



**BIOLOGICAL  
SOCIETY**

## **Comparison of Shotgun Fragment Library Approaches for Functional Enhancer Screen of the *Six3* Gene**

Carmen Jones  
100194921

Supervisor:  
Dr Timothy Grocott

A thesis submitted for The Degree of Master of Science by Research

March 2022

Word Count: 27103

## Abstract

Cyclopia is defined as the failure of the eye field to successfully divide into two. This is encompassed by the broader brain defect, Holoprosencephaly (HPE), where the brain lacks any ventral identity, and the forebrain fails to divide.

Key signalling pathways such as Sonic Hedgehog (Shh) and Nodal have been identified as critical for the ventralization of the neural plate and prosencephalic regions of the forebrain during embryo development. Specific mutations within key genes of these pathways have identified as them as causative genes for the Cyclopic/HPE phenotype.

Despite extensive knowledge surrounding these causative genes, little is known about the regulatory elements, enhancers, that control the spatial-temporal expression of these genes. In recent years many different techniques have been developed to identify and characterise potential regulatory elements.

Using the Cyclopia gene network as a disease model, different methods for detecting regulatory enhancer elements were tested. This project aimed to construct a 3D shotgun fragment reporter library for the causative Holoprosencephaly gene, *Six3*. Utilising the use of BAC clones, our approach focused on optimising the random amplification of fragments followed by a modified Golden Gate cloning strategy to circumvent the loss of material associated with traditional methods for creating BAC clone fragment libraries.

Here we successfully edited our reporter vector to contain 5 unique barcode sequences for rapid detection of enhancer activity using RT-PCR. We successfully modified our Golden Gate Cloning strategy to produce hundreds of colonies ready for library construction. Using a candidate approach, a putative enhancer sequence was identified for the *cMyc* gene. Using phylogenetic foot printing, ECR2 was identified and using RT-PCR this sequence was amplified. This ECR was then cloned into the pTK-Citrine reporter vector to validate the enhancer's spatio-temporal activity *in vivo* using the model organism, *Gallus gallus* (chick).

## **Access Condition and Agreement**

Each deposit in UEA Digital Repository is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the Data Collections is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form. You must obtain permission from the copyright holder, usually the author, for any other use. Exceptions only apply where a deposit may be explicitly provided under a stated licence, such as a Creative Commons licence or Open Government licence.

Electronic or print copies may not be offered, whether for sale or otherwise to anyone, unless explicitly stated under a Creative Commons or Open Government license. Unauthorised reproduction, editing or reformatting for resale purposes is explicitly prohibited (except where approved by the copyright holder themselves) and UEA reserves the right to take immediate 'take down' action on behalf of the copyright and/or rights holder if this Access condition of the UEA Digital Repository is breached. Any material in this database has been supplied on the understanding that it is copyright material and that no quotation from the material may be published without proper acknowledgement.

## Abbreviations

ATAC – assay for transposase-accessible chromatin using sequencing

BAC – Bacterial Artificial Chromosome

Bmp – Bone Morphogenic Protein

Cdon – Cell adhesion molecule-downregulated by oncogenes

ChIP-seq – Chromatin Immunoprecipitation followed by deep Sequencing

Cnot1 - CCR4-Not transcription complex, subunit 1

CRE – *Cis*-Regulatory Element

ECR – Evolutionary Conserved Region

EFTF – Eye Field Transcription Factor

Foxa2 – Forkhead Box A2

Gli2 - Gli-Kruppel family Member 2

NR – Neural Retina

OV – Optic Vesicle

PIC – Preinitiation complex

Ptch1 – Patched 1

RPE – Retinal Pigmented Epithelia

RT-PCR – Real Time-Polymerase Chain Reaction

Shh – Sonic Hedgehog

Six3 – Six Homeobox 3

Smo – Smoothened

Stag2 - Stromal Antigen 2

TF – Transcription Factor

Tgf $\beta$  – Transforming Growth Factor Beta

Tgif1 – Thymine-Guanine Interacting Factor 1

TSS – Transcription Start Site

# Contents

Abstract.....	2
Abbreviations.....	3
List of Figures.....	6
List of Tables.....	8
Acknowledgements.....	9
Declaration.....	10
Chapter 1: Introduction.....	11
1.2.1 Neural Induction.....	12
1.2.2 Anterior-Posterior Patterning of the Neural Plate.....	12
1.2.3 Eye Field Specification (within the anterior neural plate).....	13
1.2.4 Eye Field Division.....	14
1.2.5 Optic Vesicle Evagination.....	14
1.2.6 Optic Cup Formation.....	15
1.3 Holoprosencephaly (HPE; MIM# 236100).....	17
1.3.1 <i>Sonic hedgehog (SHH)</i> (MIM# 600725; 7q36; HPE3; see Table 1.3).....	20
1.3.2 <i>Zic2</i> (MIM# 603073; 13q32; HPE5; see Table 1.3).....	21
1.3.3 <i>Six homeobox 3 (Six3)</i> (MIM# 603714; 2p21; HPE2; see Table 1.3).....	22
1.3.4 Transforming Growth Factor-Beta-Induced Factor 1 ( <i>Tgif1</i> ) (MIM# 602630; 18p11.3; HPE4 Table 1.3).....	23
1.3.5 Causative genes involved in the Shh signalling pathway ( <i>Ptch1, Gli2, Cdon, Disp1</i> ).....	24
1.3.6 Other causative genes ( <i>Stag2, Cnot1</i> ).....	26
1.4 Regulatory Elements.....	27
1.4.1 Promoters.....	28
1.4.2 Enhancers.....	28
1.4.3 Identification and Characterisation of Enhancers.....	31
1.5 Project aims.....	34
Chapter 2: Methods and Materials.....	36
Chapter 3: Shotgun Library Construction.....	43
3.1 Introduction.....	43
3.1.1 Fragmentation.....	44
3.1.2 Cloning.....	44
3.2 Results.....	46
3.2.1 Proof of Concept: Validating Group Testing Strategy.....	46
3.2.1.1 Dilution of Fluorescent Plasmid.....	46

3.2.1.2 RT-PCR for detection of reporter transcript .....	47
3.3 Discussion.....	49
Chapter 4: Modification of the pTK Citrine Vector .....	50
4.1 Introduction .....	50
4.2 Results .....	51
4.3 Discussion.....	52
4.3.1 Optimising the pTK-Citrine reporter vector for group-testing of shotgun reporter libraries .....	52
Chapter 5: Library Construction via TA Cloning .....	53
5.1 Introduction .....	53
5.2 Results .....	53
5.3 Discussion.....	57
5.3.1 Reporter Library Construction via TA cloning .....	57
Chapter 6: Library Construction via Golden Gate Cloning of randomly-primed PCR-amplified sub-fragments .....	59
6.1 Introduction .....	59
6.2 Results .....	59
6.2.1 Optimisation of Library 1 PCR .....	59
6.2.2 Successful edit of pTK-Citrine Vector to include BsmBI adapter .....	68
6.2.3 Optimisation of Golden Gate Cloning Protocol .....	68
6.3 Discussion.....	69
Chapter 7: Candidate approach .....	71
7.1 Introduction .....	71
7.2 Results .....	71
7.3 Discussion.....	78
Chapter 8: General Discussion .....	79
8.1 Construction of a <i>Six3</i> shotgun reporter library using BAC clones .....	79
8.2 Evaluation of Reporter Library Construction via Golden Gate Assembly .....	80
8.3 Alternative strategies for constructing a shotgun reporter library .....	83
8.4 Conclusion.....	84
Supplementary Data .....	86
References .....	87

## List of Figures

Figure 1.2: Scheme of vertebrate eye development. ....	12
Figure 1.2.2.1: Neural Development.....	13
Figure 1.2.3: Diagram showing the different tissues making up the prechordal mesendoderm.....	13
Figure 1.2.5.1: Scheme showing regionalization of vertebrate neural plate and neural tube.....	14
Figure 1.2.6.1: Schematic representation of vertebrate eye development. ....	16
Figure 1.2.6.2: Schematic showing asymmetric division of c-Myc cells in the CMZ.....	17
Figure 1.3: Signalling pathways involved in HPE.....	18
Figure 1.3.1: Expression of Shh along the ventral midline axis. ....	20
Figure 1.3.2: Expression of Zic2 gene at multiple stages of development. ....	21
Figure 1.3.3: Six3 expression data shown at different developmental stages. ....	22
Figure 1.3.4: Expression of Tgif1 data at multiple stages of development ....	23
Figure 1.3.5.1: Expression of Ptch1 at multiple stages of chick development ....	24
Figure 1.3.5.2: Expression of Cdon along the midline axis. ....	25
Figure 1.3.6: Expression of Cnot1 along the midline axis. ....	26
Figure 1.4.2.1: The characteristics of poised and active enhancers.....	29
Figure 1.4.2.2: Schematic showing enhancer-promoter interactions via the 'looping' model. ....	30
Figure 1.4.2.3: Models of promoter-enhancer communication.....	31
Figure 1.4.3: Schematic showing process for ATAC-seq.....	33
Figure 2.1: Diagram outlining the Candidate vs Unbiased enhancer discovery pipeline. Candidate approach outlined on the left.....	42
Figure 3.1: Construction of shotgun reporter library .....	43
Figure 3.2.1.1: Electroporation of titrated reporter plasmid.....	47
Figure 3.2.1.2.1: Reporter transcript detection by RT-PCR.....	48
Figure 3.2.1.2.2: Detection of reporter transcript by Touchdown RT-PCR.....	49
Figure 4.1: The pre-edited ptk-Citrine vector map. ....	50
Figure 4.2: Modified pTK-Citrine vectors.....	51
Figure 5.2.1: Successful identification of a single BAC clone (RP11-771F21) covering the entirety of the Six3 gene.....	54
Figure 5.2.2: Polished RP11-771F21 BAC fragments. ....	54
Figure 5.2.3: PvuII digested pTK-Citrine-Bacrcode 1 vector. ....	55
Figure 5.2.4: Colony PCR for Six3 library creation using TA cloning method. 10 colonies were selected for colony PCR.....	56
Figure 5.3.1: Overview of Library preparation for TA cloning protocol.....	58
Figure 6.2.1.1: Amplification of fragments for Six3 library.....	60
Figure 6.2.1.2: Testing template concentration for Library 1 PCR.....	60

Figure 6.2.1.3: Library 1 PCR with template DNA <7ng .....	61
Figure 6.2.1.4: Library1 PCR with varying MgCl2 concentrations.....	62
Figure 6.2.1.5: Library1 PCR with varying primer concentrations. ....	62
Figure 6.2.1.6: Library1 PCR with template DNA eluted in elution buffer.....	63
Figure 6.2.1.7: Assay for elimination of primer dimer formed during Library 1 PCR tested using digestion method.....	64
Figure 6.2.1.8: Primer artifacts cloned into reporter vector .....	65
Figure 6.2.1.9: Primer artifact and concatemer formation during Library 1 PCR.....	67
Figure 6.2.2.1 Sequencing data showing successful edit of BsmBI adapter added to vector. ....	68
Figure 6.2.3.1: Clones for Six3 library mapped to human genome .....	69
Figure 7.2.1: Putative cMyc enhancer predicted using ECR conservation and TF binding predictions. ....	72
Figure 7.2.2: Amplification of candidate cMyc enhancer ECR performed by RT-PCR. ....	73
Figure 7.2.3: Colony PCR results for GG cloned candidate cMyc enhancer. Lane.....	74
Figure 7.2.4: Sequencing data confirms cloned candidate cMyc enhancer maps to the hg38 genome .....	74
Figure 7.2.5: Electroporations of cMyc enhancer reporter construct. ....	75
Figure 7.2.6: Still frames from time-lapse movies showing expression of fluorescent cMyc enhancer construct .....	77
Figure 8.2.1: Timeline for screening 150+ kb region of the genome for enhancer activity using GG cloning approach.....	82

## List of Tables

Table 1.2: Key developmental stages in vertebrate eye development .....	11
Table 1.3: List of causative genes for HPE.....	19
Table 2.1 Primers used to insert barcode edits into pTK-Citrine vector (Chen & Streit, 2015).....	36
Table 2.2: Oligonucleotides used in Library 1 PCR.....	37
Table 2.3: Oligonucleotides used in Library 2 PCR.....	38
Table 2.4: Oligonucleotides used for Colony PCR.....	39
Table 2.5: Oligonucleotides used for reporter transcript detection.....	39
Table 2.6: Oligonucleotides used for ECR amplification .....	41
Table S1: Table showing dilutions of Meox1 Reporter .....	86

## Acknowledgements

I would like to thank my Supervisor Dr Timothy Grocott for his invaluable support, guidance and feedback on various reports, presentations and day-to-day learning throughout this project. I would also like to thank Dr Geoff Mok, Shannon Weldon, Emily Smith and Cathrine Hyde for their help, knowledge, encouragement and general good vibes throughout this period. Finally, I would like to thank Nathan Wing for his calming presence during our study sessions.

## Declaration

I declare that this is an original piece of work conducted under the supervision of Timothy Grocott at the University of East Anglia. All data presented are my own work unless referenced otherwise.

# Chapter 1: Introduction

## 1.1 Preface to Introduction

Cyclopia is a congenital disorder resulting in the formation of a single central eye orbit. This disorder falls under the most common brain defect, Holoprosencephaly, where the forebrain fails to split into two cerebral hemispheres due to disruption in the either Shh signalling or the Nodal/TGF $\beta$  pathway during key stages of development. Such severe eye and brain malformations have diverse impacts ranging from compromised vision to lethality. Previous research has identified nine causative genes associated with the Cyclopic phenotype. Regulatory elements, such as enhancers, have been shown to have far reaching effects when it comes to transcriptional regulation of genes in disease. Therefore, we are interested in identifying and characterising these enhancer sequences and the functional role they have. However, enhancers have dynamic roles within the genome and a flexibility to their defining characteristics that hinder assays. Utilising *Gallus gallus* (chick) as a model for amniote eye development, this project looks at both candidate and unbiased approaches for functional enhancer screens, and outlines the advantages and pitfalls associated with each approach. What is currently known about Holoprosencephaly, regulatory elements, and current methods used for enhancer identification is outlined below.

## 1.2 Milestones of Eye development

The eye is our primary sense organ and a major conduit through which we experience the world around us. There is a huge medical relevance to studying eye development, as defects in the eye that arise early in gestation are causative for  $\frac{1}{4}$  of all childhood blindness. Eye development is a complex process that can be split into many overlapping stages: (i) patterning and splitting of the eye field from the developing forebrain; (ii) formation and patterning of the optic vesicle; (iii) organisation of the optic vesicle to give optic cup; (iv) differentiation and specification of retina; (v) specification of lens. **Table 1.2** below outlines the timeframe for key developmental stages of eye development in humans and in chick (Hamburger-Hamilton staging) (Hamburger & Hamilton, 1951).

**Table 1.2: Key developmental stages in vertebrate eye development**

Eye developmental stage	Days gestation (human)	Embryonic stage (chick)
Eye field specification	<22	HH4
Optic vesicle evagination	22	HH9
Lens Placode formation	28	HH12
OV and LP invagination	32	HH14
OC, start of retinal neurogenesis	33	HH15

The optic vesicle is derived from neural ectoderm and lens derived from non-neural ectoderm. Both intrinsic and extrinsic signals between these tissues are carefully regulated in a specific spatial-temporal manner to successfully co-ordinate eye morphogenesis. This report will review the initial steps of eye development focusing on patterning and splitting of the eye field, these stages are highlighted in red (**Figure 1.2**).

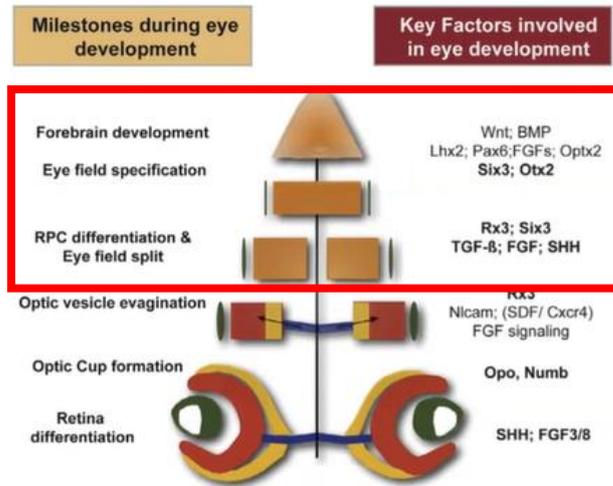


Figure 1.2: Scheme of vertebrate eye development. The central line represents both a timeline of development and the ventral midline. Orange represents eye field/retina. Blue represents optic stalk. Yellow represents pigmented retinal epithelium. Red represents neural retina. Green represents lens. Red box highlighting initial developmental stages. Adapted from (Sinn & Wittbrodt, 2013).

### 1.2.1 Neural Induction

While the first morphological indication of eye development occurs later in development when the bilateral optic vesicles evaginate from the diencephalon (forebrain) (Chow & Lang, 2001). Many steps required beforehand as the eyes are an extension of the central nervous system (CNS). In vertebrate embryos, eye development begins shortly after the onset of gastrulation, with the single eye field developing anteriorly to the primitive node, (Hensen's node in birds; Spemann's organiser in amphibians), in the anterior neural plate (Sinn & Wittbrodt, 2013). First, neural induction is required to give neural plate. This is also the first step in forebrain development. The ectoderm forms neural ectoderm in the dorsal region caused by a downregulation of Bone Morphogenetic Protein (Bmp) and Fibroblast Growth Factor (Fgf) signalling in competent ectoderm by the node producing neuralizing factors (Delaune et al, 2005). Molecules such as Chordin, Noggin, Follistatin, Cerberus and *xnr3* antagonize the Bmp signalling pathways in the process of neural induction (Nieuwkoop, 1963; Weinstein & Hemmati-Brivanlou, 1999).

### 1.2.2 Anterior-Posterior Patterning of the Neural Plate

Following neural induction, the neural plate, is then anteriorised by inhibition of Wnt/ $\beta$ -catenin signalling via Wnt antagonists (Lupo et al, 2014) (**Figure 1.2.2.1**). The forebrain acts as a default positional identity with the Wnt/ $\beta$ -catenin signalling activated in a dose-dependent gradient with high levels of *Wnt1*, *Fgf8* and retinoic acid (RA) in more posterior regions (**Figure 1.2.2.1A**) (Andrews & Nowakowski, 2019). This splits the neuroectoderm into four main domains of the forebrain (telencephalon (rostral) & diencephalon (caudal)), midbrain (mesencephalon), hindbrain (rhombencephalon) and the spinal cord (**Figure 1.2.2.1B**).

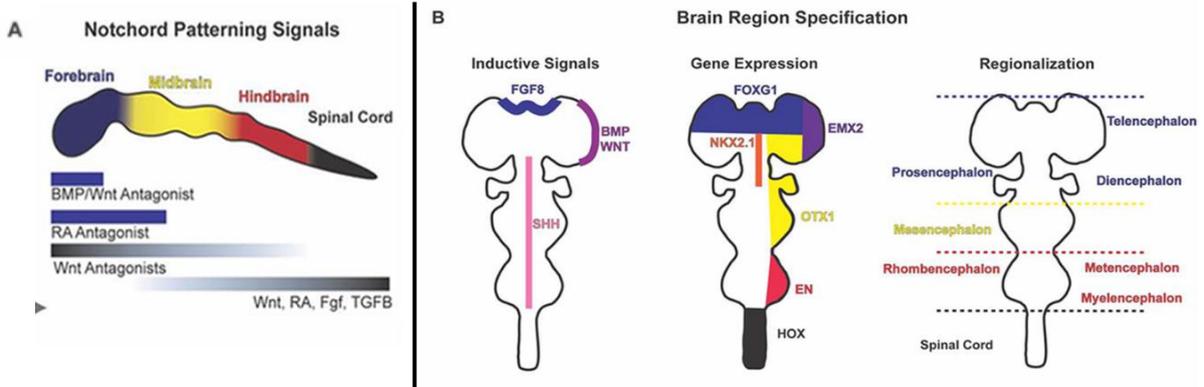


Figure 1.2.2.1: Neural Development. (A) During early development, notochord cells induce neuroepithelial identity of the overlying epithelium. The notochord becomes patterned along the rostral/caudal axis. (B) Morphogens pattern the neural tissue resulting in the discrete TF domains which develop into distinct regions of the CNS. Adapted from (Andrews et al, 2019).

### 1.2.3 Eye Field Specification (within the anterior neural plate)

The eye field is situated in the anterior neural plate, with the inhibition of Bmp signalling leading to an up-regulation of the transcription factor (TF) Orthodenticle homeobox 2 (*Otx2*) and posterior Wnt signals restricting *Otx2* to the anterior neural plate region. The eye field develops as part of the medial prosencephalon, which is now respecified to a more posterior diencephalic fate following Wnt/ $\beta$ -catenin signalling (Heisenberg et al, 2001). *Otx2* is required for forebrain specification and acts as a marker for anterior neural plate (Zuber et al, 2003). This then causes the co-expression of other TFs, known as eye field transcription factors (EFTFs), in the anterior neural plate region which characterise the eye field. These EFTFs include, but are not limited to, *Paired box 6 (Pax6)*, *Retinal homeobox protein (Rx)*, *Six3 homeobox 3 (Six3)* and *LIM homeobox 2 (Lhx2)* which are all known to regulate one another (Zuber et al, 2003). The sequence in which these genes are expressed varies from species to species.

The prechordal mesendoderm (Figure 1.2.3), which underlies the anterior neural plate (i.e. the medial prosencephalic neural plate), is required for the splitting of the eye field and the development of the hypothalamus (ventral diencephalon). Sonic hedgehog (Shh) protein secreted from these axial mesendoderm tissues act as ventralizing factors forming the ventral midline.

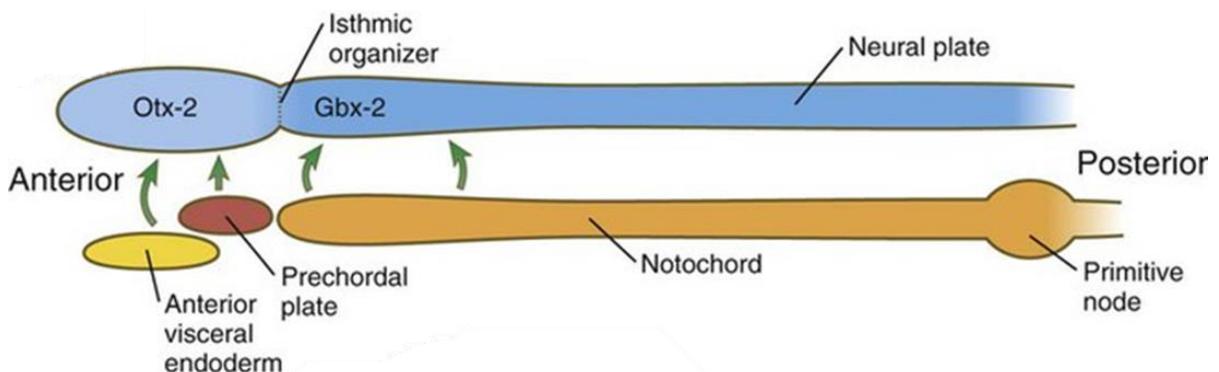


Figure 1.2.3: Diagram showing the different tissues making up the prechordal mesendoderm. The prechordal mesendoderm is the rostral-most portion of the axial mesendoderm; it is continuous with the notochord, which is the corresponding axial mesendodermal structure that underlies more caudal portions of the neural plate. Green arrows show Shh signals.

### 1.2.4 Eye Field Division

Further *Shh* expression in the medial neural plate and the ventral neural tube act to pattern the embryo. With the ventralization of the medial neural tube downregulating certain EFTFs and upregulating others. In *Xenopus* the EFTF *Pax6* is expressed throughout the eye field with *Shh* signalling at the midline downregulating *Pax6* ventrally but up-regulating *Pax2* (Zuber et al, 2003). This gives two lateral domains of *Pax6* resulting in splitting of the eye field. It has been shown in mouse explants (Shimamura & Rubenstein, 1997) that the *Shh* signals emanating from these prechordal tissues also induce *Nkx2.1* expression in the medial portion of the anterior neural plate to give hypothalamus. The inhibition of *Shh* signalling in the more dorsal regions of the prosencephalon gives telencephalon. However, in amniote animal models, such as chick, the EFTF *Pax6* is not expressed uniformly across the anterior neural plate during eye field patterning, but is activated later when optic vesicle outgrowth occurs (Li et al, 1994). This then raises two questions: what drives optic vesicle outgrowth if not *Pax6* and what actually causes the eye field to split and generate paired optic vesicles?

*Pax6* is known across the animal kingdom as the 'master control gene' for eye development, its function is highly conserved being both necessary and sufficient for eye development (Gehring, 1996). *Pax6* can induce ectopic eyes which supports the idea of its critical role in eye development (Halder et al, 1995). Loss-of-function mutations in mice can lead to the 'small eye' phenotype when in the heterozygous state and complete loss of the eye in the homozygous state (Hill et al, 1991). With similar mutations in human known to cause the ocular syndrome anophthalmia (no eye), microphthalmia (small eye) and aniridia (where the iris is undeveloped or missing) (Glaser et al, 1992; Hanson et al, 1993).

### 1.2.5 Optic Vesicle Evagination

Once the eye field has split the presumptive optic vesicles evaginate from this neuroepithelium towards the overlying surface ectoderm (**Figure 1.2.5.1**). This occurs before the most anterior ends of the neural tubes, known as the anterior neural ridge, fuse. This process is thought to be initiated by induction of specific EFTFs including *Rax* (*Rx3* in fish, *Rx1* in *Xenopus*) (Mathers, 19997) and *Six3* (Carl et al, 2002) throughout the neural ectoderm within the anterior neural plate.

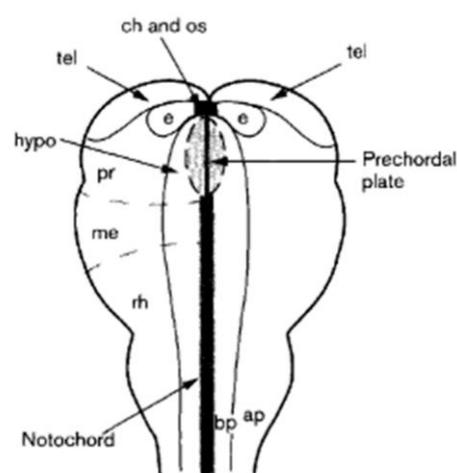
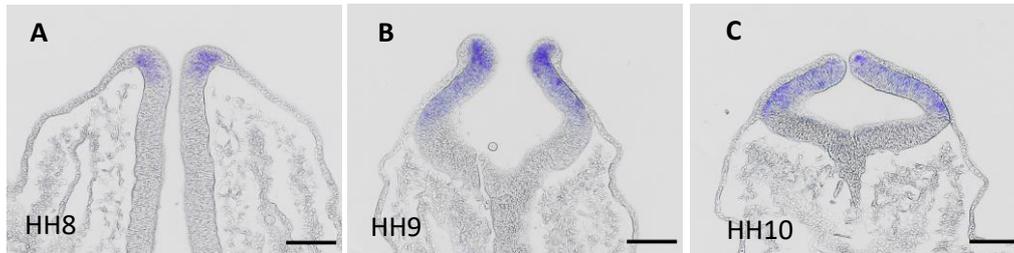


Figure 1.2.5.1: Scheme showing regionalization of vertebrate neural plate and neural tube. Shows axial tissues and the approximate location of several forebrain primordia. *Shh* signalling along the ventral midline induces hypothalamus (hypo) with optic vesicles evaginating from the lateral walls. (Rubenstein & Beachy, 1998).

Which somehow results in outgrowth of optic vesicles. This process is poorly understood in amniotes. However, it has recently been thought that there is a putative growth zone in the most anterior part of the neural folds regulated by the gene *c-Myc* (Grocott et al, unpublished data) (**Figure 1.2.5.2**). Proliferation of cells within this growth zone of the anterior neural folds could be responsible for the enlargement of the optic vesicles. However, it is unknown whether these cells are even retinal progenitor cells (RPCs).



**Figure 1.2.5.2: Horizontal sections of *c-Myc in situ*.** Expression can be seen in the anterior neural folds of the neural ectoderm tissue. (A) Stage HH8, (B), Stage HH9, (C) Stage HH10 of chick development.

The proto-oncogenes, *c-myc* and *n-myc* have well known roles in regulating cell proliferation, cycle progression, growth and survival (Eilers & Eisenman, 2008). Neural crest cells, known for maintaining their self-renewing multipotency, were shown to be regulated by *c-Myc* (Kerosuo & Bronner, 2016).

Later in development of the optic cup, it has been shown that *c-Myc* is a marker for slow-dividing progenitor cells, helping cells maintain more stem-cell-like characteristics in the peripheral margins while *n-Myc* is a marker for subsequent progenitor cells (Xue & Harris, 2012). It is thought that *c-Myc* cells in the anterior neural folds may delineate paired growth zones that contribute to optic vesicle outgrowth, as shown in **Figure 1.2.5.2**, where *c-Myc in situ* confirmed expression of the gene in the neural ectodermal tissue of the anterior neural folds (Grocott et al, unpublished data).

In fish optic vesicle outgrowth seems to be driven by specific migratory behaviour of retinal progenitor cells (RPCs) (Rembold et al, 2006) rather than proliferation. For fish optic vesicle formation begins during neurulation when cells, including RPCs, of the neural plate migrate towards the midline to form the neural keel (Martinez-Morales & Wittbrodt, 2009). The RPCs once converged at the midline, turn around to migrate into the developing optic vesicles with lateral diencephalic cells moving anteriorly into this area of evagination where they intercalate, epithelialize and contribute to optic vesicle enlargement (England et al, 2006) (Rembold et al., 2006). It is thought that downstream targets of *Rx* genes influence cells' migration, however the exact transcriptional targets of *Rx* genes have yet to be identified. However, the fate map generated by England et al supports our growth zone hypothesis. With telencephalon fated cells arising from a thin crescent shape region in the anterior of the embryo. In comparison with chick this would equate to optic vesicle outgrowth from the anterior neural folds, driven by Myc+ cells.

### 1.2.6 Optic Cup Formation

The neural epithelium tissue that is now the optic vesicle makes contact with the surface ectoderm (**Figure 1.2.6.1C**). This period of close contact where an exchange of inductive signals from the surrounding tissues occurs is extremely important to coordinate self-assembly of the eye to give retina and induce lens (Gunhaga, 2011). The mechanics behind the morphogenesis from optic vesicle to optic cup and eventually more complex 3D structures of retina and lens is still not completely understood.

When the distal portion of the optic vesicle makes contact with the surface ectoderm it results in the specification of the lens placode, and further differentiation and invagination leads to the induction of lens and a fully formed vertebrate camera eye. Contact with the surface ectoderm causes the optic vesicles to invaginate inwards, to form the bi-layered optic cup (**Figure 6D**). The inner layer of the optic cup differentiates into neural retina, which further differentiates into photoreceptors and six other major cell types.

The outer layer forms the thinner retinal pigmented epithelium layer. Regionalization of the optic vesicle into neural retina (NR) and the retinal pigmented epithelia (RPE) occurs concurrently with dorsal-ventral/proximal-distal patterning along the evaginating optic vesicle and concludes with the closure of the optic fissure to form the optic cup (Zagozewski et al, 2014).

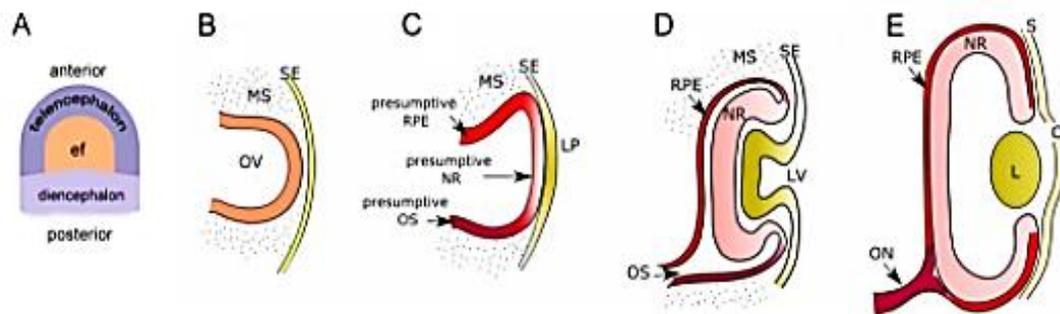


Figure 1.2.6.1: Schematic representation of vertebrate eye development. (A) Specification of the eye field (ef) within the anterior neural plate. (B) Formation of the optic vesicle (OV). (C) Specification of the retinal pigmented epithelium (RPE), neural retina (NR) and optic stalk (OS) domains within the OV and formation of the lens placode (LP) from the surface ectoderm (SE). (D) Formation of the optic cup (OC) and the lens vesicle (LV). (E) Mature optic cup and lens. (Taken from Adler & Canto-Soler, 2007).

Perturbations in mechanical signalling have resulted in the optic cup failing to fold and fuse, which causes the developmental disorder Coloboma (Miesfeld et al, 2015). In aquatic vertebrates, where these two layers of NR and RPE join at the lip of the cup, is a stem cell niche called the ciliary marginal zone (CMZ) (Fischer et al, 2013), controlled by the *c-Myc* gene (**Figure 1.2.6.2**). This process is the same as described earlier for the putative growth zone thought to be found in the anterior neural folds, but in the CMZ. Stem cells within this niche divide asymmetrically giving rise to a *c-Myc* expressing stem cell-like cells and a fast-dividing *n-Myc* expressing transient amplifying cell. These *n-Myc* cells go on to form more mature central retina structures, highlighted in **Figure 1.2.6.2**. So, all subsequent growth of the eye occurs from the peripheral optic cup lip, as shown in chick (Kitamoto & Hyer, 2010).

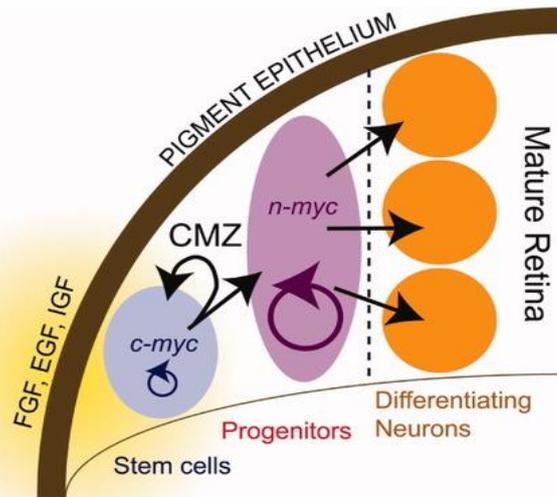


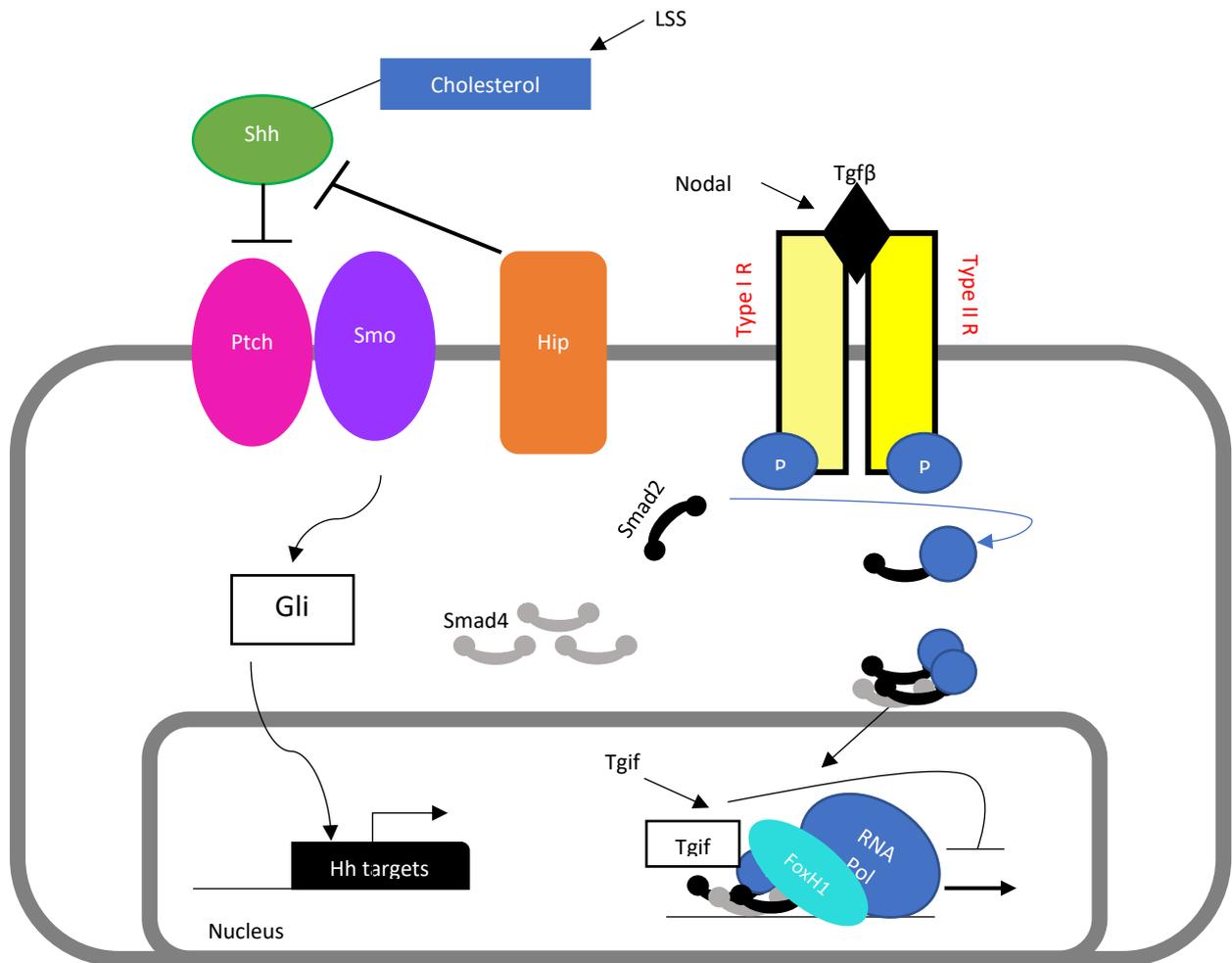
Figure 1.2.6.2: Schematic showing asymmetric division of *c-Myc* cells in the CMZ. Cells residing in the peripheral margins maintain their stem-cell like characteristics (marked by *c-Myc*) and cycle slowly while progenitor cells (marked by *n-Myc*) cycle more quickly and differentiate into mature retina. (Adapted from Xue & Harris, 2012).

### 1.3 Holoprosencephaly (HPE; MIM# 236100)

Holoprosencephaly (HPE) is a congenital disorder broadly described as the failure of the forebrain (prosencephalon) to divide into two cerebral hemispheres. There are four main classes of HPE: alobar, semilobar, lobar and middle interhemispheric variant, in order of increasing separation of the brain and therefore most-to-least severe. HPE is the most common congenital anomaly of the brain to occur during embryogenesis (Aguilella et al, 2003). It occurs in 1 in 10,000-16,000 births with the infant being still born or dying shortly after birth (Dubourg et al, 2007) and the cause of 1 in 250 miscarriages (Winter et al, 2003). The most severe HPE cases are usually detected by routine ultrasounds during pregnancy. Genetic counselling is used to determine the severity of HPE, upon which termination is often advised for the most severe forms. The most severe form, alobar HPE, is where Cyclopia is categorised.

Cyclopia is caused by the failure of the eye field to split into bi-laterally symmetric optic vesicles during normal eye development. This results in the formation of a singular central eye orbit, and in most cases, facial abnormalities. Such as a proboscis, the fleshy part of the nose which fails to descend between the divided eye field; median or bilateral cleft lip/palate in severe forms and ocular hypotelorism or solitary median maxillary central incisor in minor forms (Winter et al, 2013).

Classical HPE is thought to be caused by the failure of the ventral forebrain neuroectoderm to transduce the Shh signal produced by the underlying region of the prechordal plate. This process can fail in three ways: if the prechordal plate simply fails to form and no Shh signal is produced; if the Shh signal produced by the prechordal plate is faulty e.g. knocked-out; or if the neuroectoderm cannot transduce the Shh signal (Geng & Oliver, 2009). However, it has since been shown that Nodal/Transforming growth factor beta ( $Tgf\beta$ ) signalling (**Figure 1.3**) is also a major signalling pathway and that when disrupted is causative for HPE (Gripp et al, 2000). While Shh signalling is required to induce cells in the initial medial eye field to a hypothalamic fate, Nodal signalling is required for the formation of the prechordal plate.



**Figure 1.3: Signalling pathways involved in HPE.** Shh signalling pathway shown on the left. The Shh protein undergoes cholesterol-mediated autocatalytic cleavage. Shh signalling is activated by a double-negative mechanism. In the presence of Shh bound to Ptch, the negative effector, Smo activity is released and as a result Smo becomes phosphorylated and activated which inhibits a downstream phosphorylation cascade which ends with a full length Gli protein (Ci for invertebrate) protein. Which translocates into the nucleus and functions as a transcriptional activator.

Nodal/TGF $\beta$  pathway shown in the right. Nodal is a TGF $\beta$  superfamily member and signals through a receptor complex including the type I and type II receptors. Receptor activation results in phosphorylation of Smad2 and Smad3, and activation of genes required for PCP induction. FoxH1 is a co-transcriptional factor of Smad2 and Smad4, involved in TGF $\beta$ , activin and nodal signalling pathways. Tgif1 can regulate downstream Nodal/TGF $\beta$  signalling. It does this repressing Smad proteins.

Not all parts of these pathways are in the same tissue or cell type. (Adapted from Dubourg et al, 2007).

It has been suggested that Cyclopia can also occur after optic vesicle evagination. Shh induces the optic stalk marker Pax2 and represses the retinal marker Pax6. The loss of Pax2 expression triggers the expansion of pax6 expression medially, inducing retinal fate at the expense of optic stalk fate, leading to fusion of the bilateral eye vesicles (Ekker et al, 1995; Macdonald et al, 1995). However, a pair of optic vesicles is still generated, therefore not fitting under the true definition of Cyclopia.

HPE can be caused by environmental or metabolic factors. In animals, toxins such as steroidal alkaloids ingested via grazing contaminated vegetation throughout gestation have been known to cause the Cyclopic phenotype (Desesso, 2019). This causative agent was thought to be disrupting the

Shh signalling pathway. Certain drugs and infections have also been associated with HPE development in humans (Repetto et al, 1990; Frenkel et al, 1990). However, the only formally recognised environmental causes are: diabetes mellitus with a 1% risk of HPE (Barr et al, 1983) and maternal alcoholism (Goswami & Kusre, 2015). However, with diabetes mellitus as an environmental factor it is often associated with chromosomal abnormalities rather than isolated HPE (Chen et al, 2015). Maternal alcoholism was thought to disturb Shh signalling, but in mice it has been shown that foetal alcohol exposure synergizes with a genetic mutation in the known causative gene *Cdon* to disrupt Nodal signalling, causing midline patterning defects similar to those seen in humans (Hong et al, 2020).

HPE can be due to chromosomal abnormalities, part of other systemic malformations or an isolated brain disorder. Chromosomal analysis led to the identification of 12 candidate regions across 11 genes (assigned HPE1 to HPE12) that contain or are thought to contain genes involved in HPE. However, this review will focus on the isolated form of HPE and the causative genes thought to be involved. The candidate region they fall within is shown in **Table 1.3**.

**Table 1.3: List of causative genes for HPE**

Gene Name	Chromosome Locus	HPE phenotype
Six3	2p21	HPE-2 (Schell et al, 1996)
Shh	7q36.3	HPE-3 (Belloni et al, 1996)
Tgif1	18p11.231	HPE-4 (Gripp et al, 2000)
Zic2	13q32.3	HPE-5 (Brown et al, 1998)
Ptch1	9q22.32	HPE-7 (Ming et al, 2002)
Gli2	2q14.2	HPE-9 (Roessler et al, 2003)
Cdon	11q24.2	HPE-11 (Bae et al, 2011)
Cnot1	16q21	HPE-12 (De Franco et al, 2019)
Stag2	Xq25	HPE-13 (Kruszka et al, 2019)

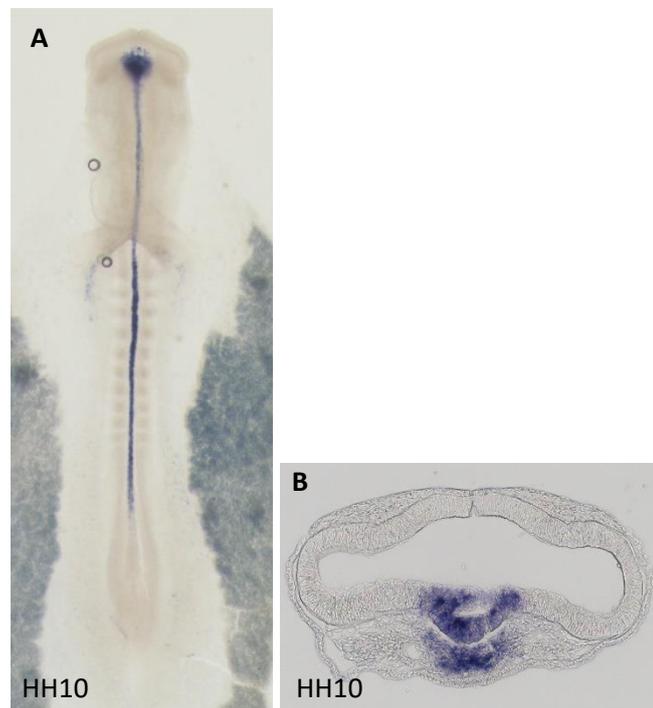
To date, nine genes have been positively implicated in the isolated form of HPE in humans: *Shh*, *Zic2*, *Six3*, *Tgif1*, *Ptch1*, *Gli2*, *Cdon*, *Cnot1* and *Stag2*. Genes thought to be involved include *Disp1*, *Chrd*, *Bmp4* and others (Dubourg et al, 2007). *Shh* was the first gene identified through recurrent chromosomal arrangements (Pfitzer and Muntefering, 1968; Pfitzer et al, 1982; Belloni et al, 1996). All subsequent genes identified were through studying the Shh signalling pathway or the Nodal/TGF $\beta$  pathway. Many are all highly expressed within or adjacent to the anterior neural folds (Grocott et al, Unpublished Data).

When looking at the genetics of isolated HPE. There are thought to be 4 main HPE genes: *Shh*, *Zic2*, *Six3* and *Tgif1*. These 4 genes account for approximately 20% of isolated HPE cases (Bendavid et al, 2006).

### 1.3.1 *Sonic hedgehog (SHH)* (MIM# 600725; 7q36; HPE3; see Table 1.3)

*Shh* is the primary gene implicated in holoprosencephaly (12.7% of HPE cases) through either large deletions in 7q36 or point mutations (Dubourg et al, 2004; Bendavid et al, 2006). With a wide variety of phenotypes observed.

As mentioned previously the secreted Shh protein acts as an inductive signal that sets up a morphogen gradient to ventralize the neural tube. *Shh* is expressed throughout the notochord and the floorplate of the neural tube during early embryo development, as shown in **Figure 1.3.1**. In later stages it is expressed in the posterior limb buds and the gut (Odent et al, 1999). The human *Shh* gene has three exons and encodes a 462 amino acid polypeptide (45kDa). This synthesised protein is a precursor molecule which is modified post-translationally through cholesterol-mediated autocatalytic cleavage to generate diffusible Shh signalling molecules (Lee et al, 1994; Porter, 2006; Magnaldo, 2002).



**Figure 1.3.1:** Expression of Shh along the ventral midline axis. **(A)** Expression of Shh in the midbrain, hindbrain, neural plate and notochord at stage HH10. **(B)** Transverse section reveals expression of Shh in ventral neural tube, floor plate and notochord. (Grocott et al, unpublished data).

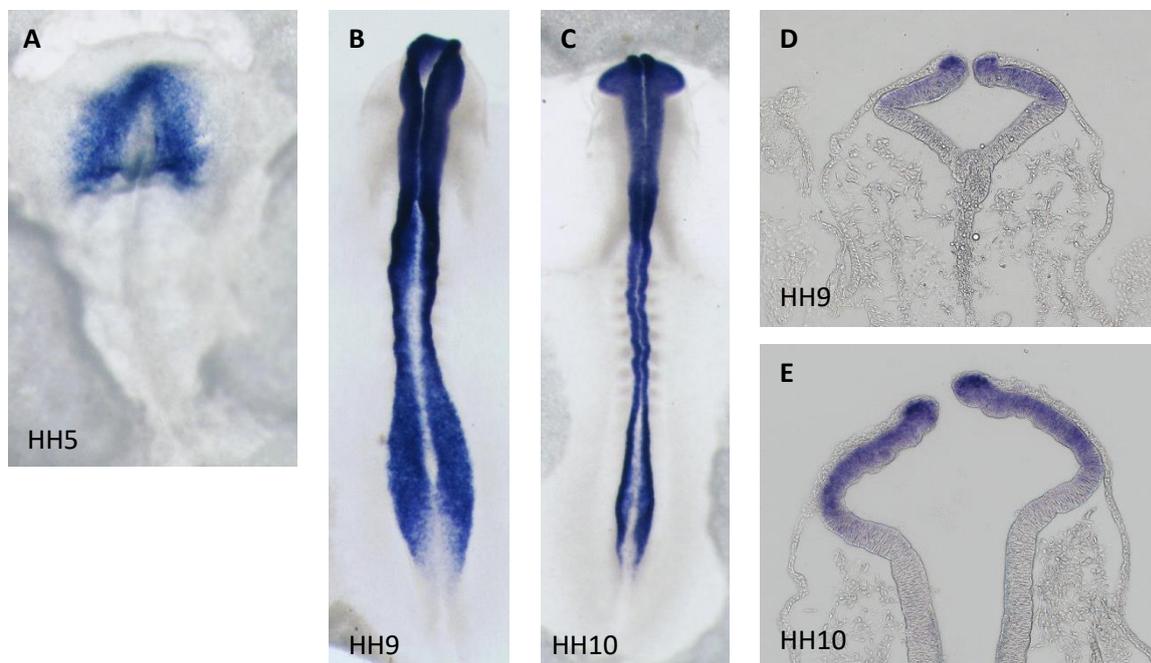
Shh is secreted from producing cells to reach the surface of target cells where it interacts and binds to the transmembrane receptor Patched (Ptch). Shh signalling is activated by a double-negative mechanism. In the absence of Shh binding to its receptor Patched (Ptch), the kinases PKA, GSK3 and CK1 phosphorylates the downstream Gli proteins, and this phosphorylation signals a proteolytic cleave of the Gli protein leaving it with an N-terminal repressor domain (Cohen et al, 2015). The Gli-repressor proteins then translocates to the nucleus and function as a transcriptional repressors.

When Shh binds to and inhibits Ptch (**Figure 1.3.1**) it results in the activation of the Shh signalling pathway, relieving inhibition of transmembrane protein Smoothed (Smo) (which becomes phosphorylated), which in turn prevents the proteolytic cleavage of Gli proteins. The resulting full-length Gli-activators then drive transcriptional activation of *Shh* target genes which include *Ptch*, *Wnt*, *Bmps* or *Gli* transcriptional factors (Cohen et al, 2015).

### 1.3.2 *Zic2* (MIM# 603073; 13q32; HPE5; see Table 1.3)

The second main HPE gene identified was *Zic2* which encodes a transcription factor that contains zinc finger DNA binding motifs similar to the Gli protein binding domains. *Zic2* can act as a transcriptional regulator when interacting with Gli proteins (Mizuguchi et al, 2001) and aids in chromatin remodelling of the cerebellum (Frank et al, 2015).

*Zic2* plays an important role during neuralation. *Zic2* is transiently expressed at the midline organiser (Primitive node in humans; Hensen's node in birds; Spemann's organiser in amphibians) (Elms et al, 2004) (**Figure 1.3.2**) and is required for the development of the mesendodermal structures such as the prechordal plate and notochord. *Zic2* is required for the epiblast cell to transit through the node and to differentiate into this axial mesendoderm and migrate to the anterior midline during mid-gastrula stage embryos (Houtmeyers et al, 2016). Nodal signalling is also required for prechordal plate establishment. However, this is at an earlier stage required for the formation of the anterior primitive streak. The *Zic2* protein is able to physically interact with stimulated Smad transcriptional mediator complexes of the Nodal signalling pathway via regulation of *Forkhead Box A2* (*Foxa2*) transcription during axial mesendoderm formation (Houtmeyers et al, 2016). This shows that *Zic2* does not act downstream of ventral neuroectoderm *Shh* signalling during neuralation but instead acts downstream of Nodal signalling during gastrulation. It is thought that *Zic2* directly controls *Foxa2* expression, which is known to control *Shh* expression (Jeong & Epstein, 2003). In the case of *Zic2* mutants the prechordal plate does not form and no *Shh* signal is produced. As the disruption to signalling happens earlier on in the developmental timeline than that in other known causative genes it can be assumed that a more severe HPE phenotype will arise. As is the case in mutated mice homozygous for the *Zic2* allele which exhibit the Cyclopic-like phenotype (Nagai et al, 2000), and a missense mutation in *Zic2* caused impaired prechordal plate development (Warr et al, 2008) and impaired Nodal signalling (Houtmeyers et al, 2016).



**Figure 1.3.2: Expression of *Zic2* gene at multiple stages of development. (A)** Expression seen in the epiblast of the neural plate at stage HH5. **(B-C)** Expression along the neural folds concentrated at the most anterior region in stages HH9 and HH10 respectively. **(D-E)** Transverse sections reveal *Zic2* expression in the anterior neural folds of the neural ectoderm in stages HH9 **(D)** and HH10 **(E)**.

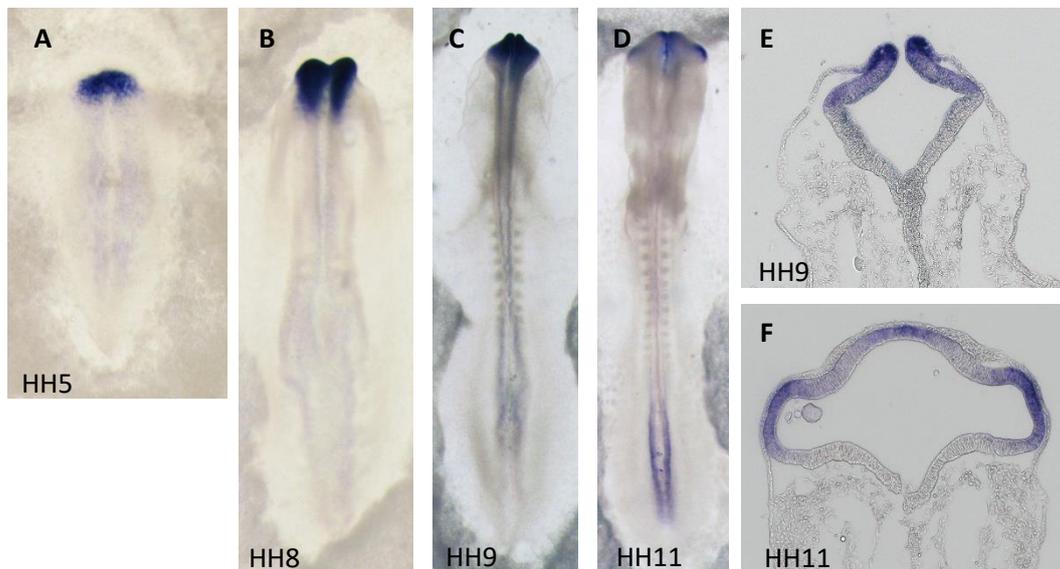
*Zic2* is also a direct regulator of *Tgif1* expression, another causative HPE gene. It can bind to a region upstream of *Tgif1* and it is strongly suggested that this binding is essential for *Zic2*-dependent transcriptional activation of *Tgif1* (Ishiguro et al, 2018). This linkage between the two genes is thought to be involved in forebrain development and HPE.

*Zic2* abnormalities account for 9.2% of HPE cases, with 75% of those in classic HPE with severe alobar or semilobar presentations (rather than non-classic MIHV presentation) (Gounongbe et al, 2020).

### 1.3.3 *Six homeobox 3 (Six3)* (MIM# 603714; 2p21; HPE2; see Table 1.3)

*Six3* is a homeobox-containing gene which is homologous to the *Drosophila sine oculis* gene; it is vital in the development of forebrain and eye development. Post neuralation, *Six3* acts as a transcriptional activator, positively regulating the transcription of *Shh* to maintain expression (Geng et al, 2008). It has also been shown to act a repressor, downregulating the expression of *Wnt* and *Bmp* and protecting the anterior neural ectoderm from their posteriorizing affects (Gestri et al, 2005; Liu et al, 2010). This is achieved by interaction with members of the *Groucho (Grg)* family of transcriptional corepressors (Zhu et al, 2002) and is critical in establishing a negative feedback regulatory loop for specification of anterior neuroectoderm. This auto-repression of *Six3* occurs by *Six3* binding to its own promoter when interacting with *Grg* proteins during development. *Six3* acting as a repressor also shows a functional role in retina and lens differentiation (Zhu et al, 2002).

*Six3* is also a direct upstream regulator of *Shh* expression. *Six3* is involved in a feedback loop with *Shh* in the ventral forebrain. *Shh* signals from the underlying axial mesoderm inducing *Six3* expression in the overlying anterior neural ectoderm (**Figure 1.3.3**). This then directly activates neural ectoderm expression of *Shh* along the ventral midline which in turn maintains *Six3* expression (Geng et al, 2008). Reduced *Six3* protein along the diencephalic ventral midline has been shown to fail to activate *Shh* expression in the neural ectoderm above the prechordal plate resulting in HPE (Lagutin et al, 2003). Mutations in *Six3* account for 1.3% of all HPE cases (Cohen, 2006). They are generally found in severe phenotypes.

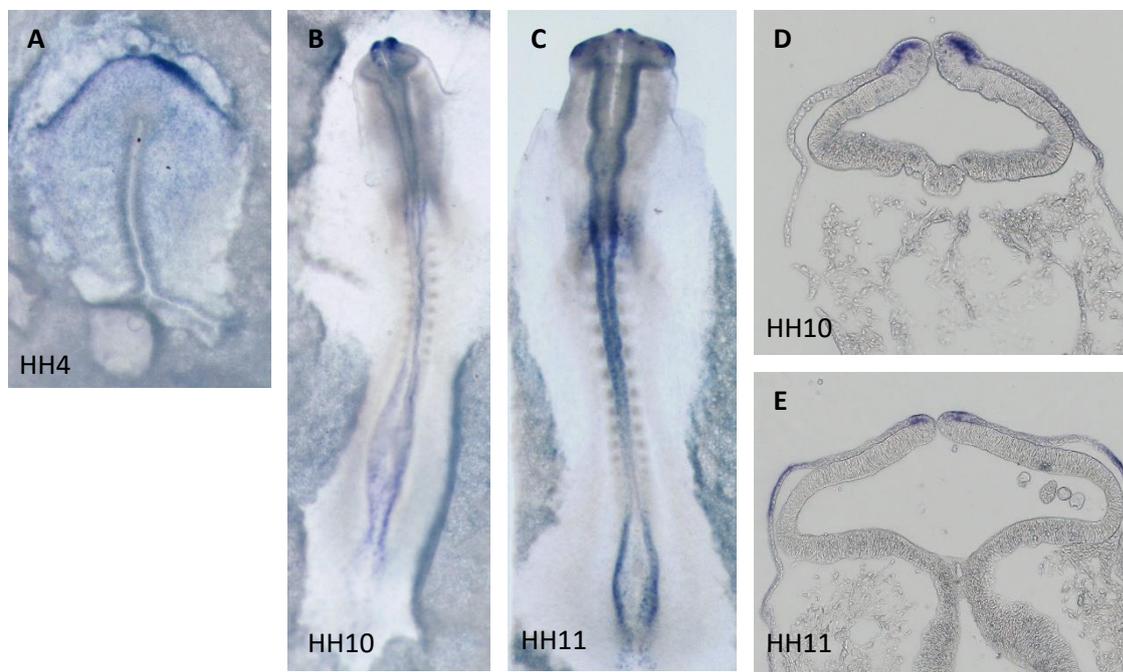


**Figure 1.3.3:** *Six3* expression data shown at different developmental stages. (A) Expression of *Six3* in the epiblast and early anterior mesoderm/endoderm at stage HH5. (B-C) Expression concentrated in the anterior neural folds at stage HH8 (B) and HH9 (C). (D) *Six3* expression seen in the forebrain and distal optic vesicles. (E) Transverse section shows *Six3* expression in the anterior neural folds in the neural ectoderm and the most anterior part of the overlying surface ectoderm in stage HH9. (F) Transverse section showing expression in the neural ectoderm of the forebrain and optic vesicle structures in stage HH11.

### 1.3.4 Transforming Growth Factor-Beta-Induced Factor 1 (*Tgif1*) (MIM# 602630; 18p11.3; HPE4 Table 1.3)

*Tgif1* is another homeodomain TF (encoding a member of the three amino acid loop extension (TALE) family) that regulates downstream signalling of Nodal/TFG $\beta$  pathways. It does this by acting as a corepressor and blocking the action of Smad proteins (**Figure 1.3.4**) (Gripp et al, 2000). *Tgif1* is perhaps the least well known main causative gene in regard to its function in neural development. *Tgif1* is also involved in the down-regulation of RA signalling by binding to the enhancer element of RA-controlling genes (Castillo et al, 2010).

As mentioned earlier *Tgif1* is a downstream target of *Zic2* (Ishiguro et al, 2018). Expression of *Zic2* and *Tgif1* overlap in the neural epithelia of the dorsal forebrain, midbrain and regions of the optic vesicle (**Figure 1.3.2 & 1.3.4**). They are both also highly expressed in earlier stages of development in the neural plate in the anterior neural folds. However, whereas *Zic2* expression is seen in the neural ectoderm, *Tgif1* expression is present in the surface ectoderm layer (**Figure 1.3.4 D-E**). When *Zic2* is knocked-down *Tgif1* protein level is reduced in the head region (Ishiguro et al, 2018).



**Figure 1.3.4: Expression of *Tgif1* data at multiple stages of development. (A)** Expression of *Tgif1* along the anterior epiblast at stage HH4. **(B)** Expression concentrated in the anterior neural folds at stage HH10 **(C)** *Tgif1* expression seen in the neural folds and presumptive lens and olfactory surface ectoderm. **(D)** Transverse section shows *Tgif1* expression in the anterior neural folds in the surface ectoderm at stage HH10. **(E)** Transverse section showing expression in the neural ectoderm of the neural folds and optic vesicles in stage HH11.

Both *Zic2* and *Tgif1* indirectly regulate Shh signalling through the regulation of Gli family proteins and retinoic acid (RA) signalling which are involved in the posteriorizing of the brain. *Tgif1* mutations leads to reduced *Tgif* function which can leave the embryo sensitized to the effects of retinoic acid, increasing the severity and penetrance of the posterior transformation phenotype (Melhuish et al, 2016). Similarly, if *Tgif2* is also mutated more severe brain developmental defect and axial patterning mutations are seen, with the case of double null embryos failing to complete gastrulation (Melhuish et al, 2016). But, at least in mice, loss of *Tgif1* function alone does not always have the severe HPE phenotypic consequences (Shen & Walsh, 2005). The variable expression of phenotypes

is thought to be influenced by the *Tgif2* gene acting redundantly when *Tgif1* expression is knocked-down. But in chick, *Tgif1* and *Tgif2* do not have the same expression patterns along the neural tube (Knepper et al, 2006).

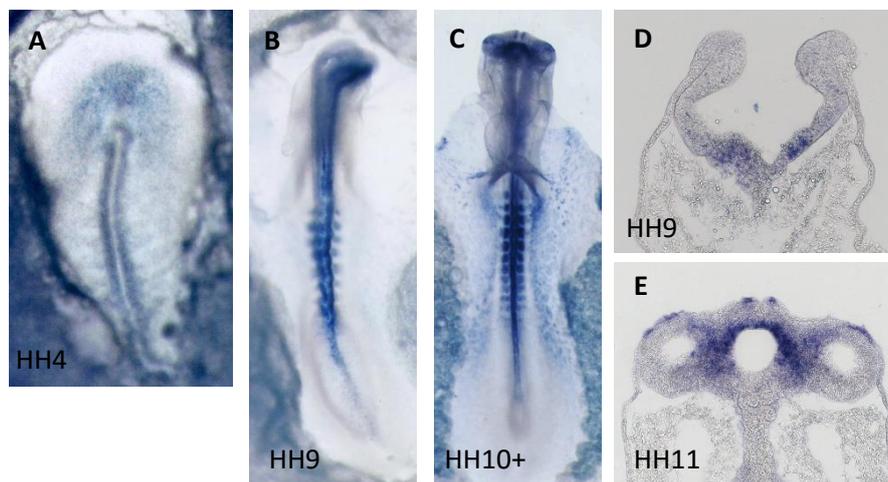
*Nkx2.2* is expressed in the ventral neural tube as a hypothalamus marker. *Tgif1* positively regulates the expression of genes expressed in the dorsal domains but not in this ventral region. This may be due to *Pax6* expression, which has a mutually repressive effect on *Nkx2.2* and a boundary may already be established that *Tgif1* cannot extend upon (Ericson et al, 1997). Mutations in *Tgif1* were also shown to decrease the ability to repress Tgf $\beta$  signalling. While it is clear *Tgif1* has an important role in neural development its exact function is still unknown.

Mutations in *Tgif1* account for 1% of HPE cases. The data suggests that *Tgif1* alterations only account for a small portion of HPE cases as only 10% of deletion mutations of the *Tgif1* gene present the HPE phenotype (Maranda et al, 2006).

### 1.3.5 Causative genes involved in the Shh signalling pathway (*Ptch1*, *Gli2*, *Cdon*, *Disp1*)

These four causative genes: *Patched1* (*Ptch1*) (MIM# 601309; 9q22; see **Table 1.3**), *Gli-Kruppel family Member 2* (*Gli2*) (MIM# 165230; 2q14; see **Table 1.3**) *Cell adhesion molecule-downregulated by oncogenes* (*Cdon*)(MIM# 608707; 11q24; see **Table 1.3**) and *Dispatched 1* (*Disp1*)(MIM# 607502; 1q41; see **Table 2**) are all part of the Shh pathway but have a very low rate of HPE causation when mutated alone.

As mentioned previously *Ptch1* is the ligand receptor of Shh (**Figure 1.3**). When unbound to Shh it acts to repress Shh signalling. Mutations in *Ptch1* can affect its ability to bind to Shh or they could enhance the repressive activity that *Ptch1* has on the signalling pathway (Ming et al, 2002). In either of these cases Shh signalling of the prechordal plate cannot occur and ventralization of the midline fails. **Figure 1.3.5.1** below shows *Ptch1* expression patterns as it acts as a Shh receptor.

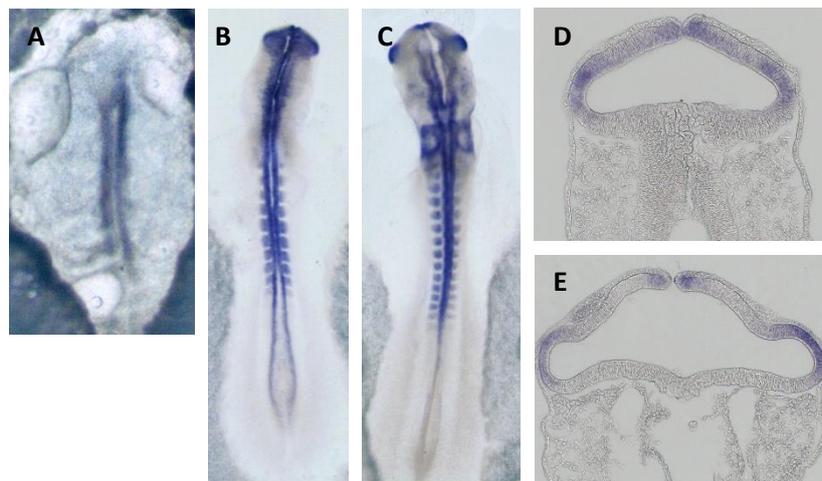


**Figure 1.3.5.1: Expression of *Ptch1* at multiple stages of chick development. (A)** Expression of *Ptch1* in the epiblast and early mesoderm anterior the Hensens node at stage HH4. **(B)** Expression concentrated along the neural tube and somites at stage HH9 **(C)** *Ptch1* expression concentrated along the neural tube, somites, forebrain and midbrain. Expression also seen in the optic vesicles at stage HH10+. **(D)** Transverse section shows *Ptch1* expression in the posterior neural folds in the neural ectoderm at stage 9. **(E)** Horizontal section showing expression in the ventral neural ectoderm of the neural tube and optic vesicles in stage HH11, in addition to the presumptive lens and olfactory surface ectoderm.

*Gli2* is a downstream target of *Shh*. It's one of the three zinc finger TF's that acts as a mediator of Shh signal transduction. In the absence of Shh signal the full-length *Gli2* is marked for phosphorylation and proteolytic cleavage of the c-terminal activator domain, creating the repressor form of the TF (Roessler et al, 2005). This repressive form then translocates into the nucleus to repress downstream Shh target genes. In the presence of Shh binding to its transmembrane receptor *Ptch1*, the inhibitory effects of Smo are released, allowing the full length *Gli2* to act as an activator, transcriptionally regulating several genes, including *Gli1* and *Ptch1* (Sasaki et al, 1999). *Gli1* acts as an activator for the Shh pathway while *Ptch1* acts as a repressor to appropriately regulate the level of activation of the Shh signalling pathway (Sigafos et al, 2021). It was shown in mice that *Gli2* acts as a downstream activator and loss of its C-terminal activator domain stalls prechordal plate development resulting in HPE (Ding et al, 1998).

*Cdon* encodes for a cell adhesion molecule that can interact with both Shh and *Ptch1*, acting as a co-receptor increasing high-affinity binding and acting as a positive Shh signal regulator (Gallardo et al, 2018). Mutations outside of the Shh-binding domain only result in HPE when *Cdon* is unable to associate with Shh and hedgehog-receptor components such as *Ptch1* (Bae et al, 2011). Mutations in *Cdon* suppress Shh signalling by disrupting *Gli* expression and therefore *Ptch1* expression, disrupting the signal and inducing HPE.

Mice with the *Cdon* mutations while exposed to alcohol in the womb develop symptoms similar to HPE (Hong and Krauss, 2012). Together foetal alcohol syndrome and *Cdon* mutations converge to change Nodal signalling in cells in early development (Hong et al, 2020), separate and before Shh signalling is involved. This suggests that *Cdon* has an upstream function independent of its role as a Shh receptor, leading to defective Nodal signalling at the primitive streak stage of development (HH3 in chick), where specification of the prechordal plate is perturbed, leading to HPE defects. **Figure 1.3.5.2** below shows strong *Cdon* expression in the anterior neural folds, similar to that shown in the other HPE genes.



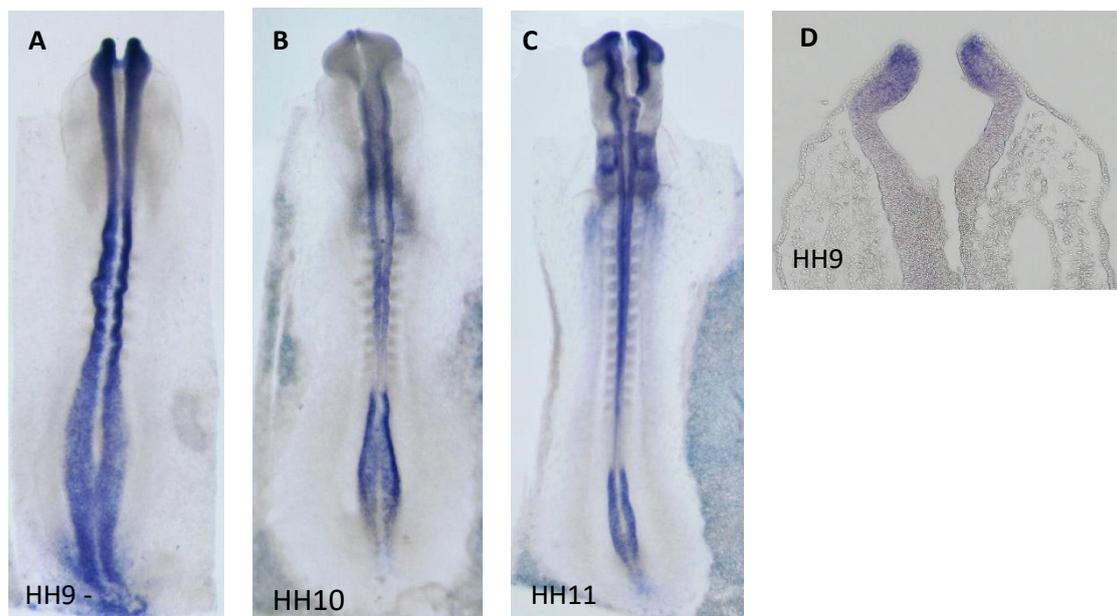
**Figure 1.3.5.2: Expression of *Cdon* along the midline axis. (A)** Expression of *Cdon* seen in the epiblast on the cells ingressing along the primitive streak at stage HH4+. **(B)** Expression concentrated along the neural tube, anterior neural folds/neural pore, distal optic vesicles, and somites at stage HH10 **(C)** *Cdon* expression concentrated in fusing anterior neural folds, distal optic vesicles, notochord and somites at stage HH11. **(D)** Transverse section shows *Cdon* expression in the anterior neural folds in the neural ectoderm and budding optic vesicles at stage HH10-. **(E)** Transverse section showing expression in the anterior neural ectoderm of the neural tube and optic vesicles in stage HH11. (Grocott et al, unpublished)

Disp1 is a transmembrane protein, like Ptch1 (Gongal et al, 2011) and is a core member of the Shh signalling pathway acting as an exporter of processed Shh ligand (Burke et al, 1999). Mutations in the mouse orthologue have resulted in HPE (Tian et al, 2008). Mutations in human *Disp1* result in primarily lobar and microforms of HPE (Mouden et al, 2016) which are rarely diagnosed without a combination of other causative gene mutations such as in *Shh*. Interestingly deletions of the 1q41-q42 region, which includes *Disp1*, have a range of clinical presentations including HPE but also opposite symptoms such as hypertelorism (increased distance between eye orbits) (Roessler et al, 2009). One homozygous mutation found implicated the gene *Fgf8*, which is also highly expressed in the anterior neural folds, similar to the HPE causative genes (Mouden et al, 2016).

Mutations in any of these four genes could work in tandem to enhance HPE severity, all focused on disrupting the Shh signalling pathway.

### 1.3.6 Other causative genes (*Stag2*, *Cnot1*)

**CCR4-Not transcription complex, subunit 1 (*Cnot1*) (MIM# 604917; 16q21; see Table 1.3)** encodes a subunit for the CCR4-NOT complex which is involved in transcriptional repression via mRNA de-adenylation (Collart & Panasenko, 2012). *Cnot1* is expressed in the neuroectoderm of the anterior neural folds (**Figure 1.3.6**). Zheng et al show that *Cnot1*, as well as other members of the Ccr4-Not complex are critical for maintaining a pluripotent state in embryonic stem cells by acting as repressors (Zheng et al 2012). *Cnot1* may play a role in allowing for the mass proliferation and growth as the forebrain but is then downregulated as the optic vesicles become more specified. Mutations in *Cnot1* were shown to impede neurological development and give HPE-like phenotypes in mice (De Franco et al, 2019). *Cnot1* was identified as a human causative gene through exome and Sanger sequencing of patient samples, where a deleterious heterozygous mutation was discovered in two unrelated patients who presented with semilobar HPE (Kruszka et al, 2019). Functional studies in chick (Grocott et al, unpublished) and mouse (Kruska et al, 2019) have shown *Cnot1* to be expressed in the anterior neural folds during neurulation and is necessary for complete forebrain division.



**Figure 1.3.6: Expression of *Cnot1* along the midline axis. (A)** Expression of *Cnot1* seen along the neural folds of the neural tube, concentrated in the anterior at stage HH9-. **(B)** Expression seen in the anterior neural folds at stage HH10. **(C)** *Cnot1* expression concentrated in anterior neural folds, forebrain, midbrain and hindbrain, optic vesicles and notochord at stage HH11. **(D)** Transverse section shows *Cnot1* expression in the anterior neural folds in the neural ectoderm at stage HH9. (Grocott et al, unpublished data).

**Stromal Antigen 2 (Stag2) (MIM# 300826; Xq25; see Table 1.3)** mutations have been found in patients with X-linked HPE13 (Kruszka et al, 2019). *Stag2* encodes for one of the four subunits of a highly conserved multiprotein complex in mammals, the cohesin complex (Brooker et al, 2014). This complex forms a ring structure which regulates sister chromatid cohesion during mitosis and meiosis (Mullegama et al, 2017). The complex is also involved in DNA replication, DNA repair and transcription.

*In vitro* knockdown of *Stag2* causes aberrant expression of HPE-causative genes *Zic2* and *Gli2* and HPE-associated genes *Smad3* and *Fgfr1* (Kruszka et al, 2019). Interestingly overexpression of *Zic2* was seen in particular. While loss-of-function of *Zic2* results in the failure of the formation of the prechordal plate structure and activation of *Foxa2*, which is required to activate *Shh*, the consequences for overexpression of *Zic2* is unclear. In *Xenopus* it has been suggested that overexpression of *Zic2* may deplete *Foxa2* levels in the Spemann organiser (Houtmeyers et al, 2016) but this not been shown in equivalent mammal models.

When the mutations in causative genes and chromosomal re-arrangements are accounted for, more than 65% of HPE cases remain unexplained, suggesting the involvement of many other genes and/or regulatory elements (Dubourg et al, 2007). Numerous other candidate genes and candidate loci have been suggested for HPE. With some already mentioned above such as, *Foxa2*, *Smad3* and *Nodal*. It seems extremely likely that many other genes will be identified in the future as causative for HPE. But while mutations in these genes alone were shown to cause the HPE phenotype a debate on multi-hit origin of HPE seems to be emerging (Ming & Muenke, 2002). This multi-hit pathology requires two or more events to occur at once e.g. mutations in multiple genes and/or environmental factors converging with genetic mutations. Environmental and genetic factors were already seen to cause severe Nodal signalling disruption in mice. This multi-hit hypothesis can also explain the wide variety of phenotypes displayed for apparently identical genetic mutations.

There is only management of HPE and genetic counselling for diagnosis. But even genetic counselling is inefficient due to the extreme phenotypic variability, the genetic heterogeneity, the multi-hit origin and the high risk of recurrence (13%) in apparently sporadic cases (Odent et al, 1998; Dubourg et al, 2007). Many questions remain unanswered around HPE and more research is needed across the whole genome to identify further candidate causes of HPE and elucidate answers explaining the molecular mechanism.

## 1.4 Regulatory Elements

In the human genome, 98% of the DNA sequence is comprised of non-protein coding 'junk DNA'. Within these non-coding sequences are *cis*-regulatory DNA elements such as promoters, enhancers, silencers and insulators. These regulatory sequences, in concert with the trans-acting factors that bind to them, play a crucial role in controlling when and how the protein-encoding information of the genome is expressed in specific cell types, conditions, and developmental stages.

The term epigenetics first came about in the 1940's (Waddington, 1942) to describe what was then considered a phenomenon of changes in phenotype, without changes to genotype. Since then the understanding of epigenetics has evolved to be defined as a mechanism by which heritable changes in gene function are caused, not by a change/mutation within the coding gene itself, but by external modifications to the DNA that affect how the cells 'read' the genes and can cause them to be switched 'on' or 'off'. Known mechanisms include DNA methylation, histone modifications and non-coding RNAs.

These regulatory elements have far-reaching effects regarding translation, gene and protein function. It is now understood that mutations within these elements can act similarly, or more discreetly, to that of mutations within their target gene. For example, the *Shh* gene is the most well characterised of the HPE causative genes; an enhancer was recently discovered, SBE7, that when knocked-out in mice, caused a complete loss of Shh signalling, not only in the prechordal plate mesoderm, but also along the ventral midline of the forebrain resulting in defects similar to those of HPE (Sagai et al, 2019). Many more diseases, such as diabetes and cancer, have in recent years been linked to the aberrant function of these regulatory elements (Lee & Yang, 2014) and as such have highlighted the importance of identifying, characterising and finally understanding the functional role they play within healthy and disease mechanisms.

#### 1.4.1 Promoters

The best characterised regulatory elements are promoters. In eukaryotes, promoter sequences typically lie immediately upstream of the transcription start sites (TSSs). This genomic sequence is also known as the core promoter, which can also commonly include a conserved promoter sequence called the TATA box. This minimal DNA sequence is sufficient to assemble the RNA polymerase II preinitiation complex (PIC) to start transcription of genomic DNA of the gene into RNA. This directs a low-basal level of transcription (Roeder et al, 1996), that can be boosted when in the presence of enhancers.

A common epigenetic modification made to promoter sequences is DNA methylation and demethylation to control gene silencing and activation, respectively. DNA methylation normally occurs on cytosine at CpG islands across the human genome (Miranda & Jones, 2007). With less than 30% of all human promoters considered to be CpG poor (Deaton & Bird, 2011).

#### 1.4.2 Enhancers

The first eukaryotic enhancer was discovered in the early 1980's when expression of a mouse immunoglobulin heavy chain gene was enhanced in *cis*, by a nearby open chromatin region (Benaerji et al, 1983). There are hundreds of thousands of putative enhancers mapped within the mammalian genomes (Shen et al, 2012; ENCODE Project Consortium, 2012), far outnumbering genes.

Enhancers are short (50-1500bp) regions of DNA that can be bound by TFs and cofactors that increase the transcriptional output of a target gene (Plank & Dean, 2014). They do this by containing clusters of short DNA motifs for which specific TFs, necessary for PIC assembly, can bind to. These TFs recruit co-activators (p300 histone acetyltransferases (HATs) and the Mediator Complex (MED)) (Shylueva et al, 2014). Co-activators often function either as histone modifiers (HDMs), ATP-dependent chromatin remodellers, or mediators of long-range crosstalk with basal transcriptional machinery at promoters. The interaction between an enhancer and a promoter can be stabilized by CCCTC-binding factor (CTCF), in cooperation with the cohesin complex and the dimerization of LOB domain-containing protein 1 (LDB1), which is recruited to chromatin through interactions with LIM proteins (**Figure 1.4.2.3A**) (Kyrchanova & Georgiev, 2021).

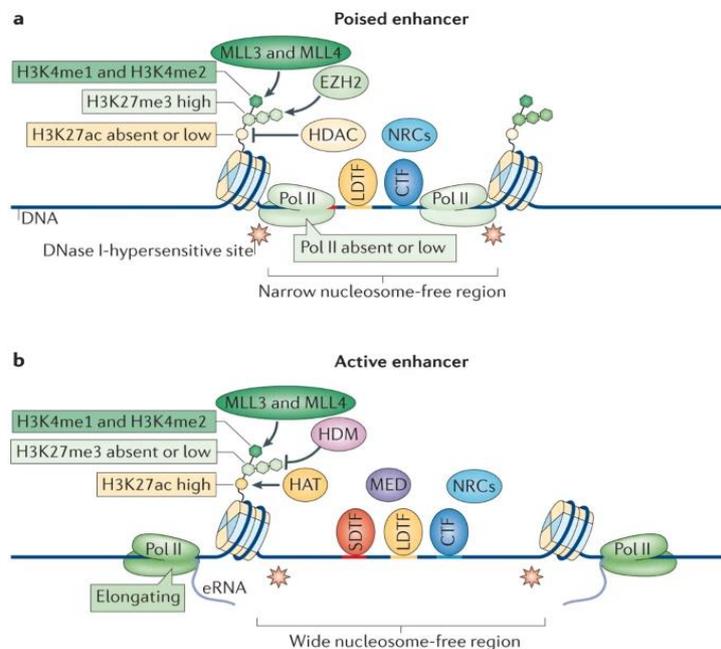
Enhancers can be broadly categorised into four states: active, poised, primed and inactive (Ernst & Kellis, 2010). These different states are required for the spatio-temporal regulation of genes within different tissues. When in their active state (**Figure 1.4.2.1B**) enhancers' signal dependent TFs (SDTFs) overcome the nucleosomal barrier by accessible chromatin remodelling establishing a nucleosome-free and DNase I-hypersensitive region of chromatin (**Figure 1.4.2.1**). They are also bound by general TFs and RNA polymerase II and can produce enhancer RNAs (eRNAs). Enhancer activation is often marked by histone modifications of the nucleosomes that flank TF binding regions

(Rasa-Iglesias et al, 2011). These modifications commonly include, but are not limited to, H3K27ac and H3K4me1 (Schoenfelder & Fraser, 2019).

Inactive enhancers can be subclassified into silenced, repressed and primed. Silenced enhancers are buried in compact chromatin depleted of transcription factor binding and histone modifications (Bozek & Gompel, 2020). Whereas repressed elements display some accessibility as they are repressed by TFs. Repressed enhancers share the H3K4me signature with active elements but have reduced or absent histone acetylation modifications (Ostuni et al, 2013).

An enhancer is generally considered to be in a primed state when its sequence specific TFs are bound but cannot produce enhancer RNAs. Primed enhancers are marked by chromatin features similar to repressed elements, H3K4me1 and H3K27me2 modifications but lack H3K27ac, which would switch them into an active state (Heinz et al, 2015; Calo & Wysocka, 2013).

Poised enhancers (**Figure 1.4.2.1A**) are similar to primed enhancers but are distinguished by the presence of the repressive H3K27me3, which must be removed to allow the transition to an active enhancer state. They are also commonly bound by Polycomb complex (Rada-Iglesias et al, 2011). These states are commonly found in pluripotent stem cells.



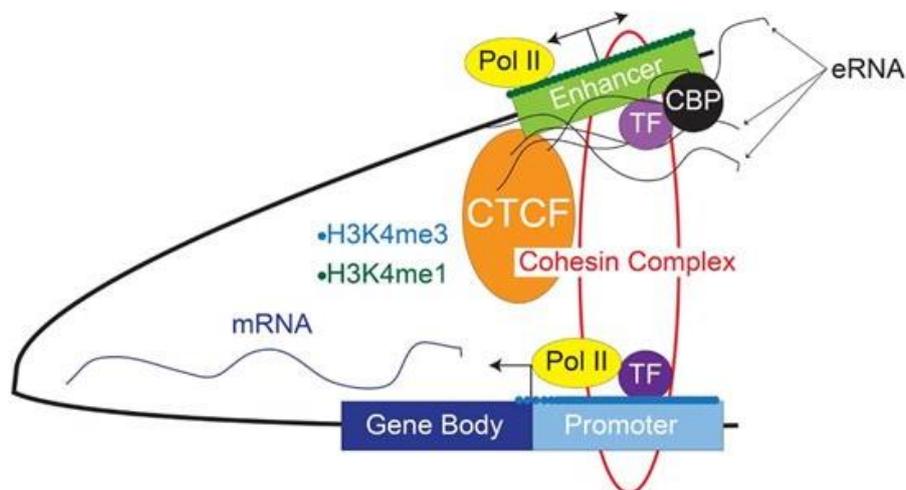
**Figure 1.4.2.1: The characteristics of poised and active enhancers. (A)** Schematic shows enhancer in a poised state. The binding of lineage determining TFs (LDTFs) and collaborative (non-specific) TFs (CTFs) can recruit coactivator proteins to remodel nucleosomes. Nucleosome-remodelling complexes (NRCs) creates a DNase I hypersensitive, nucleosome free region. The histone methyltransferases (MLL3 & MLL4) deposit the histone marks: H3K4me1 and H3K4me2. EZH2, (a component of the Polycomb complex), transfers repressive H3K27me3 marks. HDAC maintains histones in a repressed, de-acetylated state. **(B)** Schematic shown chromatin features found at active enhancers. SDTFs associate with recognition motifs and LDTFs in response to various cues. Further nucleosome displacement occurs widening the Dnase I hypersensitive site. Co-activators (HDM) remove H3K27me3 marks. HATs deposit the H3K27ac mark, recruiting the Mediator complex. Elongating Pol II results in bidirectional transcription and the generation of eRNAs, which is unique to active enhancer activity. (Adapted from Heinz et al, 2015).

While significant attention is given to features associated with active enhancers, H3K27ac, it is not exclusively what marks an active enhancer, and genomic regions that are not necessarily enhancers

have been known to exhibit these features. As such there is an increasing need to identify enhancers using multiple parameters. The co-activator p300 has other recognised acetylation marks for enhancers (Jin et al, 2011). Furthermore, other histone acetylation complexes have been associated with enhancers, and are probably occupied by multiple complexes, just like promoters.

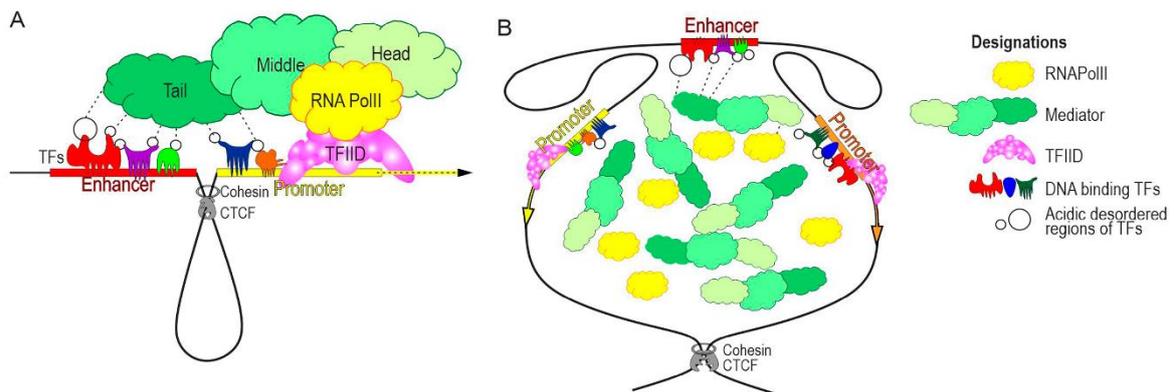
Historically enhancers were often thought to be in close proximity to their target gene(s) and their promoters, and for most housekeeping genes this is still true. However, developmental genes often have multiple enhancers located at variable distances from their promoter (Kyrchanova & Georgiev, 2021). For instance, the *Shh* gene has been shown to have multiple enhancers required for different stages of development and scattered along the chromosome, with the limb-specific ZRS enhancer located 1Mbp (1,000,000 bp) away from the gene (Lettice et al, 2002). Multiple models have been proposed to explain long-range enhancer-promoter interactions.

The ‘looping model’ (Figure 1.4.2.2) has gained favour in recent years (Ong & Corces, 2011) allowing direct interactions between enhancer and promoter. In this model the enhancer is brought into close proximity to the promoter by looping out on chromatin loops which are stabilised by protein-protein interactions (Vernimmen & Bickmore, 2015) (Figure 1.4.2.2). This allows the enhancer-associated factors to transiently bind to the TFs (which form a platform) increasing their local concentration and function at the promoter (Heinz et al, 2015). This model has been confirmed via the emergence of chromosome conformation capture (3C) technologies and since then 4C and 5C derivatives (Calo & Wysocka, 2013).



**Figure 1.4.2.2:** Schematic showing enhancer-promoter interactions via the ‘looping’ model. Enhancer is located X distance away from *cis*-acting gene and are brought into close contact with the genes they regulate through chromatin looping; mediated by both CTCF and the cohesin complex. The enhancer is bound by HAT (CBP). Active transcription occurs producing eRNAs. (Adapted from Rao, 2020).

Other models have developed from this original theory, such as the emerging view that TFs dynamically bind and dissociate from their targets (Kyrchanova & Georgiev, 2021). Forming aggregates that can interact with the subunits of the Mediator complex and basal transcriptional machinery (Figure 1.4.2.3B). Generally, enhancers are found within the same trans-activating domain (TAD) as their target promoter (Cavalheiro et al, 2021). The rate of transcription increases with the concentration of transcriptional complexes increasing near enhancers, as promoters can more efficiently recruit these complexes to initiate transcription.



**Figure 1.4.2.3: Models of promoter-enhancer communication. (A)** Classic model of promoter activation through close contact with enhancer. The interaction between enhancer and promoter is stabilised by CTCF, the cohesin complex and proteins. The TFs bind to the enhancer and a promoter to form a platform for the transient recruitment of the mediator complex. The mediator complex transfers RNA polymerase II to the promoter transcription factor IID (TFIIID) complex and accelerates further transcription initiation steps to induce a short transcriptional pulse (burst). The enhancer strength is directly correlated with the efficiency of mediator recruitment to chromatin. **(B)** Model of enhancer-promoter communication through formation of hubs. Enhancer and promoter are still brought relatively close to one another by interactions between CTCF/cohesin sites. TF's form aggregates, activation domains within TFs efficiently interact with subunits of the Mediator and RNA polymerase II complexes. As a result, the concentration of transcriptional complexes increases near enhancers, and promoters can more efficiently recruit these complexes to initiate transcription. (Adapted from Kyrchanova & Georgiev, 2021).

### 1.4.3 Identification and Characterisation of Enhancers

Studying enhancer function is currently an area of great interest. Precisely identifying regulatory elements is key to deciphering the mechanisms underlying transcriptional regulation but also understanding their role in evolution and disease. However, the identification of enhancers is challenging for several reasons; they are often found in the introns of genes neighbouring to the ones they regulate, with no set distance from the promoter. In invertebrates' enhancers have been found outside of their promoters topologically associated domain (TAD) (Ing-Simmons et al, 2021). TADs are a form of structural DNA organisation, where regions of the genome are grouped together in a loop extrusion model. Until recently it has largely been thought that enhancers and their targets genes reside within the same TAD (Long et al, 2016) largely due to the unknown impact TAD formation has on transcriptional bursting (Yokoshi et al, 2020).

Enhancers are unlike a promoter which can be identified by sequencing the 5' end of its mRNA (Bulger & Groudine, 2013). The spatial-temporal activity of enhancers is usually restricted to a select tissue t(or even just a few cells) at specific developmental time stamps. Whilst there are common modifications that can be used to identify enhancers, they are not exclusive to all or unique to the enhancer population.

The flexibility which defines an enhancer is what hinders the attempt to locate and characterise all of them within the genome. Enhancer detection therefore relies on a mix of imperfect assays that measure chromatin structure and sequence functionality.

One criterion employed for predicting putative enhancers is the use of phylogenetics and the conservation of non-coding sequences. The assumption is that Evolutionary Conserved Regions

(ECRs) of non-coding sequences, that have been preserved throughout evolution, imply a regulatory function. The higher the conservation, the greater the importance of the regulatory elements function. While this principle could be considered a 'basic' or 'established' approach to finding regulatory sequences this approach has only come about in the last 20 years as multiple fully sequenced genomes are now available for comparison. However, this method hinges on the same genes within different species requiring the same enhancer activity, which is not always the case (Hare et al, 2008; Swanson et al, 2010). Prediction of ECRs can be paired with tools such as JASPAR (JASP Team 2021) TF binding site predictions, which can help support the theory of an ECR being a genuine enhancer.

However, comparisons of genome-wide transcription factor binding patterns across species indicate that a large proportion of enhancers are species-specific (Bulger & Groudine, 2011) and even with the supporting evidence of specific TF binding sites, it does not mean the DNA motif can or does functionally bind TFs. Therefore, predicting enhancers solely on sequence conservation will result in a high-false positive rate, and fail to identify other enhancers entirely.

The use of genome wide approaches and high-throughput sequencing assays has improved significantly in the last decade. As mentioned above DNase I hypersensitivity is a characteristic associated with CREs and therefore can be used as a means of detection for enhancers. When combined with deep sequencing in adult cell lines this method has been successful in mapping active regulatory regions to the genome (Song et al, 2011; Thurman et al, 2012).

Profiles of transcription (by RNA-seq), chromatin accessibility (ATAC-Seq), and epigenetic chromatin marks (ChIP-seq or CUT&Tag) can also be studied to uncover temporal changes in gene expression, chromatin structure and chromatin states indicative of potential enhancers (Creighton et al, 2010; Rada-Iglesias et al, 2011) . However, these methods also have their pitfalls; the ChIP-seq approach is dependent on TFs being bound to the DNA sequence, which is not always the case, functionally significant, or straightforward to correlate binding with the relevant target gene (Rada-Iglesias et al, 2011). Another issue is accessing sufficient amounts of appropriate material, which is not always available at specific developmental timeframes when performing *in vivo* experiments.

ATAC-seq is a relatively new method (Buenrostro et al, 2013) for identification of open chromatin regions. It was developed to supplement MNase-seq and DNase-seq by reducing experimental time, simplifying the procedure (reducing errors) and reducing sample size (Sun et al, 2019). ATAC-seq most commonly uses the hyperactive transposase 5 (Tn5) through a 'cut and paste' mechanism (**Figure 1.4.3**) (Buenrostro et al, 2015). Whereby it fragments and tags the unprotected regions DNA with sequence adapters (**Figure 1.4.3B**). These tagged DNA fragments are amplified and then sequenced using next generation sequencing. This generates reads that can be mapped to the genome and shown as 'peaks', with peak length referencing degree of chromatin accessibility.

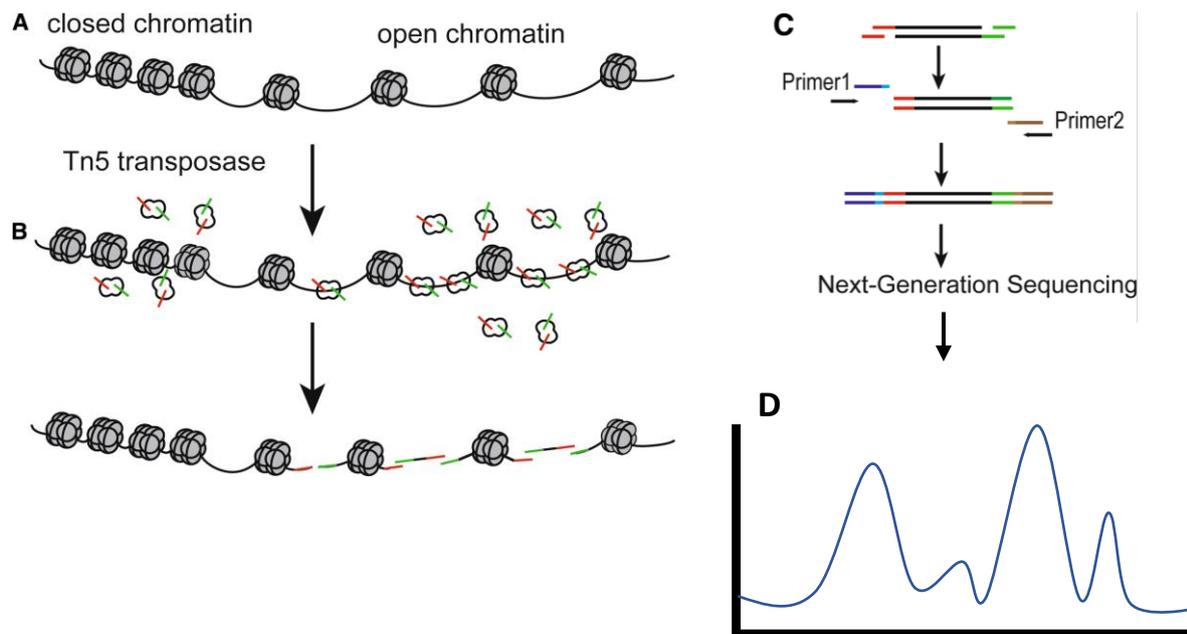


Figure 1.4.3: Schematic showing process for ATAC-seq. (A) Showing open and closed chromatin. (B) The Tn5 transposase is more abundant in open chromatin than closed chromatin. The Tn5 cuts the open chromatin and 'pastes' the 2 adapters to either end of the fragmented DNA. Green = adapter 1. Red = adapter 2. (C) Universal Primer 1 and 2 are used for PCR amplification and barcode edit for and generation of libraries for sequencing. (D) Next generation sequencing is used to generate 'reads' which are mapped to the genome as open chromatin 'peaks'. (Adapted from Sun et al, 2019).

Disadvantages of ATAC-seq include the adapter sequences at either end of the cut DNA fragment being random, leading to a 50% probability that both adapters on the same fragment are the same, and therefore unusable for amplification (Sun et al, 2019). The Tn5 transposase is known to bind and cleave at TF binding sites (Meyer & Liu, 2014), allowing detection of averaged TF footprints across the whole accessible genome, but not at single-enhancer resolution (Mok et al, 2021).

Single-cell ATAC-seq (scATAC-seq) and Omni-ATAC-seq have been further developed to examine chromatin organisation in heterogenous cell populations and eliminate mitochondrial interference and reduce background noise to obtain high quality data, respectively.

These indirect predictions of CREs are then usually studied using a functional enhancer assay (Nelson & Wardle, 2013). The potential enhancer sequence is cloned into a reporter vector and injected into an embryo showing the spatiotemporal pattern of expression driven by the conserved element. Parallel functional assays studied across multiple species can also be a good way of showing conserved expression of an enhancer, by the enhancer driving reporter gene expression in the same domain across the different species. However, as mentioned above there is often poor correlation between sequence conservation and functional conservation (Nelson & Wardle, 2013). Whether it be indirect or genome-wide approaches used to predict enhancers, the cloning of individual enhancer candidates into a reporter vector is a bottle neck in any assay.

While these potential enhancer sequences may be sufficient to drive gene expression, are they necessary? Studies have shown that when the conserved region that drives gene expression is knocked-out no change in phenotype (and gene expression) was seen (Ahituv et al, 2007). Since then it has also been shown that multiple enhancers can act on one gene and deletion of a conserved

region can be compensated by the presence of 'shadow' enhancers (Frankel et al, 2010; Perry et al, 2010; Patwardhan et al, 2012).

When used in combination, indirect approaches can lead to an increased understanding of sequence requirements and inform indirect methods of detection and while identification and characterisation of CREs remains crucial, assumptions regarding the properties of CRE's cannot be absolutely defined. Our ability to identify CRE's is improving with time and ongoing projects, such as The Encyclopaedia of DNA Elements (ENCODE), aim to map all regulatory elements within the human genome using a multitude of techniques, such as the ones described above (The ENCODE Project Consortium, 2012). But the rigorous study of the regulation of any given gene is still not a trivial undertaking.

## 1.5 Project aims

Although much is known about the causative genes associated with HPE, the enhancer elements responsible for their target genes transcription are yet to be fully identified and characterised. Enhancer elements have been found for *Six3* (Lee et al, 2017; Conte & Bovolenta, 2007), *Zic2* (Roessler et al, 2012) and *Gli2* (Minhas et al, 2015) in model organisms such as mice and zebrafish. However, these enhancers were involved in later stages of development, not identified in the primary neuralation stages so key for HPE presentation. Whole-genome wide studies have estimated the human genome contains ~400,000 enhancers (ENCODE Project Consortium, 2012) and counting, with the average human gene regulated by approximately 20 enhancers. As such it is unlikely the same enhancer sequences already found are responsible for the temporal and spatial expression of our genes of interest within our specific developmental window. Whether these enhancer sequences are even conserved across to the human and chick genome is also up for debate (Nelson & Wardle, 2013). With the exception of *Shh* where all enhancer sequences are thought to have been found and characterised (Sagai et al, 2019) future screens should yield positive hits for novel putative enhancers.

The overall aim of this project was to construct a shotgun fragment-based library for the causative *Six3* gene. To do this an unbiased approach was employed. Using bacterial artificial chromosomes (BAC) clones, a genomic shotgun reporter library made up of randomly amplified fragments was constructed. This unbiased approach to identifying enhancer elements should, in theory, increase efficacy of positive hits as compared to other approaches for identifying cis-regulatory elements. One such approach is a candidate approach, by looking at evolutionary conserved regions (ECRs) and using phylogenetic foot-printing to predictively identify putative enhancers for these known disease genes. An ECR containing a putative enhancer for the *cMyc* gene was identified using this candidate approach.

However, this approach historically has a relatively low efficacy and is a time exhaustive approach, even when used in combination with ATAC-sequencing data. As the genome is seeded with accessible elements it is not conducive to look at only one type point (e.g.TF binding sites) it is also necessary to look at heterotopic and heterochronic comparisons to see when the elements are open and closing and in which cells. However, to look at a good range of these parameters is costly and just not viable for everybody. As such this unbiased molecular approach has the potential as a successful complimentary method to these epigenetic discovery tools.

This project also looked at employing *Gallus gallus* (chick) embryos as a good model for amniote eye development. The potential enhancer elements were studied using a functional assay screen. The candidate enhancer identified for the *cMyc* gene was cloned into the pTK-Citrine reporter plasmid. Electroporated into HH4 and HH8 embryos and its spatiotemporal activity validated.

The objectives of this project were:

- To modify the pTK-Citrine Vector to contain unique barcode sequences to allow for rapid functional screening of potential enhancer sequences
- To compare and optimise TA and/or Golden Gate cloning strategies to create an unbiased shotgun library
- Optimise fragmentation and preparation of BAC clone ready for cloning
- To construct a shotgun fragment-based reporter library for the *Six3* gene
- To optimise the Golden Gate cloning strategy to produce hundreds of colonies needed to screen the entire BAC clone
- To use a candidate approach to investigate a potential enhancer for the *cMyc* gene

## Chapter 2: Methods and Materials

### Modification of pTK-Citrine Vector to Include Unique Barcode Edits

To increase the efficiency and sensitivity of our enhancer screen, we modified the original reporter vector pTK-Citrine vector following Chen & Streits protocol (Chen & Streit, 2015). Primers (**Table 2.1**) were used to introduce 16 nucleotides, downstream of the MCS and minimal pTK promoter, to generate 5 uniquely barcoded vectors. Reaction contained: The ptk-Citrine plasmid was subjected to 20 thermal cycles (98°C for 30 sec; 95°C for 10 sec; 60°C for 6 mins; 72°C for 10 min). The reaction mix contained 5ng ptk-Citrine as template, 5X Phusion HF buffer (NEB), Phusion DNA polymerase (NEB), 0.5 µM primer pair, 10 mM dNTPs, and sigma water to make up a final reaction volume of 50 µL. Additional DpnI digest at 37 °C, 60 min was performed, followed by transformation into competent DH5-α *E. coli* cells and plated on Carbenicillin & X-gal plates.

**Table 2.1 Primers used to insert edits into pTK-Citrine vector (Chen & Streit, 2015).**

Edit	Primer	Sequence (5' – 3')
Barcode edit 1	T1F1	CAGTTTTCAAGCCGGAgtaagtatcaaggttacaagacag
	T1R1	TCCGGCTTGAAAAGTgacgaccaacttctgcagttaag
Barcode edit 2	T2F1	TGATACACCGAGTCGTgtaagtatcaaggttacaagacag
	T2R1	ACGACTCGGTGTATCAacgaccaacttctgcagttaag
Barcode edit 3	T3F1	AGCTCTTCGCAAAGTGgtaagtatcaaggttacaagacag
	T3R1	CACTTTGCGAAGAGCTacgaccaacttctgcagttaag
Barcode edit 4	T4F1	CAGCTTACTCGTAAGGgtaagtaagtatcaaggttacaagacag
	T4R1	CCTTACGAGTAAGCTGacgaccaacttctgcagttaag
Barcode edit 5	T5F1	ACGATGAAGCCTTGTCgtaagtaagtatcaaggttacaagacag
	T5R1	GACAAGGCTTCATCGTacgaccaacttctgcagttaag
BsmBI	Fwd	TCGTCTCCCTGGGGCTAGCCCCGGCTCGAGATCTGC
	Rev	GGCTAGCCCAGGGAGACGACCGGTGGAAAG

### Modification of pTK-Citrine Vector to Include BsmBI adapter sequence

To allow for Golden Gate cloning we edited the vector now containing the unique barcode edits to contain the BsmBI adapter sequence. Primers (**Table 2.1**) were used to introduce the BsmBI sequences. Reaction contained: The ptk-Citrine plasmid-Barcode edit X was subjected to 20 thermal cycles (98°C for 30 sec; 95°C for 10 sec; 60°C for 6 mins; 72°C for 10 min). The reaction mix contained 5ng vector as template, 5X Phusion HF buffer (NEB), Phusion DNA polymerase (NEB), 0.5 µM primer pair, 10 mM dNTPs, and sigma water to make up a final reaction volume of 50 µL. Additional DpnI digest at 37 °C, 60 min was performed, followed by transformation into competent DH5-α *E. coli* cells and plated on Carbenicillin & X-gal plates.

### BAC Clone Preparation

Once received BAC clones were streaked out on chloramphenicol plates. Selected colonies were

grown up in low sodium LB cultures ~24hrs at 37 °C, 180 rpm. DNA extracted using Machery-Nagel Nucleobond Xtra Midi Kit (ID: 740410.10) and eluted in 100 µL sigma water. 100 ng/µL sent for sequencing (SourceBioscience) to confirm the identity of the BAC clone grown.

### Preparation of Template DNA for Library 1 PCR

BAC DNA was separated from host bacterial DNA using 1% TAE agarose gel. Supercoiled DNA was extracted and purified using Qiagen QIAEX II Gel Extraction Kit (Cat no. 20021). Eluted in 20 µL elution buffer to give ~5-12 ng/µL of template DNA.

### Library 1 PCR

Template DNA was subjected to 5 ‘initial’ thermal cycles (98°C for 30 sec; 95°C, for 30 sec; 25°C for 30 sec and 72°C for 8 min with a ramp rate of 0.1°C/sec). Followed by 35 cycles (95°C for 30 sec; 55°C for 30 sec; 72°C for 8 min). The initial cycles have a low annealing temperature to allow for random priming, whereas later cycles included a more typical annealing temperature to allow for full-length amplification of the initial randomly generated fragments. A “standard reaction” contained 5ng of purified BAC DNA as template DNA, 2.5 mM MgCl, 1x Buffer HF, 0.5 mM dNTPs, 0.5 µM Lbrry1 Fwd primer (**Table 2.2**), 10 µM Lbrry1 Rev primer (**Table 2.2**), Phusion High-Fidelity DNA Polymerase and sigma water in a final volume of 50uL. Where indicated different cycling settings and different concentration of template DNA, MgCl, primers were used to optimise this PCR. After thermal cycling the products were analysed on a TAE agarose gel (2%). Fragments were size selected, extracted and purified using a Qiagen Gel Extraction kit (Cat no. 28704) and eluted in 20 µL sigma water.

The Lbrry1 primers were designed from the M13 primers (**Table 2.4**) to contain random nucleotides to allow for random amplification of different sized fragments. These primers were designed to be used before Lbrry2 primers (**Table 2.3**) which would insert BsmBI adapter sites, to allow for Golden Gate Cloning. The Lbrry1b primers combined the Lbrry1 and Lbrry2 primers to random amplify and insert the cloning adapters by having having the following design:

5’ – A6 > BsmBI u/s adapter > M13-Fwd > N8 > AT – 3’

**Table 2.2: Oligonucleotides used in Library 1 PCR**

Primer	Sequence (5’ – 3’)
Lbrry1-Fwd	TGTA AACGACGGCCAGTNNNNNNNNAT
Lbrry1-Rev	CAGGAAACAGCTATGACNNNNNNNNAT
Lbrry1b-Fwd	AAAAAACGTCTCCCCACTCTAA AACGACGGCCAGTNNNNNNNNAT
Lbrry1b-Rev	AAAAAACGTCTCCAACACAGGAAACAGCTATGACNNNNNNNNAT

### Library 2 PCR

Size selected Library 1 products was subjected to 40 thermal cycles (98°C for 30 sec; 95°C for 30 sec; 55°C for 30 sec; 72°C for 7 min). A “standard reaction” contained ~5ng of size selected Lbrry1 template DNA, 5x Buffer HF, 10 mM dNTPs, 0.5 µM Lbrry2 Fwd primer (**Table 2.3**), 0.5 µM Lbrry2

Rev primer (**Table 2.3**), Phusion High-Fidelity DNA Polymerase and sigma water in a final volume of 50  $\mu$ L. PCR products were purified using Qiagen PCR Cleanup (Cat no. 28104).

**Table 2.3: Oligonucleotides used in Library 2 PCR**

Primer	Sequence (5' – 3')
Lbrry2-Fwd	AAAAAACGTCTCCCCAGTGTAACGACGGAAAGT
Lbrry2-Rev	AAAAAACGTCTCCAACACAGCAAACAGCTATGACC

### BsmBI Digest

To reduce the presence of concatemers in our library preparation, BsmBI digestion was performed on Library 1 PCR products. The Library 1 products were purified using Qiagen PCR Cleanup (Cat no. 28104) and digested (55 °C for 2hr). The reaction mix contained QiAquick eluate (purified PCR products), BsmBI-v2 (NEB), 10x Buffer 3.1 (NEB) and sigma water made up to a final volume of 25  $\mu$ L.

### Golden Gate Cloning

Standard Golden Gate protocol: Insert and Vector were cloned at a ratio of 1:1. The reaction mix contained insert (75ng), pTK-Citrine reporter vector (75ng), T4 DNA ligase (Promega), T4 ligase buffer (Promega), BsmBI-v2 (NEB) and sigma water made up to a final volume of 20  $\mu$ L. This reaction was subjected to 25 thermal cycles (37°C for 2 min; 16°C for 5 min), followed by a digestion step (55°C for 30 mins), before the enzyme was denatured (80°C for 5 min). Where indicated different concentrations of vector and insert and cycling conditions were used to optimise this PCR for library construction.

### Plasmid Safe reaction

To remove any linear DNA, one volume of the Golden Gate reaction was incubated (37°C for 60 min; 70°C for 30 min). The reaction mix contained GG reaction, 10x Plasmid Safe Buffer (Lucigen), ATP (25 mM), ATP-dependent DNase (Lucigen) in a final volume of 12.5  $\mu$ L.

### Transformation

One volume of reaction was added to DH5- $\alpha$  *E. coli* chemically competent cells, in a 1:10 ratio. Heat-shock transformation protocol as follows: on ice for 30 min; 42°C for 30 sec; on ice for 2 min. Grown in preheated (37°C) SOC outgrowth for 1hr shaking at 250 rpm. Spun down 5,000 rpm in a conventional tabletop centrifuge at room temp (15-25°C). Resuspended and plated on X-gal and carbenicillin plates.

### Colony PCR

Colony PCR was used to confirm the correct sized product had be cloned or amplified using PCR methods. In a "standard" procedure selected colonies were subjected to 30 thermal cycles (98°C for 30 sec; 98°C for 30 sec; 55°C for 30 sec; 72°C for 3 min). The reaction mix contained the dissolved colony, 2x BioMix Red (BioLine), 0.5  $\mu$ M selected primer pair and sigma water to a final volume of 25  $\mu$ L. After thermal cycling the products were analysed on a TAE agarose gel (1%). Selected colonies were also grown in LB cultures ~16hrs, DNA extracted using Qiagen QIAprep Spin Miniprep Kit (Cat No. / ID: 27104), and 100 ng/ $\mu$ L sent for sequencing (SourceBioscience) to confirm identity.

**Table 2.4: Oligonucleotides used for Colony PCR**

Primer	Sequence (5' – 3')
M13-Fwd	TGTA AACGACGGCCAGT
M13-Rev	CAGGAAACAGCTATGACC
Ctrn-Fwd	GAACGGCATCAAGGTGAACT
Ctrn-Rev	CTTGTACAGCTCGCTCCATGC

### Chicken Embryo Culture and Electroporation

Fertilised chicken eggs (Henry Stewart & Co.) were incubated at 37 °C. Embryos were staged according to Hamburger and Hamilton (Hamburger & Hamilton, 1951) staging. Stage HH3 embryos were collected in Ringers solution on filter paper rings based on easy-culture method (Chapman et al, 2001). Embryos were transferred into albumin-agar dishes, ventral side up, ready for electroporation. Plasmid DNA was injected at 3 sites, anterior and on either side of the primitive streak. Plasmid DNA electroporated using 4 x 50 ms pluses at 9 V, at an interval of 200 ms. Thin albumin was used to seal the lids of albumin-agar dishes and embryos were cultured at 37 °C until desired stage was achieved.

### RNA extraction and Reverse transcription of cDNA synthesis

RNA was extracted from whole embryos using QIAGEN RNeasy Plus Mini Kit, with an additional DNase digest step while on the column. Eluted in 20 µL sigma water. RT+ and RT- samples were generated from 1 µg of RNA per sample in addition to Random Primers and sigma water; incubated at 70 °C for 10 min. Followed by the addition of 5x buffer, 100 mM DTT, 10mM dNTP, RNasin and Superscript Reverse Transcriptase (water in RT- samples); incubated at 42 °C for 60 min to generate cDNA.

### Reporter transcript detection RT-PCR

cDNA (1µg) was subjected to 30 thermal cycles (98°C for 30 sec; 55°C for 30 sec; 72°C for 1 min). A “standard reaction” contained 1 µg of cDNA, 2x BioMix Red, 0.5 µM Reporter Fwd primer (**Table 2.5**), 0.5 µM Reporter Rev primer (**Table 2.5**) and sigma water in a final volume of 10 µL. Were specified different Reporter Rev primers were used.

**Table 2.5: Oligonucleotides used for reporter transcript detection**

Primer	Sequence (5' – 3')
Rprtr-Fwd	TGCAGAAGTTGGTCGTGAG
Rprtr-Rev	TGGTGCAGATGAACTTCAGG
Rprtr-Rev2	GGGTCTGTAGTTGAAGTC
Chrry-Fwd	CCGCGTTACTCCACAG
Chrry-Rev	CCTCACATTGCCAAAAGACG

## **Touchdown PCR**

cDNA (1 $\mu$ g) was subjected to 10 'initial' cycles (98°C for 30 sec; 65°C for 30 sec, Dec 1 °C/Cyc ; 72°C for 1:30 min). Followed by 20 cycles (98°C for 30 sec; 55°C for 30 sec; 72°C for 1:30 min). A "standard reaction" contained 1  $\mu$ g of cDNA, 2x BioMix Red, 0.5  $\mu$ M Reporter Fwd primer (**Table 2.5**), 0.5  $\mu$ M Reporter Rev primer (**Table 2.5**) and sigma water in a final volume of 10  $\mu$ L. Were specified different 10  $\mu$ M Reporter Rev primers were used.

## **Preparation of Insert for TA Cloning**

BAC DNA was separated from host bacterial DNA using Plasmid Safe (DNase) digest methods. BAC DNA was incubated (37 °C, 16hrs; 70 °C, 30 min) according to Plasmid-Safe ATP-Dependent DNase Lucigen User Manual. The reaction suspended the BAC DNA at ~ 0.3  $\mu$ g/ $\mu$ L in a final volume of 250  $\mu$ L. The BAC DNA was fragmented by mechanical shearing of the pushing and sucking of the reaction through a 30-gauge needle 300 times. The sheared BAC DNA was concentrated down using ethanol precipitation methods and eluted in 20  $\mu$ L sigma water. Ends were polished by pooling BAC DNA with 2x BioMix Red (BioLine) in a 1:1 ratio and incubated (37 °C, 2hr). Fragments were size selected using agarose gel electrophoresis methods. Insert was extracted and purified using a Qiagen Gel Extraction kit (Cat no. 28704) and eluted in 20  $\mu$ L sigma water, ready for TA cloning.

## **Preparation of Vector for TA Cloning**

The un-modified pTK-Citrine vector was digested using the PvuII enzyme. The reaction contained the vector (2.5  $\mu$ g), PvuII-HF (NEB), CutSmart Buffer 10x (NEB) and sigma water in a final volume of 50  $\mu$ L. The reaction was incubated (37 °C, 2hrs). The digested vector was purified on a column using Qiagen PCR Cleanup (Cat no. 28104) and eluted in 20  $\mu$ L water. The vector was polished in a T-tailing reaction. The reaction mix contained the vector, 5x Buffer (Promega), 10  $\mu$ M dTTP, GoTaq (Promega), MgCl<sub>2</sub> and sigma water in a final volume of 30  $\mu$ L. The vector was analysed using agarose gel electrophoresis methods. Vector was extracted and purified using a Qiagen Gel Extraction kit (Cat no. 28704) and eluted in 20  $\mu$ L sigma water, ready for TA cloning.

## **TA Cloning**

Insert and vector were pooled together at a ratio of 3:1. The reaction contained the vector (25ng), insert, T4 Ligase (Promega), 2x Rapid Ligation Buffer (Promega) and sigma water to a final volume of 10  $\mu$ L. A positive control using the pGem vector was employed to see if inserts were cloneable. The ligation reactions were incubated (on bench, 2hr) and 5  $\mu$ L reaction transformed into 100  $\mu$ L DH5- $\alpha$  E. coli chemically competent cells.

## **Amplifying ECR region of human *cMyc* locus**

ECR2 was amplified using primers with specific adapters to allow cloning into reporter vector using Golden Gate cloning protocol. Human genomic DNA (200ng) was thermal cycled for 35 cycles (98°C for 30 sec; 98°C for 30 sec; 55°C for 30 sec; 72°C for 4 min). "A standard reaction" contained 200ng of Human genomic DNA as template, 5x Buffer HF, 10 mM DNTPs, 0.5  $\mu$ M Fwd primer (**Table 2.6**), 0.5  $\mu$ M Rev primer (**Table 2.6**), Phusion High-Fidelity DNA Polymerase (NEB) and sigma water in a final volume of 50  $\mu$ L. PCR products were analysed using agarose gel electrophoresis and purified using Qiagen Gel Extraction kit (Cat no. 28704).

**Table 2.6: Oligonucleotides used for ECR amplification**

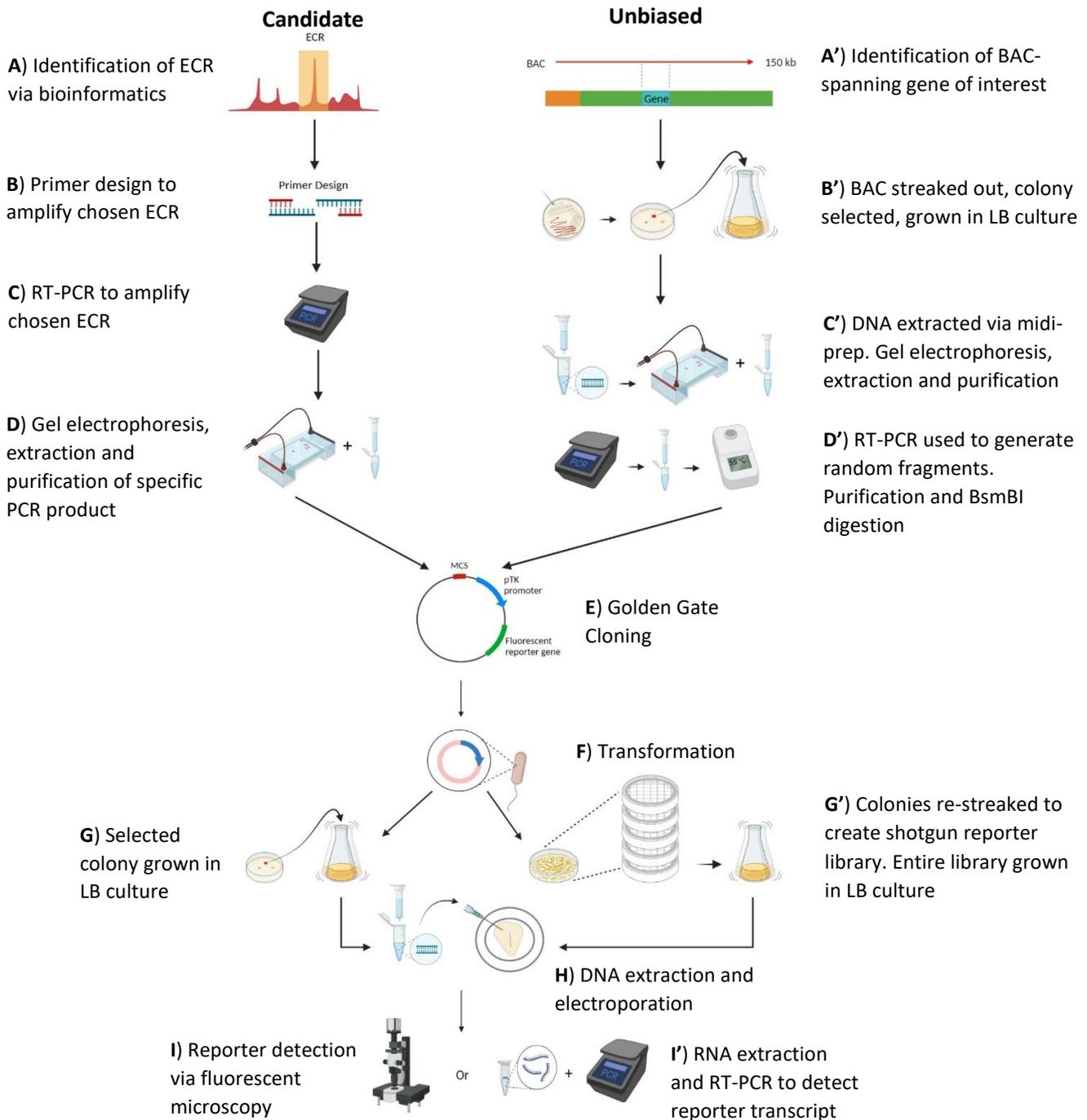
<b>Primer</b>	<b>Sequence (5' – 3')</b>
Myc-ECR2-Fwd	AAAAAA CGTCTC C CCAG TGTGCCCAACATTCTGACAT
Myc-ECR2-Rev	AAAAAA CGTCTC C AACA TCATGTCAGTGGCTGCTTTC

### **Live imaging of enhancer reporter**

Stage HH3Embryos were cultured in six well culture plates and were time-lapse recorded on a Zeiss Observer inverted widefield fluorescence microscope using Zeiss Zen Blue control software. Still images in fluorescent and brightfield were taken from the time-lapse video and analysed in Fiji/ImageJ.

### **Summary of Enhancer Discovery Pipeline**

The steps required for both candidate and unbiased approaches for enhancer screens are outlined in the figure below (**Figure 2.1**).



**Figure 2.1: Diagram outlining the Candidate vs Unbiased enhancer discovery pipeline. Candidate approach outlined on the left. (A)** Phylogenetic foot-printing used to identify ECRs. **(B)** Primers designed to amplify chosen ECR. **(C)** ECR was amplified via RT-PCR using specific primers that adding BsmBI adapters to allow for GG cloning. **(D)** ECR purified using gel electrophoresis, extraction and purification methods. **(E)** GG cloning protocol was employed to ligate ECR into fluorescent pTK-Citrine reporter vector. **(F)** Transformation of reporter vector into competent *E. coli* cells to amplify the plasmid DNA. **(G)** Reporter plasmid grown in LB culture at 37 °C for ~16hrs. **(H)** DNA extracted and purified using midi-prep kit and electroporated into stage HH3 embryos. **(I)** Fluorescent reporter activity, confirming putative enhancer sequence, viewed using confocal microscope. Unbiased approach outlined on the right. **(A')** BAC clone spanning gene of interest selected using UCSC genome browser. **(B')** BAC clone streaked on chloramphenicol plates. Colony selected and grown in low sodium LB culture at 37 °C for ~24hrs. **(C')** BAC DNA extracted via midi-prep kit. BAC human gDNA isolated from host *E. coli* gDNA using gel electrophoresis, extraction and purification methods. **(D')** Gel extracted BAC DNA used as template for Library 1 PCR to produce randomly amplified products. Products purified using column and undergone BsmBI digest to remove concatemers. Fragments size selected using gel electrophoresis, extraction and purification methods. **(E)** Modified GG cloning protocol followed to insert size selected fragments into fluorescent pTK-Citrine reporter vector. **(F)** Transformation of reporter vector into competent *E. coli* cells to amplify the plasmid DNA. **(G')** Reporter library constructed, each clone of the library grown up in 5 mL LB cultures ~16hrs. **(H)** DNA extracted and purified using midi-prep kit and electroporated into stage HH3 embryos. **(I')** RNA extracted from embryos, cDNA synthesis performed, and RT-PCR utilised to detect reporter activity.

## Chapter 3: Shotgun Library Construction

### 3.1 Introduction

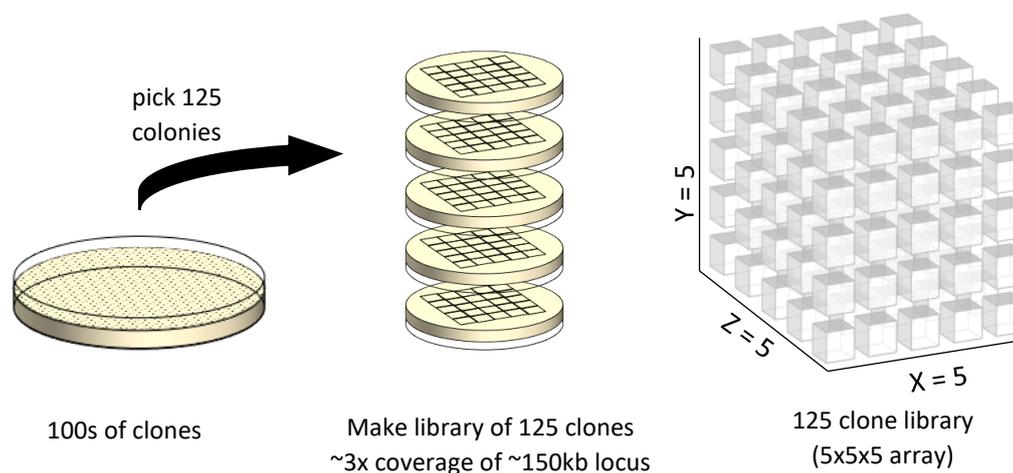
Massive parallel reporter assays quantified using deep sequencing or NanoString technology (Nam and Davidson, 2012) provide a powerful tool in which to mass-screen potential enhancer sequences. But a major disadvantage to these methods includes the limited fragment length (< 300 bp) produced through chemical synthesis.

An alternate approach to deep sequencing is the production of shotgun reporter gene libraries. Where large portions of the genome can be screened in a short amount of time. Shotgun libraries are made up of smaller fragments of DNA (with a standard size), broken down from a larger target DNA e.g, a section of the human genome; end sequencing these randomly generated smaller fragments to reassemble the initial target sequence.

To generate libraries made up of large (1kb+) DNA fragments the use of Bacterial Artificial Chromosome (BAC) clones can be employed. BACs are a vector specifically designed for the cloning of large DNA fragment based on the replication of the F factor in *Escherichia coli* (Shizuya et al, 1992). Increasing the size of the fragments within the randomly generated library, increases the size and coverage of the portion of the genome screened, with the random sub-fragments then cloned into a minimal promoter-reporter vector.

Our approach looked at creating a more efficient, low-cost, high-throughput method for shotgun library construction and screening. With this method, multiple enhancers linked to diseases could then identified by systematic functional screening in parallel to genome wide approaches. Exploiting group testing strategy (Mutesa et al, 2020) to streamline screening of the reporter library.

Each colony generated is a molecular clone, made up of a fluorescent reporter plasmid (pTK-Citrine), and a random fragment insert. Construction of the genomic shotgun reporter library consisting of 125 clones can be visualised as a 5x5x5 array, with each clone given a X, Y,Z co-ordinate (**Figure 3.1**).



**Figure 3.1:** Construction of shotgun reporter library. Utilising group testing strategy to create a 3D (5x5x5) array of 125 clones, each with a clone co-ordinate within the library.

This shotgun reporter library can then be screened by co-electroporating 25 clones within the same group.

### 3.1.1 Fragmentation

Historically, shotgun libraries are produced by shearing the start DNA; either through enzyme digestion or with physical force (Rohwer et al, 2001).

The use of a nebulizer is commonly employed to physically shear DNA (Roe, 2004). The nebulizer utilises compressed nitrogen or air to force the target DNA repeatedly through a small hole, thus producing randomly sheared double stranded DNA molecules either containing 3'- or 5' overhangs as well as blunt ends (Knierim et al, 2011). Nebulization while popular due to its low cost, is not a suggested method if working with limited starting material due to its high percentage loss of DNA (Genohub Blog, 2016).

Other methods to physically shear DNA is the use of a sonicator or an appropriate gauge needle attached to a hypodermic syringe. In sonication the DNA is suspended in solution and subjected to ultrasonic waves. These vibrations create gaseous bubbles which disrupt and shear the larger DNA fragments (Roe, 2004). Using the syringe and needle, the DNA is mechanically shared through the repetitive pushing and sucking of the solution through the needle at a rate determined by your desired length (Matsumata et al, 2005).

Enzymatic digestion is an alternative method for randomly shearing DNA. There are commercial enzymatic fragmentation kits available, such as the NEBNext™ (New England Biolabs). Kits such as this one generally contains a double stranded DNA fragmentase consisting of two enzymes. One to generate random nicks in the double stranded DNA and the other to cut the strand opposite the nicks. This ensures the double stranded DNA fragments contain short overhangs with 5'-phosphates and 3'-OH groups. A DNA ligase is also present in these kits to repair single stranded nicks.

Methods have been developed using random primers to amplify the target DNA creating library fragments (Head et al, 2014). The random primers are short oligonucleotides that have a random sequence, generally 6-8 bp in length. These random sequences bind in an unbiased fashion to the target DNA.

Following fragmentation of the DNA, either through mechanical or enzymatic methods, the standard procedure for library construction includes end-polishing, addition of adapter sequences, size selection and PCR amplification steps.

The invention of the Tn5 tagmentation protocols (Adey et al, 2010) allow for transposase-mediated adapter insertion and fragmentation to occur in a single *in vitro* reaction. Kits, such as Nextera (Illumina), contain a hyperactive derivative of the Tn5 Transposase (Goryshin & Reznikoff, 1998) which work through a 'cut and paste' mechanism. Tn5 is able to excise itself from donor DNA and integrate itself into a target genome DNA sequence (Reznikoff, 2002). As a result of the insertion, 9 bp of the target DNA is duplicated. In commercial kits free synthetic mosaic end (ME) adaptors have replaced the two inverted 19-bp end sequences (ESs) found in wildtype Tn5. Therefore, the engineered enzyme has dual activity; fragmentating the DNA and adding specific ME adaptors to the 5' end of target DNA (Adey et al, 2010). These adaptor sequences can be used to amplify the insert DNA by PCR.

### 3.1.2 Cloning

Once fragments have been generated, they must go through end-polishing and adaptor-ligation steps to be inserted into a reporter plasmid for library construction. There are various methods by which this can be achieved, some of which are outlined briefly below.

Blunt-end cloning involves the ligation of insert and plasmid vector whose terminal ends have no overhanging bases i.e. 'sticky-ends' (Ausubel et al, 1989). It is considered one of the easiest methods for cloning double stranded DNA into a plasmid vector due to its little to no preparation of the insert. Thereby avoiding any subsequent purification steps and loss of insert material for cloning. However, it comes with its own set of challenges as it cannot benefit from the hydrogen bond stabilization derived from the complimentary overhanging bases used in cohesive-end bonding. The successful ligation of insert to vector relies on a temporary association between 5'-phosphate and 3'-OH groups (Liu & Schwartz, 1992).

Preparation of the vector and insert are required for the removal or filling of overhangs to create blunt ends (Sambrook & Russell, 2012). The vector can be digested with a restriction enzyme that gives a blunt end if the multiple cloning site (MCS) contains a restriction enzyme with this feature (e.g., EcoRV) (Motohahsi, 2019). This is the recommended method of vector preparation for blunt-end cloning. It is also possible to digest with a restriction enzyme that generates overhangs, however this then requires the subsequent step of removal or filling of nucleotides. PCR can also be utilised as a method for vector preparation. Linearized plasmid can be enriched by amplification using high-fidelity polymerase and PCR primers designed with their 5' ends at the desired insertion sites (Wang et al, 2004). The linear PCR product can be separated from supercoiled plasmid template by agarose gel electrophoresis, extraction and purification methods. Note the plasmid is generally dephosphorylated for in order to prevent re-circularization (Liu & Schwartz, 1992).

Preparation of the insert for blunt-end cloning can vary due to the method of fragmentation used (Knierim et al, 2011). Inserts produced by PCR using a high-fidelity polymerase that leave blunt ends requires no preparation. Inserts produced via physical fragmentation such as shearing, or sonication generally will not have blunt ends. As such a polishing reaction is required either to remove DNA overhangs or to fill in missing bases if there is a 3'-OH group available for priming. Note that for blunt-ended cloning to be successful the insert needs to be phosphorylated on the 5' end (Liu & Schwartz, 1992).

Advantages to blunt-end cloning include the versatility of the method due to the insert not requiring any restriction sites for successful cloning. This does not pose any sequence restrictions onto the insert and vector selected. As the insert generally needs little to no preparation for blunt-end cloning this can speed up the experimental time. Disadvantages to blunt-end cloning include it being 10-100x less efficient than cohesive or 'sticky'-end cloning (Sambrook & Russell, 2012). This can lead to fewer colonies being produced and a high percentage of the recombinant colonies produced containing the insert in the wrong orientation. It should be noted that plasmid re-circularization is a common occurrence during blunt-end cloning. This low number of colonies containing the insert in the correct orientation is insufficient to generate a library.

Restriction endonucleases cleave DNA at highly specific nucleotide sequences (NEB, 2022). As such they can be used to create complimentary or 'sticky' 5' or 3' overhangs in both vector and insert. This method is known as cohesive-end cloning due to the hydrogen bonding that occurs between both DNA ends prior to ligation. The main advantages of cohesive-end cloning include the specificity of type II restriction enzyme, which allows for directional cloning of the insert (Loukianov et al, 1997), and the hydrogen bond interaction prior to ligation which increases the efficiency (Sambrook & Russel, 2012).

Preparation of the vector includes selecting a plasmid containing the three basic key features for cloning: the origin of replication, the MCS containing a number of restriction enzyme recognition

sites, and an antibiotic resistance gene (Sambrook & Russel, 2012). The vector is then digested by the chosen enzyme(s), dephosphorylated and purified ready for the ligation reaction.

For cohesive-end cloning the insert is designed to include the restriction sites that also occur in the vectors MCS, but not found elsewhere. The insert is then digested separately from the vector by the chosen enzyme(s) creating complimentary overhangs. The insert is then purified and subsequently ligated into the vector by a covalent reaction, initially aided by hydrogen bonding between the 'sticky' ends.

The advantages of cohesive-end cloning far outweigh the disadvantages in most cases, with problems such as nonspecific cutting (star activity) fixed by changing reaction conditions such as sufficient reaction buffer etc. A problem that can occur with this method is the amount of steps and rounds of purification needed of both vector and plasmid can lead to insufficient material to clone with.

Other noteworthy cloning methods include TA cloning (**Chapter 5**), which is a simple non-directional cloning of PCR products, and Golden Gate cloning (**Chapter 6**), which allows for directional assembly of multiple DNA fragments. These methods are described in more detail in Chapters 5 and 6 respectively.

Traditional cloning methods to create shotgun libraires have certain limitations, such as bias when fragmenting the DNA through partial digestion. The restriction enzyme used preferentially cuts certain DNA sequences creating biased libraries. Large quantities of genomic DNA are required for successful library construction. Blunt-end ligation is inefficient and leads to a low number of clones, insufficient to generate a library.

Our approach looked at combining random primed PCR to create library fragments with Golden Gate cloning methods to both amplify and attach adapter sequences for cloning in a one-step reaction. We also look at comparing this with more traditional fragmentation and TA cloning methods for successful library preparation. Our approach looked at creating a more efficient, low-cost, high-throughput method for shotgun library construction and screening. With the hope this method could be used to identify the possible multiple enhancers linked to diseases through systematic functional screens of the genome.

## 3.2 Results

### 3.2.1 Proof of Concept: Validating Group Testing Strategy

#### 3.2.1.1 Dilution of Fluorescent Plasmid

To exploit the group testing strategy employed in this report we first tested if it was possible to detect reporter activity of 1 in 25 co-electroporated plasmids when viewed under a fluorescent microscope. As each singular plasmid is diluted down. To model this we diluted down a known *Meox1* reporter construct (*Meox1*-Reporter) (Mok et al, 2021), that expressed fluorescent Citrine in the somites, using the reporter plasmid with no insert (empty reporter plasmid) in do a series of titrations (**S1**).

These serial dilutions (1:1, 1:5, 1:25, 1:125) were electroporated, along with a control plasmid (Cherry), into stage HH3 chick embryos of the tissues surrounding the primitive streak. Cherry is a red fluorescent protein (RFP) driven by the ubiquitous  $\beta$ -actin promoter. At HH8-10 RFP expression is observed throughout the embryo (**Figure 3.2.1.1 A, C, E, G**), showing the electroporation's were

successful. *Meox1* expression can clearly be seen in the somites for the 1:1, 1:5 and 1:25 dilutions (Figure 3.2.1.1 B, D, F). With no *Meox1* expression seen in the 1:125 dilution (Figure 3.2.1.1 H).

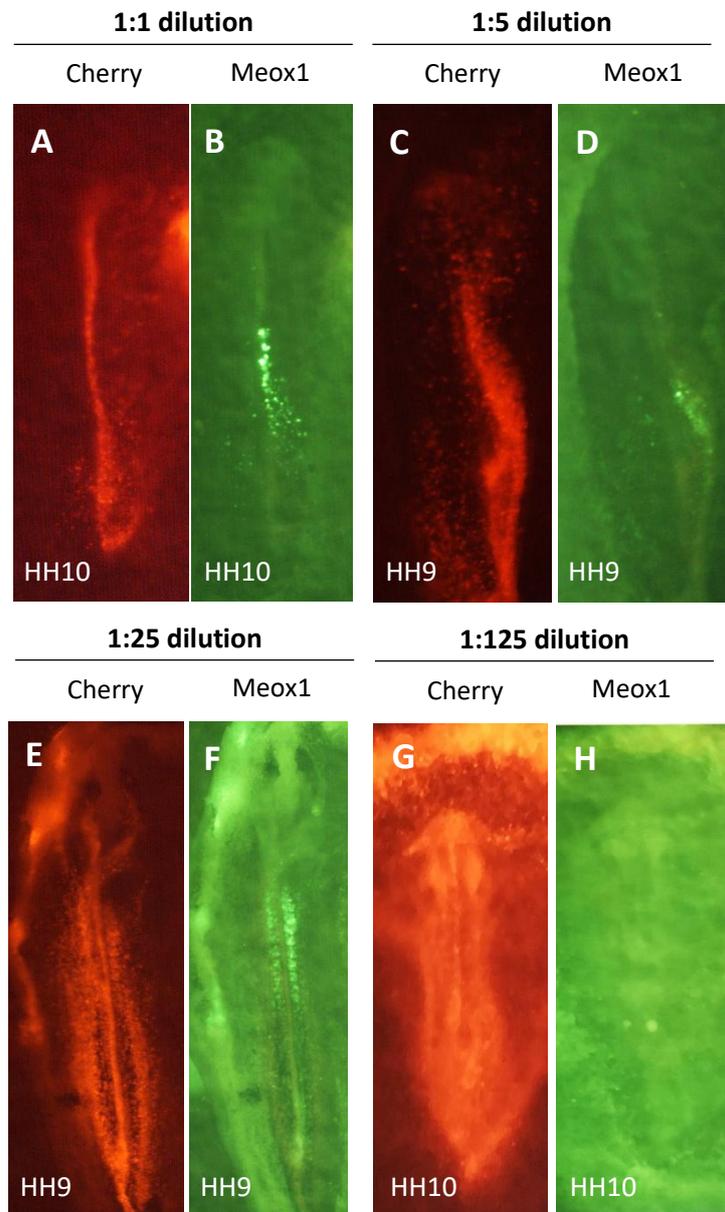


Figure 3.2.1.1: Electroporation of titrated reporter plasmid. Cherry is expressed for the control plasmid using a B-actin promoter and indicated the location of successful electroporation (A, C, E, G), while Enhancer *Meox1*-driven Citrine is visible only in the somites of 1:1, 1:5 and 1:25 dilution (B, D, F). No fluorescence present for 1:125 dilution (H).

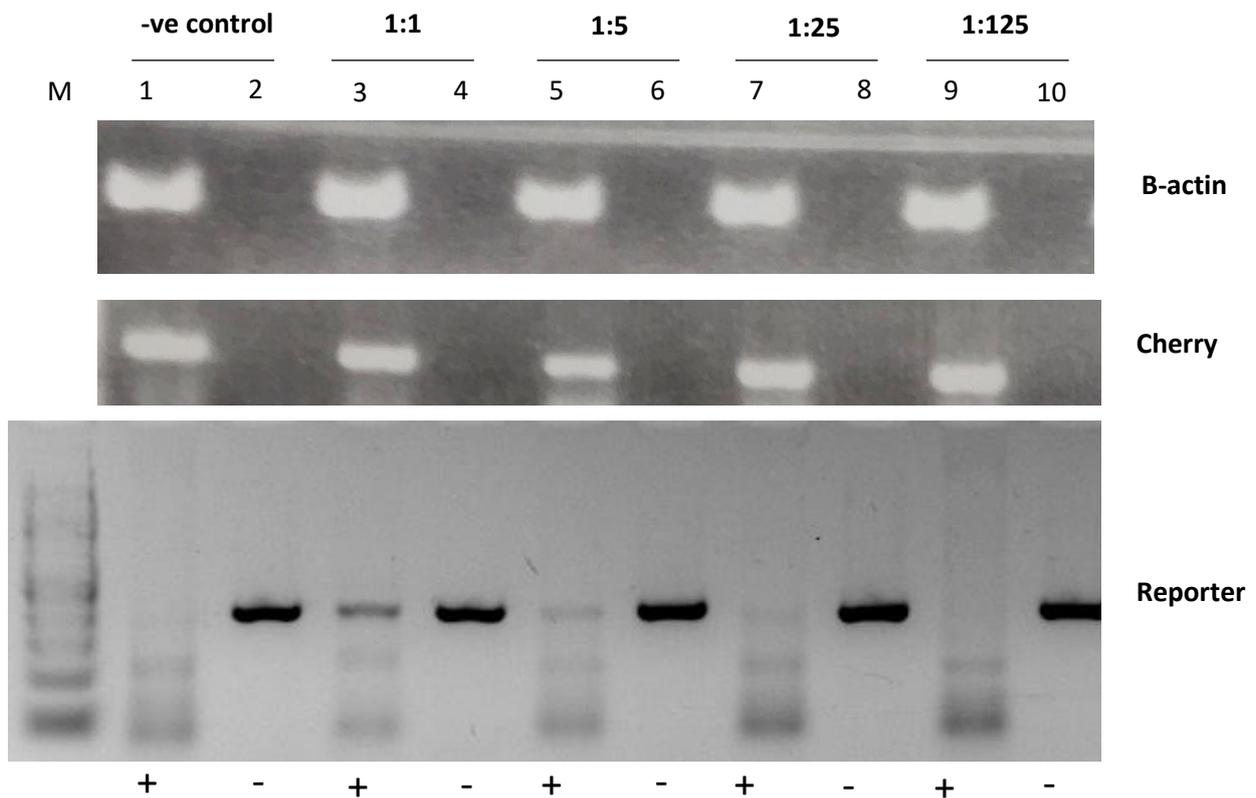
This result suggested that co-electroporating 25 plasmids (one, X, Y or Z group) (Figure 3.2.1.1) into a single embryo is a viable option for screening reporter activity via fluorescent microscopy.

### 3.2.1.2 RT-PCR for detection of reporter transcript

Visual inspection may not work for all/weak enhancers, as such embryos showing strong control RFP fluorescence were dissected, RNA isolated and 1 µg of cDNA was used as template for a standard 30 cycle RT-PCR with specific primers (Table 2.4) to confirm the expression of the reporter transcript. This result showed us it was possible to detect the reporter transcript (240bp) for the 1:1, 1:5 and

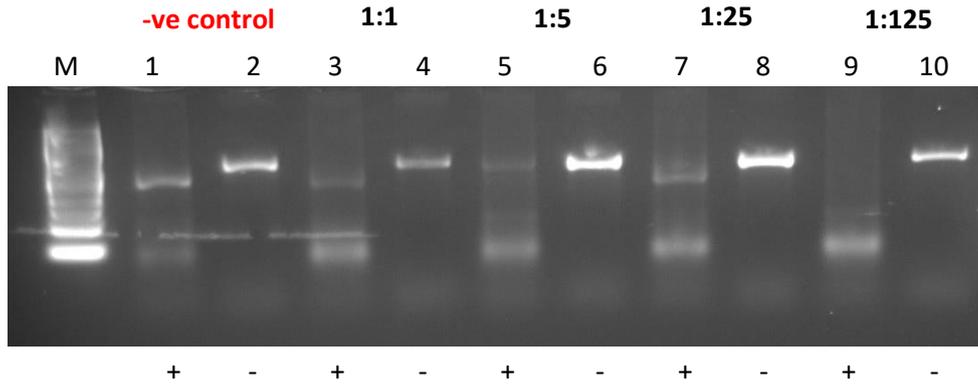
1:25 dilutions but not the 1:125 (**Figure 3.2.1.2.1**). However, a band (550bp) can also be seen in the RT- lane, this is not contamination or the presence of cDNA, we believe this is the electroporated vector.

There is no cDNA in the RT- for the polymerase to preferentially amplify, as evidenced by the empty B-actin RT- lanes, so the only template left to amplify is the plasmid itself. This is consistent with the observed product size as the primers were designed to flank the synthetic intron to allow for distinction between the cDNA and the vector as the intron is spliced out.



**Figure 3.2.1.2.1: Reporter transcript detection by RT-PCR.** Positive control B-actin shows cDNA levels are not consistent across the lanes. Lane M shows 100bp maker. Lanes 1, 3, 5, 7 & 9 are RT + lanes. Lane 1 is negative control. Lanes 3, 5, 7 & 9 show titrated reporter plasmid containing cDNA (RT+). Lower unspecified bands are present in the RT+ lanes but they can clearly be distinguished from the higher molecular weight transcript (417bp). Lanes 2, 4, 6, 8 and 10 are negative control lanes, absent of cDNA (RT-). Bands present in the RT- lanes from vector (550 bp).

A new reverse primer (**Table 2.4**) and a touchdown PCR was performed to try and remove the unspecified bands seen in the RT+ and the RT- lanes. The results (**Figure 3.2.1.2.2**) showed a band corresponding the vector (550bp) was still present in the RT- but the same pattern could be seen for the mRNA band (417bp) in the RT+ lane, whereby it was visible until the 1:125 dilution.



**Figure 3.2.1.2.2:** Detection of reporter transcript by Touchdown RT-PCR. Lane 1 & 2, negative control; lane 3 & 4, 1:1 dilution; lane 5 & 6 1:5 dilution; lane 7 & 8 1:25 dilution; lane 9 & 10 1:125 dilution. Band present in RT+ (cDNA present) that corresponds to mRNA (417bp) (Lanes 3, 5, 7). Band present in RT- (cDNA absent) that corresponds to vector (550 bp) (Lanes 2, 4, 6, 8, 10). 100bp size marker included (M).

The negative control lane has been highlighted in red (**Figures 3.2.1.2.1 & 3.2.1.2.2**) as when this assay was performed, it could not be confirmed whether this cDNA used was actually a negative control. As indicated by the band present in the RT+ this most likely is not a negative control and as such needs to be repeated to validate detection of reporter transcript assay.

These results give us confidence that we can detect transcripts diluted 1:25, validating our group testing strategy methods.

### 3.3 Discussion

Our results indicate that we can use RT-PCR to detect reporter transcripts following a titrated 1:25 electroporation. However, due to the presence of multiple bands in the RT- PCRs indicative of plasmid DNA being amplified we wanted to edit the pTK-Citrine plasmid to include a more suitable priming site that straddles the synthetic intron therefore only mRNA would be detected and not plasmid DNA. This also presented the opportunity to create 5 plasmids, each with a unique barcode sequence (Chen & Streit, 2005) to further accelerate the screening process.



Traditionally fluorescent reporter assays are limited by their means of detection, as it requires high levels of the reporter plasmid to be present to be seen by a fluorescent microscope. Factor in the co-electroporation of multiple (25) plasmids into a single embryo, the signal of each reporters' fluorescent proteins is diluted 1:25. As such the modified reporter vectors, containing the barcode edits increases the sensitivity and efficiency of enhancer validation.

Our approach used the barcode sequences based on a previous study by Chen & Streit (2005). Unfortunately, we could not use the vectors from this paper as they were designed for bunt-end cloning. We plan to utilise a Golden Gate Cloning protocol (**Chapter 6**) which requires vectors containing BsmBI restriction sites. Our vector also had the benefit of blue/white colony selection.

## 4.2 Results

A 16bp barcode (Chen & Streit, 2005) was introduced downstream of the minimal promoter and immediately upstream of the synthetic intron. Five different barcodes were inserted into individual vectors (**Table 2.1**), generating five unique edited vectors. This was verified by Sanger sequencing (**Figure 4.2**).



**Figure 4.2: Modified pTK-Citrine vectors.** Unique DNA barcodes (16bp) are inserted upstream of the chimeric intron. The forward (T1F1, T2F1...etc) and reverse primers (T1R1, T2R1.....etc) used for this modification are labelled in purple on the forward and reverse strands respectively. Primer sequence shown in capital letters correspond to the barcode sequence. Nucleotides with small case are nucleotides required for primer pairing. Barcode edit 1 (**A**), edit 2 (**B**), edit 3 (**C**), edit 4 (**D**) and edit 5 (**E**) successfully inserted into the vector. Raw chromatogram data showing sequencing software's interpretation of nucleotides (**A'**, **B'**, **C'**, **D'**, **E'**). Barcode 1 (**A'**), 2 (**B'**), 3 (**C'**), 4 (**D'**), 5 (**E'**) highlighted in blue. Probability of A, T, G, C nucleotide shown by green, red, black and blue.

It should be noted that for Barcode-edit 3 (**Figure 4.2 C**), an extra nucleotide, C, was added into the sequence. It is unknown if this nucleotide is actually present or is a sequencing error. The software used to match the reverse primer, which ends in gttagg, is unable to perfectly match with the reverse strand (**Figure 4.2 C**). However, the presence of this nucleotide does not affect the function of this barcode or the reporter gene.

The successful modification of these vectors allows for rapid detection of the plasmid via PCR. This is a sensitive and cost-effective way to assess enhancer activity that contributes to the efficiency of our enhancer screening assay. The PCR primer sequences required are taken from the Chen & Streit paper (Chen & Streit, 2015) and shown in **Table 2.1**.

## 4.3 Discussion

### 4.3.1 Optimising the pTK-Citrine reporter vector for group-testing of shotgun reporter libraries

The modifications added into the pTK-Citrine vector, highlighted in **Chapter 4.2**, resulted in the creation of 5 new reporter plasmids, each containing their own unique barcode edit. The edits taken from the Streit and Chen paper (Streit & Chen, 2015) include a further 4 barcodes. This method can easily be expanded to provide additional barcode vectors if required. This barcode strategy allows for rapid, cost-effective detection for enhancer validation via PCR. This assay can be used in conjunction with bioinformatic predictions and genome-wide experiments for efficient validation of enhancers within a functional screen. This customised method provides an alternative to the bottleneck of generating transgenic animals (Kvon, 2015).

## Chapter 5: Library Construction via TA Cloning

### 5.1 Introduction

TA cloning is a simple method that allows nondirectional cloning of PCR products. TA cloning relies on the ability of adenine (A) and thymine (T), located on the compositable 'sticky ends' of the insert and vector, to weakly associate through hydrogen bonding while the present ligase enzyme repairs the phosphodiester backbone, covalently binding the insert and vector together.

The insert is prepared by attaching a single deoxyadenosine (A) to the 3' end of the DNA. This is generally done using Taq polymerase due to the terminal activity of this non-proofreading enzyme (Clark, 1988); unlike other various DNA polymerases, such as Phusion, which possess 3' to 5' exonuclease activity for proofreading and produce blunt-end DNA fragments (Wang et al, 2004).

Vectors can be prepared for cloning by generating a T-overhang at the 3' end. These vectors are sometimes referred to as 'T-vectors' and can be bought commercially. To prepare the T-overhang, the vector can be digested by any restriction enzyme that creates a blunt end and by incubation with Taq polymerase in the presence of dideoxythymidine triphosphate (dTTP), but in the absence of all other nucleotides, a deoxythymidine is incorporated at the 3'-terminus of a linearized blunt end vector by using terminal deoxynucleotidyl-transferase (Motohashi, 2019). It is also possible to prepare the vector with the well known type II restriction enzyme, XcmI, which has an asymmetric recognition site designed to produce the 3'-end T overhang (Scutte et al, 1997).

### 5.2 Results

Utilising the UCSC genome browser, a BAC clone was identified, RP11-771F21, that spanned a 169kb region of human chromosome 2 which encompassed the *Six3* locus (**Figure 5.2.1**). Traditional shotgun library methods were adapted to attempt construction of a *Six3* shotgun reporter library. The BAC clone was subjected to a Plasmid-Safe DNase digest to eliminate the presence of host *E. coli* genomic DNA. The BAC DNA was randomly fragmented by mechanical shearing caused by the pushing and sucking of the plasmid safe treated BAC clone suspension through a 30-gauge needle 300 times.

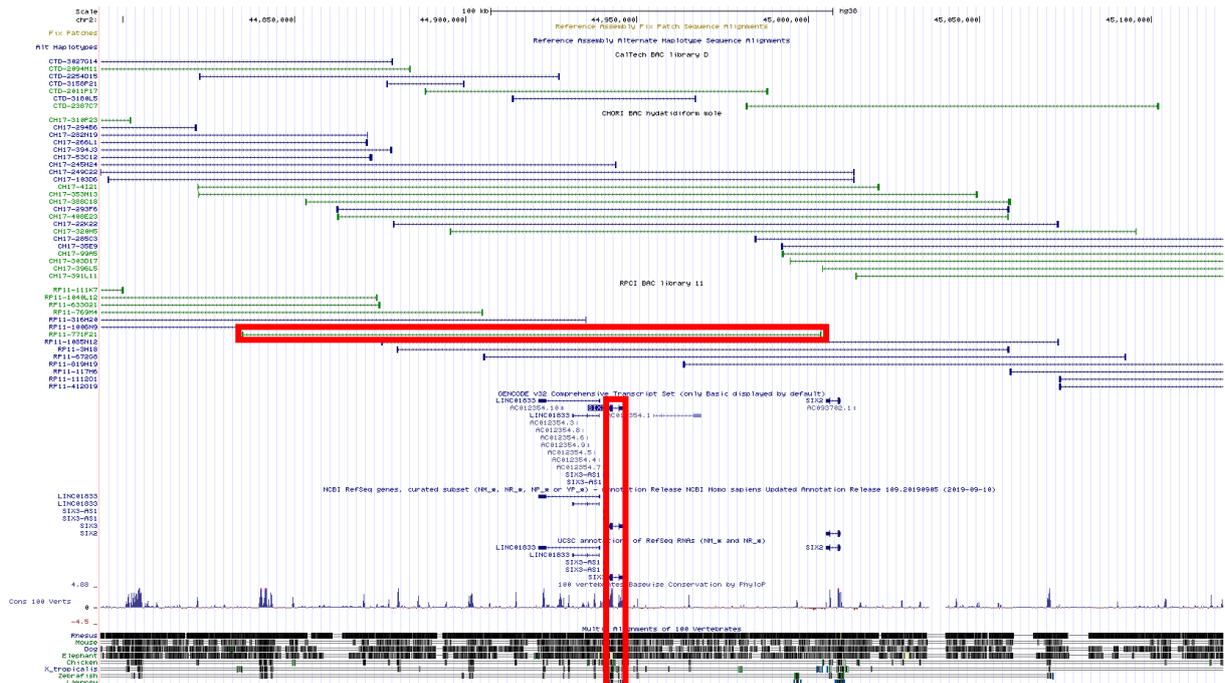


Figure 5.2.1: Successful identification of a single BAC clone (RP11-771F21) covering the entirety of the *Six3* gene. Snapshot of the UCSC genome (hg38) browser showing the *Six3* gene (chr2:44,871,054-45,016,720) within the genome, highlighted by vertical red box. BAC clones from Caltech, CHORI and RPC1 libraries shown spanning this region of the genome shown above in blue and green. BAC clones shown in green indicate verification of the clone by fluorescent *in situ* hybridization. Clone RP11-771F21, highlighted in red horizontal box, was identified to span both upstream and downstream of the *Six3* gene.

The 3' overhangs of the fragments were filled-in A-tailed using Taq polymerase which preferentially leaves a single adenosine overhang at the 3' ends. It is worth noting that any 5' overhangs would not be polished so this might hinder ligation. The fragments were size selected by agarose gel electrophoresis, fragments spanning 2-8kb were excised (Figure 5.2.2) and purified using Qiagen Gel Extraction Kit.

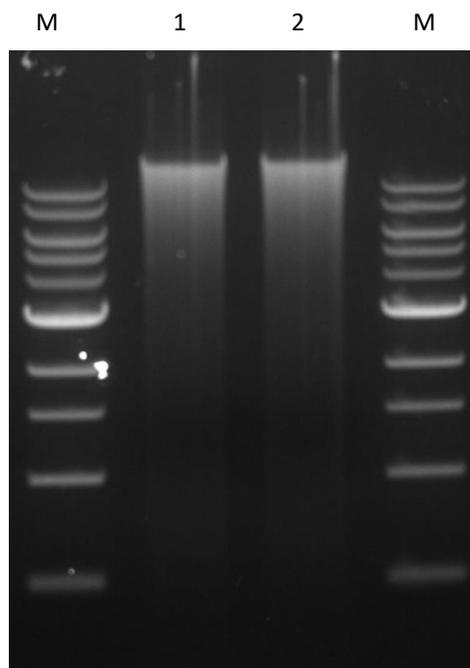
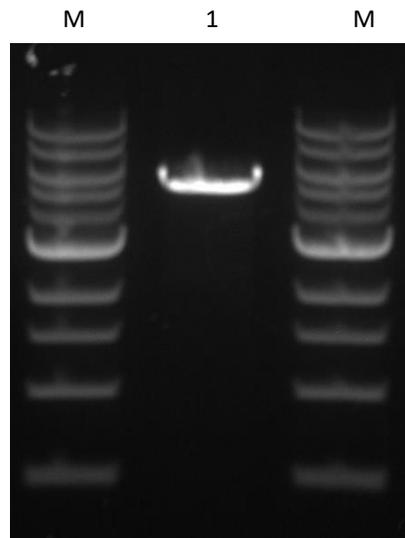


Figure 5.2.2: Polished RP11-771F21 BAC fragments. Lane 1 & 2, randomly fragmented RP11-771F21 BAC fragments. Fragments seen across lanes 1 & 2 and range from +10kb to 500 bp. The fragments seem to be concentrated in larger than 10kb fragments. 1kb size marker included either side of the amplified fragments (M).

The vector was prepared by digestion with the PvuII enzyme, this gave an expected product size of 4.9kb (**Figure 5.2.3**), purified using a column and polished in a T-tailing reaction using GoTaq.



**Figure 5.2.3:** PvuII digested pTK-Citrine-Bacrcode 1 vector. Lane 1, digested product at 4.9kb in size. 1kb size marker included (M).

These 3' A-tailed fragments were inserted into the 3' T-tailed pTK-Citrine-Barcode 1 reporter vector using TA cloning reaction conditions. With an average insert size of 4.5kb or larger, this would comprise a library that fully covers the *Six3* gene region and give more than 3x coverage of the original BAC clone.

Unfortunately, upon inspection of the colonies generated from TA cloning, very few (~20) white colonies were generated. Colony PCR was performed on 10 of these clones (**Figure 5.2.4**). The controls present during this cloning protocol revealed the insert was unclonable when in the presence of the positive control pGEM-T Easy Vector. The negative control (ptk-Citrine-Barcode 1 vector + no insert) showed a high number of clones re-ligated themselves.

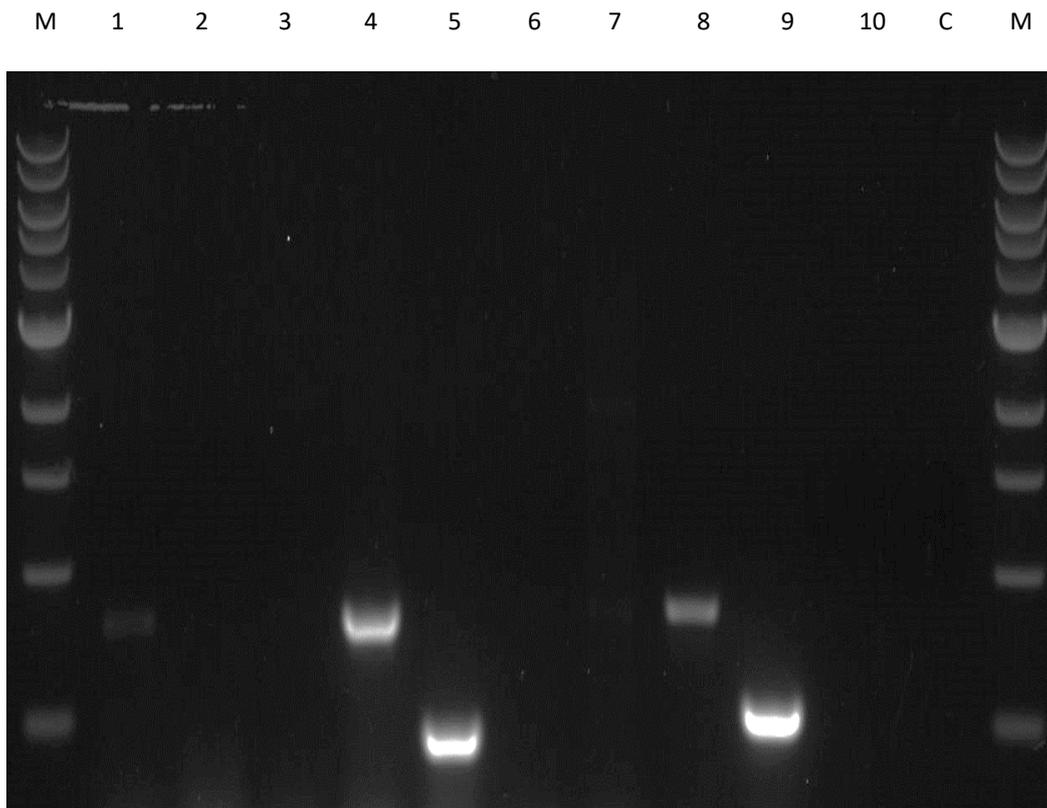


Figure 5.2.4: Colony PCR for Six3 library creation using TA cloning method. 10 colonies were selected for colony PCR. Lanes 1, 4, 7 and 8 showed colonies that gave a PCR product  $\sim 825$ bp. Lanes 5 and 9 showed colonies that gave products  $< 500$ bp. Lanes 2, 3, 6 and 10 showed colonies that gave no product. Negative control (C) was included. 1kb size marker included either side of the selected colonies (M).

The colony PCR showed two of the clones contained an insert  $\sim 825$  bp, which corresponds to the original *LacZ* gene. While these colonies appeared to be white instead of blue, a possible explanation could be following the PvuII digestion and T-tailing reaction, the ligase re-ligated the plasmid back together but with an extra nucleotide present, causing a frameshift and therefore non-functional *LacZ* gene. Six colonies gave no PCR product, this could indicate no insert, or the insert present was too large to be generated with the extension time used for this PCR, which was 4 minutes.

Colonies that generated no visible PCR products or products  $< 500$  bp in size were sent for sequencing. Sequencing data confirmed that the clones producing smaller PCR products contain artifacts that retain plasmid sequence but are missing portions or the entirety of the *LacZ* gene. This gave a non-functional *LacZ* gene, and therefore a white colony, but with no insert present. The sequencing data for the colonies that generated no product showed not only was the *LacZ* gene missing but the sequence for the reverse primer.

The sequencing data confirmed that no insert was present, and the white colonies generated were deletion artifacts. This information was not surprising as, the positive pGEM-T Easy control indicated that the insert used was not cloneable with the conditions used. That artifacts of the vector were generated for the TA cloning method showed the preparation of the insert and vector was not optimal, as artifacts were not produced during the pGEM-T Easy control.

## 5.3 Discussion

### 5.3.1 Reporter Library Construction via TA cloning

Construction of a shotgun reporter library for the Six3 gene had different challenges depending on the strategy used to build the library. Using more traditional shotgun library construction methods, highlighted in **Chapter 5**, fragmentation of the BAC and TA cloning was not an efficient strategy conducive to building a library containing large inserts. Subsequent colony PCR and sequencing data confirmed the preparation of the insert and vector was not optimal, causing artifacts to be produced. A limiting factor of using T-tailing is only the 3' overhangs are filled. This meant that cloning would only be successful if single nucleotide 3' overhangs were present on both ends ( $1/4$  of fragments); and that the 3' overhangs had an adenosine (A) nucleotide on both ends ( $1/4 \times 1/4 = 1/16$  of fragments). Therefore, the chance of any one single fragment being cloneable in our assay was:  $1/4 \times 1/16 = 1/64$ .

It should also be noted that the chance of any fragment having a 3' overhang on one end only ( $1/4$  of fragments) and that nucleotide being an A (also  $1/4$ ) is:  $1/4 \times 1/4 = 1/16$ . Being more abundant, these  $1/16$  single-3'-A-fragments would naturally outcompete the desired  $1/64$  double-3'-A-fragments and so would competitively inhibit the generation of fully ligated plasmids.

Other enzymes are available that would polish both 3' and 5' overhangs, which could then be specifically A-tailed in a subsequent reaction in which only dATP is present.

The TA cloning protocol relies on an abundance of fragments being present, so enough of them will be cloneable. However, even with optimal preparation of both vector and insert this approach is labour-intensive and requires microgram quantities of DNA to digest to generate a single final recombinant molecule (Matsumura, 2018).

While this has been a reliable protocol to produce fragment-based libraries in the past, to ensure a high-quality genomic DNA library is constructed, the DNA fragments are limited to below 1kb on average. This process, highlighted in **Figure 5.3.1**, is time-consuming and has many parameters, possibly affecting the quality of the library due to the multiple steps required. The loss of material associated with each subsequent step is high, increasing the amount of starting material required to generate a library of considerable size.

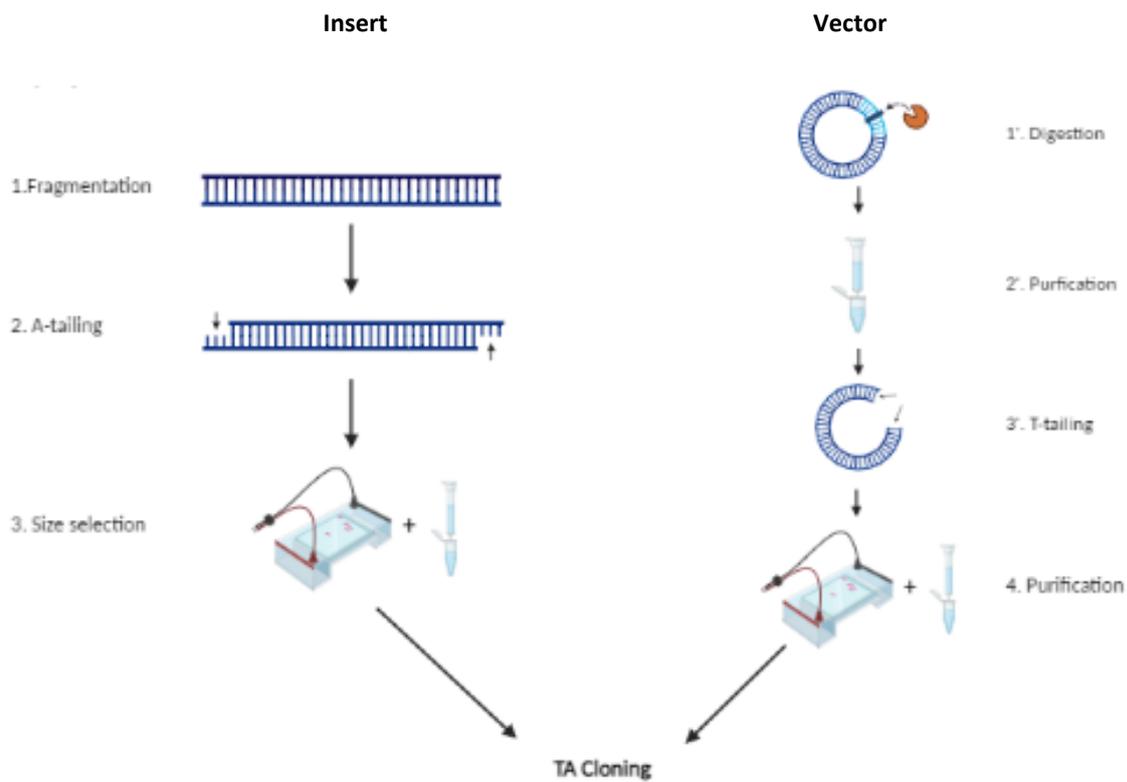


Figure 5.3.1: Overview of Library preparation for TA cloning protocol. On the left is preparation of the insert. 1. Fragmentation of the BAC clone by mechanical shearing. 2. A-tailing reaction using Taq polymerase to give sticky ends complementary to vector. 3. Size selection of fragments using gel electrophoresis, extraction and purification methods. On the right is preparation of the vector. 1'. Blunt-digestion of the pTK-Citrine vector by PvuII enzyme. 2'. PCR purification using column. 3'. 3' overhangs filled by T-tailing reaction. 4. Purification of vector using gel electrophoresis, extraction and purification methods. Once insert and vector are prepared TA cloning protocol is employed.

More favourable results were seen utilising an alternative Golden Gate cloning strategy for shotgun library construction. As such this TA cloning strategy was not pursued.

## Chapter 6: Library Construction via Golden Gate Cloning of randomly-primed PCR-amplified sub-fragments

### 6.1 Introduction

Shotgun libraries are made up of fragments of DNA (with a standard size) spanning the genome in an unbiased fashion. Historically, shotgun libraries are produced by shearing the start DNA; either through enzyme digestion or with physical force (Rohwer et al, 2001). The DNA fragments produced are then size selected on a gel, and the ends polished complimentary to subsequent blunt-end ligation (Matsumata et al, 2005).

This approach has certain limitations, such as bias when fragmenting the DNA through partial digestion. The restriction enzyme used preferentially cuts certain DNA sequences creating biased libraries. Large quantities of genomic DNA are required for successful library construction. Blunt-end ligation is inefficient and leads to a low number of clones, insufficient to generate a library.

Golden Gate cloning is a molecular assembly approach that utilizes the simultaneous digestion and ligation of single or multiple fragments in an ordered and scarless fashion. It does this by employing Type IIS restriction enzymes, such as BsmBI, and T4 DNA ligase for rounds of digestion and ligation.

Our approach focuses on generating shotgun reporter gene libraries from BAC clones, using PCR to randomly prime and amplify these BAC fragments. Then combined with Golden Gate cloning to create a high-throughput, low-cost functional screen for enhancer activity.

### 6.2 Results

#### 6.2.1 Optimisation of Library 1 PCR

##### Purification of Library 1 PCR

One of the first parameters thought to be optimised in this report was the use of gel extracted and purified BAC DNA as template rather than directly using midi-prepped BAC DNA. Similar to the plasmid safe digest used for the TA cloning method, this was to eliminate the host *E. coli* DNA from the human genomic BAC clone, so the human genomic DNA is being used as the template. This should ensure that any subsequent amplification and library preparation steps will create a Human shotgun genomic library instead of a bacterial one. However, using our method of purification is insufficient when paired with substantial PCR cycles.

##### Library 1 PCR

Our original plan to randomly amplify fragments of the human genome, and then clone the specified size of insert into a reporter construct was going to require two separate amplifications. The first amplification, called Library 1 PCR, was to randomly amplify fragments using the RP11-771F21 BAC DNA as template. This amplification gave us a range of fragments, which when viewed on a gel was shown as a smear, seemingly concentrated between the 1.5-2kb region (**Figure 6.2.1.1**). Then following gel extraction and purification of fragments sized between 3-4kb, this was used as template for the second amplification step, called Library 2 PCR. This second step was to amplify the size-selected fragments and to add the BsmBI adapter sequences to allow for Golden Gate cloning. The Library 2 PCR gave no amplification of fragments when viewed on a gel (**Figure 6.2.1.1B**). We believe this is due to incomplete separation of the higher molecular weight fragments and primer dimer during size selection following the Library 1 PCR. Meaning, little or no higher molecular weight fragments were present in the template used for the Library 2 PCR.

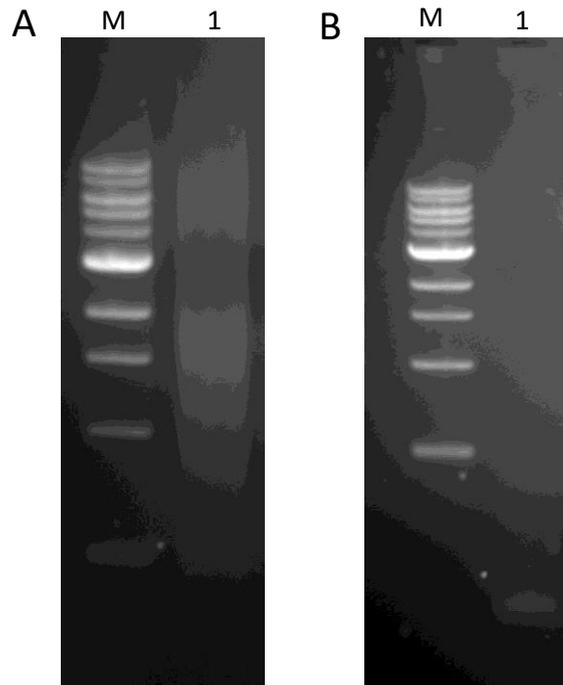


Figure 6.2.1.1: Amplification of fragments for Six3 library. (A) Lane 1, Library 1 PCR of random amplification of BAC fragments. (B) Lane 1, Library 2 PCR with no amplification seen. 1kb size marker included (M).

Primers were designed (**Table 2.2**) to allow for the Library 1 PCR to both amplify and insert BsmBI adapters in a one-step reaction. Therefore, fragments generated by the Library 1 PCR, once purified, could be directly used as insert for Golden Gate cloning. Following unsuccessful attempts of the Library 1 PCR, where no amplification of fragments was seen, optimization of template concentration was determined. Purified BAC-771F21 DNA was subjected to 40 thermal cycles using the random amplification conditions, except the quantity of template DNA in each 50 $\mu$ L reaction was varied as 7 ng, 11ng and 15ng. This experiment showed the brightest smear, when 7ng of template was used, a faint smear for 11 ng and no smear for 15 ng of template (**Figure 6.2.1.2**).

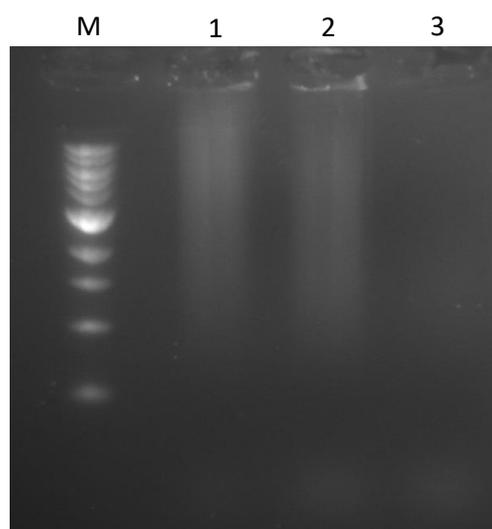
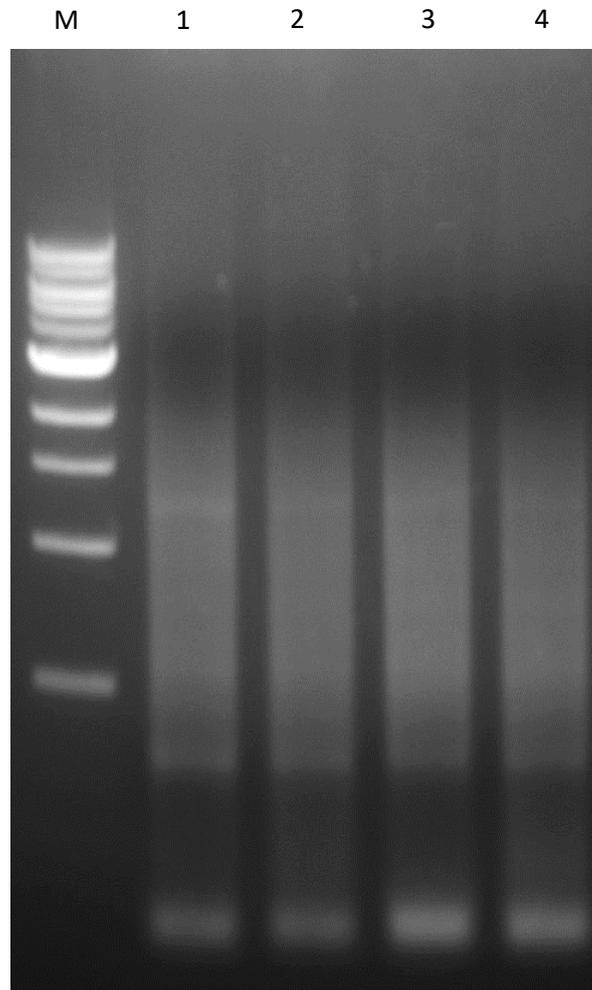


Figure 6.2.1.2: Testing template concentration for Library 1 PCR. Lane 1, 7 ng template used with random amplification present; lane 2, 11 ng template used reduced random amplification seen; lane 3, 15 ng template used no amplification seen. 1kb size marker included (M).

This shows a correlation between template concentration and rate of amplification, with a cut-off point at >10ng of template, in accordance with the manufacturer's instructions (NEB). With 7ng or below the ideal concentration for random amplification using Phusion Polymerase (**Figure 6.2.1.3**).



**Figure 6.2.1.3:** Library 1 PCR with template DNA <7ng. Lane 1; 7 ng template. Lane 2, 5 ng template. Lane 3, 3 ng template. Lane 4, 1 ng template. Amplification of random sized fragments seen for all concentrations of template. 1kb size marker included (M).

The Library 1 PCR the generated a range of fragments ranged between >500 bp to 4kb, but seemingly concentrated between 500bp-1kb region, with molecular weights higher than 4kb absent when viewed on a gel. To determine if different reaction components were limiting the range and concentration of fragments generated, 5 ng of purified BAC-771f21 DNA was subjected to 40 cycles using the standard random amplification reaction conditions but different concentrations of 1.5 mM, 2.5 mM, 3.5 mM and 4.5 mM MgCl<sub>2</sub> was used. This result showed a brighter smear when a higher concentration of MgCl<sub>2</sub> was present. This shows a positive correlation between rate of amplification and the increased concentration of MgCl<sub>2</sub> used (**Figure 6.2.1.4**), indicating MgCl<sub>2</sub> as a limiting factor. But still no change was seen in the amplification of higher molecular weight fragments.

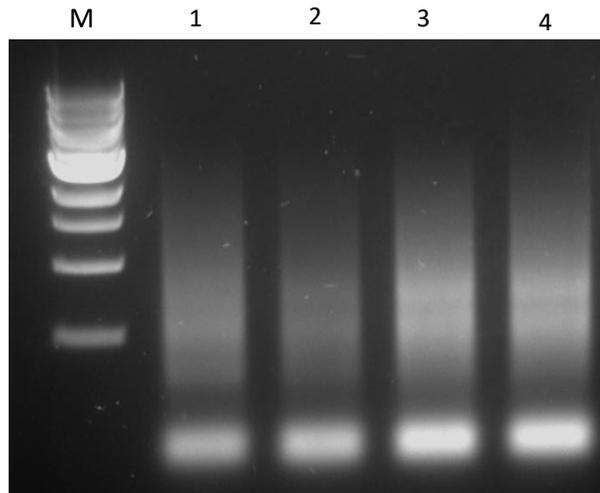


Figure 6.2.1.4: Library1 PCR with varying  $MgCl_2$  concentrations. Lane 1, 1.5 mM  $MgCl_2$ ; lane 2, 2.5 mM  $MgCl_2$ ; lane 3, 3.5 mM  $MgCl_2$ ; lane 4, 4.5 mM  $MgCl_2$ . Amplification present for all  $MgCl_2$  concentrations 1kb size marker included (M).

This led us to hypothesise that primer concentration could also be a limiting factor affecting amplification. To test this we subjected 5 ng of purified BAC-771f21 DNA to 40 cycles using the standard random amplification reaction conditions with 4.5  $\mu M$  of  $MgCl_2$  in each reaction but increased primer concentration to 1  $\mu M$ , 5  $\mu M$  and 10  $\mu M$ . This result showed us random amplification of fragments within the normal range of >500 bp - <4kb region when normal primer concentration was used (**Figure.6.2.1.5**). However, reduced random amplification was seen when 1  $\mu M$  of primers was used, with the amplification only producing lower molecular weight fragments. No random amplification was seen when 5 and 10  $\mu M$  of primers was used. This result suggests that primer concentration is not acting as a limiting factor in this reaction and that higher concentration of primers could be chelating the  $MgCl_2$  present in the reaction, which is a cofactor needed for Phusion polymerase to amplify successfully.

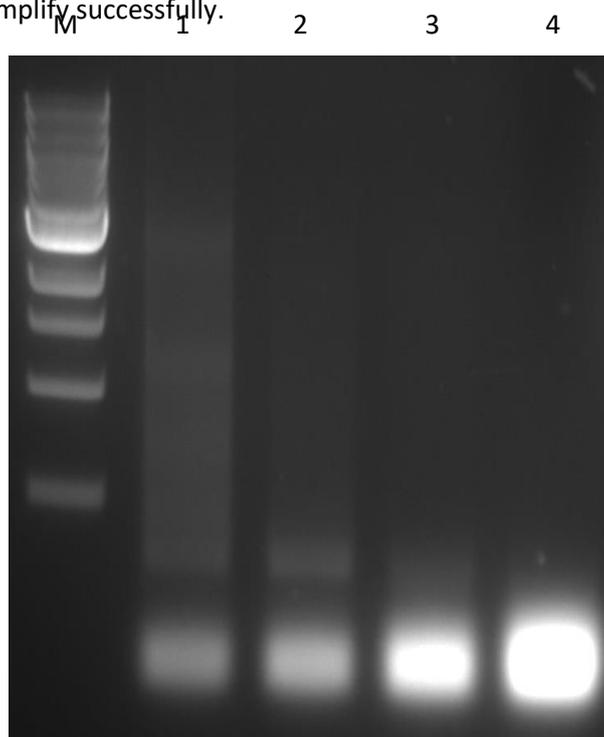


Figure 6.2.1.5: Library1 PCR with varying primer concentrations. Lane 1, 0.5  $\mu M$ , amplification of fragments present; lane 2, 1  $\mu M$ , minimal amplification present; lane 3, 5  $\mu M$ , no amplification present ; lane 4, 10  $\mu M$ , no amplification present. 1kb size marker included (M).

One of the main issues during this project was the inconsistency of the Library 1 PCR. The smears generated were not concentrated at the same molecular weights, or just completely absent. This was thought to be caused by an underlying issue with the template DNA. It appeared when the BAC DNA was eluted in water and stored at 4 °C it was degrading in less than two weeks. As such one of the final optimisations made to the Library 1 PCR was the use of template DNA that has been eluted and stored in elution buffer rather than sigma water. Elution buffer consists mostly of Tris-Cl which acts as a buffering agent when solubilizing the DNA, whilst also protecting it from degradation, such as extremes of pH. When template DNA stored in elution buffer was subjected to 40 cycles using the standard random amplification reaction conditions, increased random amplification was seen (**Figure 6.2.1.6**). This result not only gave a brighter smear, but it showed increased amplification of the higher molecular weight fragments, above 4kb, that had been absent from previous Library 1 PCRs.

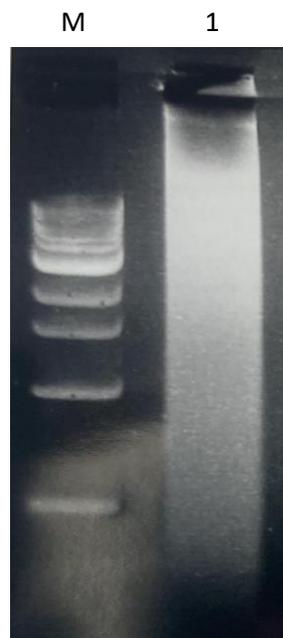


Figure 6.2.1.6: Library1 PCR with template DNA eluted in elution buffer. Lane 1, random amplification of fragments present and concentrated at higher molecular weights (>4kb). 1kb size marker included (M).

However sequencing data received post GG cloning showed the smear generated for these higher molecular weight fragments (**Figure 6.2.1.6**), were not made up of randomly amplified fragments that mapped to the human genome surrounding the Six3 locus. Instead, it showed much smaller fragments inserted into the reporter vector. These shorter fragments were revealed to be primer dimer or larger concatemers, formed from the Library 1 Forward and Reverse primers (**Table 2.2**) annealing to one another, creating a long chain during the PCR, and thus showing as higher molecular weight products. These concatemers contained multiple BsmBI adapter sites, which during the Golden Gate cloning strategy the restriction enzyme would then cleave the concatemers into much smaller fragments, which were subsequently ligated into the vector.

A 1997 paper by Brownie et al, discovered that primer dimer is an unavoidable result after 30 cycles when primers are used in high concentrations, even if there is limited or no complementarity between the two primers (Brownie et al, 1997). It should be noted that both the Forward and Reverse primers used for the Library 1 PCR contained an 8 random nucleotide sequence, that could

result in complimentary between the primers. Brownie et al showed the use of a single primer instead of a Forward and Reverse primer can help eliminate primer dimer from PCRs (Brownie et al, 1997). A method for eliminating concatemer formation during the Library 1 PCR was then investigated.

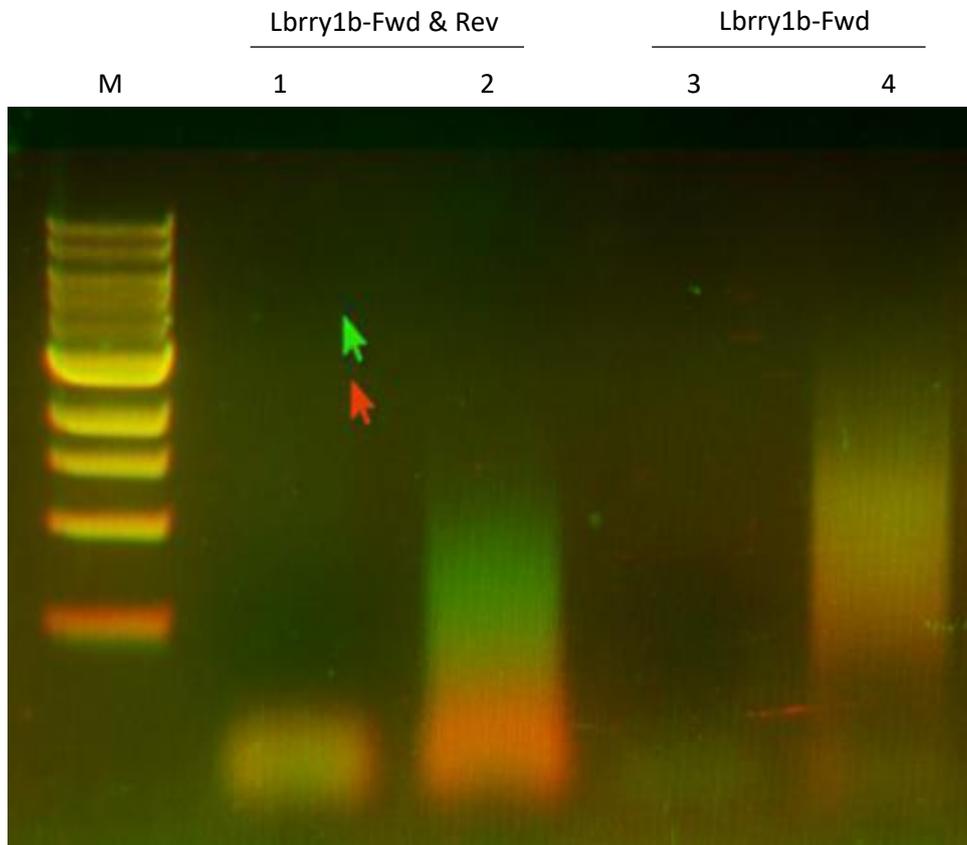
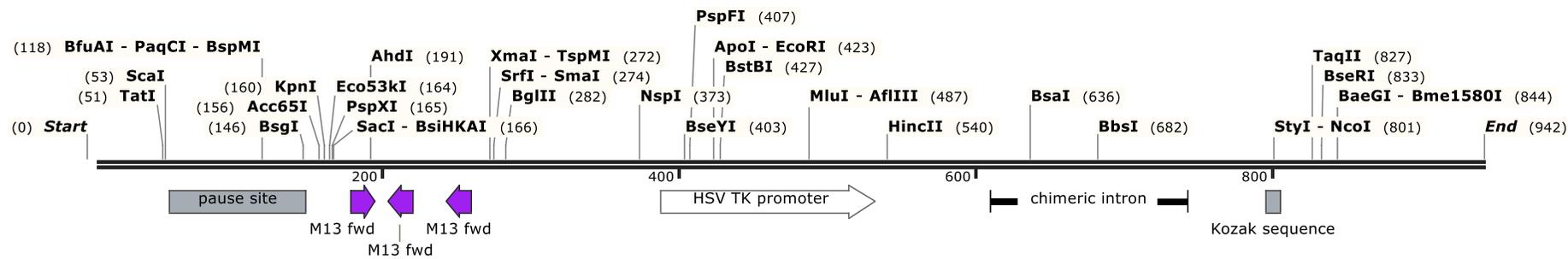


Figure 6.2.1.7: Assay for elimination of primer dimer formed during Library 1 PCR tested using digestion method. Lane 1, negative control (no template). Lane 2, template present. Lane 3, negative control (no template). Lane 4, template present. Green = undigested. Red = digested. 1kb size marker included (M). 2 min extension time used.

The figure above (**Figure 6.2.1.7**) shows a method for removing concatemers post Library 1 PCR, but pre size selection and purification of the cloneable insert. The figure shows when both Forward and Reverse primers are present concatemers are formed, shown by the green smear. This green smear then collapses once subjected to BsmBI digestion, shown by the read smear, and representing the small fragments cloned previously. Whereas when one primer is present, the smear shows only a small drop in molecular weight going from undigested to digested. This supports the use of one primer over two, to eliminate primer artifacts.

While **Figure 6.2.1.7** shows limited change in the size of fragments produced pre and post digestion, when in the presence of the Forward primer alone, further experiments showed the importance of this crucial pre-cloning digestion step. When the BsmBI digest step is skipped in the Library construction pipeline, primer artifacts are again cloned into the reporter vector, confirmed by sequencing data (**Figure 6.2.1.8**). The figure below shows multiple M13 primer sites inserted upstream of the minimal promoter, where the potential enhancer sequence should have been cloned instead. Suggesting that this insert is an artifact comprised of 3 primers.



**Six3 Lbrry\_F\_Concatemers**  
942 bp

Figure 6.2.1.8: Primer artifacts cloned into reporter vector. Sequence data revealed multiple M13 primer sites cloned upstream of the minimal promoter.

Further sequencing data showed the true unpredictable nature of concatemer formation. Even with the use of a single primer it is still possible to generate concatemers (**Figure 6.2.1.9**), although the probability is reduced. **Figure 6.2.1.9 A** shows a smaller digested product of what was once presumably a larger concatemer, digested during the Golden Gate cloning protocol by BsmBI. To avoid cloning these short concatemer fragments, the PCR product can be pre-digested before gel extracting and purifying.

**Figure 6.2.1.9 B** shows random priming of a PCR product generated in a previous PCR cycle, with a BsmBI site between the two primers. It should be noted that the Lbrry1-Fwd primers annotated in the figure are not the full-length sequence, the full length sequence contains 8 x Ns. The left-most primer has randomly annealed to the reverse strand of a previous PCR product, coincidentally its annealing site is at the end of the product so overlaps the previously incorporated primer. This could have occurred during the first few cycles when the annealing temp was lower, but not during the first cycle as there wouldn't be a previous PCR product to anneal to. As this example included an internal BsmBI site, this concatemer formation can be avoided by pre-digesting the Lbrry PCR before gel extracting. The rest of the sequence (to the right of the concatemer) was thought to be a genuine inert but when BLAST searched it matches with *E. coli* chromosome.

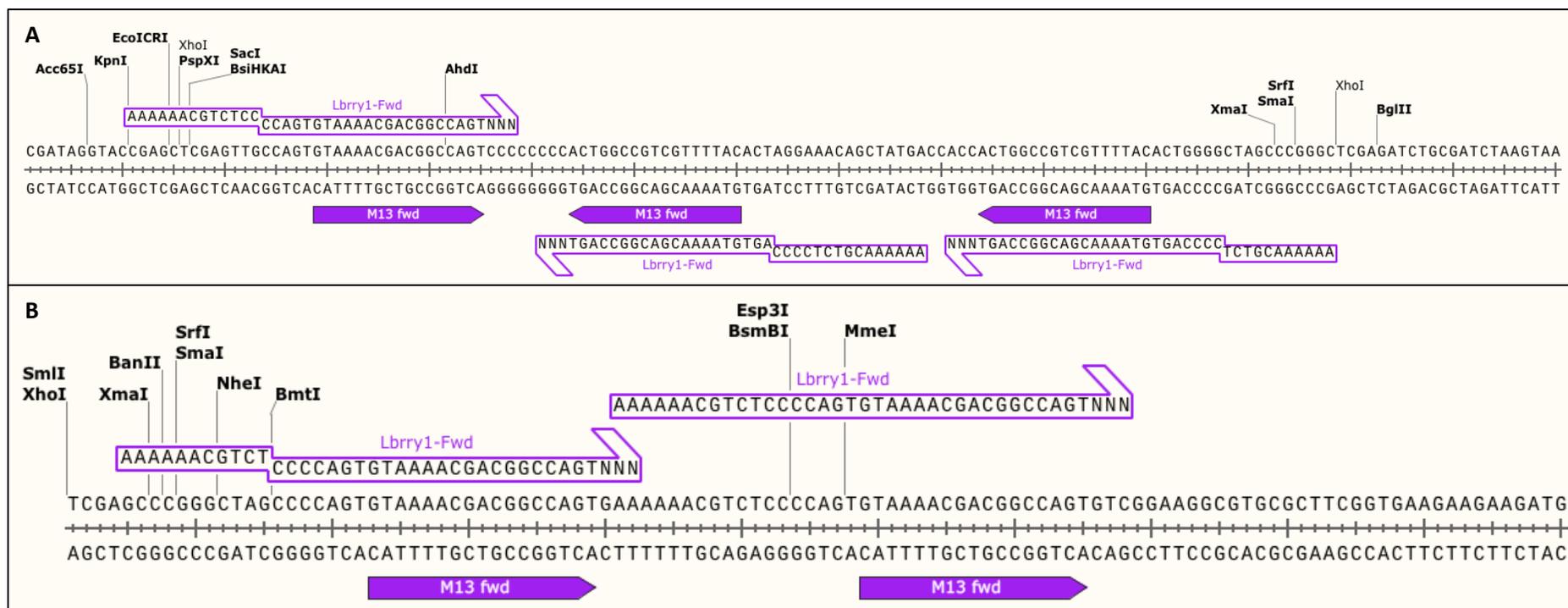


Figure 6.2.1.9: Primer artifact and concatemer formation during Library 1 PCR. Sequencing data showing examples of cloned concatemers. (A) Sequence data showing part of a larger concatemer that has been digested before cloning (hence no BsmBI sites present). Generated from PCR using only the forward primer. The first two primers on the left have been annealed together due to the 8 x N section being all Cs for one primer (top strand) and all Gs for the other. The third primer (on the right) is shown to be annealed to the 5' end of the second primer. (B) Sequence data showing random priming of previously generated PCR product. The left-most primer is shown annealed to the reverse strand of the previous PCR product, coincidentally its annealing site is at the end of the product so overlaps the previously incorporated primer. It should be noted that the Lbrry1-Fwd primers annotated on the figures are not the full-length primer sequences, actual primer sequence contains 8 x Ns.

### 6.2.2 Successful edit of pTK-Citrine Vector to include BsmBI adapter

The two BsmBI sites in pTK-Citrine yield different cohesive ends that enable directional cloning of inserts. Enhancers are non-directional, so directional cloning is not required but forward and reverse primers are required to add both adapter sequences to allow for cloning. For the Forward primer to be used solely in the Library 1 PCR, a second edit must be performed on the vector to then permit Golden Gate cloning. This edit successfully modified the downstream site to generate the same cohesive end as the upstream site. This was confirmed by the sequence data (**Figure 6.2.2.1**).

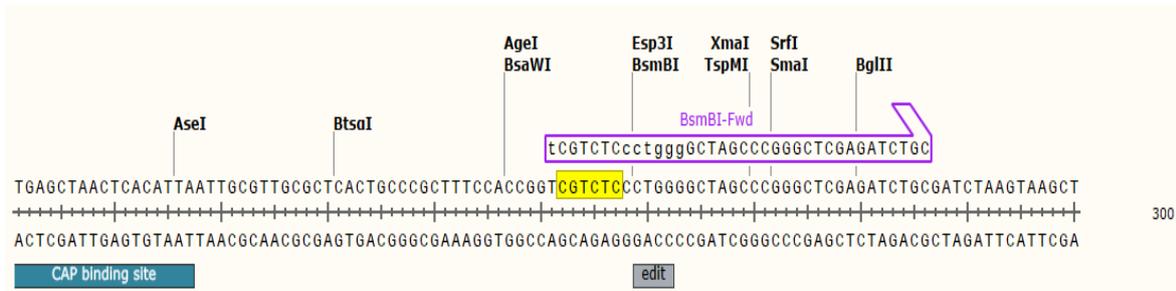


Figure 6.2.2.1 Sequencing data showing successful edit of BsmBI adapter added to vector. BsmBI recognition site highlighted in yellow. Edit labelled by grey box.

### 6.2.3 Optimisation of Golden Gate Cloning Protocol

The basis of this Golden Gate cloning is to allow for simultaneous and directional assembly of the insert (randomly amplified fragments) into a vector (pTK-citrine) using a Type IIS restriction enzyme, BsmBI. This creates a fluorescent reporter construct, whereby potential enhancer sequences, if present, can drive transcription from the minimal pTK promoter and translate the fluorescent protein. This protein can then be viewed under a fluorescent microscope, or its transcript can be detected using RT-PCR.

The cycling for the GG protocol works by having rounds of digestion and ligation, a plasmid safe reaction and bacterial transformation, followed by *LacZ* blue/white colour selection. Colonies that no longer contain the *LacZ* gene, and should contain our insert, will give white colonies. For the construction of a 3D genomic library to be possible, hundreds of colonies need to be generated at once. Each a molecular clone, with a different ~4kb fragment from the original BAC clone inserted. To achieve this many clones the standard 30 thermal cycles was increased to 60 cycles. To reduce the chance of the *LacZ* gene being re-ligated into the vector the cycling was finished on a 30-minute digestion step to decrease the amount of undesirable blue colonies produced. Both these changes have increased our colony count from dozens to hundreds. These changes have been successful in optimising the GG cloning protocol ready for Library construction.

The successful transformation of GG cloned reporter constructs gave hundreds of colonies. Colony PCR was performed, and selected clones were sent for sequencing. Two clones were successfully mapped to the human genome of chromosome 2 (**Figure 6.2.3.1**). These results showed we have successfully mapped clones corresponding to the original RP11-77121 BAC clone. However, as it can be seen in **Figure 6.2.3.1** (highlighted in red) the inserts cloned into the reporter plasmid are much shorter than the desired ~4kb fragments, both < 300 bp.

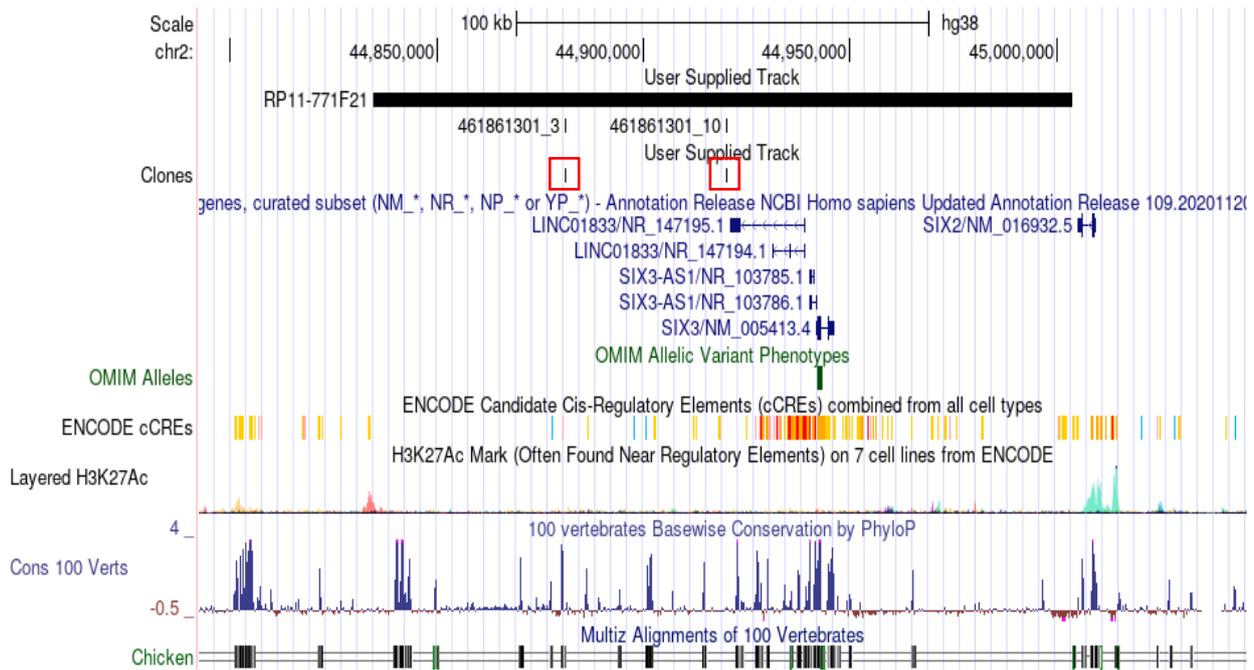


Figure 6.2.3.1: Clones for Six3 library mapped to human genome. Two clones (highlighted in red) mapped to the original RP11-771F21 BAC in the USCS genome browser Hg38 version.

Clones sequenced from all other Library construction attempts, confirmed host bacterial genomic DNA had again been cloned into the reporter constructs. We believe this repeat issue of cloning bacterial genome rather than BAC DNA is due to linear fragments of bacterial chromosome DNA being preferentially amplified over intact supercoiled BAC DNA. As supercoiled BAC DNA is used as template for the Library 1 PCR, the random amplification of this template is failing. The changes made to the PCR settings, (low annealing temp, 40 cycles, increased  $MgCl_2$ ,...etc), encourages high rates of amplification, allowing trace amounts of bacterial genomic DNA to be amplified and mistaken for template amplification.

Other parameters of the Library 1 PCR that were thought to be optimised throughout this report include: the extension time for the amplification of the fragments, the amount of initial cycles needed to allow for amplification of larger fragments, and the pre-extension ramp rate present in the first 5 steps to allow for primer annealing. An 8-minute extension time was deemed optimal to yield higher molecular weight fragments, with shorter extension times of 4 and 6 minutes failing to generate a smear at all. However, upon sequencing data confirming the presence of host *E. coli* DNA rather than the BAC DNA, the template preparation and PCR parameters need further optimisation to ensure amplification of the BAC clone over the host bacterial DNA.

### 6.3 Discussion

Multiple parameters were optimised for our GG cloning approach to efficiently generate a shotgun reporter library for enhancer activity. Among the successes was the preparation of the vector for cloning. Further optimisation is required for preparation of the insert.

The pTK-Citrine vector allows for fast throughput testing of cloning efficiency using a selectable antibiotic marker, ampicillin, and the ability to correctly identify colonies; containing the cloned insert; through blue/white colour selection of colonies grown on LB/Carb/X-gal plates. Additionally, the barcode modifications made to this vector, **Chapter 4**, allow for RT-PCR detection of positive reporter constructs that are co-electroporated with (and thus titrated by) a larger number of negative reporter constructs.

The sequencing data in **Chapter 6.2.1** showed that concatemers and primer artifacts formed during the shotgun PCR were subsequently being cloned into the reporter plasmid (**Figure 6.2.1.9**). This highlights the importance of the BsmBI pre-digest of the insert to ensure the presence of these artifacts are minimised prior to size-selection by agarose gel electrophoresis.

One of the limitations of all Golden Gate assemblies is the inverse proportionality between the number of inserts (complexity of assembly) and the number of colonies produced during transformation (number of transformants). However, this is less of an issue with our approach due to it being single insert cloning strategy. Further developments were made to overcome the limitations of Golden Gate assembly, outlined by Kucera & Cantor, to achieve maximal transformation levels in our approach (Kucera & Cantor, 2018). The efficiency levels were pushed even higher by increasing the standard 30 cycles to 60, generating hundreds of white colonies, with low blue colony generation, required for successful library construction.

One of the main limitations within our GG cloning approach was the presence of host bacterial DNA cloned into the reporter vector, even when multiple rounds of purification were employed to combat this. This issue was traced back to the preparation of the template used for the initial shotgun PCR. Multiple factors were optimised for the shotgun PCR, for the random amplification of fragments in the production of the shotgun library, outlined in **Chapter 6**. Starting with identifying the optimal concentration for the DNA template at <10 ng for a 50 µL reaction for optimal amplification using Phusion DNA polymerase.

In comparison to the barcoded reporter vectors created by Streit & Chen our approach modifies vectors that are compatible with high-efficiency Golden Gate cloning, rather than blunt-ended or TA cloning. This is key to streamlining the construction of shotgun libraries. We succeeded in modifying the pTK-Citrine vector to include the BsmBI adapter sequences as well as the unique barcodes. As such Golden Gate cloning strategy was employed to try and generate a library for the *Six3* gene however, this was unsuccessful.

## Chapter 7: Candidate approach

### 7.1 Introduction

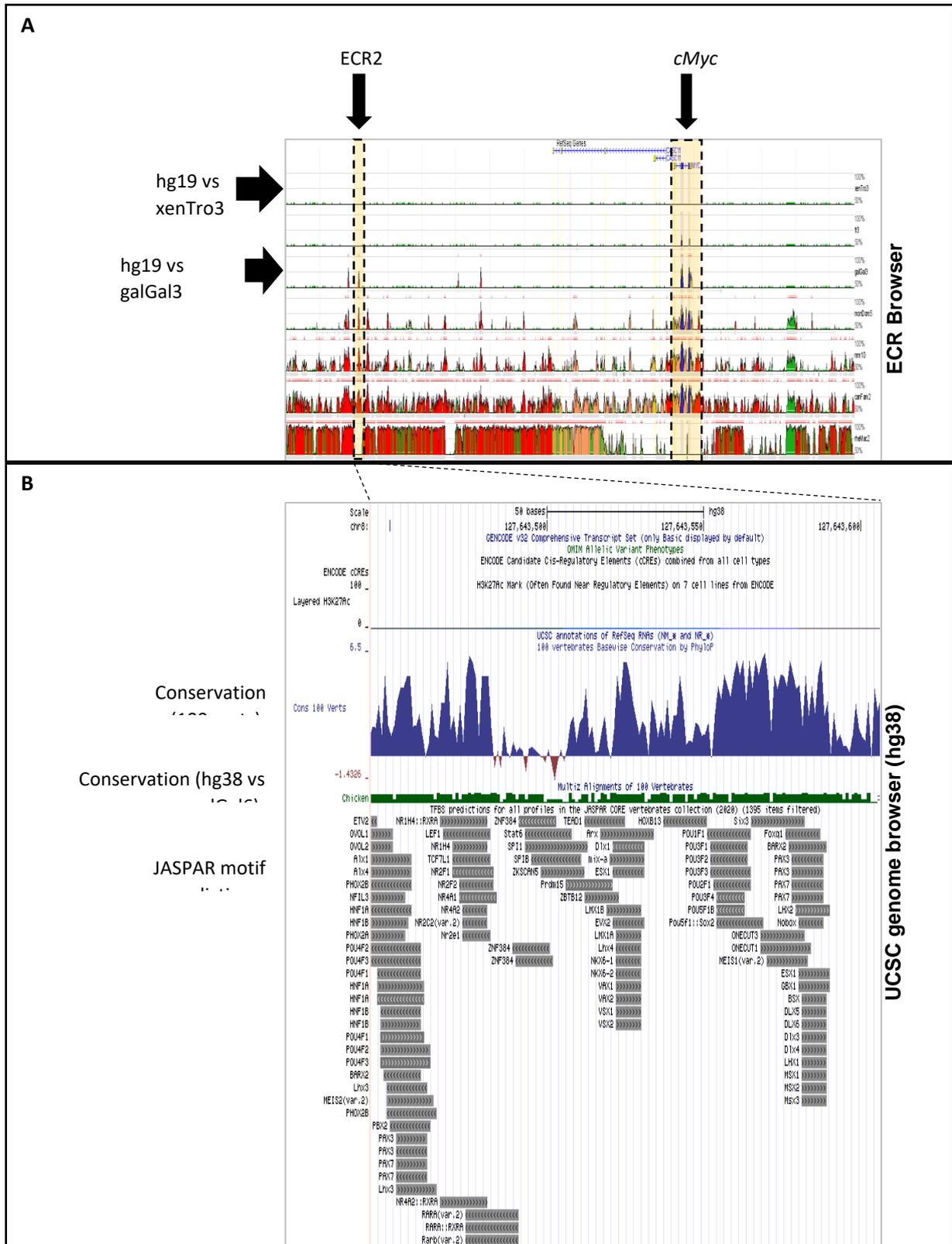
A candidate approach focuses on a specific (candidate) gene and locating potential enhancers for that gene, rather than searching the genome for potential enhancer sequences and then identifying their cis-regulated gene at a later stage. Candidate approaches can take the form of many different epigenetic tools, but our approach focused on using phylogenetics, the conservation of non-coding sequences and predicted TF binding sites for these evolutionary conserved regions (ECRs).

Our approach focused on using a candidate approach for identifying potential enhancers that drive the *cMyc* gene. As mentioned earlier in **Chapter 1.2.5**, *c-Myc* is a proto-oncogene that has a role in regulating cell proliferation, cycle progression, growth and survival (Eilers & Eisenman, 2008). *In situ* data from Grocott et al, shows *c-Myc* expression in the neural ectodermal tissue of the anterior neural folds (**Figure 1.2.5.2**) (Grocott et al, unpublished data). This is of interest to us due to the expression of the HPE causative genes also being present in these tissues at the same developmental timeframe.

### 7.2 Results

Using a candidate approach, an ECR of the genome was investigated to contain a putative enhancer for the *cMyc* gene. While *cMyc* is not considered one of the causative HPE genes, it has a similar expression pattern concentrated in the anterior neural fold region (**Figure 1.2.5.2**).

To identify ECRs the human (Hg19) and chick (galGal3) genomes were compared using the ECR browser. Sequence conservation was analysed, focusing on a region encompassing the *cMyc* gene (hg 19 chr8:128634272-128801494). Further analysis for each ECR was carried out using the JASPAR tool in the UCSC genome browser, showing the TF binding site predictions. ECR2 (hg19 chr8:128655690-128655851) (**Figure 7.2.1**) was chosen for further evaluation based on its sequence conservation, and interesting predicted TF binding sites such as *Six3*, which is a key causative gene in the Holoprosencephaly network. Other interesting predicted TFs, that were also well conserved, included: *Lhx2*, *Nkx6.2*, *Vax1*, *Vax2*, *Vsx1*, *Vsx2*. All these genes are either involved in brain or eye development.



**Figure 7.2.1: Putative cMyc enhancer predicted using ECR conservation and TF binding predictions. (A)** Shows sequence conservation of the hg19 genome compared to other species genomes. ECRs potentially containing enhancers shown by red, pink (ECR inside introns) and yellow (untranslated region parts of exons) peaks in the browser. The cMyc protein (blue regions) is encoded by the end of exon 1, the width of exon 2 and the beginning of exon 3. The cMyc gene is highlighted by dashed box and pink background. Conservation between the hg19 genome and xenTro3 and galGal3 are highlighted by bold black arrows on the left. Little sequence conservation can be seen between the Human and XenTrop genomes, more conservation is seen between Human and Chick genomes. The ECR thought to contain the putative enhancer highlighted by dashed box and yellow background and labelled ECR2 with black arrow. **(B)** Shows further analysis of chosen ECR. Predicated TF binding motifs listed using JASPAR tool in UCSC genome browser. Tracks for conservation between the hg38 and galGal6 genomes shown by blue peaks.

Primers were designed (**Table 2.6**) to amplify this chosen region of the genome. Human genomic DNA (200ng) was subjected to 35 thermal cycles of the ECR amplification Phusion reaction conditions. The PCR product generated was the correct size for the amplified ECR at 1333 bp (**Figure 7.2.2**). This length of sequence was chosen to be amplified to ensure important flanking sequences were not missed.

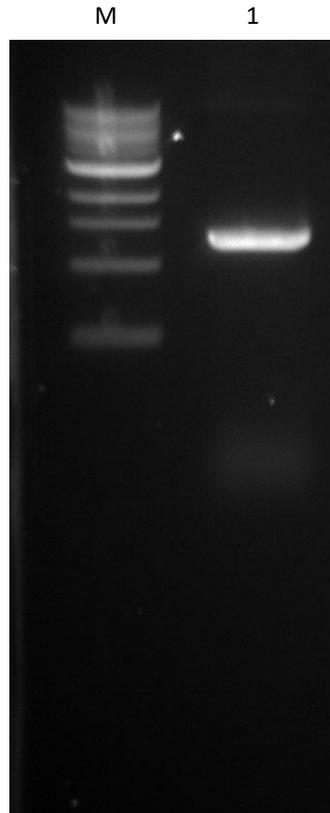


Figure 7.2.2: Amplification of candidate cMyc enhancer ECR performed by RT-PCR. Lane 1, amplified ECR, PCR product of 1333 bp generated. 1kb size marker included (M).

This product was gel extracted, purified and used as the insert for standard Golden Gate cloning into a non-modified ptk-Citrine vector (**Figure 4.1**). To validate whether the cloning and subsequent transformation was successful, colony PCR was performed (**Figure 7.2.3**) and sequencing data was collected for the chosen clones (**Figure 7.2.3, lanes 2-4**). Both the colony PCR and sequencing data for these chosen clones, verified the reporter vector included the correct amplified ECR insert.

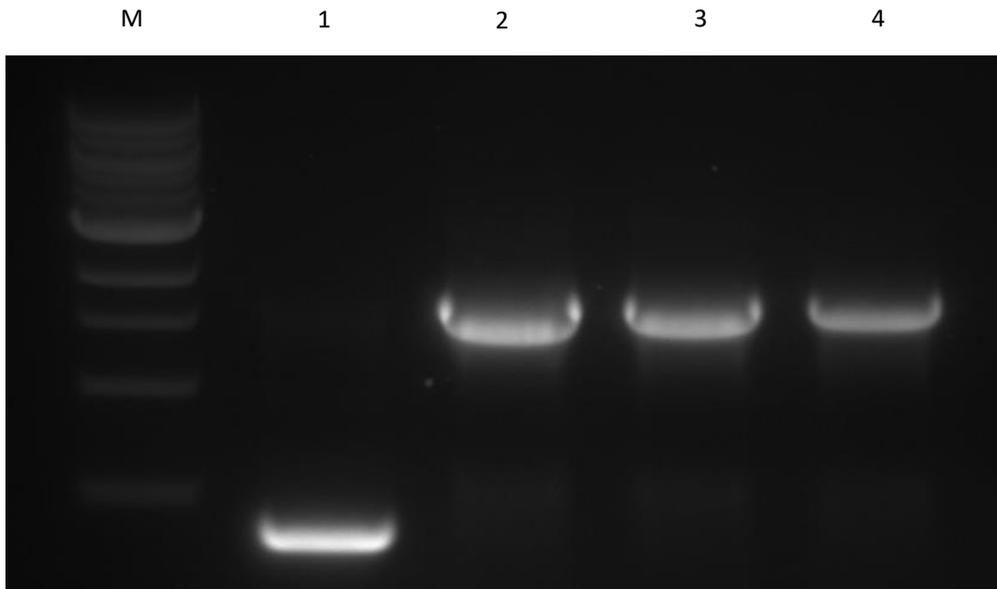


Figure 7.2.3: Colony PCR results for GG cloned candidate cMyc enhancer. Lane 1, reporter plasmid of 373bp; lanes 2-4, ECR insert of 1333 bp. 1kb size marker included (M).

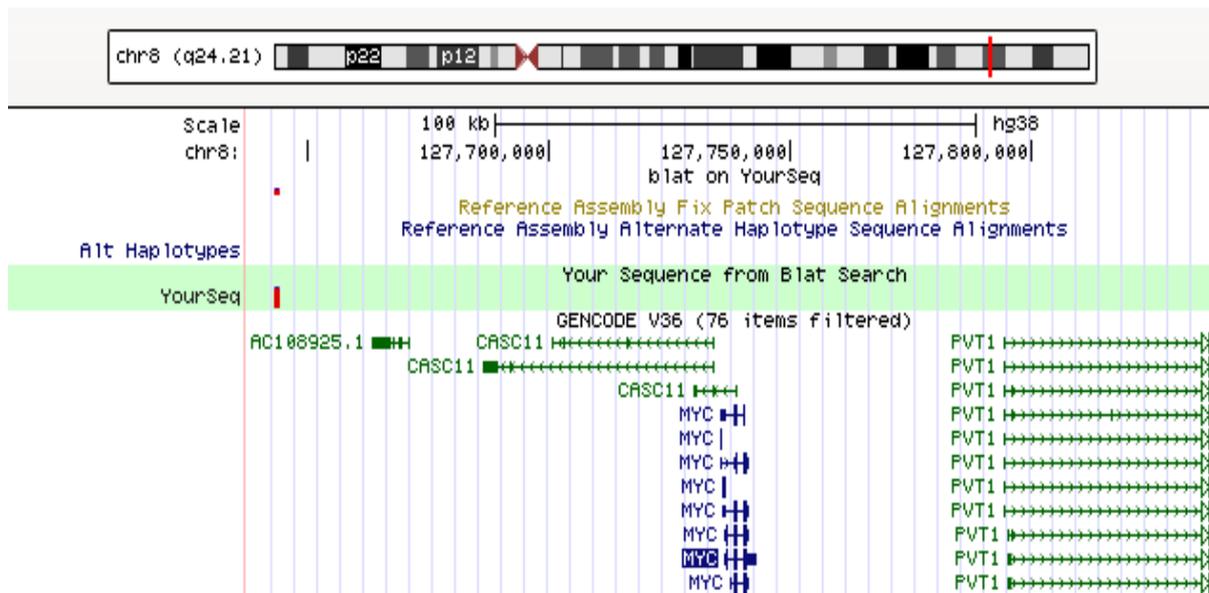
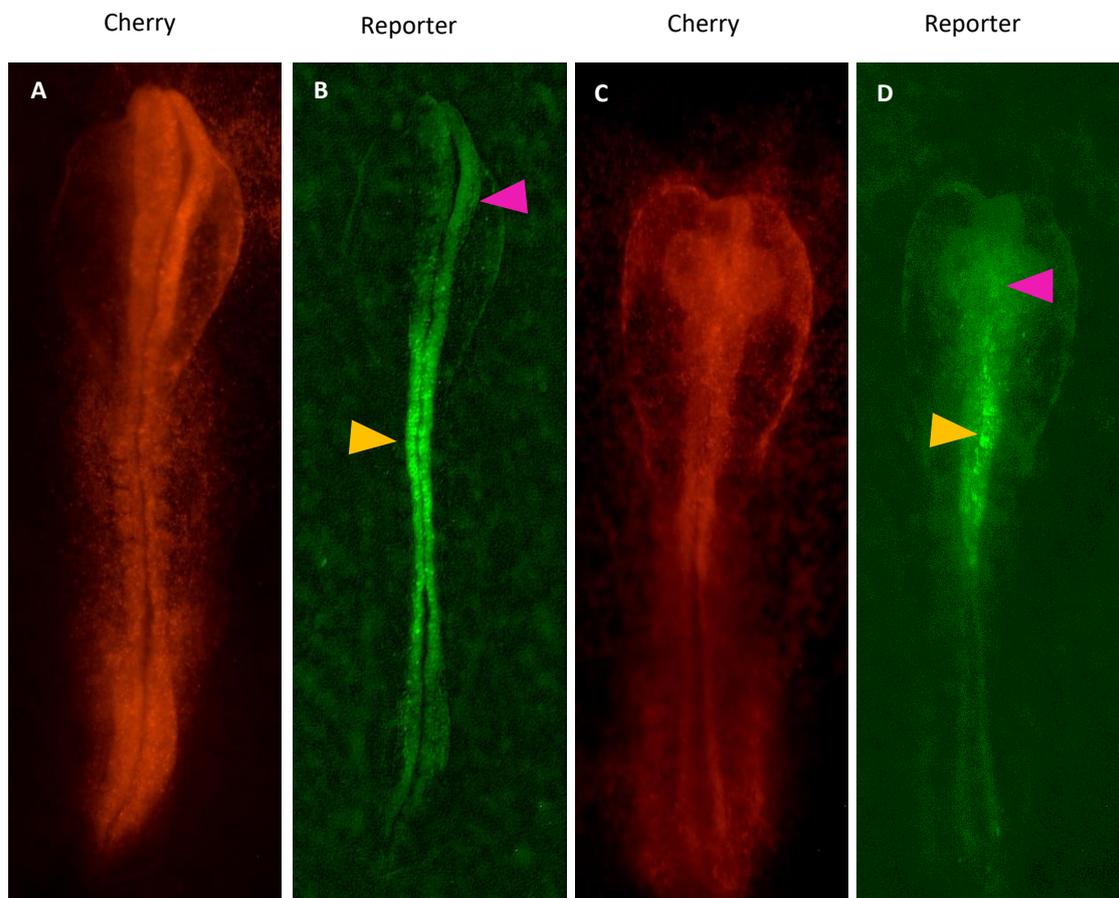


Figure 7.2.4: Sequencing data confirms cloned candidate cMyc enhancer maps to the hg38 genome. Putative enhancer mapped to the human genome, shown as 'YourSeq' and highlighted by green bar. *Myc* gene (blue) can be seen on chr8 in relation to the potential enhancer sequence.

The reporter construct was then electroporated into stage HH3 chick embryos, alongside positive control plasmid, Cherry. Cherry expression is driven by the  $\beta$ -actin promoter, active in most cells, and therefore a good choice for a positive control. Injection of the constructs was targeted to the tissue surrounding the primitive streak.

Of the five embryos electroporated, the cMyc candidate enhancer showed reporter activity was seen in two of the embryos (Figure 7.2.5). Strong Cherry expression can be seen uniformly throughout both the embryos. Reporter expression was seen along the neural crest in both stage HH8 and HH9 embryos (Figure 7.2.5 B, D). Faint expression can also be seen in the mid/hindbrain region of the stage HH9 embryo (Figure 7.2.5 D).



**Figure 7.2.5: Electroporations of cMyc enhancer reporter construct.** (A) Cherry expression is seen throughout the stage HH8 embryo. (B) Fluorescent reporter activity seen along the neural crest (yellow arrow) of the stage HH8 embryo. Faint expression seen along one side neural folds (pink arrow). (C) Cherry is expressed throughout the stage HH9 embryo. (D) Reporter expression again visualised along the neural crest (yellow arrow) and in the mid/hindbrain region (pink arrow) of the stage HH9 embryo. Embryos imaged dorsal side up.

The reporter activity seen did not fit within our predicted regions. Our preliminary *in situ* data (**Figure 1.2.5.2**) supported the presence of this putative growth zone located in the anterior neural fold. As such, any potential enhancer found for cMyc was expected to have an overlapping expression pattern. However, it has been seen before that enhancers taken out of context can have a broader expression pattern compared to their more restricted target genes (Mok et al, 2021). Another reason for the lack of expression seen in the anterior neural folds, could be due to the mass proliferation occurring in this region, diluting down the fluorescent reporter plasmid. Autofluorescence of the forming neural tube also occurs from HH8 onwards, which can mask any green fluorescence in this region. As such a far-red fluorescent protein, such as mKate2, can be used instead.

The cMyc gene itself is also known as a critical factor for cell cycle progression in the premigratory neural crest cells (Kerosuo & Bronner et al, 2016; Bellmeyer et al, 2003) so expression in the neural crest is not surprising.

Further electroporations were carried out to target to this putative growth zone. Time-lapse movies were generated starting at stage HH4 embryos developing until HH8. The movies are shown in a series of still images below (**Figure 7.2.6**). Weak Cherry expression is seen in the epiblast along one side of the embryo (**Figure 7.2.6 Aii**), fluorescence grows brighter as more Cherry is expressed in  $\beta$ -actin-containing cells (**Figure 7.2.6 Bii-Gii**) from the constitutive promoter. By stage HH8 strong

Cherry fluorescence is seen in the anterior neural folds and head process. This indicated that targeting of these cells were successful but the presence of autofluorescence should also be considered. However, while reporter fluorescence is seen in the epiblast cells as they ingress into the primitive streak and undergo EMT (**Figure 7.2.6 Ai-Di**), no *cMyc* enhancer expression can be seen extending anteriorly as Cherry does. A few cells moving towards the midline can also be seen showing reporter expression (**Figure 7.2.6 Eii-Gii**), suggesting the reporter gene may be active in the neural fold, however this mid/hindbrain region is still more posterior than expression patterns generated from in situs show.

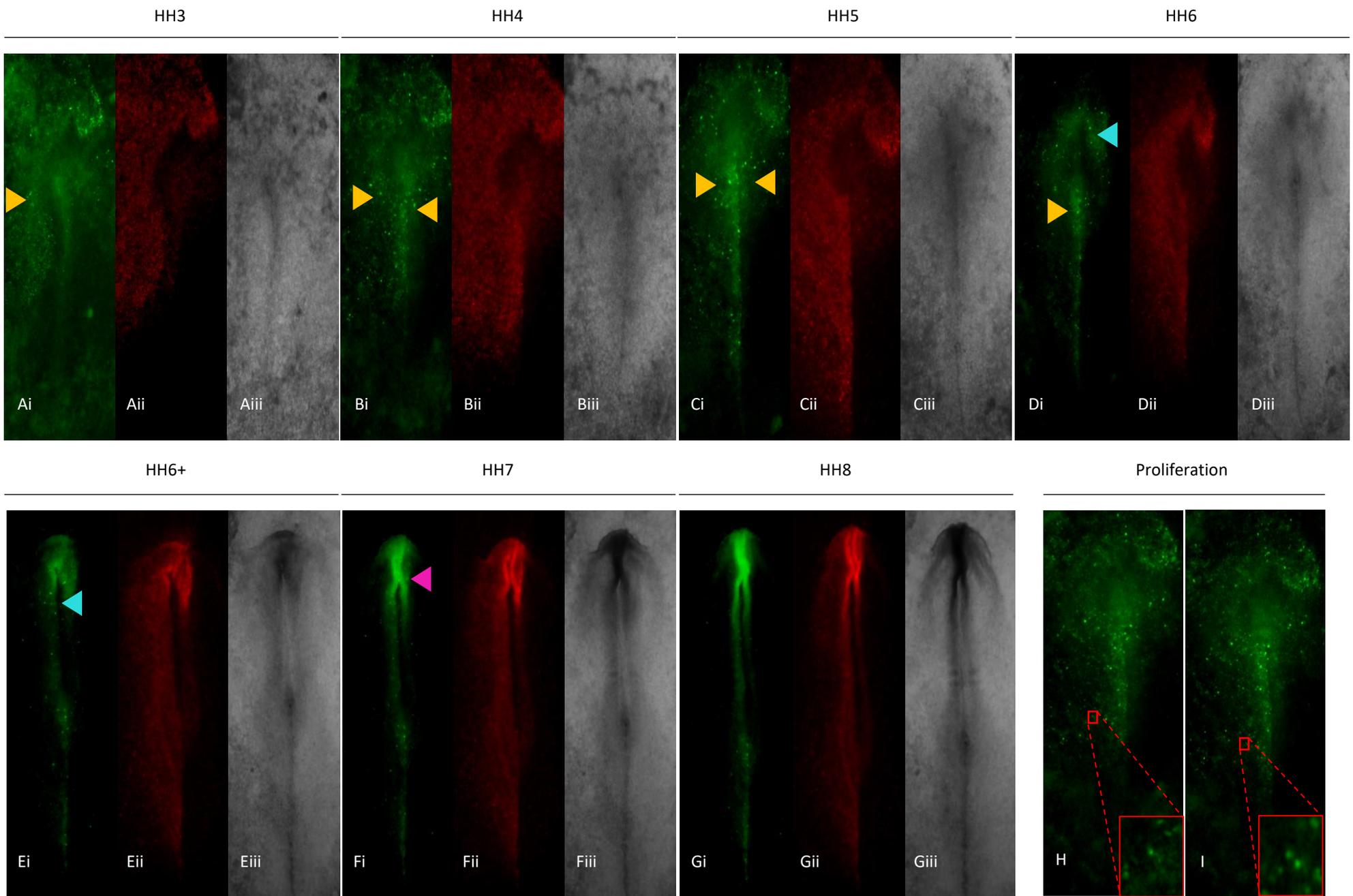


Figure 7.2.6: Still frames from time-lapse movies showing expression of fluorescent cMyc enhancer construct. Embryos were imaged overnight (22hrs) until stage HH8 embryos were achieved. (i) Green fluorescent field showing Citrine expressing cells. (ii) Red fluorescent field showing Cherry control. (iii) Brightfield images. (A-C) Embryo displaying fluorescent reporter activity in the epiblast and presumptive mesoderm in cells ingressing towards the primitive streak during gastrulation. (D-G) Embryo elongating. A few *cMyc*s show reporter fluorescence anterior to the primitive streak as the head process develops. Specific fluorescence fades as neural tube folds into neural folds. (H-I) Zoomed in image of fluorescent epiblast cell diving into two in stage HH4 embryo. Specific fluorescence seen by 'bright spots'. Autofluorescence seen as homogenous glow. Yellow arrows label epiblast. Blue arrows label neural crest. Pink arrows label neural folds.

### 7.3 Discussion

Using a candidate approach outlined in **Figure 2.1** our lab was able to identify a 1333 bp sequence Evolutionary Conserved Region (ECR) within the human genome (hg38) located upstream of the *cMyc* gene on chromosome 8. This region is conserved in chick (galGal3), dog (canFam2), mouse (mm10) and opossum (monDam5), (**Figure 7.2.1**). When the reporter construct containing the putative enhancer sequence was electroporated into stage HH4 chick embryos, strong fluorescence could be seen in the neural crest region of stage HH8 and HH9 embryos (**Figure 7.2.5**). These data when combined with the time-lapse images, (**Figure 7.2.6**), which displayed fluorescent reporter activity in the epiblast and presumptive mesoderm cells, suggests the CRE is active in pluripotent cell populations. *cMyc* is a known early marker for pre-migratory neural crest cells (Kerosuo & Bronner, 2016), and is also a Yamanaka factor. The Yamanaka factors, made up of 4 TFs (including *cMyc*), were discovered to critically regulate the developmental signalling pathways in embryonic stem cells to maintain their pluripotent states (Takahashi & Yamanaka, 2006). It was shown these factors regulate the pluripotency and differentiation of embryonic stem cells through multiple developmental signalling pathways (Liu et al, 2008; Prakash & Wurst, 2007; Doroquez & Rebay, 2006), but the complex interplay and crosstalk between these networks means the signalling network responsible for maintaining pluripotency and differentiation in this stem cell population remains unknown.

The time-lapse still images showed a few fluorescent cells migrating towards the midline and then disappearing in a region which corresponds to the anterior neural folds (**Figure 7.2.6 D-G**). This suggests that the enhancer sequence, driving reporter gene expression, may be active in the neural fold. The putative *cMyc* growth zone, located in the anterior neural folds, as shown in our preliminary *in situ* data (**Figures 1.2.5.2**), was a hypothesized area of activity for our CRE. However, the cell population within this region undergoes massive proliferation and any subsequent targeting of this cell population is an issue. Dilution of the fluorescent reporter will occur and therefore the fluorescent cells will no longer be visible by microscopy detection methods. While targeting of this anterior neural fold area was an ongoing issue for this assay, this limitation may apply when performing a functional assay for enhancers associated with other genes expressed within this region e.g., the HPE causative genes.

## Chapter 8: General Discussion

The aim of this project was to generate a high-throughput functional assay to identifying enhancers associated with the HPE gene network. This report highlights the advantages and the limitations associated with both candidate and unbiased approaches for enhancer screens. Multiple reporter plasmids were successfully modified to contain 5 unique barcode edits, required for rapid detection. The Golden Gate cloning method was improved to generate hundreds of colonies required for library construction. However, further research is required into the use of BAC clones to construct 3D shotgun reporter library, with the end goal of multiple libraries constructed for disease genes and efficiently screened for enhancers. Using a candidate approach a putative enhancer was found for the *cMyc* gene, shown to drive expression along the neural crest. Further characterisation of this *cMyc* enhancer is needed to understand its functional role within the gene network.

### 8.1 Construction of a *Six3* shotgun reporter library using BAC clones

We successfully identified a single BAC clone covering 169 kb of human chromosome 2, which includes the *Six3* locus. The BAC system, in comparison to its YAC counterpart, is highly stable, capable of cloning large inserts and useful to construct a total genomic library with high stability (Asakawa et al, 1997). This allows for the construction of a high-resolution physical map for a portion of the genome. Historically BAC libraries are made up of 10s or 100s of BACs, mapping large regions of the genome, even whole chromosomes, in high resolution. However, instead of BAC reporter construct containing the regulatory region of the gene of interest our constructs will contain fragments of the BAC, with a coverage of 2-3x achieved.

The major advantage of a BAC sub-fragment plasmid library is the efficiency in which a large region of the genome is screened for enhancer activity. With genome wide approaches, such as ATAC-seq, taking years of labour-intensive analysis to screen a similar size of the genome.

As such a reliable way to identify regulatory elements may be to systematically screen genomic regions and then identify important foci through sequence conservation. With analogous genome-wide and sequencing approaches taken gain a greater understanding of how the gene is regulated.

However, while in the 90's BACs were a celebrated tool for high resolution genome mapping, this resource is being stripped away. One of the only UK suppliers, Source Bioscience, will no longer be providing BACs for order. As such the means for BAC library construction and these low-cost high throughput assays will be that much harder to perform.

A common dominator linking both cloning approaches was the preparation of the BAC clone. A crucial step in the construction of any BAC library is the isolation of BAC DNA from the host *E. coli* DNA. Different strategies were employed throughout our cloning processes to achieve the highest possible yield of BAC DNA. With the Golden Gate cloning strategy agarose gel electrophoresis and extraction was employed. However, while gel electrophoresis may be the standard for purifying DNA products, much of the DNA is lost in the process, resulting in modest yields. This was a major pitfall in the preparation of the BAC clone for fragmentation, as a large volume of BAC was required as starting material. Hence using random primed PCR to produce fragments for library construction is preferable when purifying the BAC using gel electrophoresis and extraction methods. As such, the TA cloning strategy adopted a Plasmid Safe DNase digest approach to break down linear DNA (host bacterium DNA), highlighted in **Chapter 5.2**. Alternative approaches to gel electrophoresis, such as DNase digest have been employed (Heavens et al, 2016), but none are widely used.

However, while DNase digest is a possible approach for the initial purification of the BAC clone, the later purification step required for the fragments inserted in the reporter vector, must also act as a size selection step. As such we are once again relying on gel electrophoresis and extraction methods to yield cloneable material. Alternative methods, such as using AMPure XP beads, requires further investigation into their sensitivity.

## 8.2 Evaluation of Reporter Library Construction via Golden Gate Assembly

A 2001 paper detailing the limitations that occur in the production of shotgun libraries when using random amplification led us to increase the concentration of the cofactor  $MgCl_2$  (Rohwer et al, 2001). The results generated (**Figure 6.2.1.4**) suggested that the standard reaction conditions being used were not optimal for achieving the maximum amounts of amplification product, as such any later shotgun PCRs performed included a  $MgCl_2$  concentration of 4.5 mM.

Under the direction of this 2001 paper (Rowher et al, 2001), primer concentration was then tested as a limiting factor, however the Library 1 PCR failed to generate any amplification when deviating from the original primer concentration (**Figure 6.2.1.5**), presumably because the primers were binding  $Mg^{2+}$ .

What is arguably one of the most important factors that can limit the success of a PCR is the DNA template itself. As such use of high-quality, purified BAC DNA stored in elution buffer, rather than ultra-pure water, was one of the most important optimisations of this approach. As mentioned in **Chapter 6**, the inconsistent amplification of higher molecular weight fragments was a frustrating obstacle of the shotgun PCR that eluded us. Upon the discovery that our BAC template seemed highly susceptible to pH degradation, the presence of Tris-Cl stabilised the DNA for longer periods and consistent amplification of higher molecular weight products were finally seen (**Figure 6.2.1.6**).

However, upon receiving sequencing data and finding the presence of host bacterium genomic DNA in our constructed library, all these optimisations are cast in a shadow. Initially our suspicions were concentrated on the initial purification of the BAC template. Since then, it has focused on the PCR preferentially amplifying relaxed linear DNA (i.e. fragmented bacterial chromosome) more efficiently than supercoiled DNA (Chen et al, 2007; Datta et al, 2016; Laghi et al, 2004). PCR is a powerful tool within research due to its high sensitivity and specificity. However, this can act as a double-edged sword, as mentioned earlier the sensitivity of PCR is strongly influenced by topological characteristics of the template. With this in mind it can be inferred that the majority, if not all, of the amplification seen in the Library 1 PCRs were of linear *E. coli* genomic DNA, rather than the supercoiled human genomic DNA of the BAC clone. Due to the changes in PCR conditions, (e.g., low annealing temp, increased  $MgCl_2$  concentration, 40 total cycles), this made the PCR even more sensitive and amplified the trace amounts of host bacterium DNA left over after the initial purification step. As such any of the optimisations recorded during **Chapter 6** will need to be adjusted accordingly to facilitate PCR of supercoiled templates.

As such it is encouraged to perform colony PCR (and/or sequencing) on all clones making up the library. This is to gauge whether the correct sized insert has been cloned into the reporter vector. Once this approach is optimised the majority of clones should contain size selected inserts of human genomic DNA sub-fragments from the original BAC clone. If one is confident the correct insert has been cloned into the reporter vector this step is optional.

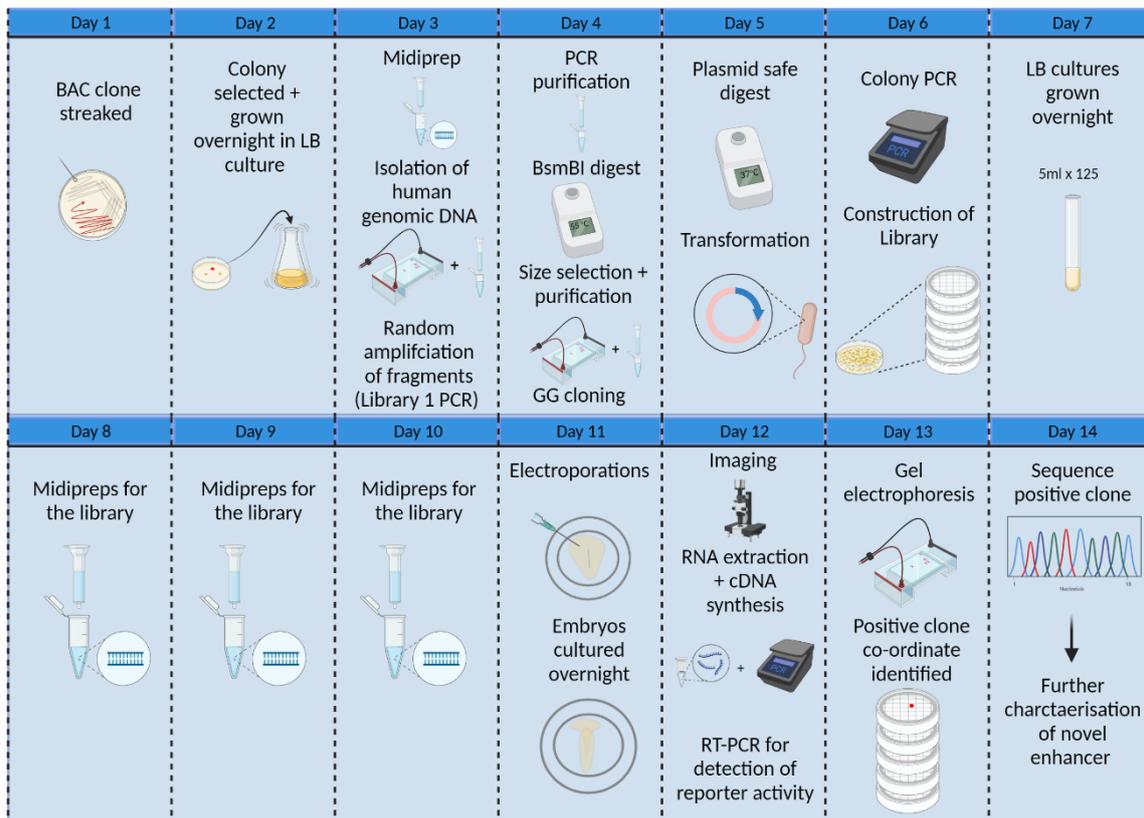
Datta et al, demonstrated an inexpensive and easy approach to linearize supercoiled DNA, with the use of restriction endonucleases (RE) in a single tube-based PCR method (RE-PCR) (Datta et al, 2016). There are two types of REs: type I and type II. Type I REs cleave the phosphodiester bonds of double stranded DNA, but not at site-specific sequences. Type II REs cleave DNA at highly specific

recognition sites. As such we are interested in the properties of type II REs and their ability in relaxing supercoiled DNA. The RE used in the Datta et al paper, EcoRI, has a 6bp recognition site of 5'-G/AATTC-3' sequence. It cuts the DNA after G, with '/' denoting where the bond is broken. EcoRI also forms sticky ends with AATT nucleotides. However, EcoRI is not a suitable RE to relax the *Six3* BAC clone RP11-771F21, as there are over 40 EcoRI sites spanning across this 16.9 kb region of the human genome. This would shred the DNA into smaller than desired fragments and is therefore not a possible candidate.

Alternatively, a more specific RE, such as NotI may be more suitable to digest the RP11-771F21 BAC clone. NotI has an 8bp recognition site of 5'-GC/GGCCGC-3'. There are only 4 NotI recognition sites along this sequence. However, they are clustered together with one site only 343bp downstream of another site, which when cleaved would create a very short fragment. The RE, MluI, should also be considered for cleavage of the RP11-771F21 BAC clone as there are only 2 restriction sites of 5'-A/CGCGT-3' equally distanced, along the 169kb region.

Other possible approaches for relaxing the supercoiled DNA is the use of engineered 'nicking' endonucleases that hydrolyses only one strand of double stranded DNA, to produce DNA that is 'nicked' rather than cleaved (Ivenso & Lillian, 2016). Or the addition of low concentration DMSO to create loose DNA regions within the negatively supercoiled DNA molecule (Lv et al, 2015). Once the BAC DNA has been relaxed it should facilitate PCR, rather than inhibit it like its supercoiled counterpart. However, it is advised to re-optimize the parameters of the shotgun PCR, different from the ones outlined in this report, so the PCR will be less sensitive and specifically amplify the desired template DNA over the trace amount of bacterial chromosome left behind post-purification. But it should be noted that once the BAC DNA is linearised this should reduce the chance of the shotgun PCR non-specifically amplifying the bacterial genome, as the more abundant, amplifiable BAC DNA should out-compete the contaminating bacterial chromosome.

Upon successful construction of the *Six3* Library, this assay would be a systematic approach for identifying functional enhancers at the *Six3* locus. Evenly scanning the entire *Six3* locus using a BAC subfragment plasmid library constructed in a pTK-Citrine reporter vector and taking advantage of the electroporation of chicken embryos using the same plasmids. The figure below, (**Figure 8.2.1**), highlights the efficiency in which a large portion of the genome can be expediently screened and tested for functional activity using our GG cloning strategy.



**Figure 8.2.1: Timeline for screening 150+ kb region of the genome for enhancer activity using GG cloning approach.** Day 1: BAC clone is streaked on chloramphenicol plates and incubated overnight. Day 2: Colony selected and grown in low sodium LB culture ~24hrs. Day 3: BAC DNA is extracted using low-plasmid copy midi-prep protocol. BAC DNA is isolated from any host *E. coli* genomic DNA via gel electrophoresis, extraction and purification methods. Purified BAC DNA is used as a template for shotgun PCR to generate randomly amplified fragments. Day 4: shotgun PCR is purified on column. Digested by BsmBI enzyme to eliminate concatemers formed. Randomly amplified fragments undergo size selection and purification by gel electrophoresis, extraction and purification methods. Modified GG cloning protocol (60 cycles) is set up overnight to insert size selected shotgun fragments into barcoded pTK-Citrine vectors. Day 5: Cloning reactions undergo Plasmid-Safe digest to remove linear DNA. Transformation of cloning reaction into competent DH5- $\alpha$  *E. coli* cells. Day 6: Colony PCR performed on clones to confirm the correct insert cloned into reporter vector. Construction of 3D reporter library. Day 7: Each colony in the library is selected and grown in 5ml LB cultures (~16hrs). Day 8-10: Midi preps performed to extract 2D groups of reporter constructs ready for functional screen. Day 11: Utilising group testing strategy, grouped reporter constructs injected into HH4 embryos, cultured until HH8. Day 12: HH8 embryos imaged for fluorescent reporter activity. RNA extraction, cDNA synthesis and RT-PCR performed on HH8 embryos to detect unique barcode sequence only present when reporter activity is present. Day 13-14: Identification of positive clone in library that contains CRE.

Studies have raised warnings on relying on sequence conservation as enhancers exist that may be unique to a branch of phylogenetic development. This approach of constructing an unbiased shotgun library of a BAC insert into a reporter vector and then to perform a functional screen in a model organism is highly efficient, as shown by Mastumata et al, Inoue et al and others (Matsumata et al, 2005; Inoue et al, 2008; Okamoto et al, 2015). Once optimised this approach has the application to rapidly identify regulatory elements for other disease pathways as a parallel functional approach to whole genome wide techniques.

### 8.3 Alternative strategies for constructing a shotgun reporter library

An alternative approach to constructing a 3D genomic library would be to exploit the Nextera library construction kit (Illumina), a well-known next-generation sequencing library prep. The preparation of DNA for high-throughput sequencing can be broken down into 3 core steps: (i) fragmentation of target sequences to desired length. (ii) attachment of adapters to the ends of target fragments, and (iii) amplification and sequencing to give reads (Head et al, 2014).

The Nextera kit combines the process of fragmentation and adapter ligation into a single step, tagmentation, using a hyperactive Tn5 transposase. This hyperactive Tn5 transposase is bound in a complex with a synthetic DNA complex and the flanked adapters for deep sequencing (Brouillette et al, 2012). However, customisation of this protocol would be required as the transposase used in the Nextera kit generally shears target sequences into 200-400 bp fragments (Feng et al, 2018). Compared to mechanical fragmentation methods, this protocol is very sensitive to the amount of DNA input. In order to obtain fragmentation of target sequence separated by the appropriate distances, the ratio of transposase complexes to sample DNA is critical (Feng et al, 2018). So, for the generation of larger fragments the amount of transposase complexes present in the reaction mix will need to be reduced. Further customisation of this kit is required to ligate our Library 1B primers (**Table 2.2**) containing the Golden Gate cloning adapter sequences, to the ends of the fragments. Rather than random adapter sequences used for next-generation sequencing, so PCR amplification and then GG cloning can be utilised instead. Or it is possible to use PCR to add Golden Gate adapters to the Nextera library. Utilising this tagmentation step in our approach could decrease the number of steps in this protocol further, therefore decreasing sample loss and streamlining the production of a high-quality genomic DNA library. Further research is needed into the use of the wild-type Tn5 transposase over the hyperactive Tn5 transposase to generate longer target fragments. However, the wild-type transposase is a protein of low-level activity with hyperactive form specifically mutated to enhance Tn5 activity (Goryshin & Reznikoff, 1998).

The use of more traditional shotgun library construction methods, of fragmentation of the BAC and TA cloning, highlighted in **Chapter 5**, were not successful. Using Taq polymerase and A-tailing was not an efficient strategy conducive to building a library containing large inserts. An alternative strategy would be to separate the polishing (e.g. by using T4 DNA Polymerase to fill in 5' overhangs and chew back 3' overhangs to form blunt ends) and A-tailing (e.g. using Taq), with a purification step in between to prevent T4 DNA Pol removing the A-tails.

### 8.3 Discovery of ECR for *cMyc* gene

Using a candidate approach we identified a 1333 bp ECR within the human genome (hg38) located upstream of the *cMyc* gene on chromosome 8. In the future further modifications can be made to the pTK-Citrine vector to include *Tol2* inverted repeats. *Tol2* is part of the transposase family, the protein can catalyse transposition of a non-autonomous *Tol2* construct. This means the terminal inverted repeats are always seen adjacent to integrated *Tol2* elements, but the transposase coding region is deleted (Kawakami, 2007). Thus, hypothetically any foreign DNA fragments can be cloned within these inverted repeats. This modification to include the *Tol2* repeats, should integrate the plasmid into the genome, whereby it will be transmitted equally and without dilution to both daughter cells via genome replication, circumventing plasmid dilution by cell proliferation.

One of the predicted TF binding sites/motifs possibly contributing to ECR2's enhancer activity is *Six3*. The expression pattern for *Six3* shows expression in the anterior neural folds, overlapping *cMyc* expression (**Figure 1.3.3**). *Six3* is thought to be acting as a transcriptional repressor (Kobayashi et al, 2001), restricting the expression of the ECR2, creating two smaller growth zones, rather than one large one. To test this hypothesis, functional experiments can be performed to mutate the *Six3* binding site. As TF binding sites are somewhat degenerate, multiple base mutation need to be introduced (G to T or A to C) to be sufficient to alter the binding site and block interaction with the

*Six3* gene. The mutational effects can be assessed by electroporating embryos with the reporter vector, now containing the mutated binding site.

Given the presence of TF binding sites in the ECR2 enhancer for eye and forebrain specific transcription factors such as *Six3* (Jeong et al, 2008), *Lhx2* (Goodbole et al, 2018), *Meis2* and others (Coy et al, 2011), a detailed characterisation of the regulatory role of ECR2 in the putative anterior neural fold growth zone is still needed. While *Six3* has been shown to indirectly suppress *cMyc* expression in tumorigenesis (Yu et al, 2017), over-expression of *Lhx2* has been known to increase the activation of downstream Wnt/ $\beta$ -catenin gene, *cMyc* (Zhou et al, 2014). These predicted TF binding sites provide an insight into the signalling pathways that might regulate ECR2 expression in the anterior neural fold region.

It should be noted the expression pattern seen for the ECR2 reporter could be restricted to epiblast and neural crest solely, and the predicted expression within the putative growth zone within the anterior neural folds will never be seen. In recent years it has been well documented that several enhancers can interact with one gene, regulating different spatial and temporal activation of a gene (Kyrchanova & Georgiev, 2021). When taken out of genomic context, enhancer sequences have also been known to cause ectopic or premature activation (Mok et al, 2021). Further investigation is required into understanding this novel CRE's functional role within the genome.

ECR2 spans a 1333 bp genomic region upstream of the *cMyc* gene. A full enhancer is usually only a few hundred base pairs in total. Generally, an enhancer consists of a core region, which determines the specificity and activity of the enhancer, and adjoining auxiliary regions, which enhance the core's activity (Uchikawa et al, 2017). Removal of these adjoining elements usually reduces the enhancer activity, but specificity is maintained. Whereas removal of the core region will fully inactivate the enhancer activity. This allows for the distinction of necessary and insufficient sequences within the enhancer region.

Identifying the minimum core essential enhancer can be accomplished by an 'enhancer bashing' assay: the stepwise trimming of base pairs from either end of the original genomic fragment (Rickels & Shilatfard, 2018). A candidate approach is traditionally considered a low-throughput assay to predict enhancers, as sequences have to be individually tested, but when used to further characterise an already confirmed CRE, it can be a useful tool. Using phylogenetic conservation of overlapping TF binding sites, 'blocks' of sequences can be selected for deletion analysis (Prasad & Paulson, 2011). If a point mutation or deletion made within a specific region decreases/terminates the level subsequent transcription of the enhancer's target gene or reporter gene, then that region can be considered necessary and may be the region interacting with the promoter or another regulatory element.

## 8.4 Conclusion

This study focused on the optimisation and construction of a 3D shotgun reporter library for the Cyclopia causing gene, *Six3*. Utilising BAC clones and group testing strategy our approach looked at creating an efficient, low-cost, high-throughput functional assay that could screen large regions of the genome for enhancer activity and analyse its spatial-temporal expression *in vivo*.

Throughout the course of comparing both unbiased and candidate approaches, a putative enhancer for the *cMyc* gene was discovered. It should be noted that while this enhancer sequence was found through a candidate approach, the positive hits associated with this type of assay is much lower than that of an unbiased approach.

The future of this project will aim at further optimising our approach to construct a 3D shotgun reporter library for genes associated with certain disease networks, followed by a rapid functional screen utilising group testing strategy. Once coordinates of the core enhancer elements have been identified, possible collaboration with clinicians is possible to examine known SNPs found within

patient samples. These known mutations can then be induced within a functional assay screen, seeing if the polymorphism is validated. Locating and identifying enhancers increases our knowledge of the gene-regulatory networks and allows for reverse-engineering of the gene circuits that regulate healthy eye and forebrain development, which is of fundamental interest.

It is crucial to identify enhancers for these causative HPE genes and they may harbour causative mutations in the human population. As such it is vital to optimise and compare enhancer detection techniques at our disposal.

## Supplementary Data

Table S1: Table showing dilutions of Meox1 Reporter

<b>Titration :</b>	<b>1:1</b>	<b>1:5</b>	<b>1:25</b>	<b>1:125</b>	<b>-ve Control</b>
<b>Cherry</b>	2 ug/ul	2 ug/ul	2 ug/ul	2 ug/ul	2 ug/ul
Empty Reporter	-	1.6 ug/ul	1.92 ug/ul	1.98 ug/ul	1 ug/ul
Meox1 Reporter	2 ug/ul	0.4 ug/ul	0.08 ug/ul	0.02 ug/ul	-
Total Reporter	2 ug/ul	2 ug/ul	2 ug/ul	2 ug/ul	2 ug/ul
Total DNA	4 ug/ul	4 ug/ul	4 ug/ul	4 ug/ul	4 ug/ul

## References

- Adey A, Morrison HG, Asan, Xun X, Kitzman JO, Turner EH, Stackhouse B, MacKenzie AP, Caruccio NC, Zhang X, Shendure J. (2010). "Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition" *Genome Biol* vol.11(12):R119. doi: 10.1186/gb-2010-11-12-r119.
- Adler R, Canto-Soler MV. (2007). "Molecular mechanisms of optic vesicle development: complexities, ambiguities and controversies" *Dev Biol* vol.305: 1-13.
- Agyeman A, Jha BK, Mazumdar T, Houghton JA. (2014). "Mode and specificity of binding of the small molecule GANT61 to GLI determines inhibition of GLI-DNA binding" *Oncotarget* vol.5(12):4492-503. doi: 10.18632/oncotarget.2046.
- Ahituv N, Zhu Y, Visel A, Holt A, Afzal V, Pennacchio LA, Rubin EM. (2007). "Deletion of ultraconserved elements yields viable mice" *PLoS Biol* vol.5, e234.
- Andrews MG, Nowakowski TJ. (2019). "Human brain development through the lens of cerebral organoid models" *Brain Res* vol.15;1725:146470. doi: 10.1016/j.brainres.2019.146470.
- Andrews MG, Nowakowski TJ. (2019). "Human brain development through the lens of cerebral organoid models" *Brain Research* vol. 1725:146470.
- Aruga J, Nagai T, Tokuyama T, Hayashizaki Y, Okazaki Y, Chapman VM, Mikoshiba K. (1996). "The mouse zic gene family. Homologues of the Drosophila pair-rule gene odd-paired" *J Biol Chem* vol.12; 271(2):1043-7.
- Asakawa S, Abe I, Kudoh Y, Kishi N, Wang Y, Kubota R, Kudoh J, Kawasaki K, Minoshima S, Shimizu N. (1997). "Human BAC library: construction and rapid screening" *Gene* vol.191(1):69-79. doi: 10.1016/s0378-1119(97)00044-9.
- Ausubel M, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JS, Struhl K. (1989). "Current protocols in molecular biology Volumes 1 and 2" John Wiley & Sons, Inc., *Mol. Reprod. Dev.*, 1: 146-146.
- Bae G-U, Domene S, Roessler E, Schachter K, Kang J-S, Muenke M, Krauss RS. (2011). "Mutations in CDON, encoding a hedgehog receptor, result in holoprosencephaly and defective interactions with other hedgehog receptors" *Am. J. Hum. Genet* vol.89: 231-230.
- Barr Jr M, Hanson JW, Currey K, Sharp S, Toriello H, Schmickel RD, Wilson GN. (1983). "Holoprosencephaly in infants of diabetic mothers" *J Pediatr* vol.102(4):565-8. doi: 10.1016/s0022-3476(83)80185-1.
- Bellmeyer A, Krase J, Lindgren J, LaBonne C. (2003). "The protooncogene c-myc is an essential regulator of neural crest formation in xenopus" *Dev Cell* vol.4(6):827-39. doi: 10.1016/s1534-5807(03)00160-6.
- Belloni E, Muenke M, Roessler E, Traverso G, Siegel-Bartelt J, Frumkin A, Mitchell HF, Donis-Keller H, Helms C, Hing AV, Heng HHQ, Koop B, Martindale D, Rommens JM, Tsui L.-C., Scherer S W. (1996). "Identification of Sonic hedgehog as a candidate gene responsible for holoprosencephaly" *Nature Genet* vol.14: 353-356.

Benabdallah NS, Gautier P, Hekimoglu-Balkan B, Lettice LA, Bhatia S, Bickmore WA. (2016). "SBE6: a novel long-range enhancer involved in driving sonic hedgehog expression in neural progenitor cells" *Open Biol* vol.6(11):160197. doi: 10.1098/rsob.160197.

Bendavid C, Dubourg C, Gicquel I, Pasquier L, Saugier-Veber P, Durou MR, Jaillard S, Frebourg T, Haddad BR, Henry C, Odent S, David V. (2006). "Molecular evaluation of fetuses with holoprosencephaly shows high incidence of microdeletions in the HPE genes" *Hum Genet* vol.119 (1-2): 1-8. 10.1007/s00439-005-0097-6.

Bendavid C, Haddad BR, Griffin A, Huizing M, Dubourg C, Gicquel I, Cavalli LR, Pasquier L, Shanske AL, Long R, Ouspenskaia M, Odent S, Lacbawan F, David V, Muenke M. (2006). "Multicolour FISH and quantitative PCR can detect submicroscopic deletions in holoprosencephaly patients with a normal karyotype" *J Med Genet* vol.43(6):496-500. doi: 10.1136/jmg.2005.037176.

Benerji J, Olson L, Schaffner W. (1983). "A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes" *Cell* vol. 33(3): 729-740.

Biolabs, New England. "Everything You Ever Wanted to Know about Type II Restriction Enzymes." NEB, <https://www.neb.com/tools-and-resources/feature-articles/everything-you-ever-wanted-to-know-about-type-ii-restriction-enzymes>. Accessed: Jan 2022.

*Birth Defects Research* vol.12 (3):219-234.

Bozek M, Gompel N. (2020). "Developmental Transcriptional Enhancers: A Subtle Interplay between Accessibility and Activity" *BioEssays* vol.42(4):1900188.

Brooker AS, Berkowitz KM. (2014). "The roles of cohesins in mitosis, meiosis, and human health and disease" *Methods Mol Biol* vol.1170():229-66.

Brouillette S, Kuersten S, Mein C, Bozek M, Terry A, Dias KR, Bhaw-Rosun L, Shintani Y, Coppens S, Ikebe C, Sawhney V, Campbell N, Kaneko M, Tano N, Ishida H, Suzuki K, Yashiro K. (2012). "A simple and novel method for RNA-seq library preparation of single cell cDNA analysis by hyperactive Tn5 transposase" *Dev Dyn* vol.241(10):1584-90. doi: 10.1002/dvdy.23850.

Brownie J, Shawcross S, Theaker J, Whitcombe D, Ferrie R, Newton C, Little S. (1997). "The elimination of primer-dimer accumulation in PCR" *Nucleic Acids Research* vol.25(16):3235-3241.

Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. (2013). "Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position" *Nat Methods* vol.10(12):1213-8. doi: 10.1038/nmeth.2688.

Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. (2015). "ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide" *Curr Protoc Mol Biol* vol.109:21.29.1-21.29.9. doi: 10.1002/0471142727.mb2129s109.

Bulger M, Groudine M. (2011). "Functional and mechanistic diversity of distal transcription enhancers" *Cell* vol.4;144(3):327-39. Erratum in: *Cell* vol.4;144(5):825. doi: 10.1016/j.cell.2011.01.024.

Burke DT, Carle GF, Olson MV. (1987). "Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors" *Science* vol.236: 806-812.

Calo E, Wysocka J. (2013). "Modification of enhancer chromatin: what, how, and why?" *Mol Cell* vol.49(5):825-37. doi: 10.1016/j.molcel.2013.01.038.

Carl M, Loosli F, Wittbrodt J. (2002). "Six3 inactivation reveals its essential role for the formation and patterning of the vertebrate eye" *Development* vol.129 (17): 4057–4063. doi: <https://doi.org/10.1242/dev.129.17.4057>.

Castillo HA, Cravo RM, Azambuja AP, Simões-Costa MS, Sura-Trueba S, Gonzalez J, Slonimsky E, Almeida K, Abreu JG, de Almeida MA, Sobreira TP, de Oliveira SH, de Oliveira PS, Signore IA, Colombo A, Concha ML, Spengler TS, Bronner-Fraser M, Nobrega M, Rosenthal N, Xavier-Neto J. (2010). "Insights into the organization of dorsal spinal cord pathways from an evolutionarily conserved raldh2 intronic enhancer" *Development* vol.137(3):507-18.

Cavalheiro GR, Pollex T & Furlong E. (2021). "To loop or not to loop: what is the role of TADs in enhancer function and gene regulation?" *Current Opinion in Genetics & Development* vol.67:119-129.

Chen C-P, Su T-H, Chern S-R, Su J-W, Lee C-C, Wang W. (2012). "Alobar holoprosencephaly, cebocephaly, and micropenis in a Klinefelter fetus of a diabetic mother" *Taiwan J Obstet Gynecol* vol. 51(4):630-4. doi: 10.1016/j.tjog.2012.09.021.

Chen J, Kadlubar FF, Chen JZ. (2007). "DNA supercoiling suppresses real-time PCR: a new approach to the quantification of mitochondrial DNA damage and repair" *Nucleic Acids Res* vol.35:1377–88.

Chen J, Streit A. (2005). "A medium-scale assay for enhancer validation in amniotes" *Developmental Dynamics* vol.244 (10): 1291-1299.

Chiang C, Litingtung Y, Lee E, Young KE, Corden JL, Westphal H, Beachy PA. (1996). "Cyclopia and defective axial patterning in mice lacking Sonic hedgehog gene function" *Nature* vol.3;383(6599):407-13. doi: 10.1038/383407a0.

Chow RL, Lang RA. (2001). "Early eye development in vertebrates" *Annu Rev Cell Dev Biol* vol.17: 255-296.

Clark, J. M. (1988). "Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases" *Nucleic Acids Res* vol.16: 9677–9686.

Cohen Jr MM. (2006). "Holoprosencephaly: Clinical, anatomic, and molecular dimensions" *Birth Defects Res. A. Clin. Mol. Teratol* vol.76:658-673.

Cohen, M et al. (2009). "Ptch1 and Gli regulate Shh signalling dynamics via multiple mechanisms" *Nature communications* vol. 6 6709. doi:10.1038/ncomms7709.

Collart MA, Panasencko OO. (2012). "The Ccr4--not complex" *Gene* vol.492(1):42-53. doi: 10.1016/j.gene.2011.09.033. Epub 2011 Oct 15. PMID: 22027279.

Conte I, Bovolenta P. (2007). "Comprehensive characterization of the cis-regulatory code responsible for the spatio-temporal expression of *olSix3.2* in the developing medaka forebrain" *Genome Biol* vol.8(7):R137. doi: 10.1186/gb-2007-8-7-r137.

Coy S, Caamaño JH, Carvajal J, Cleary ML, Borycki AG. (2011). "A novel Gli3 enhancer controls the Gli3 spatiotemporal expression pattern through a TALE homeodomain protein binding site" *Mol Cell Biol* vol.31(7):1432-43. doi: 10.1128/MCB.00451-10.

Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, et al. (2010). "Histone H3K27ac separates active from poised enhancers and predicts developmental state" *Proc. Natl. Acad. Sci* vol.107:21931–21936.

Datta S, Budhaliya R, Chatterjee S et al. (2016). "Enhancement of PCR Detection Limit by Single-Tube Restriction Endonuclease-PCR (RE-PCR)" *Mol Diagn Ther* vol.20: 297–305.

De Franco E, Watson RA, Weninger WJ, Wong CC, Flanagan SE, Caswell R, Green A, Tudor C, Lelliott CJ, Geyer SH, Maurer-Gesek B, Reissig LF, Lango Allen H, Caliebe A, Siebert R, Holterhus PM, Deeb A, Prin F, Hilbrands R, Heimberg H, Ellard S, Hattersley AT, Barroso I. (2019). "A Specific CNOT1 Mutation Results in a Novel Syndrome of Pancreatic Agenesis and Holoprosencephaly through Impaired Pancreatic and Neurological Development" *Am J Hum Genet* vol.104(5):985-989. doi: 10.1016/j.ajhg.2019.03.018.

Deaton A, Bird A. (2011). "CpG islands and the regulation of transcription" *Genes & Development* vol.25(10):1010-1022.

Delaune E, Lemaire P, Kodjabacjan L. (2005). "Neural induction in *Xenopus* requires early FGF signalling in addition to BMP inhibition" *Development* vol.132(2)L 299-310.

Desesso JM. (2019). "Of embryos and tumors: Cyclopia and the relevance of mechanistic teratology"

Ding Q, Motoyama J, Gasca S, Mo R, Sasaki H, Rossant J, Hui CC. (1998). "Diminished Sonic hedgehog signaling and lack of floor plate differentiation in *Gli2* mutant mice" *Development* vol.125(14):2533-43.

Doroquez DB, Rebay I. (2006). "Signal integration during development: mechanisms of EGFR and Notch pathway function and cross-talk" *Crit Rev Biochem Mol Biol* vol.41(6):339-85. doi: 10.1080/10409230600914344.

Dubourg C, Bendavid C, Pasquier L et-al. (2007). "Holoprosencephaly" *Orphanet Journal of Rare Diseases* vol.2(1): 8. doi:10.1186/1750-1172-2-8.

Dubourg C, Lazaro L, Pasquier L, Bendavid C, Blayau M, Le Duff F, Durou MR, Odent S, David V. (2004). "Molecular screening of SHH, ZIC2, SIX3, and TGIF genes in patients with features of holoprosencephaly spectrum: Mutation review and genotype-phenotype correlations" *Hum Mutat* vol.24 (1): 43-51. 10.1002/humu.20056.

Eilers M, Eisenman RN. (2008). "Myc's broad reach" *Genes Dev* 22: 2755– 2766.

Ekker SC, Ungar AR, Greenstein P, von Kessler DP, Porter JA, Moon, RT, et al. (1995). "Patterning activities of vertebrate hedgehog proteins in the developing eye and brain" *Curr. Biol* vol.5: 944–955. doi: 10.1016/S0960-9822(95)00185-0.

Elms P, Scurry A, Davies J, Willoughby C, Hacker T, Bogani D, Arkell R. (2004). "Overlapping and distinct expression domains of *Zic2* and *Zic3* during mouse gastrulation" *Gene Expr Patterns* vol. 4(5):505-11.

ENCODE Project Consortium. (2012). "An integrated encyclopedia of DNA elements in the human genome" *Nature* vol.489, 57–74.

England SJ, Blanchard GB, Mahadevan L, Adams RJ. (2006). "A dynamic fate map of the forebrain shows how vertebrate eyes form and explains two causes of cyclopia" *Development* vol.133:4613-4617.

Engler C, Kandzia R, Marillonnet S. (2008). "A one pot, one step, precision cloning method with high throughput capability" *PLoS One* vol.3(11):e3647. doi: 10.1371/journal.pone.0003647.

- Ernst, J. & Kellis, M. (2010). "Discovery and characterization of chromatin states for systematic annotation of the human genome" *Nature Biotech* vol.28: 817–825.
- Feng K, Costa J, Edwards JS. (2018). "Next-generation sequencing library construction on a surface" *BMC Genomics* vol.19: 416.
- Fischer AJ, Bosse JL, El-Hodiri HM. (2013). "The ciliary marginal zone (CMZ) in development and regeneration of the vertebrate eye" *Experimental Eye Research* vol.116:199-204.
- Fornes O, Castro-Mondragon JA, Khan A, et al. (2019). "JASPAR 2020: update of the open-access database of transcription factor binding profiles" *Nucleic Acids Res.* doi: 10.1093/nar/gkz1001.
- Frank CL, Liu F, Wijayatunge R, Song L, Biegler MT, Yang MG, Vockley CM, Safi A, Gersbach CA, Crawford GE, West AE. (2015). "Regulation of chromatin accessibility and Zic binding at enhancers in the developing cerebellum" *Nat Neurosci* vol.18(5):647-56. doi: 10.1038/nn.3995.
- Frankel N, Davis GK, Vargas D, Wang S, Payre F, Stern DL. (2010). "Phenotypic robustness conferred by apparently redundant transcriptional enhancers" *Nature* vol.466: 490–493.
- Frenkel LD, Gaur S, Tsoia M, Scudder R, Howell R, Kesarwala H. (1990). "Cytomegalovirus infection in children with AIDS" *Rev Infect Dis* vol.12 Suppl 7: S820-6.
- Gallardo V, Bovolenta P. (2018). "Positive and negative regulation of Shh signalling in vertebrate retinal development" *F1000Res* vol.7:F1000.
- Gehring, W. J. (1996). "The master control gene for morphogenesis and evolution of the eye. Genes Cells" *Genes to Cells* vol.1: 11-15.
- Geng X, Guillermo O. (2009). "Pathogenesis of holoprosencephaly" *The Journal of clinical investigation* vol. 119,6 (2009): 1403-13. doi:10.1172/JCI38937.
- Geng X, Speirs C, Lagutin O, Inbal A, Liu W, Solnica-Krezel L, Jeong Y, Epstein DJ, Oliver G. (2008). "Haploinsufficiency of Six3 fails to activate Sonic hedgehog expression in the ventral forebrain and causes holoprosencephaly" *Dev Cell* vol.15(2):236-47. doi: 10.1016/j.devcel.2008.07.003.
- Glaser T, Walton DS, Maas RL. (1992) "Genomic structure, evolutionary conservation and aniridia mutations in the human PAX6 gene" *Nat. Genet* vol. 2: 232– 39.
- Godbole G, Shetty AS, Roy A, D'Souza L, Chen B, Miyoshi G, Fishell G, Tole S. (2018). "Hierarchical genetic interactions between FOXP1 and LHX2 regulate the formation of the cortical hem in the developing telencephalon" *Development* vol.145(1):dev154583. doi: 10.1242/dev.154583.
- Goryshin IY, Reznikoff WS. (1998). "Tn5 in vitro transposition" *J Biol Chem* vol.273(13):7367-74. doi: 10.1074/jbc.273.13.7367.
- Goryshin IY, Reznikoff WS. (1998). "Tn5 in vitro transposition" *J Biol Chem* vol.273(13):7367-74. doi: 10.1074/jbc.273.13.7367. PMID: 9516433.
- Gounongbé C, Marangoni M, Gouder de Beauregard V, Delaunoy M, Jissendi P, Cassart M, Désir J. Middle interhemispheric variant of holoprosencephaly. (2020). "First prenatal report of a ZIC2 missense mutation" *Clin Case Rep* vol.30;8(7):1287-1292. doi: 10.1002/ccr3.2896.
- Graw J. (2010). "Eye Development" *Current Topics in Developmental Biology* vol.90:343-386.

- Gripp KW, Wotton D, Edwards MC, Roessler E, Ades L, Meinecke P, Richieri-Costa, A, Zackai EH, Massague J, Muenke M, Elledge SJ. (2000). "Mutations in TGIF cause holoprosencephaly and link NODAL signalling to human neural axis determination" *Nature Genet* vol.25: 205-208.
- Gunhaga, L. (2011). "The lens: a classical model of embryonic induction providing new insights into cell determination in early development" *Phil. Trans. R. Soc. B* vol.366, 1193-1203. doi:10.1098/rstb.2010.0175.
- Halder G, Callaerts P, Gehring WJ. (1995). "Induction of ectopic eyes by targeted expression of the eyeless gene in *Drosophila*" *Science* vol.267: 1788– 92.
- Hamburger V, Hamilton HL. (1951). "A series of normal stages in the development of the chick embryo" *Dev Dyn.* 1952 Dec;195(4):231-72. doi: 10.1002/aja.1001950404.
- Hanson IM, Seawright A, Hardman K, Hodgson S, Zaletayev D, et al. (1993) "PAX6 mutations in aniridia" *Hum. Mol. Genet* vol.2: 915– 20.
- Hare EE, Peterson BK, Iyer VN, Meier R, Eisen MB. (2008). "Sepsid even-skipped enhancers are functionally conserved in *Drosophila* despite lack of sequence conservation" *PLoS Genet* vol.4(6):e1000106.
- Head SR, Komori HK, LaMere SA, Whisenant T, Van Nieuwerburgh F, Salomon DR, Ordoukhanian P. (2014). "Library construction for next-generation sequencing: overviews and challenges" *Biotechniques* vol.56(2):61-4, 66, 68, passim. doi: 10.2144/000114133.
- Heavens D, Vasoya D, Bennet G, Musk H, Lipscombe J, Sampath D, Leggett R, Rogers A, Waugh R, Ayling S, Clark MD. (2016). "Back to BACs? A High-Throughput, low cost BAC sequencing pipeline" From the The Genome Analysis Centre, Norwich Research Park, United Kingdom.
- Heinz S, Romanoski CE, Benner C, Glass CK. (2015). "The selection and function of cell type-specific enhancers" *Nature Reviews Molecular Cell Biology* vol.16: 144-154.
- Heisenberg CP, Houart C, Take-uchi M, Rauch GJ, Young N, Coutinho P, Masai I, Caneparo L, Concha ML, Geisler R, et al. (2001). "A mutation in the Gsk3-binding domain of zebrafish Masterblind/Axin1 leads to a fate transformation of telencephalon and eyes to diencephalon" *Genes Dev* vol.15: 1427-1434.
- Hill RE, Favor J, Hogan BL, Ton CC, Saunders GF, et al. (1991). "Mouse small eye results from mutations in a paired-like homeobox- containing gene" *Nature* vol.354: 522– 25.
- Hong M, Christ A, Christa A, Willnow TE, Krauss RS. (2020). "Cdon mutation and fetal alcohol converge on Nodal signaling in a mouse model of holoprosencephaly" *eLife* 9:e60351. DOI: <https://doi.org/10.7554/eLife.60351>.
- Hong M, Christ A, Christa A, Willnow TE, Krauss RS. (2020). "Cdon mutation and fetal alcohol converge on Nodal signaling in a mouse model of holoprosencephaly" *Elife* vol.9:e60351. doi: 10.7554/eLife.60351.
- Hong M, Krauss RS. (2012). "Cdon mutation and fetal ethanol exposure synergize to produce midline signaling defects and holoprosencephaly spectrum disorders in mice" *PLoS Genet* vol.8(10):e1002999.

Houtmeyers R, Tchouate Gainkam O, Glanville-Jones HA, Van den Bosch B, Chappell A, Barratt KS, Souopgui J, Tejpar S, Arkell RM. (2016). "Zic2 mutation causes holoprosencephaly via disruption of NODAL signalling" *Human Molecular Genetics* vol.25(15):3946–3959.

Ing-Simmons E, Vaid R, Bing XY, et al. (2021). "Independence of chromatin conformation and gene regulation during Drosophila dorsoventral patterning" *Nat Genet* vol.53: 487–499.

Inoue T, Inoue YU, Asami J, Izumi H, Nakamura S, Krumlauf R. (2008). "Analysis of mouse Cdh6 gene regulation by transgenesis of modified bacterial artificial chromosomes" *Developmental Biology* vol.315(2):506-520.

Ishiguro A, Hatayama M, Otsuka MI, Aruga J. (2018). "Link between the causative genes of holoprosencephaly: Zic2 directly regulates Tgif1 expression" *Sci Rep* vol.8(1):2140. doi: 10.1038/s41598-018-20242-2.

Ivenso ID, Lillian TD. (2016). "Simulation of DNA Supercoil Relaxation" *Biophysical Journal* vol.110(10): 2176-2184.

Jeong Y, Epstein D.J. (2003). "Distinct regulators of Shh transcription in the floor plate and notochord indicate separate origins for these tissues in the mouse node" *Development* vol.130:3891–3902.

Jeong Y, Leskow F, El-Jaick K, et al. (2008). "Regulation of a remote Shh forebrain enhancer by the Six3 homeoprotein" *Nat Genet* vol.40: 1348–1353.

Jin Q, Yu LR, Wang L, Zhang Z, Kasper LH, Lee JE, Wang C, Brindle PK, Dent SY, Ge K. (2011). "Distinct roles of GCN5/PCAF-mediated H3K9ac and CBP/p300-mediated H3K18/27ac in nuclear receptor transactivation" *EMBO J* vol.30(2):249-62.

Kawakami, K. (2007). "Tol2: a versatile gene transfer vector in vertebrates" *Genome Biol* vol.8: S7.

Kerosuo L, Bronner ME. (2016). "cMyc Regulates the Size of the Premigratory Neural Crest Stem Cell Pool" *Cell Rep* vol.17(10):2648-2659. doi: 10.1016/j.celrep.2016.11.025.

Khudyakov J, Bronner-Fraser M. (2009). "Comprehensive spatiotemporal analysis of early chick neural crest network genes" *Dev Dyn* vol.238(3):716-23. doi: 10.1002/dvdy.21881.

Kim J, Kato M, Beachy PA. (2009). "Gli2 trafficking links Hedgehog-dependent activation of smoothened in the primary cilium to transcriptional activation in the nucleus" *Proc. Natl. Acad. Sci. U. S. A* vol.106:21666-21671.

Kitamoto J, Hyer J. (2010). "The expression of Wnt2b in the optic cup lip requires a border between the pigmented and nonpigmented epithelium" *Mol Vis* vol.14(16):2701-17.

Knepper JL, James AC, Ming JE. (2006). "TGIF, a gene associated with human brain defects, regulates neuronal development" *Dev Dyn* vol.235(6):1482-90. doi: 10.1002/dvdy.20725.

Knierim E, Lucke B, Schwarz JM, Schuelke M, Seelow D. (2011). "Systematic Comparison of Three Methods for Fragmentation of Long-Range PCR Products for Next Generation Sequencing" *PLOS ONE* vol.6(11): e28240.

Knierim E, Lucke B, Schwarz JM, Schuelke M, Seelow D. (2016). *6 Methods to Fragment Your DNA / RNA for Next-Gen Sequencing* Available at: <https://blog.genohub.com/2016/07/22/fragmentation-of-dna-rna-for-next-gen-sequencing-library-prep/> (Accessed: 21 November 2021).

- Kobayashi M, Nishikawa K, Suzuki T, Yamamoto M. (2001). "The homeobox protein Six3 interacts with the Groucho corepressor and acts as a transcriptional repressor in eye and forebrain formation" *Dev Biol* vol.232(2):315-26. doi: 10.1006/dbio.2001.0185.
- Kruszka P, Berger SI, Weiss K, Everson JL, Martinez AF, Hong S, Anyane-Yeboah K, Lipinski RJ, Muenke M. (2014). "A CCR4-NOT Transcription Complex, Subunit 1, CNOT1, Variant Associated with Holoprosencephaly" *Am J Hum Genet* vol.104(5):990-993. doi: 10.1016/j.ajhg.2019.03.017.
- Kruszka P, Berger SI, Weiss K, Everson JL, Martinez AF, Hong S, Anyane-Yeboah K, Lipinski RJ, Muenke M. (2019). "A CCR4-NOT transcription complex, subunit 1, CNOT1, variant associated with holoprosencephaly" *Am. J. Hum. Genet* vol.104: 990-993.
- Kucera R, Cantor E. (2018). "Breaking through the Limitations of Golden Gate Assembly—The Co-Evolution of Test Systems, Engineered Enzymes and Understanding Ligase Fidelity" From the New England Biolabs, Inc.
- Kvon EZ. (2015). "Using transgenic reporter assays to functionally characterize enhancers in animals" *Genomics* vol.106(3):185-192.
- Kyrchanova O, Georgiev P. (2021). "Mechanisms of Enhancer-Promoter Interactions in Higher Eukaryotes" *Int. J. Mol. Sci* vol.22(2): 671.
- Laghi L, Randolph AE, Malesci A, Boland CR. (2004). "Constraints imposed by supercoiling on in vitro amplification of polyomavirus DNA" *J Gen Virol* vol.85:3383–8.
- Lagutin OV, Zhu CC, Kobayashi D, Topczewski J, Shimamura K, Puelles L, Oliver G. (2003). "Six3 repression of Wnt signaling in the anterior neuroectoderm is essential for vertebrate forebrain development" *Genes and Development* vol.17: 368–379.
- Larin Z, Monaco AP, Lehrach H. (1991). "Yeast artificial chromosome libraries containing large inserts from mouse and human DNA" *Proc. Natl. Acad. Sci.* vol.88: 4123–4127.
- Lee B, Yoon J, Tri Lam D, Yoon J, Baek K, Jeong Y. (2017). "Identification of a conserved cis-regulatory element controlling mid-diencephalic expression of mouse Six3" *Genesis* vol.55(3). doi: 10.1002/dvg.23017.
- Lee JJ, Ekker SC, von Kessler DP, Porter JA, Sun BI, Beachy PA. (1994). "Autoproteolysis in hedgehog protein biogenesis" *Science* vol.2;266(5190):1528-37. doi: 10.1126/science.7985023.
- Lee TI, Young RA. (2014). "Transcriptional regulation and its misregulation in disease" *Cell* vol.152(6):1237-51. doi: 10.1016/j.cell.2013.02.014.
- Lettice LA, Heaney SJ, Purdie LA, Li L, de Beer P, Oostra BA, Goode D, Elgar G, Hill RE, de Graaff E. (2003). "A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly" *Hum Mol Genet* vol.12(14):1725-35.
- Li H, Tierney C, Wen L, Wu JY, Rao Y. (1997). "A single morphogenetic field gives rise to two retina primordia under the influence of the prechordal plate" *Development* vol.124, 603–615.
- Liu X, Huang J, Chen T, et al. (2008). "Yamanaka factors critically regulate the developmental signaling network in mouse embryonic stem cells" *Cell Res* vol.18: 1177–1189.
- Liu ZG & Schwartz LM. (1992). "An efficient method for blunt-end ligation of PCR products" *Biotechniques* vol.12(28): 30.

- Long HK, Prescott SL, Wysocka J. (2016). "Ever-changing landscapes: transcriptional enhancers in development and evolution" *Cell* vol.167:1170-1187.
- Loukianov E, Loukianova T, Periasamy M. (1997). "Efficient cloning method that selects the recombinant clones" *Biotechniques* vol.23(2):292-5. doi: 10.2144/97232st05.
- Lupo G, Bertacchi M, Carucci N, Augusti-Tocco G, Biagioni S, Cremisi F. (2014). "From pluripotency to forebrain patterning: an in vitro journey astride embryonic stem cells" *Cell Mol Life Sci* vol.71(15):2917-30. doi: 10.1007/s00018-014-1596-1.
- Lupo G, Harris W, Lewis K. (2006). "Mechanisms of ventral patterning in the vertebrate nervous system" *Nat Rev Neurosci* vol.7: 103–114.
- Lv B, Dai Y, Liu Ju, Zhuge Q, Li D. (2015). "The Effect of Dimethyl Sulfoxide on Supercoiled DNA Relaxation Catalyzed by Type I Topoisomerases" *BioMed Research International* vol. 2015, Article ID 320490, 8 pages.
- Macdonald R, Barth A, Xu Q, Holder N, Mikkola I, Wilson SW. (1995). "Midline signalling is required for Pax gene regulation and patterning of the eyes" *Development* vol.121:3267-3278.
- Macdonald R, Barth KA, Xu Q, Holder N, Mikkola, I, and Wilson SW. (1995). "Midline signalling is required for Pax gene regulation and patterning of the eyes" *Development* vol.121, 3267–3278.
- Magnaldo, T. (2002). "SHH (Sonic hedgehog)" *Atlas Genet Cytogenet Oncol Haematol* vol.6(2):117-121.
- Martinez-Morales JR, Wittbrodt J. (2009). "Shaping the vertebrate eye" *Current Opinion in Genetics & Development* vol.19(5):511-517.
- Mathers PH, Grinberg A, Mahon KA, Jamrich M. (1997). "The Rx homeobox gene is essential for vertebrate eye development" *Nature* vol.387: 603–607.
- Matisse MP, Epstein DJ, Park HL, Platt KA, Joyner AL. (1998). "Gli2 is required for induction of floor plate and adjacent cells, but not most ventral neurons in the mouse central nervous system" *Development* vol.125(15):2759-70.
- Matsumata M, Uchikawa M, Kamachi Y, Kondoh H. (2005). "Multiple N-cadherin enhancers identified by systematic functional screening indicate its Group B1 SOX-dependent regulation in neural and placodal development" *Developmental Biology* vol.286(2):601-617.
- Matsumura, I. (2015). "Why Johnny can't clone: Common pitfalls and not so common solutions" *BioTechniques* vol.59(3): IV-X111. doi:10.2144/000114324.
- Meyer CA, Liu XS. (2014). "Identifying and mitigating bias in next-generation sequencing methods for chromatin biology" *Nat Rev Genet* vol.15(11):709–21.
- Miesfeld JB, Gestri G, Clark BS, Flinn MA, Poole RJ, Bader JR, Besharse JC, Wilson SW, Link BA. (2015). "Yap and Taz regulate retinal pigment epithelial cell fate. *Development* vol.142(17):3021-32. doi: 10.1242/dev.119008.
- Minhas R, Pauls S, Ali S, Doglio L, Khan MR, Elgar G, Abbasi AA. (2015). "Cis-regulatory control of human GLI2 expression in the developing neural tube and limb bud" *Dev Dyn* vol.244(5):681-92. doi: 10.1002/dvdy.24266.

- Miranda TB, Jones PA. (2007). "DNA methylation: the nuts and bolts of repression" *J. Cell Physiol* vol.213: 384-390.
- Mizugishi K, Aruga J, Nakata K, Mikoshiba K. (2001). "Molecular properties of Zic proteins as transcriptional regulators and their relationship to GLI proteins" *J Biol Chem* vol.276(3):2180-8.
- Mok GF, Folkes L, Weldon SA, et al. (2021). "Characterising open chromatin in chick embryos identifies cis-regulatory elements important for paraxial mesoderm formation and axis extension" *Nat Commun* vol.12: 1157.
- Motohashi, K. (2019). "A novel series of high-efficiency vectors for TA cloning and blunt-end cloning of PCR products" *Sci Rep* vol.9: 6417.
- Mouden C, Dubourg C, Carré W, Rose S, Quelin C, Akloul L, Hamdi-Rozé H, Viot G, Salhi H, Darnault P, Odent S, Dupé V, David V. (2016). "Complex mode of inheritance in holoprosencephaly revealed by whole exome sequencing" *Clin Genet* vol.89(6):659-68.
- Mullegama SV, Klein SD, Mulatinho MV, Senaratne TN, Singh K; UCLA Clinical Genomics Center, Nguyen DC, Gallant NM, Strom SP, Ghahremani S, Rao NP, Martinez-Agosto JA. (2017). "De novo loss-of-function variants in STAG2 are associated with developmental delay, microcephaly, and congenital anomalies" *Am J Med Genet A* vol.173(5):1319-1327. doi: 10.1002/ajmg.a.38207. Epub 2017 Mar 11.
- Mutesa L et al. Preprint at <https://arxiv.org/abs/2004.14934> (2020).
- Nagai T, Aruga J, Minowa O, Sugimoto T, Ohno Y, Noda T, Mikoshiba K. (2000). "Zic2 regulates the kinetics of neurulation" *Proc Natl Acad Sci* vol.15; 97(4):1618-23.
- Nam J, Davidson EH. (2012). "Barcoded DNA-tag reporters for multiplex cis-regulatory analysis" *PLoS One* 7:e35934.
- Nieuwkoop PD. (1963). Pattern formation in artificially activated ectoderm (*Rana pipiens* and *Ambystoma punctatum*) *Dev. Biol* vol.7: 255– 79.
- Odent S, Atti-Bitach T, Blayau M, Mathieu M, Aug J, Delezo de AL, Gall JY, Le Marec B, Munnich A, David V, Vekemans M. (1999). "Expression of the Sonic hedgehog (SHH) gene during early human development and phenotypic expression of new mutations causing holoprosencephaly" *Hum Mol Genet* vol.8(9):1683-9. doi: 10.1093/hmg/8.9.1683. PMID: 10441331.
- Odent S, Le Marec B, Munnich A, Le Merrer M, Bonaiti-Pellie C. (1998). "Segregation analysis in nonsyndromic holoprosencephaly" *Am J Med Genet* vol.77 (2): 139-143.
- Okamoto R, Uchikawa M, Kondoh H. (2015). "Sixteen additional enhancers associated with the chicken *Sox2* locus outside the central 50-kb region" *Develop Growth Differ* vol.57: 24-39.
- Ong C, Corces V. (2011). "Enhancer function: new insights into the regulation of tissue-specific gene expression" *Nat. Rev. Genet* vol.12:283-293.
- Ostuni R, Piccolo V, Barozzi I, Polletti S, Termanini A, Bonifacio S, Curina A, Prosperini E, Ghisletti S, Natoli G. (2013). "Latent enhancers activated by stimulation in differentiated cells" *Cell* vol.152(1-2):157-71. doi: 10.1016/j.cell.2012.12.018.

Patwardhan RP, Hiatt JB, Witten DM, Kim MJ, Smith RP, May D, Lee C, Andrie JM, Lee SI, Cooper GM, et al. (2012). "Massively parallel functional dissection of mammalian enhancers in vivo" *Nat. Biotechnol* vol.30: 265–270.

Pera EM, Kessel M. (1997). "Patterning of the chick forebrain anlage by the prechordal plate" *Development* vol.4162, 4153–4162.

Perry MW, Boettiger AN, Bothma JP, Levine M. (2010). "Shadow enhancers foster robustness of *Drosophila* gastrulation" *Curr. Biol* vol.20:1562–1567.

Pfizer P, Muntefering H. (1968). "Cyclopism as a hereditary malformation" *Nature* vol.217: 1071-1072.

Pfizer P, Splitt M, Muntefering H, Friesencker JE. (1982). "Familiaere Haufung von Cyclopie ueber mehrere generationen" *Verh. Dtsch. Ges. Path* vol.66: 169-172.

Pinard R, de Winter A, Sarkis GJ, Gerstein MB, Tartaro KR, Plant RN, Egholm M, Rothberg JM, Leamon JH. (2006). "Assessment of whole genome amplification-induced bias through high-throughput, massively parallel whole genome sequencing" *BMC Genomics* vol.7:216. doi: 10.1186/1471-2164-7-216.

Plank JL, Dean A. (2014). "Enhancer function: mechanistic and genome-wide insights come together" *Mol. Cell* vol.55:5–14.

Porter FD. Cholesterol precursors and facial clefting. (2006). "*J Clin Invest* vol.116(9):2322-5. doi: 10.1172/JCI29872.

Prakash N, Wurst W. (2007). "A Wnt signal regulates stem cell fate and differentiation in vivo" *Neurodegener Dis* vol.4(4):333-8. doi: 10.1159/000101891.

Prasad MS, Paulson AF. (2011). "A combination of enhancer/silencer modules regulates spatially restricted expression of cadherin-7 in neural epithelium" *Developmental Dynamics* vol.240(7):1756-1768.

Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J. (2011). "A unique chromatin signature uncovers early developmental enhancers in humans" *Nature* vol.470(7333):279-83.

Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J. (2011). "A unique chromatin signature uncovers early developmental enhancers in humans." *Nature* vol.470(7333):279-83.

Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J. (2011). "A unique chromatin signature uncovers early developmental enhancers in humans" *Nature* vol.470:279–283.

Rembold M, Loosli F, Adams RJ, Wittbrodt J. (2006). "Individual cell migration serves as the driving force for optic vesicle evagination" *Science* vol.313:1130-1134.

Repetto M, Maziere JC, Citadelle D, Dupuis R, Meier M, Biade S, Quiec D, Roux C. (1990). "Teratogenic effect of the cholesterol synthesis inhibitor AY 9944 on rat embryos in vitro" *Teratology* vol.42 (6): 611-618. 10.1002/tera.1420420605.

Reznikoff WS. (2002). "Tn5 transposition" In *Mobile DNA II*. Craig N, Craigie R, Gellert M, Lambowitz AM. (eds). Washington, DC: American Society of Microbiology vol.1:403–421.

Rickels R, Shilatifard A. (2018). "Enhancer Logic and Mechanics in Development and Disease" *Trends in Cell Biology* vol.28(8):608-630.

- Roe B.A. (2004). "Shotgun Library Construction for DNA Sequencing. In: Zhao S., Stodolsky M. (eds) *Bacterial Artificial Chromosomes*" *Methods in Molecular Biology*™ Humana Press vol 255 (1):171.
- Roeder RG. (1996). "The role of general initiation factors in transcription by RNA polymerase II" *Trends In Biochemical Sciences* vol. 21(9): 327-335.
- Roessler E, Hu P, Hong SK, Srivastava K, Carrington B, Sood R, Petrykowska H, Elnitski L, Ribeiro LA, Richieri-Costa A, Feldman B, Odenwald WF, Muenke M. (2012). "Unique alterations of an ultraconserved non-coding element in the 3'UTR of ZIC2 in holoprosencephaly" *PLoS One* vol.7(7):e39026. doi: 10.1371/journal.pone.0039026.
- Rohwer F, Seguritan V, Choi DH, Segall AM, Azam F. (2001). "Production of shotgun libraries using random amplification" *Biotechniques* vol.31(1):108-12, 114-6, 118. doi: 10.2144/01311rr02. PMID: 11464504.
- Rubenstein JLR, Beachy PA. (1998). "Patterning of the embryonic forebrain" *Current Opinion in Neurobiology* vol.8(1):18-26.
- Sagai T, Amano T, Maeno A, Ajima R, Shiroishi T. (2019). "SHH signaling mediated by a prechordal and brain enhancer controls forebrain organization" *Proc Natl Acad Sci U S A* vol.116(47):23636-23642. doi: 10.1073/pnas.1901732116.
- Sambrook J, Russell DW, Eds. (2012). "Molecular Cloning: A Laboratory Manual. 4th ed" Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- Sasaki H, Nishizaki Y, Hui C, Nakafuku M, Kondoh H. (1999). "Regulation of Gli2 and Gli3 activities by an amino-terminal repression domain: implication of Gli2 and Gli3 as primary mediators of Shh signaling" *Development* vol.126(17):3915-24.
- Schoenfelder S, Fraser P. (2019). "Long-range enhancer–promoter contacts in gene expression control" *Nat Rev Genet* vol.20: 437–455. <https://doi.org/10.1038/s41576-019-0128-0>.
- Schutte BC, Ranade K, Pruessner J, Dracopoli N. (1997). "Optimized conditions for cloning PCR products into an XcmI T-vector" *Biotechniques* vol.22: 40–44.
- Shen J, Walsh CA. (2005). "Targeted disruption of Tgif, the mouse ortholog of a human holoprosencephaly gene, does not result in holoprosencephaly in mice" *Mol Cell Biol* vol.25(9):3639-47.
- Shen Y. et al. (2012). "A map of the cis-regulatory sequences in the mouse genome" *Nature* vol.488, 116–120.
- Shimamura K, Rubenstein JLR. (1997). "Inductive interactions direct early regionalization of the mouse forebrain" *Development* vol.124:2709-2718.
- Shizuya H, Birren B, Kim UJ, Mancino V, Slepak T, Tachiiri Y, Simon M. (1992). "Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector" *Proc. Natl. Acad. Sci.* vol.89: 8794–8797.
- Shizuya H, Birren B, Kim UJ, Mancino V, Slepak T, Tachiiri Y, Simon M. (1992). "Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector" *Proc Natl Acad Sci U S A* vol.89(18):8794-7. doi: 10.1073/pnas.89.18.8794.

- Shlyueva D, Stampfel G, Stark A. (2014). "Transcriptional enhancers: from properties to genome-wide predictions" *Nature Reviews Genetics* vol.15: 272–286.
- Sigafoos AN, Paradise BD, Fernandez-Zapico ME. (2021). "Hedgehog/GLI Signaling Pathway: Transduction, Regulation, and Implications for Disease" *Cancers (Basel)* vol.13(14):3410.
- Sinn R, Wittbrodt J. (2013). "An eye on eye development" *Mechanisms of Development* vol.130 (6–8): 347-358.
- Song L, Zhang Z, Gräsfeder LL, Boyle AP, Giresi PG, Lee BK., Sheffield NC, Gräf S, Huss M, Keefe D, et al. (2011). "Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity" *Genome Res* vol.21: 1757–1767.
- Spemann H, Mangold H. (1924). "über Induktion von Embryonalanlagen durch Implantation artfremder Organisatoren" *Archiv für Mikroskopische Anatomie und Entwicklungsmechanik*. 10.1007/bf02108176.
- Sun Y, Miao N, Sun T. (2019). "Detect accessible chromatin using ATAC-seq, from principle to applications" *Hereditas* vol.156: 29.
- Swanson CI, Evans NC, Barolo S. (2010). "Structural rules and complex regulatory circuitry constrain expression of a Notch- and EGFR-regulated eye enhancer" *Dev Cell* vol.18(3):359-70.
- Takahashi K, Yamanaka S. (2006). "Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors" *Cell* vol.126(4):663-76. doi: 10.1016/j.cell.2006.07.024.
- Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC, Stergachis AB, Wang H, Vernot B, et al. (2012). "The accessible chromatin landscape of the human genome" *Nature* vol.489: 75–82.
- Uchikawa M, Nishimura N, Iwafuchi-Doi M, Kondoh H. (2017). "Enhancer Analyses Using Chicken Embryo Electroporation. In: Sheng G. (eds) Avian and Reptilian Developmental Biology" *Methods in Molecular Biology* vol 1650. Humana Press, New York, NY. [https://doi.org/10.1007/978-1-4939-7216-6\\_12](https://doi.org/10.1007/978-1-4939-7216-6_12).
- Uchikawa M, Takemoto T, Kamachi Y, Kondoh H. (2004). "Efficient identification of regulatory sequences in the chicken genome by a powerful combination of embryo electroporation and genome comparison" *Mechanisms of Development* vol.121(9):1145-1158.
- Varga Z, Wegner J, Westerfield M. (1999). "Anterior movement of ventral diencephalic precursors separates the primordial eye field in the neural plate and requires cyclops" *Development* vol.126: 5533-5546.
- Vergara MN, Canto-Soler MV. (2012). "Rediscovering the chick embryo as a model to study retinal development" *Neural Dev* vol.7: 22.
- Vernimmen D, Bickmore WA. (2015). "The Hierarchy of Transcriptional Activation: From Enhancer to Promoter" *Trends Genet* vol.31(12):696-708.
- Waddington C. (1942). "CANALIZATION OF DEVELOPMENT AND THE INHERITANCE OF ACQUIRED CHARACTERS" *Nature* vol.150(3811): 563-565.
- Wang Y, Prosen DE, Mei L, Sullivan JC, Finney M, Vander Horn PB. (2004). "A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance in vitro" *Nucleic Acids Res* vol.18;32(3):1197-207. doi: 10.1093/nar/gkh271.

Wang Y, Prosen DE, Mei L, Sullivan JC, Finney M, Vander Horn PB. (2004). "A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance in vitro" *Nucleic Acids Res* vol.8;32(3):1197-207. doi: 10.1093/nar/gkh271.

Warr N, Powles-Glover N, Chappell A, Robson J, Norris D, Arkell RM. (2008). "Zic2-associated holoprosencephaly is caused by a transient defect in the organizer region during gastrulation" *Hum Mol Genet* vol.17(19):2986-96.

Weinstein, DC, Hemmati-Brivanlou A. (1999), "Neural induction" *Annual Review of Cell and Developmental Biology* vol.15: 411-433.

Williams RM, Suaka-Spengler T. (2021). "Rapid and efficient enhancer cloning and in vivo screening using the developing chick embryo" *STAR Protocols* vol.2(2): 100507.

Williamson I, Lettice LA, Hill RE, Bickmore WA. (2016). "Shh and ZRS enhancer colocalisation is specific to the zone of polarising activity" *Development* vol.143 (16): 2994–3001. doi: <https://doi.org/10.1242/dev.139188>.

Winter TC, Kennedy AM, Woodward PJ. (2015). "Holoprosencephaly: a survey of the entity, with embryology and fetal imaging" *Radiographics* vol.35 (1): 275-90. doi:10.1148/rg.351140040.

Wysocka J, Swigut T, Milne TA, Dou Y, Zhang X, Burlingame AL, Roeder RG, Brivanlou AH, Allis CD. (2005). "WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development" *Cell* vol.17; 121(6):859-72.

Xue XY, Harris WA. (2012). "Using myc genes to search for stem cells in the ciliary margin of the Xenopus retina" *Dev Neurobiol* vol.72, 475-490.

Yokoshi M, Segawa K, Fukaya T. (2020). "Visualizing the Role of Boundary Elements in Enhancer-Promoter Communication" *Molecular Cell* vol.78(2):224-235.e5.

Yu Z, Sun Y, She X, Wang Z, Chen S, Deng Z, Zhang Y, Liu Q, Liu Q, Zhao C, Li P, Liu C, Feng J, Fu H, Li G, Wu M. (2017). "SIX3, a tumor suppressor, inhibits astrocytoma tumorigenesis by transcriptional repression of AURKA/B" *J Hematol Oncol* vol.10(1):115. doi: 10.1186/s13045-017-0483-2.

Yun S, Saijoh Y, Hirokawa KE, Kopinke D, Murtaugh LC, Monuki ES, Levine EM. (2009). "Lhx2 links the intrinsic and extrinsic factors that control optic cup formation" *Development* vol.136:3895-3906. doi:10.1242/dev.041202.

Zagozewski JL, Zhang Q, Eisenstat D.D. (2014). "Genetic regulation of vertebrate eye development" *Clinical Genetics* vol.86 (5): 452-460.

Zheng X, Dumitru R, Lackford BL, Freudenberg JM, Singh AP, Archer TK, Jothi R & Hu G. (2012). "Cnot1, Cnot2, and Cnot3 Maintain Mouse and Human ESC Identity and Inhibit Extraembryonic Differentiation" *STEM CELLS* vol.30: 910-922.

Zhou F, Gou S, Xiong J, et al. (2014). "Oncogenicity of LHX2 in pancreatic ductal adenocarcinoma" *Mol Biol Rep* vol.41: 8163–8167.

Zuber ME, Gestri G, Viczian AS, Barsacchi G, Harris WA. (2003). "Specification of the vertebrate eye by a network of eye field transcription factors" *Development* vol.130(21):5155-67. doi: 10.1242/dev.00723.

