progenitors, where its putative role remains elusive.

To elucidate the mechanistic and temporal mode of *foxd3* action, we analysed its genome-wide binding profiles across multiple stages of zebrafish development. To this end, we have developed and used our novel biotin ChIP-seq approach allowing to specifically biotinylate factors of interest, such as *foxd3* protein, *in vivo*. *Foxd3* binding maps interpreted within the context of NC-specific transcriptional and epigenomic profiles of *foxd3*-mutant and control cells reveal that *foxd3* acts bimodally throughout the NC ontogeny. We showed that first *foxd3* functions globally as a pioneer factor to prime the onset of NC specification and migration genes by re-arranging their chromatin landscape. Strikingly, *foxd3* then gradually switches from an activator to its well-described role as a transcriptional repressor and potentially uses different partners for each role.

By employing two enhancer reporter zebrafish lines, imaging and *-omics* tools, we have also shown that auto-regulation of *foxd3* gene locus is involved during early cell-fate decisions and embryo patterning. In the cranial embryo regions, positive auto-regulatory *foxd3* feedback is driving *foxd3* expression that is needed for the NC and nervous system development. This auto-regulatory loop is ensured by a set of *foxd3 cis*-regulatory elements that exhibit NC-like regulatory signatures. Contrarily, *foxd3* is auto-regulated by a different set of enhancers and, as a result, is expressed to lower levels during the apparent neuromesodermal progenitor specification in the caudal embryo regions.

Taken together, this thesis demonstrates that *foxd3* needs to be tightly auto-regulated during early cell lineage decisions. Also, later in the NC development, *foxd3* switches from a chromatin activator to a repressor in order to maintain NC multipotency and define its cell fates.

Abstract	iii
1 General introduction	13
1.1 Chromatin and its dynamics in animal development	13
1.1.1 Global chromatin organisation	15
1.1.2 Enhancers	17
1.1.3 Topologically associating domains and gene expression	18
1.1.4 Histone variants	20
1.1.5 Pioneering factors	22
1.2 The neural crest	24
1.2.1 NC ontogeny from a GRN perspective	26
1.2.2 Distinctive axial NC competence and associated NC derivatives	31
2 Bimodal Action of foxd3 is Required to Specify Neural Crest Cells and Maintain their Multipotency <i>in vivo</i>	35
2.1 Foxd3 - a critical transcription factor for an early embryo development	35
2.2 Mutant lines to study gene regulatory roles of foxd3 throughout the NC development	38
2.3 Knockdown of foxd3 Results in Decreased Expression of NC Specifier Genes and Increased Expression of Genes Controlling NC Migration and Differentiation . . .	42
2.4 Application of the Biotin ChIP-seq method to identify direct foxd3 gene targets throughout the NC development	45
2.4.1 Generation of the foxd3-Avi transgenic fish line	46
2.4.2 Biotin-ChIP confirms direct bimodal action of foxd3 on the NC gene regulation	48
2.5 foxd3 affects chromatin accessibility of distal <i>cis</i> -regulatory elements	53
2.6 Ectopic Expression of foxd3 Modifies the Chromatin Landscape in Early Embryos	58
2.7 Distinct gene regulatory mechanisms underlie the bimodal action of foxd3	60

2.8	Discussion	67
2.8.1	Bimodal action of <i>foxd3</i> on NC gene regulation	67
2.8.2	Co-pioneering factors	69
2.8.3	Co-repressing partners	71
2.8.4	Summary	72
3	Auto-regulation of the <i>foxd3</i> genetic locus during cell lineage specification	73
3.1	Different types of transcription factor auto-regulatory loops for a tight gene expression control	73
3.2	The genetic locus of <i>foxd3</i> reveals its putatively auto-regulated <i>cis</i> -elements . . .	76
3.3	Comparison of transcription factor binding sites between Enh6 and Enh4 reveals underlying differential regulation	79
3.4	Stable transgenic Enh6-GFP/Enh4-GFP zebrafish line imaging confirms different axial activity patterns	81
3.5	Defining transcriptional signatures of Enh6+ and Enh4+ cells	87
3.6	<i>foxd3</i> + single cell RNA-seq reveals trunk NMP-related cell populations	95
3.7	Further defining Enh6+ and Enh4+ cells based on their accessible DNA regions .	100
3.8	Surveying Enh6+ and Enh4+ cell derivatives in the zebrafish larvae	104
3.9	Predicted Enh4 activating and repressing transcription factors	109
3.10	Discussion	111
3.10.1	Foxd3 is involved in positive and negative auto-regulatory loops during cell fate decisions	111
3.10.2	Enh4 - a novel marker of the zebrafish neuromesodermal progenitors? . .	113
3.10.3	Stable reporter lines to understand anterior and posterior nervous system development	117
3.10.4	Summary	118
4	Conclusions and Future Directions	119
5	Materials and Methods	125
5.1	General methods	125
5.1.1	Zebrafish husbandry	125
5.1.2	Zebrafish embryo microinjections	125
5.1.3	Cell dissociation from zebrafish embryos and FAC-sorting	126

5.2	Chapter 2 Methods	126
5.2.1	<i>foxd3</i> -mutant vs control RNA-seq	126
5.2.2	<i>foxd3</i> Biotin ChIP-seq	127
5.2.3	<i>foxd3</i> -mutant vs control ATAC-seq	130
5.2.4	Foxd3 ectopic expression assay followed by ATAC-seq	131
5.2.5	<i>Ac/Ds</i> -mediated enhancer screening assay	132
5.3	Chapter 3 Methods	134
5.3.1	Tg(FoxD3:enh6-GFP) ^{ox109} and Tg(FoxD3:enh4-GFP) ^{ox110} line generation	134
5.3.2	JASPAR and HOCOMOCO TFBS analysis	134
5.3.3	Transgenic zebrafish line embryo imaging and analysis	134
5.3.4	Enh6+/- and Enh4+/- cell RNA-seq	135
5.3.5	Enh6+/- and Enh4+/- cell ATAC-seq	136
5.4	10x Chromium single cell RNA-seq	138
5.4.1	RNA extraction, library preparation and sequencing	138
5.4.2	Single cell RNA-seq bioinformatics processing	138
5.5	Quantification and statistical analysis	139
	Appendix	141
1	Other Enh6+/- and Enh4+/- ATAC-seq <i>k</i> -mean clusters (Section 3.7)	141
	Bibliography	143
	Acknowledgments	177

List of Figures

1.1	Overview of Neural Crest derivatives	26
1.2	Regulation of neural crest formation	27
1.3	The gene regulatory network underlying NC ontogeny	28
1.4	Overview of the NC anterior-posterior axis derivatives	32
2.1	Zebrafish embryo stages examined in this study.	38
2.2	Logic behind the <i>foxd3</i> gene trap lines generated by Tatiana Hochgreb-Hägele and Marianne E. Bronner.	40
2.3	Experimental strategy for obtaining <i>foxd3</i> -mutant (yellow) and <i>foxd3</i> -control (green) NC cells for comparative transcriptomics and epigenomics analyses.	41
2.4	Transcriptome changes in the <i>foxd3</i> -mutant cells during different stages of NC development.	44
2.5	Potential <i>foxd3</i> -mutant rescue mechanism by paralogous NC factors. . . .	45
2.6	Generation of the Foxd3-Avi transgenic fish line.	47
2.7	<i>foxd3</i> biotin ChIP-seq method	49
2.8	<i>foxd3</i> biotin ChIP-seq tracks and <i>foxd3</i> binding dynamics across different stages of neural crest ontogeny.	50
2.9	Biotin ChIP-seq analysis supports a direct bi-modal <i>foxd3</i> regulatory action on NC gene expression.	52
2.10	Foxd3-mediated repression starts at 5-6 somite stage.	53
2.11	Absence of <i>foxd3</i> causes reduced accessibility of distal gene elements. . . .	54
2.12	Gene expression levels correlated to the uncovered open chromatin elements.	55
2.13	Reduced chromatin accessibility leads to a reduced gene expression in the absence of <i>foxd3</i>	55

2.14	foxd3-controlled NC gene enhancer cluster 3.	57
2.15	<i>Cis</i> -regulatory elements from <i>k</i> -cluster 3 show NC-specific reporter activity <i>in vivo</i>	58
2.16	Experimental strategy for foxd3 overexpression <i>in vivo</i>	59
2.17	Ectopically delivered foxd3 effect on the NC chromatin remodelling. . . .	61
2.18	Strategy for <i>in silico</i> analysis to reveal putative foxd3 complexes during its mediated chromatin opening and compaction.	62
2.19	Putative foxd3 co-factors acting as activators during early NC development.	63
2.20	Putative foxd3 co-factors acting as repressors or competitors during late NC development.	64
2.21	Identified putative co-factors needed for foxd3-mediated chromatin com- paction.	66
2.22	Currently proposed mechanism of the bimodal action of foxd3 during neu- ral crest differentiation.	72
3.1	Foxd3 gene locus and active auto-regulated <i>cis</i> -regulatory elements. . . .	77
3.2	Ectopically delivered foxd3 reduces Enh4 and increases Enh6 chromatin accessibility.	78
3.3	Transcription factor binding site (TFBS) analysis underlying Enh6 and Enh4 DNA sequences using JASPAR database.	80
3.4	Transcription factor binding site (TFBS) analysis underlying Enh6 and Enh4 DNA sequences using HOCOMOCO database.	80
3.5	<i>Tg(FoxD3:enh6-GFP)^{ox109}</i> line crossed to <i>Gt(foxd3-mcherry)^{ct110aR}</i> line: lateral view of the whole-mount live embryo at 5 somite stage.	82
3.6	<i>Tg(FoxD3:enh6-GFP)^{ox109}</i> line crossed to <i>Gt(foxd3-mcherry)^{ct110aR}</i> line: live embryo dorsal view sections of neural tube at 5 somite stage.	83
3.7	<i>Tg(FoxD3:enh4-GFP)^{ox110}</i> line crossed to <i>Gt(foxd3-mcherry)^{ct110aR}</i> line: lateral view of the whole-mount live embryo at 6 somite stage.	84
3.8	<i>Tg(FoxD3:enh4-GFP)^{ox110}</i> line crossed to <i>Gt(foxd3-mcherry)^{ct110aR}</i> line: dorsal view of the whole-mount live embryo at 7-8 somite stage.	85

3.9	<i>Tg(FoxD3:enh4-GFP)^{ox110}</i> line crossed to <i>Gt(foxd3-mcherry)^{ct110aR}</i> line: live embryo dorsal view sections of neural tube at 5 somite stage.	86
3.10	Principal Component Analysis (PCA) of RNA-seq samples and replicates.	87
3.11	Profiling Enh6+ cells.	88
3.12	Profiling Enh4+ cells.	90
3.13	Comparing Enh6+ cells with Enh4+ cells.	92
3.14	Schematics of hox gene expression in a zebrafish larva.	93
3.15	Differences in Bmp, Fgf and Wnt signalling molecule expression between Enh6+ and Enh4+ cells.	94
3.16	<i>Gt(foxd3-citrine)ct110a</i> single cell RNA-seq data reveals distinctive ante- rior and posterior cell types	97
3.17	<i>Gt(foxd3-citrine)ct110a</i> single cell RNA-seq data identifies different neu- romesodermal progenitor-related cell populations in zebrafish embryo . . .	98
3.18	<i>Gt(foxd3-citrine)ct110a</i> single cell RNA-seq data identifies NC-like neu- ronal progenitors	99
3.19	Surveying and comparing open chromatin regions within the <i>foxd3</i> locus between Eh6+, Enh6-, Enh4+ and Enh4- cells.	101
3.20	<i>k</i> -mean clustering reveals two Enh6+ sample-specific clusters.	102
3.21	<i>k</i> -mean clustering reveals two Enh4+ sample-specific clusters.	103
3.22	<i>Tg(FoxD3:enh6-GFP)^{ox109}</i> line crossed to <i>Gt(foxd3-mcherry)^{ct110aR}</i> line: whole-mount live larvae imaging at 2 days post fertilisation (dpf).	106
3.23	<i>Tg(FoxD3:enh4-GFP)^{ox110}</i> line crossed to <i>Gt(foxd3-mcherry)^{ct110aR}</i> line: whole-mount live larvae imaging at 2 days post fertilisation (dpf).	107
3.24	Enh6-GFP or Enh4-GFP co-localisation with <i>foxd3</i> -mCherry fluorescence at 5-6 somite and 2 days post fertilisation stages.	108
3.25	Enh4 activators and repressors.	110
5.1	Overview Biotin ChIP-seq method	129
5.2	Information on scRNA-seq features	139

1	Additional 1-6 k -mean clusters	141
2	Additional 7-13 k -mean clusters	142

1.1 Chromatin and its dynamics in animal development

Given that all cells of an organism share an identical deoxyribonucleic acid (DNA) sequence, one could wonder how cells acquire differential gene expression profiles, which fundamentally drive cellular differentiation ultimately resulting in an entire organism comprised of various tissues. A central paradigm of early embryo development is that an integration of an embryo genetic makeup and extracellular signals modulate gene activity at a chromatin level [1]. This regulation must be dynamic in order to ensure that only required genes are switched on or off to balance between cellular pluripotency and cell fate decisions [2]. Chromatin serves a number of different functions. First of all, it allows long strands of DNA to be compactly packaged within cells. Such DNA packaging is important not only structurally but also functionally. For instance, it facilitates DNA arrangement into compartments known as topologically associated domains (TADs) and metaTADs that are thought to be important for local and long-range DNA interactions, respectively [3, 4]. In general, chromatin creates an inhibitory barrier to a range of DNA binding factors involved in cellular processes such as DNA repair, replication and transcription [5, 6]. As a result, only those DNA elements that are found in an open chromatin conformation are functionally active. A level of chromatin accessibility is derived by very dynamic and highly regulated processes that involve a number of different effector proteins, which can either modulate ('writers' and 'erasers') or recognise ('readers') epigenetic marks found on chromatin.

The most significant dynamic chromatin changes are found on diverse *cis*-regulatory elements, such as enhancers, promoters, silencers and insulators. These *cis*-regulatory elements can acquire either activating or repressing histone and DNA modifications in

a regulated manner. Chromatin modifications associated with enhancer regions are of particular importance in animal development and disease [7]. One of the reasons is that enhancers can be established by the deposited particular histone marks in a highly cell-specific manner. These marks include post-translational modifications of specific lysine residues on histone 3: histone 3 lysine 4 mono-methylation (H3K4me1) and histone 3 lysine 27 acetylation (H3K27ac) for active enhancers and H3K4me1 and histone 3 lysine 27 tri-methylation (H3K27me3) for poised inactive enhancers (with a potential to get activated). Given that active enhancers initiate or markedly increase the transcription of their regulated genes and that multiple enhancers can regulate a single gene locus, as well as that a single enhancer can contribute to multiple gene expression, the significance of slight chromatin changes driving cellular differentiation is evident.

Nonetheless, histone modifications on their own have little impact on gene expression and they rather work in concert with proteins that interact with such histone marks ('readers') [6]. In a nutshell, chromatin 'reader' proteins work together to recognise and integrate cellular information in a combinatorial manner to drive and instigate change in chromatin organisation by recruiting chromatin remodellers. In turn, this influences transcriptional machinery access to coding DNA regions, which leads to a functional change of gene expression. Overall, due to changes in epigenomic signatures, cells acquire slightly different gene expression states, which is an underlying mechanism of cellular differentiation. However, relatively little is still known about the regulation of chromatin dynamics involved in precisely-controlled animal developmental processes.

The increased understanding of epigenome importance in cellular differentiation, organism development and complex disease predisposition has led to a launch of multiple projects on epigenetics, such as ENCODE and Roadmap Epigenomics [8, 9]. The ENCODE project consortium pioneered the big data study of the human epigenome and its signatures in cultured cells, while The Roadmap Epigenomics Project moved the analysis one step further by employing primary healthy and diseased human tissues and cells [10]. This combined effort to tackle numerous regulatory processes and pathways orchestrated by differential chromatin states during the course of development has already started

to shed light upon molecular mechanisms underlying both normal gene expression and aetiology of various diseases [11].

1.1.1 Global chromatin organisation

DNA is an essential macromolecule that stores a vast amount of genetic data crucial for organism development and continuous cellular maintenance. However, a common problem underlying big data is that their storage requires a large space. DNA is not an exception. A single cell contains around 2-3 metres of DNA, which needs to be efficiently packaged within a 5-10 μm diameter nucleus [12]. This is achieved by organising DNA into a highly ordered chromatin structure. A basic unit of chromatin is a nucleosome, which is comprised of positively charged proteins that have a high affinity for negatively charged DNA molecules. A single nucleosome is an octamer, which consists of two of each H2A, H2B, H3 and H4 histone proteins, and is wrapped-up twice by 147bp of DNA [13]. Another histone protein, H1, functions as a linker protein between neighbouring nucleosomes, which is important for nucleosome spacing, it also prevents ATP-dependent chromatin remodelling and overall facilitates chromatin condensation [14, 15]. On the whole, nucleosome repeats create a flexible 10 nm chromatin fibre that is commonly portrayed as beads on a string. On the other hand, the structure and even existence of higher-order fibres, such as 30 nm fibres, is somewhat controversial [12, 16]. Recently, it was proposed that 30 nm fibers are *in vitro* artefacts as a result of low-salt buffer conditions and cannot be identified *in vivo* neither in interphase nor in mitotic chromosomes [12]. In such case, chromatin structure would be more loose and thus more dynamic than previously anticipated.

On a large scale, chromatin is divided into two major types - heterochromatin and euchromatin. Heterochromatin is gene poor, very condensed and hence nearly inaccessible to transcription machinery. On the other hand, euchromatin is gene-rich, less condensed and its dynamic regulation determines levels of gene transcription [17]. A highly condensed state of heterochromatin was first thought to be primarily governed by histone and DNA epigenetic modifications, which are general histone hypo-acetylation, H3K9

and H3K27 methylation, a strong heterochromatin-protein 1 (HP-1) association with H3K9me and H3K27me, and DNA methylation on cytosine residues [18]. Remarkably, many studies have revealed that heterochromatin formation is driven by phase-phase separation mechanisms. For instance, it was shown that human heterochromatin protein HP-1 α , brought by the H3K9 trimethylation mark (H3K9me₃), oligomerises via its chromoshadow domain (CSD) upon phosphorylation of the N-terminal extension of the HP-1 α . This leads to liquid droplet formation, which overall establishes a less-permissive phase-separated heterochromatin state [19]. Similarly, components of the Polycomb repressive complex 1 (PRC1), associated with condensed chromatin formation, were also shown to form nuclear condensates due to a phase-separation, leading a silent chromatin state [20, 21]. Moreover, very recent findings revealed how nucleosome structure, at a mono-nucleosome level, also aids higher chromatin order arrangement. Here, the importance of Swi6 (yeast HP1 homologue) binding to nucleosomes was further emphasised. Upon its binding, usually buried histone residues of a mono-nucleosome get exposed to the outside causing opportunistic weak multivalent inter-nucleosome interactions contributing to chromatin compaction [22].

Heterochromatin is further subdivided into constitutive and facultative heterochromatin. Constitutive heterochromatin is a repetitive type of chromatin state that is found in very condensed chromosomal parts, such as telomeres or centromeres in all cells. Therefore, it facilitates vital cellular functions, such as chromatid separation during mitosis or keeping chromosomes intact [23]. Of note, constitutive heterochromatin is not dynamically regulated. Conversely, the formation of facultative heterochromatin is developmentally controlled in order to silence only particular genes in a cell-specific manner, which overall assists cellular differentiation. This is largely achieved by PRC2 that deposits repressive H3K27me₃ marks on targeted genes [24]. In some cases, non-coding RNAs play a role in facultative heterochromatin generation. For instance, long non-coding RNA (lncRNA) *Xist* promotes facultative heterochromatin-mediated silencing of the X chromosome in homogametic sex organisms [25]. Lastly, euchromatin is characterised by high levels of H3 and H4 acetylation, and H3K4 methylation [26, 27]. Taken together, a wide range of

different chromatin remodellers and histone modification enzymes work either agonistically or antagonistically to balance between euchromatin and facultative heterochromatin in a dynamic and cell-specific way. A general trend is that undifferentiated embryonic stem cells (ESCs) exhibit decondensed and highly dynamic chromatin states, whereas differentiated cells acquire higher levels of heterochromatic regions positioning only cell type signature genes in euchromatic regions [28].

1.1.2 Enhancers

Enhancers are distal *cis*-regulatory elements that control activation and transcription of their target genes. They can be either in a close proximity or as far as 1 Mb-away from their regulated promoters [29]. Even though promoters are important for the gene expression, the spatio-temporal control of gene expression is ensured by these distal elements in a developing animal and across different cell types. Notably, they control gene transcription in a tissue-specific manner, that can be visualised by different enhancer reporter techniques. For instance, this can be done by cloning an enhancer of interest upstream of a minimal promoter and a reporter, such as green fluorescent protein (GFP) or lacZ, followed by testing the functionality of the engineered construct by assaying reporter expression *in vivo* [30, 31]. Active enhancers are found in an accessible chromatin regions and they are flanked by nucleosomes decorated with 'permissive' marks, such as H3K4me1 and H3K27ac, deposited by lysine methyltransferases (e.g., MLL) and acetylases (e.g., p300), respectively [26, 32, 33]. These active enhancer signatures can be identified using various booming genomics-based techniques. For instance, accessible chromatin regions can be detected with Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) or DNase I hypersensitive sites sequencing (DNase-seq) methods [34, 35]. Similarly, active enhancer marks can be identified using Chromatin Immunoprecipitation coupled with massively parallel Sequencing (ChIP-Seq) or similar methods. Additionally, enhancers can also be found in inactive or poised states due to mutually antagonistic actions of PRC2 and H3K27 acetyltransferases, which compete for H3K27 trimethylation or acetylation, respectively [36]. Poised or 'ready-to-go' en-

hancers are important for an efficient cellular differentiation due to their fast activation followed by H3K27me3 replacement with H3K27ac [37]. Importantly, a well-regulated developmental gene, such as Indian hedgehog (Ihh), is usually controlled by multiple enhancers. Their different combinations can function in an additive manner upon an enhancer-specific TF binding, which subsequently ensures a regulated gene expression in a right place and at a right time, and is thus crucial for embryo patterning and cellular differentiation [38].

1.1.3 Topologically associating domains and gene expression

Given that some of the enhancers are found hundreds of kilobases or even several megabases away from their regulated transcription start site (TSS), it is instinctive to assume that they need to be brought in a close proximity to their targets. Indeed, newly developed chromosome capture techniques based on proximity-ligation methods, such as Hi-C, 3C, 4C or Capture-C, have revealed a well-defined, compartmentalised, higher-order chromatin structures, usually referred to as topologically associating domains (TADs) [39]. Those approaches have demonstrated that DNA is packaged not randomly but in a pre-determined 3D structure within nuclei. Elements located within a single TAD are known to frequently interact with one other, but they get isolated from the other TADs and their elements [40]. This 3D compartmentalisation is achieved via insulator proteins at the TAD boundaries, such as CTCF and cohesin rings [41]. One of the possible mechanisms of how these protein complexes achieve building chromatin compartments can be explained by a loop extrusion model [42, 43]. This model proposes that chromatin gets progressively extruded through a cohesin ring until it encounters two CTCF molecules bound with convergent orientations that consequently anchors an extruded loop in a stable conformation leading to a TAD formation [42, 43]. Moreover, cohesin loading and cohesin release protein complexes, SCC2/SCC4 and WAPL, respectively, were shown to be also important for controlling size of an extruded loop by dynamically regulating cohesin's residence time on DNA [44]. Interestingly, recent findings have shown that not only cohesin-related proteins mediate higher chromosome

organisation. Veritably, PRC complex proteins (PRC1/PRC2) also drive long-range chromatin interactions associated with gene repression, which are normally counteracted by cohesin loading [45]. Importantly, it was pointed out that transcription itself does not advance the 3D chromatin contact formations, but the generated enhancer-promoter loops are thought to be essential for transcription [46, 47]. Indeed, it is plausible to imagine that such 3D chromatin compartments create local 'transcription factories' filled with gathered necessary transcription machinery that as a result enhances transcription efficiency within a busy nucleus [48].

However, how dynamic and how critical TAD formation is for developmental gene regulation is still a matter of debate. On a large scale, TADs are known to be rather constant. However, extensive differences were found at the intra-TAD level amongst distinct cell types or during embryonic development [49, 50]. Indeed, some of the known interactions are only structural, while others are functional [51]. For instance, it was shown that breaking a TAD boundary using CRISPR-based deletion, that brings *Epha4* limb enhancers closer to a *Pax3* gene, results in a *Pax3a* overexpression in developing mouse limbs in an *Epha4*-like expression pattern [52]. However, progressively more studies are showing that perturbing such higher-order chromatin structures have minimal effects on gene expression. For instance, one study has employed highly rearranged balancer chromosomes containing numerous inversions, duplications and deletions that disrupted multiple TADs in a developing fly embryo. Strikingly, these manipulations have revealed that such promoter-enhancer disorder did not affect expression of majority of genes, indicating that only a small portion of genes is sensitive to TAD re-arrangements [53]. Similarly, a well described *Shh* gene TAD perturbation was recently shown not to have major effects on its expression and have no obvious defects on mice embryo limb development [54]. Therefore, whether TAD formation was evolutionary crucial for a defined and precise gene control during embryo development, or whether it only serves a structural framework with a few exceptions for some gene regulation, or whether it plays a whole another role, still needs be investigated further.

1.1.4 Histone variants

In addition to canonical histones (H2A, H2B, H3 and H4), other histone variants exist in both animal and plant cells and they seem to play a significant role in maintaining stem cell identity as well as promoting cellular differentiation. Despite the fact that variant histones are very similar in terms of amino acid sequences to their canonical counterparts, there are several intriguing differences between them. First of all, canonical histone genes are found in clusters, while histone variant genes are expressed from low or even single copy genes [55]. Furthermore, canonical histone proteins are nearly always translated during the S-phase of the cell cycle, get incorporated into chromatin in a DNA replication-dependent manner and are found ubiquitously distributed throughout the genomes. On the other hand, the expression of variant histone proteins is constant during the cell cycle and their deposition is DNA replication-independent. Additionally, variant histones are not ubiquitously dispersed and are rather found in distinct genome places, such as TSSs or enhancers [55, 56]. Remarkably, canonical histone messenger ribonucleic acids (mRNAs) lack both introns and tail polyadenylation, whereas non-canonical histone mRNAs are polyadenylated, contain introns and some of them can even be alternatively spliced to generate distinct isoforms [57]. Variants of H1, H2A, H2B and H3 have been identified, with the most significant and best studied ones known as H2A.Z, macroH2A and H3.3 histones [58].

Histone H2A.Z is generally deposited by the ATP-dependent chromatin remodelling complexes SWR1 and INO80 at nucleosome-depleted regions of TSS. Multiple *in vivo* studies showed that its deletion is lethal in mice, *Drosophila*, and *tetrahymena* [59–61]. In addition, H2A.Z histone knock-down using RNA interference technology also causes faults in *Xenopus* gastrulation [62]. Strikingly, histone H2A.Z is very well conserved between species (around 80% of sequence conservation between most species), whereas conservation of the canonical H2A histones within the same compared species was found to be lower (around 60% sequence conservation) [63]. This indicates that H2A.Z carries a distinct function to that exhibited by the H2A proteins. In general, H2A.Z represses gene expression in ESCs, while activates certain genes in differentiating cells [57]. One

of the mechanisms of H2A.Z gene silencing is coupled to the interaction with Polycomb group (PcG) proteins. Chromatin immunoprecipitations (ChIP) of H2A.Z proteins using promoter arrays has revealed H2A.Z enrichment on various key developmental promoters in ESCs, such as Hox, Pax, Sox, Tbx, Fox and Wnt gene promoters. Its distribution patterns highly resembled those of Suz12 protein patterns, which is a part of PRC2. Markedly, 93% of H2A.Z regions were co-occupied by Suz12 [64]. Additionally, *in vitro* studies uncovered that this histone variant stabilises mononucleosomes and in this way contributes to compacting chromatin that overall leads to H2A.Z-marked gene silencing [55].

Remarkably, H2A.Z dispersal patterns systematically change over a period of differentiation [58]. One of the key mediating factors of a highly dynamic nature of H2A.Z protein was determined to be an extended C-terminal acidic patch that resides on the nucleosome surface. In contrast to the H2A protein, it contains three acidic amino acid residues [65]. Once H2A.Z gets incorporated into a nucleosome, its acidic amino acid patch is configured on a nucleosome surface. This causes a slight intrinsic destabilisation between H2A.Z and H2B dimers, as well as with H3-H4 tetramer, and also changes H1 protein binding to a nucleosome [65]. Initial *in vivo* studies revealed that disruption of H2A.Z acidic patch leads to early developmental defects in *Xenopus laevis* and *Drosophila* [62, 66]. Recent studies in mouse ESCs elucidated that this acidic patch of H2A.Z is necessary to couple chromatin dynamics and developmental gene expression regulation during cell lineage commitments [67]. One study that looked at mice ESC differentiation into endoderm/hepatic progenitor cells, revealed an important cooperation between H2A.Z and Forkhead box A (Foxa2) protein that mediates this lineage commitment. Their depicted differentiation model proposes that Foxa2 gets expressed upon growth-factor stimulation and then is recruited to nucleosomes containing H2A.Z histone variant proteins. This H2A.Z-Foxa2 interaction serves as a base for a recruitment of chromatin remodelling complex comprised of Nap111 (nucleosome disassembly/assembly chaperone protein), SWI/SNF and INO80. As a result, H2A.Z-Foxa2 marked DNA regions get their nucleosomes evicted that overall promotes gene expression and cellu-

lar differentiation. Particularly, promoter, exon and 5'UTR regions were enriched for H2A.Z-Foxa2-mediated nucleosome depletion [68]. Given that Foxa proteins were previously described as 'pioneer' factors in liver development, the proposed ESC differentiation model at least partially clarifies the mechanism behind an initiation of chromatin opening in differentiating cells [69].

Similarly, H3.3 histone variant was found to be important for epigenetic landscape shaping and transcriptional gene regulation [70]. In terms of a protein structure, H3.3 is also very similar to H3 canonical variants, differing only by 4-5 amino acids [70]. These small changes, for instance, serine 31, specifically found in H3.3 that can be phosphorylated, introduce different properties in comparison to the canonical H3 [71]. Usually, H3.3 is associated with transcriptionally active sites, which at least partially may be due to H3.3 nucleosomes being devoid of HP-1 linker histone [72]. Also, its protein structure is less stable to that of H3 proteins, making H3.3 positive nucleosomes more dynamic and easier to get displaced from DNA [73]. Conceivably for this reason, H3.3 nucleosomes are enriched at active gene bodies, promoters and enhancers contributing to open chromatin states [74, 75]. At the mouse pre-implantation stage, loss of H3.3 causes an increase in chromosome condensation resulting in frequent events of aneuploidy due to faulty chromosome segregation during first cell divisions [76]. Notably, even though absence of H3.3 does cause gene expression changes, studies conducted in the mouse ESCs overall demonstrated that they still remained pluripotent and global transcriptional output was not affected [77]. This, therefore, suggest that H3.3 might contribute to gene activation, but it is dispensable. Interestingly, H3.3 enrichment is associated not only with active genes, but also with some silent chromatin regions, such as telomeres and pericentric heterochromatin, where its function is not entirely as clear yet [74, 78].

1.1.5 Pioneering factors

During cell differentiation, an 'old' cell state needs to be erased, while a 'new' one gets established. For this to happen, requisite gene expression circuits need to be re-wired, which is initiated by chromatin remodelling. As described above, chromatin is found in

either active or repressive state. It can further be sub-divided to: type A - active chromatin decorated with permissive chromosome marks (H3K4me, H3K27ac and H3K9ac); type L - unprogrammed or 'low signal' chromatin (active or repressive chromatin marks are not abundant); and type R - repressive chromatin that is marked by repressive histone modifications (H3K27me or H3K9me) according to the review by Iwafuchi-Doi and Zaret [79]. While the type A chromatin can be accessed by all TFs, L- and R-type chromatin cannot be. Thus, for a cell differentiation to be initiated, the L-chromatin needs to be programmed towards a new fate and this can be, at least partially, achieved by chromatin pioneering factors. Unlike non-pioneering TFs, pioneering TFs can bind not only 'naked' DNA, but also DNA wrapped around nucleosomes, upon which, the nucleosomes get displaced leading to a chromatin 'opening'. Hence afterwards, A-chromatin is formed and other tissue-specific TFs can start initiating gene expression cascade responsible for definition of a new fate [79]. Two of the first uncovered pioneering factors were FoxA and GATA. They were shown to occupy liver-specific enhancers ahead of hepatic cell fate induction during embryonic endoderm tissue differentiation [80]. Later studies have supported that FoxA and GATA (enhanced by FoxA binding), can bind nucleosomal DNA assembled *in vitro*, whereas other hepatic TFs were not able to do so [81]. Importantly, Fox family members have 'winged helix' DNA-binding domain, which highly resembles that of nucleosome linker histone H1 [82, 83]. Based on these observations, it was experimentally validated that H1 gets displaced upon FoxA binding [79]. This gave further insights into how FoxA pioneering activity combined with histone linker removal leads to chromatin opening upon a targeted nucleosome binding. Moreover, other studies have also shown that FoxA, in comparison to other non-pioneer TFs, moves much slower in the nucleus suggesting its longer residence time on nucleosomal DNA giving it a better opportunity to co-operatively bind with other TFs or chromatin remodellers [84, 85]. An outstanding example of how pioneering factors can access L-chromatin and induce cellular reprogramming came more than a decade ago from a famous Yamanaka study showing how four TFs, Oct3/4, Sox2, Klf4 and c-Myc, were able to induce pluripotent stem cell (iPSC) state in fully differentiated fibroblast cells [86]. Indeed, Oct3/4, Sox2

and Klf4 were shown to be chromatin pioneer factors and were able to bind L-chromatin together or individually, while c-Myc was able to bind L-chromatin only by co-operating with other Yamanaka factors [87, 88]. Importantly, differently from FoxA pioneer factor that is able to recognise its full DNA binding motif even on nucleosomes, Oct3/4, Klf4, also basic helix-loop-helix type pioneer factors, such as Ascl1, recognise only a partial DNA binding motif when they bind to nucleosomes [88].

Recently, a study employing newly developed methods and computational tools, Cut&Run and Enhanced Chromatin Occupancy (EChO), respectively, have provided novel insights on TF chromatin binding modes in living cells, which could have not been addressed until now due to technical difficulties [89]. Firstly, it was reported that during multiple time points of endoderm differentiation from ESCs, known pioneer factors, such as FOXA2, GATA4, EOMES and SOX17, access targeted DNA by both nucleosomal and direct DNA binding. Strikingly, it was also shown that it is not chromatin configuration ('open' vs 'closed') that determines pioneer factor binding, but rather strength of underlying DNA binding motifs. In other words, pioneer factor-bound nucleosomes are not necessarily found in a closed chromatin confirmation. More importantly, pioneer factors favour low-affinity binding motifs during nucleosomal DNA occupation phase in comparison to a direct DNA binding. Also, contrarily to a previous belief, the notion that pioneer factors bind to nucleosomes prior to lineage-specification non-pioneer TFs was shown to be incorrect. Overall, this paints a much more dynamic, non-hierarchical chromatin regulation picture, ruled by various kinetic and thermodynamic constraints during *in vivo* developmental transitions, which was not anticipated before due to *in vitro* study discrepancies.

1.2 The neural crest

An embryonic stem cell-like population giving rise to a plethora of derivatives, transiently appearing after gastrulation, was discovered more than 150 years ago by Wilhelm His in the chick embryo and was initially called 'Zwischenstrang' (intermediate cord) [90]. The observed cell population was located above the closing neural tube and underneath

the superficial skin ectoderm. These cells displayed a remarkable ability to emerge from their place of birth, migrate laterally along the mid-dorsal line in aggregated streams of cells and subsequently give rise to the spinal ganglia and cranial nerves [90, 91]. Due to these phenotypical observations and nervous system development contributions, this transient cell population was later called the neural crest (NC) [91]. Ever since then, these distinctive attributes of NC have inspired the curiosity of developmental biologists for over 150 years revealing numerous discoveries of this remarkable embryonic cell population.

NC is multipotent stem cell-like population of cells, which get induced at the neural plate border between neural and non-neural ectoderm during gastrulation [92]. Early signals of the neural plate border specification consist of complex cross-talks between BMP, Wnt, Notch and FGF signalling pathways emanating from both the envrioning ectoderm and underlying mesoderm germ layers [93]. Subsequently, NC cell specification is marked by early expression of, so-called NC specifier genes, such as *Snail1/2*, *Sox8/9*, *FoxD3*, *Twist*, *AP2 α* and *Ets1* [94]. After specification, NC cells undergo epithelial-to-mesenchymal transition (EMT), that enables them to change cell adhesion and cell junction properties and delaminate from the neuroepithelium. Following the neural tube exit, NC cells then extensively migrate along well established canonical pathways to various, sometimes very distant sites within the embryo to give rise to a plethora of derivative cell types, including neurons and glia of the peripheral and autonomic nervous system, myofibroblasts forming the heart septum, smooth muscle cells of major cranial blood vessels, majority of the pigment cells in the body, as well as bone, cartilage and connective tissue elements of the craniofacial skeleton and many others (see Figure 1.1) [95].

Particularly, the evolution of vertebrates, especially a novel patterning of the craniofacial skeleton, is tightly associated with development of the neural crest [96]. Indeed, the development of so called vertebrate 'new head' is linked with the emergence of cranial NC (CNC) as well as ectodermal placodes [97, 98]. Firstly, CNC is known to give rise to connective tissue, osteocytes, chrondrocytes and smooth muscle cells, thus largely

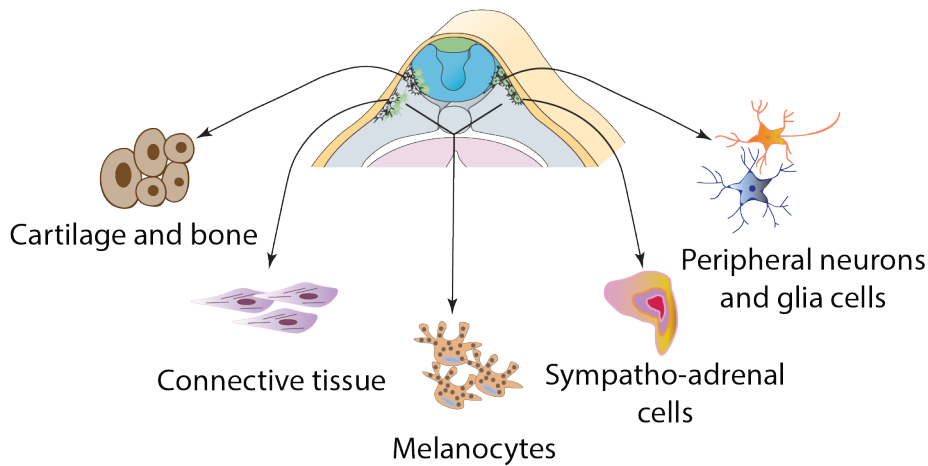


Figure 1.1 Overview of Neural Crest derivatives.

contributing to the cranial skeleton development in vertebrates [99, 100]. Importantly, even though NC does not directly contribute to vertebrate brain development, CNC provides BMP-antagonising signals that are important for fore-/midbrain development [101, 102]. Additionally, in the forebrain, CNC also gives rise to pericytes and meninges, which originate from mesodermal progenitors elsewhere in a body [103]. Moreover, the CNC is of particular interest for defining the jaw development in vertebrates due to a large NC contribution to the construction of this structure [104]. Therefore, understanding ancestry of jaw development has captured evolutionary biologists' attention for more than half a century. For instance, studying a basal jawless vertebrate - lamprey - is starting to reveal the ancestral and possibly core modules of NC gene regulatory network (GRN) allowing us to discriminate additional NC GRN modules, which give rise to higher vertebrate structures, such as jaw [105–107].

1.2.1 NC ontogeny from a GRN perspective

NC induction and specification

NC induction occurs during gastrulation and it takes place within the neural plate border between neural and non-neural ectoderm (see Figure 1.2). The key players allowing for this programme to commence are the extracellular signals, in particular WNT, FGF and intermediate levels of BMP tightly controlled by BMP antagonists, such as chordin,

noggin or follistatin, emanating from paraxial mesoderm [108]. At the GRN level, these induction signals culminate as the regulatory inputs controlling the expression of TFs specifying neural plate border. For example, Tcf/Lef that is a responsive mediator of Wnt signalling pathway, co-binds with lineage specifiers, such as the stem-cell factor FoxD3, at NC gene loci, hence initiating the NC induction [109]. As a result, the gradual induction of NC progenitors progresses in anterior-to-posterior (AP) direction during embryo neurulation [110]. Ultimately, described signals at the neural plate border

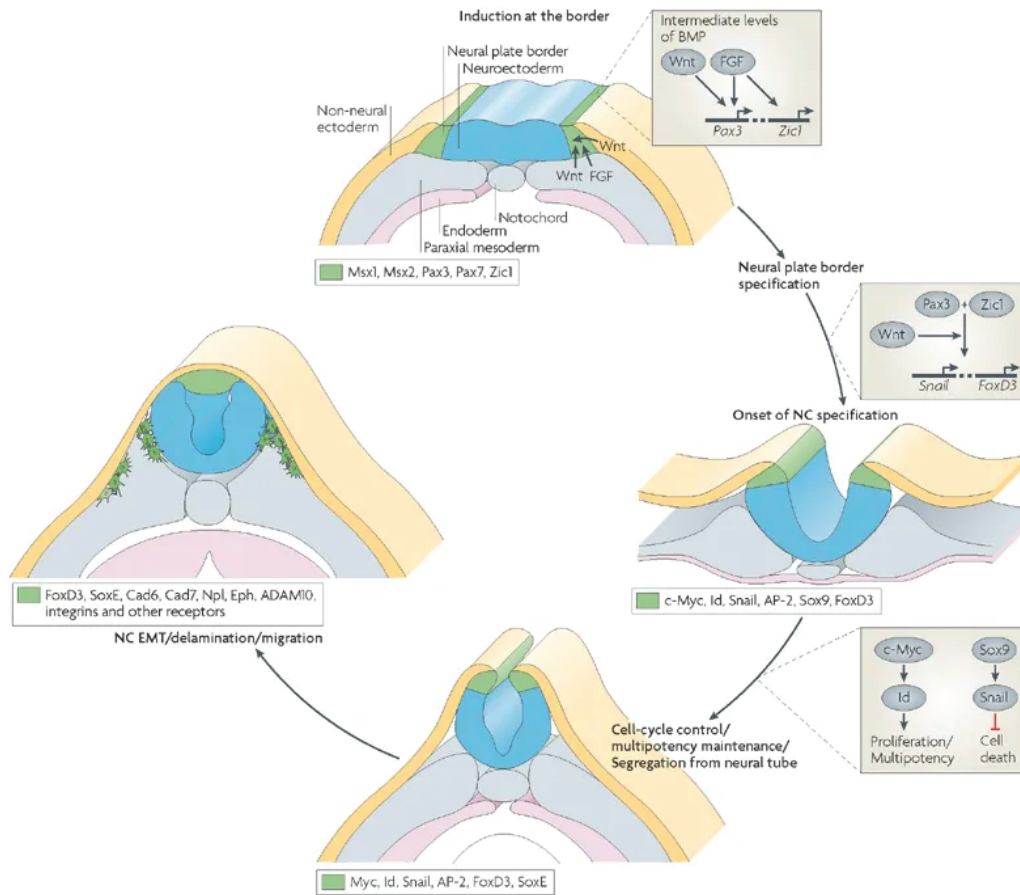


Figure 1.2 Regulation of neural crest formation. Adapted from the Sauka-Spengler and Bronner-Fraser, 2008 review [108].

activate expression of critical NC inducer genes, such as *Pax3/7*, *Zic1/2* and *Msx1/2*. Subsequently, these TFs activate expression of the *bone fide* NC specifier gene expression, such as, *Snai1*, *FoxD3*, *TFAP2s*, *Pax7*, *Ets1* and others (reviewed in [95, 108, 111], see Figure 1.3). The recent findings, including work later described in my thesis, are also

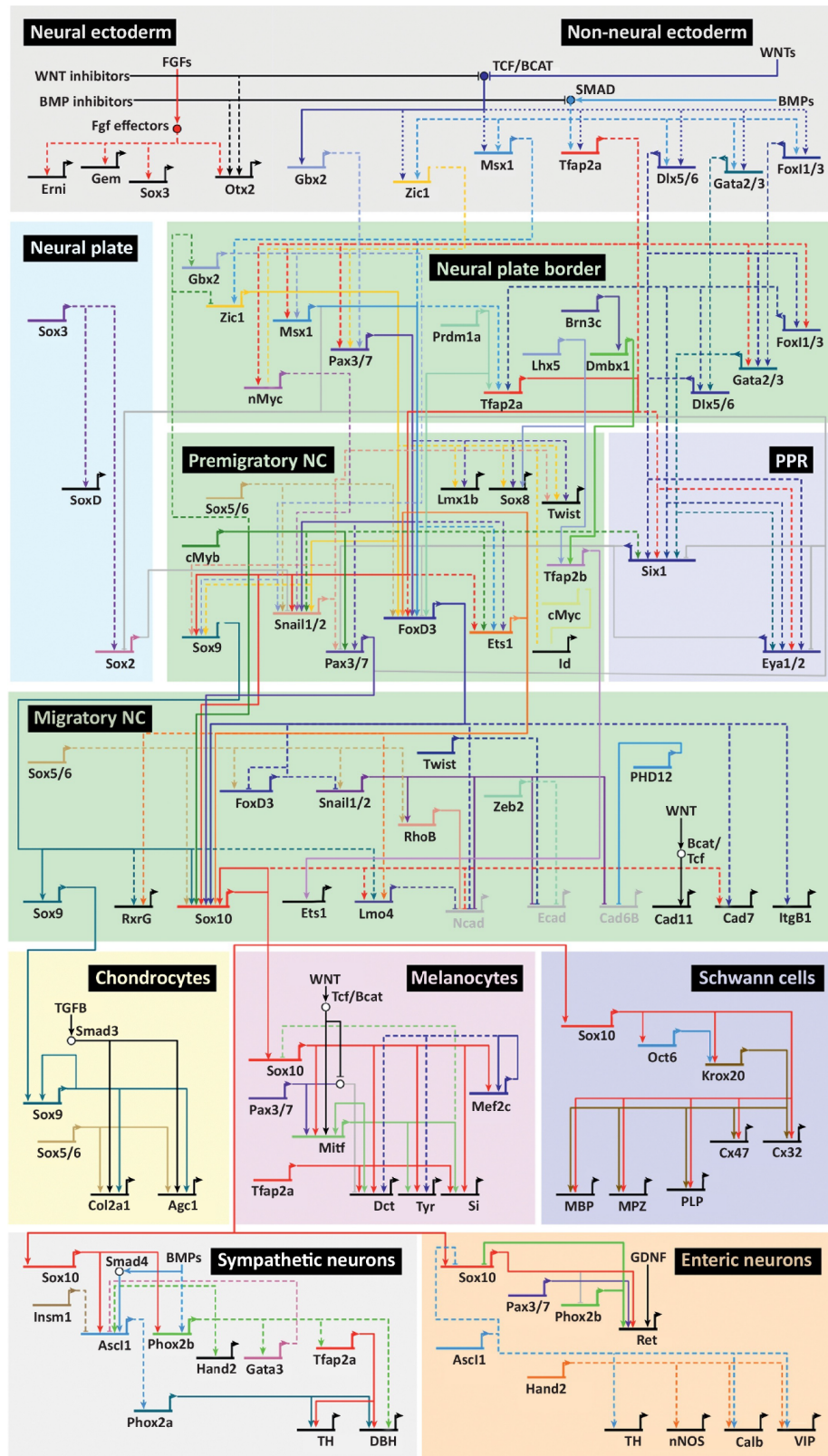


Figure 1.3 The gene regulatory network underlying NC ontogeny. Adapted from the Martik and Bronner, 2017 review [111].

pointing to some of these regulators, such as Pax7, Tfap2 α and FoxD3 contribute to NC specification via chromatin pioneering mechanisms, thus driving cellular transitions [112–114]. Importantly, some of these pioneer factors, such as FoxD3 and Tfap2 α are very likely to co-operate, since at very early NC stages (70% epiboly in zebrafish) their expression is complementary and double mutants completely abolish NC derivatives [115]. Notably, currently described hierarchy of the NC induction/specification GRN is not definitive and hence is perpetually updated, particularly at the current age of booming *-omics* approaches allowing unbiased genome-wide interrogation of NC regulators and GRN across the vertebrate taxa [116, 117].

NC delamination and migration

Once NC is specified, it has a notable ability to undergo EMT. Firstly, NC cells reshape their cellular properties, which allows them to delaminate from a neural tube, acquiring mesenchymal properties and migrating to sometimes very distant sites in a developing embryo [108]. Evidently, for a NC EMT to take place, extracellular matrix (ECM) needs to be remodelled and adherens junctions need to be dismantled. This is achieved by a change of cadherin type molecules that transitioning NC cells express. Cadherin switch happens at the beginning of neural induction when cells switch from expressing type I cadherins (E-cadherin) and start expressing the type II cadherins (Cadherin 6/7/10/11) [118]. Upon the onset of NC EMT, N-cadherin expression decreases and also expression of weaker type II cadherins gets induced. Remarkably, these steps of EMT are controlled by the previously described NC specifiers (see Figure 1.3). For instance, at the trunk NC regions, this is controlled by Snail, FoxD3 and Sox9/10 TFs, while at the cranial regions these TF work together with Ets1, LSox5 and p53 [119–123]. Importantly, Wnt signalling was also shown to contribute to Snail/2 transcription, that is crucial for the initial NC delamination via both, type I and type II cadherin, repression [124, 125]. The second part of NC EMT is ensured by upregulation of other migration factors, such as Twist and Zeb2, which downregulate E-cadherin expression in delaminating cells allowing their further separation from each other [126, 127]. Moreover, other *bona fide* NC specifiers,

such as FoxD3 and Sox10 were shown to also repress N-cadherin expression [119, 128]. Additionally, FoxD3 was also found to further contribute to EMT by repressing another transmembrane protein, Tspan18, important for tight cell adhesion establishment and maintenance [129, 130]. Furthermore, delaminating NC cells also express metalloproteases, such as ADAM10/13 and MMP16, which cleave cadherins allowing basement membrane degradation and hence, this way, contributing to the ECM remodelling that is needed for the migration to happen [131–133].

Upon acquiring the mesenchymal fates, NC start receiving and integrating positive and negative signals coming from the surrounding tissues via the cell-to-cell and cell-ECM contact interactions [134]. In the cranial regions, NC assembles into three main migrating streams due to inhibitory class 3 semaphorin signals secreted by the otic vesicle, rhombomeres and eye regions [134]. Ephrin signalling also largely contributes to dividing NC cells into distinctive streams and invading only particular tissue depending on the expressed Eph signalling molecules [134]. For instance, EphA4+/EphB1+ receptor tyrosine kinases expressing NC cells invade only the third and the fourth branchial arches (BAs) and repulse ephrin-B2 ligand positive NC cells, which end up migrating into the second BA [135]. Conversely, the positive cues guiding NC cells are stromal cell derived factor 1 (Sdf1), vascular endothelium growth factor (VEGF), FGF2/8 and glial cell-derived neurotrophic factor (GDNF) [136–140]. In the trunk regions, NC cells migrate in distinctive ventromedial or dorsolateral pathways, again guided by paraxial mesoderm signals and Eph signalling molecule compatibility, respectively. Importantly, ventromedially migrating crest largely contributes to the trunk peripheral nervous system (PNS) development, while dorsolaterally migrating crest gives rise to pigment cells [141, 142]. Of note, Slit-Robo chemorepulsive interactions also play an important role for the trunk NC guidance [143].

Moreover, provided signals in between migrating NC cells leading to either contact inhibition of locomotion or cellular co-attraction are also important for a proper NC cell migration. The contact inhibition of locomotion that helps NC cell dispersal is primarily driven by non-canonical Wnt/planar cell polarity (PCP) pathway activating small RhoA

GTPase and N-cadherin contact drives inhibition of Rac1 GTPase, that overall results in migrating NC cell polarisation [144, 145]. On the other hand, NC cells mostly migrate in large groups and are co-attracted to each other via C3a-dependent chemotaxis [146]. Strikingly, recent results have revealed that migrating non-lineage restricted trunk NC cells have a previously overlooked function during embryo development. NC cells were shown to respond to interleukin-1 β signals coming from apoptotic cells within central and peripheral nervous systems. This way NC cells get recruited and subsequently, similarly to macrophages, phagocytose cellular debris around the neural tube and spinal cord [147].

1.2.2 Distinctive axial NC competence and associated NC derivatives

Very elegant early NC fate mapping studies, involving distinctly looking quail cells being transplanted into chicken embryos, revealed that NC displays axial-specific heterogeneity [148, 149]. This was later supported by multiple studies showing that, indeed, despite NC being multipotent, it gives rise to different derivatives depending on the axial level of its origin (see Figure 1.4, reviewed in [103]). For example, only cranial NC contributes to cartilage and bone within head regions, whereas cardiac NC, part of the vagal NC, uniquely contributes to the outflow tract of a developing heart [150–152]. Interestingly, all compartments of NC can give rise to melanocytes [111]. The axial-specific heterogeneity is influenced by expression of different combinations of TFs. For example, the homeotic *Dmbx1*, *Cdx* and *Otx2* factors instruct the specification of cranial NC identity, whereas anterior *Hox* genes (e.g., *HoxA2*, *HoxB2*) and Retinoic Acid receptors pattern trunk NC, which highlights principles of GRN modularity in an axial-specific manner [153]. Importantly, by understanding these distinctive axial GRN modules, we can transform one NC cell fate for another. For example, ectopic delivery of cranial-specific module to the trunk NC, allowed its reprogramming resulting in cranial-specific chondroblast differentiation [151]. NC profiling at a single-cell level revealed further transcriptional regulatory differences underlying cranial and trunk NC fate restrictions that emerges during cell delamination in a bifurcated rather than in explicit transcriptional manner [116]. Strik-

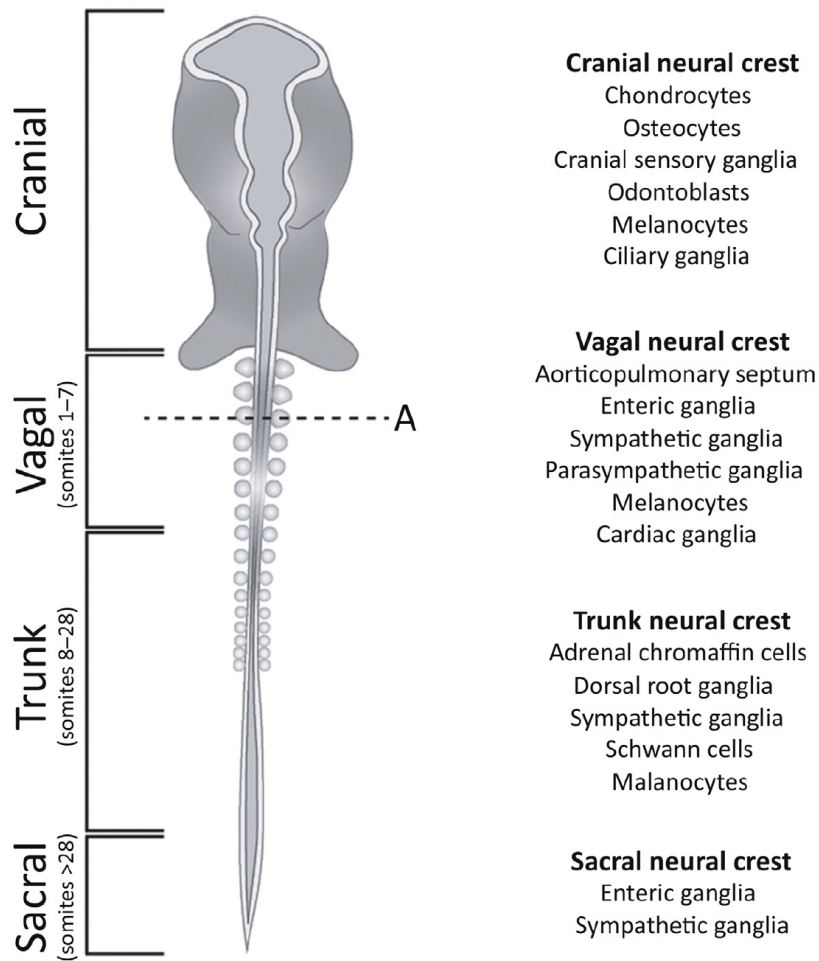


Figure 1.4 Overview of the NC anterior-posterior axis derivatives. Adapted from the Martik and Bronner, 2017 review [111].

ingly, overexpression of usually cranial *Twist1* regulator in the chick trunk, resulted in trunk crest reprogramming towards mesenchymal route, which was not achieved in the previous trunk reprogramming study [116, 151]. Notably, cranial NC was shown to acquire its positional identity already at the pre-migratory NC stages at the chromatin level [154]. Distinctive sets of positional gene promoters and enhancers get bivalently marked (H3K4me2 and H3K27me3 marks, the latter is deposited by *Ezh2*, part of PRC2) in a pre-patterned manner within the progenitor cells. Marked elements are maintained at such poised state during the cranial crest migration, later resulting in different sets of genes getting expressed at post-migratory stages, due to a repressive H3K27me3 mark switch to an active H3K27ac mark [154]. Similarly, our laboratory has also revealed that

NC is pre-patterned towards either canonical NC cell fates or neuronal-restricted cell fates already at pre-migratory NC stages via activation of different sets of enhancers in the progenitors [117].

Taken together, it is evident that tight NC regulation is critical at multiple steps of a vertebrate embryo development. Not surprisingly, multiple birth malformations and developmental disorders are associated with faults of NC development [155–158]. Moreover, a number of different cancers employ similar principles of cellular migration and self-renewal or even re-activate NC programmes [159]. Therefore, studying basic, conserved principles of NC ontogeny, holds a potential for stem-cell like therapies and cancer treatments.

Bimodal Action of *foxd3* is Required to Specify Neural Crest Cells and Maintain their Multipotency *in vivo*

In this work, we investigate the molecular mechanisms by which *foxd3* influences neural crest (NC) development from its induction through NC specification to the NC progenitor differentiation by taking advantage of fluorescent *foxd3* mutant zebrafish lines. By employing cell-specific genome-wide profiling, we were able to characterise the transcriptional and epigenetic landscape of *foxd3*-mutant and control cells *in vivo*. We have also generated and employed new transgenic zebrafish lines for the biotin ChIP-seq method to further investigate direct and indirect *foxd3*-mediated gene regulatory roles during NC ontogeny and better part of this work was already published. In summary, we demonstrate that *foxd3* drives several independent chromatin-organising mechanisms, switching from activator to repressor roles to orchestrate multiple regulatory programmes during NC formation. First, it starts with priming early NC specification by regulating essential signalling pathways, then maintains NC multipotency by controlling stem cell programmes and prevents premature migration and differentiation into neuronal NC derivatives [113].

2.1 *Foxd3* - a critical transcription factor for an early embryo development

The forkhead/winged helix transcription factor FoxD3 is an important multifaceted stem cell factor and genetic FoxD3 knockouts result in embryonic lethality [160]. It functions during several stages of embryonic development with a predominant role of maintaining uncommitted cells in their multi-/pluripotent states [160–162]. For instance, at early stages of development, FoxD3 is thought to maintain pluripotency of both trophoblast progenitors and epiblast cells in the mouse embryo [160, 161]. Furthermore, absence of

FoxD3 causes premature mouse ESCs differentiation into mesendodermal lineages while ectodermal lineage markers get reduced [163]. Interestingly, FoxD3 was recently demonstrated to play a role not only during pluripotency maintenance, but also during cell fate transitions of such pluripotent cells [164–166]. During the transition from the naïve ESCs to the primed pluripotency epiblast-like cells, FoxD3 represses enhancers by recruiting H3K4 demethylase, *Lsd1*, resulting in a decrease of active enhancer marks and an increase in inactive enhancer marks. Such changes allow the early epiblast cells to exit the initial naïve state [165]. On the other hand, in another *in vitro* study, FoxD3 was also shown to drive early epiblast differentiation (ESCs) into the late epiblast (EpiCs) stage of primed-pluripotency [166]. In this case, FoxD3 was shown to perform a dual role: in the ESCs it acts as a repressor by recruiting histone deacetylases (HDACs) to the target sites, while in the EpiCs, FoxD3 occupies new set of enhancers. It first primes them for future activation by Brg1-mediated nucleosome removal, followed by HDACs recruitment to keep newly-primed enhancers in the poised state [166]. Seemingly, at early and late epiblast stages, FoxD3 keeps embryonic cells at the ground state of pluripotency as well as contributes to the control of early cell fate transitions. Intriguingly, this forkhead transcription factor was also shown to have similar functions during the NC development. FoxD3 is required for the NC progenitor maintenance and its absence leads to a catastrophic decrease in NC derivatives, such as loss of craniofacial cartilage skeleton and cranial nerves of the peripheral nervous system, iridophore pigment cell reduction, reduction in size of dorsal root ganglia and absence of NC-derived parts of the enteric nervous system (ENS) in both mice and zebrafish embryos [162, 167–169]. Essentially, similar to its role in ESCs, in the NC FoxD3 maintains multipotency of NC progenitors by preventing their premature differentiation [162, 169]. However, FoxD3 was also implicated in controlling cell fate decisions at several NC derivative lineage nodes. A general observation is that FoxD3 needs to be gradually down-regulated in the progressively fate-restricted NC lineages [128, 170, 171]. However, at some NC lineage crossroads, FoxD3 plays more active gene regulatory roles. First, FoxD3 is involved in the neuronal/glia cell lineage segregation from the melanocyte cell lineage within the ecto-

dermally derived multipotent NC cells [172–174]. One of the shown mechanisms for this lineage split is a direct or indirect FoxD3-driven repression of Microphthalmia-associated transcription factor (MITF), one of the main transcriptional drivers of the melanogenesis programme [173, 175]. FoxD3 is also involved in the neuronal versus glial cell fate segregation in sensory ganglia and acts as melanophore/iridophore switch in bipotent pigment precursors [172, 176]. FoxD3 favours glial cell differentiation at the expense of dorsal root ganglia sensory neurons or enteric nerves and acts by inhibiting neuronal programmes in the bipotent neuronal/glial progenitors [169, 172, 176]. Therefore, it is evident that FoxD3 is the 'guardian' of the NC multipotency, with the occasional take up of a leader position during some of the NC cell fate transitions. However, the comprehensive mechanisms of such different roles throughout the NC ontogeny are yet to be resolved. One of the potential explanations could be that, even though it was initially identified as the transcriptional repressor, depending on the context, FoxD3 may also affect transcription positively. Emerging evidence in the context of EpiCs, where FoxD3 was shown to be capable of priming enhancers by removing nucleosomes together with Brg1 during early to late epiblast transition [164, 166, 177–179], is indeed pointing to its gene activating capabilities. Moreover, given that FoxD3 forkhead DNA binding domain, similar to that of FoxA chromatin pioneering proteins, is comprised of three helices and two large loops ('wings'), and is remarkably akin to the winged-helix structures of linker histone H1 that avidly binds nucleosomes, it is possible that FoxD3 is capable of an independent nucleosome removal and, hence, regulatory DNA priming during cell fate transitions [82]. Additionally, such pioneer factors have also been suggested to induce nucleosome repositioning, possibly by recruiting co-factors, such as hyperdynamic histone variants, like H2A.Z and H3.3 and other chromatin and DNA modifying proteins, allowing transcriptional complexes to bind to *cis*-regulatory elements [180–182]. Intrigued by these outstanding questions, we sought to understand how foxd3 could be playing multiple roles across different stages of NC development using zebrafish embryos as the NC model system. Investigated embryonic stages are illustrated in the Figure 2.1.

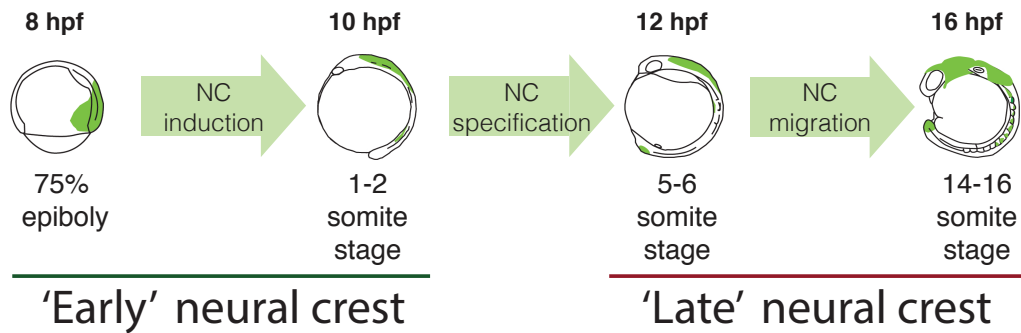


Figure 2.1 Zebrafish embryo stages examined in this study. 75% epiboly – a gastrulation stage during which embryonic shield and hypoblast are formed. 1–2 and 5–6 somite stages – induced and specified premigratory neural crest (NC), respectively. 14–16ss – migratory NC. 75% epiboly and 1–2ss will be referred as 'Early' NC, while 5–6ss/14–16ss – as 'Late' NC in this chapter. *Foxd3* expression is labelled in green; hpf – hours post fertilization.

2.2 Mutant lines to study gene regulatory roles of *foxd3* throughout the NC development

One of the most powerful means to understand a function of a gene of interest is to employ reverse genetics tools, such as ones for gene knock-out or knock-down strategies. For example, some of the previous studies that also tried untangling the *foxd3* regulatory role in NC development, have employed the *foxd3* anti-sense morpholino oligonucleotides [169]. Even though the exemplary study provided stepping-stone findings of *foxd3* necessity for NC derivative, such as jaw structures, glia, dorsal root ganglia neurons, enteric neurons, and iridophore pigment cell formation, morpholino-based gene knock-downs were called into question by the developmental biology community [183]. This was due to effect discrepancies between morphants and genetic knock-outs, where in some cases morphants displayed much more severe phenotypes and hence morpholino-specificity was questioned. Moreover, recently shown genetic compensation in mutants, but not morphants, raises another level of concern and hence tight regulations and controls need to be taken into account, when using morpholinos [184]. In addition to *foxd3*-morpholino studies, two mutant zebrafish lines, called *mos* and *sym1*, were identified in large-scale mutagenesis screens more than a decade ago [167, 168]. The sympathetic mutation 1 (*sym1*) was generated using *N*-ethyl-*N*-nitrosourea (ENU)-based mutagenesis and was identified as embryonic lethal *foxd3* mutation that leads

to absence of sympathetic neurons and defects in craniofacial development. This mutation is a frame-shift mutation causing a *foxd3* protein truncation. It was predicted to cause a loss of the second winged domain, nuclear localisation signal (NLS) and C-terminal transcriptional effector domain of the *foxd3* protein that all together disrupt *foxd3* localisation to the nucleus and binding to its DNA target sequences [167]. However, another study showed that *sym1* mutation was not a null, but rather a hypomorphic mutation, that only reduces and not completely abolishes *foxd3* function [185]. The mother superior (*mos*) mutant did not entirely complement the *sym1* phenotype. *mos* mutation was mapped outside the *foxd3* coding region, hence the mutation possibly perturbs one of the *foxd3* crucial *cis*-regulatory elements [168]. Therefore, given these concerns, for our *foxd3*-mediated NC gene regulatory studies, we have decided to employ the novel *foxd3* gene trap lines taking advantage of incorporated fluorescence reporters [186]. These lines were identified in the FlipTrap Tol2-mediated gene trap screen [187]. The gene trap cassette containing a citrine fluorophore sequence flanked by *rassf8* splice acceptor and donor sites, was integrated into the endogenous coding region of the *foxd3* gene just upstream of the forkhead domain. This cassette insertion generates a truncated cytoplasmic *foxd3* protein with no NLS or DNA-binding domain fused to a citrine reporter and a short tail of a nearby gene *alg6*. This stable zebrafish line is referred to as *Gt(foxd3-citrine)^{ct110a}* (see Figure 2.2 A). The FlipTrap cassette was designed to also carry a mCherry fluorophore sequence with a stop signal and polyA tail, in an inverted direction, flanked by loxPV sites (see Figure 2.2 B). Therefore, once *Gt(foxd3-citrine)^{ct110a}* line was established, the integrated cassette was flipped upon *cre*-mRNA injection to mediate Cre-loxP recombination and hence generate a *Gt(foxd3-mcherry)^{ct110aR}* line (Figure 2.2 B). Notably, heterozygous *foxd3* mutants, with either mCherry or citrine fluorophore, develop normally, without any signs of haploinsufficiency [186]. The advantage of using these two lines was that each fluorophore expression recapitulated endogenous *foxd3* expression, allowing us to isolate both, *foxd3*-expressing cells (C, citrine+/WT) as control cells and corresponding *foxd3*-null cells (CC, citrine+/mCherry+) for our experimental comparisons. The strategy behind the exper-

2 Bimodal Action of *foxd3* is Required to Specify Neural Crest Cells and Maintain their Multipotency *in vivo*

experimental design for transcriptomics and epigenomics analyses comparing gene expression and genome accessibility between *foxd3*-mutants (CC) and control (C) embryos is illustrated in Figure 2.3. In short, the *Gt(foxd3-citrine)^{ct110a}* and *Gt(foxd3-mcherry)^{ct110aR}* lines were crossed, embryos were sorted and dissociated into single cell suspensions. Then embryonic cells obtained from corresponding clutches were isolated using Fluorescence Activated Cell Sorting (FACS) to acquire Citrine-only expressing *foxd3*+ cell samples (C) as control and *foxd3*-mutant cell samples expressing both fluorophores (Citrine and Cherry; CC).

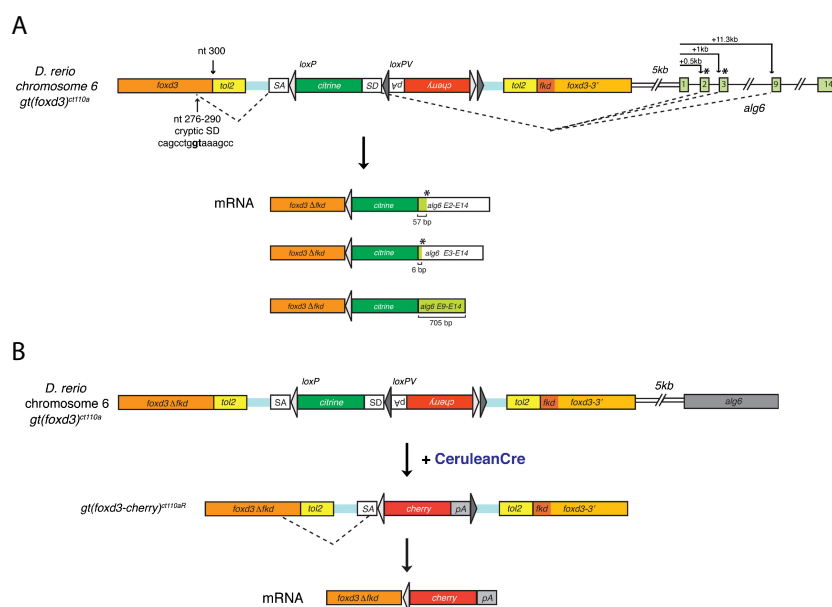


Figure 2.2 Logic behind the *foxd3* gene trap lines that were engineered by Tatiana Hochgreb-Hägele and Marianne E. Bronner. Figure adapted from [186].

(A) A diagram of the *foxd3*-citrine FlipTrap cassette used for the *Gt(foxd3-citrine)^{ct110a}* line. The cassette was integrated within the coding region of *foxd3*. Splice acceptor site (SA) of the citrine internal exon recognises a cryptic splice donor site within *foxd3* sequence, while its splice donor site (SD) splices to acceptor sites of exons 2, 3 or 9 of a nearby gene *alg6*. Such splicing events produce chimeric proteins carrying a very short N-terminal part of FoxD3 (FoxD3-ΔFkd) with no NLS signal, citrine and an Alg6 tail. Splicing of citrine with exon 2 or 3 of *alg6* introduces a premature stop codon (*), while splicing with exon 9 does not affect the coding frame. (B) A diagram of the of Cre-loxP recombination in *Gt(foxd3-citrine)^{ct110a}* zebrafish line to generate a *Gt(foxd3-mcherry)^{ct110aR}* line. The integrated FlipTrap cassette contains inverted mCherry (depicted as a redbox) and polyA (pA) sequences, flanked by heterotypic loxP (light gray arrow box) and loxPV (dark gray arrow box) sites. Upon cerulean cre-mRNA injection into the *Gt(foxd3-citrine)^{ct110a}* embryos, recombination of the loxP sites causes a deletion of citrine and mCherry-pA flip into a correct orientation. In this case, flipped construct does not splice to the *alg6* gene and the end result of these cryptic splicing events is a mutant FoxD3-ΔFkd-mCherry fusion protein.

2.2 Mutant lines to study gene regulatory roles of *foxd3* throughout the NC development

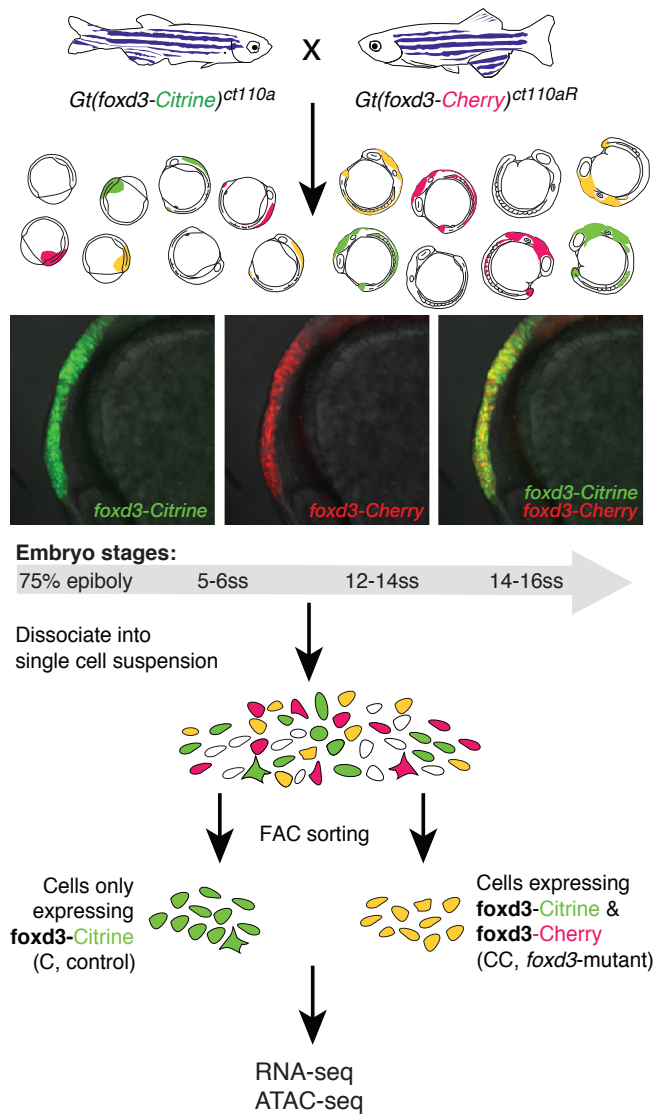


Figure 2.3 Experimental strategy for obtaining *foxd3*-mutant (yellow) and *foxd3*-control (green) NC cells for comparative transcriptomics and epigenomics analysis. Mutant (Citrine/Cherry; CC) and control (Citrine only; C) NC cells were isolated by FACS from in-crosses of heterozygote *foxd3*-mCherry and *foxd3*-Citrine fluorescent *foxd3* transgenic fish at 75% epiboly, 5-6ss, and 14ss embryonic stages.

2.3 Knockdown of *foxd3* Results in Decreased Expression of NC Specifier Genes and Increased Expression of Genes Controlling NC Migration and Differentiation

At 75% epiboly, differential expression analysis between *foxd3*-mutant (CC) and control samples (C) yielded comparable numbers of upregulated and downregulated genes (Figure 2.4 A). Interestingly at later NC developmental stages (5-6ss and 14-16ss), a larger part of differentially expressed genes consisted of upregulated rather than downregulated genes, suggesting that more genes are putatively repressed rather than activated by *foxd3* at this point in comparison to earlier stages (Figure 2.4 A). This might indicate a possible transient change between activator and repressor roles of *foxd3* during NC ontogeny. Notably, even though sets of either upregulated or downregulated genes were largely distinct at different stages, there was some overlap between 5-6ss and 14-16ss, especially amongst the genes that get de-repressed in *foxd3* mutants (Figure 2.4 A). Further Gene Ontology (GO) analysis focusing on over-represented terms amongst differentially regulated genes indicated that at 75% epiboly, *foxd3* putatively represses cell metabolism pathways, such as ribosome biogenesis, RNA processing and translation genes. It also appeared to negatively control genes involved in progenitor cell adhesion and migration (e.g. *nrp2a*, *nrp1b*, *slit1a*, * $p < 0.05$) (Figure 2.4 A-B). On the other hand, GO terms associated with the downregulated genes showed that at the same time *foxd3* is also potentially priming genes involved in tissue-specific programmes, such as neural plate border or otic placode formation (e.g. *gata2a*, *gata5*, *ets1*, *six1a/b*, *tfap2a/c*) (Figure 2.4 A-B). Most strikingly, *foxd3*-mutant cells (CC) showed downregulation of a large proportion of known NC induction and specification genes (** $p < 0.01$), many of which were *bona fide* NC transcription factors (~40%) and signalling or cell junction and adhesion molecules (~25%) (Figure 2.4 A-C). Remarkably, some of the downregulated NC specification factors, such as *prdm1* and *tfap2a/c*, have been previously shown to act upstream of *foxd3* (Figure 2.4 A) [108, 188, 189]. Hence, our findings are challenging the previously proposed NC gene regulatory network (GRN) and puts *foxd3* higher up in the NC regulatory hierarchy. Overall, statistical over-representation of the entire

2.3 Knockdown of *foxd3* Results in Decreased Expression of NC Specifier Genes and Increased Expression of Genes Controlling NC Migration and Differentiation

set of genes downregulated at 5-6ss in *foxd3*-mutant samples yielded highly significant association with neural crest and stem cell development GO terms, also terms linked to onset of EMT, cell adhesion changes and NC cell migration (Bonferroni $**p < 0.01$, Figure 2.4 B). Additionally, some of the enriched terms were also linked to NC derivative fates, such as melanocyte differentiation, cranial skeleton morphogenesis, autonomic and enteric nervous system development due to downregulation of the core NC TFs (*pax3a*, *sox9b*, *sox10*, *tfap2a*) that act both in NC specification and later in NC differentiation. Importantly, downregulation of NC specifiers was confirmed by *in situ* hybridisation (Figure 2.4 D). Another interesting observation that emerged from our differential gene expression analysis was that several paralogues from the same gene family, including the *fox* gene family, were differentially expressed in the mutant cells. For instance, some of the key NC factors, such as *snai1b* or *twist1b* were downregulated, whilst their corresponding paralogues, *snai2* and *twist2*, were upregulated (Figure 2.5). This suggested that multi-layered compensatory mechanisms were critical for robustness of the neural crest regulatory programme and embryonic development in general. First, in the absence of *foxd3*, other *fox* genes are upregulated to preserve *foxd3*-dependent gene regulation. Second, the downregulation of genes that are normally activated by *foxd3*, is rescued by the paralogue gene upregulation to compensate for such loss. This embryo developmental robustness is in line with our phenotypical *foxd3*-mutant observations explaining why zebrafish neural crest development does not completely terminate during NC specification; and clearly visible defect and failure to settle and differentiate into NC derivatives is observable only at 14ss onwards. Zebrafish mutant larvae die by 15 days post fertilisation (dpf), most likely due to defects in axial mesodermal structures [186].

Altogether, our transcriptomics results showed that *foxd3* might be playing different gene regulatory roles (i.e. both activating and repressing), possibly depending on the temporal context as we observed a continuously increased number of putatively repressed genes throughout the NC development. Most importantly, in the absence of a functional *foxd3* protein, much of the NC specification module is absent at 5-6ss. Additionally, we also found a potential mechanism for the mutant compensation due to the upregulation of

2 Bimodal Action of *foxd3* is Required to Specify Neural Crest Cells and Maintain their Multipotency *in vivo*

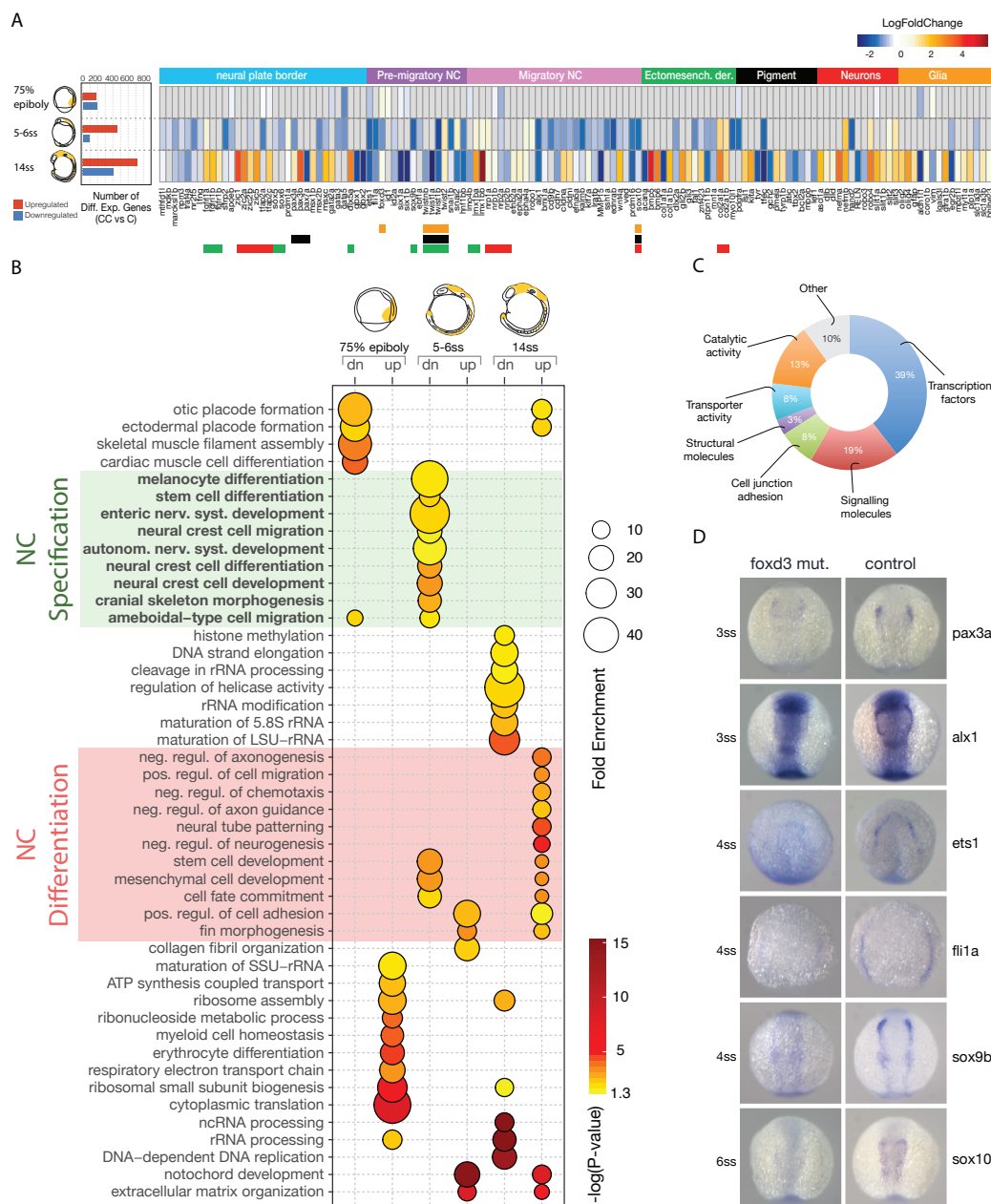


Figure 2.4 Transcriptome changes in the *foxd3*-mutant cells during different stages of NC development. **(A, left)** Bar plots comparing numbers of upregulated and downregulated genes in *foxd3*-mutant and control NC at 75% epiboly, 5-6 and 14 somite stage (ss) zebrafish embryo stages. **(A, right)** Heatmap showing fold change in expression of known NC genes between *foxd3*-mutant and control cells at 75% epiboly, 5-6ss and 14ss. Selected genes are grouped to reflect known NC-GRN structure. **(B)** Bubble plot highlighting fold enrichment and $-\log(P\text{-value})$ (Benjamini-Hochberg corrected) for the most significant gene ontology biological process terms associated to either downregulated (dn) or upregulated (up) genes in *foxd3*-mutant versus control embryos. **(C)** The pie chart showing distribution of gene ontology molecular protein function of the down-regulated genes in *foxd3* mutants. **(D)** *In situ* hybridization of 3-6ss zebrafish embryos (dorsal view) showing decrease/loss of expression of NC specifier genes in *foxd3*-mutants.

2.4 Application of the Biotin ChIP-seq method to identify direct *foxd3* gene targets throughout the NC development

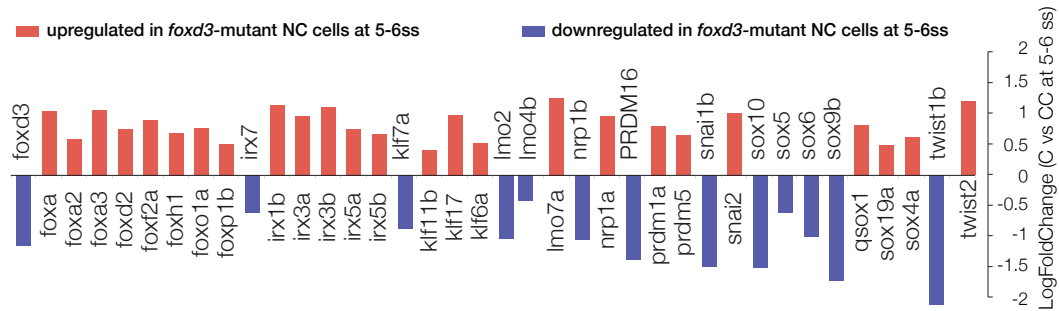


Figure 2.5 Potential *foxd3*-mutant rescue mechanism by paralogous NC factors. Bar plot representing fold change in expression of NC factors showing that paralogues are differentially regulated in control and *foxd3*-mutant NC cells.

other Fox proteins and NC factor paralogues that all together partially rescued NC specification programme (Figure 2.5). On the whole, genes governing stem cell and NC specification programmes get downregulated, while NC specification and migration genes become de-repressed in the mutant NC. This suggests that *foxd3* first largely contributes to NC specification (either directly or indirectly) by participating in targeted gene activation, and then switches to a repressor role to prevent pre-mature NC migration and differentiation.

2.4 Application of the Biotin ChIP-seq method to identify direct *foxd3* gene targets throughout the NC development

In order to understand whether the observed bimodal action of *foxd3* is direct or indirect in terms of NC gene regulation, we employed our biotin ChIP-seq method (described in Chapter 5.2.2) with the goal of tracking genome-wide dynamics of direct *foxd3* binding from early steps of NC induction (75% epiboly, 1-2ss) and specification (5-6ss) to migratory NC stages (14ss). To achieve this goal, we first needed to obtain the required transgenic zebrafish lines, i.e., one line ubiquitously expressing nuclear biotin ligase (NLS-BirA) and the other line expressing an Avi-tagged version of the *foxd3* protein in an endogenous-like manner. Hence, upon crossing these two lines, *foxd3* protein would be biotinylated *in vivo* and co-precipitated using streptavidin beads to identify its genome-wide binding maps during different stages of NC ontogeny. The BirA driver

line expressing the ubiquitous biotin ligase under the control of ubiquitous *ubb* promoter targeted to the nucleus, *Tg(ubiq:NLS-BirA-2A-Cherry)^{ox114}*, was generated using the Tol2-mediated plasmid transgenesis by Vanessa Chong-Morrison for previously developed binary biotagging approach [190].

2.4.1 Generation of the *foxd3*-Avi transgenic fish line

To obtain the Avi-tagged *foxd3*-expressing zebrafish line, we used a transposon-mediated BAC transgenesis method overviewed in Chapter 5 (Materials and Methods) and Figure 2.6.

We chose the CH211-196F13 *Danio rerio* BAC clone (size of 203kbp) that harbours a full open reading frame (ORF) of the single exon *foxd3* gene and the upstream DNA regions of more than 200kbp. Using this BAC, we captured not only the *foxd3* promoter region but also the associated *cis*-regulatory elements. In this manner, we have exploited a native *foxd3* genomic context to drive the Avi-tagged *foxd3* transgene in a correct spatiotemporal fashion and avoided common problems of plasmid transgenesis, which frequently results in incomplete gene expression patterns or very high, non-physiological expression levels due to multiple integration sites [191]. A generated recombination cassette (*foxd3*-Avi-2A-Citrine) containing 5' and 3' homology arms to the *foxd3* locus within the BAC was used to replace a wild-type *foxd3* gene with the Avi-tagged version. Of note, the Tol2-mediated BAC transgenesis ensures efficient single copy integration into the genome, which is of importance given that critical *foxd3* regulatory roles could be affected by its expression levels. Moreover, to further preserve a natural *foxd3* gene content in the transgenic fish, we used a heterozygous *Gt(foxd3-mcherry)^{ct110aR}* *foxd3* mutant fish line for *foxd3*-Avi BAC together with the Tol2 transposase mRNA injections into single cell embryos. As described previously, *Gt(foxd3-mcherry)^{ct110aR}* fish line carries one functional *foxd3* allele, while the coding region of a second *foxd3* allele is interrupted by the fluorescent mCherry reporter gene resulting in a truncated *foxd3*-mCherry fusion protein with no DNA-binding forkhead domain (Figure 2.2 B). The use of heterozygote embryos for injections avoided issues of the 3rd allele/excess *foxd3* expression during

2.4 Application of the Biotin ChIP-seq method to identify direct *foxd3* gene targets throughout the NC development

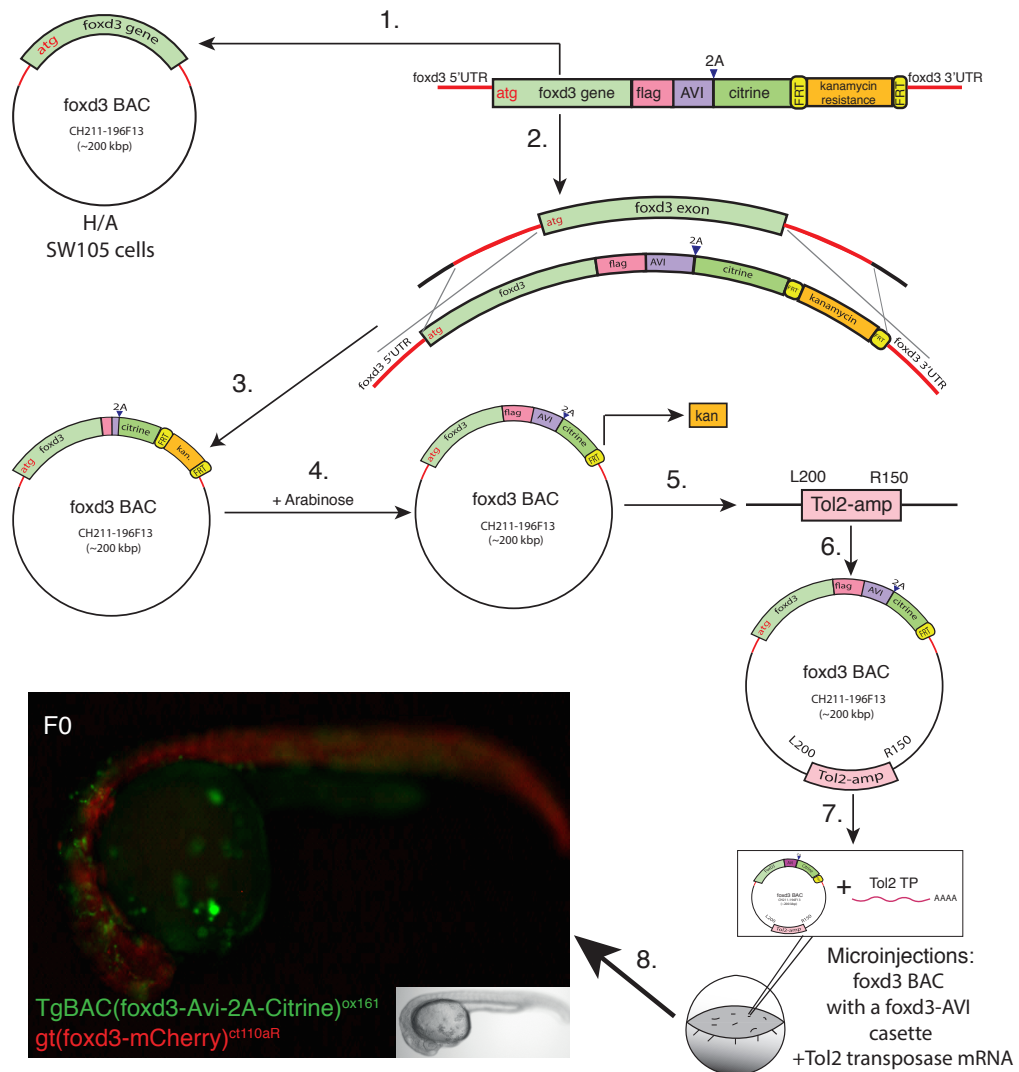


Figure 2.6 Generation of the Foxd3-Avi transgenic fish line. (1) A *foxd3*-flag-avi-2A-citrine-FRT:kan:FRT-polyA construct carrying homologous 5' and 3' overhangs to the *foxd3* CH211-196F13 *Danio rerio* BAC was electroporated into heat-activated SW105 cells that were transformed with this BAC. This bacteria strain harbours a heat-inducible λ Red-encoded recombination system, (2, 3) which was used for the endogenous *foxd3* gene within the BAC to exchange with the engineered construct via DNA recombineering. After positive bacteria selection by kanamycin resistance, (4) kanamycin resistance cassette was flipped out upon arabinose addition, which induced flippase expression in SW105 cells. (5, 6) An iToL2 cassette containing minimal sequences required for the BAC transposition into the zebrafish genome (L200 and R150) together with the ampicillin resistance gene was recombined into the *foxd3*-flag-avi-2A-citrine-polyA BAC. Positive cells were selected for ampicillin resistance. (7) BAC DNA was injected into a blastomere at single-cell stage *Gt(foxd3-mcherry)^{ct110aR}*/wild-type embryos together with the Tol2 transposase RNA. During BAC transposition into the genome, ampicillin-resistance gene was separated from the rest of the BAC and was not incorporated in the genomic sites. (8) Injected embryos exhibited citrine fluorescence in neural crest tissues. They were raised to adulthood (F₀) and were screened for the transgene germ line integrations (F₁).

critical early embryonic stages. Using this strategy, we intended to obtain transgenic animals with only two functional *foxd3* gene copies (selection by a double fluorescence: mCherry (*foxd3*-null), Citrine (*foxd3*-Avi-2A-Citrine) transgenic allele and a wild-type *foxd3* allele) (Figure 2.7 A). Injected embryos (F_0) exhibited Citrine fluorescence in a NC specific manner (Figure 2.6, step 8). Embryos with double Citrine/mCherry fluorescence were selected for germline integration screening and reared to reproductive age. We then have identified a robust *foxd3*-Avi-citrine founder, exhibiting germline integration of the recombineered BAC. Positive F_1 embryos displayed strong Citrine expression, fully overlapping mCherry activity (Figure 2.7 A). In addition, resultant embryos appeared phenotypically normal and healthy indicating that the transgene did not have a negative physiological effect on the embryonic development. F_1 offspring (double mCherry/Citrine positive) were selected and stable *foxd3*-Avi carriers were raised to adulthood (Figure 2.7 A).

2.4.2 Biotin-ChIP confirms direct bimodal action of *foxd3* on the NC gene regulation

The established Avi-effector transgenic zebrafish line, *TgBAC(foxd3-Avi-2A-Citrine)^{ox161}*, expressing Avi-tagged *foxd3* protein in an endogenous fashion (Figure 2.7 A), was crossed to the ubiquitous BirA driver, *Tg(ubiq:NLS-BirA-2A-Cherry)^{ox114}*, expressing the biotin ligase, BirA, targeted to the nucleus (Figure 2.7 B). The resulting progeny of 75% epiboly, 1-2ss, 5-6ss and 14ss was collected for biotin ChIP-seq (Figure 2.8 A), with BirA-only expressing embryos (5-6ss) used as control. Obtained biotin *foxd3* ChIP-seq tracks were clear and distinctive, emphasising the power of this method to generate high signal and low background data sets (Figure 2.8 A). Biotin ChIP-seq was performed on ~120,000 - 200,000 NC cells for each stage of interest and generated libraries were sequenced using NextSeq® 500/550 High Output Kit v2 (75 cycles) on the NextSeq500 sequencing platform. Sequenced reads were trimmed to remove adaptors and unsatisfactory quality reads ($q < 30$), duplicated reads were also removed and input samples were randomly downsampled to the number of corresponding ChIP'ed sample reads. Peaks were called

2.4 Application of the Biotin ChIP-seq method to identify direct *foxd3* gene targets throughout the NC development

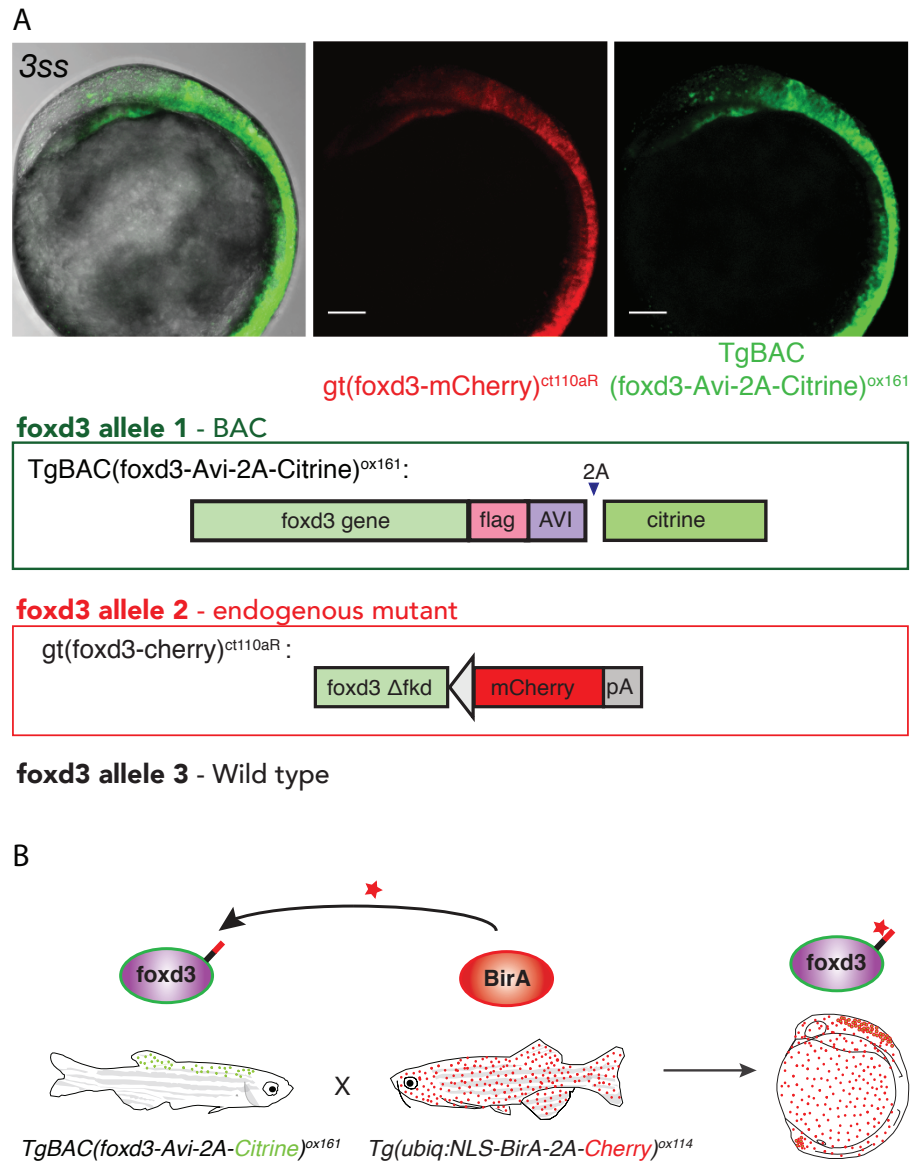


Figure 2.7 *foxd3* biotin ChIP-seq method (A) Lateral view of the transgenic *TgBAC(foxd3-Avi-2A-Citrine)^{ox161}* embryo showing overlap of Citrine and mCherry (*Gt(foxd3-mCherry)^{ct110aR}*) reporters. Scale bars correspond to 100µm. The *TgBAC(foxd3-Avi-2A-Citrine)^{ox161}* line was generated by injecting engineered *foxd3* BAC into heterozygous *foxd3*-mutant background (*Gt(foxd3-mCherry)^{ct110aR}*) in order to re-establish two *foxd3* functional alleles (*foxd3-Avi-2A-Citrine*/WT) in resulting transgenic animals. The gene flip-trap line contains a disrupted *foxd3*-coding allele with the mCherry-cassette resulting in an inactive fusion *foxd3-mCherry* protein with no DNA-binding forkhead box domain. In the *TgBAC(foxd3-Avi-2A-Citrine)^{ox161}* construct design, the citrine coding region was separated from the *foxd3-Avi* gene by a linker viral 2A sequence that mediates a ribosome skipping mechanism leading to production of two separate proteins: a *foxd3-Avi* protein and a Citrine protein from a single transcript. (B) Experimental strategy for *foxd3* protein biotinylation *in vivo*. Transgenic zebrafish expressing *Avi*-tagged *foxd3* or ubiquitous nuclear biotin ligase (NLS-BirA) were crossed to obtain embryos expressing biotinylated *foxd3* for the biotin ChIP-seq method.

2 Bimodal Action of *foxd3* is Required to Specify Neural Crest Cells and Maintain their Multipotency *in vivo*

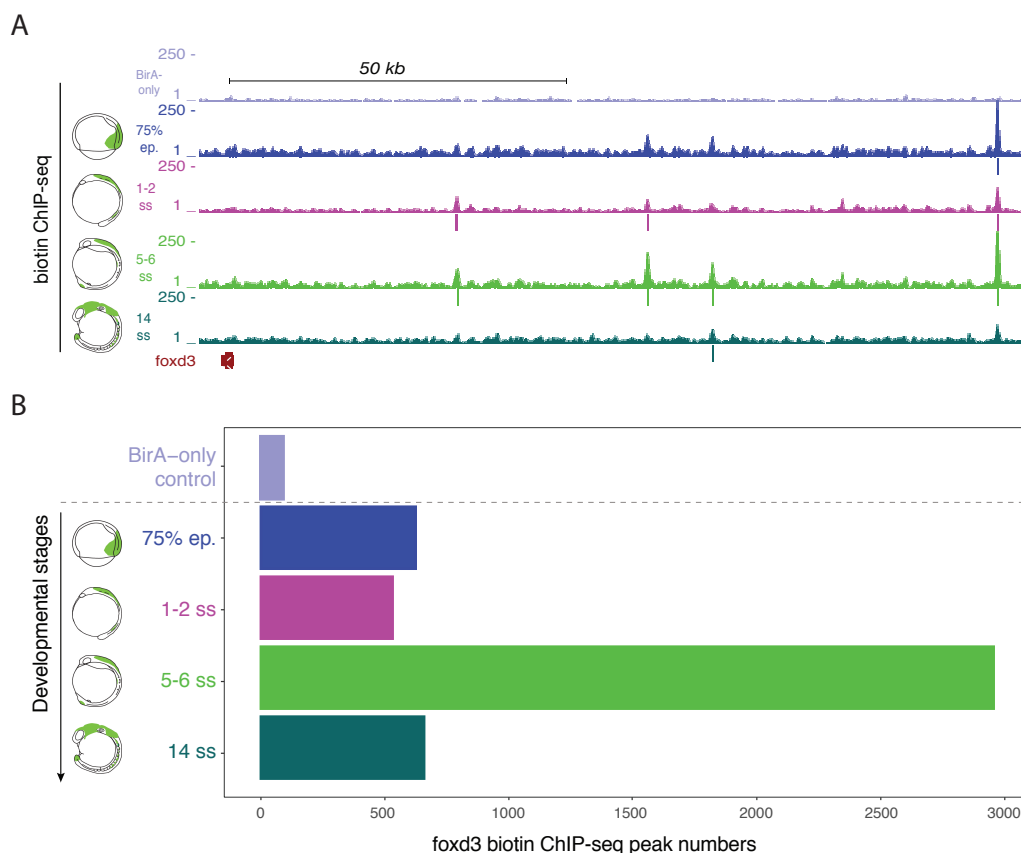


Figure 2.8 *foxd3* biotin ChIP-seq tracks and *foxd3* binding dynamics across different stages of neural crest ontogeny. (A) The genome browser screenshot indicates mapped *foxd3* biotin ChIP-seq data sets at 75% epiboly (in blue), 1-2 somite stage (in pink), 5-6 somite stage (in light green) and 14 somite stage (in dark green) zebrafish embryos within the *foxd3* regulatory locus itself for a putative *foxd3* autoregulation control that will be discussed in a greater detail in Chapter 3. Biotin ligase (BirA)-only ChIP-seq control at 5-6 somite stage is shown in purple (top track), where no Avi-tagged *foxd3* protein was expressed. Positions of called peaks are indicated as vertical lines underneath each *foxd3* biotin ChIP-seq track. (B) Bar plots comparing numbers of *foxd3* biotin ChIP-seq peaks at 75% epiboly, 1-2, 5-6 and 14 somite zebrafish embryo stages. Top bar shows a peak number of a BirA-only ChIP-seq control at 5-6 somite stage.

using Homer findPeaks script with -size 200 -minDist 1500 parameters. This revealed 624 *foxd3*-bound regions at 75% epiboly, 531 peaks at 1-2ss NC, 2955 - at 5-6ss NC and 658 - at 14ss NC, and only 89 non-specific peaks identified in the BirA-only control sample none of which overlapped specific biotin ChIP peaks, thus confirming the specificity of our biotin ChIP-seq method (Figure 2.8 B). Notably, on average we detected five times more of *foxd3*-occupied peaks at 5-6ss when compared to the other two earlier and one later tested NC developmental stages. Given that our transcriptomics analysis suggested that at early NC stages, *foxd3* is activating the NC programme, while at later stages it is

repressing its pre-mature progression, we hypothesise that 5-6ss is a highly dynamic and transitory stage, where *foxd3* exerts its bi-modal function the most, i.e., it still occupies NC initiation programme regions while simultaneously binding to the NC differentiation loci, where it assumes the repressive mode. Each ChIP peak was then annotated to the nearest expressed gene based on our transcriptomics analysis at each corresponding or later developmental stage and subsequently intersected with either down-regulated or upregulated genes that were identified by our *foxd3*-mutant analysis to pin down which genes are activated/repressed by *foxd3* directly. We found that 14.3% of down-regulated genes (total number of significantly downregulated genes in the mutant was 91) at 5-6ss (Figure 2.9 A) and 30.8% of upregulated genes (total number of upregulated genes in the mutant was 724) at 14ss (Figure 2.9 B) were normally directly bound by *foxd3*, suggesting that the identified sets of genes are activated and repressed by *foxd3*, respectively. Statistical over-representation of genes downregulated in the mutant at 5-6ss that were also directly bound by *foxd3* (hence suggesting that *foxd3* binding is crucial for their expression) revealed a significant association with neural crest, stem cell and mesenchymal cell development, NC cell migration, and regulation of neurogenesis (** $p < 0.01$, Figure 2.9 D). Important genes in NC development, such as *pax3a*, *tfap2a*, *nrp2a* and *foxd3* itself, appeared to be positively regulated by the upstream action of *foxd3* at early pre-migratory NC stages. Similarly, expression of transcription factors *id2a* and *gata3*, and signalling cytokines *wnt4a* and *mdkb* (all implicated in NC neuronal lineages), also appeared to be activated by *foxd3* (Figure 2.9 A, C, D). Conversely, by 14ss stage *foxd3*-facilitated gene repression was directed at various genes involved in cell fate commitment, including *foxa*, *olig2*, *tfap2c* and *hey2*, wnt signalling genes (e.g. *wnt3*), and neuronal differentiation factors, such as *slit2/3*, *neurod4*, *gli2b*, *otx2*, and *efna1b* (Figure 2.9 B-D). Notably, 61 out of 223 direct *foxd3* target genes (at 5-6ss & 14ss) were found to be up-regulated in *foxd3* mutants already at 5-6ss (Figure 2.10). This further supports our hypothesis that *foxd3*-mediated activation and repression are transient processes which co-occur at this pre-migratory NC stage.

Overall, our *foxd3* biotin ChIP-seq data in both pre-migratory and migratory NC fur-

2 Bimodal Action of *foxd3* is Required to Specify Neural Crest Cells and Maintain their Multipotency *in vivo*

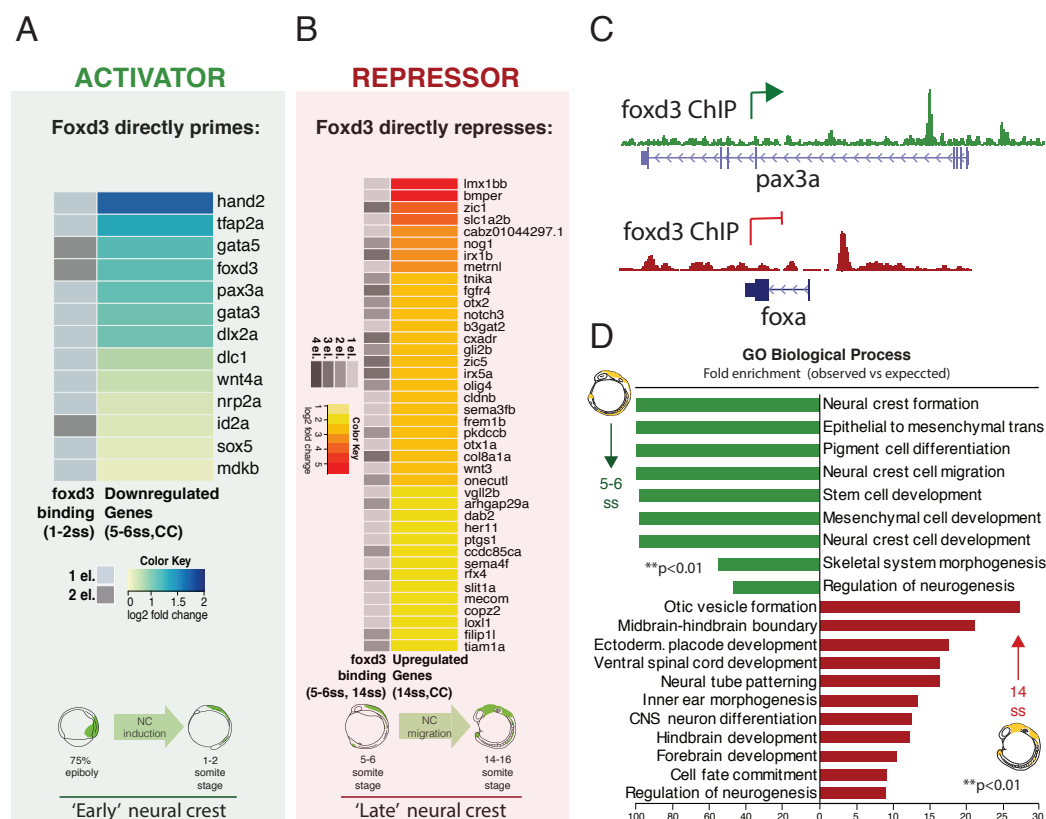


Figure 2.9 Biotin ChIP-seq analysis supports a direct bi-modal *foxd3* regulatory action on NC gene expression. (A) A Heatmap displaying downregulated genes, based on log2 fold change of differential gene expression at 5-6ss in *foxd3* mutant (CC) NC, that were found to be directly occupied by *foxd3* at 1 or 2 associated gene loci at 1-2ss. (B) A Heatmap displaying top 50 most up-regulated genes, based on log2 fold change of differential gene expression at 14ss in *foxd3* mutant (CC) NC, that were found to be directly occupied by *foxd3* at 1, 2, 3 or 4 associated *cis*-regulatory elements at 5-6ss/14ss. Total number of directly repressed genes by *foxd3* was - 223. (C) Example tracks from *foxd3* biotin ChIP-seq experiments showing *foxd3* binding to *pax3a* and *foxa* gene loci, which were found to be downregulated and upregulated in *foxd3* mutant (CC) NC, respectively. (D) Bar plots showing significantly enriched (**p<0.01) gene ontology (GO) terms underlying downregulated genes at 5-6ss in green and upregulated genes at 14ss in red in *foxd3*-mutant embryos that were bound by *foxd3* at 1-2ss and at 5-6ss/14ss, respectively.

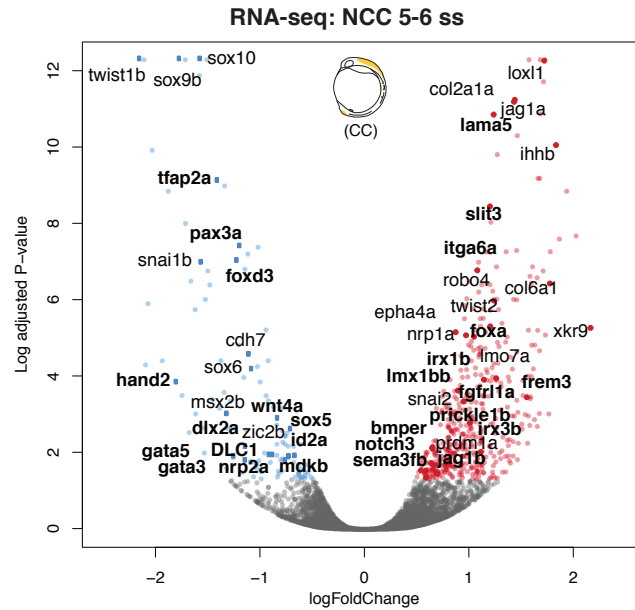


Figure 2.10 Foxd3-mediated repression starts at 5-6 somite stage. Volcano plot highlighting that most NC specifiers are downregulated in *foxd3*-mutant NC at 5-6ss. On the left side (downregulated genes) - genes directly bound by foxd3 at 1-2ss are marked in bold. On the right side (upregulated genes) - genes that are still upregulated at 14ss and are directly bound by foxd3 at 5-6ss and 14ss are marked in bold.

ther confirms a direct activation of a large portion of NC specification genes, followed by direct repression of cell differentiation genes, particularly to prevent premature differentiation into neuronal lineages by a bimodal action of foxd3. It argues for foxd3 being one of the major factors required for the NC programme induction and then specified NC multipotency maintenance.

2.5 *foxd3* affects chromatin accessibility of distal *cis*-regulatory elements

Given such a dramatic downregulation of the NC programme in the absence of foxd3, we speculated that for such a major cellular transition to happen, it was likely that major chromatin remodelling events needed to take place, in order to drive cells from one state to another. We therefore inquired whether foxd3, similarly to other fox proteins, such as FoxA1 and FoxA2, which play chromatin pioneering roles during endoderm specification [192], could also be involved in chromatin remodelling of the associated NC loci. To address this question, we performed Assay for Transposase-Accessible Chromatin

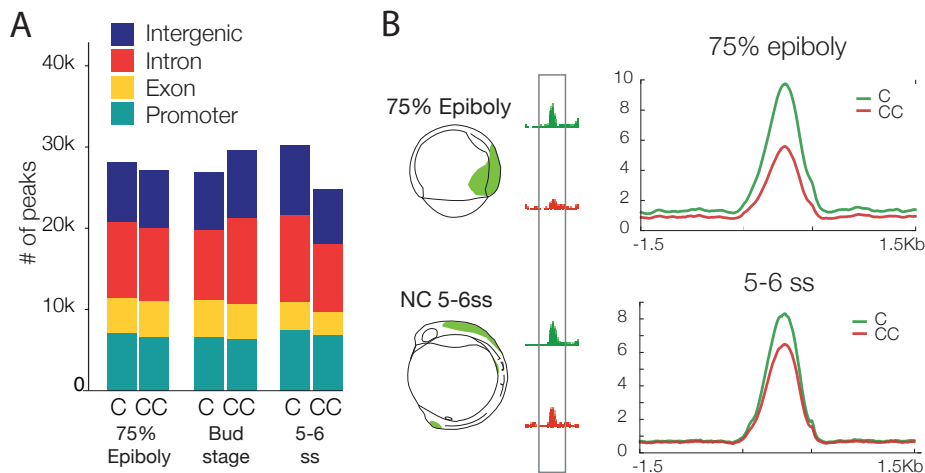


Figure 2.11 Absence of *foxd3* causes reduced accessibility of distal gene elements. (A) Stacked bar plots depicting genomic locations of ATAC-seq peaks at 75% epiboly, bud stage (1-2ss) and 5-6ss. (B) Merged profiles of selected 3,565 elements showed a decreased chromatin accessibility in *foxd3*-mutants in comparison to control samples. Notably, an accessibility defect was more pronounced at 75% epiboly stage than at 5-6ss, suggesting a compensatory mechanism taking place in order to rescue this epigenetic phenotype upon *foxd3* deletion.

in combination with massively parallel sequencing (ATAC-seq [34]) in a tissue-specific manner using the comparative *foxd3*-mutant vs control approach described in Figure 2.3. The exception was made at 1-2ss, where cells were acquired by dissection rather than by FAC-sorting.

Firstly, we analysed the ATAC-seq peaks and determined that majority of the peaks lie outside promoter regions (see Figure 2.11 A). Interestingly, the number of ATAC-seq peaks and their genomic distribution at either distal (intronic and intergenic) or proximal (promoter) was very comparable across the analysed developmental stages. Also, we observed that the *foxd3* depletion did not affect the overall distribution of peaks according to genomic location ($p=0.8743$ and 0.614 for 75% epiboly and 5-6ss, respectively). However, we found that the distal elements, and not promoters, show overall reduced accessibility in the *foxd3*-mutants versus the control samples. This defect in accessibility is more pronounced at early embryonic stages and get somewhat compensated over the NC development (see Figure 2.11 B). To further investigate whether gene expression correlates with the open chromatin states of either promoters or distal enhancer elements, or both, we assigned the ATAC-seq peaks to the nearest expressed genes from our transcrip-

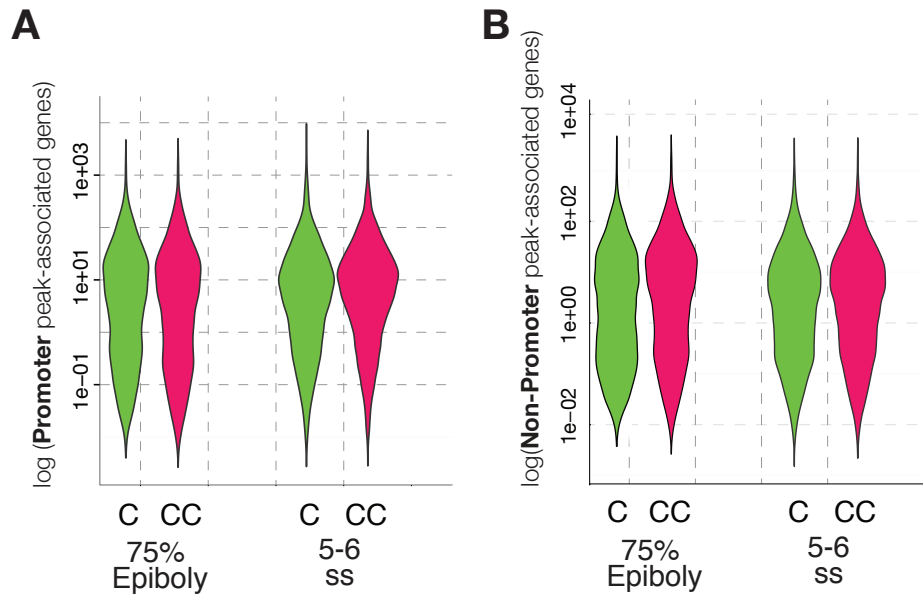


Figure 2.12 Gene expression levels correlated to the uncovered open chromatin elements. Violin plots show correlation between putative promoter and *cis*-regulatory elements with gene expression levels. **(A)** The bimodal promoter peak distribution of gene expression is only associated at 75% epiboly, but not 5-6 somite stage. **(B)** The bimodal enhancer element distribution is associated with both developmental stages.

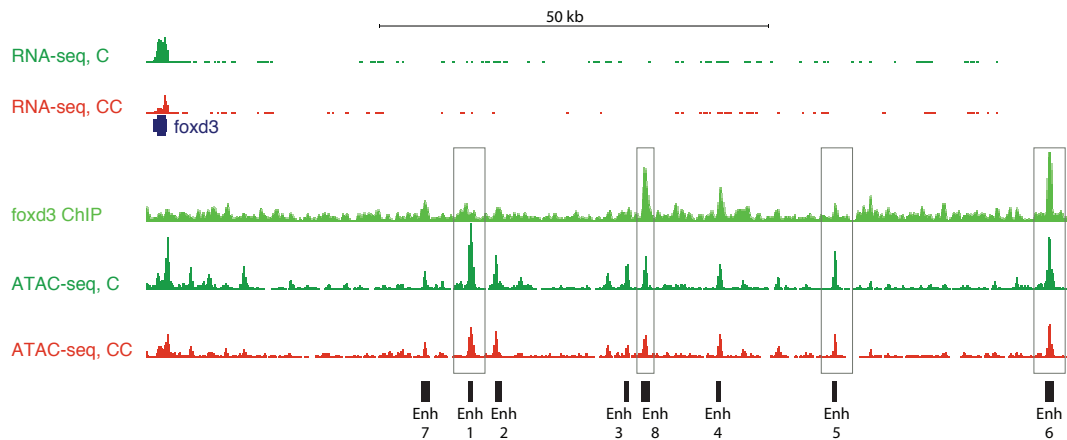


Figure 2.13 Reduced chromatin accessibility leads to a reduced gene expression in the absence of *foxd3*. Genome browser screenshot showing RNA-seq and ATAC-seq profiles at 5-6 somite stage in *foxd3*-mutant (red) and *foxd3*-control cells (green) within the *foxd3* locus itself. The light green track depicts foxd3 binding to this locus at the same developmental stage. Light grey-framed boxes highlight regions of reduced chromatin accessibility in mutant versus control cells that are normally bound by foxd3. Black boxes underneath the tracks indicate predicted foxd3 enhancers that regulate its expression and will be discussed later in this and the following chapters.

tomics results. We observed a bimodal distribution of gene expression levels associated with putative enhancer elements at all analysed stages (see Figure 2.12 A). However,

we only noted the bimodal distribution of gene expression with putative promoters at 75% epiboly stage, but not the later stage (see Figure 2.12 B). This means that, past 75% epiboly stage, enhancers take over the control of the associated gene expression. Most importantly, we wanted to understand whether observed reduced accessibility of such *cis*-regulatory elements in the absence of *foxd3* could explain the reduced NC gene expression levels at 5-6ss that we discovered in the *foxd3*-mutant transcriptome analysis. We saw that genes downregulated in the *foxd3*-mutants, such as *foxd3* itself, harboured distal elements with reduced accessibility in the mutant (CC) when compared to the control (C) cell chromatin landscapes (see Figure 2.13). To do this systematically and genome-wide, we performed *k*-means clustering of acquired ATAC-seq peaks using seqMINER method [193]. This approach identified 8 different clusters and singled out a cluster 3 as the cluster of highest importance (for detailed clustering results please see the published study [113]), as it showed specific enrichment of zebrafish gene expression ontology terms linked to NC and neural plate development (Bonferroni $**p < 0.01$) (see Figure 2.14 A). Importantly, elements from the cluster 3 showed reduced chromatin accessibility in the *foxd3*-mutant NC cells in comparison to the control NC cells (see Figure 2.14 A). Notably, we have cloned and tested a number of identified putative enhancer elements associated with NC genes, such as, *sox10*, *foxd3*, *pax3a* and *ets1* from this cluster and showed their NC-specific activity *in vivo*, further highlighting their putative contribution to spatiotemporal NC gene regulation in developing embryos (Figure 2.15). Interestingly, we found that Cluster 3 elements associated with both, neural crest and neural programmes, suggesting that they control both of those developmental processes (Figure 2.14). However, our mutant analysis suggested that *foxd3* only influences the NC programme during early NC differentiation stages. Therefore, we decided to further subcluster the cluster 3. By doing this, we acquired two subclusters - cluster 3.1 and cluster 3.2 (Figure 2.14 B-E). Strikingly, cluster 3.1 elements associated with NC and neural NC derivative programmes, while cluster 3.2 elements solely associated with the neural differentiation programmes (Bonferroni, $**p < 0.01$) (Figure 2.14 C, E). Evidently, chromatin accessibility of the cluster 3.2 elements was not affected in the *foxd3*-mutants,

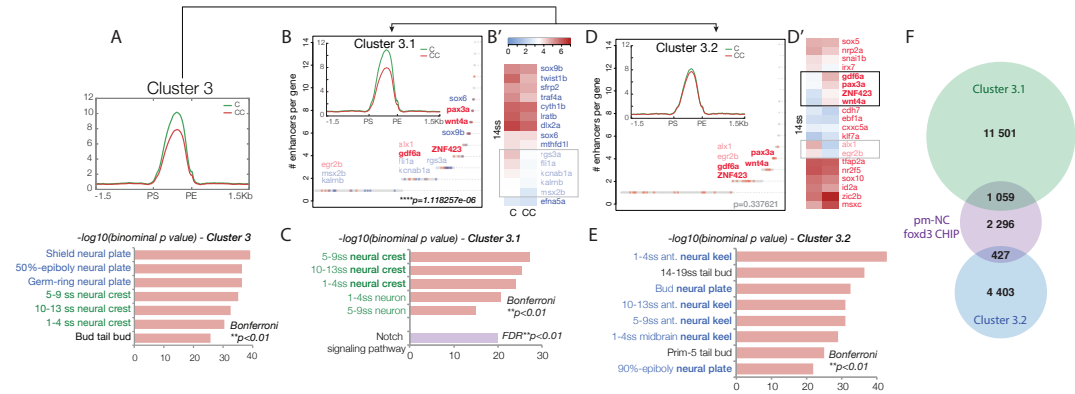


Figure 2.14 *foxd3*-controlled NC gene enhancer cluster 3. (A, top) Mean density map of a merged profile of raw counts for *k*-means cluster 3 featuring elements with differential accessibility and signal levels in *foxd3*-mutant and controls at 5-6ss. (A, bottom) Functional annotation by GREAT associates *k*-cluster 3 with neural crest and neural programme specifications. (B and D) Plots representing genes assigned to *k*-cluster 3.1 (B) and *k*-cluster 3.2 (D) ranked by the number of associated elements. (B' and D') Heatmaps showing later expression (14ss) of NC genes that were depleted in the 5-6ss mutant NC. Genes controlled only by 3.1 elements (in blue) are shown in (B') and those harbouring both 3.1 and 3.2 elements (in red) are depicted in (D'). Genes that remained downregulated at 14ss are labelled in light colour print, and those that instead were overexpressed are shown in bold. (C and E) Functional annotation by GREAT associates *k*-cluster 3.1 with neural crest specification or neuronal differentiation (C) and *k*-cluster 3.2 with neural-only programmes (neural plate/tube development) (E) (Bonferroni; ***p* < 0.01). (F) Venn diagram showing a number of elements from *k*-clusters 3.1 (in green) and 3.2 (in blue) that are directly bound by *foxd3* at premigratory NC (pm-NC) stages (in purple): 75% epiboly, 1-2ss, and 5-6ss ChIP-seq peaks.

while cluster 3.1 elements showed reduced accessibility signatures in the absence of *foxd3* (Figure 2.14 B, D). This suggests that neural *cis*-regulatory modules controlled by the cluster 3.2 elements may not require *foxd3* activity for their initiation.

To further interrogate whether cluster 3.1 elements can explain downregulation of the NC module genes upon *foxd3* removal, we assigned elements from *k*-cluster 3.1 and from *k*-cluster 3.2 to the genes expressed at each corresponding developmental stage and then ranked those genes by the number of associated elements from either cluster (see Figure 2.14 B and D). Remarkably, only cluster 3.1 elements significantly associated with genes that found to be downregulated in our *foxd3*-mutant analysis (*****p*=1.12E-60) (see Figure 2.14 B). Moreover, no other *k*-cluster, including 3.2, showed significant association to genes either up- or downregulated in *foxd3*-mutant at 5-6ss (see Figure 2.14 D). We also observed, that expression of a large number of genes that were downregulated at 5-6ss in mutant (CC) samples when compared to control (C), recovered to normal levels by 14ss,

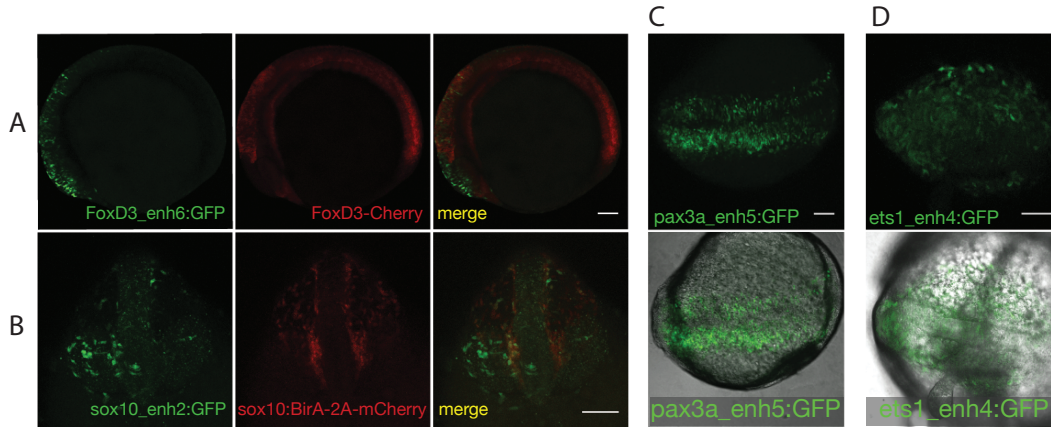


Figure 2.15 *Cis*-regulatory elements from *k*-cluster 3 show NC-specific reporter activity *in vivo*. **(A)** Lateral view of an embryo injected with *foxd3*-enh6-GFP reporter construct into the genetic background of *Gt(foxd3-mCherry)^{ct110aR}*. **(B)** Frontal view of an embryo injected with *sox10*-enh2-GFP reporter construct into the genetic background of *sox10:BirA-2A-mCherry*. **(C and D)** Fluorescent and bright-field overlay images showing dorsal view of embryos injected with *pax3a*-enh5-GFP and *ets1*-enh4-GFP reporter constructs respectively. Scale bars correspond to 100µm.

regardless of cluster 3.1 or 3.2 control (see Figure 2.14 B', D'). Importantly, we found that almost one third of all *foxd3* ChIP peaks at pre-migratory NC stages belonged to cluster 3.1 (28%). However, 11.3% of ChIP peaks also were part of the cluster 3.2. This suggests, that *foxd3* is either still required for the neuronal programme control, but not for the associated chromatin opening, or that *foxd3* is binding and repressing elements from this cluster in a portion of NC cells in order to prevent their differentiation into neuronal derivatives as we have predicted before (Section 2.4.2).

Altogether, our ATAC-seq analysis showed that *foxd3*, both directly and indirectly, affects chromatin accessibility of a number of genetic loci. Importantly, *k*-means clustering method singled out cluster 3.1 as the *bona fide* NC enhancer cluster driving NC specification genes at premigratory NC stages. Defects in the chromatin accessibility of these enhancers resulted in the decrease of NC specifiers' expression in *foxd3* mutants.

2.6 Ectopic Expression of *foxd3* Modifies the Chromatin Landscape in Early Embryos

Our experiments suggested that *foxd3* is able to modify chromatin that subsequently translates into targeted gene activation or repression in developing zebrafish embryos.

2.6 Ectopic Expression of foxd3 Modifies the Chromatin Landscape in Early Embryos

To further strengthen our proposed hypothesis, we tested whether foxd3, when delivered ectopically into foxd3-naïve cells that have never expressed this factor, would be able to induce NC programme via chromatin remodelling mechanisms. To achieve this we overexpressed foxd3 by injecting foxd3-mRNA into single cell-staged zebrafish embryos (see Figure 2.16) and allowed the injected and control embryos develop until 50% epiboly stage, when NC programme is normally not induced yet. Embryos were dissected to only collect ventral halves of embryos, where foxd3 would normally not be expressed (i.e. endogenous foxd3-naïve cells). However, all cells from embryos injected with the foxd3-mRNA translated this mRNA into a functional protein, hence producing ectopically foxd3-primed cells. We then dissociated cells from the dissected embryo halves and performed ATAC-seq to investigate how foxd3 affected chromatin landscapes in comparison to the control samples (dissected ventral halves from non-injected embryos), that did not express foxd3. By comparing the datasets from foxd3-injected embryos to controls,

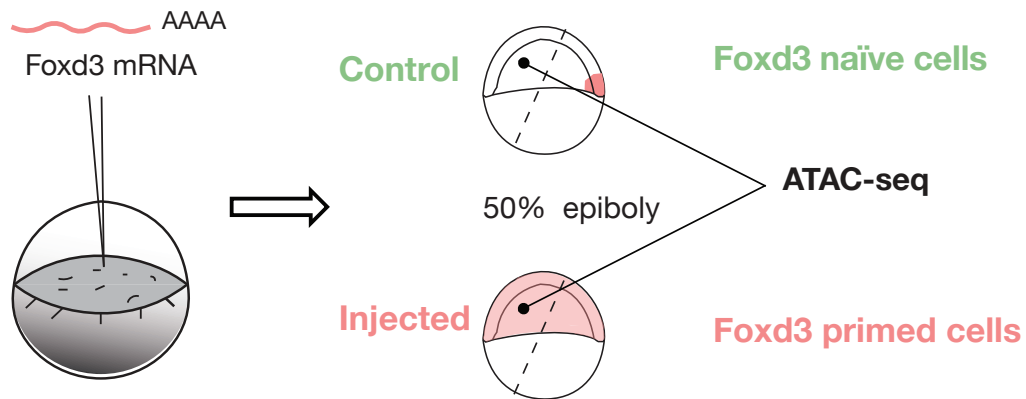


Figure 2.16 Experimental strategy for foxd3 overexpression *in vivo*. Upon foxd3 mRNA injection into single cell stage embryos, embryos were collected for ATAC-seq at 50% epiboly stage. Embryos were dissected (dashed lines) to only collect "foxd3-naïve" cells that do not normally express foxd3. Native and ectopic foxd3 expression is illustrated in dark pink and lighter pink, respectively.

we failed to detect any newly opened genomic regions that were not normally accessible in the foxd3-expressing cells. First, this suggests that foxd3 activity remains specific to a defined set of putative downstream targets and it does not randomly bind DNA when it is in abundance. Second, this also suggests that the NC chromatin landscape is somewhat shaped at this stage and that foxd3 is not able to re-sculpt chromatin from scratch for

further NC developmental stages on its own at 50% epiboly. However, we observed that ectopic *foxd3* was overall capable of re-arranging chromatin landscapes targeting regions normally bound by *foxd3*. When we intersected our injected vs control ATAC-seq signal with all known *foxd3*-binding regions uncovered by our previous *foxd3* biotin ChIP-seq analysis, we observed that ectopic *foxd3* expression increased quantity of cells that now had opened chromatin elements around a big portion of these regions, as compared to controls (several *k*-means clusters showing this signature were merged) (see Figure 2.17 A). Similarly, we uncovered a large number of elements from the previously identified cluster 3.1 containing NC gene enhancers now opened in a larger number of cells in injected embryos in comparison to the control embryos (see Figure 2.17 B). The biological processes that *foxd3* overexpression was priming or enhancing further were associated with early neural plate, neural crest and formation of NC derivatives, and in particular with Wnt and BMP signalling (Figure 2.17 A', B').

Notably, a very illustrative example of how *foxd3* is able to rearrange chromatin bi-directionally, is found within the *foxd3* genomic locus itself. Here, we identified *cis*-regulatory elements showing changes in accessibility in both directions upon *foxd3* overexpression, compacting an element previously open and ectopically opening an element normally accessible in *foxd3*-expressing cells (Figure 2.17 C). This example further highlights a bimodal action of *foxd3* to both activate and repress genes by priming or decommitting different *cis*-regulatory elements.

2.7 Distinct gene regulatory mechanisms underlie the bimodal action of *foxd3*

Interestingly, ectopic *foxd3* expression alone failed to open late NC *cis*-regulatory regions pre-maturely. Therefore, we suspect that *foxd3* most likely requires *in vivo* interacting partners to exert its bimodal activity in terms of chromatin remodelling. To predict which partners *foxd3* might be using during its mediated chromatin opening or compaction, we exploited our *foxd3* biotin ChIP-seq data. We separated *foxd3* ChIP-seq peak sets into 'early' and 'late' NC stages (see Figure 2.1), to look at putatively formed

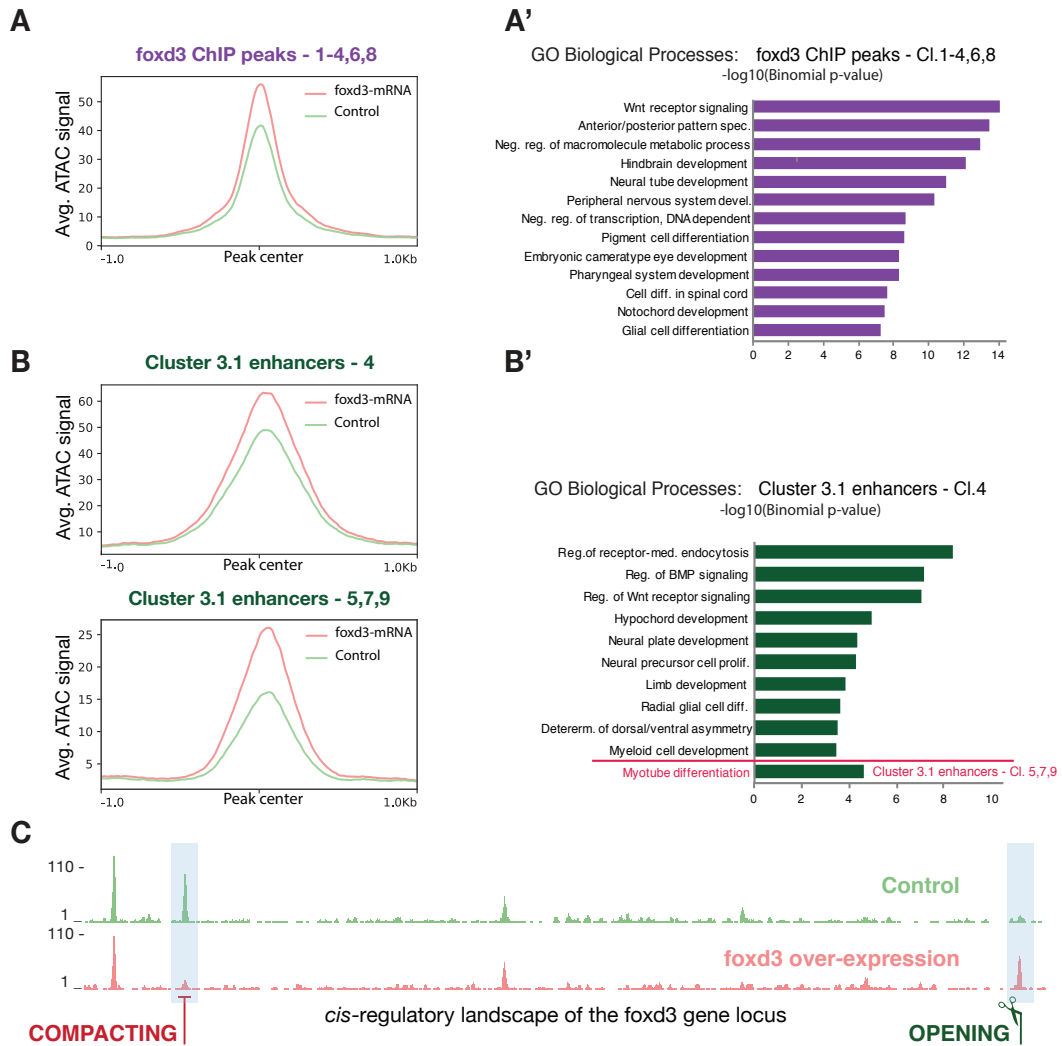
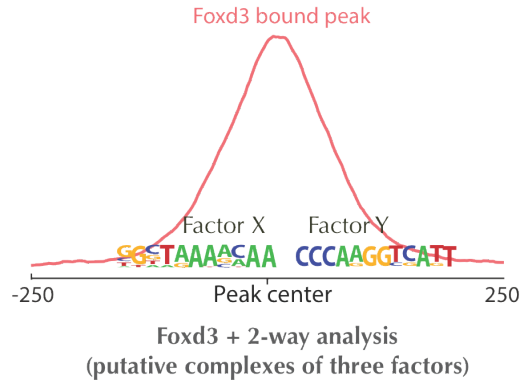


Figure 2.17 Ectopically delivered *foxd3* effect on the NC chromatin remodelling. (**A** and **B**) Mean density maps of merged profiles for *k*-means clusters featuring elements with differential accessibility between the *foxd3* mRNA injected (in pink) and control (in green) 50% epiboly-staged embryos using either *foxd3* binding coordinates or *k*-cluster 3.1 elements as a reference. (**A**) These regulatory sequences were shown to be directly bound by *foxd3* by our biotin ChIP-seq analysis (merged 1-4,6,8 *k*-mean subclusters (out of total eight subclusters) showing similar enhanced accessibility (profiles merged). (**B**) These regulatory sequences belong to the core NC enhancer cluster 3.1 (subcluster 3.1-4 elements and merged comparable signature subclusters 3.1-5, 7, 9 out of ten 3.1 subclusters). (**A' and B'**) Bar plots showing gene ontology (GO) biological processes derived from GREAT analysis that are associated with the regulation of DNA-elements showing increased accessibility upon *foxd3* overexpression. (**B'**) subcluster 3.1-4 elements associated with a single GO term, that is represented below subclusters 3.1-5,7,9 GO terms, under the red line. (**C**) Genome browser screenshot depicting region ~60 kb upstream from the *foxd3* transcription start site. Green and pink ATAC-seq tracks represent genome accessibility from control and *foxd3* overexpressing embryonic cells respectively. Blue boxes indicate either relative loss (red arrow) or acquisition (green arrow) of chromatin accessibility upon *foxd3* overexpression.

Figure 2.18 Strategy for *in silico* analysis to reveal putative *foxd3* complexes during its mediated chromatin opening and compaction. To uncover putative complexes working together with *foxd3*, we exploited our data to look for underlying transcription factor motifs that co-occur together a significant number of times (2-way analysis) under *foxd3*-bound loci.



complexes during *foxd3* activating and repressive modes, respectively. The classification of *foxd3*-bound regulatory elements solely using one of the bimodal actions was based on the assumption that these roles are temporally distinct, with *foxd3* mainly acting as an activator during early stages (75% epiboly, 1-2ss), and as a repressor at later stages of NC ontogeny (5-6ss, 14ss). As discussed previously, we have some evidence that at the 5-6ss *foxd3* is acting as both, activator and repressor. Therefore, the set of 'late' NC peaks was obtained by removing those peaks that were already found at earlier stages. This allowed us to focus separately on the 'early', *foxd3*-activated, and the 'late', *foxd3*-repressed, genomic loci.

Our strategy for such *in silico* analysis is represented in Figure 2.18. Briefly, we performed motif discovery and their characterisation using Homer *de novo* motif search algorithms within given *foxd3* biotin ChIP-seq peaksets from either 'early' or 'late' NC stages. We retained only statistically significant motifs and annotated them manually based on both Homer results and levels of gene expression at a given NC developmental stage to increase the prediction accuracy. We then computed frequencies of all possible two-way combinations of all selected and annotated motifs and selected only those combinations that were enriched at $\chi^2 P < 5 \times 10^{-3}$. This way, we were able to uncover putative complexes of three partners (*foxd3* + 2 other proteins) during the 'activating' and 'repressing' states. After *de novo* TF binding motif enrichment analysis on 'early' and 'late' *foxd3*-bound NC genomic regions, we selected 20 (out of 20) from the early and 27 (out of 32) from the late significantly enriched sequence motif co-occurrences from

a final *in silico* 2-way TF combinatorial analysis. Different motifs annotated as the same TF were pooled into single ones to generate graphical plots (see Figures 2.19 A and 2.20 A). We identified different putative foxd3 co-factors that helped us to understand how it could exert its bimodal regulation on both chromatin remodelling and subsequent gene expression.

We uncovered multiple TFs, such as, Sox3, Zic, Tfap2, Pou6f1, zinc-finger protein, Zbtb3 and Smad4, as likely candidates acting as foxd3 co-activators at early NC stages (see 2.19). Similarly, we were also able to uncover putative foxd3 co-repressors acting during late NC stages. Amongst the identified factors were Otx2, Sox10, E2F6, Rxr, bZIP and Nr2c2ap (see Figure 2.20 A). Notably, some of the uncovered late factors were known

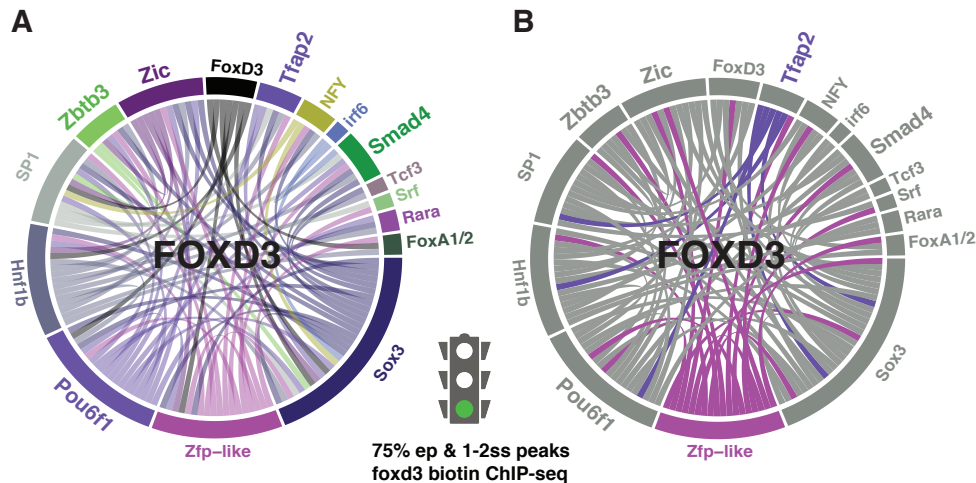
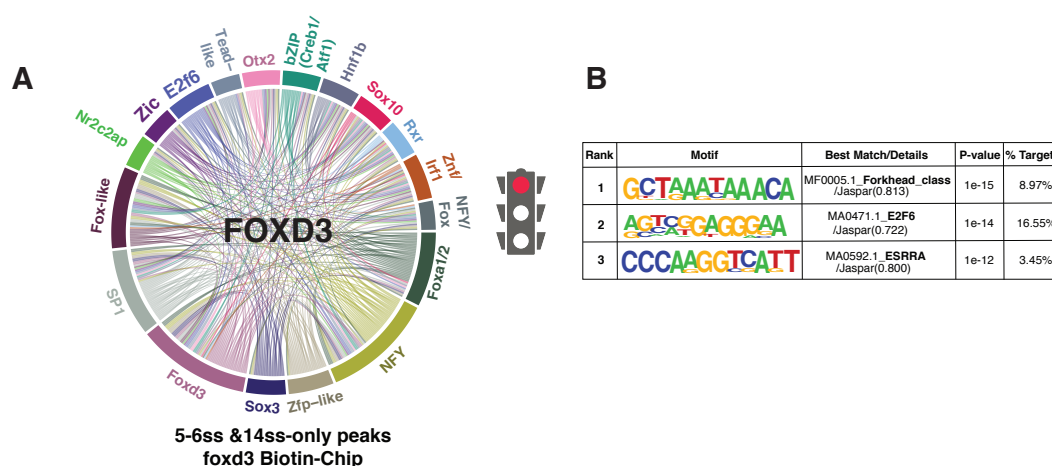


Figure 2.19 Putative foxd3 co-factors acting as activators during early NC development. (A) Spider plot showing statistically significant TF motif co-occurrences on the 'early NC' foxd3-bound activating elements from the 75% epiboly and 1-2 somite stages biotin ChIP-seq data sets. (B) The same spider plot as in (A) highlighting foxd3 interaction with zinc finger proteins (Zfp-like), Tfap2 and their associated partners.

lineage specification factors, hence it is equally likely that some of the found TFs are not working with foxd3 to repress NC differentiation fates, but rather are competing with it for binding to the same loci to drive differentiation instead. Interestingly, our 'early' and 'late' NC analysis generated distinctive fox-family motifs characterising early activation and late repressing roles. We identified foxd3 and other fox motifs at lower frequency during early NC stages (activator) in comparison to late NC stages (repressor) (see Figures 2.19 A and 2.20 A). This is consistent with previous findings showing that

2 Bimodal Action of foxd3 is Required to Specify Neural Crest Cells and Maintain their Multipotency *in vivo*

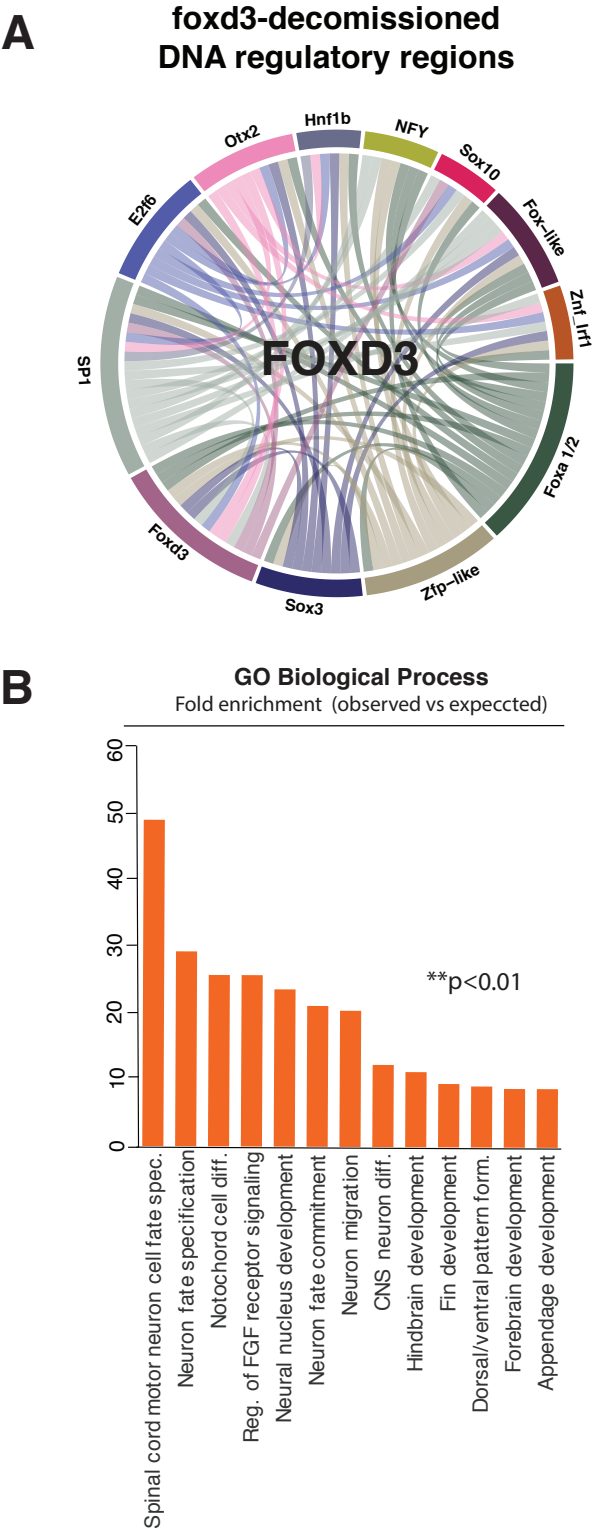


Due to our inconclusive results poorly distinguishing whether identified co-factors from late NC stages, work together with *foxd3* to repress differentiation or compete against *foxd3* to drive corresponding differentiation, we performed additional data mining to uncover the most probable co-repressors of *foxd3*. To this end, we restricted our data sets to scan only those *foxd3*-occupied genomic loci that were associated with genes that were significantly upregulated in the 14ss *foxd3*-mutant NC for *de novo* TF binding motif discovery. Interestingly, only three significantly enriched motifs were identified: *foxd3* (MF0005.1-Forkhead-class/Jaspar(0.813), $p=1e-15$), *e2f6* (MA0471.1-E2F6/Jaspar(0.722), $p=1e-14$) and *Esrra* (MA0592.1-ESRRA/Jaspar(0.800), $p=1e-12$) (Figure 2.20 B). E2F6 was found to co-occupy the highest proportion of *foxd3*-bound

repressive peaks (16.55%). Next, we decided to focus on the mechanism underlying direct chromatin compaction mediated by *foxd3*. To this end, we needed to identify elements that were first opened, then bound by *foxd3* and subsequently closed in developing embryos. Indeed, we identified a set of genomic regions that were accessible during early NC stages (75% epiboly, 1-2ss), then bound by *foxd3* from 5-6ss and ultimately closed by 14-16ss, suggesting a direct *foxd3*-facilitated chromatin compaction. In total, we found 314 such compacted genomic loci, which were assigned to 293 different genes. Importantly, 44% of these genes were found to be up-regulated in *foxd3*-mutants at 14ss, suggesting that these genes are normally regulated and kept repressed by *foxd3* via chromatin compaction mechanisms at their *cis*-regulatory regions. In order to understand which co-factors might be working with *foxd3* to compact these loci, again we scanned these loci to look for enriched TF motifs underlying these 314 genomic regions. We found 19 significant TF motif enrichment using the same initial set of 32 motifs (as were used for late NC stage analysis in Figure 2.20 A) that were subsequently annotated to 12 different TFs (Figure 2.21 A). We again noted frequent fox-family motif discovery, as the top five different fox/*foxd3* motifs were enriched on 85% of all repressed loci. Interestingly, 98.1% of all repressed loci exhibited Hnf1b a/b motif sequence, while other predominantly enriched motifs underlying *foxd3*-mediated co-repression were *otx2* (92.7%), *sox10* (79.9%), two NFY motifs (78.7% and 79.9%) and *e2f6* (70.7%). We have also performed 2-way combinatorial analysis to look which of these identified TF motifs co-occur together under the selected 314 genomic regions. The results revealed that fox motifs had the highest number of statistically significant co-occurrences, suggesting that multiple fox/*foxd3* proteins might be accumulating together to drive chromatin compaction and associated gene repression. Other significant *in silico* co-interactions were *foxd3* co-occurrence with *sp1*, *zfp*-like, *e2f6*, *otx2*, *sox3*, NFY, *znf-irf1*, *hnf1b a/b* and *sox10* motifs (Figure 2.21 A). As mentioned previously, *otx2*, *sox3* and *sox10* factors are known to be involved in NC differentiation and neural development and they may not co-operate with *foxd3* to mediate repression, but rather compete for underlying binding sequences to promote multipotent NC cell differentiation [194–196]. On the other hand,

2 Bimodal Action of *foxd3* is Required to Specify Neural Crest Cells and Maintain their Multipotency *in vivo*

Figure 2.21 Identified putative co-factors needed for *foxd3*-mediated chromatin compaction. (A) Spider plot showing significant 2-way TF co-interactions on the regulatory elements that get directly decommissioned by *foxd3* (were compacted upon a direct *foxd3* binding.) (B) Bar plot showing gene ontology (GO) biological terms significantly enriched (** $p < 0.01$) to the nearest expressed genes of the same DNA regulatory elements from (A) that get directly compacted upon *foxd3* binding by 14 somite stage.



TF e2f6 is a known transcriptional repressor that is a part of Polycomb Repressive protein Complexes 1 and 2 (PRC1/PRC2) [197–199]. We acquired four different known e2f6 motifs and scanned them again under the *foxd3*-bound loci which associated with the genes that were upregulated in the *foxd3*-mutants at 14ss. As a result, we discovered that 82.3% of them were significantly enriched for e2f6 motif (Chi-square test *** $p < 0.0004$). Based on these results and the literature search, we propose that e2f6 is a likely candidate to be one of the *foxd3* co-partners needed to maintain NC multipotency via chromatin compaction and gene repression mechanisms. Additionally, we have performed GREAT analysis to identify which genes and biological processes the compacted 314 loci associated with. This revealed a significant enrichment for neuronal fate specification terms (** $p < 0.01$, Figure 2.21 B), including genes such as *olig2/4*, *mnx1*, *fgf8a*, *gli1*, *neurog1* and *robo1*. This again confirms our previous observations that *foxd3* prevents premature NC differentiation into neuronal cell fates (see Figures 2.4 B and 2.9 D).

Collectively, our *in silico* results revealed that different co-factors might be involved with *foxd3* during its mediated gene activation and repression. Firstly, for the chromatin de-compaction and consecutive gene activation, *foxd3* seems to be co-binding *cis*-regulatory elements together with other co-pioneering factors, such as, *tfap2*. This overall is important to re-arrange chromatin in a permissive fashion for the NC programme induction. On the other hand, during later NC stages, *foxd3* seems to be repressing genes and keeping chromatin in a repressive state around elements controlling NC differentiation into derivatives, such as neuronal precursors. This is achieved by its own promiscuous binding to targeted loci or association with other repressors, such as e2f6.

2.8 Discussion

2.8.1 Bimodal action of *foxd3* on NC gene regulation

Previous studies on FoxD3 have identified this factor as having multiple and diverse roles across different stages of embryo development. The primary functions of this forkhead transcription factor were firstly described to repress transcription and via this mechanism to maintain pluripotency of early epiblast and trophoblast cells [160, 161, 177]. However,

the most recent studies of FoxD3 action underlying gene regulation in developmental biology context, provided new evidence that this factor might be working not only as a transcriptional repressor, but also as a potential activator via chromatin reorganisation mechanisms [166]. Moreover, it was shown to be important not only for pluripotency maintenance, but also for cellular transitions [165, 166]. Similar findings hinting for *foxd3* having comparable roles, such as multipotency maintenance and driving cell fate transitions, were also proposed to take place during NC development [162, 169, 176]. However, the exact mechanisms underlying such processes were not fully understood. Therefore, we aspired to tackle these outstanding questions of how *foxd3* might be controlling induction and differentiation of this remarkable multipotent neural crest cell population in zebrafish embryos.

To explore it, we employed *foxd3*-mutant zebrafish lines to compare NC transcriptomes and epigenomes between the control and mutant embryos. This strikingly revealed that *foxd3* was not acting only as a transcriptional repressor, but also as an activator in terms of gene expression control. Even more remarkably, we identified that the core of NC induction module was downregulated or absent at early stages of NC development, when functional *foxd3* was not expressed. On the other hand, during later NC stages, genes driving multipotent NC transition into migratory stages or neuronal fates, were upregulated, suggesting that normally *foxd3* is repressing these programmes, most likely to keep NC cells in a more naïve state. Our further results confirmed, that *foxd3* is controlling these two distinctive programmes both directly and indirectly, hence reinforcing the feed-forward loops governing, first NC induction, and then its continuation in a multipotent state. Importantly, by studying dynamics of chromatin 'opening' across several stages, we identified a set of NC enhancers (cluster 3.1), a substantial portion of whose accessibility was dependent on a direct *foxd3* binding. We found that their accessibility was reduced in the absence of *foxd3* that in turn leads to downregulation of the core NC genes, explaining how *foxd3* is activating the NC programme via chromatin opening processes (Figure 2.14 B). Similarly, we also found a set of elements, that get directly compacted by *foxd3* and associate with the upregulated genes in the *foxd3*-mutants (Figure 2.21).

Therefore, we concluded that *foxd3* acts as a bimodal chromatin and gene modulator throughout the NC ontogeny. Overall our results suggest that temporal boundaries between the activating and the repressive roles of *foxd3*, are not sharp, but instead, occur concomitantly, with a gradual shift towards the repressive activity after NC specification. Such bimodal activity of *foxd3* first enables early NC fate transitions from more naïve cellular states and then maintains induced NC cells in multipotent phase.

It is now widely appreciated that regulation of gene expression is largely controlled by co-operative interactions between different transcription factors that are dependent on underlying DNA binding motifs [200]. For example, another well-known Fox pioneering factor, *FoxA*, was also shown to both promote gene expression, and also to co-occupy the enhancers of silenced genes like *cdx2* together with transcriptional repressors such as Rfx and type II nuclear hormone receptor [201]. With this in mind, we looked at broad regions of *foxd3*-bound DNA loci at different stages of NC development in order to predict putative *foxd3* partners based on present TF motifs underlying those loci. First surprising result from this study was, that we did not observe an extensive enrichment of fox family motifs from our *foxd3* biotin ChIP-seq sets at early NC stages (when *foxd3* is mostly playing a priming role), whilst fox motifs were the most highly enriched out of all motifs at *foxd3*-bound DNA sequences at later NC stages (when *foxd3* primarily acts as a repressor) (Figures 2.19 and 2.20 A). For all that, this is consistent with previous studies showing that some co-binding pioneer transcription factors exhibit only a partial DNA sequence motif recognition when binding to nucleosomes, but a full recognition when binding to non-chromatinised, naked DNA [88, 89]. This observation further supports our hypothesis that *foxd3* is a bimodal NC regulator, which progressively changes from a co-pioneering factor towards a repressor during the course of the NC ontogeny.

2.8.2 Co-pioneering factors

Our computational combinatorial TF analysis has identified putative *foxd3* co-binding factors required for the pioneering activity (see Figure 2.19). One of the observed factors, *zbtb3*, is a zinc-finger and BTB/POZ domain containing transcription factor, a homolog

of *Drosophila* GAGA factor (GAF) that also can act as both activator and repressor [202]. *Zbtb3* was shown to be critical in early embryonic development and stem cell self-renewal by promoting *Nanog* expression in mice [203]. Interestingly, its fly homolog GAF, was multiple times demonstrated to affect chromatin organisation, including promoting nucleosome removal by associating with chromatin remodelling complexes, such as NURF (nucleosome remodelling factor) or FACT (facilitates chromatin transcription) [204]. Overall, our combinatorial partner analysis revealed that in addition to this zinc-finger proteins, generally ZFPs are likely to be co-involved with *foxd3* during early NC development (Figure 2.19). Intriguingly, BRG1, a catalytic subunit of chromatin remodelling SWI/SNF complex, is known to be attracted to targeted chromatin regions via its N-terminal ZFP-interaction domain [205]. Moreover, as was mentioned before, *FoxD3* was shown to directly interact with *Brg1* to prime enhancers during the ESCs differentiation into EpiCs [166]. Therefore, our *in silico* results offer a potential missing link to a mechanism of *foxd3*/ZFP-based *Brg1* recruitment in a sequence-dependent manner to the associated *cis*-regulatory elements that subsequently leads to nucleosome depletion and enhancer activation that prime cellular transitions. In addition, *smad4*, that was identified as a potential co-pioneering factor by our analysis, was also previously implicated in chromatin remodelling and is associated with *Brg1* interactome [206]. Other known NC factors, such as *zic* and *tfap2*, also seem to be playing a co-pioneering role together with *foxd3* in early NC development. Of note, the co-operative activity of *foxd3* with *tfap2a*, was previously shown to be critical for NC induction, with the two factors acting in concert to fine-tune Wnt and Bmp signalling gradients during gastrulation [115]. Curiously, neural stem cell differentiation factor, *pou6f1*, also appears to be significantly enriched in our combinatorial analysis [207]. Notably, factors, such as *sp1*, *NFY* and *hnf1b* seem not to be activation or repression-specific and rather - general, promoter-like, ubiquitous factors at both early and late NC stages.

2.8.3 Co-repressing partners

We identified a putative novel NC regulator, *e2f6*, that co-represses NC differentiation genes together with *foxd3*. E2F family members are very important players in cell cycle progression with E2F1-E2F3a factors acting as transcriptional activators, and E2F3b-E2F8 factors as transcriptional repressors [208]. Distinctively from other E2F members, E2F6 lacks retinoblastoma (Rb) tumor suppressor binding domain and is ubiquitously expressed throughout all cell cycle stages [199]. It exerts its repressive functionality through polycomb repressive complex recruitment in a DNA sequence-targeted fashion [199, 209, 210]. Notably, a different study that was using similar strategy to ours to uncover TF motif co-occurrences underlying FoxA2 binding sites, which were associated with upregulated genes in FoxA1/2 mutants, also identified E2F6 as a potential co-repressor in mouse liver cells [79]. In spite of *e2f6* motif enrichment underlying direct *foxd3* repression gene targets, we did not observe changes in *e2f6* gene expression levels across different NC stages failing to explain its 'late' NC repressive mode. However, we noticed a continuous increase of *bmi1a* and *bmi1b* expression throughout NC development progression (5.9 and 2.6 fold increase from 75% epiboly to 14ss for *bmi1a* and *bmi1b*, respectively). Given that *e2f6* is found in the Bmi1-containing PRC1 complex, we hypothesise that *bmi1a/b* is also a crucial component of a progressive *e2f6-foxd3*-mediated gene repression protein complex, however this needs to be tested in the future [199].

Disclaimer: The majority of the work presented in this chapter was conducted by me, with the exception of *foxd3*-mutant vs control ATAC-seq and RNA-seq data collection, which was carried out by Dr. Ruth Williams and some of the associated analysis of these data sets, which were carried out by Dr. Daria Gavriouchkina.

2.8.4 Summary

This chapter of my doctoral thesis presents evidence that during NC ontogeny *foxd3* may act as a switch from 'permissive' to 'repressive' chromatin organisation. We also showed that bimodal *foxd3* activity is involved during differential gene expression control by remodelling chromatin of different regulatory elements. Importantly, our results identified distinctive co-transcription factors at different NC stages and thus different transcriptional contexts, linking them to the mechanisms underlying *foxd3* bimodality to either activate or repress genes. Our *in vivo* and *in silico* results suggest that for the pioneering action of *foxd3* required for activation of NC gene expression, other co-pioneering factors, such as *zic*, *tfap2*, zinc-finger proteins are essential. On the other hand, we also singled out a novel NC co-repressor, *e2f6*, that appears to work hand-in-hand with *foxd3* to repress NC differentiation. Interestingly, our *in silico* analysis found multiple different co-occurring fox motifs during during late NC stages, pointing to a possible *foxd3* protein aggregation during its repressive mode. Please see the summary in Figure 2.22.

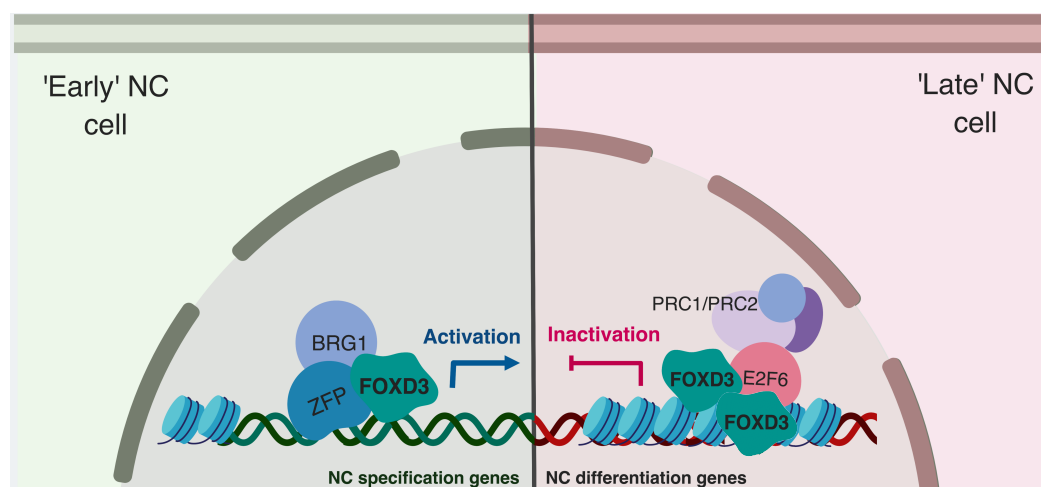


Figure 2.22 Currently proposed mechanism of the bimodal action of *foxd3* during neural crest differentiation. Our current *in vivo* and *in silico* results suggest this model, where during early NC induction, *foxd3* binds together with other co-pioneering factors, or zinc finger proteins to recruit ATP-dependent chromatin remodeller, Brg1, to displace nucleosomes underlying NC *cis*-regulatory elements. This activity promotes NC gene expression and hence its induction and specification. Once NC programme is set, during later NC stages, *foxd3* prevents premature NC differentiation into derivatives, such as neuronal precursors, by co-recruiting transcriptional co-repressors, for example members of the Polycomb repressive complexes. This activity, in turn, compacts chromatin of elements controlling NC differentiation gene expression, keeping late NC cells in the multipotent state, if *foxd3* is still expressed. These proposed result-driven hypotheses will have to be tested and validated in future experiments.

Auto-regulation of the *foxd3* genetic locus during cell lineage specification

This chapter investigates two *cis*-regulatory elements identified upstream of the *foxd3* coding region, suggesting their role in controlling *foxd3* gene expression. Interestingly, our biotin ChIP-seq results revealed that both of these elements were directly bound by *foxd3*, hinting at a possible auto-regulatory loop. Importantly, one of these elements is regulated positively (called Enh6 from now on), while the other one (Enh4) is regulated negatively by *foxd3* in the NC. As previously detailed in Chapter 2, *foxd3* is capable of acting as both an activator and a repressor in terms of chromatin remodelling. Therefore, these two elements provide a paradigm for understanding how *foxd3* achieves its bimodal action by zooming into a single locus and analysing its specific enhancers. To this end, we have generated two transgenic enhancer reporter zebrafish lines by incorporating each of the found putative *foxd3* enhancers upstream of the EGFP reporter. This confirmed that both elements are indeed active *in vivo*, but their spatial activity was critically different. By using imaging, transcriptomics and epigenomics methods, we revealed that cells expressing Enh6 maintain *foxd3* expression and are likely of the neural crest origin, while Enh4 positive cells downregulate *foxd3* expression and are likely specified as neuromesodermal progenitors (NMps).

3.1 Different types of transcription factor auto-regulatory loops for a tight gene expression control

Embryonic development is a well-orchestrated sequence of events that is primarily initiated in the nucleus once the zygotic genome activation takes place [211]. The expression of the master developmental transcription factors that control a myriad of downstream

effector genes, needs to be tightly controlled. Even slight mis-regulations perturbing the dosage of such a factor can cause profound developmental defects [212, 213]. Notably, one of the most efficient and responsive ways to tightly control expression of developmental transcription factors is via their auto-regulatory mechanisms [214]. Transcriptional feed-forward loops are usually multilayered in order to ensure a fail-safe gene expression control. Interestingly, a single TF could also switch from an auto-activation mode to an auto-repression mode depending on its cellular concentration levels and spatiotemporal localisation within a developing embryo [215, 216]. Of note, the auto-repressive and auto-activating loops can be both direct and indirect.

In case of a positive auto-regulation, both direct or indirect, TF maintenance can be achieved in the absence of other transient factors, such as morphogens, that initially onset gene transcription. Such positive feedback loops maintain high TF levels allowing persistent downstream gene regulation required during a given developmental time window to ensure appropriate cell fate decisions [214]. In addition, positive feedback loops can also be employed for a rapid and augmented expression of a given transcription factor [217]. In direct feedback loops, TFs positively regulate themselves by a directly binding to either their own promoters or to associated gene enhancer(s) [218, 219]. Certainly, the feedback loops employing gene auto-regulation via different enhancers can at least partially explain how a single TF can regulate itself differently depending on its spatial and temporal localisation within a developing organism. In indirect positive feedback loops, a TF does not directly bind and activate its own genomic locus, but instead it positively regulates its upstream activators [214]. More complex examples of indirect positive feedback loops can involve double negative feedback loops - an interplay between two repressors. For instance, during *C. elegans* neuronal cell fate decisions, miR-273, that is activated by a *cog-1* TF, represses another TF, *die-1*, that usually activates *lxy-6* microRNA, which in turn negatively regulates *cog-1*. This complex double negative feedback loop ensures that a bipotential unstable neuronal precursor can transition into two different stable neuronal derivative states [220]. Interestingly, mathematical models for several different biological systems predict that positive feedback loops amplify the

noise and destabilise activation systems in a non-linear fashion, thus allowing them to synchronise gene expression and acquire multistationarity (two or more steady states) that is essential for differentiation [221–223].

On the other hand, negative feedback loops are known to be involved in homeostasis maintenance ensuring a stability within a given system [223]. For example, BRCA1 is an important tumour suppressor, which is critical for DNA repair upon its damage. Under normal conditions, BRCA1 is auto-repressing transcription from its own promoter, together with E2F1 and Rb transcriptional regulators, thus keeping its expression at basal levels. However, upon genotoxic stress, BRCA1 is displaced from its promoter allowing a boost of *BRCA1* gene transcription in order to repair any consequent DNA damage [224]. Of note, oppositely to positive feed-forward loops, negative auto-regulation decreases stochasticity in gene expression, which results in a fast and energy efficient gain of stability in a given system, in comparison to positively or non-auto-regulated systems [225–227]. Additionally, negative auto-regulatory feedback can not only allow a steady state maintenance, but also a termination of self-regulated gene expression altogether, if its binding is persistent [214]. Not surprisingly, these rapid ON and OFF states are important and common during embryo patterning, such as seen during the segmentation of somites, where Hes gene expression oscillations are controlled via self-regulatory loops [228, 229]. Another example is that of the retina patterning, where negative auto-regulatory feedback by a homeodomain TF, Defective proventriculus, Dve, achieved via auto-repression of its enhancer, results in specification of distinct photoreceptors (PRs) [230]. A *Drosophila* compound eye retina consists of around 800 ommatidia containing eight PRs. The outer PRs are adapted to detect motion, while the inner PRs express colour-sensitive Rhodopsin (Rh) to detect colour. Their differential specification is at least partially dependent on Dve TF expression levels: outer PRs have high levels of Dve where it represses expression of the colour-sensitive Rh, while inner PRs have none to low levels of Dve depending on a PR subtype allowing colour-sensitive Rh expression. A Dve enhancer, yR7, is only activated in inner PRs, where Dve is auto-repressing it, hence minimising or abolishing its own expression in inner PRs [230].

By and large, auto-regulatory loops have been shown to play important roles during embryonic development. However, most of the feed-forward circuit studies thus far have been conducted in bacteria and yeast models, which have a rather simple genomic regulation [231]. On the other hand, multi-cellular organisms have much more complex genomes, containing promoters and *cis*-regulatory regions (such, as enhancers, silencers, insulators) [232]. Naturally, to start understanding more complex regulatory interplay, the first studies were aimed at promoter-templated regulatory loops. Nonetheless, during development, it is not promoters, but *cis*-regulatory regions that control gene expression in a spatiotemporal manner, which drives differentiation and embryonic patterning [233]. Therefore, to gain deeper insight into how developmental master TFs control development and are controlled themselves, we will need to decipher their multi-layered auto-regulation revolving around such distal regulatory DNA elements.

3.2 The genetic locus of *foxd3* reveals its putatively auto-regulated *cis*-elements

To understand the upstream regulation of the *foxd3* gene, we analysed in detail our *foxd3* biotin ChIP-seq and ATAC-seq data at the *foxd3* genomic locus across different stages of zebrafish embryonic development described in the Chapter 2. In doing so, we identified 8 putative *foxd3* enhancers based on accessible chromatin tracks from the ATAC-seq results (Figure 3.1 B). Notably, 4 identified putative *cis*-regulatory elements belonged to Cluster 3.1 and 3 elements - to Cluster 3.2, which were described in the Chapter 2 (Figures 2.14 and 3.1 A). We found that putative enhancers - enh7, enh8, enh4, enh5 and enh6 were bound by *foxd3* itself, suggesting its auto-regulatory feedback via these elements, which is common feature of important developmental TFs (Figure 3.1 A) [214]. We have cloned all 8 elements upstream an E1b minimal promoter driving EGFP reporter using a construct adapted for the Ac/Ds maize transposition that was re-purposed for enhancer screening and zebrafish transgenesis in our lab [234]. We then injected these enhancer-reported constructs into zebrafish zygotes and screened F₀ embryos for fluorescence. We were able to show that Enh4 (618bp length), Enh6 (884bp) and Enh8 (913bp) elements

3.2 The genetic locus of *foxd3* reveals its putatively auto-regulated *cis*-elements

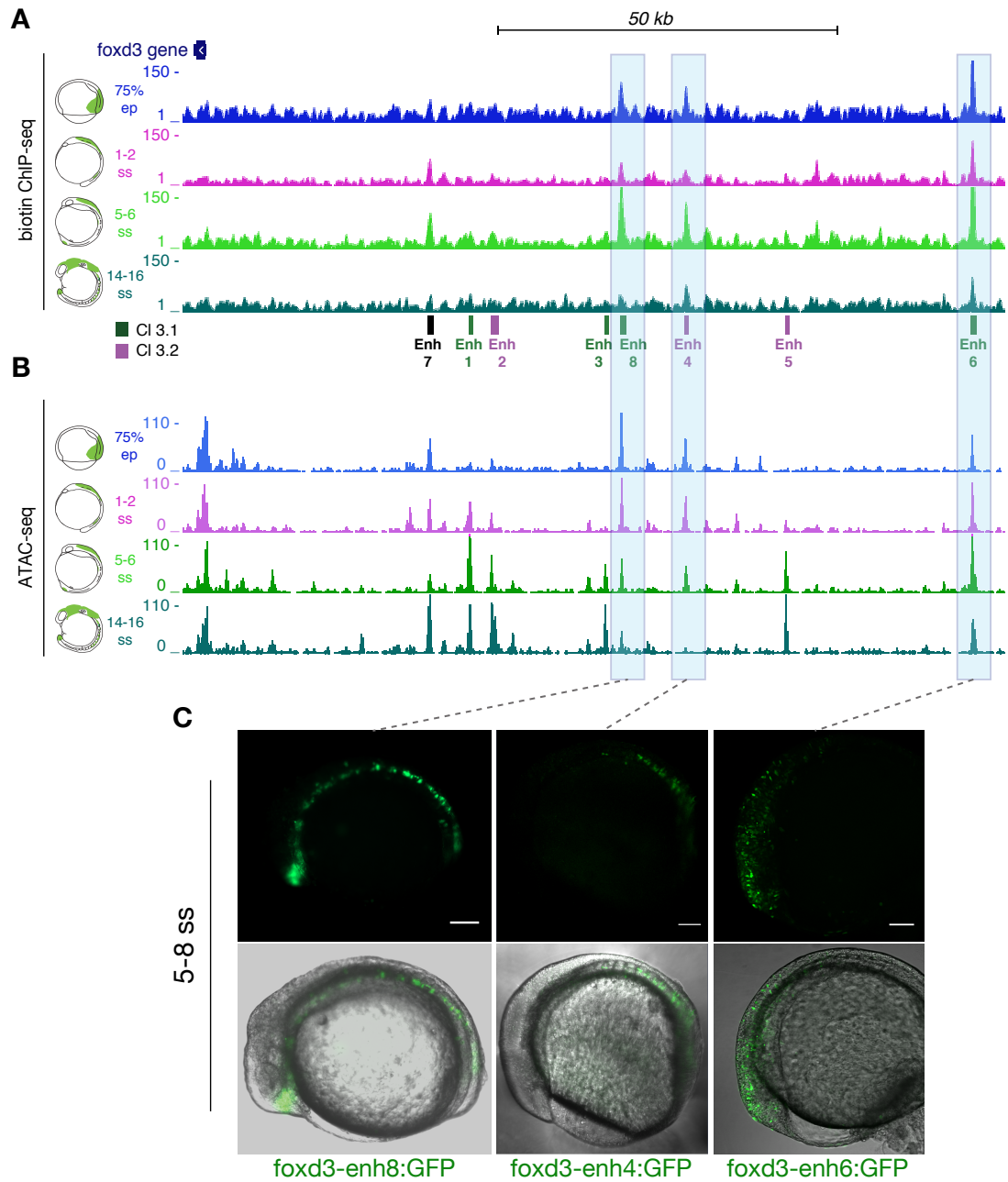


Figure 3.1 *Foxd3* gene locus and active auto-regulated *cis*-regulatory elements. (A) Foxd3 Biotin ChIP-seq at 75% epiboly (blue), 1-2 somite (purple), 5-6 somite (light green) and 14 somite (dark green) stages. Green vertical boxes highlight Cluster 3.1 elements and purple vertical boxes - Cluster 3.2 elements. The top scale bar corresponds to 50 kilobase (kb) distance. (B) ATAC-seq from *foxd3*+ cells at 75% epiboly (blue), 1-2 somite (purple), 5-6 somite (light green) and 14-16 somite (dark green) stages. (C) Lateral fluorescence microscopy images and overlaid brightfield images of *F₀* embryos (positioned in the anterior-to-posterior direction from left to right) at 5-8 somite stages that were injected with Ac/Ds enhancer-GFP reporter plasmids carrying Enh8, Enh4 and Enh6 elements. Scale bars correspond to 100 μ m.

were active *in vivo* (see Figure 3.1 C). In particular, we were intrigued by Enh6 and Enh4 due to their axial expression differences: Enh6-GFP reporter was expressed in cranial regions, while Enh4 reported activity in trunk regions (Figure 3.1 C). We speculated that these enhancers could be controlling *foxd3* expression in the distinctive cranial and trunk NC subpopulations, as was previously found to be the case in chicken embryos, where NC1 enhancer element was shown to control *FoxD3* expression at the cranial and NC2 element at the vagal and trunk levels [235].

Importantly, we observed that, while Enh6 element is bound by *foxd3* and its chromatin accessibility persists, Enh4 element is also bound by *foxd3*, but its accessibility is abolished by 14-16 somite stage (ss) (see Figure 3.1 A-B). Also, as described in the previous chapter, open chromatin conformation of Cluster 3.1 elements, such as Enh6, was shown to be *foxd3*-binding dependent, while Cluster 3.2 elements, such as Enh4, did not require *foxd3* for their opening. Therefore, this would argue that Enh6, that is found in Cluster 3.1, needs *foxd3* for its activation or chromatin accessibility maintenance. On the other hand, Enh4 seems not to be activated, but rather repressed upon *foxd3* binding. Next,

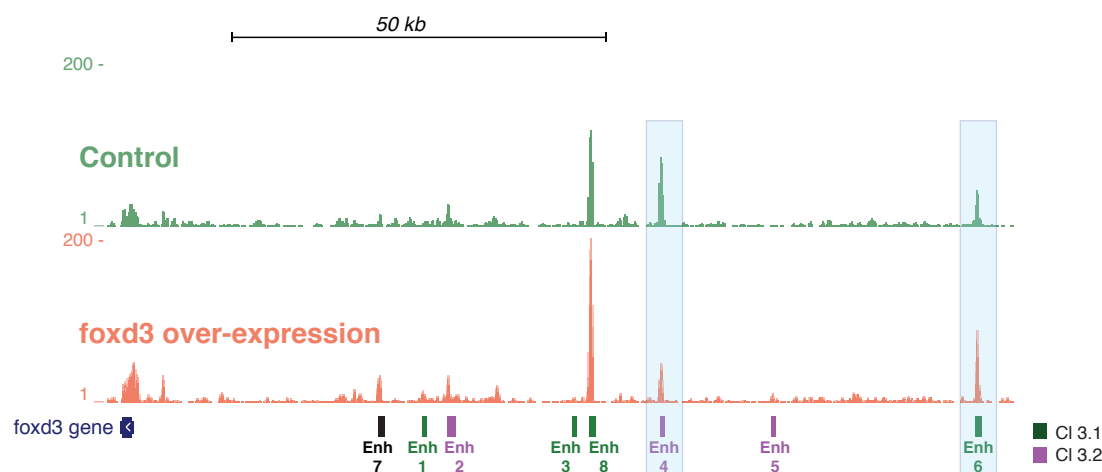


Figure 3.2 Ectopically delivered *foxd3* reduces Enh4 and increases Enh6 chromatin accessibility. Genome browser screenshot depicting region upstream of the *foxd3* transcription start site, containing putative *foxd3* *cis*-regulatory elements. Green and light orange ATAC-seq tracks represent genome accessibility from control and *foxd3* overexpressing embryonic cells from 50% stage embryos, respectively. Also see Figures 2.16 and 2.17 for more details.

we have confirmed that the observed Enh4 enhancer region was compacted upon *foxd3* over-expression in 50% epiboly embryos, which were dissected to perform ATAC-seq on

3.3 Comparison of transcription factor binding sites between Enh6 and Enh4 reveals underlying differential regulation

ectopically *foxd3*-primed cells, as was described in Section 2.6. In short, upon ectopic *foxd3* delivery via *foxd3* mRNA injection into zebrafish zygotes, we observed a reduced Enh4 and increased Enh6 chromatin accessibility in experimental embryos in comparison to control embryos (see Figure 3.2).

Altogether, we identified one putative *foxd3* enhancer, Enh6, 114kb upstream from the *foxd3* TSS, that is active in cranial embryo regions at 5-6ss and is positively auto-regulated by *foxd3* in terms of permissive chromatin confirmation. On the other hand, we have also uncovered another *cis*-regulatory element, Enh4, 70kb upstream of the *foxd3* TSS that is active in the embryo trunk at 5-6ss. Importantly, chromatin region underlying this element seems to be compacted upon *foxd3* binding, which could indicate a negative auto-regulatory *foxd3* loop.

3.3 Comparison of transcription factor binding sites between Enh6 and Enh4 reveals underlying differential regulation

First, to understand the regulatory codes underlying these two different *cis*-regulatory elements, one activated and the other one possibly repressed by *foxd3* in cranial zebrafish regions at 5-6ss, we explored which transcription factor binding sites (TFBS) were significantly enriched within each element. To do so, we have employed two largest databases containing predicted and validated TF Position Weight Matrices (PWMs), JASPAR and HOCOMOCO, to identify putative TFs binding to and regulating either Enh6 or Enh4 [236, 237]. To highlight the most highly enriched as well as differentially present TFBSs underlying Enh6 and Enh4 DNA sequences, we have prioritised and plotted those uncovered TFBSs that were found: (a) at least 3 times within either Enh6 or Enh4 and (b) each uncovered differentially present TF motif had at least two or more instances in the Enh6 sequence as compared to Enh4 or vice versa. Using JASPAR PWMs, we have found a number of putative TFBSs; the most abundant ones for the Enh6 were GATA1/2, TFAP2A, ETS1, PDX1, various SOX and FOX motifs, while for the Enh4 - ZEB1, LHX6/8/9, MEIS1, YY1, SNAI2, RAX2, ETS1 (Figure 3.3). Similarly, using HOCOMOCO PWMs, we uncovered MBD2, SP1/2, various FOX, SOX, AP2, RFX motifs

3 Auto-regulation of the *foxd3* genetic locus during cell lineage specification

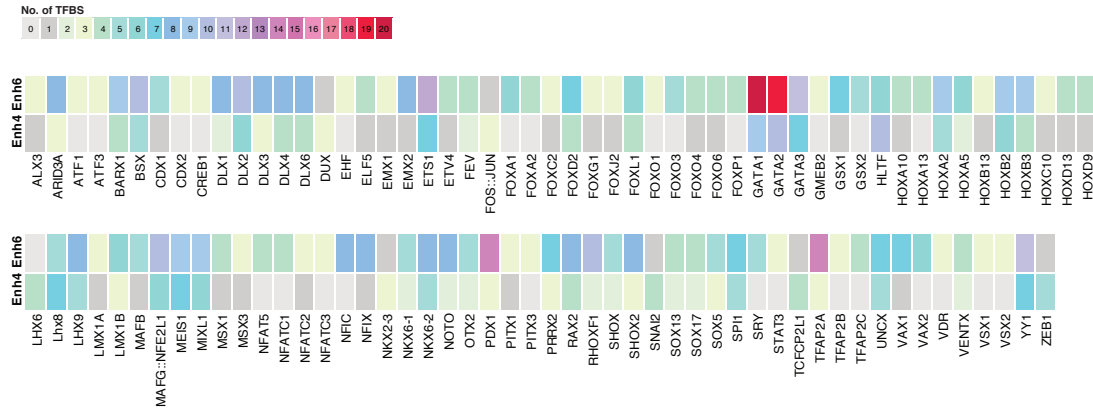


Figure 3.3 Transcription factor binding site (TFBS) analysis underlying Enh6 and Enh4 DNA sequences using JASPAR database. Predicted TFs were listed alphabetically and a number of each predicted TFBS (from 0 to 20) was illustrated using different colours.

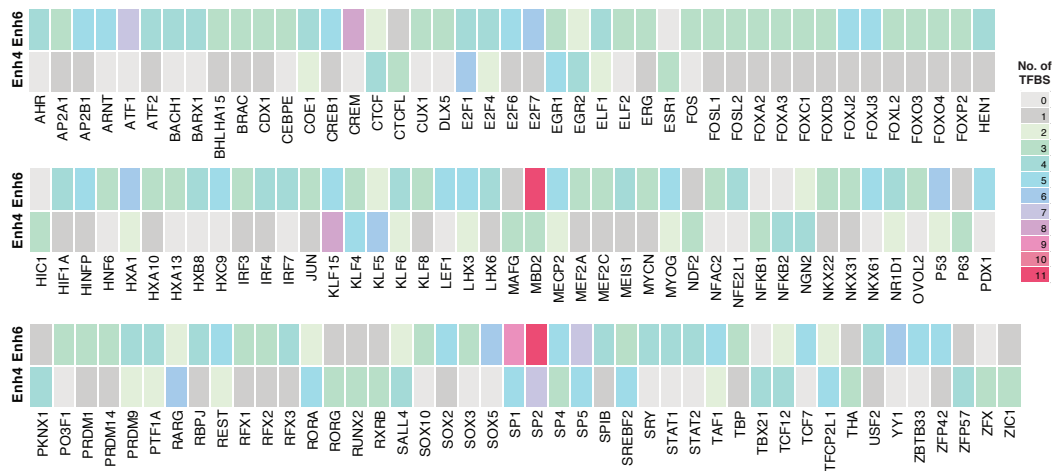


Figure 3.4 Transcription factor binding site (TFBS) analysis underlying Enh6 and Enh4 DNA sequences using HOCOMOCO database. Predicted TFs were listed alphabetically and a number of each predicted TFBS (from 0 to 11) was illustrated using different colours.

within Enh6 sequences, and ZIC1, various KLF, TCF, TBX21, SALL4, RORA/G motifs underlying Enh4 sequences (Figure 3.4). Interestingly, both elements were enriched for different TFs involved in neurogenesis (e.g., LHX, ZIC, TCF, RORA/G, OTX2). However, Enh6 sequence exhibited more NC-like TF binding signatures (i.e., Sox10, Tfp2, Fox, Ets1) in comparison to the ones underlying the Enh4 sequence, suggesting that Enh4+ cells might not be of NC origin.

3.4 Stable transgenic Enh6-GFP/Enh4-GFP zebrafish line imaging confirms different axial activity patterns

The results of our TFBS analysis advocated that Enh6 might be following the NC GRN regulation logic and hence would likely be active in NC cells, whereas the trunk element Enh4, that might be repressed in the NC, would remain active in different cell types. Therefore, to test this hypothesis, we established stable transgenic lines expressing Enh6-GFP or Enh4-GFP reporters to investigate their adequate activity rather than focusing on transient and mosaic somatic expression (as was shown in Figure 3.1 C). To this end, we selected positive F₁ embryos carrying germ-line transgene integrations and raised them to adulthood for a stable line establishment. Enh6-GFP line was annotated as *Tg(FoxD3:enh6-GFP)^{ox109}* and Enh4-GFP - as *Tg(FoxD3:enh4-GFP)^{ox110}*.

We imaged both stable transgenic lines and confirmed that Enh6 was active in cranial regions (see Figures 3.5 and 3.6), while Enh4 was active more caudally (see Figures 3.7 and 3.8) in a developing zebrafish embryo at 5-8 somite stages. To investigate the putative *foxd3* enhancer relationships to *foxd3* gene expression, we have crossed either *Tg(FoxD3:enh6-GFP)^{ox109}* or *Tg(FoxD3:enh4-GFP)^{ox110}* enhancer reporter line to *Gt(foxd3-mcherry)^{ct110aR}* *foxd3* expression reporter line. Of note, *foxd3* is known to be expressed not only in the NC, but also mesoderm (e.g. somites), tailbud, also in a small portion of hindbrain, spinal cord, cranial ganglia cells at later stages [185, 186].

As a result, we have observed that cells displaying Enh6-reporter activity largely overlapped cranial *foxd3*+ NC regions, except for the most anterior part, where only cells with Enh6 activity could be detected at 5-6ss (Figure 3.5). By analysing sections of the neural tube regions at a higher resolution we observed a positive correlation between *foxd3* and Enh6-driven EGFP expression at the same developmental stage (Pearson's $r = 0.47$, $SD = \pm 0.06$; Figure 3.22). Conversely, Enh4 reporter construct was not active in the cranial regions of zebrafish embryos, suggesting that *foxd3* occupancy of this element, as detected in our biotin ChIP experiments, might be repressing it at the cranial level at 5-6ss (Figure 3.7). However, Enh4-GFP reporter was broadly expressed in the more posterior regions surrounding the tailbud, as well as in the presomitic mesoderm (PSM)

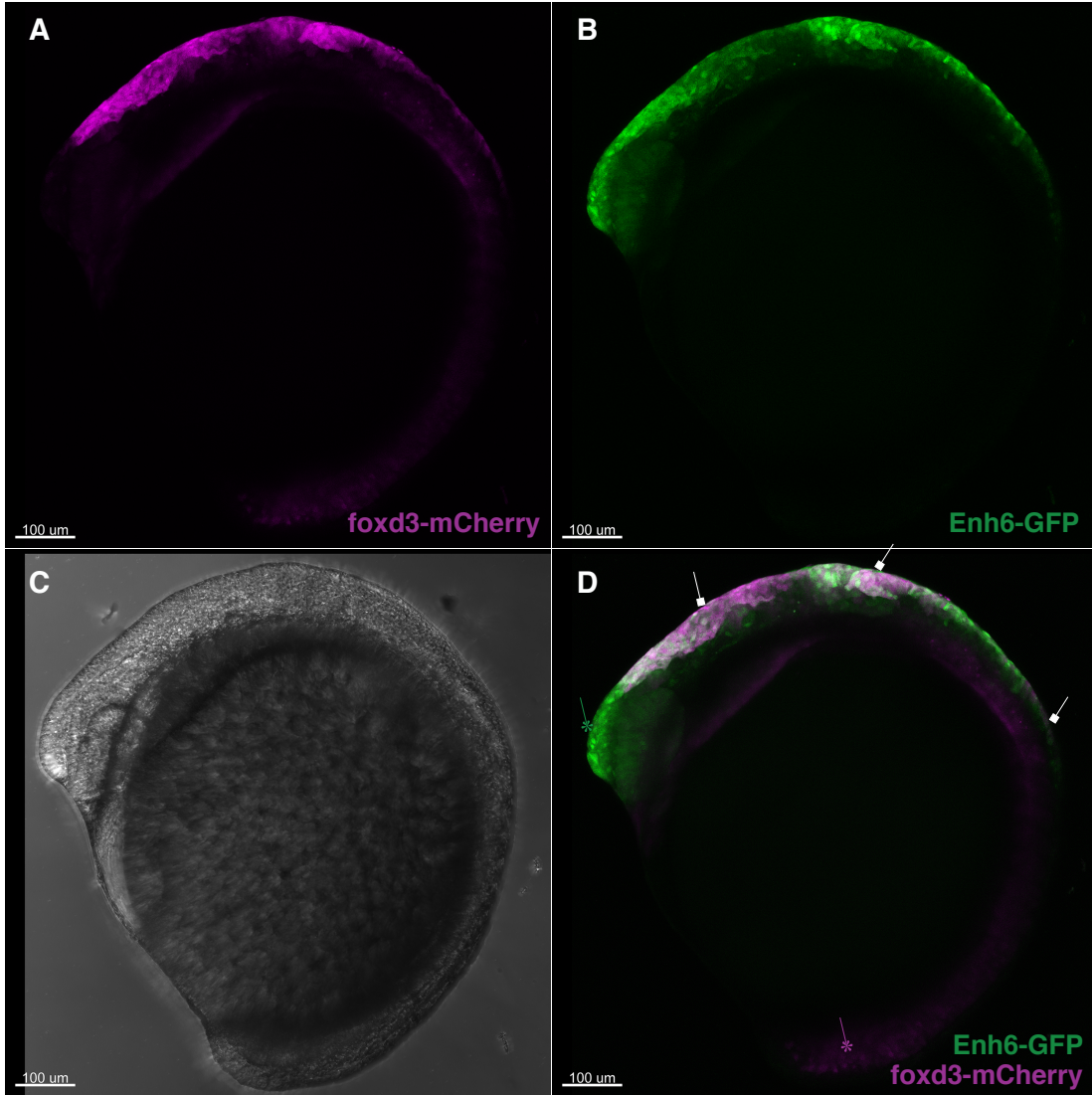


Figure 3.5 *Tg(FoxD3:enh6-GFP)^{ox109}* line crossed to *Gt(foxd3-mcherry)^{ct110aR}* line: lateral view of the whole-mount live embryo at 5 somite stage. (A) foxd3 expression as seen from *Gt(foxd3-mcherry)^{ct110aR}* reporter line fluorescence (coloured in magenta). (B) Enh6 activity as seen from *Tg(FoxD3:enh6-GFP)^{ox109}* line fluorescence. (C) Brightfield image of the imaged embryo. (D) Image overlay of (A) and (B) images. White arrows highlight overlapping fluorescence; green arrow - *Tg(FoxD3:enh6-GFP)^{ox109}* enhancer reporter line-only fluorescence; magenta arrow - *Gt(foxd3-mcherry)^{ct110aR}* reporter line-only fluorescence. Scale bars correspond to 100μm.

and the developing somites (Figures 3.7 and 3.8). When trunk neural tube regions were analysed at a higher resolution, we observed that foxd3-mCherry expressing cells overlapped Enh4-EGFP expressing cells to a significantly smaller extent than Enh6-EGFP expressing cells (Pearson's $r = 0.24$, $SD = \pm 0.18$, $p = 0.0001$ - Welch's t -test; Figures 3.9, 3.24).

3.4 Stable transgenic Enh6-GFP/Enh4-GFP zebrafish line imaging confirms different axial activity patterns

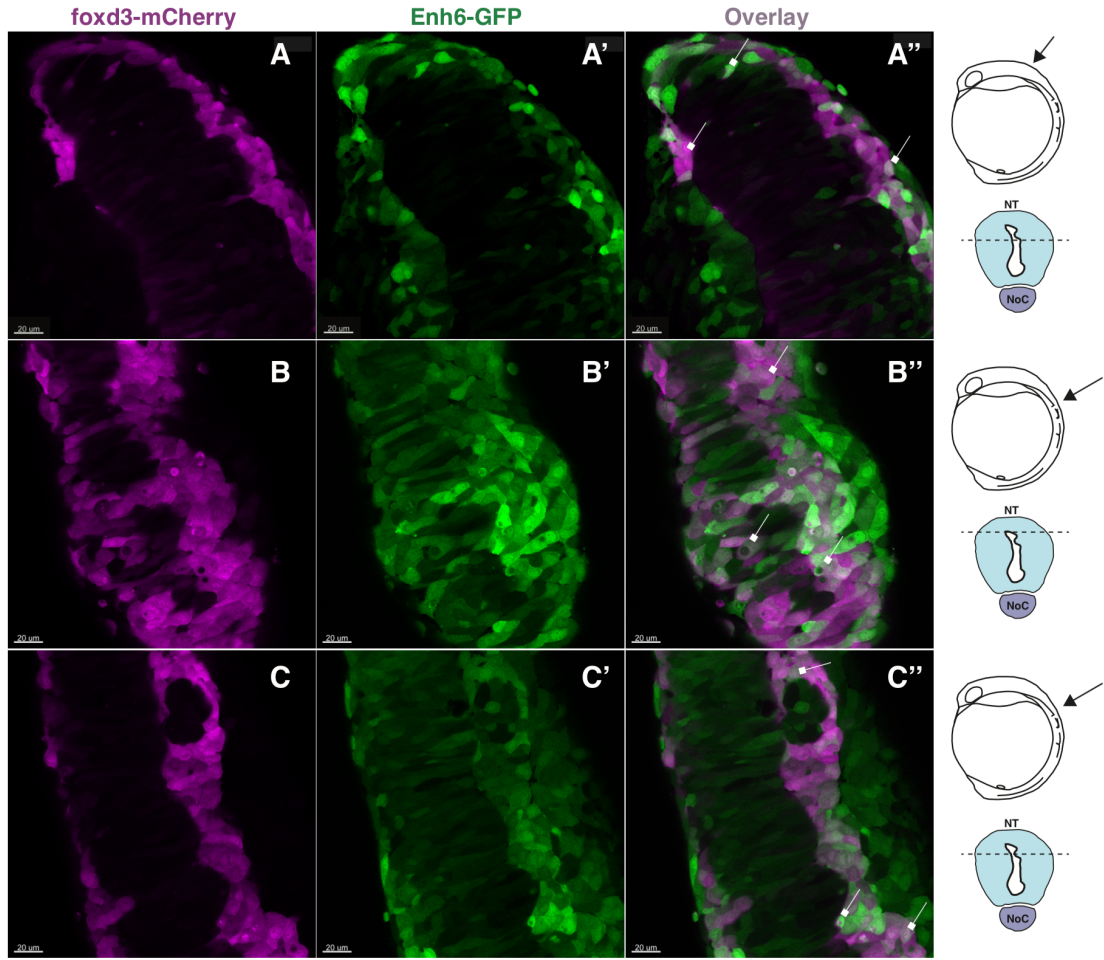


Figure 3.6 *Tg(FoxD3:enh6-GFP)^{ox109}* line crossed to *Gt(foxd3-mcherry)^{ct110aR}* line: live embryo dorsal view sections of neural tube at 5 somite stage. (A), (B), (C) *foxd3* expression as seen from *Gt(foxd3-mcherry)^{ct110aR}* reporter line fluorescence (coloured in magenta) across different sections. (A'), (B'), (C') Enh6 activity as seen from *Tg(FoxD3:enh6-GFP)^{ox109}* line fluorescence across corresponding sections as above. (A''), (B''), (C'') Image overlays of (A) - (A'), (B) - (B') and (C) - (C') images. The schematics on the right illustrate corresponding acquired image positions within an embryo. NT- neural tube, NoC - notochord. White arrows highlight overlapping cells expressing both - *foxd3*-mCherry and Enh6-GFP reporters. Scale bars correspond to 20μm.

On the whole, we have shown that Enh6 and Enh4 follow different gene regulatory instructions and are found to be active in distinctive cell populations in the zebrafish embryo during pre-migratory NC stages. First, Enh6 is positively regulated by a direct *foxd3* binding (Figure 2.1 A) and also shows motif enrichment for binding by other *bona fide* NC TFs (Figures 2.3 and 2.4). Furthermore, live embryo imaging has revealed that *foxd3* and Enh6-driven EGFP reporter were co-expressed in a large portion of *foxd3*+ cells (Figures 2.5 and 2.6). Conversely, Enh4 region, which is directly bound by *foxd3*

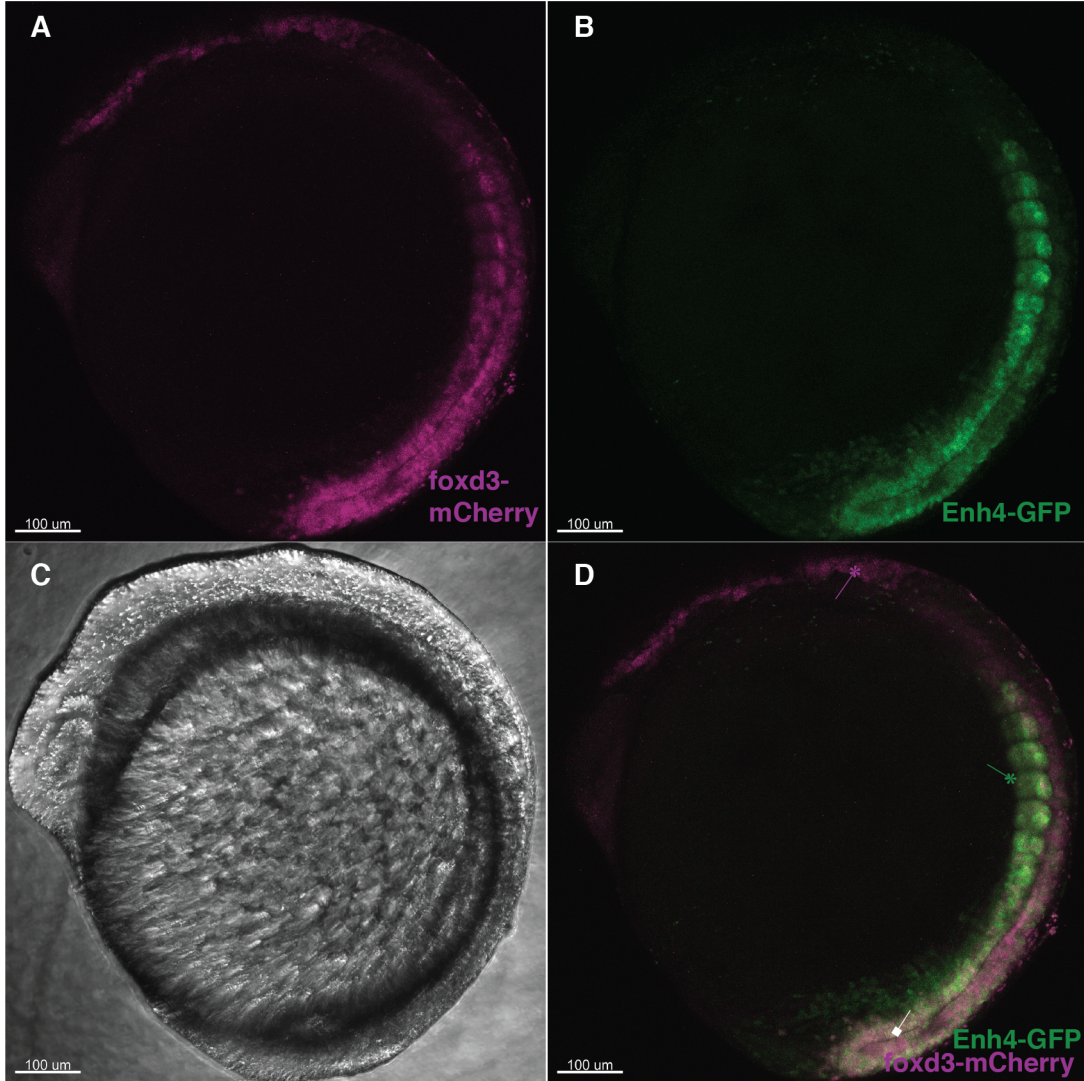


Figure 3.7 *Tg(FoxD3:enh4-GFP)^{ox110}* line crossed to *Gt(foxd3-mcherry)^{ct110aR}* line: lateral view of the whole-mount live embryo at 6 somite stage. (A) foxd3 expression as seen from *Gt(foxd3-mcherry)^{ct110aR}* reporter line fluorescence (coloured in magenta). (B) Enh4 activity as seen from *Tg(FoxD3:enh4-GFP)^{ox110}* line fluorescence. (C) Brightfield image of the imaged embryo. (D) Image overlay of (A) and (B) images. White arrow highlights overlapping fluorescence; green arrow - *Tg(FoxD3:enh4-GFP)^{ox110}* enhancer reporter line-only fluorescence; magenta arrow - *Gt(foxd3-mcherry)^{ct110aR}* reporter line-only fluorescence. Scale bars correspond to 100μm.

at 5-6ss (Figure 2.1 A) and gets compacted upon direct foxd3 binding (Figures 2.1 and 2.2), is likely to be negatively regulated by this factor. In addition, Enh4 TFBS analysis did not identify NC TFs to be putatively binding to this region, suggesting that Enh4 is not controlled in a NC fashion. Moreover, we demonstrated that Enh4 is not active in the cranial NC, suggesting that either this element is repressed in the cells that express foxd3 abundantly in this region and/or cranial cells lack the expression of the upstream

3.4 Stable transgenic Enh6-GFP/Enh4-GFP zebrafish line imaging confirms different axial activity patterns

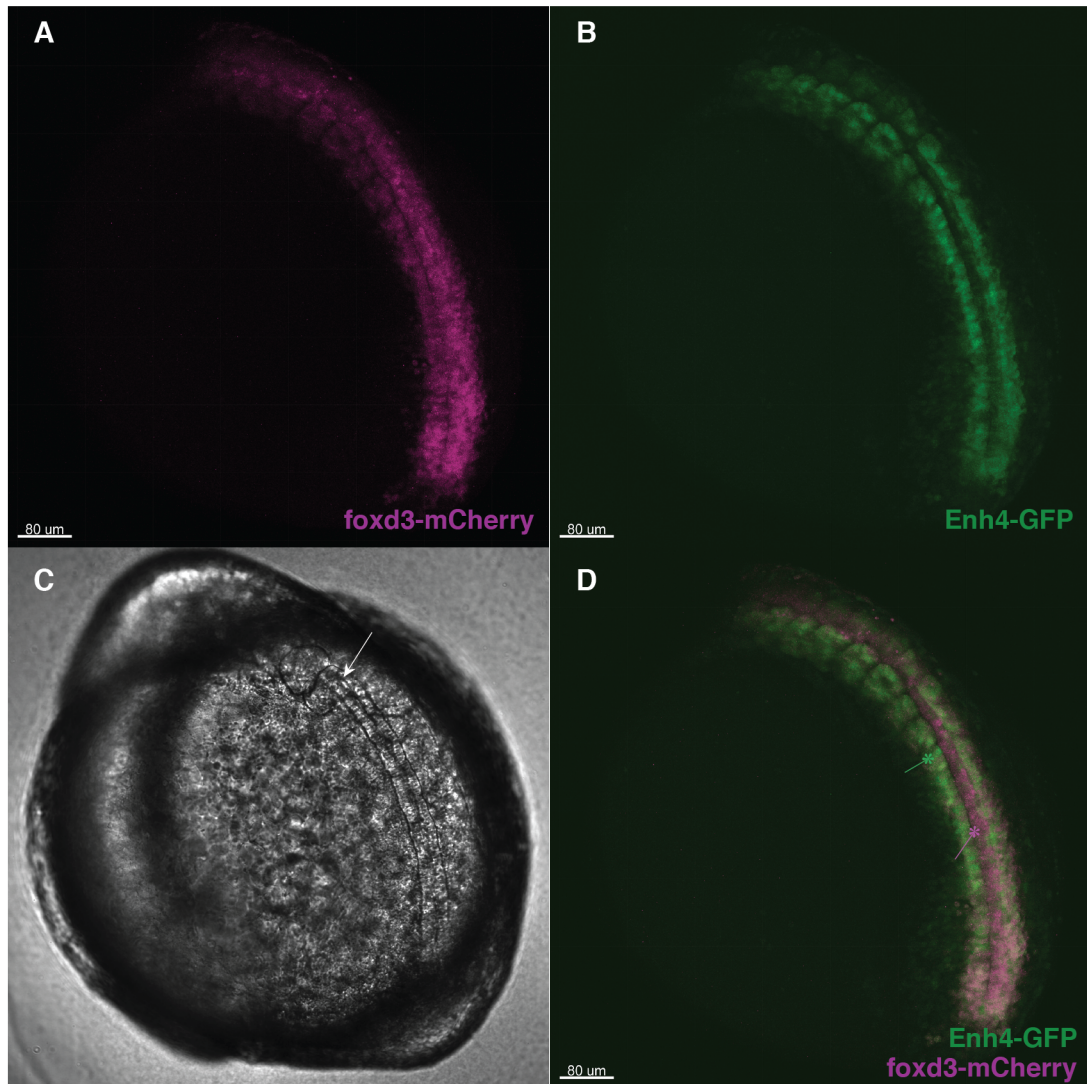


Figure 3.8 *Tg(FoxD3:enh4-GFP)^{ox110}* line crossed to *Gt(foxd3-mcherry)^{ct110aR}* line: dorsal view of the whole-mount live embryo at 7-8 somite stage. (A) *foxd3* expression as seen from *Gt(foxd3-mcherry)^{ct110aR}* reporter line fluorescence (coloured in magenta). (B) Enh4 activity as seen from *Tg(FoxD3:enh4-GFP)^{ox110}* line fluorescence. (C) Brightfield image of the imaged embryo. White arrow points to the neural tube. (D) Image overlay of (A) and (B) images. Green arrow - *Tg(FoxD3:enh4-GFP)^{ox110}* enhancer reporter line-only fluorescence; magenta arrow - *Gt(foxd3-mcherry)^{ct110aR}* reporter line-only fluorescence. Scale bars correspond to 100μm.

regulatory core factors driving this element. The latter possibility is less likely given that many of the putative regulatory factors controlling *enh4* (Figures 2.3 and 2.4) are expressed both in cranial and trunk regions. Finally, imaging embryos issued from crosses of *Tg(FoxD3:enh4-GFP)^{ox110}* enhancer reporter line and *Gt(foxd3-mcherry)^{ct110aR}* genetrapp reporter line, showed little overlap between cells expressing the reporters (Figure 2.9, further emphasising a negative relationship between *foxd3* and Enh4. We therefore

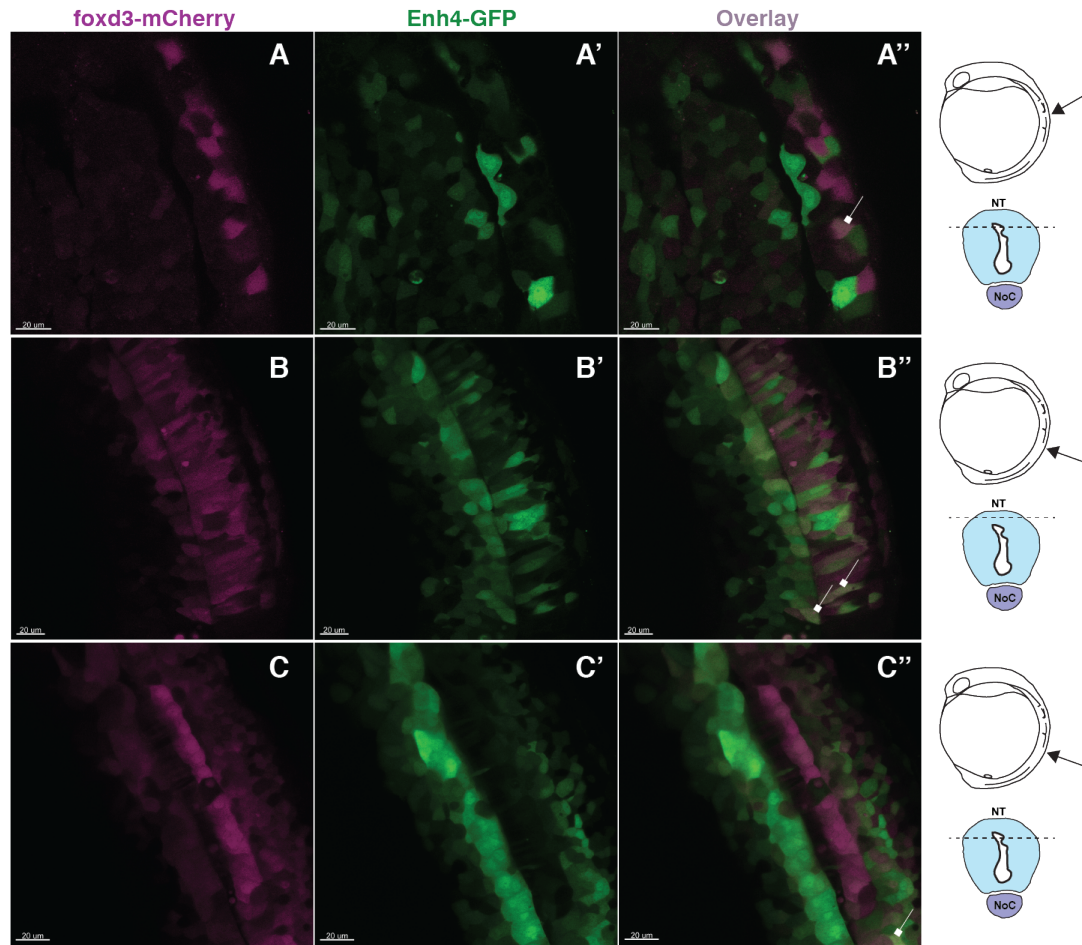


Figure 3.9 *Tg(FoxD3:enh4-GFP)^{ox110}* line crossed to *Gt(foxd3-mcherry)^{ct110aR}* line: live embryo dorsal view sections of neural tube at 5 somite stage. (A), (B), (C) *foxd3* expression as seen from *Gt(foxd3-mcherry)^{ct110aR}* reporter line fluorescence (coloured in magenta) across different sections. (A'), (B'), (C') *Enh4* activity as seen from *Tg(FoxD3:enh4-GFP)^{ox110}* line fluorescence across corresponding sections as above. (A''), (B''), (C'') Image overlays of (A) - (A'), (B) - (B') and (C) - (C') images. The schematics on the right illustrate corresponding acquired image positions within an embryo. NT- neural tube, NoC - notochord. White arrows highlight overlapping cells expressing both - *foxd3-mCherry* and *Enh4-GFP* reporters. Scale bars correspond to 20μm.

hypothesised that *foxd3* represses *Enh4* in the cranial NC. Intriguingly, *Enh4* was abundantly active in the pre-somitic mesoderm and somites of a 5-6ss developing embryo, suggesting that *foxd3* might need to be downregulated or repressed via *Enh4* in these mesodermal cells.

3.5 Defining transcriptional signatures of Enh6+ and Enh4+ cells

Our imaging results based on the phenotypic observations indicated that Enh6+ and Enh4+ cells comprise distinctive cell populations. While Enh6+ cells were likely to be of cranial ectodermal/NC origin, Enh4+ cells appeared to show mesodermal cell characteristics. In order to verify these hypotheses and understand the nature of Enh6+ and Enh4+ cell populations, we undertook an unbiased method to explore their transcriptomes at 5-6ss.

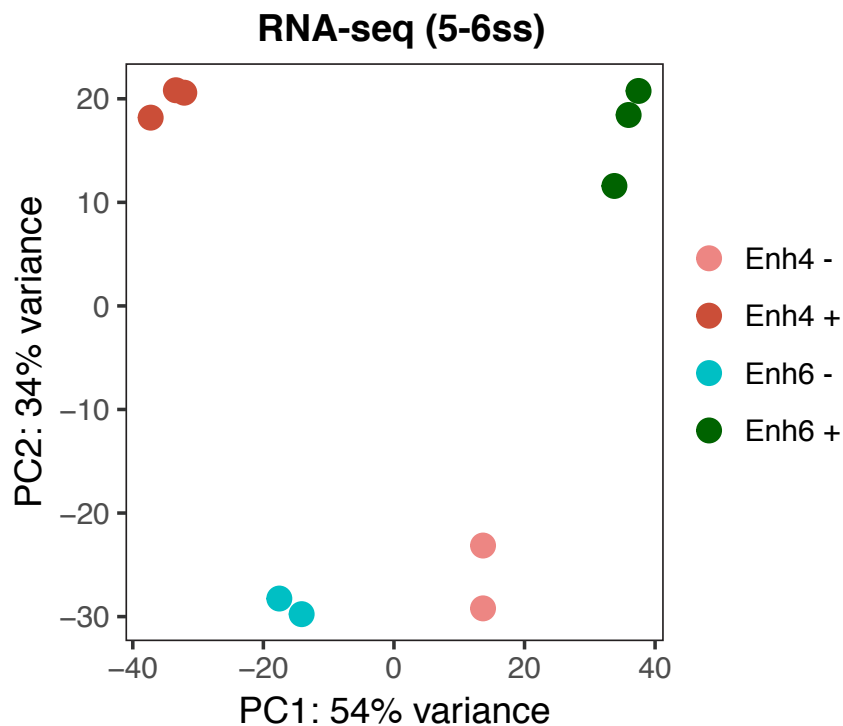


Figure 3.10 Principal Component Analysis (PCA) of RNA-seq samples and replicates. Triplicate samples from either *Tg(FoxD3:enh6-GFP)^{ox109}* or *Tg(FoxD3:enh4-GFP)^{ox110}* embryos were dissociated and EGFP-fluorescent embryonic cells were isolated using Fluorescence-activated cell sorting (FACS) method. Corresponding negative (GFP-) cells were also collected in duplicates as negative control samples. RNA-seq libraries for all samples were generated and sequenced. PCA analysis shows four distinctive groups based on differential gene expression (DeSeq2): Enh6+, Enh6-, Enh4+ and Enh4- groups.

To achieve this, we outcrossed *Tg(FoxD3:enh4-GFP)^{ox110}* or *Tg(FoxD3:enh6-GFP)^{ox109}* reporter lines to a wild-type zebrafish background and sorted embryos for the GFP signal at 5-6ss. We then dissociated whole positive embryos and sorted GFP+ and GFP- cells from each clutch using FACS in order to acquire four different sample groups: Enh6+,

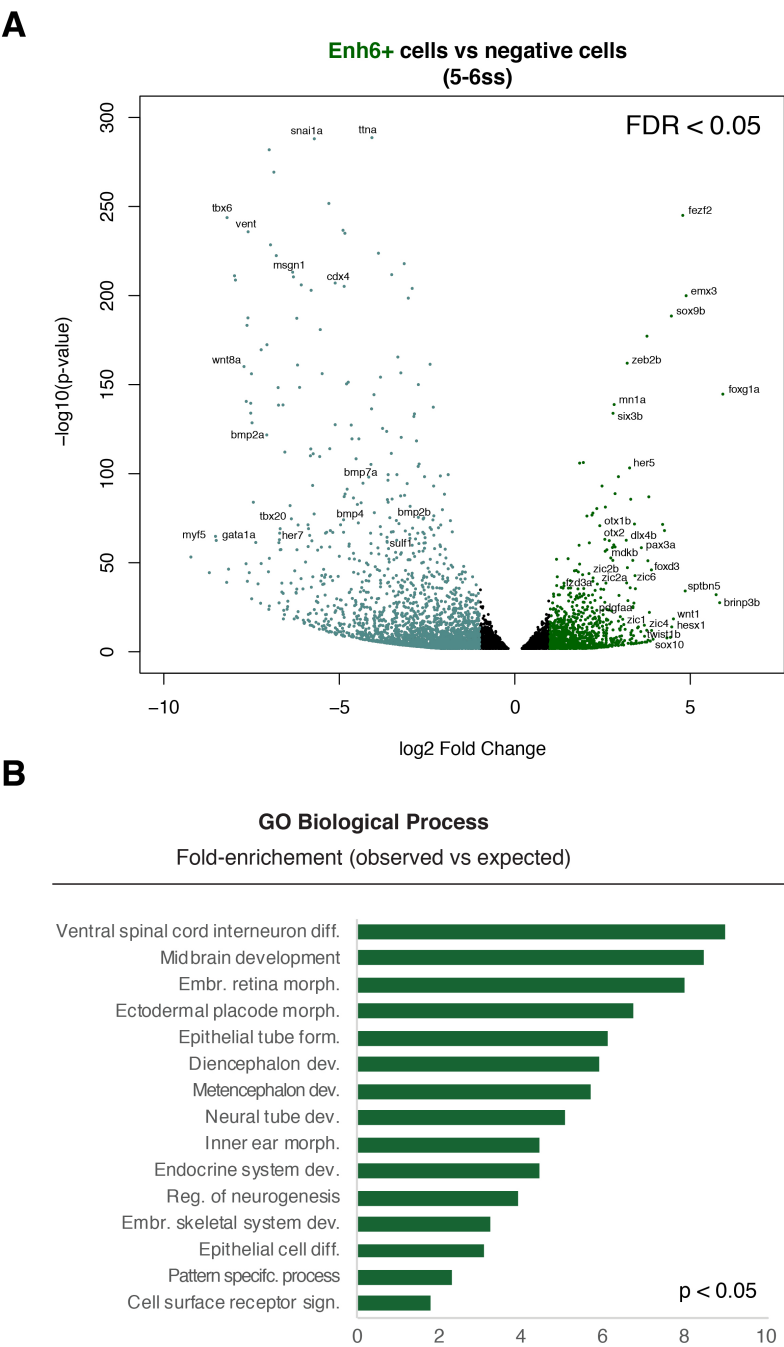


Figure 3.11 Profiling Enh6+ cells.(A) Volcano plot depicting differential expression analysis performed between Enh6+ and Enh6- cells. Only those genes that were found to be below false discovery rate (FDR) of 0.05 were taken into account. Genes on the right hand side from 0 were found enriched in the Enh6+ cell population when compared to Enh6- cells, with black dots representing genes with moderate Fold Changes ($0 < \log_2\text{fold} < 1$) and dark green dots - highly enriched genes with ($\log_2\text{fold} > 1$). Genes on the left hand side from 0 were depleted in the Enh6+ cells when compared to Enh6- cells, with black dots representing genes with ($0 < \log_2\text{fold} > -1$) change and teal green dots - strongly depleted genes at ($\log_2\text{fold} < -1$) change. (B) Bar plot showing selected significantly enriched ($p < 0.05$) gene ontology (GO) terms underlying genes that were enriched in the Enh6+ samples when compared to Enh6- samples ($\log_2\text{fold} > 1$).

Enh6-, Enh4+ and Enh4- for the bulk RNA-seq. Principal component analysis (PCA) based on gene expression separated all 4 groups and clustered replicates together, showing our data reproducibility and also further suggesting that Enh6+ and Enh4+ cells are different (see Figure 3.10).

To first comprehend what are the functions of cells with active Enh6 (Enh6+ cells), we compared molecular signatures of Enh6-positive cells to the cells that did not have an active Enh6 reporter within same embryos. After performing differential gene expression analysis, we identified significant 6514 differentially expressed genes (FDR < 0.05) (see Figure 3.11 A). Importantly, as predicted, we found that Enh6+ cells were expressing NC signature genes (e.g., *tfap2a*, *pax3a*, *sox10*, *sox9b*, *zic1*, *twist1b*, *pdgfaa*). We also confirmed that *foxd3* expression was enriched in the Enh6+ cells as compared to the rest of the embryonic cells (av. 422 normalised counts), further supporting a positive feedback loop between *foxd3* expression and its predicted Enh6. The significantly enriched Gene Ontology (GO) terms associated to the Enh6+ cell population reflected the biological processes involved in the formation of otic vesicle or retina morphogenesis (e.g., *tfap2a*, *gjf3*, *dlx4b*, *egr1*, *irx7*, *fozg1a*) (see Figure 3.11 B). Overall, GO terms were also enriched for the neural tube development (e.g., *eng2a*, *her5*, *shroom3*, *wnt1*, *pax2a*, *otx1/2*, *gli2b*), which could explain the GO terms also pointing to the central nervous system (CNS) development (i.e., forebrain, midbrain, hindbrain, spinal cord) (Figure 3.11). GO terms were overall associated with regulation of neurogenesis (e.g., *nkx2.2a/b*, *lhx4*, *notch1b*, *slit1a*, *zic5*, *cntn2*, *sema3fa*, *robo2*). Given that NC is known to give rise to peripheral nervous system (PNS) and enteric nervous system (ENS), but not CNS [238, 239], we hypothesise, that Enh6 gets activated earlier in the ectoderm that contributes to both, NC and neural tube, generation, the latter giving rise to the CNS [240, 241].

To gain insight into the nature of biological processes Enh4+ cells were involved in, we performed differential gene expression analysis comparing cells expressing Enh4-driven GFP reporter versus negative cells from the same embryo clutches, as we did for the Enh6+ versus Enh6- profiling. In doing so, we were able to assess what differentiates Enh4+ cells from the rest of the embryonic cells at 5-6ss. As a result, we have iden-

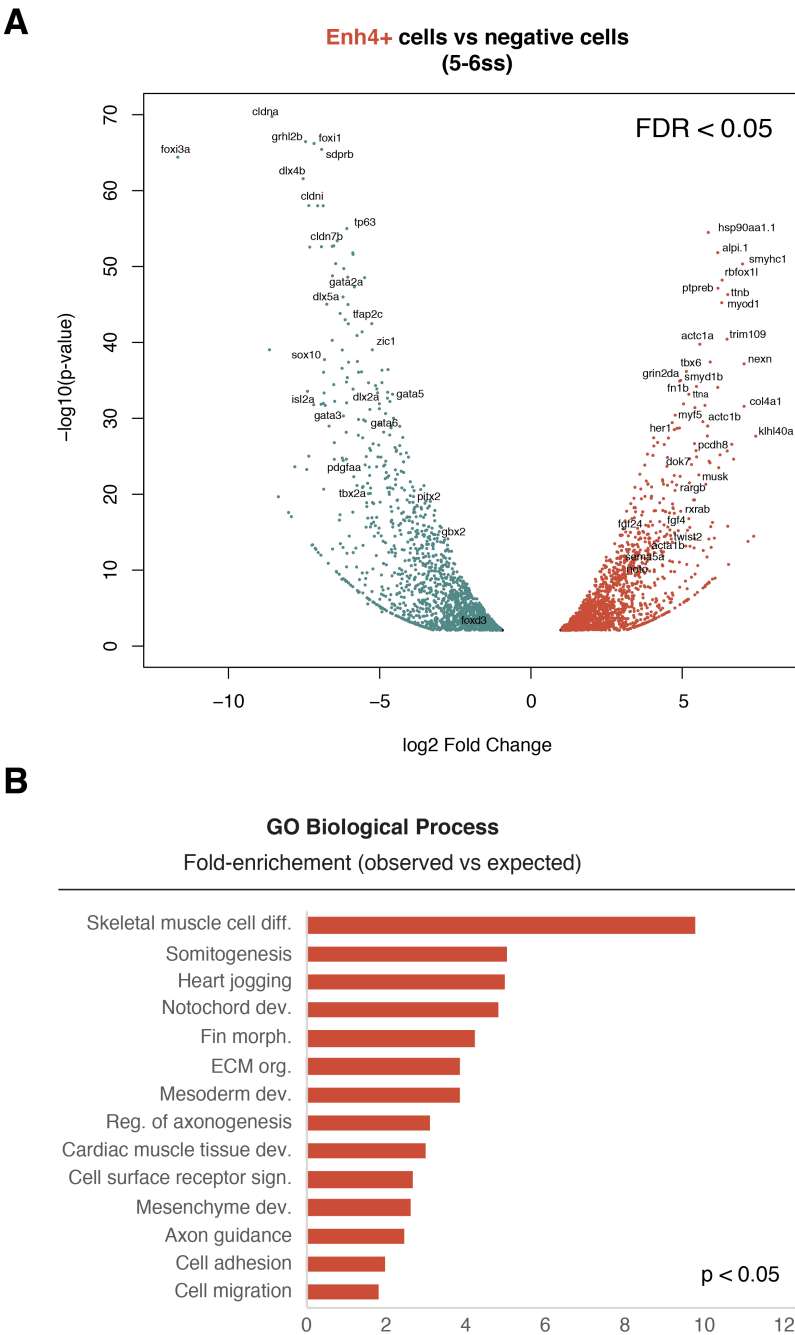


Figure 3.12 Profiling Enh4+ cells.(A) Volcano plot depicting differential expression analysis performed between Enh4+ and Enh4- cells. Only those genes that were found to be below false discovery rate (FDR) of 0.05 were taken into account. Genes on the right hand side from 0 were found enriched in the Enh4+ cell population when compared to Enh4- cells, with black dots representing genes with intermediate log Fold Changes ($0 > \log_2\text{fold} < 1$) and red dots - highly enriched genes with ($\log_2\text{fold} > 1$). Genes on the left hand side from 0 were found depleted in the Enh4+ cells when compared to Enh4- cells, with black dots representing genes with ($0 < \log_2\text{fold} < -1$) and teal green dots - genes with ($\log_2\text{fold} < -1$). (B) Bar plot showing selected significantly enriched ($p < 0.05$) gene ontology (GO) terms underlying genes that were enriched in the Enh4+ samples when compared to Enh4- samples ($\log_2\text{fold} > 1$).

tified 3522 significant differentially expressed genes (FDR < 0.05). Contrary to what we have shown for the Enh6+ gene expression enrichment vs negatives, we found that levels of the *foxd3* expression were overall low (av. 37.2 normalised counts) and depleted in the Enh4+ cells as compared to their GFP-negative controls (which of course included highly expressing anterior cranial cells, albeit in a much smaller proportion) (see Figure 3.12 A). Moreover, most of canonical *bona fide* NC markers, such as *sox10*, *tfap2a*, *pax3a*, *pdgfaa*, *twist1a* or *zic1*, were also depleted in the Enh4+ cell population (see Figure 3.12 A). These results indicated two things. First, Enh4 activity showed a negative correlation with the endogenous *foxd3* gene expression, further supporting our hypothesis of an underlying negative auto-regulatory feedback loop. Second, Enh4+ cell population did not exhibit *bona fide* NC gene expression signature, which suggested that they were not NC cells. Notably, the GO terms, underlying enriched genes in the Enh4+ cell population versus Enh4- cells, confirmed our imaging observations that Enh4 element might be active in the mesoderm and somites (see Figures 3.7, 3.8 and 3.12 B). Mesoderm derivatives, such as somites, notochord, skeletal and cardiac muscles were the most enriched GO biological process terms underlying the genes significantly enriched in the Enh4+ cell population (e.g., *myod1*, *myf5*, *klhl40a/b*, *tbx18*, *tbxta/b*, *mespaa/b*, *cdx4*, *noto*, *lama1/4*, *wnt3a*, *hmcn2*). Intriguingly, we have also detected enrichment in neural genes associated with regulation of axonogenesis and axon guidance (e.g., various *sema* factors, *slit3*, *ntn1a*, *plxnb1a/b*, *tcf3b*, *her1*, *tbr1b*, *lhx1a*, *tbx6*, *epha2b*, *prickle1b*, *gli2a*) (see Figure 3.12).

Recently, a bipotent cell population, coined neuromesodermal progenitors (NMps), that is capable of giving rise to both, somites and spinal cord cells, was discovered in mammalian embryos [242, 243]. Several studies have shown that NMps in amniotes exhibit self-renewing properties and continuously deposit spinal cord and paraxial mesoderm during posterior axis elongation, long after gastrulation is completed [243–245].

Unlike in amniotes, in anamniotes, such as zebrafish, the contribution of NMp self-renewal and pool expansion to the elongation of the body axis is still a subject of discussion, mainly due to mechanistic differences in body axis elongation between the taxa [246].

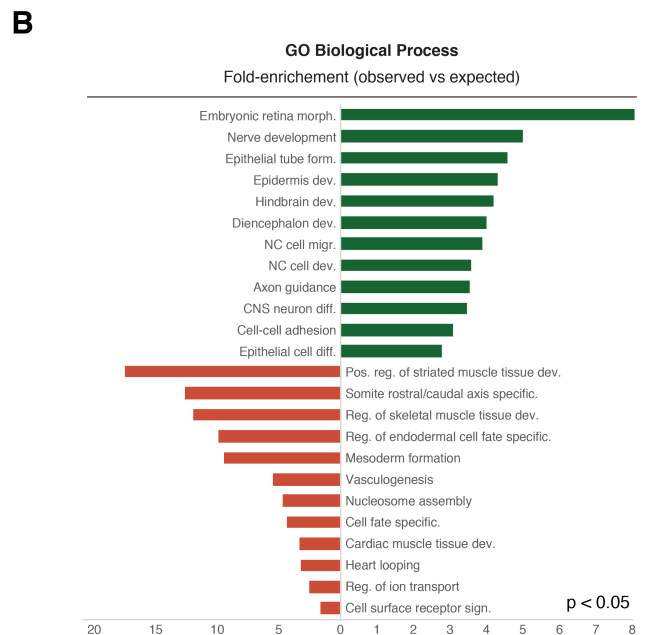


Figure 3.13 Comparing Enh6+ cells with Enh4+ cells.(A) Volcano plot depicting differential expression analysis performed between Enh6+ and Enh4+ cells. Only those genes that were found to be below false discovery rate (FDR) of 0.05 were taken into account. Genes in black on the right hand side from 0 represent genes between ($0 > \log_2\text{fold} < 1$) change and in dark green - genes at ($\log_2\text{fold} > 1$) change, hence enriched in the Enh6+ cells when compared to Enh4+ cells. Genes in black on the left hand side from 0 represent genes between ($0 < \log_2\text{fold} > -1$) change and in red - genes at ($\log_2\text{fold} < -1$) change, hence enriched in the Enh4+ cells when compared to Enh6+ cells. (B) Bar plot in green showing selected significantly enriched ($p < 0.05$) gene ontology (GO) terms underlying genes that were enriched in the Enh6+ samples when compared to Enh4+ samples and GO terms in red shows underlying genes that were enriched in the Enh4+ samples when compared to Enh6+ samples ($\log_2\text{fold} > 2$ for Enh6 enrichment or $\log_2\text{fold} < -2$ for Enh4 enrichment).

Amniotes first generate head structures during early gastrulation and posterior structures are extended only later in development, whereas zebrafish generates both head and trunk structures during convergence and extension movements in gastrulation [246]. Nonetheless, NMP population was confirmed to be conserved in zebrafish embryos [247, 248]. However, current imaging-based fate-mapping results argue that, differently from mouse and chicken embryos, in zebrafish, NMP pool quickly segregates into mesodermal and neural cell fates during gastrulation giving rise to a subset of spinal cord cells at all axial levels [248]. Then, similarly to mice embryos, a second NMP pool arises in the tailbud of developing zebrafish embryos [247, 248]. However, this pool has been shown to be stable, not expanding and is comprised of largely quiescent cells contributing to the most sacral part of the tail [248].

The exact markers of such NMPs are not entirely established yet, but the NMP cells have been shown to co-express *sox2*, *T/brachyury* (*tbxta*), *cdx1/2/4*, *tbx6*, *msgn1* [249–252]. Our bulk RNA-seq analysis of Enh4+ population has confirmed medium to high expression levels of these genes (average normalised counts): *sox2* - 4,012, *tbxta* - 50,090, *cdx4* - 1,262,684, *tbx6* - 18,660, *msgn1* - 64,206. Thus, Enh4 reporter is active in cells exhibiting both mesodermal and neural signatures, but given that the RNA-seq analysis was performed at the population level, we cannot argue that these genes are co-expressed in single Enh4+ cells that would represent the NMPs. Therefore Enh4+ cells (1) could be NMPs, (2) could have arisen from a single bipotent NMP pool or (3) could be a mixture of both (derivatives of early NMPs and tailbud NMPs).

Next, we have performed differential gene expression analysis between Enh6+ and Enh4+ cells. Overall, we acquired 7282 significant differentially expressed genes (FDR < 0.05), highlighting substantial differences between these two cell populations. Similarly to the



Figure 3.14 Schematics of hox gene expression in a zebrafish larva. Figure adapted from BJ Swalla’s review ‘Building divergent body plans with similar genetic pathways’[253].

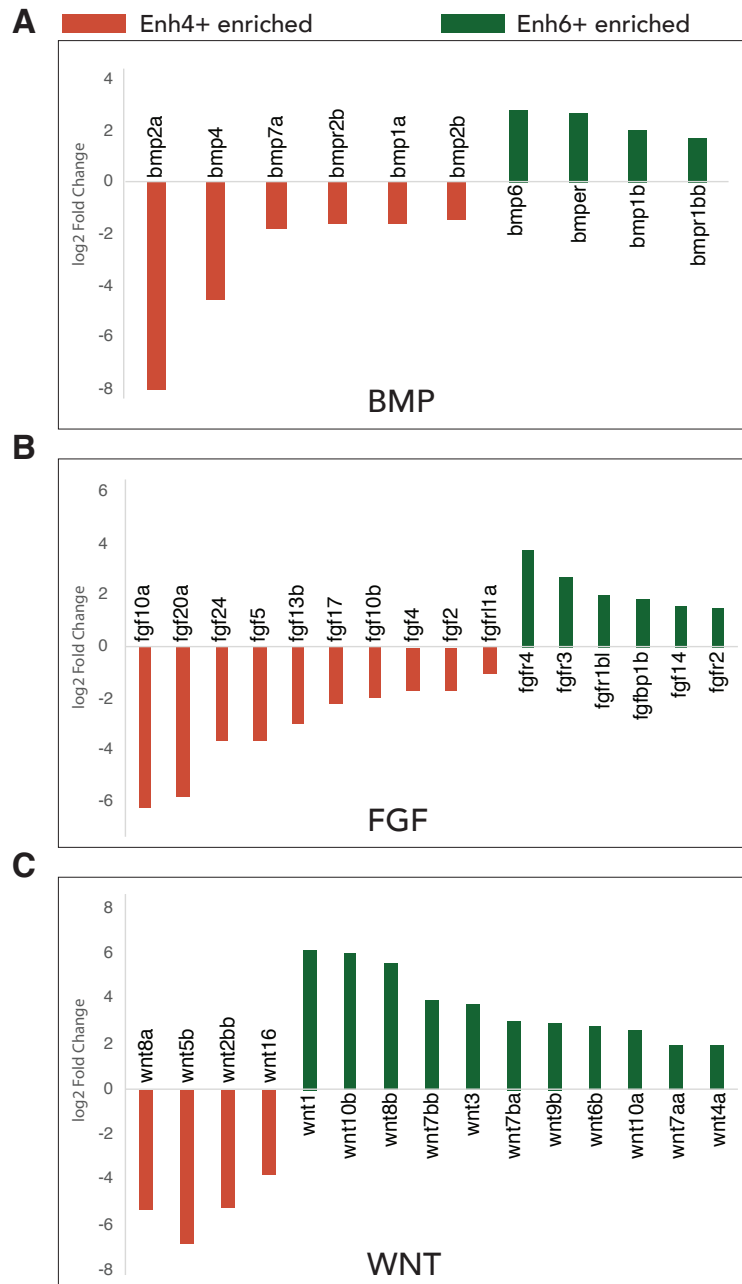


Figure 3.15 Differences in Bmp, Fgf and Wnt signalling molecule expression between Enh6+ and Enh4+ cells. (A) Differential RNA expression of bone morphogenetic proteins (BMP) and associated molecules between Enh4+ cells (in red) and Enh6+ cells (in green). (B) Differential RNA expression of fibroblast growth factors (FGFs) and associated molecules between Enh4+ cells (in red) and Enh6+ cells (in green). (C) Differential RNA expression of Wnt signalling pathway molecules between Enh4+ cells (in red) and Enh6+ cells (in green).

previous results, this also further emphasised a NC-like gene signature in the Enh6+ cells, but not in the Enh4+ cells (see Figure 3.13). Also, *foxd3* expression was enriched

in the Enh6+ cell population (Figure 3.13 A). Hence, this is indicative of Enh6 having a positive and Enh4 a negative effect on *foxd3* gene expression *in vivo*. Of note, *hox* gene expression further supports different axial localisation of Enh6+ and Enh4+ cells. Enh6+ cells were enriched for cranial *hox* genes patterning the cranial NC and rhombomeres/hindbrain: *hox1a*, *hoxb1a* and *hoxb2a* [254–256]. On the other hand, Enh4+ cells expressed more caudal and spinal cord *hox* genes (listed from the most enriched to the least enriched in terms of log2 fold change when compared to Enh6+ cells, ($10.3 < \log_2\text{fold} > 3$)): *hoxd12a*, *hoxa13b*, *hoxd10a*, *hoxa10b*, *hoxa11b*, *hoxc10a*, *hoxd9a*, *hoxa9b*, *hoxb10a*, *hoxa9a*, *hoxd11a*, *hoxb7a*, *hoxc10a*, *hoxa11a*, *hoxc11b*, *hoxc9a*, *hoxc6b*, *hoxd13a*, *hoxc11a*, *hoxc8a*, *hoxc5a*, *hoxb9a*, *hoxc3a*, *hoxc6a*, *hoxb13a*, *hoxb8a*, *hoxc12b*, *hoxb6a*, *hoxc4a*, *hoxd3a*, *hoxc13b*, *hoxb8b*, *hoxb6b*, *hoxb5b*, *hoxb4a*, *hoxb5a*, *hoxa13a* [257]. See *hox* gene expression diagram in a zebrafish larva (adapted from [253], Figure 3.14). Intriguingly, Enh6+ and Enh4+ cells also expressed different EMT factor orthologues: Enh6+ cells expressed *twist1a/b* and *snai1b* orthologues, while Enh4+ cells expressed *twist2* and *snai1a*. Other interesting observations included differences in expression patterns of BMP, FGF and Wnt signalling factors between the two cell populations (see Figure 3.15). Bmp and fgf molecules were more prevalent in the Enh4+ cells, while the receptors involved in these signalling pathways were more abundant in the Enh6+ cells (Figure 3.15 A-B). On the other hand, RNA expression of various Wnt signalling molecules was higher in the Enh6+ cells in comparison to Enh4+ cells (Figure 3.15 C).

3.6 *foxd3*+ single cell RNA-seq reveals trunk NMP-related cell populations

We next probed the transcriptional states of *foxd3*-expressing cells at 5-6ss in order to investigate the diversity of cell subpopulations expressing this factor at both cranial and trunk levels and heterogeneity within those in an unbiased fashion. To this end, we performed single-cell RNA-sequencing (scRNA-seq) of the entire Citrine-positive population within the 5-6ss embryos issued from *Gt(foxd3-citrine)ct110a* transgenic line. The embryos were dissociated and Citrine-positive cells were isolated by FACS, before loading

onto the Chromium 10X Genomics chip. 2,455 cells distributed into eight clusters as visualised in the UMAP plot, forming two distinct entities (see Figure 3.16). Clusters 0, 1, 2, 4 and 7 clearly showed anterior NC identity, with a broad distribution of canonical neural crest markers (*sox10*, *pax3a*, *foxd3*) as well as markers of anterior neural-NC identity (*otx2*) [117] (Figure 3.16). These cells expressed canonical NC markers reasonably uniformly in the zebrafish as we previously reported using smartseq2 scRNA-seq [113]. On the other hand, the molecular basis of their heterogeneity in an amniote model (chick) has also recently been described with some differences [117]. Interestingly, the other distinct grouping of 3 single cell clusters (Clusters 3, 5 and 6) had a clear posterior identity, as demonstrated by exclusive expression of posterior *hox* genes, for example, *hoxb10a* (Figure 3.16). Moreover, the assessment of top enriched genes within this posterior cell population, indicated enrichment in NMP markers with clear split between RA degrading and producing enzymes, *cyp26a1* (NMP maintenance/pro-neural) and *aldh1a2* (pro-mesodermal), respectively, expressing cells as previously shown [251]. All cells expressed *cdx4*, but there was a clear distinction between putative bipotent progenitors that co-expressed *tbxta* and *sox2*, and clearly lacked expression of bona fide NC markers *sox10*, *pax3a*, *foxd3*. Cluster 6 cells were enriched in definitive mesoderm markers, such as, *myl10* and *myod1* (shown later in Figure 3.17), as well as *aldh1a2*, suggesting these could represent newly forming somites that also expressed Citrine in our transgenic line. Differentiation pseudotime analysis of our data showed a very early bifurcated split of the cranial NC clusters and posterior NMP/mesodermal clusters, with all the cells belonging to corresponding clusters confounded (see Figure 3.16, last plot).

To increase the resolution of the analysis and gain the insight into possibly developmental trajectories of the putative posterior NMP population in zebrafish, we performed further subclustering on three posterior single-cell clusters (Clusters 3, 5 and 6) (see Figure 3.17). This analysis identified 6 new subclusters dividing this population of cells into more naïve progenitors and those forming mesodermal population. Umap plots showing expression of individual markers indicates the presence of bipotent NMPs with their canonical signature (*tbxta*, *sox2*, *msgn1*, *bmp2a*, *fgf8*) within the central single cell cluster

3.6 *foxd3*+ single cell RNA-seq reveals trunk NMP-related cell populations

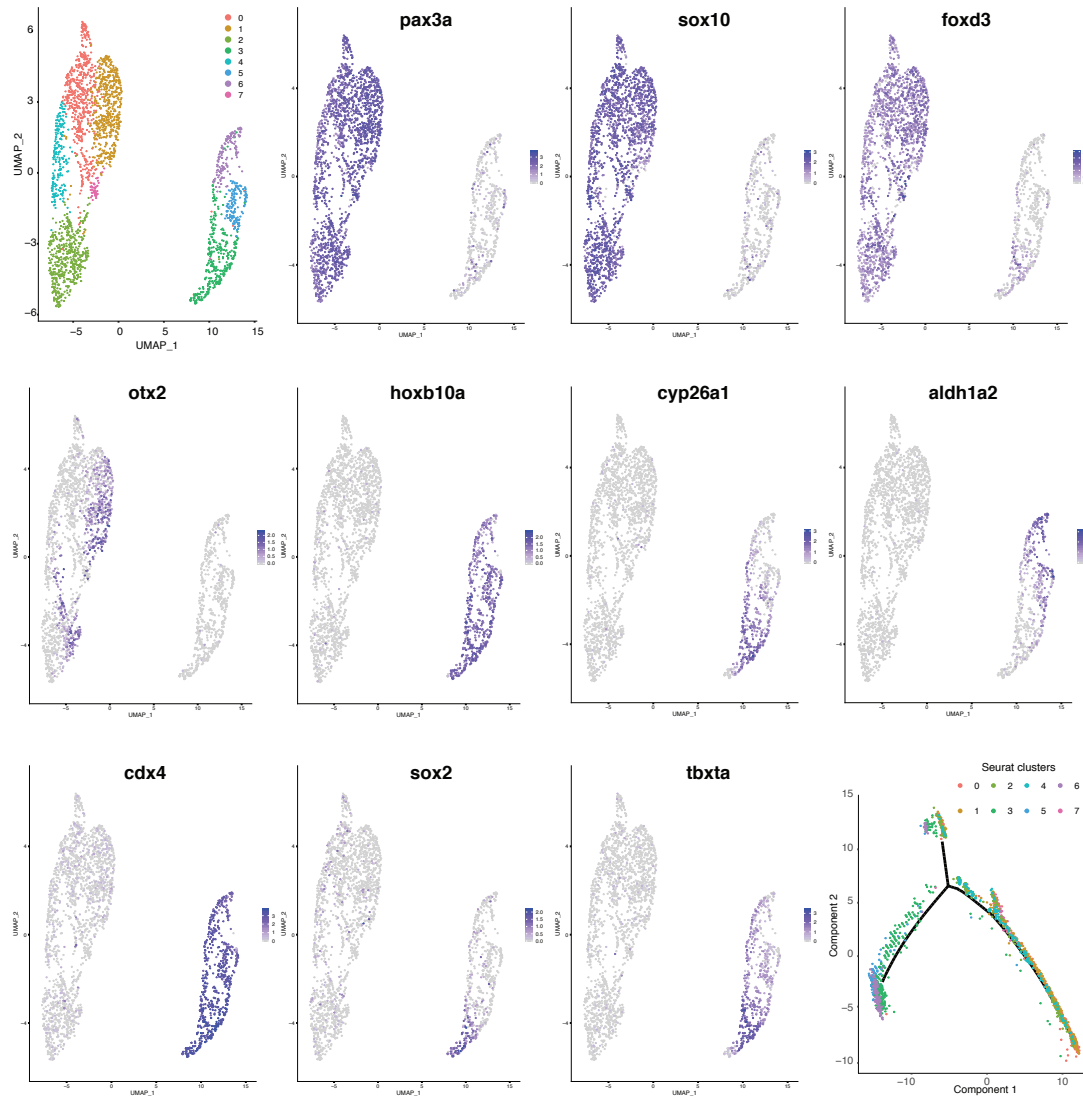


Figure 3.16 *Gt(foxd3-citrine)ct110a* single cell RNA-seq data reveals distinctive anterior and posterior cell types. Top left plot - displayed are 2,455 cells clustered based on their transcriptomes as revealed by Seurat analysis. Following plots highlight neural crest and nerome-sodermal progenitor-related gene expression levels within the clustered cells. Bottom right plot displays Monocle analysis results of the above Seurat 0-7 cluster cell differentiation in pseudo-time trajectories.

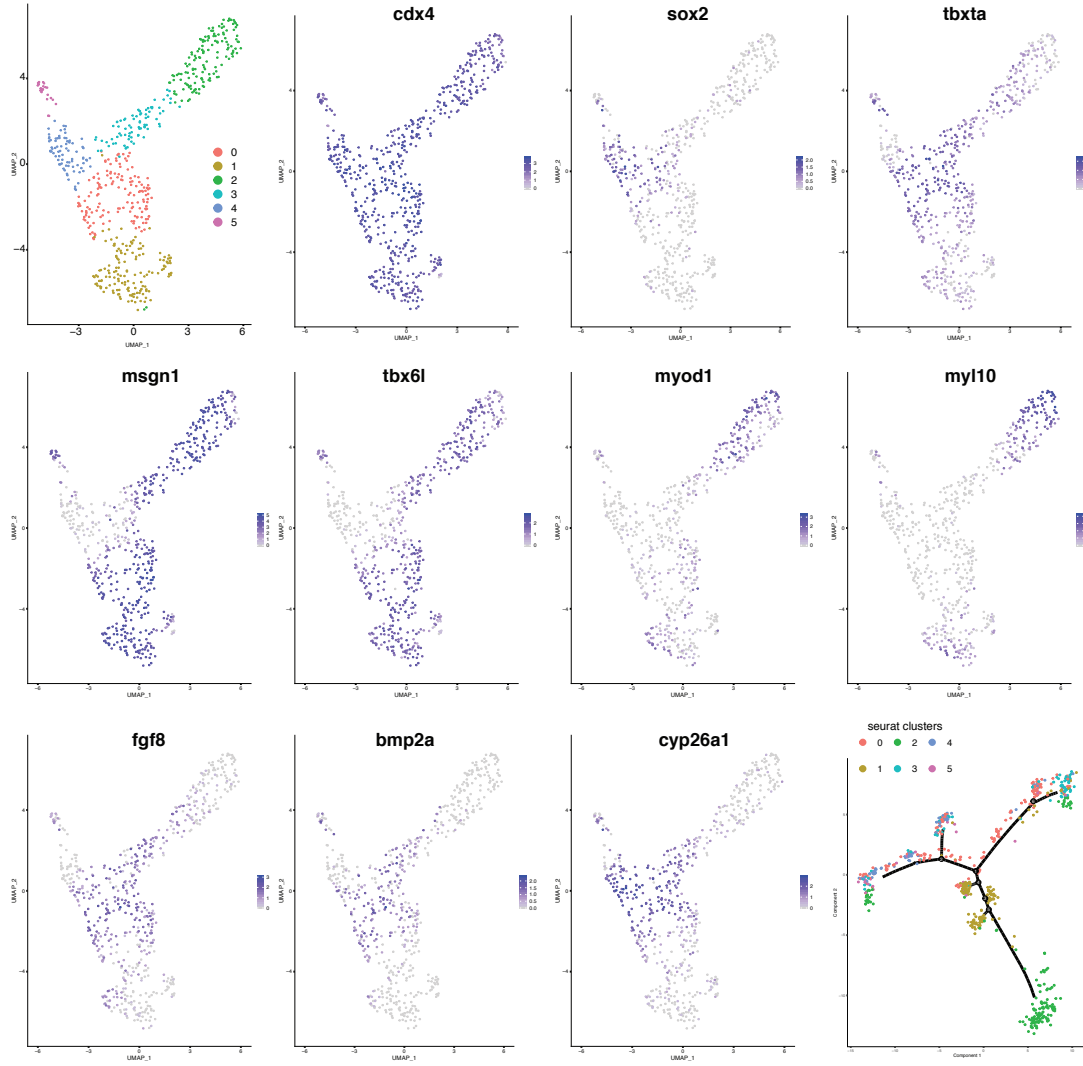


Figure 3.17 *Gt(foxd3-citrine)ct110a* single cell RNA-seq data identifies different neuromesodermal progenitor-related cell populations in zebrafish embryo. Top left plot - displayed are cells from the previously identified Seurat 3, 5 and 6 clusters (Figure 3.16), further subclustered based on their transcriptomes as revealed by Seurat analysis. Following plots highlight neuromesodermal progenitor-related gene expression levels within the subclustered cells. Bottom right plot displays Monocle analysis results of the above Seurat 0-5 subcluster NMP-related cell differentiation in pseudo-time trajectories.

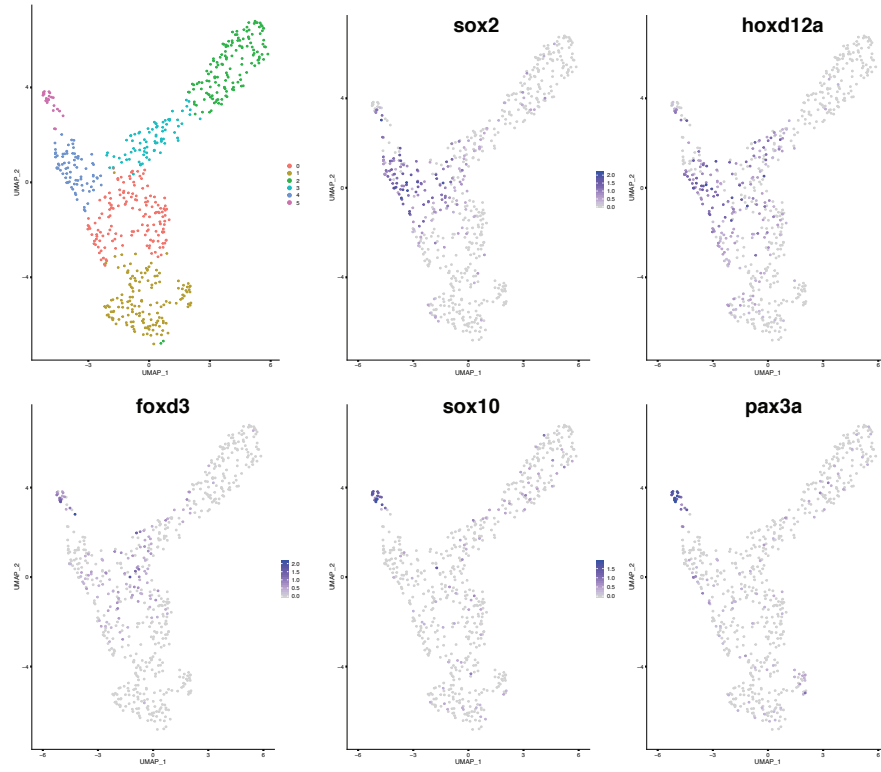


Figure 3.18 *Gt(foxd3-citrine)ct110a* single cell RNA-seq data identifies NC-like neuronal progenitors. Top left plot - displayed are cells from the previously identified Seurat 3, 5 and 6 clusters (Figure 3.16), further subclustered based on their transcriptomes as revealed by Seurat analysis. Following plots highlight *sox2* and posterior *hox* expression, as well as neural crest regulator expression.

(see Figure 3.17). The expression of NMP/pro-mesodermal markers (*msgn1*, *tbx6l* *tbxta*), as well as definitive mesodermal markers (*myod1*, *myl10*), was detected in cells belonging to two distinct subclusters (1 and 2, Figure 3.17). Pseudotime analysis of the posterior subclusters indicated an early bifurcated split of the subcluster 2 and likely independent mesodermal origin of these cells (probably paraxial mesoderm-derived somites) (Figure 3.17). Furthermore, we observed the concurrent migration of other subclusters, a distinct lower branching in the trajectory representing mesodermal progenitors likely originating from the NMps (subcluster 1, Figure 3.17), and two more branches with multiple cluster contributions that could not be resolved at this point. Interestingly, we have also discovered a small population of putative neuronal progenitors, with more posterior identity (*hox* genes) and expression of several canonical NC markers (Figure 3.18), suggesting that at neuronal differentiation stages NMps could give rise to NC-like neural progeni-

tors, possibly autonomic neurons or glia. Further pseudotime analysis spanning larger developmental time window, will be necessary to define these trajectories in greater detail.

Importantly, single cell data from *foxd3*-expressing cells further support our hypothesis of Enh4+ cells being of a NMP origin. We have identified cells with low *foxd3* transcript levels, expressing posterior *hox* genes and canonical NMP genes. This identified cell population from embryonic *foxd3*+ cells encompassing both, cranial and trunk cells, is very likely to correspond to the Enh4+ cells due to their analogous axial positioning and gene expression patterns.

3.7 Further defining Enh6+ and Enh4+ cells based on their accessible DNA regions

Having characterised the transcriptomes of Enh6+ and Enh4+ cells, we next set out to compare their accessible chromatin landscapes. Notably, this approach is poised to identify not only the present state of the cells in terms of gene expression, but also predict which programmes the cells are commencing by remodelling and opening underlying gene chromatin. To achieve this, we have employed similar experimental strategy as for the RNA-seq experiments. Again, we crossed *Tg(FoxD3:enh4-GFP)^{ox110}* or *Tg(FoxD3:enh6-GFP)^{ox109}* reporter lines to a wild-type zebrafish line and sorted embryos for the GFP signal at 5-6ss. As before, we then dissociated whole positive embryos and sorted GFP+ and GFP- cells from each clutch using FACS in order to acquire the same four different sample groups: Enh6+, Enh6-, Enh4+ and Enh4- for the bulk ATAC-seq.

First, we were able to verify that Enh6 was indeed accessible in Enh6+ cells, while Enh4 was open in Enh4+ cells (integrated plasmid reads were excluded, see Figure 3.19). Strikingly, ATAC signals underlying Enh4 and Enh6 were almost absent in Enh6+ and Enh4+ cells, respectively (see Figure 3.19). This indicates that at 5-6ss, the majority of the Enh6+ cells repress Enh4, while most of the Enh4+ cells also keep Enh6 in closed chromatin conformation. This suggests that these two elements are controlled by distinctive regulatory mechanisms at the chromatin level in different cell types, which possibly

3.7 Further defining Enh6+ and Enh4+ cells based on their accessible DNA regions

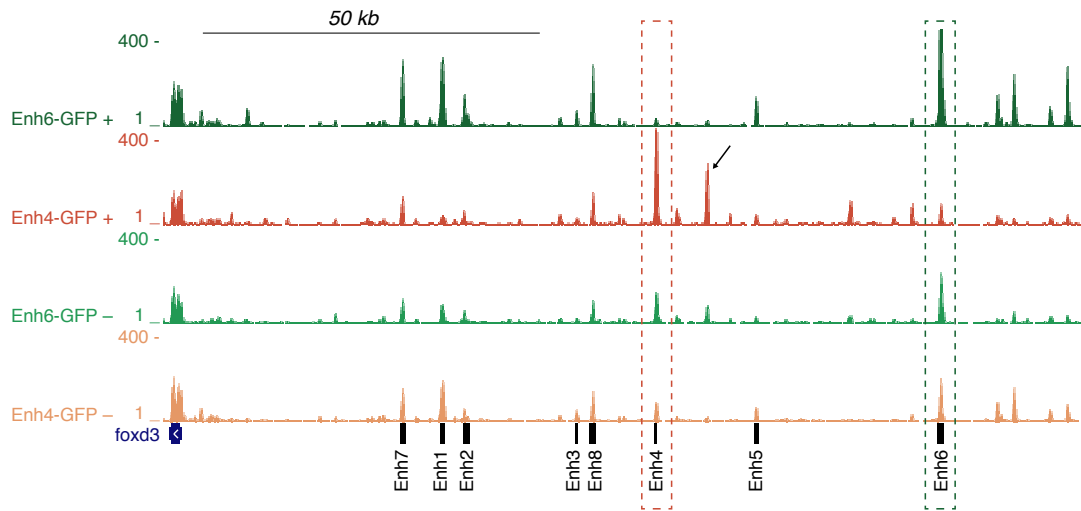


Figure 3.19 Surveying and comparing open chromatin regions within the *foxd3* locus between Enh6+, Enh6-, Enh4+ and Enh4- cells. ATAC-seq tracks from cells expressing Enh6 reporter and their negatives, and from cells expressing Enh4 reporter and their negatives. ATAC-seq reads were corrected for the transgenic Enh4 or Enh6 plasmid mapping. Black arrow highlights an Enh4+ cell population-specific open chromatin region, which was not previously observed from our NC-specific ATAC-seq data sets.

depends on available *foxd3* protein concentrations. Interestingly, we have also observed another Enh4+ cell population-specific peak ~8kb upstream of the Enh4. This could be another *cis*-regulatory element regulating *foxd3* gene expression in the Enh4+ cells, but not in the NC cells.

We next identified differentially open chromatin regions between Enh6+ and Enh4+ cells. To achieve this, we called peaks on all samples and their replicates choosing only the most significant peaks ($p < 10^{-10}$) that were found to be present in at least two out of three replicates for the downstream analysis. The consensus peak sets generated for each sample were merged to obtain a unified coordinate matrix to be used for the *k*-mean clustering algorithm [193]. *K*-means clustering was performed to identify distinct cohesive groups of elements (clusters) showing differential accessibility patterns and regions between the compared sample groups. As a result, we acquired 13 meaningful clusters (see Appendix 1 for all plotted heatmaps and associated GO terms). Two of the 13 clusters (Cluster 6-1 and Cluster 6-2) showed enriched ATAC signal underlying Enh6+ sample group (see Figure 3.20) and other two clusters were enriched in Enh4+ samples (Cluster 4-1 and Cluster 4-2, see Figure 3.21).

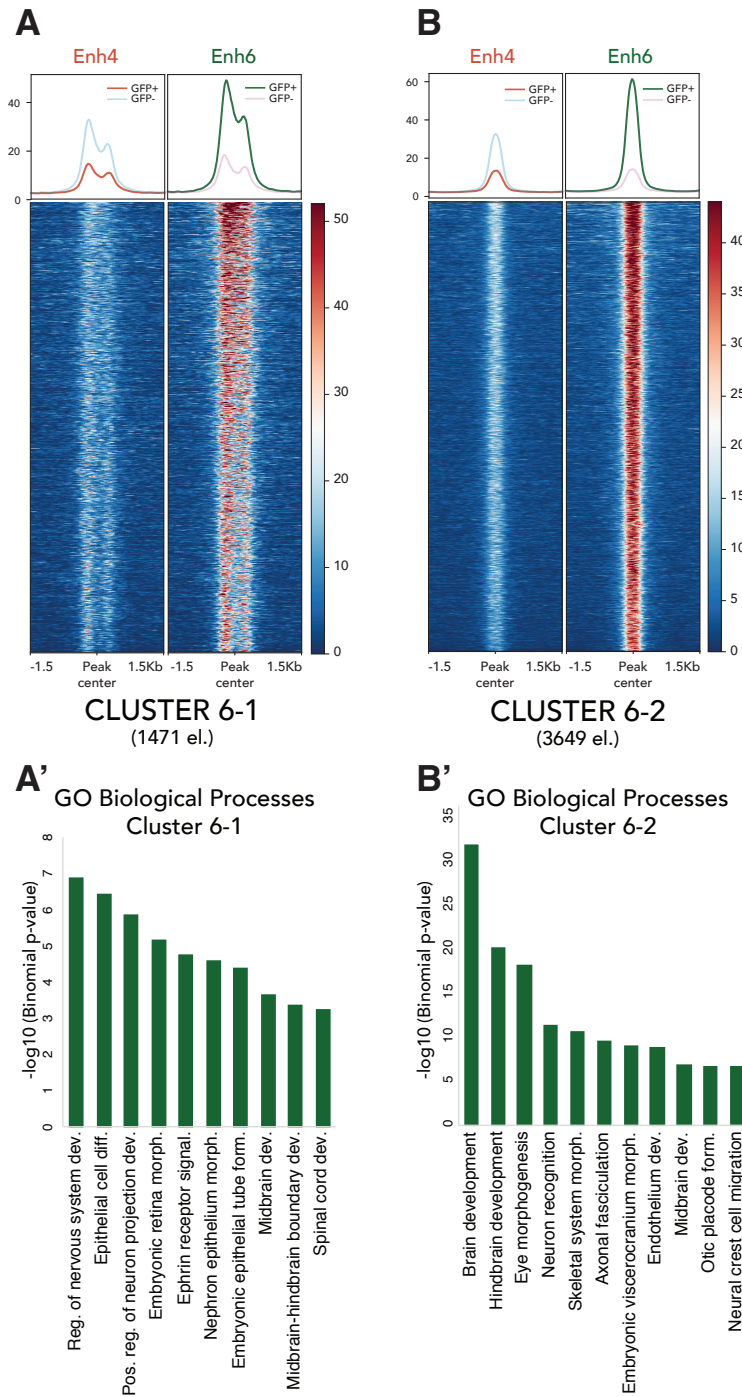


Figure 3.20 *k*-mean clustering reveals two *Enh6*⁺ sample-specific clusters. (**A** and **B**) ATAC-seq average profiles (top) and heatmaps (bottom) displaying Cluster 6-1 containing 1471 elements and Cluster 6-2 containing 3649 elements (1.5kb each side from the ATAC peak center) that were found to be accessible in the *Enh6*⁺ samples, but not in the most of *Enh4*⁺ cells. Profiles: light blue corresponds to *Enh4*⁺ cells, red - *Enh4*⁻ cells, pink - *Enh6*⁻ cells, green - *Enh6*⁺ cells. Heatmap legends indicate average ATAC signals. (**A'** and **B'**) Bar plots showing selected significantly enriched ($p < 0.01$) gene ontology (GO) terms underlying genes that were associated with Cluster 6-1 (**A**) and Cluster 6-2 (**B**) elements by GREAT analysis.

3.7 Further defining Enh6+ and Enh4+ cells based on their accessible DNA regions

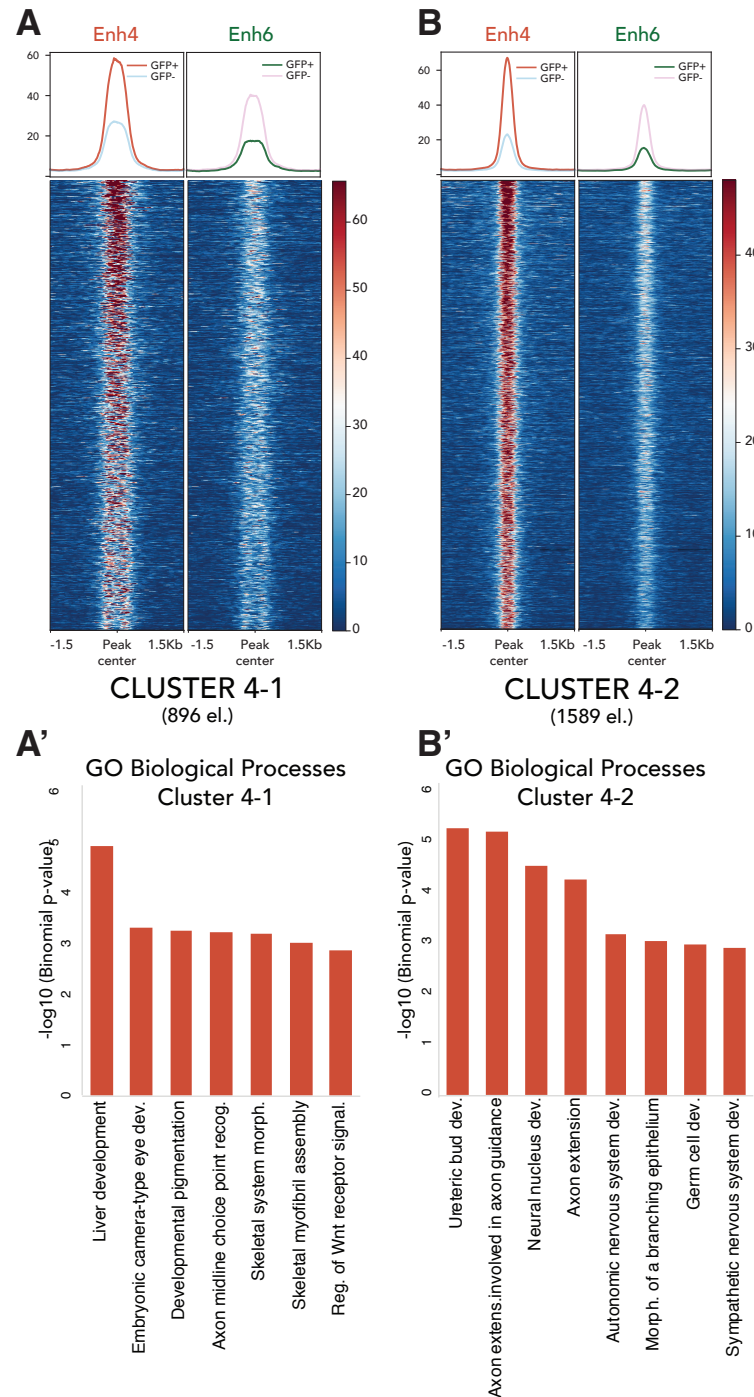


Figure 3.21 *k*-mean clustering reveals two Enh4+ sample-specific clusters. (**A** and **B**) ATAC-seq average profiles (top) and heatmaps (bottom) displaying Cluster 4-1 containing 896 elements and Cluster 4-2 containing 1589 elements (1.5kb each side from the ATAC peak center) that were found to be accessible in the Enh4+ samples, but not in the most of Enh6+ cells. Profiles: light blue corresponds to Enh4- cells, red - Enh4+ cells, pink - Enh6- cells, green - Enh6+ cells. Heatmap legends indicate average ATAC signals. (**A'** and **B'**) Bar plots showing selected significantly enriched ($p < 0.05$) gene ontology (GO) terms underlying genes that were associated with Cluster 4-1 (**A**) and Cluster 4-2 (**B**) elements by GREAT analysis.

Both Enh6+ cell specific clusters, Cl. 6-1 and Cl. 6-2 were associated with various nervous system development GO terms. The Cl. 6-2 appears to include elements of both, NC derivative and CNS formation (see Figure 3.20 B'). For instance, NC derivative GO terms were enriched for the NC derivative formation due to Enh6+ cell-specific open elements associated with genes, such as *sox3*, *sox9a/b*, *bmp6*, *pax2a*, *zeb1b*, *zic1* and *pax6a/b*. On the other hand, Cl. 6-1 elements were mainly enriched for regulation of nervous system development, such as ephrin receptor signalling, spinal cord development and midbrain development. Genes associated with these terms included *epha4/b*, *gli1/3*, *her6/9/11*, *olig2*, *slit3*, *eng2a*, *lhx1a*. Overall, these ATAC-seq results together with the RNA-seq results from the cranial Enh6+ cell population at 5-6ss, advocate that Enh6+ cells comprise not only NC cells, but also progenitors of the CNS.

Similarly, by looking at Enh4+ cell chromatin landscapes we were also able to at least partially distinguish Enh4+ cell-specific clusters, one of which showed a more pronounced mesodermal signature (Cl. 4-1, Figure 3.21 A) and the other one - a more neuronal signature (Cluster 4-2, Figure 3.21 B). Cl. 4-1 elements were assigned to genes, such as, *raraa/b*, *gata6*, *hsp90aa1.1*, *alcama*, *meis1/2a*. Cl. 4-2 neural GO terms were associated with *prickle1b*, *sema5a*, *nrp1a*, *gfra1a/b*, *fzd3b*, *olig2*, *isl2a*. These results further put the emphasis on Enh4+ cell population at 5-6ss having their chromatin permissively arranged for both, mesodermal and neural gene expression.

3.8 Surveying Enh6+ and Enh4+ cell derivatives in the zebrafish larvae

To finally validate whether the predicted cell fates of 5-6ss embryonic Enh6+ and Enh4+ cells eventuate, we imaged live zebrafish larvae at 2dpf from *Tg(FoxD3:enh6-GFP)^{ox109}* and *Tg(FoxD3:enh4-GFP)^{ox110}* lines crossed to the *Gt(foxd3-mcherry)^{ct110aR}* reporter line. By crossing enhancer reporter lines to the *foxd3* reporter line, we were again able to understand the underlying relation between *foxd3* expression and its predicted enhancers later in the zebrafish development. We identified that a small portion of *foxd3* expressing cells overlapped Enh6+ expression (see Figure 3.22 B", C", D"; Pearson's $r = 0.29$, $SD = \pm 0.12$; Figure 3.22). As for the Enh4 and *foxd3* co-expression, we saw

almost no overlapping cells despite *foxd3*+ and Enh4+ cells being in a close proximity (see Figure 3.23 B", C", D"; Pearson's $r = 0.1$, $SD = +/-0.13$; Figure 3.22). The underlying trend of *foxd3*-Enh6 and *foxd3*-Enh4 co-expression was similar at both compared stages (5-6ss and 2dpf): *foxd3* expressing cells were overlapping Enh4 expressing cells to a significantly smaller degree than Enh6-expressing cells (Welch's *t*-test; Figure 3.24). This indicates a persistent positive Enh6-*foxd3* and negative Enh4-*foxd3* auto-regulation during an embryonic development.

Tg(FoxD3:enh6-GFP)^{ox109} line imaging at 2dpf revealed that indeed Enh6+ cells give rise to the CNS: forebrain, midbrain, hindbrain and spinal cord (see Figure 3.22 A). They were largely absent from the *foxd3*+ mesodermal tissues (Figure 3.22 D-D"), but contributed to phenotypically apparent neurons/glia cells in the eye, hindbrain and spinal cord regions (see Figure 3.22 B-D"). These findings further support our previous RNA-seq and ATAC-seq interpretations of Enh6+ cells giving rise to both NC and CNS derivatives. Certainly, the exact cell types observed here still need to be validated in future.

Similarly, we were able to additionally support our hypothesis proposing that Enh4+ cells are able to give rise to both, mesodermal and neuronal cells, possibly indicating that we have identified a genetically stable reporter line for the NMps in zebrafish. Figure 3.23 A highlights that Enh4+ cells are found in mesodermal tissues and in nerve cells along the zebrafish trunk and tail. Interestingly, we identified what phenotypically appears as Enh4+ neurons in the midbrain regions of a zebrafish larvae (see Figure 3.23 B-B"). Given that at 5-6ss we observed Enh4+ cells only in the trunk/sacral embryo regions, this would suggest that either Enh4+ cells migrate towards cranial regions during development or Enh4 gets activated in the cranial areas at later developmental stages, hence this needs to be investigated in future. We also reiteratively observed phenotypically apparent Enh4+ nerve cells next to *foxd3*+ nerve cells, which never overlapped, suggesting distinctive nerve cell origins along the tail regions of a zebrafish larvae (Figure 3.23 C-D"). Similarly, we saw mesodermal regions, which were mutually exclusive for the *foxd3* and Enh4 reporters (Figure 3.23 D-D", white arrow), again suggesting a distinc-

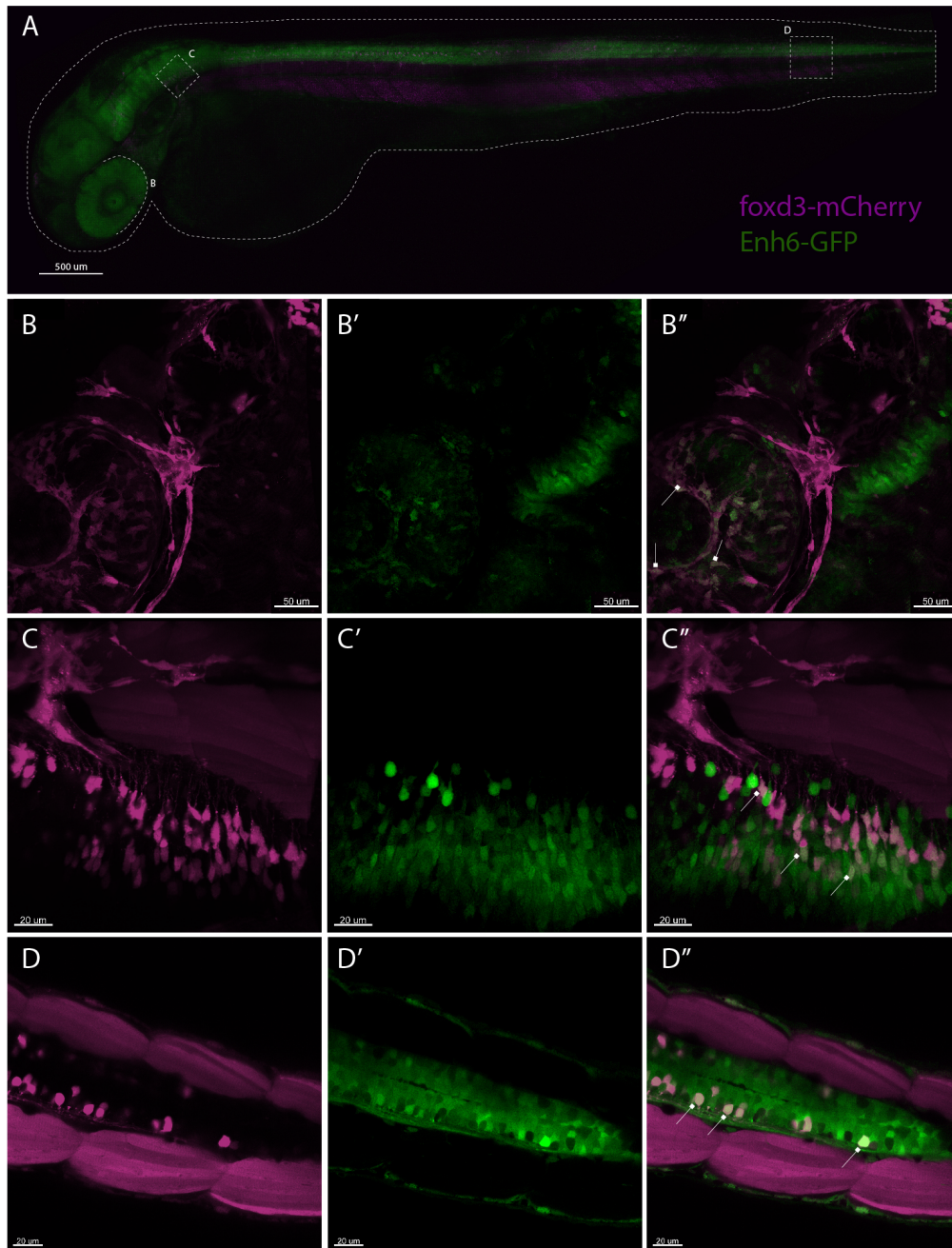


Figure 3.22 *Tg(FoxD3:enh6-GFP)^{ox109}* line crossed to *Gt(foxd3-mcherry)^{ct110aR}* line: whole-mount live larvae imaging at 2 days post fertilisation (dpf). (A) Lateral view of an entire zebrafish larva at 2dpf expressing *foxd3-mcherry* (coloured in magenta) and *Enh6-GFP* (in green). Letters B, C, D correspond to approximate positions within a zebrafish of the sections listed below (note, sections below were taken in a different larvae). Scale bar corresponds to 500μm. (B), (C), (D) *foxd3* expression as seen from *Gt(foxd3-mcherry)^{ct110aR}* reporter line fluorescence (coloured in magenta) across different sections - (B) - eye region, (C) - hindbrain region, (D) - tail. (B'), (C'), (D') *Enh6* activity as seen from *Tg(FoxD3:enh6-GFP)^{ox109}* line fluorescence across corresponding sections as above. (B''), (C''), (D'') Image overlays of (B) - (B'), (C) - (C') and (D) - (D') images. White arrows highlight overlapping cells expressing both - *foxd3-mCherry* and *Enh6-GFP* reporters. Scale bars correspond to 20μm.

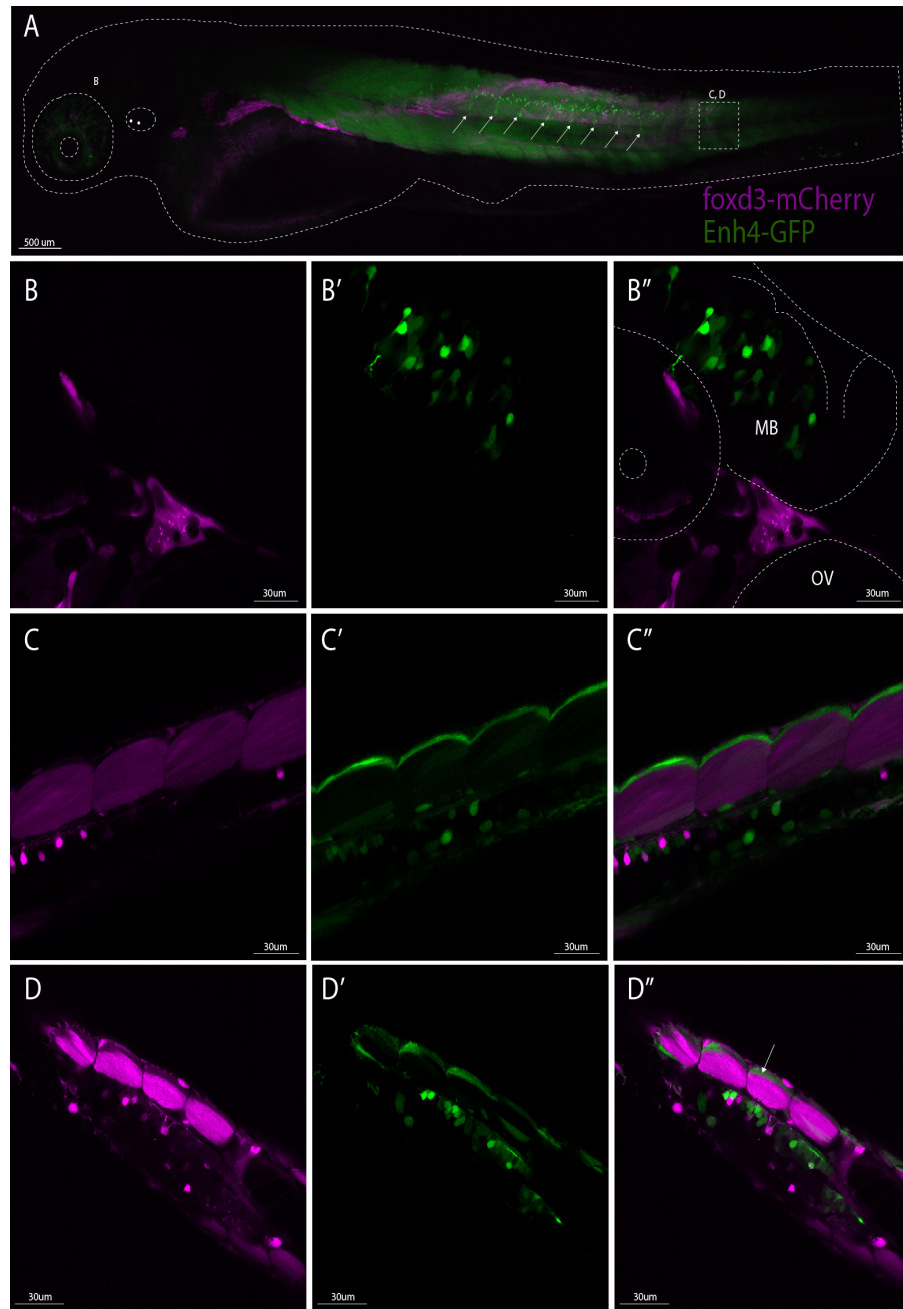


Figure 3.23 *Tg(FoxD3:enh4-GFP)^{ox110}* line crossed to *Gt(foxd3-mcherry)^{ct110aR}* line: whole-mount live larvae imaging at 2 days post fertilisation (dpf). (A) Lateral view of an entire zebrafish larva at 2dpf expressing *foxd3-mcherry* (coloured in magenta) and *Enh4-GFP* (in green). Letters B, C, D correspond to approximate positions within a zebrafish of the sections listed below (note, sections below were taken in a different larvae). White arrows point to phenotypically apparent evenly spaced neurons. Scale bar corresponds to 500µm. (B), (C), (D) *foxd3* expression as seen from *Gt(foxd3-mcherry)^{ct110aR}* reporter line fluorescence (coloured in magenta) across different sections - (B) - eye/midbrain (MB) region above the otic vesicle (OV), (C-D) - tail. (B'), (C'), (D') *Enh4* activity as seen from *Tg(FoxD3:enh4-GFP)^{ox109}* line fluorescence across corresponding sections as above. (B''), (C''), (D'') Image overlays of (B) - (B'), (C) - (C') and (D) - (D') images. White arrow in D'' points to mutually exclusive *Enh4-GFP* and *foxd3-mCherry* reporter expression in muscle. Scale bars correspond to 30µm.

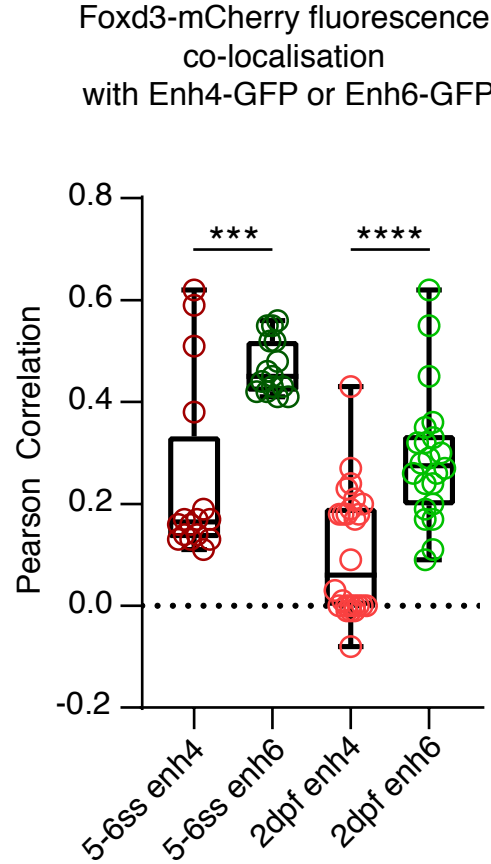


Figure 3.24 Enh6-GFP or Enh4-GFP co-localisation with *foxd3*-mCherry fluorescence at 5-6 somite and 2 days post fertilisation stages. Box plot with all data points showing GFP and mCherry pixel co-localisation between 40x optical sections of 5-6ss and 2dpf zebrafish offspring originated from *Gt(foxd3-mcherry)^{ct110aR}* line and either *Tg(FoxD3:enh6-GFP)^{ox109}* or *Tg(FoxD3:enh4-GFP)^{ox110}* line incrossing. Pearson correlation was acquired using Fiji Coloc2 plug-in, set for a bisection threshold regression calculation for compared images. Stars indicate significance levels (**p=0.0001, **** p<0.0001) comparing mean variances between Enh4-GFP and *foxd3*-mCherry fluorescence co-localisation against Enh6-GFP and *foxd3*-mCherry fluorescence co-localisation at 5-6ss and 2dpf using Welch's *t*-test.

tive origin of mesoderm, further supporting the idea that Enh4+ cells are another source of both mesodermal and neuronal cells in a developing embryo. Certainly, this needs to be further supported and validated by additional transcriptomics, epigenomics and imaging evidence. However, given our current results and hypotheses, it appears that the underlying negative Enh4 and *foxd3* feedback loop is present in a subset of *foxd3*+ cells. It likely ensures *foxd3* downregulation to minimum expression levels, which could be important for a putative zebrafish NMp cell lineage emergence.

3.9 Predicted Enh4 activating and repressing transcription factors

Thus far we have identified an apparent negative auto-regulation between Enh4 and *foxd3*. In NC cells, where *foxd3* is abundantly expressed and is regulated by Cl. 3.1 *cis*-regulatory elements, Enh4 is kept repressed upon a direct binding of *foxd3*. Meanwhile, we have also identified a pool of cells, which maintains Enh4 in an active state that in turn results to downregulation of *foxd3* expression to very low or absent levels. Interestingly, in the majority of these cells, Cl. 3.1 elements, within the *foxd3* locus, were not found to be open (Figure 3.19). However, two different putative *cis*-regulatory elements, one of which is Enh4, are found to be open upstream of the *foxd3* promoter, that are not accessible in the NC cells (Figure 3.19). This suggests a divergent *foxd3* locus regulation in different cell types. We have also proven that Enh4+ pool of cells is not of NC origin and rather gives rise to both, mesodermal and neuronal cell fates, possibly hinting to a neuromesodermal progenitor origin. Hence, conceivably, Enh4-*foxd3* negative auto-regulatory feedback loop is needed to tightly regulate *foxd3* expression that directly or indirectly leads to NC/NMp lineage fork split early in development.

In order to understand how Enh4 is regulated, giving us insights into how it is activated during NMp development and repressed in NC cells, we revisited our TFBS analysis (see Figures 3.3 and 3.4). Our Hocomoco and JASPAR results revealed that Enh4 DNA sequence is predicted to be bound by *lhx*, *zeb*, *snai2*, *tcf*, *nkx*, *egr*, *esr*, *klf*, *maf*, *ror*, *rar*, *rxr*, *sall*, *tbx*, *runx*, *zic* family TFs. Presumably, some of the predicted TFs would activate or keep it activated; while others would do the opposite - repress or keep Enh4 repressed, possibly together with *foxd3*. To shed light on the Enh4 activation vs repression, we integrated our RNA-seq results from the Enh4+ and Enh6+ cells. We have shown that Enh6+ cells are found in the anterior parts, while Enh4+ cells - at the posterior parts of a 5-6ss embryo. Moreover, Enh6+ cells were expressing *foxd3*, a repressor of Enh4, at higher levels than Enh4+ cells. Additionally, Enh4 element was not accessible in Enh6+ cells. Taken all of this together, we can argue that Enh4 is repressed in Enh6+ cells. Therefore, TFs that were highly expressed in the Enh6+ and not Enh4+ cells, but were enriched in the Enh4 TFBS analysis, are very likely to be the repressors of Enh4, at least

3 Auto-regulation of the *foxd3* genetic locus during cell lineage specification

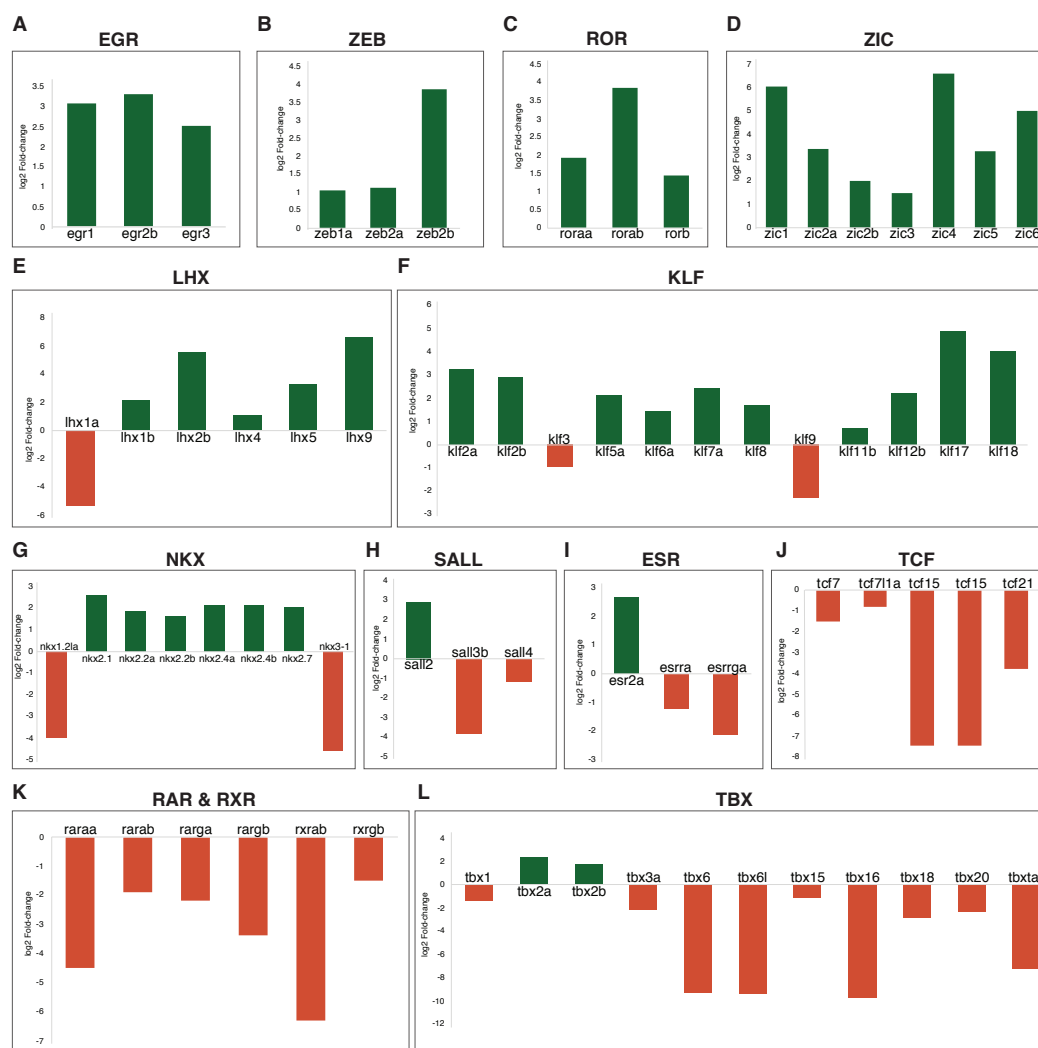


Figure 3.25 Enh4 activators and repressors. Figure highlights the most significant factors uncovered by the TFBS analysis underlying Enh4 DNA sequence. Listed bar plots **A-L** depict log₂ fold change of differentially expressed identified factors in the Enh6+ vs Enh4+ cells. The TFs, which were upregulated in the Enh6+ cells are marked in green, while the TFs, depleted in the Enh6+ cells and hence upregulated in the Enh4+ cells, are marked in red.

in the Enh6+ cells. On the other hand, those factors that were enriched for the Enh4 TFBSs and were found to be upregulated in the Enh4+ cells in comparison to the Enh6+ cells, could be the TFs that activate or keep Enh4 activated. This way, we uncovered that *egr*, *zeb*, *ror*, *zic*, *lhx*, *klf* and *nkx* family members, listed in the Figure 3.25 A-G, are the candidates of Enh4 repression in the NC cells. Intriguingly, our genome-wide analysis described in the previous chapter, has also identified *zic* family members as putative co-factors needed for the *foxd3*-mediated gene repression (See Figure 2.20 A). Contrarily,

sall, esr, tcf, tbx and RA receptors are the most likely Enh4 activators (see Figure 3.25 H-L).

3.10 Discussion

3.10.1 Foxd3 is involved in positive and negative auto-regulatory loops during cell fate decisions

This chapter of my DPhil thesis has revealed that in addition to a bi-modal action of foxd3 during the NC ontogeny, its own upstream regulation is perplexing and involves both, positive and negative, transcriptional auto-regulatory loops. To demonstrate that, we have employed two transgenic reporter zebrafish lines, expressing GFP fluorophore under a control of two different putative foxd3 enhancers. One of these enhancers, Enh6, belongs to a cluster of previously defined NC enhancers (Cl. 3.1) that requires foxd3 for chromatin accessibility establishment or maintenance (see Section 2.5 and Figures 2.14 and 3.2). Another enhancer, Enh4, belongs to a different enhancer cluster (Cl. 3.2), the chromatin accessibility of which is not dependent on foxd3. Moreover, Cl. 3.2 elements associate to genes controlling neural fate programmes (see Figure 2.14 E). What is more, both putative foxd3 enhancers were found to be directly bound by foxd3 all the way from 75% epiboly to 14-16 somite stages (See Figure 3.1 A). However, while the accessibility of Enh6 is maintained upon foxd3 binding in foxd3+ cells, Enh4 gets compacted by 14-16ss in the same cells (see Figure 3.1 B). Same was proven by foxd3 overexpression experiments, where ectopic foxd3 increased a number of cells having open Enh6, but decreased a number of cells having accessible Enh4 at 50% epiboly (see Figure 3.2). Collectively, this argues that in the NC cells, foxd3 positively auto-regulates itself via a portion of Cl. 3.1 enhancers, such as Enh6. However, Enh4 is decommissioned in the NC cells, which might be important for the NC cell fate specification step.

Further imaging analysis of the embryos expressing Enh6 and Enh4 reporters exposed that Enh6+ cells reside in the cranial regions. On the other hand, Enh4 reporter was expressed in the caudal regions of an embryo at 5-6ss. This, together with transcriptomics analysis of Enh6+ and Enh4+ cells, led us to a conclusion that a portion of Enh6+ cells

constitute *foxd3*+ NC cells. Here, *Enh4* is kept inaccessible upon a direct binding of *foxd3* during its repressive mode.

We have also characterised the trunk *Enh4*+ cells, where *Enh4* is not repressed as it is in the cranial NC. Strikingly, we demonstrated that the NC *foxd3* enhancers (Cl. 3.1) are not accessible in the majority of *Enh4*+ cells (see Figure 3.19). Interestingly, we have identified another open putative *cis*-regulatory element, in addition to *Enh4*, upstream of the *foxd3* promoter in the *Enh4*+ cells. This suggests that *foxd3* is regulated via different set of *cis*-regulatory elements in the *Enh4*+ cells in comparison to the NC cells. Even though *Enh4*+ cells do express *foxd3* as was shown by our transcriptomics analysis, the levels of expressed *foxd3* are very low. Veritabily, we were not able to frequently see *foxd3*-mCherry and *Enh4*-GFP expression overlap in our live embryo images, especially at the larvae stages. Because of the observed anti-correlation of *Enh4* activity and *foxd3* expression, we initially had doubts whether *Enh4* really regulates the *foxd3* gene promoter. However, we noted only 8 genes (*pgm1*, *efcab7*, *si:dkey-148h10.5*, *alg6*, *atg4c*, *angptl3*, *usp1*, *mafaa*), excluding *foxd3*, when looking at 500kb each side from the *Enh4*. Genes *efcab7*, *si:dkey-148h10.5*, *alg6*, *atg4c*, *angptl3* and *usp1* were not expressed or were expressed at extremely low levels in the *Enh4*+ cells. *Mafaa* was expressed, but also at rather low levels (average < 80 normalised reads). *Pgm1* (phosphoglucosyltransferase-1 enzyme) was expressed, but was enriched in *Enh4*- and *Enh6*+ cells when compared to *Enh4*+ cells. We also confirmed that *Enh4* is not transcribed in NC or whole embryo cells at 16ss (data from [190], not shown). Given these observations and the fact that *foxd3* is the nearest expressed gene, *foxd3* still seemed as a likely candidate gene to be regulated by the *Enh4* element. Therefore, we propose that *Enh4* is driving *foxd3* expression in the *Enh4*+ cells, however *foxd3* is auto-repressing itself via this element, which could be important for lineage segregation and embryo patterning. As was mentioned before, similar auto-negative loop takes place during the fly compound eye retina patterning, where TF *Dve* is autorepressing itself via *yR7* enhancer in the inner PRs permitting colour-sensitive Rh expression. In the outer PRs, *yR7* enhancer is not active and different *cis*-regulatory elements control high *Dve* expression that leads to a repression of colour-

sensitive Rh transcription and subsequent motion-sensitive PR development [230].

Overall, our data suggests that an interplay between positive and negative *foxd3* auto-regulatory loops aids axial embryo patterning and cell fate decisions. We have showed that the anterior embryo regions, that give rise to cranial NC and neuroectoderm, express high levels of *foxd3*, where it positively auto-regulates itself via *Enh6*. However, at the posterior embryo regions, *Enh6* is not active and here *foxd3* gets downregulated to low levels via a negative auto-regulatory loop involving *Enh4* element. These cells do not get specified as NC cells but rather arise from what we believe are neuromesodermal progenitors. Our TFBS analysis has also identified putative key players, such as *egr*, *zic* and *klf* gene family members, as the *Enh4* repressors in the *Enh6*⁺ cells. This, as a result, prevents the negative *Enh4-foxd3* feedback from being initiated and hence allows NC specification and development in terms of *foxd3*-mediated regulation. Some of these factors are likely able to achieve this by co-operating with *foxd3* itself. Contrarily, we have pinpointed *RA*, *sall4*, *tcf* and *tbx* gene family members as possible *Enh4* activators. Notably, *Sall4* and *RA* was found to be implicated during the NMP ontology, offering a putative mechanism of how these factors induce our newly observed negative *foxd3* auto-regulatory feedback loop allowing *foxd3* downregulation for the NMP development [258, 259].

3.10.2 *Enh4* - a novel marker of the zebrafish neuromesodermal progenitors?

A decade ago a bipotent cell population residing at the caudal embryo regions, now called neuromesodermal progenitors, was discovered. It was shown to have self-renewal and bipotent capacity to differentiate into both, somite-forming paraxial mesodermal and spinal cord neural derivatives during the embryo axis elongation [242]. In the chick and mice late gastrula and tail bud stage embryos, these cells are localised to caudal lateral epiblast, adjacent node streak border and the chordoneural hinge [244, 260–264]. The discovery of these cells has challenged a traditional view arguing that the nervous system develops from neuroectoderm via the 'activation-transformation' mechanism [243]. This long-lived dogma postulated that anterior nervous system features get induced first at

cranial regions during gastrulation and later, caudal nervous system acquire posterior features only after anterior system-like induction [265, 266]. For the NMP induction, FGF (e.g. *Fgf4/8*), BMP and Wnt (e.g. *Wnt3a*, *Wnt8a*) signalling pathways were shown to be instrumental [266]. In addition, during early embryonic stages, the node and primitive streak cells also generate RA by RA-synthesising enzyme, *Aldh1a2*, that induces NMP cell population [259]. Later, Wnt expression, *Wnt3a*, for example, is important for mesoderm formation [247, 267]. This is achieved by Wnt and FGF signalling-mediated *Cdx* induction, which posteriorises NMPs by promoting posterior *hox* gene expression [268]. *Cdx* also induces expression of T/Brachyury (T/*bra*) that on a long term is important for mesoderm induction [269]. On the other hand, for the neural cell formation, *Sox2* expression needs to be upregulated and T/Brachyury expression needs to be down-regulated [252].

The GRN underlying NMPs and their differentiation into mesodermal and neural lineages was comprehensively presented in a study covering both, *in vivo* (mice) and *in vitro* experiments [251]. It revealed that early NMPs express T/*bra* (*tbxta*), *Sox2*, *Cdx2/4*, *Nkx1.2*, *Sp8*. Of note, *Cyp26a*, a RA-degrading enzyme, that is induced by *Cdx*, suppresses RA levels at the tailbud that, which as a result, together with canonical Wnt signalling, allows NMP maintenance [247, 251]. NMPs were also shown to express 5'Hox genes and other genes, such as *Wnt8a*, *Fst*, *Cdh1*, *Rxrg* at early embryonic (E8.5) mice stages. By E9.5 stage, more posterior Hox gene expression was identified alongside *Myo5b*, *Foxd1*, *Pde4d* and others. NMPs were also shown to transit towards presomitic mesoderm differentiation via mesodermal progenitor cells (MPCs). This happens by NMPs commencing *Msgn1* and *Tbx6* gene expression in addition to T/*Bra* and *Nkx1.2* genes. Subsequently, to adopt the PSM fate, MPCs need to downregulate T/*Bra* and *Nkx1.2* genes. Strikingly, the induced *Msgn1*+/*Tbx6*+ mesoderm then starts expressing *Aldh1a2* generating RA that feeds back onto the tailbud NMP pool, where T/*Bra* gets inhibited and *Sox2* expression gets upregulated leading to neural differentiation.

We have established transgenic *Tg(FoxD3:enh4-GFP)^{ox110}* zebrafish line harbouring an element (*Enh4*) that was shown to likely be a *foxd3* *cis*-regulatory element involved in

the *foxd3* auto-repression loop that leads to its downregulation in the *Enh4+* cells. Interestingly, we have observed that at early embryonic developmental stages, soon after the AP axis establishment (5-6ss), *Enh4*-GFP reporter is expressed in the caudal embryo regions. To further characterise the *Tg(FoxD3:enh4-GFP)^{ox110}* line and understand what *Enh4+* cells might be, we performed bulk RNA-seq for the initial insights. This way, we have identified genes involved at all described NMP developmental stages, suggesting that *Enh4+* cells might contain NMps, be NMP derivatives or both. For example, we have shown that *Enh4+* cells express *sox2*, *tbxta* (brachyury), *cdx4*, *wnt8a*, *nkx1.2la* and posterior *hox* genes that correspond to the NMP pool. We have also noted *cyp26a1* RNA enrichment in the *Enh4+* cells, suggesting that the NMP pool might be maintained via RA degradation by this enzyme in the zebrafish, similarly to what was observed in mice and in *in vitro* NMps. However, so far, it was shown that RA is dispensable for zebrafish axis elongation and somitogenesis [270]. Again, given that anamniotes generate their head and trunk due to convergence extension, RA might not be critical for the early NMP induction that quickly segregates into neuronal and mesodermal lineages [246, 248]. However, the tailbud NMP pool that is thought to be not self-renewing and rather being only depleted during zebrafish embryogenesis, might, at least partially, follow the rules governing NMP differentiation into neuronal versus mesodermal fates described in amniotes [248, 251, 252]. Remarkably, our transcriptomics analysis has uncovered substantial mesodermal gene expression, such as *tbxta*, *cdx4*, *msgn1* and *tbx6*, involved during NMP to MPC to presomitic mesoderm transitions. This indicates that *Enh4+* cells, residing in the zebrafish posterior at 5-6ss, possibly comprise NMps, MPCs and PSM. Interestingly, we have also noted high *aldh1a2* (average 17,286 read counts in *Enh4+* cells normalised to *Enh4-* counts) and RA pathway gene expression in our *Enh4+* dataset (see Figure 3.25 K). This could be indicative of a specified *msgn1*+/*tbx6*+ mesoderm producing RA that feeds back onto the NMP pool to downregulate *tbxta* and induce *sox2* expression resulting into neuronal programme initiation as was described in Gouti's M. et. al. study [251]. Indeed, we were able to observe neuronal gene expression, including *sox2*, in our *Enh4+* RNA-seq dataset, but the overall balance was significantly outweighed by mesodermal

gene expression. Notably, our 2dpf Enh4-GFP zebrafish imaging demonstrated, that Enh4 reporter is active in both, muscle and neural cells (Figure 3.23). Taken together, we hypothesise a possibility, where similarly to mice embryos, zebrafish also maintains a tailbud NMP pool via RA downregulation, that first gives rise to PSM, which then initiates RA production that then leads non-autonomously induced neuronal cell fate within a NMP pool. Moreover, similarly to mice, *msgn1* was also shown to be critical for zebrafish mesoderm differentiation. Strikingly, *msgn1* overexpression led to down-regulation of canonical *wnt8*, which might be important for the NMP pool maintenance [251, 271]. Our results demonstrated that Enh4+ cells express high levels of Wnt8a (see Figure 3.15 C) hinting to another important and conserved factor for the zebrafish tailbud NMP population regulation.

Importantly, we have also performed single cell RNA-seq on *foxd3*+ cells coming from all axial embryo levels, in order to survey distinctive cells populations residing along the AP axis in unbiased manner. Remarkably, we were able to clearly discriminate canonical cranial NC cell populations from the trunk cells, which expressed same genes as the Enh4+ cells. In other words, we identified a caudal cell population, which exhibited NMP and their early derivative transcriptional signatures, suggesting that they correspond to the Enh4+ cell population. This way, we were able to further support our hypotheses, which were stated above. We have identified a small pool of naïve NMP cells, which co-expressed *cdx4*, *sox2*, *tbxta* and RA degrading enzyme, *cyp26a1*, which is possibly important for this pool maintenance (see Figures 3.16 and 3.17) at 5-6ss. Single cell data has also supported a pronounced mesodermal cell population, which expressed *cdx4*, *tbxta*, *tbx6l*, *msgn1* that corresponds to MPCs (Figure 3.17). Similarly, we also identified definitive mesodermal cells, which expressed high levels of *myod1* and *myl10* (Figure 3.17), but low or absent levels of *tbxta*. Importantly, some of the apparent MPCs/mesodermal cells were also expressing RA synthesising enzyme, *aldh1a2* (Figure 3.16). This supports our hypothesis that newly formed mesoderm is signalling back to the NMP pool resulting in neuronal programme induction, as considered above. Indeed, at 5-6ss we did not observe many cells committing to neuronal fates, and we only noted some of the cells co-

expressing *tbxta*, *sox2*, *msgn1*, *fgf8* and strangely, NC markers, such as *sox10* and *pax3a* (see Figures 3.17 and 3.18). Therefore, it will be important to understand, whether the resident naïve NMPs give rise to neurons at later developmental stages, upon receiving signals from induced mesoderm, or whether they rather stay dormant, as was shown in the previous study looking at NMP development in zebrafish [248]. Notably, our imaging analysis of *Tg(FoxD3:enh4-GFP)^{ox110}* zebrafish larvae at 2dpf clearly indicated presence of both, mesodermal and neuronal derivatives, suggesting that the identified zebrafish NMP pool is capable of generating nerve cells at some point of the NMP ontogeny.

3.10.3 Stable reporter lines to understand anterior and posterior nervous system development

We have also generated a *Tg(FoxD3:enh6-GFP)^{ox109}* zebrafish line, which oppositely to the *Tg(FoxD3:enh4-GFP)^{ox110}* enhancer reporter line, drove GFP fluorescence in the cranial regions of a developing zebrafish embryo at 5-6ss. From the imaging, transcriptomics and chromatin accessibility analysis, it was apparent that a portion of Enh6+ cells overlapped the *foxd3*+ cranial NC cells and later its derivatives. However, we have also observed that the accessibility and gene expression signatures of this enhancer associated with the CNS, such as forebrain or midbrain-hindbrain barrier, development. Moreover, imaging the Enh6 reporter line at 2dpf further supported these findings by elucidating Enh6+ element being active in the larvae CNS, for example, midbrain and hindbrain (see Figure 3.22). For a long time, the nervous system induction doctrine postulated that neuroectodermal cells at the embryo anterior are destined to differentiate into the CNS and some of these induced cells then give rise to the posterior nervous system by a subsequent acquisition of a more posterior identity [265, 266]. However, recently this doctrine was updated by new results showing that the regionalisation of the AP axis takes place first, upon which these posteriorised cells remodel their chromatin distinctively from the anterior cells during nervous system induction, suggesting distinctive origins and induction pathways of the neural cells along the AP axis [272]. Therefore, the nervous system induction and development in vertebrates needs to be revisited, especially in the light

of neural progenitor (NP) regionalisation. In this DPhil thesis, we have presented two transgenic zebrafish lines reporting activity of different DNA elements in different cells types, one of which is suggestive of giving rise to the CNS and NC-derived PNS at the anterior embryo regions, while the other one generating spinal cord neurons at the caudal embryo regions, possibly marking NMps. For instance, we have shown some of the FGF, Wnt and BMP signalling pathway molecules, known to be involved in NP induction, to be differently expressed between the two lines (Figure 3.15). We have uncovered an enriched BMP2/4/7 expression in the Enh4+ cells in comparison to Enh6+ cells, which have been previously shown to be important for the trunk development and their associated pathway perturbation led to a development of enlarged brain [273, 274]. On the other hand, BMP6, which was enriched in the Enh6+ cells, is known to be important for the CNS development and functioning [275]. Therefore, these two enhancer reporter fish lines can be exploited and adapted as suitable tools to tackle principles and evolutionary conservation questions underlying NP induction and development along the AP axis.

3.10.4 Summary

This chapter presents evidence pointing to underlying *foxd3* auto-regulation, including both, positive and negative feedback loops. It is achieved by *foxd3* auto-regulating itself via different *cis*-regulatory elements. One of these elements follows the NC-regulatory logic, resulting in high *foxd3* expression that is crucial for the NC lineage induction and development. On the other hand, the auto-repressive element, needs to be silenced in the NC cells to prevent the negative feedback loop from commencing. However, it gets activated by different signals, that are likely to play a role in the NMp induction, in the caudal embryo regions. This, as a result, leads to *foxd3* downregulation via the initiated negative auto-regulatory loop. Whether this observed negative feedback loop is crucial for the NMp induction or development remains to be investigated in future.

Conclusions and Future Directions

For centuries, embryonic development was found to be perplexing, mysterious and simply astonishing. Naturally, in order to begin understanding the complex principles behind this imperfectly perfect process, we tried to compartmentalise ideas, reduce complexity, draw explicit conclusions, make up hierarchies. Undeniably, we were also limited by the available tools and technologies. Thus now we are finally starting to turn towards mathematics, biophysics, biomechanics and kinetics. We switch back and forth between reductionist and holistic approaches; this way relentlessly rediscovering new ways of sculpting tissues, understanding new cell functions, uncovering enormous intra- and inter-cellular heterogeneity, and appreciating vast interconnected complexity within biological components governed by biophysical chemical principles.

My DPhil thesis investigated an upstream and downstream regulation of a stem-cell and neural crest factor, *foxd3*, during zebrafish embryonic development. We have demonstrated that during very early NC stages, *foxd3* regulates itself via positive and negative auto-regulatory loops to either increase or decrease its own expression, respectively. We have followed the data in an unbiased fashion to discover how this differential *foxd3* auto-regulation translates into driving (or being a consequence of) cellular transitions. Surprisingly, we have revealed, that *foxd3*, via its *cis*-regulatory element, auto-represses itself in response to other signals, such as retinoic acid signalling, giving rise to neuromesodermal progenitors. This bipotential cell population was already substantially characterised and described in mice and chick embryos, and applying the accrued knowledge has allowed researchers to generate mouse and human NMps *in vitro* [153, 276, 276]. However, our grasp on zebrafish NMP population is still rather limited [247, 248]. Therefore, understating NMP population in anamniotes and how it compares to the amniote NMP

population, holds important answers to questions about NMp evolution and ultimately underlying cellular and organismal functions. Our work in progress, is advocating that zebrafish NMp population is governed by very akin gene regulatory principles previously described in mice embryos [251, 277, 277]. Importantly, our results are also suggesting a close relationship between NMps and posterior NC progenitors, possibly governed by the same bi-modal transcription factor, *foxd3*. We have shown that abundance levels of *foxd3* might be critical for a NMp/NC progenitor fate restriction towards NMps or posterior/sacral NC. We have demonstrated that NMps express very low levels of *foxd3* in comparison to canonical cranial NC cells. By using unbiased single cell RNA-seq characterisation of the Citrine-positive cells isolated from the *Gt(foxd3-Citrine)ct110a* transgenic embryos, we were able to confirm presence of the NMp pool and its early derivatives from the trunk embryo regions, corresponding to the posterior *foxd3* enhancer (Enh4) positive cells. It will be very informative to compare this control wild type *foxd3* dataset to the single cell transcriptome coming from *foxd3*-mutant cells issued from crosses of *Gt(foxd3-Citrine)ct110a* and *Gt(foxd3-mcherry)ct110aR* transgenic lines. We are poised to address questions, such as, whether *foxd3* is needed for the generation of a putative common NC/NMp progenitor pool, as well as flexibility and regulatory plasticity of the transcriptional switch upon perturbation. For instance, lack of *foxd3* could lead to decrease or elimination of both cell types or conversely, the absence of *foxd3* could cause the expansion of the NMp pool at the expense of NC progenitor generation. Similarly, we would also need to validate the observed *foxd3*-Enh4 auto-repressive loop and unravel whether this auto-repression is important for the NMp lineage formation or is just a consequence of a cellular transition. The idea of a common NC/NMp progenitor was also recently suggested by an *in vitro* study demonstrating that human pluripotent stem cells induced for different axial progenitors, encompass a trunk-like population of cells with mixed early NC/NMp transcriptional signatures [153]. Moreover, cranial and trunk NC exhibit quite different properties and generate distinctive derivatives. Particularly, cranial NC bias for mesenchymal fates results in facial cartilage and bone formation, whereas trunk NC displays a bias for neuronal fates [116]. Furthermore, efforts on trying

to re-program the trunk NC towards cranial NC were only partially successful, with only transient and not definitive mesenchymal property activation in the reprogrammed trunk regions [151]. This poses a question, whether trunk NC has a distinctive origin to that of a cranial NC, and whether it is rather generated from a common NMP/NC progenitor. Therefore, further studies using *Tg(FoxD3:enh4-GFP)^{ox110}* and *foxd3*-mutant gene trap lines, will allow us to prove or disprove this NMP/NC precursor pool hypothesis. Moreover, integrating chromatin, bulk and single cell RNA-seq data sets across different stages of zebrafish development, will also allow us to build genetic circuits underlying putative NC/NMP progenitor establishment and the lineage split, as well as, to start building a regulatory framework underlying the NMP ontogeny in zebrafish.

On the other hand, in a different line of investigation, we have provided a strong evidence that *foxd3* was a critical NC induction and specification regulator. During the NC formation, *foxd3* is regulated by a set of *cis*-regulatory elements, that follow a NC regulatory logic. Importantly, some of these NC elements are positively auto-regulated by *foxd3* itself. We showed that first, directly and indirectly, *foxd3* is activating the NC programme by remodelling underlying NC gene chromatin. In the absence of *foxd3*, as we have identified by the *foxd3* mutant analysis, *cis*-regulatory elements of the *bona fide* NC genes are found to be less accessible resulting in significantly decreased expression of the associated genes. Notably, we have predicted and supported previous observations that *foxd3* might be working together with other chromatin co-pioneer factors, such as *tfap2a* [115]. We have also uncovered a putative *foxd3* and zinc finger protein interaction, which possibly recruits chromatin remodeller, Brg1, via its N-terminal zinc-finger protein recognition domain [166, 205]. Remarkably, FoxD3 was first described as a transcriptional repressor [177]. In line with this, we have also demonstrated that upon *foxd3*-mediated NC specification, *foxd3* gradually assumes such repressive role. It represses transcription, again via chromatin remodelling mechanisms, of factors driving NC differentiation into mesenchymal and neuronal fates. Hence, we hypothesise, that this way, *foxd3* maintains multipotency of the specified NC and prevents its premature differentiation into derivatives. To achieve this, *foxd3* is again likely to interact with context-specific partners, i.e.,

other co-repressors. We have uncovered that DNA elements, which are directly repressed by foxd3 during late NC stages, are also very highly enriched for e2f6 motifs suggesting e2f6 protein co-binding. This factor (e2f6) is a transcriptional repressor known to interact with repressive polycomb complexes (PRC1/PRC2) [199, 209, 210]. Our work thus offers insights into a potential mechanism by which polycomb complex proteins could be targeted to DNA and silence underlying chromatin in a foxd3/e2f6 dependent manner. Certainly, these *in silico* predicted interactions will need to be validated in future, for instance, by employing synthetic TF pair co-binding approaches, such as, consecutive affinity-purification systematic evolution of ligands by exponential enrichment (CAP-SELEX) [278], *in vivo* NC-targeted e2f6 perturbations or by co-immunoprecipitation assays followed by Mass spectrometry. This way, we will be able to identify which factors co-bind with foxd3 *in vivo* and also determine their heterodimeric DNA binding motifs *in silico*. Subsequently, we will be able to annotate the uncovered two-factor motifs back to our current datasets in order to comprehend which genes are controlled by which foxd3-heterodimeric TF pairs during its activation or repression mode. Most importantly, we will need to investigate whether either ectopically forced or perturbed anticipated interactions, will have a functional effect on NC induction and the maintenance of NC multipotency.

Furthermore, we have also observed potential foxd3-DNA binding differences between its chromatin pioneering and repression states. While we were able to uncover foxd3 canonical DNA binding motifs using our biotin ChIP-seq data from the late NC stages, our endeavours to do the same during its pioneering state were less successful. This advocates that different binding modes and motif affinities are involved in foxd3 binding to 'naked' DNA (during repressive activities) versus nucleosomal DNA (during pioneering activities). This is in line with other fox proteins, such as FoxN3, displaying ability to bind to two very different motifs that are determined by DNA shape [279]. Moreover, pioneer TFs were shown multiple times to occupy only partial canonical or lower affinity DNA binding motifs, when they bind to nucleosomal DNA [88, 89]. Therefore, it will be necessary to confirm that foxd3 also binds to alternate DNA binding motifs on

linear and nucleosomal DNA. To uncover alternate motifs, we will again employ synthetic approaches, such as Nucleosome CAP-SELEX, to compare foxd3 binding to linear stretches of DNA and chromatinised DNA [280]. We are also optimising the Cleavage Under Targets and Release Using Nuclease (CUT&RUN) method developed by Steven Henikoff's group for zebrafish embryos [281]. We will then apply the newly developed enhanced chromatin occupancy (EChO) algorithm to further identify either direct foxd3-DNA binding or foxd3-nucleosomal DNA binding at different DNA elements and during different NC developmental stages from an *in vivo* context [89].

My doctoral thesis work has revealed an immense complexity underlying upstream and downstream regulation of a critical NC transcription factor, foxd3, during early embryonic development. Particularly, we uncovered opposing modes by which a single factor might act during different cellular transitions, to enable cellular switch between NC progenitor and derivative fates. However, even though by exploiting versatile zebrafish genetics tools and *-omics* approaches, I have, at least partially, answered numerous questions, I was left with even more outstanding ones. Yet, I do not consider this ending to be resentful. Veritably, I do not consider this being an end in a first place. It is curiosity that continuously drives scientists to dig deeper, to understand a little bit more, to discover what was previously not known. If we run out of questions, we run out of mysteries and science all together. Inevitably, science must go on.

Materials and Methods

5.1 General methods

5.1.1 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 [282]. Embryos were staged as described previously [283]. 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.

5.1.2 Zebrafish embryo microinjections

In order to achieve high efficiency of germline integration of an injected plasmid carrying a transgene, all microinjections were carried into a blastomere of one-cell stage embryos within 15 minutes of egg laying. Phenol-red was always incorporated into a final injection mix for microinjection visualisation. Specific details for different experiments are described later in this chapter.

5.1.3 Cell dissociation from zebrafish embryos and FAC-sorting

Selected embryos were dissociated with collagenase (20mg/ml in 0.05% trypsin, 0.53nM EDTA in 1X HBSS) in 300µl volume at 30°C for 5-8 mins with intermittent pipetting to achieve a single cell suspension. The reaction was stopped by adding dissociations into 4ml of Hanks buffer (10mM Hepes (pH 8), 2.5mg/ml of Bovine serum albumin (BSA) in 1X HBSS) followed by gentle trituration with a glass serological pipette. Cells were centrifuged at 500g for 10mins and re-suspended in 5ml of Hanks buffer, passed through a 0.22µm filter and centrifuged at 750g for 10min, pelleted cells were resuspended in ~200-500µl of Hanks buffer. Dead cells were stained with 7-AAD (A1310, ThermoFisher Scientific) and were excluded from collections. Fluorescent positive or negative cells were sorted and collected using BD FACSAria II SORP, BD FACSARIA III or BD FACSAria Fusion2.

5.2 Chapter 2 Methods

5.2.1 *foxd3*-mutant vs control RNA-seq

Bulk RNA extraction, library preparation and sequencing

FAC sorted cells were washed with PBS and stored in lysis buffer (RNAqueous-Micro Total RNA Isolation Kit, Ambion, AM1931) at -80°C. RNA was extracted using Ambion RNAqueous Micro Total RNA isolation kit and checked on Bioanalyser. Samples with RIN > 7 were used to prepare cDNA using Takara Clontech SmartSeq2 V4 kit (634889). Sequencing libraries were prepared using Illumina Nextera XT library preparation kit (FC-131-1024). 75% epiboly-stage cell libraries (Citrine-expressing and Citrine-Cherry-expressing) were sequenced using 80 bp reads using Illumina Nextseq500 platform. 5-6ss cell libraries expressing *foxd3* (Citrine-expressing, Citrine-Cherry-expressing) were sequenced using 50bp reads using Illumina Hiseq2000 platform. 14ss cell libraries (Citrine-expressing and Citrine-Cherry-expressing) were sequenced using 80bp reads using Illumina NextSeq500 platform. All sequenced libraries were paired-end.

RNA-seq bioinformatics processing

Reads were trimmed to remove low quality bases using sickle tool when necessary. Read quality was evaluated using FastQC (Babraham). Mapping to GRCz10/danRer10 assembly of the zebrafish genome downloaded from UCSC Genome Browser was performed using STAR2.4.2a.(2) [284]. Read counts were obtained using subread FeatureCounts(v1.4.6-p4) [285] with standard parameters using a gene model gtf derived from Ensembl annotation downloaded from UCSC genome browser. Gene model for ENSDARG00000095311 (the antisense transcript of FoxD3), was removed from gene models. Differential Expression analysis was carried out using in DESeq2(v.1.14.1) or (v.1.18.1) [286].

5.2.2 foxd3 Biotin ChIP-seq

Generation of Avi-tagged foxd3 transgenic line

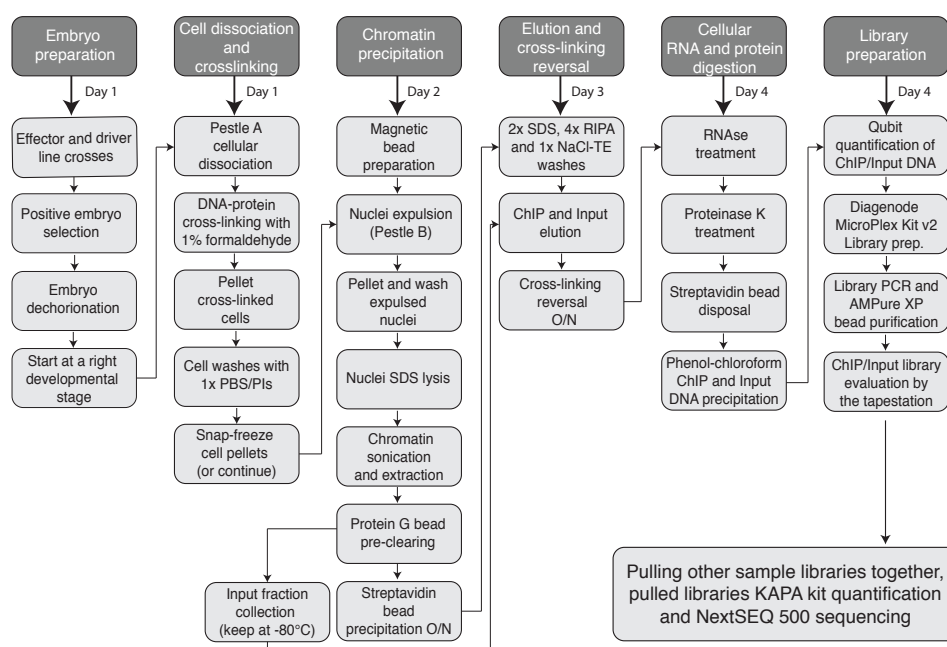
Tol2-mediated BAC transgenesis, as described in [190, 191], was used to generate TgBAC(foxd3-Avi-2A-Citrine)^{ox161} transgenic line. pGEM Avi-2A-Citrine-SV40pA-FRT-Kan-FRT recombination donor construct was generated by amplifying Avi-2A-Citrine cassette using Pfu polymerase (Pfu UltraII Hoststart PCR Master Mix, Agilent Technologies) and cloning it into the donor plasmid (#89890, Addgene) using InFusion kit (InFusion HD Cloning kit, Clontech). The full donor cassette contains a FLAG epitope, a TEV protease recognition sequence, an in-frame 48 bp Avi-Tag, a *Citrine* reporter followed by a polyA tail and a Kanamycin selection cassette flanked by flippase recognition target (FRT) sequences (that was flipped-out after selection and before zebrafish microinjections). *Citrine* reporter is separated from the *foxd3*-Avi gene by a viral linker, 2A, sequence that mediates ribosome skipping, thus allowing for co-expression of both components from a single transcript [287]. Genomic context of the *Danio rerio* BAC clone CH211-196F13 (203kbp) was used for recombineering, as it harbours the full single exon ORF of the *foxd3* gene and the upstream regions (>200kb), thus encompassing not only the *foxd3* promoter, but also associated *cis*-regulatory elements. The *foxd3* gene within the BAC was fused to the Avi-tag producing C-terminally Avi-tagged

foxd3 (foxd3-Avi-2A-Citrine) expressed in endogenous-like fashion. The final BAC (200 pg) and Tol2 RNA (40pg, *in vitro* transcribed using mMACHINE SP6 Transcription Kit, AM1340) were injected into Gt(foxd3-mcherry)^{ct110aR} heterozygous *foxd3*-mutant background to avoid foxd3 overexpression. F0 embryos were raised and screened for founders. Positive F1 embryos were raised for future experiments.

NLS-BirA zebrafish transgenic line *Tg(ubiq:NLS-BirA-2A-Cherry)^{ox114}* expresses 3X HA epitope, nuclear localisation signal (NLS) sequence fused to BirA, viral 2A sequence and a *Citrine* reporter gene under the control of ubiquitous *ubb* promoter.

FoxD3 Biotin-ChIP, library preparation and sequencing

Foxd3 Biotin-ChIP was performed on 700 for 75% epiboly, 350 for 1-2ss, 320 5-6ss experimental and BirA-only, 390 for 14ss whole embryos (~128,000 of interest) were used for a corresponding stage foxd3 Biotin-ChIP. Embryos were manually dechorionated, cells were dissociated with 20 strokes using pestle A in isotonic nuclei extraction buffer (NEB: 0.5% NP40, 0.25% Triton X, 10 mM Tris-HCl (pH 7.5), 3 mM CaCl₂, 0.25 M sucrose, 1mM DTT, 0.2 mM PMSF, 1X Proteinase inhibitor (PI) in a glass homogeniser and cross-linked using 1% formaldehyde at room temperature for 10 min. Fixation was quenched with 125 mM of glycine for 5min, cross-linker was washed-out by 3X pellet washes with 1X PBS (with 1X PI, 1 mM DTT and 0.2 mM PMSF) centrifuging at 2000g for 4min at 4°C. Pellets were re-suspended in NEB. Cell nuclei were expelled with 20 strokes using pestle B in a glass homogeniser, pelleted and washed with 1X PBS (with 1X PI, 1 mM DTT and 0.2 mM PMSF). Nuclei were lysed in SDS lysis buffer (0.7% SDS, 10mM EDTA, 50 mM Tris-HCl (pH 7.5), 1X PI). Cross-linked chromatin was sonicated at 12A 10X (10s ON 30s OFF) followed by 8A 4X (30s ON 30s OFF). Sheared chromatin samples were pre-cleared in pre-blocked Protein G beads (Dynabeads Protein G, Life Technologies, 10004D) for 1 hour at 4°C. 1/20 of biotinChIP was collected as an input fraction and stored at -80°C. Pre-cleared chromatin samples were incubated on pre-blocked streptavidin beads (Dynabeads M-280 streptavidin beads, Life Technologies, 11205D) o/n at 4°C. Beads were washed with SDS Wash Buffer (2% SDS, 10mM Tris-



FoxD3 Biotin-ChIP bioinformatics processing

Reads were trimmed for quality using sickle when necessary and mapped using bowtie (v.1.0.0). Duplicates were removed using MarkDuplicates (picard-tools/1.83). Input reads were normalised to the same number of ChIP reads by random down-sampling using samtools-1.1 (BirA - 18,545,346, 75%ep - 28,663,648, 1-2ss - 15,989,800, 5-6ss - 24,270,738, 14ss - 32,322,010 unique reads). Bigwig files were generated using an enhanced Perl script courtesy of Prof Jim Hughes. Peak calling was performed using Homer (v.4.7) [288] findPeaks script using -size 200-minDist 1500 parameters. Peaks were annotated to a nearest expressed gene at a given developmental stage in NC cell population. GO analysis was performed on acquired gene lists in pantherdb.org using statistical overrepresentation binomial test for complete biological processes. Motif discovery and characterisation was performed using Homer screening for *de novo* motifs within given foxd3 Biotin-ChIP-seq peaks. Significantly enriched motifs were annotated manually based on Homer results and levels of gene expression in NC at a corresponding developmental stage. All possible combinations of two motifs were computed using a custom R (v. 3.2.1) script (courtesy of Ivan Candido-Ferreira). Homer/4.7 annotatePeaks.pl script was utilised to screen all *de novo* motifs in foxd3 Biotin-ChIP-seq peaks co-occurring in windows of 500bps centred around peaks. A custom Python3 script using the Pandas package (courtesy of Ivan Candido-Ferreira) was used to calculate the frequency of 2-way motif combinations within foxd3-bound ChIP peaks. Combinations enriched at $\chi^2 P < 5 \times 10^{-3}$ with FDR correction for multiple hypothesis (vs testing co-motif enrichment frequencies against random DNA sequence background) were retained. Two-way networks were plotted using the 'Circlize' package in R.

5.2.3 foxd3-mutant vs control ATAC-seq

ATAC, library preparation and sequencing

FAC sorted cells were lysed (10mM Tris-HCl, pH7.4, 10mM NaCl, 3mM MgCl₂, 0.1% Igepal) and tagmented using Nextera DNA kit (Illumina FC-121-1030) for 30min at 37°C. Tagmented DNA was amplified using NEB Next High-Fidelity 2X PCR Master

Mix for 11 cycles. Tagmentation efficiency and final libraries were assessed using Agilent High Sensitivity TapeStation D1000 ScreenTape (5067- 5584) and Reagents (5067- 5585). They were sequenced using NextSeq 500/550 High Output Kit v2 (75 cycles, paired-end 40 bp run) on NextSeq 500 sequencing system.

***foxd3*-mutant vs control ATAC-seq bioinformatics processing**

Reads were trimmed for quality using sickle when necessary and mapped using Bowtie (v.1.0.0). Bigwig files were generated using an enhanced Perl script (courtesy of Prof Jim Hughes). Peak calling was performed as described previously [34]. Briefly, BAM files were sorted by name and paired end bed files were obtained using bedtools(v.2.15.0) bamtobed -bedpe. Reads that were not properly paired were discarded and paired reads were displaced by +4 bp and -5 bp for transposition correction. Reads were extended to a read length of 100bp. Peak calling was performed using MACS2 callpeak -f BED -shiftsize=100 -nomodel -slocal 1000 parameters [289]. To obtain mappable data, a synthetic 40bp-long single end fastq dataset was generated and mapped using bowtie(v.1.0.0) using -m 1 parameter. Bedgraph files were obtained using bedtools genomeCoverageBed -bg -split function. MACS2-called peaks that overlapped with regions which in the mappable did not correspond to read size (40bp) were discarded. Identification of peaks corresponding to TSS/promoter, intergenic, intronic and TES locations was carried out using Homer (v.4.8)[288] annotatePeaks.pl script. Only peaks present in both replicates were retained, using bedtools intersect function to generate reference ATAC-seq ensembles for each stage. *k*-means clustering of ATAC-seq signal was carried out using SeqMINR software as described [193], using non-promoter reference ATAC-seq peaks from 5-6ss samples.

5.2.4 Foxd3 ectopic expression assay followed by ATAC-seq

40pg of *foxd3* mRNA was injected into single cell wild-type fertilised embryos. 50% epiboly embryos were dissected to obtain cells that do not express *foxd3* intrinsically: ~12,000 cells were used per each experimental/control sample in triplicates. Cells were

dissociated with 0.05% trypsin to a single cell suspension, centrifuged and re-suspended in Hanks buffer. Cells were lysed as was described above (Section 5.2.3) and tagged using Nextera DNA kit (Illumina FC-121-1030) for 15 min at 37°C, reactions were quenched with 50mM EDTA for 30 minutes at 50°C. Tagmented fragments were purified using MinElute PCR Purification Kit (Qiagen, 28004). Tagmented DNA was amplified using NEB Next High-Fidelity 2X PCR Master Mix for 15 cycles. Tagmentation efficiency and final libraries were assessed using Agilent High Sensitivity TapeStation D1000 ScreenTape (5067- 5584) and Reagents (5067- 5585). They were sequenced using NextSeq 500/550 High Output Kit v2 (75 cycles, paired-end 40 bp run) on NextSeq 500 sequencing system.

Foxd3 over-expression ATAC-seq bioinformatics processing

Reads were processed as above (Section 5.2.3). Duplicated reads were removed using MarkDuplicates (picard-tools/1.83). All samples were randomly down-sampled to the lowest-read containing sample (10,443,726) using samtools-1.1. Processed experimental BAM files were merged together as well as control BAM files. *k*-means clustering of ATAC-seq signal was carried out using SeqMINR software as described [193], using 3.1 enhancer cluster ATAC-seq peaks and all pulled foxd3 Biotin-ChIP peaks as references for clustering. Averaged ATAC signal plots were generated using deepTools (v.2.2.2) on the selected *k*-means clusters.

5.2.5 *Ac/Ds*-mediated enhancer screening assay

All enhancer inserts were generated by amplifying genomic DNA using KAPA Long Range HotStart PCR kit (Kapa Biosystems) and cloned into the E1b:GFP:AC-DS vector (102417 Addgene, linearised with *NheI* using the InFusion kit (638909, Takara Bio). Fertilised single-cell embryos were injected with 30 pg of plasmid DNA and 25 pg of *in vitro* transcribed *Ac* mRNA (using mMESSAGE mMACHINE SP6 Transcription Kit, AM1340 on a *Bam*HI-linearised pAC-SP6 vector) [234, 290]. Injected embryos were imaged on a Zeiss780 LSM inverted confocal microscope equipped with EC Plan-Neofluar 10X/0.30 NA WD=5.2 (Zeiss) objective or using a Zeiss Axio Scope.A1 equipped with

5x/0.15 NA N-Achroplan or 10X/0.3 NA EC Plan-Neofluar objectives (Zeiss) at desired developmental stages.

The raw and processed data generated in this study have also been submitted to GEO (GEO record GSE106676).

5.3 Chapter 3 Methods

5.3.1 Tg(FoxD3:enh6-GFP)^{ox109} and Tg(FoxD3:enh4-GFP)^{ox110} line generation

The *Ac/Ds*-mediated transgenesis of the lines was carried out as described above (Section 5.2.5). F0 embryos were raised and screened for founders. Positive F1 embryos were raised for future experiments.

5.3.2 JASPAR and HOCOMOCO TFBS analysis

TFBS analysis on the Enh6 or the Enh4 DNA sequences, which were cloned into the E1b-GFP-AC-DS vector used for the transgenesis, was performed using both, HOCOMOCO (<http://hocomoco11.autosome.ru>) and JASPAR (7th release, <http://jaspar.genereg.net/>) databases [236, 237]. For the JASPAR analysis, Vertebrata taxonomic group database was used, set to 87% relative profile score threshold. Candidate TFs were then selected based on identified multiple predicted binding sites for the same TF family for either Enh4 or Enh6 sequences. For the HOCOMOCO analysis, mouse mononucleotide P-value 0.0001 TFBS model motif in Homer format was scanned and annotated using Homer(v.4.8) `annotatePeaks.pl` [288]. Candidate TFs were then selected based on identified multiple predicted binding sites for the same TF family for either Enh4 or Enh6 sequences. For final TFBS plotting: to highlight the most enriched and also differential TFBSs underlying Enh6 and Enh4 DNA sequences, we have prioritised and plotted those uncovered TFBSs that were found: (a) at least 3 times within either Enh6 or Enh4 and (b) each uncovered TF had at least two more sites between Enh6 and Enh4 or Enh4 and Enh6 sequences.

5.3.3 Transgenic zebrafish line embryo imaging and analysis

Embryos at 5-6ss or larvae at 2dpf from Tg(FoxD3:enh6-GFP)^{ox109} or Tg(FoxD3:enh4-GFP)^{ox110} line incrosses to Gt(foxd3-mcherry)^{ct110aR} line were manually dechorionated if needed, anesthetized with tricaine and mounted in 1% low-melting agarose for live imaging. Embryos or larvae were imaged on a Zeiss780 LSM inverted confocal microscope

equipped with Plan-APO Chromat 10x/0.45 NA WD=2.0 mm and LD C-APO Chromat 40x/1.1 NA WD=0.62 mm objectives or on a Zeiss780 LSM upright confocal microscope equipped with Plan-APO Chromat 10x/0.45 NA WD=2.0 mm objective. Images were processed using Fiji or Imaris image analysis software. Enh6-GFP or Enh4-GFP co-localisation with foxd3-mCherry fluorescence was performed on 40x magnification optical slices (5-6ss: 15-16 images, 2dpf: 22-26 images). Pearson correlation was acquired using Fiji Coloc2 plug-in, set for a bisection threshold regression calculation for compared images. Welch's statistical *t*-test was applied to compare Enh6-GFP and foxd3-mCherry with Enh4-GFP and foxd3-mCherry co-localisation mean variances at either 5-6ss or 2dpf . Results were plotted in GraphPad prism.

5.3.4 Enh6+/- and Enh4+/- cell RNA-seq

Bulk RNA extraction, library preparation and sequencing

FAC sorted cells (5000 cells for each sample: 3X Enh6+, 2X Enh6-, 3X Enh4+ and 2X Enh4-) were washed with PBS and stored in lysis buffer (RNAqueous-Micro Total RNA Isolation Kit, Ambion, AM1931) at -80°C. RNA was extracted using Ambion RNAqueous Micro Total RNA isolation kit and checked on TapeStation using Agilent High Sensitivity RNA Tape (5067-5579) and buffer (5067-5580). cDNA was prepared using Takara Clontech SMART-Seq v4 Ultra Low Input RNA Kit for Sequencing (634890) using 8 cycles of amplification and purified using Agencourt AMPure XP beads (Beckman Coulter, A63880). 1ng of each sample cDNA was used for library preparation. Sequencing libraries were prepared using Illumina Nextera XT library preparation kit (FC-131-1096) and purified using Agencourt AMPure XP beads. The library quality was checked using Agilent High Sensitivity TapeStation D1000 ScreenTape and Reagents. All libraries were pooled into a single pool and sequenced using NextSeq 500/550 High Output Kit v2 (75 cycles, paired-end 40 bp run) on NextSeq 500 sequencing system.

RNA-seq bioinformatics processing

Reads were trimmed to remove low quality bases using trim_galore/0.4.1 (Babraham). Read quality was evaluated using FastQC (Babraham). Mapping to GRCz10/danRer10 assembly of the zebrafish genome downloaded from UCSC Genome Browser was performed using rna-star/2.4.2a [284]. Read counts were obtained using subread FeatureCounts (subread/1.6.2) [285] with standard parameters using a gene model gtf derived from Ensembl annotation downloaded from UCSC genome browser. Differential Expression analysis (Enh6+ vs Enh6-, Enh4+ vs Enh4- and Enh6+ vs Enh4+) was carried out using DESeq2 (1.24.0) [286]. Gene Ontology Biological Terms were acquired using PANTHER Danio Rerio terms (used parameters: Statistical overrepresentation test, GO biological process complete, Binomial Test using the Bonferroni correction for multiple testing) (<http://www.pantherdb.org/>).

5.3.5 Enh6+/- and Enh4+/- cell ATAC-seq

ATAC, library preparation and sequencing

FAC sorted cells (15,000 - 25,000 for each sample: 3X Enh6+, 3X Enh6-, 3X Enh4+ and 3X Enh4-) were lysed (10mM Tris-HCl, pH7.4, 10mM NaCl, 3mM MgCl₂, 0.1% Igepal) and tagged using Nextera DNA kit (Illumina FC-121-1030) (tagmentation - 0.7µl of Tn5 enzyme in 10µl reaction) for 30min at 37°C and then reactions were quenched with 50mM EDTA for 30 minutes at 50°C. Tagmented fragments were purified using MinElute PCR Purification Kit. Tagmented DNA was amplified using NEB Next High-Fidelity 2X PCR Master Mix for 12-13 cycles and purified using Agencourt AMPure XP beads. Tagmentation efficiency and final libraries were assessed using Agilent High Sensitivity TapeStation D1000 ScreenTape and Reagents. All libraries were pooled into a single pool and sequenced using NextSeq 500/550 High Output Kit v2 (75 cycles, paired-end 40 bp run) on NextSeq 500 sequencing system.

ATAC-seq bioinformatics processing

Reads were trimmed to remove low quality bases using trim_galore/0.4.1 (Babraham). Read quality was evaluated using FastQC (Babraham). Reads were mapped to the danRer10 genome using bowtie/1.1.2. Bigwig files were generated using an enhanced Perl script (courtesy of Prof Jim Hughes). Duplicates were removed using Picard tools MarkDuplicates (picard-tools/1.83). Sample reads were equalised to the same number by random down-sampling using samtools-1.1 (to 24,661,382 reads, except for one of Ehh4-samples that had 11,457,938 reads). Peak calling was performed as described previously [34]. Briefly, BAM files were sorted by name and paired end bed files were obtained using bedtools(v.2.15.0) bamtobed -bedpe. Reads that were not properly paired were discarded and paired reads were displaced by +4 bp and -5 bp for transposition correction. Peak calling was performed using MACS2 (macs2/2.1.2) callpeak -nomodel -shift -37 -extsize 73 -slocal 1000 -B -q 0.05 parameters [289]. Only the most significant peaks ($p < 10^{-10}$) that were found to be present in at least two out of three replicates were used for the downstream analysis (consensus peaks) using bedtools intersect function. After generating the consensus peak sets for each sample, they were merged for a unified coordinate matrix that was used for the k -mean clustering to identify distinctive clusters showing different accessibility regions between the compared sample groups. k -means clustering of ATAC-seq signal was carried out using SeqMINR software as described [193]. k -mean cluster heatmaps and profiles were generated using deeptools/3.0.1 [291]. Gene ontology terms of the generated clusters were acquired using GREAT tool (version 3.0.0; Zebrafish: Wellcome Trust Zv9 genome assembly) (<http://great.stanford.edu/great/public-3.0.0/html/index.php>). To visualise foxd3 locus for the ATAC-seq results, reads mapping to the enhancer reporter plasmid were excluded by: first mapping all acquired reads (using bowtie/1.1.2.) to a plasmid sequence carrying 150bp of a cloned Enh6 or Enh4 sequence from each side of a plasmid, and then taking all the unmapped reads for the downstream mapping to the danRer10 genome as described above.

5.4 10x Chromium single cell RNA-seq

5.4.1 RNA extraction, library preparation and sequencing

Embryos issued for *Gt(foxd3-Citrine)ct110a* line were dissociated to single cell suspension using a standard protocol as described above, and individual Citrine-positive cells were collected by FACS into 2 μ l Hanks buffer. Approximately 10,000 cells to be used for droplet-based single-cell RNA-seq were loaded onto the Chromium 10X ChIP and the Chromium Single Cell 3' Library and Gel Bead Kit v3 (Cat.No. PN-1000092) was used following manufacturers instructions. Libraries were quantified using Qubit and Kappa qPCR and 4nM 10X scRNA-seq library was sequenced on a NextSeq500 platform (Illumina) using NextSeq 500/550 High Output Kit v2.5 (150 Cycles) (Cat.No. 20024907) in 28 x 8 x 0 x 91 mode.

5.4.2 Single cell RNA-seq bioinformatics processing

Single-cell RNA-seq raw base call files were demultiplexed using Cellranger (v3.0.0) [292] and mapped to the danRer10 transcriptome that contained the Citrine transcript and compiled using the Cellranger 'mkref' function. 4,980 estimated number of cells were obtained with mean reads per cell of 79,129, median genes per cell of 3,754, median UMI counts per cell of 18,975 and total number of genes detected was 23,343. 97.4% of reads had valid barcodes with a Q30 of 97% and 88.8% of the reads mapped confidently to the zebrafish genome. Downstream analysis was carried out using Seurat package (v3.0.0) in R [293]. Matrices were filtered to remove barcodes with fewer than 1000 genes, more than 5000 genes expressed and carried a high percentage of UMIs from mitochondrial features (greater than 8.5%) (Figure 5.2). 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 = 0.4, dims = 1:18). Ten cell clusters were obtained and visualised on a Umap plot. The data was subset to include the Citrine positive cells at the cranial and trunk region. Clusters containing specified hindbrain neural crest were not included for

visualisation purposes. 2,455 selected cells were further subclustered in 8 clusters (3 posterior and 5 anterior) (resolution = 0.4, dims = 1:15). The only NMP-related clusters from these 8 clusters were also singled out for visualisation purposes.

Differentiation pseudotime trajectories were reconstructed using R package Monocle (v2.10.1) [294] using data from the previously normalised and clustered Seurat objects as described above. Dimensional reduction was performed with the DDRTree algorithm [295] using highly varied genes as inputs. Cell trajectories were then reconstructed using orderCells function.

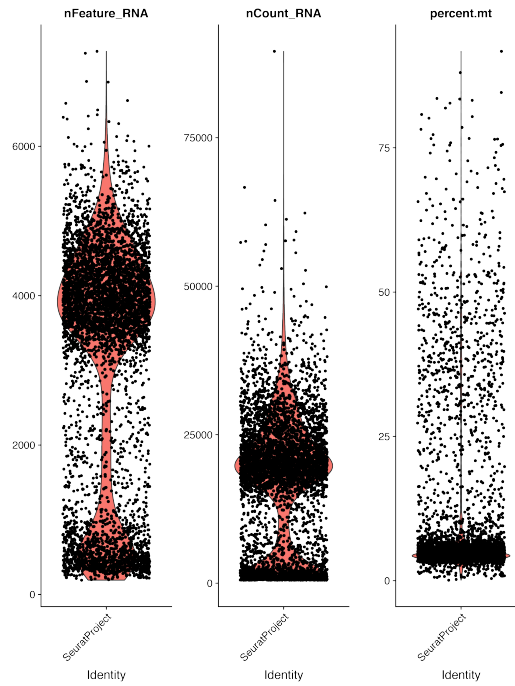


Figure 5.2 Information on scRNA-seq features.

5.5 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.

1 Other Enh6+/- and Enh4+/- ATAC-seq *k*-mean clusters (Section 3.7)

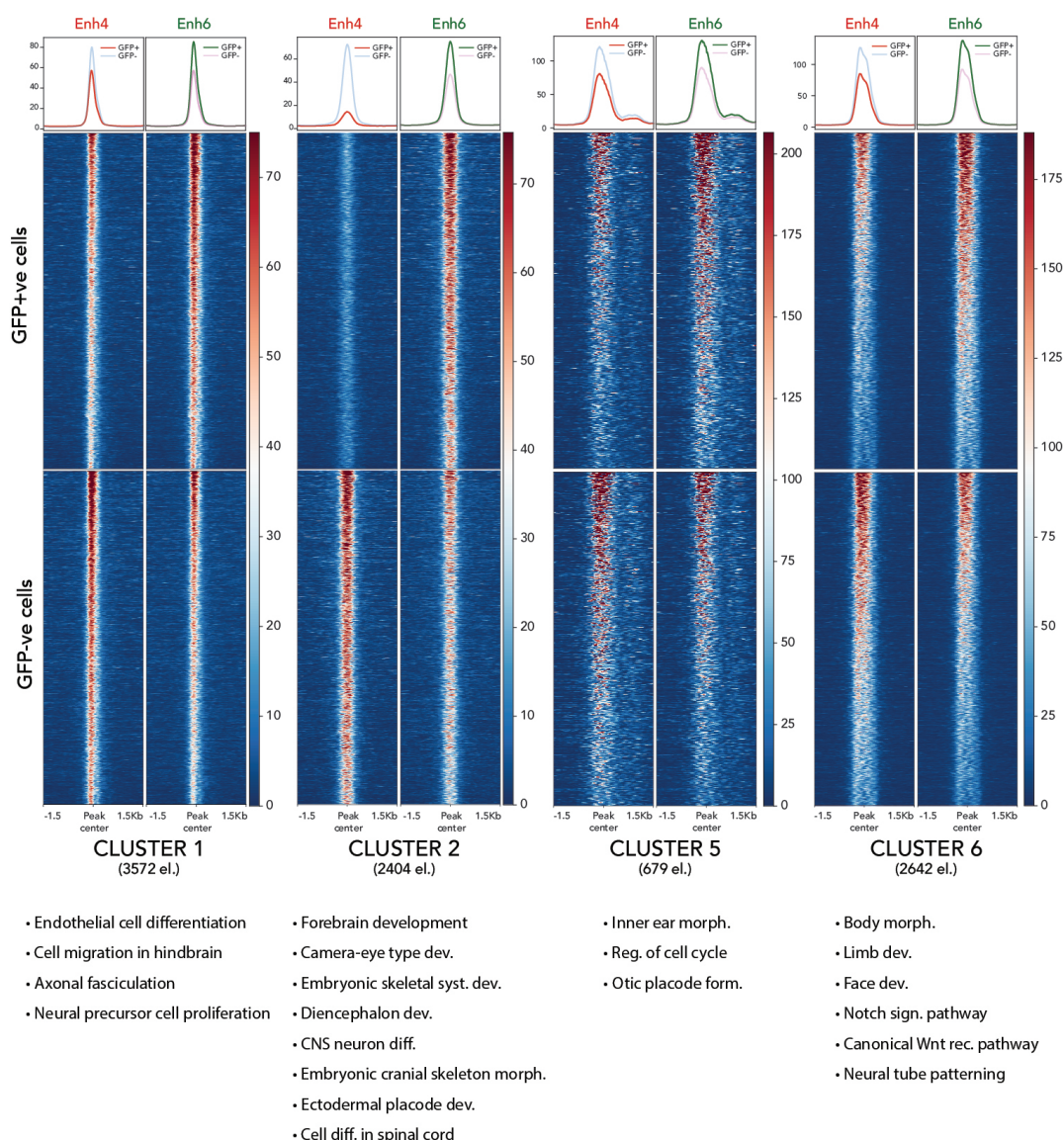


Figure 1 Additional 1-6 *k*-mean clusters to Clusters 6-1, 6-2, 4-1 and 4-2. Profiles: light blue corresponds to Enh4- cells, red - Enh4+ cells, pink - Enh6- cells, green - Enh6+ cells. Heatmap legends indicate average ATAC signals.

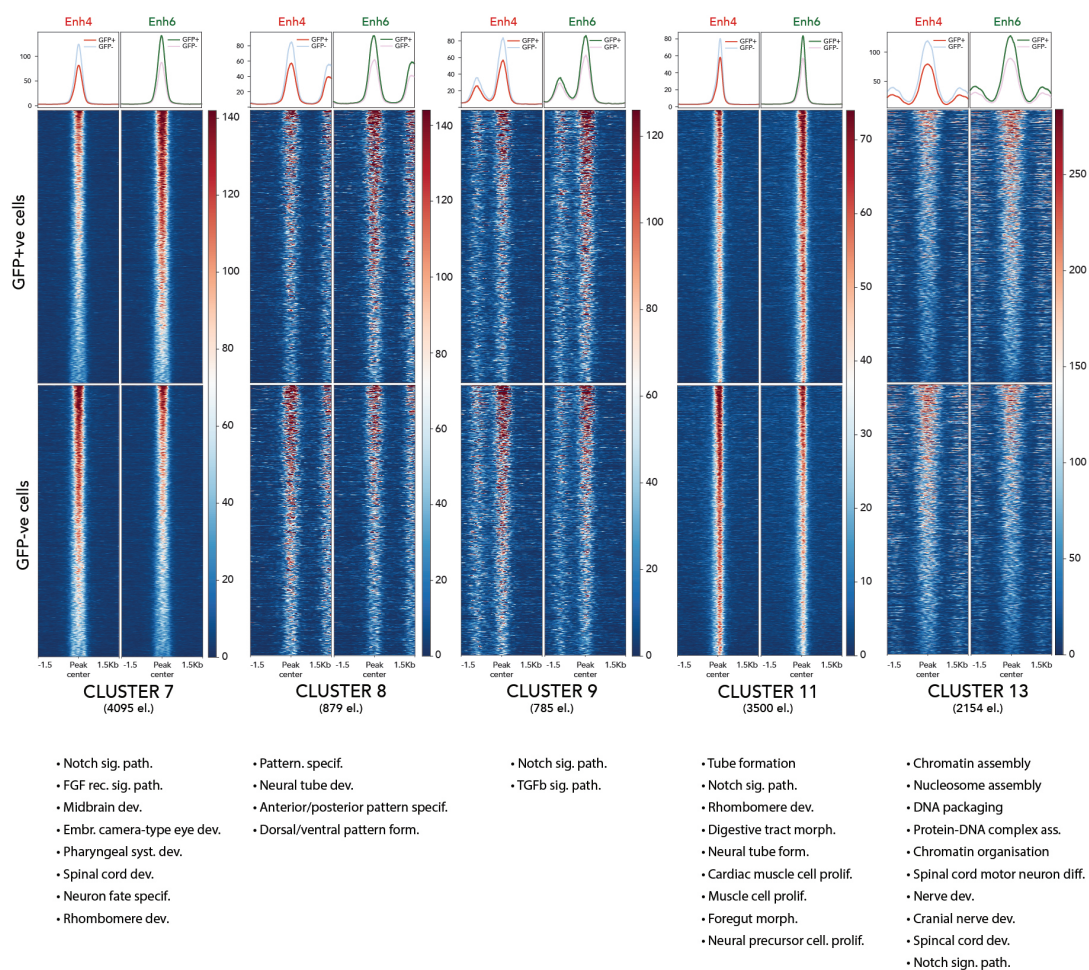


Figure 2 Additional 7-13 *k*-mean clusters to Clusters 6-1, 6-2, 4-1 and 4-2. Profiles: light blue corresponds to Enh4- cells, red - Enh4+ cells, pink - Enh6- cells, green - Enh6+ cells. Heatmap legends indicate average ATAC signals.

Bibliography

- [1] C. H. Waddington. *Organisers and genes*. [S.l.] : [s.n.], 1947 [1940].
- [2] A. Boija and M. Mannervik. A time of change: Dynamics of chromatin and transcriptional regulation during nuclear programming in early drosophila development. *Mol Reprod Dev*, 82(10):735–46, 2015.
- [3] J. R. Dixon, S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu, and B. Ren. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, 485(7398):376–80, 2012.
- [4] J. Fraser, C. Ferrai, A. M. Chiariello, M. Schueler, T. Rito, G. Laudanno, M. Barbieri, B. L. Moore, D. C. Kraemer, S. Aitken, S. Q. Xie, K. J. Morris, M. Itoh, H. Kawaji, I. Jaeger, Y. Hayashizaki, P. Carninci, A. R. Forrest, C. A. Semple, J. Dostie, A. Pombo, M. Nicodemi, and F. Consortium. Hierarchical folding and reorganization of chromosomes are linked to transcriptional changes in cellular differentiation. *Mol Syst Biol*, 11(12):852, 2015.
- [5] M. E. Minard, A. K. Jain, and M. C. Barton. Analysis of epigenetic alterations to chromatin during development. *Genesis*, 47(8):559–72, 2009.
- [6] S. G. Swygert and C. L. Peterson. Chromatin dynamics: interplay between remodeling enzymes and histone modifications. *Biochim Biophys Acta*, 1839(8):728–36, 2014.
- [7] E. Calo and J. Wysocka. Modification of enhancer chromatin: what, how, and why? *Mol Cell*, 49(5):825–37, 2013.
- [8] B. E. Bernstein, J. A. Stamatoyannopoulos, J. F. Costello, B. Ren, A. Milosavljevic, A. Meissner, M. Kellis, M. A. Marra, A. L. Beaudet, J. R. Ecker, P. J. Farnham,

- M. Hirst, E. S. Lander, T. S. Mikkelsen, and J. A. Thomson. The nih roadmap epigenomics mapping consortium. *Nat Biotechnol*, 28(10):1045–8, 2010.
- [9] The encode (encyclopedia of dna elements) project. *Science*, 306(5696):636–40, 2004.
- [10] C. E. Romanoski, C. K. Glass, H. G. Stunnenberg, L. Wilson, and G. Almouzni. Epigenomics: Roadmap for regulation. *Nature*, 518(7539):314–6, 2015.
- [11] J. Ernst and M. Kellis. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat Biotechnol*, 28(8):817–25, 2010.
- [12] K. Maeshima, R. Imai, S. Tamura, and T. Nozaki. Chromatin as dynamic 10-nm fibers. *Chromosoma*, 123(3):225–37, 2014.
- [13] K. Luger, A. W. Mader, R. K. Richmond, D. F. Sargent, and T. J. Richmond. Crystal structure of the nucleosome core particle at 2.8 a resolution. *Nature*, 389(6648):251–60, 1997.
- [14] G. Schafer, C. R. McEvoy, and H. G. Patterson. The *saccharomyces cerevisiae* linker histone hho1p is essential for chromatin compaction in stationary phase and is displaced by transcription. *Proc Natl Acad Sci U S A*, 105(39):14838–43, 2008.
- [15] Y. Fan, T. Nikitina, J. Zhao, T. J. Fleury, R. Bhattacharyya, E. E. Bouhassira, A. Stein, C. L. Woodcock, and A. I. Skoultschi. Histone h1 depletion in mammals alters global chromatin structure but causes specific changes in gene regulation. *Cell*, 123(7):1199–212, 2005.
- [16] P. J. Robinson and D. Rhodes. Structure of the ‘30 nm’ chromatin fibre: a key role for the linker histone. *Curr Opin Struct Biol*, 16(3):336–43, 2006.
- [17] K. L. Huisinga, B. Brower-Toland, and S. C. Elgin. The contradictory definitions of heterochromatin: transcription and silencing. *Chromosoma*, 115(2):110–22, 2006.
- [18] H. Tamaru. Confining euchromatin/heterochromatin territory: jumonji crosses the line. *Genes Dev*, 24(14):1465–78, 2010.

- [19] A. G. Larson, D. Elnatan, M. M. Keenen, M. J. Trnka, J. B. Johnston, A. L. Burlingame, D. A. Agard, S. Redding, and G. J. Narlikar. Liquid droplet formation by hp1a suggests a role for phase separation in heterochromatin. *Nature*, 547(7662):236–240, 2017.
- [20] A. J. Plys, C. P. Davis, J. Kim, G. Rizki, M. M. Keenen, S. K. Marr, and R. E. Kingston. Phase separation of polycomb-repressive complex 1 is governed by a charged disordered region of cbx2. *Genes Dev*, 33(13-14):799–813, 2019.
- [21] R. Tatavosian, S. Kent, K. Brown, T. Yao, H. N. Duc, T. N. Huynh, C. Y. Zhen, B. Ma, H. Wang, and X. Ren. Nuclear condensates of the polycomb protein chromobox 2 (cbx2) assemble through phase separation. *J Biol Chem*, 294(5):1451–1463, 2019.
- [22] S. Sanulli, M. Trnka, V. Dharmarajan, R. Tibble, B. Pascal, A. Burlingame, P. Griffin, J. Gross, and G. Narlikar. Disorganization of the histone core promotes organization of heterochromatin into phase-separated droplets. *bioRxiv*, page 473132, 2018.
- [23] H. D. Le, K. M. Donaldson, K. R. Cook, and G. H. Karpen. A high proportion of genes involved in position effect variegation also affect chromosome inheritance. *Chromosoma*, 112(6):269–76, 2004.
- [24] G. Li and L. Zhou. Genome-wide identification of chromatin transitional regions reveals diverse mechanisms defining the boundary of facultative heterochromatin. *PLoS One*, 8(6):e67156, 2013.
- [25] N. Brockdorff. The role of xist in x-inactivation. *Curr Opin Genet Dev*, 8(3):328–33, 1998.
- [26] A. Eberhardter 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.

- [27] L. Di Stefano, J. A. Walker, G. Burgio, D. F. Corona, P. Mulligan, A. M. Naar, and N. J. Dyson. Functional antagonism between histone h3k4 demethylases in vivo. *Genes Dev*, 25(1):17–28, 2011.
- [28] E. Meshorer, D. Yellajoshula, E. George, P. J. Scambler, D. T. Brown, and T. Misteli. Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. *Dev Cell*, 10(1):105–16, 2006.
- [29] T. Sagai, M. Hosoya, Y. Mizushima, M. Tamura, and T. Shiroishi. Elimination of a long-range cis-regulatory module causes complete loss of limb-specific shh expression and truncation of the mouse limb. *Development*, 132(4):797–803, 2005.
- [30] 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.
- [31] D. Shlyueva, G. Stampfel, and A. Stark. Transcriptional enhancers: from properties to genome-wide predictions. *Nat Rev Genet*, 15(4):272–86, 2014.
- [32] 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.
- [33] 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.
- [34] 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.
- [35] L. Song and G. E. Crawford. Dnase-seq: a high-resolution technique for mapping

- active gene regulatory elements across the genome from mammalian cells. *Cold Spring Harb Protoc*, 2010(2):pdb.prot5384, 2010.
- [36] R. Margueron and D. Reinberg. The polycomb complex prc2 and its mark in life. *Nature*, 469(7330):343–9, 2011.
- [37] 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.
- [38] A. J. Will, G. Cova, M. Osterwalder, W. L. Chan, L. Wittler, N. Brieske, V. Heinrich, J. P. de Villartay, M. Vingron, E. Klopocki, A. Visel, D. G. Lupianez, and S. Mundlos. Composition and dosage of a multipartite enhancer cluster control developmental expression of *ihh* (indian hedgehog). *Nat Genet*, 49(10):1539–1545, 2017.
- [39] J. O. Davies, A. M. Oudelaar, D. R. Higgs, and J. R. Hughes. How best to identify chromosomal interactions: a comparison of approaches. *Nat Methods*, 14(2):125–134, 2017.
- [40] G. Andrey and S. Mundlos. The three-dimensional genome: regulating gene expression during pluripotency and development. *Development*, 144(20):3646–3658, 2017.
- [41] 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.
- [42] 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.

- [43] 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.
- [44] 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.
- [45] J. Rhodes, A. Feldmann, B. Hernandez-Rodriguez, N. DiÂaz, J. Brown, N. Fursova, N. Blackledge, P. Prathapan, P. Dobrinic, M. Huseyin, A. Szczurek, K. Kruse, K. Nasmyth, V. Buckle, J. Vaquerizas, and R. Klose. Cohesin disrupts polycomb-dependent chromosome interactions. *bioRxiv*, page 593970, 2019.
- [46] R. J. Palstra, M. Simonis, P. Klous, E. Brasset, B. Eijkelkamp, and W. de Laat. Maintenance of long-range dna interactions after inhibition of ongoing rna polymerase ii transcription. *PLoS One*, 3(2):e1661, 2008.
- [47] W. Deng, J. Lee, H. Wang, J. Miller, A. Reik, P. D. Gregory, A. Dean, and G. A. Blobel. Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell*, 149(6):1233–44, 2012.
- [48] H. Sutherland and W. A. Bickmore. Transcription factories: gene expression in unions? *Nat Rev Genet*, 10(7):457–66, 2009.
- [49] B. M. Javierre, O. S. Burren, S. P. Wilder, R. Kreuzhuber, S. M. Hill, S. Sewitz, J. Cairns, S. W. Wingett, C. Varnai, M. J. Thiecke, F. Burden, S. Farrow, A. J. Cutler, K. Rehnstrom, K. Downes, L. Grassi, M. Kostadima, P. Freire-Pritchett, F. Wang, H. G. Stunnenberg, J. A. Todd, D. R. Zerbino, O. Stegle, W. H. Ouwehand, M. Frontini, C. Wallace, M. Spivakov, and P. Fraser. Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. *Cell*, 167(5):1369–1384.e19, 2016.

- [50] P. Freire-Pritchett, S. Schoenfelder, C. Varnai, S. W. Wingett, J. Cairns, A. J. Collier, R. Garcia-Vilchez, M. Furlan-Magaril, C. S. Osborne, P. Fraser, P. J. Rugg-Gunn, and M. Spivakov. Global reorganisation of cis-regulatory units upon lineage commitment of human embryonic stem cells. *Elife*, 6, 2017.
- [51] J. E. Phillips-Cremins, M. E. Sauria, A. Sanyal, T. I. Gerasimova, B. R. Lajoie, J. S. Bell, C. T. Ong, T. A. Hookway, C. Guo, Y. Sun, M. J. Bland, W. Wagstaff, S. Dalton, T. C. McDevitt, R. Sen, J. Dekker, J. Taylor, and V. G. Corces. Architectural protein subclasses shape 3d organization of genomes during lineage commitment. *Cell*, 153(6):1281–95, 2013.
- [52] D. G. Lupianez, 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.
- [53] Y. Ghavi-Helm, A. Jankowski, S. Meiers, R. R. Viales, J. O. Korb, and E. E. M. Furlong. Highly rearranged chromosomes reveal uncoupling between genome topology and gene expression. *Nat Genet*, 51(8):1272–1282, 2019.
- [54] I. Williamson, L. Kane, P. S. Devenney, I. M. Flyamer, E. Anderson, F. Kilanowski, R. E. Hill, W. A. Bickmore, and L. A. Lettice. Developmentally regulated shh expression is robust to tad perturbations. *Development*, 146(19):dev179523, 2019.
- [55] P. Chen, J. Zhao, and G. Li. Histone variants in development and diseases. *J Genet Genomics*, 40(7):355–65, 2013.
- [56] W. Albig and D. Doenecke. The human histone gene cluster at the d6s105 locus. *Hum Genet*, 101(3):284–94, 1997.
- [57] B. Bitterer and R. Schneider. Histone variants: key players of chromatin. *Cell Tissue Res*, 356(3):457–66, 2014.

- [58] V. Turinetto and C. Giachino. Histone variants as emerging regulators of embryonic stem cell identity. *Epigenetics*, 10(7):563–73, 2015.
- [59] A. van Daal and S. C. Elgin. A histone variant, h2avd, is essential in drosophila melanogaster. *Mol Biol Cell*, 3(6):593–602, 1992.
- [60] X. Liu, B. Li, and GorovskyMA. Essential and nonessential histone h2a variants in tetrahymena thermophila. *Mol Cell Biol*, 16(8):4305–11, 1996.
- [61] R. Faast, V. Thonglairoam, T. C. Schulz, J. Beall, J. R. Wells, H. Taylor, K. Matthaai, P. D. Rathjen, D. J. Tremethick, and I. Lyons. Histone variant h2a.z is required for early mammalian development. *Curr Biol*, 11(15):1183–7, 2001.
- [62] P. Ridgway, K. D. Brown, D. Rangasamy, U. Svensson, and D. J. Tremethick. Unique residues on the h2a.z containing nucleosome surface are important for xenopus laevis development. *J Biol Chem*, 279(42):43815–20, 2004.
- [63] J. Zlatanova and A. Thakar. H2a.z: view from the top. *Structure*, 16(2):166–79, 2008.
- [64] M. P. Creighton, S. Markoulaki, S. S. Levine, J. Hanna, M. A. Lodato, K. Sha, R. A. Young, R. Jaenisch, and L. A. Boyer. H2az is enriched at polycomb complex target genes in es cells and is necessary for lineage commitment. *Cell*, 135(4):649–61, 2008.
- [65] R. K. Suto, M. J. Clarkson, D. J. Tremethick, and K. Luger. Crystal structure of a nucleosome core particle containing the variant histone h2a.z. *Nat Struct Biol*, 7(12):1121–4, 2000.
- [66] M. J. Clarkson, J. R. Wells, F. Gibson, R. Saint, and D. J. Tremethick. Regions of variant histone his2avd required for drosophila development. *Nature*, 399(6737):694–7, 1999.
- [67] V. Subramanian, A. Mazumder, L. E. Surface, V. L. Butty, P. A. Fields, A. Alwan, L. Torrey, K. K. Thai, S. S. Levine, M. Bathe, and L. A. Boyer. H2a.z acidic patch

- couples chromatin dynamics to regulation of gene expression programs during esc differentiation. *PLoS Genet*, 9(8):e1003725, 2013.
- [68] Z. Li, P. Gadue, K. Chen, Y. Jiao, G. Tuteja, J. Schug, W. Li, and K. H. Kaestner. Foxa2 and h2a.z mediate nucleosome depletion during embryonic stem cell differentiation. *Cell*, 151(7):1608–16, 2012.
- [69] K. Zaret. Developmental competence of the gut endoderm: genetic potentiation by gata and hnf3/fork head proteins. *Dev Biol*, 209(1):1–10, 1999.
- [70] L. Shi, H. Wen, and X. Shi. The histone variant h3.3 in transcriptional regulation and human disease. *J Mol Biol*, 429(13):1934–1945, 2017.
- [71] S. B. Hake, B. A. Garcia, M. Kauer, S. P. Baker, J. Shabanowitz, D. F. Hunt, and C. D. Allis. Serine 31 phosphorylation of histone variant h3.3 is specific to regions bordering centromeres in metaphase chromosomes. *Proc Natl Acad Sci U S A*, 102(18):6344–9, 2005.
- [72] U. Braunschweig, G. J. Hogan, L. Pagie, and B. van Steensel. Histone h1 binding is inhibited by histone variant h3.3. *Embo j*, 28(23):3635–45, 2009.
- [73] C. Jin and G. Felsenfeld. Nucleosome stability mediated by histone variants h3.3 and h2a.z. *Genes Dev*, 21(12):1519–29, 2007.
- [74] A. D. Goldberg, L. A. Banaszynski, K. M. Noh, P. W. Lewis, S. J. Elsaesser, S. Stadler, S. Dewell, M. Law, X. Guo, X. Li, D. Wen, A. Chapgier, R. C. DeKever, J. C. Miller, Y. L. Lee, E. A. Boydston, M. C. Holmes, P. D. Gregory, J. M. Greally, S. Rafii, C. Yang, P. J. Scambler, D. Garrick, R. J. Gibbons, D. R. Higgs, I. M. Cristea, F. D. Urnov, D. Zheng, and C. D. Allis. Distinct factors control histone variant h3.3 localization at specific genomic regions. *Cell*, 140(5):678–91, 2010.
- [75] D. Ray-Gallet, A. Woolfe, I. Vassias, C. Pellentz, N. Lacoste, A. Puri, D. C. Schultz, N. A. Pchelintsev, P. D. Adams, L. E. Jansen, and G. Almouzni. Dynamics of histone h3 deposition in vivo reveal a nucleosome gap-filling mechanism for h3.3 to maintain chromatin integrity. *Mol Cell*, 44(6):928–41, 2011.

- [76] M. E. Torres-Padilla, A. J. Bannister, P. J. Hurd, T. Kouzarides, and M. Zernicka-Goetz. Dynamic distribution of the replacement histone variant h3.3 in the mouse oocyte and preimplantation embryos. *Int J Dev Biol*, 50(5):455–61, 2006.
- [77] L. A. Banaszynski, D. Wen, S. Dewell, S. J. Whitcomb, M. Lin, N. Diaz, S. J. Elsaesser, A. Chagier, A. D. Goldberg, E. Canaani, S. Rafii, D. Zheng, and C. D. Allis. Hira-dependent histone h3.3 deposition facilitates prc2 recruitment at developmental loci in es cells. *Cell*, 155(1):107–20, 2013.
- [78] P. W. Lewis, S. J. Elsaesser, K. M. Noh, S. C. Stadler, and C. D. Allis. Daxx is an h3.3-specific histone chaperone and cooperates with atrx in replication-independent chromatin assembly at telomeres. *Proc Natl Acad Sci U S A*, 107(32):14075–80, 2010.
- [79] M. Iwafuchi-Doi, G. Donahue, A. Kakumanu, J. A. Watts, S. Mahony, B. F. Pugh, D. Lee, K. H. Kaestner, and K. S. Zaret. The pioneer transcription factor foxa maintains an accessible nucleosome configuration at enhancers for tissue-specific gene activation. *Mol Cell*, 62(1):79–91, 2016.
- [80] R. Gualdi, P. Bossard, M. Zheng, Y. Hamada, J. R. Coleman, and K. S. Zaret. Hepatic specification of the gut endoderm in vitro: cell signaling and transcriptional control. *Genes Dev*, 10(13):1670–82, 1996.
- [81] 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. *Mol Cell*, 9(2):279–89, 2002.
- [82] K. L. Clark, E. D. Halay, E. Lai, and S. K. Burley. Co-crystal structure of the hnf-3/fork head dna-recognition motif resembles histone h5. *Nature*, 364(6436):412–20, 1993.
- [83] V. Ramakrishnan, J. T. Finch, V. Graziano, P. L. Lee, and R. M. Sweet. Crystal structure of globular domain of histone h5 and its implications for nucleosome binding. *Nature*, 362(6417):219–23, 1993.

- [84] T. Sekiya, U. M. Muthurajan, K. Luger, A. V. Tulin, and K. S. Zaret. Nucleosome-binding affinity as a primary determinant of the nuclear mobility of the pioneer transcription factor foxa. *Genes Dev*, 23(7):804–9, 2009.
- [85] M. Iwafuchi-Doi and K. S. Zaret. Pioneer transcription factors in cell reprogramming. *Genes Dev*, 28(24):2679–92, 2014.
- [86] K. Takahashi and S. Yamanaka. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, 126(4):663–76, 2006.
- [87] A. Soufi, G. Donahue, and K. S. Zaret. Facilitators and impediments of the pluripotency reprogramming factors’ initial engagement with the genome. *Cell*, 151(5):994–1004, 2012.
- [88] A. Soufi, M. F. Garcia, A. Jaroszewicz, N. Osman, M. Pellegrini, and K. S. Zaret. Pioneer transcription factors target partial dna motifs on nucleosomes to initiate reprogramming. *Cell*, 161(3):555–568, 2015.
- [89] M. P. Meers, D. H. Janssens, and S. Henikoff. Pioneer factor-nucleosome binding events during differentiation are motif encoded. *Mol Cell*, 75(3):562–575.e5, 2019.
- [90] W. His. *Untersuchungen über die erste Anlage des Wirbelthierleibes: Die erste Entwicklung des Menschen im Ei*. His, Wilhelm, 1831-1904 Untersuchungen über die erste Anlage des Wirbelthierleibes. Vogel, Leipzig,, 1868.
- [91] A. M. MARSHALL. Memoirs: The development of the cranial nerves in the chick. *Quarterly Journal of Microscopical Science*, s2-18(69):10–40, 1878.
- [92] 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.
- [93] M. L. Basch and M. Bronner-Fraser. Neural crest inducing signals. *Adv Exp Med Biol*, 589:24–31, 2006.

- [94] 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. *Dev Biol*, 366(1):10–21, 2012.
- [95] M. Simoes-Costa and M. E. Bronner. Establishing neural crest identity: a gene regulatory recipe. *Development*, 142(2):242–57, 2015.
- [96] S. A. Green, M. Simoes-Costa, and M. E. Bronner. Evolution of vertebrates as viewed from the crest. *Nature*, 520(7548):474–482, 2015.
- [97] C. Gans and R. G. Northcutt. Neural crest and the origin of vertebrates: a new head. *Science*, 220(4594):268–73, 1983.
- [98] R. Glenn Northcutt. The new head hypothesis revisited. *J Exp Zool B Mol Dev Evol*, 304(4):274–97, 2005.
- [99] A. Graham. The neural crest. *Curr Biol*, 13(10):R381–4, 2003.
- [100] P. C. Donoghue, A. Graham, and R. N. Kelsh. The origin and evolution of the neural crest. *Bioessays*, 30(6):530–41, 2008.
- [101] H. C. Etchevers, G. Couly, C. Vincent, and N. M. Le Douarin. Anterior cephalic neural crest is required for forebrain viability. *Development*, 126(16):3533–43, 1999.
- [102] S. Creuzet, B. Schuler, G. Couly, and N. M. Le Douarin. Reciprocal relationships between fgf8 and neural crest cells in facial and forebrain development. *Proc Natl Acad Sci U S A*, 101(14):4843–7, 2004.
- [103] H. C. Etchevers, E. Dupin, and N. M. Le Douarin. The diverse neural crest: from embryology to human pathology. *Development*, 146(5), 2019.
- [104] R. Cerny, M. Cattell, T. Sauka-Spengler, M. Bronner-Fraser, F. Yu, and D. M. Medeiros. Evidence for the prepattern/cooption model of vertebrate jaw evolution. *Proc Natl Acad Sci U S A*, 107(40):17262–7, 2010.

- [105] T. Sauka-Spengler, D. Meulemans, M. Jones, and M. Bronner-Fraser. Ancient evolutionary origin of the neural crest gene regulatory network. *Dev Cell*, 13(3):405–20, 2007.
- [106] N. Nikitina, T. Sauka-Spengler, and M. Bronner-Fraser. Dissecting early regulatory relationships in the lamprey neural crest gene network. *Proc Natl Acad Sci U S A*, 105(51):20083–8, 2008.
- [107] D. Hockman, V. Chong-Morrison, D. Gavriouchkina, S. Green, 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. *bioRxiv*, page 275875, 2018.
- [108] T. Sauka-Spengler and M. Bronner-Fraser. A gene regulatory network orchestrates neural crest formation. *Nat Rev Mol Cell Biol*, 9(7):557–68, 2008.
- [109] J. R. Dutton, A. Antonellis, T. J. Carney, F. S. Rodrigues, W. J. Pavan, A. Ward, and R. N. Kelsh. An evolutionarily conserved intronic region controls the spatiotemporal expression of the transcription factor sox10. *BMC Dev Biol*, 8:105, 2008.
- [110] J. F. Colas and G. C. Schoenwolf. Towards a cellular and molecular understanding of neurulation. *Dev Dyn*, 221(2):117–45, 2001.
- [111] M. L. Martik and M. E. Bronner. Regulatory logic underlying diversification of the neural crest. *Trends Genet*, 33(10):715–727, 2017.
- [112] A. Mayran, K. Khetchoumian, F. Hariri, T. Pastinen, Y. Gauthier, A. Balsalobre, and J. Drouin. Pioneer factor pax7 deploys a stable enhancer repertoire for specification of cell fate. *Nat Genet*, 50(2):259–269, 2018.
- [113] 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 inÂ vivo. *Dev Cell*, 47(5):608–628.e6, 2018.

- [114] M. Fernandez Garcia, C. D. Moore, K. N. Schulz, O. Alberto, G. Donague, M. M. Harrison, H. Zhu, and K. S. Zaret. Structural features of transcription factors associating with nucleosome binding. *Mol Cell*, 75(5):921–932.e6, 2019.
- [115] W. D. Wang, D. B. Melville, M. Montero-Balaguer, A. K. Hatzopoulos, and E. W. Knapik. Tfp2a and foxd3 regulate early steps in the development of the neural crest progenitor population. *Dev Biol*, 360(1):173–85, 2011.
- [116] R. Soldatov, M. Kaucka, M. E. Kastriti, J. Petersen, T. Chontorotzea, L. Englmaier, N. Akkuratova, Y. Yang, M. Haring, 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.
- [117] R. M. Williams, I. Candido-Ferreira, E. Repapi, D. Gavriouchkina, U. Senanayake, J. Telenius, S. Taylor, J. Hughes, and T. Sauka-Spengler. Reconstruction of the global neural crest gene regulatory network in vivo. *bioRxiv*, page 508473, 2018.
- [118] 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.
- [119] 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.
- [120] 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.
- [121] S. Perez-Alcala, M. A. Nieto, and J. A. Barbas. Lsox5 regulates rhob expression in the neural tube and promotes generation of the neural crest. *Development*, 131(18):4455–65, 2004.

- [122] 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.
- [123] E. Thevenneau, J. L. Duband, and M. Altabef. Ets-1 confers cranial features on neural crest delamination. *PLoS One*, 2(11):e1142, 2007.
- [124] J. Vallin, R. Thuret, E. Giacomello, M. M. Faraldo, J. P. Thiery, and F. Broders. Cloning and characterization of three xenopus slug promoters reveal direct regulation by lef/beta-catenin signaling. *J Biol Chem*, 276(32):30350–8, 2001.
- [125] J. I. Yook, X. Y. Li, I. Ota, C. Hu, H. S. Kim, N. H. Kim, S. Y. Cha, J. K. Ryu, Y. J. Choi, J. Kim, E. R. Fearon, and S. J. Weiss. A wnt-axin2-gsk3beta cascade regulates snail1 activity in breast cancer cells. *Nat Cell Biol*, 8(12):1398–406, 2006.
- [126] E. H. Barriga, P. H. Maxwell, A. E. Reyes, and R. Mayor. The hypoxia factor hif-1a controls neural crest chemotaxis and epithelial to mesenchymal transition. *J Cell Biol*, 201(5):759–76, 2013.
- [127] C. D. Rogers, A. Saxena, and M. E. Bronner. Sip1 mediates an e-cadherin-to-n-cadherin switch during cranial neural crest emt. *J Cell Biol*, 203(5):835–47, 2013.
- [128] M. Dottori, M. K. Gross, P. Labosky, and M. Goulding. The winged-helix transcription factor foxd3 suppresses interneuron differentiation and promotes neural crest cell fate. *Development*, 128(21):4127–38, 2001.
- [129] C. L. Fairchild and L. S. Gammill. Tetraspanin18 is a foxd3-responsive antagonist of cranial neural crest epithelial-to-mesenchymal transition that maintains cadherin-6b protein. *J Cell Sci*, 126(Pt 6):1464–76, 2013.
- [130] C. L. Fairchild, J. P. Conway, A. T. Schiffmacher, L. A. Taneyhill, and L. S. Gammill. Foxd3 regulates cranial neural crest emt via downregulation of tetraspanin18 independent of its functions during neural crest formation. *Mech Dev*, 132:1–12, 2014.

- [131] C. McCusker, H. Cousin, R. Neuner, and D. Alfandari. Extracellular cleavage of cadherin-11 by adam metalloproteases is essential for xenopus cranial neural crest cell migration. *Mol Biol Cell*, 20(1):78–89, 2009.
- [132] I. Shoval, A. Ludwig, and C. Kalcheim. Antagonistic roles of full-length n-cadherin and its soluble bmp cleavage product in neural crest delamination. *Development*, 134(3):491–501, 2007.
- [133] L. Roth, R. Kalev-Altman, E. Monsonego-Ornan, and D. Sela-Donenfeld. A new role of the membrane-type matrix metalloproteinase 16 (mmp16/mt3-mmp) in neural crest cell migration. *Int J Dev Biol*, 61(3-4-5):245–256, 2017.
- [134] R. Mayor and E. Theveneau. The neural crest. *Development*, 140(11):2247–51, 2013.
- [135] A. Smith, V. Robinson, K. Patel, and D. G. Wilkinson. The epha4 and ephb1 receptor tyrosine kinases and ephrin-b2 ligand regulate targeted migration of branchial neural crest cells. *Curr Biol*, 7(8):561–70, 1997.
- [136] E. C. Olesnick Killian, D. A. Birkholz, and K. B. Artinger. A role for chemokine signaling in neural crest cell migration and craniofacial development. *Dev Biol*, 333(1):161–72, 2009.
- [137] Y. Kubota and K. Ito. Chemotactic migration of mesencephalic neural crest cells in the mouse. *Dev Dyn*, 217(2):170–9, 2000.
- [138] N. Trokovic, R. Trokovic, and J. Partanen. Fibroblast growth factor signalling and regional specification of the pharyngeal ectoderm. *Int J Dev Biol*, 49(7):797–805, 2005.
- [139] R. McLennan, J. M. Teddy, J. C. Kasemeier-Kulesa, M. H. Romine, and P. M. Kulesa. Vascular endothelial growth factor (vegf) regulates cranial neural crest migration in vivo. *Dev Biol*, 339(1):114–25, 2010.

- [140] H. M. Young, C. J. Hearn, P. G. Farlie, A. J. Canty, P. Q. Thomas, and D. F. Newgreen. Gdnf is a chemoattractant for enteric neural cells. *Dev Biol*, 229(2):503–16, 2001.
- [141] R. N. Kelsh, M. L. Harris, S. Colanesi, and C. A. Erickson. Stripes and belly-spots – a review of pigment cell morphogenesis in vertebrates. *Semin Cell Dev Biol*, 20(1):90–104, 2009.
- [142] B. R. Kuo and C. A. Erickson. Regional differences in neural crest morphogenesis. *Cell Adh Migr*, 4(4):567–85, 2010.
- [143] L. Jia, L. Cheng, and J. Raper. Slit/robo signaling is necessary to confine early neural crest cells to the ventral migratory pathway in the trunk. *Dev Biol*, 282(2):411–21, 2005.
- [144] C. Carmona-Fontaine, H. K. Matthews, S. Kuriyama, M. Moreno, G. A. Dunn, M. Parsons, C. D. Stern, and R. Mayor. Contact inhibition of locomotion in vivo controls neural crest directional migration. *Nature*, 456(7224):957–61, 2008.
- [145] E. Theveneau, L. Marchant, S. Kuriyama, M. Gull, B. Moepps, M. Parsons, and R. Mayor. Collective chemotaxis requires contact-dependent cell polarity. *Dev Cell*, 19(1):39–53, 2010.
- [146] C. Carmona-Fontaine, E. Theveneau, A. Tzekou, M. Tada, M. Woods, K. M. Page, M. Parsons, J. D. Lambris, and R. Mayor. Complement fragment c3a controls mutual cell attraction during collective cell migration. *Dev Cell*, 21(6):1026–37, 2011.
- [147] Y. Zhu, S. C. Crowley, A. J. Latimer, G. M. Lewis, R. Nash, and S. Kucenas. Migratory neural crest cells phagocytose dead cells in the developing nervous system. *Cell*, 179(1):74–89.e10, 2019.
- [148] N. M. Le Douarin. *Particularites du noyau interphasique chez la caille Japonaise (Coturnix coturnix japonica)*. 1969.

- [149] C. Le Lievre. Role des cellules mesectodermiques issues des cretes neurales cephaliques dans la formation des arcs branchiaux et du squelette visceral. *Journal of Embryology and Experimental Morphology*, 31(2):453–477, 1974.
- [150] 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.
- [151] M. Simoes-Costa and M. E. Bronner. Reprogramming of avian neural crest axial identity and cell fate. *Science*, 352(6293):1570–3, 2016.
- [152] T. L. Creazzo, R. E. Godt, L. Leatherbury, S. J. Conway, and M. L. Kirby. Role of cardiac neural crest cells in cardiovascular development. *Annu Rev Physiol*, 60:267–86, 1998.
- [153] T. J. Frith, I. Granata, M. Wind, E. Stout, O. Thompson, K. Neumann, D. Stavish, P. R. Heath, D. Ortmann, J. O. Hackland, K. Anastassiadis, M. Gouti, J. Briscoe, V. Wilson, S. L. Johnson, M. Placzek, M. R. Guarracino, P. W. Andrews, and A. Tsakiridis. Human axial progenitors generate trunk neural crest cells in vitro. *Elife*, 7, 2018.
- [154] 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.
- [155] H. C. Etchevers, J. Amiel, and S. Lyonnet. Molecular bases of human neuro-cristopathies. *Adv Exp Med Biol*, 589:213–34, 2006.
- [156] A. Keyte and M. R. Hutson. The neural crest in cardiac congenital anomalies. *Differentiation*, 84(1):25–40, 2012.
- [157] P. A. Trainor. Craniofacial birth defects: The role of neural crest cells in the etiology and pathogenesis of treacher collins syndrome and the potential for prevention. *Am J Med Genet A*, 152A(12):2984–94, 2010.

- [158] R. S. Greenberg, H. K. Long, T. Swigut, and J. Wysocka. Single amino acid change underlies distinct roles of h2a.z subtypes in human syndrome. *Cell*, 178(6):1421–1436.e24, 2019.
- [159] K. L. Gallik, R. W. Treffy, L. M. Nacke, K. Ahsan, M. Rocha, A. Green-Saxena, and A. Saxena. Neural crest and cancer: Divergent travelers on similar paths. *Mech Dev*, 148:89–99, 2017.
- [160] L. A. Hanna, R. K. Foreman, I. A. Tarasenko, D. S. Kessler, and P. A. Labosky. Requirement for foxd3 in maintaining pluripotent cells of the early mouse embryo. *Genes Dev*, 16(20):2650–61, 2002.
- [161] D. M. Tompers, R. K. Foreman, Q. Wang, M. Kumanova, and P. A. Labosky. Foxd3 is required in the trophoblast progenitor cell lineage of the mouse embryo. *Dev Biol*, 285(1):126–37, 2005.
- [162] L. Teng, N. A. Mundell, A. Y. Frist, Q. Wang, and P. A. Labosky. Requirement for foxd3 in the maintenance of neural crest progenitors. *Development*, 135(9):1615–24, 2008.
- [163] Y. Liu and P. A. Labosky. Regulation of embryonic stem cell self-renewal and pluripotency by foxd3. *Stem Cells*, 26(10):2475–84, 2008.
- [164] J. L. Plank-Bazinet and N. A. Mundell. The paradox of foxd3: how does it function in pluripotency and differentiation of embryonic stem cells? *Stem Cell Investig*, 3:73, 2016.
- [165] P. Respuela, M. Nikolic, M. Tan, P. Frommolt, Y. Zhao, J. Wysocka, and A. Rada-Iglesias. Foxd3 promotes exit from naive pluripotency through enhancer decommitment and inhibits germline specification. *Cell Stem Cell*, 18(1):118–33, 2016.
- [166] R. Krishnakumar, A. F. Chen, M. G. Pantovich, M. Danial, R. J. Parchem, P. A. Labosky, and R. Blelloch. Foxd3 regulates pluripotent stem cell potential by simultaneously initiating and repressing enhancer activity. *Cell Stem Cell*, 23(2):306–307, 2018.

- [167] R. A. Stewart, B. L. Arduini, S. Berghmans, R. E. George, J. P. Kanki, P. D. Henion, and A. T. Look. Zebrafish *foxd3* is selectively required for neural crest specification, migration and survival. *Dev Biol*, 292(1):174–88, 2006.
- [168] M. Montero-Balaguer, M. R. Lang, S. W. Sachdev, C. Knappmeyer, R. A. Stewart, A. De La Guardia, A. K. Hatzopoulos, and E. W. Knapik. The mother superior mutation ablates *foxd3* activity in neural crest progenitor cells and depletes neural crest derivatives in zebrafish. *Dev Dyn*, 235(12):3199–212, 2006.
- [169] J. A. Lister, C. Cooper, K. Nguyen, M. Modrell, K. Grant, and D. W. Raible. Zebrafish *foxd3* is required for development of a subset of neural crest derivatives. *Dev Biol*, 290(1):92–104, 2006.
- [170] 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.
- [171] 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):585–95, 2010.
- [172] E. Nitzan, E. R. Pfaltzgraff, P. A. Labosky, and C. Kalcheim. Neural crest and schwann cell progenitor-derived melanocytes are two spatially segregated populations similarly regulated by *foxd3*. *Proc Natl Acad Sci U S A*, 110(31):12709–14, 2013.
- [173] A. J. Thomas and C. A. Erickson. *Foxd3* regulates the lineage switch between neural crest-derived glial cells and pigment cells by repressing *mitf* through a non-canonical mechanism. *Development*, 136(11):1849–58, 2009.
- [174] R. Kos, M. V. Reedy, R. L. Johnson, and C. A. Erickson. The winged-helix transcription factor *foxd3* is important for establishing the neural crest lineage and repressing melanogenesis in avian embryos. *Development*, 128(8):1467–79, 2001.
- [175] K. Curran, J. A. Lister, G. R. Kunkel, A. Prendergast, D. M. Parichy, and D. W.

- Raible. Interplay between *foxd3* and *mitf* regulates cell fate plasticity in the zebrafish neural crest. *Dev Biol*, 344(1):107–18, 2010.
- [176] N. A. Mundell, J. L. Plank, A. W. LeGrone, A. Y. Frist, L. Zhu, M. K. Shin, E. M. Southard-Smith, and P. A. Labosky. Enteric nervous system specific deletion of *foxd3* disrupts glial cell differentiation and activates compensatory enteric progenitors. *Dev Biol*, 363(2):373–87, 2012.
- [177] S. Yaklichkin, A. B. Steiner, Q. Lu, and D. S. Kessler. *Foxd3* and *grg4* physically interact to repress transcription and induce mesoderm in xenopus. *J Biol Chem*, 282(4):2548–57, 2007.
- [178] D. Liber, R. Domaschensz, P. H. Holmqvist, L. Mazzearella, A. Georgiou, M. Leleu, A. G. Fisher, P. A. Labosky, and N. Dillon. Epigenetic priming of a pre-b cell-specific enhancer through binding of *sox2* and *foxd3* at the esc stage. *Cell Stem Cell*, 7(1):114–26, 2010.
- [179] D. J. Sweet. *Foxd3*: A repressor, an activator, or both? *Cell Stem Cell*, 18(1):1–2, 2016.
- [180] T. Chen and S. Y. Dent. Chromatin modifiers and remodellers: regulators of cellular differentiation. *Nat Rev Genet*, 15(2):93–106, 2014.
- [181] F. Spitz and E. E. Furlong. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet*, 13(9):613–26, 2012.
- [182] K. S. Zaret and J. S. Carroll. Pioneer transcription factors: establishing competence for gene expression. *Genes Dev*, 25(21):2227–41, 2011.
- [183] D. Y. R. Stainier, E. Raz, N. D. Lawson, S. C. Ekker, R. D. Burdine, J. S. Eisen, P. W. Ingham, S. Schulte-Merker, D. Yelon, B. M. Weinstein, M. C. Mullins, S. W. Wilson, L. Ramakrishnan, S. L. Amacher, S. C. F. Neuhauss, A. Meng, N. Mochizuki, P. Panula, and C. B. Moens. Guidelines for morpholino use in zebrafish. *PLoS Genet*, 13(10):e1007000, 2017.

- [184] M. A. El-Brolosy, Z. Kontarakis, A. Rossi, C. Kuenne, S. Gunther, N. Fukuda, K. Kikhi, G. L. M. Boezio, C. M. Takacs, S. L. Lai, R. Fukuda, C. Gerri, A. J. Giraldez, and D. Y. R. Stainier. Genetic compensation triggered by mutant mrna degradation. *Nature*, 568(7751):193–197, 2019.
- [185] L. L. Chang and D. S. Kessler. Foxd3 is an essential nodal-dependent regulator of zebrafish dorsal mesoderm development. *Dev Biol*, 342(1):39–50, 2010.
- [186] 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.
- [187] I. A. Trinh, T. Hochgreb, M. Graham, D. Wu, F. Ruf-Zamojski, C. S. Jayasena, A. Saxena, R. Hawk, A. Gonzalez-Serricchio, A. Dixon, E. Chow, C. Gonzales, H. Y. Leung, I. Solomon, M. Bronner-Fraser, S. G. Megason, and S. E. Fraser. A versatile gene trap to visualize and interrogate the function of the vertebrate proteome. *Genes Dev*, 25(21):2306–20, 2011.
- [188] W. Li and R. A. Cornell. Redundant activities of tfap2a and tfap2c are required for neural crest induction and development of other non-neural ectoderm derivatives in zebrafish embryos. *Dev Biol*, 304(1):338–54, 2007.
- [189] D. R. Powell, L. Hernandez-Lagunas, K. LaMonica, and K. B. Artinger. Prdm1a directly activates foxd3 and tfap2a during zebrafish neural crest specification. *Development*, 140(16):3445–55, 2013.
- [190] L. A. Trinh, V. Chong-Morrison, D. Gavriouchkina, T. Hochgreb-Hagele, 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 Rep*, 19(2):425–440, 2017.
- [191] S. Chatterjee and T. Lufkin. Fishing for function: zebrafish bac transgenics for functional genomics. *Mol Biosyst*, 7(8):2345–51, 2011.

- [192] M. Iwafuchi-Doi. The mechanistic basis for chromatin regulation by pioneer transcription factors. *Wiley Interdiscip Rev Syst Biol Med*, 11(1):e1427, 2019.
- [193] T. Ye, A. R. Krebs, M. A. Choukrallah, C. Keime, F. Plewniak, I. Davidson, and L. Tora. seqminer: an integrated chip-seq data interpretation platform. *Nucleic Acids Res*, 39(6):e35, 2011.
- [194] F. Beby and T. Lamonerie. The homeobox gene *otx2* in development and disease. *Exp Eye Res*, 111:9–16, 2013.
- [195] T. J. Carney, K. A. Dutton, E. Greenhill, M. Delfino-Machin, P. Dufourcq, P. Blader, and R. N. Kelsh. A direct role for *sox10* in specification of neural crest-derived sensory neurons. *Development*, 133(23):4619–30, 2006.
- [196] C. T. Dee, C. S. Hirst, Y. H. Shih, V. B. Tripathi, R. K. Patient, and P. J. Scotting. *Sox3* regulates both neural fate and differentiation in the zebrafish ectoderm. *Dev Biol*, 320(1):289–301, 2008.
- [197] S. Gaubatz, J. G. Wood, and D. M. Livingston. Unusual proliferation arrest and transcriptional control properties of a newly discovered *e2f* family member, *e2f-6*. *Proc Natl Acad Sci U S A*, 95(16):9190–5, 1998.
- [198] M. Leseva, K. E. Santostefano, A. L. Rosenbluth, T. Hamazaki, and N. Terada. *E2f6*-mediated repression of the meiotic *stag3* and *smc1b* genes during early embryonic development requires *ezh2* and not the de novo methyltransferase *dnmt3b*. *Epigenetics*, 8(8):873–84, 2013.
- [199] J. M. Trimarchi, B. Fairchild, J. Wen, and J. A. Lees. The *e2f6* transcription factor is a component of the mammalian *bmi1*-containing polycomb complex. *Proc Natl Acad Sci U S A*, 98(4):1519–24, 2001.
- [200] M. Kato, N. Hata, N. Banerjee, B. Futcher, and M. Q. Zhang. Identifying combinatorial regulation of transcription factors and binding motifs. *Genome Biol*, 5(8):R56, 2004.

- [201] J. Xu, S. D. Pope, A. R. Jazirehi, J. L. Attema, P. Papathanasiou, J. A. Watts, K. S. Zaret, I. L. Weissman, and S. T. Smale. Pioneer factor interactions and unmethylated cpg dinucleotides mark silent tissue-specific enhancers in embryonic stem cells. *Proc Natl Acad Sci U S A*, 104(30):12377–82, 2007.
- [202] S. Kumar. Remote homologue identification of drosophila gaga factor in mouse. *Bioinformation*, 7(1):29–32, 2011.
- [203] B. Ye, B. Liu, L. Yang, X. Zhu, D. Zhang, W. Wu, P. Zhu, Y. Wang, S. Wang, P. Xia, Y. Du, S. Meng, G. Huang, J. Wu, R. Chen, Y. Tian, and Z. Fan. controls self-renewal of embryonic stem cells via activating expression of transcription factor. *EMBO J*, 37(8), 2018.
- [204] N. L. Adkins, T. A. Hagerman, and P. Georgel. Gaga protein: a multi-faceted transcription factor. *Biochem Cell Biol*, 84(4):559–67, 2006.
- [205] S. Kadam and B. M. Emerson. Transcriptional specificity of human swi/snf brg1 and brm chromatin remodeling complexes. *Mol Cell*, 11(2):377–89, 2003.
- [206] S. Ross, E. Cheung, T. G. Petrakis, M. Howell, W. L. Kraus, and C. S. Hill. Smads orchestrate specific histone modifications and chromatin remodeling to activate transcription. *EMBO J*, 25(19):4490–502, 2006.
- [207] L. Zhang, X. Ju, Y. Cheng, X. Guo, and T. Wen. Identifying tmem59 related gene regulatory network of mouse neural stem cell from a compendium of expression profiles. *BMC Syst Biol*, 5:152, 2011.
- [208] C. Attwooll, E. Lazzerini Denchi, and K. Helin. The e2f family: specific functions and overlapping interests. *EMBO J*, 23(24):4709–16, 2004.
- [209] C. Attwooll, S. Oddi, P. Cartwright, E. Prosperini, K. Agger, P. Steensgaard, C. Wagener, C. Sardet, M. C. Moroni, and K. Helin. A novel repressive e2f6 complex containing the polycomb group protein, epc1, that interacts with ezh2 in a proliferation-specific manner. *J Biol Chem*, 280(2):1199–208, 2005.

- [210] H. Ogawa, K. Ishiguro, S. Gaubatz, D. M. Livingston, and Y. Nakatani. A complex with chromatin modifiers that occupies e2f- and myc-responsive genes in g0 cells. *Science*, 296(5570):1132–6, 2002.
- [211] D. Jukam, S. A. M. Shariati, and J. M. Skotheim. Zygotic genome activation in vertebrates. *Dev Cell*, 42(4):316–332, 2017.
- [212] M. Tamura, T. Amano, and T. Shiroishi. The hand2 gene dosage effect in developmental defects and human congenital disorders. *Curr Top Dev Biol*, 110:129–52, 2014.
- [213] D. Y. Nishimura, C. C. Searby, W. L. Alward, D. Walton, J. E. Craig, D. A. Mackey, K. Kawase, A. B. Kanis, S. R. Patil, E. M. Stone, and V. C. Sheffield. A spectrum of foxc1 mutations suggests gene dosage as a mechanism for developmental defects of the anterior chamber of the eye. *Am J Hum Genet*, 68(2):364–72, 2001.
- [214] S. T. Crews and J. C. Pearson. Transcriptional autoregulation in development. *Curr Biol*, 19(6):R241–6, 2009.
- [215] D. K. Papadopoulos, K. Skouloudaki, Y. Engstrom, L. Terenius, R. Rigler, C. Zechner, V. Vukojevic, and P. Tomancak. Control of hox transcription factor concentration and cell-to-cell variability by an auto-regulatory switch. *Development*, 146(12), 2019.
- [216] R. P. Ngondo and P. Carbon. Transcription factor abundance controlled by an auto-regulatory mechanism involving a transcription start site switch. *Nucleic Acids Res*, 42(4):2171–84, 2014.
- [217] W. Han, Y. Yu, K. Su, R. A. Kohanski, and L. Pick. A binding site for multiple transcriptional activators in the fushi tarazu proximal enhancer is essential for gene expression in vivo. *Mol Cell Biol*, 18(6):3384–94, 1998.
- [218] M. Martinez, M. Hinojosa, D. Trombly, V. Morin, J. Stein, G. Stein, A. Javed, and S. E. Gutierrez. Transcriptional auto-regulation of runx1 p1 promoter. *PLoS One*, 11(2):e0149119, 2016.

- [219] S. Aota, N. Nakajima, R. Sakamoto, S. Watanabe, N. Ibaraki, and K. Okazaki. Pax6 autoregulation mediated by direct interaction of pax6 protein with the head surface ectoderm-specific enhancer of the mouse pax6 gene. *Dev Biol*, 257(1):1–13, 2003.
- [220] J. Johnston, R. J., S. Chang, J. F. Etchberger, C. O. Ortiz, and O. Hobert. MicroRNAs acting in a double-negative feedback loop to control a neuronal cell fate decision. *Proc Natl Acad Sci U S A*, 102(35):12449–54, 2005.
- [221] S. Palani and C. A. Sarkar. Transient noise amplification and gene expression synchronization in a bistable mammalian cell-fate switch. *Cell Rep*, 1(3):215–24, 2012.
- [222] N. T. Ingolia and A. W. Murray. Positive-feedback loops as a flexible biological module. *Curr Biol*, 17(8):668–77, 2007.
- [223] R. Thomas and M. Kaufman. Multistationarity, the basis of cell differentiation and memory. i. structural conditions of multistationarity and other nontrivial behavior. *Chaos*, 11(1):170–179, 2001.
- [224] A. De Siervi, P. De Luca, J. S. Byun, L. J. Di, T. Fufa, C. M. Haggerty, E. Vazquez, C. Moiola, D. L. Longo, and K. Gardner. Transcriptional autoregulation by brca1. *Cancer Res*, 70(2):532–42, 2010.
- [225] N. Rosenfeld, M. B. Elowitz, and U. Alon. Negative autoregulation speeds the response times of transcription networks. *J Mol Biol*, 323(5):785–93, 2002.
- [226] A. Becskei and L. Serrano. Engineering stability in gene networks by autoregulation. *Nature*, 405(6786):590–3, 2000.
- [227] Y. Dublanche, K. Michalodimitrakis, N. Kummerer, M. Foglierini, and L. Serrano. Noise in transcription negative feedback loops: simulation and experimental analysis. *Mol Syst Biol*, 2:41, 2006.

- [228] H. Hirata, S. Yoshiura, T. Ohtsuka, Y. Bessho, T. Harada, K. Yoshikawa, and R. Kageyama. Oscillatory expression of the bhlh factor *hes1* regulated by a negative feedback loop. *Science*, 298(5594):840–3, 2002.
- [229] A. C. Oates, L. G. Morelli, and S. Ares. Patterning embryos with oscillations: structure, function and dynamics of the vertebrate segmentation clock. *Development*, 139(4):625–39, 2012.
- [230] J. Yan, C. Anderson, K. Viets, S. Tran, G. Goldberg, S. Small, and J. Johnston, R. J. Regulatory logic driving stable levels of defective proventriculus expression during terminal photoreceptor specification in flies. *Development*, 144(5):844–855, 2017.
- [231] S. M. Kielbasa and M. Vingron. Transcriptional autoregulatory loops are highly conserved in vertebrate evolution. *PLoS One*, 3(9):e3210, 2008.
- [232] J. J. Riethoven. Regulatory regions in dna: promoters, enhancers, silencers, and insulators. *Methods Mol Biol*, 674:33–42, 2010.
- [233] H. K. Long, S. L. Prescott, and J. Wysocka. Ever-changing landscapes: Transcriptional enhancers in development and evolution. *Cell*, 167(5):1170–1187, 2016.
- [234] V. Chong-Morrison, F. C. Simoes, U. Senanayake, D. S. Carroll, P. R. Riley, and T. Sauka-Spengler. Re-purposing ac/ds transgenic system for crispr/dcas9 modulation of enhancers and non-coding rnas in zebrafish. *bioRxiv*, page 450684, 2018.
- [235] 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.
- [236] A. Khan, O. Fornes, A. Stigliani, M. Gheorghe, J. A. Castro-Mondragon, R. van der Lee, A. Bessy, J. Cheneby, S. R. Kulkarni, G. Tan, D. Baranasic, D. J. Arenillas, A. Sandelin, K. Vandepoele, B. Lenhard, B. Ballester, W. W. Wasserman, F. Parcy, and A. Mathelier. Jaspas 2018: update of the open-access database of transcription

- factor binding profiles and its web framework. *Nucleic Acids Res*, 46(D1):D1284, 2018.
- [237] I. V. Kulakovskiy, I. E. Vorontsov, I. S. Yevshin, R. N. Sharipov, A. D. Fedorova, E. I. Rumynskiy, Y. A. Medvedeva, A. Magana-Mora, V. B. Bajic, D. A. Papatzenko, F. A. Kolpakov, and V. J. Makeev. Hocomoco: towards a complete collection of transcription factor binding models for human and mouse via large-scale chip-seq analysis. *Nucleic Acids Res*, 46(D1):D252–D259, 2018.
- [238] J. M. Newbern. Molecular control of the neural crest and peripheral nervous system development. *Curr Top Dev Biol*, 111:201–31, 2015.
- [239] V. Sasselli, V. Pachnis, and A. J. Burns. The enteric nervous system. *Dev Biol*, 366(1):64–73, 2012.
- [240] K. Shyamala, S. Yanduri, H. C. Girish, and S. Murgod. Neural crest: The fourth germ layer. *J Oral Maxillofac Pathol*, 19(2):221–9, 2015.
- [241] P. E. Ludwig and M. Varacallo. *Neuroanatomy, Central Nervous System (CNS)*. StatPearls Publishing StatPearls Publishing LLC., Treasure Island (FL), 2019.
- [242] E. Tzouanacou, A. Wegener, F. J. Wymeersch, V. Wilson, and J. F. Nicolas. Redefining the progression of lineage segregations during mammalian embryogenesis by clonal analysis. *Dev Cell*, 17(3):365–76, 2009.
- [243] D. Henrique, E. Abranches, L. Verrier, and K. G. Storey. Neuromesodermal progenitors and the making of the spinal cord. *Development*, 142(17):2864–75, 2015.
- [244] N. Cambray and V. Wilson. Two distinct sources for a population of maturing axial progenitors. *Development*, 134(15):2829–40, 2007.
- [245] M. J. McGrew, A. Sherman, S. G. Lillico, F. M. Ellard, P. A. Radcliffe, H. J. Gilhooley, K. A. Mitrophanous, N. Cambray, V. Wilson, and H. Sang. Localised axial progenitor cell populations in the avian tail bud are not committed to a posterior hox identity. *Development*, 135(13):2289–99, 2008.

- [246] B. Steventon and A. Martinez Arias. Evo-engineering and the cellular and molecular origins of the vertebrate spinal cord. *Dev Biol*, 432(1):3–13, 2017.
- [247] B. L. Martin and D. Kimelman. Canonical wnt signaling dynamically controls multiple stem cell fate decisions during vertebrate body formation. *Dev Cell*, 22(1):223–32, 2012.
- [248] A. Attardi, T. Fulton, M. Florescu, G. Shah, L. Muresan, M. O. Lenz, C. Lancaster, J. Huiskens, A. van Oudenaarden, and B. Steventon. Neuromesodermal progenitors are a conserved source of spinal cord with divergent growth dynamics (doi: 10.1242/dev.166728). *Development*, 145(121), 2018.
- [249] F. J. Wymeersch, Y. Huang, G. Blin, N. Cambray, R. Wilkie, F. C. Wong, and V. Wilson. Position-dependent plasticity of distinct progenitor types in the primitive streak. *Elife*, 5:e10042, 2016.
- [250] A. Tsakiridis, Y. Huang, G. Blin, S. Skylaki, F. Wymeersch, R. Osorno, C. Economou, E. Karagianni, S. Zhao, S. Lowell, and V. Wilson. Distinct wnt-driven primitive streak-like populations reflect in vivo lineage precursors. *Development*, 142(4):809, 2015.
- [251] 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.
- [252] F. Koch, M. Scholze, L. Wittler, D. Schifferl, S. Sudheer, P. Grote, B. Timmermann, K. Macura, and B. G. Herrmann. Antagonistic activities of sox2 and brachyury control the fate choice of neuro-mesodermal progenitors. *Dev Cell*, 42(5):514–526.e7, 2017.
- [253] B. J. Swalla. Building divergent body plans with similar genetic pathways. *Heredity (Edinb)*, 97(3):235–43, 2006.

- [254] 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.
- [255] 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.
- [256] T. F. Schilling, V. Prince, and P. W. Ingham. Plasticity in zebrafish hox expression in the hindbrain and cranial neural crest. *Dev Biol*, 231(1):201–16, 2001.
- [257] P. Philippidou and J. S. Dasen. Hox genes: choreographers in neural development, architects of circuit organization. *Neuron*, 80(1):12–34, 2013.
- [258] 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.
- [259] V. Ribes, I. Le Roux, M. Rhinn, B. Schuhbaur, and P. Dolle. Early mouse caudal development relies on crosstalk between retinoic acid, shh and fgf signalling pathways. *Development*, 136(4):665–76, 2009.
- [260] J. M. Brown and K. G. Storey. A region of the vertebrate neural plate in which neighbouring cells can adopt neural or epidermal fates. *Curr Biol*, 10(14):869–72, 2000.
- [261] T. Iimura and O. Pourquie. Collinear activation of hoxb genes during gastrulation is linked to mesoderm cell ingression. *Nature*, 442(7102):568–71, 2006.
- [262] I. Olivera-Martinez, H. Harada, P. A. Halley, and K. G. Storey. Loss of fgf-dependent mesoderm identity and rise of endogenous retinoid signalling determine cessation of body axis elongation. *PLoS Biol*, 10(10):e1001415, 2012.

- [263] H. Kondoh and T. Takemoto. Axial stem cells deriving both posterior neural and mesodermal tissues during gastrulation. *Curr Opin Genet Dev*, 22(4):374–80, 2012.
- [264] M. Delfino-Machin, J. S. Lunn, D. N. Breitkreuz, J. Akai, and K. G. Storey. Specification and maintenance of the spinal cord stem zone. *Development*, 132(19):4273–83, 2005.
- [265] P. D. Nieuwkoop and G. v. Nigtevecht. Neural activation and transformation in explants of competent ectoderm under the influence of fragments of anterior notochord in urodeles. *Journal of Embryology and Experimental Morphology*, 2(3):175–193, 1954.
- [266] C. D. Stern. Neural induction: 10 years on since the 'default model'. *Curr Opin Cell Biol*, 18(6):692–7, 2006.
- [267] R. J. Garriock, R. B. Chalamalasetty, M. W. Kennedy, L. C. Canizales, M. Lewandoski, and T. P. Yamaguchi. Lineage tracing of neuromesodermal progenitors reveals novel wnt-dependent roles in trunk progenitor cell maintenance and differentiation. *Development*, 142(9):1628–38, 2015.
- [268] 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.
- [269] 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.
- [270] M. Berenguer, J. J. Lancman, T. J. Cunningham, P. D. S. Dong, and G. Duester. Mouse but not zebrafish requires retinoic acid for control of neuromesodermal progenitors and body axis extension. *Dev Biol*, 441(1):127–131, 2018.
- [271] R. Fior, A. A. Maxwell, T. P. Ma, A. Vezzaro, C. B. Moens, S. L. Amacher, J. Lewis, and L. Saude. The differentiation and movement of presomitic mesoderm

- progenitor cells are controlled by mesogenin 1. *Development*, 139(24):4656–65, 2012.
- [272] V. Metzis, S. Steinhauser, E. Pakanavicius, M. Gouti, D. Stamataki, K. Ivanovitch, T. Watson, T. Rayon, S. N. Mousavy Gharavy, R. Lovell-Badge, N. M. Luscombe, and J. Briscoe. Nervous system regionalization entails axial allocation before neural differentiation. *Cell*, 175(4):1105–1118.e17, 2018.
- [273] B. Reversade, H. Kuroda, H. Lee, A. Mays, and E. M. De Robertis. Depletion of *bmp2*, *bmp4*, *bmp7* and *spemann* organizer signals induces massive brain formation in *xenopus* embryos. *Development*, 132(15):3381–92, 2005.
- [274] B. Reversade and E. M. De Robertis. Regulation of *admp* and *bmp2/4/7* at opposite embryonic poles generates a self-regulating morphogenetic field. *Cell*, 123(6):1147–60, 2005.
- [275] Y. Hayashi, S. Mikawa, C. Ogawa, K. Masumoto, F. Katou, and K. Sato. *Bmp6* expression in the adult rat central nervous system. *J Chem Neuroanat*, 98:41–54, 2019.
- [276] M. Gouti, A. Tsakiridis, F. J. Wymeersch, Y. Huang, J. Kleinjung, V. Wilson, and J. Briscoe. In vitro generation of neuromesodermal progenitors reveals distinct roles for wnt signalling in the specification of spinal cord and paraxial mesoderm identity. *PLoS Biol*, 12(8):e1001937, 2014.
- [277] A. Javali, A. Misra, K. Leonavicius, D. Acharyya, B. Vyas, and R. Sambasivan. Co-expression of *tbx6* and *sox2* identifies a novel transient neuromesoderm progenitor cell state. *Development*, 144(24):4522–4529, 2017.
- [278] A. Jolma, Y. Yin, K. R. Nitta, K. Dave, A. Popov, M. Taipale, M. Enge, T. Kivioja, E. Morgunova, and J. Taipale. Dna-dependent formation of transcription factor pairs alters their binding specificity. *Nature*, 527(7578):384–8, 2015.
- [279] J. M. Rogers, C. T. Waters, T. C. M. Seegar, S. M. Jarrett, A. N. Hallworth, S. C. Blacklow, and M. L. Bulyk. Bispecific forkhead transcription factor *foxn3*

- recognizes two distinct motifs with different dna shapes. *Mol Cell*, 74(2):245–253.e6, 2019.
- [280] F. Zhu, L. Farnung, E. Kaasinen, B. Sahu, Y. Yin, B. Wei, S. O. Dodonova, K. R. Nitta, E. Morgunova, M. Taipale, P. Cramer, and J. Taipale. The interaction landscape between transcription factors and the nucleosome. *Nature*, 562(7725):76–81, 2018.
- [281] P. J. Skene and S. Henikoff. An efficient targeted nuclease strategy for high-resolution mapping of dna binding sites. *Elife*, 6, 2017.
- [282] M. Westerfield. *The zebrafish book : a guide for the laboratory use of zebrafish (danio rerio)*. 4th ed. Univ. of Oregon Press, Eugene., Eugene, Or., 2000.
- [283] 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.
- [284] A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, and T. R. Gingeras. Star: ultrafast universal rna-seq aligner. *Bioinformatics*, 29(1):15–21, 2013.
- [285] Y. Liao, G. K. Smyth, and W. Shi. featurecounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, 30(7):923–30, 2014.
- [286] M. I. Love, W. Huber, and S. Anders. Moderated estimation of fold change and dispersion for rna-seq data with deseq2. *Genome Biol*, 15(12):550, 2014.
- [287] J. H. Kim, S. R. Lee, L. H. Li, H. J. Park, J. H. Park, K. Y. Lee, M. K. Kim, B. A. Shin, and S. Y. Choi. High cleavage efficiency of a 2a peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. *PLoS One*, 6(4):e18556, 2011.
- [288] 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):576–89, 2010.
- [289] Y. Zhang, T. Liu, C. A. Meyer, J. Eeckhoutte, D. S. Johnson, B. E. Bernstein, C. Nusbaum, R. M. Myers, M. Brown, W. Li, and X. S. Liu. Model-based analysis of chip-seq (macs). *Genome Biol*, 9(9):R137, 2008.
 - [290] A. Emelyanov, Y. Gao, N. I. Naqvi, and S. Parinov. Trans-kingdom transposition of the maize dissociation element. *Genetics*, 174(3):1095–104, 2006.
 - [291] F. Ramirez, D. P. Ryan, B. Gruning, V. Bhardwaj, F. Kilpert, A. S. Richter, S. Heyne, F. Dunder, and T. Manke. deeptools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res*, 44(W1):W160–5, 2016.
 - [292] G. X. Zheng, J. M. Terry, P. Belgrader, P. Ryvkin, Z. W. Bent, R. Wilson, S. B. Ziraldo, T. D. Wheeler, G. P. McDermott, J. Zhu, M. T. Gregory, J. Shuga, L. Montesclaros, J. G. Underwood, D. A. Masquelier, S. Y. Nishimura, M. Schnall-Levin, P. W. Wyatt, C. M. Hindson, R. Bharadwaj, A. Wong, K. D. Ness, L. W. Beppu, H. J. Deeg, C. McFarland, K. R. Loeb, W. J. Valente, N. G. Ericson, E. A. Stevens, J. P. Radich, T. S. Mikkelsen, B. J. Hindson, and J. H. Bielas. Massively parallel digital transcriptional profiling of single cells. *Nat Commun*, 8:14049, 2017.
 - [293] A. Butler, P. Hoffman, P. Smibert, E. Papalexi, and R. Satija. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol*, 36(5):411–420, 2018.
 - [294] B. Langmead, C. Trapnell, M. Pop, and S. L. Salzberg. Ultrafast and memory-efficient alignment of short dna sequences to the human genome. *Genome Biol*, 10(3):R25, 2009.
 - [295] X. Qiu, Q. Mao, Y. Tang, L. Wang, R. Chawla, H. A. Pliner, and C. Trapnell. Reversed graph embedding resolves complex single-cell trajectories. *Nat Methods*, 14(10):979–982, 2017.

Acknowledgments

Firstly, I would like to thank my superb mentor, Prof Tatjana Sauka-Spengler. Since the first day I met her, I was fascinated by her brilliant mind and passion for science. I cannot stress enough, how lucky I was to be supervised and supported by Prof Tatjana at every step of my DPhil. She has encouraged me, motivated me and provided me with so many opportunities to participate in numerous conferences, workshops and courses, such as the MBL Woods Hole Embryology course, which has truly reshaped my work habits and passion for developmental biology. Additionally, Prof Tatjana has helped me to build my self-confidence and made me believe that I could do so much more than I ever thought was possible.

I am also very grateful to the rest of the lab members. You have been the most supportive and intelligent group of people! By being surrounded by such great people, I have learnt that the most important thing in academia is to lift each other up and work for a common goal. I would especially like to thank Dr. Upeka Senanayake and Dr. Daria Gavriouchkina, who have provided me with the initial 'wet' lab and bioinformatics training, respectively. I would also like to thank Dr. Filipa Simoes, who is a wonderful colleague, inspiring role model, and a dear friend. And of course I would like to thank my partner in science crime, Ivan Candido-Ferreira, who is not only a pure genius, but is also a pure soul.

During my DPhil time, I was very blessed to meet many radiant people who have become good friends and made my time in Oxford unforgettable. First and foremost, I want to celebrate my friendship with the most inspiring and kind woman, Ioanna Rota, who has become like a sister to me. Our continuous philosophical discussions and strive to empower each other have truly made me a better person. I would also like to thank my good friends, Liliana Barbieri, Dr. Pablo Carravilla, Falk Schneider and Robin Lorenz, who were always there for me and brought joy and sanity during even the most insane and hard moments. Especially, I would like to thank Dr. Erdinc Sezgin, who was with me during my happiest and also hardest moments. Our rich conversations on life and cosmos have genuinely sparked extra curiosity and passion for science within me. Above all, he has helped me to shape my scientific and personal values, which, I am sure, will remain with me for the rest of my days and will help me to do great things.

I am also very grateful for the zebrafish facility team, who has done a brilliant job, provided continuous help and made my research possible.

Finally, I would like to thank my mum, Rima Lukoseviciene, who not for once had a doubt in me or my capabilities. I am also magnificently grateful to my beloved sister, Justina Lukoseviciute, and my best friend, Justina Stukaite. Our shared moments, adventures, laughter and selfless support to each other, have definitely made my DPhil experience lighter and even more enjoyable.