A study of cis-regulatory elements in *Xenopus* and planarians

Marta Marín-Barba

Thesis submitted for the degree of Doctor of Philosophy

University of East Anglia

School of Biological Sciences

Norwich

United Kingdom

September 2017

Word count: 43845

Abstract

Multicellular organisms have the ability to produce and maintain several cell types despite the DNA molecule being the same in all cells. For that, a complex network of cis-regulatory elements (CRE) and transcription factors is involved in controlling cellular fate. Among the different approaches that allow the identification of CREs, the novel ATAC-sequencing is becoming an increasingly used tool due to its easy-to-perform protocol. In this study, this technique has been optimized to be used in two species (*Xenopus laevis* and *Schmidtea mediterranea*) to elucidate (1) the epigenetic mechanism behind *Xenopus* neural crest (NC) specification and differentiation, and (2) the establishment of anteroposterior identity during planarian regeneration.

ATAC-sequencing has been carried out on *Xenopus* animal caps induced to give rise to NC or neural tissue. Comparison of this data along with non-injected ectodermal animal caps has led to the characterization of the epigenome involved in *Xenopus* neural crest formation. Results show that 4,528 NC open regions or potential enhancers are present during specification whereas 852 NC potential enhancers are specific for differentiation. In these NC open regions, the transcription factors Zic, Meis and Sox10 play an important role. Five enhancers driving expression of key neural crest genes (*cmyc*, *foxd3*, *id3*, *snai2* and *sox10*) have been identified.

ATAC-sequencing has also been successfully validated on dissected blastemas from 48h regenerating planarians. The analysis has revealed nervous system-related transcription factor motifs such as Prep or NeuroD-1 present in anterior specific open regions whereas posterior polarity-related transcription factor binding sites such as Islet or HoxD3 are present in posterior open regions.

Key words: ATAC-sequencing, enhancer, transcription factor, *Xenopus* neural crest, planarian, anteroposterior polarity.

Abstract……..……**ii**
List of contents..……**iii**
List of figures……**vii**
List of tables……..…..**ix**
Abbreviations……...**x**
Acknowledgments……..……..**xii**

List of contents

Chapter I: Introduction
1.1. Genome architecture……**2**
1.1.1. Histone modifications……...3
1.1.2. Genomic regulatory elements……...5
1.1.2.1. Enhancers……5
1.1.2.2. Promoters……...7
1.1.3. Evolution and conservation of cis-regulatory elements……....9
1.1.4. Empirical approaches to study cis-regulatory elements……..10
1.1.4.1. Identification of cis-regulatory elements……..……..11
1.1.4.2. Validation of cis-regulatory elements……...12
1.1.4.3. Validation of promoter-enhancer interaction……..13
1.2. *Xenopus* as a model organism……..**13**
1.2.1. *Xenopus* neural crest formation……..15
1.2.1.1. Neural crest induction……..……16
1.2.1.2. Neural crest specification……..…..18
1.2.1.3. Neural crest migration……19
1.2.1.4. Neural crest differentiation……21
1.2.2. Epigenetic regulation of the neural crest cells……..21
1.2.2.1. DNA and chromatin modifications related with NC……21
1.2.2.2. Cis-regulatory elements associated with NC……..24
1.3. *Schmidtea mediterranea* as a model organism……**25**
1.3.1. Neoblasts, the planarian stem cells……26
1.3.2. Planarian regeneration……..…28
1.3.2.1. Wound healing……28
1.3.2.1. Cell death……..…29
1.3.2.3. Neoblasts proliferation……30
1.3.2.4. Head vs tail regeneration……...30
1.3.3. Epigenetic regulation of the regeneration……..33

1.4. Aims and objectives……..34

Chapter II: Materials and methods

2.1. *Xenopus* materials and methods……..……..37

2.1.1. *X. tropicalis* embryo obtainment……..37

2.1.2. *X. laevis* embryo obtainment……...37

2.1.3. Dejellying embryos……38

2.1.4. Embryo fixation……..38

2.1.5. Chemical screen……38

2.1.6. Imaging……41

2.1.7. Transformation using DH5α cells……41

2.1.8. DNA Midiprep……41

2.1.9. Probe synthesis……41

2.1.10. Whole mount *in situ* hybridization (WISH)……..42

2.1.11. Capped RNA synthesis……..43

2.1.12. Animal cap assay……43

2.1.13. ATAC-sequencing……...43

2.1.14. RNA extraction and cDNA synthesis……44

2.1.15. Quantitative PCR……45

2.1.16. Embryo morpholino injection……45

2.1.17. ChIP-sequencing……46

2.1.18. I-SceI meganuclease trangenesis……..46

2.2. Planarians materials and methods……...47

2.2.1. Organisms maintenance……47

2.2.2. Double stranded RNA (dsRNA) synthesis……47

2.2.3. Injections……47

2.2.4. ATAC-sequencing……48

2.2.5. ChIP-sequencing……...48

2.3. *In silico* methods……50

2.3.1. RNA-sequencing data analysis……50

2.3.2. ATAC-sequencing data analysis……..…..50

Chapter III: Epigenomic regulation of *Xenopus* neural crest cells

3.1. Rationale……..54

3.2. Epigenetic modifiers involved in neural crest formation……55

3.2.1.Chemical screen using a library of epigenetic inhibitors in *X. tropicalis*…….55

3.2.2. Analysis of compounds associated with pigment loss……57

3.2.3. Discussion……..59
3.3. Transcriptome analysis of *X. laevis* neural crest cells……62
3.3.1. NC-induced animal cap samples……62
3.3.2. Comparison between stage 9 and stage 13 neural crest samples……...66
3.3.3. Comparison between stage 13 and stage 18 neural crest samples……72
3.3.4. Time of expression of neural crest genes……..78
3.3.5. Discussion……...82
3.4. ATAC-sequencing to identify cis-regulatory elements……..84
3.4.1. ATAC-sequencing protocol optimization……84
3.4.2. ATAC-sequencing samples……85
3.4.3. Comparison between ATAC-sequencing samples……..94
3.4.4. Analysis of the ectodermal cis-regulatory elements……95
3.4.5. Analysis of the neuroectoderm cis-regulatory elements……98
3.4.6. Analysis of the neural crest cis-regulatory elements……100
3.4.7. Neural crest enhancer dynamic……100
3.4.8. Analysis of neural crest transcription factor motifs present in intergenic neural crest open regions……103
3.4.9. Analysis of pluripotency transcription factor motifs present in intergenic neural crest open regions……105
3.4.10. Analysis of the cmyc as a key neural crest transcription factor in cis-regulatory elements……106
3.4.11. Selection of neural crest enhancers……...111
3.4.12. Discussion……..116

Chapter IV: Epigenomic regulation of the anteroposterior regeneration in planarians
4.1. Rationale……123
4.2. Analysis of the cis-regulatory elements at 0hr and 48hR……124
4.3. Analysis of the cis-regulatory elements at 12hR……128
4.4. Discussion……131

Chapter V: General discussion
5.1. About ATAC-sequencing……..134
5.2. About *Xenopus* neural crest formation……135
5.3. About planarian anteroposterior polarity……138
5.4. Future work……...139
5.5. Where this research sits in a wider scientific context……...140

References..……142
Appendix 1: Tables of top motifs presents in the potential enhancers of the analyzed samples……..……161
Appendix 2: The positive transcriptional elongation factor (P-TEFb) is required for neural crest specification……...……...178
Appendix 3: Unraveling the mechanisms that determine the uptake and metabolism of magnetic single and multicore nanoparticles in a *Xenopus laevis* model ……191

List of figures

Figure 1: Genome architecture……2
Figure 2: Scheme of the main histone modifications involved in gene regulation……4
Figure 3: Epigenetic features of primed, poised and active enhancers……7
Figure 4: Key chromatin marks involved in active, poised and repressed promoters. ……8
Figure 5: Representation of the ATAC-sequencing protocol……12
Figure 6: *Xenopus* life cycle……14
Figure 7: *Xenopus* Neural Crest formation processes……15
Figure 8: Gene regulatory network controlling the formation of the neural plate border……18
Figure 9: Gene regulatory network controlling the specification of NC cells……19
Figure 10: Gene regulatory network controlling the EMT of NC cells……20
Figure 11: Planarian anatomy……26
Figure 12: Current model of neoblast differentiation……28
Figure 13: Proposed model for establishment of anteroposterior identities after amputation……33
Figure 14: Animal cap assay……43
Figure 15: Schematic representation of the dsRNA injection methodology……48
Figure 16: Results of the screening using epigenetic inhibitors……56
Figure 17: Phenotypes observed after HDAC and HAT inhibitors screen……58
Figure 18: Screening of a selection of HAT and HDAC inhibitors followed by WISH……59
Figure 19: Animal cap samples used for RNA sequencing experiment……64
Figure 20: Analysis of RNA sequencing samples……65
Figure 21: Gene comparison between NC stage 9 and stage 13 samples ……66
Figure 22: Genes differentially expressed between stage 9 and stage 13 NC caps……67
Figure 23: Gene comparison between NC stage 13 and stage 18 samples……72
Figure 24: Genes differentially expressed between stage 13 and stage 18 NC caps……73
Figure 25: Expression profile of neural plate border genes……79
Figure 26: Expression profile of neural crest specifiers genes……80
Figure 27: Heatmap showing the expression of pluripotency and NC factors……81
Figure 28: Examples of tagmentation prolifes obtained with bioanalyser……85
Figure 29: Expression prolife of samples used in ATAC-sequencing……87
Figure 30: Bioanalyser profile of the stage 13 ATAC-sequencing samples……88

Figure 31: Bioanalyser profile of the stage 18 ATAC-sequencing samples……..89
Figure 32: Metrics of neural crest samples……91
Figure 33: Metrics of neuroectoderm (NE) samples……..92
Figure 34: Metrics of ectodermal samples……..93
Figure 35: Venn diagrams comparing peaks of open regions present in the different samples ……..95
Figure 36: Analysis of ectoderm cis-regulatory elements……97
Figure 37: Analysis of neuroectoderm cis-regulatory elements……...99
Figure 38: Analysis of Neural crest cis-regulatory elements……..101
Figure 39: Enhancer dynamics……102
Figure 40: Analysis of neural crest motifs present in NC enhancers……104
Figure 41: Analysis of pluripotency motifs present in NC enhancers……...106
Figure 42: Model displaying the mechanism by which *cmyc* is paused in neural crest cells……..……107
Figure 43: *cmyc* is a paused gene and CDK9 morpholino causes *cmyc* to become more paused during development……..……..109
Figure 44: ATAC-sequencing data provide a link between transcriptional elongation of *cmyc* and expression of NC derivatives……...……111
Figure 45: Potential *cmyc* enhancer……..……..112
Figure 46: Potential *snai2* enhancer……..……...……113
Figure 47: Potential *foxd3* enhancer……..……...……114
Figure 48: Potential *id3* enhancer……..……115
Figure 49: Potential *sox10* enhancer……..……...……116
Figure 50: ATAC-sequencing samples at 0hR and 48hR……..……124
Figure 51: Analysis of the 0hR and 48hR ATAC-sequencing samples……127
Figure 52: Schematic representation of the ATAC-sequencing samples at 12h regeneration…..……..129
Figure 53: Bioanalyser profile of the ATAC-sequencing 12hR samples……...130
Figure 54: ChIP-qPCR validation……131
Figure 55: Impact of ATAC-sequencing in the literature……135

List of tables

Table 1: Bibliographic summary of the chromatin modifiers involved in NC formation ..... ……..22

Table 2: Epigenetic inhibitors tested in the chemical screen……38

Table 3: *Xenopus* qPCR primers……..……..45

Table 4: Planarian qPCR primers……..……49

Table 5: Top genes differentially expressed in stage 9 caps when compared with stage 13 NC-induced caps……69

Table 6: Top genes differentially expressed in stage 13 neural crest caps when compared with stage 9 caps……70

Table 7: Top genes differentially expressed at stage 13 neural crest animal caps when compared with stage 18……..……75

Table 8: Top genes differentially expressed at stage 18 neural crest animal caps when compared with stage 13 caps ……..……76

Table i: Top motifs found in the specific NC intergenic open regions stage 13 using Homer……..……..161

Table ii: Top motifs found in the specific NC intergenic open regions stage 18 using Homer……..……..163

Table iii: Top motifs found in the specific ectoderm intergenic open regions stage 18 using Homer……..……...165

Table iv: Top motifs found in the specific neuroectoderm intergenic open regions stage 18 using Homer……..167

Table v: Top motifs found in the specific 0hR intergenic open regions using Homer…...……..170

Table vi: Top motifs found in the specific anterior 48hR intergenic open regions using Homer……..……...173

Table vii: Top motifs found in the specific posterior 48hR intergenic open regions using Homer……..……...176

Abbreviations

0hR	0 hours regeneration
12hR	12 hours regeneration
48hR	48 hours regeneration
5hmC	5- hydroximethylcitosine
5mC	5-methylcitosine
ATAC-sequencing	Assay for transposase-accessible chromatin using sequencing
BMPs	Bone morphogenetic proteins
BSA	Bovine serum albumin
ChIP	Chromatin-immunoprecipitation
cNeoblasts	Clonogenic neoblasts
CNS	Central nervous system
CREs	Cis-regulatory elements
DNMT	DNA methyltransferase
dsRNA	Double strand RNA
EMT	Epithelial to mesenchymal transition
FAIRE-sequencing	Formaldehyde-assisted isolation of regulatory elements sequencing
FBS	Fetal bovine serum
FGFs	Fibroblast growth factors
FISH	Fluorescence in situ hybridization
HAT	Histone acetyltransferase
HCNRs	Highly conserved non-coding regions
HDAC	Histone deacetylases
hNCCs	Human neural crest cells
KDM	Lysine demethylases
KMT	Lysine methyltransferases
MBT	Mid-blastula transition
MMPs	Matrix metalloproteinases
MMR	Marc's Modified Ringer's
MN	Mononucleosome
MO	Morpholino
NC	Neural crest
NF	Nucleosome free
NPB	Neural plate border
PCA	Principle component analysis
PCCs	Position control cells
PMSG	Pregnant mare serum gonadotropin
PNS	Peripheral nervous system
PRC2	Polycomb repressive complex 2
qPCR	quantitative PCR
RNAi	RNA interference
RNAP II	RNA polymerase II
RPKM	Reads per kilobase of transcript per million of mapped reads

TADs	Topological associated domains
TFs	Transcription factors
TPM	Transcripts per million
TSS	Transcription start site
WISH	Whole mount *in situ* hybridization

Acknowledgments

Thanks to everyone who has been involved in this piece of work. Firstly, to my supervisor, Grant Wheeler, for giving me the opportunity of doing this phD and for his guidance through these four years. To my secondary supervisors, Andrea Münsterberg and A. Ganesan, for their helpful inputs on my project. To Ruth Williams and Andrea Wills, for helping me with the ATAC-sequencing and to Takuya Nakayama, for helping me with the transgenesis protocol. To the Wheeler-Münsterberg lab, especially to Chris, Vicky, Kim, Angels, Nicole and Katy for teaching me all the frog stuff and making less hard the time in the lab. Particularly, to Inès, for all the experiences we have had and hotel rooms shared during these four years. Also, to Estefanía, Bea, Laura and Amalia, thanks for this last year, for making me see the bright side of the thesis, and having lots of fun (even in the gym).

To the PIs involved in the DevCom network, especially to Gert Veenstra (RU) and Jose Luís Gómez-Skarmeta (CABD), for the interest they have shown in this project and for welcoming me in their labs. To Saatrje Hontelez (RU), Marta Magrí and Jose María Santos (CABD) for teaching me all I know about ChIP-sequencing. Also to the DevCom people, especially to Rita, Mateo and Matthew – it has been nice to meet you.

To Emili Saló and Teresa Adell (UB), for allowing me to apply the techniques I have learnt in planarian, which was not the plan. Above all, thanks for listening to my ideas and trusting me. Very especially to Eudald, for being willing to help me with the planarians (and cutting thousands of blastemas). Also to Miquel, Jose, Elena for making enjoyable the work in the lab. To Gus, for having taught me how to work in the lab and for listening to my stories (I know, I speak too fast). To Alex, for having been my lab-buddy and still keeping in touch.

To Elsa, Carla, Marcos, Lluís, Anna, Marta F., Miriam, Neus, Marta Ll., Berta, Eli, Irene, Ari and Clara to making me feel as I have never left.

To my parents, because I know that my being here has not been easy for them and they have still encouraged me not to give up. *Gracias por darme ánimos y llenarme la maleta de jamón y queso.*

To my best sister, for always being there, visiting me a lot and helping me with this thesis (trying to understand what an “enancer” is).

To Andrés, for being the best things that has happened to me during these years and giving me chocolate when nothing was working.

CHAPTER I:
INTRODUCTION

1.1. Genome architecture

With the origin of multicellular organisms, the control of gene expression gets more sophisticated as such organisms have to produce and maintain different cell types with the same genome information. Moreover, the eukaryotic genome has to be tightly compact to fit inside the nuclei and, at the same time, to allow the accessibility of the DNA at regulatory regions to ensure proper gene expression (Perino & Veenstra 2016). The 3D organization of the genome acquires, then, more importance as it has an impact on transcription, DNA replication, chromatin stability and splicing. The different hierarchies of organization involve: (1) chromosome territories, (2) chromatin compartments, (3) topological associated domains (TADs), (4) chromatin/histone modifications and (5) enhancer-promoter loops (Figure 1).

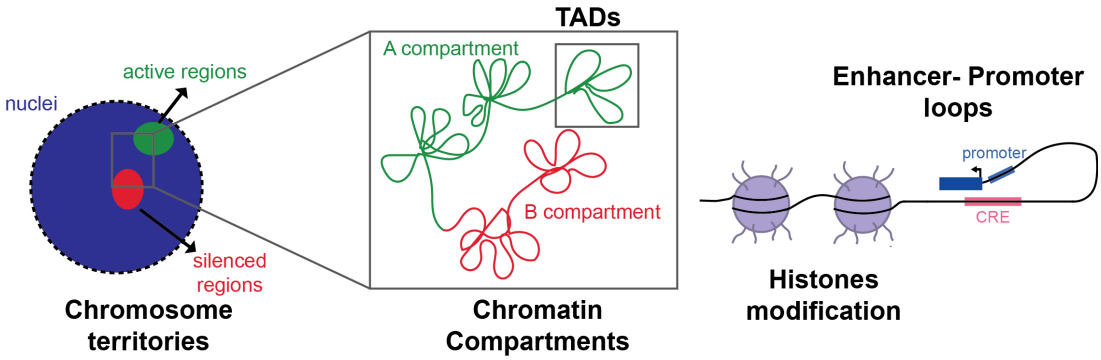


Figure 1: Genome architecture. The chromosome territories, chromatin compartments, TADs, histone modifications and enhancer-promoter loops are shown.

The chromosomes have a preferred location in the nuclei: active regions tend to be in the periphery physically separated from permanently silence regions (Figure 1). However, the location of the genome in the nuclei seems to be cell-specific rather than having a role in gene expression (Cremer & Cremer 2001). Genomic regions of approximately 5Mb determine the chromatin compartments: the active A compartment and the inactive B compartment. Within compartments, the topological associated domains are defined. Unlike the A/B compartments, TADs are very rigid and do not change between different cell types and developmental stages. TADs determine the regulatory space within which genes and regulatory elements such as enhancer, promoters and insulators interact forming loops. The loops that define TADs are highly enriched in CTCF and cohesin proteins (Perino & Veenstra 2016; Sewitz et al. 2017). At a lower scale, chromatin is a highly dynamic and regulated macromolecular complex comprised of DNA and nucleosomes, which are octamers containing two each of the histone proteins H2A, H2B, H3

and H4. 146 base pairs of genomic DNA are wrapped around each nucleosome with a 20-50bp link between them. The repetition of nucleosomes builds up the chromatin, which can be subdivided into two main domains regarding its activity: the heterochromatin, tightly packed and primarily containing inactive genes, and the euchromatin, which has a relaxed conformation and is where active transcription takes place. The N-terminal tail of the histone proteins can be post-translationally modified. These modifications include among others: acetylation, methylation, phosphorylation and ubiquitination (Sawan et al. 2008; Dawson & Kouzarides 2012). The last level of gene regulation is composed of cis-regulatory elements (CREs). CREs are sequences present in the DNA molecule that regulate gene expression. The most common elements are the promoters, at the transcription start site (TSS), and the enhancers. The regulatory landscapes of a gene is composed of all the distal cis-regulatory elements that regulate a specific promoter (Maeso et al. 2017).

As this study focuses on identifying cis-regulatory elements in two different scenarios, here, histone modifications and regulatory elements such as enhancers and promoters are going to be further explained.

1.1.1. Histone modifications

Histone proteins can be post-transcriptional modified by several chemical groups. Histone modifications are dynamically added and removed by chromatin-modifying enzymes in a highly regulated manner (Figure 2). These can (1) enhance or weaken non-covalent interactions between DNA and histones, due to changes in the net charge of nucleosomes, and (2) act as a platform for recruitment of other proteins with domains that specifically recognize these modifications. For these reasons, histone modifications modulate the accessibility to certain loci and therefore play a critical role in all DNA-based regulation processes such as transcription, DNA repair and replication (Lawrence et al. 2016).

The chromatin-modifying enzymes can be divided into three classes regarding their role: (1) writers, capable of adding chemical modifications to DNA or histones; (2) erasers, able to remove these chemical modifications and finally (3) readers, which bind to specific modifications. The specificity of the readers is due to critical residues in their binding pocket, which confer

selectivity for particular modifications, as well as residues outside the binding pocket that confer histone-sequence preferences (Berger et al. 2009).

The best-characterized epigenetic modifications are DNA methylation, histone methylation and histone acetylation (Figure 2). First, the methylation of the fifth position of cytosine (5-methylcitosine, 5mC) at CpG islands on the promoters is associated with silenced coding and non-coding genes during cell linage. DNA methyltransferases (DNMT) are the enzymes responsible of adding the methyl group whereas a complex cascade of TET enzymes is required for removing the methyl group. Secondly, lysine methylation of histone residues can be associated with activation or repression of genes depending on the position: methylation of H3K4, H3K36 and H3K79 are modifications associated with activation whereas methylation of H3K9, H3K27 and H3K20 mark repression. The enzyme families lysine methyltransferase (KMT) and lysine demethylase (KDM) are responsible of adding and removing the methyl group, respectively. On the other side, proteins with chromo and PHD domains bind specifically to lysine methylated positions, such as polycomb proteins, which binds to H3K27me to repress gene expression. Finally, acetylation of histone residues is associated with activation of transcription since this modification affects the electrical charge of histones. The histone proteins are negatively charged, and therefore this modification weakens the interaction between histone and DNA allowing the accessibility of transcription factors. Histone acetyltransferases (HAT) and histone deacetylases (HDAC) are responsible of this modification and proteins with bromo domains can bind specifically to acetylated histones, such as the transcription initiation factor TAF1 which binds to H4K5ac to stabilize TBP-TATA box complex (Kouzarides 2007).

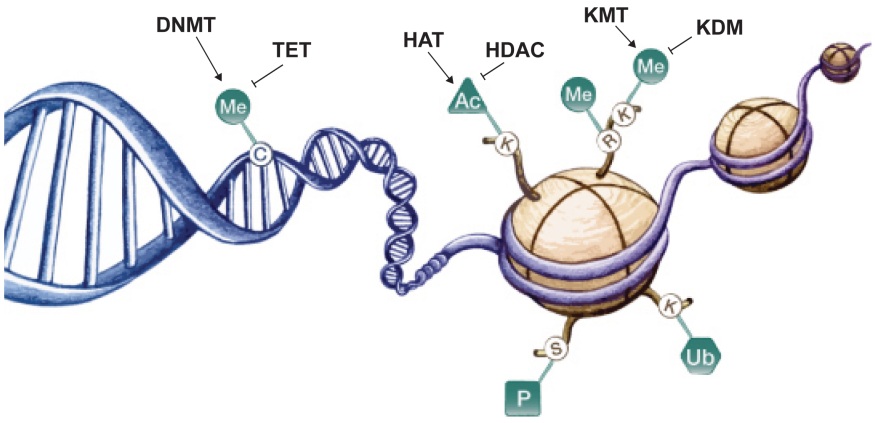


Figure 2: Scheme of the main histone modifications involved in gene regulation. Me=methylation, Ac=acetylation, P=phosphorylation and Ub=ubiquitination. Modified from Promega website.

1.1.2. Genomic regulatory elements

1.1.2.1. Enhancers

Enhancers are defined as small sequences of DNA, about a few hundred base pairs that act as a platform for transcription factors, which in turn can recruit other cofactors and nucleosome-remodeling complexes, thus regulating gene expression. Enhancers can be located 3', 5', intronically or even exonically relative to their target genes (Suryamohan & Halfon 2015). Different enhancers can exhibit different regulatory behaviors. Some genes have multiple enhancers each controlling specific spatiotemporal aspects of their target gene expression; some other genes have secondary enhancers, which drive similar expression conferring robustness; some other genes are regulated by so-called superenhancers, which contain several motifs for important transcription factors such as Oct4, Sox2, and Nanog as well as the coactivator Mediator. Superenhancers are composed by multiple concatenated enhancers within 12.5kb and have the ability to activate high levels of transcription of the target genes, usually key cell-type specific genes (Whyte et al. 2013; Spitz & Furlong 2012; Maeso et al. 2017).

Transcription factors (TFs) are an important player in cis-regulation. TFs recognize a 6-12 degenerate DNA sequence, called a motif. Enhancers often contain several TFs motifs and the cooperation between TFs is necessary to activate gene expression. There are several mechanisms whereby TFs cooperate and increase the retention time of the TFs at the enhancer: (1) direct protein-protein binding of TFs that binds to adjacent DNA sequences; (2) several TFs co-bind to an enhancer regulator (i.e. p300); (3) the binding of the first TF produces a conformational change which expose the binding site for a second TF, the first TF acts therefore as a pioneer factor; (4) the binding of the first TF prevents the remodeling of the nucleosome allowing the second TF to bind to the enhancer; and finally (5) the binding of the first TF bends the DNA in a way that the second TF is then able to bind to its motif. Despite the importance of TFs in activating enhancers, the majority of TF binding events are not directly affecting gene activation but rather with nucleosome positioning (Spitz & Furlong 2012).

Developmental enhancers, in order to ensure accuracy on gene expression, undergo progressive changes to become active. Enhancer activation involves transition from inactive to primed and then from primed to active or poised stage (Figure 3). Before enhancer activation, these are in a compact state and, therefore, inaccessible. Enhancer priming facilitate activation as decreases the remodeling

time upon activation signals (Figure 3A). The key element is the pioneer factor which binds directly to DNA and triggers nucleosome repositioning, chromatin remodeling and prevent DNA methylation, which is related to gene silencing. Examples of pioneer factors are FoxA1, involved in liver specification, and Zelda, only present in *Drosophila* and required for maternal to zygotic transition. (Caravaca et al. 2013; Liang et al. 2008). Primed enhancers are enriched in H3K4me1, which serves as a platform for Tip60/400 that catalyzes replacement of H2A with H2A.Z. Histone variants H2A.Z and H3.3, opposite to canonical histones are hyper dynamic and the DNA-histone binding of such histone variants is less tight. Both primed and active enhancers contain 5- hydroximethylcytosine (5hmC). Silenced regions contain 5-methylcytosine; to transform them into active regions the methyl group needs to be removed through TET enzymes generating 5hmC (Calo & Wysocka 2013; Spitz & Furlong 2012)

Active enhancers (Figure 3C) are characterized by several histone modifications such as H3K4me1, H3K79me3 and H3K27ac. The histone acetyltransferase p300/CBP, often used as a marker for active enhancers, acetylates H3K27, which recruit proteins with a bromodomain such as chromatin remodelers and transcription regulators. H3K27ac not only acts as a platform for proteins with bromodomain but also decreases nucleosome stability and chromatin decompactation due to changes in net charge of histones. Multiple cell-type specific and generic transcription factors bind to the enhancers together with cofactors, such as histone modifiers, chromatin remodeling complexes and mediators of long-range crosstalk with basal transcription machinery in the promoters. By looping, enhancer associated factors are directly delivered to the promoter, helped by CTCF and cohesion, in order to potentiate either the formation of the preinitiation complex or the transition from initiation to elongation (Plank & Dean 2014; Calo & Wysocka 2013).

During differentiation, the enhancers are poised to ensure rapid activation upon TFs binding or cellular signals. Poised enhancers (Figure 3B) are characterized by the presence of p300, H3K4me1 and H3K27me3, preventing acetylation of H3K27ac related with activation. Polycomb repressive complex 2 (PRC2), also present in poised enhancers, triggers H3K27 methylation (Calo & Wysocka 2013; Spitz & Furlong 2012).

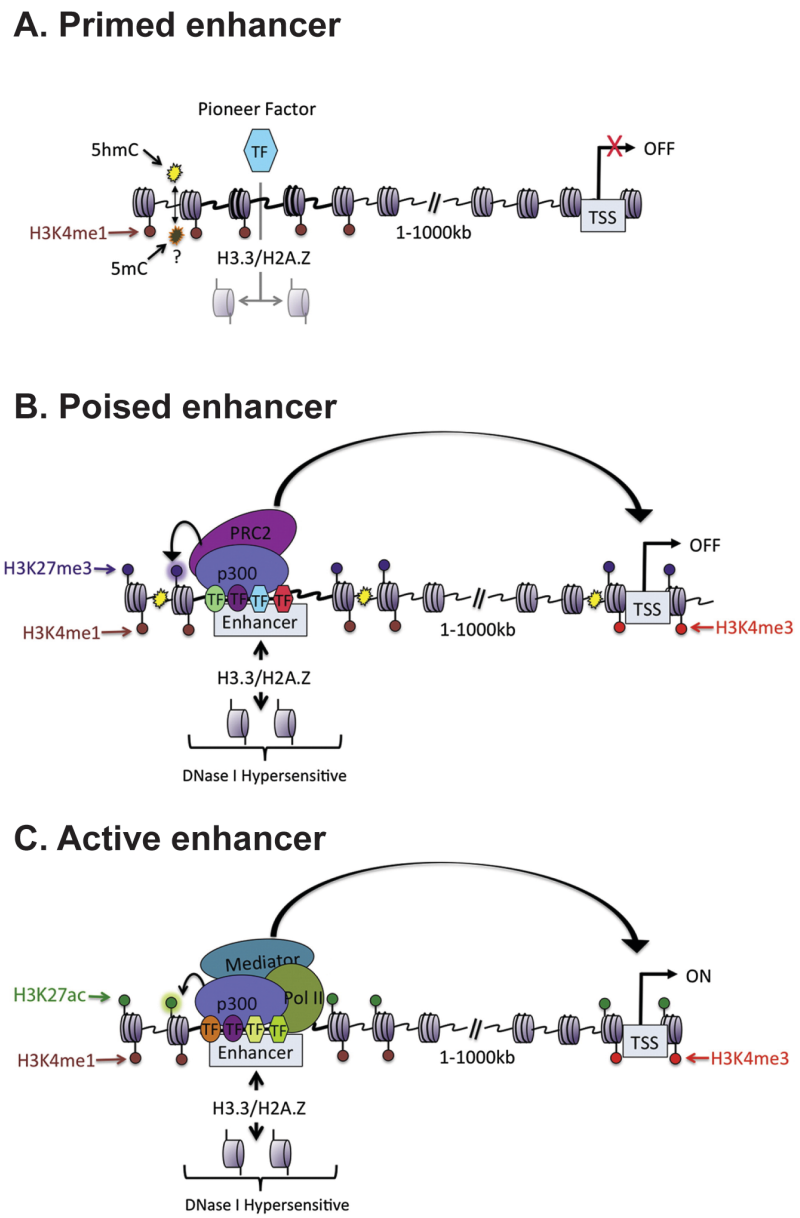


Figure 3: Epigenetic features of primed, poised and active enhancers. Modified from Calo & Wysocka (2013).

1.1.2.2. Promoters

Promoters are defined as DNA sequences proximal to the TSS (transcription start site) of a gene where the RNA polymerase binds to the DNA to initiate the transcription. About 70% of the annotated promoters in vertebrates contain CpG islands, an enrichment of C and G bases in the DNA sequence which are predominantly non-methylated (Deaton & Bird 2011).

Active promoters are characterized by H3K4me3 and H3K4me2, RNA polymerase II (RNAP II) deposition, histone acetylation such as H3K9ac or

H3K14ac, hypersensivity to DNAse I digestion and hypomethylation at the DNA level (Figure 4A). H3K4me3 is a key mark for identifying promoters, acting as a docking site for (1) transcription machinery; (2) histone acetyltransferases, needed for making the chromatin accessible; (3) histone demethylases that remove repressive marks such as H3K36me3 or H3K9me3 typical from heterochromatin; and (4) metyltransferases that add new H3K4me3 to the next nucleosomes (Li et al. 2007; Wozniak & Strahl 2014). The histone modification H3K27me3, triggered by the PRC2, defines repressed promoters together with H3K9me3 (Figure 4C). Inactive promoters also contain DNA methylated regions (Zhou et al. 2011). Some developmental genes show both active (H3K4me3) and repressive (H3K27me3) marks so they can be rapidly induced by extracellular signals; these are so-called poised promoters. RNAP II can be present but they do not produce mature mRNAs (Figure 4B)(Zhou et al. 2011).

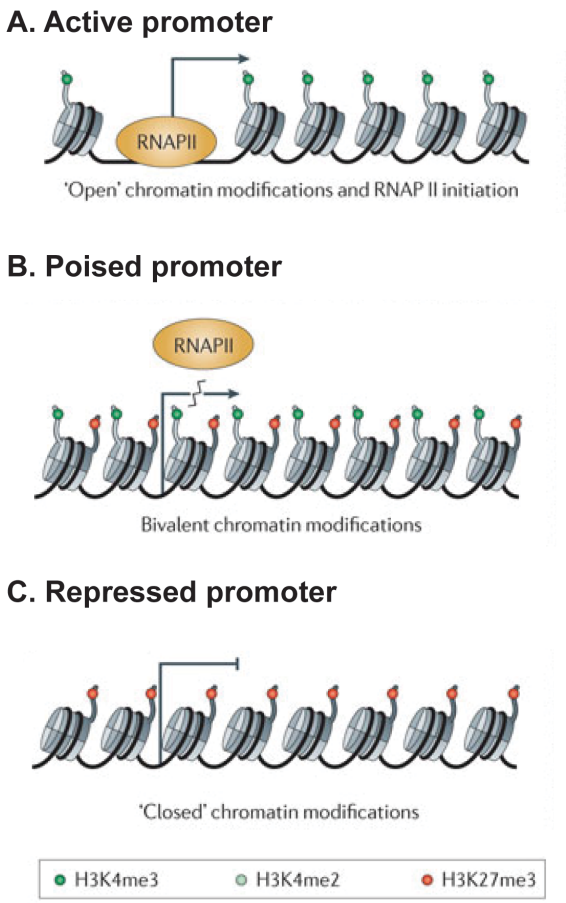


Figure 4: Key chromatin marks involved in active, poised and repressed promoters. Modified from Zhou et al. (2011).

1.1.3. Evolution and conservation of cis-regulatory elements

Gene regulatory networks involved in Hox genes patterning, establishment of body axes or nervous system regionalization are well conserved during evolution. Unlike the high degree of coding sequence conservation, cis-regulatory elements do not display a great conservation among phyla. Nevertheless, there are a few highly conserved non-coding regions (HCNRs) (Maeso et al. 2013).

HCNRs are enhancers present in vertebrate genomes involved in regulating key developmental genes. They are characterized by being short (75-100bp) and with low similarity among species (60-70%), making them technically difficult to identify. In the chordate phylum, there have only been identified 50 HCNRs common between amphioxus, the ancestor of the vertebrates, and human genomes, despite the similarities in body plan organization (Putnam et al. 2008; Woolfe et al. 2005).

Studies performed in several organisms, from flies to mammalian, reveal that regardless of the low sequence conservation of cis-regulatory elements, the signatures of enhancers are evolutionary conserved. Active enhancers correlate with H3K27ac, H3K4me1, high DNA hypersensitive sites and p300/CBP. Techniques to identify such regions include: chromatin immunoprecipitation using H2K27ac or p300 antibodies or DNAse I or ATAC-sequencing that identify hypersensitive regions. Moreover, the transcription binding sites present in the enhancers are more conserved than the broader sequence that contains them (Heintzman et al. 2007; Bogdanović et al. 2012; Neph et al. 2012).

Also relevant are enhancers turned on during late gastrula that drive expression of genes during the phylotypic stages, where the maximum morphological similarities between vertebrates are found (Irie & Kuratani 2011). It has been demonstrated that these putative enhancers display the highest sequence conservation compared with enhancers active at other stages (Maeso et al. 2013).

When talking about evolution of cis-regulatory elements, mention has to be made on the conservation of TADs. TADs are present in bilaterian animals; plants or yeast either do not contain this 3D genome organization or it is not regulated by the same molecular mechanism (Dekker & Heard 2015). The appearance of TADs correlates with the presence of CTCFs facilitating enhancer-promoter interactions (Heger et al. 2012). Alteration of TAD boundaries, either by mutation, insertion or

deletion, can change the integrity of the regulatory landscape creating or eliminating enhancers-promoter contacts thus activating or repressing genes. In fact, whole sequence analysis of cancer cells reveals high degree of mutations in CTCF sites (Hnisz et al. 2016). Studies of microsynteny (gene order) conservation show also a high degree of conservation in bilaterians. This is due to the presence of CREs in the introns of neighboring genes forcing the conservation of the regulatory landscape (Irimia et al. 2012).

The high frequency of interactions inside a TAD promotes the formation of new CREs through (1) *de novo* formation, (2) modification of preexisting CREs or (3) insertion and exaptation of transposable elements. Nevertheless, not all the new-formed potential CREs can act as functional CREs: the new enhancers need to connect to a promoter within the same TAD that would respond to new enhancers. Those enhancers driving the expression in similar pattern as the preexisting enhancers, have more chances of affecting gene expression thus (1) quantitatively affecting gene expression or (2) refining a preexisting expression pattern (Maeso et al. 2017).

An example of how changes in TADs can influence evolution is the role of the HoxD in limb formation. In vertebrates there are two hox clusters flanked by several CREs organized in two TADs: one at 3' of the gene cluster and another at 5'. The 3' CREs become active first and are involved in proximal structures whereas the 5' CREs are related with distal structures such as the hands and the digits. The central part of the cluster (*hoxd8-hoxd11*) switches from being influenced for one TAD to the other, and thus allowing the formation of the wrist and ankle. Interestingly, in amphioxus there is just one hox cluster regulated by a single TAD located at 3' CREs. This suggests that the incorporation of the inputs coming from the 5' regulatory region has allowed the formation of new structures in higher vertebrates such as the digits (Lonfat & Duboule 2015; Acemel et al. 2016).

1.1.4. Empirical approaches to study cis-regulatory elements

The main experimental methods for whole genome CREs discovery are based on chromatin-immunoprecipitation (ChIP) or chromatin accessibility techniques (DNase- MNase- FAIRE- and ATAC-sequencing). Subsequently, the potential CREs identified by next-generation sequencing can be validated by low throughput (reporter gene assay) or high throughput (enhancer-FACS-sequencing and CRE-sequencing) reporter assays. Also, to demonstrate the physical interaction

of enhancers and promoters image based (FISH) and chromatin conformation (3C, 4C, 5C and HiC) techniques can be used (Suryamohan & Halfon 2015).

1.1.4.1. Identification of cis-regulatory elements

The most frequently used technique to identify enhancers is ChIP-sequencing. This technique is based in the *in vivo* crosslinking of DNA-protein complexes followed by pull-down using a specific antibody and high throughput DNA sequencing. This method allows the detection of both direct and indirect whole genome binding sites events. The key point of this technique lies on the antibody selection: the use of an antibody recognizing a specific TF of interest would reveal this TFs binding sites on the genome, even though the binding might be not functional, whereas the use of an antibody against the co-factor p300 or the histone modifications H3K4me1 and H3K27ac would unravel the active (and poised) enhancers on the genome without any prior knowledge of the TFs involved (Visel et al. 2009; Rada-Iglesias et al. 2012).

One of the characteristics of active enhancers is that they are depleted of nucleosomes and therefore the DNA is accessible. This is the rationale behind several techniques based on open chromatin. The DNAase-sequencing technique is based on the enzymatic activity of DNase I to cleave accessible DNA (Boyle et al. 2008). MNase-sequencing relies on micrococcal nuclease activity that reveals nucleosome position therefore indirectly the accessible regions (Barski et al. 2007). Alternatively, formaldehyde-assisted isolation of regulatory elements (FAIRE-seq) isolates nucleosome-free DNA by a phenol extraction step (Giresi et al. 2007). Each of these techniques followed by deep sequencing allow the identification of open chromatin regions and the transcription factors footprint (Neph et al. 2012). However, not all open regions are active enhancers; accessible regions can also be found on primed enhancers, promoter regions, CTCF binding sites and in regions bound by repressive transcription factors. Because of that, it is advisable to combine the two different types of CRE identification for a more conclusive result (Shlyueva et al. 2014).

Also based on open chromatin, the assay for transposase-accessible chromatin using sequencing (ATAC-sequencing) stands out compared to the above mentioned methods because of the simplicity of its protocol and the small amount of material that it requires (Buenrostro et al. 2015; Buenrostro et al. 2013). ATAC-sequencing is based on a modified Tn5 transposase enzyme that *in vivo* fragments open regions of the DNA and at the same time tags these fragments for sequencing

(Figure 5). This technique can be performed in 500 to 50,000 nuclei, which makes it very appealing for certain model organisms or cell subpopulation. Unlike other high throughput techniques, the library preparation is performed with a simple PCR and purification. Tn5 cuts around nucleosomes therefore the expected fragment size to sequence is about 300bp.

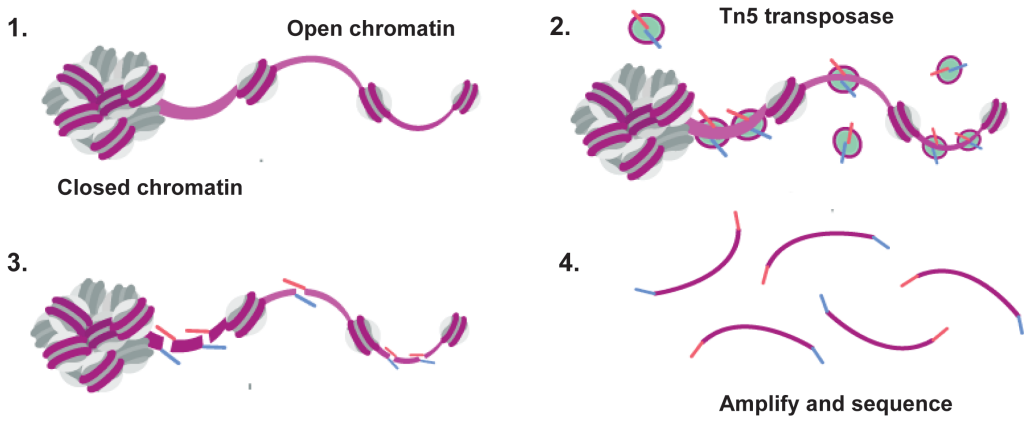


Figure 5: Representation of the ATAC-sequencing protocol. Modified from Abcam website.

1.1.4.2. Validation of cis-regulatory elements

After identification of candidate CREs, *in vivo* validation can be carried out using reporter gene assays. The idea behind this is to interrogate the ability of a certain sequence to direct transcription. The reporter gene vector contains: (1) CRE to test, (2) minimal promoter that alone drive little or no transcription and (3) reporter gene whose expression should be easy to detect such as gfp, luciferase, lacZ or β-galactosidase (Suryamohan & Halfon 2015).

Traditionally, cloned sequences were tested in model organisms such as zebrafish, flies or mice. The vector is randomly integrated in the genome and the enhancer activity is determined by the localization and the presence of the reporter gene at RNA or protein level. There are several modifications of this technique such as the enhancer-FACS-sequencing design for *Drosophila* that allows the identification of the cell type the enhancer is acting in (Gisselbrecht et al. 2013). This low throughput method is not suitable for enhancer screening and requires a small selection of CREs to test (Shlyueva et al. 2014). With the introduction of next generation sequencing strategies, a lot of effort has been made to adapt reporter gene assays. Techniques such as CRE-sequencing incorporate a barcode in the vector and the amounts of reads sequenced indicate the activity of the potential enhancer (Mogno et al. 2013).

1.1.4.3. Validation of promoter-enhancer interaction

Enhancers are known to directly contact promoters. Although it is assumed that enhancers drive the expression of neighboring genes it is necessary to experimentally address this question. The first method is the fluorescence *in situ* hybridization (FISH), which uses probes against the promoter and the enhancer to test the co-localization of both elements. This technique allows single cell resolution but is a low throughput technique limited by the resolution of the microscopy (Ronshaugen & Levine 2004).

Another set of techniques that aim to identify promoter-enhancer interactions in a high throughput manner is the so-called chromosome conformation capture-based methods (3C, 4C, 5C and HiC). The idea behind these techniques is to crosslink the chromosomal contacts generating a chimeric DNA molecule of promoter-enhancers that reflect long-range contacts. The sequencing of this molecule using specific primers is called 3C or 4C; depending if primers for promoter and enhancer or only primers for enhancers are used, respectively. These allow the validation of specific enhancer-promoter interactions. Moreover, the development of 5C and HiC reveals all the interactions in a region or in the whole genome, respectively (van Steensel & Dekker 2010). A variant of this method is the Chromatin Interaction Analysis by Paired-End Tag Sequencing (ChIA-PET), which identifies the enhancer-promoter contacts in which a specific transcription factor is involved. HiC is currently used to detect TADs within the genome (Fullwood et al. 2009).

1.2. *Xenopus* as a model organism

Xenopus, and in particular *X. laevis* and *X. tropicalis,* are African clawed frogs commonly used to study developmental biology. Since the 1950s, *X. laevis* has been widely used in research as these organisms can produce thousands of easy-to-maintain eggs that develop synchronically. The eggs are relatively large (1.4-1.5mm) and can tolerate dissections and injections, even into specific blastomeres, of large amounts of material; thus making it a good model to examine gene function (Harland & Grainger, 2011). In the mid-1990s, *X. tropicalis* became more popular for genetic studies because of its diploid genome unlike the *X. laevis* pseudotetraploid genome (Schmitt et al. 2014).

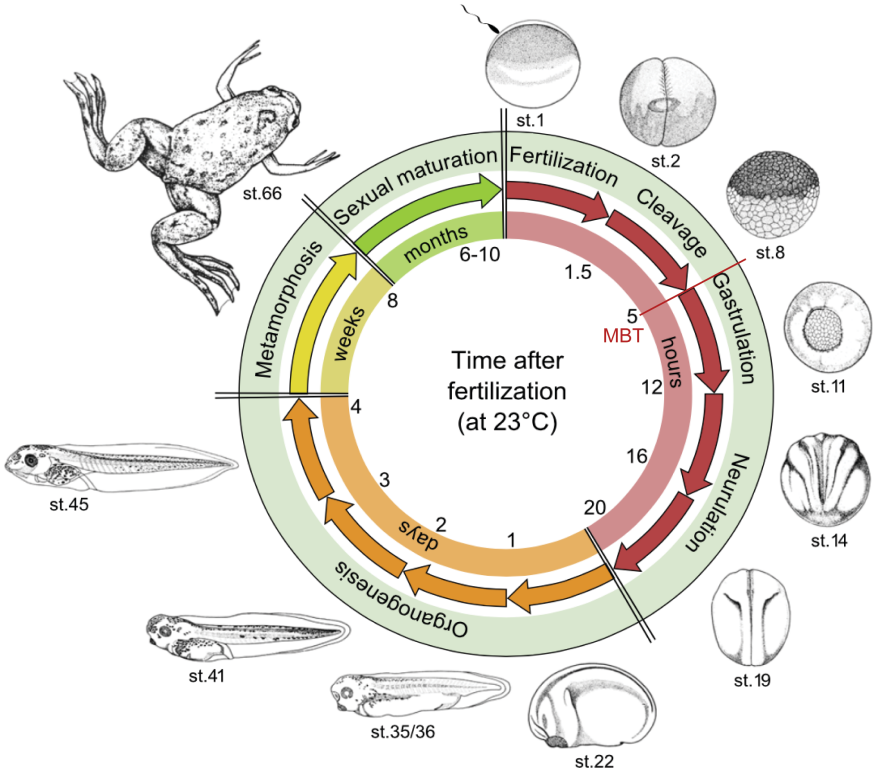


Figure 6: *Xenopus* life cycle. The embryos are stages based on the standard Niewkoop and Faber fate map. Timing of *Xenopus* development is temperature dependent. Modified from Schmitt et al. (2014).

An unfertilized egg of *Xenopus* has two easy-to-distinguish regions: (1) the animal pole, darkly pigmented, and (2) the vegetal pole, lightly pigmented and containing high amount of maternal messenger RNA (Sive et al. 2000). *Xenopus* development starts after the fertilization has occurred on the animal pole of the embryo (Figure 6). 90 minutes after fertilization, the first cleavage takes place giving rise to two identical blastomeres followed by 12 rapid synchronous cell divisions consisting in DNA replication and mitosis. Until the so-called mid-blastula transition (MBT) coinciding with stage 8-9 according to the Nieuwkoop and Faber atlas, the embryo relies on maternal RNA and proteins. Following MBT, zygotic gene transcription takes over and the cell cycle is desynchronized by adding the gap phases G1 and G2 of the cell cycle (Schmitt et al. 2014).

After fertilization, the embryonic cells divide generating a cavity inside called blastocoel. At stage 10, the blastula embryo begins to undergo gastrulation, a complex process of cell rearrangements that includes the invagination of the dorsal cells, opposite to the sperm entry point, and epiboly of the cells from the animal pole thus giving rise to the three germ layers (endoderm, mesoderm, ectoderm) that will form different tissues in the embryo (Gilbert & Barresi 2016). Shortly after

gastrulation, neurulation takes place to form the central nervous system (Figure 6). The dorsal part of the gastrula embryo, the neural plate, folds to form the neural tube while the cells on the edge of the neural plate will give rise to the preplacodal regions and the neural crest. Following neurulation, organogenesis occurs to form the different organs of the embryo such as the heart, the liver or the kidney (Sive et al. 2000).

1.2.1. Xenopus neural crest formation

The neural crest (NC) cells are a multipotent and highly migratory cell population unique to vertebrates that give rise to several derivatives such as enteric ganglia, neuroendocrine cells, melanocytes, neurons, craniofacial skeleton and connective tissue. This cell population arises from the neural plate border, which is elevated by the neural folds and finally forms the dorsal part of the neural tube. The NC then undergo epithelial to mesenchymal transition (EMT) and migrate away to their correct destination where they differentiate into the several NC-derivatives (Pegoraro & Monsoro-Burq 2013) (Figure 7). Recently, the NC has been defined as a subset of blastula cells that retain pluripotency properties whereas other cells become fate-restricted. Evidence suggests that neural crest key genes are expressed at pluripotent (blastula) stage. Also the induction of NC is sufficient to confer pluripotency capabilities (Buitrago-Delgado et al. 2015; Hoppler & Wheeler 2015).

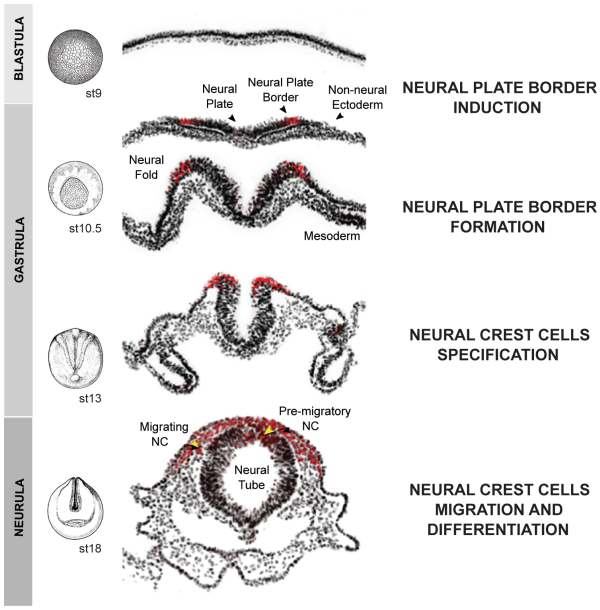


Figure 7: *Xenopus* neural crest formation processes.

Neural crest cells are induced in the neural plate border during gastrulation thanks to the interaction of several signaling pathways such as BMP, FGF and Wnt together with Notch/Delta, Retinoic acid, Hedgehog and Endothelin. These signals come from the neural plate, non-neural ectoderm and the underlying mesoderm. As a result, the neural plate border (NPB) is induced. The NPB contains a multi-progenitor cell population that is capable of giving rise to neural crest cells as well as ectodermal placodes, epidermal cells, roof plate cells and sensory neurons of the central nervous system (Simões-Costa & Bronner 2015). The ectodermal placodes generate different components of the sensory organs of the head together with the cranial NC derivatives. The NPB specifiers (*msx, pax3/7, zic, dlx5/6, hairy, tfap2a, gbx2* and *meis3*) are expressed in the prospective NC cells. Later on, during neurulation (stage 15-20), the interactions between neural plate border specifiers and the initial signaling pathways lead to the specification of the NC located in the elevating neural folds and dorsal neural tube. These cells are characterized by the expression of NC specifiers (*snai2, foxd3, sox9/10, twist, cmyc* and *tfap2a*) that drive the expression of genes involved in EMT, delamination, migration and, differentiation into the several NC derivatives (Stuhlmiller & García-Castro 2012; Pegoraro & Monsoro-Burq 2013) (Figure 7).

1.2.1.1. Neural crest induction

In order to induce the NC cells, simultaneous gradients of mainly BMP, FGF and Wnt are required in two different phases. During gastrulation, Wnt sourced from the underlying mesoderm and non-neural ectoderm and FGF also from the mesoderm must be present together with low levels of BMP, due to the presence of BMP antagonists, thus allowing NC induction (Figure 8A). Later during neurulation, BMP from the underlying mesoderm, Wnt and Notch are required to maintain the NC population (Stuhlmiller & García-Castro 2012).

Bone morphogenetic proteins (BMPs) are secreted proteins that bind to their cellular receptor activating gene expression through the canonical Smad signaling pathway. A dual function of *bmp4/7* is required: a first inhibition during gastrulation to induce the NC cells and intermediate levels during neurulation to maintain them. The BMP gradient is generated during neurulation, with high levels at the non-neural ectoderm and low levels the neural plate, promoted by BMP antagonists (Chordin, Follistatin and Noggin) secreted from the mesoderm (Steventon et al. 2009; Stuhlmiller & García-Castro 2012). Other BMP regulators in the neural-NC border have been described: *snw1* and *tumor necrosis-factor receptor associated*

factor 4 (*traf4*) activate BMP whereas *tsukushi* is a negative regulator (Kuriyama et al. 2006; Wu et al. 2011).

Wnt signaling involving both the canonical and non-canonical pathway is required for the NC induction. Canonical *wnt1, wnt3a,* expressed in the neural plate and *wnt8*, expressed in the mesoderm, can induce the neural plate border and the neural crest markers when combined with BMP antagonists (Saint-Jeannet et al. 1997; Chang & Hem 1998; LaBonne & Bronner-Fraser 1998). *frizzled7* initially and later *frizzled3* are the Wnt receptors involved in NC specification (Abu-Elmagd et al. 2006; Deardorff et al. 2001). On the other hand, the non-canonical *wnt11* acts through *microtubule-associated regulatory kinase* and it is also required for the NC formation (Ossipova & Sokol 2011).

Fibroblast growth factors (FGFs) can activate several intracellular signaling cascades. *fgf8* secreted from the mesoderm is sufficient to induce neural crest (Monsoro-Burq et al. 2003). Fgf signaling can also induce *wnt8* in the mesodermal tissues during gastrulation (Hong et al. 2008).

The interactions between these signaling pathways turn on the expression of specific NPB genes. These transcription factors positively regulate each other to define the neural plate border domain and stabilize this regulatory state to ensure their continued expression (Figure 8B). *tfap2a* activates the expression of *pax3* and *msx1* in the border with the neural plate (de Crozé et al. 2011). *pax3* and *zic1* are independently regulated but synergize in a Wnt dependent manner to induce the neural crest cells. The expression of these genes also regulates the initial signaling pathways: *hairy* and *msx1* are needed for FGF and BMP signaling and *pax3* and *tfap2a* are required for FGF and Wnt signaling allowing NC formation (Pegoraro & Monsoro-Burq 2013; Monsoro-Burq et al. 2005). All these interactions between NPB markers and signaling pathways are required for the correct NC induction.

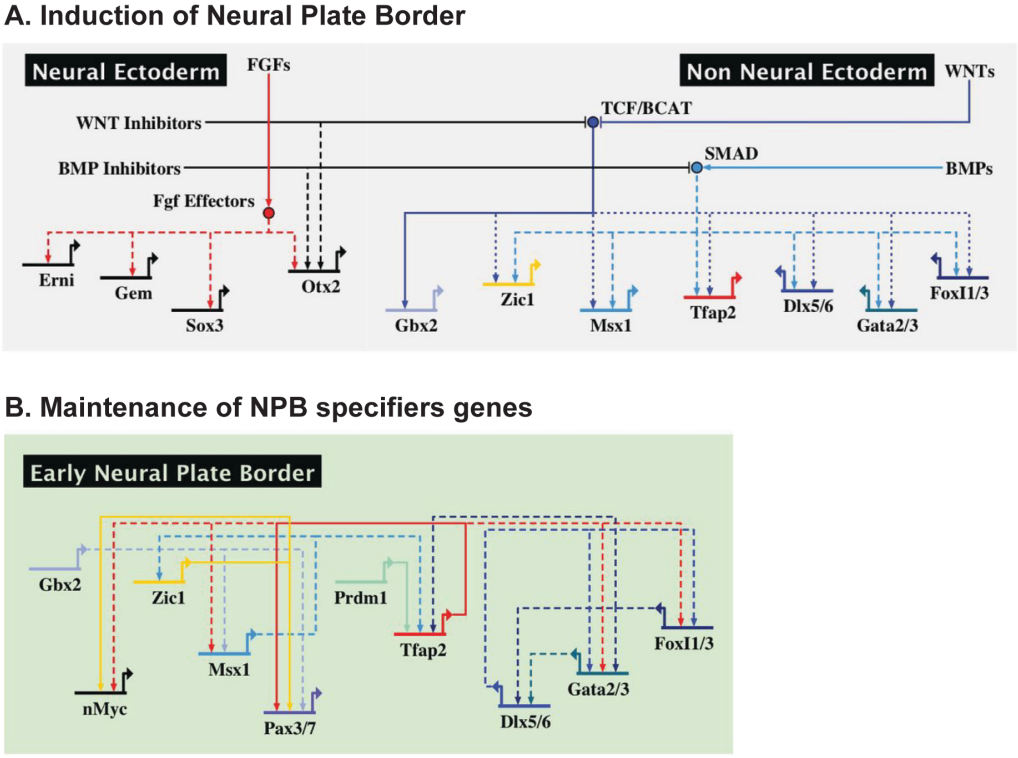


Figure 8: Gene regulatory network controlling the formation of the neural plate border. (A) Wnt, BMP and FGF signals induction of NPB. (B) Interaction between NPB required for the maintenance of NPB genes. Direct interactions are indicated with solid lines, whereas dashed lines show possible direct interactions inferred from gain- and loss-of-function studies. Modified from Simões-Costa & Bronner (2015).

1.2.1.2. Neural crest specification

The interaction between NPB markers together with Wnt and BMP signaling from surrounding tissues leads to the specification of the NC characterized by the expression of neural crest specifiers (Figure 9). These genes are involved in three main processes that characterize the NC cells. NC specifiers: (1) regulate the balance between proliferation and specification. *hairy2* and *stat3* are known to enhance proliferation at the expense of specification and subsequent differentiation the NC cells. To keep a balance, these genes also activate *id3*, which in turn promote differentiation and negatively regulates *hairy2* and *stat3* (Nichane et al. 2008; Nichane et al. 2009). *sox9* has also been involved in survival as lost of function experiments lead to the downregulation of *snail1*, an anti-apoptotic factor, causing massive apoptosis of trunk NC cells (Cheung et al. 2005).

NC specifiers also (2) maintain their expression by activating each other so that the regulatory system is stabilized (Figure 9) (Simões-Costa & Bronner 2015; Sauka-Spengler & Bronner-Fraser 2008; Pegoraro & Monsoro-Burq 2013). A subset of NC specifiers (*foxd3*, *ets1*, *snail1*/2) are first expressed to define NC identity.

These transcription factors are epigenetically activated by the neural plate border genes *pax3*, *zic*, *msx* and *tfap2a*. It has been hypothesized that *tfap2a* together with *nrf1/2* would bind NC enhancers inducing the chromatin to open allowing *pax3*, *zic* and *msx* to bind the regulatory regions of the NC genes for their proper expression (Rada-Iglesias et al. 2012; Plouhinec et al. 2014).

Due to the interaction between NPB markers, initial signaling pathways and cross-activation between NC specifiers, a full suite of genes is expressed (*foxd3*, *snail1*/2, *twist*, *pax3*/7, *sox9/10*, *ets1*, *cmyc*, *tfap2a* and *id3*) in the neural crest population, some of which are newly expressed and others whose expression is maintained form NPB cells (Simões-Costa & Bronner 2015).

Finally, NC specifiers (3) activate the expression of genes that change the adhesive properties, shape and motility of NC cells thus allowing delamination from the neuroepithelium and migration through highly regulated pathways to their final destination where these cells will differentiate (Simões-Costa & Bronner 2015; Sauka-Spengler & Bronner-Fraser 2008; Pegoraro & Monsoro-Burq 2013).

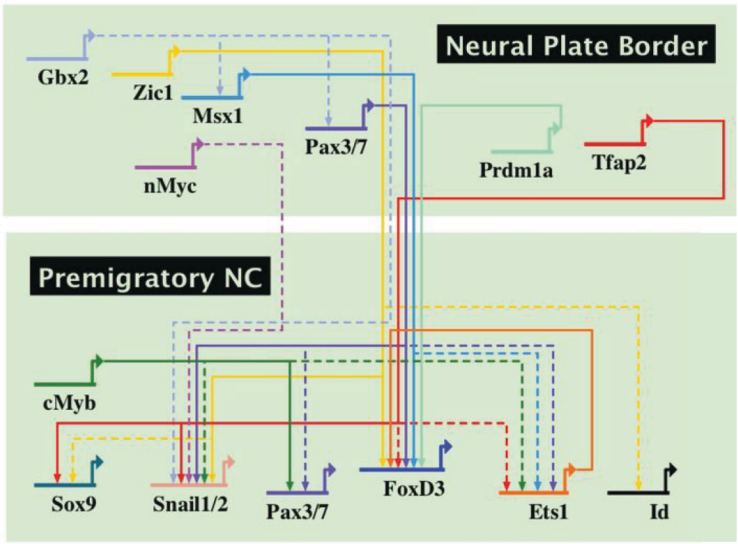


Figure 9: Gene regulatory network controlling the specification of NC cells. Direct interactions are indicated with solid lines, whereas dashed lines show possible direct interactions inferred from gain- and loss-of-function studies. Modified from Simões-Costa & Bronner (2015).

1.2.1.3. Neural crest migration

Premigratory NC cells are polarized in the dorsal neural tube. To achieve EMT, a loss of apical-basal polarity and tight junctions is required along with a shift from cadherins type I, associated with adhesion, to type II (Figure 10) (Sauka-

Spengler & Bronner-Fraser 2008). Several NC transcription factors have been associated with the switch form type I (*E-cadherin*, *N-cadherin*, *cadherin-6b*) to type II (*cadherin7/11*) cadherins. *snai2* can bind to *E-cadherin* promoter to repress its expression and it also inhibits *cadherin-6b*. Also, *foxd3* can inhibit *N-cadherin* expression (Taneyhill et al. 2007; Cano et al. 2000). Conversely, type II *cadherin 7* is upregulated by *foxd3* and *snai2* (Chu et al. 2006). The delamination process also involves downregulation of tight-junctions by *snai2* and increase of gap-junctions (Ikenouchi et al. 2003). Besides delamination from the neural tube, the NC cells also have to disaggregate into individual cells. This is regulated by *twist* and its regulator *hif1α*, which represses *E-cadherin* in the delaminating cells (Barriga et al. 2013).

To allow migration, cells must also disturb basement membrane and invade the extracellular matrix, which consist of collagens, fibronectin, laminins, vitronectin and proteoglycans. To do this, NC cells express matrix metalloproteinases (MMPs) on their surface or secrete them. In order to regulate the disruption of the cell matrix whilst keeping a balance, the NC cells express *mmps* and their inhibitors (Sauka-Spengler & Bronner-Fraser 2008). In that sense, there have been identified several *mmps* involved in the NC cells migration. For example, *mmpP17b* is involved in general NC migration; *adam13*, in migrating cranial NC cells and *mmp2* in NC-derived cardiac cells. (Leigh et al. 2013; Alfandari et al. 2001; Cai et al. 2000). While cells are migrating, they respond to their environment therefore neighboring cells are involved in directing migration and targeting the correct position (Sauka-Spengler & Bronner-Fraser 2008).

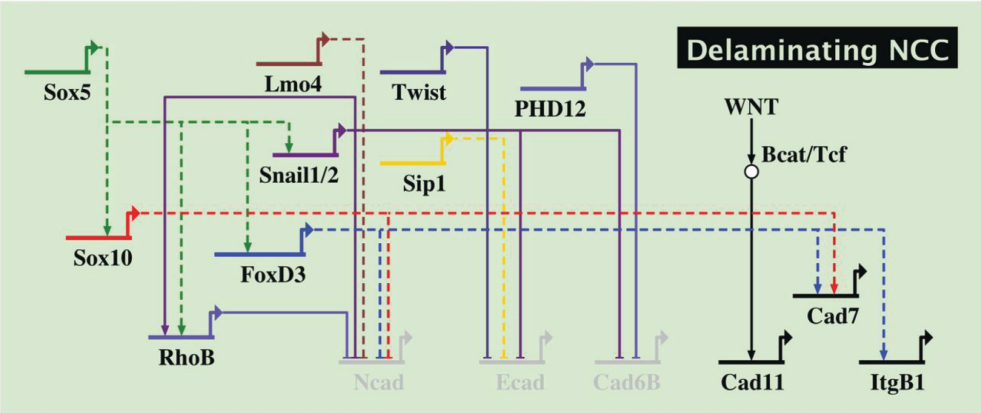


Figure 10: Gene regulatory network controlling the EMT of NC cells. Direct interactions are indicated with solid lines, whereas dashed lines show possible direct interactions inferred from gain- and loss-of-function studies. Modified from Simões-Costa & Bronner (2015).

1.2.1.4. Neural crest differentiation

Migrating neural crest cells express transcription factors involved in both migration and differentiation that will lead to the formation of several derivatives. There are four main NC-derivatives: (1) cephalic NC cells that give rise to craniofacial structures such as bones, cartilage, connective tissue, pigment cells and cephalic peripheral nervous system (PNS), (2) cardiac NC cells that contribute to the heart outflow track, (3) truck NC cells that contribute to pigment cells, PNS and adrenal glands, and (4) enteric NC cells that form the gut PNS (Mayor & Theveneau 2013).

1.2.2. Epigenetic regulation of the neural crest cells

1.2.2.1. DNA and chromatin modifications related with NC

Several modifications have been associated with NC formation in either chicken, mice, zebrafish or *Xenopus.* Table 1 summarizes the main bibliographic evidences for each of them. A couple of examples include the lysine demethylase *jmjD2a*, the adapter *phd12* that directly interact with HDAC and finally, the DNA methyltransferase *dnmt3*.

The lysine demethylase *jmjD2a* is expressed in the neural plate, neural border and the NC cells, and removes H3K9me3 and H3K36me3 *in vivo*. By inhibiting *jmjD2a*, the expression of *sox10* is down regulated suggesting that *jmjD2A* is required for H3K9me3 demethylation at *sox10* locus to activate its expression (Strobl-Mazzulla et al. 2010). Also, *snai2* represses *cadherin-6b* expression in premigratory NC cells through an epigenetic mechanism. *snai2* binds to *phd12* recruiting *hdac/sin3a* at the *cadherin-6b* promoter to repress its expression. Hence, normal levels of *cadherin-6b* are found allowing the NC cells to undergo EMT (Strobl-Mazzulla & Bronner 2012). Moreover, there is cooperation between DNA methyltransferase and Histone methyltransferase in zebrafish development. The DNMT *dnmt3* together with the HMT *g9a* bind to the *lef1* promoter allowing the correct neural and eye differentiation (Rai et al. 2010) (Table 1).

Table 1: Bibliographic summary of the chromatin modifiers involved in NC formation.

	Enzyme	ON/ OFF	Downstream targets (role in NC development)	Organism	Reference
NC specification	*aepb2* (PRC2) (KMT)	Off	*snai2* (migration), *sox10* (NC specification)	Mice	Kim et al. (2011)
	brg1 (SWI/SNF family)	On	*snai2* (NC specification)	Zebrafish	Eroglu et al. (2006)
	chd7 (chromo)	On	*foxd3* (NC specification)	Mice NC-derived stem cells	Fujita et al. (2016)
	chd7-pbaf (chromo)	On	*sox9*, *twist*, *slug* (NC specification)	*Xenopus*	Bajpai et al. (2010)
	ezh2 (PRC2) (KMT)	Off	Cooperation with *snai2* (NC specification, migration, craniofacial and cartilage formation)	*Xenopus*	Tien et al. (2015)
	hdac	Off	*bmp4*, *cadherin 6B*, *sox9*, *sox10*,*hnk1* (Neural tube closure and NC specification	Chicken	Murko et al. (2013)
	hdac1-ets1	Off	*id3* (NC formation)	*Xenopus*	Wang et al. (2015)
	jmjD2A (KDM)	On	*sox10* (NC specification)	Chicken	Strobl-Mazzulla et al. (2010)
	prd12 (HMT)	Off	*foxd3*, *slug*, *sox8* (NC formation)	*Xenopus*	Matsukawa et al. (2015)
NC migration	*dnmt3B*	Off	*sox10* (NC migration)	Chicken	Hu et al. (2014)
	hdac-sin3a-phd12	Off	*cadherin-6b* (EMT)	Chicken	Strobl-Mazzulla & Bronner (2012)

	Enzyme	ON/ OFF	Downstream targets (role in NC development)	Organism	Reference
NC differentiation	*dnmt3A*	Off	*gbx2* (Ear formation)	Chicken	Roellig & Bronner (2016)
	ezh2 (PRC2) (KMT)	Off	*hox* genes (craniofacial formation)	Mice	Schwarz et al. (2014)
	g9a (KMT)/*dnmt 3*	Off	*lef1* (craniofacial, brain and retina formation)	Zebrafish	Rai et al. (2010)
	hdac1	Off	*foxd3* (melanophore formation)	Zebrafish	Ignatius et al. (2008)
	hdac1	Off	*dlx2/3*, *hoxb3a* (craniofacial formation)	Zebrafish	Ignatius et al. (2013)
	hdac3	Off	Cardiac smooth muscle differentiation	Mice	Singh et al. (2011)
	hdac4	Off	Craniofacial formation	Zebrafish	DeLaurier et al. (2012)
	hdac8	Off	*otx2*, *lhx1* (craniofacial formation)	Mice	Haberland et al. (2009)
	kat6a	On	Nitric Oxide (craniofacial formation)	Zebrafish	Kong et al. (2014)
	kat6a	On	*hox* genes (craniofacial formation)	Zebrafish	Miller et al. (2004)
	kdm6a	On	*msx* (craniofacial formation)	Zebrafish	Qi et al. (2010)
	ring1b/rnf2 (PRC1)	Off	(Chondrocyte differentiation)	Zebrafish	van der Velden et al. (2013)

1.2.2.2. Cis-regulatory elements associated with NC

Enhancers are cis-regulatory elements required for cell specification and differentiation. Enhancers (1) coordinate the expression of important genes for a given cell type, and (2) act as a transcription factor-binding platform recognized by major lineage specifiers and effectors of signaling pathways (Remeseiro et al. 2016). It is not surprising then that the NC cells, as a major feature in vertebrates, have an extensive network of cis-regulatory elements. A general *in silico* study of human induced neural crest cells (hNCCs) identified *tfap2a* and *nrf1*/2 as important transcription factors found in most active NC enhancers, alongside p300 and H3K24ac in the surrounding nucleosomes, thus allowing permissive chromatin in hNC enhancers and therefore transcription activation of NC genes (Rada-Iglesias et al. 2012).

Since the first study in 1994, which characterized the mouse *wnt1* enhancer driving expression in cranial and trunk NC cells (Echelard et al. 1994), more research has been done aiming to identify enhancers important for the expression of key NC genes. For instance, an *ets1* enhancer (ECR1) was found to be a 694bp sequence located upstream of the *ets1* TSS and conserved in birds and mammals. This region can drive expression of *ets1* in cranial NC cells and contains TF binding sites for *sox9*, *ets1*, *tfpa2a*, *foxd3*, *pax7* and *msx1*/2, all involved in activating *ets1*. Interestingly, it contains a *cmyc* binding site that represses *ets1* expression in the hindbrain (Barembaum & Bronner 2013). *gbx2* is an important gene required for NC induction. Analysis of the *Xenopus gbx2* upstream landscape revealed a 500bp regulatory region that contains a ß-*catenin*/*TCF* binding site, indicating the role of Wnt signaling in *gbx2* activation (Li et al. 2009). The *pax3* enhancer is conserved between mouse and human and is located within the 1.6kb region upstream *pax3* TSS. It is formed by two 200bp regulatory regions (NCE1 and NCE2) spaced by 156bp. NCE2 has the ability to drive *pax3* expression in the dorsal neural tube and neural crest cells and contains a *tead2* binding site (Milewski et al. 2004). Another study identified *pax3* enhancers in zebrafish. IR1 and IR2 are two regulatory regions found in *pax3* introns that drive gene expression in the neural plate border and dorsal tube. The response to signaling pathways was analysed revealing that whereas IR1 requires Wnt and FGF for its function and it is upregulated by BMPs, IR2 is FGF dependent but its function is inhibited by BMP (Garnett et al. 2012). The *pax7* enhancer (P7eNBPE) directs expression of *pax7* in neural plate border, neural folds and early migrating NC, but it is not responsible for its later expression. P7eNBPE has been described in chicken and has very little conservation with other

organisms, being postulated as a specie-specific enhancer. It is located upstream of *pax7* and contains a *cmyb* binding site, revealing a new role for *cmyb* as a *pax7* regulator (Vadasz et al. 2013). *foxd3* contains two regulatory elements (NC1 and NC2) conserved in chicken, mice, human and *Xenopus* that together recapitulate *foxd3* expression but they show differences in regulation: NC1 drives expression in cranial neural crest cells and contains binding sites for *pax7*, *msx1*/2 and *ets1* whereas NC2 is necessary for trunk *foxd3* expression and is regulated by *pax7*, *msx1*/*2* and *zic1* (Simões-Costa et al. 2012). Several regulatory regions for *sox10* have been identified: Sox10E1 and Sox10E2 located 1.1kb and 2.1kb downstream of *sox10*, respectively, and Sox10L8 6.1kb upstream of the TSS. Both Sox10E1 and Sox10E2 regulate *sox10* expression in cranial NC cells in a rostocaudal temporal manner. These regulatory regions are conserved in chicken, opossum, mouse, rat, human but not in *Xenopus* or zebrafish. Considerable analyses have been carried out on Sox10E2 regulatory region. *sox9*, *ets1* and *cmyb* are required for initial *sox10* expression but *cmyb*, *pea3*, *sox8* are necessary for ear formation at later stages (Betancur et al. 2010b; Betancur et al. 2011). The Mef2c-F10N enhancer conserved in mice, human, rat, rabbit and chicken but not in aquatic species, drives the expression of *ß-galactosidase* in migrating NC cells during early stages and later in cranial NC cells, peripheral nervous system and melanocytes overlapping with *msx1/2* and *pax3/7* (Aoto et al. 2015). Finally, the *mitf* enhancer is a 298bp located 14.5kb from the TSS containing two functional *sox10* binding sites. It is known to be conserved in mice and human and can drive the expression of *mitf* gene in pigment cells (Watanabe et al. 2002). As very few studies report enhancers conserved in *Xenopus*, it would be required to identify cis-regulatory elements involved in *Xenopus* neural crest formation.

1.3. *Schmidtea mediterranea* as a model organism

The ability to regenerate body structures, studied for the first time by Ferchault de Réaumur in 1712, is a widespread phenomenon across the animal kingdom. In each phyla there are animals with the ability to regenerate limbs, organs and/or other appendices (Li et al. 2015). *Schimdtea mediterranea* are freshwater planarians extensively used in biological studies in order to understand regenerative body processes. Several characteristics make of *S. mediterranea* a great model to study regeneration: (1) this species can regenerate a whole new organism within days, and (2) they are the only species able to regenerate a

complete and functional nervous system *de novo* (Cebrià 2007). Moreover, (3) their genome is partially sequenced, there are available transcriptomes, and a wide range of molecular techniques such as RNA interference, *in situ* hybridization, immunohistochemistry and cell sorting are easy to carry out in these organisms (Pearson et al. 2009; Sanchez Alvarado & Newmark 1999; Hayashi et al. 2006).

Planarians are Platyhelminthes belonging to the lophotrochozoa clade (Baguñà & Riutort 2004). They are acoelomates and possess the three germ layers. Planarians have anteroposterior and dorsoventral well-defined body axis and bilateral symmetry. The main organs and tissues present in the organism are (Figure 11): (1) a central nervous system (CNS), which consists of two cephalic ganglia and two longitudinal nerve cords connected by commissural neurons, (2) a digestive system composed by three blind gut branches and a pharynx that allows ingestion and ejection of food, (3) a muscular system formed by longitudinal, diagonal and circular fibres, (4) an excretory system, a dorsal network of flame cells, (5) a multiciliar epithelium that allows locomotion, and (6) the parenchyma, i.e. the space between the gut and the epidermis (Saló 2006).

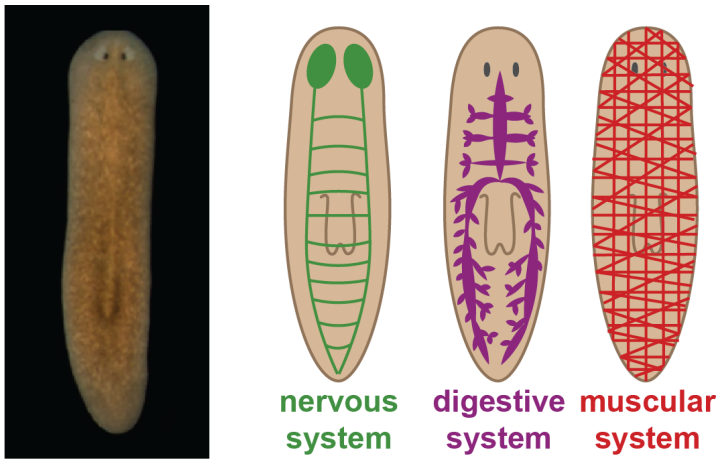


Figure 11: Planarian anatomy. An image of a wild type organism and a schematic representation of the nervous system, digestive system and muscular system are shown.

1.3.1. Neoblasts, the planarian stem cells

The regeneration of the planarian *S. mediterranea* occurs thanks to the induction of neoblast proliferation, which forms a non-pigmented tissue, called blastema, where cells differentiate in order to form all missing structures. The neoblasts represent 20-30% of the cells present in an organism and are located in the parenchyma surrounding the intestinal branches except in the region anterior to the photoreceptors and the pharynx. Neoblasts have a large nucleus/cytoplasm

ratio and they are the only mitotic cells in the organism; therefore when they are depleted by irradiation, regeneration is impaired (Zhu & Pearson 2016; Saló 2006).

There are two models which aim to explain the neoblasts lineage development: (1) the 'naïve model' suggests that there is an homogeneous population of stem cells and the lineage progression occurs in the progeny, whereas (2) the 'specialized model' sustains that there is a lineage restricted population of dividing cells (Rink 2013; Reddien 2013). This question has been addressed in the last decades leading to the conclusion that there is a small population of pluripotent stem cells together with lineage-restricted neoblasts (Figure 12). A recent study has identified a small population of clonogenic neoblasts (cNeoblasts) (Wagner et al. 2011). When one cNeoblast is injected into an irradiated planarian, it is able to replace all host cells and restore the regeneration capabilities of the organism. Other studies analysing the transcription profile of neoblasts led to the identification of four populations of linage-restricted neoblasts: gamma neoblasts (Υ) which give rise to intestine progeny; sigma neoblasts (σ) which give rise to eyes, pharynx and muscle and massively contribute to the 6h and 48h proliferation peaks after amputation; zeta neoblasts (ζ) which give rise to the epidermis; and finally nu neoblasts (ν) which give rise to the brain. Modelling the transition between the different neoblast subpopulations leads to a model where cNeoblasts can generate σ-Neoblasts which can produce Υ- ζ- and ν-neoblasts (Molinaro & Pearson 2016; Van Wolfswinkel et al. 2014).

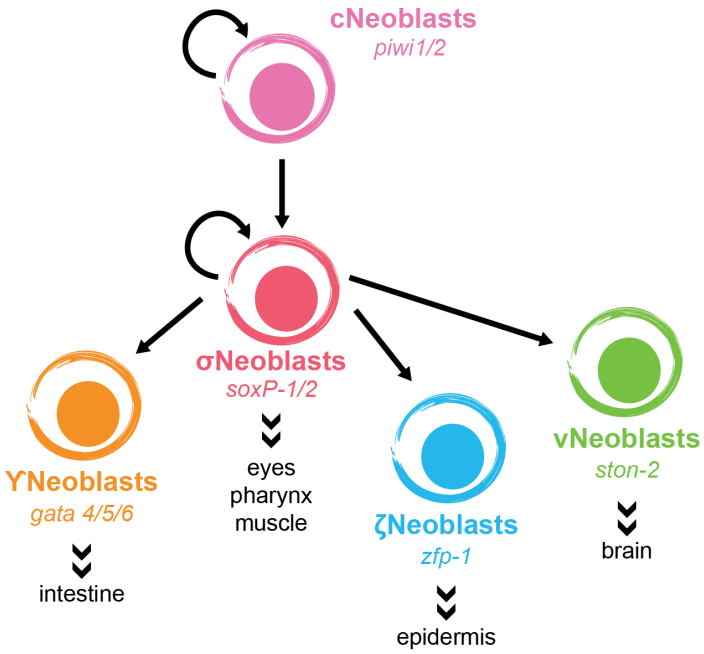


Figure 12: Current model of neoblast differentiation. cNeoblasts, characterized by *piwi1/2* expression, self-renew and give rise to the sigma neoblasts (σNeoblasts), characterized by the expression of *soxP-1/2*, which give rise to the eye, pharynx and muscle cells. σNeoblasts can give rise to the gamma neoblasts (ϒNeoblasts) that form the intestine, the zeta neoblasts (ζNeoblasts) that give rise to the epidermis and the nu neoblasts (νNeoblasts) that give rise to the brain. The ϒ, ζ, ν Neoblasts are characterized by the expression of *gata4/5/6*, *zfp1* and *ston-2*, respectively.

1.3.2. Planarian regeneration

It is known that regeneration in *S. mediterranea* involves several processes, taking place all at the same time: (1) wound healing, (2) cell death, (3) neoblast proliferation, (4) patterning, and (5) regeneration of organs and tissues (Saló 2006). The most relevant processes for the purpose of this thesis are going to be described focusing on the specification of anteroposterior patterning.

1.3.2.1. Wound healing

Wound healing and scar formation is a common process for all organisms after damage. However, not all organisms are able, after that, to restore missing structures. In planarians, there is a muscle contraction 30 seconds after injury in order to reduce the area and time of exposure of the inner tissues (Saló 2006). Following that, the external epithelium extends to cover the wound. Grafting experiments conclude that the contact between the dorsal and ventral epithelium triggers the regenerative response as well as the contact between areas with different anteroposterior identities (Salo & Baguñà 1985; Kato et al. 1999).

The kinetics of gene expression during wound healing has lead to the identification of four gene populations (Wenemoser et al. 2012). W1 genes are expressed within the 30min following amputation and correspond to genes such as *jun-1*, *fos-1* and *egr-3* known as 'immediate early response'. W2 genes are induced between 3h and 12h and fall under the category of secreted factors involved in cell positioning identity such as *wnt1*, *nlg1* and *follistatin*. W3 genes are induced between 6h and 12h and correspond to remodeling matrix factors. Finally, W4 are expressed in neoblasts and are necessary for the differentiation of new tissues, for example, *runt-1* is required for the formation of the nervous system (Sandmann et al. 2011).

It has also been shown that different mechanisms are activated in response to different types of injuries, mainly depending on whether tissue is lost (i.e. amputation or notching vs incision or puncture). *Follistatin* inhibiting *activin* is a pathway just activated in response to missing tissues thus activating the second wave of neoblast proliferation that allows regeneration of tissues. JNK signaling also promotes neoblast proliferation just on tissue-lost injures (Gaviño et al. 2013; Almuedo-Castillo et al. 2014).

1.3.2.2. Cell death

For the regeneration process to occur it is important to have a balance between cell death and proliferation. Cell death is essential for remodeling tissues as well as generating proportionate organs and tissues after amputation. On the other side, cell proliferation and differentiation are key to generate new tissues.

Two mechanisms of cell death have been associated with regeneration in planarians: apoptosis and autophagy. Regarding apoptosis, there are two main peaks of apoptosis after amputation. The first peak occur at 4 hours post-amputation and it is localized at the wound-site whereas the second peak at 3 days post-amputation is a systemic response involved in tissue remodeling. The depletion of neoblasts does not affect the kinetics and magnitude of the apoptotic response, suggesting that the cell death is on differentiated cells and neoblast-independent (Pellettieri et al. 2010).

The second cell death mechanism associated with regeneration is autophagy. The study of *death-associated protein-1*, *dap-1*, a positive regulator of autophagy, let to the conclusion that autophagy is necessary for tissue remodeling. *dap-1* is expressed in non-mitotic neoblasts and its expression is increased upon

amputation. The morphologic analysis of the cells expressing dap-1 indicates that these cells are undergoing autophagy and its inhibition affects tissue remodeling (González-Estévez et al. 2007).

1.3.2.3. Neoblasts proliferation

The proliferation of neoblast is a key process for generating new organs and tissues. There is a bimodal proliferation response to damage. The first peak at 6h post-injury is a systemic response to any type of damage. It is thought to be generated by neoblasts on G2 or late S phase of mitosis. Therefore, the signals promoting proliferation affect the transition between G2 to M phase. The second proliferation peak at 48h post-injury only occurs when there is tissue loss and it is localized on the wound-site. It involves migration of the neobasts close to the affected area to the post-blastema where they proliferate, forming the blastema. The neoblasts responsible for this second peak are thought to be in G0 or G1 phase when the amputation occurs so the second activation of proliferation is associated with signals affecting the progression from G0/G1 to M phase. Between these two peaks of cell division there is a minimum of 18h post-amputation (Saló & Baguñà 1984; Wenemoser & Reddien 2010).

Two signaling pathways have been found to be involved in controlling neoblast proliferation after injury. mTOR pathway is necessary to activate the first mitotic peak. The JNK pathway determine the transition between G2 to M phase, involved in both proliferating peaks, and it is also required to induce the wound-site apoptotic peak. Therefore, *jnk* knockdown organisms have an increase in neoblasts proliferating in the first mitotic peak, and an earlier and more sustained second peak, together with a decrease in apoptosis resulting in organisms unable to restore body proportions (González-Estévez et al. 2012; Almuedo-Castillo et al. 2014).

1.3.2.4. Head vs tail regeneration

After amputation, planarians must accomplish three main milestones to restore their head or tail: (1) determine the amount of material that is missing and if it has anterior or posterior identity; (2) form the anterior and posterior pole *de novo*, which will act as a instructive center; and finally (3) reconstruct tissues and organs (Owlarn & Bartscherer 2016).

Canonical Wnt signaling is the main pathway involved in anteroposterior patterning. Wnts are a family of secreted glycoproteins that bind to a Frizzled receptor, activate Dishevelled, which inactivates the ß-catenin destruction complex,

mainly composed of Axin, APC and GSK-3. This stabilizes ß-catenin, which can then translocate to the nucleus and activate transcription of target genes in association with TCF/Lef transcription factors. A Wnt morphogen gradient, high levels promoting tail and low levels required for head formation, has been proposed to be responsible for cell positioning. Inhibition of Wnt signaling produced, for example, by silencing *ß-catenin,* leads to a fully anteriorized organism, the phenotype of which is dose dependent: complete inhibition result in a 'radial-like hypercephalized' animal while lower inhibition result in 'two-head phenotype' with or without ectopic eyes. Activation of the Wnt pathway by *APC* or *axins* knockdowns therefore produce a 'two-tail phenotype' (Iglesias et al. 2008; Iglesias et al. 2011).

Although Wnt is essential for anteroposterior polarity, other pathways such as the Hedgehog signaling are involved in establishing and maintaining polarity. Knockdown of positive regulators of hedgehog (*hedgehog*, *smoothened*, *gli-1*) fails to regenerate a tail whereas depletion of negative regulators of Hh (*patched*, *suppressor of fused*) fails to regenerate the head (Rink et al. 2009; Yazawa et al. 2009).

The main elements for the head vs tail decision are *wnt1*, a Wnt ligand, and *notum*, an inhibitor of the Wnt signaling pathway. *wnt1* knockdown produces a 'two-head phenotype'. In intact planarians, it is expressed in a row of discreet cells in the posterior dorsal midline (Figure 13). After 6h post-amputation, *wnt1* is expressed in individual cells in both the anterior and posterior wound-sites, but its expression gets restricted to the posterior part of the animal at 48h, in a similar pattern to an intact organism. *notum* is a secreted protein that hydrolyzes the Wnt molecule in a way that Wnt is unable to bind its receptor. It is expressed in a few cells at the anterior pole of the organism. After amputation it is also expressed in both blastemas, preferentially in the anterior one, and at 48h its expression gets restricted to the most anterior tip of the blastema. Silencing *notum* generates a 'two-tail organism'. The first *wnt1* and *notum* expression in both blastemas is not affected by depletion of neoblast, suggesting their expression is stem cell independent and wound-induced (Figure 13) (Owlarn & Bartscherer 2016; Petersen & Reddien 2011).

The current model to reestablish anteroposteror polarity has two phases: first, the identity establishment (0h-24h after amputation) and later, the maintenance of the identity, previously determined (from 24h to 72h) (Figure 13). The first gene expressed after amputation is *hh* that activates *wnt1* expression in both wound-

sites. Later *wnt1* will be restricted to the posterior wound allowing the formation of a tail. *wnt1* activate, through *ß-catenin*, *notum*, in both wound-sites. In the anterior blastema, *notum* indirectly inhibits *ß-catenin* allowing head formation whereas in posterior, *notum* expression, through an unknown mechanism, is reduced leading to the formation of the tail (Almuedo-Castillo et al. 2012).

A lot of effort has been put to identify genes involved in the pole formation and re-patterning once the identity has been determined. In the posterior wound-site, *islet*, a LIM homeobox transcription factor; *pbx*, a TALE class homeodomain protein; and *pitx*, a paired class homeobox/pituitary homeobox are required for maintaining *wnt1* expression and the transcription of other posterior Wnts such as *wnt11.1* and *wnt11.2* and posterior Hox genes (*hoxD* and *abd-ba*) (Hayashi et al. 2011; Chen et al. 2013; Blassberg et al. 2013; März et al. 2013). The inhibition of any of these fail to regenerate the tail, similar to the phenotypes observed when *ß-catenin* or *wnt1* are inhibited. Although these genes are important for posterior regeneration, they are not only expressed in the posterior pole but also in the anterior. Moreover, inhibition of these genes also leads to cyclops planarians, which are planarians with one eye, suggesting that these genes are also involved in anterior regeneration.

On the anterior wound-site, *pbx* and another TALE class homeodomain protein, prep; *foxD*, a Forkhead-family transcription factor; *zic1*, a member of the Zinc finger of the cerebellum protein family; and *follistatin*, a secreted regulator of TGF-β proteins; are involved in anterior pole formation. These genes are co-expressed in a small subset of cells in the midline regions alongside *notum*, on the new-forming anterior pole, and are important for maintaining the late expression of *notum* and genes involved in maintaining anterior identity such as the frizzled receptor, *sFRP-1,* and *ndl-4* (*nou darake-like gen 4*). The inhibition of any of these genes leads to defects in anterior structures such as the brain (Chen et al. 2013; Felix & Aboobaker 2010; Scimone et al. 2014; Vásquez-Doorman & Petersen 2014; Roberts-Galbraith & Newmark 2013).

All this positional information is active in the adult, as it is not only important for regeneration but also for cellular turnover in intact animals. A recent study identified the subepidermal muscle cells as the ‘position control cells’ (PCCs) secreting all the positional genes. The muscles cells are therefore involved not only in contraction but provide body coordinates for maintaining the adult body plan. PCCs are in charge of specifying the identity of new cells during homeostasis and

after injury. These cells change the expression of the positional genes, such as *prep* and *ndl-4* in anterior and *wnt11.1* and *wnt11.2* in posterior, they are expressing in order to ensure the new cells acquire the correct identity to regenerate the missing structures (Witchley et al. 2013).

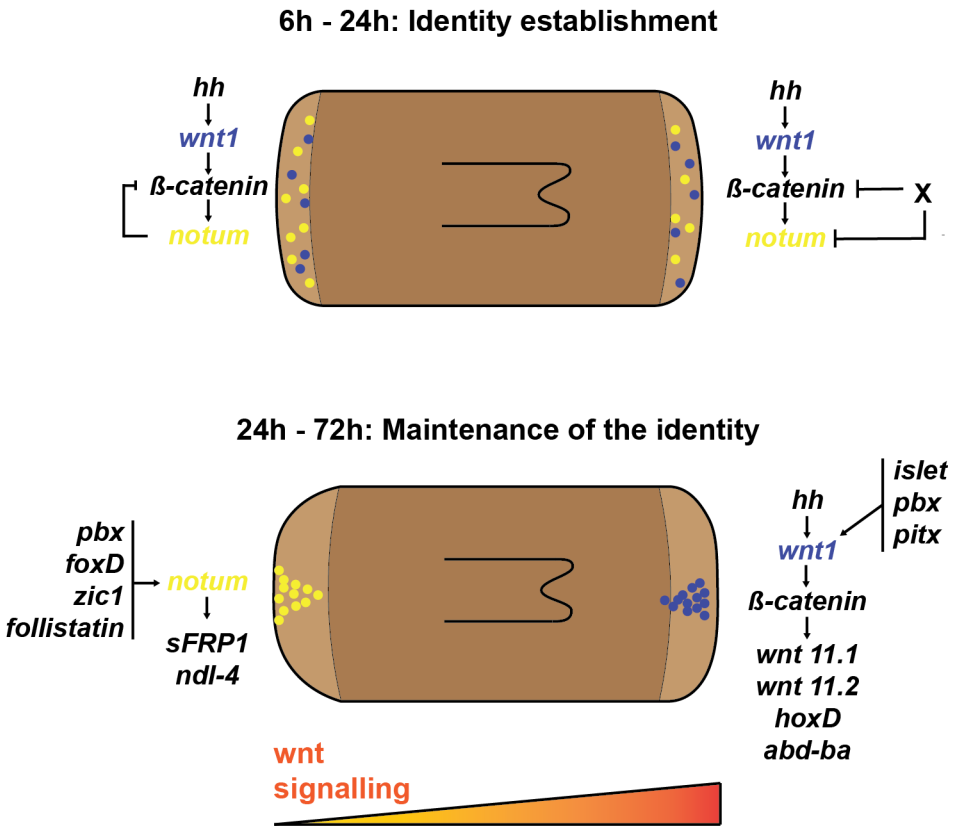


Figure 13: Proposed model for establishment of anteroposterior identities after amputation. (A) Identity establishment. After wound-healing, *wnt1* activates *ß-catenin* which in turn activates *notum*. *notum* inhibits the function of *ß-catenin* in the anterior pole but not in the posterior pole where an unknown element inhibits *notum*. (B) Maintenance of the identity. After 2 days post-amputation, the downstream targets of canonical Wnts such as *sFRP1* and *ndl-4* in anterior and the posterior Wnts and hox are activated. The anterior identity is maintained thanks to the action of *pbx*, *foxD*, *zic1* and *follistatin* and the posterior to *islet*, *pbx* and *pitx*. Anterior to the left and posterior to the right.

1.3.3. Epigenetic regulation of the regeneration

The study of the enhancers associated to the regeneration is a novel field. Only three enhancers, which activate gene expression after injury, have been identified in different organisms. In zebrafish, the len enhancer is located 7kb upstream the *lepb* TSS and activate gene expression after fin or heart amputation (Kang et al. 2016). Despite not being conserved in other organisms, a mice

transgenic line containing the len enhancer with a minimal promoter is able to drive reporter expression in amputated digit and damaged hear ventricles. A *bmp5* enhancer has been identified in mice that drive *bmp5* expression only after adult bone or soft tissue injury. This region is different from the enhancers that drive *bmp5* expression during mice development (Guenther et al. 2015). Finally, an enhancer close to a Wnt gene-encoding locus in *Drosophila* has been identified which activate expression upon injury, allowing the regeneration of the imaginal discs. After L3, polycomb-mediated epigenetic silencing of this enhancer takes place impairing the regeneration of mature imaginal discs (Harris et al. 2016).

In planarians, the CRE involved in regeneration have not been identified yet. However, a recent study shows that the epigenetic modifiers and histone modifications are conserved in planarians. These are mainly expressed in the parenchyma, where the neoblasts are present (i.e. around the digestive system and not in the pharynx or anterior to the photoreceptors) or in the neural system (i.e. in the brain and nerve cords) (Robb & Sanchez Alvarado 2014). *HDAC1-1* is expressed in neoblasts and when inhibited the regeneration is impaired. The inhibition of *HDAC1-1* could be causing cell-cycle arrest preventing neoblast to proliferate or differentiate (Robb & Sanchez Alvarado 2014). Furthermore, the role of the H3K4 methyltransferases Set1 and MLL1/2 in regeneration have been studied (Duncan et al. 2015). The results indicated that Set1 and MLL1/2 have different biological roles: whereas Set1 targets promoters of genes involved in neoblast maintenance, MLL1/2 targets promoters of genes involved with motility.

1.4. Aims and objectives

Cis-regulatory elements in the genome are known to control animal development. These elements contain multiple target sites for transcription factors and collectively control expression of developmental gene regulatory networks. Therefore, CRE analysis provides a method to study transcription factor interactions. In this study, ATAC-sequencing will be used to identify cis-regulatory elements in the genome to answer two biological questions: (1) how NC genes are regulated in *Xenopus* embryos and (2) how the anteroposterior polarity of planarians is established during regeneration?

The first biological question will be addressed by characterizing the epigenome during neural crest formation. For that, the specific objectives are:

- To identify epigenetic modifiers which specifically affect NC formation.
- To validate NC-induced animal caps as a method to obtain an enriched population of NC cells through transcriptomic analysis.
- To optimize the ATAC-sequencing protocol for animal cap samples.
- To identify neural crest exclusive regulatory elements at two different stages during NC formation: NC specification (stage 13) and NC differentiation (stage18).
- To identify overrepresented motifs in NC enhancers that could drive expression of NC genes.
- To identify enhancers of key NC genes and test their functionality.

In order to address the second biological question (i.e. to characterize the genomic regulation of the anterior and posterior poles of planarians, which define head versus tail identity), the specific objectives to accomplish are:

- To ensure that data analysis of the ATAC-sequencing is relevant in planarians.
- To identify exclusive open regions for anterior and posterior pole formation during regeneration.
- To combine ATAC-sequencing and ChIP-sequencing data to identify active enhancers involved in anterior and posterior poles of planarians at 12 hours regeneration.
- To identify overrepresented transcription factor DNA-binding sequence motifs in anterior vs posterior 12h regenerating blastemas.

CHAPTER II:
MATERIALS AND METHODS

2.1. *Xenopus* materials and methods

2.1.1. X. tropicalis embryo obtainment

Both male and female *X. tropicalis* adults were primed 24h before eggs are required with 10 units of Chorulon (human Chorionic Gonadotropin) into the dorsal lymph sac and let them rest at 26°C overnight without being fed. *X. tropicalis* males were euthanized by immersing it in 1g of ethyl 3-aminobenzoate methane sulfonate dissolved in 200ml reversed osmotic water in a 500ml beaker and placed in the dark for more than 1 hour at 4°C. The testes were surgically removed just before the egg collection and store them in Leibovitz's L-15 Medium supplemented with 10% Calf serum at 4°C.

Females were induced 5-6 hours before the eggs were required by injecting 200 units of Chorulon (human Chorionic Gonadotropin) into the dorsal lymph sac and left at 26°C. *X. tropicalis* eggs were obtained by gently applying pressure to the abdomen to encourage eggs to be released. Immediately, eggs were fertilized with a mashed section of testes dissolved in Leibovitz's L-15 Medium supplemented with 10% Calf serum and let at room temperature for 5 minutes. After that embryos were immersed in 0.05x MMR (Marc's Modified Ringer's) (50mM NaCl, 1mM KCl, 0.5mM MgCl2, 1mM CaCl2, 2.5mM HEPES pH 7.5) for 20 minutes at room temperature and rinsed several times with 0.05x MMR. Embryos were stored in a BSA (bovine serum albumin) coated Petri dish in 0.05X MMR at 26°C.

2.1.2. X. laevis embryo obtainment

Adult females were primed by injecting 100units of PMSG (pregnant mare serum gonadotropin) into one pericardial sac 5-10 days before eggs were required and left them isolated at room temperature without being fed. 12-18h before eggs collection, adult female were induced by injecting 250units of Chorulon in each pericardial sac and left overnight at 18°C. *X. laevis* eggs were obtained by gently squeezing the animals with a massaging action and collecting them in a 25ml petri dish every hour over a period of 6 hours.

Before eggs collection, a *X. laevis* male were euthanized by immersing in 1g of ethyl 3- aminobenzoate methane sulfonate dissolved in 300ml reversed osmotic water in a 500ml beaker. The beaker is wrapped in foil and placed in the dark for more than 1 hour at 4°C. Then, the testes were surgically removed and stored in 80% FBS (fetal bovine serum) in 1x MMR (100mM NaCl, 2mM KCl, 1mM MgCl2,

2mM CaCl2, 5mM HEPES pH 7.5) at 4°C. Having the *X. laevis* eggs in a 25ml Petri dish, a proportion of testes was dissolved with 1ml of 1xMMR, distributed over the embryos and left for 5 minutes at 18°C. Embryos were then immersed in 0.1X MMR for 20min at 18°C. After this embryos were rinsed with 0.1X MMR and stored in a BSA coated Petri dish in 0.1X MMR.

2.1.3. Dejellying embryos

Xenopus embryos are enveloped in a transparent jelly coat, which needs to be removed post fertilization. 0.1XMMR (*X. laevis*) or 0.05X MMR (*X. tropicalis*) were poured off the embryos and replaced with 2% L-Cysteine pH 8 for approximately 7 minutes. Embryos were subsequently washed in 1x MMR and 0.1x MMR (*X. laevis*) or 0.05X MMR (*X. tropicalis*).

2.1.4. Embryo fixation

Once *Xenopus* embryos reached the appropriate stage they were fixed using MEMFA (3.7% formaldehyde, 1XMEM salts), then washed in PBST (PBS, 0.1% Tween) and dehydrate in a serial dilution of Ethanol and kept at 4°C.

2.1.5. Chemical screen

The compounds (Table 2) were kept at 10mM with DMSO solvent at 4°C or -20°C, depending of the specificities of the compounds. 7 embryos per well were placed in a 24 well-plate and incubated with the appropriate drug dilution from stage 15 to stage 40 or from stage 9 to stage 15 (Tomlinson et al. 2009; Schmitt et al. 2014).

Table 2: Epigenetic inhibitors tested in the chemical screen.

Drug	Target	Source
ZM 447439	Aurora Kinase	Selleckchem (Prof Rob. Field)
Danusertib (PHA-739358)	Aurora Kinase	Selleckchem (Prof Rob. Field)
AT9283	Aurora Kinase	Selleckchem (Prof Rob. Field)
Barasertib (AZD1152-HQPA)	Aurora Kinase	Selleckchem (Prof Rob. Field)
SNS-314 Mesylate	Aurora Kinase	Selleckchem (Prof Rob. Field)
CYC116	Aurora Kinase	Selleckchem (Prof Rob. Field)
JNJ-7706621	Aurora Kinase	Selleckchem (Prof Rob. Field)

PHA-680632	Aurora Kinase	Selleckchem (Prof Rob. Field)
CCT129202	Aurora Kinase	Selleckchem (Prof Rob. Field)
TAK-901	Aurora Kinase	Selleckchem (Prof Rob. Field)
AMG-900	Aurora Kinase	Selleckchem (Prof Rob. Field)
PF-03814735	Aurora Kinase	Selleckchem (Prof Rob. Field)
GSK1070916	Aurora Kinase	Selleckchem (Prof Rob. Field)
CCT137690	Aurora Kinase	Selleckchem (Prof Rob. Field)
M344	Aurora Kinase	Selleckchem (Prof Rob. Field)
Decitabine	DNA Methyltransferase	Selleckchem (Prof Rob. Field)
Azacitidine	DNA Methyltransferase	Selleckchem (Prof Rob. Field)
5-Aza-2-deoxicytidine	DNA Methyltransferase	Sigma
5-azacitidine	DNA Methyltransferase	Prof. A.Ganesan
PFI-1 (PF-6405761)	Epigenetic Reader Domain	Selleckchem (Prof Rob. Field)
NU9056	HAT	Tocris
L002	HAT	Tocris
PU141	HAT	Prof. Jung
PU139	HAT	Prof. Jung
CPTH2	HAT	Prof. A.Ganesan
(S)-ESBA	HAT	Santa Cruz Biotecnology
H3-CoA-20-TAT	HAT	PhD, David J. Meyers
M3-3	HAT	Sigma
C626	HAT	Abcam
CPTB	HAT	Abcam
MG149	HAT	Prof. FJ. Dekker
Trichostatin A (TSA)	HDAC	Selleckchem (Prof Rob. Field)
Vorinostat (SAHA, MK0683)	HDAC	Selleckchem (Prof Rob. Field)
Entinostat (MS-275)	HDAC	Selleckchem (Prof Rob. Field)
Belinostat (PXD101)	HDAC	Selleckchem (Prof Rob. Field)
PCI-24781 (Abexinostat)	HDAC	Selleckchem (Prof Rob. Field)
LAQ824 (Dacinostat)	HDAC	Selleckchem (Prof Rob. Field)

Quisinostat (JNJ-26481585)	HDAC	Selleckchem (Prof Rob. Field)
Mocetinostat (MGCD0103)	HDAC	Selleckchem (Prof Rob. Field)
Valproic acid sodium salt (Sodium valproate)	HDAC	Selleckchem (Prof Rob. Field)
CUDC-101	HDAC	Selleckchem (Prof Rob. Field)
Droxinostat	HDAC	Selleckchem (Prof Rob. Field)
MC1568	HDAC	Selleckchem (Prof Rob. Field)
Pracinostat (SB939)	HDAC	Selleckchem (Prof Rob. Field)
PCI-34051	HDAC	Selleckchem (Prof Rob. Field)
Givinostat (ITF2357)	HDAC	Selleckchem (Prof Rob. Field)
AR-42	HDAC	Selleckchem (Prof Rob. Field)
Tubastatin A HCl	HDAC	Selleckchem (Prof Rob. Field)
PI3K/HDAC Inhibotor 1	HDAC	Selleckchem (Prof Rob. Field)
CI994 (Tacedinaline)	HDAC	Selleckchem (Prof Rob. Field)
Sodium Phenylbutyrate	HDAC	Selleckchem (Prof Rob. Field)
Rocilinostat (ACY-1215)	HDAC	Selleckchem (Prof Rob. Field)
Scriptaid	HDAC	Selleckchem (Prof Rob. Field)
Burkholdac B	HDAC	Prof. A.Ganesan
FG-4592	HIF	Selleckchem (Prof Rob. Field)
2-Methoxyestradiol (2-MeOE2)	HIF	Selleckchem (Prof Rob. Field)
IOX2	HIF	Selleckchem (Prof Rob. Field)
GSK J4 HCl	Histone demethylases	Selleckchem (Prof Rob. Field)
Entacapone	Histone Methyltransferase	Selleckchem (Prof Rob. Field)
BIX 01294	Histone Methyltransferase	Sigma
Pargiline	Histone Methyltransferase	Prof. A.Ganesan
SRT1720	Sirtuin	Selleckchem (Prof Rob. Field)
Resveratrol	Sirtuin	Selleckchem (Prof Rob. Field)
EX 527 (Selisistat)	Sirtuin	Selleckchem (Prof Rob. Field)
Quercetin	Sirtuin	Selleckchem (Prof Rob. Field)
Sirtinol	Sirtuin	Selleckchem (Prof Rob. Field)

2.1.6. Imaging

Images of embryos were captured using Q imaging 01-MP3.3-RTV-CLR-10 camera mounted on a Zeiss Stemi 5V6 microscope. Images were then processed using Inkscape/Illustrator and Photoshop.

2.1.7. Transformation using DH5α cells

1µl of DNA plasmid was added to 100µl of DH5α competent cells and kept on ice 30min. After that, cells were incubated 90s at 42°C and incubated again 5min on ice. 300µl of LB was added and incubate 1h at 37°C at 200rpm. Finally, 100-200µl was plated onto a LB with carbicillin. These were left at 37°C overnight to form colonies.

2.1.8. DNA Midiprep

Colonies from the transformation were taken to be incubated with 50 ml LB with 50µl of carbicillin, or the appropriate antibiotic, shaking overnight at 37°C. Plasmid DNA was isolated using the Qiagen Midiprep plasmid purification kit according to the manufactures instructions. The concentration was measured with a nanodrop (Thermofisher).

2.1.9. Probe synthesis

Plasmids containing the gene of interest, *foxd3* and *zic1* (plasmids obtained from Prof. Yoshiki Sasai and Dr. Jung Aruga, respectively), were linearized by restriction enzyme using 2µl of the appropriate enzyme (BamHI and EcoRI, respectively), 5-10µg of plasmid, 5µl of the restriction enzyme buffer in a total volume of 50µl. The reaction was incubated overnight at 37°C. The linearized plasmid was then purified through Ethanol precipitation by adding 5µl of NaAc 3M pH 5.2, 250µl of Ethanol and let overnight at -20°. The mixture was centrifuged and the pellet was resuspended in 15µl of water.

To synthesis the DIG-probe the following reaction was used: 1µg of purified linearized plasmid, 4µl of 5x Transcription Buffer (Promega), 2µl of DTT (100mM), 1µl of dNTP containing DIG labelled UTP (Roche), 1µl RNAsin (Promega) and 2µl of RNA polymerase made to a final volume of 20µl. The polymerases to transcribe the probes for *foxd3* and *zic1* were T7 and T3, respectively. The reaction was incubated

3h at 37°C and the probes were purified using G50 column, according to the manufacturers instructions.

2.1.10. Whole mount in situ hybridization (WISH)

X. tropicalis embryos were rehydrated in PBST (PBS, 0.1% Tween). All the washes were performed 5min at room temperature in a rocker. Embryos were then incubated with 10µg/ml ProteinaseK for 2min. Embryos were rinse twice 5min in 0.1M triethanolamine pH 7-8 and then washed 25µl acid anhydride/10ml 0.1M triethanolamine and finally in 50µl acid anhydride/10ml 0.1M triethanolamine. Embryos were then washed twice in PBST for 5min and incubated in 3.7% formaldehyde/PBST for 20min at room temperature after which they were washed 3 times in PBST for 5min and incubated 1hour at 60°C in hybridization buffer (50% Formamide, 5XSSC, 1mg/ml Torula RNA, 100µg/ml Heparin, 1X Debharts solution, 0.1% Tween, 0.1g/100ml Chaps, 10mM EDTA). Embryos were incubated with the probe in hybridization buffer (1/500) overnight at 60°C.

Probes were removed and stored at -20°C. Embryos were washed 20min at 60°C three times in 2X SSC and two times in 0.2X SSC (SSC 20X: 175g NaCl, 88.2g Sodium citrate total volume 1l and pH 7). Embryos were then washed twice in MABT (100mM Maleic acid, 150mM NaCl, pH 7.5, 0.1% Tween) 30min at room temperature and then transferred into 2% BMB Blocking/MABT for 1hour at room temperature with rocking. Blocking solution was replaced with antibody solution (2% BMB, 20% goat serum in 1XMAB) containing anti-digoxigenin (1:2000) and incubated overnight at 4°C with rocking.

Antibody solution was removed and embryos were washed in MABT five times for 60min each at room temperature with rocking. Embryos were then washed twice in alkaline phosphate buffer (100mM TrisHCl pH 9.5, 50mM $MgCl_2$, 100mM NaCl, 0.1% Tween). Embryos were then put in NBT/BCIP with alkaline phosphate buffer (67.5µl of 75mg/ml NBT and 52.5µl of 50ml/ml of BCIP made up to 15ml of alkaline phosphate buffer) until the desire level of colour was observed. Embryos were then put in 5XTBST (125ml 1M TrisHCl pH 7.5, 40g NaCl, 1g KCl 50ml of Tween made up to 500ml). Embryos were kept overnight in 5XTBST to remove the background and then were photographed. To store them, embryos were kept in methanol at 4°C.

2.1.11. Capped RNA synthesis

10µl of plasmid DNA containing *X. laevis noggin* and *wnt1* (plasmids obtained from Dr. Dave Hsu and Dr. Oliver Destree, respectively) were digested using 5µl of restriction enzyme buffer, 2µl of restriction enzyme (NotI for *noggin* and XbaI for *wnt1*) made up to 50µl with water. The reaction was incubated overnight at 37°C and then purified using Qiagen PCR Purification kit. For capped RNA synthesis, 1µg of linearized plasmid, 10µl of NTP/CAP, 2µl of buffer, 2µl of Sp6 RNA polymerase made up to 20µl was incubated 2hours at 37°C. 1µl of Turbo DNAse was added for 15min at 37°C. The capped RNA underwent lithium chloride precipitation and quantified.

2.1.12. Animal cap assay

Embryos were injected at two cell stage with ether *noggin* (to induce neuroectoderm), *noggin* and *wnt*1 (to induce neural crest) or non-injected (which gives ectoderm). These were left to develop until stage 9 when the animal pole was cut using sharp forceps after removing the vitelline membrane. Caps were cut in filtered-sterile 1XMMR and transferred to plates containing 0.7XMMR, 1mg/ml BSA, 100µg/l gentamycin until non-dissected sibling embryos had reached stage 13 or 18 (Figure 14).

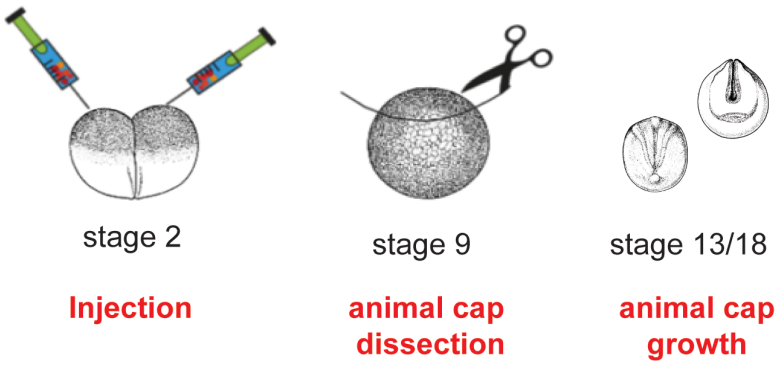


Figure 14: Animal cap assay. Embryos were injected at two-cell stage and left to develop until stage 9 when the animal pole was dissected. Caps were left develop until sibling embryos reached stage 13 or 18.

2.1.13. ATAC-sequencing

One *X. laevis* animal cap was incubated in a 1ml eppendorf with 500µl of liberase/trypsin solution (1:10) at room temperature for 10 minutes, pipetting every 3 minutes to dissociate the tissue. The solution was transferred to a 50ml falcon tube

containing 4ml Hank's solution (1xHBSS, 10mM HEPES, 0.25%BSA). The eppenford was rinsed with 1ml of Hank's solution to recover more material. The falcon was centrifuged 800g 8min, the supernatant was discarded and the pellet was resuspended in 4ml of Hank's solution. The solution was passed through a 40µm cell strainer and centrifuged again 8min at 800g. The supernatant was removed and the pellet was resuspended in 1ml of cold 1XPBS and transferred to a new eppendorf. The sample was centrifuged 500g, 10min at 4°C, the supernatant was removed and the pellet was resuspended in 50µl of cold 1XPBS. The sample was centrifuged again 500g, 10min at 4°C, the supernatant was removed and resuspended in 50µl of cold lysis buffer (5µl of Tris HCl pH 7.4, 1.5µl of $MgCl_2$, 10µl of 10% NP40, 1µl of NaCl 5M made up to 500µl with nuclease free water). 5µl of the resuspended sample were taken to count the number of nuclei in a Neubauer chamber. The volume corresponding to 50,000cells was transferred into a new eppendorf. The sample was centrifuged 500g, 10min at 4°C and resuspended in transposition reaction (10µl of 2X TD Buffer, 1µl Tn5, 9µl nuclease free water). The reaction was incubated 30min at 37°C. 2.4µl EDTA 0.5M were added and incubated at 55°C for 30min. After that, 1.2µl $MgCl_2$ was added and the samples were purified using Qiagen Mini Cleanup Kit, according to the manufacturers, and resuspended in 10µl of nuclease free water. The DNA fragments were amplified by carrying out the following PCR reaction: 10µl of transposase DNA, 10µl on nuclease free water, 2.5µl of 25µM adapter 1, 2.5µl of 25µM adapter 2.X and 25µl of NEBNext High Fidelity PCR 2X and cycling as follow: 75°C for 5min, 98°C for 30s, 16 cycles of 98°C for 10s, 63°C for 30s and 72°C for 1min, and hold at 4°C. The PCR reaction was purified using Qiagen Mini PCR Purification and resuspended in 10µl of nuclease free water. AMPure beads purification SRI1X was carried out and the libraries were quantified using Qubit.

2.1.14. RNA extraction and cDNA synthesis

Embryos were snap frozen in liquid nitrogen and kept at -80°C before RNA extraction. RNA extraction was performed using High Pure RNA Isolation kit (Roche), according to the manufacturers and samples were resuspended in 30µl of water. DNAse treatment on column was carried out. 1µg of RNA was taken to synthetize cDNA using the Maxima First Strand cDNA synthesis (Themofisher).

2.1.15. Quantitative PCR

The quantitative PCR (qPCR) reaction mixture was as follows: 2µl cDNA, 1µl of 2.5µM Primer forward and reverse dilution, 5µl of Syber green and 2µl of water. The PCR reaction cycled 50°C for 2min, 95°C for 10min, and 40 cycles of 95°C for 15s and 60°C for 1min. Amplification curves were analysed using 7500 sequence detection software.

Table 3: *Xenopus* qPCR primers.

Name	Forward primer	Reverse primer
FoxD3	TCTCTGGGGCAATCACACTC	GTACATTTGTTGATAAAGGG
Snai2	TCCCGCCACTGAAAATGCCACGATC	CCGTCCTAAAGATGAAGGGTATTCCTG
Sox2	GAGGATGGACACTTATGCCCAC	GGACATGCTGTAGGTAGGCGA
Keratin	GGAGCTAGCCAACCATGAAC	CTGCCAGCTTGGCATTATCT
Gapdh	CTTTGATGCTGATGCTGGAA	GAAGAGGGGTTGACAGGTGA
Xbra	GAATGGTGGAGGCCAGATTA	TTCATTCTGGTATGCGGTCA
MyoD	TACACTGACAGCCCCAATGA	TGCAGAGGAGAACAGGGACT
Id3	CAAGGGACCAGGTATGGATG	CCTGGCACCAACTCTTTCAG
Ocd	CATGGCATTCTCCCTGAAGT	TGGTCCCAAGGCTAAAGTTG
cMyc	GAAACACCCATCAGCAGACC	CTCTTCCTCGTCGCAGTCT
cMyc promoter 1	TGGTTCCTAGCTGTCAGTGG	TCCATGCAGGAGACAGAGAG
cMyc promoter 2	GGGCAGGATTTCTCACTTTC	ATAGCAGCTCCACGTCTCCT
cMyc promoter 3	GCTGCTGCTGCTATTACTGC	TAAATCCACAGGCTGTCGTG
cMyc gene body 1	ATCTGAGAAGCTGGCATCCT	GTGGACTCTGGCACTGAGAA§
cMyc gene body 2	CTCAGTGAAATGCTGGGAGA	TAGAAAGCATTTGGCGTCTG
cMyc gene body 3	CTGAGGCTTGTGGTTCTTCA	TTTGAGCAACAGCTGGAAAC

2.1.16. Embryo morpholino injection

X. tropicalis embryos undergoing injections were placed at two cell stage in 3% Ficoll PM400 (Sigma). Both blastomeres were injected with 100ng standard control morpholino or Cdk9-a morpholino, in the animal pole, using a 14nl calibrated

needle and oxygen free nitrogen was used with a Harvard apparatus injector (Medical Systems Research products). The injector was set at P_{out}=90, $P_{balance}$=0.6 and P_{inject}=16. 2 hours after injection the ficoll was removed and replaced with 0.5XMMR with gentamycin (1:1000) and left to develop at 26°C.

2.1.17. ChIP-sequencing

Approximately 50 embryos per sample were fixed at stage 16 in 1% formaldehyde for 30min. Embryos were then washed in 125mM glycine solution 30 minutes and then transferred gradually to 0.1X MMR. Embryos were homogenized in sonication buffer (20mM TrisHCl pH 8, 70mM KCl, 1mM EDTA, 10% glycerol, 5mM DTT, 0.125% NP40, protease inhibitors (Roche #04693132001) and stored at -80C. After sonication, chromatin samples were incubated with RNA Pol II antibody (diagenode C15200004) 1ug per 15 embryo equivalent in IP buffer overnight (50mM TrisHCl pH 8, 100mM NaCl, 2mM EDTA, 1%NP40, 1mM DTT and protease inhibitors) followed by incubation with Dynabeads Protein G for 1h. Beads where subsequently washed in ChIP 1 buffer (IP buffer + 0.1% deoxycholate), ChIP 2 buffer (ChIP 1 buffer + 400mM NaCl), ChIP 3 buffer (ChIP 1 buffer + 250mM LiCl) and TE buffer. Material was digested with proteinase K and DNA was purified using a QIAGEN kit. Q-PCR was performed pre and post library preparation for quality checking and DNA libraries were prepared using Kapa Hyper Prep kit (Kapa Biosystems).

2.1.18. I-SceI meganuclease trangenesis

The protocol was obtained form Ogino et al. (2006). The plasmid containing the enhancer, minimal promoter and the reporter gene were first digested. 2µl of 10X I-SceI buffer, 2µl of I-SceI (New England Bio Labs), 8µl of plasmid at 0.1µg/µl and 8µl of water were incubated at 37°C for 40min. The reaction was injected in 1cell stage *X. laevis* embryos. The embryos were injected in a period of 45min after being fertilized. Embryos were kept in 3% Ficoll at 12°C until they reached 4 cell stage and then transferred to 0.3X MMR with 50µg/ml gentamycin and kept at 18°C until they underwent gastrulation. Then the embryos were placed at 22°C. The injector was set at P_{out}=0.7 and P_{inject}=10.

2.2. Planarians materials and methods

2.2.1. Organisms maintenance

Planarians used in these experiments were from a clonal strain of the *S.mediterranea* BCN-10 biotype and were maintained at 18°C in PAM 1X water (93.5µg NaCl, 246.5µg $MgSO_4x7H_2O$, 100.8µg $NaHCO_3$, 7.5µg KCl, 20.35µg $MgCl_2x6H_2O$, 147µg $CaCl_2x2H_2O$ made up in 1l) and feed once-twice a week with organic beef liver. Planarians were starved for 1 week and were 4 to 6 mm in length when used for experiments.

2.2.2 Double stranded RNA (dsRNA) synthesis

dsRNA was prepared from *in vitro* transcription reactions using PCR-generated templates. For that primers for *notum* (TCGAGTGATTTGTGGTCTGG and TGAAGCTAGATTTATGTGAAAAACCA) and *wnt1* (AACAATCAAATATTCGTCA and TGTTAGGAAAGGTCGGTTGC) were used. For the in vitro transcription, 1µg of amplified DNA, 2µl of transcription buffer 10X, 2µl of 10mM NTP, 1µl RNAse inhibitor, 1µl of RNA polymerase made up to 20µl were incubated 4h at 37°C. This reaction was set to transcribe the sense and the antisense RNA. Both strands were mixed and the dsRNA was purified through phenol – chlorophorm extraction.

2.2.3. Injections

dsRNA was injected in the prepharingeal region using a Drummond Scientific Nanoject. 3632-nl of dsRNA was injected for three consecutive days. After that, planarians were amputated pre and postpharingeal (one round of injections). Two rounds or injections were required. Control animals were injected with water or a dsRNA of *gfp* (see section 4.3).

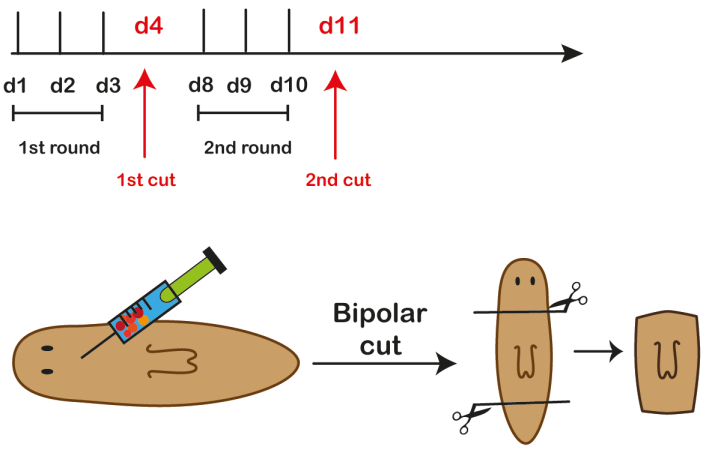


Figure 15: Schematic representation of the dsRNA injection methodology. Planarians were injected three days in the prepharyngeal region and then performed a bipolar cut. After three days the same protocol was carried out.

2.2.4. ATAC-sequencing

The same ATAC-sequencing protocol as for *Xenopus* (see 2.1.13) was followed. The organisms were washed in 2%L-Cystein pH 7 for 2min and then transferred to a petri dish containing CMFH (2.56mM $NaH_2PO_4x2H_2O$, 14.28mM NaCl, 10.21mM KCl, 9.42mM $NaHCO_3$, 1%BSA, 0.5%Glucose, 15mM HEPES pH 7.3). 20 blastemas were cut and transferred into an eppependorf. CMFH was removed and blastemas were dissociated in a solution of liberase/CMFH (1:10).

2.2.5. ChIP-sequencing

2000 anterior and posterior blastemas were fixed in groups of 100 blastemas. For that, organisms were place in 1M $MgCl_2$ for 15-30s rocking and then transferred to PBS. Following that, blastemas were fixed in 1.85%formaldehyde for 15mintutes rocking at room temperature. Glycine to a final concentration of 0.125M was added to quench the formaldehyde and it was incubated 5min at room temperature. Blastemas were washed three times in cold PBS1X and the samples were stored at -80°C. Samples were homogenize in 5ml of cell lysis buffer (10mM Tris-HCl pH 7.5, 10mM NaCl, 0.3% NP-40, Complete 1X). Samples were centrifuged at 3500rpm, 5min at 4°C, the supernatant was discard and the pellet resuspended in 660µl Nuclear lysis buffer (50mM Tris-HCl pH 7.5, 10mM EDTA, 1% SDS, Complete 1X) diluted by adding 1.340ml of ChIP Dilution buffer (16.7mM Tris-HCl pH 7.5, 1.2mM EDTA, 167mM NaCl, 0.01% SDS, 1.1% Triton X-100). Samples were split in 1ml aliquots and sonicated in a Covaris M220 with settings: duty factor=10%, PIP=75W, 100=cycles/burst, time=10'. Samples were centrifuge 5min, max speed at 4°C and

the supernatant were collected. 20µl were kept as an input control. 5/3x200µl of sonicated chromatin were incubated with 2µl of H2K27ac (ab4729) at 4°C, overnight rotating. Samples were incubated with 50µl of Dynabeads protein G resuspended in ChIP Dilution Buffer 1h at 4°C. Beads were washed twice in Wash Buffer 1(20mM Tris-HCl pH 7.5, 2mM EDTA, 150mM NaCl, 1% SDS, 1% Triton X-100), twice in Wash Buffer 2 (20mM Tris-HCl pH 7.5, 2mM EDTA, 500mM NaCl, 0.1% SDS, 1% Triton X-100) and twice in Wash Buffer 3 (10mM Tris-HCl pH 7.5, 1mM EDTA, 250mM LiCl, 1%NP-40, Na-deoxycholate 1%) and twice in 10mM Tris-HCl pH 8. Beads were resuspended in TAGmentation reaction mix (10 mM Tris-HCl pH 8.0, 5 mM $MgCl_2$, 10 % w/v dimethylformamide), 1µl of Tn5 was added and incubated for 1min. Beads were then washed twice in Wash Buffer 1 and once in TE 1X. The samples were eluted 15min in 100µl of Elution Buffer (50mM $NaHCO_3$ pH 8.8, 1%SDS) and 180µl of Elution Buffer was added to the input sample. 10µl of 4M NaCl and 0.5µl of 10mg/ml proteinase K were added to the ChIP and input sample and it was incubated 4-6h at 65°C. DNA was purified using Minelute columns (Qiagen). The library was prepared by PCR: 19µl of DNA, 1µl of 25µM of adapter1 and 2.X, 25µl of NEBNext High-Fidelity 2X PCR Master Mix and 4µl of water which was cycled 98°C for 30s, 98°C for 10s 63°C for 30's 72°C for 30 and 72°C for 5min. The number of cycles was empirically determined. The libraries were purified using Minelute columns (Qiagen). 1/100 of each ChIP sample, 1/10, 1/100, 1/1000 of the inputs were used in a qPCR with positive and negative primers according to 2.1.15.

Table 4: Planarian qPCR primers.

Name	Forward primer	Reverse primer
neg	AATCAAAATTGTTGGGGGTTT	TGACTCAGCTGGGTTCTTCA
Wnt1	TTCCATCTTTTATTCTCAGAGTTTG	TTGATTGGATAAAAATGAGGAGTT
follistatin	GACAAAATTATGCAATTCTTTTTACA	TTTCCTTTGGTAATCGAAACAA
notum	TATTGCCTATTGCCCCTTTG	GAAGCACCCTGCTGCTACTC

2.3. *In silico* methods

2.3.1. RNA-sequencing data analysis

Xenopus RNA-sequencing data performed using Kallisto 0.43.0 against the XL_9.1_v1.8.3.2.primaryTranscripts.fa transcriptome.

```
>kallisto index -i Xl_transcriptome.idx XL_9.1.fa
>kallisto quant -i Xl_transcriptome.idx -o <output> <R1> <R2>
```

The TPM were analysed with edgeR. The PCA plot, scatter plot of differentially expressed genes according a p-value=0.5, and the selection of 100 top differently expressed genes were generated using the following commands:

```
> data <- readDGE(files, columns = c(1,4))
> keep <- rowSums(cpm(data)>1) >= 2
> data <- data[keep, , keep.lib.sizes=FALSE]
> y <- calcNormFactors(data)
> y$samples
> plotMDS(y) #PCA plot
> Stage <- factor(c("st9","st9","st9","st13","st13","st13"))
> Frog <- factor(c(1,2,3,1,2,3))
> design <- model.matrix(~Stage+Frog)
> rownames(design) <- colnames(y)
> data <- estimateDisp(y,design)
> fit <- glmFit(data,design)
> lrt <- glmLRT(fit,coef=2)
> topgenes <- topTags(lrt,n=100,sort.by = "PValue", p.value = 0.05)
> top <- rownames(topgenes)
> table_topgenes <- cpm(y)[top,]
> summary(dt<- decideTestsDGE(lrt, p.value = 0.05))
> isDE <- as.logical(dt)
> DEnames <- rownames(data)[isDE]
> plotSmear(lrt, de.tags = DEnames) # Smear plot of differential
expressed genes
> write.table(topgenes, file = "file.txt", sep="\t") #Generate the
table of top differential expressed genes
```

2.3.2. ATAC-sequencing data analysis

Both *Xenopus* and planarian data was analysed using the following commands. The *Xenopus* data was analysed using the *X. laevis* genome Xla.v91.repeatMasked.fa and the features file XL_9.1_v1.8.3.2.gff3. The planarian data was analysed using the *S. mediterranea* genome SmedAsxl_genome_v1.1.nt and the features file SmedAsxl_genome_v1.1.all.gff. Reads were first analysed using bowtie1 with the argument X2000, the distance between the two reads can be up to 2000bp, as the default is 250bp. For *X. laevis* the -m1 argument was used to detect only reads that map in one place.

```
> bowtie-build <genome.fa> <ebwt_base>
> bowtie -X2000 –m1 <indexed genome> -1 <R1.fastq>  -2 <R2.fastq> -S
<output.sam>
```

The sam files were converted to bam, the mapped reads were filtered and the forward reads were then shifted 4bp and reverse reads, 5bp, due to the Tn5 properties

```
> samtools view -b -T <genome.fa> -o <output.bam> -S <input.sam>
> samtools sort <input.bam> <outout.bam>
> samtools view -b -F 4 -o <output.bam> <input.bam>
>   awk   '{if($6=="+"){$2=$2+4}   else   if   ($6=="-")   {$3=$3-5}
print$1"\t"$2"\t"$3"\t"$4"\t"$5"\t"$6}' <input.bam> > <output.bam>
```

Taking the output of this, the tracks to be visualised on the genome browser were generated using F-seq. First individual files for each chromosome/scaffold were generated and then combined in a single bigwig file that can be loaded to the UCSC genome browser.

```
> bedtools bamtobed –i  <input.bam> > <output.bed>
> fseq -of wig -f 0 -o <output_index> <input.bed>
> cat 0h/*.wig | grep -v track > <combinedwig.wig>
> wigToBigWig <combinedwig.wig> <chrom.sizes> <output>
```

To perform peak calling using MACS2 the following command was followed. The peaks common in the three replicated were selected. Peaks overlapping 50%, were considered the same peak.

```
>callpeak -t <input.bam> -n <output_index> -f BAM -g <genome_size>

>intersectBed -wo -a <rep1.narrowPeak> -b <rep2.narrowPeak> | awk
'BEGIN{FS="\t";OFS="\t"}{s1=$3-$2; s2=$13-$12; if (($21/s1 >= 0.5) ||
($21/s2>=  0.5))  {print  $0}}'  |  cut  -f  1-10  |  sort  |  uniq
|intersectBed   -wo   -a   stdin   -b   <rep3.narrowPeak>   |awk
'BEGIN{FS="\t";OFS="\t"}{s1=$3-$2; s2=$13-$12; if (($21/s1 >= 0.5) ||
($21/s2  >=  0.5))  {print  $0}}'  |  cut  -f  1-10  |sort  |  uniq  >
<sample.narrowPeak>
```

To identify peaks shared between different samples in order to generate the Venn diagrams and to identify peaks present in the different gene location the following commands were followed:

a. Peaks present in A and B but not in C:

```
>intersectBed -wo -a <A.narrowPeak> -b <B.narrowPeak> | awk
'BEGIN{FS="\t";OFS="\t"}{s1=$3-$2; s2=$13-$12; if (($21/s1 >= 0.5)
|| ($21/s2 >= 0.5)) {print $0}}' | cut -f 1-10 | sort | uniq |
intersectBed   -wao   -a   stdin   -b   <C.narrowPeak>   |   awk
'BEGIN{FS="\t";OFS="\t"}{s1=$3-$2; s2=$13-$12; if (($21 == 0) ||
($21 == 0)) {print $0}}' | cut -f 1-10 | sort | uniq | wc –l
```

b. Peaks present A but not in B or C:

```
> intersectBed -wao -a <A.narrowPeak> -b <B.narrowPeak> | awk
'BEGIN{FS="\t";OFS="\t"}{s1=$3-$2; s2=$13-$12; if (($21 == 0) ||
($21 == 0)) {print $0}}' | cut -f 1-10 | sort | uniq |
intersectBed -wao -a stdin -b <C.narrowPeak> | awk
'BEGIN{FS="\t";OFS="\t"}{s1=$3-$2; s2=$13-$12; if (($21 == 0) ||
($21 == 0)) {print $0}}' | cut -f 1-10 | sort | uniq | wc –l
```

c. Peaks present in A, B and C

```
> intersectBed -wo -a <A.narrowPeak> -b <B.narrowPeak> | awk
'BEGIN{FS="\t";OFS="\t"}{s1=$3-$2; s2=$13-$12; if (($21/s1 >= 0.5)
|| ($21/s2>= 0.5)) {print $0}}' | cut -f 1-10 | sort | uniq
|intersectBed -wo -a stdin -b <C.narrowPeak> |awk
'BEGIN{FS="\t";OFS="\t"}{s1=$3-$2; s2=$13-$12; if (($21/s1 >= 0.5)
|| ($21/s2 >= 0.5)) {print $0}}' | cut -f 1-10 |sort | wc –l
```

d. Peaks present in gene location

```
> intersectBed -u -a <input.narrowPeak> -b <genomic_location.bed >
<output.bed>
```

Finally, the motifs present in then different peak were identified using Homer:

```
> findMotifsGenome.pl <narrowPeak> <genome.fa> <outout/> -size given
```

To calculate the coverage in the different peaks to analyse the enhancer dynamic, bedtools coverage was used:

```
> bedtools coverage -a <narrowPeak> -b <replicate.bed > <output.bed>
```

CHAPTER III:
EPIGENOMIC REGULATION OF *XENOPUS* NEURAL CREST CELLS

3.1. Rationale

The main purpose of this chapter was to identify epigenetic elements relevant for *Xenopus* neural crest formation, i.e. for neural crest specification and/or differentiation. This has been subdivided in the following objectives.

The first objective was the identification of epigenetic modifiers involved in *Xenopus* NC formation. To do this, we propose to carry out a chemical screen using a library of small molecules that are known inhibitors of DNA and histone modifiers. This type of technique is a simple first approach to selecting candidate genes, as it requires no prior knowledge. Moreover, the *Xenopus* model organism is ideal for this approach as the different phenotypes are easy to visualize (Tomlinson et al. 2009; Kälin et al. 2009).

The second objective was the identification of cis-regulatory elements relevant for NC formation in *Xenopus*. As it has been pointed out, there is little conservation of CREs, excepting a small number of highly conserved non-coding regions, and there are few studies aiming to identify neural crest regulatory elements in *Xenopus*, most of them being performed in chick or mouse embryos (for instance see Betancur et al. 2010; Milewski et al. 2004). Moreover, most of the studies have been performed analysing potential regulatory regions next to candidate genes rather than carrying out a whole genome analysis (Simões-Costa et al. 2012; Vadasz et al. 2013; see section 1.2.2.2 for more examples). The main reason behind this is that the neural crest cell population represents a small percentage of the total embryo and most of the whole genome techniques require 3-5 orders of magnitude more material (Buenrostro et al, 2013). The development of new techniques based on the transposase Tn5 enzyme such as ATAC-sequencing and ChIPmentation, which allow whole genome analysis with a small amount of starting material, will enable cell type analysis. According to Buenrostro et al. 2013, DNase-sequencing requires 3-5 orders of magnitude more cells than ATAC-sequencing, generating similar results. Therefore, it is of interest to perform a whole genome analysis of neural crest cells to identify enhancers involved in this cell type formation. In order to obtain information specifically about NC cells, ATAC-sequencing is going to be performed not only in NC-induced animal caps but also in neuroectoderm and ectoderm-induced animal caps. The following bioinformatics analysis will reveal putative NC-specific enhancers and their putative upstream transcription factors.

3.2. Epigenetic modifiers involved in neural crest formation

3.2.1.Chemical screen using a library of epigenetic inhibitors in X. tropicalis

Chemical screens are assays that utilize small molecules to alter the function of specific gene products and infer the role of such a gene in development, based on the obtained phenotype. These are easy-to-perform, cheap and can be used to inhibit a gene product at specific time points. Moreover, they can be used to discover new drugs and test their toxicity levels. The use of *Xenopus* is advantageous because hundreds of eggs are obtained frequently, they can take compounds from the media and the phenotypes are easy-to-score (Tomlinson et al. 2009; Schmitt et al. 2014).

A library of 66 inhibitors of epigenetic modifiers (Table 2) was used in a chemical screen performed in *X. tropicalis* embryos. Those compounds include, among others, inhibitors for HAT, DMTS, HMT families, the importance of which has been previously mentioned in relation to NC formation. In order to see the phenotype after the treatment with the inhibitors, embryos at stage 15 were incubated with the compounds for three days at 26°C until they reached stage 38-40, when the phenotype was scored (Figure 16A). The phenotypic categories were assigned according to the main defect the embryo presented, designating one phenotype for each compound. The phenotypes (Figure 16B) were distributed into the following categories: (1) normal development, when the treated embryo looked similar to the control, treated with the appropriate amount of solvent; (2) general development, when the embryos showed general body shape alteration or the development was pronouncedly delayed compared with the control; (3) edema formation, when the embryo presented an increase of fluid around the heart cavity; (4) melanocyte morphology, when pigment cells did not present the dendritic shape characteristic to this cell type; (5) melanocyte migration, when pigment cells presented an abnormal pattern, usually located in stripes along the dorsal spine; (6) pigmentation defects, when embryos did not show any pigmentation or it was very reduced; (7) craniofacial defects, when embryos had obvious defects in the craniofacial structures; and (8) eye morphology defects, when the eye structures were not present or had an abnormal morphology. Compounds were first added at a final concentration of 50μM. If after the screen, embryos did not show any phenotype the concentration of the drugs was increased to 100μM; in contrast, if embryos were dead or very damaged the concentration was reduced to 25μM and 10μM. The screen was replicated twice to ensure the phenotypes were due to the

presence of the drug in the media. Out of the 66 compounds, 28 were associated to normal development; 18 to general development; 4 to edema formation; 3 to melanocyte morphology; 2 to melanocyte migration; 6 to pigmentation defects; 3 to craniofacial defects and 2 to eye morphology defects (Figure 16C).

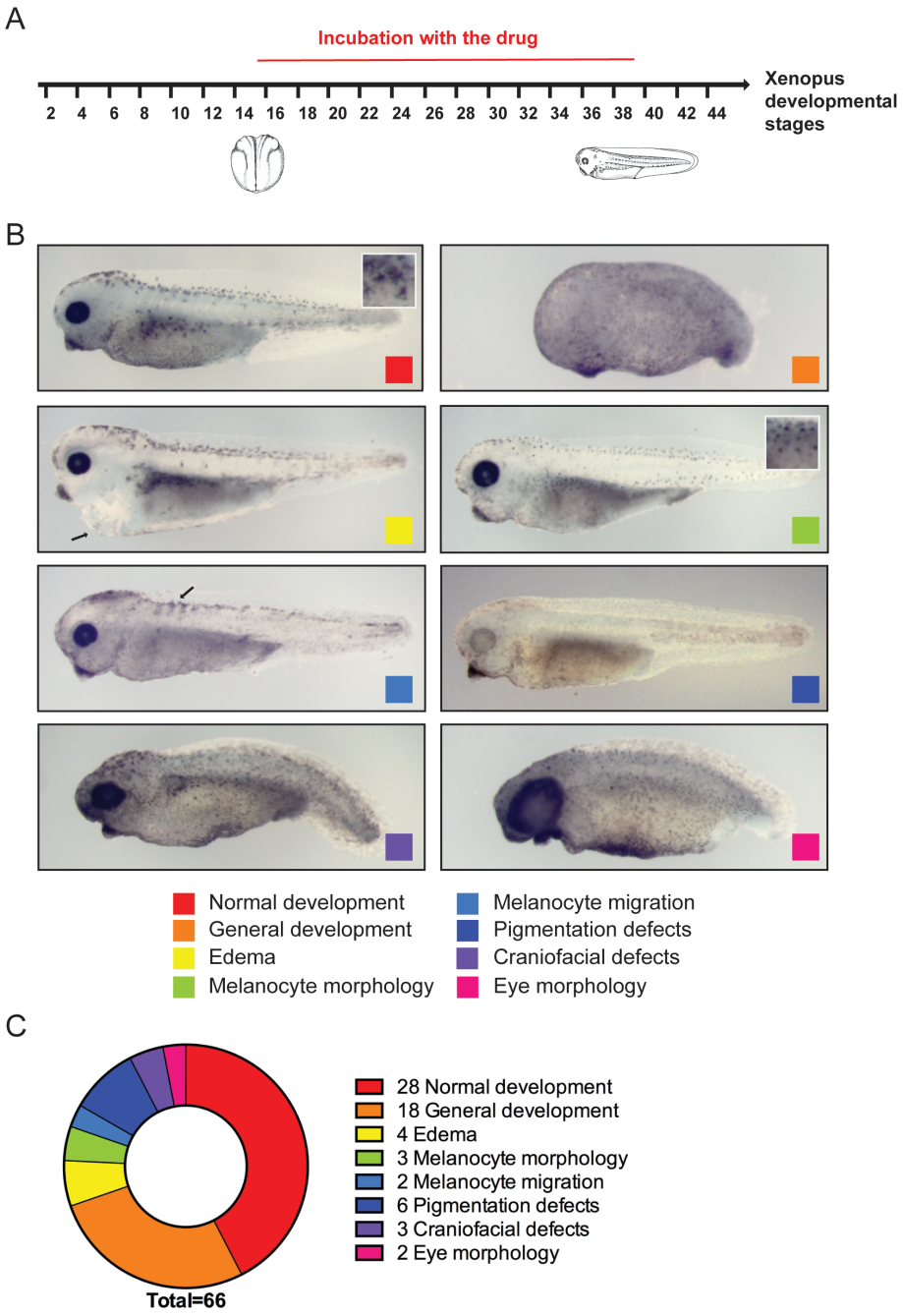


Figure 16: Results of the screening using epigenetic inhibitors. (A) Strategy followed for the experiments, embryos were incubated with the drug dissolved in the media from stage 15 to stage 38-40. (B) Phenotypic categories in which embryos are distributed with an example of the phenotype. White square shows a closer look to the melanocytes. Arrows show the main defect. (C) Distribution of the drugs into the different phenotypic categories.

3.2.2. Analysis of compounds associated with pigment loss

It was noticed that all the compounds associated with pigment loss were related to changes in histone acetylation, either histone acetyltransferases (HAT) or histone deacetylases (HDAC). Therefore, further characterization of these families of inhibitors was carried out to determine the optimal concentration and affects on NC formation (Figure 17 and Figure 18).

Regarding HDACs (Figure 17A), SAHA, also called Vorinostat, is a general inhibitor of HDAC that when applied at 40μM showed complete loss of pigmentation (Richon et al. 1998). At increasing concentration toxic defects such as bent spine and reduced size are observed. HDACs are mainly divided in four families depending on sequence homology: class I (HDAC 1-3, 8), class II (HDAC 4-7, 9-10), class III (SIRT1-7) and class IV (HDAC11) (Dawson & Kouzarides 2012). Burkholdac B is a compound that targets exclusively HDAC class I (Benelkebir et al. 2011). At 50μM it also showed pigment and developmental defects similar to SAHA, suggesting that SAHA and Burkholdac B might be inhibiting the same enzymes. The other compounds that inhibit HDAC mainly affect general development.

Regarding HATs (Figure 17B), there are many available compounds targeting several members of this family. PU139 (Gajer et al. 2015) is a general HAT inhibitor, which showed the most severe loss-of-pigmentation phenotype even in the eyes. The nuclear HATs are divided into three families: GNAT (containing KAT2A/B members), MYST (containing Tip60 member) and CPB/p300 (Dawson & Kouzarides 2012). CPTH2 (Chimenti et al. 2009) inhibits KAT2B a member of GNAT family producing at high doses a partial loss of pigmentation coupled with melanocyte morphology defects. Tip60 is inhibited by the compound NU9056 (Coffey et al. 2012) and barely affected pigmentation. PU141 (Gajer et al. 2015) inhibits the p300/CBP family and also reduced embryo pigmentation. L002 compound inhibits p300 whereas CTPB enhances p300 (Yang et al. 2014; Balasubramanyam et al. 2003). The treatment with CTPB at 100μM did not show any pigmentation defect while the treatment with L002 at 12μM clearly reduced the pigmentation of the embryo. C646 (Bowers et al. 2010) is another p300/CBP inhibitor but its inhibition mainly affected the development of the embryo.

To address if the defects on pigmentation were caused by an alteration of early neural crest formation, embryos at stage 8-9 were treated with the optimum concentration previously determined and allowed to develop until stage 15. Then a

whole mount *in situ* hybridization (WISH) using probes against a neural plate border gene, *zic1* and a neural crest specifier, *foxd3*, was carried out (Figure 18). None of the embryos did show defects in either *zic1* or *foxd3* expression suggesting that theses enzymes are affecting neural crest cells at later stages or they are affecting the production of melanin, melanogenesis (Figure 18).

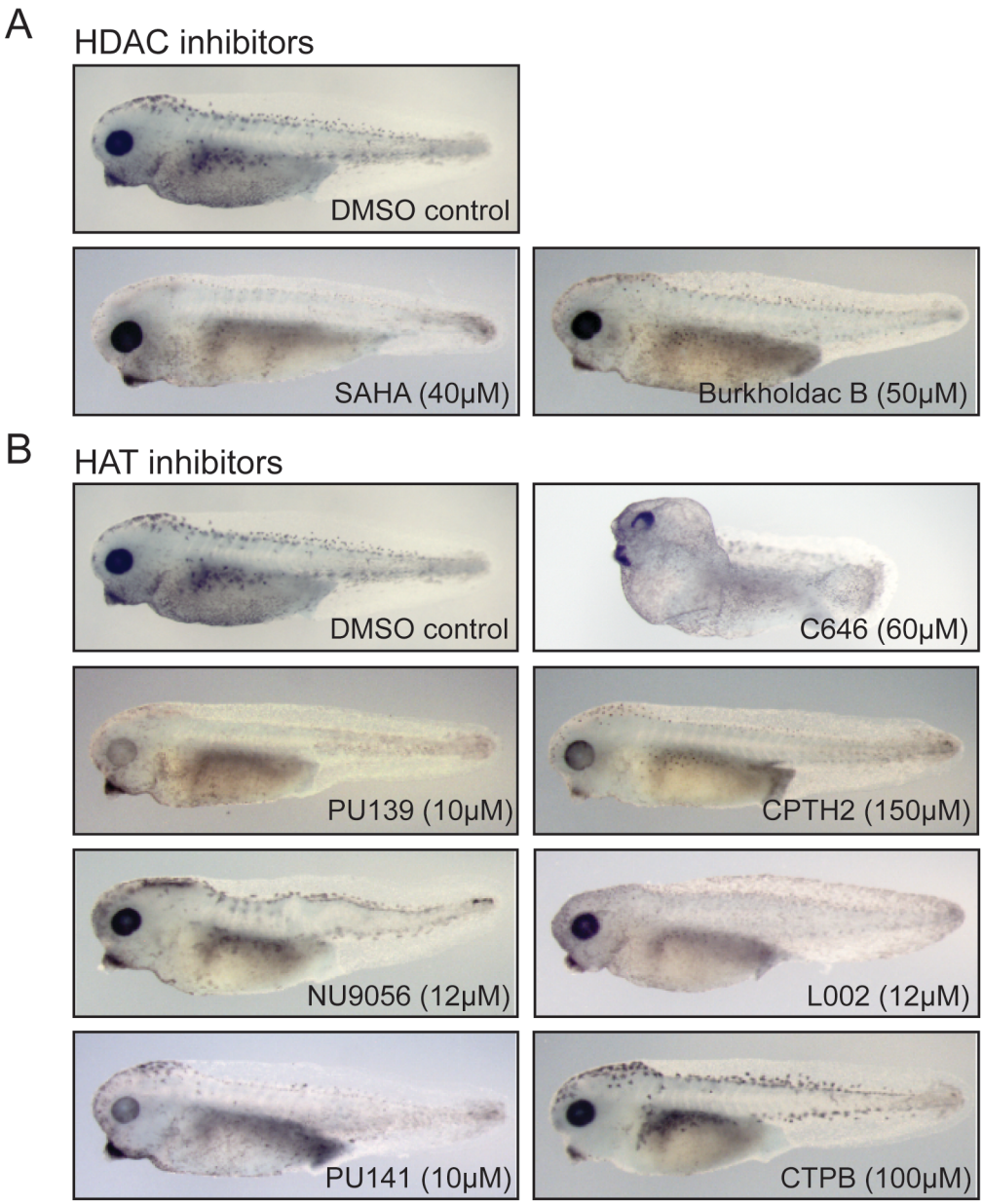


Figure 17: Phenotypes observed after HDAC and HAT inhibitors screen. (A) HDAC inhibitors. (B) HAT inhibitors.

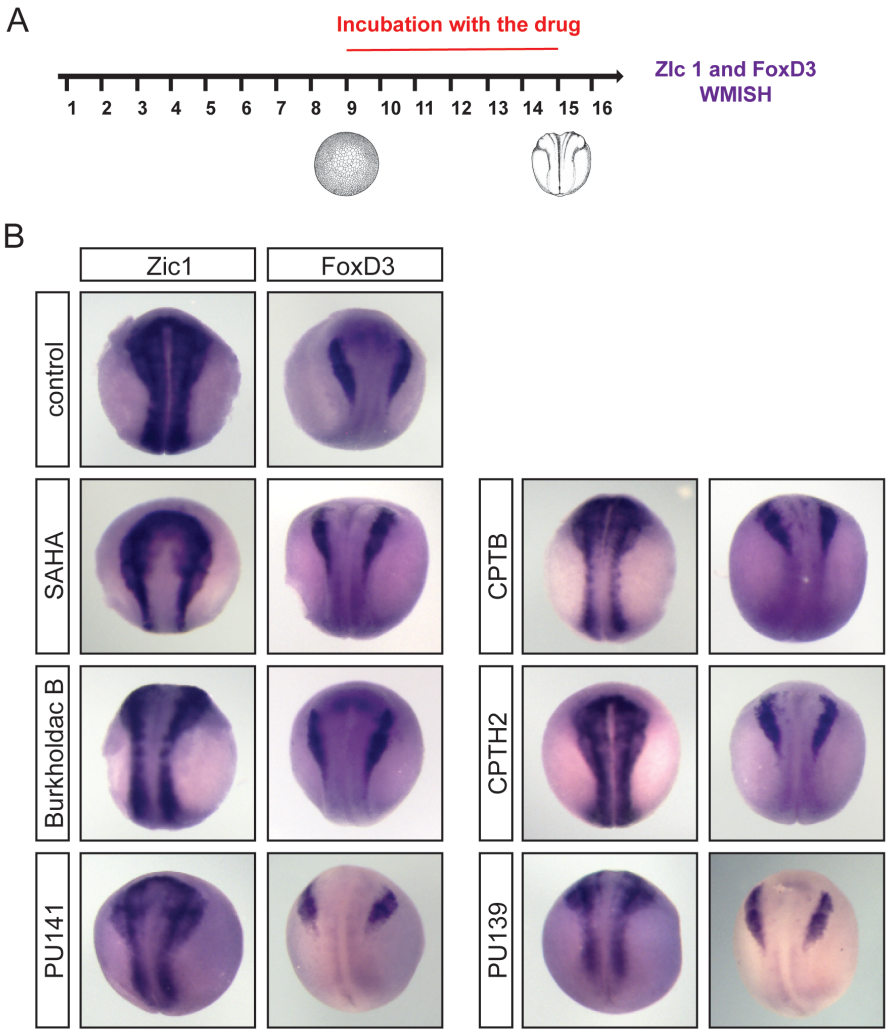


Figure 18: Screening of a selection of HAT and HDAC inhibitors followed by WISH. (A) Strategy followed, embryos were treated with drugs from stage 9 to stage 15 and then performed a WISH using *foxd3* and *zic1* markers. (B) Expression of *zic1* and *foxd3* after treatment with a selection of HDAC and HAT inhibitors. 7 embryos were used per condition.

3.2.3. Discussion

After the treatment with the library of 66 epigenetic inhibitors, 28 small molecule inhibitors did not cause any obvious phenotype in embryos; 18 were associated with developmental defects; 6 with pigment defects; 4 with edema formation; 3 with melanocytes morphology; 3 with craniofacial defects; 2 with melanocytes migration and lastly, 2 with eye morphology.

The categories overrepresented were mainly normal development and developmental defects. The abundance of first category has many explanations, which are also considerations to take into account when performing chemical

screens: (1) the drug, dissolved into the media, might not be taken up by the embryos; (2) the compounds, tested in human cells, might not recognize the *Xenopus* form of the target enzymes or might have more affinity for another enzyme; (3) there might be a compensation mechanism that does not allow the visualization of any obvious phenotype; (4) the drugs are added in a certain windows of time the target might not be having a crucial role in; and finally, (5) the concentration to inhibit certain enzymes might be higher than 150μM that is the maximum concentration above which the phenotypic effects can be due to the solvent. It was experimentally determined that toxic effects were observed when 1.5% of DMSO was added to the media. In order to prevent the phenotypic effects of the solvent, the amount of solvent was always smaller than 1.5% therefore the maximum concentration of drug screened was 150μM.

The second overrepresented category was defects in normal development, which reflect the importance and broad role that epigenetic modifiers have in development. It has been previously mentioned the importance of p300, as one of the co-factors that correlates with active enhancers, in general development. As expected, the inhibition of p300 through the compound C646, showed major defects in development at low concentration.

The third represented category is pigment reduction or absence. This is because scoring for pigment changes is relatively easy in this type of assays. As the pigment cells derive form the NC cells and our study focuses on NC formation, a further analysis was carried out with these compounds, which were all associated with changes in acetylation. The first thing to notice is the range of optimal concentrations varies from 10μM to 150μM. The amount of compound needed to inhibit the target depends on the amount of compound that is effectively absorbed by the embryo, the amount target enzyme present in the organism and the different affinity of the compound for its target.

The *zic1* and *foxd3* WISH after compound treatment did not show any affects on NC specification. It has to be mentioned that the initial screening was performed from stage 15 to stage 38-40, where the specification of NC cells has already occurred. Therefore it does make sense that the pigment reduction is not due to defects in early NC formation, but rather to the differentiation of melanocytes or melanogenesis. To determine if melanocytes are properly formed, it would be required to perform *mitf* or *tyrosinase* WISH, which are markers for pigment cell projenitors.

The general inhibitor of HDAC (SAHA) produces a similar phenotype to the HDAC class I inhibitor (BurkholdacB), suggesting than the phenotype produced by SAHA is caused by inhibition of class I HDACs (HDAC 1-3, 8). The lack of pigmentation in these organisms is consistent with the results published by Ignatius et al. (2008), where the function of *hdac1* in zebrafish melanocytes is characterized. Ignatius and collegues (2008) showed that the inhibition of *hdac1* produces a reduction on the number of melanocytes. Hdac1 prevents Foxd3 from blocking *mitf* expression therefore allowing melanogenesis; when this is disrupted by HDAC inhibition, Foxd3 represses pigment formation.

Regarding HATs, the results are not as easy to interpret. First, both C646 and L002 claim to inhibit specifically p300 although the phenotypes were quite different: whereas C646 affected general development, L002 affected mainly pigment cells. Second, there are members of both GNAT family (KAT2B) and p300/CBP family (P300), which show a loss-of-pigmentation phenotype. Although it is clear that histone acetyltransferases are relevant for melanocyte formation or melanogenesis, it remains unclear which specific subfamily is responsible for these effects. Another possibility is that there is no specificity and the inhibition of HATs does not allow the transcription of certain genes related with this cell type. As mentioned before, acetylation of histones is associated with active transcription due to repulsion between negatively charged DNA and acetylated histone residues which are also negatively charged allowing a relaxed nucleosome conformation making the DNA accessible to transcription related proteins.

There are a few studies reporting the importance of HATs for NC development. Two studies show that Kat6a (a member of the MYST subfamily) is relevant for proper craniofacial formation (Miller et al. 2004; Kong et al. 2014) . The MYST member inhibited in our screening, Tip60, did not show any major craniofacial defect, suggesting that Kat6a is specifically involved in head formation, at least in zebrafish.

The strategy followed did not produce the desired results, which were to identify small molecules affecting NC specification. No specific early NC affects were detected. This could be in part due to the time of incubation with the drug: the drugs should have been added at an earlier stage, such as stage 8-9, the embryos grown until stage 38-40 to check for defects in pigment cells or craniofacial structures. However, compounds were affecting melanocyte formation and/or melanogenesis, neural crest migration and craniofacial structure formation were

identified. This could be followed up by further characterizing the phenotypes with techniques such as Alcian blue staining, which stains cartilage and can reveal craniofacial malformations or WISH with melanocytes markers to investigate the pigment phenotypes. It would also be interesting to deeply characterize the phenotypes caused by an alteration on the histone acetylation in the context of gene expression. For that, ChIP-sequencing targeting RNAPII or specific histone acetylation modifications such as H3K27ac would be required.

We were not able to detect any specificity in epigenetic modifiers driving NC formation using this selection of small molecules. We then decided to determine if there was any specificity in CREs and motifs presents in those regulatory regions associated with NC formation.

3.3. Transcriptome analysis of *X. laevis* neural crest cells

3.3.1. NC-induced animal cap samples

In order to dissect the regulatory landscape of NC cells it was first necessary to analyse the transcriptome of this cell population. RNA-sequencing of different stages of NC formation, stage 9, stage 13 and stage 18 has been performed. The comparison between stage 9 and 13 would allow us to identify transcriptional changes required for NC specification whereas the comparison between stage 13 and 18 would reveal transcriptional changes involved in NC differentiation.

The neural crest cell population represents a very small percentage of the total number of cells present in the embryo; therefore to undergo genomic analysis specific for this cell population it is crucial to obtain an enriched population of NC cells. For that, the animal cap assay is a particularly good approach in *Xenopus*. Animal cap assay is a useful technique in *Xenopus* as it can be induced to become several tissue types, depending on the inducing signals provided (Smith 1995). To perform an animal cap assay, embryos are injected at 2-cell stage with the inducing signals, at stage 9 the animal pole of the embryo is dissected and allowed to develop until sibling non-dissected embryos reach the appropriate developmental stage; in our case stage 9, 13 and 18. Figure 14 shows a diagram of the animal cap assay. To obtain a NC population, embryos need to be injected with mRNA encoding a Wnt ligand, *Wnt1,* and a BMP antagonist; in our case *noggin* was used as a BMP antagonist (Garcia-Morales et al. 2009). The concentration of *Wnt1* and

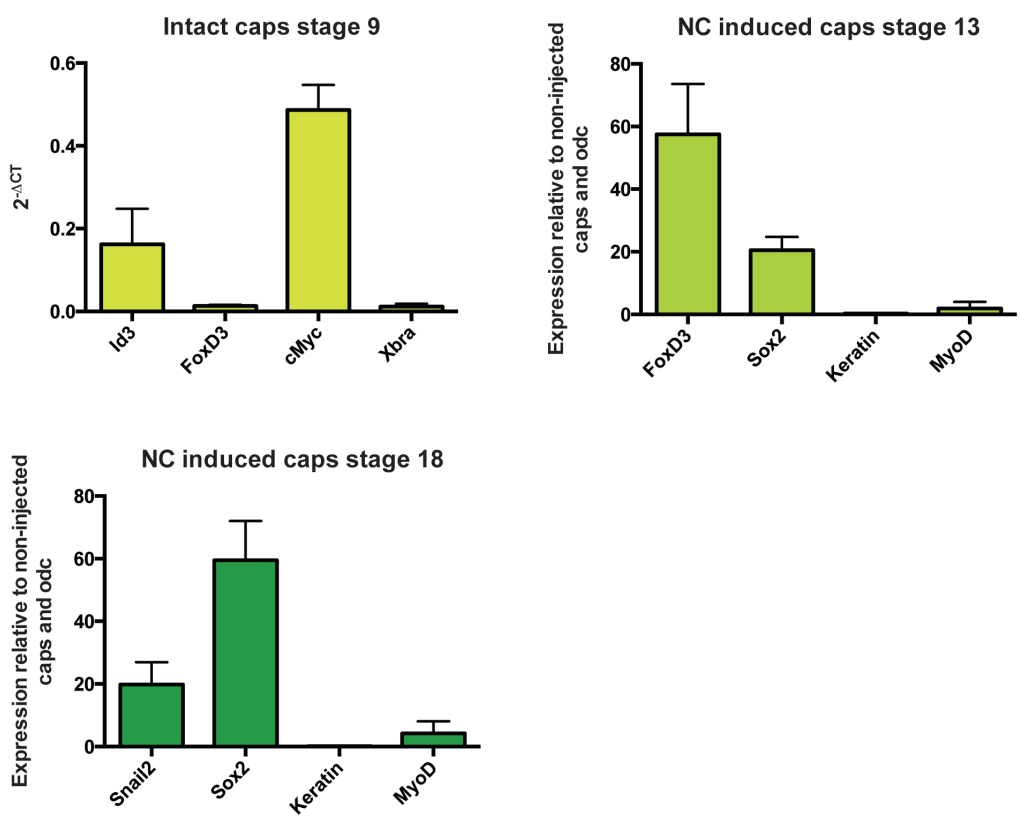
noggin sufficient to induce neural crest cell fate was found to be 100pg of *Wnt* and 500pg of *noggin*. In addition, embryos were also injected with 200pg of *gfp* to check the caps were properly injected.

The samples were analysed to ensure the expression of NC (*foxd3*, *snai2*, *id3*, *cmyc*) as well as neural (*sox2*), ectodermal (*keratin*) and mesodermal (*myoD*, *xbra*) markers was appropriate. For that, at the appropriate stage, RNA extraction followed by cDNA synthesis was performed in animal caps. In all the cases, the expression of *myoD*, *Xbra*, was very low, suggesting that there was no contamination with the underlying mesodermal tissue during the animal cap dissection (Figure 19A). This was also confirmed as mesodermal-contaminated caps display an elongated shape in comparison with rounded non-contaminated caps. The expression of *keratin* was also almost null indicating that the embryos were properly injected to induce NC. *sox2* is a early neural marker expressed in neural plate and later necessary for peripheral nervous system formation, which partly derive from NC cells. Moreover, the injection of *noggin* on its own is sufficient enough to induce neuroectoderm. This explains the presence of *sox2* in the NC samples. The expression of NC markers clearly confirms that the samples are enriched in NC cells (Figure 19A).

The RNA samples were sent for sequencing to the Earlham Institute. The requirements were 4 µg of material and 260/280 and 260/230 rations over 2 indicating no protein or phenol contamination of the samples (Figure 19B). RNA from 40-50 caps was extracted and 1ug of that was used to synthetize cDNA and test the samples for the correct gene expression by qPCR (Figure 19A). RNA samples were sequenced in a HiSeq2500 platform with a 125bp paired-end metric.

The analysis of the RNA samples was performed using Kallisto, for mapping the reads. About 80% of the reads were aligned against the XL_9.1_v1.8.3.2 available transcriptome (Figure 20A). EdgeR was used to perform differential expression analysis. In order to check the similarities between replicates and conditions, principle component analysis (PCA) was carried out (Figure 20B). We can observe that the three replicates cluster together. Also, that there are more differences between stage 9 and stage 13 rather than stage 13 and 18. This will be later reflected in the number of genes differentially expressed between those samples. There is one sample at stage 18, marked with *, that is separated form the other replicates in the PCA plot. This sample has been discarded for the further analysis.

A



B

Sample	ng/ul	volume	260/280	260/230
st9 cap rep.1	286.7	14µl	2.14	2.15
st9 cap rep.2	515.2	8µl	2.11	2.24
st9 cap rep.3	437.6	10µl	2.16	2.18
st13 NC cap rep.1	192.4	20µl	2.11	1.94
st13 NC cap rep.2	384.3	11µl	2.15	2.09
st13 NC cap rep.3	427.8	10µl	2.16	2.16
st18 NC cap rep.1	388.2	11µl	2.14	2.02
st18 NC cap rep.2	362.2	11µl	2.15	2.11
st18 NC cap rep.3	363.9	8µl	2.15	2.22

Figure 19: Animal cap samples used for RNA-sequencing experiment. (A) qPCR results showing the expression of NC markers (*foxd3*, *id3*, *cmyc*, *snai2*), neuroectoderm marker (*sox2*), ectodermal marker (*keratin*) and mesodermal markers (*xbra*, *myoD*) in animal caps at stage 9 and induced to become neural crest at stages 13 and 18. n=3 represent mean+SD. Samples were normalized against odc in non-induced caps. (B) Characteristics of the samples sent for sequencing.

A

Sample	aligned reads
st9 cap rep.1	76.83%
st9 cap rep.2	80.97%
st9 cap rep.3	80.32%
st13 NC cap rep.1	77.33%
st13 NC cap rep.2	77.54%
st13 NC cap rep.3	78.19%
st18 NC cap rep.1	77.70%
st18 NC cap rep.2	79.33%
st18 NC cap rep.3	81.52%

B

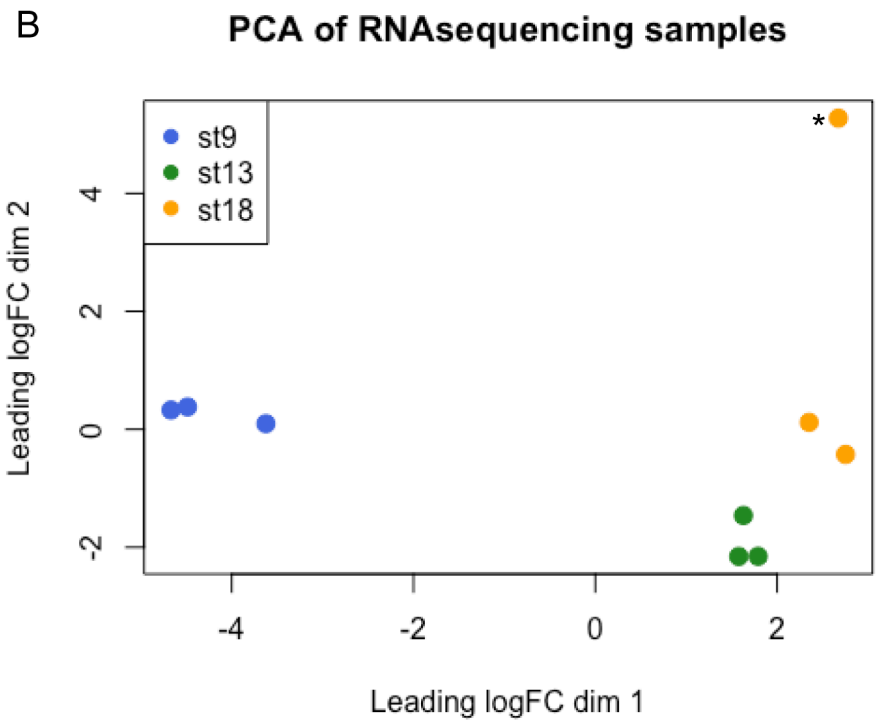


Figure 20: Analysis of RNA-sequencing samples. (A) Percentage of aligned reads using kallisto (B) PCA plot of the RNA samples. * indicates sample that was discarded for further analysis.

3.3.2. Comparison between stage 9 and stage 13 neural crest samples

Comparison between stage 9 and stage 13 was first carried out to identify required changes to specify NC cells. With a p-value <0.05, 7595 genes do not change between the two stages, 7597 genes were upregulated at stage 13 and 7171 genes were upregulated at stage 9 (Figure 21). To follow the analysis, genes were ranked according to the p-value and the top 100 genes were selected and separated in two lists, those upregulated at stage 9 and those upregulated at stage13 (Hierarchical cluster shown in Figure 22A). From the total 100 top genes, 38 were upregulated at stage 9 and after removing those without homology to any described gene and S and L duplications, 24 genes were selected. On the other side, 62 were upregulated at stage 13 and after discarding the duplicated homologues and the genes for which no homology was found to any annotated gene, 41 genes were selected. The total 65 genes were split into categories according their function and/or where they are expressed (Figure 22B, C). It was not possible to carry out GO term analysis because the *Xenopus* genes are poorly annotated and most of the genes appear unclassified; it was decided to manually annotate and classify the genes into categories according to their function and/or where they are express. These categories are broken down into a list of individual genes comprising them in Table 5 and Table 6.

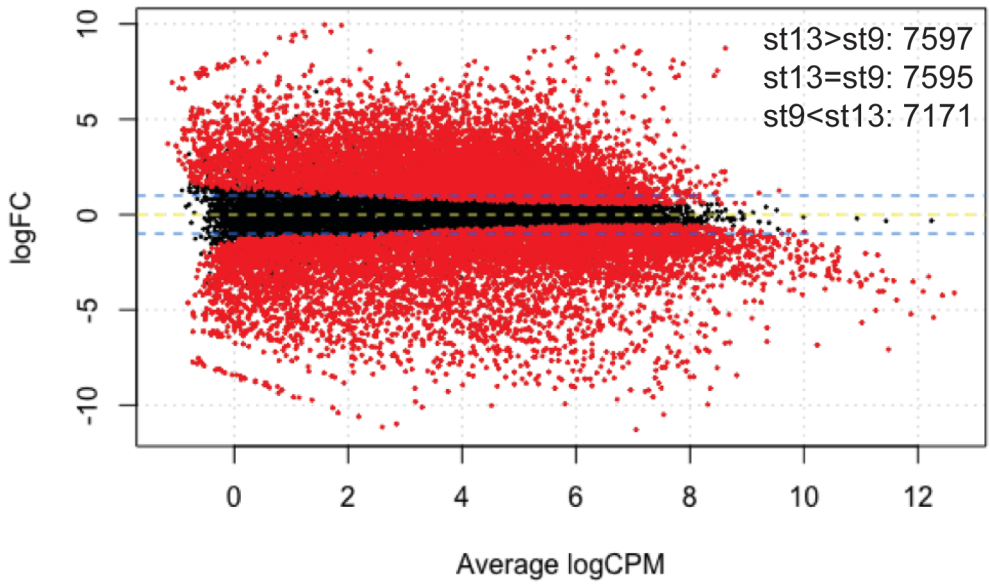


Figure 21: Gene comparison between NC stage 9 and stage 13 samples. Red dots mark differentially expressed genes. Black dots represent equally expressed genes and yellow line indicates no change blue lines indicate 10 fold change.

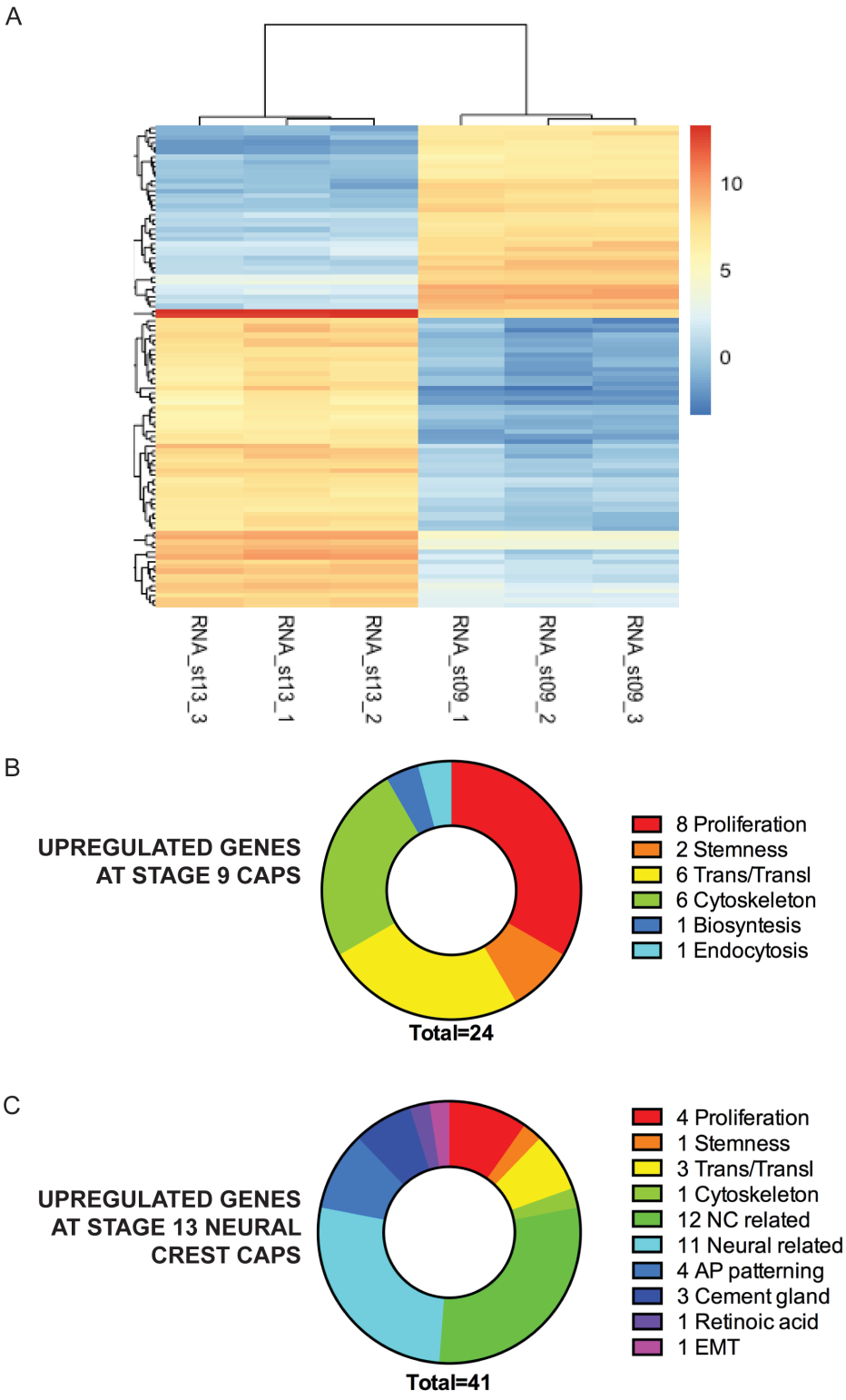


Figure 22: Genes differentially expressed between stage 9 and stage 13 NC caps. (A) Heatmap of the top 100 genes most differentially expressed between caps at stage 9 and NC caps stage 13. Hierarchical clustering performed using the complete lineage method. (B) Pie chart showing the categories of genes upregulated at stage 9 caps. (B) Pie chart showing the categories of genes upregulated at stage 13 neural crest caps. The number of genes in each category is annotated in the legend.

The genes upregulated in the stage 9 cap (Figure 22B, Table 5) samples are related mainly with: (1) proliferation; (2) cytoskeleton; (3) transcription and translation regulation and (4) stemness. The main category of upregulated genes is proliferation that together with stemness is consistent with the multipotency identity of the animal pole of the blastula embryo. At stage 8-9, the mid-blastula transition occurs, where zygotic gene transcription is activated. This could explain the upregulation of genes involved in transcription and translation regulation such as *h2ab*. Finally, just after stage 9, gastrulation starts, from stage 10 to stage 14, in which a complex process of cell rearrangements takes place. This can be the reason why a large set of cytoskeleton related genes such as *rapgef3*, involved in actin rearrangements, and *arhgap*, involved in regulating actin dynamics, are highly expressed at stage 9 compared with stage 13 when the gastrulation has already occurred.

The genes upregulated at stage 13 neural crest caps (Figure 22C, Table 6) are mainly involved in: (1) neural crest formation; (2) neural formation; (3) anteroposterior patterning; (4) proliferation; (5) cement gland and (6) transcription and translation regulation. As expected, neural plate borders associated genes such as *pax3*, *meis3* and *gbx2* are upregulated when comparing stage 9 versus stage 13 neural crest caps. There are also some brain markers upregulated such as *pinhead*, *dmbx1* or *irx3*. This might be due to the fact that *Wnt3* and *noggin* are also used to induce hindbrain (Dibner et al. 2001) and that *noggin* alone also induce neuroectodermal tissue. There are also genes involved in AP patterning of the neural plate and cement gland, which marks the most anterior part of the embryo such as *gli3*, *vill* and *ag1*.

Table 5: Top genes differentially expressed in stage 9 caps when compared with stage 13 NC-induced caps. Transcription factors relevant for ATAC-sequencing are highlighted in bold.

Categories	Gene name	Description	p-value
Proliferation	*cyclin A1*	Cell cycle regulator	1.36E-121
	cyclin B1	Cell cycle regulator	7.47E-84
	cyclin B5	Cell cycle regulator	1.42E-83
	cell division cycle associated 7 like	Positive regulation of cell proliferation	2.90E-74
	f-box protein 43	E3 ubiquitin ligase, mediates degradation of cyclins	9.18E-70
	NDC80 kinetochore complex component	Centromere-associated protein	4.50E-69
	BTG family member 3	Anti-proliferation factor	9.08E-69
	proline rich 14-like	Tethers heterochromatin to the nuclear laminar	1.90E-61
Stemness	***pou5f3.3***	Oct60 homolog	1.93E-108
	dand5	Involved in pluripotency maintenance	2.13E-90
Transcription and Translation Regulation	*cytoplasmic polyadenylation element binding protein 1*	Regulates cytoplasmic polyadenylation	2.21E-103
	foxi2	Transcription activator	1.95E-96
	h2ab	Variant histone of conventional H2A	3.06E-80
	zinc finger protein 36, C3H type-like 2, gene 1	Promotes poly(A) tail removal	7.12E-78
	h1foo	H1 histone variant oocyte specific	3.64E-66
	protein associated with topoisomerase II homolog	Translational repressor	9.69E-62
Cytoskeleton	*rap guanine nucleotide exchange factor 3*	Required for the actin rearrangement	6.14E-74
	rho GTPase activating protein	Regulates intracellular actin dynamics	8.47E-65
	microtubule associated protein 1 light chain 3 gamma	Involved in autophagy and membrane trafficking	8.52E-65
	exportin 6	Mediates the nuclear export of actin	1.29E-64
	FH2 domain containing 1	Contributes to the coordination of microtubules with actin fibers	3.45E-64
	Wiskott-Aldrich syndrome like	Regulates actin polymerization	4.68E-61
Biosynthesis	*UDP-N-acetylglucosamine pyrophosphorylase 1*	Nucleotide-sugar biosynthesis	7.23E-65
Endocytosis	*sH3-domain kinase binding protein 1*	Involved in endocytosis	1.10E-64

Table 6: Top genes differentially expressed in stage 13 neural crest caps when compared with stage 9 caps. Transcription factors relevant for ATAC-sequencing analysis are highlighted in bold.

Category	Gene name	Description	p-value
Neural crest related	*laminin subunit beta 1*	Involved in migration	1.30E-108
	pax3	NC specifier	2.35E-100
	zeb2b	Found in tissues differentiated from the neural crest	7.66E-100
	mmp14	Involved in migration	2.86E-80
	meis3	NC specifier	3.73E-80
	gbx2	NC specifier	1.35E-79
	axin2	Upregulated in neural crest precursors	1.92E-72
	complement component 3	Involved in migration	1.84E-71
	sp5	Early regulator of neural crest	2.41E-64
	osterix	Required for cranial neural crest-derived	2.20E-65
	mn1	Required for palate development	2.97E-88
Proliferation	*cyclin B3*	Cell cycle regulator	3.49E-75
	thymine DNA glycosylase	Base excision repair enzyme	2.40E-62
	retinoblastoma-like 1	Tumor suppressor	6.17E-88
	topoisomerase II beta	Manages DNA tangles and supercoils	6.88E-88
Transcription and Translation regulation	*heterogeneous nuclear ribonucleoprotein R*	Processes mRNA precursors	2.42E-79
	eukaryotic translation initiation factor 4 gamma, 3	Aids the assembly of the ribosome onto the mRNA	8.03E-65
	eukaryotic translation elongation factor 1-alpha	Facilitates translational elongation	5.79E-64
Neural related	*pinhead*	When inhibited causes microcephaly	5.91E-99
	diencephalon/ mesencephalon homeobox 1	Involved in brain and sensory organ development.	5.81E-86
	olfactomedin 4	Involved in nerve tissue formation	1.39E-80
	spalt-like transcription factor 2	Plays a role in neurogenesis, neuronal differentiation and eye development	7.24E-73
	iroquois homeobox 3	Involved in neural patterning	1.71E-72
	eukaryotic translation initiation factor 4A2	Involved in neural patterning	1.81E-66

	zygotic DNA replication licensing factor mcm3	Associated with stem cell pluripotency retinal cells	1.64E-64
	tubulin beta class I	Involved in brain development	2.13E-64
	retinal pigment epithelium-specific protein 65kDa		2.91E-62
	cadherin EGF LAG seven-pass G-type receptor 1	Involved in neural tissue	4.98E-60
Anteroposterior patterning	*nuclear receptor subfamily 6 group A member 1*	Required for anteroposterior patterning	1.66E-87
	forkhead box B1	Early anteroposterior patterning of the neuroectoderm	2.61E-63
	transcriptional enhancer factor TEF-1-like	Localized anterior region of the neural plate	6.74E-62
	gli3	Involved in neuroectoderm patterning	1.07E-59
	rax	Involved in anterior neural plate	4.03E-62
Cement gland	*anterior gradient 1*	Expressed in cement gland	5.96E-70
	arfGAP with dual PH domains 1	Expressed in cement gland	1.12E-68
	vill	Expressed in cement gland	1.84E-68
Stemness	*msi1*	Neural stem cell regulatory gene	7.83E-63
Retinoic acid	*retinoic acid receptor gamma*		4.21E-80
Cytoskeleton	*epiplakin 1*	Involved in cytoskeletal architecture organization	1.80E-79
EMT	*GLI pathogenesis-related 2*	Positive regulation of epithelial to mesenchymal transition	2.94E-66

3.3.3. Comparison between stage 13 and stage 18 neural crest samples

The differential expression of stage 13 compared with stage 18 neural crest animal cap was carried out. With a p-value <0.05, there are 16075 genes equally expressed, 2052 genes upregulated at stage 13 and 2928 genes upregulated at stage 18 (Figure 23). To follow the analysis, genes were ranked according to the p-value and the top 100 genes were selected and separated in two lists, those upregulated at stage 13 and those upregulated at stage18 (Hierarchical cluster shown in Figure 24A). Out of the 100 top genes differentially expressed, 24 were upregulated at stage 13. After removing the genes with no homology to any described gene and duplicated homologous genes, only 15 genes were selected. On the other side, 76 genes were upregulated in stage 18 neural crest caps. From those, 44 were selected after removing duplicated homologous genes and genes with no homology found to any described gene (Figure 24A). The selected genes were manually split into categories according to their function and/or expression pattern (Figure 24A, B).

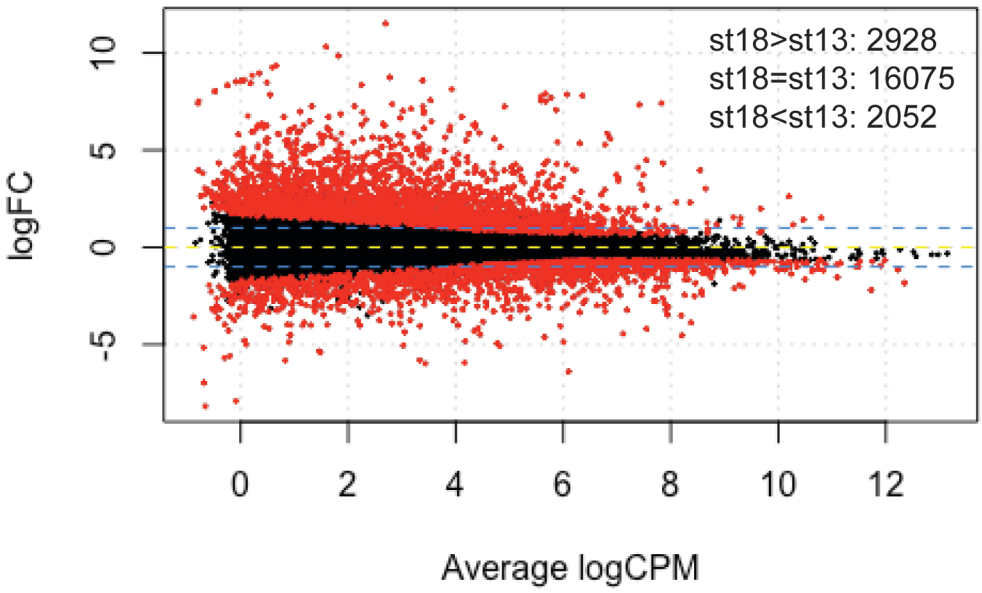


Figure 23: Gene comparison between NC stage 13 and stage 18 samples. Red dots mark differentially expressed genes; black dots represent equally expressed genes, yellow line indicates no change blue lines indicate 10 fold change.

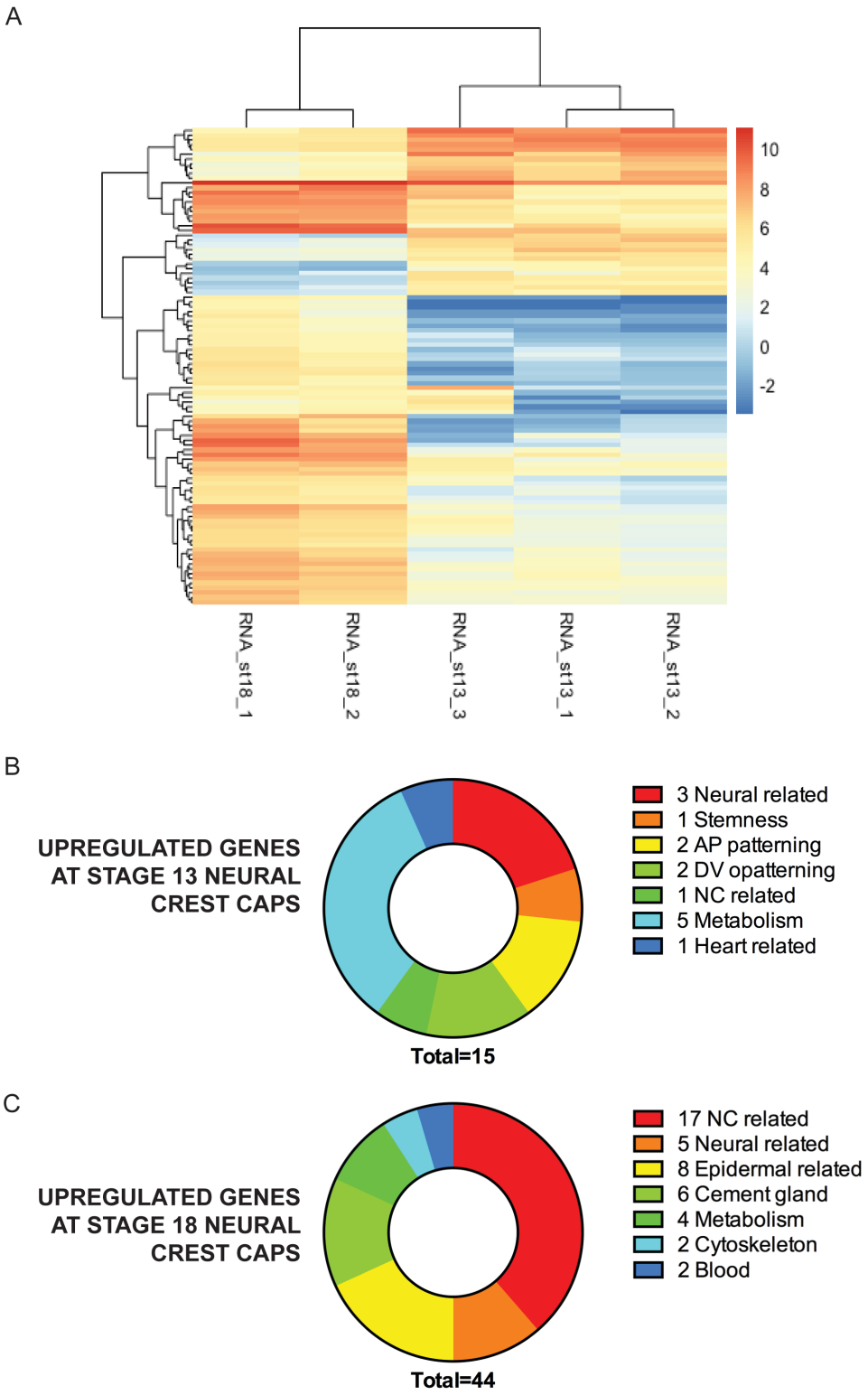


Figure 24: Genes differentially expressed between stage 13 and stage 18 NC caps. (A) Heatmap of the top 100 genes most differentially expressed between NC caps stage 13 and stage 18. Hierarchical clustering performed using the complete lineage method. (B) Pie chart showing the categories of genes upregulated at stage 13 neural crest caps. (B) Pie chart showing the categories of genes upregulated at stage 18 neural crest caps. The number of genes in each category is annotated in the legend.

The genes upregulated at stage 13, when compared to stage 18 (Figure 24B, Table 7), are mainly related to: (1) neural formation; (2) anteroposterior patterning; (3) dorsoventral patterning and (4) metabolism. At later stages the patterning of the different tissues has already been established which explains the presence of genes such as *foxb1* and *foxd4l1.1* involved in patterning establishment upregulated at stage 13. It might be surprising not to find genes involved with NC specification but it should be considered that there might be a change in expression, but not significant enough to appear in the top 100 genes or NC genes remain on in stage 13 and 18 animal caps. The only NC related gene is *ventx2*, more related with maintaining pluripotency in NC specified cells.

Alternatively, the genes upregulated at stage 18 (Figure 24C, Table 8), mainly correspond to the following categories: (1) neural crest related; (2) neural related; (3) epidermal related; (4) cement gland and (5) metabolism (Figure 23C, Table 7). Not surprisingly, several NC markers such as *sox9*, *sox8*, *foxd3* and *snai2* are found significantly upregulated at stage18 when compared with stage 13. There are also some neural genes such as *en2* and epidermal genes such as *krt*. This can reflect the heterogeneity of cells within the neural crest animal caps. There is also an upregulation in genes expressed in the cement gland such as *gale* or *xa-1*.

Table 7: Top genes differentially expressed at stage 13 neural crest animal caps when compared with stage 18 caps. Transcription factors relevant for ATAC-sequencing analysis are highlighted in bold.

Category	Gene name	Description	p-value
Stemness	***pou5f3.2***	Oct25 homolog	1.10E-34
Neural related	*tdgf1.3*	Downregulation causes forebrain defects.	9.73E-45
	pinhead	Downregulation cause microcephaly	6.55E-37
	geminin	Induces neural progenitor formation	7.27E-29
Neural crest related	***ventx2***	Maintains NC multipotency	2.71E-23
Anteroposterior patterning	*sp5l*	Involved in neuroectoderm patterning	4.48E-35
	forkhead box B1	Early anteroposterior patterning of the neuroectoderm	4.25E-21
Dorsoventral patterning	*forkhead box D4 like 1, gene 1*	Transcriptional repressor with a role in dorsal axis formation	1.01E-23
	admp	Anti-dorsalizing morphogenic protein	5.86E-21
Metabolism	*histidine ammonia-lyase*	Converts histidine into ammonia	7.54E-42
	htxA	Oxidoreductase enzymes	7.77E-40
	monocarboxylate transporter 4-like	Carries lactate and pyruvate, across biological membranes.	1.08E-27
	slc16a3	Carries lactate and pyruvate, across biological membranes	4.10E-27
	insulin	Regulates the metabolism of carbohydrates, fats and protein	8.57E-23
Heart related	*LIM domain 7 protein*	Involved in heart development	1.28E-24

.

Table 8: Top genes differentially expressed at stage 18 neural crest animal caps when compared with stage 13 caps. Transcription factors relevant for ATAC-sequencing analysis are highlighted in bold.

Category	Gene name	Description	p-value
Neural crest related	*sox8*	Induces neural crest progenitors	9.94E-57
	tp63	Expressed in branchial arches	2.22E-36
	sox9	Induces neural crest progenitors	6.50E-35
	snai2	NC marker	4.01E-33
	bmp1	Expressed in premigratory NC	6.74E-33
	tbx3	Expressed in NCC	1.11E-28
	pdgfra	Involved in NCC migration	1.72E-28
	plakophilin 3	Involved in NCC migration	1.67E-25
	tfap2b	NC specifier	8.64E-25
	kremen1	Required for neural crest induction	2.64E-24
	foxd3	NC specifier	5.26E-24
	ezr	Expressed in neural plate cell	1.15E-23
	capn8	Involved in NCC migration	4.05E-22
	castor zinc finger 1	Expressed in neuroblastoma cell lines	9.22E-22
	pcdh8	Involved in cephalic NC migration	1.16E-21
	protogenin	Prevents premature apoptosis of cephalic neural crest cells	8.45E-21
	gata2	Sympathetic neuron development	2.10E-20
Neural related	*engrailed homeobox 2*	Involved in midbrain formation	2.95E-38
	huntingtin-associated protein 1	Involved in vesicular trafficking	1.81E-22
	cmahp	Involved in brain growth	3.44E-22
	diacylglycerol kinase, eta	Associated with bipolar disorder	4.46E-20
	camk2g	Involved in neuronal plasticity	1.79E-28
	nr2f5	Expressed in hindbrain	1.73E-30
	fam171B	Expressed in brain	5.94E-20
Epidermal related	*keratin, type I cytoskeletal 47 kDa*	Epidermal fibrous structural proteins	1.67E-47
	x-epilectin	Posterior epidermal marker	7.43E-36

	sciellin	Epidermis terminal differentiation marker	1.42E-23
	ahnak	Localized in the cell surface of human epidermis	4.81E-23
	periplakin	Localized in the epidermal cornfield envelope	1.48E-22
	krt5.7	Epidermal fibrous structural proteins	1.60E-22
	krt	Epidermal fibrous structural proteins	6.11E-22
	otogelin	Expressed in goblet cells	8.30E-60
Cement gland	*nkx3-1*	Expressed in cement gland	3.52E-31
	xa-1	Expressed in cement gland	1.91E-30
	fam3D	Expressed in cement gland	7.43E-25
	hatching enzyme	Expressed in cement gland	3.67E-24
	gale	Expressed in cement gland	3.65E-20
	anterior gradient 2	Involved in cement gland differentiation	2.77E-20
Metabolism	*phosphodiesterase 4C*	Hydrolyzes, cAMP	7.77E-22
	galnt6.2	Involved in glycosylation	2.58E-24
	a-kinase anchoring protein 9	Interact with signal transduction pathways	3.31E-22
	s100 calcium binding protein A1	Involved in several cellular processes	8.02E-29
Cytoskeleton	*filamin a*	Stress fiber component	1.19E-22
	calpain 2, gene2	Modulates convergent extension movements	6.80E-20
Blood	*etv6*	Regulates embryonic hematopoiesis	5.19E-22
	gata2	Involved in globulin production	2.10E-20
Bones	*transmembrane protein 119*	Promotes the differentiation of myoblasts into osteoblasts	1.72E-20
	acid phosphatase 5, tartrate resistant	Involved in osteoclast migration	4.25E-20
Dorsoventral patterning	*sizzled*	When overexpressed can dorsalize embryos	3.29E-20

3.3.4. Time of expression of neural crest genes

The expression of neural crest genes across the different stages of neural crest development has been addressed by analysing the changes in expression, i.e., transcript per million (tpm), in the different stages of *Xenopus* NC formation. The aim of this is to determine which genes behave in the same way and try to correlate time-related expression patterns with the function of these genes.

The expression of neural plate border genes (Figure 25) i.e. *dlx5*, *dlx6*, *id3, meis3*, *msx1*, *pax3*, *tfap2a*, *zic1* and *zic 5* was analysed. For all of them, there is a common massive increase in expression at stage 13 compared with stage 9 and after that the expression is either maintained, increased or decreased. The most dramatic change in expression corresponds to *pax3* and *msx1.L* whose expression increase by 200 tpm. The change in expression of other genes such as *zic5*, *dlx5*, *dlx6* is not as drastic but all of them increase their expression at stage 13.

Following that, the expression of neural crest specifiers was also analysed (Figure 26). For genes such as *foxd3*, *cmyc*, *nmyc*, *msx2*, *pax7*, *snai2*, *sox5*, *sox8*, *sox9*, *sox10* and *twist* there is almost no expression or some expression at stage 9 and 13 but a massive increase of expression at stage 18. Genes such as *snai2* and *sox10* raise their expression from 0 to 360 and 97 tmp, respectively whereas genes like *sox5*, *pax7* and *nmyc* are barely expressed.

Another thing to notice is that for most of the genes, both *X. laevis* L and S homologues behave the same way. This might reflect that both homologous genes are functionally equivalent. However there are a few cases for which the expression varies. *snai2* and *sox9* have an homologue that shows very little expression across the different stages i.e. *snai2.S* and *sox9.L*.

Moreover, the pluripotency factors have been analysed as it has been suggested that specified neural crest cells at stage 13 share the expression of some pluripotency factors from blastula cells (Figure 27). In Buitrago-Delgado et al. (2015) they conclude that NC stage at 13 and stage 9 blastula both express pluripotency factors such as Pou5f1 xenopus homologs (*pou5f3.1*, *pou5f3.2* and *pou5f3.3*), Sox2 *Xenopus* homologs (*sox2* and *sox3*), Nanog homolog (*ventx1.2*). Both stages also express neural crest markers such as *id3*, *sox5*, *tfap2a*, *foxd3* and *snai1*. In order to confirm if our RNA-sequencing results support that hypothesis, the expression of these genes has been analysed. The least expressed genes in stage 9 and stage

13 are *sox5*, *ventx1.2* and *foxd3*. Most of the genes are expressed in all the stages, stage 9, 13 and 18. However, the level of expression of those genes at stage 13 is more similar to stage 18 than to stage 9. From the pluripotency factors, the Oct4 *Xenopus* forms show an interesting pattern: whereas *pou5f3.2* and *pou5f3.3* are highly expressed at stage 9 and their expression is reduced across the time, *pou5f3.1* has a peak of expression at stage 13 (Figure 27).

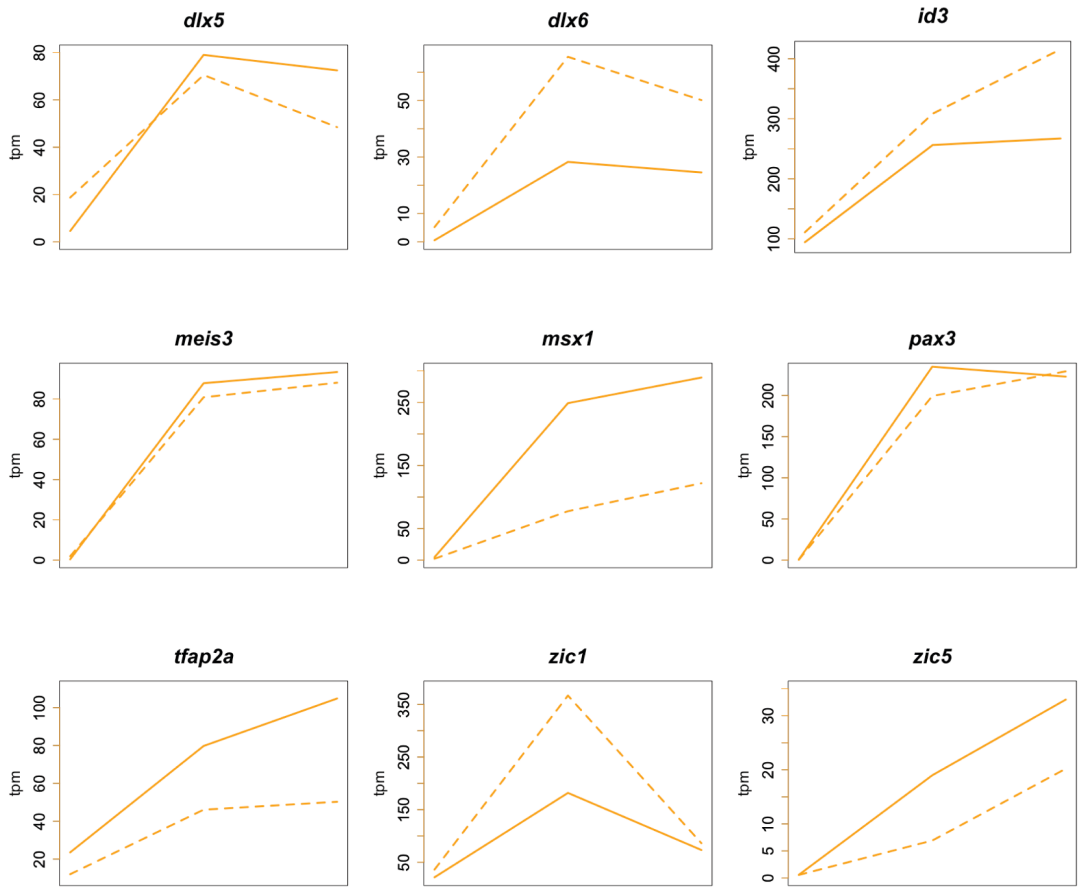


Figure 25: Expression profile of neural plate border genes. The mean of the transcripts per million (tpm) of each gene and homologues is plot across the three different stages: stage 9 (left), stage 13 (middle) and stage 18 (right). Solid line indicates L and dash line indicates S homologues.

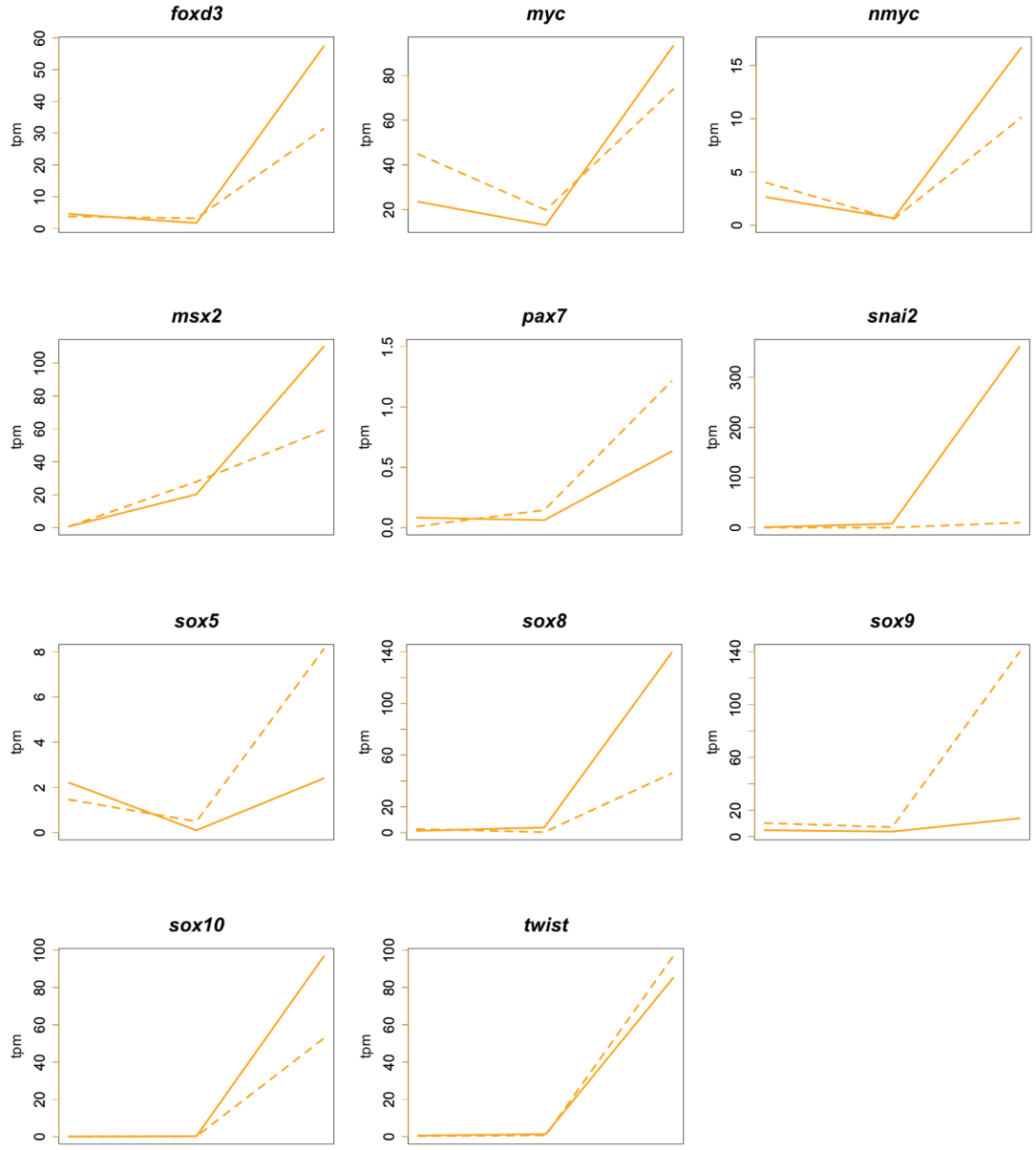


Figure 26: Expression profile of neural crest specifiers genes. The mean of the transcripts per million (tpm) of each gene and homologues is plot across the three different stages: stage 9 (left), stage 13 (middle) and stage 18 (right). Solid line indicates L and dash line indicates S homologues.

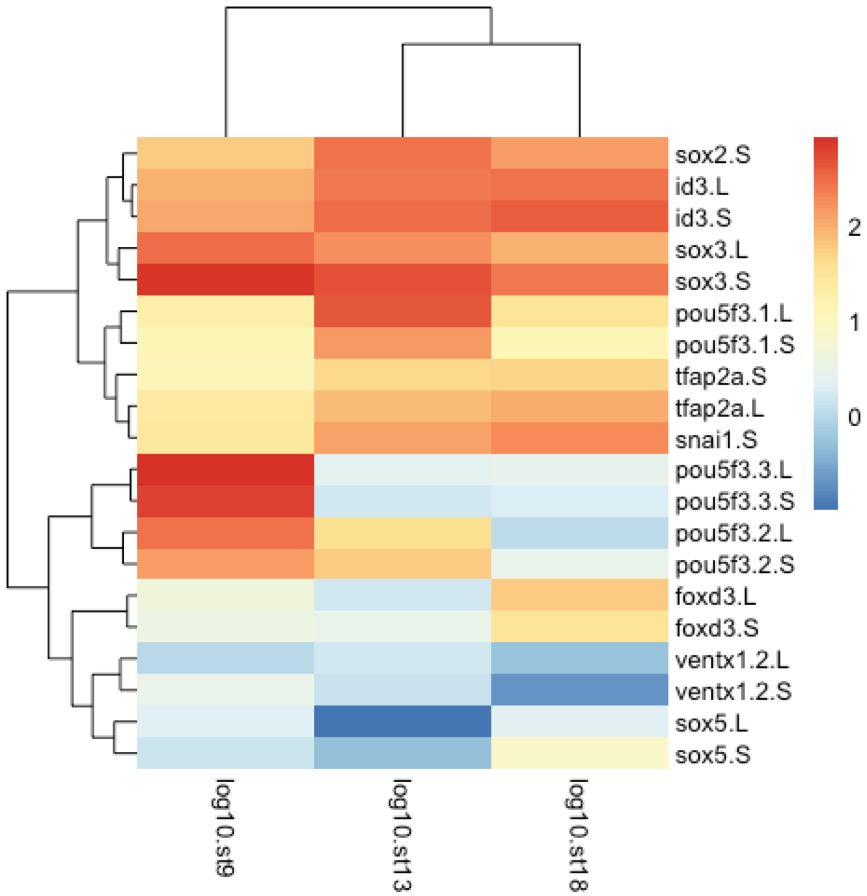


Figure 27: Heatmap showing the expression of pluripotency and NC factors. The log10 of the mean of transcripts per million has been plotted. Selection of genes has been made according to the genes described in Buitrago-Delgado et al. (2015).

3.3.5. Discussion

The aim of the RNA-sequencing was to validate a method to obtain a cell population enriched in neural crest cells, the neural crest-induced animal caps. This would allow us to study the NC population in detail without needing to use whole embryos, where the NC regulation would be masked by whole embryo regulation. The study of the transcriptional changes across the different stages of neural crest formation is going to be later integrated with the epigenomic data to have a better understanding of the NC regulome (see section 3.4).

The injection of 500pg of *noggin* and 100pg of *Wnt1* is sufficient to induce a neural crest population. The results demonstrate that the NC population obtained across the different stages behaves the same way as the neural crest cells in the embryo. The neural plate border markers have a peak of expression at stage 13 (Figure 25), when they regulate each other to define the neural plate border domain and together with the initial signaling pathways, Wnt, BMP and FGF, allow the expression of NC specifiers. NC specifiers, on the other hand, have a peak of expression at stage 18 (Figure 26), and these drive the expression of genes involved in EMT, delamination, migration and, differentiation into the several NC derivatives.

The analysis of the top genes most differentially expressed also shows that NPB genes such as *pax3*, *gbx2* and *meis3* are significantly upregulated when compared with stage 9 samples, where there is mainly an upregulation of genes such as cyclins and *pou5f3.3* involved in proliferation and stemness. When comparing stage 18 with stage 13, there is also significant increase in NC specifier genes such as *sox9*, *snai2* and *foxd3*. This confirms that the neural crest-induced animal cap is a valid method to obtain NC cells and so can be used to obtain information about the regulatory regions involved in neural crest formation.

However, there are also neural and epidermal markers found to be expressed in these samples. This can be explained, as animal cap is an organoid model to obtain neural crest population based on RNA injection, and there might be a heterogenic population where some cells are not fully committed to be NC or are committed down alternative routes. The injection of *noggin* alone is used to induce neural tissue whereas the non-injection controls form ectodermal tissue. For that, it would be interesting to perform single cell sequencing of the animal caps to determine the percentage of fully NC cells compared with the other cell types that

might be present on the animal cap and to have a better idea of how cells are being committed and specified to become neural crest cells.

The comparison between samples shown in the PCA plot (Figure 20) shows that there are more similarities between stage 13 and 18 rather that between stage 9 and stage 13. This could be due to the fact that stage 9 caps have not been injected with *noggin* and *Wnt1*. It would be of interest to compare not-injected caps and NC-induced caps at stage 9 to really answer the hypothesis raised by Buitrago-Delgado et al. (2015) that claims NC cells retain multipotent properties rather than differentiate into this cell type. Our data confirm that stage 9 and stage 13 neural crest share expression of pluripotency and neural crest genes although the level of expression is not comparable.

This transcriptome data can also be interpreted in the context of *Xenopus laevis* genome evolution, where whole genome duplication has led to an allotetraploid genome and so two copies of each gene, designated L and S. Although the developmental genes have mostly been retained, some changes in spatiotemporal pattern of expression suggesting subfunctionalization have been identified (Session et al. 2016). In the majority of NC genes, both homologous genes are expressed and show a similar patter of expression. However, *sox9* and *snai2* show differences of expression: only the *snai2.L* and the *sox9.S* homologues are expressed. This indicates a possible subfunctionaliztion of the two homologues, which could be further analysed by designing WISH probes and specific primers for each homologue. On the other hand, *Xenopus* has three copies of Oct4, which show a possible subfunctionalization: whereas *pou5f3.2* and *pou5f3.3* have a peak of expression at stage 9, *pou5f3.1* has its peak of expression at stage 13. This is especially interesting in this context as Session et al. (2016) suggest that this subfunctionalization is due to changes in the epigenetic regulation of the different homologous genes.

3.4. ATAC-sequencing to identify cis-regulatory elements

3.4.1. ATAC-sequencing protocol optimization

Section 3.3 has confirmed that the neural crest-induced animal cap is a good method to obtain an enriched population of neural crest cells. The main goal of this study was to identify regulatory regions specific for NC formation and, if possible, identify differences between CRE activated at stage 13 (specification) and stage 18 (differentiation). Because of the small amount of required material, the new method of ATAC-sequencing was chosen over the different techniques that identify enhancers such as ChIP-sequencing and DNase-sequencing (Shlyueva et al. 2014).

The first step was to optimize the protocol first published by Buenrostro and collegues (2015). The main modifications in the protocol are listed below. (1) To dissociate the tissue with a collagenase/trypsin solution and pass it through a cell strainer to obtain a single cell population. This way the cells are easier to lyse and therefore the enzyme can better reach the nuclei. (2) To increase the NP40 amount to 0.2% rather than 0.1% in the lysis buffer. NP40 is a non-denaturing detergent; therefore it increases cell permeability without disrupting protein complexes. The increase in NP40 facilitates the access to the nuclei. (3) To count the number of nuclei that will be incubated with the enzyme and add a maximum of 50,000 cells. On average, a cap stage 18 contains 100,000 cells whereas a cap stage 13 contains approximately 50,000 cells. It was decided to do ATAC-sequencing on single animal caps and not to pool several caps as that could reduce the background. (4) To reduce the volume of transposase reaction to 20µl instead of 50µl. (5) To inactivate the enzyme through two methods: firstly, incubating the reaction 30min at 55°C, temperature at which the enzyme is inactivated and secondly, adding EDTA which sequesters magnesium ions, that act as a cofactor for the Tn5 enzyme (Picelli et al. 2014; Buenrostro, et al. 2015). After the incubation at 55°C with EDTA, the same amount of $MgCl_2$ needs to be added as it is required for the following PCR reaction. That way the enzyme is inactivated after 30 minutes incubation at 37°C and overtagmentation is prevented. (6) To purify the libraries generated with AMPure Beads SRI 1X, to eliminate small fragments present in the library.

During the optimization experiments, several profiles of tagmentation were obtained (Figure 28). Not-tagmented samples are defined as having no clear peak

whereas overtagmented samples contain a peak around 200bp. In the later, taking into account that the primers used to generate the library are 50bp long, the amplified region is not informative. In properly tagmented samples a peak around 300bp is obtained and occasionally a nucleosome pattern can be seen, which is characterized by a periodicity of 200bp (Figure 28C). This peak corresponds to the size of a regulatory region.

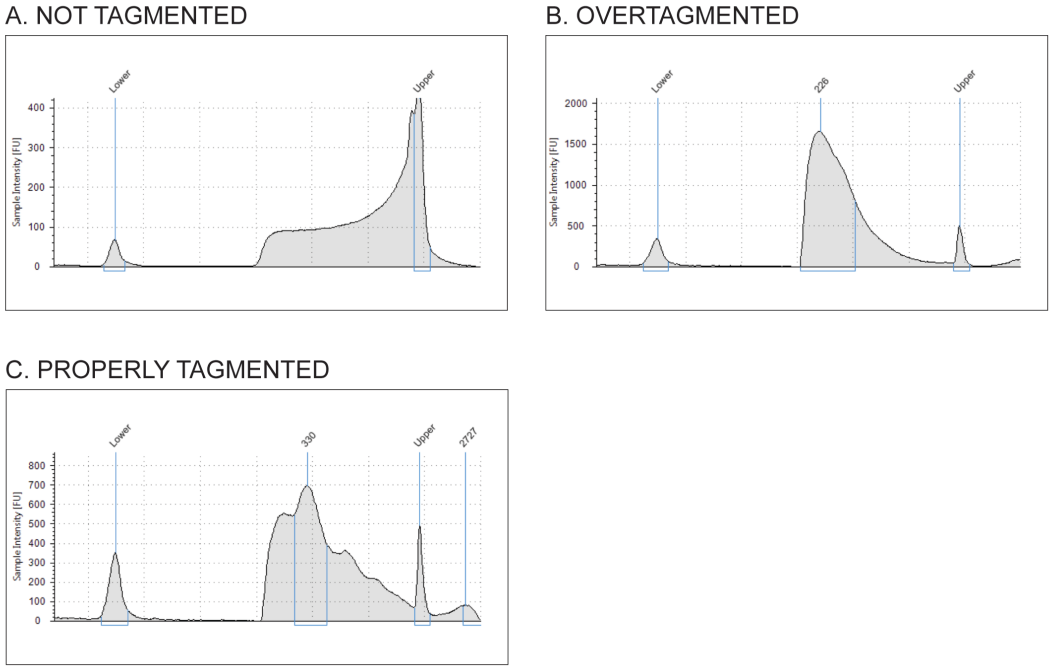


Figure 28: Examples of tagmentation profiles obtained with bioanalyser. (A) Not tagmented sample, there is no clear peak. (B) Overtagmented sample, there is a peak around 200bp. (C) Properly tagmented sample with a peak around 300bp and a nucleosome profile.

3.4.2. ATAC-sequencing samples

As mentioned previously, the aim of this study was to obtain regulatory regions specific for neural crest formation. As controls, animal caps injected with *noggin* to give neuroectoderm and non-injected animal caps were also processed. The comparison between CREs in the three samples would allow the identification of NC specific enhancers.

Neural crest-induced animal caps were obtained by injection of 500pg of *noggin* and 100pg of *Wnt1* at stage 2 and dissection of the animal pole at stage 9. Neuroectodermal caps were obtained by injection of 500pg of *noggin* and ectodermal caps were obtained without injecting any inducing agent. All of them,

however, were injected with 200pg of *gfp* to ensure the embryos were properly injected in case of neural crest and neuroectoderm, and in the case of ectoderm as a control of the injection process.

In each experiment, a subset of sibling caps was collected to check that the expression profile of the samples was appropriate as determined by RNA extraction, cDNA synthesis and qPCR. The expression of NC markers (*foxD3* for early stages and *snai2* for later stages), neuroectodermal markers (*sox2*), ectodermal marker (*keratin*) and mesodermal markers (*myoD*) was evaluated (Figure 29). None of the samples contained *myoD* indicating that there was no mesodermal contamination during the dissection process. Ectodermal caps just expressed *keratin* (Figure 29A, B). Neuroectodermal caps showed mainly *sox2* expression (Figure 29C, D). Finally neural crest caps expressed *foxD3* and *snai2* confirming that they were NC-induced. Note they also expressed *sox2* which is both a neural marker and a marker for stemness (Figure 29E, F) (Zhang & Cui 2014).

For each condition three replicates were obtained. The bioanalyser analysis and the sequencing using a MiSeq was performed in collaboration with the Sauka-Spengler Lab at the University of Oxford, obtaining approximately 15M reads per sample. All the samples were properly tagmented, as a peak at 300bp, approximately was obtained, except the neuroectoderm stage 18 replicate 3 that later will be discarded for being overtagmented (Figure 30 and Figure 31).

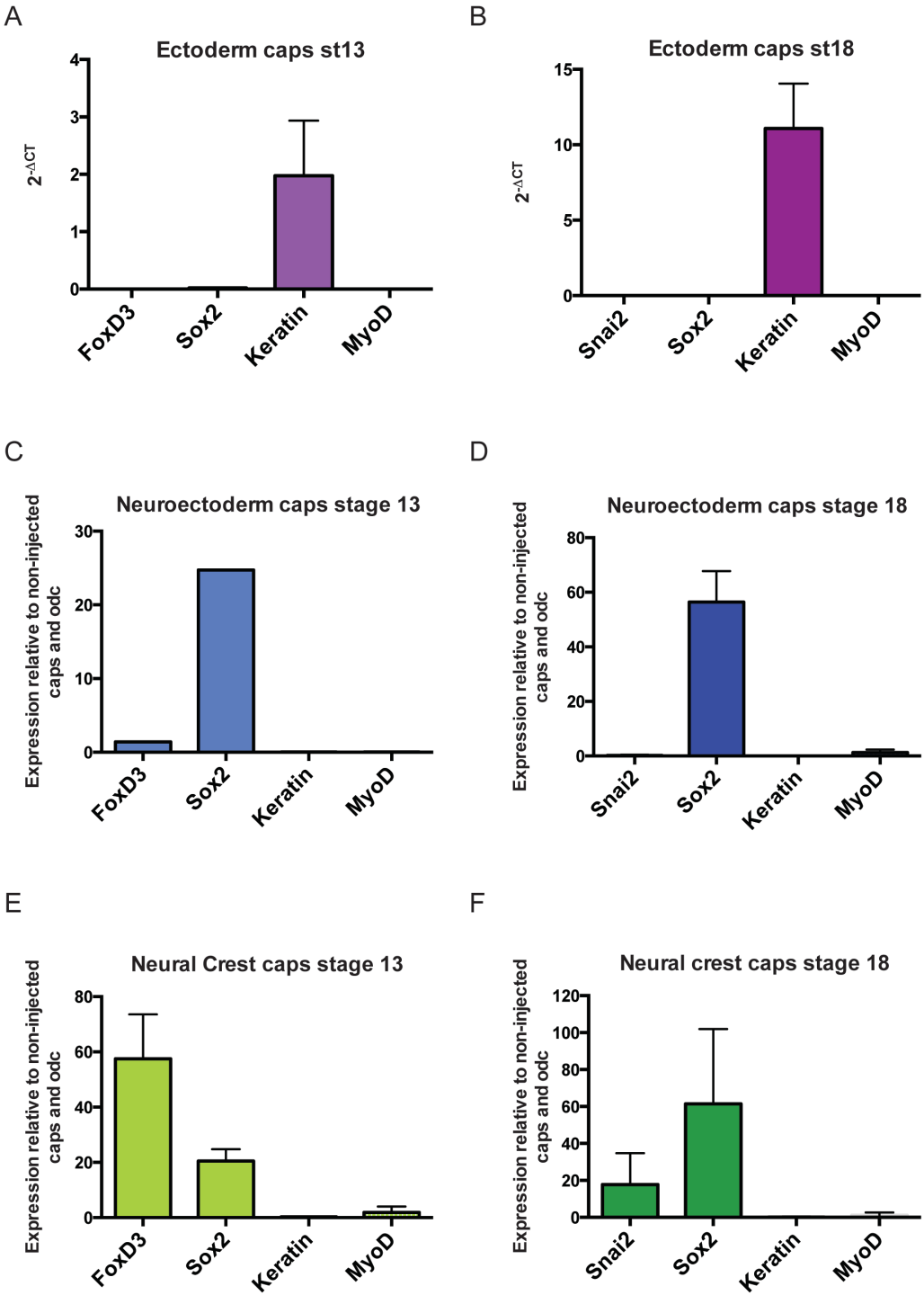


Figure 29: Expression prolife of samples used in ATAC-sequencing. (A) Ectodermal caps stage 13. (B) Ectodermal caps stage 18. (C) Neuroectodermal caps stage 13. (D) Neuroectodermal caps stage. (E) Neural crest caps stage 13. (F) Neural crest caps stage 18. The expression of NC markers (*foxd3* and *snai2*), neuroectoderm markers (*sox2*), ectoderm markers (*keratin*) and mesoderm marker (*myoD*) was evaluated by qPCR. NC and neuroectoderm samples were normalized against the *odc* in non-injected samples and ectoderm samples were normalized against *odc* expression. Mean+SD is represented.

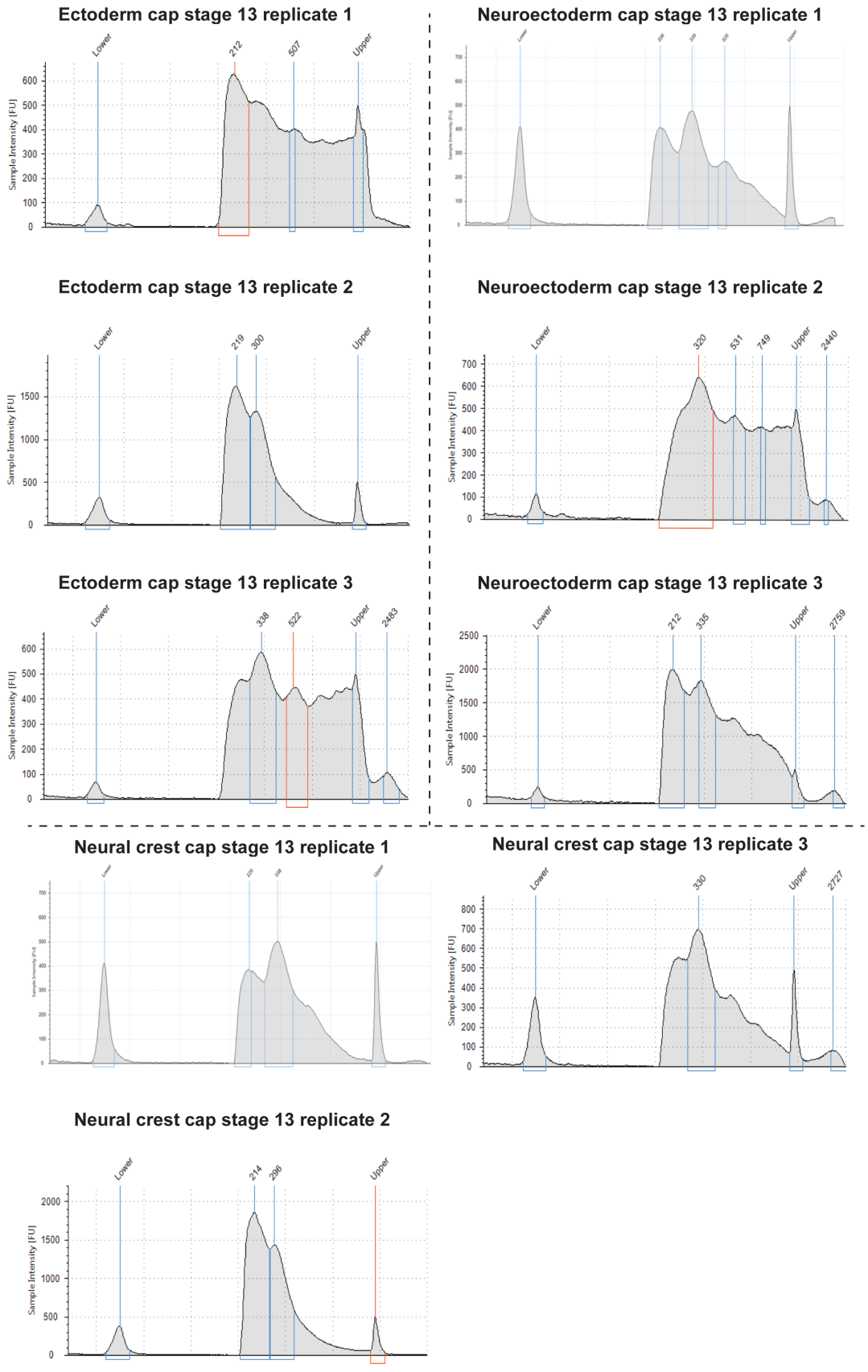


Figure 30: Bioanalyser profile of the stage 13 ATAC-sequencing samples.

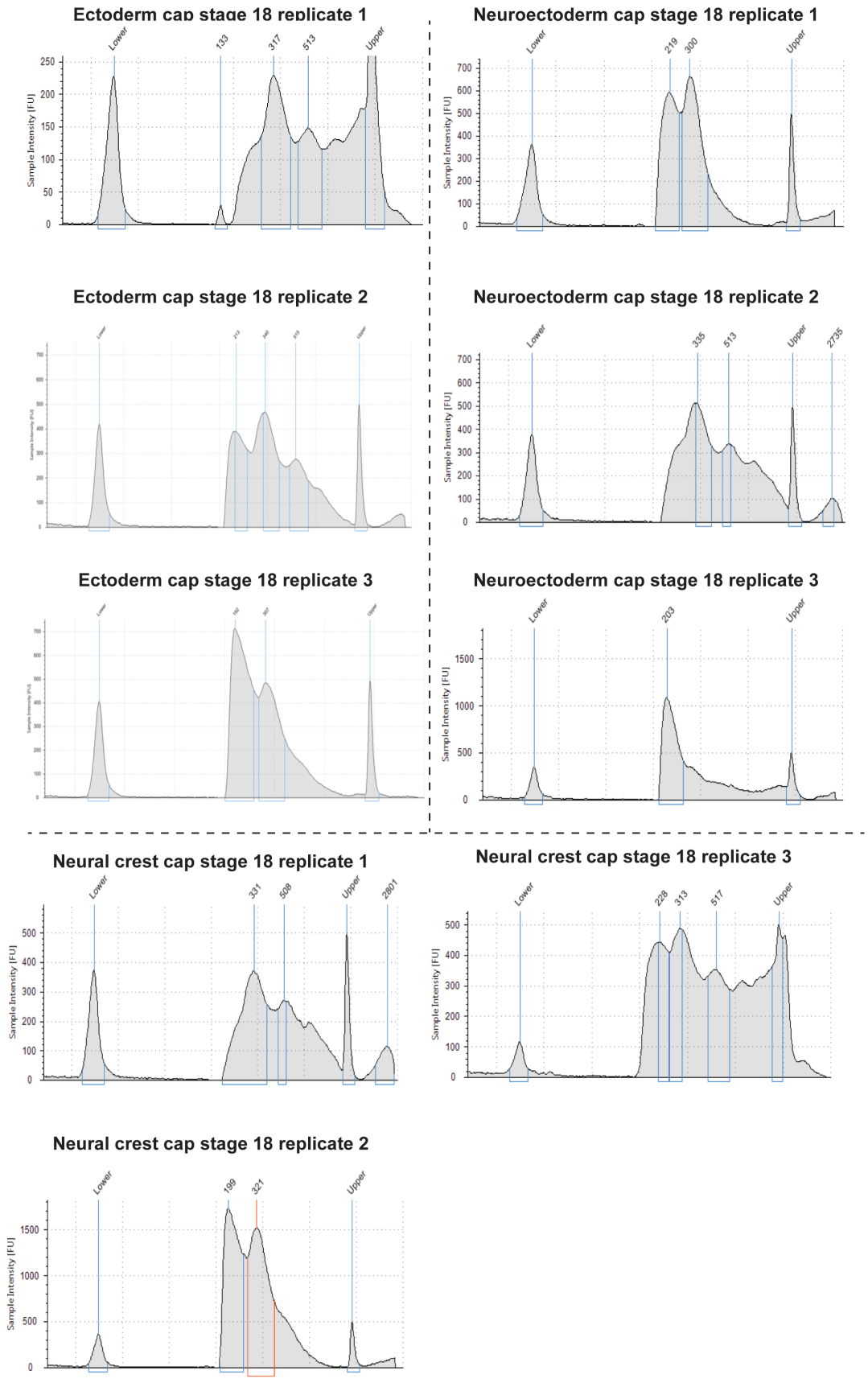


Figure 31: Bioanalyser profile of the stage 18 ATAC-sequencing samples.

After checking the quality of the samples using FastQC, the data analysis was performed. Reads were aligned to the *X. laevis* genome version 9.1 using Bowtie1 and just those pair-reads with a unique mapping site were kept. According to Encode an alignment rate of 80% is acceptable. The percentage of properly aligned reads was between 41-72%. This could be due to the fact that reads mapping in more than one place were discarded. The *X. laevis* genome is still not optimally assembled and the genome is duplicated thus favoring the multiple mapping of the reads.

The insert size distribution of the samples was analysed to check the number of reads that map on the nucleosome free (NF) region and on the mono-nucleosome (MN) fraction (Figure 32-34). The NF fraction is smaller than 100bp and, as it is free of nucleosome, is more accessible to transcription factors. The MN fraction comprises DNA fragments between 180 to 250bp and it is protected with nucleosomes. Enhancers contain a central region with no nucleosomes or highly remodeled nucleosomes flanked with mononucleosomes with highly modified histone tails (Hoffman et al. 2013).

Most of the samples display a similar metric and also similar to the one reported in Buenrostro et al. (2013). There is an increase at 50bp and then a decrease in the distribution of reads and a more subtle increase in the MR region. The metrics also show a periodicity of approximately 10.5bp, which is equal to the helical pitch of DNA. The ratio between NF and MN is close to 2, suggesting that the samples are enriched in the nucleosome free fraction and therefore in cis-regulatory elements, similar to the one reported in Buenrostro et al. (2013). The peaks were then called in the whole sample using MACS2.

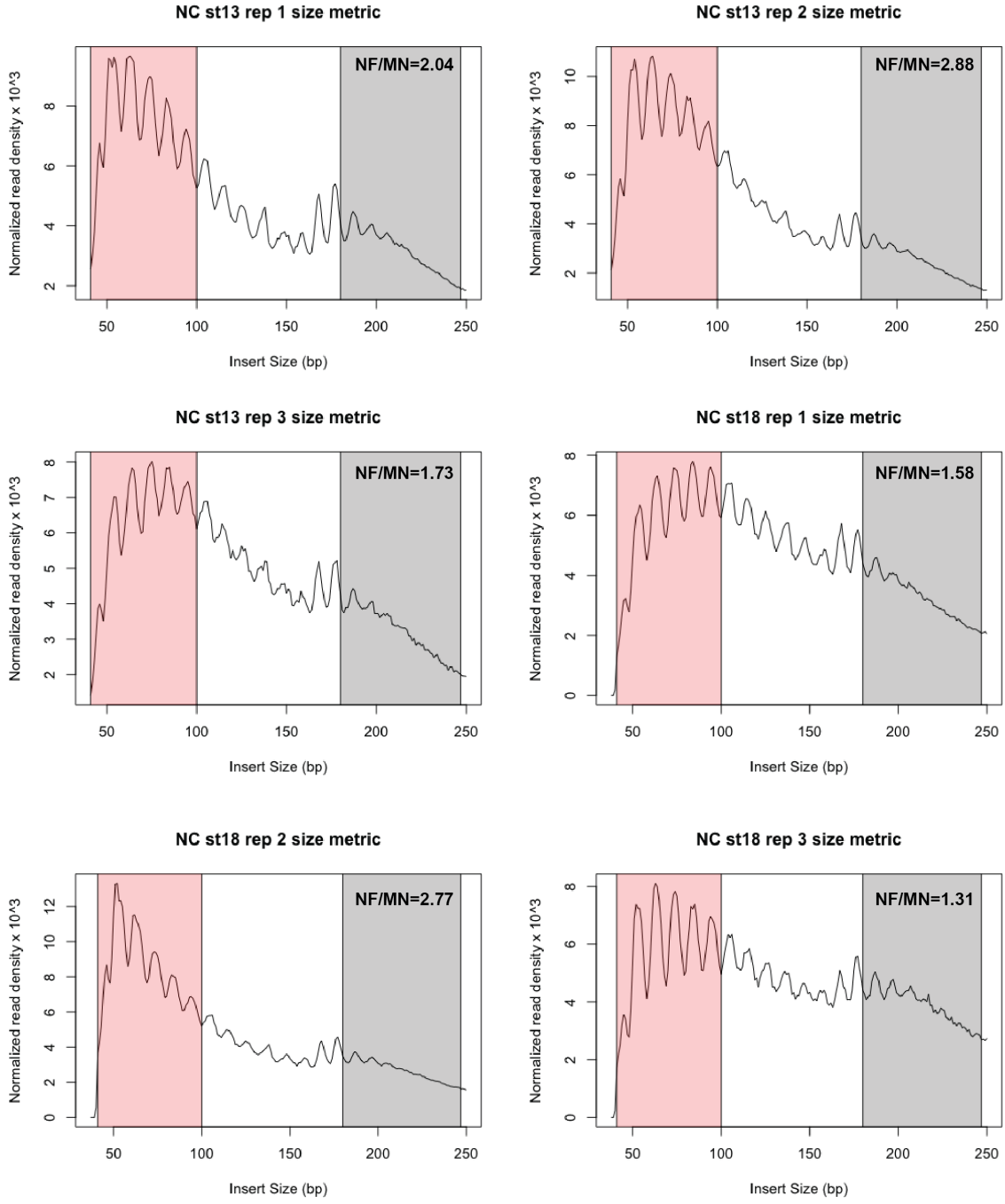


Figure 32: Metrics of neural crest samples. The plot represents the normalized reads in function of the different insert size. Red rectangle marks nucleosome free (NF) region and black rectangle mark mononucleosome (MN) region. The ratio between NF and MN is shown.

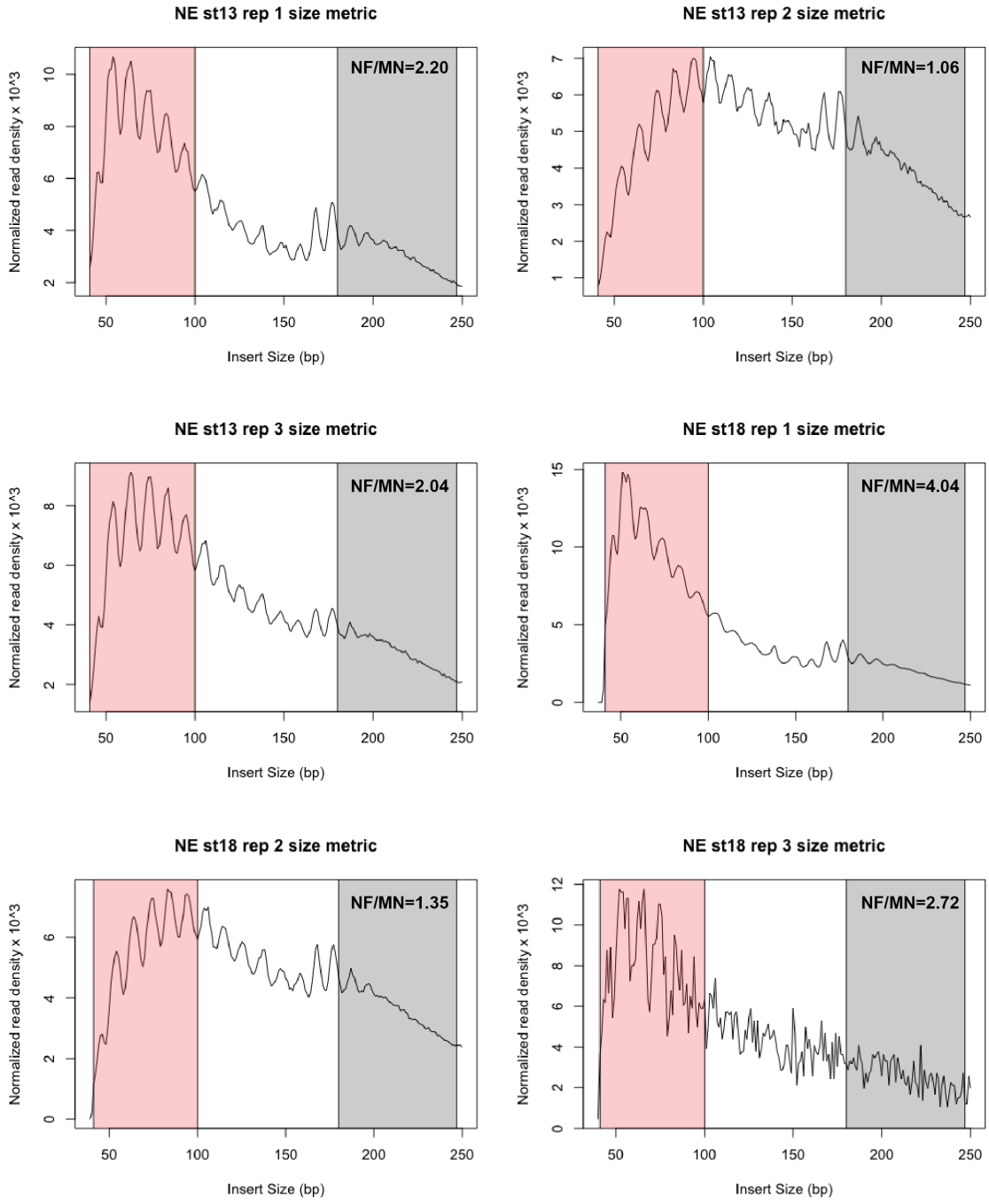


Figure 33: Metrics of neuroectoderm (NE) samples. The plot represents the normalized reads in function of the different insert size. Red rectangle marks nucleosome free (NF) region and black rectangle mark mononucleosome (MN) region. The ratio between NF and MN is shown.

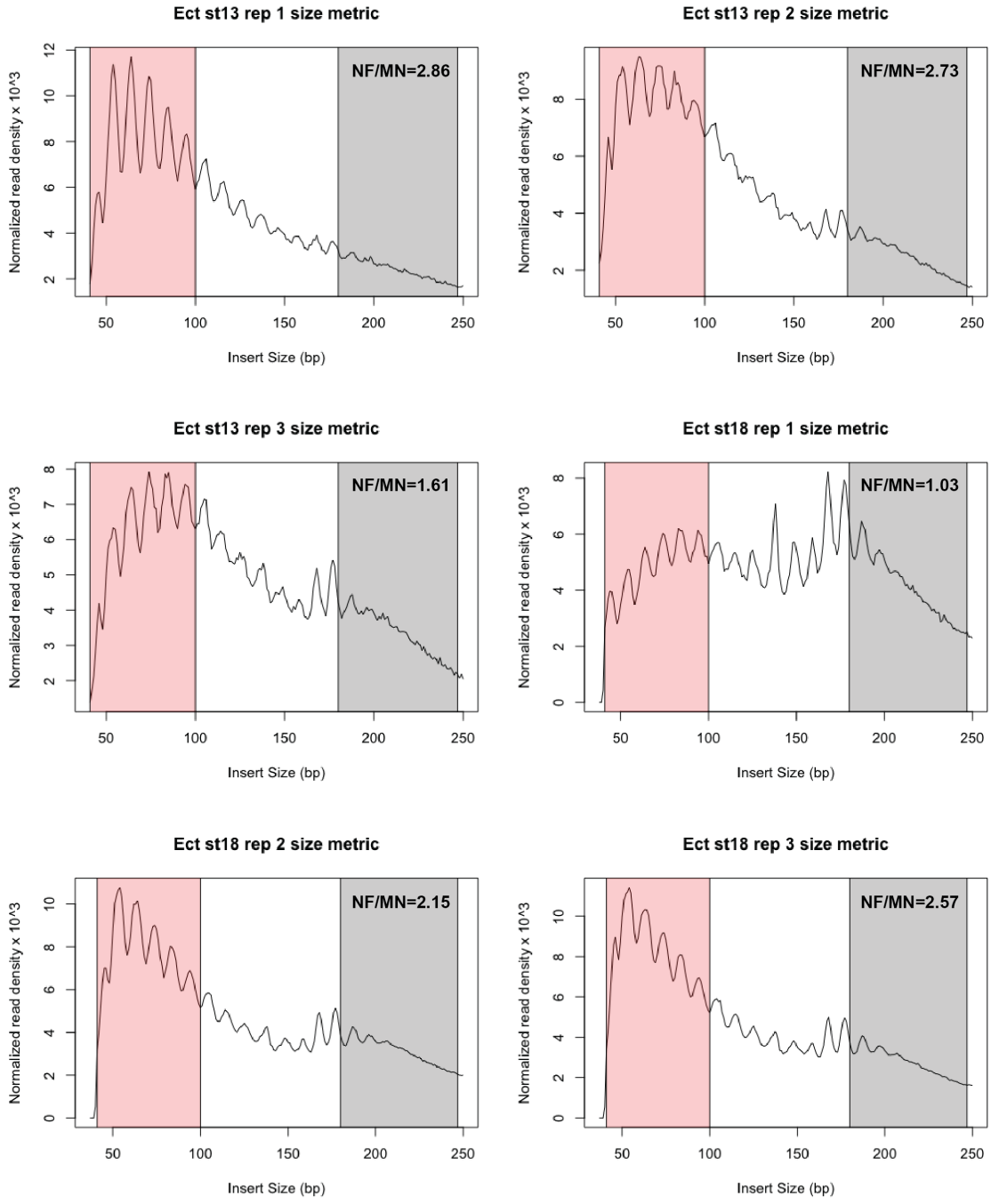


Figure 34: Metrics of ectodermal samples. The plot represents the normalized reads in function of the different insert size. Red rectangle marks nucleosome free (NF) region and black rectangle mark mononucleosome (MN) region. The ratio between NF and MN is shown.

3.4.3. Comparison between ATAC-sequencing samples

In order to compare the different samples, first, the three replicates were combined; only those peaks overlapping 50% in all the samples were selected for the following analysis, as recommended in the Encode guidelines version 7 (Encode 2017). Using this, at stage 13 there are 17,997 peaks present in the NC sample, 20,274 peaks in the neuroectoderm sample and 12,245 in the ectoderm sample. The number of peaks common in the different samples was then analysed. There are 5,592 peaks common in the three samples, 1,116 peaks common between NC and ectoderm samples, 1,816 peaks common between neuroectoderm and ectoderm samples and lastly, 5,264 peaks are common between NC and neuroectoderm samples. Finally, there are 6,025 peaks specific for NC, 7,602 peaks for neuroectoderm and 3,721 for ectoderm (Figure 35A).

The same analysis was carried out for the stage 18 samples. 11,705 are present in the NC sample, 46,515 peaks in the neuroectodermal sample and 23,923 peaks in the ectodermal sample (Figure 35B). There are 7,541 peaks common for the three samples, 2,618 peaks common between NC and neuroectoderm samples, 9,176 peaks common between neuroectoderm and ectoderm samples and 334 peaks common between NC and ectoderm samples. Finally, there are 1,212 peaks specific for NC samples, 27,181 peaks specific for neuroectoderm samples and 6,872 peaks are specific for ectoderm samples. There are a lot of peaks exclusive for neuroectoderm sample and this could be due to the fact that there are only two replicates in this sample.

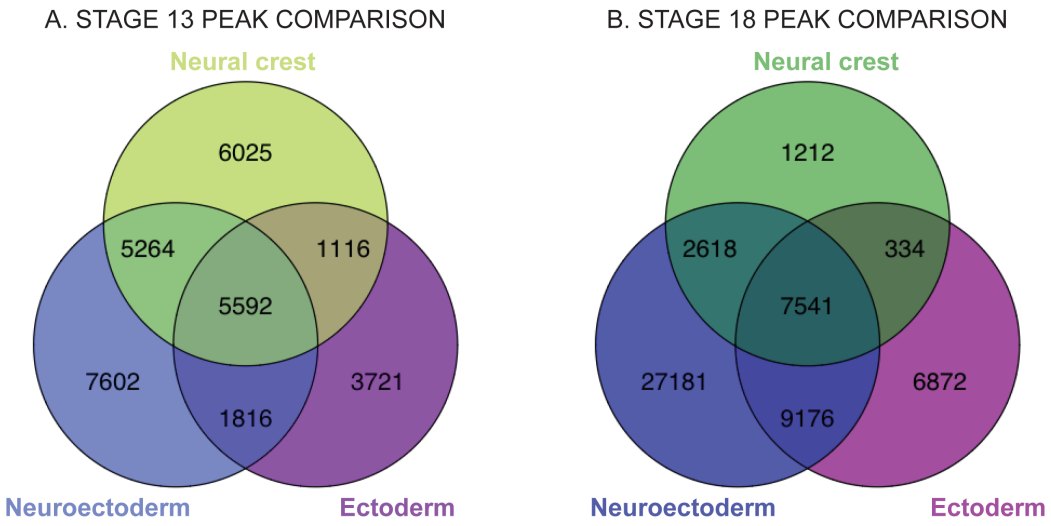


Figure 35: Venn diagrams comparing peaks of open regions present in the different samples. Peaks overlapping 50% were considered the same. (A) Comparison of NC, neuroectoderm and ectoderm at stage 13. (B) Comparison of NC, neuroectoderm and ectoderm at stage 18.

3.4.4. Analysis of the ectodermal cis-regulatory elements

The ectoderm samples were analysed with the aim of ensuring the experiment was well designed and the expected results were obtained (Figure 36). The behavior of the ectoderm marker *keratin* across the different samples was studied to make sure the promoter of this gene showed a significant increase of reads on the ectodermal samples, suggesting activation (Figure 36A). Compared with the neural crest and neuroectoderm samples, the ectodermal samples show an increase of reads across the gene and in the promoter. This confirms that the ectodermal samples behave as expected.

The specific peaks present in the ectoderm samples were then compared between the two stages. Out of the 3,721 specific peaks present at stage 13 and 6,872 peaks present at stage 18, there are 523 peaks common to the two samples, 3,198 peaks only present at stage 13 and 6,349 peaks only present at stage 18 (Figure 36B).

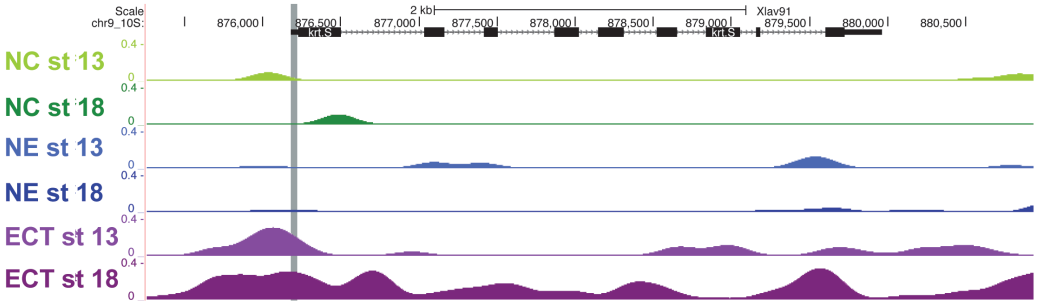
Peaks present in the different samples were located in several genome locations. In order to identify which peaks correspond to enhancers, the genomic location of each peak was determined. Those selected correspond to intergenic regions as this relates to possible cis-regulatory elements. Several chromatin marks

and a high degree of conservation have been found in the first intron of genes, indicating that the presence of enhancers in the first intron (Park et al. 2014). However the *X. laevis* annotation does not allow the distinction between exons and introns and therefore these potential enhancers were not considered. The genome was divided into four different locations: (1) genes, including exon and intron regions, (2) promoter regions defined as TSS + /– 200bp, (3) proximal enhancer regions that were defined as peaks in a range of 2kb upstream the TSS and finally, (4) distal enhancers which are peaks further than 2kb from the TSS and not present inside a gene. The analysis of the ectodermal samples reveals that 54% of the stage 18 peaks correspond to intergenic regions, in particular 41.9% to distal enhancers and 13.1% to proximal enhancers, 11.2% correspond to promoters and 33.8% to genic regions. On the other side, 59.3% of the stage 13 ectoderm peaks correspond to intergenic regions, in particular, 48.6% to distal enhancers and 10.7% to proximal enhancers, 8.5% correspond to promoters whereas 32.2% are inside gene coordinates (Figure 36C).

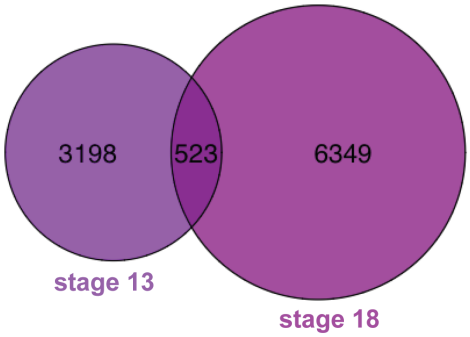
The intergenic regions, 2,556 and 4,697 for stage 13 and stage 18, respectively, were selected to continue with the analysis, as they could be potential enhancers. The motifs present in the stage 18 intergenic regions were analysed using Homer in order to detect which transcription factors are bound to the peaks and could drive expression of genes involved in the formation of non-neural ectoderm. The most relevant transcription factors are *gata2*, *gata3* and *six1* (Figure 36D). The complete list of motifs found in the ectodermal stage 18 intergenic regions is displayed in Appendix 1 Table iii. The transcription factors *gata2* and *gata3* are involved in the anteroposterior segmentation of the non-neural ectoderm. In *Xenopus* and Zebrafish *gata2* and *gata3* are expressed anteriorly in the prospective head region, *gata2* is expressed in the middle region and none is expressed in the posterior region that will become the tail (Read et al. 1998). Later on during development, *gata2* is involved in the formation of erythropoiesis progenitors whereas *gata3* is required for luminal epithelial cells differentiation in mammary glands (Dalgin et al. 2007; Chou et al. 2010). The motifs for both transcription factors are found in stage 18 ectoderm peaks; 14.36% of the peaks contain *gata2* motifs whereas 24.24% contain *gata3*. Also, 4.5% of the peaks contain *six1* motif. This transcription factor is crucial for the pre-placodal ectoderm formation, together with its co-activator *eya1*, which give rise to the cranial sensory organs and ganglia. A transcriptome analysis of the pre-placodal region determined

that *six1* acts upstream of other transcription factors involved in sensory neurons differentiation such as *atoh1* and *neurog1* (Riddiford & Schlosser 2016).

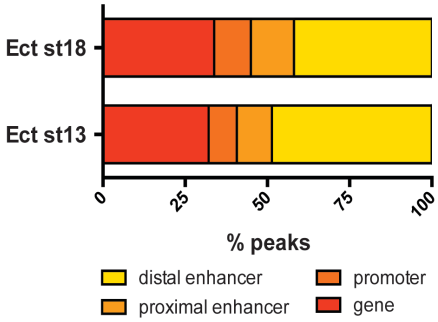
A. GENOMIC LOCI OF A ECTODERM MARKER: *keratin*



B. COMPARISON OF PEAKS BETWEEN STAGE 13 AND STAGE 18 ECTODERM



C. DISTRIBUTION OF PEAKS IN DIFFERENT GENOMIC LOCATIONS



D. MOTIFS PRESENT IN THE INTERGENIC STAGE 18 PEAKS

Motif	TF	P-value	targets with motifs	Reference
	*gata2*	10^{-39}	14.36%	Read et al. (1998)
	*gata3*	10^{-25}	24.24%	Read et al. (1998)
	*six1*	10^{-5}	4.50%	Riddiford and Schlosser (2016)

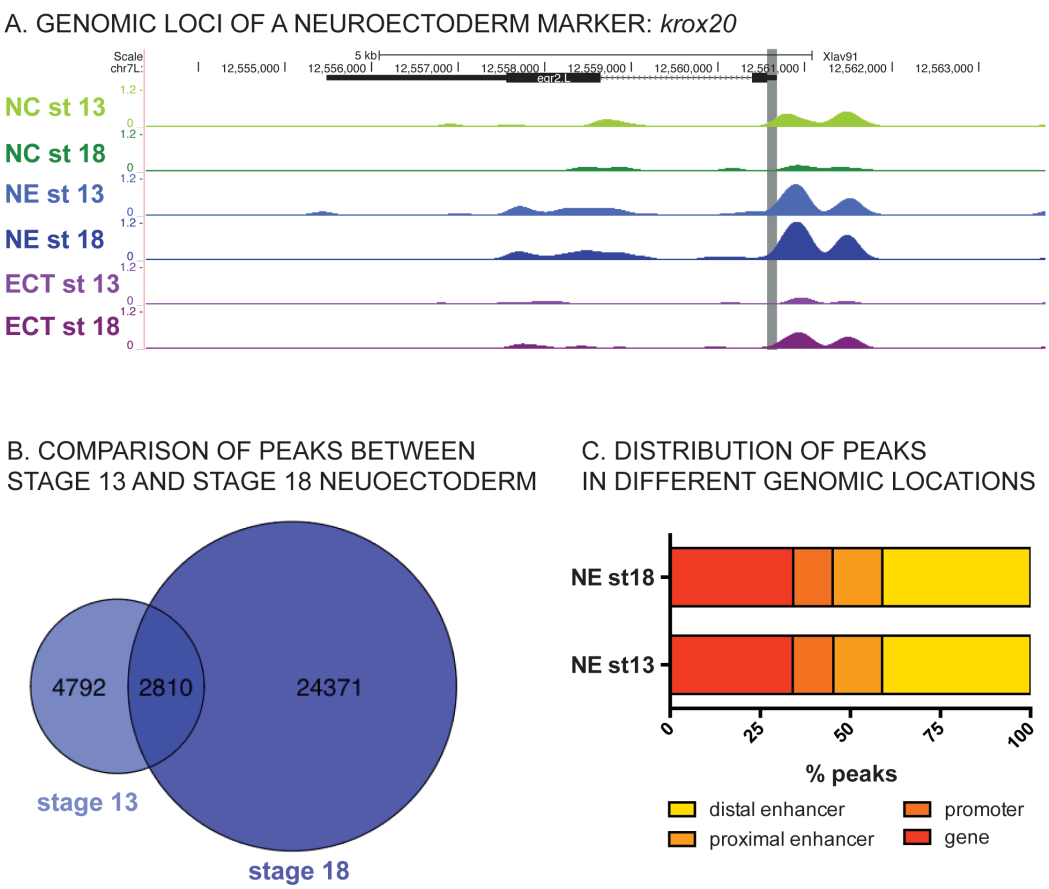
Figure 36: Analysis of ectoderm cis-regulatory elements. (A) Genomic location of an ectodermal marker, *keratin*, showing a representative track for each condition, neural crest caps stage 13 (NC st13), neural crest caps stage 18 (NC st18), neuroectoderm caps stage 13 (NE st13), neuroectoderm caps stage 18 (NE st18), ectodermal caps stage 13 (ECT st13) and ectodermal caps stage 18 (ECT st18). Grey rectangle marks the promoter of the gene. (B) Venn diagram showing the distribution of ectodermal specific peaks in the two stages analysed, stage 13 and stage 18. (C) Distribution of the ectoderm specific peaks into four different genomic locations, gene, promoter defined as TSS+/- 200bp, proximal enhancer, defined as 2kb upstream the TSS, and distal enhancer. (D) Significant ectoderm motifs present in the stage 18 intergenic ectoderm peaks.

3.4.5. Analysis of the neuroectoderm cis-regulatory elements

In a similar way to the ectodermal samples, the neuroectoderm samples were also analysed to check if expected results were obtained (Figure 37). The analysis of the *krox20* loci, also known as *egr2*, revealed that there was an enrichment of reads in the promoter. Although the neural crest and ectoderm samples also showed some reads in the promoter region, they were not as abundant as the neuroectodermal samples (Figure 37A).

The analysis of the specific peaks revealed that among the 7,602 stage 13 peaks and the 27,181 stage 18 peaks, there are 4,792 exclusive for the neuroectoderm stage 13, 24,371 exclusive for stage 18 and 2,810 are common between the two stages (Figure 37B). The analysis of the peak location revealed that both stage 13 and stage 18 have a similar proportion of peaks in each location. The 54% of both samples correspond to intergenic regions, in particular, 41% to distal enhancers and 13% to proximal enhancers, 10% correspond to promoters and 34% to gene locations (Figure 37C).

The intergenic regions for stage 13 and stage 18, 5,246 and 19,147, respectively were identified as a potential enhancers involved in the formation of the neural tissue. Analysis of the motifs present in stage 18 cis-regulatory elements revealed that *nkx6.1*, *otx2* and *sox3* are statistically significant present on those peaks (Figure 37D). The complete list of motifs found in the neuroectoderm stage 18 intergenic regions is displayed in Appendix 1 Table iv. The transcription factor *nkx6.1*, which is present in 57.45% of the peaks, is required for mid-hindbrain boundary. This gene acts upstream of *en2* and *Wnt1* which are both necessary for the mid-hindbrain boundary (Ma et al. 2013). 21.44% of the peaks contain *otx2*-binding sites. *otx2* is required to establish the presumptive forebrain, where the eye field is specified, and midbrain (Gammill & Sive 2001; Zuber et al. 2003). Finally, *sox3* is found in 46.89% of the intergenic stage 18 peaks. This gene is expressed in the presumptive central nervous system (Penzel et al. 1997).



D. MOTIFS PRESENT IN THE INTERGENIC STAGE 18 PEAKS

Motif	TF	P-value	targets with motifs	Reference
	*nkx6.1*	10^{-68}	57.45%	Ma et al. (2013)
	*otx2*	10^{-140}	21.44%	Gammill and Sive (2001)
	*sox3*	10^{-491}	46.89%	Penzel et al. (1997)

Figure 37: Analysis of neuroectoderm cis-regulatory elements. (A) Genomic location of an neuroectodermal marker, *krox20* also known as *egr2*, showing a representative track for each condition, neural crest caps stage 13 (NC st13), neural crest caps stage 18 (NC st18), neuroectoderm caps stage 13 (NE st13), neuroectoderm caps stage 18 (NE st18), ectodermal caps stage 13 (ECT st13) and ectodermal caps stage 18 (ECT st18). Grey rectangle marks the promoter of the gene. (B) Venn diagram showing the distribution of neuroectodermal specific peaks in the two stages analysed, stage 13 and stage 18. (C) Distribution of the neuroectoderm specific peaks into four different genomic locations, gene, promoter defined as TSS+/- 200bp, proximal enhancer, defined as 2kb upstream the TSS, and distal enhancer. (D) Significant neuroectoderm motifs present in the stage 18 intergenic neuroectoderm peaks. The p-value, percentage of target sequences with the motif and a bibliographic reference relating the transcription factor to the formation of neuroectoderm are shown.

3.4.6. Analysis of the neural crest cis-regulatory elements

The analysis of the *snai2* locus reveals that only in the neural crest samples there is a significant enrichment of reads in the promoter region compared with the neuroectoderm and ectoderm samples. This observation, together with the *keratin* and *krox20* loci, gives more confidence in the generated data (Figure 38A).

The comparison between stages showed that 85 peaks are common between the two stages whereas 5,940 peaks are exclusive for stage 13 and 1,127 exclusive for stage 18 (Figure 38B). Furthermore, the study of the peak location of the NC samples revealed that at stage 13 there are 46.3% of peaks located in intergenic regions, with 23% in distal enhancers and 23.3% in proximal enhancers, 21.5% correspond to promoters and 32.2% to genes. At stage 18, 62% of the peaks correspond to intergenic regions, 53.4% to distal enhancers and 8.7% to proximal enhancers, 7.2% correspond to promoters and 30.7% to genic regions (Figure 38C). To continue with the analysis, the intergenic peaks, 852 at stage 13 and 4,528 at stage 18, were used.

3.4.7. Neural crest enhancer dynamic

The ATAC-sequencing analysis allows the identification of different enhancer dynamics among the NC specific peaks (Figure 39). The most frequent enhancer dynamics found are CRE present exclusively at stage13, 2,460 peaks, and at stage 18, 715 peaks (Figure 39B and C). From the peaks present at both stages, the coverage in each peak was determined to identify those peaks differentially enriched in either one stage or the other (Figure 39D-F). Out of the 41 common intergenic peaks, most of them, 39 peaks, show a similar coverage in both stages. Just one peak is significantly more expressed at stage 13, the closest gene of which is *pcdh8.2.L* at 28kb. On the other side, a peak located at 300bp of *slc40a.S* shows a significant increase in the coverage at stage 18 compared with stage 13 (Figure 39 D and E). However, according to the RNA-sequencing data generated, none of those genes are expressed in NC-induced caps.

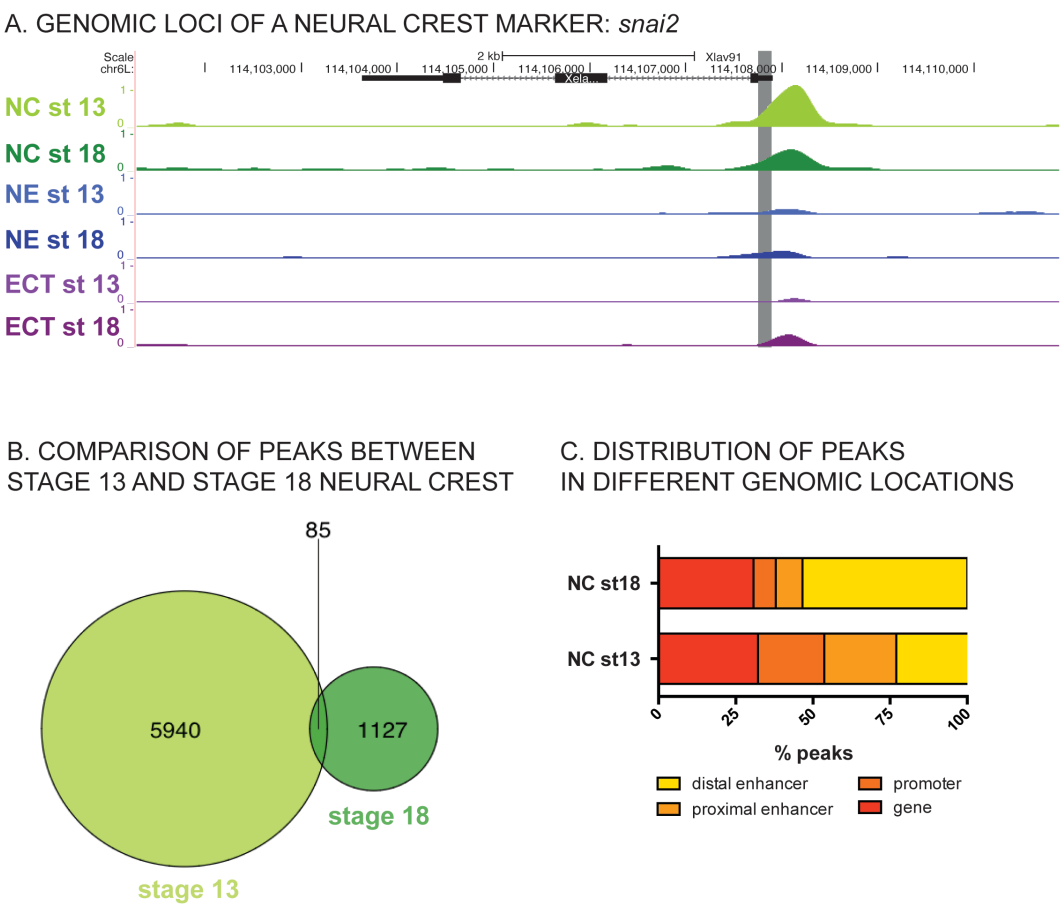


Figure 38: Analysis of neural crest cis-regulatory elements. (A) Genomic location of an neural crest marker, *snai2*, showing a representative track for each condition, neural crest caps stage 13 (NC st13), neural crest caps stage 18 (NC st18), neuroectoderm caps stage 13 (NE st13), neuroectoderm caps stage 18 (NE st18), ectodermal caps stage 13 (ECT st13) and ectodermal caps stage 18 (ECT st18). Grey rectangle marks the promoter of the gene. (B) Venn diagram showing the distribution of neural crest specific peaks in the two stages analysed, stage 13 and stage 18. (C) Distribution of the neural crest specific peaks into four different genomic locations, gene, promoter defined as TSS+/- 200bp, proximal enhancer, defined as 2kb upstream the TSS, and distal enhancer.

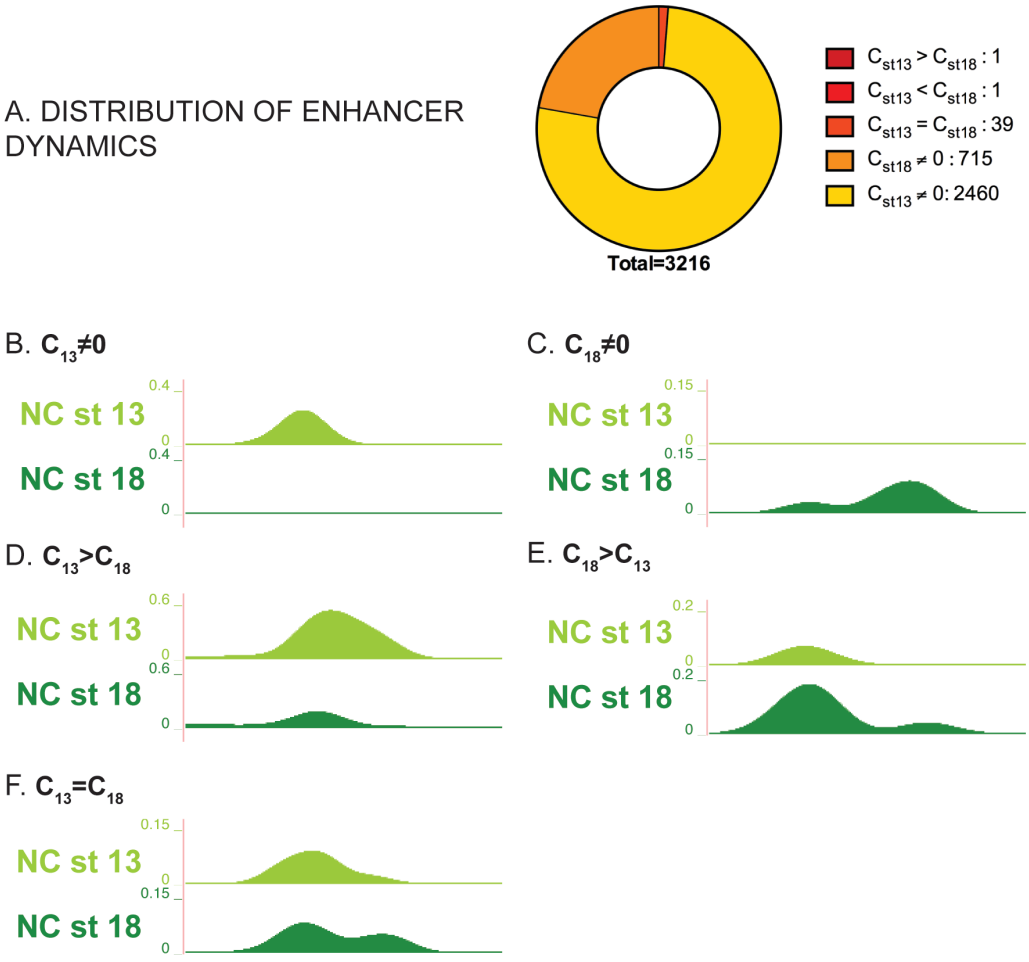


Figure 39: Enhancer dynamics. (A) Pie chart showing the distribution of the different enhancer dynamics in the intergenic NC peaks. The number of peaks in each category is annotated in the legend. (B) Example of a peak only present at stage 13. (C) Example of a peak only present at stage 18. (D) Example of a peak present at stage 13 and 18 with the coverage higher at stage 13, close to *pcdh8.2*. (E) Example of a peak present at stage 13 and 18 with the coverage higher at stage 18, close to *slc40a*. (F) Example of a peak present at stage 13 and 18 with the coverage equal in both stages.

3.4.8. Analysis of neural crest transcription factor motifs present in intergenic neural crest open regions

The presence of neural crest-related motifs in the intergenic NC peaks was evaluated using Homer (Heinz et al. 2010). The complete list of motifs found in the stage 13 and 18 intergenic regions are displayed in Appendix 1 Tables i and ii. This data reveals NC specific transcription factors involved in activating other NC genes, through binding to their enhancers. The RNA-sequencing data was used to support this data as it confirmed if specific transcription factors were being expressed at these stages.

Figure 40 shows that a *zic* motif is present in 18.39% of the intergenic open regions at stage 13 and 17.46% at stage 18 with a p-value smaller than 10^{-10}. The presence of zic bound to NC enhancers is likely to happen as *zic1* is expressed in both stages at high levels, according to the RNA-sequencing. *pax3* and *meis* show similar results. A *pax3* motif is present in 4.88% of the intergenic peaks at stage 13, and in 4.37% of the peaks at stage 18, in both cases with a p-value smaller than 10^{-2}. *meis* motif is present in 28.91% and 26.19% of the stage 13 and stage 18 intergenic peaks, respectively, also with a p-value smaller than 10^{-2}. Both *pax3* and *meis3* are expressed at the two stages, *pax3* with more than 200 tpm and *meis3* between 100-200 tpm in both stages. A different case is *pax7*, this motif appears in a low percentage of NC peaks and moreover this gene shows almost no expression across the different stages. The *sox10* motif is present in 26.23% and 28.04% of peaks with a significant p-value at stage 13 and 18, respectively. However, while *sox10* is highly expressed at stage 18, it is almost not detected at stage 13 suggesting that *sox10* plays an important role in NC enhancer activation at stage 18. *sox9* motif is present in 12.04% of the stage 13 peaks and 15.34% of the stage 18 peaks, in both cases with a p-value between 10^{-10} and 10^{-2}. *sox9* is expressed at both stages. *nmyc* is expressed at stage 18, with differences between the S and the L form. The *nmyc* motif is only present significantly in 9.48% of the inergenic NC regions at stage 13. *cmyc* also shows a similar pattern, although *cmyc* is highly expressed in both stages, its motif is only significantly present in 7.32% of the stage 13 intergenic regions. *tfap2a* motif is only significantly present in 8.60% of the stage 18 peaks where both homologous genes are highly expressed.

In conclusion, based on the ATAC-sequencing combined wih the RNA-sequencing data, the NC genes relevant for the stage 13 enhancer activation are *zic1*, *pax3*, *meis3* and c*myc*. At stage 18, *zic1*, *pax3 meis3*, *sox10*, *sox9* and *tfap2a*

seem to be the most relevant transcription factors for the activation of stage 18 enhancer.

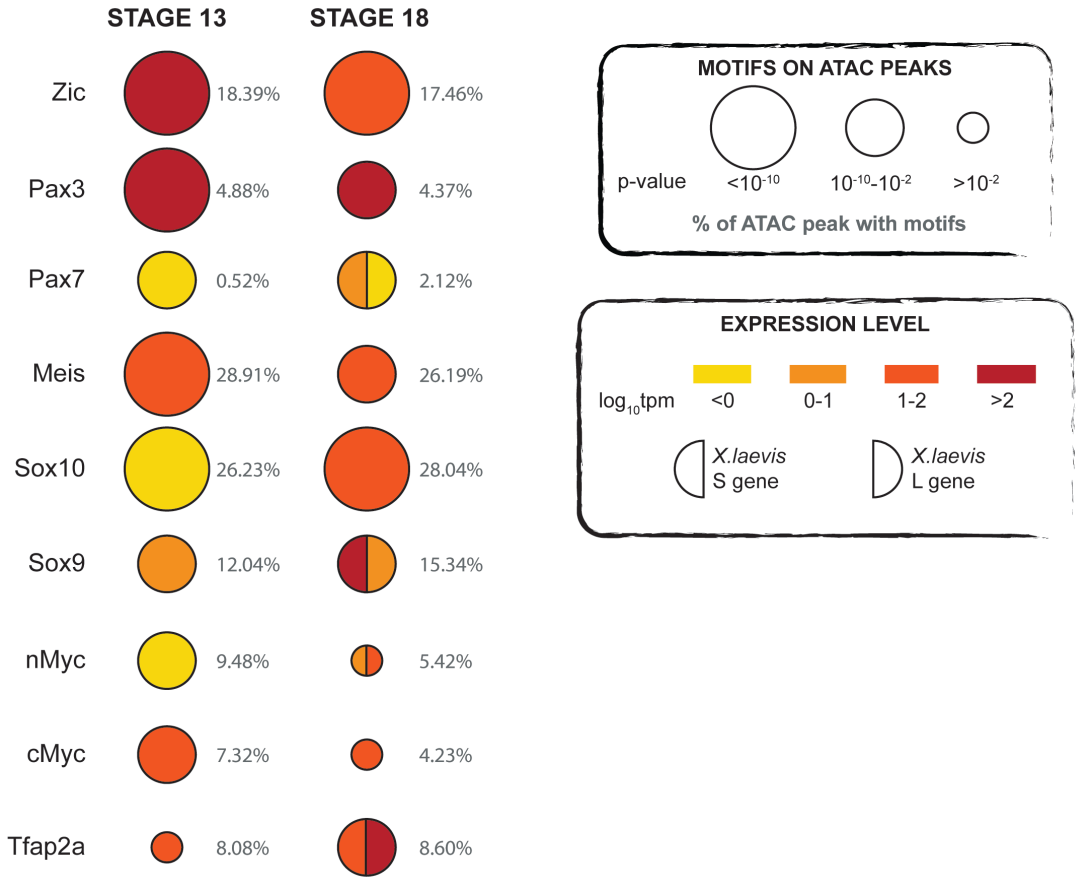


Figure 40: Analysis of neural crest motifs present in NC enhancers. The diameter of the circles represents arbitrary the significance of motifs found in the intergenic peaks using Homer (Heinz et al. 2010). Small circles represent p-value greater then 10^{-2}, mid-size circles represent p-value between 10^{-2} and 10^{-10} and big circles represent p-values small than 10^{-10}. Next to the circle the percentage of the motif in the peaks selection is shown. The level of expression of each gene is shown by the color of the circle: yellow marks no expression, less than 1 transcript per million (tpm) and log10tpm <0, orange marks low expression, between 1 and 10 tpm and log10tpm between 0 and 1, dark orange marks moderate expression, between 10 and 100 tpm, and log10tpm between 1-2 and brown marks high expression, more than 100 tpm, and log10tpm grater than 2. The left half of the circle represents the S homologous of the genes and the right half of the circle the L homologous of the gene.

3.4.9. Analysis of pluripotency transcription factor motifs present in intergenic neural crest open regions

Using a similar strategy and the same peak selection used in section 3.4.8, the importance of pluripotency factors in neural crest formation has been evaluated. The presence of *oct4*, *sox2* and *nanog* motifs was studied in the context of their homologs expression during NC formation (Figure 41).

The *oct4* motif is present in 11.70% of the stage 13 NC and in 4.89% of the stage 18 NC intergenic peaks, both significantly. The *oct4 Xenopus* homologs, *pou5f3.2* and *pou5f3.3* are all highly expressed at stage 9. At stage 13 *pou5f3.1* is highly expressed whereas *pou5f3.2* and *pou5f3.3* have moderate and low expression. At stage 18, *pou5f3.1* is moderately expressed and *pou5f3.2* and *pou5f3.3* are low expressed. The *sox2* motif is present in 16.31% of the stage 13 peaks and 19.05% of the stage18 peaks. There are two *Xenopus* homologs, *sox2* and *sox3*. The expression of the *sox2*.S form is moderate-high across all stages whereas the L form has not been annotated yet if shown to be present. *sox3* shows a similar pattern to *sox2*, *sox3* is expressed also moderate-high across all stages and its motif is present in 28.55% and 31.35% of the stage 13 and stage 18 neural crest intergenic peaks, respectively. Finally, the *nanog* motif is significantly present in 56.15% of the stage 13 peaks but not at all at stage 18. However, the expression of the *nanog* homolog, *ventx1,* is low across all the stages.

To sum up, the pluripotency motifs are present in the specific NC peaks across the different stages of this cell type formation and the *Xenopus* homologs are relatively highly expressed also across the different stages. Although they are expressed and the motifs are present across the different stages, they seem more relevant at stage 13 than at stage 18.

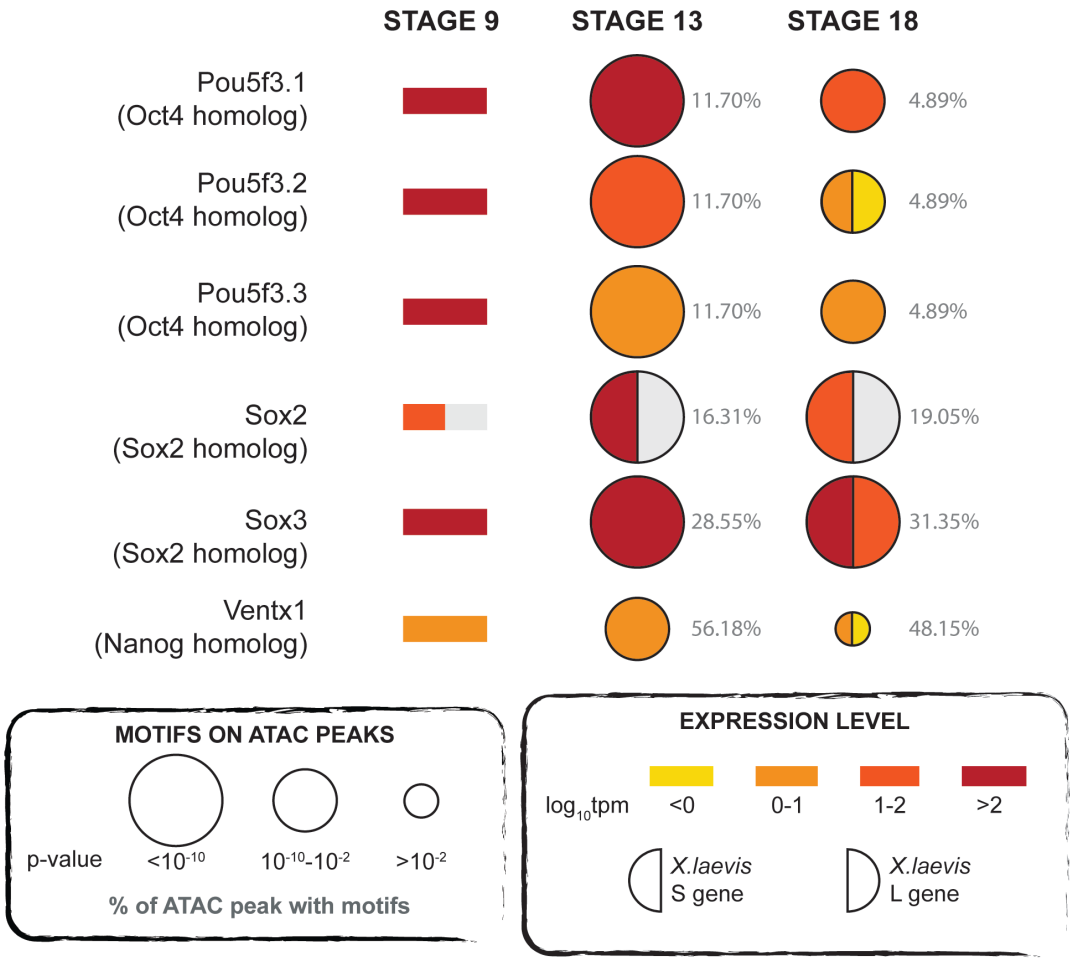


Figure 41: Analysis of pluripotency motifs present in NC enhancers. The size of the circles represents the significance of motifs found in the intergenic peaks using Homer (Heinz et al. 2010). Small circles represent p-value greater then 10^{-2}, mid-size circles represent p-value between 10^{-2} and 10^{-10} and big circles represent p-values small than 10^{-10}. Next to the circle the percentage of the motif in the peaks is shown. The level of expression of each gene is shown by the color of the circle: yellow marks no expression, less than 1 transcript per million (tpm) and log10tpm<0, orange marks low expression, between 1 and 10 tpm and log10tpm between 0 and 1, dark orange marks moderate expression, between 10 and 100 tpm and log10 between 1-2 and brown marks high expression, more than 100 tpm and log10tpm grater than 2. The left half of the circle represents the S homologue of the genes and the right half of the circle the L homologue of the gene.

3.4.10. Analysis of the cmyc as a key neural crest transcription factor in cis-regulatory elements

Recent work in the lab (see Appendix 2) suggested that the neural crest specifier *cmyc* is poised at early stages of NC formation (Figure 42). The RNA polymerase II is recruited in the promoter region of *cmyc*, upstream of the TSS, where it is bound by the pause factors, DRB-sensitivity inducing factor (DSIF), negative elongation factor (NELF) and Gdown1. In order to start *cmyc* transcription,

the transcription elongator factor (P-TEFb), mainly composed of cyclin dependent kinase 9 (CDK9) and CYCLINT1, needs to be recruited. The P-TEFb complex then phosphorylates the pause factors and the RNA polymerase II thus releasing the pause stage and allowing transcriptional elongation. This work showed that the loss of the P-TEFb components, CDK9 and CYCLINT1, leads to the loss of *cmyc* and therefore to the expression of the NC specifiers such as *sox10* (Hatch et al. 2016).

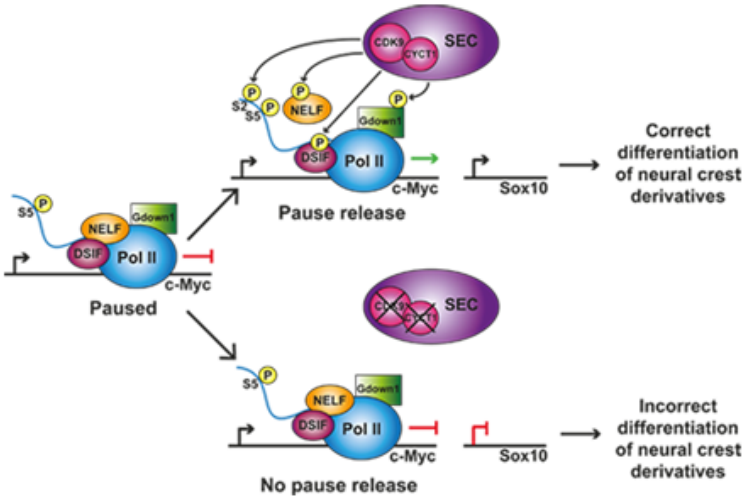


Figure 42: Model displaying the mechanism by which *cmyc* is paused in neural crest cells. On the left *cmyc* is shown in a paused state with Pol II bound to NELF, DSIF and Gdown1. We have shown that a knockdown of the P-TEFb components *cdk9* and *cyclinT1* leads to a loss of *cmyc* expression and a subsequent loss of *sox10* expression and incorrect development of neural crest derivatives (bottom scenario). This suggests that in a wild type neural crest cell, P-TEFb is promoting transcriptional elongation of *cmyc*. P-TEFb in its active form bounds to the SEC will phosphorylate DSIF, NELF, Gdown1 and serine2 of the C terminal of RNAP II. This promotes pause release on the *cmyc* gene and allows the subsequent expression of *sox10* leading to correct synchronous development of neural crest derivatives (top scenario). S2=Serine2, S5=Serine5, P=phosphorylation mark, NELF=Negative elongation factor, DSIF=DRB sensitivity inducing factor, Pol II=RNA polymerase II, CDK9=Cyclin dependent kinase 9, SEC=Super elongation complex. Figure extracted from Hatch et al. (2016).

In order to confirm that *cmyc* is a paused gene and that after the lost of function of the PTEF-b components, the pausing stage is not released; ChIP-sequencing of RNA polymerase II on control and CDK9 morpholino (MO) injected embryos was performed. *X. tropicalis* embryos were injected at 2-cell stage with 100ng of CDK9 morpholino or control morpholino and fixed for ChIP-sequencing. This experiment was carried out at Radboud University under the supervision of Dr. Saartje Hontelez and the experiments were analysed with help from Dr. Saartje Hontelez, Dr. Simon van Heeringen and Prof. Gert Jan C. Veenstra. This work was published in Hatch et al. (2016), see Appendix 2.

The results of the RNA polymerase II ChiP-qPCR showed that after the *cdk9* inhibition, there is an increase in the amount of reads in both the *cmyc* promoter and the gene body compared with the control (Figure 43A). In the control situation, almost all the regions are below 10 fold change occupancy compared with a gene desert region whereas in the CDK9 morpholino, all the regions are above 10 fold change occupancy compared with the input sample even some promoter regions are enriched over 20 times. This is also confirmed by the visualization of the ChIP data on the genome browser (Figure 43B). The overall effect of *cdk9* inhibition was also analysed by plotting the rpkm (reads per kilobase of transcript per million of mapped reads) on the gene body vs the promoter region, defined as TSS +/- 150bp. When comparing control and *cdk9* MO plots, there is a shift in the y-axis indicating that there is an accumulation of reads in the promoter region. The *cmyc* gene is one of the genes more sensitive to the transcription pausing release, while the rpkm on its gene body is increased from 5 to 10, the rpkm on its TSS goes from 11 to 50 (Figure 43C). The pausing index values for all annotated genes were determined by calculating rpkm TSS +/-150bp divided by the rpkm on the gene body. The histogram shows that there is an overall increase in the pausing index mainly due to the increase in promoter occupancy (Figure 43D).

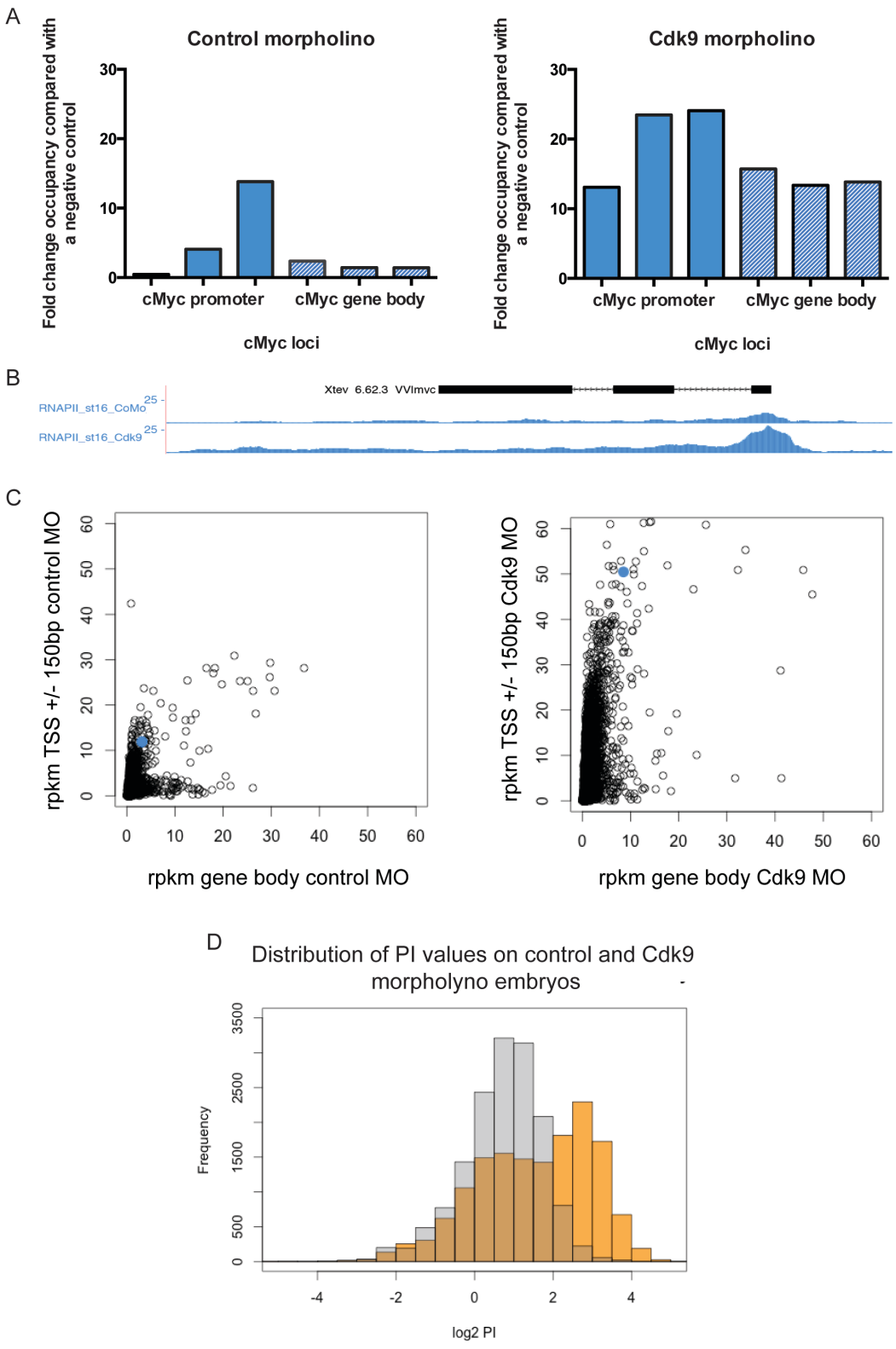
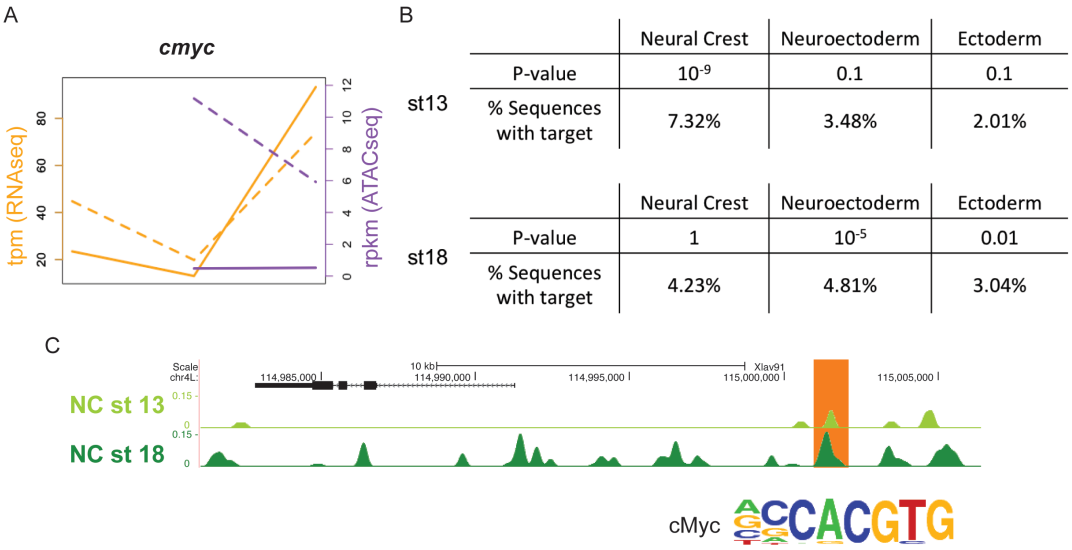


Figure 43: *cmyc* is a paused gene and Cdk9 morpholino causes *cmyc* to become more paused during development. (A) RNAPII ChIP-qPCR results showing fold change of accupancy compared with a negative primer design in a gene desert in the input sample. Solid blue bars show enrichment in promoter region and stripped blue bars show enrichment in the gene body. (B) RNAPII profile of the *cmyc* gene at stage 16 from embryos injected with control or Cdk9s morpholino visualized in the UCSC genome browser. C) ChIP-

sequencing scatter plots of rpkm on the gene body versus the transcriptional start site (TSS) in Control (left panel) and Cdk9 morpholino-injected embryos (right panel). The *cmyc* gene is shown by the blue spot. (D) Histogram plots showing the distribution of the Pausing Index values (rpkm TSS +/-150 bp divided by the rpkm on the gene body) in log2 for control (grey) and Cdk9 morpholino-injected embryos (orange).

Although the previous results indicate that *cmyc* is a poised gene in *Xenopus*, the experiments were performed in whole embryo due to the starting material limitation. The ATAC-sequencing data combined with the RNA-sequencing data on animal caps allow the visualization of the poised stage (Figure 44A). Initially there is a high accessibility in the promoter region, reflected by ATAC-sequencing peak at the promoter. Active promoters and are characterized by being accessible and occupied by RNAPII. Although the promoter is active, the transcription is not effective, as reflect the RNA-sequencing data. After releasing the pausing stage, there is an increase in the transcript production coupled with a decrease in the promoter accessibility, as the transcription machinery is being processive and not recruited in the promoter. The analysis of *cmyc.*S shows that there is an increase in the transcription level at stage 18 compared with stage 13, there are 20 tpm at stage 13 but 74 tpm at stage 18. Conversely, there is a decrease in the promoter accessibility; the number of reads decreases from 11 to 6 rpkm. However, the L homologue does not show similar results suggesting different functionalization of the L and S homologue. This confirms that the poised stage observed in whole embryo in part occurs in NC-induced animal cap cells (Figure 44A).

The presence of *cmyc*-binding sites in intergenic enhancers was also evaluated (Figure 44B). *cmyc* binding-sites are significantly present in 7.32% of the NC enhancers at stage 13 but not in neuroectoderm and ectoderm enhancers. However, at stage 18 *cmyc*-binding sites are not present at neural crest enhancers but they are in neuroectoderm and ectoderm enhancers. This serves as a mechanism by which *cmyc* activates NC specifiers such as *sox10*. *cmyc*-binding sites are not present in stage 13 or stage 18 promoters. As demonstrated in Hatch et al. (2016) the loss of *cmyc*, through inhibiting *cdk9*, leads to defects in NC specifiers expression such as *sox10* and *foxd3*. Moreover, a potencial enhancer of *sox10* contains *cmyc*-binding sites (Figure 44C). This provides a link between transcriptional elongation of *cmyc* and expression of NC derivatives to allow correct formation of NC structures.



	Neural Crest	Neuroectoderm	Ectoderm
P-value	10^{-9}	0.1	0.1
% Sequences with target	7.32%	3.48%	2.01%

	Neural Crest	Neuroectoderm	Ectoderm
P-value	1	10^{-5}	0.01
% Sequences with target	4.23%	4.81%	3.04%

Figure 44: ATAC-sequencing data provides a link between transcriptional elongation of *cmyc* and expression of NC derivatives. (A) Transcription expression on *cmyc*.S and L and accessibility in *cmyc*.S and L promoter at stage 9, 13 and 18. Solid line indicate L form dashed line indicate S form. Tpm= transcripts per million, rpkm=reads per kilobase transcript per million of mapped reads. (B) Table comparing *cmyc*-binding sites across different stages and samples. The p-value and percentage of sequences with *cmyc* motif is reported. (C) Visualization of ATAC-sequencing data at stage 13 and 18 in the UCSC genome browser at *sox10* location. The orange rectangle shows a potential active enhancer in which *cmyc* binding motif is present.

3.4.11. Selection of neural crest enhancers

A detailed analysis of five enhancers was carried out because of their potential role in activating expression of crucial NC genes i.e. *cmyc*, *snai2*, *foxd3*, *id3* and *sox10*. This selection was done taking into account that those enhancers were NC exclusive, close to a neural crest related gene and that the motifs present on those peaks were other NC genes whose interaction has been previously described.

A potential *cmyc* cis-regulatory element (Figure 45) is found 1880bp downstream the *cmyc*.L gene transcription terminal site. It is 2Kb long containing two potential enhancers. It contains motifs for *cmyc*, *zic*, *ets1*, *tfap2a*, *nmyc*, *tcf* and *smad*. The relationship of some of these transcription factors with *cmyc* has been reported. As a NC specifier, it is downstream of the Wnt and BMP signaling pathways, this could be through the Wnt effector Tcf and the BMP effector Smad motifs, both present in this enhancer (Simões-Costa & Bronner 2015). *zic1*, together with *pax3*, are involved in activating NC specifiers such as *snail2*, *foxd3*, *twist1*.

Although the direct regulation of *cmyc* via *zic1* has not been established yet, this would not be surprising. *ets1* and *nmyc* have also been related to NC formation. However, their regulation of *cmyc* has not been demonstrated (Wang et al. 2015; Wakamatsu et al. 1997). The appearance of a *cmyc*-binding motif suggests that *cmyc* could autoregulate itself. Finally, *tfap2a* has been suggested as a key element for NC enhancer activation (Rada-Iglesias et al. 2012). Moreover, in the motif analysis performed on human NC enhancers, motifs such as Ets, Zic, Tcf3, Smad4, Ebox (cMyc) appeared significantly upregulated (Rada-Iglesias et al. 2012). There is not much known about the genes upstream of *cmyc*. However, all this evidences suggest that this *cmyc* enhancer located 1880bp downstream of the *cmyc.L* transcription terminal site could be involved in regulating c*myc* expression in neural crest cells. A potential enhancer is found intronically. However, it does not contain any NC-transcription factor motif, suggesting it is not involved in the NC expression of *cmyc*.

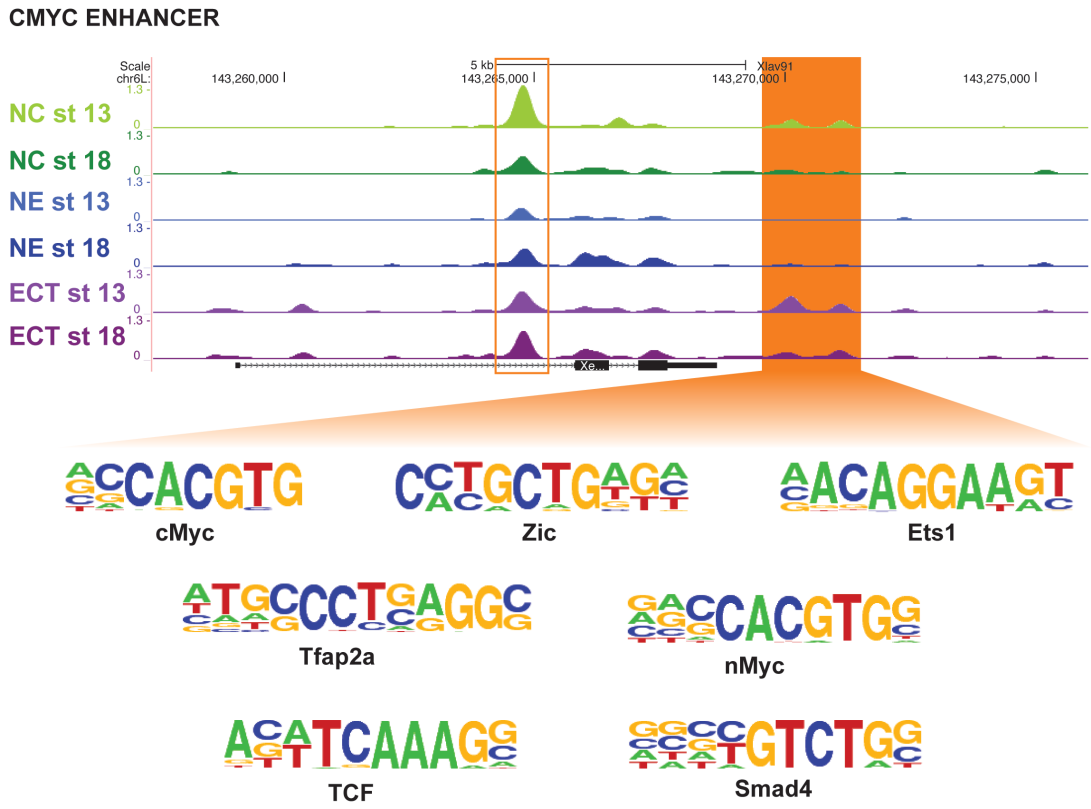


Figure 45: Potential *cmyc* enhancer. Genomic location and motifs present in the *cmyc.L* enhancer are shown. A second potential enhancer is framed in orange.

The second enhancer is the *snai2.L* enhancer located 7kb downstream of the TSS and 900bp long. It contains *tfap2a*, *sox10*, *foxd3*, *cmyb* and *smad4* binding motifs and this region is mainly open in NC cell type at both stages and ectoderm

stage 18 (Figure 46). The transcription factors, whose motifs are found in the enhancers, have been previously related with *snai2*. The presence of *tfap2a* in this potential CRE reinforces the potentially important role of this enhancer in NC cells (Rada-Iglesias et al. 2012). The expression of *snai2* is regulated by *sox10* (Honoré et al. 2003). *foxd3* and *snai2* have largely overlapping expression pattern. However overexpression of *foxd3* induces ectopic expression of *snai2* indicating that *foxd3* regulates *snai2* expression (Sasai et al. 2001). In neuroblastoma cells, a tumor derived from neural crest cells, *cmyb* activates *snai2* expression (Tanno et al. 2010). Finally, *snai2* is induced by BMP stimulation (Larman et al. 2012). This all supports the potential role of this enhancer in *snai2*.L expression therefore allowing correct formation of NC derivatives.

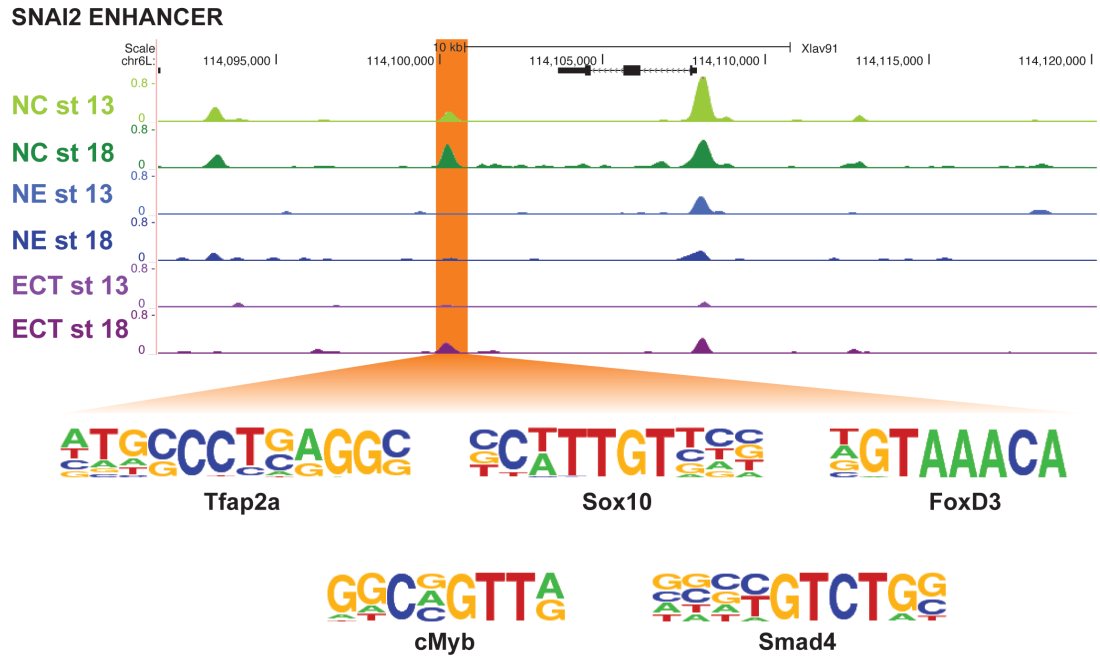


Figure 46: Potential *snai2* enhancer. Genomic location and motifs present in the *snai2*.L enhancer are shown.

A potential *foxd3* enhancer is located 1kb upstream of *foxd3*.S and it is 960bp long. This enhancer contains *sox5*, *foxd3*, *pax7*, *six1*, *smad4* and *meis* binding sites, which have all been reported as *foxd3* regulators (Figure 47). Both *six1* and *sox5* genes inhibit *foxd3* expression; the inhibition of *sox5* and *six1* expand the expression domain of *foxd3* (Perez-Alcala et al. 2004; Brugmann et al. 2004). Contrary, this expression of *foxd3* is highly reduced when *meis3* is inhibited (Gutkovich et al. 2010). *pax3/7* directly regulates *foxd3* expression through binding to two regulatory regions that drive expression in chick cranial and trunk NC cells (Simões-Costa et al. 2012). Finally, it has been suggested that *foxd3* negatively

auto-regulates itself as its expression is reduced after exogenous protein injection in *Xenopus* embryos. Also, the early expression of *foxd3* is activated by BMP signaling pathway (Pohl & Knöchel 2001). The appearance of these motif binding sites support the hypothesis that this potential *foxd3*.S enhancer could be involved in NC formation. Upstream of the TSS there is another potential enhancer only present at stage 13 (Figure 47). However, the motif analysis does not reveal any NC-transcription factor binding sites.

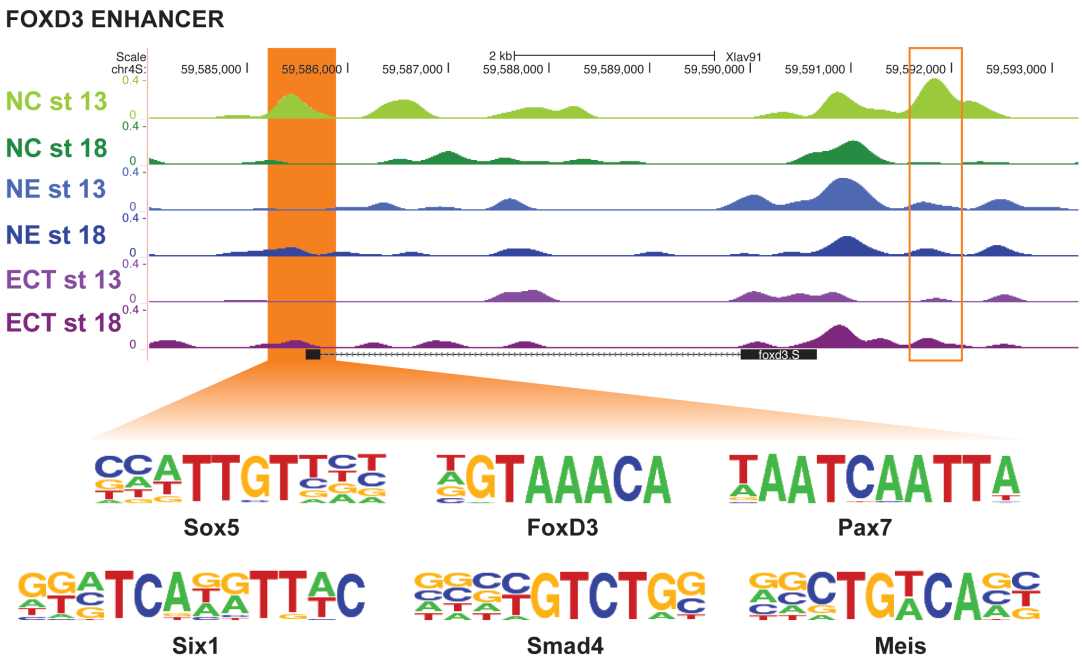


Figure 47: Potential *foxd3* enhancer. Genomic location and motifs present in the *foxd3.S* enhancer are shown. A second potential enhancer is framed in orange.

The potential *id3* enhancer is found in chromosome 2S, 1.9kb upstream the *id3* TSS (Figure 48). This CRE is 980bp and contains binding sites for the two major *id3* regulator, *cmyc* and BMP signaling. BMPs induce Id proteins thus inhibiting neural differentiation (Ying et al. 2003). Also, *id3* is a *cmyc* effector and its inhibition affects the formation of NC cells (Light et al. 2005). The appearance of these two motifs supports the potential involvement of this enhancer in *id3* expression.

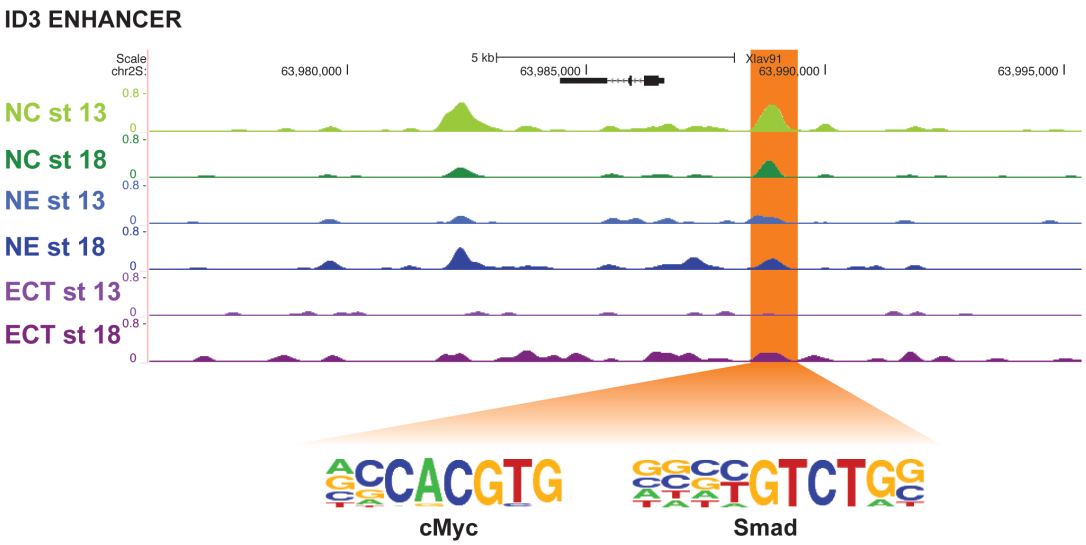


Figure 48: Potential *id3* enhancer. Genomic location and motifs present in the *id3*.S enhancer are shown.

Finally, a potential *sox10* CRE is located 9.6kb upstream the *sox10*.L TSS (Figure 49). It is 5,176bp long and it contains three potential enhancers. This regulatory region contains, among others transcription factors-binding sites, *pax7, sox5*, *sox10*, *tfap2a*, *smad*, *tcf and sox9*, all of them previously described as sox10 regulators. During mouse embryogenesis, a *sox10* enhancer contains transcription factor binding sites for *sox*, *tcf*, *pax* and *tfap2a* (Werner et al. 2007). Also, *tfap2a* is found in human NC enhancers as a key factor for NC enhancer activation (Rada-Iglesias et al. 2012). *sox9* activates *sox10* expression through a cranial enhancer Sox10E2 in chick embryos (Betancur et al. 2010b). BMP and Wnt maintain *sox10* expression in cells, in the absence of both, these cells die or dedifferentiate (Kléber et al. 2005). The appearance of these motifs that relates NC genes with the potential enhancer reinforces the potential role of this CRE in *sox10* regulation. The epigenetic profile upstream the sox10 TSS suggest that there might be a super-enhancer. There are several enhancers concatenated within 12.5Kb and there are transcription factor-binding sites for Oct4, Sox2 and Nanog, both characteristics of super-enhancers (Pott & Lieb 2014). The presence of mediator, Med1, cannot be detected through Homer motif analysis. It would be required to perform ChIP-qPCR to detect if this potential super-enhancer contains mediator-binding sites.

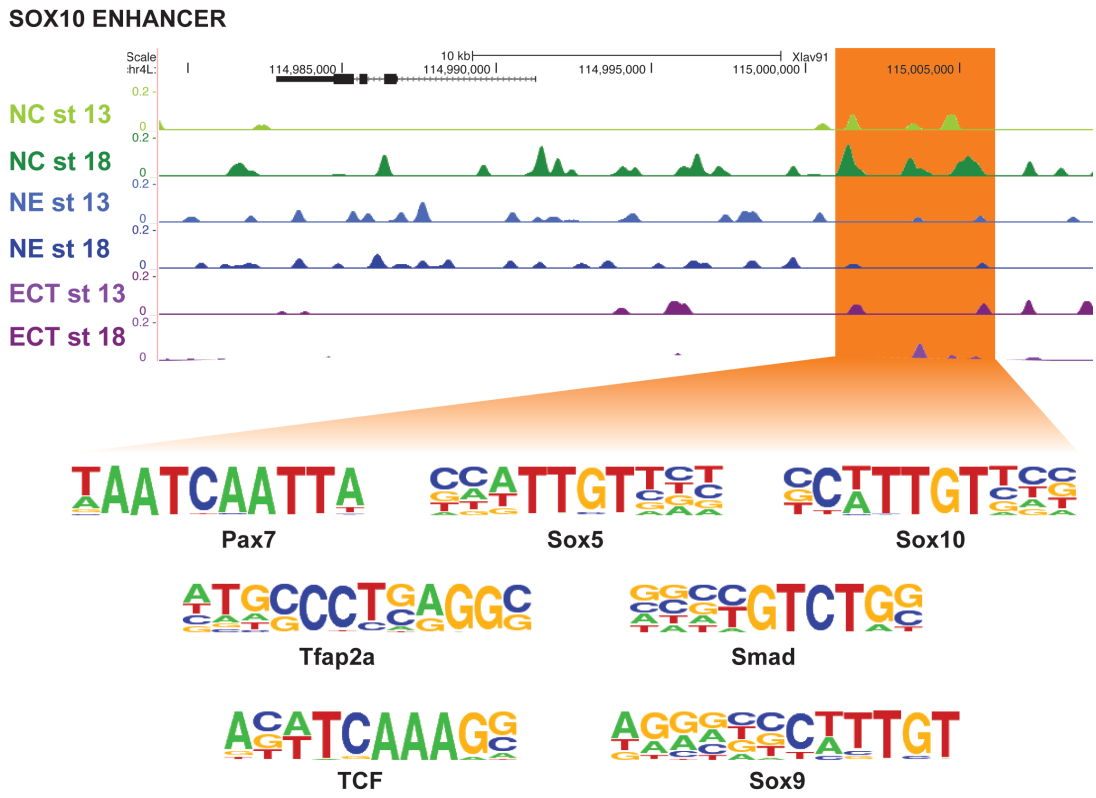


Figure 49: Potential *sox10* enhancer. Genomic location and motifs present in the *sox10*.L enhancer are shown.

3.4.12. Discussion

The aim of the study was to characterize the epigenome during the NC formation, comparing stage 13, where this cell type is specified, and stage 18, where the differentiation of NC cells occurs. For that, NC-specific enhancers at both stages and the potential transcription factors involved in NC enhancer activation have been identified. Also enhancers, which could drive expression of key NC genes, have been described.

The main problems encountered were related with the protocol and its analysis. The most important point to take into account when performing ATAC-sequencing is the number of cells used in the reaction, a large number of cells can inhibit the enzyme, the optimum being between 10-50,000 cells. Because ATAC-sequencing is a quite recent technique there is no clear protocol and consensus about the data analysis. However, Encode constantly updates the recommended pipeline to follow (Encode 2017).

Looking at the results, the ATAC-sequencing samples are quite variable; an average of 20% of the peaks of a sample (minimum of 7% and maximum of 32%) are common between the three replicates and taken into account for the analysis. This is due to the technique *per se* and the analysis. Because Tn5 is an enzyme that cuts unprotected regions of the genome and this is quite dynamic, the peaks are not in a fixed position and when adopting the premise that 50% of the peak should overlap, a lot of them are discarded. Actually, the neuroectoderm stage 18 sample, which consist of 2 replicates, contains twice more peaks than the rest of the samples, which consist of three replicates. This suggests that the increase of replicates reduces dramatically the number of common peaks and therefore it is necessary to analyse several replicates to draw conclusions about the ATAC-sequencing data. The distribution of peaks at stage 18 will likely vary if another replicate is incorporated at the neuroectoderm sample.

At stage 13, about 30% of the open regions are common between the three cell types (NC, neuroectoderm and ectoderm) and could be regulating housekeeping genes whereas 30% are cell type-specific (Figure 35). Not surprisingly, neural crest and neuroectoderm share more than 50% of the regulatory elements in both stages revealing the great similarities between these two cell types or a mixed population inside the animal cap. This could be due to the fact that both cell types are induced with the same amount of *noggin* and the neural crest cap could contain some neuroectodermal cells or that, being generated in the same environment, the transcriptome and epigenome at early stages are similar. It would be necessary then to perform single cell analysis to analyse the NC-induced animal cap. Contrary to this, the ectoderm sample does not share as many regulatory elements with the NC or neuroectoderm samples.

The analysis of the neuroectoderm and ectoderm peaks confirms that the experiment was well performed and gives more confidence when drawing conclusions about the NC data. The neuroectoderm peaks contain motifs for transcription factors such as *nkx6.1*, *otx2* and *sox3*, all involved in anterior nervous system formation (Ma et al. 2013; Gammill & Sive 2001; Penzel et al. 1997). The ectodermal peaks, in contrast, contain *gata2/3*, *six1* binding sites all involved in the formation of ectodermal derivatives (Read et al. 1998; Riddiford & Schlosser 2016).

It was noticed that at stage 18, there are not many NC specific open regions when compared with stage 13 (i.e.1,212 vs 6,025). Also the comparison of NC open regions between the two stages reveals that there are only 85 common regions.

Looking at the distribution of peaks in the stage 18 sample, there is a small percentage that relates with open or active promoters. All this suggests that the chromatin becomes progressively more closed as differentiation proceeds. This is also reflected by the great proportion of pluripotency transcription factors motifs found at stage 13 but not as present at stage 18.

When analysing the enhancer dynamics of the NC enhancers, most of them are present in either one stage or in both stages. Not many enhancers being activated or closed are observed. As mentioned before, these two stages, specification and differentiation are quite different at epigenetic level, as shown by ATAC-sequencing. However, specification and differentiation are not similar at transcription level, as revealed by the RNA-sequencing. In order to distinguish the transition between the two stages, ATAC-sequencing at intermediate stages should be performed.

The open regions in the different samples were distributed according to their genomic location in a gene, including exons and introns; a promoter, defined as TSS +/- 200bp; a proximal enhancer, arbitrarily defined as peaks located 2kb upstream; and distal enhancers located farther than 2kb from the TSS and not inside the gene coordinates. About 25% of the open regions are located inside the genes. The *X. laevis* genome annotation does not allow the distinction between exons and introns, therefore potential enhancers located in the first intron are not considered, thus underestimating the number of enhancers. Also, reads across the gene body are observed which could be related to active transcription taking place in the gene. Although the distinction of proximal and distal enhancer is made, both of them are taken into account for the motif analysis. However, whereas the target of the distal enhancers is not easy to predict, proximal enhancers are quite likely to interact with the closest promoter. On average, 50% of the specific peaks are potential enhancers. However, not all open regions are enhancers, they can also be insulators or silencers (Kolovos et al. 2012). It is then advisable to complement ATAC-sequencing data with other epigenetic marks such as p300 or H3K27ac, to confirm which peaks are truly active enhancers.

To determine which genes are involved in activating those enhancers and therefore add information to the gene regulatory network established for the *Xenopus* neural crest, it is required to analyse the motifs overrepresented in the active enhancers. The motifs present at stage 13 and stage 18 are mainly NC related genes. At stage 13, the most relevant transcription factors according to the

expression level, the abundance of motifs and the p-value are Meis (p-value=10^{-16}) and Zic (p-value=10^{-29}); at stage 18 Zic (p-value=10^{-13}), Meis (p-value=10^{-7}) and Sox10 (p-value=10^{-13}). Zic and Meis proteins therefore seem important for activating NC enhancers during the different stages of NC formation whereas Sox10 is more relevant for activating late stages of NC formation. Although it is not possible to determine which Zic and Meis proteins are involved in NC formation, based on the literature *zic1* and *meis3* are the more plausible option (Plouhinec et al. 2014; Gutkovich et al. 2010). It would be of interest to perform ChIP-sequencing for these transcription factors to determine which enhancers those genes are activating and if there is a core of transcription factors involved in activating NC enhancers.

A similar study performed in human derived NC cells suggested Tfap2a and Nr2f1/2 as the key elements in the NC enhancer activation (Rada-Iglesias et al. 2012). Moreover, it was found that the inhibition of *nr2f1* in *Xenopus* triggers craniofacial defects. Surprisingly, the Nrf2/1 motif is not found on the NC enhancers that have been identified and moreover, *nr2f1* expression is very low in the NC-induced animal caps according the RNA-sequencing data generated. On the other hand, in that study 10% of the active enhancers contain Tfap2a motifs; this percentage being similar to the 8.6% that found in our data at stage 18. However, it was surprising not to find the *tfap2a* motif present at stage 13. Perhaps, *tfap2a* is more relevant for late stages of NC cells formation and/or cranial NC cells. For that it would be required to perform a similar study but at later stages on NC development.

Our results support the hypothesis that multipotent cells at the blastula stage remain multipotent to become NC cells (Buitrago-Delgado et al. 2015). The presence of motifs for pluripotency factors such as Oct4, Sox2 and Nanog homologs in NC enhancers, both at transcript and motif level, is quite significant at stage 13. Oct4 and Nanog homologs do not seem as relevant during the differentiation of NC cells indicating that there is a decrease on the multipotency abilities as differentiation occurs. Sox2 homologs, however, are maintained at stage 18 potentially because the function of this gene is required later on during brain formation (Zhang & Cui 2014).

The combination of ChIP-sequencing and ATAC-sequencing revealed that the poised stage release of *cmyc* allows the processive transcription of this gene, which in turn can activate downstream genes through binding to their enhancers. However, *cmyc* is not the pioneer factor for NC genes as it can only bind to open

chromatin that contains active histone modifications (Iwafuchi-Doi & Zaret 2014). cMyc cannot bind to the NC promoters, as they do not contain *cmyc* binding sites. In order to confirm this hypothesis and identify the *cmyc* downstream targets, it will be necessary to carry out ChIP-sequencing using a *cmyc* antibody. *cmyc* motif analysis indicates that it is required during NC specification but it is not necessary later during differentiation. However, at stage 18, *cmyc* is present in neuroectoderm enhancers as it is necessary for proliferation and self-renewal of neural stem cells (Kerosuo et al. 2008).

Another objective of the project was to identify enhancers that drive the expression of key NC genes, conferring tissue specificity to the linked genes. For this purpose enhancers close to NC genes, i.e *cmyc*, *foxd3*, *snai2*, *id3* and *sox10*, which are mainly open in NC samples and containing NC related motifs, were selected. Most regulatory regions are not evolutionarily constrained (Bulger & Groudine 2010). The conservation in the Tibetean frog, zebrafish, chick and mice of the candidate enhancers has been analysed using UCSC liftOver. The selection of enhancers is only partially conserved in frogs, both *X. tropicalis* and the Tibetean frog *Nanorana parkeri*, suggesting they are conserved only on the Anura order. Among the genes selected, only a chick *sox10* and *foxD3* enhancers have been previously described (Simões-Costa et al. 2012; Betancur et al. 2010b). The *sox10* and *foxd3* characterized enhancers are not conserved in *Xenopus* although they contain respectively *sox9* and *pax7* binding sites. The fact that the binding sites are conserved but it is not the broader DNA sequencing that contains them, is a known fact in evolutionary studies (Heintzman et al. 2007; Bogdanović et al. 2012; Neph et al. 2012). It would then be interesting to check if those enhancers can drive not only expression in *Xenopus* embryos but in distally related organisms such as zebrafish and chick.

The validation of the enhancer activity needs to be tested, for that I-SceI meganuclease method is going to be carried out in *X. laevis* (Ogino et al. 2006). Briefly, a construct with the enhancer to test, a minimal promoter and a reporter gene such as *gfp* flanked by I-SceI sites needs to be generated. The construct is digested and injected at 1-cell stage *X. laevis* embryo. The embryo is grown until the appropriate stage and the fluorescence is checked under the microscope. The method has been optimized using a gata2 enhancer and a pax3 promoter provided by the Grainger lab. An efficiency of 10% of full transgenic embryos is obtained, similar to the reported by Ogino et al. (2006).

Also, the contribution of the transcription factors whose motifs are found in the enhancers should be validated. To answer that, first, the different transcription factors should be inhibited, by morpholino injection, and the level of transcription of the target gene should be analysed. Second, a construct with the enhancer containing a mutation in a specific binding site should be injected and tested for the ability to drive *gfp* expression.

To sum up, this study has allowed the characterization of *Xenopus* NC enhancers and the identification of enhancers that could drive expression of *snai2*, *foxD3*, *id3*, *sox10* and *cmyc*.

CHAPTER IV:
EPIGENOMIC REGULATION OF THE ANTEROPOSTERIOR REGENERATION IN PLANARIANS

4.1. Rationale

Regeneration and tissue renewal are essential processes of adult animals that must be tightly controlled, since its dysfunction leads to common illnesses as cancer. The study of regeneration is also important to understand and deliver treatments to regenerate organs injured or damaged due to heart attacks, liver damage or nerve injury. During regeneration the proper amount of new cells must be produced and their fate is to be properly re-specified and they are to be integrated into an already developed body plan. The aim of this chapter was to study the genomic regulation of the regeneration of anterior and posterior poles of planarians, which define head vs tail identity. For this, data obtained from high throughput genomic technologies such as ATAC and ChIP-sequencing have been performed. This project has been carried out in collaboration with Emili Saló's lab in the UB (Spain) and José Luis Gómez-Skarmeta's lab in the CABD (Spain).

The first objective was to set up the ATAC-sequencing technique in planarians, as the identification of CRE in *S. mediterranea* is a new field. For that, ATAC-sequencing in 0h and 48h regenerating (0hR and 48hR) blastemas was carried out to check the feasibility of the project, through the identification of specific transcription factors motifs in open regions known to be required for anterior or posterior specification. The current planarian genome contains 15,046 scaffolds and an average scaffold length of 77,506bp and thus the mappability of the reads could have been difficult. Also, as a distantly related organism, how successful the prediction of motifs present in the enhancers would be unclear.

The second objective, once the technique was validated, was to identify anterior and posterior specific active enhancers at 12 hours of regeneration (12hR) and analyse the motifs present in those. For this aim, ATAC-sequencing of anterior and posterior 12hR blastemas has been carried out. To select only open regions related with polarity ATAC-sequencing on 12hR anterior blastemas of *notum* RNAi organism, which forms a tail instead of a head, and 12hR posterior blastema of *wnt1* RNAi organism, which forms a head instead of a tail, has been carried out. To narrow the relevant active regulatory elements in such regions, H3K27ac ChIP-sequencing on 12hR blastemas has also been performed. Potential transcription factors that bind to the active enhancers will be identified as well as preliminary attempts to link them with their target genes. With this approach, we aim to bring to light more elements to reconstruct the GRN underlying the anterior and posterior regeneration.

4.2. Analysis of the cis-regulatory elements at 0hR and 48hR

As the ATAC-sequencing protocol was previously optimised (see section 3.4.1), the same protocol was followed but changing the initial cell dissociation procedure, according to a specific protocol for planarians. For that, organisms were first cut in the postpharyngeal region. For the 0 hour regeneration (0hR) sample, 20 anterior and posterior regions were pooled. For the 48 hours regeneration (48hR) samples, organisms were left for 48h and then 20 anterior and 20 posterior blastemas were dissected. Figure 50 illustrates the dissections carried out. The bioanalyser profiles of the three samples showed a peak around 300bp confirming the samples were good for sequencing (Figure 50B).

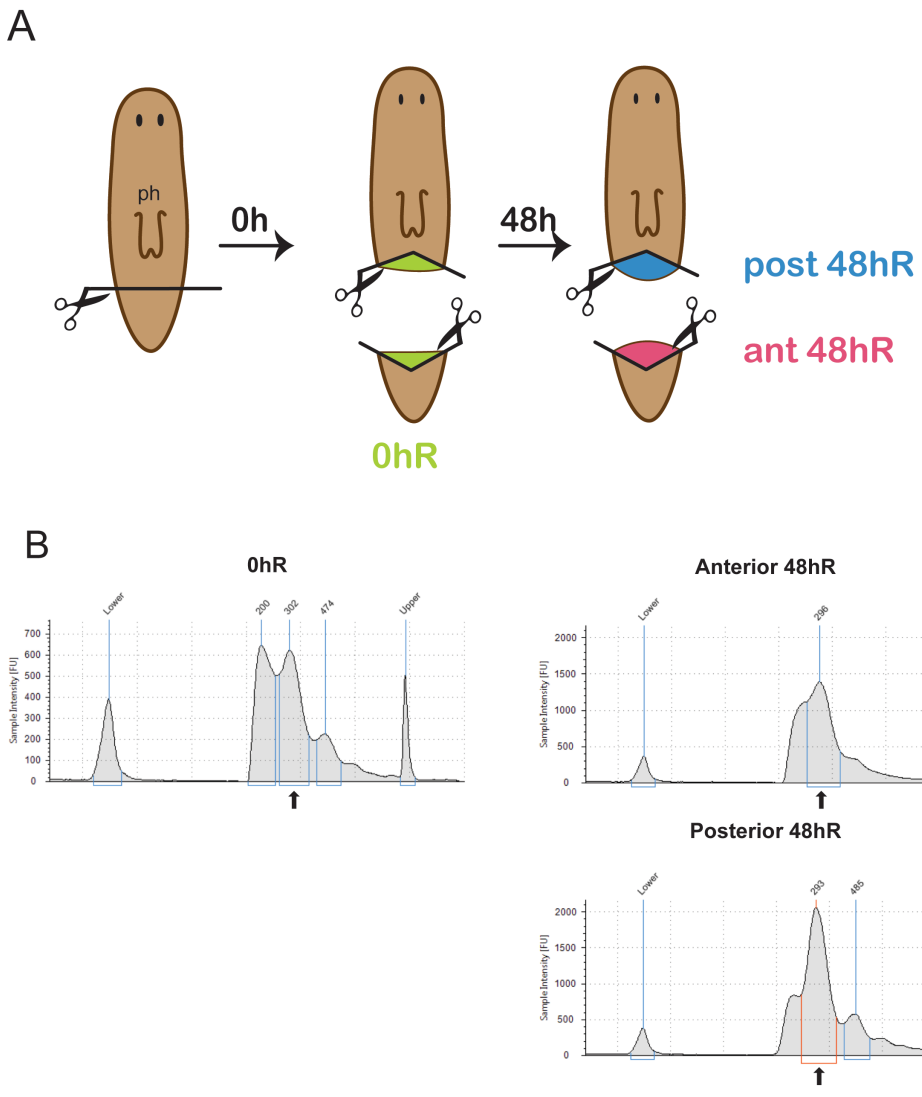


Figure 50: ATAC-sequencing samples at 0hR and 48hR. (A) Schematic representation of the samples. Ph=pharynx. (B) Bioanalyser profile of the samples. The peak at approximately 300bp is indicated by an arrow.

The samples were analysed in a similar way to the *Xenopus* samples, but in this case there was just one replicate per sample, as it was just a first attempt to check if the bioinformatics analysis afterwards gave relevant results. Despite the planarian genome not being well-assembled -containing 15,046 scaffolds and an average scaffold length of 77,506bp- the reads aligned properly. Between 59% and 67% of the reads aligned to the genome. This is low compared to the 80% threshold, which is the percentage of mapped reads advised by Encode. However, this threshold has been established taking into account well annotated genomes such as human or *Drosophila*. Although the reads reached only 67%, and because of the specificities of the planarian genome, it was decided to continue with the analysis.

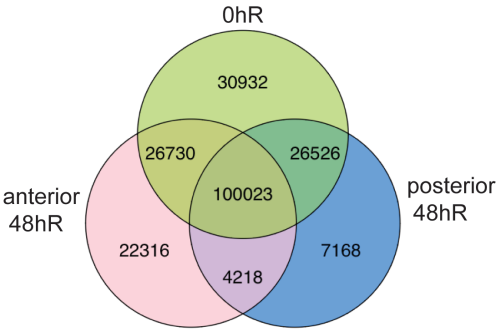
The comparison between samples reveals that 20% of 0hR peaks, 15% of the anterior 48hR and 5% of the posterior 48hR are specific for each sample (Figure 51A). Most of the peaks are common between the three samples. The anterior and posterior 48hR share just a few open regions, indicating that the change of fate has already occurred. When analysing the location of the open regions, more than 70% are potential proximal or distal enhancers (Figure 51B). However, because the genome is not completely assembled, these values might be overrepresented.

Because in planarians there is currently no possibility to generate transgenic organisms to test enhancers, the most interesting thing to look at is the transcription factors that activate those enhancers. For that, motif analysis on the different intergenic regions was carried out (Figure 51C-E). The complete list of top motifs, ranked according to the p-value and percentage of open regions with each motifs are displayed in Appendix 1 Tables v-vii. At 0hR the intergenic open regions are enriched in *gata4*, *hnf4* and *foxA*, all transcription factors associated with the digestive system of the planarians (Figure 51C). Considering the location of the cuts, these all make sense. *gata4* is expressed in the digestive track of planarians and it is required for the homeostasis and differentiation of the intestinal structures (Flores et al. 2016). The hepatocyte nuclear factor 4, *hnf4*, is an intestine associated gene and it is required for the gut differentiation (Scimone et al. 2014). Finally, *foxA* is expressed in the pharynx and surrounding tissue and it is necessary for the pharynx regrowth (Adler et al. 2014). The motifs present at anterior 48hR intergenic regions are involved in nervous system regeneration, i.e. *prep*, *neuroD*-1 and *lhx2* (Figure 51D). Moreover, these transcription factors are upregulated at 48hR after anterior amputation (Scimone et al. 2014). *prep* is required to

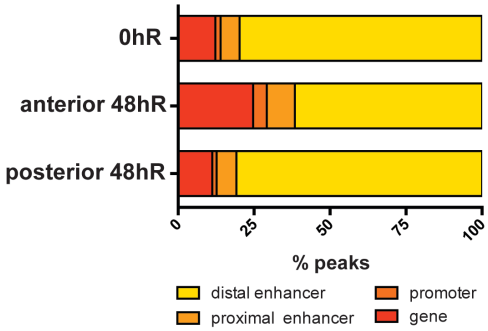
differentiate and pattern anterior structures as its inhibition produces defects in the regeneration of brain and eyes (Felix & Aboobaker 2010). *neuroD-1* and *lhx2* are nervous system genes involved in neural specification (Cowles et al. 2013; Scimone et al. 2014). The posterior 48hR intergenic regions are enriched in *islet*, *hoxD* and *egr-1* motifs (Figure 51E). *islet* is one of the key genes that determines posterior identity as its inhibition causes a 'tail-less' phenotype. In the posterior blastema *islet* maintains *wnt1* expression (Hayashi et al. 2011). *hoxD* is the only posterior hox gene found in planarians and it is expressed in the posterior blastema of regenerating planarians (Iglesias et al. 2008). Finally, *egr-1* is a marker of zeta-neoblasta which gives rise to the epidermal derivatives (Van Wolfswinkel et al. 2014).

The results demonstrate that despite the incompleteness of the genome, it is possible to obtain relevant information performing ATAC-sequencing in planarians and analysing the motifs present in the potential enhancers. However, these motifs are involved with late stages of regeneration, once the polarity is established. In order to analyse how the polarity is determined, it is required to perform a similar study but at earlier stages of the planarian regeneration, at 12hR, the time point previous to the establishment of the anterior and posterior poles.

A: Peak comparation between samples



B: Location of the exclusive peaks



C: Relevant motifs present in 0hR enhaners

Motif	TF	P-value	targets with motifs	Reference
	*gata4*	10^{-13}	10.00%	Flores et al (2016)
	*hnf4*	10^{-11}	1.59%	Scimone et al. (2014)
	*foxa*	10^{-7}	10.13%	Adler 2014

D: Relevant motifs present in anterior 48hR enhaners

Motif	TF	P-value	targets with motifs	Reference
	*prep*	10^{-23}	10.87%	Felix et al.(2010)
	*neuroD-1*	10^{-10}	3.14%	Cowles et al. (2013)
	*lhx2*	10^{-59}	24.88%	Scimone et al. (2014)

E: Relevant motifs present in posterior 48hR enhaners

Motif	TF	P-value	targets with motifs	Reference
	*islet*	10^{-4}	16.00%	Hayashi et al. (2011)
	*hoxD*	10^{-2}	14.53%	Iglesias et al. (2008)
	*egr-1*	10^{-8}	1.70%	Wolfswinkel et al. (2014)

Figure 51: Analysis of the 0hR and 48hR ATAC-sequencing samples. (A) Venn diagram showing the distribution of peaks in the three samples. (B) Bar plot showing the location of the exclusive peaks. The categories are genes, promoters defined as TSS+/- 200bp, proximal enhancer located 2Kb upstream the promoter and distal enhancers. (C) Motifs overrepresented in the 0hR exclusive intergenic peaks. (D) Motifs overrepresented in the anterior 48hR exclusive intergenic peaks. (E) Motifs overrepresented in the anterior 48hR exclusive intergenic peaks. The motifs, transcription factor associated, p-value, %of peaks with the motifs, and a bibliographic reference for the transcription factors are shown.

4.3. Preliminary analysis of the cis-regulatory elements at 12hR

To identify the enhancers and transcription factors involved in the formation of the anterior and posterior poles, ATAC-sequencing at 12hR has been performed (Figure 52). Posterior and anterior blastemas at 12hR were collected for ATAC-sequencing. To select only regulatory elements involved in the polarity, anterior blastemas of *notum* RNAi animals (which regenerate a tail instead of a head) and posterior blastemas of *wnt1* RNAi planarians (which regenerate a head instead of a tail) were also collected for ATAC-sequencing. Three replicates per sample were done. All the samples showed a peak around 300bp when analysing the fragment length profile using a bioanalyser (Figure 53).

The comparison between anterior and posterior blastemas, as well as with the corresponding RNAi sample, will allow the identification of cis-regulatory elements specifically involved in posterior and anterior regeneration. In order to obtain RNAi organisms, planarians are injected for three days with double stranded RNA specific for the gene, then cut pre and postpharyngeal, and after three days the injection procedure is repeated. The synthesis of the double stranded RNA and injections were carried out by Eudald Carreras at the University of Barcelona.

In order to identify, out of the open regions, the active enhancers, ChIP-sequencing with H3K27ac of wild-type blastemas at 12hR has been carried out. This technique requires around 2,000 blastemas at 12hR, those were collected by Eudald Carreras and the experiment has been performed at the Centro Andaluz de Biologia del Desarollo, Spain. Two replicates of anterior and posterior 12hR blastemas were obtained. The ChIP-qPCR validation carried out is not conclusive enough (Figure 54). The anterior samples show enrichment in the *notum* promoter compared with a gene dessert (Figure 54A). However, the posterior samples did not show enrichment for any *wnt1*, *follistatin* and *notum* promoter. It was expected to obtain *wnt1* or *follistatin* enrichment in the posterior samples (Figure 54B). This could be because the *follistatin* and *wnt1* primers have been designed upstream the TSS and might not be in the promoter regions, as they are not clearly defined. Also, the negative primer could be an open region specific for the posterior sample. It might have been useful to design two or three more negative primers. Despite this, because the anterior samples look good and there is no prior information about the enhancers and promoters activated during regeneration, it was decided to sequence the four samples.

All the ATAC-sequencing samples and ChIP-sequencing samples are currently being sequenced in the BGI Hong Kong.

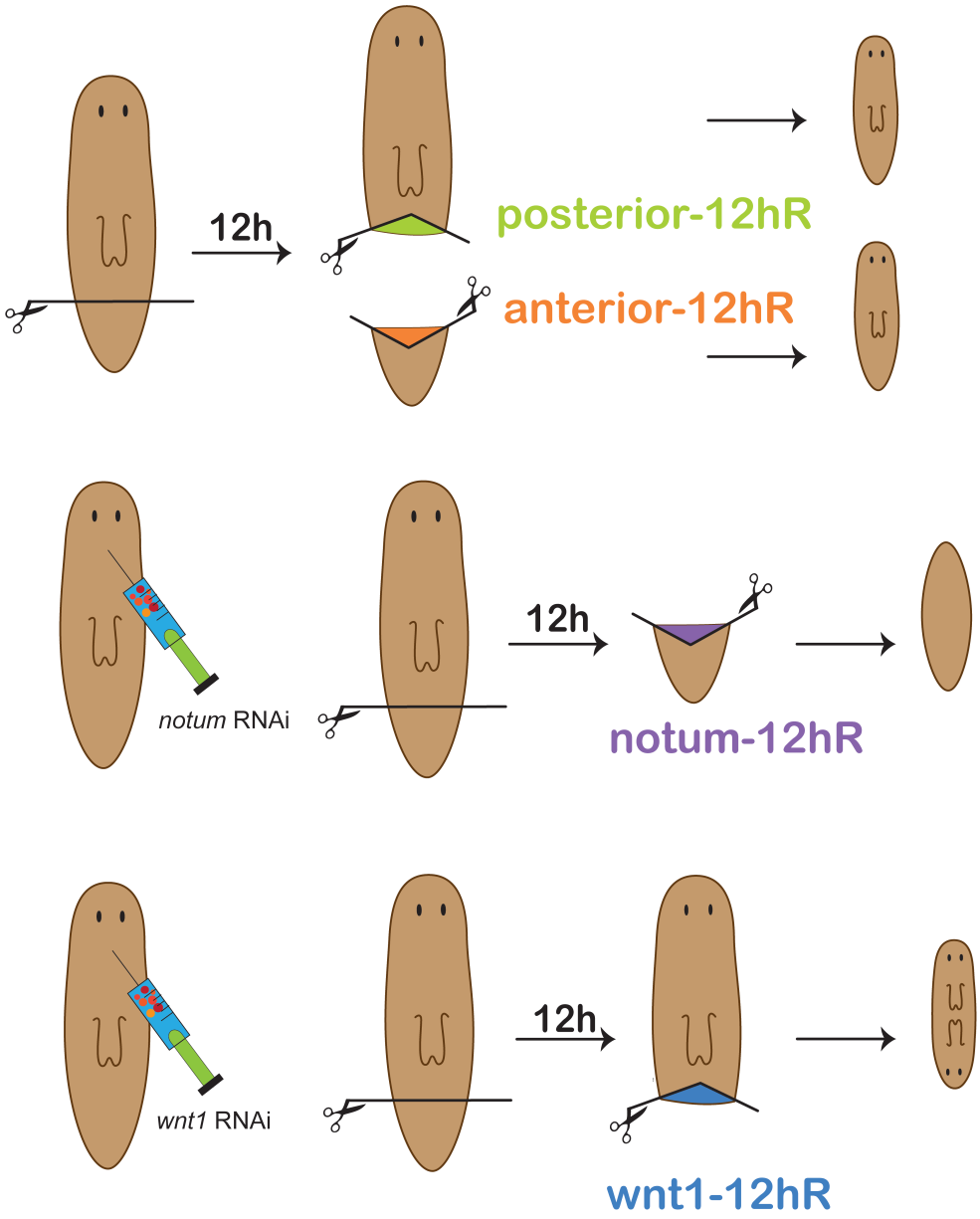


Figure 52: Schematic representation of the ATAC-sequencing samples at 12h regeneration. Posterior 12hR, anterior 12hR, notum 12hR and wnt1 12hR.

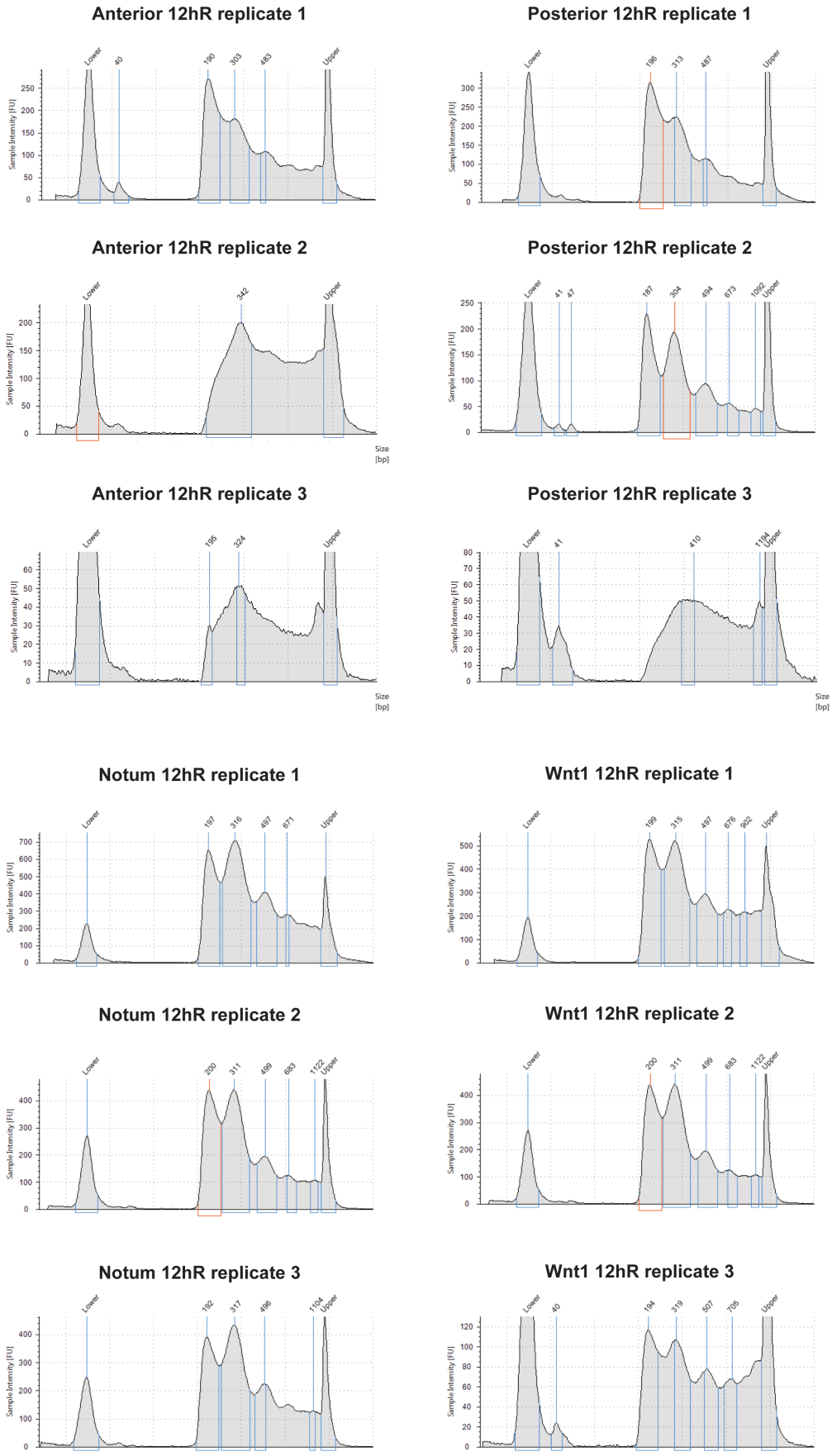


Figure 53: Bioanalyser profile of the ATAC-sequencing 12hR samples.

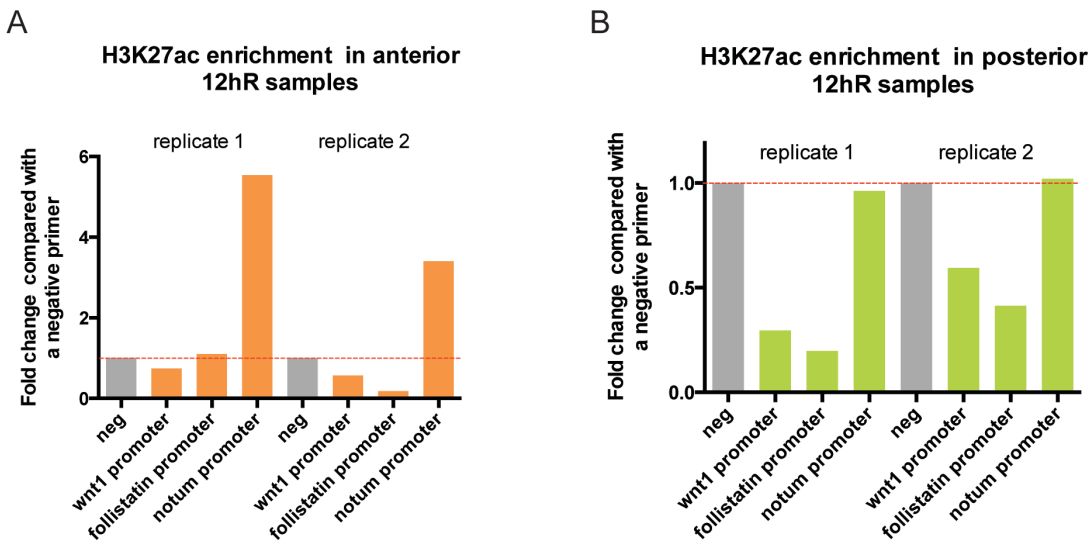


Figure 54: ChIP-qPCR validation. Enrichment of H3K27ac in *wnt1*, *follistatin* and *notum* promoter compared with a gene dessert location. (A) Anterior samples. (B) Posterior samples.

4.4. Discussion

The aim of this study was firstly to determine the potential of the ATAC-sequencing technique in planarians and secondly to study the genomic regulation of the anterior and posterior poles, since they define head vs tail identity, through integrating ATAC-sequencing and ChIP-sequencing methodology.

The ATAC-sequencing study carried out at 0hR revealed transcription factor binding sites for genes related with digestive system. This is consistent with the dissected area in which ATAC-sequencing was performed. On the other side, ATAC-sequencing carried out on anterior and posterior blastemas at 48hR revealed motifs for nervous system and posterior regeneration, respectively. This confirms that despite the not well-assembled genome and the motif data being from distantly related organisms, it is possible to identify enhancers and potential transcription factors that can bind to them. This opens a window of high-throughput methodologies to answer relevant questions in the regenerative field in planarians.

Knowing the potential of this technique, it was decided to study the genomic regulation of the anterior and posterior poles. The study at 48hR revealed late effectors of the anterior and posterior regeneration. To discover the regulatory

elements involved in the specification of the poles, it was required to perform a similar study at an early time point, i.e. 12hR. The anterior pole is currently defined by the expression of *notum* (a Wnt inhibitor), and the posterior pole is defined by the expression of *wnt1*. To narrow the number of potential enhancers to those exclusively related with pole identity, ATAC-sequencing on *notum* and *wnt1* RNAi organisms was also performed. Furthermore, as the functional validation of the enhancers in the planarian animal model is not possible, ChIP-sequencing with H3K27ac, which reveals active enhancers, has also been performed. The integration of both data will provide a list of cis-regulatory elements (promoters and enhancers), and potential transcription factors bound to those elements, specifically involved in the anterior and posterior specification. The full list of candidate elements will be categorized according to genomic position, signalling pathway and biological process. Special attention will be paid to the elements found to be related with the Wnt signalling pathway.

CHAPTER V:
GENERAL DISCUSSION

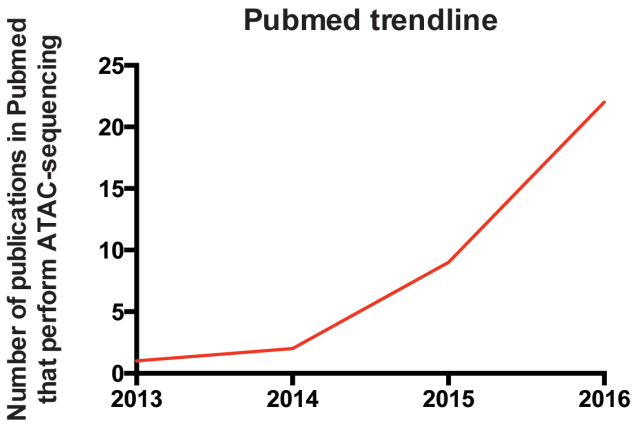
5.1. About ATAC-sequencing

Cis-regulatory elements in the genome are known to control animal development. These elements contain multiple target sites for transcription factors and collectively control the expression of developmental gene regulatory networks. Therefore, CRE analysis provides a method to study transcription factors interactions.

ATAC-sequencing is a relatively new high throughput technique that thanks to the Tn5 transposase reveals open regions in the genome. The advantages of this technique are: (1) the low amount of starting material required, (2) the low complexity of the protocol, and (3) the no need for any antibody steps, enabling it to be performed in any organism. Disadvantages of the method include (1) the high variability in the samples, (2) the rudimentary bioinformatics analysis and (3) the necessity to test the enhancer functionality to determine if they are active enhancers or to combine the ATAC-sequencing data with p300 or H3K27ac ChIP-sequencing data, which specifically identify active enhancers. Despite its draw backs, since the protocol was published in 2013, the number of papers containing ATAC-sequencing, according to Pubmed, has been rising dramatically (Figure 55A). The method has been used to identify active enhancers in for example, *Drosophila* dorsoventral patterning, mouse rod photoreceptors or human myeloid differentiation and many other tissues and cells, examples shown in Figure 5B.

In this study, ATAC-sequencing has been carried out in order to identify cis-regulatory elements in the genome to answer two biological questions: (1) how neural crest genes are regulated in *Xenopus* embryos and (2) how the anteroposterior polarity of planarians is re-established during regeneration.

A



B

Studies in which ATAC-sequencing has been carried out to identify enhancers	Reference
cMyb knockdown vs control in K562 cell line	(Fuglerud et al., 2017)
Mammalian preimplantation embryos	(Wu et al., 2016)
Human Myeloid Differentiation	(Ramirez et al., 2017)
$CD8^+$ T cells	(Mognol et al., 2017)
Mouse Rod Photoreceptors	(Hughes et al., 2017)
Early mouse embryo	(Simon et al., 2017)
Human pancreatic islets	(Varshney et al., 2017)
Adult mouse visual cortex	(Gray et al., 2017)
Drosophila dorsoventral patterning	(Koenecke et al., 2016)

Figure 55: Impact of ATAC-sequencing in the literature. (A) Pubmed trendline of the ATAC-sequencing citation. (B) Main publications that carries out ATAC-sequencing to identify enhancers.

5.2. About *Xenopus* neural crest formation

Chapter III focused on the investigation of the regulatory elements involved in NC formation. For this, several approaches were tested. Firstly, a chemical screen with specific epigenetic inhibitors was carried out on *Xenopus* embryos. Out of the 66 tested compounds, distinct phenotypes were identified such as developmental defects, edema formation or pigment defects. Among these, the pigmentation phenotype was of more interest, as the pigment cells derive from the neural crest. All the compounds associated with pigment cell phenotypes were histone acetyltransferases and histone deacetylases, suggesting that histone acetylation plays a role in pigmentation. However, when testing the expression of

neural crest markers (i.e. *zic1* and *foxd3*) by WISH no alteration was detected; this suggested that the HAT and HDAC compounds rather than affecting NC specification affected later stages such as melanocyte formation or melanogenesis. Several hypotheses have been given for the fact of not finding compounds affecting NC specification, the main one being that the drugs should have been added earlier. Thus, in order to test this hypothesis, it would be needed to repeat the chemical screen adding the drugs during gastrulation (i.e. between stages 9 and 13). Also, in order to determine if the reason for the lack of pigmentation was the melanocyte formation or the melagonesesis process, it would be required to visualise if the melanocytes are properly formed through for example WISH of *mitf* or *tyrosinase*, markers for melanocyte cells.

As an alternative approach, it was decided to identify regulatory elements on the DNA and transcription factors associated, responsible for activating NC genes. To address this, ATAC-sequencing was carried out on NC-induced animal caps. Firstly, transcriptome analyses were performed on different stages (i.e. stage 9, 13 and 18) of organoid-like NC-induced animal caps in order to check that these caps resembled NC cells present in the embryo. This transcriptome analysis showed that the expression of NC markers followed the expected temporal pattern of expression according to the literature (Betancur et al. 2010a). With this study, it could be possible to not only look at already described genes but also look for potential new NC genes. As the model was proven to work, it was then possible to study another level of genomic regulation analysing the cis-regulatory elements of such animal caps. However, as it may not be like the *in vivo* situation, it would be necessary to confirm the obtained results at whole embryo level.

ATAC-sequencing as described by Buenrostro et al. (2013) was first carried out on numerous animal caps, but the fragment length profile analysed by bioanalyser did not show an enrichment in the regulatory region fraction (i.e. 300 bp); thus indicating that proper tagmentation was not occurring. A number of optimization steps were introduced in the protocol the main ones being: (1) the tissue was dissociated with trypsin/collagenase, and (2) cells were counted to not exceed a maximum number of 50,000 cells (i.e. to not overload the Tn5 transposase enzyme).

In order to make sure that specific NC open regions were identified via ATAC-sequencing, neuroectoderm and ectoderm-induced animal caps were also used. For each cell type two stages were sampled (i.e. stage 13 and stage 18),

which were particularly interesting for the study as this is when NC specification is originated and differentiation is starting. Three replicates were obtained for each stage and cell type. When doing the bioinformatics analysis three restrictions were applied to constrain, as much as possible, the potential NC enhancers. That is: (1) due to the genome dynamics and the activity of the transposase generating variability among the replicates, for each of the cell type and stage, only those open regions which were present in the three replicates were considered for further bioinformatics analyses; (2) as the open regions revealed by the transposase may relate not only to enhancers but also to promoters and gene locations, only peaks in the intergenic regions, which is the main location for enhancers, were taken into account and (3) the peaks in the intergenic regions of each of the cell types were compared and only those present in NC were considered potential enhancers. Thus, the analysis of the samples allowed us to identify 4528 NC potential specific enhancers at stage 13 and 852 at stage 18 (Figure 38).

When comparing the potential enhancers at the two different stages, it was noticed: (1) that not many of them were present in both, and (2) that not many enhancer transitions, from more to less active or accessible or vice versa, were detected. With this in mind, it would seem appropriate to carry out the same ATAC-sequencing protocol at intermediate stages.

The transcription factor motifs present in the selected potential enhancers were identified using Homer (Heinz et al. 2010). This analysis revealed: (1) most of the enhancers contained potential transcription factors binding sites which have been shown to be specific for NC genes, supporting that the NC regulatory network is complex; (2) Zic and Meis motifs are particularly enriched in both stages and Sox10 only at stage 18. Also, at stage 13, motifs for pluripotency factors are found, which support the hypothesis by Buitrago-Delgado et al. (2015) that indicates that NC cells at stage 13 retain the pluripotency capabilities.

Finally, by combining Homer observations and visualizing the ATAC-sequencing tracks in the UCSC genome browser, five candidate enhancers were identified, based on (1) their proximity to key NC genes, (2) the enhancer specificity, and (3) their NC transcription factors binding sites. These enhancers could potentially drive expression of *cmyc*, *snai2*, *foxd3*, *id3* and *sox10*. These enhancers are only conserved in *X. tropicalis* and *Nanorana parkeri* indicating that these are anura-specific. In order to further confirm the functionality of these enhancers, I-SceI meganuclease transgenesis needs to be carried out.

Our work suggests that this type of study allows the obtainment of large amounts of information about the epigenome of a single cell type and the regulation of specific genes in a cell specific context, which is especially interesting in organisms such as *Xenopus*, which do not have a great level of conservation with other organisms.

5.3. About planarian anteroposterior polarity

Chapter IV focused on the investigation of the regulatory elements involved in the formation of the planarian anterior and posterior poles, which define head vs tail identity. For this, the ATAC-sequencing approach was carried out. It was first required to test how successful this technique would be in planarians taking into account (1) that the published genome is not well-assembled, -currently it is formed by 15,046 scaffolds, with an average scaffold length of 77,506bp-, and (2) that it was uncertain if the transcription factor motifs identified in higher organisms would be conserved in planarians.

Thus, in order to test the feasibility of the ATAC-sequencing in planarians, a first attempt was carried out with three types of regenerating blastemas: at 0hR, anterior 48hR and posterior 48hR, and only one replicate of each. These time points were chosen according to published literature, as they have been widely studied. The bioinformatics analysis of these samples was done with the same methodology as developed for the *Xenopus* analysis, that is: only the exclusive intergenic peaks were selected. With this, potential transcription factor binding sites were obtained. No further analysis (i.e. selection of candidate enhancers) was done, as their functional validation in planarians is not possible. Results of this preliminary study showed that the 0hR blastemas contain motifs for digestive system related genes whereas the anterior and posterior 48hR blastemas contain motifs for nervous system and posterior patterning related genes, respectively. This observation, which is in line with all the published literature, proved that the ATAC-sequencing protocol and informatics analysis, as followed, could be successfully used in planarians.

ATAC-sequencing was then applied to planarians in order to unravel specification of anterior and posterior poles. For this, an earlier regeneration time point was required (i.e. 12hR). Moreover, to select only potential enhancers involved in anterior and posterior identity, ATAC-sequencing of *notum* RNAi organisms, which form a tail instead of a head, and *wnt1* RNAi, which forms a head instead of a tail, was also incorporated in the analysis. As established previously,

three replicates per condition were obtained. The combination of the anterior and posterior samples taking into account the *notum* and *wnt1* RNAi, will generate a list of potential enhancers involved in pole identity. To select from those only the active enhancers, the data will be combined with ChIP-sequencing results for a marker of active enhancer, H3K27ac, performed on anterior and posterior 12hR blastemas. All the ATAC-sequencing and ChIP-sequencing samples have been obtained and they are currently being sequenced. As a result of the analysis, a list of potential transcription factors involved in the anterior vs posterior polarity will be obtained. The transcription factors will be validated through RNAi methodology and characterization of the phenotype.

5.4. Future work

This research has allowed us to firstly standardize a protocol (both experimental and bioinformatic) for ATAC-sequencing, which has been applied to two different species (*Xenopus laevis* and *Schmidtea mediterranea*), and secondly to identify potential enhancers and/or their associated transcription factors in *Xenopus* and planarians. However, these results are only the first step in the characterisation of the genome regulatory regions in the context of *Xenopus* NC and planarian regeneration. Thus, in order to fully understand these processes more research needs to be done.

For instance, in order to complete the *Xenopus* study it is required to test the functionality of the enhancers through I-SceI meganuclease transgenesis. For those positive hits, the contribution to the transcription factors whose motifs are found in the enhancers should be validated. To test that (1) the different transcription factors should be inhibited, by morpholino injection, and the level of transcription of the target gene should be analysed and (2) a construct with the enhancer containing a mutation in a specific binding site, should be introduced either by DNA injection or transgenesis and tested the ability to drive *gfp* expression. Ultimately, a transgenic line for a specific NC enhancer could be generated which would facilitate the study of NC cells. For example, cell sorting based techniques to isolate specific populations of cells possibly for single cell sequencing or chemical screen testing the fluorescence after compounds treatment could be carried out. The data generated can be also used to analyse any gene of interest to identify potential enhancers, not only NC but also neuroectoderm and ectoderm genes. Finally, the relevance of the NC transcription factors motifs found in most of the enhancers, i.e, Zic and Meis, as a key element for NC enhancer activation should be tested. For

that, these transcription factors alone or in combination could be used to induce NC in animal caps and the comparison of the epigenetic landscape with the obtained with *noggin* and *Wnt1* induction would reveal the percentage of enhancers activated due to the action of Zic and Meis, and their potential role as a pioneer factors.

With respect to the planarian project, the ATAC-sequencing and ChIP-sequencing data needs to be processed and the list of potential transcription factors involved in specification of anterior and posterior polarity needs to be generated. Those transcription factors can be validated by RNAi methodology and analysing the phenotype in the regeneration context. Furthermore, it would be of interest to perform a similar analysis at different time points to understand how the specification of the planarian poles occurs.

5.5. Where this research sits in a wider scientific context

This research has been designed in order to bring light into the broad genomic fields of the *Xenopus* NC formation and the planarian regeneration, and for that, the novel ATAC-sequencing methodology has been implemented. It is worth noticing that traditionally, gene studies both in *Xenopus* and planarian are based on placing a gene of interest inside a gene regulatory network, which only suggests indirect regulatory interactions. Instead, the strength of the high throughput methods such as ATAC or ChIP-sequencing, resides on the fact that the technique enables us to directly reveal interactions between enhancers and promoters, transcription factors and genes; thus giving a global picture of genomic interactions. Moreover, this study has also considered single cell populations to be used with the high throughput techniques. The advantage that we see in performing ATAC or ChIP-sequencing with cell populations (i.e. animal caps or dissected blastemas) is that the direct interactions revealed are less prone to background noise, allowing us to draw more robust conclusions which can be further tested.

In the neural crest field it is known that organisms such as mice, chicken, zebrafish or humans have enhancers, which are not conserved in *Xenopus*. For instance, our five candidate *Xenopus* NC enhancers have not been found in any of these organisms. We think that, from an evolutionary point of view, it would be of interest to perform similar high throughput studies in the aforementioned organisms and to analyse the evolution of the NC regulatory regions. NC development has been extensively studied in *Xenopus*. Our results will be shared with the community and used to further understand the dynamics of NC development.

Lastly, regarding the planarian field, our research is one of the first studies to apply high throughput techniques for approaching the regeneration process. Only Duncan et al. (2015) have performed ChIP-sequencing in whole organisms planarians with the aim of identifying targets of two H3K4 methyltransferases. Hence, our research, which sets up a standardized protocol, could be a precedent for other studies aiming to use the same methodology. This is also the first time that planarian regeneration is approached with such a technique. Thus, these results, which take into account new elements (i.e, promoters and enhancers), would generate a lot of information about how the anteroposterior identity is defined. Also, as ATAC-sequencing has not been so much used in any other regenerative organisms, the identification of transcription factors and enhancers associated to regeneration, could be relevant not only for the planarian but for the regenerative field. For example, in Kang et al. (2016) a zebrafish regeneration enhancer can drive the expression of genes in injured mammalian tissue.

To sum up, the application of ATAC-sequencing is a relatively novel technique, which can generate robust information to build up better gene regulatory networks.

References

Abu-Elmagd, M., Garcia-Morales, C., Wheeler, G.N., 2006. Frizzled7 mediates canonical Wnt signaling in neural crest induction. *Developmental Biology*, 298(1), pp.285–298.

Acemel, R.D., Tena, J.J., Irastorza-Azcarate, I., Marlétaz, F., Gómez-Marín, C., De La Calle-Mustienes, E., Bertrand, S., Diaz, S.G., Aldea, D., Aury, J.M., Mangenot, S., Holland, P. W.H., Devos, D.P., Maeso, I., Escrivá, H., Gómez-Skarmeta, J. L. 2016 A single three-dimensional chromatin compartment in amphioxus indicates a stepwise evolution of vertebrate Hox bimodal regulation. *Nature Genetics*, 48(3), pp 336-341

Adler, C.E., Seidel, C.W., McKinney, S., Sanchez Alvarado, A., 2014. Selective amputation of the pharynx identifies a FoxA-dependent regeneration program in planaria. *eLife*, 3, pp.1-22.

Alfandari, D., Cousin, H., Gaultier, A., Smith, K., White, J.M., Darribère, T., DeSimone, D. W., 2001. Xenopus ADAM 13 is a metalloprotease required for cranial neural crest-cell migration. *Current Biology*, 11(12), pp.918–930.

Almuedo-Castillo, M., Crespo, X., Seebeck, F., Bartscherer, K., Salo, E., Adell, T., 2014. JNK Controls the Onset of Mitosis in Planarian Stem Cells and Triggers Apoptotic Cell Death Required for Regeneration and Remodeling. *PLoS Genetics*, 10(6), e1004400.

Almuedo-Castillo, M., Sureda-Gómez, M., Adell, T., 2012. Wnt signaling in planarians: New answers to old questions. *International Journal of Developmental Biology*, 56(1–3), pp.53–65.

Aoto, K., Sandell, L., Butler Tjaden, N.E., Yuen, K.C., Noack Watt, K.E., Black, B.L., 2015. Mef2c-F10N enhancer driven b-galactosidase (LacZ) and Cre recombinase mice facilitate analyses of gene function and lineage fate in neural crest cells. *Developmental Biology*, 402(1), pp.3–16.

Baguñà, J., Riutort, M., 2004. The dawn of bilaterian animals: the case of acoelomorph flatworms. *BioEssays*, 26(10), pp.1046-1057.

Bajpai, R. Chen, D., Rada-Iglesias, A., Zhang, J., Xiong, Y., Helms, J., Chang, C., Zhao, Y., Swigut, T., Wysocka, J., 2010. CHD7 cooperates with PBAF to control multipotent neural crest formation. *Nature*, 463(7283), pp.958–62.

Balasubramanyam, K., Swaminathan, V., Ranganathan, A., Kundu, T.K., 2003. Small molecule modulators of histone acetyltransferase p300. *J Biol Chem*, 278(21), pp.19134-19140.

Barembaum, M., Bronner, M.E., 2013. Identification and dissection of a key enhancer mediating cranial neural crest speci fi c expression of transcription factor, Ets-1. *Developmental Biology*, 382(2), pp.567–575.

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

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., Zhao, K., 2007. High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell*, 129(4), pp.823–837.

Benelkebir, H., Donlevy, A.M., Packham, G., Ganesan, A., 2011. Total synthesis and stereochemical assignment of burkholdac B, a depsipeptide HDAC inhibitor. *Org Lett*., 13(24) pp.6334-6337.

Berger, S.L., Kouzarides, T., Shiekhattar, R., Shilatifard, A., 2009. An operational definition of epigenetics. *Genes and Development*, 23, pp.781–783.

Betancur, P., Bronner-Fraser, M., Sauka-Spengler, T., 2010a. Assembling Neural Crest Regulatory Circuits into a Gene Regulatory Network. *Annual Review of Cell and Developmental Biology*, 26(1), pp.581–603.

Betancur, P., Bronner-Fraser, M., Sauka-Spengler, T., 2010b. Genomic code for Sox10 activation reveals a key regulatory enhancer for cranial neural crest. *PNAS*, 107(8), pp.3570–3575.

Betancur, P., Sauka-Spengler, T., Bronner, M., 2011. A Sox10 enhancer element common to the otic placode and neural crest is activated by tissue-specific paralogs. *Development*, 138(17), pp.3689–3698.

Blassberg, R.A., Felix, D.A., Tejada-Romero, B., Aboobaker, A.A., 2013. PBX/extradenticle is required to re-establish axial structures and polarity during planarian regeneration. *Development*, 140(4), pp.730–739.

Bogdanović, O. Fernandez-Miñán, A., Tena, J.J., De La Calle-Mustienes, E., Hidalgo, C., Van Kruysbergen, I., Van Heeringen, S., J. Veenstra, G.J., Gómez-Skarmeta, J.L., 2012. Dynamics of enhancer chromatin signatures mark the transition from pluripotency to cell specification during embryogenesis. *Genome Research*, 22(10), pp.2043–2053.

Bowers, E.M., Yan, G., Mukherjee, C., et al., 2010. Virtual ligand screening of the p300/CBP histone acetyltransferase: Identification of a selective small molecule inhibitor. *Chem Biol*, 17(5), pp-471-482.

Boyle, A.P., Davis, S., Shulha, H.P., Meltzer, P., Margulies, E.H., Weng, Z., Furey, T.S., Crawford, G.E., 2008. High-Resolution Mapping

and Characterization of Open Chromatin across the Genome. *Cell*, 132(2), pp.311–322.

Brugmann, S., Pandur, P.D., Kenyon, K.L., Pignoni, F., Moody, S., 2004. Six1 promotes a placodal fate within the lateral neurogenic ectoderm by functioning as both a transcriptional activator and repressor. *Development*, 131(23), pp.5871–5881.

Buenrostro, J.D., Wu, B., Chang, H.Y., Greenleaf, W.J., 2015. ATAC-seq: A method for assaying chromatin accessibility genome-wide. *Current Protocols in Molecular Biology*, 109, p.21-29.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., Greenleaf, W.J., 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature Methods*, 10(12), pp.1213–1218.

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

Bulger, M., Groudine, M., 2010. Enhancers: The abundance and function of regulatory sequences beyond promoters. *Developmental Biology*, 339(2), pp.250–257.

Cai, D.H., Vollberg, T.M., Hahn-Dantona, E., Quigley, J., Brauer, P.R., 2000. MMP-2 expression during early avian cardiac and neural crest morphogenesis. *Anatomical Record*, 259(2), pp.168–179.

Calo, E., Wysocka, J., 2013. Modification of Enhancer Chromatin: What, How, and Why? *Molecular Cell*, 49(5), pp.825–837.

Cano, A., Pérez-Moreno, M.A., Rodrigo, I., Locascio, A., Blanco, M.J., del Barrio, M.G., Portillo, F., Nieto, M.A., 2000. The transcription factor snail controls epithelial-mesenchymal transitions by repressing E-cadherin expression. *Nature cell biology*, 2(2), pp.76–83.

Caravaca, J.M., Donahue, G., Becker, J.S., He, X., Vinson, C., Zaret, K.S., 2013. Bookmarking by specific and nonspecific binding of FoxA1 pioneer factor to mitotic chromosomes. *Genes and Development*, 27(3), pp.251–260.

Cebrià, F., 2007. Regenerating the central nervous system: How easy for planarians! *Development Genes and Evolution*, 217(11–12), pp.733–748.

Chang, C., Hem, A., 1998. Neural Crest Induction by Xwnt7B in Xenopus. *Developmental Biology*,134, pp.129–134.

Chen, C.-C.G., Wang, I.E., Reddien, P.W., 2013. Pbx Is Required for Pole and Eye Regeneration in Planarians. *Development*, 140(4), pp.719–729.

Cheung, M., Chaboissier, M.C., Mynett, A., Hirst, E., Schedl, A., Briscoe, J., 2005. The transcriptional control of trunk neural crest induction, survival, and delamination. *Developmental Cell*, 8(2), pp.179–192.

Chimenti, F., Bizzarri, B., Maccioni, E., et al., 2009. A Novel Histone Acetyltransferase Inhibitor Modulating Gcn5 Network: Cyclopentylidene-[4-(4′-chlorophenyl)thiazol-2-yl)hydrazone. *JMedChem*, 52(2), pp:530-536.

Chou, J., Provot, S., Werb, Z., 2010. Gata3 in Development and Cancer Differentiation: Cells GATA Have It! *J Cell Physiol*, 222(1), pp.42–49.

Chu, Y.S., Eder, O., Thomas, W.A., Simcha, I., Pincet, F., Ben-Ze'ev, A., Perez, E., Thiery, J.P., Dufour, S., 2006. Prototypical type I E-cadherin and type II cadherin-7 mediate very distinct adhesiveness through their extracellular domains. *Journal of Biological Chemistry*, 281(5), pp.2901–2910.

Coffey, K., Blackburn, T.J., Cook, S., et al., 2012. Characterisation of a Tip60 Specific Inhibitor, NU9056, in Prostate Cancer. *PLoS One*, 7(10), e45539

Cowles, M.W., Brown, D.D.R., Nisperos, S.V., Stanley, B.N., Pearson, B.J., Zayas, R.M., 2013. Genome-wide analysis of the bHLH gene family in planarians identifies factors required for adult neurogenesis and neuronal regeneration. *Development*, 140(23), pp.4691–4702.

Cremer, T., Cremer, C., 2001. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nature reviews genetics*, 2, pp.292–301.

de Crozé, N., Maczkowiak, F., Monsoro-Burq, A.H., 2011. Reiterative AP2a activity controls sequential steps in the neural crest gene regulatory network. *PNAS*, 108(1), pp.155–60.

Dalgin, G., Goldman, D.C., Donley, N., Ahmed, R., Eide, C.A., Christian, J.L., 2007. GATA-2 functions downstream of BMPs and CaM KIV in ectodermal cells during primitive hematopoiesis. *Developmental Biology,* 310, pp.454–469.

Dawson, M.A., Kouzarides, T., 2012. Cancer epigenetics: From mechanism to therapy. *Cell*, 150(1), pp.12–27.

Deardorff, M.A., Tan, C., Saint-Jeannet, J.P., Klein, P.S., 2001. A role for frizzled 3 in neural crest development. *Development*, 128(19), pp.3655–3663.

Deaton, A.M., Bird, A.P., 2011. CpG islands and the regulation of transcription. *Genes and Development*, 25(10), pp.1010–1022.

Dekker, J., Heard, E., 2015. Structural and functional diversity of Topologically Associating Domains. *FEBS Letters*, 589(20), pp.2877–2884.

DeLaurier, A., Nakamura, Y., Braasch, I., Khanna, V., Kato, H., Wakitani, S., Postlethwait, J.H., Kimmel, C.B., 2012. Histone deacetylase-4 is required during

early cranial neural crest development for generation of the zebrafish palatal skeleton. *BMC Developmental Biology*, 12(1), p.16.

Dibner, C., Elias, S., Frank, D., 2001. XMeis3 protein activity is required for proper hindbrain patterning in Xenopus laevis embryos. *Development*, 128, pp.3415–3426.

Duncan, E.M., Chitsazan, A.D., Seidel, C.W., Sanchez Alvarado, A., 2015. Set1 and MLL1/2 Target Distinct Sets of Functionally Different Genomic Loci In Vivo. *Cell Reports,* 13, pp.2741–2755.

Echelard, Y., Vassileva, G., McMahon, a P., 1994. Cis-acting regulatory sequences governing Wnt-1 expression in the developing mouse CNS. *Development*, 120(8), pp.2213–2224.

ENCODE Consortium Analysis Working Group (2017). ATAC-seq Data Analysis Pipeline. Retrieved from https://www.encodeproject.org/documents/0eb389f9-d23d-4053-b25b-1e2826ee5a86/@@download/attachment/ATACpipelineV7.pdf

Eroglu, B., Wang, G., Tu, N., Sun, X., Mivechi, N.F., 2006. Critical role of Brg1 member of the SWI/SNF chromatin remodeling complex during neurogenesis and neural crest induction in zebrafish. *Developmental Dynamics*, 235(10), pp.2722–2735.

Felix, D.A., Aboobaker, A.A., 2010. The TALE class homeobox gene Smed-prep defines the anterior compartment for head regeneration. *PLoS Genetics*, 6(4), e1000915.

Flores, N.M., Oviedo, N.J., Sage, J., 2016. Essential role for the planarian intestinal GATA transcription factor in stem cells and regeneration. *Developmental Biology*, 418(1), pp.179–188.

Fuglerud, B.M., Lemma, R.B., Wanichawan, P., Sundaram, A.Y.M., Eskeland, R., Gabrielsen, O.S., 2017. A c-Myb mutant causes deregulated differentiation due to impaired histone binding and abrogated pioneer factor function. *Nucleic Acids Research*, (20), pp.1–16.

Fujita, K., Ogawa, R., Ito, K., 2016. CHD7, Oct3/4, Sox2, and Nanog control FoxD3 expression during mouse neural crest-derived stem cell formation. *FEBS Journal*, 283(20), pp.3791–3806.

Fullwood, M.J., Liu, M.H., Pan, Y.F., et al., 2009. An oestrogen-receptor-α-bound human chromatin interactome. *Nature*, 462(7269), pp.58–64.

Gajer, J.M., Furdas, S.D., Gründer, A., et al., 2015. Histone acetyltransferase inhibitors block neuroblastoma cell growth in vivo. *Oncogenesis*, 4(2), pp.1-10.

Gammill, L.S., Sive, H., 2001. otx2 expression in the ectoderm activates anterior neural determination and is required for Xenopus cement gland formation. *Developmental Biology*, 240, pp.223–236.

Garcia-Morales, C., Liu, C.H., Abu-Elmagd, M., Hajihosseini, M.K., Wheeler, G.N., 2009. Frizzled-10 promotes sensory neuron development in Xenopus embryos. *Developmental Biology*, 335(1), pp.143-155.

Garnett, A.T., Square, T.A., Medeiros, D.M., 2012. BMP, Wnt and FGF signals are integrated through evolutionarily conserved enhancers to achieve robust expression of Pax3 and Zic genes at the zebrafish neural plate border. *Development*, 139(22), pp.4220–4231.

Gaviño, M.A., Wenemoser, D., Wang, I.E., Reddien, P.W., 2013. Tissue absence initiates regeneration through Follistatin-mediated inhibition of Activin signaling. *eLife*, 2013(2), pp.1–13.

Gilbert, S.F., Barresi, M.J.F., 2016. *Developmental Biology* 11th edition. Sinauer Associates, (ed.), Sunderland, MA.

Giresi, P.G., Kim, J., Mcdaniell, R.M., Giresi, P.G., Kim, J., Mcdaniell, R.M., Iyer, V.R., Lieb, J.D., 2007. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome research*, 17, pp.877–885.

Gisselbrecht, S.S., Barrera, L.A., Porsch, M., et al., 2013. Highly parallel assays of tissue-specific enhancers in whole Drosophila embryos. *Nature Methods*, 10(8), pp.774–780.

González-Estévez, C., Felix, D.A., Smith, M.D., Paps, J., Morley, S.J., James, V., Sharp, T., Aboobaker, A.A., 2012. SMG-1 and mTORC1 act antagonistically to regulate response to injury and growth in planarians. *PLoS Genetics*, 8(3), e1002619.

González-Estévez, C., Felix, D.A., Aboobaker, A.A., Saló, E., 2007. Gtdap-1 promotes autophagy and is required for planarian remodeling during regeneration and starvation. *PNAS*, 104(33), pp.13373–13378.

Gray, L.T., Yao, Z., Nguyen, T.N., Kim, T.K., Zeng, H., Tasic, B., 2017. Layer-specific chromatin accessibility landscapes reveal regulatory networks in adult mouse visual cortex. *eLife*, 6, pp.1–30.

Guenther, C.A., Wang, Z., Li, E., Tran, M.C., Logan, C.Y., Nusse, R., Pantalena-Filho, L., Yang, G.P., Kingsley, D.M., 2015. A distinct regulatory region of the Bmp5 locus activates gene expression following adult bone fracture or soft tissue injury Catherine. *Bone*, 77, pp.31–41.

Gutkovich, Y.E., Wang, Z., Li, E., Tran, M.C., Logan, C.Y., Nusse, R., Pantalena-Filho, L., Yang, G.P., Kingsley, D.M., 2010. Xenopus Meis3 protein lies at a nexus downstream to Zic1 and Pax3 proteins, regulating multiple cell-fates during early nervous system development. *Developmental Biology*, 338(1), pp.50–62.

Haberland, M., Mokalled, M.H., Montgomery, R.L., Haberland, M., Mokalled, M.H., Montgomery, L., Olson, E.N., 2009. Epigenetic control of skull morphogenesis by histone deacetylase 8 service. *Genes and Development*, 23, pp.1625–1630.

Harland, R.M., Grainger, R.M., 2011. Xenopus research: Metamorphosed by genetics and genomics. *Trends in Genetics*, 27(12), pp.507–515.

Harris, R.E., Setiawan, L., Saul, J., Hariharan, I.K., 2016. Localized epigenetic silencing of a damage-activated WNT enhancer limits regeneration in mature Drosophila imaginal discs. *eLife*, 5, pp.1–28.

Hatch, V.L., Marín-Barba, M., Moxon, S., et al., 2016. The positive transcriptional elongation factor (P-TEFb) is required for neural crest specification. *Developmental Biology*, 416(2), pp.361–372.

Hayashi, T., Motoishi, M., Yazawa, S., Itomi, K., Tanegashima, C., Nishimura, O., Agata, K., Tarui, H., 2011. A LIM-homeobox gene is required for differentiation of Wnt-expressing cells at the posterior end of the planarian body. *Development*, 138(17), pp.3679–3688.

Hayashi, T., Asami, M., Higuchi, S., Shibata, N., Agata, K., 2006. Isolation of planarian X-ray-sensitive stem cells by fluorescence-activated cell sorting. *Development Growth and Differentiation*, 48(6), pp.371–380.

Heger, P., Marin, B., Bartkuhn, M., Schierenberg, E., Wiehe, T., 2012. The chromatin insulator CTCF and the emergence of metazoan diversity. *PNAS*, 109(43), pp.17507–17512.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., Glass, C.K., 2010. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Molecular Cell*, 38(4), pp.576–589.

Heintzman, N.D., Stuart, R.K., Hon, G., et al., 2007. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature Genetics*, 39(3), pp.311–318.

Hnisz, D., Weintraub, A.S., Day, D.S., et al., 2016. Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science*, 351(6280), pp.1454–1458.

Hoffman, M.M., Ernst, J., Wilder, S.P., et al., 2013. Integrative annotation of chromatin elements from ENCODE data. *Nucleic Acids Research*, 41(2), pp.827–841.

Hong, C.-S., Park, B.-Y., Saint-Jeannet, J.-P., 2008. Fgf8a induces neural crest indirectly through the activation of Wnt8 in the paraxial mesoderm. *Development*, 135, pp.3903–3910.

Honoré, S.M., Aybar, M.J., Mayor, R., 2003. Sox10 is required for the early development of the prospective neural crest in Xenopus embryos. *Developmental Biology*, 260(1), pp.79–96.

Hoppler, S., Wheeler, G.N., 2015. It's about time for neural crest. *Science*, 150(3811), pp.563–565.

Hu, N., Strobl-Mazzulla, P.H., Simoes-Costa, M., Sánchez-Vásquez, E., Bronner, M.E., 2014. DNA methyltransferase 3B regulates duration of neural crest production via repression of Sox10. *PNAS*, 111(50), pp.1–6.

Hughes, A.E.O., Enright, J.M., Myers, C.A., Shen, S.Q., Corbo, J.C., 2017. Cell Type-Specific Epigenomic Analysis Reveals a Uniquely Closed Chromatin Architecture in Mouse Rod Photoreceptors. *Scientific Reports*, 7, pp.1-16.

Iglesias, M., Almuedo-Castillo, M., Aboobaker, A.A., Saló, E., 2011. Early planarian brain regeneration is independent of blastema polarity mediated by the Wnt/β-catenin pathway. *Developmental Biology*, 358(1), pp.68–78.

Iglesias, M., Gomez-Skarmeta, J.L., Saló, E., Adell, T., 2008. Silencing of Smed-betacatenin1 generates radial-like hypercephalized planarians. *Development*, 135(7), pp.1215–1221.

Ignatius, M.S., Moose, H.E., El-Hodiri, H.M., Henion, P.D., 2008. Colgate/Hdac1 Repression of Foxd3 Expression Is Required To Permit Mitfa-Dependent Melanogenesis. *Developmental Biology*, 313(2), pp.568–583.

Ignatius, M.S., Unal Eroglu, A., Malireddy, S., Gallagher, G., Nambiar, R.M., Henion, P.D., 2013. Distinct Functional and Temporal Requirements for Zebrafish Hdac1 during Neural Crest-Derived Craniofacial and Peripheral Neuron Development. *PLoS ONE*, 8(5), pp.1–14.

Ikenouchi, J., Matsuda, M., Furuse, M., Tsukita, S., 2003. Regulation of tight junctions during the epithelium-mesenchyme transition: direct repression of the gene expression of claudins/occludin by Snail. *Journal of Cell Science*, 116, pp.1959–67.

Irie, N., Kuratani, S., 2011. Comparative transcriptome analysis reveals vertebrate phylotypic period during organogenesis. *Nature Communications*, 2, pp.248-254.

Irimia, M., Tena, J.J., Alexis, M.S., et al., 2012. Extensive conservation of ancient microsynteny across metazoans due to cis -regulatory constraints. *Genome research*, 2, pp.2356–2367.

Iwafuchi-Doi, M., Zaret, K.S., 2014. Pioneer transcription factors in cell fate specification. *Genes and Development*, 28, pp.2679–2692.

Kälin, R.E., Bänziger-Tobler, N.E., Detmar, M., Brändli, A.W., 2009. An in vivo chemical library screen in Xenopus tadpoles reveals novel pathways involved in angiogenesis and lymphangiogenesis. *Blood*, 114(5), pp.1110–1122.

Kang, J., Hu, J., Karra, R., et al., 2016. Modulation of tissue repair by regeneration enhancer elements. *Nature*, 532(7598), pp.201–206.

Kato, K., Orii, H., Watanabe, K., Agata, K., 1999. The role of dorsoventral interaction in the onset of planarian regeneration. *Development*, 126(5), pp.1031–1040.

Kerosuo, L., Piltti, K., Fox, H., Angers-Loustau, A., Hayry, V., Eilers, M., Sariola, H., Wartiovaara, K., 2008. Myc increases self-renewal in neural progenitor cells through Miz-1. *Journal of Cell Science*, 121(23), pp.3941–3950.

Kim, H., Kang, K., Ekram, M.B., Roh, T.Y., Kim, J., 2011. Aebp2 as an epigenetic regulator for neural crest cells. *PLoS ONE*, 6(9), e25174.

Kléber, M., Lee, H.Y., Wurdak, H., Buchstaller, J., Riccomagno, M.M., Ittner, L.M., Suter, U., Epstein, D.J., Sommer, L., 2005. Neural crest stem cell maintenance by combinatorial Wnt and BMP signaling. *Journal of Cell Biology*, 169(2), pp.309–320.

Koenecke, N., Johnston, J., Gaertner, B., Natarajan, M., Zeitlinger, J., 2016. Genome-wide identification of Drosophila dorso-ventral enhancers by differential histone acetylation analysis. *Genome Biology*, 17(1), pp.196-215.

Kolovos, P., Knoch, T.A., Grosveld, F.G., Cook, P.R., Papantonis, A., 2012. Enhancers and silencers: an integrated and simple model for their function. *Epigenetics and Chromatin*, 5(1), pp.1-8.

Kong, Y., Grimaldi, M., Curtin, E., Dougherty, M., Kaufman, C., White, R.M., 2014. Neural crest development and craniofacial morphogenesis is coordinated by nitric oxide and histone acetylation. *Chemistry and Biology*, 21(4), pp.488–501.

Kouzarides, T., 2007. Chromatin Modifications and Their Function. *Cell*, 128(4), pp.693–705.

Kuriyama, S., Lupo, G., Ohta, K., Ohnuma, S., Harris, W.A., Tanaka, H., 2006. Tsukushi controls ectodermal patterning and neural crest specification in Xenopus by direct regulation of BMP4 and X-delta-1 activity. *Development*, 133(1), pp.75–88.

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

Larman, B.W., Karolak, M.J., Lindner, V., Oxburgh, L., 2012. Distinct Bone Morphogenetic Proteins activate indistinguishable transcriptional responses in nephron epithelia including Notch target genes. *Cell signal*, 24(1), pp.257–264.

Lawrence, M., Daujat, S., Schneider, R., 2016. Lateral Thinking: How Histone Modifications Regulate Gene Expression. *Trends in Genetics*, 32(1), pp.42–56.

Leigh, N.R., Schupp, M.O., Li, K., Padmanabhan, V., Gastonguay, A., Wang, L., Chun, C.Z., Wilkinson, G.A., Ramchandran, R., 2013. Mmp17b Is Essential for Proper Neural Crest Cell Migration In Vivo. *PLoS ONE*, 8(10), pp.1–12.

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

Li, B., Carey, M., Workman, J.L., 2007. The Role of Chromatin during Transcription. *Cell*, 128(4), pp.707–719.

Li, Q., Yang, H., Zhong, T.P., 2015. Regeneration across metazoan phylogeny: Lessons from model organisms. *Journal of Genetics and Genomics*, 42(2), pp.57–70.

Liang, H.-L., Nien, C., Liu, H., Metzstein, M.M., Kirov, N., Rushlow, C., 2008. The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila. *Nature*, 456(7220), pp.400–403.

Light, W., Vernon, A.E., Lasorella, A., Lavarone, A., LaBonne, C., 2005. Xenopus Id3 is required downstream of Myc for the formation of multipotent neural crest progenitor cells. *Development*, 132(8), pp.1831–1841.

Lonfat, N., Duboule, D., 2015. Structure, function and evolution of topologically associating domains (TADs) at HOX loci. *FEBS Letters*, 589(20), pp.2869–2876.

Ma, P., Xia, Y., Ma, L., Zhao, S., Mao, B., 2013. Xenopus Nkx6.1 and Nkx6.2 are required for mid-hindbrain boundary development. *Development Genes and Evolution*, 223(4), pp.253–259.

Maeso, I., Irimia, M., Tena, J.J., Casares, F., Gomez-Skarmeta, J.L., 2013. Deep conservation of cis-regulatory elements in metazoans. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368(1632), pp.1-10.

Maeso, I., D Acemel, R., Gomez-Skarmeta, J.L., 2017. Regulatory landscapes in development and evolution. *Current Opinion in Genetics and Development*, 43, pp.17–22.

März, M., Seebeck, F., Bartscherer, K., 2013. A Pitx transcription factor controls the establishment and maintenance of the serotonergic lineage in planarians. *Development*, 140(22), pp.4499–4509.

Matsukawa, S., Miwata, K., Asashima, M., Michiue, T., 2015. The requirement of histone modification by PRDM12 and Kdm4a for the development of pre-placodal ectoderm and neural crest in Xenopus. *Developmental Biology*, 399(1), pp.164–176.

Mayor, R., Theveneau, E., 2013. The neural crest. *Development*, 140(11), pp.2247–2251.

Milewski, R.C., Chi, N.C., Li, J., Brown, C., Lu, M.M., Epstein, J.A., 2004. Identification of minimal enhancer elements sufficient for Pax3 expression in neural crest and implication of Tead2 as a regulator of Pax3. *Development*, 131(4), pp.829–837.

Miller, C.T., Maves, L., Kimmel, C.B., 2004. moz regulates Hox expression and pharyngeal segmental identity in zebrafish. *Development*, 131(10), pp.2443–2461.

Mogno, I., Kwasnieski, J.C., Cohen, B.A., 2013. Massively parallel synthetic promoter assays reveal the in vivo effects of binding site variants. *Genome Research*, 23, pp.1908–1915.

Mognol, G.P., Spreafico, R., Wong, V., Scott-Browne, J.P., Togher, S., Hoffmann, A., Hogan, P.G., Rao, A., Trifari, S., 2017. Exhaustion-associated regulatory regions in CD8 $^{+}$ tumor-infiltrating T cells. *PNAS*, pp.1-10.

Molinaro, A.M., Pearson, B.J., 2016. In silico lineage tracing through single cell transcriptomics identifies a neural stem cell population in planarians. *Genome Biology*, 17(1), p.87-104.

Monsoro-Burq, A.-H., Fletcher, R.B., Harland, R.M., 2003. Neural crest induction by paraxial mesoderm in Xenopus embryos requires FGF signals. *Development*, 130(14), pp.3111–3124.

Monsoro-Burq, A.H., Wang, E., Harland, R., 2005. Msx1 and Pax3 cooperate to mediate FGF8 and WNT signals during Xenopus neural crest induction. *Developmental Cell*, 8(2), pp.167–178.

Murko, C., Lagger, S., Steiner, M., Seiser, C., Schoefer, C., Pusch, O., 2013. Histone deacetylase inhibitor Trichostatin A induces neural tube defects and promotes neural crest specification in the chicken neural tube. *Differentiation*, 85(1–2), pp.55–66.

Neph, S., Vierstra, J., Stergachis, A.B., et al., 2012. An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature*, 489(7414), pp.83–90.

Nichane, M., de Crozé, N., Ren, X., Souopgui, J., Monsoro-Burq, A.H., Bellefroid, E.J., 2008. Hairy2-Id3 interactions play an essential role in Xenopus neural crest progenitor specification. *Developmental Biology*, 322(2), pp.355–367.

Nichane, M., Ren, X., Bellefroid, E.J., 2009. Self-regulation of Stat3 activity coordinates cell-cycle progression and neural crest specification. *The EMBO Journal*, 29(1), pp.55–67.

Ogino, H., McConnell, W.B., Grainger, R.M., 2006. High-throughput transgenesis in Xenopus using I-SceI meganuclease. *Nature Protocols*, 1(4), pp.1703–1710.

Ossipova, O., Sokol, S.Y., 2011. Neural crest specification by noncanonical Wnt signaling and PAR-1. *Development*, 138(24), pp.5441–50.

Owlarn, S., Bartscherer, K., 2016. Go ahead, grow a head! A planarian's guide to anterior regeneration. *Regeneration*, 3(3), pp.139–155.

Park, S., Hannenhalli, S., Choi, S., 2014. Conservation in first introns is positively associated with the number of exons within genes and the presence of regulatory epigenetic signals. *BMC Genomics*, 15(1), p.526.

Pearson, B.J., Eisenhoffer, G.T., Gurley, K.A., Rink, J.C., Miller, D.E., Sanchez Alvarado, A., 2009. Formaldehyde-based whole-mount in situ hybridization method for planarians. *Developmental Dynamics*, 238(2), pp.443–450.

Pegoraro, C., Monsoro-Burq, A.H., 2013. Signaling and transcriptional regulation in neural crest specification and migration: Lessons from xenopus embryos. *Developmental Biology*, 2(2), pp.247–259.

Pellettieri, J., Fitzgerald, P., Watanabe, S., Mancuso, J., Green, D.R., Sanchez Alvarado, A., 2010. Cell death and tissue remodeling in planarian regeneration. *Developmental Biology*, 338(1), pp.76–85.

Penzel, R., Oschwald, R., Chen, Y., Tacke, L., Grunz, H., 1997. Characterization and early embryonic expression of a neural specific transcription factor xSOX3 in Xenopus laevis. *The International Journal of Developmental Biology*, 41(5), pp.667–77.

Perez-Alcala, S., Nieto, M.A., Barbas, J. a, 2004. LSox5 regulates RhoB expression in the neural tube and promotes generation of the neural crest. *Development*, 131(18), pp.4455–4465.

Perino, M., Veenstra, G.J.C., 2016. Chromatin Control of Developmental Dynamics and Plasticity. *Developmental Cell*, 38(6), pp.610–620.

Petersen, C.P., Reddien, P.W., 2011. Polarized activation of notum at wounds inhibits Wnt signaling to promote planarian head regeneration. *Science*, 332(6031), pp.852–855.

Picelli, S., Björklund, A.K., Reinius, B., Sagasser, S., Winberg, G., Sandberg, R., 2014. Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Research*, 24(12), pp.2033–2040.

Plank, J.L., Dean, A., 2014. Enhancer function: Mechanistic and genome-wide insights come together. *Molecular Cell*, 55(1), pp.5–14.

Plouhinec, J.L., Roche, D.D., Pegoraro, C., et al., 2014. Pax3 and Zic1 trigger the early neural crest gene regulatory network by the direct activation of multiple key neural crest specifiers. *Developmental Biology*, 386(2), pp.461–472.

Pohl, B.S., Knöchel, W., 2001. Overexpression of the transcriptional repressor FoxD3 prevents neural crest formation in Xenopus embryos. *Mechanisms of Development*, 103(1–2), pp.93–106.

Pott, S., Lieb, J.D., 2014. What are super-enhancers? *Nature Genetics*, 47(1), pp.8–12.

Putnam, N.H., Butts, T., Ferrier, D.K., et al., 2008. The amphioxus genome and the evolution of the chordate karyotype. *Nature*, 453(7198), pp.1064–1071.

Qi, H.H., Sarkissian, M., Hu, G.-Q., et al., 2010. Histone H4K20/H3K9 demethylase PHF8 regulates zebrafish brain and craniofacial development. *Nature*, 466(7305), pp.503–7.

Rada-iglesias, A., Bajpai, R., Prescott, S., Brugmann, S.A., Swigut, T., 2012. Epigenomic Annotation of Enhancers Predicts Transcriptional Regulators of Human Neural Crest. *Stem Cell*, 11(5), pp.633–648.

Rai, K., Jafri, I.F., Chidester, S., James, S.R., Karpf, A.R., Cairns, B.R., Jones, D.A., 2010. Dnmt3 and G9a cooperate for tissue-specific development in zebrafish. *Journal of Biological Chemistry*, 285(6), pp.4110–4121.

Ramirez, R.N., El-Ali, N.C., Mager, M.A., Wayman, D., Conesa, A., Mortazavi, A., 2017. Dynamic Gene Regulatory Networks of Human Myeloid Differentiation. *Cell Systems*, 4(4), pp.416–429.

Read, E.M., Rodaway, A.R.F., Neave, B., Brandon, N., Holder, N., Patient, R.K., Walmsley, M.E., 1998. Evidence for non-axial A/P patterning in the nonneural ectoderm of Xenopus and zebrafish pregastrula embryos. *International Journal of Developmental Biology*, 42(6), pp.763–774.

Reddien, P.W., 2013. Specialized progenitors and regeneration. *Development*, 140(5), pp.951–957.

Remeseiro, S., Hörnblad, A., Spitz, F., 2016. Gene regulation during development in the light of topologically associating domains. *Developmental Biology*, 5(2), pp.169–185.

Richon, V.M., Emiliani, S., Verdin, E., Webb, Y., Breslow, R., Rifkind, R.A., Marks, P.A., 1998. A class of hybrid polar inducers of transformed cell differentiation inhibits histone deacetylases. *Proc Natl Acad Sci*, 95(6) pp.3003-3007.

Riddiford, N., Schlosser, G., 2016. Dissecting the pre-placodal transcriptome to reveal presumptive direct targets of Six1 and Eya1 in cranial placodes. *eLife*, 5, pp.1–33.

Rink, J.C., Gurley, K.A., Elliott, S.A., Sanchez Alvarado, A., 2009. Planarian Hh signaling regulates regeneration polarity and links Hh pathway evolution to cilia. *Science*, 326(5958), pp.1406–1410.

Rink, J.C., 2013. Stem cell systems and regeneration in planaria. *Development Genes and Evolution*, 223(1–2), pp.67–84.

Robb, S.M.C., Sanchez Alvarado, A., 2014. Histone Modifications and Regeneration in the Planarian Schmidtea mediterranea. *Current Topics in Developmental Biology*, 108, pp.71–93.

Roberts-Galbraith, R.H., Newmark, P. a, 2013. Follistatin antagonizes activin signaling and acts with notum to direct planarian head regeneration. *PNAS*, 110(4), pp.1363–8.

Roellig, D., Bronner, M.E., 2016. The epigenetic modifier DNMT3A is necessary for proper otic placode formation. *Developmental Biology*, 411(2), pp.294–300.

Ronshaugen, M., Levine, M., 2004. Visualization of trans-homolog enhancer-promoter interactions at the Abd-B hox locus in the Drosophila embryo. *Developmental Cell*, 7(6), pp.925–932.

Saint-Jeannet, J.P., He, X., Varmus, H.E., Dawid, I.B., 1997. Regulation of dorsal fate in the neuraxis by Wnt-1 and Wnt-3a. *PNAS*, 94(25), pp.13713–13718.

Saló, E., 2006. The power of regeneration and the stem-cell kingdom: Freshwater planarians (Platyhelminthes). *BioEssays*, 28(5), pp.546–559.

Saló, E., Baguñà, J., 1985. Proximal and distal transformation during intercalary regeneration in the planarian Dugesia(S)mediterranea. *Roux's Archives of Developmental Biology*, (194), pp.364–368.

Saló, E., Baguñà, J., 1984. Regeneration and pattern formation in planarians. *J. Embryol exp. Morph.*, 83, pp.63–80.

Sanchez Alvarado, A., Newmark, P.A., 1999. Double-stranded RNA specifically disrupts gene expression during planarian regeneration. *Developmental Biology*, 96, pp.5049–5054.

Sandmann, T., Vogg, M.C., Owlarn, S., Boutros, M., Bartscherer, K., 2011. The head-regeneration transcriptome of the planarian Schmidtea mediterranea. *Genome Biology*, 12(8), pp.1-19.

Sasai, N., Mizuseki, K., Sasai, Y., 2001. Requirement of FoxD3-class signaling for neural crest determination in Xenopus. *Development,* 128(13), pp.2525–2536.

Sauka-Spengler, T., Bronner-Fraser, M., 2008. A gene regulatory network orchestrates neural crest formation. *Molecular cell biology*, 9(7), pp.557–68.

Sawan, C., Vaissière, T., Murr, R., Herceg, Z., 2008. Epigenetic drivers and genetic passengers on the road to cancer. *Mutation Research*, 642(1–2), pp.1–13.

Schmitt, S.M., Gull, M., Brändli, A.W., 2014. Engineering Xenopus embryos for phenotypic drug discovery screening. *Advanced Drug Delivery Reviews*, 69–70, pp.225–246.

Schwarz, D., Varum, S., Zemke, M., Schöler, A., Baggiolini, A., Draganova, K., Koseki, H., Schübeler, D., Sommer, L., 2014. Ezh2 is required for neural crest-derived cartilage and bone formation. *Development*, 141(4), pp.867–77.

Scimone, M.L., Kravarik, K.M., Lapan, S.W., Reddien, P.W., 2014. Neoblast specialization in regeneration of the planarian schmidtea mediterranea. *Stem Cell Reports*, 3(2), pp.339–352.

Scimone, M.L., Lapan, S.W., Reddien, P.W., 2014. A forkhead Transcription Factor Is Wound-Induced at the Planarian Midline and Required for Anterior Pole Regeneration. *PLoS Genetics*, 10(1).

Session, A.M., Uno, Y., Kwon, T., et al., 2016. Genome evolution in the allotetraploid frog Xenopus laevis. *Nature*, 538(7625), pp.336–343.

Sewitz, S.A., Fahmi, Z., Lipkow, K., 2017. Higher order assembly: folding the chromosome. *Current Opinion in Structural Biology*, 42, pp.162–168.

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

Simões-Costa, M., McKeown, S.J., Tan-Cabugao, J., Sauka-Spengler, T., Bronner, M.E., 2012. Dynamic and Differential Regulation of Stem Cell Factor FoxD3 in the Neural Crest Is Encrypted in the Genome. *PLoS Genetics*, 8(12).

Simões-Costa, M., Bronner, M.E., 2015. Establishing neural crest identity: a gene regulatory recipe. *Development*, 142(2), pp.242–57.

Simon, C.S., Downes, D.J., Gosden, M.E., et al., 2017. Functional characterisation of cis-regulatory elements governing dynamic *Eomes* expression in the early mouse embryo. *Development*, 144, pp.1249-1260.

Singh, N., Trivedi, C.M., Lu, M., Mullican, S.E., Lazar, M.A., Epstein, J.A., 2011. Histone deacetylase 3 regulates smooth muscle differentiation in neural crest cells and development of the cardiac outflow tract. *Circulation Research*, 109(11), pp.1240–1249.

Sive, H.L., Grainger, R.M., Harland, R.M., 2000. *Early Development of Xenopus Laevis. A Laboratory Manual,* C. S. H. L. Press, NY.

Smith, J.C., 1995. Mesodern-inducing factors and mesodermal patterning. *Current Opinion in Cell Biology*, 7(6), pp.856–861.

Spitz, F., Furlong, E.E.M., 2012. Transcription factors: From enhancer binding to developmental control. *Nat. Rev. Genet.*, 13(9), pp.613–626.

van Steensel, B., Dekker, J., 2010. Genomics tools for unraveling chromosome architecture. *Nature Biotechnology*, 28(10), pp.1089–1095.

Steventon, B., Araya, C., Linker, C., Kuriyama, S., Mayor, R., 2009. Differential requirements of BMP and Wnt signalling during gastrulation and neurulation define two steps in neural crest induction. *Development*, 136(5), pp.771–9.

Strobl-Mazzulla, P.H., Bronner, M.E., 2012. A PHD12-Snail2 repressive complex epigenetically mediates neural crest epithelial-to-mesenchymal transition. *Journal of Cell Biology*, 198(6), pp.999–1010.

Strobl-Mazzulla, P.H., Sauka-Spengler, T., Bronner-Fraser, M., 2010. Histone demethylase JmjD2A regulates neural crest specification. *Developmental Cell*, 19(3), pp.460–468.

Stuhlmiller, T.J., García-Castro, M.I., 2012. Current perspectives of the signaling pathways directing neural crest induction. *Cellular and Molecular Life Sciences*, 69(22), pp.3715–3737.

Suryamohan, K., Halfon, M.S., 2015. Identifying transcriptional cis-regulatory modules in animal genomes. *Developmental Biology*, 4(2), pp.59–84.

Taneyhill, L. a, Coles, E.G., Bronner-Fraser, M., 2007. Snail2 directly represses cadherin6B during epithelial-to-mesenchymal transitions of the neural crest. *Development*, 134(8), pp.1481–1490.

Tanno, B., Sesti, F., Cesi, V., Bossi, G., Ferrari-Amorotti, G., Bussolari, R., Tirindelli, D., Calabretta, B., Raschellà, G., 2010. Expression of slug is regulated by c-Myb and is required for invasion and bone marrow homing of cancer cells of different origin. *Journal of Biological Chemistry*, 285(38), pp.29434–29445.

Tien, C.-L., Jones, A., Wang, H., Gerigk, M., Nozell, S., Chang, C., 2015. Snail2/Slug cooperates with Polycomb repressive complex 2 (PRC2) to regulate neural crest development. *Development*, 2, pp.1–10.

Tomlinson, M.L., Rejzek, M., Fidock, M., Field, R.A., Wheeler, G.N., 2009. Chemical genomics identifies compounds affecting Xenopus laevis pigment cell development. *Molecular BioSystems*, 5(4), p.376.

Vadasz, S., Marquez, J., Tulloch, M., Shylo, N.A., García-Castro, M.I., 2013. Pax7 is regulated by cMyb during early neural crest development through a novel enhancer. *Development*, 140(17), pp.3691–702.

Varshney, A., Scott, L.J., Welch, R.P., et al., 2017. Genetic regulatory signatures underlying islet gene expression and type 2 diabetes. *PNAS*, 114(9), pp.2301–2306.

Vásquez-Doorman, C., Petersen, C.P., 2014. zic-1 Expression in Planarian Neoblasts after Injury Controls Anterior Pole Regeneration. *PLoS Genetics*, 10(7).

van der Velden, Y.U., Wang, L., Querol Cano, L., Haramis, A.P.G., 2013. The polycomb group protein ring1b/rnf2 is specifically required for craniofacial development. *PloS one*, 8(9), p.e73997.

Van Wolfswinkel, J.C., Wagner, D.E., Reddien, P.W., 2014. Single-cell analysis reveals functionally distinct classes within the planarian stem cell compartment. *Cell Stem Cell*, 15(3), pp.326–339.

Visel, A., Blow, M.J., Li, Z., et al., 2009. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature*, 457(7231), pp.854–858.

Wagner, D.E., Wang, I., Reddien, P.W., 2011. Clonogenic neoblasts are pluripotent adult stem cells that underlie planarian regeneration. *Science*, 332(6031), p.811-.

Wakamatsu, Y., Watanabe, Y., Nakamura, H., Kondoh, H., 1997. Regulation of the neural crest cell fate by N-myc: promotion of ventral migration and neuronal differentiation. *Development*, 124(10), pp.1953–1962.

Wang, C., Kam, R.K.T., Shi, W., et al., 2015. The Proto-oncogene transcription factor Ets1 regulates neural crest development through histone deacetylase 1 to mediate output of bone morphogenetic protein signaling. *Journal of Biological Chemistry*, 290(36), pp.21925–21938.

Watanabe, K.-I., Takeda, K., Yasumoto, K.-I., et al., 2002. Identification of a distal enhancer for the melanocyte-specific promoter of the MITF gene. *Pigment cell research*, 15(3), pp.201–211.

Wenemoser, D., Lapan, S.W., Wilkinson, A.W., Bell, G.W., Reddien, P.W., 2012. A molecular wound response program associated with regeneration initiation in planarians. *Genes and Development*, 26(9), pp.988–1002.

Wenemoser, D., Reddien, P.W., 2010. Planarian regeneration involves distinct stem cell responses to wounds and tissue absence. *Developmental Biology*, 344(2), pp.979–991.

Werner, T., Hammer, A., Wahlbuhl, M., Bösl, M.R., Wegner, M., 2007. Multiple conserved regulatory elements with overlapping functions determine Sox10

expression in mouse embryogenesis. *Nucleic Acids Research*, 35(19), pp.6526–6538.

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

Witchley, J.N., Mayer, M., Wagner, D.E., Owen, J.H., Reddien, P.W., 2013. Muscle cells provide instructions for planarian regeneration. *Cell Reports*, 4(4), pp.633–641.

Van Wolfswinkel, J.C., Wagner, D.E., Reddien, P.W., 2014. Single-cell analysis reveals functionally distinct classes within the planarian stem cell compartment. *Cell Stem Cell*, 15(3), pp.326–339.

Woolfe, A., Goodson, M., Goode, D.K., 2005. Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biology*, 3(1), pp.116-130.

Wozniak, G.G., Strahl, B.D., 2014. Hitting the “mark”: Interpreting lysine methylation in the context of active transcription. *Biochimica et Biophysica Acta*, 1839(12), pp.1353–1361.

Wu, J., Huang, B., Chen, H., et al., 2016. The landscape of accessible chromatin in mammalian preimplantation embryos. *Nature*, 534(7609), pp.652–657.

Wu, M.Y., Ramel, M.C., Howell, M., Hill, C.S., 2011. SNW1 is a critical regulator of spatial BMP activity, neural plate border formation, and neural crest specification in vertebrate embryos. *PLoS Biology*, 9(2).

Yazawa, S., Umesono, Y., Hayashi, T., Tarui, H., Agata, K., 2009. Planarian Hedgehog/Patched establishes anterior-posterior polarity by regulating Wnt signaling. *PNAS*, 106(52), pp.22329–22334.

Yang, H., Pinello, C.E., Luo, J., et al., 2014. Small-molecule inhibitors of acetyltransferase p300 identified by high-throughput screening are potent anticancer agents. *Molecular Cancer Therapy*, 12(5), pp.610–620.

Ying, Q.L., Nichols, J., Chambers, I., Smith, A., 2003. BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. *Cell*, 115(3), pp.281–292.

Zhang, S., Cui, W., 2014. Sox2, a key factor in the regulation of pluripotency and neural differentiation. *World Journal of Stem Cells*, 6(3), pp.305–311.

Zhou, V.W., Goren, A., Bernstein, B.E., 2011. Charting histone modifications and the functional organization of mammalian genomes. *Nature Reviews Genetics*, 12(1), pp.7–18.

Zhu, S.J., Pearson, B.J., 2016. (Neo)blast from the past : new insights into planarian stem cell lineages. *Current Opinion in Genetics and Development*, 40, pp.74–80.

Zuber, M.E., Gestri, G., Viczian, A.S., Barsacchi, G., Harris, W.A., 2003. Specification of the vertebrate eye by a network of eye field transcription factors. *Development*, 130(21), pp.5155–5167.

Appendix 1: Tables of top motifs presents in the potential enhancers of the analyzed samples

Table i: Top motifs found in the specific NC intergenic open regions stage 13 using Homer. * Indicates the *Xenopus* homolog. The transcription factors mentioned in the Chapter 3 are highlighted. The TPM data is extracted form the RNA-sequencing data. Motifs shown have a p-value <0.01 and are sorted by the % of target sequences with motifs. Tpm=transcripts per million. L and S refer to the long and short version of the gene.

Motif Name	P-value	% of Target with Motif	TPM st13 NC (L/S)
Nanog(Homeobox)	1.00E-06	56.18%	none/none
Tgif2(Homeobox)	1.00E-06	46.10%	40.36/124.49
PHA-4(Forkhead)	1.00E-05	45.58%	**foxa4* 2.52/0.95
Tgif1(Homeobox)	1.00E-07	44.82%	28.24/12.50
Tbx5(T-box)	1.00E-04	43.02%	0.12/0.04
Nkx6.1(Homeobox)	1.00E-07	39.98%	0.23/0.10
Smad3(MAD)	1.00E-03	36.43%	0.30/0.14
KLF14(Zf)	1.00E-112	36.39%	none/none
Ptf1a(bHLH)	1.00E-02	33.55%	0.04/
Lhx3(Homeobox)	1.00E-32	32.39%	0.42/0.12
Isl1(Homeobox)	1.00E-07	30.51%	17.88/14.87
Eomes(T-box)	1.00E-09	29.67%	0.15/0.07
Meis1(Homeobox)	1.00E-16	28.91%	4.10/0.47
Sox3(HMG)	1.00E-18	28.55%	175.23/461.03
NFY(CCAAT)	1.00E-54	26.47%	**nfya* 211.62/48.27
Sox10(HMG)	1.00E-15	26.23%	0.23/0.16
Sox6(HMG)	1.00E-13	26.03%	0.30/0.11
Znf263(Zf)	1.00E-03	25.95%	none/none
PABPC1(?)	1.00E-02	24.23%	5044.78/1382.64
KLF5(Zf)	1.00E-74	23.67%	27.23/11.38
Maz(Zf)	1.00E-16	23.43%	6.41/18.71
Foxo1(Forkhead)	1.00E-02	23.15%	0.82/0.48
IBL1(bHLH)	1.00E-02	22.03%	*Thy1 0.01/
Smad4(MAD)	1.00E-03	21.47%	**smad4*.1 75.81/274.89 **smad4*.2 3.80/6.81
Lhx2(Homeobox)	1.00E-23	21.15%	5.21/3.79
Ascl1(bHLH)	1.00E-03	20.55%	0.30/0.21
Lhx1(Homeobox)	1.00E-15	20.39%	0.10/0.21
E2A(bHLH)	1.00E-03	20.27%	**tcf3* 60.36/222.10

ERG(ETS)	1.00E-03	19.83%	0.02/0
ETV1(ETS)	1.00E-04	18.83%	26.36/24.64
Ap4(bHLH)	1.00E-07	18.59%	**ap4e1* 0.71/2.73
FOXA1(Forkhead)	1.00E-10	18.47%	0.00/0.42
Zic(Zf)	1.00E-29	18.39%	**zic1* 181.82/366.90
SPCH(bHLH)	1.00E-09	17.59%	**foxp2* 0.11/0.06
Fli1(ETS)	1.00E-05	17.39%	0.39/0.00
Sox15(HMG)	1.00E-12	17.23%	515/
MyoG(bHLH)	1.00E-09	17.07%	none/none
Tcf12(bHLH)	1.00E-08	16.99%	37.55/23.91
Sox2(HMG)	1.00E-18	16.31%	/276.76
GSC(Homeobox)	1.00E-04	16.19%	3.52/1.52
PU.1-IRF(ETS:IRF)	1.00E-07	15.79%	**spi1* 0.14/0.22
Rbpj1(?)	1.00E-04	15.75%	4.91/7.16
MITF(bHLH)	1.00E-09	15.67%	0.03/0.10
HIF-1b(HLH)	1.00E-02	15.59%	**arnt* 5.01/10.43
NPAS2(bHLH)	1.00E-05	15.39%	0.24/
MyoD(bHLH)	1.00E-10	15.35%	**myod1* 0.14/0.94
FOXM1(Forkhead)	1.00E-07	15.19%	2.69/19.89
ZNF467(Zf)	1.00E-11	14.71%	none/none
ETS1(ETS)	1.00E-04	14.67%	1.22/1.50
Tcf21(bHLH)	1.00E-06	14.63%	0.19/0.88
GABPA(ETS)	1.00E-04	14.27%	28.23/56.36
Atf1(bZIP)	1.00E-07	13.91%	53.64/41.32
LIN-39(Homeobox)	1.00E-02	13.87%	none/none
HOXD13(Homeobox)	1.00E-02	13.71%	0.01/0.10
Pdx1(Homeobox)	1.00E-04	13.39%	none/none
Etv2(ETS)	1.00E-03	12.99%	0.94/0.08
Pit1(Homeobox)	1.00E-02	12.95%	**pou1f1* none/
Myf5(bHLH)	1.00E-08	12.79%	1.63/0.84
TEAD4(TEA)	1.00E-10	12.75%	none/none
FoxL2(Forkhead)	1.00E-07	12.75%	0.12/0.06
Foxa2(Forkhead)	1.00E-05	12.51%	0.40/0.15
Sox9(HMG)	1.00E-04	12.04%	3.78/7.13
Sox4(HMG)	1.00E-05	11.92%	0.92/4.32
Oct4(POU,Homeobox)	1.00E-31	11.76%	**pou5f3.1* 435.03/150.06 **pou5f3.2* 39.68/59.31 **pou5f3.3* 2.45/1.61

Klf9(Zf)	1.00E-59	11.04%	2.42/0.20
Otx2(Homeobox)	1.00E-02	11.04%	144.69/65.24
Oct6(POU,Homeobox)	1.00E-27	10.72%	**pou3f1* 8.14/3.96
Elk1(ETS)	1.00E-03	10.72%	6.97/0.97
KLF10(Zf)	1.00E-21	10.68%	14.55/50.38
Elk4(ETS)	1.00E-04	10.56%	2.20/15.76
Egr1(Zf)	1.00E-13	10.32%	1.41/1.76
Atf7(bZIP)	1.00E-04	10.32%	3.06/4.12
ELF5(ETS)	1.00E-02	10.28%	0.11/0.11
ZNF189(Zf)	1.00E-02	10.16%	none/none
EBF1(EBF)	1.00E-02	10.12%	**ebf1-like* /0.20
CLOCK(bHLH)	1.00E-07	10.08%	3.18/11.53
Max(bHLH)	1.00E-06	10.04%	71.81/
DPL-1(E2F)	1.00E-05	9.92%	none/none
Cdx2(Homeobox)	1.00E-03	9.92%	4.03/4.40
n-Myc(bHLH)	1.00E-04	9.48%	0.66/0.57
HOXA9(Homeobox)	1.00E-02	9.44%	0.14/0.09
Foxh1(Forkhead)	1.00E-02	9.16%	1.64/0.41
USF1(bHLH)	1.00E-06	8.80%	5.51/15.65
Brn1(POU,Homeobox)	1.00E-25	8.48%	none/none
Klf4(Zf)	1.00E-33	7.96%	0.10/0.50
Usf2(bHLH)	1.00E-12	7.84%	102.20/
Atf2(bZIP)	1.00E-02	7.84%	0.37/0.18
E2F4(E2F)	1.00E-04	7.76%	27.38/42.68
Hoxc9(Homeobox)	1.00E-04	7.76%	0.24/0.23
c-Jun-CRE(bZIP)	1.00E-02	7.40%	0.87/0.65
c-Myc(bHLH)	1.00E-09	7.32%	13.03/19.81

Table ii: Top motifs found in the specific NC intergenic open regions stage 18 using Homer. * Indicates the *Xenopus* homolog. The transcription factors mentioned in the Chapter 3 are highlighted. The TPM data is extracted form the RNA-sequencing data. Motifs shown have a p-value <0.01 and are sorted by the % of target sequences with motifs. Tpm=transcripts per million. L and S refer to the long and short version of the gene.

Motif Name	**P-value**	**% of Target with Motif**	**TPM st18 NC (L/S)**
Tgif2(Homeobox)	1.00E-03	42.99%	34.21/63.13
Tgif1(Homeobox)	1.00E-02	40.74%	30.44/14.50
Sox3(HMG)	1.00E-17	31.35%	63.89/168.25
Sox10(HMG)	1.00E-13	28.04%	64.49/54.13

Sox6(HMG)	1.00E-11	27.38%	0.11/0.41
Meis3(Homeobox)	1.00E-07	26.19%	75.46/76.55
Lhx3(Homeobox)	1.00E-05	25.66%	0.22/0.04
Pdx1(Homeobox)	1.00E-22	23.02%	0/0
LIN-39(Homeobox)	1.00E-13	21.43%	none/none
Sox15(HMG)	1.00E-12	19.97%	129.92/-
Sox2(HMG)	1.00E-17	19.05%	/92.7
Zic(Zf)	1.00E-13	17.46%	*zic1** 48.84/57.57
Lhx1(Homeobox)	1.00E-02	15.74%	0.18/0.31
Sox9(HMG)	1.00E-09	15.34%	9.34/93.33
KLF14(Zf)	1.00E-03	14.95%	3.42/0.50
Lhx2(Homeobox)	1.00E-02	14.95%	18.85/9.01
Sox4(HMG)	1.00E-07	14.02%	3.24/4.47
HOXA9(Homeobox)	1.00E-08	13.89%	0.41/0.35
FOXA1(Forkhead)	1.00E-02	13.10%	5.87/2.29
Hoxc9(Homeobox)	1.00E-10	11.51%	0.28/0.43
HNF6(Homeobox)	1.00E-04	10.71%	none/none
AP-2gamma(AP2)	1.00E-02	10.32%	148.37/193.11
AP-2alpha(AP2)	1.00E-05	8.60%	92.49/42.20
HOXA2(Homeobox)	1.00E-33	7.67%	2.97/4.73
Tcf4(HMG)	1.00E-08	7.28%	none/none
BATF(bZIP)	1.00E-02	6.88%	0/0.13
Hoxb4(Homeobox)	1.00E-11	6.35%	0.76/0.20
Egr1(Zf)	1.00E-02	5.82%	1.03/1.14
Pbx3(Homeobox)	1.00E-05	5.56%	0.28/1.78
Pknox1(Homeobox)	1.00E-05	5.42%	21.77/17.04
PAX3(Homeobox)	1.00E-04	4.37%	148.75/153.01
Tbx20(T-box)	1.00E-04	3.84%	1.41/
Tcf3(HMG)	1.00E-03	3.84%	77.99/214.77
OCT4(POU)	1.00E-02	3.31%	**pou5f3.1* 47.03/11.87
			**pou5f3.2* 0.89/2.67
			**pou5f3.3* 2.14/1.44
VDR(NR)	1.00E-02	3.17%	3.60/13.85
Nur77(NR)	1.00E-03	3.04%	none/none
PBX1(Homeobox)	1.00E-03	2.25%	49.07/26.91
Pax7(Homeobox)	1.00E-03	2.12%	0.42/1.47
TR4(NR)	1.00E-03	1.85%	none/none

Table iii: Top motifs found in the specific ectoderm intergenic open regions stage 18 using Homer. The transcription factors mentioned in the Chapter 3 are highlighted. Motifs shown have a p-value<0.01 and are sorted by the % of target sequences with motifs.

Motif Name	P-value	% of Target with Motif
Nkx6.1(Homeobox)	1.00E-13	40.82%
Isl1(Homeobox)	1.00E-05	27.95%
Lhx3(Homeobox)	1.00E-18	27.90%
GATA3(Zf)	1.00E-21	24.24%
Eomes(T-box)	1.00E-03	23.84%
AP-2gamma(AP2)	1.00E-180	22.82%
Sox3(HMG)	1.00E-03	20.61%
Foxo1(Forkhead)	1.00E-02	20.45%
KLF14(Zf)	1.00E-47	19.75%
Sox6(HMG)	1.00E-02	19.70%
Lhx2(Homeobox)	1.00E-21	19.67%
Meis1(Homeobox)	1.00E-02	19.22%
Sox10(HMG)	1.00E-03	19.06%
AP-2alpha(AP2)	1.00E-180	18.77%
Lhx1(Homeobox)	1.00E-11	18.71%
Gata4(Zf)	1.00E-34	18.42%
FOXA1(Forkhead)	1.00E-07	17.43%
Smad4(MAD)	1.00E-03	16.74%
KLF5(Zf)	1.00E-79	16.55%
TEAD4(TEA)	1.00E-78	16.15%
GRHL2(CP2)	1.00E-217	15.40%
NFY(CCAAT)	1.00E-05	14.63%
Maz(Zf)	1.00E-05	14.44%
Gata2(Zf)	1.00E-39	14.36%
FOXM1(Forkhead)	1.00E-04	13.96%
Rbpj1	1.00E-05	13.85%
Gata1(Zf)	1.00E-39	13.45%
Sox15(HMG)	1.00E-03	13.19%
Tbet(T-box)	1.00E-04	12.44%
ZNF416(Zf)	1.00E-04	12.36%
FoxL2(Forkhead)	1.00E-02	11.64%
Foxa2(Forkhead)	1.00E-04	11.40%
TEAD2(TEA)	1.00E-72	11.19%
EBF1(EBF)	1.00E-21	10.86%

Sox2(HMG)	1.00E-02	10.62%
Bcl6(Zf)	1.00E-02	10.01%
Sox4(HMG)	1.00E-02	9.42%
NFAT(RHD)	1.00E-03	9.21%
CEBP(bZIP)	1.00E-02	8.76%
STAT4(Stat)	1.00E-02	8.68%
Oct4(POU,Homeobox)	1.00E-14	8.44%
Oct6(POU,Homeobox)	1.00E-20	8.38%
AP-1(bZIP)	1.00E-03	7.34%
KLF10(Zf)	1.00E-22	7.21%
Zelda(Zf)	1.00E-05	7.15%
FOXP1(Forkhead)	1.00E-08	6.67%
Atf3(bZIP)	1.00E-02	6.67%
BATF(bZIP)	1.00E-05	6.27%
Brn1(POU,Homeobox)	1.00E-16	6.25%
Stat3+il21(Stat)	1.00E-03	6.22%
Rfx5(HTH)	1.00E-11	6.14%
Egr1(Zf)	1.00E-06	5.74%
Fra1(bZIP)	1.00E-03	5.58%
Klf9(Zf)	1.00E-27	5.34%
Klf4(Zf)	1.00E-33	5.29%
Oct2(POU,Homeobox)	1.00E-05	4.86%
Phox2a(Homeobox)	1.00E-02	4.59%
Six1(Homeobox)	1.00E-05	4.54%
Stat3(Stat)	1.00E-07	4.51%
Tcf4(HMG)	1.00E-05	4.51%
p63(p53)	1.00E-05	4.35%
Pbx3(Homeobox)	1.00E-04	3.36%
E-box	1.00E-03	3.26%
Fosl2(bZIP)	1.00E-08	3.23%
Pknox1(Homeobox)	1.00E-04	3.18%
c-Myc(bHLH)	1.00E-02	3.04%
Chop(bZIP)	1.00E-02	3.02%
Hoxb4(Homeobox)	1.00E-02	2.88%
STAT1(Stat)	1.00E-04	2.70%
STAT5(Stat)	1.00E-04	2.62%
Sp1(Zf)	1.00E-17	2.59%
Pho2(bHLH)	1.00E-02	2.43%
PAX5(Homeobox)	1.00E-15	2.24%

Rfx1(HTH)	1.00E-03	2.19%
Jun-AP1(bZIP)	1.00E-07	2.16%
Tcf3(HMG)	1.00E-02	2.16%
CTCF(Zf)	1.00E-03	2.08%
JunD(bZIP)	1.00E-02	2.03%
GATA3(Zf),DR4	1.00E-10	1.92%
Egr2(Zf)	1.00E-09	1.74%
p53(p53)	1.00E-06	1.60%
Rfx2(HTH)	1.00E-05	1.39%
Brn2(POU,Homeobox)	1.00E-02	1.39%
GATA3(Zf),DR8	1.00E-02	1.07%
FOXA1 (Forkhead,NR)	1.00E-03	0.93%
Pax7(Homeobox)	1.00E-08	0.69%
E2F7(E2F)	1.00E-02	0.69%

Table iv: Top motifs found in the specific neuroectoderm intergenic open regions stage 18 using Homer. Highlighted the transcription factors mentioned in the Chapter 3. Motifs shown have a p-value <0.01 and are sorted by the % of target sequences with motifs.

Motif Name	P-value	% of Target with Motif
Nanog(Homeobox)	1.00E-06	67.13%
Nkx6.1(Homeobox)	1.00E-68	57.45%
Lhx3(Homeobox)	1.00E-243	47.40%
Sox3(HMG)	1.00E-491	46.89%
Isl1(Homeobox)	1.00E-137	46.61%
Sox6(HMG)	1.00E-366	43.34%
Sox10(HMG)	1.00E-439	43.26%
CRX(Homeobox)	1.00E-31	43.08%
Trl(Zf)	1.00E-08	39.24%
Lhx2(Homeobox)	1.00E-414	38.36%
Lhx1(Homeobox)	1.00E-320	37.87%
Eomes(T-box)	1.00E-06	36.01%
EGL-5(Homeobox)	1.00E-34	35.04%
Sox15(HMG)	1.00E-475	33.76%
Sox2(HMG)	1.00E-662	32.64%
Foxo1(Forkhead)	1.00E-09	29.37%
Meis1(Homeobox)	1.00E-10	29.23%
GSC(Homeobox)	1.00E-126	28.39%

GATA3(Zf)	1.00E-03	27.35%
Sox4(HMG)	1.00E-346	25.47%
NFY(CCAAT)	1.00E-50	24.77%
Smad4(MAD)	1.00E-09	24.49%
Maz(Zf)	1.00E-44	24.42%
Sox9(HMG)	1.00E-218	23.69%
HOXD13(Homeobox)	1.00E-08	23.16%
Otx2(Homeobox)	1.00E-140	21.44%
Pdx1(Homeobox)	1.00E-21	21.10%
Rbpj1	1.00E-03	19.99%
PU.1-IRF(ETS:IRF)	1.00E-19	19.47%
FOXA1(Forkhead)	1.00E-02	19.32%
Ap4(bHLH)	1.00E-02	19.20%
FOXM1(Forkhead)	1.00E-03	19.02%
Gata4(Zf)	1.00E-05	18.84%
Tbet(T-box)	1.00E-03	17.43%
RXR(NR),DR1	1.00E-16	16.74%
Zic(Zf)	1.00E-42	16.70%
KLF5(Zf)	1.00E-50	16.52%
HOXA9(Homeobox)	1.00E-09	14.70%
NFAT(RHD)	1.00E-07	14.68%
CEBP(bZIP)	1.00E-05	13.98%
Gfi1b(Zf)	1.00E-06	13.80%
CES-1(Homeobox)	1.00E-34	13.77%
HNF6(Homeobox)	1.00E-15	13.29%
Phox2a(Homeobox)	1.00E-137	12.90%
Foxh1(Forkhead)	1.00E-08	12.89%
TEAD(TEA)	1.00E-26	12.59%
MafA(bZIP)	1.00E-04	12.50%
Oct6(Homeobox)	1.00E-54	11.76%
Hoxc9(Homeobox)	1.00E-05	10.95%
AP-1(bZIP)	1.00E-06	10.88%
PRDM1(Zf)	1.00E-20	10.60%
Atf3(bZIP)	1.00E-09	10.39%
BATF(bZIP)	1.00E-17	9.37%
FOXP1(Forkhead)	1.00E-13	8.93%
Max(bHLH)	1.00E-03	8.87%
Brn1(POU,Homeobox)	1.00E-49	8.85%
KLF10(Zf)	1.00E-41	8.82%

Egr1(Zf)	1.00E-16	8.78%
n-Myc(bHLH)	1.00E-03	8.59%
Fra1(bZIP)	1.00E-10	8.47%
Esrrb(NR)	1.00E-06	8.34%
Tcf4(HMG)	1.00E-31	8.22%
NFkB-p65(RHD)	1.00E-02	7.77%
STAT6(Stat)	1.00E-02	7.46%
Klf9(Zf)	1.00E-104	7.30%
ELT-3(Gata)	1.00E-02	7.28%
IRF4(IRF)	1.00E-03	7.27%
Oct2(POU,Homeobox)	1.00E-11	7.21%
Rfx5(HTH)	1.00E-04	6.59%
HNF4a(NR)	1.00E-02	6.19%
FXR(NR)	1.00E-05	6.08%
Pbx3(Homeobox)	1.00E-29	5.87%
DMRT1(DM)	1.00E-03	5.71%
E2F6(E2F)	1.00E-12	5.70%
PAX5(Paired,Homeobox)	1.00E-16	5.58%
Hoxb4(Homeobox)	1.00E-31	5.49%
PAX3(Homeobox)	1.00E-25	5.42%
Pknox1(Homeobox)	1.00E-20	5.34%
Klf4(Zf)	1.00E-32	5.20%
MafF(bZIP)	1.00E-04	5.04%
E-box	1.00E-06	4.99%
Oct4(Homeobox)	1.00E-58	4.92%
c-Myc(bHLH)	1.00E-05	4.81%
DMRT6(DM)	1.00E-03	4.77%
Nur77(NR)	1.00E-39	4.67%
PGR(NR)	1.00E-02	4.35%
Tcf3(HMG)	1.00E-11	4.35%
GRE(NR),IR3	1.00E-02	3.87%
Pax8(Paired,Homeobox)	1.00E-08	3.84%
Pho2(bHLH)	1.00E-04	3.80%
Fosl2(bZIP)	1.00E-07	3.72%
VDR(NR),DR3	1.00E-03	3.70%
ARE(NR)	1.00E-04	3.60%
Rfx1(HTH)	1.00E-21	3.60%
Sp1(Zf)	1.00E-43	3.56%
Tbx20(T-box)	1.00E-07	3.37%

PAX6(Homeobox)	1.00E-93	3.32%
PU.1:IRF8(ETS:IRF)	1.00E-03	3.02%
FHY3(FAR1)	1.00E-06	2.90%
BORIS(Zf)	1.00E-02	2.86%
EKLF(Zf)	1.00E-07	2.80%
CTCF(Zf)	1.00E-09	2.73%
ABF1	1.00E-04	2.61%
Jun-AP1(bZIP)	1.00E-09	2.61%
TR4(NR),DR1	1.00E-52	2.56%
Hnf1(Homeobox)	1.00E-03	2.37%
Rfx2(HTH)	1.00E-40	2.26%
RFX(HTH)	1.00E-53	2.13%
Brn2(POU,Homeobox)	1.00E-06	2.05%
Egr2(Zf)	1.00E-18	2.00%
HOXA2(Homeobox)	1.00E-08	1.99%
PBX1(Homeobox)	1.00E-11	1.98%
RORgt(NR)	1.00E-06	1.97%
Pax7(Homeobox)	1.00E-76	1.66%
GLI3(Zf)	1.00E-02	1.28%
TCFL2(HMG)	1.00E-06	1.26%
ZNF519(Zf)	1.00E-03	1.23%
E2F7(E2F)	1.00E-04	1.16%
ZNF41(Zf)	1.00E-06	0.61%
EBNA1(EBV-virus)	1.00E-06	0.31%
REST-NRSF(Zf)	1.00E-07	0.29%
ZBTB33(Zf)	1.00E-05	0.29%
p53(p53)	1.00E-02	0.23%

Table v: Top motifs found in the specific 0hR intergenic open regions using Homer. The transcription factors mentioned in the Chapter 4are highlighted. Motifs shown have a p-value<0.01 and are sorted by the % of target sequences with motifs.

Motif Name	P-value	% of Target with Motif
Isl1(Homeobox)	1.00E-17	20.56%
EGL-5(Homeobox)	1.00E-03	20.54%
LIN-39(Homeobox)	1.00E-02	19.04%
Nkx2.1(Homeobox)	1.00E-16	16.83%
Bapx1(Homeobox)	1.00E-43	15.76%

GATA3(Zf)	1.00E-17	15.67%
KANADI1(Myb)	1.00E-03	15.65%
Eomes(T-box)	1.00E-07	15.19%
Nkx2.5(Homeobox)	1.00E-44	14.60%
AR-halfsite(NR)	1.00E-16	14.04%
Nkx3.1(Homeobox)	1.00E-14	13.68%
Nkx2.2(Homeobox)	1.00E-45	12.32%
Sox3(HMG)	1.00E-02	11.83%
LHY(Myb)	1.00E-02	11.69%
FoxL2(Forkhead)	1.00E-17	10.87%
Foxo1(Forkhead)	1.00E-03	10.25%
FOXA1(Forkhead)	1.00E-07	10.13%
Gata4(Zf)	1.00E-13	10.04%
NFY(CCAAT)	1.00E-03	9.04%
Trl(Zf)	1.00E-05	8.80%
Mef2b(MADS)	1.00E-03	7.53%
Sox15(HMG)	1.00E-08	7.53%
Hoxc9(Homeobox)	1.00E-07	6.91%
Erra(NR)	1.00E-06	6.86%
Foxa2(Forkhead)	1.00E-06	6.59%
ZNF264(Zf)	1.00E-86	6.26%
EHF(ETS)	1.00E-07	6.23%
ETV1(ETS)	1.00E-02	5.84%
Gata2(Zf)	1.00E-13	5.79%
Rfx5(HTH)	1.00E-92	5.74%
Smad2(MAD)	1.00E-02	5.66%
ERG(ETS)	1.00E-06	5.64%
Sox2(HMG)	1.00E-05	5.44%
E2A(bHLH),near_PU.1	1.00E-15	5.13%
Gata1(Zf)	1.00E-06	4.71%
Sox4(HMG)	1.00E-03	4.49%
PQM-1(?)	1.00E-03	4.47%
SPDEF(ETS)	1.00E-04	4.44%
FOXP1(Forkhead)	1.00E-14	3.95%
PRDM1(Zf)	1.00E-43	3.89%
ELF5(ETS)	1.00E-09	3.76%
ABF1	1.00E-02	3.56%
Gfi1b(Zf)	1.00E-07	3.48%
ZFX(Zf)	1.00E-09	3.30%

FHY3(FAR1)	1.00E-75	3.28%
Etv2(ETS)	1.00E-08	3.27%
Esrrb(NR)	1.00E-05	3.02%
Egr1(Zf)	1.00E-58	2.97%
KLF14(Zf)	1.00E-02	2.65%
ZBTB12(Zf)	1.00E-44	2.43%
Nr5a2(NR)	1.00E-02	2.12%
GEI-11(Myb?)	1.00E-76	2.06%
Egr2(Zf)	1.00E-85	1.97%
IRF1(IRF)	1.00E-50	1.64%
HNF4a(NR),DR1	1.00E-11	1.59%
Oct4(POU,Homeobox)	1.00E-06	1.50%
IRF2(IRF)	1.00E-43	1.44%
Rfx1(HTH)	1.00E-30	1.40%
Rfx2(HTH)	1.00E-23	1.35%
SpiB(ETS)	1.00E-11	1.23%
ISRE(IRF)	1.00E-61	1.17%
PAX5(Homeobox),	1.00E-32	1.08%
FXR(NR),IR1	1.00E-05	1.07%
HOXA2(Homeobox)	1.00E-02	1.05%
HRE(HSF)	1.00E-02	0.98%
Dorsal(RHD)	1.00E-02	0.95%
Jun-AP1(bZIP)	1.00E-05	0.79%
Ets1-distal(ETS)	1.00E-03	0.78%
MafK(bZIP)	1.00E-10	0.69%
ZNF136(Zf)	1.00E-02	0.69%
ZNF675(Zf)	1.00E-06	0.62%
Bach1(bZIP)	1.00E-07	0.36%
NF-E2(bZIP)	1.00E-16	0.36%
GLI3(Zf)	1.00E-07	0.35%
Nrf2(bZIP)	1.00E-12	0.28%
TR4(NR),DR1	1.00E-04	0.23%
ZNF382(Zf)	1.00E-02	0.11%

Table vi: Top motifs found in the specific anterior 48hR intergenic open regions using Homer. Highlighted the transcription factors mentioned in the Chapter 4. Motifs shown have a p-value<0.01 and are sorted by the % of target sequences with motifs. * Indicates the planarian homolog

Motif Name	P-value	% of Target with Motif
Nkx6.1(Homeobox)	1.00E-12	51.13%
Nanog(Homeobox)	1.00E-44	50.78%
PHA-4(Forkhead)	1.00E-63	42.68%
Lhx3(Homeobox)	1.00E-106	39.29%
Tgif2(Homeobox)	1.00E-22	31.90%
Tgif1(Homeobox)	1.00E-21	31.01%
LIN-39(Homeobox)	1.00E-15	30.12%
Lhx1(Homeobox)	1.00E-67	27.08%
EGL-5(Homeobox)	1.00E-09	26.92%
Isl1(Homeobox)	1.00E-45	26.84%
Lhx2(Homeobox)	1.00E-59	24.88%
CRX(Homeobox)	1.00E-05	22.99%
MYB(HTH)	1.00E-14	21.41%
Pdx1(Homeobox)	1.00E-24	20.12%
Pit1(Homeobox)	1.00E-15	19.98%
HNF6(Homeobox)	1.00E-32	19.70%
CES-1(Homeobox)	1.00E-39	19.53%
SCL(bHLH)	1.00E-06	19.15%
Eomes(T-box)	1.00E-02	19.01%
SUT1?	1.00E-02	18.30%
GATA3(Zf)	1.00E-02	18.06%
NFY(CCAAT)	1.00E-239	17.60%
BMYB(HTH)	1.00E-06	16.78%
RBFox2(?)	1.00E-41	16.20%
Bapx1(Homeobox)	1.00E-20	15.83%
Tbx5(T-box)	1.00E-03	15.46%
AMYB(HTH)	1.00E-11	15.06%
Foxo1(Forkhead)	1.00E-26	14.89%
FOXA1(Forkhead)	1.00E-13	14.88%
Smad3(MAD)	1.00E-17	14.54%
Phox2a(Homeobox)	1.00E-45	14.46%
FoxL2(Forkhead)	1.00E-07	14.14%
GSC(Homeobox)	1.00E-05	13.87%

FOXM1(Forkhead)	1.00E-10	13.41%
Gata4(Zf)	1.00E-05	11.49%
HOXA9(Homeobox)	1.00E-03	11.42%
BMAL1(bHLH)	1.00E-12	11.09%
Meis1/*prep (Homeobox)	1.00E-23	10.87%
Ptf1a(bHLH)	1.00E-16	10.53%
Hoxc9(Homeobox)	1.00E-05	10.40%
Olig2(bHLH)	1.00E-05	10.31%
Foxa2(Forkhead)	1.00E-13	10.00%
HIF-1b(HLH)	1.00E-04	9.79%
Trl(Zf)	1.00E-34	9.59%
Foxh1(Forkhead)	1.00E-13	9.37%
AP-1(bZIP)	1.00E-118	9.03%
SPCH(bHLH)	1.00E-19	8.83%
Atf3(bZIP)	1.00E-122	8.80%
IBL1(bHLH)	1.00E-13	8.64%
BATF(bZIP)	1.00E-117	8.59%
Fra1(bZIP)	1.00E-134	8.56%
NFAT(RHD)	1.00E-03	8.40%
Atf1(bZIP)	1.00E-03	8.07%
Tal1	1.00E-02	7.33%
Smad4(MAD)	1.00E-07	6.81%
PIF4(bHLH)	1.00E-08	6.71%
HEB(bHLH)	1.00E-19	6.69%
Fli1(ETS)	1.00E-02	6.41%
Oct4(POU,Homeobox)	1.00E-02	6.37%
EHF(ETS)	1.00E-07	6.23%
Smad2(MAD)	1.00E-02	6.23%
Gata2(Zf)	1.00E-03	6.22%
PU.1-IRF(ETS:IRF)	1.00E-06	6.18%
HLH-1(bHLH)	1.00E-26	6.08%
Rbpj1(?)	1.00E-06	6.08%
Oct6(POU,Homeobox)	1.00E-06	6.02%
NPAS2(bHLH)	1.00E-14	5.96%
FOXP1(Forkhead)	1.00E-13	5.86%
ERG(ETS)	1.00E-11	5.84%
ETV1(ETS)	1.00E-03	5.77%
PIF5ox(bHLH)	1.00E-04	5.51%
MITF(bHLH)	1.00E-04	5.40%

Gata1(Zf)	1.00E-04	5.34%
TEAD(TEA)	1.00E-02	5.20%
Ascl1(bHLH)	1.00E-07	4.72%
Pax7(Paired,Homeobox)	1.00E-15	4.69%
Ap4(bHLH)	1.00E-16	4.55%
Brn1(POU,Homeobox)	1.00E-08	4.50%
Atf7(bZIP)	1.00E-02	4.27%
ETS1(ETS)	1.00E-09	4.23%
Atoh1(bHLH)	1.00E-09	4.21%
E2A(bHLH)	1.00E-10	4.05%
MafF(bZIP)	1.00E-06	4.03%
Fosl2(bZIP)	1.00E-97	3.97%
PRDM1(Zf)	1.00E-07	3.82%
Hoxb4(Homeobox)	1.00E-13	3.73%
KLF10(Zf)	1.00E-98	3.71%
MafA(bZIP)	1.00E-03	3.60%
STAT6(Stat)	1.00E-03	3.58%
DPL-1(E2F)	1.00E-03	3.57%
Esrrb(NR)	1.00E-05	3.57%
MyoG(bHLH)	1.00E-13	3.57%
E-box	1.00E-25	3.56%
Etv2(ETS)	1.00E-09	3.54%
PAX3 (Paired,Homeobox)	1.00E-17	3.53%
SFP1	1.00E-07	3.46%
Max(bHLH)	1.00E-08	3.45%
Tcf21(bHLH)	1.00E-14	3.43%
ELF5(ETS)	1.00E-02	3.38%
CLOCK(bHLH)	1.00E-09	3.31%
NeuroD1(bHLH)	1.00E-10	3.14%
n-Myc(bHLH)	1.00E-06	3.08%

Table vii: Top motifs found in the specific posterior 48hR intergenic open regions using Homer. The transcription factors mentioned in the Chapter 4 are highlighted. Motifs shown have a p-value<0.01 and are sorted by the % of target sequences with motifs.

Motif Name	P-value	% of Target with Motif
Nanog(Homeobox)	1.00E-05	32.82%
Lhx3(Homeobox)	1.00E-03	21.49%
Isl1(Homeobox)	1.00E-04	16.00%
HOXD13(Homeobox)	1.00E-02	14.57%
GATA3(Zf)	1.00E-03	13.30%
Nkx2.1(Homeobox)	1.00E-03	12.37%
Bapx1(Homeobox)	1.00E-13	11.75%
Nkx2.5(Homeobox)	1.00E-09	10.69%
Nkx3.1(Homeobox)	1.00E-02	10.03%
Nkx2.2(Homeobox)	1.00E-04	8.12%
Gata4(Zf)	1.00E-02	7.88%
SEP3(MADS)	1.00E-02	7.32%
Mef2a(MADS)	1.00E-03	3.31%
Rfx5(HTH)	1.00E-12	3.02%
Nr5a2(NR)	1.00E-09	3.00%
Esrrb(NR)	1.00E-05	2.85%
PRDM1(Zf)	1.00E-02	2.52%
Pax7(Paired,Homeobox)	1.00E-02	2.41%
MafF(bZIP)	1.00E-04	2.39%
Nr5a2(NR)	1.00E-08	2.29%
Egr1(Zf)	1.00E-08	1.70%
GATA3(Zf),DR4	1.00E-09	1.37%
Fosl2(bZIP)	1.00E-05	1.15%
HNF4a(NR),DR1	1.00E-02	1.10%
Egr2(Zf)	1.00E-12	1.01%
Jun-AP1(bZIP)	1.00E-08	0.97%
ZNF136(Zf)	1.00E-06	0.94%
IRF1(IRF)	1.00E-03	0.89%
GEI-11(Myb?)	1.00E-05	0.84%
Rfx1(HTH)	1.00E-02	0.81%
MafK(bZIP)	1.00E-10	0.76%
IRF2(IRF)	1.00E-02	0.69%
Bach2(bZIP)	1.00E-03	0.66%
ISRE(IRF)	1.00E-05	0.63%

Bach1(bZIP)	1.00E-09	0.53%
NF-E2(bZIP)	1.00E-15	0.53%
Nrf2(bZIP)	1.00E-13	0.43%
NRF(NRF)	1.00E-03	0.40%
NRF1(NRF)	1.00E-04	0.28%
ZNF692(Zf)	1.00E-02	0.25%
ZNF382(Zf)	1.00E-02	0.20%

Appendix 2:

The positive transcriptional elongation factor (P-TEFb) is required for neural crest specification

Victoria L. Hatch, Marta Marín-Barba, Simon Moxona, Christopher T. Ford, Nicole J. Ward, Matthew L. Tomlinson, Ines Desanlis, Adam E. Hendry, Saartje Hontelez, Ila van Kruijsbergen, Gert Jan C. Veenstra, Andrea E. Münsterberg, Grant N. Wheeler.

Appendix 3:

Unravelling the mechanisms that determine the uptake and metabolism of magnetic single and multicore nanoparticles in a *Xenopus laevis* model

M. Marín-Barba, H. Gavilán, L. Gutiérrez, E. Lozano-Velasco, I. Rodríguez-Ramiro, G. N. Wheeler, C. J. Morris, M. P. Morales and A. Ruiz

Developmental Biology 416 (2016) 361–372

Contents lists available at ScienceDirect

Developmental Biology

journal homepage: www.elsevier.com/locate/developmentalbiology

The positive transcriptional elongation factor (P-TEFb) is required for neural crest specification

Victoria L. Hatch [a,1], Marta Marin-Barba [a], Simon Moxon [a,2], Christopher T. Ford [a], Nicole J. Ward [a], Matthew L. Tomlinson [a], Ines Desanlis [a], Adam E. Hendry [a], Saartje Hontelez [b], Ila van Kruijsbergen [b], Gert Jan C. Veenstra [b], Andrea E. Münsterberg [a], Grant N. Wheeler [a,*]

[a] School of Biological Sciences, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, UK
[b] Radboud University, Department of Molecular Developmental Biology, Faculty of Science, Radboud Institute for Molecular Life Sciences, Nijmegen, The Netherlands

ARTICLE INFO

Article history:
Received 12 October 2015
Received in revised form
6 May 2016
Accepted 8 June 2016
Available online 23 June 2016

Keywords:
Neural crest cells
Leflunomide
Transcriptional elongation
c-Myc
P-TEFb
Xenopus
Polymerase pausing
Cdk9
CyclinT1

ABSTRACT

Regulation of gene expression at the level of transcriptional elongation has been shown to be important in stem cells and tumour cells, but its role in the whole animal is only now being fully explored. Neural crest cells (NCCs) are a multipotent population of cells that migrate during early development from the dorsal neural tube throughout the embryo where they differentiate into a variety of cell types including pigment cells, cranio-facial skeleton and sensory neurons. Specification of NCCs is both spatially and temporally regulated during embryonic development. Here we show that components of the transcriptional elongation regulatory machinery, CDK9 and CYCLINT1 of the P-TEFb complex, are required to regulate neural crest specification. In particular, we show that expression of the proto-oncogene *c-Myc* and *c-Myc* responsive genes are affected. Our data suggest that P-TEFb is crucial to drive expression of *c-Myc*, which acts as a 'gate-keeper' for the correct temporal and spatial development of the neural crest.

1. Introduction

RNA polymerase pausing and transcriptional elongation is important in the specification of many types of cells including embryonic stem (ES) cells. In order for genes to undergo transcription, RNA polymerase II (Pol II) must be recruited to the promoter region (Hochheimer and Tjian, 2003). Subsequent to this, RNA Pol II may undergo regulation to temporally control the expression of genes in specific cell types. Such regulation includes the promoter proximal pausing of Pol II, the process by which Pol II undergoes stalling or poising at the 5′ end of a gene just downstream of the transcription start site (Core and Lis, 2008). This occurs due to binding of pause factors, DRB-sensitivity inducing factor (DSIF), negative elongation factor (NELF) and Gdown1 (Cheng et al., 2012; Wada et al., 1998; Yamaguchi et al., 1999). In order for transcriptional elongation to commence the positive transcription elongation factor (P-TEFb) must be recruited to the transcription elongation complex. P-TEFb is known to be a complex composed of cyclin dependent kinase 9 (CDK9) and cyclin T1 (Bres et al., 2008; Fujita et al., 2009; Jonkers and Lis, 2015; Kohoutek, 2009). To initiate transcriptional elongation, P-TEFb phosphorylates serine2 of the c-terminal domain (CTD) of Pol II and the SPT5 subunit of DSIF. After this phosphorylation, NELF dissociates from the complex and transcriptional elongation begins (Fujita et al., 2009). This process occurs when P-TEFb is in its active form and part of a collective of several proteins, termed the super elongation complex (SEC), which is known to regulate transcriptional elongation (Luo et al., 2012; Zhou et al., 2012).

The SEC is recruited to the promoter region of genes when rapid transcription is required in response to specific developmental signals, changes to available nutrition or even acute temperature change. It is because of this that the genes primarily regulated by RNA polymerase pausing are master regulator genes and/or rapid response genes such as heat shock and serum inducible genes (Fuda et al., 2009; Luo et al., 2012). It has also been

* Corresponding author.
E-mail address: grant.wheeler@uea.ac.uk (G.N. Wheeler).
[1] Present address: Institute of Molecular Biology gGmbH (IMB), 55128 Mainz, Germany.
[2] Present address: The Genome Analysis Centre (TGAC), Norwich Research Park, Norwich NR4 7UH, UK.

http://dx.doi.org/10.1016/j.ydbio.2016.06.012
0012-1606/

suggested that the main function of SEC recruitment to paused genes is not necessarily to allow rapid induction but also to synchronise genes in development (Boettiger and Levine, 2009). Work in *Drosophila* and ES cells has shown that transcription elongation is not important globally within the genome but is found to regulate 10–40% of genes. Some of these genes have been identified as crucial developmental control genes (Gaertner and Zeitlinger, 2014; Zeitlinger et al., 2007).

The neural crest is a transient embryonic cell population found exclusively in vertebrates. Neural crest cells (NCCs) migrate from the dorsal neural tube to various parts of the embryo where they differentiate to give rise to a diverse set of cell types (Milet and Monsoro-Burq, 2012; Pegoraro and Monsoro-Burq, 2013; Sauka-Spengler and Bronner-Fraser, 2008). These include enteric ganglia, neuroendocrine cells, melanophores, and craniofacial skeletal and connective tissue, thus NCCs are multi-potent progenitors. Their induction and specification begins during the early gastrula stage of development and continues throughout organogenesis. This process depends on multi-module gene regulatory networks acting between the neural plate, non-neural ectoderm and paraxial mesoderm (Buitrago-Delgado et al., 2015; Hoppler and Wheeler, 2015; Huang and Saint-Jeannet, 2004; Sauka-Spengler and Bronner-Fraser, 2008). Multiple signals and transcription factors are responsible for neural crest (NC) properties such as multipotency, induction, specification, migration and differentiation (Sauka-Spengler and Bronner-Fraser, 2008).

Once the neural plate border is fully specified and NCCs have been induced, neural crest specifiers begin to act (Milet and Monsoro-Burq, 2012). These include the transcription factors *Snail1*, *Snail2* (*Slug*), *Sox8*, *Sox9*, *Sox10*, *FoxD3*, and *c-Myc*. After specification morphological changes in the neural crest cells including changes to shape, adhesion and cell motility enable the cells to migrate from the dorsal neural tube (Sauka-Spengler and Bronner-Fraser, 2008). Neural crest specifiers can be separated into two groups; early and late NC specifiers. Expression of early neural crest specifiers, including *c-Myc*, and *Snail1*, is first identified at the neural plate border, whilst expression of slightly later neural crest specifiers such as *Snail2* and *FoxD3* is seen in pre-migrating neural crest and can continue throughout migration. Only after these have all started to be expressed does expression of members of the SoxE family such as *Sox10* start (Sauka-Spengler et al., 2007). It has been suggested that the early neural crest specifier *c-Myc* may have a role in maintaining neural crest cell multipotency and proliferation by maintaining a balance between cell death and proliferation following NC specification (Bellmeyer et al., 2003; Kee and Bronner-Fraser, 2005; Light et al., 2005).

Using chemical genetic screens we have previously identified small molecules that affect pigment cell development (Tomlinson et al., 2005, 2009a, 2009b). One of these, the small molecule leflunomide, inhibits pigment cell development in *Xenopus* and Zebrafish (Elworthy et al., 2003; Tomlinson et al., 2009b; White et al., 2011). In addition, leflunomide inhibits melanoma tumour growth in tissue culture and in mouse xenografts. Leflunomide inhibits the enzyme dihydroorotate dehydrogenase, a rate limiting enzyme in the pyrimidine biosynthesis pathway. It acts to inhibit transcriptional elongation of genes required for neural crest development and melanoma growth by limiting the pool of pyrimidines in the cell (White et al., 2011).

Here we further investigate leflunomide's ability to inhibit transcriptional elongation during neural crest development and show that it prevents expression of both, early and late neural crest specifiers such as *Snail2*, *Sox10* and *FoxD3* and *c-Myc*. As leflunomide acts through inhibiting transcriptional elongation we speculated that regulation of transcriptional elongation could be crucial for neural crest development. We therefore characterised the temporal and spatial expression of components of the Positive Transcription Elongation Factor B (P-TEFb) complex, *Cdk9* and *CyclinT1*, during neural crest development in *Xenopus laevis*. CDK9 and CYCLINT1 loss-of-function analyses by morpholino knockdown closely mimic the phenotypes induced by leflunomide treatment. Interestingly, we identify *c-Myc* as a gene susceptible to regulation of transcriptional elongation in the early NC and show that it is able to rescue CDK9 and CYCLINT1 morphant phenotypes. This is backed up by ChIP sequencing data of *Xenopus* embryos which shows the *c-Myc* gene to be one of the most 'paused' genes in the genome. We thus propose regulation at the level of transcriptional elongation to be important in NC development through the driving of *c-Myc* expression.

2. Results

2.1. The expression of the P-TEFb complex components during Xenopus laevis development

We have previously shown that inhibition of transcriptional elongation leads to defects in neural crest development (White et al., 2011). An important regulator of transcriptional elongation is the P-TEFb complex. We therefore decided to investigate the role of the components of the P-TEFb complex during *Xenopus* development. The P-TEFb complex is comprised of CDK9 and CYCLINT1. Expression analysis showed *Cdk9.S*, *Cdk9.L and CyclinT1* to be expressed in the neural plate and neural plate border of neurula stage embryos (Fig. 1). In tailbud stage embryos at stages 26 and 32 expression was seen in the NC derived branchial arches. To determine if their expression overlapped with the NC, we carried out double *in situ* hybridisations at st16 for all three P-TEFb components with the neural crest marker Sox10. Expression of the *Cdk9.S*, *Cdk9.L* and *CyclinT1* overlapped with expression of *Sox10*, indicating that the P-TEFb complex is expressed in neural crest cells (Fig. 1).

2.2. Leflunomide treatment affects the development of neural crest derivatives

Leflunomide was identified as an inhibitor of NC as a result of chemical screens carried out with the aim to look for small molecules that affect pigment cell development. Our previous results concentrated on the effects of leflunomide on zebrafish neural crest (White et al., 2011). Here we look at the effects of leflunomide in more detail on *Xenopus* neural crest. Leflunomide inhibits the formation of *Xenopus laevis* melanophores in a dose dependent manner and causes a loss of melanophores when added to the embryo after gastrulation (Fig. 2A and B). Initially, a leflunomide dose response was used to establish the general phenotype induced by the compound. DMSO treated embryos served as a control. Embryos were treated at a range of concentrations from stage 15 and analysed at stage 38 (Fig. 2Ai). Leflunomide treatment led to a loss of melanocytes in the underside of the tail, head and lateral stripe of the embryo (Fig. 2Ai black arrow heads). This phenotype was dose dependent, with 60 μM leflunomide demonstrating the most severe phenotypes occurring in a significant number of embryos (Fig. 2Bi). Treating *Xenopus* embryos at earlier stages showed a stronger effect on pigment cell development along with a shortened axis (Fig. 2Aii and Bii).

2.3. Leflunomide inhibits transcription of neural crest specifier genes

Next, we looked at how leflunomide treatment affects the expression of genes known to play a role at different stages of neural crest development. Embryos were treated with 60 μM leflunomide or DMSO from fertilisation and examined by *in situ* hybridisation

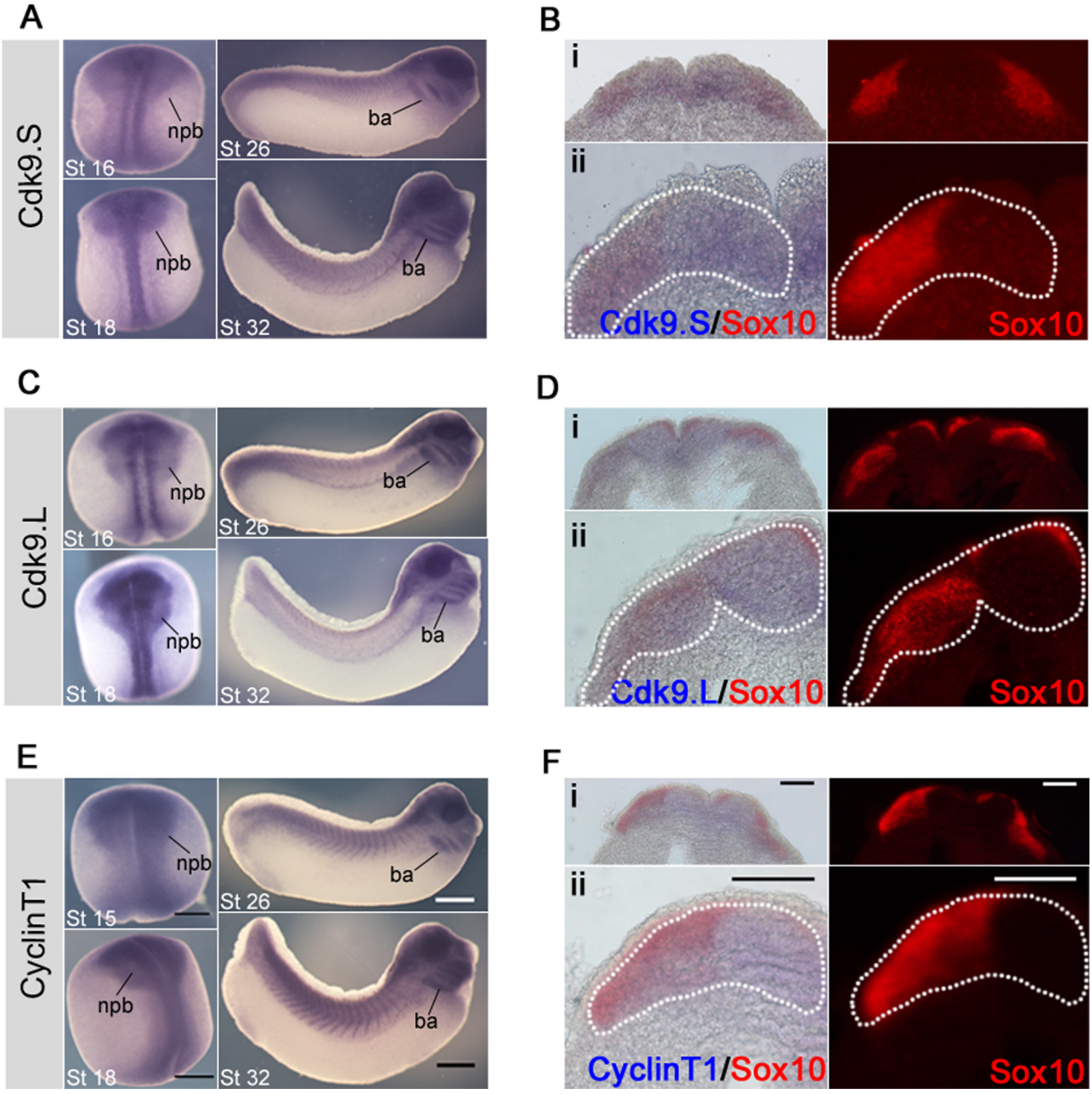
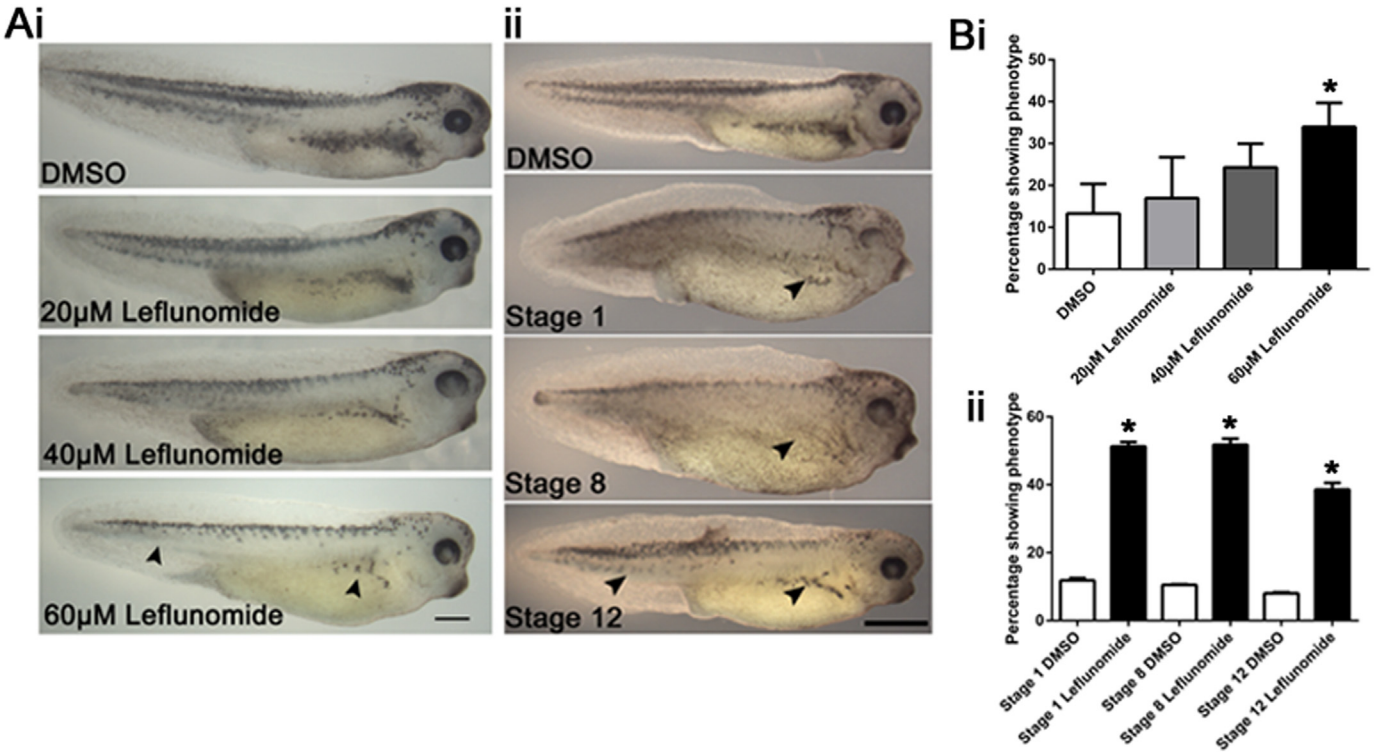
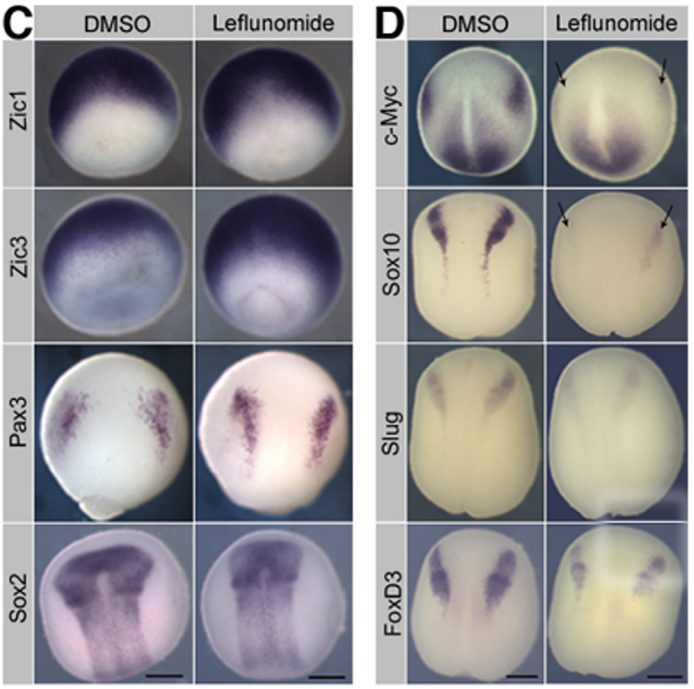


Fig. 1. The expression of the P-TEFb complex components during *Xenopus laevis* development. Transcripts for the P-TEFb components were detected using NBT/BCIP and expression of *CDK9.S*, *CDK9.L* and *CyclinT1* appeared blue in the sectioned neurula stage embryo. *Sox10* was detected using fast red and fluorescence microscopy. (A) *CDK9.S* expression at stage 16 and 18 in the neural plate and neural plate border and expression at stages 26 and 32 in the branchial arches. Scale bar = 100 μm (Bi) 10 × magnification section of a stage 16 embryo showing *CDK9.S* expression in the neural plate and neural plate border (blue) and *Sox10* expression in the neural crest (red). (ii) 20 × magnification section shows overlap of the blue *CDK9.S* expression with the red *Sox10* expression. Expression of *CDK9.S* is outlined with a dashed white line and overlaid onto the *Sox10* expression image. Scale bar = 100 μm (C) *CDK9.L* expression at stage 16 and 18 in the neural plate and neural plate border and expression at stages 26 and 32 in the branchial arches. Scale bar = 100 μm (Di) 10 × magnification section of a stage 16 embryo showing *CDK9.L* expression in the neural plate and neural plate border (ii) 20 × magnification section shows overlap of the blue *CDK9.L* expression with the red *Sox10* expression. Expression of *CDK9.L* is outlined with a dashed white line and overlaid onto the *Sox10* expression image. Scale bar = 100 μm (E) *CyclinT1* expression at stage 15 and 18 in the neural plate and neural plate border and expression at stages 26 and 32 in the branchial arches. Scale bar = 100 μm (Fi) 10 × magnification section of a stage 15 embryo showing *CDK9.S* expression in the neural plate and neural plate border (blue) and *Sox10* expression in the neural crest (red). (ii) 20 × magnification section shows overlap of the blue *CDK9.S* expression with the red *Sox10* expression. Expression of *CDK9.S* is outlined with a dashed white line and overlaid onto the *Sox10* expression image. Scale bar = 100 μm. Npb = Neural plate border, ba = branchial arches.

for different markers of neural crest development (Fig. 2C and D). The expression of the neural plate border markers *Zic1*, *Zic3* and *Pax3* did not change after leflunomide treatment when compared to the DMSO control (Fig. 2C). This suggests that leflunomide does not interfere with these early stages of neural crest development. There was also no significant change seen in *Sox2* expression throughout the neural plate (Fig. 2C) suggesting that neural development is largely unaffected by leflunomide treatment. We next tested genes known to play a role in neural crest specification. The earliest neural crest specifier gene, which showed a loss in expression was *c-Myc*. This gene gave an interesting expression pattern as it appeared that leflunomide-mediated inhibition of *c-Myc* expression was restricted to the areas fated to become neural crest (Fig. 2D black arrows). Loss of *c-Myc* expression was not observed in more posterior neural domains. Later neural crest specifiers, such as *Sox10*, were also significantly reduced after





E

Gene name	baseMean DMSO	baseMean Leflunomide	p-value
FoxD3	1292.18	52.97	2.20E-17
Sox10	134.80	1.04	3.48E-08
Sox9	1327.81	426.85	0.00032
Sox8	394.78	100.74	0.00042
Snail2 (Slug)	222.42	51.93	0.0014
c-myc	1052.42	662.60	0.134
Snail1	1788	2315	0.37
Sox1	10.59	10.39	1
Sox2	1667.70	1445.67	0.63
Zic1	5664.60	5783.72	0.94
Zic3	4624.69	6414.12	0.25
Pax3	2368.67	2709.59	0.64
Hsp90aa1.1	287.90	62.31	0.00033
Hsp105	3296.89	1363.62	0.0026
Hsp110	4241.47	1932.75	0.0068

Fig. 2. Leflunomide treatment affects the development of neural crest derivatives expression of neural crest genes. (Ai) Top to bottom shows DMSO treated control embryo, 20 μM, 40 μM and 60 μM leflunomide treated *Xenopus laevis* embryos. Leflunomide is applied at stage 15. Scale bar = 500 μm (ii) Top to bottom shows DMSO treated control embryo, 60 μM leflunomide applied at stage 1, 8 and 12. Phenotype is observed at stage 38. Black arrow heads indicate pigment cell loss. Scale bar = 300 μm (Bi) Graph showing the percentage of embryos displaying a pigment loss phenotype after treatment of DMSO n = 155, 20 μM n = 158, 40 μM n = 164 and 60 μM leflunomide n = 141. *$P < 0.05$ by Kruskall Wallis statistical test. (ii) Graph showing the percentage of embryos displaying a pigment loss phenotype after treatment of DMSO and 60 μM leflunomide applied at stage 1 (DMSO n = 60 leflunomide n = 72), 8 (DMSO n = 67 leflunomide n = 73) and 12 (DMSO n = 63 leflunomide n = 67). *$P < 0.05$ by Kruskall Wallis statistical test. (C) Whole mount *in situ* hybridisation carried out on embryos treated with either DMSO or 60 μM leflunomide from stage 4 until stage 13 (*Zic1, Zic3* and *Pax3*) or stage 15 (*Sox2*). The neural plate border markers *Zic1, Zic3* and *Pax3* show no change in expression. Similarly, no change is seen for neural plate marker *Sox2*. Scale bar = 200 μm (D) Whole mount *in situ* hybridisation carried out on embryos treated with either DMSO or 60 μM leflunomide from stage 4 until stage 13 (*c-Myc*) or stage 15 (*Sox10, Slug/Snail2* and *FoxD3*). Specific loss of *c-Myc* expression is seen in the anterior region of the embryo (black arrows) and no loss of expression is seen in the posterior neural tissue. Loss of *Sox10* expression can be seen (black arrows) and some loss or alteration of expression on Slug/*Snail2* and *FoxD3*. Scale bar = 200 μm (E) RNA sequencing carried out on animal caps injected with Wnt-8 and noggin and treated with DMSO or leflunomide. The change in expression seen in specific genes found in the animal cap sample by RNA sequencing is shown. Green = neural crest specifying genes, Orange = neuroectoderm genes, Blue = neural plate border specifiers, Grey = heat shock protein. The complete list of genes affected can be seen in Supplementary data 1.

leflunomide treatment (Fig. 2D black arrows) and also shown to be only affected in the NC as expression in the otic vesicle (ear) was not affected (Fig. S8). In addition we examined the neural crest specifiers Slug/*Snail2* and *FoxD3* (Fig. 2D). Many embryos showed a partial loss or an alteration in the expression of these genes. Quantification of these results (Fig. S1A) showed that *c-Myc* and *Sox10* show the strongest loss of expression. These results were confirmed by qRT-PCR on whole embryos (Fig. S1B). Using this technique we found that neural plate border markers *Zic1*, *Zic3*, *Pax3* and the neural plate marker *Sox2* displayed no significant change in expression level (Fig. S1B). Similarly to the *in situ* hybridisation data we saw a significant decrease in expression of neural crest specifying genes *c-Myc*, *Sox10*, *Slug/Snail2*, and *Sox9* (Fig. S1B).

2.4. RNA sequencing carried out on animal caps treated with leflunomide

To characterise the tissue specific effects of leflunomide treatment we carried out RNA sequencing on animal cap explants induced to become neural crest, neuroectoderm or ectoderm. Animal caps were cut at stage 9 and immediately immersed into either 60 μM leflunomide or DMSO and allowed to develop until stage 15. The genes predominantly down-regulated in the neural crest animal cap sample were genes involved in neural crest specification and general development (Fig. 2E and Fig. S1C). Most of these genes, were previously identified to be affected by *in situ* hybridisation such as *c-Myc*, *Sox10*, *FoxD3*, Slug/*Snail2* and *Sox9*. Interestingly Snail1 expression was not affected by leflunomide. In agreement with our *in situ* data, we saw no change in neural plate markers, such as *Sox1* and *Sox2*, or neural plate border markers, such as *Zic1*, *Zic3* and *Pax3*. Other categories of down regulated genes included genes involved in eye development and metabolism, both of which are closely associated with neural crest development. Interestingly we saw down-regulation of heat shock genes (*Hsp90aa1.1*, *Hsp105* and *Hsp110*), which were some of the first genes suggested to be controlled by RNA polymerase pausing and transcriptional elongation (Andrulis et al., 2000; Keegan et al., 2002; Lis et al., 2000). The neuroectoderm sample showed a decrease in the expression of genes, which restrict neuroectoderm fate *e.g.* Grainyhead-like 3, suggesting an increase in neuroectoderm cell fate. The ectoderm sample showed no significant increase or decrease in any other genes suggesting that regulation of transcriptional elongation is not important for ectoderm development. These results indicate that interference with transcriptional elongation in this system mostly affects neural crest development. We found expression of *c-Myc* to be slightly down-regulated in the neural crest sample with no change seen in the neuroectodem sample and it only shows low expression in the ectoderm. This result agrees with the *in situ* hybridisation results obtained (Fig. 2D). The small decrease in the NC sample is likely due to a masking of any decrease by contaminating neural tissue, which is always found in the NC animal caps. A full list of all genes affected in all three sample types can be found in Supplementary data 1. Overall, this data confirmed our findings that leflunomide acts to inhibit NC development and that this could be due to inhibition of transcriptional elongation.

2.5. Knockdown of the P-TEFb components leads to defects in neural crest development

To determine the requirement of the P-TEFb complex and subsequent transcriptional elongation for NCC development, we carried out knockdown studies of the individual components. We obtained morpholinos (MO) for CDK9.S, CDK9.L and CYCLINT1. Morpholino sequences and controls are shown in Fig. S2. CDK9 MOs were injected individually and in combination at the one or two cell stage. Embryos injected with these morpholinos were left to develop until stage 38. The phenotype obtained after knockdown of the P-TEFb components was comparable to that obtained after leflunomide treatment. The MO-injected embryos displayed neural crest defects including loss of pigment cells (Fig. 3A and Fig. S3A) comparable to those shown in Fig. 2 and disorganisation of cranial cartilage (Fig. S3B).

We next carried out *in situ* hybridisation using neural plate border genes, neural crest specifier genes and a neural plate marker. Embryos were injected with a standard control Morpholino, CDK9.S morpholino or the CYCLINT1 morpholino along with 300 pg of *LacZ* cRNA into one cell of a two-cell stage, with the non-injected half serving as control (Fig. 3B, Fig. S3D and Fig. S7). The expression of neural plate border markers *Zic1*, *Zic3* and *Pax3* were not affected (Fig. 3B). The Neural plate marker *Sox2* and other neural genes were not affected (Fig. 3B and S5). *c-Myc* and *Sox10* however, showed a striking knockdown after CDK9.S-MO or CYCLINT1-MO (Fig. 3B). Loss of *c-Myc* was seen specifically in the anterior patches, which give rise to NC (black arrows Fig. 3B). Complete loss of *Sox10* expression was seen in the NC (black arrows Fig. 3B). These results were confirmed using qRT-PCR (Fig. S4C). The same experiment was carried out after injection of the CDK9.L MO, or a combination of the CDK9.S and CDK9.L MOs and this data is shown in Fig. S4A and B. When embryos were allowed to progress to stage 26 an effect on Sox10 expression in the cranial NC can be seen but no change in expression of Sox10 is seen in the otic vesicle (Fig. S8). In Leflunomide treated embryos Sox10 expression in the otic vesicle is also not affected while there is a loss of expression in the cranial NC and pigment cells (Fig. S8). Other neural crest markers Slug/*Snail2*, *FoxD3* and *Sox9* show some alteration and partial loss of expression, but very little complete loss of expression (Fig. 3B and Fig. S4). In support of the RNA-Seq data expression of Snail1, which is also an early specifier of NC, was not affected by knockdown of P-TEFb components (Fig. S7). The phenotypes after P-TEFb knock-down were strikingly similar to those obtained after leflunomide treatment (Fig. 2C and D).

c-Myc is also known to regulate the development of trigeminal placodes in the anterior region of the embryo. We therefore looked at the expression of placodal markers by WISH. The placode markers we used include *Tbx2*, *NeuroD* and *ElrD*, which are expressed in the trigeminal placodes and the neural tube (Green and Vetter, 2011; Schlosser and Ahrens, 2004; Showell et al., 2006). We were able to detect a loss of expression of all three markers, specifically in the trigeminal placode region, but not in any of the other tissues in which they are expressed (Fig. 3C and Fig. S4D).

2.6. c-Myc is a paused gene during development

As *c-Myc* is the earliest gene to be affected by Leflunomide and the knockdown of the P-TEFb components, we looked at the RNA pol II occupancy profile of *c-Myc* and compared it to the promoter histone modification H3K4me3 and the processive elongation mark H3K36me3. These data suggest that *c-Myc* is in a paused or poised state during mid to late gastrula just before the NC are specified (Fig. 4A). We next carried out RNA Pol II ChIP-seq on embryos injected with CDK9.S morpholino compared to embryos injected with control morpholino. Knockdown of Cdk9 causes an increase in RNA Pol II occupance, the effect of which is the strongest on the promoter, suggesting that there is an increase in pausing of *c-Myc* (Fig. 4A, blue tracks). We determined the pausing index values for all annotated genes and Fig. 4Bi shows that there is an overall increase in pausing. This would be expected given that transcriptional elongation is effectively being suppressed. Fig. 4Bii highlights that the loss of CDK9 causes an increase in the amount of RNA Pol II bound to the promoter compared to the rest

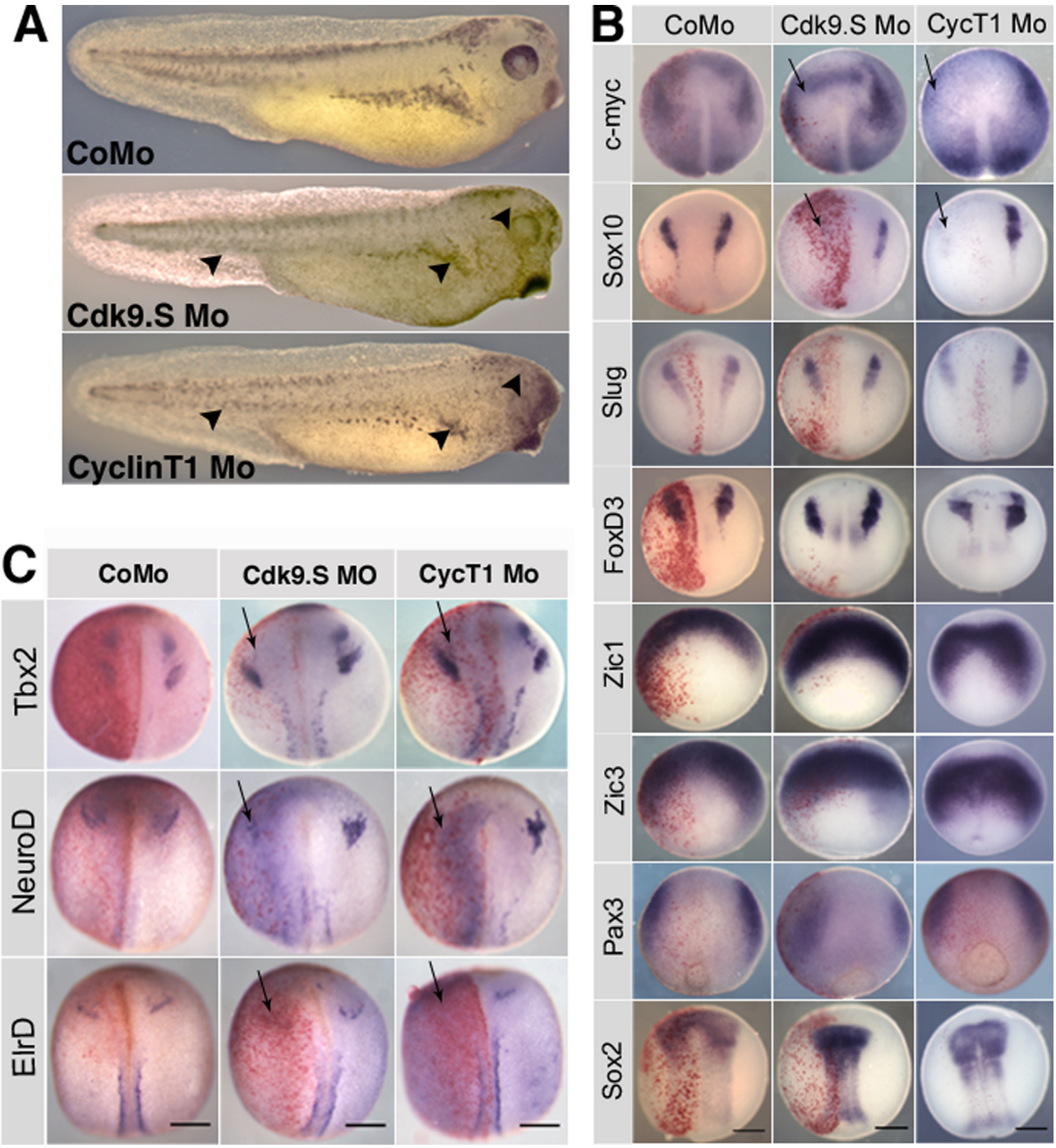


Fig. 3. Knockdown of P-TEFb components leads to loss of expression of neural crest specifying genes. (A) All embryos were injected with 100 ng standard control morpholino, 100 ng CDK9.S morpholino and 60 ng CYCLINT1 morpholino in both cells of a two cell stage embryo and fixed at stage 38. Loss of melanocytes in the tail head and lateral stripe are indicated by black arrow heads. (B) All embryos were injected in one cell of a two cell stage embryo with either a control morpholino CoMo, 100 ng CDK9.S morpholino or 60 ng CYCLINT1 morpholino. 300 pg of *Lac-Z* cRNA was co-injected with morpholinos and β-gal activity was detected using a red gal stain to monitor lineage tracing. *in situ* hybridisation was then carried out on these embryos for neural crest specifying, neural plate border and neural plate genes. *c-Myc* (stage 13) shows tissue specific loss of expression (black arrows). *Sox10* (stage 15) shows loss of expression (black arrows). Slug/*Snail2* expression (stage 15) and *FoxD3* expression (stage 15) appear partially lost or altered. *Zic1* expression (stage 12), *Zic3* expression (stage 12), *Pax3* expression (stage 12) and *Sox2* expression (stage 15) appear unchanged. Quantification of these results can be found in Fig. S5D. Scale bar = 200 μm. (C) Knockdown of P-TEFb components leads to a loss of trigeminal placode markers. All embryos were injected in one cell of a two cell stage embryo with either 100 ng control morpholino (CoMo), 100 ng CDK9.S morpholino or 60 ng CYCLINT1 morpholino. 300 pg of *Lac-Z* cRNA was co-injected with morpholinos and β-gal activity was detected using a red gal stain to monitor lineage tracing. *in situ* hybridisation was then carried out on these embryos for trigeminal placode markers. *Tbx2* expression (stage 18) shows loss of expression specifically in the trigeminal placode region (black arrow). Expression in the otic placode and dorsal root ganglia in not altered. *NeuroD* expression (stage 18) shows specific loss of expression in the trigeminal placode region (black arrow). *ElrD* expression (stage 18) is lost specifically in the trigeminal placode region (black arrow). Expression in the dorsal root ganglia remains unaltered.

of the gene body. The position of c-Myc is shown by the blue spot. While the RPKM on its gene body is increased from 5 to 10, the RPKM on its TSS goes from 11 to 50. This suggests that c-Myc is particularly sensitive to control of expression at the level of transcriptional elongation and is therefore in a more 'paused' state.

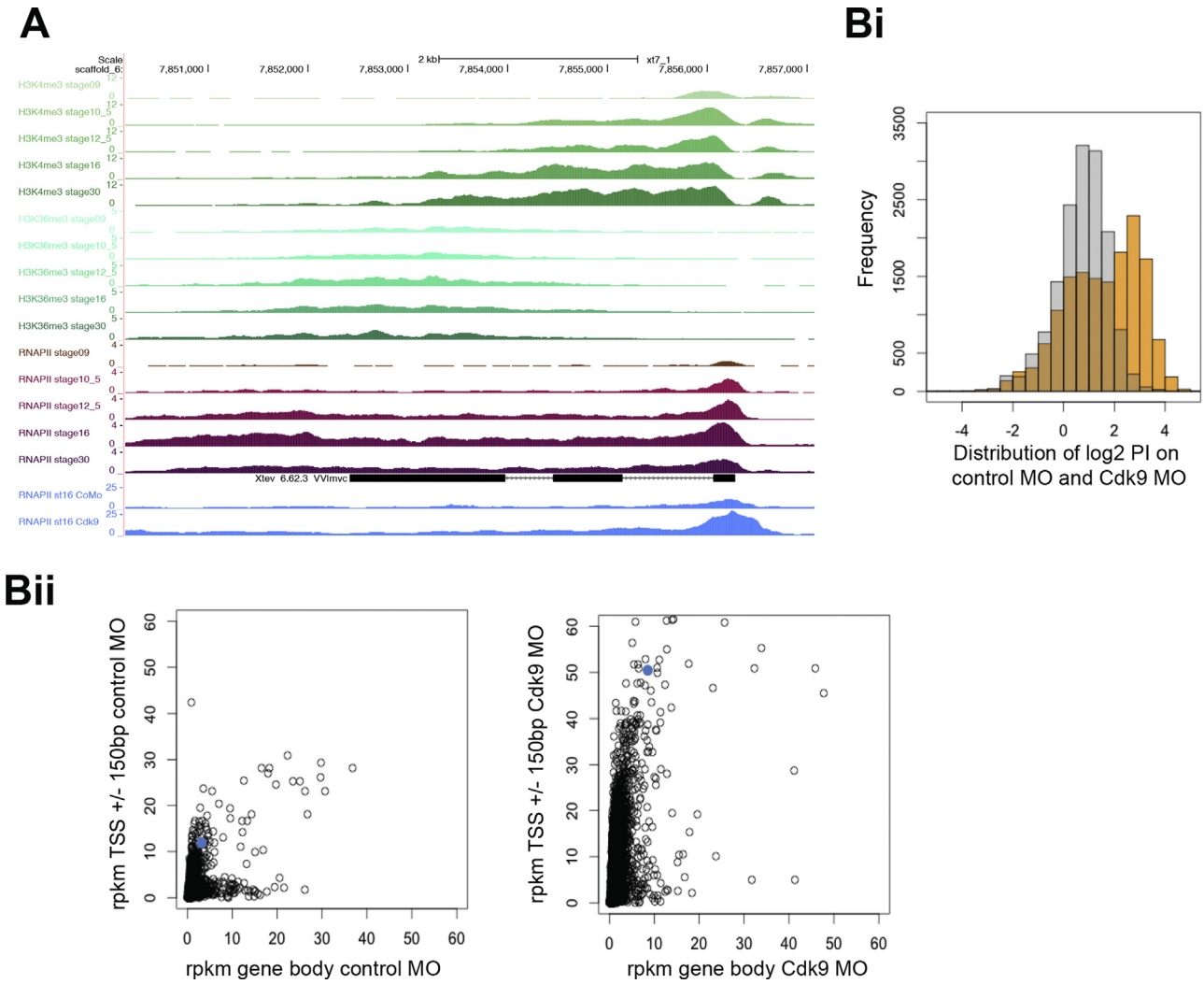


Fig. 4. c-Myc is a paused gene and CDK9 MO causes c-Myc to become more paused during development. (A) Histone methylation and RNA polymerase II (RNAPII) binding at the c-Myc gene. Profiles of H3K4me3 (marking active promoters and 5′ ends of gene bodies, green), H3K36me3 (showing active transcription, light green) and RNA Pol II (RNAPII, purple) (Hontelez et al., 2015) are visualised using the UCSC Genome Browser for the c-Myc gene on scaffold_6:7844903–7863621 (*X. tropicalis* genome assembly version 7.1). The transcriptional profile shows *c-Myc* to be a paused gene in *Xenopus* blastula and gastrula-stage embryos. RNA Pol II profile of the *c-Myc* gene at stage 16 from embryos injected with control MO or Cdk9s MO (blue). Note that the *c-Myc* gene is proportionately more paused in the cdk9 MO treated embryos. (B) RNA Pol II ChiP-seq on whole embryos injected with Control MO or Cdk9.S MO. i) Histogram plots showing the distribution of the Pausing Index values (RPKM TSS $\pm$ 150 bp divided by the RPKM on the gene body) in $\log_2$ space for Control MO-injected (grey) and Cdk9.S MO-injected embryos (orange). The populations were shown to be significantly different from each other (Mann-Whitney P-value $< 2.2 \times 10^{-16}$), indicative of RNAPII redistribution. ii) *c-Myc* is more highly paused in Cdk9 MO treated embryos. Scatter plots of RPKM on the gene body *versus* the transcriptional start site (tss) in Control MO (left panel) and Cdk9.S MO-injected embryos (right panel). The loss of Cdk9 causes an increase in the amount of RNA Pol II bound to the promoter compared to the rest of the gene body. The *c-Myc* gene is shown by the blue spot. Genes with the strongest increase in gene body RNAPII have the lowest increase in PI, suggesting that RNAPII stalls either at the promoter or in the gene body. The effects of RNAPII inhibition on transcript levels is shown in Fig. 3B.

2.7. c-Myc is regulated by transcriptional elongation in neural crest cells

We postulated that *c-Myc* could be the primary gene regulated by transcriptional elongation in neural crest cells. We therefore examined if loss of *Sox10* expression after P-TEFb morpholino injection could be rescued by co-injecting *c-Myc* RNA. For this experiment CDK9.S morpholino and CYCLINT1 morpholinos were injected alone or with 100 pg of *c-Myc* and 500 pg of *c-Myc* RNA. As before, knockdown of CDK9.S and CYCLINT1 led to a high number of embryos displaying loss of *Sox10* expression (Fig. 5A and B). Co-injection of these morpholinos with 100 pg of *c-Myc* resulted in a reduction of the number of embryos displaying a loss of expression phenotype. The majority of embryos displayed a partial loss phenotype or wild type phenotype. After injection of 500 pg of *c-Myc* RNA a greater number of embryos displayed a wild type phenotype (Fig. 5A and B). Injection of 500 pg of *c-Myc* alone had very little effect on *Sox10* expression and injection of the standard control MO did not affect *Sox10* expression (Fig. 5B). Injection of c-Myc RNA into one blastomere was also shown to rescue leflunomide treated embryos on the injected side (Fig. S6).

3. Discussion

The molecular mechanisms controlling NCC specification and differentiation are complex and consist of many layers of regulation to ensure the cells decide upon the correct lineage and form the correct cell types. Other multipotent cell types such as embryonic stem cells and haematopoietic stem cells are thought to undergo regulation at the level of transcriptional elongation (Bai et al., 2010; Lin et al., 2011). It has previously been shown that *Foggy* mutant Zebrafish, which are deficient in the SPT5 subunit of DSIF, display pigmentation and cardiac defects suggesting a

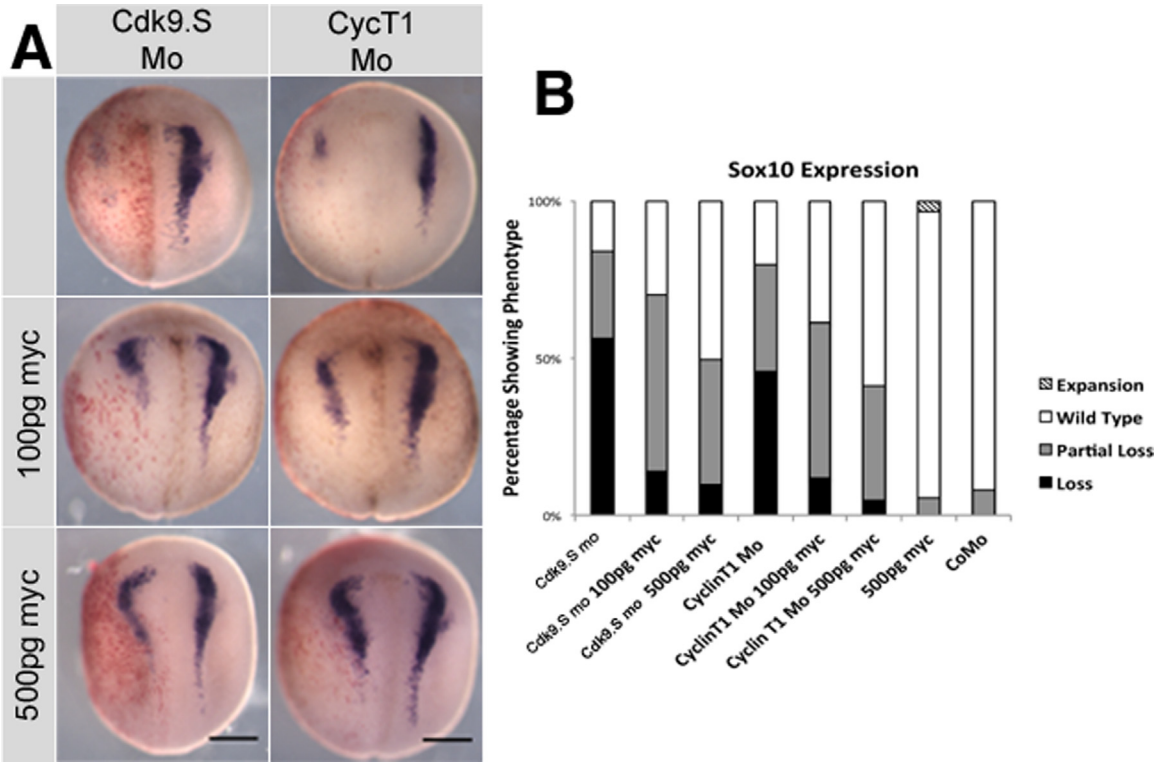


Fig. 5. c-Myc can rescue the effect of Cdk and CyclinT1 morpholino knockdown (A) Embryos were co-injected into one cell of a two cell stage embryo with a morpholino and *c-Myc* RNA. 300 pg of *Lac-Z* cRNA was co-injected with morpholinos and RNA. β-gal activity was detected using a red gal stain to monitor lineage tracing. Embryos then underwent *Sox10 in situ*. Injection of 100 ng Cdk9.S morpholino and 60 ng CyclinT1 morpholino results in a loss of *Sox10* expression on the injected side. Co-injection of Cdk9.S and CyclinT1 morpholino with 100 pg *c-Myc* RNA results in a partial rescue of *Sox10* expression. Co-injection of Cdk9.S and CyclinT1 morpholino with 500 pg *c-Myc* RNA results in a greater rescue of *Sox10* expression. Scale bar=200 μm. These results are quantified in Fig. S6D. (B) Graph showing the percentage of embryos displaying an expansion, wild type, partial loss or loss of *Sox10* expression pattern by *in situ* hybridisation after injection of Cdk9.S morpholino alone (n=51) and co-injected with 100 pg (n=114) and 500 pg of *c-Myc* RNA (n=61). CyclinT1 morpholino alone (n=46) and co-injected with 100 pg (n=61) and 500 pg of *c-Myc* RNA (n=52). 500 pg of *c-Myc* RNA alone (n=34) and 100 ng of a standard control morpholino alone (n=33).

disruption in neural crest development (Keegan et al., 2002). In addition the *Xenopus tropicalis* mutant cyd vicious, which displays pigmentation defects, is also mutated in the *Spt5* gene (Lyle Zimmerman, personal communication). Here we provide evidence for direct regulation of the specification of NC at the level of transcriptional elongation by the P-TEFb complex. In addition, we show that in the NC, *c-Myc* is the crucial gene regulated at the level of transcriptional elongation.

The methods used in this study were firstly to use leflunomide as an inhibitor of transcriptional elongation and secondly, through the use of morpholino knockdowns of components of the P-TEFb complex to block transcriptional elongation. Both of these methods gave comparable results (Figs. 2 and 3, Fig. S1–4). Both demonstrate a disruption in the development of neural crest derivatives including melanophores, sensory neurons and cranio-facial cartilage. When looking at the levels of expression of genes involved in the neural crest gene regulatory network the two genes demonstrating the greatest level of knockdown by WISH, qRT-PCR or RNAseq were *c-Myc,* an early NC specifier and *Sox10,* a NC specifier downstream of *c-Myc* also expressed in the migrating NC. Furthermore, we showed that the loss of *Sox10* gene expression in the NC was rescued by injecting embryos with *c-Myc* RNA (Fig. 5).

To investigate whether *c-Myc* expression was directly regulated by the P-TEFb complex and transcriptional elongation, we looked at cell types other than neural crest cells, which are known to have *c-Myc* within their gene regulatory network. We reasoned that if *c-Myc* is a direct target a loss in expression of markers for these other cell types should be observed. Our *in situ* hybridisation data suggested that *c-Myc* expression in the placodal regions is lost after knockdown of P-TEFb components (Fig. 3B), as opposed to more posterior neural expression, which was not affected. We therefore looked at the trigeminal placodes, which are similar to NC and which express *c-Myc* in the early stages of their gene regulatory network (Bellmeyer et al., 2003; Grocott et al., 2012). We were able to detect a loss of expression of three markers, specifically in the trigeminal placode region, but not in any of the other tissues in which they are expressed (Fig. 3C and Fig. S4D). This provides further evidence for *c-Myc* itself being directly regulated by transcriptional elongation in specific tissues. Interestingly we also see a loss of Rohon-Beard cell markers (Runx1, Islet1 and Ngnr1) upon knockdown of the P-TEFb complex (Fig. S4E). It is currently not known whether c-Myc signalling is required for the development of these cells. The combination of the loss of neural crest cells, trigeminal placodes and Rohon-Beard cells causes embryos treated with leflunomide or CDK9 morpholino to become insensitive to touch response assays as shown in (Fig. S3C and Supplementary movies 1–3).

Our ChIP-Seq analysis (Fig. 4) shows c-Myc to be a paused gene during early development and that the inhibition of transcriptional elongation by Cdk9 morpholino causes *c-Myc* to become even more paused compared to the rest of the genome. This is consistent with studies using human 293 T embryonic kidney cell lines, which have also shown *c-Myc* to be a paused gene (Takahashi et al., 2011). Rahl et al. (2010) have previously shown in tissue culture that c-MYC can directly regulate the activity of P-TEFb. This suggests the activation of *c-Myc* by signals such as WNT (Bellmeyer et al., 2003) could lead to a positive feedback system where c-MYC activates P-TEFb which in turn promotes

transcriptional elongation of the *c-Myc* gene. Our rescue experiments in Fig. 5 and Fig. S6 suggests c-Myc overexpression can overcome the effect of loss of P-TEFb activity. However we cannot exclude the possibility that the increased amount of c-MYC in these embryos may increase any residual P-TEFb activity to overcome the loss of activity from the MO. Either way it still underlines the crucial role for regulation at the level of transcriptional elongation and the role of *c-Myc* in this process.

Our RNA-Seq results using animal cap explants show that when the three cell types NC, neuroectoderm and ectoderm have inhibited transcriptional elongation the only cell type to show specific changes was the NC. We show down regulation of *c-Myc* expression in the neural crest sample and observed little change in the neural and ectoderm samples (Supplementary data 1). This gives us further reason to suggest that regulation of *c-Myc* expression by transcriptional elongation is occurring only in specific tissues, in this case the NC.

Overall, these results provide evidence that *c-Myc* is the 'gatekeeper' regulated by transcriptional elongation in both, the NC and trigeminal placodes, and that expression of the downstream marker *Sox10* in the NC is lost as a secondary effect of this. This is consistent with previous analysis of microarrays from melanoma tumours where many *c-Myc* responsive genes were affected by leflunomide treatment (White et al., 2011). Our results are summarised as a model in Fig. 6. We propose that the *c-Myc* gene is held in a paused state within NCCs and not in all tissues. This paused state is then specifically released in a wild type NC cell by the P-TEFb complex to allow normal *c-Myc* transcription and therefore promote synchronous transcription of downstream genes such as *Sox10* and the subsequent differentiation of neural crest derivatives (Fig. 6 top route). MYC is known to actually stimulate recruitment of P-TEFb to promoters of active genes thus leading to increased expression in a process called transcriptional amplification (Rahl and Young, 2014). The results obtained from our P-TEFb knockdown studies and leflunomide treated embryos suggest that when the function of P-TEFb is impaired *c-Myc* is no longer expressed at a high enough level and this in turn leads to incorrect expression of downstream NC specifiers such as *Sox10* and the synchronicity of maintaining multipotency, proliferation and differentiation of neural crest is lost (Fig. 6 bottom route). Recently it has been shown that regulation of transcriptional elongation to be important for genes regulating cell cycle and signal transduction (Williams et al., 2015). *c-Myc* is an important regulator for such processes and as we show here it is clearly important in the regulation of the neural crest.

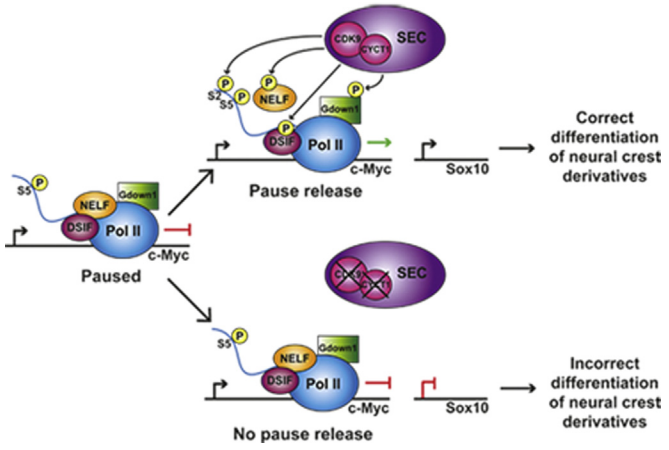


Fig. 6. Model displaying the mechanism by which *c-Myc* is paused in neural crest cells. On the left *c-Myc* is shown in a paused state with Pol ll bound to NELF, DSIF and Gdown1. We have shown that a knockdown of the P-TEFb components CDK9 and CYCLINT1 leads to a loss of *c-Myc* expression and a subsequent loss of *Sox10* expression and incorrect development of neural crest derivatives (bottom scenario). This suggests in a wild type neural crest cell, P-TEFb is promoting transcriptional elongation of *c-Myc*. P-TEFb in its active form bound to the SEC will phosphorylate DSIF, NELF, Gdown1 and serine2 of the C terminal of RNA Pol ll. This promotes pause release on the *c-Myc* gene and allows the subsequent expression of *Sox10* leading to correct synchronous development of neural crest derivatives (top scenario). S2=Serine2, S5=Serine5, P=phosphorylation mark, NELF=Negative elongation factor, DSIF=DRB sensitivity inducing factor, Pol ll=RNA polymerase II, CDK9=Cyclin dependent kinase 9, SEC=Super elongation complex.

To conclude, we provide evidence that the P-TEFb activation of transcriptional elongation is an important step regulating the correct expression of neural crest genes and that this is mediated through the action of *c-Myc*. It is possible that a polymerase pausing step is acting to prime *c-Myc* for transcription to ensure the correct temporal expression of NC genes and therefore a synchronous development of this cell type into its derivatives. Polymerase pausing has previously been shown to be important in cells, which undergo synchronous gene expression and development. Compared to stochastic uncoordinated gene expression in the whole embryo, the development of the NC requires such synchronous gene expression (Boettiger and Levine, 2009). *c-Myc* has been shown here to be particularly sensitive to inhibition of transcriptional elongation and so it is possible that *c-Myc* is a paused gene in prospective NC cells (Bai et al., 2010). There has been much debate in recent literature as to whether this phenomenon of RNA polymerase pausing and regulation of transcriptional elongation is (1) common to all genes and (2) whether it might be specific to certain cell types (Luo et al., 2012). Our data contributes to this debate by providing evidence that regulation of transcriptional elongation is focussed on c-Myc and is particularly important in developing neural crest.

4. Materials and methods

4.1. Embryos, capped RNA synthesis, morpholinos, microinjection and animal capping

All experiments were performed in compliance with the relevant laws and institutional guidelines at the University of East Anglia. The research has been approved by the local ethical review committee according to UK Home Office regulations. *Xenopus laevis* embryos were obtained as previously described (Harrison et al., 2004). Embryos were staged using the Nieuwkoop and Faber normal table of *Xenopus* development (Nieuwkoop, 1994). Full length cDNAs for *Xenopus laevis CDK9.S* (IMAGE:5156812) and *CyclinT1* (IMAGE:7207389) were obtained from the IMAGE Clone collection. CDK9.L was obtained from Dr Paul Mead. The cDNA clones were all completely sequenced before subcloning into relevant vectors. Capped RNA was synthesised using the SP6 mMESSAGE mMACHINE™ (Ambion) kit for *Wnt1*, *Noggin*, *c-Myc* (from Carole Labonne), *LacZ* and mutagenesis rescue constructs for *CDK9.S*, *CDK9.L* and *CyclinT1*. Mutagenesis was carried out using the QuickChange II XL Site-Directed Mutagenesis kit (Agilent Technologies). Morpholino oligos (MOs) were designed and made by Gene Tools. Morpholino sequences are: Standard control (CoMO) 5′-CCTCTTACCTCAGTTACAATTTATA-3′, *CDK9.S* 5′-ATGGCCAAGAACTACGACTCGGTGG-3′, *CDK9.L* 5′-ATGGTAAAGAACTACGACTCTGTTG-3′ and *CyclinT1* 5′-GACATCGTACTTTATGGCAACAAAC-3′. These were tested by *in vitro* translation using the TNT® Coupled Reticulocyte Lysate System (Promega) following manufacturers instructions. Capped RNA and morpholinos were co-injected with 300 pg *LacZ*. Microinjections were carried out in 3% Ficoll PM400 (Sigma). Embryos were injected at the 2 cell stage in either 1 or both cells. β-Galactosidase activity was detected by colorimetric reaction using Red-Gal (Sigma). Embryos were staged and fixed for 2 h at room temperature in MEMFA. Animal capping

was carried out as previously described (Garcia-Morales et al., 2009).

4.2. Wholemount in situ hybridisation

Single and double *in situ* hybridisation was carried out as previously described (Harrison et al., 2004). Probes were synthesised for *c-Myc* (IMAGE:4884683), *Sox10*, *Zic1*, *Zic3*, *Pax3*, *Sox2*, Slug/*Snail2*, *FoxD3*, *Sox9*, *CDK9.S*, *CDK9.L* and *CyclinT1*.

4.3. Cryosectioning

Embryos were fixed in MEMFA for 1 h, washed twice for 5 min in PBST and then put in 30% sucrose for 4 h at room temperature. Embryos were then transferred to cryomoulds filled with optimal cutting temperature (OCT) compound and left at 4 °C overnight. Embryos were then positioned appropriately for sectioning and left at −20 °C overnight. Embryos were sectioned using the LEICA CM 1950 Cryostat and sections were placed on 5% TESPA slides. Slides were then washed for 5 min in PBST and covered with a coverslip using hydromount. Images were taken using a Zeiss CCD upright microscope with colour camera.

4.4. Alcian blue staining

Embryos were fixed for 1 h at room temperature, dehydrated with 5 washes of 100% ethanol (EtoH) for 5 min each at RT and then left in Alcian blue staining solution for 3 nights. After this they were washed 3 times for 15 min in 95% ethanol at RT then rehydrated in 2% KOH. This was done gradually using 10 min washes of 75% EtoH in 2% KOH, 50% EtoH in 2% KOH, 25% EtoH in 2% KOH then 3 2% KOH washes. Embryos were then stored in glycerol to make embryos more transparent. Facial cartilage was then dissected out before imaging.

4.5. RNA extraction, PCR, Q-PCR and RNA sequencing

For the PCR, Q-PCR and RNA-sequencing experiments animal caps were cut at stage 8 and allowed to develop to stage 15. Approximately 200 caps were required for each condition tested to generate enough RNA for sequencing. Caps were collected in batches of approx. 30 and RNA purified. Each batch was tested by RT_PCR or QPCR to make sure the caps were expressing the relevant NC, neural or ectodermal markers. Total RNA was extracted using TRIzol® (15596-026, Life technologies, UK) and DNAse treated. Reverse transcription was performed using SuperScript® II reverse transcriptase (18064-014, Invitrogen, CA, USA), with 1 μg total RNA and random primers (C1181, Promega, Southampton, UK). Brightwhite 96 well plates (BW-FAST) and high quality optical plate seals (BW-ADVSEAL) were used (PrimerDesign Ltd, Southampton, UK). Gene-specific nucleotide sequences were detected using Precision™ FAST-LR 2× mastermix (PrimerDesign Ltd, Southampton, UK). Quantitative RT-PCR was performed in a final volume of 20 μL using a 7500 FAST real-time PCR instrument (Roche) under the following cycling conditions: 95 °C for 20 s, 40 cycles at 95 °C for 3 s, 60 °C for 30 s. After cycling, a melting curve was recorded between 60 °C and 95 °C under the following 95 °C for 15 s, 60 °C for 1 min, 95 °C for 15 s, 60 °C for 15 s with a ramp rate of 0.11 °C s^{-1}. Detection primers were designed using Primer3 software (http://biotools.umassmed.edu/bioapps/primer3_www.cgi) and custom designed by PrimerDesign Ltd, UK where indicated (Rozen and Skaletsky, 2000). Their oligonucleotide sequences are provided in Supplementary Table 1. In-house designed primers were synthesised by eurofins (Norwich, UK) and custom primers were designed by and purchased from PrimerDesign Ltd (PrimerDesign Ltd, Southampton, UK).

For the RNA seq experiments to determine whether we get a similar loss of neural crest specifying genes compared to the whole embryo data, the animal cap samples were initially tested by *in situ* hybridisation and RT-PCR. RNA sequencing was carried out to investigate which genes were affected by leflunomide across the whole genome in the three sample types.

For RNA sequencing total RNA was fragmented before undergoing cDNA library preparation. cDNA was end-repaired, A-tailed and adapter-ligated before amplification and size selection. The prepared libraries undergo quality control by running on a gel and using a bioanalyser to check library fragment size distribution before 100 bp paired end sequencing over one lane of a flow cell. Samples were multiplexed 2 × 3 and ran on one lane each to yield approximately 25 million paired reads per sample. Reads were aligned to the *Xenopus laevis* transcriptome (xlaevisMRNA.fasta) downloaded from Xenbase (release 28/09/12) (Bowes et al., 2008; Langmead and Salzberg, 2012). Mapping was carried out with bowtie2 on paired-end samples using default parameters. A custom Perl script was used to parse the bowtie results and count the number of sequences in each sample mapping to every mRNA sequence in the transcriptome file. The table of raw abundances for each mRNA was then read into R and differential expression between untreated and leflunomide treated samples was calculated using the DEseq package (Anders and Huber, 2010). Any transcript with a p-value of ≤ 0.05 was classed as being differentially expressed between untreated and leflunomide treated samples. The data has been uploaded to the NCBI SRA, accession number PRJNA325078.

The data for quantitative RT-PCR experiments were analysed using the 7500FAST software. Measured Ct values for quantitative RT-PCR technical and biological replicates for each gene of interest were analysed using the NORMA-gene algorithm (Heckmann et al., 2011). NORMA-gene reduces systematic and artificial between-replicate bias utilising the entire data-set of the target genes being studied. NORMA-gene is applicable to small data sets greater than five genes of interest and is used as our primary analysis tool as it produces equal or better normalisation compared to the delta-delta ct method. Leflunomide affects gene transcription thus impacting upon metabolic and neural crest pathways. Therefore, using reference *vs.* target normalisation is not appropriate to measure gene expression.

4.6. ChIP-sequencing

Morpholinos were injected at 2 cell stage embryos and incubated at 26 °C in 0.05× Marc's Modified Ringer's (MMR). Approximately 50 embryos per sample were fixed at stage 16 in 1% formaldehyde for 30 min. Samples were then washed in 125 mM glycine solution 30 min and then transfer gradually to 0.1× MMR. Embryos were homogenized in sonication buffer (20 mM TrisHCl pH=8, 70 mM KCl, 1 mM EDTA, 10% glycerol, 5 mM DTT, 0.125% NP40, protease inhibitors (Roche #04693132001) and stored at −80 °C. ChIP was prepared as previously described (Jallow et al., 2004). After sonication, chromatin samples were incubated with RNA Pol II antibody (diagenode C15200004) 1 μg per 15 embryo equivalent in IP buffer overnight (50 mM TrisHCl pH=8, 100 mM NaCl, 2 mM EDTA, 1%NP40, 1 mM DTT and protease inhibitors) followed by incubation with Dynabeads Protein G for 1 h. Beads where subsequently washed in ChIP 1 buffer (IP buffer +0.1% deoxycholate), ChIP 2 buffer (ChIP 1 buffer +400 mM NaCl), ChIP 3 buffer (ChIP 1 buffer +250 mM LiCl) and TE buffer. Material was digested with proteinase K and DNA was purified using a QIAGEN kit. Q-PCR was performed pre and post library preparation for quality checking and DNA libraries were prepared using Kapa Hyper Prep kit (Kapa Biosysistem). Sequencing was done on the Illumina HiSeq2000 platform and reads were mapped to the

reference *X. tropicalis* genome JGI7.1 using BWA allowing one mismatch. The CDK9 MO tracks are uploaded on the GEO database and have GEO accession number GSE82153.

4.7. Small molecule compound treatment

The small molecule compound leflunomide (Sigma) was dissolved in DMSO to make a 10 mM stock. This was then diluted directly into the embryos media 0.1 × MMR. DMSO was added to the media as a control (Tomlinson et al., 2012).

Author contributions

VH, CF, NW, MT, MM, ID, SH, IK and AH performed experiments. VH, CF, SM, MT, GV, MM, AM and GW designed experiments, discussed and analysed data. GW directed the research. VH and GW wrote the manuscript.

Acknowledgements

We would like to thank members of the Münsterberg and Wheeler labs past and present for their help and support during this project. We would also like to thank Stefan Hoppler, Tim Grocott and Richard White for helpful discussions, Paul Mead for sending us his *Cdk9* plasmid and other colleagues for various plasmids. VH and GW acknowledge the help provided by the Cold Spring Harbor *Xenopus* course. This work was supported by a BBSRC, United Kingdom Grant (BB/I022252) to GW, a UEA Dean's studentship to VH, by the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme FP7 under REA Grant agreement number 607142 (DevCom) to MM and ID and by the UEA, United Kingdom, the John Innes Centre, United Kingdom and AstraZeneca, United Kingdom (R18873) to AH.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.ydbio.2016.06.012.

References

Anders, S., Huber, W., 2010. Differential expression analysis for sequence count data. Genome Biol. 11, R106.
Andrulis, E.D., Guzman, E., Doring, P., Werner, J., Lis, J.T., 2000. High-resolution localization of Drosophila Spt5 and Spt6 at heat shock genes in vivo: roles in promoter proximal pausing and transcription elongation. Genes Dev. 14, 2635–2649.
Bai, X., Kim, J., Yang, Z., Jurynec, M.J., Akie, T.E., Lee, J., LeBlanc, J., Sessa, A., Jiang, H., DiBiase, A., Zhou, Y., Grunwald, D.J., Lin, S., Cantor, A.B., Orkin, S.H., Zon, L.I., 2010. TIF1gamma controls erythroid cell fate by regulating transcription elongation. Cell 142, 133–143.
Bellmeyer, A., Krase, J., Lindgren, J., LaBonne, C., 2003. The protooncogene c-myc is an essential regulator of neural crest formation in xenopus. Dev. Cell 4, 827–839.
Boettiger, A.N., Levine, M., 2009. Synchronous and stochastic patterns of gene activation in the Drosophila embryo. Science 325, 471–473.
Bowes, J.B., Snyder, K.A., Segerdell, E., Gibb, R., Jarabek, C., Noumen, E., Pollet, N., Vize, P.D., 2008. Xenbase: a Xenopus biology and genomics resource. Nucleic Acids Res. 36, D761–D767.
Bres, V., Yoh, S.M., Jones, K.A., 2008. The multi-tasking P-TEFb complex. Curr. Opin. Cell Biol. 20, 334–340.
Buitrago-Delgado, E., Nordin, K., Rao, A., Geary, L., LaBonne, C., 2015. NEURODEVELOPMENT. Shared regulatory programs suggest retention of blastula-stage potential in neural crest cells. Science 348, 1332–1335.
Cheng, B., Li, T., Rahl, P.B., Adamson, T.E., Loudas, N.B., Guo, J., Varzavand, K., Cooper, J.J., Hu, X., Gnatt, A., Young, R.A., Price, D.H., 2012. Functional association of Gdown1 with RNA polymerase II poised on human genes. Mol. Cell 45, 38–50.
Core, L.J., Lis, J.T., 2008. Transcription regulation through promoter-proximal pausing of RNA polymerase II. Science 319, 1791–1792.
Elworthy, S., Lister, J.A., Carney, T.J., Raible, D.W., Kelsh, R.N., 2003. Transcriptional regulation of mitfa accounts for the sox10 requirement in zebrafish melanophore development. Development 130, 2809–2818.
Fuda, N.J., Ardehali, M.B., Lis, J.T., 2009. Defining mechanisms that regulate RNA polymerase II transcription in vivo. Nature 461, 186–192.
Fujita, T., Piuz, I., Schlegel, W., 2009. The transcription elongation factors NELF, DSIF and P-TEFb control constitutive transcription in a gene-specific manner. FEBS Lett. 583, 2893–2898.
Gaertner, B., Zeitlinger, J., 2014. RNA polymerase II pausing during development. Development 141, 1179–1183.
Garcia-Morales, C., Liu, C.H., Abu-Elmagd, M., Hajihosseini, M.K., Wheeler, G.N., 2009. Frizzled-10 promotes sensory neuron development in Xenopus embryos. Dev. Biol. 335, 143–155.
Green, Y.S., Vetter, M.L., 2011. EBF factors drive expression of multiple classes of target genes governing neuronal development. Neural Dev. 6, 19.
Grocott, T., Tambalo, M., Streit, A., 2012. The peripheral sensory nervous system in the vertebrate head: a gene regulatory perspective. Dev. Biol. 370, 3–23.
Harrison, M., Abu-Elmagd, M., Grocott, T., Yates, C., Gavrilovic, J., Wheeler, G.N., 2004. Matrix metalloproteinase genes in Xenopus development. Dev. Dyn. 231, 214–220.
Heckmann, L.H., Sorensen, P.B., Krogh, P.H., Sorensen, J.G., 2011. NORMA-Gene: a simple and robust method for qPCR normalization based on target gene data. BMC Bioinform. 12, 250.
Hochheimer, A., Tjian, R., 2003. Diversified transcription initiation complexes expand promoter selectivity and tissue-specific gene expression. Genes Dev. 17, 1309–1320.
Hontelez, S., van Kruijsbergen, I., Georgiou, G., van Heeringen, S.J., Bogdanovic, O., Lister, R., Veenstra, G.J., 2015. Embryonic transcription is controlled by maternally defined chromatin state. Nat. Commun. 6, 10148.
Hoppler, S., Wheeler, G.N., 2015. DEVELOPMENTAL BIOLOGY. It's about time for neural crest. Science 348, 1316–1317.
Huang, X., Saint-Jeannet, J.P., 2004. Induction of the neural crest and the opportunities of life on the edge. Dev. Biol. 275, 1–11.
Jallow, Z., Jacobi, U.G., Weeks, D.L., Dawid, I.B., Veenstra, G.J., 2004. Specialized and redundant roles of TBP and a vertebrate-specific TBP paralog in embryonic gene regulation in Xenopus. Proc. Natl. Acad. Sci. USA 101, 13525–13530.
Jonkers, I., Lis, J.T., 2015. Getting up to speed with transcription elongation by RNA polymerase II. Nat. Rev. Mol. Cell Biol. 16, 167–177.
Kee, Y., Bronner-Fraser, M., 2005. To proliferate or to die: role of Id3 in cell cycle progression and survival of neural crest progenitors. Genes Dev. 19, 744–755.
Keegan, B.R., Feldman, J.L., Lee, D.H., Koos, D.S., Ho, R.K., Stainier, D.Y., Yelon, D., 2002. The elongation factors Pandora/Spt6 and Foggy/Spt5 promote transcription in the zebrafish embryo. Development 129, 1623–1632.
Kohoutek, J., 2009. P-TEFb-the final frontier. Cell Div. 4, 19.
Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359.
Light, W., Vernon, A.E., Lasorella, A., Iavarone, A., LaBonne, C., 2005. Xenopus Id3 is required downstream of Myc for the formation of multipotent neural crest progenitor cells. Development 132, 1831–1841.
Lin, C., Garrett, A.S., De Kumar, B., Smith, E.R., Gogol, M., Seidel, C., Krumlauf, R., Shilatifard, A., 2011. Dynamic transcriptional events in embryonic stem cells mediated by the super elongation complex (SEC). Genes Dev. 25, 1486–1498.
Lis, J.T., Mason, P., Peng, J., Price, D.H., Werner, J., 2000. P-TEFb kinase recruitment and function at heat shock loci. Genes Dev. 14, 792–803.
Luo, Z., Lin, C., Shilatifard, A., 2012. The super elongation complex (SEC) family in transcriptional control. Nat. Rev. Mol. Cell Biol. 13, 543–547.
Milet, C., Monsoro-Burq, A.H., 2012. Neural crest induction at the neural plate border in vertebrates. Dev. Biol. 366, 22–33.
Nieuwkoop, P., Faber, J., 1994. Normal Table of *Xenopus laevis* (Daudin). Garlan Publishing, New York.
Pegoraro, C., Monsoro-Burq, A.H., 2013. Signaling and transcriptional regulation in neural crest specification and migration: lessons from xenopus embryos. Wiley Interdiscip. Rev. Dev. Biol. 2, 247–259.
Rahl, P.B., Young, R.A., 2014. MYC and transcription elongation. Cold Spring Harb. Perspect. Med. 4, a020990.
Rahl, P.B., Lin, C.Y., Seila, A.C., Flynn, R.A., McCuine, S., Burge, C.B., Sharp, P.A., Young, R.A., 2010. c-Myc regulates transcriptional pause release. Cell 141, 432–445.
Rozen, S., Skaletsky, H., 2000. Primer3 on the WWW for general users and for biologist programmers. Methods Mol. Biol. 132, 365–386.
Sauka-Spengler, T., Bronner-Fraser, M., 2008. A gene regulatory network orchestrates neural crest formation. Nat. Rev. Mol. Cell Biol. 9, 557–568.
Sauka-Spengler, T., Meulemans, D., Jones, M., Bronner-Fraser, M., 2007. Ancient evolutionary origin of the neural crest gene regulatory network. Dev. Cell 13, 405–420.
Schlosser, G., Ahrens, K., 2004. Molecular anatomy of placode development in *Xenopus laevis*. Dev. Biol. 271, 439–466.
Showell, C., Christine, K.S., Mandel, E.M., Conlon, F.L., 2006. Developmental expression patterns of Tbx1, Tbx2, Tbx5, and Tbx20 in Xenopus tropicalis. Dev. Dyn. 235, 1623–1630.
Takahashi, H., Parmely, T.J., Sato, S., Tomomori-Sato, C., Banks, C.A., Kong, S.E., Szutorisz, H., Swanson, S.K., Martin-Brown, S., Washburn, M.P., Florens, L., Seidel, C.W., Lin, C., Smith, E.R., Shilatifard, A., Conaway, R.C., Conaway, J.W., 2011. Human mediator subunit MED26 functions as a docking site for transcription elongation factors. Cell 146, 92–104.
Tomlinson, M.L., Field, R.A., Wheeler, G.N., 2005. Xenopus as a model organism in

developmental chemical genetic screens. Mol. Biosyst. 1, 223–228.
Tomlinson, M.L., Hendry, A.E., Wheeler, G.N., 2012. Chemical genetics and drug discovery in Xenopus. Methods Mol. Biol. 917, 155–166.
Tomlinson, M.L., Rejzek, M., Fidock, M., Field, R.A., Wheeler, G.N., 2009b. Chemical genomics identifies compounds affecting *Xenopus laevis* pigment cell development. Mol. Biosyst. 5, 376–384.
Tomlinson, M.L., Guan, P., Morris, R.J., Fidock, M.D., Rejzek, M., Garcia-Morales, C., Field, R.A., Wheeler, G.N., 2009a. A chemical genomic approach identifies matrix metalloproteinases as playing an essential and specific role in Xenopus melanophore migration. Chem. Biol. 16, 93–104.
Wada, T., Takagi, T., Yamaguchi, Y., Ferdous, A., Imai, T., Hirose, S., Sugimoto, S., Yano, K., Hartzog, G.A., Winston, F., Buratowski, S., Handa, H., 1998. DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. Genes Dev. 12, 343–356.
White, R.M., Cech, J., Ratanasirintrawoot, S., Lin, C.Y., Rahl, P.B., Burke, C.J., Langdon, E., Tomlinson, M.L., Mosher, J., Kaufman, C., Chen, F., Long, H.K., Kramer, M., Datta, S., Neuberg, D., Granter, S., Young, R.A., Morrison, S., Wheeler, G.N., Zon, L.I., 2011. DHODH modulates transcriptional elongation in the neural crest and melanoma. Nature 471, 518–522.
Williams, L.H., Fromm, G., Gokey, N.G., Henriques, T., Muse, G.W., Burkholder, A., Fargo, D.C., Hu, G., Adelman, K., 2015. Pausing of RNA polymerase II regulates mammalian developmental potential through control of signaling networks. Mol. Cell 58, 311–322.
Yamaguchi, Y., Takagi, T., Wada, T., Yano, K., Furuya, A., Sugimoto, S., Hasegawa, J., Handa, H., 1999. NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. Cell 97, 41–51.
Zeitlinger, J., Stark, A., Kellis, M., Hong, J.W., Nechaev, S., Adelman, K., Levine, M., Young, R.A., 2007. RNA polymerase stalling at developmental control genes in the *Drosophila melanogaster* embryo. Nat. Genet. 39, 1512–1516.
Zhou, Q., Li, T., Price, D.H., 2012. RNA polymerase II elongation control. Annu. Rev. Biochem. 81, 119–143.

Nanoscale

PAPER

View Article Online
View Journal

Cite this: DOI: 10.1039/c7nr06020c

Unravelling the mechanisms that determine the uptake and metabolism of magnetic single and multicore nanoparticles in a *Xenopus laevis* model†

M. Marín-Barba,‡[a] H. Gavilán,‡[b] L. Gutiérrez,[c] E. Lozano-Velasco,[a] I. Rodríguez-Ramiro,[d] G. N. Wheeler,[a] C. J. Morris,[e] M. P. Morales[b] and A. Ruiz*[e]

Multicore superparamagnetic nanoparticles have been proposed as ideal tools for some biomedical applications because of their high magnetic moment per particle, high specific surface area and long term colloidal stability. Through controlled aggregation and packing of magnetic cores it is possible to obtain not only single-core but also multicore and hollow spheres with internal voids. In this work, we compare toxicological properties of single and multicore nanoparticles. Both types of particles showed moderate *in vitro* toxicity (MTT assay) tested in Hep G2 (human hepatocellular carcinoma) and Caco-2 (human colorectal adenocarcinoma) cells. The influence of surface chemistry in their biological behavior was also studied after functionalization with *O,O'*-bis(2-aminoethyl) PEG (2000 Da). For the first time, these nanoparticles were evaluated in a *Xenopus laevis* model studying their whole organism toxicity and their impact upon iron metabolism. The degree of activation of the metabolic pathway depends on the size and surface charge of the nanoparticles which determine their uptake. The results also highlight the potential of *Xenopus laevis* model bridging the gap between *in vitro* cell-based assays and rodent models for toxicity assessment to develop effective nanoparticles for biomedical applications.

Received 14th August 2017,
Accepted 6th December 2017

DOI: 10.1039/c7nr06020c

rsc.li/nanoscale

Introduction

Iron oxide magnetic nanoparticles (IOMNPs) are extensively studied nowadays for their potential in biomedical applications.[1,2] Their surface chemistry can be modified, adding functionality to the material and enabling their use for gene therapy, tissue regeneration and drug delivery, to specifically target tumours using external magnetic fields.[3] Their magnetic properties can be exploited for magnetic resonance imaging and magnetic-fluid hyperthermia, which has raised hope for improved imaging techniques[4] and cancer treatment.[5,6] However, in spite of their potential, few of these biomaterials have reached clinical practice.[7]

A crucial issue for magnetic nanoparticle utilization in biomedicine and their approval by regulatory agencies depends on their biotransformation and toxicity. The course and fate of the nanoparticles once they deliver their imaging or therapeutic objective needs to be studied. Thus, the metabolism of the particles and degradation by-products must be assessed and assured. Since iron is involved in diverse cellular processes,[8] organisms display mechanisms to transport and store iron in non-toxic forms.[9] Therefore, IOMNPs are predicted to be safely incorporated into biological systems. Increasing evidence demonstrates that IOMNPs trigger iron-coping mechanisms in cells and that the degradation products of these materials are incorporated into normal iron metabolic routes.[10–16] However, nanoparticle coating has been shown to be a determinant of the IOMNPs uptake and degradation pathways.[17,18] In order to establish solid conclusions about IOMNPs toxicity and biodistribution and their cellular effects, it would be ideal to have biologically pertinent models.

Xenopus laevis is an amphibian model ideal to study the course and fate of the nanoparticles as it is an easy and flexible

[a]*School of Biological Sciences, University of East Anglia, Norwich Research Park, Norwich, Norfolk, NR4 7TJ, UK*
[b]*Instituto de Ciencia de Materiales de Madrid, Sor Juana Inés de la Cruz 3, 28049 Madrid, Spain*
[c]*Department Química Analítica, Instituto de Nanociencia de Aragón, Universidad de Zaragoza and CIBER-BBN, 50018 Zaragoza, Spain*
[d]*School of Medicine, University of East Anglia, Norwich Research Park, Norwich, Norfolk, NR4 7TJ, UK*
[e]*School of Pharmacy, University of East Anglia, Norwich Research Park, Norwich, Norfolk, NR4 7TJ, UK. E-mail: G.Ruiz-Estrada@uea.ac.uk*
†Electronic supplementary information (ESI) available. See DOI: 10.1039/c7nr06020c
‡These authors contributed equally to this work.

system in which to evaluate vertebrate embryology, basic cell and molecular biology, genomics, neurobiology and toxicology.[19] Recently, this model was used to identify the risk of exposure to contaminated water.[20] The mortality, malformations and growth inhibition of *Xenopus laevis* were studied, after challenge with metal oxide-based nanomaterial contaminants (γ-Fe_2O_3, TiO_2, ZnO and CuO). Notably, γ-Fe_2O_3 did not pose risks to amphibian populations up to 1 mg mL^{-1} where it causes developmental abnormalities. In contrast, the other nanoparticles caused gastrointestinal, spinal, and other abnormalities at concentrations of ~3 mg L^{-1}. Furthermore, it was shown that nanoparticle sizes above 200 nm had toxic effects.[21] Despite available literature on the effects of many compounds on the larval development of *X. laevis* in environmental studies, there are only few reports evaluating nanoparticles toxicity and biodistribution designed for biomedical applications,[22,23] and there is still a lack of knowledge bridging biotransformation studies in cell-based assays with data generated from rodent *in vivo* systems.

In this work, we report the effect of γ-Fe_2O_3 magnetic single and multicore nanoparticles suitable for bioapplications in *X. laevis* embryos. The particles have uniform size in the nanoscale and are coated with biocompatible shells. IOMNPs colloids used in this work can be classified as single-core (with only one magnetic core per particle) and multicore (with several magnetic cores per particle).[24] Single-core superparamagnetic nanoparticles have been proposed as ideal tools in biomedicine, since they display longer circulation times post-injection.[25] Indeed, this may favour their uptake in leaky vasculature regions such as tumors.[26] In contrast, for some medical applications, such as bioseparation or magnetic fluid hyperthermia, it can be advantageous to use larger multicore particles that have a large magnetic moment per particle.[27] Few *in vitro* and *in vivo* studies have been dedicated to the comparison of single-core and multicore nanoparticles[28] and there is a lack of knowledge still on how the aggregation of magnetic cores forming multicore nanoparticles affects the nanoparticle uptake and degradation.

Here we report for the first-time toxicity and metabolism of single-core and multicore nanoparticles using a combination of *in vitro* cell culture models and an *in vivo X. laevis* embryo model. The application of AC magnetic susceptibility measurements has proven to be an excellent tool to study magnetic nanoparticles in complex matrices, since the diamagnetic signal coming from tissues does not interfere with the superparamagnetic signal of the NPs.[29] This way it was possible to follow the signal of the magnetic nanoparticles in animal models[30–32] and to quantify the iron content. This is one of the greatest challenges in the nanomaterials area nowadays, *i.e.* determining how best to characterize nanoparticles and follow their transformation/degradation.[33] The study of *in vivo* fate of IOMNPs is imperative to develop successful biomedical applications. In this paper, we present for the first time, a study, by means of AC magnetic susceptibility measurements, of the intake of iron-containing particles in *X. laevis* embryos.

Results and discussion

Nanoparticles synthesis and characterization

Two different types of iron oxide nanoparticles were synthesized in this work. Single-core nanoparticles (SC) were obtained by thermal decomposition of the iron(III) oleate precursor in 1-octadecene (Fig. 1A and B). Particles were 13 nm (±1 nm) in diameter, uniform in size (Fig. 1C), relatively round and well dispersed, owing to the presence of oleic acid around the particles. In a different approach, multicore nanoparticles (MC) were obtained by polyol mediated reduction of iron(III) chloride. MC are composed of spherical 142 nm (±23 nm) nanoparticles with a well-defined size and shape (Fig. 1D and F). These MC nanoparticles consist of smaller cores of approximately 10 nm. HRTEM and X-ray diffraction patterns have already been reported.[34] Particle structure and size were selected intentionally, since they are two key parameters that directly influence *in vivo* biological behavior. The size of intravenously injected nanoparticles greatly affects their *in vivo* biodistribution, *e.g.* particles from 60 to 150 nm in size are taken up by the reticuloendothelial system leading to rapid uptake in the liver and spleen in humans. Intravenously injected nanoparticles with diameters of 10–40 nm allow longer blood circulation and can cross capillary walls, and they are often phagocytosed by macrophages which traffic to lymph nodes and bone marrow.[35] However, how these parameters affect the greater picture of toxicity and biodegradability mechanisms is still poorly understood.

Iron oxide nanoparticles obtained by thermal decomposition are hydrophobic. To stabilize them in aqueous media and make them suitable for biological applications, oleic acid on the surface of the nanoparticle was removed with DMSA *via* ligand exchange reaction (SC-DMSA). Polyol mediated processes render hydrophilic nanoparticles along with polyvinylpyrrolidone (PVP40), present in the reaction, however an extra capping agent like citric acid enhances the electrostatic repulsion and facilitates the final dispersion of the MC nanoparticles in aqueous media (MC-Cit). Hydrodynamic sizes are always higher than sizes measured by TEM, indicating the presence of the coating or some degree of agglomeration after surface modification, but both types of particles remain below 200 nm (34 and 181 nm were obtained for SC-DMSA and MC-Cit, respectively), an important requirement for biomedical applications. Both particles have high negative surface charge at pH 7 (Z-potential equal to −38 and −25 mV for SC-DMSA and MC-Cit, respectively). In order to evaluate the influence of the surface charge in particle absorption and biodegradation, we conjugated covalently a diamine PEG derivative to the carboxylic groups (from both DMSA and citric acid) to the surface of the nanoparticles. After PEG conjugation, average hydrodynamic size at pH 7 increased from 34 to 65 nm and from 181 to 183 nm for SC-DMSA and MC-Cit, respectively (SC-DMSA-PEG and MC-Cit-PEG) and net surface charge decreased from approximately −38 to −24 mV for SC-DMSA-PEG samples and from −25 to −18 mV for MC-Cit-PEG. Colloidal properties of aqueous suspensions of the nano-

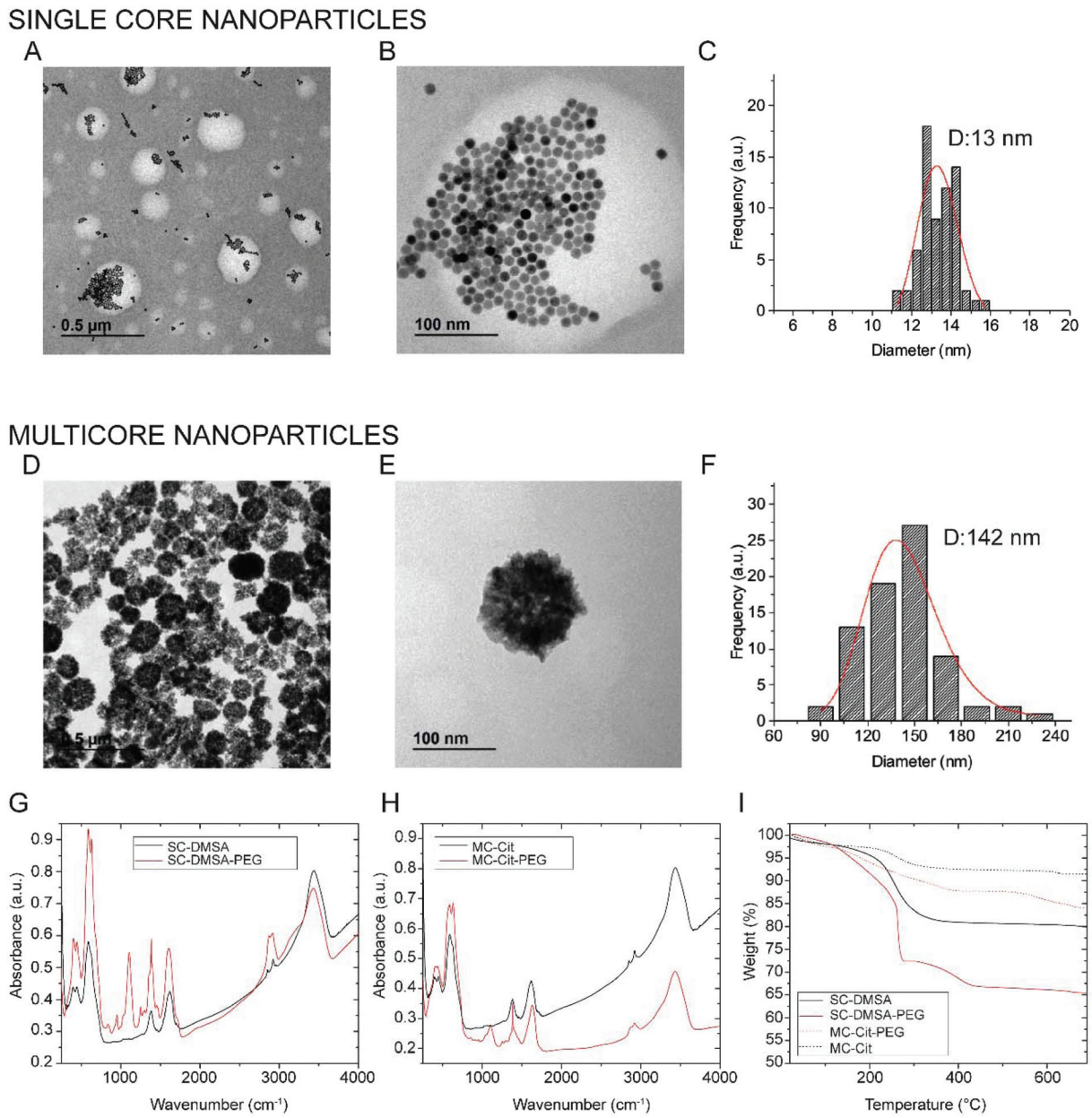


Fig. 1 Transmission electron microscopy (TEM) images of single and multicore nanoparticles. (A, B) SC-DMSA. (D, E) MC-Cit. (C, F) Size distribution histograms. Red lines indicate the log-normal fitting function of TEM particle size data. Physical chemical characterization of the nanoparticles after conjugation with PEG. FTIR spectra of (G) single-core nanoparticles. (H) Multicore nanoparticles. (I) Thermogravimetric analysis of four formulations.

particles at pH 7 are summarized in Table 1. Also, hydrodynamic sizes of the particles have been studied in different biological media (Fig. 1S†).

Nanoparticle surface modification was also confirmed by FTIR (Fig. 1G and H). For all the samples, the typical bands of metal skeleton vibration (Fe–O) in the region of 550–600 cm^{-1} and a broad peak between 3000 and 3500 cm-1 due to surface –OH groups are observed. After PEG conjugation, some peaks appeared at 1354 and 1102 cm^{-1}, indicating asymmetric and symmetric stretching of C–O–C, and out-of-plane bending of the –CH of the PEG chains at 956 cm^{-1}. TGA of the unconju-

Table 1 Colloidal properties of aqueous suspensions of the nanoparticles at pH 7. Hydrodynamic sizes, PDI (= standard deviation/mean size) and surface charge of single and multicore nanoparticles after PEG conjugation

Sample	Hydrodynamic size (nm)	PDI	ζ-Potential (mV)
SC-DMSA	34	0.054	−38 ± 12
SC-DMSA-PEG	65	0.084	−24 ± 7
MC-Cit	181	0.201	−25 ± 9
MC-Cit-PEG	183	0.225	−18 ± 9

gated nanoparticles reveals a weight loss of ~15% and 8% for SC-DMSA and MC-Cit respectively, due to the removal of physical and chemical water and capping molecules (DMSA or citric acid) (Fig. 1I). Particles modified with diamine PEG nonetheless showed a larger amount of conjugated polymer (~20% in the case of SC-DMSA-PEG and 13% for MC-Cit-PEG) which indicates greater reaction efficiency. In the case of SC-DMSA-PEG, the presence of ~2 molecules of PEG per nm^2 was calculated. In the case of MC-Cit-PEG is not possible to make an accurate calculation of the number of PEG molecules bounded to the surface because is not a homogeneous spherical surface due to their multicore structure. A notable decrease in the surface charge of the particles is observed after PEG conjugation, however not all carboxyl groups are modified during the reaction. The surface is not saturated probably due to the steric hindrance caused by the PEG chains, and therefore we have a limited balance of the negative surface charge coming from the DMSA or citric acid.

Toxicity *in vitro*

In vitro toxicological characterization of the nanoparticles was evaluated through the degree of cell survival by means of the standard MTT assay. The cell lines used, Hep G2 and Caco-2, represent the liver and the intestine, which are two important target organs to encounter the NPs after oral exposure or intravenous injection in a biomedical application. The analysis of cytotoxicity after incubation with the nanoparticles showed that viability of cell culture is not significantly reduced by the presence of the nanoparticles up to 160 μg mL^{-1} Fe concentration after 24 h of treatment (80–100% viability compared with the control) (Fig. 2A and B). At iron concentrations of 320 μg mL^{-1}, SC decreased the viability of both cell lines down to values of 20%. In the case of MC, viability percentage of both cell lines is in the range of 60–80%, indicating lower toxicity even at high iron concentrations. The effect of PEG coating is not very clear at high iron concentration. In the case of Hep G2 cells, PEG functionalization improved the cytotoxicity for SC while in Caco-2 it had no effect. For MC, PEG functionalization reduces the viability to 60% in Caco-2 cells (Fig. 2A and B). Even though this method needs further refinement and standardization for toxicity evaluation of nanomaterials, it offers a quick, inexpensive and high-throughput methodology to perform *in vitro* cytotoxicity screens.[36]

In order to characterize the absorption and biotransformation of the nanoparticles, we determined by ELISA the ferritin formation in Caco-2 cells as a measure of iron uptake (Fig. 2C). Single-core nanoparticles seem to be internalized by the cells more efficiently than multicore nanoparticles as evidenced by a 5.4-fold increase in intracellular ferritin (*cf.* control) compared with a 3.7-fold increase for MC-Cit probably due to the smaller size of the SC. PEGylated MC particles failed to alter ($p > 0.05$) the ferritin levels in Caco-2 cells. The surface modification with PEG in both types of nanoparticles reduces the internalization by the cells. The coating with short-chain PEG molecules of iron oxide nanoparticles was found to shield their surface charge and decreases their internalization.[37]

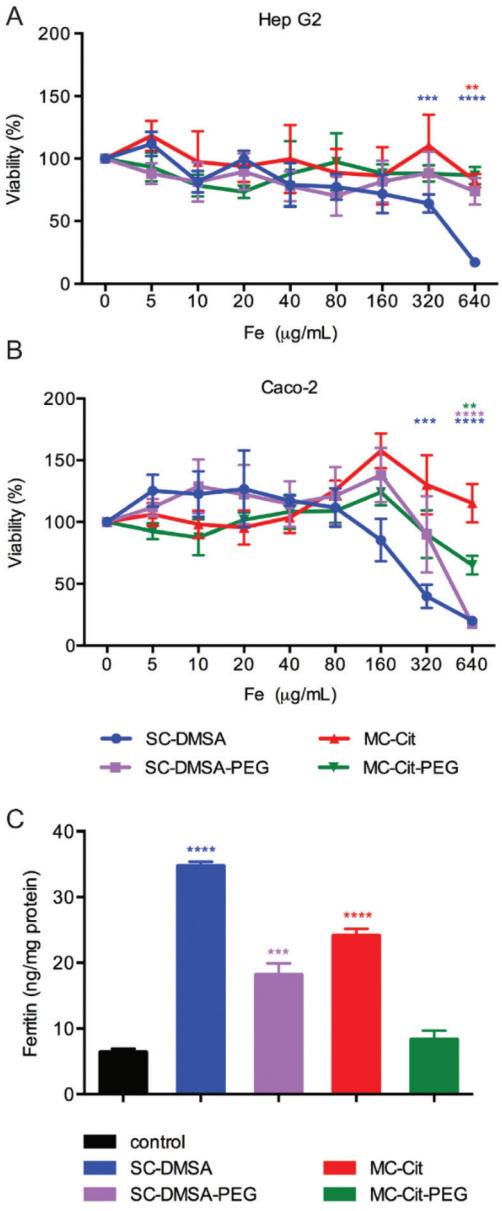


Fig. 2 Evaluation of cell viability by MTT assay. (A) Hep G2 cells. (B) Caco-2 cells. Cells were treated with the nanoparticles for 24 h. (C) Ferritin formation measured by ELISA in Caco-2 cells exposed to 40 μg mL^{-1} of single or multicore nanoparticles. Data represent means ± SD ($n = 3$). * shows statistically-significance compared with the control (*t* test, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$).

In vivo evaluation of nanoparticles

To characterize the general toxicity of the nanoparticles, embryos were treated with four different concentrations, 0.25, 0.5, 0.75 and 1 mg mL^{-1}. In these experiments, nanoparticles were added at stage 38 and survival was evaluated every 24 h until embryos reach stage 45, after approximately 72 h of exposure. None of the conditions was found lethal for the

embryos, therefore we decided to perform the rest of the experiments comparing a low and a high dose of nanoparticles (0.5 and 1 mg mL^{-1}). The use of the *Xenopus* model has advantages with respect to other animal models. First, embryos develop externally, allowing experiments to be performed prior to, or directly following fertilization. Also, they have a rapid embryo growth and development. A larval tadpole has a fully functional set of organs, and it can be examined to determine if any experimental intervention (in this case a solution containing nanoparticles) has had an effect.

Tadpoles treated with SC-DMSA and MC-Cit at 1 mg mL^{-1} displayed some general body toxicity compared with the control, characterized by defects in embryo body shape like bent spine or enlarged ventral fin (Fig. 3B). In the case of SC-DMSA, 20% of the animals showed enlarged ventral fin and MC-Cit provokes a significant increase (27%) of embryos with bent spine.

Interestingly, the effect of the PEG coating for each type of NPs reduced the toxicological effect as evidenced in a 2.6-fold decrease of embryos with bent spine and a 2-fold decrease of mild gut inflammation for MC; and a significant reduction in the enlarged ventral fin phenotype for SC treated tadpoles (Fig. 3A). In the case of embryos treated with 0.5 mg mL^{-1}, defects were subtler. These results suggested a dose-dependent uptake of the NPs by the embryos. We next evaluated gut morphology as it is the main organ involved in NPs' absorption after oral ingestion. Different degrees of inflammation were observed according to the dose of the nanoparticles, but also

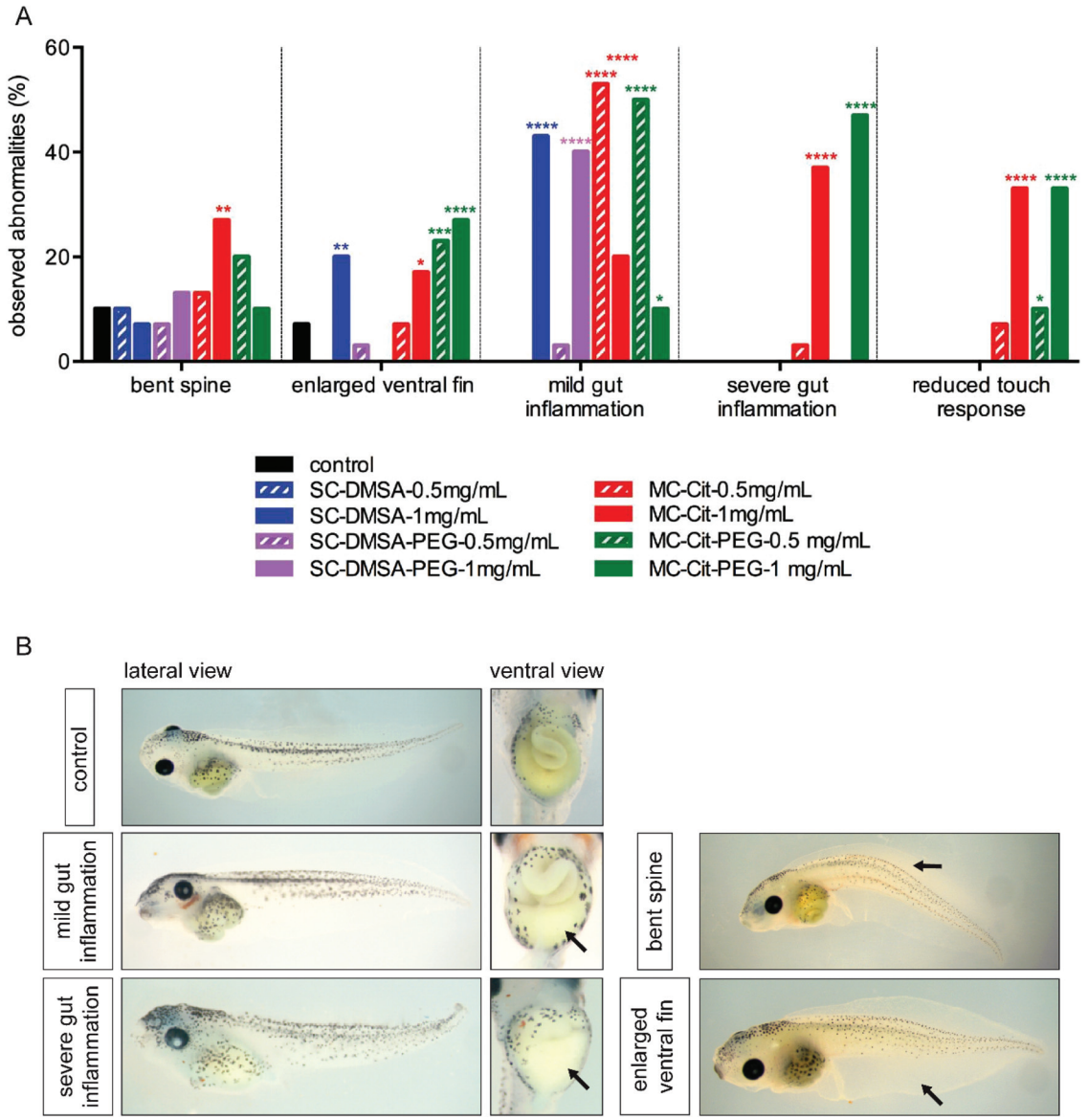


Fig. 3 *Xenopus laevis* phenotypic nanotoxicity assay after 72 h of exposure from stage 38–45. (A) Quantification of embryonic abnormalities. Histograms shown are of 30 embryos per group (chi-square test, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$). (B) Representative images from the *X. laevis* phenotypic nanotoxicity assay.

according to their chemical structure. Mild inflammation is observed in 43% of animals treated with SC-DMSA at 1 mg mL^{-1}. The intestine coiled structure is preserved although it is visibly enlarged. Embryos treated with MC-Cit at the same dose showed more drastic effects (20% mild and 36% severe inflammation). In the case of severe inflammation, the gut is not maintained or not well formed, as we are unable to see the intestinal loops. Regarding the effects on inflammation we didn't observed significant differences in PEGylated particles compared with the uncoated ones. In addition, 30% of the animals treated with MC showed a significant decrease of touch response after 72 hours of exposure (Fig. 3A).

Electron microscopy analysis of pharynx sections (Fig. 4), taken from different embryos, affords the localization of the ingested particles. The presence of the NPs in the tissue sample is evidenced by the Energy Dispersive X-ray (EDS) elemental mapping which delivers spots with brighter contrast in areas of NP accumulation, due to the higher Z atomic number of iron. NP intake is visibly detected in the images of pharynx sections of the embryos because the EDS spectra clearly showed K electron shell (K-α and K-β lines) of iron, whereas in the case of control these peaks are absent (Fig. 4). EDS analysis confirmed that only at the dose of 1 mg mL^{-1}, SC-DMSA and MC-Cit, could be detected in the upper body tissue sections, in contrast with the control. PEGylated nanoparticles (SC or MC) were undetectable in all organisms. We also observed SC-DMSA in the gills of the embryos (Fig. 2S†). These results agree with the size-dependent deposition in the intestine and/or the gills in Zebrafish observed for silver nanoparticles where particles from 10–20 nm showed increased bioavailability compared with 140 nm particles.[38] In our case, the deposition of SC-DMSA in the gills is directly related to the smaller size and larger surface area compared with MC-Cit, providing increased adherence, penetration, and deposition in the organism.

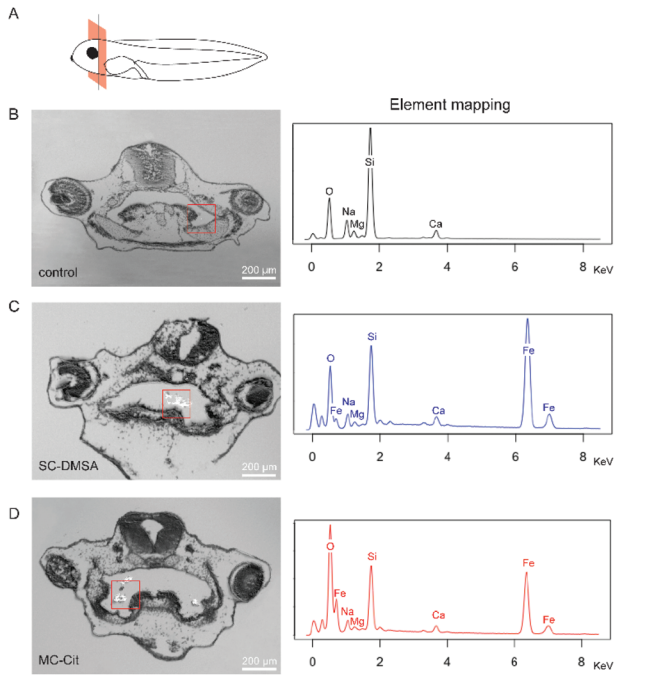


Fig. 4 Scanning electron microscopy (SEM) images of pharynx sections from different embryos. (A) Schematic representation of the transversal section analyzed. (B) Control. (C) Embryos exposed to SC-DMSA nanoparticles (1 mg mL^{-1}, 72 h). (D) Embryos exposed to MC-Cit nanoparticles (1 mg mL^{-1}, 72 h). Right panel shows elemental analysis of red selected areas in the images obtained by Energy Dispersive X-ray Spectrometry (SEM-EDS).

Iron quantification. AC magnetic susceptibility and ICP-OES

In order to quantify the iron present in the samples through SEM-EDS analysis, the application of quantitative correction procedures is needed, which are sometimes referred to as matrix corrections.[39] As SEM only provides local information, the quantitative analysis of NP intake, was performed through magnetic characterization of freeze-dried tissue samples and iron elemental analysis of acid digested embryos. Moreover, these techniques allow the comparison of the accumulation depending on the coating and the surface charge (DMSA, citric acid, and PEG functionalization).

In order to evaluate the accumulation of the nanoparticles in the animals, we performed AC susceptibility of pools of seven animals. Magnetic measurements, especially AC magnetic susceptibility, are extremely sensitive being able to distinguish the contribution from the magnetic nanoparticles from that of other endogenous iron-containing species, usually present in a bigger concentration but with weaker magnetic signal than the nanoparticles. The presence of magnetic nanoparticles can be identified by a maximum in the in-phase magnetic susceptibility (χ') accompanied by a maximum at slightly lower temperatures in the out-of-phase susceptibility (χ''). The temperature location of the maxima depends on the nanoparticle (material, size, aggregation degree, *etc.*). The height of the maxima is a surrogate measurement of the concentration of nanoparticles.

In our case, a dose-dependent amount of nanoparticles is observed in the animals treated with particles without PEG coating (Fig. 5), independently if they are SC or MC. Interestingly PEG coated nanoparticles, both SC and MC, were below the detection limits of the technique. These results are in agreement with what we observed by SEM-EDS. In this work, the presence of ferritin, the iron storage protein with an out-of-phase susceptibility maxima located at around 10 K,[18] has not been observed. The absence of a substantial paramagnetic contribution, usually observed at the lowest temperatures, indicates that the concentration of free iron atoms is very low, indicating that NP degradation process is limited if not absent in the timescale of these experiments.

Through AC magnetic susceptibility measurements, dose-dependent amounts of nanoparticles (SC-DMSA and MC-Cit) were detected. In addition, a quantitative analysis of NP intake was performed by ICP-OES (Fig. 6). Elemental analysis results evidenced a dose-dependent iron absorption, which is in agreement with the iron content detected by AC magnetic susceptibility. The absorption of SC-DMSA detected in the

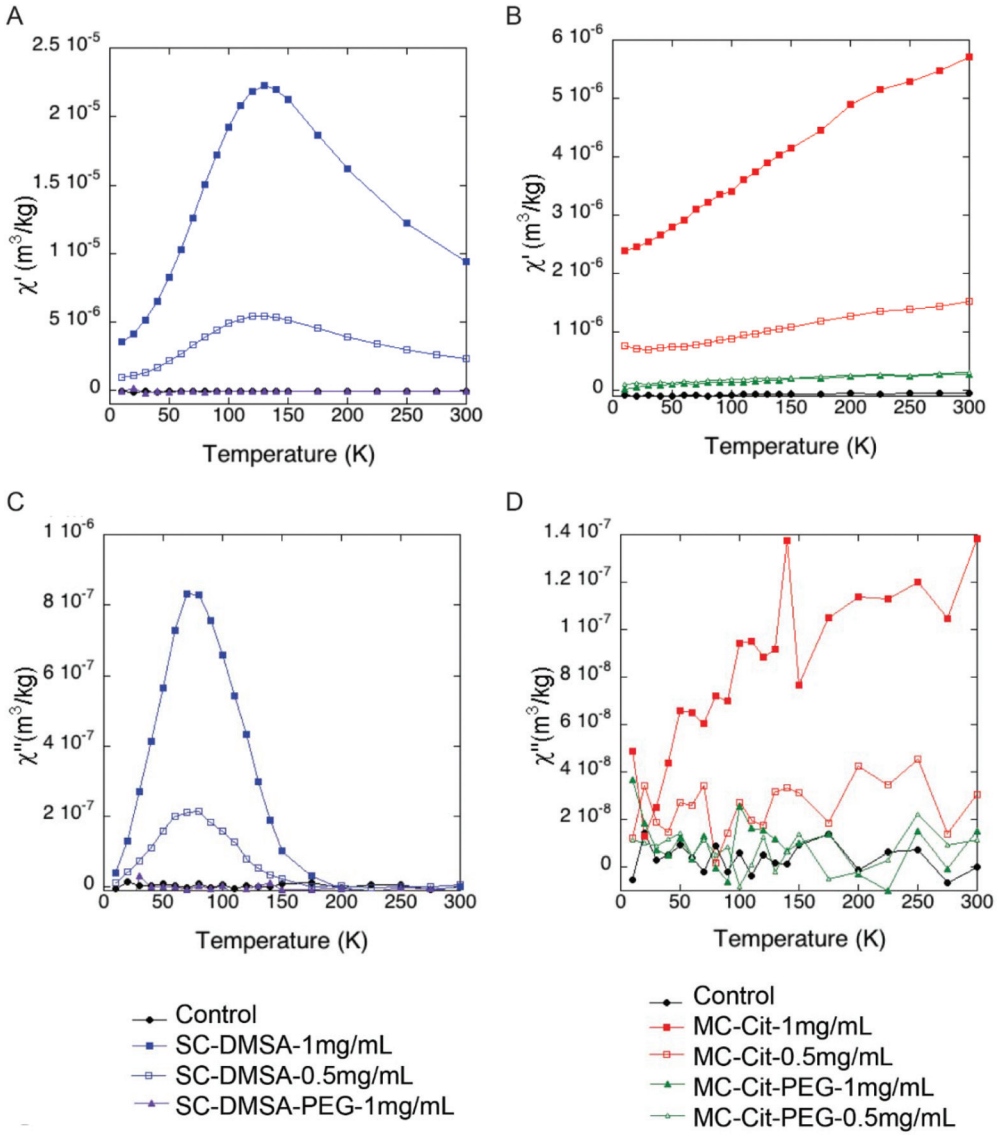


Fig. 5 AC magnetic susceptibility measurement of embryo tissue samples. In-phase ($\chi'(T)$) and out-of-phase ($\chi''(T)$) components of the magnetic susceptibility, per mass of sample, corresponding to freeze-dried tissues from different embryos after exposure (72 h): (A, C) embryos exposed to SC. (B, D) Embryos exposed to MC (n = 7 per group).

embryos has been confirmed by ICP-OES, a three or five-fold increase in comparison with MC-Cit for iron concentrations of 0.5 and 1 mg mL^{-1}, respectively. Interestingly, the iron content in the animals treated with PEGylated NPs is comparable to the control indicating a very low absorption of these NPs.

One possible explanation for the visual disturbance of the intestine structure of the embryos treated with MC-Cit-PEG, could be that the particles are rapidly excreted by the organism and therefore the iron content in these animals is seen as low. As we previously observed in Caco-2 cells, PEG functionalization reduces net surface charge of the particles and this effect could reduce the absorption of the particles in the gut favouring their excretion. Surface functionalization with PEG has been reported to enable particles to diffuse through mucus in the gastrointestinal tract at a rate similar to diffusion through water. It has been hypothesized that this occurs by reduction of particle–mucus interactions.[40] There is also evidence that PEG-coating can decrease macrophage or HeLa uptake of iron particles where the polymer increases uptake hindrance compared with uncoated NPs.[41] Another possible explanation for the morphological alterations observed at higher doses of nanoparticles could be the presence of polyol used during the MC synthesis. Other studies have shown that polyol-made nanoparticles produce *in vitro* (10 μg mL^{-1}) and *in vivo* (0.8 mg kg^{-1}) toxicity.[42]

Iron metabolism

Since iron resulting from NP degradation is predicted to be processed by iron metabolic pathways, we next studied the expression of different genes involved in iron metabolism and

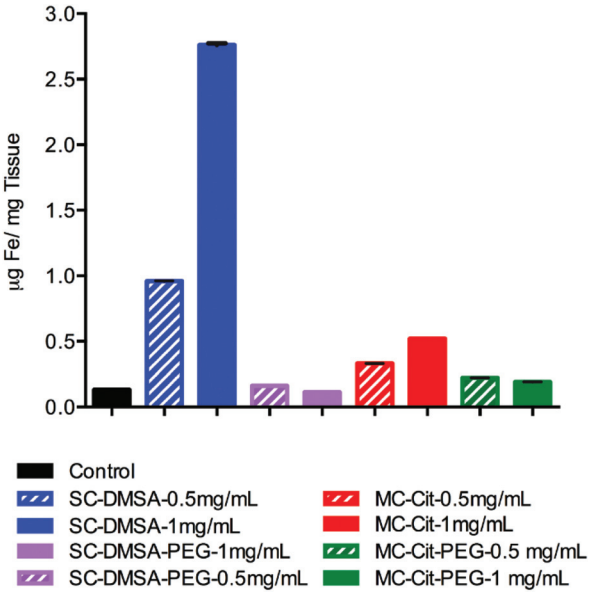


Fig. 6 Iron quantification though ICP-OES of acid digested tissues from embryos treated at different concentrations of IOMNPs. Striped pattern: 0.5 mg mL^{-1}. Solid pattern: 1 mg mL^{-1} (n = 7 per group).

oxidative stress. The expression of genes linked to both processes was assessed by real-time quantitative PCR (RT-qPCR) in embryos treated with SC-DMSA and MC-Cit after 72 h of treatment with the nanoparticles at 1 mg mL^{-1}, taking into account that only at highest doses could we detect a significant amount of absorbed nanoparticles (Fig. 7A).

In the case of SC-DMSA treated embryos, *transferrin* and *dmt1* increased 2.6 and 2-fold respectively (*cf.* control) coupled with a 70% downregulation of *hepcidin*. In contrast, the MC-Cit treated embryos displayed marginal increases (p < 0.05) in *hepcidin*, and *transferrin* and *dmt1* levels remained unchanged. In both cases the transcription of *ferritin* is increased after nanoparticle treatment, 2.9-fold change for SC-DMSA and 1.9-fold change for MC-Cit treated embryos (Fig. 7A).

From these data, we can describe differences in the metabolic activation pathway of single-core or multicore nanoparticles. After iron ingestion, ferrous ions (Fe^{2+}) are absorbed in the enterocytes through the divalent metal transporter-1 (DMT1). In the apical membrane of the enterocytes, the cytochrome B facilitates the reduction of ferric to ferrous ions enabling the absorption of iron.

In the basolateral membrane of the enterocytes the ferroportin transporter is located together with hephaestin which oxidize ferrous ions and allows transferrin to bind iron and carry it in the bloodstream through the different organs. Hepcidin negatively regulates the entry of iron into circulation through inhibiting ferroportin in case of iron overload (Fig. 8A). Ferritin is the most important protein involved in iron storage within cells, and the levels of Fe^{2+} present in the organism regulates its expression.[43]

From magnetic measurements and elemental analysis, we detected a higher absorption of SC-DMSA compared with MC-Cit. We propose that the rate of particle internalization into the embryos will dictate the time at which iron metabolic pathway activation is measurable. In the case of SC-DMSA

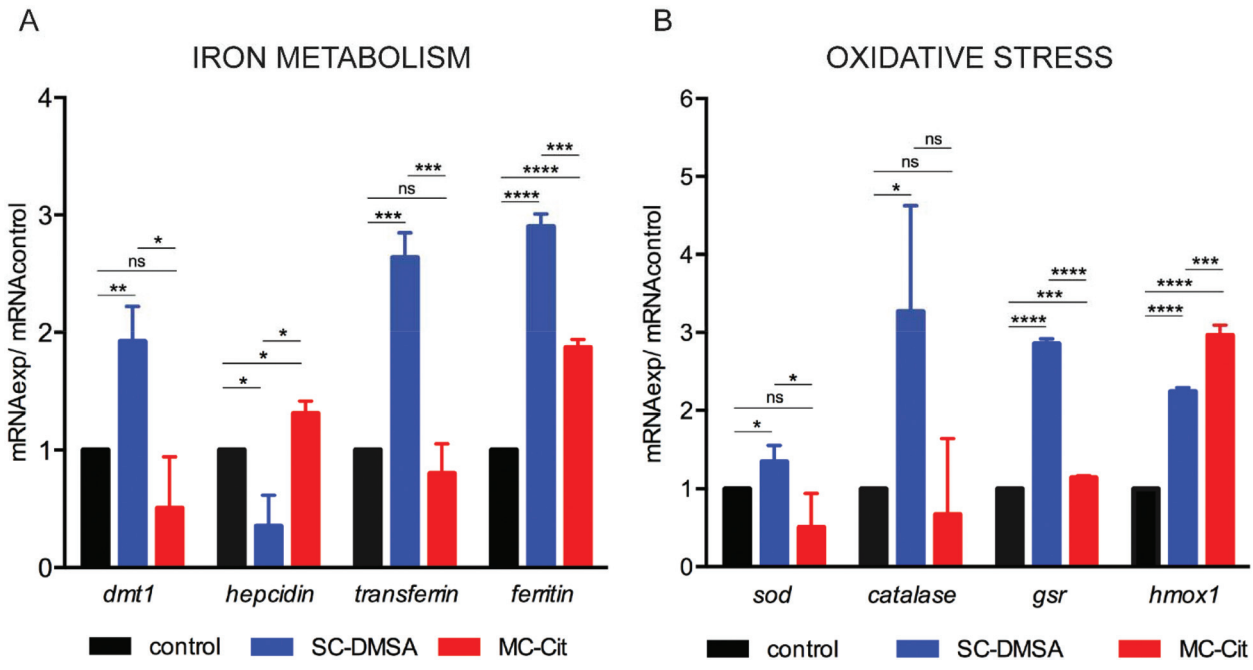


Fig. 7 Effect on the expression of different genes in *X. laevis* embryos exposed to SC-DMSA or MC-Cit after 72 h at 1 mg mL^{-1}. (A) Genes involved in iron metabolism. (B) Genes involved in oxidative stress response (n = 7 per group). * shows statistically-significance compared with the control (t test, ns > 0.05, *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001).

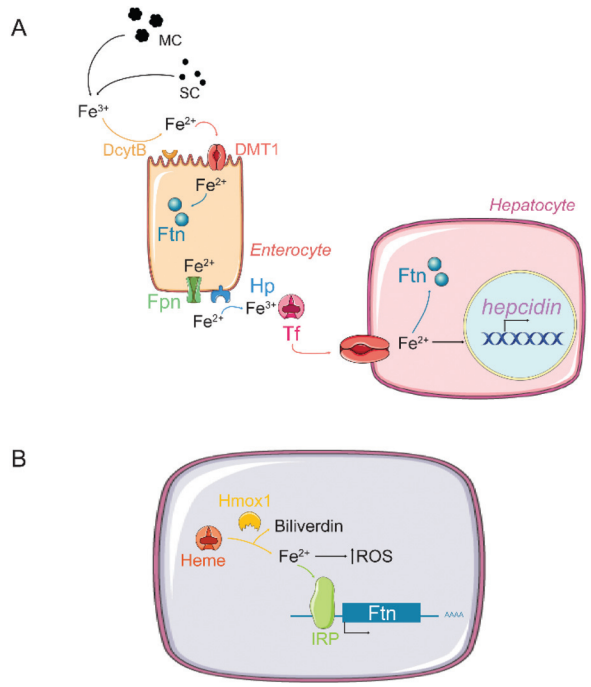


Fig. 8 Schematic representation of iron metabolic pathway. (A) Enterocyte uptake and transfer of iron: ferrous ions (Fe^{2+}) are absorbed in the enterocytes through the divalent metal transporter-1 (DMT1). In the apical membrane of the enterocytes, the cytochrome B (DcytB) facilitates the reduction of ferric to ferrous ions enabling the absorption of iron. In the basolateral membrane of the enterocytes the ferroportin transporter (Fpn) is located together with hephaestin (Hp), which oxidize ferrous ions and allows transferrin to bind iron and carry it in the blood-stream through the different organs. In the liver iron will be stored in complex with ferritin (Ftn). In the case of iron overload Hepcidin transcription is activated and it will regulate the entry of iron into circulation through inhibiting ferroportin. (B) Regulation of ferritin mRNA translation by Hmox1 and IRP: iron overload induces ROS generation. Hmox1 is induced after oxidative stress, thus degrading heme group to form biliverdin and ferrous ions. The production of Fe^{2+} leads to the activation of iron regulatory protein (IRP) which is able to control the translation of iron sensitive proteins such as ferritin. In the absence of iron, IRP binds to ferritin mRNA and inhibits its translation. However, when iron ions are available, they bind to IRP and release it from ferritin mRNA, thus allowing its translation.

treated embryos, *dmt-1*, *transferrin* and *ferritin* have the highest levels of transcription. The increase in ferritin in embryos treated with SC-DMSA compared with MC-Cit, follows the same pattern observed in Caco-2 cells where the protein was detected by ELISA suggesting a higher iron bioavailability from single-core nanoparticles. *Hepcidin* is downregulated for SC-DMSA, which implies a saturation of the metabolic pathway of iron. In the case of MC-Cit only *ferritin* and *hepcidin* are upregulated corroborating a slower activation of the iron metabolic pathway depending on the amount of internalized particles after 72 h of treatment.

Oxidative stress

In embryos treated with SC-DMSA or MC-Cit at 1 mg mL^{-1} we observed a 2.2 and 2.96-fold increase in *hmox1* (*cf.* control) respectively after 72 h of exposure (Fig. 7B). Increases in transcript levels of *hmox1* have been described in macrophages treated with iron oxide hybrids nanomaterials.[44] Hmox1 is induced after oxidative stress, reactive oxygen species or heavy metals, thus degrading heme group to form biliverdin and Fe^{2+}. The production of Fe^{2+} leads to the activation of iron regulatory protein (IRP) which is able to control the translation of iron sensitive proteins such as ferritin. In the absence of iron, IRP binds to ferritin mRNA and inhibits its translation. However, when iron ions are available, they bind to IRP and release it from ferritin mRNA, thus allowing its translation (Fig. 8B).[45] The increase in Fe^{2+} in the embryos 72 h after the treatment, activates iron response proteins, which activates the translation of proteins involved in iron metabolism such as ferritin. We show *in vivo* the activation by Hmox-1 of iron regulatory proteins and their effect in the transcription of ferritin mRNA associated to the treatment with iron oxide nanoparticles (Fig. 7A and B).

Other oxidative stress response genes showed an increase in expression in the case of embryos treated with SC-DMSA (*sod* 1.35-fold, *catalase* 3.27-fold and *gsr* 2.85-fold) compared with MC-Cit (*sod* 0.5-fold, *catalase* 0.67-fold and *gsr* 1.14-fold) (Fig. 7B). These results suggest the activation of protective mechanisms depending on nanoparticle uptake in order to reduce reactive oxygen species (ROS) generated during the biodegradation process. This is the early response of the organism when nanoparticle clearance induces the formation of ROS. However, we are not in the presence of an acute state of oxidative stress when a general decrease of antioxidant enzymes takes place.[46]

With these results, we provide evidence that an early developmental vertebrate model such as *Xenopus laevis* is a rapid and inexpensive system for NP toxicity assessment, compared with adult mammalian models. Because most theranostic applications need nanoparticles to be taken up by the cells, intracellular nano-biodegradation in an *in vitro* model needs to correlate with *in vivo* observations. In the case of IONPs degradation *in vitro* and their bioavailability by the cells translated into an increase of ferritin levels (Fig. 2C), should correlate with the corresponding activation of the iron metabolic pathway as shown in Fig. 7A.

Here we demonstrated for the first time the study, by means of AC magnetic susceptibility measurements, of the intake of iron-containing particles in *X. laevis* embryos. We postulate that the rate of particle internalization into the embryos will dictate the time at which iron metabolic pathway activation is measurable. Also, the NP deposition and uptake likely depends on the physico-chemical characteristics of the nanoparticles in terms of structure, size, and chemical modification of the surface. We can consider *Xenopus* as the bridge between cell-based assays and mammalian models taking into account that the activation of the iron metabolic pathway, especially with SC-DMSA, correlates with previous results in different murine models tracking IONPs biodegradation.[18,32,47] The significant increase of transcript levels of *ferritin* or *dmt1* correlates with previous observations in a Wistar rat animal

model where these proteins were quantified by ELISA in tissue samples from the liver, spleen and kidneys, 24 hours after SC-DMSA intravenous injection.[18] Also, multicore nanoparticles degradation has been described *in vitro* recently, starting with the dissociation of the multigrain structure and following with the massive dissolution of the iron oxide cores.[48] These results highlight the necessity to study at longer times the nanoparticle biotransformation in *X. laevis*.

Conclusions

In this work, the toxicity and metabolization of DMSA, citric acid and PEG coated single and multicore iron oxide magnetic nanoparticles have been studied in Hep G2 and Caco-2 cell lines, and in an amphibian animal (*Xenopus laevis*) model during its embryo development. The viability of both cell lines is preserved in all cases when treated with the nanoparticles up to an iron concentration of 160 μg mL^{-1}. Above that concentration, only PEGylated nanoparticles didn't cause toxicity on Hep G2 cells but Caco-2 cells were more sensitive. In the case of the *in vivo* viability, although none of the single-core nor multicore coated nanoparticles were found lethal for the embryos, more dramatic effects were observed in the treatment with multicore nanoparticles. Single-core nanoparticles are absorbed (three or five-fold increase for doses of 0.5 and 1 mg mL^{-1}, respectively) in comparison with multicore ones suggesting size-dependent differences in the deposition an uptake of the nanoparticles. Monitoring nanoparticle biodegradation is critical for therapeutic efficiency and safety issues. Depending on their structural organization (single or multicore) and surface chemistry, the embryos will trigger the iron metabolic pathway to varying degrees. These findings advance the understanding of iron oxide nanoparticle metabolization in an early developmental vertebrate model like *Xenopus laevis*, which offers a quick, inexpensive and high-throughput alternative prior toxicity assessment of nanotherapeutics in rodent models.

Materials and experimental methods

Materials

Commercial products: iron(III) chloride hexahydrate, oleylamine, oleic acid, *meso*-2,3-dimercaptosuccinic acid (DMSA), 1-octadecene, toluene, dimethyl sulfoxide, *O*,*O*′-bis(2-aminoethyl)PEG, 2000 Da, polyvinylpyrrolidone PVP40, sodium acetate trihydrate, and citric acid (Sigma-Aldrich), sodium oleate (Riedel-de Haen), *n*-hexane (Scharlau), ethyl-3-(3-dimethylaminopropyl)-carbodiimide, ethylene glycol (Fluka), and ethanol (Panreac). Dialysis tubing cellulose membranes were purchased from Sigma-Aldrich and washed prior to use.

Nanoparticles synthesis

Synthesis of iron oxide single-core nanoparticles (SC). The synthesis of iron oxide nanoparticles has been based on a previous work described in the literature.[49] The reaction was carried out under nitrogen. In a round-bottomed flask (500 mL), equipped with a mechanical stirrer (glass stirrer shaft), thermometer, entry for nitrogen flow and reflux condenser, iron(III) oleate (4.54 g) was mixed with oleic acid (0.724 g) in 1-octadecene (50 mL). The mixture was stirred (100 rpm) and slowly heated (5 °C min^{-1} for T < 100 °C, and 2 °C min^{-1} for T > 100 °C) until reflux (325 °C) with a heating mantle. The heating mantle was withdrawn, and the system was cooled down to room temperature. The resulting black mixture was washed with ethanol several times by centrifugation (RCF = 7500) and magnetic decantation. The resulting dried black solid was redispersed in hexane.

Synthesis of iron oxide multicore nanoparticles (MC). The synthesis of iron oxide nanoparticles has been based on a previous work described in the literature[50] but the experimental procedure and the concentration of NaAc have been varied. Typically, 2.62 mmol iron(III) chloride were dissolved with ultrasound in 109 mL of ethylene glycol. Then, 140 mmol PVP40 were added slowly under vigorously magnetic stirring (>1000 rpm) and mild heating until completely dissolved. Then, 26.2 mmol of $NaAc{\cdot}3H_2O$ were added to the solution. The mixture was sealed in a Teflon-lined autoclave (125 mL) and maintained at 200 °C for 16 h for solvothermal crystallization, following cooling inside oven. The precipitated solid product was washed with ethanol and distilled water through centrifugation (RCF = 7500) several times.

Surface modification

Ligand exchange with dimercaptosuccinic acid on single-core NPs (SC-DMSA). For DMSA ligand exchange, a standard procedure was used.[51] In a typical experiment, ethanol (20 mL) was added to a volume of SC dispersed in hexane containing a mass of Fe_3O_4 of 50 mg. The mixture was sonicated and then placed on a magnet to separate the liquid from the black solid residue of nanoparticles. The residue was washed with more ethanol (5 × 10 mL) following the same procedure, until the discarded liquid had a clean appearance. The remaining black residue was dispersed in toluene (20 mL) and the dispersion added to a solution of DMSA (90 mg) in dimethyl sulfoxide (5 mL). The resulting black suspension was then shaken in a laboratory tube rotator. After 2 days, SC-DMSA nanoparticles were precipitated as a black powder stuck to the glass tube and the liquid phase was transparent and pale yellow. The liquid was discarded and the nanoparticles were washed with ethanol (4 × 10 mL), sonicating and centrifuging (RCF = 7500). The final black solid was air dried and redispersed in distilled water. KOH 1 M was added to increase the pH to 10 and HNO_3 0.01 M was used to lower the pH to 7. The dispersion was then placed in a cellulose membrane and dialyzed for 5 days in distilled water, to remove any excess of unreacted DMSA and any other small impurities that may be present in the dispersion without being attached to the nanoparticles.

Surface coating with citric acid on multicore NPs (MC-Cit). For citric acid coating a standard procedure was used.[24,52] First, sample volume equivalent to 20 mg of Fe was adjusted to

pH 2 and then dispersed in 13 mL of a solution of citric acid 0.1 M. Afterwards, the mixture was heated at 80 °C for 30 min. The solution was centrifuged and washed with distilled water. Finally, the pH was adjusted first to 11 with KOH 1 M and then to 7 with HNO_3 0.01 M.

Surface coating with polyethylene glycol (SC-DMSA-PEG and MC-Cit-PEG). PEG conjugation reaction has been based on a previous work described in the literature.[37] Amine-functionalized PEG was attached to SC-DMSA or MC-Cit *via* an ethyl-3-(3-dimethylaminopropyl)-carbodiimide (EDC)-mediated coupling reaction using the polymer *O*,*O*′-bis(2-aminoethyl) PEG, 2000 Da (Fig. 3S†). PEGylation reaction was carried out in a refrigerated ultrasonic bath; an aqueous solution containing 10 mg SC-DMSA or MC-Cit was mixed with the PEG derivative (4 mg). The total amount of EDC (1 mg) was divided into five aliquots, one of which was added every 1 h, and the fifth 4 h after the previous addition. The molar ratio of COOH groups/amine groups/EDC was 1 : 1.5 : 1; pH was adjusted to 6, and the mixture was sonicated (4 h, 25 °C), followed by extensive dialysis.

Nanoparticle structural characterization

The particle sizes and morphologies were determined by transmission electron microscopy (TEM) with a JEM1010 microscope (JEOL, Peabody, USA) operating at 100 kV. Samples were prepared by placing a drop of the uncoated particles suspended in water onto a carbon coated copper grid and allowing it to dry at room temperature. The size distributions were determined by manual measurement of more than 100 particles using the public domain software ImageJ. The presence of the coating and the washing process was also confirmed and studied by Fourier transform infrared spectroscopy (FTIR) in the range of 4000–250 cm^{-1} by use of a Bruker (USA) IFS 66VS. The samples for FTIR were prepared diluting the dried powder in KBr at 2% by weight and pressing it into a pellet. The presence of the coating was also studied by thermogravimetric analyses (TGA). They were performed in a Seiko TG/DTA 320U thermobalance, whose temperature scanning range is from room temperature up to 900 °C. For this work, samples were heated from room temperature to 700 °C at 10 °C min^{-1} under an air flow of 100 mL min^{-1}. Platinum pans were used and α-Al_2O_3 was used as reference. Colloidal properties were studied in a Zetasizer Nano S, from Malvern Instruments (UK). The hydrodynamic size was determined by Dynamic Light Scattering (DLS) and the zeta potential was measured as a function of pH at 25 °C, using HNO_3 and KOH to change the pH of the suspensions. Hydrodynamic size is given as the intensity-weighted mean.

Cell culture

Caco-2 cells (HTB-37) and HepG2 (HB-8065) were obtained from American Type Culture Collection (Manassas, VA, USA) and stored in liquid nitrogen. Hep G2 (human hepatocellular carcinoma) and Caco-2 (human colorectal adenocarcinoma) cells were cultured as mono-layers in Dulbecco's modified Eagle medium supplemented with 2% penicillin–streptomycin and 10% fetal bovine serum, in a humidified incubator (37 °C, 5% CO_2). For toxicity experiments, cells were seeded in 96-well plates (approximately 1×10^4 cells per well, 200 μL per well). For iron uptake experiments, Caco-2 cells between passages 30–36 were seeded onto collagen-coated 12-well plates (Bio-Greiner, UK) at a density of 2×10^5 cells per well suspended in 1 mL of supplemented DMEM which was replaced every 2 days. Cells were used on confluence at days 13–15 post-seeding. In order to ensure a low basal media iron levels, 24 hours prior to the initiation of the nanoparticles treatments, the DMEM medium was replaced by Eagle's minimum essential medium (MEM) without foetal bovine serum supplemented with 10 mmol L^{-1} PIPES [piperazine-*N*,*N*′-bis-(2-ethanesulfonic acid)], 26.1 mM $NaHCO_3$, 19.4 mmol L^{-1} glucose, 1% antibiotic–antimycotic solution, 11 μmol L^{-1} hydrocortisone, 0.87 μmol L^{-1} insulin, 0.02 μmol L^{-1} sodium selenite (Na_2SeO_3), 0.05 μmol L^{-1} triiodothyronine and 20 μg L^{-1} epidermal growth factor.[53] The day of the experiment, the nanoparticles were diluted in the low-iron MEM to obtain a 40 μg mL^{-1} final iron concentration and subsequently Caco-2 cells were exposed for 24 hours with the treatments. Ferritin formation was measured 24 h after treatment. Cells were rinsed with Milli-Q (18.2 MΩ) H_2O and subsequently lysed by scraping in 100 μL (12 well plates) of CelLytic M (Sigma-Aldrich, Gillingham, UK). Cell lysates were kept on ice for 15 min and stored at −80 °C. For analysis, samples were thawed and centrifuged at 14 000*g* for 15 min. Cellular debris was discarded and the supernatant containing the proteins was analysed for ferritin using the Spectro Ferritin ELISA assay (Ramco Laboratories Inc., Stafford, TX, USA). The ferritin concentration in the samples was determined using a microplate reader at an excitation wavelength of 500 nm according to the manufacturer's protocol. Ferritin concentrations were normalized to total cell protein using the Pierce Protein BCA protein assay (ThermoFisher Scientific, Loughborough, UK).

Cytotoxicity assay (MTT)

Cell viability was determined using the standard 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl tetrazolium bromide (MTT) assay 24 h after exposure to NP. Cells were seeded in 96-well plates at (approximately 1×10^4 cells per well, 200 μL per well). In total 60 wells were seeded per plate (6 rows × 10 columns) as the outer wells were left empty to avoid errors due to evaporation. Each row was used as a replicate (3 wells per condition) and serial dilutions went across the columns of the plate. Cells were left to grow until 70–80% confluency. NP-containing medium was removed after 24 h, cells were rinsed three times with PBS, and MTT solution in medium (final MTT concentration 50 μg mL^{-1}) was added and incubated (2 h, 37 °C). The MTT solution was removed without disturbing cells, 0.2 mL per well of DMSO and 0.025 mL per well of Sorensen buffer were added, the plates were shaken gently to dissolve formazan crystals, and the absorbance was read on a microplate reader at 550 nm. Cell viability (%) was calculated as $[(A - B)/A \times 100]$, where A and B are the absorbance of control and treated cells, respectively. Values represent mean ± SD ($n = 3$).

In vivo test

All experiments were performed in compliance with the relevant laws and institutional guidelines at the University of East Anglia. The research has been approved by the local ethical review committee according to UK Home Office regulations. *Xenopus laevis* embryo toxicity assays were carried out as described before.[54,55] Briefly, adult females were primed with PMSG (Pregnant Mare's Serum Gonadotropin) and induced with Chorulon. Eggs were obtained manually and fertilized in a petri dish by adding male sperm (male testis incubated with 2 mL 1× MMR (Marc's modified ringers) + 8 mL FBS). The fertilized embryos were dejellied *via* cysteine, and the embryos were plated in BSA-coated petri dishes covered with 0.1× MMR. Embryos were left at 23 °C until they reached stage 38, and then plated in 24-well plates (7 embryos per well) in 0.1× MMR medium containing the nanoparticles and incubated at 23 °C. Non-treated embryos were used as control. The mortality and the morphological changes of the embryos were recorded every 24 h until embryos reached stage 45.

Histological evaluation

When the embryos reached the appropriate stage, they were fixed in MEMFA (3.7% Formaldehyde, 1× MEM salts and DEPC H_2O) overnight at 4 °C. Samples were washed in PBS, dehydrated and kept in 100% ethanol. To embed embryos in wax, they were directly washed in a 65 °C oven first in histoclear, then in 1 : 1 histoclear : wax and finally wax 30 min to 1 h each wash. Then the embryos were and placed in molds with wax. Embryos were sectioned using a microtome generating 10 μm slices. Slices were then hydrated in water and dried to be analyzed by Scanning electron microscopy.

Magnetic characterization

The resulting freeze-dried samples were placed into gelatin capsules for their magnetic characterization in a Quantum Design MPMS-XL SQUID magnetometer with an AC susceptibility option. The measurements were performed with AC amplitude of 0.41 Oe, in the temperature range between 1.8 and 300 K and at a frequency of 11 Hz.

Iron content analysis

Groups of 7 embryos were pooled. Animals were weighed and lyophilized 72 hours in a Telstar lyoquest lyophilizer, and the iron content was measured by Inductively Coupled Plasma-Optical Emission Spectrometry (ICP-OES) in an Optima 2100 DV from PerkinElmer, after acid digestion, with concentrated HNO_3 during 1 h; or kept freeze-dried for magnetic characterization, respectively. The sample manipulation was performed using disposable plastic material to avoid ferromagnetic contamination.

RNA extraction and quantitative PCR

Groups of 7 embryos were snap frozen in liquid nitrogen. RNA was extracted using High Pure RNA isolation kit (Roche) and 1 μg of RNA was taken to synthetize cDNA using Maxima First Strand cDNA synthesis kit (Thermofisher). RT-PCR was performed using SYBER Green detection method. Primers were designed targeting both copies of *X. laevis* genes. Gapdh have been used as a control.

Primers	Sequence 5′–3′
fth1_F	tggagtaacaccctggaagc
fth1_R	aggatcaaccttgtcggatg
tf_F	agaaagggcaagtgggtttt
tf_R	tctggcaaagtgacaacagc
dmt1_F	cagaggatgaaacgcactca
dmt1_R	atcctgccactgatccagac
hepcidin_F	aaatcaaccccaatctgctg
hepcidin_R	gtttgttgattgccgaaggt
hmox1_F	ggagacctctcaggtggaca
hmox1_R	atggagttcatacgggaacg
sod2_F	tgtgcaggctcagtgtttgt
sod2_R	gctgcagagcaccataatca
gsr_F	gcaaagaggagaaggtggtg
gsr_R	cggaggaagtcggatgaata
cat1_F	cttctgcccagatgcttttc
cat1_R	agttgcccagagcgacttta
gapdh_F	ctttgatgctgatgctggaa
gapdh_R	gaagagggttgacaggtga

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

M. Marín-Barba has been supported by the People Program (Marie Curie Actions) of the European Union's Seventh Framework Program FP7 under REA grant agreement number 607142 (DevCom). AR, CJM and GNW acknowledge support from the 7-People Framework – Marie Curie Industry and Academia Partnerships & Pathways scheme (DNA-TRAP project, grant agreement nr. 612338). This work was also supported by the European Commission Framework Program 7 (NanoMag project, No. 604448) and by the Spanish Ministry of Economy and Competitiveness (Mago project, No. MAT2014-52069-R). H. Gavilán has carried out this work while undertaking a doctoral program on Advanced Chemistry at the Complutense University of Madrid. L. Gutiérrez acknowledges financial support from the Ramón y Cajal subprogram (RYC-2014-15512). E. Lozano-Velasco acknowledges support from Marie Curie fellowship (705089-MIR-CHROM-C). FTIR spectroscopy and thermogravimetric and chemical analysis were carried out in the support laboratories of Instituto de Ciencia de Materiales de Madrid (ICMM/CSIC). Authors acknowledge the facilities and the scientific and technical assistance, especially that of Bertrand Leze from the SEM service of the University of East Anglia.

References

1 J. Roger, J. N. Pons, R. Massart, A. Halbreich and J. C. Bacri, Some Biomedical Applications of Ferrofluids, *Eur. Phys. J.: Appl. Phys.*, 1999, **5**, 321–325.
2 W. Schutt, C. Gruttner, U. Hafeli, M. Zborowski, J. Teller, H. Putzar and C. Schumichen, Applications of Magnetic Targeting in Diagnosis and Therapy-Possibilities and Limitations: A Mini-Review, *Hybridoma*, 1997, **16**, 109–117.
3 R. Mejias, S. Perez-Yague, L. Gutierrez, L. I. Cabrera, R. Spada, P. Acedo, C. J. Serna, F. J. Lazaro, A. Villanueva, P. Morales Mdel, *et al.*, Dimercaptosuccinic Acid-Coated Magnetite Nanoparticles for Magnetically Guided in Vivo Delivery of Interferon Gamma for Cancer Immunotherapy, *Biomaterials*, 2011, **32**, 2938–2952.
4 S. Laurent, D. Forge, M. Port, A. Roch, C. Robic, L. Elst and R. N. Vander; Muller, Magnetic Iron Oxide Nanoparticles: Synthesis, Stabilization, Vectorization, Physicochemical Characterizations, and Biological Applications, *Chem. Rev.*, 2008, **108**, 2064–2110.
5 C. Alexiou, W. Arnold, R. J. Klein, F. G. Parak, P. Hulin, C. Bergemann, W. Erhardt, S. Wagenpfeil and A. S. Lübbe, Locoregional Cancer Treatment with Magnetic Drug Targeting, *Cancer Res.*, 2000, **60**, 6641–6648.
6 A. Jordan, R. Scholz, P. Wust, H. Fähling and R. Felix, Magnetic Fluid Hyperthermia (MFH): Cancer Treatment with AC Magnetic Field Induced Excitation of Biocompatible Superparamagnetic Nanoparticles, *J. Magn. Magn. Mater.*, 1999, **201**, 413–419.
7 R. Weissleder, M. Nahrendorf and M. J. Pittet, Imaging Macrophages with Nanoparticles, *Nat. Mater.*, 2014, **13**, 125–138.
8 R. Crichton, Intracellular Iron Metabolism and Cellular Iron Homeostasis, in *Inorganic Biochemistry of Iron Metabolism*, John Wiley & Sons, Ltd, 2002, pp. 167–190.
9 R. Crichton, *Intracellular Iron Storage and Biomineralization*, in *Inorganic Biochemistry of Iron Metabolism*, John Wiley & Sons, Ltd, 2002, pp. 133–165.
10 A. Balakumaran, E. Pawelczyk, J. Ren, B. Sworder, A. Chaudhry, M. Sabatino, D. Stroncek, J. A. Frank and P. G. Robey, Superparamagnetic Iron Oxide Nanoparticles Labeling of Bone Marrow Stromal (Mesenchymal) Cells Does Not Affect Their "Stemness", *PLoS One*, 2010, **5**, 1–8.
11 M. Geppert, M. C. Hohnholt, S. Nürnberger and R. Dringen, Ferritin up-Regulation and Transient ROS Production in Cultured Brain Astrocytes after Loading with Iron Oxide Nanoparticles, *Acta Biomater.*, 2012, **8**, 3832–3839.
12 M. C. Hohnholt, M. Geppert and R. Dringen, Treatment with Iron Oxide Nanoparticles Induces Ferritin Synthesis but Not Oxidative Stress in Oligodendroglial Cells, *Acta Biomater.*, 2011, **7**, 3946–3954.
13 J. Gu, H. Xu, Y. Han, W. Dai, W. Hao, C. Wang, N. Gu, H. Xu and J. Cao, The Internalization Pathway, Metabolic Fate and Biological Effect of Superparamagnetic Iron Oxide Nanoparticles in the Macrophage-like RAW264.7 Cell, *Sci. China: Life Sci.*, 2011, **54**, 793–805.
14 J. M. Rojas, L. Sanz-Ortega, V. Mulens-Arias, L. Gutiérrez, S. Pérez-Yagüe and D. F. Barber, Superparamagnetic Iron Oxide Nanoparticle Uptake Alters M2 Macrophage Phenotype, Iron Metabolism, Migration and Invasion, *Nanomedicine*, 2016, **12**, 1127–1138.
15 V. Mulens-Arias, J. M. Rojas, S. Pérez-Yagüe, M. P. Morales and D. F. Barber, Polyethylenimine-Coated SPIONs Trigger Macrophage Activation through TLR-4 Signaling and ROS Production and Modulate Podosome Dynamics, *Biomaterials*, 2015, **52**, 494–506.
16 F. Mazuel, A. Espinosa, N. Luciani, M. Reffay, R. Le Borgne, L. Motte, K. Desboeufs, A. Michel, T. Pellegrino, Y. Lalatonne, *et al.*, Massive Intracellular Biodegradation of Iron Oxide Nanoparticles Evidenced Magnetically at Single-Endosome and Tissue Levels, *ACS Nano*, 2016, **10**, 7627–7638.
17 J. M. Rojas, H. Gavilán, V. del Dedo, E. Lorente-Sorolla, L. Sanz-Ortega, G. B. da Silva, R. Costo, S. Perez-Yagüe, M. Talelli, M. Marciello, *et al.*, Time-Course Assessment of the Aggregation and Metabolization of Magnetic Nanoparticles, *Acta Biomater.*, 2017, **58**, 181–195.
18 A. Ruiz, L. Gutierrez, P. Cáceres-Velez, D. Santos, S. Chaves, M. L. Fascineli, M. P. Garcia, R. B. Azevedo and M. P. Morales, Biotransformation of Magnetic Nanoparticles as a Function of the Coating in a Rat Model, *Nanoscale*, 2015, **7**, 16321–16329.
19 C. James-Zorn, V. G. Ponferrada, K. A. Burns, J. D. Fortriede, V. S. Lotay, Y. Liu, J. Brad Karpinka, K. Karimi, A. M. Zorn and P. D. Vize, Xenbase: Core Features, Data Acquisition, and Data Processing, *Genesis*, 2015, **53**, 486–497.
20 S. Nations, M. Wages, J. E. Cañas, J. Maul, C. Theodorakis and G. P. Cobb, Acute Effects of Fe2O3, TiO2, ZnO and CuO Nanomaterials on Xenopus Laevis, *Chemosphere*, 2011, **83**, 1053–1061.
21 S. Nations, M. Long, M. Wages, J. Canas, J. D. Maul, C. Theodorakis and G. P. Cobb, Effects of ZnO Nanomaterials on Xenopus Laevis Growth and Development, *Ecotoxicol. Environ. Saf.*, 2011, **74**, 203–210.
22 C. Webster, D. Di Silvio, A. Devarajan, P. Bigini, E. Micotti, C. Giudice, M. Salmona, G. N. Wheeler, V. Sherwood and F. Baldelli Bombelli, An Early Developmental Vertebrate Model for Nanomaterial Safety: Bridging Cell-Based and Mammalian Toxicity Assessment, *Nanomedicine*, 2016, **11**, 643–656.
23 M. Giannaccini, M. Giannini, M. P. Calatayud, G. F. Goya, A. Cuschieri, L. Dente and V. Raffa, Magnetic Nanoparticles as Intraocular Drug Delivery System to Target Retinal Pigmented Epithelium (RPE), *Int. J. Mol. Sci.*, 2014, **15**, 1590–1605.
24 L. Gutierrez, R. Costo, C. Gruttner, F. Westphal, N. Gehrke, D. Heinke, A. Fornara, Q. A. Pankhurst, C. Johansson, S. Veintemillas-Verdaguer, *et al.*, Synthesis Methods to Prepare Single- and Multi-Core Iron Oxide Nanoparticles

for Biomedical Applications, *Dalton Trans.*, 2015, **44**, 2943–2952.
25 A. K. A. Silva, A. Espinosa, J. Kolosnjaj-Tabi, C. Wilhelm and F. Gazeau, Medical Applications of Iron Oxide Nanoparticles, in *Iron Oxides*, Wiley-VCH Verlag GmbH & Co. KGaA, 2016, pp. 425–472.
26 V. Torchilin, Tumor Delivery of Macromolecular Drugs Based on the EPR Effect, *Adv. Drug Delivery Rev.*, 2011, **63**, 131–135.
27 S. Dutz, M. Kettering, I. Hilger, R. Müller and M. Zeisberger, Magnetic Multicore Nanoparticles for Hyperthermia-influence of Particle Immobilization in Tumour Tissue on Magnetic Properties, *Nanotechnology*, 2011, **22**, 265102.
28 J. Kolosnjaj-Tabi, L. Lartigue, Y. Javed, N. Luciani, T. Pellegrino, C. Wilhelm, D. Alloyeau and F. Gazeau, Biotransformations of Magnetic Nanoparticles in the Body, *Nano Today*, 2016, **11**, 280–284.
29 F. J. Lázaro, L. Gutiérrez, A. R. Abadía, M. S. Romero and A. López, Biological Tissue Magnetism in the Frame of Iron Overload Diseases, *J. Magn. Magn. Mater.*, 2007, **316**, 126–131.
30 L. Gutierrez, A. R. Abadia, M. S. Romero, C. Quintana, M. P. Morales, C. Patino and R. Arranz, Bioinorganic Transformations of Liver Iron Deposits Observed by Tissue Magnetic Characterisation in a Rat Model, *J. Inorg. Biochem.*, 2006, **100**, 1790–1799.
31 F. J. Lázaro, L. Gutiérrez, A. R. Abadía, M. S. Romero, A. López and M. J. Muñoz, Whole Tissue AC Susceptibility after Superparamagnetic Iron Oxide Contrast Agent Administration in a Rat Model, *J. Magn. Magn. Mater.*, 2007, **311**, 460–463.
32 M. Martin, A. Rodriguez-Nogales, V. Garces, N. Galvez, L. Gutierrez, J. Galvez, D. Rondon, M. Olivares and J. M. Dominguez-Vera, Magnetic Study on Biodistribution and Biodegradation of Oral Magnetic Nanostructures in the Rat Gastrointestinal Tract, *Nanoscale*, 2016, **8**, 15041–15047.
33 P. Mulvaney, W. J. Parak, F. Caruso and P. S. Weiss, Standardizing Nanomaterials, *ACS Nano*, 2016, **10**, 9763–9764.
34 B. Jia and L. Gao, Morphological Transformation of Fe3O4 Spherical Aggregates from Solid to Hollow and Their Self-Assembly under an External Magnetic Field, *J. Phys. Chem. C*, 2008, **112**, 666–671.
35 L. LaConte, N. Nitin and G. Bao, Magnetic Nanoparticle Probes, *Mater. Today*, 2005, **8**, 32–38.
36 M. Mahmoudi, H. Hofmann, B. Rothen-Rutishauser and A. Petri-Fink, Assessing the in Vitro and in Vivo Toxicity of Superparamagnetic Iron Oxide Nanoparticles, *Chem. Rev.*, 2012, **112**, 2323–2338.
37 A. Ruiz, G. Salas, M. Calero, Y. Hernández, A. Villanueva, F. Herranz, S. Veintemillas-Verdaguer, E. Martínez, D. F. Barber and M. P. Morales, Short-Chain PEG Molecules Strongly Bound to Magnetic Nanoparticle for MRI Long Circulating Agents, *Acta Biomater.*, 2013, **9**, 6421–6430.
38 O. J. Osborne, S. Lin, C. H. Chang, Z. Ji, X. Yu, X. Wang, S. Lin, T. Xia and A. E. Nel, Organ-Specific and Size-Dependent Ag Nanoparticle Toxicity in Gills and Intestines of Adult Zebrafish, *ACS Nano*, 2015, **9**, 9573–9584.
39 J. I. Goldstein, D. E. Newbury, P. Echlin, D. C. Joy, A. D. Romig, C. E. Lyman, C. Fiori and E. Lifshin, Qualitative X-Ray Analysis, in *Scanning Electron Microscopy and X-Ray Microanalysis: A Text for Biologists, Materials Scientists, and Geologists*, Springer US, Boston, MA, 1992, pp. 341–364.
40 H. M. Yildiz, C. A. McKelvey, P. J. Marsac and R. L. Carrier, Size Selectivity of Intestinal Mucus to Diffusing Particulates Is Dependent on Surface Chemistry and Exposure to Lipids, *J. Drug Targeting*, 2015, **23**, 768–774.
41 J. Xie, C. Xu, N. Kohler, Y. Hou and S. Sun, Controlled PEGylation of Monodisperse Fe3O4 Nanoparticles for Reduced Non-Specific Uptake by Macrophage Cells, *Adv. Mater.*, 2007, **19**, 3163–3166.
42 A. Hanini, A. Schmitt, K. Kacem, F. Chau, S. Ammar and J. Gavard, Evaluation of Iron Oxide Nanoparticle Biocompatibility, *Int. J. Nanomed.*, 2011, **6**, 787–794.
43 S. Waldvogel-Abramowski, G. Waeber, C. Gassner, A. Buser, B. M. Frey, B. Favrat and J.-D. Tissot, Physiology of Iron Metabolism, *Transfus. Med. Hemother.*, 2014, **41**, 213–221.
44 D. Elgrabli, W. Dachraoui, H. De Marmier, C. Ménard-Moyon, D. Bégin, S. Bégin-Colin, A. Bianco, D. Alloyeau and F. Gazeau, Intracellular Degradation of Functionalized Carbon Nanotube / Iron Oxide Hybrids Is Modulated by Iron via Nrf2 Pathway, *Sci. Rep.*, 2017, **7**, 40997.
45 R. S. Eisenstein and H. N. Munro, Translational Regulation of Ferritin Synthesis by Iron, *Enzyme*, 1990, **44**, 42–58.
46 M. R. Balas, I. M. Din Popescu, A. Hermenean, O. L. Cinteză, R. Burlacu, A. Ardelean and A. Dinischiotu, Exposure to Iron Oxide Nanoparticles Coated with Phospholipid-Based Polymeric Micelles Induces Biochemical and Histopathological Pulmonary Changes in Mice, *Int. J. Mol. Sci.*, 2015, **16**, 29417–29435.
47 R. Mejías, L. Gutiérrez, G. Salas, S. Pérez-Yagüe, T. M. Zotes, F. J. Lázaro, M. P. Morales and D. F. Barber, Long Term Biotransformation and Toxicity of Dimercaptosuccinic Acid-Coated Magnetic Nanoparticles Support Their Use in Biomedical Applications, *J. Controlled Release*, 2013, **171**, 225–233.
48 F. Mazuel, A. Espinosa, G. Radtke, M. Bugnet, S. Neveu, Y. Lalatonne, G. A. Botton, A. Abou-Hassan and C. Wilhelm, Magneto-Thermal Metrics Can Mirror the Long-Term Intracellular Fate of Magneto-Plasmonic Nanohybrids and Reveal the Remarkable Shielding Effect of Gold, *Adv. Funct. Mater.*, 2017, **27**, 1605997.
49 G. Salas, C. Casado, F. J. Teran, R. Miranda, C. J. Serna and M. P. Morales, Controlled Synthesis of Uniform Magnetite Nanocrystals with High-Quality Properties for Biomedical Applications, *J. Mater. Chem.*, 2012, **22**, 21065.
50 Q. Sun, Z. Ren, R. Wang, W. Chen and C. Chen, Magnetite Hollow Spheres: Solution Synthesis, Phase Formation and Magnetic Property, *J. Nanopart. Res.*, 2011, **13**, 213–220.

51 Y. Jun, Y.-M. Huh, J. Choi, J.-H. Lee, H.-T. Song, S. KimKim; Yoon, K.-S. Kim, J.-S. Shin, J.-S. Suh, *et al.*, Nanoscale Size Effect of Magnetic Nanocrystals and Their Utilization for Cancer Diagnosis via Magnetic Resonance Imaging, *J. Am. Chem. Soc.*, 2005, **127**, 5732–5733.

52 M.-S. Martina, J.-P. Fortin, C. Ménager, O. Clément, G. Barratt, C. Grabielle-Madelmont, F. Gazeau, V. Cabuil and S. Lesieur, Generation of Superparamagnetic Liposomes Revealed as Highly Efficient MRI Contrast Agents for in Vivo Imaging, *J. Am. Chem. Soc.*, 2005, **127**, 10676–10685.

53 R. P. Glahn, O. A. Lee, A. Yeung, M. I. Goldman and D. D. Miller, Caco-2 Cell Ferritin Formation Predicts Nonradiolabeled Food Iron Availability in an In Vitro Digestion/Caco-2 Cell Culture Model, *J. Nutr.*, 1998, **128**, 1555–1561.

54 M. L. Tomlinson, A. E. Hendry and G. N. Wheeler, Chemical Genetics and Drug Discovery in Xenopus, in *Xenopus Protocols: Post-Genomic Approaches*, ed. S. Hoppler and P. D. Vize, Humana Press, Totowa, NJ, 2012, pp. 155–166.

55 M. Mamusa, L. Sitia, F. Barbero, A. Ruyra, T. D. Calvo, C. Montis, A. Gonzalez-Paredes, G. N. Wheeler, C. J. Morris, M. McArthur, *et al.*, Cationic Liposomal Vectors Incorporating a Bolaamphiphile for Oligonucleotide Antimicrobials, *Biochim. Biophys. Acta, Biomembr.*, 2017, **1859**, 1767–1777.