Identification of *Cis*-Regulatory elements specific for the Neural Crest during *Xenopus* development

Anne Cathrine Hyde

Thesis submitted for the degree of Master of Science by Research

University of East Anglia

School of Biological Sciences

Norwich

United Kingdom

10 January 2022

Word count: 30,952

Abstract

The neural crest (NC) is a multipotent cell population that develops early in the embryo. The NC cells (NCC) migrate throughout the embryo and contribute to a range of different tissues including the craniofacial skeleton, the heart, and many different neurons. To achieve this multipotency, the NC is highly regulated by *cis*-regulatory elements (CREs) which can decide the special and temporal expression of genes required for NC formation. Some CREs specific for NC have been identified in a several different species, but few CREs specific for NC development have been identified in *Xenopus*.

CREs are believed to exist in open regions of the genome. Here we use an assay for transposase accessible chromatin using sequencing (ATAC-seq) to identify these open regions in *Xenopus* animal cap tissue induced to be NC. We have identified putative CREs from this data using bioinformatic approaches and then tested the activity of these putative CREs in a reporter assay by creating transgenic *Xenopus* embryos using a I-SceI meganuclease method.

We initially identified and cloned 20 putative CREs, and of these 12 showed either neural or specific NC developmental expression and are therefore believed to regulate different genes previously associated with NC development. These include putative CREs for *sox9* and *snai2*.

The regulation of NC development is very complex and to understand diseases associated with NC we need to understand this regulation. We have identified a good method for cloning the putative CREs identified in ATAC-seq data into a reporter vector for creating transgenic *Xenopus* embryos. This method has allowed us to identify several putative CREs specific for NC development.

Table of Contents

List of figures

List of tables

Acknowledgements

I would like to thank Professor Grant Wheeler for being a supportive supervisor who has given me an invaluable experience where I have learnt to work independently on my own project. You have given me great feedback on results and presentations and helped me think critically about both. I would also like to thank Dr. Claudia Buhigas for analysing the ATAC-seq data and helping me with the bioinformatics I performed in this project, and Victor Martinez Heredia for teaching me most of what I needed to know. Further I would like to thank Professor Andrea Münsterberg, Dr Tim Grocott and Dr Geoff Mok for their support and thought throughout my project. Lastly, I would like to thank everyone in the Wheeler lab for obtaining an endless supply of *Xenopus* embryos for my project.

Declaration

I declare that this thesis is an original piece of work under the supervision of Professor G. N. Wheeler. The work presented is my work unless stated otherwise, and follows the University of East Anglia's Policy on plagiarism and academic dishonesty. The ATAC-seq data collection was performed by Dr Marta Marín-Barba and the analysis of that data was performed by Dr Claudia Buhigas. With exception to the ATAC-seq data collection, some preliminary analysis by Dr Marta Marín-Barba, and some in *situ* images in the discussion, the work that is presented in this thesis has not been published or submitted for a qualification at this, or any other university.

List of abbreviations

A-P: anterior-posterior

AE: anterior expression

ANF: anterior neural fold

ATAC-seq: assay for transposase accessible chromatin using sequencing

B: brain

BA: branchial arches

BLAST: basic local alignment search tool

BMP: bone morphogenetic protein

ChIP-seq: chromatin immunoprecipitation and sequencing

CNC: cranial neural crest

CNS: central nervous system

CRE: *cis*-regulatory element

CRISPR: clustered regularly interspaced short palindromic repeats

D-V: dorsal-ventral

DPE: downstream promoter element

DS: dying cells

E: eye

ECT: ectoderm

EMT: epithelial-to-mesenchymal

END: endoderm

eRNA: enhancer derived RNA

EXRC: European Xenopus resource centre

F: fluorescence

FGF: fibroblast growth factor

GE: general expression

GFP: green fluorescent protein

GRN: gene regulatory network

HAT: histone acetyl transferase

HINT: Hmm-based IdeNtfication of transcription factor footprints

Inr: Initiator motif

KRAB: kruppel-associated box repressor

M: muscle

MES: mesoderm

MMR: Marc’s modified ringer

MNC: migrating neural crest

NC: neural crest

NCC: neural crest cell

NE: neuroepithelium

NF: neural fold

NOT: notochord

NP: neural plate

NPB: neural plate border

NT: neural tube

OT: otic vesicle

PCR: polymerase chain reaction

PNF: posterior neural fold

PPE: pre-placodal ectoderm

PR: pronephros

REMI: restriction endonuclease mediated integration

S: spotted

TAD: topologically associating domains

TBP: TATA-box binding protein

TF: transcription factor

TFBS: transcription factor binding site

TFIID: transcription factor IID

UCSC: University of California Santa Cruz

WH: whole half

WISH: wholemount in *situ* hybridisation

WNF: within neural fold

Chapter 1: Introduction

1.1 Vertebrate development

The vertebrates originated on the Earth about half a billion years ago and have diverged extensively since then becoming large and complex organisms including fish, amphibians, reptiles, birds, and mammals (York and McCauley, 2020, Kardong, 2015). However, even though they show such diversity, their development and body plan remain fairly similar as seen in

Figure 1 (Wolpert, 2019). In this thesis we focus on frog (*Xenopus*) development.

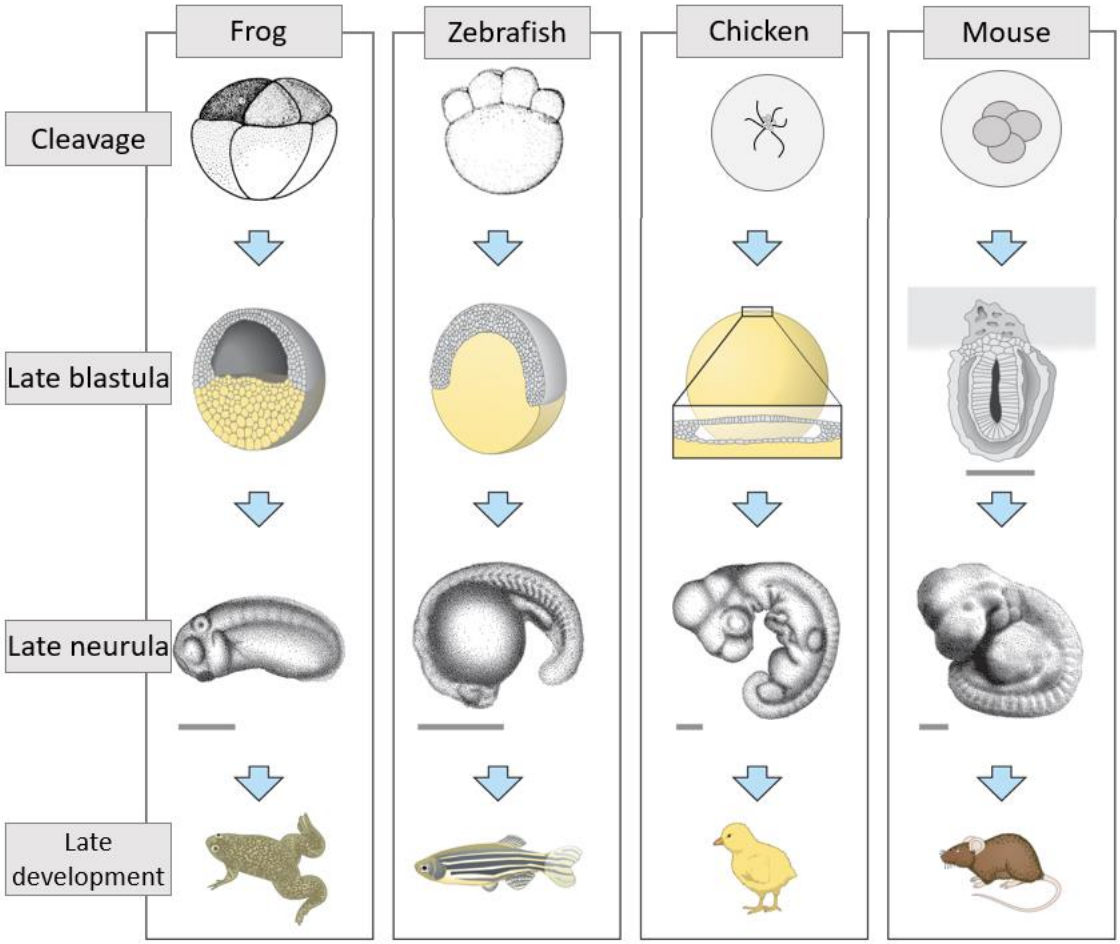


Figure 1: Comparison of vertebrate embryo development for common model organisms; frog, zebrafish, chicken, and mouse. There are several differences between vertebrates as they develop, however, they all go through the same stages. Staging is based on the frog. (First row) The frog cleaves with the vegetative cells, whilst the fish and chicken cleave on top of the yolk. The mouse (mammals) cleaves all cells, then the cells specialise to become inner cell mass (embryo) or trophectoderm (placenta). (Second row) Cross-section of late blastula embryos just before gastrulation. Yellow soe yolk cells, whilst the mouse has implanted into the uterine wall and developed extra-embryonic tissues. (Third row) After neurulation and the formation of the neural tube the embryos look very similar

as the neural tube, notochord, somites, head and tail are present. The scale bar is 1 mm. (Bottom row) Even though they develop somewhat similar, the adult organisms are very different. (The chicken is not an adult hen or rooster). Figure adapted from Gilbert (2014), Kimmel (1995), Nieuwkoop and Faber (1994), and Wolpert (2019) (Gilbert, 2014, Kimmel et al., 1995, Nieuwkoop and Faber, 1994, Wolpert, 2019).

1.1.1 Early *Xenopus* development

Once a *Xenopus* egg is fertilized, the zygote goes through several rapid divisions including the animal and the vegetal regions, where the cells do not grow, only multiply (Figure 1). Once the zygote has between 16 and 64 cells it is called a morula. When the zygote has 128 cells a fluid filled cavity develop called the blastocoel and it is now referred to as a blastula, which in frog is a sphere of cells, where each cell is called a blastomere (Wolpert, 2019, Gilbert, 2014, Nieuwkoop and Faber, 1994). The next stage of development is cell fating and gastrulation. For the frog, due to fate mapping, we know that the marginal zone of the blastula will become the mesoderm (MES) which gives rise to internal structures. The animal pole region becomes ectoderm (ECT), which later covers the whole embryo, whilst the vegetal regions become endoderm (END) (Figure 2). The blastula now gastrulates and is referred to as a gastrula, and the cells are no longer proliferating but moving and rearranging themselves into the correct positions of the body plan. This process can first be seen when the blastopore starts to form as a small depression at the surface of the dorsal blastula. The depression then moves laterally and ventrally to form a circle around the edge of the ECT which moves by a process called involution over the MES and END internalising the two tissues. During gastrulation, the MES is already starting its specification into structures such as the notochord and somites in dorsal regions. Kidney in the intermediate regions and the heart in anterior regions (Figure 2) (Gilbert, 2014, Nieuwkoop and Faber, 1994, Wolpert, 2019). Gastrulation is followed by neurulation in which the neural tube (NT) forms. From this stage the head, trunk, and tail form in an anterior-posterior (A-P) axis, whilst the backbone and the belly forms in the dorsal-ventral (D-V) axis (De Robertis and Kuroda, 2004). The embryo is now referred to as a neurula and can in the frog be visualised by the formation of neural folds (NF) at the edge of the neural plate (NP) on the dorsal side of the ECT, opposite to the closing blastopore. The NF rise and fold towards the midline of the embryo where they fuse creating the NT which separates from the epidermis above. The anterior NT gives rise to the brain, whilst the rest of the NT will later become the spinal cord (Kimmel et al., 1995, Nieuwkoop and Faber, 1994, Wolpert, 2019). Neurulation in the frog will be discussed in more detail in subchapter 1.2.

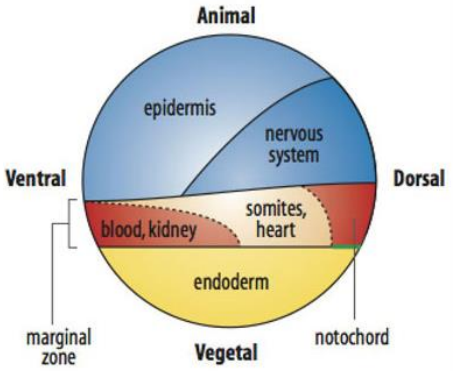


Figure 2: Patterning of the embryonic tissues in a Xenopus embryo during gastrulation. In the animal region, the ectoderm (blue) is patterned into the epidermis and the nervous system. The marginal zone containing the mesoderm (red and peach) is patterned into the notochord dorsally in the embryo, somites and heart in the mid, and blood and kidney ventrally. The endoderm (yellow) becomes the gut. Green line is the blastopore. Picture taken from (Wolpert, 2019).

1.2 The neural crest

The neural crest (NC) is a migratory embryonic cell population that was first described and studied by Wilhelm His in the chick in 1868 (Dupont, 2018, York and McCauley, 2020). It has since then been considered a hallmark of the vertebrates as it is responsible for the superior sensory organs, sophisticated organisation of the brain and the powerful jaw (Hoppler and Wheeler, 2015). Because of this it is believed that the origin of the NC was a seminal event in early vertebrate history distinguishing the vertebrates from invertebrate chordate relatives (York and McCauley, 2020). However, this has caused a challenge when studying NC cells, as one cannot use the classical organisms such as yeast or flies to study them. Further, in vertebrate models such as the mouse embryo development occurs internally when the NC is induced, and the embryo is very small and challenging to work with at this stage. Therefore, models such as the chick, zebrafish and frog have become common models to use when studying the NC (Gammill and Bronner-Fraser, 2003). The NC has classically been studied by transplantation experiments in amphibians, and in birds by quail-chick chimeras, showing its migratory and differentiation abilities (Simões-Costa and Bronner, 2015). In this thesis, NC induction and development will be investigated in relation to frog (*Xenopus*) development, and in the following subchapters I have attempted to separate the process of NC induction into patterning of the dorsal ECT, defining the NP border (NPB), NC induction and NC migration, although several of the signaling pathways overlap in the different stages.

1.2.1 Patterning of the dorsal ectoderm

The induction and subsequent migration of the NC cells (NCC) occur through a series of complex molecular interactions. As the blastopore closes at the end of gastrulation the dorsal

ECT of the *Xenopus* embryo is patterned into three tissues. The neuroepithelium (NE), also called the neural ECT or the NP; the nonneural ECT which will from the epidermis; and the region between these two which will for the most part form the NC and placodes (Figure 3). (Hoppler and Wheeler, 2015, Le Douarin, 1980, Prasad et al., 2019, Gammill and Bronner-Fraser, 2003). Placodes will be discussed in more detail in subchapter 1.3. The division of these three tissues starts with the specification of the NPB which can be established by a combination of different signals. (1) The prospective ventral non-neural ECT and ventral MES secrete bone morphogenetic protein (BMP). The involuting dorsal MES and the NP secrete BMP antagonists such as Chordin, Cerberus, Noggin and Follistatin. This causes a dorsoventral gradient of BMP in the ECT, where it is absent in the NP promoting a neural fate, intermediate at the NPB, and high in the ECT causing the activation of ectodermal markers (Gammill and Bronner-Fraser, 2003, LaBonne and Bronner-Fraser, 1998, Prasad et al., 2019). (2) Further to the BMP gradient, the dorsolateral marginal zone or the prospective paraxial MES situated underneath the prospective NC secrete Wnt (Wnt3a and Wnt8a) and fibroblast growth factor (FGF) which together with the intermediate concentration of BMP cause the NCC fate (Gammill and Bronner-Fraser, 2003, LaBonne and Bronner-Fraser, 1998, Monsoro-Burq et al., 2005, Prasad et al., 2019, Steventon et al., 2009). Later, BMP signalling is re-activated by Wnt3a at this now created ‘zone of competence’, which is the area in the ECT where the NC can from, to maintain the NC marker expression. This way, the concentration of BMP is ideal for further activation of NC specifiers (Gammill and Bronner-Fraser, 2003, Pla and Monsoro-Burq, 2018, Prasad et al., 2019).

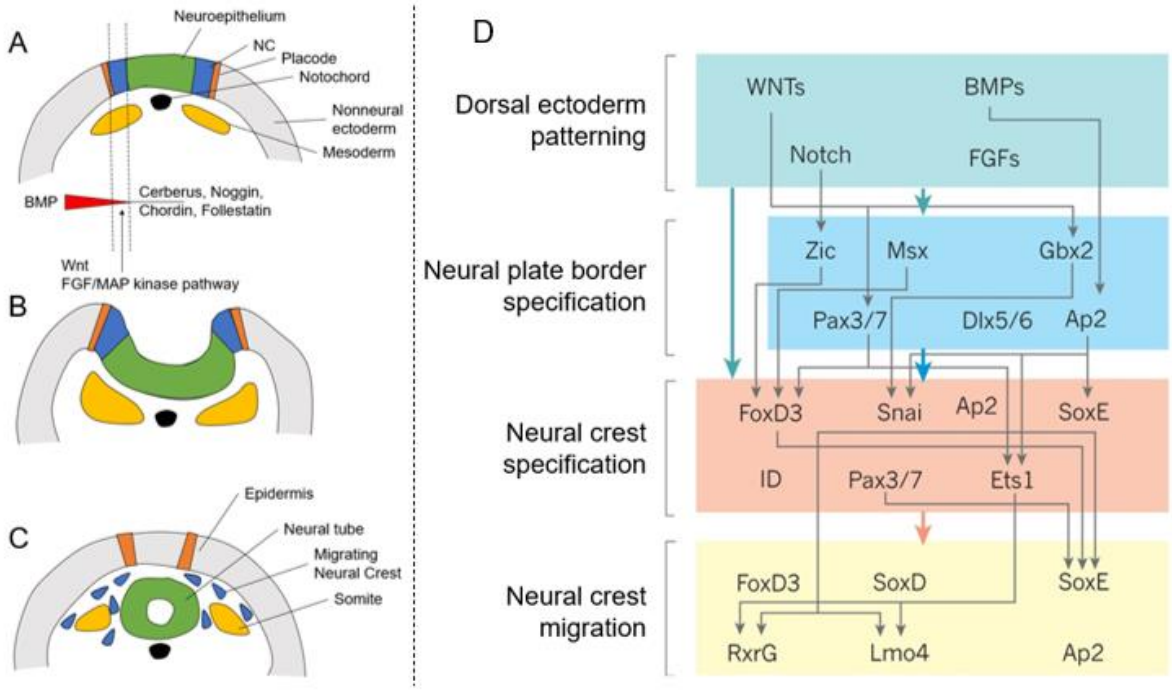


Figure 3: Induction of the neural crest. A: The region for NC cell specification is obtained by the concentration of BMP being high enough to overcome the antagonistic effect of neuroepithelium specifiers. But not so high that it specifies nonneural ectoderm. B: With the regions of the NP specified the NP invaginates to form the NT. C: The NC goes through EMT and migrates though out the developing body. D: Part of the gene regulatory network involved in neural crest development and migration from (Green et al., 2015).

1.2.2 Neural plate border specifiers

The NPB now contains cells that are not yet NCC but that have the potential to become NCC, as well as ectodermal placodes, epidermal cells, roof plate cells and sensory neurones of the central nervous system (CNS) (Simões-Costa and Bronner, 2015). A brief overview of the gene regulatory network (GRN) for NC induction can be seen in Figure 3D. As explained in the previous paragraph, a carefully regulated concentration of BMP causes expression of NPB specifiers in the 'zone of competence' in the patterned dorsal ECT. These include transcription factors (TFs) such as Zic1, Msx1/2, Pax3/7, Dlx5, Gbx2, AP2a And SP5. Many of these NPB specifiers are originally expressed not only in the NPB but in portions of the NP and the prospective epidermis as well until the boundary sharpens (Simões-Costa and Bronner, 2015). The carefully regulated concentration of BMP cause induction of *msx1, pax3* and *zic1* expression, and low levels of BMP can cause d*lx5, zic3* and *hes4* expression (Prasad et al., 2019, Monsoro-Burq et al., 2005). *Dlx5* expression causes NP specification, however, it has been shown that ectopic expression of *dlx5* does not cause NC marker expression (Gammill and Bronner-Fraser, 2003). Further, Wnt can directly cause *gbx2* expression, and Gbx2 can then cause expression of *max1* and *pax3*. FGF signalling is required for induction of *hes4*, and

as well as Wnt, FGF is required for *pax3*, *tfap2a* and *mesx1* activation which are important TFs for NPB specification (Pla and Monsoro-Burq, 2018). *Pax7* is known as one of the earliest NC markers and its expression can be regulated by FGF, Wnt and retinoids as well as *pax7* expression being activated by *pax3* expression, which can also activate its own *pax3* expression (Maczkowiak et al., 2010). Pax3 can together with Zic1 cause a NC fate, and they have specifically been associated together with regulating *axin2* and *cyp26c1* expression which in turn regulate *Wnt* expression. Pax3 and Zic1 can also cause activation of *znf703*, which regulates *snai2* and *sox10* (Prasad et al., 2019, Gammill and Bronner-Fraser, 2003, Pla and Monsoro-Burq, 2018, Hong and Saint-Jeannet, 2017, Plouhinec et al., 2014). This complicated network of protein interactions emphasises that several proteins are involved in the formation of the NPB, and they cross-regulate each other and interact in an intricate GRN.

1.2.3 Neural crest specifiers

With the NPB specified, the TFs present in the 'zone of competence' together with other signalling events induce the cells' fate to become pre-migratory and migratory NC through the expression of NC specifiers. These NC specifiers include *sox8, sox9, sox10, foxD3, snai1/2, cMyc* and *id* genes. Continued expression of *pax3/7* and *AP2a* is also seen. Here we will go through some of the NC specifiers and how they function to create the NCC. SoxE TFs (Sox8/9/10) are expressed in the NPB after NC induction. Sox9 and Sox10 are high mobility group (HMG)-domain transcriptional activators (Gammill and Bronner-Fraser, 2003). The expression of SoxE genes is variable in different species and that is why here as stated I focus on *Xenopus*. SoxE gene expression is very similar and overlapping. During NC induction Sox9 and Sox10 repress differentiation by maintaining the multipotency state of the cells. Further, in later stages Sox10 is involved in terminal differentiation of the NC into melanocyte sand glia, whilst Sox9 initiates the ectomesenchyme differentiation (Aoki et al., 2003, Prasad et al., 2019). *Sox9* expression can be induced by AP2α and a combination of Gbx2 and Zic1, and both *sox9* and s*ox10* can be repressed by Id TFs, such as Id3. Further, Sox9 and Slug (Snai2) can activate s*ox10* expression (Gammill and Bronner-Fraser, 2003, Pla and Monsoro-Burq, 2018, Prasad et al., 2019). Slug, also known as Snai2, is a zinc-finger transcriptional repressor, in a family with Snai1. They are paralogous TFs that have evolved due to a gene duplication event in the protochordate vertebrates. They are some of the most well-known markers of the NC and *snai1* has been observed in the embryos as early as the blastula stage. *Snai2* is first detected in the NPB where it is involved with specification and migration of NC. *Snai2* can be regulated by Wnt8 and intermediate levels of BMP, as well as Notch signalling through the Hairy2 TF (Gammill and Bronner-Fraser, 2003, Pla and Monsoro-Burq, 2018, Prasad et al., 2019). In fact, the *snai2* promoter contains a lymphoid enhancer-binding

factor/T-cell factor (LEF/TCF) binding site which mediates Wnt signalling (Gammill and Bronner-Fraser, 2003, Vallin et al., 2001). NPB specifiers such as Zic1, Msx1 and Pax3/7 also regulate *snai2* in the presence of Wnt signalling in *Xenopus* (Prasad et al., 2019). Downstream of BMP and cMyc signalling, Id (Inhibitor of differentiation) TFs are involved in maintaining the multipotency state of NC by regulating expression of genes for proliferation and differentiation (Prasad et al., 2019). As with Id and Snai TF FoxD3 maintain NC multipotency by preventing early differentiation, and it has been proposed that FoxD3 can bind to nucleosomes and open compact chromatin to assist other TFs acting as a pioneer factor (Gammill and Bronner-Fraser, 2003). Pax7, Msx1/2, and Ets1 act as upstream regulators of *foxD3* in cranial and trunk NC, and Zic1 specifically regulates *foxD3* in the trunk NC. Further, Notch signalling through Hairy2, Msx1, and a combination of Zic1, Pax3/7, and Wnt induce *foxD3* expression, and as a further layer of regulation f*oxD3* in *Xenopus* is autoregulatory (Gammill and Bronner-Fraser, 2003, Prasad et al., 2019, Pla and Monsoro-Burq, 2018, Sato et al., 2005).

However, not only TFs regulate the NC. Adaptor proteins that do not interact with DNA, but rather with the TFs themselves are involved in the signalling cascades. In *Xenopus*, Ajuba Lin proteins act as co-repressors of *snai1* and *snai2* during NC formation, and ectopic expression of LIM domain TF LMO4 activate *snai1* and *snai2* expression in the NPB and the nonneural ectoderm (Prasad et al., 2019, Pla and Monsoro-Burq, 2018).

1.2.4 Neural crest migration

The NC is a migratory cell population that can differentiate into structures throughout the vertebrate body. Some structures include the cartilage and the bones of the craniofacial skeleton, chromaffin cells of the adrenal medulla, melanocytes of the skin, many sensory neurones and glia and tooth and heart primordia (Figure 4) (York and McCauley, 2020). However, the NC is not believed to be a defined population of cells before they migrate. Before this time, they are thought to be a population of cells in the NFs that have the potential to become migratory NCC (Gammill and Bronner-Fraser, 2003). When the NCC is ready to migrate, they go through an epithelial-to-mesenchymal transition (EMT) to delaminate from the NE. Migration then occurs in a rostrocaudal wave with the cranial NC (CNC) going though EMT first followed by the trunk NC (Prasad et al., 2019). CNC all delaminate at the same time, whilst the trunk NC progressively delaminate from anterior to posterior. In frog this migration starts whilst the NP is still open. The main TFs involved in EMT in the NC include Snai1/2, FoxD3, and Sox9/10, as well as Ets1 and Sox5 in the cephalic regions. Snai1 induces expression of *snai2/slug, foxD3*, and *ets1*. The major event in EMT is the transition from E-cadherins to N-cadherins to weaken cell-cell adhesion. This event is controlled by the

stabilization of hypoxia inducible factor 1 α (Hif-1α) through Twist (Szabo and Mayor, 2018). However, Snai1/2 have been implicated in the direct repression of E-cadherin and claudins/occludin causing adheren junction break up and destruction of tight junctions (Gammill and Bronner-Fraser, 2003). After NC delamination CNC migrate through the tight spaces between the epidermal and mesodermal layers in a large sheet that splits into distinct streams. The position of the streams corresponds to the Hox gene positions along the anterior to posterior axis of the NT. The trunk NC migrate in thinner streams aligning with the somite development along the anterior to posterior axis. This NC follow one of two routes, (1) the ventromedial path along the NT and notochord or (2) the dorsolateral path immediately blown the dorsal ECT (Szabo and Mayor, 2018, Raible et al., 1992). The NC continuously interact with their surrounding tissues as they migrate. Signalling through Eph/ephrins prevent NC migration into specific regions and keep the cells on their paths (Mellott and Burke, 2008, Szabo and Mayor, 2018, Shellard and Mayor, 2016). Another way for the NC to migrate is by chemotaxis, either secreted by the surrounding tissue or from the NC itself (Szabo and Mayor, 2018, Shellard and Mayor, 2016). The migration of the NC has been thoroughly investigated by Le Douarin who devised a method to make chick-quail chimeras. The two birds had different organisation of the heterochromatin during interface, and their cells are therefore easily distinguished in grafted chimeras and can be followed when migrating (Le Douarin, 2004). In her studies she substituted different regions of the NP, and after the careful activation of NC genes, she could follow where the different regions along the anteroposterior axis migrated to, and what the NC become (Le Douarin, 2004).

The regulatory network that governs the expression of the NC genes is a highly complex system that is difficult to decode due to it being a spatial and temporal continuum. As has been shown in this subchapter it is very difficult to study NC development in sections depending on what developmental stage the embryo is going through (Williams et al., 2019, York and McCauley, 2020).

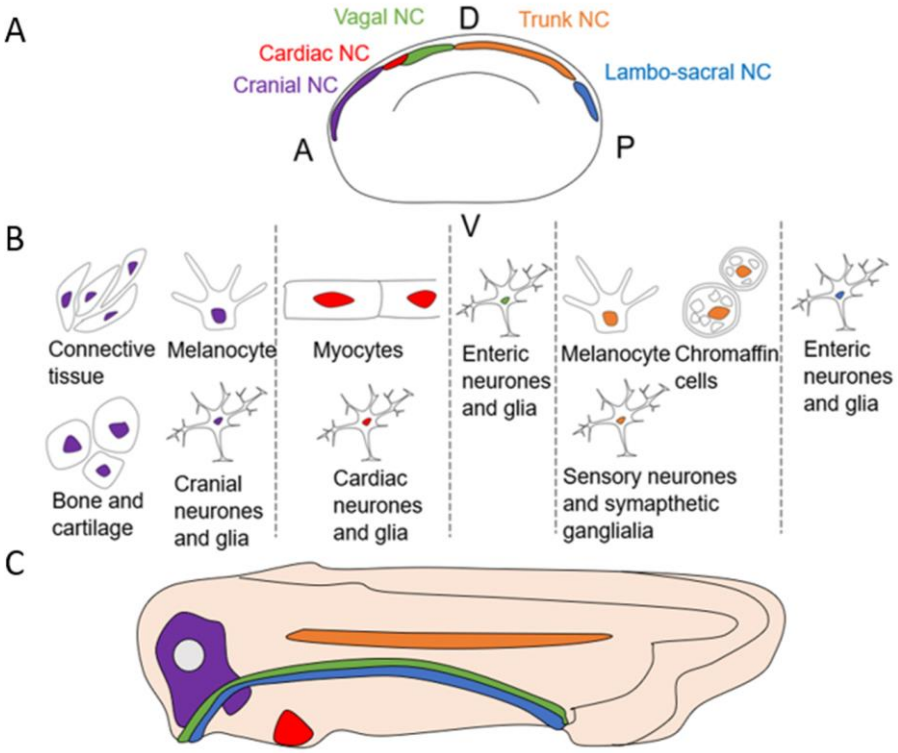


Figure 4: Migration of the neural crest. A: The region along the NT where the NC forms already determines what progenitor cells it can become. B: Some cell types the NC can become depending on where along the NT they develop. C: Locations in the tailbud the NC migrates to.

1.2.5 Diseases of the neural crest

As explored, the NC has a highly complicated GRN. If the GRN is disrupted, and there is a defect in the specification, migration, or differentiation of the NC, it can lead to a wide array of congenital disorders. These diseases are commonly known as neurocristopathies and can affect several different organs. Many of them are characterised by malformations in the head, the heart, or the peripheral nervous system (Sánchez-Gaya et al., 2020, Vega-Lopez et al., 2018). Craniofacial malformations are among the most common neurocristopathies and can be seen in more than 700 different syndromes. One example is where this is the case is craniosynostosis, where one or more of the sutures in the cranium of an infant fuses prematurely. Another common neurocristopathy is the well-known cleft palate phenotype (Kabbani and Raghuveer, 2004, Vega-Lopez et al., 2018). More specific neurocristopathies include Hirschsprung diseases which affects 1 in 5000 live births and is the main genetic cause for functional intestinal obstruction. It is caused by a lack of enteric ganglia along the intestine due to a defect in NC development (Amiel and Lyonnet, 2001). Bamforth-Lazarus Syndrome is also a neurocristopathies that affects 1 in 4000 live births. It has been linked to among other mutations in *foxe1* causing loss of thyroid function, cleft palate, and spiky hair (Vega-Lopez et al., 2018). However, neurocristopathies also include tumours, such as melanomas and

neuroblastomas, which are extracranial solid tumours that occur during childhood (Nakagawara et al., 2018, Vega-Lopez et al., 2018).

Although many neurocristopathies are linked to mutations in genes important for NC development, some neurocristopathies are not linked to mutations in specific genes, but rather the non-protein coding region of the DNA assumed to house most of the regulatory regions (Sánchez-Gaya et al., 2020, Pennacchio et al., 2013). These can be direct mutations to the sequence which alter the expression of the gene the regulatory region governs over, or it can be due to changes in the chromatin structure causing a disruption in the access to the regulatory region, meaning the regulatory region and the target gene can no longer communicate. One example of this scenario was found in a patient with branchio-oculo-facial syndrome, which is a neurocristopathy characterised by severe facial, hearing, and cutaneous anomalies. Although this phenotype is commonly caused by disruption of the *TFAP2A* gene, this one patient had both alleles of *TFAP2A* intact, but had a long heterozygous inversion that caused a disconnection between one *TFAP2A* allele and a cognate enhancer. This led to *TFAP2A* monoallelic and haploinsufficient expression in CNC (Krijger and de Laat, 2016, Maurano et al., 2012, Sánchez-Gaya et al., 2020, Laugsch et al., 2019).

1.3 Placodes

A second type of tissue that is a key marker for vertebrates are the placodes which give rise to many of the sensory structures in the vertebrate head such as the cranial ganglia and organs of special sense. They can generally be divided into adenohypophyseal, olfactory, lens, trigeminal, epibranchial, and lateral line placodes. They develop in the region referred to as the pre-placodal ECT (PPE), in the non-neural ECT just lateral to the plate border (Figure 3). The pattern of the placodes looks like a horseshoe, wrapping itself around the anterior NP. Later they divide into smaller clusters of cells representing their subsequent progenitors. They then undergo either invagination or delamination to then differentiate into various sensory cells (York and McCauley, 2020). The induction of the PPE uses many of the same molecules as the induction of the NC such as the BMP, FGF and Wnt pathways and their specification is controlled by s*ix* and *eya*. Even though there are several differences between NC and placodes, such as NC migration though EMT and the placodes mostly forming structures by invagination, there are also many similarities. These similarities are important as both cell types work intricately together to form the vertebrate head (York et al., 2020, Graham and Shimeld, 2013).

1.4 Chromatin organisation

To fit the genome of a eukaryotic organism into a cell, regulate biochemical activities such as enhancer regulation, and establish cell identity during development, the genome needs to be packaged into chromatin. Firstly, the DNA is tightly bound around a group of proteins called histones forming a nucleosome. The nucleosome consists of eight histone molecules - a histone octamer - two of each H2A, H2B, H3 and H4, and about 147 base pairs of DNA wrapped around it twice (Figure 5). A ninth histone, H1, can be found on the outside of the nucleosome holding it all together. A quarter part of the histone is made up of arginine or lysine, which are very positive amino acids, therefore they have high affinity for the negatively charged DNA. Further, the histones have a flexible amino-terminal tail that contains several lysine and arginine residues that extends away from the nucleosome (Berg et al., 2015). Post-transcriptional modifications to these tails can have significant effects on the chromatin conformation, and therefore they are readily modified by histone modifiers which can be recruited by regulatory sequences, TFs, and co-activator complexes. Histone modifiers may add acetyl, methyl, phosphates, or any other modifications to the histones or histone tails. Histones can also be swapped out for other variants of histones. This allows for functional diversity in the nucleosome (Buenrostro et al., 2013, Perino and Veenstra, 2016, Zovkic, 2020). All these modifications to the histones alter the charge of the nucleosome and therefore its ability to keep the DNA wrapped around it. It can cause an opening of the chromatin to allow access to *cis*-regulatory regions or closing the chromatin so genes cannot be transcribed (Buenrostro et al., 2013, Perino and Veenstra, 2016).

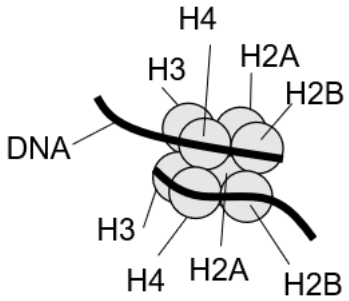


Figure 5: Nucleosome structure. Eight histone molecules (H2A, H2B, H3, and H4) are wrapped by DNA forming a nucleosome to compact the chromatin into the nucleus.

Within the chromatin, the DNA is divided into topologically associating domains (TADs) which are 3D regions of high interaction. They can be about 1 Mb in size and neighbouring TADs are separated by boundary elements bound by clusters of isolator molecules such as the zinc finger TF CTCF (Figure 6A). The formation of a TAD is proposed to occur though a model called the loop extrusion model. In this model the TAD forms a loop by the cohesin molecule

which forms a ring around the DNA making the loop. The cohesin then moves along the DNA until it encounters the boundary specified by CTCF (Figure 6B). This brings distant DNA elements in closer proximity to one another, allowing for more efficient interactions (Ibrahim and Mundlos, 2020, Sikorska and Sexton, 2020, Bolt and Duboule, 2020). If there are deletions, inversions or translocations in the genome that disrupt the 3D organisation of the TAD it can causes disease due to for example loss-of-function of an enhancer as that enhancer and its target promoter are disconnected, or gain-of-function as an enhancer gains regulatory ability over a new gene (Ibrahim and Mundlos, 2020). However, even though the general idea of TADs say that gene regulation is restricted to within them, there are observations within gene regulation that contradict this model. If a cell is depleted of cohesin and CTCF, there is little effect on the gene expression, suggesting that the TAD is not as important as one would think (Ibrahim and Mundlos, 2020, Ing-Simmons et al., 2021). Further, when looking at the well-studied dorsoventral patterning in the *Drosophila* Ing-Simmons (2021) saw that even though one would expect the TADs to be different in different cell types as chromatin states and gene expressions are different, the 3D chromatin organisation remained mostly the same in all cells, further suggesting that cell specific gene regulation does not rely on specific chromatin organisation. However, they still proposed that enhancers and target promoters were found in regulatory domains (Ing-Simmons et al., 2021).

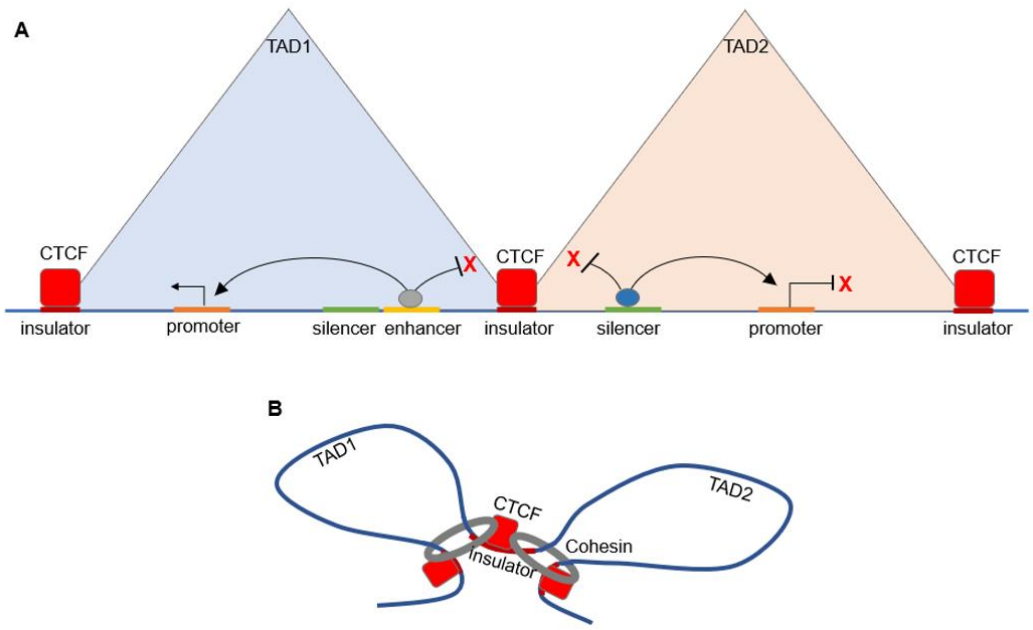


Figure 6: Proposed TAD structure. A TAD is restricted by CTCF proteins bound to insulator elements. Within the TAD enhancers and silencers affect promoters, however they cannot generally affect the promoters in neighbouring TADs. B: TADs are believed to have a 3D structure generated by cohesin creating a loop extrusion structure where cohesin wraps around the DNA and makes a loop, cohesin then moves along the DNA until it encounters CTCF proteins bound to insulator elements which marks the border between TADs.

1.5 Genes can be controlled by *Cis*-regulators

In a complex organism, all cells contain the same genome, and a high degree of regulation allows for cells to be different and to make different tissues. This regulation termed epigenetic regulation was first explored by Waddington in 1942 where he brought together ideas of regulatory networks where changes in one gene, the *genotype*, during early development, may cause greater defects, or change the *phenotype* (Waddington, 2012). Much of this regulation is done in *cis*, meaning that non-protein coding, functional DNA elements regulate gene expression on the same chromosome they exist on (Savarese and Grosschedl, 2006). These *cis*-regulatory elements (CREs) are often found in the non-protein coding genomic regions or introns which makes up about 98% of the DNA. They are bound by different proteins such as TFs and long non-coding RNAs acting in *trans*, meaning the factors can be transcribed from a different chromosome (Elgar and Vavouri, 2008, Sánchez-Gaya et al., 2020). Promoters and enhancers are the major form of CREs (Sánchez-Gaya et al., 2020, Savarese and Grosschedl, 2006). Promoters are found at the start of a gene sequence and are bound by transcriptional regulators that cause initiation of transcription. These factors are highly conserved and include the RNA polymerase II among other general TFs (Sánchez-Gaya et al., 2020). Enhancers are regulatory sequences that can be found far away from the promoter and are bound by TFs causing interactions with the promoter, and regulation of gene transcription (Panigrahi and O'Malley, 2021). Insulators and silencer sequences as well as more complex locus control are also included as CREs (Sánchez-Gaya et al., 2020). These CREs will be explored in more detail in the following paragraphs.

1.5.1 Enhancers

An enhancer is a form of CRE that governs the spatiotemporal and quantitative expression of a target gene by interacting with the gene's promoter (Sánchez-Gaya et al., 2020, Karnuta and Scacheri, 2018, Panigrahi and O'Malley, 2021). The concept of enhancers was first explored by Banerji in 1981 where they identified a 72 base pair long DNA sequence that enhanced the expression of β-globin in HeLa cells (Banerji et al., 1981). However, the way enhancers work is still to this date not fully understood (Panigrahi and O'Malley, 2021). It is believed that enhancer can act independently of the distance from the target genes promoter as exampled by the discovery of an enhancer for *Shh* located one million bases from the gene itself (Lettice et al., 2003). Further, enhancers are believed to be independent of their orientation to the gene's promoter, and although they are mostly found in the intergenic and intronic regions of the DNA, they can also on occasions be found within exons (Krijger and de Laat, 2016). Enhancers are short non-protein coding sequences of about 100-1000 bp in

length and multiple enhancers may cluster together to form a super enhancer. What defines the enhancers are the binding sites for regulatory proteins (Panigrahi and O'Malley, 2021). The enhancer is also governed by the underlying DNA topology, meaning what state the enhancer is found in. These states are governed by histone modifications causing them to be either decommissioned, poised, primed or active. Closed chromatin usually has a high affinity for nucleosomes, and when TFs bind, they can rapidly undergo nucleosome depletion (Long et al., 2016, Sánchez-Gaya et al., 2020, Karnuta and Scacheri, 2018, Pennacchio et al., 2013, Panigrahi and O'Malley, 2021). The architecture and the function of enhancers will be further discussed in the following paragraphs.

1.5.1.1 Enhancer architecture and grammar

Enhancers can be built in several different ways referred to as enhancer architecture. This architecture is built of different enhancer grammar which refers to how the enhancer is organized. This involves binding affinity, number, spacing, orientation, order, local DNA shape and type of TF motifs (Long et al., 2016). There are three major models for enhancer architecture (Panigrahi and O'Malley, 2021) shown in Figure 7. In the billboard model of enhancer architecture TFs bind the enhancer either individually or as modules which can act independently from each other, the spacing, and the orientation that they are in (Arnosti and Kulkarni, 2005, Panigrahi and O'Malley, 2021, Vockley et al., 2017) (Figure 7A). Here it is more important that certain TFs are present rather than how they bind, and they may even contain TF binding sites (TFBSs) that are not a match to the consensus sequences allowing a certain specificity in the enhancer (Long et al., 2016). Therefore, there is lower conservation required within this type of enhancer. Evidence for this module was seen when 5000 synthetic liver enhancers were tested in a reporter assay. The synthetic enhancers were constructed of different combinations of 12 liver specific TFs and showed that the organisation of the enhancer had less effect on regulation (Smith et al., 2013, Vockley et al., 2017). In the TF collective TFs can be recruited by TFBSs as well as protein-protein interactions. This means that similar sequences can have a very different group of TFs binding indirectly, or different sequences motifs may have similar TFs binding (Figure 7B). This allows for higher evolvability again leading to a lower conservation (Panigrahi and O'Malley, 2021). These enhancers can be visualised by identifying TFs at a sequence that has no TFBSs for those TFs (Long et al., 2016). An example of a TF collective is seen in the dorsal mesoderm of the *Drosophila* embryo. Five TFs pMad, dTCF, Doc, Pnr and Tin are recruited to an enhancer for heart function in the only region they are all co-expressed, and this recruitment occurs even in the absence of any consistent motif grammar to recruit these specific TFs (Junion et al., 2012). Lastly there is the strict enhanceosome model. It has several interdependent modules, and if the sequence, order, or spacing of the modules are changed the enhancer stops working properly. Here TFs

bind cooperatively as a nucleoprotein complex, and the bound proteins may possess a high degree of cooperation which is why they stop working if the sequence is changed (Figure 7C). These enhancers are usually highly conserved (Arnosti and Kulkarni, 2005, Long et al., 2016, Panigrahi and O'Malley, 2021). One example of an enhansceosome is the mammalian interferon beta CRE. It is 65 bp long and the binding proteins HMG-I(Y) act as an architect and govern the precise binding of Rel family NF-κB proteins, the ATF-2/c-jun heterodimer, and proteins from the interferon regulatory family (INF) to the sequence. Once activated the enhancer cause activation of the IFNβ gene promoter. However, if any of the binding sites are mutated or moved, the enhancer stops working and IFNβ is not transcribed (Arnosti and Kulkarni, 2005, Thanos and Maniatis, 1995).

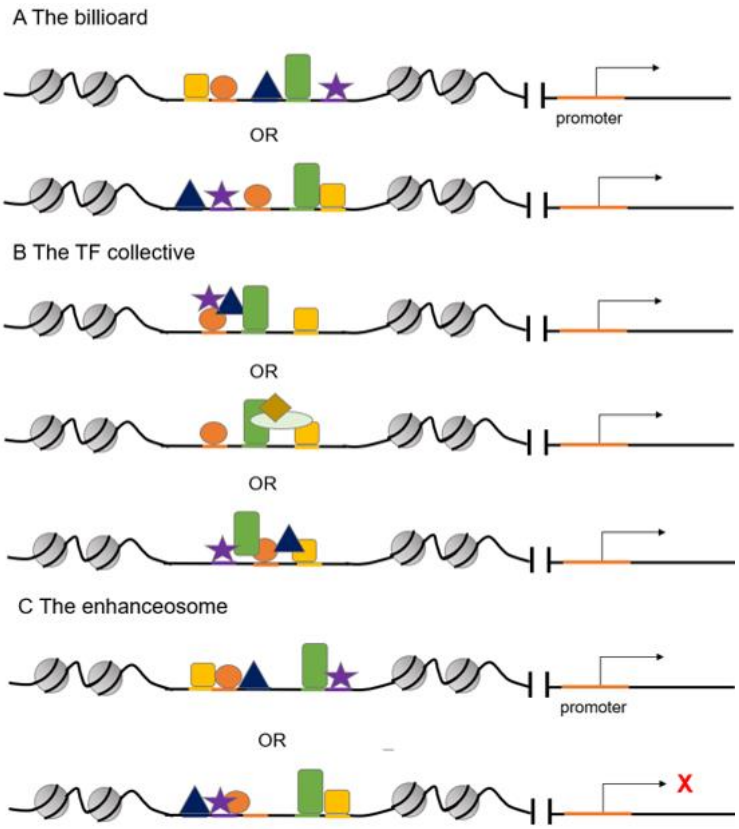


Figure 7: Three models for enhancer architecture. A: The billboard model suggest that TFs bind individually or as modules which can act independently of each other, orientation and spacing. It is more important at all TFs are present than how they are present. B: In the TF collective TFs bind the enhancer or interact with the enhancer through protein-protein interactions. This means the similar enhancer may different TFs bound, or different sequences may have the same TFs bound. C: The enhanceosome can only work if the correct TFs bind in the correct order. It is dependent of orientation, spacing and interactions.

1.5.1.2 Enhancer accessibility

When an enhancer is needed to cause activation of a gene, it must first enter an open state in the chromatin. This is one reason why several TFs are needed to bind at an enhancer, as a single TF is not believed to have enough effect on the closed chromatin confirmation to make it overcome the intrinsic affinity it has for the histone, meaning the TF cannot access the underlying DNA. Therefore, several TFs are needed to cooperatively overcome the energetic barrier of nucleosome eviction to allow for activation of the enhancer. This model is referred to as “direct cooperativity”. Alternatively, TFs can independently compete together with a histone for access to the DNA in a model referred to as “indirect cooperation” or “collaborative competition (Long et al., 2016). Further, pioneer factors such as TFAP2A associated with NC gene accessibility (Rothstein and Simoes-Costa, 2020), or master regulators may bind nucleosomal DNA at developmental enhancers and prime them for activation by recruiting chromatin remodelers such as histone acetyl transferases (HAT), which facilitate the removal of post-transcriptional modifications on histones (Long et al., 2016, Panigrahi and O’Malley, 2021).

1.5.1.3 Enhancer activity

Once the enhancer is accessible, other TFs, coregulators, chromatin remodelers and modifiers, reader proteins and in some cases RNA polymerase II are recruited to the enhancer. There can be a collection of hundreds of proteins at the enhancer which keeps the enhancer nucleosome deficient, a feature that can be used to identify enhancers (Panigrahi and O’Malley, 2021). The enhancer now must physically interact with the promoter of its target gene to cause transcription. However, in particular in higher eukaryotes, the enhancer can be found a considerable distance away from the target gene. Several models have been proposed for enhancer-promoter interaction which can be seen in Figure 8. These include the linking model where a transcriptional activator bound to the enhancer recruit proteins to form a chain that links with the promoter (Figure 8A). Although, it can be considered inconvenient to have such long chains of proteins within the wide genome (Panigrahi and O’Malley, 2021). Alternatively, this model may propose that the chromatin between the promoter and the enhancer is being reorganized to cause activation of the promoter. Again quite inconvenient for long distances, therefore, the linking model is more likely to be used where there are short distances between the enhancer and the promoter (Pennacchio et al., 2013). A second model is the tracking model, where a transcriptional activator that is recruited to the enhancer moves along the chromatin and looks for the promoter. Alternatively, it could be the RNA polymerase II that is recruited to upstream enhancers and tracks along the chromatin, pulling the enhancer with it until it meets the promoter (Figure 8B). This model however is challenging as it would heavily rely on motor proteins that would collide with the RNA polymerase II in genes that are

being transcribed especially if the enhancer is found in an intron, but also if the enhancer is not regulating the gene closest to it (Furlong and Levine, 2018, Panigrahi and O'Malley, 2021). Therefore, if this model is being used, it is likely used on enhancers that are in close proximity to the promoter (Pennacchio et al., 2013). Lastly, looping is the more favoured model. This classic model says that the enhancer recruits its cell- and condition-specific TFs then loop in a 3D space to allow the TFs to interact with proteins in the target promoter that the TFs have affinity for either directly or indirectly via co-activators (Figure 8C). This causes chromatin remodelling and the recruitment of the basal transcriptional machinery at the promoter causing enhanced transcription of the target gene (Furlong and Levine, 2018, Gasperini et al., 2020, Panigrahi and O'Malley, 2021, Pennacchio et al., 2013). The space in which looping occurs is proposed in some cases to be controlled by CTCF proteins within the TAD, although it has also been reported that enhances can regulate target genes on the other side of a CTCF boundary (Furlong and Levine, 2018).

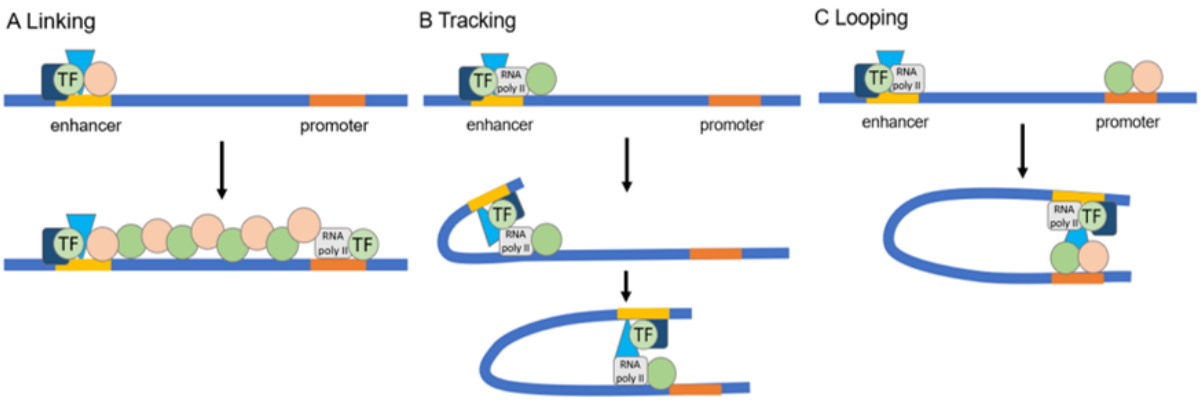


Figure 8: Three models for enhancer activity adapted from Furlong and Levine, 2018. A: The linking model propose that a chain of proteins link the enhancer and the promoter together to cause promoter activation by the enhance. B: The tracking model suggest that the transcription imitation complex form at the enhancer, then either a motor protein or RNA polymerase II mole along the DNA to identify the promoter. C: The looping model suggest that proteins bind at the enhancer and the promoter which have affinity for each other causing them to interact.

1.5.1.4 Multiple actions by enhancers

Depending on an enhancer's grammar and architecture it will function in different ways. Some genes such as developmental genes, can be regulated by multiple enhancers. All the enhancers may be active in one tissue regulating when and how much of the gene is being expressed. Alternatively, different enhancers may be active in different tissues causing the expression of the same gene in alternative locations at alternative times to control the developmental pattern of the gene expressed. Further, one enhancer may control the expression of several genes. Here, the genes may compete for which gene is going to be expressed. This system is used for olfactory neurones in the sensory system to make sure

only one olfactory neurone gene is expressed in each cell. Alternatively, one enhancer may activate all the genes it regulates, however not cause expression at maximum level (Krijger and de Laat, 2016, Zeitlinger, 2020). Lastly, one enhancer may have several different TFBSs so that different cell types can use the same enhancer, but express the different TFs that will allow for a different effect temporally or in different cells from the same enhancer (Panigrahi and O'Malley, 2021).

Further, enhancers are often referred to as non-coding or non-protein coding elements. However, in 2005 when analysing the transcriptional landscape in the mouse genome, several non-coding RNAs were identified and proposed to be transcribed from enhancers (Carninci et al., 2005). Tens of thousands of distinct enhancer derived RNAs (eRNAs) have been suggested to be found in vertebrate cells, outnumbering the number of mRNAs. They can be between 180 bp – 2 kb long, largely non-spliced and non-polyadenylated (Panigrahi and O'Malley, 2021). The turnover of eRNAs is quite quick and even though the function of eRNAs have yet to be discovered, they have been proposed to stabilise the enhancer-promoter interactions (Kolovos et al., 2012, Panigrahi and O'Malley, 2021).

1.5.2 Promoters, insulators, and silencers

Promoters are found at the 5' end of a gene, and the core promoter contains the transcriptional start site as well as several DNA elements where regulatory proteins can bind. These include the TATA box which can be bound by TATA-box binding proteins (TBP) which is a part of the Transcription Factor IID (TFIID) complex which recruits RNA polymerase II and mediates the assembly of the pre-initiation complex which starts transcription of the gene. Other DNA element within the core promoter is the Initiator (Inr) motif and the downstream promoter element (DPE), which in the absence of a TATA-box also can recruit the TFIID (Haberle and Stark, 2018, Riethoven, 2010). However, the level of transcription that can occur by the regulatory proteins at the promoter alone is usually very low, and therefore the transcription is commonly enhanced by other regulatory proteins bound further away from the gene. These CREs include enhancers (Latchman, Haberle and Stark, 2018), but also the proximal promoter element which can be found a few hundred base pair upstream from the core promoter and may also have binding sites for regulatory proteins (Maston et al., 2006).

Insulators can perform long range effects on genes by creating domains for regulation to work in. These domains are as previously discussed the TADs. Insulators can be positioned between a promoter and an enhancer or silencer, blocking the effect or the regulatory element on the gene (Figure 6). Alternatively, insulators can also block the spread of repressive chromatin into a region that needs to be open (Brasset and Vaury, 2005, West et al., 2002).

Insulator elements work in an orientation and position dependent manner (Maston et al., 2006). Further, they must be bound by insulator binding proteins to be functional as enhancer-blocking proteins. Several binding proteins have been identified in Drosophila, including su(Hw), BEAF, Zw5 and GAGA. However, only one have been identified in vertebrates yet, the CTCF protein (West et al., 2002).

Silencer elements can be bound by repressive TFs and thereby cause repression at a gene promoter and inhibit the transcription of the target gene. They generally work independently of their orientation towards the gene, and they can be located several kb away from the gene it regulates, or within the intron of the gene (Maston et al., 2006). They can often be found positioned next to enhancers in the genome where the activity of the regulatory element can quickly be change from activation to repression (Doni Jayavelu et al., 2020) (Figure 6). An example of this phenomenon is the regulation of the k*it* gene. *Kit* encodes a receptor tyrosine kinase essential for normal haematopoiesis, and a loss of *kit* leads to embryonic death. An enhancer for *kit* is bound by GATA2, forming an enhancer-promoter loop. If GATA1 expression activated, GATA2 is moved from the enhancer and the enhancer-promoter loop is disabled. GATA1 acts as a repressor binding to silencer element in proximity to the GATA2 bound enhancer (Jing et al., 2008). However, silencers can also be part of the proximal promoter element or work as an independent unit (Maston et al., 2006).

1.6 Conclusion

Vertebrate development consists of a complicated network of molecules and genes interacting to cause correct gene expression at the right time and place. Much of this regulation has stayed conserved though out evolution, but equally so a lot of it has also changed owing to the variability of organisms. If this network is disrupted it can lead to disease. Many diseases have been linked to mutations within genes, however, a considerable number of diseases have not been able to be linked to mutations within a specific gene, but rather to mutations in the non-protein coding regions of the DNA. In the non-protein coding DNA several *cis*-regulatory sequences for genes can be found which are necessary for proper gene expression. Therefore, mutations in the non-protein coding DNA, may alter the grammar of a *cis*-regulatory region causing mis-expression of the gene linked to the disease. To understand these diseases better, we need to identify these regulatory sequences. In this project, we look for *cis*-regulatory sequences that regulate genes required for neural crest development. The neural crest is a multipotent tissue that migrates throughout the early embryo and contributes to several tissues including the cranial skeleton, skin, and heart to mention a few. Therefore, identifying *cis*-regulators for neural crest specific genes would allow us to understand diseases linked to the neural crest better.

1.7 Research aims and objectives

Hypothesis: Open regions in the *X. laevis* genome identified in the ATAC-seq as exclusively open in neural crest are *cis*-regulator elements specific for neural crest development.

Although researchers have been identifying CREs for many years now, few CREs specific for NC development have been identified. In a previous PhD project in the Wheeler lab, Marta Marín-Barba performed a method to identify open region in the chromatin called ATAC-seq on *X. laevis* animal caps induced to become NC, ECT and NE (Marín-Barba and Wheeler, unpublished results).

In this project we aim to:

- Identify CRE's using bioinformatic analysis of ATAC-seq data.
- Analyse the CRE sequences for putative TFBSs.
- Clone putative CRE's into a reporter plasmid.
- Test the CRE's activity by using a *Xenopus* transgenics reporter assay in *X. laevis* and *X. tropicalis*.

Chapter 2: Materials and Methods

2.1 Xenopus as a model

In this project we use *Xenopus laevis* and *Xenopus tropicalis*, also known as the African Clawed frog as a model organism. *Xenopus* is a common model organism in developmental biology due to being easily maintained and bred, they lay large clutches of eggs externally, and has a rapid development. Further, 79% of identified human disease genes have verified orthologues in *Xenopus*, and together with *Mus muluscus and Gallus domesticus* it is one of the best-known vertebrate models (Buisson et al., 2015, Cannatella and de Sá, 1993).

2.1.1 Embryo harvesting

X. laevis and X. tropicalis male and female adults were provided by European Xenopus Resource Centre (EXRC) based at the University of Portsmouth. Female were induced by injection of ovulation-stimulating hormone in the dorsal lymph sac, and a male *X. laevis* was anaesthetised by 1 g MS-222 in 300 ml water for two hours at 4°C, then dissected for its testis. The testis was stored in testes buffer at 4°C. *X. tropicalis* sperm was stored in -80°C, then thawed at 37°C before used. Both procedures were done by a trained license holder. The females were encouraged to lay eggs by gentle squeezing into a 90 mm x 16.2 mm petri dish where the eggs were fertilised in vitro by half of one harvested testis crushed with pestle to release sperm in 1 x MMR (1 x Marc's Modified Ringer: 100 mM NaCl, 2 mM KCl, 1 mM $MgCl_2$, 2 mM $CaCl_2$, 5 mM HEPES, pH 7.5) for *X. laevis*, or thawed sperm resuspended in 125 µl 0.1 x MMR for *X. tropicalis.* The tube was then rinsed with another 125 µl 0.1 x MMR to release all the sperm and added to the eggs. The eggs were then spread out in the petri dish and left to incubate for 5 minutes for *X. laevis* and 10 minutes at 26 °C for *X. tropicalis*. The eggs were flooded with 0.1 x MMR for both species and incubated for 20 minutes before the solution was substituted for 2 % cysteine to de-jelly the embryos for 7 minutes. The embryos were washed twice with 1 x MMR and twice with 0.1 x MMR for *X. laevis* and twice with 0.1 x MMR and 0.05 x MMR for *X. tropicalis*, then placed in a 150 mm x 15 mm petri dish coated with BSA and stored in 0.1 x MMR or 0.05 x MMR. Embryos were grown in incubators between 18 - 23°C for *X. laevis* and 20 - 26°C for *X. tropicalis* and staged according to the Nieuwkoop and Faber's Normal Table of *X. laevis* (Nieuwkoop and Faber, 1994).

2.2 Identification of open chromatin regions

The ATAC-seq from the NC, NE and ECT were aligned in the UCSC genome browser to the *X. laevis* v9.2 genome obtained from Xenbase.org and differentially compared by a

bioinformatician. The open chromatin regions that were significantly different in the NC were identified and chosen for further analysis such as conservation analysis. Lastly the ATAC-seq data was compared with RNA-seq data by a bioinformatician.

2.3 Predicting transcription factor binding sites

For CREs to be activated they need to be bound by proteins such as TFs. To increase the significance of hypothesis that the open chromatin regions we choose to study were in fact CREs, we predicted what TFs may bind to the cloned sequences. We used the UCSC genome browser's own predictions as to the TF sites in the sequences we were studying, and HINT-ATAC data produced by Dr Claudia Buhigas was used to predict TFs bound to the sequence at the time of the Tn5 transposase cleavage (Mok et al., 2021).

2.4 Making transgenic construct

The transgenic constructs containing the putative CREs were made by following the procedure explained by Ogino (2006) (Ogino et al., 2006a, Ogino et al., 2006b) with some modifications, using a construct given to us by Dr Robert Granger.

2.4.1 Cloning of open chromatin regions

Primers for the chosen open chromatin regions can be seen in appendix 3. Appendix 5 has a detailed pipeline protocol for choosing primers all the way to making the transgenic embryos. The primers were designed with either Not1 or Sbf1 restriction enzyme sites for future ligation purposes. The efficiency of the primers was tested by PCR using 6 µl Biomix TAQ polymerase (meridian bioscience), 0.5 µl of forward and reverse primer, 7.5 µl SW H_2O, and 1 µl of 200ng/µl *X. laevis* genomic DNA and following the PCR program in appendix 1. *X. laevis* genomic DNA was obtained from stage 28 *X. laevis* embryos frozen and stored at -80°C following Invitrogen by Thermo Fisher Scientific PureLink™ Genomic DNA mini kit. To visualise the cloned regions 5 µl of the PCR was electroporated on an ethidium bromide 1% agarose gel.

Primers deemed to be efficient were used to clone open regions using high fidelity Phusion or KAPA polymerase. When using Phusion (New England BioLabs) 5 µl 5xbuffer HF, 0.5 µl 10mM dNTPs, 0.75 µl DMSO, 0.5 µl Mg, 0.25 µl Phusion polymerase, 14.5 µl SW H_2O, 1 µl 200ng/µl *X. laevis* genomic DNA and 1.25 µl forward and reverse primer were mixed and the PCR program in appendix 1 was followed. 5 µl of the cloned regions were visualised on an ethidium bromide 1% agarose gel. When using KAPA long range hot start ready mix polymerase 13.75 µl SW H_2O, 7.75 µl KAPA ready mix (Invitrogen by ThermoFisher Scientific),

1.25 µl forward and reverse primer and 1 µl of 200ng/µl *X. laevis* genomic DNA were mixed and the PCR program in appendix 1 was followed. 5 µl cloned fragment was visualised on an ethidium bromide 1% agarose gel. Regions that did not clone with either Phusion or KAPA were attempted cloned with low fidelity TAQ polymerase (meridian bioscience). To troubleshoot poor cloning gradient PCRs were done with a temperature gradient of 50-60 °C using TAQ. This caused changes to the original TAQ PCR program where regions 63 and 217 were PCR with an annealing temperature of 60 °C. For all cloned constructs, the leftover 20 µl PCR product was purified follow the QIAquick® PCR Purification Kit and eluding in 30µl EB buffer.

2.4.2 Ligation into transgenic vector

The purified PCR products were digested for 1 hour at 37°C by mixing 10.5 µl SW H_2O, 2.5 µl Cutsmart (RE buffer), 10 µl PCR product 1 µl Not1 (New England Bio Labs) and 1 µl Sbf1 (New England Bio Labs). The enzymes were deactivated at 80°C for 15 minutes.

The transgenic vector 1960 given to us by Dr Robert Granger seen in appendix 2 was digested for 1 hour at 37°C by mixing 40 µl SW H_2O, 5 µl Cutsmart (RE buffer), 3 µl of ~500ng/ µl transgenic vector, 1 µl Not1 (New England Bio Labs) and 1 µl Sbf1 (New England BioLabs). The reaction was terminated at 80°C for 15 minutes. All the digestion was mixed with 10 µl loading dye and visualised on an ethidium bromide 1% agarose gel. The 6,300 bp backbone was cut out of the gel and purified following the QIAEX®II Gel extraction kit.

The purified digested transgenic vector backbone and the digested cloned DNA was ligated together by mixing 1 µl ligation buffer, 1 µl T4 ligase, transgenic vector backbone and cloned region in a 1:6 ratio, SW H_2O to make up a 10 µl reaction. The reaction was left at 4°C overnight.

2.4.3 Making a negative control for transgenesis

The regulatory region from plasmids 1960 and 1965 given to us by Dr Robert Granger were removed to make a negative control for the transgenic vector. This was done by a PCR reaction that amplified two sides of the plasmids. Primers for both plasmids can be seen in Table 1Table 2. Plasmids were amplified using Phusion HF (New England BioLabs) by mixing 5 µl 5x buffer, 0.5 µl 10mM dNTPs, 0.75 µl DMSO, 0.5 µl Mg, 1 µl of about 75ng/µl plasmid, 0.25 µl Phusion, 14-5 µl SW H_2O, 1.25 µl of each primer, and following the PCR program in appendix 1. 5 µl of each two PCR reactions for each plasmid were mixed with 0.5 µl DpnI to

removed excess circular plasmid by incubating for 2 h at 37°C. 2 µl of this reaction was then transformed as described in subchapter 2.4.4.

Table 1: Primers for amplification of both sides of plasmid 1960 to remove the regulatory region.

(1)	Amp_Fw	5'- GGAGCTGAATGAAGCCATACCAAACGACGAGCGTGA-3'
(1)	1960_Rv2	5'- ggccgcCACCGCGGTGGATC-3'
(2)	1960_Fw2	5'- tgcaGGTCGACCATAGTGA-3'
(2)	Amp_Rv	5'- TCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCC-3'

Table 2: Primers for amplification of both sides of plasmid 1965 to remove the regulatory region.

(1)	Amp_Fw	5'- GGAGCTGAATGAAGCCATACCAAACGACGAGCGTGA-3'
(1)	1965_Rv	5'- ggccgcCACCGCGGTGGAT-3
(2)	1962_Fw	5'- CGGGCTGCactgcaGGTCGA-3'
(2)	Amp_Rv	5'- TCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCC-3'

2.4.4 Transformation of the ligation product

To obtain enough sample for transgenics and to sequence the ligation to ensure the ligation had worked, the ligated plasmid was transformed into competent *Escherichia coli* bacteria using 50 µl competent bacteria and 5 µl ligated plasmid. The solution was left on ice for 30 minutes, then heat shock at 42°C for 90 seconds and left on ice for 2 minutes. 1 ml of LB was added to the competent bacteria which were then left in a 37°C heat shaker at 450 rpm for 1 hour 30 minutes. The bacteria were pelleted by centrifuging at 7,000 rpm for 5 minutes, then 800 µl of the supernatant was removed and the pelleted bacteria was resuspended in the remaining 200 µl supernatant and plated on a plate of LB agar with 1% carbenicillin. The bacteria grew overnight at 37°C.

To identify successfully transformed competent cells a colony PCR was performed by picking 5 colonies with a 10 µl pipet tip and mixing them vigorously with 30 µl SW H_2O. The colony was PCR amplified by mixing 10.5 µl biomix TAQ polymerase (meridian bioscience), 1 µl 1960_Fw primer (5'- GCCAACTCTGACTCTAGAACT-3'), 1 µl GATA_Rv primer (5'- CGAACCACTTTGTACAAGAAAGC3'), 9.5 µl SW H_2O, 3 µl colony water, and the PCR program in appendix 1 was followed. To visualise the PCR amplified regions 10 µl of the PCR was run on an ethidium bromide 2% agarose gel. One of the 5 colonies with the correct insert size was grown in 5 ml LB with 5 µl carbenicillin and 10 µl of the colony water at 37°C at 180 rpm overnight. The next day 4 ml of the colony was spun down and purified using the QIAprep® Spin Miniprep Kit (250). To confirm the insert, 15 µl of 50ng/µl purified plasmid was

sequenced using 2 µl of the 1960_Fw primer at Eurofins genomics. Transgenic vectors with confirmed inserts were grown in 100 ml LB with 100 µl carbenicillin and 100 µl colony overnight at 37°C at 180 rpm. The transgenic vector was then purified by spinning down 50 ml of the colonies and follow the Invitrogen PureLink™ Fast Low Endotoxin Midi Plasmid Purification Kit by Thermo Fisher Scientific. Alternatively, 3 ml of the grown colonies were spun down and the protocol for the PureYield™ Plasmid Miniprep system by Promega was followed, or 200 ml of colonies were grown and the EndoFree® Plamid Maxi Kit by Qiagen was followed.

2.5 I-Scel meganuclease transgenesis

The protocol for the transgenic reaction was described in Ogino (2006) (Ogino et al., 2006a). The transgenic vector was digested with 1 µl I-Scel (New England Bio Labs), 1 µl Cutsmart (RE buffer), 4 µl SW H_2O, and 4 µl of 0.1µg/µl or 0.2µg/µl transgenic vector at 37°C for 40 minutes. The reaction was then put on ice. One cell-stage *X. laevis* embryos were placed in 3% Ficoll and injected in the animal pole with 10nl of the reaction. A one cell-stage embryo was therefore injected with either 400pg DNA or 800pg DNA. One cell-stage *X. tropicalis* embryos were treated like the *X. laevis*, however were only injected with 4nl of the 0.2µg/µl reaction, meaning they were injected with 320pg of the plasmid. When the embryos stared to cleave, the solution was changed to 0.1 x MMR for *X. laevis* and 0.05 x MMR for *X. tropicalis*. The embryos grew at 18 - 20°C or 20 – 26°C until stage 30 when they were terminated.

2.6 Imaging

Embryos with satisfactory fluorescent expression at gastrula or neurula stages were imaged and allowed to continue to grow, however, tailbud and tadpole embryos with satisfactory fluorescent expression were fixed in MEMFA (40 ml dH_2O, 5 ml MEMF salt, 5 ml 36% formaldehyde) for 2 hours at RT. The MEMFA was washed off by two 5 minute washes with DepC-PBST (600ml Depc-PBS, 600µl Tween20). Fluorescent images were taken on a 2% agarose plate using a Leica MZ 16F fluorescent microscope with a Leica DFC300 FX camera using green excitation, and the Fluorescent Leica Firecam software Version 3.4.1. (Leica microsystems Ltd).

Chapter 3: Identification, bioinformatical analysis and cloning of differential peaks

3.1 Introduction

As discussed in subchapter 1.4 the genome is packed into a highly regulated chromatin state, where some parts of the genome are closed and non-accessible, and other regions are open and accessible. CREs need to be accessible by TFs and other regulatory proteins to cause regulation of the gene it is associated with. Therefore, to predict putative CREs, regions that are open in the genome need to be identified. The method used in this project is called Assay for Transposase Accessible Chromatin using Sequencing (ATAC-Seq), which was developed by Buenrostro and colleagues in 2013 when they used it to sequence open chromatin regions in B-cells. The method takes advantage of the hyperactive Tn5 transposase which has adaptors for high-throughput DNA sequencing. The Tn5 integrate its adaptors in accessible chromatin. The steric hindrance in less accessible chromatin makes integration of Tn5 less probable, and therefore the fragments that will be amplified by a PCR reaction and sequenced by high-throughput sequencing are likely to be from open chromatin regions (Figure 9) (Buenrostro et al., 2013).

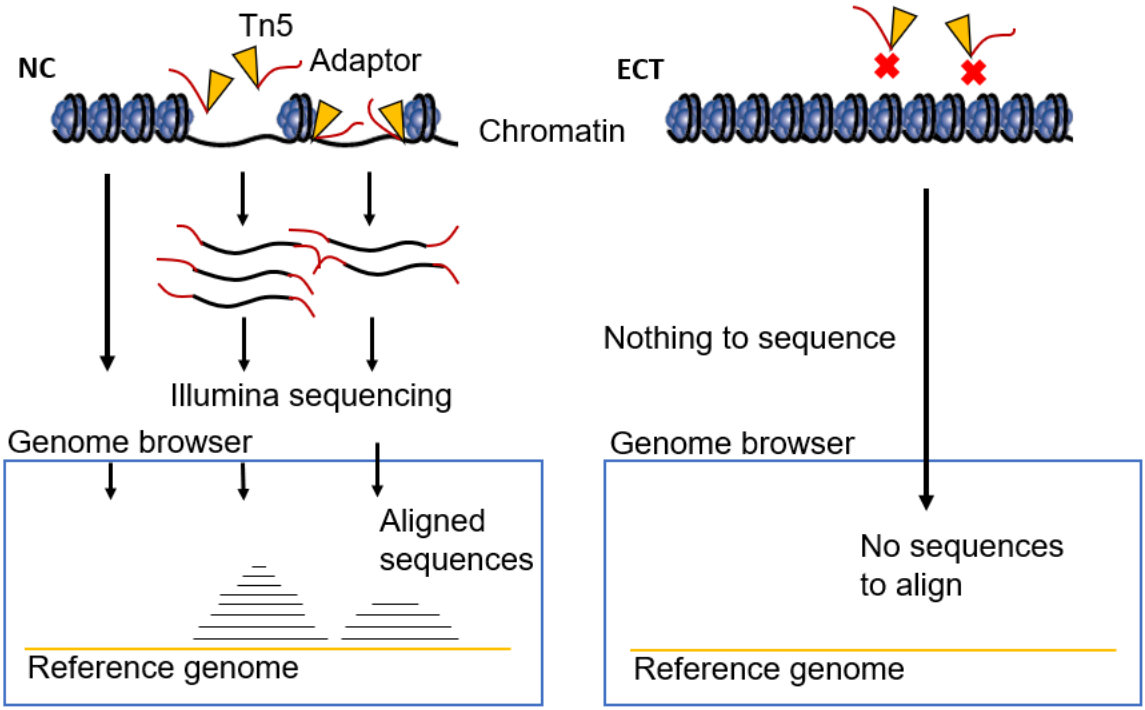


Figure 9: Schematic diagram of ATAC-sequencing showing a neural crest (NC) and an ectodermal (ECT) sample. The NC has open regions in the chromatin which are accessible by the Tn5 transposase which integrates into the open regions, cleave them out of the genome and ligate on adaptors for high-throughput sequencing by illumina. The sequences are aligned to a reference genome in a genome browser. The more sequences the higher the peak. If the chromatin is not open as in the start of the NC and the ECT sample, the Tn5 cannot integrate and cleave the genome. Nothing can be sequenced and there are no sequences to align to the reference genome, meaning there is no peak.

CREs can be identified by other methods not including ATAC-seq. However, this is a difficult task, as CREs such as enhancers lack a strong genetic-defining feature (Betancur et al., 2010, Sánchez-Gaya et al., 2020). Most of the current knowledge of the NC GRN comes largely from transperturbation experiments using morpholinos to knock down mRNA activity, although this does not show much about the regulatory sequences that the GRN uses to regulate genes (Betancur et al., 2010). To study the regulatory sequences, researchers have used methods such as random deletions, which was performed by among others Natoli in 1997 among others, when he identified a 1.6 kb regulatory sequences upstream of *pax3* in mouse (Degenhardt et al., 2010b, Li et al., 1999, Natoli et al., 1997). Further, by looking at conservation in 2010 Betancur identified two CREs for *sox10* in chicken (Betancur et al., 2010). As well as using 'indirect methods', or using the genomic sequence to identify CREs, 'direct methods' by using epigenetic profiling can also be used to identify CREs (Nelson and Wardle, 2013). By using for example ChIP-seq (chromatin immunoprecipitation and sequencing) one can identify epigenetic markers such as H3K4mel and H3K27ac which are commonly associated with active enhancers or identify co-activators such as p300 (Heintzman

et al., 2009, Karnuta and Scacheri, 2018, Sánchez-Gaya et al., 2020, Krijger and de Laat, 2016). Other attempts to use various sequencing techniques to study open chromatin required millions of cells to start, and these methods are complex and time consuming and they cannot probe the interplay of nucleosome positioning, chromatin accessibility and TF binding simultaneously. They are also multistep processes, with less accuracy and more opportunities for mistakes. Whereas ATAC-seq is a one protocol process and therefore there are fewer possibilities for error and it requires less starting material than previous methods that sequence open chromatin. Therefore, ATAC-seq is now beginning to be the preferred method to find regulatory regions in the genome (Buenrostro et al., 2013, Simon et al., 2012, Song and Crawford, 2010).

Even though ATAC-seq as a method has been available since 2013, an adaptation for the method in *Xenopus* was not published before 2019 (Bright and Veenstra, 2019, Buenrostro et al., 2013), and therefore, there are not many studies that have used ATAC-seq in *Xenopus* yet. In 2020, Esmaeili and colleagues used Buenrosto's ATAC-seq method to unravel the reason for loss of competency to early inductive signals such as Wnt during dorsal specification in *Xenopus* embryos. They hypothesised that it was due to a change in chromatin architecture that made the chromatin inaccessible. They looked at different stages of tissue competency for dorsal induction, but also MES and NC induction, by making animal cap ectodermal, mesodermal and NC explants. The ATAC-seq data suggested that ventral cells lose the ability to activate dorsal genes as there is a loss of chromatin accessibility at the promoter of dorsal genes. However, these cells retained a competency to activate genes for A-P patterning by Wnt (Esmaeili et al., 2020). Further, they also looked at the chromatin states role in Wnt ability to activate NC induction, and showed that the chromatin surrounding Wnt target genes such as *snai1/2* and *twist* remains open at early gastrula, but also during early neurula, showing that another mechanism must be responsible for the loss of response (Esmaeili et al., 2020). This work showed that ATAC-seq is an efficient way to analyse open chromatin in *Xenopus*.

Although there are not many studies yet using ATAC-seq for *Xenopus*, ATAC-seq has been used in other organisms. Williams and colleagues (2019) wanted to identify genomic regulators controlling the NC program and constructed a high-resolution map of open chromatin regions using ATAC-seq in chick NC tissue as well as non-NC tissue. They found several differentially accessible elements in the NC sample around the NC specifier *snai2* within a 700 kb TAD and identified five active enhancers for *snai2* (Williams et al., 2019). Further, nine enhancers for s*ox10* were identified in zebrafish by using ATAC-seq on zebrafish melanomas. One of the enhancers could drive reporter expression in a subset of NC cells

(Cunningham et al., 2021). Lastly, RNA-seq and ATAC-seq was used in chick to investigate the transcriptome and the chromatin accessibility in the paraxial mesoderm. Here they identified two enhancers upstream of TCF15, and HINT-ATAC revealed several interesting TF binding sites (Mok et al., 2021). These papers show that even though ATAC-seq has not been widely used in *Xenopus* yet, it is being used frequently in many other model organisms.

In this project a PhD student in the Wheeler lab, Dr Marta Marín-Barba, performed ATAC-seq on *X. laevis* animal cap tissue induced to become either NC, NE or ECT at stage 13 and 18. She injected a two-cell stage embryo with Pax3 and Zic1 to induce a NC fate, just Pax3 to cause a neural fate, and no injection caused an ECT fate. She further performed RNA-seq on the same samples to see what TFs and genes were expressed and did ATAC-seq to identify open regions in the chromatin, and then she compared their expression. By doing this she identified open chromatin regions around genes relevant to NC development which could be CREs (Marín-Barba and Wheeler, unpublished results).

Here we use these data to predict CREs. To then visualise whether a predicted CRE is active during development, a reporter construct with a green fluorescent protein (GFP) that can be transcribed if the CRE can cause activation of gene expression is used. The vector used here was a gift from the Dr Robert Granger at the University of Virginia. It contained a ligation site for the CRE using Not1 and Sbf1 restriction enzyme sites, a *GATA2* minimal promoter and the GFP gene (Figure 10) (Ogino et al., 2006b). In this project 19 of the 20 CREs were cloned and ligated into the reporter construct.

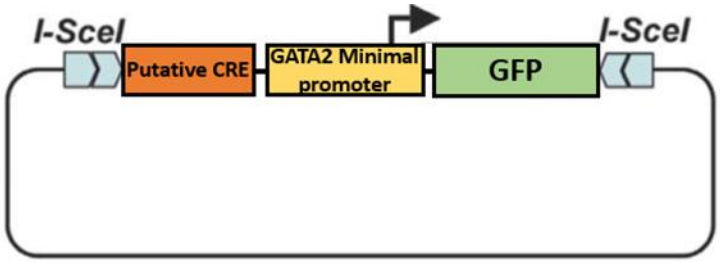


Figure 10: Simplified image of reporter construct used in transgenesis, adapted from Ogino et al., 2006b. The construct is injected into one-cell stage embryos with the I-SceI meganuclease which integrates the construct into the genome.

In this chapter we will take this large data set and aim to:

- Analyse the differential analysis of ATAC-seq data.
- Identify open regions exclusive to the NC sample in the UCSC genome browser.
- Predict transcription factor binding sites in the regions chosen to be studied.

- Clone the identified CREs into a transgenic construct.

It is important to note that much of the bioinformatic analysis of the ATAC-seq data set was carried out by a bioinformatician in the Wheeler lab, Dr Claudia Buhigas, with myself observing and providing assistance.

3.2 Results

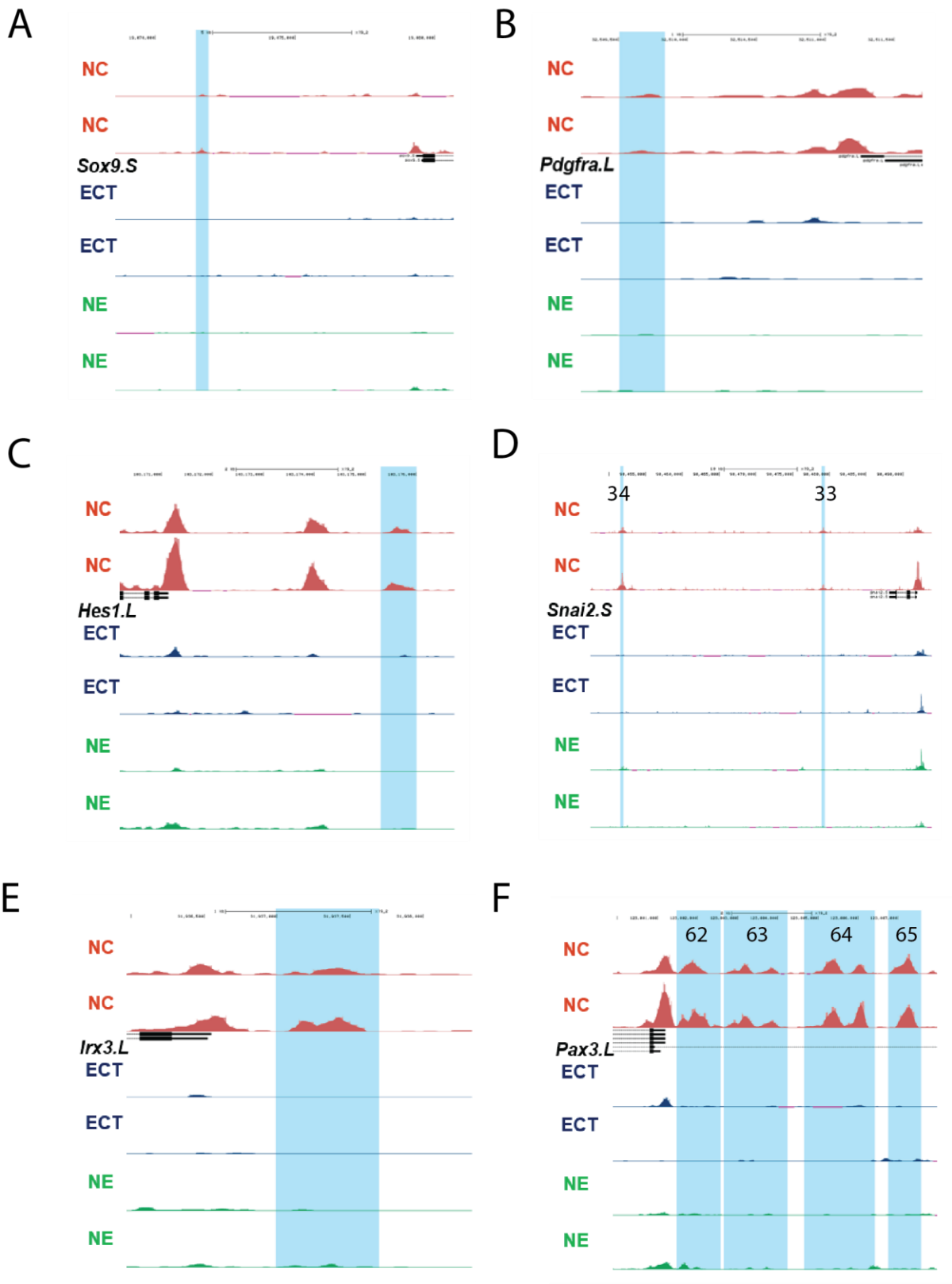
3.2.1 Identifying CREs using the UCSC genome browser

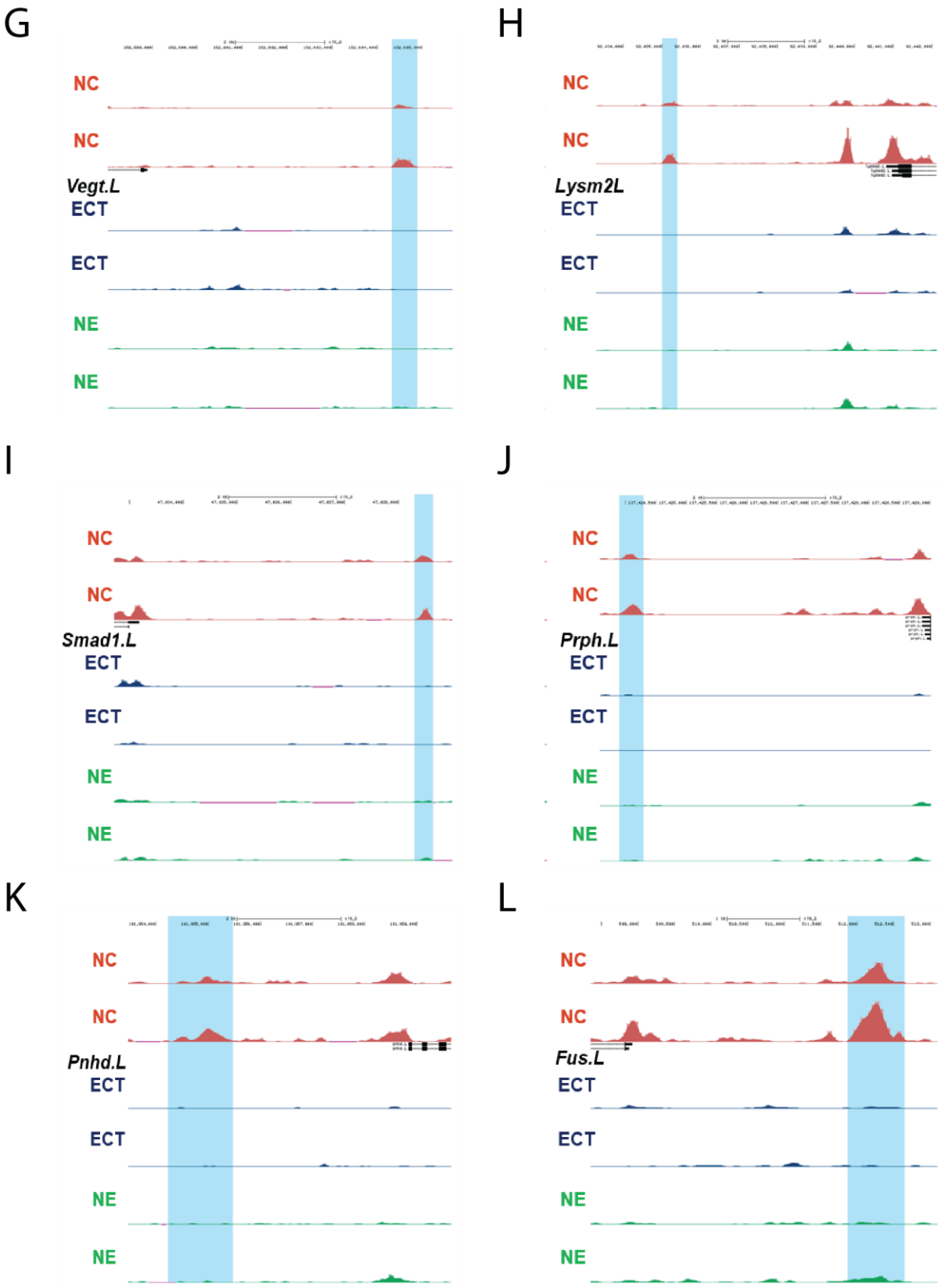
A differential analysis between three sets of different ATAC-seq data for each induced NC, NE and ECT tissue taken from animal caps of *X. laevis* at stage 13 was carried out by Dr Buhigas. 1068 regions within a 6,000 bp region around the genes of interest were identified as significantly different in the three NC sample compared to the three ECT and NE samples. From this further analysis identified 665 open regions which were deemed to be so close to the gene of interest that they most likely are promoters. The remaining 403 open regions were evaluated in the UCSC genome browser and 20 open regions, now referred as CREs, that were exclusively open in the NC sample were chosen for further evaluation (Table 3). The CREs seen in the genome browser can be seen in Figure 11, and the coordinates for the regions can be seen in appendix 3. The CREs were also analysed by looking at the conservation of the sequences by comparing them to the *X. tropicalis*, chicken, mouse, and human genomes. All the 20 CREs had high conservation with *X. tropicalis*, but no conservation with any other species. Lastly, RNA-seq from the same stage 13 NC induced animal cap was also analysed. Of the 20 CREs associated genes identified for further investigation *sox9, hes1, VegT, Lymed2, evx1, and Myf5* were not found in the data.

Table 3: CREs identified in the differential analysis that were studied further in the transgenic analysis.

CRE	Figure 11	Associated gene	Size of open region	Role in NC	Distance from TSS
22	A	*sox9.S*	455bp	NC migration	8 kb upstream
26	B	*pdgfra.L*	331bp	NC migration	1.5 kb upstream
31	C	*hes1.L*	731bp	NC induction at NPB	4.2 kb upstream
33	D	*snai2.S*	465bp	NC specification and migration	13 kb downstream
34	D	*snai2.S*	465bp		40 kb downstream

52	E	*irx3.L*	715bp	Cranial NC development	500 bp upstream
62	F	*pax3.L*	1115bp	NC induction	500 bp upstream
63	F	*pax3.L*	1595bp		1.7 kb upstream
64	F	*pax3.L*	1781bp		3.6 kb upstream
65	F	*pax3.L*	825bp		5.7 kb upstream
101	G	*vegt.L*	569bp	Repressed in NC. Needed for mesoderm and ectoderm formation	5.5 kb upstream
105	H	*lysmd2.L*	398bp	Central nervous system	5.8 kb upstream
116	I	*smad1.L*	350bp	NC induction	5.3 kb upstream
138	J	*prph.L*	404bp	NC induction to peripheral sensory neurons	5 kb upstream
160	K	*pnhd.L*	1242bp	Cranial development	3.3 kb upstream
180	L	*fus.L*	793bp	Neuronal development	3 kb upstream
182	M	*wnt8a.L*	615bp	NC induction	3.2 upstream
183	M	*wnt8a.L*	681bp		4 kb upstream
192	N	*evx1.L*	821bp	Patterning of embryo during development	5.7 downstream
217	O	*myf5.L*	771bp	Early cranial NC development	4 kb upstream


A
NC
NC
Sox9.S
ECT
ECT
NE
NE
B
NC
NC
Pdgfra.L
ECT
ECT
NE
NE
C
NC
NC
Hes1.L
ECT
ECT
NE
NE
D
34
33
NC
NC
Snai2.S
ECT
ECT
NE
NE
E
NC
NC
Irx3.L
ECT
ECT
NE
NE
F
62
63
64
65
NC
NC
Pax3.L
ECT
ECT
NE
NE


G
NC
NC
Vegt.L
ECT
ECT
NE
NE
H
NC
NC
Lysm2L
ECT
ECT
NE
NE
I
NC
NC
Smad1.L
ECT
ECT
NE
NE
J
NC
NC
Prph.L
ECT
ECT
NE
NE
K
NC
NC
Pnhd.L
ECT
ECT
NE
NE
L
NC
NC
Fus.L
ECT
ECT
NE
NE

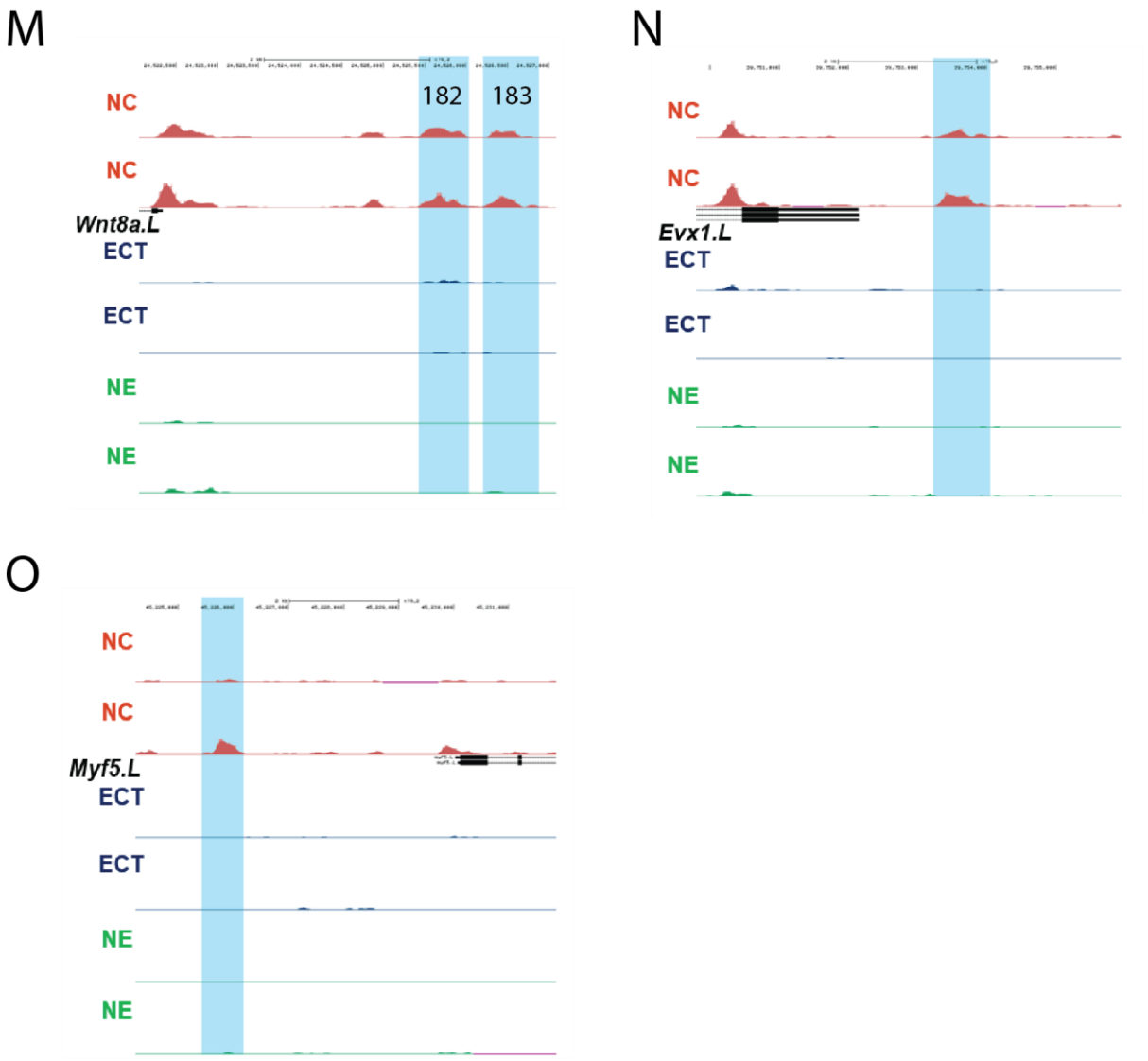


Figure 11: Cloned peaks identified in ATAC-seq tracks viewed in the UCSC Genome Browser. CREs for cloning are highlighted in blue. NC: Neural crest, ECT: Ectoderm, NE: Neuroectoderm. A: CRE 22 associated with sox9. B: CRE 26 associated with pdgfra. C: CRE 31 associated with hes1. D: CREs 33 and 34 associated with snai2. E: CRE 52 associated with irx3. F: CREs 62, 63, 64, 65 associated with pax3. G: CRE 101 associated with vegT. H: CRE 105 associated with lysme2. I: CRE 116 associated with smad1. J: CRE 138 associated with prph. K: CRE 160 associated with pnhd. L: CRE 180 associated with fus. M: CREs 182 and 183 associated with wnt8a. N: CRE 192 associated with evx1. O: CRE 217 associated with mtf5.

3.2.2 Predicting transcription factor binding sites in CREs

TFBSs within the CRE sequences were predicted by using the UCSC genome browser which used Xenbase.org to make its predictions (Karimi et al., 2018), and HINT-ATAC which used JASPAR for its predictions. A list of TFBSs predicted by the UCSC genome browser was compared to a list of NC related TFs from *Transcriptional control of Neural Crest Development* by Nelms and Labosky, (Nelms and Labosky, 2010), and the NC related TFs were noted in Table 4. Further, sequences potentially bound by TFs at the time of Tn5 digestion at stage 13

were identified in a HINT-ATAC analysis by a Dr Buhigas and compared to the same list of NC TFs from the previous analysis. The results can also be seen in Table 1Table 4 .

Table 4: Transcription factor binding sites in the CRE sequences studied. In bold are TFs that are in common for both categories.

CRE	Related gene	UCSC predictions	HINT-ATAC
22	*sox9.S*	gsc, pax6, runx1, unx2	pitx2, pou2f1, pou5f1, zic1, zic4, zic5
26	*pdgfra.L*	none in genome browser	no hint-atac
31	*hes1.L*	dlx2, dlx3, hoxa2, hoxa3, hoxb2, lbx1, pou3f1, sp1	sp3, sp4,
33	*snai2.S*	**runx1**,	dlx3, dlx5, dlx6, lhx1, msx1, msx2, **runx1**, tcf7
34	*snai2.S*	foxd3, foxe1, lef1, tcf12, tcf4, zic3, zic4,	gata2, gata6, pax3, pax7, phox2a, tcf7, zic1, zic2 (zic1::zic2),
52	*Irx.L*	hand1, hand2, hoxa2, hoxb2, hoxd3, lef1, **sox2, sox4**, sox9, **tcf7**,	**sox2, sox4**, sox10, **tcf7**,
62	*pax3.L*	myc, mycn, pax6, rxra, smad4, sp1, zic3, zic4	mef2a, mef2c, mef2d, myb, pbx1,
63	*pax3.L*	hmga2, mnt, myc, mycn, usf1, usf2, zic2, zic4, zic5	zic1, zic3,
64	*pax3.L*	alx1, gata4, pbx1, tcf12, tcf4,	ascl1, hoxb9, pax3, pax9, stat3, tcf12
65	*pax3.L*	none in genome browser	gata4, prdm1, stat3
101	*vegt.L*	hoxd4, **pou3f1,** smad3,	foxc1, foxc2, foxd2, foxf2, foxj3, mafb, pou2f1, **pou3f1**, pou3f2
105	*lysmd2.L*	phox2a, phox2b, zic3, zic4,	barx1, dlx2, dlx5, hoxa7, hoxb2, hoxb6, hoxb7, hoxd3, lhx1, msx1, msx2, tlx2
116	*smad1.L*	**gsc,** hoxa3, hoxb4, mnt, **pitx2,** smad3, **zic3, zic4,**	foxc1, **gsc**, gsc2, hand2, mef2a, **pitx2**, smad4, zic1, zic2, **zic3, zic4,** zic5

138	*prph.L*	runx1, **sox2,** tlx1, tlx3,	nr5a1, nr5a2, pax9, phox2b, prdm1, **sox2**, sox4, sox10
160	*pnhd.L*	meis1, pknox1,	foxo1, lef1, mef2a, mef2c, mef2d, mybl2, nr2f2, phox2a, pou2f1, pou3f1, pou3f2, pou4f1, prdm1, runx1, sox4, stat3, tcf12, tcf7, zic2
180	*fus.L*	hoxa1, **lbx1**, **pou3f2,** runx1, runx2, sox2, sox4, sox5, **sox8, sox9,** zic1,	alx1, alx3, alx4, dlx6, emx1, emx2, hoxb3, **lbx1**, lbx2, pou2f1, pou3f1, **pou3f2**, pou4f1, prrx1, prrx2, **sox8, sox9**, tead2, tlx2
182	*wnt8a.L*	**dlx6, lhx1**, **msx1, msx2,**	dlx3, dlx5, **dlx6**, foxf1, lbx2, **lhx1**, **msx1, msx2,** pou2f1, pou3f1, pou3f2, sox10, sox4, tead2
183	*wnt8a.L*	gata2, mecom, sox9	gata3, gata6, mafb, phox2b, pou3f1, pou3f2
192	*evx1.L*	alx1, emx2, emx2, hoxa1, hoxb2, lhx1, msx1, msx2, pax3, pbx1, runx2, sox2, sox4	mybl2, phox2a, pou2f1, pou4f1, pou5f1,
217	*myf5.L*	foxc2, foxd1, lef1, mef2a, mef2c, sp1, tcf4, tcf7	pou4f1, tcf12,

3.2.3 Cloning of open regions containing CREs

Primers for the 20 CREs were generated (appendix 3), and the regions were PCR amplified out of the *X. laevis* genome (Figure 12). Region 65 was challenging to PCR and clone, as it only amplified with low fidelity polymerase TAQ, and the correct size PCR product would not ligate into the plasmid. This is believed to be due to the forward primer having several targets in the *X. laevis* genome, as seen when BLAST searches were carried out, even though the reverse primer is precise. Region 65 was therefore not proceeded with. The PCR products were ligated into the reporter vector as described in the materials and methods chapter, in the position showed in Figure 10, and the plasmids were amplified by transforming them into competent *E. coli* cells. The plasmids were sequenced by Eurofins genomics which confirmed that the region had been correctly cloned (appendix 4), and that they were ready for injection.

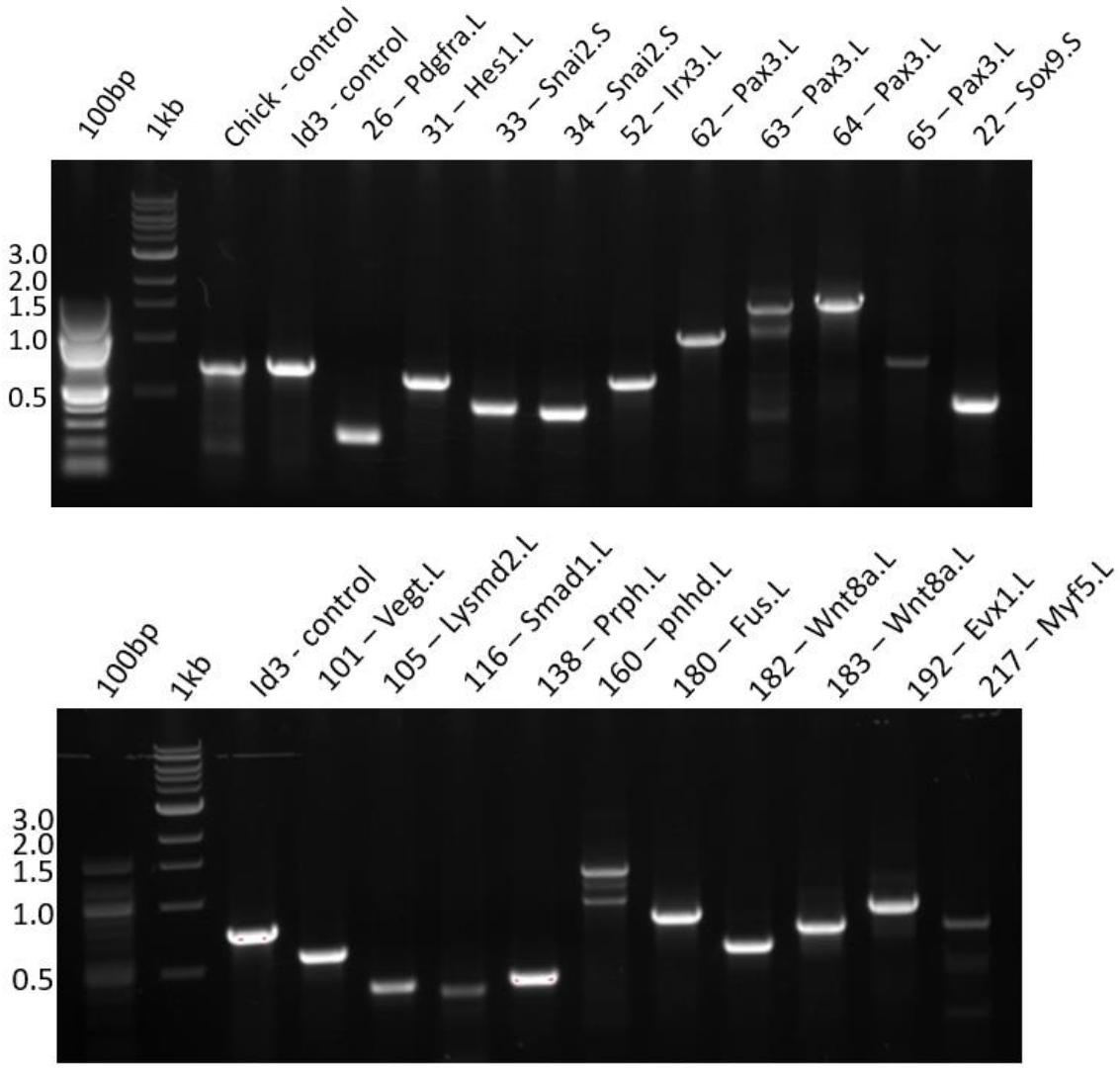


Figure 12: Low fidelity TAQ PCR to test primers designed to clone out open chromatin regions. It shows that all primers amplify the correct size associated with their open region.

3.3 Discussion

The differential analysis of the ATAC-seq data from stage 13 induced animal cap NC, NE and ECT samples identified 403 regions that were significantly more open in the NC compared to the NE and ECT. Of the 403 open regions 20 regions predicted to be putative CREs and referred to as CREs in this project, were found to be exclusive to the NC sample, meaning that no peaks were visible in the NE and ECT samples as can be seen in Figure 11. It was postulated that this exclusiveness would mean that the CREs would be more likely to be NC specific. This was therefore the reason for this being the first criteria for choosing the peaks. The corresponding RNA-seq for the stage 13 induced animal cap NC sample was also analysed, and it was identified that *sox9, hes1, vegT, lymed2, evx1, and myf5* were not present

in the RNA-seq. For *vegT* this was expected as VegT is one of the key markers for endoderm formation in the blastula, where it activates Nodal signaling (Sudou et al., 2016). VegT has no associated function in NC development, and it has been seen that inhibiting the expression of the zygotic splice variant of *vegT* (Antipodean) has no effect on NC induction when it is critical for paraxial mesoderm patterning (Klymkowsky et al., 2010). The CRE close to *vegT* could however be another type of CRE, even though this project is specifically looking for enhancers, this region of the genome could be open to be a silencer, or an insulator as discussed in subchapter 1.5.2. Further, *vegT*, is the gene that is the closest to the CRE, however, CREs can be found several thousand bp away from genes they regulate, so it might not regulate *vegT*. For *sox9* the results of the RNA-seq were strange. Sox9, as discussed in subchapter 1.2 is well known to be involved in the induction of the NPB and in the NC induction and should be present in the NC sample at stage 13. The lack of Sox9 in the sample may suggest that the RNA-seq data is not as reliable as we would hope.

When investigating the CRE sequences in further detail to look for putative TFBSs, several NC related TF were identified (Table 4). These include many of the TF discussed in subchapter 1.2 as being involved in NPB and NC induction. Many of the sequences contained Pax3 and Pax7 binding sites, as well as Sox8, Sox9 and Sox10 binding sites. Several of the Zic proteins binding sites have also been predicted, as well as Lef1 together with different Tcf sites associated with Wnt signaling (Santiago et al., 2017). The specifics of the TF binding sites and how they relate to the gene associated with the CRE in this study will be further investigated in subchapter 4.3.

However, there is one observation when looking at the predicated TFBSs worth noting at this stage and that is the low overlap between TFBSs predicted by the genome browser and the HINT-ATAC. The HINT-ATAC might not pick up on all the predicted sites in the browser as not all TFs may be bound to the CRE at the time of Tn5 treatment. However, by this logic one would assume that the TFBSs that were in the HINT-ATAC would be in the genome browser, but this is not the case. In some cases, both analyses suggest the same type of TF such as GATA2 or GATA3, but not both. These TFs have very similar binding sites so it might be that the same GATA can bind. Further, we could look at the data and say that the TFBSs predicted by the HINT-ATAC is more reliable, as it is experimental data. However, at this stage of the investigation, any TFBSs may be worth looking into if they are interesting enough.

Further, experimental data showns that the method devised in this project to clone the CREs is a good method. The gel electrophoresis image (Figure 12) showed that all designed primers amplified sequences in the *X. laevis* genome of the correct size. 17 of the regions were then

successfully ligated into the reporter vector as shown by sequencing (appendix 4). CREs 63, 65 and 217 did not PCR amplify correctly with the high-fidelity KAPA or Phusion, as shown by the sequencing not matching the predicted sequence. However, 63 and 217 did amplify with TAQ and were ligated into the reporter plasmid correctly (appendix 4). CRE 65 did amplify but the correct sequence did not ligate into the reporter vector. This was possibly due to the forwards primer not being specific enough. This highlights that when making the primers for the CREs to be PCR amplified out of the *X. laevis* genome, checking their specificity in BLAST towards the *X. laevis* genome will increase the chances of this protocol. Even through these challenges these results show that the method designed to construct the reporter constructs was an efficient method.

Chapter 4: Transgenic expression in developing *X. laevis* and *X. tropicalis* embryos

4.1 Introduction

We hypothesise that the open regions identified in the differential analysis in Chapter 3: are *cis*-regulatory elements (CREs) such as enhancers. Therefore, to analyse the activity of the CREs they were microinjected into *Xenopus* embryos using a transgenic method. If the CREs are regulatory they will cause activation of the minimal promoter for *GATA2* in the reporter construct which will then cause expression of GFP in the cells that the CRE is active. The reason to use transgenics was to avoid or minimise mosaic expression of the CREs.

The transgenic method followed was using the I-SceI meganuclease, a method published by the Granger lab in 2006 for *Xenopus* (Ogino et al., 2006a). The technique was first developed for the fish Medaka (*Oryzias latipes*) (Thermes et al., 2002), then adapted for *X. tropicalis* by Ogino et al., in 2006, however it also functions in *X. laevis*. The I-SceI meganuclease is a mitochondrial intron-encoded homing endonuclease isolated from the yeast *Saccharomyces cerevisiae* (Jacquier and Dujon, 1985). The plasmid construct used in this transgenesis has the restriction sites for I-SceI flanking the CRE, the minimal promoter, and the reporter gene (Figure 10). This construct is co-injected into a one-cell stage embryo together with the meganuclease (Figure 13). Ogino and colleagues concluded that they saw about 30% of the injected embryos express the transgene in a promoter-dependent manner in *X. tropicalis* and about 20% in *X. laevis*. The transgenic embryos express the transgene on both the left and the right side of the embryo or on one of the sides and the expression is not mosaic. The process of how the construct flanked by the I-SceI restriction sites is integrated into the genome is unknown as neither the Medaka or the *Xenopus* genome contain the 18 base pair recognition site for I-SceI, however, the technique's simplicity and accuracy has made it a valuable tool (Ogino et al., 2006a, Ogino et al., 2006b, Ishibashi et al., 2012, Yergeau et al., 2010).

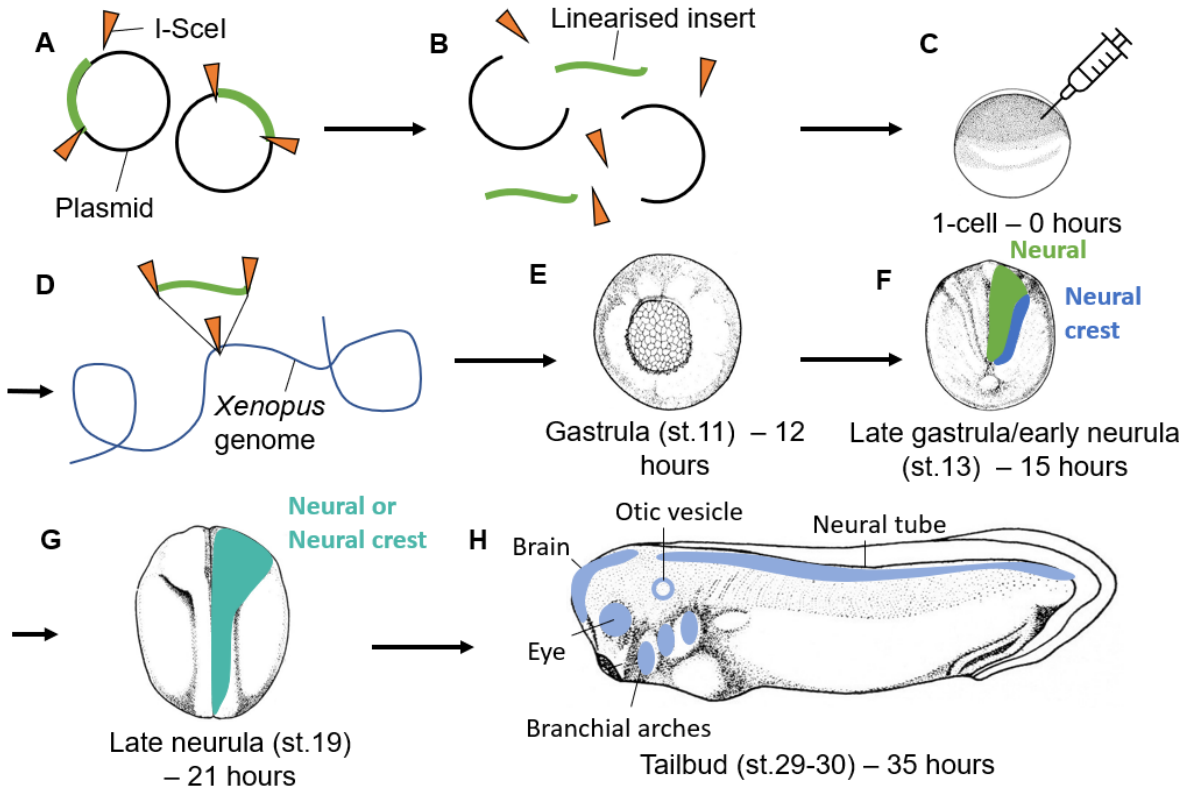


Figure 13: Schematic diagram of transgenic process. A: On the day of injections, the reporter plasmid is digested together with the I-SceI meganuclease. B: The I-SceI meganuclease linearises the CRE, minimal promoter and reporter gene (GFP). C: The digestion reaction is microinjected into a one-cell stage embryo in the animal cap. D: The linearized construct is randomly integrated into the Xenopus genome by a yet unknown process. E: Embryos are observed from late gastrula stage throughout neurula for reporter gene expression. F: Expression can be seen either as neural or as neural crest. G: At late neurula stages it is no longer possible to distinguish neural from neural crest expression along the neural fold. H: Expression can be seen in some neural crest and placodal related structures in the tailbud embryos, such as the brain, the eye, the otic vesicle, the neural tube, and the branchial arches. Pictures adapted from (Nieuwkoop and Faber, 1994).

The I-SceI meganuclease transgenesis method is a technically simpler alternative to the restriction endonuclease mediated integration (REMI). REMI involves isolation of sperm nuclei and egg extracts which are then combined with linearized plasmid DNA to introduce a transgene and a restriction enzyme to generate chromosomal breaks to promote recombination of the transgene into the genome. This is then transplanted into an unfertilized egg (Amaya and Kroll, 2010, Amaya and Kroll, 1999, Wheeler et al., 2000). The integration of the plasmid occurs the earliest out of the transgenic methods, and therefore the chance of mosaic animals is smaller than with the linearised plasmids (Yergeau et al., 2010). Disadvantages with this method is that it uses a bigger needle that the I-SceI transgenesis method which can be stressful for the embryo. Further, there is a lower percentage of success with REMI for tadpoles to develop normally (20-40% cleave, and only 5-40% then develop

after). REMI also require a good level of skill in sperm nuclear transplantation (Ogino et al., 2006b) and GN Wheeler personal communication).

A third method of transgenesis that is preferred in the model zebrafish (*Danio rerio*) is the Tol2 system. Tol2 is a transposase from the medaka fish, and it does not have any target site preference. It is believed to show the highest rate of genomic integration into the germline and is therefore the preferred system to make a transgenic line (Suster et al., 2009). In zebrafish it has a 50% success rate where the I-SceI meganuclease only has a 30% rate of integration when microinjected into a one-cell stage embryo. The system is used for expression of reporter genes under the control of promoters, but also for gene and enhancer trapping, and it has been used to rescue mutant phenotypes (Kawakami, 2005). It has also been shown that if a plasmid with a Tol2 protein is microinjected into a one-cells stage *Xenopus* embryo together with Tol2 mRNA, about 30% or more of the embryos can show reporter gene expression. Further there is good integration of the system into the germline. However, there is a potential for the integration of the DNA into the genome to be mosaic, and therefore, this method may not be suitable to study founder animals (F_0) (Hamlet et al., 2006, Yergeau et al., 2010).

In this chapter we will use the I-SceI to generate transgenic embryos with our CRE's. We will:

- Determine the optimal transgenic methodology using I-SceI meganuclease
- Analyse transgenic reporter expression by CREs

4.2 Results

4.2.1 Optimizing the transgenesis of *X. laevis* and *X. tropicalis*

To identify whether the transgenic method would work we first performed injections with several different control constructs. Control plasmid 1965 was given to us by the Grainger lab and contains a characterised enhancer for *foxe3* found in the lens controlling the *GATA2* minimal promoter. It is the same plasmid as 1960 however it only contains the Not1 restriction enzyme site and not the Sbf1 site so it cannot be used to make new transgenic constructs. One challenge arose when embryos injected with 1965 showed little GFP expression. 1.01% (2/198) of the injected embryos show neural expression at neural stages which can be seen in Figure 14C. This expression disappeared at tailbud stages. A further 8.08% (16/198) of the injected embryos showed spotty fluorescence associated with dying cells. A second control plasmid contained a CRE associated with *Id3* identified by Dr Marín-Barba (2017). It showed GFP expression in 4.65% of the embryos with a very low number of 43 surviving embryos and

2 of them showing neural expression. These 2 embryos also showed no expression at tailbud stage. 1 of the 43 embryos (2.33%) showed a different expression pattern that was not neural, and 5 of the 43 embryos (11.63%) showed spotty expression associated with dying cells (Figure 14A-B). Further, a negative control containing the 1960 backbone with no insert was injected into the one-cell staged embryos. Plasmid 1960 from the Granger lab was used to make the CRE plasmids. The negative construct showed no specific fluorescence at any stages in a normally developing embryos, and out of 173 surviving embryos, 15.61 % (27/173) showed spotty expression associated with dying cells. An example of an embryo injected with the 1960 negative control plasmid can be seen in Figure 15 where there is no GFP expression until stage 30 where there is non-specific spotty expression in the abdomen.

To perform the transgenics in the embryos, the protocol written by Ogino (2006) (Ogino et al., 2006a) was followed. However, as explored, when injecting the control plasmid 1965, the *Id3* CRE, as well as CRE 34 identified in this project with the suggested 260pg there was little GFP expression. Therefore, the protocol had to be optimized. One-cell stage embryos were injected in the animal cap with either 400pg or 800pg of the plasmids together with I-SceI meganuclease, and their development was observed. The two concentrations were similar in survival and fluorescence (Figure 16 and Figure 17). One week embryos injected with CRE 1965 at 400pg only gave about 6% more survival than 800pg, and there was only fluorescent expression in embryos injected with 400pg (Figure 16). The same week embryos injected with 800pg of either the *Id3* CRE or CRE 34, although there were no embryos injected with 400pg, there was fluorescent expression observed at 800pg and compared to the wild type (WT) survival the CRE for Id3 had 66% more death and CRE 34 had 30% more death where 400pg of 1965 had 26% more death (Figure 16A). The next week 800pg injections of 1965, the CRE for *Id3* and CRE 34 gave 6%, 20% and 16% more survival than 400pg injected respectfully. Further, 800pg injections gave 13% and 26% more fluorescence for the CRE for Id3 and CRE 34 respectfully. However, for 1965, 400pg injections gave 13% more fluorescence than 800pg (Figure 16). Further, the GFP expression seen with CRE 34 was similar between the two concentrations as seen in Figure 17, where both concentrations have expression around the neural tube, but none of them had very consistent patterns.

Further, during the first two weeks of injections there was strong GFP expression in the abdomen of the embryos at several stages. The same GFP expression was seen with different CRE constructs as seen with construct 34 and 52 in Figure 18. To try and understand this expression, the time of digestion was changed from once a week to 40 minutes before injection as suggested by Ogino (Ogino et al., 2006a). This appeared to remove the abdominal pattern seen in the embryos, however, it was seen to reappear in some embryos on occasions.

Lastly, it had been observed in chick embryos that there was poorer survival when injecting them with plasmids prepared by an endotoxin low midi kit compared to an endotoxin free maxi kit. Therefore, endotoxin low or free mini, midi and maxi kits were all used to purify CRE 22 and injected to identify whether one kit was better than the other. However, this was found to not influence survival compared to the WT when working with *Xenopus*, and all three kits worked at the same rate (no data shown).

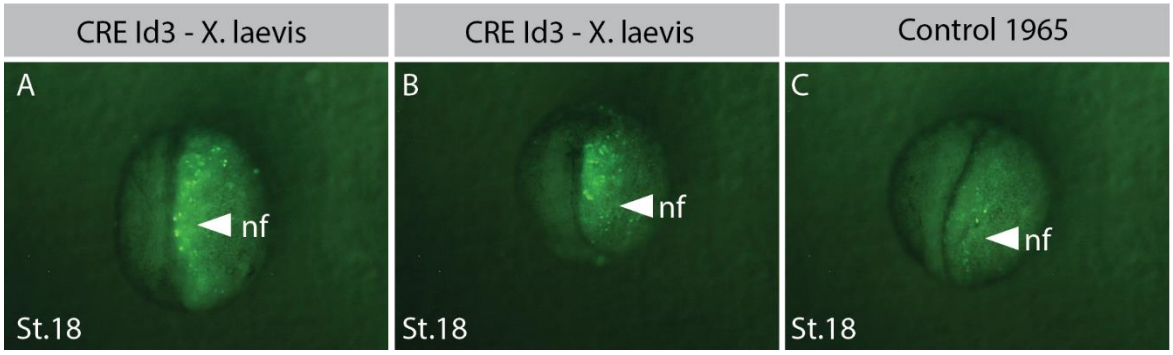


Figure 14: Fluorescence observed with control plasmids. A and B: CRE for Id3 identified by Marta (2017). Fluorescence can be observed along the neural fold (nf). C: Control plasmid 1965 given to us from the Granger lab. Faint fluorescence can be seen along the neural fold (nf). A-C: Dorsal view, anterior top, posterior bottom.

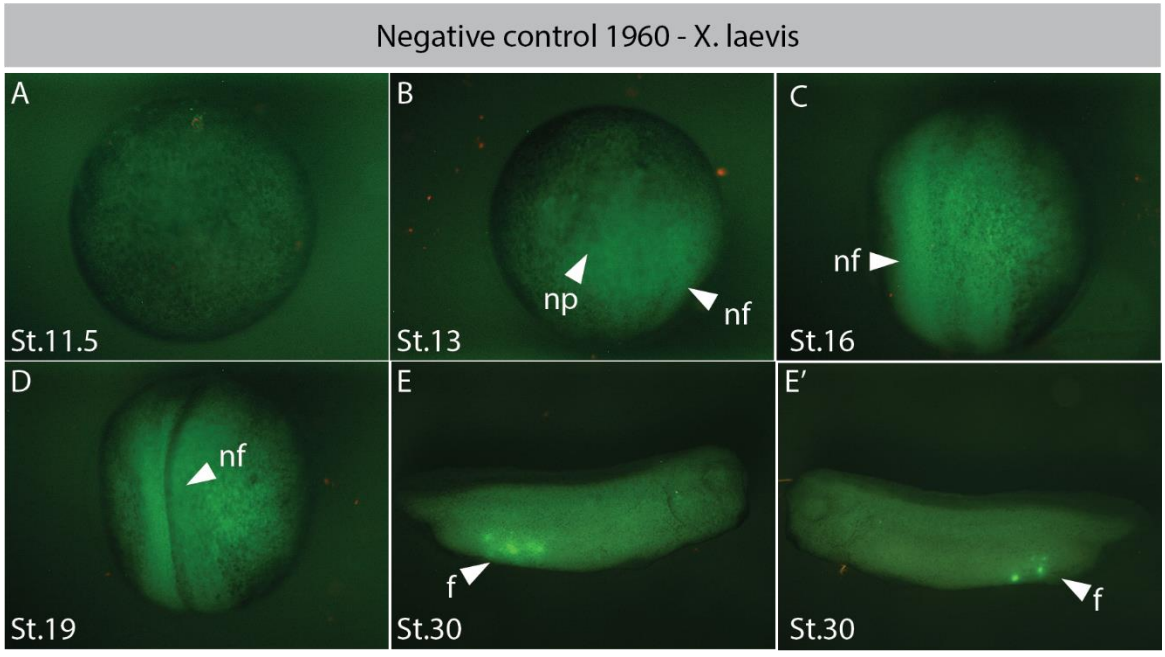


Figure 15: Plasmid 1960 used to create transgenic plasmids, with no insert acting as a negative control show no reporter gene (GFP) expression from late gastrula through to late neurula (A-D). A: Looking down at anterior of late gastrula. B: Early neurula show neural plate (np) and the formation of the neural folds (nf). C: The neural folds move towards the dorsal midline of the embryo. D: The neural folds fuse at the dorsal midline to form the neural

tube. During tailbud stages (E and E') spotty GFP expression is observed in the posterior ventral parts of the embryos (f). A-D: Dorsal view, anterior top, posterior bottom. E: Lateral view, anterior right, posterior left, dorsal top, ventral bottom. E': Lateral view, anterior left, posterior right, dorsal top, ventral bottom.

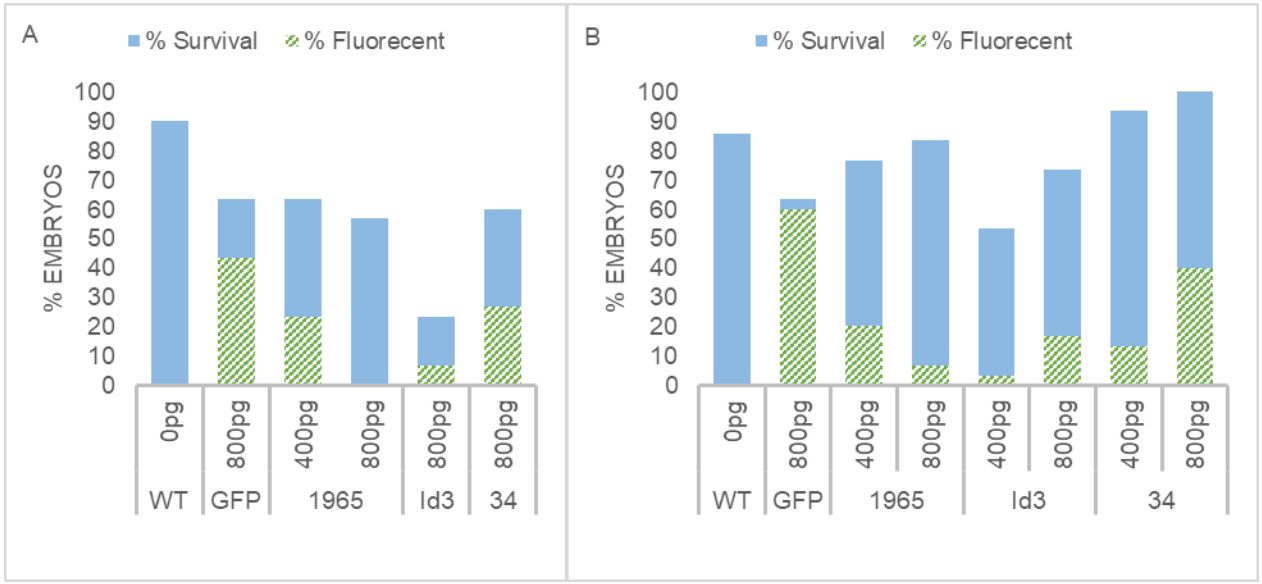


Figure 16: Comparing survival rates and fluorescence between 400pg and 800pg injections in X. laevis. Both graphs represent one week each with good survival rates. A: 400pg injections resulted in 6% higher survival than 800pg for embryos injected with 1965 and 3% more fluorescence. B: 800pg injections resulted in 6-20% higher survival than 400pg and 13-26% higher fluorescence.

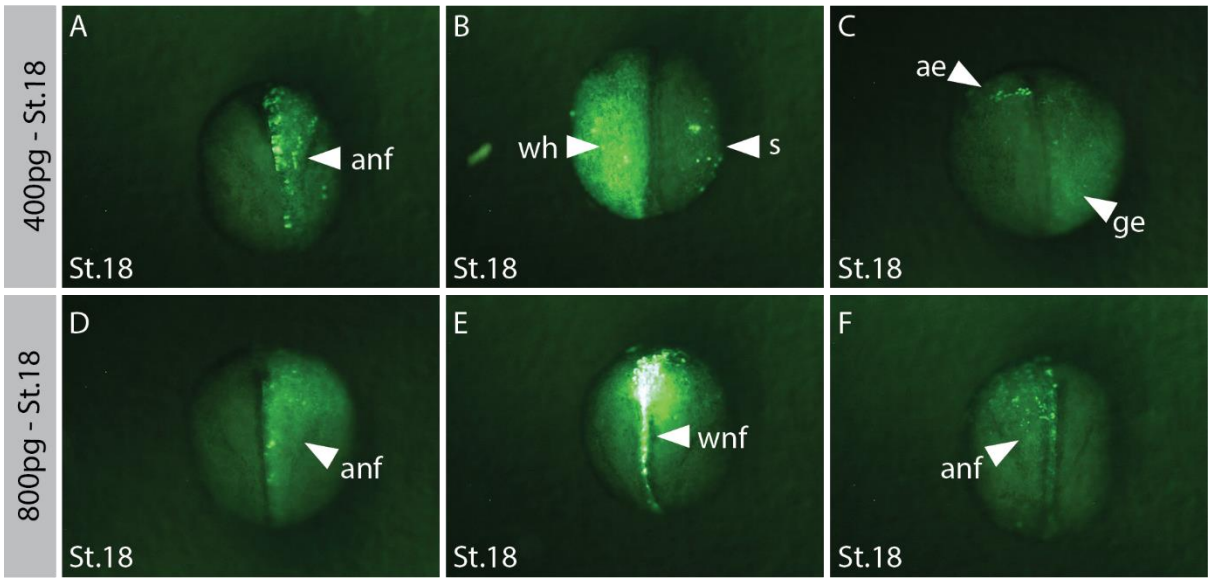


Figure 17: 400pg injections vs 800pg injections of CRE 34. A-C: With 400pg injected there is varied expression. A: Expression is seen anterior long the neural fold (anf). B: Expression is seen in the whole half (wh) on the left of the embryo and spotted (s) in the right half. C: Weak expression anterior (ae) in the embryo and general expression (ge) in the posterior half on the right of the embryo. D-F: With 800pg injected there is also varied expression. D: Expression along the anterior neural fold (anf). E: Strong expression within the neural fold (wnf) from anterior to

posterior. F: Weak spotted expression along the anterior neural fold (anf). A-F: Dorsal view, anterior top, posterior bottom.

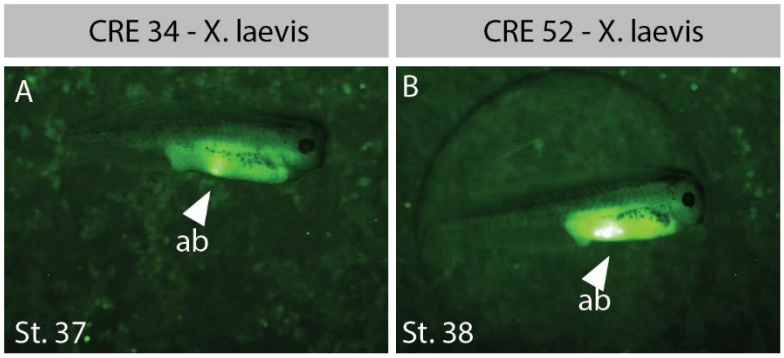


Figure 18: Expression observed as a spot or spots in the abdomen ventrally in tadpoles injected with different CREs. A: CRE 34 associated with snai2. One spot seen in abdomen (ab). B: CRE 52 associated with irx3. Several spots seen in abdomen (ab) at the same location as in A. A-B: Lateral view, anterior right, posterior left, dorsal top, ventral bottom.

4.2.2 Reporter expression by CREs

The reporter constructs were injected into one-cell staged embryos, then the embryos were grown until stage 11.5 when they were first visualised under a fluorescent microscope. They were then observed until stage 20 and reobserved at stage 27-30. From this it was discovered that twelve of the constructs give transgenic fluorescent expression. These were CREs 22, 26, 31, 33, 34, 52, 62, 63, 138, 180, 183, and 192 (Table 3). These expression patterns will be discussed further in the following sections. The other seven constructs that were cloned did not show any fluorescent expression.

4.2.2.1 CRE 22

CRE 22 can be found 8 kb upstream of the *sox9* gene on the small chromosome of *X. laevis*. It is 455 bp long, has good conservation with *X. tropicalis*, and expression from the reporter gene is seen when injected into both *X. laevis* and *X. tropicalis*. This region was selected for analysis before the ATAC-seq data was analysed in comparison with the RNA-seq data and it was therefore later discovered that Sox9 is not found in the corresponding NC RNA-seq data. However, reporter gene expression can still be seen, and *sox9* expression has been observed in *Xenopus* embryos from gastrula stage onwards (Spokony et al., 2002). In this project, the reporter expression is first observed at the lateral edge of the neural plate at stage 13 (Figure 19A). As the neural border folds inwards towards the dorsal midline, the fluorescent expression migrates with the fold until it fuses at the midline to form the neural tube (Figure 19B-E', Figure 20, Figure 20A-C). The expression is most consistent at stage 20 where it is

weaker at the posterior dorsal midline and thickens anterior along the dorsal midline, although the expression does in most transgenic embryos not extend past the neural fold at the anterior end (Figure 19E'). 6.60% (43/652) of the surviving injected embryos were transgenic. The described expression pattern was seen in 4.45% (32/652) of those embryos. At tailbud stage 30, 3.07% (20/652) of the embryos showed fluorescent expression along the neural tube where it is stronger anterior in the neural tube and brain, and often expression can be seen in the eye and otic vesicle (Figure 19F-F', Figure 20D, Figure 21). Some embryos show expression in the branchial arches (Figure 20D, Figure 21C). Figure 21 shows variable expression at stage 30 which has all been classed as neural expression.

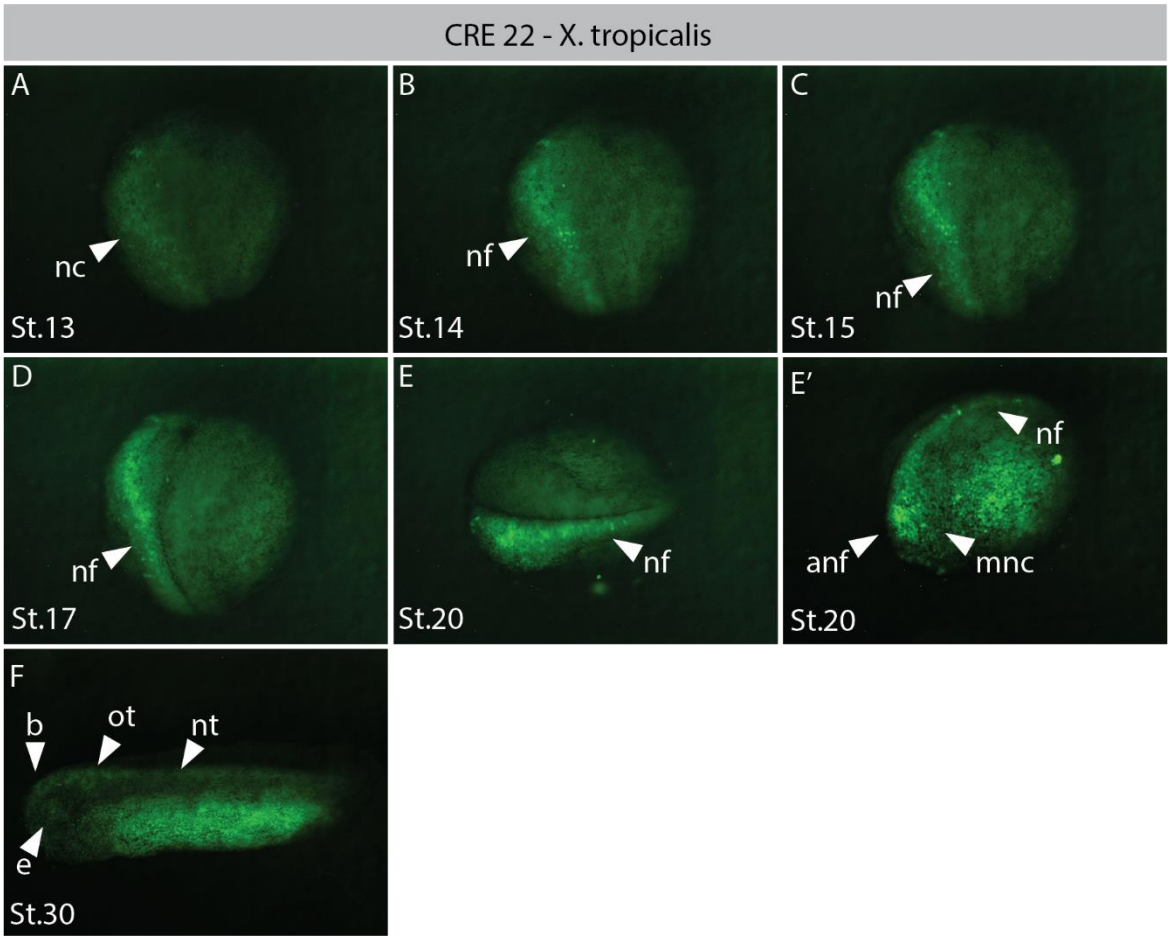


Figure 19: Expression pattern of GFP by CRE 22 associated with sox9 seen in X. tropicalis. A: Week fluorescence appeared at early neurula at the edge of the neural plate where the neural folds (nf) form, possibly in the neural crest (nc). B-D: The expression in the neural crest moves with the neural fold towards the dorsal midline. E-E': Expression can be seen along the fusing neural fold and the anterior neural fold (anf) in the embryo. Possible migrating neural crest (mnc) from the anterior of the embryo (E'). F: Week GFP expression can be seen in the neural tube (nt), otic vesicle (ot), brain (b) and the eye (e). A-D: Dorsal view, anterior top, posterior bottom. E: Dorsal view, anterior left, posterior right. E': Lateral view, anterior left, posterior right, dorsal top, ventral bottom. F: Lateral view, anterior left, posterior right, dorsal top, ventral bottom.

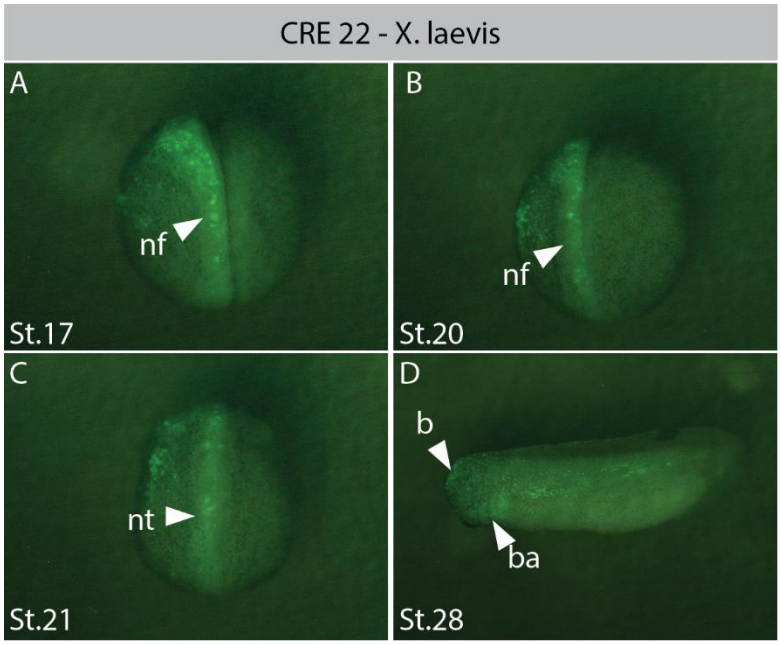


Figure 20: Expression pattern of GFP by CRE 22 associated with sox9 seen in X. laevis. A-C: Expression in mid neurula along the neural fold (nf) as the nf fuses to become the neural tube. D: Expression in the brain (b) and the branchial arches (ba). A-C: Dorsal view, anterior top, posterior bottom. D: Lateral view, anterior left, posterior right, dorsal top, ventral bottom.

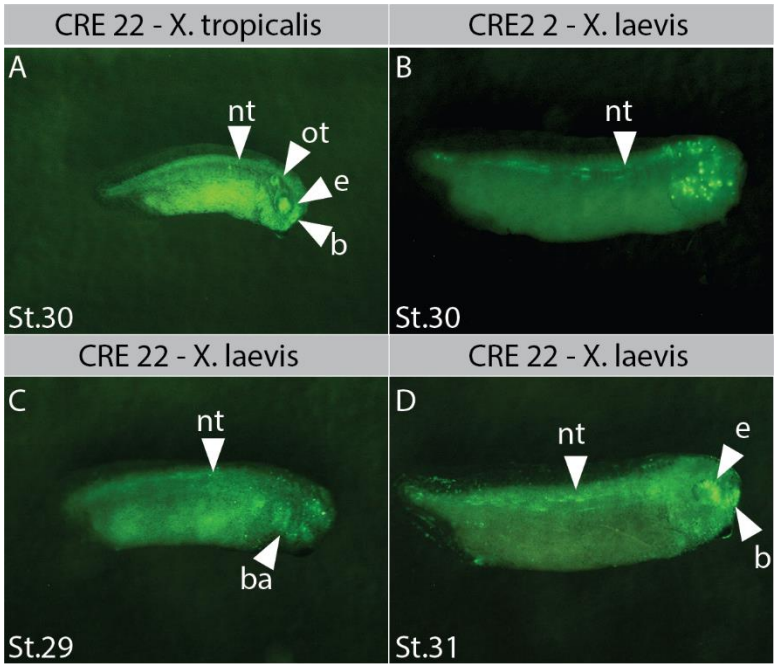


Figure 21: Variable late expression by CRE 22. A: X. tropicalis embryo showing expression in the neural tube (nt), the otic vesicle (ot), eye (e) and the brain (b). B: X. laevis embryos showing expression in the neural tube (nt) and spotted expression in the head. C: X. laevis embryo showing expression in the nt and the branchial arches (ba). D: X. laevis embryo showing expression in the neural tube (nt), eye (e) and the brain (b). A-D: Lateral view, anterior right, posterior left, dorsal top, ventral bottom.

4.2.2.2 CRE 26

CRE 26 can be found 1.5 kb upstream of the *pdgfra* gene on the long chromosome, and the region is 331 bp long. Expression of the reporter gene was not seen when the construct has been injected into *X. tropicalis,* only in *X. laevis*. Fluorescent expression can first be seen at stage 12.5 in the posterior neural plate originating from the closing blastopore. There may be stronger expression at the lateral edge of the neural plate, however, this is not very clear (Figure 22A). The expression can in stage 14 be seen along the neural fold where it migrates with the neural fold towards the dorsal midline (Figure 22B). At stage 16 the expression does not extend above the anterior neural fold, however, at stage 19 the expression has extended over the anterior neural fold. At stage 16 and 19 the posterior expression along the neural fold become weaker (Figure 22C-D). This expression pattern can be seen in 6.74% (13/193) of the 9.33% (18/193) transgenic embryos, where the other expression was non consistent with the non-neural ectoderm. (Figure 22A-D, Figure 23). Not all embryos show identical transgenic expression, therefore, Figure 23 show some different expression pattern that are classed as neural. At tailbud stage 30 the fluorescence is weak, and many of the transgenic embryos show no fluorescence. However, it was occasionally observed in the muscle or the presomitic mesoderm along the dorsal posterior regions and possible expression in the branchial arches and the eye in 1.04% (2/193) of the embryos (Figure 22E).

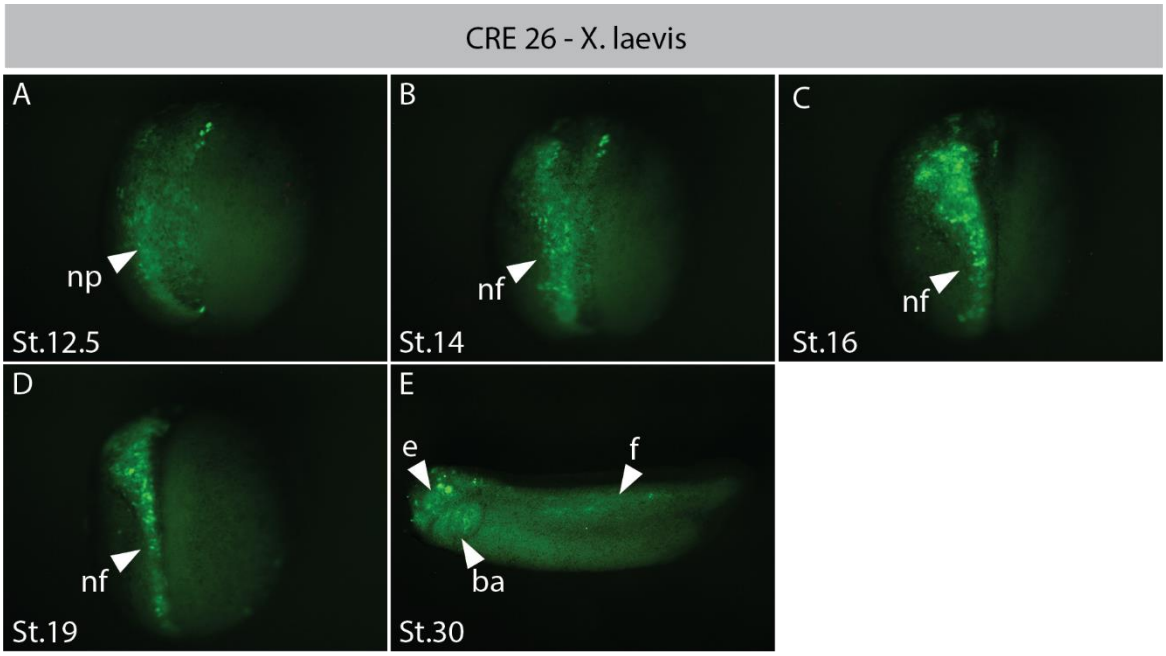


Figure 22: Expression pattern of GFP by CRE 26 associated with pdgfra seen in X. laevis. A: Expression in the posterior neural plate (np), possible posterior neural crest at the edge of the expression. B: The expression moves with the neural cold towards the dorsal midline. C-D: Strong expression along the neural fold. E: Expression in eye (e) and the branchial arches (ba), as well as unknown expression along the posterior of the neural tube (f) which

could be the presomitic mesoderm A-D: Dorsal view, anterior top, posterior bottom. E: Lateral view, anterior left, posterior right, dorsal top, ventral bottom.

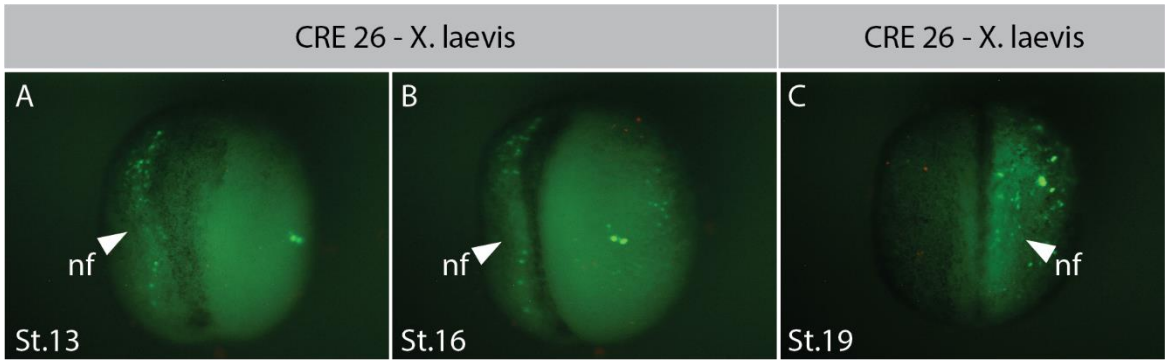


Figure 23: Variable expression seen with CRE 26 associated with pdgfra seen in X. laevis which has also been categorised as neural. A-B: Same embryo at different stage. A: Spotted expression possibly at the edge of the neural plate along the neural fold (nf). B: The spotted expression moves with the neural fold towards the dorsal midline. C: Different embryo with week expression along the neural fold. A-C: Dorsal view, anterior top, posterior bottom.

4.2.2.3 CRE 31

CRE 31 can be found 4.2 kb upstream of the *hes1* gene on the long chromosome. The region is 731 bp long, and Hes1 was not identified in the RNA-seq, like Sox9. However, *hes1* expression has been identified during early embryo development (Vega-López et al., 2015). In this project expression can be seen with transgenic *hes1 X. laevis* embryos. The expression can first be identified at stage 13 in the neural plate (Figure 24A). As the neural folds form the expression follow the neural fold towards the dorsal midline. The expression is weak posterior along the neural tube then becomes strong and spreads out anterior along the neural tube where it extends over the anterior neural fold (Figure 24B-C). This expression has been seen in 2.02% (5/248) of the injected embryos where there were 2.42% (6/248) transgenic embryos (Figure 24A-C). At stage 30 strong expression in the eye and some expression in the branchial arches is seen in 0.81% (2/248) of the embryos, and in one embryo there is also expression in the otic vesicle (Figure 24D).

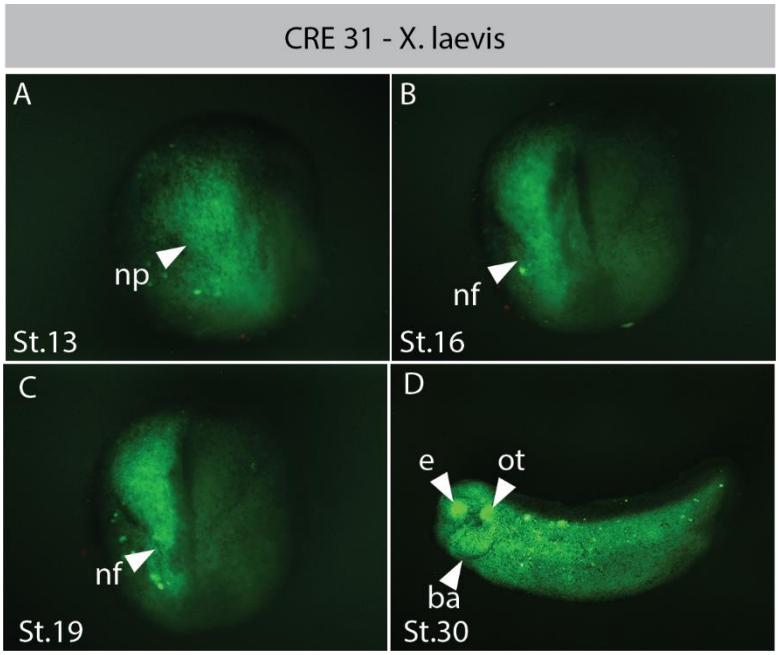


Figure 24: Expression pattern of GFP by CRE 31 associated with hes1 seen in X. laevis. A: Expression seen in the neural plate (np) at early stages. B-C: The expression moves with the neural fold (nf) towards the dorsal midline. D: In the tailbud there is strong expression the eye (e) and otic vesicle (ot). There is also some expression the branchial arches (ba). A-C: Dorsal view, anterior top, posterior bottom. E: Lateral view, anterior left, posterior right, dorsal top, ventral bottom.

4.2.2.4 CRE 33

CRE 33 can be found 13 kb downstream of the *snai2* gene on the short chromosome, but it is also conserved in the long chromosome and with *X. tropicalis.* The region is 465 bp long. Expression is first seen in the neural plate at the start of neurulation (Figure 25A). The expression is then seen at the edge of the neural fold where it is thicker and spreads out anterior in the embryo, and the expression extends over the anterior neural fold. Posterior in the embryo the expression along the neural fold is thinner, and it does not extend to the far posterior of the embryo (Figure 25B). However, when the neural folds fuses to form the neural tube, the expression spreads out along the A-P axis (Figure 25C). This expression pattern is seen in 2.64% (6/227) of the injected embryos where 4.41% (10/227) of the injected embryos were transgenic. In tailbud stages the expression becomes quite variable. In Figure 25D there is expression in the neural tube and the otic vesicle. However, as seen in Figure 26, some tailbud embryos also show expression in the eye and notochord. These stage 30 expressions are examples of what has been classed as neural meaning that 1.76% (4/227) of the embryos show neural expression at stage 30.

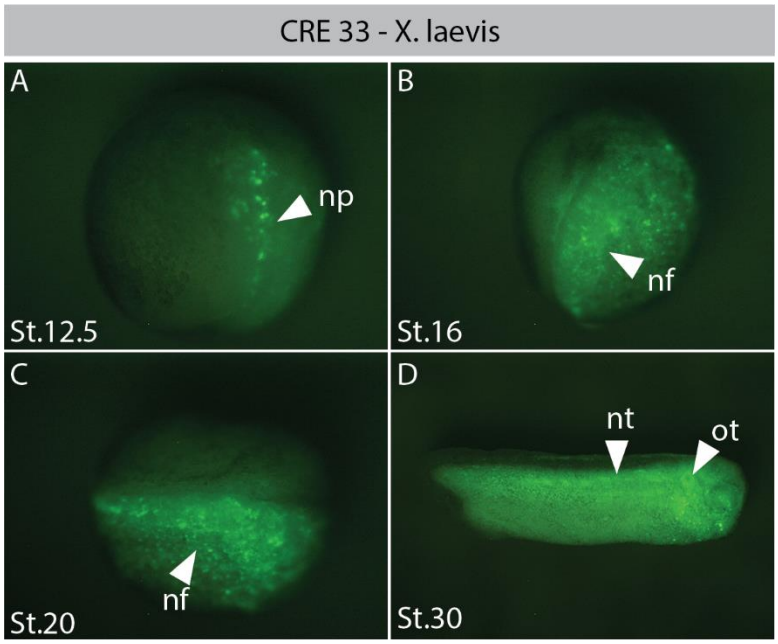


Figure 25: Expression pattern of GFP by CRE 33 associated with snai2 seen in X. laevis. A: Expression seen in the neural plate (np) at early stages. B: Expression seen in the neural fold (nf) as it migrates towards the dorsal midline. C: Expression along the neural fold at the dorsal midline as it fuses to form the neural tube. D: Expression can be seen in the neural tube (nt) and the otic vesicle (ot). A-B: Dorsal view, anterior top, posterior bottom. C: Dorsal view, anterior right, posterior left. D: Lateral view, anterior right, posterior left, dorsal top, ventral bottom.

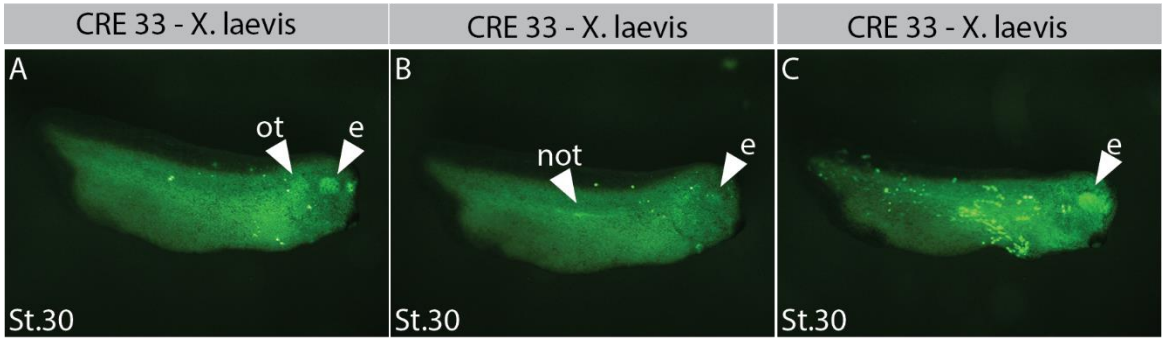


Figure 26: Variable expression seen in stage 30 CRE 33. A: Expression in the otic vesicle (ot) and eye (e). B: Expression in the notochord (not) and eye (e). C: Strong expression in the eye (e). There is also spotted and scattered expression throughout the embryo. A-C: Lateral view, anterior right, posterior left, dorsal top, ventral bottom.

4.2.2.5 CRE 34

CRE 34 is found 40 kb downstream from the *snai2* gene on the short chromosome. It is also conserved on the long chromosome of *X. laevis* and in *X. tropicalis.* The region is also 465 bp, however it is not the same sequence as CRE 33. The CRE is active in both *X. laevis* and *X. tropicalis* as seen by reporter activity in both species. This CRE shows some variation in expression pattern. However, the most reoccurring pattern seen in 6.39% (46/720) of the

10.28% (74/720) transgenic embryos can first be observed at late gastrula stage 12.5 either in the neural plate or the lateral edge of the neural plate where the neural crest originates, from the posterior closing blastopore (Figure 27A, Figure 28A). The expression further concentrates along the neural folds and migrates with the folds towards the dorsal midline where the expression is seen all along the neural tube, thinner posterior in the embryo and thicker anterior in the embryo. The expression folds around the anterior end of the neural fold (Figure 27B-D, Figure 28B-C). In 3.75% (27/720) there is still expression at tailbud stage 30, showing that many of the embryos lose their fluorescence at this stage. However, the embryos that are fluorescent at this stage show the fluorescence mostly along the neural tube, and in some case the eye, otic vesicle, brain, and the notochord (Figure 27E). Figure 29 show some variable expression seen at stage 30 which have all been classified as neural and show some of the mentioned expression patterns above.

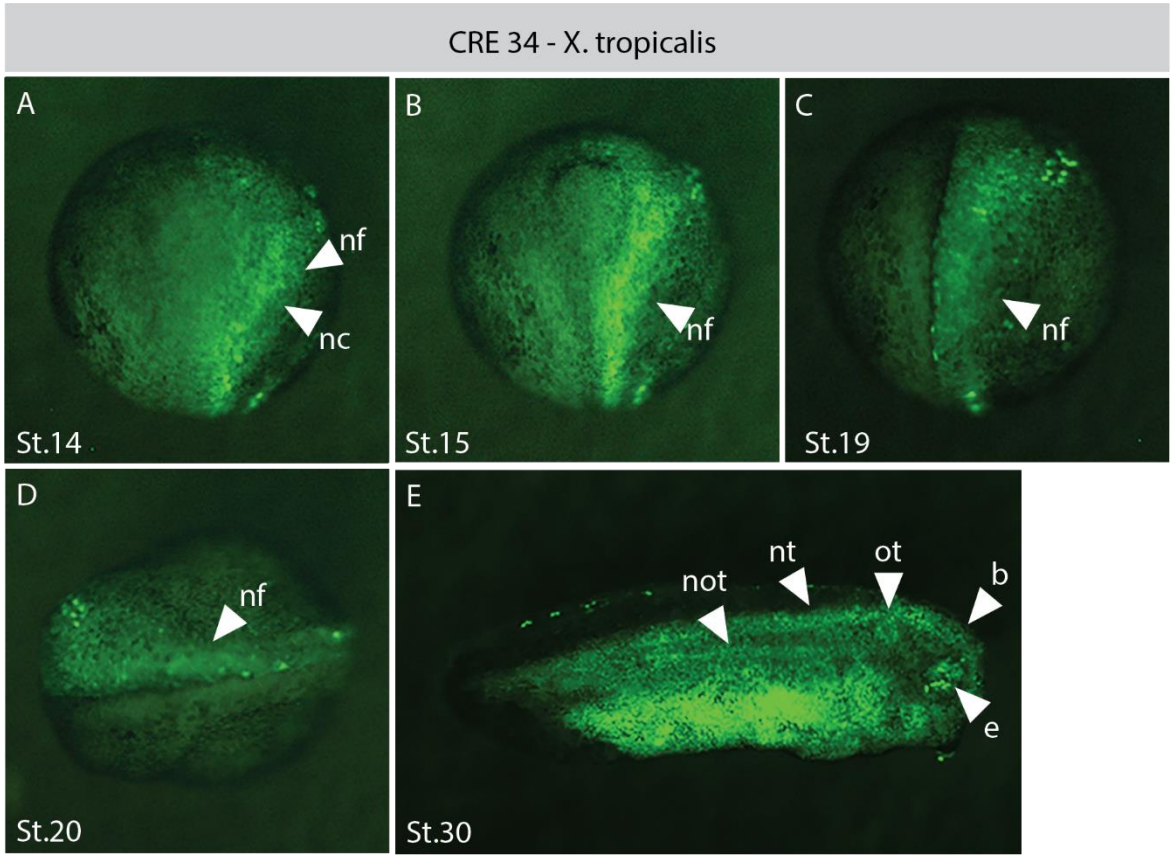


Figure 27: Expression pattern by CRE 34 associated with snai2 seen in X. tropicalis. A: Week fluorescence appear at early neurula from the midline to the posterior in the embryo at the edge of the neural plate where the neural folds (nf) form, possibly in the neural crest (nc). B-C: The expression in the neural crest moves with the neural fold towards the dorsal midline. D: Expression along the fusing neural fold (nf) and the anterior neural fold in the embryo. E: Week GFP expression can be seen in the notochord (not), neural tube (nt), otic vesicle (ot), brain (b) and the eye (e). A-C: Dorsal view, anterior top, posterior bottom. C: Dorsal view, anterior left, posterior right. D: Lateral view, anterior right, posterior left, dorsal top, ventral bottom.

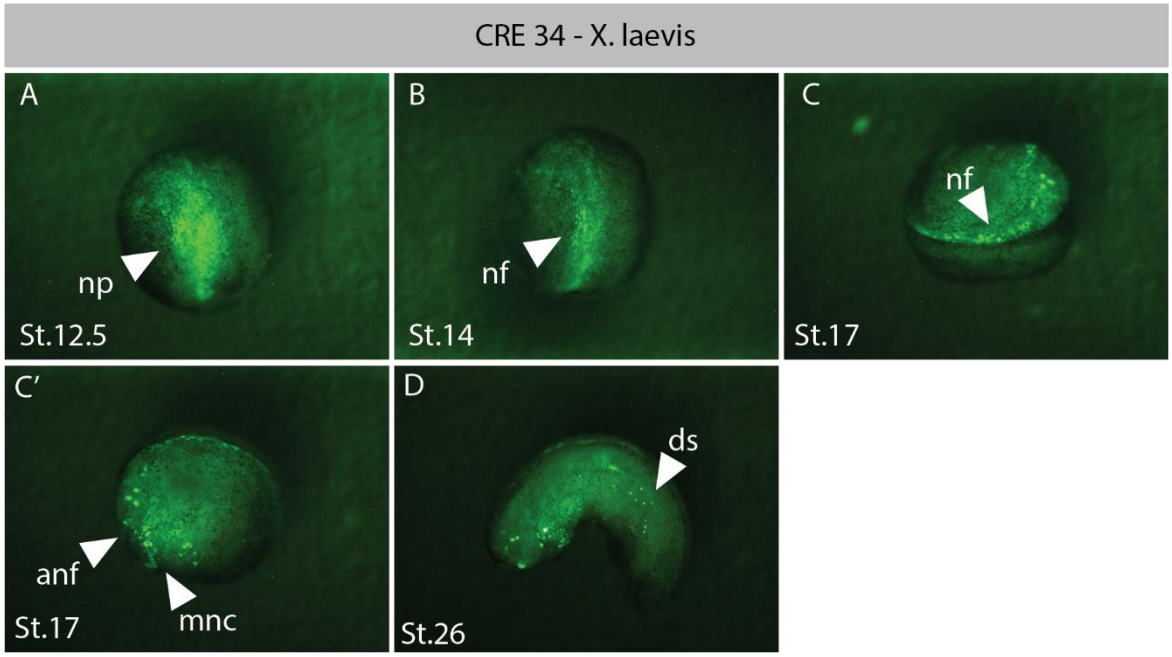


Figure 28: Expression pattern by CRE 34 associated with snai2 seen in X. laevis. A: Strong expression at early neurula from the midline to the posterior in the embryo in the neural plate (np). B: The expression concentrates along the neural fold (nf) and moves towards the dorsal midline. C-C': Expression along the fusing neural fold and the anterior neural fold (anf) in the embryo. C': Possible migrating neural crest (mnc) or dying cells. D: The expression in the older embryo is scattered and not specific, probably dying cells (ds).

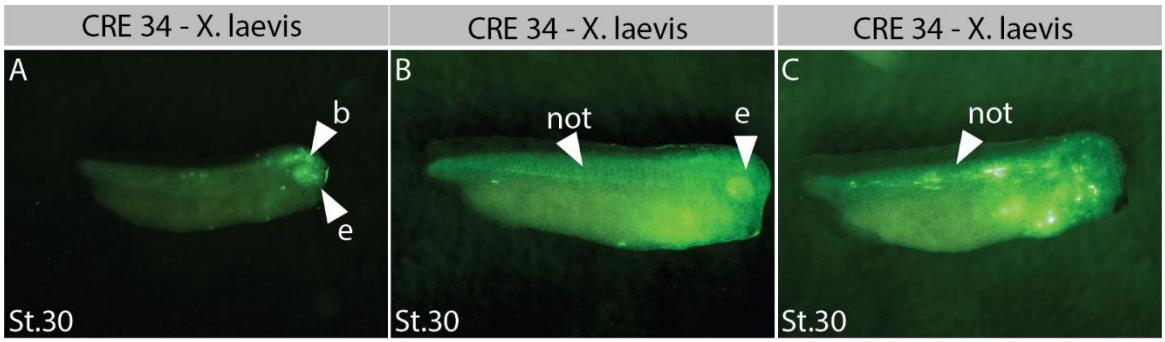


Figure 29: Variable expression seen in stage 30 CRE 34. A: Expression in the brain (b) and eye (e). B: Expression in the notochord (not) and eye (e). C: Expression in the notochord (not). A-C: Lateral view, anterior right, posterior left, dorsal top, ventral bottom.

4.2.2.6 CRE 52

CRE 52 is found 500 bp upstream from the *irx3* gene on the long chromosome. Expression is only seen in *X. laevis* and not when injected in *X. tropicalis*. Expression can first be observed during early neurulation in the neural plate (Figure 30A). The expression then follows the neural fold as it moves towards the dorsal midline, where it fuses to from the neural tube

(Figure 30B). At stage 19 strong expression can be seen from the posterior end to the anterior end along the neural tube with the expression being slightly thicker anterior. However, it does not extend past the anterior neural fold (Figure 30C). This expression is seen in 3.16% (12/380) of the injected embryos where 4.47% (17/380) of the injected embryos were transgenic. At stage 30, 1.84% (7/380) of the embryos have weak expression in the head and along the neural tube as seen in Figure 31A-B. However, some of the stage 30 embryos who have had similar expression patterns to what has been described show expression in what appears to be muscle along the anterior neural tube (Figure 31D-C).

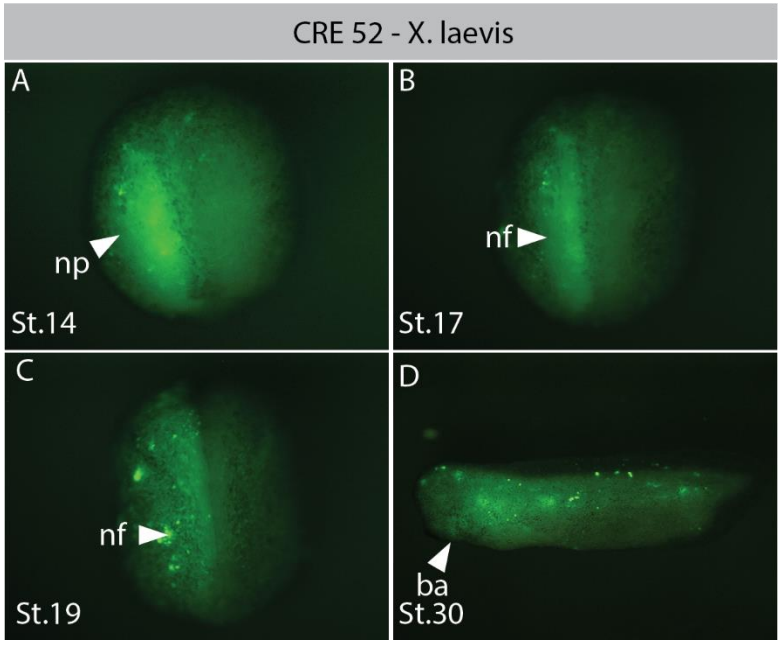


Figure 30: Expression pattern by CRE 52 associated with irx3 seen in X. laevis. A: Expression seen in the neural plate (np) at early stages. B: Expression seen in the neural fold (nf) as it migrates towards the dorsal midline. C: Expression along the neural fold at the dorsal midline as it fuses to form the neural tube. D: Possible expression in the branchial arches (ba), but also spotted expression. A-C: Dorsal view, anterior top, posterior bottom. D: Lateral view, anterior left, posterior right, dorsal top, ventral bottom.

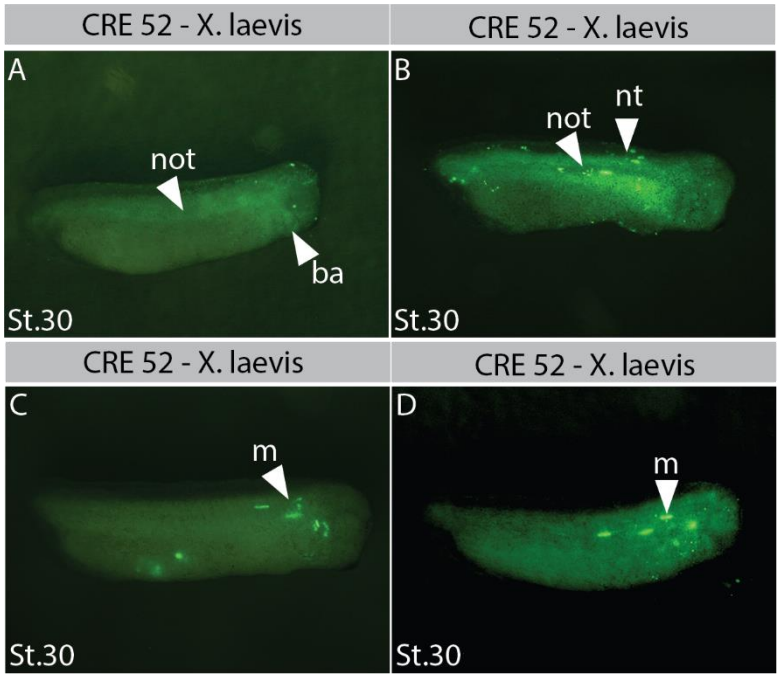


Figure 31: Variable expression seen in stage 30 CRE 52. A: Expression in the notochord (not) and possibly in the branchial arches (ba). B: Expression in the notochord (not) and the neural tube (nt). C-D: Expression in muscle (m) along the anterior neural tube. A-D: Lateral view, anterior right, posterior left, dorsal top, ventral bottom.

4.2.2.7 CRE 62 and 63

CRE 62 is found 500 bp upstream of the *pax3* gene on the long chromosome and is 1115 bp long. CRE 63 is found 1.7 kb upstream from the *pax3* gene on the long chromosome, it consists of two peaks, and is 1595 bp. Expression of both CREs can be observed in both *X. laevis* and *X. tropicalis*. The expression patterns seen by 62 and 63 are very similar, though 63 gives cleaner and more consistent reporter expression in *X. laevis*. For both CREs there is a general expression that can first be observed at stage 12.5 at late gastrula in the neural plate, possibly laterally in the edge of the neural plate in the NC (Figure 32A, Figure 33A, Figure 34A and Figure 35A). The expression then follows the neural fold as it moves towards the dorsal midline where the expression extends from the posterior neural tube to anterior where the expression covers a greater area anterior. CRE 62 have less expression posterior than CRE 63. In both cases the expression can be seen to pass the anterior neural fold (Figure 32B-F, Figure 33B-C, Figure 34B-D, Figure 35B-C''). This expression has been observed in 7.69% (25/325) of the 9.85% (32/325) transgenic embryos for CRE 62 and 8.41% (9/107) of the 10.28% (11/107) transgenic embryos for CRE 63. At stage 30 two different expression patterns can be seen. One of these patterns show expression along the neural tube and sometimes there is some fluorescence in the eye (Figure 32G, Figure 35D). For CRE 62 3.69% (12/325) of the embryos show this pattern. However, 1.85% (6/325) of the embryos also show strong expression in the neural tube, brain and eye in a pattern associated with

pax6 expression (Figure 33D, Figure 36A). Similar expression patterns are seen in CRE 63 with 1.87% (2/107) of the embryos showing expression in the neural tube but not the eye, and 2.80% (3/107) showing the classic *pax6* pattern (Figure 34E-E'', Figure 36B). A comparison between the pattern believed to be *pax6* expression seen with CRE 62 and 63 in albino *X. laevis* embryos can be seen in Figure 36.

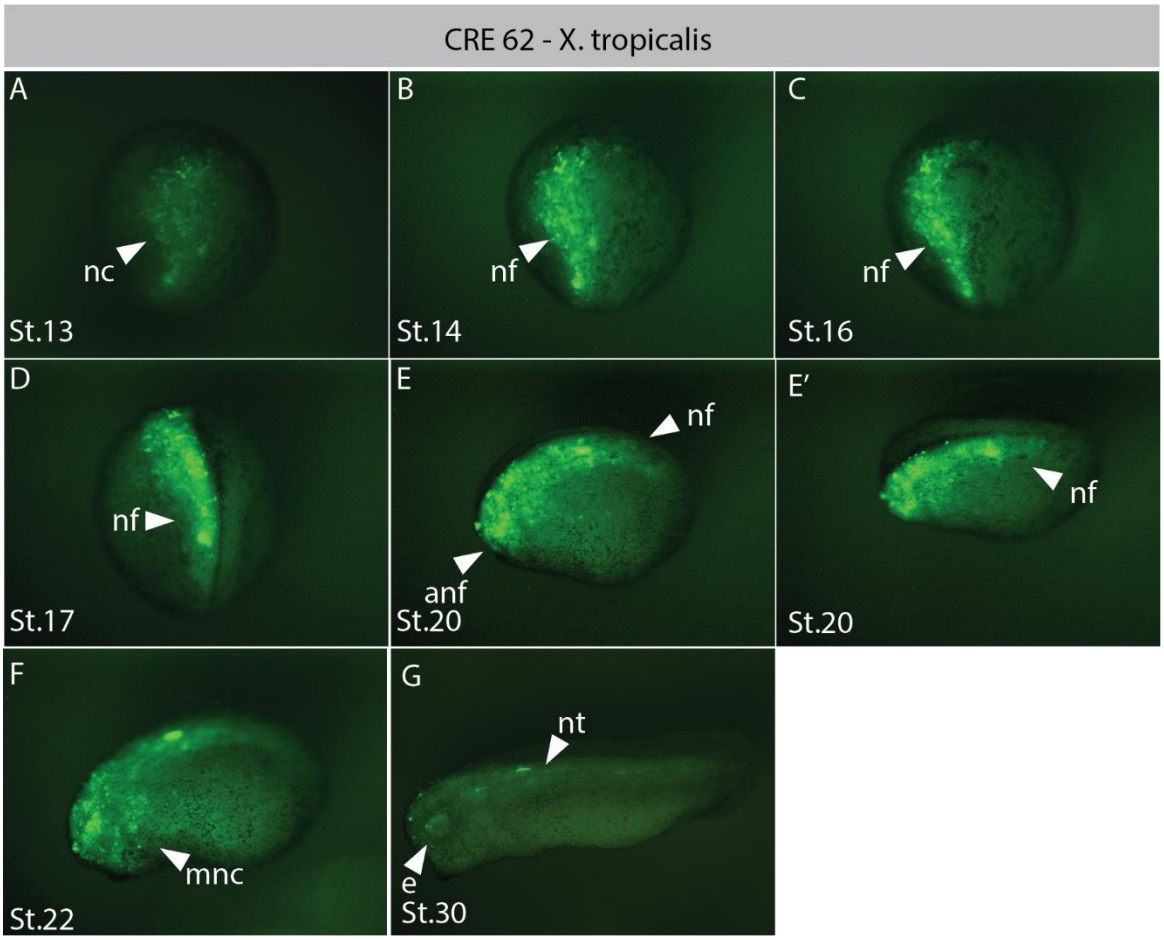


Figure 32: Expression pattern by CRE 62 associated with pax3 seen in X. tropicalis. A: Expression seen at the edge of the neural plate in early stages, possibly in the neural crest (nc). B-D: Expression seen in the neural fold (nf) as it migrates towards the dorsal midline. E-E': Expression along the neural fold at the dorsal midline as it fuses to form the neural tube, and in the anterior neural fold (anf). F: Expression in the neural tube and the anterior of the embryo. Possible migrating neural crest (mnc). G: Expression in the eye (e) and the neural tube (nt) which could also be muscle. A-D: Dorsal view, anterior top, posterior bottom. E: Lateral view, anterior left, posterior right, dorsal top, ventral bottom. E': Dorsal view, anterior left, posterior right. F: Lateral view, anterior left, posterior right, dorsal top, ventral bottom. G: Lateral view, anterior left, posterior right, dorsal top, ventral bottom.

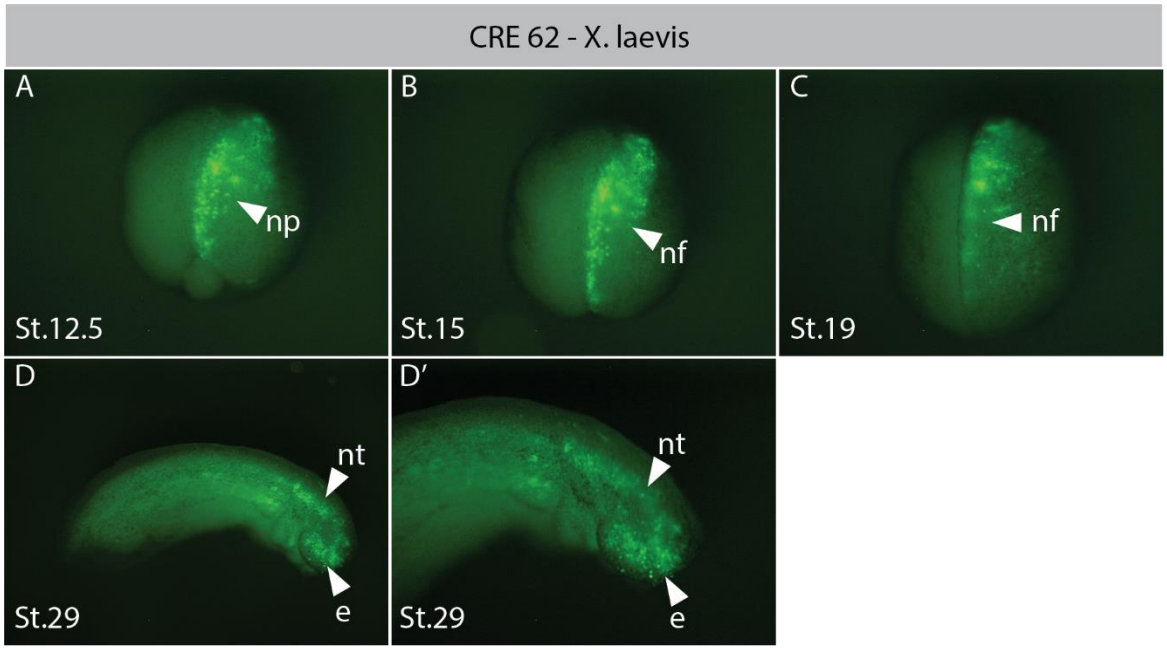


Figure 33: Expression pattern by CRE 62 associated with pax3 seen in X. laevis. A: Expression seen in the neural plate (np). B-C: Expression seen in the neural fold (nf) as it migrates towards the dorsal midline. D: Expression in the eye (e) and neural tube (nt). D': Close up of anterior of D. A-C: Dorsal view, anterior top, posterior bottom. D-D': Lateral view, anterior right, posterior left, dorsal top, ventral bottom.

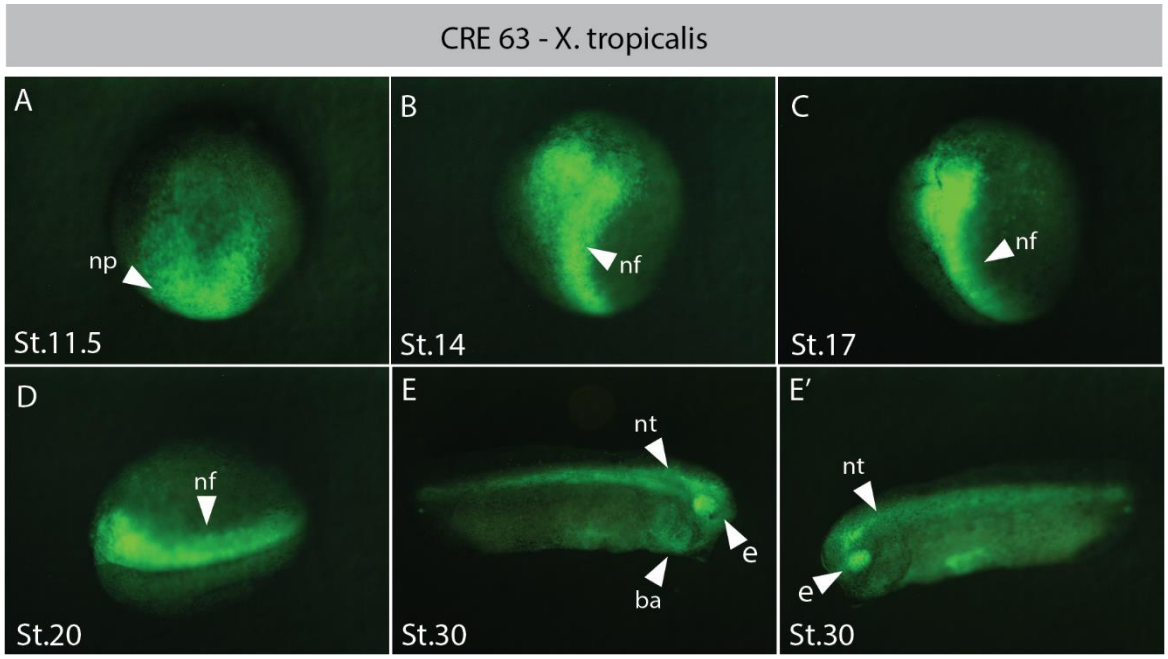


Figure 34: Expression pattern by CRE 63 associated with pax3 seen in X. tropicalis. A: Expression seen in the neural plate (np). B-D: Expression seen in the neural fold (nf) as it migrates towards the dorsal midline. E-E': Expression in the eye (e) and neural tube (nt). E: Possible expression in the branchial arches (ba). A: Uncertain of orientation. B-C: Dorsal view, anterior top, posterior bottom. D: Dorsal view, anterior left, posterior right. E: Lateral view, anterior right, posterior left, dorsal top, ventral bottom. E': Lateral view, anterior left, posterior right, dorsal top, ventral bottom.

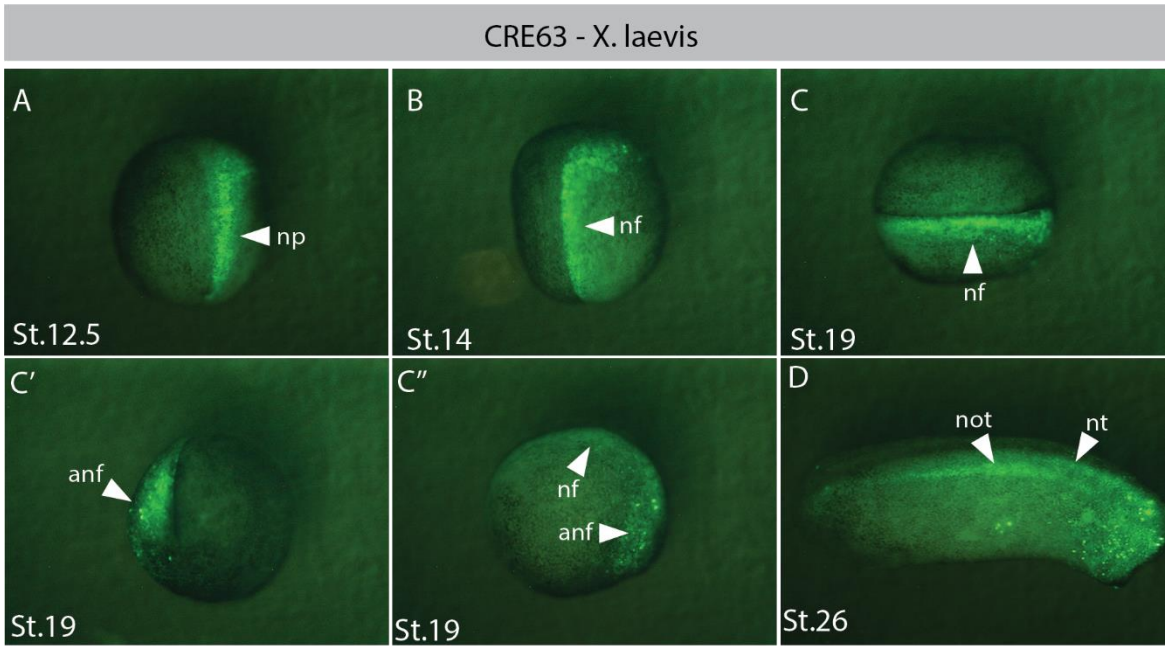


Figure 35: Expression pattern by CRE 63 associated with pax3 seen in X. laevis. A: Expression seen in the neural plate (np). B-C: Expression seen in the neural fold (nf) as it migrates towards the dorsal midline. C': Expression in the anterior neural fold, not extending over the anterior neural fold (anf). C'': Expression in the anterior neural fold (anf) and the neural fold (nf). D: Expression in the notochord (not) and neural tube (nt). A-B: Dorsal view, anterior top, posterior bottom. C: Dorsal view, anterior right, posterior left. C': Anterior view, dorsal top, ventral bottom. C'': Lateral view, anterior right, posterior left, dorsal top, ventral bottom. D: Lateral view, anterior right, posterior left, dorsal top, ventral bottom.

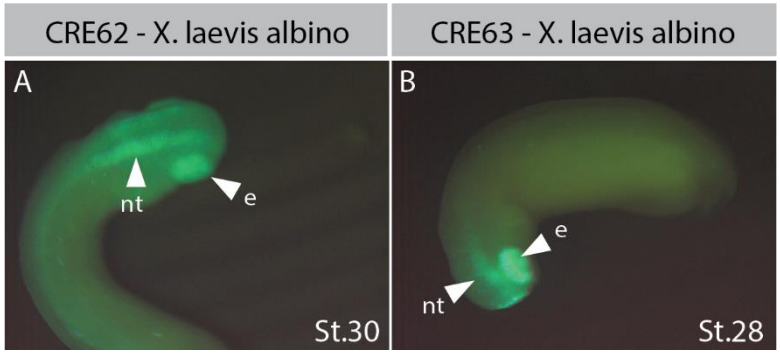


Figure 36: Alternative late expression comparison between CRE 62 and CRE 63 seen in albino X. laevis embryos. A: Expression in the eye (e) and the neural tube (nt). B: Expression in the eye (e) and the anterior neural tube (nt). A: Dorsal view, anterior top, posterior bottom. B: Dorsal view, anterior bottom, posterior top.

4.2.2.8 CRE 138

CRE 138 is found 5 kb upstream from the *prph* gene on the long chromosome and is 404bp long. Expression has only been observed in *X. laevis* where it first appears at around stage 13 in the neural plate, however, the expression is medial with less expression anterior and

posterior in the embryos (Figure 37A). As the neural fold moves towards the dorsal midline the expression moves with it and concentrates along the fusing neural folds (Figure 37B). At around stage 19 the expression moves further posterior and anterior, along the forming neural tube, with thin expression posterior, and spread-out expression anterior in the embryo, however the expression does not extend past the anterior neural fold (Figure 37C). This expression pattern has been observed in 6.41% (10/156) of the 9.62% (15/156) transgenic embryos. At tailbud stage in 4.49% (7/156) of the embryos, expression can be seen in the otic vesicle, the branchial arches and in some cases the neural tube. Many of the embryos have strong expression in the eye. Further there is some expression further dorsal to the neural tube that could be the notochord (Figure 37D).

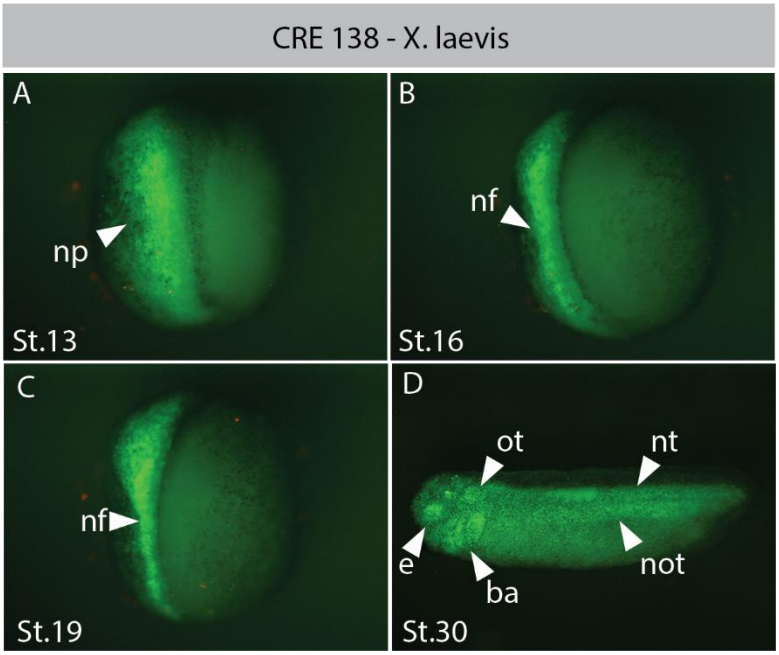


Figure 37: Expression pattern by CRE 138 associated with prph seen in X. laevis. A: Expression seen in the neural plate (np) at early stages. B: Expression seen in the neural fold (nf) as it migrates towards the dorsal midline. C: Expression along the neural fold at the dorsal midline as it fuses to form the neural tube. D: Expression in the eye (e), otic vesicle (ot), neural tube (nt), branchial arches (ba) and possible the notochord (not). A-C: Dorsal view, anterior top, posterior bottom. D: Lateral view, anterior left, posterior right, dorsal top, ventral bottom.

4.2.2.9 CRE 180

CRE 180 is found 3 kb upstream from the *fus* gene on the long chromosome. It is 793 bp long. Expression is only seen when injected into *X. laevis* where it first can be observed at the start of neurulation in the neural plate (Figure 38A). The expression then concentrates along the neural fold from mid posterior to mid anterior (Figure 38B). When the neural fold fuse at the dorsal midline, the expression extends further anterior, but not posterior (Figure 38C). This

expression pattern can be seen in 10.63% (17/160) of the embryos where 13.13% (21/160) of the embryos are transgenic. In tailbud stages there is expression in the neural tube or in stripes along the neural tube which might be muscle expression. Several embryos also show expression in one stripe under the neural tube which might be the pronephros or the notochord. Further, several embryos have expression in the brain and strong expression in the eye. Some embryos have expression in the branchial arches and the otic vesicle (Figure 38D-D', Figure 39A-D). 8.13% (13/160) of the tailbud embryos have been classed as having neural expression.

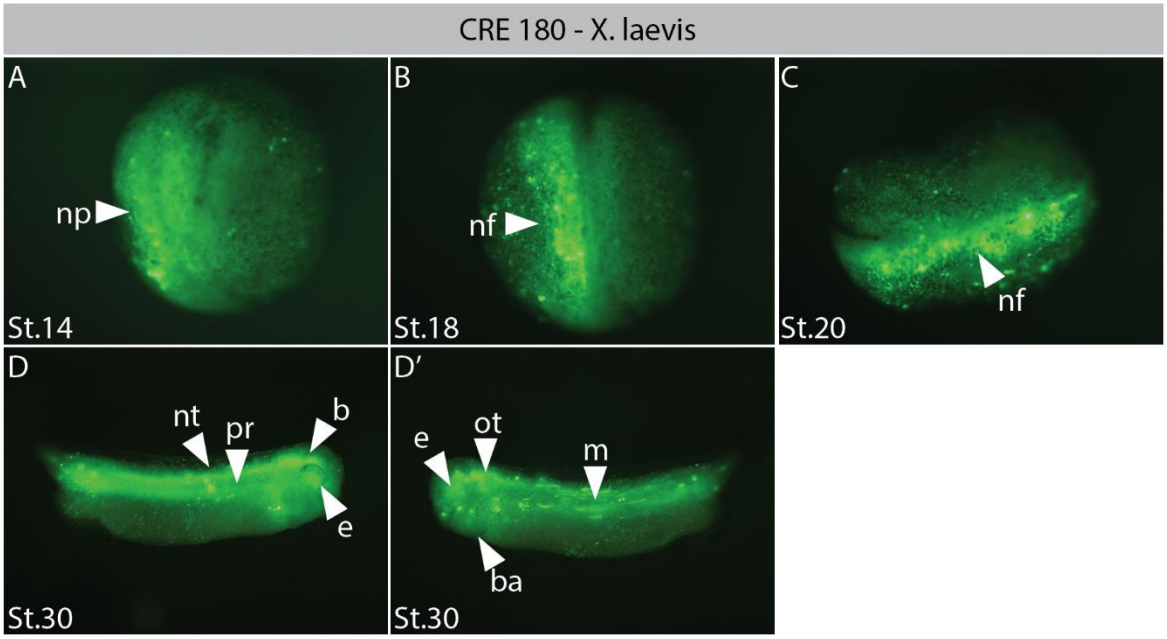


Figure 38: Expression pattern by CRE 180 associated with fus seen in X. laevis. A: Expression seen in the neural plate (np) at early stages. B: Expression seen from mid anterior to mid posterior along the neural fold (nf) as it migrates towards the dorsal midline. C: Expression along the neural fold at the dorsal midline as it fuses to form the neural tube. D: Expression in the neural tube (nt), possibly the pronephros (pr) although it could also be the notochord, brain (b) and the eye (e). D': Expression in the eye (e), otic vesicle (ot), branchial arches (ba) and possible some muscle (m) along the neural tube. A-B: Dorsal view, anterior top, posterior bottom. C: Dorsal view, anterior left, posterior right. D: Lateral view, anterior right, posterior left, dorsal top, ventral bottom. D': Lateral view, anterior left, posterior right, dorsal top, ventral bottom.

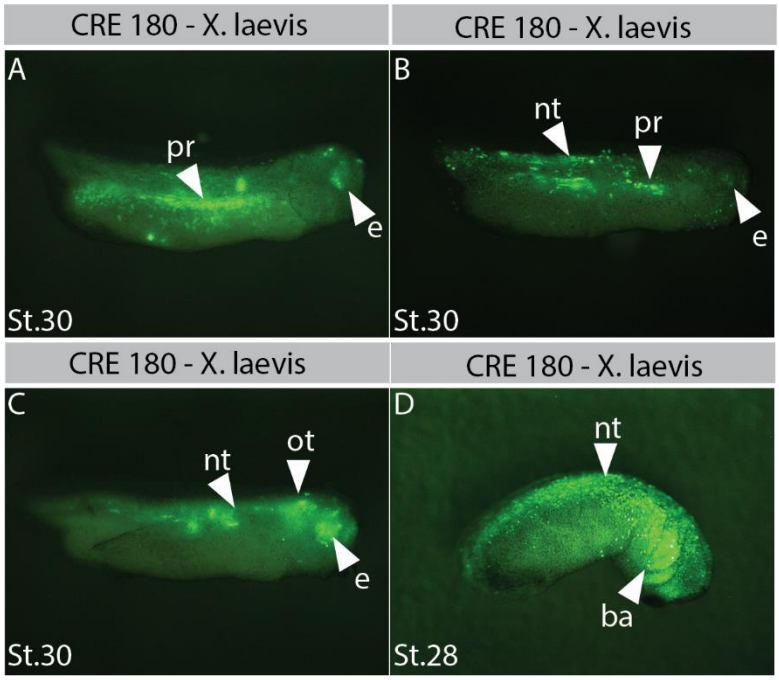


Figure 39: Variable expression seen in tailbud for CRE 180. A: Expression in the possibly the pronephros (pr) and eye (e). B: Expression in the neural tube (nt), possibly the pronephros (pr) and the eye (e). C: Expression in neural tube (nt), although it is stripy so could be muscle, otic vesicle (ot) and the eye (e). D: Expression in the neural tube (nt) and the branchial arches (ba). A-D: Lateral view, anterior right, posterior left, dorsal top, ventral bottom.

4.2.2.10 CRE 183

CRE 183 is found 4 kb upstream from the *wnt8a* gene on the long chromosome, and is 681bp long. Expression of the reporter gene can first be observed at the start of neurulation in the neural plate, possible in the neural crest, although this is difficult to tell apart (Figure 40A). The expression is later found in the neural fold as it forms and migrates towards the dorsal midline. The expression extends from the posterior of the embryo to mid anterior and does not extend past the neural fold, and the thick expression is even all along the neural fold (Figure 40B). This expression pattern is seen in 6.15% (12/184) of the 9.24% (17/184) transgenic embryos. Only 2.27% (5/184) of the tailbud stage embryos show any fluorescence, and two different expression patterns have been seen. In Figure 40C and Figure 41A there is expression in the notochord and also possibly the branchial arches in Figure 40C. Whilst the three other embryos classed as neural all show expression the eye, as seen in Figure 41B, this image also show other NC related structures such as the otic vesicle, brain and neural tube expression.

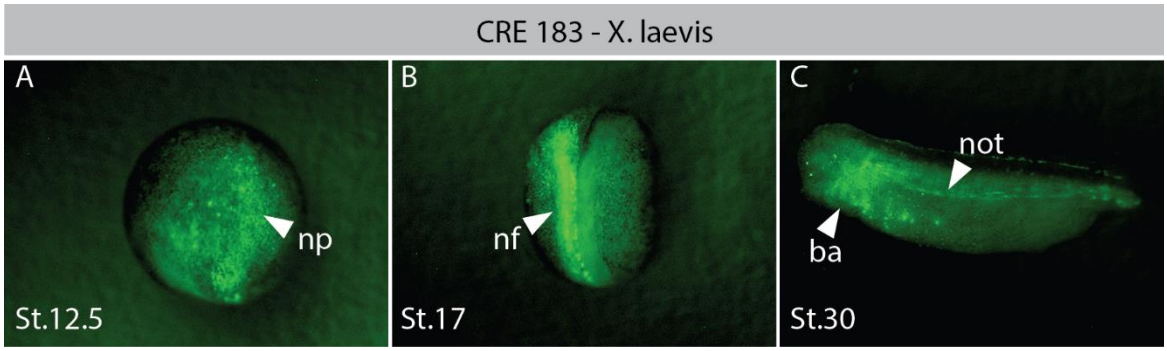


Figure 40: Expression pattern by CRE 183 located close to wnt8a seen in X. laevis. A: Expression seen in the neural plate (np) at early stages, possibly at the edge of the neural plate in the neural crest. B: Expression along the neural fold (nf) from posterior to mid anterior. C: Expression in possibly the notochord (not), or maybe muscle, and the branchial arches (ba). A-B: Dorsal view, anterior top, posterior bottom. C: Lateral view, anterior left, posterior right, dorsal top, ventral bottom.

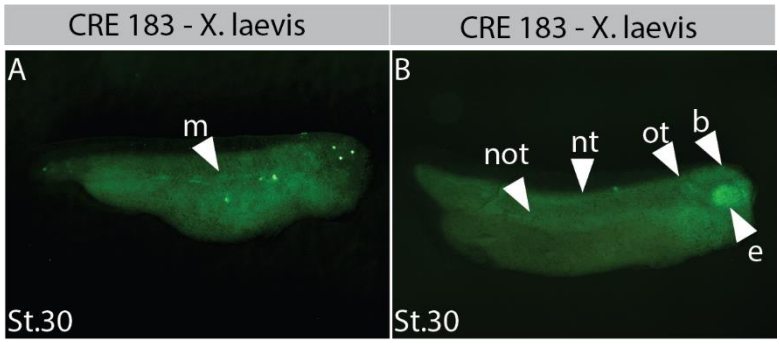


Figure 41: Variable expression seen in stage 30 CRE 183. A: Expression in the notochord or possibly muscle (m). B: Expression in the notochord (not), neural tube (nt), otic vesicle (ot), brain (b) and eye (e). A-B: Lateral view, anterior right, posterior left, dorsal top, ventral bottom.

4.2.2.11 CRE 192

CRE 192 is found 5.7 kb downstream from the *evx1* gene on the long chromosome, and it is 821bp long. *Evx1* expression is not found in the RNA-seq, however, it has been found to be expressed from the midblastula transition until late tailbud in other investigations (Barro et al., 1994). In this project CRE 192 expression has been observed in *X. laevis*, first occurring at the start of neurulation at around stage 13 along the neural fold (Figure 42A). The expression moves with the neural fold towards the dorsal midline as the embryo further develop. Most of the expression is concentrated anterior in the embryo with weak and thin expression along the posterior neural fold (Figure 42B). This expression pattern is seen in all the transgenic embryos, which are 7.83% (9/115) of the surviving embryos. Only 3.48% (4/115) of the tailbud embryos show expression. In one of these embryos seen in Figure 42C, there is expression in the neural tube, otic vesicle, brain, eye, and branchial arches. The other three embryos had

expression in parts of these, one had expression in the eye, one in the branchial arches and one in the neural tube and notochord.

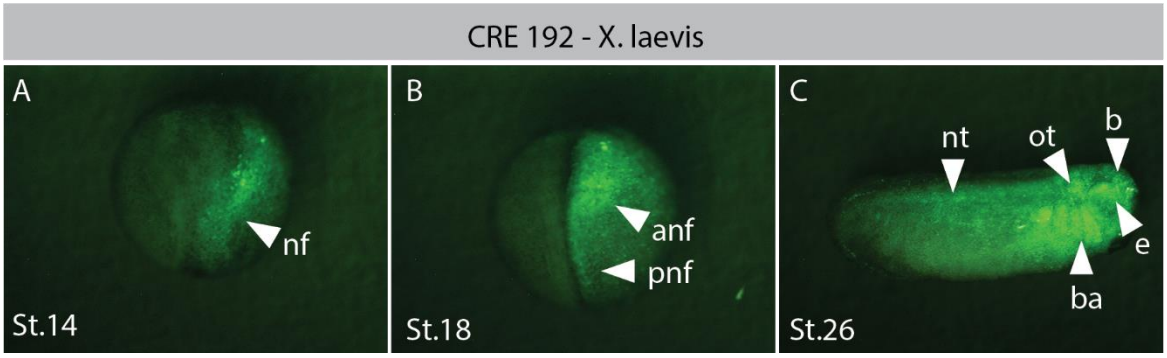


Figure 42: Expression pattern by CRE 192 associated with evx1 seen in X. laevis. A: Expression in the neural fold (nf) at early stages, possibly at the edge of the neural plate in the neural crest. B: Expression along the neural fold (nf), thin expression in the posterior neural fold (pnf) and thick expression in the anterior neural fold (anf). C: Expression in possibly the neural tube (nt), the otic vesicle (ot), the brain (b) the eye (e) and the branchial arches (ba). A-B: Dorsal view, anterior top, posterior bottom. C: Lateral view, anterior left, posterior right, dorsal top, ventral bottom.

4.3 Discussion

4.3.1 Optimizing transgenics

At the start of the transgenesis period of this project there were several complications that had to be resolved. One such complication was that with the control plasmids 1965 and the CRE for Id3 or for early injected CREs identified in the project there was no expression of the reporter gene at early neurula. Instead, at the start of tailbud stages the embryos started to express the reporter gene in the abdomen (Figure 18). To attempt to resolve this problem new digestions of the reporter construct on the same day as the injections were performed and the expression in the abdomen disappeared. Reading further into Ogino's methods revealed that they strongly suggest making the digestion reaction 40 minutes before injections and no earlier as the digestion is not a normal DNA digestion, and even though we do not know yet how the exogeneous DNA gets integrated into the genome it is suggested that after the digestion proteins are required to bind at the digestion ends, and this is not stable for long periods of time (Ogino et al., 2006a). As the digestions were not made fresh each time, and rather frozen down and used a week later, this might suggest why I got this ectopic expression in the abdomen of tailbud and tadpole embryos.

Further to this complication, as stated in the previous paragraph, Ogino suggest making the digestion fresh for each injection and they suggest injecting the one cell-stage eggs as soon as possible after fertilization, completing injections within 45 min from fertilization as the timing of the integration of the injected DNA into the host genome is restricted to early periods of the one cell-stage (Ogino et al., 2006a). However, as in our lab we cannot predict when the embryo will be fertilized and ready to inject, the reactions are usually made during the first collection of eggs, and the reactions are used throughout the rest of the day. The fertilization of and washing of the eggs then take up to two hours. One observation is that most of the embryos are only half fluorescent even though they are injected at one-cell stage. Therefore, the time of the digestion may be the reason why there is only see reporter expression in one half of the embryo as the reporter construct gets integrated into the DNA at a two-cell stage rather than a one-cell stage. Although this does not necessarily compromise the experiments as the two halves develop equally.

In addition, for the transgenesis we chose to inject 800pg (*X. laevis*) or 320pg (*X. tropicalis*) of the plasmid construct into the one-cell stage embryo. This was decided as it did not seem to have a toxic effect on the embryo, and there was little or no reporter expression when lower concentrations were injected. However, this does contradict the procedure designed by Ogino (Ogino et al., 2006a). They suggest injecting 2-6nl of the 100ng/µl reaction to *X. laevis* and only 2nl of the same reaction in *X. tropicalis*, as an increase may lead to a greater fraction of transgenic embryos but also to more dead or abnormal embryos. However, they do state that the capacity for the transgenic embryos to tolerate the injected DNA also depends on the health of the parental frog, and that this might change the injections slightly. Further, they suggest that a higher concentration of the plasmid may lead to greater ectopic expression as not all the plasmids may be digested and integrated (Ogino et al., 2006a). Therefore, this is a possible complication to look out for and that might be observed in the transgenic embryos.

4.3.2 CRE 22, 33 and 34 have similar reporter gene expression patters during neurulation

As discussed in chapter 1.2, Sox9 is involved in NC induction and development in *Xenopus*. One of the CREs that have given the most consistent expression is CRE 22 which we have associated with *sox9* as it is closest to this gene. Further, the expression pattern observed by CRE 22 (Figure 19, Figure 20) corresponds to what has been described of *sox9* expression by Spoknoy and colleagues in the developing *Xenopus* embryos (Figure 43) (Spokony et al., 2002). They identified *sox9* expression shortly after gastrulation at the lateral edges of the neural plate. This is commonly looked at as the NC forming regions. They also observed *sox9*

expression in the sensory layer of the ectoderm adjacent to the NC, known to be the prospective otic placode. The expression in the otic placodal/vesicle continues throughout embryogenesis (Spokony et al., 2002). As the NT forms Spokony and colleagues show that the *sox9* expression is strong along the closing neural fold, not passing the neural fold anterior, and there is little expression in the most posterior regions of the embryo. The *sox9* expression then migrate towards the anterior of the embryo into the cranial structures. During later tailbud development, (stage 25 and onwards) the *sox9* expression is downregulated in trunk NC and persist in cranial NC (CNC) as they populate the pharyngeal arches, otic placodes, developing eye, genital ridge, and notochord (Figure 43) (Spokony et al., 2002). For CRE 22, this expression pattern is seen in some embryos with expression in the eye, and the otic vesicle in Figure 43. There is also expression in the brain in this embryo which does correspond to the image in Spokony's paper, however, this might not be the brain, and rather the craniofacial skeleton (Figure 43D). The embryo in Figure 19F-F' and repeated in Figure 43C-D also has expression in the neural tube, however, if cross sectioned this could be expression in the notochord rather than the neural tube. Further, the embryo in Figure 20D shows expression in the branchial arches (pharyngeal arches). The embryos in Figure 21 also show expression in the mentioned structures, where the expression in the neural tube as said could be expression in the notochord. One reason for why the stage 30 tailbuds show such a differential expression may be because the CRE may not be functioning at stage 30. It was identified as open in stage 13 induced NC samples, and there is no data as to whether the CRE is normally open at stage 30. Therefore, the GFP expression observed may be perdurant GFP protein from earlier development, and some cells may have had more GFP produced and therefore the expression could have persisted for a longer time. The expression pattern for *sox9* described by Spokony in 2002 corresponds with the pattern observed by the CRE 22 associated in our study with *sox9*, therefore it is likely that the CRE does regulate *sox9* expression (Figure 43).

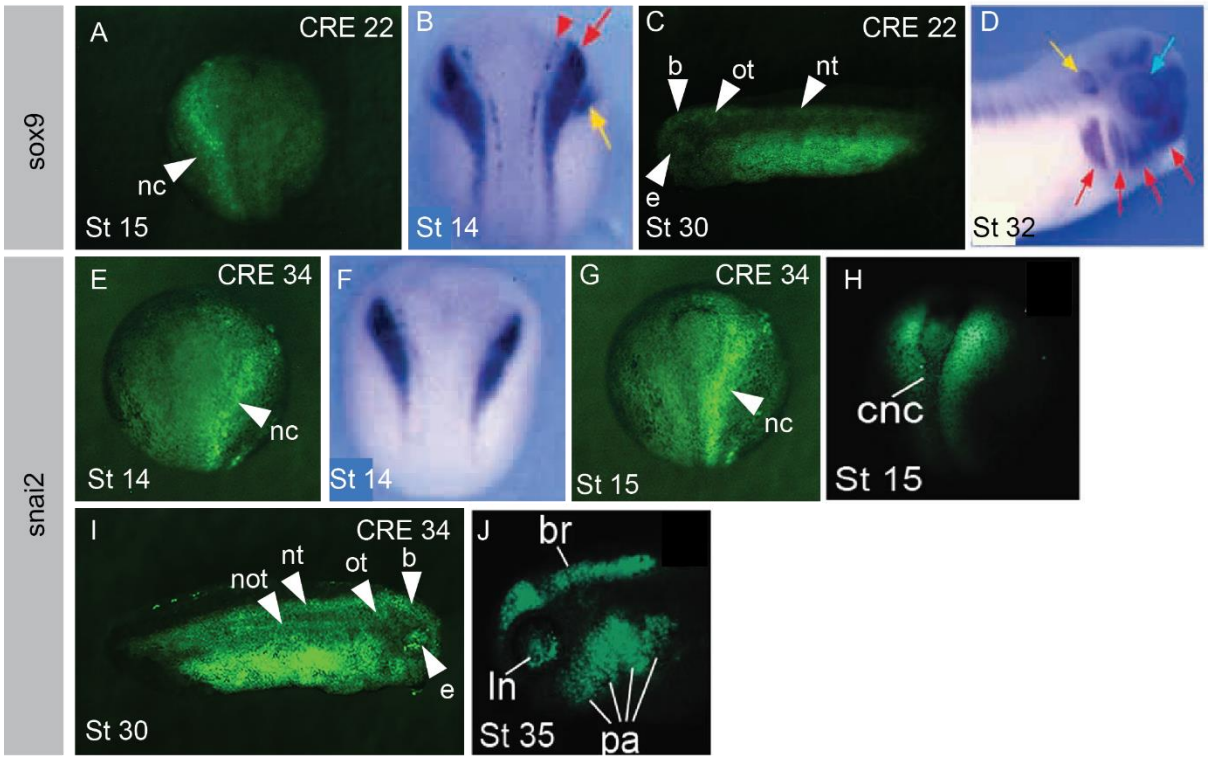


Figure 43: Comparison between CRE 22 and 33, WISH from Spokony and colleagues (2002), and promoter/enhancer for snai2 by Li and colleagues (2019). A: CRE 22 associated with sox9 show expression the neural crest (nc). B: WHIS sox9 expression in the neural crest (red arrow and arrowhead) and the placodes (yellow arrow), by Spokony et al (2002). C: Tailbud expression by CRE 22 in the brain (b), otic vesicle (ot), neural tube (nt) and eye (e). D: WISH tailbud sox9 expression in the head by Spokony et al (2002). Expression seen in the pharyngeal arches (red arrows), otic vesicle, (yellow arrow), and the eye (blue arrow), and restricted areas of the brain. E: CRE 34 associated with snai2 show expression in the neural crest (nc). F: WHIS snai2 expression in the neural crest by Spokony et al (2002). G: CRE 34 expression in the neural crest (nc). H: Fluorescent expression caused by a promoter/enhancer for snai2 discovered by Li and colleagues (2019), showing expression in the cranial neural crest (cnc). I: CRE 34 expression in tailbud. Expression in the notochord (not), neural tube (nt), otic vesicle (ot), brain (b), and eye (e). J: Tailbud expression by enhancer/promoter for snai2 discovered by Li et al (2019) show expression in brain (br), eye (In), and pharyngeal arches (pa). A-B, E-H: Dorsal view, anterior top, posterior bottom. C, J: Lateral view, anterior left, posterior right, dorsal top, ventral bottom. D, I: Lateral view, anterior right, posterior left, dorsal top, ventral bottom.

Spokony (2002) further identified that the *sox9* expression was spatially and temporally co-expressed with *snai2* (*slug*) after gastrulation along the lateral edges of the neural plate (Figure 43F) (Spokony et al., 2002). Mayor (1995) described *snai2* (*XSlu*) expression from stage 12 in the lateral aspects of the embryo (Mayor et al., 1995). At stage 14 the anterior folds were strongly labelled and the posterior folds weakly labelled. At stage 16 the anterior NC was organised into promigratory aggregates. This expression is indeed like the expression of *sox9* (Figure 43). Further, Li (2019) cloned a promoter/enhancer 3.9 kb from *snai2* in *X. tropicalis* and described the expression pattern to be in premigratory neural crest at stage 15 and in the

migrating CNC at stages 19-22 (Figure 43H, J) (Li et al., 2019, Vallin et al., 2001). The expression pattern shown by CRE 33 and 34 which we have associated with s*nai2* are similar and compared in Figure 43. As described for *snai2* expression, CREs 33 and 34 cause reporter expression possibly at the edge of the neural plate as seen in Figure 27A for CRE 34 (*X. tropicalis*). For Figure 25A which show CRE 33 and Figure 28A showing Cre 34 (*X. laeivs*) expression it is uncertain whether the expression is in the neural plate or at the edge of the neural plate which then could be NC. The expression pattern is then similar to *sox9* with expression aggregating in the anterior of the embryo with weak expression along the posterior neural tube. Li (2019) then shows that in stage 35 there is expression in the lens, the brain, and the branchial arches (Li et al., 2019). With CREs 33 and 34 the reporter expression at stag 30 has been variable. However, many embryos do show expression in the eye (Figure 26 for 33, and Figure 27E and Figure 29A-B for 34), where some of the embryos with CRE 34 also showing expression in the brain. Many of the stage 30 embryos for both CREs 33 and 34 also show expression in the neural tube or the notochord, which has not been described by others. Although, as explored for *sox9*, this variability in stage 30 expression might be a result of residual GFP as the CRE might no longer be active at this stage. When analysing the expression pattern for CREs 33 and 34 it corresponds with the described patterns that should be expected. Therefore, it is likely that CREs 33 and 34 are regulators for *snai2*, however, further investigations would be needed such as a wholemount in *situ* hybridisation for GFP to see exactly where the GFP may be expressed.

Further evidence for the regulatory ability of the CREs towards NC induction and development can come from the bioinformatical analysis of TF binding sites in the sequences. Among others TF binding sites found in CRE 22 are Zic1/4/5 as identified by HINT-ATAC. Zic1 has been shown to target among others *sox9* for regulation (Bae et al., 2014), Zic4 has been identified in the NPB (Fujimi et al., 2006), and Zic5 has been identified in the prospective NC (Nakata et al., 2000). Further, the UCSC genome browser predicted a Pax6 TFBS, which could be expected as Spokony (2002) identified *sox9* expression in the eye (Spokony et al., 2002), and Pax6 is a key TF for the development of the eye. However, as it has been shown that in mouse Pax6 act as a repressor to *sox9* rather than an activator (Cohen-Tayar et al., 2018) there is a possibility that it does not account for the activation of the CRE in the eye.

Even though the activity patterns for the three CREs 22, 33 and 34 are very similar, the TFBS identified in their sequences are very different (Table 4). For CRE 33 there are among others Dlx3/5/6. The Dlx family are important TFs for positioning the NPB between the neural and non-neural ectoderm (Woda et al., 2003). For CRE 34, there are Zic1/2 binding sites predicted like CRE 22. Further, Pax3 and Pax7 binding sites have also been predicted by HINT-ATAC,

which are important TFs for NC specification as discussed in subchapter 1.2. The UCSC genome browser also predict Lef1 and Tcf4/12 binding sites with the HINT-ATAC identifying a Tcf7 binding site, suggesting that Wnt signaling is involved in this CRE (Santiago et al., 2017), something they also identified in the promoter/enhancer for *snai2* identified by Li (2019) (Li et al., 2019).

4.3.3 CRE 26 may regulate *pdgfra* expression

CRE 26 is closest to the gene *pdgfra*, which is the Platelet Derived Growth Factor Receptor Alpha, and is important for CNC development. Defects in pdgfra signaling can cause among others the Patch mutant min mice which is characterised by white patches of hair in the trunk (Smith and Tallquist, 2010). Other features of this mutant include aberrant NC cell migration, deficiencies in connective tissue in many organs and defects in non-neuronal derivatives of the crest cells (Li et al., 1996). *Pdgfra* has been shown to be a downstream target of Pax3 and Zic1 signaling which cause a NC fate. It is first activated and expressed at stage 10 in *Xenopus*. However, the temporal expression and function of *pdgfra* has not been explored in *Xenopus* much further. Bae (2014) performed *Wholemount In Situ Hybridisations* (WISH) on two sages; neurula and tailbud and observed pdgfra expression in CNC in the neurula. In tailbud they saw expression in the branchial arches and the presomitic mesoderm (Bae et al., 2014). These patterns may correspond to the expression shown by CRE 26 as compared in Figure 44. There are no WISH images or descriptions for early neurula expression of *pdgfra*. However, Bae (2014) did identify *pdgfra* expression in stage 10 embryos and in Figure 22A-B there is expression from stage 12.5 in the neural plate. At stage 16 in Figure 22C the pattern seen by the GFP showed strong expression anterior in the embryo with less and slightly weaker expression posterior in along the forming neural tube. This pattern may be what Bae (2014) identified. Figure 22 has expression in the branchial arches as identified by Bae (2014), it also has expression along the posterior of the neural tube, and this might be the presomitic mesoderm (Bae et al., 2014).

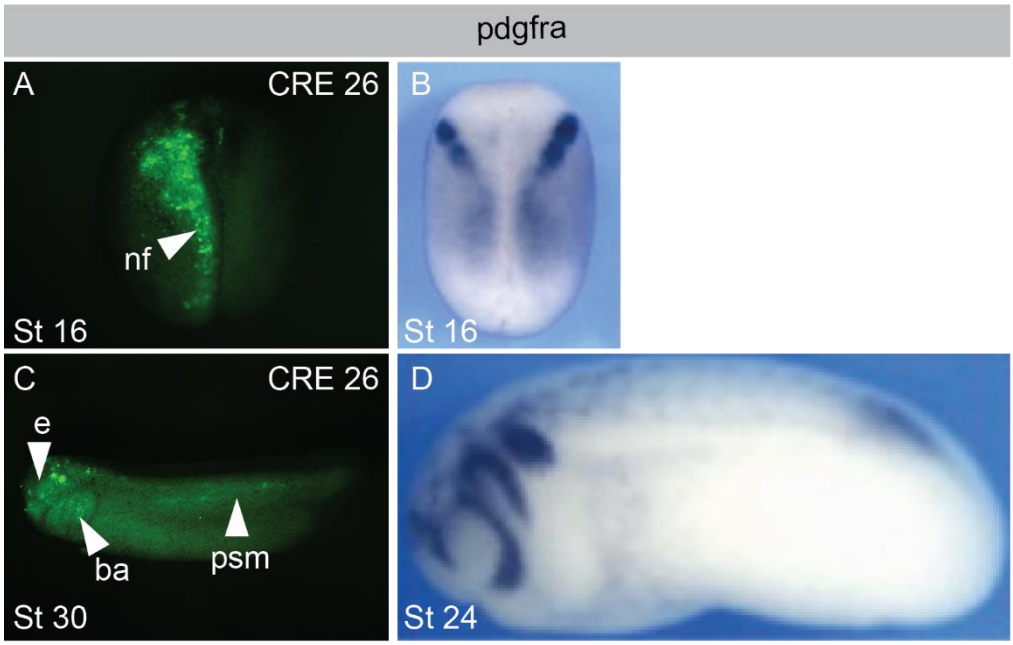


Figure 44: Comparison between CRE 26 and WISH for pdgfra by Bae and colleagues (2014). A: CRE 26 expression along the neural fold (nf), this is possibly the neural crest. B: Pdgfra expression by Bae et al (2014) along the neural fold in the neural crest. C: CRE 26 expression in the eye (e) branchial arches (ba) and possibly in the presomitic mesoderm (psm). D: WISH for pdgfra by Bae et al (2014) expression seen the presomitic mesoderm and the branchial arches. A-B: Dorsal view, anterior top, posterior bottom. C-D: Lateral view, anterior left, posterior right, dorsal top, ventral bottom.

As described in subchapter 1.5.1.3 a CRE might not regulate the closest gene. Therefore, CRE 26 may in fact regulate other genes showing the expression observed. Other genes include *gsx2* which is 50 kb away from the region. Gsx2 or Gsh2 in Xenopus, is expressed in the anterior endoderm and in the developing nervous system from a neural plate stage. It is involved in complex patterning of the brain (Illes et al., 2009), however *gsx2* does not seem to be involved in NC development, so it is unlikely that the CRE is regulating gsx2 (Pei et al., 2011). The other gene within 200 kb of the CRE is *chic2* which is 70 kb away, but there is no published expression on this gene. Further, no TFBS have been identified in CRE 26, although, this does not mean this it is not a CRE as the architecture and grammar of CREs are highly variable. Even though there is little research done on the temporal expression and function of *pdgfra*, the patterns caused by its expression do appear to be like the expression caused by CRE 26. Therefore, it is likely that CRE 26 is a regulator of *pdgfra*.

4.3.4 CRE 31 may regulate *hes1* or *sox2*

Hes1 or *hairy1* as it is also referred to in *Xenopus* is a part of the HES family (hairy and enhancer-of-split) of TFs (Vega-López et al., 2015). *Hes1* is a pleiotropic gene which several different roles during development, from being part of forming the boundary between the

neuroectoderm and the non-neural ectoderm during late gastrulation and early neurulation to maintaining proliferation of neural progenitor cells and to playing a role in gliogenesis and neuronal protection (Hardwick and Philpott, 2019, Vega-López et al., 2015). It has also been identified as a downstream effector of Notch signaling together with several other HES genes where they act as negative regulators of neural differentiation (Murai et al., 2011). In *Xenopus hes1* is expressed in the neural plate and the pronephric mesoderm. It defines dorsally and ventrally the neural crest boundaries and is expressed in the prospective NC and in the prospective placodes surrounding it during neurulation (Murai et al., 2011, Vega-López et al., 2015). Not much else is known about the expression of *hes1* in the tailbud, however unpublished images by Aldo Ciau-Uitz on Xenbase.org suggest that the tailbud embryo has strong expression in the eye, and otic vesicle, as well as weak expression in the head and neural tube. CRE 31 has been associated with the gene *hes1* due to its proximity as it is found 4.2 kb upstream of the gene. It shows expression in the neural plate during early neurulation, which concentrates in the neural fold as the neural tube starts to form. There is stronger expression anterior than there is posterior (Figure 24A-C). However, except for the very earliest expression noted to be in the neural plate by Vega-Lopez (2015), none of the observed expression by CRE 31 during neurulation corresponds to the anterior expression observed for *hes1*. The expression does not mark the border between the neural and non-neural ectoderm, and there is no defined expression in the pre-placodal region. The expression seen by CRE 31 during neurulation is neural. It is possible that the early expression of GFP in the neural plate is covering up this specific expression, however this seems unlikely. During the tailbud stages the observed expression corresponds to the expression noted in Xenbase.org, with strong expression in the eye and the otic vesicle, as well as weaker expression in the head which has been annotated as possibly being the branchial arches in Figure 24D and weak expression along the neural tube. It is very possible that the expression seen by CRE 31 is a random integration. The number of observed embryos with this pattern is extremely low (5/248) and there has been one other transgenic embryo with a different expression pattern. When looking at other genes around the CRE, it is possible that the CRE regulates *sox2*. The *sox2* gene is found 1,100 kb away from the CRE, however CREs have been reported to work over vast distances (Lettice et al., 2003). *Sox2* is expressed widely in developing neuroepithelial cells and mark neural progenitors (Rogers et al., 2009). It is first expressed in the dorsal ectoderm in the early gastrula when neural induction first takes place. During neurulation, *sox2* is expressed in the neuroectoderm from anterior to posterior, as the neural folds move towards the dorsal midline. *Sox2* is then expressed in the developing neural tube, and in the tailbud there is expression along the neural tube, in the eye, forebrain and weak expression in the head. Throughout development the *sox2* expression is pan-neural with expression in the central nervous system, NC, and the placodes (Mizuseki et al., 1998, Rogers

et al., 2009). However, the expression seen by CRE 31 may not cover the whole neural plate and in Figure 24B there does not seem to be expression in the midline within the forming neural tube. Later expression by CRE 31 however, is found in the eye, head, and neural tube. Therefore, it is possible that the CRE regulates *sox2*, however it is not clear. Several TFs for NC development have been identified in CRE 31 further indicating that it is a real CRE and not a random integration (Table 4). These include Dlx2/3, where Dlx TFs are generally very important for the NPB, but also the differentiating NC (Dai et al., 2013, Woda et al., 2003).

4.3.5 CRE 52 is not likely to regulate *irx3*

Irx3 is an *Iroquois* gene which is associated with the development of the border at the posterior diencephalon. At early neurula the expression pattern for *irx3* is associated with two patches at either side of the midline from the posterior forebrain to the spinal cord. As the embryo goes though neurulation the expression along the dorsal midline disappears and two lateral lines anterior in the embryo persists at the border for the diencephalon. In tailbud embryos, expression can be seen in the midbrain and hindbrain, otic vesicle and the pronephros (Rodríguez-Seguel et al., 2009). This pattern is not observed with CRE 52 which has been associated with *irx3* in this project due to its proximity to the gene (Table 3). CRE 52 showed expressions in the neural plate then along the forming neural tube (Figure 30A-C). The tailbud embryo had different expression patterns as well, some showing the branchial arches, notochord, and neural tube, whilst others showed stripy expression seen in muscle (Figure 30D-D', Figure 31). It is possible that the expression seen by CRE 52 is a random integration event, however, the fact that 3.16% of the embryos have similar patterns and only 1.32% of the embryos had other patterns, may make this theory sound unlikely. Further, several TFBSs have been predicted in this sequence that are related to NC development. Several of the Sox genes including *sox2/4/9/10*, as well as a l*ef1* and *tcf7* site to mention some. Another thought is that the CRE regulates another gene as explored previously, a CRE does not need to be in proximity to its target gene to function (Lettice et al., 2003). 120 kb from CRE 52 is the *Fto* gene. *Fto* has been linked with obesity in the adult mice, and loss of *Fto* result in neural crest cell defects (Osborn et al., 2014). However, little is known about is involvement in development. One thing that is known about *Fto* is that it is regulated by Wnt signaling (Osborn et al., 2014). This is interesting as CRE 52 as mentioned has a Lef1 and Tcf7 predicted TFBS, which are associated with Wnt signalling. Therefore, although it now seems unlikely that CRE 52 regulates *irx3*, it is possible that it regulates *fto*, but very little is known about *fto* during development and more research is required.

4.3.6 CRE 62 and 63 have similar expression patterns

Pax3 is one of the TFs that are essential for NPB and NC specification, and together with Zic1 it can initiate expression of early NC markers in the blastula ectoderm to cause NC specification (Milet et al., 2013, Plouhinec et al., 2014). Therefore, it is a well-studied and described TF. In the *Xenopus* gastrula and neural plate stage, *pax3* expression is first observed in broad domains in the posterior and lateral neural plate (Bang et al., 1997). The expression then becomes more defined at the border between the neuroectoderm and the non-neural ectoderm during early neurulation (stage 12). The expression can later (stage 16) be observed in the prospective neural folds in the prospective NC and hatching gland. After neurulation the expression remains in the neural tube, and the superficial hatching gland cells, however it is turned off in the migrating NC (Alkobtawi et al., 2018, Bang et al., 1997). CREs 62 and 63 have been associated with *pax3* due to their proximity to the gene. Further, the expression pattern caused by the CREs during early development are similar to what has been observed for *pax3* as can be seen in Figure 45 for CRE 62. At early neurula there is expression in from the posterior towards the anterior in the neural plate (Figure 32A for 62, Figure 34A and Figure 35A for 63), at times this expression can arguably be more lateral in the plate than medial. The expression caused by the CREs further concentrates in the neural fold as development proceeds (Figure 32B and Figure 33A-B for 62, Figure 34B and Figure 35B for 63). The expression then concentrates along the neural tube from the posterior to the anterior with more expression anterior than posterior (Figure 32C-F and Figure 33C for 62, Figure 34C-D and Figure 35C-C'' for 63). In Figure 32F there is some scattered expression anterior which could be migrating NC however, this can also be non-specific expression commonly seen in dying cells. An interesting observation is made in the expression in the tailbuds for both CREs. A larger percentage show expression in the neural tube with little or no expression in the eye similar to what is observed for p*ax3* except there is no clear expression in the hatching glands (Figure 32G for 62, Figure 35D for 63). However, interestingly there is also a considerable percentage of embryos (1.85% for 62 and 2.80% for 63) that show a pattern associated with p*ax6* tailbud stage expression (Figure 33D-D', Figure 34E-E', Figure 36). *Pax6* is a key regulator of retinal development (Gehring and Ikeo, 1999), and it is very important for the developing central nervous system (Manuel et al., 2015). It is first expressed after completing gastrulation and entering early neurulation (stage 12.5) as two stripes along the midline from anterior to posterior and as a crescent in the anterior of the embryo. By stage 14 the expression in the anterior crescent has intensified and now outlines the anterior-most edge of the neural plate and the neural ridge (Figure 45). This region will later in development form the telencephalon, olfactory bulb, and part of the diencephalon. The

lateral stripes running anterior to posterior along the midline will form part of the myelencephalon (Hirsch and Harris, 1997). At stage 23 the distinct *pax6*-like expression that is seen in late stage tailbuds injected with CREs 62 and 63 starts to appear (Figure 45G-H). Before this time the expression observed in the injected embryos follow the neural fold rather than two strips along the midline and even though the expression is seen to pass the anterior neural fold, it does not appear to form a crescent above the neural plate (Figure 45). The early tailbud expression for *pax6* shows two bilateral stripes which are the developing hindbrain consisting of the metencephalon and the myelencephalon. There is also expression in the forebrain in the developing telencephalon and diencephalon, as well as intense expression in the eye. At stage 28 these expressions are intensified, however there is little expression in the mesencephalic region (Hirsch and Harris, 1997). This strong expression as two bilateral stripes in the head of the embryo is indeed seen a smaller percentage of tailbud embryos injected with CRE 62 and 63. (Figure 33D-D' for 62, Figure 34E-E' for 63, Figure 36 for comparison). So, the question stands; how does the neural or neural crest expression during neural development caused by CREs 62 and 63 cause a small percentage of tailbud embryos to express a classic *pax6* pattern?

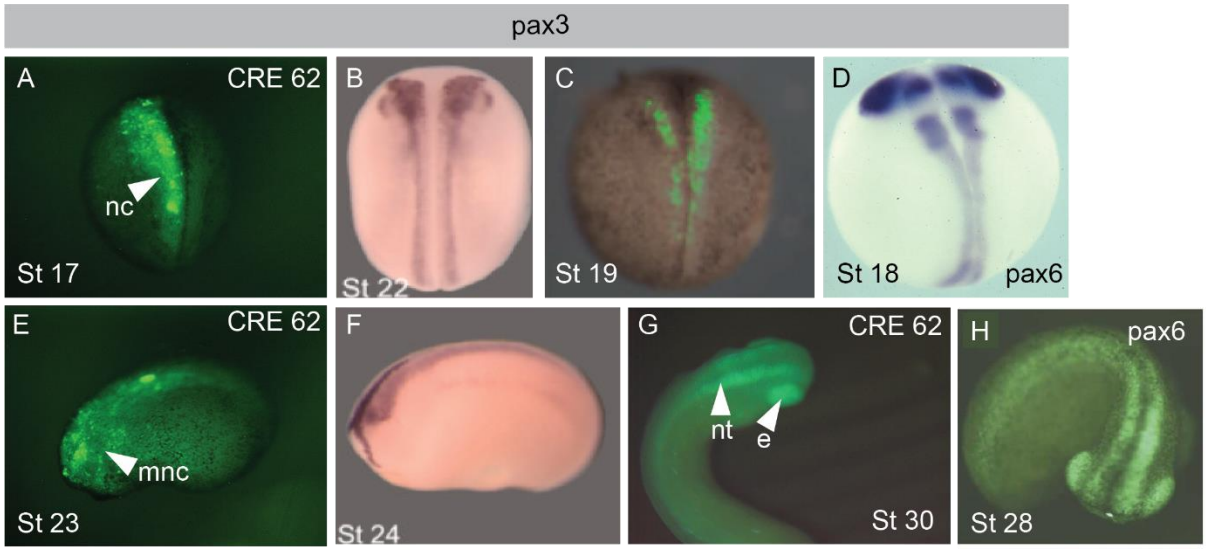


Figure 45: Comparison between CRE 62 expression, WISH and enhancer for pax3 by Alkobatwi and colleagues (2018), WISH for pax6 by Harland labs on Xenbase.org, and transgenic Xenopus for pax6 expression by Ogino and colleagues (2006). A: Expression by CRE 62 in the neural crest (nc) along the neural tube. B: WISH of pax6 showing expression along the neural tube by Alkobatwi et al (2018). C: Expression along the neural tube by an enhancer for pax3 discovered by Alkobatwi and colleagues (2018). D: WHIS expression of pax6 by Harland labs on Xenbase.org along the neural tube and anterior as a crescent in the embryo. E: Expression by CRE 62 in the neural tube and possibly in migrating neural crest (mnc). F: WISH of pax3 in the neural tube and anterior in the embryo by Alkobatwi and colleagues (2018). G: Expression in another embryo by CRE 62 showing expression in the eye (e) and neural tube (nt). H: Expression caused by the pax6 promoter by Ogino and colleagues (2006). A-

D: Dorsal view, anterior top, posterior bottom. E-F: Lateral view, anterior left, posterior right, dorsal top, ventral bottom. G: Dorsal view, anterior top, posterior bottom. H: Dorsal view, anterior right, posterior left.

There are several speculative theories as to why this is the case. One theory could be that the CREs have a dual function where they during neurulation regulate *pax3* expression, then during tailbud they regulate *pax6* expression (Krijger and de Laat, 2016). However, p*ax3* is found on chromosome 5 in *Xenopus*, and *pax6* is found on chromosome 4, so this is no longer a *cis*-regulation in this case. Enhancers can regulate genes on different homologues of a chromosome in trans. This is called transvection. However, this does not explain the phenomenon seen as the genes are on different chromosomes, not just homologues (Fukaya and Levine, 2017, Tsai et al., 2018a). Therefore, this theory is not plausible. Another theory comes from the eRNAs, where it has been speculated whether eRNAs can bind promoters to regulate gene transcription in trans. This theory is strengthened by Tsai's (2018) observations of ^{DDR}eRNA transcribed from mouse chromosome 7 being able to act in trans on chromosome 1 and regulate *Myogenin* (Panigrahi and O'Malley, 2021, Tsai et al., 2018b). It might be that CREs 62 and 63 are transcribed as eRNAs and then regulate *pax6*. It is known that pax3 and pax6 interact greatly during early development as indirect repressors of each other in chicken. p*ax6* is initially expressed in the pre-placodal region, as a crescent surrounding the cranial neural plate, at the forbrain level. *FGFs* and *wnts* expressed by the midbrain and the midbrain/hindbrain boundary then cause expression of *pax3* which expands ventro-laterally into the pre-placodal region which repress *pax6* in that region. Miss expression of *pax6* has also been seen to repress *pax3* expression in the prospective ophthalmic trigeminal placodes (Wakamatsu, 2011). This interaction is further emphasised by looking at the TFBS observed within CRE 62 which has a predicted pax6 binding site. However, even though pax3 and pax6 interact during development, this does not account for why a CRE on a different chromosome from *pax6* would show a *pax6* expression pattern in the tailbud.

It is also possible that there is an error with the method, although one could expect to see this problem in more of the injections and not just the two CREs for *pax3*. However, when the Granger lab made their plasmid for I-SceI mutagenesis that was being used in this investigation, they did first use it to make a transgenic *pax6 X. tropicalis* (Ogino et al., 2006b). There could be contamination from this group in our sample. But then again, it would be expected that the *pax6* pattern would be observed in embryos injected with other constructs. Lastly, the CRE 62 and 63 have some highly NC specific TFBS such as Zic3/4 for CRE 62 and Zic1/2/3/4/5 for CRE 63. And both have myc and mycn to mention a few, further strengthening the neural or neural crest expression patterns observed in the neurula injected

embryos. The two CREs for *pax3* show some interesting patterns that are both expected for *pax3* and not. Therefore, more investigations into this strange expression are needed.

Further, it is important to note that a 2.9 kb region upstream of the *pax3* promoter has been identified previously in *X. laevis* by doing a comparative analysis between Xenopus and other vertebrate species by Alkobtawi and colleagues. They found a region that had ≥29.6 % homology. They generated a transgenic line of frogs using this region in a Tol2 method and found that it caused expression of GFP restricted to the neural folds in the neurula embryos (Alkobtawi et al., 2018). Combining CRE 62 and 63 together forms a 2.7 kb fragment which is likely to be the same regions found by Alkobtawi. However, here we have identified that the two regions can separately cause expression of the reporter genes. Further, in mice a 1.6 kb sequence upstream of *pax3* proximal to the promoter have also been identified and studied with great care, showing the importance of this regulatory region for *pax3* throughout evolution (Degenhardt et al., 2010a).

4.3.7 CRE 138 show some similarities to the expression of *prph*

Prph or *peripherin*, is also referred to as *XIF3* in Xenopus, and is a type III neuronal intermediate filament protein, which is expressed in neurones of the peripheral nervous system (PNS), as well as some brain stem cells, and spinal cord neurones (Eriksson et al., 2008). These cells are descendants of the NC (Umehara et al., 2020). However, much is not known about the early expression or function of *prph* in *Xenopus*. Sharpe (1989) described the expression of *prph* in *Xenopus* at stage 18. They saw intense labelling of *prph* outside the anterior neural tube in what they believe to be the migrating neural crest, which would by stage 22 migrate to the cranial ganglia. They also observed that at stage 18, *prph* could be found as two patches ventrolateral on either side of the neural tube, which they believed could be motor neurones. At stage 18 they saw no expression along the neural tube posterior in the embryo. In stage 26 tailbuds there is prominent staining in the head and dorsal tissues of the embryo, and at stage 32 they saw staining in cranial ganglia such as the posterior hindbrain, and diffuse expression along the neural tube (Sharpe et al., 1989). Further, Green and Vetter (2011) described the expression of *prph* at stage 15 in the trigeminal placodes and primary neurones along the neural tube. Whilst at stage 23 and 28 they saw expression in the trigeminal placodes, olfactory placodes, spinal cord, retina, and many of the brain domains. However, there is also spotty expression just after the head from dorsal to ventral in their images which we in the Wheeler lab believe to be neural crest migrating to the sympathoadrenal regions (Green and Vetter, 2011). CRE 138 has been associated with *prph* due to its proximity (Table 3). It has early expression in the mid neural plate regions which then at stage 16 concentrates

along the neural tube, stronger anterior than posterior (Figure 37). It can be speculated that the expression is not within the neural tube and just at the outside of it seen in Figure 37B. This could correspond to the expression observed by Sharpe (1989), however, this could only be confirmed by sectioning. At stage 19 (Figure 37C), there is only thin expression posterior along the neural tube, contradicting what was seen by Sharpe (1989), but not Green and Vetter (2011), and strong and wide expression in the anterior of the embryo, which could be precursors to the anterior structures observed by both papers. However, the expression observed in stage 30 differs from the observed expression, where in Figure 37D, there is expression in the eye as Green and Vetter (2011) observed. The expression in the otic vesicle, posterior neural tube and posterior notochord has not been described previously. Further, there is expression in what could be the branchial arches. This could be the migrating neural crest observed in Green and Vetter (2011). However, the expression caused by CRE 138 is not identical to the observed expression caused by the *prph* gene does not mean that the CRE does not regulate *prph*. As stated, the precursors of the cells that will express *prph*, are neural crest cells, and the CRE has been found to be open at stage 13 which is the stage when NC is specified. CRE 138 might just be one of the CREs needed for the *prph* expression pattern.

Further, there are a few genes located around CRE 138 which have not yet been characterised, so the CRE could be a regulator for one of those genes. Lastly, some TFBS related to NC have been identified within the sequence of CRE 138. This includes Sox2/4/10, but some lesser known TFs such as Nr5a1 and Nr5a2 involved in adrenal medulla development, and NC development and the pharyngeal arches respectfully (Nelms and Labosky, 2010). CRE 138 show some similarities with the already described expression of *prph*, and it may show an expression pattern associated with NC or neural expression. Therefore, it is possible that CRE 138 does regulate the *prph* gene, however, more investigation is needed, such as exploring the temporal expression of *prph* in more detail, but also exploring the function of the other genes located within 100 kb of the CRE.

4.3.8 CRE 180 show similarities to the expression of *fus*

Fus or *fused in Sarcoma* is part of the protein family FET. It regulates various aspects of RNA metabolism and processing, as well as micro-RNA biogenesis. *Fus* is a possible regulator of the early transcriptional machinery in *Xenopus* development, where it is among other necessary for proper mRNA splicing of key genes for regulation of *Xenopus* development (Bibonne et al., 2013, Dichmann and Harland, 2012). Low *fus* expression has been observed as early as gastrulation in the ectoderm and some in the mesoderm. During neurulation it is restricted to anterior and posterior neural regions and the epidermal ectoderm, where

transverse sections show expression specifically in the neural tube and the lateral plate mesoderm. There is no expression in the notochord and weak expression in the somatic mesoderm. At the tailbud stage there is expression in the brain, spinal cord, eye, otic vesicle, branchial arches, pronephros, proctodaeum, and the tailbud (Bibonne et al., 2013). For CRE 180, there is expression in the general neural plate at early neurulation (Figure 38A), which is explained by Bibonne (2013) with *fus* being expressed in the general ectoderm at the start before it concentrated along the neural tube from the anterior to the posterior as seen in Figure 38B. There could also be expression in the lateral plate mesoderm however, this can only be observed with sectioning. As seen in Figure 38D-D' tailbud embryos show expression in most of the structures observed by Bibonne (2013) such as the brain and spinal cord, eye and otic vesicle, branchial arches, possibly in the pronephros and the expression extends all the way to the tailbud. Other stage 30 embryos show some of these structures but not all (Figure 39A-D). This may have been described previously because the CRE may only be active at the region we have observed it to be open at stage 13. Although the expression in Figure 38D-D' is very strong suggesting that the GFP is not residual GFP but rather GFP being expressed at the time. Further, several TFBSs related to NC have been mapped to the sequence of CRE 180, such as Sox2/4/5/8/9, Dlx6, Zic1, and Pou2f1 which plays a role in the specification and differentiation of NC (Nelms and Labosky, 2010). Further, there are other genes within 100 kb to CRE 180 that have not yet been characterised. It may be that CRE 180 regulates one of these genes. The expression of the reporter gene caused by CRE 180 show similarities to the expression pattern of *fus*. Therefore, it is possible that CRE 180 regulates *fus* expression.

4.3.9 CRE 183 cannot be associated with any gene

CRE 183 is one of two CREs that were identified in proximity to the *wnt8a* gene and was the only one of the two that caused reproducible expression. However, *wnt8a* is not expressed in the NC. Even though several *wnts* such as *wnt1* and *wnt3a* are expressed in the neural tube, *wnt8* is expressed in the ventral and lateral mesoderm and the involuted mesoderm that lies under the prospective neural plate (Wu et al., 2003). The expression pattern observed with CRE 183 however is found in the neural plate and along the neural tube in the neurula embryo (Figure 40A-B). Further the expression in the tailbud embryos were quite variable with possible expression in the notochord or muscle being a common factor (Figure 40C, Figure 41). It is likely that if CRE 183 is a CRE it regulates a different gene. However, not many of the genes within 200 kb of the CRE are characterised and those that are not associated with NC development. Therefore, it is not possible to say what genes could be related to the expression observed by CRE 183. It is also possible that the expression observed may be a random integration even. Although the pattern seen at neurula stage was quite consistent other

expression patterns that were not consistent were also observed, be that not at a very high percentage (2.72%). Further, very few TFBS were mapped to the sequence of CRE 183. Some include GATA2/3/6 which relate to patterning of neurones rather than the development of NC (Nelms and Labosky, 2010). The expression caused by CRE 183 is a strange one that cannot now be related to any gene. Further investigation into other genes surrounding the CRE is required.

4.3.10 CRE 192 show a neural or neural crest expression

CRE 192 has been associated with *evx1* due to its proximity to the gene. *Evx1* is also called *xhox3* in *Xenopus* and is a homeobox gene implicated in the patterning of the anterior to posterior axis. It is one of the first genes to be expressed from the midblastula transition, and here it is expressed in a graded fashion along the axis in the mesoderm. This expression persists until early neurulation where early neurones start to express *evx1* at high levels. Later expression can be seen in the developing central nervous system, such as the hindbrain (Barro et al., 1994). The tailbud stages show expression in the tailbud as well as the eye, otic vesicle, hindbrain, and neural tube (Beck and Slack, 1998). Other than this, there is not much information of the expression pattern caused by *evx1*. Therefore, understanding it in relation to the expression pattern caused by CRE 192 is difficult. CRE 192 shows early expression at the edge of the neural plate in the neural folds as they move towards the dorsal midline (Figure 42A). It is difficult to know if there is any mesoderm expression without sectioning, however, the expression resembles other neural expressions seen with for example *snai2*. The late neurula has little expression posterior which contradict Beck and Slack (1998) who explains *evx1* as a key marker of the posterior tailbud. Further, there is strong expression in the whole anterior of the embryo which could be the precursors of the central nervous system (Figure 42B). The tailbud expression for CRE 192 share some similarities to *evx1.* There is expression in the brain and the neural tube as well as in the eye. However, there is also strong expression in the branchial arches (Figure 42C), and expression has been observed in the notochord. Seeing that there are some discrepancies, and the numbers are quite low, it could suggest that there is random integration occurring, although, no other expression patterns have been observed by this CRE, showing the consistency of the early expression. One other complication with *evx1* is that *evx1* has not previously been associated with the development of the NC. It may be that there is not enough research on its connections to the NC at this point of time, however taken together with the fact that *evx1* is not found in the RNA-seq it is possible that CRE 192 regulates another gene. Within 200 kb of CRE 192 there are several other targets that could explain the expression pattern seen by the CRE, these include h*oxa2* and *hoxa3* which have very similar expression patterns (Nelms and Labosky, 2010). *Hoxa3* is

first expressed in the early neurula in the developing spinal cord and migrating NC. It creates and anterior boundary at rhombomere 5. In the tailbud, expression is persistent in the spinal cord, but it also migrates into the third pharyngeal arch (Lee et al., 2013). However, even this expression pattern does not correspond to the one observed by CRE 192. Where there is clear expression in the anterior of the embryo during neurulation (Figure 42A-B). There is also anterior expression in the head in the tailbud (Figure 42C), however this might just be because the expression is slightly mosaic. Other than the expression pattern there are several TFBSs related to the NC found in the sequence of CRE 192. The UCSC genome browser predicts that TFs such as Pax3, Sox2/4, Msx1/2 and even Hoxa1 which also has a similar expression pattern to *hoxa2* and *hoxa3* (McNulty et al., 2005). The expression pattern caused by CRE 192 is confusing. However, it appears to show neural and possible neural crest expression.

Chapter 5: General discussion

Specific regulation of gene expression is vital for proper spatial and temporal expression of genes during development. One way to cause this regulation is though *cis*-regulatory elements (CREs). In this investigation we have used ATAC-sequencing of *X. leavis* animal cap tissue induced to become NC, NE or ECT and identified open regions in the chromatin exclusive to the NC sample. These open regions have the potential to act as CREs. 20 of the open regions, annotated as CREs were chosen for further study by a reporter assay. Twelve of these CREs gave reoccurring reporter gene expression in a neural or neural crest pattern. The specific pattern for the twelve CREs have been discussed in detail in subchapter 4.2.2.

5.1 Low percentage of transgenic embryos

Making transgenic organisms is a challenging procedure which yielded low integration, and a low percentage of success. However, it is the most effective way to study expression throughout development in organisms such as *Xenopus* and Zebrafish as with the DNA integrated into the genome, there is a lower chance of mosaic organisms, leading to a more accurate representation of activity. There are several ways to randomly integrate DNA into a genome. In *Xenopus*, initially REMI was used where a piece of foreign DNA is integrated into the sperm genome, which is then injected into the egg to simulate fertilization. Even though this method results in the least mosaic organisms as the DNA is integrated into the genome even before fertilization, it also has one of the lowest percentages of success. 20-40% of the embryos will develop normally after cleavage, and as low as 5% or as high as all the 40% may then express the transgene. Grouped together with the fact that REMI requires a high skill level as the needle for the sperm transplantation is larger than the needle for other methods and therefore has a higher chance of killing the egg (Amaya and Kroll, 2010, Amaya and Kroll, 1999, Ogino et al., 2006b, Wheeler et al., 2000, Yergeau et al., 2010). The method preferred in Zebrafish transgenics is the Tol2 system which uses the Tol2 transposase to randomly integrate microinjected linearized DNA into the genome. This system, which also has been reported to function in *Xenopus*, has the highest success rate for creating transgenic organisms with a 50% success. A negative side of the Tol2 system is that it may have the latest integration of the DNA, meaning that there is a higher chance of mosaic animals, and the method is therefore more commonly used to create lines, not to study the F_0 generation (Hamlet et al., 2006, Suster et al., 2009, Yergeau et al., 2010). The I-SceI meganuclease method may be a good middle ground. Even though the I-SceI enzyme has been reported to be challenging to work with as it must be stored at -80°C. The enzyme linearizes the DNA that

is to be integrated before the digestion reaction is microinjected into the fertilized egg. As the enzyme does not need to be transcribed it can start to integrate the linearized DNA at once, therefore there is a lower chance of mosaic animals. The percentage of success is also higher than REMI with 30% success reported for *X. tropicalis,* and 20% success reported for *X. laevis* (Ishibashi et al., 2012, Ogino et al., 2006a, Ogino et al., 2006b, Yergeau et al., 2010). For this reason, I-SceI is the method that has been used in this investigation.

However, the success rate for the I-SceI method in the current project is far below the reported success by Ogino (2006) (Ogino et al., 2006a). In this project, when grouping *X. laevis and X. tropicalis* embryos together we have seen between 2 – 13% transgenic organisms. This percentage was also obtained after modifying the protocol made my Ogino (2006) as the concentration of the plasmid they injected did not cause transgenic expression for us. However, altering the concentration of the plasmid, but not altering the concentration of the enzyme, might have unbalanced the reaction. This might mean that some of the plasmid has not been digested, and this could have caused an additional spotty/mosaic expression to the transgenic expression due to the limited diffusion and therefore unequal inheritance by daughter cells. This spotty expression was observed by Ogino (2006) in a small number of cells, and it would present itself as stripes in tailbud embryos (Alkobtawi et al., 2018, Ogino et al., 2006a). This could also have caused a lower fluorescence and embryos that were transgenic might have been miscounted. Similar patterns to Ogino (2006) may be seen in among others tailbud embryos injected with CRE 180 (Figure 38D', Figure 39). However, these stripes may also be muscle. Further, the low percentage may be caused by our inexperience with the procedure to begin with, lowering the chance of success. One way that this could have been further investigated is to co-inject the plasmid with another fluorescent mRNA such as Cherry. By doing this we would know the exact number of embryos that were injected with the digestion, and this could end up increasing the percentage. Another change to improve this could be to alter the vector by adding a crystalline promoter controlling another fluorescent protein. The crystalline promoter is lens specific and would confirm whether the linearized DNA has integrated into the genome (Huang et al., 1999). Lastly, the number of transgenic embryos may be affected by the frog quality. During the project, some frogs were sick, which could have lowered the quality of the eggs, and therefore their survival or ability to integrate the linearized plasmid (Ogino et al., 2006a). Even though the percentage of transgenic embryos may be considered low, the results may still be valid. The greater percentage of transgenic embryos for the CREs that cause transgenic expression is consistent. For that reason, further investigation into the twelve CREs is needed.

5.2 What is acceptable variable expression?

As well as the percentages being low, a further consideration to assess when looking at the results from this investigation is that within the low percentages the expression seen by the CREs are slightly varied with each construct. The strength of the expression, the width and length of expression as well as the continuity of expression is slightly varied during neurula stages. Then during tailbud stages the variation of expression can be even greater. The variation within tailbud stage expression is to be expected as the CREs were identified as accessible during neural stage 13. It is very possible that the CREs are not normally accessible or active during tailbud stages and therefore in many cases it is possible that the expression that can be seen in tailbuds are residual GFP from the neurula stage (Li et al., 2018). Further, in the above subchapter, it was discussed that there might be mosaic expression due to undigested plasmid. As well as changing the percentage of success this might have caused variable expression patterns. Some tailbud embryos from the current investigation have the stripy expression seen in Ogino (2006). An example of this can be seen in Figure 38D'. Here the pattern has been annotated as muscle, however, it is possible after comparing to Ogino (2006) that it is mosaic expression (Ogino et al., 2006b). Because of this additional mosaic pattern, embryos may still have been classed as having the same expression pattern, even though it is variable.

It is also important to note that the integration of the reporter construct is random, and this may also cause a wide range of different expression patterns. As explained in subchapter 1.4 the genome is organised into a chromatin state that is highly regulated. By randomly integrating the new DNA may be influenced by positional effects. This is where the DNA is not in its normal chromosomal context, and therefore it may be influenced by other CREs. In some cases, the expression may be seen as very low or not present at all. In other cases, the expression patterns might be different from what is expected as the reporter construct might be under control of a stronger CRE near the integration site (Krijger and de Laat, 2016, Liu, 2013). For most of the CREs there were some expression patterns that were not reoccurring but observed and counted. The percentage of these embryos were always lower that the reoccurring pattern, and therefore the reoccurring pattern was believed to be real. Further, also as explained in subchapter 1.5.1.4 when looking specifically at enhancers it was explained that enhancers may not function as one unit. Several enhancers may act together to cause expression of a gene. In this investigation we have looked at CREs on their own and this might not fully represent their activity thereby causing a different expression from what we would expect or a partial expression of the known expression for the closest gene (Krijger and de Laat, 2016, Panigrahi and O'Malley, 2021, Zeitlinger, 2020). This means that the CREs that cause transgenic expression might not cause their native expression, and some of the

CREs that gave no transgenic expression might still be enhancers, but they cannot work outside their specific context.

5.3 Conclusion about CREs in the project

This project started with investigating 20 CREs in the lab to assess whether they could cause expression of a reporter gene. Twelve of the CREs showed reporter expression and of these we can speculate about what gene they regulate. We cannot say for certain if the CREs regulate these genes as further investigation for all the CREs would have to take place. The predicted targets for the twelve CREs can be seen in Table 5. However, it is important to note that around most of the CREs there are genes annotated that have not been characterised. It is possible that these genes are involved with NC development and could therefore also be targets.

In general, the pattern that all the CREs show are what we refer to as neural. Expression of the reporter gene is initiated during late gastrula or early neurula either at the edge of the neural plate or in the neural plate. Then as the embryo develops, this expression concentrates along the neural folds and move with the folds towards the dorsal midline. In the tailbud the expression is very varied for the different CREs, however some common expression is in the neural tube and head, and many show expression in the eye.

Even though most of the CREs investigated in this project were associated with genes that have been linked to NC development, not all the CREs showed a specific NC pattern in this project. This might be due to the low percentage and high variability in clarity of the expression. It might be that with a more optimized protocol and a higher number of injected embryos that more of the CREs show a more consistent NC pattern. However, from the results gathered in this project it would be most interesting in relation to the NC to continue the investigation into CRE 22, 26, 33 and 34, 62 and 63 as wells as 192 due to their potential NC specificity (Table 5).

Table 5: Summary of predicted target genes for the twelve CREs in that cause reporter expression.

CRE	Closest gene	Possible gene target	NC positive
22	*sox9*	*sox9*	✓
26	*pdgfra*	*pdgfra*	✓
31	*hes1*	*hes1 or sox2*	?
33	*snai2*	*snai2*	?
34	*snai2*	*snai2*	✓

52	*irx3*	*fto*	X
62	*pax3*	*pax3 or pax6*	✓
63	*pax3*	*pax3 or pax6*	✓
138	*prph*	*prph*	?
180	*fus*	*fus*	?
183	*wnt8a*	*None*	X
192	*evx1*	*hoxa2 or hoxa3*	✓/?

5.4 Future work on this project

The conclusion about the target genes of the CREs from this investigation is a vague one as by the research performed, we cannot make any definitive conclusions. Therefore, there are several experiments that would strengthen the conclusions made. First, building up the numbers of transgenic embryos is important, as the numbers are quite low. To emphasis if the numbers are correctly counted, sectioning of the embryos to see where dorsally the expression is would clarify what the CREs regulate. We could also perform WISH for GFP to clearly visualise where in the embryo the GFP is expressed. Further, we would see if the CREs are CREs, and what part of the sequence is important for CRE function. To do this we would mutate interesting predicted TFBSs identified by the bioinformatical analysis in the injected plasmid. The plasmid would be microinjected into embryos and the expression caused by the mutated CRE would be re-assessed to see if the level of expression is changed or lost. To further emphasise if the CRE is a CRE one could use ChIP-Seq to identify chromatin modifications associated with CREs in the CRE sequences (Taminato et al., 2016). Once there is enough reassurance that the CRE most likely is a CRE, and we have a prediction of the functional units of the CRE we can look at their significance in the embryo by doing CRISPR/Cas9 knockouts to make a loss-of-function mutation. We can then assess if the loss of the CRE influences the development of the NC or the expression of the predicted target gene, or other genes (C.Tobias et al., 2021, Mok et al., 2021). Alternatively we can use the CRISPR/Cas9::KRAB method where a repressor domain such as Kruppel-associated box (KRAB) repressor is fused to the Cas9 causing inhibition of activity thereby not altering the genomic DNA, but still achieving the wanted effect which is to identify whether the CRE regulates a specific gene. This method was used by both Williams and Mok to identify important regions in their CREs in chick (Mok et al., 2021, Williams et al., 2019, Parsi et al., 2017).

5.5 Considerations with *cis*-regulatory reporter assays

When attempting to identify CREs such as enhancers, there are several things to consider when looking at the results from the reporter assay. It is highly likely that most of the chosen regions will display a negative result, meaning there will be no reporter gene expression, or rather, no specific reporter gene expression. However, this negative result might not represent the region we are testing. It may be that the CRE cannot cause expression of the reporter gene by itself, and in its natural environment it may require other CREs to also be present to cause expression (Long et al., 2016, Nelson and Wardle, 2013). Further, the regulatory region may be a repressor rather than a positive regulator, also causing a negative result. To test whether the CRE is a repressor one can do an enhancer-blocker assay as done by Royo and colleagues where they inserted an insulator element between a known enhancer and a minimal promoter (Royo et al., 2011). Another challenge with reporter assays can be that the CRE cannot cause expression of the reporter gene when interacting with a minimal promoter, and it might rather require the cognate promoter of the gene it controls to be able to cause expression. To test this one can use a plasmid with a minimal promoter and another plasmid with the cognate promoter to identify whether this is the case (Nelson and Wardle, 2013). Lastly, we do not know what the CRE is. As explored earlier we can see if it is a repressor by an enhancer-blocker assay, however, if it is a positive regulator, we might be inclined to call it an enhancer. Although, if the positive CRE is found close to the gene it might be a proximal promoter rather than an enhancer. Proximal promoter elements can be a few hundred base pairs from the core promoter elements (100-200 bp), and in some cases they can cause cell specific regulation (Lodish, 2000, Maston et al., 2006). Some of the CREs identified in this project, including CRE 52 and 62 are only 500 bp away from the promoter of the gene closest to them, therefore, it is possible that they are proximal promoters although they are further away that what proximal promoters are suggested to be.

Chapter 6: References

ALKOBTAWI, M., RAY, H., BARRIGA, E. H., MORENO, M., KERNEY, R., MONSORO-BURQ, A.-H., SAINT-JEANNET, J.-P. & MAYOR, R. 2018. Characterization of Pax3 and Sox10 transgenic Xenopus laevis embryos as tools to study neural crest development. *Developmental Biology,* 444**,** S202-S208.

AMAYA, E. & KROLL, K. 2010. Production of transgenic Xenopus laevis by restriction enzyme mediated integration and nuclear transplantation. *J Vis Exp*.

AMAYA, E. & KROLL, K. L. 1999. A Method for Generating Transgenic Frog Embryos. *In:* SHARPE, P. T. & MASON, I. (eds.) *Molecular Embryology: Methods and Protocols.* Totowa, NJ: Humana Press.

AMIEL, J. & LYONNET, S. 2001. Hirschsprung disease, associated syndromes, and genetics: a review. *Journal of Medical Genetics,* 38**,** 729-739.

AOKI, Y., SAINT-GERMAIN, N., GYDA, M., MAGNER-FINK, E., LEE, Y.-H., CREDIDIO, C. & SAINT-JEANNET, J.-P. 2003. Sox10 regulates the development of neural crest-derived melanocytes in Xenopus. *Developmental Biology,* 259**,** 19-33.

ARNOSTI, D. N. & KULKARNI, M. M. 2005. Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards? *Journal of Cellular Biochemistry,* 94**,** 890-898.

BAE, C.-J., PARK, B.-Y., LEE, Y.-H., TOBIAS, J. W., HONG, C.-S. & SAINT-JEANNET, J.-P. 2014. Identification of Pax3 and Zic1 targets in the developing neural crest. *Developmental Biology,* 386**,** 473-483.

BANERJI, J., RUSCONI, S. & SCHAFFNER, W. 1981. Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. *Cell,* 27**,** 299-308.

BANG, A. G., PAPALOPULU, N., KINTNER, C. & GOULDING, M. D. 1997. Expression of Pax-3 is initiated in the early neural plate by posteriorizing signals produced by the organizer and by posterior non-axial mesoderm. *Development,* 124**,** 2075-2085.

BARRO, O., JOLY, C., CONDAMINE, H. & BOULEKBACHE, H. 1994. Widespread expression of the Xenopus homeobox gene Xhox3 in zebrafish eggs causes a disruption of the anterior-posterior axis. *Int J Dev Biol,* 38**,** 613-22.

BECK, C. W. & SLACK, J. M. W. 1998. Analysis of the developing Xenopus tail bud reveals separate phases of gene expression during determination and outgrowth. *Mechanisms of Development,* 72**,** 41-52.

BERG, J. M., TYMOCZKO, J. L., GATTO, G. J. & STRYER, L. 2015. *Biochemistry*.

BETANCUR, P., BRONNER-FRASER, M. & SAUKA-SPENGLER, T. 2010. Genomic code for Sox10 activation reveals a key regulatory enhancer for cranial neural crest. *Proceedings of the National Academy of Sciences of the United States of America,* 107**,** 3570-3575.

BIBONNE, A., NÉANT, I., BATUT, J., LECLERC, C., MOREAU, M. & GILBERT, T. 2013. Three calcium-sensitive genes, fus, brd3 and wdr5, are highly expressed in neural and renal territories during amphibian development. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research,* 1833**,** 1665-1671.

BOLT, C. C. & DUBOULE, D. 2020. The regulatory landscapes of developmental genes. *Development,* 147.

BRASSET, E. & VAURY, C. 2005. Insulators are fundamental components of the eukaryotic genomes. *Heredity,* 94**,** 571-576.

BRIGHT, A. R. & VEENSTRA, G. J. C. 2019. Assay for Transposase-Accessible Chromatin-Sequencing Using Xenopus Embryos. *Cold Spring Harb Protoc,* 2019.

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. *Nat Methods,* 10**,** 1213-8.

BUISSON, I., LE BOUFFANT, R., FUTEL, M., RIOU, J. F. & UMBHAUER, M. 2015. Pax8 and Pax2 are specifically required at different steps of Xenopus pronephros development. *Dev Biol,* 397**,** 175-90.

C.TOBIAS, I., E.ABATTI, L., D.MOORTHY, S., SHANELLEMULLANY, TIEGHTAYLOR, NAWRAHKHADER, A.FILICE, M. & A.MITCHELL, J. 2021. Transcriptional enhancers: from prediction to functional assessment on a genome-wide scale. *Genome,* 64**,** 426-448.

CANNATELLA, D. C. & DE SÁ, R. O. 1993. Xenopus Laevis as a Model Organism. *Systematic Biology,* 42**,** 476-507.

CARNINCI, P., KASUKAWA, T., KATAYAMA, S., GOUGH, J., FRITH, M. C., MAEDA, N., OYAMA, R., RAVASI, T., LENHARD, B., WELLS, C., KODZIUS, R., SHIMOKAWA, K., BAJIC, V. B., BRENNER, S. E., BATALOV, S., FORREST, A. R. R., ZAVOLAN, M., DAVIS, M. J., WILMING, L. G., AIDINIS, V., ALLEN, J. E., AMBESI-IMPIOMBATO, A., APWEILER, R., ATURALIYA, R. N., BAILEY, T. L., BANSAL, M., BAXTER, L., BEISEL, K. W., BERSANO, T., BONO, H., CHALK, A. M., CHIU, K. P., CHOUDHARY, V., CHRISTOFFELS, A., CLUTTERBUCK, D. R., CROWE, M. L., DALLA, E., DALRYMPLE, B. P., DE BONO, B., DELLA GATTA, G., DI BERNARDO, D., DOWN, T., ENGSTROM, P., FAGIOLINI, M., FAULKNER, G., FLETCHER, C. F., FUKUSHIMA, T., FURUNO, M., FUTAKI, S., GARIBOLDI, M., GEORGII-HEMMING, P., GINGERAS, T. R., GOJOBORI, T., GREEN, R. E., GUSTINCICH, S., HARBERS, M., HAYASHI, Y., HENSCH, T. K., HIROKAWA, N., HILL, D., HUMINIECKI, L., IACONO, M., IKEO, K., IWAMA, A., ISHIKAWA, T., JAKT, M., KANAPIN, A., KATOH, M., KAWASAWA, Y., KELSO, J., KITAMURA, H., KITANO, H., KOLLIAS, G., KRISHNAN, S. P. T., KRUGER, A., KUMMERFELD, S. K., KUROCHKIN, I. V., LAREAU, L. F., LAZAREVIC, D., LIPOVICH, L., LIU, J., LIUNI, S., MCWILLIAM, S., BABU, M. M., MADERA, M., MARCHIONNI, L., MATSUDA, H., MATSUZAWA, S., MIKI, H., MIGNONE, F., MIYAKE, S., MORRIS, K., MOTTAGUI-TABAR, S., MULDER, N., NAKANO, N., NAKAUCHI, H., NG, P., NILSSON, R., NISHIGUCHI, S., NISHIKAWA, S., et al. 2005. The Transcriptional Landscape of the Mammalian Genome. *Science,* 309**,** 1559-1563.

COHEN-TAYAR, Y., COHEN, H., MITIAGIN, Y., ABRAVANEL, Z., LEVY, C., IDELSON, M., REUBINOFF, B., ITZKOVITZ, S., RAVIV, S., KAESTNER, K. H., BLINDER, P., ELKON, R. & ASHERY-PADAN, R. 2018. Pax6 regulation of Sox9 in the mouse retinal pigmented epithelium controls its timely differentiation and choroid vasculature development. *Development,* 145.

CUNNINGHAM, R. L., KRAMER, E. T., DEGEORGIA, S. K., GODOY, P. M., ZAROV, A. P., SENEVIRATNE, S., GRIGURA, V. & KAUFMAN, C. K. 2021. Functional in vivo characterization of sox10 enhancers in neural crest and melanoma development. *Communications Biology,* 4**,** 695.

DAI, J., KUANG, Y., FANG, B., GONG, H., LU, S., MOU, Z., SUN, H., DONG, Y., LU, J., ZHANG, W., ZHANG, J., WANG, Z., WANG, X. & SHEN, G. 2013. The effect of overexpression of Dlx2 on the migration, proliferation and osteogenic differentiation of cranial neural crest stem cells. *Biomaterials,* 34**,** 1898-1910.

DEGENHARDT, K. R., MILEWSKI, R. C., PADMANABHAN, A., MILLER, M., SINGH, M. K., LANG, D., ENGLEKA, K. A., WU, M., LI, J., ZHOU, D., ANTONUCCI, N., LI, L. & EPSTEIN, J. A. 2010a. Distinct enhancers at the Pax3 locus can function redundantly to regulate neural tube and neural crest expressions. *Dev Biol,* 339**,** 519-27.

DEGENHARDT, K. R., MILEWSKI, R. C., PADMANABHAN, A., MILLER, M., SINGH, M. K., LANG, D., ENGLEKA, K. A., WU, M., LI, J., ZHOU, D., ANTONUCCI, N., LI, L. & EPSTEIN, J. A. 2010b. Distinct enhancers at the Pax3 locus can function redundantly to regulate neural tube and neural crest expressions. *Developmental biology,* 339**,** 519-527.

DICHMANN, D. S. & HARLAND, R. M. 2012. fus/TLS orchestrates splicing of developmental regulators during gastrulation. *Genes Dev,* 26**,** 1351-63.

DONI JAYAVELU, N., JAJODIA, A., MISHRA, A. & HAWKINS, R. D. 2020. Candidate silencer elements for the human and mouse genomes. *Nature Communications,* 11**,** 1061.

DUPONT, J. C. 2018. Historical perspective on neuroembryology: Wilhelm His and his contemporaries. *Genesis,* 56**,** e23218.

ELGAR, G. & VAVOURI, T. 2008. Tuning in to the signals: noncoding sequence conservation in vertebrate genomes. *Trends in genetics : TIG,* 24**,** 344-352.
ERIKSSON, K. S., ZHANG, S., LIN, L., LARIVIÈRE, R. C., JULIEN, J.-P. & MIGNOT, E. 2008. The type III neurofilament peripherin is expressed in the tuberomammillary neurons of the mouse. *BMC Neuroscience,* 9**,** 26.
ESMAEILI, M., BLYTHE, S. A., TOBIAS, J. W., ZHANG, K., YANG, J. & KLEIN, P. S. 2020. Chromatin accessibility and histone acetylation in the regulation of competence in early development. *Dev Biol,* 462**,** 20-35.
FUJIMI, T. J., MIKOSHIBA, K. & ARUGA, J. 2006. Xenopus Zic4: Conservation and diversification of expression profiles and protein function among the Xenopus Zic family. *Developmental Dynamics,* 235**,** 3379-3386.
FUKAYA, T. & LEVINE, M. 2017. Transvection. *Current biology : CB,* 27**,** R1047-R1049.
FURLONG, E. E. M. & LEVINE, M. 2018. Developmental enhancers and chromosome topology. *Science,* 361**,** 1341-1345.
GAMMILL, L. S. & BRONNER-FRASER, M. 2003. Neural crest specification: migrating into genomics. *Nat Rev Neurosci,* 4**,** 795-805.
GASPERINI, M., TOME, J. M. & SHENDURE, J. 2020. Towards a comprehensive catalogue of validated and target-linked human enhancers. *Nature reviews. Genetics,* 21**,** 292-310.
GEHRING, W. J. & IKEO, K. 1999. Pax 6: mastering eye morphogenesis and eye evolution. *Trends in Genetics,* 15**,** 371-377.
GILBERT, S. F. 2014. *Developmental biology,* Sunderland, MA, USA, Sinauer Associates, Inc., Publishers.
GRAHAM, A. & SHIMELD, S. M. 2013. The origin and evolution of the ectodermal placodes. *Journal of anatomy,* 222**,** 32-40.
GREEN, S. A., SIMOES-COSTA, M. & BRONNER, M. E. 2015. Evolution of vertebrates as viewed from the crest. *Nature,* 520**,** 474-482.
GREEN, Y. S. & VETTER, M. L. 2011. EBF factors drive expression of multiple classes of target genes governing neuronal development. *Neural Development,* 6**,** 19.
HABERLE, V. & STARK, A. 2018. Eukaryotic core promoters and the functional basis of transcription initiation. *Nature reviews. Molecular cell biology,* 19**,** 621-637.
HAMLET, M. R., YERGEAU, D. A., KULIYEV, E., TAKEDA, M., TAIRA, M., KAWAKAMI, K. & MEAD, P. E. 2006. Tol2 transposon-mediated transgenesis in Xenopus tropicalis. *Genesis,* 44**,** 438-45.
HARDWICK, L. J. A. & PHILPOTT, A. 2019. N-terminal phosphorylation of xHes1 controls inhibition of primary neurogenesis in Xenopus. *Biochemical and Biophysical Research Communications,* 509**,** 557-563.
HEINTZMAN, N. D., HON, G. C., HAWKINS, R. D., KHERADPOUR, P., STARK, A., HARP, L. F., YE, Z., LEE, L. K., STUART, R. K., CHING, C. W., CHING, K. A., ANTOSIEWICZ-BOURGET, J. E., LIU, H., ZHANG, X., GREEN, R. D., LOBANENKOV, V. V., STEWART, R., THOMSON, J. A., CRAWFORD, G. E., KELLIS, M. & REN, B. 2009. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature,* 459**,** 108-112.
HIRSCH, N. & HARRIS, W. A. 1997. Xenopus Pax-6 and retinal development. *Journal of Neurobiology,* 32**,** 45-61.
HONG, C.-S. & SAINT-JEANNET, J.-P. 2017. Znf703, a novel target of Pax3 and Zic1, regulates hindbrain and neural crest development in Xenopus. *Genesis (New York, N.Y. : 2000),* 55**,** 10.1002/dvg.23082.
HOPPLER, S. & WHEELER, G. N. 2015. It's about time for neural crest. *Science,* 348**,** 1316.
HUANG, H., MARSH-ARMSTRONG, N. & BROWN, D. D. 1999. Metamorphosis is inhibited in transgenic Xenopus laevis tadpoles that overexpress type III deiodinase. *Proceedings of the National Academy of Sciences,* 96**,** 962-967.
IBRAHIM, D. M. & MUNDLOS, S. 2020. Three-dimensional chromatin in disease: What holds us together and what drives us apart? *Curr Opin Cell Biol,* 64**,** 1-9.

ILLES, J. C., WINTERBOTTOM, E. & ISAACS, H. V. 2009. Cloning and expression analysis of the anterior parahox genes, Gsh1 and Gsh2 from Xenopus tropicalis. *Developmental Dynamics,* 238**,** 194-203.

ING-SIMMONS, E., VAID, R., BING, X. Y., LEVINE, M., MANNERVIK, M. & VAQUERIZAS, J. M. 2021. Independence of chromatin conformation and gene regulation during Drosophila dorsoventral patterning. *Nature Genetics,* 53**,** 487-499.

ISHIBASHI, S., LOVE, N. R. & AMAYA, E. 2012. A Simple Method of Transgenesis Using I-Sce I Meganuclease in Xenopus. *In:* HOPPLER, S. & VIZE, P. D. (eds.) *Xenopus Protocols: Post-Genomic Approaches.* Totowa, NJ: Humana Press.

JACQUIER, A. & DUJON, B. 1985. An intron-encoded protein is active in a gene conversion process that spreads an intron into a mitochondrial gene. *Cell,* 41**,** 383-394.

JING, H., VAKOC, C. R., YING, L., MANDAT, S., WANG, H., ZHENG, X. & BLOBEL, G. A. 2008. Exchange of GATA Factors Mediates Transitions in Looped Chromatin Organization at a Developmentally Regulated Gene Locus. *Molecular Cell,* 29**,** 232-242.

JUNION, G., SPIVAKOV, M., GIRARDOT, C., BRAUN, M., GUSTAFSON, E. H., BIRNEY, E. & FURLONG, EILEEN E. M. 2012. A Transcription Factor Collective Defines Cardiac Cell Fate and Reflects Lineage History. *Cell,* 148**,** 473-486.

KABBANI, H. & RAGHUVEER, T. S. 2004. Craniosynostosis. *Am Fam Physician,* 69**,** 2863-70.

KARDONG, K. V. 2015. *Vertebrates : comparative anatomy, function, evolution,* New York, NY, McGraw-Hill Education.

KARIMI, K., FORTRIEDE, J. D., LOTAY, V. S., BURNS, K. A., WANG, D. Z., FISHER, M. E., PELLS, T. J., JAMES-ZORN, C., WANG, Y., PONFERRADA, V. G., CHU, S., CHATURVEDI, P., ZORN, A. M. & VIZE, P. D. 2018. Xenbase: a genomic, epigenomic and transcriptomic model organism database. *Nucleic acids research,* 46**,** D861-D868.

KARNUTA, J. M. & SCACHERI, P. C. 2018. Enhancers: bridging the gap between gene control and human disease. *Human Molecular Genetics,* 27**,** R219-R227.

KAWAKAMI, K. 2005. Transposon tools and methods in zebrafish. *Developmental Dynamics,* 234**,** 244-254.

KIMMEL, C. B., BALLARD, W. W., KIMMEL, S. R., ULLMANN, B. & SCHILLING, T. F. 1995. Stages of embryonic development of the zebrafish. *Dev Dyn,* 203**,** 253-310.

KLYMKOWSKY, M. W., ROSSI, C. C. & ARTINGER, K. B. 2010. Mechanisms driving neural crest induction and migration in the zebrafish and Xenopus laevis. *Cell adhesion & migration,* 4**,** 595-608.

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 & Chromatin,* 5**,** 1.

KRIJGER, P. H. & DE LAAT, W. 2016. Regulation of disease-associated gene expression in the 3D genome. *Nat Rev Mol Cell Biol,* 17**,** 771-782.

LABONNE, C. & BRONNER-FRASER, M. 1998. Neural crest induction in Xenopus: evidence for a two-signal model. *Development,* 125**,** 2403-2414.

LATCHMAN, D. S. Transcriptional Gene Regulation in Eukaryotes. *eLS.*

LAUGSCH, M., BARTUSEL, M., REHIMI, R., ALIRZAYEVA, H., KARAOLIDOU, A., CRISPATZU, G., ZENTIS, P., NIKOLIC, M., BLECKWEHL, T., KOLOVOS, P., VAN IJCKEN, W. F. J., ŠARIĆ, T., KOEHLER, K., FROMMOLT, P., LACHLAN, K., BAPTISTA, J. & RADA-IGLESIAS, A. 2019. Modeling the Pathological Long-Range Regulatory Effects of Human Structural Variation with Patient-Specific hiPSCs. *Cell Stem Cell,* 24**,** 736-752.e12.

LE DOUARIN, N. M. 1980. The ontogeny of the neural crest in avian embryo chimaeras. *Nature,* 286**,** 663-669.

LE DOUARIN, N. M. 2004. The avian embryo as a model to study the development of the neural crest: a long and still ongoing story. *Mechanisms of Development,* 121**,** 1089-1102.

LEE, Y.-H., WILLIAMS, A., HONG, C.-S., YOU, Y., SENOO, M. & SAINT-JEANNET, J.-P. 2013. Early development of the thymus in Xenopus laevis. *Developmental Dynamics,* 242**,** 164-178.

LETTICE, L. A., HEANEY, S. J. H., PURDIE, L. A., LI, L., DE BEER, P., OOSTRA, B. A., GOODE, D., ELGAR, G., HILL, R. E. & DE GRAAFF, E. 2003. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Human Molecular Genetics,* 12**,** 1725-1735.

LI, I. M. H., LIU, K., NEAL, A., CLEGG, P. D., DE VAL, S. & BOU-GHARIOS, G. 2018. Differential tissue specific, temporal and spatial expression patterns of the Aggrecan gene is modulated by independent enhancer elements. *Scientific Reports,* 8**,** 950.

LI, J., LIU, K. C., JIN, F., LU, M. M. & EPSTEIN, J. A. 1999. Transgenic rescue of congenital heart disease and spina bifida in Splotch mice. *Development,* 126**,** 2495-2503.

LI, J., PERFETTO, M., MATERNA, C., LI, R., THI TRAN, H., VLEMINCKX, K., DUNCAN, M. K. & WEI, S. 2019. A new transgenic reporter line reveals Wnt-dependent Snai2 re-expression and cranial neural crest differentiation in Xenopus. *Scientific Reports,* 9**,** 11191.

LI, L., SCHATTEMAN, G. C., OPPENHEIM, R. W., LEI, M., BOWEN-POPE, D. F. & HOUENOU, L. J. 1996. Altered development of spinal cord in the mouse mutant (Patch) lacking the PDGF receptor α-subunit gene. *Developmental Brain Research,* 96**,** 204-209.

LIU, C. 2013. Strategies for designing transgenic DNA constructs. *Methods Mol Biol,* 1027**,** 183-201.

LODISH, H. F. 2000. *Molecular cell biology,* New York, W.H. Freeman.

LONG, H. K., PRESCOTT, S. L. & WYSOCKA, J. 2016. Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. *Cell,* 167**,** 1170-1187.

MACZKOWIAK, F., MATÉOS, S., WANG, E., ROCHE, D., HARLAND, R. & MONSORO-BURQ, A. H. 2010. The Pax3 and Pax7 paralogs cooperate in neural and neural crest patterning using distinct molecular mechanisms, in Xenopus laevis embryos. *Developmental Biology,* 340**,** 381-396.

MANUEL, M. N., MI, D., MASON, J. O. & PRICE, D. J. 2015. Regulation of cerebral cortical neurogenesis by the Pax6 transcription factor. *Frontiers in Cellular Neuroscience,* 9.

MASTON, G. A., EVANS, S. K. & GREEN, M. R. 2006. Transcriptional regulatory elements in the human genome. *Annu Rev Genomics Hum Genet,* 7**,** 29-59.

MAURANO, M. T., HUMBERT, R., RYNES, E., THURMAN, R. E., HAUGEN, E., WANG, H., REYNOLDS, A. P., SANDSTROM, R., QU, H., BRODY, J., SHAFER, A., NERI, F., LEE, K., KUTYAVIN, T., STEHLING-SUN, S., JOHNSON, A. K., CANFIELD, T. K., GISTE, E., DIEGEL, M., BATES, D., HANSEN, R. S., NEPH, S., SABO, P. J., HEIMFELD, S., RAUBITSCHEK, A., ZIEGLER, S., COTSAPAS, C., SOTOODEHNIA, N., GLASS, I., SUNYAEV, S. R., KAUL, R. & STAMATOYANNOPOULOS, J. A. 2012. Systematic localization of common disease-associated variation in regulatory DNA. *Science (New York, N.Y.),* 337**,** 1190-1195.

MAYOR, R., MORGAN, R. & SARGENT, M. G. 1995. Induction of the prospective neural crest of Xenopus. *Development,* 121**,** 767-777.

MCNULTY, C. L., PERES, J. O. N., BARDINE, N., VAN DEN AKKER, W. M. R. & DURSTON, A. J. 2005. Knockdown of the complete Hox paralogous group 1 leads to dramatic hindbrain and neural crest defects. *Development,* 132**,** 2861-2871.

MELLOTT, D. O. & BURKE, R. D. 2008. Divergent roles for Eph and ephrin in avian cranial neural crest. *BMC developmental biology,* 8**,** 56-56.

MILET, C., MACZKOWIAK, F., ROCHE, D. D. & MONSORO-BURQ, A. H. 2013. Pax3 and Zic1 drive induction and differentiation of multipotent, migratory, and functional neural crest in Xenopus embryos. *Proceedings of the National Academy of Sciences,* 110**,** 5528-5533.

MIZUSEKI, K., KISHI, M., MATSUI, M., NAKANISHI, S. & SASAI, Y. 1998. Xenopus Zic-related-1 and Sox-2, two factors induced by chordin, have distinct activities in the initiation of neural induction. *Development,* 125**,** 579-587.

MOK, G. F., FOLKES, L., WELDON, S. A., MANIOU, E., MARTINEZ-HEREDIA, V., GODDEN, A. M., WILLIAMS, R. M., SAUKA-SPENGLER, T., WHEELER, G. N., MOXON, S. & MÜNSTERBERG, A. E. 2021. Characterising open chromatin in chick embryos identifies cis-regulatory elements

important for paraxial mesoderm formation and axis extension. *Nature Communications,* 12**,** 1157.

MONSORO-BURQ, A. H., WANG, E. & HARLAND, R. 2005. Msx1 and Pax3 cooperate to mediate FGF8 and WNT signals during Xenopus neural crest induction. *Dev Cell,* 8**,** 167-78.

MURAI, K., PHILPOTT, A. & JONES, P. H. 2011. Hes6 Is Required for the Neurogenic Activity of Neurogenin and NeuroD. *PLOS ONE,* 6**,** e27880.

NAKAGAWARA, A., LI, Y., IZUMI, H., MURAMORI, K., INADA, H. & NISHI, M. 2018. Neuroblastoma. *Jpn J Clin Oncol,* 48**,** 214-241.

NAKATA, K., KOYABU, Y., ARUGA, J. & MIKOSHIBA, K. 2000. A novel member of the Xenopus Zic family, Zic5, mediates neural crest development. *Mechanisms of Development,* 99**,** 83-91.

NATOLI, T. A., ELLSWORTH, M. K., WU, C., GROSS, K. W. & PRUITT, S. C. 1997. Positive and negative DNA sequence elements are required to establish the pattern of Pax3 expression. *Development,* 124**,** 617-626.

NELMS, B. L. & LABOSKY, P. A. 2010. Transcriptional Control of Neural Crest Development. *Colloquium Series on Developmental Biology,* 1**,** 1-227.

NELSON, A. C. & WARDLE, F. C. 2013. Conserved non-coding elements and cis regulation: actions speak louder than words. *Development,* 140**,** 1385-1395.

NIEUWKOOP, P. D. & FABER, J. 1994. *Normal Table of Xenopus Laevis (Daudin): A Systematical and Chronological Survey of the Development from the Fertilized Egg Till the End of Metamorphosis*, Garland Pub.

OGINO, H., MCCONNELL, W. B. & GRAINGER, R. M. 2006a. High-throughput transgenesis in Xenopus using I-SceI meganuclease. *Nat Protoc,* 1**,** 1703-10.

OGINO, H., MCCONNELL, W. B. & GRAINGER, R. M. 2006b. Highly efficient transgenesis in Xenopus tropicalis using I-SceI meganuclease. *Mech Dev,* 123**,** 103-13.

OSBORN, D. P. S., ROCCASECCA, R. M., MCMURRAY, F., HERNANDEZ-HERNANDEZ, V., MUKHERJEE, S., BARROSO, I., STEMPLE, D., COX, R., BEALES, P. L. & CHRISTOU-SAVINA, S. 2014. Loss of FTO antagonises Wnt signaling and leads to developmental defects associated with ciliopathies. *PloS one,* 9**,** e87662-e87662.

PANIGRAHI, A. & O'MALLEY, B. W. 2021. Mechanisms of enhancer action: the known and the unknown. *Genome Biology,* 22**,** 108.

PARSI, K. M., HENNESSY, E., KEARNS, N. & MAEHR, R. 2017. Using an Inducible CRISPR-dCas9-KRAB Effector System to Dissect Transcriptional Regulation in Human Embryonic Stem Cells. *In:* WAJAPEYEE, N. & GUPTA, R. (eds.) *Eukaryotic Transcriptional and Post-Transcriptional Gene Expression Regulation.* New York, NY: Springer New York.

PEI, Z., WANG, B., CHEN, G., NAGAO, M., NAKAFUKU, M. & CAMPBELL, K. 2011. Homeobox genes Gsx1 and Gsx2 differentially regulate telencephalic progenitor maturation. *Proceedings of the National Academy of Sciences,* 108**,** 1675-1680.

PENNACCHIO, L. A., BICKMORE, W., DEAN, A., NOBREGA, M. A. & BEJERANO, G. 2013. Enhancers: five essential questions. *Nature Reviews Genetics,* 14**,** 288-295.

PERINO, M. & VEENSTRA, G. J. 2016. Chromatin Control of Developmental Dynamics and Plasticity. *Dev Cell,* 38**,** 610-20.

PLA, P. & MONSORO-BURQ, A. H. 2018. The neural border: Induction, specification and maturation of the territory that generates neural crest cells. *Developmental Biology,* 444**,** S36-S46.

PLOUHINEC, J.-L., ROCHE, D. D., PEGORARO, C., FIGUEIREDO, A. L., MACZKOWIAK, F., BRUNET, L. J., MILET, C., VERT, J.-P., POLLET, N., HARLAND, R. M. & MONSORO-BURQ, A. H. 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**,** 461-472.

PRASAD, M. S., CHARNEY, R. M. & GARCIA-CASTRO, M. I. 2019. Specification and formation of the neural crest: Perspectives on lineage segregation. *Genesis,* 57**,** e23276.

RAIBLE, D. W., WOOD, A., HODSDON, W., HENION, P. D., WESTON, J. A. & EISEN, J. S. 1992. Segregation and early dispersal of neural crest cells in the embryonic zebrafish. *Dev Dyn,* 195**,** 29-42.

RIETHOVEN, J.-J. M. 2010. Regulatory Regions in DNA: Promoters, Enhancers, Silencers, and Insulators. *In:* LADUNGA, I. (ed.) *Computational Biology of Transcription Factor Binding.* Totowa, NJ: Humana Press.

RODRÍGUEZ-SEGUEL, E., ALARCÓN, P. & GÓMEZ-SKARMETA, J. L. 2009. The Xenopus Irx genes are essential for neural patterning and define the border between prethalamus and thalamus through mutual antagonism with the anterior repressors Fezf and Arx. *Developmental Biology,* 329**,** 258-268.

ROGERS, C. D., HARAFUJI, N., ARCHER, T., CUNNINGHAM, D. D. & CASEY, E. S. 2009. Xenopus Sox3 activates sox2 and geminin and indirectly represses Xvent2 expression to induce neural progenitor formation at the expense of non-neural ectodermal derivatives. *Mechanisms of development,* 126**,** 42-55.

ROTHSTEIN, M. & SIMOES-COSTA, M. 2020. Heterodimerization of TFAP2 pioneer factors drives epigenomic remodeling during neural crest specification. *Genome research,* 30**,** 35-48.

ROYO, J. L., HIDALGO, C., RONCERO, Y., SEDA, M. A., AKALIN, A., LENHARD, B., CASARES, F. & GÓMEZ-SKARMETA, J. L. 2011. Dissecting the transcriptional regulatory properties of human chromosome 16 highly conserved non-coding regions. *PloS one,* 6**,** e24824-e24824.

SÁNCHEZ-GAYA, V., MARINER-FAULÍ, M. & RADA-IGLESIAS, A. 2020. Rare or Overlooked? Structural Disruption of Regulatory Domains in Human Neurocristopathies. *Frontiers in genetics,* 11**,** 688-688.

SANTIAGO, L., DANIELS, G., WANG, D., DENG, F.-M. & LEE, P. 2017. Wnt signaling pathway protein LEF1 in cancer, as a biomarker for prognosis and a target for treatment. *American journal of cancer research,* 7**,** 1389-1406.

SATO, T., SASAI, N. & SASAI, Y. 2005. Neural crest determination by co-activation of Pax3 and Zic1 genes in Xenopus ectoderm. *Development,* 132**,** 2355-2363.

SAVARESE, F. & GROSSCHEDL, R. 2006. Blurring cis and trans in Gene Regulation. *Cell,* 126**,** 248-250.

SHARPE, C. R., PLUCK, A. & GURDON, J. B. 1989. XIF3, a Xenopus peripherin gene, requires an inductive signal for enhanced expression in anterior neural tissue. *Development,* 107**,** 701-14.

SHELLARD, A. & MAYOR, R. 2016. Chemotaxis during neural crest migration. *Semin Cell Dev Biol,* 55**,** 111-8.

SIKORSKA, N. & SEXTON, T. 2020. Defining Functionally Relevant Spatial Chromatin Domains: It is a TAD Complicated. *Journal of Molecular Biology,* 432**,** 653-664.

SIMÕES-COSTA, M. & BRONNER, M. E. 2015. Establishing neural crest identity: a gene regulatory recipe. *Development (Cambridge, England),* 142**,** 242-257.

SIMON, J. M., GIRESI, P. G., DAVIS, I. J. & LIEB, J. D. 2012. Using formaldehyde-assisted isolation of regulatory elements (FAIRE) to isolate active regulatory DNA. *Nature protocols,* 7**,** 256-267.

SMITH, C. L. & TALLQUIST, M. D. 2010. PDGF function in diverse neural crest cell populations. *Cell adhesion & migration,* 4**,** 561-566.

SMITH, R. P., TAHER, L., PATWARDHAN, R. P., KIM, M. J., INOUE, F., SHENDURE, J., OVCHARENKO, I. & AHITUV, N. 2013. Massively parallel decoding of mammalian regulatory sequences supports a flexible organizational model. *Nature genetics,* 45**,** 1021-1028.

SONG, L. & CRAWFORD, G. E. 2010. DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. *Cold Spring Harb Protoc,* 2010**,** pdb.prot5384.

SPOKONY, R. F., AOKI, Y., SAINT-GERMAIN, N., MAGNER-FINK, E. & SAINT-JEANNET, J.-P. 2002. The transcription factor Sox9 is required for cranial neural crest development in Xenopus. *Development,* 129**,** 421-432.

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 (Cambridge, England),* 136**,** 771-779.

SUDOU, N., GARCÉS-VÁSCONEZ, A., LÓPEZ-LATORRE, M. A., TAIRA, M. & DEL PINO, E. M. 2016. Transcription factors Mix1 and VegT, relocalization of vegt mRNA, and conserved

endoderm and dorsal specification in frogs. *Proceedings of the National Academy of Sciences,* 113**,** 5628-5633.

SUSTER, M. L., KIKUTA, H., URASAKI, A., ASAKAWA, K. & KAWAKAMI, K. 2009. Transgenesis in Zebrafish with the Tol2 Transposon System. *In:* CARTWRIGHT, E. J. (ed.) *Transgenesis Techniques: Principles and Protocols.* Totowa, NJ: Humana Press.

SZABO, A. & MAYOR, R. 2018. Mechanisms of Neural Crest Migration. *Annu Rev Genet,* 52**,** 43-63.

TAMINATO, T., YOKOTA, D., ARAKI, S., OVARA, H., YAMASU, K. & KAWAMURA, A. 2016. Enhancer activity-based identification of functional enhancers using zebrafish embryos. *Genomics,* 108**,** 102-107.

THANOS, D. & MANIATIS, T. 1995. Virus induction of human IFNβ gene expression requires the assembly of an enhanceosome. *Cell,* 83**,** 1091-1100.

THERMES, V., GRABHER, C., RISTORATORE, F., BOURRAT, F., CHOULIKA, A., WITTBRODT, J. & JOLY, J.-S. 2002. I-SceI meganuclease mediates highly efficient transgenesis in fish. *Mechanisms of Development,* 118**,** 91-98.

TSAI, A., SINGER, R. H. & CROCKER, J. 2018a. Transvection Goes Live—Visualizing Enhancer-Promoter Communication between Chromosomes. *Molecular Cell,* 70**,** 195-196.

TSAI, P.-F., DELL'ORSO, S., RODRIGUEZ, J., VIVANCO, K. O., KO, K.-D., JIANG, K., JUAN, A. H., SARSHAD, A. A., VIAN, L., TRAN, M., WANGSA, D., WANG, A. H., PEROVANOVIC, J., ANASTASAKIS, D., RALSTON, E., RIED, T., SUN, H.-W., HAFNER, M., LARSON, D. R. & SARTORELLI, V. 2018b. A Muscle-Specific Enhancer RNA Mediates Cohesin Recruitment and Regulates Transcription In trans. *Molecular Cell,* 71**,** 129-141.e8.

UMEHARA, Y., TOYAMA, S., TOMINAGA, M., MATSUDA, H., TAKAHASHI, N., KAMATA, Y., NIYONSABA, F., OGAWA, H. & TAKAMORI, K. 2020. Robust induction of neural crest cells to derive peripheral sensory neurons from human induced pluripotent stem cells. *Scientific Reports,* 10**,** 4360.

VALLIN, J., THURET, R., GIACOMELLO, E., FARALDO, M. M., THIERY, J. P. & BRODERS, F. 2001. Cloning and Characterization of ThreeXenopus Slug Promoters Reveal Direct Regulation by Lef/β-Catenin Signaling *. *Journal of Biological Chemistry,* 276**,** 30350-30358.

VEGA-LÓPEZ, G. A., BONANO, M., TRÍBULO, C., FERNÁNDEZ, J. P., AGÜERO, T. H. & AYBAR, M. J. 2015. Functional analysis of Hairy genes in Xenopus neural crest initial specification and cell migration. *Dev Dyn,* 244**,** 988-1013.

VEGA-LOPEZ, G. A., CERRIZUELA, S., TRIBULO, C. & AYBAR, M. J. 2018. Neurocristopathies: New insights 150 years after the neural crest discovery. *Dev Biol,* 444 Suppl 1**,** S110-S143.

VOCKLEY, C. M., MCDOWELL, I. C., D'IPPOLITO, A. M. & REDDY, T. E. 2017. A long-range flexible billboard model of gene activation. *Transcription,* 8**,** 261-267.

WADDINGTON, C. H. 2012. The epigenotype. 1942. *Int J Epidemiol,* 41**,** 10-3.

WAKAMATSU, Y. 2011. Mutual repression between Pax3 and Pax6 is involved in the positioning of ophthalmic trigeminal placode in avian embryo. *Development, Growth & Differentiation,* 53**,** 994-1003.

WEST, A. G., GASZNER, M. & FELSENFELD, G. 2002. Insulators: many functions, many mechanisms. *Genes Dev,* 16**,** 271-88.

WHEELER, G. N., HAMILTON, F. S. & HOPPLER, S. 2000. Inducible gene expression in transgenic Xenopus embryos. *Current Biology,* 10**,** 849-852.

WILLIAMS, R. M., CANDIDO-FERREIRA, I., REPAPI, E., GAVRIOUCHKINA, D., SENANAYAKE, U., LING, I. T. C., TELENIUS, J., TAYLOR, S., HUGHES, J. & SAUKA-SPENGLER, T. 2019. Reconstruction of the Global Neural Crest Gene Regulatory Network In Vivo. *Developmental Cell,* 51**,** 255-276.e7.

WODA, J. M., PASTAGIA, J., MERCOLA, M. & ARTINGER, K. B. 2003. Dlx proteins position the neural plate border and determine adjacent cell fates. *Development (Cambridge, England),* 130**,** 331-342.

WOLPERT, L. 2019. *Principles of development*, Oxford University Press.

WU, J., SAINT-JEANNET, J.-P. & KLEIN, P. S. 2003. Wnt–frizzled signaling in neural crest formation. *Trends in Neurosciences,* 26**,** 40-45.

YERGEAU, D. A., KELLEY, C. M., ZHU, H., KULIYEV, E. & MEAD, P. E. 2010. Transposon transgenesis in Xenopus. *Methods (San Diego, Calif.),* 51**,** 92-100.

YORK, J. R. & MCCAULEY, D. W. 2020. The origin and evolution of vertebrate neural crest cells. *Open Biology,* 10**,** 190285.

YORK, J. R., YUAN, T. & MCCAULEY, D. W. 2020. Evolutionary and Developmental Associations of Neural Crest and Placodes in the Vertebrate Head: Insights From Jawless Vertebrates. *Frontiers in Physiology,* 11.

ZEITLINGER, J. 2020. Seven myths of how transcription factors read the cis-regulatory code. *Current Opinion in Systems Biology,* 23**,** 22-31.

ZOVKIC, I. B. 2020. Epigenetics and memory: an expanded role for chromatin dynamics. *Curr Opin Neurobiol,* 67**,** 58-65.

Appendices

Appendix 1: PCR programs

1.1: Biomix TAQ polymerase PCR program

	Temp	Time	Repeats
Initial denaturation	94°C	2 min	1
Denaturation	94°C	30 sec	35
Annealing	56°C	30 sec	
Elongation	72°C	2 min	
Final extension	72°C	5 min	1
Hold	10°C	Forever	1

1.2: Phusion polymerase PCR program

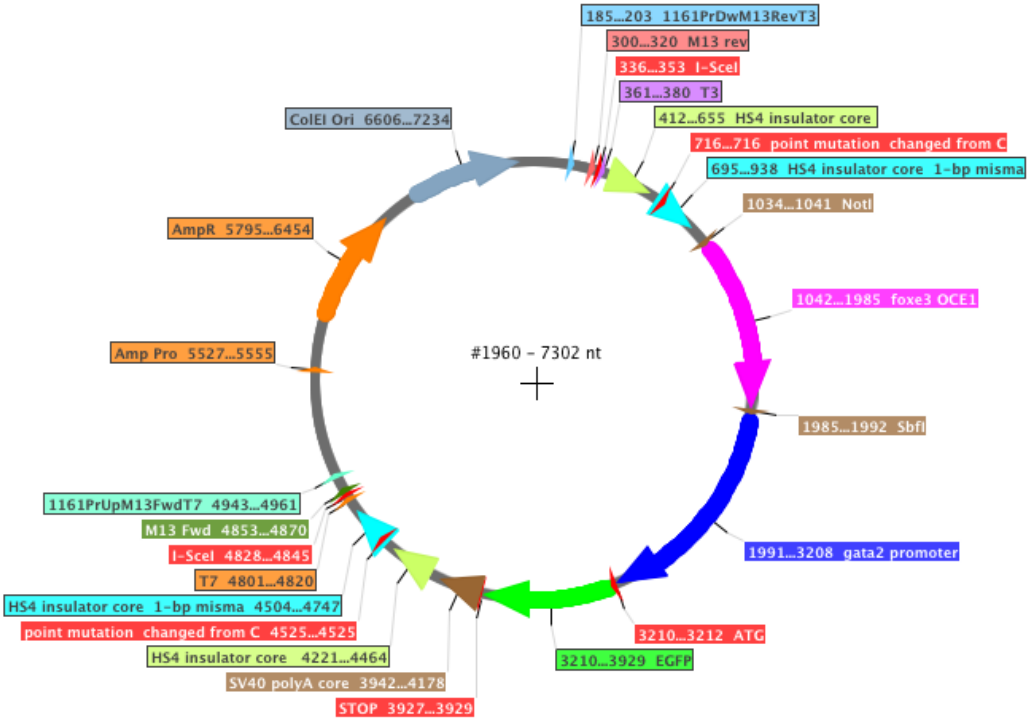
	Temp	Time	Repeats
Initial denaturation	98°C	30 sec	1
Denaturation	98°C	10 sec	Repeat 35 times
Annealing	58°C	30 sec	
Elongation	72°C	2 min	
Final extension	72°C	5 min	1
Hold	10°C	Forever	1

1.3: KAPA long range hot start ready mix PCR program

	Temp	Time	Repeats
Initial denaturation	94°C	3 min	1
Denaturation	94°C	15 sec	Repeat 10 times
Annealing	55°C	15 sec	
Elongation	68°C	3 min (1kb/min)	
Denaturation	94°C	15 sec	Repeat 25 times
Annealing	63°C	15 sec	
Elongation	68°C	3 min (1kb/min)	
Final extension	72°C	4 min	1

Hold	10°C	Forever	1

Appendix 2: 1960 plasmid map.



Appendix 3: Primers for cloned open chromatin regions.

Gene	Primer name	Primer (5'-3')	Fragment size	Fragment coordinates
sox9.S	22.1	TTGCGGCCGCTTTGTATTATCCCAAAGGGCT	455bp	chr9_10S:19,071,476-19,071,930
	22.2	TTCCTGCAGGAATGGTCTCTCTGTGGTTACA		
pdgfra.L	26.1	TTGCGGCCGCATTAGGTACTGGGGGTTACTG	331bp	chr1L:32,509,511-32,509,841
	26.2	TTCCTGCAGGGAGCAGGTTATTATTTGCCCT		
hes1.L	31.1	TTGCGGCCGCTGAGTATGCCCTCCTGCTTAT	731bp	chr5L:103,175,287-103,176,017
	31.2	TTCCTGCAGGACTAGTGTAGAGTGTGTCTCT		
snai2.S	33.1	TTGCGGCCGCCATGAGTTTTTCACTAAGCACA	465bp	chr6S:90,478,943-90,479,407
	33.2	TTCCTGCAGGGCAGTGCTATAAAACAGTCTG		
snai2.S	34.1	TTGCGGCCGCAGGTACTGTTTTACGTAGCAC	465bp	chr6S:90,451,590-90,452,054
	34.2	TTCCTGCAGGAACTAGAGAGTCTGCTCTTGG		
irx3.L	52.1	TTGCGGCCGCATAAGGGAACGTTGCTTTAGC	715bp	chr4L:31,936,991-31,937,705
	52.2	TTCCTGCAGGCGGTAGTACCATAGGATCATC		
pax3.L	62.1	TTGCGGCCGCATTTATTCTCCCCTCCCCTAC	1115bp	chr5L:123,001,476-123,002,590
	62.2	TTCCTGCAGGATTCTCATTCCAGTGATGAGG		
pax3.L	63.1	TTGCGGCCGCATTGCAATTGGTGTTTTCAGC	1595bp	chr5L:123,002,656-123,004,249
	63.2	TTCCTGCAGGTTCAATGGTGTCTGAAGCCAG		
pax3.L	64.1	TTGCGGCCGCTCTAATCAATCATGGGGCGAC	1781bp	chr5L:123,004,663-123,006,443
	64.2	TTCCTGCAGGGAACTGCTAATAGGTAACACG		
pax3.L	65.1	TTGCGGCCGCGGCCTACTATTGTGGCTTTGT	825bp	chr5L:123,006,762-123,007,586
	65.2	TTCCTGCAGGTGGTAGAGCAATAAGGCCCAT		
Vegt.L	101.1	TTGCGGCCGCAGGGAAACCAACACTTCCAAA	569bp	chr1L:152,604,333-152,604,901
	101.2	TTCCTGCAGGTGGAAGAACGAAGCAATGTGT		
Lysmd2.L	105.1	TTGCGGCCGCTAAACTGCAATGAAACACGAGG	398bp	chr3L:92,435,018-92,435,415
	105.2	TTCCTGCAGGACTGAACTAGGTATCAAATTGAATTTG		
Smad1.L	116.1	TTGCGGCCGCAACACAGGCGACTTTTCATT	350bp	chr1S:47,038,298-47,038,647

	116.2	TTCCTGCAGGGACAAAATGTACAAGGTTTTTCCT		
prph.L	138.1	TTGCGGCCGCCAGAACTTGGGTAACTGCTGA	404bp	chr2L:137,423,898-137,424,301
	138.2	TTCCTGCAGGATTATGGCCAATAGAAATGAGAAGG		
Pnhd.L	160.1	TTGCGGCCGCGACAGGTTTTAGGTGGTGTAT	1242bp	chr1L:101,954,238-101,955,479
	160.2	TTCCTGCAGGAGCAATATTAGGCACTCTGTT		
fus.L	180.1	TTGCGGCCGCGAAGCCTGAACAGTAGAACGA	793bp	Scaffold87:511,870-512,662
	180.2	TTCCTGCAGGCAAACAGAGCATCCCATAGGC		
wnt8a.L	182.1	TTGCGGCCGCCTGAGCTGAGATCTTGTTTCC	615bp	chr3L:24,525,434-24,526,048
	182.2	TTCCTGCAGGAGGACTGCATTAGTTATGGAC		
wnt8a.L	183.1	TTGCGGCCGCCTCTTTATGGGTGTGCAAACT	681bp	chr3L:24,526,208-24,526,888
	183.2	TTCCTGCAGGTGCCTTTCTTTTTGTGGACTG		
evx1.L	192.1	TTGCGGCCGCCAGGTCGAGCCTAATGTTGTA	821bp	chr6L:39,753,237-39,754,057
	192.2	TTCCTGCAGGATTGGTTATCAGCCCATTTCT		
Myf5.L	217.1	TTGCGGCCGCATAACACCTAGAGGCAAGACA	771bp	chr3L:45,225,434-45,226,204

Appendix 4: Sequencing of the cloned plasmid inserts.

Region	Sequence (5'-3')
22	GGCCGCTTTGTATTATCCCAAAGGGCTTCATATTAAAGAAGTACAACTTGAT ATATATGCTGCAAACACACAAGACCCCATGTTGGGTGTGTAACATTCTTAG AGCTAATCAAGTTTGTGGCATTATCAGAATATCTCCTGGGGTTATGTATAGT TAATGACTTGATATCTATAAGTCTCTTAAAGGTCGTACCCACAGGCGTTCTG ACTTGTAGTGGATTTTGACACAAGCTGAGACACCATTGGGTACAGAGCCTA AGGTTTACATCTCCCAGGGTTTTTGAATTGCTATTACTCACCCCATGTTGTG TCCTACAGGGTGATAATCCCCTGAAAACTCAGTGTTGGATGCAAATTGCAG CAAGCAGTCATGTTGGCCACTGTTACAATTATTGGTTACAATTATTGGTGAT TGTACCACAGTTCTGTAACCACAGAGAGACCATTCCTGCA
26	GGCCGCATTAGGTACTGGGGGTTACTGTACCCCTTTCCCTCTTGTTCTTGA CTATGTCAGTGCAAGTCACATAGAAGGAATACGGTATCCTGTTAGCATTCC TGTCAATAGACTGTTCATGTATTAGGCTCCAATTATTTCCTGCAGAAGCATA TGGAAGCAATAGCAGCGCTGGGCACTCACCCGGAGTTTGGATTGTGGTAT AGTGCAGAGGGAAAAAGTCACAGCTGAGGGCAAATAATAACCTGCTCCCT GCA
31	GGCCGCTGAGTATGCCCTCCTGCTTATATCACTTTTTTTGTAAGTCTTGAGA TCGTATTTACTTTGGTAGTACTCATATTGGACTTGTGCATATCCCCTTTTGA AGGTTCTAATAATACAACAATAAAGGTGTAGCTTTTGAGACTATGCAGAGTC CTGCTGTCCCTTTAAGTAAGAATGCTTGTGGGTATTTGTATGGAGTGTTGC AGCACCCTGTGTGTCTCACACCCATCTCAGAGGGGAAGCGGCTAAACAAA CAGTTATATGAGAGCTGCTGTTCCACCGCACACTTCTTGTGGATTCAGAAG AGATACTGAGAACTGAAAATTGCAAAGGCGCCTGCTCTCATTGACACCTCT CAAAGCAATTATCTCATTGCCTACAAAGGGCAGCGGAGTGTGTGGAAAACA CAGATAGTTACAGTAACACATAACTGGGACTTGAGAGGATTAATTCATTATC AGCTCACTGTTTAACAACCCCCGCCTCCAAAGTCACACACAATTGAAATAC CATGTTATAACAACAACATTGCAACACGCATGATGGTAGCATAAGAGATGC TGCTGCTAATGAGTTAATTAGAGAGTGCACTATGGGAATGGCCATTAAGAG ACACACTCTACACTAGTCCTGCA
33	GGCCGCCATGAGTTTTTCACTAAGCACAGTCAACAAATAGATCTTGTATTCT GCAGAATACCCATAAGGCAATAAAGAAGACACAGAAAACAACTTATTTATTA GAAGTAGGTCAACCTTTTTAGGGACTGATGGAATCATGCACAAGGAGAAGA CAATCACAGCACTGAGATCCTTCTTAATGTTATTTGTTCCACTTAATGGTCA CATGCTGACCAATTACTTCTCAGTCAGCAGGATTCAGTGGCATTATACCAC CTCTTATCCATGCTCTCTCTTTGTGCTTTAGCACTTCAGGCAAAATGTTTAG GCCAGGGGCATTTTTAATCAAAGACTTACTGGGACACTGTCTTTATGAGAC ATACATATATATATTGTTTGTTAGGTCAGTGGCTCTGATATCCACATGGGTT GGCTGTTCTGCATAGCCCACAGACTGTTTTATAGCACTGCCCTGCA
34	GGCCGCAGGTACTGTTTTACGTAGCACTGGGCGACTAATATCCCCCGAAT CTTACCATGTGCCACCAGCCTTAATTGTGACATGCCAATAAAATGACAAAC AGCAGAAACACTTCAAAGCTTAACACAGGTTGATAAAGTAGCCAAGATTGA TTAGCAAGACTGGTCAAGGAAGTGTGTATGGAATGAAAGGTACAATGTTTT ATCTGTACAGCACCTGCTTGTTTGTTTTGTACTCAGGCCATTTGTCTTAAAA GGCCTCCAAAAAGACAGAGACAAGAATCAATTGTCCCAAGTTCCACAAACA GCAGGGGTCTTGAAAGAACAAAGCTGATGTGGCAGTTCTGCGTGTCCTGC TTTTCATCAGAGTAGTATCAGGTTCCCTGCTGTGGACTGATAGGAGCCAAG AGCAGACTCTCTAGTTCCTGCA
52	GGCCGCATAAGGGAACGTTGCTTTAGCAGTTATTATTCTGGCTTGGCCTTT TCCGGAATCCTTCATCCCCTTGGACTGGTTCCCCTTTAGCCCTGGGACCGT ATGCAAATGCAGATTAGGGCTAATGACAGAGGGTTGGCAAATTCTCTTGCA TTTACATTGTATCTGCAGCGCTTCTTAACTCTTTTACAAAGGGGATATCGAA TTAGTCAAAACGATCAATGGTCGAAATAGAGGCTTCATAGTCGCCTAAAAA GGAACTTCAAAGAGTTTGAGGTCCCGGCTTTACATCAAAGCTGGGGGAAA

	GCAGGCAAATTGAAATCCTAACAAAGGAAAGCCCGCTGCCCTTTGGGCGC TGTGTAAATTTGTAGTGTAAGGGGAGAGACAATAGGAGGCACTTGGATTAA ACCCATTAACACCTTTTTGTCTTGGTGTTAAACTGTAATTAGAAAATTTACGA TCACTTTACTCTGCTCTCGTGACACAGGATACCTGCACTTGAAGCAGCAAT GTGCCGTGTGTATCTGGCAAACTGCGCGCAAAATAGCTCCTAAGGATGCA AGGTGTCTCCCAATGAGGATTAGGGGCAATTTCTTATATACATTTTTTATTA CCAGTAGATGATCCTATGGTACTACCGCCTGCA
62 – Does not include SbfI site as seq too long for sequencing	GGCCGCATTTATTTTCCCCTCCCCTACCCCAATTCTCTCCCTACACATACAC ACACACACACACACACACATACACACACACACACACACACACAGTAATGTAGC AGCCAGGCGCACTTGTGTGATTCCCAGCTTAGAGTGGGCGGACTGGGGG CAGCATTTGTTGAAAGGAATTTATAGCAGTGGCCAGAGAGCTGGATAAATA GAGGGAGAGCAGTTGAAGCAGTGGATGGAGGATCCCCAGCAGGAGATGA CGCCAGGAAGGCTATAAAGCCTGTCTCTCCTTTCCATAAGGTGCCATTGTC CGACCCCATTTATTTATAACATGGTTAATGCCAGGCAGCAGATACACTGGA GCAGATCCCTTTATTTGCAATGGGCCATTCAAGTGACTGTCACTCCTGCCA CTGAACTCATTCTCTGCCTTCATCAATCAAAGGGATGGGACACAGGCAGCT TTGGCCTTTGAAATAAAAGAAACAATGACTTCTGTCTTTAACCCTTTAGGCA GGGCACGGAGACCCCCGCTACTTGCAGAGCTGCGCCTTATCCGGCACTCA AGGGGTCAAATCCCGCTGTTCCCTCTACAAAATCAAGAGAGGGGAAAAAC ACAAACAACCCCCCCATTTGAATCACAATTTTTCCTCCCCTTGTCACATCTA CAAAGCGACTAGGGTCAGGGATCCCCCCTCCGTTTTCTTCATTATCCTAGT TTGGGAAGCCTATGTGTTACTGCTGGCTACAGGCTTGATAAAGGGCCCGA AACGTTGCCTTGTTTGTGAAATGCTTATGGGATAAATAAATCACCTATTTTC ACGGAGCAGTTATTTTTTGNNNNGCTTCTGGATTTTGACTTTTACTGCTGAG TACAGGGAATATCAAATTAAATATCACCCAGGATCGCGCTGCTCATTTTG
63 - Does not include NotI site	CGTGAAACAAAACTGCTGGCTTCGACACCATTGAACCTGCAGGTTCAATGG TGTCTGAAGCCAGTTCCCCCTCAGCCGACAGCACTTACCATTAGAGGGGC AGCAAAGTCAGCCACTGCGCTGCCTGTAGCAGCACAGACAGATAGACTAT AGAACCAATAAAAGAGTTCAAACCACTATCTAAAATGGACATTTAAACTGAG AGCTATTACCCACGGGAGCTGAAAACACCAATTGCAATGCGGCCGCATTN NNNNTGGTGTTTTCAGCCTCCAGTTACTGGACTTTGTTGTAGGATCCCAGT GTTTTAGTTTGAGGCGCACTGAGTTGCAGCAAATAATTCTCACCCATGTAA GTGTATCCAGTGCCAAGCGGCCCCCTCGTGTGAAAGTTTCGTTTAAAAAGC TACGGAAAAGCAAATAGTAGCTGGAAATGATAGATGTGTGAGGGTGGTGA CAGGGCATGGAGAATTGGGGACTAAGCTTAAAACTGCTCCCTACAGAATG GAGACGTGTCGCCCAGTAGTCAGTCTCTCCCTCCTCTCATAAATCTCAAGA TTTCAACGTATTCTCCTTGTGCCACAAGTTACACCACTAAGGTACTTAGTTA TTTTCATGGAATCTGGGAGCAATGCCCTATGGATGAAGTTTCTTGTGCTTG AGTCCATGGGTGCCCAGTTGGTTTCAGTTCTCCTTCCAATGGCAGAGCCA GGGGGATACAAACCTCTCCAGCCAAGTTCCAGTTTTTAGAGTACAACGGG NTTTTTATGAGATAAACCATGAGATCAGTTTAGTCNCGGACTCAATATTTCC CATCTCCCGTTGTGAGTCCCAGCAAGCTGACAGTTCTCCCAAAAGGGGGG GCTAAAATACAACTTCCAGCAACCC
64	GGCCGCTCTAATCAATCATGGGGCGACATGGCAGAAGCTAAATTATTAGCC CTAGACTTTCTCAGCTCCTAACCGATACATAAAATTGGCTCATGATTGCAGT CAGCAGTATACAAGTCGTGCATAGCTATTTACTAAACATTCCTCAAGTCATT GTATTACATGGAGATAACCATAAATTATATCTGATTGTCAAGAATTTACAGTT TGGGCTTAAAGTCTTTGAGATACAACATGGCTATTATGGGCAAGTGAAAGC AGACTTGTGTAGTTCAGTTCAGTGAGGGTGGATTCCAGGACTGGCAGTGC TACAGGGGCTAGTGTTACTTTCTATCAGGACTTGGGTTTGTATCAGCCAAA GTGGATCCCAGAGCCATTGCTAAATCAACCTGGAACTCTGCTTACTGTAGA AAGTCAAGATTTTTAAGGCCATGGTCGCCTTGTAAAGAAGTTCTTAGGAGG ATTCCAGTACAGAACTAATTGTGTGAATGGGAGAACACAGCTCAAGCCTGG TCTGCACATAAAGCAGCTGGGGAGCACAAGCCATCAGTGCGAGAGGTGG

	GAGACAAATGATCAGCAACCAGCAAACGTCCCCTAACAAATTGCCATGAAA TGAAAATGTATCAGTCCCCTCACCGTGTCATATTGTGAGTGCAGAGAGATT GCTGGGGGGATGAGCGAATTACTGCTGCCACTCTCACATGGTAACTTAACT CCACTGTGCCCCCTCAACAGGCCTCGCCTTATCAGCCCATCAGGGGTTTG TAGGACTGACATTTCATTTGAATGTGACTTCTCAGCAGAAGCCCCCGACTG AACCTGCAAACTACAGGACATGCTTCATCAGTTTTATCTTGGCCTTTTCACT GTAGCGCATTCCCTGACTCCTGTATTCTTCATTTCTGGCCCCACTAACATG ATATTGCAAGTAAAGTGCCTGCCAGGTGGCAATAGAGATTTATTCCTGCAT ATATTACATGGCATCACCAAACTCAAGCACCTACAAGACATTTAGAAAGGAT TATAANAAGTGTATTGCCATGAATAAACTGCATCCTAGGTATTAAAGGAAAA ATCGATGTTATTGNGCAATGCAAGCTAGGATCAGTTTTAGGCCAGGGCATA NAGGGCATTTGANNCCCAACATCCCAGGGGTCCAATGAGGGAA
101	GGCCGCAGGGAAACCAACACTTCCAAACGAACTGTAGCATTGGATTCCTG ACCTGTAAAACCCTAGGATGCCCAATAACGTACCGTGTACACACCGTGTGT TATAAGCAGAGAACTGACTCACATGCTGTCTACAAAGAACTAACCCCCAGA GAGTTCAGACTGCAAGGTAAATAAGGGTATGTTAATGCCAAACTCTTCTTC CAGACATTTGCACTACAACCAACACTGCTCATTATGCTGATTTTGTAGGCAT ACCAAAGCACCCAGGACACCTTCTATAAAGCTAAGGAAATGACCATTAGAG ACAACAAATGTTCAGAGGGTGTGATTAACTATTCCTAAACTGCCCATATCAT TTAGAAGACCTTTTGCATCTGCCTAGTATTTCTCATGGGACTGTACTGTACA TGCAGCCTTAGGACAGCAGCCGATGGGGAGATTTGTTGGTGCGATCAAAA TACACAACTGCAAGTGACAAATCTGACAAAAACGCATCTCTGTTGGAAATA ATTGAAATTGGCAGCGGTAATACCAACACATTGCTTCGTTCTTCCACCTGC A
105	GGCCGCTAAACTGCAATGAAACACGAGGTATGCTGAGTTGGTTGCAATCTA TGCAGTCCATAAGTATGTGATGGATGTATTTCTATGCATTCTGTAGGCAGC ATTGACTCAGCAAAATGCTATGTCAGTGGGGAATATCTGTCCCATCTGCAT TGTTGGCCATTTGCCAGTGCTACAGACACAAGTGAGAGGTAGAAAACCATT AGGCCTAATTGCATATCAAAAGTGACATAGAGGAATCTTGACCACTGCAGG GGCCACCTTCAGAAACCCTCCCTTTCCTCTGGGGCCCAGTACAACATTTAG CATCTCTTGTGTTATGCAAAACCAAATTCAAATTCAATTTGATACCTAGTTCA GTCCTGCA
116	GGCCGCAACACAGGCGACTTTTCATTATAGCAGATGGGAAGACACTTGGA GGCAGTTCGGGGAGATTGTCACCCAGAAGAAGAGGGGATTAGTCACCAGG CAACTAAATCTCCCTGAATCTCCTCGTGTGGACTATCCCTTAGTCTTCTGTA CTTCAGGCTCTCCGGCTCCTGGTTGTTAGCGTCTGTGCTCCAGACAGCAG GGTGTTAATGACTCCCTGCAGTGACAAGCAAAGAAGCAAAGCCTTTTCTCT CATTCTGCTTGCTTGGTAAATATAGTAGATCCTTTTATGTTTCTCAAGGGAC AAGGAAAAACCTTGTACATTTTGTCCCTGCA
138	GGCCGCCAGAACTTGGGTAACTGCTGATGAAGAAATCTTTATAGGCAGATT TCACCCGCAACATCATGCTCCCCTTGCGTTGTTAGTGCTTTCTCAAGATGT CCGTGTTTAAGTTTTACCATGTCCTTGCATTCAAAGTACAATGAATTAATCT GGGAGCAGCTTCTTAACAAACTTGTCTTTGTGTGCTGTGACCATGAACAGT CTATAAACACCCTGTGGTTCCTTTGTTTTGGTATTCGGACTCTTGGTGAGTT CAGAATCCACACTGTTTGTGTCAAATAAACTTATTTTCCTGTGTCCTCACAT TTCTCTATTTCTCTGGCCTGCTTCAATGCTGCCAAGATGCCCGTTAGAACG GCCTTCTCATTTCTATTGGCCATAATCCTGCA
160 - Does not include SbfI site as seq too	GGCCGCGACAGGTTTTAGGTGGTGTATTTTTGGATTCAAGCTATTTTCAGG GTCGAGGTATAATAAATCTAGAATATTCAAGTTTTTTTTAAAGAAAACTCGA GCAATTTTTTTTTTAACTCAAACTTTTGTGGAAAAACAACTCGAACCTTAATA AATCTGCCCCAAAGTGTTCTTATGCTTATGCTCATCTGTTGTACTTCCTCCT CGGTCCCTTACAAACACTAGACAAGTGGCCCCTAAGACTTTGTTGTTGCCC TCTGGTGGTCAGTCAGAGTGAGTTTTGGCCACTGTACGTGACAACTACACC AATCTCTCAGCCACGCATTCCTCTAAAGGACTGAAATCAGCATCTTAAATCT

long for sequencing	GCCCATGTGTGACCTGAAATTGTAAATTGTAGAGATCACAAATTTTCCTGCT GTTTACAGCAATTAATTTTTATAGGCTGTACAGGAAAAGAACCTATGGAAGA CCGACATACATAACATGAAATGTGTGTAGTTTTTACCAAAGCATCTCATACA CCATATTTTATTTAGCAAAGAGAACATGCTAAAACATTTGCTTTAAACCAAAT ATGTGTTCATTTCCATATAGAAAACAAATTACCTGAAATGCAAATATATGAG AAATGTGATTGTTCCACTGCATCTTTAATGTAAGAGCTCTAAATTTAGTTAG CATAAAGAGGCTTTGAAGTTTCCAAAGATCCATACAAGACACCTATATTAGT CTCAGTAAAAGTAAAACAGATCAAAGACTCAAGTCTCTGCTATTCCCTGCA GAGAAAGACTAAGAATATAATTCACATTAGATTAGTTAGAAACAAAAGACTG AAATGGTATCCTGCTACCCTTCCATGACTACAGTATTTAAAAAAAAAAAGGT TAAACTAGCGGCGGATAAAAGAAGAAAAAAAAAGGAAAAATATGAGCCACA GCTCTGTACTAACCACAGAGCATTTCTTCCCAGAATGTTTTGCATTGTACCT AAGCAAAAGTGTATTTGTTCCCATGATGTTGCTAGGCA
180	GGCCGCGAAGCCTGAACAGTAGAACGAAATAGCCTGCGAAAAGGCAAAAT GTAACAAAAAATCCAGCAATTTCCTTCTACTGTGAATGCGCCTGCCCTGGG AAATTTGAAAAGCGAAGAAGAGGATTGCTCCGTGGTTCTCGCTGGCAGAA CCGCAGGCTGGTGTAGTTTTTTGCAAGTAAATGTGAATACTAAACAAGAAT CACCCCATGTGCTTGCTATGAGCCATTGTTTGTAATGTTGTTGGTCTGGGG CTGCGCATTCCATTTAATTAGCATCTCAATTCAATAAGCAGAAAGGCTTTGC CAGCTGTTAGCAGAGAGAAGGTAGAAGCCATAGACAGCCATTGATTCCAAA TTCATGTAAAATGCATGAGCTGTGCAGGCACCATAAGGCCACCCCATAGG GTGAACTAGTCACAATAATAATGGGTTTGACTGCTCCAGGCTGTGCATTGT GCACTCTGTAATATTATGTGACAAAAAACATGGCAGGGCCAGAGAATGGTC TATAATCAGTAAAAGAAGTCTGATTGGGATCAATTGACATTGGTTTTGCCTT ACTCTAATGCTGAGTGTATATCGAGTGTCCGTTCCATAATGAAGGGTTCCC CAACCTTTTTCACCCATTAGTCACATTCAAATGAAAAAGGAGTTGGGGAGC AAGACAAGCATGGAAAGGTCTCTGAAGATGCCCAATAAGGACTGTGGTTG GCTATTTAGTTGGCCTATGCAGACTGGCAGCCTATGGGATGCTCTGTTTGC CTGCA
182	GGCCGCCTGAGCTGAGATCTTGTTTCCAGCACACAGGACCCAACGTCACT GTTTGAACCCACTTGACAAGATACAACTGCTGCCTTCAATGTGCCAGAGAG CAGAACAAGCCGTTTAATGGGAAATTCCCTCAAAACAGTGGGCAACTATGC TAATTTTTACACACTACATGTGCATAATTGTACCCTCAGGTTATATGCAGTA CCACCTTGCAATACAGTTTACTGCTAGGCTACAGACGGTAATAGAGGCATA TGGGCCCCAGTGTAGAATTTGCCTTTGGGCCACGTTCTTTGTTTCCTTGTA ATAAACATTCCAACCAATTACCACAGGTTTCTCATTCCTAAATCCCCCCCCC AAATGCCTCAAAGCCTGTTATAGAGATGATTGGCAACTTCTTTACAGCCTTT GGTTCAGGTCAGGGCCCCGGGAGCATTTTTGATTCCAACATCTCCTGTGG CCCTAAGTTCAATAATTAGATCACATAAGGGGCCCATGAAGAGTTGTCAGC TGAGTCCATAACTAATGCAGTCCTCCTGCA
183	GGCCGCCTCTTTATGGGTGTGCAAACTACTAAGCGCACGATCGATTCACCA CCTACTCAGTTAGAGGCCTCAATTTCTTGTCTTTCTGATCGTAGCATATATG CCGCTGCTTGTCTTGTAATATACATACAGTTTAGTACACACGCTGGGAAGA GTGCCTTACATTTCCTTTTGTTATCTGAATGGGGAAACAGACCATTCCAGG AGTATCACCTCCCACGTCTCTGGCCATTGTACATAAAAGGCTCCAGATAAA GAACAGCAGAGTAGCAAAGATAAATACTTCCTCTATAGCCTCATCAATTTGC CTCAGAAGGGGAAAAAAACCTTCATGACTCCAAAAAGCAATCAGATCTCGT CCCTTTATCCTCTACCCCCAGCTTTCTTTCTACATAATTTGCTAATCGTGTC TCATTTGCCTTTCATTCTGTCTTTTTTTTGACTGGCAATAACTTAAACATTTC ATTAAATAAAACTGCAGTAACTTGTTATTAAAGGATATTCTCCCAGCTTTCAA TACGGGGTCATCGACCCCAGAAACCAAAAAGCTTTAGATCTGGGAGGCTA CAGTTTAACTGTTCTCTCTTTTCAAATTCTCTTTATTGAAATTAATAATTAATC CTGCAGACTTTTTGACAGAAGTTTTAGACCAACAGTCCACAAAAAGAAAGG CACCTGCA

192	GGCCGCCAGGTCGAGCCTAATGTTGTAGCAGGATAGTGACCAGCTGTTTT TTTTGATAGAAATAATAATATTGAATGAACATTGTGCCAATTTGGCTACGAA CACCGAAATGGGGAGGGTAAAGCATAATAATCCCCCCCCCCCCCCCAGACA ACATCAGCTCCCTCCCTGCCTGCCCCGGCTGTTAGCTGGTGATTGATTGC CTTATTAAAGCGTGTTCTTGTAAGTGTGACCAAACTGATTGCAAAGTACATG TTTAGAATAAAGCTTGCCCTAACAGGCGAATAAATTGGGAGGCTGGGCGT GACTTGCTGGTTAATAATTTATCATTCAGCTACAGGGATCGGTTAATCGAAT TTGCATGGTGTACAGTTACTAAAGTTACTTGAGCAATTCAGCTTCAATTGCT TAAAGTTAGCAAGAGGAACCCCTGGTATCCATTCTCTATAAAAAAAAAAAAA GCCCTCAGCCAAGATACATCACTGGCAACTGCAGATCTTAACCCTGCTGCC CCATCTATCATACGAACATACCAGCATCCTTGCCTCTTCCCCTTTGTTTTCA ATTATCCCTGTGCTCCATTTGTTACATTCTATTCCTTGAGTTGAGCATTGAC TGGAAATAGCCGCTACCATATGTCAGCCTGTCCCCCAAGCCTGCAAGCTA CAATAGTTTCGGGAACAGAGCCCGGACCTTTGTTTCAAAAACATGTTCCTT GCTATACAATCTGGGGACAAAGATGCTAATTAGCTTTTAACTCCACAAGTGT TGATGTGAGCCTCCACCACAGAACAAGAAATGGGCTGATAACCAATCCTGC A
217	ACCGCGGNTGGCGGCCGCATAACACCTGGGCAAGACAATATAGTATAAAC AATATATGCATTGCAAAAATTAAATATAAATATAACAATAGTAAATATAAATA TAATACTGAAAATAAAAAGTAATATAGAAATAGAAATTGGTTGCATTTAATAA GTCTGTAAATGTCAAAAGCAATGTAAATAAAGCTGATTTGTGTTGGCCTGA GTTGAAATGAATAAGAAAAGGAGCTCCCACTCCAAACTAATTCTCTGCAGG TGGCCCTGATATGCATGTTATATGTGAGAGATGTGGGTGGGGAAGCCATA AGCCTCCTGAACGTGTATGTGTCTCTGGGTAGCAGGAACTAGCAACTCTG CCTGTTCTCAAAGCTGCAAAAGGTATTAATCACTTTGATCTCTCCTGGGAGT CCAAACAGAAACCAATGGCTTCTTCAGCTGTGTTTTATCCCTGTGCATCCT GTAGCAGGTACTCATATGTTTGGCCTCAATGTGTGGGGTTTCCAGACTCAC ACAAACACTTCACACAATGCTCTAAAAAAACCGCCCATTGAACAGCTCAGA TTGTGGATGACCGAATATTAATAACCACGCCAATGTATTGCTCGTGTGTCTA TCCACTTCTCAGTTTAGATTGTAAGCTCTTATGGGCAGGGCCCTCTTTACCT CCTGTATCAGCCAGTAATTGTCCTGCAGGTCGACCATAGTGACTGGATATG TTGTGTTTTACAGTATTATGTAGTCTGTTTTTTATGCAAAATCTAATTTAATAT ATTGATATTTATATCATTTTACGTTTCTCGTTCAGCTTTCTTGTACAAAGTGG TTCGATTCGAGGTCGACGGTATCGATAAGCTTGCGCGATATTCATTAATAG AATAGAGGCATTTTAATACATTTCTGCACAATTAAAAATTAAATATAATCCTG CAAGTCTATAATTA

Appendix 5: Detailed enhancer pipeline protocol.

Enhancer pipeline protocols

You need:

- Genomic DNA
- Plasmid backbone
- Inserts

Genomic DNA

1. Harvest st. 38 *Xenopus laevis* embryos and place 10 in each tube. Remove all water and keep on ice. If they are not being used straight away store at -80°C.

2. Thaw the embryos in room temperature.
3. Turn on heat block for 55°C.
4. Follow Invitrogen by Thermo Fisher Scientific PureLink™ Genomic DNA mini kit.
    a. Add 180µl digestion buffer and 20µl proteinase K to the tube. Ensure that the tissue is completely immersed in the buffer mix.
    b. Incubate the samples at 55°C for 1 hour, with occasional vortexing until the lysis is complete. Use orange needles with 1ml syringe to break the cells up.
    c. To remove any particulate materials, centrifuge the lysate at max speed for 3 mins at RT. Transfer the supernatant to a clean microcentrifuge tube.
    d. Add 20µl RNaseA to the lysate, mix well by brief vortexing and incubate at RT for 2-5 min.
    e. Add 200µl PureLink Genomic lysis/binding buffer and mix well by vortexing to a yield homogenous solution.
    f. Add 200µl 100% ethanol to the lysate. Mix well by vortexing to yield a homogeneous solution.
    g. The next steps are for PURIFICATION:
    h. Add the lysate to the PureLink Spin Column (~640µl) and centrifuge at 10000 rpm (?) (10K G) for 1 min at RT. Discard the collection tube and place the spin column into a clean purelink collection tube.
    i. Add 500µl wash buffer 1 (with added ethanol) to the column and centrifuge for 1 min at RT. Discard the collection tube and place spin column into a clean purelink collection tube.
    j. Add 500µl wash buffer 2 (with added ethanol) to the column. Centrifuge at max speed for 3 mins at RT. Discard the collection tube.
    k. Place the spin column in a sterile 1.5ml microcentrifuge tube and add 30µl elusion buffer to the column. Incubate at RT for 2mins. Centrifuge for 1 min at RT at max speed.
5. Nanodrop.

<u>Plasmid backbone preparation</u>

If needed transform the plasmids (#1960, #1962, #1965, #1966) to get an increased concentration of the plasmid. Purify using miniprep, midiprep or low endotoxin midiprep kits.

1. Set up a digestion reaction for #1960 or #1962 in PCR tubes. 50µl reaction. Amount of DNA depends on concentration (ng/µl) you have obtained.

SW H_2O	40µl
Cutsmart (RE buffer)	5µl
DNA	3µl
Sbf1	1µl
Not1	1µl

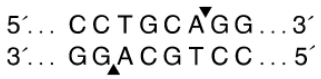
2. Run a PCR program that goes for 37°C for 1 hour, and 80°C for 15 min (last to deactivate the enzymes). Then 10°C forever for storage until you use it. OR use a heat cabinet at 37°C and then a heat block at 80°C.
3. Run all the sample on a gel. 50µl digestion + 10µl dye (5:1). Use 5µl of the ladders. 100V for 40 min.
4. Weigh the 1.5ml microcentrifuge tubes before you add the gel to it.
5. Image gel, then us a UV light block, wearing protective gear, and use a razer to cut out the bands taking as little extra agarose as possible. Place band in pre-weighed tubes.
6. Weigh the tubes with the gel in them.
7. Follow gel extraction kit. QIAEX®II Gel extraction kit (150).
    a. The QIAII beads are in the fridge.
8. Nanodrop. The nanodrop might not look good but this is okay. 5-8ng/µl is fine.

<u>Make primers for open regions</u>

Find about a 21bp sequence at the top strand at the start and add a Not1 restriction enzyme (RE) site at its 5' end and two extra bases for increased RE binding (TTGCGGCCGC) with a melting temperature of 55-56°C. Do the same for the end of the sequence at the bottom strand and add a Sbf1 site and two extra bases at the 5' end (TTCCTGCAGG).

Not1

Sbf1



When the primers arrive add the recommended amount of SW H_2O. Dilute a part of the stock 1:10. Add 10µl of the stock to 90µl of SW H_2O. This is the solution you will be using in your experiments.

<u>Testing primers</u>

Test the primers first using the low fidelity TAQ polymerase.

1. Use a ready mix and mix the following in PCR tubes. 15.5µl in total.

Biomix	6µl
Primer 1	0.5µl
Primer 2	0.5µl
SW H_2O	7.5µl
Genomic DNA ~200ng/µl	1µl

2. Run the following PCR program.

	Temp	Time	Repeats
Initial denaturation	94°C	2 min	1
Denaturation	94°C	30 sec	35
Annealing	56°C	30 sec	
Elongation	72°C	2 min	
Final extension	72°C	5 min	1
Hold	10°C	Forever	1

3. Run 10µl on a 1% gel. The Biomix has dye in it.

<u>PCR with high fidelity Phusion polymerase</u>

Here you need a negative control as well as your samples. The negative control can be using only 1 primer and water as a substitute to the 2nd primer, OR, you can add the primers but no DNA.

1. Mix the following. You can also make a master mix, divide it, and then just add the primers separately. Makes a 25 µl reaction.

5xbuffer HF	5µl
10mM dNTPs	0.5µl
DMSO	0.75µl
Mg	0.5µl
Phusion polymerase	0.25µl
SW H_2O	14.5µl
DNA ~ 200ng/µl	1µl
Primer 1	1.25µl
Primer 2	1.25µl

2. Run the following PCR program.

	Temp	Time	Repeats
Initial denaturation	98°C	30 sec	1
Denaturation	98°C	10 sec	Repeat 35 times
Annealing	58°C	30 sec	
Elongation	72°C	2 min	
Final extension	72°C	5 min	1
Hold	10°C	Forever	1

3. Run 5µl PCR reaction with 1µl dye on a gel. Keep the other 20µl in the freezer whilst the gel is running.

<u>PCR with Kapa long range hot start ready mix</u>

1. Mix the following up to 25µl. Can make a master mix first: 22.5µl in each tube before primers.

	Protocol online	Emily (Works fine)
SW H_2O	?	13.75µl
Kapa ready mix	12.5µl (x1)	7.75µl
Primer 1	1.25µl	1.25µl
Primer 2	1.25µl	1.25µl
DNA ~200ng/µl	?	1µl

2. Run the following PCR program.

	Temp	Time	Repeats
Initial denaturation	95°C	3 min	1
Denaturation	95°C	30 sec	Repeat 35 times
Annealing	?°C	30 sec	
Elongation	72°C	1 min/kb	
Final extension	72°C	1 min/kb	1
Hold	10?°C	Forever	1

OR from Emily:

	Temp	Time	Repeats
Initial denaturation	94°C	3 min	1
Denaturation	94°C	15 sec	Repeat 10 times
Annealing	55°C	15 sec	
Elongation	68°C	3 min (1kb/min)	
Denaturation	94°C	15 sec	Repeat 25 times
Annealing	63°C	15 sec	
Elongation	68°C	3 min (1kb/min)	
Final extension	72°C	4 min	1
Hold	10°C	Forever	1

3. Run 5µl on gel. Save the other 20µl for further cloning.

PCR purification

1. Follow the QIAquick® PCR Purification Kit (50) protocol and elude in 30µl EB buffer.
2. Nanodrop.

Digestion reaction for PCR product

1. Mix the following in PCR tubes. Here you can make a master mix then add the DNA in separately. Makes a 25µl reaction. Two ways of mixing shown. Second works fine.

SW H_2O	0.5µl	10.5µl
Cutsmart (RE buffer)	2.5µl	2.5µl
PCR product	20µl	10µl
Sbf1	1µl	1µl
Not1	1µl	1µl

2. Run a PCR program that goes for 37°C for 1 hour, and 80°C for 15 min (last to deactivate the enzymes). Then 10°C forever for storage until you use it.

Ethanol precipitation – this step might be redundant

This is done to clean up the sample to remove buffers and enzymes used in the digestion.

1. Add 2.5µl sodium acetate and 75µl 100% ethanol to the 25µl digestion reaction (0.1:1µl and 3xvolum of ethanol). Mix well by inverting. This can be stored in the freezer until the next day.
2. Turn on cold centrifuge (fast cool, 4°C).
3. Transfer samples to 1.5ml microcentrifuge tubes and centrifuge for 30min at 4°C.
4. Remove supernatant.
5. Add 500µl of 70% ethanol, not need to resuspend, and centrifuge for another 5 minutes at 4°C.
6. Remove all the supernatant. Be careful not to disturb the pellet, but the more liquid that is gone the quicker the next step.
7. Airdry pellets so they turn from white to transparent.
8. Resuspend the pellet in 20µl SW H_2O.
9. Nanodrop again.

Ligation

You need more insert than vector. A 1:6 ratio worked fine. Negative control here is a sample that has all the reagents but no insert. Water is used instead.

1. Mix the following. It makes a 10µl solution.

SW H_2O	?µl - dependent on other ratios
Ligation buffer	1µl
Plasmid	1 (ratio) - 2µl of 8 ng/µl was good. Work from that
Insert	6 (ratio)
T4 ligase	1µl

2. Leave at RT for 3h then put in freezer if not going straight to the next step, OR, leave in cold room at 4°C overnight then put in freezer if not going straight to the next step.

Transformation

Here you need some controls. A negative control is the negative control from the ligation (only plasmid backbone, no insert). The positive control is the original plasmid, #1960.

You need:

- To work around a Bunsen burner.
- Make LB agar plates with carbenicillin. They need 20 minutes to set.

1. Heat up a 200ml LB agar in the microwave until completely melted (2 min 30 sec).
2. Cool under the tap until touchable.
3. Add 200µl carbenicillin. Mix well.
4. Add to petri dishes. You can make 12 plates from one flask.
5. Label with date, initials, and carb. If using immediately you can label with what you will palate on it too.

- Collect ice and collect competent cells from -80°C. Keep on ice.
- Turn on heat block at 42°C.
- Turn on heat shaker at 37°C at 400 rpm.

1. Vortex and spin the competent cells.
2. Mix 50µl competent cells with 5µl ligated plasmid. Use only 1µl for the positive control. Vortex, spin, and leave on ice for 30 minutes.
3. Heat shock the competent cells by placing then in the heat block for 90 secs, tapping them at the side every 10-15 secs, and placing in a new hole every time.
4. Put the competent cells back on ice for 2 mins.
5. Add 1ml of LB to the competent cells. Work around the Bunsen burner.
6. Place the competent cells in the heat shaker for 1h 30min.
7. Spin the competent cells at 7000 rpm for 5mins, pelleting the bacteria.
8. Remove 800µl of the supernatant.
9. Resuspend the pellet in the remaining 200µl. Plate, make sure you spread over the whole plate.
10. Incubate the plate upside down at 37°C overnight.

<u>Colony PCR</u>

Can do it with LB plates or with water. This method describes the water way.

As there is no X-gal selection in this vector chose extra colonies to test PCR. For example, chose 5 colonies from each plate.

Positive control can be the positive transformation and the negative control can be the negative ligation/transformation or just water.

1. In PCR tubes place 30µl of SW H_2O.
2. Use P2 pipet tip to pick up one colony and place the pipet in one of the PCR tubes with water and mix thoroughly.
3. In new PCR tubes mix the following. Here you can make a master mix without the colony dilution first and then divide 22µl into each PCR tube.

Biomix -NEED NAME	10.5µl
Primer 1	1µl
Primer 2	1µl
SW H_2O	9.5µl
Colony dilution	3µl

4. Run the following PCR program.

	Temp	Time	Repeats
Initial denaturation	94°C	2 min	1
Denaturation	94°C	30 sec	35
Annealing	56°C	30 sec	
Elongation	72°C	2 min	
Final extension	72°C	5 min	1
Hold	10°C	Forever	1

5. Run a sample (10µl) on a gel to see if the ligation/transformation has worked.

<u>Making a mini-culture for sequencing</u>

6. Grow 10µl of the colony water in 5ml LB with 5µl of carbenicillin overnight at 37°C at 180 rpm.
7. Miniprep 4ml of the culture using the QIAprep® Spin Miniprep Kit (250).
    a. Spin 2ml of the 5ml culture in a 2ml microcentrifuge tube at 6000 rpm for 5 min. Discard the supernatant. Add another 2ml culture in the same microcentrifuge tube and spin at 6000 rpm for another 5 min. Discard the supernatant.
    b. Add 250µl of P1 resuspension buffer (found in fridge) and resuspend the pellet by pipetting.
    c. Add 250µl of P2 lysis buffer. Mix by inverting the tube 4-6 times or by pipetting. Incubate at RT for 5 min, but no longer. This kit usually has LyseBlue in the P2 so the solution will turn blue and gloopy.
    d. Add 350µl of N3 neutralization buffer buffer and mix immediately by inverting the tubes 4-6 times. The solution will turn white.
    e. Centrifuge at 13000 rpm for 10 minutes.

f. Transfer 800µl of the supernatant to the column provided. Be careful not to disturb the pellet.
g. Centrifuge at 13000 rpm for 1 minute. Discard the supernatant.
h. Add 750µl of PE binding buffer. Centrifuge at 13000 for 1 minute. Discard the supernatant. Spin for another minute to remove residual waste.
i. Transfer the column to a clean 1.5ml microcentrifuge tube.
j. Add 30µl elusion buffer (EB) to the centre of the filter and incubate at RT for 1 minutes. Centrifuge at 13000 rpm for 1 minute to elude the DNA.

8. Nanodrop.
9. Send 15µl of 50-100ng/µl together with 2µl of one primer for sequencing.

$$\frac{X\ ng/\mu l}{50\ ng/\mu l} = Y$$

$$\frac{1}{Y} = Z \times 15\mu l = \mu l\ miniprep$$

$$15 - \mu l\ miniprep = SW\ H_2O\ you\ need\ to\ mix\ with\ miniprep$$

a. https://eurofinsgenomics.eu/en/ecom/checkout/login-register/?returnUrl=intLinkSIT_TubeSeq&msg=3
b. Log in -> Chose nr of samples -> define type eg. Plasmid -> chose whether premix or not -> add one number then click arrow under to auto fill in the other samples -> add to cart -> remember to put in your email to get results -> and remember to add PO number in.

Making a midi-culture for injection

1. Add 100µl of carbenicillin to 100ml of LB, and 100µl of your 5ml culture.
2. Grow overnight at 37°C at 180 rpm.
3. Next day do a midiprep following an Invitrogen PureLink™ Fast Low Endotoxin Midi Plasmid Purification Kit by Thermo Fisher Scientific.
    a. Pellet max 50ml of bacterial culture. (Carmen does 100ml the kit says max 50ml). Centrifuge at 4°C at 6000 rpm for 15 mins.
    b. Prepare the vacuum by washing through it with RO water. Place higher than vacuum machine and make sure the vacuum is working.
    c. Add 8ml of Resuspension buffer and resuspend by vortexing.
    d. Add 8ml of Lysis buffer and mix by inverting 6 times. Incubate for 3 mins at RT. The solution will turn dark purple and viscous.

e. Add 8ml of Precipitation buffer and mix by inverting 6 times. The solution will turn yellow when neutralising is complete.
f. Load the lysate into a syringe and leave for 5 mins until the precipitate has floated to the top.
g. Remove the lock and insert the plunger. Filter the lysate into a new 50ml tube. Do not use excessive force.
h. Add 8ml of binding buffer to the clarified lysate and mix by inverting 10 times.
i. Insert the column in the vacuum and add the lysate to the column assemblage. Turn on the vacuum until all the lysate has run through.
j. Take out the column and reapply it. Unscrew the column assemblage and the cap and discard.
k. Wash with 800µl Wash buffer 1, turn on vacuum, let the liquid go though, take off column and reapply.
l. Wash with 800µl Wash buffer 2. Turn on vacuum, let liquid go through, take off column and reapply, redo this step once more.
m. Place column in clean microcentrifuge tube. Centrifuge at 10000 rmp for 1 min to remove any residual fluid.
n. Place column in a new microcentrifuge tube, add 50µl elusion buffer. Incubate at RT for 2 mins. Centrifuge for 10000 rmp for 1 min to elude the DNA.

4. Nanodrop.

OR

Do an endotoxin free miniprep (PureYield Plasmid Miniprep Systems from Promega). It will usually give about 300 ng/µl, which is enough for frog.

1. Grow up 7 ml of your culture at 37°C at 180 rpm.
2. Do two minipreps to get 40 µl.
3. Centrifuge 1.5 ml of the bacterial culture for 1 minute at maximum speed in a microcentrifuge. Discard the supernatant.
4. Add an additional 1.5 ml of bacterial culture to the same tube and repeat step 3.
5. Add 600µl water (or TE buffer) to the cell pellet, and resuspend completely.
6. Add 100 µl of cell lysis buffer, and mis by inverting the tubes 6 times.
7. Add 350 µl of cold (4-8°C) Neutralization solution, and mix by inverting.
8. Centrifuge at maximum speed in a microcentrifuge for 3 minutes.
9. Transfer the supernatant (about a 1 ml) to a PureYield Microcolumn without disturbing the cell debris pellet.

10. Place the minicolumn in a collection tube and centrifuge at maximum speed for 30 seconds. Discard the flowthrough.
11. Add 200 µl of Endotoxin Removal Wash (ERB) to the minicolumn. Centrifuge at maximum speed for 30 seconds.
12. Add 400 µl of Column Was Solution (CWC) (With added 95% ethanol). Centirfuge at maximum speed for 1 minute.
13. Transfer the minicolumn to a clean 1.5 ml microcentrifuge tube, and add 20 µl Elution buffer or nuclease-free water directly to the filter. Let it stand for two minutes.
14. Centrifuge at max speed for 30 seconds to elude the plasmide DNA. Store at -20°C.
15. Nanodrop.

Injection

1. Pull needles.
2. Digest the vector by mixing the following:

10xI-Scel buffer (Cutsmart)	2µl
I-Scel meganuclease	2µl
Plasmid (0.1µg/µl)	8µl
SW H_2O	8µl

$$\frac{100ng/\mu l}{Xng/\mu l} x 8\mu l = Y\ \mu l\ plasmid$$

$$8\mu l\ -\ Y\ \mu l\ plasmid\ = \mu l\ SW\ H_2O\ to\ add\ to\ Y\mu l\ plasmid\ get\ 0.1\mu g/\mu l\ plasmid$$

- To get 200ng/µl divide 200 by X.

3. Incubate at 37°C for 40 mins. Keep on ice. Leftovers can be frozen down and reused.
4. Inject into one-cells staged embryos in a period of 45mins after fertilization. For laevis the injector should be set to P_{out}=0.7, P_{inject}=10 according to Marta.
    a. However, when you cut your needle and fill it with water calibrate it so that you inject 10nl into a 1-cell stage laevis embryo. This is done by zooming in as far as you can into the needle, then using a measuring eye peace make sure the bobble you inject is 13. For tropicalis inject 4nl which is 5.2 on the graticule.
5. Keep injected embryos in 3% Fricoll at 12°C until they are at a 4 cell stage. Or keep in Fricoll for 30 min at 18°C.

Fricoll – mix using magnetic stirrer

Fricoll	3g
1xMMR	30ml
RO (green tap) water	70ml

6. Transfer the embryos to 0.3 MMR with 50µg/ml gentamycin and keep them at 18°C until gastrulation.
    a. Or change the solution to 0.1 MMR and move to the incubator of your choice.
7. Them move to 22°C. When you inject, if you put the embryos at 18°C overnight they will be st.12 (gastrula) the next morning. If you put them at 23°C overnight they will be stage 14 (neurula) the next morning.
8. When you have satisfactory GFP expression fix the embryos for 2 hours at RT or overnight at 4°C in MEMFA. Was 2x5min in DepC-PBST and keep in fridge. Do not was with ethanol as this washes away the GFP expression.