



# 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.

# **Chromatin Dynamics at the Sonic Hedgehog Locus:**

**A study using limb derived Sonic  
Hedgehog inducible cell lines to  
investigate chromatin architecture.**

Adam Douglas

PhD by Research

University of Edinburgh

2016

MRC-Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, University  
of Edinburgh, Western General Hospital, Crewe Rd, Edinburgh,

United Kingdom EH4 2XU



Medical Research Council Human Genetics Unit  
MRC Institute of Genetics and Molecular Medicine  
[www.igmm.ac.uk](http://www.igmm.ac.uk)

# Declaration

I have conducted all work in this thesis unless otherwise stated. In these cases, the person/persons responsible have been fully acknowledged. All work has been conducted to obtain the qualification of “PhD by Research” and has not been used to obtain another degree or professional qualification.

Adam Douglas, September 2016

# Acknowledgements

Firstly, I would like to thank my supervisor Professor Robert E Hill for all his help and support throughout my 4 years within his laboratory. His open door policy of allowing me to walk straight into his office to chat helped to provide a friendly and enjoyable work environment. I am also appreciative of his encouragement to acquire additional skill sets - namely a grounding in bioinformatics - by attending various courses and seminars. The three weeks I spent at the University of Leipzig were particularly enjoyable.

Next, I would like to extend my gratitude to my project supervisor Dr Silvia Peluso who provided an exciting and innovative PhD project which allowed me to work on state-of-the art biological techniques such as chromatin conformation capture and new variations of fluorescent *in situ* hybridisation (FISH). I wish Silvia the best of luck in her scientific career and look forward to reading her future publications.

I also owe a massive thank you to Shelagh Boyle who assisted me in all things FISH related. Her patience, kindness and assistance when helping me to optimise tricky procedures went above and beyond the call of duty and I am deeply indebted to her. I am looking forward to the post-viva afternoon tea!! In the same way I must extend my thanks to Dr Iain Williamson who trained me to perform 5C and also provided the primers and some of the laboratory reagents needed to perform such experiments. Most importantly he is a Liverpool FC fan so we were always going to get along!!!

I must also thank each member of the Hill laboratory for their assistance at various stages throughout my studies. Dr Laura Lettice and Fiona Kilanowski provided me with the CRISPr/Cas9 ES cell mutants and Laura also assisted me in dissecting embryonic mouse limb buds on several occasions. Anna Thornburn's help was essential when setting up mice breedings and ensuring my pregnant mice were well looked after. I must also mention and thank Alison Hill for teaching me about all things tissue culture related. In addition, Carlo, Fay, Zoe, Joan, Giulia and the rest of both C3 and C4 provided an excellent working environment and I hope to keep in touch with everyone after I leave.

Beyond the Hill laboratory I would like to thank the Technical Services department for providing a number of reagents and liquid nitrogen on demand. Sometimes a good long football chat with Sean, Gaz, Derek, Jim or David helped to maintain my sanity on those days when everything seemed to go wrong. Agnes and Stuart were also essential in helping to

sequence both 5C and 4C libraries. I will miss my daily chats with Agnes and wish her a happy and fulfilling retirement.

On a personal note I would like to thank both my friends and family for all their love and support throughout the last four years. My girlfriend Fiona was a constant source of encouragement and understanding and a great comfort when the going got tough. My mother Glenda, father Derek and brother Mark have supported my studies from day one and have always been there when I have needed them most. I also owe a huge debt of thanks to my extended family: Alex; Barbara; Brian; Derek; Derrick; Flora; Georgia-meg; Margaret; Ruth; and Sam.

But most of all,

Mum and dad this is for you,

Adam

# Abstract

Enhancers are cis-regulatory sequences which promote the expression of target genes in a spatial and temporal fashion. They can be located within genes or between them and can act at distances of over 1 Mb. There are several different mechanisms by which enhancers regulate gene expression. Some, such as those regulating the *Hox* genes, are located close to each other in the genome in a structure referred to as a regulatory archipelago. These come together and act in combination to regulate gene expression, with different enhancer combinations resulting in different patterns of expression. On the other hand, enhancers can act individually, with designated enhancers responsible for regulating the expression of the same gene in different tissues or at different stages of development. Indeed, this is the case for the Sonic Hedgehog gene (*Shh*) where several different enhancers located within a gene sparse region referred to as a gene desert, act separately leading to *Shh* expression in areas such as the brain, the lungs, the notochord and neural tube and the limbs.

Within the developing mouse embryo, *Shh* is expressed over roughly a two day period from E10 to E12 in a posterior distal region referred to as the Zone of Polarising Activity (ZPA). Ectopic expression in anterior regions has been observed in some common congenital diseases which affect the limbs, sometimes resulting in the formation of extra digits. The reason for this mis-expression is largely due to defects in the *Shh* limb enhancer commonly referred to as the Zone of Polarising Activity Regulatory Sequence (ZRS). Mutations within this highly conserved sequence create additional protein binding sites thus activating the enhancer in the wrong locations. The associated diseases are known collectively as the ZRS associated syndromes and can range from the less severe phenotype of preaxial polydactyly type II (characterised by an extra digit near the thumb) to the more severe Werner Mesomelic Syndrome (WMS), where patients present with a clear displacement of their tibia.

The mechanism by which the ZRS functions is yet to be fully elucidated, with current studies producing conflicting data. What is known, is that the region encapsulating the *Shh* gene is highly compact, with both the gene and its enhancers located in a highly conserved Topological Associated Domain (TAD) as proven by Hi-C experiments. The boundaries of this domain are likely created by the binding of the protein CTCF to specified binding sites located at the either end of the locus. This restricts the ability of the enhancers to regulate the expression of genes outside the TAD.

To study the exact mechanism by which the ZRS is activated and regulates *Shh* expression, the Hill laboratory has used cultured cell lines derived from the posterior regions of an E11.5 limb bud. Gene expression in these cells is highly reflective of the posterior limb bud, with the key exception being *Shh*, which is not expressed. However, using different drug treatments or biological manipulations *Shh* can be activated thereby making this the perfect system to analyse the mechanisms leading to *Shh* activation.

In this investigation the cell lines were used to determine how the position of the ZRS changes upon activation. Using techniques such as Fluorescent *in situ* hybridisation (FISH) with either fosmid probes or directly labelled probes called MYtags, it was confirmed that the *Shh* locus is indeed highly compact in both *Shh* expressing and non-expressing cells. However, no differences were observed in terms of the distance between the ZRS and *Shh* between these two conditions in our cell lines. Next, both carbon copy chromosome conformation capture (5C) and circular chromosome conformation capture (4C) were used to look at changes to the *Shh* locus in different conditions. This confirmed Hi-C experiments and other recent publications suggesting that *Shh* is located within a TAD, the position of which is highly conserved between different conditions and cell lines. Furthermore, treatments activating the *Shh* gene resulted in significant deviations to the chromatin interactions within the locus suggesting a repositioning of structures when the gene is active.

It is believed that the use of *Shh* inducible limb derived cell lines will prove extremely useful in future scientific endeavours to study the mechanisms of mammalian limb development. These provide a quick and easy means of accessing large numbers of *Shh* expressing cells, a feature which is increasingly important in an era where large cell numbers are needed for conducting chromosome conformation capture experiments such as Hi-C, 5C and 4C.

# Lay Abstract

During embryonic development genes are controlled by sequences called enhancers which are able to turn genes on and off at very specific times. This is extremely important in the process of limb development as the specific times when enhancers are active determine the precise number of fingers and toes in fully formed mammals. One of the main genes orchestrating limb development is named after the popular fictional computer character Sonic Hedgehog and is expressed at the base of the developing limb in healthy embryos. However, DNA mutations can lead to expression in other locations within the limb resulting in congenital malformations. Most commonly extra digits are observed - a condition referred to as polydactyly or the fusion of digits, referred to as syndactyl. This condition occurs not just in humans but also in other mammals such as mice, chickens and cats. Indeed, Ernest Hemmingway was well known for owning a polydactylous cat with 6 digits, descendants of which are now referred to as Hemingway's cats.

Understanding the mechanisms by which enhancers operate to produce a fully developed limb is a major focus of scientific research today. Indeed, members of the Hill laboratory focus particularly on the enhancer which regulates Sonic Hedgehog. They study the effect of mutations, its sequence composition and its relationship with the Sonic Hedgehog gene when it is being expressed and when it is not. The focus of this investigation was to determine how the position of the Sonic Hedgehog enhancer changes when the Sonic Hedgehog gene is turned off and when it is turned on. To do this, cells taken from an embryonic limb bud were grown and used to perform various laboratory based techniques. Consequently, it was found that the Sonic Hedgehog enhancer maintains a very close relationship with the gene in all conditions and is maintained in a very compact structure.



# Table of Contents

|                                                                                                       |             |
|-------------------------------------------------------------------------------------------------------|-------------|
| <b>Declaration</b> .....                                                                              | <b>ii</b>   |
| <b>Acknowledgements</b> .....                                                                         | <b>iii</b>  |
| <b>Abstract</b> .....                                                                                 | <b>v</b>    |
| <b>Lay Abstract</b> .....                                                                             | <b>vii</b>  |
| <b>List of Figures</b> .....                                                                          | <b>xiii</b> |
| <b>List of Tables</b> .....                                                                           | <b>xv</b>   |
| <b>List of Abbreviations</b> .....                                                                    | <b>xvii</b> |
| <b>Chapter 1</b> .....                                                                                | <b>1</b>    |
| <b>1.1 Hedgehog Signalling</b> .....                                                                  | <b>2</b>    |
| 1.1.1 The Hedgehog Protein; Discovery of the Hedgehog Protein and its Processing in the Nucleus ..... | 2           |
| 1.1.2 The Hedgehog Signalling Pathway .....                                                           | 5           |
| <b>1.2 Signalling in the limb bud</b> .....                                                           | <b>7</b>    |
| 1.2.1 Stages of Limb Development .....                                                                | 7           |
| 1.2.2 Initiation of the Limb Bud .....                                                                | 8           |
| 1.2.3 Formation of the Proximal-Distal Axis .....                                                     | 11          |
| 1.2.4 Formation of the Anterior-Posterior Axis .....                                                  | 15          |
| 1.2.5 Communication Between the ZPA and the AER is Essential for Limb Development.....                | 16          |
| 1.2.6 Formation of the Dorso-Ventral Axis .....                                                       | 18          |
| 1.2.7 Evolution of the Tetrapod Limb .....                                                            | 19          |
| <b>1.3 Sonic Hedgehog</b> .....                                                                       | <b>21</b>   |
| 1.3.1 <i>Shh</i> related limb abnormalities .....                                                     | 21          |
| 1.3.2 <i>Shh</i> signalling in the Neural Tube.....                                                   | 24          |
| 1.3.3 <i>Shh</i> signalling in Craniofacial Development.....                                          | 24          |
| 1.3.4 <i>Shh</i> signalling in the Lungs .....                                                        | 25          |
| 1.3.5 <i>Shh</i> signalling in the Gut.....                                                           | 25          |
| 1.3.6 Other regions involving <i>Shh</i> signalling .....                                             | 25          |
| 1.3.7 The <i>Shh</i> Regulatory Domain.....                                                           | 25          |
| 1.3.8 The Mechanism of the ZRS .....                                                                  | 26          |
| <b>1.4 Enhancers</b> .....                                                                            | <b>30</b>   |
| 1.4.1 Regulatory Locus Composition .....                                                              | 30          |
| 1.4.2 Super Enhancers and Stretch Enhancers .....                                                     | 32          |
| 1.4.3 Evolution of Enhancers .....                                                                    | 33          |

|                  |                                                                          |           |
|------------------|--------------------------------------------------------------------------|-----------|
| 1.4.3.1          | <b>Highly Conserved and Ultra-Conserved Enhancer Sequences</b>           | <b>33</b> |
| 1.4.3.2          | <b>Conservation of Function without Conservation of Sequence</b>         | <b>35</b> |
| 1.4.3.3          | <b>Conservation of Sequence vs Conservation of Function</b>              | <b>36</b> |
| 1.4.3.4          | <b>Mechanisms of Ultra-Conservation</b>                                  | <b>37</b> |
| 1.4.3.5          | <b>Evolution of Enhancers</b>                                            | <b>37</b> |
| 1.4.3.6          | <b>Enhancer Mutations and Disease</b>                                    | <b>38</b> |
| 1.4.4            | Enhancer Identification                                                  | 39        |
| ➤                | Chromosome Conformation Capture (3C) Techniques                          | 40        |
| ➤                | Fluorescent <i>in situ</i> hybridisation (FISH)                          | 41        |
| ➤                | DNase1-seq, MNase-seq, FAIRE-seq and ATAC-seq                            | 42        |
| ➤                | Reporter Assays, STARR-seq and Enhancer Trap                             | 43        |
| ➤                | ChIP-seq                                                                 | 44        |
| <b>1.5</b>       | <b>Chromatin Structure</b>                                               | <b>46</b> |
| 1.5.1            | Topological Associated Domains (TADs)                                    | 46        |
| 1.5.2            | Identifying Mechanisms of Loop Formation through Polymer based Modelling | 48        |
| 1.5.3            | Mutations at TAD boundaries cause disease                                | 49        |
| 1.5.4            | Lamin Associated Domains (LADs)                                          | 53        |
| 1.5.5            | Laminopathies                                                            | 55        |
| 1.5.6            | Radial Positioning in the Nucleus and Chromosome Territories             | 55        |
| 1.5.7            | Transcription Factories                                                  | 57        |
| 1.5.8            | 10 nm Fibre (beads-on-a-string) and 30 nm Fibre Chromatin Structures     | 59        |
| 1.5.9            | Higher Order Chromatin Structure                                         | 60        |
| <b>1.6</b>       | <b>Project Aims</b>                                                      | <b>61</b> |
| <b>Chapter 2</b> |                                                                          | <b>62</b> |
| <b>2.1</b>       | <b>Materials</b>                                                         | <b>63</b> |
| <b>2.2</b>       | <b>General Laboratory Solutions</b>                                      | <b>66</b> |
| <b>2.3</b>       | <b>General Laboratory Procedures</b>                                     | <b>67</b> |
| 2.3.1            | Isolation of RNA                                                         | 67        |
| 2.3.2            | Assessment of RNA purity, quantity and integrity                         | 67        |
| 2.3.3            | Reverse Transcription Reaction                                           | 68        |
| 2.3.4            | Quantitative Real-Time PCR (qPCR)                                        | 68        |
| 2.3.5            | Phenol Chloroform Extractions                                            | 70        |
| 2.3.6            | DNA Ethanol Precipitations                                               | 70        |
| 2.3.7            | Agarose Gel Electrophoresis                                              | 71        |
| 2.3.8            | Generation and Statistical Analysis of RNA-seq data                      | 71        |
| <b>2.4</b>       | <b>Fluorescent <i>in situ</i> hybridisation (FISH)</b>                   | <b>72</b> |
| 2.4.1            | Cell Culture and Cell Fixation                                           | 72        |
| 2.4.2            | Purification of Fosmid DNA                                               | 72        |
| 2.4.3            | Nick Translation Reactions (NTR)                                         | 73        |

|                       |                                                                              |            |
|-----------------------|------------------------------------------------------------------------------|------------|
| 2.4.4                 | 3D-FISH.....                                                                 | 74         |
| 2.4.5                 | Deconvolution Microscopy .....                                               | 75         |
| 2.4.6                 | Preparation of cells for metaphase spreads. ....                             | 76         |
| 2.4.7                 | Preparation of cells for 2D-FISH.....                                        | 76         |
| 2.4.7                 | 2D-FISH.....                                                                 | 77         |
| <b>2.5</b>            | <b>Chromosome Conformation Capture Techniques .....</b>                      | <b>78</b>  |
| 2.5.1                 | Preparation of 3C libraries .....                                            | 78         |
| 2.5.2                 | Preparation of 4C libraries .....                                            | 80         |
| 2.5.3                 | Amplification of 4C libraries.....                                           | 81         |
| 2.5.4                 | Sequencing of 4C libraries.....                                              | 82         |
| 2.5.5                 | Preparation of 5C libraries .....                                            | 83         |
| 2.5.5.1               | <b>Titration of amplified 3C libraries .....</b>                             | <b>83</b>  |
| 2.5.5.2               | <b>Annealing Reaction.....</b>                                               | <b>85</b>  |
| 2.5.5.3               | <b>Ligation Reaction.....</b>                                                | <b>100</b> |
| 2.5.5.4               | <b>5C library amplification .....</b>                                        | <b>101</b> |
| 2.5.5.5               | <b>Purification, Quantification and Sequencing of Amplified 5C libraries</b> | <b>103</b> |
| 2.5.6                 | Analysis of 4C and 5C data .....                                             | 103        |
| <b>2.6</b>            | <b>Cell Culture Techniques .....</b>                                         | <b>106</b> |
| 2.6.1                 | Immortalised Cell lines.....                                                 | 106        |
| 2.6.2                 | Cell Passaging .....                                                         | 106        |
| 2.6.3                 | Drug Treatments .....                                                        | 106        |
| 2.6.4                 | Mycoplasma Testing .....                                                     | 107        |
| <b>2.7</b>            | <b>Immunofluorescence.....</b>                                               | <b>108</b> |
| <b>2.8</b>            | <b>Mouse Work .....</b>                                                      | <b>109</b> |
| 2.8.1                 | Mouse Dissections .....                                                      | 109        |
| <b>Chapter 3.....</b> | <b>110</b>                                                                   |            |
| <b>3.1</b>            | <b>Introduction .....</b>                                                    | <b>111</b> |
| 3.1.1                 | The 14fp cell line .....                                                     | 112        |
| 3.1.2                 | Additional Cell Lines.....                                                   | 113        |
| 3.1.3                 | Alternatives to TSA.....                                                     | 113        |
| <b>3.2</b>            | <b>Aims.....</b>                                                             | <b>117</b> |
| 3.2.1                 | General Aims.....                                                            | 117        |
| 3.2.2                 | Experimental Aims .....                                                      | 117        |
| <b>3.3</b>            | <b>Experimental Procedures.....</b>                                          | <b>118</b> |
| 3.3.1                 | 14fp Cells: Preparation of RNA for sequencing.....                           | 118        |
| 3.3.2                 | E11.5 Limb Cells: Preparation of RNA for sequencing .....                    | 118        |
| 3.2.1                 | RNA-sequencing: Analysis Pipeline .....                                      | 118        |
| <b>3.4</b>            | <b>Results and Discussion.....</b>                                           | <b>119</b> |

|            |                                                                                                                                                                      |            |
|------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------|
| 3.4.1      | Processing of RNA sequencing data.....                                                                                                                               | 119        |
| 3.4.2      | Quality Control of RNA sequencing data .....                                                                                                                         | 120        |
| 3.4.3.     | Genome wide FPKM changes for cell line and limb dissection experiments ..                                                                                            | 120        |
| 3.4.4      | TSA treatment of 14fp cells affects a variety of cellular processes .....                                                                                            | 124        |
| 3.4.5      | Gene expression differences between E11.5 proximal and distal limb sections reflect the position of the different signaling centers within the developing limb bud.. | 132        |
| 3.4.6      | A subset of genes are differentially expressed in both cell line and limb dissection experiments. ....                                                               | 139        |
| 3.4.7      | 88fp cells react similarly to 14fp cells when treated with TSA .....                                                                                                 | 145        |
| <b>3.5</b> | <b>Conclusions .....</b>                                                                                                                                             | <b>147</b> |
|            | <b>Chapter 4.....</b>                                                                                                                                                | <b>148</b> |
| <b>4.1</b> | <b>Introduction .....</b>                                                                                                                                            | <b>149</b> |
| <b>4.2</b> | <b>Aims.....</b>                                                                                                                                                     | <b>152</b> |
| 4.2.1      | General Aims.....                                                                                                                                                    | 152        |
| 4.2.2      | Experimental Aims .....                                                                                                                                              | 152        |
| <b>4.3</b> | <b>Results .....</b>                                                                                                                                                 | <b>153</b> |
| 4.3.1      | 14fp, MD cells and 88fp cells are all hyperdiploid .....                                                                                                             | 153        |
| 4.3.2      | <i>Shh</i> -ZRS co-localisation does not change in 14fp, 88fp and MD cell lines on TSA treatment.....                                                                | 155        |
| 4.3.3      | The localisation of probes at the Nuclear Periphery is unaffected by TSA treatment.....                                                                              | 161        |
| 4.3.4      | Knock down of PEA3 induces SHH expression, but has no effect on <i>Shh</i> -ZRS co-localisation .....                                                                | 163        |
| 4.3.5      | TSA treatment causes <i>Shh</i> to loop out of its Chromosome Territory .....                                                                                        | 166        |
| 4.3.6      | Analysis of Chromatin Compaction at the <i>Shh</i> locus in 14fp cells .....                                                                                         | 168        |
| 4.3.7      | Analysis of Chromatin Compaction at the <i>Shh</i> locus in ES cells.....                                                                                            | 173        |
| <b>4.4</b> | <b>Discussion .....</b>                                                                                                                                              | <b>179</b> |
| <b>4.5</b> | <b>Conclusions .....</b>                                                                                                                                             | <b>182</b> |
|            | <b>Chapter 5.....</b>                                                                                                                                                | <b>183</b> |
| <b>5.1</b> | <b>Introduction .....</b>                                                                                                                                            | <b>184</b> |
| <b>5.2</b> | <b>Aims.....</b>                                                                                                                                                     | <b>186</b> |
| 5.2.1      | General Aims.....                                                                                                                                                    | 186        |
| 5.2.2      | Experimental Aims .....                                                                                                                                              | 186        |
| <b>5.3</b> | <b>Results .....</b>                                                                                                                                                 | <b>187</b> |
| 5.3.1.     | 5C quality control and processing.....                                                                                                                               | 187        |
| 5.3.2      | Optimisation of 5C Procedure.....                                                                                                                                    | 193        |
| 5.3.3      | The TAD containing the <i>Shh</i> regulatory domain is conserved across the cell lines                                                                               | 193        |

|            |                                                                                                                     |            |
|------------|---------------------------------------------------------------------------------------------------------------------|------------|
| 5.3.4      | <i>Shh</i> and the ZRS show high interaction frequencies in <i>Shh</i> expressing and non-expressing cells .....    | 194        |
| 5.3.5      | In 14fp and 88fp cells TSA treatment reduces the interaction frequency of regions located around <i>Lmbr1</i> ..... | 195        |
| 5.3.6      | Modelling of 5C data.....                                                                                           | 201        |
| 5.3.7      | Preliminary 4C data suggests a specific interaction at the <i>Shh</i> gene of TSA treated 14fp cells .....          | 205        |
| <b>5.4</b> | <b>Discussion .....</b>                                                                                             | <b>209</b> |
| <b>5.5</b> | <b>Conclusions .....</b>                                                                                            | <b>211</b> |
|            | <b>Chapter 6.....</b>                                                                                               | <b>212</b> |
| <b>6.1</b> | <b>14fp Cells - a model system to study <i>Shh</i> limb bud expression .</b>                                        | <b>213</b> |
| <b>6.2</b> | <b>Model of TSA induced <i>Shh</i> expression in 14fp cells.....</b>                                                | <b>214</b> |
| <b>6.3</b> | <b>Compaction of the <i>Shh</i> regulatory region is unaffected by <i>Shh</i> expression .....</b>                  | <b>215</b> |
| <b>6.4</b> | <b>Interactions within the <i>Shh</i> regulatory region.....</b>                                                    | <b>217</b> |
| <b>6.5</b> | <b>Future Work.....</b>                                                                                             | <b>218</b> |
|            | <b>Chapter 7.....</b>                                                                                               | <b>221</b> |
|            | <b>Chapter 8.....</b>                                                                                               | <b>227</b> |
|            | <b>Chapter 9.....</b>                                                                                               | <b>244</b> |

# List of Figures

## Chapter 1: Introduction

|                 |    |
|-----------------|----|
| Figure 1.1..... | 6  |
| Figure 1.2..... | 9  |
| Figure 1.3..... | 13 |
| Figure 1.4..... | 23 |
| Figure 1.5..... | 28 |
| Figure 1.6..... | 45 |
| Figure 1.7..... | 51 |
| Figure 1.8..... | 58 |

## Chapter 2: Materials and Methods

|                 |     |
|-----------------|-----|
| Figure 2.1..... | 104 |
|-----------------|-----|

## Chapter 3: Using Sonic Hedgehog Inducible Cell Lines as a Model System to Study the Limb

|                  |     |
|------------------|-----|
| Figure 3.1.....  | 115 |
| Figure 3.2.....  | 122 |
| Figure 3.3.....  | 123 |
| Figure 3.4.....  | 127 |
| Figure 3.5.....  | 129 |
| Figure 3.6.....  | 131 |
| Figure 3.7.....  | 135 |
| Figure 3.8.....  | 137 |
| Figure 3.9.....  | 141 |
| Figure 3.10..... | 144 |
| Figure 3.11..... | 146 |

## Chapter 4: Chromatin Compaction of the *Shh* locus

|                  |     |
|------------------|-----|
| Figure 4.1.....  | 151 |
| Figure 4.2.....  | 154 |
| Figure 4.3.....  | 157 |
| Figure 4.4.....  | 159 |
| Figure 4.5.....  | 162 |
| Figure 4.6.....  | 164 |
| Figure 4.7.....  | 167 |
| Figure 4.8.....  | 170 |
| Figure 4.9.....  | 171 |
| Figure 4.10..... | 172 |
| Figure 4.11..... | 175 |
| Figure 4.12..... | 176 |
| Figure 4.13..... | 177 |

|                  |     |
|------------------|-----|
| Figure 4.14..... | 178 |
| Figure 4.15..... | 181 |

## **Chapter 5: Conformational Changes within the *Shh* TAD**

|                  |     |
|------------------|-----|
| Figure 5.1.....  | 190 |
| Figure 5.2.....  | 191 |
| Figure 5.3.....  | 192 |
| Figure 5.4.....  | 196 |
| Figure 5.5.....  | 197 |
| Figure 5.6.....  | 198 |
| Figure 5.7.....  | 199 |
| Figure 5.8.....  | 200 |
| Figure 5.9.....  | 202 |
| Figure 5.10..... | 203 |
| Figure 5.11..... | 204 |
| Figure 5.12..... | 207 |

## **Chapter 6: Discussion**

|                 |     |
|-----------------|-----|
| Figure 6.1..... | 220 |
|-----------------|-----|

## **Chapter 7: Appendices**

|                 |     |
|-----------------|-----|
| Figure 7.1..... | 223 |
| Figure 7.2..... | 224 |
| Figure 7.3..... | 225 |
| Figure 7.4..... | 226 |

# List of Tables

## Chapter 2: Using Sonic Hedgehog Inducible Cell Lines as a Model System to Study the Limb

|                 |     |
|-----------------|-----|
| Table 2.1.....  | 69  |
| Table 2.2.....  | 69  |
| Table 2.3.....  | 70  |
| Table 2.4.....  | 73  |
| Table 2.5.....  | 75  |
| Table 2.6.....  | 80  |
| Table 2.7.....  | 82  |
| Table 2.8.....  | 82  |
| Table 2.9.....  | 84  |
| Table 2.10..... | 84  |
| Table 2.11..... | 85  |
| Table 2.12..... | 85  |
| Table 2.13..... | 86  |
| Table 2.14..... | 86  |
| Table 2.15..... | 87  |
| Table 2.16..... | 101 |
| Table 2.17..... | 102 |
| Table 2.18..... | 102 |
| Table 2.19..... | 102 |
| Table 2.20..... | 103 |

## Chapter 3: Using Sonic Hedgehog Inducible Cell Lines as a Model System to Study the Limb

|                |     |
|----------------|-----|
| Table 3.1..... | 121 |
| Table 3.2..... | 126 |
| Table 3.3..... | 128 |
| Table 3.4..... | 130 |
| Table 3.5..... | 134 |
| Table 3.6..... | 136 |
| Table 3.7..... | 138 |
| Table 3.8..... | 142 |
| Table 3.9..... | 143 |

## Chapter 4: Chromatin Compaction of the *Shh* locus

|                |     |
|----------------|-----|
| Table 4.1..... | 156 |
|----------------|-----|



## **Chapter 5: Conformational Changes within the *Shh* TAD**

|                |     |
|----------------|-----|
| Table 5.1..... | 188 |
| Table 5.2..... | 189 |
| Table 5.3..... | 206 |

## **Chapter 7: Appendices**

|                |     |
|----------------|-----|
| Table 7.1..... | 222 |
|----------------|-----|

# List of Abbreviations

## A

|      |                                            |
|------|--------------------------------------------|
| A    | Adenine                                    |
| AC   | Astrocytes                                 |
| ADLD | Autosomal dominant leukodystrophy          |
| AER  | Apical ectodermal ridge                    |
| ATAC | Assay for transposase-accessible chromatin |
| ATP  | Adenosine triphosphate                     |

## B

|     |                                        |
|-----|----------------------------------------|
| BRF | Edinburgh biomedical research facility |
| BSA | Bovine serum albumin                   |

## C

|          |                                                             |
|----------|-------------------------------------------------------------|
| C        | Cytosine                                                    |
| CBS      | CTCF binding sites                                          |
| CD       | Campomelic dysplasia                                        |
| cDNA     | Complementary DNA                                           |
| CEU      | Cohesive end units                                          |
| ChIA-PET | Chromatin interaction analysis by paired-end tag sequencing |
| ChIP     | Chromatin immunoprecipitation                               |
| CNE      | Conserved non-coding element                                |
| CNS      | Central nervous system                                      |
| CNV      | Copy number variants                                        |
| CRISPR   | Clustered regularly interspaced short palindromic repeats   |
| CT       | Chromosome territory                                        |
| CTCF     | CCCTC-binding factor                                        |
| Cys      | Cysteine                                                    |

## D

|       |                                              |
|-------|----------------------------------------------|
| DAFs  | Derived allele frequencies                   |
| DamID | DNA adenine methyltransferase identification |

|       |                                    |
|-------|------------------------------------|
| DAPI  | 4',6-diamidino-2-phenylindole      |
| Dig   | Digoxigenin                        |
| DHSs  | DNase1 hypersensitivity sites      |
| DMEM  | Dulbecco's modified eagle's medium |
| DMSO  | Dimethyl sulfoxide                 |
| DNA   | Deoxyribonucleic acid              |
| Dox   | Doxycycline                        |
| dNTPS | Deoxyribonucleic acid solution     |
| DTT   | Dithiothreitol                     |

## E

|        |                                           |
|--------|-------------------------------------------|
| E      | Embryonic day                             |
| EDMD   | Emery-Dreifuss muscular dystrophy         |
| EDTA   | Ethylenediaminetetra - acetic acid        |
| ESI-MS | Electrospray ionisation mass spectrometry |
| ER     | Endoplasmic reticulum                     |
| ESCs   | Embryonic stem cells                      |

## F

|       |                                                        |
|-------|--------------------------------------------------------|
| FAIRE | Formaldehyde assisted isolation of regulatory elements |
| FCS   | Foetal calf serum                                      |
| FDR   | False discovery rate                                   |
| FISH  | Fluorescent <i>in situ</i> hybridisation               |
| FITC  | Fluorescein isothiocyanate                             |
| FPKM  | Fragments per kilobase of sequence mapped              |
| FTIs  | Farnesyltransferase inhibitors                         |

## G

|   |                                                      |
|---|------------------------------------------------------|
| g | grams/relative centrifugal force (context dependent) |
|---|------------------------------------------------------|

|                 |                                        |                 |                                      |
|-----------------|----------------------------------------|-----------------|--------------------------------------|
| G               | Guanine                                | mESCs           | Mouse embryonic stem cells           |
| gDNA            | Genomic DNA                            | mEpiSCs         | Mouse epiblast stem cells            |
| GI              | Gastrointestinal                       | mg              | Milligrams                           |
| GO              | Gene ontology                          | min             | Minutes                              |
| <b><u>H</u></b> |                                        | ml              | Millilitre                           |
| hESCs           | Humans embryonic stem cells            | mm              | Millimetre                           |
| HGPS            | Hutchinson-Gilford progeria syndrome   | MM              | Multiple myeloma                     |
| HPE             | Holoprosencephaly                      | MNs             | Motor neurons                        |
| HPLC            | High performance liquid chromatography | <b><u>N</u></b> |                                      |
| hr              | Hours                                  | ncHARs          | Non-coding human accelerated regions |
| HS              | Hypersensitive sites                   | ng              | Nanograms                            |
| HTS             | Hass-type polysyndactly                | NPCs            | Neural progenitor cells              |
| <b><u>I</u></b> |                                        | NTR             | Nick translation reaction            |
| IFN $\gamma$    | Interferon gamma                       | NVPs            | Nodal vesicular parcels              |
| INM             | Inner nuclear membrane                 | <b><u>O</u></b> |                                      |
| IRES            | Internal ribosome entry site           | OD              | Optical density                      |
| <b><u>K</u></b> |                                        | ONM             | Outer nuclear membrane               |
| Kb              | Kilobases                              | <b><u>P</u></b> |                                      |
| kDa             | Kilodalton                             | PBS             | Phosphate buffered saline            |
| <b><u>L</u></b> |                                        | PCR             | Polymerase chain reaction            |
| L               | Litre                                  | PFA             | Paraformaldehyde                     |
| LAD             | Lamin associated domain                | PG              | Hox paralogous groups                |
| LASs            | Lamina associated sequences            | Pmol            | Picomols                             |
| LEF             | Loop extruding factor                  | PPD2            | Preaxial polydactyly type 2          |
| LHED            | Local Hopping Enhancer Detection       | PRDs            | Positive regulatory domains          |
| LPM             | Lateral plate mesoderm                 | PRS             | Pierre robin sequence                |
| LPSs            | Large punctuate structures             | P/S             | Penicillin-streptomycin              |
| LSS             | Laurin-Sandrow syndrome                | <b><u>Q</u></b> |                                      |
| <b><u>M</u></b> |                                        | Q               | Phred quality score                  |
| M               | mol/dm <sup>3</sup>                    | qPCR            | Quantitative real-time PCR           |
| MB              | Megabases                              | <b><u>R</u></b> |                                      |
| <b><u>M</u></b> |                                        | RA              | Retinoic acid                        |
| <b><u>M</u></b> |                                        | RNA             | Ribonucleic acid                     |
| <b><u>M</u></b> |                                        | RNAi            | RNA interference                     |

|                 |                                                                      |                 |                                                 |
|-----------------|----------------------------------------------------------------------|-----------------|-------------------------------------------------|
| rRNA            | Ribosomal RNA                                                        | U               | Units                                           |
| rpm             | Revolutions per minute                                               |                 |                                                 |
| RT              | Reverse transcriptase/real time/room temperature (context dependent) |                 |                                                 |
|                 |                                                                      | <b><u>V</u></b> |                                                 |
| <b><u>S</u></b> |                                                                      | VPA             | Valporic acid                                   |
| SAXs            | Small angle X-ray scattering                                         | <b><u>W</u></b> |                                                 |
| SD              | Segmental duplications                                               | W               | Watts                                           |
| SDS             | Sodium dodecyl sulfate                                               | WMS             | Werner mesomelic syndrome                       |
| Seq             | Sequencing                                                           | WT              | Wild type                                       |
| ShRNAi          | Short hairpin RNAi                                                   |                 |                                                 |
| SNP             | Single nucleotide polymorphism                                       |                 |                                                 |
| SSC             | Saline-Sodium Citrate                                                |                 |                                                 |
| STARR           | Self-transcribing active regulatory region sequencing                | <b><u>Z</u></b> |                                                 |
| STD             | Salmon testis DNA                                                    | ZPA             | Zone of Polarising Activity                     |
|                 |                                                                      | ZRS             | Zone of Polarising Activity Regulatory Sequence |
| <b><u>T</u></b> |                                                                      |                 |                                                 |
| T               | Thymine                                                              |                 |                                                 |
| TAD             | Topological associated domain                                        |                 |                                                 |
| TAE             | Tris-acetate- EDTA                                                   | 3C              | Chromosome conformation capture                 |
| TALEN           | Transcription activator-like effector nucleases                      | 4C              | Circular chromosome conformation capture        |
| TE              | Tris-EDTA                                                            | 5C              | Chromatin conformation capture carbon copy      |
| THPSTPT         | Tibial hypoplasia-polysyndactyly-triphalangeal thumb                 | 2D              | Two dimensional                                 |
| TSA             | Trichostatin A                                                       | 3D              | Three dimensional                               |
|                 |                                                                      | μl              | Microlitre                                      |
| <b><u>U</u></b> |                                                                      | μm              | Micrometre                                      |
|                 |                                                                      | μM              | Micromolar                                      |

# Chapter 1

---

## Introduction

# 1.1 Hedgehog Signalling

## 1.1.1 The Hedgehog Protein; Discovery of the Hedgehog Protein and its Processing in the Nucleus

The Hedgehog group proteins are currently regarded as one of the most important classes of signalling molecules present in both vertebrates and invertebrates. Although significant leaps in understanding of both their mechanism of action and role during development have been made, our knowledge of this group still remains in its relative infancy. The hedgehog mutation was first observed in the Nobel Prize winning experiment conducted in 1980 by Nüsslein-Volhard and Wieschaus (Nüsslein-Volhard and Wieschaus, 1980) where a large scale mutagenesis screen was conducted in an attempt to identify the genes responsible for regulating the segmental pattern of the *Drosophila* embryo. One of the resulting mutants displayed denticle structures across the entirety of each independent body segment, thereby giving it a prickly appearance which the author's believed resembled that of a hedgehog's spikes and hence led to the identifying name *hedgehog*. Several other mutants such as *gooseberry* and *patch* were also identified, with a common feature being the deletion of a posterior region of each segment and its replacement with the mirror image duplication of the anterior portion of the segment. Together these were referred to as the segment polarity mutants. Two other classes of mutants were also characterised from this screen: the pair-rule mutants such as *even-skipped* and *runt*, where only alternate segments were affected; and, the Gap mutants such as *Krüppel* and *knirps* where substantial numbers of segments were simultaneously affected. Together these mutants defined a class of genes that would later be recognised as instrumental in patterning both the vertebrate and invertebrate embryo.

However, after its initial discovery in the *Drosophila*, it was over a decade before variants were discovered in mammals. In 1993 three papers (Krauss et al., 1993, Riddle et al., 1993, Echelard et al., 1993) were published in the same volume of the journal *Cell* detailing the discovery of three mammalian genes homologous to the hedgehog gene from *Drosophila*: *Sonic hedgehog* (*Shh*); *Indian Hedgehog* (*Ihh*); and, *Desert Hedgehog* (*Dhh*). In addition, zebrafish also have the additional hedgehog genes *tiggywinkle*, *echidna* and *qiqihar* hedgehog. Of these, *Shh* became the most studied probably due to its importance in regulating key developmental processes at specifically defined times during development of the vertebrate embryo. Some of the most important developmental roles of *Shh* are discussed in the following sections.

Over subsequent years the structure, function and mechanism of the hedgehog proteins were studied in greater detail at both a molecular and cellular levels. It is now known that *Drosophila* hedgehog protein exists initially as a 45 kDa pro-protein which undergoes an auto-cleavage reaction (Lee et al., 1994) using the Hog domain at the carboxyl terminal to produce both an amino-terminal (HH-N) and carboxyl terminal (HH-C) polypeptide, with only the HH-N polypeptide involved in the further signalling processes. This occurs through the production of a thioester intermediate which subsequently undergoes nucleophilic attack on the thioester carbonyl group by cholesterol. The resulting HH-N species formed is therefore a cholesterol adduct (Porter et al., 1996). Using ESI-MS and HPLC, Pepinsky *et al.* (Pepinsky et al., 1998) later showed that human SHH - expressed in both insect and mammalian cell lines - was palmitoylated on the  $\alpha$ -amino group on Cys-24 of the HH-N polypeptide, in addition to cholesterol. However, this did not occur on every peptide, with the authors concluding that both cholesterol only (HH-N) and cholesterol with palmitoyl (HH-Np) polypeptides exist. The identity of the protein responsible for this palmitoyl addition was published concomitantly by several different research groups and can therefore be identified by the numerous synonyms: *skinny hedgehog (ski)* (Chamoun et al., 2001); *sightless* (Lee and Treisman, 2001); *central missing* (Amanai and Jiang, 2001); or, *rasp* (Micchelli et al., 2002). In humans this palmitoyltransferase, which catalyses the addition of a palmitate group to SHH, is known as Hedgehog acyltransferase (HHAT). These proteins belong to a class of enzymes collectively referred to as the membrane-bound O-acyl transferases (MBOAT). Buglino and Mesh (Buglino and Resh, 2008) showed that a palmitate group, most likely provided by a palmitoyl-CoA substrate, was added to the N-terminus of SHH through an amide linkage and that the addition of this group was independent of the previous auto-cleavage reactions and the cholesterol addition. They also performed immunofluorescence experiments to show that the HHAT protein was localised in both the endoplasmic reticulum (ER) and golgi apparatus suggesting that the palmitoylation of SHH occurs in the secretory pathway.

The exit of this cholesterol modified form of hedgehog from the cell is orchestrated by the 12 pass transmembrane protein Dispatched (Disp) which shares a large amount of sequence homology to the hedgehog target protein Patched (Ptc). In studies of the wing imaginal disc (Burke et al., 1999), *disp*<sup>-/-</sup> cells retain the cholesterol adduct within the cell and are unable to release it for long range signalling processes. This contrasts with wild type cells which can effectively spread the hedgehog signal to their neighbouring cells. Burke *et al.* (Burke et al., 1999) highlighted that the presence of cholesterol is critical to the release of this auto-

processed, post-translationally modified protein. When cholesterol was swapped with an alternative lipid anchor Disp was unable to remove this protein from the cell.

The addition of cholesterol to hedgehog proteins seems to have a key role in regulating the location of hedgehog signalling. Gallet *et al.* (Gallet et al., 2006) showed that the cholesterol addition in *Drosophila* cells was essential in targeting the modified hedgehog protein to the plasma membrane. After release from the cell cholesterol addition was also needed for the formation of large hedgehog multimers necessary for the transport of hedgehog as large punctuate structures (LPSs) to long range targets. For example long-range hedgehog signalling was shown in *Drosophila* in both the ventral ectoderm and embryonic dorsal epidermis (Gallet et al., 2006). Without cholesterol, hedgehog cannot form dimers and shows no long range activity.

The hedgehog cholesterol addition and the protein Dispatched were shown to play a key role in localising Hh-Np to the apical membrane (Gallet et al., 2003) in cells of the *Drosophila* ectoderm. Without cholesterol, hedgehog is unable to form LPSs within the cell and is located only at the basolateral surface. Likewise, in *disp* mutants LPSs were not formed and Hh-Np was only present at the basolateral surface in small quantities.

Research continues into the removal of hedgehog and its subsequent transport to other cells situated both close by and at longer distances. What has become clear is that lipoproteins in both invertebrates and vertebrates are necessary for long range Hh/Shh signalling processes (Palm et al., 2013). In *Drosophila* the lipoprotein lipophorin is necessary for long range Hh signalling in imaginal discs (Panáková et al., 2005) and this is mediated by the binding of lipophorin to glypicans such as Dally and Dally-like through both GPI anchors and herapan sulfate moieties (Eugster et al., 2007). A similar process occurs in vertebrate SHH signalling. However, this is only one of a variety of mechanisms leading to the eviction of HH from the cell, with evidence for HH exiting as multimers and through the recently discovered shedding mechanism (Ohlig et al., 2011). Also, HH can leave in vesicles specifically referred to as NVPs (Nodal Vesicular Parcels). In this case, SHH and retinoic acid (RA) are transported within NVPs in the nodal flow and, therefore, contribute to left-right asymmetry in the developing vertebrate embryo (Tanaka et al., 2005). In vertebrates, the protein SCUBE2, an ortholog of the zebrafish protein *you* has been shown to assist in the secretion of SHH from the cell. SCUBE2 binds to the cholesterol on SHH thereby maintaining it in a soluble state enabling easy release into the extracellular matrix (Tukachinsky et al., 2012, Creanga et al.,

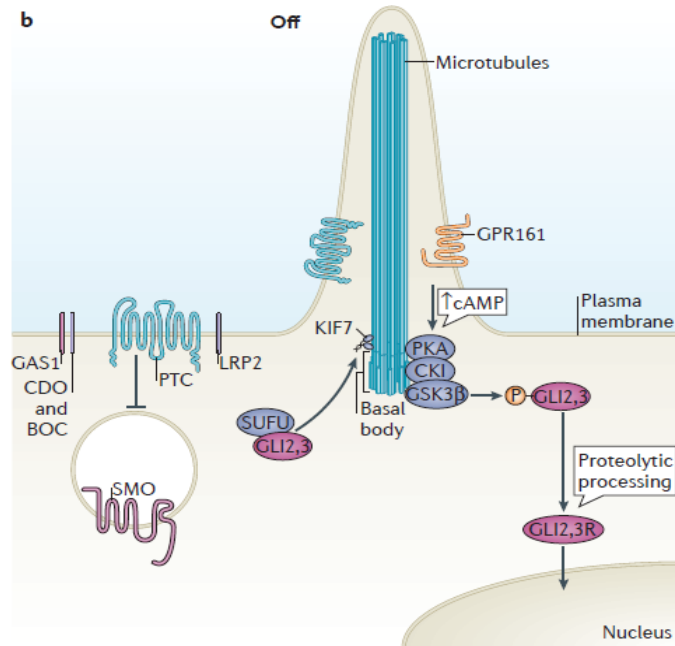


2012). It is hypothesised that SHH is passed from Dispatched to SCUBE2 and then released from the cell where it then forms multimers for long range transport.

### 1.1.2 The Hedgehog Signalling Pathway

SHH signalling has been tightly conserved throughout evolution utilising very similar mechanisms from *Drosophila* to mammals. The main difference however, is that all SHH signalling is conducted in the primary cilium in mammals whereas this structure is not needed for HH signalling in *Drosophila*. In mammals SHH binds to the patched receptor (PTCH1) which is in complex with associated co-receptors GAS1, CDO and BOC thereby nullifying its inhibition of smoothened (SMO). Entry into the cilium is facilitated by the binding of phosphorylated SMO to  $\beta$ -arrestin and KIF3A. Release of SMO inhibition results in the anterograde movement of KIF7 from the base to the tip of the cilium, which may act to release GLI3 from SUFU allowing GLI3-A to enter the nucleus. Loss of SHH signalling maintains SMO in an inhibitory state which maintains inhibition of GLI3 by SUFU. KIF7 remains at the base of the cilia in complex with GLI proteins where, after phosphorylation and  $\beta$ -TRCP recognition, they are partially degraded into their repressor form (GLI3-R) (Anderson et al., 2012, Briscoe and Théron, 2013). Cos2, the *Drosophila* homolog of KIF7, has lost its kinesin motor activity thereby relinquishing the need for primary cilia in Hedgehog signalling. The mammalian SHH signalling pathway is depicted in Figure 1.1.

(A)



(B)

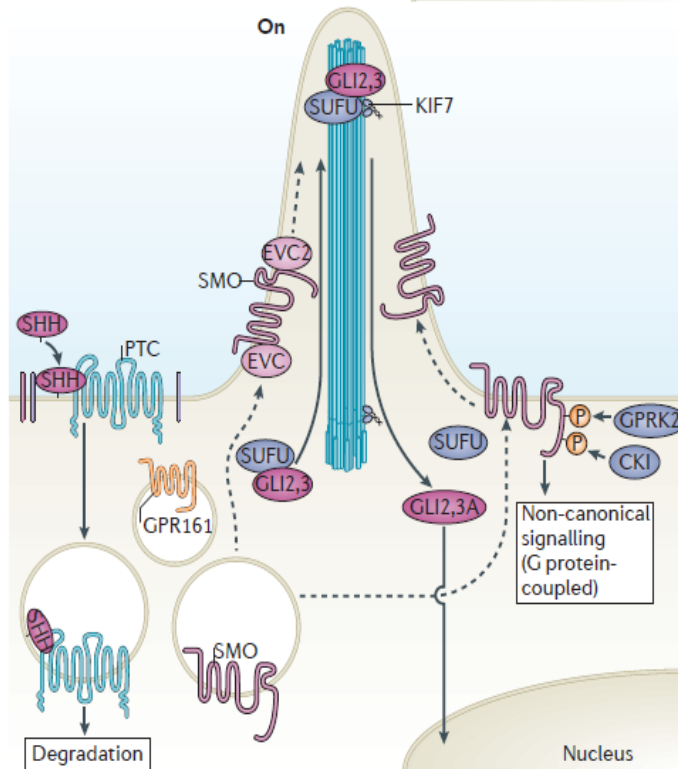


Figure 1.1 – The SHH signalling pathway in mammals in (A) non-expressing and (B) expressing cells. (A) In non-expressing cells patched represses smoothed. KIF7 remains at the base of cilia, leaving GLI3 to be degraded into its repressor form. (B) In *Shh* expressing cells patched loses its inhibitory effect on smoothed and moves into the cilia via β-arrestin and KIF3A. KIF7 travels up the cilia allowing the release of GLI3 from SUFU which can then travel to the nucleus in its activator form. Figure from (Briscoe and Théron, 2013).

## 1.2 Signalling in the limb bud

The developing mouse embryonic limb bud is an ideal model system to study the mechanisms by which morphogens act. The limb is patterned along three separate axes: the proximal-distal axis; the anterior-posterior axis; and, the dorsal-ventral axis. This results in three separate limb sections; the most proximal of which is referred to as the stylopod and (when referring to the mouse forelimb) contains the humerus. The zeugopod follows - containing the radius and ulna - and finally, the autopod contains the carpals, metacarpals and phalanges (Figure 1.3(A)). The length of time active, the distance travelled within the limb bud and the amounts of signalling molecules present are all equally important factors in the formation of a fully functional limb. In addition, an intricate cross talk is necessary between different signalling molecules involved in patterning the three limb axes. This section will discuss some of the key signalling processes which occur within the embryonic mouse limb bud with a special focus on the secreted morphogen SHH which plays a key role in establishment of the anterior-posterior limb axis.

### 1.2.1 Stages of Limb Development

The majority of mouse limb formation occurs over five days from E9.5 to E14.5 (Figure 1.2(A)). During this period a small limb bud emerges from the side of the body trunk and grows outwards; the proximal skeletal elements developing first followed by the more distal elements at later stages. The process of forelimb and hindlimb formation is very similar with hindlimb development trailing that of the forelimb by roughly half a day (Martin, 1990).

The forelimb bud is first visible at E9.5 occurring adjacent to somites 7-13 with the hindlimb apparent at E10 beside somites 27-31. However, it is not until E11.5 that the beginning of cartilaginous structures can be detected in the most proximal regions of the forelimb. A day later the humerus, radius and ulna are all visible along with digits 3, 4 and 5. The apical ectodermal ridge (AER) - a structure important in regulating the signalling processes within the limb bud (Chapter 1.2.3) found at the growing tip - is also fully visible by E11.5, with a small thickening of ectodermal cells visible a day earlier. An invasion of myogenic cells from the dermomyotome, which will later develop into muscle, migrate towards the base of the limb bud at E10.5 with muscle formation evident the following day. Around that time, nerve cells infiltrate the limb and begin to spread distally. By E14 all digits are clearly visible and cells are beginning to die in the mesenchyme separating these structures. The nerves have spread to the fingers and toes and the tendons can be seen. Finally, by E14.5 the interdigital

mesenchyme has disappeared leaving a full complement of digits and toes (although these still lack the most distal phalangeal structures) (Martin, 1990).

## 1.2.2 Initiation of the Limb Bud

The limb bud initially forms as outgrowth of the lateral plate mesoderm (LPM) in mice and in chicks. Development of the three limb axes is dependent on the production of *Fgf10*, a signalling molecule needed to initiate the *Fgf-Shh* feedback loop which defines limb growth (Chapter 1.2.4). The expression of *Fgf10* is driven by retinoic acid (RA) expressed in both the paraxial and lateral plate mesoderm. The process is initiated by an as yet unknown molecule originating in the paraxial mesoderm activating RA and leading to *Tbx5* (*Tbx4* in the hind limb) (Figure 1.2(B)) expression in the LPM of the forelimb (Nishimoto et al., 2015). In chick limb buds, a foil barrier placed between the paraxial mesoderm and the LPM between stages 12 and 13 resulted in a loss of *Fgf10* expression in the LPM but *Tbx5* levels either side of the barrier were unaffected. Expression of *Fgf10* was rescued when a RA soaked bead was placed on the distal side of the barrier. The conclusions from these experiments were that RA from the LPM was initially needed to activate *Tbx5/4* (in accordance with the *Hox* and  $\beta$ -*catenin/LEF/TCF* genes) in a process referred to as limb induction. At later stages, *Tbx5/4* then acts in tandem with RA to activate *Fgf10* in a process referred to as limb initiation (Nishimoto et al., 2015).

The positions at which limb buds develop along the body axis are also tightly controlled. The T-box transcription factors mark these areas with *Tbx5* expressed in the forelimb LPM and *Tbx4*, *Pitx1* and *Islet1* marking the hindlimb. Their expression is controlled through the collinear activation of *Hox* genes in a progressively rostral to caudal orientation. *Hox* paralogous groups (PG) 4 and 5 expressed in the forelimb LPM activate *Tbx5* while *Hoxc8*, 9 and 10 expressed in more caudal regions repress *Tbx5*. In contrast, *Hoxc9* is a positive regulator of *Pitx1* involved in hindlimb development (Nishimoto and Logan, 2016).

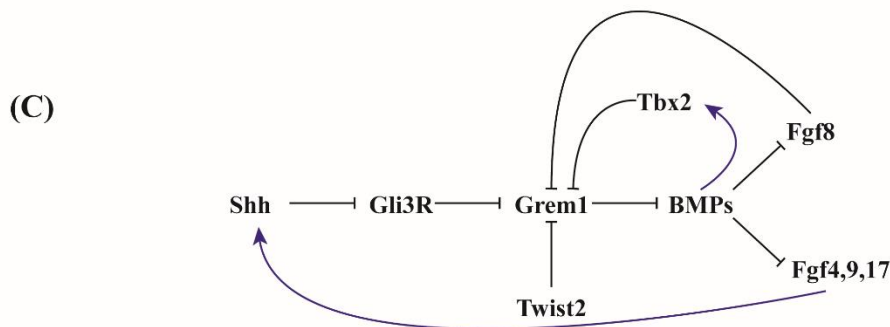
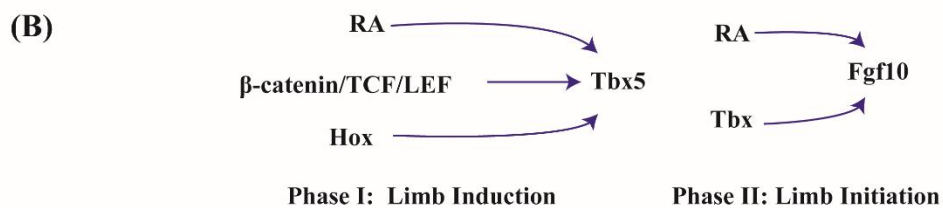
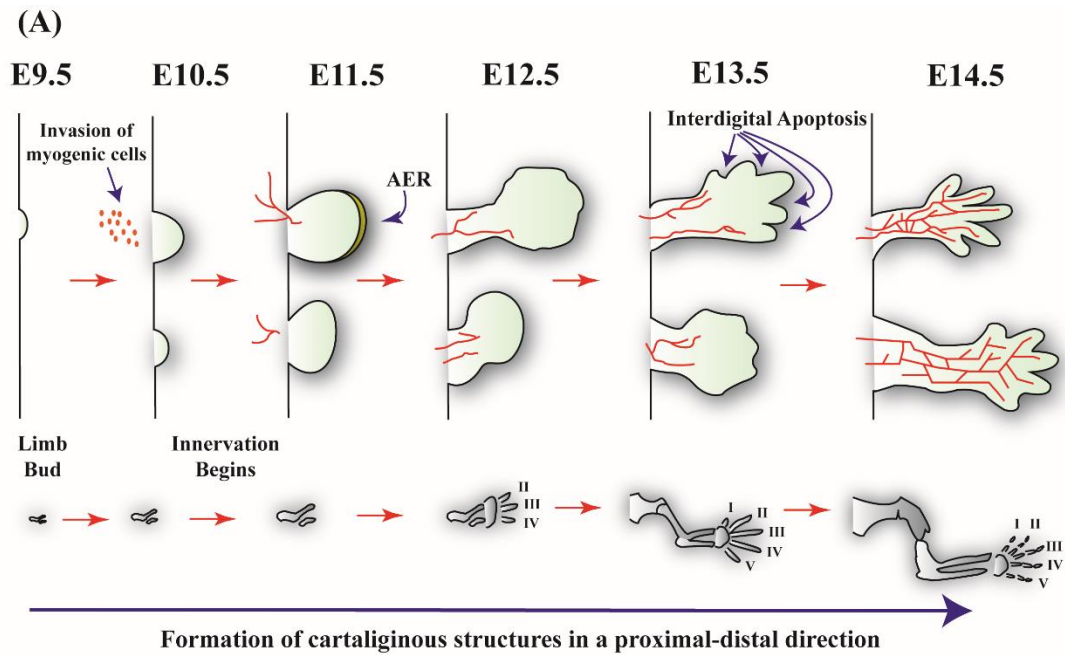


Figure 1.2 – The major stages of limb development occur over 5 days from E9.5 to E14.5 during which a series of complex interconnecting signalling pathways are established leading to the growth of the limb along three axes. (A) The forelimb bud is visible at E9.5 and has substantially increased in size a day later. At this point, myogenic progenitor cells begin to invade the limb bud to initiate muscle formation. Nerves can be seen invading the forelimb at E11.5 - a process referred to as innervation (nerves shown as red lines) - and grow in a proximal to distal direction. By E12.5 the beginnings of digits II, III and IV are

visible while digits I and V appear a day later. Interdigital mesenchyme is removed by apoptosis at E13.5 leaving fully formed digits at E14.5. The formation of the cartilaginous structures of the forelimb are shown along the bottom of this diagram. (B) The key signalling processes involved in both limb induction and limb initiation. Limb induction occurs first resulting with the production of TBX5. Together with RA this acts to induce *Fgf10* which is subsequently involved in a self regulating feedback loop with *Fgf8* resulting in limb outgrowth. (C) The signalling pathways involving the *Fgfs* produced in the AER and *Shh* produced in the ZPA are tightly connected.

### 1.2.3 Formation of the Proximal-Distal Axis

The proximal-distal limb axis of the developing limb bud runs from the shoulder to the tips of the fingers. Elongation of the limb bud along this axis is driven by the apical ectodermal ridge (AER), a thickening of ectoderm at the tip of the growing limb bud. Various models have been proposed to explain the mechanism by which the limb is specified along this axis leading to the formation of the three limb sections. Initially the “progress zone” model was favoured in which the AER maintains a small group of cells in an undifferentiated state at the tip of the developing limb - the so-called progress zone. As the limb grows and cells divide, some cells are displaced from this region allowing them to develop into a particular structure. Cells which leave at an early stage develop into more proximal structures while those which are maintained within the progress zone for a longer time period form the distal features. Experiments in support of this theory were largely based on the effect of removing the AER on limb development. For example, when the AER of chick limb buds was removed at an early stage only proximal structures were formed with a complete lack of distal features. On the other hand, when the AER was removed at later stages distal structures had also developed (Summerbell, 1974, Summerbell et al., 1973). It was proposed that a signal from the AER was responsible for maintaining the progress zone but had no role in specification of the tissues. Instead, the time at which cells were evicted from the progress zone was critical in determining the anatomical feature which developed. Evidence for this hypothesis came from experiments where the AERs from different chick limbs at different developmental stages were removed, switched and re-grafted (Rubin and Saunders, 1972). For the most part, these grafts had little influence on the specification of tissues along the proximo-distal axis. This result contrasts another model where different signals emitted from the AER pattern the limb at progressive developmental stages. It was nearly 20 years after the “progress zone” model was hypothesised before signalling molecules capable of replicating the function of the AER were identified. In two separate studies FGF-4 and FGF-2 were both shown to drive normal proximo-distal limb outgrowth in chick limbs when implanted at the distal edge of the limb after removal of the AER (Niswander et al., 1993, Fallon et al., 1994).

In the following years the importance of AER-FGFs to proximo-distal limb architecture was confirmed (Mariani et al., 2008). Whilst, the “progress zone” model - which argued that cells were specified after they leave the progress zone – continued to be investigated, increasing evidence suggested that limb cells were specified at a very early stage after which the three limb sections expand sequentially in a proximal to distal direction, the “early specification”

model (Dudley et al., 2002). The “two-signal” model (Zeller et al., 2009) incorporates the opposing functions of retinoic acid (RA) produced in the proximal regions and the FGFs produced in the progress zone. RA activates *Meis1* and *Meis2* thereby defining the proximal portions of the developing limb. These signals are inhibited by FGFs within the progress zone. As the limb grows the influence of RA decreases distally eventually being completely inhibited by AER-FGFs. At this point the zeugopod becomes specified (Mercader et al., 2000). The “two-signal” model supports the hypothesis that the limb bud cells may be specified early as the initial limb bud is immediately exposed to both RA and FGFs which influence its development.

The “differentiation front” model incorporates aspects of both the “two-signal” and “progress zone” models to interpret proximodistal patterning (Tabin and Wolpert, 2007). In this theory, the AER-FGFs in the distal limb produce an area of undifferentiated cells - the undifferentiated zone - with its border referred to as the differentiation front. As the cells cross the differentiation front, the influence of the AER-FGFs is lost and they become committed to differentiation. The most proximal cells of the developing limb will pass through the differentiation front first, followed by progressively more distal cells and hence the limb is patterned in a proximal to distal fashion. However, this model differs from the “progress zone” model as the cells in the undifferentiated zone do not rely on an internal clock within this region to relay positional information. Instead, this information is attained by dynamic gene expression patterns that arise as the limb expands. Like the “two-signal model”, the AER-FGFs and RA induced *Meis* gene expression form overlapping regions which eventually separate as the limb expands and leave a middle region unaffected by both signals. *Hoxa11* begins to dominate the distal limb and forms a boundary to the *Meis* expressing regions as both genes are probably inhibitory of each other. *Hoxa13* then becomes expressed in the distal most cells and begins to dominate the region as *Hoxa11* cells progress through the differentiation front. Now the zeugopod is identified exclusively through *Hoxa11* expression in comparison to the autopod expressing *Hoxa13*, while the posterior stylopod is characterised by *Meis* gene expression (Tabin and Wolpert, 2007).



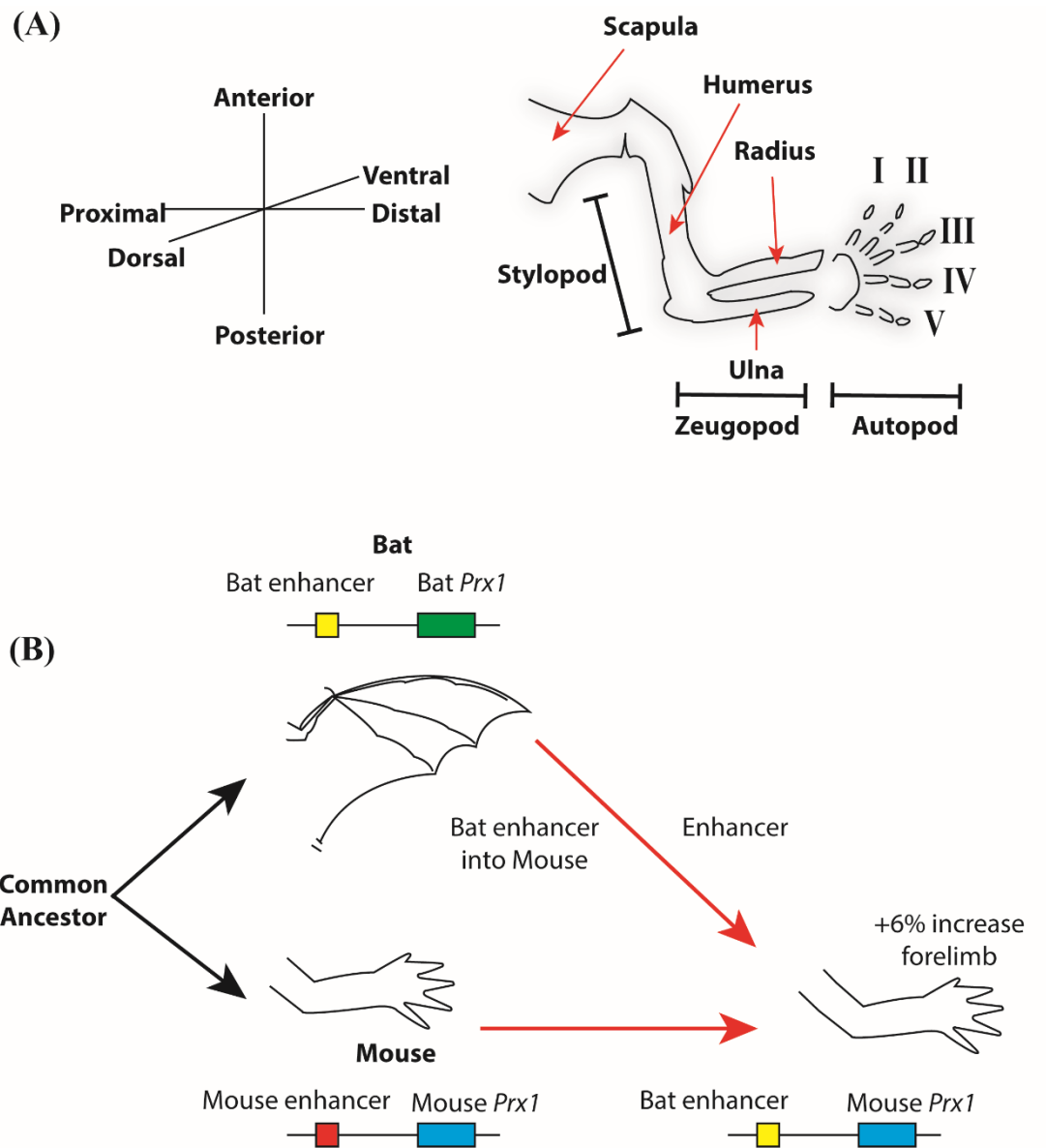


Figure 1.3 – (A) The limb is patterned along three separate axes; the proximo-distal axis running from the shoulder to the tips of the digits, the anterior-posterior axis which runs from the thumb to the little finger and the dorso-ventral axis which runs from the back of the hand to the palm. The bones from the resulting fully grown limb are partitioned into three discrete sections. For the forelimb, the stylopod contains the humerus, the zeugopod contains the radius and ulna and the autopod contains all the bones of the hand plate as well as the phalanges making up the digits. A similar pattern is observed for the hindlimb. This diagram only highlights the various features of the forelimb. (B) The common lab mouse and fruit bat are both descended from a common ancestor in existence roughly 100

million years ago. Both have evolved very different limb structures which can partly be explained through the divergence of enhancer sequences from this common ancestor. Enhancers for the *Prx1* gene are an example of this evolutionary process. When the *Prx1* mouse enhancer was swapped with the *Prx1* bat enhancer mouse forelimbs increased in size.

#### 1.2.4 Formation of the Anterior-Posterior Axis

Prior to *Shh* expression the limb is initially pre-patterned based on the antagonistic relationship between the two transcription factors GLI3-R and dHAND along the anterior-posterior axis. GLI3-R inhibits dHAND expression in the anterior portion of the limb bud while promoting the expression of other factors such as *Alx4* that define the anterior identity. In contrast, dHAND inhibits the expression of *Gli3-R* and other anteriorly expressed genes from spreading to posterior regions and contributes to the expression of posteriorly expressed genes such as *Shh* (te Welscher et al., 2002). Expression of *Shh* within posterior regions thereby establishes a new signalling centre referred to as the Zone of Polarising Activity (ZPA). The ZPA was initially identified through grafting experiments using chick limb buds where the posterior portion of the limb bud from one animal was fused to the anterior portion of another resulting in mirror image duplications of the digits. SHH was confirmed as the molecule driving this polarizing activity within the ZPA; that is when SHH expressing cells were implanted into the anterior regions of chick limb buds, phenotypes similar to the ZPA grafts were produced (Riddle et al., 1993). The conclusion from these experiments was that SHH was essential - not only for specifying the number of digits produced but also the positional identity of these digits. However, what remained unknown was the mechanism by which SHH acted within the limb bud to achieve proper digit formation.

It was immediately hypothesised that SHH may act as a morphogen within the limb bud. Morphogens are small signalling molecules capable of regulating the positional identity of cells within tissues based on the quantity of signal exposed to each cell (Crick, 1970, Wolpert, 1969). Defined signal levels within a morphogen gradient will specify cell fates, with for example, a higher concentration of morphogen eliciting a different effect than that of a lower concentration.

In this case SHH produced within the ZPA would diffuse throughout the limb bud in a posterior to anterior direction and the identity of each digit would be defined by the concentration of SHH at that particular location. The highest concentration of SHH would lead to the formation of the little finger (digit 5 in mice) and the lowest concentration to the thumb (digit 1 in mice). Experiments using irradiated ZPA grafts prior to the discovery of SHH, had suggested that a diffusible morphogen specifies the digits. For example, Smith *et al.* (Smith et al., 1978) showed that when chick ZPA cells were subjected to an increasing dosage of radiation and then grafted onto the anterior of another limb bud, the ability to form

digits 4 to 2 was inversely proportional to the amount of radiation. The conclusion from these experiments was that increasing the radiation directly affected the quantity of a potential signal from the grafted cells thereby reducing its ability to form digits needing high signal levels. Further experiments using irradiated ZPA grafts (Smith, 1980) confirmed that the time at which duplicated digits develop agrees with a morphogen model for limb patterning.

However, this morphogenic model of anterior-posterior patterning within the limb bud was revised in a report by Harfe *et al.* (Harfe *et al.*, 2004) that proposed that digit formation was dependent on both a spatial and temporal gradient of SHH. This study highlighted the observation that digits 3 to 5 are descended from SHH expressing cells while digit 2 is not - although it is still dependent on SHH to form. The authors suggested that as cells in the posterior region expand, some move away from the SHH expressing domain of the ZPA. This is a sequential process, where the length of time cells are exposed to SHH defines whether they will develop into digits 3, 4 or 5. Digit 2 is dependent on the diffusion of the SHH morphogen anteriorly, as the cells which produce this digit have never been part of the ZPA.

In contrast to the spread of SHH in a posterior to anterior direction, GATA6 is shown to have a higher concentration in the anterior hindlimb which decreases posteriorly. Deletion of GATA6 resulted in the spread of *Shh* expression anteriorly, while increased GATA6 resulted in a drop in *Shh* expression. In subsequent experiments, the *Shh* limb enhancer, the ZRS (Chapter 1.3.8), was shown to possess binding sites for GATA6. When a ZRS-lacZ construct was used in transgenic experiments a combination of GATA6 and FOG2 reduced its activity. The conclusions from this study were therefore that GATA6 binding at the ZRS inhibited *Shh* expression thereby restricting it from spreading ectopically (Kozhemyakina *et al.*, 2014).

### **1.2.5 Communication Between the ZPA and the AER is Essential for Limb Development**

Although development along the three limb axes are usually analysed individually, it is important to be aware of a complex cross-talk between the individual components of each pathway. For example, an interconnecting relationship between the AER-FGFs and SHH produced in the ZPA is important in both anterior-posterior and proximo-distal axis development. Early studies showed that a combination of FGF4 and RA induces *Shh* expression. SHH and FGF4 then enter into a positive feedback loop, each maintaining the expression of the other without the need for RA (Niswander *et al.*, 1994). Induction of *Fgf4*

expression by SHH was later shown to be mediated through the BMP antagonist gremlin (Figure 1.2(C)). BMP is an inhibitor of *Fgf4* expression but inhibition of BMP by gremlin subsequently allows *Fgf4* expression within the AER (Zúñiga et al., 1999). In the chick limb bud this feedback loop is turned off by E6. At this stage, SHH producing cells and their descendants no longer produce gremlin. At first this is not a problem as the SHH protein is able to diffuse through cells and activate gremlin in the more anterior responsive cells. However, as the pool of gremlin non-expressing cells increases, the SHH protein has further to travel and eventually is unable to reach the responsive tissue. As a result, FGF4 is no longer produced and the SHH-FGF4 feedback loop breaks down (Scherz et al., 2004). In 2008 this model was updated to incorporate the inhibitory function of *Fgf8* on gremlin (Verheyden and Sun, 2008). Here, *Shh* induced gremlin expression leads to the induction of *Fgfs* 4, 8, 9 and 17. At this stage the SHH-FGF4 feedback loop is functional and induction of both genes occurs. However, FGF8 is able to inhibit gremlin and as time progresses levels of FGF8 increase. As a result, gremlin expression is inhibited thus allowing the BMPs to inhibit the AER-*Fgfs*. Loss of the *Fgf4* expression thereby leads to a drop in *Shh* expression and the termination of the self-regulating feedback loop. In addition, both TBX2 (Farin et al., 2013) and TWIST2 (Wade et al., 2012) have been shown to inhibit gremlin; hence both assist to terminate the SHH-FGF4 feedback loop.

The integration of the *Hox* (*HoxA;D*) genes at different developmental stages into the SHH-FGF feedback loop provides an additional layer of complexity to this system. At early stages these can promote the expression of *Fgf10* which subsequently induces *Fgf8*. Also, the *HoxA;D* genes initiate gremlin expression independent of *Shh*. Later on (around E11.5 in mice), these genes promote the spread of gremlin to anterior positions of the limb bud and the maintenance of *Shh* expression. Expression of *Hox* genes at key developmental stages is therefore essential in regulating the development of both the anterior-posterior and proximo-distal axes (Sheth et al., 2013).

## 1.2.6 Formation of the Dorso-Ventral Axis

The dorso-ventral limb axis runs from the back of the hand to the palm and, like the other two axes, requires signals emanating from the AER for proper formation. Among these are the *Wnt* family of signalling molecules. *Wnt7a* is expressed in the dorsal ectoderm and is responsible for regulating the spatiotemporal expression of *Lmx1* in the dorsal mesenchyme. Misexpression of *Lmx1* ventrally causes a “Double Dorsal” limb pattern (Riddle et al., 1995). Interestingly, *Wnt7a* signals through a  $\beta$ -catenin independent pathway in contrast to *Wnt3a* which functions in both establishing the AER as well as regulating the expression of key genes such as *Bmp2*, *Fgf4* and *Fgf8* (Kengaku et al., 1998). However, there is an interconnection between this *Wnt7a* signalling pathway and the previously discussed SHH-FGF4 feedback loop (Chapter 1.2.4). Removal of the dorsal ectoderm and therefore WNT7A results in a loss of *Shh* signalling and the formation of some proximal skeletal elements which can be rescued by the addition of *Shh* expressing cells to the posterior distal region of the limb (Yang and Niswander, 1995). This restriction of *Wnt7a* to the dorsal ectoderm is mediated by *En-1* expression which is expressed in both the ventral ectoderm and mesenchyme and represses the spread of *Wnt7a* to these regions. Ectopic *En-1* in the dorsal ectoderm results in a loss of *Wnt7a* expression and the malformation of the AER. This impacts the SHH-FGF4 feedback loop which leads to a reduction in *Shh* expression and defects in the developing proximodistal skeleton (Logan et al., 1997). A further study looking at the role of BMP signalling in dorso-ventral patterning showed that a conditional knock out of the *Bmpr* gene responsible for the type I BMP receptor (BMPR-IA) resulted in a loss of ventral structures. This correlated with a reduction in *En-1* and an increase in *Wnt7a* and *Lmx1b*. Expression of BMPs such as *Bmp7* occur in ventral regions near the AER at early limb bud development stages. It seems, therefore, that BMPs are responsible for activating *En-1* in the ventral ectoderm which is subsequently needed to pattern the ventral limb (Ahn et al., 2001). The order in which BMPR-IA and  $\beta$ -catenin are utilised in the signalling cascade leading to AER formation remains to be fully elucidated and has been debated in two conflicting reports (Soshnikova et al., 2003, Barrow et al., 2003). However, BMPR-IA is undoubtedly involved in regulating *Fgf8* expression, as continued stimulation of the receptor results in *Fgf8* expression ectopically.

The transcription factors SP6 and SP8 are also a vital component of the signalling cascade resulting in correct dorso-ventral limb patterning. Both factors act in a redundant manner; however, SP8 is the more important of the two due to its higher expression within the developing limb. SP6/SP8 act downstream of  $\beta$ -catenin to regulate the *Fgf8* expression needed

in AER formation and, loss of both SP6/SP8 therefore results in a loss of *Fgf8* expression. As *Fgf10* is in a self-regulating feedback loop with *Fgf8*, *Fgf10* is initially expressed but quickly lost leading to the disappearance of the limb bud at E11.5. SP6/SP8 are also responsible for regulating the correct expression of *En-1*, with mutants displaying *Wnt7a* in ventral regions. BMP signalling in the absence of SP6/SP8 cannot maintain the correct level of *En-1* (Haro et al., 2014).

Dorso-ventral patterning of the limb is therefore reliant on signals within the AER. However, there is some evidence that the early limb is patterned prior to the emergence of this structure. In experiments using quail-chick chimeras, both the mesoderm and ectoderm within and surrounding the presumptive wing were examined. The ectoderm within this region was shown to give rise only to the AER. Dorsal limb ectoderm derived from the ectoderm covering the paraxial and intermediate mesoderm, while ventral limb ectoderm derived from lateral somatopleural ectoderm. Further experiments also suggested that a signal emitted from the somites functions to pattern the dorsal limb, while a signal emitted from the lateral somatopleure patterns the ventral limb. In this way, dorso-ventral patterning is established prior to the formation of the AER (Michaud et al., 1997). One hypothesis is that patterning of the limb bud is initiated by the signals from both the somites and lateral somatopleure but, once the AER is operational it takes over the role of polarising the limb.

### **1.2.7 Evolution of the Tetrapod Limb**

The purpose of this section is to provide a basic compendium of some of the key evolutionary processes leading to the formation of a human limb, with a focus on certain genetic changes that have occurred at different phylogenetic branch points. These events have been chosen as they relate to differences in both anterior-posterior and proximal-distal limb patterning (Chapter 1.2.2-1.2.4) as well as changes in enhancer activity (Chapter 1.2).

As mentioned previously, the mouse limb is composed of the three different sections; the stylopod, zeugopod and autopod (Chapter 1.2). In fact, the same basic limb structure can be applied to all vertebrates of the superclass Tetrapoda. However, the same structure cannot be applied to describe the limbs of all living animals. For instance, within actinopterygii, teleost fish show a clear absence of autopod structures. Instead, these fish have proximal endoskeletal elements referred to as radials which are followed by long fin rays more distally (Wagner and Chiu, 2001). Within sarcopterygians, components of both the stylopod and zeugopod can be distinguished and begin to resemble the human form in correlation with how closely they are

related (Schneider and Shubin, 2013). Striking structural similarities are obvious when comparing to fossils such as those of the *Acanthostega* which display these features in addition to distal bones. It therefore seems reasonable that the evolutionary fin to limb transition occurred through a loss of fin rays and an expansion of endoskeletal structures (Schneider and Shubin, 2013).

Within tetrapods, the structure of mammalian limbs is highly diverse. This is best highlighted by the contrasting structure of both the fruit bat (*Carollia perspicillata*) and mouse (*Mus musculus*) limbs where both animals share a common ancestor from roughly 80-100 million years ago. Bat limbs have become adapted for flight with the forelimb containing digits of increased length separated by wing membrane which is also attached to the hindlimb (Cretekos et al., 2001). The reasons for these differences are hypothesized to be the result of diversified patterning mechanisms between both animals. Indeed, this has been shown in one study where the enhancers responsible for expressing the *Prx1* gene in mice were replaced with the *Prx1* enhancers from bats (Figure 1.3(B)). The resulting mice showed increased digit length of the forelimbs caused by increased *Prx1* expression in the forelimb bones at E17.5 (Cretekos et al., 2008).

The idea that differences in enhancer activity affect gene expression thereby resulting in evolutionary changes to morphological features, is of particular interest. In addition to changes in the activity of enhancers, the emergence of new enhancers and the removal of existing ones is likely to have evolutionary consequences. Indeed, a study looking at changes in the number and location of enhancers in humans, rhesus macaque and mouse limbs at different time points during development identified 2,915 limb specific enhancers unique to humans. The genes associated with these new enhancers were related to processes such as bone morphogenesis and growth suggesting that the presence of these features directed the evolution of human limbs (Cotney et al., 2013).



## 1.3 Sonic Hedgehog

*“...the one gene that my students remember is Sonic hedgehog”*

*Lewis Wolpert (Ingham and Vicente, 2014)*

### 1.3.1 *Shh* related limb abnormalities

Analyses of SHH related limb phenotypes have typically been undertaken with mice and chickens - two model systems which provide a quick and decipherable tool to analyse patterning mechanisms within the developing embryonic limb bud. In mice, *Shh* expression is first detected in the limb bud at embryonic day E9.75 and is completely absent by E12. As mentioned previously, control of *Shh* expression within the ZPA of the limb is maintained through the ZRS, a highly conserved 780 bp sequence. This is situated 800 kb in mice, and 1 Mb in humans, away from the *Shh* gene in intron 5 of the *Lmbr1* gene (previously called *C7orf2* in humans). Many different mutations have been identified within the ZRS leading to a number of limb phenotypes. These are collectively referred to as the “ZRS associated syndromes” (Wieczorek et al., 2010). Several of these mutations give rise to ectopic *Shh* expression with other mutations believed to have a similar effect. In mice, ectopic expression in anterior regions of the limb bud results in the production of extra digits at this location - a condition referred to as pre-axial polydactyly. This trait contrasts with other enhancer associated point mutations which usually lead to a reduction in gene expression. In humans, a spectrum of limb phenotypes has been observed which depends upon the type of ZRS mutation with point mutations, microduplications, inversions and deletions within the *Shh* locus having been identified. As such, the “ZRS associated syndromes” can be categorised into three main categories. Isolated point mutations causing preaxial polydactyly type 2 (PPD2; MIM #174500) are referred to as Type1a syndromes while Type 1b syndromes refer solely to point mutations at position 404 within the ZRS causing Werner Mesomelic Syndrome (WMS) (tibial hypoplasia-polysyndactyly-triphalangeal thumb, THPSTPT; MIM# 188770). Typically these mutations have presented as a G > A or G > C but recently a G > T was also identified (Girisha et al., 2014). Intriguingly, a 402C > T mutation has been identified where heterozygotes have TP and PPD, but homozygotes have WMS (VanderMeer et al., 2014). In addition a 406A > G mutation has also been linked to WMS (Norbnop et al., 2014). Type II syndromes do not involve point mutations but include the duplications which result in

conditions such as Hass-type polysyndactyly (HTS) among others (Wieczorek et al., 2010). Duplications of small regions can have particularly serious effects causing diseases such as Laurin-Sandrow syndrome (LSS) (Lohan et al., 2014).

By 2012, 15 single point mutations had been identified within the human ZRS with each one contributing specifically to limb defects (Anderson et al., 2012). In addition, mutations have been identified in other species such as the so-called “Hemingways’ cats” which display supernumerary digits (Lettice et al., 2008). It was proposed that some of these mutations may alter transcription factor binding sites within the ZRS; either by producing new sites for transcription factors which enhance *Shh* expression, or, by removing sites to which repressor proteins bind. Indeed, this was observed for the ETS transcription factor binding sites within the ZRS. The correct balance of ETS factors binding to the ZRS is important for maintaining *Shh* expression in posterior regions of the limb bud. ETS1 and GABP $\alpha$  are positive regulators of *Shh* expression and have 5 known binding sites within the ZRS while ETV4 and ETV5 repress *Shh* expression and have 2 known ZRS binding sites (Figure 1.4). The so-called “AUS” mutation transforms an ETV4/5 binding site into an ETS/GABP $\alpha$  site thereby resulting in *Shh* expression ectopically (Lettice et al., 2012).

The deletions identified relating to *Shh* limb phenotypes do not incorporate the ZRS itself. However, this does not rule out the possibility that their absence affects the mechanism by which the ZRS enhances *Shh* expression. One of these deletions upstream of the ZRS results in Acheiropodia, a condition defined by the absence of both hands and feet (Ianakiev et al., 2001). Another affects a putative silencer located 240 kb upstream of the gene, the absence of which results in PPD-hypertrichosis (Petit et al., 2016). The authors suggest that this region could be involved in the interactions with other cis-regulatory elements and the *Shh* promoter, with the removal of the silencer possibly affecting the looping structures bringing these regions together. However, this sequence is more likely to act as a silencer as it contains GATA binding sites which is significant as GATA6 has been shown to repress *Shh* activity in the anterior regions of mouse limb buds.

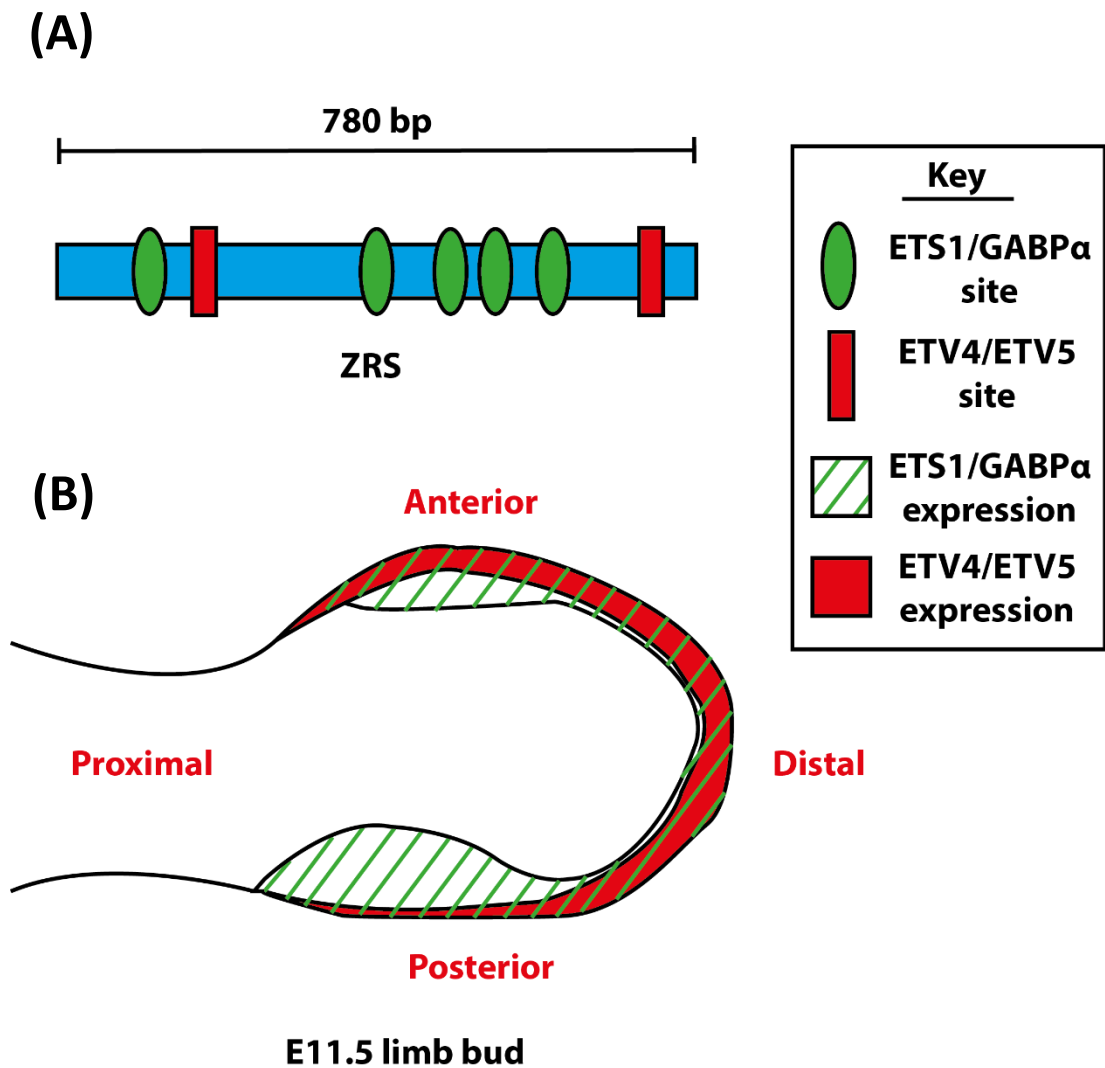


Figure 1.4 – ETS factors control the location of *Shh* expression within the E11.5 limb bud. (A) The ZRS is 780 bp in length and contains five ETS1/GABPα binding sites and two ETV4/ETV5 binding sites. ETS1 and GABPα are positive regulators of *Shh* expression while ETV4 and ETV5 are negative regulators. A greater ratio of ETS1/GABPα: ETV4/ETV5 transcription factors in the posterior region of the limb bud coincides with the expression of *Shh* in this region. In contrast reduced levels of ETS1/GABPα in the anterior regions reflect a lack of *Shh* expression at this location.

### 1.3.2 *Shh* signalling in the Neural Tube

SHH produced in the notochord induces the floor plate of the neural tube to also produce SHH. A gradient of SHH is established with highest concentration at the ventral side of the neural tube and lowest concentration at the dorsal side. As a result, a high concentration of GLIA builds up ventrally and decreases distally, while a high concentration of GLIR builds up distally and decreases ventrally. The ratio of GLIA:GLIR is responsible for determining cell fates along the dorso-ventral axis (Cohen et al., 2013). Cells receiving the highest dose of SHH and therefore GLIA develop floor plate cells, with lower concentrations inducing motor neurons. Cells receiving the lowest concentration of SHH and therefore highest concentrations of GLIR at the ventral neural tube become ventral neurons (Choudhry et al., 2014).

### 1.3.3 *Shh* signalling in Craniofacial Development

SHH plays a key role in orchestrating various aspects of craniofacial development. In humans, *Shh* mutations can cause holoprosencephaly (HPE) characterised by a failure of the two cranial hemispheres to split. Patients can have a range of different phenotypes depending on the severity of the disease, which can range from cyclopia in extreme cases to microcephaly and ocular hypotelorism in more moderate cases (Nanni et al., 1999). *In situ* hybridisation experiments have identified *Shh* expression in craniofacial features as early as E8.5 where it is maintained in the pharyngeal endoderm and diencephalon (Paiva et al., 2010). Its role in pharyngeal arch development was confirmed through the use of *Shh* mutants. In these mice development of the first pharyngeal arch is stunted and becomes fused by E9.5 - possibly as a result of abnormal neural crest cell migration (Yamagishi et al., 2006). From E12 - E13.5 there is evidence for the involvement of *Shh* in the development of the teeth and oral epithelium (Paiva et al., 2010).

At later developmental stages (E16.5 onwards) *Shh* is involved in the formation of sutures within the skull. Expression occurs predominantly in the midline suture mesenchyme with a possible involvement in coronal suture development. Through the induction of *Msx2*, *Shh* is involved in the process of mesenchyme proliferation. This contrasts with the role of *Ihh* which has a proven function in osteoblastogenesis eventually resulting in the fusion of sutures (Pan et al., 2013).

### 1.3.4 *Shh* signalling in the Lungs

In mice, lung organogenesis can be divided into four phases with *Shh* signalling evident at each of these stages: Pseudoglandular stage (E12.5 - E16.6); canalicular stage (E16.5 - E17.5); saccular stage (E17.5 - P0); and, alveolar stage (P0 - P14). SHH is first observed at E10.5 and is clearly localised at the distal end of branchial tubules at E13.5, thereby highlighting its role in orchestrating lung bud branching (Miller et al., 2001, Kugler et al., 2015). This was confirmed by the *Shh* null mouse which showed a complete loss of branching and developed lungs consisting of only a very basic sac structure (Kugler et al., 2015). By the canalicular stage, SHH is detected only in discrete locations such as the respiratory epithelium.

### 1.3.5 *Shh* signalling in the Gut

The gastrointestinal (GI) tract or alimentary canal is heavily reliant on *Shh* signalling for proper development. Indeed, SHH expression is observed at various locations and times within the embryonic gut. In mice, *in situ* hybridisations experiments have identified an uneven distribution of SHH between the forestomach and hindstomach epitheliums at E11.5, with the forestomach showing much higher expression. This patterning persists until E14.5 (Bitgood and McMahon, 1995). As shown by *Shh*<sup>-/-</sup> mutants, *Shh* expression is important in the formation of the tracheoesophageal septum as mutants present with a fusion of the trachea and oesophagus; two tubes which are normally morphologically distinct by E10.5 (Litingtung et al., 1998). Restriction of SHH from the pancreatic epithelium is equally important as experiments in chicks suggest that SHH inhibits the formation of the pancreatic buds (Hebrok, 2003). By E18.5, SHH is present in the epithelium of the stomach, small intestine and colon (Ramalho-Santos et al., 2000).

### 1.3.6 Other regions involving *Shh* signalling

*Shh* signalling is also involved in several other features in mammals including whiskers, hair and the urogenital system (Bitgood and McMahon, 1995).

### 1.3.7 The *Shh* Regulatory Domain

As previously mentioned, the enhancer responsible for regulating *Shh* expression in the limb is known as the Zone of Polarising Activity Regulatory Sequence (ZRS). It is located 1 Mb upstream of *Shh* within intron 5 of the *Lmbr1* gene (Figure 1.5). The region between *Shh* and

*Lmbr1* contains only a single gene (*Rnf32*) and is referred to as a gene desert. This contains several other enhancers which regulate *Shh* expression in different tissues: MACS1, MFCS4 and MRCS1 are each responsible for different aspects of epithelial *Shh* expression (Sagai et al., 2009), SFPE1 and SFPE2 regulate floor plate expression in the spinal cord and hindbrain (Jeong et al., 2006), SBE1 ensures ventral midbrain and diencephalon expression; and SBE2 – SBE4 control forebrain expression.

The development of the Local Hopping Enhancer Detection (LHED) System has proved useful in studying the mechanisms of enhancer function within a locus (Kokubu et al., 2009). This technique involves inserting a transposon connected to a *lacZ* reporter gene into a chromosome. The sleeping beauty transposase enzyme is expressed allowing the transposon to “hop” to a different location within the locus. In this way the activity of different enhancers can be studied by analysis of the *lacZ* expression profiles in different tissues.

Anderson *et al.* (Anderson et al., 2014) have recently utilised this system to study the mechanism of enhancers within the *Shh* locus. They created a range of *lacZ* insertions with four of these mapping to the gene desert between *Shh* and *Rnf32*. When limb bud development was analysed, *lacZ* expression was higher for inserts located close to the ZRS and *Shh* with a decrease in expression in the intervening locations. Interestingly, *Rnf32* - which is not regulated by any of the known *Shh* enhancers - contained an insert in intron 6 which replicated *Shh* expression in the CNS, brain, gut and limb. The authors suggest that regions within this gene are susceptible to the action of the *Shh* enhancers but activation of the *Rnf32* gene itself is prevented - possibly through some protective mechanism within its promoter.

### **1.3.8 The Mechanism of the ZRS**

Research into the mechanism by which the ZRS interacts with *Shh* remains in its infancy. Amano *et al.* (Amano et al., 2009) analysed E10.5 embryos containing a ZRS knockdown. Genes within this region maintained a normal expression pattern while SHH expression was diminished. This highlighted the specificity at which the ZRS controls *Shh* expression. 3D-FISH was also conducted using cells taken from three different regions along the anterior-posterior axis of an E10.5 limb bud. One of these regions contained cells exclusively from the ZPA. A probe specific for the ZRS and another for *Shh* were used to analyse the interaction between both sequences within the three regions. In the ZPA cells, 18 % of probes co-localised compared to the 4 % recorded for cells derived from the middle of the limb bud. However, in anterior cells 11 % of probes co-localised, which was much higher than expected. In the same

investigation Amano *et al.* (Amano et al., 2009) used chromosome conformation capture (3C) to confirm the interaction between the ZRS and *Shh*. They found that in E10.5 embryos there was a specific interaction between the ZRS and *Shh* which was later diminished by E12.5. From this research, a model was proposed suggesting that cells within the ZPA can exist within three different chromosome conformation states. In cells from the middle of the limb bud, the ZRS is situated far from *Shh*. In both anterior and posterior cells, the ZRS and *Shh* are in close proximity. However, the *Shh* locus from ZPA cells loops out of the chromosome territory (CT) leading to *Shh* expression. This is not observed for anterior cells, which contain the ZPA within the CT.

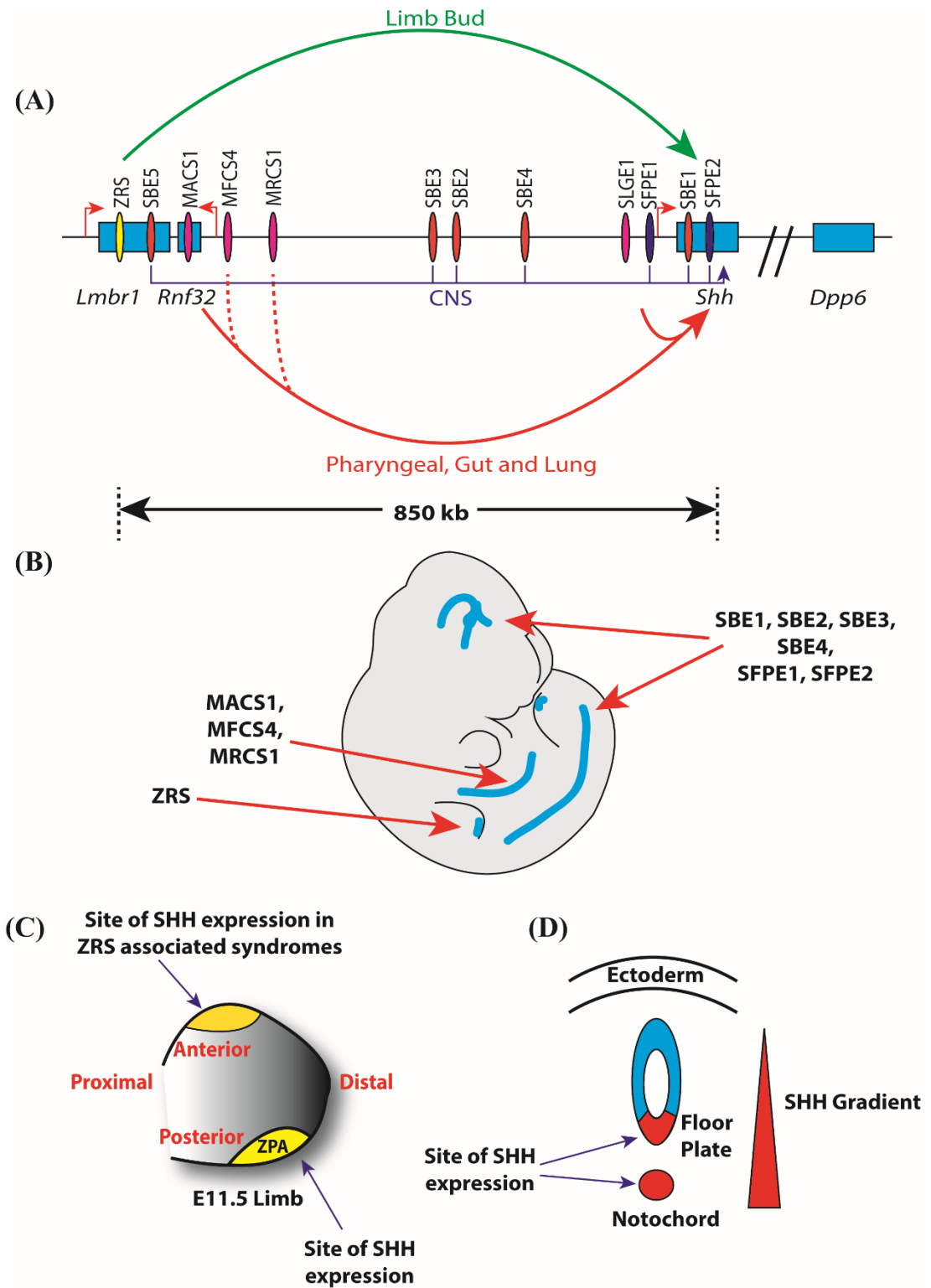


Figure 1.5 - (A) In mice, the distance between the *Shh* gene and the ZRS is roughly 850 kb. The intervening region contains a plethora of different enhancers which regulate *Shh* expression in different spatial and temporal patterns. The ZRS (yellow oval) regulates SHH in the developing limb bud, while other subgroups of enhancers regulate *Shh* expression



throughout the CNS (orange and purple ovals) and in the epithelia of the lungs, gut and pharynx (pink ovals). Genes are depicted as blue boxes, with *Rnf32* the only gene separating *Lmbr1* and *Shh*. *Dpp6* is on the other side of *Shh* from the ZRS at a roughly equidistant position. (B) An E10.5 embryo with *Shh* expression highlighted in blue. The enhancers responsible for *Shh* expression at these locations are indicated. (C)(D) The two most studied locations of *Shh* expression are the developing limb bud and the notochord and neural tube. In the limb bud (C) *Shh* is expressed in the ZPA located in the posterior distal region, while ectopic expression can sometimes be observed in anterior regions leading to a diverse range of phenotypes. *Shh* expression from both the notochord and floor plate of the neural neural tube (D) creates a gradient of *Shh* along the neural tube, highest at the ventral end and lowest at the dorsal end. The concentration of *Shh* at each position along the neural tube regulates the cellular identity of cells along the dorso-ventral axis of the neural tube.

## 1.4 Enhancers

*“I think one of the maybe disappointing things in the sequencing of the human genome was that actually, we didn’t have very many more genes than a worm has and I guess we felt affronted by that. We thought we were a bit more complicated. So, the realisation since that time has actually been, it’s not how many genes you’ve got, but what you do with them, how you control them.”*

*Professor Wendy Bickmore, Director of the MRC Human Genetics Unit, the University of Edinburgh – in an interview from the Naked Scientists, 2013. Available from: <http://www.thenakedscientists.com/HTML/interviews/interview/1000180/>*

### 1.4.1 Regulatory Locus Composition

Prior to sequencing the human genome in 2001, scientists believed that the added complexity of humans over other animals was due to a greater number of protein coding genes. Indeed some predicted that the total number of human genes could exceed 100,000 (Pennisi, 2003). However, by 2014, studies predicted that protein coding genes within humans could total as little as 20,000 or less. It was therefore clear that additional information within the genome must be contributing to human complexity. In the intervening period, the importance of a class of non-coding sequences - referred to as enhancers – was gaining increasing attention. Enhancers are cis regulatory sequences which can bind to transcription factors and enhance gene transcription (Douglas and Hill, 2014). They can operate over large genomic distances with some (such as the ZRS regulating *Shh* limb bud expression) located up to 1 Mb from the target gene. Enhancers can regulate gene expression in several different ways. Some operate individually and regulate the expression of a particular gene at a particular developmental stage. In contrast other enhancers come together in groups with different combinations yielding different expression profiles.

One of the most well studied examples of enhancers acting in combination are the *HoxD* enhancers which form a “regulatory archipelago” (Montavon et al., 2011). This consists of a plethora of individual enhancer sequences spread out over a large area which come together at defined times to regulate transcription of target genes. Enhancers regulating the *HoxD* genes are located across two different topological associated domains (TADs) (Chapter 1.5.1). Early expressed genes are regulated by enhancers within the telomeric TAD which eventually

become inactivated over time. This coincides with an increase in activity of enhancers within the centromeric TAD which activate genes expressed at later stages. In this way the expression of the *HoxD* genes is tightly controlled (Andrey et al., 2013).

Enhancers regulating the expression of *Fgf8* also act in combination and have been referred to as holo-enhancer. These enhancers are located throughout the *Fgf8* locus, each with individual enhancer activities, which combine to produce a defined *Fgf8* expression profile which does not necessarily reflect their independent activities (Marinić et al., 2013). In contrast to the HoxD regulatory archipelago and the *Fgf8* holo-enhancer the *Shh* gene is regulated by individual enhancers which operate at defined times to regulate expression in different tissues (Chapter 1.3.7).

## 1.4.2 Super Enhancers and Stretch Enhancers

In recent years, the term “super enhancer” has been used to describe a subset of highly active enhancers. In the initial study (Whyte et al., 2013), enhancers were identified based on their ability to bind *Oct4*, *Sox2* and *Nanog* simultaneously. For analytical purposes, closely spaced enhancers (within 12.5 kb) were combined and considered as a single unit. Enhancers were then ranked based on the extent of MED1 ChIP-seq binding from lowest to highest, the values scaled from 0 to 1 and plotted on x axis of a graph with the scaled values again plotted on the y axis. This produced a line which an obvious deviation in incline at a defined x coordinate. This region where the tangent of the curve is equal to 1.0 is the division point separating normal enhancers from super enhancers. Only those with an x coordinate corresponding to a tangent  $> 1.0$  were considered as super enhancers.

Therefore, by definition the main difference between super enhancers and normal enhancers is the ability to bind high levels of MED1 above a defined threshold. However, the MED1 cut-off value used to separate enhancer categories is not derived from an experimental finding, but is merely an analytical tool to process the data. Therefore, questions still remain over whether super enhancers are a discrete class of non-coding elements functioning differently from normal enhancers or are simply a more powerful enhancer. Whatever the case, super enhancers seem to play a key role in maintaining the identity of a cell, such as in ESCs where they regulate the expression of factors promoting pluripotency such as *Oct4*, *Sox2* and *Nanog*. Super enhancers in these cells are also enriched in binding sites for KLF4 and ESRRB (Whyte et al., 2013) as well as NRFA2, PRDM14, TCF2L1, SMAD3, STAT3, and TCF3 (Hnisz et al., 2013) which all play a role in maintaining the identity of the cell. It seems therefore that super enhancers are involved in an auto-regulatory feedback loop, where the gene products they activate bind back to the super enhancer thus promoting their own expression.

Super enhancers can play a role in regulating certain diseases. When placed next to an oncogene, they can accelerate gene expression resulting in cancers such as multiple myeloma (MM). In one study (Lovén et al., 2013), super enhancers were identified based on the binding of both MED1 and BRD4 in MM1.S tumor cells. Treatment with the drug JQ1 inhibits BRD4 which disrupts the super enhancers and limits the expression of oncogenes. The positioning of super enhancers close to oncogenes can occur through chromosome rearrangements. This has been shown for the MYC locus which is brought into the vicinity of super enhancers resulting in increased gene expression (Affer et al., 2014) . The methylation status of super

enhancers can also result in aberrant expression of gene targets with both hypomethylation and hypermethylation of these sequences capable of causing tumours (Heyn et al., 2016).

“Stretch enhancers” have also been described as large enhancers greater than 3 kb in length (Parker et al., 2013). Like super enhancers, stretch enhancers are important in regulating genes responsible for maintaining a particular cellular identity. Diseases associated with a particular cell type also showed an enrichment of disease causing SNPs in stretch enhancers compared to normal enhancers.

Analysis of the transition of pluripotent naive mESCs to epiblast stem cells (mEpiSCs) identified a subset of enhancers displaying the active histone marks H3K4me1 and H3K27ac (Chapter 1.2.10) that were unused in the mESCs but were later used in mEpiSCs to regulate pluripotency genes that were commonly expressed in both cell types (Factor et al., 2014). These enhancers were referred to as “seed enhancers” while the enhancers originally regulating gene expression in the mESCs that become inactive are referred to as the naive-dominant enhancers. The reason why these genes change the enhancers that regulate them (i.e. the switch from using naive-dominant to seed enhancers) during the progression of mESCs to mEpiSCs has not yet been fully deciphered. However, it is clear that seed enhancers do play an important role in later processes with many involved as components of super enhancers or stretch enhancers in adult tissue.

### **1.4.3 Evolution of Enhancers**

#### **1.4.3.1 Highly Conserved and Ultra-Conserved Enhancer Sequences**

The primary method of identifying novel enhancers is through sequence comparison of different genomes. Identification of conserved non-coding sequence is a primary indication that a regulatory element may be conserved between species and is therefore of functional significance. However not all enhancers have maintained tight sequence conservation between different organisms throughout evolution. Indeed, enhancers with similar functional properties that lack strict sequence conservation have been identified. On the other hand, the so called “ultra-enhancers” which contain 100 % sequence conservation for at least 200 bp (Bejerano et al., 2004) suggest that precise sequence information is important for these elements to function. Therefore, why do some enhancers require precise, uncompromising

sequence information to function whilst others can operate efficiently without such stringent requirements?

The association between conserved enhancer sequences and developmental genes was again highlighted by Plessy *et al.* (Plessy *et al.*, 2005) who attempted to identify common elements in zebrafish from a test sample of 104 murine enhancers by sequence comparison methods. Only 10.5 % of the test set showed conserved function, all of which acted upon developmental genes such as *dlx2*, *pax6*, *dlx5*, *pax2*, *hoxc8*, *hoxd4* and *nkx2.5*. It seems therefore that this tendency for highly conserved non-coding regions to regulate developmental genes is not unique to ultra-conserved enhancers, a point which was further emphasised by Visel *et al.* (Visel *et al.*, 2008). In this study, transgenic assays were conducted using both ultra-conserved enhancers and other highly conserved enhancers that lacked ultra-conservation. No significant difference was found in their functionality with both subsets predictably acting upon developmental genes. Pennacchio *et al.* (Pennacchio *et al.*, 2006) used high sequence conservation between human and pufferfish and ultra-conservation between humans and rodents as a method of identifying potential enhancers. Of the 167 sequences tested in transgenic reporter assays, the majority of identified enhancers were involved in nervous system development at E11.5. In an interesting side note, the authors compared their data to that provided by an additional ChIP-on-ChIP study (Lee *et al.*, 2006). This showed that 4 enhancers active at E11.5 also function as gene silencers in embryonic stem cells. Having this important dual role in development may explain the need for such high sequence conservation throughout evolution of these sequences.

These studies therefore suggest a critical role for ultra-conserved sequences in regulating gene expression. Removal of an ultra-conserved enhancer region would therefore be expected to cause massive phenotypic effects. However, in a study where four ultra-conserved regions were knocked out in mice (Ahituv *et al.*, 2007) no significant phenotype was observed. These ultra-conserved regions were specifically chosen because they mimicked the expression of nearby genes in *in vivo* reporter assays and were therefore believed to be acting as enhancers. The question therefore arises - why are such highly conserved sequences, which seem to act as enhancers, present in an organism when their removal seems to have no detrimental effect? Also, why is such high sequence conservation maintained? The authors suggest that these knock out mice may still exhibit mild phenotypic effects which show no obvious phenotype in a laboratory environment. Also, the deleted elements may exhibit a level of redundancy with other enhancers able to take over their functional roles.

### 1.4.3.2 Conservation of Function without Conservation of Sequence

Although sequence comparison techniques remain the primary method of enhancer identification, these methods alone do not reveal all of the functional non-coding regulatory elements acting within the genome. Some enhancers show relatively poor sequence conservation between species yet remain mechanistically functional. For example, the *even-skipped* (*eve*) enhancers in *Drosophila* and sepsids share poor sequence similarity but, when non-coding information from sepsid species are introduced into *Drosophila*, the *eve* expression patterns remain similar (Hare et al., 2008b, Hare et al., 2008a). Only small sequence blocks of roughly 20 - 30 bp remain highly conserved between the different species. These correspond to overlapping and adjacent pairs of transcription factor binding sites.

Similar experiments (Fisher et al., 2006) were conducted in which human conserved non-coding sequences regulating the *RET* gene were introduced into zebrafish via a transposon based assay. These human sequences recapitulated the zebrafish *RET* expression patterns despite lacking sequence conservation. The authors suggest two possible reasons for this. Firstly, the non-coding sequences from both organisms have evolved to function in a similar fashion despite lacking orthology. Alternatively, these sequences are orthologous but have undergone substantial rearrangement of transcription factor binding motifs while maintaining a similar function.

*Sparkling* (*spa*), an enhancer responsible for regulating *dPax2* expression in *Drosophila* cone cells, shows significant sequence differences between species with a rearrangement of specific transcription factor binding sites, while remaining functionally conserved (Swanson et al., 2011). When chimeric enhancers were created containing opposing halves from both *D. melanogaster* (*mel*) and *D. pseudoobscura* (*pse*) *spa*, the *mel*<sup>5'</sup> + *pse*<sup>3'</sup> construct was inactive highlighting that the *pse*<sup>3'</sup> could not substitute for the *mel*<sup>3'</sup>. Therefore, substantial binding site rearrangement has occurred over a relatively quick evolutionary timescale. This study went on to identify novel binding site motifs within both *mel spa* and *pse spa* which differ in both number and location within these enhancers. As the function of these regions is identical, the distribution of factors binding to these motifs must be variable between both regions. However, although binding site location was interchangeable between species, distances between some motifs were conserved, perhaps highlighting why enhancer function could be maintained despite differences in transcription factor distribution.

### 1.4.3.3 Conservation of Sequence vs Conservation of Function

It seems that two very different types of enhancers are present within metazoans. Some require precise sequence information to function, with this sequence highly conserved throughout evolution. Others share a less stringently conserved sequence between species but maintain small pockets of short, highly conserved nucleotides.

The “enhanceosome” and “billboard” models have been proposed as a paradigm of enhancer function and their viability reviewed extensively (Arnosti and Kulkarni, 2005, Rubinstein and de Souza, 2013). In the enhanceosome model, transcription factors bind collectively to binding sites within the enhancer to form a large nucleoprotein complex. Consequently, minor nucleotide alterations within the enhancer can disrupt protein binding suggesting that high sequence conservation within this region is important. Alternatively, in the billboard model transcription factor binding sites do not have such strict positional restraints. Separate nucleoprotein complexes form and act independently to ensure the correct enhancer activity. Several biological examples of both models have been identified.

*Ifn $\beta$*  represents a well-studied example of a gene regulated by an enhanceosome (Thanos and Maniatis, 1995). Its enhancer is made up of 4 positive regulatory domains (PRDs) responsible for binding three transcription factors; NF- $\kappa$ B, IRF-1 (later found to be IRF-3 and IRF-7 [20]) and ATF-2/c-Jun. In PRD swap experiments, virus stimulated IFN $\beta$  expression was reduced and was less specific, while other transcription factors such as TNF $\alpha$  were able to influence expression. The authors concluded that the precise organisation of PRDs was required for wild type expression and specificity. When a half helical turn of random DNA was inserted between these PRDs gene expression was greatly reduced. However, when the half helical turn was replaced by a full turn, higher gene expression levels resumed. Various combinations of DNA insertions were tested between PRDs leading to the conclusion that the precise orientation of PRDs was necessary to allow the combinational binding of transcription factors to maintain wild type gene expression levels. Therefore, IFN $\beta$  expression relies on the formation of a large nucleoprotein enhanceosome complex created by specific transcription factor binding at its enhancer.



#### **1.4.3.4 Mechanisms of Ultra-Conservation**

As discussed previously, various hypotheses have been suggested to explain the mechanism of ultra-conservation. Purifying (negative) selection at CNEs or decreased mutational rates are two opposing theories which have been put forward. However, recent studies (Drake *et al.*, 2006, Katzman *et al.*, 2007) argue that purifying selection is the dominant driving force behind these highly conserved regions and they are not mutational cold spots. A common theme in both studies is the analysis of derived allele frequencies (DAFs) to investigate the differences between conserved and nonconserved regions. Drake *et al.* (Drake *et al.*, 2006) analysed DAFs within CNEs in three separate human populations. An excess of rare derived alleles of SNPs (< 10 %) within these populations compared to nonconserved regions suggests that CNEs undergo purifying selection. Katzman *et al.* (Katzman *et al.*, 2007) looked at DAFs specifically within ultra-conserved elements in humans and compared these to nonsynonymous sites. They concluded that the purifying selection is three times stronger at ultra-conserved regions compared to nonsynonymous sites.

In 2006, Derti *et al.* (Derti *et al.*, 2006) investigated whether ultra-conserved elements were maintained within segmental duplications (SDs) and copy number variants (CNVs). They compared several sets of ultra-conserved elements with both data sets comprising SDs and CNVs from various sources. Ultra-conserved elements were shown to be severely reduced in both SDs and CNVs with three separate models proposed to explain these findings. Firstly, ultra-conserved elements may oppose DNA reorganisation. Secondly, these ultra-conserved regions are mutated or removed after duplication events. Finally, inclusion of ultra-conserved elements within a duplicated region produces a lethal phenotype or decreases fitness levels of an organism. The authors suggest that ultra-conserved elements may be dosage sensitive and both maternal and paternal copies pair together through some mechanism to maintain the correct copy number. If this mechanism is disrupted, by perhaps the presence of additional copies of the ultra-conserved element, then the fitness of the organism is detrimentally affected. Therefore, precise sequence information is needed to maintain this pairing, with deviations resulting in the eventual removal of the sequence from the population.

#### **1.4.3.5 Evolution of Enhancers**

Carter and Wagner (Carter and Wagner, 2002) suggest that enhancer evolution is simply dependent upon population size. They present a model whereby enhancer sequence variation is created by two deleterious mutations which together maintain enhancer function. In smaller

populations evolution rate is lower leading to tighter sequence conservation. Large populations have a faster evolution rate leading to increased sequence turnover. This may explain why invertebrate species with large population sizes and fast generation times have significantly greater sequence variability in enhancer regions compared to vertebrates.

Glassford and Rebeiz (Glassford and Rebeiz, 2013) analysed the sequence changes of the *Nepriylisin-1 (Nep1)* enhancer from the *Drosophila yakuba/Drosophila santomea* ancestor to the modern day *Drosophila santomea*. 4 mutations within this enhancer have evolved over 400,000 years. By creating all possible combinations of the 4 mutations leading to this current sequence and analysing the effects of these mutations iteratively by in vivo reporter assays, the authors were able to determine likely pathways for evolution of this enhancer. This study highlighted differing factors which affect enhancer evolution. For example, the addition of some mutations had no effect on their own in some pathways unless they were combined with an additional mutation. This shows that epistasis plays a critical role in enhancer evolution. The authors suggest that this enhancer displays aspects of an enhanceosome mechanism as protein-protein binding is necessary for optimum function. Conversely as no mutation causes significant detrimental effects the *Nep1* enhancer also shows characteristics of a Billboard model.

Non-coding human accelerating regions (ncHARs) have become of increasing interest to identify human specific enhancer sequences. These are non-coding regions which show high sequence comparison between mammals yet exhibit various alterations between humans and chimpanzees. These specific changes may regulate human specific gene expression profiles, unique from other mammals. In one study (Capra et al., 2013) a large sample of ncHARs were identified and subjected to several biological and computational based analysis methods leading to the conclusion that many of these regions serve as developmental enhancers. Of these, a subset of enhancers exhibited different expression profiles from chimpanzee reference sites, suggesting that these ncHARs displayed enhancer activity unique to humans.

#### **1.4.3.6 Enhancer Mutations and Disease**

Mutations in promoters and the coding regions of developmental genes have long been shown to cause a reduction in gene expression and an associated phenotype. Indeed, complete deletions of such regions have proven to be the cause of many congenital abnormalities. In the same way, mutations within enhancer sequences are gaining increasing recognition as an alternative culprit for a reduction in gene activity leading to disease. Such mutations in enhancers are responsible for deregulating the normal spatiotemporal expression of *Pax6* and

*Sox9*. *Pax6* is a key developmental regulator which displays a central role in eye formation and one of the main consequences of human *PAX6* misexpression is the disease aniridia. One of these mutations, a G > T transversion within the SIMO enhancer downstream of *PAX6*, was shown to affect the binding of PAX6 protein to the enhancer; thereby prohibiting autoregulation (Bhatia et al., 2013). In addition, a 681 kb deletion downstream of *PAX6* which deletes the 3' regulatory enhancers plus five other genes, has been linked to the disease ocular coloboma (Guo et al., 2013).

Campomelic Dysplasia (CD, MIM114290) defined by skeletal defects and male sex reversal is caused by mutations within *SOX9*. This gene is situated within a gene desert which contains a number of enhancers (Bien-Willner et al., 2007). Many of these identified as Sox9 enhancers in chromosome translocations which alter the position of the enhancers within the locus thereby affecting their interaction with the gene promoter resulting in an associated abnormality. Alterations of regions close to the gene have a more damaging effect compared to changes 375 kb to 932 kb upstream, suggesting that the former region contains enhancers with more critical functions (Yao et al., 2015). However microduplications within the more distant region can cause 46, XX *SRY*-negative DSD (Hyon et al., 2015, Xiao et al., 2013).

Pierre Robin Sequence (PRS, OMIM 261800) is caused by translocations occurring within a 160 kb region in the gene desert between *KCNJ2* and *SOX9*. In the same study, deletions both 5' and 3' of the *SOX9* gene were implicated in causing PRS as well as a T to C point mutation within a highly conserved non-coding element. (Benko et al., 2009). Interestingly, duplications within this gene desert do not cause PRS or CD, but result in brachydactyly-anonychia (Kurth et al., 2009).

#### **1.4.4 Enhancer Identification**

As mentioned previously, sequence comparison methods were used extensively to identify enhancers. However, this is not always an effective approach, with many enhancers showing significant sequence deviations between different species throughout evolution. Therefore, using sequence comparison approaches exclusively to identify enhancers may result in an incomplete analysis. Other additional, complementary techniques must also be applied to ensure a complete analysis such as:

## ➤ Chromosome Conformation Capture (3C) Techniques

Chromosome Conformation Capture (3C) has traditionally been used to study chromosome interactions (Hagège et al., 2007, Dekker et al., 2002, Miele et al., 2006). In this technique cross-linked chromatin is digested with a restriction enzyme after which closely located DNA is ligated together followed by the reversal of cross-linking. Primers designed to both of the digested fragments are used in PCR reactions to determine which DNA sites interact (Dekker et al., 2002). However, this procedure relies on prior knowledge or a hypothesis of interacting partners and is known as a “one-to-one” method. In 2006, two papers were published (Zhao et al., 2006, Simonis et al., 2006) describing a procedure (referred to as circular chromosome conformation capture (4C)) which built upon the original 3C protocol. This technique differed by the incorporation of a second digestion and ligation step to create small DNA circles where a known DNA sequence is ligated to an unknown sequence with which it interacts *in vivo*. Inverse PCR using primers designed within the known sequence were used to create a 4C library which was subsequently analyzed using microarrays. The advantage of this procedure was that the entire set of interactions of a sole DNA sequence could be examined, a so called “one-to-all” application.

Since then, a number of different protocols diversifying from the core 3C and 4C principles have been developed to analyse chromatin interactions (de Wit and de Laat, 2012, Duan et al., 2012, Simonis et al., 2009, Sexton et al., 2012). One of these was published in 2013 (Stadhouders et al., 2013) and described a variant of chromosome conformation capture referred to as multiplexed 3C-seq. This protocol is similar to 4C - involving two rounds of digestion and ligation – but, in contrast to traditional 4C technology, the amplified libraries are sequenced at the end of the procedure instead of using microarrays. Multiplexed 3C-seq allows the analysis of the interaction between a single “bait” sequence and the rest of the genome; i.e. a “many-to-all” strategy. This provides a distinct advantage over 3C-qPCR and 4C as several bait sequences can be analysed in a single experiment. Multiplexed 3C-seq also provides genome wide coverage and can thus provide information on *trans* contacts as well as those in *cis*.

Chromatin Confirmation Capture Carbon Copy (5C) has been used to look at all the genomic interactions occurring within a locus (Dostie et al., 2006). In this technique, 3C libraries are made as first described in the original 3C protocols but following ligation sequences designed to anneal to the ends of each restriction fragment are added. After a short annealing period,

Taq DNA ligase is added which ligates the annealed primers together. Amplification of these primers by PCR followed by high throughput sequencing and bioinformatic analysis, allows the construction of a detailed interaction map for every restriction fragment.

Traditional 3C provided a means of confirming suspected sequences are acting as enhancers. If a test sequence shows an interaction with a promoter, then it is likely that it may act as an enhancer *in vivo*. 4C and 5C are more exploratory measures which can identify all the sequences interacting with a tested gene. In this way, new enhancers can be identified.

Variations of the original 3C technique are summarised in Figure 1.6.

### ➤ **Fluorescent *in situ* hybridisation (FISH)**

Fluorescent *in situ* hybridization (FISH) is a technique which can be used to confirm observations obtained from chromatin conformation capture experiments (although this is not always the case (Williamson et al., 2014)). There are many variations but a basic protocol involves labelling specific DNA probes with molecules such as biotin or digoxigenin (dig). These are then re-suspended in hybridization mix and denatured at a high temperature. At the same time, formaldehyde fixed cells are treated with RNase and then denatured at a high temperature. The two are then combined allowing the tagged probes to hybridize to their corresponding sequences within the cells. The following day the cells are exposed to a series of fluorescently labelled antibodies which bind to either biotin or the dig and can be visualized using a fluorescent light microscope (Levsky and Singer, 2003, Price, 1993).

FISH experiments have proved useful in identifying or confirming enhancer promoter interactions. By fluorescently labelling an enhancer with one fluorophore and a promoter with another, the distance between both regions can be determined. If differing experimental conditions or cell types lead to changes in distances between enhancers and promoters, assumptions can be made about the activity of an enhancer in each case. For example; if a drug treatment results in a decreased distance between an enhancer and promoter, one may assume that the enhancer has actively moved to the promoter - a process which may drive transcription.

To date, many variations of FISH are utilised depending on the biological question. Basic 2D-FISH allows measurements in only the x and y planes while 3D-FISH allows additional

measurements through the z plane. Three colour FISH typically involves labelling the nuclei with DAPI and the DNA probes with two other fluorophores, while four colour FISH allows the use of three fluorophores in addition to DAPI. RNA FISH can also be conducted where RNA probes are utilised instead of DNA.

### ➤ **DNase1-seq, MNase-seq, FAIRE-seq and ATAC-seq**

Interphase chromatin forms a highly compact structure allowing the roughly two metres of DNA present in each human cell to be condensed within the nucleus. At the very basic level, this compaction is maintained by repeating units of histone complexes with DNA wrapped around them – referred to as nucleosomes. Within the genome, nucleosomes are a ubiquitous feature yet there are portions where these structures are depleted. These regions coincide with areas of active enhancers and expressing genes, where transcription factors and other proteins bind. Various techniques are used to identify such areas of open chromatin.

DNase1-seq (Song and Crawford, 2010) utilizes the enzyme DNase1 to cut between nucleosome compact regions releasing the active chromatin - the so-called DNase hypersensitive (HS) sites. After further processing and the ligation of adaptors, these short pieces of DNA can be sequenced using next generation technologies. The output from such an experiment is a list of sequences which can be aligned to a reference genome thus providing a visual map of open chromatin. In contrast to DNase1-seq, MNase-seq (Cui and Zhao, 2012) can be used to identify precise nucleosome positions. This uses MNase to digest the open chromatin leaving only the nucleosome bound DNA behind. This can then be isolated, amplified and sequenced with the analysis conducted in a similar manner to before. The mapped sequences should thus contrast directly with the results from DNase1-seq as they are highlighting areas of nucleosome occupancy. An alternative method of identifying open chromatin is FAIRE (Formaldehyde Assisted Isolation of Regulatory Elements). This involves the cross linking of DNA to chromatin which is then sheared, isolated and sequenced. The information gained from such an experiment is similar to that from DNase1-seq and often the two are often used in combination. One advantage of FAIRE is that formaldehyde fixation captures the chromatin state immediately prior to cell death whereas in DNase1-seq several steps are needed to permeabilize the cell which may cause the build-up of artifacts (Giresi et al., 2007).

Finally, ATAC-seq (Assay for Transposase-Accessible Chromatin using sequencing) is the most recent method designed to analyse open chromatin. This uses the Tn5 transposase to

integrate sequencing adaptors into open chromatin. One of the main advantages of this technique is that between 500 and 5,000 cells can be used as a starting point in contrast to the millions of starting cell numbers required in the previously described techniques. In this way tissue samples can be tested easily instead of growing cells in culture, a process which may affect their epigenomic state and thus the positioning of their nucleosomes (Buenrostro et al., 2013).

### ➤ Reporter Assays, STARR-seq and Enhancer Trap

The presence of open chromatin, although a strong indicator of active DNA, does not provide a quantitative measure of the strength of an enhancer. Traditionally, this has been measured using transgenic reporter assays. One simple approach may be to place a suspected enhancer sequence upstream of a minimal promoter and a reporter gene such as lacZ. If the enhancer is functional, the lacZ gene will be expressed and the level and location of this expression can be visualized by *in situ* hybridization experiments (or by staining for  $\beta$ -galactosidase which is the product of the lacZ gene). However, the use of such experiments becomes increasingly more time consuming when large numbers of suspected enhancers need to be tested. Indeed, such an approach is practically impossible when large libraries of sequences are to be tested. One potential genome-wide approach that has been developed for such a purpose is STARR-seq (Self-Transcribing Active Regulatory Region Sequencing). Here, libraries of suspected enhancer sequences are placed downstream of a minimal promoter and cloned using a given cell line. An active enhancer will act on the promoter increasing transcription of itself. The RNA from these cells is isolated, reverse transcribed and sequenced. When this is mapped to reference genome, only the active enhancers from the initial library are identified (Arnold et al., 2013). This technique therefore avoids a labor intensive means of screening multitudes of sequences for enhancer activity.

The enhancer trap technique utilises transposase enzymes to integrate promoter bound reporter genes randomly into the genome. It was first utilised in *Drosophila* (O'Kane and Gehring, 1987) before later adaptations allowed its use in vertebrates (Balciunas et al., 2004). Only promoters in close proximity to an enhancer will allow reporter gene expression. The reporter expression data can be used in coordination with gene expression data to identify the boundaries within which an enhancer can function (Chapter 1.3.2).

## ➤ ChIP-seq

In recent years, ChIP-seq has proved an excellent tool not only to identify enhancers but to also provide an indication of their activity. The basis of this technique lies in the fact that defined protein complexes are known to bind at enhancers which subsequently play a role in activating transcription at promoters. Probing for these complexes thus identifies a region containing an enhancer. One of the most common proteins tested is the acetyltransferase, P300. Using a P300 antibody, ChIP-seq has identified tissue specific enhancers in several studies with their activity confirmed in transgenic assays (Visel et al., 2009, Blow et al., 2010). This has identified enhancers that display poor sequence correlation between species and were not directly apparent.

In addition, the histones surrounding enhancers display intricate epigenetic patterns which correlate with the activity of the enhancer they surround. For example, in human ES cells (hESCs) active enhancers were associated with both H3K4me1 and H3K27ac and designated Class I, while Class II enhancers replaced H3K27ac with H3K27me3 and were termed “poised”. Higher levels of RNA pol II were associated with the active enhancers thereby inferring transcription while SUZ12 - a component of the polycomb PCR2 complex - was associated with the poised enhancers suggesting a more repressive state (Rada-Iglesias et al., 2011). In a later study in mouse ES cells (mESCs), three classes of enhancer were described with further subdivisions within each class. Similar to hESCs, active enhancers were classified based on high levels of both H3K4me1 and H3K27ac, while poised enhancers had both H3K4me1 and H3K27me3. Intriguingly the third class - termed intermediate enhancers - were identified purely by the presence of H3K4me1 without H3K27ac or H3K27me3. The activity of the genes nearest these enhancers was highest when near those classified as active, followed by intermediate and lastly by the poised enhancers. Additionally, other histone marks such as H3K36me3 provide an alternative classifier of active enhancers while H3K9me3 marks poised enhancers (Zentner et al., 2011). When ChIP is conducted using antibodies for these types of histone marks, the location and activity status of such enhancers can be determined. These types of experiments are particularly informative when analysing multiple tissue types simultaneously. For example, if studying the differentiation of ESCs to a particular cell type, the epigenetic status of enhancers could be compared across both cell lines. If, for instance, a particular enhancer changes from a poised to active state between the ESCs and differentiated cells, it implies that this enhancer is involved in the differentiation process.



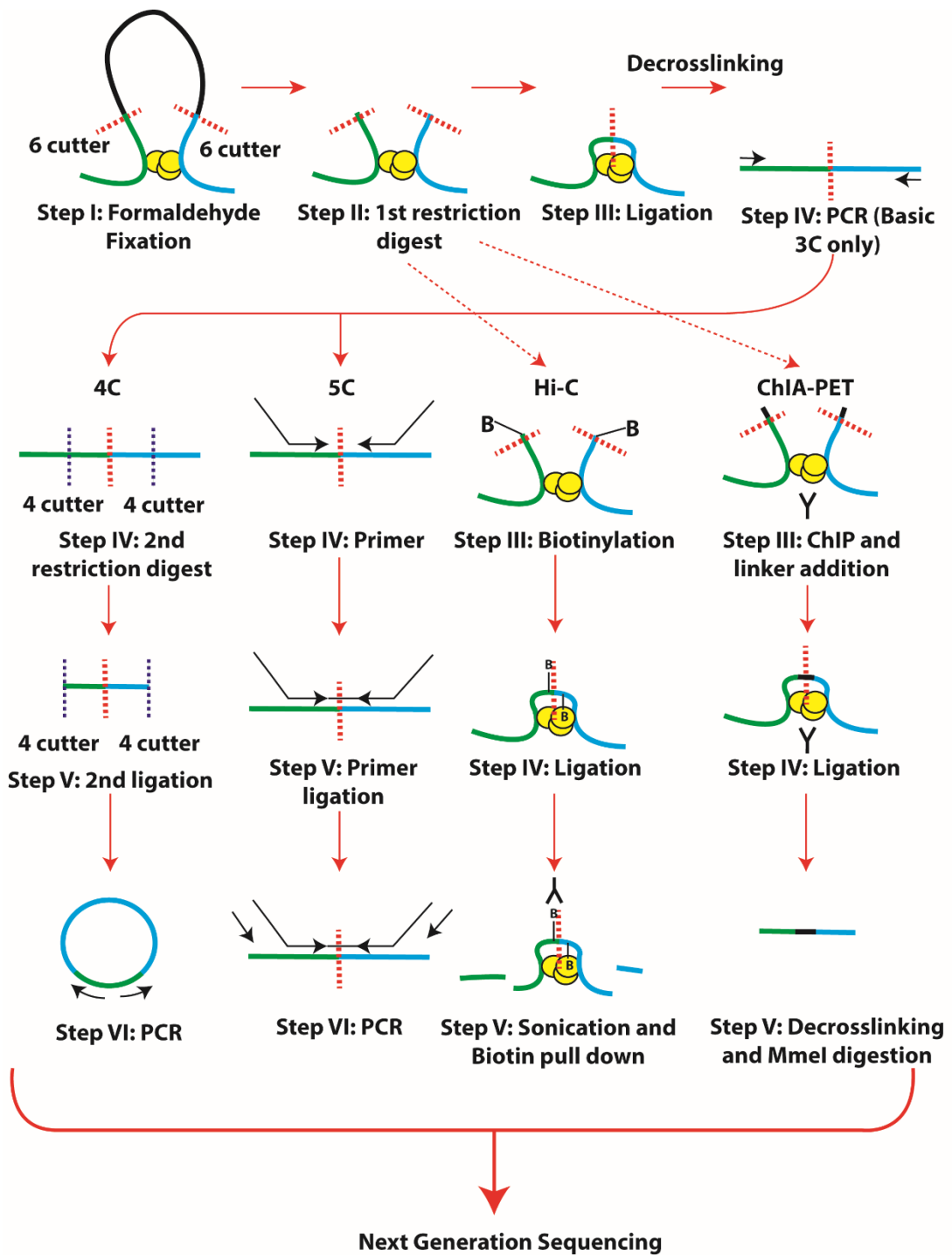


Figure 1.6 – Summary of the keys stages of the commonly used chromosome conformation capture techniques; 3C, 4C, 5C, Hi-C and ChIA-PET. A multitude of other 3C style techniques are available but are not analysed in this figure. Both 4C and 5C involve the construction of a 3C library before further steps are conducted. Hi-C and ChIA-PET only follow the original 3C protocol until the first restriction enzyme digestion after which both protocols diverge. In all protocols the finished libraries are sequenced using next generation technologies.

## 1.5 Chromatin Structure

### 1.5.1 Topological Associated Domains (TADs)

*“The DNA can be imagined as a long thread that is subdivided by walls into various sections or TADs. All elements within a section – genes, regulators, transcription factors, polymerases and many others – can interact freely. The walls separate the sections and shield them from neighbouring activity”.*

*Professor Stefan Mudlos, Genetic diseases shift boundaries within the genome,  
Available from: <https://www.mpg.de/9229203/dna-tads>*

The advent and subsequent development of the “C” technologies provided a novel way of looking at chromatin structure and folding at a genome wide level. Hi-C in particular allowed scientists to view the multitude of interactions a gene can make both *in cis* and *in trans* and the boundaries along the DNA preventing further interactions from occurring. From these experiments the term “Topological Associated Domain (TAD)” was derived to describe regions of DNA, typically around 1 Mb in length, in which the interactions within the region are maintained. Chromatin from all cells is divided into such regions with the same TAD boundaries consistent across different species and cell types, but also showing diverse intra-TAD interaction profiles dependent on the activity of genes within the TAD (Dixon et al., 2012). The reason DNA is divided into such regions and the exact time at which this occurs in development is an active area of research. At present, increasing evidence has implicated CCCTC-binding factor (CTCF) in marking the boundaries of TADs. For some time CTCF was thought to behave solely as an insulating protein shielding genes from the activating effect of enhancers (Bell et al., 1999). However, recent reports suggest an architectural role where CTCF can bring distant acting regions into close proximity. Indeed, 3C has shown that CTCF is necessary for creating loops between two distant regions, with its removal resulting in the destabilisation of these structures (Splinter et al., 2006). The assumption is, therefore, that by bringing two regions together in a loop, CTCF can mediate the interaction between an enhancer and promoter leading to gene activation. However, other enhancers excluded from the loop are physically unable to interact with genes within the loop. In this way, CTCF is acting as an insulator. The loops created by the binding of CTCF are the TADs observed from Hi-C. This basic model of CTCF facilitating DNA loops, which bring distant sites together,

becomes more complicated when other architectural proteins are also considered. ChIP-seq experiments showed that CTCF shares binding sites with both cohesin and mediator, two proteins known to interact with a predicted involvement in DNA looping (Kagey et al., 2010, Ramalho-Santos et al., 2000). The size of the DNA loops created by these proteins is dependent on the combination of proteins present. For example, in one study using multipotent neural progenitor cells (NPCs) and mouse ES cells, the largest loop structures were formed when CTCF and SMC-1 combine (> 1 Mb) while the smallest loops formed when MED12 and SMC-1 were present (<100 kb). Loops ranging from 600 - 1 Mb occur when MED12 binds on its own while a combination of the three factors produces loops of < 300 kb (Phillips-Cremins et al., 2013). The combination of architectural proteins present is therefore necessary in determining the size of DNA loops which are manifested as TADs or the small sub-domains within TADs referred to as sub-TADs.

Recent studies suggest that the orientations of CTCF binding sites (CBS) along the genome are involved in the construction of TADs. CBSs in the forward orientation are likely to interact with those in the opposite orientation leading to the formation of a loop. TAD boundaries form when a reverse CBS is placed alongside a forward site. Indeed, CRISPR/Cas 9 experiments where CBSs have been re-orientated in the human  $\beta$ -globin locus show that novel interactions are possible when reverse-forward CBSs have been transformed into forward-forward, with the new forward CBS interacting with a reverse CBS previously inaccessible (Guo et al., 2015).

This system of correctly orientated CBSs along the genome creating TADs seems to be conserved across species suggesting a universal method of packaging DNA. Indeed, highly conserved CBSs are located at TAD boundaries, while CBSs with weaker sequence conservation tend to be species specific and are more likely to fall within TADs. These unique sites are probably involved in novel interactions leading to gene expression profiles which differ from the other species (Vietri Rudan et al., 2015).

## 1.5.2 Identifying Mechanisms of Loop Formation through Polymer based Modelling

Using Hi-C and ChIP-seq for chromatin architectural proteins, various studies investigated the mechanism by which TAD loops are initially formed. One of the most promising models is known as the “Chromatin Extrusion” model. In this scenario, loop extruding factors (LEFs) bind to the DNA, probably as a dimer, with one unit moving along the genome in a forward direction and the other in reverse. In this way, DNA is looped through the LEFs without producing a knot. The loop will continue to grow in size until it reaches a boundary element. The most likely candidates for LEFs are cohesin while the boundary elements may be the CBSs (Fudenberg et al., 2016). The means by which cohesin and CTCF coordinate to loop out DNA have not yet been elucidated. It can be imagined that a cohesin dimer binds and moves along the DNA until it reaches CTCF already bound to two separate correctly orientated CBSs thereby stopping the cohesin unit in its path. Alternatively, cohesin and CTCF may bind to DNA together and the combined structure is stopped upon reaching the CBSs (Sanborn et al., 2015, Yardımcı and Noble, 2015). In contrast, the cohesin complex may not bind within the TAD, but two separate complexes may bind at the TAD boundaries along with RNA pol II. As the DNA is transcribed, it is looped through both complexes, only stopping when they meet (Razin et al., 2016).

Recently, the location of both CTCF binding and DNase1 hypersensitivity sites (DHSs) have been used to create models of chromatin structure. In these examples, chromatin is modelled as a polymer consisting of multiple beads, with protein bridges able to form between individual CTCF binding sites or DHSs. The accuracy of these conformations are established by comparisons with experimental chromatin conformation capture data. These models were applied to both the  $\alpha$ - and  $\beta$ -globin loci, producing structures with interaction profiles correlating highly with Capture-C data. The authors suggest that removing CTCF bridges from the models may provide an in-silico approach to analyse the effect of CTCF binding site deletions. In this way systemically removing aspects of the model may provide an insight into their biological significance (Brackley et al., 2016).

The aforementioned studies have all used a defined set of parameters to create simple chromatin models where the contact frequencies within a defined region can be compared to heat maps from experimental 3C type data. An alternative approach is to use these heat maps directly to induce chromatin structure.

### 1.5.3 Mutations at TAD boundaries cause disease

As previously raised, TADs play an important role in restricting the accessibility of a defined set of genes to a selection of enhancers which dictate the spatiotemporal expression of these genes within an organism. The protection role TADs provide is established through the boundary regions which prohibit enhancers in other TADs from acting. Over the last several years there has been increasing interest in the effect of disrupting TAD boundaries and the correlation of this with disease (Figure 1.7).

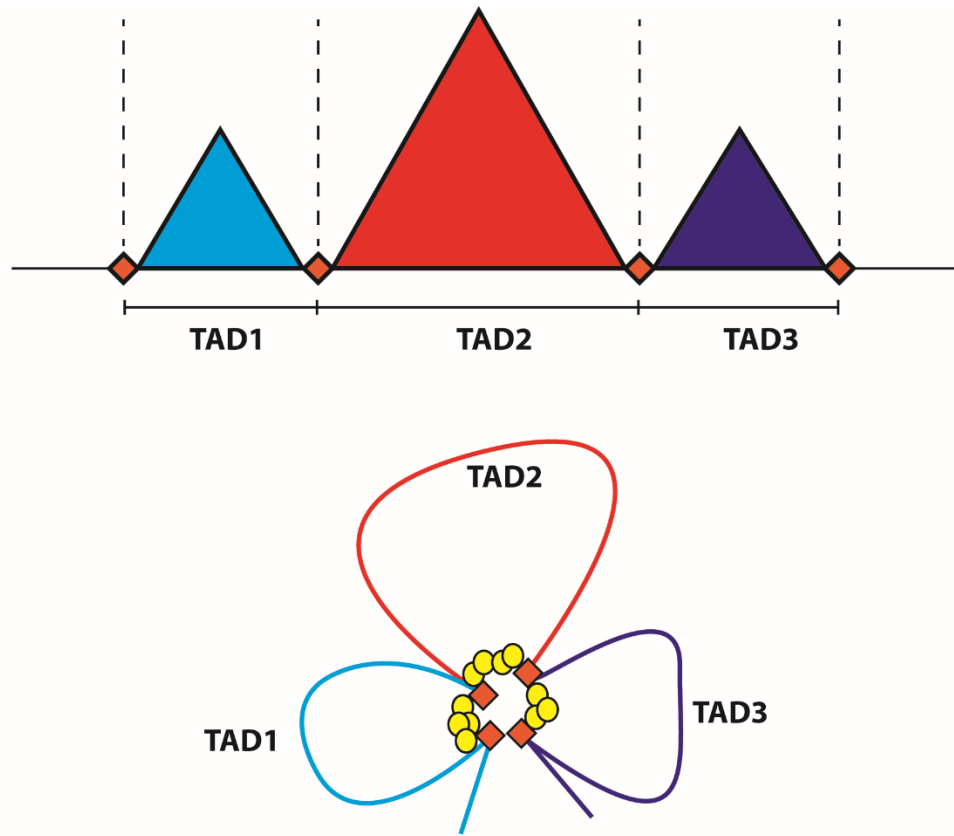
In a recent study, defects at human TAD boundaries were shown to result in limb abnormalities (Lupiáñez et al., 2015). The region containing three TADs was examined with the genes *WNT6* and *IHH* contained in the first and most centromeric TAD; *EPHA4* in the second TAD; and, *PAX3* in the third most telomeric TAD. Genomic deletions incorporating the boundary between the *EPHA4* containing TAD and the *PAX3* containing TAD resulted in the phenotype Brachydactyly, while inversions and duplications at the boundary between the *WNT6-IHH* TAD and the *EPHA4* TAD resulted in F-syndrome. These mutations were replicated in mouse models using the CRISPR/Cas9 system and the interaction profiles of the genes within the TADs analysed by 4C. Interestingly, novel interactions were observed for the *WNT6* and *IHH* genes with regions in the adjacent TAD when the boundary region was altered. New interactions were also detected for the *PAX3* gene with the *EPHA4* TAD when the TAD boundary separating the two was removed. A cluster of common interactions was observed for all these mutant 4C profiles to a region of 150 kb within the *EPHA4* TAD; a site containing several *EPHA4* enhancers. The disruption of the *EPHA4* TAD boundaries therefore allows the genes in adjacent TADs such as *WNT6*, *IHH* and *PAX3* to adopt enhancers to which they have previously been insulated, resulting in aberrant gene expression profiles and mutant phenotypes. Deletions at TAD boundaries have also been linked to other diseases such as autosomal dominant leukodystrophy (ADLD) where a ~660 kb deletion removes a boundary element of the *LMNB1* containing TAD. This allows enhancers from the adjacent TAD to make previously inhibited interactions with *LMNB1* resulting in increased Lamin B1 expression in the brain (Giorgio et al., 2015). In one study, Mesomelic dysplasia Savarirayan type was shown to be caused by large deletions across three TAD boundaries. This is believed to bring a number of enhancers into the vicinity of *ID4*, altering its expression thereby producing the associated skeletal abnormalities (Flöttmann et al., 2015). This mechanism, where enhancers - previously shielded from a gene by a boundary element - come into contact

with the gene when the boundary is removed and alters its expression levels, has been termed “enhancer adoption”.

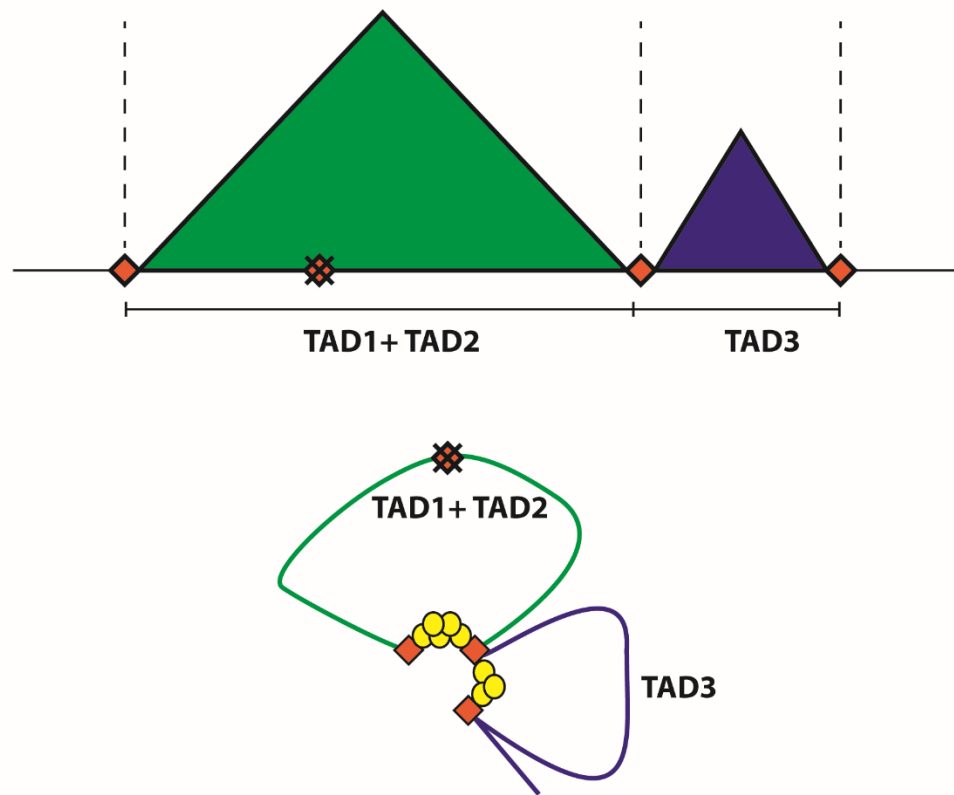
The method of using the CRISPR/Cas9 system to re-position TAD boundaries by deleting CBSs has been used in several other studies to determine if ectopic interactions that result from these mutations alter gene expression. For example, one such investigation focussed on the role CBSs had in demarcating TADs during the retinoic acid induced differentiation of ES cells (ESCs) to motor neurons (MNs). In WT MNs, a CBS situated between *Hoxa5-6* separates two adjacent TADs, with *Hoxa1-5* fully expressed in one TAD and *Hox7-13* repressed in the other. Deletion of the CBS shifts the TAD boundary caudally to the next CBS resulting in increased *Hoxa7* and *Hoxa9* expression (Narendra et al., 2015).

Recently the effect of shifting TAD boundaries has also been shown as a cause of cancer (Valton and Dekker, 2016). For example, gain-of-function *IDH* mutant gliomas show increased DNA methylation which, when present at CBSs, reduces the binding of CTCF. Among the many loci affected, the reduction in CTCF binding between the adjacent TADs containing *FIP1L1* and *PDGFRA* allows the interaction of *FIP1L1* enhancers with *PDGFRA* thereby leading to increased expression of this oncogene (Flavahan et al., 2016). Several other examples have been identified where the removal of a CBS increases expression of a proto-oncogene (Hnisz et al., 2016) or structural genomic rearrangements such as duplications and deletions have re-positioned enhancers in close proximity to an oncogene resulting in cancer (Northcott et al., 2014).

(A)



(B)



**Figure 1.7 – (A, top) The output from most 3C type experiments is an interaction frequency map which depicts how frequently different portions of the genome interact with each other. When presented graphically, these frequently present as triangles as only certain regions of the genome are capable of interacting with each other. These regions are referred to as Topological Associated Domains (TADs) and are separated by boundary elements. The cause of these boundaries is an area of ongoing research but the increasing evidence suggests that the binding of the protein CTCF to its binding sites (orange diamonds) is sufficient to stop interactions between TADs. (A, bottom) In reality, TADs probably appear as loops of DNA, where all sequences within the loop are capable of interacting with each other. (B, top) Using the CRISPR/Cas9 system, some laboratories have systematically removed CTCF binding sites and studied the effect on the associated TADs. In some cases, this has resulted in the merging of two adjacent TADs allowing enhancers previously inaccessible in one TAD to interact with genes in the other TAD thereby resulting in an aberrant gene expression. In interaction frequency maps this is shown by the merging of two triangles. The removal of TAD boundaries has been identified as a source of several diseases due to enhancers from foreign TADs activating genes at inappropriate times. (B, bottom) Removal of such a boundary is likely to affect the looping structure seen normally in the genome.**



### 1.5.4 Lamin Associated Domains (LADs)

In eukaryotic cells, the nuclear envelope separates the cytoplasm from the nucleoplasm and is composed of two separate membranes: the inner nuclear membrane (INM); and, the outer nuclear membrane (ONM). The surface of the INM pointing into the nucleoplasm is surrounded by a mesh of intermediate filaments referred to as the lamin proteins (Wilson and Foisner, 2010). This is known as the nuclear lamina. In mammals, the three genes *LMNA*, *LMNB1* and *LMNB2* are responsible for regulating a series of different lamin isoforms (Dittmer and Misteli, 2011). The most common are Lamin A, Lamin B1, Lamin B2 and Lamin C. The roles of these lamin proteins in regulating gene expression is an area of ongoing biological research.

In 2008, a study was conducted to identify all genomic sequences interacting with the nuclear lamina in human fibroblasts. DNA adenine methyltransferase identification (DamID), using a chimera of DNA adenine methyltransferase (Dam) and human lamin B1, identified 1300 regions binding to the nuclear lamina. These sequences - referred to as Lamin Associated Domains (LADs) - ranged in size from 0.1 - 10 Mb and were associated with areas of low transcription based on the decreased levels of RNA polymerase II and the activating histone mark H3K4me2 at gene promoters. The presence of CTCF and active gene promoters acting away from the LADs, near the LAD boundaries was hypothesised to prevent the spread of euchromatin into the LAD and the spread of heterochromatin out of the LAD (Guelen et al., 2008).

This study was quickly followed by another looking at the change in LADs accompanying the transition of mouse embryonic stem cells (mESCs) to neural precursor cells (NPCs) and then to astrocytes (ACs). Overall the position of LADs was highly maintained with a conservation of between 73 - 87 % between cell types. However, there were some important differences. Genes involved in maintaining the pluripotency of the ESCs such as *Nanog* and *Klf4* were repositioned towards the nuclear lamina during the differentiation process which correlated with a decrease in their expression. However, a reduced gene expression accompanying an association with the nuclear lamina was evident only in certain cases. When ESC to NPC differentiation was examined, nuclear lamina repositioning occasionally had little effect on gene expression. However, these “silent” genes which either moved towards or away from the nuclear lamina and had no effect on gene expression during the transition from ESCs to NPCs, did show expression differences during the subsequent transition of NPCs to ACs. The

authors suggested this first movement primes the genes to be activated at later stages (Peric-Hupkes et al., 2010).

At this stage it was known that LADs were associated with a repressed chromatin state, with genes located in these areas showing a reduction in gene expression compared to genes located away from the nuclear lamina into the nucleoplasm. However, the mechanism promoting this relationship between LADs and gene expression was not understood. Were poorly expressed genes simply repositioned to the nuclear lamina or was there an active process that shuttles specific genes to the nuclear lamina where gene expression is subsequently reduced? In a series of experiments using constructs containing either LAD or non-LAD DNA, Zullo *et al.* showed that the expression of a gene could be reduced when relocated to the nuclear lamina. The specific sequences in LADs responsible for this were enriched for the sequence motif GAGA and referred to as lamina associating sequences (LASs). The protein CKROX along with HDAC3 and LAP2 $\beta$  were shown to bind to LASs with knockdown of CKROX reducing the ability of LADs to position at the nuclear lamina (Zullo et al., 2012). In addition, studies using *Drosophila* had previously shown the Lamin LamDm<sub>0</sub> mutants had increased expression in testis specific gene clusters in the regions *60D1* and *22A1*. Removal of LamDm<sub>0</sub> by RNAi also resulted in the movement of these genes from the lamina into the nucleoplasm (Shevelyov et al., 2009). Using TALENs, Therizols *et al.* showed that transcriptional activation of the genes *Ptn*, *Sox6* and *Nrp1* in ESCs was sufficient to drive repositioning of chromatin towards the nuclear centre. However, by introducing an acidic peptide at these genes which causes chromatin decondensation with the absence of gene transcription, there was still a movement of chromatin away from the nuclear periphery. This suggests that chromatin decondensation alone is sufficient to relocate chromatin from the nuclear lamina to the nucleoplasm (Therizols et al., 2014).

Interestingly, in mESCs the presence of lamin proteins was not essential for the formation of LADs. This was proven by replacing Lamin B1 with Emerin (EMD) for DamID experiments. Emerin is also bound at the nuclear lamina and produces near identical DamID profiles to those produced with Lamin B1. Knock out of both Lamin B1 and B2 in mESCs produced almost identical Dam-EMD profiles to those obtained from wild type cells. The expression of genes within LADs was also unaffected suggesting that these proteins were not essential for the repressive nature observed at LADs. To eliminate the possibility that Lamin A and C were compensating for the removal of Lamin B1 and B2 in LAD formation, these were knocked down by RNAi. Again Dam-EMD profiles were near identical to the wild type profiles. Therefore, this study suggests that an alternative mechanism is responsible for the repression

of genes at LADs in mESCs which does not utilize lamin proteins to clamp genes to the nuclear lamina. This is in contrast to differentiated cells and human fibroblasts where the presence of lamin proteins at LADs is essential for gene repression at the nuclear lamina (Amendola and van Steensel, 2015).

### **1.5.5 Laminopathies**

In many cell types, the presence of lamin proteins at the nuclear lamina plays an essential role in regulating gene expression (Chapter 1.3.5). Disruption of these proteins therefore seems a likely source of gene expression irregularities possibly resulting in disease. Indeed, many diseases have been identified with defective lamin proteins being the root cause. Collectively, these diseases are termed “laminopathies”.

Many laminopathies have been identified and extensively reviewed (Worman and Bonne, 2007) with the source of many of these diseases routed to defects in either Lamin A or Lamin B proteins. Emery-Dreifuss muscular dystrophy (EDMD) is one of these diseases characterised by abnormalities in both skeletal and cardiac muscle. A study using *C.elegans* looked at the effect of a Y59C mutation in LMN-1, a mutation which has been observed in human EDMD patients. This prevented the release of chromatin linked to muscle specific genes at the nuclear periphery thereby reducing their expression. The same circumstances may be occurring in humans (Mattout et al., 2011). Another well studied laminopathy is Hutchinson-Gilford progeria syndrome (HGPS), a premature aging disease resulting in average life span of 13 years (Merideth et al., 2008). The source of HGPS is an irregular form of Lamin A caused by a C > T mutation in exon 11 at residue 1824. This mutant is known as progerin and - unlike the Lamin A protein - it maintains the farnesyl group post translational modification. (Cao et al., 2007). This affects progerin by anchoring it to the nuclear membrane and preventing its release. Treatment with farnesyltransferase inhibitors (FTIs) can remedy some of the cellular effects seen in HGPS by allowing release of progerin from the nuclear membrane (Capell et al., 2005). In one study, progerin was shown to interact with a unique subset of genes from lamin A causing abnormal gene expression changes which may contribute to the phenotypes observed in HGPS patients (Kubben et al., 2012).

### **1.5.6 Radial Positioning in the Nucleus and Chromosome Territories**

After mitosis chromosomes undergo a process of decompaction and present as less ordered chromatin within the nucleus. For a long time, it was not known whether chromatin originating

from different chromosomes was allowed to intermingle or whether it was restricted to a series of distinct domains. Over the last several decades the latter option has been proved with interphase chromosomes occupying the so-called chromosome territories (CTs). In one of the key experiments leading to this conclusion, conducted by Zorn and colleagues, cells were irradiated with a UV laser and then monitored for DNA damage. Only certain chromosomes showed damage at any one time thus demonstrating how unlikely it was that chromosome regions are intertwined. If this were the case, multiple sites of damage on many different chromosomes would have been observed (Zorn et al., 1979).

Later research highlighted that the organization of CTs within the nucleus was also structured. In general, a correlation exists between the density of genes within a chromosome and its location within the nucleus. FISH experiments using chromosome paints in both human lymphoblast and fibroblast cell lines showed that the more gene rich chromosomes such as HSA1, -16, -17, -19 and -22 were located much closer towards the centre of the nucleus than the gene poor chromosomes such as HSA2, -4, -13 and -18 which are located much closer to the nuclear periphery (Boyle et al., 2001) (Figure 1.8). This pattern seems applicable to most cell lines as a study looking at the nuclear position of the gene poor chromosome 18 and the gene dense chromosome 19 showed that the CT of chromosome 19 was situated closer towards the nuclear interior in a variety of different cell types (Cremer et al., 2003).

So, how does this concept of chromatin territories feed into what is already known about transcriptional regulation? In 2004, Chambeyron and Bickmore conducted a study looking at the regulation of the *HoxB* gene cluster during differentiation of ES cells. Retinoic acid (RA) induced expression of the *Hoxb* genes in a co-linear fashion beginning with the expression of *Hoxb1* after 4 days and ending with the expression of *Hoxb9* after 10 days. Interestingly, activating histone marks were evident across the entire gene cluster 4 days after RA addition and a global decompaction of the region was evident after 2 days. The authors concluded that both of these events served to poise the genes for activation without activating them, as some genes containing activating histone marks 4 days after RA addition were not expressed. Instead, the looping out of poised genes from the CT resulted in the subsequent gene expression. Genes along the *HoxB* cluster are looped out in order from *Hoxb1* to *Hoxb9* thereby reflecting the order in which they are expressed (Chambeyron and Bickmore, 2004). The result of this looping out of genes from their CT may be to allow their interaction with transcription factories (Chapter 1.3.5).

Interestingly in a follow up study, this time looking at the *HoxD* gene cluster instead of the *HoxB* cluster it was found that the process of decondensation did not always correlate with dissociation of genes from their CT. Indeed, the order of events seems to depend on the anatomical position in which the genes are expressed. For example, the mouse tail bud showed both decondensation and looping out of genes while the limb bud showed decondensation while maintaining genes within the CT (Morey et al., 2007).

### 1.5.7 Transcription Factories

In an experiment conducted in 1996 transcription in HeLa cells was shown to be localised to a set of roughly 2,100 discrete “transcription factories” within the nucleus each with a diameter of around 71 nm (Iborra et al., 1996). The authors of this study suggested that contrary to previous thought suggesting RNA polymerases were believed to move along DNA regulating transcription, the RNA polymerases may occupy fixed positions within transcription factories and instead it is the DNA template which moves over the enzyme. Since then there has been a stream of publications concerning transcription factories. In 2004, Osborne and colleagues showed that multiple genes situated at different locations along a chromosome could come together at the same transcription factory. Using both RNA FISH, DNA FISH and RNA immuno-FISH, *Hbb-1* and *Eraf* (two genes expressed in erythrocyte cells) showed an increased interaction at the same RNA polymerase II site for transcriptionally active alleles. Alleles that were transcriptionally inactive were located outside these regions. As transcription is known to be a discontinuous process, the authors concluded that the repeated movement of genes into and out of a transcription factory is likely to result in these pulses of transcription (Osborne et al., 2004). Interestingly a complete loss of transcription does not alter the number of transcription factories within the nucleus suggesting that they are distinct non-transient structures (Mitchell and Fraser, 2008).

The next question addressed was whether all transcription factories are equivalent or are different factories responsible for the transcription of different types of genes? This was answered recently when Xu and Cook showed that promoters specific for RNA polymerases type I, II and III are localised to different transcription factories. Genes with introns also re-located to transcription factories different to those occupied by genes without introns (Xu and Cook, 2008). Accordingly, transcription factories are specialised compartments which are responsible for the regulation of only specific subsets of genes.

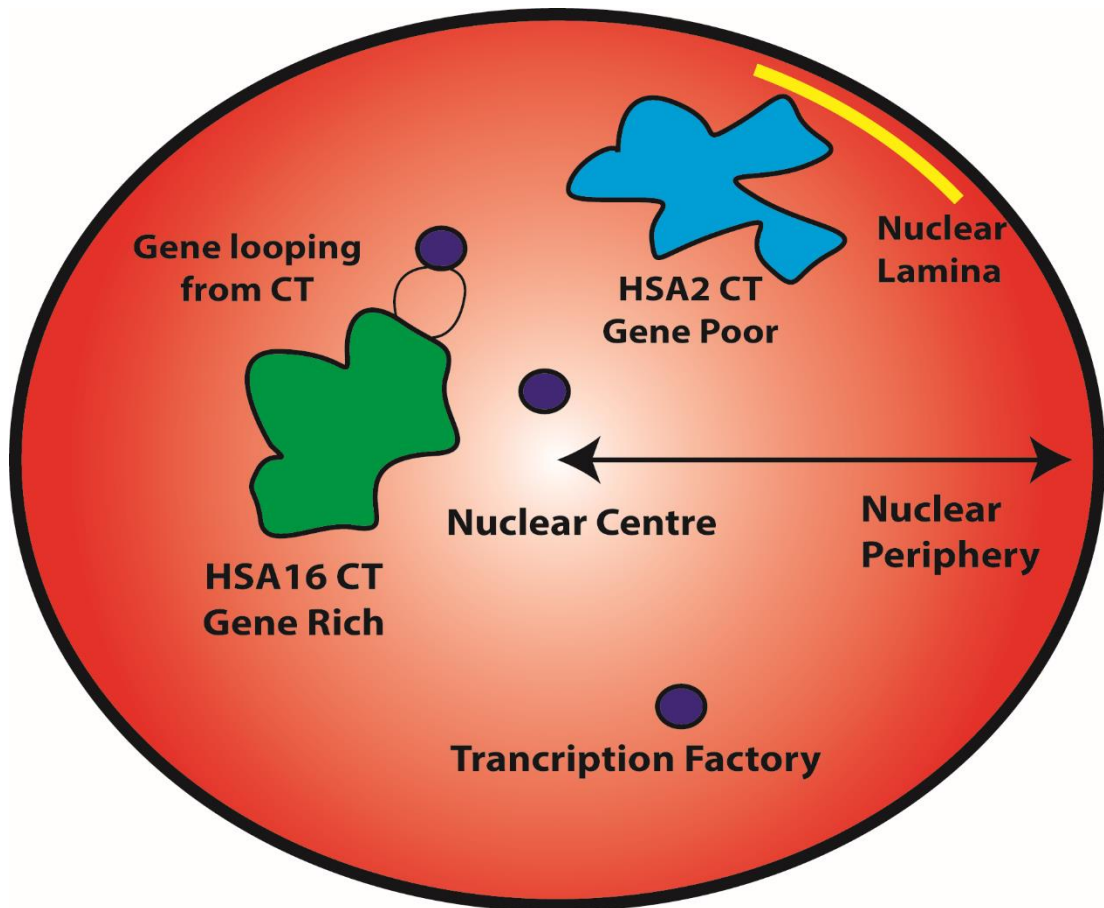


Figure 1.8 – Recent evidence has shown that the nucleus of eukaryotic cells is not as disorganised as previously thought but has a distinct structure. After mitosis the chromosomes of interphase nuclei decondense and occupy discrete chromosome territories (CTs). In general, gene rich CTs such as the human HSA 16 CT (shown in green) are located closer to the nuclear interior while gene poor CTs such as HSA2 (shown in blue) are located closer to the periphery. At the periphery, chromatin in contact with the lamina referred to as LADs (lamina associated domains) generally has inactivated genes which can become activated by withdrawing from this region. Sometimes, genes within CTs have been shown to loop out from their CT and make contact with transcription factories (purple circle). These consist of RNA polymerase molecules in close proximity which regulate transcription.

### 1.5.8 10 nm Fibre (beads-on-a-string) and 30 nm Fibre Chromatin Structures

In humans, the nucleus of each cell holds the genomic DNA which (when extended) is roughly two metres in length. To fit into such a small compartment, DNA must be tightly packaged - a process hypothesised to occur through multiple processes. At the basic level, DNA is wrapped around histone proteins which - due to their positive charge - negate the repulsive forces of long negatively charged DNA chains. Histone octamers, consisting of an H3-H4 tetramer bound to two H2A-H2B dimers, wrap around 147 bp of DNA (Ausió, 2015). Evidence of chains of histones entangled by DNA and separated by linking DNA is widespread, with the “beads-on-a-string” like structure (Olins and Olins, 1974) referred to as the 10 nm fibre. The discovery of a more compact 30 nm fibre *in vitro* suggested that the 10 nm fibre may integrate into this structure to aid compaction. However, scientific understanding of the 30 nm fibre remains in its infancy with debate ongoing as to whether such a structure actually exists *in vivo*. The way in which the 30 nm fibre folds is also contested with a one start helix solenoid model and a two start helix zigzag model the most likely scenarios (Maeshima et al., 2014). The crystal structure of a tetranucleosome favours the latter model (Schalch et al., 2005). However, increasing evidence suggests that an important structural feature is the interaction between acidic amino acids on H2A with the H4 N-terminal tail; disruption of which results in destabilisation of the entire 30 nm fibre (Tremethick, 2007). A number of extrinsic factors may influence the structure of the fibre during the isolation process and it is possible that both structures may exist depending on the conditions. The lack of *in vivo* evidence for the 30 nm fibre remains a worry with some researchers even suggesting that it is merely an artefact caused by low salt concentrations which are dissimilar to the physiological conditions (Maeshima et al., 2014).

Evidence for the existence of the 30 nm fibre in mitotic chromosomes has centred on results obtained from small-angle X-ray scattering (SAXS) which has detected peaks believed to represent the 30 nm fibre (Langmore and Paulson, 1983). However, the results were a direct contradiction of those obtained from cryo-EM studies which showed the absence of such structures. Further analysis of the SAXS data highlighted that the 30 nm fibre peak was actually caused by contaminating ribosomes and when these were removed the peak was no longer detected. However, a peak is still detected in chicken erythrocyte nuclei which do not have ribosomes suggesting that this structure may be present in erythrocytes. The authors of this study suggested that mitotic chromosomes therefore, are not composed of 30 nm fibres but are assembled based on the irregular cooperation of 10 nm fibres (Nishino et al., 2012).

### 1.5.9 Higher Order Chromatin Structure

Whilst it seems that the basic packaging unit of chromosomes are irregularly orientated 10 nm fibres (Chapter 1.3.5) with the potential involvement of 30 nm fibres, what are the additional higher order packaging structures which allow the chromatin to be so highly condensed? In 1978 Sedat and Manuelidis proposed what would later become known as the hierarchical helical folding model (Sedat and Manuelidis, 1978) where the 30 nm fibre folds into a 100 nm fibre which subsequently folds into a 500 nm fibre, with the process continuing until the largest chromatin fibres are formed. A year later, the radial loop model was also hypothesized (Marsden and Laemmli, 1979) which suggested that chromatin is first organized into loops where the base of each loop is brought together to form a central core. However, both of these models were based on the idea that the 30 nm fibre is present *in vivo*. The current model referred to as the polymer melt model, suggests that - in low chromatin concentrations - intra-molecular interactions are favoured allowing the formation of a 30 nm fibre. However, as the chromatin concentration is increased the number of inter-molecular interactions also increases and the irregularly folded chromatin 10 nm fibres interdigitate (Maeshima et al., 2010). This gives chromatin a liquid like property as it can adopt certain conformations depending on the conditions (Maeshima et al., 2016). This model may apply to both mitotic and interphase chromatin. Accumulation of clumps of interdigitated irregularly packaged chromatin may come together to form “chromatin liquid drops” - clumps of chromatin where transcription can only occur on the outside of the drop while the inner regions are shielded (Maeshima et al., 2014). These structures may correlate with the previously discussed TADs and form the basis of higher order chromatin structure.



## 1.6 Project Aims

The aim of this investigation was to gain an increased understanding of the mechanism by which the ZRS activates the *Shh* gene. To do this chromosome conformation capture and fluorescent *in situ* hybridisation (FISH) techniques were both employed using *Shh* inducible cell lines (14fp cells) which were derived from mouse E11.5 limb buds. These cell lines allowed the study of the ZRS in a controlled environment where the *Shh* gene could be activated by the addition of the chemical trichostatin A (TSA) or by other means such as the addition of the protein GABP $\alpha$  or knock down of the protein PEA3. The addition of TSA to the 14fp cells results in increased H3K27ac, an activating histone mark (Creighton et al., 2010), at the ZRS. This does not occur when either ES cells or control cells from the mandibular region (MD cells) of the E11.5 mouse are treated with TSA. This suggests that TSA is activating the *Shh* gene directly through the ZRS.

By studying both *Shh* expressing and non-expressing cells the following questions could be tested:

- How compact is the *Shh* locus compared to other regions of similar size?
- Is compaction of the locus altered when *Shh* is expressed?
- Is there an increased co-localisation between the ZRS and *Shh* gene in *Shh* expressing cells?
- Does the ZRS make different contacts within the *Shh* regulatory region upon *Shh* expression?

However, before attempting to answer such questions the use of these *Shh* inducible cell lines as a model system to study the limb bud was to be further examined. In particular RNA sequencing experiments needed to be conducted to ensure genes expressed within the cell lines were similar to those expressed within the E11.5 limb bud. Using RNA obtained from both TSA treated and untreated (*Shh* expressing and non-expressing) cell lines and proximal and distal (containing the ZPA where *Shh* is expressed) limb sections the following questions could be tested:

- Do the untreated cell lines express similar genes to the limb sections?
- Do the TSA treated cell lines express similar genes to the distal limb bud?
- Does TSA alter other pathways other than those involved in limb initiation and development?

# **Chapter 2**

---

## **Materials and Methods**

## 2.1 Materials

### Chemicals and Reagents

- Trizol Reagent (Life Technologies, cat no. 15596026)
- Glycogen (Roche, cat no. 10901393001)
- Lightcycler 480 SYBR Green I Master Mix (Roche, cat no. 0488735200)
- DMEM (Life Technologies, cat no. 41966029)
- Paraformaldehyde (Sigma Aldrich, cat no. P6418)
- Triton X-100 (Sigma Aldrich, cat no. 9002-93-1)
- Chloramphenicol (Sigma Aldrich, cat no. C0378)
- Phenol/chloroform/isomayl alcohol (25:24:1 (vol/vol/vol); pH 8; Sigma-Aldrich, cat. no. 77617)
- Chloroform/isoamyl alcohol (24:1 (vol/vol); Sigma-Aldrich, cat. no. C0549)
- RNaseA (Roche, cat no.10 109 169 001)
- Glucose (AnalR, cat no. 101174Y)
- Tris(hydroxymethyl)aminomethane (AnalR, cat. No 103156X)
- Potassium Acetate (Fisher Scientific, cat. no. 127-09-3)
- Glacial Acetic Acid (Sigma Aldrich, cat no.33209)
- Magnesium Sulfate ( $Mg_2SO_4$ ) (Sigma Aldrich, cat no. 10034-99-8)
- Dithiothreitol (DTT)
- Formamide (Sigma Aldrich, cat. no. F7503)
- Cot I (Fisher Scientific, cat. no. 18440016)
- Sonicated Salmon Sperm (Sigma Aldrich, cat. no. 68938-01-2)
- Dextan Sulfate (Sigma Aldrich, cat. no. 42867)
- Tween 20 (Sigma Aldrich, cat. no. P1379)
- Marvel Dried Skimmed Milk
- DAPI (Sigma Aldrich, cat. no. D8417)
- Formaldehyde solution 37 % (vol/vol) (Merck, cat. no 1039992500)
- Glycine (Fisher Scientific, cat. no. G/0800/60)
- Hind III (Roche, cat. no. R0104M, NEB, cat. no. R3104S)
- T4 DNA ligase (NEB, cat. no. M0202M)
- T4 DNA ligase buffer (NEB, cat. no. B0202S)
- Sodium Acetate (Fisher Scientific, cat. no. S/2080/60)
- Igepal CA-630 (Sigma Aldrich, cat. no. I8896)

- Complete EDTA free protease inhibitor tablets (Roche, cat. no. 11873580001)
- MluCI (NEB, cat. no. R0538L)
- Salmon testis DNA (STD) (Sigma Aldrich, cat. no. D7656)
- Taq ligase (NEB, cat. no. M0208L)
- 10 × Taq ligase buffer (NEB, cat. no. B0208S)
- Proteinase K (Roche, cat. no. 03115836001)
- Trisodium Citrate (Fisher Scientific, cat. no. S/3320/60)
- Methanol (Fisher Scientific, cat. no. M/4000/17)
- Potassium Chloride (Sigma Aldrich, cat. no. 31248)
- Colcemid (KaryoMAX, cat. no. 15210-040)
- Trichostatin A (TSA) (Cayman Chemical Company, cat. no. 89730)
- Dimethyl sulfoxide (DMSO) (Sigma Aldrich, cat. no. D2650)
- dNTPs (Fisher Scientific, cat. no. 10297018)
- Propan-2-ol (Fisher Scientific, P/7500/15)
- Ethanol (Fisher, E/0650DF/15)
- Bovine Serum Albumin (BSA) (Sigma Aldrich, cat. no. A3294)
- Taq DNA polymerase (NEB, cat. no. M0273S)
- Vectashield (Vector Laboratories, cat. no. H-1200)
- Hi-Pure low EEO Agraose (Biogene, cat. no. 300-300)
- DnaseI (Roche, cat. no. 10104159001)
- DNA polymerase 1 (Fisher, EP0041)

### **Kits**

- QuantiTect Reverse Transcription Kit (Qiagen, cat no. 205313)
- MinElute PCR purification kit (Qiagen, cat no. 28004)
- Lonza-Mycoalert Mycoplasma Detection Kit (Lonza, cat no. LT07-318)
- QIA quick PCR Purification Kit (Qiagen, cat no. 28106)
- Expand Long Template PCR System (Roche, cat no. 11681834001)

### **Equipment**

- 22 × 22 mm coverslip (Fisher Scientific, cat. no. 12-541B)
- 22 × 40 mm coverslip (Fisher Scientific, cat. no. 12-543A)

- 22 × 50 mm coverslip (Fisher Scientific, cat. no. 12-543C)
- 40 µM cell strainer (BD Falcon, cat. no. 352340)
- quadriPERM 16 × 12 chamber (REF 946077308)
- Quick Spin Columns for Radiolabelled Purification (Roche, cat. no. 11273965001)
- Sorvall Legend X1R centrifuge
- Eppendorf® Safe-Lock microcentrifuge tubes (Sigma Aldrich, cat. no. T9661)
- Thermo Scientific Heraeus Fresco 21 Centrifuge
- DNA Engine Tetrad 2 Peltier Thermal Cycler
- Grant Instruments Grant 26 Ltr Expert Unstirred Water Bath Stainless Steel
- Thermo Scientific Savant DNA 120 Speedvac Concentrator
- Peq Power 300V Peqlab Power supply
- Mettler PM4600 DeltaRange Balance
- Spectrolinker™ XL-1000 Series UV Crosslinker
- LightCycler® 480 Instrument II (Roche)

### **Antibodies**

- SHH Antibody [RM0128-4A37] (AB86462) (Abcam)
- Alexa Fluor® 488 (Thermo Scientific, cat no. A-11094)
- FITC Anti-Digoxigenin (Sigma Aldrich, cat no. 11207741910)
- FITC Anti-Sheep
- Biotinylated Anti-Avidin (Vector Laboratories, cat no. BA-0300)
- Avidin Texas Red (Vector Laboratories, cat no. A-2016)
- Rhodamine Anti-Digoxigenin
- Texas Red Anti-Sheep (Vector Laboratories, cat no. TI-6000)
- Avidin FITC (Vector Laboratories, cat no. A-2011)

## 2.2 General Laboratory Solutions

The following solutions were made by the Human Genetics Unit Technical Services department and used throughout this investigation:

- **20 × SSC:** For a 1 L solution - 175.3 g Sodium Chloride, 88.2 g Sodium Citrate, made up in distilled water
- **50 X TAE:** For a 1 L solution – 242 g Tris Base, 57.1 ml Glacial Acetic Acid, 100 ml 0.5 M EDTA (pH 8.0), made up in distilled water
- **0.5 M EDTA:** For a 1 L solution – 186.1 g EDTA, ~20 g Sodium Hydroxide, made up in distilled water
- **5 M NaOH:** For a 1 L solution – 200 g Sodium Hydroxide, made up in distilled water
- **5 M NaCl:** For a 1 L solution – 292.2 g Sodium Chloride, made up in distilled water
- **Trypsin:** For a 1 L solution – 2g trypsin 1:250, 5 ml Phenol Red, 0.06 g Penicillin, 0.13 g Streptomycin, made up in PBS, pH to 7.8 with sodium hydrogen carbonate
- **Versene:** For a 1 L solution – 10 Dulbecco tablets, 0.4 g Sodium EDTA, 5 ml 0.2 % Phenol Red, made up in distilled water
- **PBS:** For a 1 L solution – 10 Dulbecco tablets, 0.1 g Sodium Azide, 20 g Bovine Serum Albumin, made up in distilled water
- **L-Glutamine:** For a 1 L solution – 30 g L-Glutamine, made up in distilled water
- **Penicillin/Streptomycin:** For a 1 L solution - 7.0 g (10 x 10<sup>8</sup> U) Penicillin, 13 g Streptomycin, made up in distilled water
- **L-Broth:** For a 1 L solution – 10 g Tryptone, 5 g Yeast Extract, 10 g Sodium Chloride, 1.0 g Glucose, made up in distilled water

## **2.3 General Laboratory Procedures**

### **2.3.1 Isolation of RNA**

RNA was isolated from cells with trizol reagent using the protocol provided by life technologies. This procedure could be carried out on cells immediately or alternatively cells could be isolated and stored in the appropriate volume of trizol at -20 °C until needed. In chromosome conformation capture experiments (4C and 5C) and in FISH experiments,  $1 \times 10^6$  cells were isolated from the rest of the cells needed to conduct the experiment and stored in 500  $\mu$ l of trizol reagent. RNA could therefore be isolated when required allowing the subsequent production of cDNA which could be used for quantitative real time PCR experiments.

In brief, for a 500  $\mu$ l trizol sample, the protocol involved leaving the solution at room temperature for 5 min followed by the addition of 100  $\mu$ l of chloroform. After thorough mixing and a further short incubation at room temperature, samples were centrifuged at 12,000 g for 15 min at 4 °C thereby producing a solution with three separate layers: a lower organic phase; an interphase; and an aqueous phase on top which contained the isolated RNA. This aqueous phase was removed and added to 10  $\mu$ g of glycogen (20 mg/ml) and 250  $\mu$ l of propan-2-ol and then centrifuged for a further 10 min at 12,000 g (4 °C). Subsequently the supernatant was removed and the sample was washed in 75 % ethanol, centrifuged for 5 min (7500 g, 4 °C) and then left for 10 min at room temperature. Finally, the remaining RNA pellet was re-suspended in 50  $\mu$ l of RNase free water and heated at 55 °C for 15 min. At this point the RNA solution was either stored at -80 °C or placed on ice and used immediately to make CDNA.

### **2.3.2 Assessment of RNA purity, quantity and integrity**

To assess the integrity of RNA samples, 1  $\mu$ l of a 50  $\mu$ l sample was run on an ethidium bromide stained 1 % agarose gel. RNA was noted to be of an acceptable quality if two dominant bands were observed representing 28S (larger molecular weight) and 18S (lower molecular weight) cytoplasmic rRNA, where the intensity of the 28S band was roughly double that of the 18S band.

The purity and quantity of RNA was determined using a NanoDrop ND-1000 spectrophotometer. Acceptable RNA samples displayed an A260/280 ratio of ~ 2.0 and A260/230 ratio of ~ 2.0-2.2. Sometimes noticeable absorbance peaks were observed around 230 nm which may be the presence of contaminating trizol within the sample which had not

been removed from a previous stage. In these cases a further ethanol precipitation was conducted where 3  $\mu$ l of sodium acetate pH 5.2 and 1  $\mu$ l of 20 mg/ml glycogen were added to the 50  $\mu$ l RNA sample in addition to 2.5 volumes of 100 % ethanol and left overnight at -20 °C. The following day the samples were centrifuged at 12,000 g for 10 min, the supernatant removed and the sample re-suspended again in 20  $\mu$ l - 50  $\mu$ l RNase free water. Samples were subsequently analysed on the NanoDrop spectrophotometer.

### **2.3.3 Reverse Transcription Reaction**

Reverse transcription was conducted using either the QuantiTect Reverse Transcription Kit (Qiagen) or the first strand cDNA synthesis kit (Roche) which is no longer available. Using the Qiagen kit consists of two separate reactions: a Genomic DNA Elimination Reaction and a Reverse-transcription reaction. The Genomic DNA Elimination Reaction involves incubating up to 1  $\mu$ g RNA in 7  $\times$  gDNA Wipeout Buffer and RNase free water in a 14  $\mu$ l reaction at 42 °C for 2 min. This mixture is then added to a reverse transcription master mix consisting of the following components: Quantiscript Reverse Transcriptase; 5  $\times$  Quantiscript RT Buffer; and RT primer mix. Tubes were then incubated at 42 °C for 15 min followed by a further incubation at 95 °C for 3 min. 30  $\mu$ l of RNase free water was added to each reaction to give a total of 50  $\mu$ l of cDNA which could then be used for quantitative real-time PCR (qPCR).

When preparing samples for qPCR, replicate reverse transcription reactions were conducted without the addition of the reverse transcription enzyme. These served to identify any contaminating genomic DNA in the samples tested.

### **2.3.4 Quantitative Real-Time PCR (qPCR)**

Quantitative real-time PCR (qPCR) was conducted using the Lightcycler 480 SYBR Green I Master protocol (Roche). In brief a master mix was made up for each series of primers used in the qPCR. 18  $\mu$ l of master mix were added to the appropriate positions in a 96 multiwell plate, followed by 2  $\mu$ l of sample cDNA. The master mix components for a single 20  $\mu$ l reaction are given in Table 2.1. The programme for the Lightcycler 480 Instrument is given in Table 2.2. A melt curve was performed for each qPCR experiment to ensure unwanted material, such as primer dimers, were not analysed. GAPDH was used as the standard housekeeping gene in all relative qPCR experiments. Lightcycler 480 software v 1.5 was used to analyse qPCR results. The advanced relative quantification setting was selected which uses



the  $\Delta\Delta CT$  method to calculate relative fold gene expression (assuming efficiency (E)=2). The primers used in qPCR reactions are detailed in Table 2.3.

| <u>Component</u>                     | <u>Reaction Volume (<math>\mu</math>l)</u> |
|--------------------------------------|--------------------------------------------|
| LightCycler® 480 SYBR Green I Master | 10                                         |
| 5 $\mu$ M Primer Mix                 | 1.2                                        |
| RNase free water                     | 6.8                                        |
| <b>Total</b>                         | <b>18</b>                                  |

**Table 2.1 – Volume of each component needed for a 1 × reaction master mix. 18  $\mu$ l of master mix are added per well in 96 multiwell plate in addition to 2  $\mu$ l of cDNA.**

| <u>Stage</u> | <u>Process</u>        | <u>Temperature (<math>^{\circ}</math>C)</u> | <u>Time (min:sec)</u> | <u>No. Cycles</u> |
|--------------|-----------------------|---------------------------------------------|-----------------------|-------------------|
| <b>1</b>     | <b>Pre-incubation</b> | 95                                          | 10:00                 | 1                 |
| <b>2</b>     | <b>Amplification</b>  | 95                                          | 00:10                 | 50                |
|              | <b>Melting</b>        | Primer Dependent                            | 00:30                 |                   |
|              | <b>Annealing</b>      | 72                                          | 00:30                 |                   |
| <b>3</b>     | <b>Melt Curve</b>     | Primer Dependent                            | 00:15                 | 1                 |
|              |                       | 95                                          | Continuous            |                   |

**Table 2.2 – Lightcycler 480 Instrument programme for relative quantification qPCR.**

| <b>Primer</b> | <b>Sequences</b>                                |
|---------------|-------------------------------------------------|
| <i>Shh</i>    | TCCACTGTTCTGTGAAAGCAG<br>GGGACGTAAGTCCTTCACCA   |
| <i>Gapdh</i>  | GGGTCCTATAAAATACGGACTGC<br>CCATTTTGTCTACGGGACGA |
| <i>Gabpa</i>  | CGGGGAGAAATTCTTTGGA<br>CTTGGCTGGCCCCAAAACATA    |
| <i>Pea3</i>   | CAGCAGGAAGCCACCACT<br>GGACTTGATGGCGATTTGTC      |

**Table 2.3 – Primer sequences used for qPCR experiments**

### **2.3.5 Phenol Chloroform Extractions**

Isolation of DNA was conducted by phenol/chloroform extractions followed by ethanol precipitations. In brief, an identical volume of phenol/chloroform/isomayl alcohol (25:24:1) (Sigma) was added to the experimental sample in a 1.5 ml tube, shaken vigorously for 15 s and centrifuged at 13,000 rpm for 5 min. The solution separates into 2 layers with the aqueous phase forming the top layer and the organic phase the lower layer. DNA maintained within this top aqueous layer was carefully removed and placed in a fresh tube. The organic phase was discarded in designated phenol waste containers.

### **2.3.6 DNA Ethanol Precipitations**

The aqueous phase from Chapter 2.3.5 was used to perform ethanol precipitations. To 100 µl aqueous solution the following components were added: 1 µl glycogen (20 mg/ml); 10 µl 3 M Sodium Acetate pH 5.2; and 250 µl ethanol. This solution was mixed thoroughly and left on dry ice for 15 min before centrifuging at 13000 rpm for 30 min (4 °C). Next, the supernatant was removed leaving a DNA pellet which was washed in 70 % ethanol and centrifuged at 13000 rpm for 2 min (4 °C). This process was repeated, removing the supernatant and washing the pellet again in 70 % ethanol. After centrifugation and removal of the supernatant the DNA

pellet was left to dry for 5 min after which it was re-suspended in MilliQ water and used as appropriate. (Typically DNA was analysed by agarose gel electrophoresis.)

### **2.3.7 Agarose Gel Electrophoresis**

0.6 % - 2.5 % agarose gels (w/v) were prepared in Tris-Acetate-EDTA (TAE) buffer. 10 mg/ml ethidium bromide was added to visualise DNA bands under UV light. Samples were added in loading buffer prepared in-house consisting of 15 % Ficol (Fisher) and Orange G (Sigma). Gels were run at 100 V for 1 hr.

### **2.3.8 Generation and Statistical Analysis of RNA-seq data**

RNA was prepared on-site and tested for purity and integrity as described in section 3.3.1. Next generation sequencing was conducted by GATC Biotech AG. Analysis of RNA-seq data was performed using the Tuxedo suite on the Galaxy main server (section 3.2.1).

## **2.4 Fluorescent *in situ* hybridisation (FISH)**

### **2.4.1 Cell Culture and Cell Fixation**

Cells were first grown to confluency in a T75 flask in complete DMEM media containing 10 % FCS; 1 % Pen/Strep; 1 % L-Glutamine; and 0.1 % Interferon Gamma. Cells were subsequently washed in 5 ml PBS and trypsinised for 15 min at 33 °C. 10 × complete DMEM media was subsequently added to neutralise the effects of the trypsin and the cells centrifuged at 1200 rpm for 5 min. 1 ml of complete DMEM media was added to the cell pellet and an appropriate volume of cell suspension added to a superfrost slide in a quadriPERM 16 × 12 chamber. 5 ml of complete DMEM media were added to each slide and the cells were allowed to adhere to the surface of the slide and grow overnight. The following day, the media for each slide was replaced with 5 ml of fresh DMEM media containing the appropriate concentration of TSA or DMSO. Cells were then incubated for the appropriate time at 33 °C as dictated by the experimental protocol.

Before fixation, cells were washed for three min in PBS three times. Slides were then transferred to 4 % paraformaldehyde (PFA) solution and left for 10 min. PFA was removed and the slides washed again in PBS three times for three min each. After the washes the slides were transferred to a 0.5 % Triton X-100/PBS solution and left for 10 min. Finally, slides were washed in PBS in the same manner as previously mentioned and left to dry. After 10 min, slides were washed briefly in MilliQ water and left to dry at room temperature. Slides containing fixed cells could either be used immediately or stored at -80 °C until needed.

### **2.4.2 Purification of Fosmid DNA**

An overnight culture was set up for each fosmid glycerol stock. Each culture consisted of 5 ml L-Broth containing 12.5 ng/μl chloramphenicol. Cultures were left overnight at 37 °C. For each fosmid being tested, 1.5 ml of culture media was extracted and spun at maximum speed in a 4 °C table-top centrifuge for 30 seconds. The supernatant was removed and replaced with 200 μl GTE Buffer containing lysozyme. The pellet was re-suspended in the GTE Buffer, vortexed and left at room temperature for 5 min. 400 μl of lysis buffer was subsequently added to this solution and mixed by inversion. At this point the solution should go clear. This was left on ice for a further 5 min. 300 μl of acetate buffer was subsequently added producing masses of white flocculent. This was left on ice for 5 min followed by centrifugation at maximum speed in a table-top centrifuge at 4 °C for 5 min. The supernatant was removed and

added to an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1, pH 8.0). The solution was mixed thoroughly and spun at maximum speed in a table-top centrifuge at 4 °C for 4 min. The aqueous upper layer was removed and placed in a new Eppendorf tube along with an equal volume of chloroform:isoamyl alcohol (24:1). The solution was mixed and centrifuged as described for the above phenol:chloroform:isoamyl alcohol extraction. The top aqueous layer was extracted and placed in a new tube with an equal volume of iso-propanol. This was stored at -20 °C for 1 hr and then centrifuged at 4 °C for 15 min at maximum speed. After the supernatant was removed the pellet was washed in 70 % ethanol, briefly centrifuged at maximum speed at 4 °C and left to dry for 5 min. This pellet was re-suspended 25 µl TE Buffer after which 2 µl 10 mg/ml RNaseA was added and the solution heated for 5 min in a 37 °C water bath. The final solution could either be stored at -20 °C or used directly in nick translation reactions. The composition of the buffers used in this section are given in Table 2.4.

| <b><u>Buffer</u></b>  | <b><u>Components</u></b>                          |
|-----------------------|---------------------------------------------------|
| <b>GTE Buffer</b>     | 50 mM Glucose, 25 mM Tris pH 8.0, 10 mM EDTA      |
| <b>Lysis Buffer</b>   | 0.2 M NaOH, 1 % SDS                               |
| <b>Acetate Buffer</b> | 3 M potassium acetate, 11.5 % glacial acetic acid |

**Table 2.4 - List of Buffers used in the purification of fosmid DNA.**

### **2.4.3 Nick Translation Reactions (NTR)**

A nick translation reaction was used to introduce either biotin or digoxigenin into each fosmid. Each 20 µl reaction contained 0.0625 mM dATP, dCTP and dGTP in addition to either 0.0625 mM bio-16-dUTP or 0.0375 mM digoxigenin-11-dUTP. (When digoxigenin probes were used 0.025 mM dTTP was also added). In addition, 1 µg of purified fosmid DNA (Chapter 2.4.2), 1 µl 1:10 DNase 1 (Roche), 1 µl DNA polymerase 1 (Invitrogen) and 2 µl nick translation salts were added to the reaction mixture. This was left at 16 °C for 90 min after which 2 µl 20 %

SDS and 3  $\mu$ l 0.5 M EDTA were added. The solution was then made up to 90  $\mu$ l in TE buffer and purified using Pharmacia Quick spin columns according to the manufacturer's instructions.

Nick Translation Salts (0.5 M Tris pH 7.5, 0.1 M  $Mg_2SO_4$ , 1 mM DTT, 0.5 mg/ml BSA fraction V Sigma).

#### **2.4.4 3D-FISH**

3D-FISH was conducted over two days with denaturing of both slides and probes conducted on the first day, an annealing stage overnight and both washing and detection conducted the following day. Fixed slides were removed from  $-80\text{ }^\circ\text{C}$  and washed briefly in  $2 \times \text{SSC}$ . The slides were then transferred to  $2 \times \text{SSC}$  containing 100  $\mu\text{g/ml}$  RnaseA and left for 1 hr at  $37\text{ }^\circ\text{C}$ . After incubation, the slides were washed briefly in  $2 \times \text{SSC}$  and then washed sequentially in 70 %, 90 % and 100 % ethanol solutions for 2 min each, after which the slides were left to air dry at room temperature. Once dry, the slides were heated at  $70\text{ }^\circ\text{C}$  for 5 min in an oven then denatured for between 15 – 30 min (experiment dependent) at  $80\text{ }^\circ\text{C}$  in 70 % formamide, pH 7.5. Subsequently, the slides were placed immediately into ice cold ethanol for 2 min then washed in 90 % and 100 % ethanol respectively (2 min per wash) and left to dry at room temperature. All slides and coverslips were warmed on a hot plate set to  $37\text{ }^\circ\text{C}$  before the addition of the probes.

To prepare the probes, 100 ng of each fosmid was added to 14  $\mu\text{g}$  CotI, 5  $\mu\text{g}$  sonicated salmon sperm DNA and 2 volumes of 100 % ethanol in a 1.5 ml Eppendorf tube. This was repeated for each slide being examined. These tubes were then transferred to a heated vacuum centrifuge and spun until no liquid remained. In the meantime, a hybridisation solution was made up consisting of 50 % deionised formamide,  $2 \times \text{SSC}$ , 20 % dextran sulfate and 1 % Tween 20. 30  $\mu$ l of this hybridisation solution was added to each tube containing the dried probe mixture, mixed well and left for 1 hr at room temperature. To denature the probes these tubes were transferred to an  $80\text{ }^\circ\text{C}$  water bath and left for 5 min. Subsequently, probes were transferred to a  $37\text{ }^\circ\text{C}$  water bath to pre-anneal for 15 min. 15  $\mu$ l of pre-annealed probe solution was added to a  $22 \times 40\text{ mm}$  coverslip and then added to a pre-warmed slide which was then sealed with rubber solution and placed in a covered tray in a  $37\text{ }^\circ\text{C}$  water bath overnight.

The following day the slides were transferred to  $2 \times \text{SSC}$  and washed at  $45\text{ }^\circ\text{C}$  for 3 min. This was repeated 3 more times. Slides were then transferred to  $0.1 \times \text{SSC}$  and washed at  $60\text{ }^\circ\text{C}$  for 3 min with this washing procedure also repeated an additional 3 times. Finally, slides were transferred to  $4 \times \text{SSC}$ , 0.1 % Tween 20 at room temperature and left for several min. Before

antibody solutions were added blocking solution was first added to the slides (4× SSC, 5 % marvel) for 5 min on a 22 × 50 mm coverslip. This was removed, replaced with the first antibody solution (Table 2.5) and the slides placed in a moistened chamber and incubated at 37 °C for 30 min. After this incubation the coverslips were removed and slides washed three times, 2 min each time in 4 × SSC, 0.1 % Tween 20 at 37 °C. Three more antibody incubations were conducted in the order detailed in Table 2.5, with each incubation lasting 30 min and conducted in a moistened chamber at 37 °C. Again slides were washed three times for 2 min each time in 4 × SSC, 0.1 % Tween 20 at 37 °C in between antibody incubations. After the final antibody incubation, slides were added to 4 × SSC, 0.1 % Tween 20 containing 50 ng/ml DAPI for 5 min. Slides were then removed and mounted in 25 µl vectashield on a 22 × 50 mm coverslip and sealed with Pang Rubber Solution. All FISH slides can be stored for several weeks and analysed when required if stored in the dark at 4 °C.

| <u>Order</u> | <u>Combination 1</u>                           | <u>Combination 2</u>                           |
|--------------|------------------------------------------------|------------------------------------------------|
| <b>1</b>     | 1:20 FITC Anti-Digoxigenin                     | 1:20 Rhodamine Anti-Digoxigenin                |
| <b>2</b>     | 1:100 FITC Anti-Sheep + 1:500 Avidin Texas Red | 1:100 Texas Red Anti-Sheep + 1:500 Avidin FITC |
| <b>3</b>     | 1:100 Biotinylated Anti-Avidin                 | 1:100 Biotinylated Anti-Avidin                 |
| <b>4</b>     | 1:500 Avidin Texas Red                         | 1:500 Avidin FITC                              |

**Table 2.5 - The order of antibody incubations used in FISH experiments.**

#### **2.4.5 Deconvolution Microscopy**

All images were acquired using a Zeiss Axioplan 2 system referred to in-house as “Granny” containing a Lumen 200 W metal halide light source (Prior Scientific Instruments) and Chroma #89014ET single excitation and emission filters (Chroma Technology Corporation). The system contains a PIFOC collar (PIFOC model P-721, Physik Instrumente, Karlsruhe, Germany) allowing acquisition of fluorescent images at 0.2 µm intervals through the z-plane. All FISH images were captured at 100 × magnification with a Hamamatsu Orca AG CCD

camera (Hamamatsu Photonics, Welwyn Garden City, UK). These were acquired using Volocity software (PerkinElmer, Waltham, MA, USA) with all images deconvolved using Volocity as well (Williamson, 2012). To do this calculated point spread functions (PSFs) for DAPI, FITC and TxRED were used as inputs for Volocity's iterative restoration procedure. To measure the distance between fluorescent spots Volocity's quantitation module was used. This placed each image in the xy plane and allowed users to manually scroll through the z axis to identify the position at which the fluorescence of each probe was at its maximum. A line could then be created from the centroid of one probe dissecting the z axis to that of another probe and the distance of this line measured. Over 100 measurements were collected for each FISH experiment.

#### **2.4.6 Preparation of cells for metaphase spreads.**

Metaphase spreads were made for 14fp, 88fp and MD cell lines. Cells were grown in complete DMEM media until 80 % confluent with a 1:100 dilution (stock 10 µg/ml) of colcemid (Karyomax) added 45 min before use. After removal of the media, trypsin was added for 15 min and cells incubated at 33 °C. × 10 volume of complete DMEM media was added and cells centrifuged for 5 min at 1200 rpm. The media was removed and cells washed in PBS before being centrifuged again for an additional 5 min at 1200 rpm. After removal of wash, cell pellets were re-suspended in 10 ml lysis buffer (0.5 % trisodium citrate/0.25 % potassium chloride) and left for 10 min in a 37 °C water bath. Following this step, cells were spun down as before and fix solution (3:1 methanol:acetic acid) added dropwise to the cell pellet with constant vortexing. Cells were spun down as before and the fix solution removed. Fresh fix solution was added again as described and the process repeated 3 times. Finally fixed cells were dropped onto microscope slides and allowed to dry before use. The remaining solution could be stored at -20 °C until needed. 2D images were obtained using the Zeiss Axioplan 2 microscope described in Chapter 2.4.5.

#### **2.4.7 Preparation of cells for 2D-FISH.**

MYtag experiments were conducted on 14fp cells using a 2D-FISH protocol. Preparation of cells for these experiments was conducted as described in Chapter 2.4.6 with the omission of the first step where colcemid is added. ES cells were obtained from the Laura Lettice and Fiona Kilanowski (Hill laboratory) and were prepared for MYtag experiments as Chapter 2.4.6 without the addition of colcemid.



### 2.4.7 2D-FISH

2D-FISH was conducted for MYtag experiments. In brief, the 3D-FISH protocol was followed (Chapter 2.4.4) with several alterations. Firstly, MYtags were not spun down with CotI and sonicated salmon sperm in a vacuum centrifuge. Instead the hybridisation mix was made as previously described, with 1  $\mu$ l of each MYtag library added to 10  $\mu$ l of this solution. Unlike the 3D-FISH procedure the hybridisation mix was not left for 1 hr. Instead after addition of the MYtags, the mix was immediately transferred to a 70 °C water bath and denatured for 5 min. After this step, the mix was placed directly on ice. Slides were RNase treated and washed in ethanol as previously described. However, denaturation was conducted in 70 % formamide, pH 7.5 for only 2 min 15 sec in a 70 °C water bath. Slides were then washed in 70 %, 90 % and 100 % ethanol as before. 12  $\mu$ l of denatured hybridisation mix was added to a 22  $\times$  22 mm coverslip and incubated overnight on the denatured slides as previously described. The following day slides were washed 5 times in 2  $\times$  SSC at 45 °C for 2 min each time. This was followed by five 2 min washes in 0.1  $\times$  SSC at 60 °C. Slides were washed in DAPI and mounted in vectashield as 3D-FISH.

## 2.5 Chromosome Conformation Capture Techniques

### 2.5.1 Preparation of 3C libraries

3C libraries were made as detailed in the protocol published by the Stoler laboratory (Stadhouders et al., 2013) and were subsequently used in either 4C or 5C experiments (Figure 2.1). Between  $1 \times 10^6$  and  $1 \times 10^7$  cells were used as a starting point for these experiments. For the cell lines, cells were grown in culture and either treated with TSA or DMSO (as a control) and left for the length of time dictated by the experiment. After this, cells were washed in 5 ml PBS and trypsinised for 15 min at 33 °C.  $10 \times$  complete DMEM media was subsequently added to neutralise the effects of the trypsin after which cells were passed through a 40  $\mu$ m cell strainer to provide a single cell solution which was then centrifuged at 1200 rpm for 5 min. The media was removed leaving the cell pellet which was re-suspended in 1 ml 10 % FCS/PBS. Using a haemocytometer, the number of cells were counted. At this point, a fraction of these cells (usually  $1 \times 10^6$  cells) were removed, spun down at 1200 rpm for 5 min and re-suspended in 500  $\mu$ l trizol solution. This was stored at -20 °C until needed. This aliquot was used to make RNA and then cDNA which was tested by qPCR for *Shh* expression. In this way the effect of TSA on the cells could be examined in each 4C/5C experiment.

The remaining cell suspension was made up to 12 ml in 10 % FCS/PBS after which 649  $\mu$ l 37 % formaldehyde solution was added (2 % final formaldehyde solution) with the tubes rotated gently for 10 min. After this, glycine solution at a final concentration of 0.125 M (made up in PBS) was added to quench the reaction which was then centrifuged at 340 g for 8 min at 4 °C. As described in the protocol, the cell pellet was then washed in PBS and centrifuged again as detailed above. In initial experiments, a large number of aggregates were formed at the stage of HindIII digestion. As a result, the length of time cells were exposed to lysis buffer was increased and the cells were also dounced. In total, cells were left in 5 ml lysis buffer (Table 2.6) for 15 min after which they were dounced 40 times and then spun down at 650 g for 5 min. After removal of lysis buffer, nuclei were once again washed in PBS, spun down at 650 g for 5 min and then frozen in liquid nitrogen before storing at -80 °C.

HindIII digestions were conducted by combining aspects of the protocols provided by the Stoler (Stadhouders et al., 2013) and de Laat (van de Werken et al., 2012). If more than  $5 \times 10^6$  cells were used in the experiment, nuclei were made up to 2 ml in  $1.2 \times$  HindIII restriction

buffer and then divided into 4 separate tubes each containing 500  $\mu$ l of solution. HindIII digestions were performed on each of these tubes which were later combined before the first ligation. SDS was added to each of these tubes to a final concentration of 0.2 %, with the tubes then shaken on a thermomixer at 1400 rpm for 1 hr. This was followed by the addition of Triton X-100 to a final concentration of 2 % (vol/vol) and further shaking for 1 hr (1400rpm). Both NEB and Roche HindIII restriction enzymes were used interchangeably in this process. 200 U HindIII was first added for 4 hr (37 °C, 1400 rpm), followed by the addition of a further 200 U with overnight incubation. The next day, another 200 U was added for 4 hr and then another 200 U for an additional 4 hr. Thus 800 U HindIII was added overall.

To check the digestion was conducted to completion, 2 x 5  $\mu$ l aliquots were taken from each of the tubes prior to digestion (undigested control) and a further 2 taken after digestion (digested control). One of each of the digested and undigested controls were combined with 90  $\mu$ l of 10 mM Tris-HCl (pH 7.5) and 2.5  $\mu$ l 20 mg/ml proteinase K and then incubated at 65 °C for 4 hr followed by a phenol/chloroform extraction and ethanol precipitation. These were referred to as the de-crosslinked digested and undigested controls. The remaining undigested and digested controls were made up to 100  $\mu$ l in 10 mM Tris-HCl (pH 7.5) and also subjected to a phenol/chloroform extraction and ethanol precipitation. These controls were referred to as the crosslinked digested and undigested controls. All samples were run on a 0.6 % agarose gel and compared. Crosslinked controls were not visible on the gel, while the de-crosslinked undigested control appeared as a high molecular weight band appearing at around 12 kb. The de-crosslinked digested controls appeared as a smear running from a high molecular weight to a lower molecular weight of around 4 kb. Providing the control experiments looked as expected the protocol was continued. In the rare occasions where HindIII digestion was incomplete additional enzyme was added (typically 400 U) overnight and the digested controls analysed again as described.

To halt HindIII digestion, a final concentration of 1.6 % SDS was added to each of the 500  $\mu$ l aliquots for 25 min at 65 °C on a thermomixer (1400 rpm). At this point, each of the 4 500  $\mu$ l aliquots were combined and 24.5 ml of 1.15 x T4 DNA ligase ligation buffer added, followed by the addition of Triton X-100 to a final concentration of 1 % (performed in 50 ml tubes.) Tubes were incubated in a 37 °C water bath for 1 hr followed by the addition of 26800 CEU T4 DNA ligase (NEB) and overnight incubation at 16 °C. At this stage, 100  $\mu$ l solution was removed and 2.5  $\mu$ l (20 mg/ml) proteinase K added at 65 °C for 4 hr. DNA was then purified by performing a phenol:chloroform extraction followed by an ethanol precipitation. The final

DNA pellet was made up to 20  $\mu$ l in MilliQ water and run out on a 0.6 % agarose gel. Successful 3C libraries appeared as a high molecular weight band while libraries with incomplete ligation still appeared as a high molecular weight smear. In cases where ligation was incomplete additional ligase was added together with ATP and the solution left at 16 °C again overnight.

1200  $\mu$ g RNase A was then added to the solution which was left in a 65 °C water bath overnight to de-crosslink the DNA. The following day, the solution was again split into 4 equal aliquots of roughly 7 ml to which an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) was added. After vigorous mixing, tubes were centrifuged at 3200 g for 15 min at RT. After careful removal of the phenol layer, the following additions were made to the remaining solution in each of the 4 tubes: 7 ml MilliQ Water; 1.5 ml 2M Sodium Acetate (pH 5.6); and 35 ml 100 % ethanol. These were mixed thoroughly and left at -80 °C for 3 hr. Following this incubation, the tubes were centrifuged for 45 min at 3200 g (RT), the supernatant removed and pellet washed in 70 % ethanol, before being centrifuged again for 15 min at 3200 g (4 °C). Upon removing the supernatant from the ethanol wash, the DNA pellets were left to dry for 20 min after which 37.5  $\mu$ l of 10 mM Tris-HCl (pH 7.5) was added to each pellet and the tubes transferred to a 37 °C water bath allowing the DNA pellets to go back into solution. Once all pellets were re-dissolved, the solutions were combined giving a final volume of 150  $\mu$ l final 3C library. When making 3C libraries for further use in 5C, 20  $\mu$ l TE buffer was added to each pellet which were later combined to give 100  $\mu$ l final 3C library. 1  $\mu$ l final 3C libraries were run on 0.6 % agarose gels as a quality control check. Final libraries produced a high molecular weight band showing that all digested HindIII fragments had been successfully re-ligated and purified.

| <b><u>Buffer</u></b> | <b><u>Component</u></b>                                                                                          |
|----------------------|------------------------------------------------------------------------------------------------------------------|
| <b>Lysis Buffer</b>  | 10 mM Tris-HCl (pH 8.0), 10 mM NaCl, 0.2 % (vol/vol) Igepal and 1 $\times$ EDTA free protease inhibitor solution |

**Table 2.6 - Components used to make the lysis buffer used in the 3C protocol.**

### **2.5.2 Preparation of 4C libraries**

Before 3C libraries were digested with a second restriction enzyme, their quantity was assessed by gel densitometry. Increasing concentrations of genomic DNA were run on a 0.6 % gel along with 1  $\mu$ l 3C library. Using imageJ software (<http://imagej.net/Welcome>) the intensity

of each band was determined and then plotted on a graph of intensity vs DNA concentration. The intensity of the 3C library band then allowed an approximate concentration of the 3C library to be established. As described in the Stoler protocol (Stadhouders et al., 2013), around 25 µg 3C library was digested with the 4-cutter restriction enzyme MluCI (using 1 U per µg DNA) overnight at 37 °C where the concentration of 3C library was 100 ng/µl. After digestion, the DNA was purified by a phenol: chloroform extraction and ethanol precipitation. In brief, an identical volume of phenol:chloroform:isoamyl alcohol (25:24:1) solution was added to the digestion reaction, mixed vigorously and spun down at RT for 15 min at 15,800 g. After the phenol layer was discarded, the following solutions were added: 40 mg glycogen; 1/10<sup>th</sup> total reaction volume of 2 M Sodium Acetate (pH 5.6); and 850 µl 100 % ethanol. The solution was mixed thoroughly and added to liquid nitrogen for around 5 s. Upon removal, tubes were centrifuged at 4 °C for 20 min at 15,800 g. Finally, the supernatant was removed and the pellet washed in 70 % ethanol. This also was centrifuged at 4 °C for 5 min at 15,800 g. Upon removing the supernatant, the pellet was dried for 15 min before 100 µl MilliQ water was added and the tubes transferred to a 37 °C water bath for an additional 15 min allowing the pellets to re-dissolve. At this point, 5 µl of sample were removed and run on a 1.5 % agarose gel to assess the quality of the second restriction digest. Successful digests appeared as a smear between 0.3 – 1 kb. Secondary ligations were conducted overnight at 16 °C in a total reaction volume of 14 ml. This consisted of the MluCI digested 3C library, 13400 CEU T4 DNA ligase (NEB) and 1.4 ml 10 × ligase buffer.

The remainder of the protocol was conducted exactly as described by the Stoler laboratory protocol (Stadhouders et al., 2013) to the point of purifying the 4C library. In brief, this involved an additional phenol/chloroform extraction and ethanol precipitation of the ligated sample to eventually produce 150 µl 4C library. DNA purification was conducted using the QIA quick PCR purification kit according to the manufacturer's protocol (<https://www.qiagen.com/gb/>). The 4C library was divided into 4 aliquots and 4 columns were used to purify each of these samples. Purified libraries were later recombined to give a total volume of 300 µl purified 4C library.

### **2.5.3 Amplification of 4C libraries.**

4C libraries were amplified using the expand long template PCR system according to the manufacturer's protocol (Roche). The primers used are provided in Table 2.7 and the PCR protocol provided in Table 2.8. Increasing concentrations of 4C library (25, 50, 100, 200 ng) were first amplified and analysed on a 1.5 % agarose gel to determine which starting

concentration to use in subsequent PCR reactions. (The aim was to find a starting concentration of 4C library which, when amplified, showed a range of products while at the same time ensuring that these products were not amplified to the point of saturation). After choosing a starting concentration of 4C library, PCRs were repeated 16 times and the resulting solutions pooled into a single tube. DNA was purified using the QIA quick PCR Purification Kit (Qiagen).

| <b>Primer</b>  | <b>Sequence</b>     |
|----------------|---------------------|
| HindIII Primer | GGGGAACTGATCACAAGA  |
| MluCI Primer   | CATCTTTTCTTGCAGGTGT |

**Table 2.7 – Primer sequences used for amplification of 4C libraries.**

| <b>Stage</b> | <b>Temperature (°C)</b> | <b>Time (min:sec)</b> | <b>No. Cycles</b> |
|--------------|-------------------------|-----------------------|-------------------|
| 1            | 94                      | 2:00                  | 1                 |
| 2            | 94                      | 0:15                  | 34                |
|              | 58                      | 1:00                  |                   |
|              | 68                      | 3:00                  |                   |
| 3            | 68                      | 7:00                  | 1                 |
| 4            | 4                       | forever               | 1                 |

**Table 2.8 - PCR program used for amplification of 4C libraries**

#### **2.5.4 Sequencing of 4C libraries**

4C libraries were sequenced by the Human Genetics Unit Technical Services department. Libraries were first sheared and sequencing adaptors ligated on. Sequencing was conducted using the Ion Proton™ Sequencer.

## **2.5.5 Preparation of 5C libraries**

### **2.5.5.1 Titration of amplified 3C libraries**

To determine the quantity of 3C library needed to perform 5C reactions, the 3C library was serially diluted into 10 separate aliquots and amplified. In brief, 5 µl 3C library was added to 5 µl MilliQ water in a 1.5 ml tube and mixed thoroughly. 5 µl of this solution was removed and replaced in a second tube already containing 5 µl MilliQ water. In this way the 3C library was first diluted to ½ strength and then to ¼ strength. This process was repeated until 10 tubes containing serially diluted 3C libraries were produced. 4 µl from each tube was removed and placed in separate wells of a 96 well plate on ice. In the meantime, master mix 1 and 2 were prepared according to Tables 2.9 and 2.10. Dilutions of primers and salmon testis DNA (STD) were conducted prior to this and left on ice for 10 min before use. 8.2 µl master mix 1 followed by 10.8 µl master mix 2 were added to the appropriate position in the plate. Lastly 2 µl of the second primer were added. PCR was conducted as Table 2.11. The PCR programme was first started and paused at the first step thereby maintaining the block at 94 °C before the plate was added. The program was resumed when the plate was added. All steps in this procedure were conducted on ice with all reagents kept as cold as possible. This helped reduce the amount of primer dimers in reactions containing sufficient amounts of 3C library. Upon completion of the PCR the products were analysed on a 1.5 % agarose gel. Well-made 3C libraries showed a product size of around 150 bp which decreased in intensity as the dilution factor increased. This correlated with an increase in primer dimer. The sequence of primers used in this PCR are given in Table 2.12.

Using ImageJ software (<http://imagej.net/Welcome>) the relative intensities of the 3C products could be determined allowing a graph of intensity vs volume of sample added (µl) to be plotted. This graph should appear as a straight line showing a direct correlation between intensity and volume of sample which curves off at higher sample volumes. The amount of 3C library used in 5C reactions can be determined from the linear portion of this graph.

| <b>Master Mix 1</b>                   | <b>Volume (µl)</b> |
|---------------------------------------|--------------------|
| 10 × HIFI                             | 2.5                |
| 50 mM Mg <sub>2</sub> SO <sub>4</sub> | 2                  |
| 25 mM dNTPs                           | 0.2                |
| 0.1 mg/ml STD                         | 1.5                |
| 5 µM primer 1                         | 2                  |
| <b>Total</b>                          | <b>8.2</b>         |

**Table 2.9 – Volume of reagents needed for master mix 1 in the titration of amplified 3C libraries protocol.**

| <b>Master Mix 2</b>            | <b>Volume (µl)</b> |
|--------------------------------|--------------------|
| MilliQ Water                   | 10.6               |
| Taq DNA Polymerase (5000 U/ml) | 0.2                |
| <b>Total</b>                   | <b>10.8</b>        |

**Table 2.10 – Volume of reagents needed for master mix 2 in the titration of amplified 3C libraries protocol.**



| Stage | Temperature (°C) | Time (min:sec) | No. Cycles |
|-------|------------------|----------------|------------|
| 1     | 94               | 5:00           | 1          |
| 2     | 94               | 0:30           | 35         |
|       | 65               | 0:30           |            |
|       | 72               | 1:00           |            |
| 3     | 94               | 1:00           | 1          |
|       | 65               | 0:45           |            |
|       | 72               | 8:00           |            |
| 4     | 4                | forever        | 1          |

**Table 2.11 – PCR program used in the titration of amplified 3C libraries protocol.**

| Primer | Sequence                         |
|--------|----------------------------------|
| USP 14 | CAGTGTGCCTCCAAACACCAAACCTAACAACC |
| USP 15 | CAAAGAACAAAGCAACCGCTGGACATAGTGG  |

**Table 2.12 – Sequence of primers used to amplify products in the titration of amplified 3C libraries protocol.**

### 2.5.5.2 Annealing Reaction

Prior to setting up the annealing reaction 5C primers (at 20  $\mu$ M starting concentration) stored at -20°C were thawed at 4 °C. In the meantime, 96.4  $\mu$ l MilliQ water was added to each position in five 96 well plates which were kept on ice at all times. Once thawed, the primer tubes were kept on ice, with 3.6  $\mu$ l of each primer added to separate positions in the 96 well plates. This was repeated until all primers were diluted in individual wells (each primer was now at a concentration of 0.73  $\mu$ M).

Next, 4  $\mu\text{l}$  of each primer dilution was added to a single 1.5 ml tube on ice giving a final volume of 1460  $\mu\text{l}$ . At this stage, each primer has a concentration of 0.002  $\mu\text{M}$ . 2 control reactions are typically performed with each 5C sample. The first control is a “no ligase” control where Taq ligase is not added after the annealing step. The second control is the “no template” control where 3C library is not added to the annealing master mix. The annealing master mix was made up as detailed in Table 2.13 with final volumes also incorporating the two control reactions. The annealing reaction was set up as detailed in Table 2.14. The volume of 3C library used in this reaction was decided from the graph of amplified 3C product intensity vs volume (Chapter 2.5.4.1). The starting concentration of STD used in this reaction was 1 mg/ml. For all cell lines and E11.5 limb cells used in this investigation the annealing reaction was conducted at 55 °C overnight. This was conducted on a PCR machine which had been set at 55 °C and paused. The primer plate could therefore be added to the machine and the annealing reaction started as quickly as possible.

A list of the 5C primers primers is provided in Table 2.15.

| <b>Annealing Master Mix</b> | <b>Volume (<math>\mu\text{l}</math>)</b> |
|-----------------------------|------------------------------------------|
| 5C Primers                  | 1.7                                      |
| NEB Buffer 4                | 1                                        |

**Table 2.13 – Volume of reagents needed for annealing master mix.**

|                            | <b>DNA (<math>\mu\text{l}</math>)</b> | <b>STD (<math>\mu\text{l}</math>)</b> | <b>Annealing Master Mix (<math>\mu\text{l}</math>)</b> |
|----------------------------|---------------------------------------|---------------------------------------|--------------------------------------------------------|
| <b>5C reaction</b>         | 5.86                                  | 1.44                                  | 2.7                                                    |
| <b>No ligase Control</b>   | 5.86                                  | 1.44                                  | 2.7                                                    |
| <b>No template control</b> | 0.00                                  | 1.44                                  | 2.7                                                    |

**Table 2.14 – Volume of reagents needed for annealing reaction.**

| Start<br>HindIII<br>fragment | End<br>HindIII<br>fragment | Primer<br>Primer | Primer<br>Position | Primer Sequence                |
|------------------------------|----------------------------|------------------|--------------------|--------------------------------|
| <b>28319935</b>              | 28324734                   | 5                | Forward            | AGCATAGAGTGTGTGTAGGTGCTGCCTAAG |
| <b>28331586</b>              | 28333893                   | 8                | Forward            | GGTAAGAGTCCCAAAGAACAGCTTGTTAAG |
| <b>28336650</b>              | 28340960                   | 10               | Forward            | AACACCTCTAGCATGATAGCACTTTGCAAG |
| <b>28344676</b>              | 28347365                   | 12               | Forward            | CAAACCCTAGAAGCCACAGGGACCAAG    |
| <b>28355883</b>              | 28370234                   | 18               | Forward            | GGAGACCCACACTAAGGGCCTCAAG      |
| <b>28373717</b>              | 28374948                   | 20               | Forward            | TGTGTTTTAGGGATGAGGGATTCTTTAAAG |
| <b>28388373</b>              | 28403099                   | 27               | Forward            | TTGATAGTGCTGTTTCCTGTGGCTAGAAG  |
| <b>28419609</b>              | 28428926                   | 31               | Forward            | TTCTAAATATAATCCAGAGAGAAGGCTAAG |
| <b>28452395</b>              | 28454456                   | 34               | Forward            | CAATAAAGGTAGAACTTGGGCCAGTGAAG  |
| <b>28463308</b>              | 28463573                   | 38               | Forward            | ATGGGGCCGGATTTAACTCAACAATCAAAG |
| <b>28469869</b>              | 28470959                   | 42               | Forward            | AAAAATAAAAAGGAGGCCTGGATTCTGAAG |
| <b>28484053</b>              | 28486216                   | 45               | Forward            | AAAATCTCCCTGGAGTCAAAGGGTTAAAAG |
| <b>28491940</b>              | 28496579                   | 47               | Forward            | CCACCCCCAGTATCTGCAACCTCAAG     |
| <b>28506236</b>              | 28509301                   | 50               | Forward            | GCAAGAGCCCACCAGGGTCAAG         |
| <b>28511000</b>              | 28511814                   | 53               | Forward            | GTGATGTCTCCCCTGTGAGCAGGAAG     |
| <b>28514357</b>              | 28515715                   | 57               | Forward            | TTCTCATGTCAAGATCCACATAATATCAAG |
| <b>28528248</b>              | 28532168                   | 59               | Forward            | TTACATATCAGCTGCTGTATCCATCACAAG |
| <b>28537154</b>              | 28541461                   | 62               | Forward            | CACAGGGCTCTTTCATAGCCTAAGAACAAG |
| <b>28542833</b>              | 28543234                   | 64               | Forward            | CACAGTCACCCCCTGTACCCCAAAG      |
| <b>28547549</b>              | 28547729                   | 68               | Forward            | GGTACTTCTACAGTGGGGGAGGATGTAAAG |
| <b>28559901</b>              | 28566954                   | 73               | Forward            | TCTATGTATAAGCCACACCAAGGAAAGAAG |
| <b>28570464</b>              | 28571000                   | 75               | Forward            | TGGCAAATCAGAAAACTCTTTGGATTAAG  |
| <b>28572085</b>              | 28576299                   | 77               | Forward            | GAGTTCAAGAGCCCCAAATCCCTCTAAG   |
| <b>28577016</b>              | 28580658                   | 79               | Forward            | GGAGGCAGAGCCTCTGAGTCACAAG      |
| <b>28581219</b>              | 28585322                   | 81               | Forward            | CTCCAGACTGAGACCTTCTGAGACCAAAG  |

|                 |          |     |         |                                  |
|-----------------|----------|-----|---------|----------------------------------|
| <b>28593412</b> | 28595934 | 84  | Forward | GGAAAGGTA CTCTGGGGTGCATCACAAAG   |
| <b>28611343</b> | 28619284 | 90  | Forward | GACCTGACACCTTTGGGATGAAAGTGTAAG   |
| <b>28620320</b> | 28627997 | 93  | Forward | ATTGGCTATGTAGATGAAGATGGTCCCAAG   |
| <b>28631581</b> | 28635856 | 95  | Forward | CGTGTACAGTTAGCTACACCCTCAAAAAG    |
| <b>28638809</b> | 28643392 | 97  | Forward | GTAGCGTTCGCGCGCCTCAAG            |
| <b>28646447</b> | 28647705 | 99  | Forward | TAGCATAGGGGTTATGGATGGACTCAAAAAG  |
| <b>28650693</b> | 28651422 | 101 | Forward | TGAGTTTTAGTAACCACTATAAAAAGAAG    |
| <b>28652424</b> | 28656081 | 103 | Forward | GCTGATTCCTTTGCTGACTGGAGTGTGAAG   |
| <b>28668895</b> | 28669869 | 107 | Forward | GTCTTCTTATCCCTGGGTAAATTGTAAAG    |
| <b>28681874</b> | 28682959 | 111 | Forward | CAGCATGGCTGTGAGGGAAAGTAGCTAAG    |
| <b>28685148</b> | 28687114 | 114 | Forward | CAATTCAGTGCCAGCCTCTCTCGGTAAG     |
| <b>28689808</b> | 28693301 | 117 | Forward | CCTGTTTGCAGTGTGTCTTCTCACAGGAAG   |
| <b>28698429</b> | 28705774 | 119 | Forward | AAAAGAAAAGAATCACCATGACTCTTAAG    |
| <b>28710086</b> | 28710865 | 121 | Forward | CATGGTAACATCGTGTGTAGATAGAAAAAG   |
| <b>28715945</b> | 28717043 | 123 | Forward | AACTTTGTGGCACCTTTCTCCTCCAGAAAG   |
| <b>28723250</b> | 28724961 | 128 | Forward | CCTTCTGTGTGATCATCTGACACATAAAAAG  |
| <b>28731809</b> | 28735435 | 130 | Forward | GTTGCAGACTGAGGGGCTCCAGAAG        |
| <b>28738818</b> | 28743466 | 132 | Forward | ATTTTCTGTGGCATTATTAGGCAGGTAAG    |
| <b>28748046</b> | 28750565 | 134 | Forward | CCTGCAGTCAGGGAACCGAGAGAAG        |
| <b>28763979</b> | 28767078 | 140 | Forward | ATATTTGGATGTTCTGTGTCAGTGGCCTGAAG |
| <b>28780564</b> | 28782634 | 144 | Forward | CCAGAGACCCCTCCATCTGCTCAAAG       |
| <b>28789547</b> | 28793709 | 147 | Forward | CATCTGATTGGCCAAGCCGCACAAG        |
| <b>28799299</b> | 28801546 | 152 | Forward | TAGAAAAGATGCTGGGAACCTCATTCTAAG   |
| <b>28805808</b> | 28806894 | 155 | Forward | GTGGAGCCATCATGGAATTGCATGGAAG     |
| <b>28837190</b> | 28843289 | 162 | Forward | AGAACCACAGGATACCCATAAGAGCCAAAG   |
| <b>28870608</b> | 28876511 | 168 | Forward | ATATCTACACCAGCTTTCTAAAATGGAAAG   |
| <b>28878135</b> | 28878784 | 170 | Forward | CTGTGATCTGAAGGTGTAAGCTGAGATAAG   |

|                 |          |     |         |                                 |
|-----------------|----------|-----|---------|---------------------------------|
| <b>28879305</b> | 28880400 | 173 | Forward | GGTGTGCAGCCAGTGTGCATATTAGACAAG  |
| <b>28882358</b> | 28883821 | 176 | Forward | GGCTGCAAAAGTTGGGTCTCATTTGTGAAG  |
| <b>28884804</b> | 28887360 | 178 | Forward | CTTGGGTAAAGCTATACTGGATGCGGCAAG  |
| <b>28889233</b> | 28895175 | 180 | Forward | TTACCTAGGTAAGTCTGTCCTTTCAGAAG   |
| <b>28923745</b> | 28923931 | 191 | Forward | TTGTTATTCCTGGGAGACTTAATTGGCAAG  |
| <b>28942022</b> | 28945936 | 197 | Forward | CCACAACCTGCCCTACGGTGTAAAG       |
| <b>28949570</b> | 28950647 | 199 | Forward | GGTATGGGATATCCTTTGGGGTTCACTAAG  |
| <b>28951389</b> | 28952477 | 201 | Forward | CTTCACTCCCTCAGTTACCAAGCCACAAAG  |
| <b>28954090</b> | 28954407 | 204 | Forward | CAGGATGCCTTAGGAGACACGAGGAAG     |
| <b>28997868</b> | 29000514 | 212 | Forward | CAGGCACCTAAGTGTTAGAGAAGTTGGAAG  |
| <b>29003766</b> | 29004021 | 214 | Forward | TACAAGTCTCATCTGAGCCCTCCAAAAAAG  |
| <b>29005071</b> | 29006498 | 216 | Forward | GATGGGTCTTTCAGAGTCTGTTCCCTGAAG  |
| <b>29008670</b> | 29011994 | 218 | Forward | TGTTACCAAAAATTTATTCTAAAAGGCAAG  |
| <b>29013867</b> | 29014066 | 220 | Forward | CAAGATTACCTGAAGTGCCGGTACAGAAG   |
| <b>29020058</b> | 29022523 | 222 | Forward | GCTCTGTGTGCCTTCAGCTCTCTGAAG     |
| <b>29028121</b> | 29029629 | 224 | Forward | AGCAGGCTTCCTCCTAGGATTATAATGAAG  |
| <b>29032469</b> | 29038542 | 226 | Forward | CTGAGTCTCAAGCAGCTAGCTTTCAGAAAG  |
| <b>29048476</b> | 29050003 | 230 | Forward | CTTCCCAGGCTTTGAAGGGAACACACTAAG  |
| <b>29051240</b> | 29058423 | 232 | Forward | GGTCCACTGGCAGCCCAAGAAG          |
| <b>29060962</b> | 29068732 | 234 | Forward | AGTGTTTAGGTTCTAAGGACATGGCAAAG   |
| <b>29081208</b> | 29082327 | 238 | Forward | AGTTTCCTTGAAAATCATGGCCAGCAAAG   |
| <b>29084039</b> | 29096781 | 241 | Forward | TCACACACTGGCAGCTGGACAAG         |
| <b>29100884</b> | 29101246 | 243 | Forward | AAGATAAAAACCTTCCATCTTAGAACAAAG  |
| <b>29109460</b> | 29111881 | 245 | Forward | TAGGAAAAGGGACTTAATTTGTCATCAAAG  |
| <b>29113253</b> | 29117196 | 247 | Forward | CGTATAGACTTAAATGAATTAGAACAAAAG  |
| <b>29130489</b> | 29131877 | 251 | Forward | AGAAGGATTGTTTTATTCTGTCCTTAAAG   |
| <b>29137043</b> | 29154044 | 254 | Forward | GTTTTTTTCTGGCACAAGCTACCACTTAAAG |

|                 |          |     |         |                                |
|-----------------|----------|-----|---------|--------------------------------|
| <b>29155708</b> | 29159386 | 256 | Forward | TTTATTCTAACACTTATCCCATCCTGCAAG |
| <b>29160129</b> | 29166826 | 259 | Forward | CTATATTTATTTAAAGTACAAAAACCTAAG |
| <b>29168467</b> | 29173627 | 261 | Forward | GCACCGAGACCAGCTTCTGAGATCAAG    |
| <b>29173866</b> | 29174415 | 263 | Forward | CATTAAACTGTGTTGAACCTATTTATTAAG |
| <b>29179396</b> | 29180591 | 265 | Forward | TGGCACCTCAAATTGAGACCTTGCTTTAAG |
| <b>29184580</b> | 29185047 | 267 | Forward | AGTGTTTCTATTTTTATGCTGAGTCCCAAG |
| <b>29185853</b> | 29188276 | 269 | Forward | TGAGAACAGTATTTTACTTAATTTGAGAAG |
| <b>29191727</b> | 29196352 | 272 | Forward | TTGAAAGCTGATTTCAAACAATGATTAAAG |
| <b>29199537</b> | 29203806 | 275 | Forward | AGCTATCATTGGTTAAAAACTGTTAGAAG  |
| <b>29206157</b> | 29209241 | 278 | Forward | GGCACATGCTGGGTCCCAGATAAG       |
| <b>29216469</b> | 29220327 | 280 | Forward | CATGGGAGGTGTCAAACGGATTGGTGAAG  |
| <b>29230618</b> | 29233018 | 283 | Forward | GCCACAGTCTGGAAGCACAGATCCAAG    |
| <b>29236381</b> | 29237108 | 285 | Forward | CCAGGGTCATCTATGTGCATGCTCACAAG  |
| <b>29240128</b> | 29243266 | 288 | Forward | TGTCATCCCTTGTCTAGTTTACTCTAAG   |
| <b>29246729</b> | 29247568 | 290 | Forward | GCGTAGTAAACACGGGGTTATAAGCTAAG  |
| <b>29250430</b> | 29250736 | 292 | Forward | GGGAGCAATTCTTAAGAGTGCTTTTCTAAG |
| <b>29259187</b> | 29262463 | 296 | Forward | CTAGGTTTCTTGGGGAGAAGGGCTATGAAG |
| <b>29264722</b> | 29268159 | 298 | Forward | GGAGCAGCCAGTAGCCCCAGAAG        |
| <b>29268268</b> | 29270862 | 300 | Forward | TGATGCTATCTCTCTCAAAGGAGGAGAAG  |
| <b>29277402</b> | 29281527 | 303 | Forward | TGGTAAGAAAATGAAATGTAAGGTACGAAG |
| <b>29292944</b> | 29296257 | 309 | Forward | GATGGATCTGGAAGGAGGGACACCAAAG   |
| <b>29309648</b> | 29310307 | 314 | Forward | CCAGCCTGTGATGACAGGTGGTAGAAG    |
| <b>29310490</b> | 29311401 | 316 | Forward | ACCTTTCCTTCTCTAACTGCCATGCAAAG  |
| <b>29317749</b> | 29319160 | 319 | Forward | GCCCCTGAAGATATAGCACAGTCTTGCAAG |
| <b>29319273</b> | 29320195 | 321 | Forward | TCAAACCATACAGCTGAATATGGAGAGAAG |
| <b>29331158</b> | 29332122 | 325 | Forward | CACAGCTCCTGCACATCTGGGGAAG      |
| <b>29347104</b> | 29352141 | 327 | Forward | CCATTTAAGGCCTGCCAGCTCAAG       |

|                 |          |     |         |                                |
|-----------------|----------|-----|---------|--------------------------------|
| <b>29356572</b> | 29358986 | 330 | Forward | AGATCTGTGCAACTTTCTTCAAGCCACAAG |
| <b>29384407</b> | 29389154 | 342 | Forward | TCACCACATGTTGAACTCTGGATGGGAAAG |
| <b>29402955</b> | 29404268 | 346 | Forward | TTTGGTTTGGTTTTCTGGGTTGAGGGGAAG |
| <b>29407972</b> | 29411033 | 350 | Forward | AGCCTCCAGCTAAATGCCAACAAATGAAAG |
| <b>29411528</b> | 29412802 | 353 | Forward | CAAATACAAAAAATCTGCCAAACAACAAAG |
| <b>29413668</b> | 29419014 | 355 | Forward | CCCTCTTCTTTTCAAAGATGGAAAGTAAG  |
| <b>29426352</b> | 29436153 | 361 | Forward | TACTTATTTGTGACATGGTATCTAAAAAG  |
| <b>29457358</b> | 29463943 | 367 | Forward | GACCACATCCTGTCTAAACCCTGCCAAG   |
| <b>29471071</b> | 29483618 | 371 | Forward | CTGGCTGTCTTTGCCACCAAACAAAG     |
| <b>29486277</b> | 29487932 | 373 | Forward | GTTCTTTCAGTGATAGATGAGAAGTGTAAG |
| <b>29495457</b> | 29497960 | 376 | Forward | TCAAGGAAGGCTACAAGAGGAGAGGCAAG  |
| <b>29514013</b> | 29518402 | 379 | Forward | GGATATTGTCTCCAATGGTTATGTTAAAAG |
| <b>29521325</b> | 29521533 | 383 | Forward | GTTGCCTGCTATTAATCTGCTTCTACAAG  |
| <b>29527531</b> | 29527778 | 386 | Forward | GTAGGCTTGAGCACTTACCAGCAGTGAAAG |
| <b>29530052</b> | 29530237 | 388 | Forward | CAGCTGAGAAGACCCAGCACAAATCCAAG  |
| <b>29534954</b> | 29536943 | 391 | Forward | AGAGCCACAGTGAACAATGTCTCTAGCAAG |
| <b>29539102</b> | 29543893 | 394 | Forward | CTAAAGGTCTTCTTGAGAGGAGCACTGAAG |
| <b>29555730</b> | 29559300 | 399 | Forward | CAGGCTTCTGCCTGCAGAACCAAG       |
| <b>29560665</b> | 29560851 | 402 | Forward | GGAGTAAATGCTACATGACTGCCTGGGAAG |
| <b>29561929</b> | 29563972 | 404 | Forward | TAGTTATCCCGTGAGAAAGGTACACTCAAG |
| <b>29570405</b> | 29571012 | 408 | Forward | GACCTGACACTGTTGATGACATTAACAAAG |
| <b>29574786</b> | 29577091 | 410 | Forward | CCTCACAACCTCTTTGTAGCAAGTGATAAG |
| <b>29577562</b> | 29579620 | 412 | Forward | ATTATTTAATGTAGTTAGTAATTTGAAAG  |
| <b>29582839</b> | 29584474 | 415 | Forward | TATATATACTTATGAACATGTTTGAAAAG  |
| <b>29585001</b> | 29590052 | 418 | Forward | AGCAAAGCATTAAAGTGTAGCAGTGGAAAG |
| <b>29595576</b> | 29595777 | 421 | Forward | TACCCTTATCAGAATGAAGTGTGCACAAAG |
| <b>29607876</b> | 29613408 | 426 | Forward | CTATGGATGTTAACCAATAAGCTACAGAAG |

|                 |          |     |         |                                |
|-----------------|----------|-----|---------|--------------------------------|
| <b>29626315</b> | 29628597 | 431 | Forward | ACATCCATAAGATTATTTCCGATGATCAAG |
| <b>29632212</b> | 29635239 | 435 | Forward | TGCTCTACCATGCGTGAAGGTGATATAAAG |
| <b>29637006</b> | 29638519 | 437 | Forward | ATAAATCTAACAAACTAAAAATCAGTCAAG |
| <b>29640676</b> | 29641133 | 439 | Forward | CTCATTTCATTACCTCACTCCTTTAGAAG  |
| <b>29642804</b> | 29649071 | 441 | Forward | CATACAAAAGCCAGTTTTCTCAATGTGAAG |
| <b>29656939</b> | 29658283 | 447 | Forward | TTCACATGTAACAAAGCCAAATTTATGAAG |
| <b>29701562</b> | 29703903 | 457 | Forward | GAGCCACTGCAGGGCTGGAAG          |
| <b>29705291</b> | 29706210 | 459 | Forward | ACATTTCTGTTGCAGGGCTATTCATGGAAG |
| <b>29709706</b> | 29713030 | 461 | Forward | GGTAACAAGAATAGCATTGAAAATTCTAAG |
| <b>29716182</b> | 29720154 | 464 | Forward | CACCAGCCACCATCAACTACACAAAGAAG  |
| <b>29727115</b> | 29730837 | 466 | Forward | CTGAATTTAGGTGTATGTTTACTCAGTAAG |
| <b>29750154</b> | 29752052 | 470 | Forward | CATAATAAAACTGAAAGGAGATGACCCAAG |
| <b>29757662</b> | 29764922 | 474 | Forward | TAGTACTTTAGCTAATGGCATTCACTGAAG |
| <b>29768741</b> | 29769326 | 477 | Forward | TGAAATGTCCACAAAGGATTATCAGGAAG  |
| <b>29770691</b> | 29772235 | 479 | Forward | GGTGAGCTAGTCAGAGCAGTGCTGAAAAG  |
| <b>29777534</b> | 29781541 | 482 | Forward | TGGCATTATATCTACTTAGGAGAAAAG    |
| <b>29782721</b> | 29784642 | 484 | Forward | GACCCCAAGGCACCCAACCTCAAG       |
| <b>29786563</b> | 29787575 | 487 | Forward | TTCTGAGCTCCTGTTTTCCCTCAGAGTAAG |
| <b>29797223</b> | 29805362 | 489 | Forward | GGTTGGAGAGTTTGGAGGCTGAACACAAG  |
| <b>29806189</b> | 29807449 | 491 | Forward | GTTTGTGTGTAGGGGATGTGATCCAAAAG  |
| <b>29809880</b> | 29815593 | 493 | Forward | ATTTGCCACTCAAATCTGCACTTTCCAAG  |
| <b>29816517</b> | 29820938 | 495 | Forward | GGAGGGGGGAGGTGGGAACAAG         |
| <b>29835164</b> | 29835981 | 497 | Forward | AAGACTCAGTTATACCAATGGTTCAATAAG |
| <b>29836157</b> | 29838218 | 499 | Forward | GGTGCAGGGAAAGTTGATAAGGGCAAAG   |
| <b>29845799</b> | 29846398 | 505 | Forward | TTCAGCTTTGTCTTGGTGTGTGGACTAAAG |
| <b>29848888</b> | 29849213 | 507 | Forward | GTACTGAGCCTAGCAGAGGAAGCTCAAG   |
| <b>29854912</b> | 29855302 | 511 | Forward | AGTGACCTAGATCTTCCCCTACACTTTAAG |



|                 |          |     |         |                                    |
|-----------------|----------|-----|---------|------------------------------------|
| <b>29859545</b> | 29863829 | 513 | Forward | CATCATCTGTGCCATTTCTTAGCCCCAAAG     |
| <b>29870040</b> | 29874795 | 517 | Forward | CTCTGAAACATGCAACCACACAGGCAAAG      |
| <b>29896489</b> | 29896823 | 521 | Forward | TTTTTTTCCCATTAGGCGAAATTTTAAAG      |
| <b>29897293</b> | 29900118 | 523 | Forward | TAAAACCTGCAGCAACTTTCCTGTGTTAAG     |
| <b>29902876</b> | 29906098 | 525 | Forward | AATATGGTTTCCTTATGATGTAGCTTTAAG     |
| <b>29907125</b> | 29909799 | 527 | Forward | TTGTGATCAAATTTACATGTCTAAGCTAAG     |
| <b>29913291</b> | 29913607 | 530 | Forward | TTCCTTAGCTTTATATATTGATGGAGAAG      |
| <b>29918393</b> | 29927218 | 533 | Forward | AGATCTCTCTACTTGTTGATTAATAGCAAG     |
| <b>29938026</b> | 29947212 | 536 | Forward | AAGTTAATTGAATGAAAATGATCAACTAAG     |
| <b>29949610</b> | 29953599 | 538 | Forward | CATGTGTTGTAGAGGCTGGGAACCACAAG      |
| <b>29958999</b> | 29960405 | 540 | Forward | CATGTGGGTTAACAGTGAGTTAAGCCCAAG     |
| <b>29963243</b> | 29963692 | 543 | Forward | TAGACATGTGAATTTTATTCTTGAGGAAG      |
| <b>29967853</b> | 29970155 | 546 | Forward | GATTTTTTTTCTAGGGGTTTATTCAAG        |
| <b>29977865</b> | 29978715 | 548 | Forward | AAGGAAAAAAAAAGAATAAGGTTTCAGAAG     |
| <b>29979320</b> | 29988808 | 550 | Forward | TGTACAGTATAAGCACCCCTATAGCCAAAG     |
| <b>29996460</b> | 30001398 | 554 | Forward | GCAACTCAGTGAAATTAACCAAAGATGAAG     |
| <b>60918191</b> | 60924238 | 2   | Forward | GGGAGGAAGATAAAAAGATGGGGATGGAA<br>G |
| <b>60932406</b> | 60937638 | 5   | Forward | CTGTCAGTCTCCTGCTGCCACTAACAAG       |
| <b>60949684</b> | 60966226 | 9   | Forward | GTGTCAAGGAGGCAGACCTTCAGGAAG        |
| <b>60970542</b> | 60976197 | 11  | Forward | AACCCATCTCTAACAATCCCTTTGTAAAAG     |
| <b>60977797</b> | 60977935 | 14  | Forward | GCAGGAGTCTAAGCCACCAGGGGAAG         |
| <b>61000572</b> | 61003268 | 17  | Forward | ACTTTTCCAAAATCCAAGCTGACTTCAAG      |
| <b>28317087</b> | 28319149 | 3   | Reverse | CTTCCAGTATTGGGTTACAGTTAATGGAGT     |
| <b>28325862</b> | 28331585 | 7   | Reverse | CTTCTCTCATCCCTACACTAACCAGGCCT      |
| <b>28333894</b> | 28336649 | 9   | Reverse | CTTGACACACCTACCCTAAGTAATCAAT       |
| <b>28340961</b> | 28344675 | 11  | Reverse | CTTAGGATGTGCCTCTACTGTGGGGG         |
| <b>28347366</b> | 28348911 | 13  | Reverse | CTTGCCAGTTTATCTAGGTAGCCTGCCAG      |

|                 |          |     |         |                                |
|-----------------|----------|-----|---------|--------------------------------|
| <b>28370235</b> | 28373716 | 19  | Reverse | CTTGGATTGGCGTGGCTCTGAGTCAT     |
| <b>28385728</b> | 28388130 | 25  | Reverse | CTTCCTTTCTGGTATCTATTGACCTTCCCT |
| <b>28416843</b> | 28419608 | 30  | Reverse | CTTCTCCTGTAAGATGGCAATTTATTTATT |
| <b>28428927</b> | 28447705 | 32  | Reverse | CTTTGTCATTCCACCATGGTTGTGGTGAAC |
| <b>28454457</b> | 28458325 | 35  | Reverse | CTTGAGCTCACATATGGGACACTCTTGACA |
| <b>28469222</b> | 28469868 | 41  | Reverse | CTTCTTAGCTAGGCCAGCATAATGTACCG  |
| <b>28476584</b> | 28484052 | 44  | Reverse | CTTAGCTCAAAATGTAGGAAATGGCCATTC |
| <b>28486217</b> | 28491939 | 46  | Reverse | CTTGTGAACAGTCCCACCAGGTCACTG    |
| <b>28504618</b> | 28506235 | 49  | Reverse | CTTAAGGTGGGGGTGACACAGTCCAAAG   |
| <b>28509323</b> | 28510999 | 52  | Reverse | CTTCAGTTTGGGTGACACATGCAGGACAAA |
| <b>28511842</b> | 28514356 | 56  | Reverse | CTTATCCTTCTGTGTCTAGTTGAAGTGG   |
| <b>28515716</b> | 28528247 | 58  | Reverse | CTTGTTGAGCTCATGTAATGCCTATGGAT  |
| <b>28534318</b> | 28537153 | 61  | Reverse | CTTGAAAAGCACAGATAAAAATGCCATTTG |
| <b>28541462</b> | 28542832 | 63  | Reverse | CTTCCTTGGGCAAGTGGTATCTTCCTTAGC |
| <b>28543235</b> | 28543365 | 65  | Reverse | CTTTGACAAAGTGATGCATCTAAGATCCTA |
| <b>28547730</b> | 28548357 | 69  | Reverse | CTTCTCATATGACTCTGGTTTCTTGGCCCA |
| <b>28566955</b> | 28570463 | 74  | Reverse | CTTAGCATGGGACTCAGAAAACAAAATAGG |
| <b>28571001</b> | 28572084 | 76  | Reverse | CTTGAAATTGAAGTATCTCTCTCAGCACCT |
| <b>28576300</b> | 28577015 | 78  | Reverse | CTTCCAGAAGATCTGCAGCAACTCTCTCTC |
| <b>28580659</b> | 28581218 | 80  | Reverse | CTTGAAGCATGTGTGGACTCCCATTCTTCT |
| <b>28585390</b> | 28593411 | 83  | Reverse | CTTGCGGTTGTAGCTAAGAGTGAATTTGAA |
| <b>28604862</b> | 28605362 | 86  | Reverse | CTTTGGGGTGGGAACAAGGAGACTTCAC   |
| <b>28619329</b> | 28620319 | 92  | Reverse | CTTGACAGAGGAGCCTAAAAGGTGACTTAA |
| <b>28627998</b> | 28631580 | 94  | Reverse | CTTTCATTGAGACTCCCTCAGCCTCAGT   |
| <b>28635857</b> | 28638808 | 96  | Reverse | CTTAGCAGTTTCTGTAAAAAATAAAAGTAC |
| <b>28643393</b> | 28646446 | 98  | Reverse | CTTGTGCAGTACTAAATCATAATGCCATAA |
| <b>28647706</b> | 28650692 | 100 | Reverse | CTTAAAAGGTACAATGATAGAAGAAATAG  |

|                 |          |     |         |                                 |
|-----------------|----------|-----|---------|---------------------------------|
| <b>28651423</b> | 28652423 | 102 | Reverse | CTTCCTATGGCTGAGAACTGCTTAGATAAT  |
| <b>28662072</b> | 28668894 | 106 | Reverse | CTTTTGGTTGTTTCAACCATTTTTCACTTA  |
| <b>28676542</b> | 28681826 | 109 | Reverse | CTTCCTCTCTCAGGAAACCAGTCTTCTGAG  |
| <b>28682960</b> | 28684009 | 112 | Reverse | CTTCCAACAGTACATTATCCTAAGCGTCTA  |
| <b>28687115</b> | 28688407 | 115 | Reverse | CTTACACCAATTAACCTTCTTAGAAGTAGAC |
| <b>28693302</b> | 28698428 | 118 | Reverse | CTTAAATCACGAAGTACTGAGGCTTACCAA  |
| <b>28705775</b> | 28710085 | 120 | Reverse | CTTATTCACAGGCCATTCTGGTAGGAACAT  |
| <b>28710866</b> | 28715944 | 122 | Reverse | CTTATGTGGAGAACTAACACCATCATAAAT  |
| <b>28717044</b> | 28718937 | 124 | Reverse | CTTTACCACTGCGGAAGGGGGAAAACA     |
| <b>28724962</b> | 28731808 | 129 | Reverse | CTTATACTGGGTGGAGGTCAATTCTGGACT  |
| <b>28735436</b> | 28738817 | 131 | Reverse | CTTAACTGGCTCTGCTCTCAGAATGAGG    |
| <b>28743467</b> | 28748045 | 133 | Reverse | CTTTAATGTTCTGGTTTGTGTTGTTCTAA   |
| <b>28759100</b> | 28763978 | 139 | Reverse | CTTTTATGTGGCTCTGCTTTTGTATTACAA  |
| <b>28770697</b> | 28780563 | 143 | Reverse | CTTGTTCCCGTACCCACATAAAAGGCC     |
| <b>28782635</b> | 28789524 | 145 | Reverse | CTTTCCCTCACCCATTGAAAGAAGGGAG    |
| <b>28794058</b> | 28797606 | 149 | Reverse | CTTACTACCAGTCCTTGTCTGTCTAATA    |
| <b>28798546</b> | 28799298 | 151 | Reverse | CTTGGCTAACATTGGACAACCCAAGTGTTT  |
| <b>28805019</b> | 28805807 | 154 | Reverse | CTTCTGGACACCCAGAAATGTGCGTCTC    |
| <b>28821671</b> | 28827662 | 158 | Reverse | CTTCTCCTAAGAACAGCTAGACCTATGCAC  |
| <b>28843290</b> | 28847606 | 163 | Reverse | CTTCAAAGCTGCAGTGCTTTGAAGTGTCTG  |
| <b>28876512</b> | 28878134 | 169 | Reverse | CTTTATTGCCAGGTCAAATGATTTAAACAT  |
| <b>28878785</b> | 28879286 | 171 | Reverse | CTTTGAAGGAGACCCTATTTCTATGTGGG   |
| <b>28880401</b> | 28880839 | 174 | Reverse | CTTTATTCACCTCTGACATGCAAGCCAACA  |
| <b>28883822</b> | 28884803 | 177 | Reverse | CTTTGGAAGGCTGGGTGGTCAGC         |
| <b>28887361</b> | 28889232 | 179 | Reverse | CTTCCAAAATTCTATTTTGGGAAAAAATGA  |
| <b>28895176</b> | 28899338 | 181 | Reverse | CTTCCAGAACTGTTTACTTCTTCTGGAG    |
| <b>28925837</b> | 28929215 | 193 | Reverse | CTTCCAGGTCACGTAAAGATATTTAGTA    |

|                 |          |     |         |                                |
|-----------------|----------|-----|---------|--------------------------------|
| <b>28945937</b> | 28949569 | 198 | Reverse | CTTAGAAGGACCCTAGAATGGTCCCCTGAA |
| <b>28950648</b> | 28951388 | 200 | Reverse | CTTTGAATGAGGTCAATAAAAATCTACCTC |
| <b>28971638</b> | 28978794 | 207 | Reverse | CTTTTCAAAAAGCCACATGAAAATCTACTA |
| <b>29000515</b> | 29003765 | 213 | Reverse | CTTATGAGTGCAAGGTCTGCCAGGTTG    |
| <b>29004022</b> | 29005070 | 215 | Reverse | CTTAGTTGTTCTTCAGTGTCAGTTACTTC  |
| <b>29006499</b> | 29008669 | 217 | Reverse | CTTTTATTAAGGCCATGGGGCCATGGAG   |
| <b>29011995</b> | 29013866 | 219 | Reverse | CTTCTAATGAACCTGCTCCTGACCGCATG  |
| <b>29014067</b> | 29020057 | 221 | Reverse | CTTCAGGAGTCTAACTGCAATAATAATGCA |
| <b>29022524</b> | 29028120 | 223 | Reverse | CTTTGCTGTTGGTCAAAGGTAGCAGCTGA  |
| <b>29029630</b> | 29032468 | 225 | Reverse | CTTCGGGCCCCGGAGGGAGG           |
| <b>29038543</b> | 29039555 | 227 | Reverse | CTTTTGAGAGGTGTGAGACAATCAAATAA  |
| <b>29050004</b> | 29051239 | 231 | Reverse | CTTCAATTTGTGAGCCTCTACAAAACCTCA |
| <b>29058424</b> | 29060961 | 233 | Reverse | CTTCCTTTTCATCTTGAATCAGCCTATAGA |
| <b>29078836</b> | 29081207 | 237 | Reverse | CTTTCTCCCAGTGATGCTGTTAGTTGTTCC |
| <b>29082328</b> | 29083134 | 239 | Reverse | CTTGCAATACTTCCAAGAGAAAGAGAACAC |
| <b>29096782</b> | 29100883 | 242 | Reverse | CTTTCAATTACTGCTCTAGCACAATGCCTG |
| <b>29101247</b> | 29109459 | 244 | Reverse | CTTGAAAAGAGGAATATTCGGGAATGGAGG |
| <b>29111882</b> | 29113252 | 246 | Reverse | CTTAGATCTGGTGGTAGAAGGGATACTGGA |
| <b>29117197</b> | 29119739 | 248 | Reverse | CTTTGCTGTCTGCTCAGAGGAGAACTCTGT |
| <b>29131878</b> | 29136802 | 252 | Reverse | CTTAGGCCATGAAGGAAGATGGCTTTGACA |
| <b>29154045</b> | 29155707 | 255 | Reverse | CTTCCAATGATCCTCTCTGCTCCAATGAAA |
| <b>29159409</b> | 29160128 | 258 | Reverse | CTTGGGGACCATTTAAAATATGTTCTAGAT |
| <b>29166827</b> | 29168466 | 260 | Reverse | CTTGCCATGTAAGATGGGATATCTGGCCAG |
| <b>29173628</b> | 29173865 | 262 | Reverse | CTTATTTATCAAGTACAGTTGCTCAAGTAT |
| <b>29174416</b> | 29179395 | 264 | Reverse | CTTGCAGCTCCAGACATTTACCCATCTCTG |
| <b>29180592</b> | 29184579 | 266 | Reverse | CTTCCTCCTTACCCCGGCTGCTAATC     |
| <b>29185048</b> | 29185852 | 268 | Reverse | CTTTGATATTACAGTGATGAATTGATATGT |

|                 |          |     |         |                                  |
|-----------------|----------|-----|---------|----------------------------------|
| <b>29188338</b> | 29191726 | 271 | Reverse | CTTAGACTTCCTTTTAGAACATTCTTGGTT   |
| <b>29198345</b> | 29199536 | 274 | Reverse | CTTATCAACAACACCTGCATTTTAAAAGAC   |
| <b>29203807</b> | 29203920 | 276 | Reverse | CTTCATGCTGGCAGACAAAGTAAATTCGGG   |
| <b>29209242</b> | 29216468 | 279 | Reverse | CTTCCTTGTAATATCTGGAATAGAAAAGAG   |
| <b>29220328</b> | 29223230 | 281 | Reverse | CTTGGAGATTTCCAGGTCTGCTTCTACATT   |
| <b>29233019</b> | 29236380 | 284 | Reverse | CTTCCATCAGATCTGTGTTGTCCAAGAATC   |
| <b>29237129</b> | 29240127 | 287 | Reverse | CTTGCTAACACCGTTTGAGGTGGAGATGC    |
| <b>29243267</b> | 29246728 | 289 | Reverse | CTTCTCCTGGTCATCATCTGGCCAGTATCA   |
| <b>29247569</b> | 29250429 | 291 | Reverse | CTTGGCTACACTGCATTTCTTGTATTATA    |
| <b>29253038</b> | 29259186 | 295 | Reverse | CTTTGAAGATTTCCAAGAGTTCAAACCCAG   |
| <b>29262464</b> | 29264721 | 297 | Reverse | CTTGAAAAGTGAAGAGATATCCTACAGAAT   |
| <b>29268160</b> | 29268267 | 299 | Reverse | CTTGCTTTGCCATTGGACCTTGTCTAG      |
| <b>29270863</b> | 29272035 | 301 | Reverse | CTTTAAAGCAACAATATGATATGCATCATC   |
| <b>29281528</b> | 29287941 | 304 | Reverse | CTTCCATCTTTTTTCATTTAGAAGCACCAG   |
| <b>29305044</b> | 29309647 | 313 | Reverse | CTTTAAATTCCTGCTTGTA AAAAATTGTTTT |
| <b>29311402</b> | 29315522 | 317 | Reverse | CTTGACCTCAGTATTCTGTACTCACTCCCC   |
| <b>29319161</b> | 29319272 | 320 | Reverse | CTTGCACGTCTCCATTCATCTTGACTAAAG   |
| <b>29321894</b> | 29331157 | 324 | Reverse | CTTAAATCTAATAAGATGAAGGAAAATAAC   |
| <b>29332123</b> | 29347103 | 326 | Reverse | CTTCCAACACCAAATGTGGTTCAATTACGC   |
| <b>29358987</b> | 29363174 | 331 | Reverse | CTTGGATGCCTTTGATCCCCTTATAACTTA   |
| <b>29401642</b> | 29402954 | 345 | Reverse | CTTCTACGTAGTAGATTTTTATATGGAATT   |
| <b>29406181</b> | 29407971 | 349 | Reverse | CTTCCCTCCTGGGAATGCTGTGTGG        |
| <b>29411034</b> | 29411488 | 351 | Reverse | CTTAGGATAGGAAGGAATTAGTGATACAGT   |
| <b>29412803</b> | 29413667 | 354 | Reverse | CTTGGAAACTGCACCATGGA ACTCTCACTT  |
| <b>29425367</b> | 29425485 | 359 | Reverse | CTTTCTTGAGAAGGAGTTTTATTCTAGTA    |
| <b>29436154</b> | 29438524 | 362 | Reverse | CTTCAGGGGAGTGAAAGAATTAAGATTTCT   |
| <b>29454397</b> | 29457357 | 366 | Reverse | CTTGCCCCACAGGGCAGGC              |

|                 |          |     |         |                                 |
|-----------------|----------|-----|---------|---------------------------------|
| <b>29466664</b> | 29471070 | 370 | Reverse | CTTTGGTATGAGGTAATAAAAAATAATTGAA |
| <b>29483619</b> | 29486276 | 372 | Reverse | CTTCGTGGTTGTATGCCTGTAATTGTGTTT  |
| <b>29492359</b> | 29495456 | 375 | Reverse | CTTCAAGGTTGCAGGTTACAAAATTATTA   |
| <b>29508039</b> | 29514012 | 378 | Reverse | CTTTGTGAGGACGTGAGTCCGGCTG       |
| <b>29519437</b> | 29521324 | 382 | Reverse | CTTTTGGATGCCATTAGCTAGACCTGAGTT  |
| <b>29527218</b> | 29527530 | 385 | Reverse | CTTAACTTACTTCACTAGATTTACTTCTGG  |
| <b>29527779</b> | 29530051 | 387 | Reverse | CTTTCTACAACTGGGAAAACCAGCCTTGG   |
| <b>29531243</b> | 29534953 | 390 | Reverse | CTTATGAAGTTTTTCAGTTAAAAGTCACAT  |
| <b>29537971</b> | 29539101 | 393 | Reverse | CTTTTACTGGTTTAATATCCTTCCAGGCTT  |
| <b>29543894</b> | 29545876 | 395 | Reverse | CTTACTTTCTGTACCTTAAAGGTGAAAA    |
| <b>29559301</b> | 29560638 | 400 | Reverse | CTTCTCCTAGCACCAACCTTATGATCCTGG  |
| <b>29560852</b> | 29561928 | 403 | Reverse | CTTGGTCTTTAATGTAGGCATGATGGGGTA  |
| <b>29563973</b> | 29566252 | 405 | Reverse | CTTGGCAGTTTGATTAATGCATGCTTAAGC  |
| <b>29567141</b> | 29570404 | 407 | Reverse | CTTGGGTACCAGTTGCATAAGCGTGG      |
| <b>29571013</b> | 29574785 | 409 | Reverse | CTCCCAACCAAGGTGGGTGGG           |
| <b>29577092</b> | 29577561 | 411 | Reverse | CTTGCGAACAGAATAAAGGACGCATTTACC  |
| <b>29582421</b> | 29582838 | 414 | Reverse | CTTAGCTTCCATTTGTTGGGAAGAGTGGTG  |
| <b>29584537</b> | 29585000 | 417 | Reverse | CTTACTTCAGAATTAGGAAAACACAAAGCA  |
| <b>29590053</b> | 29595530 | 419 | Reverse | CTTCCATTCTGAGTCTAGTGACTTAAAGG   |
| <b>29606038</b> | 29607875 | 425 | Reverse | CTTGTCACTGCCATTACCTGACTGTGC     |
| <b>29622314</b> | 29623626 | 429 | Reverse | CTTCTTATTGAAAAAATTGAAATTTTCTCT  |
| <b>29628598</b> | 29629851 | 432 | Reverse | CTTCATATCTAACATTTGACTCATTGAGAG  |
| <b>29635240</b> | 29637005 | 436 | Reverse | CTTAGAACATTAGGAATCATCACTAACTCT  |
| <b>29638520</b> | 29640675 | 438 | Reverse | CTTGCTGCCCTCACCTGTGTTGGTAA      |
| <b>29641134</b> | 29642803 | 440 | Reverse | CTTGCTTTTGTGTAGGGATTTTACTTCTCT  |
| <b>29652798</b> | 29652907 | 444 | Reverse | CTTGCCACAGCCTAGTTTGAGCCTTAGG    |
| <b>29677935</b> | 29691941 | 453 | Reverse | CTTGAGTGAGGCTTTATTTTATGCAATTGG  |

|                 |          |     |         |                                   |
|-----------------|----------|-----|---------|-----------------------------------|
| <b>29703904</b> | 29705290 | 458 | Reverse | CTTCAATAAACAAGTGAGGGCGGGTGCA      |
| <b>29706211</b> | 29709705 | 460 | Reverse | CTTAGAATACTTAGTACTTGGTTTACACCT    |
| <b>29713093</b> | 29716181 | 463 | Reverse | CTTGGAAGTCTTTTCTCAGGAAAGAAAGGACAG |
| <b>29720155</b> | 29727114 | 465 | Reverse | CTTTGGTGTGTAGTACTTTGGAGTGATAAG    |
| <b>29741658</b> | 29748329 | 468 | Reverse | CTTTGGGGATCCTCTGTAAGTGGTTGCTC     |
| <b>29755324</b> | 29755456 | 472 | Reverse | CTTTCTCAATTGATTCTGAAAGGGAAAATG    |
| <b>29766525</b> | 29768740 | 476 | Reverse | CTTCCTCGTTCAGATGGGGTCTGG          |
| <b>29769327</b> | 29770690 | 478 | Reverse | CTTGAGGTTTTCATATCAACAAGGCTCAGT    |
| <b>29774284</b> | 29777533 | 481 | Reverse | CTTGTCTTTGAGAGCCCGGTGCCT          |
| <b>29781542</b> | 29782720 | 483 | Reverse | CTTTTTCCAAGTCAGGTTAGTAAAAGCAGA    |
| <b>29787576</b> | 29797222 | 488 | Reverse | CTTATACCTTTATGACCCATAATGCACAGA    |
| <b>29805363</b> | 29806188 | 490 | Reverse | CTTGATTTCTCAAACACTTATTGATTCGT     |
| <b>29807450</b> | 29809879 | 492 | Reverse | CTTTCTAAAAAGGGAGGAAGGAAATCCAAG    |
| <b>29815594</b> | 29816516 | 494 | Reverse | CTTCCCTGCTGTAGGGGAGAGCG           |
| <b>29820939</b> | 29835163 | 496 | Reverse | CTTGATGAAGATCTTGGCATGGCAATGCAC    |
| <b>29835982</b> | 29836156 | 498 | Reverse | CTTCTTTGCGGAATTCCTAGGACGCTAATG    |
| <b>29838219</b> | 29839237 | 500 | Reverse | CTTGCCGCCAGGAAGCTAATTCCTC         |
| <b>29844732</b> | 29845798 | 504 | Reverse | CTTTCAGAAATGGAAGACTTTTTTAATTCT    |
| <b>29846399</b> | 29848887 | 506 | Reverse | CTTGAATGAGGAAACATAGGCTGAGAGGCC    |
| <b>29849214</b> | 29850560 | 508 | Reverse | CTTTCTACAGCCATGCCATTTACAGAAGCC    |
| <b>29851352</b> | 29854911 | 510 | Reverse | CTTCTACCTGCTAACAACTTTCTCCTGTC     |
| <b>29855303</b> | 29859544 | 512 | Reverse | CTTGCGCCTTTCCAGCAGTCTTGAAACATA    |
| <b>29863830</b> | 29864649 | 514 | Reverse | CTTATCCACCATGGCCCAAGATTATCTTT     |
| <b>29874796</b> | 29884457 | 518 | Reverse | CTTAACCAAGTAACACTACCAACTGGAGAC    |
| <b>29896824</b> | 29897292 | 522 | Reverse | CTTCATCTGACCTCGTTTGACGAAGCTC      |
| <b>29906099</b> | 29907124 | 526 | Reverse | CTTCCAAGTTTCTGAAGCATCCTCACCAGA    |
| <b>29911938</b> | 29913290 | 529 | Reverse | CTTTTGACTAGTGGACTTTATCTGCTCTCA    |

|                 |          |     |         |                                 |
|-----------------|----------|-----|---------|---------------------------------|
| <b>29913932</b> | 29918392 | 532 | Reverse | CTTCAGTTATACAAAGTTCTTTTCATGCCCC |
| <b>29931042</b> | 29938025 | 535 | Reverse | CTTAGGTATAAAAGAAAATCTTTAAATACC  |
| <b>29947213</b> | 29949609 | 537 | Reverse | CTTTCTGACCAAAGTTGACAACAGCACTAT  |
| <b>29953600</b> | 29958998 | 539 | Reverse | CTTTTGTTTATGGTCTGGAATGTACCCATT  |
| <b>29960412</b> | 29963242 | 542 | Reverse | CTTCAGCCCCTTGCCTTGATGCC         |
| <b>29963693</b> | 29963925 | 544 | Reverse | CTTTGAAAGACCCATGTCATAGTACGTGTT  |
| <b>29970156</b> | 29977864 | 547 | Reverse | CTTAGGCTCAAAGATGACTGCAGAGGAGAG  |
| <b>29978716</b> | 29979319 | 549 | Reverse | CTTGCAAGCTAGCTAGCCCAAGGGATAC    |
| <b>29988809</b> | 29991479 | 551 | Reverse | CTTTCAATTTTTCCACATCCTGCCCAACAC  |
| <b>29995277</b> | 29996459 | 553 | Reverse | CTTTCCAGTATTGCTGGAGCTGTAGTCCGA  |
| <b>30001399</b> | 30005000 | 555 | Reverse | CTTTCCAGGCATAAAGCAGAGACAGGCAAA  |
| <b>60917307</b> | 60918190 | 1   | Reverse | CTTTGAGACTAGACCGAAGTCTCCAGAATC  |
| <b>60937639</b> | 60942991 | 6   | Reverse | CTTGAGGAGACAGCACTGCTGGTAGATAG   |
| <b>60966227</b> | 60970541 | 10  | Reverse | CTTACTGTGGGCTACCCATTTGTA CTCTTA |
| <b>60976198</b> | 60977387 | 12  | Reverse | CTTTTATTATAAGAGATCTTAGCTAATGA   |
| <b>60981479</b> | 61000571 | 16  | Reverse | CTTTTTTCCAAATAGTAACCCACAGGGCCA  |

**Table 2.15 – List of 5C primers**

### **2.5.5.3 Ligation Reaction**

The following day, the ligation buffer master mix was made up as detailed in Table 2.16. Due to the NADH content, the Taq DNA ligase × 10 buffer was kept at -80 °C and thawed at 65 °C for 10 min prior to use. The addition of the ligation buffer master mix to the annealing reaction was conducted by opening the lid of the PCR machine and adding the solution directly to the wells of the PCR plate while still at 55 °C. Once the solution was added, the lid was closed and the ligation reaction left for a further hour. Finally, after this incubation the plate was removed and the ligase enzyme denatured by heating the plate at 65 °C for 10 min.



| <b>Ligation Buffer Master Mix</b> | <b>Volume (µl)</b> |
|-----------------------------------|--------------------|
| Taq DNA ligase                    | 0.25               |
| Taq DNA ligase × 10 buffer        | 2.5                |
| MilliQ Water                      | 17.25              |

**Table 2.16 - Volume of reagents needed for ligation buffer master mix**

#### **2.5.5.4 5C library amplification**

The 5C library was amplified using two separate master mixes as detailed in Tables 2.17 and 2.18. Each PCR was performed in a 96 well plate by first adding 5.7 µl master mix 1 followed by 16.3 µl master mix 2 and lastly 3 µl 5C library per reaction. The PCR amplification program is detailed in Table 2.19. PCR was conducted using either 26 or 28 cycles in stage 2 and comparing the results. The resulting products were analysed by agarose gel electrophoresis using a 2.5 % gel. The amplified product appeared as a single band at approximately 120 bp. Typically, no bands were present in the “no template” control while sometimes a faint smear could be seen in the “no ligase” control appearing below the expected molecular weight of the product. If the 5C library produced a band of the correct size the PCR was conducted an additional 3 times and the products combined. The number of cycles chosen for stage 2 in this PCR was chosen by observing the agarose gel from the previous PCR and determining the number of cycles which produced the most product while keeping primer dimers and other non-specific products to a minimum. Typically, 26 cycles were chosen; however sometimes more were necessary (Chapter 5.3.2). The A and P1 key primers are the sequencing adaptors needed for Ion Torrent sequencing. The sequence of these primers is provided in Table 2.20.

| <b>Master Mix 1</b>                   | <b>Volume (µl)</b> |
|---------------------------------------|--------------------|
| 10 × HIFI                             | 2.5                |
| 25 mM dNTPs                           | 0.2                |
| 20 µM A key Primer                    | 0.5                |
| 20 µM P1 key Primer                   | 0.5                |
| 50 mM Mg <sub>2</sub> SO <sub>4</sub> | 2                  |

**Table 2.17 - Master Mix 1 used in the 5C library amplification protocol.**

| <b>Master Mix 2</b> | <b>Volume (µl)</b> |
|---------------------|--------------------|
| Taq polymerase      | 0.2                |
| MilliQ Water        | 16.1               |

**Table 2.18 - Master Mix 2 used in the 5C library amplification protocol.**

| <b>Stage</b> | <b>Temperature (°C)</b> | <b>Time (min:sec)</b> | <b>No. Cycles</b> |
|--------------|-------------------------|-----------------------|-------------------|
| 1            | 95                      | 5:00                  | 1                 |
| 2            | 95                      | 0:30                  | 26/28             |
|              | 60                      | 0:30                  |                   |
|              | 72                      | 0:30                  |                   |
| 3            | 95                      | 0:30                  | 1                 |
|              | 60                      | 0:30                  |                   |
|              | 72                      | 8:00                  |                   |
| 4            | 4                       | forever               | 1                 |

**Table 2.19 – The PCR program used for amplification of 5C products.**

| <b>Primer</b> | <b>Sequence</b>                |
|---------------|--------------------------------|
| A KeyPrimer   | CTGAGTCGGAGACACGCAGGGATGAGATGG |
| P1 Key Primer | CCTCTCTATGGGCAGTCGGTGAT        |

**Table 2.20 – The sequence of the Ion Torrent adaptors used for amplification of the 5C libraries.**

### **2.5.5.5 Purification, Quantification and Sequencing of Amplified 5C libraries**

Amplified 5C libraries were purified using the MinElute PCR purification kit as described in the manufacturer's protocol (<https://www.qiagen.com>). Libraries were eluted in 15 µl Buffer EB (10 mM TrisHCl, pH 8.5) and stored at -20 °C until required. To determine the quantity of 5C product, samples were run on the Agilent Bioanalyser using the DNA 1000 kit. Samples were then diluted in MilliQ water to 2600 pmol/L and sent to the in-house sequencing department. Here, samples were sequenced as described by the Bickmore laboratory (Williamson et al., 2016). In brief, they were diluted to 26 pmol/L and run using an Ion 316™ Chip on the Ion PGM™ Sequencer.

### **2.5.6 Analysis of 4C and 5C data**

All 4C experiments were analysed by our collaborators in the Semple laboratory using the package r3C-seq. Normalisation procedures are described in (Thongjuea et al., 2013). In summary reads were normalised using a power-law reference distribution to generate reads per million (RPM). In both 4C and 5C experiments significant cis interactions were determined data by calculating a Z-score for each interaction (section 5.3.3). The associated p-values and q-values (incorporating false discovery rate) were used to gauge the significance of each interaction.

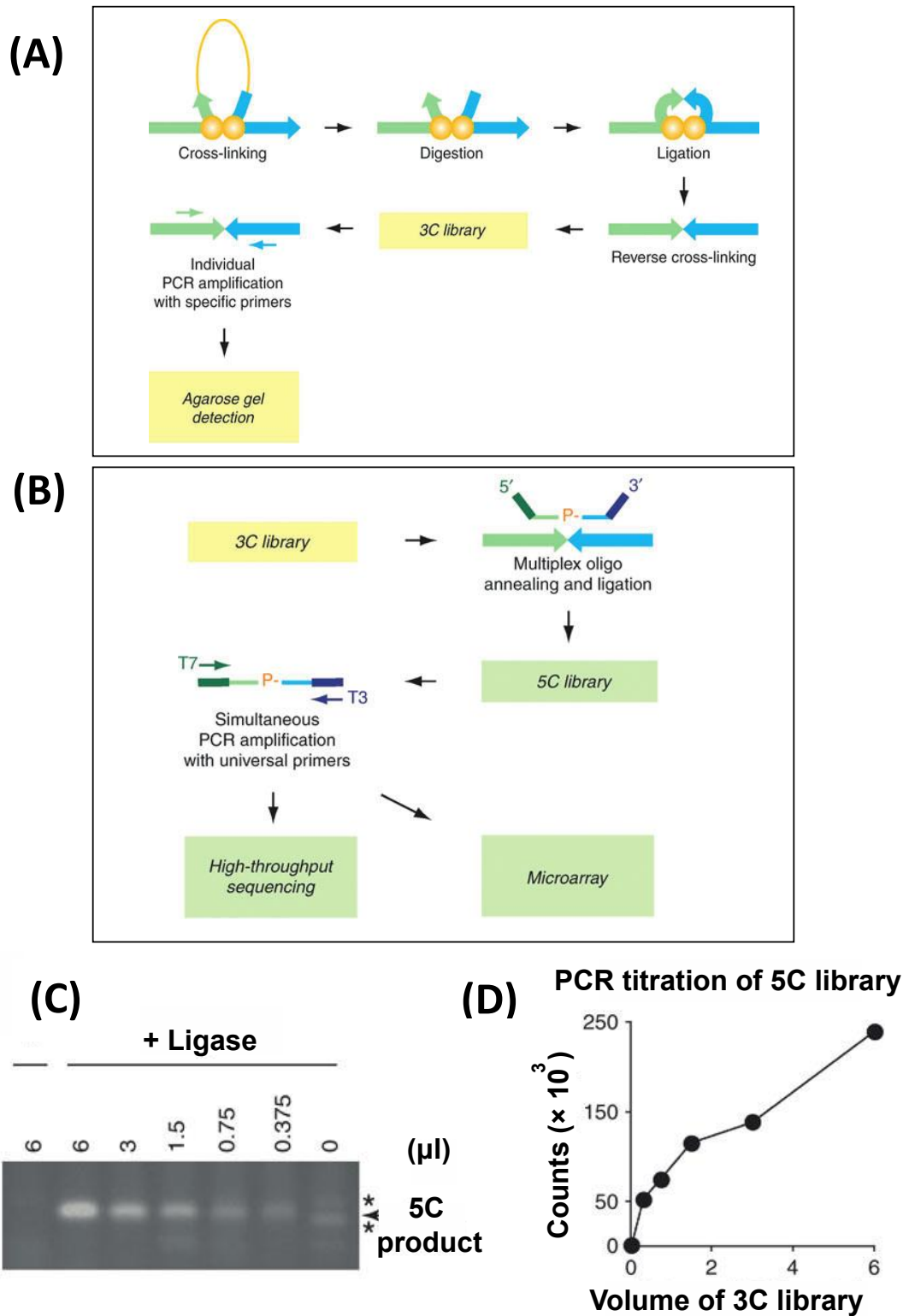


Figure 2.1 – Overview of the 5C protocol and quality control experiments. (A) The first step of the 5C protocol is to create a 3C library. Cross-linking agents such as formaldehyde

are used to join two regions in close physical proximity. Samples are then digested with a restriction enzyme (typically a 6-cutter such as HindIII) followed by a ligation reaction, where the closely spaced regions are ligated together. After reversing the cross-linking, PCR can be used to determine if two regions interact. (B) On completion of a 3C library, 5C primers are annealed to the ends of each restriction fragment. Primers situated next to each other are ligated together using the enzyme Taq DNA ligase. Finally, PCR is conducted using primers which anneal to the ends of each 5C primer. The PCR primers also contain sequencing adaptors for Ion Torrent sequencing thereby allowing the amplified 5C product to undergo next generation sequencing. (C) To examine the quality of the 3C library and determine the quantity needed for 5C experiments, the 3C library is amplified using two primers which anneal to two regions known to interact. Several reactions are performed, each using a decreasing starting volume of 3C library. For high quality 3C samples the intensity of amplified product (shown as a single band on an agarose gel) should decrease linearly with decreasing volume of initial 3C library. (D) The intensity of the bands for each 5C product can then be plotted on a graph vs the initial starting volume of 3C library used in each PCR reaction. The volume of starting 3C library used in the proceeding 5C experiments was selected from the linear portion of this graph. Figure adapted from (Dostie et al., 2006).

## 2.6 Cell Culture Techniques

### 2.6.1 Immortalised Cell lines

14fp, 88fp and MD cell lines were already in use by the Hill laboratory prior to this investigation. These are all immortalised cells derived from the offspring of an immortamouse (H-2KbtsA58) × CD1 cross. The immortamouse contains a temperature sensitive large tumor antigen (permissive at 33 °C) driven by the H-2Kb promoter which is induced upon addition of interferon (Jat et al., 1991). 14fp and 88fp cells are derived from the posterior region of an E11.5 limb bud, while MD cells are derived from the mandibular regions. 14fp and 88fp cells were obtained by members of the Hill laboratory while MD cells were obtained from the Jackson laboratory. For the 14fp and 88fp cell lines, cells were dissected and left for 20 min in trypsin (0.2 g/l)/versene solution after which they were transferred to tissue culture plates containing DMEM media with the following additions: interferon gamma (IFN $\gamma$ ); fetal calf serum (FCS); and penicillin-streptomycin (P/S). These were maintained at 33 °C in tissue culture incubators.

### 2.6.2 Cell Passaging

All immortalised cells were grown no later than passage 18 as beyond this point they grow much more quickly - perhaps becoming transformed. Cells were grown at 33 °C in 3 % O<sub>2</sub> and 5 % CO<sub>2</sub> in tissue culture incubators. The culture media used was DMEM which also contained: interferon gamma (IFN $\gamma$ ); fetal calf serum (FCS); penicillin-streptomycin (P/S); and L-glutamine (referred to as complete DMEM media). To split the cells, the media was first removed from flasks and the cells washed in PBS. After removal of the PBS wash, a small amount of trypsin (0.2 g/l)/versene solution was applied (usually around 1 ml per T75 flask) and left covering the cells for 15 min at 33 °C. Cell flasks were then shaken gently allowing the cells to be stripped from the base of the flask and into solution. Complete DMEM media was then added to 10 × the volume of trypsin/versene solution. The total solution was then divided into separate flasks with extra complete DMEM solution added as appropriate.

### 2.6.3 Drug Treatments

In many experiments immortalised cells were treated with either Trichostatin A (TSA) or DMSO as a control. TSA solution was ordered as a powder and re-suspended in DMSO to 1 mM upon arrival. Both TSA and DMSO solutions were stored at -20 °C when not in use. For

1  $\mu$ M TSA treatments stock TSA solution was diluted 1/1000 in complete DMEM media and added to the cells.

#### **2.6.4 Mycoplasma Testing**

Immortalised cell lines were tested periodically for mycoplasma by the Human Genetics Unit Technical Services department using the Lonza-Mycoalert Mycoplasma Detection Kit (LT07). No positive results were ever detected.

## 2.7 Immunofluorescence

Immunofluorescence was conducted on cells grown on superfrost slides which had been fixed, permeabilized and frozen at  $-80\text{ }^{\circ}\text{C}$  as described in Chapter 2.4.1. Frozen slides were thawed at room temperature and washed in PBS<sup>+</sup> blocking solution (1 % BSA/PBS) for 30 min. For detection of SHH expression, 0.75  $\mu\text{l}$  of the C-terminal antibody AB86462 was diluted in 50  $\mu\text{l}$  PBS<sup>+</sup> blocking solution, added to a  $22 \times 50\text{ mm}$  coverslip and left on the slide at  $4\text{ }^{\circ}\text{C}$  overnight. The following day, the coverslip was removed and slides washed four times (2 min each) in PBS<sup>+</sup> blocking solution. The secondary antibody Alexa Fluor® 488 was diluted 1:1500 in PBS<sup>+</sup> added to a  $22 \times 50\text{ mm}$  coverslip and left on the slide for 30 min in the dark. Upon removal, slides were washed for 2 min in PBS<sup>+</sup> blocking solution an additional 4 times. Lastly, slides were stained with a 1/1000 dilution of DAPI (stock 50  $\mu\text{g/ml}$ ) in PBS for 3 min and mounted in vectashield. Images were obtained using the in-house Zeiss epifluorescence imaging microscope “Limelight” using iVision software.



## **2.8 Mouse Work**

### **2.8.1 Mouse Dissections**

All mice embryos at stage E11. 5 were obtained from the Edinburgh Biomedical Research Facility (BRF). All mice used were the species CD1 which were originally obtained from Charles River Laboratories.

# Chapter 3

---

## **Using Sonic Hedgehog Inducible Cell Lines as a Model System to Study the Limb**

### 3.1 Introduction

One of the main ambitions of the Hill laboratory is to elucidate the processes leading to induction of the *Shh* gene in the mouse embryonic limb bud. In particular, there is an interest in how the ZRS, the enhancer responsible for activating *Shh* in the limb bud functions. Among many lines of inquiry, we are interested in: changes in histone marks at the ZRS upon induction of the *Shh* gene; the specificity of the ZRS towards *Shh* in both *Shh* expressing and non-expressing cells; the compaction of the ZRS-*Shh* region in *Shh* expressing and non-expressing cells; and the transcription factors and other proteins binding to the ZRS in *Shh* expressing and non-expressing cells. However, the E11.5 mouse embryonic limb bud consists of an extremely complex series of interconnecting signalling pathways. *In vivo* manipulation of individual components from a single pathway can have a knock on effect on several others and therefore dissecting the series of events within a single pathway can prove challenging. Therefore, to look at the process of *Shh* induction by the ZRS in isolation it was decided that a more controlled system should be examined. A *Shh* inducible system was suggested, where cultured cell lines could be used to analyse the ZRS and *Shh* when the gene was not expressed and then how both components were affected upon activation of the gene. The results from such experiments could be then be compared to *in vivo* experiments using E11.5 limb bud tissue.

In addition, using cultured cell lines to study gene activation can have several advantages. For example, modern techniques used to analyse chromatin structure such as chromosome conformation capture (3C) usually require a large number of cells in the range of  $10^6$  -  $10^7$ . In the case of the mouse limb SHH is expressed over a period of only a few days (E9.75 – E12, Chapter 1.3.1) and is produced in a small sub-section of the limb referred to as the zone of polarising activity (ZPA) which is found in the posterior distal region. At E11.5, when SHH expression is at its maximum, we estimate the limb bud consists of roughly 200,000 – 300,000 cells with the ZPA occupying only roughly a twentieth of this area. Therefore, if performing mouse dissections to isolate SHH expressing cells, several mouse litters may be required. Normal dissections are also inaccurate as it is extremely difficult to separate the ZPA from the remainder of the limb bud. At best the limb can be dissected into different sections i.e. into proximal and distal limb sections or anterior and posterior sections, where the ZPA is contained in either the distal or posterior portions. However, using cultured cell lines it is relatively easy to harvest large numbers of cells relatively quickly and these types of experiments can be conducted over a much smaller time scale.

### 3.1.1 The 14fp cell line

The 14fp cell line are immortalised cells (cells derived from the immortal mouse containing the temperature sensitive T antigen (Chapter 2.6.1) arising from the posterior region of an E11.5 limb bud. These cells are grown in culture at 33 °C with 3 % O<sub>2</sub> and 5 % CO<sub>2</sub> in the presence of interferon gamma (IFN $\gamma$ ); fetal calf serum (FCS); penicillin-streptomycin (P/S); and L-glutamine. They can be grown up to around passage 18 after which clear indications of transformation are evident. PCR has shown that some genes expressed within the E11.5 limb bud are also expressed within the 14fp cells (such as Hand2 and Gremlin) suggesting that the cells are largely reflective of early embryonic limb bud cells (Figure 3.1(A)). One key exception is that no detectable levels of SHH are evident.

Peluso *et al.* (manuscript in preparation) have shown that treatment of 14fp cells with trichostatin A (TSA) is sufficient to induce a low level expression of *Shh*. TSA is a histone deacetylase inhibitor capable of inhibiting class I and class II HDACs (HDACs: 1, 3, 4, 6, 10). It has also been shown to affect the expression of a number of genes in chick limb buds (Zhao *et al.*, 2009). In 14fp cells, increasing the concentration of TSA up to 1  $\mu$ M increases the output of SHH expressed. Above this concentration, there is an observable increase in cell death. As a result, a concentration of 1  $\mu$ M TSA was chosen for all 14fp treatments. At this concentration, time course experiments have shown *Shh* expression to increase incrementally up to 24 hrs after which expression decreases and returns to basal levels by 36 hr (Figure 3.1(B)).

Chromatin immunoprecipitation (ChIP) experiments have shown that in these cells the Zone of Polarising Activity Regulatory Sequence (ZRS) - the enhancer responsible for SHH expression in the limb bud - is in a poised state as indicated by histone H3K4 monomethylation (H3K4me1) and the binding of the protein P300 which, in combination are considered indicators of an enhancer poised for activation (Creyghton *et al.*, 2010). On TSA addition, - H3K27ac, a marker of active enhancers - is also detected at the ZRS. At this stage, the enhancer histone marks at the ZRS reflect those observed in the distal region of the E11.5 limb bud (Silvia Peluso, manuscript in preparation).

### 3.1.2 Additional Cell Lines

The 14fp cell lines are the most well characterized cells used in this investigation as ChIP experiments have determined the activity status of the ZRS both before and after TSA treatment. The 88fp cell line, derived from the same region and the same stage as the 14fp cells, is also immortalized and can induce *Shh* upon TSA treatment when grown in similar conditions. These cells have been used in experiments to determine if the properties of the 14fp cells are unique to this cell line or, alternatively, are shared among immortalized limb cells originating from the limb.

The MD cell line are immortalized cells that originate from the mandible of the E11.5 mouse embryo. SHH is not expressed in this region in mouse embryonic tissue and the MD cell line does not induce *Shh* expression upon TSA treatment. When compared to the 14fp cells, there is negligible H3K4me1 and H3K27ac at the ZRS when TSA is either absent or present, suggesting that the enhancer is inactive within these cell lines. The MD cell line is therefore a good control for this investigation (Silvia Peluso, manuscript in preparation).

### 3.1.3 Alternatives to TSA

In addition to TSA drug treatment, Silvia Peluso (project supervisor) has shown that either knockdown of the transcription factor PEA3 or upregulation GABP $\alpha$  is sufficient to induce SHH expression in 14fp cells (Figure 3.1(C)(D)). Immunoprecipitation experiments have identified PEA3 as a binding partner of HDAC2 and GABP $\alpha$  as a binding partner of P300. Both GABP $\alpha$  and PEA3 are members of the ETS factor family and have binding sites within the ZRS (Chapter 1.3.1). Within the limb bud GABP $\alpha$  is expressed throughout the distal edge with a larger expression domain in posterior regions overlapping the ZRS than anterior regions. PEA3 (also known as ETV4) is expressed throughout the distal edge with expression distributed equally between anterior and posterior sites. Both proteins have been shown to regulate *Shh* expression within the limb bud with GABP $\alpha$  increasing levels of SHH and ETV4 reducing SHH. The balance of both proteins within the limb bud is therefore essential for restricting *Shh* expression to the ZPA. The importance of maintaining this equilibrium is illustrated by ZRS mutants where a ETV4 binding site has been substituted for a GABP $\alpha$  site. These mutants display ectopic *Shh* expression in anterior regions.

In this chapter RNA-seq experiments were conducted on both TSA treated and control 14fp cells. ChIP experiments suggest that TSA induces the *Shh* gene through the ZRS. As a result,

the addition of TSA to 14fp cells has the potential to act as a good model system for studying the mechanism by which the ZRS activates the *Shh* gene. RNA-seq experiments were conducted on these cells to investigate the effect TSA has on other genes (perhaps activating additional developmental enhancers) and to determine if the addition of TSA to the 14fp cell activates non-expressing genes that would normally be active in the distal limb bud.

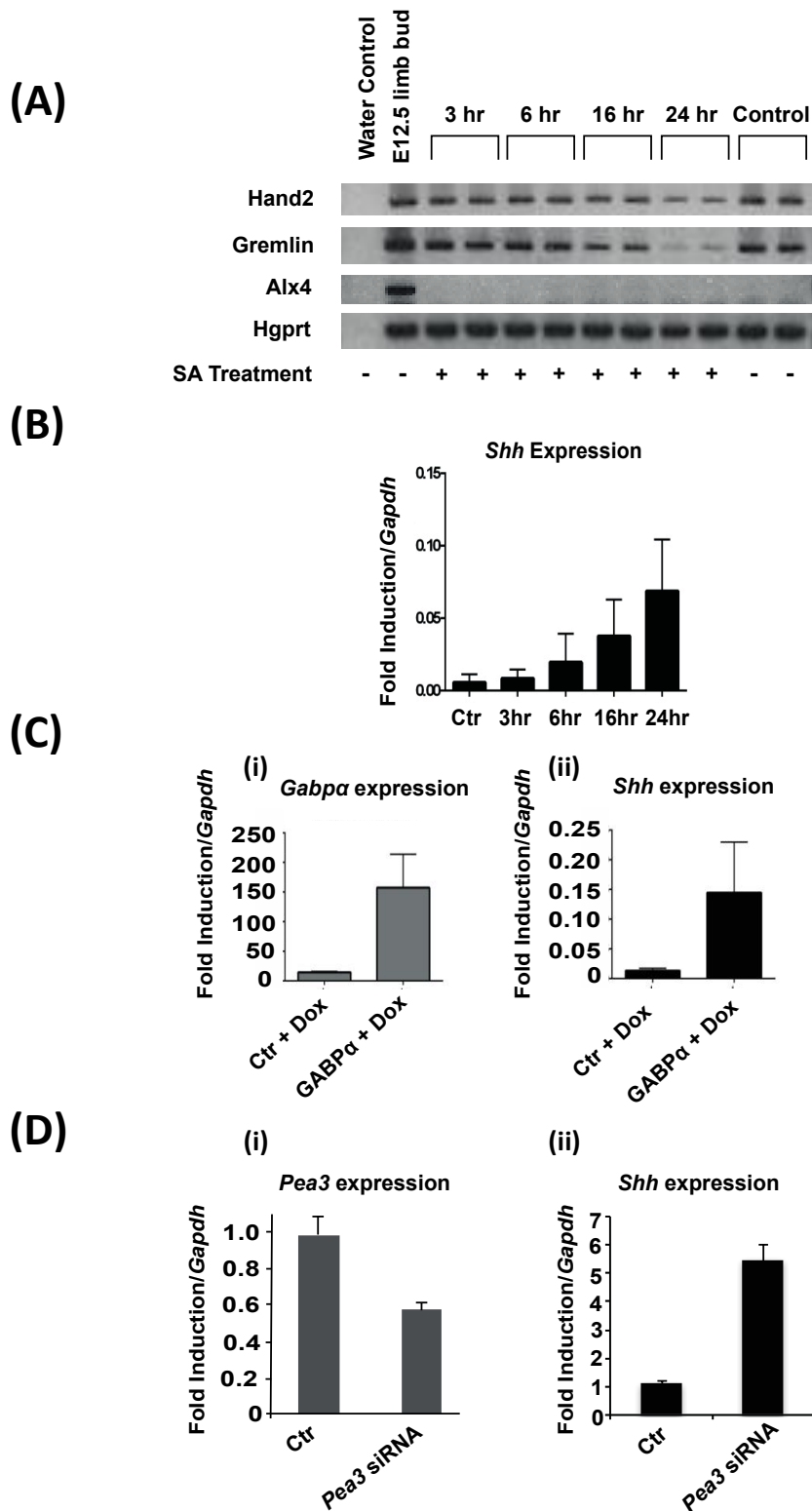


Figure 3.1 – Project background – data provided by Christine Mordstein and Silvia Peluso. (A) Agarose gel electrophoresis of RT-PCR samples showing that after 24 hr TSA treatment *Gremlin* and *Hand2* expression decreases in 14fp cells, while the housekeeping gene *Hgprt* is unaffected. *Alx4* is not detected in TSA treated and control 14fp cells. (B-D) RT-qPCR was used to test the expression of *Shh*, *Gabpa* and *Pea3* in 14fp cells under different

conditions. (B) *Shh* expression increases upon TSA treatment up to 24 hr and returns to base line levels by 36 hr (36 hr time point omitted). (C) A GABP $\alpha$  producing, doxycycline inducible vector was introduced into 14fp cells. An increase in *Gabpa* expression (i) coincided with an increase in *Shh* expression (ii). (D) Knock-down of *Pea3* levels (i) coincided with an increase in *Shh* expression (ii).



## 3.2 Aims

### 3.2.1 General Aims

The aim of this section was to characterize the 14fp cells further by using RNA-seq to determine how closely they resemble an embryonic limb bud at day E11.5. If the cells expressed the same genes as the E11.5 limb bud, they could be used as to the model system to investigate limb signaling pathways. We also looked at different aspects of the 88fp cell lines to see how they compare with those of the 14fp cells.

### 3.2.2 Experimental Aims

The experimental aims of this section are as follows:

- To investigate gene expression differences between TSA treated (*Shh* expressing) and control (*Shh* non-expressing) 14fp cells.
- To investigate gene expression differences between distal (*Shh* expressing) and proximal (*Shh* non-expressing) portions of the E11.5 limb bud.
- To investigate gene expression differences between the 14fp cells (control and TSA treated) and the distal and proximal E11.5 limb bud sections.
- To investigate the concentrations of TSA needed to regulate the maximum SHH expression in 88fp cells.
- To investigate the time point where SHH expression is at its maximum after treatment with the optimum concentration of TSA in 88fp cells.
- To investigate the effect of TSA treatment on *Pea3* and *Gabpa* levels in 14fp and 88fp cells.

## **3.3 Experimental Procedures**

### **3.3.1 14fp Cells: Preparation of RNA for sequencing**

RNA for both treated and untreated 14fp cells was provided by Silvia Peluso. This involved treating 14fp cells with either 1  $\mu$ M TSA or DMSO (acting as a control) for 24 hrs. At this point cells were harvested by scraping and RNA was isolated using a standard trizol extraction protocol (Chapter 2.3.1). The integrity of RNA was tested by agarose gel electrophoresis, and analysed using the nanodrop spectrophotometer and by running on an Agilent Technologies 2100 Bioanalyser. Only RNA with an OD 260/280 ratio greater than 1.8; an 260/230 ratio greater than 1.7 and an RNA integrity number of 8 or above was considered for sequencing. This process was repeated on a different day, thereby providing biological replicates of both control and TSA treated 14fp cells.

### **3.3.2 E11.5 Limb Cells: Preparation of RNA for sequencing**

RNA for both proximal and distal limb sections was provided by Silvia Peluso. In summary, the limb buds from several CD1 embryos at E11.5 were dissected in PBS and divided into both proximal and distal sections. All proximal sections were collated into the same vial and this was repeated for all distal sections. Both mouse forelimbs and hindlimbs were combined for these experiments. The RNA extraction procedure and RNA quality checks were performed as Chapter 3.3.1. Two biological replicates of distal limb sections and two biological replicates of proximal limb sections were sent for sequencing.

### **3.2.1 RNA-sequencing: Analysis Pipeline**

Raw sequencing reads were first checked for sequencing quality using the tool FastQC. After checking, they were changed into a suitable format for use in further processing steps using FastQC Groomer. After trimming poor quality reads (using FastQ quality trimmer tool) and the removal of sequencing adaptors (using Clip tool), Tophat was used to align the processed reads to the mouse mm9 genome (the Tophat settings are provided in Appendices – Table 1). Differential expression was conducted using the Cuffdiff tool with the default settings. Finally, CummeRbund was used to visualize the output from Cuffdiff. This process was repeated for both cell line and limb dissection experiments.

## 3.4 Results and Discussion

### 3.4.1 Processing of RNA sequencing data

RNA was prepared and sequenced by GATC Biotech AG according to Chapter 3.3.1. The Inview Transcriptome Explore package was used where mRNA was selected by poly(A) enrichment and then subsequently fragmented. cDNA was then prepared using random hexamer priming. After adapter ligation, Illumina sequencing was conducted yielding single end 50 bp reads. For the cell line experiments, more than 20 million reads were obtained for each sample (Table 3.1) with an alignment rate to the mm9 genome of greater than 80 % (after trimming and removal of adapter sequences). For the limb bud experiments, more than 25 million reads were obtained for each sample with a mapping rate of greater than 95 %. In terms of read number, the greatest variability between replicates was observed for the control 14fp samples, with one replicate yielding 88 million reads and the other only 25 million reads.

The output from Cuffdiff gene differential expression testing was filtered in two parts. Firstly, for gene IDs which had passed the Cuffdiff statistical testing procedures and secondly for gene IDs with significant changes in FPKM (Fragments per Kilobase of Sequence Mapped) as determined by the Cuffdiff statistical testing. The resulting lists were then ranked in either ascending order or descending order of  $\log_2$ fold change in FPKM values. The top 20 hits were extracted and presented in this section. For cell line experiments a positive change in  $\log_2$ fold FPKM relate to a significant upregulation in gene expression from the control 14fp cells to the TSA treated cells. In contrast a negative value relates to a significant reduction in gene expression from the control 14fp cells to the TSA treated cells. A similar rationale can be applied to the limb bud experiments where a positive change in  $\log_2$ fold FPKM relates to a significant upregulation in gene expression from the proximal limb dissection to the distal limb dissection and a negative  $\log_2$ fold FPKM relates to a significant reduction in gene expression. In total, between control and TSA treated 14fp cells 7,872 genes showed significant differential expression while 1,500 genes showed significant differential expression between the proximal and distal sections of the E11.5 limb bud.

Using GOrilla, an online bioinformatics tool which specializes in clustering genes according to their associated GO (Gene Ontology) terms to determine which GO terms are enriched within a sample, we analysed the biological processes associated with the differentially expressed genes in both cell lines and limb bud experiments. The top 20 GO terms (ranked

according to p-value) associated with differentially expressed genes from cell line and limb bud experiments are also presented in this section. The input to GOrilla was a list of differentially expressed genes ranked in descending order of  $\log_2$ fold FPKM.

### **3.4.2 Quality Control of RNA sequencing data**

It was important to confirm that changes in the differential expression of genes in these experiments were due to either: (a) the addition of TSA to 14fp cells; or, (b) to whether genes were being expressed in proximal or distal limb sections; i.e., it was important to confirm that the expression of genes in replicate samples were similar to each other and that large variation between replicates was not affecting the analysis. Therefore, to check that replicate samples had similar gene expression profiles to each other data was plotted on both mds (multidimensional scaling) and dendrogram graphs (Figure 3.2(A)(B)). Both of these plots show that the replicates for each condition in both sets of experiments cluster closer together than they do with other conditions. For example, in the limb cell experiments replicates of the proximal limb dissections cluster closer to each other than they do to the replicates of the distal dissections. This confirms that the major source of variation in these experiments is between the different conditions and not between replicates within a condition.

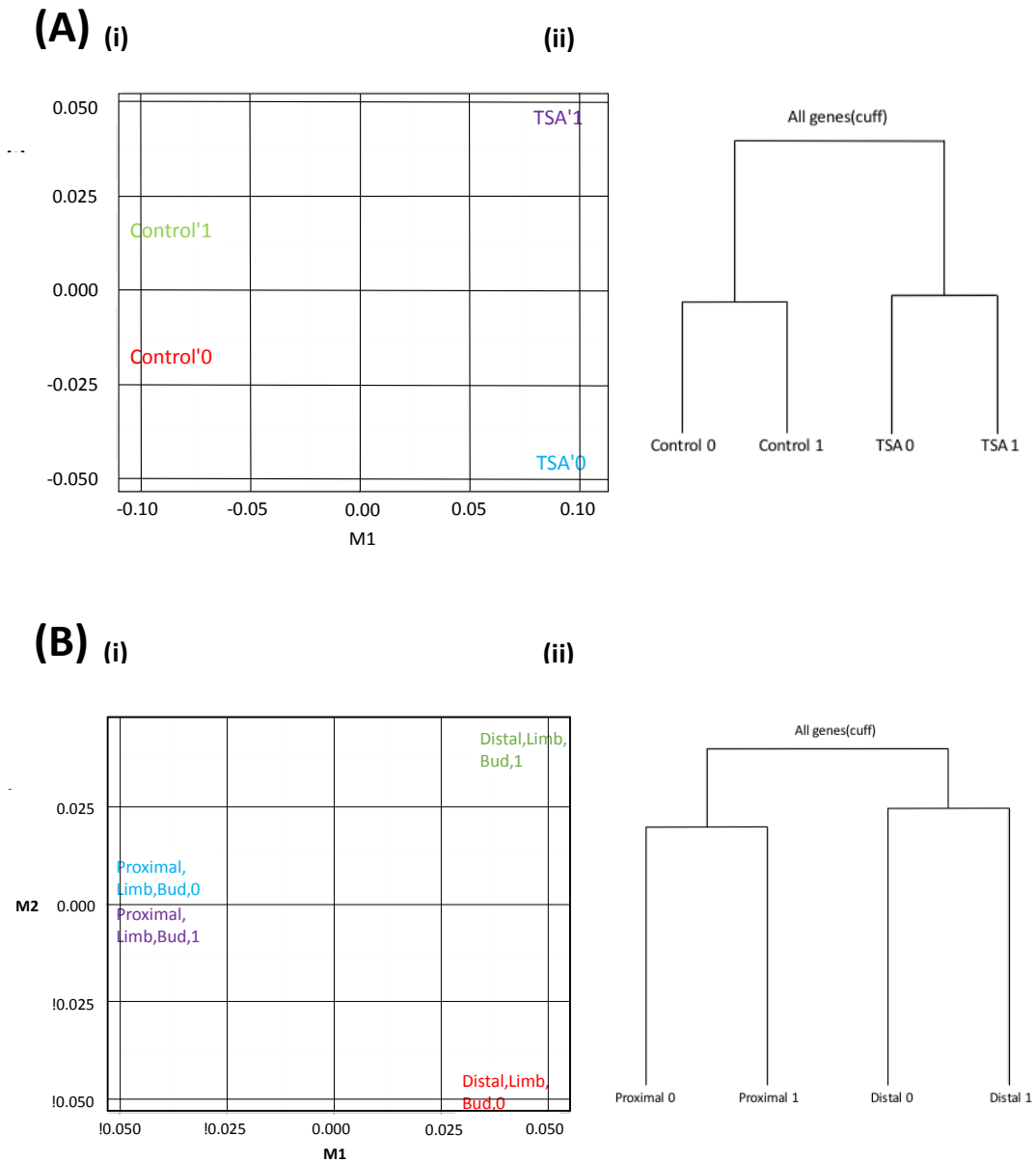
### **3.4.3. Genome wide FPKM changes for cell line and limb dissection experiments**

Analysis of the boxplots for both cell lines and limb dissection experiments suggests that the global distribution of  $\log_{10}$ FPKM values show little variation between both treated and untreated 14fp cell lines and between proximal and distal limb dissections (the median  $\log_{10}$ FPKM value – depicted as the horizontal line within each boxplot – is similar for each of the four samples). This suggests that for the cell lines there is no observable collective increase or decrease in gene expression upon TSA treatment (assuming that TSA treatment causes no significant increase or decrease in the amount of RNA produced per cell – this could be determined by repeating the experiment with standardized controls (Lovén et al., 2012)). Also for the limb dissections there is no observable collective difference in the expression of all genes between proximal and distal sections. The distribution of  $\log_{10}$ FPKM values is greater in the 14fp cell line experiments than in the limb dissection experiments as seen by the greater interquartile range (determined by the distance between the top and bottom of each boxplot) on boxplots for both treated and untreated 14fp cells than boxplots for the separate limb regions

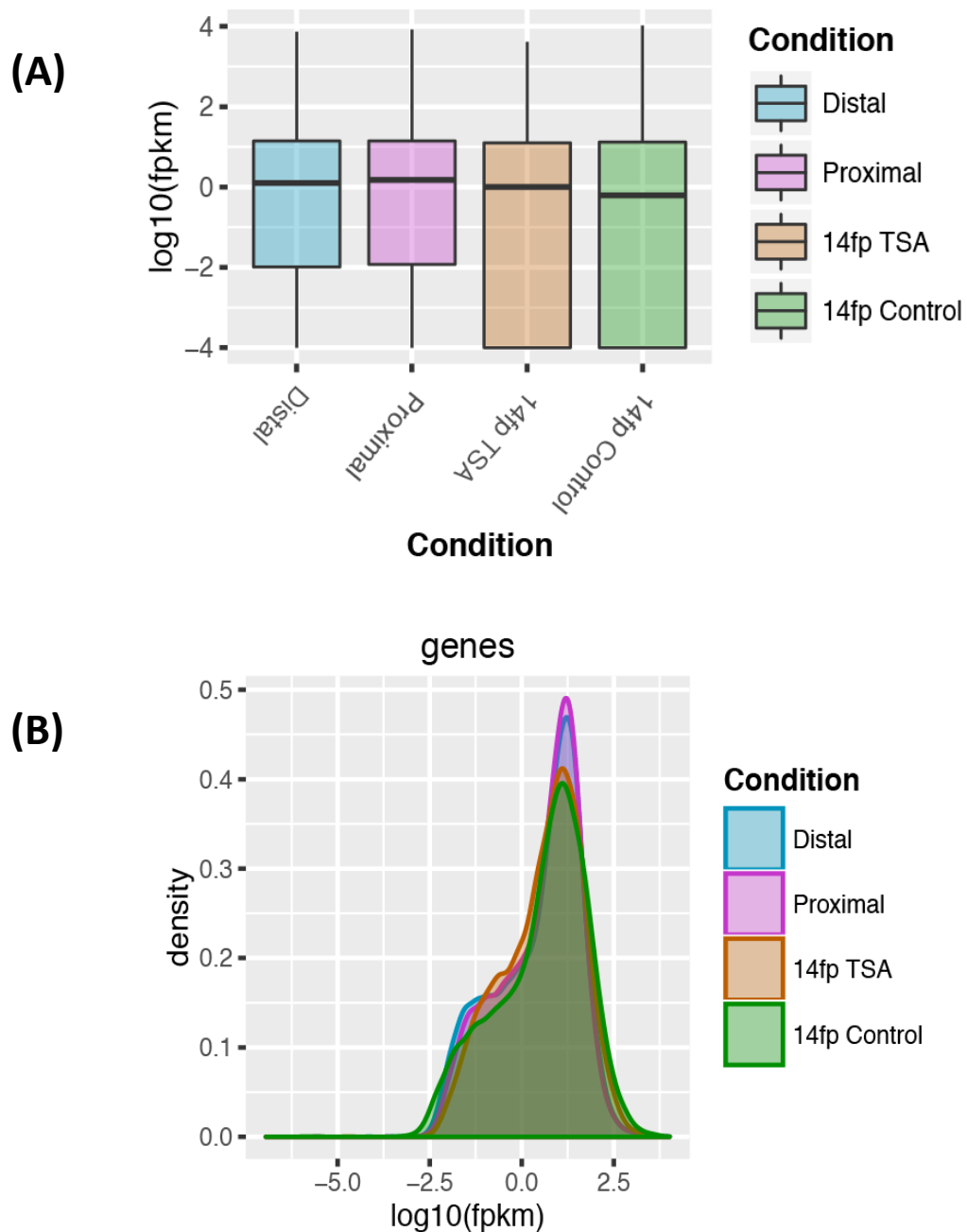
(Figure 3.3(A)). The density plots for all samples are similar and show that a large number of genes have a positive  $\log_{10}$ FPKM (as indicated by the peak of the graph), while there are also a substantial number of genes with a negative  $\log_{10}$ FPKM (Figure 3.3(B)).

| <b>Sample</b>                        | <b>No. Generated Reads</b> | <b>No. Mapped Reads</b> |
|--------------------------------------|----------------------------|-------------------------|
| <b>14fp Control Replicate 1</b>      | 88061533                   | 69502864                |
| <b>14fp Control Replicate 2</b>      | 24927241                   | 23648587                |
| <b>14fp TSA Replicate 1</b>          | 54252778                   | 43886851                |
| <b>14fp TSA Replicate 2</b>          | 67963171                   | 46952553                |
| <b>Proximal Limb Bud Replicate 1</b> | 38861118                   | 31834984                |
| <b>Proximal Limb Bud Replicate 2</b> | 28744255                   | 23551802                |
| <b>Distal Limb Bud Replicate 1</b>   | 38830914                   | 31473416                |
| <b>Distal Limb Bud Replicate 2</b>   | 40000755                   | 32650196                |

**Table 3.1 – The number of reads generated and the number of reads mapped to the mm9 genome for each sample in the RNA-sequencing experiments.**



**Figure 3.2 – Multidimensional scaling (mds) and dendrogram plots for both (A) cell line experiments and (B) limb dissection experiments (A)(i) TSA and control samples clearly separate along the first dimension of the mds plot. (ii) TSA replicates cluster closer together than to the control replicates on the dendrogram. (B)(i) Proximal and distal limb sections clearly separate along the first dimension of the mds plot. (ii) Distal section replicates cluster closer together than to the proximal replicates on the dendrogram.**



**Figure 3.3 – Boxplots (A) and density plots (B) for both cell line and limb dissection experiments. (A) Boxplots show that the median  $\log_{10}$ FPKM value (middle line in the boxes) is very similar between all samples. The distribution of  $\log_{10}$ FPKM values for all genes in both proximal and distal limb sections is much narrower than the distributions for treated and untreated cells, as indicated by the smaller size of the boxes (and thus a lower interquartile range). (B) The density plots for all samples show that a large number of genes have a positive  $\log_{10}$ FPKM value (as indicated by the peak), while there are also a substantial number of genes with negative  $\log_{10}$ FPKM values (shoulder of the peak).**

### 3.4.4 TSA treatment of 14fp cells affects a variety of cellular processes

Upon analysis of the top 20 genes with significantly upregulated expression (Table 3.2 and Figure 3.4) and the top 20 genes with significantly reduced expression (Table 3.3 and Figure 3.5), it is obvious that the processes other than those affecting the limb bud signaling pathways are affected. The top hit in Table 3.2 is the gene *Atp1a3* (8.36944 log<sub>2</sub>FPKM fold change, q=0.00045389) which is known to play a role in sodium and potassium ion transport. Analysis of the top GOrilla hits (Table 3.4) shows that the process of “cellular sodium ion homeostasis” (q=1.32E-07) is enriched which in addition to *Atp1a3* also includes the two other genes *Tesc* (7.18772 log<sub>2</sub>FPKM fold change, q=0.00265304) and *Atp1b2* (7.19875 log<sub>2</sub>FPKM fold change, q=0.000163866) which both appear in Table 3.2.

Unlike the limb experiments, where the GO term “limb morphogenesis” is in the top 20 Gorilla hits (Table 3.7), this GO term was not in the top 20 Gorilla hits for the cell line experiments (Table 3.4). This suggests that although signaling pathways involved in patterning the limb bud are affected by TSA treatment, the expression of genes involved in other processes (such as sodium and potassium ion transport) are altered to a greater extent.

We selected a number of genes known to play a role in embryonic limb development and analyzed changes in gene expression. Most importantly, we observed an increase in the expression of the *Shh* gene (Figure 3.10). However, this increase was very small with the gene failing the Cuffdiff statistical testing procedures using default parameters. When these parameters were altered so that a reduced number of alignments could be included in the statistical tests, differential expression of *Shh* was deemed significant. Of note, there is also a significant upregulation of *Fgf8* (q=0.000163866) and a significant downregulation of *Grem1* (-2.37148 log<sub>2</sub>FPKM fold change, q=0.000163866) which are both involved in *Shh* signaling pathways. A decrease in *Grem1* expression is at odds to the distal limb bud where *Shh* induces *Grem1* expression. As stated previously TSA is an inhibitor of class I and class II HDACs. HDACs 1 to 11 are expressed in 14fp cells with HDACs 1, 6, 7, 8, 10 and 11 showing significant differential expression. On the other hand, HDACs 2, 3, 4 and 9 do not show significant differential expression. However, there is substantial expression of all of these genes in both treated and untreated cells (Figure 3.6).

Other investigations (Towers et al., 2008) have analysed the effect of histone deacetylases such as TSA when implanted on beads into chick limb buds. One such study looked at the effect of Valpoirc Acid (VPA) on the processes of chondrogenesis in E12 forelimb buds and



confirmed that VPA treatment resulted in a significant decrease in both SOX9, SOX5 and SOX6 (Paradis and Hales, 2013). However, we found no statistical difference in the differential expression of these genes between TSA treated and untreated 14fp cells.

One observation from treating 14fp cells with TSA is the large extent of cell death that has occurred within 24 hrs. This correlates with the well-known role of TSA in regulating apoptosis. It was hypothesized, therefore, that TSA treatment may lead to the upregulation of several apoptotic markers. In particular, one study looking at the effect of TSA on pancreatic cancer cell lines noticed a significant increase in expression of the apoptosis inducing gene *Bcl2l11* (*Bim*) and reduction in expression of the anti-apoptotic genes *Bcl2l1* (*Bcl-xl*) and *Bcl2l2* (*Bcl-w*) (Moore et al., 2004). Indeed, a similar effect was observed on the treatment of 14fp cells with TSA as *Bcl2l11* was significantly induced (2.68475 log<sub>2</sub>FPKM fold change, q=0.000163866) and both *Bcl2l1* (-1.70833 log<sub>2</sub>FPKM fold change, q=0.000163866) and *Bcl2l2* (-1.04631 log<sub>2</sub>FPKM fold change, q=0.000722654) were significantly reduced. TSA has also been shown to slow cell growth by restricting progression through the cell cycle at the G1 or G2 stage. For example, components regulating progression through these stages such as cyclin D1 and CDK4 were downregulated, while p21<sup>waf1</sup> and p53 were upregulated in one study looking at the effect of TSA in human breast epithelial cells (Park et al., 2008). We found a significant reduction in gene expression of *Cdk4*, *Cdk6*, *cyclin D*, *cyclin E*, *p53* and *p21* upon TSA treatment suggesting that this stage of the cell cycle is severely compromised.

| Gene            | Locus                     | log <sub>2</sub> (fold change) | q value     |
|-----------------|---------------------------|--------------------------------|-------------|
| <b>Atp1a3</b>   | chr7:25763187-25790914    | 8.36944                        | 0.00045389  |
| <b>Elavl3</b>   | chr9:21819448-21856467    | 7.51933                        | 0.00566822  |
| <b>Padi2</b>    | chr4:140462274-140513817  | 7.34784                        | 0.00640266  |
| <b>Prss16</b>   | chr13:22094044-22101610   | 7.29626                        | 0.00410193  |
| <b>Nptx1</b>    | chr11:119400032-119409134 | 7.25427                        | 0.00453523  |
| <b>Atp1b2</b>   | chr11:69413251-69419462   | 7.19875                        | 0.000163866 |
| <b>Tesc</b>     | chr5:118477832-118511879  | 7.18772                        | 0.00265304  |
| <b>Rassf4</b>   | chr6:116583025-116623854  | 7.09386                        | 0.0137202   |
| <b>Bex2</b>     | chrX:132601103-132602775  | 7.07583                        | 0.0334972   |
| <b>Nrxn2</b>    | chr19:6418737-6533217     | 6.99842                        | 0.000163866 |
| <b>Slc40a1</b>  | chr1:45964914-45982439    | 6.97537                        | 0.0378811   |
| <b>Epas1</b>    | chr17:87153203-87232750   | 6.96579                        | 0.000163866 |
| <b>Pou3f1</b>   | chr4:124334888-124337899  | 6.96488                        | 0.00898437  |
| <b>Clu</b>      | chr14:66587319-66600382   | 6.95934                        | 0.000163866 |
| <b>Dpysl5</b>   | chr5:31014267-31101742    | 6.83975                        | 0.000163866 |
| <b>Vgf</b>      | chr5:137506164-137509221  | 6.80705                        | 0.000163866 |
| <b>Cadm3</b>    | chr1:175264384-175297826  | 6.73671                        | 0.00265304  |
| <b>Lama1</b>    | chr17:68046604-68171985   | 6.73339                        | 0.0080163   |
| <b>Irf8</b>     | chr8:123260275-123280592  | 6.70344                        | 0.000163866 |
| <b>Mapk8ip2</b> | chr15:89284341-89292878   | 6.68748                        | 0.00850558  |

**Table 3.2 - The top 20 genes which show the greatest positive change in log<sub>2</sub>fold FPKM values between untreated and TSA treated 14fp cells. i.e. the top 20 ranked genes which show an increase in expression upon TSA treatment. The FDR adjusted q-value for each gene is provided. These values were taken from the Cuffdiff output folder “gene differential expression testing”.**

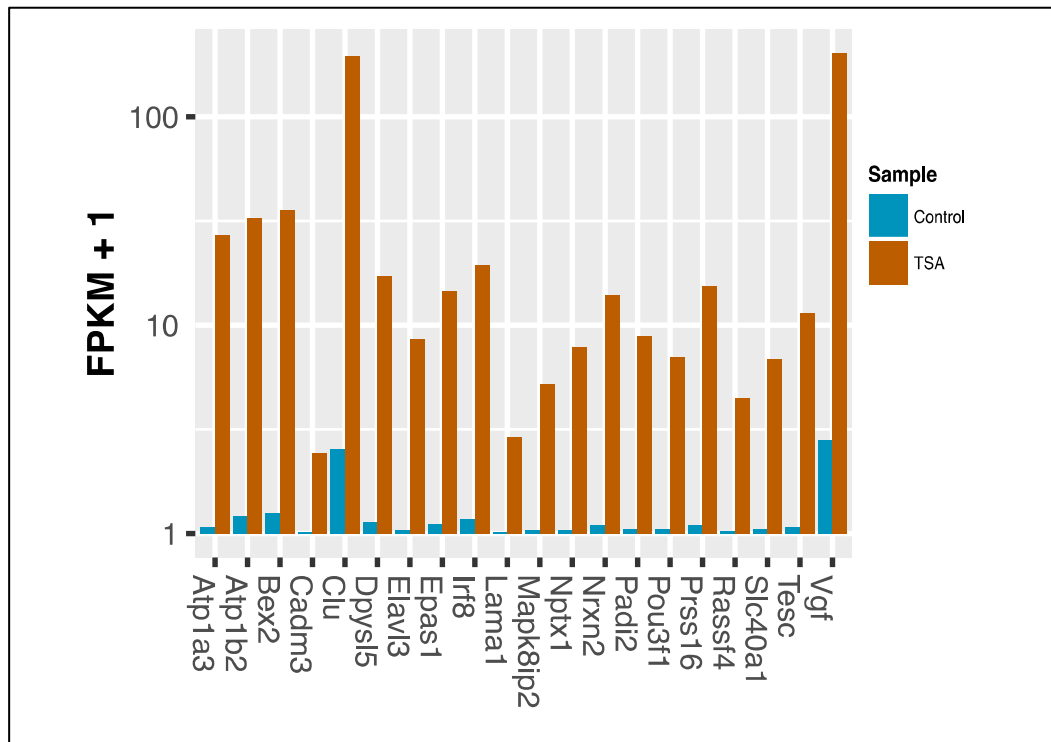


Figure 3.4 – Bar chart displaying the change in FPKM value of the top 20 most positively differentially expressed genes from Table 3.2. i.e. the top 20 ranked genes which show an increase in expression upon TSA treatment. FPKM values are displayed as a log scale along the vertical axis. Expression of genes from untreated 14fp cells are highlighted in blue, while the expression of genes from TSA treated cells are represented as orange bars.

| Gene             | Locus                     | log <sub>2</sub> (fold change) | q value     |
|------------------|---------------------------|--------------------------------|-------------|
| <b>Wisp2</b>     | chr2:163646569-163658883  | -7.12989                       | 0.000163866 |
| <b>Fam150a</b>   | chr1:6349411-6384812      | -7.08516                       | 0.00463929  |
| <b>Tgfb1i1</b>   | chr7:135390384-135397226  | -6.65208                       | 0.000163866 |
| <b>Peg12</b>     | chr7:69606756-69609396    | -6.27441                       | 0.011576    |
| <b>Calhm2</b>    | chr19:47206721-47212784   | -5.84305                       | 0.000163866 |
| <b>Rcbtb2</b>    | chr14:73542316-73583861   | -5.82715                       | 0.000163866 |
| <b>Lsp1</b>      | chr7:149646774-149701914  | -5.75427                       | 0.000163866 |
| <b>Clcf1</b>     | chr19:4214391-4222615     | -5.70572                       | 0.00577033  |
| <b>Wisp1</b>     | chr15:66722954-66754761   | -5.58081                       | 0.000163866 |
| <b>Masp1</b>     | chr16:23451857-23520663   | -5.50122                       | 0.000163866 |
| <b>Il1rl1</b>    | chr1:40496493-40522259    | -5.48584                       | 0.000163866 |
| <b>Anxa8</b>     | chr14:34864315-34915940   | -5.44822                       | 0.000163866 |
| <b>Krt7</b>      | chr15:101242833-101258237 | -5.42591                       | 0.00577033  |
| <b>Serpinb9b</b> | chr13:33119282-33132427   | -5.41885                       | 0.00149063  |
| <b>Clec2d</b>    | chr6:129130632-129136553  | -5.38279                       | 0.000163866 |
| <b>Efs</b>       | chr14:55535379-55545625   | -5.37985                       | 0.000163866 |
| <b>Hdac7</b>     | chr15:97614795-97674933   | -5.29212                       | 0.0420699   |
| <b>Gm53</b>      | chr11:96112973-96125798   | -5.28774                       | 0.00536537  |
| <b>Ptpmt1</b>    | chr2:90750869-90758207    | -5.24323                       | 0.000163866 |
| <b>Irgm1</b>     | chr11:48678750-48684848   | -5.12431                       | 0.000163866 |

**Table 3.3 - The top 20 genes which show the greatest negative change in log<sub>2</sub>fold FPKM values between untreated and TSA treated 14fp cells. i.e. the top 20 ranked genes which show a decrease in expression upon TSA treatment. The FDR adjusted q-value for each gene is provided. These values were taken from the Cuffdiff output folder “gene differential expression testing”.**

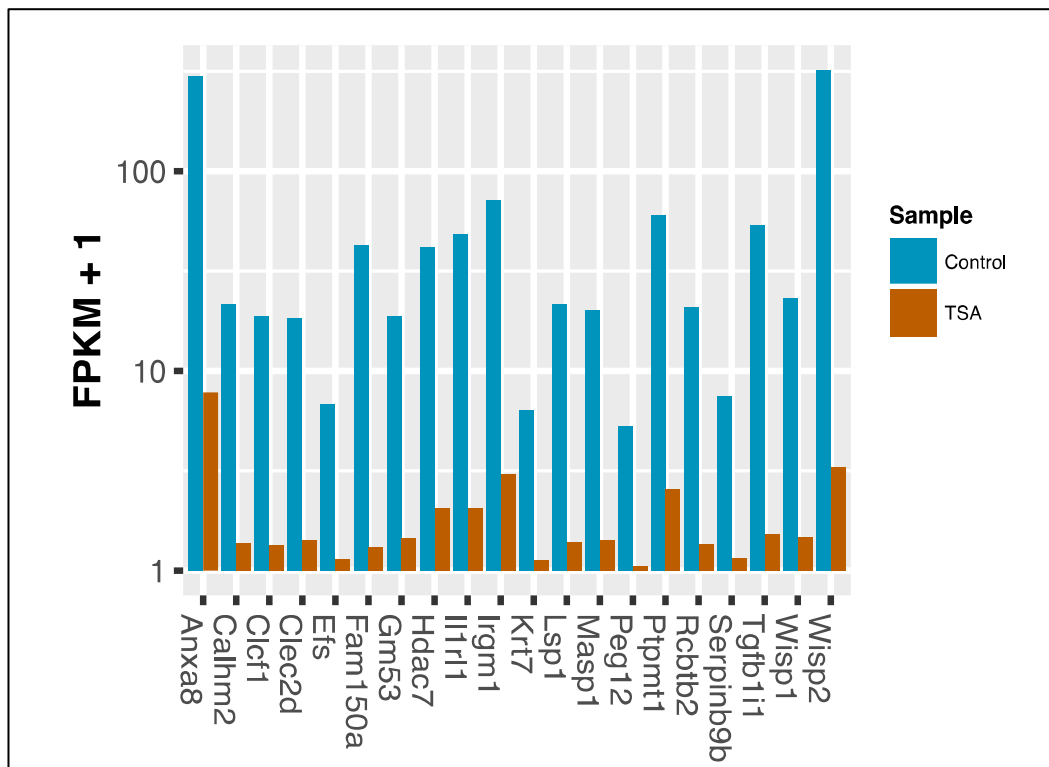


Figure 3.5 – Bar chart displaying the change in FPKM value of the top 20 most negatively differentially expressed genes from Table 3.3 i.e. the top 20 ranked genes which show a decrease in expression upon TSA treatment. FPKM values are displayed as a log scale along the vertical axis. Expression of genes from untreated 14fp cells are highlighted in blue, while the expression of genes from TSA treated cells are represented as orange bars.

| Description                                                     | P-value  | FDR q-value |
|-----------------------------------------------------------------|----------|-------------|
| signal transduction                                             | 5.39E-10 | 6.36E-06    |
| modulation of synaptic transmission                             | 3.10E-09 | 1.83E-05    |
| multicellular organismal process                                | 3.57E-09 | 1.40E-05    |
| behavior                                                        | 7.63E-09 | 2.25E-05    |
| system process                                                  | 2.45E-08 | 5.76E-05    |
| cellular sodium ion homeostasis                                 | 1.32E-07 | 2.59E-04    |
| sodium ion homeostasis                                          | 2.16E-07 | 3.64E-04    |
| regulation of synaptic plasticity                               | 2.28E-07 | 3.37E-04    |
| regulation of synapse structure or activity                     | 3.01E-07 | 3.94E-04    |
| regulation of neurotransmitter levels                           | 3.81E-07 | 4.49E-04    |
| single-organism cellular process                                | 3.97E-07 | 4.25E-04    |
| neurological system process                                     | 4.36E-07 | 4.28E-04    |
| single-organism behavior                                        | 6.12E-07 | 5.55E-04    |
| G-protein coupled receptor signaling pathway                    | 8.32E-07 | 7.01E-04    |
| metal ion transport                                             | 8.33E-07 | 6.55E-04    |
| single-multicellular organism process                           | 8.60E-07 | 6.34E-04    |
| single-organism process                                         | 1.12E-06 | 7.74E-04    |
| signaling                                                       | 1.38E-06 | 9.03E-04    |
| homophilic cell adhesion via plasma membrane adhesion molecules | 2.63E-06 | 1.63E-03    |
| cell differentiation                                            | 2.67E-06 | 1.57E-03    |

**Table 3.4 – Top 20 GOrilla hits for 14fp cell line experiments sorted by p-value, where the input dataset was a list of differentially expressed genes ranked in descending order of log<sub>2</sub>fold FPKM values i.e. the differentially expressed gene with the highest positive log<sub>2</sub>fold FPKM was positioned at the top of the input list.**

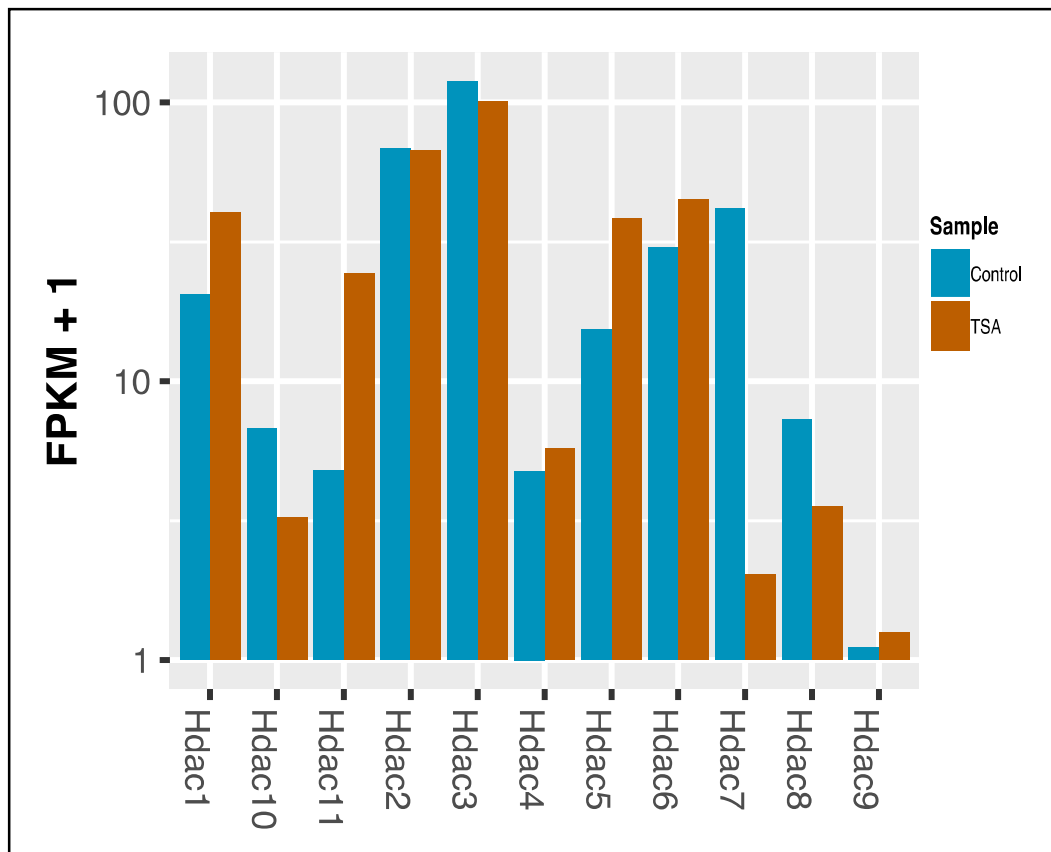


Figure 3.6 - Bar chart displaying the change in FPKM values of HDACs 1 to 11 in the 14fp cell line experiments. FPKM values are displayed as a log scale along the vertical axis. Expression of genes from untreated 14fp cells are highlighted in blue, while the expression of genes from TSA treated cells are represented as orange bars.

### 3.4.5 Gene expression differences between E11.5 proximal and distal limb sections reflect the position of the different signaling centers within the developing limb bud.

The top 20 genes with significantly upregulated expression (ranked according to  $\log_2$ FPKM fold change values) in the distal limb compared to the proximal limb are set out in Table 3.5 and highlighted in Figure 3.7. The top 20 genes with significantly reduced expression (ranked according to  $\log_2$ FPKM fold change values) in the distal limb compared to the proximal limb are provided in in Table 3.6 and highlighted in Figure 3.8. The distal limb section includes the ZPA where *Shh* is expressed and, the majority of the AER and underlying mesenchyme (the progress zone). Therefore, genes involved in both the *Shh* signaling pathway and the AER *Fgf-Shh* feedback loop were expected to have altered expression values between limb bud regions. Indeed, unlike the cell line experiments there is a significant upregulation of *Grem1* (2.5044  $\log_2$ FPKM fold change,  $q=0.000902763$ ) and a much greater increase in *Shh* expression (2.88697  $\log_2$ FPKM fold change,  $q=0.000902763$ ). This accords with *Shh*'s role to positively regulate *Grem1*. In contrast to cell line experiments which show a decrease in *Fgf10* there is a significant increase in *Fgf10* (1.55067  $\log_2$ FPKM fold change,  $q=0.000902763$ ). However, like the cell lines there is a significant increase in *Fgf8* (3.24119  $\log_2$ FPKM fold change,  $q=0.000902763$ ). This is expected as *Fgf8* is known to be expressed in the AER (Lewandoski et al., 2000). The gene with greatest significant negative  $\log_2$ FPKM fold change was *Pax1* (-5.71561  $\log_2$ FPKM fold change,  $q=0.000902763$ ). In previous studies, *Pax1* was shown to be expressed in anterior posterior regions of the E12 limb bud (Timmons et al., 1994). This matches our results and confirms that genes marking the proximal portions of the limb are still active in our dissections. In addition, the genes *Meis1* (-2.23785  $\log_2$ FPKM fold change,  $q=0.000902763$ ) and *Meis2* (-2.49135  $\log_2$ FPKM fold change,  $q=0.000902763$ ) which have previously been shown to be expressed in the posterior limb (Zeller et al., 2009) showed a significant decrease in expression in our distal dissections compared to the proximal dissections. In contrast, the known distally expressed gene *Cyp26b1* (Yashiro et al., 2004) showed a significant increase in expression in distal dissections (1.28247  $\log_2$ FPKM fold change,  $q=0.000902763$ ). This again confirms that the limb bud has been accurately dissected as the expression domains of well characterized limb genes match the literature.

The Gorilla result from this analysis was as expected with terms such as “limb morphogenesis” appearing; however, “male genitalia development”, “neural tube closure” were all listed as top



hits (Table 3.7) due to the presence of significantly expressed genes such as *Shh* and *Hox* genes appearing in several of these categories.

| Gene                 | Locus                    | log <sub>2</sub> (fold change) | q-value     |
|----------------------|--------------------------|--------------------------------|-------------|
| <b>Evx2</b>          | chr2:74493672-74497476   | 4.53072                        | 0.000902763 |
| <b>Gja3</b>          | chr14:57654496-57676782  | 4.40322                        | 0.000902763 |
| <b>Aox3</b>          | chr1:58169979-58257296   | 4.15424                        | 0.000902763 |
| <b>Hoxd13</b>        | chr2:74506366-74509655   | 4.14145                        | 0.000902763 |
| <b>5730457N03Rik</b> | chr6:52258382-52268372   | 3.76511                        | 0.000902763 |
| <b>Evx1</b>          | chr6:52258382-52268372   | 3.43016                        | 0.000902763 |
| <b>Gpx2</b>          | chr12:77893321-77896541  | 3.41893                        | 0.00170037  |
| <b>Vwde</b>          | chr6:13135609-13174965   | 3.38107                        | 0.000902763 |
| <b>Fgf8</b>          | chr19:45811287-45817374  | 3.24119                        | 0.000902763 |
| <b>Sv2b</b>          | chr7:82259780-82476305   | 3.20648                        | 0.000902763 |
| <b>Eomes</b>         | chr9:118387306-118395250 | 3.13133                        | 0.000902763 |
| <b>Hoxa13</b>        | chr6:52208851-52210874   | 2.9647                         | 0.000902763 |
| <b>Scn11a</b>        | chr9:119662882-119734574 | 2.96021                        | 0.000902763 |
| <b>Tgm3</b>          | chr2:129838109-129876135 | 2.91075                        | 0.0031293   |
| <b>Shh</b>           | chr5:28783379-29045749   | 2.88697                        | 0.000902763 |
| <b>Nlrp10</b>        | chr7:116065366-116073672 | 2.73749                        | 0.0075471   |
| <b>Igfbpl1</b>       | chr4:45822378-45839699   | 2.62534                        | 0.000902763 |
| <b>Hoxd12</b>        | chr2:74513086-74515762   | 2.53079                        | 0.000902763 |
| <b>Cbln1</b>         | chr8:89992751-89996491   | 2.5122                         | 0.000902763 |
| <b>Grem1</b>         | chr2:113588831-113598805 | 2.5044                         | 0.000902763 |

**Table 3.5 - The top 20 genes which show the greatest positive change in log<sub>2</sub>fold FPKM values in distal over proximal limb regions. i.e. the top 20 ranked genes which show an increase in expression in distal regions compared to proximal regions. The FDR adjusted q-value for each gene is provided. These values were taken from the Cuffdiff output folder “gene differential expression testing”.**

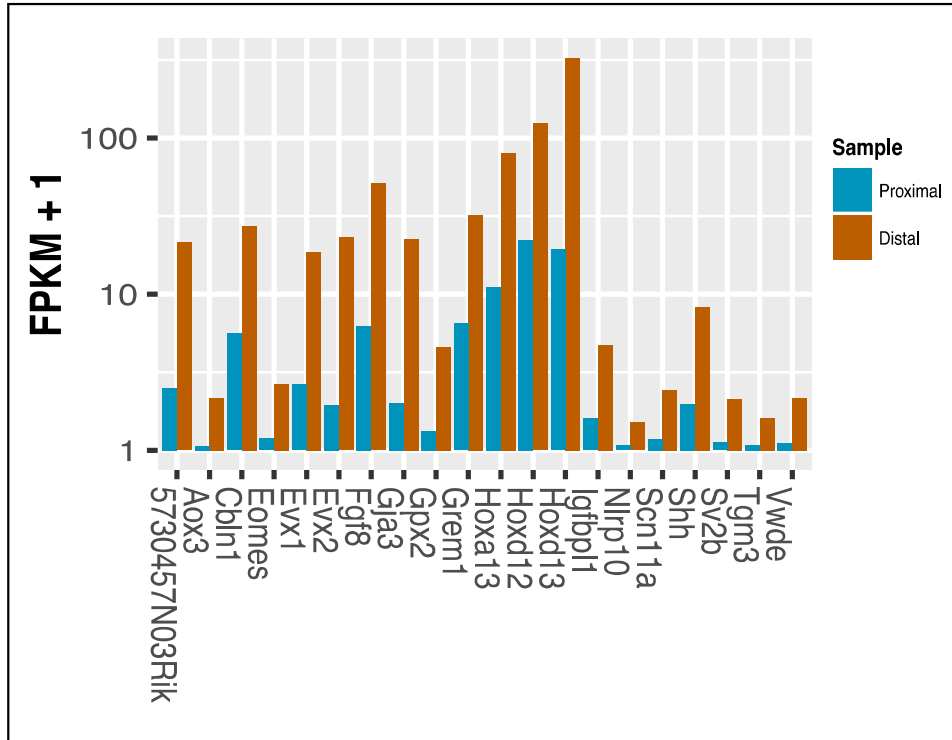


Figure 3.7 – Bar chart displaying the change in FPKM value of the top 20 most positively differentially expressed genes from Table 3.5. i.e. the top 20 ranked genes which show an increase in expression in distal regions compared to proximal regions. FPKM values are displayed as a log scale along the vertical axis. Expression of genes from proximal limb dissections are highlighted in blue, while the expression of genes from distal limb dissections are represented as orange bars.

| Gene           | Locus                    | log <sub>2</sub> (fold change) | q value     |
|----------------|--------------------------|--------------------------------|-------------|
| <b>Pax1</b>    | chr2:147190729-147200784 | -5.71561                       | 0.000902763 |
| <b>Rfx4</b>    | chr10:84218792-84369283  | -5.18115                       | 0.00381167  |
| <b>Dcc</b>     | chr18:71413285-72510723  | -5.01257                       | 0.000902763 |
| <b>Col14a1</b> | chr15:55139304-55352358  | -4.98327                       | 0.000902763 |
| <b>Zcchc5</b>  | chrX:104032420-104035982 | -4.47706                       | 0.000902763 |
| <b>Gpr17</b>   | chr18:32091160-32118273  | -4.38011                       | 0.00381167  |
| <b>Pkdcc</b>   | chr17:83614622-83624409  | -4.01958                       | 0.000902763 |
| <b>Calb2</b>   | chr8:112666437-112692106 | -3.95627                       | 0.000902763 |
| <b>Tnnt2</b>   | chr1:137732962-137748838 | -3.83998                       | 0.000902763 |
| <b>Sox10</b>   | chr15:78985342-78994920  | -3.81487                       | 0.000902763 |
| <b>Barx1</b>   | chr13:48758404-48761876  | -3.7916                        | 0.0153649   |
| <b>Lmod3</b>   | chr6:97188521-97202774   | -3.78148                       | 0.0144063   |
| <b>Foxd3</b>   | chr4:99322989-99325362   | -3.77032                       | 0.00381167  |
| <b>Trim55</b>  | chr3:19544459-19591599   | -3.75539                       | 0.0114033   |
| <b>Zfp804a</b> | chr2:81893814-82100035   | -3.73861                       | 0.000902763 |
| <b>Rxfp2</b>   | chr5:150821249-150884761 | -3.73358                       | 0.0123733   |
| <b>Msc</b>     | chr1:14743428-14746047   | -3.73266                       | 0.000902763 |
| <b>Mybph</b>   | chr1:136090024-136132008 | -3.6845                        | 0.0239337   |
| <b>Col6a5</b>  | chr9:105758399-105862974 | -3.64411                       | 0.000902763 |
| <b>Myh3</b>    | chr11:66891801-66915793  | -3.59298                       | 0.000902763 |

**Table 3.6 - The top 20 genes which show the greatest negative change in log<sub>2</sub>fold FPKM values in distal over proximal limb regions. i.e. the top 20 ranked genes which show a decrease in expression in distal regions compared to proximal regions The FDR adjusted q-value for each gene is provided. These values have been taken from the Cuffdiff output folder “gene differential expression testing”.**

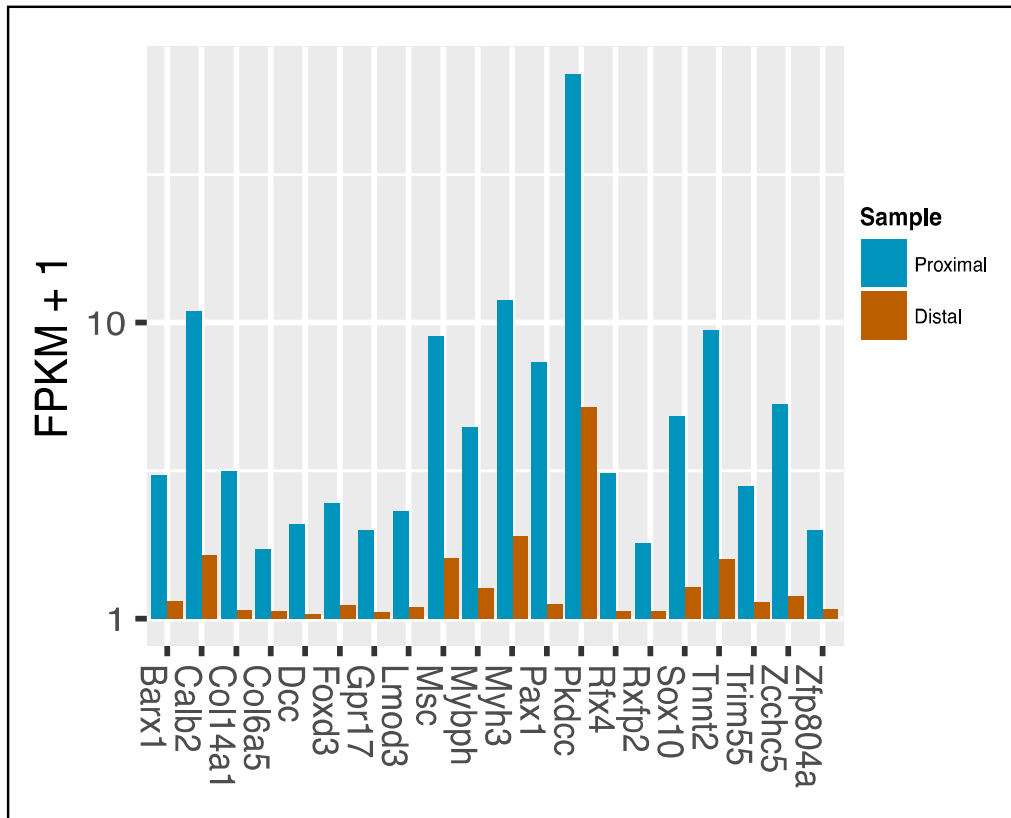


Figure 3.8 – Bar chart displaying the change in FPKM value of the top 20 most negatively differentially expressed genes from Table 3.6. i.e. the top 20 ranked genes which show a decrease in expression in distal regions compared to proximal regions. FPKM values are displayed as a log scale along the vertical axis. Expression of genes from proximal limb dissections are highlighted in blue, while the expression of genes from distal limb dissections are represented as orange bars.

| Description                                            | P-value  | FDR q-value |
|--------------------------------------------------------|----------|-------------|
| organonitrogen compound metabolic process              | 3.46E-07 | 2.58E-03    |
| male genitalia development                             | 4.92E-06 | 1.83E-02    |
| reproductive system development                        | 4.92E-06 | 1.22E-02    |
| nitrogen compound metabolic process                    | 1.08E-05 | 2.01E-02    |
| pattern specification process                          | 2.41E-05 | 3.60E-02    |
| genitalia development                                  | 3.72E-05 | 4.62E-02    |
| determination of bilateral symmetry                    | 5.71E-05 | 6.08E-02    |
| determination of left/right symmetry                   | 5.71E-05 | 5.32E-02    |
| cerebral cortex neuron differentiation                 | 6.29E-05 | 5.21E-02    |
| tube closure                                           | 6.69E-05 | 4.99E-02    |
| nucleobase-containing small molecule metabolic process | 1.23E-04 | 8.34E-02    |
| specification of symmetry                              | 1.24E-04 | 7.73E-02    |
| neural tube closure                                    | 1.43E-04 | 8.19E-02    |
| small molecule metabolic process                       | 1.48E-04 | 7.89E-02    |
| appendage morphogenesis                                | 1.90E-04 | 9.47E-02    |
| limb morphogenesis                                     | 1.90E-04 | 8.88E-02    |
| organophosphate metabolic process                      | 2.22E-04 | 9.75E-02    |
| single-organism metabolic process                      | 2.23E-04 | 9.23E-02    |
| alpha-amino acid metabolic process                     | 2.26E-04 | 8.87E-02    |
| branching involved in prostate gland morphogenesis     | 2.47E-04 | 9.20E-02    |

**Table 3.7 – Top 20 GOrilla hits for limb bud experiments sorted by p-value where the input dataset was a list of differentially expressed genes ranked in descending order of log<sub>2</sub>fold FPKM values i.e. the differentially expressed gene with the highest positive log<sub>2</sub>fold FPKM was positioned at the top of the input list.**

### 3.4.6 A subset of genes are differentially expressed in both cell line and limb dissection experiments.

In total, 1500 genes showed significant differential expression in the limb bud experiments which is much fewer than observed in the cell line experiments where 7,872 genes showed significant differential expression. However, a proportion of these genes overlapped between experiments suggesting that the effect of TSA treatment on 14fp cells affected some of the same signaling processes that are unevenly distributed across the limb bud. In total the cell line experiments shared 147 differentially expressed genes with a significant positive  $\log_2$ fold FPKM with the limb bud experiments. A total of 272 differentially expressed genes with a significant negative  $\log_2$ fold FPKM were found in common (Figure 3.9).

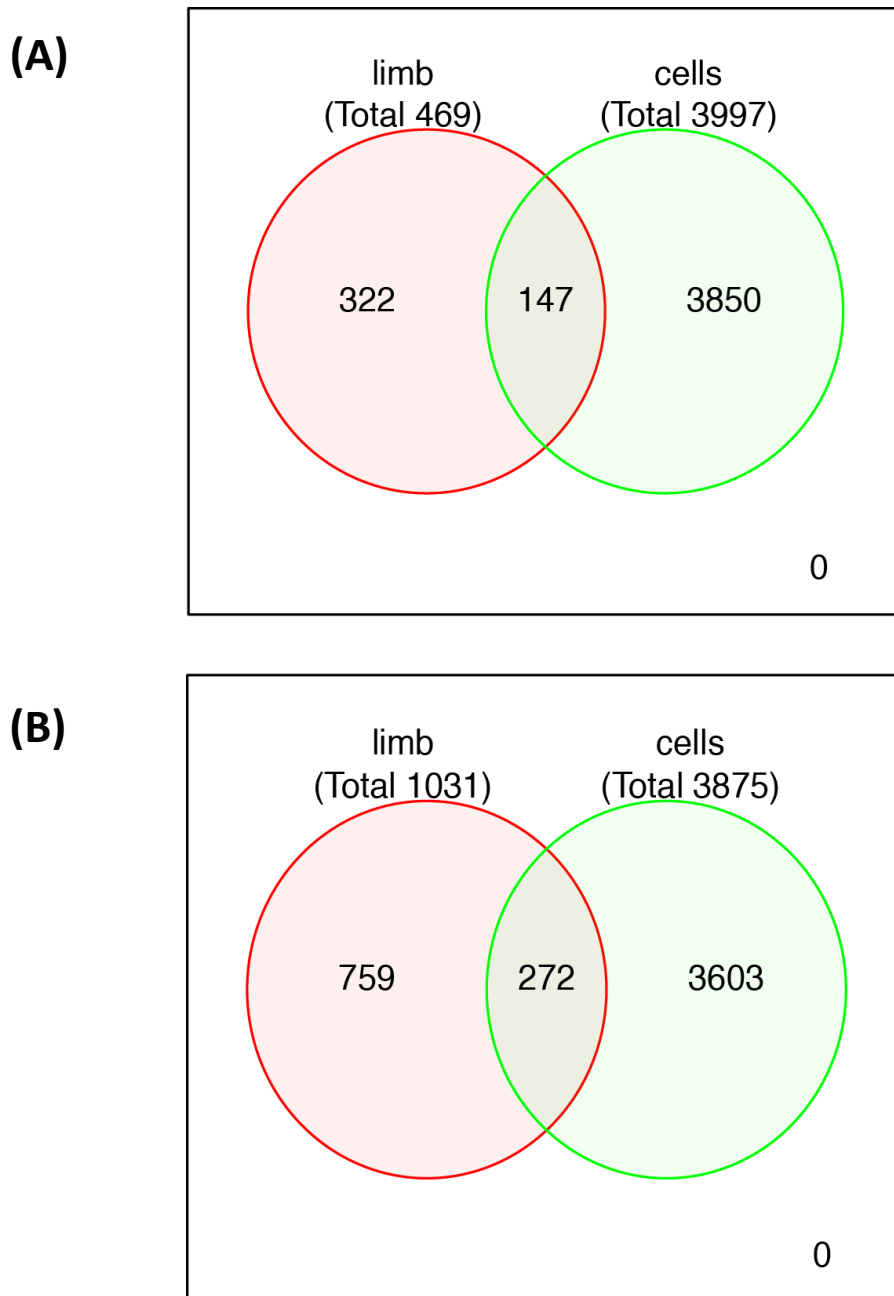
The 147 differentially expressed genes with a significant positive  $\log_2$ fold FPKM between both experiments were then filtered according to the  $\log_2$ fold FPKM value from the limb dissection experiments. The 20 most positively differentially expressed genes (ranked according to FPKM values from limb bud experiments) are presented in Table 3.8 and the top 20 most negatively differentially expressed genes (ranked according to FPKM values from limb bud experiments) are presented in Table 3.9. In Table 3.8, *Fgf8* has the second highest  $\log_2$ fold FPKM value, showing that both TSA treated 14fp cells and the distal limb have an enrichment of this gene. Therefore, one of the main effects of TSA treatment is to induce expression of genes normally expressed within the AER. Of note, *Pdzd2* (Tsui, 2014) - a suspected inhibitor of the *Shh* signaling pathway - is also significantly upregulated.

Of particular interest is how gene expression of the ETS factors (ETS1, GABP $\alpha$ , ETV4 (Pea3) and ETV5) (Chapter 1.3.1) are altered in the 14fp cell lines on TSA treatment and how this compares to the different compartments of the limb. Between the proximal and distal limb sections, there was no statistical difference in *Ets1* and *Gabp $\alpha$*  levels, while there were only small deviations to *Etv4* and *Etv5* ( $< 1.5 \log_2$ fold FPKM). In the 14fp cells, there was only a small deviation of *Ets1* expression levels ( $< 1.5 \log_2$ fold FPKM) and no significant change to *Gabp $\alpha$* . However, there was a substantial drop in *Etv4* levels on TSA addition ( $-1.73994 \log_2$ fold FPKM,  $q=0.000163866$ ) and an increase in *Etv5* ( $2.20157 \log_2$ fold FPKM,  $q=0.000163866$ ). This is interesting as downregulation of *Etv4* on TSA addition correlates with the induction of *Shh*. As mentioned previously, Silvia Peluso (manuscript in preparation) has shown that knock-down of *Etv4* (Chapter 3.1.3) in 14fp cells can also induce *Shh* expression and that *Etv4* binds to HDAC2. Therefore, it may be possible that the activation of *Shh* on

TSA treatment is a secondary effect caused by reduction of *Etv4* levels. Decreased levels of ETV4 binding to the ZRS in combination with increased H3K27ac also caused by TSA treatment would then allow the enhancer to activate *Shh* expression. The reduction of *Etv4* levels before *Shh* expression may explain why it takes 24 hr after TSA treatment before *Shh* is expressed to its maximum instead of this process occurring instantaneously. A time course of *Pea3* levels after TSA treatment may shed more light on the order of the events resulting in *Shh* expression. However, the addition of TSA caused no significant changes to *Hdac2* levels, thereby confirming that changes to the expression levels of this gene are not resulting in *Shh* expression. The increase in expression of *Etv5*, an inhibitor of *Shh*, may counteract the effect of reducing *Etv4* and explain why only such a small quantity of *Shh* is expressed.

In mice, a total of 26 different ETS factors are expressed. We examined the remainder of these in addition to ETS1, GABP $\alpha$ , ETV4 and ETV5 to determine if there is a significant variation in their gene expression between either proximal and distal limb sections or between 14fp cells before and after TSA treatment. In the limb experiments, there was no significant change in any of the other ETS factors except *Ets2* (which showed a log<sub>2</sub>fold FPKM change of less than 1.5). However, in the cell lines experiments, in addition to the changes already mentioned, a number of other ETS factors showed altered expression profiles. *Elf2*, *Elf4*, *Fli1*, *Erf*, *Elk3* and *Etv6* all showed a significant reduction in expression on addition of TSA with log<sub>2</sub>fold FPKM changes greater than -1.5. (i.e. ~ a 3-fold decrease or greater). *Etv3* was the only gene to show a significant increase in expression, however this was with a log<sub>2</sub>fold FPKM of less than 1.5.





**Figure 3.9 – Venn diagrams showing the number of genes with (A) a significant positive  $\log_2$ fold FPKM change for the cell line experiments (red circle) and the E11.5 limb bud experiments (green circle) and (B) a significant negative  $\log_2$ fold FPKM change for the cell line experiments (red circle) and the E11.5 limb bud experiments (green circle). Areas where the circles overlap indicate common genes which show the same pattern of  $\log_2$ fold FPKM changes between experiments. Gene numbers are stated in each circle.**

| Gene    | Locus                     | log <sub>2</sub> (fold change) | q value     |
|---------|---------------------------|--------------------------------|-------------|
| Gja3    | chr14:57654496-57676782   | 4.40322                        | 0.000902763 |
| Fgf8    | chr19:45811287-45817374   | 3.24119                        | 0.000902763 |
| Eomes   | chr9:118387306-118395250  | 3.13133                        | 0.000902763 |
| Igfbpl1 | chr4:45822378-45839699    | 2.62534                        | 0.000902763 |
| Cbln1   | chr8:89992751-89996491    | 2.5122                         | 0.000902763 |
| Pmaip1  | chr18:66618257-66625212   | 2.4633                         | 0.000902763 |
| Smpd3   | chr8:108776447-108861888  | 2.38778                        | 0.000902763 |
| Epha8   | chr4:136485333-136512731  | 2.12861                        | 0.000902763 |
| Srms    | chr2:180940267-180947876  | 2.09892                        | 0.00243874  |
| Chst8   | chr7:35459486-35597730    | 1.95936                        | 0.000902763 |
| Pdzd2   | chr15:12286808-12522311   | 1.95711                        | 0.000902763 |
| Enpp1   | chr10:24361216-24431908   | 1.94541                        | 0.000902763 |
| Slco4a1 | chr2:180195682-180209557  | 1.86905                        | 0.000902763 |
| Sgms2   | chr3:131021903-131047841  | 1.84697                        | 0.000902763 |
| Gldc    | chr19:30172936-30249908   | 1.82455                        | 0.000902763 |
| Lmo7    | chr14:102129144-102333910 | 1.80197                        | 0.000902763 |
| Gbx2    | chr1:91824536-91827751    | 1.79543                        | 0.0129105   |
| Dgkk    | chrX:6356431-6525489      | 1.79234                        | 0.000902763 |
| Hey1    | chr3:8663358-8667038      | 1.7818                         | 0.000902763 |
| Lmo2    | chr2:103798151-103822035  | 1.76301                        | 0.000902763 |

**Table 3.8 - Top 20 genes with the greatest positive change in log<sub>2</sub>fold FPKM values shared between both cell line and limb experiments. In this table genes are ranked according to log<sub>2</sub>fold FPKM values from limb bud experiments. i.e the genes which show an increase in expression on TSA treatment and also show an increase in expression in the distal limb compared to the proximal limb were compared – the top 20 genes which showed significant upregulation in both experiments is provided.**

| Gene                 | Locus                     | log <sub>2</sub> (fold change) | q value     |
|----------------------|---------------------------|--------------------------------|-------------|
| <b>Pkdcc</b>         | chr17:83614622-83624409   | -4.01958                       | 0.000902763 |
| <b>Tnnt2</b>         | chr1:137732962-137748838  | -3.83998                       | 0.000902763 |
| <b>Myl4</b>          | chr11:104411976-104448533 | -3.3695                        | 0.000902763 |
| <b>Lox</b>           | chr18:52676891-52689362   | -3.30191                       | 0.000902763 |
| <b>Osr1</b>          | chr12:9581247-9588306     | -3.24676                       | 0.000902763 |
| <b>Matn4</b>         | chr2:164215128-164240948  | -2.8342                        | 0.000902763 |
| <b>Rapsn</b>         | chr2:90875783-90885886    | -2.79298                       | 0.000902763 |
| <b>Tnc</b>           | chr4:63620818-63708049    | -2.6887                        | 0.000902763 |
| <b>Dkk2</b>          | chr3:131748255-131843268  | -2.67357                       | 0.000902763 |
| <b>Col12a1</b>       | chr9:79446797-79566485    | -2.62952                       | 0.000902763 |
| <b>Sgcd</b>          | chr11:46792284-47192804   | -2.51521                       | 0.000902763 |
| <b>4933436C20Rik</b> | chr8:94849931-94880019    | -2.44038                       | 0.000902763 |
| <b>Ephx1</b>         | chr1:182919686-182947626  | -2.43408                       | 0.0158466   |
| <b>Acta2</b>         | chr19:34315580-34329826   | -2.4279                        | 0.000902763 |
| <b>Baalc</b>         | chr15:38765454-38782810   | -2.36165                       | 0.00170037  |
| <b>Sytl2</b>         | chr7:97497208-97558949    | -2.34326                       | 0.000902763 |
| <b>Hic1</b>          | chr11:74978066-74983757   | -2.32203                       | 0.000902763 |
| <b>Lrrc32</b>        | chr7:105642731-105650340  | -2.32061                       | 0.000902763 |
| <b>Slc24a3</b>       | chr2:145068346-145467675  | -2.22692                       | 0.000902763 |
| <b>Hoxb3</b>         | chr11:96184439-96209244   | -2.22167                       | 0.000902763 |

**Table 3.9 - Top 20 genes with the greatest negative change in log<sub>2</sub>fold FPKM values shared between both cell line and limb experiments. In this table genes are ranked according to log<sub>2</sub>fold FPKM values changes seen in limb bud experiments. i.e the genes which show a decrease in expression on TSA treatment and also show a decrease in expression in the distal limb compared to the proximal limb were compared – the top 20 genes which showed significant downregulation in both experiments is provided.**

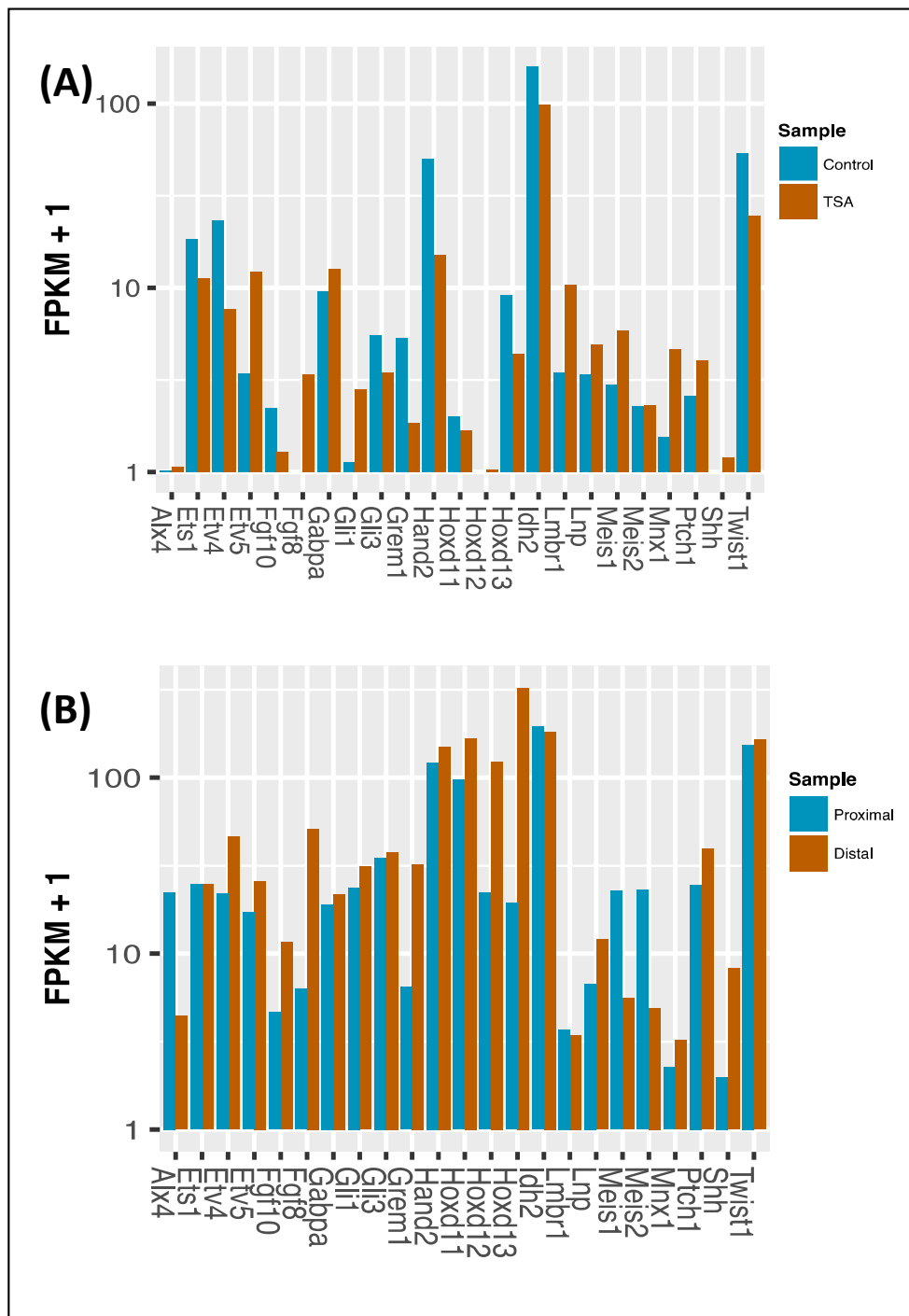


Figure 3.10 - Bar chart displaying the change in FPKM values of some genes known to play a role in limb bud development in (A) cell line experiments and (B) the E11.5 limb bud experiment. FPKM values are displayed as a log scale along the vertical axis. In (A) expression of genes from control cells are highlighted in blue, while the expression of genes from TSA treated cells are represented as orange bars. In (B) expression of genes from proximal limb dissections are highlighted in blue, while the expression of genes from distal limb dissections are represented as orange bars.

### 3.4.7 88fp cells react similarly to 14fp cells when treated with TSA

To determine if induction of *Shh* expression upon TSA treatment was a unique feature of the 14fp cell line or if it could be applied to other limb derived cell lines, we examined the effect of TSA treatment on the 88fp cell line. Firstly increasing concentrations of TSA (1  $\mu$ M, 2  $\mu$ M, 4 $\mu$ M and 8 $\mu$ M) were added to the separate flasks of 88fp cells and left for 24 hr (this time point was chosen as it is the point of maximum *Shh* expression on TSA treatment of 14fp cells). Gene expression was tested by quantitative real time PCR using the housekeeping gene GAPDH for normalisation (Figure 3.11A). Similar to the 14fp cells, increasing the concentration of TSA above 1  $\mu$ M had no effect on *Shh* expression. A 1  $\mu$ M concentration of TSA was subsequently chosen to use in further experiments as it yields the maximum *Shh* expression with less cell death than higher concentrations of TSA. A time course was subsequently conducted where *Shh* expression was monitored at intervals of 3 hr, 6 hr, 16 hr and 24 hr after treatment with 1  $\mu$ M TSA. Like the 14fp cells, there is an incremental increase in *Shh* expression up to 24 hr (Figure 3.11B).

Levels of *Pea3* and *Gabp $\alpha$*  were also analysed after 24 hr using different concentrations of TSA. In accordance with 14fp cells, *Pea3* decreases upon TSA treatment (Figure 3.11(C)) with a significant difference between control and 1  $\mu$ M TSA treated cells (student's t-test,  $p = 0.0152$ ). A significant increase in *Gabp $\alpha$*  expression was observed with 4  $\mu$ M TSA (student's t-test,  $p = 0.0351$ ) but not with any other TSA concentration. However, as only two biological replicates were used in this experiment and there are large error bars on the expression bar chart this result has not yet been confirmed (Figure 3.11(D)).

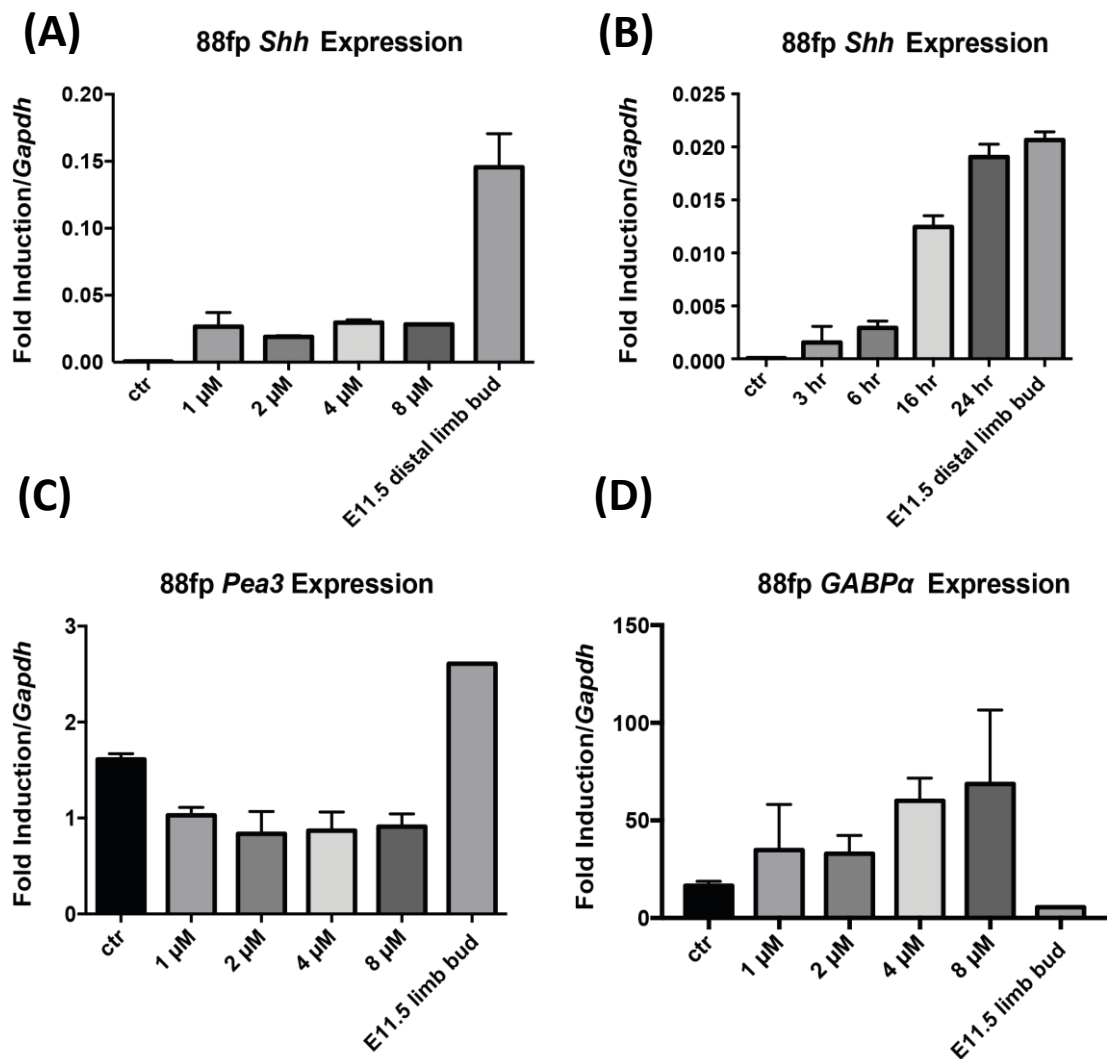


Figure 3.11 – 88fp cells respond to TSA treatment in a similar manner to 14fp cells (Silvia Peluso, manuscript in preparation). (A-D) RT-qPCR was used to test the expression of *Shh*, *Gabpa* and *Pea3* in 88fp cells under different conditions. (A) The effect of increasing concentrations of TSA (1, 2, 4 and 8  $\mu$ M) on *Shh* expression were tested by qPCR. (B) Using qPCR the effect of the addition of 1  $\mu$ M TSA on *Shh* expression was tested at different time points (3hr, 6hr, 16hr and 24hr). (C)(D) The effect of increasing TSA concentrations (1, 2, 4 and 8  $\mu$ M) on levels of *Pea3* (C) and and *Gabp $\alpha$*  (D) were tested by qPCR. *Shh* expression from the distal limb bud was tested as a comparison for all qPCR experiments. *Gapdh* was used for normalisation.

### 3.5 Conclusions

RNA sequencing of control and TSA treated 14fp cells confirms that expression of a large number of genes is affected by TSA treatment. *Shh* is expressed on addition of TSA, but is produced in very small quantities. This result has been confirmed by both qPCR and immunofluorescence (Silvia Peluso, manuscript in preparation). The other affected genes reside within a number of different signaling pathways and are involved in regulating many different processes. However, a number of developmentally important genes expressed within either the distal or proximal limb sections are also expressed in the 14fp cells. It is therefore concluded that the 14fp cell line is a good model system to study the signaling pathways involved in embryonic limb development. In addition, the 88fp cell line responds in a similar fashion to the 14fp cells with maximum SHH expression obtained using 1  $\mu$ M TSA for 24 hr. It was therefore decided to utilize both cell lines for the remainder of the investigation.

# Chapter 4

---

## Chromatin Compaction of the *Shh* locus



## 4.1 Introduction

The mechanism by which enhancers interact with promoters to regulate gene expression remains an ongoing area of research. As enhancers can interact with genes up to 1 Mb away it is widely accepted that under certain circumstances the two sequences can come together by the “looping out” of intervening DNA. The key question is how does an enhancer locate a promoter to facilitate such loops? In some cases, this may occur through a simple diffusion process where an enhancer comes into contact with a promoter by chance and binds due to the attraction of transcription factors and other proteins at both of these sequences. In other cases, a scanning process has been suggested where the enhancer moves along the DNA only stopping when it reaches the promoter (reviewed in (Pennacchio et al., 2013)). In addition, the interaction between proteins bound at boundary elements such as CTCF may also bring enhancers and promoters into close proximity (Chapter 1.5.1).

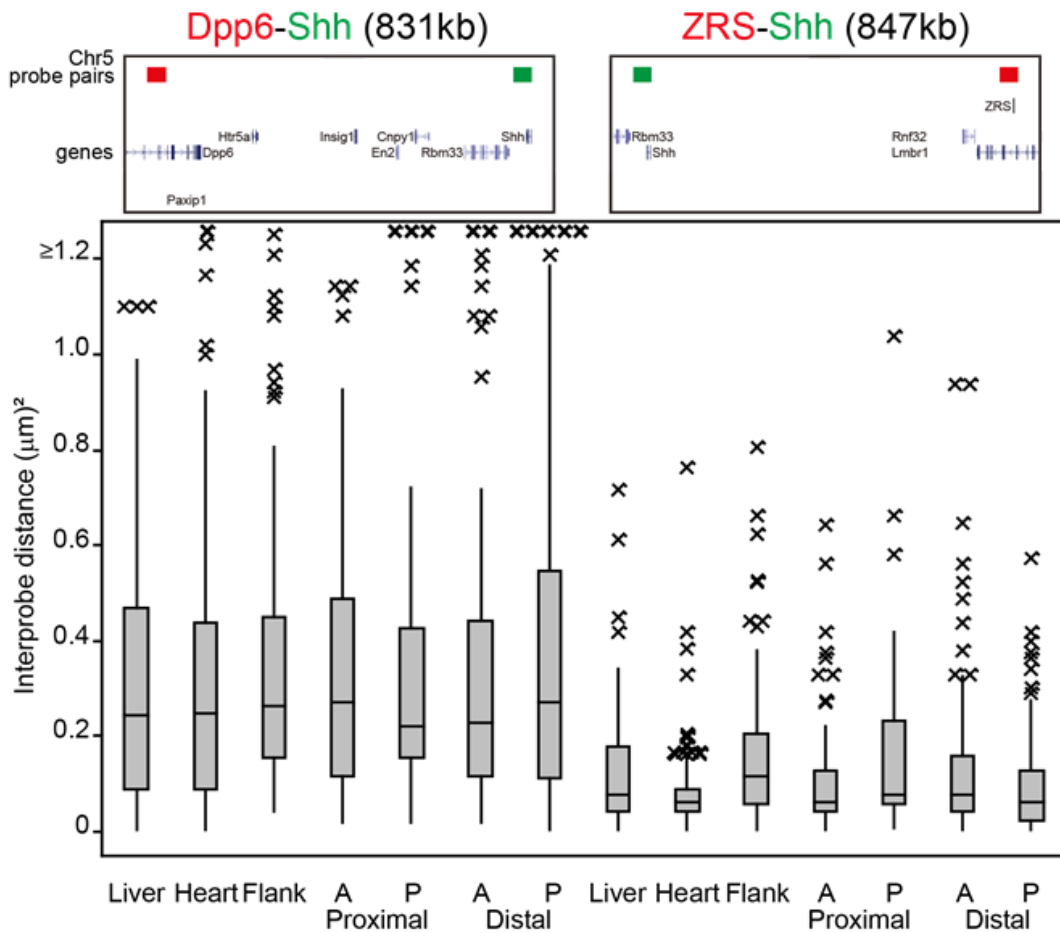
In the limb bud, the Sonic hedgehog regulatory domain is guarded by five main CTCF protein binding sites which are believed to be of central importance in the construction of a Topological Associated Domain (TAD) containing this region. Two of these binding sites are located around the *Shh* gene while the other three are located within or around the *Lmbr1* gene. Evidence of this TAD has been shown by Hi-C experiments where the boundaries of the TAD are located near these CTCF binding sites. The binding of CTCF to these binding sites may result in the proteins interacting with each other to create a loop-like structure. Regions within the *Shh* regulatory domain which are inside the loop are thus restricted from interacting with regions outside the loop and thus the *Shh* enhancers are restricted from acting upon genes located outside of the TAD. Of course this is a simplistic view, as several sub-TADs located within the *Shh* regulatory region may appear as smaller loops within the larger looped structure.

Recent experiments have looked at the effects of removing CTCF binding sites on the structures of TADs (Chapter 1.5.3). In short, such alterations have shifted the boundaries of TADs allowing enhancers to interact with genes with which they had previously been restricted. In humans this can occur naturally resulting in a variety of different diseases (Lupiáñez et al., 2015). It was hypothesised that removing CTCF binding sites situated near the boundaries of the TAD containing the *Shh* regulatory domain may have similar effects. In addition, previous investigations have shown that the *Shh* containing TAD is highly compact

in both *Shh* expressing and non-expressing tissues (Figure 4.1). Perhaps, such compaction is caused by CTCF protein binding and removal of the binding sites may then relax the structure.

Although the *Shh* containing TAD maintains a compact structure in a variety of tissues, there may be alterations within the TAD that are dependent on the expression status of the *Shh* gene. As mentioned previously (Chapter 1.3.8) one investigation (Amano et al., 2009) noted a greater co-localisation between the ZRS and *Shh* in anterior and posterior portions of the limb bud compared to the middle region. In contrast, a recent collaboration between the Hill and Bickmore laboratories showed that the ZRS and *Shh* co-localise more closely only within the ZPA where the gene is expressed (Williamson et al., 2016). This suggests a mechanism whereby the ZRS moves into contact with *Shh* within the ZPA that then activates the gene.

In collaboration with the Bickmore laboratory a series of ES cell lines have been created each with a different deletion of one of the five CTCF binding sites located at the *Shh* containing TAD boundaries (Lettice and Williamson, unpublished). These deletions were generated using the CRISPR/Cas9 system. Using these cell lines, I have performed a series of experiments to analyse the effects of systematically removing different CTCF binding sites on the compaction of the *Shh* locus. In addition, using the *Shh* inducible cell lines (Chapter 3), I have looked at the co-localisation of the ZRS and *Shh* under several different conditions that activate the gene. The aim of these experiments was to determine if the ZRS behaves in a similar fashion in the cell lines as it does in the ZPA of an E11.5 mouse limb bud.



**Figure 4.1 – Boxplots from 3D-FISH experiments showing the interprobe distance between either *Shh* and *Dpp6* or *Shh* and the ZRS in different embryonic mouse tissues. The distance between *Shh* and *Dpp6* in the genome is roughly equidistant to that between *Shh* and the ZRS (the *Shh-Dpp6* region therefore serves as a good control when analysing compaction of the *Shh-ZRS* region). “A” and “P” refer to anterior and posterior regions of the limb bud. All tissue sections are taken from and E11.5 mouse embryo. Outliers are indicated with asterisks. For each tissue section, the interprobe distance between *Shh* and *Dpp6* is greater than that between *Shh* and the ZRS suggesting that the *Shh-Dpp6* region is less compact in all tissues analysed. Figure provided by Iain Williamson.**

## 4.2 Aims

### 4.2.1 General Aims

I am interested in how the TAD containing the *Shh* regulatory domain is affected under different conditions. Firstly, I examined how the conformation of the locus changes between *Shh* expressing and non-expressing 14fp, 88fp and MD cells. Secondly, in an attempt to determine why the *Shh* locus is so compact, I analysed how the structure of the TAD changes when key CTCF binding sites located at the boundaries of the locus are removed.

### 4.2.2 Experimental Aims

The experimental aims of this section are as follows:

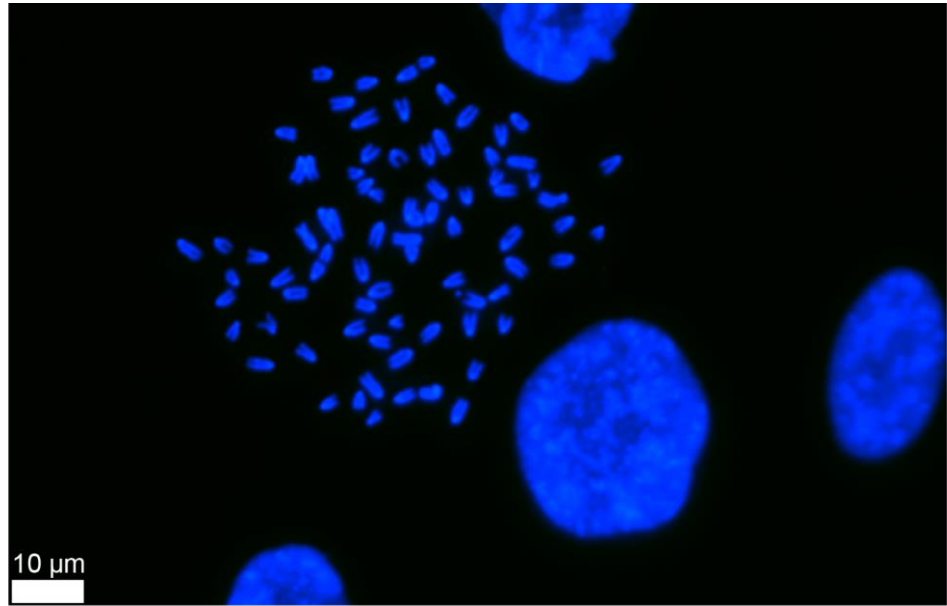
- Perform FISH on 14fp, 88fp and MD cell lines using probes for both *Shh* and the ZRS to examine how the distance between these points changes on the addition of TSA.
- Perform FISH on 14fp cells using probes for both *Shh* and the ZRS to examine how the distance between these points changes when *Pea3* is knocked down.
- Perform FISH on 14fp cells using a probe for *Shh* and a chromosome 5 paint to examine if the *Shh* gene loops out from its chromosome territory on the addition of TSA.
- Use custom designed MYtags (fluorescent oligonucleotides) to compare the compaction of the genomic region from *Shh* to the ZRS and the region from *Shh* to *Dpp6* in both TSA treated and untreated 14fp cells.
- Use custom designed MYtags to compare the compaction of the region from *Shh*-ZRS and the region from *Shh*-*Dpp6* in both wild type and mutant ES cells which each contain a deletion of a specific CTCF binding site.

## 4.3 Results

### 4.3.1 14fp, MD cells and 88fp cells are all hyperdiploid

Before conducting FISH experiments, the ploidy of the 14fp, MD and 88fp cells were examined by counting chromosomes in metaphase spreads (Figure 4.2). Typically, 20 spreads or more were captured and analysed for each cell line. For the 14fp cells, the majority of spreads contained between 77 - 83 chromosomes (74 %) and were considered as tetraploid (tetraploid mice nuclei should contain exactly 80 chromosomes; however, chromosome numbers in this range were considered to be tetraploid due to experimental variation i.e. chromosomes lying on top of each other). The remainder of the spreads contained more than 40 chromosomes but less than 77 and were considered as simply hyperdiploid. Similar results were obtained with the MD and 88fp cell lines, with slightly greater numbers considered hyperdiploid in comparison to tetraploid. In further experiments where a chromosome 5 paint was used (Chapter 4.3.5) in experiments using 14fp cells, four chromosome territories could be distinguished in some of the cells thereby confirming that the cells contained 4 copies of chromosome 5. It was important to study the ploidy of each cell line before conducting FISH experiments so that the number of FISH probes binding within each nucleus could be predicted. For example, as the majority of the 14fp cells were deemed to be tetraploid this suggested that in FISH experiments, four *Shh* probes and four ZRS probes should be observed in each nucleus.

14fp Cells



MD Cells

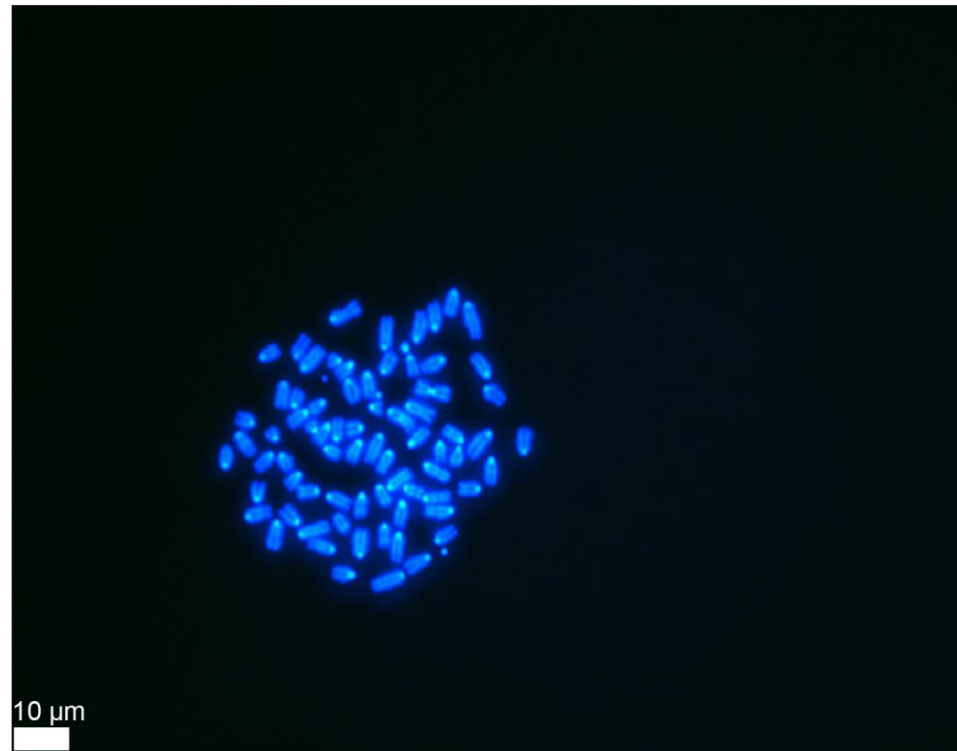


Figure 4.2 – Metaphase spreads of 14fp and MD cell lines showing chromosomes labelled with blue DAPI staining.

### 4.3.2 *Shh*-ZRS co-localisation does not change in 14fp, 88fp and MD cell lines on TSA treatment.

In recent investigations, the ZRS was shown to co-localise more closely to the *Shh* gene within the ZPA of the E11.5 limb bud where SHH is produced (Chapter 4.1). To determine if the ZRS co-localised more closely to *Shh* in the immortalised limb cell lines (Chapter 3), 3D-FISH experiments were performed using probes for both *Shh* and the ZRS (Table 4.1).

Fosmid probes for both *Shh* and the ZRS were provided by the Bickmore laboratory and were used to perform 3D-FISH experiments in 14fp, 88fp and MD cell lines. 14fp and 88fp cells both express *Shh* on TSA treatment, while MD cells do not and therefore provide a suitable control. Each cell line was treated with either 1  $\mu$ M TSA or DMSO (as a control) for 24 hr before fixation. In 14fp and 88fp cell lines experiments, SHH expression was checked by either qPCR or immunofluorescence using a SHH antibody (Figure 4.3). Immunofluorescence of the 14fp cells shows that the majority of the cells contain an increased fluorescent signal when treated with TSA. The antibody used to detect SHH binds to the C-terminal SHH peptide, a processed form of the protein which does not leave the cell (unlike the N-terminal SHH peptide which can be detected in the extracellular matrix) (Chapter 1). Therefore, the assumption is that the TSA treated cells are each producing SHH protein and not taking it up from the extracellular matrix where it has been produced from other cells.

The distance between the *Shh* and ZRS probes was measured for each condition in each cell line. Over 100 measurements were recorded for each condition in each experiment (Figures 4.2 and 4.3). Mann-Whitney U tests confirmed that there was no statistical difference in the distance between the ZRS and *Shh* after TSA treatment in all three cell lines (p-values > 0.05). In a study analysing the chromatin positioning of HoxD enhancers, FISH probes < 200nm apart were considered to be co-localised (Williamson et al., 2012). FISH measurements were therefore binned into 200 nm groups and Fisher exact tests performed to determine if there was a significant change in the co-localised signals upon TSA treatment. For the three cell lines, p-values > 0.05 were obtained confirming that TSA treatment does not cause a detectable change in the co-localisation of *Shh* and the ZRS.

It should be noted that the expression of *Shh* on TSA treatment in the 88fp cells is much lower than the *Shh* expression observed in the limb bud controls (Figure 4.4). *Shh* expression in the 14fp cells on TSA treatment is also much lower than limb bud controls (Figures 4.7). Perhaps

the reason why no significant differences were observed in the co-localisation of *Shh* and the ZRS in the cell lines was because *Shh* expression was too low.

| <b>Fosmid</b> | <b>Start</b> | <b>End</b> |
|---------------|--------------|------------|
| <i>Shh</i>    | 28754458     | 28795879   |
| SBE4          | 29107140     | 29147593   |
| ZRS           | 29611727     | 29653695   |

**Table 4.1 – Location on chromosome 5 of fosmids used in FISH experiments.**



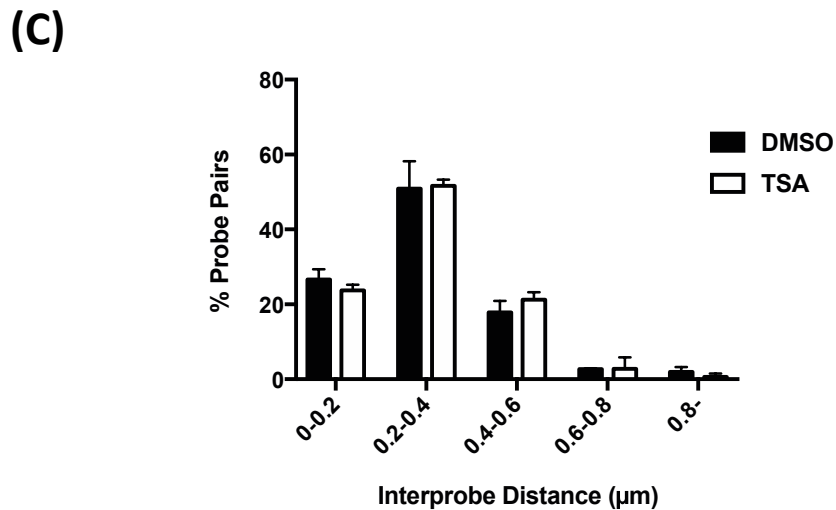
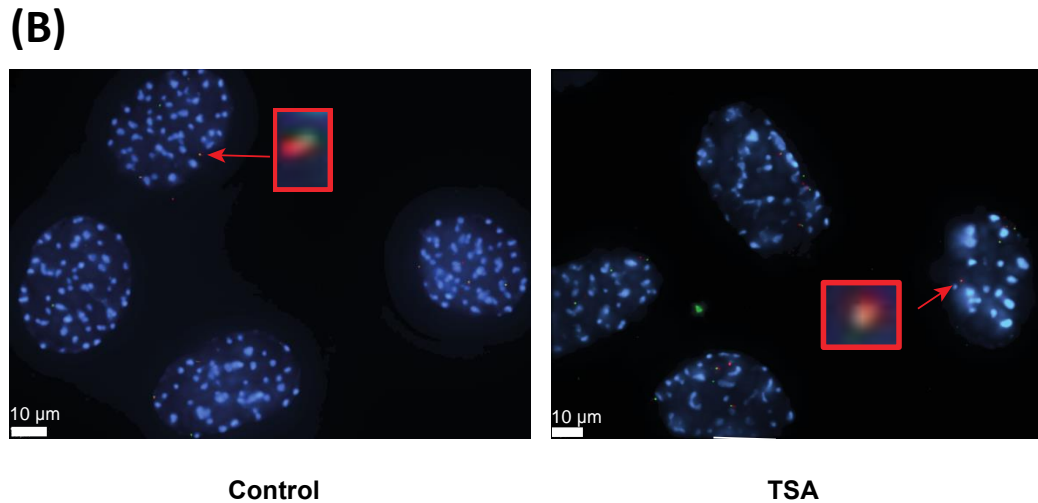
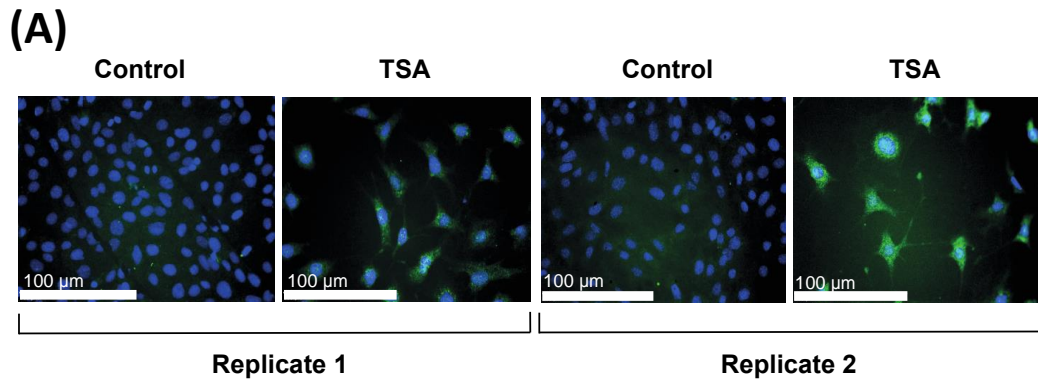


Figure 4.3 – TSA treatment did not affect co-localisation of *Shh* and the ZRS in 14fp cells. 3D-FISH was conducted in duplicate using probes for both *Shh* and the ZRS. (A) SHH expression was checked by immunofluorescence for both replicates using a SHH primary antibody yielding a green signal when viewed under a fluorescent microscope. There was an observable increase in green signal in TSA treated cells compared to control cells (Nb The contrast of these images has been altered so the signal can be viewed more clearly in

this document. However, the same manipulations were conducted on images for both TSA treated and control 14fp cells and therefore the difference in green signal between both conditions should remain unchanged). (B) Images from 3D-FISH experiments from one of the replicates for both control and TSA treated 14fp cells. Nuclei are stained with DAPI and appear blue. Both images show a zoomed in image of the red (*Shh*) and green (ZRS) FISH probes. (C) The measurements from 3D-FISH data for both control and TSA treated 14fp cells were grouped into 200 nm “bins” and plotted as a bar chart. Signals were considered co-localised if they fell into the first bin and were less than or equal to 200 nm apart. There was no significant difference between co-localised signals in the 14fp cells upon TSA treatment (as tested by Fisher exact tests) and no significant difference in compaction (as tested by Mann-Whitney U tests – see Figure 4.4(C)). Over 100 measurements were recorded for control and TSA treated cells for both replicates. Error bars represent the standard deviation between two replicates.

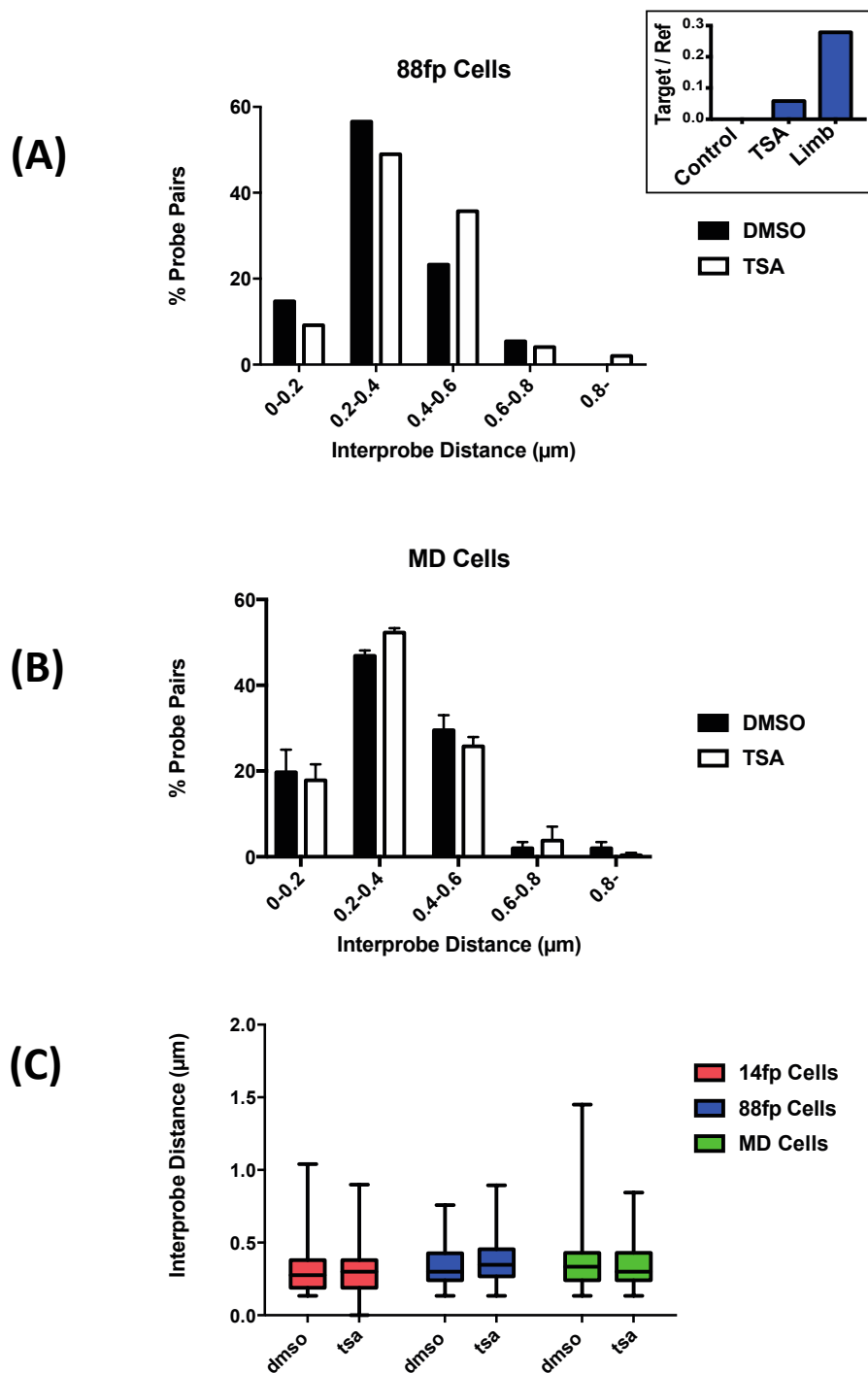


Figure 4.4 - TSA treatment did not affect co-localisation of *Shh* and the ZRS in 88fp and MD cell lines. The measurements from 3D-FISH data for both control and TSA treated 88fp and MD cells were grouped into 200 nm “bins” and plotted as bar charts. Signals were considered co-localised if they fell into the first bin and were less than or equal to 200 nm apart. (A)(B) There was no significant difference between co-localised signals in the 88fp and MD cells upon TSA treatment (as tested by Fisher exact tests). The expression of *Shh* in 88fp TSA treated cells was tested by qPCR (A, upper). In these graphs error bars represent the standard deviation between two replicates. (C) The compaction of 14fp, 88fp

and MD cells was examined by plotting the distributions of *Shh*-ZRS distances on boxplots and testing the statistical difference between distributions using the Mann-Whitney U test. In this plot the measurements for replicate samples are combined with each boxplot representing the total number of measurements. No significant change in compaction were observed between the cell lines or between TSA treated and control cells. Over 100 measurements were recorded for control and TSA treated cells for both replicates.

### 4.3.3 The localisation of probes at the Nuclear Periphery is unaffected by TSA treatment

Several investigations have linked the localisation of genes at the nuclear periphery to a repression of gene expression (Chapter 1) (Zullo et al., 2012). It was hypothesised that TSA treatment may alter the number of *Shh* and ZRS probes located at the nuclear periphery, perhaps moving them towards the nuclear interior which is associated with an increase in gene expression (Therizols et al., 2014). To investigate this hypothesis, FISH data obtained from treating 14fp cells with TSA (as well as the control cells) was used to look at the location of both *Shh* and ZRS probes within the nucleus. A nuclear erosion script was used which divides the nuclear section visualised into 5 regions of equal area (Figure 4.5). Region 1 incorporates the nuclear periphery while region 5 incorporates the nuclear centre. To determine if the difference in the number of probes located in region 1 was significantly different between control and TSA treated cells the Fisher exact test was used. There was no significant difference between the number of probes located at the nuclear periphery (p-value > 0.05) between samples. This suggests that TSA treatment has no effect on the positioning of the *Shh* locus at the nuclear periphery.

However, one of the caveats of using this nuclear erosion script was that the position of probes within the nucleus could only be examined along only two dimensions (i.e. probe position was examined along the x and y axes but not along the z axes). Therefore, in theory some probes which appear in region 5 (Figure 4.5) and are considered to be located within the nuclear centre may actually be nearer to the nuclear periphery if the z axis was considered.

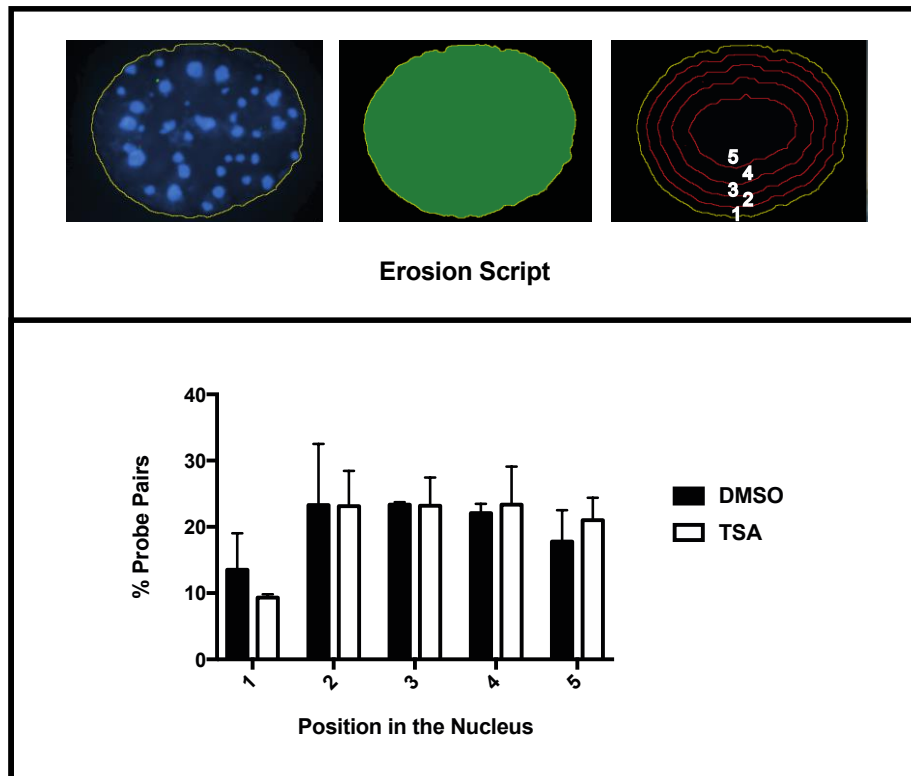


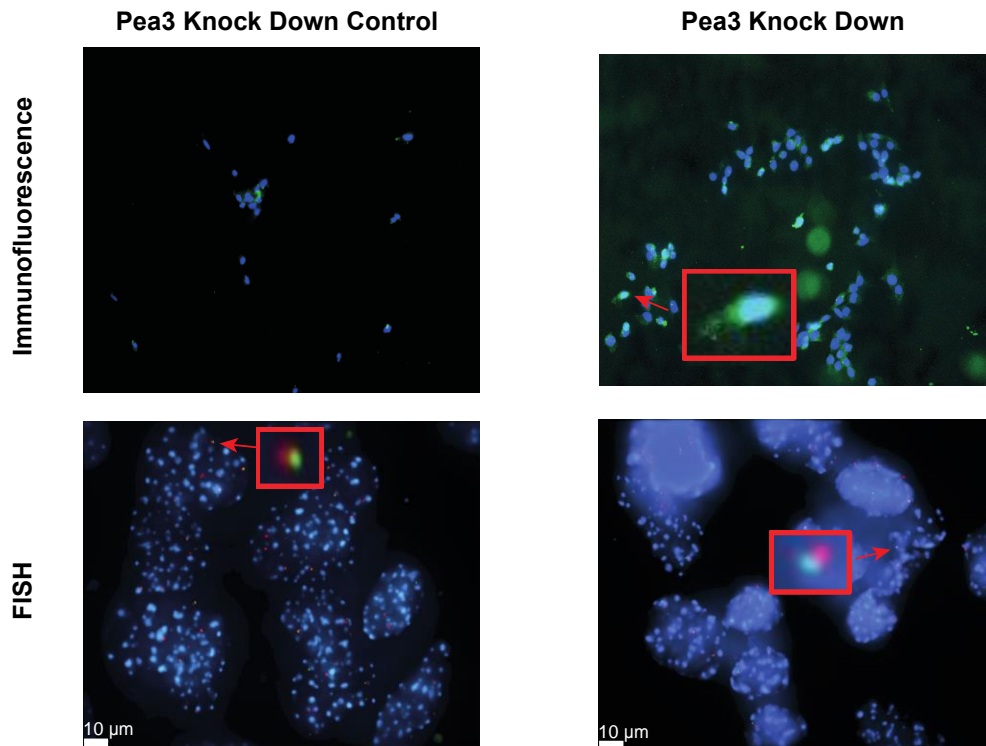
Figure 4.5 – *Shh* and ZRS probes do not move away from the nuclear periphery upon TSA treatment. A nuclear erosion script was used to analyse the position of *Shh* and ZRS probes from the 14fp cell FISH experiments within the nucleus. This divided the nucleus into 5 regions each of equal area. The first region included the nuclear periphery while the fifth was situated at the nuclear centre (regions labelled 1-5). The number of probes located in each region was determined. There was no significant difference in the number of probes located at the nuclear periphery in 14fp control and TSA treated cells as tested by using the Fisher exact test.

#### 4.3.4 Knock down of PEA3 induces SHH expression, but has no effect on *Shh*-ZRS co-localisation

In order to determine if additional experimental methods of inducing *Shh* expression - other than the addition of TSA - had an effect on the co-localisation of the ZRS and *Shh* 3D-FISH was performed on both control 14fp cells and 14fp cells with knocked down levels of *Pea3*. This was achieved by using *Pea3* ShRNAi mediated through use of a lentivirus (performed by Silvia Peluso). PEA3 protein (also known as ETV4) was previously shown to bind to the ZRS and act as a negative regulator of *Shh* expression. In combination with other ETS factors, this restricts *Shh* expression to the ZPA of the E11.5 mouse embryo. In the 14fp cells, Silvia Peluso (manuscript in preparation) showed that knock down of *Pea3* induces *Shh* expression (Chapter 3.1.3).

In this experiment, the effect of *Pea3* knock down on *Shh* expression was analysed by performing immunofluorescence on both control 14fp cells and cells with knocked down levels of PEA3 using a SHH antibody (performed by Silvia Peluso). Cells with reduced levels of PEA3 showed an increase in SHH expression compared to the controls (Figure 4.6(A)) as indicated by an increase in the fluorescent signal confirming that reducing PEA3 levels increases SHH expression. However, there was no significant difference in the distance between *Shh* and the ZRS when comparing control cells and cells with reduced levels of PEA3 (p-value > 0.05). In addition, there was no significant difference in the proportion of co-localising signals (p-value > 0.05) between both conditions. Therefore, it was concluded that although knock-down of PEA3 increases SHH expression, it has no effect on the distance between the *Shh* gene and the ZRS.

(A)



(B)

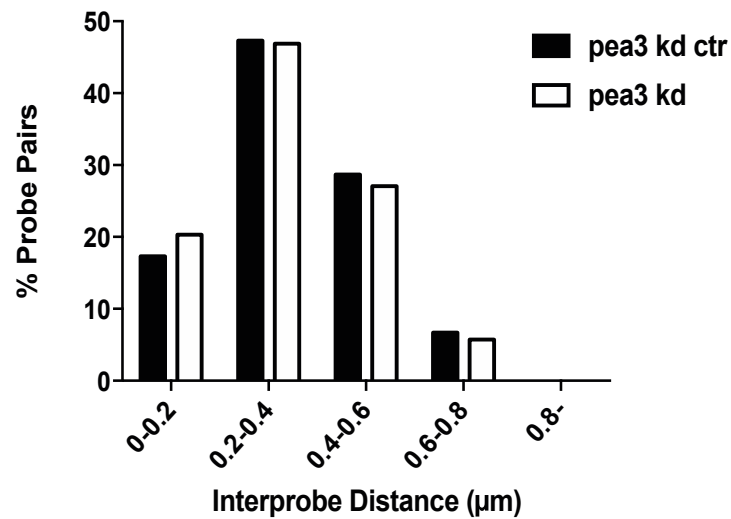


Figure 4.6 – Reduced levels of PEA3 increase *Shh* expression but have no effect on *Shh*-ZRS co-localisation. 3D-FISH was performed on control 14fp cells and 14fp cells with knocked down *Pea3*. (A, upper) Immunofluorescence was used to check that knock down of *Pea3* resulted in an increase in SHH expression. This was tested using a SHH primary antibody.



An increase in green signal was detectable in *Pea3* knock down 14fp cells suggesting an increase in *Shh* expression. (Nb The contrast of these images has been altered so the signal can be viewed more clearly in this document. However, the same manipulations were conducted on images for both TSA treated and control 14fp cells and therefore the difference in green signal between both conditions should remain unchanged). (A, lower) Images from 3D-FISH experiments from both the *Pea3* knock down control and *Pea3* knock down 14fp cells. Nuclei are stained with DAPI and appear blue. Both images show a zoomed in image of the red (*Shh*) and green (ZRS) FISH probes. (B) The measurements from 3D-FISH data for both *Pea3* knock down control and *Pea3* knock down 14fp cells were grouped into 200 nm “bins” and plotted as bar charts. Signals were considered co-localised if they fell into the first bin and were less than or equal to 200 nm apart. There was no significant difference between co-localised signals in the *Pea3* knock down control and *Pea3* knock down 14fp cells (as tested by Fisher exact tests). Over 100 measurements were recorded for each condition.

#### 4.3.5 TSA treatment causes *Shh* to loop out of its Chromosome Territory

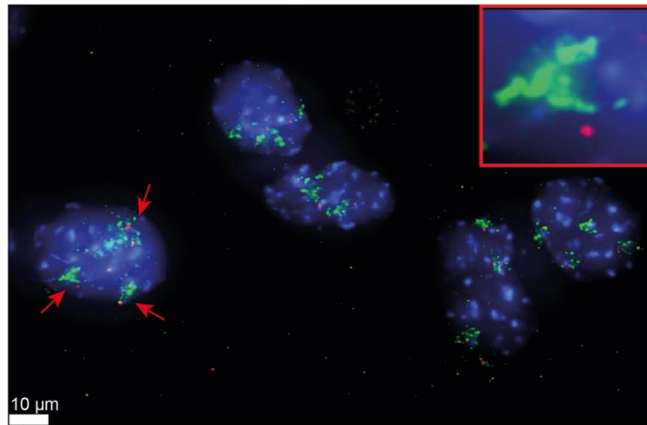
3D-FISH experiments on the different cell lines showed that TSA treatment did not have an effect on co-localisation of *Shh* and the ZRS. One reason for this may be that the cells were expressing *Shh* at a very low level and, differences in co-localisation were too minor to be detected by FISH. Alternatively, the *Shh* gene may be activated through additional mechanisms on addition of TSA.

It was hypothesised that if TSA had no effect on *Shh*-ZRS co-localisation, it may activate the *Shh* gene by causing it to loop out of its chromosome territory into the contact of transcription factories (Chambeyron and Bickmore, 2004). This was tested by performing FISH on 14fp cells using a fosmid probe for the *Shh* gene and a chromosome 5 paint (*Shh* is located within chromosome 5) on both control and TSA treated samples. Chromosome paints consist of different nucleic acid probes containing fluorophores which can hybridize to specific DNA sequences. In this case, the probes bind to chromosome 5 which can then be visualised using a fluorescence microscope.

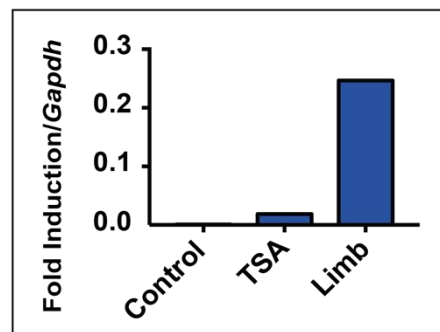
The experiment was repeated in duplicate with > 50 chromosome territories examined for each condition and replicate (Figure 4.7). *Shh* expression of TSA treated cells was checked by qPCR. Fisher exact tests were performed on the combined control and TSA distributions. There was a significant difference between control and TSA treated samples (p-value = 0.0364), thereby suggesting that TSA treatment caused an increase in the number of *Shh* probes located outside the chromosome territories rather than inside the territories. This increase in localisation outside of the chromosome territory may result in expression of the *Shh* gene.

To confirm that the looping out of the *Shh* gene from its chromosome territory is a result of *Shh* expression and not simply an effect of TSA, similar experiments are being repeated with MD cells as a control. If TSA has no effect on *Shh* looping in MD cells the looping out of the *Shh* gene in 14fp cells can be attributed to the expression of *Shh*.

(A)



(B)



(C)

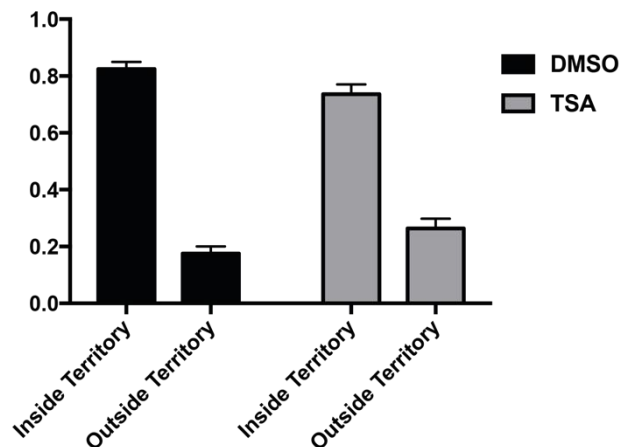


Figure 4.7 – *Shh* loops out of its chromosome territory upon TSA treatment. (A)(B) FISH experiments were conducted using a chromosome 5 paint (which fluoresces green under the fluorescent microscope) and a fosmid probe for *Shh* (red). (A) In the 14fp cell line distinct chromosome 5 territories could be visualised (red arrows) with most cells containing between three and four of these territories. (A, inset) A zoomed-in view of one of the chromosome 5 territories with the *Shh* gene positioned outside it. (B) SHH expression was tested for in this experiment by qPCR. (C) Bar graphs showing the proportion of *Shh* probe located either inside or outside the chromosome 5 territories in control and TSA treated 14fp cells. There was a significant increase in the proportion of probes located outside the chromosome 5 territories upon TSA treatment. This was repeated in duplicate with more than 50 chromosome territories analysed per sample.

#### 4.3.6 Analysis of Chromatin Compaction at the *Shh* locus in 14fp cells

The data presented in Chapter 4.3.1 – 4.3.5 shows that in the immortalised limb cell lines (14fp, 88fp and MD cells) there is no change in *Shh*-ZRS co-localisation upon *Shh* expression. However, like E11.5 limb cells (Figure 4.1), the *Shh*-ZRS region in the cell lines is highly compact with roughly 20 % of probes co-localised (interprobe distance  $\leq 200\text{nm}$ ) even when *Shh* is not expressed. To look at compaction of the *Shh* locus within the 14fp cell line in more detail, MYtag technology was utilised where individual nucleic acids of roughly 40-60 bp are directly labelled with a fluorescent probe. Together MYtag libraries (consisting of thousands of fluorescently labelled nucleic acid probes) can be used in FISH experiments (instead of fosmid probes) to tile areas of the genome and can be visualised under a fluorescent microscope. The purpose of this was to examine the structure of the locus containing *Shh* and the ZRS in comparison to a neighbouring region of similar length (i.e. the region between *Shh* and *Dpp6*). It was hoped that using MYtags would help to visualise local architectural/structural features within the *Shh*-ZRS TAD; in particular, those not visualised by 3D-FISH.

In these experiments, eight regions of 20 kb were tiled using MYtags between the *Shh* gene and the ZRS (Chr5: 28,783,380 - 29,704,930) with a 100 kb spacing between each region (Figure 4.11(B)). As a control, another region of equal size ranging from *Shh* to *Dpp6* (Chr5: 27,861,830 - 28,783,380) was tiled using MYtags in a similar fashion. In preliminary experiments (with ES cells) the *Shh*-ZRS MYtags contained an ATTO-488 fluorophore, while the *Shh*-*Dpp6* MYtags contained an ATTO-647N fluorophore. However, these fluoresced with only a faint signal under the microscope and it was difficult to discriminate between the 20 kb MYtag spots. These libraries were subsequently re-ordered with the *Shh*-ZRS fluorophore replaced by ATTO-594 and the *Shh*-*Dpp6* fluorophore replaced by 6-FAM. The subsequent images obtained were much brighter. All MYtag experiments were conducted using a 2D-FISH protocol as described in the Methods Chapter.

For both control and TSA treated cells, the number of spots observed for the *Shh*-ZRS region (appear as red spots) and the *Shh*-*Dpp6* region (appear as green spots) were counted (Figure 4.8). In total, spots for over 100 different sets of *Shh*-ZRS and *Shh*-*Dpp6* loci for both control and treated cells were counted. On examining spot numbers, it became obvious that for any given loci one large spot may consist of several smaller spots which cannot be individually identified. In this way, counting the number of spots may not be completely reflective of the

relative compaction of a locus. Therefore, in addition to counting the number of spots the total area covered by the spots, per locus was also measured.

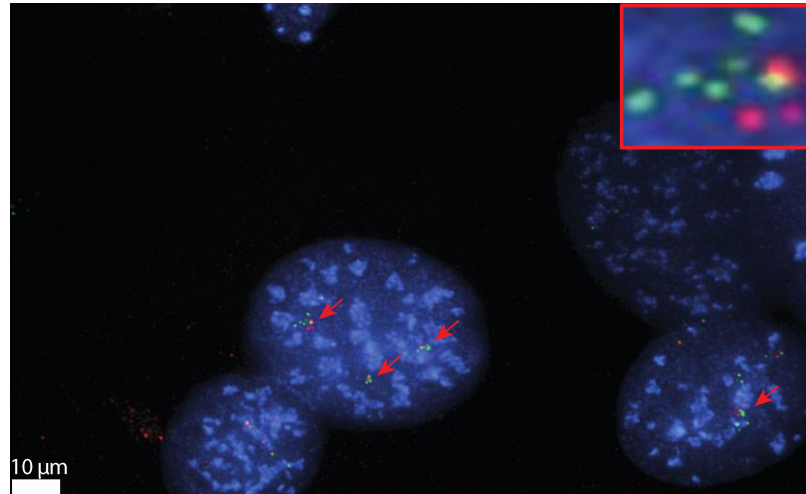
In both TSA treated and control 14fp cells, there was a significant difference in the number of spots counted for the *Shh-ZRS* loci compared to the *Shh-Dpp6* loci (p-values < 0.0001). The *Shh-Dpp6* loci showed consistently more spots than the *Shh-ZRS* loci suggesting that the latter are much more compact (Figure 4.9). The boxplots for *Shh-ZRS* and *Shh-Dpp6* loci for both treated and untreated 14fp cells are almost identical to those obtained in the ES cell experiments (Chapter 4.3.8). This suggests that the extent to which both the *Shh-ZRS* and *Shh-Dpp6* loci are compacted is similar across different cell lines. However, there was no significant difference (p-value > 0.05) in the number of spots counted for the *Shh-ZRS* loci between TSA treated and control 14fp cells, thereby suggesting that TSA treatment does not alter the compaction of this locus.

Again, when spot area was analysed there was a significant difference between the area of red (*Shh-ZRS*) and green (*Shh-Dpp6*) spots in both control and TSA treated cells (p-values < 0.0001). The area of green spots was greater in both cases suggesting that the *Shh-Dpp6* locus was less compact. In contrast, there was no significant difference between the area of red spots between control and TSA treated cells (p-value > 0.05). Therefore, in this case analysis of both spot area and spot number suggested that TSA treatment did not alter the compaction of *Shh-ZRS* region (Figure 4.10).

## 14fp Cells

(A)

DMSO



(B)

TSA

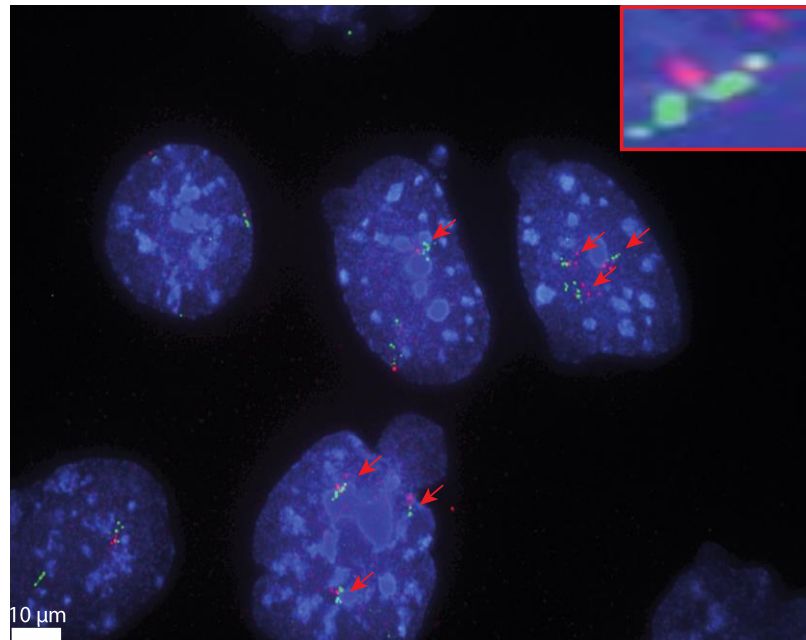


Figure 4.8 - Images showing the MYtag spots in (A) Control and (B) TSA treated 14fp cells. Inset in each image is an enhanced view of MYtags binding at one locus. Red arrows point to the position of Mytags in the cells.

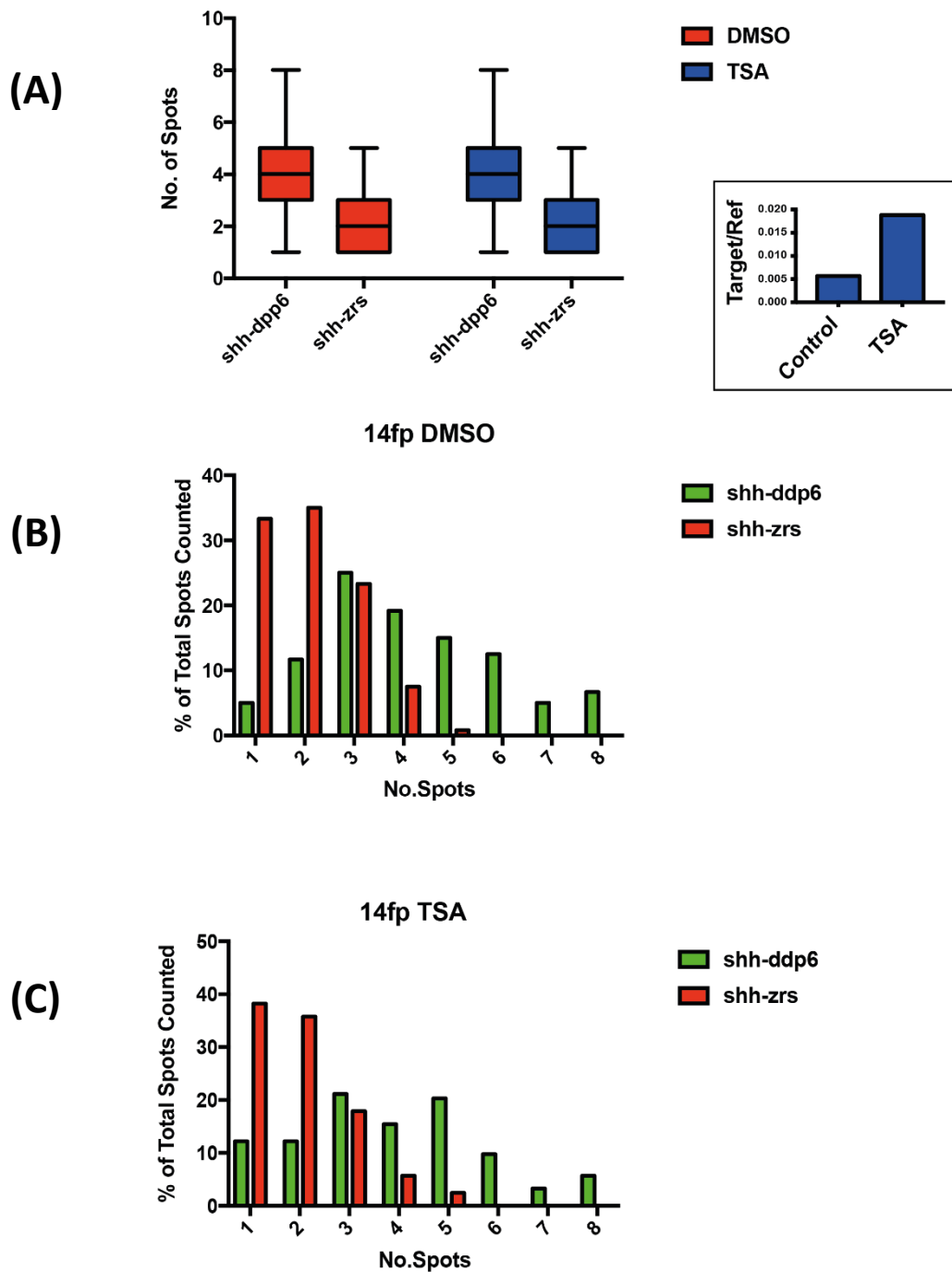


Figure 4.9 – (A) Boxplots showing the distribution of the number of red and green spots counted for control and TSA treated 14fp cells ( $n > 100$ ). (B)(C) The number of red and green spots counted for control (B) and TSA treated (C) 14fp cells in MYtag experiments plotted as a bar charts. The number of spots counted per locus (horizontal axis) is plotted against the percentage of the total spots counted (vertical axis). Red bars represent red spots counted for the MYtags binding to the region between *Shh* and *Lmbr1*, while green bars represent green spots counted for the MYtags binding to the region between *Shh* and *Dpp6*.

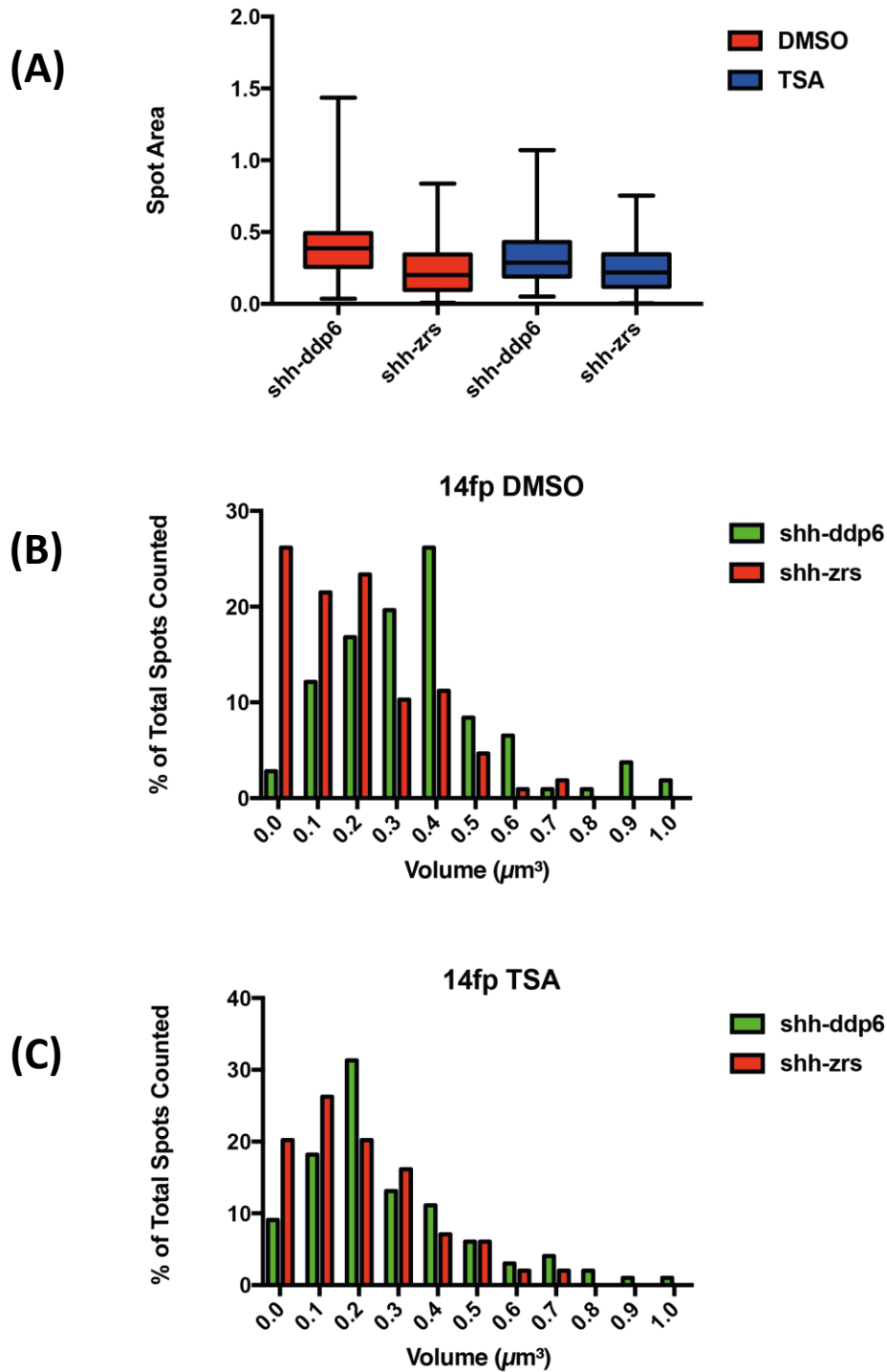


Figure 4.10 – (A) Boxplots showing the distribution of red and green spot areas counted for control and TSA treated 14fp cells ( $n > 100$ ) (B)(C) The area of red and green spots counted for control (B) and TSA treated (C) 14fp cells in MYtag experiments plotted as a bar charts. The area of spots counted per locus (horizontal axis) is plotted against the percentage of the total spots counted (vertical axis). Red bars represent red spots counted for the MYtags binding to the region between *Shh* and *Lmbr1*, while green bars represent green spots counted for the MYtags binding to the region between *Shh* and *Dpp6*.



### 4.3.7 Analysis of Chromatin Compaction at the *Shh* locus in ES cells

Chapter 4.3.6 showed that the *Shh*-ZRS region is much more compact than the *Shh-Dpp6* region which is of similar size. However, what caused this high degree of compaction was unknown. It was hypothesised that the binding of the protein CTCF to specific binding sites at the boundaries of the TAD containing *Shh* may help to create such a compact structure. To test this hypothesis, FISH experiments were conducted (similar to those detailed in Chapter 4.3.6) with MYtags on ES cells containing specific deletions of CTCF binding sites at the boundaries of the *Shh* containing TAD.

The ES cell mutants were provided by Williamson and Lettice (manuscript in preparation). Functional CTCF binding sites were identified by Chip-seq as part of the ENCODE project, with the 5 regions used within the *Shh* regulatory region appearing as peaks in multiple cell lines. The ‘Optimized CRISPR Design tool’ (<http://crispr.mit.edu/>) was used to design guide RNAs which flanked either side of the peaks (regions ranging from 750 bp to 1.4 kb). The resultant oligos were cloned into px458, which carries Cas9 from *S. pyogenes* in addition to 2A-EGFP, and a cloning backbone for the sgRNA (Addgene). E14Tg2A ES cells were electroporated with both the vectors for one CTCF site and allowed to grow for 48 hours before being FACS sorted for the GFP contained within the vector. Fluorescent cells were then plated at low density until clones became apparent (approximately 10 days). Homozygous clones were identified by PCR screening for the presence of the shorter deleted allele and the absence of wildtype one and expanded for further analysis. Each of these cell lines therefore had a deletion of a single CTCF binding site located at the boundaries of the *Shh* locus. The locations of the CTCF deletions are labelled 1-5 on Figure 4.11(B). For the remainder of this document, these mutant cell lines are referred to as ES mutant 1-5.

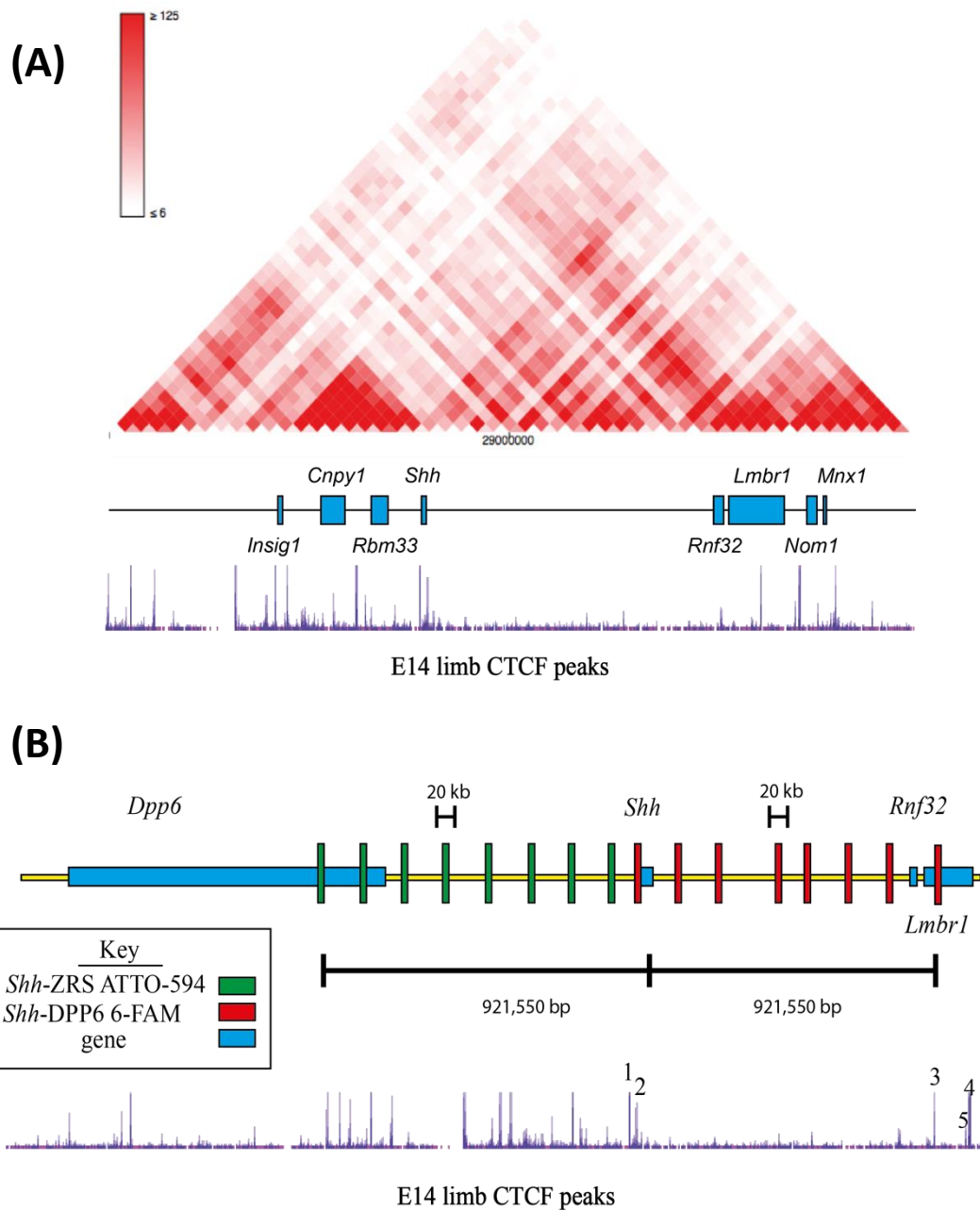
For each cell line, the number of spots observed for the *Shh*-ZRS region (appear as red spots) and the *Shh-Dpp6* region (appear as green spots) were counted. In total, spots for over 100 different sets of *Shh*-ZRS and *Shh-Dpp6* loci (70 loci for ES mutant 3) were counted. As all the ES cells were diploid, roughly 50 different nuclei were examined. In total, a maximum of 6 red spots in a single locus were observed, while all eight green spots could be observed in a single locus in all cell lines. This suggests that compared to the *Shh-Dpp6* region, the *Shh*-ZRS region is more compact.

On viewing the boxplots of both spot number and spot area for the *Shh*-ZRS and *Shh-Dpp6* regions, it is immediately obvious that the *Shh*-ZRS region is much more compact than the

*Shh-Dpp6* region in both E14 wild type ES cells and all the mutant cell lines (Figure 4.12). Mann-Whitney U tests were used to analyse the distribution of spot numbers between *Shh-ZRS* and *Shh-Dpp6* regions across all the cell lines. A p-value of <0.0001 was produced for both wild type and mutant lines, thereby confirming that there is a highly significant difference between the number of observed spots for the *Shh-ZRS* locus compared to the number of spots observed for the *Shh-Dpp6* locus across all cell lines studied. When the *Shh-ZRS* spot number distributions were compared to those of each of the mutants separately, there was no statistical difference. Therefore, when analysing in terms of spot number, there is no notable change in compaction of the *Shh-ZRS* locus in ES cells when CTCF sites binding the locus are removed independently.

In contrast to the spot number analysis, there was a significant difference in the total area (Figure 4.12(C)) of *Shh-ZRS* spots in the wild type cells compared to each of the CTCF mutant cell lines (p-value < 0.0001) with the wild type cells being less compact. In this experiment, analysis of both spot number and spot area produce conflicting results and, therefore, it cannot be confirmed if the CTCF deletions have an effect on compaction. Perhaps analysing spot area is more sensitive than analysing spot number. Performing chromosome conformation capture experiments (3C) on each of the cell lines and comparing the results to these MYtag experiments, may help to determine whether analysing spot area or spot number is more accurate.

The distributions of spot number and spot area are shown in Figures 4.13 and 4.14 respectively.



**Figure 4.11 – (A)** The heat map obtained from Hi-C showing the topological associated domain (TAD) containing the *Shh* gene in mouse ESCs at 40 kb resolution (<http://promoter.bx.psu.edu/hi-c/>). Interactions between regions are depicted by a red pixel, with a more intense red colour signifying a stronger interaction. Regions of CTCF binding obtained from Encode E14 ChIP data are shown below the heat map (<https://genome.ucsc.edu/ENCODE/>). **(B)** The region spanning from *Lmbr1* to *Dpp6* showing the position MYtag clusters bind. 8 MYtag clusters bind between *Shh* and *Lmbr1* and are labelled with the ATTO-594 fluorophore appearing red under the microscope while 8 MYtag clusters bind between *Shh* and *Dpp6* and are labelled with the fluorophore 6-FAM appearing green under the microscope. Each cluster is 20 kb in length with roughly 100 kb spacing between each cluster. Regions of CTCF binding obtained from Encode E14 ChIP data are shown below the heat map with labels 1 – 5 signifying the position of the CTCF binding site deletions studied in ES cell experiments.

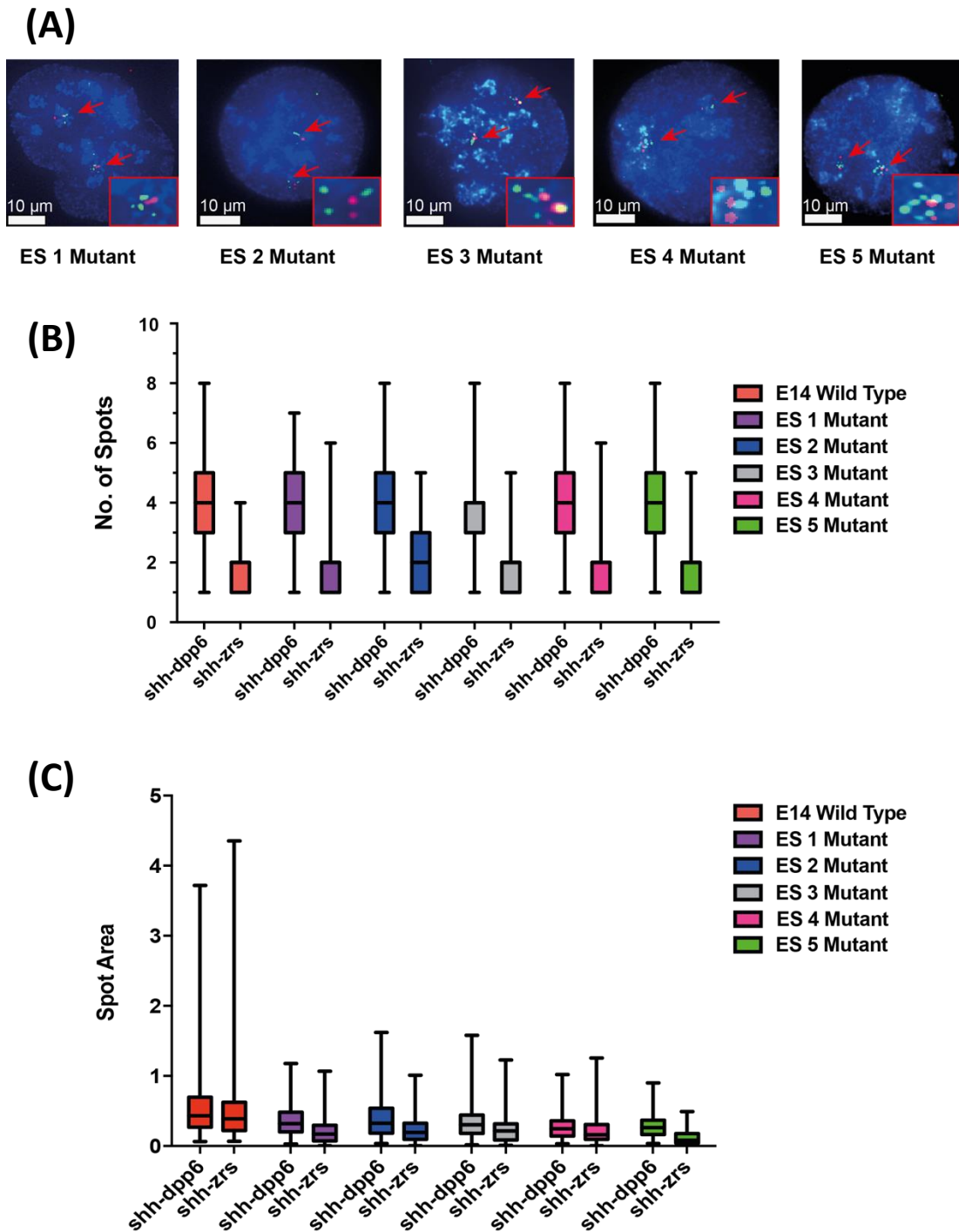


Figure 4.12 – (A) Images showing the MYtag spots in the 5 mutant ES cell lines. Red spots are Mytags binding to the region between *Shh* and *Dpp6* while green spots are MYtags binding to the region between *Shh* and *Lmbr1*. Red arrows point to the position of the MYtags. (Inset) Enhanced image of MYtags binding at one locus per mutant cell line. (B) Boxplots showing the distribution of the number of red and green spots counted for wild type and mutant ES cells ( $n > 70$ ). (C) Boxplots showing the distribution of calculated spot area of red and green spots for wild type and mutant ES cells ( $n > 70$ ).

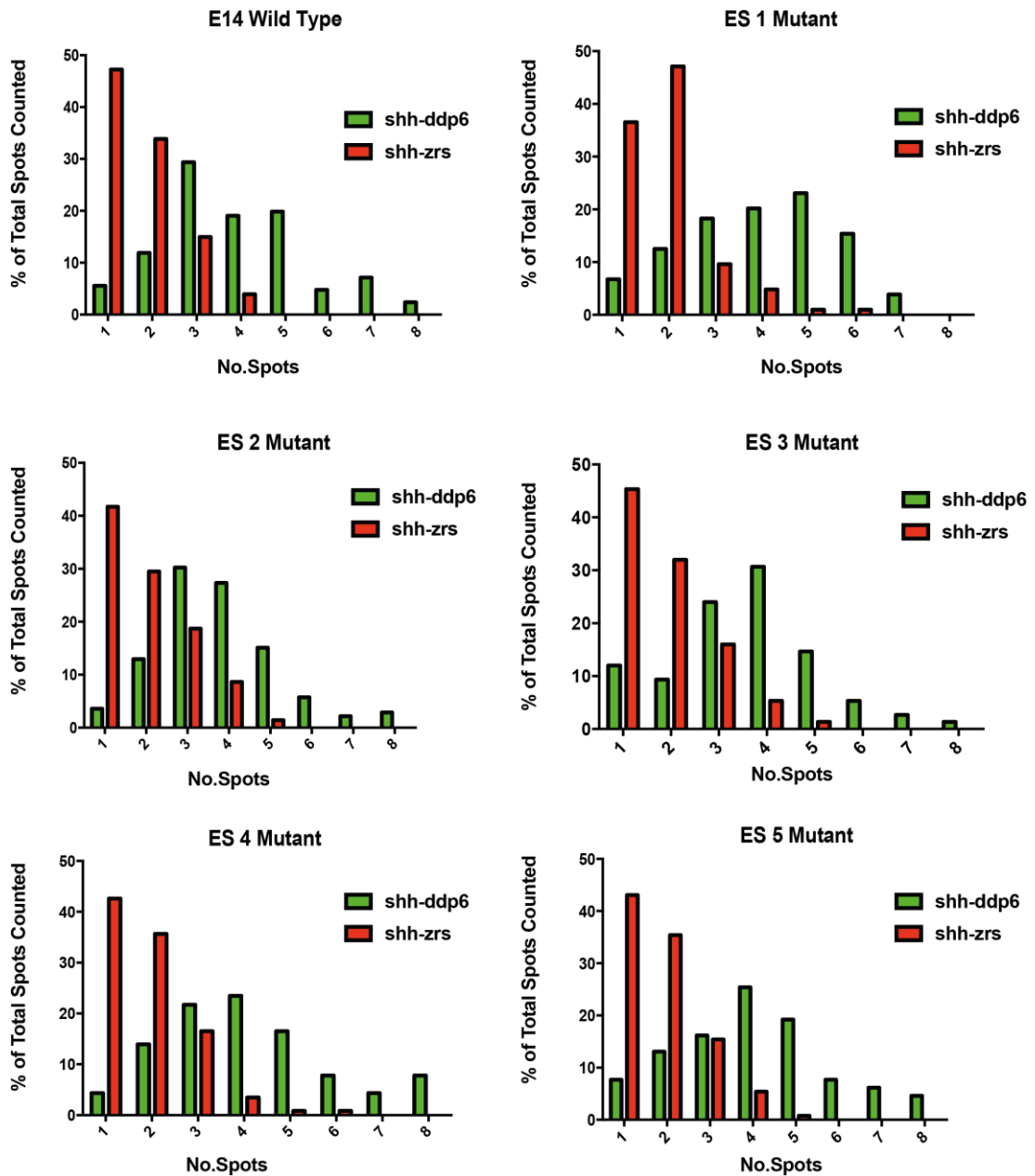


Figure 4.13 – The number of red and green spots counted for E14 wild type and mutant ES cells in the MYtag experiments plotted as a bar charts. The number of spots counted per locus (horizontal axis) is plotted against the percentage of the total spots counted (vertical axis) for that particular cell line. Red bars represent red spots counted for the MYtags binding to the region between *Shh* and *Lmbr1*, while green bars represent green spots counted for the MYtags binding to the region between *Shh* and *Dpp6*.

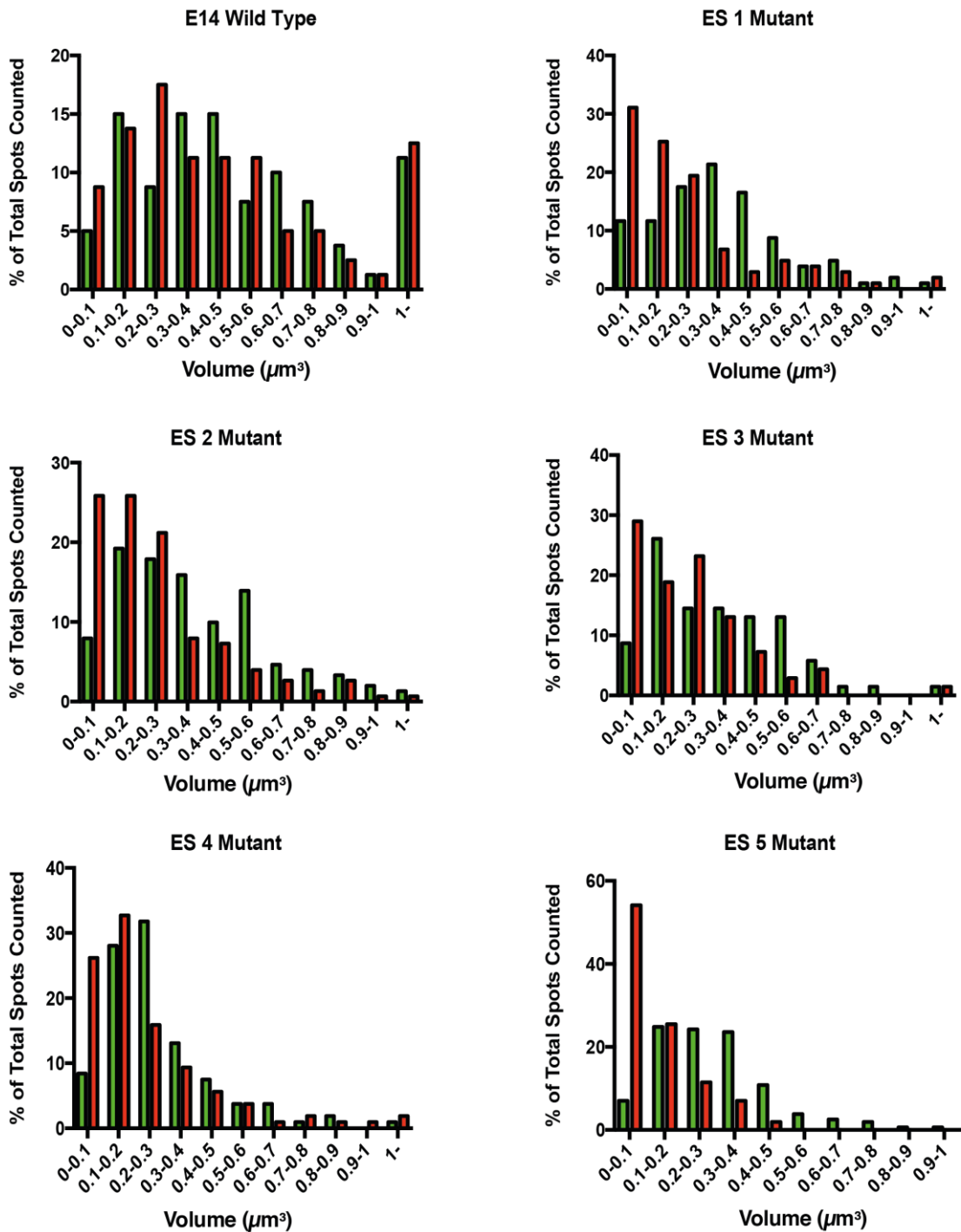


Figure 4.14 - The area of red and green spots counted for E14 wild type and mutant ES cells in the MYtag experiments plotted as a bar charts. The area of spots counted per locus (horizontal axis) is plotted against the percentage of the total spots counted (vertical axis) for that particular cell line. Red bars represent the areas of red spots for the MYtags binding to the region between *Shh* and *Lmbr1*, while green bars represent the areas of green spots for the MYtags binding to the region between *Shh* and *Dpp6*.

## 4.4 Discussion

In this chapter, 3D-FISH experiments showed that the distance between the ZRS and *Shh* is unaffected in 14fp, 88fp and MD cells when treated with TSA. This is also the case when PEA3 levels are reduced (an alternative method of inducing *Shh* expression) in the 14fp cell lines. The distance of these FISH probes from the nuclear periphery is also unaffected upon TSA treatment. These results contrast recent FISH data in which *Shh* and ZRS signals co-localise more closely in the ZPA of the E11.5 limb bud where *Shh* is expressed, compared to other regions of the limb where *Shh* is not expressed (Williamson et al., 2016). The reason for these conflicting results is unclear but it is possible that TSA may activate the gene through some alternative mechanism (discussed further in Chapter 5 and Silvia Peluso - manuscript in preparation). It should also be noted that compared to the ZPA of an E11.5 limb bud, the expression of *Shh* in the 14fp cells upon TSA treatment is relatively low. Although TSA is activating the *Shh* gene in 14fp cells, interactions may be occurring at a low level at any one time and, therefore, the co-localization of the ZRS *Shh* is not detectable by FISH.

An additional method of inducing *Shh* expression is to increase expression of the protein GABP $\alpha$  (Chapter 3). At present experiments are ongoing to determine if increasing levels of GABP $\alpha$  has an effect on the co-localisation of the *Shh* gene and the ZRS. To do this, 14fp cells are transfected with a *Gabpa* expression vector (Figure 4.15) capable of producing GABP $\alpha$  on treatment with doxycycline (dox) (Silvia Peluso - manuscript in preparation) and tested for *Shh*-ZRS co-localisation by 3D-FISH experiments. Attempts are being made to optimise this experiment by FACS sorting only the mCherry positive GABP $\alpha$  expressing cells before performing FISH. In this way, the distances between the ZRS and *Shh* can be examined in cells exclusively expressing GABP $\alpha$ .

Interestingly, there was a significant increase in the proportion of *Shh* probes located outside the nuclear territory compared to those located inside the territory upon TSA treatment. This accords with published data suggesting that within the ZPA, *Shh* relocates outside of its chromosome territory causing SHH production. Preliminary FISH data (see Appendix) using probes for both *Shh* and SBE4 in 14fp cells, where SBE4 is located between *Shh* and the ZRS in the locus, also suggests that the distance between *Shh* and SBE4 is greater than that between *Shh* and the ZRS. This also agrees with recent findings (Williamson et al., 2016) suggesting that in all cells, *Shh* and the ZRS are held tightly together perhaps forming a loop which moves SBE4 further away from both points (Appendices - Figure 1).

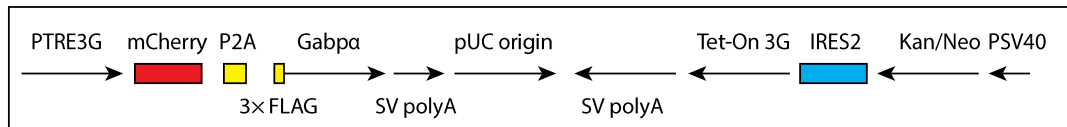
To analyse compaction of the 14fp cells, directly labelled fluorescent probes called MYtags were used to tile the regions spanning from *Shh* to the ZRS, and then from *Shh* to the *Dpp6*, where the distance between *Shh*-ZRS is roughly the same as that between *Shh*-*Dpp6*. The *Shh*-ZRS region was significantly more compact than the *Shh*-*Dpp6* region but, compaction of the *Shh*-ZRS region was unaffected by treatment with TSA. Again, this confirms that *Shh*-ZRS locus maintains a baseline level of compaction regardless of cell type, tissue or the expression status of the *Shh* gene. It should be noted that when using MYtags a complete complement of 8 spots were observed for the *Shh*-*Dpp6* region but only a maximum of 6 spots for the *Shh*-ZRS region. It was rationalised that the reason for this observation was that the *Shh*-ZRS region was more compact and therefore the spots were lying on top of each other so that only a select few could be observed. Of course, it may be that some of the MYtags did not anneal correctly and this is why 6 spots - instead of the full complement of 8 - were observed. To confirm that this is not the case, Fibre-FISH - which involves stretching out DNA - could be performed using the MYtags. If 8 spots are observed for the *Shh*-ZRS region, this would confirm that all MYtags are capable of binding and, the reason only 6 are seen in the ES cell and 14fp experiments is due to the compact nature of the chromatin.

MYtags were also used to analyse how compaction of the *Shh* locus is affected when key CTCF binding sites at the boundaries of the TAD containing the locus are deleted. To carry out this examination, ES cells were used each containing deletions of a separate CTCF binding site (deletions obtained using the CRISP/Cas9 system). It was found that in both wild type and mutant cell lines, the *Shh*-ZRS region was always more tightly compact than the region between *Shh*-*Dpp6*. However, on analysis of spot number, there was no significant difference in compaction of the *Shh*-ZRS loci between the different cell lines. In contrast, when spot area was analysed wild type cells were shown to be less compact than each of the mutant cell lines. The reasons for this are unclear as deletion of CTCF sites were expected to make the locus less compact. To confirm that the locus is becoming more compact on deletion of CTCF sites 3D-FISH experiments could be conducted on both wild type and CTCF mutant ES cells. Measuring the distance between various probes located across the *Shh* locus in these cell lines should confirm whether the mutants are less compact.

However, independent of spot number or spot area analysis, these results confirm recently published data that the *Shh* locus is highly compact even in tissues not expressing *Shh*. When considering spot number analyses, the reason why deletion of the CTCF binding sites did not affect compaction of the *Shh* locus is unclear. However, deleting one CTCF site at a time still



leaves 4 sites intact at the boundaries of the locus in addition to other sites located further away which may be enough to maintain the chromatin structure. It is postulated that if this experiment was repeated in ES cells containing double or triple deletions of TAD boundary CTCF binding sites, there would be a more pronounced relaxation of the chromatin structure and a greater number of MYtag spots observed for the *Shh*-ZRS region.



**Figure 4.15 - The expression vector used to express GABP $\alpha$  on dox administration. Cells are selected by growing in medium containing G418 solution. Those containing the vector express the *neo* gene thereby producing neomycin and are resistant. The placement of the IRES2 site within the vector also allows the production of the Tet-On 3G protein in addition to neomycin from the same PSV40 promoter. In the presence of dox, Tet-On 3G can bind to the PTRE3G promoter resulting in the production of mCherry and FLAG tagged GABP $\alpha$  with both proteins produced separately due to the presence of the intervening P2A self-cleaving peptide.**

## 4.5 Conclusions

In summary, (when considering spot number analyses) compaction of the *Shh* regulatory region was unaffected in ES cells with single deletions of CTCF binding sites located at the boundaries of the TAD containing *Shh*. Compaction of the locus was also not affected when 14fp cells were treated with TSA. 14fp, 88fp and MD cell lines showed no significant changes in *Shh*-ZRS co-localisation upon TSA treatment and, in addition, reducing PEA3 levels or increasing GABP $\alpha$  levels had no effect on *Shh*-ZRS co-localisation in 14fp cells. In contrast, TSA treatment caused *Shh* to loop out of its chromosome territory in 14fp cells which possibly explains the production of SHH.

# Chapter 5

---

## Conformational Changes within the *Shh* TAD

## 5.1 Introduction

Chapter 4 details the use of FISH experiments to explore compaction of the *Shh* locus and co-localisation of *Shh* and the ZRS in different cell lines. Firstly, 3D-FISH experiments using fosmid probes confirmed that, at least in the cell lines (14fp, 88fp and MD cell lines), there is no difference in the proximity of *Shh* and the ZRS in both TSA treated cells and untreated cells. Secondly, experiments performed in 14fp and ES cells using MYtags confirmed that the region from *Shh* to the ZRS is much more compact than that between *Shh* and *Dpp6*. However, these experiments gave little insight into the effect *Shh* expression has on *Shh* gene interactions within the entire *Shh* regulatory domain. To identify these additional changes within the locus (if any), we used a combined approach of carbon copy chromosome conformation capture (5C) with 3D modelling and circular chromosome conformation capture (4C). 3C techniques are commonly used to confirm results from FISH data and vice versa. However, although these techniques are usually complementary, there are situations where results obtained from such procedures do not correlate. For example, in one investigation an increase in interaction frequency between two 5C restriction fragments was observed for a region shown to become less compact in FISH experiments (Williamson et al., 2014). Therefore, we were also interested in how the results from both 4C and 5C would compare with the previously obtained FISH data and the predicted structural models of the *Shh* regulatory domain.

Chromatin conformation capture techniques have already been used to study the *Shh* locus in several different tissues. In the limb bud 3C confirmed FISH data that the ZRS and *Shh* are in close proximity in both anterior and posterior portions of the limb (Amano et al., 2009). 5C experiments using cells from the embryonic limb, head and body confirmed Hi-C data that the *Shh* regulatory region is contained within a TAD and the structure of this TAD is conserved across different tissues (Williamson et al., 2016).

As mentioned previously (Chapter 1), the activity of enhancers within the *Shh* regulatory region has been examined using the sleeping beauty transposon system (Anderson et al., 2014). The sleeping beauty transposon carrying LacZ reporter gene inserts randomly into the locus allowing the spatiotemporal activity of the various enhancers to be assessed. This showed that the ZRS acted upon regions very close to itself in either direction. Inserts in the gene desert showed increasing LacZ expression the closer they were positioned to *Shh*. Therefore, using 5C we proposed to examine whether the frequency of ZRS contacts increases closer to *Shh* (in

agreement with the Sleeping Beauty system) and whether there are any other substantial interactions within the gene desert.

## 5.2 Aims

### 5.2.1 General Aims

The aim of this section was to perform chromosome conformation experiments on the *Shh* inducible and control cell lines (14fp, 88fp and MD cells) to identify interaction differences. The interactions within the TAD containing the *Shh* gene will be examined to determine the changes that may occur in the activation of *Shh* expression.

### 5.2.2 Experimental Aims

The experimental aims of this section are as follows:

- Perform 5C on both TSA treated and control (*Shh* expressing and non-expressing) 14fp, 88fp and MD cell lines.
- Identify significant interaction differences between the cell lines using in silico 4C profiles.
- Identify significant interaction differences between TSA treated and control cells (*Shh* expressing and non-expressing) of the same cell line using in silico 4C profiles.
- Use the Autochrom3D modelling tool to create 3D models of the 5C data.
- Perform 4C on both TSA treated and control 14fp cell lines.

## 5.3 Results

### 5.3.1. 5C quality control and processing

5C experiments were conducted as described in Chapter 2. Gel pictures showing quality control checks at various stages of the protocol are provided in Figure 5.1. All 5C libraries were created by first producing a 3C library by digesting formaldehyde fixed cells with HindIII restriction enzyme followed by ligation with T4 DNA ligase. 5C primers (provided by the Bickmore laboratory) designed to anneal to the ends of HindIII fragments were annealed, ligated and amplified to produce the final 5C library. In total 365 primers were designed to anneal across a region encompassing the entire *Shh* TAD (Chr5: 28,317,087-30,005,000). All 5C libraries were sequenced using in-house Ion Torrent PGM sequencing. The number of raw sequence reads are provided in Table 5.1. These were processed using an automated pipeline provided by the Dostie laboratory which filtered out poor quality reads ( $Q < 30$ ) and aligned remaining reads to the mouse mm9 genome. The output of this tool was an interaction frequency matrix which can be used uploaded onto the my5C website (<http://my5c.umassmed.edu/welcome/welcome.php>) and used to analyse 5C heatmaps.

In addition to primers designed across the *Shh* regulatory region, primers were also designed for a region containing *Usp22* gene (Chr11: 60,917,307-61,003,268) which in previous studies shows no interaction with the *Shh* regulatory domain and therefore acts as a good control region for 5C experiments (Berlivet et al., 2013). High quality 5C experiments showed interactions within the *Usp22* gene which did not interact with the *Shh* containing TAD. Poor quality experiments showed interactions between the TADS containing the *Shh* and *Usp22* genes and subsequently discarded. The quality of 5C experiments could also be examined using R packages such as HiTC, the output of which is shown in Figure 5.2. This shows that all interactions are maintained within a single chromosome with negligible *trans* interactions across different chromosomes (Figure 5.2(A)). Also, the interactions of a single point are greatest with those closest to it with the interaction frequency decreasing with points which are further away (Figure 5.2(B)).

All 5C experiments were normalized by dividing reads counts per primer pair by the total read number and multiplying the values by 1000. This produced a matrix with each value in the matrix representing the frequency with which two primers interact (termed the interaction frequencies). These were analysed using my5C. For initial analysis (example in Figure 5.3) primer pair interactions were binned into 30 kb regions producing 57 “bins” spanning the

region from Chr5: 28317087 – 30005000. Interactions were also studied in 10 kb and 50 kb bins (Figure 5.5). Unaltered heat maps are included in the Appendices. In some experiments single primer pair interactions produced interaction frequencies with a much greater value (>100 fold) than the average interaction frequency. These were removed from the interaction frequency matrices before normalization. Other studies have removed single interaction frequency values in a similar way (Smith et al., 2016). Pearson coefficients were calculated to determine how similar the interaction frequency matrices of each 5C experiment were (Table 5.2). These values range from zero to one, with a value of zero suggesting no correlation and a value of one suggesting a perfect correlation between two samples. Replicate 5C samples showed a high degree of similarity with Pearson coefficient values > 0.69 (Table 5.2). Upon removal of singletons, these values also increased showing an even greater degree of similarity between replicates.

| <b>Sample</b>       | <b>Biological Replicate 1</b> | <b>Biological Replicate 2</b> |
|---------------------|-------------------------------|-------------------------------|
| <b>88fp Control</b> | 923,410                       | 4,201,036                     |
| <b>88fp TSA</b>     | 1,584,078                     | 3,856,256                     |
| <b>14fp Control</b> | 3,173,784                     | 2,863,822                     |
| <b>14fp TSA</b>     | 4,604,748                     | N/A                           |
| <b>MD Control</b>   | 4,181,036                     | N/A                           |
| <b>MD TSA</b>       | 1,671,580                     | 5,144,862                     |

**Table 5.1 - Number of raw sequence reads from Ion PGM sequencing for the 5C experiments.**



| Sample             | 88fp Control Bio 1 | 88fp TSA Bio 1 | 14fp Control Bio 1 | 14fp TSA Bio 1 | MD Control Bio 1 | MD TSA Bio 1 | 88fp Control Bio 2 | 88fp TSA Bio 2 | 14fp Control Bio 2 | MD TSA Bio 2 |
|--------------------|--------------------|----------------|--------------------|----------------|------------------|--------------|--------------------|----------------|--------------------|--------------|
| 88fp Control Bio 1 | 1                  | 0.875716745    | 0.637326919        | 0.83115879     | 0.738442604      | 0.798823022  | 0.711805209        | 0.56811442     | 0.539499937        | 0.670148406  |
| 88fp TSA Bio1      | 1                  | 1              | 0.750823341        | 0.919086463    | 0.853210952      | 0.880584622  | 0.827359493        | 0.693258798    | 0.667054357        | 0.798941321  |
| 14fp Control Bio 1 | 1                  | 1              | 1                  | 0.864456814    | 0.91940296       | 0.745806469  | 0.827456549        | 0.783325613    | 0.88999909         | 0.685901203  |
| 14fp TSA Bio 1     | 1                  | 1              | 1                  | 1              | 0.930068257      | 0.893060522  | 0.899838221        | 0.790844487    | 0.78366076         | 0.835168586  |
| MD Control Bio 1   | 1                  | 1              | 1                  | 1              | 1                | 0.895292552  | 0.939396556        | 0.82722482     | 0.889922088        | 0.861063897  |
| MD TSA Bio 1       | 1                  | 1              | 1                  | 1              | 1                | 1            | 0.922933825        | 0.743376185    | 0.773809468        | 0.909123488  |
| 88fp Control Bio 2 | 1                  | 1              | 1                  | 1              | 1                | 1            | 1                  | 0.890162865    | 0.892809344        | 0.930218166  |
| 88fp TSA Bio 2     | 1                  | 1              | 1                  | 1              | 1                | 1            | 1                  | 1              | 0.830696877        | 0.771968366  |
| 14fp Control Bio 2 | 1                  | 1              | 1                  | 1              | 1                | 1            | 1                  | 1              | 1                  | 0.771968366  |
| MD TSA Bio 2       | 1                  | 1              | 1                  | 1              | 1                | 1            | 1                  | 1              | 1                  | 1            |

### Key






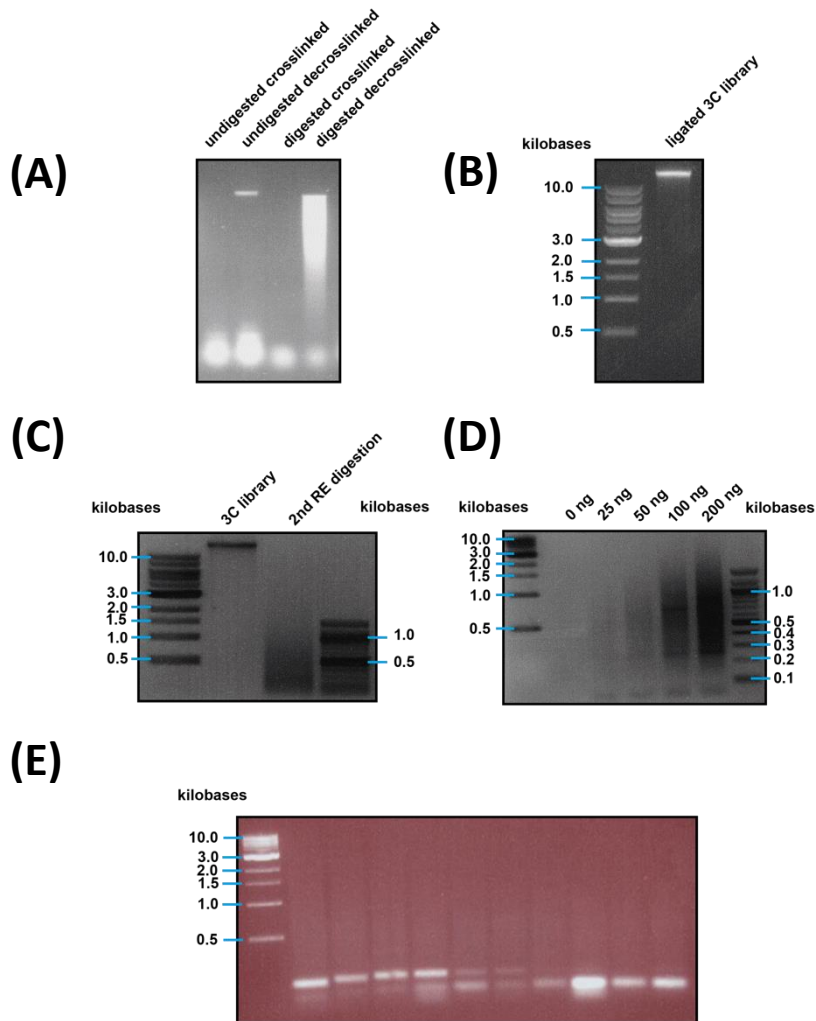
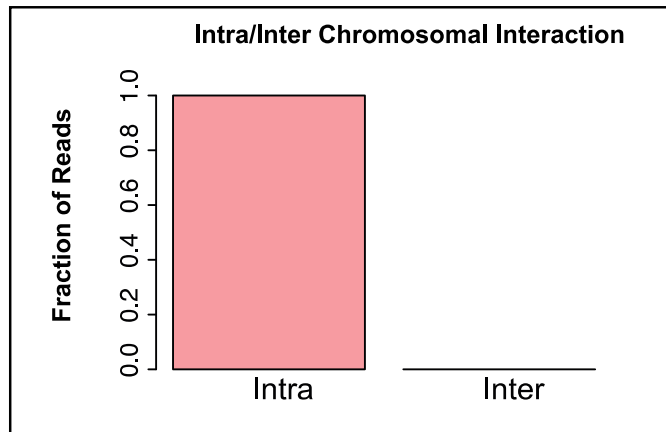
| Colour                                                                              | Pearson Coefficient |
|-------------------------------------------------------------------------------------|---------------------|
|  | 0.9-1.0             |
|  | 0.8-0.9             |
|  | 0.7-0.8             |
|  | 0.6-0.7             |
|   | 0.5-0.6             |

Table 5.2 - Calculated Pearson coefficient values showing the similarity between different samples and replicates. Pearson calculations were conducted using unaltered interaction frequency matrices for each sample. Pearson coefficient values can range from 0 to 1, with highly similar samples having values close to 1 and dissimilar samples having values closer to 0.



**Figure 5.1 - Agarose gels showing quality checks at various stages of the 4C/5C protocols. (A) Undigested decrosslinked formaldehyde fixed cells appear as a high molecular weight band while decrosslinked Hind III digested cells appear as a high molecular weight smear (running from around 12 – 4 kb). (B) 3C libraries appear as a high molecular weight band. (C) In 4C the 2<sup>nd</sup> restriction digest (in this case with enzyme MluCI) results in a low molecular weight smear running from around 0.3 – 1 kb. (D) Different starting concentrations of 4C library (0 – 200 ng) are tested by PCR to assess which concentration produces maximum product while at the same time avoiding saturation. This concentration is used in subsequent PCRs. (E) In the 5C protocol, 3C libraries are tested by PCR using two primers on HindIII fragments which only come together on ligation of successfully digested cells. The 3C libraries are first serially diluted with PCR conducted on increasingly lower starting concentrations. PCRs run on an agarose gel show a decrease in amplified PCR product with an eventual increase in primer dimer. Using gel densitometry, the intensity of this product can be plotted against starting 3C library concentration. The starting 3C library concentration used to make 5C libraries is chosen from the linear portion of this graph.**

(A)



(B)

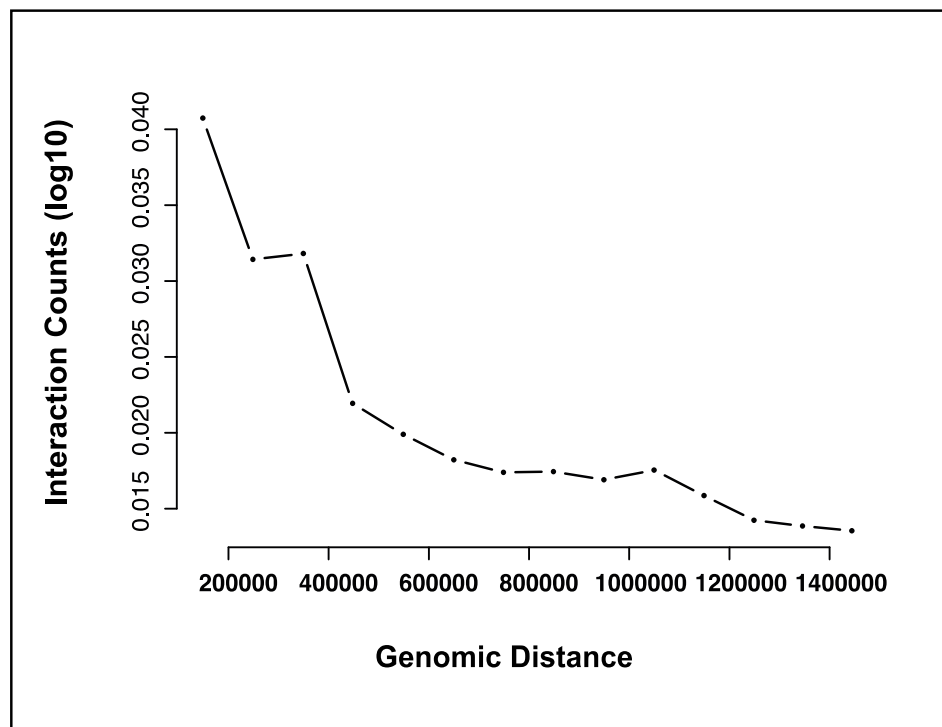


Figure 5.2 – Quality control checks used to assess the 5C experiments. (A) A typical 5C experiment shows only intra molecular interactions and no *trans* interactions. (B) For a given 5C primer the number of interactions with other primers decrease with distance.

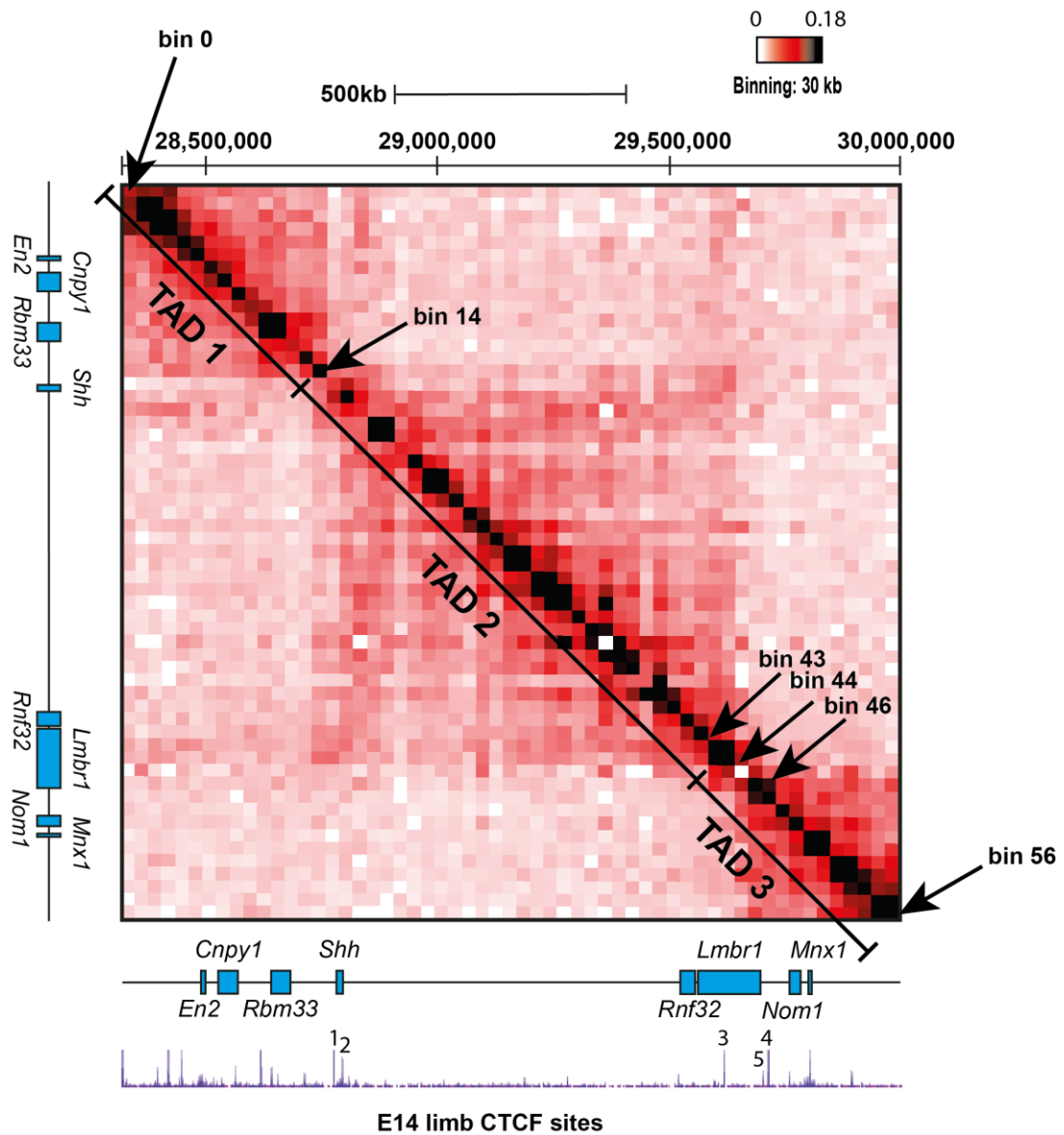


Figure 5.3 – A typical heat map obtained from the my5C website (in this case for the TSA treated 88fp cell line with interactions binned into 30 kb regions). Each pixel represents an interaction between two regions. Interactions occurring at low frequency appear white, while more frequent interactions appear red. The most frequent interactions appear black. A dissecting diagonal line appears in black signifying the high interaction frequency of regions located next to each other. This line highlights the genomic position of the binned fragments i.e. bin 0 (beginning Chr5: 28317087) begins at the top left of the heat map while bin 56 appears in the bottom right corner (ending Chr5: 30005000). All cell lines studied show 3 large TADs (visible as 3 large red squares). The first TAD contains the genes *En2*, *Cnpy1* and *Rbm33*, the second TAD contains the genes *Shh*, *Rnf32* and part of *Lmbr1*, while the third TAD contains the remainder of *Lmbr1*, *Nom1* and *Mnx1*. Each heat map consists of two triangles displaying the same interactions reflected along the diagonal. Only one of these triangles needs to be analysed and subsequent heat maps are depicted as a single triangle. E14 limb CHIP-seq peaks from ENCODE are displayed and labelled as Chapter 4.

### 5.3.2 Optimisation of 5C Procedure

Before sequencing, 3 µl of each 5C library was amplified through 26 cycles of PCR and analysed using the Agilent Bioanalyser. However, on several occasions this did not produce enough amplified 5C library for sequencing and left a substantial amount of primer dimers which may affect sequencing. Therefore, for certain libraries PCR was conducted using an increased starting volume of 5C library from 3 µl to 6 µl and the number of cycles was increased from 26 cycles to 28, 30 and occasionally 32 cycles. This produced substantially more 5C product with less primer dimer for sequencing.

### 5.3.3 The TAD containing the *Shh* regulatory domain is conserved across the cell lines

5C experiments were conducted using TSA treated and control 14fp, 88fp and MD cell lines. The aim of these experiments was to determine how the interactions of the ZRS and the *Shh* gene are altered between *Shh* expressing and non-expressing cells. As mentioned in the previous chapters, the 14fp and 88fp cell lines express *Shh* on TSA treatment while the MD cells do not and are therefore used as a control. Each 5C experiment produced a heat map with the same general features as Figure 5.3 (when interactions are grouped into 30 kb bins). In this heat map, three main TADs are obvious. The first spans from bin 0 to bin 14 (Chr5: 28317087 – 28767087); the second from bin 14 to bin 44 (Chr5:28737087 – 29667087); and, the third from bin 45 to bin 56 (chr5:29667087 – 30005000). The first TAD contains the genes *En2*, *Cnpy1* and *Rbm33*; the second TAD contains *Shh*, *Rnf32* and part of the *Lmbr1* gene; and, the third contains the remainder of *Lmbr1*, *Nom1* and *Mnx1*. The position of the TAD boundaries correlates closely with CTCF binding sites. For example, limb E14 CTCF peaks located near the *Shh* gene are very close to the boundary between TADs 1 and 2 and, peaks at the *Lmbr1* gene are close to the boundary between TADs 2 and 3 (these CTCF binding sites are those which have been deleted in MYtag ES cell experiments and the labelling of peaks is consistent with Chapter 4). Interestingly, a smaller region of high interactions is observed between bins 43 and 46 (Chr5:29607087 – 29727087) which lies between TADs 2 and 3. This may be a smaller TAD formed by the interactions of CTCF protein bound at sites 3, 4 and 5. (Nb E14 limb CTCF binding sites were obtained from ENCODE at UCSC (<https://genome.ucsc.edu/ENCODE/>)).

### 5.3.4 *Shh* and the ZRS show high interaction frequencies in *Shh* expressing and non-expressing cells

The interaction between *Shh* and the ZRS was analysed in both TSA treated and control cells (arrows, Figure 5.4). High interaction frequencies were observed between both regions in all cells regardless of *Shh* expression and thus confirm 3D-FISH data showing the same levels of co-localisation between *Shh* and the ZRS (Chapter 4). To confirm that such interactions were significant, interactions from each 5C experiment were grouped into bins of either 10, 30 or 50 kb. The Z-score for each bin was calculated according to the following equation (Baù and Marti-Renom, 2011):

$$\text{Z-score} = (\text{averageLog}_{10} - \text{currentEntryLog}_{10} / \text{stdevLog}_{10})$$

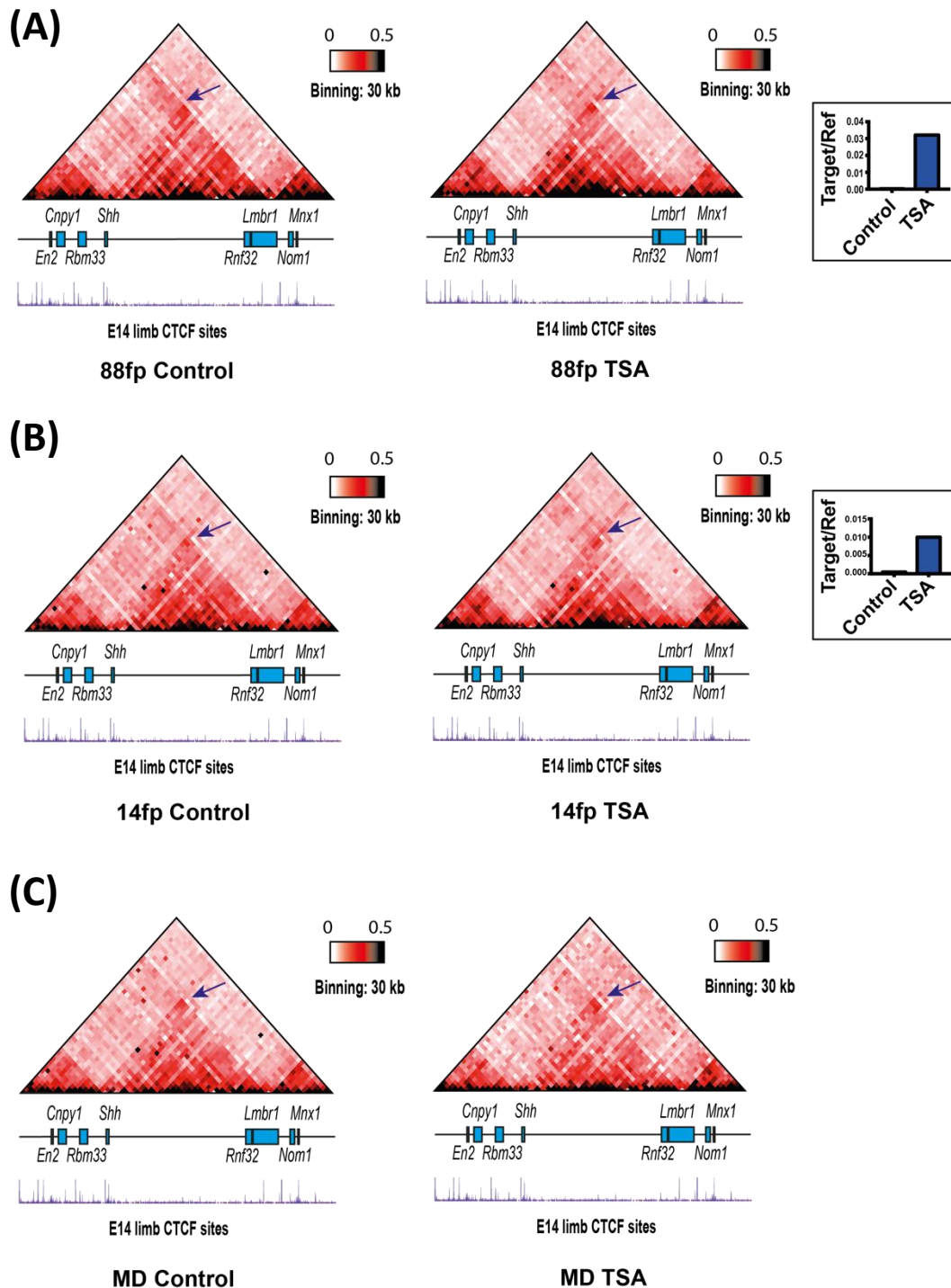
This calculation involves subtracting the  $\log_{10}$  interaction frequency value for each interaction from the average  $\log_{10}$  interaction frequency of the entire dataset and dividing by the standard deviation of the  $\log_{10}$  interaction frequencies. Using the Z-scores, the associated p-value of each interaction was calculated with the assumption that the data was normally distributed. Only interactions with a p-value of  $< 0.001$  were considered significant and are shown in Figure 5 for single replicates of both TSA treated and control 14fp, 88fp and MD cell lines. Strikingly, at 30 kb resolution the interaction between bins containing *Shh* and the ZRS are significant in each cell line (arrows in Figure 5.5) regardless of *Shh* expression, thereby confirming that both regions are in close contact.

*In-silico* 4C graphs were plotted from the 5C data using either *Shh* or the ZRS as the “bait” (Figures 5.6 and 5.7). When *Shh* was used, the increase in interaction frequency at the ZRS was obvious compared to the surrounding sequence. When the ZRS was used as “bait”, there was an obvious enrichment of interactions at *Shh* with virtually no interactions in the interceding regions. The neural enhancer SBE4, which is located roughly in the middle of the sequence separating *Shh* and the ZRS, was also examined. The interaction frequency between *Shh* and the ZRS was always greater than the interaction frequency between *Shh* and SBE4 or between the ZRS and SBE4. This suggests that in the cell lines, *Shh* and the ZRS are always in much closer contact than *Shh* and SBE4 and, also agrees with a looping model where *Shh* and the ZRS are brought together resulting in the SBE4 being pushed away from this location in the loop.

These trends were confirmed when the experiments were repeated with an additional biological replicate for each cell line and treatment (Figure 7). However, replicates for TSA treated 14fp cells and control MD cells were of poor quality and must be repeated. These have been omitted from this analysis.

### **5.3.5 In 14fp and 88fp cells TSA treatment reduces the interaction frequency of regions located around *Lmbr1*.**

To assess the differences between *Shh* expressing and non-expressing cells for each of the cell lines, normalised interaction matrices for control cells were subtracted from normalised interaction matrices from TSA treated cells (Figure 5.8). For the 14fp and 88fp cell lines, TSA treatment resulted in decrease in interaction frequency within TAD2 of regions included in bins 39 to 44 (chr5:29487087 – 29667087) (when interactions are grouped into 30 kb bins). This region includes the genes *Rnf32* and *Lmbr1*. An increase in interaction frequency of regions included in bins 14 to 38 (chr5: 28737087 – 29487087) was observed upon TSA treatment. This affect was not observed in MD cells.



**Figure 5.4 - Heat maps from 5C experiments for TSA treated and control (A) 88fp cells (B) 14fp cells and (C) MD cells. The intensity of interactions represents the frequency of interactions and ranges from white to red to black with increasing frequencies. All heat maps show interactions binned into 30 kb regions. Purple arrows identify the interaction between bins containing *Shh* and the ZRS. E14 limb ChIP-seq peaks from ENCODE are displayed. Each heat map represents a single biological replicate.**



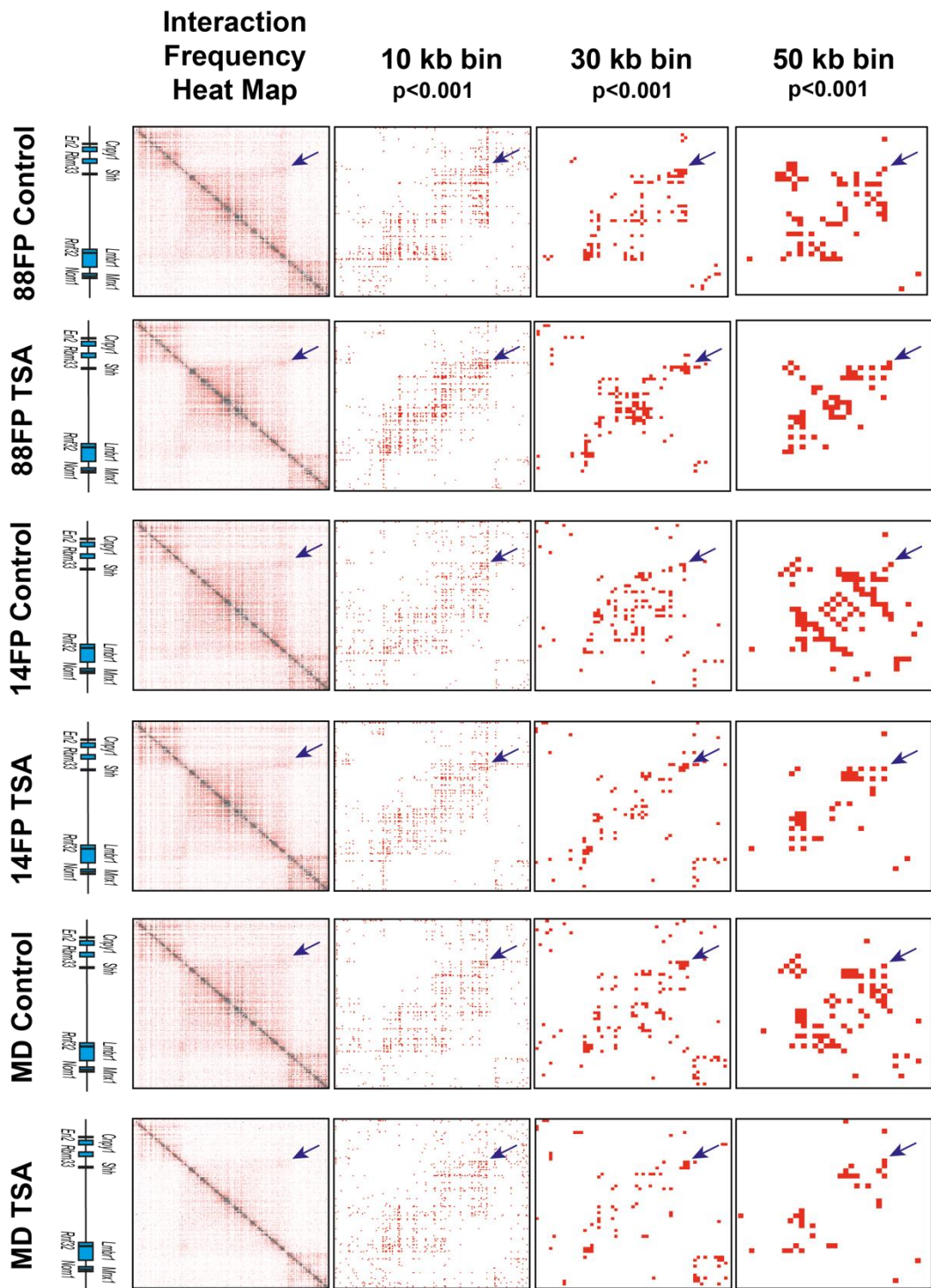


Figure 5.5 – The interactions for each 5C experiment were grouped into bins of either 10, 30 or 50 kb and the Z-scores for each bin calculated as described in Chapter 5.3.3. Using the Z-scores the p-value for each interaction was calculated assuming the data was normally distributed. For the binned interaction data above, only significant interactions are displayed. The normalized interaction frequency heat maps for each experiment showing all interactions are also displayed. Purple arrows point to the interaction between the ZRS and *Shh*.

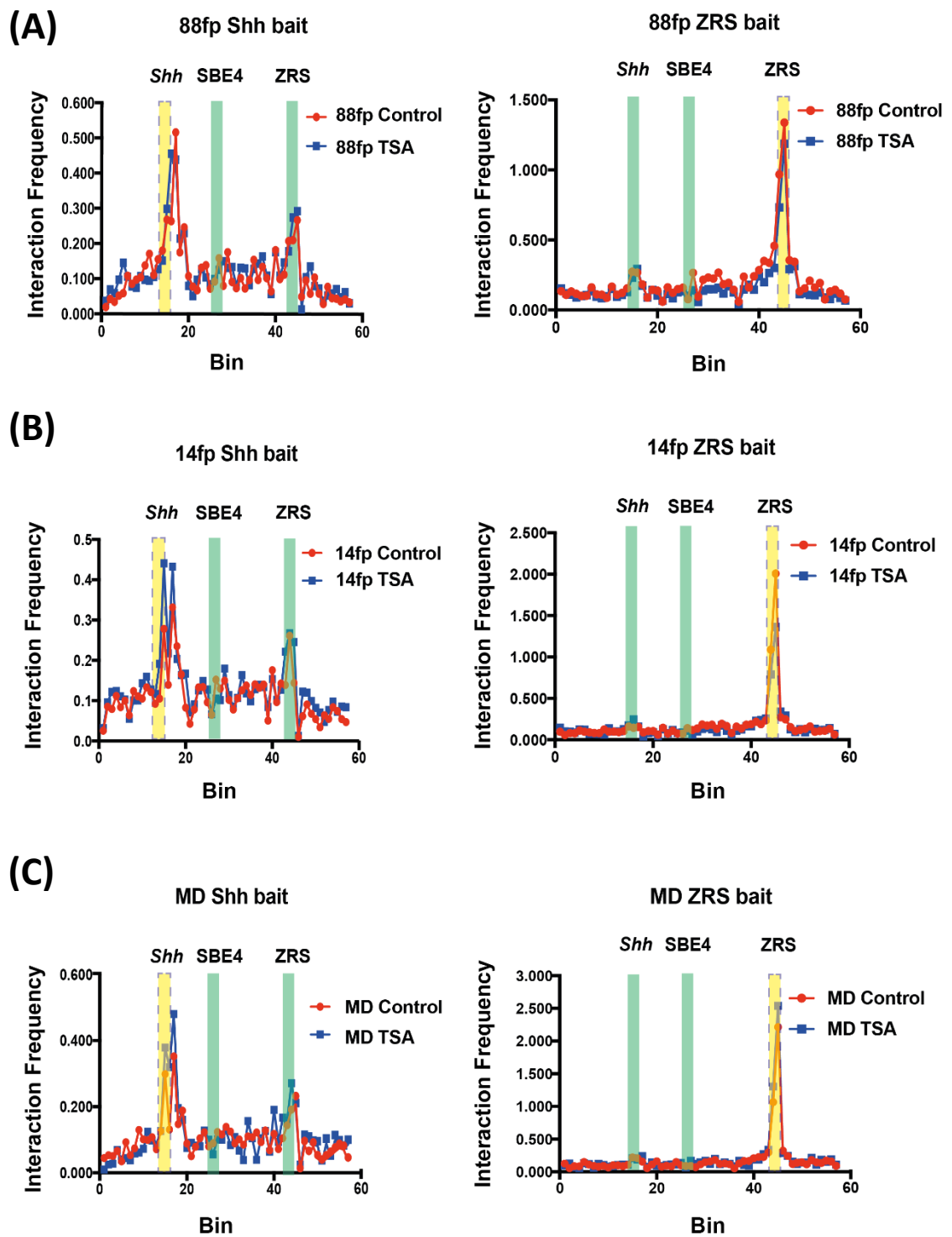
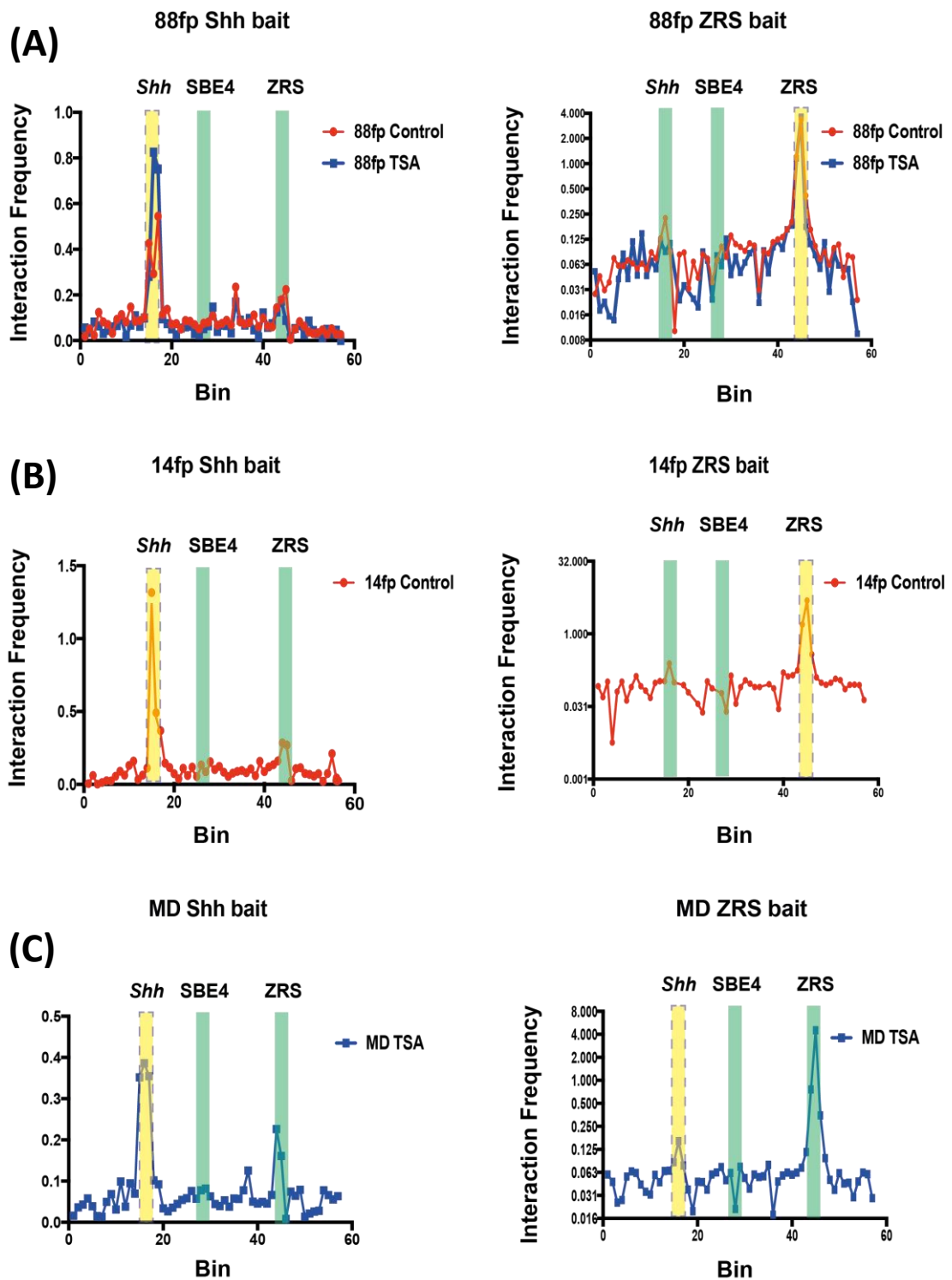
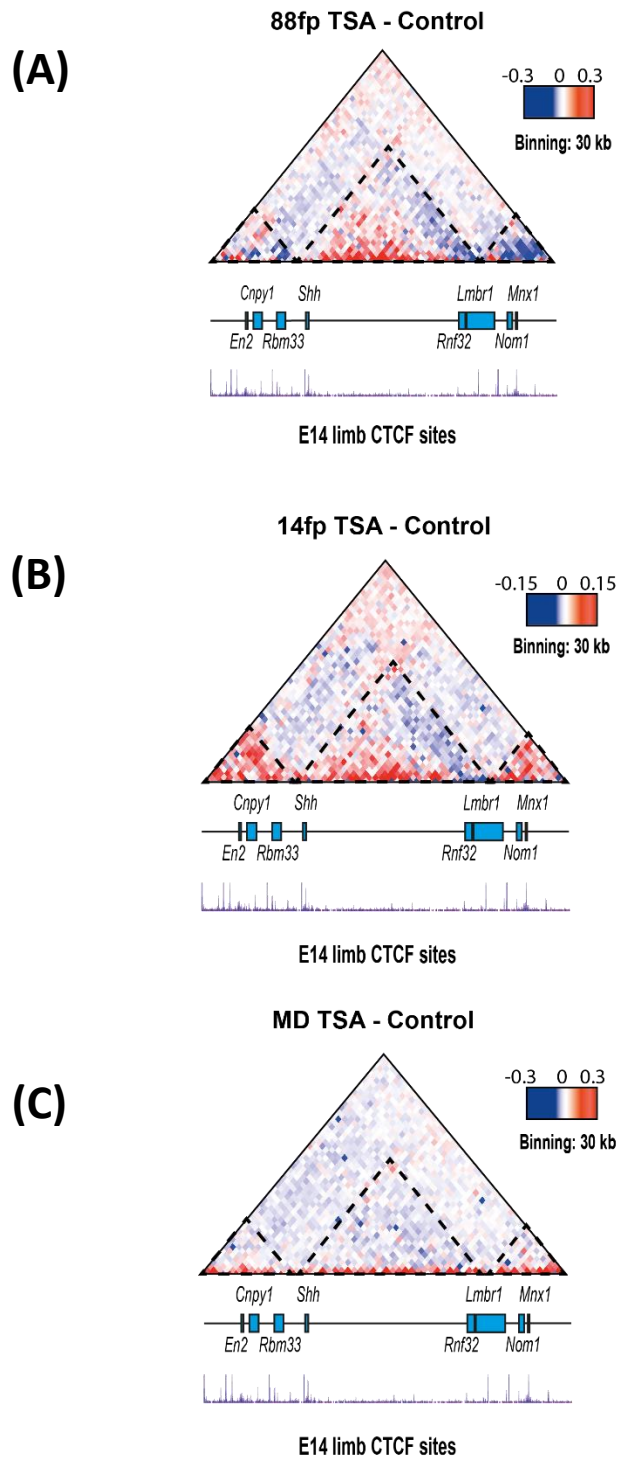


Figure 5.6 – *In silico* 4C profiles obtained from the 5C heat maps using either *Shh* or the ZRS as the bait sequence for (A) 88fp cells (B) 14fp cells and (C) MD cells. Interactions of control cells are depicted as red lines while interactions of TSA treated cells are depicted as blue lines. The bait sequence is highlighted in yellow in each graph while other regions of interest are highlighted in green. These profiles are obtained from the 1<sup>st</sup> biological replicate for each cell line.



**Figure 5.7 – *In silico* 4C profiles obtained from the 5C heat maps using either *Shh* or the ZRS as the bait sequence for (A) 88fp cells (B) 14fp cells and (C) MD cells. Interactions of control cells are depicted as red lines while interactions of TSA treated cells are depicted as blue lines. The bait sequence is highlighted in yellow in each graph while other regions of interest are highlighted in green. These profiles are obtained from the 2<sup>nd</sup> biological replicate for each cell line. TSA treated 14fp cells and control MD cells produced poor quality heat maps and thus *in silico* 4C profiles have been omitted.**



**Figure 5.8 – Heat maps showing the difference in interaction frequencies in control and TSA treated (A) 88fp (B) 14fp and (C) MD cell lines. Enrichment of TSA interactions appear as red pixels while enriched interactions within control cells appear in blue. More intense colors depict greater interaction frequencies. All heat maps show interactions between binned into 30 kb regions. E14 limb ChIP-seq peaks from ENCODE are displayed. Each heat map represents the difference in interactions between single biological replicates.**

### 5.3.6 Modelling of 5C data

Working with our collaborators in the Semple laboratory, (PhD student Ben Moore) the normalised and binned (30 kb bins) interaction frequency matrix obtained from the my5C website was used to create a model of the 5C data (Figure 5.9). The “pairwise” file from my5C was converted to a format suitable for inserting into the online modelling tool Autochrom3D (Peng et al., 2013) (R script provided in Appendices (Moore, B)). 10 simulations (Figure 5.10) were created for both TSA treated and control 14fp, 88fp and MD cell lines (using biological replicate 1 for each) and the distances between *Shh* and the ZRS compared by Mann-Whitney U tests. Within these models, *Shh* lies across two beads and so the distances between both of these beads and the bead containing the ZRS were analysed. For both 14fp and 88fp cells, there was no significant difference in the distance between the ZRS and *Shh* when measuring from either of the *Shh* beads ( $p > 0.05$ ). These results agree with 3D-FISH data (Chapter 4) where TSA treatment does not affect the distance between *Shh* and ZRS fosmid probes (Figure 5.11(A)(B)). However, there was a significant difference in the distance between *Shh* and the ZRS in MD cells when measuring from either of the *Shh* beads ( $p < 0.0001$ ) (Figure 5.11 (A)(B)). This contrasts 3D-FISH data where no significant change is observed in *Shh* and ZRS fosmid probe distances on TSA treatment.

Another output from Autochrom3D was the radius of gyration (ROG) for each simulation which can serve as an indicator for compaction (all radius of gyration values calculated as described by (Peng et al., 2013)). Small ROG values indicate a compact structure while larger values suggest a decompact structure. In this investigation, the ROG was calculated for each simulation and Mann-Whitney U tests used to determine if there was a significant difference in values between cell lines (Figure 5.11(C)). Interestingly, for each cell line there was a significant increase in the ROG upon TSA treatment, suggesting a decrease in compaction. This contrasts the MYtag FISH experiments where no significant difference in compaction was observed in the 14fp cell line upon TSA treatment.

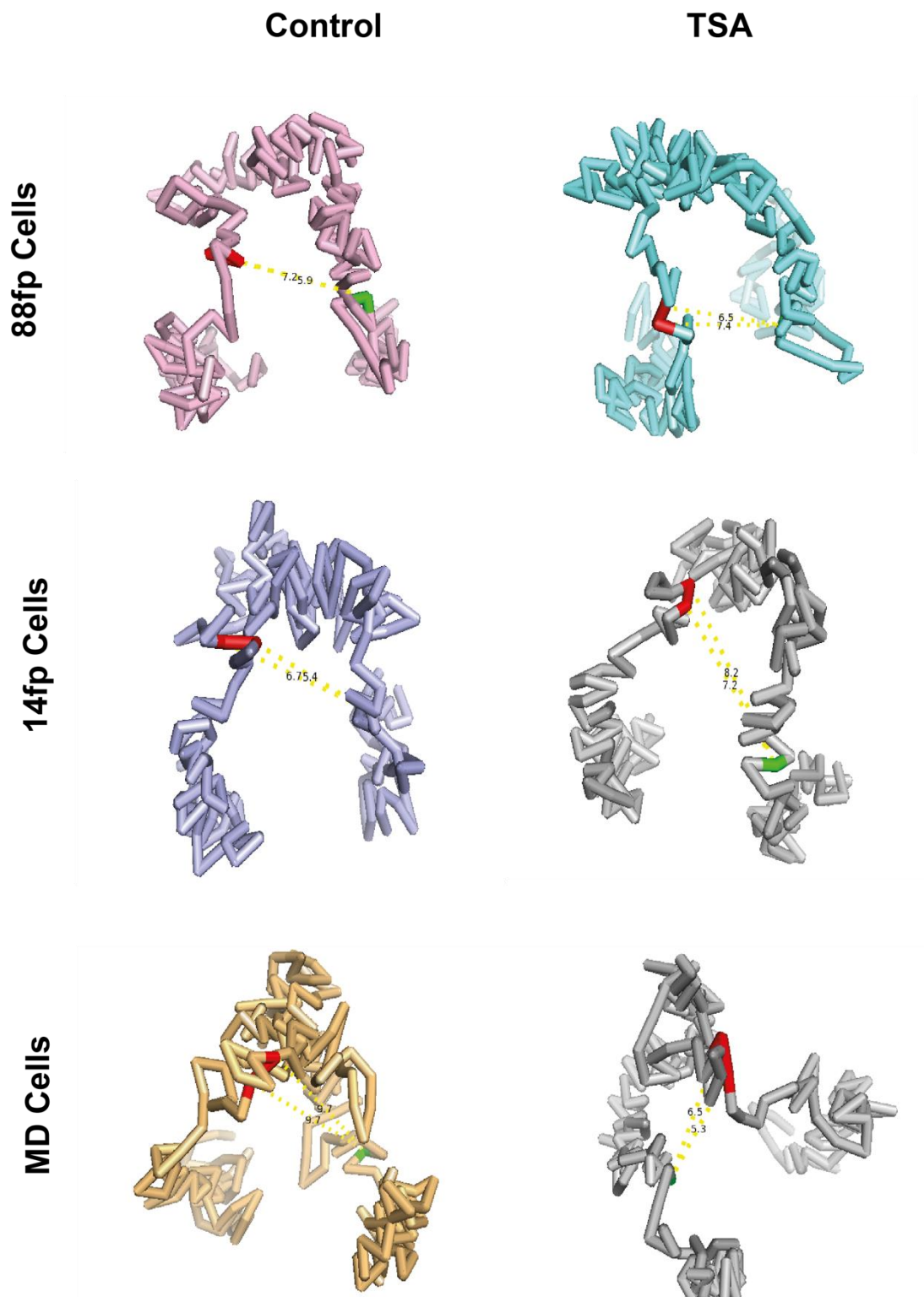
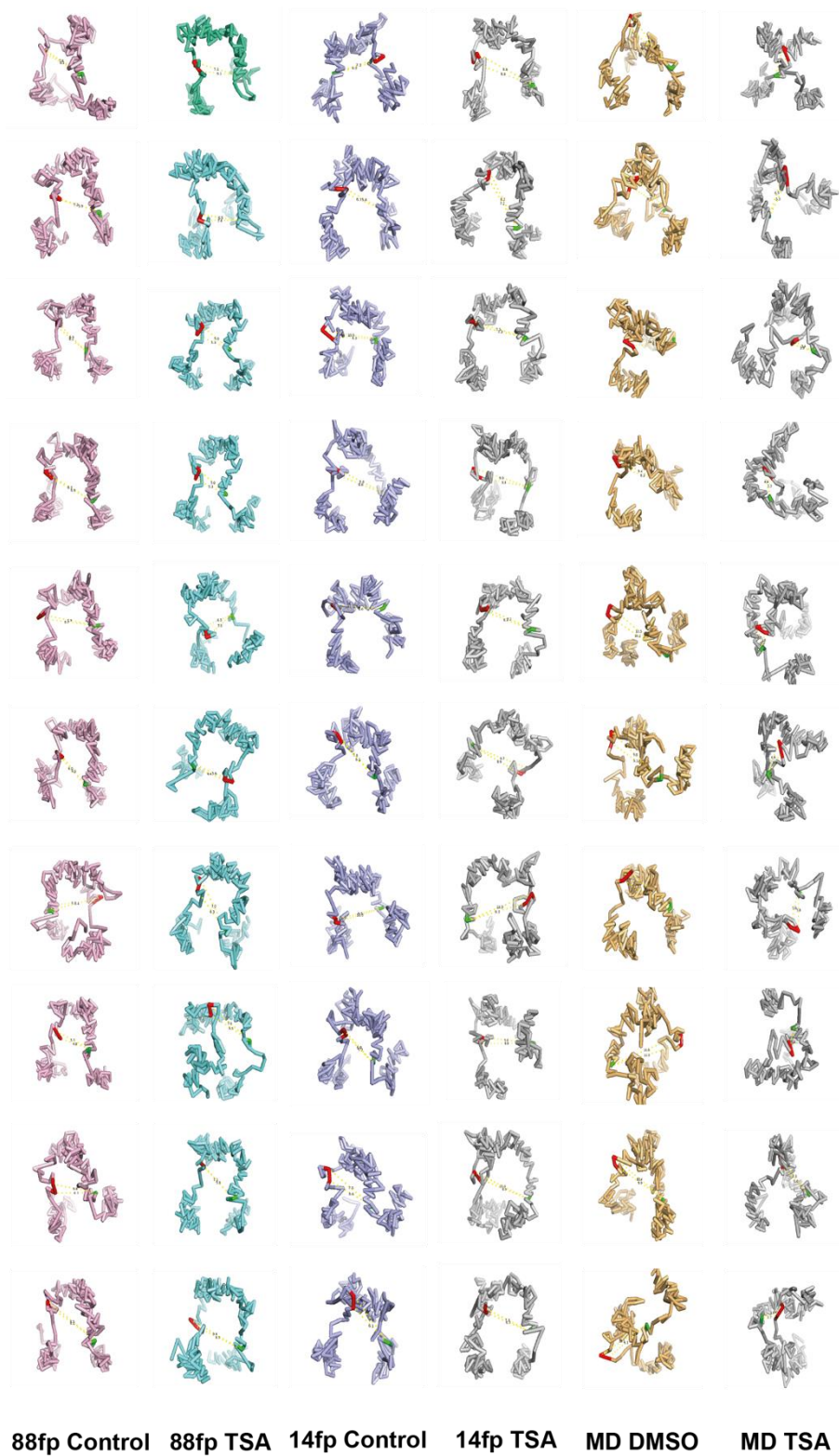


Figure 5.9 – 5C models of TSA treated and control 88fp, 14fp and MD cells by the online modelling tool Autochrom3D. *Shh* is located across 2 beads and is highlighted in red, while the ZRS is highlighted in green. The distance from either of the *Shh* beads to the ZRS is depicted by a dotted yellow line.



**Figure 5.10 – Simulations were repeated 10 times each for TSA treated and control 88fp, 14fp and MD cell lines.**

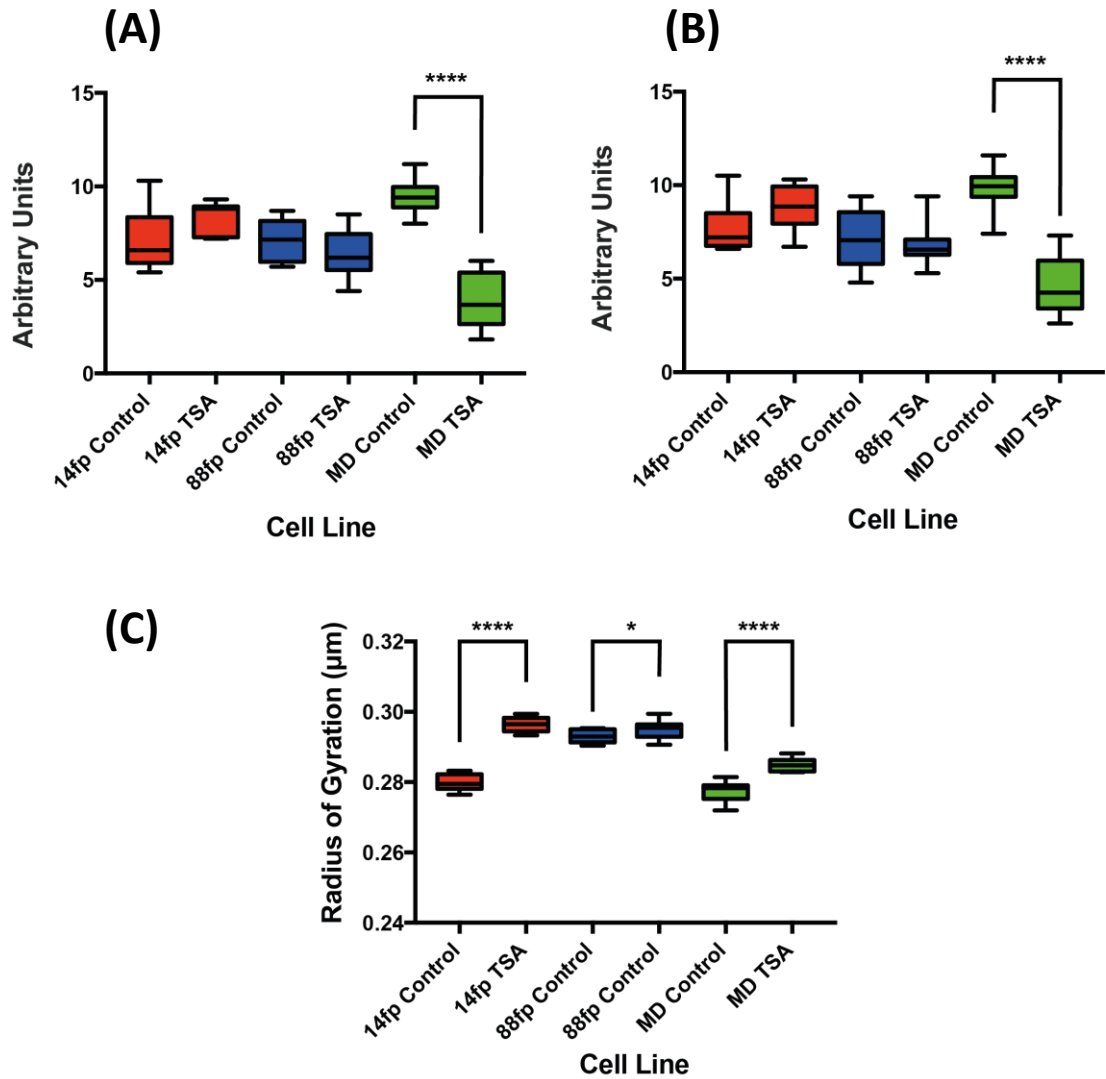


Figure 5.11 – The distance between (A) the first *Shh* bead and the ZRS and (B) the second *Shh* bead and the ZRS in TSA treated and control 14fp, 88fp and MD cells was determined for each of the 10 simulations and plotted as a boxplot. Significant differences in the distance between these points was determined by Mann-Whitney U tests (\* =  $p \leq 0.05$ , \*\*\*\* =  $p \leq 0.0001$ ). (C) Relative compaction of each simulation was determined as a function of the radius of gyration ( $\mu\text{m}$ ) with measurements for each of the simulations plotted as mentioned above. For each cell there was a significant decrease in compaction upon TSA treatment as indicated by an increase in the radius of gyration.



### 5.3.7 Preliminary 4C data suggests a specific interaction at the *Shh* gene of TSA treated 14fp cells

4C (also referred to as 3C-seq) was conducted according to the procedure published by the Soler laboratory and detailed in the Methods (Peng et al., 2013, Stadhouders et al., 2013). The aim of these experiments was to study ZRS and *Shh* interactions within the *Shh* regulatory domain using a technique with a higher resolution than 5C. (i.e. It is not always possible to design primers to the ends of each restriction fragment used in 5C experiments. As a result, 4C experiments are normally of a higher resolution).

The bait sequence used was the 1.7 kb HindIII fragment containing the ZRS. Cells were fixed in 2 % formaldehyde and the primary digestion conducted using Hind III. After the first ligation, 3C libraries were digested with MluCI followed by an another ligation to produce 4C libraries. PCR primers were designed within the bait sequence before the restriction sites and thereby amplify a region of known sequence before progressing into unknown sequence. This procedure was repeated with 14fp cells treated with TSA for both 24 hr and 18 hr and for DMSO treated control cells as well. After amplification and purification 4C libraries were sequenced using the in-house Ion Proton™ (Ion Torrent) system. To our knowledge this is the first time Ion Torrent technology has been used to sequence 4C libraires. The number of reads obtained for each 4C sample are given in Table 5.3.

4C analysis was conducted in collaboration with the Semple laboratory and this analysis has also been described in the thesis provided by PhD student Ben Moore (unpublished). In brief, reads containing the end of the bait sequence reading through the second restriction site into the unknown sequence were isolated and filtered so that only those with high quality score ( $Q > 30$ ) were maintained. These were aligned to the mm9 genome and the BAM files used as an input for the r3Cseq package which was used to normalise the data and identify significant *cis* and *trans* interactions.

The results from 4C experiments are provided in Figure 5.12. In 14fp cells treated with TSA for 18 hr, there was a significant interaction at the *Shh* gene (FDR q-value  $< 5 \times 10^{-10}$ ) which was not observed in the control cells. This interaction is also observed in 14fp cells treated for 24 hr (FDR q-value  $< 5 \times 10^{-10}$ ). However, in control cells, the adjacent 3' HindIII fragment shows a significant interaction (FDR q-value  $< 5 \times 10^{-5}$ ) which is not observed in the treated cells. Cells treated for 18hr, and to a lesser extent those treated for 24 hr, showed a decreased number of significant interactions across the locus. We rationalize that TSA treatment may

cause a reduction in the ZRS interactions across the *Shh* regulatory region with the enhancer now highly specific for the *Shh* gene, a mechanism which may lead to SHH production. As this technique involved the use of Ion Torrent sequencing instead of Illumina sequencing which is commonly used in other publications, we asked our collaborators to assess the quality of our data. Both sequence duplication levels and the percentage of *cis* interactions were analysed. Sequence duplication levels ranged from 62.8 % to 84.4 % and therefore fall below the 95 % expected as described by Soler laboratory. There was also a significant reduction in *cis* interactions below the expected 50 %, ranging from 7 – 20 % across the three samples. However, as expected the highest number of bait specific interactions occurred along chromosome 5 (in *cis*), with other chromosomes (in *trans*) showing less bait specific interactions. Both analyses suggest a low signal-to-noise ratio within the data where non-specific interactions have also been examined. This may be because 4C libraries were sequenced from the side of the second restriction digest site and not from the primary restriction site as described in the protocol. This was conducted because the HindIII fragment used for 4C is extremely AT rich and suitable PCR primers could only be designed at this specific site. Sequencing from this end of the library is likely to have increased background noise. Repeat experiments are currently ongoing with biological replicates, performing both Ion Torrent and Illumina sequencing to examine how the output from both procedures compare and to increase the quality of the data.

| 4C Sample     | No. Reads (millions) |
|---------------|----------------------|
| 14fp Control  | 10                   |
| 14fp TSA 18hr | 8.8                  |
| 14fp TSA 24hr | 24.2                 |

**Table 5.3 – Number of reads obtained for each 4C sample.**

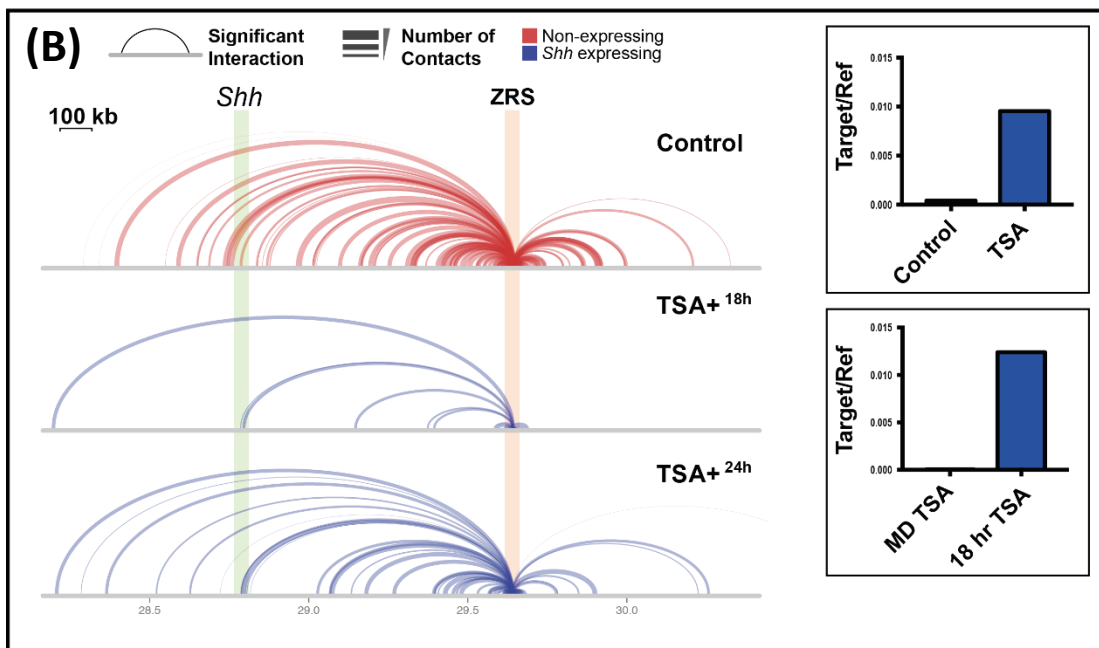
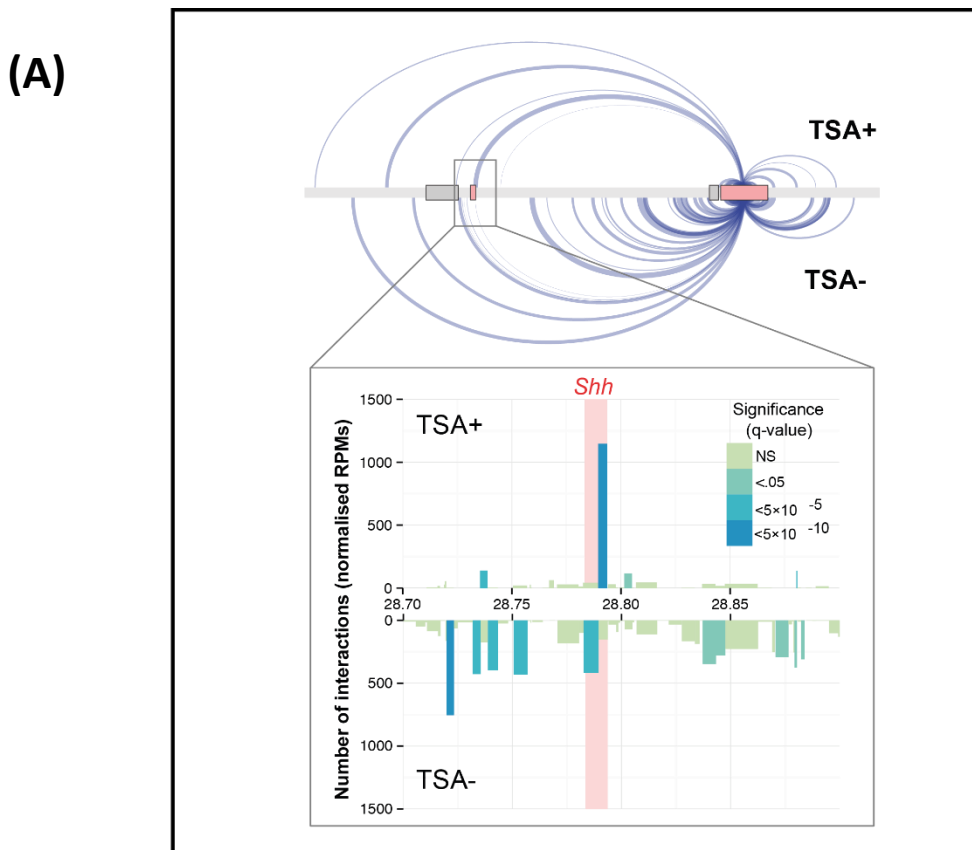


Figure 5.12 – Figures provided by Ben Moore (PhD Student, Semple laboratory). (A) Differences in ZRS interactions at the *Shh* gene upon TSA treatment (18 hr) in 14fp cells. *Lmbr1* (containing the ZRS) is depicted by the large pink rectangle and *Shh* as the small pink rectangle. Significant interactions between the ZRS and other chromatin regions are shown as purple arches (FDR q-value  $< 5 \times 10^{-5}$ ). The zoomed in image shows ZRS interactions

within a 200 kb region containing the *Shh* gene. The width of each bar represents the size of a HindIII restriction fragment. (B) Significant interactions of the ZRS with other chromatin regions in 14fp cells treated with TSA for 18 hr and 24 hr as well as untreated control cells. Significant interactions between the ZRS and other chromatin regions are shown as red (control cells) and purple (TSA treated cells) arches. Expression of *Shh* in TSA treated 14fp cells was checked by qPCR.

## 5.4 Discussion

5C experiments were conducted on TSA treated and control 14fp, 88fp and MD cell lines. In these experiments, the *Shh* gene was contained within a topological associated domain (TAD), the boundaries of which remained constant across all cell lines. These boundary regions correlate with five E14 ChIP-seq CTCF protein binding peaks (Figure 5.3), thereby suggesting a role of CTCF in establishing the TAD (the role of the CTCF protein binding to these sites to create a compact region is discussed in Chapter 4). Some studies have used polymer based models to study chromatin structure where proteins such as CTCF bind to a designated binding site and make contact with another CTCF protein to create a loop (Brackley et al., 2016). These models are then compared to chromatin conformation capture data to determine their validity. In some cases, interactions from the model replicate interactions from the experimental data to a high degree of accuracy thereby confirming the importance of CTCF in establishing the chromatin structure. Recently collaborative work has begun with the Marenduzzo laboratory to determine if some of these CTCF polymer models can replicate interactions seen in our 5C data.

A significant interaction was detected between *Shh* and the ZRS in each cell line regardless of *Shh* expression. This agrees with the recently published 5C data showing that the *Shh* gene and ZRS are always in close proximity in different tissues (Williamson et al., 2016). However, chromatin models using the online software Autochrom3D suggested that in each cell line the chromatin became less compact on TSA treatment (there was a significant increase in the radius of gyration upon TSA treatment in each cell line). This is at odds with previous FISH data using MYtags which suggested that TSA does not affect compaction of this region. However, previous investigations have shown that TSA treatment in other cell lines causes a reversible decompaction of chromatin (Tóth et al., 2004). In this investigation the decrease in chromatin compaction upon TSA treatment depicted by the models may be a result of the parameters selected for modelling. Going forward it may be beneficial to use an additional modelling program and compare the results to those obtained using Autochrom3D.

Replicate experiments need to be repeated for the 4C experiments before concrete conclusions can be deduced. However, it is interesting that in both 14fp cells treated for either 18 hr and 24 hr with TSA, there is a decrease in significant interactions across the *Shh* regulatory region compared to the control untreated cells. It would be intriguing if one of the effects of TSA was to increase the specificity of the ZRS for the *Shh* gene, resulting in a decrease in non-specific interactions across the locus. However, as stated, more experiments need to be

conducted to confirm this hypothesis. On comparison of both 4C and 5C experiments, it is obvious that the ZRS contacts regions close to, or at, the *Shh* gene in both expressing and non-expressing cells. 4C experiments in *Shh* expressing cells do show a specific interaction at the ZRS which is not significant in non-expressing cells (Figure 5.12). This interaction is not obvious from 5C experiments. 4C experiments also show interactions between the ZRS and regions passed the *Shh* gene and into the adjacent TAD, in all cells regardless of *Shh* expression. These significant ZRS interactions were not detected by 5C. To resolve the differences between 5C and 4C, all 4C experiments must be completed in duplicate. If the discrepancies still remain FISH experiments could be conducted to determine if specific ZRS interactions identified by either 4C or 5C are real.

## 5.5 Conclusions

In summary, 5C experiments show that there is a significant interaction between the ZRS and *Shh* gene in both *Shh* expressing and non-expressing cells suggesting that both regions are always in close contact. Modelling of the 5C data using the software Autochrom3D suggests that on the addition of TSA the *Shh* locus becomes less compact. This contrasts with FISH data (Chapter 4) which shows *Shh*-ZRS co-localisation is unaffected by the addition of TSA. Preliminary 4C data suggests that interactions the ZRS makes within the *Shh* regulatory region are altered on the addition of TSA. Further experiments are underway to confirm these findings.

# Chapter 6

---

## Discussion



## 6.1 14fp Cells - a model system to study *Shh* limb bud expression

In this investigation 14fp cells were used – cells derived from the posterior portion of an E11.5 limb bud (which incorporates the ZPA where *Shh* is expressed) – to study the mechanism of *Shh* expression within the limb. Preliminary PCR experiments (Silvia Peluso, manuscript in preparation) showed that the 14fp cells retained expression of some genes expressed within the posterior limb such as *Hand2* and *Gremlin* but *Shh* expression was not maintained (Chapter 3.1). However, the addition of trichostatin A (TSA), an inhibitor of class I and II HDACs, to these cells resulted in the production of SHH with maximum expression obtained 24 hr after addition of TSA. Furthermore, upon TSA treatment the *Shh* gene appeared to be activated by the limb bud specific enhancer the ZRS (zone of polarizing activity regulatory sequence). In untreated cells the ZRS contained the histone marks H3K4me1 but not H3K27ac. In other investigations enhancers displaying such modifications have been described as poised. However, on addition of TSA there was an accumulation of H3K27ac at the ZRS in addition to the H3K4me1 already present (Silvia Peluso, manuscript in preparation) – enhancers with such histone marks are considered active (Creighton et al., 2010). Therefore, it seems that in 14fp cells addition of TSA changes the activity of the ZRS from a poised to active state which subsequently activates the *Shh* gene.

In order to examine how closely the cultured 14fp cells resemble the cells within an E11.5 limb bud and, to determine the effects of TSA on gene expression within 14fp cells, RNA sequencing experiments were conducted on the following samples: TSA treated and untreated 14fp cells; and, cells dissected from the proximal and distal regions of an E11.5 limb bud. Although there were obvious differences in the expression of many genes, 14fp cells maintained the expression of some key developmental genes critical to patterning the E11.5 limb bud. For example, the ETS factors - ETS1, GABP $\alpha$ , ETV4 and ETV5 - were all present in the untreated 14fp cells. These are extremely important in restricting *Shh* expression to the ZPA. *Gremlin* and *Hand2* were also expressed and showed a reduction in expression upon TSA treatment - confirming PCR experiments. Most importantly *Shh* was absent in the untreated cell line but was induced upon TSA treatment. Although *Shh* expression is low, immunofluorescence and qPCR experiments have confirmed that *Shh* is induced upon TSA treatment (Chapters 4 -5 and Silvia Peluso, manuscript in preparation).

The addition of TSA and other HDAC inhibitors to cultured cell lines has previously been reported to have wide ranging effects on genes expressed in many different pathways. In cancer cell lines, TSA affects cell cycle progression and therefore slows growth rate and increases apoptosis. In the 14fp cells, addition of TSA also altered the expression of cell cycle components such as *Cdk4* and *Cdk6* and genes involved in apoptotic pathways such as *Bcl2l11* and *Bcl2l1*. Therefore, for the 14fp cells it was clear that TSA addition affected multiple signaling pathways and was not specific to regulating the expression of genes involved in limb development. For this reason, induction of *Shh* expression by both increasing GABP $\alpha$  and reducing PEA3 levels was also examined in this investigation (discussed later).

## 6.2 Model of TSA induced *Shh* expression in 14fp cells

The mechanism by which TSA induces *Shh* expression has not yet been fully elucidated. However, several experiments have identified key steps in this process. Silvia Peluso (manuscript in preparation) has shown that knock down of *Pea3* or an increase in GABP $\alpha$  levels can induce *Shh* expression in 14fp cells. Indeed, in this investigation knock down of *Pea3* was used to induce *Shh* expression in FISH experiments (Chapter 4). Co-immunoprecipitation experiments also identified PEA3 as a binding partner for HDAC2 and GABP $\alpha$  as a binding partner for P300 (a histone acetyltransferase). Within the ZRS there are five binding sites for ETS1/GABP $\alpha$  and two binding sites for ETV4(PEA3)/ETV5 (Lettice et al., 2012) and ChIP experiments have shown that PEA3, HDAC2, GABP $\alpha$  and P300 are enriched at the ZRS (Silvia Peluso, manuscript in preparation) in 14fp cells. Therefore, in a non-induced state I propose that the ZRS binds PEA3 which recruits HDAC2 and GABP $\alpha$  which recruits P300.

As mentioned previously, TSA is a histone deacetylase inhibitor and, therefore, on addition of TSA to 14fp cells, HDAC2 is inhibited. It is hypothesized that this could have two possible consequences: either HDAC2 inhibition causes loss of PEA3 binding to the ZRS resulting in an increase in the ratio of positive: negative ETS regulators and an increase in *Shh* expression; or, HDAC2 inhibition disrupts binding between HDAC2 and PEA3. In the latter event, PEA3 remains bound to the ZRS without HDAC2. In other cell lines PEA3 can be acetylated by P300 and act as a positive regulator of gene expression (Guo et al., 2011). Therefore, in 14fp cells a loss of binding between PEA3 and HDAC2 could allow PEA3 to be acetylated by P300

thereby shifting the function of PEA3 from a negative regulator of *Shh* expression to a positive regulator. Human PEA3 has been shown to interact with MED25, a component of the mediator complex which is involved in transcriptional regulation (Verger et al., 2013). To identify whether a similar interaction between these two proteins occurs in the mouse limb bud, ChIP-seq with a MED25 antibody in TSA treated and control 14fp cells could be attempted.

ChIP experiments (Silvia Peluso manuscript in preparation) in 14fp cells confirm a reduction in PEA3 levels at the ZRS on TSA addition. This result argues in favor of the first mechanism in which HDAC2 inhibition prevents binding of PEA3 to the ZRS. However, the RNA sequencing experiments (Chapter 3) showed that TSA reduced *Pea3* expression levels, so it is also possible that the reduction of PEA3 bound to the ZRS on TSA treatment is due to the reduced levels of available PEA3 protein. A combination of all these mechanisms may be occurring upon TSA addition, leading to *Shh* expression.

The PRC2 complex is involved in repressing gene expression with studies in humans (Boyer et al., 2006) showing that the PRC2 components EZH2/EED interact with HDAC2 at these repressed sites (van der Vlag and Otte, 1999). It would be interesting to determine if EZH2/EED binds to HDAC2 in the 14fp cell line before and after TSA treatment. It is hypothesized that in the absence of TSA, HDAC2 binds to EZH2/EED with the PRC2 complex repressing *Shh* expression. Upon TSA treatment, inhibition of HDAC2 prevents the association of EZH2/EED with the ZRS, thereby relieving PRC2 induced *Shh* gene repression. In prostate cancer cell lines, the addition of TSA inhibited EZH2 regulated gene repression (Varambally et al., 2002). Therefore, in the 14fp cells, this hypothesis could be tested by performing ChIP-seq experiments using antibodies against EED and EZH2 in TSA treated and control 14fp cells.

### **6.3 Compaction of the *Shh* regulatory region is unaffected by *Shh* expression**

To study compaction of the *Shh* locus 3D-FISH was conducted on TSA treated and control 14fp, 88fp and MD cell lines. The 14fp and 88fp cells both express *Shh* on TSA treatment while *Shh* expression is not detected in MD cells – the MD cells therefore act as a suitable control cell line. In FISH experiments, no significant difference was observed in the percentage of co-localised *Shh* and ZRS probes across all cell lines independent of whether

*Shh* was being expressed or not. This was a surprise, as recent publications have shown that there is a significant increase in co-localisation between the ZRS and *Shh* in the ZPA of the mouse E11.5 limb bud where *Shh* is expressed (Williamson et al., 2016). This suggests a mechanism in which the ZRS moves into contact with the *Shh* gene to regulate *Shh* expression of cells within the ZPA. However, in the 14fp cell line the expression of *Shh* upon TSA treatment is low compared to *Shh* expression within an E11.5 limb bud (see qPCR control experiments – Chapters 4 and 5). It is possible that the effect of TSA addition to the cell lines cannot be detected by FISH experiments. Alternatively, the mechanism of *Shh* induction in the 14fp and 88fp cell lines is different to that of cells in the ZPA (Chapter 6.2). To study compaction of the *Shh* locus in more detail, fluorescently labelled MYtags were used with one library used to tile the region between *Shh* and the ZRS and another to tile the region between *Shh* and *Dpp6*. The *Shh*-ZRS region was much more compact than the *Shh*-*Dpp6* region but no significant difference in compaction was observed between the *Shh*-ZRS regions in TSA treated and control 14fp cells (compaction was determined by counting MYtag spot number and by spot area).

Together these observations suggest that compaction of the *Shh* locus within the cell lines does not change on addition of TSA. This is interesting as previous investigations in other cell lines have shown that TSA can cause reversible decompaction of large chromatin regions (Tóth et al., 2004). The fact that compaction of this locus is unaffected by TSA treatment suggests that there are intrinsic mechanisms that cause compaction and these forces are unaffected by the addition of TSA. It was hypothesized that the binding of the protein CTCF to sites located near both the *Shh* gene and the ZRS may hold the locus in a tightly compact position. Bound CTCF proteins would interact with each other bringing the ZRS into contact with *Shh*. To test this hypothesis Laura Lettice and Iain Williamson designed and constructed ES cells using the CRISPR/Cas9 system, which singly deleted each of the CTCF binding sites located next to *Shh* or the ZRS. In total five mutant lines were created, each with a different CTCF binding site mutation. Laura and Iain have since gone on to use these cell lines for other experiments (manuscript in preparation). In this investigation, the two MYtag libraries were used to assess compaction in both wild type (WT) ES cells and the CTCF binding site mutants. Like the 14fp cells, the *Shh*-*Dpp6* region was much less compact than the *Shh*-ZRS region in the WT and mutant cells. There was no significant difference in the number of spots for the *Shh*-ZRS region between WT and mutant cells, suggesting that compaction of the locus does not change when single CTCF binding sites are removed. However, on analysis of spot area there was in

increase in compaction of the *Shh*-ZRS region in the mutants compared to wild type cells. The reason for the contrasting results between both analyses is unclear.

The fact that no change in compaction was observed when analyzing spot number for the *Shh*-ZRS region in the ES cell mutants was surprising as I believe that CTCF binding is integral for maintaining the region in a compact state. Using MYtags to tile the *Shh* locus in 14fp cells and ES cells is an innovative and a potentially very informative tool to look at compaction but in this investigation there were a number of caveats. Firstly, although all MYtag spots were observed for the *Shh-Dpp6* library, only a maximum of 6 spots (out of a possible 8 spots) were observed for the *Shh*-ZRS library at any one time. Therefore, it was unclear whether the compact nature of the chromatin always prevented the final two spots from being visible or whether the MYtags were not fully annealing in the FISH experiments. In order to confirm that the latter scenario is not occurring, the MYtag libraries need to be tested on decondensed chromatin which should show all 8 spots for the *Shh*-ZRS region if all MYtags are binding properly. Also, all MYtag experiments were conducted using a 2D-FISH protocol as it was found that the intensity of the fluorescent spots was greater when cells were treated in this way compared to 3D-FISH experiments. However, this meant images could not be taken through three dimensions (i.e. images through separate z-stacks were not possible) which may mean the spot number analysis may be less accurate. Probably the biggest limitation was that using only one MYtag library to tile a region meant that no positional information was provided in these experiments i.e. for the the *Shh-Dpp6* region 8 spots were sometimes visible for a single locus but the exact position at which each spot was bound to the DNA was unknown. Originally we had planned to hybridize fosmid probes for the *Shh* gene and the ZRS along with MYtags in the same FISH experiment; however, this has yet to be optimized. FISH experiments using fosmid probes on the ES mutants are currently being conducted (Iain Williamson and Laura Lettice, manuscript in preparation).

## 6.4 Interactions within the *Shh* regulatory region

To look at interactions within the *Shh* regulatory region upon *Shh* expression, 5C was conducted in 14fp and 88fp cell lines while MD cells were used as a control. In all cell lines, regardless of *Shh* expression, an increased interaction frequency was observed for the *Shh* gene with the ZRS. This suggests that within the cells *Shh* and the ZRS are in close proximity even when the gene is not being expressed. Heat maps for all the cell lines confirmed that *Shh* is contained within a topological associated domain (TAD) which also contains the genes *Rnf32* and part of *Lmbr1*. 4C provided an additional way of analyzing ZRS interactions at a higher

resolution with preliminary experiments suggesting that there were less significant interactions across the locus on TSA treatment. Further experiments need to be conducted to confirm this result.

## 6.5 Future Work

In the immediate future replicates for the 5C experiments, TSA treated 14fp cells; and control MD cells are needed before publication. Also, replicates are needed for the 14fp cell 4C experiments. 4C should also be conducted on the MD cell line which would act as a suitable control. Towards the end of this investigation, both 5C and FISH experiments were planned using the *Shh*gfpcre mouse (<https://www.jax.org/strain/005622>). This mouse expresses enhanced green fluorescent protein (EGFP) in all regions where *Shh* is also expressed. Initial experiments had begun where the green fluorescent cells from an E11.5 limb bud (visible in the ZPA) were separated from the non-fluorescent cells by fluorescent associated cell sorting (FACs). In these experiments, cells were first fixed in formaldehyde before FACs and then lysed and stored at -80 °C until needed. The plan was to combine all GFP positive nuclei into one sample and then perform 5C. As a control, the GFP negative nuclei could also be combined and used for 5C. The aim of these experiments was to determine if there were significant interaction differences between cells both within and outside the ZPA in E11.5 limb buds. FISH experiments on both GFP positive and negative cells could also be used to identify the distance between the ZRS and *Shh* gene in both *Shh* expressing and non-expressing limb tissue. However, as mentioned previously the recent paper by Williamson and Lettice (Williamson et al., 2016) performed FISH experiments on both anterior and posterior (containing the ZPA) limb sections and these showed that the ZRS and *Shh* gene co-localise more closely in the posterior region. Therefore, the major question being examined by the *Shh*gfpcre mouse FISH experiments has largely been answered.

A number of future FISH experiments are also planned using the MYtags. Firstly, it would be interesting to look at chromatin compaction using mouse tissue. Compaction of the *Shh* locus in the E11.5 limb bud could be compared to other expressing tissue such as the lungs, brain and gut, as well non-expressing tissue such as the mandible and heart. Also, it would be interesting to perform MYtag FISH experiments on E11.5 limb bud sections to examine how compaction changes in either: an anterior to posterior orientation or, proximal to distal orientation i.e. from regions containing the ZPA where *Shh* is expressed to regions not expressing *Shh*.

Lastly, to provide confirmation that the ZRS activates the *Shh* gene in 14fp cells when treated with TSA, experiments could be conducted using CRISPr to delete the ZRS. 14fp cells with a ZRS deletion would not be expected to induce *Shh* expression. The compaction of the *Shh* locus could also be tested using 3D-FISH to examine the effect of a ZRS deletion in 14fp cells.

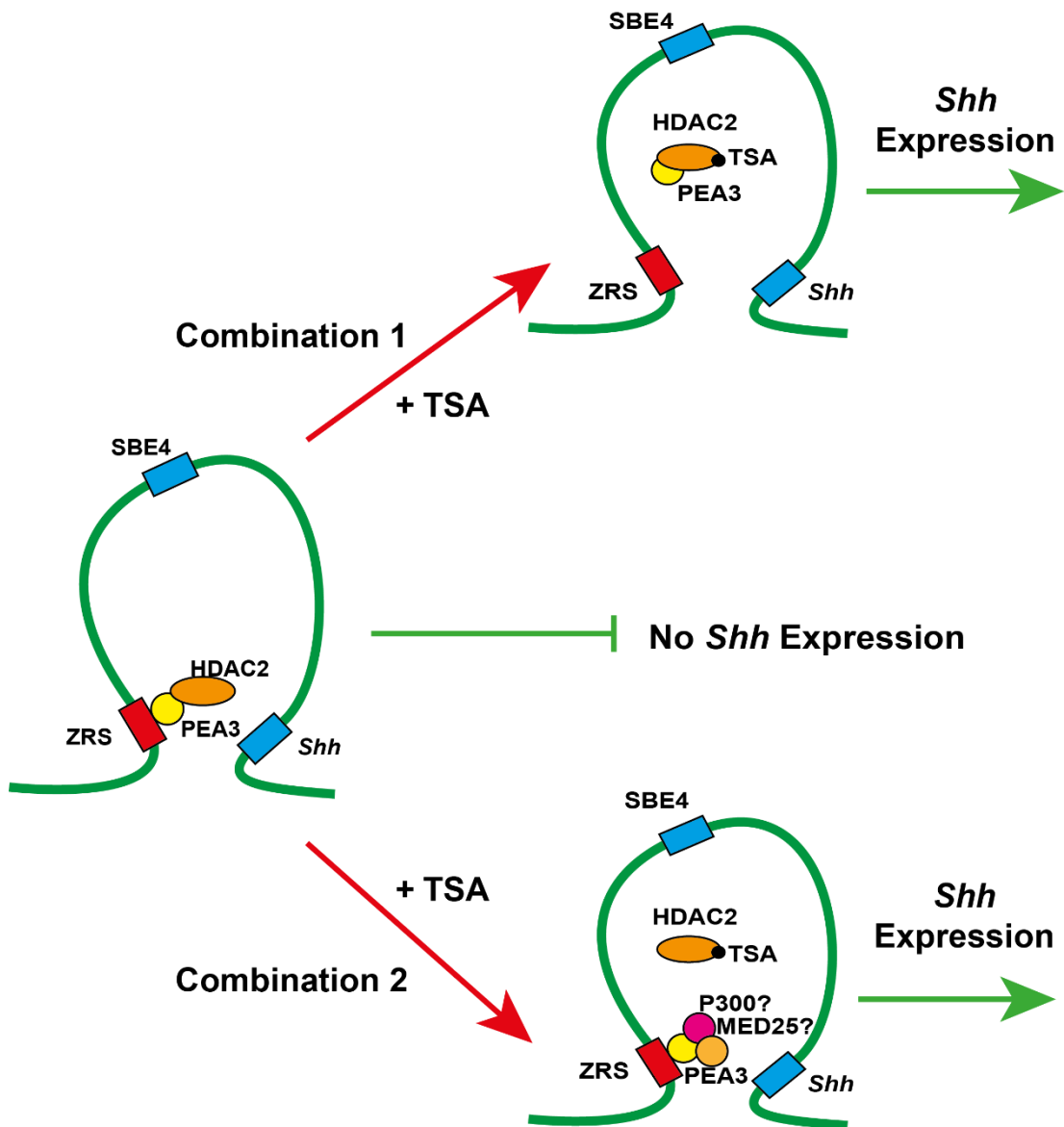


Figure 6.1 – Possible mechanisms of TSA induced *Shh* expression in 14fp cells. In untreated cells, the *Shh* gene and the ZRS are in close proximity, while other enhancers such as SBE4 are located further away from the gene. The ZRS contains the protein PEA3 bound to HDAC2. On addition of TSA there are two possibilities. Firstly, as TSA is an HDAC inhibitor it binds directly to HDAC2 and therefore prevents the binding of negative regulator PEA3 to the ZRS. As a result, there is a shift in the ratio of positive:negative regulators bound to the ZRS resulting in *Shh* expression. Alternatively, TSA treatment inhibits HDAC2 and it is no longer bound at the ZRS but PEA3 binding is unaffected and it remains bound without HDAC2. In this scenario it is possible that the PEA3 protein is itself acetylated by histone acetyltransferases such as P300, thereby changing its role from a negative regulator of *Shh* expression to a positive regulator. In humans MED25 can bind to PEA3. If this occurs in the 14fp cells it may aid inactivation of the *Shh* gene.



# Chapter 7

---

## Appendices

| <b>Parameter</b>                                                                              | <b>Value</b>     |
|-----------------------------------------------------------------------------------------------|------------------|
| <b>Max realign edit distance</b>                                                              | 1000             |
| <b>Max edit distance</b>                                                                      | 2                |
| <b>Library Type</b>                                                                           | FR<br>Unstranded |
| <b>Final read mismatches</b>                                                                  | 2                |
| <b>Use bowtie -n mode</b>                                                                     | No               |
| <b>Anchor length (at least 3)</b>                                                             | 8                |
| <b>Maximum number of mismatches that can appear in the anchor region of spliced alignment</b> | 0                |
| <b>The minimum intron length</b>                                                              | 70               |
| <b>The maximum intron length</b>                                                              | 500,000          |
| <b>Allow indel search</b>                                                                     | Yes              |
| <b>Max insertion length.</b>                                                                  | 3                |
| <b>Max deletion length.</b>                                                                   | 3                |
| <b>Maximum number of alignments to be allowed</b>                                             | 20               |
| <b>Minimum intron length that may be found during split-segment (default) search</b>          | 50               |
| <b>Maximum intron length that may be found during split-segment (default) search</b>          | 500,000          |
| <b>Number of mismatches allowed in each segment alignment for reads mapped independently</b>  | 2                |
| <b>Minimum length of read segments</b>                                                        | 25               |
| <b>Output unmapped reads</b>                                                                  | No               |
| <b>Do you want to supply your own junction data</b>                                           | No               |
| <b>Use coverage-based search for junctions</b>                                                | Auto             |
| <b>Use Microexon Search</b>                                                                   | No               |
| <b>Do Fusion Search</b>                                                                       | No               |
| <b>Set Bowtie2 settings</b>                                                                   | No               |
| <b>Specify read group?</b>                                                                    | No               |

**Table 1 – The Tophat parameters for RNA-seq analysis.**

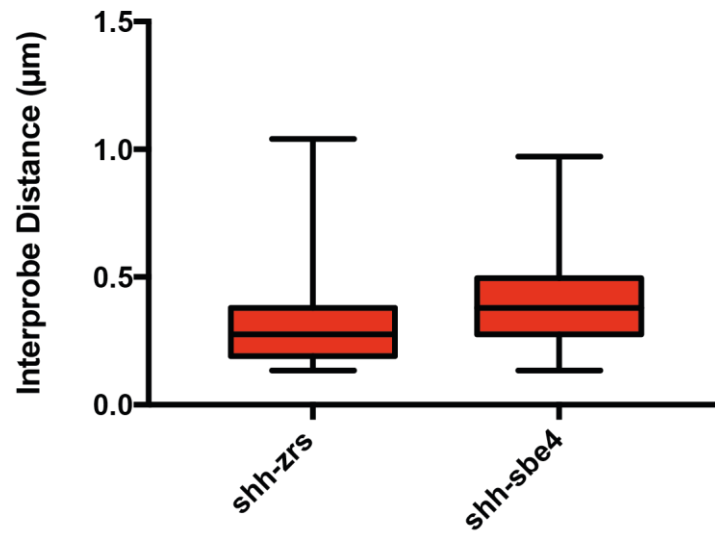


Figure 1 – Preliminary 3D-FISH data shows *Shh*-ZRS probes are more tightly compact than *Shh*-SBE4 probes in 14fp cells suggesting that the *Shh* regulatory region is contained within a loop (p-value < 0.0001, Mann-Whitney U test). Over 100 measurements were recorded for *Shh*-SBE4 and *Shh*-ZRS experiments.

```

# read in pairwise output from my5c
pw <- read.table("~/Desktop/mddmsformodelling.txt", skip=5, sep="\t", header=F)

# parse start and end co-ordinates from fragment IDs
s1 <- as.numeric(gsub(".*?:(\\d+)-.*", "\\1", pw[,1]))
e1 <- as.numeric(gsub(".*-(\\d+)$", "\\1", pw[,1]))
s2 <- as.numeric(gsub(".*?:(\\d+)-.*", "\\1", pw[,3]))
e2 <- as.numeric(gsub(".*-(\\d+)$", "\\1", pw[,3]))

# average start and ends to get midpoint
out <- data.frame(chr1="chr5", p1=as.integer(rowMeans(cbind(s1, e1))),
                 chr2="chr5", p2=as.integer(rowMeans(cbind(s2, e2))),
                 interaction=pw[,2])

# remove NAs
out <- out[complete.cases(out),]

# remove control region (if present)
out <- out[out$p1 < 60e6 & out$p2 < 60e6,]

# write to tab-delimited file
write.table(out, "~/Desktop/mddsmofinal112.txt", sep="\t",
           row.names=F, col.names=F, quote=F)

```

Figure 2 – R script used to convert pairwise file output from my5C into an input format for Autochrom3D (prepared by PhD student Ben Moore, Semple laboratory)

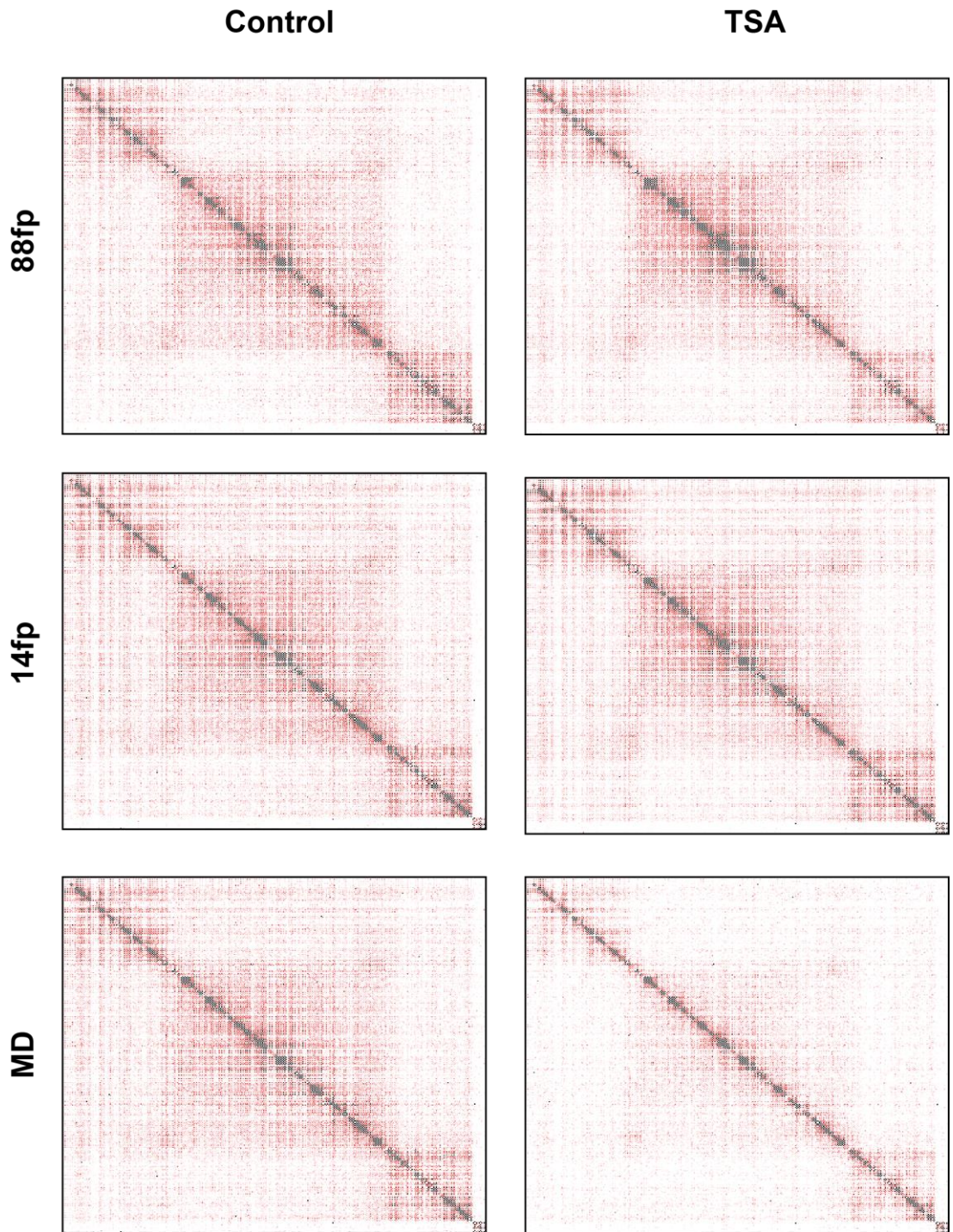
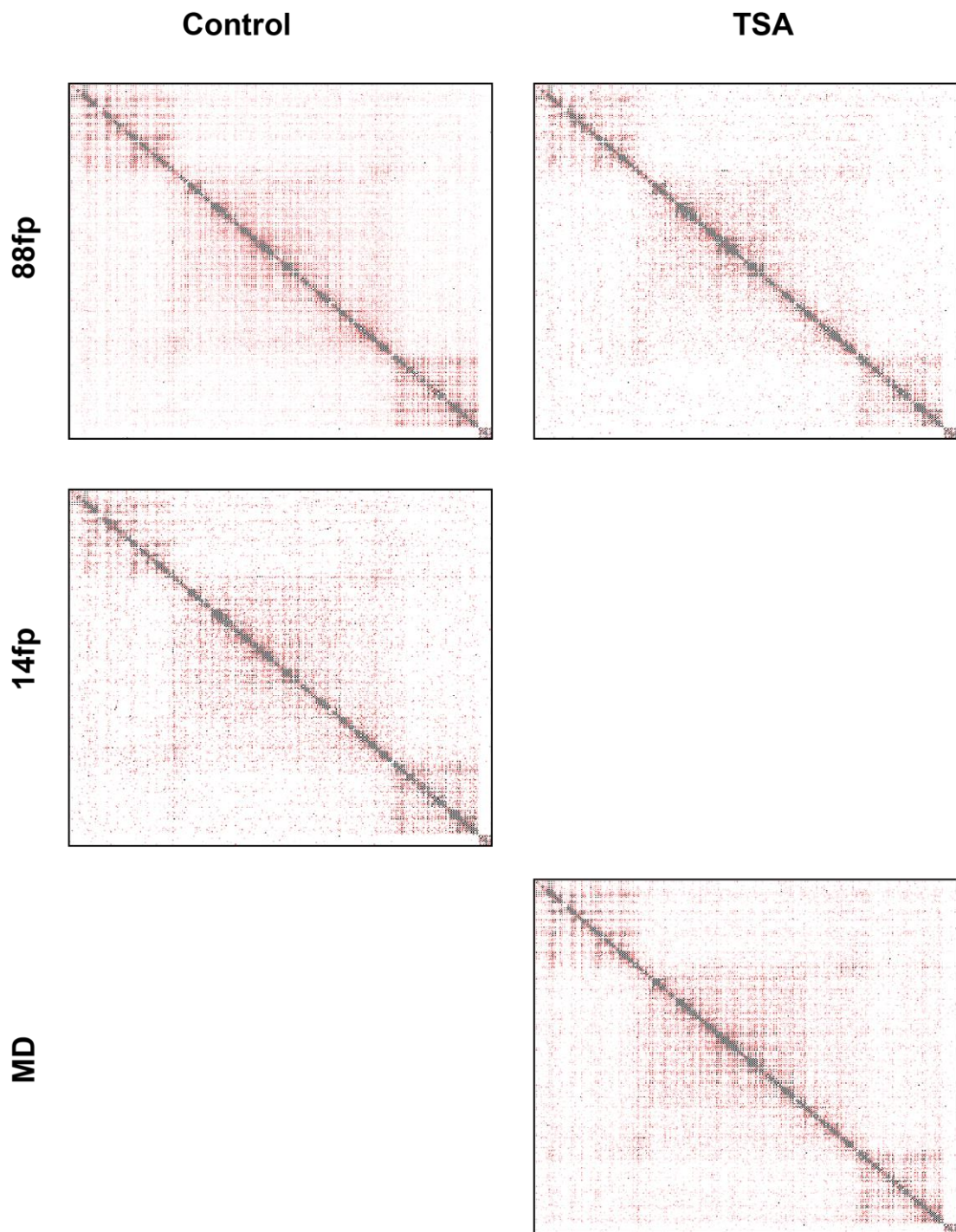


Figure 3 – Unprocessed heat maps of the first biological replicate for TSA treated and control 14fp, 88fp and MD cell lines.



**Figure 4 – Unprocessed heat maps of the second biological replicate for TSA treated and control 14fp, 88fp and MD cell lines. TSA treated 14fp cells and control MD cells were of poor quality and are omitted.**

# Chapter 8

---

## Bibliography

- AFFER, M., CHESI, M., CHEN, W. D., KEATS, J. J., DEMCHENKO, Y. N., TAMIZHMANI, K., GARBITT, V. M., RIGGS, D. L., BRENTS, L. A., ROSCHKE, A. V., VAN WIJER, S., FONSECA, R., BERGSAGEL, P. L. & KUEHL, W. M. 2014. Promiscuous MYC locus rearrangements hijack enhancers but mostly super-enhancers to dysregulate MYC expression in multiple myeloma. *Leukemia*, 28, 1725-35.
- AHITUV, N., ZHU, Y., VISEL, A., HOLT, A., AFZAL, V., PENNACCHIO, L. A. & RUBIN, E. M. 2007. Deletion of ultraconserved elements yields viable mice. *PLoS Biol*, 5, e234.
- AHN, K., MISHINA, Y., HANKS, M. C., BEHRINGER, R. R. & CRENSHAW, E. B. 2001. BMPR-IA signaling is required for the formation of the apical ectodermal ridge and dorsal-ventral patterning of the limb. *Development*, 128, 4449-61.
- AMANAI, K. & JIANG, J. 2001. Distinct roles of Central missing and Dispatched in sending the Hedgehog signal. *Development*, 128, 5119-27.
- AMANO, T., SAGAI, T., TANABE, H., MIZUSHINA, Y., NAKAZAWA, H. & SHIROISHI, T. 2009. Chromosomal dynamics at the Shh locus: limb bud-specific differential regulation of competence and active transcription. *Dev Cell*, 16, 47-57.
- AMENDOLA, M. & VAN STEENSEL, B. 2015. Nuclear lamins are not required for lamina-associated domain organization in mouse embryonic stem cells. *EMBO Rep*, 16, 610-7.
- ANDERSON, E., DEVENNEY, P. S., HILL, R. E. & LETTICE, L. A. 2014. Mapping the Shh long-range regulatory domain. *Development*, 141, 3934-43.
- ANDERSON, E., PELUSO, S., LETTICE, L. A. & HILL, R. E. 2012. Human limb abnormalities caused by disruption of hedgehog signaling. *Trends Genet*, 28, 364-73.
- ANDREY, G., MONTAVON, T., MASCREZ, B., GONZALEZ, F., NOORDERMEER, D., LELEU, M., TRONO, D., SPITZ, F. & DUBOULE, D. 2013. A switch between topological domains underlies HoxD genes collinearity in mouse limbs. *Science*, 340, 1234167.
- ARNOLD, C. D., GERLACH, D., STELZER, C., BORYŃ, Ł., RATH, M. & STARK, A. 2013. Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science*, 339, 1074-7.
- ARNOSTI, D. N. & KULKARNI, M. M. 2005. Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards? *J Cell Biochem*, 94, 890-8.
- AUSIÓ, J. 2015. The shades of gray of the chromatin fiber: recent literature provides new insights into the structure of chromatin. *Bioessays*, 37, 46-51.
- BALCIUNAS, D., DAVIDSON, A. E., SIVASUBBU, S., HERMANSON, S. B., WELLE, Z. & EKKER, S. C. 2004. Enhancer trapping in zebrafish using the Sleeping Beauty transposon. *BMC Genomics*, 5, 62.
- BARROW, J. R., THOMAS, K. R., BOUSSADIA-ZAHUI, O., MOORE, R., KEMLER, R., CAPECCHI, M. R. & MCMAHON, A. P. 2003. Ectodermal Wnt3/beta-catenin signaling is required for the establishment and maintenance of the apical ectodermal ridge. *Genes Dev*, 17, 394-409.
- BAÙ, D. & MARTI-RENOM, M. A. 2011. Structure determination of genomic domains by satisfaction of spatial restraints. *Chromosome Res*, 19, 25-35.
- BEJERANO, G., PHEASANT, M., MAKUNIN, I., STEPHEN, S., KENT, W. J., MATTICK, J. S. & HAUSSLER, D. 2004. Ultraconserved elements in the human genome. *Science*, 304, 1321-5.
- BELL, A. C., WEST, A. G. & FELSENFELD, G. 1999. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell*, 98, 387-96.
- BENKO, S., FANTES, J. A., AMIEL, J., KLEINJAN, D. J., THOMAS, S., RAMSAY, J., JAMSHIDI, N., ESSAFI, A., HEANEY, S., GORDON, C. T., MCBRIDE, D., GOLZIO, C., FISHER, M., PERRY, P., ABADIE, V., AYUSO, C., HOLDER-ESPINASSE, M., KILPATRICK, N., LEES, M. M., PICARD, A., TEMPLE, I. K., THOMAS, P., VAZQUEZ, M. P., VEKEMANS, M., ROEST



- CROLLIUS, H., HASTIE, N. D., MUNNICH, A., ETCHEVERS, H. C., PELET, A., FARLIE, P. G., FITZPATRICK, D. R. & LYONNET, S. 2009. Highly conserved non-coding elements on either side of SOX9 associated with Pierre Robin sequence. *Nat Genet*, 41, 359-64.
- BERLIVET, S., PAQUETTE, D., DUMOUCHEL, A., LANGLAIS, D., DOSTIE, J. & KMITA, M. 2013. Clustering of tissue-specific sub-TADs accompanies the regulation of HoxA genes in developing limbs. *PLoS Genet*, 9, e1004018.
- BHATIA, S., BENGANI, H., FISH, M., BROWN, A., DIVIZIA, M. T., DE MARCO, R., DAMANTE, G., GRAINGER, R., VAN HEYNINGEN, V. & KLEINJAN, D. A. 2013. Disruption of autoregulatory feedback by a mutation in a remote, ultraconserved PAX6 enhancer causes aniridia. *Am J Hum Genet*, 93, 1126-34.
- BIEN-WILLNER, G. A., STANKIEWICZ, P. & LUPSKI, J. R. 2007. SOX9<sup>cre1</sup>, a cis-acting regulatory element located 1.1 Mb upstream of SOX9, mediates its enhancement through the SHH pathway. *Hum Mol Genet*, 16, 1143-56.
- BITGOOD, M. J. & MCMAHON, A. P. 1995. Hedgehog and Bmp genes are coexpressed at many diverse sites of cell-cell interaction in the mouse embryo. *Dev Biol*, 172, 126-38.
- BLOW, M. J., MCCULLEY, D. J., LI, Z., ZHANG, T., AKIYAMA, J. A., HOLT, A., PLAJSER-FRICK, I., SHOUKRY, M., WRIGHT, C., CHEN, F., AFZAL, V., BRISTOW, J., REN, B., BLACK, B. L., RUBIN, E. M., VISEL, A. & PENNACCHIO, L. A. 2010. ChIP-Seq identification of weakly conserved heart enhancers. *Nat Genet*, 42, 806-10.
- BOYER, L. A., PLATH, K., ZEITLINGER, J., BRAMBRINK, T., MEDEIROS, L. A., LEE, T. I., LEVINE, S. S., WERNIG, M., TAJONAR, A., RAY, M. K., BELL, G. W., OTTE, A. P., VIDAL, M., GIFFORD, D. K., YOUNG, R. A. & JAENISCH, R. 2006. Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature*, 441, 349-53.
- BOYLE, S., GILCHRIST, S., BRIDGER, J. M., MAHY, N. L., ELLIS, J. A. & BICKMORE, W. A. 2001. The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. *Hum Mol Genet*, 10, 211-9.
- BRACKLEY, C. A., BROWN, J. M., WAITHE, D., BABBS, C., DAVIES, J., HUGHES, J. R., BUCKLE, V. J. & MARENDUZZO, D. 2016. Predicting the three-dimensional folding of cis-regulatory regions in mammalian genomes using bioinformatic data and polymer models. *Genome Biol*, 17, 59.
- BRISCOE, J. & THÉRON, P. P. 2013. The mechanisms of Hedgehog signalling and its roles in development and disease. *Nat Rev Mol Cell Biol*, 14, 416-29.
- BUENROSTRO, J. D., GIRESI, P. G., ZABA, L. C., CHANG, H. Y. & GREENLEAF, W. J. 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods*, 10, 1213-8.
- BUGLINO, J. A. & RESH, M. D. 2008. What is a palmitoyltransferase with specificity for N-palmitoylation of Sonic Hedgehog. *J Biol Chem*, 283, 22076-88.
- BURKE, R., NELLEN, D., BELLOTTO, M., HAFEN, E., SENTI, K. A., DICKSON, B. J. & BASLER, K. 1999. Dispatched, a novel sterol-sensing domain protein dedicated to the release of cholesterol-modified hedgehog from signaling cells. *Cell*, 99, 803-15.
- CAO, K., CAPELL, B. C., ERDOS, M. R., DJABALI, K. & COLLINS, F. S. 2007. A lamin A protein isoform overexpressed in Hutchinson-Gilford progeria syndrome interferes with mitosis in progeria and normal cells. *Proc Natl Acad Sci U S A*, 104, 4949-54.
- CAPELL, B. C., ERDOS, M. R., MADIGAN, J. P., FIORDALISI, J. J., VARGA, R., CONNEELY, K. N., GORDON, L. B., DER, C. J., COX, A. D. & COLLINS, F. S. 2005. Inhibiting farnesylation of progerin prevents the characteristic nuclear blebbing of Hutchinson-Gilford progeria syndrome. *Proc Natl Acad Sci U S A*, 102, 12879-84.

- CAPRA, J. A., ERWIN, G. D., MCKINSEY, G., RUBENSTEIN, J. L. & POLLARD, K. S. 2013. Many human accelerated regions are developmental enhancers. *Philos Trans R Soc Lond B Biol Sci*, 368, 20130025.
- CARTER, A. J. & WAGNER, G. P. 2002. Evolution of functionally conserved enhancers can be accelerated in large populations: a population-genetic model. *Proc Biol Sci*, 269, 953-60.
- CHAMBEYRON, S. & BICKMORE, W. A. 2004. Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription. *Genes Dev*, 18, 1119-30.
- CHAMOUN, Z., MANN, R. K., NELLEN, D., VON KESSLER, D. P., BELLOTTO, M., BEACHY, P. A. & BASLER, K. 2001. Skinny hedgehog, an acyltransferase required for palmitoylation and activity of the hedgehog signal. *Science*, 293, 2080-4.
- CHOUDHRY, Z., RIKANI, A. A., CHOUDHRY, A. M., TARIQ, S., ZAKARIA, F., ASGHAR, M. W., SARFRAZ, M. K., HAIDER, K., SHAFIQ, A. A. & MOBASSARAH, N. J. 2014. Sonic hedgehog signalling pathway: a complex network. *Ann Neurosci*, 21, 28-31.
- COHEN, M., BRISCOE, J. & BLASSBERG, R. 2013. Morphogen interpretation: the transcriptional logic of neural tube patterning. *Curr Opin Genet Dev*, 23, 423-8.
- COTNEY, J., LENG, J., YIN, J., REILLY, S. K., DEMARE, L. E., EMERA, D., AYOUB, A. E., RAKIC, P. & NOONAN, J. P. 2013. The evolution of lineage-specific regulatory activities in the human embryonic limb. *Cell*, 154, 185-96.
- CREANGA, A., GLENN, T. D., MANN, R. K., SAUNDERS, A. M., TALBOT, W. S. & BEACHY, P. A. 2012. Scube/You activity mediates release of dually lipid-modified Hedgehog signal in soluble form. *Genes Dev*, 26, 1312-25.
- CREMER, M., KÜPPER, K., WAGLER, B., WIZELMAN, L., VON HASE, J., WEILAND, Y., KREJA, L., DIEBOLD, J., SPEICHER, M. R. & CREMER, T. 2003. Inheritance of gene density-related higher order chromatin arrangements in normal and tumor cell nuclei. *J Cell Biol*, 162, 809-20.
- CRETEKOS, C. J., RASWEILER, J. J. & BEHRINGER, R. R. 2001. Comparative studies on limb morphogenesis in mice and bats: a functional genetic approach towards a molecular understanding of diversity in organ formation. *Reprod Fertil Dev*, 13, 691-5.
- CRETEKOS, C. J., WANG, Y., GREEN, E. D., MARTIN, J. F., RASWEILER, J. J. & BEHRINGER, R. R. 2008. Regulatory divergence modifies limb length between mammals. *Genes Dev*, 22, 141-51.
- CREYGHTON, M. P., CHENG, A. W., WELSTEAD, G. G., KOOISTRA, T., CAREY, B. W., STEINE, E. J., HANNA, J., LODATO, M. A., FRAMPTON, G. M., SHARP, P. A., BOYER, L. A., YOUNG, R. A. & JAENISCH, R. 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A*, 107, 21931-6.
- CRICK, F. 1970. Diffusion in embryogenesis. *Nature*, 225, 420-2.
- CUI, K. & ZHAO, K. 2012. Genome-wide approaches to determining nucleosome occupancy in metazoans using MNase-Seq. *Methods Mol Biol*, 833, 413-9.
- DE WIT, E. & DE LAAT, W. 2012. A decade of 3C technologies: insights into nuclear organization. *Genes Dev*, 26, 11-24.
- DEKKER, J., RIPPE, K., DEKKER, M. & KLECKNER, N. 2002. Capturing chromosome conformation. *Science*, 295, 1306-11.
- DETI, A., ROTH, F. P., CHURCH, G. M. & WU, C. T. 2006. Mammalian ultraconserved elements are strongly depleted among segmental duplications and copy number variants. *Nat Genet*, 38, 1216-20.
- DITTMER, T. A. & MISTELI, T. 2011. The lamin protein family. *Genome Biol*, 12, 222.

- DIXON, J. R., SELVARAJ, S., YUE, F., KIM, A., LI, Y., SHEN, Y., HU, M., LIU, J. S. & REN, B. 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, 485, 376-80.
- DOSTIE, J., RICHMOND, T. A., ARNAOUT, R. A., SELZER, R. R., LEE, W. L., HONAN, T. A., RUBIO, E. D., KRUMM, A., LAMB, J., NUSBAUM, C., GREEN, R. D. & DEKKER, J. 2006. Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res*, 16, 1299-309.
- DOUGLAS, A. T. & HILL, R. D. 2014. Variation in vertebrate cis-regulatory elements in evolution and disease. *Transcription*, 5, e28848.
- DRAKE, J. A., BIRD, C., NEMESH, J., THOMAS, D. J., NEWTON-CHEH, C., REYMOND, A., EXCOFFIER, L., ATTAR, H., ANTONARAKIS, S. E., DERMITZAKIS, E. T. & HIRSCHHORN, J. N. 2006. Conserved noncoding sequences are selectively constrained and not mutation cold spots. *Nat Genet*, 38, 223-7.
- DUAN, Z., ANDRONESCU, M., SCHUTZ, K., LEE, C., SHENDURE, J., FIELDS, S., NOBLE, W. S. & ANTHONY BLAU, C. 2012. A genome-wide 3C-method for characterizing the three-dimensional architectures of genomes. *Methods*, 58, 277-88.
- DUDLEY, A. T., ROS, M. A. & TABIN, C. J. 2002. A re-examination of proximodistal patterning during vertebrate limb development. *Nature*, 418, 539-44.
- ECHELARD, Y., EPSTEIN, D. J., ST-JACQUES, B., SHEN, L., MOHLER, J., MCMAHON, J. A. & MCMAHON, A. P. 1993. Sonic hedgehog, a member of a family of putative signaling molecules, is implicated in the regulation of CNS polarity. *Cell*, 75, 1417-30.
- EUGSTER, C., PANÁKOVÁ, D., MAHMOUD, A. & EATON, S. 2007. Lipoprotein-heparan sulfate interactions in the Hh pathway. *Dev Cell*, 13, 57-71.
- FACTOR, D. C., CORRADIN, O., ZENTNER, G. E., SAIKHOVA, A., SONG, L., CHENOWETH, J. G., MCKAY, R. D., CRAWFORD, G. E., SCACHERI, P. C. & TESAR, P. J. 2014. Epigenomic comparison reveals activation of "seed" enhancers during transition from naive to primed pluripotency. *Cell Stem Cell*, 14, 854-63.
- FALLON, J. F., LÓPEZ, A., ROS, M. A., SAVAGE, M. P., OLWIN, B. B. & SIMANDL, B. K. 1994. FGF-2: apical ectodermal ridge growth signal for chick limb development. *Science*, 264, 104-7.
- FARIN, H. F., LÜDTKE, T. H., SCHMIDT, M. K., PLACZKO, S., SCHUSTER-GOSSLER, K., PETRY, M., CHRISTOFFELS, V. M. & KISPERT, A. 2013. Tbx2 terminates shh/fgf signaling in the developing mouse limb bud by direct repression of gremlin1. *PLoS Genet*, 9, e1003467.
- FISHER, S., GRICE, E. A., VINTON, R. M., BESSLING, S. L. & MCCALLION, A. S. 2006. Conservation of RET regulatory function from human to zebrafish without sequence similarity. *Science*, 312, 276-9.
- FLAVAHAN, W. A., DRIER, Y., LIAU, B. B., GILLESPIE, S. M., VENTEICHER, A. S., STEMMER-RACHAMIMOV, A. O., SUVÀ, M. L. & BERNSTEIN, B. E. 2016. Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature*, 529, 110-4.
- FLÖTTMANN, R., WAGNER, J., KOBUS, K., CURRY, C. J., SAVARIRAYAN, R., NISHIMURA, G., YASUI, N., SPRANGER, J., VAN ESCH, H., LYONS, M. J., DUPONT, B. R., DWIVEDI, A., KLOPOCKI, E., HORN, D., MUNDLOS, S. & SPIELMANN, M. 2015. Microdeletions on 6p22.3 are associated with mesomelic dysplasia Savarirayan type. *J Med Genet*, 52, 476-83.
- FUDENBERG, G., IMAKAEV, M., LU, C., GOLOBORODKO, A., ABDENNUR, N. & MIRNY, L. A. 2016. Formation of Chromosomal Domains by Loop Extrusion. *Cell Rep*, 15, 2038-49.

- GALLET, A., RODRIGUEZ, R., RUEL, L. & THEROND, P. P. 2003. Cholesterol modification of hedgehog is required for trafficking and movement, revealing an asymmetric cellular response to hedgehog. *Dev Cell*, 4, 191-204.
- GALLET, A., RUEL, L., STACCINI-LAVENANT, L. & THÉRON, P. P. 2006. Cholesterol modification is necessary for controlled planar long-range activity of Hedgehog in *Drosophila epithelia*. *Development*, 133, 407-18.
- GIORGIO, E., ROBYR, D., SPIELMANN, M., FERRERO, E., DI GREGORIO, E., IMPERIALE, D., VAULA, G., STAMOULIS, G., SANTONI, F., ATZORI, C., GASPARINI, L., FERRERA, D., CANALE, C., GUIPPONI, M., PENNACCHIO, L. A., ANTONARAKIS, S. E., BRUSSINO, A. & BRUSCO, A. 2015. A large genomic deletion leads to enhancer adoption by the lamin B1 gene: a second path to autosomal dominant adult-onset demyelinating leukodystrophy (ADLD). *Hum Mol Genet*, 24, 3143-54.
- GIRESI, P. G., KIM, J., MCDANIELL, R. M., IYER, V. R. & LIEB, J. D. 2007. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res*, 17, 877-85.
- GIRISHA, K. M., BIDCHOL, A. M., KAMATH, P. S., SHAH, K. H., MORTIER, G. R., MUNDLOS, S. & SHAH, H. 2014. A novel mutation (g.106737G>T) in zone of polarizing activity regulatory sequence (ZRS) causes variable limb phenotypes in Werner mesomelia. *Am J Med Genet A*, 164A, 898-906.
- GLASSFORD, W. J. & REBEIZ, M. 2013. Assessing constraints on the path of regulatory sequence evolution. *Philos Trans R Soc Lond B Biol Sci*, 368, 20130026.
- GUELEN, L., PAGIE, L., BRASSET, E., MEULEMAN, W., FAZA, M. B., TALHOUT, W., EUSSEN, B. H., DE KLEIN, A., WESSELS, L., DE LAAT, W. & VAN STEENSEL, B. 2008. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature*, 453, 948-51.
- GUO, B., PANAGIOTAKI, N., WARWOOD, S. & SHARROCKS, A. D. 2011. Dynamic modification of the ETS transcription factor PEA3 by sumoylation and p300-mediated acetylation. *Nucleic Acids Res*, 39, 6403-13.
- GUO, H., DAI, L., HUANG, Y., LIAO, Q. & BAI, Y. 2013. A large novel deletion downstream of PAX6 gene in a Chinese family with ocular coloboma. *PLoS One*, 8, e83073.
- GUO, Y., XU, Q., CANZIO, D., SHOU, J., LI, J., GORKIN, D. U., JUNG, I., WU, H., ZHAI, Y., TANG, Y., LU, Y., WU, Y., JIA, Z., LI, W., ZHANG, M. Q., REN, B., KRAINER, A. R., MANIATIS, T. & WU, Q. 2015. CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell*, 162, 900-10.
- HAGÈGE, H., KLOUS, P., BRAEM, C., SPLINTER, E., DEKKER, J., CATHALA, G., DE LAAT, W. & FORNÉ, T. 2007. Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nat Protoc*, 2, 1722-33.
- HARE, E. E., PETERSON, B. K. & EISEN, M. B. 2008a. A careful look at binding site reorganization in the even-skipped enhancers of *Drosophila* and sepsids. *PLoS Genet*, 4, e1000268.
- HARE, E. E., PETERSON, B. K., IYER, V. N., MEIER, R. & EISEN, M. B. 2008b. Sepsid even-skipped enhancers are functionally conserved in *Drosophila* despite lack of sequence conservation. *PLoS Genet*, 4, e1000106.
- HARFE, B. D., SCHERZ, P. J., NISSIM, S., TIAN, H., MCMAHON, A. P. & TABIN, C. J. 2004. Evidence for an expansion-based temporal Shh gradient in specifying vertebrate digit identities. *Cell*, 118, 517-28.
- HARO, E., DELGADO, I., JUNCO, M., YAMADA, Y., MANSOURI, A., OBERG, K. C. & ROS, M. A. 2014. Sp6 and Sp8 transcription factors control AER formation and dorsal-ventral patterning in limb development. *PLoS Genet*, 10, e1004468.
- HEBROK, M. 2003. Hedgehog signaling in pancreas development. *Mech Dev*, 120, 45-57.

- HEYN, H., VIDAL, E., FERREIRA, H. J., VIZOSO, M., SAYOLS, S., GOMEZ, A., MORAN, S., BOQUE-SASTRE, R., GUIL, S., MARTINEZ-CARDUS, A., LIN, C. Y., ROYO, R., SANCHEZ-MUT, J. V., MARTINEZ, R., GUT, M., TORRENTS, D., OROZCO, M., GUT, I., YOUNG, R. A. & ESTELLER, M. 2016. Epigenomic analysis detects aberrant super-enhancer DNA methylation in human cancer. *Genome Biol*, 17, 11.
- HNISZ, D., ABRAHAM, B. J., LEE, T. I., LAU, A., SAINT-ANDRÉ, V., SIGOVA, A. A., HOKE, H. A. & YOUNG, R. A. 2013. Super-enhancers in the control of cell identity and disease. *Cell*, 155, 934-47.
- HNISZ, D., WEINTRAUB, A. S., DAY, D. S., VALTON, A. L., BAK, R. O., LI, C. H., GOLDMANN, J., LAJOIE, B. R., FAN, Z. P., SIGOVA, A. A., REDDY, J., BORGES-RIVERA, D., LEE, T. I., JAENISCH, R., PORTEUS, M. H., DEKKER, J. & YOUNG, R. A. 2016. Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science*, 351, 1454-8.
- HYON, C., CHANTOT-BASTARAUD, S., HARBUZ, R., BHOURI, R., PERROT, N., PEYCELON, M., SIBONY, M., ROJO, S., PIGUEL, X., BILAN, F., GILBERT-DUSSARDIER, B., KITZIS, A., MCELREAVEY, K., SIFFROI, J. P. & BASHAMBOO, A. 2015. Refining the regulatory region upstream of SOX9 associated with 46,XX testicular disorders of Sex Development (DSD). *Am J Med Genet A*, 167A, 1851-8.
- IANAKIEV, P., VAN BAREN MJ, DALY, M. J., TOLEDO, S. P., CAVALCANTI, M. G., NETO, J. C., SILVEIRA, E. L., FREIRE-MAIA, A., HEUTINK, P., KILPATRICK, M. W. & TSIPOURAS, P. 2001. Acheiropodia is caused by a genomic deletion in C7orf2, the human orthologue of the Lmbr1 gene. *Am J Hum Genet*, 68, 38-45.
- IBORRA, F. J., POMBO, A., JACKSON, D. A. & COOK, P. R. 1996. Active RNA polymerases are localized within discrete transcription 'factories' in human nuclei. *J Cell Sci*, 109 ( Pt 6), 1427-36.
- INGHAM, P. & VICENTE, C. 2014. An interview with Phil Ingham. *Development*, 141, 2363-5.
- JAT, P. S., NOBLE, M. D., ATALIOTIS, P., TANAKA, Y., YANNOOTSOS, N., LARSEN, L. & KIOUSSIS, D. 1991. Direct derivation of conditionally immortal cell lines from an H-2Kb-tsA58 transgenic mouse. *Proc Natl Acad Sci U S A*, 88, 5096-100.
- JEONG, Y., EL-JAICK, K., ROESSLER, E., MUENKE, M. & 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, 761-72.
- KAGEY, M. H., NEWMAN, J. J., BILODEAU, S., ZHAN, Y., ORLANDO, D. A., VAN BERKUM, N. L., EBMEIER, C. C., GOOSSENS, J., RAHL, P. B., LEVINE, S. S., TAATJES, D. J., DEKKER, J. & YOUNG, R. A. 2010. Mediator and cohesin connect gene expression and chromatin architecture. *Nature*, 467, 430-5.
- KATZMAN, S., KERN, A. D., BEJERANO, G., FEWELL, G., FULTON, L., WILSON, R. K., SALAMA, S. R. & HAUSSLER, D. 2007. Human genome ultraconserved elements are ultraselected. *Science*, 317, 915.
- KENGAKU, M., CAPDEVILA, J., RODRIGUEZ-ESTEBAN, C., DE LA PEÑA, J., JOHNSON, R. L., IZPISÚA BELMONTE, J. C. & TABIN, C. J. 1998. Distinct WNT pathways regulating AER formation and dorsoventral polarity in the chick limb bud. *Science*, 280, 1274-7.
- KOKUBU, C., HORIE, K., ABE, K., IKEDA, R., MIZUNO, S., UNO, Y., OGIWARA, S., OHTSUKA, M., ISOTANI, A., OKABE, M., IMAI, K. & TAKEDA, J. 2009. A transposon-based chromosomal engineering method to survey a large cis-regulatory landscape in mice. *Nat Genet*, 41, 946-52.
- KOZHEMYAKINA, E., IONESCU, A. & LASSAR, A. B. 2014. GATA6 is a crucial regulator of Shh in the limb bud. *PLoS Genet*, 10, e1004072.

- KRAUSS, S., CONCORDET, J. P. & 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*, 75, 1431-44.
- KUBBEN, N., ADRIAENS, M., MEULEMAN, W., VONCKEN, J. W., VAN STEENSEL, B. & MISTELI, T. 2012. Mapping of lamin A- and progerin-interacting genome regions. *Chromosoma*, 121, 447-64.
- KUGLER, M. C., JOYNER, A. L., LOOMIS, C. A. & MUNGER, J. S. 2015. Sonic hedgehog signaling in the lung. From development to disease. *Am J Respir Cell Mol Biol*, 52, 1-13.
- KURTH, I., KLOPOCKI, E., STRICKER, S., VAN OOSTERWIJK, J., VANEK, S., ALTMANN, J., SANTOS, H. G., VAN HARSEL, J. J., DE RAVEL, T., WILKIE, A. O., GAL, A. & MUNDLOS, S. 2009. Duplications of noncoding elements 5' of *SOX9* are associated with brachydactyly-anonychia. *Nat Genet*, 41, 862-3.
- LANGMORE, J. P. & PAULSON, J. R. 1983. Low angle x-ray diffraction studies of chromatin structure in vivo and in isolated nuclei and metaphase chromosomes. *J Cell Biol*, 96, 1120-31.
- LEE, J. D. & TREISMAN, J. E. 2001. *Sightless* has homology to transmembrane acyltransferases and is required to generate active Hedgehog protein. *Curr Biol*, 11, 1147-52.
- LEE, J. J., EKKER, S. C., VON KESSLER, D. P., PORTER, J. A., SUN, B. I. & BEACHY, P. A. 1994. Autoproteolysis in hedgehog protein biogenesis. *Science*, 266, 1528-37.
- LEE, T. I., JENNER, R. G., BOYER, L. A., GUENTHER, M. G., LEVINE, S. S., KUMAR, R. M., CHEVALIER, B., JOHNSTONE, S. E., COLE, M. F., ISONO, K., KOSEKI, H., FUCHIKAMI, T., ABE, K., MURRAY, H. L., ZUCKER, J. P., YUAN, B., BELL, G. W., HERBOLSHEIMER, E., HANNETT, N. M., SUN, K., ODOM, D. T., OTTE, A. P., VOLKERT, T. L., BARTEL, D. P., MELTON, D. A., GIFFORD, D. K., JAENISCH, R. & YOUNG, R. A. 2006. Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell*, 125, 301-13.
- LETTICE, L. A., HILL, A. E., DEVENNEY, P. S. & HILL, R. E. 2008. Point mutations in a distant sonic hedgehog cis-regulator generate a variable regulatory output responsible for preaxial polydactyly. *Hum Mol Genet*, 17, 978-85.
- LETTICE, L. A., WILLIAMSON, I., WILTSHIRE, J. H., PELUSO, S., DEVENNEY, P. S., HILL, A. E., ESSAFI, A., HAGMAN, J., MORT, R., GRIMES, G., DEANGELIS, C. L. & HILL, R. E. 2012. Opposing functions of the ETS factor family define *Shh* spatial expression in limb buds and underlie polydactyly. *Dev Cell*, 22, 459-67.
- LEVSKY, J. M. & SINGER, R. H. 2003. Fluorescence in situ hybridization: past, present and future. *J Cell Sci*, 116, 2833-8.
- LEWANDOSKI, M., SUN, X. & MARTIN, G. R. 2000. *Fgf8* signalling from the AER is essential for normal limb development. *Nat Genet*, 26, 460-3.
- LITINGTUNG, Y., LEI, L., WESTPHAL, H. & CHIANG, C. 1998. Sonic hedgehog is essential to foregut development. *Nat Genet*, 20, 58-61.
- LOGAN, C., HORNBRUCH, A., CAMPBELL, I. & LUMSDEN, A. 1997. The role of *Engrailed* in establishing the dorsoventral axis of the chick limb. *Development*, 124, 2317-24.
- LOHAN, S., SPIELMANN, M., DOELKEN, S. C., FLÖTTMANN, R., MUHAMMAD, F., BAIG, S. M., WAJID, M., HÜLSEMANN, W., HABENICHT, R., KJAER, K. W., PATIL, S. J., GIRISHA, K. M., ABARCA-BARRIGA, H. H., MUNDLOS, S. & KLOPOCKI, E. 2014. Microduplications encompassing the Sonic hedgehog limb enhancer ZRS are associated with Haas-type polysyndactyly and Laurin-Sandrow syndrome. *Clin Genet*, 86, 318-25.

- LOVÉN, J., HOKE, H. A., LIN, C. Y., LAU, A., ORLANDO, D. A., VAKOC, C. R., BRADNER, J. E., LEE, T. I. & YOUNG, R. A. 2013. Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell*, 153, 320-34.
- LOVÉN, J., ORLANDO, D. A., SIGOVA, A. A., LIN, C. Y., RAHL, P. B., BURGE, C. B., LEVENS, D. L., LEE, T. I. & YOUNG, R. A. 2012. Revisiting global gene expression analysis. *Cell*, 151, 476-82.
- LUPIÁÑEZ, D. G., KRAFT, K., HEINRICH, V., KRAWITZ, P., BRANCATI, F., KLOPOCKI, E., HORN, D., KAYSERILI, H., OPITZ, J. M., LAXOVA, R., SANTOS-SIMARRO, F., GILBERT-DUSSARDIER, B., WITTLER, L., BORSCHIWER, M., HAAS, S. A., OSTERWALDER, M., FRANKE, M., TIMMERMANN, B., HECHT, J., SPIELMANN, M., VISEL, A. & MUNDLOS, S. 2015. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell*, 161, 1012-25.
- MAESHIMA, K., HIHARA, S. & ELTSOV, M. 2010. Chromatin structure: does the 30-nm fibre exist in vivo? *Curr Opin Cell Biol*, 22, 291-7.
- MAESHIMA, K., IDE, S., HIBINO, K. & SASAI, M. 2016. Liquid-like behavior of chromatin. *Curr Opin Genet Dev*, 37, 36-45.
- MAESHIMA, K., IMAI, R., TAMURA, S. & NOZAKI, T. 2014. Chromatin as dynamic 10-nm fibers. *Chromosoma*, 123, 225-37.
- MARIANI, F. V., AHN, C. P. & MARTIN, G. R. 2008. Genetic evidence that FGFs have an instructive role in limb proximal-distal patterning. *Nature*, 453, 401-5.
- MARINIĆ, M., AKTAS, T., RUF, S. & SPITZ, F. 2013. An integrated holo-enhancer unit defines tissue and gene specificity of the Fgf8 regulatory landscape. *Dev Cell*, 24, 530-42.
- MARSDEN, M. P. & LAEMMLI, U. K. 1979. Metaphase chromosome structure: evidence for a radial loop model. *Cell*, 17, 849-58.
- MARTIN, P. 1990. Tissue patterning in the developing mouse limb. *Int J Dev Biol*, 34, 323-36.
- MATTOUT, A., PIKE, B. L., TOWBIN, B. D., BANK, E. M., GONZALEZ-SANDOVAL, A., STADLER, M. B., MEISTER, P., GRUENBAUM, Y. & GASSER, S. M. 2011. An EDMD mutation in *C. elegans* lamin blocks muscle-specific gene relocation and compromises muscle integrity. *Curr Biol*, 21, 1603-14.
- MERCADER, N., LEONARDO, E., PIEDRA, M. E., MARTÍNEZ-A, C., ROS, M. A. & TORRES, M. 2000. Opposing RA and FGF signals control proximodistal vertebrate limb development through regulation of Meis genes. *Development*, 127, 3961-70.
- MERIDETH, M. A., GORDON, L. B., CLAUSS, S., SACHDEV, V., SMITH, A. C., PERRY, M. B., BREWER, C. C., ZALEWSKI, C., KIM, H. J., SOLOMON, B., BROOKS, B. P., GERBER, L. H., TURNER, M. L., DOMINGO, D. L., HART, T. C., GRAF, J., REYNOLDS, J. C., GROPMAN, A., YANOVSKI, J. A., GERHARD-HERMAN, M., COLLINS, F. S., NABEL, E. G., CANNON, R. O., GAHL, W. A. & INTRONE, W. J. 2008. Phenotype and course of Hutchinson-Gilford progeria syndrome. *N Engl J Med*, 358, 592-604.
- MICCHELLI, C. A., THE, I., SELVA, E., MOGILA, V. & PERRIMON, N. 2002. Rasp, a putative transmembrane acyltransferase, is required for Hedgehog signaling. *Development*, 129, 843-51.
- MICHAUD, J. L., LAPOINTE, F. & LE DOUARIN, N. M. 1997. The dorsoventral polarity of the presumptive limb is determined by signals produced by the somites and by the lateral somatopleure. *Development*, 124, 1453-63.
- MIELE, A., GHELDOF, N., TABUCHI, T. M., DOSTIE, J. & DEKKER, J. 2006. Mapping chromatin interactions by chromosome conformation capture. *Curr Protoc Mol Biol*, Chapter 21, Unit 21.11.
- MILLER, L. A., WERT, S. E. & WHITSETT, J. A. 2001. Immunolocalization of sonic hedgehog (Shh) in developing mouse lung. *J Histochem Cytochem*, 49, 1593-604.

- MITCHELL, J. A. & FRASER, P. 2008. Transcription factories are nuclear subcompartments that remain in the absence of transcription. *Genes Dev*, 22, 20-5.
- MONTAVON, T., SOSHIKOVA, N., MASCREZ, B., JOYE, E., THEVENET, L., SPLINTER, E., DE LAAT, W., SPITZ, F. & DUBOULE, D. 2011. A regulatory archipelago controls Hox genes transcription in digits. *Cell*, 147, 1132-45.
- MOORE, P. S., BARBI, S., DONADELLI, M., COSTANZO, C., BASSI, C., PALMIERI, M. & SCARPA, A. 2004. Gene expression profiling after treatment with the histone deacetylase inhibitor trichostatin A reveals altered expression of both pro- and anti-apoptotic genes in pancreatic adenocarcinoma cells. *Biochim Biophys Acta*, 1693, 167-76.
- MOREY, C., DA SILVA, N. R., PERRY, P. & BICKMORE, W. A. 2007. Nuclear reorganisation and chromatin decondensation are conserved, but distinct, mechanisms linked to Hox gene activation. *Development*, 134, 909-19.
- NANNI, L., MING, J. E., BOCIAN, M., STEINHAUS, K., BIANCHI, D. W., DIE-SMULDERS, C., GIANNOTTI, A., IMAIZUMI, K., JONES, K. L., CAMPO, M. D., MARTIN, R. A., MEINECKE, P., PIERPONT, M. E., ROBIN, N. H., YOUNG, I. D., ROESSLER, E. & MUENKE, M. 1999. The mutational spectrum of the sonic hedgehog gene in holoprosencephaly: SHH mutations cause a significant proportion of autosomal dominant holoprosencephaly. *Hum Mol Genet*, 8, 2479-88.
- NARENDRA, V., ROCHA, P. P., AN, D., RAVIRAM, R., SKOK, J. A., MAZZONI, E. O. & REINBERG, D. 2015. CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. *Science*, 347, 1017-21.
- NISHIMOTO, S. & LOGAN, M. P. 2016. Subdivision of the lateral plate mesoderm and specification of the forelimb and hindlimb forming domains. *Semin Cell Dev Biol*, 49, 102-8.
- NISHIMOTO, S., WILDE, S. M., WOOD, S. & LOGAN, M. P. 2015. RA Acts in a Coherent Feed-Forward Mechanism with Tbx5 to Control Limb Bud Induction and Initiation. *Cell Rep*, 12, 879-91.
- NISHINO, Y., ELTSOV, M., JOTI, Y., ITO, K., TAKATA, H., TAKAHASHI, Y., HIHARA, S., FRANGAKIS, A. S., IMAMOTO, N., ISHIKAWA, T. & MAESHIMA, K. 2012. Human mitotic chromosomes consist predominantly of irregularly folded nucleosome fibres without a 30-nm chromatin structure. *EMBO J*, 31, 1644-53.
- NISWANDER, L., JEFFREY, S., MARTIN, G. R. & TICKLE, C. 1994. A positive feedback loop coordinates growth and patterning in the vertebrate limb. *Nature*, 371, 609-12.
- NISWANDER, L., TICKLE, C., VOGEL, A., BOOTH, I. & MARTIN, G. R. 1993. FGF-4 replaces the apical ectodermal ridge and directs outgrowth and patterning of the limb. *Cell*, 75, 579-87.
- NORBNOP, P., SRICHOMTHONG, C., SUPHAPEETIPORN, K. & SHOTELERSUK, V. 2014. ZRS 406A>G mutation in patients with tibial hypoplasia, polydactyly and triphalangeal first fingers. *J Hum Genet*, 59, 467-70.
- NORTHCOTT, P. A., LEE, C., ZICHER, T., STÜTZ, A. M., ERKEK, S., KAWAUCHI, D., SHIH, D. J., HOVESTADT, V., ZAPATKA, M., STURM, D., JONES, D. T., KOOL, M., REMKE, M., CAVALLI, F. M., ZUYDERDUYN, S., BADER, G. D., VANDENBERG, S., ESPARZA, L. A., RYZHOVA, M., WANG, W., WITTMANN, A., STARK, S., SIEBER, L., SEKERCIN, H., LINKE, L., KRATOCHWIL, F., JÄGER, N., BUCHHALTER, I., IMBUSCH, C. D., ZIPPRICH, G., RAEDER, B., SCHMIDT, S., DIESSL, N., WOLF, S., WIEMANN, S., BRORS, B., LAWRENZ, C., EILS, J., WARNATZ, H. J., RISCH, T., YASPO, M. L., WEBER, U. D., BARTHOLOMAE, C. C., VON KALLE, C., TURÁNYI, E., HAUSER, P., SANDEN, E., DARABI, A., SIESJÖ, P., STERBA, J., ZITTERBART, K., SUMERAUER, D., VAN SLUIS, P., VERSTEEG, R., VOLCKMANN, R., KOSTER, J., SCHUHMANN, M. U., EBINGER, M., GRIMES, H. L., ROBINSON, G. W., GAJJAR, A., MYNAREK, M., VON HOFF, K.,



- RUTKOWSKI, S., PIETSCH, T., SCHEURLIN, W., FELSBURG, J., REIFENBERGER, G., KULOZIK, A. E., VON DEIMLING, A., WITT, O., EILS, R., GILBERTSON, R. J., KORSHUNOV, A., TAYLOR, M. D., LICHTER, P., KORBEL, J. O., WECHSLER-REYA, R. J. & PFISTER, S. M. 2014. Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma. *Nature*, 511, 428-34.
- NÜSSLEIN-VOLHARD, C. & WIESCHAUS, E. 1980. Mutations affecting segment number and polarity in *Drosophila*. *Nature*, 287, 795-801.
- O'KANE, C. J. & GEHRING, W. J. 1987. Detection in situ of genomic regulatory elements in *Drosophila*. *Proc Natl Acad Sci U S A*, 84, 9123-7.
- OHLIG, S., FARSHI, P., PICKHINKE, U., VAN DEN BOOM, J., HÖING, S., JAKUSCHEV, S., HOFFMANN, D., DREIER, R., SCHÖLER, H. R., DIERKER, T., BORDYCH, C. & GROBE, K. 2011. Sonic hedgehog shedding results in functional activation of the solubilized protein. *Dev Cell*, 20, 764-74.
- OLINS, A. L. & OLINS, D. E. 1974. Spheroid chromatin units (v bodies). *Science*, 183, 330-2.
- OSBORNE, C. S., CHAKALOVA, L., BROWN, K. E., CARTER, D., HORTON, A., DEBRAND, E., GOYENECHEA, B., MITCHELL, J. A., LOPES, S., REIK, W. & FRASER, P. 2004. Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat Genet*, 36, 1065-71.
- PAIVA, K. B., SILVA-VALENZUELA, M., MASSIRONI, S. M., KO, G. M., SIQUEIRA, F. M. & NUNES, F. D. 2010. Differential Shh, Bmp and Wnt gene expressions during craniofacial development in mice. *Acta Histochem*, 112, 508-17.
- PALM, W., SWIERCZYNSKA, M. M., KUMARI, V., EHRHART-BORNSTEIN, M., BORNSTEIN, S. R. & EATON, S. 2013. Secretion and signaling activities of lipoprotein-associated hedgehog and non-sterol-modified hedgehog in flies and mammals. *PLoS Biol*, 11, e1001505.
- PAN, A., CHANG, L., NGUYEN, A. & JAMES, A. W. 2013. A review of hedgehog signaling in cranial bone development. *Front Physiol*, 4, 61.
- PANÁKOVÁ, D., SPRONG, H., MAROIS, E., THIELE, C. & EATON, S. 2005. Lipoprotein particles are required for Hedgehog and Wingless signalling. *Nature*, 435, 58-65.
- PARADIS, F. H. & HALES, B. F. 2013. Exposure to valproic acid inhibits chondrogenesis and osteogenesis in mid-organogenesis mouse limbs. *Toxicol Sci*, 131, 234-41.
- PARK, H., LEE, Y. J., KIM, T. H., LEE, J., YOON, S., CHOI, W. S., MYUNG, C. S. & KIM, H. S. 2008. Effects of trichostatin A, a histone deacetylase inhibitor, on the regulation of apoptosis in H-ras-transformed breast epithelial cells. *Int J Mol Med*, 22, 605-11.
- PARKER, S. C., STITZEL, M. L., TAYLOR, D. L., OROZCO, J. M., ERDOS, M. R., AKIYAMA, J. A., VAN BUEREN, K. L., CHINES, P. S., NARISU, N., BLACK, B. L., VISEL, A., PENNACCHIO, L. A., COLLINS, F. S., PROGRAM, N. C. S., AUTHORS, N. I. O. H. I. S. C. C. S. P. & AUTHORS, N. C. S. P. 2013. Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proc Natl Acad Sci U S A*, 110, 17921-6.
- PENG, C., FU, L. Y., DONG, P. F., DENG, Z. L., LI, J. X., WANG, X. T. & ZHANG, H. Y. 2013. The sequencing bias relaxed characteristics of Hi-C derived data and implications for chromatin 3D modeling. *Nucleic Acids Res*, 41, e183.
- PENNACCHIO, L. A., AHITUV, N., MOSES, A. M., PRABHAKAR, S., NOBREGA, M. A., SHOUKRY, M., MINOVITSKY, S., DUBCHAK, I., HOLT, A., LEWIS, K. D., PLAJSER-FRICK, I., AKIYAMA, J., DE VAL, S., AFZAL, V., BLACK, B. L., COURONNE, O., EISEN, M. B., VISEL, A. & RUBIN, E. M. 2006. In vivo enhancer analysis of human conserved non-coding sequences. *Nature*, 444, 499-502.
- PENNACCHIO, L. A., BICKMORE, W., DEAN, A., NOBREGA, M. A. & BEJERANO, G. 2013. Enhancers: five essential questions. *Nat Rev Genet*, 14, 288-95.

- PENNISI, E. 2003. Human genome. A low number wins the GeneSweep Pool. *Science*, 300, 1484.
- PEPINSKY, R. B., ZENG, C., WEN, D., RAYHORN, P., BAKER, D. P., WILLIAMS, K. P., BIXLER, S. A., AMBROSE, C. M., GARBER, E. A., MIATKOWSKI, K., TAYLOR, F. R., WANG, E. A. & GALDES, A. 1998. Identification of a palmitic acid-modified form of human Sonic hedgehog. *J Biol Chem*, 273, 14037-45.
- PERIC-HUPKES, D., MEULEMAN, W., PAGIE, L., BRUGGEMAN, S. W., SOLOVEI, I., BRUGMAN, W., GRÄF, S., FLICEK, P., KERKHOVEN, R. M., VAN LOHUIZEN, M., REINDERS, M., WESSELS, L. & VAN STEENSEL, B. 2010. Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. *Mol Cell*, 38, 603-13.
- PETIT, F., JOURDAIN, A. S., HOLDER-ESPINASSE, M., KEREN, B., ANDRIEUX, J., DUTERQUE-COQUILLAUD, M., PORCHET, N., MANOUVRIER-HANU, S. & ESCANDE, F. 2016. The disruption of a novel limb cis-regulatory element of SHH is associated with autosomal dominant preaxial polydactyly-hypertrichosis. *Eur J Hum Genet*, 24, 37-43.
- PHILLIPS-CREMINS, J. E., SAURIA, M. E., SANYAL, A., GERASIMOVA, T. I., LAJOIE, B. R., BELL, J. S., ONG, C. T., HOOKWAY, T. A., GUO, C., SUN, Y., BLAND, M. J., WAGSTAFF, W., DALTON, S., MCDEVITT, T. C., SEN, R., DEKKER, J., TAYLOR, J. & CORCES, V. G. 2013. Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell*, 153, 1281-95.
- PLESSY, C., DICKMEIS, T., CHALMEL, F. & STRÄHLE, U. 2005. Enhancer sequence conservation between vertebrates is favoured in developmental regulator genes. *Trends Genet*, 21, 207-10.
- PORTER, J. A., YOUNG, K. E. & BEACHY, P. A. 1996. Cholesterol modification of hedgehog signaling proteins in animal development. *Science*, 274, 255-9.
- PRICE, C. M. 1993. Fluorescence in situ hybridization. *Blood Rev*, 7, 127-34.
- RADA-IGLESIAS, A., BAJPAI, R., SWIGUT, T., BRUGMANN, S. A., FLYNN, R. A. & WYSOCKA, J. 2011. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature*, 470, 279-83.
- RAMALHO-SANTOS, M., MELTON, D. A. & MCMAHON, A. P. 2000. Hedgehog signals regulate multiple aspects of gastrointestinal development. *Development*, 127, 2763-72.
- RAZIN, S. V., GAVRILOV, A. A., VASSETZKY, Y. S. & ULIANOV, S. V. 2016. Topologically associating domains: gene warehouses adapted to serve transcriptional regulation. *Transcription*, 0.
- RIDDLE, R. D., ENSINI, M., NELSON, C., TSUCHIDA, T., JESSELL, T. M. & TABIN, C. 1995. Induction of the LIM homeobox gene *Lmx1* by WNT7a establishes dorsoventral pattern in the vertebrate limb. *Cell*, 83, 631-40.
- RIDDLE, R. D., JOHNSON, R. L., LAUFER, E. & TABIN, C. 1993. Sonic hedgehog mediates the polarizing activity of the ZPA. *Cell*, 75, 1401-16.
- RUBIN, L. & SAUNDERS, J. W. 1972. Ectodermal-mesodermal interactions in the growth of limb buds in the chick embryo: constancy and temporal limits of the ectodermal induction. *Dev Biol*, 28, 94-112.
- RUBINSTEIN, M. & DE SOUZA, F. S. 2013. Evolution of transcriptional enhancers and animal diversity. *Philos Trans R Soc Lond B Biol Sci*, 368, 20130017.
- SAGAI, T., AMANO, T., TAMURA, M., MIZUSHINA, Y., SUMIYAMA, K. & SHIROISHI, T. 2009. A cluster of three long-range enhancers directs regional *Shh* expression in the epithelial linings. *Development*, 136, 1665-74.
- SANBORN, A. L., RAO, S. S., HUANG, S. C., DURAND, N. C., HUNTLEY, M. H., JEWETT, A. I., BOCHKOV, I. D., CHINNAPPAN, D., CUTKOSKY, A., LI, J., GEETING, K. P., GNIRKE, A.,

- MELNIKOV, A., MCKENNA, D., STAMENOVA, E. K., LANDER, E. S. & AIDEN, E. L. 2015. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc Natl Acad Sci U S A*, 112, E6456-65.
- SCHALCH, T., DUDA, S., SARGENT, D. F. & RICHMOND, T. J. 2005. X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature*, 436, 138-41.
- SCHERZ, P. J., HARFE, B. D., MCMAHON, A. P. & TABIN, C. J. 2004. The limb bud Shh-Fgf feedback loop is terminated by expansion of former ZPA cells. *Science*, 305, 396-9.
- SCHNEIDER, I. & SHUBIN, N. H. 2013. The origin of the tetrapod limb: from expeditions to enhancers. *Trends Genet*, 29, 419-26.
- SEDAT, J. & MANUELIDIS, L. 1978. A direct approach to the structure of eukaryotic chromosomes. *Cold Spring Harb Symp Quant Biol*, 42 Pt 1, 331-50.
- SEXTON, T., KURUKUTI, S., MITCHELL, J. A., UMLAUF, D., NAGANO, T. & FRASER, P. 2012. Sensitive detection of chromatin coassociations using enhanced chromosome conformation capture on chip. *Nat Protoc*, 7, 1335-50.
- SHETH, R., GRÉGOIRE, D., DUMOUCHEL, A., SCOTTI, M., PHAM, J. M., NEMEC, S., BASTIDA, M. F., ROS, M. A. & KMITA, M. 2013. Decoupling the function of Hox and Shh in developing limb reveals multiple inputs of Hox genes on limb growth. *Development*, 140, 2130-8.
- SHEVELYOV, Y. Y., LAVROV, S. A., MIKHAYLOVA, L. M., NURMINSKY, I. D., KULATHINAL, R. J., EGOROVA, K. S., ROZOVSKY, Y. M. & NURMINSKY, D. I. 2009. The B-type lamin is required for somatic repression of testis-specific gene clusters. *Proc Natl Acad Sci U S A*, 106, 3282-7.
- SIMONIS, M., KLOUS, P., HOMMINGA, I., GALJAARD, R. J., RIJKERS, E. J., GROSVELD, F., MEIJERINK, J. P. & DE LAAT, W. 2009. High-resolution identification of balanced and complex chromosomal rearrangements by 4C technology. *Nat Methods*, 6, 837-42.
- SIMONIS, M., KLOUS, P., SPLINTER, E., MOSHKIN, Y., WILLEMSSEN, R., DE WIT, E., VAN STEENSEL, B. & DE LAAT, W. 2006. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat Genet*, 38, 1348-54.
- SMITH, E. M., LAJOIE, B. R., JAIN, G. & DEKKER, J. 2016. Invariant TAD Boundaries Constrain Cell-Type-Specific Looping Interactions between Promoters and Distal Elements around the CFTR Locus. *Am J Hum Genet*, 98, 185-201.
- SMITH, J. C. 1980. The time required for positional signalling in the chick wing bud. *J Embryol Exp Morphol*, 60, 321-8.
- SMITH, J. C., TICKLE, C. & WOLPERT, L. 1978. Attenuation of positional signalling in the chick limb by high doses of gamma-radiation. *Nature*, 272, 612-3.
- SONG, L. & CRAWFORD, G. E. 2010. DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. *Cold Spring Harb Protoc*, 2010, pdb.prot5384.
- SOSHNIKOVA, N., ZECHNER, D., HUELSKEN, J., MISHINA, Y., BEHRINGER, R. R., TAKETO, M. M., CRENSHAW, E. B. & BIRCHMEIER, W. 2003. Genetic interaction between Wnt/beta-catenin and BMP receptor signaling during formation of the AER and the dorsal-ventral axis in the limb. *Genes Dev*, 17, 1963-8.
- SPLINTER, E., HEATH, H., KOOREN, J., PALSTRA, R. J., KLOUS, P., GROSVELD, F., GALJART, N. & DE LAAT, W. 2006. CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev*, 20, 2349-54.
- STADHOUDERS, R., KOLOVOS, P., BROUWER, R., ZUIN, J., VAN DEN HEUVEL, A., KOCKX, C., PALSTRA, R. J., WENDT, K. S., GROSVELD, F., VAN IJCKEN, W. & SOLER, E. 2013. Multiplexed chromosome conformation capture sequencing for rapid genome-scale

- high-resolution detection of long-range chromatin interactions. *Nat Protoc*, 8, 509-24.
- SUMMERBELL, D. 1974. A quantitative analysis of the effect of excision of the AER from the chick limb-bud. *J Embryol Exp Morphol*, 32, 651-60.
- SUMMERBELL, D., LEWIS, J. H. & WOLPERT, L. 1973. Positional information in chick limb morphogenesis. *Nature*, 244, 492-6.
- SWANSON, C. I., SCHWIMMER, D. B. & BAROLO, S. 2011. Rapid evolutionary rewiring of a structurally constrained eye enhancer. *Curr Biol*, 21, 1186-96.
- TABIN, C. & WOLPERT, L. 2007. Rethinking the proximodistal axis of the vertebrate limb in the molecular era. *Genes Dev*, 21, 1433-42.
- TANAKA, Y., OKADA, Y. & HIROKAWA, N. 2005. FGF-induced vesicular release of Sonic hedgehog and retinoic acid in leftward nodal flow is critical for left-right determination. *Nature*, 435, 172-7.
- TE WELSCHER, P., FERNANDEZ-TERAN, M., ROS, M. A. & ZELLER, R. 2002. Mutual genetic antagonism involving GLI3 and dHAND prepatterns the vertebrate limb bud mesenchyme prior to SHH signaling. *Genes Dev*, 16, 421-6.
- THANOS, D. & MANIATIS, T. 1995. Virus induction of human IFN beta gene expression requires the assembly of an enhanceosome. *Cell*, 83, 1091-100.
- THERIZOLS, P., ILLINGWORTH, R. S., COURILLEAU, C., BOYLE, S., WOOD, A. J. & BICKMORE, W. A. 2014. Chromatin decondensation is sufficient to alter nuclear organization in embryonic stem cells. *Science*, 346, 1238-42.
- THONGJUEA, S., STADHOUDERS, R., GROSVELD, F. G., SOLER, E. & LENHARD, B. 2013. r3Cseq: an R/Bioconductor package for the discovery of long-range genomic interactions from chromosome conformation capture and next-generation sequencing data. *Nucleic Acids Res*, 41, e132.
- TIMMONS, P. M., WALLIN, J., RIGBY, P. W. & BALLING, R. 1994. Expression and function of Pax 1 during development of the pectoral girdle. *Development*, 120, 2773-85.
- TOWERS, M., MAHOOD, R., YIN, Y. & TICKLE, C. 2008. Integration of growth and specification in chick wing digit-patterning. *Nature*, 452, 882-6.
- TREMETHICK, D. J. 2007. Higher-order structures of chromatin: the elusive 30 nm fiber. *Cell*, 128, 651-4.
- TSUI, M. G. 2014. *PDZD2, a candidate for brachydactyly type A1, encodes a secreted protein that negatively modulates hedgehog signaling* Ph.D. THESIS, The University of Hong Kong.
- TUKACHINSKY, H., KUZMICKAS, R. P., JAO, C. Y., LIU, J. & SALIC, A. 2012. Dispatched and scube mediate the efficient secretion of the cholesterol-modified hedgehog ligand. *Cell Rep*, 2, 308-20.
- TÓTH, K. F., KNOCH, T. A., WACHSMUTH, M., FRANK-STÖHR, M., STÖHR, M., BACHER, C. P., MÜLLER, G. & RIPPE, K. 2004. Trichostatin A-induced histone acetylation causes decondensation of interphase chromatin. *J Cell Sci*, 117, 4277-87.
- VALTON, A. L. & DEKKER, J. 2016. TAD disruption as oncogenic driver. *Curr Opin Genet Dev*, 36, 34-40.
- VAN DE WERKEN, H. J., DE VREE, P. J., SPLINTER, E., HOLWERDA, S. J., KLOUS, P., DE WIT, E. & DE LAAT, W. 2012. 4C technology: protocols and data analysis. *Methods Enzymol*, 513, 89-112.
- VAN DER VLAG, J. & OTTE, A. P. 1999. Transcriptional repression mediated by the human polycomb-group protein EED involves histone deacetylation. *Nat Genet*, 23, 474-8.
- VANDERMEER, J. E., LOZANO, R., SUN, M., XUE, Y., DAENTL, D., JABS, E. W., WILCOX, W. R. & AHITUV, N. 2014. A novel ZRS mutation leads to preaxial polydactyly type 2 in a

- heterozygous form and Werner mesomelic syndrome in a homozygous form. *Hum Mutat*, 35, 945-8.
- VARAMBALLY, S., DHANASEKARAN, S. M., ZHOU, M., BARRETTE, T. R., KUMAR-SINHA, C., SANDA, M. G., GHOSH, D., PIENTA, K. J., SEWALT, R. G., OTTE, A. P., RUBIN, M. A. & CHINNAIYAN, A. M. 2002. The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature*, 419, 624-9.
- VERGER, A., BAERT, J. L., VERREMAN, K., DEWITTE, F., FERREIRA, E., LENS, Z., DE LAUNOIT, Y., VILLERET, V. & MONTÉ, D. 2013. The Mediator complex subunit MED25 is targeted by the N-terminal transactivation domain of the PEA3 group members. *Nucleic Acids Res*, 41, 4847-59.
- VERHEYDEN, J. M. & SUN, X. 2008. An Fgf/Gremlin inhibitory feedback loop triggers termination of limb bud outgrowth. *Nature*, 454, 638-41.
- VIETRI RUDAN, M., BARRINGTON, C., HENDERSON, S., ERNST, C., ODOM, D. T., TANAY, A. & HADJUR, S. 2015. Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell Rep*, 10, 1297-309.
- VISEL, A., BLOW, M. J., LI, Z., ZHANG, T., AKIYAMA, J. A., HOLT, A., PLAJSER-FRICK, I., SHOUKRY, M., WRIGHT, C., CHEN, F., AFZAL, V., REN, B., RUBIN, E. M. & PENNACCHIO, L. A. 2009. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature*, 457, 854-8.
- VISEL, A., PRABHAKAR, S., AKIYAMA, J. A., SHOUKRY, M., LEWIS, K. D., HOLT, A., PLAJSER-FRICK, I., AFZAL, V., RUBIN, E. M. & PENNACCHIO, L. A. 2008. Ultraconservation identifies a small subset of extremely constrained developmental enhancers. *Nat Genet*, 40, 158-60.
- WADE, C., BRINAS, I., WELFARE, M., WICKING, C. & FARLIE, P. G. 2012. Twist2 contributes to termination of limb bud outgrowth and patterning through direct regulation of *Grem1*. *Dev Biol*, 370, 145-53.
- WAGNER, G. P. & CHIU, C. H. 2001. The tetrapod limb: a hypothesis on its origin. *J Exp Zool*, 291, 226-40.
- WHYTE, W. A., ORLANDO, D. A., HNISZ, D., ABRAHAM, B. J., LIN, C. Y., KAGEY, M. H., RAHL, P. B., LEE, T. I. & YOUNG, R. A. 2013. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell*, 153, 307-19.
- WIECZOREK, D., PAWLIK, B., LI, Y., AKARSU, N. A., CALIEBE, A., MAY, K. J., SCHWEIGER, B., VARGAS, F. R., BALCI, S., GILLESSEN-KAESBACH, G. & WOLLNIK, B. 2010. A specific mutation in the distant sonic hedgehog (SHH) cis-regulator (ZRS) causes Werner mesomelic syndrome (WMS) while complete ZRS duplications underlie Haas type polysyndactyly and preaxial polydactyly (PPD) with or without triphalangeal thumb. *Hum Mutat*, 31, 81-9.
- WILLIAMSON, I. 2012. *Differential chromatin topology and transcription factor enhancer binding regulate spatiotemporal gene expression in limb development*. PhD by Research, University of Edinburgh.
- WILLIAMSON, I., BERLIVET, S., ESKELAND, R., BOYLE, S., ILLINGWORTH, R. S., PAQUETTE, D., DOSTIE, J. & BICKMORE, W. A. 2014. Spatial genome organization: contrasting views from chromosome conformation capture and fluorescence in situ hybridization. *Genes Dev*, 28, 2778-91.

- WILLIAMSON, I., ESKELAND, R., LETTICE, L. A., HILL, A. E., BOYLE, S., GRIMES, G. R., HILL, R. E. & BICKMORE, W. A. 2012. Anterior-posterior differences in HoxD chromatin topology in limb development. *Development*, 139, 3157-67.
- WILLIAMSON, I., LETTICE, L. A., HILL, R. E. & BICKMORE, W. A. 2016. Shh and ZRS enhancer co-localisation is specific to the zone of polarizing activity. *Development*.
- WILSON, K. L. & FOISNER, R. 2010. Lamin-binding Proteins. *Cold Spring Harb Perspect Biol*, 2, a000554.
- WOLPERT, L. 1969. Positional information and the spatial pattern of cellular differentiation. *J Theor Biol*, 25, 1-47.
- WORMAN, H. J. & BONNE, G. 2007. "Laminopathies": a wide spectrum of human diseases. *Exp Cell Res*, 313, 2121-33.
- XIAO, B., JI, X., XING, Y., CHEN, Y. W. & TAO, J. 2013. A rare case of 46, XX SRY-negative male with approximately 74-kb duplication in a region upstream of SOX9. *Eur J Med Genet*, 56, 695-8.
- XU, M. & COOK, P. R. 2008. Similar active genes cluster in specialized transcription factories. *J Cell Biol*, 181, 615-23.
- YAMAGISHI, C., YAMAGISHI, H., MAEDA, J., TSUCHIHASHI, T., IVEY, K., HU, T. & SRIVASTAVA, D. 2006. Sonic hedgehog is essential for first pharyngeal arch development. *Pediatr Res*, 59, 349-54.
- YANG, Y. & NISWANDER, L. 1995. Interaction between the signaling molecules WNT7a and SHH during vertebrate limb development: dorsal signals regulate anteroposterior patterning. *Cell*, 80, 939-47.
- YAO, B., WANG, Q., LIU, C. F., BHATTARAM, P., LI, W., MEAD, T. J., CRISH, J. F. & LEFEBVRE, V. 2015. The SOX9 upstream region prone to chromosomal aberrations causing campomelic dysplasia contains multiple cartilage enhancers. *Nucleic Acids Res*, 43, 5394-408.
- YARDIMCI, G. G. & NOBLE, W. S. 2015. Predictive model of 3D domain formation via CTCF-mediated extrusion. *Proc Natl Acad Sci U S A*, 112, 14404-5.
- YASHIRO, K., ZHAO, X., UEHARA, M., YAMASHITA, K., NISHIJIMA, M., NISHINO, J., SAIJOH, Y., SAKAI, Y. & HAMADA, H. 2004. Regulation of retinoic acid distribution is required for proximodistal patterning and outgrowth of the developing mouse limb. *Dev Cell*, 6, 411-22.
- ZELLER, R., LÓPEZ-RÍOS, J. & ZUNIGA, A. 2009. Vertebrate limb bud development: moving towards integrative analysis of organogenesis. *Nat Rev Genet*, 10, 845-58.
- ZENTNER, G. E., TESAR, P. J. & SCACHERI, P. C. 2011. Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. *Genome Res*, 21, 1273-83.
- ZHAO, W., DAI, F., BONAFEDE, A., SCHAFER, S., JUNG, M., YUSUF, F., GAMEL, A. J., WANG, J. & BRAND-SABERI, B. 2009. Histone deacetylase inhibitor, trichostatin A, affects gene expression patterns during morphogenesis of chicken limb buds in vivo. *Cells Tissues Organs*, 190, 121-34.
- ZHAO, Z., TAVOOSIDANA, G., SJÖLINDER, M., GÖNDÖR, A., MARIANO, P., WANG, S., KANDURI, C., LEZCANO, M., SANDHU, K. S., SINGH, U., PANT, V., TIWARI, V., KURUKUTI, S. & OHLSSON, R. 2006. Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat Genet*, 38, 1341-7.
- ZORN, C., CREMER, C., CREMER, T. & ZIMMER, J. 1979. Unscheduled DNA synthesis after partial UV irradiation of the cell nucleus. Distribution in interphase and metaphase. *Exp Cell Res*, 124, 111-9.

- ZULLO, J. M., DEMARCO, I. A., PIQUÉ-REGI, R., GAFFNEY, D. J., EPSTEIN, C. B., SPOONER, C. J., LUPERCHIO, T. R., BERNSTEIN, B. E., PRITCHARD, J. K., REDDY, K. L. & SINGH, H. 2012. DNA sequence-dependent compartmentalization and silencing of chromatin at the nuclear lamina. *Cell*, 149, 1474-87.
- ZÚÑIGA, A., HARAMIS, A. P., MCMAHON, A. P. & ZELLER, R. 1999. Signal relay by BMP antagonism controls the SHH/FGF4 feedback loop in vertebrate limb buds. *Nature*, 401, 598-602.

# Chapter 9

---

## Published Papers



## **Paper 1**

DOUGLAS, A. T. & HILL, R. D. 2014. Variation in vertebrate cis-regulatory elements in evolution and disease. *Transcription*, 5, e28848.

## **Paper 2**

Douglas, Adam, Peluso, Silvia, Lettice, Laura A, and Hill, Robert E (Aug 2016) Cis-Regulatory Mutations in Human Disease. In: eLS. John Wiley & Sons Ltd, Chichester. <http://www.els.net> [doi: 10.1002/9780470015902.a0024920]

## **Paper 3**

MORT, R. L., FORD, M. J., SAKAUE-SAWANO, A., LINDSTROM, N. O., CASADIO, A., DOUGLAS, A. T., KEIGHREN, M. A., HOHENSTEIN, P., MIYAWAKI, A. & JACKSON, I. J. 2014. Fucci2a: a bicistronic cell cycle reporter that allows Cre mediated tissue specific expression in mice. *Cell Cycle*, 13, 2681-96.