

THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e. g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

- This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
- A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
- This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
- The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
- When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

Transcriptional dynamics of the *Sonic hedgehog* gene

Katy Alexandra Graham

A thesis submitted for the degree of Doctor of Philosophy

The University of Edinburgh

2022

INSTITUTE OF GENETICS & CANCER

Declaration

I hereby declare that this thesis has been composed by me for the degree of PhD at the University of Edinburgh. All work was performed by me, unless otherwise stated in the text, and all sources of information have been acknowledged in the text. The work has not been submitted for any other degree or professional qualification.

Katy Alexandra Graham

November 2022

Acknowledgements

First and foremost, I would like to thank my two incredible supervisors Bob and Laura. Bob, ‘thank you for giving me the opportunity to carry out my PhD project in your lab’. To Laura, thank you for all the endless chats, answering my ‘this may be a stupid question’ questions without getting exasperated and always knowing what I meant when I couldn’t remember the technical term for something and would opt for ‘thingamabob’ instead. Your help, guidance and (most importantly) patience both in the lab and out has been invaluable. To Bob, I was only joking, you do get a bigger thank you than that! Thank you for delaying your retirement yet again so that I could be your final ‘final’ student. I truly appreciated that your door was always open to chat and your constant enthusiasm about my work. Thank you for always challenging me, all your support and encouragement. I will miss working with you both a lot!

A big thank you must go to all the past and present members of the Hill lab and W2, many of who have slowly departed while I finished this PhD: Lauren, Portia, Paul, Lorraine, Silvia, Alison, Fiona and Carlo. Thank you for making the lab such a friendly place to work and for all your help over the years. I also need to thank Iain for all his help with DNA-FISH. Matt, Laura and James from imaging also need a huge thank you for all the training and help during this project. I also want to give a huge thank you to Hannes Becher for all his help with the stats.

To my friends from cohort: Jenny, Kirsty, Veronica and ‘mystery’ Alex. Thank you for all the lunch time chats and laughs, the nights out and just generally attempting to keep each other sane during our PhDs. We’ve certainly had quite a time of it, I don’t think any of us could have foreseen having to contend with a pandemic on top of the usual stress of a PhD. That being

said, I couldn't have wished for a better bunch to have been through it with. Your friendship, laughter and support has been amazing.

To Pinja, I really didn't know what a best friend was until I met you. Thank you for just being you and always being there for me. I'm sorry I decided not to move to London, but I promise I'll visit more! To James, Craig, Andrew, Bernadette, Ross and Sean: thank you for becoming a second family to me these last few years. My life has become brighter and even more fun with you all in it.

To my very large, loud and crazy extended family: Granny, Nana, Jacqueline & Robert, Sandra & Stuart, Angus & Sorrelle, Christie, Olivia, Georgia, Arianna, Elouisa, Ilaina and 'Baby' Robert (I have named you in order of age before you start any arguments about favouritism!). Thank you for all your support, laughs and love, I'm truly lucky to have you all. To my Papa, who didn't quite get to see me finish this, thank you for always encouraging my pursuit of education while simultaneously teaching me to love a party – I miss you every day.

To my siblings both old and new: Sophie, Andrew, Jack and Megan. Getting to grow up with you as children and now as adults has been one of life's privileges. Sophie, thank you for always making sure I had a healthy dose of competition. Andrew, thank you for always making me laugh. Jack, thank you for always making everyone feel like they are the most interesting person in the room (and for putting up with Sophie). Megan, thank you for becoming one of my closest friends and confidantes even though there's nothing you hate more than a big sisssssster (and for putting up with Andrew). I want to thank you all for all the fun times, and your support in the hard times. I love

you all dearly. Also, to my furry, four-legged siblings: Holly, Cookie and Bumble, thank you for all the kisses, cuddles and walks.

Last but certainly not least, to my mum and dad, Jeananne and Gilly. Thank you for your unwavering belief in me, even when my belief in myself falters. You have spent my whole life going above and beyond for all of us. I would not be where I am today or achieved what I have if it weren't for both of you. I don't think I'll ever be able to put into words how much I love you both and how grateful I am to be your daughter. Thank you for everything.

Abstract

Enhancers are capable of driving gene expression over linearly vast distances, allowing precise patterns of spatiotemporal gene expression. They are able to do this independent of orientation to the promoter, and a single gene often has multiple enhancers. There is still limited understanding of how developmental enhancers drive transcription. It must be a highly regulated process, previous evidence has shown that alterations in expression levels can result in developmental malformations. Furthermore, there is debate surrounding the mechanisms of how enhancers interact with their distal promoters. The models currently most popular in the field are looping and transcriptional hubs.

Sonic hedgehog (Shh) gene expression is a good model to further our understanding of both developmental transcriptional regulation and distal enhancer-promoter interactions. *Shh* expression is regulated by many tissue and spatial specific enhancers with, in some instances, single enhancers driving expression in single embryonic domains. The enhancers are all located within a single TAD, at a range of distances from the *Shh* promoter. Over the years, my lab has taken a special interest in the limb enhancer, ZRS. The ZRS drives transcription in the distal posterior mesenchyme of the developing limb bud in a domain called the ZPA. We have been able to identify a network of activator and repressor binding sites within the ZRS that restricts transcription in the absence of histological boundaries providing an interesting model for me to explore mechanisms for how a developmental enhancer drives transcription.

Throughout this thesis I will address three main aims. Firstly, I will establish transcriptional characteristics at the wild-type *Shh* locus. Before I start

exploring how an enhancer drives transcription using different mouse models, I first need a strong understanding of what transcription looks like in wild-type animals. I addressed this using nascent RNA-FISH. Using nascent RNA-FISH I have been able to determine the bursting frequency of *Shh* in the ZPA. Furthermore, this technique has allowed me to ascertain if an active enhancer can be transcribed through. Meanwhile, the use of RNAscope has allowed me to establish the overall pattern of expression across the ZPA. Determining whether there is a clear-cut boundary or a gradient type pattern.

Secondly, I will decipher the role of discrete functional elements of the ZRS in transcription. Previous work from members of my lab has identified different binding sites located throughout the ZRS. For example, there are four known Hox binding sites. Mutations in these sites are known to cause down-regulation of *Shh.* I used mutants for these sites to determine the action of HOXD proteins on the ZRS and how this impacts transcription. Furthermore, I have investigated how pioneer factor binding of the ZRS influences transcriptional characteristics. This was done by investigating how the mutation of a Lim homeodomain binding site impacted transcriptional characteristics of *Shh*. To contrast, I then explored how up-regulatory mutations of the ZRS effect transcription by using mice where a ZRS repressor site, the WMS, was disrupted. This work revealed that HOXD proteins, Limb homeodomain proteins and proteins binding the WMS all have different influences on *Shh* transcription, revealing a range of different roles.

Finally, I will explore long-range regulation of distal enhancer-promoter interactions. There have been multiple models proposed to explain how enhancers interact with their target promoters. Two of these models for explaining long range regulation are the looping model and the transcriptional hub model. These two models can be differentiated by the action of a single enhancer on multiple promoters, the looping model predicts promoter choice

while the transcriptional hub predicts simultaneous promoter activation. To test these models, I looked at the ability of *Shh* enhancers to drive transcription of multiple promoters in different contexts. Firstly, I used a mouse line carrying a LacZ reporter integrated within the *Shh* TAD. This provided a second internal promoter where expression is driven by *Shh* enhancers in their cognate tissues. Secondly, I performed experiments examining activation of two endogenous genes in adjacent TADs, *Shh* and *Mnx1.*

Lay Summary

When a gene is turned “on” it goes through a process called transcription that ultimately produces the protein that the gene codes for. An enhancer is a switch in our DNA that controls where and when a gene is used in our body and how much of the protein is made; thus effectively controlling gene output. It was once believed that the more complex an organism is, the more genes it must have in its DNA. However, with the advent of high throughput, next-generation DNA sequencing technology, this notion does not hold up. The human genome contains approximately 20,000 genes, a number which does not differ greatly from a well-studied, but much simpler organism called *C. elegans*, a roundworm. One mechanism that separates organisms in complexity is how genes are controlled. This is the role played by enhancers. The process by which enhancers switch genes “on” is complex and there is still much of the process that we do not understand. The main aim of my thesis is to explore some of the mechanisms used by enhancers to control genes.

Sonic hedgehog (*Shh*) is a gene used in the embryo that is crucial for normal development of the brain, face, gut, heart and limbs. It is crucial in many different organs and has many different enhancers, each responsible for switching the gene on in different tissues at different times. The enhancers are located at a variety of distances along the DNA from the *Shh* gene; some extremely close by, others long distances away. These factors make *Shh* and its enhancers a good model for studying how enhancers switch genes on, especially from great distances.

Currently there are a few different proposed mechanisms for how enhancers control their target genes and switch them on over vast distances. In my

thesis I find firstly, that the Shh gene has a defined periodicity; that is, the gene is on for short bursts followed by a quiescent period such that this on-off rate determines the amount of protein made. The enhancer controls this on-off rate. Secondly, I investigate how this on-off rate defined by the enhancer, is conveyed to the gene and I discuss several mechanisms highlighted by my studies.

List of figures

Figure 1.1. Levels of chromatin organisation 7
Figure 1.2 Cohesin-driven loop extrusion 12
Figure 1.3. Transcriptional bursting 17
Figure 1.4. Mechanisms of enhancer promoter interaction 32
Figure 1.5 SHH signalling pathway 36
Figure 1.6. Shh regulatory locus 39
Figure 1.7. Shh regulation in the developing limb bud 46
Figure 2.1 Schematic of genes probed in RNA-FISH and DNA-FISH 57
Figure 3.1. RNA FISH set up and controls 72
Figure 3.2 Burst frequencies of Shh, 5' Lmbr1 and ZRS by nascent RNA-FISH of ZPA and floorplate tissue. 80
Figure 3.3 Expression of homologous Shh alleles in ZPA and floorplate tissue at E10.5 and E11.5 85
Figure 4.1 Hox site deletion results in decreased bursting frequency in the ZPA 102
Figure 4.2 Phenotypic investigation of mutating LhxA and LhxB sites in the ZRS 106
Figure 4.3 Mutation of LhxB binding site in the ZRS results in reduced burst frequency in the ZPA 112
Figure 4.4 Adjacent insertion of two adenosines (+AA) to the WMS repressor site results in increased Shh expression. 117
Figure 5.1. The LHED transposon system and SBLac526 mice 131
Figure 5.2. RNA-FISH reveals that Shh enhancers rarely co-activate Shh and a lacZ reporter gene 138
Figure 5.3. RNA-FISH of ∆35 and WT littermate embryonic tissue shows ZRS can co-activate Shh and Mnx1 144
Figure 5.4. DNA-FISH of WT and ∆35 tissues suggests a transcriptional hub for enhancer-promoter interaction 148
Figure 6.1 Cis expression of genes with no shared regulatory elements 162
Figure 6.2 Co-expression of genes outside of the Shh TAD 165
Figure 6.3. Model for co-ordinated expression across a TAD 171

List of tables

Table 2.1 General Reagents 51
Table 2.2. Genotyping thermal cycler programme..... 53
Table 2.3. Primer sequences for genotyping..... 53
Table 2.4 RNA-FISH probe locations in the mouse genome. Note all probe sets are a pool of 48 unique 20 or 22-mer fluorescently tagged oligo sets... 55
Table 2.5. Details of fosmid probes..... 59
Table 2.6. Antibody details..... 62
Table 2.7. RNAscope probe details..... 65

Abbreviations

AER	Apical ectodermal ridge
ATAC	Assay for Transposase-Accessible Chromatin
CNS	Central Nervous System
CRISPR	clustered regularly interspaced short palindromic repeats
CT	Chromosome territory
CTCF	CCCTC-binding factor
DISP1	Dispatched1
DNA	Deoxyribonucleic acid
EMSA	Electrophoretic Mobility Shift Assay
ESC	Embryonic stem cell
ETS	E26 transformation-specific
ETV	ETS translocation variant
FGF	Fibroblast growth factor
FISH	Fluorescent *in situ* hybridisation
GFP	Green fluorescnet protein
HOX	Homeobox protein
HWE	Hardy-Weinberg Equilibrium
LAD	Lamina-associated domain
LD	Linkage Disequilibrium
LHED	Local Hopping Enhancer Detection System
MACS1	Mammal amphibian conserved sequence 1
MFCS4	Mammal fish conserved sequence 4
MGE	Medial ganglionic emminence
PCA	Principal component analysis
PCR	Polymerase chain reaction
PPD	Preaxial polydactyly
PTCH1	Patched1
PTM	Post translational modification
RNA	Ribonucleic acid
SBE	Sonic hedgehog forebrain enhancer
SEM	Standard error of the mean
SFPE1	Sonic hedgehog floorplate enhancer 1
SFPE2	Sonic hedgehog floorplate enhancer 2
SHH	Sonic hedgehog
SMC	Structural Maintenance of Chromosomes
SMO	Smoothened
SNP	Single nucleotide polymorphism

TAD	Topologically associated domain
WMS	Werner Mesomelic Syndrome
WT	wild-type
ZFN	Zinc finger nuclease
ZPA	Zone of polarising activity
ZRS	ZPA regulatory sequence

Contents

Declaration ii
Acknowledgements iii
Abstract vii
Lay Summary x
List of figures xii
List of tables xiii
Abbreviations xiv
Chapter 1 Introduction 1
1.1. Genome organisation 1
1.1.1. Chromatin organisation 1
1.1.1.1. Chromosome territories 3
1.1.1.2. A/B compartments 4
1.1.1.3. Topologically associated domains and loops 8
1.1.2. TADs in gene regulation 13
1.2. Transcription is a discontinuous process 15
1.2.1. Transcriptional bursting and its visualisation 15
1.2.2. Regulation of transcriptional bursting 17
1.2.2.1. Proximal regulation 17
1.2.2.2. Distal regulation 18
1.3. Gene regulation during development 20
1.3.1. Overview of developmental enhancers 21
1.3.1.1. Methods of enhancer identification 24
1.3.1.2. Enhancer transcription and eRNAs 26
1.3.2. Mechanisms of long-range enhancer promoter interaction 28
1.3.2.1. Tracking and linking 29
1.3.2.2. Looping 29

1.3.2.3. Transcriptional hubs 31
1.4. *Sonic hedgehog* 33
1.4.1. *Shh* discovery 33
1.4.2. Shh signalling 34
1.4.3. Shh regulation 37
1.4.4. Shh regulation in the developing limb bud 40
1.5. Aims of PhD thesis 48
Chapter 2 Materials and Methods 51
2.1. General Reagents 51
2.2. Mouse tissue 51
2.2.1. Embryonic tissue dissection 51
2.2.2. Genotyping 52
2.2.2.1. DNA Extraction 52
1.1.3.1. Genotyping PCR 52
2.2.3. Tissue embedding and sectioning 53
2.3. RNA fluorescent *in situ* hybridisation 54
2.3.1. RNA-FISH on cryosections 54
2.3.2. RNA-FISH imaging and analysis 57
2.4. DNA fluorescent *in situ* hybridisation 58
2.4.1. Fosmid probe preparation 58
2.4.2. Fosmid probe purification 59
2.4.3. Nick Translation 60
2.4.4. 3D FISH on cryosections 61
2.4.5. Imaging and analysis of DNA-FISH 62
2.5. RNAscope of cryosections 63
2.6. Statistical Methods 65
2.6.1. Student's T-test, Mann-Whitney U test and Pearson's correlation 65

2.6.2. Hardy-Weinberg equilibrium for monoallelic expression 65
2.6.3. Testing co-ordinated expression of alleles using Linkage Disequilibrium 66
Chapter 3 Transcriptional characteristics at the wild-type *Shh* locus 67
3.1. Introduction 67
3.2. Determining the bursting characteristics of *Shh* 69
3.2.1. Tissue set up and controls for RNA FISH 69
3.2.2. *Shh* transcription occurs in bursts 74
3.2.3. ZRS is transcribed through when it is an active enhancer 75
3.2.4. Analysis of expression of homologous *Shh* alleles 81
3.3. Discussion and Conclusions 87
3.3.1. Monoallelic expression as a method of transcriptional regulation 88
3.3.2. Transcription through an active enhancer suggests a transient enhancer-promoter interaction complex 91
Chapter 4 Functional elements of the ZRS and their effects on transcription 95
4.1. Introduction 95
4.2. Down-regulatory ZRS mutations and transcription 98
4.2.1. Hox site removal causes reduction in transcriptional bursting. 98
4.2.2. Removal of a Lim homeodomain binding motif effects transcriptional bursting in the ZPA 104
4.3. Up-regulatory ZRS mutations and transcription 113
4.3.1. Inserting additional adenosines adjacent to the WMS repressor site does not affect ZPA bursting frequency 113
4.3.2. Inserting additional adenosines adjacent to the WMS repressor site causes ectopic *Shh* transcriptional bursting in the anterior of the developing limb bud 115

4.4. Discussion and conclusions 119
4.4.1. HoxD binding of the ZRS has two different actions 120
4.4.2. Interruption to Lim homeodomain binding reduces *Shh* bursting frequency 122
4.4.3. Disruption to the WMS repressor site leaves *Shh* expression in the ZPA unaffected 124
Chapter 5 Long-range regulation of distal enhancer-promoter interactions 125
5.1. Introduction 125
5.2. Action of a single enhancer on multiple target promoters 127
5.2.1. Introduction to the SBLac526 mouse line 128
5.2.2. Enhancers cannot simultaneously co-activate *Shh* and LacZ reporter gene in cognate tissues 133
5.2.3. Enhancers have a set activity level when interacting with *Shh* and *lacZ* 135
5.2.4. ZRS can simultaneously co-activate *Shh* and *Mnx1* in developing limb bud 139
5.2.5. DNA-FISH of ∆35 mouse tissues 145
5.3. Conclusions and discussion 150
5.3.1. *Shh* enhancers cannot co-activate *Shh* and *lacZ* in their cognate tissues 151
5.3.2. ZRS can co-activate *Shh* and *Mnx1* in the developing limb bud 153
5.3.3. Comparison of results 154
5.3.4. Future Directions 156
Chapter 6 Co-ordinated expression of genes across TAD boundaries 158
6.1. Introduction 158
6.2. Cis co-expression of *Shh* and *Lmbr1* 159

6.3. *C*o-expression of *Shh*, *Nom1* and *Rbm33 when in-Cis.* 163
6.4. Conclusions and Discussion 167
6.4.1. Co-ordinated expression of genes within a TAD 167
6.4.2. Housekeeping genes from different TADs are not co-ordinated with *Shh* 168
6.4.3. A potential model for co-ordinated expression across a TAD 169
Chapter 7 Concluding remarks 172
References 176

Chapter 1 Introduction

1.1. Genome organisation

1.1.1. Chromatin organisation

In humans approximately 6 billion base pairs (bp), equivalent to two meters of DNA, is compacted to fit inside a single nucleus (Nurk *et al.*, 2022). The average human nuclei are 10 μM in diameter (Sun, Shen and Yokota, 2000). This compaction must not prevent the DNA from being accessible for various cellular processes such as DNA replication and repair and for gene transcription. A gene is a segment of DNA that encodes protein which expresses a given trait such as eye or hair colour. Diploid organisms have chromosome pairs, known as homologous chromosomes which are roughly the same size and shape as each other. This results in each gene having two variants which are referred to as alleles. The location of a gene within a chromosome is known as a locus.

The basis for the formation of chromatin is the nucleosome. Composed of DNA wrapped around various histone proteins, the nucleosome results in a 'beads-on-a-string', 10 nm fibre structure (Finch and Klug, 1976). The histones form an octamer consisting of two copies of each core histone protein (H2A, H2B, H3 and H4). 147 bp of DNA then wraps around this octamer, forming a nucleosome core. This structure repeats every ~200 $\pm$ 40 bp (Luger *et al.*, 1997). In the 20-40 bp that separates each nucleosome a linker histone, H1, binds stabilising the higher order structure (Woodcock, Skoultchi and Fan, 2006).

Regions of DNA have different functions and therefore, it is necessary for varying levels of accessibility throughout, for example, enhancers must be accessible for transcription factor binding. Posttranslational modifications (PTMs) of histones, in particular their amino- and carboxy-terminal tails, is a

method by which this is controlled. Examples of such modifications are methylation, acetylation, phosphorylation, sumoylation, or ubiquitinylation (Berger, 2007; Kouzarides, 2007; Taverna *et al.*, 2007). Decompaction of nucleosomes can be driven by the acetylation of specific lysines on the histone tail, for example (Kouzarides, 2007). These alterations help to determine the activity of a given region of DNA and how it interacts with the nucleosomes. In fact, histones can even be exchanged for different histone variations that play crucial roles in processes such as transcriptional regulation and DNA repair (Kouzarides, 2007; Henikoff and Smith, 2015). One example of this would be the histone variant H2A.Z which is present in regulatory regions of DNA. It acts by counteracting heterochromatic silencing (Henikoff and Smith, 2015). Given the various PTMs and their effects on whether a region of DNA is deemed active or inactive, chromatin is arguably the basic regulatory unit of genomes.

Historically, it was believed that the 10 nm fibre of chromatin was then arranged into a 30 nm fibre (Finch and Klug, 1976; Thoma, Koller and Klug, 1979), with a variety of hierarchical topologies suggested for this structure (Wong, Victor and Mozziconacci, 2007; Wu, Bassett and Travers, 2007). However, in recent years there has been debate around whether this is the structure that supersedes the 10 nm fibre. More recent studies using cryo-electron microscopy on vitrified cells have found that there is no discernible higher order structures in the chromosome (Maeshima, Hihara and Eltsov, 2010). In fact, several studies have not been able to find evidence of the 30 nm fibre (Fussner *et al.*, 2012; Ricci *et al.*, 2015; Nozaki *et al.*, 2017). Some work has proposed that rather than the 30 nm fibre, nucleosomes are organised into domains of varying sizes (Ricci *et al.*, 2015; Nozaki *et al.*, 2017). In these instances larger domains correspond with heterochromatic regions, and smaller domains associate with regions of active transcription (Ricci *et al.*, 2015). Whereas other studies have been unable to find any higher order structure, suggesting that the nucleosomal fibre exists in a

highly disordered state (Eltsov *et al.*, 2008; Maeshima, Hihara and Eltsov, 2010; Fussner *et al.*, 2012). Therefore, given the recent evidence, it appears that chromatin may assume a disordered and more flexible structure that varies depending on genomic activity in a given region (Eltsov *et al.*, 2008; Maeshima, Hihara and Eltsov, 2010; Fussner *et al.*, 2012; Ricci *et al.*, 2015; Nozaki *et al.*, 2017).

When we begin to look at organisation in the order of multi-kilobase to megabase pairs many different methods of chromatin organisation have been identified. I will now explore some of these levels of organisation such as chromosome territories, A/B compartments, chromatin loops and topologically associated domains (TADs).

1.1.1.1. Chromosome territories

Early work by Carl Rabl and Theodor Boveri (Boveri, 1909; Rabl, 1885) identified that during interphase mammalian chromosomes take on a defined space in the nucleus which allowed minimisation of interchromosomal contact (reviewed; Cremer and Cremer, 2010). Since these early papers chromosome territories (CTs) have recently been visualised using several methods such as DNA FISH and chromosome painting (Boyle *et al.*, 2001, 2011; Bolzer *et al.*, 2005; Lieberman-Aiden *et al.*, 2009).

Evidence suggests that CTs could have a functional purpose for transcriptional activity (Figure 1(i)). Previously, it was shown that chromosomes tend to either localise at the nuclear periphery or centre, based on their gene density. Croft et al., (1999) previously showed that human chromosome 19, which is relatively small and gene rich is more likely to be found in the centre of the nucleus than chromosome 18, which is a similar size but gene poor. In fact, there is evidence of gene-rich

chromosomes localising in the centre of the nucleus, and gene-poor chromosomes in the periphery in multiple studies using various methodologies (Croft *et al.*, 1999; Boyle *et al.*, 2001; Bolzer *et al.*, 2005; Lieberman-Aiden *et al.*, 2009).

Chromatin folding into CTs act as a form of regulation of gene expression. Indeed genes embedded deep within a CT are more difficult to activate (Gibcus and Dekker, 2013). Furthermore, it has been shown that some genes such as the *Hoxb* locus decondense and extrude out of their CT in a synchronised manner with the start of gene expression (Chambeyron and Bickmore, 2004). However, there is also contrasting evidence that suggests expressing alleles outside their territory are no more likely to be transcriptionally active than alleles inside the CT (Morey, Kress and Bickmore, 2009). Therefore, it appears that CTs may play a role in regulation of transcription of some genes but not all. It has been suggested that perhaps it is not the position within a territory that is important but the mobility of chromatin. Increased mobility results in increased likelihood of a gene coming into contact with transcriptional factories and therefore transcriptional activation (Morey, Kress and Bickmore, 2009).

1.1.1.2. A/B compartments

Initial mammalian Hi-C experiments showed a checkerboard pattern, revealing a network of intra- and interchromosomal contacts within the genome (Lieberman-Aiden *et al.*, 2009; Imakaev *et al.*, 2012). It highlighted that multi-megabase regions interact with each other, even across large distances along chromosomes. Principal component analysis (PCA) of the Hi-C data identified that there are two types of chromosomal compartments: A and B (Lieberman-Aiden *et al.*, 2009) (Figure 1(ii)). The A compartments are mainly associated with euchromatin, which is decondensed, accessible and associated with active transcription (Lieberman-Aiden *et al.*, 2009).

Histones in these regions have modifications associated with active chromatin (H3K36me3) and poised chromatin (H3K27me3). Furthermore, regions have demonstrated DNase I hypersensitivity, an indicator that chromatin has lost its condensed structure making it more accessible (Lieberman-Aiden *et al.*, 2009). On the other hand, B compartments have been associated with condensed, inaccessible heterochromatin, with high frequencies of the H3K9me3 histone modification which is a silencing marker (Lieberman-Aiden *et al.*, 2009). Interestingly, findings have also suggested that B compartments are associated with Lamina Associated Domains (LADs) which suggests proximity to the periphery of the nucleus (Guelen *et al.*, 2008; Ryba *et al.*, 2010). Further work using DNA-FISH has confirmed that B compartments are associated more frequently with the nuclear periphery and active A compartments with the interior of the nucleus (Wang *et al.*, 2016).

Subsequent work suggests that A and B compartments are long-range contact patterns and newer higher resolution maps identify six types of sub-compartments (Rao *et al.*, 2014). These sub-compartments can be associated with the A and B compartments. The A sub-compartments, termed A1 and A2 are associated with activating chromatin marks such as H3K36me3, H3K79me2, H3K27ac, and contain highly expressed genes (Rao *et al.*, 2014). The B compartment has four associated sub-compartments (B1, B2, B3, B4) that all exhibit slightly different characteristics. B1 is positively associated with the H3K27me3 mark and negatively with H3K36me3, suggesting that these compartments contain facultative heterochromatin – chromatin that has the opportunity to adopt open or closed conformation (Rao *et al.*, 2014). Meanwhile B2 and B3 appear to be enriched at the nuclear lamina and lack the above histone marks (Rao *et al.*, 2014). The fourth sub-compartment, B4, is the smallest and was first identified on chromosome 19. These sub-compartments are characterised by the large numbers of KRAB-ZFN genes they contain and having both activating

chromatin marks and heterochromatin associated marks (Rao *et al.*, 2014). Presence of sub-compartmentalisation suggests that chromatin organisation into domains of similar chromatin states is possibly important for the genomic functions of these compartments.

Clearly, both A and B compartments have separate functional roles. The A compartment is indicative of active chromatin while the B compartment is indicative of repressive chromatin structure. However, it is still an active debate about whether the compartments cause genomic regions to be active/repressive or whether the compartments form because the genomic regions are active/repressive. There have been contrasting findings on the matter. Some studies have shown that the presence of a gene in a LAD results in loss of transcriptional activity (Reddy *et al.*, 2008; Akhtar *et al.*, 2013), favouring the argument that compartments influence chromatin state. However, another study demonstrated that the synthetic decompaction of a gene located in the B compartment was sufficient to induce movement into the nuclear interior, but not sufficient to activate gene expression (Therizols *et al.*, 2014). This suggests that transcription and chromatin compaction are not intrinsically linked.

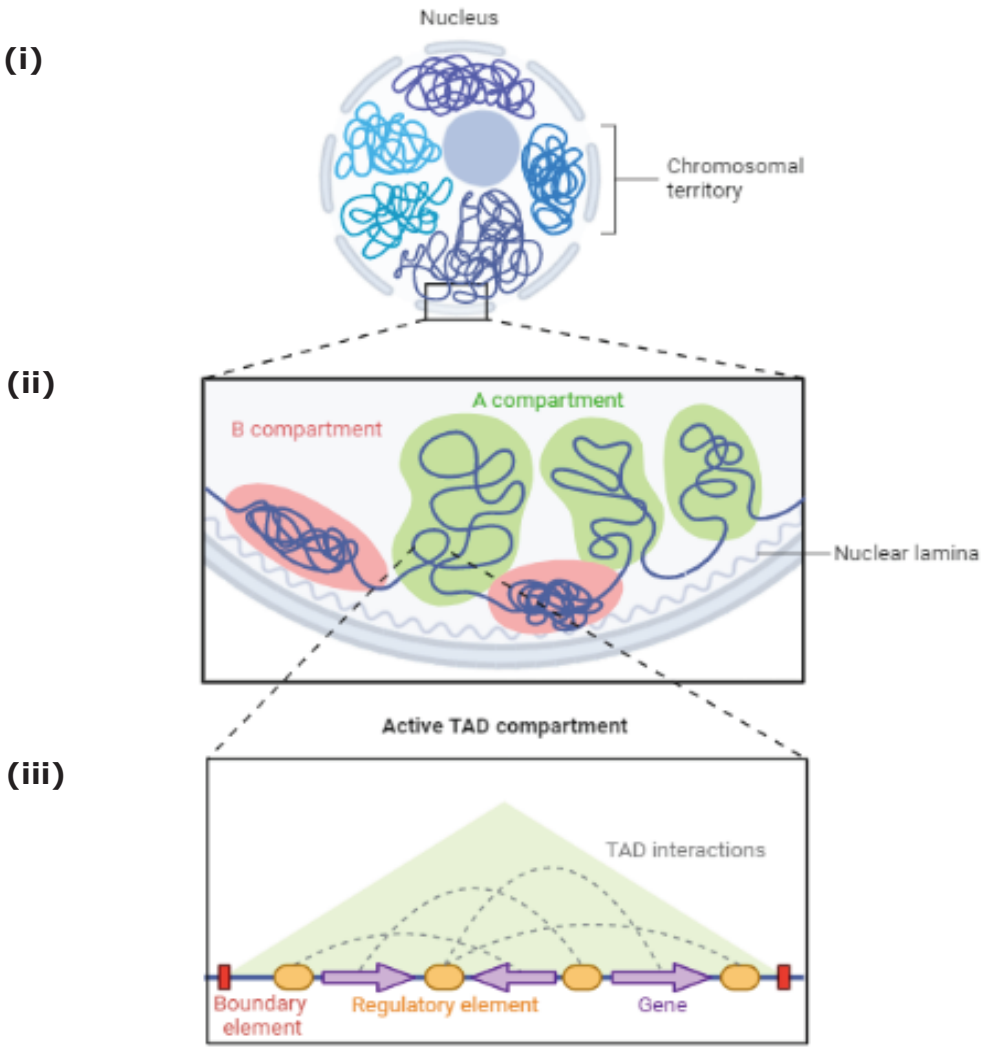


Figure 1.1. Levels of chromatin organisation

(i) Chromosomes occupy a distinct space within the nucleus, forming a chromosome territory. **(ii)** Within chromosome territories chromatin is organised into compartments of active A compartments and repressed B compartments. **(iii)** Chromatin is then organised into TADs which allow preferential interactions between genes and regulatory elements within it. Adapted from “Chromosome Organization in Nucleus: TADs”, by BioRender.com (2022). Retrieved from https://app.biorender.com/biorender-templates.

1.1.1.3. Topologically associated domains and loops

Early evidence noted the presence of ~1 Mb chromatin regions in the genome, captured by imaging 3D FISH (Albiez *et al.*, 2006) (Figure 1(iii)). More recently, the use of 5C and Hi-C to look at chromatin organisation at the sub-megabase level in human, mouse and *Drosophila melanogaster* identified self-interacting regions of chromatin (Dixon *et al.*, 2012; Hou *et al.*, 2012; Nora *et al.*, 2012; Sexton *et al.*, 2012). These regions have been termed topologically associated domains (TADs). Fluorescently labelled probes highlighted that loci interacted more frequently within TADs than between them (Nora *et al.*, 2012). Furthermore, TADs have been verified as structures in single cells via super-resolution imaging using oligopaints and single cell Hi-C (Nagano *et al.*, 2013; Flyamer *et al.*, 2017; Bintu *et al.*, 2018). Prior to these findings TADs had only been identified through population-averages of large cell numbers in Hi-C maps. Therefore, single cell methods have confirmed that TADs do exist at the single cell level. Humans and mice have ~2000 TADs in their genome, with each having a median size of ~1 Mb, covering more than 90% of the genome (Dixon *et al.*, 2012; Ea *et al.*, 2015). TADs have also been shown to be conserved across multiple cells types (Dixon *et al.*, 2012; Rao *et al.*, 2014).

TADs are separated by regions of low interaction which are considered boundaries (Nora *et al.*, 2012). The areas surrounding these boundaries have now been identified to have some defining features. TAD boundaries are typically populated by transcription start sites, housekeeping genes, tRNA genes, short interspersed nuclear elements (SINEs) and activating histone marks (Dixon *et al.*, 2012; Nora *et al.*, 2012; Sexton *et al.*, 2012). CTCF binding sites are another underlying feature of the TAD structure and are enriched at the TAD boundaries (Dixon *et al.*, 2012). CTCF site orientation is crucial to their function, with boundary sites in a convergent orientation (Sanborn *et al.*, 2015). Altering enhancer-promoter distance within

a TAD does not affect gene expression, however inversions of regions of the TAD prevent regulatory contacts (Symmons *et al.*, 2016a), highlighting the importance of CTCF site orientation. CTCF is one of the main architectural proteins responsible for organising the genome into TADs, though it should be noted that CTCF does not bind all TAD boundaries (Dixon *et al.*, 2012).

Later work using high-resolution 5C analysis has shown that TADs are divided further into sub-TADs (Phillips-Cremins *et al.*, 2013). Where TAD structure appears to be conserved between cell types (Dixon *et al.*, 2012; Rao *et al.*, 2014), sub-TADs are not (Dixon *et al.*, 2012; Phillips-Cremins *et al.*, 2013), suggesting that these cell type specific contacts are involved in gene regulation during cellular differentiation (Berlivet *et al.*, 2013; Andrey *et al.*, 2017; Bonev *et al.*, 2017). Using mouse embryonic stem cells (ESCs) one group showed that the sub-TAD structures surrounding developmentally important genes (e.g. *Oct4, Nanog, Nestin, Sox2*) changed during differentiation into neural progenitor cells (Phillips-Cremins *et al.*, 2013). Indeed another study has shown that in varying tissues different sub-TAD structures exist at the HoxA locus allowing tissue specific interaction between enhancers and gene sequences (Berlivet *et al.*, 2013). Clearly, the presence of these sub-TADs plays a role in gene regulation. Previously sub-TADs were shown to have weaker insulating properties than TAD structures (Norton *et al.*, 2018), perhaps in line with suggestions that sub-TAD structures are driven mainly by transcription factors and chromatin modifications rather than architectural proteins (Berlivet *et al.*, 2013; Andrey *et al.*, 2017; Bonev *et al.*, 2017). This would allow for a more transient structure based on gene activity, accounting for changes between cell and tissue types.

As previously mentioned TAD boundaries tend to be enriched with CTCF sites (Phillips-Cremins *et al.*, 2013), which appear to function as insulator elements. Accordingly, when a CTCF site is placed between an enhancer

and target promoter, enhancer activity becomes blocked by CTCF protein binding (Bell and Felsenfeld, 2000; Hark *et al.*, 2000). Therefore, TAD boundaries insulate against inappropriate interactions between enhancers and promoters found in different TADs (Barrington, Finn and Hadjur, 2017). Furthermore, CTCF is not the only insulating factor found at TAD boundaries. Other studies found that the structural maintenance of chromosomes (SMC) cohesin complex (Figure 2A), and constitutively active genes were enriched at TAD boundaries (Dixon *et al.*, 2012; Phillips-Cremins *et al.*, 2013; Sanborn *et al.*, 2015; Vietri Rudan *et al.*, 2015). In fact, these findings led to the proposal that TADs are formed by cohesin-driven loop extrusion (Figure 2B).

The prevailing model for explaining how TADs are formed is the cohesion-driven loop extrusion model, which relies on the enrichment of CTCF sites at TAD boundaries as well as the SMC cohesin complex (Dixon *et al.*, 2012; Phillips-Cremins *et al.*, 2013; Rao *et al.*, 2014; Vietri Rudan *et al.*, 2015). In fact, previous studies showed that most CTCF sites within the genome are occupied by cohesion (Parelho *et al.*, 2008; Rao *et al.*, 2014). Furthermore, ChIP-seq has shown that cohesin bound peaks are normally found adjacent to CTCF bound sites at TAD boundaries (Sanborn *et al.*, 2015; Fudenberg *et al.*, 2016). Hence, the loop-extrusion model proposes that a cohesin ring acts as a loop-extrusion factor which translocates along DNA forming a loop until it reaches CTCF sites, which acts as boundary elements (Fudenberg *et al.*, 2016; Barrington, Finn and Hadjur, 2017). This argument is bolstered further by the high correlation of convergent CTCF binding observed at TAD boundaries (Rao *et al.*, 2014). Further work has shown that movement of cohesin along DNA can be blocked by CTCF (Davidson *et al.*, 2016). It has also been shown that without CTCF and cohesin loop domains cannot form (Gassler *et al.*, 2017; Nora *et al.*, 2017; Rao *et al.*, 2017). Current evidence supports the loop extrusion model for TAD formation. Furthermore, it is believed the extrusion of DNA brings linearly distal promoters and enhancers into close proximity, therefore helping drive transcription.

Using oligopaints and 3D super resolution optical reconstruction microscopy (STORM) Bintu and colleagues were able to measure pairwise interactions between chromatin segments and compare to ensemble Hi-C measurements (Bintu *et al.*, 2018). Interestingly, in doing so they identified that there is variation in TAD-like structures from cell to cell by showing that chromatin regions from different ensemble sub-TADs, while showing clear segregation in some cells, show large areas of overlap in others (Bintu *et al.*, 2018). Furthermore, it was found that these TAD-like structures occurred at all cell cycle stages, with similar variation in the structures at all phases. These findings suggest that TADs are dynamic structures between individual cells, adopting varying conformational states and not static structures. Further evidence to support the argument that TAD structure is highly dynamic comes from studies where cohesin or proteins involved in cohesin loading to DNA are depleted. Several studies have shown through the use of auxin-inducible degron systems, that the depletion of cohesin results in loss of TAD structures (Rao *et al.*, 2017; Wutz *et al.*, 2017). Furthermore, depletion of the cohesin loading protein, NIPBL, similarly resulted in disruption to TAD structure (Schwarzer *et al.*, 2017). These findings suggest the need for continual loading of cohesin in order for chromatin loops and therefore, TADs to form. This again argues that TAD structure is dynamic.

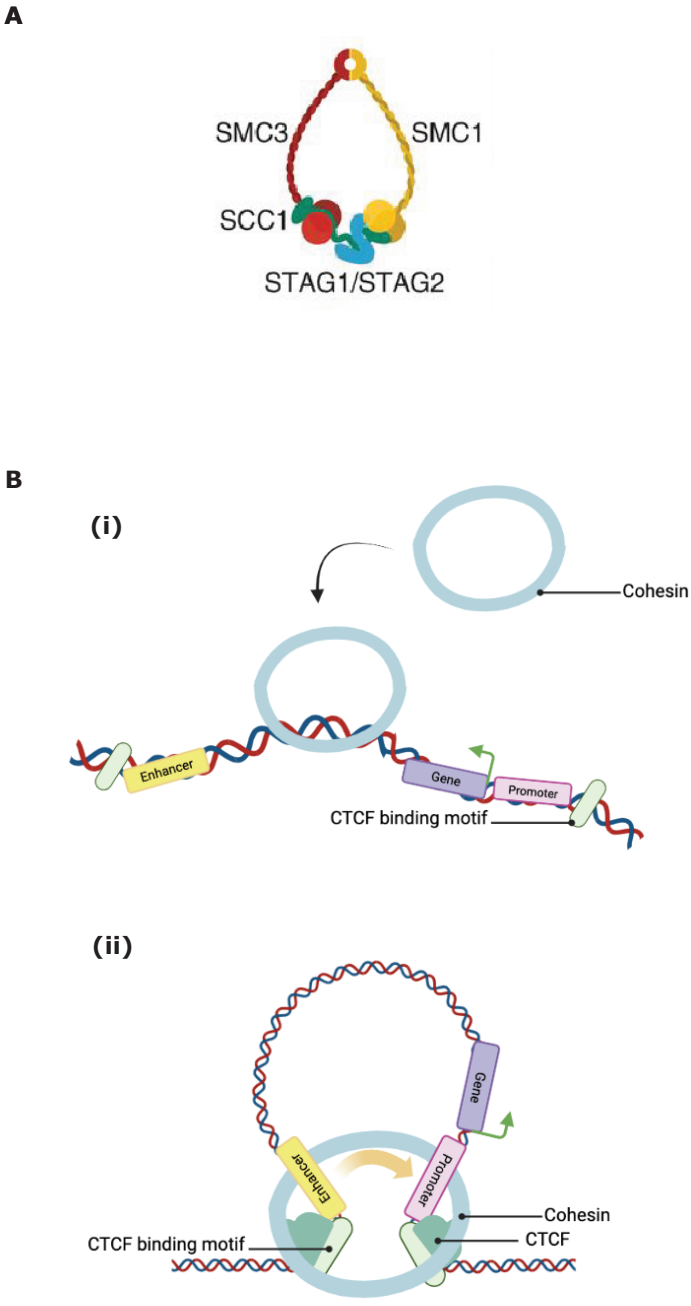


Figure 1.2 Cohesin-driven loop extrusion

A Diagram of the SMC cohesin complex highlighting the different subunits (adapted from Davidson *et al.*, 2019). **B** A simplified schematic of the loop extrusion process. (i) Cohesin is loaded on to chromatin. (ii) Cohesin then translocates across the chromatin fibre, extruding DNA out and driving loop formation until it reaches CTCF bound sites, bringing an enhancer into proximity with its target promoter. Adapted from “Transcriptional Regulation

by CTCF and Cohesin", by BioRender.com (2022). Retrieved from https://app.biorender.com/biorender-templates.

1.1.2. TADs in gene regulation

Regulatory landscapes of key developmental genes overlap with TAD structures (Andrey *et al.*, 2017). Therefore, it seems likely that TADs may play a key role in gene regulation. However, the role of TADs in gene regulation is a highly debated topic, with much contrasting evidence.

Primarily it is believed that TADs act to cause preference for local enhancer-promoter interactions and insulate against inappropriate interactions between enhancers and promoters in differing TADs. An early example of this arose in work by Lupiáñez et al. (Lupiáñez *et al.*, 2015) who were looking at the *Epha4* TAD. They systematically recreated in mouse mutations from humans found to cause limb malformations. They found that interruption of CTCF-associated boundary domains caused aberrant interactions between *Epha4* enhancers and *Wnt6,* driving ectopic expression in the limb (Lupiáñez *et al.*, 2015). Work from another group (Nora *et al.*, 2012) looking at the genes involved in X-chromosome inactivation also highlight the importance of TADs insulating aberrant interactions. *Xist* and *Tsix* are located in separate TADs with a strong boundary. When this boundary was removed there was gene misregulation and increased interactions between the two TADs (Nora *et al.*, 2012). These examples highlight the function and importance of TADs in gene regulation.

Further evidence for the role of TADs in gene regulation was demonstrated by Symmons et al (Symmons *et al.*, 2014, 2016a). They showed that an enhancer can drive expression of a reporter gene when it is located in the same TAD, however, when the enhancer was moved outside the TAD

boundary its activity was redirected to a promoter in its new TAD (Symmons *et al.*, 2014, 2016a). In fact, the act of enhancer adoption has been seen in several cases involving *Sonic hedgehog* (*Shh*). Enhancer adoption being genomic rearrangements that result in an enhancer driving expression of a non-target promoter. One inversion caused down-regulation of *Shh* in the limb at normal developmental stages, however become up-regulated at later time points suggesting it was driven by non-native enhancers resulting in severe malformations (Niedermaier *et al.*, 2005). A different study showed that genomic rearrangement resulted in limb malformation caused by *Shh* being relocated into a different TAD where an "adopted" enhancer drove *Shh* limb expression (Lettice *et al.*, 2011). These examples highlight that disruption to TAD boundaries can cause human disease and therefore demonstrates importance of TADs in gene regulation.

However, contrasting evidence has begun to emerge suggesting that in some cases TAD disruption has no real effects on gene expression. For example, increasingly large deletions were required to interrupt the TAD boundaries in the HoxD locus. The group were finally able to disrupt the TAD with a 400 kb deletion; however, they found there was no effect on limb bud development. Enhancers that drove distal and proximal limb expression were still able to act independently (Rodríguez-Carballo *et al.*, 2017). Other evidence has shown that complete depletion of CTCF sites at the *Sox9–Kcnj2* TAD *in vivo*, caused fusion of neighbouring TADs but had no effect on gene expression (Despang *et al.*, 2019). In fact, multiple studies that examine cohesin and CTCF depletion genome-wide have reported limited effects on gene expression (Haarhuis *et al.*, 2017; Nora *et al.*, 2017; Rao *et al.*, 2017; Kubo *et al.*, 2021).

A few recent studies have looked at the removal of CTCF sites from the *Shh* TAD and what effect this has. Deletion of CTCF sites within the *Shh* TAD has

found that there is redundancy within the CTCF sites (Paliou *et al.*, 2019; Williamson *et al.*, 2019a). However, there is disagreement about the effects of CTCF deletion on *Shh* expression levels. While evidence from my lab suggests that deletion of single CTCF sites results in no quantifiable change in *Shh* levels (Williamson *et al.*, 2019a). Work from others has quantified a 50% decrease in *Shh* expression, however, this was upon removal of several sites on an already sensitized background (Paliou *et al.*, 2019). Both work from our lab and others agrees that limited deletion of CTCF sites alone results in no phenotype in mice (Paliou *et al.*, 2019; Williamson *et al.*, 2019a).

It is clear that there is still much not understood about the role of TADs in gene regulation. More work is needed to further dissect the role of CTCF and cohesin in TAD formation and the role of TADs in enhancer-promoter interactions.

1.2. Transcription is a discontinuous process

1.2.1. Transcriptional bursting and its visualisation

For many years the technology available to us, such as Northern blotting, RT-PCR and microarrays made measurements based of large populations of cells. This made transcription appear to be a smooth continuous process. With the development of new technology cell to cell variation in protein and transcript levels could be observed, highlighting the importance of variation for determining cell fates (Elowitz *et al.*, 2002; Levsky *et al.*, 2002). It has now been observed in many different organisms and cell types that transcription is discontinuous occurring in “bursts” or “pulses” (Golding *et al.*, 2005; Suter *et al.*, 2011; Little, Tikhonov and Gregor, 2013; Bothma *et al.*, 2014; Chong *et al.*, 2014; Ochiai *et al.*, 2014; Bahar Halpern *et al.*, 2015). In fact, some investigations have found that bursting is the predominant mode of gene expression in mammalian cells (Dar *et al.*, 2012; Little, Tikhonov and Gregor,

2013; Bahar Halpern *et al.*, 2015). In most cases it appears that bursts of transcription are random, followed by distinct periods of inactivity (Raj *et al.*, 2006). Bursting can be split into two main parameters: burst size and burst frequency. Burst size is the number of mRNA molecules produced in a single burst. Burst frequency is how often a burst is occurring in time (Figure 3).

Experimentally, bursts have been visualised in several different ways. Methods have varied depending on whether live cells or fixed/lysed cells have been used. A first method of visualisation in live cells is gene traps with a luciferase reporter (Suter *et al.*, 2011). The main advantages of this method is that you can measure the burst frequency directly. However, this does mean that bursting characteristics are only observed at the level of protein expression. Therefore, instances where transcripts are not converted to protein or if there are quick fluctuations these will be missed. One way to overcome these disadvantages is by directly labelling transcripts with bacteriophage coat proteins such as MS2 or PP7 which are subsequently bound by fluorescent reporters such as GFP. This has been done in several instances, allowing single mRNA molecules to be visualised through microscopy (Bertrand *et al.*, 1998; Chubb *et al.*, 2006; Bensidoun *et al.*, 2016). Though it should be noted that certain assumptions need to be made using this technique; there has to be an estimate of how many GFP molecules occupy a stem loop and also an estimate of the intensity of a single GFP molecule is assumed (Chubb and Liverpool, 2010).

Most data that has been gathered by visualising bursty transcription has come from the use of bacteriophage coat proteins in lives cells and single molecule or nascent RNA-FISH (smRNA-FISH) in fixed cells. This technique allows for the visualisation of single transcripts at the cellular level at a set point in time (Femino *et al.*, 1998; Mueller *et al.*, 2013; Halpern and Itzkovitz, 2016). The main advantage of this technique is that it allows for simultaneous

detection, localization and quantification to be performed, without making the assumptions necessary for labelling transcripts in live cells with bacteriophage coat proteins. The main disadvantage of this technique is it is indirect for measuring bursting characteristics, some form of modelling must be used in tandem due to the fixed nature of the cells. However, this method does allow for bursting to be looked at in tissue, therefore gathering new insights into how bursting characteristics could impact development.

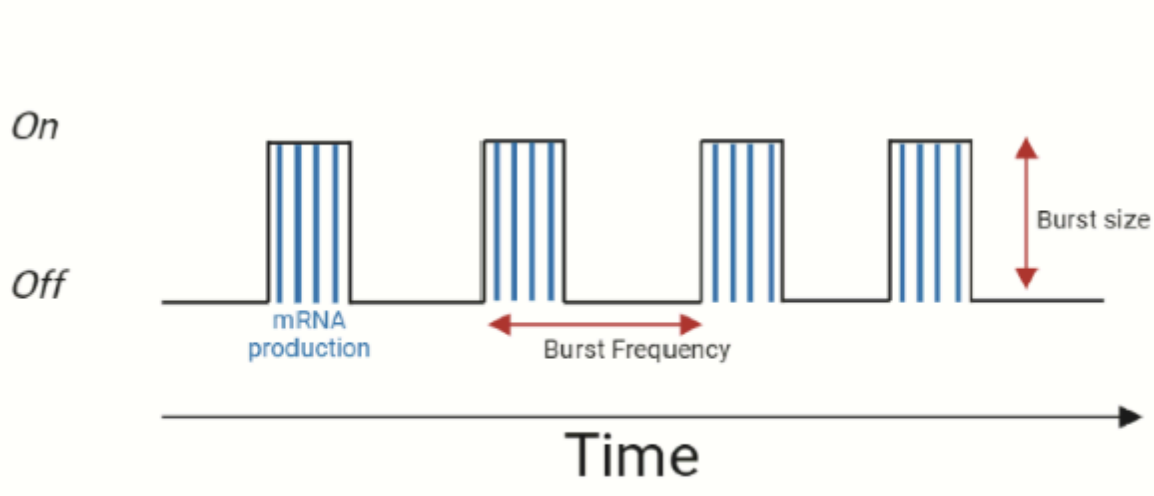


Figure 1.3. Transcriptional bursting

Transcriptional bursting can be divided into two main parameters. Burst size is determined by the number of mRNA molecules in a single burst. Burst frequency is determined by how often in time bursts are occurring.

1.2.2. Regulation of transcriptional bursting

1.2.2.1. Proximal regulation

The most obvious rate limiting factor to transcriptional regulation is the sequence-specific binding of transcriptional activators, followed by assembly

of the RNA Polymerase II (Pol II) Pre-initiation complex (PIC) at transcription start sites. This series of binding events essentially induces a transcriptional burst. Evidence suggests that bursting is an activator-centric event (Senecal *et al.*, 2014; Stavreva *et al.*, 2019). Indeed, using smRNA-FISH to look at bursting in the protooncogene *c-Fos* it has been shown that the frequency of bursts is positively correlated levels of transcription factors (Senecal *et al.*, 2014). Furthermore, another group used fluorescently-labelled MS2 imaging to determine that the burst frequency of a glucocorticoid-responsive gene positively correlated with transcription factor levels (Stavreva *et al.*, 2019). Work has also shown that the make-up of a promoter can affect both burst frequency and size. By inserting an additional CCAAT domain in transgenes to drive luciferase expression it was shown that not only did the number of mRNA molecules produced in a burst increase, but the "off" periods between bursts decreased (Suter *et al.*, 2011). This work is supported further when looking at the promoter element of an MHC class I gene. The systematic deletion of different core promoter elements found that this effected both burst size and frequency, with different elements playing differing roles (Hendy *et al.*, 2017). Interestingly, using allele sensitive single cell RNA sequencing also showed that core promoter elements affect burst size (Larsson *et al.*, 2019). Overall, these findings show that different characteristics of transcriptional bursting are influenced by transcriptional activators recruited to the transcriptional start site.

1.2.2.2. Distal regulation

Enhancers have long been shown to be crucial in gene expression in many processes such as development and circadian rhythms (Banerji, Rusconi and Schaffner, 1981; Fang *et al.*, 2014). They can often be located at far linear distances from their target promoters, in the order of megabases apart (Lettice *et al.*, 2003). There are currently two prevailing models for how enhancers communicate with their promoters, which will be discussed in more detail in section 1.3.3. In brief, the "looping" model believes the

enhancer and promoter come into close contact and form a stable complex with co-factors and Pol II (Thanos and Maniatis, 1995). The “transcriptional hub” model, states that while looping of intervening DNA does bring the enhancer and promoter into relative proximity, the structure is transient and encourages concentration of transcription factors and Pol II PIC to the surrounding environment or “hub”, with no need for contact between the enhancer and promoter (Pennacchio *et al.*, 2013; Dufourt *et al.*, 2018). If the looping model is valid, then presumably the degree of stability of the complex could lead to an increase in re-initiation events therefore having the potential to increase burst size.

Indeed, multiple studies have shown that enhancers increase burst frequency (Walters *et al.*, 1995; Bartman *et al.*, 2016; Chen *et al.*, 2018). β-globin burst frequency and size has been shown to increase during erythroid maturation. To test the bursting characteristics of the β-globin enhancer Bartman et al. (Bartman *et al.*, 2016) forced a chromatin loop between the enhancer and promoter using a zinc finger fusion protein that binds both elements. They were able to show that the β-globin enhancer predominantly encodes burst frequency. Interestingly, this approach suggests that close proximity is necessary for transcriptional activation. Furthering this finding, Chen and colleagues looked at the even skipped locus, *eve*, in *Drosophila melanogaster* using live cell MS2-PP7 imaging (Chen *et al.*, 2018). They were able to show that enhancer-promoter proximity was essential for transcriptional activation. These findings help argue for the looping model of enhancer promoter activation.

However, the need for proximity between the enhancer and promoter for transcription has not been documented in every gene. For example, using single-cell RNA sequencing and comparing WT mouse ESCs and ESCs containing a *Sox2* enhancer deletion, it was shown that the enhancer was

crucial in effecting burst frequency but had no significant effect on burst size (Larsson *et al.*, 2019). Interestingly, work from a different group (Alexander *et al.*, 2019) was able to uncouple the *Sox2* burst frequency from enhancer-promoter proximity. This supports the transcriptional hub model by showing proximity between the enhancer and promoter is not necessary to induce transcriptional bursting. Therefore, there is conflicting arguments about the role of enhancers in transcriptional bursting and how they influence transcriptional activation and further investigation is required.

1.3. Gene regulation during development

Historically, it was believed that the more complex an organism is the more genes it must have. However, evidence now contradicts this. What mediates an organism's complexity is its ability to regulate gene expression. We know now that only ~2% of the human genome is protein coding (The ENCODE Project Consortium, 2012). Therefore, it is *cis*-regulatory elements that drive the formation of the many different cell types and tissues required to form a complex multicellular organism. There are several categories of *cis*-regulatory elements such as promoters, enhancers, repressors and insulators. It is enhancers that control a gene's spatiotemporal expression, making them crucial to development. Mammalian genomes are currently estimated to contain anywhere between hundreds of thousands of enhancers to over a million (The ENCODE Project Consortium, 2012).

Below I shall define what an enhancer is and mention some key characteristics. In particular I will discuss our current understandings on how enhancers control gene expression with a focus on long-range enhancer-promoter interactions and the current debates surrounding the topic.

1.3.1. Overview of developmental enhancers

Looking at the sea urchin 2A histone gene, a “modulator” of its transcription was found upstream of the transcriptional start site which upregulates transcription of the gene but was not sufficient to illicit transcriptional initiation alone (Grosschedl and Birnstiel, 1980). This is perhaps the earliest example of an enhancer. However, key characteristics about how enhancers function were still unknown. Indeed, only the following year a 72 bp sequence motif ~200 bp upstream of the early genes of SV40 was found to be essential for transcription (Benoist and Chambon, 1981; Gruss, Dhar and Khoury, 1981). Utilising this 72 bp sequence allowed key characteristics of enhancers to come to light. By inserting this sequence upstream of other promoters, such as β-globin, it was shown that the sequence could cause upregulation of gene expression in a *cis*-acting, orientation and position independent manner (Banerji, Rusconi and Schaffner, 1981; Moreau *et al.*, 1981); thus, providing the earliest definition of an enhancer.

Some enhancers tend to be highly conserved, for example those of developmental genes (Lettice *et al.*, 2003). Strikingly, there are examples of ultra-conserved elements that are 100% conserved from rodents to humans (Bejerano *et al.*, 2004). Conservation between species can be utilised for comparative studies to help identify enhancers (Nobrega *et al.*, 2003; Woolfe *et al.*, 2004). Another interesting feature of enhancers is the long distances they can be located from their target genes. One extreme example is found among the regulators of the *SOX9* gene in which enhancers were identified 1.45 Mb away (Long *et al.*, 2020). This poses questions about the mechanisms by which enhancers and promoters can come into close proximity in 3D space. Many genes have multiple enhancers that drive precise spatiotemporal expression in a tissue specific manner. This results in complex expression patterns, especially in the case of developmental genes (Grosveld *et al.*, 1987; Lettice *et al.*, 2003; Spitz, Gonzalez and Duboule, 2003). Some estimates suggest on average each human gene has ~10-20

enhancers each (The ENCODE Project Consortium, 2012). Often for developmental genes their regulatory landscapes are located in gene deserts, which is regions of the genome that are gene poor (Visel *et al.*, 2009; Shen *et al.*, 2012). Possibly, this is so that there is space for enhancers that are driving the developmental genes.

Further research found that enhancers are composed of clusters of transcription factor binding motifs, with transcription factors binding in a sequence specific and cell type specific manner (Long, Prescott and Wysocka, 2016). These transcription factors are key in transcriptional initiation. Pioneer factors are a type of transcription factor crucial in chromatin activation (Iwafuchi-Doi and Zaret, 2016; Mayran and Drouin, 2018; Larson, Marsh and Harrison, 2021). They have the unique ability of identifying and binding their target DNA sequence in 'inactive' or 'closed' chromatin. They then recruit nucleosome remodelling complexes therefore establishing domains of accessible chromatin (Long, Prescott and Wysocka, 2016; Zaret, 2020). Their ability to do this facilitates the subsequent binding cascades of additional transcription factors which define target gene expression. Therefore, pioneer factors are at the top of the gene regulatory networks which control developmental transitions (Larson, Marsh and Harrison, 2021).

Taking in to account what we now know about enhancers, the current definitions states that an enhancer is a *cis*-acting element that upregulates expression of its target gene *in vivo* when bound by cell-type specific transcription factors and contains the associated chromatin features.

Perhaps a final important feature of enhancers to note is that they can act additively or redundantly. As mentioned above a single gene can have multiple enhancers and this is often conducive to complex expression

patterns, especially during development (Grosveld *et al.*, 1987; Lettice *et al.*, 2003; Spitz, Gonzalez and Duboule, 2003). This is enhancers behaving in an additive manner and means that if a single enhancer is lost then gene expression is affected in only the enhancers target tissue. For example, deletion of the limb enhancer for the *Sonic hedgehog* gene results in complete loss of *Shh* expression in the limb but other tissues are unaffected (Sagai *et al.*, 2005). Conversely, enhancers can often behave redundantly allowing for phenotypic robustness (Osterwalder *et al.*, 2018). Redundant enhancers are also known as secondary or shadow enhancers. They are pervasive feature of developmental regulatory networks. One study (Cannavò *et al.*, 2016) predicted over 1000 secondary enhancers in *Drosophila melanogaster*. In mice 10 different limb enhancers were deleted individually with no phenotypic effect (Osterwalder *et al.*, 2018), highlighting the importance of redundancy in developmental systems. In some instances it has been shown that redundant enhancers maintain normal expression upon removal of the predominant enhancer (Carleton, Berrett and Gertz, 2017). Seemingly, in most cases the predominant enhancer is the one located most distally from the promoter (Song, Sharan and Ovcharenko, 2019). However, it has also been shown that on a sensitized background the removal of one redundant enhancer is enough to have a phenotypic effect suggesting that these enhancers have an additive effect on expression levels (Osterwalder *et al.*, 2018). Finally, it should be mentioned that a subset of enhancers has been classified as ‘super-enhancers’. These are clusters of enhancers that are strongly associated with high occupancy of transcriptional co-activators such as Mediator (Hnisz *et al.*, 2013; Whyte *et al.*, 2013). Super enhancers are associated with cell fate decisions. There is some debate whether they truly represent a novel way to regulate gene expression (Hay *et al.*, 2016).

1.3.1.1. Methods of enhancer identification

With the discovery of enhancers came the need for methods for their identification. Initial efforts made use of extremely conserved sequences of DNA as a way of identifying putative enhancers. One study identified sequences of extreme evolutionary conservation from *Fugu* to human, and ultra-conserved sequences in human-mouse-rat (Pennacchio *et al.*, 2006). They were then able to use transgenic mouse reporter assays to demonstrate *in vivo* enhancer function of 45% of these sequences (Pennacchio *et al.*, 2006). Similarly, ultra-conserved and extremely conserved sequences between human and rodent were compared with transgenic mouse reporter assays to determine if developmental enhancers were more prevalent in either group (Visel *et al.*, 2008). It was found that developmental enhancers were equally present in both groups (Visel *et al.*, 2008).

However, despite conservation being a useful tool for identifying enhancers, it has been proposed that divergences in enhancer sequences underlie differences in transcriptional activation between species (Jubb *et al.*, 2016). Furthermore, evidence suggests that there are a substantial proportion of regulatory elements which are not conserved enough to be detected by comparative techniques (The ENCODE Project Consortium, 2007). In order for enhancers to drive transcription, they need to be accessible for TF binding and to interact with their target promoters. To do this higher order remodelling of chromatin is necessary, and technologies have been developed that utilise these changes for enhancer identification (reviewed in Gasperini, Tome and Shendure, 2020).

One alternative approach to comparative genomics is chromatin immunoprecipitation (ChIP). This technique chemically cross-links protein-DNA interactions using formaldehyde. After cross-linking has occurred

chromosomes are then fragmented and antibodies for specific proteins can be used to pull down DNA attached to this protein. One example of a this technique being used to identify enhancers has been by pulling down DNA cross-linked to the acetyltransferase and transcriptional activator, p300 (Visel *et al.*, 2009; Blow *et al.*, 2010). By pairing ChIP with sequencing (ChIP-Seq) enhancers have been able to be identified in a tissue specific manner (Visel *et al.*, 2009; Blow *et al.*, 2010). The development of other high-throughput methods such as DNase-Seq and ATAC-Seq also created new methods for enhancer identification. These techniques use DNase hypersensitivity and open chromatin, respectively, as a proxy for enhancers (Boyle *et al.*, 2008; Buenrostro *et al.*, 2015). More recently, CUT&RUN has also been developed (Skene and Henikoff, 2017). This technique overcomes high background seen in ChIP techniques as the technique is conducted in low cell numbers *in vivo*. The cells are immobilised on beads, permeabilised and incubated with antibody targeting the DNA binding protein of choice. DNA is then digested by the addition of protein A - MNase fusion protein and calcium ions. DNA is then allowed to diffuse out of the cells and the supernatant can be collected and sequenced (Skene and Henikoff, 2017).

Overall, early methods of enhancer identification involved comparative genomics. However, more recent high-throughput technologies have allowed for enhancers to be identified where they are not highly conserved, and these technologies allow for this to be done in a tissue specific manner. These new throughput technologies identify putative enhancers, with their function still needing to be shown *in vivo.* Suspected enhancers need to be confirmed *in vivo* as the chromatin accessibilities and histone marks that the above techniques are based on are not necessarily unique to enhancers.

The current gold standard for validating enhancer sequences is the enhancer reporter assay (Heintzman *et al.*, 2007; Visel *et al.*, 2009). This technique

involves a reporter construct, which contains a reporter gene, such a *lacZ* or GFP, a minimal promoter unable to drive expression of the reporter gene alone, and a putative enhancer sequence. The effect this putative enhancer sequence has on the reporter gene can then be measured in cells lines or in developing embryos. The advantage of using developing embryos is that the tissue specific and temporal function of the putative enhancer sequence can be observed (Visel *et al.*, 2007). It should be noted, however, despite this method being the gold standard it is low-throughput and time consuming.

A technique that has overcome the low-throughput shortcomings of traditional enhancer reporter assays is Massively Parallel Reporter Assays (MPRA). This technique was originally developed to identify promoter sequences (Patwardhan *et al.*, 2009), but has since been adapted for enhancer identification (reviewed in Inoue and Ahituv, 2015). Each putative enhancer sequence is cloned to drive expression of a barcoded DNA sequence under a minimal promoter, RNA sequence analysis can then be done for the barcoded sequence. However, this technique also has its disadvantages. Firstly, it is expensive and synthesising large oligonucleotide libraries is complex. Furthermore, this technique assumes that any enhancer can drive any promoter, of course, this is also true of traditional enhancer reporter assays.

1.3.1.2. Enhancer transcription and eRNAs

Despite many enhancers being located in non-coding regions of the genome, it has been shown that enhancer regions are actively transcribed (de Santa *et al.*, 2010; Kim *et al.*, 2010). These enhancer-derived RNAs have been termed eRNAs. The structure of eRNAs can vary immensely and there is much debate over their function.

eRNA production begins with the recruitment of the transcription factors and coactivators necessary for enhancer formation and gene activation (Kaikkonen *et al.*, 2013). The binding of activated transcription factors to the enhancer locus promotes nucleosome remodelling. Remodelling allows further histone modifiers to be recruited and for the locus to be opened further by acetylation of H3K27, in turn allowing for the binding of RNA Pol II (Kaikkonen *et al.*, 2013). eRNA transcription, elongation and processing then follows (Kaikkonen *et al.*, 2013). It is thought that the phosphorylation of the CTD of RNA Pol II and an abundance of poly(A) signals shortly downstream of the transcription start site are crucial for eRNA transcription (Arnold, Wells and Li, 2020). RNA Pol II found at enhancers have been shown to have higher levels of phosphorylation at the Ser5 residue of the CTD than at Ser2. This is indicative of shorter elongation times (Koch *et al.*, 2011). Prolonged elongation is likely prevented by the high abundance of poly(A) signals found immediately downstream of the enhancer transcription start site, in a bidirectional manner (Andersson *et al.*, 2014). In fact, polyadenylation of eRNAs leads to RNA Pol II instability and termination of transcription (Henriques *et al.*, 2018).

In the first instance eRNAs were believed to be bidirectional, non-polyadenylated structures of less than ~2 kb in length (Kim *et al.*, 2010). It has since been shown that eRNAs can be unidirectional and polyadenylated, with polyadenylated transcripts tending to be longer than their non-polyadenylated counterparts (Koch *et al.*, 2011). However, a genome-wide analysis has shown that eRNAs are largely short (median 346 nucleotides), bidirectionally transcribed, unspliced and non-polyadenylated (Andersson *et al.*, 2014). In fact only a small proportion of eRNAs have been shown to be long (~5 kb), polyadenylated, and specifically transcribed from active enhancers (Koch *et al.*, 2011).

Much contrasting evidence exist for the function of eRNAs. Several studies have shown that overexpression of eRNAs leads to an increase in target mRNAs (Alvarez-Dominguez *et al.*, 2017; Shii *et al.*, 2017; Jiao *et al.*, 2018). This suggests a causal role of eRNAs in transcription, rather than them being a by-product of noisy transcription. eRNAs have been implicated in the recruitment of cohesin, however, contrasting studies have suggested eRNA knockdown leaves cohesin recruitment unaffected (reviewed by Arnold, Wells and Li, 2020). By using siRNA one group was able to knockdown eRNAs and found that there was a decrease in enhancer interaction with its nearest promoter (Li *et al.*, 2013). Using RNA-pulldowns, RNA immunoprecipitation (RIP) and RIP-qPCR they were able to show that eRNAs interact with cohesin (Li *et al.*, 2013). eRNA knockdowns then showed decreased cohesin recruitment to those enhancers (Li *et al.*, 2013). eRNAs have also been implicated in 'trapping' transcription factors to increase their binding affinity to DNA (Sigova *et al.*, 2015). Sigova and colleagues showed that reduced transcription of regulatory elements resulted in loss of occupancy of transcription factor YY1 (Sigova *et al.*, 2015). Overall, these findings all implacte a functional role of eRNAs and enhancer transcription.

1.3.2. Mechanisms of long-range enhancer promoter interaction

As mentioned above, enhancers can be located linearly far away from their target promoter, as far as 1Mb or more (Lettice *et al.*, 2003; Long *et al.*, 2020). Interestingly, it has been shown that these large distances are not necessary for communication between enhancers and promoters. Enhancers located close to their promoters are still capable of driving expression of distal reporter genes (Anderson *et al.*, 2014; Symmons *et al.*, 2016b). Furthermore, large deletions that bring enhancers and promoters into closer proximity do not impact enhancer function (Williamson *et al.*, 2019a). Therefore, it remains an active debate how distal enhancers can communicate with promoters over vast distances. There have been several

models proposed to explain distal enhancer-promoter interaction. Mainly these models are: tracking, linking, looping and transcriptional hubs.

1.3.2.1. Tracking and linking

In short, tracking suggests that following the binding of transcriptional machinery to the enhancer it can then translocate along the DNA bringing the enhancer into contact with its target promoter (Furlong and Levine, 2018) (Figure 4A). Linking is an expansion of this hypothesis, proposing that protein-protein oligomers bridge the space between the enhancer and promoter (Furlong and Levine, 2018) (Figure 4B). While both these models seem feasible for a relatively proximal interaction and may play a role in distal interactions, it is unlikely they can solely account for interactions over large distances.

1.3.2.2. Looping

The two prevailing models in the field are certainly looping and transcriptional hubs. The looping model brings the promoter and enhancer close together in 3D space promoting physical contact with the intervening DNA “looping” out (Barrington, Finn and Hadjur, 2017). This model suggests direct contact between the enhancer and promoter is necessary for transcriptional activation (Figure 4C). Cohesin-driven loop extrusion is often proposed as the mechanism by which loops are formed. Several key factors contribute to this. Promoters and enhancers are often found at the base of chromatin loops, associating them with gene activation (Rao *et al.*, 2014). Furthermore, as discussed in section 1.1.1.3 genes with highly correlated expression patterns are likely to be located in the same TAD (Nora *et al.*, 2012; Dixon, Gorkin and Ren, 2016) and regulatory landscapes of key developmental genes overlap with TAD structures (Andrey *et al.*, 2017). The dynamic nature of cohesin translocating along DNA would allow enhancers and promoters to come into close contact. There is compelling evidence to suggests that

depletion of CTCF/cohesin using auxin-inducible degron systems leaves gene expression relatively unaffected (Nora *et al.*, 2017; Rao *et al.*, 2017). This suggests that cohesin-driven loop extrusion is not essential for enhancer-promoter interactions. However, cohesin-driven loop extrusion is not the only mechanism believed to induce chromatin loops. Tracking, as mentioned above, is also a possible mechanism of inducing loops in chromatin (Benabdallah and Bickmore, 2015). Though it seems unlikely this is the method used for distal enhancers and promoters. Also, CTCF sites and binding does not characterise every TAD boundary. There has also been evidence of YY1 and LDB1 being involved in the formation of loops (Schoenfelder *et al.*, 2010; Bonev *et al.*, 2017; Krivega and Dean, 2017). Therefore, cohesion-driven loop extrusion cannot be the only mechanism for enhancer-promoter interaction.

Previous studies have shown that forcing a loop between β-globin and one of its enhancers, LCR, is sufficient to induce gene expression (Deng *et al.*, 2014; Bartman *et al.*, 2016). This forced looping was done by fusing Zinc Finger (ZF) proteins designed to the β-globin promoter locus and the LCR locus with self associating domains of the Ldb1 protein (Deng *et al.*, 2014; Bartman *et al.*, 2016). The ZF proteins bind to the promoter and enhancer locus and then the interaction between the Ldb1 domains force them together. As mentioned in 1.2.2.2 this suggests proximity is necessary for transcriptional activation. Further evidence of the need for proximity for transcriptional activation comes from Chen and colleagues (2018) and has already been discussed above. Also discussed above, evidence shows that proximity is not necessary for the gene expression in the *Sox2* locus (Alexander *et al.*, 2019). Using a live-cell imaging approach where they measures the 3D distance between *Sox2* and its enhancer, Alexander *et al.* (2019) were able to show that *Sox2* transcription was not dependent on proximity between the gene and its enhancer. Clearly, evidence in support of

the looping model is not without its contradictions and therefore further research is needed to investigate distal-enhancer promoter interactions.

1.3.2.3. Transcriptional hubs

Prior to the use of live cell imaging in this field, transcription factories were another suggested model. It was believed that there were pre-existing foci (the “factories”) at fixed locations throughout the loci and that once transcription factors bind the enhancer there is mobilization of the gene to one of these points (Jackson *et al.*, 1998). These foci contained an abundance of the proteins necessary for transcription. It was subsequently shown that these foci are not fixed and assemble *de novo* (Cisse *et al.*, 2013). Transcriptional hubs model expands on this model and incorporates elements of several of the proposed models for enhancer-promoter interaction. The transcriptional hub model acknowledges a more dynamic process where pre-looped topologies act as hubs for transcription inducing proteins to accumulate (Furlong and Levine, 2018) (Figure 4D).

Unlike the looping model, the hub model dictates that the enhancer and promoter need not make contact, only that they be brought into relative proximity of each other. The hub hypothesis dictates that: 1) transcription occurs in bursts, 2) a single enhancer can co-activate transcription of linked reporter genes and, 3) distances separating enhancers and promoters during activation are proximal (~200-400nm) but not intimate (<200nm) (Lim and Levine, 2021). An example recently reported in *Drosophila melanogaster* showed that reporter genes placed equidistant on either side of an enhancer results in simultaneous co-activation of the reporter genes (Fukaya, Lim and Levine, 2016). However, despite showing that enhancers impact gene activity, as measured by burst frequency, it was not explored whether asymmetrical distance of such reporter genes from the enhancer effects co-activation (Fukaya, Lim and Levine, 2016). Co-activation is consistent with

the transcriptional hub model as it suggests there is no need for intimate contact between the enhancer and promoter to drive expression. Though it should be noted that to date co-activation by a single enhancer has only been shown with reporter genes and therefore it cannot be stated whether this phenomenon is endogenous. Large distances separating the enhancers and promoters during activation has also been shown recently (Williamson *et al.*, 2016; Benabdallah *et al.*, 2019). Again, this is consistent with the hub model as it shows no need for close proximity between enhancer and target promoter. Further research is needed to determine if a looping or transcriptional hub model explains enhancer-promoter interaction, or indeed if these two mechanisms are mutually exclusive.

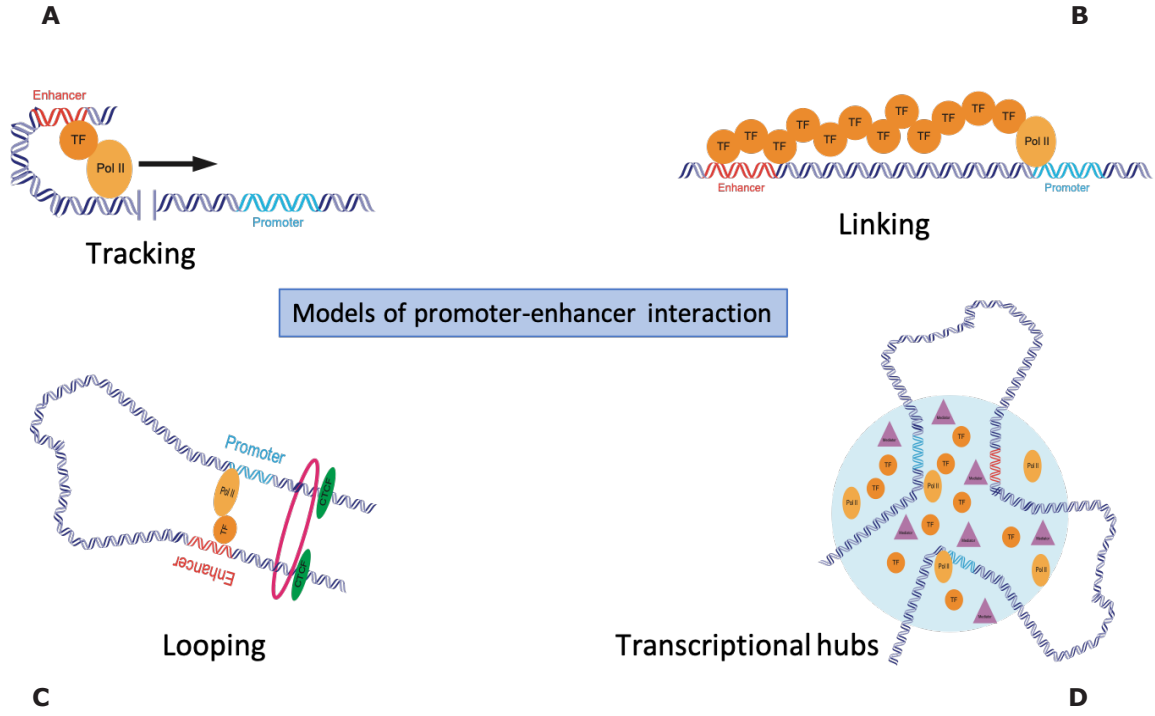


Figure 1.4. Mechanisms of enhancer promoter interaction

A Transcriptional machinery binds at the enhancer and then translocates along the intervening DNA until it reaches the target promoter. **B** Protein-protein oligomers bridge the space between the enhancer and target promoter. **C** Looping explained in detail in Figure 2. **D** Pre-looped topologies

act as hubs for transcription leading to accumulation of transcriptional proteins surrounding enhancer and target promoters.

1.4. *Sonic hedgehog*

1.4.1. *Shh* discovery

The *hedgehog* (*Hh*) gene was discovered in 1980 in *Drosophila melanogaster* (Nüsslein-Volhard and Wieschaus, 1980). Mutagenesis of *hedgehog* revealed that mutations in this gene are embryonic lethal. Drosophila have pointy projections, called denticles, on their bodies that are usually striped due to segmentation. *Hedgehog* mutants lost the gaps between stripes that made them resemble hedgehogs, giving the gene its name (Nüsslein-Volhard and Wieschaus, 1980). *Hedgehog* was shown to play a role in polarising the embryo for segmentation, therefore first suggesting its role in tissue patterning in embryonic development.

Later, three homologues were discovered in vertebrates, *Indian hedgehog (Ihh), Desert hedgehog (Dhh)* and *Sonic hedgehog (Shh)* (Echelard *et al.*, 1993; Krauss, Concordet and Ingham, 1993; Riddle *et al.*, 1993). *Dhh* has been shown to play a crucial role in spermatogenesis (Bitgood, Shen and McMahon, 1996), whereas *Ihh* is involved in skeletal morphogenesis and chondrocyte differentiation (Vortkamp *et al.*, 1996).

Shh is the most widely expressed in the developing embryo out of the three vertebrate genes (Echelard *et al.*, 1993). It is expressed in the central nervous system (CNS), the notochord, craniofacial tissue, the gut and epithelial linings and in the limb bud (Echelard *et al.*, 1993). In early experiments cells expressing *Shh* in the posterior limb bud were grafted to non-expressing tissue in the anterior limb bud (Riddle *et al.*, 1993). This was

sufficient to induce growth of extra digits, giving the earliest indication that SHH acts as a morphogen in specifying cells fates (Riddle *et al.*, 1993). A morphogen is signalling molecule whose concentration defines a cell's fate(Tabata and Takei, 2004). The idea of SHH acting as an extracellular signal was further evidenced in work by Chiang and colleagues (Chiang *et al.*, 1996). They were able to show that knocking out *Shh* was embryonic lethal and that embryonic defects stretched beyond areas where *Shh* is normally transcribed (Chiang *et al.*, 1996). In doing so they further demonstrated the idea that through its signalling pathway, SHH protein helps establish embryonic patterning.

SHH has also been proposed to have mitogenic properties. A mitogen is a molecule capable of inducing or accelerating cell division (mitosis)(Hazra, Ghosh and Hazra, 2017). By generating conditional knock-outs of *Shh* in the mouse at different timepoints Zhu and colleagues (Zhu *et al.*, 2008) suggested that Shh acts as a morphogen very early on, after this stage it is crucial for driving cell division. They found that progressively earlier *Shh* removal resulted in progressively less digits, with the remaining digits having relatively normal development (Zhu *et al.*, 2008). They suggested that this is because the earlier *Shh* is removed, the less time it has to establish a sufficient number of cells in the developing limb bud. Again, these findings highlight the importance of *Shh* and Shh signalling in embryonic patterning and development.

1.4.2. Shh signalling

In mammals the SHH signalling pathway largely occurs in the primary cilium (Murdoch and Copp, 2010). Within this pathway SHH protein is secreted as a ligand. Firstly, a full length 45 kDa precursor SHH undergoes autoproteolytic cleavage (Lee *et al.*, 1994). Cleavage results in a 25 kDa C-terminus (SHH-C) and a 19 kDa N-termius (SHH-N) (Lee *et al.*, 1994) (Figure 1.5A). It is

SHH-N domain that has signalling properties, however, cleavage of SHH-C has been shown to be crucial for covalent attachment of a cholesterol molecule to SHH-N (Porter *et al.*, 1995, 1996; Porter, Young and Beachy, 1996). Prior to being secreted from cells, the SHH-N must undergo further modification, including palmitoylation (Porter *et al.*, 1995; Pepinsky *et al.*, 1998; Hardy and Resh, 2012). Following modification the SHH-N is secreted out of the cell through transmembrane protein, Dispatched 1 (DISP1) (Burke *et al.*, 1999) (Figure 1.5). Once secreted SHH-N binds to the transmembrane protein, Patched 1 (PTCH1). PTCH1 is normally an inhibitor of the protein Smoothened (SMO). SHH-N binding to PTCH1 results in a conformational change of PTCH1 causing the release of SMO (Chen and Struhl, 1998; Murone, Rosenthal and De Sauvage, 1999) (Figure 1.5B). Once released SMO relocates to the primary cilium. The release and relocation of SMO allows for a downstream signalling cascade, eventually resulting in activation of glioma-associated (GLI) transcription factors and ultimately resulting in changes in gene expression (Chen and Struhl, 1998; Murone, Rosenthal and De Sauvage, 1999).

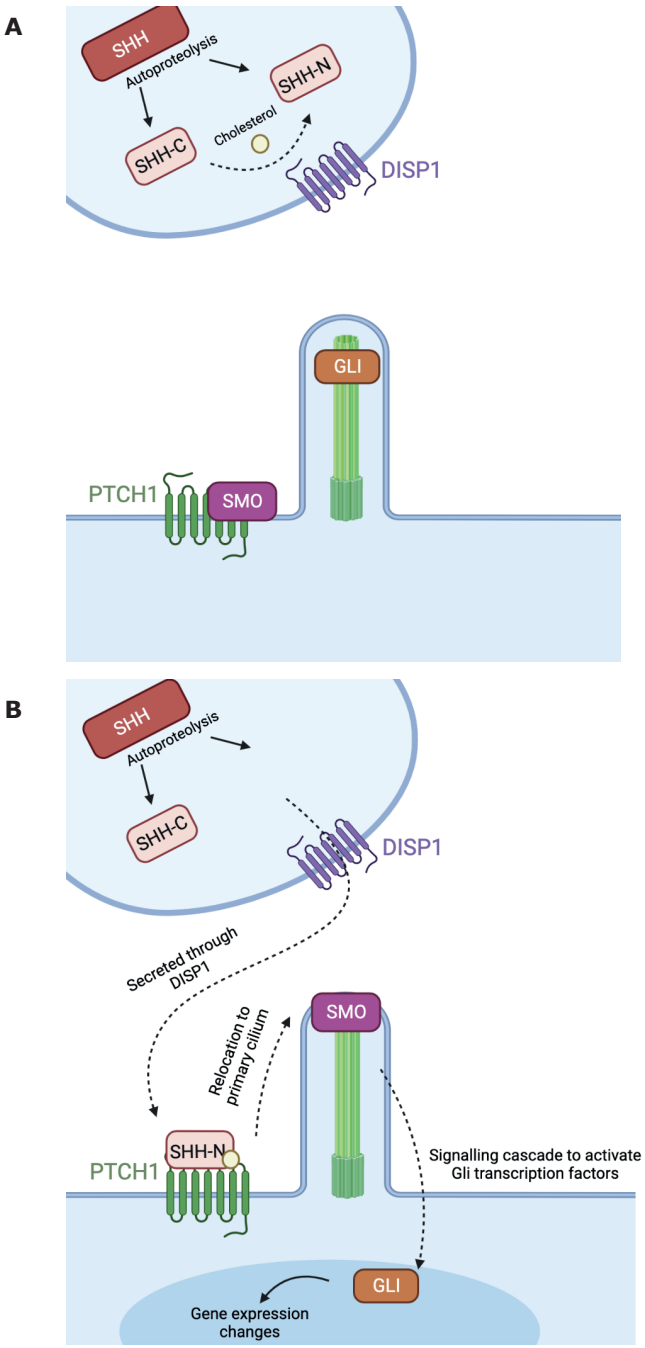


Figure 1.5 SHH signalling pathway

A SHH precursor protein undergoes autoproteolytic cleavage, resulting in SHH-C and SHH-N domains. A cholesterol molecule then covalently attaches to the SHH-N domain. **B** The modified SHH-N domain is secreted through the transmembrane protein, DISP1, and binds to PTCH1. Binding to PTCH1 results in the release of SMO which relocates to the primary cilium. Activated

SMO initiates a downstream signalling cascade that ultimately results in activation of Gli transcription factors and therefore gene expression changes.

1.4.3. Shh regulation

Shh has a vast regulatory landscape, with many tissue specific enhancers which drive its spatiotemporal expression (Anderson *et al.*, 2014) (Figure 5). This landscape is located within a single TAD that stretches ~1Mb in mouse (Williamson *et al.*, 2016). These enhancers are located within the gene itself, in an upstream gene desert and even within the intron of another gene. Spatial and temporal expression of *Shh* can be driven by enhancers from as far as 800-1000kb away from the promoter (Anderson *et al.*, 2014).

Different enhancers drive expression of *Shh* in different tissues. As previously mentioned, the main tissues where *Shh* is expressed is the CNS, gut and epithelial linings and the limb buds (Echelard *et al.*, 1993). There are currently nine known enhancers that drive expression in the CNS that we are aware of. Using transgenic reporter assays the first of these were discovered; *Shh* floorplate enhancer 1 (SFPE1), *Shh* floorplate enhancer 2 (SFPE2) and *Shh* brain enhancer 1 (SBE1) (Epstein, McMahon and Joyner, 1999). SFPE1 and SFPE2 behave redundantly and drive expression of *Shh* in the floorplate and the spinal cord, however, evidence suggests these regulate through the binding of different transcription factors (Epstein, McMahon and Joyner, 1999; Jeong *et al.*, 2006) (Figure 1.6). SFPE1 is located ~9 kb upstream of the transcriptional start site (Epstein, McMahon and Joyner, 1999). Interestingly, SBE1 and SFPE2 are located within an intron of *Shh* (Epstein, McMahon and Joyner, 1999). Currently they are the only *Shh* enhancers reported to be located within *Shh*. SBE1 is responsible for expression in the ventral midbrain and caudal diencephalon (Jeong *et al.*, 2006).

Jeong and colleagues (Jeong *et al.*, 2006) deployed an enhancer trap across the 1 Mb *Shh* regulatory locus that lead to the discovery of a cluster of brain enhancers located ~320 – 430 kb upstream of *Shh*. These enhancers are called SBE2, SBE3 and SBE4 (Figure 1.6). The expression patterns driven by these enhancers are complex and have overlapping regions. SBE2 and SBE4 both drive expression in the ventral diencephalon, however, SBE2 is the sole driver in the rostral diencephalon (Jeong *et al.*, 2006). Along with SBE3, SBE4 also drives expression in the ventral telencephalon. However, SBE3 drives only in the subventricular region of the medial ganglionic eminence (MGE) whereas, SBE4 drives in both the subventricular and ventricular region of the MGE (Jeong *et al.*, 2006).

Currently there are three other known SBEs; SBE5, SBE6 and SBE7. When looking for regions with similar motif signatures to SBE1, SBE5 was discovered (Yao *et al.*, 2016). Located in intron 15 of *Lmbr1* SBE5 drives *Shh* expression in the midbrain, showing partial redundancy with SBE1 (Yao *et al.*, 2016). Another more recent discovery is SBE6 which drives expression in the floorplate and hindbrain (Benabdallah *et al.*, 2016). Finally, SBE7 is responsible for expression in the prechordal plate (Sagai *et al.*, 2019; Crane-Smith *et al.*, 2021).

Four known enhancers, MRCS1, MFCS4, MACS1 and SLGE control *Shh* expression in the gut and epithelial linings (Sagai *et al.*, 2009) (Figure 1.6). Located in intron 8 of *Rnf32*, MACS1 drives expression in the epithelial linings of the gut including the laryngotracheal tube, stomach and lungs (Sagai *et al.*, 2009). Interestingly there is a non-conserved secondary enhancer, SLGE, that has the same expression pattern as MACS1 except in the laryngotracheal tube (Tsukiji, Amano and Shiroishi, 2014). MRCS1 and MRCS4 exhibit distinct expression patterns in the epithelial linings of the oral cavity (Sagai *et al.*, 2009).

Interestingly, *Shh* expression in the developing mouse limb bud is controlled by a single enhancer, the ZRS (Lettice *et al.*, 2003) (Figure 1.6). Regulation of *Shh* in the developing limb bud will be discussed in more detail below.

All of these enhancers are located across the *Shh* regulatory landscape and not just near the boundaries of the *Shh* TAD. The enhancers at the centre of the *Shh* TAD do not coincide in location with CTCF sites (Paliou *et al.*, 2019; Williamson *et al.*, 2019a). This poses questions as to how these internal enhancers could be regulated by the cohesin-CTCF model. More recently evidence has started to suggest that condensates and the transcriptional hub model may be involved in regulation of such *Shh* enhancers (Benabdallah *et al.*, 2019).

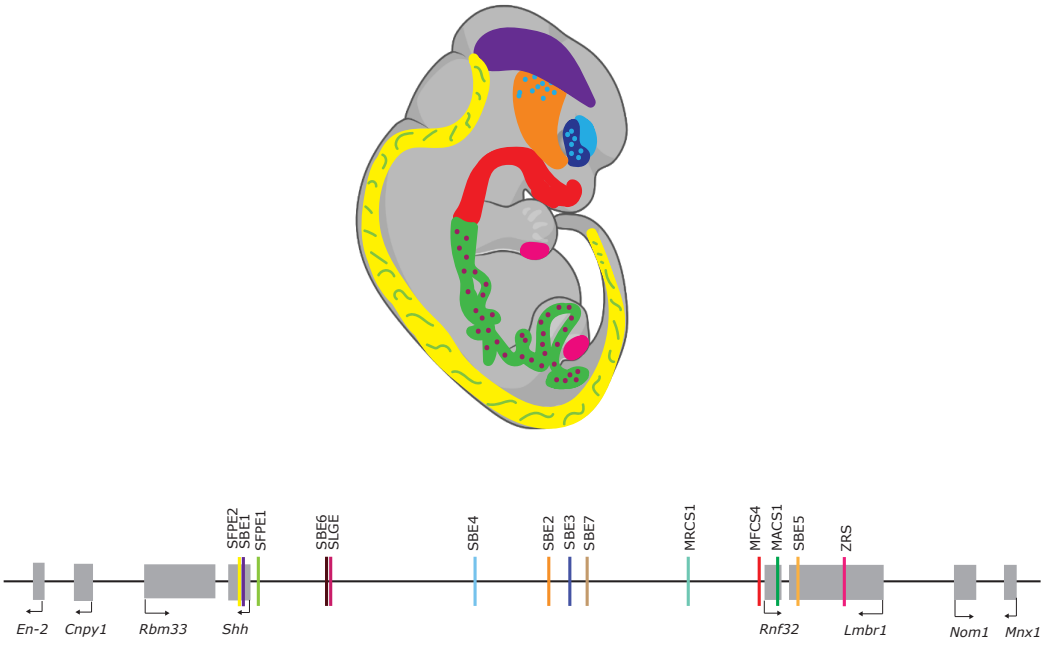


Figure 1.6. Shh regulatory locus

Shh has many tissue specific enhancers which span the gene desert between *Shh* and *Lmbr1.* Genes are shown in grey with arrows indicating

which direction they are transcribed. Enhancers are shown as various coloured bars, coordinated with the diagram of the embryo above to show where some key enhancers to this thesis drive expression.

1.4.4. Shh regulation in the developing limb bud

In the developing limb bud, *Shh* expression is restricted to an 'organising region', called the zone of polarising activity (ZPA), found in the distal posterior mesenchyme (Riddle *et al.*, 1993) (Figure 1.7A). Embryonic organisers are defined as a region of cells that can induce cell fate and define structural patterns in adjacent embryonic cells (Anderson *et al.*, 2016). A classic example is the Spemann-Mangold organiser, which is responsible for the induction of neural tissues along the anteroposterior axis during amphibian development, most self-organising development systems are probably under the influence of a local organiser (Waddington and Schmidt, 1933; Anderson *et al.*, 2016; Martinez Arias and Steventon, 2018).

The ZPA regulatory sequence (ZRS) is located in intron 5 of the ubiquitously expressed gene, *Lmbr1* (Lettice *et al.*, 2003). While *Lmbr1* is ubiquitously expressed, its function is currently unknown. Early works implicated it in limb development, giving rise to its name (Heus *et al.*, 1999; Ianakiev *et al.*, 2001). However, it is now known that these early mutations giving rise to acheiropodia and preaxial polydactyly were caused by disruption to the ZRS. Around 770bp in length, the ZRS is highly conserved retaining ~70% of its sequence from fish to mice. It is a single element which consists of several discrete functional features, including HOXD binding sites, ETS sites, a WMS repressor site and an Ebox for HAND2 binding (Lettice *et al.*, 2017) (Figure 1.7B). Mutations within the ZRS exclusively effect the limb and therefore have no effect on viability (Sagai *et al.*, 2005), making it a useful model for understanding enhancer function. These features of the ZRS shall be discussed in more detail below.

Interestingly, the ZRS has also been proposed to interact with a gene in a neighbouring TAD, *Mnx1* (Williamson *et al.*, 2019). *Mnx1* is expressed in the ZPA, suggesting shared regulation. Removal of the TAD boundary separating the ZRS and *Mnx1* results in the increased contact between the two via DNA-FISH (Williamson *et al.*, 2019). Mnx1 is key in normal motor neuronal differentiation and is primarily expressed in the neurons (von Steyern *et al.*, 1999). It currently has no known function in limb development and therefore, it is hypothesised that its expression in the ZPA is due to off-target ZRS activity.

Members from the Hill lab have identified a collection of ETS binding sites within the ZRS (Lettice *et al.*, 2012). They showed that the sites were bound by two groups of ETS transcription factors which have differing roles. Firstly, GABPα/ETS1 binding was shown to adjust the positional boundary of *Shh* expression. Whereas, ETV4/5 binding was shown to repress *Shh* expression in the anterior of the limb bud. Later work on the ZRS then identified that it has two main domains of activity. This was done by systematically dissecting the ZRS using transgenic assays and the endogenous locus (Lettice *et al.*, 2014). It was shown that the 5' end of the ZRS is crucial for spatiotemporal activity, functioning mainly in the short range. The 3' end however is instrumental to promote long-range activity. Without the 3' end the chromosomal conformational changes correlated with gene activity cannot be made.

Werner mesomelic syndrome (WMS) is characterised by preaxial polydactyly (PPD) of the hands and feet but is uniquely associated with short limb dwarfism because of tibial hypoplasia. This phenotype is linked to a 5-bp site in the ZRS, called the WMS site (Anderson *et al.*, 2012) (Figure 1.7B). Work

from the Hill lab, which introduced various mutations to the ZRS by CRISPR mutagenesis, showed that the WMS site is responsible for repression and restriction of *Shh* expression (Lettice *et al.*, 2017). Firstly, introducing a point mutation (Cu) found in a Cuban family into mice mimicked WMS. Deletion of the 5bp site also had a similar phenotype to homozygous Cu mice. In both cases *Shh* expression occurred in an ectopic site in the distal anterior limb bud resulting in preaxial polydactyly. Introducing insertions adjacent to the WMS repressor site also caused PPD. A single +A insertion caused the lengthening of first digit and occasional extra phalange in the hindlimb with no forelimb phenotype. However, by adding two additional adenosines, +AA, PPD was caused in the forelimb and extra digits occurred centrally in the hindlimbs. This all highlights the importance of the WMS repressor site for restricting *Shh* expression to the ZPA. In the same paper multiple Hox binding sites were identified and shown to determine levels of *Shh* expression (Lettice *et al.*, 2017).

The ZRS is known to contain at least four Hox sites (Figure 1.7B). These sites have been shown to bind HOXD9, 10 and 11 proteins with high affinity and HOXD12 and 13 with lower affinity (Lettice *et al.*, 2017). Previous work has shown that removal of an increasing number of these Hox sites using CRISPR technology results in increasingly severe phenotypes, ranging from partial digit loss to complete digit loss (Lettice *et al.*, 2017). Amongst Hox sites 1-3 the WMS repressor site is located. Deleting the WMS site and one Hox site restores a WT phenotype (Lettice *et al.*, 2017). This suggested that Hox site binding helps define the amount of Shh being expressed in the ZPA, as removal of a site could cancel out ectopic expression.

Using ChIP-qPCR it was identified that there was enrichment of H3K4me1, a marker of enhancer activity, across the ZRS in the distal limb bud (Peluso *et al.*, 2017). There is no *Shh* transcription in the anterior of the limb bud (Riddle

et al., 1993). However, upon further dissection of the limb it was shown that there was no discernible difference in H3K4me1 mark between the distal anterior and distal posterior limb bud (Peluso *et al.*, 2017). This was surprising and suggests that the ZRS is broadly primed as an active enhancer in tissue where it is not driving *Shh* expression. Further investigation using a limb derived cell line and short-term organ culture identified that Fibroblast Growth Factor (FGF) signalling plays a crucial role in priming the ZRS. FGFs are growth factors. Growth factors are diffusible signalling molecules that play an important role in cell growth and proliferation. Furthermore, ChIP-on-chip analysis identified GAPBα to be the activator of *Shh* expression mediating the acetylation status of H3K27. Therefore, in the ZPA the ZRS has the following histone marks H3K4me1 and H3K27ac. When FGF signalling was inhibited, GABPα was released from the ZRS, and the enhancer was no longer poised. ETV4 had previously been shown to repress *Shh* expression outside the ZPA (Lettice *et al.*, 2012). Further research has now shown that ETV4 is induced by FGF signalling from the AER(Peluso *et al.*, 2017). Therefore, FGFs are excreted from the AER, resulting in the upregulation of ETV4 which restricts *Shh* expression in the anterior of the limb bud (Figure 1.7A).

Early experiments showed that tissue grafted from the ZPA to the anterior limb bud led to additional digits being formed (Saunders and Gasseling, 1968). This led to the proposal that an unknown factor, now known to be SHH, was being secreted from the ZPA which was responsible for digit patterning (Wolpert, 1969). Therefore, Shh signalling from the ZPA is involved in defining digit number and identity.

The limb bud is initially polarized by the expression of the transcription factors GLI3 and HAND2, in the anterior and posterior of the limb bud respectively (Te Welscher *et al.*, 2002). As mentioned above FGFs from the

apical ectodermal ridge (AER) regulate *Shh* expression (Anderson *et al.*, 2012) and recently it was shown that FGFs are essential for priming the ZRS for *Shh* expression broadly across the distal limb mesenchyme (Peluso *et al.*, 2017). Evidence has suggested that 5'HOXD factors and HAND2 are crucial for *Shh* expression in the ZPA (Galli *et al.*, 2010). Members of the ETS transcription factor family then go on to establish position of the boundary of Shh expression. It is the transcription factor ETV4 which keeps the ZRS transcriptionally inactive in the anterior of the limb, while GAPBα acts as an activator of the ZRS in the ZPA (Peluso *et al.*, 2017). These interactions are crucial for specific positional expression of *Shh* in the limb bud (Figure 1.7A).

For many years it was believed that a concentration gradient of Shh is responsible for digit patterning (Wolpert, 1969). However, there is debate surrounding this with several opposing models. It has been suggested that both varying Shh concentration and Shh exposure time are necessary to specify both digit number and identity (Ahn and Joyner, 2004; Harfe *et al.*, 2004). More recently the biphasic model has been proposed. The biphasic model suggests that SHH first acts as a morphogen defining digit identity and then acts as a mitogen promoting growth (Zhu *et al.*, 2008). Data from my lab supports the SHH concentration gradient model, suggesting that digit number is dependent on *Shh* expression levels (Lettice *et al.*, 2014, 2017) (Figure 1.7B).

Restricted, focal expression of *Shh* is crucial for defining digit pattern; thus, it is important to understand how spatiospecific expression is established. There are different possibilities to answer this. *Shh* expression may respond to a single focal signal which would result in a gradient of *Shh* expression; or perhaps the genetic network for ZRS binding defines a threshold of *Shh* expression which would show a sharp, sudden drop in expression rather than a gradual decrease in expression over many cells of the ZPA. *Drosophila* gap

genes are expressed in clearly defined areas of the embryo, much like *Shh* is in the limb. Research on these genes highlights the importance of a genetic network in defining the boundaries of expression (Rivera-Pomar and Jäckle, 1996; Briscoe and Small, 2015). Therefore, current evidence supports the threshold response hypothesis, but this has not yet been formally shown or quantified in mammals.

Due to previous research characterising *Shh* and the ZRS, the system can be used as a model to explore how developmental enhancers drive expression of their target genes. Firstly, because of its well defined binding motifs it is possible to look at how the transcription factors binding these sites impact transcription. Both the overall pattern across the ZPA and bursting characteristics. ZRS is located ~850 kb from the *Shh* promoter, within a single TAD. Therefore, it can also function as an important model for understanding long-range enhancer-promoter interactions.

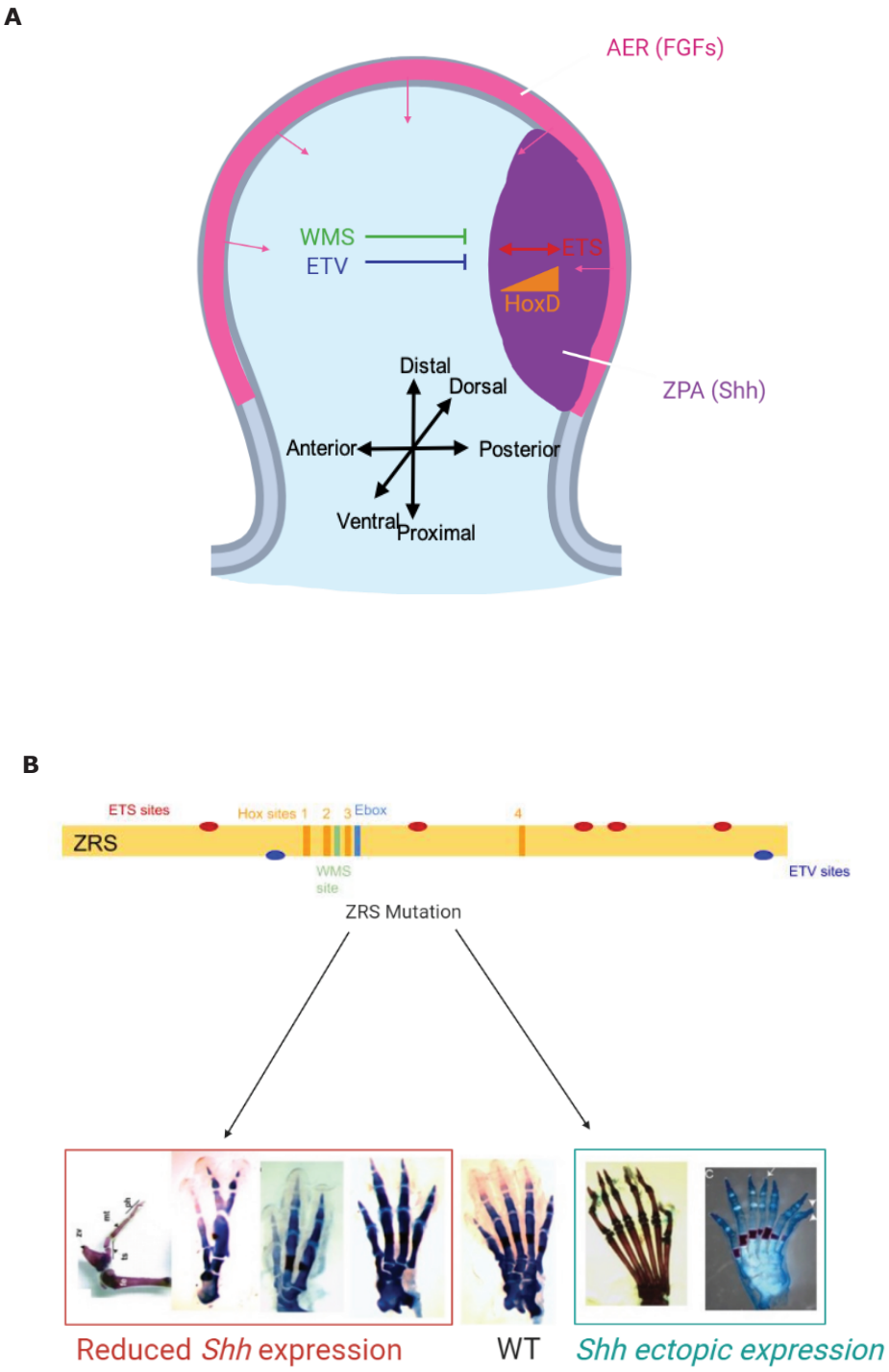


Figure 1.7. Shh regulation in the developing limb bud

A Diagram displaying the role of different transcription factors in setting up *Shh* expression in the developing limb bud. Black arrows indicate the orientation of the limb bud. **B** Diagram of the ZRS indicating various

transcription factor binding sites. Images below show limb phenotypes based on reduced or ectopic *Shh* expression.

1.5. Aims of PhD thesis

We still have limited understanding of how developmental enhancers drive transcription. It is a highly regulated process; previous evidence has shown that alterations in expression levels of some developmental genes can result in structural malformations in the embryo. Different transcriptional factors are proposed to have different roles in driving transcription, however, how they achieve this is also not understood. Furthermore, there is debate surrounding the mechanisms of how enhancers interact with their distal promoters. The models currently most popular in the field are looping and transcriptional hubs. *Shh* is a potential model to dissect these hypotheses and further our understanding of promoter-enhancer interactions. *Shh* expression is regulated by many tissue and spatial specific enhancers with, in some instances, single enhancers driving expression in single tissues. More specifically the Hill lab has characterised the limb enhancer, ZRS, to determine different transcription factor binding sites contained within its sequence. Importantly, they have established a number of animal resources that would be appropriate for studying the transcriptional dynamics of the *Shh* locus.

Therefore, this project had two main overall aims.

1. to determine how enhancers and their components drive transcription at the single cell level.
2. to explore how developmental enhancers drive transcription of their target genes over vast linear distances. I aimed to do this through the study of *Shh* and the ZRS as a model system.

To begin to address the first aim I first wanted to understand what the pattern of transcription is of *Shh* in the ZPA in WT. I did this by asking the following questions:

- Does *Shh* transcription occur in bursts?
- Can ZRS be transcribed through as part of a larger *Lmbr1* transcript, when it is actively driving *Shh* transcription?
- Is there a pattern for *Shh* transcription at allelic level (monoallelic vs biallelic)?

Once the bursting characteristics of WT *Shh* transcription was established, I then sought to understand what role the components of the ZRS played in setting up this pattern . PPD is brought about by the development of additional digit(s) anterior (thumb side) of the hand, PPD is also a characteristic of WMS. These malformations are caused by excess expression of *Shh* during development and are often caused by small or even point mutations within the ZRS. This indicates that even small alterations to the ZRS can lead to obvious phenotypes. Therefore, binding of the ZRS by different transcription factors must be specific and tightly controlled. While it is understood that mutating these binding sites results in up or down regulation of *Shh*, it is not currently understood how these TFs are determining different characteristics of transcription . More specifically I sought to understand the effect of different ZRS binding TFs on transcriptional bursting characteristics. In order to do this I addressed the following questions:

- How does Hox site binding influence transcriptional characteristics at the *Shh* locus?
- How does Lim homeodomain binding influence transcriptional characteristics at the *Shh* locus?
- How does removal of a repressor site from the ZRS influence transcriptional characteristics at the *Shh* locus?
- Does the binding of these different types of transcription factors drive transcriptional bursting of *Shh* in different ways?

My second aim was to explore enhancer promoter interaction over large linear distances, and more specifically I wanted to understand the method by which enhancers and promoters interact. To do this I asked the following questions:

- Can a single enhancer drive transcription of multiple genes simultaneously?
- Does the expression levels of an enhancer's target gene change when another target promoter is introduced?
- Does TAD structure influence co-expression of multiple genes at a similar locus?

My hypothesis was that if a transcriptional hub model were valid a single enhancer could simultaneously drive transcription of multiple genes, without affecting the enhancers activity level i.e. the transcriptional output of the multiple genes.

Throughout this thesis the primary method for answering these questions will be nascent RNA-FISH, which allows for the visualisation of transcription as a single cell level, without losing spatial context of where the cell is located.

Chapter 2 Materials and Methods

2.1. General Reagents

Table 2.1 General Reagents

Reagent	Supplier	Cat. number
Agarose	BioGene	300-300
Bovine serum albumin (BSA)	Sigma-Aldrich	A3294
Dextran sulphate	Fisher Scientific	BP1585-100
Diethyl pyrocarbonate (DEPC)	Sigma-Aldrich	D5758
Ethanol (EtOH)	Sigma-Aldrich	E/0650DF/17
Ethylenediaminetetraacetic acid (EDTA)	Fisher Scientific	E9884
Formamide	Sigma-Aldrich	G1890
Glycerol	Sigma-Aldrich	F7503
Isopropanol	Sigma-Aldrich	326961000
Paraformaldehyde (PFA)	Fisher Scientific	16005
Phenol-chloroform	Sigma-Aldrich	S77617
Proteinase K	Sigma-Aldrich	P8044
Triton X-100	Sigma-Aldrich	T8787
Tween 20	Sigma-Aldrich	P1379

2.2. Mouse tissue

2.2.1. Embryonic tissue dissection

All mouse lines described were housed in a University of Edinburgh facilities. They were stored in either the BRF (Biomedical Research facility) which is a semi-barrier mouse facility or the Transgenic Unit which is a full-barrier mouse facility. Embryos were dissected out of uteri in ice-cold PBS, washed

in PBS and then fixed in 4% PFA overnight at 4°C. They were then stored at 4°C inn PBS. Staging of embryos was done based on several criteria including eye pigmentation, approximate size of embryo and approximate shape of limb buds.

2.2.2. Genotyping

2.2.2.1. DNA Extraction

All genotyping was conducted on embryos staged between E10.5 to E11.5. Yolk sacs were placed in 200 µl lysis buffer (100 mM Tris-HCl, pH 8.5; 5 mM EDTA; 0.2% SDS; 200 mM NaCl) and 10 µg/mL proteinase K overnight at 55°C. To extract DNA, samples were briefly vortexed and equal volume of isopropanol was added. DNA was then spooled out and resuspended in 250 µl TE buffer and stored at -20°C.

1.1.3.1. Genotyping PCR

Extracted DNA was put in to PCR reactions with a final volume of 25 µL using Platinum Taq (Fisher Scientific, #10966-034), as per manufacturer's instructions. PCR was performed on a thermal cycler (Table 2.2) and reaction products were run on a 1% agarose gel for analysis of genotype. Primer sequences found in Table 2.3.

Temperature (°C)	Time (seconds)	Cycles
94	120	
94	30	35
58	30	
72	90	

Table 2.2. Genotyping thermal cycler programme

Table 2.3. Primer sequences for genotyping

Mouse Line	Forward sequence	Reverse sequence	Primer for
Δ110	GGATTAAGAGGTTAGCTCCTG	GCGCATTTGCCACTAACAC	ZRS
Δ48			
+AA			
SBLac2-67	CAACTTAATCGCCTTGCAGCAC	CGCTGATTTGTGTAGTCGGTT	LacZ
	AACACCACACCTCAGACACAT	AGTTTGTTGCAGTTTTGGGGG	2-67 Insert site
Δ35	CACATATCACACATATCCAGG	TCAGTAGAGAGTCAGTAACC	WT locus
		AATCATGATTACAGATGAGTG	Deletion

2.2.3. Tissue embedding and sectioning

Fixed embryos were transferred through a sucrose gradient: 10% sucrose in PBS until sunk to bottom of tube, 20% sucrose in PBS until sunk, 30% sucrose until sunk. Embryos were then snap frozen on dry ice in the correct orientation for sectioning, in OCT (VWR, #AGR1180). Whole embryo cryosections were then cut at 8 µm thick (Leica, #CM1950), mounted on Superfrost Plus slides (Thermo Scientific, #J1800AMNZ) and subsequently stores at -20°C short term or -80°C long term. Superfrost Plus slides contain

a permanent positive charge that allow frozen sections to be drawn to their surface electrostatically, covalent bonds then form between the slides and the formalin fixed sections. Therefore, no slide-coating procedures are necessary for the sections to adhere to the slides.

2.3. RNA fluorescent *in situ* hybridisation

2.3.1. RNA-FISH on cryosections

RNA FISH was carried out using custom Stellaris® RNA FISH probes designed against nascent *Shh* mRNA as previously described (Williamson et al., 2019), nascent *5' Lmbr1* mRNA, intron 5 of *Lmbr1* across the ZRS (called ZRS from here on), *lacZ*, nascent *Mnx1*, nascent Nom1 and nascent Rbm33 (Table 2.4). A schematic of these genes can be seen in Figure 2.1. As per manufacturer's instructions (www.biosearchtech.com/stellarisprotocols), fixed-frozen tissue sections from E11.5 embryos were washed in 1xPBS and permeabilised in 70% ethanol overnight at 4°C. Slides were washed in wash buffer (2× SSC, 10% deionised formamide). RNA FISH probes were diluted in Stellaris RNA FISH hybridisation buffer (#SMF-HB1-10) to 125 nM and hybridised to slides overnight in a 1X PBS humidified chamber at 37°C. Slides were washed twice for 30 min in wash buffer (2× SSC, 10% deionised formamide) at 37°C, counterstained with 5 ng/ml DAPI, washed in 1× PBS and mounted in Vectashield (Vector Laboratories).

For all mouse strains analysed, both floorplate and forelimb tissue were assessed. For *LhxB*^{edit} hindlimbs were also analysed. In SBLac526 mice hind gut tissue was assessed in addition to the floorplate and forelimb tissue.

Table 2.4 RNA-FISH probe locations in the mouse genome. Note all probe sets are a pool of 48 unique 20 or 22-mer fluorescently tagged oligo sets.

Probe set	**Mouse genome coordinates (mm10)**	**Distance probes distributed over (bp)**
Shh	chr 5: 28461307 - 28467201	5894
ZRS	chr 5: 29313166 - 29316440	3274
5' Lmbr1	chr 5: 29440125 - 29446297	6172
Mnx1	chr 5: 29679192 - 29682620	3428
LacZ	insert site: chr 5: 29319991	3074
Nom1	chr 5: 29641402 - 29646634	5232
Rbm33	chr 5: 28522491 - 28528573	6082


Scale
Fosmids
Genes
RNA-FISH Probes
1 Mb
mm9
28,500,000
29,000,000
29,500,000
30,000,000
Shh fosmid
ZRS fosmid
Mnx1 fosmid
Insig1
En2
Rbm33
Cnpy1
Shh
Rnf32
Nom1
Ube3c
Lmbr1
Dnajb6
Mnx1

Figure 2.1 Schematic of genes probed in RNA-FISH and DNA-FISH

Schematic of the Shh locus on chromosome 5 in mice, including the genes probed in RNA-FISH and DNA-FISH (blue). The location of where fosmid probes hybridise in DNA-FISH can be seen in red. The location where RNA-FISH probes hybridise are shown in orange, located under the gene to which they hybridise.

2.3.2. RNA-FISH imaging and analysis

All RNA-FISH images were captured as follows with the exception of those in experiments for +AA mice, WT mice probed for *Nom1* and *Rbm33*, where the microscope was replaced. Slides were imaged using a Photometrics Coolsnap HQ2 CCD camera and a Zeiss AxioImager A1 fluorescence microscope with a Plan Apochromat 100x 1.4NA objective, a Nikon Intensilight Mercury based light source (Nikon UK Ltd, Kingston- on-Thames, UK) and Chroma 4 (89000 Sedat Quad ET) single excitation and emission filters with the excitation and emission filters installed in Prior motorised filter wheels. A piezoelectrically driven objective mount (PIFOC model P-721, Physik Instrumente GmbH & Co, Karlsruhe) was used to control movement in the z dimension. Images were taken of varying tissues as stated in relevant results sections. Either two or three embryos of each genotype were used, and two sections of each embryo were captured per tissue. Images of the limb were taken so that the edge of the limb bud was visible and cells with active transcription sites were distributed throughout, ensuring non-expressing limb tissue was not analysed. Images were captured as *z* stacks with step size set at 0.2μm. Hardware control, image capture and analysis were performed using Nikon Nis-Elements software. Images were deconvolved using a calculated point spread function with the constrained iterative algorithm of Volocity (PerkinElmer Inc, Waltham MA). Image analysis was carried out using the Quantitation module of Volocity. Every

nuclei in a given image was analysed in order for a burst frequency to be extrapolated.

The following: +AA mice, WT mice prob for *Nom1* and *Rbm33*, were imaged with a new microscope. Epifluorescent images were acquired using a Teledyne Photometrics Prime BSI sCMOS camera (Teledyne Photometrics Ltd, Tuscon, AZ) and a Zeiss AxioImager Z2 fluorescence microscope with a Plan Apochromat 100x 1.46NA objective lens. Fluorescence illumination was provided by an Excelitas Xylis LED light source (Excelitas Technologies Corp. 200 West Street, 4th Floor East, Waltham, MA 02451 United States) using either Zeiss (Carl Zeiss UK, Cambridge, UK) or Semrock (IDEX Health & Science, LLC, Center of Excellence, 1180 John Street, Rochester, New York 14586, United States) single band filter cubes. Image capture and deconvolution were performed using Zeiss Zen 3.5 acquisition software. Images were deconvolved using a calculated PSF in Zeiss Zen 3.5 Acquisition software (Carl Zeiss UK, Cambridge, UK). Image analysis was then carried out in Imaris (version 9.6.0). Spots were detected using Imaris spot detection feature.

2.4. DNA fluorescent *in situ* hybridisation

2.4.1. Fosmid probe preparation

The following fosmid preparation steps were carried out on my behalf by Iain Williamson. Fosmids are DNA vectors that contain the origin of replication site of *E. coli* F-plasmids. The main advantage of fosmids is that they can stably contain large DNA fragments. Fosmids were ordered from BacPac Genomics as agar stabs. E. coli were streaked onto chloramphenicol (12.5 ng/µL) agar plates and left to incubate overnight at 37 ˚C. Individual colonies were picked and incubated overnight shaking at 225 rpm in 5 mL L-broth containing chloramphenicol. Glycerol stocks of fosmid probes were made by

added 170 µL of glycerol to 830 µL fosmid culture and stored at -80°C. For details of fosmids used see Table 2.5.

Table 2.5. Details of fosmid probes

Region	Whitehead ID	Coordinates	Size (bp)
Shh	WI1-0574O18	Chr 5: 28754458-28795879	41421
ZRS	WI1-1047E14	Chr 5: 29611727-29653695	41968
Mnx1	WI1-1204B06	Chr 5: 29791124-29827491	36368

2.4.2. Fosmid probe purification

To isolate fosmid DNA a pipette tip of glycerol stock was set up in an overnight culture of 5 mL L-broth with chloramphenicol (12.5 ng/µL) shaking at 300 rpm. Cultures were then pelleted by spinning down at 4500 rpm for 5 mins, the supernatant was then removed. Pellet was resuspended in 200 µL GTE (50 mM glucose, 25 mM Tris pH 8 and 10 mM EDTA) containing lysozyme and 2 µL RNase A (20 mg/ml; Invitrogen, #12091021) and left at room temperature for 5 mins. 400 µL of lysis buffer (0.2 M NaOH, 1% SDS) was then added. Solution was mixed by inversion until clear and left on ice for 5 mins. The reaction was then stopped by adding 300 µL acetate buffer (5 M potassium acetate, 11.5% (v/v) glacial acetic acid) mixing gently and leaving on ice for 5 mins. Samples were then centrifuged at 13,500 rpm for 5 mins at 4 ˚C. Supernatant was then transferred to a fresh Eppendorf and an equal volume of phenol:chloroform was added (~700 µL). Samples were then mixed by inversion and centrifuged at 13,500 rpm for 4 mins at 4 ˚C. The top layer was then transferred to a fresh Eppendorf and equal volume of

chloroform was added. Samples were mixed well and centrifuged at maximum speed for 4 mins at 4 ˚C. Top layer was then transferred to a fresh Eppendorf and equal volume of isopropanol was added. Samples were then stored at -20 ˚C for at least one hour. Samples then centrifuged for 15 mins at 13,500 rpm at 4 ˚C, supernatant was removed, and pellet was washed in 70% ethanol. Supernatant was removed, and pellet was allowed to dry before being resuspended in 30 µL TE buffer. Fosmid miniprep DNA was quantified using the Quibit dsDNA broad range assay (ThermoFisher, #Q32850) with the Quibit 4 fluorometer and then stored at -20°C.

2.4.3. Nick Translation

Fosmid DNA was labelled in a nick translation reaction. For directly labelled probes reactions consisted of 2.5 µL dATP, dCTP and dGTP (0.5 mM, ThermoFisher, #10297018), 2 µL 10x nick translation salts (0.5M Tris pH 7.5, 0.1 M $MgSO_4$, 1 mM DTT, 0.5 mg/mL BSA fraction V (Merck, #10735086001)), 1 µL DNase I (diluted 1:5 in ice cold water, Merck, #471628001), 1 µL DNA polymerase I (ThermoFisher, #18010017) and 2.5 µL either Green496-dUTP (Enzo, #ENZ-42831) or Chromatide Alexa Fluor 594-5-dUTP (ThermoFisher, #C11400) . For digoxygenin labelled probes, 1.5 µL digoxigenin-11-dUTP (Merck, #3359247910) and 1 µL dUTP (0.5 mM) were added instead of the directly labelled dTTP. The reaction was made up to 20 µL with distilled water.

Reactions were mixed and briefly spun down to collect liquid in bottom of the Eppendorf tube, they were then stored at 16 ˚C for 90 mins. The nick translation was then stopped by adding 2 µL of 20% SDS and 3 µL 0.5 M EDTA. The final volume was then made up to 90 µL with TE buffer and reaction was spun through Quick Spin columns (Roche, #11273973001) as described by the manufacturer’s instructions.

2.4.4.3D FISH on cryosections

Slides were selected so sections were mid-way through the proximal limb bud and so that tissue from the AER was still present to allow for orientation of the posterior distal mesenchyme and locating ZPA tissue. Since embryos were mounted whole these sections also contained floorplate tissue. Slides were washed in 1X PBS for 5 mins at room temperature with gentle shaking. They were then baked at 60 ˚C before post fixation in 4% PFA at 4 ˚C for 15 mins. Tissue was then dehydrated by leaving for 5 mins each in 50%, 70% and 100% ethanol, before being left to dry for 5 mins. Slides were then boiled in 1X target retrieval (Advanced Cell Diagnostics, #322000) for 5 mins. Slides were immediately cooled by washing in distilled water twice and then rinsed twice in 100% ethanol and allowed to dry. Three drops of Protease III (Advanced Cell Diagnostics, #322337) were then placed on slides. Slides were parafilmed and placed in a humidity chamber at 40 ˚C for 30 mins. Slides were washed briefly in distilled water, twice. Then they were washed briefly in 2X SSC.

Slides were placed in denaturing solution (70% formamide/2X SSC, pH to 7.5) at 80 ˚C for 15 mins. They were then quickly transferred to ice cold 70% ethanol for 2 mins, 90% and 100% ethanol for 2 mins each and allowed to airdry. Approximately 80 – 100 ng of fosmid probe was used per slide by preparing with 6 µg CotI DNA (Invitrogen, #18440016) per fosmid and 5 µg autoclaved salmon sperm (Invitrogen, #15632011) per fosmid with 2 volumes of 100% ethanol and spin vacuuming until probe is dry. Approximately 30 µL of hybridisation mix (50% formamide, 10% dextran sulphate, 2X SSC, 1% Tween 20 in ddH2O) was then used to resuspend each fosmid probe. Probes were left to dissolve in hybridisation mix for 1 hour and then vortexed well and briefly spun down. Fosmid probes were then denatured at 80 ˚C for 5 mins before transferring to 37 ˚C for 15 mins to pre-anneal. Fosmid probe/

hybridisation mixture was then placed on each slide, which were then coverslipped, sealed with rubber solution and left at 37 ˚C in the dark overnight.

Slides were then washed four times in 2X SSC at 45 ˚C for 3 mins, four times in 0.1X SSC at 60 ˚C for 3 mins and then transferred into 4X SSC, 0.1% Tween20. For biotin and digoxigenin labelled probes, slides were incubated with blocking buffer (4X SSC with 5% Marvel) at room temperature for 5 mins. Antibodies used to detect biotin and digoxigenin labelled probes were diluted in blocking buffer and centrifuged at maximum speed for 15 mins at 4°C. Slides were incubated with antibodies for 1 hour at 37°C in a humidified chamber with three washes for 2 mins in 4X SSC containing 0.1% Tween 20 between different antibody incubations. Slides were stained with 0.5 µg/mL DAPI, mounted using Vectashield and stored at 4°C prior to imaging. Antibody details can be found in Table 2.6.

Table 2.6. Antibody details

Antibody	**Concentration**	**Product details**
Biotinylated anti-avidin	1:100	Vector labs
Strepavidin Cy5	1:10	Amersham
Texas Red anti-sheep	1:100	Abcam
Rhodamine anti-DIG FAB fragments	1:20	Roche

2.4.5. Imaging and analysis of DNA-FISH

Epifluorescent images were acquired using a Teledyne Photometrics Prime BSI sCMOS camera (Teledyne Photometrics Ltd, Tuscon, AZ) and a Zeiss

AxioImager Z2 fluorescence microscope with a Plan Apochromat 100x 1.46NA objective lens. Fluorescence illumination was provided by an Excelitas Xylis LED light source (Excelitas Technologies Corp. 200 West Street, 4th Floor East, Waltham, MA 02451 United States) using either Zeiss (Carl Zeiss UK, Cambridge, UK) or Semrock (IDEX Health & Science, LLC, Center of Excellence, 1180 John Street, Rochester, New York 14586, United States) single band filter cubes. Image capture and deconvolution were performed using Zeiss Zen 3.5 acquisition software. Images were deconvolved using a calculated PSF in Zeiss Zen 3.5 Acquisition software (Carl Zeiss UK, Cambridge, UK). Image analysis was then carried out in Imaris (version 9.6.0). Spots were detected using Imaris spot detection feature, and distances measured from the centre points of detected spots. ~90 nuclei from each tissue and embryo were analysed.

2.5. RNAscope of cryosections

Manufacturer's protocol was followed (acdbio.com). In brief, slides were washed in 1X PBS for 5 mins at room temperature with gentle shaking. They were then baked at 60 ˚C before post fixation in 4% PFA at 4 ˚C for 15 mins. Tissue was then dehydrated by leaving for 5 mins each in 50%, 70% and 100% ethanol, before being left to dry for 5 mins. Slides were then boiled in 1X target retrieval (Advanced Cell Diagnostics, #322000) for 5 mins. Slides were immediately cooled by washing in distilled water twice and then rinsed twice in 100% ethanol and allowed to dry at room temperature overnight. Three drops of Protease III (Advanced Cell Diagnostics, #322337) were then placed on slides. Slides were parafilmed and placed in a humidity chamber at 40 ˚C for 30 mins. Slides were washed briefly in distilled water, twice.

Probe solution was prepared by first warming probes to 40 ˚C for 10 mins and then allowing to cool to room temperature (probe details found in Table

2.7). C1 and C2 probes were then mixed in a 50:1 ratio, to make approximately 100 µL per slide. Probe solution was distributed on each slide and they were parafilmed. Slides were left for 2 hours at 40 ˚C in a humidity chamber before being washed twice in 1X Wash Buffer (Advanced Cell Diagnostics 310091). The following steps were conducted using the RNAScope® Fluorescent Multiplex Detection Reagents (Advanced Cell Diagnostics 320851). Approximately four drops of Amp 1-FL were placed on slides for 30 mins at 40 ˚C in a humidity chamber. Slides were washed twice in 1X Wash Buffer, then four drops of Amp 2-FL were placed on for 15 mins at 40 ˚C. Again, slides were washed twice in 1X Wash Buffer before four drops of Amp 3-FL were place and left for 30 mins at 40 ˚C. Finally, slides were washed twice in 1X Wash Buffer and Amp 4-FL (A, B or C) were placed on slides for 15 mins at 40 ˚C. Slides were washed twice before Concanavalin A, Alexa Fluor™ 647 Conjugate (Thermo Fischer Scientific C21421) was place on slides for 5 mins to stain cell membranes. Finally, this was rinsed off using distilled water and DAPI stained cells for 30 secs. Slides were mounted in ProLong Gold Antifade Mountant (Thermo Fischer Scientific P36930).

Slides were imaged using a Photometrics Coolsnap HQ2 CCD camera and a Zeiss AxioImager A1 fluorescence microscope with a Plan Apochromat 100x 1.4NA objective, a Nikon Intensilight Mercury based light source (Nikon UK Ltd, Kingston- on-Thames, UK) and Chroma 4 (89000 Sedat Quad ET) single excitation and emission filters with the excitation and emission filters installed in Prior motorised filter wheels. A piezoelectrically driven objective mount (PIFOC model P-721, Physik Instrumente GmbH & Co, Karlsruhe) was used to control movement in the z dimension.

Table 2.7. RNAscope probe details

Probe Name	Cat no.
Mm-Shh-C2	314361-C2
Mm-Ptch1	402811
B-E.coli-lacZ	538051

2.6. Statistical Methods

2.6.1. Student's T-test, Mann-Whitney U test and Pearson's correlation

Student's T-test, Mann-Whitney U test and Pearson's correlation were all calculated using the core 'stats' package in R version 3.5.1 (2018-07-02) (R Core Team, 2018).

2.6.2. Hardy-Weinberg equilibrium for monoallelic expression

To determine if monoallelic expression of genes was significant, in collaboration with Hannes Becher (Bioinformatics core statistician, IGC), we decided that HWE could be adapted. HWE is the measurement of whether two alleles segregate randomly by looking at allele frequencies within a population. Therefore, HWE is normally calculated using genotypes (AA, aA, aa). From my nascent RNA-FISH data bursting alleles can be analogous to alleles segregating in a population. When looking at bursting of homologous genes within a single cell there will be either 0, 1 or 2, signals. Therefore, AA = 2 bursts, Aa = 1 burst, aa = 0 bursts. My data for bursts present in each cell was converted into a genotype. From there the R package 'genetics' (Warnes *et al.*, 2021) could be used to calculate HWE and test it by a Chi-squared test.

2.6.3. Testing co-ordinated expression of alleles using Linkage Disequilibrium

Again, in collaboration with Hannes Becher we decided that calculations for determining Linkage Disequilibrium (LD) could be adapted to measure if two alleles burst in a co-ordinated fashion. In doing so Chi-squared could be used to determine the significance of co-ordinated bursting of genes. LD is the non-random association of two alleles at different loci (Slatkin, 2008). In this instance bursting of two genes is analogous to linkage of two alleles. To calculate if two bursting genes are linked we converted the bursting of each gene into a genotype as noted above. We could then test for LD using the R package 'genetics' (Warnes *et al.*, 2021), significance was quantified using a Chi-squared test.

Chapter 3 Transcriptional characteristics at the wild-type *Shh* locus

3.1. Introduction

There is still much that is not understood about how enhancers drive specific gene expression in a spatiotemporal manner. *Shh* is a well-studied gene with many known enhancers that drive expression in several developing tissues (Riddle *et al.*, 1993; Epstein, McMahon and Joyner, 1999; Jeong *et al.*, 2006). More specifically its limb enhancer, the ZRS, is the sole driver of *Shh* expression in the mouse limb (Lettice *et al.*, 2003; Sagai *et al.*, 2005). The ZRS is able to restrict expression of *Shh* to an area called the ZPA in the absence of histological boundaries and is composed of discrete functional elements and binding motifs that allow it to define the area of the ZPA. Mutation within these functional units can result in increasingly severe phenotypes (Lettice *et al.*, 2012, 2014, 2017). Overall, the ZRS and the ZPA is set up as a strong model for understanding how the individual parts of an enhancer contribute to transcriptional activation in a precise point in space and time.

Furthermore, the ZRS is located within intron 5 of a ubiquitously expressed housekeeping gene, *Lmbr1*. This poses questions about mechanisms of transcriptional activation and enhancer-promoter communication. For example, traditional models of transcriptional activation suggests occupancy of enhancers by an array of transcriptional machinery components (Thanos and Maniatis, 1995). However, if this is the case how can an enhancer located in a ubiquitously expressed gene, such as the ZRS in *Lmbr1*, be transcribed through, allowing for a full *Lmbr1* transcript to be expressed whilst occupied by an array of proteins? In this chapter I will look at this using nascent RNA-FISH with probes for *5′ Lmbr1* and probes designed across the ZRS in intron 5 of *Lmbr1*. These probes sets are located ~60 kb apart from each other. By using

both probe sets I aim to determine if transcription of *Lmbr1* is terminated at the point of the ZRS. It is important to note that when I discuss ZRS transcription in this chapter I am not referring to eRNAs, but transcription through the ZRS as part of the larger *Lmbr1* transcript.

It has now been well documented that transcription is not the continuous process we once believed it to be, and in fact occurs in bursts (Raj *et al.*, 2006; Zoller, Little and Gregor, 2018). In a developing system that is sensitive to variation in gene expression levels, transcription must be tightly regulated to drive normal development. One such system involves *Shh.* Reduced levels of *Shh* transcripts in the developing limb bud can lead to varying degrees of digit loss while ectopic expression can lead to additional digits (Anderson *et al.*, 2012). In order to dissect the role enhancers play in transcriptional bursting, its regulation and how this allows a precise transcriptional pattern to be set up, we must first understand what the pattern of expression is in a wild-type (WT) system. Does *Shh* expression indeed occur in bursts? Are there allelic patterns in *Shh* transcription i.e. biallelic or monoallelic expression?

In this chapter I will determine the bursting characteristics of *Shh* at the WT locus, extrapolating a bursting frequency and determine if an active enhancer can be transcribed through. Furthermore, I will look for patterns in *Shh* bursting at the allelic level to explore if there is a possible method of transcriptional regulation determined by the number of bursting alleles in a cell. My main approach for answering these questions will be nascent RNA-FISH.

3.2. Determining the bursting characteristics of *Shh*

3.2.1. Tissue set up and controls for RNA FISH

I decided to use nascent RNA-FISH on ZPA tissue to examine bursting characteristics of *Shh*. Nascent RNA-FISH allows for the direct visualisation of transcripts in tissue. I chose to look at the nascent transcripts as this would allow visualisation of the active transcription sites only. In nascent RNA-FISH, probes are designed to hybridise to the transcript introns. An RNA-FISH approach will result in either 1 or 2 spots of fluorescent signal per cell, indicating either 1 or 2 alleles being transcribed in a cell. Since introns will be spliced out and degraded, the signal detected will be dependent on intron half-life and not on the length of the gene, resulting in visualisation of active transcription sites. By looking at all the cells in a set area and recording the active transcription sites in those cells, I extrapolated a burst frequency based on the percentage of cells bursting at any one time.

In order to maintain consistency, I decided to section embryos so the limb bud was cut along the anterior-posterior axis. This results in sequential sections moving from the distal tip to the proximal end of the limb bud (Figure 3.2A). Sectioning by this method also allows a histological marker to be identified and allowed for me to determine location within the limb bud. The AER appears as a thickening of cells around the edge of the sections and is a strong indicator that ZPA tissue is adjacent (Figure 3.2A,B). Furthermore, sectioning in this orientation on whole embryos allowed for both neural tube and floorplate tissue to be contained within each section. The floorplate is a tissue where *Shh* is highly expressed and therefore will be used as a control tissue for later experiments in this thesis. To ensure that that cells outside of the ZPA boundary were not assessed images were captured so that RNA-FISH signal was distributed throughout the cells in the image (i.e. no large areas of no expression). All cells present in an image were analysed,

allowing me to discern what percentage of cells were bursting at any one time.

The nature of sections means there is the potential that nuclei will be cut, and instances of partial nuclei will be present in any image. Therefore, every nucleus in an image may not have two alleles present. Furthermore, I am looking at a developing tissue and cells will potentially be rapidly dividing, giving rise to cells with potentially 4 alleles. I ascertained what proportion of cells had 0-4 alleles present by performing DNA-FISH on WT E11.5 embryo, sectioned in the same orientation and thickness, 8 μM, as would be done for RNA-FISH. I selected sections that still had AER tissue present in order to orientate myself and know I was looking at cells within ZPA tissue. These regions have previously been shown to express *Shh* by immunohistochemistry (Laura Lettice, personal communication). The AER is easily identified in tissue sections as a thickened area of cells on the posterior margin of the limb bud (Figure 3.2B). DNA-FISH was then performed using DNA fosmids for *Shh*, *Mnx1* and across the ZRS (Figure 3.2C). I chose to count the number of *Mnx1* alleles present in each cell as the signal from this channel was brightest with the least background. I found that I could detect 2 alleles in 64.2% of cells in the ZPA (Figure 3.2D). I could detect 1 allele in 28.5% of cells and 0 alleles in 3.6% of cells. A small number of cells also had 3 and 4 alleles present, 2.9% and 0.7% respectively. In the floorplate similar numbers were observed. 29.6% of cells had one allele, 64.4% had two, 2.2% had three alleles, and none had four. As mentioned, this variation in allele number may be caused by cutting nuclei during sectioning and by cells at different stages of the cell cycle. These data highlights that there is very small numbers of cycling cells and large numbers with at least one allele. The majority of cells still have both alleles present despite sectioning. Since all sections were done in the same manner, that allows burst frequency calculations to be comparable, as the number of alleles lost to sectioning will be similar between sections.

Taking the above factors into account I performed nascent RNA-FISH in WT E10.5 and E11.5 embryo tissue sections. As shown in Figure 1E, RNA-FISH signal is identifiable by differentially coloured puncta within nuclei, with each coloured spot representing a different loci.

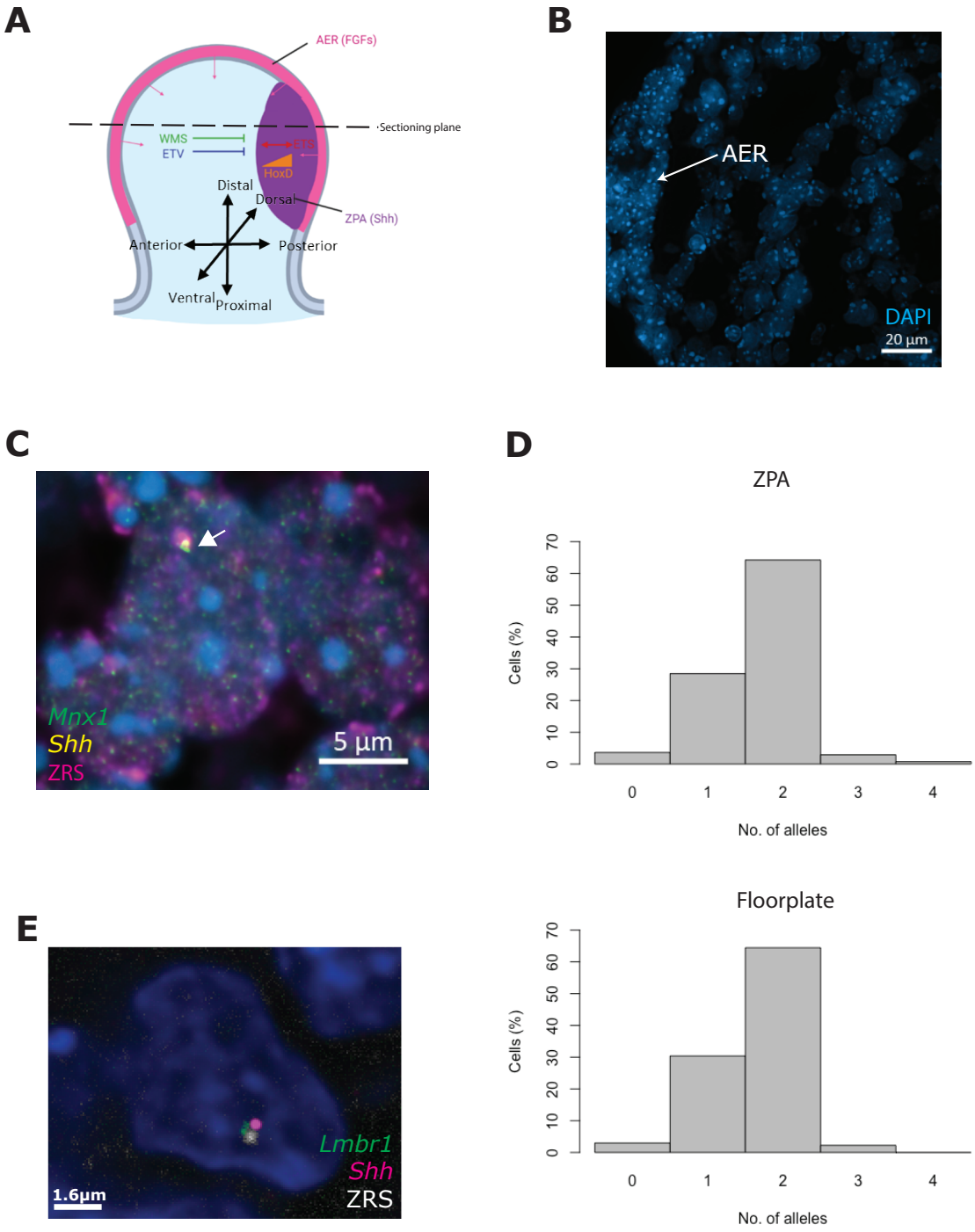


Figure 3.1. RNA FISH set up and controls

A Diagram showing the key features of *Shh* expression in the limb bud. The dashed line indicates how limb buds were sectioned. Black arrows show the different axis that intersect the limb bud. **B** Representative image of DAPI staining on limb bud nuclei. Arrow points to the AER to demonstrate how I orientated myself in the tissue to find ZPA tissue for DNA-FISH. Scale bar = 20 µM. **C** Representative image of DNA-FISH. White arrow points to DNA-FISH signal. *Mnx1* signal is shown in green, *Shh* signal is shown in yellow

and ZRS signal is shown in magenta. DAPI staining shown in blue. Scale bar = 5 μM. **D** Histograms showing the distribution of number of *Mnx1* alleles present in nuclei analysed during DNA-FISH in ZPA and floorplate tissue. **E** Representative image of RNA-FISH in a WT E10.5 embryo. ZPA cell is shown. *Lmbr1* signal shown in green. *Shh* signal shown in magenta. ZRS signal shown in white. Scale bars = 1.6 μM.

3.2.2. *Shh* transcription occurs in bursts

As mentioned in section 1.2, transcriptional bursting has two key parameters: burst frequency and burst size. Nascent RNA-FISH enables extrapolation of burst frequency. *Shh* is expressed in the ZPA at developmental stages E10.5 and E11.5 (Riddle *et al.*, 1993). Therefore, nascent RNA-FISH was conducted on both E10.5 and E11.5 wild-type fixed-frozen tissue (*Figure* 3.3A). As well as forelimb ZPA tissue, floorplate tissue was also analysed as the ZRS does not drive *Shh* expression in this tissue and therefore, will act as a control tissue in later experiments. Due to the orientation of whole embryos during sectioning the flooeplate of the neural tube was present in the same sections as the limb buds. Three probe sets were used. The first were designed by Lauren Kane to the 5' intron of *Shh* (Williamson *et al.*, 2019a). The second two sets I designed using Stellaris® RNA FISH Probe Designer. These sets target intron 1 at the 5' end of *Lmbr1* and intron 5 of *Lmbr1* across the ZRS.

Shh transcription occurs in bursts (*Figure* 3.3). At E10.5 49.9% of cells were observed with *Shh* transcription sites in the ZPA. This number fell slightly to 43.6% at E11.5. However, this difference in expression between the two stages was found not to be significant (independent t-test, *p=0.2342*). *Shh* expression was also observed in the floorplate. As expected higher numbers of active *Shh* transcription sites were observed than in the ZPA. At E10.5 72.8% of cells observed were expressing *Shh*, this number was 74.7% at E11.5. Firstly, these figures confirm that *Shh* transcription is occurring in bursts. There is not an active transcription site in every cell. These findings also show that there is high consistency between in the data between stages and sections.

To confirm transcription had occurred in cells where there were not active transcription sites I conducted RNAScope to detect mature transcripts for

Shh and *Ptch1* (Figure 3.3E). *Ptch1* is directly upregulated by *Shh* expression and therefore is a good marker of *Shh* expression when only low levels of *Shh* can be detected. RNAscope is a proprietary method of smRNA-FISH that amplifies signal so that reportedly one punctum in an image, is equivalent to one mRNA transcript. Mature *Shh* transcripts were detected in almost every ZPA cell (Figure 3.3E), despite nascent RNA-FISH detecting only ~43% of cells bursting for *Shh*. This shows that transcription is occurring in bursts as cells contain mature transcripts, despite not necessarily being under active transcription. In the future, automated image analysis techniques could be developed to quantify transcripts detected by RNAscope, therefore detecting subtle changes in gene expression at the single cell level without losing spatial specific information about where the transcripts are expressed from in a tissue.

3.2.3. ZRS is transcribed through when it is an active enhancer

There are many enhancers located within introns of genes. One study predicted that of the ~36,500 enhancers found in HeLa cells, 37.9% were located in introns (Heintzman *et al.*, 2009). A number of these enhancers are present in genes who are being actively transcribed, such as the ZRS in *Lmbr1* in developing limb bud tissue. Since the ZRS is an active enhancer in this tissue there is the possibility it will be occupied by multiple proteins such as transcription factors and other transcriptional machinery components, suppressing transcription. Therefore, there are two possible outcomes expected from probing ZRS through nascent RNA-FISH: 1) there are active transcription sites confirming its transcription; 2) there is no active transcription site. No ZRS transcription site could be the result of there being no *Lmbr1* expression. A *5' Lmbr1* probe will act as a control to verify that *Lmbr1* is active. If I observe *5' Lmbr1* expression but no ZRS expression this suggests that *Lmbr1* expression has been prematurely terminated. Likely termination would be caused by occupation of the ZRS, while it is an active

enhancer, by proteins. Therefore, preventing transcriptional machinery that is transcribing *Lmbr1* to pass the through the ZRS.

ZRS is capable of being transcribed through whilst *Shh* is actively being transcribed, suggesting that occupation of the ZRS by different TFs and other transcriptional machinery for driving *Shh* transcription does not prevent transcription through the enhancer itself (Figure 3.3A,C). Approximately the same percentage of ZRS active transcription sites are observed in E10.5 (31.3%) and E11.5 (30.4%) embryos in the ZPA (Figure 3.3C). As we would expect this also correlates with the percentage of 5' *Lmbr1* transcription sites observed; 38.9% in E10.5 embryos and 31.5% in E11.5 embryos (Figure 3.3D). It should be noted that the probes for ZRS and *Lmbr1* are located ~60 kb apart. Given this linear distance it is possible with transcriptional elongation rates and intronic half-life that *Lmbr1* and ZRS signals will not be observed together, or if they are the signal could be from separate transcriptional bursts (Figure 3.3F). The elongation rate and intron half-life are untested at this locus.

The presence of both *Lmbr1* and ZRS transcription sites reveals that despite the fact that, in this context, the ZRS should be occupied by a number of proteins forming a transcriptional coactivator complex driving *Shh* expression, it is still possible for *Lmbr1* to be transcribed through the ZRS. In fact, out of all the cells observed 22.4% of E10.5 and 17.63% of E11.5 ZPA cells were expressing ZRS and *Shh* at the same time on the same chromosome. This leads to several possibilities: large protein structures can be transcribed through, meaning the ZRS and the transcriptional proteins bound to it are a transient active structure, or perhaps close physical contact between the ZRS and transcriptional proteins is not necessary to drive transcription of *Shh*. The possibility that close contact between the ZRS and

transcriptional proteins is not necessary to drive transcription of *Shh* supports a transcriptional hub model, which will be discussed further in chapter 5.

Interestingly, there is not equal expression of ZRS and 5' *Lmbr1* in the floorplate at both stages (Figure 3.3C,D). There is a marked reduction of ZRS expression compared to 5' *Lmbr1*. At E10.5 32.8% of cells are observed with 5' *Lmbr1* transcription sites, while only 20% express ZRS (paired t-test, $p = 0.01$). Similarly, at E11.5 28.6% of cells express 5' *Lmbr1* with only 15.6% expressing ZRS, however this difference is not significant (paired t-test, $p=0.08$). This difference at E10.5 may be due to a high rate of run off transcripts or suggests suppression of ZRS transcription in floorplate tissue. As we see relatively even expression of both alleles in the ZPA (*Figure* 3.3C,D) it is clear this apparent suppression is not due to the linear distance between 5' *Lmbr1* and the ZRS (~60kb). Run off transcripts are created due to abortive transcription prior to the transcription complex being released from the promoter, which could result in the appearance of more 5' transcript than 3' transcript. However, it is believed that these transcripts are only tens of bp long (Revyakin *et al.*, 2006). Therefore, this is unlikely the cause of the different levels on *Lmbr1* and ZRS signals here as the *5' Lmbr1* probes are located ~2000bp from the start of the transcript.

It is also likely that this observation could be due to a technical reason. For example, there may have been higher background than usual in that channel during imaging, resulting in loss of signal during the deconvolution process. Out of focus light causes an image to blur and fluorescent signal to be unclear. Deconvolution is an image processing technique that improves the sharpness of an image and removes the background, unfocussed light. However, in light channels that are more auto fluorescent, such as the green channel, this can result in weaker signal being removed from the image also. Due to the observation of reduced transcription through the ZRS not being

seen in ZPA or E11.5 floorplate, it is highly likely that this observation is due to technical reasons such as autofluorescence.

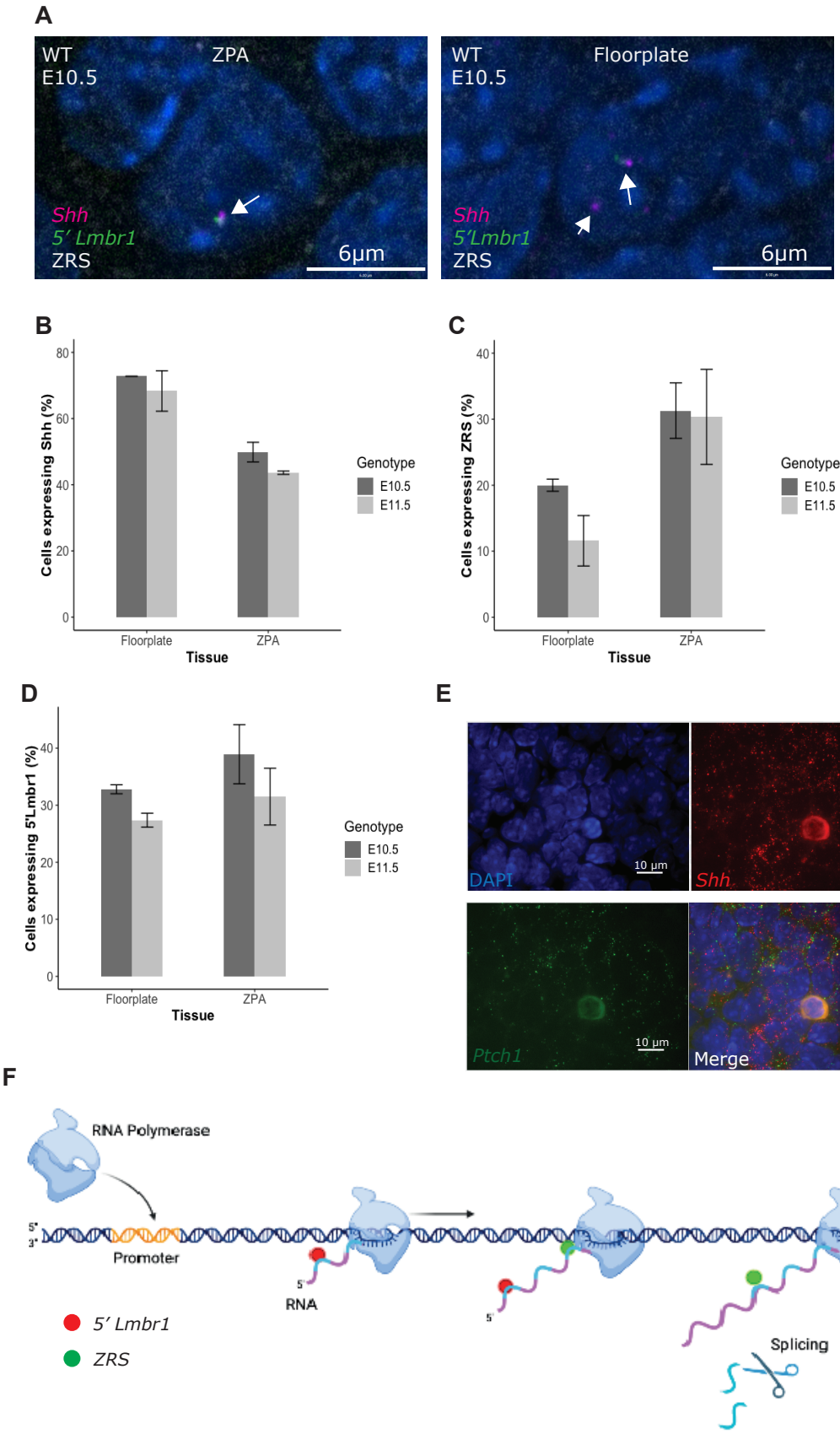

A
WT
E10.5
ZPA
Shh
5′ Lmbr1
ZRS
6µm
WT
E10.5
Floorplate
Shh
5′Lmbr1
ZRS
6µm
B
Cells expressing Shh (%)
Floorplate
ZPA
Tissue
Genotype
E10.5
E11.5
C
Cells expressing ZRS (%)
Floorplate
ZPA
Tissue
Genotype
E10.5
E11.5
D
Cells expressing 5'Lmbr1 (%)
Floorplate
ZPA
Tissue
Genotype
E10.5
E11.5
E
DAPI
10 µm
Shh
10 µm
Ptch1
10 µm
Merge
10 µm
F
RNA Polymerase
Promoter
RNA
5′ Lmbr1
ZRS
Splicing

Figure 3.2 Burst frequencies of Shh, 5' Lmbr1 and ZRS by nascent RNA-FISH of ZPA and floorplate tissue.

A Representative image of nascent RNA-FISH in E10.5 ZPA (left) and floorplate (right) tissue in. Probes for *Shh* (magenta), *5' Lmbr1* (green) and ZRS (white) were used and RNA-FISH signal of these probes are indicated by white arrows. Scale bars = 6 μM. B-D) Percentage of cells expressing (**B**) *Shh*, (**C**) *5'Lmbr1* and, (**D**) ZRS were recorded from both E10.5 and E11.5 wild-type embryos. Error bars calculated by SEM, obtained from two or three tissue sections from two embryos. For each embryonic stage there is n = ~300 nuclei (sum of all nuclei from all sections of both embryos analysed). **E** RNAscope for mature *Shh* and *Ptch1* transcripts in E11.5 ZPA tissue. DAPI shown in blue, *Shh* shown in red. Scale bars = 10 μM. **F** Diagram showing different possible ways in which RNA-FISH signals may occur on a *Lmbr1* transcript. The red spot represents a *5' Lmbr1* signal, the green spot represents a ZRS signal. Purple and blue wavy lines represent transcripts, with the blue indicating introns and purple indicating exons. These probes are designed to areas ~60 kb apart so when *5' Lmbr1* and ZRS are viewed together it is not necessarily on the same transcript.

3.2.4. Analysis of expression of homologous *Shh* alleles

Upon initial examination of how many transcription sites are present in each cell, it appears that actively transcribing ZPA cells may have a preference for monoallelic expression at both embryonic stages for all three alleles probed for (Figure 3.4). Expression of a single *Shh* allele was detected in 42.2% of cells at E10.5 and 39.0% at E11.5 in the ZPA. Whereas expression of two alleles was only detected in 7.6% of E10.5 and 4.6% of E11.5 ZPA cells (Figure 3.4A). Similarly, in the floorplate, 53.4% of cells express one *Shh* allele at E10.5 and 57.9% at E11.5 (Figure 3.4A). Interestingly, this trend is also prominent for *Lmbr1* and ZRS. The ZRS bursts from one allele in 27.7% of cells, versus from two alleles in 3.6% of cells in the ZPA at E10.5 (Figure 3.4B). At E11.5, 29.2% of cells burst from one ZRS allele and 1.2% burst from two (Figure 3.4B). In the ZPA 31.3% of cells express one allele of *Lmbr1* compared to 7.6% expressing from two alleles at E10.5 (Figure 3.4C). This is mirrored at E11.5 where 30.2% of cells burst from one *Lmbr1* allele and 1.6% burst from two (Figure 3.4C). Similar trends are seen in the floorplate (Figure 3.4B, C). It is important to note, that the above observations do not take into account the alleles that will be lost in the sectioning process, as discussed in 3.2.1.

To examine if monoallelic expression is a significant trend I consulted with Hannes Becher (Bioinformatics core statistician, IGC) and we adapted Hardy-Weinberg Equilibrium (HWE). This theorem states that if alleles of a gene segregate at random (i.e. independently from each other) then genotype frequencies in a population should be as follows (where A and a are arbitrary genotypes):

$$AA = p^2$$

$$Aa = 2pq$$

$$aa = q^2$$

Where p is the frequency of A in the population and q is the frequency of a. In my data I have count values of either 0, 1 or 2 alleles bursting in a cell. These can therefore be analogous to aa, Aa and AA respectively. I can test for HWE under the assumption that alleles segregating independently is analogous to transcriptional bursts happening independently for the two alleles of one gene. HWE can be tested using a Chi-squared test. The null hypothesis is that Hardy-Weinberg equilibrium holds using the Chi-Square method. Therefore, if p-value ≤ 0.05 then bursting alleles are not in HWE and therefore do not burst in a monoallelic manner. It should be noted that at this stage this is a preliminary analysis. Unfortunately, it does not yet consider the DNA-FISH data from section 3.2.1 which determines how many nuclei have been cut and are therefore missing alleles.

In E10.5 ZPA the *Shh* alleles are in HWE, as are the two ZRS alleles (Chi-squared test, *Shh p-value = 0.62, ZRS p-value = 0.47*). Each allele of *Shh* is bursting independently, meaning the two *Shh* alleles in a cell are observed bursting at the same time less than would be expected based on their burst frequency, and therefore is exhibiting monoallelic expression. This is also the case with the two ZRS alleles. Interestingly, the *Lmbr1* alleles are not in HWE (Chi-squared test, *p-value = 0.02*). This is unusual considering the ZRS is contained within *Lmbr1* and therefore it would be expected their results would be similar. Sections from two embryos were analysed during RNA-FISH, further investigation found that *Lmbr1* not being in HWE results from the measurements in the second embryo, the first embryo is in HWE (Chi-squared test, *Lmbr1 Embryo 1 p-value = 0.51, Embryo 2 p-value = 0.02*). Breaking it down even further, each individual section was analysed (two sections for each embryo) and it was found every individual section was in HWE. This is analogous to the Walund effect. The Walund effect is when two or more subpopulations are in a Hardy–Weinberg equilibrium but have different allele frequencies. This will be explored further in the discussion. For all three alleles at E10.5 in the floorplate tissues the null hypothesis can be

rejected, meaning they are in HWE. (Chi-squared test, *Shh p-value = 0.16, Lmbr1 = 0.56, ZRS = 1*).

At E11.5 in the ZPA *Shh* is also in HWE (Chi-squared test, *p-value = 0.18*). However, *Lmbr1* and ZRS are not (Chi-squared test, *Lmbr1 p-value = 0.006, ZRS = 0.04*). Again, this can be attributed to the Walund effect. When looking at *Lmbr1* individual sections, only one is not in HWE. In the case of ZRS all individual sections are in HWE. In the floorplate at this stage *Lmbr1* and ZRS are both in HWE (Chi-squared test, *Lmbr1 p-value = 0.09, ZRS = 0.09*). Interestingly, *Shh* is not in HWE again because of variation in one section.

Overall, these data suggest that there is a preference for monoallelic expression at the *Shh* locus. Transcriptional bursting has often been described as a stochastic process (Elowitz *et al.*, 2002; Raj *et al.*, 2006; Larsson *et al.*, 2019; Sun and Zhang, 2019). However, the observations shown here suggest that there may be a level of transcriptional regulation that favours single allele activation. Possibly, this regulation is to overcome the noisy nature of transcription and maintain transcript levels at a desirable level for normal development. However, it is interesting that *Lmbr1,* a housekeeping gene, follows the same trend for monoallelic expression. Perhaps housekeeping genes are required at relatively low levels across a variety of tissues and therefore are expressed in a monoallelic manner. Therefore, suppression of two *Shh* alleles being expressed may be a method of preventing negative dosage effects of *Shh*, such as polydactyly.

As mentioned, due to time constraints, a caveat of these calculations is that they do not take into account the number of nuclei that have been cut during the sectioning process, resulting in the loss of alleles. Therefore, it must also be considered that the observation of monoallelic expression is the result of

~29% (Figure 3.2D) of cells only having a single allele present, which could create an impression of higher monoallelic expression than is actually occurring. Therefore, the limitation of this experiment shall be discussed in greater detail below.

Another assumption made here is that all bursts are captured by RNA-FISH. There is the possibility that some transcripts do not have enough RNA-FISH probes hybridised to them for a signal to be observed by microscopy. Again, this would give the impression of higher monoallelic expression than is actually occurring.

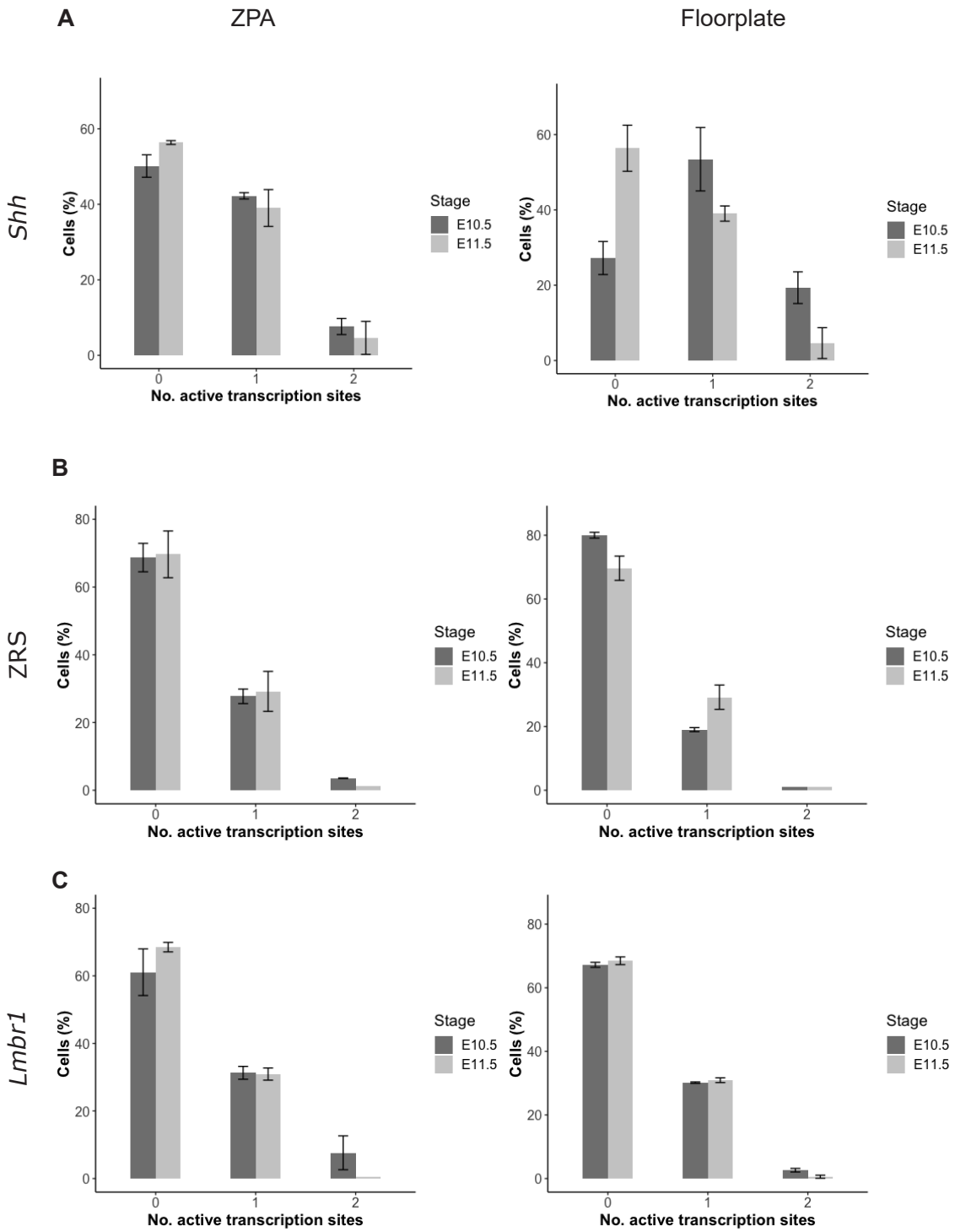


Figure 3.3 Expression of homologous Shh alleles in ZPA and floorplate tissue at E10.5 and E11.5

Histograms showing the number of active *Shh* **(A)**, ZRS **(B)** and *Lmbr1* **(C)** transcription sites present in cells in ZPA (left) and floorplate tissue (right) at

both E10.5 and E11.5. Either 0, 1 or 2 sites were present in analysed nuclei. Error bars calculated by SEM. N = ~300 nuclei.

3.3. Discussion and Conclusions

In this chapter I have utilised nascent RNA-FISH technology to explore several key questions around transcription of a highly regulated developmental gene, *Shh*. Firstly, I set up for RNA-FISH to be conducted in a controlled manner, with consistency between tissue sections in section 3.2.1. I set out my criteria for determining that I was observing the correct tissue using histological markers and determined how many nuclei were losing alleles through the sectioning process (Figure 3.2). I then went on to show that *Shh* transcription is occurring in bursts by demonstrating that there are not active transcription sites in every ZPA cell, but there are mature transcripts, indicating transcription has occurred (Figure 3.3A,B). Transcriptional bursting is a process thought to be stochastic and noisy in nature (Raj *et al.*, 2006; Little, Tikhonov and Gregor, 2013; Larsson *et al.*, 2019).

Following this I explored whether an enhancer crucial for *Shh* limb expression and located in a ubiquitously expressed gene could be transcribed while actively driving *Shh* transcription. Interestingly, *Lmbr1* transcribed through the ZRS when it is an active enhancer which could be suggestive of a transcriptional hub model for enhancer promoter interactions (Figure 3.3C). Finally, I analysed bursting of homologous alleles. Preliminary results suggest that *Shh* is expressed in a monoallelic manner, which suggests suppression of both alleles being expressed simultaneously (Figure 3.4). However, it should be noted these results do not take into account the number of alleles lost in the tissue sectioning process. Overall, these findings suggest that, transcriptional bursting, a process documented to be stochastic, potentially has methods of regulation and that enhancer-promoter interaction is transient which supports a transcriptional hub model rather than the traditional looping model.

3.3.1. Monoallelic expression as a method of transcriptional regulation

Previous works have highlighted adverse phenotypes that occur when *Shh* expression levels are affected (Sagai *et al.*, 2005; Lettice *et al.*, 2014; Crane-Smith *et al.*, 2021). More specifically mutations within the ZRS that cause a reduction in *Shh* levels can cause digit loss and syndactyly and increased levels can cause polydactyly (Lettice *et al.*, 2017). Therefore, expression levels of *Shh* must be highly regulated to maintain normal development. Conversely, it is well documented that transcription is a noisy process (Chubb *et al.*, 2006; Raj *et al.*, 2006; Chubb and Liverpool, 2010; Larsson *et al.*, 2019). This poses a conundrum where a noisy process must have some level of regulation. Here I have shown that transcription of *Shh* occurs in bursts, despite its need for transcript levels to be maintained in a regulated manner.

One possibility for how transcript levels are regulated is the suppression of two alleles from bursting simultaneously as shown in this chapter (Figure 3.4). Indeed, there has been previous evidence to support monoallelic expression as a method of regulation. Early work on eukaryotic transcriptional bursting involved using MS2-loops in *Dictyostelium* (Chubb *et al.*, 2006). The group looked at a developmental gene, *dscA*, and found that reinitiation of transcription is more common than *de novo* initiation of transcription. This implies some form of transcriptional memory, perhaps caused by a pre-existing transcriptional complex or favourable chromatin confirmation (Chubb *et al.*, 2006). This could imply that one allele is more likely to burst than the other. However, this work looked at this at the cellular level, determining a cell that had already burst was more likely to again, rather than the allelic level.

Later works were able to determine which alleles were bursting and if there was an allelic imbalance. One study explored random monoallelic expression of ~1300 autosomal genes (Zwemer *et al.*, 2012). Random monoallelic expression is an epigenetic phenomenon where a gene is more likely to be expressed from either the maternal or paternal allele. They found >10% the investigated genes displayed random monoallelic expression (Zwemer *et al.*, 2012). In another investigation using allele-specific RNA-FISH Symmons *et al.* (2019), were able to look at various genes where monoallelic expression had been reported. Using their allelic-specific data and mathematical modelling they were able to show that instances of monoallelic expression are due to low levels of expression and not through some greater regulatory mechanism (Symmons *et al.*, 2019). Bearing this in mind it is possible that the monoallelic expression I have observed is not the result of a regulatory mechanism. However, Symmons and colleagues (2019) only looked at three genes where monoallelic expression had been reported. In fact, other works have suggested that monoallelic expression of a gene is largely a trait of genes that specify cell fates (Nag *et al.*, 2015). Suggesting that for genes where transcript levels must be highly regulated, monoallelic expression is a method of controlling levels. As previously noted, *Shh* is a morphogen responsible for specifying cell fates and therefore falls into this category.

As mentioned earlier, I have conducted these experiments on tissue sections and therefore some nuclei will be cut, potentially losing copies of alleles. Using DNA-FISH I showed that approximately 30% of nuclei in the floorplate and ZPA did not have two alleles present (Figure 3.2D). This number represents the maximum number of alleles loss but does not necessarily represent the actual number of alleles lost. For this to be the exact number of alleles lost there is the assumption that the experimental conditions were perfect. Firstly, it assumes that the DNA fosmids have penetrated the nuclei and bound to the DNA with an 100% success rate. Fosmid probes are large in size, for example the *Shh* probe used is ~40 kb, this can limit the success

of hybridizing. Also, the deconvolution process can lead to high background fluorescence obscuring the real signal. Therefore, the number of nuclei I have reported having lost alleles is a “worst case scenario” and the actual figure could be smaller. Despite this, the number of lost alleles should be considered when looking at monoallelic expression. Unfortunately, due to time constraints this was not possible, and therefore these results could be skewed, making monoallelic expression appear more apparent than it is in actuality. Hannes Becher (IGC core statistician) will develop a method to calculate maximum likelihood projections that will take alleles lost in the sectioning process into consideration.

In a few instances the Walund effect has occurred making it appear as if the whole population of alleles are not bursting independently. This occurs when subpopulations are in HWE but have different allele frequencies. There could be several explanations for this occurring in my data. Firstly, there could be small discrepancies in the staging of the embryos. While I followed a set criterion for choosing embryos (detailed in materials and methods), there was no way for me to know the exact time of fertilization and therefore it is possible that embryos analysed were at slightly different stages. Another possibility is that fluorescent background could be higher in some tissue that others, resulting in loss of signal during deconvolution. This effect has been seen most prominently for *Lmbr1*, and probes for this allele were imaged in the green channel, which is usually the channel with the highest autofluorescence.

Overall, I propose that monoallelic expression could be a mechanism for transcriptional regulation for *Shh.* Expression levels of *Shh* are crucial for ensuring normal development and therefore, in the face of a stochastic process, monoallelic expression helps maintain low expression levels. However, further investigation would be needed to show this definitively. One

way to begin to explore this would be to look at allele specific expression. This would have to be done at the single-cell level to avoid population average results. Therefore, scRNA-seq or a new allele specific method of live-cell imaging (Maass *et al.*, 2018), could look at allele specific bursting. However, the issue with these methods is that they rely on there being SNPs present in a gene to differentiate the alleles. In *Shh* there are no SNPs in the coding region which could make these techniques possible.

3.3.2. Transcription through an active enhancer suggests a transient enhancer-promoter interaction complex

Currently, looping, a proposed method of enhancer-promoter interaction, suggests that a stable transcriptional complex is formed between the enhancer and promoter, with binding at both regulatory elements (Lim and Levine, 2021). The ZRS is located in a ubiquitously expressed gene, *Lmbr1*, and therefore should be transcribed through as part of a larger *Lmbr1* transcript. However, in the ZPA the ZRS is actively driving transcription of *Shh* and therefore, if the looping model is valid will be occupied by a number of transcription factors and other transcriptional proteins. Here, I have shown that the 5' end of *Lmbr1* is expressed in the ZPA (Figure 3.3D). Furthermore, I have shown that this transcription extends through the ZRS, even in a context where *Shh* is being transcribed from the same chromosome. For this to be possible, if there is a stable and intimate complex between enhancer and promoter it must be transient to allow transcriptional machinery to pass through the ZRS. Importantly, if large transcription factor complexes form at an enhancer it seems likely that there are mechanisms that enable the transcription machinery to pass. One potential caveat of these results is that if *Shh* is being transcribed, then activation must have occurred at an earlier time point. Therefore, in nascent RNA-FISH when *Shh* and ZRS have been

observed together, this maybe not necessarily be while ZRS is activating *Shh*.

Previously it has been shown that enhancers can be transcribed and there has been increasing evidence of enhancer RNAs (eRNAs). Transcription occurs in a higher percentage of the genome than is protein coding (reviewed: Natoli and Andrau, 2012). In fact, in 2010 it was shown that enhancer regions are actively transcribed (de Santa *et al.*, 2010; Kim *et al.*, 2010). The role of these eRNAs is debated. Some suggest that eRNAs are just a by-product of noisy transcription. Multiple studies have shown that exogenous overexpression of eRNAs lead to an increase in target mRNAs (Alvarez-Dominguez *et al.*, 2017; Shii *et al.*, 2017; Jiao *et al.*, 2018), suggesting a causal role of eRNAs in transcription. It is perhaps that eRNAs are transcribed to allow a conformational changed in the chromatin to make enhancers accessible. Evidence has implicated eRNAs in models of enhancer-promoter interaction (reviewed: Arnold, Wells and Li, 2020). Some have suggested that eRNAs help recruit cohesion complex subunits, however contrasting studies have suggested cohesin recruitment is unaffected by eRNA knock-down (Arnold, Wells and Li, 2020). One study showed that eRNAs helped “trap” transcription factors to increase their binding affinity to DNA (Sigova *et al.*, 2015). Therefore, enhancers are frequently transcribed, and this transcription may have a functional role.

Unpublished work from the Hill lab by Laura Lettice (personal correspondence) has shown that introducing a transcriptional stop cassette in *Lmbr1* upstream of the ZRS does not impact *Shh* expression. This uncouples the expression of *Lmbr1* and the ZRS being necessary for *Shh* expression but does not rule out the presence of eRNAs. It could still be possible that ZRS eRNAs are produced by a separate method to *Lmbr1* transcription. I think it unlikely that eRNAs present in nuclei I have observed will be shown

as RNA-FISH signal due to probes spanning over 3 kb with the ZRS located centrally. The ZRS is ~750 bp long and therefore an insufficient number of probes will hybridise to this area to produce a signal without transcription of the adjacent sequences. Furthermore, if RNA-FISH detected eRNAs I would expect to see ZRS RNA-FISH signal at a higher frequency than *5′ Lmbr1* RNA-FISH signal, this is not the case. RNA-FISH signal for *5'Lmbr1* probes and probes across the ZRS are observed in ~30% of ZPA cells (*Figure* 3.3C,D). Therefore, ZRS signal observed is most likely part of the *Lmbr1* transcript. However, it has been shown that eRNAs can be transcribed bi-directionally using a technique called Cap analysis of gene experession (CAGE), which aimed to identify putative enhancers by their eRNAs (Murakawa *et al.*, 2016). Therefore, it is possible that the way the ZRS probes have been designed, transcription of *Lmbr1* could be terminating, but it appears it has not due to bidirectional eRNA transcripts. A future experiment to determine this would be to design 3' *Lmbr1* RNA-FISH probes and see if they are observed at a similar frequency to *5′ Lmbr1* probes.

The ZRS RNA-FISH probes that have been used here are designed to a region of ~3 kb with the ZRS located centrally. To ensure that transcription is progressing past this point I could design probes to the 3’ of *Lmbr1*. It is possible that transcription could be terminating at the ZRS. However, I find this unlikely as not enough probes hybridise upstream of the ZRS to cause a signal that could be captured using fluorescent microscopy.

The ZRS being transcribed has implications in predicting modes of enhancer-promoter interaction which will be discussed further in chapter 5. I think these results imply that a transcriptional hub model for enhancer-promoter interaction is more likely than a looping model. The hub model allows for transient structures and does not require stable contact between the enhancer and promoter, this would allow the ZRS to be transcribed through

whilst it is driving transcription. Whereas the looping model requires close contact between the enhancer and promoter and occupation of different transcriptional proteins, forming a large complex.

Chapter 4 Functional elements of the ZRS and their effects on transcription

4.1. Introduction

The ZRS is the sole driver of *Shh* transcription in the developing mouse limb bud (Lettice *et al.*, 2003, 2014; Anderson *et al.*, 2014), which makes it an interesting and well-suited model for understanding how regulatory elements within an enhancer can impact the spatiotemporal transcriptional pattern of a gene. Work from the Hill lab has previously identified a network of activators and repressors that bind the ZRS and determined what role these proteins play in setting up limb development (Lettice *et al.*, 2012, 2017; Peluso *et al.*, 2017). This network of activators and repressors is essential in defining the ZPA, which is not restricted by any overt histological boundaries to its anterior edge.

To summarise, FGFs broadly prime the distal limb bud for *Shh* expression (Peluso *et al.*, 2017). Binding of the ETV sites and the WMS repressor site in the ZRS in the anterior of the limb bud prevent the ZRS driving *Shh* expression in this area (Lettice *et al.*, 2012). Binding of the ETS sites defines the boundary of the ZPA while Hox site binding determines the levels of *Shh* expressed (Lettice *et al.*, 2017). Therefore, the ZRS is composed of different functional units that determine the action of the enhancer as a whole. While we know details of how the ZPA is defined via binding of different TF binding sites within the ZRS, it is not yet fully understood what role these binding sites are playing at the transcriptional level. Previous research has shown how mutations in these binding sites down- or up-regulate *Shh* expression (Lettice *et al.*, 2012, 2017; Peluso *et al.*, 2017). However, it is unknown how these mutations change transcriptional bursting characteristics. Previous findings have suggested that enhancers encode for the frequency of bursts, while core promoters encode burst size (Bartman *et al.*, 2016; Larsson *et al.*,

2019). However, previous research hasn't investigated in detail how individual transcription factors impact these bursting characteristics. Investigating this aspect further could highlight subtle differences in the actions of different TFs and therefore, how their loss of binding could impact disease.

The ZRS is bound by transcription factors that have different mechanisms. In this chapter I will look at both mutations in Hox binding sites and Lim homeodomain binding sites. As already mentioned Hox site binding establishes levels of *Shh* transcribed. Lim homeodomain proteins act as pioneer factors (Silvia Peluso, manuscript in preparation). Pioneer factors are a type of transcription factor crucial in chromatin activation (Iwafuchi-Doi and Zaret, 2016; Mayran and Drouin, 2018; Larson, Marsh and Harrison, 2021). They have the ability of being able to identify and bind their target DNA sequence in 'inactive' or 'closed' chromatin. This, therefore establishes domains of accessible chromatin for subsequent transcription factor binding (Zaret, 2020). While pioneer factors are a type of transcription factor, not all TFs are pioneer factors as they do not have the ability to bind inactive chromatin. Therefore, these different types of transcription factor could influence bursting characteristics differently. Removal of increasing numbers of Hox sites causes increasingly severe digit loss however difference in expression levels are not detectable by qPCR (Lettice *et al.*, 2017). Therefore, I also decided to investigate the role of individual Hox sites on bursting frequency.

The floorplate of the neural tube is a tissue that highly expresses *Shh*, but is not driven by the ZRS, instead it *Shh* expression is driven by SFPE1 and SFPE2 (Epstein, McMahon and Joyner, 1999). Expression of *Shh* in the floorplate is important for the dorso-ventral patterning of the neural tube. However, for the purposes of the experiments in the following chapters the

floorplate will be used as a control tissue, as mutations in the ZRS should not impact *Shh* expression in this tissue.

By using nascent RNA-FISH I plan to dissect how down-regulatory and up-regulatory ZRS mutations effect the transcriptional characteristics described in Chapter 3. I shall explore whether different modes of transcription factor activity have different effects on transcriptional bursting. I will look at what impact individual Hox sites play in defining levels of *Shh* transcripts. Finally, I will also explore what role repressor site binding contributes to bursting characteristics.

4.2. Down-regulatory ZRS mutations and transcription

4.2.1. Hox site removal causes reduction in transcriptional bursting

The ZRS is composed of an array of different binding sites for transcriptional activators and repressors. For example, the ZRS is known to contain at least four Hox sites. These sites bind HOXD9, 10 and 11 proteins with high affinity and HOXD12 and 13 with lower affinity (Lettice *et al.*, 2017). Previous work has shown that removal of an increasing number of these Hox sites results in increasingly severe digit loss (Lettice *et al.*, 2017). Lettice et al. (2017) created two mouse lines, ∆48 and ∆110, to show the phenotypic effects of sequential Hox site loss. These lines were created using CRISPR-Cas9 technology, through pronuclear injection of plasmid DNA. Amongst Hox sites 1-3 there is also the WMS repressor site (Figure 4.1A). Binding of this site has been shown to repress ectopic *Shh* expression (Lettice *et al.*, 2017). A 20bp deletion of the ZRS which removes Hox site 3 and the WMS site results in no discernible phenotype, even when placed on a sensitized background (Lettice *et al.*, 2017). This suggests that removal of a Hox site, cancels out ectopic *Shh* expression caused by WMS site removal. Removal of an additional Hox site, Hox site 2 in mutant ∆48 causes partial digit loss. Further deletion of Hox site 1 in mutant ∆110 caused digit loss in (Figure 4.1B). Looking at relative expression levels in the ZPA showed a significant decrease in *Shh* in both ∆48 and ∆110 compared to WT. However, there was no significant difference in levels between ∆48 and ∆110 themselves (Lettice *et al.*, 2017). This poses questions about how a more severe phenotype occurs despite relative expression levels seemingly remaining the same, and what actions binding of Hox proteins to Hox binding sites is having on *Shh* expression.

I decided to use these two mutants, Δ48 and Δ110, to determine what role HoxD proteins have on the transcriptional characteristics of *Shh* expression in the developing limb. In doing so I could highlight how Hox proteins work and could provide insights into the mechanisms by which transcription factors drive expression. Hox proteins are particularly interesting as subsequent loss of sites could uncover more subtle actions of transcription factors on an enhancer. Investigation was conducted by using nascent RNA-FISH on E11.5 embryos to look at the burst frequency in the ZPA in Δ48, Δ110 and WT littermates. I also examined floorplate tissue as a control tissue. The ZRS drives *Shh* expression in the ZPA only. Therefore, *Shh* expression in the floorplate should be unaffected by the Δ48 and Δ110 mutations. Probes were used for *Shh* and *5' Lmbr1*. Probing for *Lmbr1* works as an internal control. While in the same TAD, *Lmbr1* expression is not driven by the ZRS. Therefore, the Δ48 and Δ110 mutations should not affect *Lmbr1* expression.

The bursting frequency of *Lmbr1* is not affected by either the Δ48 or Δ110 ZRS mutations (Figure 4.1C and D). This is prevalent in both the ZPA and the floorplate tissue and reinforces the notion that the ZRS does not regulate *Lmbr1* expression. A significant drop in cells containing active transcription sites was observed in Δ48 ZPA tissue (Figure 4.1E). In WT littermates *Shh* expression occurred in ~43% of cells. This number was reduced to ~17% in Δ48 homozygotes (independent t-test, *p-value = 0.0005*). In the floorplate *Shh* expression remains constant when compared to WT littermates. Furthermore, in Δ110 ZPA tissue, there is again a significant drop in *Shh* expression. In WT embryos *Shh* expression occurred in ~33% of ZPA cells, this was reduced to ~11% in Δ110 homozygotes (independent t-test, *p-value = 0.003*). The difference seen in WT *Shh* expression in Δ48 and Δ110 littermates is not significant (independent t-test, *p-value = 0.143*). In Δ110 floorplate tissue *Shh* expression remains unaffected by the ZRS mutation, as expected (Figure 4.1F). Notably there is a difference in WT *Shh* expression between WT littermates of Δ48 and Δ110 mice, a likely reason for this could

be slight differences in embryo stages. It is not possible to know the exact time of conception of the embryos. While certain parameters were used to stage embryos e.g. eye colouration and limb bud size, this cannot guarantee that embryos are the exact same stage.

It has previously been shown that the removal of an additional Hox site in the ∆110 mice compared to the ∆48 mice has a more severe phenotype: loss of a full digit compared to partial loss (Lettice *et al.*, 2017). However, qPCR did not reveal a difference in expression levels. Nascent RNA-FISH presented the opportunity to look at changes in expression in a more precise manner. The use of qPCR can result in the inclusion of non-expressing tissue due to difficulties dissecting small tissues. Through the use of nascent RNA-FISH I could ensure only ZPA tissue was analysed, allowing more subtle changes in expression to be observed. This technique has successfully been used previously for measuring changes in expression (Williamson *et al.*, 2019b; Crane-Smith *et al.*, 2021). Therefore, subtle expressional changes could explain the difference in phenotype between the two mutants. In ∆48 ZPA tissue ~17% of ZPA cells were expressing *Shh*. This number was ~11% in ∆110 ZPA cells. I found that there was no difference in *Shh* expression levels between ∆48 and ∆110 mice (independent t-test, *p-value = 0.78*). Therefore, despite this being a more sensitive technique, I was still unable to find a difference that could explain the increasingly severe phenotype.

The RNA-FISH technique used here would only show a difference in bursting frequency. This could possibly explain the difference in phenotype between the two mutants. Potentially, the burst frequency may not have changed, but the overall number of cells which have the ability to burst had, implying that HOXD binding to the ZRS may have two roles in *Shh* transcription. Firstly, Hox binding may determine the frequency at which *Shh* is expressed in

nuclei and secondly, may determine the overall number of cells which are capable of doing so, possibly due to sensitivity to an exogenous factor.

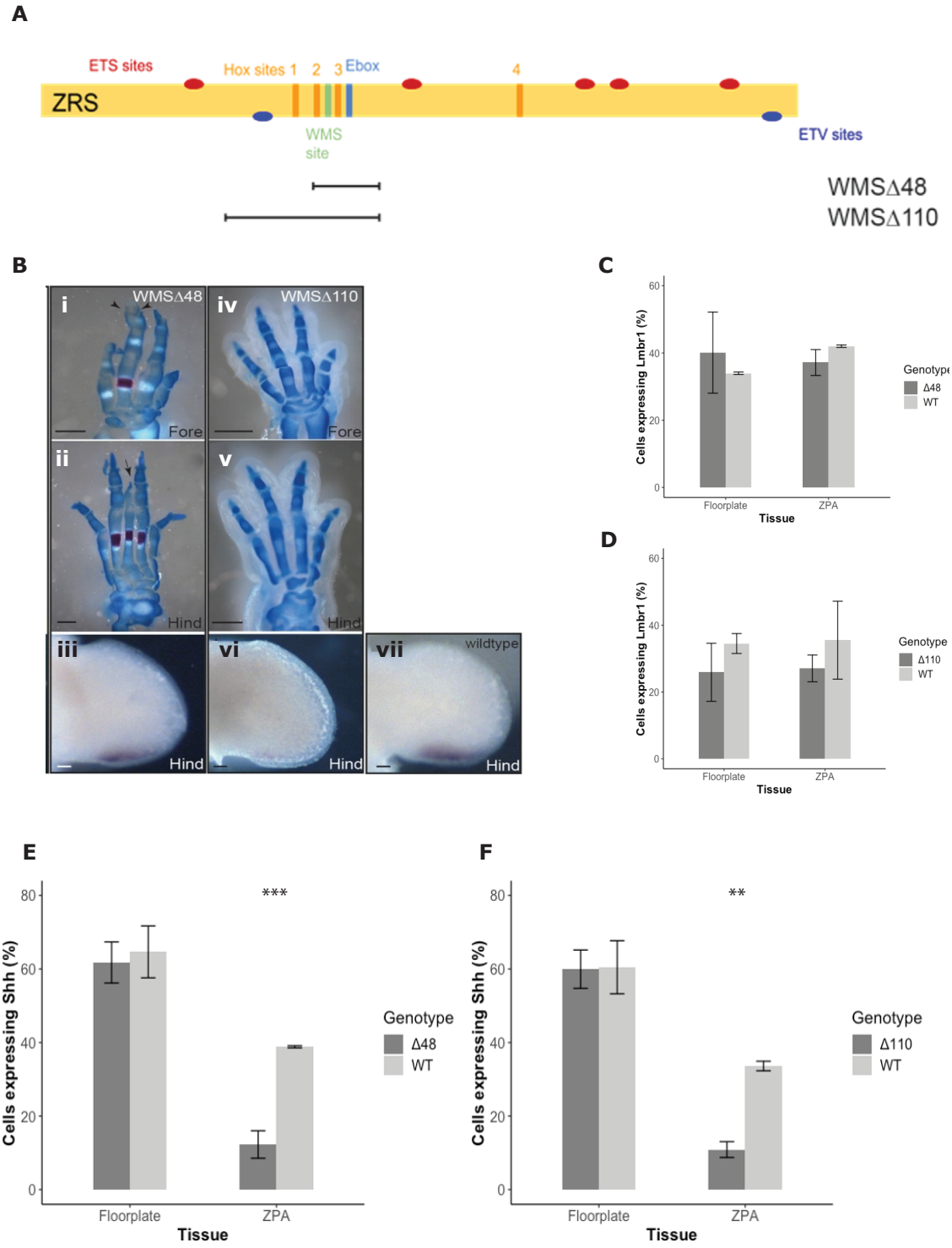


Figure 4.1 Hox site deletion results in decreased bursting frequency in the ZPA

A Diagram of the ZRS showing the relative locations of various binding sites, and the Δ48 and Δ110 deletions. **B** i & ii display skeletal staining of fore- and hindlimbs from mice homozygous for the Δ48 deletion. Arrows indicate where

there is digit loss and two terminal phalanges (i) and partial digit loss (ii). iii shows the *Shh* mRNA expression pattern in ∆48 hind limb bud. iv and v display skeletal staining of ∆110 fore- and hindlimbs respectively. Showing complete loss of a digit. vi shows *Shh* mRNA expression in ∆110 hind limb bud. Vii shows *Shh* mRNA expression in WT hind limb bud. Scale bars, i, ii, iv, v = 500 μM, iii, iv, vii = 100 μM. Images in **(A)** and **(B)** are adapted from Lettice *et al.*, (2017). Percentage of cells expressing *Lmbr1* in ∆48 **(C)** and ∆110 **(D)** E11.5 embryos. Percentage of cells expressing *Shh* in ∆48 **(E)** and ∆110 **(F)** E11.5 embryos. Error bars calculated by SEM, obtained from two tissue sections from each of two embryos. For each genotype n = ~250 nuclei (sum of all nuclei from all sections of both embryos analysed). *<0.05, **<0.01, ***<0.001 (independent t-test).

4.2.2. Removal of a Lim homeodomain binding motif effects transcriptional bursting in the ZPA

Previous research in the Hill lab has shown that ZRS priming occurs broadly across the distal limb bud mesenchyme, even outside of the ZPA, in tissue transcriptionally inactive for *Shh* (Peluso *et al.*, 2017). FGFs were identified as being crucial for this priming to occur, however, little is understood about how priming is activated and how this is linked to FGFs (Peluso *et al.*, 2017). Silvia Peluso (manuscript in preparation) conducted CUT&RUN and ATAC-seq on both distal and proximal limb bud cells for H3K4me1 and used this data to identify differential peaks in the sequencing data that were unique to the distal mesenchyme of the limb bud. By performing a Homer motif search on the unique sequences, binding motifs for the Lim homeodomain family, Isl1, Lhx1, Lhx2 and Lhx3 (unpublished results) were the most prominent within sequences.

Two of the Lim homeodomain binding sites were identified in the ZRS, which were named LhxA and LhxB (Figure 4.2A). Paul Devenney was able to edit these sites by CRISPR technology, utilising tracrRNA, Cas9 and an IDT repair oligo (eu.idtdna.com). Transgenic mice were created by pronuclear injection of plasmids into zygotes which are then implanted to a pseudopregnant mouse. Two mouse lines were created: $LhxA^{edit}$ and $LhxB^{edit}$. In each of these lines the motif binding sequence was changed from ATTA to GCCG (Figure 4.2A) to ensure loss of binding site activity without affecting spacing of the ZRS.

Transgenic embryos were also created by injecting constructs containing the mutated ZRS driving expression of a LacZ reporter gene. This allows for the observation of relative expression levels by β-gal staining. Staining was carried out by Laura Lettice on E11.5 embryos (Figure 4.2B). By this method

there appears to be a slight reduction in transgenic expression in LhxA^{edit} mice, in both the forelimbs and hindlimbs. LhxB ^{edit} mice had almost complete loss of transgenic expression in forelimbs and hindlimbs. Laura then also conducted whole-mount *in situ* hybridisations to look at *Shh* mRNA expression in both LhxA^{edit} and LhxB ^{edit} E11.5 embryos and their WT littermates (Figure 4.2C). In these *in situs* there is no discernible loss of *Shh* expression between the WT and LhxA^{edit} embryos. However, in the LhxB ^{edit} embryos there is a drastic reduction in *Shh* mRNA levels.

In order to determine if these observed differences in *Shh* expression result in a phenotype, Laura performed skeletal staining on E16.5 limbs (Figure 4.2D). No identifiable phenotype in LhxA^{edit} mice or, more surprisingly, LhxB ^{edit} mice was detected. However, while there was no skeletal phenotype, there was evidence of soft tissue syndactyly in the LhxB ^{edit} mice (Figure 4.2E). This phenotype occurred in both forelimbs and hindlimbs and showed no preference for particular digits.

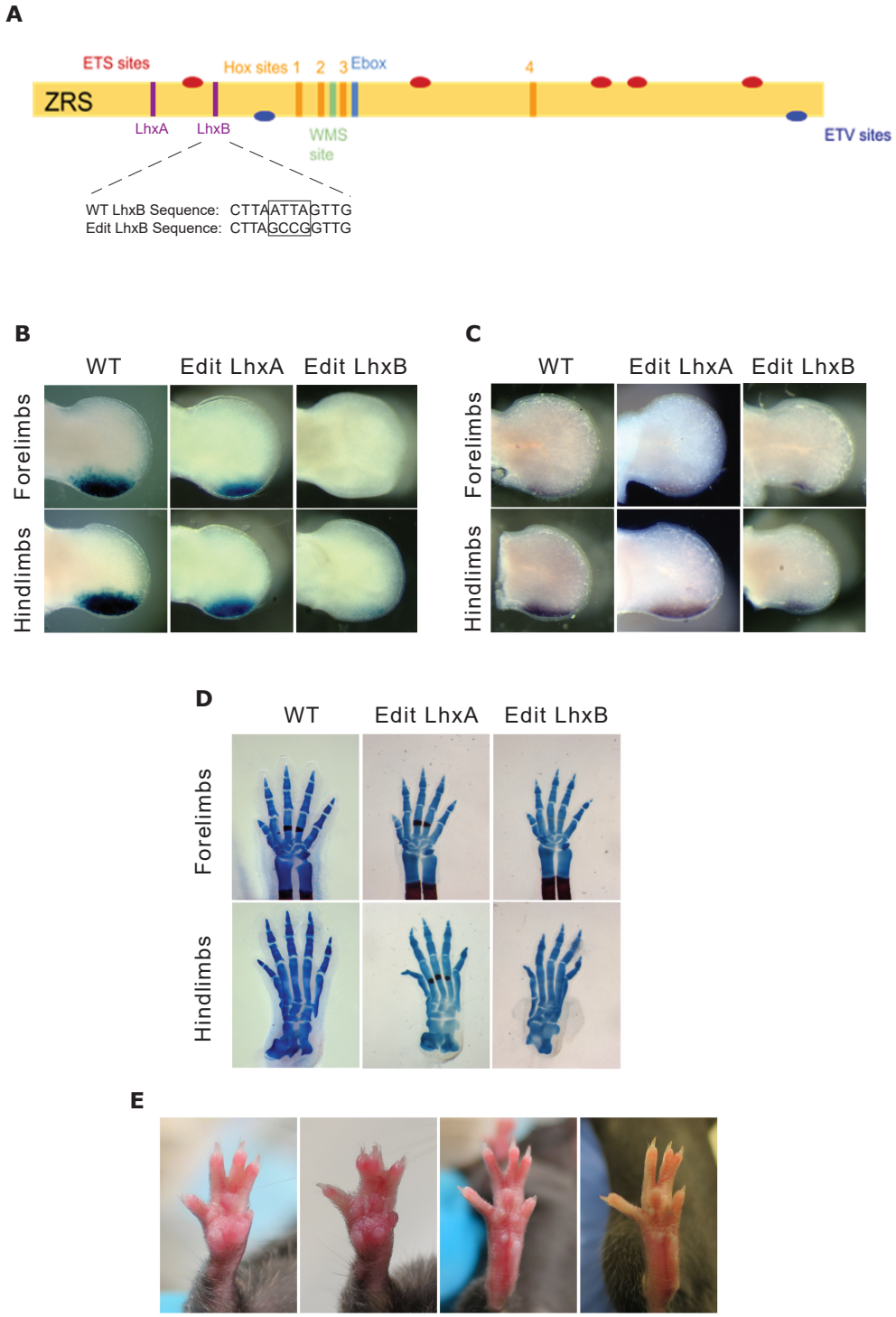


Figure 4.2 Phenotypic investigation of mutating LhxA and LhxB sites in the ZRS

A Graphical representation of ZRS showing relative locations of various binding sites, including those of LhxA and LhxB. Also included is the WT LhxB sequence and the new mutated sequence. **B** Limb buds from

transgenic embryos (E11.5) containing LhxA and LhxB mutations in the ZRS which is driving LacZ expression. **C** *In situ* hybridizations for *Shh* mRNA in E11.5 limb buds where the ZRS contains LhxA and LhxB mutation. **D** Skeletal staining of WT, LhxA and LhxB mutated ZRS mice. **E** Soft tissue syndactyly of LhxB homozygous mice. First two images show forelimbs and second two images show hindlimbs.

Given that LhxB^{edit} mice were the only ones that displayed a qualitative loss in *Shh* expression and any phenotype, I decided to explore this further using nascent RNA-FISH. Using RNA-FISH technology would allow me to identify a quantifiable loss in *Shh* expression. Furthermore, Lim homeodomain proteins are pioneer factors and therefore may have a different action on the ZRS compared to transcription factors such as the HOXD proteins, which act as transcription factors.

I performed RNA-FISH on three LhxB^{edit} embryos and three heterozygous littermates (Figure 4.3A). Heterozygotes were used as opposed to WT littermates as this was the tissue available to me, under time constraints. Two sections from each embryo were analysed. Floorplate tissue was also analysed as a control. Both forelimb and hindlimb ZPA tissue was analysed. *Lmbr1* was also used as an internal control.

When looking at RNA-FISH for *5' Lmbr1* I observed that there was no change in expression in the floorplate tissue between LhxB^{edit} and heterozygote embryos (Figure 4.3B) and confirms, as expected, that the mutation has no effect on *Lmbr1* transcription in the floorplate tissue. However, in the forelimb and hindlimb tissue, it initially appears that there may be a slight reduction in *Lmbr1* expression. In the hindlimb there is a reduction from 37.4% of cells expressing *Lmbr1* in the heterozygote to 29.6% of cells in the mutant. In the forelimb the heterozygote expresses *Lmbr1* in 35.8% of ZPA cells, whereas in the mutant this number is 26.5%. However, this difference was found to not be statistically significant (independent t-tests: hindlimb *p-value = 0.3*, forelimb *p-value = 0.07*). However, these numbers are still comparable to *Lmbr1* expression seen in WT in chapter 3 where at E11.5 28.6% of ZPA cells express *Lmbr1.*

As expected, *Shh* expression showed no difference in floorplate expression levels between the LhxB^{edit} tissue and heterozygous tissue (Figure 4.3C). Therefore, this shows once again that there is no variance between sections and embryos. When looking at forelimb tissue I recorded a drastic drop in *Shh* active transcription sites in the mutant when compared to the heterozygote. In the heterozygote 35% of cells expressed *Shh*, in the mutant this number dropped to 17.4% (independent t-test, *p-value = 0.01*). A similar drop was observed in the hindlimb. In the heterozygote hindlimb 35.1% of cells expressed *Shh*, in the mutant this number dropped to 18.9% (independent t-test, *p-value = 0.02*). These results are consistent with the soft tissue syndactyly phenotype observed in LhxB^{edit} mice. The fore- and hindlimb suffer a similar decrease in *Shh* expression, which would explain why the soft tissue syndactyly is not specific to either limb type.

Interestingly, when I compare *Shh* expression levels in LhxB^{edit} to expression levels in Δ48 and Δ110 mice they are the same (LhxB^{edit}: ~17%, Δ48: ~17%, Δ110: ~11%). This is unusual considering the differences in phenotype severity. LhxB^{edit} mice display no skeletal phenotype, only mild soft tissue syndactyly. Whereas, Δ48 and Δ110 mice have partial and full digit loss respectively. I expected that similar *Shh* bursting frequencies would result in similar phenotypes. A possible explanation for this was discussed earlier in 4.2.1. Possibly nascent RNA-FISH is captured differences in burst frequency but lost information about the total number of cells capable of expressing *Shh*. Given the results I have shown here it is likely that LhxB^{edit} effects the bursting capability of *Shh* while having no effect on the area of the ZPA, whereas the Hox site mutations effect the burst frequency, and the area of the ZPA in increasing magnitude depending on number of Hox sites removed. This finding indicates a difference on the action pioneer factors and regular transcription factors have on transcriptional activation. Hox proteins potentially effect both burst frequency and ZPA area, whereas pioneer factors such as Lim homeodomain proteins affect only burst frequency.

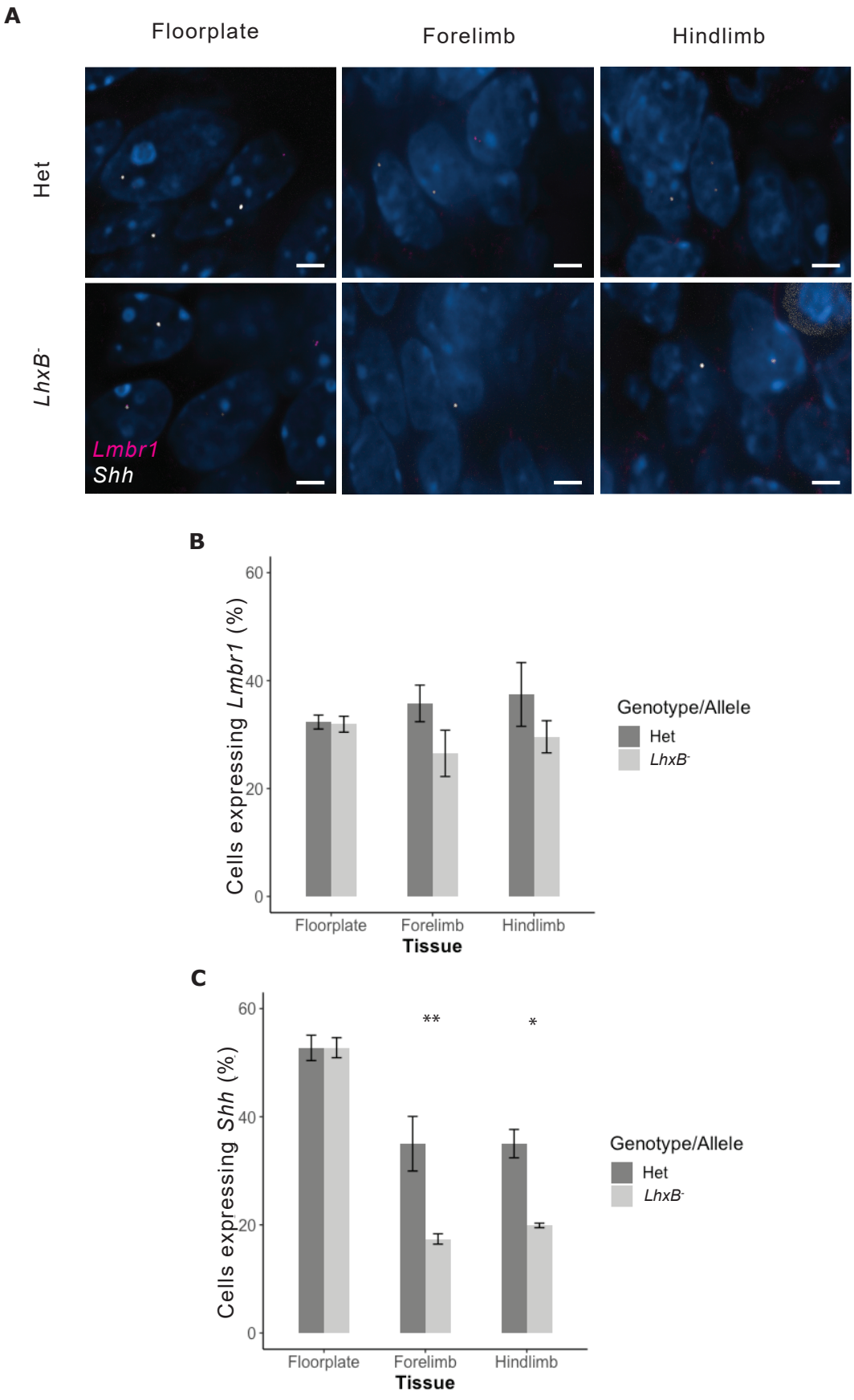

A
Floorplate
Forelimb
Hindlimb
Het
LhxB-
Lmbr1
Shh
B
Cells expressing Lmbr1 (%)
60
40
20
0
Floorplate
Forelimb
Hindlimb
Tissue
Genotype/Allele
Het
LhxB-
C
Cells expressing Shh (%)
60
40
20
0
**
*
Floorplate
Forelimb
Hindlimb
Tissue
Genotype/Allele
Het
LhxB-

Figure 4.3 Mutation of LhxB binding site in the ZRS results in reduced burst frequency in the ZPA

A Representative images of nascent RNA-FISH from both LhxB$^{+/edit}$ and LhxB ^{edit} in the floorplate, forelimb and hindlimb. DAPI staining is shown in blue, *Shh* in white and *Lmbr1* in magenta. Scale bars = 3.2 μM. **B** Percentage of cells expressing *5' Lmbr1* in floorplate, forelimb and hindlimb tissue. **C** Percentage of cells expressing *Shh* in floorplate, forelimb and hindlimb tissue. Error bars were calculated by SEM, obtained from two tissue sections from each of three embryos. For each genotype n = ~350 nuclei (sum of all nuclei from all sections of all embryos analysed). *<0.05, **<0.01, ***<0.001 (independent t-test).

4.3. Up-regulatory ZRS mutations and transcription

4.3.1. Inserting additional adenosines adjacent to the WMS repressor site does not affect ZPA bursting frequency

As previously mentioned Lettice et al. (2017) created mice containing a +AA insertion adjacent to the WMS repressor site. The WMS repressor site is crucial for repressing *Shh* expression outside of the ZPA. The mice develop a phenotype in the heterozygote through ectopic *Shh* expression in the distal anterior of the developing limb bud (Figure 4.4). Figure 4.4A demonstrates the skeletal formation of a WT mouse hindlimb, and Figure 4B shows the WT *Shh* mRNA expression pattern in an E11.5 limb bud. In comparison Figure 4C-E displays phenotypes of differing severities in +AA heterozygote hindlimbs. Fusion and duplication of internal digits (asterisks) can be seen in the heterozygote forelimb (Figure 4.4C) and bending of the hindlimb is also prevalent due to tibial shortening (Figure 4.4D).

The most severe phenotype is shown in Figure 4.4E, where the hindlimb shows bifurcation at the tip of the extra preaxial digit (arrowheads) and the centrally located extra digit (arrow). Figure 4.4F shows the *Shh* mRNA expression pattern in a +AA heterozygote E11.5 limb buds. When compared to the expression pattern in Figure 4.4B, there is ectopic expression in the whole distal tip of the limb bud that is not present in the WT. This results in the PPD and extra centrally located digit. Furthermore, in Figure 4.4G *Shh* mRNA expression in both fore- and hind- limb bud is shown. Ectopic expression around the distal edge and in the anterior hindlimb bud has become more exaggerated. However, there does not seem to be many discernible differences between expression in the ZPA when comparing to the WT, heterozygote and homozygote.

I used this mutant to explore the impact of an up-regulatory ZRS mutation on the transcriptional characteristics of *Shh* within the ZPA. I have now established that down-regulatory mutations to Hox sites and a pioneer factor binding site both decrease the bursting frequency of *Shh* within the ZPA. I will now determine if up-regulatory mutations, in contrast, cause an increase in burst frequency in the ZPA.

I conducted nascent RNA-FISH for *Shh* on forelimbs and floorplate of E11.5 +AA homozygotes and WT littermates (Figure 4.4H). Upon initial examination it appears that there is lower *Shh* expression in the WT littermates compared to the +AA homozygotes in the floorplate. In the WT we see expression in ~70% of cells and in the +AA homozygote this number is ~85%; however, this difference is not significant (independent t-test, *p-value = 0.07*). This confirms that the mutation in the ZRS does not impact *Shh* expression in the floorplate.

Interestingly I found that there is not a significant increase in the instances of active transcription sites in the ZPA between the WT and +AA homozygotes (Figure 4.4H; independent t-test, *p-value = 0.52*). In the WT there is active transcription in ~43% of cells. In the +AA homozygote ~50% of cells are expressing *Shh*. This is in contrast to that seen in down-regulatory ZRS mutations where *Shh* bursting frequency changes dramatically (Section 4.2). Though when we compare the *in situ* hybridisations seen in Figure 4B, F & G perhaps this finding is not surprising. These figures show the mRNA expression in WT, +AA heterozygote and homozygote limb buds. While there is clear ectopic expression around the distal edge of the heterozygote the expression in the posterior of the limb bud does not appear to have changed. This suggests *Shh* expression in the ZPA isn't affected by the mutation.

It is not currently understood what protein binds to the WMS repressor site, however, these findings make it clear that it has a different action on the ZRS and *Shh* transcription compared to Hox proteins and Lim homeodomain proteins. These findings suggest that WMS binding is crucial for the repression of *Shh* transcription in the entire distal edge of the limb bud outside of the ZPA. However, data from *in situ* hybridisations and RNA-FISH imply it does not have role in the suppression of *Shh* in the internal anterior of the limb bud. Overall, this suggests that the ZRS has new additional activity that is driving ectopic expression in developing limb bud but does not display an increase in activity in the ZPA.

4.3.2. Inserting additional adenosines adjacent to the WMS repressor site causes ectopic *Shh* transcriptional bursting in the anterior of the developing limb bud

I also looked at the ectopic *Shh* expression site in the anterior of the limb bud when conducting nascent RNA-FISH on these embryos. I ensured I was observing expression in the same anterior region of the limb bud due to the orientation of the embedded embryo. All embryos were embedded as previously described in section 3.2.1. I would choose to image the anterior that was near the end of, but still had AER present. This tissue was directly opposite of the ZPA tissue imaged.

As expected there was no *Shh* transcription sites observed in the anterior of the WT limb bud (Figure 4.4H). However, in the +AA homozygote I found that ~27% of anterior limb bud cells contained active transcription sites for *Shh* (independent t-test, *p-value = 0.006)*. This shows that the +AA insertion is causing ectopic transcription.

The data shown here suggest that perhaps two things are happening. Firstly, as previously discussed, the entire distal limb bud is primed for *Shh* expression. Disruption to the WMS repressor site results in the distal anterior limb bud now expressing *Shh*. However, these data suggest that the WMS site disruption has no effect on bursting frequency in the ZPA (the distal posterior limb bud). This highlights the WMS site functions as a region-specific repressor, in that it appears to only be impacting expression in one area of the limb bud. From previous work, *in situ* hybridisations have shown only ectopic expression in the distal edges of the limb bud (Figure 4.4F,G) (Lettice *et al.*, 2017), therefore perhaps whatever protein binds the WMS is upregulated by secreted factors from the AER, which is found on the distal edges of the limb bud.

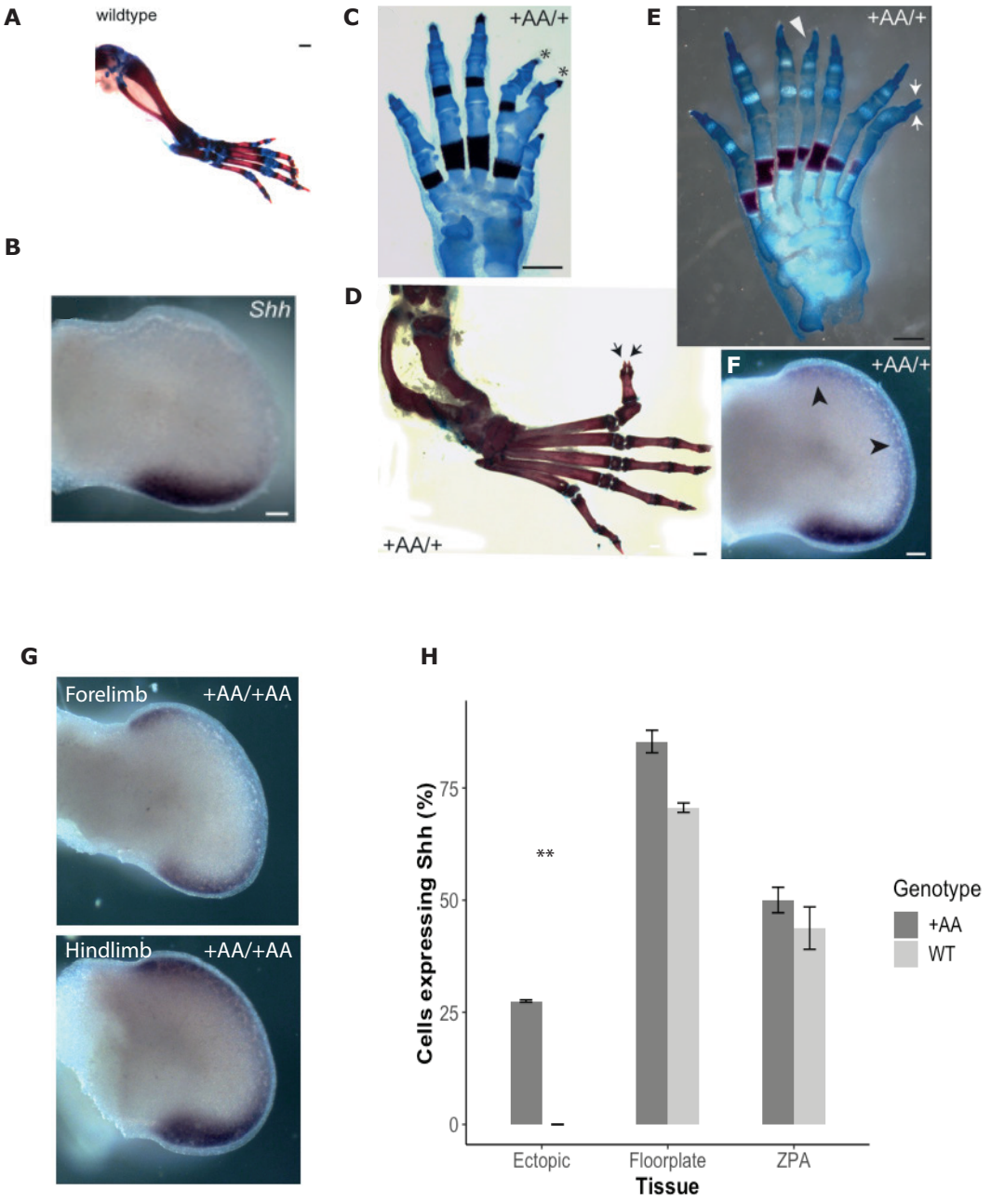


Figure 4.4 Adjacent insertion of two adenosines (+AA) to the WMS repressor site results in increased Shh expression.

A Skeletal staining of the wild-type hindlimb **B** wild-type *Shh* mRNA expression in an E11.5 hindlimb bud. Scale bar = 500μM. **C-E** Skeletal staining of +AA heterozygote hindlimbs showing differing phenotypes. Fusion and duplication of internal digits (asterisks) can be seen in the heterozygote forelimb. Bifurcation at the tip of extra preaxial digits (arrowheads) and

additional central digit (arrow). Images A - E adapted from Lettice et al.(Lettice *et al.*, 2017) Scale bars = 100 μM (B & F), 500μM (C & E), 1mm (D) **G** The *Shh* expression pattern in +AA homozygote E11.5 fore- and hindlimb buds. *In situ* hybridisations were carried out by Laura Lettice. **H** Percentage of cells expressing *Shh* in WT and +AA homozygotes tissues. Error bars calculated by SEM, obtained two embryos with two sections from each. For each genotype for each tissue n = 200 nuclei (sum of all nuclei from all sections of all embryos analysed). *<0.05, **<0.01, ***<0.001 (independent t-test).

4.4. Discussion and conclusions

In this chapter I have explored the action of three different transcription factor types on the *Shh* limb enhancer, ZRS. Firstly, I looked at Hox site binding, which has previously been reported to control levels of *Shh* expressed in the ZPA (Lettice *et al.*, 2017). I have proposed that HoxD proteins have two actions on the ZRS. Firstly, I showed that removal of two or more Hox sites resulted in a decrease in *Shh* burst frequency in the ZPA. Interestingly, I observed no difference between ∆48 and ∆110 *Shh* burst frequency. Looking at previous *in situs* (Figure 4.1B, Lettice et al., 2017) perhaps the area of the ZPA is decreasing which would not have been quantified by this analysis, further quantitative investigation is required. If confirmed it appears Hox site binding to the ZRS may also influence the number of cells that have the ability to burst.

I then studied the action of Lim homeodomain proteins, which are pioneer factors, on the ZRS. Mutation in the LhxB site resulted in an appreciable down-regulation of *Shh* expression in both the forelimbs and the hindlimbs. Interestingly, despite expression levels of *Shh* now being similar to those seen in the ∆48 and ∆110 mice, the phenotype observed was minimal by comparison. No skeletal malformations were observed, only a mild soft tissue syndactyl that showed no preference for particular digits or limbs. This highlights the different action of a Lim homeodomain proteins on expression compared to HoxD proteins. Perhaps, Lim homeodomain proteins influence the capability of an allele to burst, but do not affect the number of cells capable of doing so. There is also the possibility that Lim homeodomain proteins have a different effect on the timing of transcription. Perhaps, loss of HoxD binding effects the length of time *Shh* is being expressed in the ZPA negatively. Less overall time expressing *Shh* in ∆48 and ∆110 embryos could result in a more severe phenotype.

Finally, I have highlighted the action of the WMS site in *Shh* repression. Interestingly interrupting the binding site does not impact bursting frequency in the ZPA or appear to affect the ZPA area based on *in situ* hybridisations. Instead, it appears that the main impact of WMS disruption is ectopic expression of *Shh* around the distal anterior edge of the developing limb bud. These findings suggest that WMS site binding plays a crucial role outside of the ZPA in restricting *Shh* expression but within the ZPA does not have an effect on *Shh* expression.

4.4.1. HoxD binding of the ZRS has two different actions

Previously work from the Hill lab has characterised that Hox sites within the ZRS are crucial for defining the amount of *Shh* expressed from the ZPA (Lettice *et al.*, 2017). In this chapter I have described results that may extend the action of Hox proteins on the ZRS further. Firstly, I have shown that mutating Hox sites results in reduced burst frequency of *Shh* in both the ∆48 and ∆110 developing limb bud. Therefore, the first action of hox sites binding is defining the levels of *Shh* produced in the ZPA by affecting burst frequency.

Despite an increased number of hox sites being lost in the ∆110 mice compared to the ∆48, and a more severe phenotype as a result, I was unable to detect a difference in *Shh* burst frequency within the ZPA. This correlated with previous findings based on qPCR results (Lettice *et al.*, 2017). There are several possible reasons why this may be occurring.

Firstly, during imaging of nascent RNA-FISH in ∆48 and ∆110 mutants it appeared that the area of expressing cells had decreased. By looking at *in situ* hybridisations by Laura Lettice it appears that the stained *Shh*

expressing regions in Δ48 and Δ110 limb buds do become progressively smaller (Lettice *et al.*, 2017). This suggests a second action of Hox site binding, defining the overall number of cells that express *Shh.* Of course, the observations here do not measure the reduction in area quantitively and therefore further investigation is needed. One method to determine this would be to examine allele specific expression. An experiment to examine this would be to cross heterozygote mice with GFP tagged *Shh* and heterozygote mice for one of the ZRS mutants, such as Δ110. Under the assumption that the ZRS is *cis*-acting then a WT ZRS will be driving expression of the Shh-GFP and a Δ110 ZRS will be driving expression of a WT *Shh*. Nascent RNA-FISH could then be used in conjunction with immunofluorescent staining. If the hypothesis that reduction in Hox site binding leads to reduction in ZPA area I would predict that GFP expression would be seen across the ZPA showing the WT expression pattern. Nascent RNA-FISH staining for WT *Shh* would be viewed in a reduced area, within a subset of the GFP stained cells.

A second possibility, not necessarily mutually exclusive to the above, is that different HoxD proteins that bind to the ZRS Hox sites have different roles. Previously Lettice and colleagues have demonstrated that the HoxD proteins that bind the ZRS have slightly different preferences for which hox site they bind through electromobility shift assays (EMSAs) (Lettice *et al.*, 2017). In the context of the Δ48 and Δ110 deletions, Hox sites 1-3 are those effected. Site 1 is preferentially bound by HoxD9 and 11, and weakly by HoxD10. Site 2 is preferentially bound by HoxD11 and weakly by HoxD10 and 11. Finally site 3 is bound by HoxD9, 10 and 11 (Lettice *et al.*, 2017). HoxD 12 and 13 were also shown to bind with lower affinity to all sites. Emerging works have highlighted the differing roles of Hox proteins, for example HoxD13 has been proposed to have pioneer-like activity, binding more inaccessible chromatin, compared to proteins like HoxD9 which bind accessible chromatin (Desanlis *et al.*, 2020). Therefore, it stands to reason that each Hox site may play slightly different roles in defining *Shh* expression.

A final possible explanation for the phenotype variations in Δ48 and Δ110 limb buds is that the method I have used does not differentiate burst sizes. The subsequent removal of an additional Hox site could result in reduced burst size, explaining the difference in phenotypes observed. Further work would be necessary to test this. One method that could determine bursting size would be allele-sensitive single-cell sequencing as described by Larsson and colleagues (Larsson *et al.*, 2019). Unfortunately, this method would not be successful in *Shh* as there are no SNPs in the mature RNA in mouse strains or close Mus species on which to differentiate the alleles. An alternative would be to use live-cell imaging utilising MS2 tagged *Shh* transcripts. The numbers of transcripts in a single burst can be estimated based on the intensity of spot fluorescence (Chubb *et al.*, 2006). Therefore, any change in transcripts produced in a single burst could be quantified.

4.4.2. Interruption to Lim homeodomain binding reduces *Shh* bursting frequency

Looking at the binding sites of Lim homeodomain proteins has highlighted the different functions of pioneer factors on an enhancer compared to activating transcription factors such as the Hox proteins. It should be noted that editing the LhxB site of the ZRS did not knockout all pioneering function, suggesting that other transcription factors that bind the ZRS can act as pioneer factors. I have demonstrated the mild syndactyly phenotype observed in $LhxB^{edit}$ homozygotes is not explained by an increase in burst frequency of *Shh* when compared to *Shh* burst frequency in Δ48 and Δ110 embryos. As previously stated limb bud development is sensitive to fluctuations in *Shh* transcript levels, therefore, this limited phenotype would imply that the overall level of *Shh* transcripts will not have changed dramatically. One possible reason could be that Lim homeodomain factor binding is having an effect on the burst frequency but not the area of the ZPA at the E11.5 stage. Perhaps the

loss of Hox sites results in reduction of time *Shh* is 'on'. This could explain the area of the ZPA appearing reduced, as *Shh* is being switched 'off' earlier.

If restriction of the ZPA area were a valid explanation for the differences in phenotype between LhxB^{edit} and ∆48 and ∆110 mice, then I would not have been able to pick up these changes analysing nascent RNA-FISH with the methods I have used. In the future, RNAscope could be utilised to determine if the area of the ZPA and therefore number of cells expressing *Shh* is affected. RNAscope allows for visualisation of mature transcripts and through collaboration with the IGC Advanced Imaging Resource I was able to determine an imaging technique that allowed for several images to be stitched into one. Had more time been available I could have quantified the number of cells that are expressing *Shh* and the number of transcripts.

Interestingly, previous work implicated Lhx2 as a mediatory protein in the Shh-Fgf feedback loop (Watson *et al.*, 2018). Lhx2 is a Lim homeodomain protein. The study found that implantation of Fgf beads into the posterior limb bud resulted in an upregulation of Lhx2. Furthermore, injection of a Lhx2 overexpression vector into developing chick forelimbs was not sufficient to induce *Shh* expression alone, however, used in conjunction with Fgf beads *Shh* expression was increased (Watson *et al.*, 2018). Additionally, reducing Lhx2 at the ZPA via knockdown using an Anti-Lhx2 morpholino, resulted in reduced Shh (Watson *et al.*, 2018). These findings are consistent with the data shown here. The data generated by other lab members has shown directly that the ZRS contains Lim homeodomain binding sites, through CUT&RUN and ATAC-Seq. My work with RNA-FISH has shown mutations of these Lim homeodomain sites decreases *Shh* expression.

4.4.3. Disruption to the WMS repressor site leaves *Shh* expression in the ZPA unaffected

Interestingly, disrupting the WMS repressor site in the ZRS did not impact *Shh* burst frequency in the ZPA suggesting that it plays a region-specific role. There was ectopic expression of *Shh* and therefore an increase in burst frequency in the distal anterior and distal margin on the limb bud. This would suggest that polydactyly and additional central digits described in previous literature (Lettice *et al.*, 2017) were caused solely by ectopic *Shh* expression.

The pattern of ectopic expression may suggest whatever factor binds the WMS site may be induced by secreted factors from the AER. The AER is the tissue directly adjacent to the ectopic *Shh* expressing area and produces a number of growth factors (Fernandez-Teran and Ros, 2008). A potential method to determine the repressor binding to the WMS site would be an electrophoretic mobility shift assay (EMSA). Members of the Hill lab have previously attempted this extensively and were unable to identify specific binding to the site.

Clearly, up-regulatory mutations of the ZRS have a different action than down-regulatory mutations. Down-regulatory mutations result in a reduction in burst frequency of *Shh* in the ZPA. However, up-regulatory mutations, such as +AA, does not lead to an increase in burst frequency in the ZPA. Instead ectopic expression occurs.

Chapter 5 Long-range regulation of distal enhancer-promoter interactions

5.1. Introduction

Vast linear distances are found between some enhancers and their promoters, such as the *Shh* promoter and the ZRS (~1Mb) (Lettice *et al.*, 2003). This creates much speculation about how they can be brought close together in 3D space to allow interaction. The two prevailing models for explaining distal interaction between enhancers and promoters are looping and transcriptional hubs (Lim and Levine, 2021). The transcriptional hub model incorporates looping as well as some form of tracking and/or linking and suggests that foci of the transcriptional machinery form *de novo* at points of enhancer-promoter proximity induced by looping. A recent review has established criteria to test experimentally for transcriptional hubs (Lim and Levine, 2021). Firstly, transcription of a gene must occur in bursts. Notably, a single enhancer must be capable of co-activating transcription of linked reporter genes. Co-activation occurs when an enhancer is able to simultaneously drive expression of two or more genes. Finally, in the transcriptional hub model distance between enhancer and promoter is not important, therefore relatively large distances (>200 nm) can separate enhancers and promoters during activation (Lim and Levine, 2021). Therefore, I shall explore each of these parameters in the context of the *Shh* locus here, to further understand the model by which enhancer and promoters communicate over large linear distances.

As mentioned in the introduction, previous experiments in *Drosophila melanogaster* identified that an enhancer is capable of driving simultaneous expression of two reporter genes (Fukaya, Lim and Levine, 2016). However, in this example reporter genes were linearly equidistant from the enhancer which does not highlight whether co-activation is possible when distances from the enhancer are asymmetrical. Furthermore, as reporter genes are

used there is no evidence of co-activation endogenously. I would hypothesise in the case of the ZRS, that if a transcriptional hub model were valid it would be able to drive expression of *Shh* and a reporter gene simultaneously. However, if it were only able to drive the expression of one gene at a time this would suggest a looping model as the mode of enhancer-promoter interaction. Therefore, the main question addressed in this chapter is: can the ZRS drive transcription of multiple genes simultaneously?

5.2. Action of a single enhancer on multiple target promoters

The transcriptional hub hypothesis postulates that a single enhancer is capable of communicating with multiple promoters simultaneously (Lim and Levine, 2021). This model assumes that direct, intimate contact between the promoter and enhancer is not necessary. Instead, pre-looped topologies, such as a TAD, cause the formation of "hubs" that contain all the necessary proteins to drive transcription (Furlong and Levine, 2018). This allows multiple genes to be expressed at once, driven by a single enhancer. Alternatively, the looping model proposes that cohesin-driven loop extrusion allows an enhancer and promoter to make close contact, allowing the enhancer to drive transcription. Therefore, an enhancer can only interact with one promoter at a time. Interestingly, a recent study has shown the presence of loop-stacking (Hafner *et al.*, 2023). Using optical reconstruction of chromatin architerture (ORCA) imaging in a multiplexed fashion the group were able to show that loops of chromatin stack together, held together by cohesin, to form a hub of CTCF border sites. Perhaps this is a method of transcriptional hub formation.

I explored this hypothesis by looking at multiple systems. Firstly, I determined if a single enhancer, ZRS, could drive *Shh* expression and that of a reporter gene, LacZ, simultaneously. I then investigated if other *Shh* enhancers, SFPE1 and MACS1, which are located at different distances from *Shh* and a *lacZ* reporter gene, were able to drive simultaneous expression of the two genes. Furthermore, I chose to look in an endogenous system to explore whether the ZRS could co-activate *Shh* and *Mnx1.*

5.2.1. Introduction to the SBLac526 mouse line

In the first instance I decided to investigate if *Shh* enhancers could co-activate multiple promoters within the *Shh* TAD in mouse embryos. To do this I used nascent RNA-FISH, applying probes designed for *Shh* and *lacZ*, on tissue from a mouse containing a LacZ reporter gene.

Previously in the Hill lab a reporter mouse was developed with the aim of mapping the *Shh* regulatory locus (Anderson *et al.*, 2014). The SBLac526 line contains a *lacZ* reporter gene within the *Shh* gene desert. The reporter gene is fused to a *Hsp68* minimal promoter which is a basal promoter and therefore requires an enhancer to drive expression.

This mouse was developed using the Local Hopping Enhancer Detection System (LHED) (Kokubu *et al.*, 2009), a system based on the Sleeping Beauty (SB) transposon. The SB system contains two key components: a synthetic gene encoding a transposase, and inverted/direct repeats which are targeted by the transposase (Ivics *et al.*, 1997). SB has a strong tendency to reinsert at a locus linked to the donor site, known as “local hopping”. The LHED system uses knock-in technology, transposon based enhancer detection, and Cre-/LoxP-mediated chromosomal engineering to allow for monitoring of enhancer activity at a targeted region of the genome (Kokubu *et al.*, 2009).

The method used by Anderson et al. (2014) to create the SBLac526 line can be seen below in Figure 5.1A. Anderson et al., (2014) designed the LHED vector to contain no *BamHI OR HindIII* restriction enzyme sites to allow any genomic region containing these sites to be chosen as a homology arm containing these unique restriction sites. In order to increase the likelihood of homologous recombination a region of ~7 kb was selected of reasonably

unique DNA. For this reason, the gene desert of the *Shh* locus was avoided as it contains large amounts of repetitive DNA. The two homology arms were designed with a prepositioned gap of ~600 bp to improve recombination efficiency (Figure 5.1A(i)). The LHED vector was then linearized in the gap region and electroporated into E14 mouse Embryonic Stem (ES) cells. PCR analysis and DNA-FISH was then used to confirm correct targeting in E14 ES cell clones (Figure 5.1A(ii)). The LHED vector was incorporated into the TAD using homologous recombination and then the LacZ reporter gene was "hopped" using transposase (Figure 5.1A(iii)). Following successful insertion of the LHED vector into the ES cell genome, ES cell clones were injected into tetraploid blastocysts which were subsequently used to create embryos wholly from ES cells (Anderson *et al.*, 2014). The use of tetraploid blastocysts is an advantage as the tetraploid cells will only contributed to the placenta and yolk sac, and not to the somatic cells of the embryo, and therefore pups produced are compromised wholly from ES cells. Cre recombinase treatment can then be used to delete remaining elements of the construct such as the neomycin and puromycin resistance genes and the PGK promoters whose control they are under. However, this step was not carried out in the SBLac526 line meaning these elements are still contained within the genome. Potentially, this means that there are up to four promoters which the *Shh* enhancers could be driving in their cognate tissues. If the enhancers can only communicate with one promoter at a time this would potentially allow for an enhancer to divideits time between more promoters. However, if this is valid any decreased expression of *Shh* is still an indicator that a *Shh* enhancer is only communicating with one enhancer at a time. It should also be noticed that the SBLac526 mice do not have a phenotype.

In Figure 5.1B there is a comparison between β-galactosidase staining of a SBLac526 embryo and *Shh in situ* hybridisation staining of a WT embryo. The reporter gene shows similar expression pattern to *Shh*, presumably because it is being driven by the relevant *Shh* enhancers in their specific

tissues. Therefore, in the ZPA *lacZ* expression is driven by the ZRS and in the floorplate by SFPE1. Figure 5.1B also gives a diagrammatic representation of where the LacZ reporter gene is located in the *Shh* TAD and approximate distances from various enhancers. Furthermore, SBLac546 embryos do not display any phenotypes in homozygotes, suggesting that manipulation of the genome to insert the reporter gene has not impacted expression of gene in the *Shh* TAD.

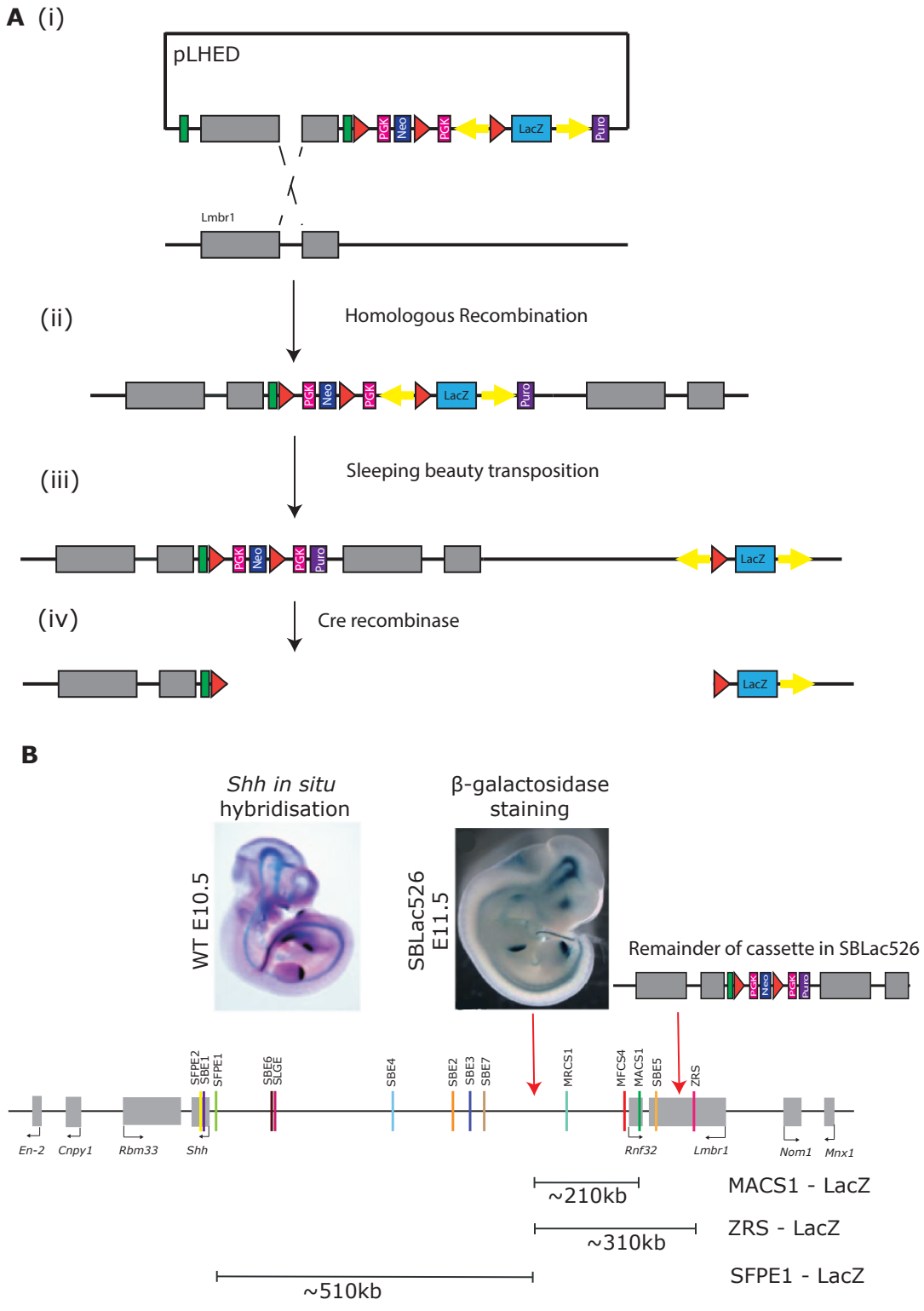


Figure 5.1. The LHED transposon system and SBLac526 mice

A (i) The LHED system comprises the pLHED vector containing a *LacZ* reporter gene (light blue rectangle), IR/DR repeats (Yellow arrows), and G418 selection cassette with promoter (dark blue rectangle), puromycin selection cassette (pink rectangle), PGK promoter (purple rectangle), *LoxP* sites (red triangle) and FRT sites (green rectangles). Firstly, a homology arm for region of interest complete with pre-positioned gap was cloned into the vector. The vector was then linearized within the gap and targeted into the genomic locus via homologous recombination which can be selected for using neomycin (ii). (iii) Once inserted within the genome the transposon elements consisting of the *LacZ* gene flanked by the SB IR/DR can be mobilised by transient expression of a transposase. Following transposition, the PGK promoter comes in contact with the puromycin resistance gene meaning puromycin can be used to select for transposition. The transposon element will then reinsert within the genome within approximately 50% of cases and the *LacZ* staining will reflect nearby enhancer function. The above was conducted in ES cells to create clones which were then injected into tetraploid blastocysts which were subsequently used to generate embryos. (iv) Cre recombinase can also be utilised to generate deletions between the LoxP sites of the original insertion site and the transposed site which can then be utilised to detect loss of function phenotypes. The LacZ gene also remains present so alterations to enhancer activity can also be analysed. (Adapted from the thesis of Eve Anderson, 2014). **B** Diagrammatic representation of the *Shh* regulatory locus showing the locations of the *lacZ* reporter gene and the remainder of the LHED cassette in the SBLac526 line. A comparison of β-galactosidase staining of SBLac526 embryo and a WT Shh *in situ* hybridisation is shown. The distances between *lacZ* and various *Shh* enhancers is also indicated.

5.2.2. Enhancers cannot simultaneously co-activate *Shh* and LacZ reporter gene in cognate tissues

To determine if ZRS can co-activate expression of *Shh* and *lacZ,* nascent RNA-FISH was performed on SBLac526 homozygotes and WT littermates at stage E10.5 (Figure 5.2A). Three different tissues, acting under different enhancers were assessed. I looked at the ZPA, floorplate and hindgut. The ZPA is driven by the ZRS, the floorplate is driven by SFPE1 and SFPE2, and the hindgut is driven by MACS1. SFPE1 and SFPE2 behave redundantly (Epstein, McMahon and Joyner, 1999). These tissues were chosen as it allowed me to determine what effect linear distance between an enhancer and promoter has on distal interactions. MACS1 is located most proximal to the *lacZ* reporter gene, at approximately 210kb away. This is followed by the ZRS at ~310kb and then SFPE1 at ~510kb.

I found that there were almost no instances where *Shh* and *lacZ* were observed being expressed in the same cell on the same chromosome in any of the three tissues tested (Figure 5.2B). Figure 5.2B shows the total percentage of cells that expressed *lacZ*, total percentage of cell expressing *Shh* and finally the total percentage of cells expressing both *lacZ* and *Shh* on the same chromosome. In the floorplate *lacZ* expression was observed in 20.2% of cells, for *Shh* expression this number was 54.2% of cells. Therefore, both genes are being expressed in floorplate tissue. However, when I look for instances of both *lacZ* and *Shh* being expressed from the same chromosome this number is only 0.6%.

Through collaboration with Hannes Becher (Bioinformatics Core Statistician at the IGC) a method to determine the significance of linked co-expression was developed. We decided that linked expression of two alleles is akin to testing for linkage disequilibrium (LD). LD is the non-random association of two alleles of different loci (Slatkin, 2008). Bursting of two alleles in this

instance is analogous to linkage of two alleles. Therefore, correlation between two alleles bursting can be measured:

$$r = \frac{p(AB) - p(A) * p(B)}{\sqrt{\left(p(A) * p(a) * p(B) * p(b)\right)}}$$

Where: *p(A)* is defined as the observed probability of allele 'A' bursting for marker 1; *p(a) = 1-p(A)* is defined as the observed probability of allele 'a' not bursting for marker 1; *p(B)* is defined as the observed probability of allele 'B' bursting for marker 2; and *p(b) = 1- p(B)* is defined as the observed probability of allele 'b' not bursting for marker 2; and *p(AB)* is defined as the probability of the marker allele pair 'AB' bursting simultaneously.

Therefore, to determine if there is any correlation between bursting of *Shh* and *lacZ* in floorplate cells the above equation was assessed. A negative correlation between bursting of these alleles was found (Chi-squared tests, *p-value = 0.0004*). This suggests that the two alleles are bursting together less frequently than would be expected at random. The transcriptional hub hypothesis predicts that SFPE1 should be able to drive transcription of both of these genes in unison which is not corroborated by these data. This suggests that the SFPE1 only interacts with one promoter at a time. While the transcriptional hub model does not postulate that an enhancer must drive expression of two genes simultaneously in every instance, it would be expected that the two genes would be observed bursting together more frequently than would be expected random. These data show the two genes being transcribed simultaneously less frequently than would be expected at random, suggesting SFPE1 is only driving one promoter at a time.

The hindgut and the ZPA show a similar trend. In the hindgut 20.2% of cells are expressing *lacZ* and 26% of cells are expressing *Shh*. However, only

0.96% of cells show co-activation of the two on the same chromosome. Similarly, in the ZPA, 26.8% of cells express *Shh* and 28.2% of cells express *lacZ,* but only 2.4% express *lacZ* and *Shh* on the same chromosome. Again, testing for LD there is a negative correlation between the alleles in both hindgut and ZPA tissue (Chi-squared test, *hindgut p-value = 0.003, ZPA p-value = 0.001*). Therefore, these data suggest that MACS1 and ZRS predominantly communicate with one promoter at a time.

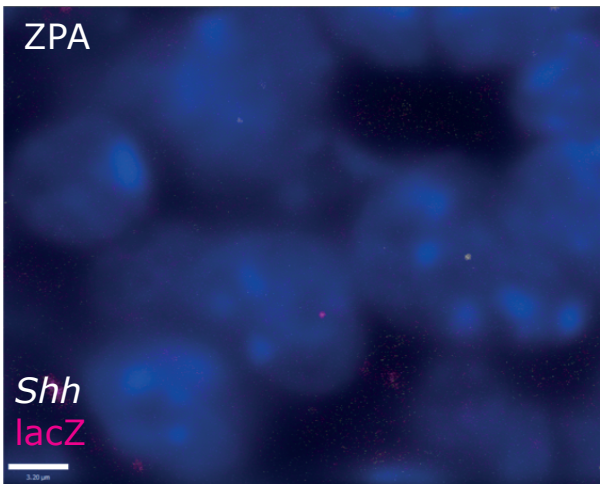
If an enhancer was driving transcription via a transcriptional hub I would expect a positive correlation between bursting *Shh* and *lacZ* alleles. A shared enhancer driving their transcription simultaneously would result in higher instances of the alleles being co-activated. These data support a looping model for distal enhancer-promoter interactions indicating that the *Shh* enhancers; SFPE1, MACS1 and ZRS, cannot co-activate *Shh* and *lacZ* in their cognate tissues and that the enhancers can only interact with one promoter at a time.

5.2.3. Enhancers have a set activity level when interacting with *Shh* and *lacZ*

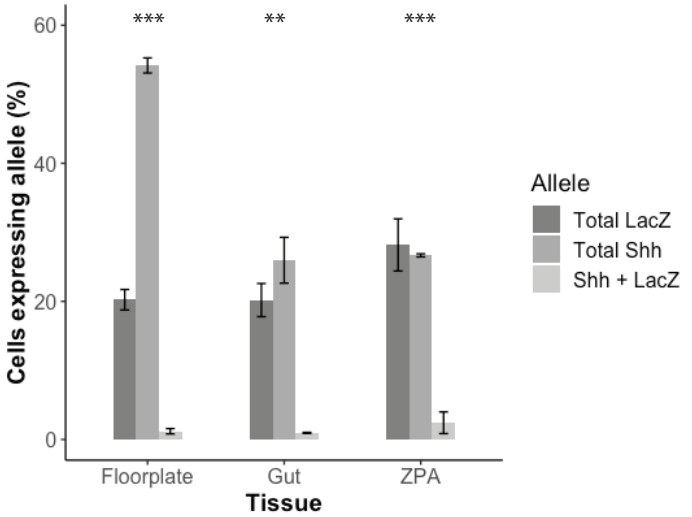
Interestingly, I noticed that when I compared the percentage of cells expressing *Shh* in the SBLac526 mice to WT littermates there was a drop in the levels of *Shh* expression in all three tissues (Figure 5.2C). In the WT ZPA and WT hindgut tissue ~47% and ~50% of cells were expressing *Shh*, respectively. In the SBLac526 mice these numbers had fallen to ~26% in both tissues (ZPA; independent t-test, *p-value = 0.008,* hindgut; independent t-test, *p-value = 0.02*). This drop in *Shh* expression is roughly equivalent to the *lacZ* expression now observed. There is a drop of ~21% *Shh* in the hindgut, with *lacZ* now being expressed in 20.2% of cells. Similarly, in the ZPA there is a ~24% drop in *Shh* expression, with *lacZ* now being expressed in 28.2% of cells. The WT floorplate ~72% of cells are expressing *Shh*, with

this number dropping to 54.2% of cells in the SBLac526 floorplate (independent t-test, *p-value* = $1.65e^{-05}$). This is ~18% fewer cells expressing *Shh*, roughly equivalent to the 20.2% of cells now expressing *lacZ*. This suggests that all three enhancers have a set activity level and that by introducing another target promoter that activity level has been split between the promoters. Again, this is suggestive of a looping model for distal enhancer-promoter interactions as it shows the enhancers activity is being split between the promoters.

A


ZPA
Shh
lacZ

B


***
**
***
60
40
20
0
Cells expressing allele (%)
Floorplate
Gut
ZPA
Tissue
Allele
Total LacZ
Total Shh
Shh + LacZ

C

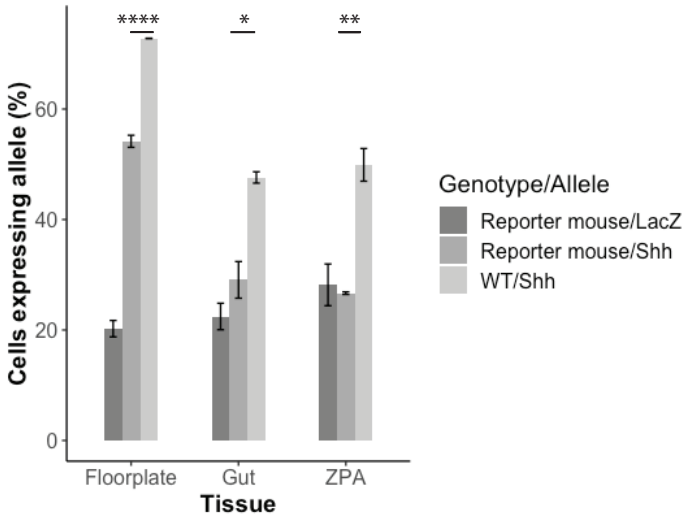

****
*
**
60
40
20
0
Cells expressing allele (%)
Floorplate
Gut
ZPA
Tissue
Genotype/Allele
Reporter mouse/LacZ
Reporter mouse/Shh
WT/Shh

Figure 5.2. RNA-FISH reveals that Shh enhancers rarely co-activate Shh and a lacZ reporter gene

A Representative image of RNA-FISH conducted in the ZPA. DAPI staining is shown in blue. *Shh* is shown in white and *lacZ* is shown in magenta. Scale bar = 3.20 μM. **B** Bar chart representing the total percentage of cells expressing *Shh* and *lacZ*, and percentage of cells expressing *Shh* and *lacZ* in the same cell on the same chromosome. Results shown for floorplate, hindgut and ZPA tissue. *<0.05, **<0.01, ***<0.001, ****<0.0001 (LD, Chi-squared test). **C** Bar chart comparing the percentage of cells expressing *Shh* and *lacZ* in SBLac526 mouse tissues to WT *Shh* expression. *<0.05, **<0.01, ***<0.001, ****<0.0001 (independent t-test). Error bars calculated by SEM, obtained from two tissue sections from two embryos. For each tissue there is n = ~300 nuclei (sum of all nuclei from all sections of both embryos analysed).

5.2.4. ZRS can simultaneously co-activate *Shh* and *Mnx1* in developing limb bud

Previous research that explored an enhancers ability to drive transcription of multiple promoters was conducted solely on reporter genes, positioned equidistant from the enhancer (Fukaya, Lim and Levine, 2016). In the experiments in section 5.2.2 I was able to examine an enhancer with its endogenous gene and a reporter. This does not tell me what would occur in a fully endogenous system, where a single enhancer is communicating with two target promoters in the same tissue.

I wanted to explore the action of an enhancer on multiple genes in an endogenous context. Previous work from the Hill and Bickmore labs has suggested that the ZRS drives transcription of not only *Shh* but a gene in an adjacent TAD, *Mnx1* (Williamson *et al.*, 2019b). *Mnx1* is expressed in the ZPA suggesting shared regulation with *Shh* in this tissue. Williamson et al. (2019) generated a 35 kb deletion using CRISPR guides to create ES cells clones which were then injected into blastocysts. This mouse line, ∆35, removed CTCF sites 4 and 5 of the *Shh* TAD, resulting in a shifted TAD boundary and increased contact between the ZRS and *Mnx1* (Figure 5.3A, B). These mice displayed no phenotype and *Shh* expression levels were unaffected (Williamson *et al.*, 2019b). By DNA-FISH they observed a decrease in distance between the ZRS and *Mnx1* alleles. They also noted a modest but not significant increase in *Mnx1* expression. Therefore, I decided to investigate whether the ZRS was capable of co-activating *Shh* and *Mnx1* in both ∆35 homozygotes and WT littermates.

I conducted RNA-FISH on E11.5 ∆35 homozygote embryos and WT littermates (Figure 5.3C). Both ZPA and floorplate tissue were assessed. I would expect there to be no *Mnx1* expression in the floorplate tissue, as this has not previously been observed in the expression pattern of *Mnx1* (Hagan

et al., 2000). As expected there is no *Mnx1* expression observed in the floorplate in either the WT or ∆35 homozygote mice. Interestingly there is no effect on *Shh* expression in the floorplate despite the disruption to the *Shh-Mnx1* TAD boundary (Figure 5.3D, E). In the WT floorplate tissue 61.4% of cells were expressing *Shh*, in the ∆35 homozygotes this number was 64.4% (independent t-test, *p-value = 0.61*). This is perhaps not surprising as the *Shh* floorplate enhancers SFPE1 and SFPE2 are located directly adjacent and within *Shh* itself. Therefore, the enhancers do not need to interact distally, and disruption to TAD structure should not affect their distance from their target promoter. Though it should be noted despite the proximity these enhancers have to the promoter, previous work from the Hill lab has shown they do have the ability to drive transcription distally (Anderson *et al.*, 2014). Therefore, this interaction with *Mnx1* may be a property of just ZRS, and not all *Shh* enhancers.

Enhancer-promoter specificity has been a topic that has garnered much interest over the years, and while not an extensive list, several key hypotheses have been proposed as an explanation. A first determinant in enhancer-promoter specificity is promoter accessibility. In order for a promoter to be activated it must be found in 'open' chromatin to allow occupation by the pre-initiation complex. A striking example of this is during spermatocyte differentiation. Using a combination of RNA-seq, CAGE and ATAC-seq to map chromatin accessibility, one group showed that expressing genes during the progression from proliferating spermatogonia to differentiating spermatocytes, displayed 'closed' promoters in the spermatogonia stage, and 'open' promoters in the spermatocyte stage. This highlights that promoter need to be in 'open' chromatin state to allow gene expression (Lu *et al.*, 2020). Of course, an accessible promoter does not necessarily mean that it can be driven by any nearby enhancer. A second determinant of enhancer-promoter specificity is the sequence of the core promoter. One classic study in drosophila used homologous recombination to

replace the gene *oaf*'s promoter with the promoter of the *dpp* gene (Merli *et al.*, 1996). Despite the *dpp* enhancers not usually being able to drive *oaf* transcription, when the *dpp* promoter was attached to the *oaf* gene, transcription could then be driven by the *dpp* enhancers (Merli *et al.*, 1996). This indicates that promoter sequence also confers enhancer-promoter specificity.

When I looked at ZPA tissue I found a few interesting observations. Firstly, in the WT I observed 19.4% of cells expressing *Mnx1* and 46% of cells expressing *Shh*. When I analysed which cells were co-expressing *Shh* and *Mnx1* I found that 10% of cells had linked expression (Figure 5.4D). Therefore, approximately half the instances in which *Mnx1* is being expressed, it is being co-expressed with *Shh*. Testing for LD the positive correlation between *Shh* and *Mnx1* is significant (Chi-squared test, *p-value* = *0.0005*). The alleles are bursting from the same chromosome more often than would be expected, suggesting that ZRS is capable of driving expression of both simultaneously.

When observing the ∆35 homozygous ZPA, I found that *Shh* expression had not changed compared to WT (Figure 5.3E) (independent t-test, *p-value* = *0.8*). This confirms previous findings from the Hill lab which suggest *Shh* expression is robust to TAD perturbations (Williamson *et al.*, 2019b). Unlike previous evidence from the Hill lab (Williamson *et al.*, 2019b), I found that there is a significant increase in *Mnx1* expression in the ZPA. In the WT ZPA 19.4% of cells expressed *Mnx1*, while in the ∆35 homozygote this number increased to 41% of cells (Figure 5.3D, E)(independent t-test, *p-value* = *0.04*). A Likely reason that I was able to identify a significant change, when Williamson et al. (2019) were unable, was that RNA-FISH is a more sensitive technique for measuring changes in expression compared to qPCR. RNA-FISH enables a single cell analysis which I argue presents a more accurate

picture of the ZPA than whole limb bud analysis by PCR. There is also an increase in the instances of *Shh* and *Mnx1* being co-activated. In the WT this number had been 10% of cells, in the ∆35 homozygotes 26.1% of cells are co-expressing *Shh* and *Mnx1* from the same chromosome. Again, LD finds bursting of *Shh* and *Mnx1* is positively correlated (Chi-squared test, *p-value* = $6.04x10^{-14}$).

Finally, considering ∆35 homozygote and WT littermates ZPA results together, the activity level of the ZRS has increased when the TAD boundary is shifted (Figure 5.3F). *Shh* expression is unaffected but *Mnx1* expression has increased, with a portion of this increase being co-activation with *Shh*. This indicates that the ZRS can drive higher levels of transcription, consistent with a transcriptional hub model which suggests that a) enhancers can talk to multiple promoters simultaneously; b) without impinging on overall transcription levels.

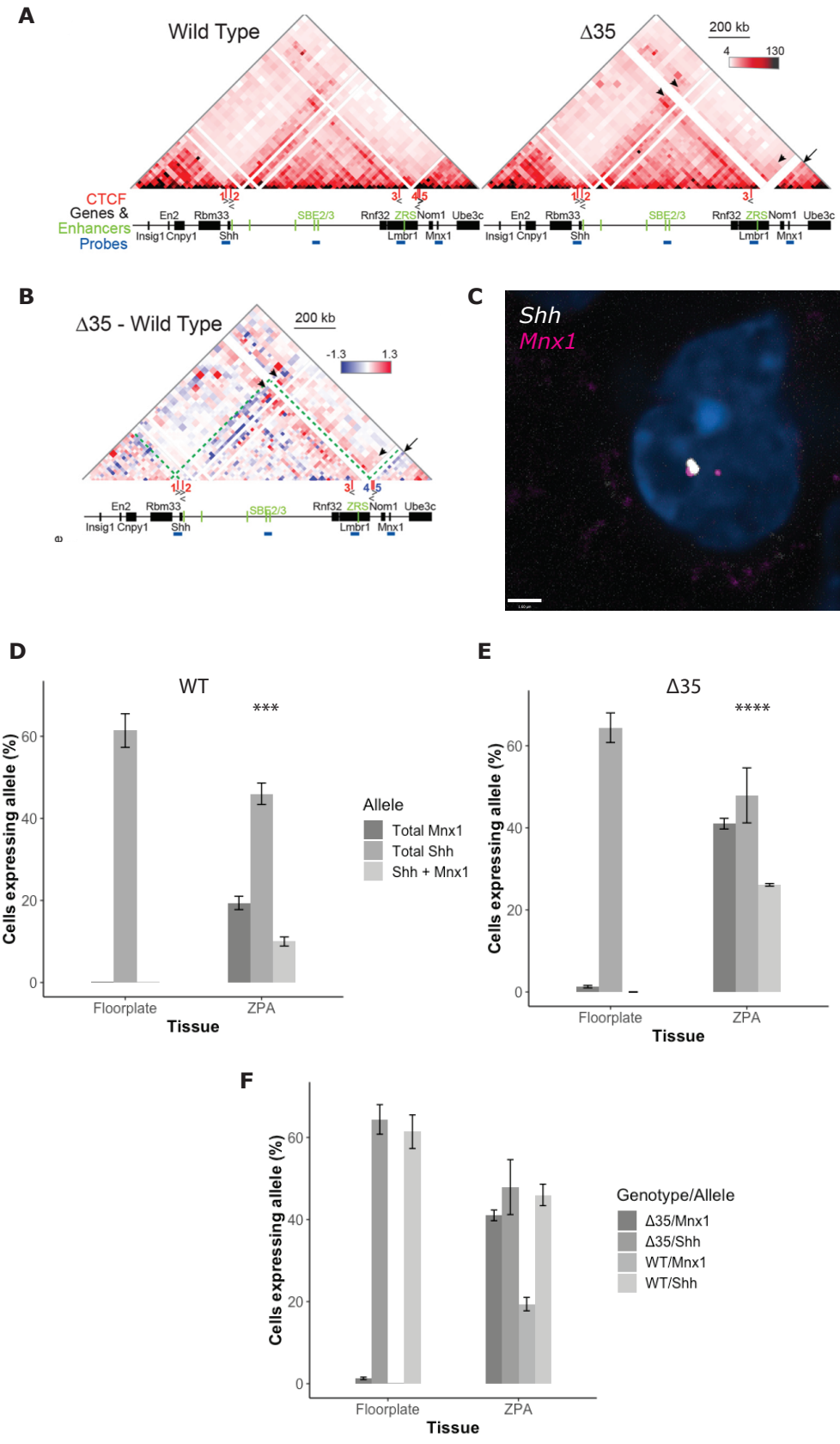

A
Wild Type
Δ35
200 kb
4
130
CTCF
Genes &
Enhancers
Probes
1
2
3
4
5
En2
Rbm33
SBE2/3
Rnf32
ZRS
Nom1
Ube3c
Insig1
Cnpy1
Shh
Lmbr1
Mnx1
B
Δ35 - Wild Type
200 kb
-1.3
1.3
C
Shh
Mnx1
ZPA
D
WT
***
Cells expressing allele (%)
Allele
Total Mnx1
Total Shh
Shh + Mnx1
Floorplate
ZPA
Tissue
E
Δ35
****
F
Genotype/Allele
Δ35/Mnx1
Δ35/Shh
WT/Mnx1
WT/Shh

Figure 5.3. RNA-FISH of ∆35 and WT littermate embryonic tissue shows ZRS can co-activate Shh and Mnx1

A 5C heatmaps of WT and ∆35 ESCs (Williamson *et al.*, 2019b). **B** Heatmap comparing Δ35 5C enrichment (red) with wild type (blue). Green dashed lines indicate the TAD boundaries, black arrows highlight loss of interactions in Δ35 cells and black arrowheads indicate enriched interactions in Δ35 cells. (A&B are adapted from Williamson et al., 2019). **C** Representative image of RNA-FISH conducted in ZPA cells. DAPI is shown in blue, probes for *Shh* in white, a *Mnx1* probe in magenta. Scale bar = 1.60 μM. **D** Total percentage of cells expressing *Shh* and *Mnx1,* and percentage of cells expressing *Shh* and *Mnx1* in the same cell on the same chromosome in WT tissue and **E** ∆35 tissue. *<0.05, **<0.01, ***<0.001, ****<0.0001 (LD, Chi-squared test). **F** Comparison of total *Shh* and *Mnx1* expression levels in WT and ∆35 tissue. Error bars calculated by SEM, obtained from two tissue sections from two embryos. For each tissue there is n = ~200 nuclei (sum of all nuclei from all sections of both embryos analysed).

5.2.5. DNA-FISH of ∆35 mouse tissues

I have shown two sets of contradictory results above. *Shh* enhancers are unable to co-activate *Shh* and *lacZ,* however, the ZRS is capable of co-activating *Shh* and *Mnx1*. To decipher reasons for these contradictory data I performed DNA-FISH on tissues from homozygous ∆35 mice and WT litter mates to determine the position of the alleles in relation to each other. Unfortunately, due to time constraints I was unable to perform DNA-FISH on SBLac526 tissue as initially planned.

However, I was able to address some key questions about interactions between ZRS, *Shh* and *Mnx1* in ∆35 and WT tissue. Firstly, I explored how distances between the alleles are altered by the 35 kb deletion perhaps giving insights into how proximity between the target genes and enhancer affects expression levels. Secondly, I examined how the distances between alleles, coupled with expression of these alleles can be suggestive of certain models of distal enhancer-promoter interaction.

For ∆35 tissue three sets of DNA fosmids were used to probe for *Shh*, *Mnx1* and ZRS. There are a few different possible outcomes I would expect to see. The transcriptional hub hypothesis suggests that the physical distance that separate enhancers from their target promoters is not important (Lim and Levine, 2021), whereas intimate contact would be predicted of a looping model (Müller, Sogo and Schaffner, 1989). Previous data from the Hill lab has shown that in the ∆35 ESCs, *Mnx1* and ZRS come significantly closer together than in the WT (Williamson *et al.*, 2019b). To prove this was not solely caused by the deletion of 35 kb between the probes, the 35 kb stretch was inverted. This caused CTCF site 4 and 5 to be in the opposite orientation, therefore shifting the TAD boundary without altering the linear distance between *Mnx1* and ZRS. It was still found that the ZRS and *Mnx1* interprobe distances were significantly shorter than in WT (Williamson *et al.*,

2019b). Therefore, I would expect the same outcome here. Though it should be noted that the Williamson et al. (2019) conducted their DNA-FISH on ESCs, which do not express *Shh*.

DNA-FISH was conducted on WT and ∆35 E11.5 embryos using the three fosmids described above (Figure 5.4). I looked at both ZPA tissue and anterior limb bud tissue as non-expressing control. First of all, looking at the interprobe distances of *Shh*/ZRS in the ZPA there is a significant decrease in ∆35 tissue compared to WT (Figure 5.4B) (Mann-Whitney *U*-test, *p-value = 2.086e-13*). The deletion of the 35 kb fragment allows *Shh* and ZRS to come closer together in 3D space. However, I have shown that *Shh* bursting frequency is not changed (Figure 5.3E), suggesting that decreased distance does not affect *Shh* mRNA levels. In the anterior limb bud there is no change in the interprobe distance between *Shh* and ZRS when comparing WT and ∆35 tissue (Figure 5.4C) (Mann-Whitney *U*-test, *p-value = 0.69*). Therefore, shifting of the TAD boundary allows ZRS to come into closer contact with *Shh* in ∆35 tissue only when it is an active enhancer, yet this has no effect on bursting frequency. Previous work has also shown there is no effect on relative *Shh* mRNA levels (Williamson *et al.*, 2019).

When looking at interprobe distances for *Mnx1* and ZRS there is no change between ∆35 and WT in the ZPA (Figure 5.4B). This is interesting as it contradicts previous data from ESCs where there was decreased distance between the two alleles (Williamson *et al.*, 2019). In section 5.2.4 I have shown that there is a significant increase in *Mnx1* bursting frequency, if the looping hypothesis is valid this should have been reflected in a decrease in distance between the alleles. Therefore, these data support a transcriptional hub hypothesis. Furthermore, the 35 kb deletion does not affect the positioning of ZRS and *Mnx1* in 3D space, meaning the distance observed

between ZRS and *Mnx1* by DNA-FISH has not decreased in ∆35 ZPA tissue compared to WT.

Interestingly the interprobe distance between *Mnx1* and ZRS increases in ∆35 tissue in the anterior limb bud (Figure 5.4C) (Mann-Whitney *U*-test, *p-value = 0.02*). This contradicts the findings of Williamson et al (2019) in ESCs (which are non-expressing cells). They found that in ∆35 ESCs *Mnx1* was in increased proximity to ZRS. These differences are possibly due the difference in model systems (tissue vs ESCs) being used.

Comparing distance between *Shh*/ZRS and Mnx1/ZRS, it is clear *Shh* is much closer to the enhancer in both the WT and ∆35 in the ZPA (Figure 5.4B). I have shown that *Shh* and *Mnx1* can be co-activated, in fact, ~50% of the instances I see *Mnx1* bursting, *Shh* is being expressed from the same chromosome at the same time (Figure 5.3D, E). Therefore, these relatively large distances between the ZRS and *Mnx1* is suggestive of a transcriptional hub model. Furthermore, there is no correlation between the *Shh*/ZRS and *Mnx1*/ZRS distances in either the WT or ∆35 (Figure 5.4D) (Pearson's correlation, *WT p-value = 0.34, ∆35 p-value = 0.12*). If the enhancer was switching between the two genes, as expected in a looping model, I would expect to see a correlation in the distances; when *Mnx1* is close to the ZRS then *Shh* is more distal and vice versa. However, this is not the case. The lack of correlation suggests that the proximity of either gene does not impact the ability of the ZRS to interact with either promoter. Again, this is suggestive of a transcriptional hub model.

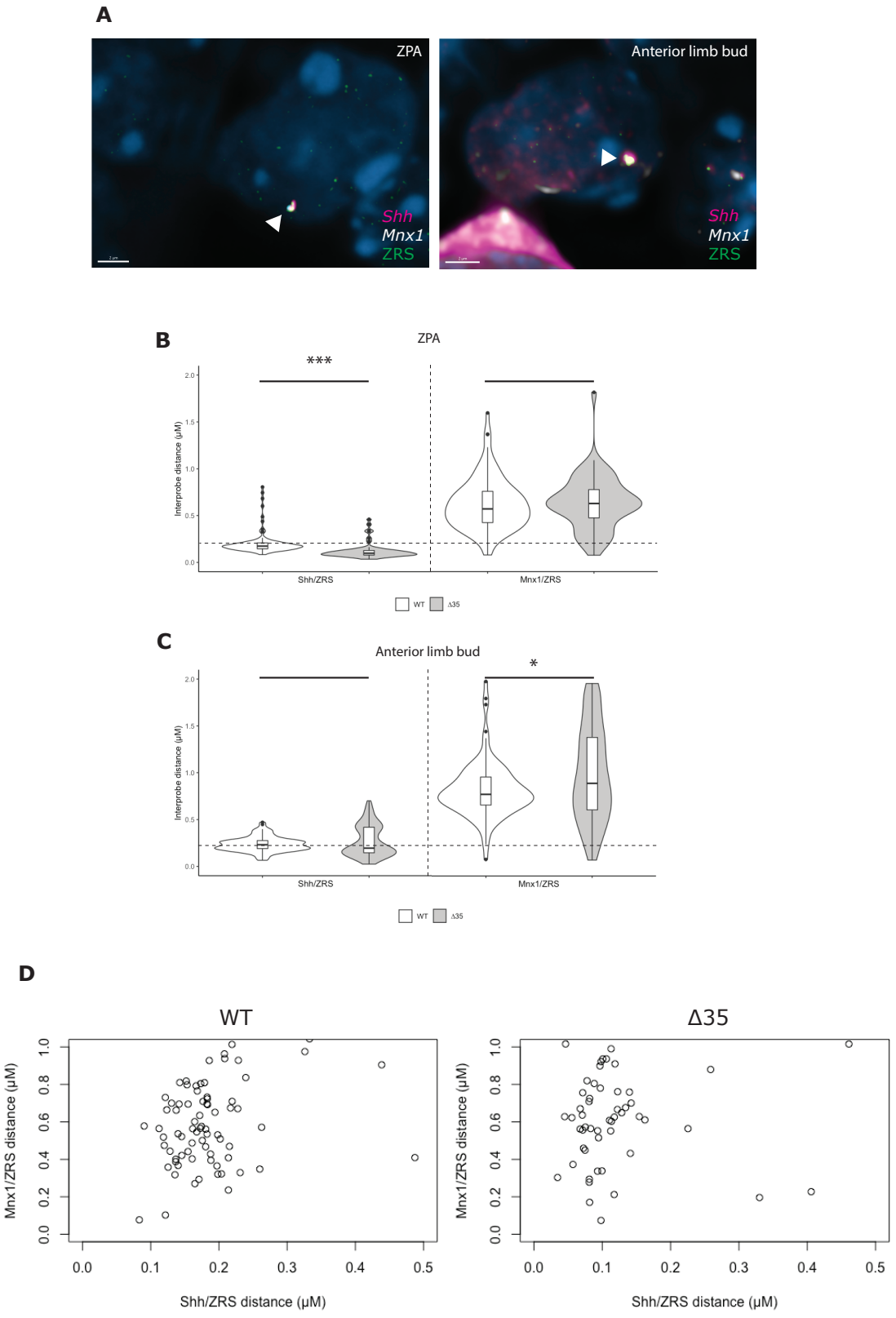


Figure 5.4. DNA-FISH of WT and Δ35 tissues suggests a transcriptional hub for enhancer-promoter interaction

A Representative images of DNA-FISH signal in ZPA and anterior limb bud cells. DAPI is shown in blue. Three fosmids were used to probe for *Shh* (magenta), *Mnx1* (white) and ZRS (green). White arrows indicate co-localisation of the three alleles. Smaller green spots in both images are background noise, as is the magenta haze seen in the anterior limb bud. Scale bars = 2 μM. **B** Violin plots show the distribution of interprobe distances between *Shh*/ZRS and *Mnx1*/ZRS in the ZPA and **C** anterior limb bud in both WT and ∆35 tissue. Horizontal dashed line indicates 200 nM. *<0.05, **<0.01, ***<0.001 (Mann–Whitney *U*-Tests). Number of distances between alleles recorded, n = 90. **D** Scatter plot mapping *Shh*/ZRS distance against *Mnx1*/ZRS distance in WT and ∆35 ZPA tissue. No correlation was found by Pearson correlation coefficient.

5.3. Conclusions and discussion

In this chapter I have explored distal enhancer-promoter interactions in the context of the *Shh* regulatory locus. One of the defining criteria of the transcriptional hub hypothesis, is that an enhancer can co-activate reporter genes simultaneously (Fukaya, Lim and Levine, 2016). However, this observation had not been investigated in an endogenous context and also in a context where the genes are not equidistant from the enhancer until now. Therefore, to decipher the mode of enhancer-promoter communication at the *Shh* locus I used nascent RNA-FISH in two different contexts to determine if co-activation was possible. Firstly, I used the SBLac526 mouse line, which contains a *lacZ* reporter gene within the *Shh* TAD gene desert, driven by the *Shh* enhancers in their cognate tissues. I found that three different *Shh* enhancers, located at varying distances from the *lacZ* and *Shh* promoter, were largely unable to co-activate *lacZ* and *Shh*. Furthermore, I showed that each of the enhancers had a set activity level, and the introduction of additional promoters caused the enhancer to split this activity between promoters. Lack of co-activation of the two genes and the set activity level of the enhancers suggested a looping model of enhancer-promoter interaction. However, this was not a fully endogenous system as a reporter gene was still used to investigate.

I next explored a fully endogenous system by looking at co-activation of *Shh* and *Mnx1* by ZRS, the *Shh* limb enhancer, in both WT and ∆35 embryos. Strikingly, these results contradicted the findings above. *Shh* and *Mnx1* are co-activated by the ZRS. When the TAD boundary separating the ZRS from *Mnx1* is disrupted the activity level of the ZRS increases. These findings indicate a transcriptional hub is the mechanism driving transcription.

Finally, I performed DNA-FISH on WT and ∆35 tissue to determine how TAD boundary deletion could impact distance between *Shh*, *Mnx1* and ZRS

alleles. I found that in ∆35 ZPA tissue that the ZRS and *Shh* became significantly closer together. Interestingly there was no change in the distance between *Mnx1* and ZRS in ∆35 ZPA tissue, despite increased *Mnx1* expression, indicating distance between the two alleles does not impact expression levels. This reinforces the hypothesis that the ZRS is communicating with the *Mnx1* promoter via a transcriptional hub. Furthermore, there is no correlation between *Mnx1/*ZRS distance and *Shh*/ZRS distance suggesting that proximity of either gene to the enhancer does not impact interaction of the enhancer with the other gene.

5.3.1. *Shh* enhancers cannot co-activate *Shh* and *lacZ* in their cognate tissues

Using nascent RNA-FISH I have shown the ZRS, SFPE1 and MACS1 are all largely unable to simultaneously co-activate *Shh* and *lacZ* under a *Hsp68* promoter. Being unable to simultaneously drive expression of multiple genes suggests that an enhancer can only communicate with one promoter at a time. This in turn would suggest a looping model for enhancer-promoter interaction. I have also shown that ZRS, SFPE1 and MACS1 appear to have a set activity level which gets divided between the promoters they interact with. These findings support a looping model for enhancer-promoter interaction.

SBLac526 mouse embryos were used to make the above observations. The main caveat to choosing the SBLac526 line is that there are still an additional two PGK promoters located in the *Shh* TAD, as described in 1.2.1 (Figure 1B). This means there is potential for the *Shh* enhancers to now interact with these promoters, in addition to the *Shh* promoter and the *Hsp86* promoter. The PGK is a strong promoter which does not require the influence of an enhancer to drive transcription. Also, the ZRS does not interact with other housekeeping promoters within the *Shh* TAD, such as that of *Lmbr1 (*Figure

4.1*C,D)*. In chapter 4, it was demonstrated that mutation of the ZRS does not alter the expression levels of *Lmbr1.* This could be verified further by looking at the expression pattern of *Lmbr1* and if it is independent of the *Shh* expression pattern. In areas where *Shh* is also expressed, which should be expected as *Lmbr1* is ubiquitously expressed, if there is no upregulation of the *Lmbr1* this indicates it is not being driven by the *Shh* enhancer in this region. This could be further assessed in another housekeeping gene within the *Shh* TAD, *Rnf32*.

Furthermore, under the transcriptional hub hypothesis(Lim and Levine, 2021) the number of promoters present should not affect the enhancers ability to drive transcription in any promoter as close contact is not required. Therefore, the presence of these promoters should not affect my ability to assess if *Shh* enhancers can co-activate *Shh* and *lacZ*. In the future this could be tested further. RNA-FISH probes could be designed to *neo* and *puro* genes that are driven by the PGK promoters, establishing if there is linked expression of the genes with *Shh*. Furthermore, there is a mouse line that retains the *lacZ* gene in the same position as the SBLac526 line but the additional elements of the construct, such as the PGK promoters, have been removed. Therefore, in this line nascent RNA-FISH for *Shh* and *lacZ* could be performed to determine if there is any change in results compared to those described here.

The set activity levels that the three enhancers appear to have is an interesting observation. It reinforces the idea that the enhancers can only communicate with one promoter at a time. Chen and colleagues (2018) have made similar observations and found that in *Drosophila melanogaster* insertion of a reporter gene into the *eve* locus resulted in a 5-20% decrease in *eve* transcription. Competition for the enhancer activity between the additional promoters was suspected to be the cause. Together these findings

suggest that a looping model may be at play. This model would explain why the introduction of a second promoter would result in a down-regulation of the target gene expression. The close contact between enhancer and promoter required by the looping model for gene activation would only allow communication with one promoter at a time.

5.3.2. ZRS can co-activate *Shh* and *Mnx1* in the developing limb bud

Contrary to the results with *lacZ* and *Shh*, the ZRS is able to co-activate *Shh* and *Mnx1*. Furthermore, when the TAD boundary insulating the ZRS from *Mnx1* is removed, the activity of the level of the ZRS increases. These findings suggest that transcription is being driven by a transcriptional hub.

Interestingly, in ∆35 ZPA tissue I found there was a significant increase in *Mnx1* expression compared to WT. This in contrast to previous findings where by qPCR a modest but not significant change in *Mnx1* mRNA levels could not be detected (Williamson *et al.*, 2019b). This highlights that nascent RNA-FISH can be utilised as a more sensitive technique for detecting changes in expression levels. By demonstrating this increase in *Mnx1* expression occurs upon a shift in TAD boundary and that *Shh* expression is unaffected, I have reinforced the data of Williamson et al. (2019) that TAD boundaries are necessary for the insulation against aberrant interactions, but are not necessary to ensure interaction between target genes and their enhancers. In fact, multiple other studies at other loci suggest that removal of CTCF sites, therefore disrupting TAD boundaries, do not impact gene expression (Rodríguez-Carballo *et al.*, 2017; Despang *et al.*, 2019).

5.3.3. Comparison of results

Obviously, the *Shh* and *lacZ* co-activation and the *Shh* and *Mnx1* co-activation results contradict each other. The ZRS is largely unable to co-activate *Shh* and *lacZ*, as are multiple other *Shh* enhancers. This supports the looping hypothesis. Conversely, ZRS is capable of co-activating *Shh* and *Mnx1*, supporting a transcriptional hub hypothesis. This perhaps suggests that the two mechanisms are not mutually exclusive. One potential reason for this could be promoter types of the different genes. *Hsp68*, the promoter attached to *lacZ* in this instance, is a basal promoter and therefore requires activation from an enhancer for transcription to proceed. Previous research has shown that the mouse ZRS is crucial for *Shh* expression in the developing limb bud. Deletion of the ZRS sequence in mice results in complete loss of *Shh* expression in the limb buds and loss of distal skeletal elements of the adult limb (Sagai *et al.*, 2005). Furthermore, a different group took the mouse ZRS and replaced it with a "serpentised" version. This serpentised ZRS was orthologous to the sequence in cobras. There was loss of *Shh* expression in the developing limb buds and severe truncation of the limbs in adult mice (Kvon *et al.*, 2016). These findings demonstrate that the *Shh* promoter is not capable of driving expression without a functional limb enhancer in limb. Potentially the *Mnx1* promoter has a different base activity level that allows it to be driven by less stable interactions. A future experiment that could be conducted to test the activity level of the *Mnx1* promoter would be to create transgenics. This would involve microinjection of two different construct into different groups of zygotes. The first set of zygotes would be injected with a construct containing the ZRS sequence and *lacZ* attached to the *Mnx1* promoter. In this construct the ZRS would be driving the expression of *lacZ* under the *Mnx1* promoter. In the transgenic embryos, it would be expected that *lacZ* expression pattern would follow the *Shh* expression pattern in the limb. The second group of constructs would have the *Mnx1* promoter and *lacZ* only. Therefore, any *lacZ* expression would be driven solely by the *Mnx1* promoter. If the promoter only construct can drive *lacZ* expression this would indicate that enhancer activity is not

necessary to drive the *Mnx1* promoter, therefore suggesting that stable interactions between the *Mnx1* promoter and an enhancer is not necessary to drive transcription.

TAD structure must also be considered. *Shh, lacZ* and ZRS interact within a single TAD whereas it is necessary for interaction across a TAD boundary to take place for ZRS to drive *Mnx1* transcription. Potentially a conformational change, such as a change in looping, is necessary for the ZRS to move into proximity of and communicate with the *lacZ* gene. This may not be necessary for ZRS to interact with *Mnx1*. ZRS interacts with *Mnx1* and *Shh* endogenously, therefore chromatin structure may already place the genes within the vicinity of the ZRS so that a hub can form and drive expression of both genes simultaneously. Indeed, through DNA-FISH I have shown that there is no correlation between *Shh*/ZRS distance and *Mnx1*/ZRS distance, implying that a conformational change is not necessary for ZRS to interact with either gene. To determine if the opposite is observed between *Shh*/ZRS and *lacZ*/ZRS distance DNA-FISH in SBLac526 ZPA tissue must be conducted. I would predict that if a conformational change occurred then there would be a correlation between distances. For example, when *Shh* is in proximity to the ZRS then *lacZ* is more distal and vice versa. A further way to investigate in the future would to be conduct DNA/RNA-FISH. Determining the position of the alleles while they are actively transcribed would give further insight into the mechanism by enhancer promoter interaction is occurring. In a looping model I would predict that when an active transcription site of one allele is present, there would be proximity between that allele and the enhancer, with the other allele being distant. In a looping model distance of either promoter from the enhancer would not predict the presence of an active transcription site.

Overall, further investigation is needed to decipher what mode of enhancer-promoter interaction is used to drive *Shh* transcription. I propose that the evidence I have shown here points towards a transcriptional hub model as in an endogenous context of the ZRS with *Shh* and *Mnx1* this is clearly the mechanism being used. Furthermore, as noted in chapter 3 ZRS is capable of being transcribed through which suggests a transient transcriptional complex supportive of a transcriptional hub model. Evidence of the looping model arises when a heterologous promoter-reporter gene is inserted within the *Shh* regulatory locus. However, I would argue that evidence of the looping model occurring would suggest that mechanisms of enhancer-promoter interaction are not a 'one size fits all' solution. It is likely different alleles interact through different methods. Perhaps the underlying TAD structure between enhancers and promoters help define which mechanism occurs to allow enhancer-promoter interactions.

5.3.4. Future Directions

Given the contradictory results of ZRS co-activation of *Shh* and *Mnx1,* and lack of co-activation of *Shh and lacZ* this gives rise to interesting experiments that could be conducted in the future.

A first experiment could be to test the effect promoter type has on the ZRS splitting its time between promoters. Here it has been shown that the ZRS cannot drive expression of *Shh* and *lacZ* but can drive expression of *Shh* and *Mnx1* simultaneously. A mouse line could be generated that places a *lacZ* reporter gene under the control of the *Mnx1* promoter. If the difference in promoter is the reason that the ZRS can simultaneously drive *Shh* and *Mnx1* in the WT, in this new mouse line I would predict that the ZRS would now be able to simultaneously drive *Shh* and *lacZ* expression. If simultaneous co-activation was observed in this situation this would have important

implications for promoter engagement into a transcriptional hub and imply the sequence of the promoter is important for enhancer-promoter selectivity.

Another interesting observation shown here is that *lacZ* insertion into the *Shh* TAD results in a drop in *Shh* burst frequency. In drosophila, the insertion of a *lacZ* reporter gene has also been shown to reduce gene expression levels (Chen *et al.*, 2018). This observation could be tested further by inserting further *lacZ* reporter genes to the *Shh* locus. If the ZRS, or other *Shh* enhancers, are diving their time between the promoters, I would predict that sequential increase in the number of *lacZ* genes in the *Shh* regulatory locus, would result in sequential decrease in *Shh* expression. A further method to determine if the ZRS is diving its time between the *Shh* promoter and *Hsp68* promoter *lacZ* is under would be to conduct DNA-FISH for ZRS, *Shh* and *lacZ*. If the ZRS' time is being divided I would predict that by DNA-FISH the ZRS would be observed in proximity to one gene or the other. This would demonstrate that the enhancer is only communicating with one promoter at a time.

Chapter 6 Co-ordinated expression of genes across TAD boundaries

6.1. Introduction

Previous studies have found that the transcriptional bursting of genes can co-fluctuate (Raj *et al.*, 2006; Sun and Zhang, 2019). One study inserted two reporter genes into CHO-AA8-Tet-off cells creating two cells lines, one line contained the reporter genes in different *trans* loci, the other contained them at closely cis *cis* loci (Raj *et al.*, 2006). Using RNA-FISH the group identified that when integrated at the neighbouring loci bursts of the genes were co-ordinated, but at separate loci bursts were random (Raj *et al.*, 2006). More recently using single-cell RNA-seq another group looked at co-fluctuation of gene expression and unexpectedly found that not only were genes at the same loci co-fluctuating but that genes up to 60 million bp (6Mb) away were also (Sun and Zhang, 2019). This long-range effect was attributed to proximity in 3D space (Sun and Zhang, 2019). Co-fluctuation of neighbouring genes is likely due to similarities in chromatin accessibilities (Raj *et al.*, 2006; Sun and Zhang, 2019).

Therefore, it is clear that genes in the same loci have a tendency to be expressed in a co-ordinated manner, meaning they have a transcriptional burst at the same time. However, there is no current examples of how TAD structure and boundaries may affect this expression. Interestingly, in the previous chapter I have shown that *Mnx1* and *Shh* are expressed in a co-ordinated manner in the ZPA, due to a shared enhancer. The *Shh* TAD and those TADs neighbouring either side contain several ubiquitously expressed housekeeping genes, these genes have no known shared regulatory elements such as enhancers. I decided to investigate if there is any co-expression of *Shh* with several of these genes, both within and outside the *Shh* TAD. In doing so I aimed to uncover how TAD structure and boundaries perhaps regulate co-expression of these genes.

6.2. Cis co-expression of *Shh* and *Lmbr1*

When looking at multiplexed RNA-FISH in a single cell, I would expect to see genes being expressed in one of two ways if multiple active transcription sites were present: *cis* or *trans*. Genes bursting in a *cis* manner are those being expressed in the same cell on the same chromosome, whereas genes bursting in a *trans* manner are expressed in the same cell but on different chromosomes. Graphical examples are shown in Figure 6.1A.

Lmbr1 is ubiquitously expressed and does not have any shared regulatory machinery with *Shh*, though *Lmbr1* does contain the *Shh* limb enhancer, ZRS, within its fifth intron. Due to the fact *Shh* and *Lmbr1* are apparently not regulated by the same elements, I would expect that their expression would be random, the genes as likely to burst in *a* cis *or trans* manner. However, proximity in 3D space and similar chromatin accessibilities have been suggested to cause co-fluctuation in the expression of genes (Raj *et al.*, 2006; Sun and Zhang, 2019). Therefore, I decided to test if being located within the same TAD could allow for *Shh* and *Lmbr1* to be co-expressed.

It should be noted that probes for *Shh* and *5' Lmbr1* are located ~900 kb apart, and ZRS and *5' Lmbr1* probes are ~60 kb apart. When performing nascent RNA-FISH in WT embryos looking at *Shh*, 5' *Lmbr1* and ZRS expression an interesting phenomenon was observed. There appears to be a preference for alleles to be co-expressed on the same chromosome (Figure 6.1B,C). For example, at E10.5 in the ZPA I detected a cumulative total of 32.1% of cells co-expressing some combination of the three alleles on the same chromosome, as opposed to 7.1% of cells showing trans transcription of the alleles. Therefore, there is coordinated expression of these alleles. Co-expression from the same chromosome is seen at both developmental stages, both in the ZPA and the floorplate.

Shh and *Lmbr1* being co-expressed is surprising because of their apparent lack of shared regulatory machinery. I observed *Shh* and *Lmbr1* being co-expressed in 21.9% and 19.4% of cells observed in E10.5 and E11.5 ZPA tissue respectively. I observed *trans* expression of the two alleles in 1.78% of E10.5 ZPA cells. In E11.5 this number is 1.16%. I found that *Shh* and *Lmbr1* are co-expressed in a *cis* manner more than would be expected at random at both time points (Chi-squared test, *E10.5 p-value = 3.95x10* $^{-12}$*, E11.5 p-value < 2.22x10* $^{-16}$). There is a clear preference for *cis* transcription of *Shh* and *Lmbr1* in the ZPA.

In the floorplate I observed 19.3% and 22.1% of cells co-expressing the two alleles in E10.5 and E11.5, respectively. In both embryonic stages in the floorplate *Shh* and *Lmbr1* are in LD, therefore, bursting of the two alleles occured more frequently when arranged in a *cis* manner (Chi-squared test, *E10.5 p-value = 0.009, E11.5 p-value = 2.5x10* $^{-5}$). Therefore, there is a preference for more interaction between enhancers and promoters when arranged in-cis in cells of the floorplate as well.

Interestingly I observed instances of *5' Lmbr1* and ZRS not being co-expressed, in 2.8% of E10.5 ZPA cells and in 0.9% of E11.5 ZPA cells. This is unusual considering they are part of the same gene. However, they are linearly far apart, separated by ~60kb. Estimates using live-cells place intron half-life at 0.4 – 7.5 mins (Audibert, Weil and Dautry, 2002). Transcriptional elongation is predicted to occur between 2 Kb/min and 6 Kb/min (Fuchs *et al.*, 2014). Using the fastest estimate of elongation, it would take ~10 mins to transcribe from *5' Lmbr1* to ZRS. This time is greater than the longest intron half-life estimates of 7.5 mins. Therefore, it is likely that when 5' *Lmbr1* and ZRS sites are observed together they are part of different transcription bursts

and therefore, we would expect to see the sites together or separate at similar rates. Testing for LD shows that bursting of the two alleles are still highly correlated (Chi-squared test, *E10.5 p-value* $< 2.22 \times 10^{-16}$, *E11.5 p-value* $< 2.22 \times 10^{-16}$).

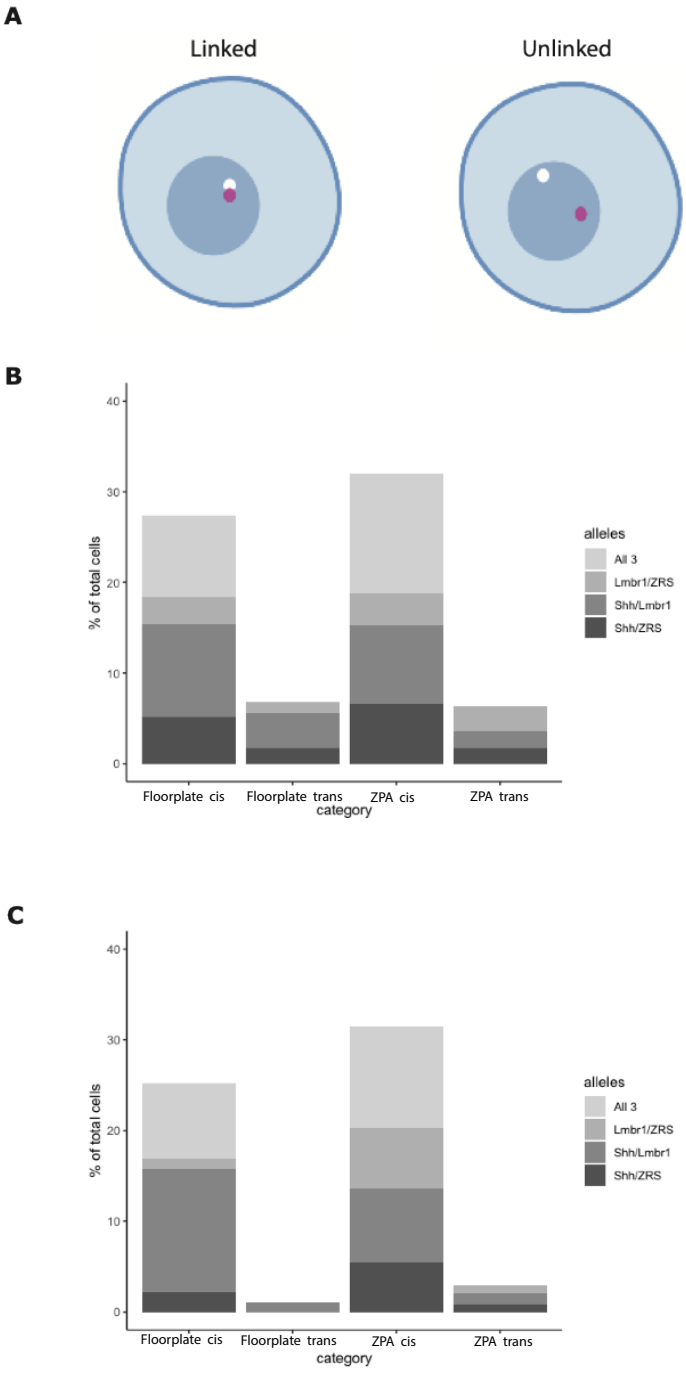


Figure 6.1 Cis expression of genes with no shared regulatory elements

A) Diagram showing how cis and trans RNA-FISH signals appear in a nucleus. Proximity is assumed to represent allele association based on DNA-FISH analysis. **B)** Proportion of E10.5 nuclei in ZPA and floorplate tissue expressing multiple alleles in a cis and trans manner. **C)** Proportion of E11.5 nuclei in ZPA and floorplate tissue expressing multiple alleles in a cis and trans manner.

6.3. *Co*-expression of *Shh*, *Nom1* and *Rbm33 when in-Cis.*

Given the co-expression of *Shh* and *Lmbr1* from the same chromosome in both ZPA and floorplate, I decided to explore if this coordination stretches outside of the TAD boundaries. Housekeeping genes are present in both of the TADs adjacent to the *Shh* TAD. I chose to investigate *Rbm33* which is located in the TAD adjacent to the *Shh* boundary and *Nom1* which is in the TAD adjacent to the *Lmbr1* boundary. I also investigated *Rnf32* expression with *Shh*, as it is another housekeeping gene within the *Shh* TAD. However, I could not identify any *Rnf32* signal in any tissues. Previous work was able to identify low levels of *Rnf32* expression in WT E11.5 embryos (Anderson *et al.*, 2014), however *in situ* hybridisations were unable to identify any specific expression.

I conducted nascent RNA-FISH on cryosections from E11.5 WT embryos (Figure 6.2A). I conducted the hybridisations so that either *Shh* and *Nom1*, or *Shh* and *Rbm33* were analysed. Again, both ZPA and floorplate tissue were analysed. I found that *Nom1* is expressed in ~20% of cells in both the floorplate and ZPA (Figure 6.2B). *Rbm33* is expressed in 29.2% of floorplate cells and in 37.4% of ZPA cells (Figure 6.2C). To determine if there is *cis* co-expression between either housekeeping gene and *Shh* I looked at the instances where both *Shh* and one of the housekeeping genes are expressed in the same cell.

First looking at *Nom1* and *Shh* expression there is a fairly even split between *cis* and *trans* co-expression (Figure 6.2D). In the floorplate 4.6% of cells show *cis* co-expression and 3.3% show *trans* co-expression. In the ZPA 2.9% of cells have cis expression of *Shh* and *Nom1*, whereas 3.9% show *trans* co-expression. These results suggest that there is not coordinated

expression of these two genes. Testing for LD finds that there is a negative correlation between bursting of *Shh* and *Nom1* in the ZPA and floorplate (Chi-squared test, *ZPA p-value = 0.0006, floorplate p-value = 9.45x10* $^{-5}$). These data suggest that bursting of these alleles on the same chromosome occurs less than expected.

Interestingly, when looking for *cis* co-expression of *Rbm33* and *Shh* results are initially less clear (Figure 6.2D). In the ZPA, 12.3% of cells show *cis* co-expression of the two genes, whereas only 3% show *trans* co-expression. When I looked at the floorplate I found 14.9% of cells had *cis* co-expression for *Rbm33* and *Shh* compared to only 3.6% *trans*. These data could be interpreted as showing coordination of expression between *Shh* and *Rbm33*. However, upon testing for LD no significant bursting co-ordination is found in either tissue (Chi-squared test, *ZPA p-value = 0.35, floorplate p-value = 0.1*). These findings suggest coordinated expression of genes is contained within a TAD when there is no shared regulatory machinery at play. *Shh* and *Mnx1* do show *cis* co-expression in the limb despite being in different TADS but this is likely due to them sharing an enhancer, ZRS.

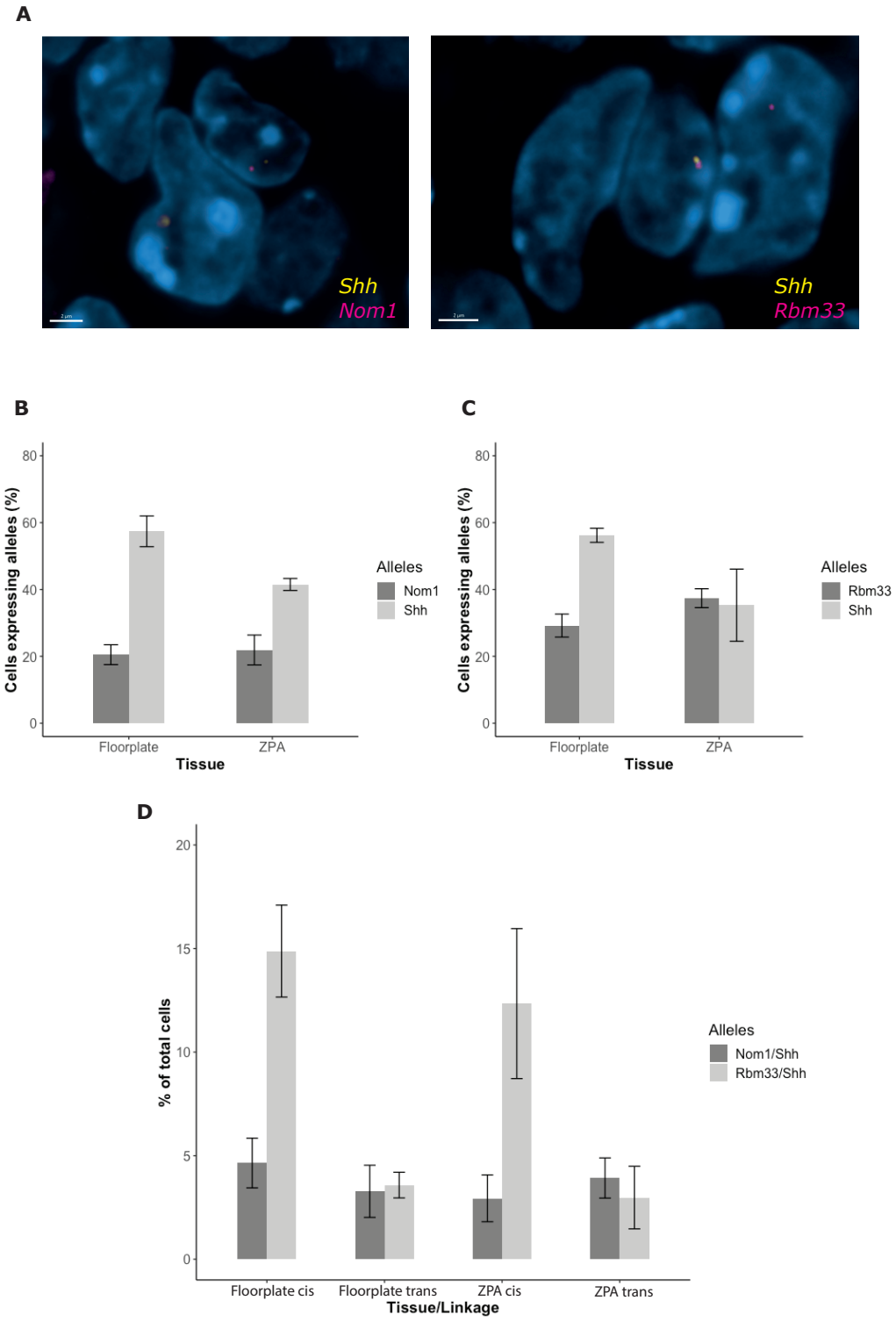


Figure 6.2 Co-expression of genes outside of the Shh TAD

A Representative images of nascent RNA-FISH in the ZPA for *Shh* (yellow), *Nom1* (left, magenta) and *Rbm33* (right, magenta). DAPI is shown in blue. Scale bars = 2 μM. **B** Bar chart showing the expression of *Shh* and *Nom1* in E11.5 WT ZPA and floorplate tissue. **C** Bar chart showing the expression of *Shh* and *Rbm33* in E11.5 WT ZPA and floorplate tissue. **D** Bar charts comparing the *cis* and *trans* co-expression of *Shh* and *Nom1*, and *Shh* and *Rbm33.* Error bars calculated by SEM, obtained from two tissue sections from two embryos. For each tissue there is n = ~200 nuclei (sum of all nuclei from all sections of both embryos analysed).

6.4. Conclusions and Discussion

In this chapter I investigated *cis* co-expression of genes within and across TAD boundaries. First, I demonstrated that despite being located ~900 kb apart *Shh* and *Lmbr1* are co-expressed within the *Shh* TAD. This was found in both the floorplate and the ZPA. To determine if co-expression extended past the TAD boundaries I looked for co-expression of both *Nom1* and *Rbm33*, each located in one of the neighbouring TADS. Interestingly I found that *Shh* and *Nom1* were negatively correlated, meaning the two alleles burst together less frequently than would be random. *Rbm33* and *Shh* bursting have no significant correlation and therefore, there is not co-ordinated expression of the two alleles. These results would initially suggest that TAD boundaries prevent co-ordinated expression extending beyond a TAD. However, I have shown in chapter 5 that *Mnx1* and *Shh* do show *cis* expression, despite a TAD boundary separating them. This is likely due to a shared enhancer.

6.4.1. Co-ordinated expression of genes within a TAD

I have reported here that transcriptional bursts of *Shh* and *Lmbr1* appear in a co-expressed manner in multiple tissues and embryonic stages. There are several possible explanations for this. Firstly, this phenomenon could be caused by ZRS being contained within *Lmbr1*. Certain models of enhancer-promoter interaction, e.g. transcriptional hubs, could explain this. The ZRS and *Shh* promoter will be interacting in a transcriptionally favourable environment in ZPA tissue. Their proximity in 3D space could therefore also be bringing the *Lmbr1* promoter into the same environment and be causing the *cis* expression I have observed. However, due to this occurring in multiple tissues, and not just in the ZPA where ZRS is an active enhancer it is unlikely that this is an explanation for the co-ordinated transcriptional bursting of these genes. It is likely the co-expression from the same chromosome is cause by favourable chromatin accessibility, which has been noted in previous studies (Raj *et al.*, 2006; Sun and Zhang, 2019). Unfortunately,

because I was unable to detect *Rnf32* signal in *Shh* expressing tissues I cannot determine if this phenomenon is apparent in other genes in the TAD.

6.4.2. Housekeeping genes from different TADs are not co-ordinated with *Shh*

Perhaps unsurprisingly, there is no co-expression of *Shh* with either *Nom1* or *Rbm33*. *Rbm33* is located in the TAD directly adjacent to *Shh*, and its expression shows no correlation with *Shh*, meaning the two alleles are expressed at random relative to each other. *Nom1* also shows no co-expression with *Shh* despite its neighbouring gene, *Mnx1,* doing so. *Mnx1* co-ordination is presumably due to a shared enhancer, ZRS. However, I believed this would have resulted in similar chromatin accessibility for *Nom1* and therefore caused co-expression, which is clearly not the case. To investigate this in the future DNA-FISH could be conducted to look at the position of *Shh, Mnx1* and *Nom1* in relation to each other. I would predict that *Nom1* would appear to be looped out of the region where *Mnx1* and *Shh* are located to explain why it is not co-ordinated.

An interesting experiment to perform in the future would be to look for co-ordinated expression of *Shh* and *Nom1* in the ∆35 mutants. The 35 kb deletion in these mice disrupt the TAD boundary separating *Shh* from *Nom1* (Williamson *et al.*, 2019b). If the deletion resulted in *Nom1* and *Shh* expression now being co-ordinated this would suggest that sharing a TAD structure results in co-expression of genes. Therefore, supporting previous findings suggesting that similar chromatin accessibilities results in co-fluctuation of gene expressions (Raj *et al.*, 2006; Sun and Zhang, 2019). Another, experiments that could further dissect the role of TAD boundaries in co-ordinated transcription would be to put *Nom1* in the *Shh* TAD. This could be done by CRISPR genome editing, directing the gene be inserted into the *Shh* TAD gene dessert through a homologous repair template. If *Shh* and

Nom1 expression became co-ordinated this would establish the TAD boundaries as an underlying feature of co-ordinated expression of genes at the same loci. However, this finding would not explain why *Mnx1*, which is located in the same TAD as *Nom1*, can have coordinated expression with *Shh* across a TAD boundary.

Another possibility could be that the genes are located within different transcriptional compartments. A way to identify if this is valid would be to conduct Hi-C in ZPA tissue. This would identify different compartments present in the tissue and indicate if the *Shh, Nom1* and *Rbm33* are located in different compartments.

6.4.3. A potential model for co-ordinated expression across a TAD

Using *Shh* and *Lmbr1* I have shown that within a TAD expression of genes appear co-ordinated suggesting they are in a shared environment favourable for transcription. Interestingly, while housekeeping genes either side of the *Shh* TAD boundary, *Nom1* and *Rbm33*, are expressed they are not co-ordinated with *Shh*. This indicates that while these genes are in an environment where active transcription can take place this environment must be distinct from *Shh*. Therefore, a TAD may indicate both a limited domain of preferred interactions and as shown here, a region of coordinated transcription. Further analysis of transcriptional bursting in other TAD domains may support this hypothesis. I propose the model shown in Figure 6.3. It suggests that the *Shh* TAD and the *Nom1* and *Rbm33* TADs are within the same transcriptionally active region perhaps residing at the periphery of the chromosomal territory. The formation of separate transcriptional hubs drives expression of genes in each TAD, accounting for why all genes are expressed but not in a co-ordinated manner. However, crosstalk between the two TADs can occur; for example, with ZRS driving transcription of *Mnx1*

despite being separated by a TAD boundary, and this crosstalk increases when the TAD boundary is weakened, such as in the Δ35 embryos where *Mnx1* expression increases. Hence boundaries do not appear to be absolute and Mnx1 expression in the limb, for an as yet undefined reason, is able to adopt activity from its neighbouring TAD. Why *Mnx1* is able to do this would need to be explored in future works.

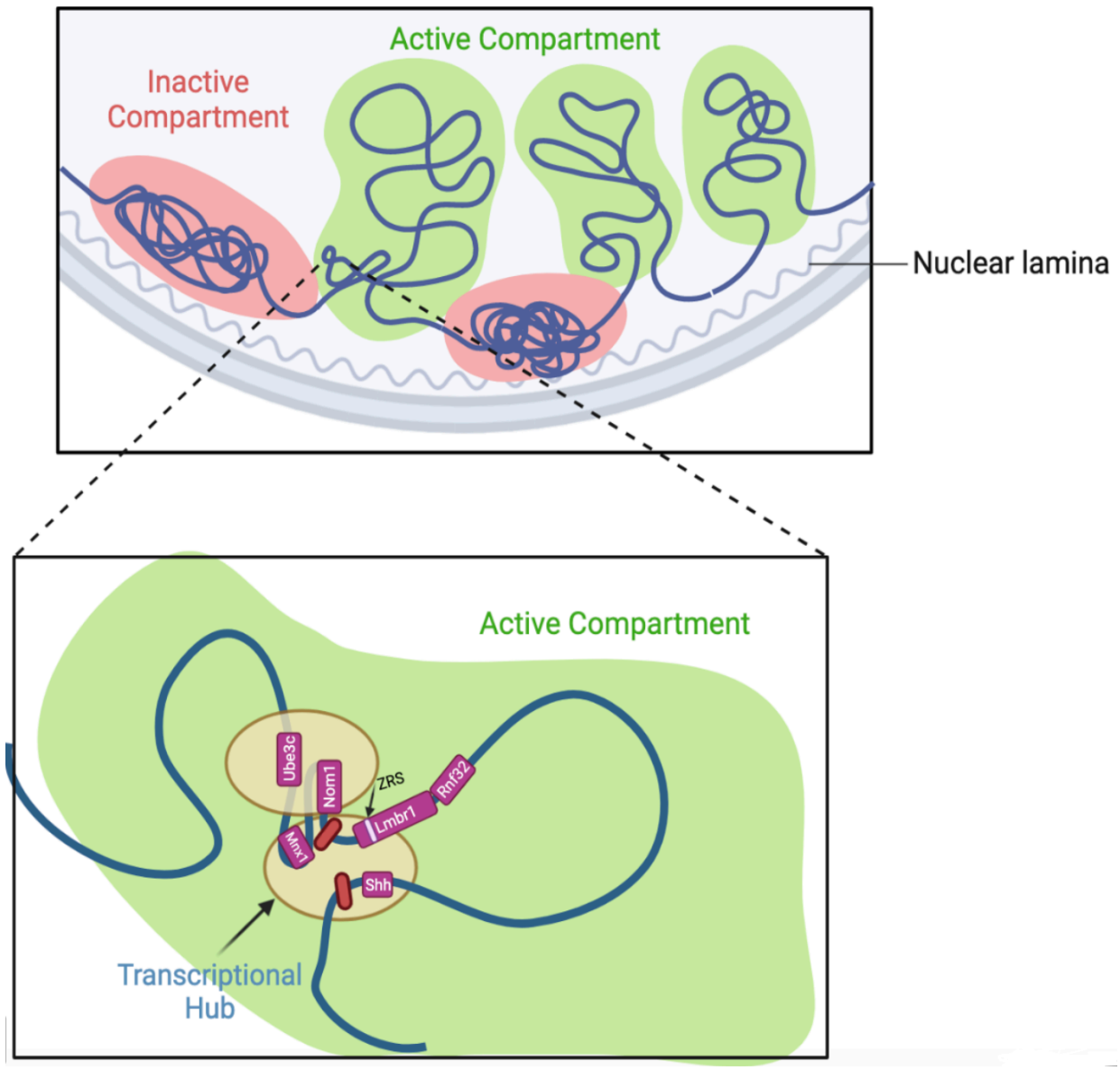


Figure 6.3. Model for co-ordinated expression across a TAD

Shh, *Nom1* and *Rbm33* are located within the same transcriptionally active compartment but are found within distinct transcriptional hubs which co-ordinated transcription within them. Genes are shown in purple, red boxes indicated TAD boundaries, ZRS is shown in white, yellow circles indicate transcriptional hubs.

Chapter 7 Concluding remarks

During my PhD I have explored several questions surrounding enhancers and their ability to drive specific spaciotemporal expression. More specifically I have used the *Shh* regulatory locus as a model to investigate long-range enhancer promoter interactions. Firstly, I utilised nascent RNA-FISH to determine that *Shh* transcription occurs in bursts. Interestingly my data suggests that both *Shh* and *Lmbr1*, a housekeeping gene that contains the *Shh* limb enhancer, are largely expressed in a monoallelic manner, meaning that only one allele of each gene was observed bursting at a time. Variations in *Shh* expression levels have been shown to have a variety of adverse phenotypes such as syndactyly and polydactyly (Anderson *et al.*, 2012; Lettice *et al.*, 2012, 2017). However, transcription is a stochastic process (Chubb *et al.*, 2006; Raj *et al.*, 2006; Chubb and Liverpool, 2010; Larsson *et al.*, 2019). Therefore, I have proposed that monoallelic expression may be a method of gene regulation to overcome the noisy nature of transcription and maintain expression levels within a range for normal development. I have also shown that the ZRS is capable of being transcribed through as part of the *Lmbr1* transcript while it is acting as an active enhancer. This poses several possibilities for enhancer-promoter interactions. If a stable and intimate complex forms between an enhancer and promoter it must be transient to allow transcription of the ZRS as part of the *Lmbr1* transcript. Alternatively, a stable complex is not required to activate transcription, allowing the ZRS to both be transcribed and drive transcription of *Shh* simultaneously.

I then explored the role of different transcription factors in stimulating *Shh* transcription in the ZPA of the developing limb bud. This allowed me to determine how a variety of transcription factors interact with ZRS. Firstly, I looked at Hox protein binding on the ZRS. Removal of two Hox sites in ∆48 mice and three Hox sites in ∆110 mice resulted in a reduction in *Shh* burst

frequency compared to WT mice. I found that despite partial digit loss versus full digit loss in Δ48 and Δ110 mice, I was unable to find a difference in burst frequency between the two mutants. In the future, it could be determined if the area of the ZPA differs in these two mutants. If it does, this would raise the possibility that Hox proteins have two actions on the ZRS: 1) defining levels of *Shh* produced in the ZPA by affecting burst frequency and, 2) determining the number of cells capable of expressing *Shh*.

Next, I looked at Lim homeodomain binding sites within the ZRS. Lim homeodomain proteins act as pioneer factors. I focussed on the LhxB site identified by Silvia Peluso (manuscript in preparation) as mutation in this site resulted in a soft tissue syndactyly phenotype. The LhxB^{edit} homozygous mice displayed a reduced burst frequency of *Shh* in the ZPA. Interestingly the burst frequency observed was at a similar level to that of Δ48 and Δ110 homozygotes, despite the phenotype of LhxB^{edit} mice being far milder. Therefore, perhaps pioneer factors influence the levels of *Shh* produced through burst frequency but do not impact the number of cells capable of bursting.

To contrast, I next investigated the role of a repressor site in the ZRS, the WMS site. It is currently unknown what factor binds to this site but disruption by inserting an adjacent +AA results in additional digits both centrally and pre-axially. Interestingly I found that disruption of the WMS site did not affect burst frequency within the ZPA. Instead there is ectopic expression of *Shh*, suggesting that the action of the WMS site is region specific. The ectopic expression pattern stretches the entire distal edge of the limb bud adjacent to the AER. Possibly the factor that binds the WMS site is induced by factors secreted from the AER.

I also investigated mechanisms of long-range enhancer-promoter interaction at the *Shh* locus. The two prevailing hypotheses for this in the literature are looping and transcriptional hubs(Fukaya, Lim and Levine, 2016; Furlong and Levine, 2018; Lim and Levine, 2021). Previous research had indicated that if a transcriptional hub model were valid then an enhancer can simultaneously drive transcription of two reporter genes (Fukaya, Lim and Levine, 2016). However, the Fukaya study was limited in that it did not look at enhancers and promoters in an endogenous context or look at examples where reporter genes were not equidistant from the enhancer. The current study has explored both of these limitations. Firstly, I looked at the ability of three different *Shh* enhancers (ZRS, SFPE1 and MACS1), all located at varying distances from *Shh* and *lacZ,* to drive transcription of both *Shh* and a *lacZ* reporter gene. Interestingly I found that the enhancers were unable to drive transcription of *Shh* and *lacZ* simultaneously, which suggests a looping model for enhancer-promoter interaction. Furthermore, results suggested that the enhancers have a set activity level and the addition of the *lacZ* promoter into the *Shh* regulatory locus resulted in the enhancer dividing this activity between the promoters.

I proceeded to further investigate this in an endogenous system. The ZRS drives transcription of *Shh* and has also been shown to up-regulated *Mnx1* in the ZPA. Therefore, I determined if the ZRS was able to drive transcription of both genes at the same time. I found that the ZRS could drive transcription of both genes simultaneously, contradicting the findings with *lacZ*. Interestingly, when the TAD boundary separating the ZRS from *Mnx1* was disrupted, the activity of the ZRS increased with no effect on *Shh* transcription. These results all point towards a transcriptional hub model. DNA-FISH also revealed that there was no correlation between *Shh/*ZRS distance and *Mnx1/*ZRS distance in 3D space, indicating that proximity of one gene to the enhancer has no impact of the distance to the other, which suggests the enhancer is not switching between the two genes. Overall, I believe that my results

indicate that looping and transcriptional hub models are not mutually exclusive and there could be various mechanisms of long-range enhancer-promoter interaction across the genome.

Finally, I decided to explore co-ordinated expression of genes across a TAD. It has been suggested in the past that genes at the same loci have a tendency for their expression to co-fluctuate(Raj *et al.*, 2006; Sun and Zhang, 2019). Therefore, I looked for co-ordinated bursting of *Shh* with *Lmbr1* to determine co-expression within TAD boundaries. Strikingly, there was clear co-ordination of the two genes bursting from the same chromosome simultaneously, despite no known shared regulatory elements. I next wanted to decipher what effect TAD boundaries could have in coordinated expression and therefore looked at bursting of *Shh* with *Nom1* and *Rbm33*. I found that *Shh* was not co-expressed with either of these housekeeping genes from neighbouring TADs. All genes were shown to be being transcribed, suggesting occupation in a transcriptional active compartment of the nucleus. However, *Shh* and *Lmbr1* could be expressed in a co-ordinated manner within a TAD, while *Shh* was not co-expressed with *Nom1* and *Rbm33* from the adjacent TAD. This suggests a second underlying mechanism co-ordinating expression, potentially a transcriptional hub. Therefore, I proposed a model in which *Shh*, *Nom1* and *Rbm33* are all located within the same transcriptionally active compartment but are in distinct transcriptional hubs which co-ordinate expression within TADs.

References

Ahn, S. and Joyner, A. L. (2004) 'Dynamic Changes in the Response of Cells to Positive Hedgehog Signaling during Mouse Limb Patterning', *Cell*. Cell Press, 118(4), pp. 505–516. doi: 10.1016/J.CELL.2004.07.023.

Akhtar, W. *et al.* (2013) 'XChromatin position effects assayed by thousands of reporters integrated in parallel', *Cell*. Elsevier B.V., 154(4), pp. 914–927. doi: 10.1016/j.cell.2013.07.018.

Albiez, H. *et al.* (2006) 'Chromatin domains and the interchromatin compartment form structurally defined and functionally interacting nuclear networks', *Chromosome Research 2006 14:7*. Springer, 14(7), pp. 707–733. doi: 10.1007/S10577-006-1086-X.

Alexander, J. M. *et al.* (2019) 'Live-cell imaging reveals enhancer-dependent Sox2 transcription in the absence of enhancer proximity', *eLife*. Edited by R. H. Singer, K. Struhl, and Z. Liu. eLife Sciences Publications, Ltd, 8, p. e41769. doi: 10.7554/eLife.41769.

Alvarez-Dominguez, J. R. *et al.* (2017) 'The Super-Enhancer-Derived alncRNA-EC7/Bloodlinc Potentiates Red Blood Cell Development in trans', *Cell Reports*. Cell Press, 19(12), pp. 2503–2514. doi: 10.1016/J.CELREP.2017.05.082.

Anderson, C. *et al.* (2016) 'A strategy to discover new organizers identifies a putative heart organizer', *Nature Communications*. Nature Publishing Group, 7. doi: 10.1038/ncomms12656.

Anderson, E. *et al.* (2012) 'Human limb abnormalities caused by disruption of hedgehog signaling', *Trends in Genetics*. Elsevier Current Trends, 28(8), pp. 364–373. doi: 10.1016/J.TIG.2012.03.012.

Anderson, E. *et al.* (2014) 'Mapping the Shh long-range regulatory domain.', *Development (Cambridge, England)*. Oxford University Press for The Company of Biologists Limited, 141(20), pp. 3934–43. doi:

10.1242/dev.108480.

Andersson, R. *et al.* (2014) ‘An atlas of active enhancers across human cell types and tissues’, *Nature*, 507(7493), pp. 455–461. doi: 10.1038/nature12787.

Andrey, G. *et al.* (2017) ‘Characterization of hundreds of regulatory landscapes in developing limbs reveals two regimes of chromatin folding’, *Genome Research*, 27(2), pp. 223–233. doi: 10.1101/gr.213066.116.

Arnold, P. R., Wells, A. D. and Li, X. C. (2020) ‘Diversity and Emerging Roles of Enhancer RNA in Regulation of Gene Expression and Cell Fate’, *Frontiers in Cell and Developmental Biology*. Frontiers Media S.A., 7, p. 377. doi: 10.3389/FCELL.2019.00377/BIBTEX.

Audibert, A., Weil, D. and Dautry, F. (2002) ‘In Vivo Kinetics of mRNA Splicing and Transport in Mammalian Cells’, *Molecular and Cellular Biology*. American Society for Microbiology, 22(19), pp. 6706–6718. doi: 10.1128/MCB.22.19.6706-6718.2002/ASSET/756FC898-B709-4FCB-A1A0-B069EF85E1AC/ASSETS/GRAPHIC/MB1920335009.JPEG.

Bahar Halpern, K. *et al.* (2015) ‘Bursty Gene Expression in the Intact Mammalian Liver’, *Molecular Cell*. Elsevier, 58(1), pp. 147–156. doi: 10.1016/j.molcel.2015.01.027.

Banerji, J., Rusconi, S. and Schaffner, W. (1981) ‘Expression of a β-globin gene is enhanced by remote SV40 DNA sequences’, *Cell*. Cell Press, 27(2), pp. 299–308. doi: 10.1016/0092-8674(81)90413-X.

Barrington, C., Finn, R. and Hadjur, S. (2017) ‘Cohesin biology meets the loop extrusion model’, *Chromosome Research*. Springer Netherlands, 25(1), pp. 51–60. doi: 10.1007/s10577-017-9550-3.

Bartman, C. R. *et al.* (2016) ‘Enhancer Regulation of Transcriptional Bursting Parameters Revealed by Forced Chromatin Looping’, *Molecular Cell*. Cell Press, 62(2), pp. 237–247. doi: 10.1016/j.molcel.2016.03.007.

Bejerano, G. *et al.* (2004) ‘Ultraconserved Elements in the Human Genome’, *Science*, 304(5675), pp. 1321–1325. doi: 10.1126/science.1098119.

Bell, A. C. and Felsenfeld, G. (2000) ‘Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene’, *Nature*, 405(6785), pp. 482–485. doi: 10.1038/35013100.

Benabdallah, N. S. *et al.* (2016) ‘SBE6: a novel long-range enhancer involved in driving sonic hedgehog expression in neural progenitor cells’, *Open biology*. Open Biol, 6(11). doi: 10.1098/RSOB.160197.

Benabdallah, N. S. *et al.* (2019) ‘Decreased Enhancer-Promoter Proximity Accompanying Enhancer Activation’, *Molecular Cell*. Cell Press, 76(3), pp. 473-484.e7. doi: 10.1016/j.molcel.2019.07.038.

Benabdallah, N. S. and Bickmore, W. A. (2015) ‘Regulatory Domains and Their Mechanisms’, *Cold Spring Harbor Symposia on Quantitative Biology*. Cold Spring Harbor Laboratory Press, 80, pp. 45–51. doi: 10.1101/SQB.2015.80.027268.

Benoist, C. and Chambon, P. (1981) ‘In vivo sequence requirements of the SV40 early promoter region’, *Nature 1981 290:5804*. Nature Publishing Group, 290(5804), pp. 304–310. doi: 10.1038/290304a0.

Bensidoun, P. *et al.* (2016) ‘Imaging single mRNAs to study dynamics of mRNA export in the yeast Saccharomyces cerevisiae’, *Methods*. Academic Press, 98, pp. 104–114. doi: 10.1016/J.YMETH.2016.01.006.

Berger, S. L. (2007) ‘The complex language of chromatin regulation during transcription’, *Nature*, 447(7143), pp. 407–412. doi: 10.1038/nature05915.

Berlivet, S. *et al.* (2013) ‘Clustering of Tissue-Specific Sub-TADs Accompanies the Regulation of HoxA Genes in Developing Limbs’, *PLOS Genetics*. Public Library of Science, 9(12), pp. 1–17. doi: 10.1371/journal.pgen.1004018.

Bertrand, E. *et al.* (1998) ‘Localization of ASH1 mRNA Particles in Living Yeast’, *Molecular Cell*. Elsevier, 2(4), pp. 437–445. doi: 10.1016/S1097-2765(00)80143-4.

Bintu, B. *et al.* (2018) ‘Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells’, *Science*. American Association for the Advancement of Science, 362(6413). doi: 10.1126/SCIENCE.AAU1783/SUPPL_FILE/AAU1783_TABLE_S2.XLSX.

Bitgood, M. J., Shen, L. and McMahon, A. P. (1996) ‘Sertoli cell signaling by Desert hedgehog regulates the male germline’, *Current biology : CB*. Curr Biol, 6(3), pp. 298–304. doi: 10.1016/S0960-9822(02)00480-3.

Blow, M. J. *et al.* (2010) ‘ChIP-Seq identification of weakly conserved heart enhancers’, *Nature Genetics*, 42(9), pp. 806–810. doi: 10.1038/ng.650.

Bolzer, A. *et al.* (2005) ‘Three-Dimensional Maps of All Chromosomes in Human Male Fibroblast Nuclei and Prometaphase Rosettes’, *PLOS Biology*. Public Library of Science, 3(5), p. e157. doi: 10.1371/JOURNAL.PBIO.0030157.

Bonev, B. *et al.* (2017) ‘Multiscale 3D Genome Rewiring during Mouse Neural Development’, *Cell*. Elsevier, 171(3), pp. 557-572.e24. doi: 10.1016/j.cell.2017.09.043.

Bothma, J. P. *et al.* (2014) ‘Dynamic regulation of eve stripe 2 expression reveals transcriptional bursts in living Drosophila embryos’, *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 111(29), pp. 10598–10603. doi: 10.1073/PNAS.1410022111/SUPPL_FILE/PNAS.1410022111.SM05.MOV.

Boveri, T. (no date) ‘Die Blastomerenkerne von Ascaris Megalocephala und die Theorie der Chromosomenindividualität’, in.

Boyle, A. P. *et al.* (2008) ‘High-Resolution Mapping and Characterization of Open Chromatin across the Genome’, *Cell*.

Elsevier, 132(2), pp. 311–322. doi: 10.1016/j.cell.2007.12.014.

Boyle, S. *et al.* (2001) 'The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells', *Human molecular genetics*. Hum Mol Genet, 10(3), pp. 211–219. doi: 10.1093/HMG/10.3.211.

Boyle, S. *et al.* (2011) 'Fluorescence in situ hybridization with high-complexity repeat-free oligonucleotide probes generated by massively parallel synthesis', *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology*. Chromosome Res, 19(7), pp. 901–909. doi: 10.1007/S10577-011-9245-0.

Briscoe, J. and Small, S. (2015) 'Morphogen rules: Design principles of gradient-mediated embryo patterning', *Development (Cambridge)*. Company of Biologists Ltd, pp. 3996–4009. doi: 10.1242/dev.129452.

Buenrostro, J. D. *et al.* (2015) 'ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide', *Current Protocols in Molecular Biology*, 109(1), pp. 21.29.1-21.29.9. doi: https://doi.org/10.1002/0471142727.mb2129s109.

Burke, R. *et al.* (1999) 'Dispatched, a novel sterol-sensing domain protein dedicated to the release of cholesterol-modified hedgehog from signaling cells', *Cell*. Cell, 99(7), pp. 803–815. doi: 10.1016/S0092-8674(00)81677-3.

Cannavò, E. *et al.* (2016) 'Shadow Enhancers Are Pervasive Features of Developmental Regulatory Networks', *Current biology : CB*. Curr Biol, 26(1), pp. 38–51. doi: 10.1016/J.CUB.2015.11.034.

Carleton, J. B., Berrett, K. C. and Gertz, J. (2017) 'Multiplex Enhancer Interference Reveals Collaborative Control of Gene Regulation by Estrogen Receptor α-Bound Enhancers', *Cell Systems*, 5(4), pp. 333-344.e5. doi: https://doi.org/10.1016/j.cels.2017.08.011.

Chambeyron, S. and Bickmore, W. A. (2004) 'Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription', *Genes & development*. Genes Dev, 18(10), pp. 1119–1130. doi:

10.1101/GAD.292104.

Chen, H. *et al.* (2018) ‘Dynamic interplay between enhancer–promoter topology and gene activity’, *Nature Genetics 2018 50:9*. Nature Publishing Group, 50(9), pp. 1296–1303. doi: 10.1038/s41588-018-0175-z.

Chen, Y. and Struhl, G. (1998) ‘In vivo evidence that Patched and Smoothened constitute distinct binding and transducing components of a Hedgehog receptor complex’, *Development (Cambridge, England)*. Development, 125(24), pp. 4943–4948. doi: 10.1242/DEV.125.24.4943.

Chiang, C. *et al.* (1996) ‘Cyclopia and defective axial patterning in mice lacking Sonic hedgehog gene function’, *Nature*, 383, pp. 407–13. Available at: https://www.nature.com/articles/383407a0.pdf (Accessed: 30 May 2019).

Chong, S. *et al.* (2014) ‘Mechanism of Transcriptional Bursting in Bacteria’, *Cell*. Elsevier, 158(2), pp. 314–326. doi: 10.1016/j.cell.2014.05.038.

Chubb, J. R. *et al.* (2006) ‘Transcriptional Pulsing of a Developmental Gene’, *Current Biology*, 16(10), pp. 1018–1025. doi: 10.1016/j.cub.2006.03.092.

Chubb, J. R. and Liverpool, T. B. (2010) ‘Bursts and pulses: insights from single cell studies into transcriptional mechanisms’, *Current Opinion in Genetics & Development*. Elsevier Current Trends, 20(5), pp. 478–484. doi: 10.1016/J.GDE.2010.06.009.

Cisse, I. I. *et al.* (2013) ‘Real-time dynamics of RNA polymerase II clustering in live human cells’, *Science*. American Association for the Advancement of Science, 341(6146), pp. 664–667. doi: 10.1126/SCIENCE.1239053/SUPPL_FILE/CISSE.SM.PDF.

Crane-Smith, Z. *et al.* (2021) ‘A Highly Conserved Shh Enhancer Coordinates Hypothalamic and Craniofacial Development’, *Frontiers in Cell and Developmental Biology*. Frontiers, 9, p. 621. doi: 10.3389/fcell.2021.595744.

Cremer, T. and Cremer, M. (2010) ‘Chromosome Territories’, *Cold Spring Harbor Perspectives in Biology*. Cold Spring Harbor Laboratory Press, 2(3), p. a003889. doi: 10.1101/CSHPERSPECT.A003889.

Croft, J. A. *et al.* (1999) ‘Differences in the Localization and Morphology of Chromosomes in the Human Nucleus’, *Journal of Cell Biology*. The Rockefeller University Press, 145(6), pp. 1119–1131. doi: 10.1083/JCB.145.6.1119.

Dar, R. D. *et al.* (2012) ‘Transcriptional burst frequency and burst size are equally modulated across the human genome’, *Proceedings of the National Academy of Sciences*, 109(43), pp. 17454–17459. doi: 10.1073/pnas.1213530109.

Davidson, I. F. *et al.* (2016) ‘Rapid movement and transcriptional re-localization of human cohesin on DNA’, *The EMBO Journal*. John Wiley & Sons, Ltd, 35(24), pp. 2671–2685. doi: 10.15252/EMBJ.201695402.

Davidson, I. F. *et al.* (2019) ‘DNA loop extrusion by human cohesin’, *Science*. American Association for the Advancement of Science, 366(6471), pp. 1338–1345. doi: 10.1126/SCIENCE.AAZ3418/SUPPL_FILE/AAZ3418S9.MOV.

Deng, W. *et al.* (2014) ‘Reactivation of Developmentally Silenced Globin Genes by Forced Chromatin Looping’, *Cell*. Cell Press, 158(4), pp. 849–860. doi: 10.1016/J.CELL.2014.05.050.

Desanlis, I. *et al.* (2020) ‘HOX13-dependent chromatin accessibility underlies the transition towards the digit development program’, *Nature Communications 2020 11:1*. Nature Publishing Group, 11(1), pp. 1–10. doi: 10.1038/s41467-020-16317-2.

Despang, A. *et al.* (2019) ‘Functional dissection of the Sox9–Kcnj2 locus identifies nonessential and instructive roles of TAD architecture’, *Nature Genetics 2019 51:8*. Nature Publishing Group, 51(8), pp. 1263–1271. doi: 10.1038/s41588-019-0466-z.

Dixon, J. R. *et al.* (2012) ‘Topological domains in mammalian genomes identified by analysis of chromatin interactions’. doi: 10.1038/nature11082.

Dixon, J. R., Gorkin, D. U. and Ren, B. (2016) ‘Chromatin Domains: the Unit of Chromosome Organization’, *Molecular cell*. NIH Public Access, 62(5), p. 668. doi: 10.1016/J.MOLCEL.2016.05.018.

Dufourt, J. *et al.* (2018) ‘Temporal control of gene expression by the pioneer factor Zelda through transient interactions in hubs’, *Nature Communications 2018 9:1*. Nature Publishing Group, 9(1), pp. 1–13. doi: 10.1038/s41467-018-07613-z.

Ea, V. *et al.* (2015) ‘Contribution of Topological Domains and Loop Formation to 3D Chromatin Organization’, *Genes 2015, Vol. 6, Pages 734-750*. Multidisciplinary Digital Publishing Institute, 6(3), pp. 734–750. doi: 10.3390/GENES6030734.

Echelard, Y. *et al.* (1993) ‘Sonic hedgehog, a member of a family of putative signaling molecules, is implicated in the regulation of CNS polarity’, *Cell*. Cell, 75(7), pp. 1417–1430. doi: 10.1016/0092-8674(93)90627-3.

Elowitz, M. B. *et al.* (2002) ‘Stochastic Gene Expression in a Single Cell’, *Science*, 297(5584), pp. 1183–1186. doi: 10.1126/science.1070919.

Eltsov, M. *et al.* (2008) ‘Analysis of cryo-electron microscopy images does not support the existence of 30-nm chromatin fibers in mitotic chromosomes in situ’, *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 105(50), pp. 19732–19737. doi: 10.1073/PNAS.0810057105/SUPPL_FILE/0810057105SI.PDF.

Epstein, D. J., McMahon, A. P. and Joyner, A. L. (1999) ‘Regionalization of Sonic hedgehog transcription along the anteroposterior axis of the mouse central nervous system is regulated by Hnf3-dependent and -independent mechanisms’, *Development (Cambridge, England)*. Development, 126(2), pp. 281–292. doi: 10.1242/DEV.126.2.281.

Fang, B. *et al.* (2014) ‘Circadian Enhancers Coordinate Multiple Phases of Rhythmic Gene Transcription In Vivo’, *Cell*. Cell Press, 159(5), pp. 1140–1152. doi: 10.1016/J.CELL.2014.10.022.

Femino, A. M. *et al.* (1998) ‘Visualization of single RNA transcripts in situ’, *Science*. American Association for the Advancement of Science, 280(5363), pp. 585–590. doi: 10.1126/SCIENCE.280.5363.585/SUPPL_FILE/975399.XHTML.

Fernandez-Teran, M. and Ros, M. A. (2008) ‘The Apical Ectodermal Ridge: morphological aspects and signaling pathways’, *Int. J. Dev. Biol.* UPV/EHU Press, 52(7), pp. 857–871.

Finch, J. and Klug, A. (1976) ‘Solenoidal Model for Superstructure in Chromatin on JSTOR’, *Proceedings of the National Academy of Sciences of the United States of America*, 73(6), pp. 1897–1901. Available at: https://www.jstor.org/stable/65584#metadata_info_tab_contents (Accessed: 8 November 2022).

Flyamer, I. M. *et al.* (2017) ‘Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition’, *Nature 2017 544:7648*. Nature Publishing Group, 544(7648), pp. 110–114. doi: 10.1038/nature21711.

Fuchs, G. *et al.* (2014) ‘4sUDRB-seq: measuring genomewide transcriptional elongation rates and initiation frequencies within cells’, *Genome biology*. BioMed Central, 15(5), p. R69. doi: 10.1186/GB-2014-15-5-R69/FIGURES/4.

Fudenberg, G. *et al.* (2016) ‘Formation of Chromosomal Domains by Loop Extrusion’, *Cell Reports*. Elsevier, 15(9), pp. 2038–2049. doi: 10.1016/j.celrep.2016.04.085.

Fukaya, T., Lim, B. and Levine, M. (2016) ‘Enhancer Control of Transcriptional Bursting’, *Cell*. Cell Press, 166(2), pp. 358–368. doi: 10.1016/j.cell.2016.05.025.

Furlong, E. E. M. and Levine, M. (2018) ‘Developmental enhancers and

chromosome topology', *Science*. American Association for the Advancement of Science, 361(6409), pp. 1341–1345. doi: 10.1126/science.aau0320.

Fussner, E. *et al.* (2012) 'Open and closed domains in the mouse genome are configured as 10-nm chromatin fibres', *EMBO reports*. John Wiley & Sons, Ltd, 13(11), pp. 992–996. doi: 10.1038/EMBOR.2012.139.

Galli, A. *et al.* (2010) 'Distinct roles of Hand2 in initiating polarity and posterior Shh expression during the onset of mouse limb bud development', *PLoS genetics*. PLoS Genet, 6(4). doi: 10.1371/JOURNAL.PGEN.1000901.

Gasperini, M., Tome, J. M. and Shendure, J. (2020) 'Towards a comprehensive catalogue of validated and target-linked human enhancers', *Nature Reviews Genetics*, 21(5), pp. 292–310. doi: 10.1038/s41576-019-0209-0.

Gassler, J. *et al.* (2017) 'A mechanism of cohesin-dependent loop extrusion organizes zygotic genome architecture', *The EMBO Journal*. John Wiley & Sons, Ltd, 36(24), pp. 3600–3618. doi: 10.15252/EMBJ.201798083.

Gibcus, J. H. and Dekker, J. (2013) 'The Hierarchy of the 3D Genome', *Molecular Cell*. Cell Press, 49(5), pp. 773–782. doi: 10.1016/J.MOLCEL.2013.02.011.

Golding, I. *et al.* (2005) 'Real-Time Kinetics of Gene Activity in Individual Bacteria', *Cell*. Elsevier, 123(6), pp. 1025–1036. doi: 10.1016/j.cell.2005.09.031.

Grosschedl, R. and Birnstiel, M. L. (1980) 'Spacer DNA sequences upstream of the T-A-T-A-A-A-T-A sequence are essential for promotion of H2A histone gene transcription in vivo.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 77(12), p. 7102. doi: 10.1073/PNAS.77.12.7102.

Grosveld, F. *et al.* (1987) 'Position-independent, high-level expression of the human β-globin gene in transgenic mice', *Cell*. Cell Press, 51(6), pp. 975–

985. doi: 10.1016/0092-8674(87)90584-8.

Gruss, P., Dhar, R. and Khoury, G. (1981) ‘Simian virus 40 tandem repeated sequences as an element of the early promoter (deletion mutants/chromatin/RNA initiation/transcriptional regulation)’, *Proc. NatL Acad. Sci. USA*, 78(2), pp. 943–947.

Guelen, L. *et al.* (2008) ‘Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions’, *Nature 2008 453:7197*. Nature Publishing Group, 453(7197), pp. 948–951. doi: 10.1038/nature06947.

Haarhuis, J. H. I. *et al.* (2017) ‘The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension’, *Cell*. Cell Press, 169(4), pp. 693-707.e14. doi: 10.1016/j.cell.2017.04.013.

Hafner, A. *et al.* (2023) ‘Loop stacking organizes genome folding from TADs to chromosomes’, *Molecular Cell*, 83(9), pp. 1377-1392.e6. doi: https://doi.org/10.1016/j.molcel.2023.04.008.

Hagan, D. M. *et al.* (2000) ‘Mutation Analysis and Embryonic Expression of the HLXB9 Currarino Syndrome Gene’, *Am. J. Hum. Genet*, 66, pp. 1504–1515.

Halpern, K. B. and Itzkovitz, S. (2016) ‘Single molecule approaches for quantifying transcription and degradation rates in intact mammalian tissues’, *Methods*. Academic Press, 98, pp. 134–142. doi: 10.1016/J.YMETH.2015.11.015.

Hardy, R. Y. and Resh, M. D. (2012) ‘Identification of N-terminal Residues of Sonic Hedgehog Important for Palmitoylation by Hedgehog Acyltransferase *’, *Journal of Biological Chemistry*. Elsevier, 287(51), pp. 42881–42889. doi: 10.1074/jbc.M112.426833.

Harfe, B. D. *et al.* (2004) ‘Evidence for an Expansion-Based Temporal Shh Gradient in Specifying Vertebrate Digit Identities’, *Cell*. Cell Press, 118(4),

pp. 517–528. doi: 10.1016/J.CELL.2004.07.024.

Hark, A. T. *et al.* (2000) 'CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus', *Nature*, 405(6785), pp. 486–489. doi: 10.1038/35013106.

Hay, D. *et al.* (2016) 'Genetic dissection of the α-globin super-enhancer in vivo', *Nature Genetics 2016 48:8*. Nature Publishing Group, 48(8), pp. 895–903. doi: 10.1038/ng.3605.

Hazra, S., Ghosh, S. and Hazra, B. (2017) 'Chapter 8 - Phytochemicals With Antileishmanial Activity: Prospective Drug Targets', in Atta-ur-Rahman (ed.). Elsevier (Studies in Natural Products Chemistry), pp. 303–336. doi: https://doi.org/10.1016/B978-0-444-63931-8.00008-4.

Heintzman, N. D. *et al.* (2007) 'Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome', *Nature Genetics 2007 39:3*. Nature Publishing Group, 39(3), pp. 311–318. doi: 10.1038/ng1966.

Heintzman, N. D. *et al.* (2009) 'Histone modifications at human enhancers reflect global cell-type-specific gene expression', *Nature*. Nature, 459(7243), pp. 108–112. doi: 10.1038/NATURE07829.

Hendy, O. *et al.* (2017) 'Differential context-specific impact of individual core promoter elements on transcriptional dynamics', *Molecular Biology of the Cell*. American Society for Cell Biology, 28(23), pp. 3360–3370. doi: 10.1091/MBC.E17-06-0408/ASSET/IMAGES/LARGE/3360FIG6.TIF.GZ.JPEG.

Henikoff, S. and Smith, M. M. (2015) 'Histone Variants and Epigenetics', *Cold Spring Harbor Perspectives in Biology*. Cold Spring Harbor Laboratory Press, 7(1). doi: 10.1101/CSHPERSPECT.A019364.

Henriques, T. *et al.* (2018) 'Widespread transcriptional pausing and elongation control at enhancers', *Genes and Development*, (32), pp. 26–41.

Heus, H. C. *et al.* (1999) ‘A Physical and Transcriptional Map of the Preaxial Polydactyly Locus on Chromosome 7q36’, *Genomics*, 57(3), pp. 342–351. doi: https://doi.org/10.1006/geno.1999.5796.

Hnisz, D. *et al.* (2013) ‘Super-enhancers in the control of cell identity and disease’, *Cell*. Cell, 155(4), p. 934. doi: 10.1016/J.CELL.2013.09.053.

Hou, C. *et al.* (2012) ‘Gene Density, Transcription, and Insulators Contribute to the Partition of the Drosophila Genome into Physical Domains’, *Molecular Cell*. Cell Press, 48(3), pp. 471–484. doi: 10.1016/J.MOLCEL.2012.08.031.

Ianakiev, P. *et al.* (2001) ‘Acheiropodia Is Caused by a Genomic Deletion in C7orf2, the Human Orthologue of the Lmbr1 Gene’, *The American Journal of Human Genetics*. Elsevier, 68(1), pp. 38–45. doi: 10.1086/316955.

Imakaev, M. *et al.* (2012) ‘Iterative Correction of Hi-C Data Reveals Hallmarks of Chromosome Organization’, *Nature methods*. NIH Public Access, 9(10), p. 999. doi: 10.1038/NMETH.2148.

Inoue, F. and Ahituv, N. (2015) ‘Decoding enhancers using massively parallel reporter assays’, *Genomics*, 106(3), pp. 159–164. doi: https://doi.org/10.1016/j.ygeno.2015.06.005.

Ivics, Z. *et al.* (1997) ‘Molecular Reconstruction of Sleeping Beauty, a Tc1-like Transposon from Fish, and Its Transposition in Human Cells’, *Cell*. Cell Press, 91(4), pp. 501–510. doi: 10.1016/S0092-8674(00)80436-5.

Iwafuchi-Doi, M. and Zaret, K. S. (2016) ‘Cell fate control by pioneer transcription factors’, *Development*. The Company of Biologists, 143(11), pp. 1833–1837. doi: 10.1242/DEV.133900.

Jackson, D. A. *et al.* (1998) ‘Numbers and organization of RNA polymerases, nascent transcripts, and transcription units in HeLa nuclei’, *Molecular Biology of the Cell*. American Society for Cell Biology, 9(6), pp. 1523–1536. doi: 10.1091/mbc.9.6.1523.

Jeong, Y. *et al.* (2006) ‘A functional screen for sonic hedgehog regulatory elements across a 1 Mb interval identifies long-range ventral forebrain enhancers.’, *Development (Cambridge, England)*, 133(4), pp. 761–72. doi: 10.1242/dev.02239.

Jiao, W. *et al.* (2018) ‘HPSE enhancer RNA promotes cancer progression through driving chromatin looping and regulating hnRNPU/p300/EGR1/HPSE axis’, *Oncogene 2018 37:20*. Nature Publishing Group, 37(20), pp. 2728–2745. doi: 10.1038/s41388-018-0128-0.

Jubb, A. W. *et al.* (2016) ‘Enhancer Turnover Is Associated with a Divergent Transcriptional Response to Glucocorticoid in Mouse and Human Macrophages’, *The Journal of Immunology*, 196(2), pp. 813–822. doi: 10.4049/jimmunol.1502009.

Kaikkonen, M. U. *et al.* (2013) ‘Remodeling of the Enhancer Landscape during Macrophage Activation Is Coupled to Enhancer Transcription’, *Molecular Cell*, 51(3), pp. 310–325. doi: https://doi.org/10.1016/j.molcel.2013.07.010.

Kim, T. K. *et al.* (2010) ‘Widespread transcription at neuronal activity-regulated enhancers’, *Nature 2010 465:7295*. Nature Publishing Group, 465(7295), pp. 182–187. doi: 10.1038/nature09033.

Koch, F. *et al.* (2011) ‘Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters’, *Nature Structural & Molecular Biology*, 18(8), pp. 956–963. doi: 10.1038/nsmb.2085.

Kokubu, C. *et al.* (2009) ‘A transposon-based chromosomal engineering method to survey a large cis-regulatory landscape in mice’, *Nature Genetics 2009 41:8*. Nature Publishing Group, 41(8), pp. 946–952. doi: 10.1038/ng.397.

Kouzarides, T. (2007) ‘Chromatin Modifications and Their Function’, *Cell*, 128(4), pp. 693–705. doi: https://doi.org/10.1016/j.cell.2007.02.005.

Krauss, S., Concordet, J. P. and Ingham, P. W. (1993) ‘A functionally conserved homolog of the Drosophila segment polarity gene hh is expressed in tissues with polarizing activity in zebrafish embryos’, *Cell*. Cell, 75(7), pp. 1431–1444. doi: 10.1016/0092-8674(93)90628-4.

Krivega, I. and Dean, A. (2017) ‘LDB1-mediated enhancer looping can be established independent of mediator and cohesin’, *Nucleic acids research*. Nucleic Acids Res, 45(14), pp. 8255–8268. doi: 10.1093/NAR/GKX433.

Kubo, N. *et al.* (2021) ‘Promoter-proximal CTCF binding promotes distal enhancer-dependent gene activation’, *Nature Structural & Molecular Biology 2021 28:2*. Nature Publishing Group, 28(2), pp. 152–161. doi: 10.1038/s41594-020-00539-5.

Kvon, E. Z. *et al.* (2016) ‘Progressive Loss of Function in a Limb Enhancer during Snake Evolution’, *Cell*. Cell Press, 167(3), pp. 633-642.e11. doi: 10.1016/j.cell.2016.09.028.

Larson, E. D., Marsh, A. J. and Harrison, M. M. (2021) ‘Pioneering the developmental frontier’, *Molecular Cell*. Cell Press, 81(8), pp. 1640–1650. doi: 10.1016/J.MOLCEL.2021.02.020.

Larsson, A. J. M. *et al.* (2019) ‘Genomic encoding of transcriptional burst kinetics’, *Nature*. Nature Publishing Group, 565(7738), pp. 251–254. doi: 10.1038/s41586-018-0836-1.

Lee, J. J. *et al.* (1994) ‘Autoproteolysis in hedgehog protein biogenesis’, *Science (New York, N.Y.)*. Science, 266(5190), pp. 1528–1537. doi: 10.1126/SCIENCE.7985023.

Lettice, L. A. *et al.* (2003) ‘A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly’, *Human Molecular Genetics*. Narnia, 12(14), pp. 1725–1735. doi: 10.1093/hmg/ddg180.

Lettice, L. A. *et al.* (2011) ‘Enhancer-adoption as a mechanism of human

developmental disease', *Human Mutation*. John Wiley & Sons, Ltd, 32(12), pp. 1492–1499. doi: 10.1002/HUMU.21615.

Lettice, L. A. *et al.* (2012) 'Opposing functions of the ETS factor family define Shh spatial expression in limb buds and underlie polydactyly.', *Developmental cell*, 22(2), pp. 459–67. doi: 10.1016/j.devcel.2011.12.010.

Lettice, L. A. *et al.* (2014) 'Development of five digits is controlled by a bipartite long-range cis-regulator', *Development*, 141(8), pp. 1715–1740. doi: 10.1242/dev.095430.

Lettice, L. A. *et al.* (2017) 'The Conserved Sonic Hedgehog Limb Enhancer Consists of Discrete Functional Elements that Regulate Precise Spatial Expression In Brief', *CellReports*, 20, pp. 1396–1408. doi: 10.1016/j.celrep.2017.07.037.

Levsky, J. M. *et al.* (2002) 'Single-cell gene expression profiling.', *Science (New York, N.Y.)*, 297(5582), pp. 836–40. doi: 10.1126/science.1072241.

Li, W. *et al.* (2013) 'Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation', *Nature*, 498(7455), pp. 516–520. doi: 10.1038/nature12210.

Lieberman-Aiden, E. *et al.* (2009) 'Comprehensive mapping of long-range interactions reveals folding principles of the human genome', *Science (New York, N.Y.)*. Science, 326(5950), pp. 289–293. doi: 10.1126/SCIENCE.1181369.

Lim, B. and Levine, M. S. (2021) 'Enhancer-promoter communication: hubs or loops?', *Current Opinion in Genetics and Development*. Elsevier Ltd, pp. 5–9. doi: 10.1016/j.gde.2020.10.001.

Little, S. C., Tikhonov, M. and Gregor, T. (2013) 'Precise Developmental Gene Expression Arises from Globally Stochastic Transcriptional Activity', *Cell*. Elsevier, 154(4), pp. 789–800. doi: 10.1016/j.cell.2013.07.025.

Long, H. K. *et al.* (2020) ‘Loss of Extreme Long-Range Enhancers in Human Neural Crest Drives a Craniofacial Disorder’, *Cell Stem Cell*. Elsevier, 27(5), pp. 765-783.e14. doi: 10.1016/j.stem.2020.09.001.

Long, H. K., Prescott, S. L. and Wysocka, J. (2016) ‘Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution’, *Cell*. Elsevier, 167(5), pp. 1170–1187. doi: 10.1016/j.cell.2016.09.018.

Lu, D. *et al.* (2020) ‘Developmental regulation of cell type-specific transcription by novel promoter-proximal sequence elements’, *Genes & Development* , 34(9–10), pp. 663–677. doi: 10.1101/gad.335331.119 .

Luger, K. *et al.* (1997) ‘Crystal structure of the nucleosome core particle at 2.8 Å resolution’, *Nature 1997 389:6648*. Nature Publishing Group, 389(6648), pp. 251–260. doi: 10.1038/38444.

Lupiáñez, D. G. *et al.* (2015) ‘Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions’, *Cell*. Cell, 161(5), pp. 1012–1025. doi: 10.1016/J.CELL.2015.04.004.

Maass, P. G. *et al.* (2018) ‘Spatiotemporal allele organization by allele-specific CRISPR live-cell imaging: SNP-CLING’, *Nature structural & molecular biology*. NIH Public Access, 25(2), p. 176. doi: 10.1038/S41594-017-0015-3.

Maeshima, K., Hihara, S. and Eltsov, M. (2010) ‘Chromatin structure: does the 30-nm fibre exist in vivo?’, *Current Opinion in Cell Biology*, 22(3), pp. 291–297. doi: https://doi.org/10.1016/j.ceb.2010.03.001.

Martinez Arias, A. and Steventon, B. (2018) ‘On the nature and function of organizers’, *Development (Cambridge)*. Company of Biologists Ltd. doi: 10.1242/dev.159525.

Mayran, A. and Drouin, J. (2018) ‘Pioneer transcription factors shape the epigenetic landscape’, *Journal of Biological Chemistry*. Elsevier, 293(36), pp. 13795–13804. doi: 10.1074/JBC.R117.001232.

Merli, C. *et al.* (1996) ‘Promoter specificity mediates the independent regulation of neighboring genes.’, *Genes & Development* , 10(10), pp. 1260–1270. doi: 10.1101/gad.10.10.1260 .

Moreau, P. *et al.* (1981) ‘The SV40 72 base repair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants.’, *Nucleic Acids Research*. Oxford University Press, 9(22), p. 6047. doi: 10.1093/NAR/9.22.6047.

Morey, C., Kress, C. and Bickmore, W. A. (2009) ‘Lack of bystander activation shows that localization exterior to chromosome territories is not sufficient to up-regulate gene expression’, *Genome Research*. Cold Spring Harbor Laboratory Press, 19(7), pp. 1184–1194. doi: 10.1101/GR.089045.108.

Mueller, F. *et al.* (2013) ‘FISH-quant: automatic counting of transcripts in 3D FISH images’, *Nature Methods 2013 10:4*. Nature Publishing Group, 10(4), pp. 277–278. doi: 10.1038/nmeth.2406.

Müller, H. P., Sogo, J. M. and Schaffner, W. (1989) ‘An enhancer stimulates transcription in Trans when attached to the promoter via a protein bridge’, *Cell*. Cell Press, 58(4), pp. 767–777. doi: 10.1016/0092-8674(89)90110-4.

Murakawa, Y. *et al.* (2016) ‘Enhanced Identification of Transcriptional Enhancers Provides Mechanistic Insights into Diseases’, *Trends in Genetics*, 32(2), pp. 76–88. doi: https://doi.org/10.1016/j.tig.2015.11.004.

Murdoch, J. N. and Copp, A. J. (2010) ‘The relationship between sonic hedgehog signaling, cilia, and neural tube defects’, *Birth Defects Research Part A - Clinical and Molecular Teratology*, 88(8), pp. 633–652. doi: 10.1002/BDRA.20686.

Murone, M., Rosenthal, A. and De Sauvage, F. J. (1999) ‘Sonic hedgehog signaling by the patched-smoothened receptor complex’, *Current biology : CB*. Curr Biol, 9(2), pp. 76–84. doi: 10.1016/S0960-9822(99)80018-9.

Nag, A. *et al.* (2015) ‘Chromatin signature identifies monoallelic gene expression across mammalian cell types’, *G3: Genes, Genomes, Genetics*. Genetics Society of America, 5(8), pp. 1713–1720. doi: 10.1534/G3.115.018853/-/DC1.

Nagano, T. *et al.* (2013) ‘Single cell Hi-C reveals cell-to-cell variability in chromosome structure’, *Nature*. Europe PMC Funders, 502(7469), pp. 59–64. doi: 10.1038/NATURE12593.

Natoli, G. and Andrau, J. C. (2012) ‘Noncoding Transcription at Enhancers: General Principles and Functional Models’, *https://doi.org/10.1146/annurev-genet-110711-155459*. Annual Reviews , 46, pp. 1–19. doi: 10.1146/ANNUREV-GENET-110711-155459.

Niedermaier, M. *et al.* (2005) ‘An inversion involving the mouse Shh locus results in brachydactyly through dysregulation of Shh expression’, *The Journal of clinical investigation*. J Clin Invest, 115(4), pp. 900–909. doi: 10.1172/JCI23675.

Nobrega, M. A. *et al.* (2003) ‘Scanning Human Gene Deserts for Long-Range Enhancers’, *Science*, 302(5644), p. 413. doi: 10.1126/science.1088328.

Nora, E. P. *et al.* (2012) ‘Spatial partitioning of the regulatory landscape of the X-inactivation centre’, *Nature*. Nature, 485(7398), pp. 381–385. doi: 10.1038/NATURE11049.

Nora, E. P. *et al.* (2017) ‘Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization’, *Cell*. Cell Press, 169(5), pp. 930-944.e22. doi: 10.1016/J.CELL.2017.05.004/ATTACHMENT/90A78F58-A750-40FB-9838-146F014D31A6/MMC7.XLSX.

Norton, H. K. *et al.* (2018) ‘Detecting hierarchical genome folding with network modularity’, *Nature Methods*, 15(2), pp. 119–122. doi: 10.1038/nmeth.4560.

Nozaki, T. *et al.* (2017) ‘Dynamic Organization of Chromatin Domains Revealed by Super-Resolution Live-Cell Imaging’. doi: 10.1016/j.molcel.2017.06.018.

Nurk, S. *et al.* (2022) ‘The complete sequence of a human genome’, *Science*, 376(6588), pp. 44–53. doi: 10.1126/science.abj6987.

Nüsslein-Volhard, C. and Wieschaus, E. (1980) ‘Mutations affecting segment number and polarity in Drosophila’, *Nature 1980 287:5785*. Nature Publishing Group, 287(5785), pp. 795–801. doi: 10.1038/287795a0.

Ochiai, H. *et al.* (2014) ‘Stochastic promoter activation affects Nanog expression variability in mouse embryonic stem cells’, *Scientific Reports 2014 4:1*. Nature Publishing Group, 4(1), pp. 1–9. doi: 10.1038/srep07125.

Osterwalder, M. *et al.* (2018) ‘Enhancer redundancy provides phenotypic robustness in mammalian development’, *Nature*, 554(7691), pp. 239–243. doi: 10.1038/nature25461.

Paliou, C. *et al.* (2019) ‘Preformed chromatin topology assists transcriptional robustness of Shh during limb development’. PNAS, 116(25), pp. 12390–12399. doi: 10.1073/pnas.1900672116.

Parelho, V. *et al.* (2008) ‘Cohesins Functionally Associate with CTCF on Mammalian Chromosome Arms’, *Cell*. Elsevier, 132(3), pp. 422–433. doi: 10.1016/j.cell.2008.01.011.

Patwardhan, R. P. *et al.* (2009) ‘High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis’, *Nature Biotechnology*, 27(12), pp. 1173–1175. doi: 10.1038/nbt.1589.

Peluso, S. *et al.* (2017) ‘Fibroblast growth factors (FGFs) prime the limb specific Shh enhancer for chromatin changes that balance histone acetylation mediated by E26 transformation-specific (ETS) factors’, *eLife*, 6. doi: 10.7554/eLife.28590.

Pennacchio, L. A. *et al.* (2006) 'In vivo enhancer analysis of human conserved non-coding sequences', *Nature*, 444(7118), pp. 499–502. doi: 10.1038/nature05295.

Pennacchio, L. A. *et al.* (2013) 'Enhancers: Five essential questions', *Nature Reviews Genetics*. NIH Public Access, pp. 288–295. doi: 10.1038/nrg3458.

Pepinsky, R. B. *et al.* (1998) 'Identification of a Palmitic Acid-modified Form of Human Sonic hedgehog *', *Journal of Biological Chemistry*. Elsevier, 273(22), pp. 14037–14045. doi: 10.1074/jbc.273.22.14037.

Phillips-Cremins, J. E. *et al.* (2013) 'Architectural Protein Subclasses Shape 3D Organization of Genomes during Lineage Commitment', *Cell*. Cell Press, 153(6), pp. 1281–1295. doi: 10.1016/J.CELL.2013.04.053.

Porter, J. A. *et al.* (1995) 'The product of hedgehog autoproteolytic cleavage active in local and long-range signalling', *Nature 1995 374:6520*. Nature Publishing Group, 374(6520), pp. 363–366. doi: 10.1038/374363a0.

Porter, J. A. *et al.* (1996) 'Hedgehog Patterning Activity: Role of a Lipophilic Modification Mediated by the Carboxy-Terminal Autoprocessing Domain', *Cell*. Cell Press, 86(1), pp. 21–34. doi: 10.1016/S0092-8674(00)80074-4.

Porter, J. A., Young, K. E. and Beachy, P. A. (1996) 'Cholesterol Modification of Hedgehog Signaling Proteins in Animal Development', *Science*. American Association for the Advancement of Science , 274(5285), pp. 255–259. doi: 10.1126/SCIENCE.274.5285.255.

R Core Team (2018) 'R: A Language and Environment for Statistical Computing'. Vienna, Austria. Available at: https://www.r-project.org/.

Rabl, C. (1885) *Uber Zelltheilung. Morphologisches Jahrbuch.* Edited by Gegenbaur C. W. Engelmann.

Raj, A. *et al.* (2006) 'Stochastic mRNA Synthesis in Mammalian Cells', *PLoS Biology*. Edited by U. Schibler, 4(10), p. e309. doi:

10.1371/journal.pbio.0040309.

Rao, S. S. P. *et al.* (2014) ‘A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping’, *Cell*. Cell Press, 159(7), pp. 1665–1680. doi: 10.1016/j.cell.2014.11.021.

Rao, S. S. P. *et al.* (2017) ‘Cohesin Loss Eliminates All Loop Domains’, *Cell*. Cell, 171(2), pp. 305-320.e24. doi: 10.1016/J.CELL.2017.09.026.

Reddy, K. L. *et al.* (2008) ‘Transcriptional repression mediated by repositioning of genes to the nuclear lamina’, *Nature 2008 452:7184*. Nature Publishing Group, 452(7184), pp. 243–247. doi: 10.1038/nature06727.

Revyakin, A. *et al.* (2006) ‘Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching’, *Science*. American Association for the Advancement of Science , 314(5802), pp. 1139–1143. doi: 10.1126/SCIENCE.1131398/SUPPL_FILE/REVYAKIN.SOM.PDF.

Ricci, M. A. *et al.* (2015) ‘Chromatin Fibers Are Formed by Heterogeneous Groups of Nucleosomes In Vivo’, *Cell*. Cell Press, 160(6), pp. 1145–1158. doi: 10.1016/J.CELL.2015.01.054.

Riddle, R. D. *et al.* (1993) ‘Sonic hedgehog mediates the polarizing activity of the ZPA’, *Cell*. Cell Press, 75(7), pp. 1401–1416. doi: 10.1016/0092-8674(93)90626-2.

Rivera-Pomar, R. and Jäckle, H. (1996) ‘From gradients to stripes in Drosophila embryogenesis: Filling in the gaps’, *Trends in Genetics*. Elsevier Ltd, pp. 478–483. doi: 10.1016/0168-9525(96)10044-5.

Rodríguez-Carballo, E. *et al.* (2017) ‘The HoxD cluster is a dynamic and resilient TAD boundary controlling the segregation of antagonistic regulatory landscapes’, *Genes & development*. Genes Dev, 31(22), pp. 2264–2281. doi: 10.1101/GAD.307769.117.

Ryba, T. *et al.* (2010) ‘Evolutionarily conserved replication timing profiles

predict long-range chromatin interactions and distinguish closely related cell types', *Genome research*. Genome Res, 20(6), pp. 761–770. doi: 10.1101/GR.099655.109.

Sagai, T. *et al.* (2005) 'Elimination of a long-range cis-regulatory module causes complete loss of limb-specific Shh expression and truncation of the mouse limb', *Development*. The Company of Biologists, 132(4), pp. 797–803. doi: 10.1242/DEV.01613.

Sagai, T. *et al.* (2009) 'A cluster of three long-range enhancers directs regional Shh expression in the epithelial linings.', *Development (Cambridge, England)*, 136(10), pp. 1665–74. doi: 10.1242/dev.032714.

Sagai, T. *et al.* (2019) 'SHH signaling mediated by a prechordal and brain enhancer controls forebrain organization', *Proceedings of the National Academy of Sciences of the United States of America*. Proc Natl Acad Sci U S A, 116(47), pp. 23636–23642. doi: 10.1073/PNAS.1901732116.

Sanborn, A. L. *et al.* (2015) 'Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 112(47), pp. E6456–E6465. doi: 10.1073/pnas.1518552112.

de Santa, F. *et al.* (2010) 'A Large Fraction of Extragenic RNA Pol II Transcription Sites Overlap Enhancers', *PLOS Biology*. Public Library of Science, 8(5), p. e1000384. doi: 10.1371/JOURNAL.PBIO.1000384.

Saunders, J. . and Gasseling, M. T. (1968) *Ectodermal-mesenchymal interactions in the origin of limb symmetry*, *Epithelial-mesenchymal Interactions*. Edited by R. Fleischmajer and R. E. Billingham. Baltimore, Maryland: Williams and Wilkins.

Schoenfelder, S. *et al.* (2010) 'Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells',

Nature genetics. Nat Genet, 42(1), pp. 53–61. doi: 10.1038/NG.496.

Schwarzer, W. *et al.* (2017) 'Two independent modes of chromatin organization revealed by cohesin removal', *Nature*, 551(7678), pp. 51–56. doi: 10.1038/nature24281.

Senecal, A. *et al.* (2014) 'Transcription Factors Modulate c-Fos Transcriptional Bursts', *Cell Reports*. Cell Press, 8(1), pp. 75–83. doi: 10.1016/J.CELREP.2014.05.053.

Sexton, T. *et al.* (2012) 'Three-Dimensional Folding and Functional Organization Principles of the Drosophila Genome', *Cell*. Cell Press, 148(3), pp. 458–472. doi: 10.1016/J.CELL.2012.01.010.

Shen, Y. *et al.* (2012) 'A map of the cis-regulatory sequences in the mouse genome', *Nature 2012 488:7409*. Nature Publishing Group, 488(7409), pp. 116–120. doi: 10.1038/nature11243.

Shii, L. *et al.* (2017) 'SERPINB2 is regulated by dynamic interactions with pause-release proteins and enhancer RNAs', *Molecular Immunology*. Pergamon, 88, pp. 20–31. doi: 10.1016/J.MOLIMM.2017.05.005.

Sigova, A. A. *et al.* (2015) 'Transcription factor trapping by RNA in gene regulatory elements', *Science*. American Association for the Advancement of Science, 350(6263), pp. 978–991. doi: 10.1126/SCIENCE.AAD3346/SUPPL_FILE/SIGOVA.SM.REVISION.1.PDF.

Skene, P. J. and Henikoff, S. (2017) 'An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites', *eLife*. Edited by D. Reinberg. eLife Sciences Publications, Ltd, 6, p. e21856. doi: 10.7554/eLife.21856.

Slatkin, M. (2008) 'Linkage disequilibrium — understanding the evolutionary past and mapping the medical future', *Nature Reviews Genetics 2008 9:6*. Nature Publishing Group, 9(6), pp. 477–485. doi: 10.1038/nrg2361.

Song, W., Sharan, R. and Ovcharenko, I. (2019) ‘The first enhancer in an enhancer chain safeguards subsequent enhancer-promoter contacts from a distance’, *Genome Biology*, 20(1), p. 197. doi: 10.1186/s13059-019-1808-y.

Spitz, F., Gonzalez, F. and Duboule, D. (2003) ‘A Global Control Region Defines a Chromosomal Regulatory Landscape Containing the HoxD Cluster’, *Cell*. Elsevier, 113(3), pp. 405–417. doi: 10.1016/S0092-8674(03)00310-6.

Stavreva, D. A. *et al.* (2019) ‘Transcriptional Bursting and Co-bursting Regulation by Steroid Hormone Release Pattern and Transcription Factor Mobility’, *Molecular Cell*. Cell Press, 75(6), pp. 1161-1177.e11. doi: 10.1016/J.MOLCEL.2019.06.042.

von Steyern, F. V. *et al.* (1999) ‘The homeodomain transcription factors Islet 1 and HB9 are expressed in adult alpha and gamma motoneurons identified by selective retrograde tracing’, *European Journal of Neuroscience*, 11(6), pp. 2093–2102. doi: https://doi.org/10.1046/j.1460-9568.1999.00631.x.

Sun, H. Bin, Shen, J. and Yokota, H. (2000) ‘Size-Dependent Positioning of Human Chromosomes in Interphase Nuclei’, *Biophysical Journal*. Cell Press, 79(1), pp. 184–190. doi: 10.1016/S0006-3495(00)76282-5.

Sun, M. and Zhang, J. (2019) ‘Chromosome-wide co-fluctuation of stochastic gene expression in mammalian cells’, *PLOS Genetics*. Public Library of Science (PLoS), 15(9), p. e1008389. doi: 10.1371/journal.pgen.1008389.

Suter, D. M. *et al.* (2011) ‘Mammalian genes are transcribed with widely different bursting kinetics’, *Science*. American Association for the Advancement of Science, 332(6028), pp. 472–474. doi: 10.1126/SCIENCE.1198817/SUPPL_FILE/SUTER.SOM.PDF.

Symmons, O. *et al.* (2014) ‘Functional and topological characteristics of mammalian regulatory domains’, *Genome Research*. Cold Spring Harbor Laboratory Press, 24(3), p. 390. doi: 10.1101/GR.163519.113.

Symmons, O. *et al.* (2016a) 'The Shh Topological Domain Facilitates the Action of Remote Enhancers by Reducing the Effects of Genomic Distances', *Developmental Cell*. Cell Press, 39(5), pp. 529–543. doi: 10.1016/j.devcel.2016.10.015.

Symmons, O. *et al.* (2016b) 'The Shh Topological Domain Facilitates the Action of Remote Enhancers by Reducing the Effects of Genomic Distances', *Developmental Cell*. Elsevier, 39(5), p. 529. doi: 10.1016/J.DEVCEL.2016.10.015.

Symmons, O. *et al.* (2019) 'Allele-specific RNA imaging shows that allelic imbalances can arise in tissues through transcriptional bursting', *PLoS Genetics*. Public Library of Science, 15(1). doi: 10.1371/journal.pgen.1007874.

Tabata, T. and Takei, Y. (2004) 'Morphogens, their identification and regulation', *Development*, 131(4), pp. 703–712. doi: 10.1242/dev.01043.

Taverna, S. D. *et al.* (2007) 'How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers', *Nature Structural & Molecular Biology*, 14(11), pp. 1025–1040. doi: 10.1038/nsmb1338.

Thanos, D. and Maniatis, T. (1995) 'Virus induction of human IFNβ gene expression requires the assembly of an enhanceosome', *Cell*. Cell Press, 83(7), pp. 1091–1100. doi: 10.1016/0092-8674(95)90136-1.

The ENCODE Project Consortium (2007) 'Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project', *Nature*, 447(7146), pp. 799–816. doi: 10.1038/nature05874.

The ENCODE Project Consortium (2012) 'An integrated encyclopedia of DNA elements in the human genome', *Nature*. Nature Publishing Group, 489(7414), pp. 57–74. doi: 10.1038/nature11247.

Therizols, P. *et al.* (2014) 'Chromatin decondensation is sufficient to alter

nuclear organization in embryonic stem cells', *Science*. American Association for the Advancement of Science, 346(6214), pp. 1238–1242. doi: 10.1126/SCIENCE.1259587/SUPPL_FILE/THERIZOLS.SM.PDF.

Thoma, F., Koller, T. and Klug, A. (1979) 'Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin', *The Journal of Cell Biology*. The Rockefeller University Press, 83(2), p. 403. doi: 10.1083/JCB.83.2.403.

Tsukiji, N., Amano, T. and Shiroishi, T. (2014) 'A novel regulatory element for Shh expression in the lung and gut of mouse embryos', *Mechanisms of Development*. Elsevier, 131(1), pp. 127–136. doi: 10.1016/J.MOD.2013.09.003.

Vietri Rudan, M. *et al.* (2015) 'Comparative Hi-C Reveals that CTCF Underlies Evolution of Chromosomal Domain Architecture', *Cell Reports*. Elsevier, 10(8), pp. 1297–1309. doi: 10.1016/j.celrep.2015.02.004.

Visel, A. *et al.* (2007) 'VISTA Enhancer Browser—a database of tissue-specific human enhancers', *Nucleic Acids Research*, 35(suppl_1), pp. D88–D92. doi: 10.1093/nar/gkl822.

Visel, A. *et al.* (2008) 'Ultraconservation identifies a small subset of extremely constrained developmental enhancers', *Nature Genetics*, 40(2), pp. 158–160. doi: 10.1038/ng.2007.55.

Visel, A. *et al.* (2009) 'ChIP-seq accurately predicts tissue-specific activity of enhancers', *Nature 2009 457:7231*. Nature Publishing Group, 457(7231), pp. 854–858. doi: 10.1038/nature07730.

Vortkamp, A. *et al.* (1996) 'Regulation of rate of cartilage differentiation by Indian hedgehog and PTH-related protein', *Science (New York, N.Y.)*. Science, 273(5275), pp. 613–622. doi: 10.1126/SCIENCE.273.5275.613.

Waddington, C. H. and Schmidt, G. A. (1933) 'Induction by heteroplastic grafts of the primitive streak in birds.', *Wilhelm Roux' Archiv fur*

Entwicklungsmechanik der Organismen, 128(3), pp. 522–563. doi: 10.1007/BF00649863.

Walters, M. C. *et al.* (1995) 'Enhancers increase the probability but not the level of gene expression.', *Proceedings of the National Academy of Sciences*, 92(15), pp. 7125–7129. doi: 10.1073/pnas.92.15.7125.

Wang, S. *et al.* (2016) 'Spatial organization of chromatin domains and compartments in single chromosomes', *Science*. American Association for the Advancement of Science, 353(6299), pp. 598–602. doi: 10.1126/SCIENCE.AAF8084/SUPPL_FILE/WANG-SM.PDF.

Warnes, G. *et al.* (2021) 'genetics: Population Genetics'. Available at: https://cran.r-project.org/package=genetics.

Watson, B. A. *et al.* (2018) 'LHX2 Mediates the FGF-to-SHH Regulatory Loop during Limb Development', *Journal of Developmental Biology*. Multidisciplinary Digital Publishing Institute (MDPI), 6(2). doi: 10.3390/JDB6020013.

Te Welscher, P. *et al.* (2002) 'Progression of vertebrate limb development through SHH-mediated counteraction of GLI3', *Science (New York, N.Y.)*. Science, 298(5594), pp. 827–830. doi: 10.1126/SCIENCE.1075620.

Whyte, W. A. *et al.* (2013) 'Master transcription factors and mediator establish super-enhancers at key cell identity genes', *Cell*. Cell, 153(2), pp. 307–319. doi: 10.1016/J.CELL.2013.03.035.

Williamson, I. *et al.* (2016) 'Shh and ZRS enhancer colocalisation is specific to the zone of polarising activity'. doi: 10.1242/dev.139188.

Williamson, I. *et al.* (2019a) ' Developmentally regulated Shh expression is robust to TAD perturbations ', *Development*. The Company of Biologists, 146(19), p. dev179523. doi: 10.1242/dev.179523.

Williamson, I. *et al.* (2019b) ' Developmentally regulated Shh expression is

robust to TAD perturbations ', *Development*. The Company of Biologists, p. dev.179523. doi: 10.1242/dev.179523.

Wolpert, L. (1969) *Positional Information and the Spatial Pattern of Cellular Differentiationt*, *Theoret. Biol*. Available at: https://www.mbl.edu/physiology/files/2014/06/Wolpert1969.pdf (Accessed: 21 May 2019).

Wong, H., Victor, J. M. and Mozziconacci, J. (2007) 'An All-Atom Model of the Chromatin Fiber Containing Linker Histones Reveals a Versatile Structure Tuned by the Nucleosomal Repeat Length', *PLoS ONE*. PLOS, 2(9). doi: 10.1371/JOURNAL.PONE.0000877.

Woodcock, C. L., Skoultchi, A. I. and Fan, Y. (2006) 'Role of linker histone in chromatin structure and function: H1 stoichiometry and nucleosome repeat length', *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology*. Chromosome Res, 14(1), pp. 17–25. doi: 10.1007/S10577-005-1024-3.

Woolfe, A. *et al.* (2004) 'Highly Conserved Non-Coding Sequences Are Associated with Vertebrate Development', *PLOS Biology*. Public Library of Science, 3(1), p. null. doi: 10.1371/journal.pbio.0030007.

Wu, C., Bassett, A. and Travers, A. (2007) 'A variable topology for the 30-nm chromatin fibre', *EMBO Reports*. European Molecular Biology Organization, 8(12), p. 1129. doi: 10.1038/SJ.EMBOR.7401115.

Wutz, G. *et al.* (2017) 'Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins', *The EMBO Journal*, 36(24), pp. 3573–3599. doi: https://doi.org/10.15252/embj.201798004.

Yao, Y. *et al.* (2016) 'Cis-regulatory architecture of a brain-signaling center predates the origin of chordates', *Nature genetics*. NIH Public Access, 48(5), p. 575. doi: 10.1038/NG.3542.

Zaret, K. S. (2020) ‘Pioneer Transcription Factors Initiating Gene Network Changes’, *https://doi.org/10.1146/annurev-genet-030220-015007*. Annual Reviews , 54, pp. 367–385. doi: 10.1146/ANNUREV-GENET-030220-015007.

Zhu, J. *et al.* (2008) ‘Uncoupling Sonic Hedgehog Control of Pattern and Expansion of the Developing Limb Bud’, *Developmental Cell*. Cell Press, 14(4), pp. 624–632. doi: 10.1016/J.DEVCEL.2008.01.008.

Zoller, B., Little, S. C. and Gregor, T. (2018) ‘Diverse Spatial Expression Patterns Emerge from Unified Kinetics of Transcriptional Bursting’, *Cell*. Cell Press, 175(3), pp. 835-847.e25. doi: 10.1016/j.cell.2018.09.056.

Zwemer, L. M. *et al.* (2012) ‘Autosomal monoallelic expression in the mouse’, *Genome Biology*. BioMed Central, 13(2), pp. 1–13. doi: 10.1186/GB-2012-13-2-R10/FIGURES/5.

frontiers in Cell and Developmental Biology

ORIGINAL RESEARCH
published: 01 April 2021
doi: 10.3389/fcell.2021.595744

A Highly Conserved *Shh* Enhancer Coordinates Hypothalamic and Craniofacial Development

Zoe Crane-Smith[1†], Jeffrey Schoenebeck[2], Katy A. Graham[1], Paul S. Devenney[1], Lorraine Rose[1], Mark Ditzell[3], Eve Anderson[1†], Joseph I. Thomson[1], Natasha Klenin[4], Deborah M. Kurrasch[4], Laura A. Lettice[1] and Robert E. Hill[1]*

[1] Medical Research Council Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, United Kingdom, [2] The Roslin Institute and Royal (Dick) School for Veterinary Studies, University of Edinburgh, Edinburgh, United Kingdom, [3] Cancer Research UK Edinburgh Centre, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, United Kingdom, [4] Department of Medical Genetics, Alberta Children's Hospital Research Institute and Hotchkiss Brain Institute, University of Calgary, Calgary, AB, Canada

OPEN ACCESS

Edited by:
Ruth Maree Arkell,
Australian National University,
Australia

Reviewed by:
Rolf Stottmann,
Cincinnati Children's Hospital Medical Center, United States
Takanori Amano,
RIKEN BioResource Research Center (BRC), Japan
Robert J. Lipinski,
University of Wisconsin-Madison, United States

***Correspondence:**
Robert E. Hill
bob.hill@igmm.ed.ac.uk

†Present address:
Zoe Crane-Smith,
Developmental Biology and Cancer Programme, UCL Great Ormond Street Institute of Child Health, University College London, London, United Kingdom
Eve Anderson,
Cancer Research UK Beatson Institute, Glasgow, United Kingdom

Specialty section:
This article was submitted to Developmental Epigenetics, a section of the journal Frontiers in Cell and Developmental Biology

Received: 17 August 2020
Accepted: 28 February 2021
Published: 01 April 2021

Enhancers that are conserved deep in evolutionary time regulate characteristics held in common across taxonomic classes. Here, deletion of the highly conserved *Shh* enhancer SBE2 (*Shh* brain enhancer 2) in mouse markedly reduced *Shh* expression within the embryonic brain specifically in the rostral diencephalon; however, no abnormal anatomical phenotype was observed. Secondary enhancer activity was subsequently identified which likely mediates low levels of expression. In contrast, when crossing the SBE2 deletion with the *Shh* null allele, brain and craniofacial development were disrupted; thus, linking SBE2 regulated *Shh* expression to multiple defects and further enabling the study of the effects of differing levels of *Shh* on embryogenesis. Development of the hypothalamus, derived from the rostral diencephalon, was disrupted along both the anterior-posterior (AP) and the dorsal-ventral (DV) axes. Expression of DV patterning genes and subsequent neuronal population induction were particularly sensitive to *Shh* expression levels, demonstrating a novel morphogenic context for *Shh*. The role of SBE2, which is highlighted by DV gene expression, is to step-up expression of *Shh* above the minimal activity of the second enhancer, ensuring the necessary levels of *Shh* in a regional-specific manner. We also show that low *Shh* levels in the diencephalon disrupted neighbouring craniofacial development, including mediolateral patterning of the bones along the cranial floor and viscerocranium. Thus, SBE2 contributes to hypothalamic morphogenesis and ensures there is coordination with the formation of the adjacent midline cranial bones that subsequently protect the neural tissue.

Keywords: *Shh*, enhancer, hypothalamic development, craniofacial formation, embryonic development

INTRODUCTION

Long-distance enhancers that are conserved across multiple classes of vertebrates are implicated in regulating early embryonic development (Nord et al., 2013). The *Shh* regulatory domain, which extends 1 Mb upstream of the gene contains a number of sequence elements identified by high sequence conservation. These elements appear to be largely responsible for the typical pattern of *Shh* expression observed at initial stages of organogenesis (Jeong et al., 2008; Anderson et al., 2014). The deep conservation and early stages of expression mediated by these *Shh* enhancers suggest a fundamental role in generating ancient, shared vertebrate characteristics. Studies to date show a modular composition of the *Shh* regulatory domain and thus, for at least some of the *cis*-regulators, there is a simple relationship of a single enhancer to a distinct spatial domain of expression (Jeong et al., 2006; Lettice et al., 2014); although, in other instances, secondary enhancers provide compensatory low levels of expression (Tsukiji et al., 2014; Letelier et al., 2018).

SHH is a signalling factor with a fundamental role in brain and craniofacial development. A number of enhancers have been identified within the *Shh* regulatory domain that are responsible for regulating expression in the midbrain and forebrain. Transgenic analysis indicates that the SBE2 enhancer (Jeong et al., 2006, 2008; Zhao et al., 2012) regulates *Shh* expression in the early developing forebrain, from the zona limitans intrathalmica (ZLI) to the medial ganglionic eminence (MGE), specifically within the rostral diencephalon (Jeong et al., 2006; Zhao et al., 2012). Of particular interest is a clinically relevant inactivating mutation in the SBE2 enhancer associated with features of semilobar holoprosencephaly (Jeong et al., 2008). Holoprosencephaly (HPE) is a developmental disorder of both the forebrain and midline facial structures that include microcephaly, midfacial hypoplasia, cleft lip and palate and moderate fusion of the hypothalamus and occasionally includes the condition of diabetes insipidus (Marcorelles and Laquerriere, 2010). More commonly, HPE is associated with heterozygous mutations that inactivate the *Shh* gene (Chiang et al., 1996; Ming and Muenke, 2002). These data suggest that disruption of SBE2 driven expression in the early developing forebrain may contribute to this disorder (Jeong et al., 2008).

Here, we deleted the SBE2 enhancer to discern its overall contribution to the *Shh* expression pattern and to establish its requirement for mediating the embryonic phenotype. *Shh* expression is continuous throughout the central nervous system, relying on a number of *cis*-regulators to generate the overall pattern (Anderson et al., 2014). The deletion of SBE2 interrupted this spatial pattern by specifically downregulating expression in the ventral portion of the rostral diencephalon; however, this reduction of SBE2 driven *Shh* expression produced no overt phenotype. We confirmed the presence of a second enhancer (Sagai et al., 2019) responsible for residual levels of *Shh* expression that compensated for the loss of SBE2 in the rostral diencephalon. Further reduction in expression levels of *Shh* was attained by crossing the SBE2 deletion mice to the line carrying the *Shh* null allele. The development of the hypothalamus and of the surrounding tissue was disrupted in the compound mutants, linking diencephalic spatiotemporal expression to multiple defects and enabling analysis of development at different expression levels. We showed, firstly, that low levels of *Shh* are sufficient for patterning in the hypothalamus along the AP axis to occur. Further reductions in the levels of *Shh*, however, led to misplaced spatial assignment of the hypophyseal lobes due to disruption of the AP boundary. Secondly, gene expression along the DV axis was, in contrast, concentration sensitive and consistent with SHH functioning as a morphogen along this developmental axis, with tissue patterning and eventual neuronal fate reliant upon varying concentrations of *Shh* signalling. Thirdly, we found that loss of SBE2 activity disrupts craniofacial development affecting mediolateral patterning of the bones along the cranial floor and palatal shelf, revealing that SBE2 activity in the brain regulates cranial and facial morphogenesis, providing further evidence of the ventral patterning role played by *Shh* (Dworkin et al., 2016; Corman et al., 2018). Hence, a single regulatory component coordinates the combined function of directing brain development with the concomitant role of ensuring physical neuroprotection.

MATERIALS AND METHODS

Embryo Production and Transgenic Targeting

All animal experiments were reviewed and approved by the University of Edinburgh Animal Welfare and Ethics Committee and were conducted with appropriate licensing under Animals (Scientific Procedures) Act 1986.

The Shh^{null}, $Shh^{\Delta SBE2}$ (described below) and SBLac526 lines were maintained on a C57BL/6J background and crossed to this strain for at least 5 generations prior to experimental collection. Embryos were harvested at various embryonic stages between E9.5 and E17.5. For all experiments described triplicate datasets were used, unless otherwise stated. Heterozygous loss of *Shh*, $Shh^{null/+}$, is known to cause no phenotypic effects (Chiang et al., 1996). Nevertheless differences between wild-type, $Shh^{\Delta SBE2/+}$ and $Shh^{null/+}$ embryos were assessed. As expected, no differences were detected between the three previously mentioned genotypes. Unless otherwise stated, control embryos were $Shh^{+/+}$ wild-types.

The expression driven by a potential enhancer region of interest was analysed in G0 transgenic embryos using a *lacZ* construct which carried sequence of the region in question, a β-globin minimal promoter and *lacZ* as has been previously described (Yee and Rigby, 1993). The DNA fragment of interest was generated using PCR with the primers GATCAT**GTCGAC**GCTCCAGGTACTGCTGTTCAG and GAT CAT**GCTGAC**ATGTGGATGGCAAGCATTGGC- the *Sal*I restriction site used in the cloning is highlighted in bold.

ATAC-Seq and FAC Sorting

Entire heads of GFP positive eGFP/Cre-*Shh* (Harfe et al., 2004) embryos were dissected and single cell suspensions were made by incubating the dissected tissue in 1:5 dilution of trypsin:versene

at 37°C for 15 min. GFP positive cells were sorted using a fluorescence activated cell sorter and ATAC-seq libraries were made from the positive cell population pooled across littermates as previously described (Buenrostro et al., 2015). Subsequently libraries were subject to size selection to exclude DNA fragments larger than 1 kb using SPRIselect magnetic beads (Beckman Coulter). Samples were sequenced on a Illumina HiSeq 4000 platform to obtain 50 bp paired-end reads. Resulting reads were trimmed using cutadapt paired-end trimming, and subsequently aligned to the mm9 mouse genome using bowtie2 paired-end alignment. Unmapped reads and those mapping to the mitochondrial genome were then removed and duplicate reads were filtered out using Picard MarkDuplicates. Reads were shifted by +4 bp for those mapping to the positive strand and −5 bp for those mapping to the negative strand. Broad peaks were then called, to identify open chromatin regions, using MACS2 callpeak, using –g mm, -f BEDPE, -B –broads and cutoff 0.01 options. The ATAC-seq data have been deposited at the NCBI Gene Expression Omnibus (GEO) database (GEO accession number: GSE158074).

SBE2 Deletion Line Production by CRISPR

CRISPR guide oligo pairs were designed flanking the SBE2 enhancer region described by Zhao et al. (2012). The pairs were annealed and cloned into the pX330 vector.

Guide sequences used were: Upstream of SBE2, 5′ AAACACA TTAAAGCCCTCCAGCG 3′ and 5′ CACCGACGCTGGA GGGCTTTAATGT 3′; Downstream of SBE2, 5′ AAACACGAG CAAGCCAACCGGAGG 3′ and 5′ CACCGCTCCGGTTGGCT TGCTCGT 3′. pX330-U6-Chimeric_BB-CBh-hSpCas9 was a gift from Feng Zhang (Addgene plasmid # 42230[1]; RRID:Addgene_42230) (Cong et al., 2013).

Transgenic mice were made by pronuclear injection of plasmid DNA into C57BL/6J × CBA F2 embryos. G0 pups were genotyped by PCR and sequencing. Pups carrying the correct deletions were then used to establish the line by mating to C57BL/6J mice.

OPT Analysis

OPT imaging was performed (Sharpe, 2003). Briefly, PFA-fixed embryos were embedded in 1% low-melting-point agarose and dehydrated by immersion in methanol for 24 h. The samples were cleared for 24 h in BABB (one part benzyl alcohol/two parts benzyl benzoate). Samples were then scanned using a Bioptonics 3001 scanner[2], images taken every 0.9° (of a 360° rotation) and were reconstructed using Bioptonics proprietary software with the outputs then being viewed with Dataviewer (Bioptonics) and Bioptonics Viewer.

Morphometric Analysis

MicroCTs of fifteen E17.5 embryos (7 controls, 8 mutants) collected from three litters were scanned post mortem using a Bruker Skyscanner 1076. Individual skull scans were downsampled 2:1 using Bruker NRecon. Reconstructed images were converted to DICOM format using Bruker DicomCT. Twenty-eight landmarks were placed on each isosurface using Stratovan CheckPoint (v19.03.04.1102 Win64). In two mutants, dysmorphology of the vomer prevented placement of corresponding landmarks and as such, the landmarks were indicated as "missing." Raw coordinates data saved in Morphologika format were processed in R (v4.0.2) using Geomorph (v3.3.2), Rvcg (v0.19.2), Morpho (v2.8), and Hotelling (v1.0-5) packages. Missing landmarks were estimated using a multivariate regression method implemented in Geomorph. Following, a Generalised Procrustes analysis (GPA) was translated, rotated and rescaled raw landmark coordinate data to minimise the sum of squared distances between specimens. The amount of rescaling required in the GPA is captured by the centroid size, a variable which is used subsequently as a proxy for specimen size. A single outlier (control, Litter 3) was removed as its Procrustes distance was >1 standard deviation from the mean. Following removal of the outlier, the GPA was rerun to produce Procrustes shape residuals. Model effects of shape, genotype, size and litter were tested using the non-parametric method of multivariate ANOVA implemented in Geomorph. Significant model effects included size, genotype and litter nested in genotype; the best model was described by coordinates ∼log(centroid size)+ genotype/litter.

In order to visualise shape differences, a randomly chosen ply surface file was warped using thin plate splines using the mean coordinate values of all fourteen specimens. The resulting mesh represented the dataset reference which was subsequently warped to using principal component minimum and maximum shape variables.

Tissue Processing

Embryos were dissected at the stages indicated and fixed overnight at 4°C in 4% paraformaldehyde (PFA) in phosphate-buffered saline (PBS). For cryosectioning embryos were washed briefly in PBS, and then passed through a sucrose gradient: 30 min in 5% sucrose in PBS, 30 min in 10% sucrose in PBS before being placed in 20% sucrose in PBS overnight at 4°C. Embryos were embedded in OCT compound and stored at −80°C. Sections were cut on a cryostat at varying thicknesses appropriate to the embryonic stage: For E11.5 at 8 μm, for E13.5 at 10 μm, and for E17.5 at 14 μm. For paraffin sectioning E10.5 embryos were paraffin embedded and subsequently 6 μm thick sections were cut.

H&E Staining

Cryosections were allowed to reach room temperature before the slides were rehydrated by passage through 100% EtOH 3× 5 min then serial passage through 90, 70, 50, and lastly 30% EtOH for a couple of min each. Slides were washed in H_2O and stained in haemotoxylin for 4 min. They were rinsed in H_2O and subsequently differentiated in acid/alcohol for a few seconds. Slides were again rinsed in H_2O then placed in saturated lithium carbonate solution for a few seconds. They were rinsed in H_2O, stained in eosin for 2 min and again rinsed in H_2O. Slides were dehydrated through 4 changes of 100% ethanol

[1] http://n2t.net/addgene:42230

[2] www.biooptonics.com

and passaged 3× through xylene for 5 min each, subsequently mounted in DPX (Cell Path, SEA-1304-00A), coverslipped and allowed to dry overnight.

lacZ Expression Analysis, Dawson Staining, and Wholemount *in situ* Hybridisation

Embryos were analysed for *lacZ* expression between E10.5 and E14.5 through staining for β-gal activity as previously described (Mackenzie et al., 1997). Skeletal preparations from E17.5 foetuses were stained simultaneously with Alizarin Red and Alcian Blue (Nagy et al., 2009a,b). Whole-mount *in situ* hybridisation on half heads was performed as previously described (Hecksher-Sørensen et al., 1998). Whole-mount *in situ* hybridisation on E9.5 day embryos was performed as previously described (Pai et al., 2015).

RNA FISH

RNA FISH was carried out using custom Stellaris® RNA FISH probes designed against nascent *Shh* mRNA as previously described (Williamson et al., 2019). As per manufacturer's instructions[3], paraffin sections from E10.5 embryos were deparaffinised in xylene, hydrated in ethanol and permeabilised in 70% ethanol overnight at 4°C. Slides were incubated in 10 μg/mL proteinase K in PBS for 20 min at 37°C followed by washes in PBS and wash buffer (2× SSC, 10% deionised formamide). *Shh* RNA FISH probes were diluted in Stellaris RNA FISH hybridisation buffer (#SMF-HB1-10) to 125 nM and hybridised to slides overnight in a humidified chamber at 37°C. Slides were washed twice for 30 min in wash buffer (2× SSC, 10% deionised formamide) at 37°C, counterstained with 5 ng/mL DAPI, washed in PBS and mounted in Vectashield (Vector Laboratories). This was performed in triplicate for both wild-type control and $Shh^{\Delta SBE2/\Delta SBE2}$ embryos.

RNA FISH Image Analysis

Image capture and analysis of RNA FISH was carried out as previously described (Williamson et al., 2019). Images were taken of ventral hypothalamic region, midbrain and non-expressing diencephalic epithelial wall tissue from three control and $Shh^{\Delta SBE2/\Delta SBE2}$ embryos. Two sections of each embryo were captured per tissue. Images were captured as *z* stacks with step size set at 0.2 μm. Hardware control, image capture and analysis were performed using Nikon Nis-Elements software. Images were deconvolved using a calculated point spread function with the constrained iterative algorithm of Volocity (PerkinElmer). Image analysis was carried out using the Quantitation module of Volocity.

In situ Hybridisation on Cryosections

For *in situ* on cryosections, the protocol was adaped from Lu et al. (2013) as described.

Briefly, *in situ* probes were placed at 37°C for 5 min, added to hybridisation buffer [1× salt (2 M NaCl, 100 mM Tris, 65 mM NaH_2PO_4, 50 mM Na_2HPO_4, and 50 mM EDTA in ddH_2O, pH to 7.5], 50% deionised formamide, 10% dextran sulfate, 1 mg/mL yeast RNA, 1× Denhardt's (2% BSA, 2% Ficcoll, and 2% PVP made up to 100 mL with ddH_2O) pre-warmed to 65°C and subsequently denatured at 95°C for 2 min. Probes were added to slides, that had been previously allowed to reach room temperature, and placed in a 65°C waterbath overnight in a chamber humidified with 5× SSC/50% formamide. Slides were subject to 3× 30 min washes at 65°C in washing solution (50% formamide, 25 mL 1× SSC and 0.1% Tween 20 made up to 500 mL Milli-Q H_2O). Subsequently, slides were washed 2× 30 min in 1× MABT (100 mM maleic acid, 167 mM NaCl in ddH_2O, pH to 7.5, then 0.1% Tween 20). Slides were blocked in 1× MABT + 2% blocking reagent + 20% heat inactivated sheep serum. A 1/2,500 dilution of anti-DIG antibody (Roche, 1109327490) in 1× MABT + 20% heat inactivated sheep serum was then added and slides were placed in a humified chamber and incubated overnight at room temperature. Slides were washed 5× for 20 min in 1× MABT. They were then washed with NTMT (100 mM NaCl, 50 mM MgCl, 100 mM Tris pH 9.5, and 0.1% Tween 20 made up in ddH_2O) for 10 min followed by NTMTL (NTMT + 5 mM Levamisole) for 10 min. Staining solution, NTMTL containing 3.75 μL/mL BCIP (Roche, 11383221001) and 5 μL/mL NBT (Roche, 11383213001), was then added and slides were placed at 37°C to allow colour to develop. Slides were immersed in H_2O, air dried and mounted in Prolong gold (Thermo Fisher Scientific, P39634).

Immunofluorescence

Cryosections were removed from the −80°C freezer and allowed to reach room temperature before being subjected to antigen retrieval using a citric acid solution (10 mM citric acid dissolved in ddH_2O, then pH to 6–6.5) at 65°C for 15 min. Subsequently, slides were washed twice for 5 min in PBS then blocked using 10% goat serum in PBT (0.1% Triton-X100 in PBS) for 1 h at room temperature. For PITX2 staining, donkey serum was used. Slides were incubated in primary antibodies diluted in blocking solution overnight at 4°C as follows: rabbit anti-CASP3 (1:500, Abcam, ab49822); mouse anti-COL2A1 (1:400, Santa Cruz, sc-52659); rabbit anti-ISL1 (1:200, Abcam, ab20670); rabbit anti-LHX3 (1:250, Abcam, ab14555); rabbit anti-NKX2-1 (1:250, Abcam, ab76013); rabbit anti-NKX2-2 (1:250, Abcam, ab191077); rabbit anti-NKX6-1 (1:250, Abcam, ab221549); rabbit anti-OLIG2 (1:200, Abcam, ab109186); rabbit anti-oxytocin (1:500, Abcam, ab212193); rabbit anti-PAX6 (1:350, Abcam, ab195045); sheep anti-PITX2 (1:200, R&D, AF7388); rabbit anti-SF1 (1:200, Abcam, ab65815); rabbit anti-TBX3 (1:200, Abcam, ab99302); rabbit anti-TH1 (1:200, Abcam, ab137869); rabbit anti-TCF4 (1:100, Cell Signalling; C48H11); rabbit anti-vasopressin (1:500, Abcam, ab213707). Slides were then washed 3 times for 10 min each in PBT and incubated for 1 h at room temperature with fluorescent secondary antibodies in PBT in the dark, secondaries as follows: goat anti-rabbit 488 (1:500, Thermo Fisher Scientific, A11034); goat anti-mouse 488 (1:500, Thermo Fisher Scientific, A11029); donkey anti-sheep 594 (1:500, Thermo Fisher Scientific, A11016). Slides were then washed 3 times for 10 min each in PBT and subsequently stained with

[3]www.biosearchtech.com/stellarisprotocols

0.1 μg/mL DAPI in PBS for 10 min, then washed in PBS twice for 5 min each. Lastly, slides were mounted using Prolong Gold (Thermo Fisher Scientific, P39634) and allowed to dry prior to imaging.

RESULTS

Loss of SBE2 Enhancer Activity Leads to Reduction in Rostral Diencephalic *Shh* Expression

To examine the role of the SBE2 enhancer during development, this element was deleted using CRISPR/Cas9 (Cong et al., 2013) to remove 1.2 kb (**Supplementary Figures S1A,B**) containing the conserved element. The SBE2 deleted allele, henceforth referred to as $Shh^{\Delta SBE2}$, was crossed to generate homozygous embryos ($Shh^{\Delta SBE2/\Delta SBE2}$) and also crossed to *Shh* null (Shh^{null}) mice to generate compound heterozygous embryos ($Shh^{null/\Delta SBE2}$). Loss of the SBE2 enhancer alone had no effect upon survival as the $Shh^{\Delta SBE2/\Delta SBE2}$ mice were viable and fertile, appeared phenotypically normal and at weaning no reduction from the expected number of mutant mice was found. Conversely, no live $Shh^{null/\Delta SBE2}$ offspring were recorded by postnatal (P) 14 days (**Figure 1B**), these compound heterozygous embryos died perinatally failing to survive beyond P4, with the cause of postnatal lethality currently undetermined. Normal ratios were observed at embryonic (E) 17.5 (**Figure 1A**).

The rostral diencephalon, extending from the ZLI to the MGE, is the target for SBE2 expression as predicted by SBE2/reporter gene analysis in transgenic embryos (Jeong et al., 2006). *Shh* mRNA expression analysis at E10.5 in $Shh^{\Delta SBE2/\Delta SBE2}$ and $Shh^{null/\Delta SBE2}$ embryos revealed levels of *Shh* expression undetected by wholemount *in situ* hybridisation in the ventral portion of the rostral diencephalon in both crosses (**Figures 1C–E** and **Supplementary Figure S2J**). This region of the diencephalon will give rise to the ventral portion of the tuberal and mammillary hypothalamus (Xie and Dorsky, 2017), and for simplicity will henceforth be referred to as the ventral hypothalamus (VH). The region lacking *Shh* expression in the SBE2 mutant embryos, however, was appreciably more restricted than the region predicted by transgenic analysis, this is likely due to overlapping expression caudally directed by other brain enhancers such as SBE3 and SBE4. As compared to wildtype, the $Shh^{null/+}$ embryos showed the same relative pattern of expression, but at a reduced level (Lettice et al., 2014).

The differences in postnatal viability between $Shh^{\Delta SBE2/\Delta SBE2}$ and $Shh^{null/\Delta SBE2}$ suggested that there was residual *Shh* expression in the viable line. Since this residual expression was undetected by *in situ* hybridisation, we used RNA FISH to directly assay transcriptional changes due to the deletion of the SBE2 enhancer. Enhancers regulate the periodicity of transcription affecting the rate of transcriptional bursting (Larsson et al., 2019) and hence, this approach, which assays the production of *Shh* nascent transcripts, provides a relative assessment of enhancer activity. We have previously used RNA FISH to focus on transcriptional changes (Williamson et al., 2019) in small expression domains within embryonic tissues; here, focusing on expressing cells (**Supplementary Figures S2H,I**) within the VH. In the control VH, about 52% of the cells (**Figure 1I**) exhibited at least one active allele while in the $Shh^{\Delta SBE2/\Delta SBE2}$ mutant, that number was approximately 23% (**Figure 1I**). The majority of actively transcribing cells (∼90%, **Supplementary Figure S2G**) had only a single active allele (**Figure 1J**). Expression of *Shh* is not under SBE2 control in the midbrain and transcriptional differences measured in this region were not significant (**Figures 1I,J**), while non-expressing neighbouring domains showed no appreciable allelic signal (<0.3%). SBE2, therefore, accounted for approximately 55–60% of the transcriptional activity in the VH.

Further reduction in activity in the $Shh^{null/\Delta SBE2}$ embryo was indicated by patched-1 (*Ptch-1*) expression. *Ptch1* expression, which is upregulated by hedgehog signalling and is a sensitive readout of *Shh* expression, was examined at E10.5 (Casillas and Roelink, 2018). Similar to the expression pattern observed for *Shh*, *Ptch1* was also downregulated in the VH upon loss of SBE2 activity (**Figures 1F–H**). However, in this instance, levels differed between the compound heterozygous and the homozygous embryos. While $Shh^{null/\Delta SBE2}$ embryos showed no detectable levels of *Ptch1* expression in the VH (**Figure 1H**, red arrow), the $Shh^{\Delta SBE2/\Delta SBE2}$ embryos presented residual levels (**Figure 1G**, green arrow). It should be noted that while at E9.5 *Shh* levels were low in the ventral diencephalon in the absence of SBE2 for both mutant conditions (**Supplementary Figures S2A–C**), that *Ptch1* expression pattern in neighbouring tissue was comparable to wildtype (**Supplementary Figures S2D–F**). We suggest that *Ptch1* dependent development is unaffected until after E9.5 and the defects occur after this stage. To determine when VH *Shh* expression is downregulated, a mouse line containing the *lacZ* reporter gene inserted into the *Shh* regulatory domain, SBLac526, was used (Anderson et al., 2014) as proxy for *Shh* expression. VH *lacZ* expression was low but detectable up to E13.5 (**Supplementary Figures S3A–C**, red arrow), with expression subsequently absent at E14.5 (**Supplementary Figure S3D**, black arrow). Hence, the effective role of SBE2 encompasses a developmental window that begins sometime after E9.5 extending to E13.5 when all regulatory activity in the VH is decreasing.

The residual levels of *Shh* expression present in the SBE2 mutant appeared capable of rescuing the postnatal lethality. Accordingly, an additional enhancer was identified in *Shh* expressing neural cells from E10.5 dissected heads. ATAC-seq (Buenrostro et al., 2015) revealed a peak corresponding to SBE2, as expected, and additionally a peak at a vertebrate conserved sequence element (position chr5:29230287-29230750 [mm9]) (**Supplementary Figure S4A**), no peaks were detected for SBE3 or SBE4, likely due to the low contribution of the telencephalic cells vs. other brain regions. When used in reporter constructs in transgenic embryos at E10.5 (**Supplementary Figures S4B–D**) the conserved sequence element showed broad expression activity extending anteriorly from the midbrain into the diencephalon (**Supplementary Figures S4B–D**). Recently, an enhancer whose primary function is to drive expression initially in the prechordal plate, and later throughout the ventral forebrain, called SBE7 (Sagai et al., 2019), was reported at the same chromosomal position. Thus, SBE7 has a primary

A Embryonic day 17.5(E17.5)

	Total	Control	$Shh^{\Delta SBE2/+}$	$Shh^{null/+}$	$Shh^{\Delta SBE2/null}$
Observed	41	10	11	8	12
Expected	-	10	10	10	10

p=0.7985

B Post-natal day 14 (P14)

	Total	Control	$Shh^{\Delta SBE2/+}$	$Shh^{null/+}$	$Shh^{\Delta SBE2/null}$
Observed	24	4	8	12	0
Expected	-	6	6	6	6

p=0.0025

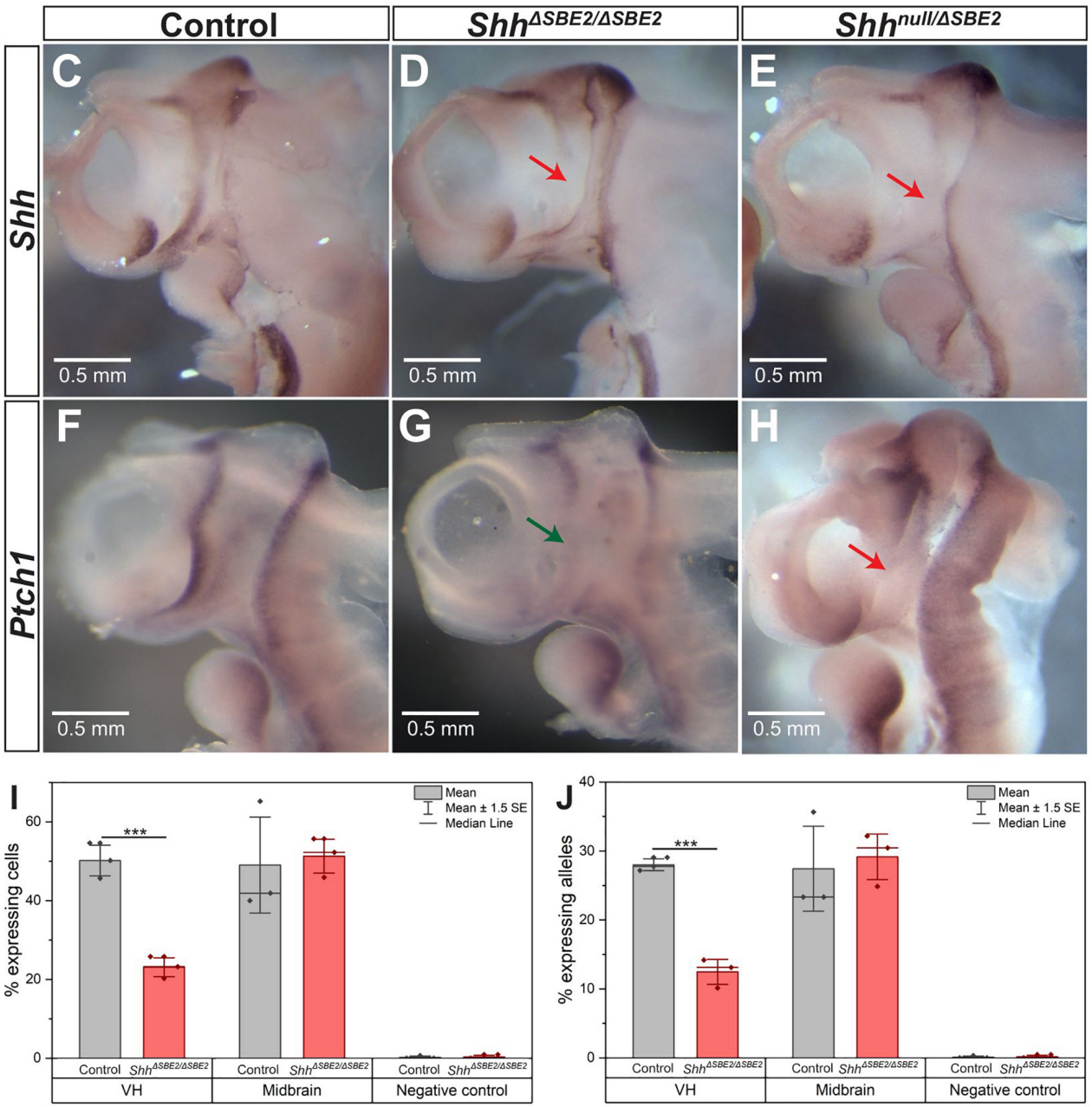


FIGURE 1 | Loss of SBE2 leads to reduced *Shh* activity in the rostral diencephalon. **(A,B)** Peri and post-natal survival for $Shh^{null/\Delta SBE2}$ compound heterozygotes is shown. Statistical significance was assessed using a chi-square test. **(C–E)** *In situ* hybridisation for *Shh* in E10.5 day heads bisected along the midline is shown for control, $Shh^{\Delta SBE2/\Delta SBE2}$ homozygotes and $Shh^{null/\Delta SBE2}$ compound heterozygotes. The red arrows point to the rostral diencephalic region where *Shh* expression is absent. **(F–H)** *In situ* hybridisation for *Ptch1* in E10.5 day bisected heads is shown. Wildtype levels of *Ptch1* expression in rostral diencephalon in control embryos, reduced levels in $Shh^{\Delta SBE2/\Delta SBE2}$ (indicated by green arrow in G) and undetectable levels in $Shh^{null/\Delta SBE2}$ (indicated by red arrow in H) is shown. Scale bars **(C–H)** appear in the bottom left hand corner. **(I)** Percentage of cells expressing at least one allele of *Shh* in the VH, midbrain or negative control tissue at E10.5 detected by RNA FISH. **(J)** Percentage of Shh alleles expressed in the tissue of the VH, midbrain or negative control at E10.5 detected by RNA FISH. **(I,J)** Statistical significance was assessed using a Student's *T*-test. Error bars represent median ± s.e.m. $p^{***} < 0.001$ ($n = 3$).

function in the prechordal plate but, either singly or with other unidentified elements, accounts for approximately 40–45% of the transcriptional activity in the VH. Compensation by *de novo* activation of other *Hh* genes, however, cannot be ruled out.

Loss of SBE2 Activity Effects Midline Craniofacial Malformations

SBE2 deletion embryos were analysed for craniofacial malformations. Unsurprisingly, given the postnatal viability of $Shh^{\Delta SBE2/\Delta SBE2}$ embryos, no gross craniofacial malformations nor that of the craniofacial skeletal elements (**Supplementary Figures S5A,B**) were observed. In contrast, at E17.5, $Shh^{null/\Delta SBE2}$ embryos exhibited external head malfomation, in which the mutant embryos presented with a more rounded head (**Supplementary Figures S6A–D**). To investigate these changes in head shape, microCT scans of control and $Shh^{null/\Delta SBE2}$ mutant embryos were subject to geometric morphometric analysis (**Supplementary Figure S7**) to visualise and test for differences between control and mutant embryos. All four possible genotypes, $Shh^{+/+}$, $Shh^{null/+}$, $Shh^{\Delta SBE2/+}$, and $Shh^{null/\Delta SBE2}$ were analysed. For all analyses subsequently presented $Shh^{+/+}$, $Shh^{null/+}$, and $Shh^{\Delta SBE2/+}$ clustered together and were indistinguishable. As such, and also due to the previously established lack of phenotypic effects in $Shh^{null/+}$ (Chiang et al., 1996), all three genotypes were used as a control for morphometric analyses. Centroid size, a proxy of skull scale, revealed that the compound heterozygous mutant embryos tended to have an overall reduction in skull size, though this was not significant (pairwise two-sample Wilcoxon, $p = 0.49$) (**Figure 2A**). However, shape variation was significantly affected by size, as well as genotype and litter provided the best regression fit ($R^2 = 0.83$, ANOVA $p < 0.001$). Principal components of allometry-free shape was also explored. We observed clear, genotype-specific separation of control and mutant mice, with the latter occupying a relatively broader expanse of morphospace (**Figure 2B**). In mutants, Principal Component 1 (54.7%) (**Figure 2B**) describes a reduction in the anteroposterior axis while the mediolateral axis expanded (**Figures 2C–H**). Procrustes variances between control and mutant mice were not significant, however, this was presumably due to limitations in statistical power. Indeed, shape differences described by PC1 were significant (Welch Two Sample *t*-test, $p = 1.5 \times 10^{-4}$). OPT analysis revealed that, in addition to presenting a perturbed overall head structure, $Shh^{null/\Delta SBE2}$ embryos displayed affected nasal cavities (**Figures 2J,L**), which were reduced in length. Moreover, a region of brain tissue, specifically the midline hypothalamic tissue, was absent in the mutant embryos (**Figures 2I,K**).

$Shh^{null/\Delta SBE2}$ embryos displayed midline craniofacial bone malformations with a high degree of variability, which, based on phenotypic severity, were classified into three categories: mild, moderate and severe (**Figures 3A–D**). All mutant embryos showed a reduction in the basisphenoid bone, ranging from the presence of only a small remnant (**Figure 3B**, pale green arrow) to complete loss (**Figures 3C,D**) and in the vomer bone, ranging from an excessive fusion (**Figure 3B**, yellow arrow) to complete absence (**Figures 3C,D**). In addition, the moderate and severe embryos displayed defects in the pterygoid and maxillary bones (**Figures 3C,D**, white and black arrows). Notably, all embryos showed some degree of midline craniofacial malformation, which indicates that *Shh* signalling directed by SBE2 is involved in the midline development of the craniofacial bones, and parallels the role *Shh* plays in early midline brain development (Chiang et al., 1996; Blaess et al., 2014).

Mislocalised Pituitary and Loss of SHH signalling Disrupts Craniofacial Development

Closer examination of the heads of mutant compound heterozygous embryos revealed that the basisphenoid bone, the most posterior defective cranial bone, was intercepted by the adenohypophyseal pituitary lobe (labelled AP in **Figures 3E–G**). This lobe of the pituitary was misplaced and formed in a more anterior position, growing within the region where the basisphenoid will form (the basis for the mislocalisation is discussed below). Developmental timing of this event suggests that the adenohypophysis disrupts this region before chondrogenesis initiates, which indicates that the adenohypophyseal tissue likely presents a physical barrier to basisphenoid development. The phenotypic variability may correlate with the extent to which the pituitary lobe intercepted the basisphenoid bone in the mutant embryos.

Further investigations into the effect of SBE2 directed expression on craniofacial development focused on mesenchymal *Ptch1* expression in the vicinity of the forebrain. In control embryos at E12.5, *Ptch1* expression was detected within a narrow layer of mesenchyme located directly adjacent to the floor of the diencephalon where *Shh* was expressed (**Figures 4A,B**). COL2A1, a type 2 collagen, is expressed during early cartilage formation (Hissnauer et al., 2010) and co-localised with *Ptch1* in control embryos (**Figures 4E,F**), indicating that the underlying facial mesenchyme is under the influence of SHH and is committed toward cartilage formation. At the same stage in the $Shh^{null/\Delta SBE2}$ embryos, *Ptch1* expression was reduced in the underlying mesenchyme (**Figures 4C,D**) and only a thin lining of cells expressing COL2A1 was seen (**Figures 4G,H**). The loss of *Ptch1* and COL2A1 staining was found to be independent of the physical disruptive effects caused by mislocalisation of the adenohypophysis, as the absence of both markers was evident in regions of the cranial base which lie anterior to disrupted adenohypophyseal tissue detected in the mutant embryos (**Figures 4C,G**). Thus, SBE2 mediated *Shh* expression contributed to an adequate cell response and bone formation in the sub–diencephalic mesenchyme along the midline of the palatal floor, and reduction of this cell population is critical for the midline craniofacial phenotype.

$Shh^{null/\Delta SBE2}$ Embryos Display Mislocalised and Deformed Pituitary Lobes Due to Disruption of the AP Hypothalamic Boundary

Pituitary lobe development was analysed at stages E11.5, E13.5, E15.5, and E17.5. In $Shh^{null/\Delta SBE2}$ mutant embryos

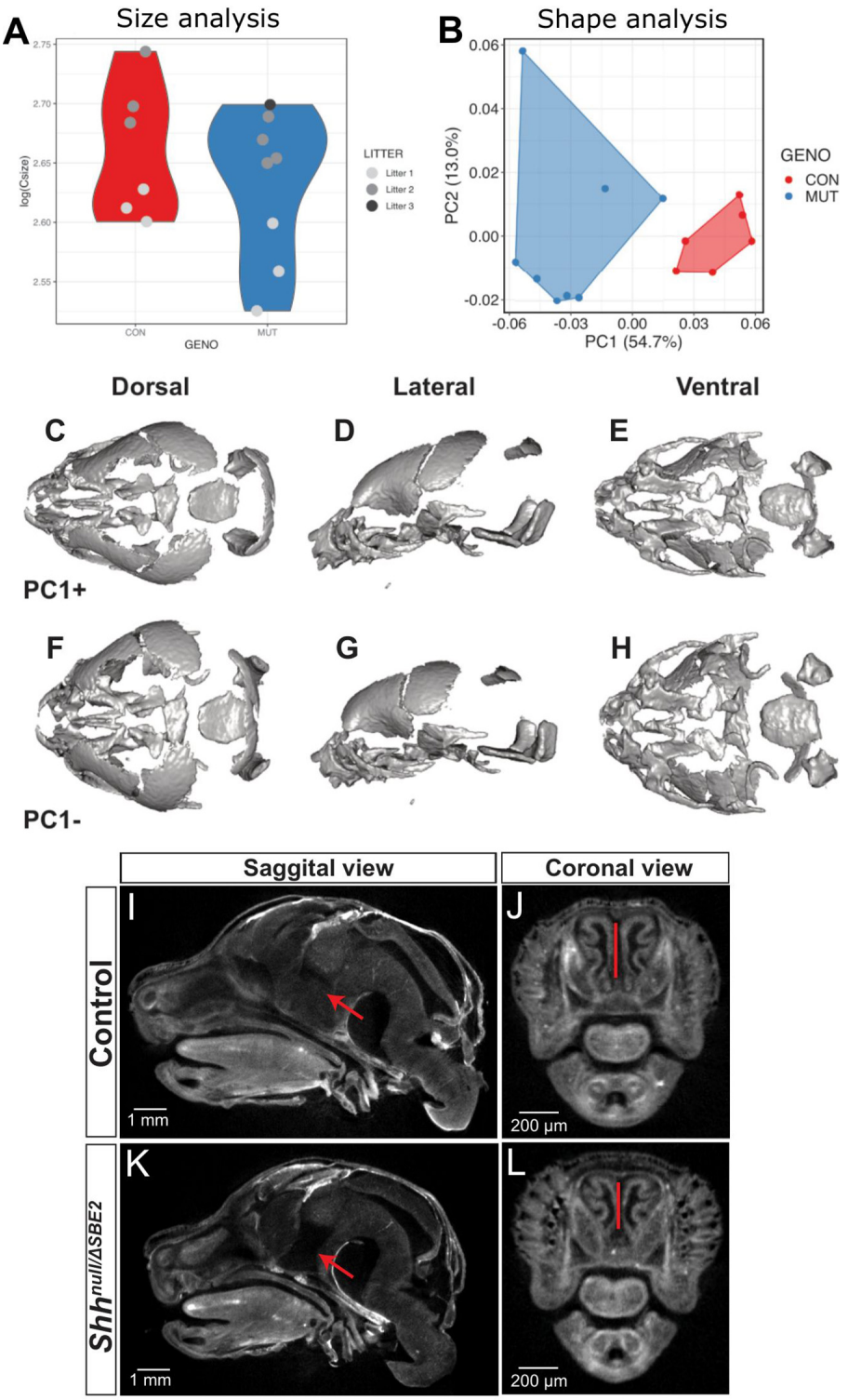


FIGURE 2 | *Shh*$^{null\,/\,\Delta SBE2}$ embryos have an altered head shape and gross malformations. **(A)** Skull size, as described by the log(centroid size) landmark configurations of control and *Shh*$^{null\,/\,\Delta SBE2}$ at E17.5. **(B)** Principal components (PC) plot of allometry-free skull shape at E17.5. **(C–H)** Dorsal, lateral and ventral views of a mean surface warped according to extreme positive PC1 **(C–E)** or negative PC1 **(F–H)** (control $n = 6$; mutant $n = 8$). **(I–L)** OPT scan stills from E17.5 day control and *Shh*$^{null\,/\,\Delta SBE2}$ heads, presenting the midline sagittal plane and the coronal nasal region. Red arrows point to the region of hypothalamic tissue which is absent in *Shh*$^{null\,/\,\Delta SBE2}$ embryos. Red lines serve to demonstrate the reduced nasal cavities found in *Shh*$^{null\,/\,\Delta SBE2}$ embryos. Scale bars are depicted in the bottom left hand corner (control $n = 2$; mutant $n = 3$).

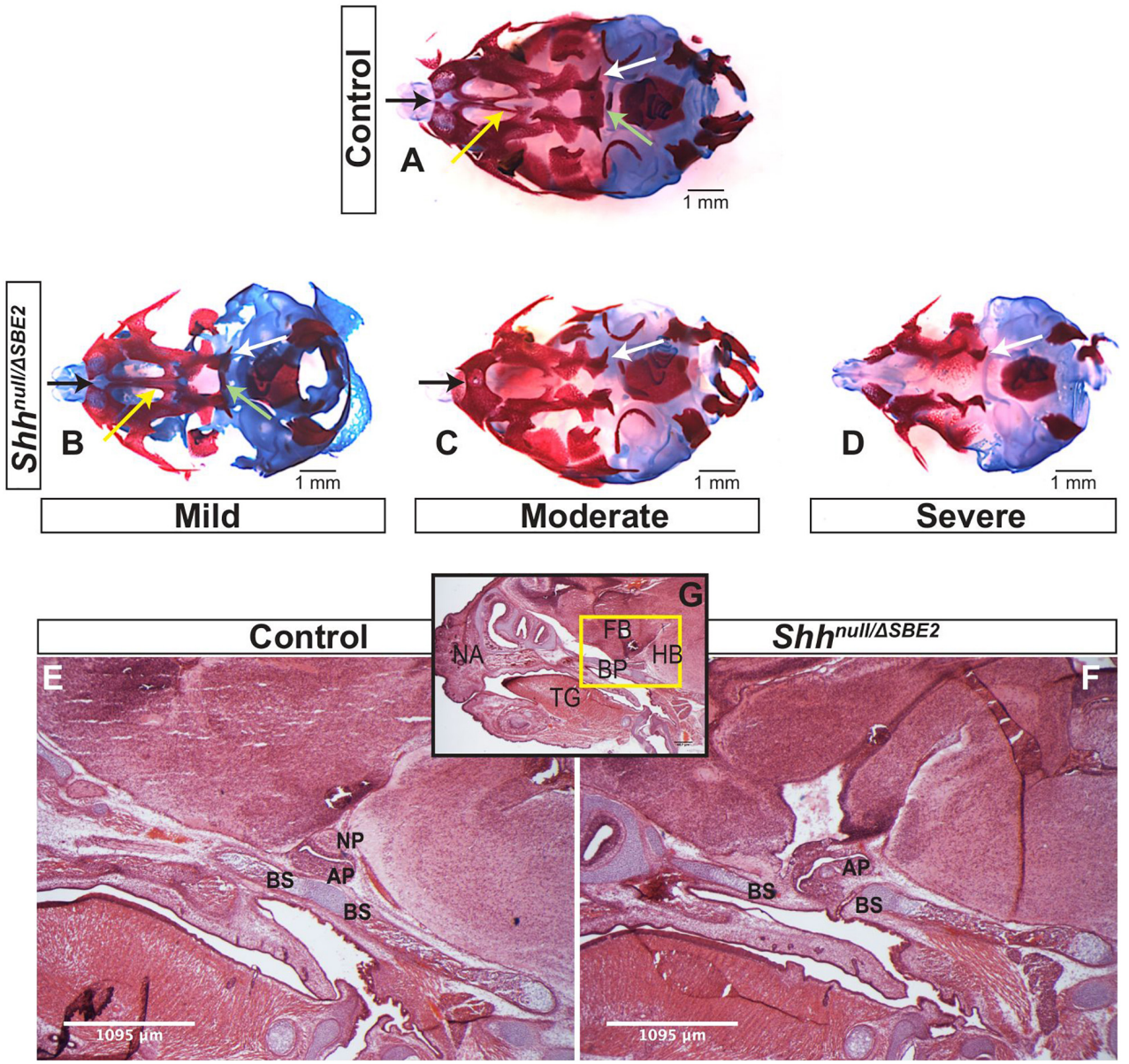


FIGURE 3 | *Shh*$^{null/\Delta SBE2}$ embryos present malformations in the bones of the cranial vault. **(A–D)** Dual staining highlighting chondrogenic (blue) and skeletal (red) craniofacial elements in control and *Shh*$^{null/\Delta SBE2}$ E17.5 heads are shown. Pale green arrows indicate the basisphenoid bone, yellow arrows indicate the vomer bone, black arrows the pre-maxillary bone and the white arrows indicate the pterygoid bone. **(E,F)** H&E staining of E17.5 day parasagittal cryosections from control and *Shh*$^{null/\Delta SBE2}$ heads. A zoomed out view of an E17.5 day head parasagittal cryosection is depicted in **(G)**. The yellow box presents the region of the head depicted in **(E,F)**. AP- adenohypophysis; NP- neurohypophysis; BS- basisphenoid bone; BP- base plate of the skull; NA- nasal prominence; FB- forebrain; HB- hindbrain; TG- tongue. Scale bars are shown in the bottom right or left hand corner (n = 3).

at E11.5, Rathke's pouch (RP), which will give rise to the adenohypophyseal lobe, and the infundibulum, which will give rise to the neurohypophyseal pituitary lobe, were both ectopically shifted ventrally (**Supplementary Figures S8A,B**). These findings reflect the previous observations by Zhao et al. (2012) using conditional *Shh* deletion mice. At E13.5, RP was severely malformed in the *Shh*$^{null/\Delta SBE2}$ mutant embryos, while the infundibulum did not separate from the neuroectodermal tissue to the same degree as control counterparts (**Figures 5A–C**) and remained ectopically located at E15.5 (**Supplementary Figures S8C,D**). Interestingly, at E17.5 the neurophypophysis appeared absent in the *Shh*$^{null/\Delta SBE2}$ mutant embryos (**Figures 3E,F**), yet staining for markers of neurohypophyseal tissue revealed that this was due to the lack of separation of the neurohypophyseal tissue from the neuroectodermal tissue of origin (**Figures 5L,N**). Moreover, caspase-3 (CASP3) staining revealed that this remnant neurohypophyseal tissue appeared to be undergoing apoptosis (**Figures 5M,O**) indicating that

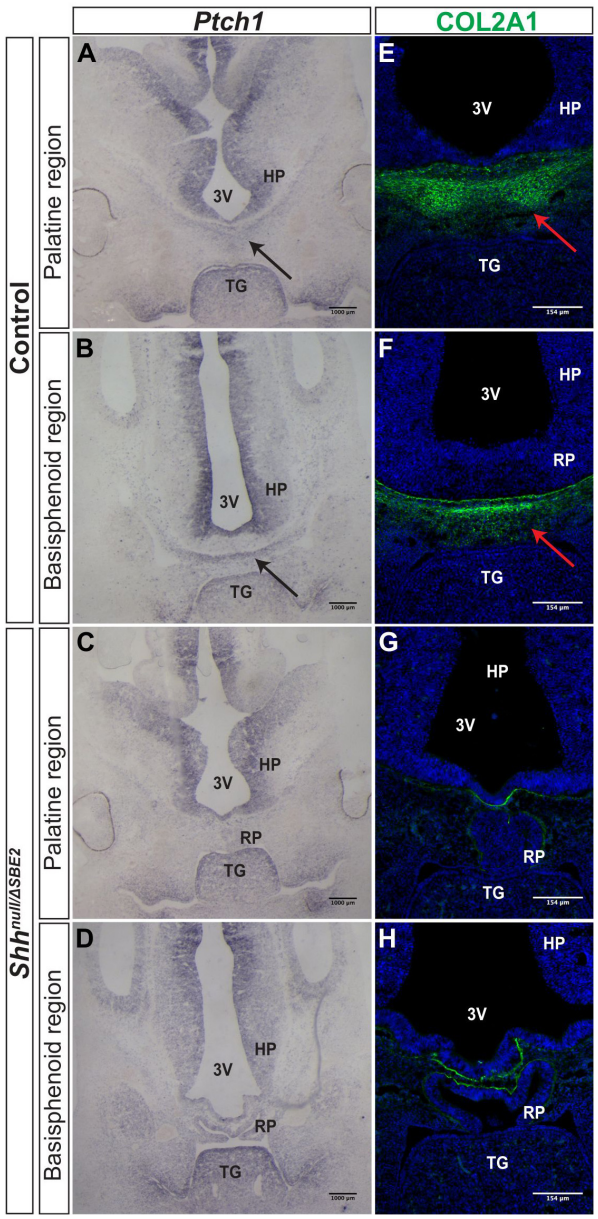


FIGURE 4 | Bone formation in the viscerocranium of $Shh^{null/\Delta SBE2}$ embryos. **(A–D)** *In situ* hybridisation for *Ptch1* in E12.5 day coronal cryosections of control and $Shh^{null/\Delta SBE2}$ embryos. **(A,C)** present the palatine region and **(B,D)** present the basisphenoid region. Black arrows point to the region of *Ptch1* expression seen below the hypothalamic tissue which is absent in $Shh^{null/\Delta SBE2}$ embryos. **(E–H)** Immunofluorescent staining for COL2A1 in E12.5 day coronal cryosections of control and $Shh^{null/\Delta SBE2}$ embryos. **(E,G)** present the palatine region and **(F,H)** present the basisphenoid region. Red arrows point to the region of COL2A1 expression seen below the hypothalamic tissue which mostly is absent in $Shh^{null/\Delta SBE2}$ embryos. 3V- third ventricle; HP- hypothalamus; TG- tongue; RP- Rathke's pouch (n = 3).

inadequate neurohypophyseal partitioning likely led to tissue degeneration. Despite the fact that both hypophyseal lobes were misspecified in the $Shh^{null/\Delta SBE2}$ embryos, early hypophyseal patterning was unperturbed (**Supplementary Figures S9C,F,I,L**).

To exclude the possibility that the observed pituitary malformation and mislocalisation were due to a developmental delay, expression of adenohypophyseal markers that display stage-specific restricted expression were analysed. ISL1 is expressed throughout the early developing adenohypophysis but later becomes restricted to the ventral most portion of the lobe (Ericson et al., 1998; Castinetti et al., 2015). This same restriction of expression is seen for TBX3, which is expressed broadly throughout the developing adenohypophysis, but by E12.5 expression is restricted to the ventral aspect of the lobe (Pontecorvi et al., 2008). For both ISL1 (**Supplementary Figures S10A,B**) and TBX3 (**Supplementary Figures S10C,D**) expression in the adenohypophysis at E12.5 in $Shh^{null/\Delta SBE2}$ embryos was found to reflect the pattern seen in control embryos, with expression only detected in the ventral most portion, indicating that the adenohypophyseal lobe was not developmentally delayed.

While patterning and early neuronal marker expression within the pituitary lobes appeared normal, at later developmental stages, E17.5, neuronal populations within varied nuclei of the VH, which establish connections with the pituitary lobes, were disrupted. Both the tyrosine hydroxylase (TH1) (**Supplementary Figures S11A,B**) and the vasopressin neuronal population (**Supplementary Figures S11E,F**) of the paraventricular nucleus (PVN) were lost in $Shh^{null/\Delta SBE2}$ embryos. Additionally, the oxytocin neuronal population of the PVN was absent while the population found within the arcuate nucleus (ARC) was either misspecified or mislocalised (**Supplementary Figures S11C,D**). To confirm that the loss of PVN and ARC neuronal populations did not result from a developmental delay in hypothalamic neuronal specification, expression of oxytocin and TH1 was analysed at E15.5. At E15.5 oxytocin expression was present in the PVN of the control embryos (**Supplementary Figure S12A**, white arrow) but absent from $Shh^{null/\Delta SBE2}$ embryos (**Supplementary Figure S12B**). Similarly, expression of TH1 was detected in the ARC of control embryos (**Supplementary Figure S12C**, yellow arrow), yet no expression was detected in $Shh^{null/\Delta SBE2}$ embryos (**Supplementary Figure S12D**). This indicates that the lack of specific gene expression in hypothalamic neuronal populations at E15.5 persists for at least two additional days up to E17.5, arguing for a loss of the neuronal population and not a developmental delay in neuronal specification. These results indicate that, in addition to the disruption of hypophyseal structure and location in development, there is also disruption of hypothalamic neuronal populations, demonstrating that SBE2 activity contributes to the role of *Shh* in hypophyseal and hypothalamic neuronal development. Interestingly, the hypothalamic populations found to be affected are known to establish connections with and regulate the hypophysis, indicating that *Shh* plays a role in coordinating hypophyseal development and subsequent neuronal regulation. The disruption of these late neuronal populations in $Shh^{null/\Delta SBE2}$ embryos revealed that early loss of SBE2 directed *Shh* activity in the VH has late effects upon development in the hypothalamus, at timepoints at which *Shh* expression is no longer present in the tissue in question (**Supplementary Figure S3**).

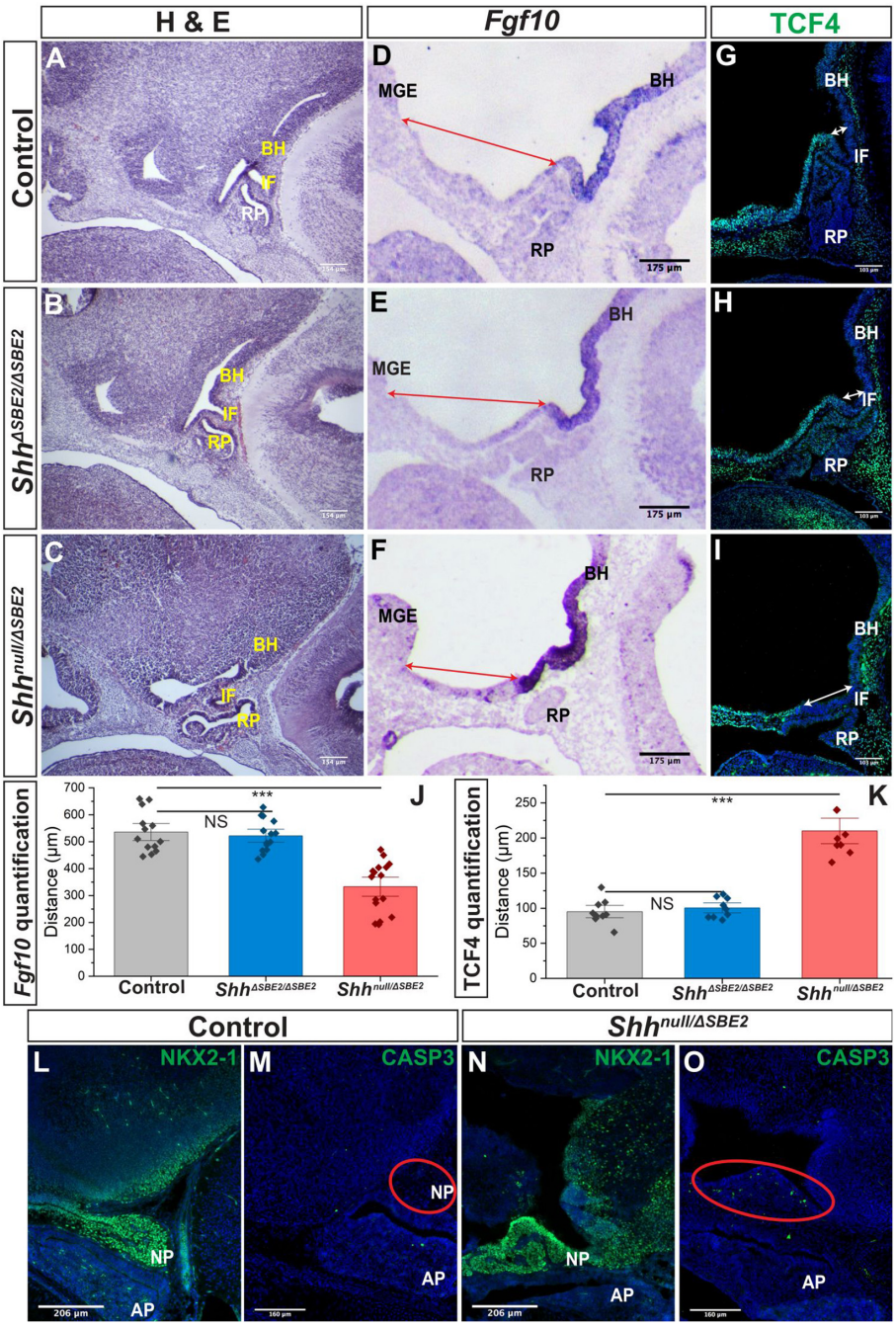


FIGURE 5 | The AP hypothalamic boundary is shifted ventrally in $Shh^{null/\Delta SBE2}$ embryos. **(A–C)** H&E staining of E13.5 day sagittal cryosections of control, $Shh^{\Delta SBE2/\Delta SBE2}$ and $Shh^{null/\Delta SBE2}$ embryos. **(D–F)** *In situ* hybridisation for *Fgf10* in E11.5 day sagittal cryosections of control, $Shh^{\Delta SBE2/\Delta SBE2}$ and $Shh^{null/\Delta SBE2}$ embryos. Red arrows represent the distance between the MGE and the anterior boundary of *Fgf10* expression. **(G–I)** Immunofluorescent staining for TCF4 in E11.5 day sagittal cryosections of control, $Shh^{\Delta SBE2/\Delta SBE2}$ and $Shh^{null/\Delta SBE2}$ embryos. White arrows represent the distance between the posterior infundibulum and the posterior boundary of TCF4 expression. **(J)** Image J measurement based quantification of the distance (in µm) depicted by the red arrows in **(D–F)** in control, $Shh^{\Delta SBE2/\Delta SBE2}$ and $Shh^{null/\Delta SBE2}$ embryos. Statistical significance was assessed using the Student's *T*-test. Error bars represent median ± s.e.m. $^{***}p < 0.001$. **(K)** Image J measurement based quantification of the distance (in µm) between depicted by the white arrows in **(G–I)** in control, $Shh^{\Delta SBE2/\Delta SBE2}$ and $Shh^{null/\Delta SBE2}$ embryos. Statistical significance was assessed using the Student's *T*-test. Error bars represent median ± S.E.M. $^{***}p < 0.001$. **(L,N)** Immunofluorescent staining for NKX2-1 in E17.5 day sagittal cryosections of control and $Shh^{null/\Delta SBE2}$ embryos. **(M,O)** Immunofluorescent staining for CASP3 in E17.5 day sagittal cryosections of control and $Shh^{null/\Delta SBE2}$ embryos. Red circles demarcate the region of cell death apparent in mutant embryos. NS, not significant; BH, basal hypothalamus; IF, infundibulum; RP, Rathke's pouch; MGE, medial ganglionic eminence. Scale bars are shown in the bottom left hand corner ($n = 3$).

In $Shh^{\Delta SBE2/\Delta SBE2}$ mutant embryos, neither the adenohypophysis nor the neurohypophysis were ectopically located and both lobes had separated from the tissue of origin at E13.5 (**Supplementary Figure S5**). Additionally, no defects were observed in early marker expression (**Supplementary Figures S9B,E,H,K**); however, mild malformations were detected predominantly in the structure of the lateral portion of adenohypophyseal lobe in $Shh^{\Delta SBE2/\Delta SBE2}$ mutant embryos (**Supplementary Figures S5C–K**).

Distinct anterior-posterior expression domains are established along the midline of the hypothalamus by E9.5. The posterior hypothalamic domain expresses factors including *Bmp4* and *Fgf10* whilst the anterior domain expresses *Six6*, *Tcf4,* and *Shh*. Mutual inhibition between these two gene expression domains exists, whereby in the absence of one domain the opposing domain will expand to occupy the hypothalamic space (Ericson et al., 1998; Takuma et al., 1998). The infundibulum expresses the posterior markers *Bmp4* and *Fgf10* and is specified dorsally to the anterior hypothalamic tissue, at the AP transition zone (Ferrand, 1972; Kawamura and Kikuyama, 1995; Takuma et al., 1998; Treier et al., 1998). Inductive cues from the ventral hypothalamic tissue play a role in orchestrating the formation of RP (Schwind, 1928; Takuma et al., 1998). As RP and the infundibulum were both shifted ventrally in the $Shh^{null/\Delta SBE2}$ mutant embryos (**Figure 5C**), verification of any other associated defects in the establishment of the AP hypothalamic boundary were sought. The posterior boundary of expression of TCF4 was shifted ventrally in the mutant embryos (**Figures 5G–I**) whereas the anterior boundary of expression of *Fgf10* was shifted ventrally, expanding expression (**Figures 5D–F**). Using ImageJ based measurements, the distance (in μm) between the boundaries of expression and known landmarks was assessed and the shift observed for both the anterior and posterior markers was found to be significant in the $Shh^{null/\Delta SBE2}$ mutant embryos (**Figures 5J,K**). These results indicated that there is a shift in the AP domain boundary with a reduction in anterior markers and gain of posterior markers. In the $Shh^{\Delta SBE2/\Delta SBE2}$ mutant embryos no shift in expression of TCF4 (**Figure 5H**) or *Fgf10* (**Figure 5E**) was observed, correlating with correct positioning of the pituitary lobes (**Figure 5C**). These results suggest that shift of the pituitary lobes occurs as a result of the aberrant specification of the AP boundary, which does not appear to be sensitive to reduced levels of SHH found in $Shh^{\Delta SBE2/\Delta SBE2}$ embryos (**Supplementary Figures S5C–F**).

Loss of SBE2 Enhancer Activity Leads to Defects in Hypothalamic Patterning in a Concentration Dependent Manner

Much of our knowledge about DV hypothalamic patterning is informed by what is known regarding DV neural tube patterning. In the developing neural tube, neural progenitors acquire distinct transcriptional identities based on length of exposure to patterning signals and DV positioning (Dessaud et al., 2007, 2010; Balaskas et al., 2012). Indeed it is the opposing action of the dorsal neural tube signals *Wnts* and BMPs and the ventral neural tube signal *Shh* that drive patterning along this axis (Burbridge et al., 2016). Subsequently, expression of distinct combinations of transcription factors drives the differentiation of the neural progenitors toward specific neural fates (Dessaud et al., 2008). In the developing hypothalamus *Shh* is initially expressed from the prechordal mesoderm; it is subsequently expressed along the floor plate and plays a role in inducing ventral hypothalamic fate (Burbridge et al., 2016; Corman et al., 2018). In contrast, BMPs are expressed from the roof of the telencephalon, regulating alar hypothalamic development (Burbridge et al., 2016). Given that *Shh* signalling from the hypothalamic floor plate was reduced in the $Shh^{\Delta SBE2/\Delta SBE2}$ and $Shh^{null/\Delta SBE2}$ mutant embryos, DV hypothalamic patterning was examined. Genes whose expression are responsive to SHH levels in the neural tube were of particular interest. No differences were detected between wild-type, $Shh^{\Delta SBE2/+}$ and $Shh^{null/+}$ embryos (**Supplementary Figure S13**), indicating that heterozygous loss of *Shh* has no effect upon DV patterning marker expression.

Expression of PAX6, known to be expressed within the alar hypothalamus, and the opposing basal hypothalamic marker NKX2-2 (Guillemot, 2007; Croizier et al., 2015) were examined. NKX2-2 expression commences early in hypothalamic development, with expression detected in the basal hypothalamic region which borders the alar zone. Later in development, this expression pattern alters within the VH, whereby expression has forked into two bilateral stripes; one that occupies the transition zone of the alar-basal hypothalamus extending into the alar hypothalamus itself, and a second stripe that resides within the basal hypothalamic territory of the VH (Burbridge et al., 2016). At E13.5, this bilateral stripe expression pattern was observed in control embryos (**Figure 6A**). However, in $Shh^{\Delta SBE2/\Delta SBE2}$ mutants the dorsal VH stripe, within the basal hypothalamus, presented patchy expression (**Figure 6B**). The $Shh^{null/\Delta SBE2}$ mutant embryos were more severely affected, with both the alar and basal VH stripes absent (**Figure 6C**). This data indicate that low levels of *Shh* expression can induce some NKX2-2 expression in $Shh^{\Delta SBE2/\Delta SBE2}$, however, these low levels are not sufficient to either specify or maintain the entirety of the NKX2-2 population.

PAX6 is a neural tube and hypothalamic marker, whose expression pattern is known to be restricted to the alar hypothalamus (Burbridge et al., 2016) and, as expected, E13.5 control embryos displayed PAX6 expression solely within the alar portion of the VH (**Figure 6D**). In $Shh^{\Delta SBE2/\Delta SBE2}$ mutant embryos there was a dorsal expansion of PAX6 expression into the basal hypothalamic region in the embryos (**Figure 6E**). In $Shh^{null/\Delta SBE2}$ embryos, this dorsal expansion of PAX6 expression was more severe, with large patches of PAX6 expression detected in the basal hypothalamus of the mutant embryos (**Figure 6F**). These results indicate that *Shh* signalling from the basal hypothalamus is required to restrict PAX6 expression solely to the alar region, and that low levels of *Shh* activity, those found in basal hypothalamus of $Shh^{\Delta SBE2/\Delta SBE2}$ mutants, are not sufficient to exclude PAX6 activity from this region.

These results for NKX2-2 and PAX6 within the VH, upon loss of SHH signalling activity, mirror what was found by Corman et al. (2018). However, these data demonstrate a novel dose dependent effect of SHH signalling whereby these transcription factors are not only able to respond to the presence, or indeed

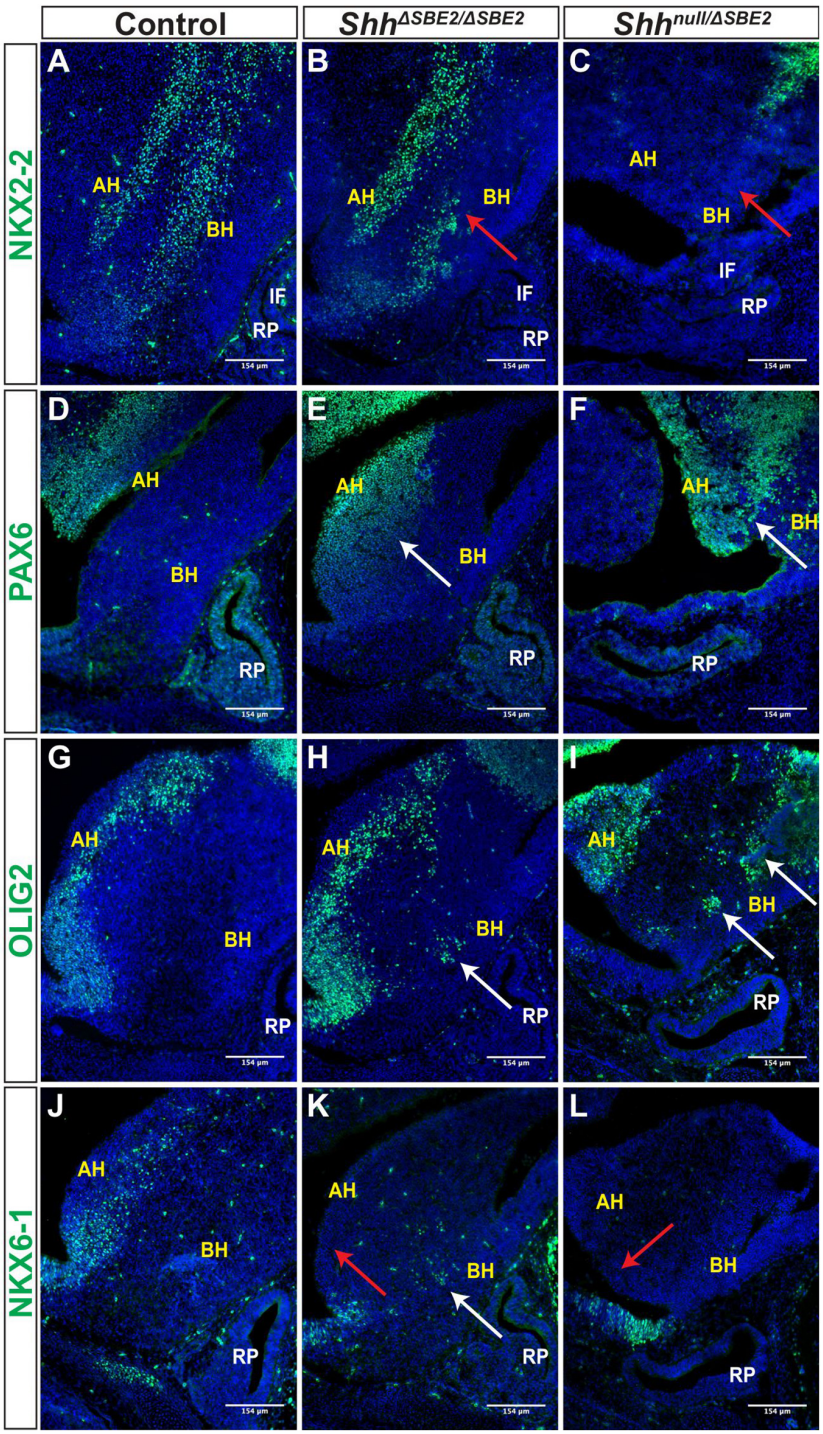


FIGURE 6 | DV hypothalamic patterning is affected by loss of SBE2. **(A–C)** Immunofluorescent staining for NKX2-2 in E13.5 day ventral hypothalamic (VH) parasagittal cryosections of control, $Shh^{\Delta SBE2/\Delta SBE2}$ and $Shh^{null/\Delta SBE2}$ embryos. Red arrows point to the region of NKX2-2 expression in the basal hypothalamus which is reduced or absent upon removal of SBE2 enhancer function. **(D–F)** Immunofluorescent staining for PAX6 in E13.5 day VH parasagittal cryosections of control, $Shh^{\Delta SBE2/\Delta SBE2}$ and $Shh^{null/\Delta SBE2}$ embryos. White arrows point to the region of ectopic PAX6 expression in the BH which is seen upon removal of SBE2 function. **(G–I)** Immunofluorescent staining for OLIG2 in E13.5 day VH parasagittal cryosections of control, $Shh^{\Delta SBE2/\Delta SBE2}$ and $Shh^{null/\Delta SBE2}$ embryos. White arrows point to the ectopic regions of expression of OLIG2 in the BH which are seen upon removal of SBE2 function. **(J–L)** Immunofluorescent staining for NKX6-1 in E13.5 day VH parasagittal cryosections of control, $Shh^{\Delta SBE2/\Delta SBE2}$ and $Shh^{null/\Delta SBE2}$ embryos. Red arrows point to the region of NKX6-1 expression in the AH which is absent upon removal of SBE2 function whilst white arrows point to the region of ectopic expression in the BH seen in $Shh^{\Delta SBE2/\Delta SBE2}$ embryos. AH- alar hypothalamus; BH- basal hypothalamus; IF- infundibulum; RP- Rathke's pouch. Scale bars are shown in the bottom right hand corner (n = 3).

absence, of *Shh* signalling, but are also sensitive to the amount of *Shh*, with the levels of *Shh* activity being key for the establishment of hypothalamic signalling domains.

Expression of the transcription factor OLIG2 is induced by SHH (Dessaud et al., 2008), and in control embryos was detected in the alar hypothalamus in a narrow region along the edge of the third ventricle (**Figure 6G**). In $Shh^{\Delta SBE2/\Delta SBE2}$ mutant embryos, however, the OLIG2 positive cells showed a subtle dorsal shift away from the edge of the third ventricle and additional expressing cells were detected in the basal hypothalamus (**Figure 6H**). In $Shh^{null/\Delta SBE2}$ mutant embryos, OLIG2 positive staining cells had dramatically shifted into the basal portion of the hypothalamus, with large areas of positive staining cells detected (**Figure 6I**). These data indicated that *Shh* activity acts to ensure adequate OLIG2 expression within the alar hypothalamus, whereupon reduction of SHH allows for OLIG2 expression within the basal hypothalamus. Unexpectedly, OLIG2 was found to be expanded into ectopic regions of the VH in mutant embryos rather than gradually reduced with increasing loss of SHH activity. However, Balaskas et al., 2012, demonstrated in the neural tube that, whilst NKX2-2 and OLIG2 are both equally responsive to SHH activity, higher levels of signalling are required for NKX2-2 expression as it is repressed by both PAX6 and OLIG2, creating a gene regulatory network between these three transcription factors. Thus, in this instance, the residual levels of *Shh* expression directed by SBE7 in $Shh^{null/\Delta SBE2}$ mutant embryos are likely sufficient to induce OLIG2 expression, yet not NKX2-2 due to the repressive activity of OLIG2 and PAX6.

NKX6-1, another transcription factor induced by SHH in the neural tube (Dessaud et al., 2008), was expressed in the alar hypothalamus in a narrow region along the edge of the third ventricle in sagittal and parasagittal regions (**Figure 6J**); but in contrast to OLIG2, these cells were only detected along the basal lateral edge of the third ventricle. In $Shh^{\Delta SBE2/\Delta SBE2}$ mutant embryos, NKX6-1 expression along the basal lateral edge of the third ventricle was undetectable (**Figure 6K**), however, isolated groups of expressing cells were detected in the basal hypothalamus (**Figure 6K**). $Shh^{null/\Delta SBE2}$ mutants showed complete loss of the region of staining along the basal lateral edge with no additional regions of expression observed (**Figure 6L**). NKX6-1 appears to require low levels of *Shh* expression to be induced, levels which are normally found along the edge of the third ventricle. In $Shh^{\Delta SBE2/\Delta SBE2}$ mutants the levels of *Shh* expression are reduced within the basal hypothalamus to a level that allows for ectopic expression of NKX6-1 within this region. In $Shh^{null/\Delta SBE2}$ mutants the complete loss of the expressing cells from the edge of the third ventricle indicates that the levels of *Shh* signalling activity within the VH, directed by SBE7, are insufficient to induce expression of NKX6-1. These data demonstrate a novel, previously undescribed role for SHH activity in the regulation of NKX6-1 expression within the hypothalamus, which mirrors the role described in the neural tube (Sander et al., 2000). Thus, providing further evidence that the signalling network which operates in the neural tube has been co-opted in hypothalamic development.

For all hypothalamic markers analysed, the extent of changes in expression correlate to the scale of alteration in levels of *Shh* expression, suggesting that *Shh* signalling is functioning in a concentration dependent manner to induce ventral patterning in the hypothalamus. The data described not only present novel examples of patterning genes regulated by SHH activity, as is the case for NKX6-1, but also expand on data presented by Corman et al. (2018). We further demonstrate that, not only do these genes require SHH activity to be adequately expressed, but they are capable of responding to varying levels of SHH. This concentration dependent effect mirrors that seen for *Shh* in the neural tube, and involves many of the same key players, indicating that *Shh* is acting as a morphogen within the VH to regulate tissue establishment within this critical brain region.

SBE2 Enhancer Activity Controls Downstream Effects Upon Hypothalamic Neuronal Populations

We next sought to assess the effects of disrupted DV patterning on neuronal populations. We first verified that there were no detectable differences in marker expression between wild-type and heterozygotes, $Shh^{\Delta SBE2/+}$ and $Shh^{null/+}$ embryos (**Supplementary Figure S14**), demonstrating that heterozygous loss of *Shh* has no gross effect upon DV neuronal population specification. To examine the effect of disrupted DV patterning upon the establishment of the different ventral hypothalamic neuronal populations, we first analysed expression of the broad ventral hypothalamic neuronal marker NKX2-1 (Ohyama et al., 2005). In control embryos at E13.5, NKX2-1 positive cells were detected across the entirety of the basal hypothalamic domain (**Figure 7A**). In $Shh^{\Delta SBE2/\Delta SBE2}$ mutant embryos, the region of NKX2-1 expressing cells was reduced and embedded within smaller domains intermingled with patches devoid of staining (**Figure 7B**). In the $Shh^{null/\Delta SBE2}$ mutants the region of expression was further reduced with larger and more numerous patches of negative cells observed (**Figure 7C**). These results indicate that whilst SHH activity is required for proper induction of the NKX2-1 neuronal population, that it is either responsive to very low levels of *Shh* signalling or another factor must also sustain the remnant expression seen in mutant embryos.

ISL1 is a melanocortinogenic neuronal marker, for which positive cells are detected throughout the hypothalamus (Kim et al., 2015). In control embryos positive staining was detected throughout the basal hypothalamic region, as expected (**Figure 7D**). In $Shh^{\Delta SBE2/\Delta SBE2}$ mutant embryos expression within the alar hypothalamus was reduced whilst within the basal hypothalamus it was unaffected (**Figure 7E**), in $Shh^{null/\Delta SBE2}$ mutants there was substantial loss of positive staining cells in both the alar and the basal hypothalamus (**Figure 7F**).

We additionally verified whether neurons that are restricted to specific nuclei of the basal hypothalamus were affected upon removal of SBE2 activity. Steroidogenic factor 1 (SF1) positive neurons are localised to the ventromedial hypothalamus (VMH) and are involved in the regulation of glucose metabolism and energy balance (Choi et al., 2013). In control embryos, as expected, SF1 positive neurons were detected in the developing VMH of the basal hypothalamus (**Figure 7G**). In $Shh^{\Delta SBE2/\Delta SBE2}$ mutant embryos the region of the basal hypothalamus expressing

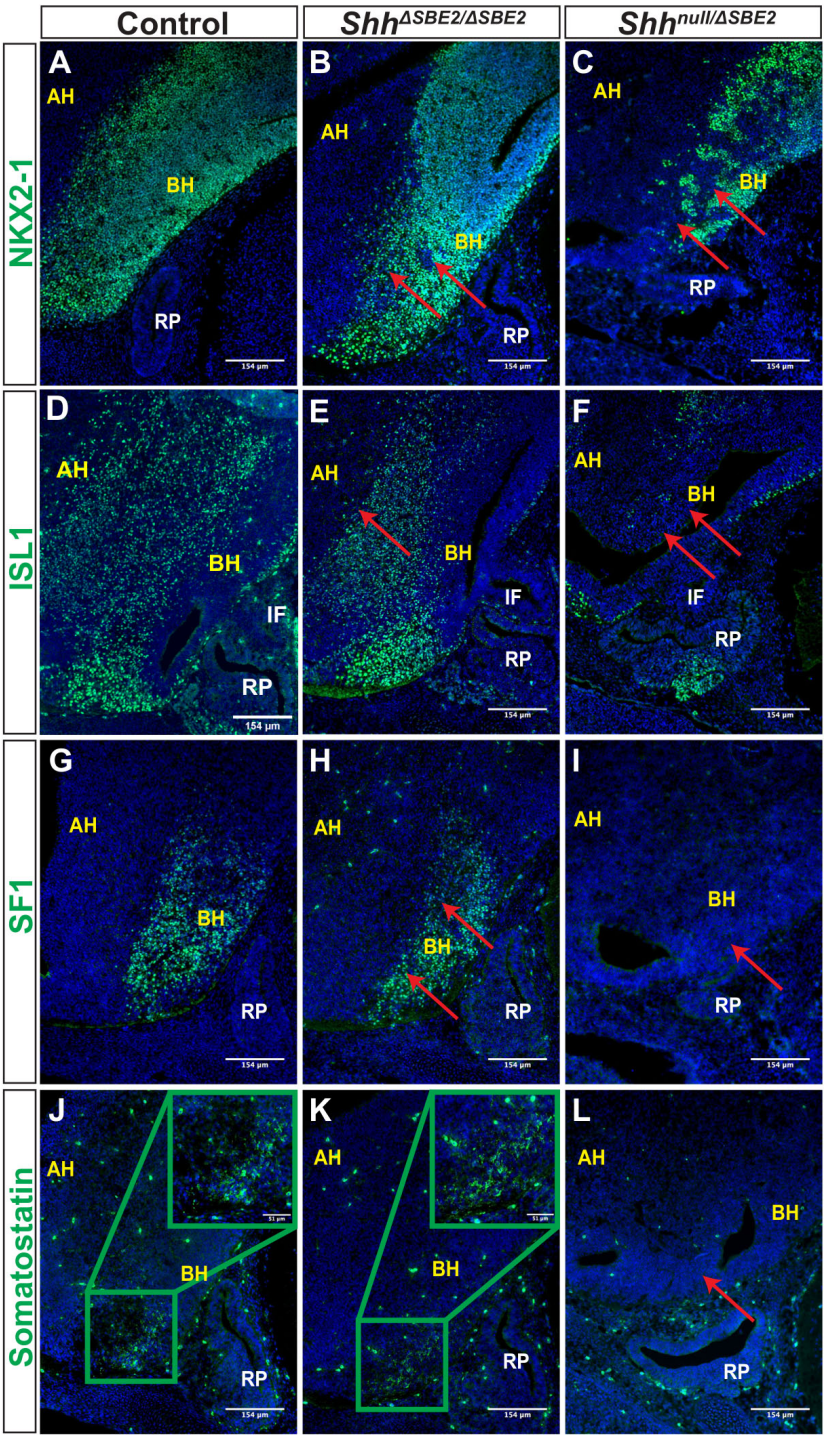


FIGURE 7 | DV hypothalamic neuronal populations are affected by loss of SBE2. **(A–C)** Immunofluorescent staining for NKX2-1 in E13.5 day VH parasagittal cryosections of control, $Shh^{\Delta SBE2/\Delta SBE2}$ and $Shh^{null/\Delta SBE2}$ embryos. Red arrows point to the regions devoid of NKX2-1 expression which are found upon removal of SBE2 function. **(D–F)** Immunofluorescent staining for ISL1 in E13.5 day VH parasagittal cryosections of control, $Shh^{\Delta SBE2/\Delta SBE2}$ and $Shh^{null/\Delta SBE2}$ embryos. Red arrows point to the regions of the BH where ISL1 is affected in $Shh^{\Delta SBE2/\Delta SBE2}$ or mostly absent in $Shh^{null/\Delta SBE2}$ embryos. **(G–I)** Immunofluorescent staining for SF1 in E13.5 day VH parasagittal cryosections of control, $Shh^{\Delta SBE2/\Delta SBE2}$ and $Shh^{null/\Delta SBE2}$ embryos. Red arrows point to the region of the BH where SF1 expression is reduced or absent upon removal of SBE2 function. **(J–L)** Immunofluorescent staining for somatostatin in E13.5 day VH parasagittal cryosections of control, $Shh^{\Delta SBE2/\Delta SBE2}$ and $Shh^{null/\Delta SBE2}$ embryos. Red arrows point to the region of the BH where somatostatin is lost in $Shh^{null/\Delta SBE2}$ embryos. A zoomed in view of the region of somatostatin expression is presented in the green boxes in the upper right hand corner of **(J,K)**. AH- alar hypothalamus; BH- basal hypothalamus; IF- infundibulum; RP- Rathke's pouch. Scale bars are shown in the bottom right hand corner (n = 3).

SF1 was reduced (**Figure 7H**), while in $Shh^{null/\Delta SBE2}$ mutant embryos there was a complete loss of SF1 positive neurons (**Figure 7I**). Somatostatin (SOM) expressing neurons are a subset of GABAergic inhibitory neurons that are detected in varied nuclei of the hypothalamus, including the VMH and the arcuate nucleus, which serve to regulate cell proliferation ultimately affecting growth (Urban-Ciecko and Barth, 2016). In control embryos, SOM neurons were detected in the basal hypothalamic region corresponding to the arcuate nucleus at E13.5 (**Figure 7J**). No differences were detected in $Shh^{\Delta SBE2/\Delta SBE2}$ mutant embryos where comparable staining was detected (**Figure 7K**); however, in $Shh^{null/\Delta SBE2}$ mutant embryos no positive staining for SOM was detected in the VH (**Figure 7L**). These results indicate that low levels of *Shh* in the VH are sufficient to retain specification of SOM neurons, but higher levels of *Shh* activity are required to sustain and/or specify the SF1 neuronal population in the basal hypothalamus. In the $Shh^{null/\Delta SBE2}$ mutant embryos both of these populations were absent which indicates that early *Shh* activity is essential for the adequate establishment of both the SF1 and SOM neuronal populations of the basal hypothalamus.

Corman et al. (2018), have previously demonstrated that loss of *Shh* signalling directed by SBE2 leads to a loss or reduction of neuronal markers. Our data expand on these findings by demonstrating a previously unknown concentration dependent requirement for *Shh* in VH neuronal development whereby some neuronal populations, such SOM, are only affected when the levels of *Shh* are vastly reduced, as in $Shh^{null/\Delta SBE2}$ mutants. In contrast, the NKX2-1 and SF1 populations are affected by moderate reductions in the levels of *Shh* and further affected by greater loss of *Shh* signalling.

DISCUSSION

Secondary Enhancer Activity Contributes to *Shh* Expression After SBE2 Inactivation

Loss of SBE2 in $Shh^{\Delta SBE2/\Delta SBE2}$ homozygous mice did not cause lethality or overt phenotypic effects in the brain and craniofacial regions due to additional enhancer activity most likely from the broadly acting element SBE7. A multistep series of events generate the normal *Shh* expression levels in the ventral diencephalon of which SBE2 activity plays a later role. SBE7 is responsible for early *Shh* expression throughout the prechordal plate a tissue that underlies the developing brain, from as early as E7.5 (Sagai et al., 2019). SBE7, subsequently, drives *Shh* expression at low levels in the ventral midline of the developing forebrain throughout later developmental stages. By E9.5 SBE2 activation, which is dependent on the earlier SBE7 activity (Sagai et al., 2019), contributes to a substantially increased level of expression specifically in the ventral diencephalon.

SBE2 activity does not appear to be obligatory since in its absence the animals are viable and fertile; raising questions about the value of this enhancer to the embryo. In contrast, the deep conservation of this element strongly argues that SBE2 provides fitness to individuals and is crucial across the vertebrate classes. Consequently, we argue that this is not an example of enhancer redundancy that commonly occurs in the mammalian genome (Osterwalder et al., 2018). Our analysis shows that expression of DV patterning genes is sensitive to a level of SHH which is perturbed in the absence of SBE2. SBE2 and SBE7 are components of a systematic temporal response required to reach sufficient SHH levels to attain the precise pattern of gene expression. Since further reduction of *Shh* levels (as in the $Shh^{null/\Delta SBE2}$ mutant) causes greater disruption to the expression of hypothalamic marker genes and perinatal lethality, it is reasonable to argue that, in the absence of SBE2, these animals are not fit in the wild. Disruption of neuronal populations seen in $Shh^{\Delta SBE2/\Delta SBE2}$ homozygous embryos, such as the reduction in the NKX2-1 and SF1 positive population, may lead to phenotypic effects on physiology, such as reproduction and energy balance. Indeed, specific depletion of the SF1 neuronal population of the VMH is known to be viable, however, the metabolic response of skeletal muscle is attenuated, revealing an effect upon metabolism (Fujikawa et al., 2016).

SBE2 Mediated Expression Links Brain and Craniofacial Development

Heterozygous inactivating mutations in the *SHH* gene in patients cause HPE (Chiang et al., 1996; Ming and Muenke, 2002) while in mice complete loss of *Shh* leads to multiple phenotypic consequences including the absence of brain and craniofacial structures (Chiang et al., 1996). While these observations highlight the links between brain and craniofacial development, the question as to whether *Shh* signalling from the neuroectoderm is directly required for the development of the craniofacial elements has remained largely unanswered. Physical interactions between the brain and face begin early at the initial stages of facial formation (Marcucio et al., 2011). There are two fundamentally different mechanisms by which development of the face and brain may be interrelated; firstly, the developing brain serves as a dynamic architectural foundation upon which the face responds and develops and secondly, signalling directly from the brain to the cranium and face regulates morphogenesis.

In fish, *Shh* signalling from both the oral ectoderm and neural tissue is required for adequate progression of craniofacial development (Wada et al., 2005). In mammalian development, *Shh* signalling from the oral ectoderm, which underlies the cranial base and also neighbours the facial skeletal elements, has been demonstrated to be key in adequate craniofacial development. Depletion of signalling from the oral ectoderm leads to defects in the maxillary and mandibular bones, the palatine bones and the basisphenoid bone among others (Jeong et al., 2004). The present study addressed the role of SBE2 forebrain expression activity on the development of adjacent tissue which would be predicted to form the midline of the cranial floor. Loss of SBE2 in the presence of the *Shh* null allele leads to both lethality and craniofacial malformations. Defects were observed in multiple skeletal elements of the cranial base including the vomer bone and the basisphenoid bone, which resides below the pituitary glands and the hypothalamic tissue, demonstrating that *Shh*

signalling from the rostral diencephalon is directly required for adequate craniofacial development.

SBE2 Affects Both AP and DV Patterning in the Hypothalamus

In addition to the craniofacial defects observed in the SBE2 mutant mouse lines, defects in hypothalamic development were also observed. This is perhaps unsurprising given that the region of *Shh* expression directed by SBE2 is the rostral diencephalon, one of the tissues from which the hypothalamus originates. However, the mode in which *Shh* regulates hypothalamic patterning and establishment appears to differ depending on the axis considered.

The AP axis was less sensitive to *Shh* levels of expression, whereby low, residual levels of signalling activity in the $Shh^{\Delta SBE2/\Delta SBE2}$ mutants were sufficient to sustain normal patterning. In contrast, further reduction of *Shh* expression in the $Shh^{null/\Delta SBE2}$ mutant embryos caused a loss of the anterior hypothalamic identity, which is promoted by *Shh*, in favour of the posterior hypothalamic identity. A shift in the anterior marker TCF4 was observed in the $Shh^{null/\Delta SBE2}$ mutant embryos, which contrasts previous observations (Zhao et al., 2012), where a shift in all other AP hypothalamic markers with the exception of TCF4 was found. In contrast, the lack of changes observed in pituitary marker expression agree with observations by Zhao et al. (2012) in SBE2-Cre Δ*Shh* mice, where pituitary markers were unperturbed, despite hypophyseal misplacement. Furthermore, SBE2-Cre Δ*Shh* mice displayed multiple invagination sites for RP, however only one site was seen in $Shh^{null/\Delta SBE2}$ embryos. The differences observed between our results and those of Zhao et al. (2012), likely reside in the different mouse lines used. Specifically, the use of SBE2-Cre to conditionally knock-down *Shh* activity may lead, firstly, to a temporal delay in the removal of *Shh* activity and, secondly, to a broader removal of *Shh* activity than that caused by direct deletion of SBE2. Notably, we also found that *Shh* is required for adequate development of hypothalamic neuronal populations which connect to the hypophyseal lobes and in turn, regulate hormonal release. These results demonstrate a dual role for *Shh* in ensuring hypophyseal function, directly in hypophyseal development and secondarily in regulating innervation.

In contrast to AP patterning, the DV hypothalamic axis was affected by *Shh* activity in a concentration dependent manner. This is apparent in the $Shh^{\Delta SBE2/\Delta SBE2}$ mutant embryos where defects in DV patterning, and also subsequent neuronal induction, were observed. These defects were far less severe than those observed in the $Shh^{null/\Delta SBE2}$ mutant embryos, where the disruption to DV development resulted in a complete loss of hypothalamic neuronal populations. The results presented regarding DV hypothalamic patterning are comparable to those described by Corman et al. (2018) where the same key neural tube patterning genes and subsequent neuronal regulators are involved. However, we have demonstrated a novel concentration dependent patterning activity of *Shh* in the hypothalamus, which has previously only been described in the neural tube. Moreover, the results demonstrate loss of neuronal populations beyond those of the dorsomedial, ventromedial and arcuate nuclei, additionally implicating neurons of the paraventricular nucleus, which play a key role in hypophyseal regulation. Corman et al. (2018) also propose that NKX2-2 is expressed in a domain between dorsal PAX6 and ventral NKX2-1; however, our data suggest that NKX2-1 is expressed broadly across the entire ventral hypothalamic region, with NKX2-2 expressed in a subdomain of the ventral hypothalamus, in regions overlapping with NKX2-1. It is possible that the ventral hypothalamic markers induced by *Shh*, including NKX2-2 are responsible for inducing expression of NKX2-1, whereby ventral hypothalamic markers that are not completely absent in the $Shh^{null/\Delta SBE2}$ mutant embryos, sustain remaining NKX2-1 expression.

Similar to the patterning process in the neural tube, the response of different DV patterning transcription factors to *Shh* was concentration dependent. For example, NKX6-1 was ventrally displaced in the $Shh^{\Delta SBE2/\Delta SBE2}$ mutants, much like OLIG2. However, it was lost altogether in $Shh^{null/\Delta SBE2}$ mutant embryos, unlike OLIG2 which was further displaced; indicating that NKX6-1 is more sensitive to loss of *Shh* signalling than OLIG2. Similarly, in the neural tube OLIG2 responds to a wider range of *Shh* signalling levels than NKX6-1. This same variable sensitivity to *Shh* signalling was seen for the neuronal populations, whereby only some of the neuronal population that were disrupted in the $Shh^{null/\Delta SBE2}$ mutant embryos were also affected in the $Shh^{\Delta SBE2/\Delta SBE2}$ mutants; for example, NKX2-1 expression required higher levels of *Shh* expression than those needed for somatostatin. Moreover, *Shh* was found to only be essential to maintain some neuronal populations, like SF1 and somatostatin populations, which were completely absent in $Shh^{null/\Delta SBE2}$ mutants, whereas other populations such as ISL1 persist despite near complete loss of VH *Shh* signalling. Thus, early SBE2 directed *Shh* signalling directs subsequent DV patterning in a concentration dependent manner acting as a morphogen within ventral hypothalamic development, in which transcription factor expression patterns are controlled by *Shh* levels which in turn instruct ventral hypothalamic neuronal populations.

SBE2 Coordinates Craniofacial Development and Patterning of the Hypothalamus

SBE2 regulates the expression of *Shh* in a region of the diencephalon in which SHH signalling coordinates development of both the diencephalon and the neighbouring tissue which forms the cranial and facial bones, skeletal elements which are in turn responsible for protecting the neural structures from external exposure. Homologues of the craniofacial bones disrupted in mouse are found in species of all vertebrate classes from fish to mammals (Maddin et al., 2016). In addition, the hypothalamus is an ancient brain centre critical for the production and release of key hormones (Xie and Dorsky, 2017). We suggest that deep in vertebrate evolution SBE2 activity arose as a means to govern VH development and as a consequence of hypothalamus proximity to craniofacial mesenchyme enabled *Shh* signalling to control formation of the bones to ensure its own protection. We have demonstrated that disruptions to coordinated development of the VH and

cranial vault in mice leads to phenotypic defects which reflect those seen for HPE patients. These results reveal not only that the interactions between the brain and the face display deeply conserved evolutionary roots, but also show the key importance of mouse models for addressing human disease phenotypes.

DATA AVAILABILITY STATEMENT

The raw data produced for this study has been deposited in the Gene Expression Omnibus, Accession number GSE158074.

ETHICS STATEMENT

All animal experiments were reviewed and approved by the University of Edinburgh Animal Welfare and Ethics Committee and were conducted with appropriate licensing under Animals (Scientific Procedures) Act 1986.

AUTHOR CONTRIBUTIONS

RH: conceptualization. LL, NK, and ZC-S: methodology. ZC-S and DK: validation. ZC-S, JS, KG, PD, LR, MD, JT, and LL: investigation. ZC-S and RH: writing—original draft. DK, JS, and LL: writing—review and editing. ZC-S, KG, JS, and LL: visualization. DK, JS, and RH: supervision. DK and RH: funding acquisition. All authors contributed to the article and approved the submitted version.

FUNDING

ZC-S, KG, PD, LR, EA, JT, LL, and RH were supported by the MRC University Human Genetics Unit grant (MM_UU_00007/8). NK and DK were funded by a CIHR operating grant to DK (MOP-275053). JS was supported by a BBSRC Institute Strategic Programme grant BBS/E/D/10002071.

ACKNOWLEDGMENTS

We would like to thank the MRC Technical Services for providing DNA sequence support. We would also like to thank the staff at the University of Edinburgh WGH BVS for expert technical assistance. Collaborative work was supported by the MRC and the University of Calgary.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcell.2021.595744/full#supplementary-material

REFERENCES

Anderson, E., Devenney, P. S., Hill, R. E., and Lettice, L. A. (2014). Mapping the Shh long-range regulatory domain. *Development* 141, 3934–3943. doi: 10.1242/dev.108480

Balaskas, N., Ribeiro, A., Panovska, J., Dessaud, E., Sasai, N., Page, K. M., et al. (2012). Gene regulatory logic for reading the Sonic Hedgehog signaling gradient in the vertebrate neural tube. *Cell* 148, 273–284. doi: 10.1016/j.cell.2011.10.047

Blaess, S., Szabó, N., Haddad-Tóvolli, R., Zhou, X., and Álvarez-Bolado, G. (2014). Sonic hedgehog signaling in the development of the mouse hypothalamus. *Front. Neuroanat.* 8:156. doi: 10.3389/fnana.2014.00156

Buenrostro, J. D., Wu, B., Chang, H. Y., and Greenleaf, W. J. (2015). ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.* 109, 21.29.1–21.29.9. doi: 10.1002/0471142727.mb2129s109

Burbridge, S., Stewart, I., and Placzek, M. (2016). Development of the neuroendocrine hypothalamus. *Compr. Physiol.* 6, 623–643. doi: 10.1002/cphy.c150023

Casillas, C., and Roelink, H. (2018). Gain-of-function Shh mutants activate Smo cell-autonomously independent of Ptch1/2 function. *Mech. Dev.* 153, 30–41. doi: 10.1016/j.mod.2018.08.009

Castinetti, F., Brinkmeier, M. L., Mortensen, A. H., Vella, K. R., Gergics, P., Brue, T., et al. (2015). ISL1 is necessary for maximal thyrotrope response to hypothyroidism. *Mol. Endocrinol.* 29, 1510–1521. doi: 10.1210/me.2015-1192

Chiang, C., Ying, L. T. T., Lee, E., Young, K. E., Corden, J. L., Westphal, H., et al. (1996). Cyclopia and defective axial patterning in mice lacking Sonic hedgehog gene function. *Nature* 383, 407–413. doi: 10.1038/383407a0

Choi, Y.-H., Fujikawa, T., Lee, J., Reuter, A., and Kim, K. W. (2013). Revisiting the ventral medial nucleus of the hypothalamus: the roles of SF-1 neurons in energy homeostasis. *Front. Neurosci.* 7:71. doi: 10.3389/fnins.2013.00071

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science (New York, N.Y.)* 339, 819–823. doi: 10.1126/science.1231143

Corman, T. S., Bergendahl, S. E., and Epstein, D. J. (2018). Distinct temporal requirements for Sonic hedgehog signaling in development of the tuberal hypothalamus. *Development (Cambridge, England)* 145:dev167379. doi: 10.1242/dev.167379

Croizier, S., Chometton, S., Fellmann, D., and Risold, P.-Y. (2015). Characterization of a mammalian prosencephalic functional plan. *Front. Neuroanat.* 8:161. doi: 10.3389/fnana.2014.00161

Dessaud, E., McMahon, A. P., and Briscoe, J. (2008). Pattern formation in the vertebrate neural tube: a sonic hedgehog morphogen-regulated transcriptional network. *Development* 135, 2489–2503. doi: 10.1242/dev.009324

Dessaud, E., Ribes, V., Balaskas, N., Yang, L. L., Pierani, A., Kicheva, A., et al. (2010). Dynamic assignment and maintenance of positional identity in the ventral neural tube by the morphogen sonic hedgehog. *PLoS Biol.* 8:e1000382. doi: 10.1371/journal.pbio.1000382

Dessaud, E., Yang, L. L., Hill, K., Cox, B., Ulloa, F., Ribeiro, A., et al. (2007). Interpretation of the sonic hedgehog morphogen gradient by a temporal adaptation mechanism. *Nature* 450, 717–720. doi: 10.1038/nature06347

Dworkin, S., Boglev, Y., Owens, H., Goldie, S., Dworkin, S., Boglev, Y., et al. (2016). The role of sonic hedgehog in craniofacial patterning, morphogenesis and cranial neural crest survival. *J. Dev. Biol.* 4:24. doi: 10.3390/jdb4030024

Ericson, J., Norlin, S., Jessell, T. M., and Edlund, T. (1998). Integrated FGF and BMP signaling controls the progression of progenitor cell differentiation and the emergence of pattern in the embryonic anterior pituitary. *Development (Cambridge, England)* 125, 1005–1015.

Ferrand, R. (1972). [Experimental study of the factors in cytological differentiation of the adenohypophysis in the chick embryo]. *Arch. Biol.* 83, 297–371.

Fujikawa, T., Castorena, C. M., Pearson, M., Kusminski, C. M., Ahmed, N., Battiprolu, P. K., et al. (2016). SF-1 expression in the hypothalamus is required for beneficial metabolic effects of exercise. *ELife* 5:e18206. doi: 10.7554/eLife.18206

Guillemot, F. (2007). Spatial and temporal specification of neural fates by transcription factor codes. *Development (Cambridge, England)* 134, 3771–3780. doi: 10.1242/dev.006379

Harfe, B. D., Scherz, P. J., Nissim, S., Tian, H., McMahon, A. P., and Tabin, C. J. (2004). Evidence for an expansion-based temporal shh gradient in specifying vertebrate digit identities. *Cell* 118, 517–528. doi: 10.1016/j.cell.2004.07.024

Hecksher-Sørensen, J., Hill, R. E., and Lettice, L. (1998). Double labeling for whole-mount in situ hybridization in mouse. *BioTechniques* 24, 914–918. doi: 10.2144/98246bm02

Hissnauer, T. N., Baranowsky, A., Pestka, J. M., Streichert, T., Wiegandt, K., Goepfert, C., et al. (2010). Identification of molecular markers for articular cartilage. *Osteoarthr. Cartil.* 18, 1630–1638. doi: 10.1016/J.JOCA.2010.10.002

Jeong, J. H., Mao, J. H., Tenzen, T., Kottmann, A. H., and McMahon, A. P. (2004). Hedgehog signaling in the neural crest cells regulates the patterning and growth of facial primordia. *Genes Dev.* 18, 937–951. doi: 10.1101/gad.1190304

Jeong, Y., El-Jaick, K., Roessler, E., Muenke, M., and Epstein, D. J. (2006). A functional screen for sonic hedgehog regulatory elements across a 1 Mb interval identifies long-range ventral forebrain enhancers. *Development* 133, 761–772. doi: 10.1242/dev.02239

Jeong, Y., Leskow, F. C., El-Jaick, K., Roessler, E., Muenke, M., Yocum, A., et al. (2008). Regulation of a remote Shh forebrain enhancer by the Six3 homeoprotein. *Nat. Genet.* 40, 1348–1353. doi: 10.1038/ng.230

Kawamura, K., and Kikuyama, S. (1995). Induction from posterior hypothalamus is essential for the development of the pituitary proopiomelacortin (POMC) cells of the toad (Bufo japonicus). *Cell Tissue Res.* 279, 233–239. doi: 10.1007/bf00318479

Kim, N., Park, C., Jeong, Y., and Song, M.-R. (2015). Functional diversification of motor neuron-specific *Isl1* enhancers during evolution. *PLoS Genet.* 11:e1005560. doi: 10.1371/journal.pgen.1005560

Larsson, A. J. M., Johnsson, P., Hagemann-Jensen, M., Hartmanis, L., Faridani, O. R., Reinius, B., et al. (2019). Genomic encoding of transcriptional burst kinetics. *Nature* 565, 251–254. doi: 10.1038/s41586-018-0836-1

Letelier, J., De La Calle-Mustienes, E., Pieretti, J., Naranjo, S., Maeso, I., Nakamura, T., et al. (2018). A conserved Shh cis-regulatory module highlights a common developmental origin of unpaired and paired fins. *Nat. Genet.* 50, 504–509. doi: 10.1038/s41588-018-0080-5

Lettice, L. A., Williamson, I., Devenney, P. S., Kilanowski, F., Dorin, J., and Hill, R. E. (2014). Development of five digits is controlled by a bipartite long-range cis-regulator. *Development* 141, 1715–1725. doi: 10.1242/dev.095430

Lu, F., Kar, D., Gruenig, N., Zhang, Z. W., Cousins, N., Rodgers, H. M., et al. (2013). Rax is a selector gene for mediobasal hypothalamic cell types. *J. Neurosci.* 33, 259–272. doi: 10.1523/JNEUROSCI.0913-12.2013

Mackenzie, M. A. F., Jordan, S. A., Budd, P. S., and Jackson, I. J. (1997). Activation of the receptor tyrosine kinase kit is required for the proliferation of melanoblasts in the mouse embryo. *Dev. Biol.* 192, 99–107. doi: 10.1006/DBIO.1997.8738

Maddin, H. C., Piekarski, N., Sefton, E. M., and Hanken, J. (2016). Homology of the cranial vault in birds: new insights based on embryonic fate-mapping and character analysis. *R. Soc. Open Sci.* 3:160356. doi: 10.1098/rsos.160356

Marcorelles, P., and Laquerriere, A. (2010). Neuropathology of holoprosencephaly. *Am. J. Med. Genet. Part C Semin. Med. Genet.* 154C, 109–119. doi: 10.1002/ajmg.c.30249

Marcucio, R. S., Young, N. M., Hu, D., and Hallgrimsson, B. (2011). Mechanisms that underlie co-variation of the brain and face. *Genesis* 49, 177–189. doi: 10.1002/dvg.20710

Ming, J. E., and Muenke, M. (2002). Multiple hits during early embryonic development: digenic diseases and holoprosencephaly. *Am. J. Hum. Genet.* 71, 1017–1032. doi: 10.1086/344412

Nagy, A., Gertsenstein, M., Vintersten, K., and Behringer, R. (2009a). Alizarin red staining of post-natal bone in mouse. *Cold Spring Harb. Protoc.* 2009:db.rot5171. doi: 10.1101/pdb.prot5171

Nagy, A., Gertsenstein, M., Vintersten, K., and Behringer, R. (2009b). Alcian blue staining of the mouse fetal cartilaginous skeleton. *Cold Spring Harb. Protoc.* 2009:db.rot5169. doi: 10.1101/pdb.prot5169

Nord, A. S., Blow, M. J., Attanasio, C., Akiyama, J. A., Holt, A., Hosseini, R., et al. (2013). Rapid and pervasive changes in genome-wide enhancer usage during mammalian development. *Cell* 155, 1521–1531. doi: 10.1016/j.cell.2013.11.033

Ohyama, K., Ellis, P., Kimura, S., and Placzek, M. (2005). Directed differentiation of neural cells to hypothalamic dopaminergic neurons. *Development (Cambridge, England)* 132, 5185–5197. doi: 10.1242/dev.02094

Osterwalder, M., Barozzi, I., Tissiéres, V., Fukuda-Yuzawa, Y., Mannion, B. J., Afzal, S. Y., et al. (2018). Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature* 554, 239–243. doi: 10.1038/nature25461

Pai, Y. J., Leung, K. Y., Savery, D., Hutchin, T., Prunty, H., Heales, S., et al. (2015). Glycine decarboxylase deficiency causes neural tube defects and features of non-ketotic hyperglycinemia in mice. *Nat. Commun.* 6:6388. doi: 10.1038/ncomms7388

Pontecorvi, M., Goding, C. R., Richardson, W. D., and Kessaris, N. (2008). Expression of Tbx2 and Tbx3 in the developing hypothalamic–pituitary axis. *Gene Expr. Patterns* 8, 411–417. doi: 10.1016/j.gep.2008.04.006

Sagai, T., Amano, T., Maeno, A., Ajima, R., and Shiroishi, T. (2019). SHH signaling mediated by a prechordal and brain enhancer controls forebrain organization. *Proc. Natl. Acad. Sci. U.S.A.* 116, 23636–23642. doi: 10.1073/pnas.1901732116

Sander, M., Paydar, S., Ericson, J., Briscoe, J., Berber, E., German, M., et al. (2000). Ventral neural patterning by Nkx homeobox genes: Nkx6.1 controls somatic motor neuron and ventral interneuron fates. *Genes Dev.* 14, 2134–2139. doi: 10.1101/gad.820400

Schwind, J. L. (1928). The development of the hypophysis cerebri of the albino rat. *Am. J. Anat.* 41, 295–319. doi: 10.1002/aja.1000410206

Sharpe, J. (2003). Optical projection tomography as a new tool for studying embryo anatomy. *J. Anat.* 202, 175–181. doi: 10.1046/J.1469-7580.2003.00155.X

Takuma, N., Sheng, H. Z., Furuta, Y., Ward, J. M., Sharma, K., Hogan, B. L., et al. (1998). Formation of Rathke's pouch requires dual induction from the diencephalon. *Development (Cambridge, England)* 125, 4835–4840.

Treier, M., Gleiberman, A. S., O'Connell, S. M., Szeto, D. P., McMahon, J. A., McMahon, A. P., et al. (1998). Multistep signaling requirements for pituitary organogenesis in vivo. *Genes Dev.* 12, 1691–1704. doi: 10.1101/gad.12.11.1691

Tsukiji, N., Amano, T., and Shiroishi, T. (2014). A novel regulatory element for Shh expression in the lung and gut of mouse embryos. *Mech. Dev.* 131, 127–136. doi: 10.1016/J.MOD.2013.09.003

Urban-Ciecko, J., and Barth, A. L. (2016). Somatostatin-expressing neurons in cortical networks. *Nat. Rev. Neurosci.* 17, 401–409. doi: 10.1038/nrn.2016.53

Wada, N., Javidan, Y., Nelson, S., Carney, T. J., Kelsh, R. N., and Schilling, T. F. (2005). Hedgehog signaling is required for cranial neural crest morphogenesis and chondrogenesis at the midline in the zebrafish skull. *Development (Cambridge, England)* 132, 3977–3988. doi: 10.1242/dev.01943

Williamson, I., Kane, L., Devenney, P. S., Flyamer, I. M., Anderson, E., Kilanowski, F., et al. (2019). Developmentally regulated Shh expression is robust to TAD perturbations. *Development (Cambridge)* 146:dev179523. doi: 10.1242/dev.179523

Xie, Y., and Dorsky, R. I. (2017). Development of the hypothalamus: conservation, modification and innovation. *Development (Cambridge, England)* 144, 1588–1599. doi: 10.1242/dev.139055

Yee, S. P., and Rigby, P. W. J. (1993). The regulation of myogenin gene expression during the embryonic development of the mouse. *Genes Dev.* 7, 1277–1289. doi: 10.1101/gad.7.7a.1277

Zhao, L., Zevallos, S. E., Rizzoti, K., Jeong, Y., Lovell-Badge, R., and Epstein, D. J. (2012). Disruption of SoxB1-dependent sonic hedgehog expression in the hypothalamus causes septo-optic dysplasia. *Dev. Cell* 22, 585–596. doi: 10.1016/J.DEVCEL.2011.12.023

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Citation: Crane-Smith Z, Schoenebeck J, Graham KA, Devenney PS, Rose L, Ditzell M, Anderson E, Thomson JI, Klenin N, Kurrasch DM, Lettice LA and Hill RE (2021) A Highly Conserved Shh Enhancer Coordinates Hypothalamic and Craniofacial Development. Front. Cell Dev. Biol. 9:595744. doi: 10.3389/fcell.2021.595744

Copyright © 2021 Crane-Smith, Schoenebeck, Graham, Devenney, Rose, Ditzell, Anderson, Thomson, Klenin, Kurrasch, Lettice and Hill. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.