THE UNIVERSITY *of* EDINBURGH

Differential chromatin topology and transcription factor enhancer binding regulate spatiotemporal gene expression in limb development

Iain Williamson

PhD by Research

University of Edinburgh

2012

Declaration

This thesis has been entirely composed by me. Apart from where stated the experiments were carried out solely by me. The work has not been submitted for any other degree or professional qualification.

Iain Williamson, July 2012

Acknowlegments

Many thanks to my two supervisors, Wendy and Bob, for giving me the opportunity to carry out my PhD project in their labs, their supervision over the past four years, and for correcting my thesis. A big thank you to Ragnhild, my day-to-day supervisor over the first 2-3 years, who instructed me on how to do most of the experimental techniques describe in this thesis and had a large part in the setting up of the HoxD project, and to Shelagh for teaching me how to do FISH, guidance in how best to analyse the data, and all the Shh ES cell work. Thanks also to Allison for help with tissue culture, Laura for limb tissue and preparation of tissue sections (and how to differentiate between anterior and posterior limb bud sections and draw an E10.5, as opposed to E11.5, forelimb bud), and Carlo for showing me how to do EMSAs. Thanks also to Liz for words of wisdom on FISH, and to Pradeep and Rob for help with ChIP-chip and data analysis. Many thanks to my student panel – Andrew and Julia – for their advice on how the project was developing and on where to focus to keep it on track. To the rest of the Bickmore and Hill labs and more generally to all on E3 and C3. To all the students, especially from my year, Gillian and Sehrish; and Victor, Sophie, Fay and Natalie whom I have shared an office with. Socially, thanks also to the Bickmore lab plus Diana, Abby, Dave, Kevin, Paul, and Carlo.

Elsewhere in the Unit, I would like to thank Paul and Matt for all microscopy-related matters, Graeme for bioinformatics, Craig for help in formatting this thesis into something fit for publishing, and all at technical services. Thanks also to the administration department who have provided great support, in particular Pauline and Sheena.

Thank you to all my family and friends who have been very supportive. To mum and Biff, Harvey, and especially to Sarah and Liam for putting up with my decade-long belated higher education. Finally, and most importantly, a huge thank you to Susan who has supported me throughout my PhD.

Abstract

Many developmental genes are located in gene-poor genomic regions and are activated by long-range enhancers located up to 1Mb away. Modification and reorganisation of chromatin structure is pivotal to such long-range gene regulation. A prerequisite for enhancer activity is the binding of transcription factors and co-factors with the interplay between activating and repressive factors determining tissue, spatial and temporal specificity.

Spatiotemporal control of sonic hedgehog (*Shh*) and the 5′ *Hoxd* genes (especially *Hoxd13)* is crucial for vertebrate limb anterior-posterior (A-P) axis and autopod patterning. *Shh* tissue specificity is controlled by multiple enhancers throughout an adjacent gene desert. The ~0.8Mb-distant limb enhancer (ZRS) bypasses nearby genes to activate only *Shh*. In contrast, limb-specific HoxD expression is regulated by multiple enhancers, with the ~200kb-distant global control region (GCR) regulatory element the most characterised. In this thesis I investigated the mechanisms of ZRS and GCR regulation of *Shh* and *Hoxd13* respectively. The model system used was immortalised cell lines derived from the anterior and posterior distal forelimb buds of E10.5 and E11.5 mouse embryos. Cell line data were confirmed in dissected limb tissue.

Increased expression of the 5′ *Hoxd* genes, particularly *Hoxd13*, correlated with the loss of the repressive, polycomb catalysed, histone modification H3K27me3 and decompaction of chromatin structure over the HoxD locus at the distal posterior forelimb bud at stage E10.5. Moreover, I show that the GCR spatially co-localises with the 5′ HoxD locus at the distal posterior region of E10.5-11 embryos. These data are consistent with the formation of a chromatin loop between *Hoxd13* and the GCR at the time and place of distal limb bud development when the GCR is required to initiate 5′ *Hoxd* gene expression. This is the first example of A-P differences in chromatin compaction and local folding in the limb.

Point mutations within the ZRS cause ectopic (anterior) *Shh* expression, which results in preaxial polydactyly (PPD). The ZRS contains multiple canonical ETS transcription factor binding motifs, and point mutations in two families with

PPD results in the formation of additional ETS binding sites. The point mutations cause the loss or reduction of ETV4/5 transcription factor binding at a non-canonical ETS binding site and enable additional binding instead of ETS1. I show that ETV4/5, ETS1 and another ETS protein GABPα all bind to the ZRS. This work has revealed the differential effect on *Shh* expression of two groups of ETS factors mediated through the ZRS. The binding of ETS1/ GABPα determines the posterior *Shh* expression domain while ETV4/5 restricts anterior *Shh* expression. Two point mutations alter the ETS-binding profile, creating an additional ETS1/ GABPα site that is sufficient to drive ectopic *Shh* expression.

DNA FISH on E11.5 forelimb and floorplate tissue sections revealed that the *Shh*-ZRS genomic locus is in a compact chromatin conformation in both *Shh*-expressing and non-expressing cells. However, I show that the ZRS co-localises with *Shh* to a significantly greater extent in the distal posterior limb bud and the floorplate compared with cells where *Shh* is not expressed.

This thesis presents novel research into long-range gene regulation during limb development, elucidating the role of chromatin re-organisation and how spatial-specific enhancer activity is determined by opposing sets of binding factors.

Table of Contents

Chapter 3: The ETS family have opposing functions in the limb that determine Shh spatial expression through the ZRS

Chapter 4: The *Shh*-ZRS regulatory region is held in a compact higher-order conformation in expressing and non-expressing cells

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

List of Tables

Chapter 2

Chapter 3

Chapter 4

Chapter 6

Abbreviations

A	adenine
ac	acetylation
AC	Family A & C ZRS mutation
AER	apical ectodermal ridge
A-P	anterior-posterior
AUS	Australian family ZRS mutation
Belg2	Belgian family 2 ZRS mutation
bp	base pair(s)
BSA	bovine serum albumin
C	cytosine
cDNA	complementary DNA
ChIP	chromatin immunoprecipitation
CpG	cytosine-phosphate-guanine
CT	chromosome territory
CTD	C-terminal domain
d	nuclear interprobe distance (μm)
D-MEM	Dulbecco's-minimal Eagle's medium
Dapi	4′,6-diamidino-2-phenylindole
dH_2O	deionised water
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
DTT	dithiothreitol
E	embryonic stage
EDTA	ethylene diamine tetraacetic acid
EGTA	ethylene glycol tetraacetic acid
EMSA	electrophoretic mobility shift assay
ES	embryonic stem
FCS	foetal calf serum
FISH	fluorescence in situ hybridization
G	guanine
GCR	global control region
GFP	green fluorescent protein
GO	gene ontology
HAT	histone acetyltransferase

HCNE	highly conserved noncoding element
HDAC	histone deacetylase
HMTase	histone methyltransferase
HRP	horseradish peroxidise
Ig	immunoglobin
LB	Luria Bertani broth
LCR	locus control region
lncRNA	long noncoding RNA
kb	kilobase pair(s)
M	molar
Mb	megabase pair(s)
me	methylation
MMU	*Mus musculus* chromosome
mRNA	messenger RNA
μg	microgram
μl	microlitre
μm	micrometre
nm	nanometre
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PEV	position-effect variegation
pFA	paraformaldehyde
PPD	preaxial polydactyly
PRC	polycomb repressive complex
PRE	polycomb response element
RNA	ribonucleic acid
RNAPII	RNA polymerase II
SBE	*Shh* brain enhancer
SDS	sodium dodecyl sulphate
SEM	standard error of the mean
SFPE	*Shh* floorplate enhancer
SNP	single nucleotide polymorphism
T	thymine
TAE	Tris-EDTA acetic acid buffer
TE	Tris-EDTA
TF	transcription factor
TFBS	transcription factor binding site
TSS	transcription start site

UV	ultra-violet radiation
WT	wild type
ZPA	zone of polarizing activity
ZRS	(zone of polarizing activity) regulatory sequence
3C	chromatin conformation capture
5C	chromatin conformation capture carbon copy
14fp	E11.5 posterior limb cell line

Chapter 1

Introduction

From the characterisation of the operon in *E. coli* (Jacob and Monod, 1961) through *Drosophila* and then mouse genetics – in tandem with human Mendelian disease genetics – the crucial role of non-coding *cis*-regulatory regions in directing gene expression has become apparent. Over the last decade, due to the rapidly increasing capabilities of genomic research, the identification of potential cis-regulatory regions has escalated; however, many of these proposed regulatory regions await functional characterisation. Multi-species conservation of not only coding, but also non-coding, genomic regions has overturned the view of the latter as being largely composed of just non-functional DNA sequences accumulated over evolutionary time. And as advances in the genomics field gather pace, in harness with biochemical and molecular biology techniques that enable more extensive (genome-wide) and in-depth (high resolution) analysis, it is becoming increasingly clear that a substantial proportion of the non-coding genome is required to control gene activity.

As well as reviewing the literature on *cis* regulation in general, and enhancers in particular, this introduction will also cover the two main themes of my project. (1) the role of chromatin in regulating gene activity; and (2) how enhancers are regulated – that is, how *trans* factors that bind to regulatory sequences form functional nucleoprotein complexes. My project focuses on long-range gene regulation in the mouse; almost exclusively on limb bud development and two regulatory regions spanning up to a megabase (Mb) in size that contain multiple *cis*-regulatory sequences controlling the key developmental genes Sonic hedgehog (*Shh*) and the HoxD cluster. Therefore, I will discuss the gene networks involved in the growth and patterning of the vertebrate limb, emphasising the importance of maintaining tight spatial and temporal regulatory control of the *Shh* and *Hoxd* genes.

1.1 Chromatin structure and nuclear organisation: how modification and conformation impinge upon transcription

The mammalian genome is over three billion DNA base pairs in size (about two metres in length) and is contained within a nucleus about six micrometres (μm) in diameter. In all eukaryotes DNA is intricately bound up with an extensive suite of chromatin proteins with a wide range of functions. The assembly of long DNA chains into chromatin serves not only for compaction but also to protect the DNA from physical and chemical damage and as a platform for DNA metabolism; including transcription. Here, I concentrate on aspects of chromatin composition and organisation that influence gene activity; first, how manipulation of chromatin structure determines whether specific genomic regions have an active or repressed configuration as it pertains to transcription; second, the three dimensional architecture of chromatin within the nucleus, and how reorganisation of higher-order structure enables activation or repression of transcription within localised regions or over long distances.

1.1.1 Transcriptional status of genomic loci determined by nucleosome dynamics, chromatin modification, and bound protein factors and co-factors

1.1.1.1 Nucleosome dynamics

The basic unit of chromatin is the nucleosome, composed of ~147 base pairs (bp) of DNA wrapped around an octamer of histone proteins and results in the 'beads-on-a-string' 10 nm chromatin fibre (20-50bp of linker DNA joins each nucleosome). Four histone proteins, H2A, H2B, H3 and H4 comprise the core histone octamer: a H3:H4 tetramer and two H2A:H2B dimers. N terminal "tails" of core histones project out from the nucleosome and are susceptible to post-translational modification. The nucleosome creates a physical barrier that the factors involved in transcription, replication and DNA repair have to overcome. Incorporation of histone variants (e.g. H2A.Z and H3.3) generally result in more labile nucleosomes and frequently occur at regulatory sites, such as promoters and enhancers, thereby easing their removal to facilitate transcription factor/ co-factor binding (Jin et al., 2009).

A study that mapped potential enhancer locations in the human genome (binding of androgen receptor to predominantly distal regulatory sites induced by treatment of cancer cells with androgen) identified nucleosome removal at androgen receptor and FOXA1 binding sites while the neighbouring nucleosomes were marked by mono- and dimethylation of the H3 tail at lysine four (H3K4me1/2) (He et al., 2010). This histone mark is characteristic of enhancers (see below). The evicted central nucleosome was more likely to contain the H2A.Z variant and the underlying genomic sequence was more highly conserved than the sequences of the flanking nucleosomes and was A/T and AA/TT/TA/AT dinucleotide rich. Previous work had observed that whereas nucleosome-enriched genomic regions contain high levels of C/G and GC dinucleotides, the opposite is true for nucleosome-depleted regions, which are AT-rich (Peckham et al., 2007; Yuan & Liu, 2008). Therefore, a combination of transcription factor binding, core histone composition and DNA sequence can determine nucleosome stability; with regulatory elements characteristically nucleosome-depleted, H2A.Z-enriched, and marked by histone modifications associated with an active chromatin state.

1.1.1.2 Histone modification marks associated with active, repressed and silent loci

Histones are subject to an extensive suite of post-translational modifications at various amino acid residues within the globular core but particularly throughout their N and C terminal "tails" (Fig. 1.1). Residues can be acetylated (lysine, K), methylated (K and arginine, R) phosphorylated (serine, S, and threonine, T), ubiquitylated (K), sumoylated (K), deiminated (R), or undergo isomerization (proline, P-*cis* > P-*trans*) (Bannister & Kouzarides 2011). Whereas acetylation of H3 and H4 residues is exclusively associated with more open, decondensed chromatin, modifications such as methylation and ubiquitination are present in both active and repressed genomic regions, depending on the residue modified and/or, in the case of methylation, the number of covalently bound groups at a specific residue. Before describing in more depth the active and repressive marks (and the enzymes responsible) found at promoters and enhancers, I will briefly discuss the concepts of euchromatin and heterochromatin.

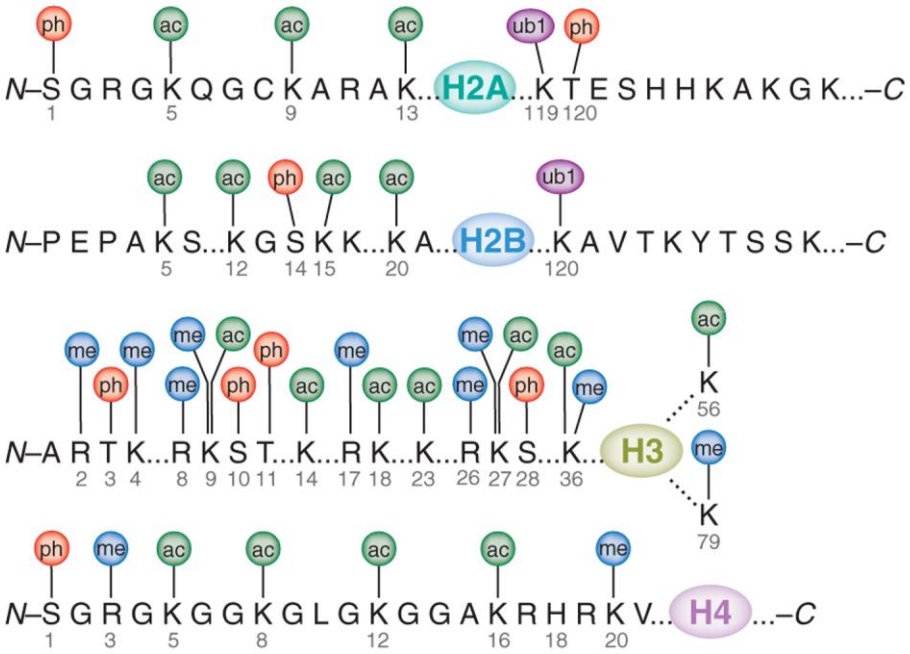

Figure 1.1. Modifications to residues on histone tails. Locations within the histone "tails" of residues that undergo post-translational modification are identified. Amino acid residues: K, lysine; R, arginine; S, serine; T, threonine. Histone modifications: me, methylation; ph, phosphorylation; ac, acetylation; ub1, ubiquitination. (From (Bhaumik & Smith, 2007))

Heterochromatin can describe either permanently compacted pericentric and telomeric chromosome regions (constitutive), or densely compacted chromatin at genomic regions in certain tissues or at varying stages in differentiation and development (facultative). In the mouse, constitutive heterochromatin is associated with repetitive DNA such as tandem satellite repeats and has characteristic histone modifications such as tri-methylation of lysines 9 and 20 of H3 (H3K9me3, H4K20me3) (Martens et al. 2005). Interspersed repeats and transposons (LINEs, SINEs) have more variable associations with different degrees of H3K9, H3K20 and H3K27 methylation dependent on the cell differentiation state. A classic example of facultative heterochromatin is the inactive X chromosome (XCI) in which the long non-coding RNA (lncRNA) *Xist* recruits the polycomb complex PRC2 (I discuss polycomb more extensively in sections 1.1.1.4-1.1.1.7), resulting in the silenced X chromosome being covered by the PRC2-catalysed H3K27me3 (Mak et al., 2004; Okamoto et al., 2004).

Euchromatic domains are generally described as "open" and are often more gene-rich, compared to consitutive heterochromatin. Multiple acetylation of lysine residues of H3 and H4 is a feature of chromatin structure over transcriptionally active genes and regulatory elements (Kuo et al. 1998), and the recruitment of histone acetyl transferases (HATs) by transcription factors bound to promoters and enhancers is well established (Bannister & Kouzarides, 1996; Mizzen et al., 1996; Ogryzko et al., 1996; Chen et al., 1997; Kouzarides, 2007). However, histone deacetylases (HDACs) – enzymes frequently identified at facultative heterochromatin – have been found on active genes (Wang et al. 2009).

Methylation of K4 on H3 is associated with active or poised regulatory elements (Zhou et al. 2010). H3K4me2/3 are found at active gene promoters with generally lower more widespread levels of the mono-methyl mark, and in addition H3K4me3 is also found at the promoters of bivalently marked genes (section 1.1.1.6). Mono- and di-methylation of H3K4 mark distal regulatory elements, although the tri-methyl modification has been identified at enhancers also (section 1.1.1.3). The histone methyltransferase (HMTase) responsible for most H3K4me3 is Set1 of the COMPASS (COMplex Proteins ASsociated with Set1) complex

(Ardehali et al. 2011) though the MLL family of HMTases also bring about H3K4me3 at specific target genes, with MLL1/2 responsible for modifying Hox promoters (Milne et al., 2002; Ernst et al., 2004; Wang et al., 2009). *Drosophila* studies suggest that the continuous presence of H3K4 methylases at gene promoters is required to maintain expression, and that the H3K4me3 mark is either not maintained following their removal or is insufficient on its own to maintain transcription (Ingham, 1985; Klymenko & Mullert, 2004). The role of the modification itself is likely to be as a marker recognised by chromatin modifying enzymes and other factors required to facilitate RNA polymerase II (RNAPII) binding to the DNA (Eissenberg & Shilatifard, 2010).

1.1.1.3 Histone modifications associated with enhancers

Histone modification patterns at enhancers are highly cell type-specific (Heintzman et al., 2009). Analysis of 30Mb of the human genome (selected by the ENCODE project), in immortalised human HeLa cells, by chromatin immunoprecipitation (ChIP) on arrays (ChIP-chip) identified H3K4me1 but not H4K3me3 as being associated with putative enhancer sites (Heintzman et al., 2007); that is, the mono-methylated mark was present at the same distal loci as the transcriptional activator and acetyltransferase p300 (Eckner et al., 1994; Merika et al., 1998; Maston et al., 2006). These results agreed with the initial ENCODE study based on multiple human cell lines (Birney et al., 2007); however, ChIP-seq analysis on human T-cells suggested that both H3K4me1 and H3K4me3 are enriched at putative enhancers (Barski et al., 2007; Wang et al., 2008). To resolve this discrepancy and establish if these histone modifications can distinguish enhancers from active promoters, Robertson et al. (2008) identified DNA-binding sites for two transcription factors that bind predominantly at distal regulatory elements: STAT1 in unstimulated and IFNγ–stimulated HeLa cells, and Foxa2 in mouse adult liver. H3K4me1 was the dominant modification at distal regulatory elements, but a sizeable proportion of these distal sites had both H3K4 methylation marks. The authors suggest that this discrepancy may either be due to different modifications associating with functionally different transcription factors at distal elements or that the variation stems from different tissues or cellular states. In human T cell differentiation

H3K4me1 was shown to be associated with enhancers whether active or not, while increased H3K4me3 correlated with RNAPII occupancy of enhancers and greater regulatory activity (Pekowska et al., 2011).

Whereas the methylases responsible for H3K4me3 are well defined (Eissenberg & Shilatifard, 2010), HMTases that specifically catalyse H3K4me1 are not known; although COMPASS is capable of mono- di- and tri-methylation of H3K4 in yeast (Wood et al., 2007). In contrast, demethylation of H3K4me1 and H3K4me3 is catalysed by different demethylases. The JARID1 family have been identified as specific demethylases of H3K4me2/3, and the JARID1 RBP2 enzyme is associated with repressed Hox genes in undifferentiated ES cells (Christensen et al., 2007; Iwase et al., 2007; Lee et al., 2007). The lysine-specific LSD1 was the first demethylase to be characterised (Shi & Whetstine, 2007). LSD1 demethylates H3K4 and K9 mono- or di-methyl marks but is unable to remove the tri-methyl modification. A recent study implicates LSD1 in silencing enhancers of genes that maintain mouse ES cell pluripotency during differentiation and following the loss of *Pou5f1* binding. LSD1 is not essential for maintaining ES cells in an undifferentiated state (Whyte et al., 2012).

Acetylation of H3K27 (H3K27ac) – a mark that can be deposited by the p300 and CBP acetyltransferases (Tie et al., 2009) – is prevalent at the transcription start sites (TSS) of active genes in mammalian cells (Wang et al., 2008), but can also identify active enhancers (Rada-Iglesias et al., 2010; Creyghton et al., 2010). Whereas enhancers active in human embryonic stem cells (hESCs) are marked not only by H3K4me1 but also H3K27ac, enhancers that activate genes involved in early embryogenesis – "poised enhancers" – are distinguished by a lack of H3K27ac and enrichment of the repressive H3K27me3 mark (Rada-Iglesias et al., 2010). A subset of poised enhancers acquires H3K27ac during neuronal differentiation. This also suggests a role for H3K27me3 in regulating enhancer activity, at least in ES cells.

1.1.1.4 The polycomb repressive complexes

H3K27me3 correlates with gene repression in development and is mediated by the polycomb repressive complex 2 (PRC2) (Fig. 1.2). Originally identified in

Drosophila, the PRC2 complex co-operates with the PRC1 (Fig. 1.2) complex to maintain gene-silencing. PRC2 tri-methylates H3K27 (Cao et al., 2002) through the Ezh methyltransferase component of the complex, a function conserved from mammals (Ezh2 and Ezh1 (Shen et al., 2008)) to flies (in *Drosophila* the homologue is E(Z) (Czermin et al., 2002)). Ezh2-catalysed H3K27me3 can be bound by the second polycomb complex, PRC1, through the chromodomain of the polycomb protein homologues (Cbxs in mammals*)* that specifically recognise H3K27me3 (Cao et al., 2002). The functions and target specificity of the different Cbx paralogues is not fully understood, but Cbx7 seems to be the main polycomb homologue important in pluripotent ES cells (Morey et al., 2012). Moreover, PRC1, in a complex with RYBP, can bind to target sites, albeit at generally low levels, in PRC2-deficient mouse ES cells, therefore revealing a H3K27me3-independent PRC1 recruitment pathway (Tavares et al., 2012). PRC1 can also be recruited to the inactive X in the absence of PRC2 (Schoeftner et al., 2006).

PRC1 can ubiquitinate H2A at lysine 119 (uH2A) through the E3 ligase activity of Ring1A/1B (de Napoles et al. 2004) (Wang et al. 2004) (Buchwald et al. 2006). Ring1B is the predominant ubiquitin E3 ligase during differentiation and early development; loss of the Ring1B results in an extensive depletion of uH2A whereas the loss of uH2A is minimal in the absence of the Ring1A (de Napoles et al., 2004). The relative importance of the two mammalian Ring proteins in early development was demonstrated by Ring1A knockout mice being viable with only minor skeletal defects, while loss of Ring1B results in embryonic lethality (Lorente et al., 2000; Voncken et al., 2003).

1.1.1.5 Targets and targeting of polycomb

Mouse embryos lacking *Eed*, *Ezh2* or *Suz12* either fail to develop following implantation or complete gastrulation (Faust et al., 1995; O'Carroll et al., 2001; Pasini et al., 2004). Extensive H3K27me3 profiles have been determined for both human and mouse ES cells (Mikkelsen et al., 2007; Pan et al., 2007; Ku et al., 2008). Several studies into the role of polycomb in mammalian development have established the involvement of both PRC2 and PRC1 in maintaining ES cell pluripotency and the co-ordination of temporal activation of key developmental

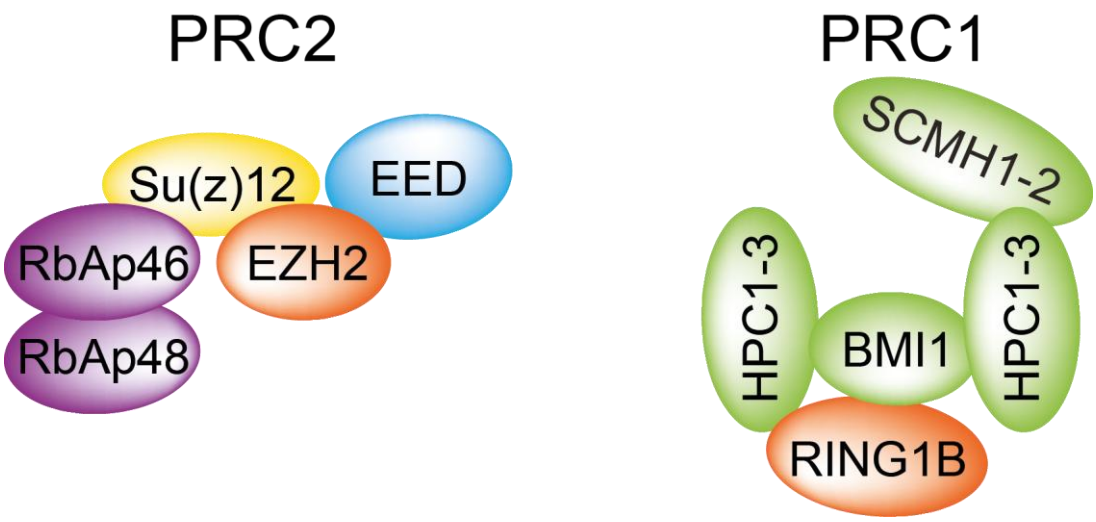

Figure 1.2. Components of the mammalian polycomb response complexes. PRC2 complexes contain four core elements: the catalytic subunit EZH2, Su(z)12, EED and nucleosome remodelling factor RbAp46/48. In PRC1 complexes the core components are: SCMH1-2, HPC1-3, BMI1 and the catalytic subunit RING1B.

genes (Azuara et al., 2006; Boyer et al., 2006; Lee et al., Jorgensen et al., 2006; Stock et al., 2007; Endoh et al., 2008) and also during the terminal differentiation of multipotent cells in embryogenesis (Caretti et al., 2004). PRC2 occupies large repressed domains containing genes encoding developmental regulators in human ES cells (Lee et al., 2006). In murine ES cells, removal of the PRC2 component Eed results in the loss of PRC1, and H3K27me3 and derepression of polycomb-target genes (Boyer et al., 2006).

Unlike in flies, where polycomb response elements (PREs) have been widely characterised (Ringrose & Paro, 2007), specific mammalian PRE complexes – with a couple of exceptions (Sing et al., 2009; Woo et al., 2010) – have yet to be identified. However, several studies noted that polycomb targets tended to correlate with the presence of CpG islands (Ku et al., 2008; Mohn et al., 2008). Moreover, using murine ES cells engineered to contain modified bacterial artificial chromosomes and novel CpG island–like sequences, free of active motifs, inactive GC-rich genomic sequences were shown to be sufficient to recruit PRC2 (Mendenhall et al., 2010). Taking this analysis further, a recent study showed that bivalent chromatin regions in ES cells (containing both H3K27me3 and H3K4me3) result from the competition between polycomb recruitment and transcriptional activation at non-methylated CpG islands (Lynch et al., 2011). Using a humanised mouse model, the authors show that the extensive recruitment of polycomb complexes and chromatin bivalency at the human CpG island-containing α-globin locus is due to *cis*-acting sequences, as a 4kb region of the human locus can establish a novel bivalent domain when inserted into the mouse locus - which lacks CpG islands.

Further means of recruiting PRC2 invoke the activity of lncRNAs – both in *cis* (*Xist* in X chromosome inactivation (Zhao et al., 2008)) and in *trans* (*HOTAIR* in repression of *HOXD* (Rinn et al., 2007)). Moreover, many small RNAs have been co-purified with PRC2 (Kanhere et al., 2010; Zhao et al., 2010; Kotake et al., 2011).

1.1.1.6 Bivalent chromatin domains

In ES cells polycomb-marked (H3K27me3) domains are pervasively marked with both H3K4me3 and H3K27me3 – not only in and around promoter regions but

throughout the genomic regions in which they are situated. Initially characterised at highly conserved noncoding elements (HCNEs) predominantly associated with genes active in development in mouse ES cells (Bernstein et al., 2006; Azuara et al., 2006), these "bivalent" modifications were subsequently identified genome-wide at developmental genes in both mouse (Mikkelsen et al., 2007) and humans ES cells (Pan et al., 2007). Many of the bivalent genes identified in the latter study are upregulated early on in differentiation and are therefore described as being in a "poised" state. Further genome-wide analysis in human T-cells demonstrated that this bivalency is not unique to ES cells (Roh et al., 2006). Loss of H3K4me3, through mutation of the DPY-30 component of COMPASS, does not affect undifferentiated ES cell renewal but is detrimental to differentiation – suggesting that H3K4me3 at bivalent genes is required for their upregulation during early development (Jiang et al., 2011).

1.1.1.7 Polycomb and mechanisms of transcriptional repression

The mechanisms by which polycomb-repressive complexes repress their target genes has still to be fully elucidated, but it appears to be at a step downstream of transcription initiation. The C-terminal domain (CTD) of the largest RNAPII subunit, RPB1, contains multiples of the same heptapeptide consensus sequence with three serine residues (Ser) phosphorylated at various stages of the RNAPII cycle. RNAPII is recruited to TSSs in a hypophosphorylated form and its release from the gene promoter is dependent upon phosphorylation of Ser 5 (S5p) by TFIIH. Phosphorylation of S5 correlates with transcription initiation, capping of the 5′ end of the nascent RNA and histone methyl transferase (HMT) recruitment. If not aborted, RNAPII-S5p reaches an intrinsic pausing site and requires phosphorylation of Ser 2 (S2p) by the P-TEFb kinase for productive elongation. Phosphorylation of Ser 7 (S7p) is present at the promoter and exons of active genes and is associated with both S5p and S2p (Brookes & Pombo, 2009). ChIP has revealed that RNAPII is present at the 'silent' polycomb targets in ES cells, but that this polymerase is in an unusual form (Stock et al. 2007). The presence of S5p, and absence of S2p, suggests that the polymerase is in a form that can initiate transcription, but that cannot elongate. Co-occupation of RNAPII-S5p, in this poised configuration, and PRC1 at bivalent genes

in murine ES cells has been elucidated (Stock et al., 2007). The poised configuration is unusual in that although S2p is absent, RNAPII cannot be detected by the 8WG16 antibody that specifically recognizes non-phosphorylated S2 residues, indicating that the residues are inaccessible due to another CTD modification or conformational change (Brookes & Pombo, 2009). Loss of PRC1 and H2Aub1 releases this poised RNAPII to allow expression of the target genes. Genome-wide, PRC and RNAPII-S5p simultaneously co-occupy both promoters and coding regions of silenced developmental genes in murine ES cells, which results in transcripts that fail to mature into mRNA (Brookes et al., 2012).

1.1.2 Higher-order chromatin structure and the regulation of transcription

Chromatin is composed of DNA and proteins in an approximately 1:2 ratio by mass (Lee & Young, 2000). At the gross nuclear level of interphase cells there are two types of chromatin packaging: tightly wound condensed heterochromatin and the relatively open structure of euchromatin. Chromatin organisation above the 10 nm fibre level is not well defined. Evidence for a 30 nm fibre largely comes from *in vitro* assays on pure polynucleosomes (Fussner et al., 2011). In vivo, the linker histone H1 has been implicated in the formation of 30 nm fibres (Allan et al., 1981; Bates et al., 1981) and is particularly prevalent in heterochromatin and depleted in ES cells (Fan et al., 2003; Fan et al. 2005). Indeed, downregulation of H1 in *Drosophila* causes heterochromatin dispersal (Lu et al., 2009), and observation by electron microscopy (e.m.) of polynucleosomes (20-40) from wild type and H1 knockout mouse ES cells indicated that reduced H1:nucleosome stoichiometry results in chromatin decondensation in mammals (Fan et al., 2005). Although *in vivo* observation by e.m. of structured 30 nm fibres has been rare (Fussner et al. 2011), regions of densely compacted chromatin can be detected by cytological analysis; and have been confirmed by biophysical fractionation of chromatin (Gilbert et al., 2004; Ghirlando & Felsenfeld, 2008). Analysis of the fractionated chromatin suggests that heterochromatin structure is heterogenous and that compact chromatin fibres are also located in euchromatin. Moreover, although open chromatin fibres are prevalent in gene-rich regions, there is no correlation of open chromatin structure and gene

expression. Rather, an open chromatin configuration implies a sub-nuclear environment conducive to gene transcription dependent upon the presence of the correct activating factors.

FISH was used to confirm the biophysical data in Gilbert et al. (2004). Squared interprobe distances (d^2) between two FISH probes have a linear relationship to the genomic distance that separates them (van den Engh et al., 1992) and can be used to compare chromatin compaction between different regions in the same cell type (Yokota et al., 1997), at the same locus between cells at different stages of differentiation (Chambeyron & Bickmore, 2004), or at the same locus in different tissues during embryogenesis (Morey et al., 2007). Structural properties of interphase chromosomes over intermediate (> 0.1-< 1.5 Mbp) and large (2-200 Mbp) scales can be ascertained by calculating statistical ratios of standard deviation/mean and median/mean from interprobe distance data sets. Ratios approximate to ~0.52 - ~0.94 describe a random walk path – in other words chromatin structure at the region being measured is flexible and in a random configuration (Sachs et al., 1995).

1.1.2.1 Nuclear location and re-location from chromosome territories correlates with expression state of some genes

Various applications of FISH, to locate whole chromosomes or specific loci within the interphase nucleus, have been employed to characterise nuclear organisation and higher-order chromatin structure. Interphase chromosomes inhabit discrete chromosome territories (CTs), with relatively gene-rich chromosomes generally located near the centre of the nucleus while more gene-poor chromosomes locate towards the periphery (Boyle et al., 2001). When inactive, the β-globin locus is associated with centromeric heterochromatin at the nuclear periphery. During successive stages of erythroid differentiation there is progressive re-location of the locus to the nuclear centre with a corresponding upregulation of gene transcription (Ragoczy et al., 2006). Activation of β-globin occurs upon movement away from the heterochromatic region but still within the nuclear periphery, albeit at levels relatively lower than subsequent expression upon re-location to the centre. Furthermore, upregulation was shown to involve association of the β-globin genes with RNAP II foci that is dependent on a group of distal regulatory elements known

as the locus control region (LCR). Artificial tethering of genes to nuclear membrane structures can reduce the transcriptional efficiency of some but not all, confirming that the nuclear periphery is not incompatible with gene expression (Finlan et al., 2008; Kumaran & Spector, 2008; Reddy et al., 2008).

Relocation of chromatin fibres associated with co-ordinately expressed gene clusters to the periphery or even the outside of visible CTs often correlates with their activation, (Volpi et al., 2000; Mahy et al., 2002b). The murine and human β-globin loci loop out of their CTs in proerythroblast cells prior to their activation (Ragoczy et al., 2003), providing evidence that looping out from the CT can reflect poised and even repressed states as well as active states. This re-localisation depends on the LCR. However, a contradictory study reported that in mouse spleen cells and human erythroblasts the β-globin locus mostly remained within the CT at various stages of differentiation, including the stages of maximum expression and those immediately preceding (poised state) (Brown et al., 2006). In an elegant experiment to determine the relative importance of LCR activity and its correlation with chromatin decondensation, spread of histone modification marks, locus repositioning away from the CT, association with RNAPII foci and gene activation by long-range regulation through chromatin loops, the human β-globin LCR was inserted into a gene-rich region of the mouse genome (Noordermeer et al., 2008). The human LCR strongly increased the frequency of this locus being re-located outside its CT, however, upregulation of gene transcription was more dependent on orientation of the LCR and its ability to interact with specific genes – i.e. through enhancer activity from long-range chromatin loops (discussed further in Section 1.2.4). And the re-location of the Hoxb cluster outside of its chromosome territory during ES cell differentiation has no effect on the expression levels of neighbouring genes (Morey et al., 2009). An oligonucleotide pool generated to recognise the complete exome of murine chromosome 2 (MMU2) by FISH revealed that the coding sequences are contained within decondensed chromatin regions away from the periphery-located CT core and towards the nuclear interior in ES cells (Boyle et al., 2011). Localised chromatin conformation, then, is more crucial in creating the conditions required for transcription to occur than nuclear position or re-location from CTs, as gene

expression also occurs within the territories (Mahy et al., 2002a) – which shows that the transcriptional machinery can access internal genes.

1.1.2.2 Long-range chromatin interactions demarcate chromosome sub-compartments

Chromatin loops, distinct from the mechanism by which chromatin fibres project out of the CT, are structures that result from long-range interactions in *cis.* Interactions within chromosomes from tens of kilobases to over megabase scales have been detected using techniques such as chromatin conformation capture (3C) and its derivatives, and FISH. The Hi-C adaptation couples the capture of sheared biotinylated 3C products by streptavidin beads with massively parallel sequencing to allow genome-wide probing of three-dimensional architecture. Long-range *cis* interactions across human p and q chromosome arms can be detected by Hi-C, whereas contacts across the centromere are limited (Lieberman-Aiden et al., 2009). This data corresponds with FISH analysis that shows p and q arms occupying mostly distinct, non-overlapping territories (Dostie & Bickmore, 2012). Hi-C contact maps of human and *Drosophila* chromosomes reveal that active and inactive domains generally segregate. Inactive domains tend to remain within CTs whereas active domains are less compact, locate to CT peripheries, and are more likely to form interchromosomal contacts with other active regions (Lieberman-Aiden et al., 2009; Yaffe & Tanay, 2011; Sexton et al., 2012). Hi-C allows the spatial organisation of entire genomes to be probed; however, DNA sequencing cost and depth limits the resolution scale to about 1 Mb (Dostie & Bickmore, 2012). Domains with a high gene density and an open chromatin configuration or high levels of transcription can also be detected together by 4C (Simonis et al., 2006), which has also been used to reveal specific enhancer-promoter interactions (Montavon et al., 2011).

In *Drosophila*, 4C and FISH assays demonstrated that polycomb-bound chromatin domains within the same chromosome arm co-localise in nuclear space, therefore implicating polycomb proteins as facilitators of long-range interactions (of polycomb targets) in *cis* (Bantignies et al., 2011). Restriction of contacts between polycomb-bound domains to discrete chromosome arms are imposed by the centromere; a

percentric inversion enabled contacts to be made between polycomb targets normally located on separate arms (Tolhuis et al., 2011).

1.1.3 Chromatin modification and higher-order reorganisation are involved in regulation of the Hox clusters

There are four mammalian Hox clusters which encode homeobox transcription factors with crucial roles in anterior-posterior (A-P) patterning of the main embryonic body axis (Fig.1.3). They show the property of colinearity: their activation occurs in a tight spatial and temporal manner from the 3′ to 5′ ends of the clusters corresponding to anterior to progressively more posterior expression domains across the main body axis (Kmita & Duboule, 2003). Two of the mammalian Hox clusters – A and D – are also involved in limb and genital development (Deschamps, 2007).

Hox are classic polycomb targets – from flies to mammals – coated by H3K27me3 in ES cells and upregulated in polycomb mutants (Azuara et al. 2006; Boyer et al. 2006; Jørgensen et al. 2006; Lee et al. 2006; Endoh et al. 2008). Upon differentiation of murine ES cells H3K27me3 is progressively lost from the HoxB and HoxD clusters from the 3′ to the 5′ direction in each loci – i.e. from the earliest activated genes to the progressively later activated genes (Eskeland et al., 2010) (Fig. 1.4A). A similar polar loss of H3K27me3 is seen during embryogenesis (Soshnikova & Duboule 2009). As the progressively more 5′ *Hoxd* genes are being expressed in the developing tail bud region of stages E8.5 to E9.5 embryos, H3K27me3 is lost as the genes become highly marked with H3K4me3. Furthermore, the 3′ genes become enriched for H3K27me3 as they are being switched off. Both these studies highlight the crucial role played by the modifying complexes – polycomb for H3K27me3 and Set1/MLL for H3K4me3 – that lay down these histone marks in Hox regulation.

1.1.3.1 Polycomb complexes determine Hox chromatin conformation and consequent gene repression

Several studies also implicate polycomb in the regulation of higher-order chromatin compaction in Drosophila (Marchetti, 2003; Francis et al., 2004). In ES cells the silent Hox clusters are maintained in a compact chromatin state (Chambeyron &

Bickmore, 2004), that then unfolds upon the activation of Hox clusters – upon ES cell differentiation (Chambeyron & Bickmore, 2004; Morey et al., 2007) (Fig. 1.4B). There is also a looping out from the CT as the genes become expressed. These levels of chromatin and nuclear organisation are recapitulated during embryogenesis (Chambeyron & Bickmore, 2004; Morey et al., 2007). However, by looking at chromatin compaction and CT organisation simultaneously using 4 colour FISH, Morey et al. (2007) also detected decompaction at HoxD loci still located within in their CTs during ES cell differentiation. These results show that although nuclear reorganisation and chromatin decompaction are conserved mechanisms involved in Hox activation, they are not one and the same process; and provide further evidence for decondensation rather than looping out of CTs as being more indicative of an active chromatin state.

The compact chromatin conformation at Hox loci in ES cells is mediated by the polycomb complexes (Eskeland et al., 2010). In *Ring1B*$^{-/-}$ ES cells both the HoxB and HoxD loci decompact and there is an upregulation of gene expression, whereas in *Ring1B*$^{+/-}$ ES cells both clusters are in a more decompact state but gene expression is not significantly increased (Eskeland et al., 2010). Chromatin unfolding, then, is a direct result of a reduction in Ring1B levels and is not a consequence of gene transcription. The addition of full length wild type (WT) *Ring1B* or a catalytically inactive version into *Ring1B*$^{-/-}$ ES cells specifically rescued chromatin compaction at the Hox loci (Eskeland et al., 2010). This paper has unravelled the key functional role of polycomb in *Hox* gene regulation and demonstrated causation between polycomb-binding, chromatin conformation and gene repression, and has shown that Ring1B can repress gene expression through chromatin compaction independent of its catalytic function.

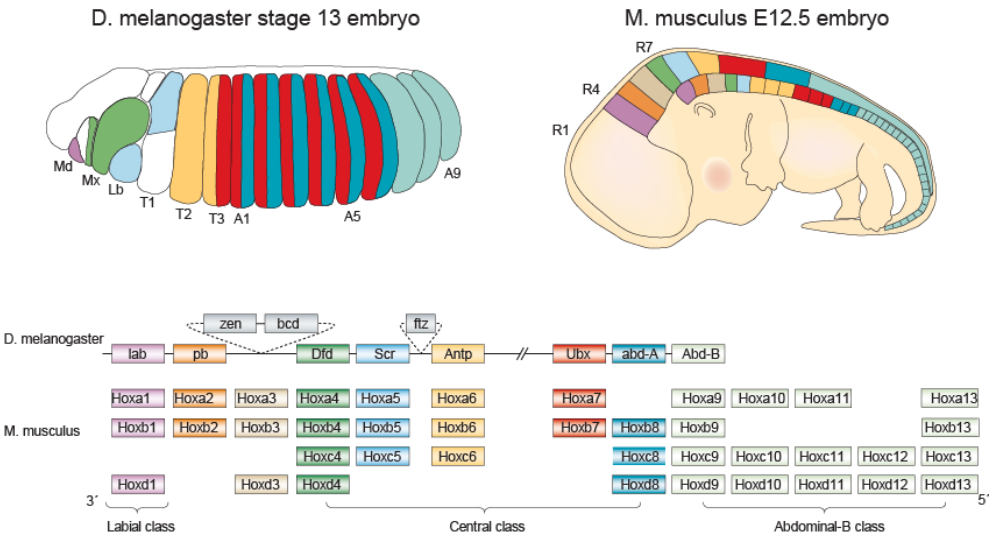

Figure 1.3. Hox expresssion domains and genomic organisation. Top left panel indicates the expression domains of Hox genes across the anterior-posterior (A-P) axis (anterior to the left) of *Drosophila melanogaster* embryos and top right panel indicates the domains across the mouse embryo. Below is a schematic representing the 3′ to 5′ sequence of Hox genes in flies and mammals, which have four clusters (A, B, C and D) with overlapping expression domains across the A-P axis. Colour coding across the embryos correspond to the different groups of genes, 3′ genes expressed in anterior regions and 5′ progressively more posterior. (Md, mandibular; Mx, maxillary; Lb, labial; T1–T3, thoracic segments; A1–A9, abdominal segments; R1, R4, R7, hindbrain rhombomeres.) (Adapted from (Pearson et al., 2005))

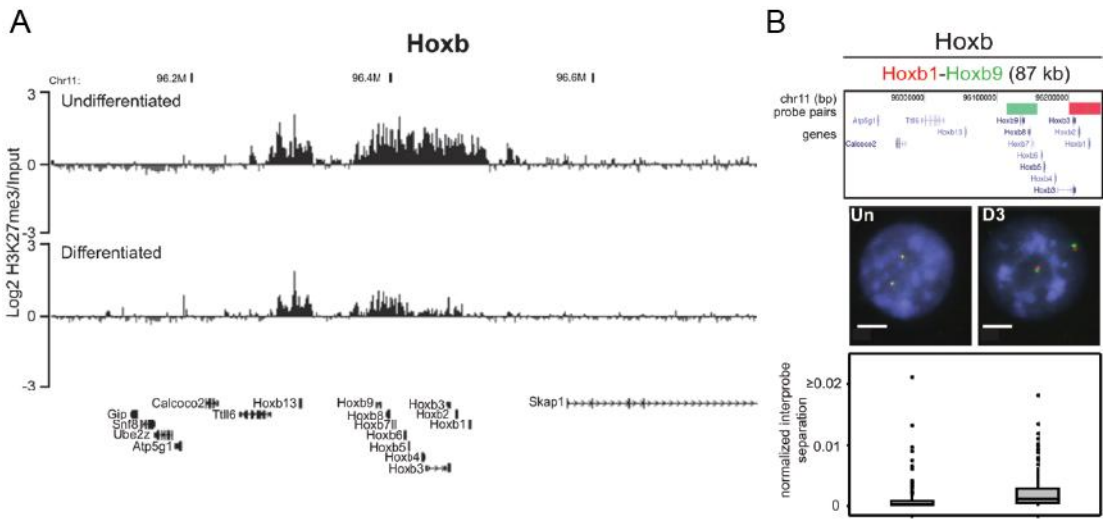

Figure 1.4. Loss of the polycomb-catalysed modification H3K27me3 and chromatin decompaction upon differentiation at the HoxB locus in ES cells. A, In pluripotent ES cells the *Hoxb* genes are silent and the cluster in covered with the repressive chromatin mark H3K27me3, as shown on these UCSC browser tracks from ChIP-chip data. Differentiation down the neuronal pathway results in the progressive 3′ to 5′ loss of the polycomb-catalysed modification, correlating with the order of *Hoxb* expression. **B,** 2D FISH analysis of the HoxB locus reveals a decompaction of chromatin structure over the locus corresponding to the loss of H3K27me3 upon differentiation of ES cells. There was no change in compaction at a control region on the same chromosome. (Adapted from Eskeland et al., 2010.)

1.2 Long-range *cis* regulation: spatial, temporal and tissue-specific control of transcription by regulatory elements

Over forty years ago (Britten & Davidson, 1969) theorised that the orders-of-magnitude differences in genome size from viruses to mammals could in large part be explained by the proliferation of regulatory sequences that drive increased biological complexity. Britten and Davidson's model proposed that there are five classes of elements required for spatial, temporal and tissue-specific coordination of gene expression: *sensor genes* (*cis*-regulatory elements), *integrator genes* (non-coding RNAs), *receptor genes* (gene promoters), *activator RNA* (transcription initiation complexes, including RNA polymerase – the proteins involved in transcription initiation had yet to be elucidated, although the authors do suggest that their proposed activator RNA may code for active proteins), and *producer genes* (protein-coding genes) [terms in italics as used in the paper, with the present terms in parentheses] . Five classes of elements, they suggest, are the minimum necessary to establish specificity of gene control over a genome shared by every cell of a multicellular organism: to respond to an external signal; to produce a second signal; transmit the second signal to receptors that are unable to respond to the first signal; reception of the second signal; and activation and transcription of the producer (protein-coding) gene.

Since its publication much of the conceptual framework about gene regulation by distal genomic elements envisaged by Britten and Davidson has proven to be sound. However, a great deal has still to be determined as to the functional roles of these elements. I will now review what is currently known about the active fields of research into distal *cis*-regulatory elements, describing current hypotheses on regulatory mechanisms and the experimental techniques being used. I also discuss the pre-genomic and genomic methods of identifying distal regulatory sequences, the evidence for these regulatory elements being crucial in evolution and development, and their implication in common/complex human disease.

1.2.1 Functional classification of *cis*-regulatory elements

Genomic regulatory elements other than the promoter have been assigned to one of three functional classes: enhancers, repressors and insulators (enhancer/repressor-blocking and /or barrier/boundary elements). The regulatory mechanisms that employ the first two are similar, whereas insulators have been described as more static genomic elements that establish regulatory domains (Gaszner & Felsenfeld, 2006). Moreover, combinations of regulatory elements that control multiple gene activity within discrete genomic regions have been identified and have been termed locus control regions. These *cis*-regulatory sequences are ~200 – 1000 base pairs (bp) in size and contain multiple transcription factor binding sites (TFBS).

1.2.1.1 Enhancers and repressors

DNA regulatory elements classified as enhancers and repressors regulate target genes in a manner autonomous of location and orientation to transcription start sites; the former having a positive effect on transcription whereas the latter negatively regulate gene activity (Fig. 1.5). Sequences identified as distal regulatory regions do not in themselves regulate gene transcription; rather they form nucleoprotein structures in complex with *trans*-acting DNA-binding factors and co-factors which modify chromatin structure and interact with the basal transcriptional machinery (Arnosti & Kulkarni, 2005). Therefore, whether a distal regulatory element functions as an enhancer or repressor is determined by the active or repressive qualities of the bound factors and co-factors. As enhancer structure and function is covered in greater detail below (sections 1.2.3 and 1.2.4) I will focus on repressor properties. Repressors can be discrete regulatory elements or be contained within a proximal promoter or a distal enhancer (Maston, Evans & Green, 2006). Possible functional models include preventing the binding of a transcriptional activator, as elucidated at the interleukin-4 promoter when the repressor element is bound by the transcriptional repressor BCL-6 (Harris et al., 2005), and instigating a repressive chromatin state through the recruitment of histone-modifying complexes such as polycomb (Srinivasan & Atchison 2004). In *Drosophila* polycomb binds at polycomb response elements (Pirrotta 1997) (Francis & Kingston 2001), conserved DNA elements that apart from

a putative site within the HoxD cluster with similar characteristics (Woo et al., 2010) have yet to be widely characterised in mammals.

Classifying many regulatory elements, which are part of complex transcriptional responses to developmental cues and environmental stresses, as either enhancers or repressors is somewhat arbitrary and simplistic. The reality is more likely to be a range of functional elements from purely active to purely repressive. For instance, the response of regulatory elements to environmental stimuli such as diet, steroid hormones and hypoxia is dependent not only on the cell type but also the transcription factors induced by the stimulus, which subsequently bind to and influence the transcriptional response (Semenza et al., 1991; Katzenellenbogen, 1996; McGrane, 2007).

1.2.1.2 Insulators

In contrast to enhancers/repressors, the activity of an insulator element is position dependent. Insulators possess enhancer-blocking and/or barrier properties (the latter mechanism establishes a boundary between active and repressive chromatin), and have been postulated to have a role in maintaining discrete regulatory domains (Fig. 1.5) (boundary elements involved in the segregation of chromosomes into topologically associating domains (Dixon et al., 2012; Nora et al., 2012)). Insulation is widely accepted as playing an important role in the epigenetic control of gene expression, particularly at imprinted loci (Gaszner & Felsenfeld 2006). Most of our knowledge of insulator function has been derived from transgenic studies, initially in *Drosophila* (Geyer et al., 1986; Bownes, 1990). The HS4 element of the chicken β-globin region was the first vertebrate insulator to be characterised: a compound element with both enhancer-blocking and barrier activity (Chung et al., 1993; Ma et al., 2011).

Although the functional mechanisms of enhancer-blocking elements have still to be fully elucidated, these insulators do need to lie between an enhancer/repressor and gene promoter(s) (Fig. 1.5). One aspect of enhancer-blocking activity may be the formation of chromatin loops that result in the positioning of enhancers and promoters in different chromatin domains, a mechanism common to vertebrates and

invertebrates (Yusufzai et al., 2004). In *Drosophila* the key component of loop formation is the suppressor of hairy wing (Su(Hw)) protein (Gerasimova et al., 2000). In mammals, CTCF proteins have been identified at many enhancer-blocking sites (Chung et al., 1993; Bell et al., 1999; Filippova et al., 2001; Chao et al., 2002). Like Su(Hw) CTCF is a large zinc finger protein and is proposed to possess similar properties: the capability to form clusters and create closed loops, or possibly to tether chromatin fibre to the nucleolar surface. Furthermore, there is evidence of a role for cohesin – a component of the structural maintenance of chromosome (SMC) complex that is essential for chromosome mechanics – in facilitating intrachromosomal loops between CTCF-binding elements at several loci of the mammalian genome (Hadjur et al., 2009; Mishiro et al., 2009; Nativio et al., 2009; Hou et al., 2010).

Barrier elements prevent stochastic, meta-stable, heritable silencing of genes situated in euchromatic domains by the advancement of heterochromatic regions – a process known as position-effect variegation (PEV) (Gaszner & Felsenfeld, 2006). In yeast, barrier activity has been linked to nucleosomal removal (Bi et al., 2004) and the recruitment of HATs and ATP-dependent nucleosome-remodelling complexes (Oki et al., 2004). Analysis of the chicken HS4 insulator suggests a similar mechanism operates in vertebrates: the element is situated at a nuclease-hypersensitive site with adjacent histones heavily marked with euchromatin modifications (H3K4 methylation, H3 and H4 acetylation) due to the recruitment of HATs and HMTs (Litt et al., 2001a; Litt et al., 2001b). Ubiquitination of H2B (H2Bub1) is also required to maintain the barrier properties of HS4; depletion of the E3 ligase RNF20/BRE1A (which mediates H2B ubiquitination at HS4) results in loss of the active chromatin modifications at the insulator and a consequent encroachment of heterochromatin modifications (Ma et al., 2011).

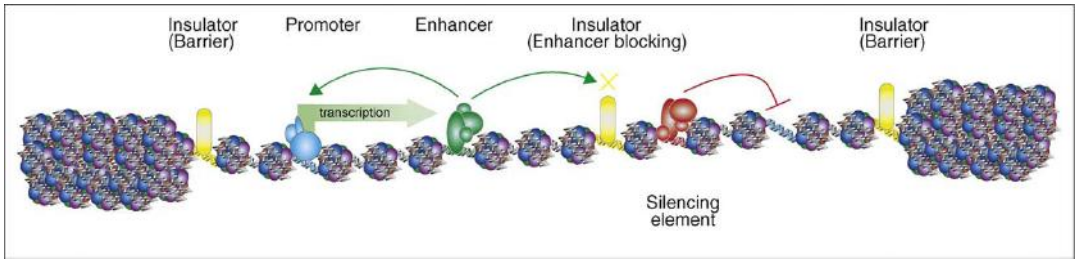

Figure 1.5. Functional classes of *cis*-regulatory elements. Transcription initiation at promoters (blue DNA) may be activated by enhancers (green DNA) or repressed by silencers (red DNA). Enhancer and silencer activity may be restricted by position-dependent insulator elements (yellow DNA), which also form barriers that prevent the spreading of repressive heterochromatin. Blue ovals are transcription factors, activating and repressing proteins are green and red ovals respectively and yellow ovals are insulating proteins. (From Heintzman and Ren, 2009.)

1.2.2 Methods of Enhancer Identification

Up until the availability of comparative genomic methods and genome-wide identification of histone modification marks and bound factors /cofactors associated with enhancers, the identification of *cis*-regulatory elements was restricted to individual genes or gene clusters and the surrounding genomic region; time-consuming and labour-intensive efforts that in many cases were initiated following the establishment of non-coding deletions/chromosome rearrangements/single nucleotide polymorphisms (SNPs) as causes of human disorders (Section 1.2.6). Translocations at the β-globin locus that displaced non-coding regions causing β-thalassemia (Kioussis et al., 1983; Semenza et al., 1984), identified distal sequences that were subsequently shown to have enhancer functions in transgenic mice (Grosveld et al., 1987). Similarly, a translocation breakpoint mapped to a locus 1 Mb from the sonic hedgehog (*Shh*) gene in a human patient with the congenital limb malformation preaxial polydactyly (PPD) corresponded to the transgenic insertion point in a polydactylous mouse known as Sasquatch (*Ssq*) (Lettice et al., 2002). Analysis showed that both chromosomal lesions occur in a similar location within the *Lmbr1* gene and genetic analysis revealed that the *Ssq* mutation interrupts a long-range *cis*-acting regulator of *Shh* (Fig. 4.5 top left-hand panel shows the *Shh-Lmbr1* region on mouse chromosome 5, and the bottom second panel from the left shows the highly conserved regulatory element (ZRS) located within intron 5 of *Lmbr1*). A highly conserved element was identified and further sequence analysis identified point mutations within this regulator and transgenic assays confirmed the regulatory sequence as a limb-specific enhancer that drives *Shh* expression (Lettice et al. 2003; Sagai et al., 2005).

The availability of sequenced genomes for an ever-expanding number of organisms has facilitated the search for regulatory elements on a genome-wide scale. This comparative genomic strategy assumes that functional non-coding elements are under selection, just as coding sequences are. One of the first papers using comparative genomics focussed on identifying enhancers in two large gene deserts (a genomic environment in which a large proportion of developmental genes reside, section 1.2.5) adjacent to the human *DACH1* gene (Nobrega et al., 2003). Thirty two

genomic regions conserved across to pufferfish (*Fugu rubripes*) were identified, and 7 of the 9 sequences tested in mouse transgenic assays were able to drive reporter gene expression in the same tissues as endogenous *DACH*. In a whole genome study comparing human and pufferfish genomes, 1400 highly conserved non-coding elements (HCNEs) were identified; the majority situated in and around developmental genes (Woolfe et al., 2005). The authors tested twenty five of these elements, located within the genomic regions of four developmentally important genes, 23 of which showed enhancer activity in transiently transfected zebrafish embryos (Ancora http://ancora.genereg.net; Engstrom et al., 2008).

Comparative genomic strategies alone cannot identify all enhancers (Blow et al., 2010). Transgenic assays can show that HCNEs have the potential to function as enhancers but at the present time can only be used to test a small number of sequences. Furthermore, those sequences that do not drive reporter expression cannot be discounted as enhancers as they may be active at other stages in development. A further note of caution in presuming that all conserved sequences are crucial for viability was raised by the deletion of four ultraconserved elements (sequences of >200 that show perfect conservation in the human, mouse and rat genomes) within sequences that show enhancer activity in transgenic assays (Ahituv et al., 2007). All four resulting mouse lines showed no severe phenotypic effects and were viable.

To complement these strategies, ChIP combined with arrays (ChIP-chip) or deep sequencing (ChIP-seq) has been used to profile histone modifications and variants that mark enhancers, indicate an open chromatin state, or identify regions bound by transcription factor or co-factors associated with enhancers. As described in sections 1.1.1.1 and 1.1.1.3, H3K4me1/2, H3K27ac, H2A.Z, DNase I sensitivity and loss of nucleosome occupancy are all markers of active enhancers (Heintzman et al., 2007; Barski et al., 2007; ENCODE Project Consortium, 2007; Xi et al., 2007; Heintzman et al., 2009; He et al., 2010). These studies all demonstrate that enhancer activity is tissue or cell-type specific. Moreover, two recent studies propose that the H3K27ac modification marks active enhancers, whereas H3K4me1 is present at both active and poised enhancers (Creyghton et al., 2010; Rada-Iglesias et al., 2011). And

enhancers have been suggested to be switched off in development by the histone demethylase LSD1 (Whyte et al., 2012).

Although potential non-coding regulatory elements have been identified through specific transcription factor-binding (Johnson et al., 2007; Chen et al, 2008; Wederell et al., 2008), this is not a practical strategy for identifying enhancers genome-wide due to the sheer number of mammalian transcription factors, many of which are tissue-specific and are activated at specific temporal stages during development. Transcriptional co-regulators, such as chromatin remodelers and mediators, are less numerous and many have been shown to be recruited specifically to enhancer sequences (Heintzman and Ren, 2009). Genome-wide occupancy of the transcriptional co-factor p300 (an acetyltransferase component of enhancer-associated protein complexes, Section 1.1.1) was assayed in several tissues from E11.5 mouse embryos and was shown to have a 5-16-fold better enhancer prediction rate compared with comparative genomics (Visel et al., 2009). The advantages of this approach are illustrated by the identification of > 3000 heart enhancers in E11.5 embryos, a large proportion of which are only weakly conserved (Blow et al., 2010).

Although multiple species comparisons and chromatin profiling has enabled the identification of many putative regulatory sequences and the testing of a selection of these in transgenic assays, the functional capability of the vast majority of these novel regulatory elements awaits confirmation *in vivo*, for instance by their targeted disruption in the genome. And, of course, these studies do not tell us anything about how enhancers function that is, how they target and activate specific genes, particularly for very distal enhancers that work over large genomic distances (Sections 1.1.2, 1.2.4, 1.2.5).

1.2.3 Enhancers in Evolution and Development

How did such *cis*-regulatory sequences evolve? Although speculative, their rise alongside the evolution of multicellular life is a reasonable proposition. There are several scenarios as to how non-coding regions arose. Whereas simple elements that bind few factors may have arisen *de novo* by small base changes, it seems unlikely

that complex enhancers could have arisen this way. Other possible routes are; the exaptation of retroposons to form clusters of protein-binding motifs at new genomic locations (Santangelo et al., 2007; Sasaki et al., 2008), or the acquisition of novel regulatory functions by promoters of archaic genes whose coding sequences have been subsequently lost. Existing enhancers may also evolve to acquire new functions in development.

What are the evolutionary consequences of mutation/deletion of existing enhancers? Ever since the first characterisation of a *cis*-regulatory sequence – the *lac* operon (Monod and Jacob, 1961) – evolutionary biologists have argued that regulatory mutations have had a major impact on phenotypic evolution (Wray, 2007). Whereas alteration of coding regions may have quite blunt effects on protein function, alterations at enhancers have the potential to subtly adjust levels of gene expression – a situation required for the evolution of quantitative traits and response to environmental stresses (Williamson et al., 2011). Moreover, the phenotypic effects of mutating enhancers can be confined to discrete tissues or developmental stages and so could allow a pre-existing gene to be harnessed to a new developmental scenario (Reibez et al., 2009). For instance, neural-specific activation of *Lunapark* and *Evx2*, situated between the HoxD cluster and the global control region (GCR), is regulated by enhancer elements within the GCR (Fig. 1.10). Co-option of the GCR to then facilitate *Hoxd13-10* expression during limb development (discussed further in Sections 1.3 and 1.5) may have been due to the accrual of mutations to produce an element able to function in the developing limb (Gonzalez et al., 2007).

A limb-specific enhancer of *Prx1* – which promotes bone elongation in the forelimb – is responsible for higher levels of transcription and more extensive expression domains at key limb bud locations and embryonic stages in the fruit bat (Cretekos et al., 2008.). Replacing the endogenous mouse enhancer with the bat enhancer culminated in mouse embryos with longer forelimbs but no morphological effects in other organs where *Prx1* is expressed. Similarly, in transgenic mouse assays human-specific point mutations within a highly conserved non-coding element (HACNS1) confer a limb expression pattern, including the presumptive anterior wrist and proximal thumb, that appears human-specific compared to that

conferred by orthologous sequences from non-human primates (Prabhakar *et al*, 2008). These studies highlight the significance of *cis*-regulatory mutations for the divergence of morphological traits in closely related species. Morphological differences can also result from the loss of tissue-specific enhancers. Human-specific deletions of mostly non-coding sequences that are highly conserved between chimpanzees and other mammals include regions near the androgen receptor locus and the neural-specific *GADD45G* (Mclean et al., 2011). These contain tissue-specific enhancers whose loss correlates with human-specific traits – the loss of secondary sexual characteristics associated with androgen hormone signalling and the expansion of specific brain regions respectively.

An intriguing aspect of the *cis*-regulation of many developmental control genes is the role of 'shadow' enhancers. These are regulatory sequences that each can drive expression of the same gene in overlapping domains. Although mostly characterised in *Drosophila* larval development, several mammalian genes (e.g *Shh* and *Sox10*) have been described with redundant enhancer activity (Jeong *et al*, 2006; Werner *et al*, 2007.) Enhancer redundancy likely ensures that the control of gene expression is robust, even in conditions of genetic or environmental stress.

1.2.4 The genomic context of enhancers

A sizeable proportion of enhancers are situated tens-to-hundreds (even thousands) of kilobases from their target genes, often in stable gene deserts: expansive gene-poor regions that tend to resist chromosomal rearrangements (Ovcharenko et al., 2005). Indeed, many genes with tight spatial and/or temporal expression domains – such as developmental genes – tend to be located adjacent to gene deserts. The absence of other neighbouring genes could help ensure regulatory specificity between enhancer and promoter. An example of problems that can arise from the promiscuous action of enhancers placed within gene-rich domains is illustrated by the ectopic activation of neighbouring genes when the β-globin locus control region (LCR) is integrated into a gene-rich domain of housekeeping genes (Section 1.1.2) (Noordermeer et al., 2008).

However, there must be specificity between some enhancers and promoters. The sonic hedgehog gene (*Shh*), which codes for a morphogen directing cell fate in several tissues during organogenesis, is adjacent to a large gene desert containing at least three enhancers that are active only within *Shh*-expressing tissues of the embryonic brain (Fig. 1.6A) (Jeong et al., 2006). But, *Shh* expression during limb development is regulated by another enhancer, known as the ZRS (zone of polarizing activity regulatory sequence), situated about one megabase away, beyond the neural enhancers and within an intron of another gene (Lettice et al., 2003) (Fig. 1.6A). The ZRS bypasses both the gene in which it is located and a neighbouring gene in order to activate only *Shh*.

1.2.5 How do long-range enhancers activate target genes?

Enhancers are clusters of cognate binding sites for transcription factors (TFs). There have been three, not entirely mutually exclusive, mechanisms proposed for enhancer-gene communication (Figure 1.6B). In the first, sequence-specific factors bound to the enhancer and/or promoter, recruit further factors – perhaps even including RNA polymerase itself – to reorganise chromatin throughout the intervening region as the signal for gene activation (Bulger and Groudine, 2010). Due to the large genomic distances involved, and even intervening promoters, this seems unlikely to be a feasible mechanism for very long-range control, such as that at *Shh*.

The second mechanism of action also involves reorganizing the intervening chromatin, but to allow for direct enhancer-promoter interaction – possibly by formation of a series of "mini" chromatin loops (Montavon et al., 2011).

The third mechanism invokes the spatial co-localisation in the nucleus of enhancer and promoter, which can then directly interact with each other if the necessary factors are bound, looping out the intervening chromosomal region. This looping may involve the delivery of RNA polymerase to the promoter, or even the delivery of factors to de-repress the target gene, for instance a histone demethylase to remove polycomb (Vernimmen et al., 2011). One attractive feature of these models is that it ensures that an enhancer can activate the expression of only one gene at a time.

However, enhancer-promoter mediated looping may be insufficient for gene activation. Spatial co-localisation of the ZRS and *Shh* is reported to occur in nuclei in both expressing and non- expressing halves of the developing limb-bud, but on the posterior side of the limb-bud the active *Shh* locus is, in addition, also extruded out from its chromosome territory (Amano et al., 2009). A similar looping out from chromosome territories is induced by an ectopic β-globin LCR (Noordermeer et al., 2008).

Does there even need to be direct physical interaction between enhancer and promoter? It is feasible that the clustered binding sites for transcription factors at enhancers act to reduce the effective nuclear search space of these proteins for a gene promoter by diffusion/non-specific binding. (Elf J et al., 2007). In this model, enhancer-promoter specificity would rely upon shared high affinity binding sites for the same set of factors. Moreover, enhancers themselves often bind RNA polymerase II (De Santa et al., 2010) and are transcribed to produce short non-coding RNAs (Kim et al., 2010). Whether these RNAs have a functional role has still to be determined. A long non-coding RNA (lncRNA) has been described that influences the neural-specific activity of a developmental enhancer. *Evf2* is a ncRNA that is transcribed across an ultraconserved enhancer situated between the *Dix5* and *Dix6* homeodomain genes and cooperates with the *Dix2* protein to increase the activity of the enhancer in a target and homeodomain specific manner (Feng et al., 2006). The authors speculate that the role of *Evf2* is to enable *Dix2* to bind the enhancer and suggest that developmentally regulated ncRNAs may function in *trans* to regulate the transcriptional activity of homeodomain proteins. The lncRNA *HOTAIR* is located in the HOXC locus and has been shown to have a repressive influence on the 5′ *HOXD* genes in human fibroblasts (Rinn et al., 2007).

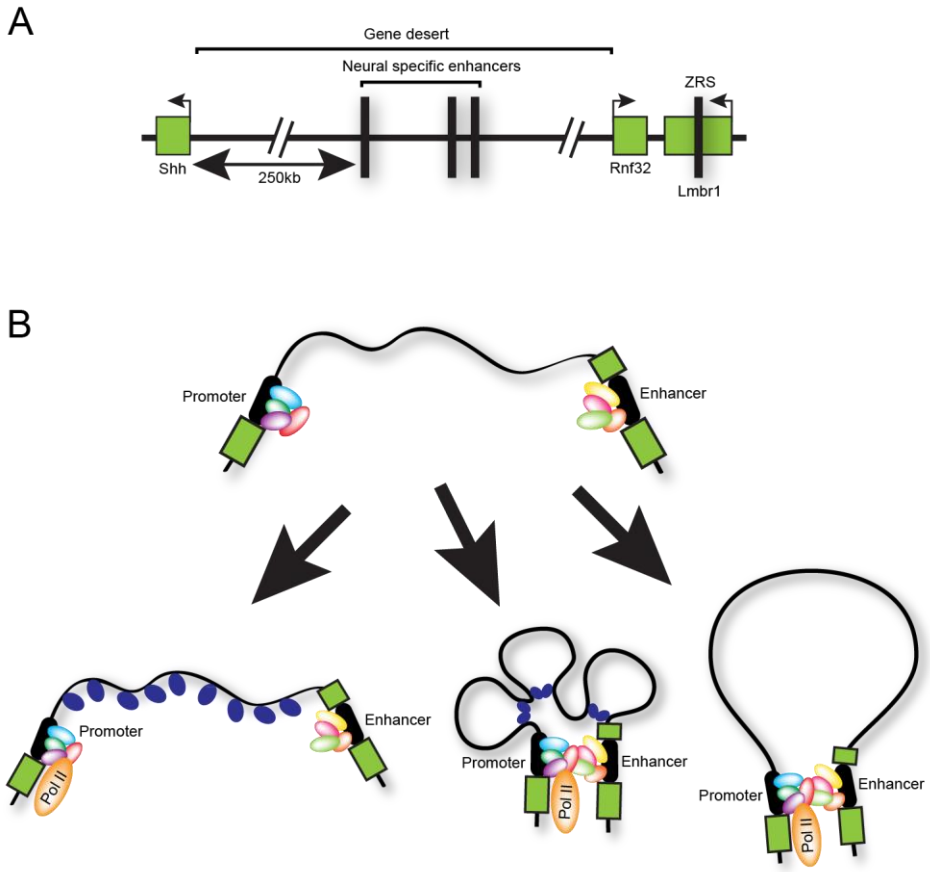

Figure 1.6. The sonic hedgehog regulatory region and possible mechanisms of enhancer activity. A, Linear representation of the *Shh* locus showing the location of long-range neural-specific and limb-specific enhancers. **B,** Models of enhancer-promoter communication. Factors bound to the enhancer and promoter recruit chromatin-modifying factors that reorganise the intervening chromatin with the factors themselves being the signal for gene activation or enabling direct enhancer-promoter interaction by the formation of chromatin “mini-loops”; or, spatial co-localisation of the enhancer and promoter, which can interact if the necessary factors are present, with a looping out of the intervening chromosomal region.

To initiate the transcription of target genes enhancers must integrate and process cellular signalling information. Two models of information processing have been proposed: "enhanceosome" and "billboard" (Arnosti & Kulkarni, 2005). In the former there is a high level of cooperative and coordinate action between the bound transcription factors to give unitary outputs; while in the latter the proteins do not operate as a single unit, but rather form multiple functional units that independently regulate gene expression. So, whereas mutations to individual enhanceosome binding sites may ablate enhancer function, the more flexible characteristics of billboard style enhancers suggests that alteration to individual binding sites would be less drastic to overall enhancer function and therefore more amenable to evolutionary adaptation.

1.2.6 Enhancers and disease

Mutations associated with β-thalassaemia were what first led to the identification of the β-globin LCR and deletions that affect an enhancer's function, or translocations that separate enhancers from their cognate promoters, have been identified as the cause of several simple Mendelian human genetic diseases (Noonan & McCallion, 2010). Deletion of a bone-specific enhancer of *SOST* (a negative regulator of bone formation) has been identified as the cause of the homozygous recessive disorder Van Buchem's disease (Loots et al., 2005). Similarly, the deletion of a regulatory sequence around 1Mb upstream of *POU3F4* is responsible for some cases of the Mendelian disorder X-linked deafness type 3 (Noonan & McCallion, 2010). Translocations over a > 1Mb genomic region that cause disruption of *SOX9* expression are associated with several congenital skeletal malformations and implicate the loss of enhancers that regulate tissue-specific expression of the gene (Gordon et al., 2009). Point mutations in the *Shh* ZRS result in a gain of function of enhancer activity and ectopic activation of *Shh* expression in the anterior limb-bud – the underlying cause of PPD (Lettice et al., 2003). Human limb abnormalities can also result from translocations around HOXD (Spitz et al., 2002). A balanced translocation 60 kb from the cluster is associated with mesomelic dysplasia, which is a shortening of the forearms and forelegs.

Almost a half of single nucleotide polymorphisms (SNPs) that show statistical associations with common or complex human disease in genome-wide association studies (GWAS) are within non-coding regions and gene deserts and so potentially involve enhancers (Noonan and McCallion, 2010). For example, a SNP within an intronic enhancer of the *RET* gene confers a greater risk of Hirschsprung disease than the major allele (Emison et al., 2005). SNPs in a gene desert 330kb upstream of the *c-myc* protooncogene are associated with greater risk of several cancers (Wasserman et al., 2010). Two studies (Tuupanen *et al*, 2009; Pomerantz *et al*, 2009) established the mechanism by which one of the SNPs in this gene desert results in colorectal cancer, revealing that it is located within an enhancer and increases its binding affinity for the TCF4 transcription factor, which activates genes through the Wnt signalling pathway. This enhancer was shown to physically interact with the *MYC* oncogene in colorectal cancer cell lines but not in fibroblast (control) cells. A SNP within an enhancer for the pigment gene *OCA2* reduces the capacity of the enhancer to interact with the *OCA2* promoter which directly influences melanocyte pigmentation (Visser et al, 2012). Why might so much of the genetic variation apparently associated with complex human disease be in enhancers? Whereas point mutations in protein coding regions can completely ablate gene function and so result in strong and penetrant phenotypes, point mutations in an enhancer may alter binding affinity for a specific TF and so could result in only subtle changes in the level, time, or place of gene expression. The resulting phenotype could therefore be more modest and subject to interaction with other genes and with the environment.

1.2.7 Mechanisms of long-range enhancer function have still to be fully elucidated

Section 1.2.5 makes clear that a comprehensive understanding of the mechanisms of long-range regulation has yet to emerge, although several possible mechanisms have been proposed based on the available evidence. Our current lack of mechanistic understanding about how enhancers regulate temporal and spatial patterns of expression from the correct target gene is impeding the exploitation of the massive

investments being made in GWAS and the functional analysis of human disease. What needs to be done to better understand how enhancers work? The dearth of knowledge on enhancer biochemistry needs to be addressed in order to gain a more complete understanding of how enhancers integrate and process signalling information that promotes transcription. Enhancer elements function as nucleoprotein complexes, with multiple associated sequence-specific and sequence non-specific binding proteins. What are these proteins and how do they act together? There could be a high level of cooperative and coordinate action between the bound TF to give unitary outputs to the transcriptional machinery, or they may rather form multiple functional units that each independently regulate gene expression. (Arnosti and Kulkarni, 2005).

The precise chronological sequence of chromatin events at enhancers is also still to be elucidated. Does any direct physical enhancer-promoter interaction only occur once factors are bound to both elements, or does the interaction itself contribute to the loading of TFs or transcriptional machinery to the promoter? Current assays for long-range physical interactions – fluorescent *in situ* hybridization (FISH) and chromosome conformation capture (3C) and its derivatives – only give snapshots or cell population averaged information, respectively. There is a need for dynamic assays of chromosome organisation; to capture long-range interactions as they occur in real time and for these to be coupled to methods for quantifying levels of gene expression in single cells, at defined times and places in differentiation and development.

1.3 Growth and development of the vertebrate distal limb

Digits are the main anatomical feature of the distal vertebrate limb (autopod). The patterning of the digits, and the processes that regulate the growth and form of the autopod, are a paradigm of organogenesis. Polarisation of the distal vertebrate limb across the anterior-posterior (A-P) axis (thumb/radius to the minimus/ulna) is a morphologically recognisable consequence of the signalling interaction of two organising centres, backed up by several inter- and intra-cellular systems, that ensure limb asymmetry. Indeed, the establishment of the vertebrate body plan relies on the control of polarity. The signaling molecule sonic hedgehog (Shh) is involved in regulating the polarity of many tissues during development (Echelard et al., 1993; Haraguchi et al., 2001). During organogenesis, activation of Shh is tightly regulated and expression is restricted to specific spatio-temporal patterns (Masuya *et al*, 2006). Early work on limb development revealed that the posterior mesenchyme of the early limb bud is the location of one of the organizing centres, known as the zone of polarizing activity (ZPA) (Fig. 1.7B) (Saunders and Gasseling, 1968). Shh was subsequently identified as the secretory protein responsible for activating the polarizing activity (Riddle et al., 1993).

Although the early work that identified the ZPA – and how these cells were the source for A-P patterning instructions – was carried out in chick limb buds, much of the more recent work has utilised the powerful new tools of mouse molecular genetics (Robert and Lallemand, 2006). These tools include fine-tuned modulation of gene expression, intra-chromosomal deletions and duplications, and lineage analysis. As the model for this project is mouse limb bud cell lines, the following description of distal limb development is mostly based on the mechanisms and timing that occurs during mouse limb growth.

1.3.1 Early limb outgrowth and the defining of the A-P axis

At embryonic day (E) 9.0 limb budding occurs, and the limb A-P axis is initially determined by two mutually antagonistic patterns of gene expression (Fig. 1.7A). Pre-patterning of limb asymmetry depends on the expression of *Hand2* in the

posterior limb bud and the truncated form of *Gli3* (Gli3R) in the anterior (Robert and Lallemand, 2006; Hill, 2007). *Hand2* induces *Shh* expression either along with or through the 5′ *Hoxd* genes, thereby positioning the ZPA at the posterior margin of the limb bud. In the anterior half of the developing limb bud Gli3R is involved in the activation of *Alx4*, which inhibits the expression of *Shh*-activating genes and is therefore crucial to the asymmetrical pre-patterning of the developing limb (Hill, 2007).

Early onset *Hoxd* expression in the limb culminates in the most 5′ genes (*Hoxd11-13*) having restricted, posterior expression domains (Fig. 1.8). Through a series of deletions and duplications within the HoxD cluster that re-located individual genes – thereby altering their order and spatial position within the cluster – Tarchini and Duboule (2006) elucidated the collinear mechanism of spatial and temporal control of this initial wave of *Hoxd* expression. Two regulatory elements – an enhancer located at the telomeric end of the *Hoxd* cluster (ELCR) and a repressor element (POST) at the centromeric end – combine to ensure that the 5′ genes are spatially restricted to the posterior limb bud, and temporally restricted to between E9.5 and E10.0, the location and time of *Shh* activation. The ELCR defines the timing of activation: 3′ genes such as *Hoxd1* are closest to the enhancer and are expressed earliest – at E9.0 and throughout the limb bud – followed in a temporal collinear manner by successive genes within the cluster. Meanwhile, the POST element also influences expression in an asymmetrical manner: only the 5′ genes have posteriorly restricted expression domains and the extent of restriction depends upon position within the cluster relative to the 5′-located repressor. Gli3R may mediate this repression as the 5′ genes are activated in the anterior limb upon inactivation of *Gli3* (Zuniga and Zeller, 1999; te Welscher et al., 2002). A hitherto unknown role for the polycomb methyltransferase subunit Ezh2 in limb bud pre-patterning was recently established. The conditional inactivation of *Ezh2* resulted in the posteriorization of limb bud identity with the loss of *Gli3* and *Alx4* and ectopic expression of *Hand2* and the 5′*Hoxd* genes in the anterior limb (Wyngaarden et al., 2011). Posterior-restricted expression of *Hoxd11*, *Hoxd12* and *Hoxd13* is crucial to ensuring limb A-P polarization: when their expression domains are anteriorized they

can induce ectopic *Shh* expression, which leads to the loss of limb asymmetry (Knezevic et al., 1997; Zakany et al., 2004; Tarchini et al., 2006).

The mutual antagonism of the *Hand2* and *Gli3* (R) expression domains then ensures that *5′Hoxd* gene activity is confined to the posterior margin of the limb bud during the initial phase of outgrowth and A-P polarization (E9.0 – E10.0). This polarity consequently ensures that activation of *Shh* and generation of the ZPA occurs at the posterior limb bud only. However, *Shh* has the potential to be anteriorly expressed, and Fgf-regulated Etv4/5 (two members of the PEA3 sub-family of Ets transcription factors that have redundant activity in the limb) have been identified as direct repressors of *Shh* in the anterior limb bud (Mao et al., 2009; Zhang et al., 2009). This implicates the proximal-distal (P-D) organizing centre the apical ectodermal ridge (AER) in not only interacting and mutually maintaining the ZPA (see below) but also actively preventing anterior *Shh* expression. In the latter study conditional inactivation of Etv4/5 in mesoderm-derived cells had no effect on P-D development but all the resulting embryos exhibited a preaxial polydactyly (PPD) phenotype. PPD is a congenital abnormality that affects the digits on the anterior side of the distal limb (I discuss the link between Shh and PPD further in Section 1.4); and ectopic, anterior expression of *Shh* was detected at E11.0 in mutant embryos. Both papers provide evidence that Etv4/5 inhibition of *Shh* is required at the beginning of limb bud development. Activation of the Etv inhibitory form at E9.0 Mao et al. (2009) resulted in ectopic expression of *Shh*, whereas upon its activation at E10.0 *Shh* expression remained restricted to the posterior margin. And in Zhang et al (2009) complete Etv inactivation was not observed until E10.0, the likely consequence of ectopic *Shh* activity and PPD only occurring in the mutant hindlimbs as hindlimb development begins at a later stage than forelimb.

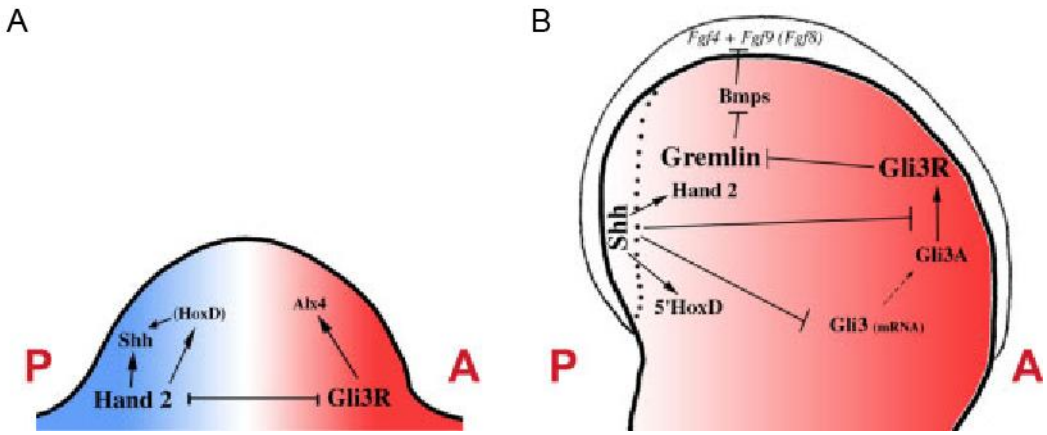

Figure 1.7. Genetic interactions during limb bud development. A, At E9.0 mutual antagonism between *Hand2* and *Gli3* (both encode transcription factors) determines the anteroposterior axis. *Gli3* is expressed in the anterior half, mostly in its processed Gli3R form and controls the expression of *Alx4*. *Hand2* either directly induces *Shh* expression or indirectly through the 5′ *Hoxd* genes in the posterior half of the limb bud. **B,** At E11.0 *Shh* maintains *Hand2* expression, plays a role in instigating a second wave of 5′ *Hoxd* gene expression crucial for digit patterning, and regulates Gli3R levels. The Bmp antagonist *Gremlin*, which is repressed by Gli3R, is expressed only in the posterior limb bud, and this prevents the down-regulation of the *Fgfs* at the AER. This indirect interaction of *Shh* with the *Fgfs* completes a feedback loop which also involves the maintenance of *Shh* expression by the *Fgfs*. (From Robert and Lallemand, 2006.)

1.3.2 Development of the vertebrate autopod

In the mouse initial growth and patterning of the distal handplate takes place from E10.0 - E12.5. This period of limb development is defined by the expression of *Shh* and a second wave of 5′ *Hoxd* gene expression under the control of long-range enhancers that both determine levels of expression and influence spatio-temporal expression domains (I discuss both enhancers in more detail in Sections 1.4 and 1.5). At the end of this crucial period the digits have been specified and can be identified within the distal limb bud. Subsequent development involves mostly chondrogenesis and the growth of pre-patterned limb structures. A multifaceted interplay of regulatory and signaling networks direct autopod specification and development, the complexity of which has still to be fully unravelled. However, I will briefly describe the salient aspects of the role played by the key players.

A signaling feedback loop between the ZPA and the AER maintains *Shh* expression (Fig. 1.7B). Signals that emanate from the AER are required for outgrowth of the limb along the P-D axis. The AER is situated in a strip of epithelial cells at the distal tip of the limb bud and is the source of fibroblast growth factor (Fgf) signaling in the limb, with Fgf4 and Fgf8 being the major factors involved in P-D patterning (Sun et al., 2002, Mariani et al., 2008). Fgf signaling is also required for the maintenance of *Shh* expression, while *Shh* indirectly maintains *Fgf* expression (Fig. 1.7B). Gli3R levels are regulated by Shh repression of *Gli3* transcription in the posterior limb bud and by restricting the processing of Gli3A to Gli3R to the anterior (Hill, 2007). Shh maintains *Hand2* expression in the posterior limb, and also *Gremlin*, a Bmp antagonist, to prevent the inhibition of Fgf signaling and the Bmp-induced regression of the AER (Laufer et al., 1994; Niswander et al., 1994; Zuniga et al., 1999; Benazet et al., 2009). Therefore, this feedback loop, linking the epithelium and the mesenchyme, co-ordinates A-P and P-D patterning and ensures correct distal growth.

Shh plays a direct role in digit specification and growth. The classic concentration morphogen model posits that digit identity is dependent on distance from the ZPA through the formation of a Shh concentration gradient across the presumptive handplate, with only digit 1 (thumb) not specified in some way by *Shh*.

However, experiments that identified the fate of *Shh*-expressing cells and cells responding to Shh signalling suggest a temporal dynamic needs to be included in a model that accounts for digit identity (Ahn and Joyner, 2004; Harfe et al., 2004). In these studies digit 2 condensation is specified by Shh signalling, whereas digits 3, 4 and 5 (most posterior) derive from the ZPA and their identity is established from progressively increasing periods of exposure to Shh.

More recent findings, in the mouse and chick, have revealed a direct link between Shh-mediated growth and patterning: the growth-morphogen model. In the chick, both inhibition of Shh signalling and blocking of cellular proliferation reduced limb bud width and, as a consequence, the number of digits (Towers et al., 2008). Crucially, though, whereas inhibition of Shh resulted in the formation of anterior digits, blocking of cellular proliferation (with *Shh* expression maintained, albeit at slightly lower levels) led to posterior digit formation (digits 3 and 4, or 4 alone). These results suggest that Shh-dependent limb bud expansion across the A-P axis is necessary to separate digit primordia for normal specification in the chick, and therefore correct *Shh* expression is crucial for digit specification and proliferation of limb mesenchymal cells. Studies in the mouse corroborate the chick study conclusions; nonetheless with some salient disparities (Zhu et al., 2008). In mouse limb development each digit has a distinct temporal requirement for Shh signalling: early (E9.5) inactivation culminated in the condensation of digit 4 only; inactivation at subsequently later stages allowed the condensations of digits 2 then 5. Therefore, the order of digit development is opposite to the susceptibility of digit loss (digit 3, then 5, 2, 4). Shh controls cell proliferation (and survival) and as a consequence signalling duration influences digit number, as in the chick. However, unlike in the chick, reduced exposure time does not result in the anteriorization of digit identity, but rather the digits maintain their identity; an outcome that the authors argue is due to digit specification in the mouse occurring at an early developmental stage. So these data imply that, in the mouse, Shh specifies A-P patterning early on in limb development and is required at later stages to promote cell cycle progression to ensure there are enough cells for complete formation of the digits.

Late onset expression of the 5′ Hoxd genes (*Hoxd13* to *Hoxd10*), by a process termed "reverse collinearity" (Nelson et al., 1996), is initiated in the *Shh*-expressing region of the posterior limb bud from around E10.5 until E12.5 (Fig. 1.8). The most 5′ *Hoxd* gene, *Hoxd13*, has the strongest expression then levels progressively decrease down to *Hoxd10*, and this varying level of expression ensures correct digit morphogenesis (Kmita, et al., 2002; Montavon, et al., 2008). Late onset 5′ *Hoxd* gene expression is controlled by enhancers located centromeric of the cluster, namely the GCR and Prox (Spitz et al., 2003; Gonzalez et al., 2007), and position within the cluster determines the strength and extent of each gene's expression domain across the distal plate (Tarchini and Duboule, 2006) (Fig. 1.8) *Hoxd13* is the only *Hoxd* gene that is expressed at the extreme, anterior, distal plate within the cells that give rise to presumptive digit 1 (thumb in humans) (Montavon et al., 2008).

My research has focussed on how *Shh* and the 5′*Hoxd* genes are regulated during this phase of mouse distal limb development, and in particular the role of the two long-range enhancers – ZRS and GCR. Therefore, I will end this introduction to my thesis by describing what is known about both these elements in more detail before listing the key aims upon which I based my research projects.

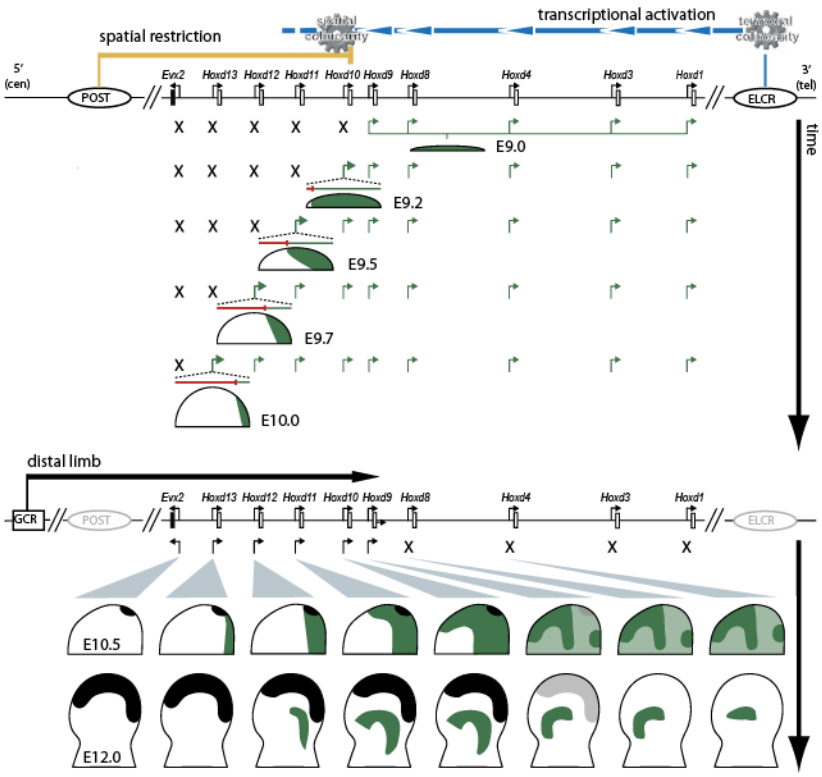

Figure 1.8. *Hoxd* expression occurs in two waves during limb development. The first wave (from E9.0 to E10.0) is temporally and spatially controlled by two regulatory elements either end of the cluster (ELCR and POST). Expression occurs in a collinear (*Hoxd1* to *Hoxd13*) manner. Late onset expression involves the 5′ genes only and is under the control of the GCR and Prox enhancers. *Hoxd13* is expressed first and to the greatest extent, with *Hoxd12*, *11* and *10* expressed in decreasing levels. This process has been termed “reverse collinearity”. (Adapted from Tarchini and Duboule, 2006.)

1.4 ZRS

The zone of polarizing activity regulatory sequence (ZRS) is the long-range *cis*-regulator of sonic hedgehog (Shh) in the limb. The enhancer lies up to 1Mb away from its target in an intron of the *Lmbr1* gene, and an 800bp sequence of the enhancer is evolutionary conserved across to chondrichthyans (Dahn, *et al*, 2007). Transgene assays demonstrated that the highly conserved ZRS sequence from *Fugu* is able to induce a ZPA in mouse (Lettice, *et al*, 2003). Intriguingly, the ZRS exerts no control over either *Lmbr1* or *Rnf32*, which lie an order of magnitude closer to the enhancer than *Shh*. Therefore, the active ZRS bypasses the gene it is located within and an adjacent gene and is able to traverse an 800kb gene desert to induce *Shh* expression (Figure 1.9).

The ZRS has a functional role in the instigation of *Shh* expression along the distal posterior mesenchyme of the limb bud to create the ZPA, as limb-specific *Shh* expression is abrogated upon its deletion (Sagai *et al*, 2005). Moreover, point mutations across the whole conserved sequence of the enhancer induce anterior, ectopic *Shh* expression that can cause PPD (Lettice *et al*, 2003; Sagai *et al*, 2004; Lettice *et al*, 2008). These point mutations occur within dissimilar DNA sequences, indicating that different DNA-protein interactions are being disrupted but with the same regulatory outcome. However, point mutations in two families with PPD result in the formation of protein-binding motifs associated with the ETS family of transcription factors, which include ETV4 and ETV5 (Fig. 3.1A) – factors with a known role in *Shh* regulation in limb development (Section 1.3.2). *In Silico* analysis by the Hill lab identified multiple ETS binding motifs, which suggests a possible role for these factors in *Shh* limb expression through the ZRS (I will discuss this further in the introduction to Chapter 3, the results chapter based on research into ETS factor binding at the ZRS).

Recent analysis of Silkie (Slk) chickens (an ancient breed with naturally polydactylous feet) revealed causative links between the mutated ZRS – both ectopic anterior *Shh* expression and more extensive Shh signaling in the posterior leg bud – and extension of the digit-forming leg bud field; all resulting in morphogenic variability of anterior digits: from an extra phalanx in digit 1 to two extra digits with

either digit 1 or digit 2 morphology (Dunn et al., 2011). Based on their results, the authors propose a mechanism based on the growth-morphogen model with the mutated ZRS driving increased *Shh* expression in an expanded posterior domain. Consequently, Shh signaling is intensified, which extends the digit-forming field. When the posterior domain is then surgically removed at this stage ectopic activation of *Shh* fails to occur and the result is the condensation of an additional anterior digit but the loss of the most posterior digit 4. Abnormal levels of Shh signaling causes the anterior expansion of ectodermal Fgf signaling which then acts upon anterior mesenchyme tissue that is receptive to *Shh*-activating signals through the mutated ZRS. Ectopic expression of *Shh* then ensures the growth of additional digit(s) with a more posterior morphology.

The length and conservation of the ZRS suggests a high degree of complexity in controlling the spatial-temporal expression of *Shh* in the limb. Its size – ~800bp – is at the upper end of well defined regulatory elements, and so an intriguing question is: does the regulatory sequence contains information that not only enables enhancer-promoter interaction and subsequent *Shh* expression but also endows the activated enhancer with the capability to function over large distances? Work on the elucidation of how the ZRS activates *Shh* suggests that there is a physical co-localisation of the enhancer and gene in both the anterior and posterior margins of the mouse limb bud at E10.5; however, a looping out of the *Shh* locus from the CT could only be detected at the posterior margin where *Shh* is expressed (Amano et al., 2009). This study indicates that higher-order chromatin conformation has a role in ZRS activity, and the multiple point mutations identified within the enhancer suggests that its functional capability is dependent on the dynamic interplay between activating and repressive binding factors. What remains to be addressed are the identity of activating and repressing factors and the mechanism by which they interact to constrain ZRS activity to the posterior limb bud. In addition, what are the motive forces behind the ZRS-*Shh* interactions? Is it a looping mechanism? gross compaction of the intervening region through irregular folding of the chromatin fibre? or is this whole region a discrete genomic locus (a “regulatory landscape”, as has also been proposed for the gene desert adjacent to HoxD, see below) being held in a tight conformation that reduces the nuclear search space for tissue-specific

enhancers of *Shh* to find and activate the gene? Each of these scenarios rely upon a deterministic, biochemical manipulation of the higher-order chromatin structure, while still leaving room for stochastic interactions as the enhancer searches for a receptive promoter.

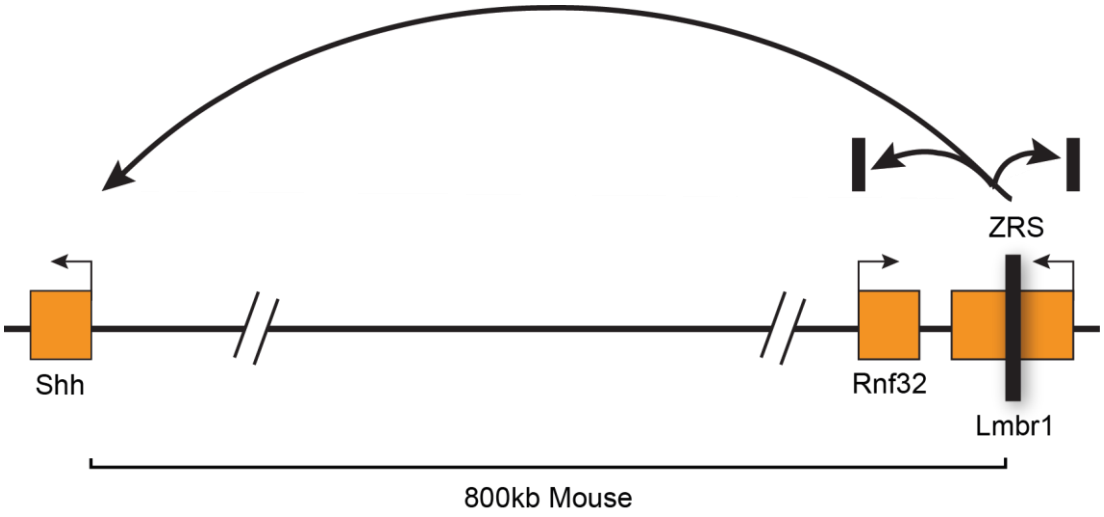

Figure 1.9. ZRS/*Shh* regulatory region. The arrows indicate that the enhancer only regulates *Shh*, bypassing the adjacent *Lmbr1* and *Rnf32* genes.

1.5 GCR

The global control region (GCR) is a series of *cis*-regulatory elements within a ~50kb region located around 200kb upstream of the HoxD cluster. These regulatory elements are highly conserved in vertebrates and are active in the developing nervous system, regulating *Lnp* and *Evx2* expression – two genes that lie between the GCR and HoxD (Spitz et al., 2003). However, one of these elements has been co-opted along with the 5′ genes during evolution of the tetrapod limb. The GCR, then, is active in the limb and genital buds and the CNS during development, but only *Evx2* and *Lnp* are expressed in the CNS (Fig. 1.10); possibly due to the activity of a boundary element located between *Evx2* and *Hoxd13* (Yamagishi *et al*, 2007).

The GCR is highly conserved between humans and mice and sequence identity is quite high compared with the *Fugu* sequences. However, the *Fugu* GCR does not drive reporter gene expression in mouse limb buds, but does drive expression in neural tissue (Spitz *et al*, 2003). These results suggest that late-phase *cis* regulation of *Hoxd* genes is unique to tetrapods. However, further transgenic analysis of GCR activity in fish species whose genomes have not undergone a duplication event (the authors propose that the additional set of *Hoxd* genes and GCR in *Fugu* could have culminated in enhancers with divergent functions) demonstrated that conserved sequences from tetrapods, zebrafish and skate can drive appendage expression in transgenic mice and zebrafish (Schneider et al., 2011). The authors suggest that development of the autopod during tetrapod evolution arose through modification of ancient gnathostome regulatory elements that led to an expansion of late phase *Hoxd* expression, possibly further enhanced by the acquisition of the Prox enhancer that has no orthologous sequence in fish (Gonzalez et al., 2007).

Although the molecular mechanisms that underlie GCR regulation of the 5′ *Hoxd* genes are unknown, the decreasing levels of expression from *Hoxd13* to *Hoxd10* indicate that positioning within the cluster is important. The use of deletion and duplication alleles confirmed the importance of gene position, the more 5′ the gene the greater its expression levels (Kmita *et al*, 2002; Tarchini and Duboule, 2006). Nevertheless, Montavon et al. (2008) identified some promoter-specific responses, with *Hoxd13* the most responsive. This quantitative colinearity

mechanism, then, with progressively weaker expression of *Hoxd* genes the further they are located away from the GCR, suggests direct promoter-enhancer interactions, possibly through a chromatin looping mechanism. Moreover, transgene experiments and genetically induced rearrangements of *Hox* loci suggest that progressive opening of the chromatin structure from 3′ to 5′ has a role in sequential gene activation collinear with the order of the genes (Kmita and Duboule, 2003; Soshnikova and Duboule, 2009). This model is supported by work done in the Bickmore lab, which showed there is chromatin decondensation followed by activation of *Hoxb* and *Hoxd* genes (Chambeyron and Bickmore, 2004; Chambeyron *et al*, 2005; Morey et al., 2007; Eskeland et al., 2010). So, with regard to late phase *Hoxd* activation in the limb are the same regulatory processes involved in addition to the activity of *cis*-regulatory elements? That is, does polycomb have a role and, as a consequence, does the HoxD locus undergo decompaction?

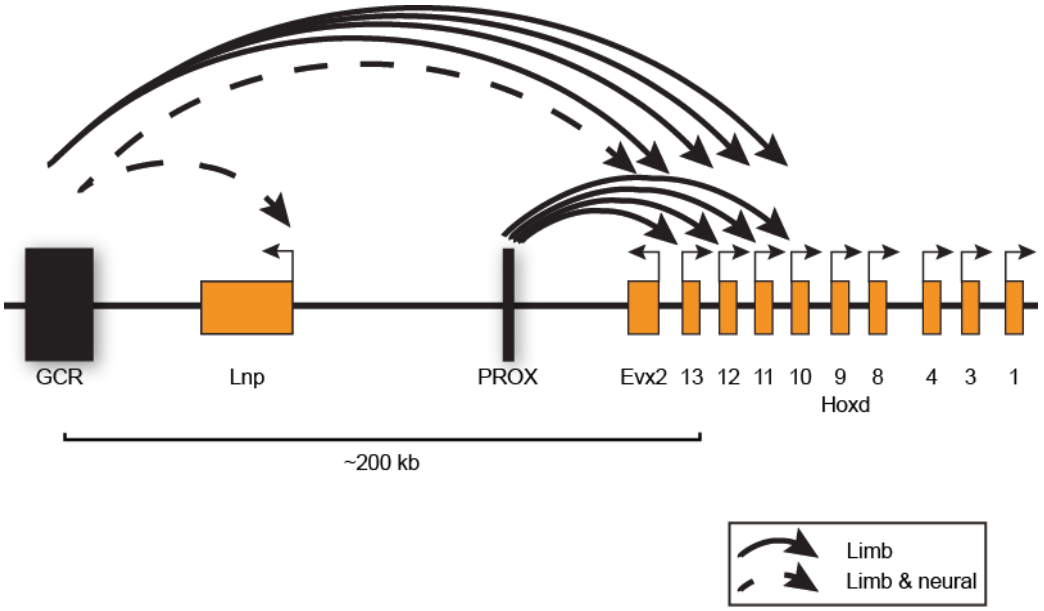

Figure 1.10. The 5′ HoxD regulatory region. The arrows indicate the genes that are activated by the Prox and/or GCR enhancers, solid arrows for genes regulated in the limb, dashed for genes regulated in limb and neural tissue.

1.6 Aims

The aim of my research is to delineate the chromatin structure and conformation at the *Shh* and *Hoxd* loci, including the long-range enhancers, in a model of the developing limb bud, and to determine whether chromatin modelling and structural alterations to chromosome territories have a role in the functional mechanisms of the ZRS and the GCR. Furthermore, I aim to characterise DNA-protein interactions at conserved and mutant binding motifs identified *in silico*, and subsequently to identify transcription factors that bind to the active ZRS.

To do this I made use of immortalised cell lines derived from the developing mouse limb bud, as well as dissected limb tissue and tissue sections, and genomic tiling arrays that span the genomic regions under study. I used chromatin immunoprecipitation (ChIP), fluorescence in situ hybridisation (FISH), expression microarrays and quantitative PCR to try and answer the following questions:

- Are modifications to chromatin structure necessary for the activation of limb-specific *Shh* and 5′*Hoxd* expression?
- Does polycomb regulate limb-specific *Hoxd* expression?
- Are alterations to chromatin conformation in each regulatory region enabling physical interaction between the enhancers and gene promoters?

For determining DNA-protein interactions at the ZRS ETS-binding sub-regions, various biochemical techniques in addition to ChIP-chip have been used to answer the following questions:

- Which protein or proteins bind(s) to the conserved and mutant binding sites?
- What is the mechanism that determines ZRS activity and how can point mutations disrupt the spatial specificity of this activity?

Chapter 2

Anterior-posterior differences in HoxD chromatin topology in limb development

2.1 Introduction

Regulated expression of Hox genes is crucial for anterior-posterior (A-P) patterning of the primary embryonic body axis, with Hox genes being first activated at gastrulation (Deschamps and van Nes, 2005). The HoxD cluster has also been co-opted more recently in evolution into regulating the growth and patterning of the limb and digits.

Polycomb repressive complexes (PRC1 and PRC2) are required to maintain Hox genes in a silent and compact chromatin state in embryonic stem (ES) cells (Boyer et al., 2006; Lee et al., 2006; Endoh et al., 2008; Stock et al., 2007; Eskeland et al., 2010). Whereas the roles of polycomb-mediated gene silencing and chromatin compaction at Hox loci have been well established in both early embryonic development and differentiation along the primary embryonic axis (Voncken et al., 2003; Faust et al., 1998; Chambeyron et al., 2005; Soshnikova and Duboule, 2009), whether polycomb-mediated histone modifications and chromatin changes are involved in Hox gene regulation in the secondary body axis is unclear.

There are two separate phases of *Hoxd* gene expression important in limb development and patterning (Zakany and Duboule 2007), with each phase regulated by two different groups of global regulatory elements situated either side of the cluster (Spitz et al., 2005). An early phase of temporal regulation results in expression of 3′ *Hoxd* genes (*Hoxd1-9*) earlier than the 5′ genes, due to collinear activation by an as yet undefined 3′-located enhancer (ELCR, Section 1.3) (Tarchini and Duboule, 2006). Combined with a proposed repressor element situated 5′ of the cluster, this restricts 5′ *Hoxd* expression to the posterior side of the distal limb bud and is required for limb outgrowth, proximal limb development, limb A/P polarity and the posterior expression of *Sonic hedgehog* (*Shh*) (Figs. 1.7, 1.8). This 3′-5′ temporal and spatial colinearity is reminiscent of regulation in the main embryonic axis, which is accompanied by progressive loss of histone H3 lysine 27 tri-methylation (H3K27me3) – the histone modification catalysed by PRC2 (Soshnikova and Duboule, 2009).

The early phase of limb *Hoxd* expression is followed by a later phase (E10.5) in the distal limb which is required for digit morphogenesis (Spitz et al., 2003). This is characterised by "quantitative colinearity" in which expression of the most 5′ gene, *Hoxd13*, is initially strongest in the posterior mesenchyme of the distal limb, with progressively less strong expression of *Hoxd12* to *Hoxd10*. The expression domains of all the 5′ genes (*Hoxd13-10*) then spread anteriorly, so that by E12.5 they are being transcribed across the handplate but with only *Hoxd13* gene expression robust enough to be detectable on the most anterior side of the distal limb – the position of the presumptive thumb (Fig. 1.8) (Montavon et al., 2008). This expression is driven by enhancer elements including; a ~40kb global control region (GCR) centromeric of the cluster, ~200kb 5′ of *Hoxd13* beyond *Evx2* and *Lnp*, and the Prox enhancer that is located between *Evx2* and *Lnp* (Fig 1.10) (Spitz et al., 2003; Tschopp and Duboule, 2011). This late phase of HoxD regulation in the distal limb is a more recent adaptation (Schneider et al., 2011) and given its distinct regulatory characteristics it was not clear whether polycomb-mediated repression is involved.

2.2 Cell lines retain distal limb identity

2.2.1 Cell lines derived from the mesenchyme of E10.5 limb buds

The activity of limb enhancer elements in the HoxD GCR first becomes apparent in posterior mesenchyme cells of the distal forelimb bud at E10.5, followed later by anterior extension of this expression zone across the distal limb (Spitz et al., 2003). Therefore I restricted my analysis to E10.5 limb buds to capture the initiation of this second wave of HoxD regulation.

Analysing chromatin structure in the small number of cells dissected from regions of the mouse embryo is challenging (Soshnikova and Duboule, 2009). However, conditionally immortalised cell lines can be derived from the transgenic "immortomouse" that expresses a temperature sensitive SV40 T antigen H-2Kb-tsA58 (simian virus 40 (SV40) large tumour (T) antigen (TAg)) (Jat et al. 1991). The temperature sensitive gene was used for producing the transgenic mice to prevent the expression of the antigen *in vivo* as it has been associated with tumourigenesis and improper development. By culturing the cells derived from the transgenic embryos at 33°C the antigen is activated and consequently restricts further differentiation (Jat et al, 1991). Such cell lines appear to retain many biological properties of the cells from which they were derived, including gene expression patterns and response to signalling pathways (Kohn et al., 2010). Two sets of cell lines were derived from dissected E10.5 forelimb buds (by A. Hill) to represent the most posterior third of the distal limb (cell lines P1 and P2) or the anterior two-thirds (A1 and A2) (Fig. 2.1).

The embryonic limb bud consists of two main cell types, mesenchyme and a layer of surface ectoderm which at the most distal limb margin forms the apical ectodermal ridge (AER). The morphology of the cell lines indicated that they were likely to be mesenchymal in origin. I confirmed this by RT-PCR analysis of RNA isolated from all four cell lines (Fig. 2.2A). *Fgf10* is expressed in limb bud mesenchyme and signals to the Fgfr2b receptor expressed in the AER to induce expression of *Fgf8* there. In contrast, *Fgfr2c* is expressed in both the mesoderm and ectoderm of the developing limb bud (Lizarraga et al., 1999; Duboc and Logan, 2011). The detection of mRNAs from *Fgf10* and *Fgfr2c*, but not *Fgf8* or *Fgfr2b* in

immortomouse derived limb bud cell lines indicates their derivation from the limb mesenchyme (Fig. 2.2A). The origin of these cells from the distal, rather than proximal, margin of the limb is indicated by the expression of *Etv4* (Mao et al. 2009). I conclude that the 4 cell lines retain developmental specific functions expected of their limb origin (Robert and Lallemand, 2006; Hill, 2007).

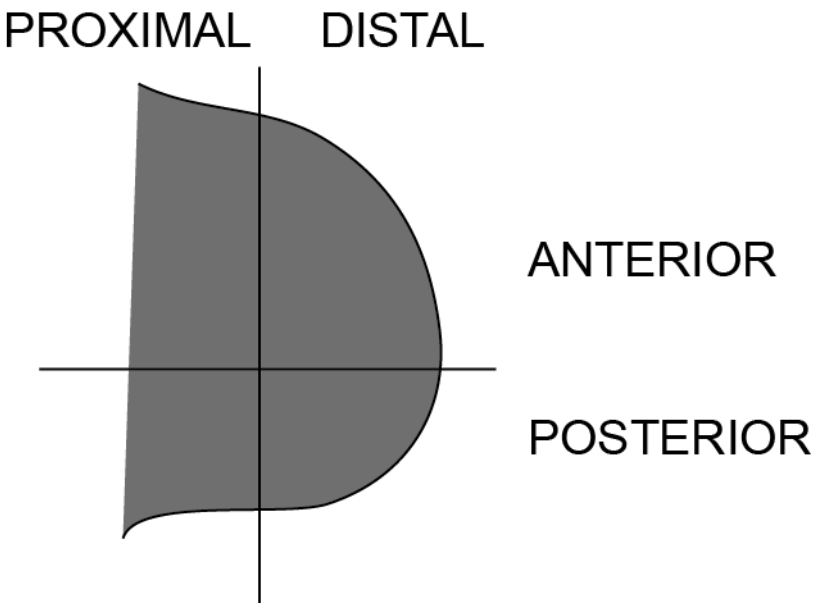

Figure 2.1. Schematic of E10.5 mouse forelimb bud. The cross-lines show where the limb buds were dissected into anterior vs posterior distal regions and used to generate cell lines from Immortomouse.

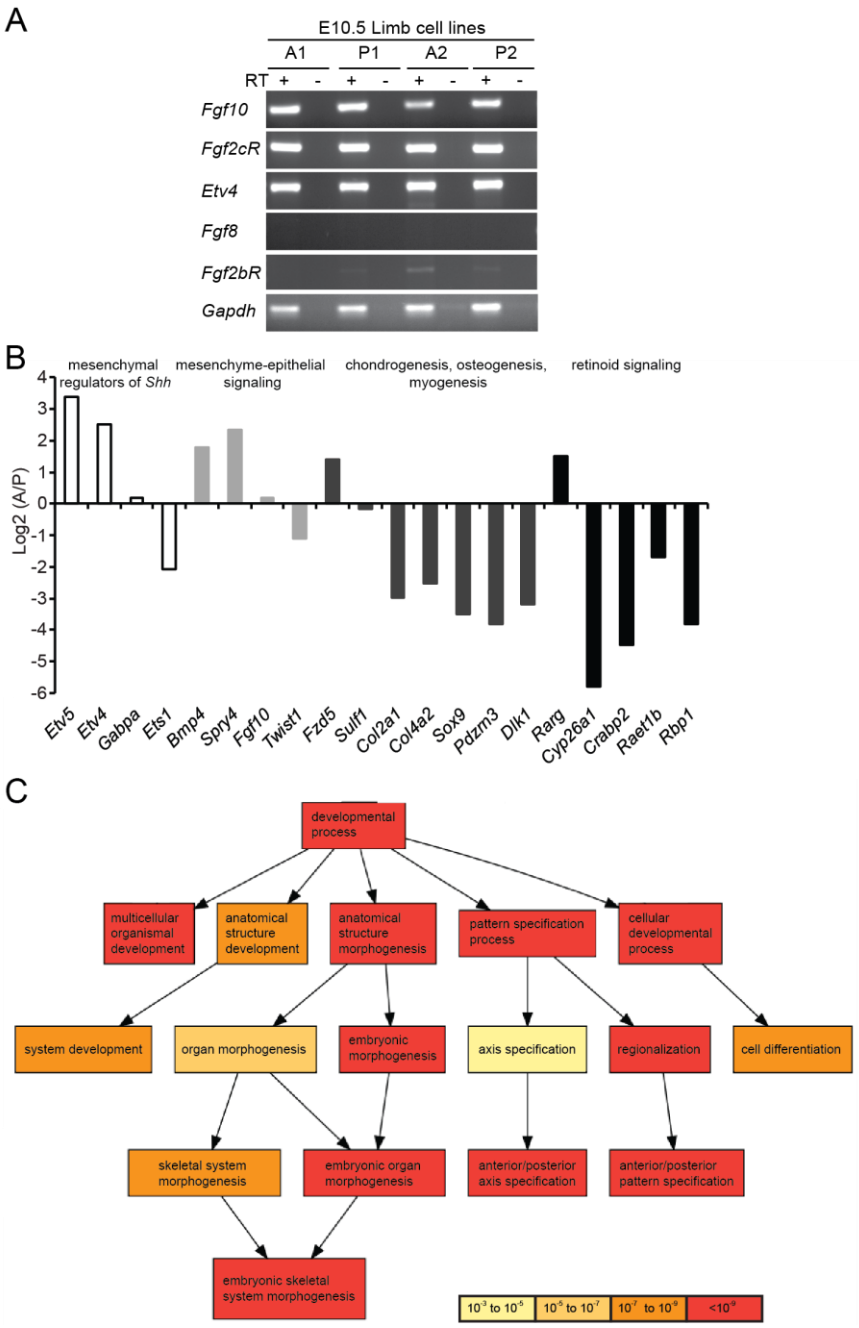

Figure 2.2. Characterisation of cell lines derived from distal posterior and anterior forelimb. A, RT-PCR to detect the expression of mesenchymal markers (*Fgf10*, *Fgf2cR*, *Etv4*) and epithelial markers (*Fgf8*, *Fgf2cR*) in the A1/P1 and A2/P2 cell lines. **B**, Selection of genes showing greater anterior or posterior expression or similar levels of expression in the A2/P2 cell lines from expression microarray analysis. Genes are categorized according to their function in limb development. *y*-axis: log2 scale showing greater anterior expression upper, greater posterior lower (A/P). **C**, Schematic of gene ontology (GO) pathway analysis of expression microarray data for genes with significantly different A/P expression.

To provide further insight into the differences between the anterior and posterior derived cell lines (A2 and P2) I analysed RNA prepared from these cell lines on mouse expression microarrays (Illumina). Array hybridisation was conducted by the Wellcome Trust Clinical Research Facility and the data were analysed in R using the limma (Smyth et al. 2005) and beadarray (Dunning et al. 2007) bioconductor packages. Probes were assessed for differential expression using limma's lmFit, eBayes and topTable function. Levels of *Etv4* and *Etv5* mRNAs – encoding transcription factors known to be expressed in the distal limb mesenchyme and important for the posterior restriction of Shh expression (Mao et al., 2009) – were higher in the anterior cells than posterior, whereas members of different groups of ETS factors that define the spatial boundary of *Shh* (Chapter 3) were higher in the posterior cells (*Ets1*) or expressed at similar levels (*Gabp*α) throughout the distal limb bud (Lettice et al., 2012) (Fig. 2.2B). Genes involved in mesenchymal-epithelial signalling and patterning of the distal limb bud were upregulated in anterior (*Bmp4*, *Spry4*) and posterior (*Twist1*) cells or expressed evenly across the A-P axis (*Fgf10*). Amongst the genes with the highest ratio of expression in the posterior cells were those involved in retinoid signalling (*Cyp26a1*, *Crabp2*, *Raet1b*, *Rbp1*), and in chondrogenesis, osteogenesis or myogenesis (*Sox9*, *Pdzrn3*, *Dlk1*, *Col2a1*, *Col4a2*) (Fig. 2.2B). However, some genes associated with retinoid signalling (*Rarg*) and skeletal morphogenesis (*Fzd5*) were more highly expressed in the anterior cells or showed similar levels of expression (*Sulf1*).

Gene Ontology analysis (GOrilla) indicated that the genes with the most significantly different A/P expression levels were most highly enriched for GO Biological Process terms ($P < 10^{-9}$) including anterior/posterior axis and pattern specification, anatomical structure morphogenesis and embryonic skeletal system morphogenesis – categories that reflect the cell lines' origins from mesenchymal tissue dissected from across the A/P axis of a developing appendage (Fig. 2.2C).

2.2.2 Immortomouse cell lines reflect differential 5′ Hoxd gene expression in distal posterior limb bud

The spatial expression domains of Hoxd genes in E10.5 limb buds are well characterised (Spitz et al., 2003; 2005; Tarchini and Duboule, 2006; Zakany and Duboule, 2007). To determine how relevant the limb cell lines are for analysis of HoxD spatial regulation in the distal limb I carried out RT-PCR and qRT-PCR on both sets of cell lines and on tissue dissected from E10.5 anterior and posterior forelimb buds.

Hoxd1 mRNA was not detected in any of the cell lines and only very faintly in anterior limb tissue (Fig. 2A). *Hoxd3* expression was detected at levels that are generally low (Fig2A) and similar between the anterior and posterior limb tissue and the corresponding cell lines – except for even lower levels in P2 cells (Fig 2B). There was no significant anterior/posterior difference in *Hoxd3* expression in either cell line pairs or limb tissue (Fig. 2C). Conversely, at the 5′ end of HoxD the levels of *Hoxd13* expression are 17 fold higher in posterior limb and limb-derived cell lines than in the anterior equivalents (Fig. 2C). A2 and P2 cell lines showed similar expression levels as detected in dissected anterior and posterior tissue, but A1 and P1 lines showed proportionally similar decreases in expression levels (Fig. 2B). I observed slightly higher *Hoxd12* expression in the posterior limb tissue than the anterior, and higher *Hoxd11* in P1 compared to A1 cells, but their expression levels were too low for reliable quantification. *Hoxd10* expression shows a high P/A difference in both cell line pairs – the apparent large difference (Fig 2C) measured in the A1/ P1 cell lines is due to the very low levels of expression in A1 – and *Hoxd8* expression shows a P/A differential (3 fold) in the second cell line pair only due to increased expression of this gene in the P2 cells, expression levels in A2, A1 and P1 being similar to limb tissue levels (Fig. 2B and C). No P/A difference in *Hoxd8* expression was detected in the E10.5 limb tissue samples. This suggests that both cell line pairs capture cells that have activated 5′ *Hoxd* genes specifically in the posterior compartment at the start of the second wave of HoxD activation in the distal limb and that Hox gene activation has a greater extension 3′ in the P2 cell line (to *Hoxd8*) than P1 (*Hoxd10*). Gene activation extends more 3′ in the posterior derived cell lines

than in the dissected limb tissue. This could be due to the outgrowth of cells from the posterior limb that have more extensive gene activation than the average or, more likely, that the regulatory mechanisms driving progressive 5′-3′ HoxD activation continued to operate for some time after tissue dissection and during cell immortalisation.

Upstream of *Hoxd13*, *Evx2* expression could not be detected in the E10.5 cell lines by PCR (data not shown) suggesting that the cells likely originate from outside of the small domain at the extreme distal margin of the E10.5 forelimb bud to which *Evx2* expression is restricted (Tarchini and Duboule, 2006). *Shh* expression was also not detected, suggesting that it is not required for maintenance of the second wave of HoxD activation, at least in cell lines, as has been previously proposed (Harfe et al., 2004).

Whilst both immortomouse derived distal limb cell line pairs show a posterior -anterior difference in 5′ HoxD activity, the P2/A2 pair show a more extensive domain of activation (*Hoxd13* to *Hoxd8*) and so were chosen for most of the subsequent study.

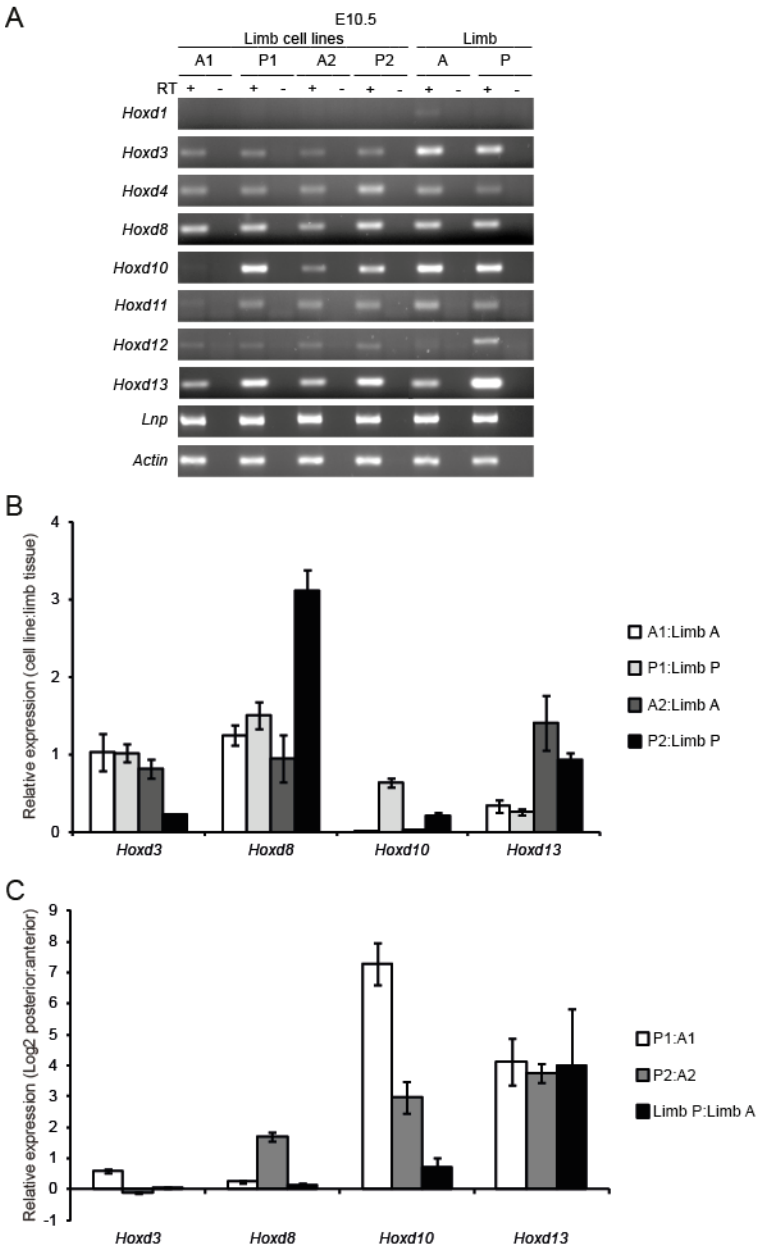

Figure 2.3. Comparison of Hoxd expression in the distal anterior and posterior limb cell lines with distal anterior and posterior forelimb tissue A, RT-PCR to detect the expression of 3′ (*Hoxd1*, *Hoxd3*, *Hoxd4*, *Hoxd8*) and 5′ (*Hoxd10*, *Hoxd11*, *Hoxd12*, *Hoxd13*) *Hoxd* genes and the adjacent *Lnp* in both sets of cell lines (A1 & P1, A2 & P2) and in E10.5 distal forelimb tissue (A & P). Primer pair sequences for each gene are indicated in Table 6.1. **B and C,** Quantitative (q)RT-PCR to compare expression levels of 3′ (*Hoxd3*, *Hoxd8*) and 5′ (*Hoxd10*, *Hoxd13*) *Hoxd* genes in both sets of cell lines and anterior or posterior distal forelimb tissue. In **B,** expression in each cell line is compared to the corresponding limb tissue. **C,** Log2 P/A expression in both sets of cell lines (white and grey bars) and in distal forelimb tissue (black bars). The *y*-axis has a Log2 scale due to the large difference in posterior/anterior expression of *Hoxd10* in the A1 & P1 cell lines.

A model for late phase HoxD activation in the distal limb bud proposes a reciprocal activation pathway involving the restriction of early phase 5′ *Hoxd* gene expression to the posterior margin of the limb bud that then ensures posterior-only activation of *Shh*, which in turn effects the initiation of secondary 5′ HoxD expression in the distal posterior limb bud (Harfe et al., 2004; Robert and Lallemand, 2006). I did not detect *Shh* expression in either the P1 or P2 posterior derived cell lines, which nonetheless maintain upregulated 5′ *Hoxd* gene expression – especially of *Hoxd13* (Fig. 2.3). However, late phase *Hoxd* gene expression still occurs in $Shh^{-/-}$ $Gli3^{-/-}$ mice, and only at reduced levels in $Shh^{-/-}$ mice (Litingtung et al., 2002). *Shh* expression prevents *Gli3* repression of the 5′ *Hoxd* genes, both by repressing *Gli3* expression posteriorly and antagonizing Gli3 processing to Gli3R, as Gli3R has been suggested to be the repressor of 5′ *Hoxd* gene expression in anterior limb (Tarchini and Duboule, 2006). It is likely that posterior *Shh* expression in the embryonic limb, up to the point of cell immortalisation, was sufficient to activate late phase HoxD expression. The subsequent strong maintenance of *Hoxd13* expression in the posterior limb-derived cell lines suggests that *Shh* is not required for maintaining HoxD activity. Loss of *Shh* expression in the cell lines could be due to the loss of interaction between the mesenchyme of the zone of polarizing activity (ZPA), which maintains A-P polarity and is defined by *Shh* expression, and the overlying AER which regulates proximal-distal patterning and growth (Laufer et al., 1994; Ogura et al., 1996).

2.3 A-P difference in polycomb over the HoxD cluster

2.3.1 Loss of H3K27me3 over the HoxD cluster in posterior cells

The polycomb complex PRC2 is fundamental to regulation of Hox gene clusters during ES cell differentiation and early embryogenesis and is responsible for blanketing HoxD with H3K27me3 (Boyer et al., 2006; Lee et al., 2006). Polarised (3′ - 5′) loss of H3K27me3 accompanies 3′ *Hoxd* gene activation during ES cell differentiation (Eskeland et al., 2010) and in the tail bud of the main embryonic axis during early embryogenesis (Soshnikova and Duboule, 2009). PRC2 function in the limb is required for cell survival and for proximodistal elongation of the limb (Wyngaarden et al., 2011), but whether differential polycomb-mediated chromatin changes are involved in regulating anterior-posterior *Hoxd* gene expression during the late phase of distal limb patterning is unclear.

From my analyses in Figure 2.2, the limb cell lines are composed of, what appears to be, a relatively homogenous mesenchymal cell population. In contrast, more heterogenous cell populations are inevitably present in dissected limb tissue, where AER- and mesenchyme- derived cells expressing different cohorts of *Hoxd* genes may mask differences in H3K27me3 levels between the anterior and posterior mesenchyme. Immunoblotting shows that total levels of H3K27me3 are similar between A2 and P2 cells (Fig. 2.4A). Using native ChIP combined with custom high-density tiling arrays covering multiple regions of the mouse genome, including 670kb encompassing HoxD, I determined the H3K27me3 profile of the A2/P2 distal forelimb-derived cell lines (Fig. 2.4B).

H3K27me3 is pervasive over HoxD in anterior cells, with both the 5′ and 3′ (*Hoxd1 – d4*) genes densely covered (Fig. 2.4B). The dip in H3K27me3 between *Hoxd4* and *Hoxd8*, is similar to that seen in ES cells (Eskeland et al., 2010). Whilst H3K27me3 still blankets *Hoxd1 - d3* in posterior (P2) cells, it is largely absent over the 5′ genes (Fig. 2.4B, upper two tracks and higher resolution tracks below displaying the *Hoxd13* and *Hoxd12* genomic region). Low levels of H3K27me3 at the HoxA and HoxB loci where observed, with only *Hoxa11-13* and *Hoxb13* highly enriched (*Hoxa13* expression is restricted to the extreme distal margin of the limb

bud and temporal activation occurs at around E10.5 (Lehoczky et al., 2004)). The polycomb target *Pax6* was blanketed with H3K27me3, however no change in anterior versus posterior cells was observed. H3K27me3 is absent over *Brd3*, a gene that codes for a bromodomain protein with a role in histone acetylation.

Analysis of all regions of the genome covered by the tiling array, which includes the other paralogous Hox clusters, showed that extensive A-P differences in H3K27me3 are specific to HoxD. Of those probes significantly enriched in H3K27me3 in anterior cells (log2 H3K27me3/input ≥ 1) (Fig. 2.5, Upper Pie chart) only 8% are from HoxD. Yet 93% of these probes with an at least two-fold (log2 ≥ 1) A/P difference in H3K27me3 enrichment are from this locus – 88% from 5′ HoxD, and 5% from 3′ HoxD (Fig. 2.5, Lower Pie Chart). In contrast, any differences of H3K27me3 coverage over the entire HoxA locus between anterior and posterior cells at stage E10.5 is minimal: the HoxA locus contains 4% of the probes significantly enriched for H3K27me3 in anterior cells but only 2% of probes with at least two-fold A/P difference in H3K27me3.

Despite their cellular heterogeneity, I also analysed H3K27me3 in E10.5 anterior and posterior dissected distal forelimbs (Fig. 2.4B, lower two tracks) by native ChIP, using an amended protocol for low cells numbers (see Chapter 6: Methods and Materials). In both limb regions a block of H3K27me3 covered the 3′ end of HoxD (*d1-d4*). Both anterior and posterior limb samples also show a second block of modification over the 5′ end of HoxD, from *Hoxd9* through to *Evx2*, but this was at a significantly higher level in anterior compared to the posterior region (*Hoxd13* – *Hoxd10* $p < 0.0001$) (Fig. 2.6A and B). Whereas in the anterior distal limb H3K27me3 remains high from *Evx2* to beyond *Hoxd12*, in the posterior H3K27me3 declines from the *Evx2* - *Hoxd13* intergenic region up to *Hoxd11* (Fig. 2.6A). This A-P difference in H3K27me3 levels in limb tissue is specific to the 5′ HoxD region (Fig. 2.6C), being significantly greater ($p < 0.0001$) than the A-P difference for all Hox loci combined. The A-P difference for the 3′ HoxD region is not statistically significant ($p = 0.57$).

Quantitative PCR (qPCR) confirmed the lower H3K27me3 levels in P2 vs A2 cells at *Hoxd10* (Fig. 2.7A), with a less dramatic decrease at *Hoxd1*. There was no

significant A-P difference in H3K27me3 at the polycomb target *Olig2* (positive control) or housekeeping gene *β-actin* (negative control). A decrease in H3K27me3 at the *Hoxd10* promoter in the posterior compared with anterior limb tissue cell population was confirmed by qPCR, whereas H3K27me3 levels at the *Hoxd1* promoter were similar in both distal limb samples (Fig. 2.7B). These data are consistent with a role for polycomb-mediated repression in regulating posterior-anterior differences in *Hoxd* gene expression during the patterning of the distal limb.

2.3.2 PRC1 levels are reduced over HoxD in distal posterior cells

The switch over from early to late phase HoxD expression in the posterior mesenchyme of the E10.5 distal forelimb, involving the switching off of 3′ *Hoxd* genes as the 5′ genes are being activated in reverse order, suggests that regulatory factors are operating in a highly dynamic milieu. Therefore, snapshots of H3K27me3 profiles at this crucial developmental time point may not fully capture this dynamism if there were a time lag in the removal of this histone modification. The PRC1 complex recognizes and binds to H3K27me3-modified chromatin to bring about chromatin compaction and gene repression (Eskeland et al., 2010). To determine whether there is differential PRC1 binding related to limb-specific Hoxd gene expression, I analysed the distribution of the PRC1 component Ring1B by ChIP in E10.5 anterior and posterior dissected limb bud tissues.

The Ring1B profile at HoxD shows a correlation with that of H3K27me3 in limb bud tissues (Fig. 2.8A, H3K27me3 profile over the HoxD locus from Fig. 2.4 shown above). There is a block of Ring1B from *Evx2* up to *Hoxd9* in distal anterior limb cells, whereas in posterior tissue this block reaches *Hoxd10* and only sparsely covers the genomic region encompassing *Hoxd13-11*. The A-P difference in Ring1B levels was significant throughout the 5′ HoxD (promoters, genes and intergenic $p < 0.0001$). Over 3′ HoxD A-P differences were only significant over the genes themselves ($p < 0.0001$) and not over the gene promoters ($p = 0.68$) nor the intergenic regions ($p = 0.07$) (Fig. 2.8B). A-P difference at the 5′ HoxD region was greater than at all the Hox loci combined ($p < 0.0001$) (Fig. 2.8C).

At the HoxB locus, *Hoxb13* has high coverage of Ring1B in anterior and posterior cells, similar to H3K27me3. Ring1B-binding over the rest of the HoxB cluster was similar in anterior and posterior tissue, except over *Hoxb9* and the adjacent intragenic region where Ring1B levels were more extensive in the anterior cells (Fig. 2.8A). As for H3K27me3, Ring1B covered the *Pax6* locus but was absent at the *Brd3* locus in both anterior and posterior tissue.

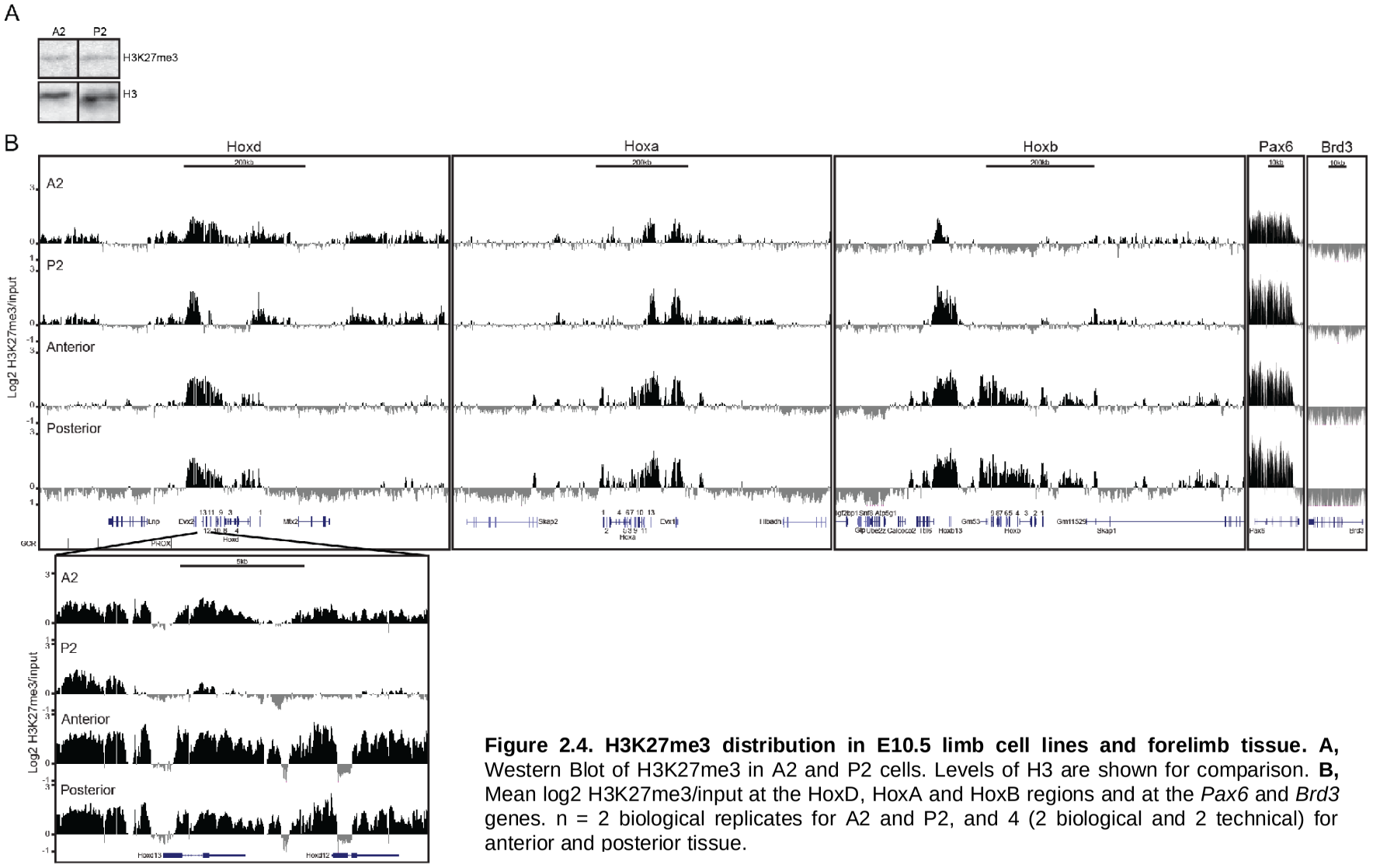

Figure 2.4. H3K27me3 distribution in E10.5 limb cell lines and forelimb tissue. A, Western Blot of H3K27me3 in A2 and P2 cells. Levels of H3 are shown for comparison. **B,** Mean log2 H3K27me3/input at the HoxD, HoxA and HoxB regions and at the *Pax6* and *Brd3* genes. n = 2 biological replicates for A2 and P2, and 4 (2 biological and 2 technical) for anterior and posterior tissue.

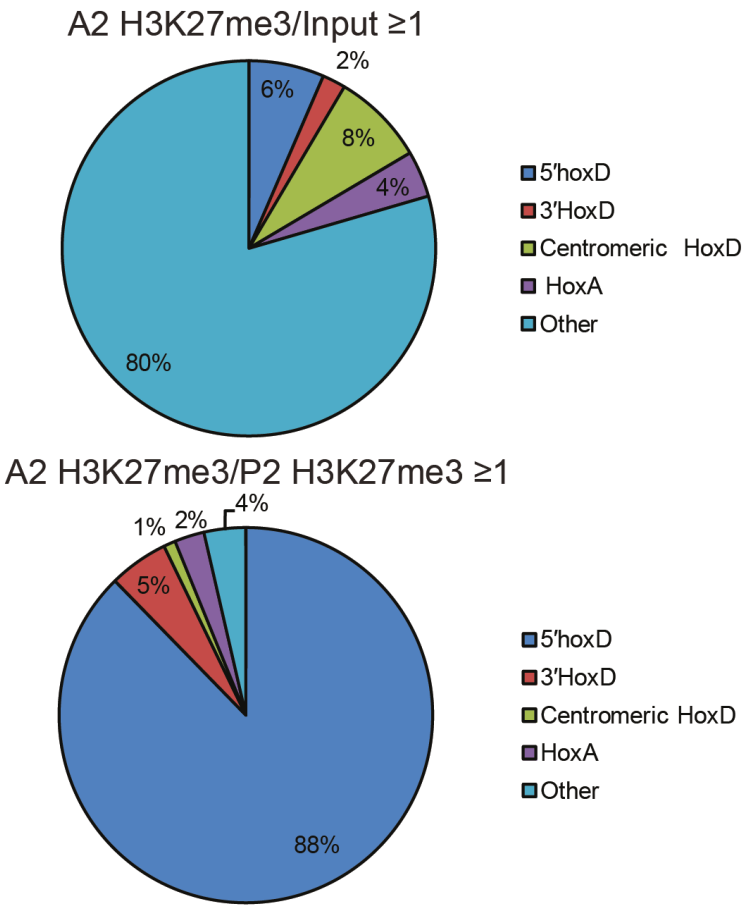

Figure 2.5. The majority of highly enriched H3K27me3 probes in A2 cells compared to P2 cells cover the 5′ HoxD genomic region. Pie charts showing the genomic distribution of different categories of probes enriched for; (top) H3K27me3 in A2 cells (log2 H3K27me3/input ≥ 1) versus (bottom) the proportion with an A2-P2 difference of log2 ≥ 1.

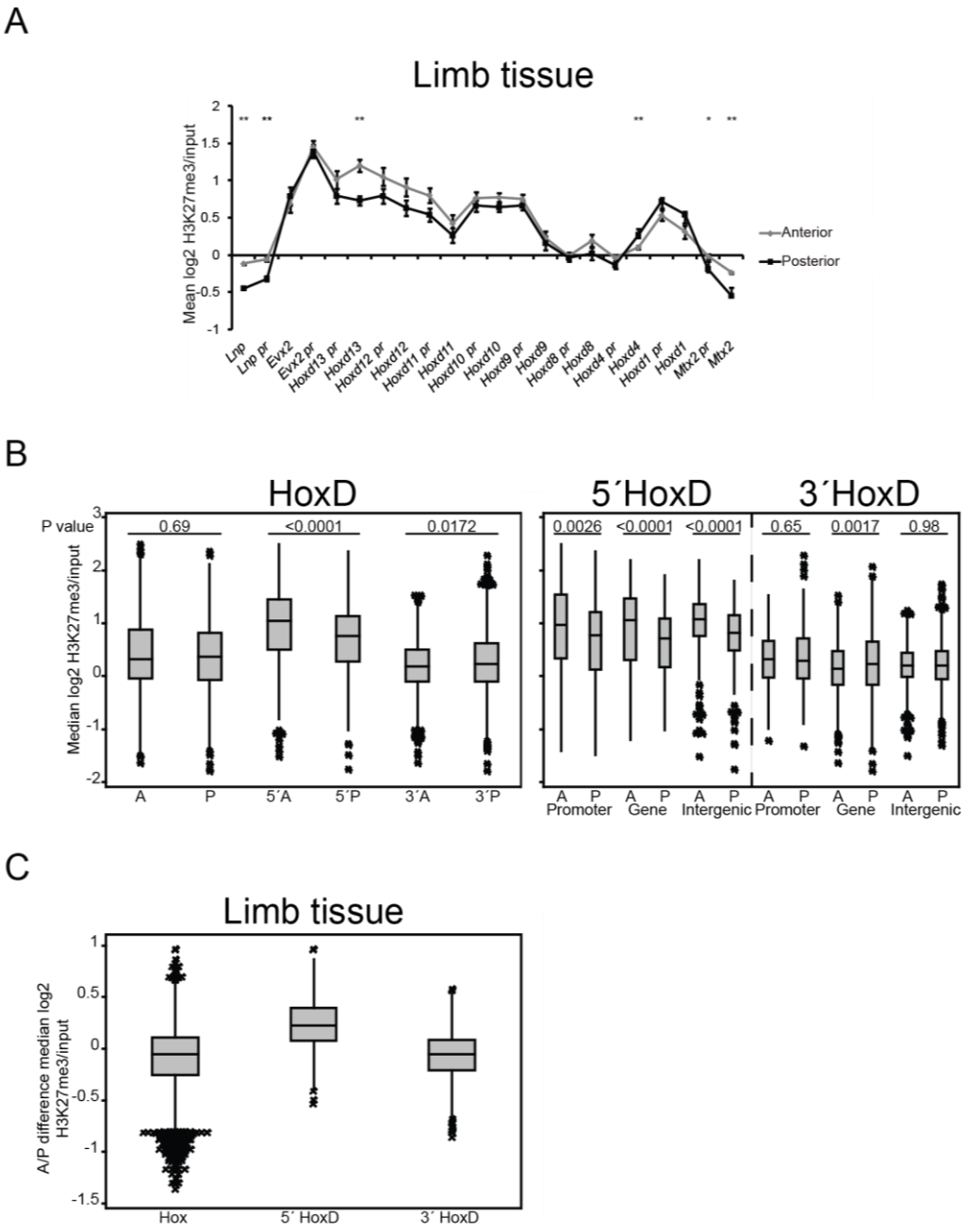

Figure 2.6. Comparison of H3K27me3 levels over 5′ and 3′ HoxD in E10.5 distal anterior and posterior limb tissue. A, Mean (+/- S.E.M.) log2 H3K27me3/input at HoxD and neighbouring genes and promoter regions in distal forelimb anterior and posterior tissue. Average log2 values were calculated from each individual probe value that covered the genomic locations of the genes and promoters. The significance of differences in H3K27me3 enrichment in anterior compared to posterior cells over each gene and promoter were examined by 2-sample t-test (* $p < 0.01$, ** $p < 0.0001$). **B,** Median log2 H3K27me3/input at the HoxD locus in distal anterior and posterior cells. The whole locus, 5′ and 3′ ends and promoter, gene and intergenic loci within each region were compared; and the statistical significance of anterior-posterior differences was examined by Mann-Whitney U tests. **C,** A-P difference in median log2 H3K27me3/input at all the Hox loci combined and the 5′ and 3′ ends of the HoxD locus. The statistical significance of differences between the combined Hox clusters and either end of the HoxD region were examined by Mann-Whitney U tests (5′ HoxD : Hox $p < 0.0001$, 3′ HoxD : Hox $p = 0.57$).

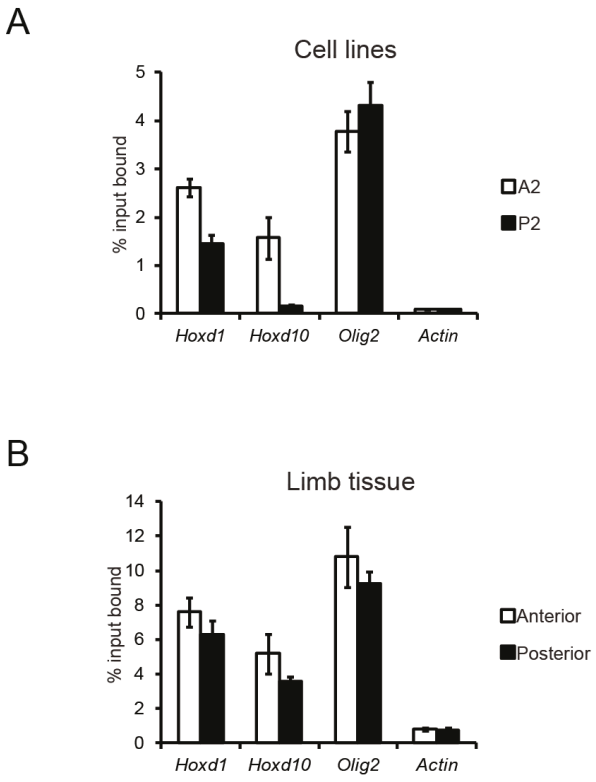

Figure 2.7. Loss of H3K27me3 at 5′ *Hoxd* promoter in posterior cells. A, ChIP for H3K27me3 at the *Hoxd1*, *Hoxd10*, *Olig2* and β-*actin* promoters by qPCR in A2 (white) and P2 (black) cells. Enrichment is shown as mean percent input bound ± SEM over 3 biological replicates. **B,** ChIP for H3K27me3 at the same gene promoters as (A) in E10.5 distal anterior (white) and posterior (black) forelimb tissue. Enrichment is shown as mean percent input bound ± SEM over 2 biological replicates.

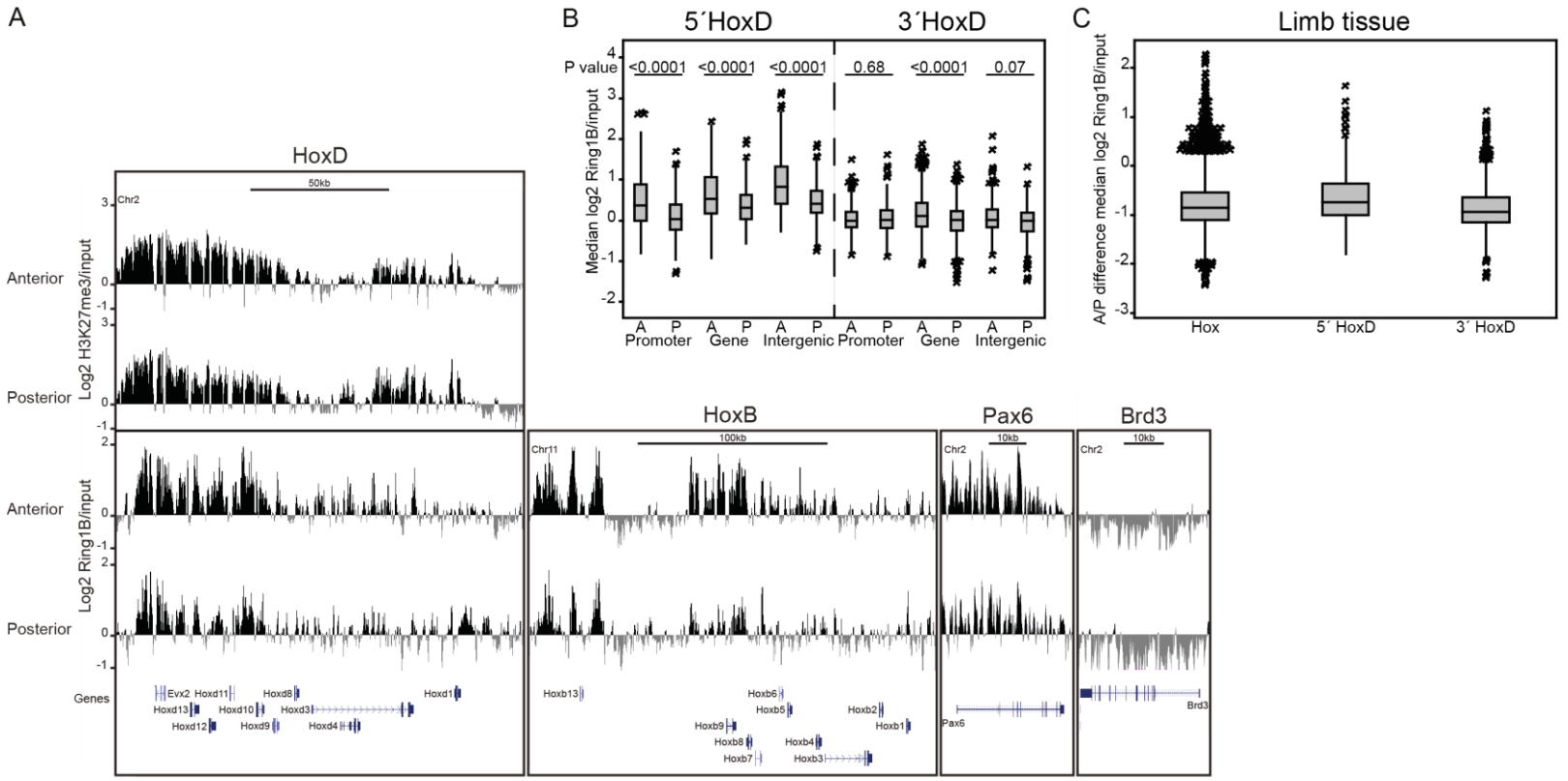

Figure 2.8. Ring1B distribution over Hox and non-Hox genomic regions in E10.5 forelimb tissue A, Mean log2 Ring1B/input at the HoxD, HoxB and β-globin regions and the *Pax6* gene (H3K27me3 tracks for HoxD locus shown above for comparison). Number of biological replicates n = 2. **B,** Median log2 Ring1B/input at the promoter, gene and intergenic regions of the HoxD locus in distal anterior and posterior forelimb tissue cells. The statistical significance of differences between each of the limb regions were examined by Mann-Whitney U tests. **C,** A-P difference in median log2 Ring1B/input at all the Hox clusters and the 5′ and 3′ ends of the HoxD locus. The statistical significance of differences between the combined Hox clusters and either end of the HoxD region were examined by Mann-Whitney U tests (5′ HoxD : Hox and 3′ HoxD : Hox $p < 0.0001$).

2.4 Anterior-posterior differences in chromatin compaction at HoxD in the distal limb bud

2.4.1 Chromatin decompaction in both posterior cell lines specifically over HoxD

Polycomb, and particularly PRC1, brings about a chromatin compaction at target loci that is detectable by measuring the spatial separation of fluorescence in situ hybridization (FISH) signals from closely apposed probe pairs (Eskeland et al., 2010) (Fig. 2.9). Using these approaches, a decompaction of HoxD is seen as Hox genes are activated during ES cell differentiation and in the tail-bud of the embryo (Morey et al., 2007). I first used 2D FISH to assay chromatin compaction (d^2) across HoxD in limb bud cell lines. Any difference in inter-probe distance (d) due to variation in nuclear size between the cell lines was also assessed by normalising d^2 by the nuclear radius (r^2). However, in practice the exact same conclusions were reached when considering interprobe separation without normalization to nuclear size (Table 2.1).

I found that chromatin across HoxD (*Hoxd3 – Hoxd13*) was significantly less compact in distal posterior cell lines compared to the distal anterior (p = 0.0002 and p = 0.03 for A1/P1 and A2/P2 respectively) (Fig. 2.10 and 2.11, and Table 2.1). This was restricted to the region with differential H3K27me3, as there is no significant A-P difference in chromatin compaction of the GCR-*Lnp* region 5′ of HoxD where the low levels of H3K27me3 detected are the same between A and P cell lines (Fig. 2.4B). Nor was there differential compaction at a control region near the *Pax6* locus (*Rcn – Rpl10*) on the same chromosome (MMU2) as HoxD, and which contains genes with no known role in limb development and no differential expression between A and P cell lines.

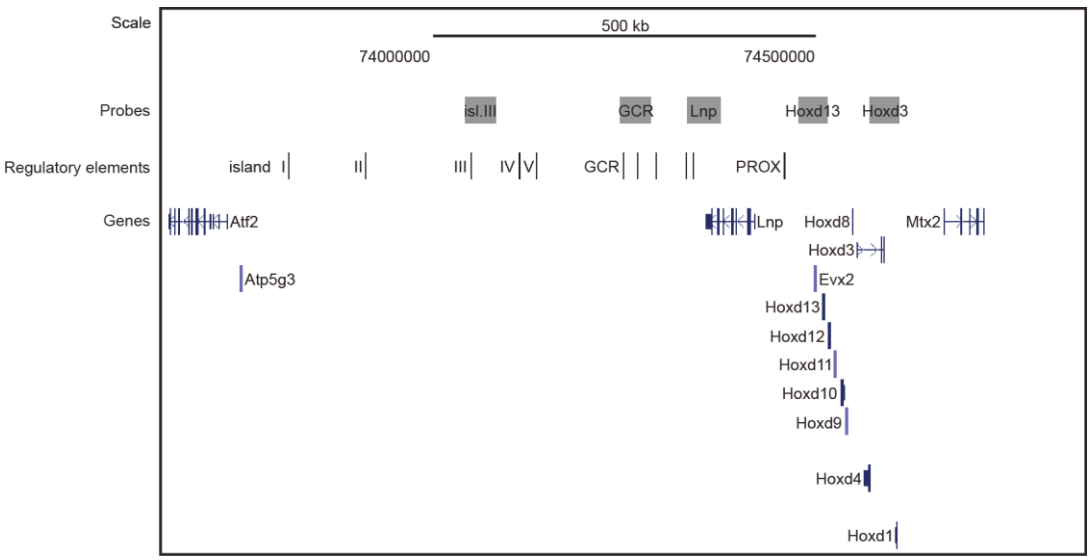

Figure 2.9 Schematic of the genomic region around HoxD. The grey boxes above the genomic region depict the probes used for FISH experiments. Highly conserved regions within the GCR locus, the position of the limb-specific Prox enhancer and the recently characterised island I-V regulatory elements (Montavon et al., 2011) are also indicated.

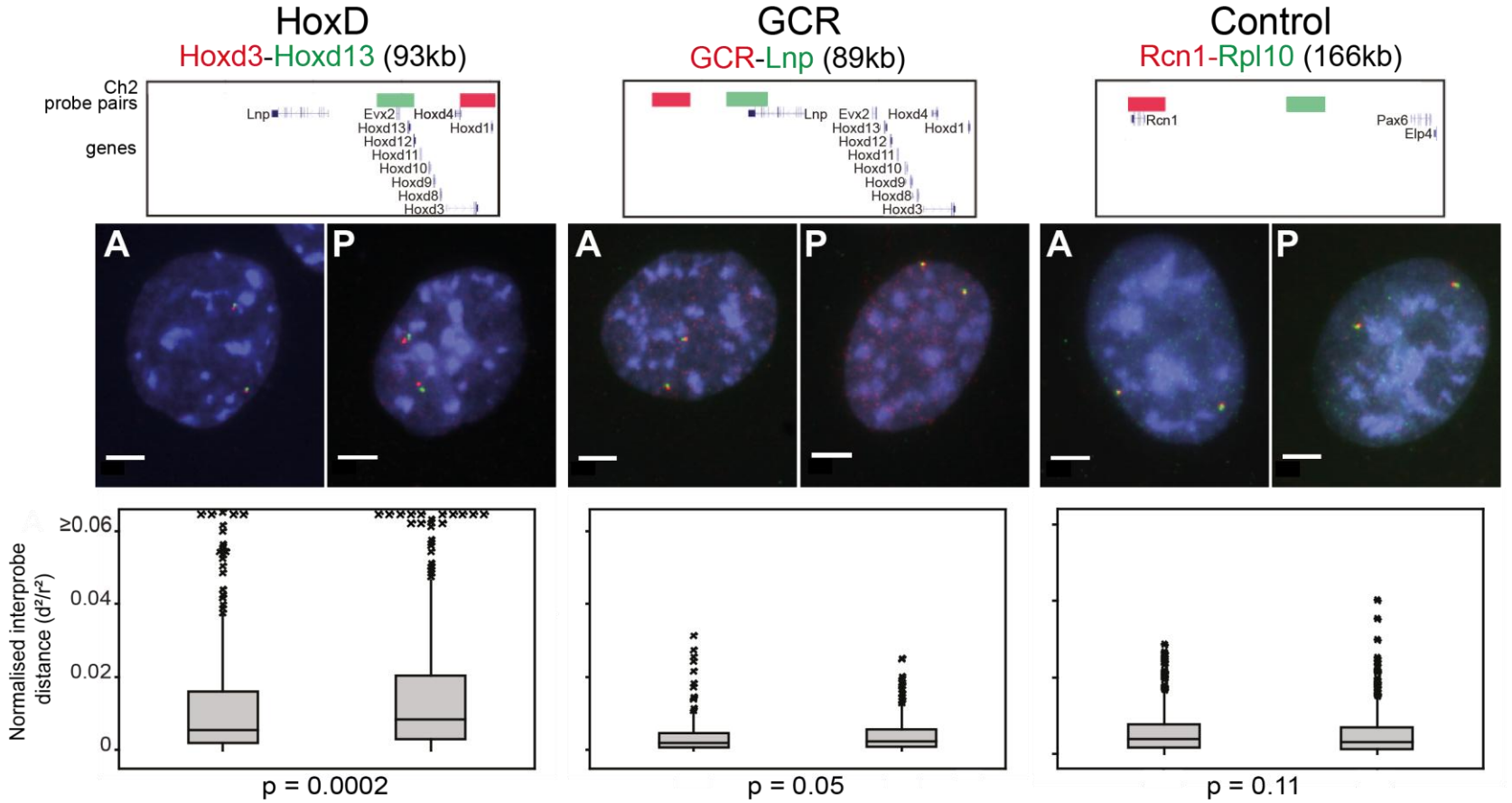

Figure 2.10. Chromatin decompaction at the HoxD locus in P1 cells. 2D FISH with probe pairs at HoxD, the genomic region centromeric to HoxD (GCR – *Lnp*) and a control region on MMU2, in MAA-fixed nuclei of anterior (A1) and posterior (P1) limb cells counterstained with DAPI (blue). Scale bar = 5 μm. Probe positions for each experiment are shown above the images. Box plots show the distribution of interprobe distances (d^2) normalized for nuclear radius (r^2) for A1 and P1 cells. The shaded boxes show the median and interquartile range of the data; asterisks indicate outliers. Number of loci: HoxD = ~400, GCR = ~300, control = ~300. The statistical significance of differences between A1 and P1 were examined by Mann-Whitney U tests: *Hoxd3-Hoxd13 p* = 0.002; GCR-*Lnp p* = 0.05; *Rcn1-Rpl10 p* = 0.11.

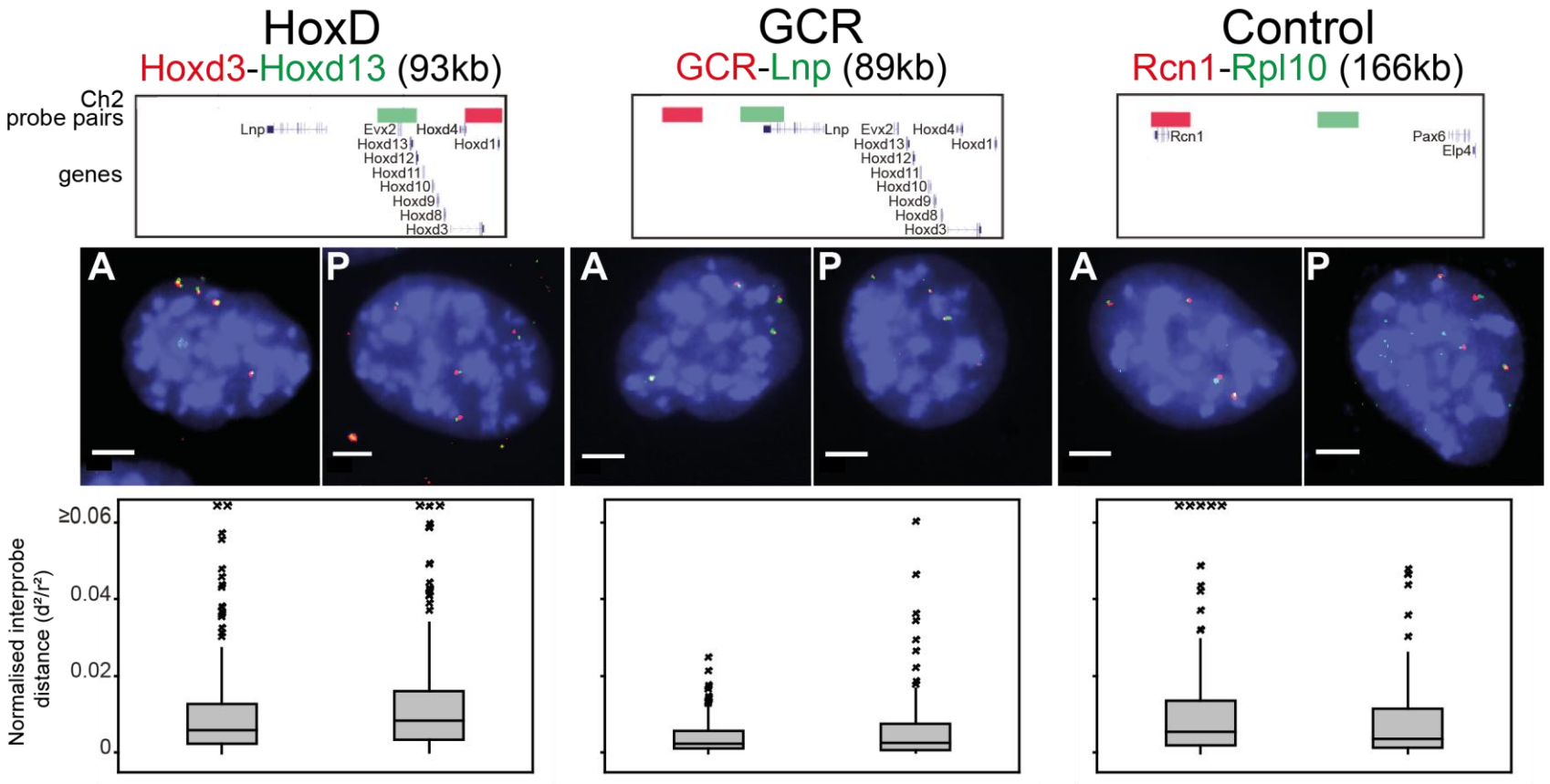

Figure 2.11. Chromatin decompaction at the HoxD locus in P2 cells. 2D FISH with probe pairs at HoxD, the genomic region centromeric to HoxD (GCR – *Lnp*) and a control region on MMU2, in MAA-fixed nuclei of anterior (A2) and posterior (P2) limb cells counterstained with DAPI (blue). Scale bar = 5 μm. Probe positions for each experiment are shown above the images. Box plots show the distribution of interprobe distances (d^2) normalized for nuclear radius (r^2) for A2 and P2 cells. The shaded boxes show the median and interquartile range of the data; asterisks indicate outliers. Number of loci: ~100 for each probe pair. The statistical significance of differences between A2 and P2 were examined by Mann-Whitney U tests: *Hoxd3-Hoxd13* p = 0.03; GCR-*lnp* p = 0.90; *Rcn1-Rpl10* p = 0.08.

Table 2.1. Normalised interprobe distance for A1 & P1 and A2 & P2 cell lines

Cell line	Hoxd3-Hoxd13 (93 kb)	GCR-Lnp (89 kb)	Rcn-Rpl10 (166 kb)
	Normalized interprobe distance (d^2/r^2)		
A1 **P1**	0.0056 0.0084 (p = 0.0002)	0.0021 0.0024 (p = 0.05)	0.0039 0.0030 (p = 0.11)
A2 **P2**	0.0060 0.0084 (p = 0.03)	0.0025 0.0027 (p = 0.90)	0.0054 0.0038 (p = 0.08)
	Squared interprobe distance (d^2) (μm^2)		
A1 **P1**	0.16 0.25 (p < 0.0001)	0.066 0.073 (p = 0.34)	0.11 0.09 (p = 0.14)
A2 **P2**	0.20 0.27 (p = 0.02)	0.09 0.08 (p = 0.88)	0.18 0.12 (p = 0.10)

Statistical analysis of data for Fig. 2.10 and 2.11. Interprobe distances are median values, *p*-values from Mann-Whitney U Tests.

2.4.2 Chromatin decompaction at HoxD is specific to the distal posterior limb bud in E11.0 embryos

I sought confirmation of differential HoxD compaction by FISH analysis in tissue sections cut through the anterior and posterior regions of the forelimbs of paraformaldehyde-fixed E10.5 – E11.0 embryos (Fig. 2.12). Here, not only was I able to compare the distal anterior and posterior regions, but also more proximal limb regions and indeed the flank region adjacent to the forelimb bud tissue where *Hoxd* genes are not expressed. As observed in the cell lines, HoxD chromatin was significantly less compact (d^2) at the distal posterior forelimb bud compared to the distal anterior (p = 0.0008) but also compared with other limb and non-limb regions analysed (proximal posterior, p = 0.02; proximal anterior, p = 0.0002; flank, p = 0.0002) (Fig. 2.13 and Table 2.2). There was no significant difference in HoxD chromatin compaction between any of the other limb regions or even between these other limb regions and the flank tissue. Likewise, no significant differences in chromatin compaction could be identified between any of the tissue regions, including the distal forelimb bud, at the GCR-*Lnp* and the *Rcn-Rpl10* control regions (Fig. 2.13 and Table 2.2). Therefore, an unfolding of chromatin at HoxD accompanies the start of the second wave of Hoxd gene activation during limb development.

Table 2.2. Interprobe distances for E11 limb bud sections

Limb region	Hoxd3-Hoxd13 (93 kb)	GCR-Lnp (89 kb)	Rcn-Rpl10 (166 kb)
	Interprobe distance (d^2) (μm^2)		
Distal posterior	0.178	0.076	0.080
Distal anterior	0.116 (p = 0.0008)	0.076 (p = 0.93)	0.081 (p = 0.10)
Proximal posterior	0.153 (p = 0.02)	0.076 (p = 0.46)	0.090 (p = 0.06)
Proximal anterior	0.098 (p = 0.0002)	0.074 (p = 0.45)	0.112 (p = 0.05)
Flank	0.130 (p = 0.0002)	0.078 (p = 0.49)	0.112 (p = 0.17)

Statistical analysis of data for Fig. 2.13. Interprobe distances are median values, *p*-values from Mann-Whitney U Tests.

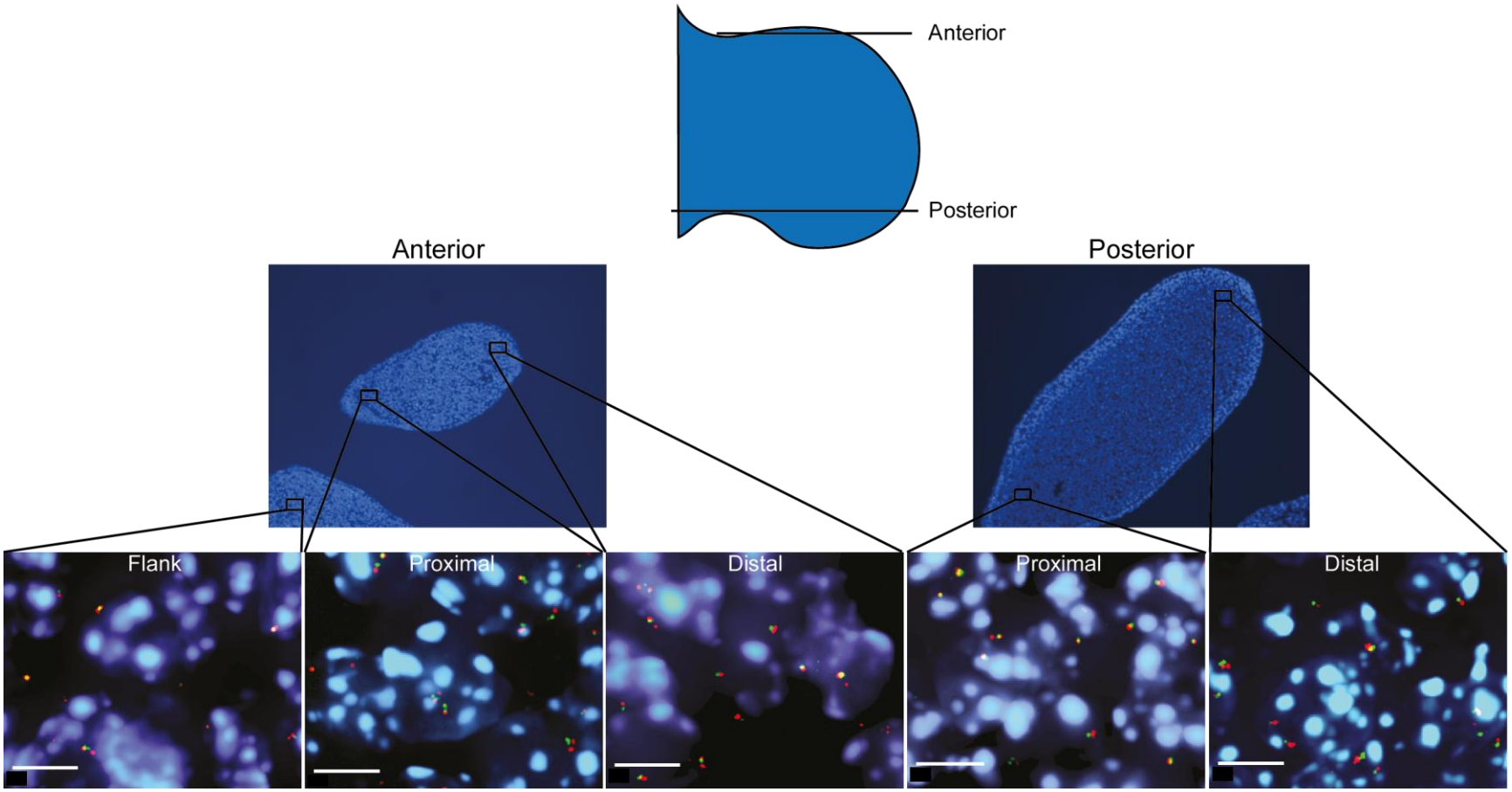

Figure 2.12. Representative sections and nuclear images. Images of anterior and posterior tissue sections for DNA FISH, counterstained with DAPI. Above is a schematic indicating the position and plane of the sections. Below are examples of nuclei from each of the limb regions and the adjacent flank region from which interprobe distances were measured. Hybridisation signals are for probe pairs at HoxD. Scale bars = 5 μm.

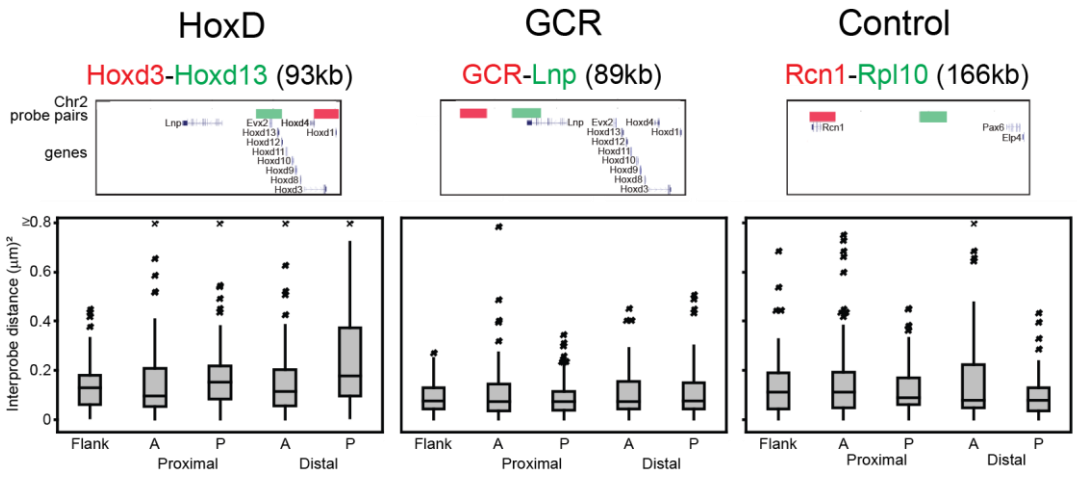

Figure 2.13. Decompaction of HoxD is specific to the distal posterior region of E11 forelimbs. Box plots show the distribution of interprobe distances (d^2) at the HoxD, GCR and control loci for the proximal and distal anterior and posterior forelimb bud and the adjacent flank. Shaded boxes show the median and interquartile range of the data; asterisks indicate outliers. Number of loci: HoxD = 100, GCR > 80, control > 80. Probe positions as indicated in Fig. 2.8. The statistical significance of differences between each of the limb regions and flank were examined by Mann-Whitney U tests: *Hoxd3-Hoxd13* – distal posterior - distal anterior, $p = 0.0008$; - proximal posterior, $p = 0.02$; - proximal anterior, $p = 0.0002$; - flank, $p = 0.0002$. GCR-*Lnp* – distal posterior - distal anterior, $p = 0.93$; - proximal posterior, $p = 0.46$; - proximal anterior, $p = 0.45$; - flank, $p = 0.49$. *Rcn1-Rpl10* – distal posterior - distal anterior, $p = 0.10$; - proximal posterior, $p = 0.06$; - proximal anterior, $p = 0.05$; - flank, $p = 0.17$.

2.5 Nuclear co-localisation of the GCR and 5′ HoxD in distal posterior cells

Developmental stage E10.5 marks the start of the later phase of distal limb development, characterised by quantitative colinearity in which strong expression of *Hoxd13* is initiated in the distal margin of the posterior mesenchyme. This is driven by enhancer elements located 5′ of HoxD, the best characterised of which is the GCR ~200kb centromeric of *Hoxd13* (Figs. 2.4B and 2.9) (Montavon et al., 2008).

One possible mechanism for long-range *cis*-regulation invokes the spatial co-localisation in the nucleus of the enhancer and target promoter (Williamson et al., 2011). Indeed, modelling of the changes in gene expression in the distal limb bud that occur as a consequence of 5′ HoxD deletions led to the suggestion that the first step in 5′ *Hoxd* gene activation in the distal posterior limb bud might be a long-range interaction (looping) between regulatory sequences such as GCR and *Hoxd13* (Montavon et al., 2008). To test this model, I analysed FISH signals for probes from the GCR and *Hoxd13* regions and determined the proportion that co-localised (d < 200nm). I analysed multiple areas within the E11.0 limb bud (Fig. 2.14A); including those where late phase gene activation initiates (distal posterior), where it is poised for activation at later embryonic stages (distal anterior), where the early phase of 3′-5′ *Hoxd* gene activation – that does not depend on the GCR – has occurred (proximal limb) and then in the control flank mesoderm.

Average GCR - *Hoxd13* interprobe distance (d^2) were significantly smaller in nuclei of the posterior distal limb bud compared to distal anterior ($p = 0.04$, Mann-Whitney U analysis) (Fig. 2.14B; Table 2.3). This was also true for comparison of GCR - *Hoxd13* interprobe distances (d^2) measured in the flank, proximal anterior and posterior (Mann-Whitney U analysis of distal posterior versus proximal posterior, $p = 0.002$; proximal anterior, $p = 0.02$; or flank, $p < 0.0001$). This was not seen for distances between GCR and a probe at the 3′ end of *Lnp* (Fig. 2.13) indicating that there is not a simple compaction of the whole region 5′ of *Hoxd13*. Moreover, the proportion of alleles where GCR and *Hoxd13* probe signals were spatially co-localised was far higher in posterior distal limb (> 30%) than in all other areas tested (Figs. 2.14C and 2.15A; Table 2.4). These data are consistent with the formation of a

chromatin loop between *Hoxd13* and the GCR regulatory module at the time and place of distal limb bud development when the GCR is required to initiate HoxD reverse co-linearity and *Hoxd13* expression (Fig. 2.16).

Recently, the 4C technique was used to identify sequences that could be captured together with *Hoxd13* in tissue from across the distal limb at E12.5 – a later stage of limb development and 5′ HoxD activation than studied here (Montavon et al. 2011). This identified multiple new potential regulatory regions in the extensive gene desert centromeric of HoxD and located far beyond the GCR. It was suggested that these elements might simultaneously interact with each other and with 5′ HoxD in the distal limb to form a compact regulatory hub (Figs. 2.9 and 2.16). One of these new regulatory elements, island III (Fig. 2.9), is located ~200kb centromeric of the GCR and was shown to interact with *Hoxd13* in limb and brain tissue (Montavon et al. 2011). To test this model, we measured the physical separation between the island III region and the GCR – a genomic distance that is equivalent to that separating GCR from *Hoxd13* (~200kb) – and also between island III and *Hoxd13* (Fig. 2.9). In E11.0 limb buds, we did not detect any significantly increased co-localisation frequency between island III and *Hoxd13* or between island III and GCR in posterior distal limb, compared to other forelimb regions (Figs 2.14C and 2.15B & C, Table 2.4). However, average interprobe distances (d^2) between *Hoxd13* and island III are similar to those between *Hoxd13* - GCR and GCR - island III despite the former being double the genomic distance (Fig. 2.14B, Table 2.3). These data suggest that the ~400kb genomic region from the 5′ end of HoxD into the centromeric gene desert is configured in a relatively compact, yet flexible, conformation (Fig. 2.16).

I did not assess the co-localisation of the Prox regulatory element (Fig. 2.9) as it is too close to the *Evx2 – Hoxd13* region to assess by FISH.

Table 2.3. Interprobe distances for HoxD regulatory regions in E11 limb bud sections

Limb region	*Hoxd13*-GCR (231 kb)	Island III-GCR (201 kb)	*Hoxd13*-island III (432 kb)
	Interprobe distance (d^2) (μm^2)		
Distal posterior	0.076	0.084	0.072
Distal anterior	0.098 (p = 0.04)	0.087 (p = 0.43)	0.067 (p = 0.96)
Proximal posterior	0.107 (p = 0.002)	0.098 (p = 0.99)	0.085 (p = 0.11)
Proximal anterior	0.098 (p = 0.02)	0.094 (p = 1.00)	0.099 (p = 0.005)
Flank	0.171 (p < 0.0001)	0.098 (p = 0.77)	0.081 (p = 0.04)

Statistical analysis of data for Fig. 2.14B. Interprobe distances are median values, *p*-values from Mann-Whitney U Tests.

Table 2.4. Co-localisation frequency of *Hoxd13* and enhancer probes for E11 limb bud sections

Limb region	*Hoxd13*-GCR (231 kb)	Island III-GCR (201 kb)	*Hoxd13*-island III (432 kb)
	Co-localisation frequency (%)		
Distal posterior	30	13	19
Distal anterior	14 (p = 0.01)	11 (p = 0.83)	14 (p = 0.30)
Proximal posterior	5 (p < 0.0001)	11 (p = 0.839)	16 (p = 0.61)
Proximal anterior	8 (p = 0.0001)	13 (p = 1.00)	13 (p = 0.22)
Flank	7 (p < 0.0001)	12 (p = 1.00)	10 (p = 0.07)

Statistical analysis of data for Fig. 2.14C. *p*-values from Fisher's Exact Tests

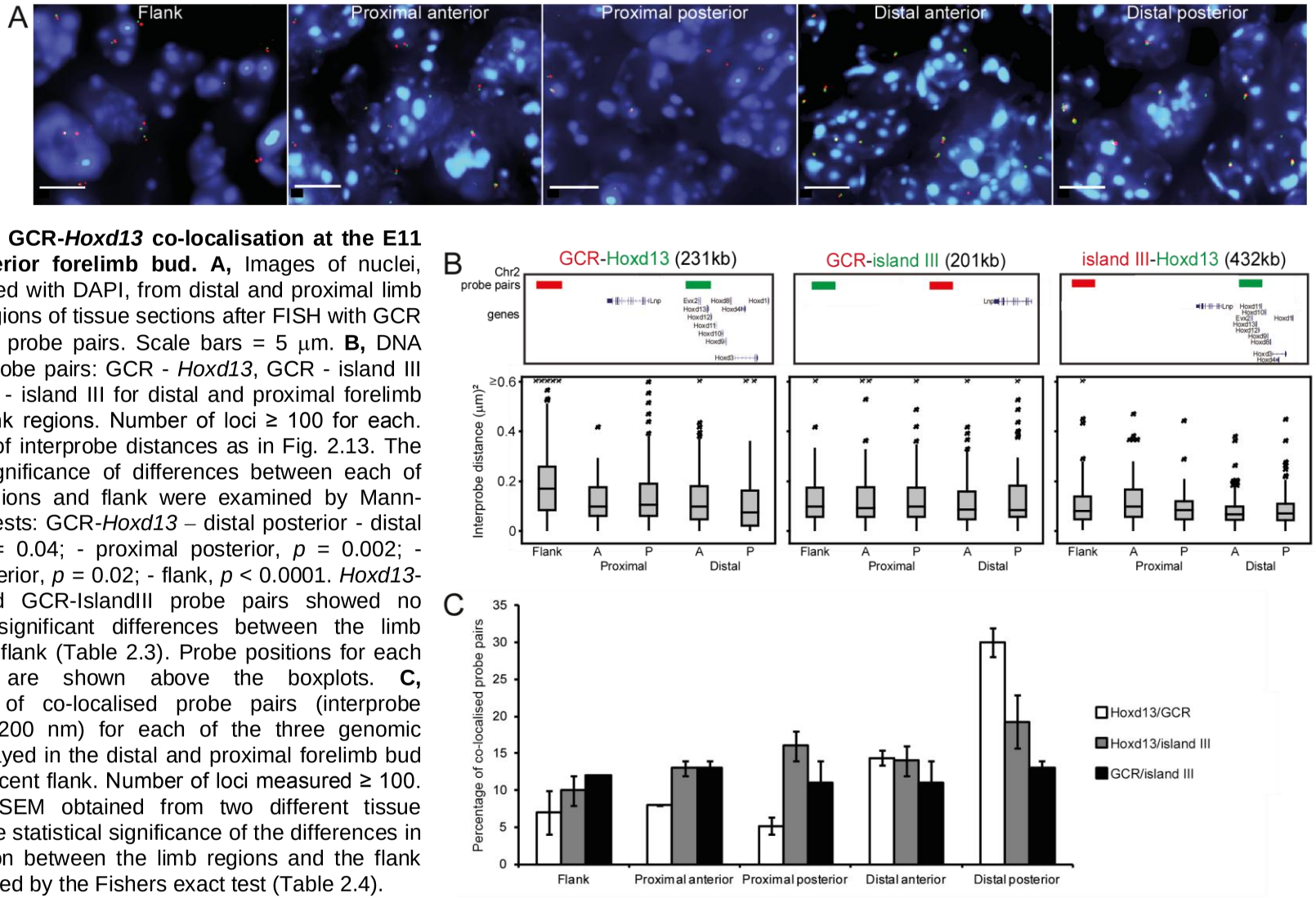

Figure 2.14. GCR-*Hoxd13* co-localisation at the E11 distal posterior forelimb bud. A, Images of nuclei, counterstained with DAPI, from distal and proximal limb and flank regions of tissue sections after FISH with GCR and *Hoxd13* probe pairs. Scale bars = 5 μm. **B,** DNA FISH with probe pairs: GCR - *Hoxd13*, GCR - island III and *Hoxd13* - island III for distal and proximal forelimb bud and flank regions. Number of loci ≥ 100 for each. Distribution of interprobe distances as in Fig. 2.13. The statistical significance of differences between each of the limb regions and flank were examined by Mann-Whitney U tests: GCR-*Hoxd13* – distal posterior - distal anterior, p = 0.04; - proximal posterior, p = 0.002; - proximal anterior, p = 0.02; - flank, p < 0.0001. *Hoxd13*-IslandIII and GCR-IslandIII probe pairs showed no statistically significant differences between the limb regions and flank (Table 2.3). Probe positions for each experiment are shown above the boxplots. **C,** Percentage of co-localised probe pairs (interprobe distance < 200 nm) for each of the three genomic regions assayed in the distal and proximal forelimb bud and the adjacent flank. Number of loci measured ≥ 100. Error bars SEM obtained from two different tissue sections. The statistical significance of the differences in co-localisation between the limb regions and the flank were examined by the Fishers exact test (Table 2.4).

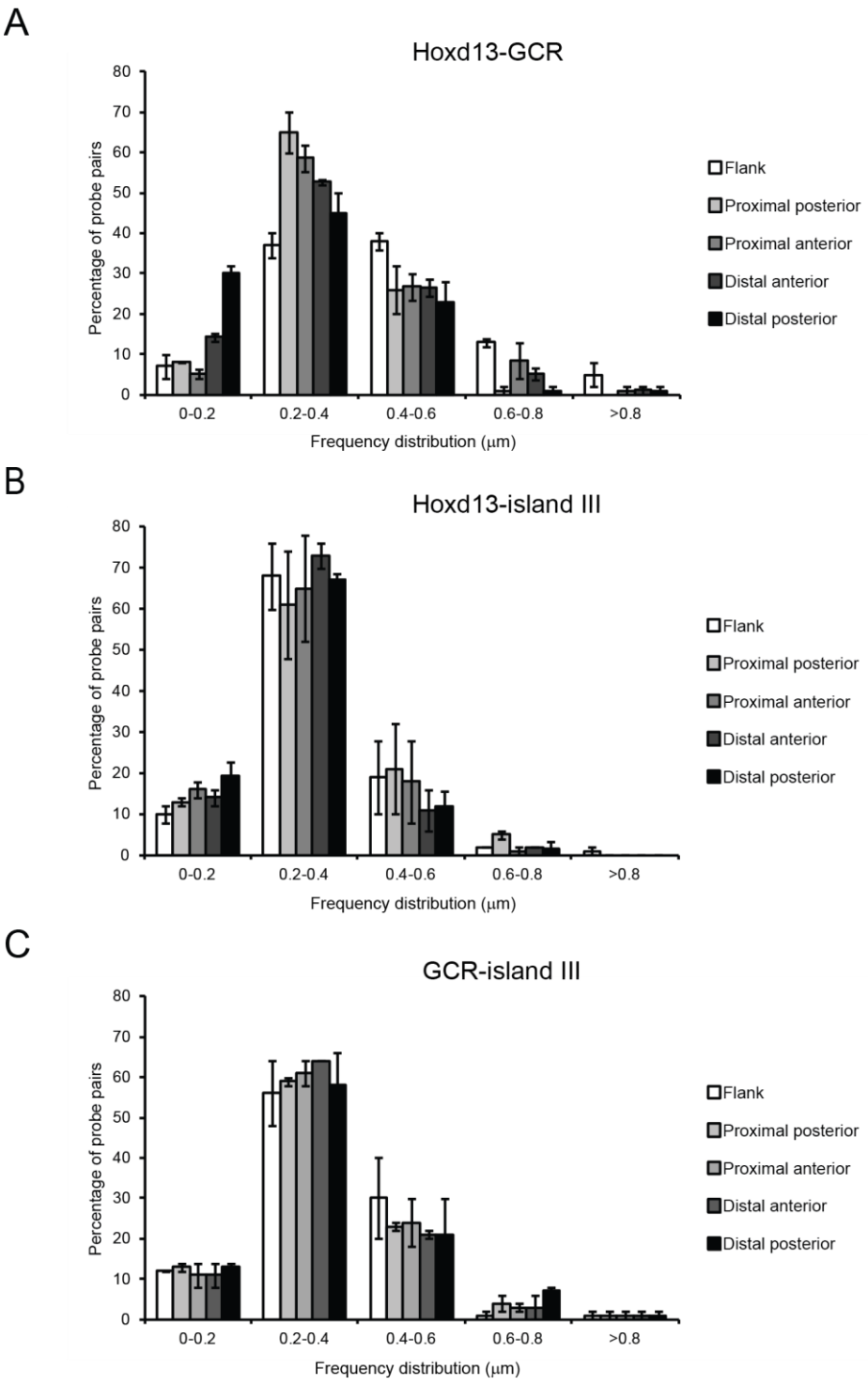

Figure 2.15. Frequency distributions of the interprobe distance (d) between the GCR enhancer, *Hoxd13* and island III probes. Frequences were calculated for every 0.2μm distance interval (*x*-axis) between **A,** GCR – *Hoxd13*, **B,** *Hoxd13* - island III and **C,** GCR - island III in nuclei derived from proximal and distal E11 forelimb bud regions and the adjacent flank region. The location and plane of the tissue sections are shown in Figure 2.11. Number of loci measured ≥ 100. Error bars represent SEM obtained from two different tissue sections.

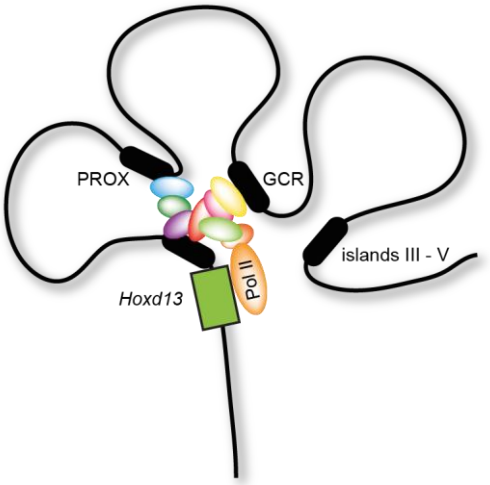

Figure 2.16. Possible higher-order chromatin conformation of the 5′ HoxD regulatory region. *Hoxd13* expression at E10.5-11 initiated following interaction with the GCR and Prox enhancers. Measurement of interprobe distance between *Hoxd13* – GCR, *Hoxd13* – island III and GCR – island III indicates that the nuclear distances separating all three regions are similar. FISH data presented here and 4C data from E12.0 embryos (Montavon et al., 2011) suggest that the locus has a compact conformation that may result from multiple chromatin mini loops enabling direct interaction between the enhancers and their target genes. But in this study, only the GCR showed significantly greater co-localisation with *Hoxd13*, implying that the island enhancers identified in the previous study, although situated within the regulatory region, are not yet active but are required for 5′ *Hoxd* expression at later developmental stages and/or across more anterior limb regions. (islands I & II not shown but located beyond the island III – V locus (Fig. 2.8). Prox previously shown to drive limb 5′ *HoxD* expression (Gonzalez et al., 2007). For ease of interpretation only *Hoxd13* shown.)

2.6 Poised/active enhancer histone modification H3K4me1 marks conserved GCR and Prox loci in distal limb anterior and posterior cells

The histone modifications H3K4me1/3 are well-established markers of regulatory elements (ENCODE consortium, 2007; Barski et al., 2007; Heintzman et al., 2007; Wang et al., 2008) with H3K4me1 the dominant modification over enhancers in most cell types (Robertson et al., 2008; Heintzman et al., 2009) (section 1.1.1.3). Although widely accepted as being an "active mark", H3K4me1 has also been identified over "poised" enhancers in human ES cells, mouse ES cells and several adult tissues (Rada Iglesias et al., 2010; Creyghton et al., 2010.; Zentner et al., 2011).

In order to try and identify active and poised enhancers that might have a role in regulating gene expression in the developing limb, I used the native ChIP protocol modified for small numbers of cells combined with custom high-density tiling arrays, to determine the H3K4me1 profile of cells dissected from the distal anterior and posterior forelimb buds of E10.5 embryos.

The HoxD cluster, and indeed the other three Hox clusters, are pervasively covered with the H3K4me1 mark in both anterior and posterior limb bud cells (Figure 2.17, upper panels, HoxA also shown), similar to H3K4me2/3 coverage in mouse primary lung fibroblasts (Bernstein et al., 2005). However, there is a dip in levels in promoter regions immediately adjacent to transcription start sites (TSSs) of the 5′ *Hoxd* genes and *Evx2*, likely to be a consequence of histone H3 lysine 4 being modified by the tri-methylation mark here (Fig. 2.17, lower middle panel) (Barski et al., 2007, Heintzman et al., 2007). In contrast the β-globin genomic region, including enhancer elements, is free of H3K4me1; as expected from limb bud cell populations composed of non-erythroid precursors.

Similar H3K4me1 modification in anterior and posterior cells at the promoter regions of 3′ and 5′ Hoxd genes was confirmed by qPCR (Fig. 2.18), although levels are greater at *Hoxd1* compared to *Hoxd10* which may be due to the progressive loss of the H3K4me3 mark at this silenced gene (Fig 2.3A). The ubiquitously-expressed β-actin gene's promoter has high levels of H3K4me1 in both cell populations, whereas levels are minimal at the ES cell pluripotency marker *Nanog* (Fig. 2.18).

The chromatin at both the GCR and Prox loci is also marked by H3K4me1, with no A-P difference (Fig. 2.17, upper left panel) – an indication of the capability of these enhancers to activate the 5′ *Hoxd* genes across the distal handplate at later developmental stages (> E11.0 up to E12.5). Analysis of these loci at a higher resolution (Fig 2.17, lower left and centre panels) revealed that the modification is present over highly conserved regions in particular, comparable to the coverage in E12.0 distal limb tissue (Montavon et al., 2011). Closer analysis of the 5′ HoxD region also shows that H3K4me1 is pervasive mostly over intergenic regions, largely correlating with conserved sequences (for example the punctate blocks immediately upstream of the *Hoxd12* and *Hoxd10* promoters), and a further block over a cluster of highly conserved sequences centromeric to the adjacent *Evx2* gene: which may be indicative of more proximal regulatory elements known to be located in and around the HoxD cluster.

Analysis of other genes that are expressed in the distal limb: *Shh* (Chapter 4), *Fgf10* and *Hoxa13* provided further evidence for the presence of H3K4me1 at enhancers active in limb development. H3K4me1 is present not only immediately upstream of the *Fgf10* promoter – the location of regulatory elements previously shown to be capable of driving reporter gene expression on the limb (Sasak et al., 2002) – but also at several conserved intergenic and 3′ loci of this large gene (Fig. 2.17, upper, second from the right, panel). The HoxA cluster is considered orthologous to HoxD and is also active in the limb. *Hoxa13* is expressed in a spatio-temporal manner similar to the late phase of 5′ *Hoxd* expression, and analysis of the genomic region upstream of the HoxA cluster identified a regulatory region similar to the GCR that controls expression of *Hoxa13* and upstream genes – but not the rest of the *Hoxa* genes – in the developing distal limb bud (Lehoczky et al., 2004; Lehoczky and Innes, 2008). This region has been located to intron 4 of the *Hibadh* gene, with further elements centromeric to this gene, and again – similar to the GCR – H3K4me1 is pervasive specifically over regions containing highly conserved sequences (Fig. 2.17, lower right panel). Further array analysis is presented and discussed in Chapter 4, including for the active limb genes *Hand2* and *Grem1* as well as *Shh*. But of note, all the known and potential regulatory elements of these genes and of the 5′ *HoxD* genes and *Hoxa13* do not show any anterior-posterior difference

in H3K4me1 modification, despite all having posteriorly-restricted domains at this developmental stage (E10.5).

Extensive analysis of genomic regions containing genes that are active during the development of the distal handplate has identified the H3K4me1 modification at both proximal and distal enhancers. In E10.5 distal limb buds, long-range elements that drive late phase 5′ *Hoxd* and *Hoxa13* expression are extensively marked by this histone modification in tissue where they are active (posterior) and in which they will be subsequently activated (anterior). Therefore, the H3K4me1 histone modification is not a suitable mark for active limb-specific enhancers.

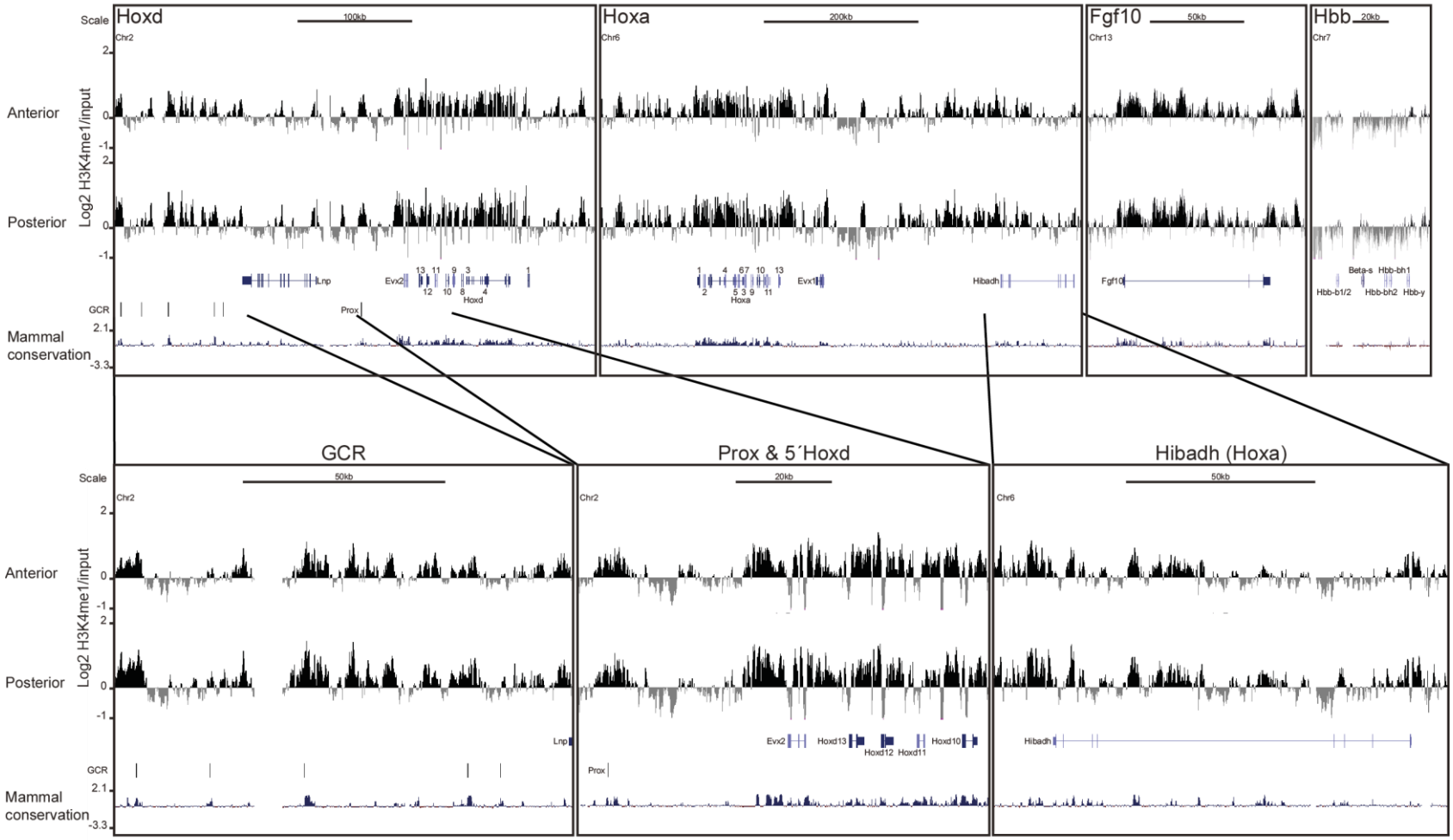

Figure 2.17. H3K4me1 distribution in E10.5 anterior and posterior forelimb tissue. Mean log2 H3K4me1/input at the HoxD, HoxA and Hbb regions and at the *Fgf10* gene. n = 2 biological replicates.

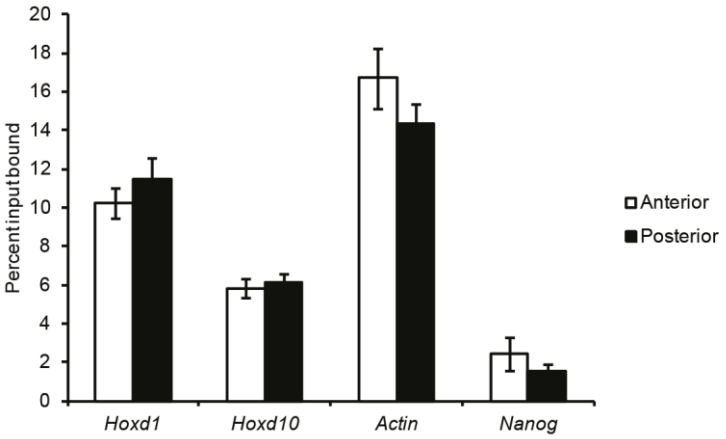

Figure 2.18. Similar H3K4me1 levels at 3′ and 5′ *Hoxd* genes in anterior and posterior limb tissue. ChIP for H3K4me1 at the *Hoxd1*, *Hoxd10* and β-*actin* promoters, and *Nanog* gene by qPCR in distal anterior (white) and posterior (black) forelimb tissue. Enrichment is shown as mean percent input bound ± SEM over 2 biological replicates.

2.7 Discussion

Limb-specific secondary activation of HoxD is initiated in the distal posterior forelimb bud at ~E10.5 in the mouse. I have confirmed that 5′ *Hoxd* gene expression levels are significantly greater in distal posterior E10.5 cells relative to distal anterior (Fig. 2.3), commensurate with the spatial and temporal aspects of late phase reverse co-linearity (Spitz et al., 2003). This A-P difference in expression levels is lost by E11.5-E12.5 as the 5′ genes are activated in more distal anterior regions, but with *Hoxd13* still the most strongly expressed (data not shown) (Montavon et al., 2008).

I have shown that immortalised cell lines derived from the anterior vs posterior E10.5 distal limb recapitulate global expression patterns consistent with their mesenchymal origin from the limb during anterior-posterior axis specification and skeletal morphogenesis (Fig. 2.2). Moreover, these cells maintain the differential *Hoxd* gene expression pattern of the temporal developmental stage from which they are derived (Fig. 2.3).

I identified differential H3K27me3 over 5′ *Hoxd* genes between anterior and posterior derived limb cell lines. In anterior cells, I found extensive H3K27me3 over *Hoxd13* towards *Hoxd10* that is absent from posterior cells (Fig. 2.4). That this contributes to the repression of 5′ *Hoxd* genes in the anterior distal limb is consistent with the ectopic anterior expression of *Hoxd11* and *Hoxd13* in E10.5 limb buds after ablation of the PRC2 component Ezh2 (Wyngaarden et al., 2011). Since *Hoxd* genes have been transcriptionally active during the earlier phase of limb development, I cannot determine whether this represents an active removal of H3K27me3 from the 5′ HoxD in cells derived from the posterior distal E10.5 limb, or whether this pattern results from the re-establishment of H3K27me3 by the PRC2 complex over this region in anterior distal cells. The presence of H3K27me3 over 3′ HoxD in both anterior and posterior derived cells and in dissected limb tissue most likely represents a re-imposing of this repressive histone mark after the initial phase of *Hoxd* gene expression in early, proximal limb bud development (Fig. 2.4).

The late phase 5′ Hoxd gene expression in the posterior distal limb bud is also accompanied by a reduction of Ring1B binding, as expected in a model where PRC1

binding is directed by recognition of H3K27me3 (Fig. 2.8) (Ku et al., 2008), although a low level of PRC1 binding is thought to be mediated by Rybp1 and is independent of H2K27me3 (Tavares et al., 2012). Enrichment for PRC1 was detected across the *Lnp* – GCR locus but levels were low and therefore likely to be background 'noise'. PRC1 is required to maintain Hox clusters in a silent, compact chromatin state in ES cells (Eskeland et al., 2010). Consistent with this I found differential chromatin compaction at HoxD both between the anterior and posterior derived limb cell lines and in limb tissue (Figs. 2.10, 2.11 and 2.13), with a greater degree of decompaction in cells from the posterior distal region than in cells measured in any other region of the limb bud at this stage of development (Fig. 2.13). These data suggest that dynamic polycomb-mediated control of HoxD expression and higher-order chromatin structure, previously described occurring along the main anterior-posterior body axis (Morey et al., 2007) (Spitz et al., 2003) is recapitulated during limb-specific *Hoxd* gene expression.

I also observed co-localisation of the GCR regulatory region and *Evx2*/*Hoxd13* specifically at the distal posterior limb bud (Figs. 2.14 and 2.15A); consistent with the formation of a chromatin loop between enhancer elements of the GCR and the 5′ HoxD at the time and place of distal limb bud development when the GCR is required to initiate HoxD reverse co-linearity and *Hoxd13* expression (Spitz et al., 2003) (Montavon et al., 2008). My analysis at single cell resolution is consistent with average interaction frequencies captured between the 5' HoxD region and GCR by 4C analysis in the limb bud (Montavon et al., 2011). However, there was no analysis of anterior vs posterior regions of the limb in that latter study. By contrast I did not find co-localisation of 5′ HoxD and a more distal regulatory region (island III) (Fig. 2.14C and 2.15B), previously identified by 4C (Montavon et al., 2011). However, that study was conducted from across limbs from a later developmental stage (E12) and therefore the co-localisation might occur later than E11. My analysis of nuclear distance does indicate that the ~400 kb region extending out from the 5' HoxD locus into the adjacent gene desert is in a generally compact chromatin conformation in all the limb regions tested (Figures 2.14A & B Table 2.3). Thus the entire regulatory domain may be configured to minimize the nuclear search space for enhancer-gene interactions to then occur at the correct time and place (Fig.

2.16). Further analyses over the entire temporal stage of late phase HoxD activity (E10.5 – E12.5) and across the A-P distal axis should clarify the spatial and temporal specificity of the multiple regulatory elements.

I found that both the conserved regions of the GCR and the Prox enhancer of HoxD have similar levels of H3K4me1 in the distal anterior and posterior limb bud (Fig. 2.17); and that HoxD itself is covered with H3K4me1, particularly intergenic regions that contain proximal regulatory elements (Figs. 2.17 and 2.18). Therefore, H3K4me1 is not only present at the enhancers of the 5′ *Hoxd* genes in limb bud tissue in which they are being activated (posterior) but also where they will be activated at a later developmental stage (anterior): in other words, where the enhancers could be described as being in a poised state (Creyghton et al., 2010; Rada Iglesias et al., 2010).

I have also shown that H3K4me1 marks long-range regulatory elements orthologous to the HoxD GCR (Lehoczky et al., 2004; Lehoczky and Innes, 2008) that activate *Hoxa13* in a similar spatial and temporal manner as the second phase of 5′ *Hoxd* expression (Zakany and Duboule, 2007). Moreover, analysis of other genomic regions covered by the high density arrays reveals that H3K4me1 is present at the long-range regulatory elements of many non-Hox genes active during distal limb bud development; including *Fgf10* (Fig. 2.17), *Shh* and proposed elements associated with *Grem1* and *Hand2* (discussed in Chapter 4). This extensive analysis of the H3K4me1 mark in distal limb bud tissues suggests that it is prevalent over both proximal and distal regulatory elements of genes that are expressed during the growth and patterning of the hand plate. An examination of H3K27ac (a proposed mark of active enhancers (Creyghton et al., 2010; Rada Iglesias et al., 2010)) across these genomic regions may differentiate between tissues where the enhancers are active (posterior) and poised (anterior).

I have demonstrated that complex higher-order chromatin dynamics is implicated in the regulation of *Hoxd* gene expression during distal limb development, with two different processes occurring – chromatin decompaction and looping. This is the first demonstration of differential chromatin compaction and enhancer-gene co-localisation across the anterior-posterior axis of the developing limb, as a

previous study analysing the spatial co-localisation of *Shh* and its limb enhancer reported similar co-localisation in both posterior (*Shh* expressing) and anterior (no *Shh* expression) limb regions (Amano et al. 2009).

Before presenting (preliminary) data on the chromatin conformation of the *Shh* regulatory region in limb and neural tissue (Chapter 4), in the next chapter I will describe work carried out by myself and others into the competing regulatory activities of two different sets of ETS transcription factors at the limb-specific enhancer of *Shh* (ZRS).

Chapter 3

The ETS family have opposing functions in the limb that determine Shh spatial expression through the ZRS

3.1 Introduction

Anterior-posterior (A-P) polarisation of the developing distal limb bud is determined by the zone of polarising activity (ZPA), a limb organising region spatially defined at the distal posterior margin by *Shh* expression and the polarising function of which is mediated by SHH (Tickle, 2006; Hill, 2007). Several models have been proposed to explain how *Shh* activity defines digit number and identity (Towers and Tickle, 2009), and recent studies have established that the protein acts as both a diffusible morphogen and mitogen to coordinate digit development by integrating growth with digit specification (Towers et al., 2008; Zhu et al., 2008; see Section 1.3). All models of *Shh* function in directing distal limb asymmetry across the A-P axis are predicated on spatial expression being restricted to the ZPA (Yang et al., 1997; Ahn and Joyner, 2004; Harfe et al., 2004).

Regulation of the spatiotemporal pattern of *Shh* expression in the developing limb bud is encoded by the ZPA regulatory sequence (ZRS) long-range enhancer, located up to 1 Mb from its target gene (Fig. 1.7) (Lettice et al., 2002; Lettice et al., 2003). Point mutations within the highly conserved ~800 bp sequence have been shown to generate ectopic anterior expression of *Shh* in several species (Lettice et al., 2003; Sagai et al., 2004, Gurnett et al., 2007; Lettice et al., 2008; Furniss et al., 2008), which is the underlying cause of the preaxial polydactyly (PPD). The severity of this skeletal malformation is variable, ranging from triphalangial thumb (three phalanges instead of two and rotation of the thumb digit to the same plane as the fingers) to the duplication of digits and tibial aplasia. PPD is a heritable abnormality that occurs in about 1:2000 human births (Hill, 2007), and has been additionally identified in mice, cats, dogs and chickens. The human congenital conditions collectively referred to as ZRS-associated syndromes (Wieczorek et al., 2010) include preaxial polydactyly type 2 (PPD2), triphalangeal thumb polysyndactyly (TPTPS), syndactyly type IV (SD4) and Werner's mesomelic syndrome (WMS) (Lettice et al., 2003; Gurnett et al., 2007; Lettice et al., 2008; Furniss et al., 2008; Semerci et al., 2009; Farooq et al., 2010). Other limb abnormalities that map near to

the ZRS include the recessive condition achieropodia (Ianakiev et al., 2001) and acropectoral syndrome (Dundar et al., 2001).

The concept of "homotypic clustering" posits that within *cis*-regulatory sequences important transcription factors will have multiple binding sites (Wagner, 1999; Lifanov et al., 2003; Gotea et al., 2010). The ZRS contains a highly conserved 7 bp motif (AGGAA$^{G}/_{A}$T) repeated five times (Fig. 3.2A), a sequence that is contained within an 8 bp consensus ETS1 motif ($^{C}/_{G}/_{A}$AGGAA$^{G}/_{A}$T) identified at proposed distal regulatory elements in T cells (Hollenhorst et al., 2009). ETS proteins form a large family of DNA-binding transcription factors related by their highly conserved ETS binding domain (Sharrocks, 2001). The 28 human *ETS* genes are divided into subfamilies based upon function that can be further classified into four groups defined by differences in binding motifs preferentially recognised in vitro (Fig. 3.1) (Wei et al., 2010; Hollenhorst et al., 2011). Nevertheless, all in vitro ETS motifs tested contain an invariant GGA core (Oikawa and Yamada, 2002). Many ETS factors are oncogenic and are involved in regulating gene expression by expediting the assembly of transcriptional machinery at specific promoters and enhancers (Hollenhorst et al., 2011). Two members of the PEA3 subfamily – ETV4 and ETV5 – have been identified as components of an *Fgf*-induced pathway that inhibits anterior expression of *Shh* in the mouse developing limb (Zhang et al., 2009; Mao et al., 2009; see also Section 1.3.1).

In addition to ETV4 and ETV5, several other ETS factors have been detected in the distal mesenchyme of early-stage mouse limbs. *Ets1* and *Ets2* are expressed in the distal posterior limb at E10.5 (see Fig. 2.2B and Section 2.2.1 for *Ets1* expression in the P2 E10.5 posterior cell line) and by E11.5 can be detected across the distal mesenchyme (Fig. 3.2C, L. Lettice) (Ristevski et al., 2002). In situ hybridisation analysis also detected *Etv4* and *Etv5* expression across the distal mesenchyme at E11.5 (Fig. 3.2C, L. Lettice). In addition *Gabp*α mRNA in the distal mesenchyme (Ristevski et al., 2004) and *Elf1* expression throughout the limb (Richardson et al., 2010) have previously been reported in the developing murine limb bud. Immunoblotting showed that these proteins are present in both the anterior and posterior limb bud regions at E11.5, and are enriched in nuclear extracts (Fig. 3.2D,

S.Peluso). Of all the known point mutations within the ZRS that engender PPD, none are located within the ETS consensus sites; but mutations that cause PPD2 in two families – Family A & C (AC) (Gurnett et al., 2007) and an Australian family (AUS) (E. Da Graaff, personal communication) – convert the surrounding sequence at the 3′ end of the ZRS into additional ETS binding sites (Fig. 3.2A). Transgenic analyses by the Hill lab and others (Lettice et al., 2003; Furniss et al., 2008; Lettice et al., 2008) have demonstrated that the ZRS can drive LacZ reporter gene expression in the expected posterior domain of the limb. The ZRS with the AUS mutation can also drive reporter expression in the anterior margin of E11.5 limb buds, and the mutation engendered an increase in the posterior expression domain deeper into the middle of the limb (Fig. 3.2B, L. Lettice).

So, the ZRS contains multiple ETS binding motifs, the ETS factors ETV4 and ETV5 have been shown to prevent ectopic anterior *Shh* expression, and yet point mutations within the ZRS that give rise to additional ETS sites results in anterior activation of *Shh* and as a consequence PPD2. In this project I was part of a lab effort to delineate whether and how the ETV factors, and possibly other ETS proteins expressed in the limb, exert control over *Shh* expression through binding to the ZRS.

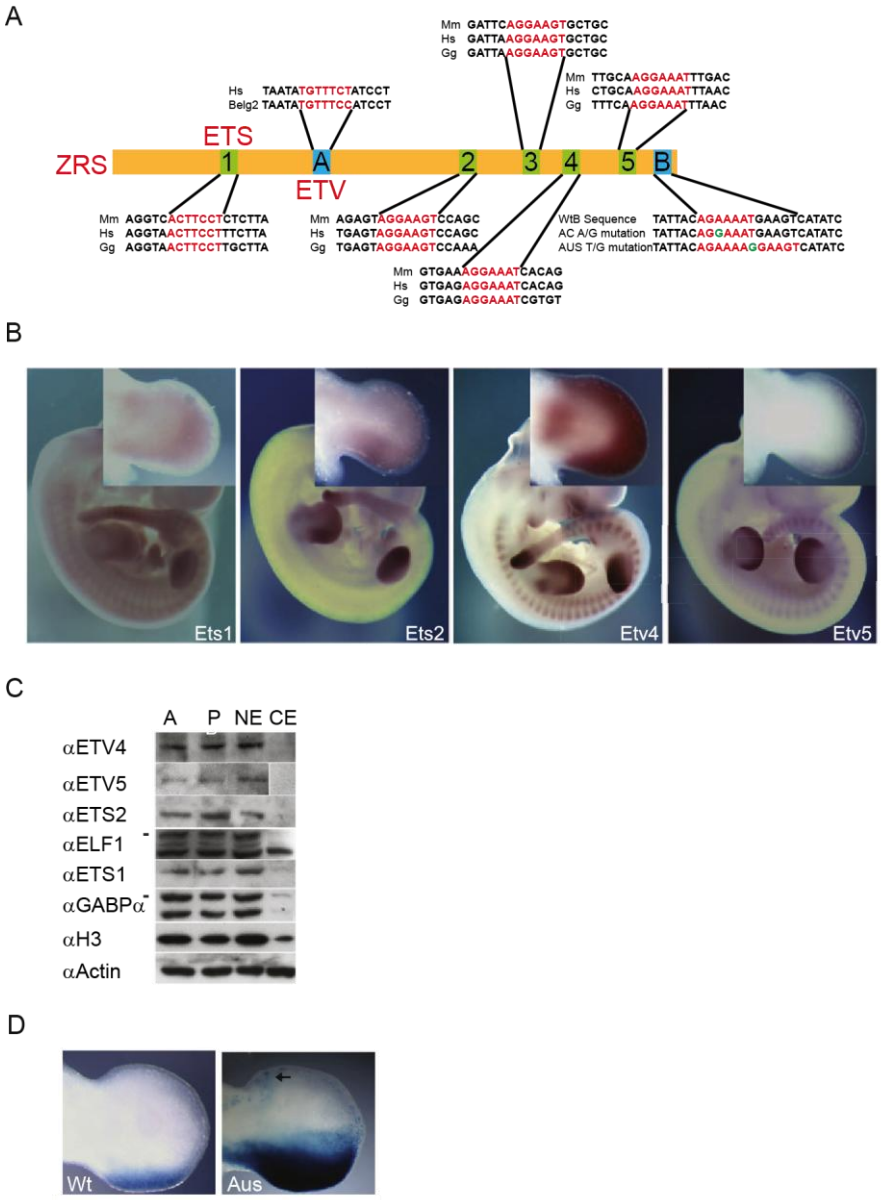

Figure 3.1. Several ETS factors are expressed in the early limb bud and point mutations within the ZRS that result in additional ETS motifs cause ectopic expression of *Shh*. A, Schematic showing the location and conserved sequences of the ETS motifs (numbered 1-5, green boxes) and the two non-canonical ETS sites we have termed ETV motifs (A & B, blue boxes). The sequence in and around ETV site B is the location of the AC and AUS mutations that create additional ETS motifs. Site A is the location of another point mutation that causes PPD in a Belgian family (Belg2). **B,** In situ hybridisation analysis for *Etv4*, *Etv5*, *Ets2* and *Ets1* in E11.5 embryos (insets show forelimb buds). **C,** Western blot analysis using antibodies raised against ETV4, ETV5, ETS2, ELF1, ETS1, GABPα, histone H3 and Actin, with nuclear extracts from E11.5 anterior and posterior limb buds. For comparison, whole limb nuclear and cytoplasmic extracts are shown. H3 and Actin are loading controls. **D,** E11.5 forelimbs from transgenic animals carrying either wild-type or mutant (AUS) ZRS reporter constructs show that the mutation results in expansion of posterior activity and ectopic anterior expression. (Modified from Lettice et al., 2012. **B** & **D** done by L. Lettice, **C** done by S. Peluso.)

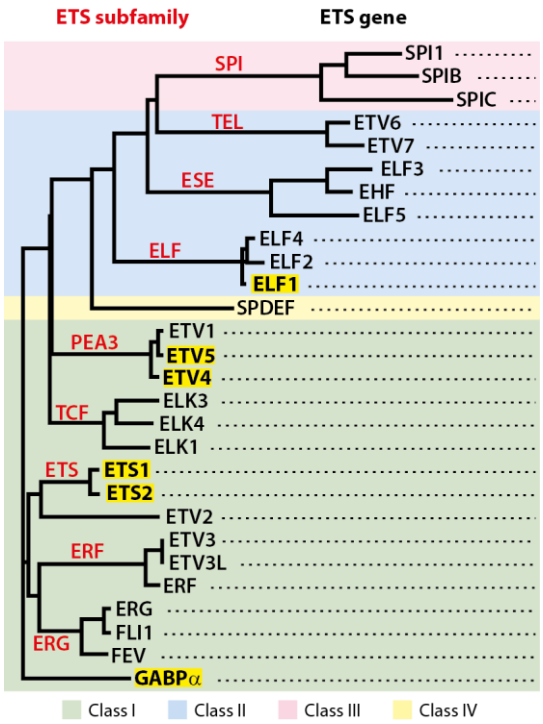

Figure 3.2. Dendogram showing relatedness of human ETS factors. Length of each horizontal line is an indication of estimated evolutionary distance. Branches labelled in red separate sub-families based on function and classes based on DNA binding motifs derived from in vitro analysis are indicated by background colours. Factors highlighted in yellow are expressed in early limb bud development and were analysed in this study. (Modified from Hollenhorst et al., 2011.)

3.2 Differential binding at the wild type and mutant sites of the ZRS and identification of binding factors in vitro

3.2.1. Cell line analysis identifies differential protein-binding profiles at WtB and AUS/AC mutation sequences

Using an immortilised cell line derived from the posterior limb buds of E11.5 embryos (termed 14fp) I delineated the protein-binding profile associated with the WtB, AC and AUS 7 bp motifs, and two of the five highly conserved ETS motifs (1 & 3). Binding profiles were determined by electrophoretic mobility shift assays (EMSAs) using biotinylated double-strand oligonucleotide (ds-oligos) probes and nuclear extracts from 14fp cells. Initial analysis focussed on the AUS and AC sites with 24 bp ds-oligos spanning the WtB or mutant sequences (Fig. 3.2A). The WtB sequence produced a single gel-shifted specific band (band1, Fig. 3.3A), with a lower non-specific band (NS) observed for most ds-oligos. The ds-oligos with the AC mutation produced a strong higher migrating band (band 2), although band 1 was still present albeit at very low levels (Fig. 3.3A). In contrast, ds-oligos with the AUS mutation produced strong band1 and band 2 shifts, therefore exhibiting a binding profile that is a combination of the WtB and AC (Fig. 3.3A); a characteristic that may be due to the positions of both point mutations within the DNA sequence with regard to the ETV site (Fig. 3.2A, see below).

To determine whether the proteins binding to the AC and AUS probes are the same, I ran a series of competition assays using unlabelled AC and AUS ds-oligos with a molar ratio of 1:50 labelled to unlabelled. Unlabelled AC and AUS oligos competed with labelled AUS probes for the proteins that cause both band 1 and band 2 (Figs. 3.3B & C, first two lanes, and previous work done during my master's project). These assays demonstrate that the same proteins are causing the band shifts. And during my Master's project I carried out a series of EMSAs based on the AUS ds-oligos with contiguous 3 bp mutations. The protein responsible for the upper band shift (band 2) binds to the ETS motif of the AUS ds-oligos as this high migrating band is abrogated upon specific mutation of the ETS binding domain.

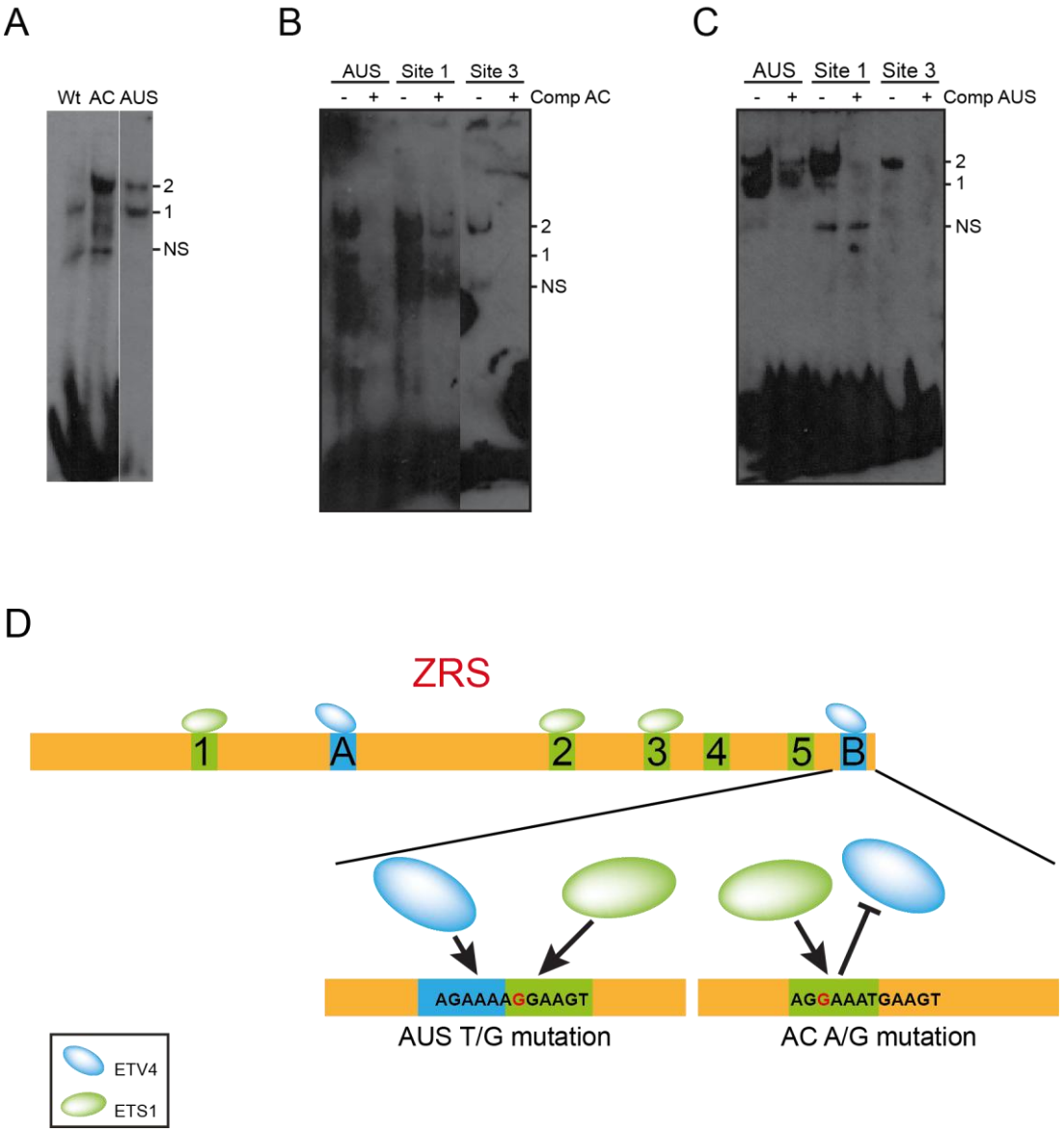

Figure 3.3. Point mutations that result in additional ZRS ETS motifs bind to the same proteins as endogenous sites. A-C EMSA analysis of nuclear extracts from an immortalised cell line (14fp) derived from E11.5 embryo posterior forelimbs. **A,** Protein-binding profiles produced by ds-oligos containing WtB, AUS, or AC sequence. WtB produced a specific band (1), AC generated a higher migrating band (2), and AUS produced a combination of the two. A non-specific band (NS) was generated by all ds-oligos in most assays. **B & C,** Competition by EMSA of the binding for ZRS sites 1 and 3, showing that unlabelled AUS and AC ds-oligos compete for the band 2 generated by both ds-oligos containing the endogenous sites. **D,** Schematic representation of ETS factor binding to ZRS sequences with canonical and non-canonical ETS motifs. In vitro analysis identified ETV4 as being able to bind a core AGAAA motif (sites A and B) while ETS1 binds the AGGAAGT motifs at sites 1, 2, 3 and the AUS and AC mutation sites. Whereas the AC mutation changes the ETV site B to an ETS motif, the AUS mutation creates an additional ETS motif adjacent to the ETV site which may result in competition between the two factors for the binding site; as suggested by the different band shift profiles of the ds-oligos containing either mutation site (schematic figure summarises data from Lettice et al., 2012).

I also analysed two of the five endogenous AGGAA$^{G}/_{A}$T sites within the ZRS: sites 1 and 3 (Fig. 3.2A). Labelled oligo probes for both generated a band that migrated to the same position as band 2 generated by the AUS and AC ds-oligos; and showed specificity for binding by competition with both AUS and AC unlabelled ds-probes (1:50 labelled to unlabelled) (Fig. 3.3B & C). Moreover, previous work showed the protein(s) that cause the band 2 shift binds the ETS motif of both probes as contiguous 3 bp mutations specifically within these sites resulted in the reduction (site 1) or loss (site 2) of this higher migrating band. A weak band that migrates to the same position as band 1 generated by WtB and AUS ds-oligos was generated by the site 1 sequence; and again was competed for by the AUS and, with a lower affinity, the AC unlabelled ds-oligos.

EMSA analysis of the posterior forelimb 14fp cell line suggested that both the AUS and AC mutations disrupt DNA-protein interactions at the WtB locus of the ZRS; the former enabling an additional protein to bind while the latter abrogates the binding of a protein normally present at WtB, which is replaced by the same protein that also binds the Aus AGGAAGT motif. Moreover, this same protein can be detected at two of the endogenous AGGAA$^{G}/_{A}$T ETS sites tested and therefore imply that the AUS and AC point mutations do indeed result in additional ETS binding sites recognised by factor(s) that bind the conserved ETS motifs. To confirm this analysis and identify the factors, further EMSA analysis was undertaken on anterior and posterior forelimb tissue dissected from E11.5 embryos.

3.2.2. Confirmation of differential binding at mutant and wild-type sites and identification of the binding ETS factors in anterior and posterior limb tissue

This section is a summary of the EMSA analysis on nuclear extract from tissue dissected from E11.5 anterior and posterior limbs carried out by other lab members (J. Wiltshire and C. de Angelis) covered in the Lettice et al (2012) (Appendix B). I discuss this work here as it was a continuation of the analysis described in the previous section and feeds into the subsequent in vivo analysis that I undertook to confirm the EMSA data (section 3.3). Both this section and section 3.4 shows the

collaborative effort that went into this work and need to be summarised in order give context to the analysis that I specifically carried out.

Briefly, upon confirming the binding profiles originally identified in the posterior cell line for WtB, AC, AUS and endogenous sites 1and 3 (and demonstrating binding of protein(s) that cause the band shift 2 at site 2), ETS1 was identified as causing band shift 2 while ETV4 caused band shift 1 (Fig.3.3A). The depletion of the protein that causes the WtB band shift following the addition of the anti-ETV4 antibody therefore indicates that this ETS factor binds to a non-canonical AGAAAAT site, the AGAAA core of which is present at the 5′ end of the ZRS that was also shown to bind ETV4 (Figs. 3.2A and 3.3D, sites A and B). This second motif is located at the position of a previously published Belg2 mutation site (Lettice et al., 2003), and analysis of the Belg2 sequence shows that it binds an additional, unidentified factor that generates an upper band (see Fig.2 of Lettice et al., 2012, included with this thesis). So, as Fig. 3.2A shows, whereas the AC mutation changes this site to a canonical AGGAAGT and therefore allows the replacement of ETV4 with ETS1, the AUS mutation leaves the AGAAA sequence intact, which may explain why this mutation could enable ETS1 to bind in addition to ETV4 (Fig. 3.3D).

Extensive EMSA analysis of both canonical and non-cannonical ETS motifs within the ZRS, both endogenous and mutant, suggests that ETS factors do bind to and therefore regulate the ZRS in the limb. These in vitro assays have shown that at least two factors from different subfamilies – ETS1 and ETV4 – are capable of binding to specific sequences within the enhancer. Consequently, I sought to determine whether ETS factors are indeed bound to the ZRS in vivo.

3.3. Identification of ETS factors that bind the ZRS in vivo

To ascertain that Ets proteins bind to the ZRS in the early limb bud I performed chromatin immunoprecipitation (ChIP) on nuclear extracts derived from whole autopods dissected from E11.5 limbs. An initial screen of the ETS factors expressed in the limb by qPCR failed to detect at appreciable levels either ELF1 or ETS2 at the ZRS (data not shown, J. Wiltshire). Therefore, further investigations using high density tiling arrays did not include ETS2, but ELF1 was analysed as a negative control. The arrays were the same as those described in the previous chapter, containing not only the genomic regions in which the Hox clusters are located, but also regions containing genes crucial to early distal limb bud growth and patterning: *Hand2*, *Grem1*, *Fgf10* and of course a 1.2 Mb region covering both *Shh* and the ZRS. In this section I will focus on ETS factor binding to the ZRS but in section 3.5 I describe in more detail the binding patterns of each of the factors tested over all the genomic regions covered by the arrays.

Previous genome-wide in vivo analysis of ETS1 binding revealed co-occupation of GABPα and ELF1 at a significant subset of sites in Jurkat T cells (Hollenhorst et al., 2007; Hollenhorst et al., 2009). Microarray analysis showed that only GABPα was bound along with ETS1 at the ZRS (Fig. 3.4, first panel, fifth and fourth tracks respectively). The GABPα binding profile suggests that it is present throughout the ZRS, whereas ETS1 binding overlapped but was more skewed to the 3′ end. Although the arrays are not of a sufficient resolution to allow the identification of specific binding sites for each of these factors, the more 3′ endogenous ETS sites contain the sequence AGGAA**A**T, while the rest are AGGAA**G**T motifs. Binding of ETS1 to the ZRS was confirmed by qPCR, showing over two-fold enrichment over IgG (Fig. 3.6). Neither GABPα nor ETS1 is bound at an appreciable level at the *Shh* gene or promoter (Fig. 3.4, third panel). ELF1 binding could not be detected at the ZRS on the arrays (Fig. 3.4, third track, first panel), corresponding to the initial qPCR screen and confirmed by qPCR on the same samples as applied to the arrays (Fig. 3.6). Indeed, the array-wide analysis suggests that in the murine embryonic limb ELF1 does not co-occupy the same sites as ETS1 and GABPα but is frequently bound at the same sites occupied by ETV5 (section

3.5, Fig. 3.9). The second panel in Fig. 3.4 shows the promoter region of an uncharacterised gene at the 3′ end of the Hba locus where there is an appreciable level of ELF1 partially overlapping a defined ETV5 peak. Moreover, this ELF1 peak corresponds to a clustering of ETS motifs (second bottom track showing the locations of ETS motifs). Both GABPα and ETS1 are completely absent from this site, as is ETV4, a pattern that develops from a more general analysis of the array data (section 3.5). Although ELF1 enrichment can be detected broadly across *Shh*, most is at background levels, with the possible exception of a minor peak within the first intron that may correspond to a cluster of ETS sites, and unlikely to be significant (Fig. 3.4, third panel, third track).

ETV4 and ETV5 have recently been shown to prevent ectopic activation of *Shh* in the anterior limb bud (Mao et al., 2009; Zhang et al., 2009). Accordingly, both ETV4 and ETV5 were enriched over the probes covering the ZRS locus (Fig. 3.4 first panel, first and second tracks respectively). ETV4 binding was characterised by two peaks either end of the ZRS, reflecting the location of the two predicted ETV sites (Fig. 3.2A). ETV5, by contrast, was enriched over a >5 kb region, with a peak over the whole of the ZRS. The extent of this enrichment is not due to non-specificity binding; as revealed when considering the whole 1 Mb *Shh*/ZRS region (Fig. 3.5) and, more generally, the rest of the regions on the array – many of which are largely free of ETV5. ETV5 binds the ZRS directly; the broad enrichment domain – which can also be observed, to a lesser extent, for ETV4 – may be due to the high density of ETV sites within this region (Fig. 3.4, first panel, bottom track showing locations of ETV motifs). Enrichment of ETV5 at the ZRS was confirmed by qPCR (Fig. 3.6). Both ETV4 and ETV5 were absent from the *Shh* gene, nevertheless a small, discrete ETV5 peak was detected at the promoter region and similar small, discrete ETV5 and ETV4 peaks were located immediately 3′ of the gene; both regions containing clusters of ETV motifs (Fig. 3.4 third panel). However, as for the other ETS factors analysed, the levels of enrichment at the *Shh* locus are quite low and therefore unlikely to be significant.

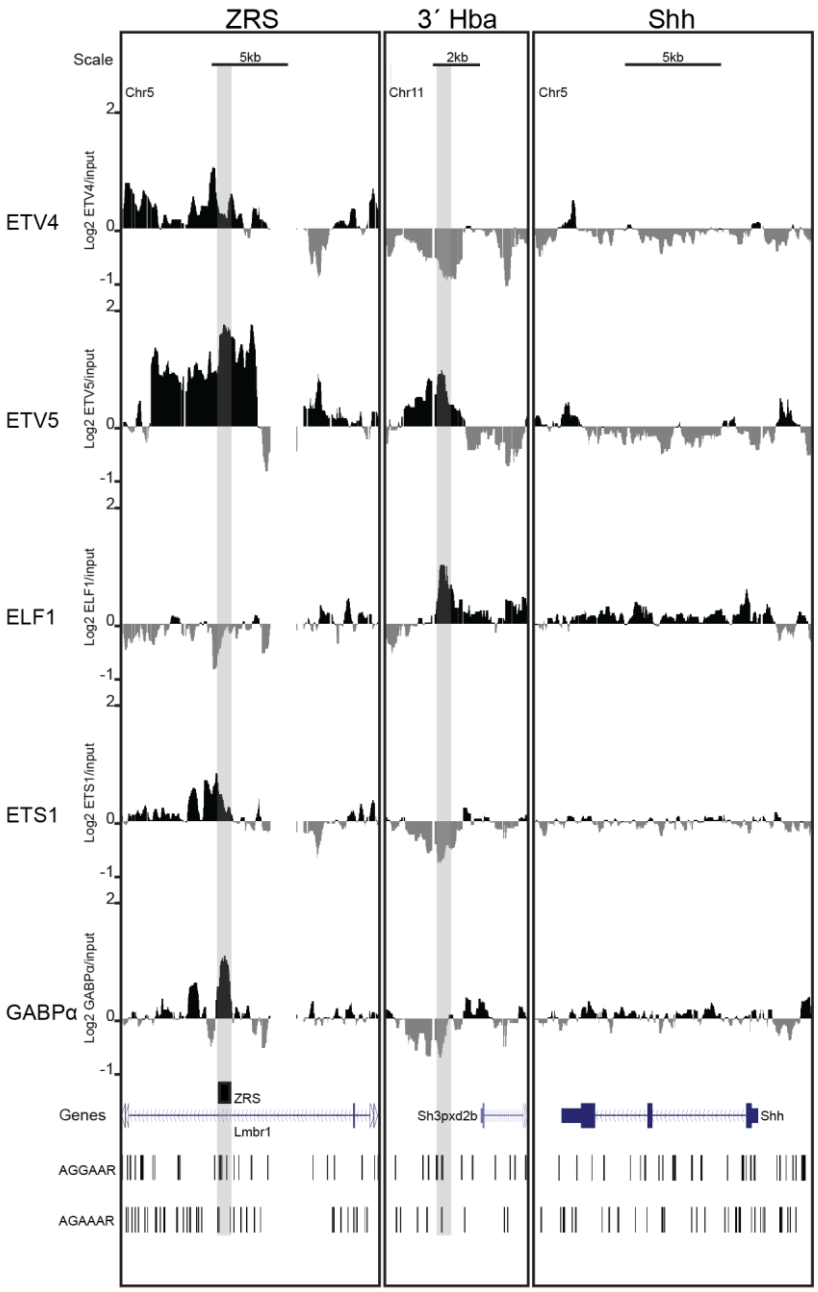

Figure 3.4. ETS factors bind to the ZRS in the distal limb buds of E11.5 embryos. ChIP using antibodies for five different ETS proteins (ETV4, ETV5, ELF1, ETS1 and GABPα) analysed by hybridising equal quantities of enriched and input labelled DNA to high density tiling arrays. Three panels displaying the binding profiles of each factor are presented: the ZRS region and two control regions containing the *Shh* gene plus promoter and a region 3′ of the α-globin locus (3′Hba). The y-axis is Log2 for each ChIP/Input DNA and the x-axis represents a segment of DNA from the array. The positions of potential ETS1/GABPα (AGGAA$^{G}/_{A}$) and ETV4/ETV5 (AGAAA$^{G}/_{A}$) binding sites are shown on the bottom two tracks.

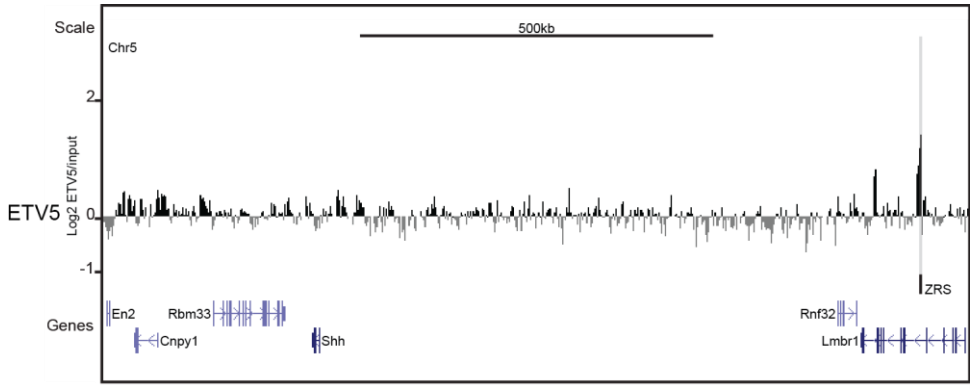

Figure 3.5. Discrete ETV5 enrichment peaks throughout the *Shh*/ZRS genomic region. Panel displaying the ETV5 binding pattern over the entire *Shh* region on chromosome 5 covered by the microarray. Enrichment peaks are specifically high over the ZRS and an intronic region at the 3′ end of the *Lmbr1* gene. Smaller, clustered peaks can also be observed in and around the genes 3′ of *Shh* and upstream of the gene itself, whereas the rest of the >1 Mb region has only background levels or no enrichment.

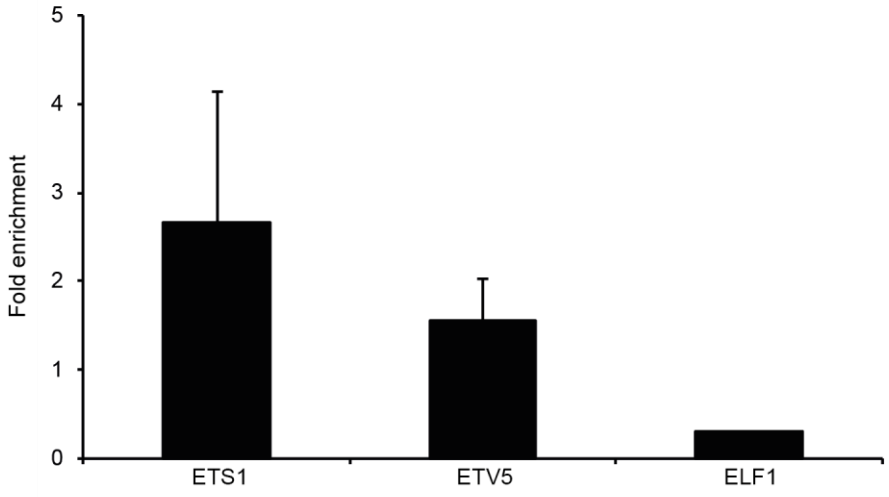

Figure 3.6. ETS1 and ETV5 binding at the ZRS confirmed by qPCR. Enrichment of ETS1 and ETV5 at the ZRS was detected by qPCR of enriched samples compared to ChIP samples using a control (IgG) antibody ± SEM over 3 biological replicates. ELF1 levels were negligible.

3.4 Transgenic assays define the mechanisms by which the opposing ETS factors regulate *Shh* expression through the ZRS in the developing limb

The transgenic experiments and analysis described in the following two sub-sections were done by L. Lettice.

3.4.1 Two mechanisms can induce ectopic *Shh* expression

There are two possible explanations for ectopic expression of *Shh* induced by the AUS mutation: the displacement of (possibly) ETV4/ETV5 at site B reduces the level of these factors, or the creation of an additional AGGAAGT (ETS) site directly drives ectopic expression (Figure 3.2A). Neither deletion nor contiguous 3bp mutation (AGA**AAA**T → AGA**GCG**T) of the ETV site caused ectopic expression in transgenic embryos (Lettice et al., 2012). Nonetheless, the loss of both sites A and B did result in ectopic reporter gene expression (Fig. 3.7D). However, the insertion of an additional AGGAAGT site within a putative neutral position of the ZRS – a variable stretch of Ts that differ in mammalian species – (Fig. 3.7A) generated ectopic expression (Fig. 3.7C); whereas the insertion of seven additional Ts (control) did not (Fig. 3.7B). Therefore, the creation of an additional ETS site (possibly bound by GABPα/ETS1) is sufficient on its own to drive ectopic expression of *Shh*.

3.4.2 Endogenous ETS sites define the *Shh* expression boundary

To study further the consequences for ZRS function of having multiple endogenous ETS sites, each was systematically inactivated (**AGG**AA$^{G}/_{A}$T → **CTT**AA$^{G}/_{A}$T) on its own or in combination with other sites and evaluated by transgenic assays (a full description of these results can be found in Lettice et al. (2012) enclosed at the end of this thesis). The wild type expression domain is shown in Fig. 3.7E. Loss of sites 1 and 3 (the sites with the highest affinity for ETS1 binding in vitro) caused the most extensive reduction in reporter-gene expression (Fig. 3.7H). Loss of additional sites did not reduce the reporter gene's expression domain further (Fig. 3.7I & J). The presence of either site 1 or site 3 on its own was sufficient to generate expression domains approaching the wild-type construct (Figs. 3.7K & L); whilst together the expression boundaries were indistinguishable from wild-type (Fig. 3.7M).

Disruption of site 1 in the transgenic construct containing the ZRS with the AUS mutation had little effect on the extensive posterior expression domain and did not prevent ectopic expression (compare Fig. 3.7N with Fig 3.2B). But ablation of sites 1 and 3 reduced the posterior expression domain to wild type boundaries (Fig. 3.7O). Removal of other sites did not cause further appreciable changes in expression levels (Fig. 3.7P), indicating that a single high affinity site is capable of generating wild-type transgenic expression. Whereas the addition of the AUS mutation results in ectopic anterior expression in the limb, the removal of all five endogenous sites results in no ectopic expression, therefore suggesting that the wild-type sites assist in the ectopic expression (Lettice et al., 2012).

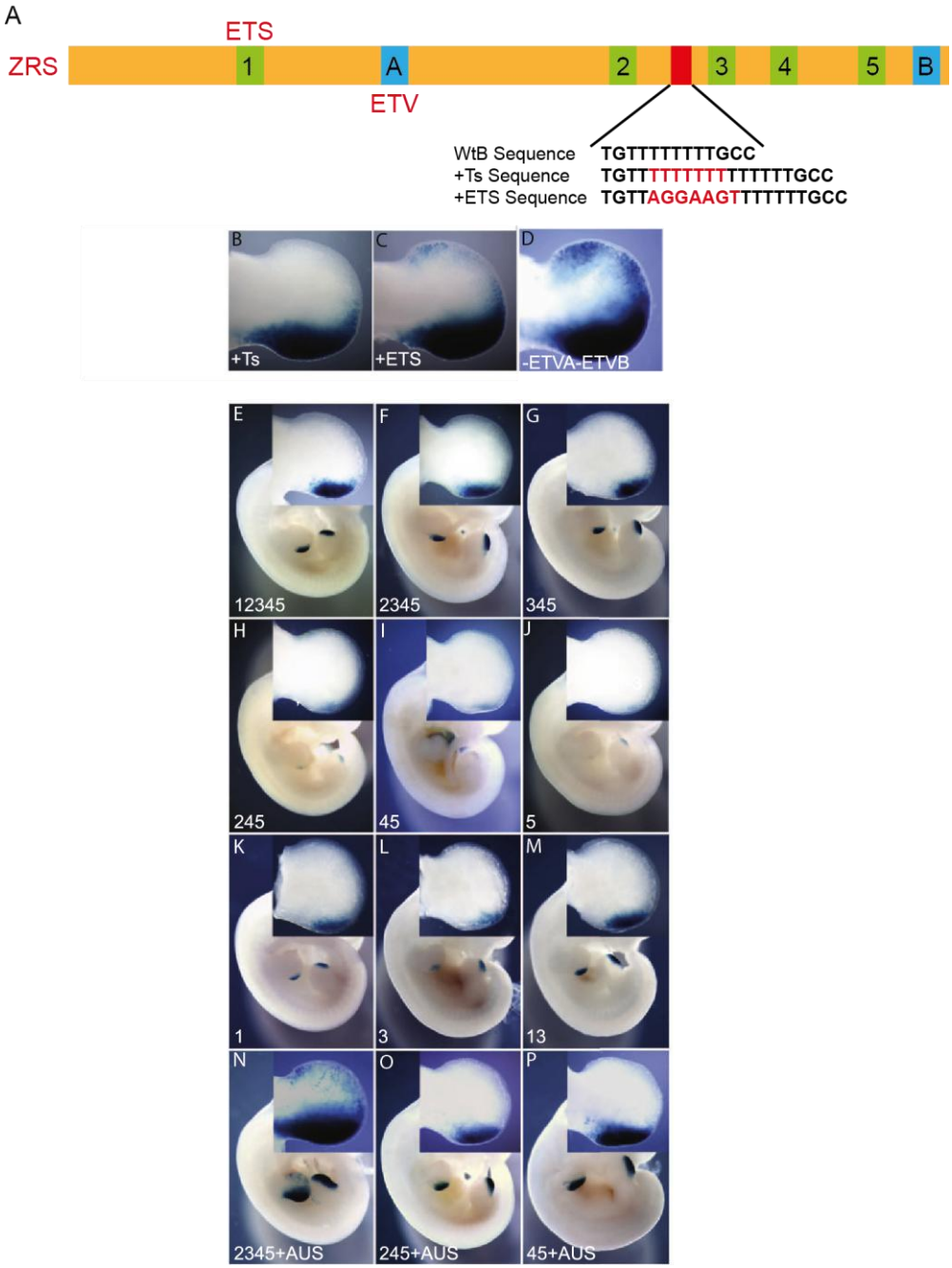

Figure 3.7. Transgenic analysis of embryos carrying mutant ZRS sequences. A, Position of the run of Ts within the ZRS, with the changes shown in red. **B, C,** Expression due to the changes shown in A. Only the addition of the ETS motif (C) results in ectopic expression of the reporter gene. **D,** Disruption of sites A and B (-ETVA -ETVB) in combination results in ectopic anterior expression. **E-M,** Examples of transgenic embryos for endogenous ETS mutations. Remaining ETS sites are depicted in the lower left-hand corner of each image. **N-P,** Transgenic embryos representative of the addition of the Aus mutation with endogenous ETS site mutations. Each transgenic image has close-up of the forelimb in the inserts. (modified from Lettice et al., 2012. B-M done by L. Lettice.)

3.5 Binding profiles of ETS factors across the tiling array genomic regions

3.5.1 ETS factors predominantly bind to intergenic and intronic regions

Recent genome-wide analysis of ETS factor binding in Jurkat T-cells demonstrated that both ELF1 and ETS1 co-localise with GABPα at a significant proportion of sites (Hollenhorst et al., 2009; Hollenhorst et al., 2011). To determine the binding patterns of each of the ETS factors analysed in this study, I quantified the proportion of significant enrichment levels located within discrete genomic territories. Thresholds were determined using the R/Bioconductor *Ringo* package (Toedling et al., 2007) and a previously published algorithm (Schwartz et al., 2006) (see Methods section 6.6.3.2). To account for variation in antibody efficiency and noise levels due to differences in sample DNA quality, enrichment thresholds were calculated separately for each antibody. These thresholds are: GABPα > 0.76, ETS1 > 0.64, ETV4 > 0.78, ETV5 > 0.85, ELF1 > 0.7 (Log2 values). Peaks were scored as containing at least three probes with enrichment above these thresholds within a block of continuous enrichment up to 500bp in size.

Of the tiling array regions analysed, probes covered ~4.7 Mb of the mouse genome in which genes active in early distal limb development are included (*Hoxa* and *Hoxd*, *Shh*, *Hand2*, *Grem1*and *Fgf10*) and ~5.3 Mb containing genes belonging to multiple categories, including maintenance of ES cell pluripotency, ES cell differentiation, and early embryo development. Division of the array data into these two broad categories revealed that the proportion of ETV5 peaks (17%) in the defined limb regions of the array is significantly less than that of the other ETS factors ($p < 0.0001$ (Fishers Exact two tail tests); ETS1, 51%; GABPα, 53%; ETV4, 42%; ELF1, 43%) (Fig. 3.8B.

Each genomic region covered by the arrays were further sub-divided into intergenic, intronic, exonic (protein-coding) and promoter regions (as for the Chapter 2 array analysis). Promoter regions were defined as the 2 kb region upstream of the TSS, therefore possibly encompassing sites that have been previously defined as proximal enhancers (Hollenhorst et al., 2009)); and the number of significantly

enriched peaks located within each of these categories were counted. ETV4 and ETV5 had significantly greater numbers of highly enriched binding sites across the full array compared to the other ETS factors (Table 3.1). Most peaks for all the factors are located in intergenic and intronic regions (from 69% of GABPα peaks to 84% of ETV5 peaks) (Fig. 3.8A, top row of pie charts). Surprisingly, the numbers of peaks located within promoter regions is very low for all factors but particularly GABPα (2%), ETS1 and ETV5 (both 5%); especially considering a previous study on ETS1 that indicated that the majority of binding occurs near TSSs (Hollenhorst et al., 2009). The fourth panel of Figure 3.9 shows the solitary GABPα peak within a promoter from the array analysis (highlighted locus on the right), located at the active limb gene *Hand2*. However, array bias has likely contributed to this discrepancy. Hollenhorst et al. (2009) measured ETS1 binding genome-wide by ChIP-seq and a large proportion of promoter-bound regions were adjacent to housekeeping genes, whereas the arrays used here cover genomic regions containing mostly developmental genes and genes involved in ES cell differentiation. Insofar as the genomic regions covered by the array, then, the ETS factors assayed predominantly bind to non-coding regions remote from gene TSSs in E11.5 mouse limb tissue.

Focussing on the array regions that include the genes involved in distal limb growth and development only reveals that the proportion of binding sites located within intergenic and intronic regions is even greater for all the ETS factors assayed apart from ELF1 (Figure 3.8A, lower row of pie charts). Over 90% of GABPα, ETS1, ETV4 and ETV5 peaks are located in non-coding regions remote from TSSs in genomic regions that include these active distal limb genes, an increase in proportion due to greater relative numbers of peaks situated within intergenic regions. These data clearly show the preponderance of bound distal non-coding sites over more proximal promoter and coding sites for all the ETS factors tested, particularly for the four ZRS-binding factors in regions covered by the array containing genes active in distal limb development.

3.5.2. High correlation between significant binding peaks and ETS/ETV motifs

The majority of enriched sites for GABPα (76%), ETS1 (65%) and ELF1 (69%) contained the ETS motif AGGAA$^{G}/_{A}$ (Fig. 3.8C, Table 3.1), which forms the core of the ETS1-specific motif identified in human T cells (Hollenhorst et al, 2009). Likewise, ETV4 and ETV5 peaks show a high correlation with the non-canonical ETV motif AGAAA$^{G}/_{A}$ identified within the ZRS: 84% and 73% respectively (Fig. 3.8C and Table 3.1). These data further validate the characterisation of AGAAA$^{G}/_{A}$ as a motif recognised and bound by members of the PEA3 sub-family in the mouse, and is especially robust due to the large number of peaks for both factors (255 and 271 for ETV4 and ETV5 respectively).

Peaks for all factors are less correlative with conserved loci (defined as sequences within factor peaks with high mammalian conservation (Fig. 3.9, bottom track)). Nevertheless for GABPα, though the total numbers are relatively low, 56% of the peak sites contain conserved sequences (Fig. 3.8C, Table 3.1). Within intron 4 of the *Fmn* gene on chromosome 2 there is a significant GABPα peak over a locus containing a conserved region and a cluster of ETS and ETV motifs (Fig. 3.9, highlighted region in first panel). Also within the 15 kb region shown here is a more widespread enrichment of ETV4 and ETV5, with a significant peak of the latter co-localised with the GABPα peak. ETS1 is barely above background while ELF1 is largely absent. Panel three of Fig. 3.9 shows a 10 kb locus within a gene desert adjacent to the *Hand2* gene that contains several conserved regions occupied by GABPα. The most significant GABPα peak (highlighted) covers conserved sequences with one ETS motif adjacent to but not within the peak. This conserved region does contain several ETV motifs; however, ETV4 enrichment does not reach the significance threshold for that factor.

Table 3.1. Genomic characteristics of ETS factor peaks

Genomic characteristics	Peaks				
	GABPα	**ETS1**	**ETV4**	**ETV5**	**ELF1**
All	45	88	255	271	98
ETS/ETV sites	34 (76)	57 (65)	213 (84)	199 (73)	68 (69)
Conserved regions	25 (56)	27 (31)	46 (18)	44 (16)	27 (28)

Number and proportion (in parentheses) of ETS factor peaks containing ETS/ETV binding motifs and mammalian regions of conservation for Figure 3.8C. (Log2 enrichment: GABPα > 0.76, ETS1 > 0.64, ETV4 > 0.78, ETV5 > 0.85, ELF1 > 0.7; peaks scored as containing at least three probes with values above these thresholds within a 500 bp window.)

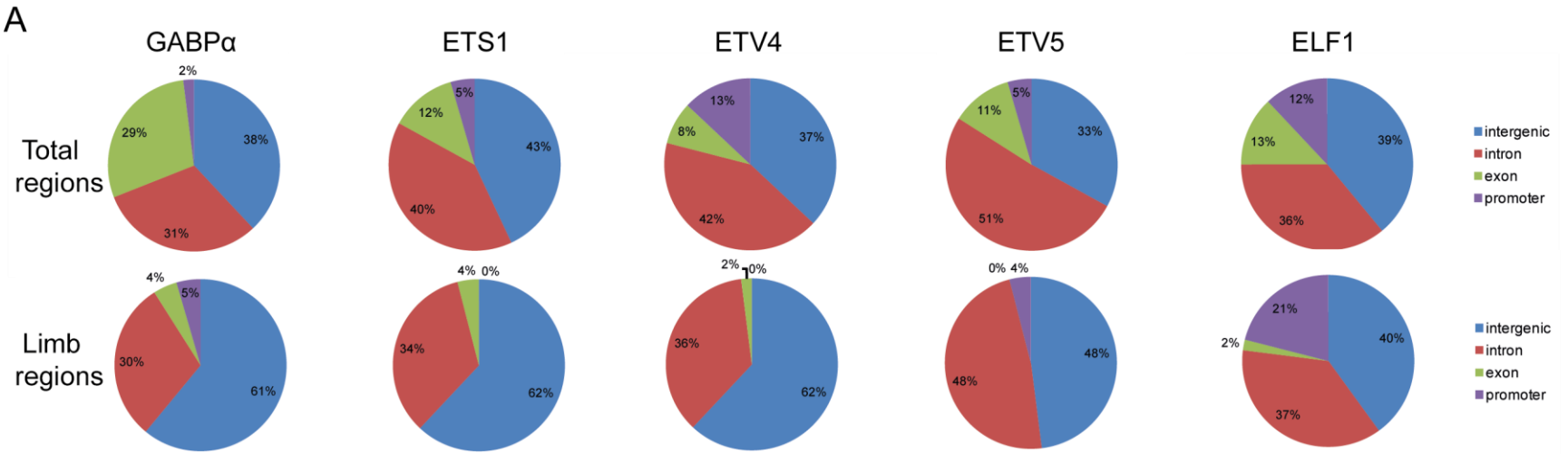

Figure 3.8. Binding profiles of ETS factors in E11.5 murine embryonic limbs. A, Proportion of peaks for each ETS factor within different genomic categories – intergenic, introns, exons and promoters (promoter regions defined as 2 kb upstream of TSS) – across all array genomic regions compared to regions containing genes that regulate distal limb bud development. Log2 enrichment thresholds: GABPα > 0.76, ETS1 > 0.64, ETV4 > 0.78, ETV5 > 0.85, ELF1 > 0.7. Peaks defined as containing at least three probes with values above these thresholds within a 500 bp window of continuous enrichment. **B,** Proportion of enriched peaks located within genomic regions containing genes involved in the patterning and growth of the early limb bud. **C,** Genomic characteristics of enriched peaks. Proportion of peak sites containing ETS/ETV motifs shown by the white bars, while black bars show proportion of peaks corresponding to regions of conservation.

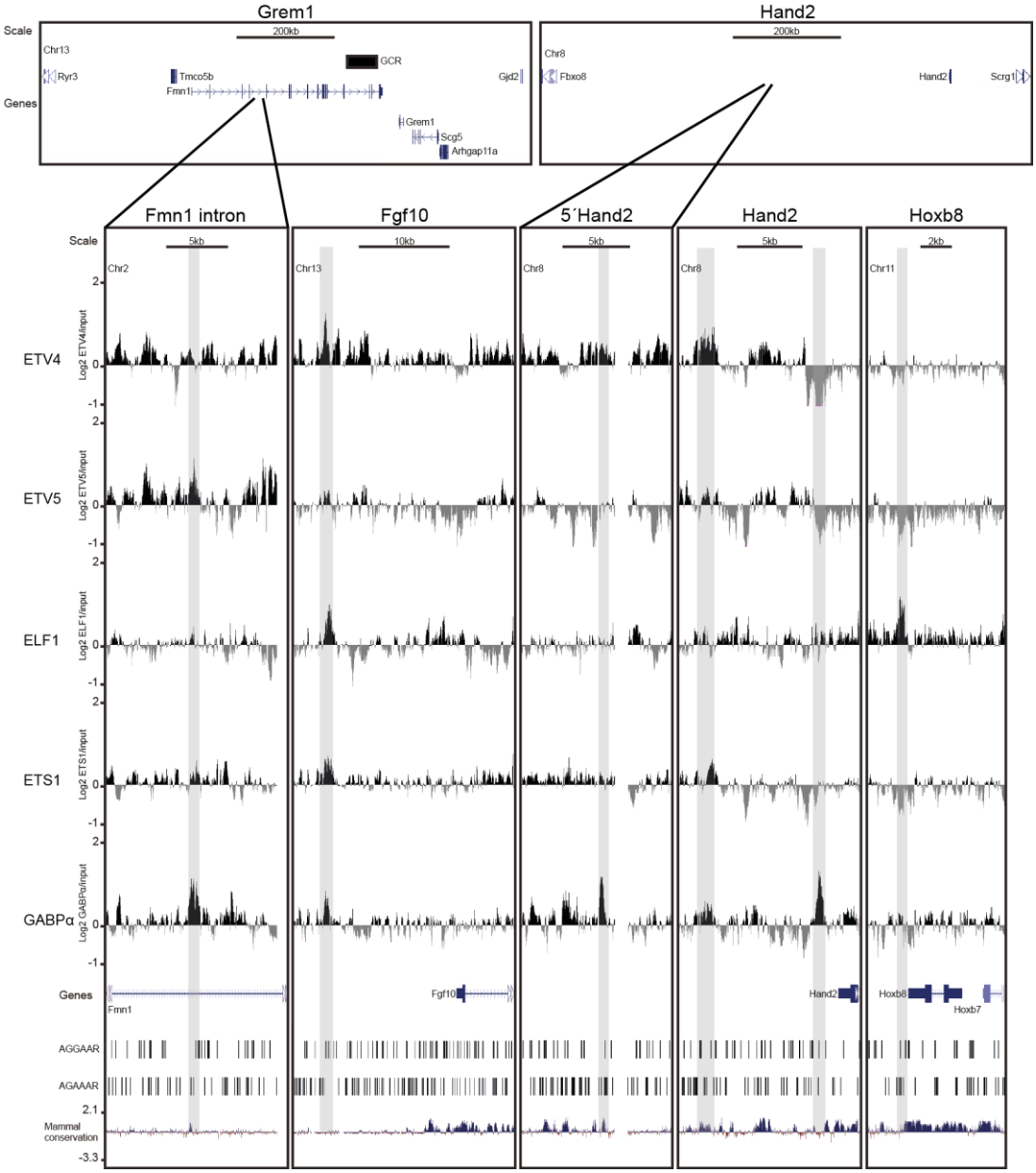

Figure 3.9. Binding profiles of ETS factors at genomic regions associated with genes expressed in the distal limb buds of E11.5 embryos. ChIP using antibodies for five different ETS proteins (ETV4, ETV5, ELF1, ETS1 and GABPα) analysed by hybridising equal quantities of enriched and input labelled DNA to high density tiling arrays. Five panels displaying the binding profiles of each factor are presented: promoter regions of *Fgf10*, *Hoxb8* and *Hand2*; an intronic region of the *Fmn* gene adjacent to *Grem1*; and an intergenic region within the gene desert upstream of *Hand2*. The y-axis is Log2 for each ChIP/Input DNA and the x-axis represents a segment of DNA from the array. The positions of potential ETS1/GABPα ($AGGAA^{G}/_{A}$) and ETV4/ETV5 ($AGAAA^{G}/_{A}$) binding sites are shown on the two tracks below the genes and regions of high mammalian conservation are shown on the bottom track (based on multiple alignments of 19 placental mammals and a further 11 non-placental mammals and non-mammal vertebrates). Full genomic regions containing *Grem1* and *Hand2* covered by the tiling arrays shown above.

3.5.3. Majority of each factors' peaks are not co-localised with the peaks of the other ETS factors

Redundant binding is a striking feature of several ETS factors (Hollenhorst et al., 2011). Comparisons of GABPα (Valouev et al., 2008) and ETS1 (Hollenhorst et al., 2009) genome-wide data reveals that most GABPα sites are also occupied by ETS1 in Jurkat cells. Further comparisons of the GABPα data and genome-wide ELF1 occupancy in Jurkat cells (Wei et al., 2010) also showed an extensive overlap between these two factors from different subfamilies (Hollenhorst et al., 2011). Therefore, I analysed the peaks of each factor in turn to determine the extent of co-localisation of other ETS proteins, in particular GABPα, ETS1 and ELF1 at the same binding regions and the separate co-localisation of ETV4 and ETV5. Only where significant peaks of each factor overlapped were considered; however this highly stringent criteria excluded many sites bound by a significant peak of one factor that had continuous blocks of enrichment of other factors which – although above background levels – were not above the significance threshold set for each factor (Fig. 3.9).

Surprisingly, given the redundancy of the two PEA3 sub-family proteins, ETV4 and ETV5 could only be detected at seven of the same binding sites (3% and 2.5% of their total peaks respectively) (Fig. 3.10A). Of the two, ETV4 has a higher proportion of co-localised binding sites, most of which are co-occupied with ETS1 (32, 12%) and to a lesser extent GABPα (10, 4%) (Figure 3.10C). In the selective genomic regions covered by the microarrays only GABPα (38%) and ETS1 (44%) had significant numbers of peaks co-localised with other factors – mostly with each other and ETV4 (Fig. 3.10C). In contrast to the analysis done on genome-wide data sets (Hollenhorst et al., 2011), ELF1 co-localisation with GABPα and/or ETS1 was minimal (1 and 2 sites respectively, with no sites co-occupied by all three) (Fig. 3.10B). The end panel of Fig. 3.9 shows an ELF1 peak within the promoter of *Hoxb8*. Apart from very sparse GABPα enrichment at background levels, no other ETS factor is present in a locus on the edge of a highly conserved region and containing adjacent ETS and ETV motif clusters.

Co-localisation in this study cannot be construed as co-occupation of the same site in the same cells, as the limb buds from which the tissue was dissected for these assays contain a heterogeneous cell population. But it does suggest that ETS1, ETV4 and to a lesser extent GABPα may bind many of the same sites in the mouse genome during limb development, although probably binding to different motifs rather than motif redundancy, as has been shown to occur at the ZRS in this study. The second panel of Fig. 3.9 shows a binding site co-localised by ETV4 and ETS1 ~10kb upstream of *Fgf10*. A significant ELF1 peak and a below threshold GABPα block were also detected over the same non-conserved genomic region containing ETS and ETV motifs located near an active limb regulatory gene.

This extensive analysis of the binding patterns of the five ETS factors to the genomic regions covered by the tiling array suggests that GABPα and ETS1 co-localise at a substantial proportion of binding sites in mouse embryonic limb cells, though not to the same extent as has been previously shown to occur in more extensive, genome-wide, analysis of human T cells (Hollenhorst et al., 2009). However, the PEA3 sub-family member ETV4 also co-localised with these two at a significant number of sites, probably by binding to the ETV motif rather than redundantly competing for the same binding sequence recognised by GABPα and ETS1. The majority of genomic locations of each factors' binding peaks contained either the ETS or the ETV binding motifs, although conserved sequences were present in a majority of peak sites for GABPα only. Most locations of significant enrichment for all the factors were within intergenic and intronic regions.

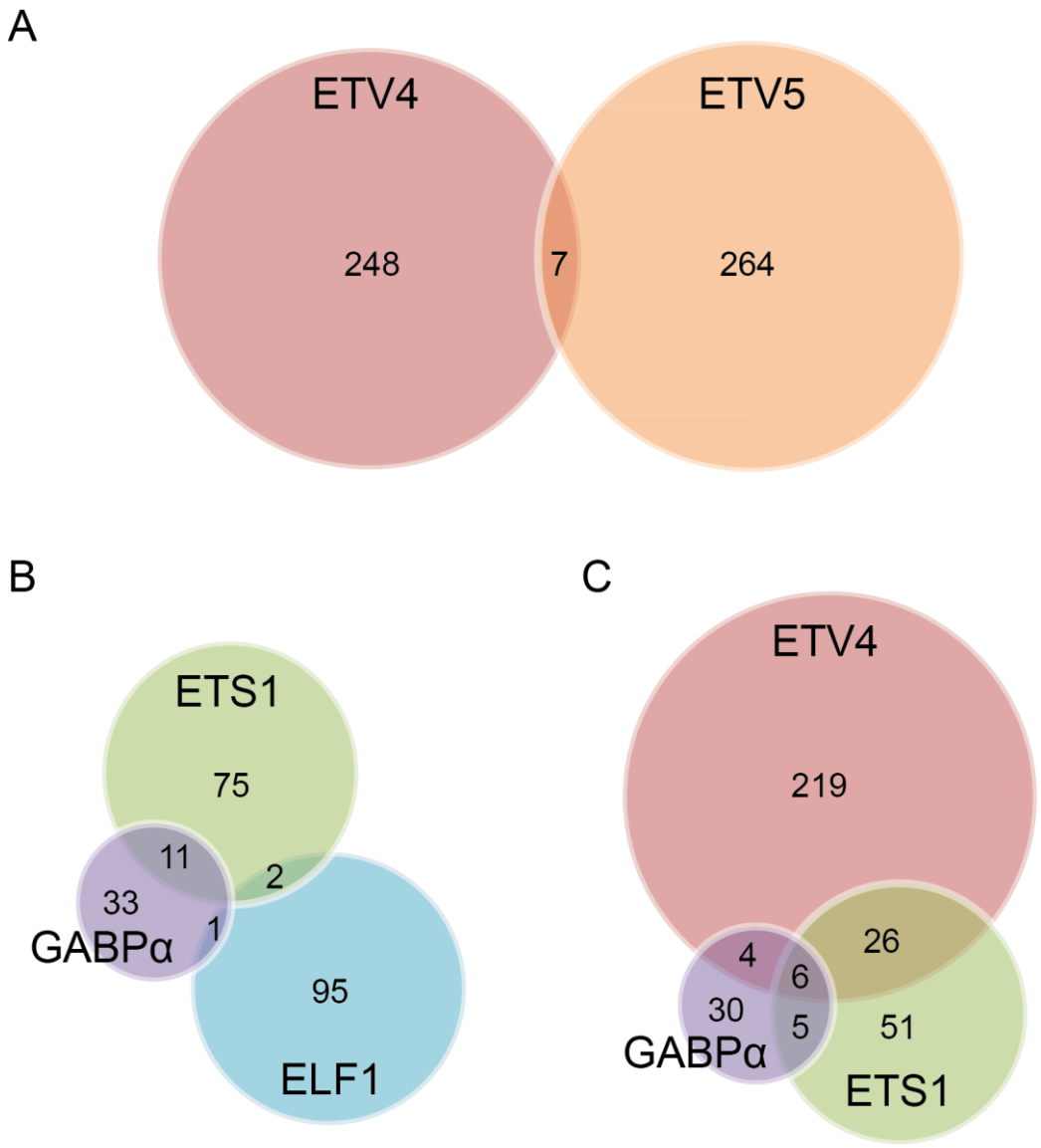

Figure 3.10. Co-localisation analysis at significant binding peaks. A – C, Venn diagrams representing the extent of co-localised ETS factors. The PEA3 sub-family factors ETV4 and ETV5 do not show an extensive overlap in binding sites (**A**). Of the three factors that generally bind to the ETS motif ELF1 does not co-localise with either GABPα or ETS1 to a significant extent, in contrast to the co-occupancy of the latter two at the same genomic sites (**B**). ETV4 also co-localised with GABPα and particularly ETS1 at a significant proportion of the latter two's binding peaks (**C**). Values within each circle are the numbers of non-localised peaks for each factor, values within each overlapping portion the number of co-localised sites shared between different factors. Size of circle and extent of overlap between circles approximately proportional to number of peaks and co-localised peaks respectively.

3.6 Discussion

3.6.1 Opposing functions of two different groups of ETS factors at the ZRS defines *Shh* spatial expression in limb buds

Extensive in vitro analysis has revealed that ETS1 binds to at least three of the five endogenous ETS motifs (sites 1, 2 and 3) and the two mutation sites (AUS and AC) within the ZRS (Figs. 3.2A, 3.4D) (Lettice et al., 2012). This AGGAA$^{G}/_{A}$T motif has been previously described as ETS1-specific in human T cell Jurkat cell lines and is predominantly located at distal regulatory regions (Hollenhorst et al., 2009). EMSA assays were highly suggestive that ETS1, and possibly other ETS factors expressed during mouse limb development – such as ETS2 from the same sub-family which has overlapping, redundant roles in mouse development (Wei et al., 2009), and two members of different sub-families (GABPα and ELF1) previously characterised as co-occupying ETS1 binding sites – bind to and regulate the activity of the ZRS. Failure to detect GABPα binding to the ETS motif was likely due to the failure of GABPα and GABPβ to form the heterotetramic complexes required for enhancing the DNA-binding affinity of the former (Batchelor et al., 1998) in a dilute in vitro environment.

EMSAs also identified ETV4 as being capable of ZRS binding, however, this factor did not bind the ETS motif but rather to the non-canonical AGAAA$^{G}/_{A}$ motif (Figs. 3.2A, 3.3D) (Lettice et al., 2012)). ETV4 and ETV5 have redundant activity in limb development and expression in the distal mesenchyme is maintained by FGF signalling from the AER that represses anterior expression of *Shh* in the limb (Mao et al., 2009; Zhang et al., 2009). The identification of binding sites recognised by one of these PEA3 sub-family members (Fig. 3.2A, sites A and B – site A corresponds to a point mutation that causes PPD in a Belgian family (Belg2)) suggested that this repressive regulatory activity may occur through the ZRS.

Therefore, EMSA analysis is indicative of two possible mechanisms by which the ZRS AC and AUS mutations result in ectopic anterior expression of *Shh*. In the first, the additional ETS site created by the point mutations enables increased binding of active factors, possibly ETS1 alone or in concert with others, which

enables the ZRS to override repressive signals in the anterior limb and induce ectopic *Shh* expression. In the second scenario, the point mutations cause the abrogation of the ETV site B (AC) or the creation of an adjacent ETS motif and the partial loss of the ETV motif (AUS) that reduces ETV4 and/or ETV5 binding at the ZRS in the anterior limb (Fig. 3.3D), thus removing the factors repressing ectopic *Shh* expression.

In order to establish the importance of the ETS and possible non-canonical ETV motifs in the regulation of ZRS activity, I performed a series of ChIP-chip assays to determine which ETS factors expressed in the developing limb bud bind to the enhancer in vivo. These assays identified four different ETS proteins: ETS1 and ETV4, previously shown to bind ZRS sequences in EMSAs, were confirmed along with GABPα and ETV5, which could not be detected in vitro – thereby highlighting the limitation of the cell-free assays. GABPα and ETS1 co-occupancy of the same binding sites has been previously characterised in human Jurkat cells at a genome-wide scale (Hollenhorst et al., 2009), and the tiling arrays used in this study show that co-localisation occurs at the ZRS in murine E11.5 embryonic limb tissue.

Although the ChIP-chip assays were not of a sufficient resolution to determine the specific binding sites of each of the ETS factors within the ZRS, the EMSA analysis indicates that ETS1, and by extension GABPα, are binding to the ETS motifs (1 - 5) and are activating factors, while ETV4 and ETV5 bind to the ETV sites (A & B) and have a repressive influence on ZRS activity. To establish the exact binding sites for each factor and whether GABPα/ETS1 and ETV4/ETV5 compete for the same sites would require performing ChIP-chip on tissue or cells derived from mice with each site knocked out to identify which site(s) are necessary for each factor to bind the ZRS. These assays would be technically difficult and time intensive, but by using reporter constructs containing the ZRS with single and multiple deletions of the ETS/ETV sites and/or with the additional AUS site the importance of each binding motif in ensuring the correct activity of the ZRS could be established.

Through multiple binding sites within the ZRS, GABPα/ETS1 determines the expression boundary of *Shh* that defines the posterior-restricted ZPA (Figs. 3.2B & 3.7E). These AGGAA$^{G}/_{A}$T motifs are common to both factors, and both bind to the ZRS in vivo (Fig. 3.4). The EMSA and transgenic results presented in this chapter suggests that sites 1 and 3 (Fig. 3.2A) have the greatest input, with both acting cumulatively to produce a wild-type expression boundary (Fig. 3.7J). Nevertheless, other sites do participate: sites 2, 4 and 5 drive reporter expression at the extreme posterior margin (Fig. 3.7F) and expand the boundary established by site 3 alone (Lettice et al., 2012).

The addition of the mutant AUS site highlights this combinatorial mechanism as it results in extension of the Shh expression boundary deeper into the centre of the limb bud (Fig. 3.2B). The AUS mutation is the most active site analysed in transgene assays: loss of site 1 still engenders posterior extension of expression and ectopic expression (Fig. 3.7K), while loss of both sites 1 and 3 is compensated for by the addition of the AUS site (Fig. 3.7L & M). The differential expression levels of GABPα but especially ETS1 in the posterior compared to anterior limb (for instance see Fig. 2.2B), combined with the in vitro and in vivo binding data and the transgenic analysis, suggests that ETS factor levels and number of binding sites act together to regulate ZRS spatial activity and consequently the extent of *Shh* expression. This activating regulatory system works in tandem with the repressive activity of a second regulatory system that involves the PEA3 factors ETV4 and ETV5 (Fig. 3.11). Again building on the in vitro binding assays, two non-canonical ETS sites termed ETV motifs A and B (Fig. 3.2 A) were identified in the ZRS that are bound by ETV4/ETV5 (Fig. 3.3D). The abrogation of both sites results in ectopic anterior expression (Fig. 3.7D).

As well as increasing the wild-type expression domain, the AUS mutation also drives anterior, ectopic expression; the known cause of PPD which is the pathogenetic outcome of this point mutation within the ZRS. A previous study suggests that the ZRS is poised for activity not only in the posterior but also the anterior limb bud margin (Amano et al., 2009). The activity of ETV4/ETV5 prevents ZRS-induced *Shh* expression in the wild-type anterior limb bud, but an additional

high affinity ETS site such as the AUS site disrupts the activating/repressing balance, overcoming the repressive influence of ETV4/ETV5 in the anterior limb to activate the poised ZRS and as a consequence effecting the loss of posterior-restricted *Shh* activation.

3.6.2 ETS factor binding profiles suggest a more general limb-specific interaction at distal regulatory regions for GABPα/ETS1 and ETV4

Despite the focus of the tiling array on a selection of genomic regions containing genes involved in ES cell differentiation and early development, the binding profiles of each of the ETS factors analysed may be indicative of a more widespread role in distal *cis*-regulation. As the objective was to elucidate the genomic characteristics preferentially bound by each factor, and not to identify specific regulatory elements, then the likelihood of missing binding sites within possible bona-fide elements due to enrichment values lower than the threshold(s) should not be crucial to the analysis. For example, extensive ETV4 enrichment covers a highly-conserved region ~10 kb upstream of *Hand2* (Fig. 3.9, left-hand highlighted region within panel 4), the location of a branchial arch-specific enhancer of *Hand2* (McFadden et al., 2000). This site corresponds to a cluster of ETV motifs, the ETV4 protein has a known repressive function in the limb (Mao et al., 2009; Zhang et al., 2009; Lettice et al., 2012) and this enhancer has no role in limb development, but enrichment was below the threshold and so this locus had been excluded from the set of ETV4 peaks.

For all five ETS factors tested, the majority of significant enrichment sites are located within intergenic and intronic regions; particularly so for the four ZRS-binding factors at the genomic regions involved in limb development (Fig. 3.8A). Only a small minority of enriched sites were located within promoters. Genome-wide characterisation of ETV4/ETV5 binding has yet to be done for any model system, so the binding profiles presented here await confirmation as to how representative they are of the genomic location of target sites. Human ETS1 binding has been characterised though, using both promoter-specific arrays (Hollenhorst et al., 2007) and whole genome ChIP-seq (Hollenhorst et al., 2009). These studies indicate that

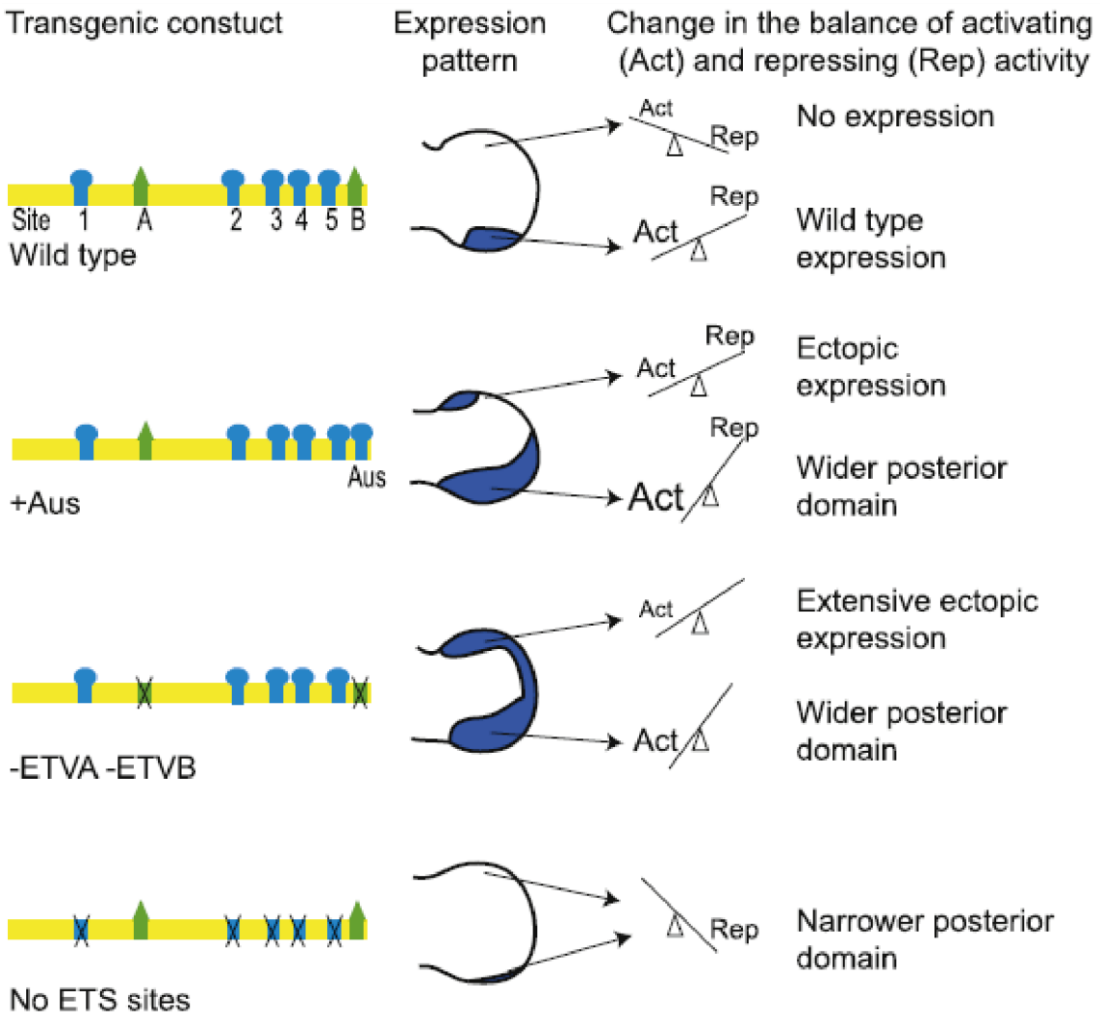

Figure 3.11. Model demonstrating the opposing functions of the two ETS groups. Summary of how four of the transgenic constructs (shown on the left) may interact with ETS factors in the limb. Expression domains for each are shown in the middle, change in activating/repressing balance on the right. (From Lettice et al., 2012)

ETS1 predominantly binds to active promoter regions, in particular of housekeeping genes, containing the highly redundant ETS motif CCGGAAGT – the redundancy of this site may be indicative of different ETS factors regulating these genes in different cell types (Hollenhorst et al., 2011). However, the genome-wide analysis of the human T cells highlighted substantial binding of ETS1 at possible distal regulatory elements containing the more specific AC**A**GGAAGT motif (containing the 7 bp site clustered in the ZRS). Therefore, the predominance of distal binding sites identified by the array data could in large part be due to the bias in the regions covered and the lack of housekeeping genes.

A striking feature of the array analysis is the high co-localisation of ETS1 and ETV4, and to a lesser extent GABPα (Fig. 3.10C). Comparisons of two different data sets for ETS1 and GABPα binding sites in Jurkat T cell lines identified extensive co-occupation of the same sites by these factors (Hollenhorst et al., 2009), which has occurred to a lesser extent in the array data presented here (Fig. 3.10C). Intriguingly, GABPα has a binding domain recognised by the co-activator histone acetyltransferase CBP/p300 (Kang et al., 2008), a protein complex frequently identified at enhancers (Chapter1, section 1.2). Therefore, GABPα may also be a marker of active regulatory elements. ELF1 does not co-localise with GABPα/ETS1, in contrast to analysis of two different genome-wide sequencing studies of GABPα (Valouev et al., 2008) and ELF1 (Wei et al., 2010) which revealed that nearly two thirds of ELF1-bound regions co-localise with GABPα sites (Hollenhorst et al., 2011). ELF1, then, does not have a role in determining ZRS activity, nor, it seems, does it have a collaborative function with GABPα/ETS1 in mouse limb embryogenesis.

Co-localisation of ETV4 at a substantial proportion of GABPα and particularly ETS1 peaks (and vice versa) may indicate that the opposing functions of these two groups elucidated at the ZRS occur at other regulatory elements. Although requiring further verification, the competing active and repressive dynamics co-ordinating ZRS activity, of which these two sets of ETS factors are an intrinsic part of, may be a common mechanism in determining spatial and temporal enhancer activity during development. Of course, it is unlikely that these factors are working

in isolation. For larger enhancer elements such as the ZRS – that have greater scope to bind complex assortments of factors – spatio-temporal activity and the capability to correctly target and regulate genes over large genomic distances will involve the input of multiple factors.

3.6.3 Conclusions

Using electrophoretic mobility shift assays (EMSAs), chromatin immunoprecipitation combined with qPCR and tiling microarrays (ChIP-qPCR, ChIP-chip) and transgenic assays, we show that ETS1 and GABPα occupy the multiple ETS sites and regulate the position of the *Shh* posterior boundary and thus define the ZPA, while multiple binding of ETV4 and ETV5 at non-canonical ETS (ETV) motifs within the ZRS prevents ectopic anterior *Shh* expression. Further, the two PPD2 mutations that result in additional ETS sites disrupts the balance between the active and repressive ETS factors and results in activation of *Shh* in the anterior limb bud and an extension of the posterior domain. From extensive analysis of the microarray data, I show that the majority of significant peaks for each ETS factor are located in non-coding regions remote from TSSs. Moreover, sixty five to eighty four percent of well defined peaks for each factor are associated with either ETS or ETV binding motifs.

3.6.4 Future directions

The ETS binding pattern at the ZRS identified by ChIP-chip cannot be considered as occurring within individual limb mesenchymal cells due to cell heterogeneity of the E11.5 dissected limb tissue. To determine the binding profile of *Shh*-expressing cells requires cell sorting, possibly by the knockin of GFP into the *Shh locus*. Subsequent analysis, either by ChIP-chip or Chip-seq, of *Shh*-expressing and non-expressing distal limb cells will clarify whether the different ETS factors do indeed vary in their binding affinity to the ZRS depending on the spatial position within the distal limb bud. Analysis by ChIP-seq would also determine genome-wide ETS factor-binding

in the distal murine limb bud, thus revealing whether the genomic binding profiles identified in the present study accurately reflect the binding patterns of these proteins during limb development.

Chapter 4

The Shh-ZRS regulatory region is held in a compact higher-order conformation in expressing and non-expressing cells

4.1 Introduction

Multiple tissue-specific enhancers regulate the expression of the sonic hedgehog gene (*Shh*), which codes for a morphogen directing cell fate during organogenesis. Proximal enhancers drive expression in the developing floorplate/neural tube, hindbrain and midbrain (SFPE1); distal enhancers (SBE4) located ~300-400 kb away control *Shh* activity in overlapping domains of the forebrain (Jeong et al., 2006), while the ~800 kb distant limb-specific ZRS enhancer drives expression within the distal, posterior limb bud (Lettice et al., 2003) (Fig. 4.1). Possible *cis*-regulatory elements in and around *Rnf32* adjacent to the ZRS-containing *Lmbr1* have been shown to drive reporter expression in *Shh*-expressing cells within oral, pharyngeal and lung-gut epithelium (Sagai et al., 2009). Unpublished data describes a further neural enhancer at the 3′ end of *Lmbr1* (D. Epstein, private correspondence).

Therefore, restriction of *Shh* activity to specific tissues at particular stages of embryogenesis requires regulation by several enhancer elements spread throughout a ~1 Mb region defined by the near complete absence of other genes (Fig. 4.1): a genomic feature known as a gene desert (Chapter 1, section 1.2.4). This genomic environment is characteristic of many developmental genes, such as the *HoxD* cluster (chapter 2) and *Hand2* (section 4.2.3). Prevention of promiscuous enhancer activity may explain why these enhancers are located in genomic regions largely devoid of protein-coding sequences but not why many are located at such large genomic distances from their target genes; although remoteness from the proximal promoter may ensure that promiscuous expression of the target gene itself does not occur. Of note, all known regulatory elements lie 5′ of *Shh*. A simple looping model enabling direct enhancer-promoter interaction (Fig. 1.6) would not require distal regulatory elements exclusively located to one side of the gene; the concept of topologically separate genomic domains, on the other hand, may account for the 5′ exclusivity of *Shh* enhancers (Dixon et al., 2012; Nora et al., 2012).

How, then, do the limb, forebrain and epithelial enhancers of *Shh* target and activate the gene? Due to the large genomic distances some alteration of higher-order chromatin conformation is likely to be involved. A recent study showed that chromosome conformational dynamics enabled nuclear co-localisation of the ZRS

with *Shh* in murine E10.5 limb cells with the capacity to express *Shh* (Amano et al., 2009). In Chapter 2 I showed that the GCR limb enhancer directly co-localises with the 5′ end of the HoxD cluster in E11.0 distal posterior limb tissue – i.e. the time and place in development where the enhancer activates 5′ *Hoxd* expression. Regulatory elements further into the gene desert centromeric of the HoxD locus have also been suggested to directly interact with *Hoxd13* at a later stage in mouse limb development (E12.5) (Montavon et al., 2011), although unlike the case for the GCR I was not able to visualize co-localisation of one of these with 5′ HoxD (Figs. 2.13B and 2.14C) . These studies suggest that some sort of chromatin looping (described in Chapter 1, section 1.2.5) enables direct enhancer-promoter interactions to occur at the *Shh* and HoxD genomic regions.

However, the mechanism whereby higher-order chromatin structure is manipulated in order for these enhancers to co-localise and possibly directly interact with the promoters of *Shh*/*Hoxd13* has yet to be determined. My DNA FISH analysis of higher-order chromatin structure over HoxD and the 5′ gene desert indicates that this ~500 kb region is in a compact yet flexible conformation (Chapter 2) – which is in accord with 4C data on the same region (Montavon et al., 2011). Therefore, chromatin structure of what could be described as a regulatory region – encompassing several regulatory elements and their target genes – is being held in such a conformation as to optimize enhancer-promoter interaction and exclude the adjacent genomic territories. Could a similar mechanism be occurring at the *Shh* regulatory region? In a series of experiments aimed at elucidating what is required for the ZRS to work over such a large genomic distance, a β-gal reporter cassette with a minimal promoter adjacent to the 5′ end of the ZRS (the minimum portion of the highly-conserved region of the enhancer able to drive reporter expression in the ZPA of transgenic mice) was inserted into the endogenous ZRS site (L. Lettice). Surprisingly, in mice heterozygous for the reporter construct, β-gal expression could be detected throughout all the embryonic tissues where *Shh* is expressed and not just in the distal limb buds as expected. These results imply that the reporter gene is being activated not only by the adjacent limb-specific ZRS but also by all the other tissue-specific *Shh* enhancers. Therefore, during embryogenesis is the ~1 Mb genomic region – from *Shh* to the ZRS – in a similar compact chromatin state as the

HoxD regulatory region, enabling all receptive promoters to be bound and activated by active enhancers? Both the gene in which the ZRS is located and the adjacent gene – *Lmbr1* and *Rnf32* (Figure 4.1) – do not appear to respond to *Shh* enhancers. *Rnf32* and *Lmbr1* are expressed at low levels throughout the embryo and are not up-regulated in the posterior limb (Hill, 2007). This lack of up-regulation in ZRS-active tissue suggests that the enhancer does not function as an alternative, tissue-specific, promoter to produce non-protein coding polyadenylated RNA transcripts reflecting *Lmbr1* structure; a capability recently identified in erythroid-specific intragenic enhancers (Kowalczyk et al., 2012).

In this chapter I will present the preliminary results from on-going research to answer the above question. This project is still in its infancy; therefore the data presented is not complete. I have focused on E11.5 limb and neural tissue to determine chromatin conformation state and possible enhancer-gene co-localisations by DNA FISH. I also present ChIP-chip data from E10.5 distal anterior and posterior limb tissue that reveals H3K27me3 and H3K4me1 histone modification states at the *Shh* and other regions containing genes active during limb development (*Hand2* and *Grem1*). Conformation of the *Shh* region in ES cells, where the gene is not expressed, was also analysed by FISH and I discuss these results too.

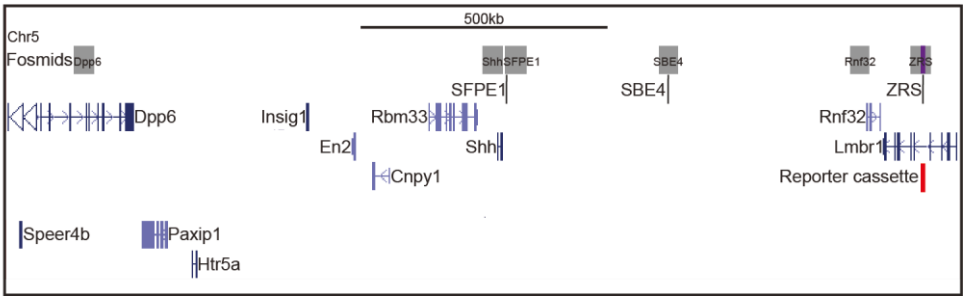

Figure 4.1. *Shh* genomic region indicating the position of fosmid probes. Bottom track shows gene locations with *Shh* situated in the middle. To the right of *Shh* is the ~800 kb gene desert with the locations of the proximal SFPE1 and distal SBE4 neural enhancers indicated, as is the location of the ZRS within *Lmbr1* at the other edge of the gene desert. Below *Lmbr1* the red box indicates the location of the β-gal reporter with 5′ ZRS and minimal promoter insertion. The grey boxes in the top track show the locations bound by the fosmid probes used in the FISH assays. The probe over the *Dpp6* gene is the equivalent distance from *Shh* as the ZRS. The reporter-specific probe is indicated in purple within the ZRS grey box.

4.2. Chromatin topology of the *Shh*-ZRS regulatory region in E11.5 forelimb buds

4.2.1. *Shh*-ZRS region is maintained in what appears to be a highly folded chromatin state throughout the forelimb bud and the adjacent flank

To determine the conformation of higher-order chromatin structure across the gene desert from *Shh* to the ZRS in Shh-expressing and non-expressing limb cells I carried out 3D FISH on tissue sections dissected from the anterior and posterior regions of E11.5 forelimb buds (Fig. 4.2). As for the HoxD analysis in Chapter 2 both distal and proximal regions of the limb bud were analysed, and I also measured compaction levels in the adjacent flank region where *Shh* expression does not occur at any time during embryogenesis. This was an important control as the anterior margin of the limb bud is also primed to express *Shh* (Lettice et al., 2003; Amano et al., 2009; Mao et al., 2009; Zhang et al., 2009). For each of these limb and flank tissues I also measured the nuclear distance (d^2) between hybridisation signals for probes covering *Shh* and the *Dpp6* gene situated the same linear genomic distance from *Shh* as the ZRS in the other direction (Fig. 4.1). This genomic region does not contain any known *Shh* regulatory elements.

The bottom two rows of FISH images in Fig. 4.2 demonstrate the difference in chromatin conformation between the genomic regions 5′ and 3′ of *Shh*. The *Shh-Dpp6* probe pairs are separate from each other in all tissue sections (Fig. 4.2 lower row), while the majority of *Shh*-ZRS probe pairs are closely apposed or overlapping in *Shh*-expressing (distal posterior limb), -primed (distal anterior), and -non-expressing (proximal limb, flank). Calculation of the median interprobe distances (d^2) confirmed the initial visual analysis. *Shh* and *Dpp6* regions have median d^2 separation values of 0.22 - 0.27 μm^2 and a wide spread of inter-probe distances as revealed by the box plots (Fig. 4.3, Table 4.1). In contrast, median d^2 values between the *Shh* and ZRS probes were 0.063 - 0.12 μm^2 (flank) and the range of interprobe distances much closer. The difference in interprobe distances between *Shh- Dpp6* and *Shh*-ZRS is highly significant for all tissues ($p < 0.0001$), whereas there are no significant differences in *Shh-Dpp6* separation distance between the different limb tissues and flank tissue (Table 4.1).

Analysis of the frequency of hybridisation signals separated by defined distance (d) intervals also demonstrates what appears to be significantly greater decompact chromatin conformation of the genomic territory 3′ of *Shh* compared to the 800 kb region up to the ZRS. The proportion of co-localized signals (< 200 nm) was less for *Shh-Dpp6* than *Shh-ZRS* in all tissues (Fig. 4.4A). Conversely, the proportion of signal pairs ≥ 600 nm apart was greater for *Shh-Dpp6* than *Shh*-ZRS: 33 - 40% compared to 5 - 10% respectively (Fig. 4.4B and C, Table 4.2). The proportions of widely separate probe pairs spanning the adjacent genomic territories are highly significantly different in all the limb regions and *Shh* non-expressing flank tissue (Table 4.2).

Comparison of overall interprobe distance (d^2) between *Shh* and the ZRS revealed that the nuclear distance between gene and enhancer is significantly shorter in cells of the distal limb bud (both expressing (posterior) and poised (anterior) tissues) compared to the adjacent flank tissue ($p = 0.04$, Fig. 4.3 and Table 4.1). The difference between proximal anterior and flank is also significant ($p = 0.005$) while comparison of proximal posterior limb tissue and flank tissue reveals no significant difference. The proximal posterior nuclei were located nearer the flank due to these sections containing limb tissue up to the shoulder region, whereas the anterior sections were at the extremity of the limb bud where it forms a bulb type shape at this stage of development (Fig. 4.2, schematic showing position of dissected tissues). As a result the proximal anterior measurements may not be a true reflection of this limb region and therefore further analysis of tissue sections containing proximal anterior tissue up to the shoulder is required. The cells I have classed as proximal anterior are located in a region of the E11.5 limb bud likely to contribute to forearm (radius/ulna) development and therefore influenced by *Shh* activity. Finally, comparisons between proximal and distal limb tissues revealed no significant difference in overall *Shh*-ZRS interprobe distance throughout the E11.5 forelimb bud; in both *Shh*-expressing and non-expressing limb cells this regulatory region is being maintained in a similar (possibly compact) chromatin state (Fig. 4.3, Table 4.1).

In E11.5 murine embryos the *Shh*-ZRS ~800 kb region appears to be configured into a compact chromatin structure in the forelimb bud and adjacent flank. ChIP-chip analysis of E10.5 distal anterior and posterior tissue (a temporal stage when *Shh* expression is established in the distal forelimb bud) revealed that, although the Shh gene itself is coated by H3K27me3, the entire regulatory region located 5′ of Shh is largely free of this histone modification (Fig. 4.5, top left-hand panel) associated with a repressive and compact chromatin state (Chapter 2) (Soshnikova and Duboule, 2009; Eskeland et al., 2010). Similarities with the HoxD locus – key developmental genes with multiple enhancers spread throughout adjacent gene deserts – suggests, as for HoxD, the entire region up to the ZRS is being held in a compact yet flexible chromatin conformation that reduces the nuclear search space for the ZRS to interact with the *Shh* promoter. The surprising result in this context is the close conformation of the locus in the flank cells, where a range of interprobe distances similar to the region 3′ of *Shh* (compare boxplots in Fig. 4.3) would be expected (further analysis of neural tissues and adjacent non-neural cells suggests that this higher-order structure may be a general configuration during murine embryogenesis, see section 4.4.1).

Table 4.1. Interprobe distances for E11.5 limb bud sections

Limb region	***Shh*-ZRS (847 kb)**	***Shh*-Dpp6 (831 kb)**
	Interprobe distance (d^2) (μm^2)	
Distal posterior	0.08	0.27
Distal anterior	0.076 (p = 0.93)	0.23 (p = 0.47)
Proximal posterior	0.076 (p = 0.18)	0.22 (p = 0.43)
Proximal anterior	0.063 (p = 0.58)	0.27 (p = 0.54)
Flank	0.12 (p = 0.04)	0.26 (p = 0.98)

Statistical analysis of data for Fig. 4.3. Interprobe distances are median values, *p*-values from Mann-Whitney U Tests.

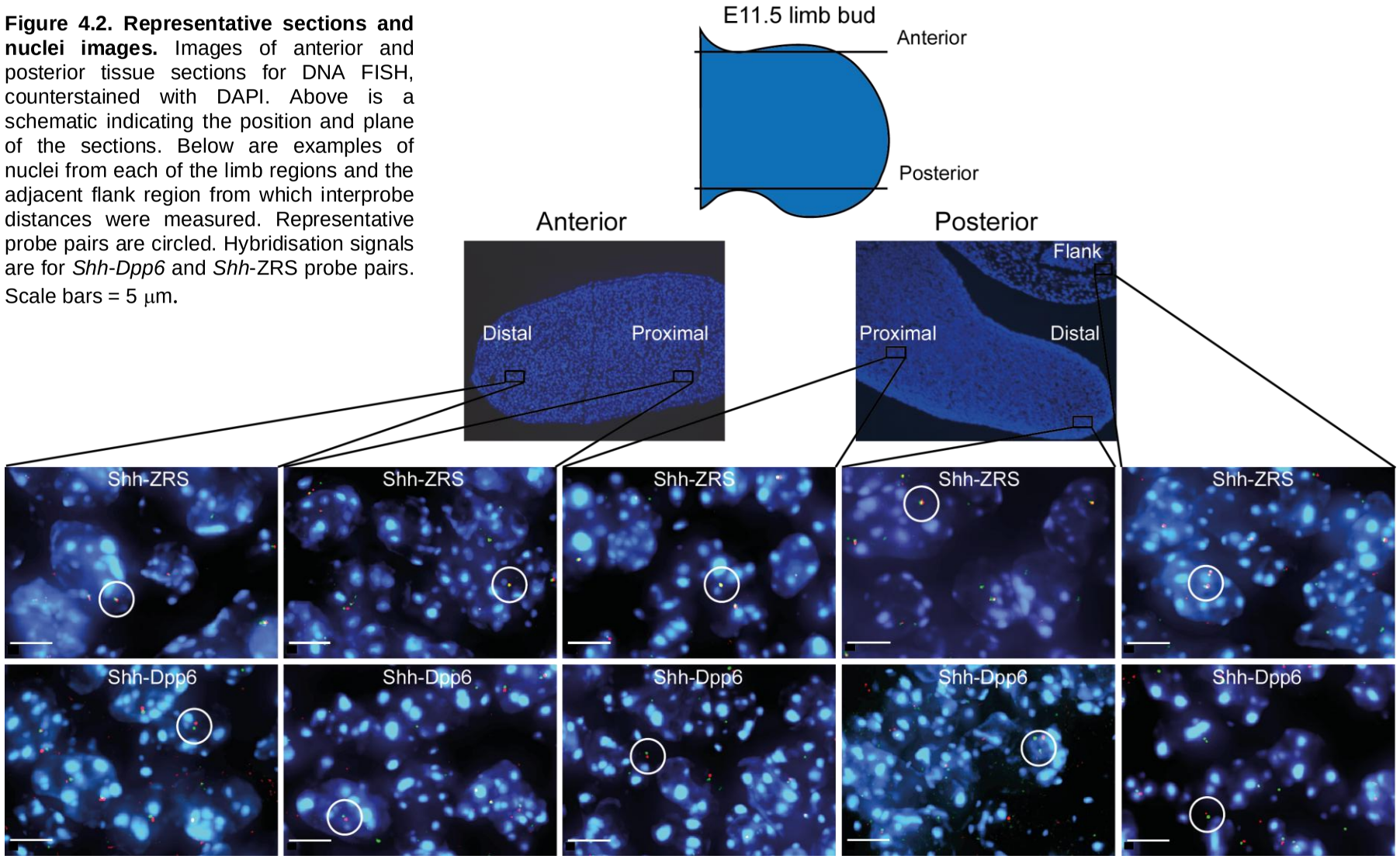

Figure 4.2. Representative sections and nuclei images. Images of anterior and posterior tissue sections for DNA FISH, counterstained with DAPI. Above is a schematic indicating the position and plane of the sections. Below are examples of nuclei from each of the limb regions and the adjacent flank region from which interprobe distances were measured. Representative probe pairs are circled. Hybridisation signals are for *Shh-Dpp6* and *Shh*-ZRS probe pairs. Scale bars = 5 μm.

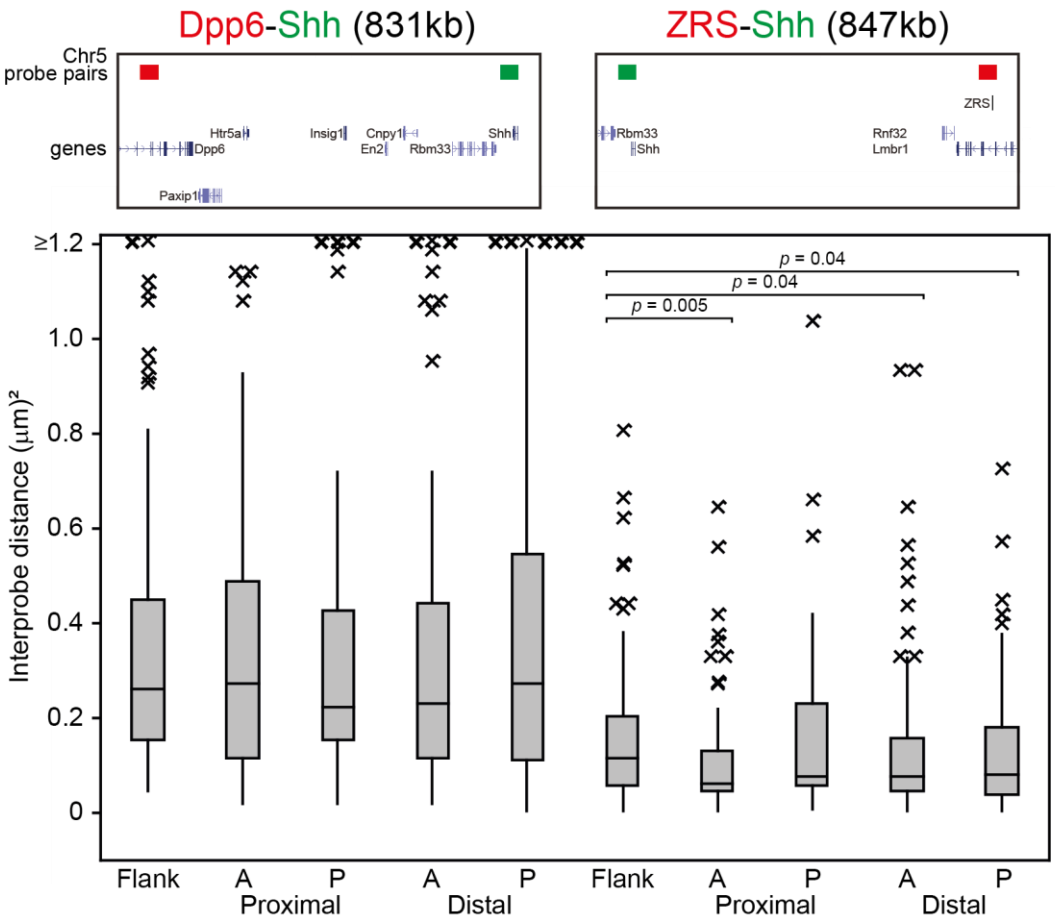

Figure 4.3. The *Shh*-ZRS regulatory region is maintained in a compact chromatin structure in *Shh*-expressing, -primed and –non-expressing E11.5 murine forelimb buds and adjacent flank. Box plots show the distribution of interprobe distances (d^2) at the adjacent *Shh-Dpp6* and *Shh*-ZRS genomic regions for the proximal and distal anterior and posterior forelimb bud and the adjacent flank. Shaded boxes show the median and interquartile range of the data; asterisks indicate outliers. Number of loci: *Shh-Dpp6* = 90, *Shh*-ZRS 90 - 97. Probe positions shown above. The statistical significance of differences between each of the limb regions and flank were examined by Mann-Whitney U tests (Table 4.1).

Table 4.2. Frequency of Shh-ZRS and *Shh-Dpp6* probes separated by ≥ 600nm for E11.5 limb bud sections.

Probe pairs	Distal posterior	Distal anterior	Proximal posterior	Proximal anterior	Flank
	Frequency (%) ≥ 600 nm				
***Shh*-ZRS (847 kb)**	9.5	10	10.5	5.5	11
***Shh*-Dpp6 (831 kb)**	40 ($p < 0.0001$)	33.5 ($p = 0.0002$)	30.5 ($p = 0.0008$)	36.5 ($p < 0.0001$)	39 ($p < 0.0001$)

Statistical analysis of data for Fig. 4.5 B and C. *p*-values from Fisher's Exact Tests.

4.2.2. Increased co-localisation of the ZRS and *Shh* in the distal posterior forelimb of E11.5 embryos

Despite the compactness of the *Shh*-ZRS region throughout the forelimb bud and flank region of E11.5 embryos, differences in co-localisation frequency (defined as d ≤ 0.2 μm) of the gene and enhancer within the various limb regions and flank could be discerned. Co-localisation within the distal posterior limb bud (23.5%, averaged over measurements from three different sets of ~30 nuclei from different sections through the same embryo) was significantly greater than within the flank region (11%): $p = 0.03$ (Fig. 4.4 A, Table 4.3). In contrast, co-localisation within the distal anterior and proximal forelimb regions was not significantly greater than the flank (distal anterior 13.5%, proximal posterior 13%, and proximal anterior 13.5%).

Comparison of co-localisation frequency within the distal posterior forelimb bud and the other limb tissues reveals that, although increased in the former, the difference is not quite significant: $p = 0.09$ for each (Fig. 4.4A, Table 4.3). However, further work is required to complete this analysis due to a possible skewing of the distal posterior data set in a manner that has lowered the proportion of co-localised probe pairs. Of the three sets of data, two had co-localisation frequencies of 30.5% and 23.5%, while the data set average for the other section was 16.5%. This latter frequency is closer to the other limb regions and could be due this section containing cells outside the ZPA and closer to the centre of the distal limb bud. The *Shh* expression domain is smaller than that of *Hoxd13* which was analysed in the same way in Chapter 2. Therefore I will continue with this analysis on other posterior

sections taken from individual embryos to determine if, indeed, a change in co-localisation across the distal limb can be identified.

Table 4.3. Co-localisation frequency of the *Shh* probe with the ZRS and *Dpp6* probes for E11.5 limb bud sections

Limb region	***Shh*-ZRS (847 kb)**	***Shh*-Dpp6 (831 kb)**
	Co-localisation frequency (%)	
Distal posterior	23.5	3.5
Distal anterior	13.5 (p = 0.09)	4.5 (p = 1.00)
Proximal posterior	13 (p = 0.09)	0.5 (p = 0.62)
Proximal anterior	13.5 (p = 0.09)	4.5 (p = 1.00)
Flank	11 (p = 0.03)	0 (p = 0.25)

Statistical analysis of data for Fig. 4.4A. p-values from Fisher's Exact Tests.

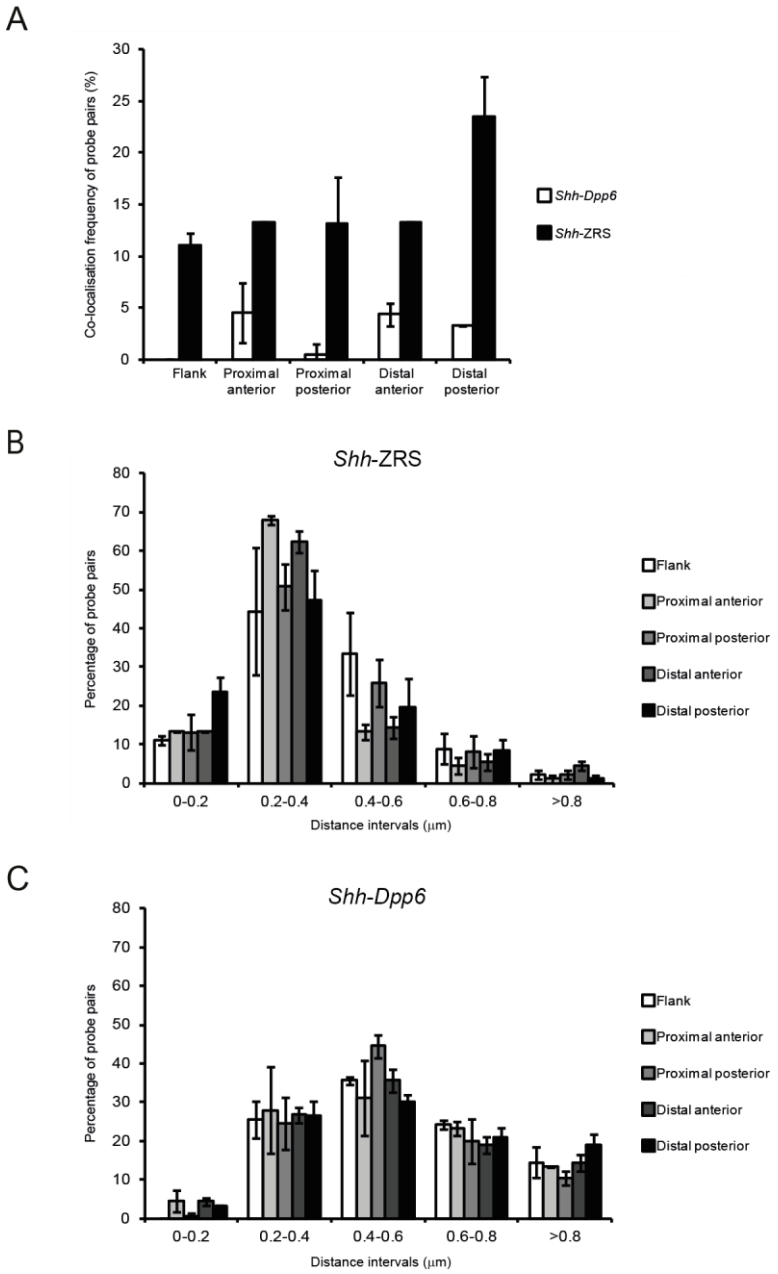

Figure 4.4. Frequency distributions of the interprobe distance (d) between *Shh* and the ZRS and *Dpp6* probes. A, Percentage of co-localised probe pairs (interprobe distance < 200 nm) for each of the two genomic regions assayed in the distal anterior, distal posterior, proximal anterior, proximal posterior forelimb bud and the flank. The statistical significance of the differences in co-localisation between the limb regions and the flank were examined by the Fishers exact test (Table 4.3). Frequences were calculated for every 0.2μm distance interval (*x*-axis) between **B,** *Shh*-ZRS, **C,** *Shh-Dpp6* in nuclei derived from proximal and distal E11.5 forelimb bud regions and the adjacent flank region. The location and plane of the tissue sections are shown in Figure 4.2. Number of loci measured ≥ 90. Error bars represent SEM obtained from two different tissue sections.

4.3. Histone modification characteristics of similar genomic regions containing active limb genes

4.3.1 H3K27me3 is largely absent from genomic regions encompassing *Shh*, *Grem1* and *Hand2*

The genomic territories that contain *Shh* and two other genes involved in anterior-posterior specification and development of the distal handplate – *Grem1* and *Hand2* (sections 1.3.1 and 1.3.2) – have many features in common. *Hand2*, like *Shh*, is expressed not only in the limb but also in other tissues during development (Srivastava et al., 1995; Fernandez-Teran et al., 2000; Ruest et al., 2003; Morikawa et al., 2005) and is also situated adjacent to a ~1 Mb gene desert containing many highly conserved loci, although they have yet to be characterised as possessing regulatory capabilities (Fig. 4.5, top right-hand panel). *Grem1* is not located near a gene desert but the genomic region of chromosome 2 where it is situated is gene-poor, with the large *Fmn1* gene spread over ~400 kb adjacent to the gene (Fig.4.5, top centre panel). A global control region (GCR) containing multiple enhancer elements that regulates *Grem1* and *Fmn1* limb expression (Vokes et al., 2008) is located across the 3′ end of the *Fmn1* gene (Fig. 4.5 bottom centre panel). Both *Hand2* and *Grem1*, then, inhabit a genomic environment similar to *Shh* and both HoxA and HoxD clusters: all are crucial developmental genes with tissue-specific regulatory elements.

I analysed H3K27me3 modification across the *Shh*, *Grem1* and *Hand2* loci by ChIP in dissected limb tissues and found that it is almost completely absent from the three genomic regions (Fig. 4.5 top panels, lower two tracks). Across the *Grem1* locus the only peaks are over an intergenic location 5′ of *Fmn1*, over a gene downstream of *Grem1*, and over the *Grem1* promoter (Fig. 4.5 top and bottom centre panels, lower two tracks). Both *Shh* and particularly *Hand2* are covered with the H3K27me3 modification but the chromatin over both adjacent gene deserts is virtually free of the mark (Fig. 4.5 top left-hand and right-hand panels, lower two tracks). Whereas an anterior-posterior difference in H3K27me3 levels can be discerned over *Hand2*, there is no difference over *Shh* (Fig. 4.5. bottom extreme left-hand and right-hand panels, lower two tracks). As for the 5′ *Hoxd* genes, *Shh* and

Hand2 have posteriorly-restricted expression domains at E10.5. The extent of *Hand2* expression across the distal posterior limb is greater than for *Shh*, which is the probable reason for the anterior-posterior difference for the former and not the latter. *Shh*-expressing cells will account for only a minority within the posterior tissue samples. Interestingly, H3K27me3 covers a large ~40 kb block at *Hand2* (Fig 4.5 bottom extreme right-hand panel), extending beyond the limits of the annotated CpG islands which have been suggested to be involved in polycomb recruitment.

4.3.2. H3K4me1 marks characterised and potential regulatory elements of the *Shh*, *Grem1* and *Hand2* genomic regions in distal posterior and anterior limb bud cells

Apart from the widespread coverage over a ~40 kb block downstream of *Hand2* that contains several highly conserved regions (Fig 4.5 bottom extreme right-hand panel), H3K4me1 has a punctate coverage over the three regions displayed in Fig. 4.5 (top three panels, upper two tracks); most of the peaks being situated in intergenic and intronic sites rather than proximal promoters. Over the *Shh*-ZRS region discrete peaks occur over the ZRS and to a lesser extent the floorplate enhancer SFPE1, but the gene itself is extensively covered (Fig. 4.5, two bottom left-hand panels, upper two tracks). The *Grem1* and *Hand2* genomic regions have greater coverage over the large *Fmn1* gene adjacent to the former and the gene desert adjacent to the latter, with many peaks corresponding to conserved sequences. For instance, intron 4 of *Fmn1*, the location of GABPα and ETV5 peaks, and a region in the middle of the *Hand2* gene desert with highly conserved sequences, that also contains a significant GABPα binding site in E11.5 limb tissue (Chapter 3, Fig. 3.9), also have blocks of H3K4me1 (Fig. 4.5 bottom panels 3 and 6, upper two tracks). Moreover, the highly conserved elements within the *Grem1* GCR are also covered with this mark, as is an intergenic region at the far end of the *Hand2* gene desert that intriguingly shows an anterior-posterior difference not only for H3K4me1 but also for H3K27me3 (Fig. 4.5 bottom panels 4 and 5). At none of the loci analysed is there a differential distribution of H3K4me1 between anterior and posterior limb tissue, despite the differential expression of these 3 genes in the limb region; supporting the

view that this modification cannot on its own be considered as a mark for active enhancers.

As discussed briefly in section 4.2.1 the lack of H3K27me3 over the *Shh*-ZRS region indicates that polycomb is not responsible for the compact state of this locus, but may have some role in the regulation of *Shh* as revealed by the array data displayed in Fig. 4.5. Likewise, *Hand2* expression is probably also influenced by polycomb activity, but from the array data these repressive complexes may have less of a role in *Grem1* regulation. Common to all three regions, though, is the general absence of this mark and the presence of the distal regulatory mark H3K4me1, often over non-coding conserved regulatory sequences. Further work to identify possible distal elements in the *Hand2* and *Grem1* regions and analysis of chromatin conformation will reveal whether regulation of these genes has a higher-order structure component similar to *Shh* and the 5′ *Hoxd* genes.

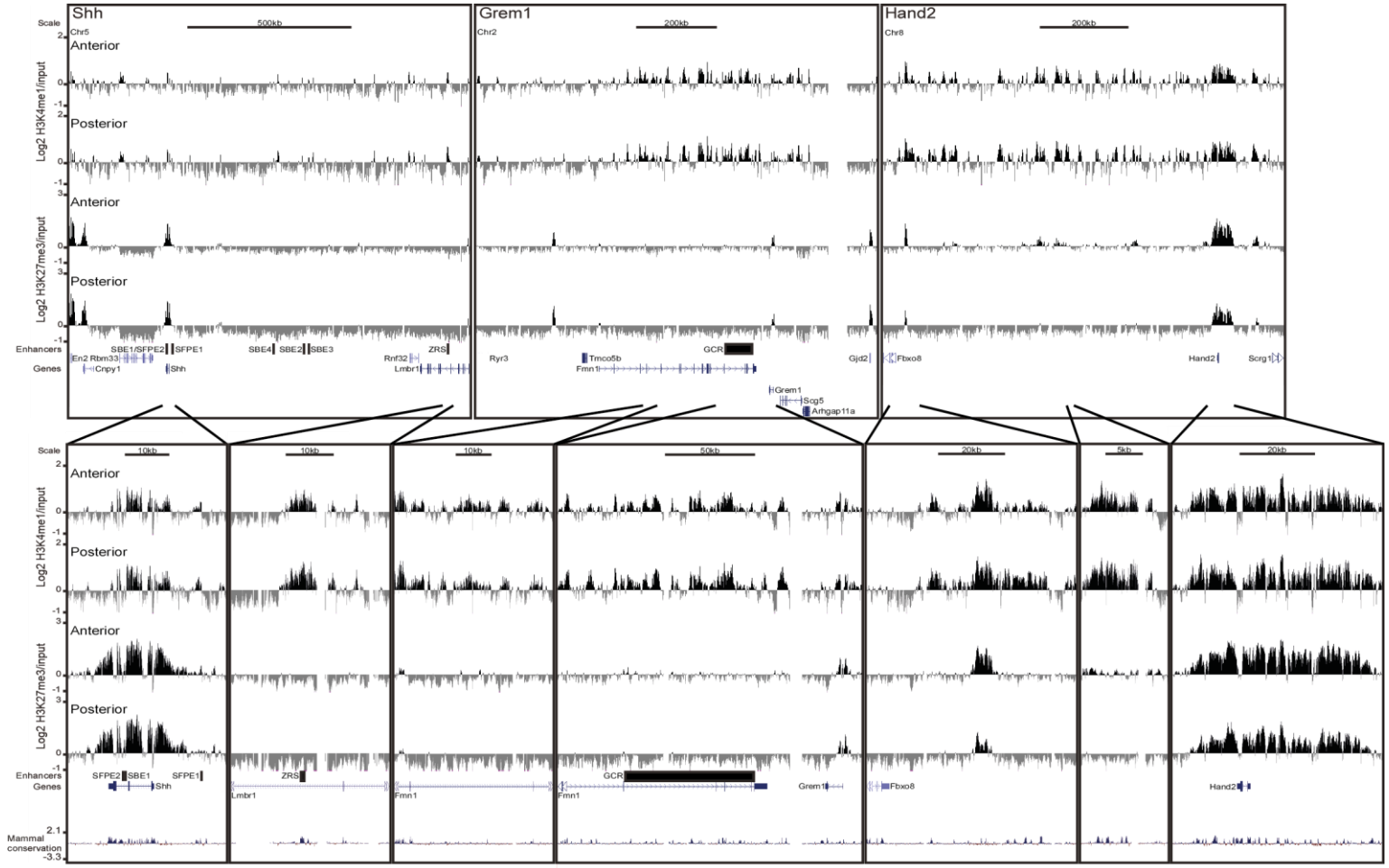

Figure 4.5. H3K4me1 and H3K27me3 distribution in E10.5 forelimb tissue. Mean log2 H3K4me1/input and H3K27me3/input at the *Shh*, *Grem1* and *Hand2* regions. n = 2 biological replicates for H3K4me1 and n = 4 replicates for H3K27me3 (2 biological, 2 technical).

4.4. Chromatin topology of the *Shh*-ZRS regulatory region in E11.5 floorplate cells

The objective of the preliminary experiments described in this section is to determine if the neural enhancers co-localise with the β-gal cassette inserted into the endogenous ZRS locus in cells where they drive *Shh* expression. DNA FISH on tissue sections derived from E11.5 embryos heterozygous for the reporter construct was carried out, initially with the fosmid probe that binds the wild type region of *Lmbr1* containing the ZRS. A 10 kb probe specific for the reporter construct (Fig. 4.1, purple region within the ZRS fosmid) has been optimised in order to do 4-colour FISH to compare co-localisation frequency of the SFPE1 and SBE4 enhancers with the reporter construct allele to co-localisation frequency with the ZRS wild type allele in floorplate and forebrain tissues. Work presented here involves the SFPE1 probe along with either the wild type ZRS probe or the *Dpp6* probe as in the limb analysis. These experiments initially focussed on determining the extent of chromatin compaction over the *Shh*-ZRS regulatory region in floorplate cells and adjacent cells where *Shh* is not expressed. The results are based on two sections through the same embryo, and although the results are intriguing, I aim to validate the data by carrying out further analysis on sections from other embryos. The sections shown here (Fig. 4.6B) have not been cut cleanly across the floorplate/notochord but rather at a more obtuse angle, making it more difficult to clearly identify *Shh*-expressing floorplate/notochord tissue (Fig. 4.6A). Therefore, I must again emphasise that results presented below can only be regarded as indicative of chromatin topology within neural cells; however the data is similar to what appears to be occurring in *Shh*-expressing and non-expressing tissues within – and adjacent to – the limb bud.

4.4.1 *Shh*-ZRS region is maintained in a compact chromatin state in floorplate and *Shh*-non-expressing cells

DNA FISH analysis in floorplate and adjacent cells suggests that the *Shh* regulatory region is in a highly compact conformation, similar to the limb bud at this stage of murine embryogenesis (Fig. 4.6). Interprobe distances (d^2) between SFPE1 and ZRS

revealed no significant difference in chromatin compaction between the *Shh*-expressing and adjacent non-expressing cells (p = 0.34) (Fig. 4.7A, Table 4.4). The region 3′ of *Shh* is in a decompact state in the cells adjacent to the floorplate similar to that in the limb and flank tissue (compare median interprobe distances in Tables 4.1 and 4.4); however this region is very decondensed in the floorplate, significantly greater than in the adjacent tissue (p = 0.0008) (Fig. 4.7A, Table 4.4). Difference in compaction is also highly significant when comparing SFPE1-ZRS and SFPE1-*Dpp6* in both sets of cells ($p < 0.0001$) (Fig. 4.7A).

Analysis of the frequency of hybridisation signals separated by defined distance (d) intervals also reveals the highly decompact chromatin state of the genomic territory 3′ of *Shh*. The proportion of signal pairs ≥ 600 nm apart was greater for SFPE1-*Dpp6* than SFPE1-ZRS: 50% and 32% compared to 1.5% and 4% in the floorplate and adjacent tissue respectively (Fig. 4.7B, Table 4.5). In both groups of cells the difference in proportion was highly significant: $p < 0.0001$. The significantly greater decompaction of the SFPE1-*Dpp6* region in floorplate compared to adjacent cells can be accounted for by the ~25% of the probe pairs in the former having an interprobe distance (d) greater than 800 nm (Fig. 4.7B).

Table 4.4. Interprobe distances for E11.5 floorplate sections

Tissue	SFPE1-ZRS (816 kb)	SFPE1-*Dpp6* (866 kb)
	Interprobe distance (d^2) (μm^2)	
Floorplate	0.06	0.37
Adjacent non-expressing	0.06 (p = 0.34)	0.24 (p = 0.0008)

Statistical analysis of data for Fig. 4.7A. Interprobe distances are median values, *p*-values from Mann-Whitney U Tests.

Table 4.5. Frequency of SFPE1-ZRS and SFPE1-*Dpp6* probes separated by ≥ 600nm for E11.5 floorplate sections.

Probe pairs	Floorplate	Adjacent non-expressing
	Frequency (%) ≥ 600 nm	
SFPE1-ZRS (816 kb)	1.5	4
SFPE1-*Dpp6* (866 kb)	50 ($p < 0.0001$)	32 ($p < 0.0001$)

Statistical analysis of data for Fig. 4.7 B. *p*-values from Fisher's Exact Tests.

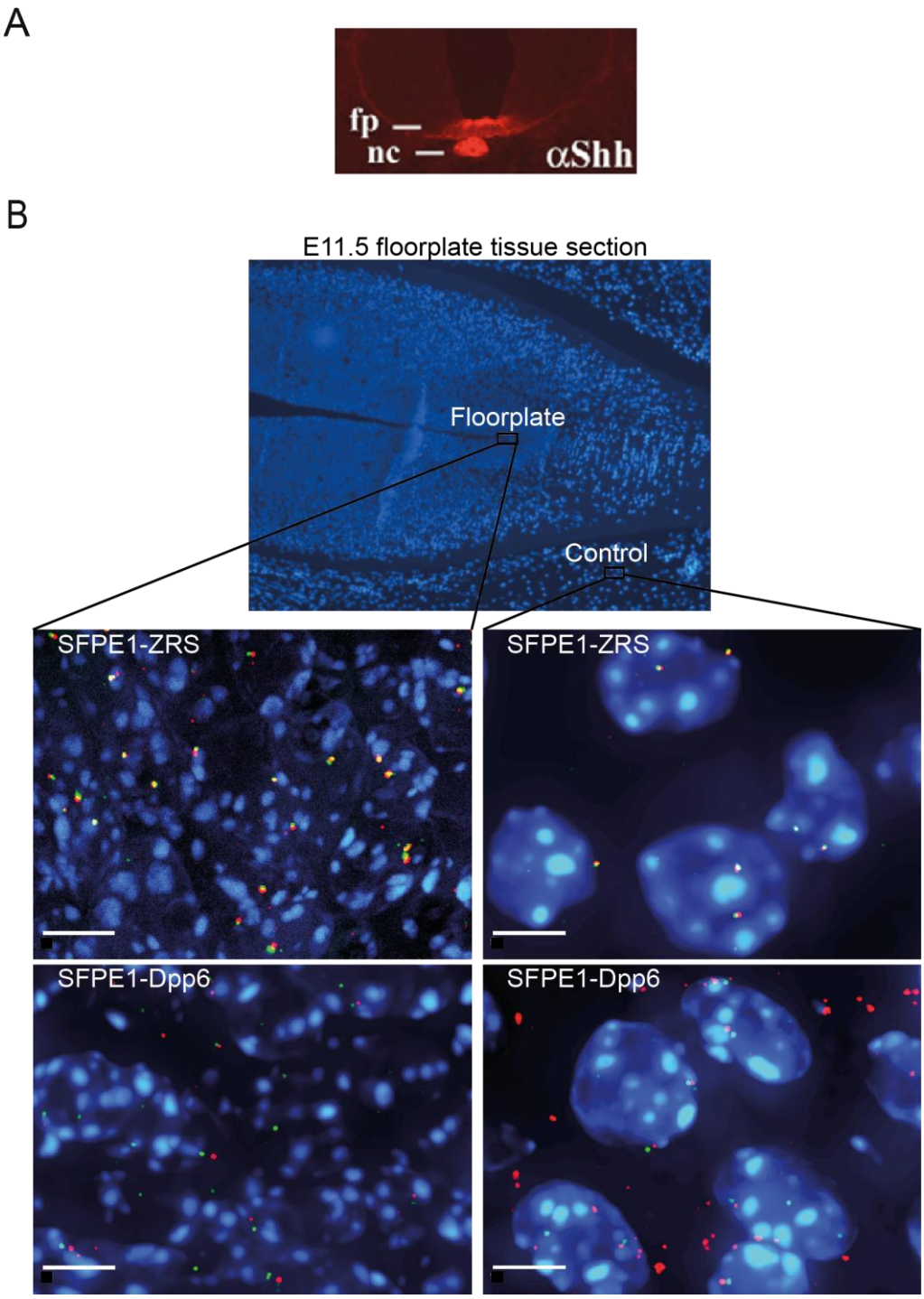

Figure 4.6. Representative section and nuclei images. A, *Shh* expression in the floorplate and neural tube (Image from Jeong et al., 2006) (fp, floorplate; nc, notochord). **B,** Image of floorplate tissue section for DNA FISH, counterstained with DAPI. Below are examples of nuclei from floorplate and the adjacent region from which interprobe distances were measured. Hybridisation signals are for SFPE1-*Dpp6* and SFPE1-ZRS probe pairs. Scale bars = 5 μm.

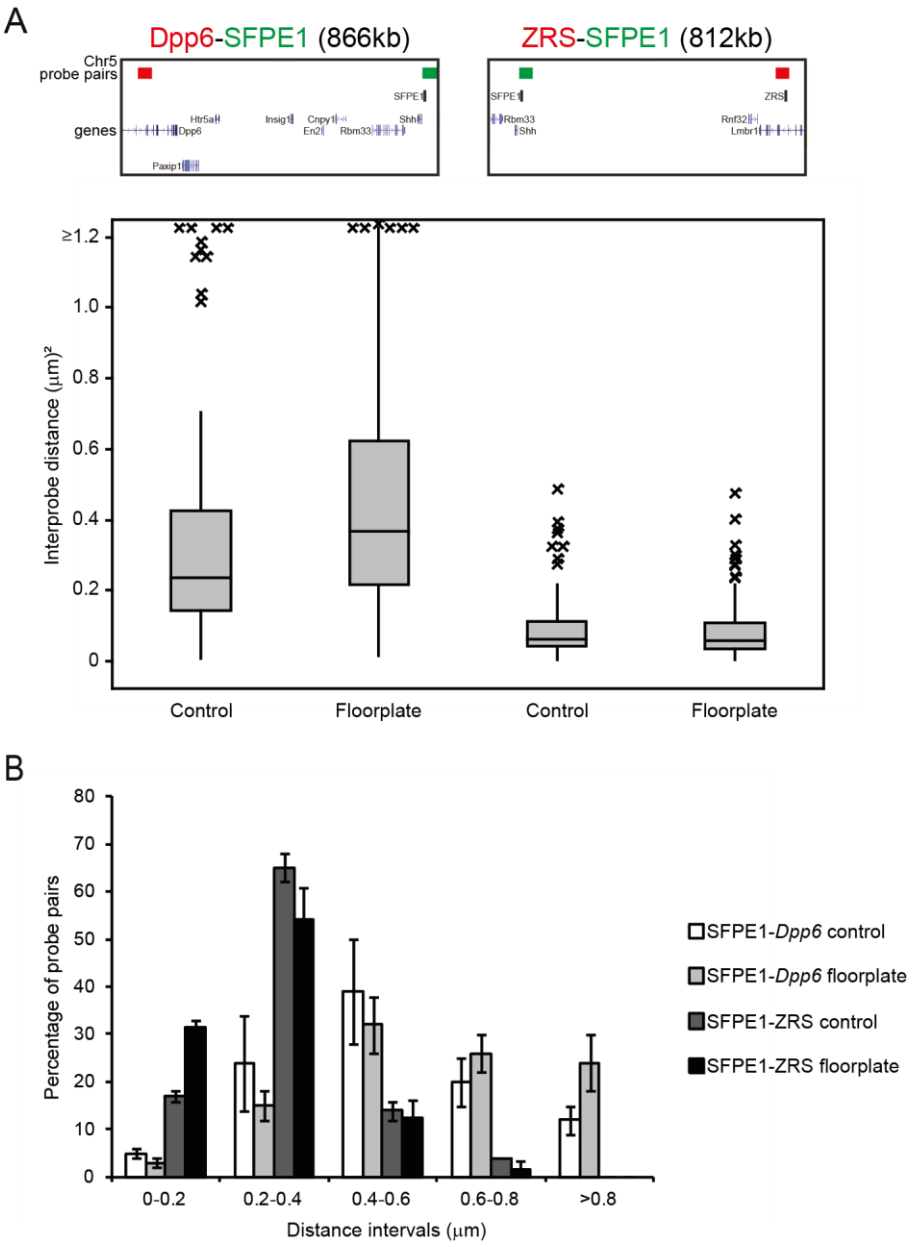

Figure 4.7. The *Shh*-ZRS regulatory region is maintained in a compact chromatin structure in possible *Shh*-expressing floorplate cells and adjacent non-expressing cells but with significantly greater co-localisation of SFPE1-ZRS in the floorplate only. A, Box plots show the distribution of interprobe distances (d^2) for the SFPE1-*Dpp6* and SFPE1-ZRS probe pairs in floorplate and adjacent tissue. Shaded boxes show the median and interquartile range of the data; asterisks indicate outliers. Number of loci: SFPE1-*Dpp6* = 100, SFPE1-ZRS ≥ 100. Probe positions shown above. The statistical significance of differences between each tissue was examined by Mann-Whitney U tests (Table 4.4). **B,** Frequences were calculated for every 0.2μm distance interval (*x*-axis) between SFPE1-ZRS and SFPE1-*Dpp6* in floorplate and adjacent *Shh*-non-expressing cell nuclei. The location of the tissue sections are shown in Figure 4.6. The statistical significance of the differences in co-localisation (< 200nm) of both probe pairs between Shh-expressing and non-expressing tissue were examined by the Fishers exact test (Table 4.6). Number of loci measured ≥ 100. Error bars represent SEM obtained from two different tissue sections.

4.4.2. Significantly greater co-localisation of SFPE1 and ZRS in floorplate tissue compared to the adjacent cells

Co-localisation of the SFPE1 floorplate enhancer probe and the probe covering the ZRS region (32% averaged over measurements from two different sections) within the *Shh*-expressing cells of the floorplate was significantly greater than within the adjacent non-expressing region (17%): $p = 0.02$ (Fig. 4.7B, Table 4.6). Frequency of co-localisation between SFPE1 and the *Dpp6* was minimal in both cell types and not significantly different ($p = 0.72$).

These data suggest that the ZRS and *Shh* loci co-localise in neural cells of the floorplate (the SFPE1 fosmid probe binds to the region immediately upstream of the *Shh* TSS, Fig. 4.1), although the ZRS has been shown to have limb-specific activity (Lettice et al., 2003; Sagai et al., 2005). Increased frequency of co-localisation may be due to the allele containing the reporter cassette. These tissue sections are from mice heterozygous for the construct, β-gal expression throughout the floorplate indicates that the floorplate enhancers (including SFPE1) are able to target and activate the reporter gene located at the endogenous ZRS locus. The compact higher-order chromatin conformation of the *Shh*-ZRS region in cells of the floorplate reduces the nuclear search space for these proximal enhancers of *Shh* (SFPE2 is located within an intron of the gene, Fig. 4.1) that may enable them to act over this large genomic distance. Further analysis of the floorplate tissue sections will involve 4-colour DNA FISH using the same probes in addition to the probe that binds to the reporter construct; in this way I will be able to discern whether the increased co-localisation frequency is due to interactions specifically between the β-gal endogenous knock-in and the floorplate enhancer.

Table 4.6. Co-localisation frequency of the SFPE1 probe with the ZRS and *Dpp6* probes for E11.5 floorplate sections

Tissue	SFPE1-ZRS (816 kb)	SFPE1-*Dpp6* (866 kb)
	Co-localisation frequency (%)	
Floorplate	32	3
Adjacent non-expressing	17 ($p = 0.02$)	5 ($p = 0.72$)

Statistical analysis of data for Fig. 4.7B. *p*-values from Fisher's Exact Tests.

4.5. *Shh* regulatory region is maintained in a compact chromatin structure in embryonic stem cells

My initial analysis of the *Shh* regulatory region in murine E11.5 embryos by FISH has revealed that the chromatin is arranged into a compact chromatin state not only in cells where the *Shh* is expressed (distal posterior limb bud, floorplate) but also in non-expressing cells (regions adjacent to the limb and floorplate). Further tissues await characterisation but in this section I will describe 2D FISH experiments that determined the chromatin conformation of the regulatory region in undifferentiated ES cells, where *Shh* expression is repressed. (2D FISH experiments on ES cells were carried out by S. Boyle.)

In these experiments the SFPE1 probe was used rather than the *Shh* probe but both recognise immediately adjacent regions, so for this analysis the SFPE1 probe was a legitimate proxy for the *Shh* genomic locus (Fig. 4.1). To account for differences in nuclear size, interprobe distance (d^2) was normalised to nuclear radius (d^2/r^2), although the non-normalised results were also comparable (Table 4.7). As for the E11.5 tissues analysed, the 800kb SFPE1-ZRS region was highly compact compared to similarly sized SFPE1-*Dpp6* interval ($p < 0.0001$) (Fig. 4.8, Table 4.7). However, the interprobe distance between SFPE1-ZRS (816 kb) is significantly shorter than that for SBE4-ZRS (515 kb) ($p < 0.0001$) (Fig. 4.8A, Table 4.7). So although the forebrain enhancer is ~300kb closer to the ZRS, their nuclear separation is greater than the distance from the *Shh* locus to the limb enhancer. Moreover, the interprobe distance between SFPE1 and SBE4 (286 kb) is also greater than the distance between SFPE1 and ZRS, though this difference is not quite significant ($p = 0.06$) (Fig. 4.8A, Table 4.7).

These data suggests that the compact nature of the *Shh* regulatory region is the result of a (small-scale) looping mechanism – possibly similar to the formation of a series of mini-loops described in section 1.2.5, Fig. 1.6 – that maintains the locus containing *Shh* and the locus where the ZRS resides in close proximity, at least in ES cells (Fig. 4.8B). And as the box plots in Fig. 4.8 show, the range of interprobe distances for each of the sub-regions between *Shh* and the ZRS, though greater than

for SFPE1-ZRS, did not reach the levels indicative of an extensive unravelling of chromatin structure characteristic of a large looping mechanism.

Table 4.7. Normalised and squared interprobe distances for OS25 ES cells

Probe pairs	Normalized interprobe distance (d^2/r^2)	Squared interprobe distance (d^2) (μm^2)
SFPE1-ZRS (816 kb)	0.0006	0.08
SFPE1-SBE4 (286 kb)	0.0009 (p = 0.06)	0.14 (p = 0.15)
SBE4-ZRS (515 kb)	0.0014 (p < 0.0001)	0.22 (p = 0.0001)
SFPE1-*Dpp6* (866 kb)	0.0097 (p < 0.0001)	1.75 (p < 0.0001)

Statistical analysis of data for Fig. 4.8. Interprobe distances are median values, *p*-values from Mann-Whitney U Tests.

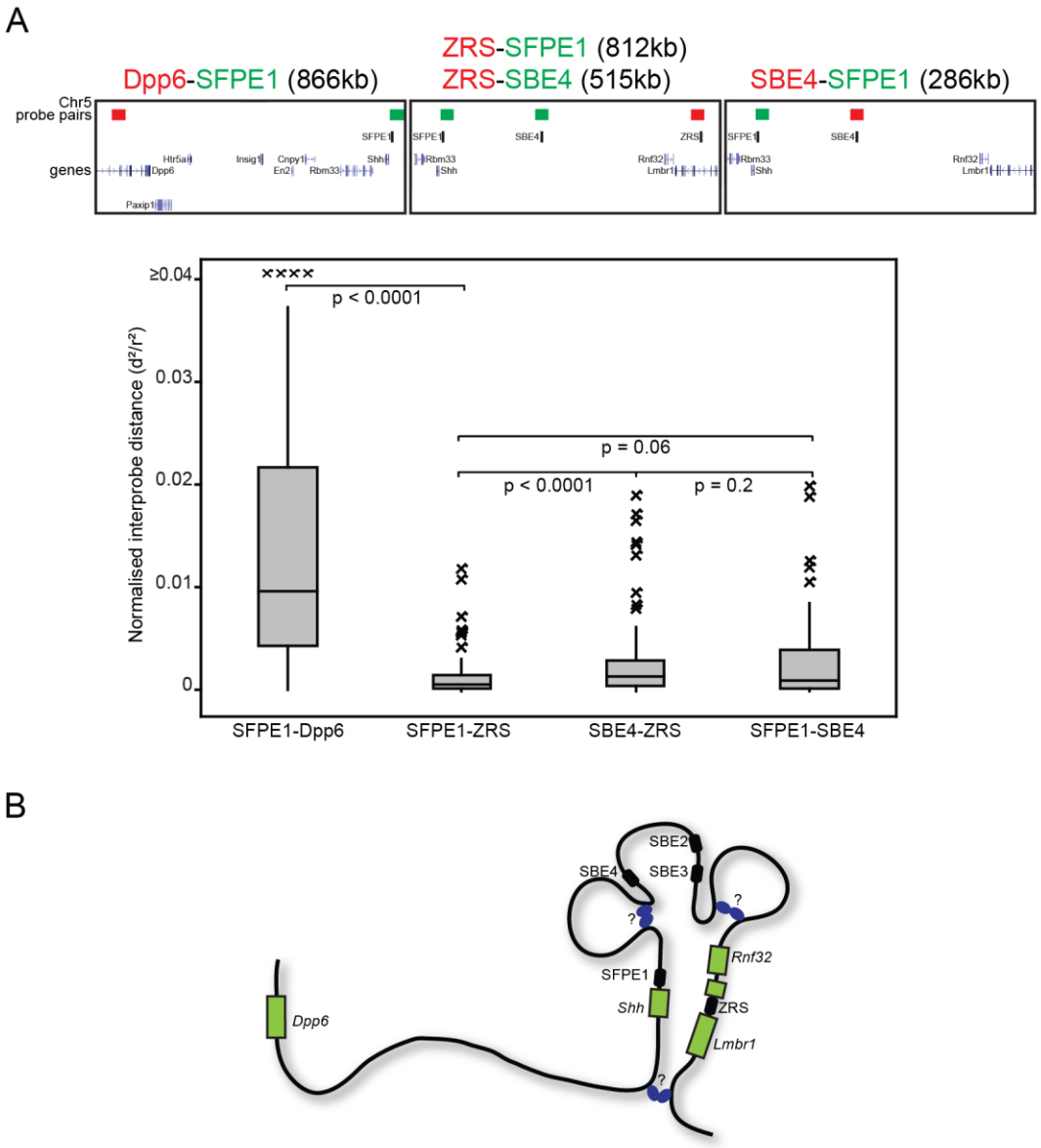

Figure 4.8. Chromatin structure of the *Shh* regulatory region is maintained in a compact conformation in ES cells. A, 2D FISH with probe pairs across the ~800 kb regions centromeric and telomeric of *Shh*, in MAA-fixed nuclei of OS25 ES cells. Probe positions for each experiment are shown above. Box plots show the distribution of interprobe distances (d^2) normalized for nuclear radius (r^2). The shaded boxes show the median and interquartile range of the data; asterisks indicate outliers. Number of loci: SFPE1-*Dpp6* = 82, SFPE1-ZRS = 93, SBE4-ZRS = 86, SFPE1-SBE4 = 86. The statistical significance of differences between each probe pair were examined by Mann-Whitney U tests (Table 4.7). **B.** Schematic of a possible higher-order chromatin configuration over the *Shh*-ZRS regulatory region that accounts for the interprobe distances measured between the various probe pairs in ES cells and embryonic tissue. Factors that could be maintaining this structure (?) have yet to be elucidated.

4.6 Discussion

4.6.1. Compact chromatin structure characterises the *Shh*-ZRS region in murine E11.5 limb and neural tissue and undifferentiated ES cells

From the preliminary data of this on-going project the most surprising finding is what appears to be a highly compact chromatin structure across the *Shh*/SFPE1-ZRS genomic region. These probe pairs have significantly shorter interprobe distances separating them than the *Shh*/SFPE1-*Dpp6* probe pairs that cover the adjacent genomic region; which is maintained in undifferentiated ES cells where *Shh* is not expressed (Fig. 4.8A) and in all the tissues so far examined in E11.5 embryos (Figs. 4.3 and 4.7B). In the E11.5 limb bud no significant difference in compaction levels of the *Shh*-ZRS regulatory region could be detected between the distal posterior where *Shh* is expressed and the distal anterior and proximal limb bud where the gene is silent. In the adjacent flank region, though, the ~800 kb region was significantly less compact but – as the box plots reveal – not to the extent of the unfolded chromatin structure throughout the 800 kb region 3′ of *Shh* (Fig. 4.3). Floorplate and adjacent cells showed no difference in compaction. To determine whether this chromatin conformation is maintained throughout all embryonic tissues at E11.5 requires further analysis of cells more remote from tissues where *Shh* is expressed. As for the HoxD locus, this compact conformation could be maintained to minimize the nuclear search space for the long-range enhancers to find the *Shh* promoter. So analysis of limb and neural tissue from embryos at a later stage in development, when the gene is no longer expressed, may reveal a more decompact chromatin conformation over the *Shh*-ZRS genomic region.

ChIP-chip data from E10.5 distal limb bud tissue revealed that the PRC2-catalysed H3K27me3 histone mark is absent from the *Shh* regulatory region, indicating that polycomb complexes do not have a role in the formation of this compact chromatin structure. In any case, the compact nature of the locus is more likely to signify an active or at least poised configuration; possibly similar to the conformation over the HoxD regulatory region described in Chapter 2 and in Montavon et al. (2011). My immediate focus in order to elucidate further the chromatin conformation of the locus is to measure the distances between *Shh* and the

forebrain enhancer SBE4 and between SBE4 and the ZRS in the limb and floorplate, as for the ES cell analysis. To determine whether the chromatin is forming a series of "mini loops" (Fig. 4.8B) I will use additional probes for loci between SBE4 and ZRS, including the *Rnf32* fosmid (Fig. 4.1), and by 3- and 4-colour FISH look for increased interprobe distances between probe pairs that indicate looping across particular regions. Further studies would involve trying to establish the probable biochemical factors required to maintain the higher-order chromatin structure in this compact conformation. Possible proteins are CTCF and cohesion, which has been identified as a facilitator in the formation of intrachromosomal loops (Hadjur et al., 2009; Mishiro et al., 2009; Nativio et al., 2009; Hou et al., 2010).

The FISH data presented for both the *Shh* and HoxD regulatory regions suggests that they form what has recently been described as "topological domains": up to megabase size genomic regions characterized by local chromatin interactions and separated by insulator elements bound by CTCF (Dixon et al., 2012). Indeed, both regions are identified as discrete topological domains in murine ES cells (Fig. 4.9). In another recent paper chromatin conformation capture carbon copy (5C) (Dostie et al., 2006) and super-resolution microscopy was used to demarcate "topologically associating domains" within a 4 Mb region of the X chromosome containing the X-inactivation centre (Nora et al., 2012). Therefore, I intend to interrogate further both regulatory landscapes using 5C in conjunction with FISH assays (possibly using high resolution microscopy for the *Shh* locus) in order to clearly define the limits of each region and confirm the gene-enhancer co-localisation results (see below). For HoxD I intend to use the immortalised cell lines (Chapter 2) for the 5C assays but these cells may not be a good system for *Shh* as it cannot be detected by PCR; however the Hill lab are attempting to activate the gene in the cell lines.

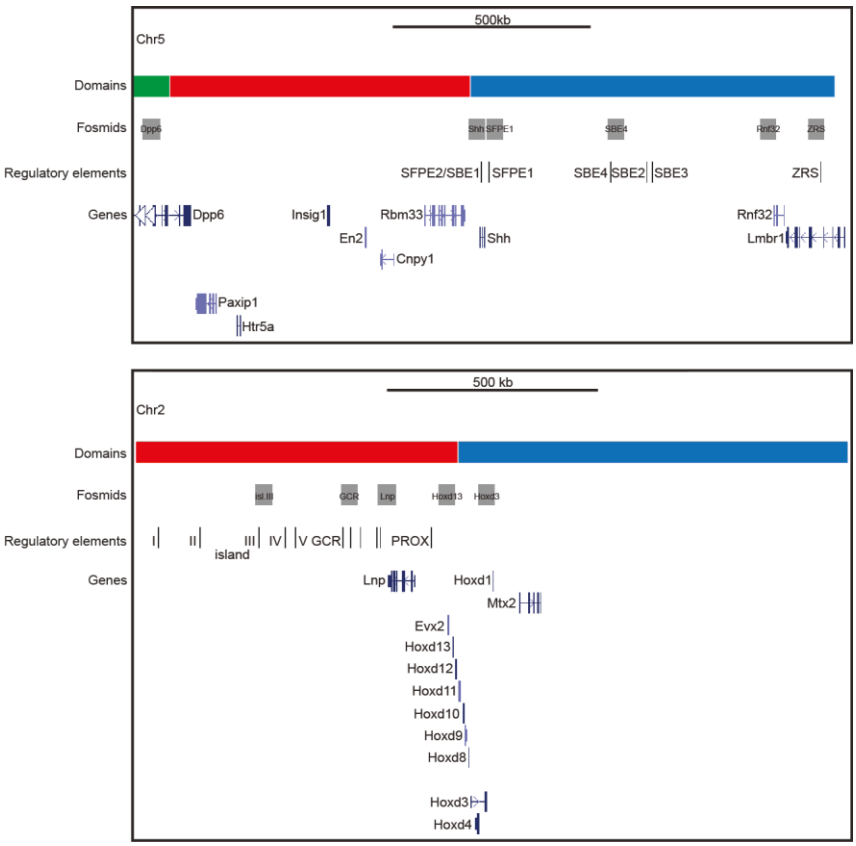

Figure 4.9. Topologically interacting domains in murine ES cells. The *Shh* and 5′ HoxD regulatory regions correlate with chromatin interaction domains identified in ES cells. Domains are shown as coloured bars along the top track. The second track shows fosmid positions used in FISH analysis of both genomic regions, with the position of regulatory elements below. Bottom track shows gene locations. (Topological domains from Dixon et al., 2012.)

4.6.2 Frequency of ZRS-*Shh* interactions was increased in the distal posterior limb and floorplate

In both *Shh*-expressing tissues analysed by FISH (limb bud and floorplate), the proportion of co-localised *Shh*/SFPE1-ZRS probes was greater than the adjacent tissue. Previous analysis of *Shh*-ZRS co-localisation in the limb, using the same criteria, found a co-localisation frequency of 18% of measured probe pairs in posterior cells compared to 10% in anterior cells dissected from E10.5 murine distal forelimbs (Amano et al., 2009). The difference in frequency was not significant between anterior and posterior cells, but co-localisation frequency in posterior cells was significantly greater than in cells dissected from the middle of the limb bud; a location where *Shh* is not normally expressed and where the authors present data showing over 40% of probe pairs separated by distances greater than 600 nm. This analysis was done on single cell suspensions, while the work I present here was done on dissected tissue sections; and in no tissues so far analysed have I found this level of probe separation. In the distal posterior and anterior tissue nuclei analysed here the frequency of co-localisation was 23.5% and 13.5% respectively, again not quite significantly different but this could be due to one of the posterior tissue sections measured (16.5% co-localisation) not containing ZPA cells. Further analysis of tissue sections across the entire forelimb will determine whether co-localisation of *Shh*-ZRS significantly increases within the ZPA. Of course, for a more accurate comparison with the earlier study, and to determine if there are temporal differences in chromatin topology over the *Shh*-ZRS regulatory region, requires repeating this analysis on E10.5 forelimb tissue. To ensure only ZPA cells are analysed, embryos containing a *Shh* allele containing an endogenous GFP knock-in could be used; immunofluorescence analysis of alternate sections would therefore allow a more precise identification of *Shh*-expressing cells.

Although expected in the distal posterior limb bud where the ZRS is active, co-localisation in the floorplate is not required to drive *Shh* expression there as two proximal enhancers have been shown to regulate expression throughout the floorplate/neural tube and hindbrain (Jeong et al., 2006). As explained in section 4.4 the tissues analysed were derived from mouse embryos heterozygous for a reporter

construct inserted into the endogenous ZRS locus (Fig. 4.1), so increased co-localisation may be indicative of the proximal enhancers having long-range capabilities and are actively driving reporter gene expression in the floorplate. Further analysis using a probe that is specific for the inserted reporter construct should be able to establish if these enhancers do have this capability and may provide insight into the intriguing question as to whether the capacity for long-range activity is inherent to all enhancers – that it is the presence of the right binding factors and/or the receptiveness of promoters within "topologically associating domains" that determine the extent of enhancer activity. However, co-localisation could also be caused by another, currently uncharacterised floorplate enhancer located near to the ZRS which is interacting with *Shh*. Or, indeed, both scenarios could occur. To determine if this proposed neural enhancer does co-localise with *Shh* in the floorplate requires FISH analysis on wild type embryos, ideally in conjunction with the 5C assays described above. However, collecting sufficient quantities of chromatin from floorplate tissue, and limb tissue, to carry out these assays would be technically challenging.

4.6.3 Histone H3 lysine 4 mono-methylation marks known and potential enhancers of genes expressed during distal limb development

The genomic regions displayed in Fig. 4.5 contain three genes – *Shh*, *Grem1* and *Hand2* – that play crucial roles in posterior specification of the distal limb bud and consequent development of the handplate (Robert and Lallemand, 2006; Hill, 2007). All three reside in chromosome regions that can be termed gene-poor (especially *Shh* and *Hand2*) and at least for *Shh* and *Grem1*, are regulated by long-range enhancers. The histone modification associated with distal regulatory elements, H3K4me1, covers discrete, often conserved, distal loci throughout each of these genomic territories; including the known limb enhancers ZRS and *Grem1* GCR. However, these enhancers have the H3K4me1 mark in anterior cells also, which indicates that this modification is also present over poised enhancers during limb development (as I showed for the *Hoxd* and *Hoxa* enhancers in Chapter 2). Additional ChIP-chip analysis of other active enhancer markers, particularly H3K27ac and p300, may reveal differences in enhancer status between the anterior and posterior limb bud.

To establish if the conserved regions with the H3K4me1 peaks, that also had GABPα peaks in the E11.5 ChIP-chip assays described in Chapter 3, are possible enhancers (at the centre of the gene desert centromeric of *Hand2* and within intron 4 of *Fmn1* (Fig. 4.5)) I intend to interrogate both regions by DNA FISH. Furthermore, I would like to establish whether *Grem1* and *Hand2* also reside in regulatory regions as suggested by their surrounding genomic landscape; in the first instance by FISH to determine chromatin conformation and then at greater resolution by the 5C technology I will use to define the HoxD and (hopefully) *Shh* regulatory landscapes.

Chapter 5

Discussion

A wealth of data has provided clear evidence that dynamic manipulation of higher-order chromatin structure impacts upon gene expression (Dostie and Bickmore, 2012). The capacity of distal enhancer elements to activate target genes in a tissue-specific manner requires modification to, and very likely altered conformation of, chromatin structure (Amano et al., 2009; Montavon et al., 2011; Cotney et al., 2012). However, specificity may be more reliant upon the binding of factors necessary to activate an enhancer; for large regulatory sequences such as the ZRS competing active and repressive factors are likely to bind in varying ratios that depend on their cellular levels or the levels of co-factors, which will determine subsequent activity (Lettice et al., 2012).

Many key developmental genes – which require a high degree of regulation to ensure correct spatio-temporal expression – inhabit gene-poor genomic territories; often adjacent to gene deserts. This thesis describes the work I have undertaken to delineate the roles of chromatin structure and conformation in the activation of *Shh* and the 5′ Hoxd genes by tissue-specific enhancers, particularly during distal limb development (Chapters 2 and 4). Both have neighbouring 500-800 kb gene deserts containing multiple enhancers (the limb-specific ZRS is of course located at the opposite end of the *Shh* gene desert within *Lmbr1*) that ensure spatial and temporal expression is robustly regulated (Spitz et al., 2003; Lettice et al., 2003; Jeong et al, 2006; Montavon et al., 2011). In addition, Chapter 3 described research into the opposing functions of two different groups of ETS factors that bind to the ZRS in the limb, thereby beginning to elucidate the biochemical processes that determine the activity of this limb-specific enhancer.

5.1. Polycomb has a role in the regulation of limb-specific Hoxd expression

In Chapter 2 I have presented the results from a thorough investigation into local chromatin conformation and histone modification over the HoxD cluster and how chromatin structure impacts upon gene expression. Work on ES cells had shown that polycomb has a direct role in chromatin compaction over the Hox clusters, and

that loss of H3K27me3 correlates with chromatin unfolding over HoxB and HoxD and progressive 3′ to 5′ expression of the genes (Eskeland et al., 2010). Chromatin compaction levels were then shown to be dependent on the polycomb PRC1 complex which recognises and binds chromatin with the H3K27me3 histone modification. I have established that similar polycomb driven processes occur during limb-specific Hoxd expression. In both E10.5 distal posterior cell lines and limb tissue there was a loss/reduction of H3K27me3 over the 5′ genes compared to anterior cell lines and limb tissue, chromatin decompaction was restricted to the distal posterior cells/tissue, and an upregulation of the 5′ genes in the distal posterior cell lines and dissected limb tissue.

During this research I have made use of immortalised cell lines set up by A. Hill and extensive expression analysis has shown their veracity as a model system for limb development (Chapter 2, section 2.2). Over the *Hoxd* genes, the general pattern of expression is reflected by the cell lines, and FISH analysis established that the chromatin is more decompact over the HoxD locus in both posterior cell lines compared to anterior cell lines – therefore recapitulating the different compaction levels identified in their tissue of origin (Chapter 2, section 2.4) (Williamson et al, in press).

5.2. Two groups of ETS factors influence limb-specific *Shh* expression through the ZRS

As functioning enhancers are nucleoprotein complexes, biochemical processes occurring during enhancer activity need to be delineated for a full understanding of enhancer function. In Chapter 3 I described research into the role of several members of the ETS family of transcription factors in regulating the spatial activity of the *Shh* limb enhancer the ZRS. By ChIP-chip I established that GABPα, ETS1, ETV4 and ETV5 bind to the ZRS, and in vitro EMSA analysis revealed that whereas ETS1 binds to the canonical AGGAA$^{G}/_{A}$T motif (Hollenhorst et al., 2011), ETV4 binds the novel AGAAA$^{G}/_{A}$ motif (Fig. 3.2A). ETV4 and ETV5 have functional redundancy during limb development and repress anterior *Shh* expression

(Mao et al., 2009; Zhang et al., 2009). Reporter assays show that deletion of both ETV ZRS sites results in ectopic expression, indicating that the repressive activity of these factors is on the ZRS and not *Shh* directly (Lettice et al., 2012).

The five AGGAA$^{G}/_{A}$T ZRS sites have varying potency for driving reporter gene expression, with sites 1 and 3 having the greatest input (Chapter 3, section 3.6.1). The AUS mutation that creates an additional ETS site immediately adjacent to the ETV site B (Fig. 3.3D) can, however, compensate for both these sites alone; and with the other sites present the posterior expression domain increases towards the centre of the distal limb bud and an anterior expression domain is produced. This work implicates GABPα and ETS1 as at least two of the ETS factors required to activate the ZRS in the distal posterior limb bud, and proposes a possible mechanism for ectopic anterior expression of *Shh* through point mutations in the ZRS creating additional binding sites for these factors; mutations that are the cause of the ZRS-associated syndrome PPD2 in two families (Gurnett et al., 2007; E. Da Graaff, personal communication).

5.3. Both 5′ HoxD and *Shh* regulatory regions are maintained in a compact higher-order chromatin conformation in expressing and non-expressing tissues

Analysis of the ~500 kb region 5′ of the HoxD cluster by FISH has revealed that the entire region is folded in a close chromatin conformation throughout the forelimb bud of E11.0 embryos (Chapter 2). Likewise, the ~800 kb *Shh*-ZRS region is being held in what looks to be a similar conformation in both limb and neural *Shh*-expressing and non-expressing tissue and, surprisingly, ES cells (Chapter 4). Nevertheless, I have been able to quantify increased enhancer-gene co-localisations within expressing tissues compared with non-expressing. I have shown the GCR and 5′ Hoxd genes spatially co-localise within the distal posterior E11.0 limb bud – thereby revealing at the single cell level the requirement for direct interaction between this enhancer and its target genes for upregulation of expression to occur (Chapter 2) (Williamson et al., in press). Another potential regulatory element

identified farther into the adjacent 5′ gene desert (island III (Montavon et al., 2011)) does not show a significant increase in co-localisation with either the GCR or *Hoxd13* in the distal posterior forelimb, which suggests that this element is not required for 5′ *Hoxd* activation at this developmental stage. Therefore, the multiple potential regulatory elements identified by Montavon et al. (2011) in E12.5 limb buds may not all be required simultaneously for second phase *Hoxd* limb expression; but rather a more spatio-temporal mechanism is at work. One possibility is that GCR/Prox initiates expression earliest and specifically in distal posterior regions, and then subsequently the other elements drive expression in progressively more anterior domains throughout the later temporal stages (>E11.0 - E12.5) of second phase Hoxd limb expression.

To gain a more complete understanding of the regulatory mechanisms controlling limb-specific *Hoxd* expression I intend to continue my investigations into the processes involved. Initially, I will use 5C, probably the chromatin conformation capture method most able to accurately quantify physical interactions of genes and regulatory elements and characterise local chromatin architecture (up to 1-2 Mb) (Dostie and Bickmore, 2012). Then, by combining 5C and FISH, I will interrogate the region further at progressively later developmental stages to build a more complete picture of chromatin structure over the 5′ HoxD regulatory region and how it is manipulated through this crucial phase of limb development when tight regulation of the 5′ genes is necessary for the correct patterning of the digits (Gonzales et al., 2007; Montavon et al., 2008; Montavon et al., 2011). My proposed model for the configuration of the regulatory region (Fig. 2.15) invokes the chromatin as a series of mini-loops, a configuration likely to require maintenance by biochemical mechanisms. Therefore, I will look for the binding of factors capable of forming chromatin loops, such as CTCF and cohesin (Hadjur et al., 2009; Mishiro et al., 2009; Nativio et al., 2009; Hou et al., 2010), by ChIP-chip. A functional role for chromatin looping in gene expression was recently shown at the β-globin locus (Deng et al., 2012). Forced interactions between the β-globin promoter and the LCR in non-expressing erythroid cells where the locus is normally in a relaxed state resulted in activation of transcription.

Several histone modifications and modifying enzymes are associated with enhancers, such as H3K4me1/2, H3K27ac and the histone acetyltransferase p300 (Heintzman et al., 2007; Barski et al., 2007; Maston et al., 2006; Rada Iglesias et al., 2010; Creyghton et al., 2010). ChIP-chip H3K4me1 data from E10.5 anterior and posterior limb tissue revealed that this modification was present across the distal limb bud at known and potential *cis*-regulatory elements of posterior-specific genes (Chapters 2 and 4). Therefore, H3K4me1 is not a good indicator for active enhancers in the limb. Acetylation of H3K27, however, has been shown to be a better mark for identifying active limb enhancers (Cotney et al., 2012). I intend to characterise the H3K27ac modification over these regions and the HoxD-associated island elements (using tiling arrays with wider coverage over the genomic regions containing limb genes) in E10.5 anterior and posterior tissue by ChIP-chip in order to determine whether this mark differentiates between active and inactive/poised enhancers. In the context of the 5′ HoxD regulatory region, these ChIP-chip assays may reveal differences in H3K27ac modification between GCR/PROX and the island elements at the E10.5-11.0 developmental stage. Consequently, I may then repeat this analysis on E11.5 and E12.5 anterior and posterior tissue in conjunction with FISH, 5C and ChIP-chip for CTCF/cohesin.

The *Shh*-ZRS FISH data from limb and neural tissue and ES cells may be indicative of higher-order chromatin structure over developmental gene regulatory regions; a conformation maintained from before the embryonic stages when the gene is active at least until the period of expression has concluded. It is rather striking that the close chromatin conformation identified in *Shh*-expressing cells (floorplate and distal posterior limb bud) is also the state of the locus in E11.5 non-expressing tissue so far analysed and even in undifferentiated ES cells where *Shh* is never expressed (Chapter 4). These data partially contradict previous analysis of this locus, which indicated that *Shh* and the ZRS genomic regions were not in close proximity in limb tissue where the gene is never expressed (Amano et al., 2009) – a conformation I have yet to identify in limb, neural and adjacent tissues. Although more widespread analysis of E11.5 tissues is required (and of earlier stages), this unchanging higher-order structure may be a general conformation of genomic regions containing developmental genes regulated by tissue-specific enhancers. Maintenance of a close

chromatin structure – possibly (as for *Shh*) from ES cells through differentiation and subsequent embryo development – ensures that regulatory elements and target gene(s) are in close proximity, reducing the nuclear search-space for the enhancers to optimise activation efficiency. Timing of activation and tissue-specificity would then rely upon the correct set of factors and co-factors being present at the enhancers and promoters. To investigate further this model of long-range regulation I will carry out the experiments described in section 4.6 of Chapter 4.

5.4. Summary

In this thesis I have presented novel data on the mechanisms of long-range *cis* regulation as imparted by distal enhancers upon target genes during mouse forelimb bud development. I have shown that two different aspects of chromatin higher-order structural manipulation occurs during 5′ Hoxd activation: chromatin decondensation across the cluster itself that correlates with a reduction of H3K27me3 and indeed the PRC1 component Ring1B (which also occurs during ES cell differentiation (Eskeland et al., 2010) and main body anterior-posterior development (Morey et al., 2007; Soshnikova and Duboule., 2009)); and co-localisation of the limb-specific GCR and the 5′ end of the HoxD cluster (Williamson et al., in press (Appendix B)). I identified several ETS factors bound to the ZRS in vitro and in vivo and was part of a lab effort to elucidate the role of these factors in regulating *Shh* limb expression through the ZRS. Finally, I have established that compact chromatin configurations occur over both the 5′ Hoxd and *Shh* regulatory regions; a conformation that may be shared by other regulatory genes occupying similar genomic landscapes.

Chapter 6

Methods & Materials

6.1 General molecular biology reagents

Solutions prepared by MRC Human Genetics Unit (HGU) Core Scientific Services:

Tris-HCl

Tris[hydroxymethyl]aminomethane (Tris base) was dissolved in sterile water and pH adjusted using HCl.

EDTA

Ethyldiaminetetra-acetic acid di-sodium salt was dissolved in sterile water, solid NaOH added to increase pH to 8.

TAE buffer, 50X stock

Tris base 242g, glacial acetic acid 57.1 ml, 0.5M EDTA to a final volume of 1 litre with distilled water.

PBS

Phosphate buffered saline: NaCl 8g, KCl 0.2g, Na_2HPO_4 1.44g, KH_2PO_4 0.24g – in a final volume of 1 litre of distilled water and adjusted to pH 7.4.

20X Saline Sodium Citrate (SSC)

NaCl 175.3g; Sodium citrate 88.2g – in a final volume of 1 litre of distilled water and adjusted to pH 7.0.

L-broth

Tryptone 10 g
Yeast extract 5 g
NaCl 10 g
Glucose 1 g

Made up to 1 litre with water and autoclaved.

Agar

15 g agar was added to 1 litre L-broth to prepare solid media.

Solutions prepared in the lab:

TE buffer

10 mM Tris.HCl (pH 8.0), 1mM EDTA

1X TAE buffer

50X TAE 20ml in a final volume of 1 litre with distilled water

6.2 Tissue culture

6.2.1 Immortomouse cell lines from anterior and posterior distal limb

The posterior third, and the anterior two thirds, of distal forelimb buds were dissected from E10.5 and E11.5 day embryos from an Immortomouse (H-2k-tsA58) x CD1 cross. The transgenic mice were created by microinjection of a DNA fragment containing the a thermolabile large tumour (T) antigen (TAg) gene from a simian virus 40 (SV40) strain (tsA58, temperature sensitive (ts)) fused to the widely active H-2k^{b} promoter (Jat et al. 1991). After washing in PBS, limb buds were treated using trypsin (0.2 g/l)/Versene for 15-20 minutes, and then gently dispersed by repeated pipetting with a pastette. Cells were then plated into 6 well plates in DMEM (Invitrogen) with 10% Foetal Calf Serum (Sigma), Penicillin/Streptomycin and 20 ng/ml γ-Interferon (Peprotech). The two pairs of E10.5 cell lines (A1/P1 and A2/P2) were derived from two litters, the E11.5 cell line (14fp) from a single litter

Cells were grown at 33°C – the permissive temperature for the temperature-sensitive T antigen. Media was changed twice per week and cells were passaged as necessary. E10.5 cells were harvested no later than passage 12 due to a notable change in growing speed which may be indicative of a progressive loss of their immortalised status and further differentiation of sub-populations of cells.

6.3 Electrophoresis

6.3.1 Gel electrophoresis of DNA samples

Separation of DNA fragments by size was done by agarose gel electrophoresis. Agarose (High Pure, BioGene) was dissolved in 1X TAE in a microwave to make 1-2% (weight/volume) gels. Ethidium bromide (Electran, 10 mg/ml) was added to the gel before setting to a final concentration of 25 μg/100 ml.

Loading buffer (see 6.3.3) was added to the DNA samples before loading and gels were run at 60-100V, visualised under UV illumination and photographed on a Gel Doc XR (BioRad) imaging system.

6.3.2 Gel electrophoresis of RNA samples

As for DNA samples but with the following additional steps: added an equal volume of formamide to the sample volume containing 1μg of RNA along with the loading buffer. Incubate samples for 10 minutes at 65°C, then 5minutes 4°C before running on a 1% gel.

6.3.3. Agarose gel loading buffer

Prepared as a 5X stock and stored at room temperature.

	Quantity	Final concentration
Glycerol	5 ml	50%
0.5M EDTA	100 μl	5 mM
Orange G	0.03 g	0.3%

To final volume of 10 ml in MilliQ water.

6.4 Analysing gene expression

6.4.1 RNA extraction

Total RNA was prepared using Trizol reagent (Invitrogen) according to manufacturer's protocol (for limb buds, dissected anterior and posterior tissue was dissociated into single cell suspensions in Trizol using a syringe fitted with a 25G (0.5 mm) needle (BD Microlance)), followed by acid phenol:chloroform:isopropyl alcohol extraction and then digested with 2U DNaseI (Ambion) for 30 minutes at 37°C.

6.4.2 cDNA synthesis

cDNA was made using a first strand synthesis kit (Roche) according to manufacturer's instructions. A negative control reaction, lacking reverse transcriptase, was also performed for each RNA sample.

6.4.3 RT (Reverse Transcription)-PCR

First strand product was amplified by PCR using gene specific primers (Table 6.1) for 32 cycles (95°C for 20 s, 60°C for 30 s, 72°C for 30 s) and products visualised by electrophoresis.

6.4.4 Quantitative real time (RT-q)PCR

RT-qPCR was carried out on the BioRad CFX96 Real-Time System using Quantitect SYBR Green detection kit (QIAGEN) and using the thermal cycler programme: 15 min Hotstart; 44 PCR cycles (95°C for 15 s, 55°C for 30 s, 72°C for 30 s). For each run, a standard curve for each primer set (Table 6.1) was obtained using anterior and posterior cell lines and limb tissue cDNAs. Fold expression in 3 biological replicates of posterior/anterior cells (cell lines and limb tissue) was calculated using the BioRad

CFX Manager software and normalized to GAPDH. Standard error of the mean (SEM) for fold change expression of P1, P2 and posterior limb tissue cells relative to A1, A2 and anterior limb tissue cells respectively and comparison of anterior and posterior cell lines' expression levels with their tissue of origin was calculated using the formula:

$$\left(\frac{\Delta Z}{Z}\right)^2 = \left(\frac{\Delta A}{A}\right)^2 + \left(\frac{\Delta B}{B}\right)^2$$

Where **Z** is fold change, and A and B are two measured transcript levels with SEM ΔA and ΔB.

Table 6.1. PCR primers for analysis of transcript levels

Gene	Oligo	Sequence	Ref
Hoxd1	d1bUp	CCACAGCACTTTCGAGTGGA	Morey et al., 2007
Hoxd1	Hoxd1r	ACTCTTTCTCTAGCTCTGTCAG	
Hoxd3	d3pA	GAACTCCAAGCAGAAGAACAG	Condie & Capecchi, 1993
Hoxd3	Hoxd3r	CAGATAGCGGTTGAAGTGGAAC	
Hoxd4	Hoxd4f	GGAGAACGAGGGAGAACCA	
Hoxd4	Hoxd4r	TCTGCTGCTGCTATGACTGC	
Hoxd8	Hoxd8f	CACTTAAATCAGAGCTCGTCTCC	
Hoxd8	Hoxd8r	GACCTCGATTCTCCTCTTCCTGG	
Hoxd10	Hoxd10f	GTGCAGGAGAAGGAAAGCAAAG	Modified from (Mizusawa et al.,
Hoxd10	Hoxd10r	TAACGCTCTTACTGATCTCTAGGC	
Hoxd11	Hoxd11f	TCCAGGCAAACGAGAGAAAC	
Hoxd11	Hoxd11r	TTGGCAAATAAGGTTTCTGGA	
Hoxd12	Hoxd12f	CAACTTGAACATGGCAGTGCAAG	
Hoxd12	Hoxd12r	CTGCTGCTTTGTGTAGGGTTTCC	
Hoxd13	Hoxd13f	AGTCCTGGACGCTAGCCAACG	
Hoxd13	Hoxd13r	GTAGACGCACATGTCCGGCTG	
Actin	Actinf	AGAGCTATGAGCTGCCTGACG	
Actin	Actinr	TGTGTTGGCATAGAGGTCTTTACG	
Lnp	LnpaUp	GTGGAAGGCTCAAGCTCAAC	
Lnp	LnpbLo	TGCTGGGCAATCTGAATATG	
GAPDH	GABDHf	ATCACCATCTTCCAGGAGCGAG	
GAPDH	GABDHr	GACCCTTTTGGCTCCACCCTTC	Modified from (Boyer et al.,
FgfR2b	FGFR2f	CTGTTCAATGTGACGGAGATGG	
FgfR2b	FGFR2r	ACAGACGCGTTGTTAYCCTCAC	
FgfR2c	FGF2Cf	CACTCTGCATGGTTGACAGTTC	
FgfR2c	FGF2Cr	CACTCTGCATGGTTGACAGTTC	
Fgf8	FGF8f	AAAGTCACACAGCGACATGTGAGG	
Fgf8	FGF8r	TCTGTGAATACGCAGTCCTTGCCT	
Fgf10	FGF10f	AGCGGGACCAAGAATGAAGACTGT	
Fgf10	FGF10r	CCTGCCATTGTGCTGCCAGTTAAA	
Etv4	Pea3f	ACCATGGAGAGCAGTGCCTTTACT	
Etv4	Pea3r	ATGCACATCCAGGGACATCTGAGT	

6.4.5 Gene expression microarrays

For expression arrays, 400 ng of total RNA from A2 and P2 cells (3 biological replicates for each) were amplified using the Illumina TotalPrep RNA Amplification Kit according to manufacturers' instructions (Ambion). Amplified, biotinylated cRNA (1.5 μg) were labelled and hybridized to Illumina MouseRef6 Gene Expression beadchip arrays and run on an Illumina Beadstation (Wellcome Trust Clinical Research Facility).

Probe level expression data were exported from the Illumina beadstudio software and analysed in R using the limma (Smyth et al. 2005) and beadarray (Dunning et al. 2007) bioconductor packages (analysis by G. Grimes). Probes that were not detected on the microarrays (detection p-values <0.01) were removed from the analysis. Probe signals were quantiled normalized to remove technical variation. Probes were assessed for differential expression using limma's lmFit, eBayes and topTable function. P-values were corrected for multiple testing using the Benjamini & Hochberg method.

To interpret functional differences in anterior-posterior expression, the 5944 significantly differential expressed (DE) genes (adjusted p-value <0.05) were ranked according to p-value and searched for enriched gene ontology (GO) terms that appear at the top of the ranked list, using the **G**ene **O**ntology en**ri**chment ana**l**ysis and visua**l**iz**a**tion tool (GOrilla) (Eden et al. 2009). Of the 13,427 total genes detected on the microarray, 10,581 genes could be mapped to GO terms in GOrilla. To avoid false positives we set a stringent *P* value of 10^{-7} as the cut off for GO terms to be considered based on the Bonferroni correction.

6.5. Protein analysis

6.5.1 Acid extraction of histones

To isolate nuclei, 3-6 x 10^6 A1, P1 and A2, P2 cells were washed in ice cold PBS and resuspended in 1 ml PBS containing 0.2% Triton X-100 and 0.2 mM PMSF. After centrifugation (12000 *g*, 20 minutes, 4°C), histones were extracted as described in (Bonaldi et al. 2004) and boiled in 100 µl SDS loading dye. For equal sample loading 1, 2.5 and 5 µl quantities from each cell line sample were run on a 17.5% SDS-PAGE gel. Histones were stored at -20°C.

6.5.2. Polyacrylamide Gel Electrophoresis (PAGE)

17.5% polyacrylamide gels were cast in NuPAGE cassettes. Histone samples and Benchmark pre-stained protein marker (Invitrogen) were loaded, and gels were run in Tris-glycine buffer (25 mM Tris, 250 mM glycine, 0.1% SDS). Gels were stained with Coomassie Blue or subjected to Western Blotting.

17.5% polyacrylamide gel

Resolving gel	
PAA 30%	3.5 ml
Tris 1.5 M pH8.8	1.5 ml
H_2O	1 ml
SDS 20%	30 µl
APS 10%	60 µl
Temed	3 µl

Stacking gel 5% (stock)	
PAA 30%	8.5 ml
Tris 2 M pH6.8	3.125 ml
H_2O	38.125 ml

Stacking gel 5%	
Stock	4 ml
APS 10%	40 µl
Temed	4 µl

6.5.3. Western Blotting

Proteins on polyacrylamide gels were transferred to Hybond-P membranes (Amersham Biosciences) in buffer containing 25 mM Tris, 190 mM glycine and 15% methanol for 1.5 hours. Membranes were blocked in 3% BSA in TBS-T (0.1%

Tween-20) overnight at 4°C, then probed for 1 hour with antibodies that detect: α-H3K27me3 (1:1000, Millipore, 07-449) and α-H3 (1:50000, Abcam, ab1791). Blots were washed 3 x 10 minutes with TBS-T and incubated with a secondary antibody-horseradish peroxidise (HRP) conjugate (1:10000) in TBS-T for 30 minutes, and then a further 3 x 10 minute washes. Blots were incubated with Amersham enhanced chemiluminescence (ECL) reagent kit for 2 minutes and exposed to autoradiography film (Amersham).

6.6. Preparation of nuclear extracts for EMSAs

The cell extraction protocol was adapted from Schreiber, *et al* (1989). The components of Buffers A and B are below. Both buffers were kept on ice throughout the procedure. A minimum of four T150 flasks of a cell line growing confluently were washed twice in warm (37°C) PBS. Following the aspiration of the second wash, ice-cold PBS was added to each flask twice and the cells were scraped into a 50ml centrifuge tube. The samples were spun down for 10 minutes at 600*g* and the PBS was removed. This centrifugation step and all subsequent steps were done at 4°C. 6ml of ice-cold PBS was added to each centrifuge tube and the pellet was re-suspended by pipetting up and down.

The cell suspension in each 50ml centrifuge tube was divided into aliquots of 1ml and transferred to 1.5ml microcentrifuge tubes and spun down for 1.5 minutes at 600 x g then resuspended in an equivalent volume of ice-cold Buffer A (vortex for 10 seconds). Following incubation for 15 minutes on ice, 10% IGEPAL (or NP-40), for a final concentration of 0.2%, was added to each suspension and vortexed for 10 seconds (immediately after the addition). The cell suspensions were incubated for a further 2 minutes on ice then spun down for 2 minutes at 13800*g*. The supernatant was aspirated and a further 0.5ml of Buffer A added to each pellet and vortexed. The suspensions were spun down 2 minutes at 13800*g* and the supernatant removed.

The volumes of the nuclear pellets were estimated by eye and 3 volumes of Buffer B was added to each pellet and vortexed. The samples were incubated on ice for 30-60 minutes with intermittent, vigorous, vortexing. All the samples were pooled and spun down for 10 minutes at 13800 *g*. The supernatant was collected and separated into aliquots and stored at -70°C. Protein concentration of each extraction was 1-2 mg/ml, as measured by nanodrop.

Protein extraction buffers

Buffer A	Buffer B
10 mM HEPES (pH 7.9)	20 mM HEPES (pH 7.9)
10 mM KCl	400 mM NaCl
0.1 mM EDTA	1 mM EDTA
0.1 mM EGTA	1 mM EGTA
dH_2O up to 50 ml	dH_2O up to 50 ml

100µl of 1M DTT and 1 protease cocktail tablet (Roche) dissolved in 1ml of dH2O added to each buffer immediately before use**.**

6.7 Electrophoretic mobility shift assays

6.7.1 Oligonucleotides

Table 6.2 lists the oligonucleotides used for EMSA analysis. The upper and lower strands were annealed by heating to ~95°C with an equimolar mixture of both single strands and slow cooling to room temperature. Upper strands were biotin-labelled before annealing (section 6.7.2), with commercially labelled wild type, Australian and Family AC upper strands also used (5′ end biotin label on the commercially labelled strands, 3′ end biotin labelling undertaken in the lab). All labelled double strand oligos were kept in 10fmol/µl stocks for the EMSA assays and the unlabelled double strand oligos for EMSA competition assays were kept in 50, 100 and 500fmol/µl stocks (x5, 10 and 50).

Table 6.2. EMSA oligos

Oligonucleotide	Sequence
Wild type up	TATTACAGAAAATGAAGTCATATC
Wild type lo	GATATGACTTCATTTTCTGTAATA
AC up	TATTACAGGAAATGAAGTCATATC
AC lo	GATATGACTTCATTTCCTGTAATA
Aus up	TATTACAGAAAAGGAAGTCATATC
Aus lo	GATATGACTTCCTTTTCTGTAATA
Site1 up	CTGAGGTCACTTCCTCTCTTAATT
Site1 lo	AATTAAGAGAGGAAGTGACCACAG
Site3 up	GAGCGATTCAGGAAGTGCTGCTTA
Site3 lo	TAAGCAGCACTTCCTGAATCGCTC

All oligos recognize conserved regions of the ZRS. Wild type oligos bind to the 3′ region where both the AC and Aus mutations occur, Sites1 & 3 contain the highly bound ETS motifs.

6.7.2 Biotin 3′ end labelling of oligonucleotides

Oligonucleotides were biotinylated by 3′ tailing with terminal deoxynucleotidyl transferase (TdT) and 1-3 biotinylated ribonucleotides (Biotin-11-UTP) using the Pierce labelling kit (Pierce Biotechnology, Inc., protocol available at www.piercenet.com, product number: 89818). The procedure is optimised for labelling 5pmol of 3′-OH ends per reaction. Tailed oligos were extracted using chloroform:isoamyl alcohol and labelling efficiency was estimated using the Dot Blot by hand spotting procedure described in the protocol. The labelled oligo strand was then annealed to its complimentary strand as described in the previous section.

6.7.3 Electrophoretic mobility shift assays

Band shift assays were carried out as described in the Pierce LightShift^{R} Chemiluminescent EMSA Kit protocol (Pierce Biotechnology, Inc., available at www.piercenet.com, product number 20148).

6.7.3.1 Preparation and pre-run polyacrylimide gel

Polyacrylimide gels (6%) containing 30% (w/v) acrylimide, 0.5X TBE, 25% (w/v) ammonium persulphate, TEMED and ultrapure water were prepared. The gels were pre-run with loading buffer in 0.5X TBE for ~60 minutes.

6.7.3.2 Binding reactions

The modified procedure for the binding reaction was as follows:

1. Prepared 15 µl binding reactions, including the labelled target oligo and the nuclear extract from a limb mesenchyme cell line. To prevent non-specific DNA-protein binding I incubated the reaction mixture containing the protein extract and the nonspecific Poly dI·dC competitor DNA for 5 minutes before adding the labelled target oligo and the unlabelled oligo if it was a competition assay. Binding reactions were incubated for 20 minutes at room temperature.

2. Added 3µl of 5X loading buffer to each reaction and mixed by pipetting.

Electrophoresis of the binding reactions, electrophoretic transfer of binding reactions to a positively charged nylon membrane (Roche) and cross-linking of oligos to the membrane were all carried out as described in the protocol.

Likewise, detection of the biotin-labelled oligo/protein complexes by chemiluminescence was done following the Pierce EMSA kit protocol. The membrane was exposed to Amersham ECL Hyperfilm (Buckinghamshire, UK), with exposure times ranging from 10 seconds to 8 minutes.

<u>Binding reaction components</u>
Binding buffer 1X
Poly dI•dC 66.7 ng/µl
Protein extract ~4-8 µg
Unlabelled oligo 1000 fmol (for competition assays)
Biotin end-labelled target oligo 20 fmol
Ultrapure water up to 15 µl

6.8. Chromatin Immunoprecipitation (ChIP)

6.8.1 Native ChIP

6.8.1.1 Native ChIP on cell lines

Nuclei were prepared and resuspended in NB-R as previously described (Gilbert et al., 2003). Nuclei corresponding to 2 x 10^7 – 1 x 10^8 cells were digested with 40-46 Boehringer units of MNase (Worthington) for 10 minutes at room temperature in the presence of 20μg RNase A to obtain a chromatin ladder enriched in tri-, tetra-, and pentanucleosomes. The reaction was stopped by adding an equal volume of Stop Buffer and incubated on ice overnight. Between 50-150 μg released chromatin were mixed with 10 μg prebound (to Protein A or G magnetic beads, Invitrogen) H3K27me3 antibody (Abcam, ab6002) or mouse IgG (Santa Cruz, sc-2025) in the presence of 100 μg BSA and incubated for 3 hours at 4°C. Beads were then washed 3x with Wash Buffer and once in TE. Bound complexes were eluted with 0.1 M $NaHCO_{3,}$ 1% SDS at room temperature. Immunoprecipitated and input DNA were purified with Proteinase K (Genaxxon) and QIAGEN PCR purification kit.

6.8.1.2 Native ChIP on limb tissue

For limb tissue, distal anterior and posterior forelimb buds were dissected from 60 – 70 E10.5 embryos and, due to the lower number of cells (~1 x 10^6), native chromatin immunoprecipitation was performed with the following modifications: cells were digested with 8 - 9 Boehringer units of MNase (Worthington) for 10 minutes at room temperature in the presence of 20μg RNase A to obtain a chromatin ladder enriched in tri-, tetra-, and pentanucleosomes. Between 10 - 30 μg released chromatin were mixed with 3 - 5 μg prebound (to Protein A or G magnetic beads, Invitrogen) H3K27me3 antibody (Millipore, 07-449) or H3K4me1 antibody (Abcam, ab8895) in the presence of 25 μg BSA and incubated for 3 hours at 4°C. Immunoprecipitated and input DNA were purified with Proteinase K (Genaxxon) and QIAGEN MinElute PCR purification kit.

6.8.1.3 Native ChIP buffers

Solutions prepared in the lab and stored at 4°C.

NBA	NBB	NBR
85 mM NaCl	85 mM NaCl	85 mM NaCl
5.5% Sucrose	5.5% Sucrose	5.5% Sucrose
10 mM TrisHCl pH 7.5	10 mM TrisHCl pH 7.5	10mM TrisHCl pH7.5
0.2 mM EDTA	0.2 mM EDTA	3 mM $MgCl_2$
0.2 mM PMSF	0.2 mM PMSF	1.5 mM $CaCl_2$
1 mM DTT	1 mM DTT	0.2 mM PMSF
1X Protease Inhibitors	0.1% NP40	1 mM DTT
	1X Protease Inhibitors	

Stop Buffer	ChiP W1	ChiP W2
215 mM NaCl	150 mM NaCl	200 mM NaCl
10 mM TrisHCl pH 8	10 mM TrisHCl pH 8	10 mM TrisHCl pH 8
20 mM EDTA	2 mM EDTA	2 mM EDTA
5.5 % Sucrose	1% NP40	1 % Triton X 100
2 % TritonX 100	1 % Nadeoxycholate	0.2 mM PMSF
0.2 mM PMSF	0.2 mM PMSF	1 mM DTT
1 mM DTT	1 mM DTT	2X Protease Inhibitors
2X Protease Inhibitors	2X Protease Inhibitors	

Elution Buffer	EX50	Sonication Buffer
0.1 M $NaHCO_3$	10mM HEPES pH7.6	2M NaCl
1 % SDS	50mM NaCl	5M Urea
	1.5mM $MgCl_2$	
	0.5mM EGTA	
	10% glycerol	
	1mM DTT	
	0.2mM PMSF	

Micrococcal Nuclease Buffer: 50% Glycerol, 10mM Tris pH7.6 and 50mM NaCl.

6.8.2 Crosslinked ChIP on limb tissue

6.8.2.1 Chromatin Immunoprecipitation for Ring1B and ETS factors

Anterior and posterior E10.5 limb tissue cells (~5 x 10^6 - ~1.5 x 10^7 posterior cells, ~1 - 3 x 10^7 anterior cells) dissected from the distal forelimb buds of 70 – 75 E10.5

embryos were fixed with 1% formaldehyde (25°C, 10 minutes) and stopped with 0.125M Glycine. Crosslinked ChIP for Ring 1B was performed as described (Stock et al., 2007). Nuclei were sonicated using a Diagenode Bioruptor (Leige, full power 30s on, 30s off, in an ice bath for 50 minutes) to produce fragments of <500bp. Between 30-50 μg released chromatin were incubated with 1μg of prebound (Protein A magnetic beads, Invitrogen) Ring1B antibody (MBL, D139-3) or mouse IgG (Santa Cruz, sc-2025) and washed and eluted as described (Stock et al., 2007). Reverse crosslinked DNA was purified as for nChIP.

For ChIP of ETS factors, cells from dissected E11.5 forelimbs were fixed and sonicated as above, except that the garments size after sonication was < 300bp. An arbitrary concentration of 350 μg chromatin was incubated with 5 μg prebound (to Protein A or G magnetic beads, Invitrogen) IgG (Santa Cruz, sc-2025) or ETS1 (Hagman), ETV4 (Abcam, ab70425), ETV5 (Abcam, ab102010), GABPα (Santa Cruz, sc-22810) or ELF1 (santa Cruz, sc-631) in the presence of 50 μg of BSA and washed and eluted as described in (Stock et al., 2007). Reverse crosslinked DNA was purified with Proteinase K (Glenaxxon) and QIAGEN PCR purification kit.

6.8.2.3. Crosslink ChIP buffers

Solutions prepared in the lab and stored at 4°C.

1% formaldehyde was prepared in PBS.

Glycine stop buffer: 0.125M glycine in PBS.

Swelling buffer	Sonication buffer	Wash sonication buffer
25 mM HEPES pH7.9	50 mM HEPES pH7.9	50 mM HEPES pH7.9
1.5 mM $MgCl_2$	140 mM NaCl	140 mM NaCl
10 mM KCl	1 mM EDTA	1 mM EDTA
0.1% NP40	1% Triton X100	1% Triton X100
0.2 mM PMSF	0.25% Na-deoxycholate	0.1% Na-deoxycholate
1 mM DTT	0.1% SDS	0.1% SDS
Protease inhibitor cocktail	0.2 mM PMSF	
	1 mM DTT	
	Protease inhibitor cocktail	

Elution buffer	Block solution	Urea buffer
0.1 M $NaHCO_3$	1 x PBS	2 M NaCl
1% SDS	0.5% BSA ($^{w}/_{v}$)	5 M Urea

Wash buffer A	Wash buffer B
50 mM HEPES pH7.9	20 mM Tris pH8.0
500 mM NaCl	1 mM EDTA
1 mM EDTA	250 mM LiCl
1% Triton X100	0.5% NP-40
0.1% Na-deoxycholate	0.5% Na-deoxycholate
0.1% SDS	

6.8.3 ChIP analysis

6.8.3.1. ChIP-qPCR

For relative quantification by real-time PCR, dilutions of MNase-digested (native) chromatin from E10.5 anterior and posterior limb cell lines and dissected limb tissue (input) were used to create standard curves for the immunoprecipitated H3K27me3 and H3K4me1 chromatin (3 biological replicates). For each replicate, PCRs were performed in triplicate using Quantitect SYBR Green detection kit with a Peltier PTC-200 thermal cycler using an inbuilt Cromo4 continuous fluorescence detector connected to Opticom 3.1 software interface (using primers described in Table 6.3). Thermal cycler programme: 15 min Hotstart; 44 PCR cycles (95°C for 15 s, 55°C for 30 s, 72°C for 30 s).

For ETS enrichment at the ZRS, qPCR was carried out using equal concentrations of input, IgG and Chip DNA. Enrichment values for ETS1, ETV5 and ELF1 from E11.5 whole limb sonicated chromatin (crosslinked) are presented as fold differences relative to IgG and normalized to input with the formula $2^{[(CtIgG-CtInput) - (CtAb-CtInput)]}$ where Ct values are threshold cycles. PCRs were performed in duplicate or triplicate on 3 biological replicates. ZRS primers listed in Table 6.3.

Table 6.3. Real-time PCR primers for ChIP analysis

Promoter/exon	Oligo name	Sequence	Ref
Hoxd1	Hoxd1promf	GAGTAACTTGACCTTCTCAGAG	
	Hoxd1promr	ATTGCGGGAGAAAGGCAGGGAAG	
Hoxd10	Hoxd10prof	TAGTAGATGTCGCTGTTGTCCG	
	Hoxd10pror	ACATGACAACCAAGCCAATGAGA	
Olig2	Olig2f	GCCTGACGCTACAGTGACAA	(Boyer et al., 2006)
	Olig2r	GGCTAATTCCGCTCAATGAA	(Boyer et al., 2006)
Actin	Actinf	CCTCGATGCTGACCCTCATCC	
	Actinr	GACACTGCCCCATTCAATGTCTC	
ZRS	ZRS3f	AGAAGAGAGTAGGAAGTCCAGCCT	
	ZRS3r	GAGCCTTCCTCTTGCCTGTGATTT	
Nanog	Nanogex2for	GAAGACCTGCCTCTTCAAGGC	
	Nanogex2rev	AGAACACAGTCCGCATCTTCT	

6.8.3.2. ChIP-chip

Ten nanograms (optimal) of input or ChIP DNA were amplified using the WGA2 whole genome amplification kit (Sigma). Amplified material was labelled with Cy3 or Cy5 by random priming according to the NimbleGen ChIP-chip protocol (Roche). In total, 2 or 3 biological replicates with dye swaps were hybridized for 20 h and washed according to manufacturer´s protocol. A custom 3x720K mouse tiling array (NimbleGen, Roche) containing 179,493 unique probes from different genomic regions, with each probe represented by 4 replicates was used (Appendix 1 lists the genomic regions covered and the coordinates based on the Mouse NCBI37 (mm9) build). Arrays were scanned on a NimbleGen MS 200 Microarray scanner (Roche) using 100% laser power and 2 µm resolution. Raw signal intensities were quantified from TIFF images using MS 200 Data Collection software.

Data processing was done by G. Grimes. Microarray data were analyzed in R/Bioconductor using the Epigenome (PROT43) protocol with the following parameters: the mean signal intensity of the 4 replicate probes on each array was taken. Loess normalization was used within arrays to correct for the dye bias, and scale normalization was used within replicates, to control inter-array variability. Log enrichment was calculated by subtracting the mean log2 input intensities from the mean log2 enriched intensities. Significant enrichment was tested using the significance analysis of microarrays (SAM) technique (Tusher et al., 2001), and the

local false discovery rate based on the SAM statistic was calculated using the Locfdr R package. A false discovery rate of 0.05 was used as the significance cut-off.

To determine the significance of the difference in H3K27me3 and Ring1B enrichment over HoxD in anterior vs posterior limb tissue cells the median log enrichment value of probes covering the locus or specific locus sub-regions (5′, 3′; promoters, genes, intergenic regions) were calculated. The statistical significance of any anterior-posterior differences was assessed using the non-parametric Mann-Whitney U-test. Mean log enrichment values for H3K27me3 were used to compare individual genes and promoter regions within HoxD and adjacent genes and statistical significance was assessed using the 2-sample Students T-test.

The R/Bioconductor *Ringo* package (Toedling et al., 2007) and a previously published algorithm (Schwartz et al., 2006) were used to determine thresholds to quantify significantly enriched ETS factor peaks. Taking into account that reporter levels must exceed a certain threshold (y_0), G. Grimes chose a mixture modelling approach as positive and negative control sets were not available. So, the distribution of reporter levels, y, were based on a mixture of two underlying distributions: the null distribution of probes in non-enriched regions and the alternative distribution in enriched regions. The resultant histogram based upon the mixture of these distributions allows the assumptions that the null distribution is symmetrically distributed around its mode, m_0, close to $y = 0$, and the alternative distribution spread more widely and skewed to the right (greater than) of m_0. Based on these assumptions, the null distribution is estimated as follows: m_0 can be identified by the midpoint of the shorth of reporter levels that fall into the -1 - 1 interval on a log2 scale. The null distribution is then estimated from the empirical distribution of m_0-$|y-m_0|$, that is, by reflecting $y < m_0$ on to $y > m_0$. From this estimated null distribution the enrichment thresholds y_0 were determined, using the 99.9% quantile (Toedling and Huber, 2008). To account for variation in antibody efficiency and noise levels due to differences in sample DNA quality, enrichment thresholds were calculated separately for each antibody. Peaks were scored as containing at least three probes with enrichment above these thresholds within a block of continuous enrichment up to 500bp in size.

Conserved sites were based on the UCSC 30-way Multiz Alignment & Conservation track which identifies conserved sequences shared by 19 placental mammals. Aligned sequences from the remaining 11 non-placental mammal vertebrates (down to teleost fish) included in the analysis can boost or decrease conservation scores.

6.9. Fluorescent in situ hybridization

6.9.1 Fosmid probe preparation

6.9.1.1 Fosmid stocks

Fosmids were streaked (Whitehead (Sanger), Table 6.4) on to agar plates containing chloramphenicol (12.5 ng/μl) and incubated overnight at 37°C. I picked individual colonies and incubated overnight in 3 ml L-broth at 37°C with shaking. I then added 170 μl of glycerol to 830 μl L-broth aliquots and stored at -70°C.

Table 6.4. Fosmid Probes

	Region	Whitehead (Sanger) Name	Ensemble name	Coordinates		Size (bp)
				Start	End	
Hoxd	*GCR*	WI1-2157A11	G135P63331H7	74242615	74282044	39429
	Lnp	WI1-482L15	G135P61870C5	74329582	74372986	43404
	Hoxd13	WI1-469P2	G135P67444A12	74474157	74513003	38846
	Hoxd3	WI1-121N10	G135P67844B8	74566983	74605438	38455
	Island III	WI1-1404J11	G135P64810D7	74040543	74081331	40788
Pax6	*Rcn*	WI1-1767E4	G135P601417F11	105221381	105259938	38557
	Rpl10	WI1-1550J22	G135P601672D2	105387894	105425103	37209
Shh	*Dpp6*	WI1-1085J14	G135P600264D6	27932527	27975636	43109
	Shh	WI1-574O18	G135P64333A4	28754458	28795879	41421
	SFPE1	WI1-552F5	G135P65882F10	28798551	28841224	42673
	SBE4	WI1-2751A6	G135P600205H10	29107140	29147593	40453
	ZRS	WI1-1047E14	G135P600929F6	29611727	29653695	41968

Names are Ensembl (r 45) (http://jun2007.archive.ensembl.org/Mus_musculus/index.html). Mouse genome assembly number: NCBI m37.

6.9.1.2 Fosmid miniprep

I cultured a pipette tip amount of fosmid overnight at 37°C in L-broth containing chloramphenocol (12.5 μg/ml). After removal of the supernatant, I added 200 μl GTE buffer containing lyzozyme and mixed vigorously, then incubated for 5 minutes

at room temperature before adding 400 µl lysis buffer, mixed by inversion and incubated again on ice for 5 minutes. I then added 300 µl acetate buffer, vortexed, and incubated for a further 5 minutes on ice. Following centrifugation at 12000*g* for 5 minutes at 4°C, I performed phenol:chloroform extraction. The fosmid prep was then washed in 70% alcohol and dried before resuspending in 25 µl TE. Rnase A (1µg) was added and incubated for 5 minutes at 37°C. Fosmid prep was stored at -20°C.

GTE buffer
50 mM glucose
25 mM Tris pH8
10 mM EDTA

Lysis buffer
0.2 M NaOH
1% SDS

Acetate buffer (100 ml)
5 M potassium acetate 60 ml
Acetic acid 11.5 ml
Distilled water 28.5 ml

6.9.1.3. Probe labelling (Nick translation)

A reaction mixture containing the following–

2 µl Nick translation salts (stock)
2.5 µl 0.5 mM dATP
2.5 µl 0.5 mM dCTP
2.5 µl 0.5 mM dGTP
2.5 µl bio-16-dUTP (Roche)
OR 1.5 µl digoxygenin-11-dUTP + 1 µl 0.5 mM dTTP
6 µl miniprep DNA (0.5-1 µg)
1 µl DNase I (1:500) (Invitrogen)
1 µl DNA polymerase I (Invitrogen)

–was incubated for 1.5 hours at 16°C. The reaction was stopped by adding 3 µl EDTA and 2 µl 20% SDS, then 65 µl TE was added before purifying through a Quick spin column (Pharmacia). Labelled probes were stored at -20°C.

Nick translation salts stock
0.5 M Tris pH7.5
0.1 M $MgSO_4$
1 mM DT
0.5 mg/ml BSA fraction V (Sigma)

6.9.2. 2D FISH

Cells were swollen in 0.5% trisodium citrate/0.25% KCl followed by fixation in methanol acetic acid (MAA – 3:1 vol/vol). Slides were incubated in 100 μg/ml RNase A in 2 x SCC for 1 hour, washed in 2 x SCC and dehydrated through an alcohol series. Slides were denatured in 70% formamide/2 x SCC for 75 s at 70°C. Between 80-120 ng of biotin- and digoxigenin-labeled fosmid probes were used per slide, with 8-12 μg of mouse Cot1 DNA (Invitrogen) and 10 μg salmon sperm DNA. Probes were denatured at 70°C for 5 minutes, reannealed with Cot1 DNA for 15 minutes at 37°C and hybridized to the denatured slides overnight at 37°C.

Slides were washed 4 x 3 minutes in 2X SSC at 45°C, 4 x 3 minutes in 0.1X SSC at 60°C and transferred to 4X SCC, 0.1% Tween 20. Following incubation for 5 minutes with blocking buffer (4X SCC, 5% Marvel) biotinylated probes were detected using fluorochrome-conjugated avidin (FITC or Texas Red) (Vector Laboratories) then biotinylated antiavidin (Vector Laboratories) followed by fluorochrome-conjugated avidin. Digoxigenin-labeled probes were detected using FITC-conjugated antidigoxigenin (Vector Laboratories) followed by FITC-conjugated anti-sheep (Vector Laboratories). Slides were counterstained in 0.5 μg/ml DAPI.

6.9.3 Mouse embryo sectioning and DNA FISH

Embryos from CD1 mice were collected at E10.5-11, fixed in 4% formaldehyde overnight at 4°C, dehydrated through a graded ethanol series, cleared in xylene and embedded in paraffin blocks. Sections were cut at 6 μm and laid on Superfrost+ slides. The slides were baked at 60°C for 20 minutes, washed four times in xylene for 10 minutes each, rehydrated through an ethanol series before being microwaved for 20 minutes in 0.1 mol/l citrate pH 6 buffer, washed in water and rinsed in 2x SCC before use. FISH was performed as described above, except: for the denaturation step 3minutes at 75°C in 70% formamide/2xSCC pH7.5 followed by 3 minutes in ice-

cold 70% ethanol; and DAPI (0.05 μg/ml) was added to the final wash (5 minutes room temperatute).

6.9.4 Image analysis

P. Perry wrote the scripts for 2D image capture and analysis and M. Pearson set up Volocity packages for 3D image capture and analysis. Both provided guidance for all the microscopy work described in this thesis.

For 2D FISH, slides were imaged using a Hamamatsu Orca AG CCD camera (Hamamatsu Photonics (UK) Ltd, Welwyn Garden City, UK) and a Zeiss Axioplan II fluorescence microscope with Plan-neofluar objectives, a 100W Hg source (Carl Zeiss, Welwyn Garden City, UK) and Chroma #83000 triple band pass filter set (Chroma Technology Corp., Rockingham, VT) with the excitation filters installed in a motorised filter wheel (Prior Scientific Instruments, Cambridge, UK). Image capture and analysis were performed using in-house scripts (P. Perry) written for IPLab Spectrum (Scanalytics Corp, Fairfax, VA). Briefly, Interprobe distance, d, was calculated by segmenting the hybridization signals and determining the xy coordinates of the centroid of each. d was then calculated from $d = [(x_1 - x_2)^2 + (y_1 - y^2)^2]$.

The squared interprobe distances (d^2) were normalised to nuclear radius squared (r^2). The statistical significance of differences in median-squared interprobe distances were assessed using the non-parametric Mann-Whitney U test to examine the null hypothesis that two sets of amine-modified oligos show the same distribution. Each data set consisted of at least 100 nuclei (200 loci) for each cell line, and for each probe combination.

For 3D analysis of tissue sections, slides were imaged with a Hamamatsu Orca AG CCD camera (Hamamatsu Photonics (UK) Ltd, Welwyn Garden City, UK), Zeiss Axioplan II fluorescence microscope with Plan-neofluar or Plan apochromat objectives, a Lumen 200W metal halide light source (Prior Scientific Instruments,

Cambridge, UK) and Chroma #89014ET single excitation and emission filters (Chroma Technology Corp., Rockingham, VT) with the excitation and emission filters installed in Prior motorised filter wheels. A piezoelectrically driven objective mount (PIFOC model P-721, Physik Instrumente GmbH & Co, Karlsruhe) was used to control movement in the z dimension. Hardware control, image capture and analysis were performed using Volocity (Perkinelmer Inc, Waltham, MA). Images were deconvolved using a calculated PSF with the constrained iterative algorithm of Volocity (Perkinelmer Inc, Waltham, MA). Image analysis was carried out using the Quantitation module of Volocity (Perkinelmer Inc, Waltham, MA). For 3D FISH between 90 – 120 loci were measured for each tissue and for each probe combination.

References

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 Biology*, *5*(9), 1906-1911.

Ahn, S., and Joyner, A.L. (2004). Dynamic changes in the response of cells to positive hedgehog signaling during mouse limb patterning. *Cell, 118*, 505–516.

Allan, J., Cowling, G., & Harborne, N. (1981). Regulation of the higher-order structure of chromatin by histones H1 and H5. *The Journal of cell biology*, *90*, 279-288.

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. *Developmental Cell*, *16*, 47-57.

Ardehali, M. B., Mei, A., Zobeck, K. L., Caron, M., Lis, J. T., & Kusch, T. (2011). Drosophila Set1 is the major histone H3 lysine 4. *The EMBO Journal*, *30*(14), 2817-2828.

Arnosti, D. N., & Kulkarni, M. M. (2005). Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards? *Journal of cellular biochemistry*, *94*(5), 890-8.

Azuara, V., Perry, P., Sauer, S., Spivakov, M., Jørgensen, H. F., John, R. M., Gouti, M., Casanova, M., Warnes, G., Merkenschlager, M., Fisher, A. G. (2006). Chromatin signatures of pluripotent cell lines. *Nature Cell Biology*, *8*(5), 532-538.

Bates, D., Butler, P., Pearson, E., & Thomas, J. (1981). Stability of the Higher-Order Structure of Chicken-Erythrocyte Chromatin in Solution. *European Journal of Biochemistry*, *119*, 469-476.

Bannister, A. J., & Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Nature*, *21*(3), 381-395.

Bannister, A. J. & Kouzarides, T. (1996). The CBP co-activator is a histone acetyltransferase. *Nature*, *384*, 641-643.

Bantignies, F., Roure, V., Comet, I., Leblanc, B., Schuettengruber, B., Bonnet, J., Tixier, V., et al. (2011). Polycomb-dependent regulatory contacts between distant Hox loci in Drosophila. *Cell*, *144*(2), 214-26.

Barski, A., Cuddapah, S., Cui, K., Roh, T., Schones, D., Wang, Z., Wei, G., Chepelev, I. & Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. *Cell*, 823-837.

Bell, A C., West, A G., & Felsenfeld, G. (1999). The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell*, *98*(3), 387-96.

Benazet, J., Bischofberger, M., Tiecke, E., Goncalves, A., Martin, J. F., et al. (2009) A self-regulatory system of interlinked signalling feedback loops controls mouse limb patterning. *Science*, *323*, 1050-1053

Bernstein, B. E., Meissner, A., & Lander, E. S. (2007). The Mammalian Epigenome. *Cell 4*, 669-681.

Bernstein, B., Kamal, M., & Lindblad-Toh, K. (2005). Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell*, *120*, 169-181.

Bernstein, B., Mikkelsen, T., Xie, X., Kamal, M., Huebert, D., Cuff, J., Fry, B., et al. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*, *125*, 315-326.

Beuchle, D., Struhl, G., & Müller, J. (2001). Polycomb group proteins and heritable silencing of Drosophila Hox genes. *Development*, *128*(6), 993-1004.

Bhaumik, S., & Smith, E. (2007). Covalent modifications of histones during development and disease pathogenesis. *Nature structural & molecular biology*, *14*(11), 1008-1016.

Bi, X., Yu, Q., & Sandmeier, J. (2004). Formation of boundaries of transcriptionally silent chromatin by nucleosome-excluding structures. *Molecular and cellular biology*, *24*(5), 2118-2131.

Birney, E., Stamatoyannopoulos, J. a, Dutta, A., Guigó, R., Gingeras, T. R., Margulies, E. H., Weng, Z., et al. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, *447*(7146), 799-816.

Blow, M. J., Mcculley, D. J., Li, Z., Zhang, T., Akiyama, J. A., Holt, A., Plajzer-frick, I., et al. (2011). ChIP-seq Identification of Weakly Conserved Heart Enhancers. *Nature Genetics*, *42*(9), 806-810.

Blumberg, B., & Evans, R. M. (1998). Orphan nuclear receptors − new ligands and new possibilities. *Genes & Development*, *12*, 3149-3155.

Bonaldi, T., Imhof, A., and Regula, J. T. (2004). A combination of different mass spectroscopic techniques for the analysis of dynamic changes of histone modifications. Proteomics *4*, 1382-1396.

Bownes, M. (1990). Preferential insertion of P elements into genes expressed in the germ-line of Drosophila melanogaster. *Molecular & general genetics : MGG*, *222*(2-3), 457-60.

Boyer, L., 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., Richard A. Young, R. A. & Jaenisch, R. (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature*, *441*, 349-353.

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. *Human molecular genetics*, *10*(3), 211-9.

Boyle, S., Rodesch, M. J., Halvensleben, H. A., Jeddeloh, J. A., & Bickmore, W. A. (2011). Fluorescence in situ hybridization with high-complexity repeat-free oligonucleotide probes generated by massively parallel synthesis. *Chromosome research*, *19*(7), 901-9.

Britten, R., & Davidson, E. (1969). Gene regulation for higher cells: a theory. *Science*, *165*(3891), 349-357.

Brockdorff, N. (2011). Chromosome silencing mechanisms in X-chromosome inactivation: unknown unknowns. *Development*, *5065*, 5057-5065.

Brookes, E., & Pombo, A. (2009). Modifications of RNA polymerase II are pivotal in regulating gene expression states. *EMBO reports*, *10*, 1213-1219.

Brookes, E., Santiago, I. de, Hebenstreit, D., Morris, K., Carroll, T., Xie, S. Q., Stock, J. K., Heidemann, Eick, M. D., Nozaki, N., Kimura, H., Ragoussis, J., Teichmann, S. A. & Pombo, A. (2012). Polycomb Associates Genome-wide with a Specific RNA Polymerase II Variant, and Regulates Metabolic Genes in ESCs. *Cell Stem Cell*, *10*, 157-170.

Brown, J. M., Leach, J., Reittie, J. E., Atzberger, A., Lee-Prudhoe, J., Wood, W. G., Higgs, D. R., et al. (2006). Coregulated human globin genes are frequently in spatial proximity when active. *The Journal of cell biology*, *172*(2), 177-87.

Buchwald, G., Stoop, P. V. D., Weichenrieder, O., Perrakis, A., van Lohuizen, M., & Sixma, T. K. (2006). Structure and E3-ligase activity of the Ring – Ring complex of Polycomb proteins Bmi1 and Ring1b. *The EMBO*, *25*(11), 2465-2474.

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

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R., Zhang, Y. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science*, *298*, *1039*-1043 doi:10.1126/science.1076997

Caretti, G., Padova, M. D., & Micales, B., Lyons, G. E. & Sartorelli, V. (2004). The Polycomb Ezh2 methyltransferase regulates muscle gene expression and skeletal muscle differentiation. *Genes & development*, *18*, 2627-2638.

Chambeyron, S., & Bickmore, W. A. (2004). Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription. *Genes & development*, *18*(10), 1119-30.

Chambeyron, S., Da Silva, N. R., Lawson, K. a, & Bickmore, W. A. (2005). Nuclear re-organisation of the Hoxb complex during mouse embryonic development. *Development*, *132*(9), 2215-23.

Chao, W., Huynh, K. D., Spencer, R. J., Davidow, L. S., & Lee, J. T. (2002). CTCF, a candidate trans-acting factor for X-inactivation choice. *Science*, *295*(5553), 345-7.

Chen, H., Lin, R. J., Schiltz, R. L., Chakravarti, D., Nash, a, Nagy, L., Privalsky, M. L., et al. (1997). Nuclear receptor coactivator ACTR is a novel histone acetyltransferase and forms a multimeric activation complex with P/CAF and CBP/p300. *Cell*, *90*(3), 569-80.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V. B., Wong, E., et al. (2008). Integration of External Signaling Pathways with the Core Transcriptional Network in Embryonic Stem Cells. *Cell*, *133*, 1106-1117.

Christensen, J., Agger, K., Cloos, P., & Pasini, D., Rose, S., Sennels, L., Rappsilber, J., Hansen, K. H., Salcini, A. E. & Helin, K. (2007). RBP2 belongs to a family of demethylases, specific for tri-and dimethylated lysine 4 on histone 3. *Cell*, *4*, 1063-1076.

Chung, J. H., Whiteley, M., & Felsenfeld, G. (1993). A 5' element of the chicken beta-globin domain serves as an insulator in human erythroid cells and protects against position effect in Drosophila. *Cell*, *74*(3), 505-14.

Cotney, J., Leng, J., Oh, S., Demare, L. E., Reilly, S. K., Gerstein, M. B., & Noonan, J. P. (2012). Chromatin state signatures associated with tissue-specific gene expression and enhancer activity in the embryonic limb. *Genome research*, *22*(6), 1069-80.

Cretekos, C. J., Wang, Y., Green, E. D., NISC Comparative Sequencing Program, Martin, J. F., Rasweiler IV,J. J. & Behringer, R. R. (2008). Regulatory divergence modifies limb length between mammals. *Genes & Development*, *22*, 141-151.

Creyghton, M., Cheng, A., Welstead, G., Kooistra, T., Carey, B., Steine, E., Hanna, J., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *PNAS*. doi:10.1073/pnas.1016071107

Czermin, B., Melfi, R., McCabe, D., Seitz, V., Imhof, A., & Pirrotta, V. (2002). Drosophila Enhancer of Zeste/ESC Complexes Have a Histone H3 Methyltransferase Activity that Marks Chromosomal Polycomb Sites. *Cell*, *111*, 185-196.

Dahn, R., D., Davis, M. C., Pappano, W. N., Shubin, N. H. (2007) Sonic hedgehog function in chondrichthyan fins and the evolution of appendage patterning. *Nature*, *445*, 311-314.

Davey, M. G., Paton, I. R., Yin, Y., Schmidt, M., Bangs, F. K., Morrice, D. R., Smith, T. G., Buxton, P., Stamataki, D., Tanaka, M., Münsterberg, A. E., James Briscoe, J., Tickle, C. & Burt, D. W. (2006). The chicken talpid3 gene encodes a novel protein essentialfor Hedgehog signaling. *Genes & Development*, *20*, 1365-1377.

de Napoles, M., Mermoud, J. E., Wakao, R., Tang, Y. A., Endoh, M., Appanah, R., Nesterova, T. B., et al. (2004). Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation. *Developmental cell*, *7*(5), 663-76.

Deng, W., Lee, J., Wang, H., Miller, J., Reik, A., Gregory, P. D., Dean, A., et al. (2012). Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell*, *149*(6), 1233-44.

De Santa, F., Barozzi, I., Mietton, F., Ghisletti, S., Polletti, S., Tusi, B. K., Muller, H., Ragoussis, J., Wei, C., Natoli, G. (2010). A large fraction of extragenic RNA Pol II transcription sites overlap enhancers. *PLoS Biology*, *8*, 1-17.

Deschamps, J. (2007). Ancestral and recently recruited global control of the Hox genes in development. *Current opinion in genetics & development*, *17*(5), 422-7.

Deschamps, J., & van Nes, J. (2005). Developmental regulation of the Hox genes during axial morphogenesis in the mouse. *Development*, *132*(13), 2931-42.

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*, 376-380.

Dorsett, D. (2011). Cohesin: genomic insights into controlling gene transcription and development. *Current opinion in genetics & development*, *21*(2), 199-206.

Dostie, J., & Bickmore, W. A. (2012). Chromosome organization in the nucleus - charting new territory across the Hi-Cs. *Current opinion in genetics & development*, *22*, 125-131

Duboc, V. and Logan, M. P. O. (2011). Regulation of limb bud initiation and limb type morphology. *Dev. Dyn.*, *240*, 1017-1027.

Dundar, M., Gordon, T. M., Ozyazgan, I., Oguzkaya, F., Ozkul, Y., Cooke, A., Wilkinson, A. G., et al. (2001). A novel acropectoral syndrome maps to chromosome 7q36. *Journal of Medical Genetics*, *38*, 304-309.

Dunn, I. C., Paton, I. R., Clelland, A. K., Sebastian, S., Johnson, E. J., Mcteir, L., Windsor, D., Sherman, A., Sang, H., Burt, D. W., Tickle, C. & Davey, M. G. et al. (2011). The Chicken Polydactyly (Po) Locus Causes Allelic Imbalance and Ectopic Expression of Shh During Limb Development. *Developmental Dynamics*, *240*, 1163-1172.

Dunning, M. J., Smith, M. L., Ritchie, M. E. and Tavaré, S. (2007). beadarray: R classes and methods for Illumina bead-based data. *Bioinformatics*, 23, 2183-2184.

Echelard, Y., Epstein, D. J., St.-Jaques, B., Shen, L., Mohler, J., McMahon, J. A., McMahon, A. P. (1993) Sonic hedgehog, a member of a family of putative signalling molecules, is implicated in the regulation of CNS polarity. *Cell 75*, 1417-1430.

Eckner, R., Ewen, M., Newsome, D., Gerdes, M., DeCaprio, J., Lawrence, J., & Livingston, D. (1994). Molecular cloning and functional analysis of the adenovirus E1A-associated 300-kD protein (p300) reveals a protein with properties of a transcriptional adaptor. *Genes & Development*, *8*, 869-884.

Eden, E., Navon, R., Steinfeld, I., Lipson, D. and Yakhini, Z. (2009). GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics*, *10*, 48.

Eissenberg, J. C., & Shilatifard, A. (2010). Histone H3 lysine 4 (H3K4) methylation in development and differentiation. *Developmental Biology*, *339*(2), 240-249.

Elf, J., Li, G. W. & Xie, X., S. (2007) Probing transcription factor dynamics at the single-molecule level in a living cell. *Science*, *316*, 1191-1194.

Emison, E. S., Mccallion, A. S., Kashuk, C. S., Bush, R. T., Grice, E., Lin, S., Portnoy, M. E., et al. (2005). A common sex-dependent mutation in a RET enhancer underlies Hirschsprung disease risk. *Nature*, *434*, 857-863.

Endoh, M., Endo, T. a, Endoh, T., Fujimura, Y.-ichi, Ohara, O., Toyoda, T., Otte, A. P., et al. (2008). Polycomb group proteins Ring1A/B are functionally linked to the core transcriptional regulatory circuitry to maintain ES cell identity. *Development*, *135*(8), 1513-24.

Engstrom, P. G., Fredman, D., Lenhard, B. (2008) Ancora: a web resource for exploring highly conserved noncoding elements and their association with developmental regulatory genes. *Genome Biol.*, *9*, R34.

Ernst, P., Mabon, M., Davidson, A. J., Zon, L. I., & Korsmeyer, S. J. (2004). An Mll-Dependent Hox Program Drives Hematopoietic Progenitor Expansion. *Current Biology*, *14*, 2063-2069.

Eskeland, R., Leeb, M., Grimes, G. R., Kress, C., Boyle, S., Sproul, D., Gilbert, N., et al. (2010). Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination. *Molecular cell*, *38*(3), 452-64.

Fan, Y., Nikitina, T., Morin-Kensicki, E., Zhao, J., Magnuson, T. R., Woodcock, C. L., & Skoultchi, A. I. (2003). H1 linker histones are essential for mouse development and affect nucleosome spacing in vivo. *Molecular and cellular biologyand cellular biology*, *23*(13), 4559-4572.

Fan, Y., Nikitina, T., Zhao, J., Fleury, T. J., Bhattacharyya, R., Bouhassira, E. E., Stein, A., et al. (2005). Histone H1 depletion in mammals alters global chromatin structure but causes specific changes in gene regulation. *Cell*, *123*(7), 1199-212.

Farooq, M., Troelsen, J.T., Boyd, M., Eiberg, H., Hansen, L., Hussain, M.S., Rehman, S., Azhar, A., Ali, A., Bakhtiar, S.M., et al. (2010). Preaxial polydactyly/triphalangeal thumb is associated with changed transcription factorbinding affinity in a family with a novel point mutation in the long-range cisregulatory element ZRS. *Eur. J. Hum. Genet.*, *18*, 733–736.

Faust, C., Schumacher, A., Holdener, B., & Magnuson, T. (1995). The eed mutation disrupts anterior mesoderm production in mice. *Development*, *121*, 273-285.

Faust, C., Lawson, K. A., Schork, N. J., Thiel, B. and Magnuson, T. (1998). The Polycomb-group gene eed is required for normal morphogenetic movements during gastrulation in the mouse embryo. *Development, 125*, 4495-4506.

Feng, J., Bi, C., Clark, B. S., Mady, R., Shah, P. & Kohtz, J. D. (2006) The *Evf-2* noncoding RNA is transcribed from the Dix-5/6 ultraconserved region and functions as a Dix-2 transcriptional activator *Genes and Development* **20**, 1470-1484.

Fernandez-Teran, M., Piedra, M. E., Kathiriya, I.S., Srivastava, D., Rodriguez-Rey, J. C., Ros, M. A. Role of dHAND in the anterior-posterior polarization of the limb bud: implications for the Sonic hedgehog pathway. *Development, 127*, 2133-2142.

Filippova, G. N., Thienes, C. P., Penn, B. H., Cho, D. H., Hu, Y. J., Moore, J. M., Klesert, T. R., et al. (2001). CTCF-binding sites flank CTG/CAG repeats and form a methylation-sensitive insulator at the DM1 locus. *Nature genetics*, *28*(4), 335-43.

Finlan, L. E., Sproul, D., Thomson, I., Boyle, S., Kerr, E., Perry, P., Ylstra, B., et al. (2008). Recruitment to the nuclear periphery can alter expression of genes in human cells. *PLoS genetics*, *4*(3), e1000039.

Francis, N. J., Kingston, R. E., & Woodcock, C. L. (2004). Chromatin compaction by a polycomb group protein complex. *Science*, *306*(5701), 1574-7.

Francis, N., & Kingston, R. (2001). Mechanisms of transcriptional memory. *Nature Reviews Molecular Cell, 2*, 409-421.

Fullwood, M. J., Liu, M. H., Pan, Y. F., Liu, J., Han, X., Bin, Y., Orlov, Y. L., et al. (2010). An Oestrogen Receptor α-bound Human Chromatin Interactome. *Nature*, *462*(7269), 58-64.

Furniss, D., Lettice, L.A., Taylor, I.B., Critchley, P.S., Giele, H., Hill, R.E. & Wilkie, A.O. (2008). A variant in the sonic hedgehog regulatory sequence (ZRS) is associated with triphalangeal thumb and deregulates expression in the developing limb. *Hum. Mol. Genet.*, *17*, 2417–2423.

Fussner, E., Ching, R. W., & Bazett-Jones, D. P. (2011). Living without 30nm chromatin fibers. *Trends in biochemical sciences*, *36*(1), 1-6.

Gaszner, M., & Felsenfeld, G. (2006). Insulators: exploiting transcriptional and epigenetic mechanisms. *Nature reviews. Genetics*, *7*(9), 703-13.

Gerasimova, T. I., Byrd, K., & Corces, V. G. (2000). A chromatin insulator determines the nuclear localization of DNA. *Molecular cell*, *6*(5), 1025-35.

Geyer, P. K., Spana, C., & Corces, V. G. (1986). On the molecular mechanism of gypsy-induced mutations at the yellow locus of Drosophila melanogaster. *The EMBO journal*, *5*(10), 2657-62.

Ghirlando, R., & Felsenfeld, G. (2008). Hydrodynamic studies on defined heterochromatin fragments support a 30-nm fiber having six nucleosomes per turn. *Journal of molecular biology*, *376*(5), 1417-25.

Gilbert, N., Boyle, S., Sutherland, H., de Las Heras, J., Allan, J., Jenuwein, T., and Bickmore, W. A. (2003). Formation of facultative heterochromatin in the absence of HP1. *Embo J, 22*, 5540-5550.

Gilbert, N., Boyle, S., Fiegler, H., Woodfine, K., Carter, N. P., & Bickmore, W. A. (2004). Chromatin architecture of the human genome: gene-rich domains are enriched in open chromatin fibers. *Cell*, *118*(5), 555-66.

Gonzalez, F., Duboule, D., & Spitz, F. (2007) Transgenic analysis of Hoxd gene regulation during digit development. *Dev.Biol.* 306, 847–859.

Gordon, C. T., Tan, T. Y., Benko, S., Fitzpatrick, D., Lyonnet, S., & Farlie, P. G. (2009). Long-range regulation at the SOX9 locus in development and disease. *Journal of medical genetics*, *46*(10), 649-56.

Gotea, V., Visel, A., Westlund, J.M., Nobrega, M.A., Pennacchio, L.A., & Ovcharenko, I. (2010). Homotypic clusters of transcription factor binding sites are a key component of human promoters and enhancers. *Genome Res.*, *20*, 565–577.

Grosveld, F., Assendelft, G. M. V., Greaves, D. R., & Kollias, G. (1987). Position-Independent , High-Level Expression of the Human Beta-Globin Gene in Transgenic Mice. *Cell*, *51*, 975-985.

Gurnett, C.A., Bowcock, A.M., Dietz, F.R., Morcuende, J.A., Murray, J.C., and Dobbs, M.B. (2007). Two novel point mutations in the long-range SHH enhancer in three families with triphalangeal thumb and preaxial polydactyly. *Am. J. Med. Genet. A.*, *143*, 27–32.

Hadjur, S., Williams, L. M., Ryan, N. K., Cobb, B. S., Sexton, T., Fraser, P., Fisher, A. G., et al. (2009). Cohesins form chromosomal cis-interactions at the developmentally regulated IFNg locus. *Nature*, *460*(7253), 410-3.

Haraguchi, R., Mo, R., Hui, C., Motoyana, J., Makino, S., Shiroishi, T., Gaffield, W., Yamada, G. (2001) Unique functions of sonic hedgehog signalling during external genitalia development. *Development*, 128, 4241-4250.

Harfe, B. D., Scherz, P. J., Nissim, S., Tian, H., McMahon, A. P. and Tabin, C. J. (2004). Evidence for an expansion-based temporal Shh gradient in specifying vertebrate digit identities. *Cell*, 118, 517-528.

Harris, M. B., Mostecki, J., & Rothman, P. B. (2005). Repression of an interleukin-4-responsive promoter requires cooperative BCL-6 function. *The Journal of biological chemistry*, *280*(13), 13114-21.

He, H. H., Meyer, C. a, Shin, H., Bailey, S. T., Wei, G., Wang, Q., Zhang, Y., et al. (2010). Nucleosome dynamics define transcriptional enhancers. *Nature genetics*, *42*(4), 343-7.

Heard, E. (2005). Delving into the diversity of facultative heterochromatin: the epigenetics of the inactive X chromosome. *Current opinion in genetics & development*, *15*(5), 482-9.

Heintzman, N. D., Hon, G. C., Hawkins, R. D., Kheradpour, P., Stark, A., Harp, L. F., Ye, Z., et al. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature*, *459*(7243), 108-12.

Heintzman, N. D., Ren, B. (2009) Finding distal regulatory elements in the human genome. *Curr. Opin. Genet. Dev.*, *19*, 541-549.

Heintzman, N., Stuart, R., Hon, G., Fu, Y., Ching, C. W., Hawkins, R. D., Barrera, L. O., Calcar, S. V., Qu, C., Ching, K. A., Wang, W., Weng, Z., Green, R. D., Crawford, G. E. & Ren, B. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature genetics*, *39*(3), 311-318.

Hill, R. E. (2007). How to make a zone of polarizing activity : Insights into limb development via the abnormality preaxial polydactyly. *Develop. Growth Diff. 49*, 439-448

Hiragami-hamada, K., Xie, S. Q., Saveliev, A., Uribe-lewis, S., Pombo, A., & Festenstein, R. (2009). The Molecular basis of heterochromatin-mediated silencing in mammals. *Epigenetics and Chromatin*, *9*, 1-17.

Hollenhorst, P. C., Chandler, K. J., Poulsen, R. L., Johnson, W. E., Speck, N. a, & Graves, B. J. (2009). DNA specificity determinants associate with distinct transcription factor functions. *PLoS genetics*, *5*(12), e1000778.

Hollenhorst, P. C., Ferris, M. W., Hull, M. a, Chae, H., Kim, S., & Graves, B. J. (2011). Oncogenic ETS proteins mimic activated RAS/MAPK signaling in prostate cells. *Genes & development*, *25*(20), 2147-57.

Hollenhorst, P. C., McIntosh, L. P., & Graves, B. J. (2011). Genomic and biochemical insights into the specificity of ETS transcription factors. *Annual review of biochemistry*, *80*, 437-71.

Hollenhorst, P. C., Shah, A. a, Hopkins, C., & Graves, B. J. (2007). Genome-wide analyses reveal properties of redundant and specific promoter occupancy within the ETS gene family. *Genes & development*, *21*(15), 1882-94.

Hou, C., Dale, R., & Dean, A. (2010). Cell type specificity of chromatin organization mediated by CTCF and cohesin. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(8), 3651-6.

Ianakiev, P., van Baren MJ, Daly, M. J., Toledo, S. P., Cavalcanti, M. G., Neto, J. C., Silveira, E. L., et al. (2001). Acheiropodia is caused by a genomic deletion in C7orf2, the human orthologue of the Lmbr1 gene. *American journal of human genetics*, *68*(1), 38-45.

Ingham, P. W. (1985). A clonal analysis of the requirement for the trithorax gene in the diversification of segments in Drosophila. *Journal of embryology and experimental morphology*, *89*, 349-65.

Iwase, S., Lan, F., Bayliss, P., de la Torre-Ubieta, L., Huarte, M., Qi, H. H., Whetstine, J. R., et al. (2007). The X-linked mental retardation gene SMCX/JARID1C defines a family of histone H3 lysine 4 demethylases. *Cell*, *128*(6), 1077-88.

Jacob, F., Monod, M. (1961) Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology*. 3, 318-356

Jat, P. S., Noble, M. D., Ataliotis, P., Tanaka, Y., Yannoutsos, N., Larsen, L. and Kioussis, D. (1991). Direct derivation of conditionally immortal cell lines from an H-2Kb-tsA58 transgenic mouse. *Proc. Natl. Acad. Sci. USA, 88,* 5096-5100.

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*, 7761-7772.

Jiang, H., Shukla, A., Wang, X., Chen, W.-yi, Bernstein, B. E., & Roeder, R. G. (2011). Role for Dpy-30 in ES cell-fate specification by regulation of H3K4 methylation within bivalent domains. *Cell*, *144*(4), 513-25.

Jin, C., Zang, C., Wei, G., Cui, K., Peng, W., Zhao, K., & Felsenfeld, G. (2009). H3.3/H2A.Z double variant-containing nucleosomes mark "nucleosome-free regions" of active promoters and other regulatory regions. *Nature genetics*, *41*(8), 941-5.

Jing, H., Vakoc, C. R., Ying, L., Mandat, S., Wang, H., Zheng, X., & Blobel, G. A. (2008). Exchange of GATA factors mediates transitions in looped chromatin organization at a developmentally regulated gene locus. *Molecular cell*, *29*(2), 232-42.

Johnson, D. S., Mortazavi, A., Myers, R. M., & Wold, B. (2007). Genome-Wide Mapping of in Vivo Protein-DNA Interactions. *Science*, *316*, 1497-1502.

Jorgensen, H., Azuara, V., Amoils, S., Spivakov, M., Terry, A., Nesterova, T., Cobb, B. S., Ramsahoye, B., Merkenschlager, M. & Fisher, A. G. (2007). The impact of chromatin modifiers on the timing of locus replication in mouse embryonic stem cells. *Genome Biology*, *8*(8). doi:10.1186/gb-2007-8-8-r169

Kang, H.-seo, Nelson, M. L., Mackereth, C. D., Schärpf, M., Barbara, J., & Mcintosh, L. P. (2009). Identification and Structural Characterization of a CBP/p300- Binding Domain from the ETS Family Transcription Factor GABPα. *Journal of Molecular Biology*, *377*(3), 636-646.

Kanhere, A., Viiri, K., Araújo, C. C., Rasaiyaah, J., Bouwman, R. D., Whyte, W. a, Pereira, C. F., et al. (2010). Short RNAs are transcribed from repressed polycomb target genes and interact with polycomb repressive complex-2. *Molecular cell*, *38*(5), 675-88.

Katzenellenbogen, B. S. (1996). Estrogen receptors: bioactivities and interactions with cell signaling pathways. *Biology of reproduction*, *54*(2), 287-93.

Kim, T., Hemberg, M., Gray, J. M., Costa, A. M., Bear, D. M., Wu, J., Harmin, D. A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., Markenscoff-Papadimitriou, E., Kuhl, D., Bito, H., Worley, P. F., Kreiman, G., Greenber, M., E. (2010). Widespread transcription at neuronal activity-regulated enhancers. *Nature*, *465*, 182-187.

Kioussis, D., Vanin, E., DeLange, T., Flavell, R., & Grosveld, F. (1983). Beta-globin gene inactivation by DNA translocation in gamma beta-thalassaemia. *Nature*, *306*, 662-666.

Klymenko, T., & Müller, J. (2004). The histone methyltransferases Trithorax and Ash1 prevent transcriptional silencing by Polycomb group proteins. *EMBO reports*, *5*(4), 373-377.

Kmita, M., Fraudeau, N., Herault, Y., Duboule, D. (2002) Serial locus deletions and duplications *in vivo* suggest a mechanism for Hoxd genes colinearity in developing limbs. *Nature*, *420*, 145-150

Kmita, M., & Duboule, D. (2003). Organizing axes in time and space; 25 years of colinear tinkering. *Science*, *301*(5631), 331-3.

Knezevic, V., De Santo, R., Schughart, K., Huffstadt, U., Chiang, C., Mahon, K. A., Mackem, S. (1997) *Hoxd12* differentially affects preaxial and postaxial chondrogenic branches in the limb and regulates sonic hedgehog in a positive feedback loop. *Development* , *124*, 4523-4536

Kohn, E. A., Du, Z., Sato, M., Van Schyndle, C. M., Welsh, M. A., Yang, Y.-A., Stuelten, C. H., Tang, B., Ju, W., Bottinger, E. P. et al. (2010). A novel approach for the generation of genetically modified mammary epithelial cell cultures yields new insights into TGFb signaling in the mammary gland. *Breast Cancer Res.*, *12*, R83.

Kotake, Y., Nakagawa, T., Kitagawa, K., Suzuki, S., Liu, N., Kitagawa, M., & Xiong, Y. (2011). Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15(INK4B) tumor suppressor gene. *Oncogene*, *30*(16), 1956-62.

Kouzarides, T. (2007). Chromatin modifications and their function. *Cell*, 693-705.

Kowalczyk, M. S., Hughes, J. R., Garrick, D., Lynch, M. D., Sharpe, J. a, Sloane-Stanley, J. A, McGowan, S. J., et al. (2012). Intragenic enhancers act as alternative promoters. *Molecular cell*, *45*(4), 447-58.

Ku, M., Koche, R., Rheinbay, E., Mendenhall, E., Endoh, M., Mikkelsen,T. S., Presser, A., Nusbaum, C., Xie, X., Chi, A. S., Adli, M., Kasif, S., Ptaszek, L. M., Cowan, C. A., Lander, E. S., Koseki, H. & Bernstein, B. E. (2008). Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS genetics*, *4*(10). doi:10.1371/journal.pgen.1000242

Kuijper, S., Beverdam, A., Kroon, C., Brouwer, A., Candille, S., Barsh, G., & Meijlink, F. (2005). Genetics of shoulder girdle formation: roles of Tbx15 and aristaless-like genes. *Development*, *132*(7), 1601-10.

Kumaran, R. I., & Spector, D. L. (2008). A genetic locus targeted to the nuclear periphery in living cells maintains its transcriptional competence. *The Journal of cell biology*, *180*(1), 51-65.

Kunderfranco, P., Mello-Grand, M., Cangemi, R., Pellini, S., Mensah, A., Albertini, V., Malek, A., et al. (2010). ETS transcription factors control transcription of EZH2 and epigenetic silencing of the tumor suppressor gene Nkx3.1 in prostate cancer. *PloS one*, *5*(5), e10547.

Kuo, M., Zhou, J., Jambeck, P., Churchill, M. E. A., & Allis, C. D. (1998). Histone acetyltransferase activity of yeast Gcn5p is required for the activation of target genes in vivo. *Genes & Development*, *12*, 627-639.

Laufer, E., Nelson, C. E., Johnson, R. L., Morgan, B. A. and Tabin, C. (1994). Sonic hedgehog and Fgf-4 act through a signaling cascade and feedback loop to integrate growth and patterning of the developing limb bud. *Cell*, *79*, 993-1003.

Lee, N., Zhang, J., Klose, R. J., Erdjument-Bromage, H., Tempst, P., Jones, R. S., & Zhang, Y. (2007). The trithorax-group protein Lid is a histone H3 trimethyl-Lys4 demethylase. *Nature structural & molecular biology*, *14*(4), 341-3.

Lee, T. I., & Young, R. A. (2000). Transcription of eukaryotic protein-coding genes. *Annual review of genetics*, *34*, 77-137.

Lee, T., Jenner, R., Boyer, L., Guenther, M., et al. (2006). Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell*, *125*, 301-313.

Leeb, M., & Wutz, A. (2007). Ring1B is crucial for the regulation of developmental control genes and PRC1 proteins but not X inactivation in embryonic cells. *The Journal of cell biology*, *178*(2), 219-229.

Lehoczky, J. A, Williams, M. E., & Innis, J. W. (2004). Conserved expression domains for genes upstream and within the HoxA and HoxD clusters suggests a long-range enhancer existed before cluster duplication. *Evolution & development*, *6*(6), 423-30.

Lehoczky, J. A. & Innes, J. W. (2008) BAC transgenic analysis reveals enhancers sufficient for *Hoxa13* and neighbourhood gene expression in mouse embryonic distal limbs and genital bud. *Evol. Dev.*, *10*, 421-432.

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 *Human Molecular Genetics*, *17*, 978-985.

Lettice, L., Horikoshi, T., Heaney, S., van Baren, M., van der Linde, H., Breedveld, G., Joosse, M., et al. (2002). Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly. *PNAS*, *99*(11), 7548-7553.

Lettice, L. A., Heaney, S. J., Purdie, L. A., Li, L., de Beer, P., Oostra, B. A., Goode, D., Elgar, G., Hill, R. E., de Graaf, E. (2003) A long-range *Shh* enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly *Human Molecular Genetics*, *12*, 1725-1735.

Lettice, L. A., Iain Williamson,I., Wiltshire, J. H., Peluso, S., Devenney, P. S., Hill, A. E., Essafi, A., Hagman, J., Mort, R., Grimes, G., Carlo L. DeAngelis, C. L., & Hill, R. E. (2012) Opposing functions of the ETS factor family define *Shh* spatial expression in limb buds and polydactyly. *Developmental Cell*, *22*, 459-467.

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

Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, *326*(5950), 289-93.

Lippman, Z., Gendrel, A., Black, M., Vaughn, M., et al. (2004). Role of transposable elements in heterochromatin and epigenetic control. *Nature*, *430*, 471-476.

Lifanov, A.P., Makeev, V.J., Nazina, A.G., & Papatsenko, D.A. (2003). Homotypic regulatory clusters in Drosophila. *Genome Res.*, *13*, 579–588.

Litingtung, Y., Dahn, R. D., Li, Y., Fallon, J. F. & Chiang, C. (2002). Shh and Gli3 are dispensable for limb skeleton formation but regulate digit number and identity. *Nature*, *418*, 979-983. doi:10.1038/nature01012.1.

Litt, M D, Simpson, M., Gaszner, M., Allis, C. D., & Felsenfeld, G. (2001b). Correlation between histone lysine methylation and developmental changes at the chicken beta-globin locus. *Science*, *293*(5539), 2453-5.

Litt, M., Simpson, M., Recillas-Targa, F., Prioleau, M.-N., & Felsenfeld, G. (2001a). Transitions in histone acetylation reveal boundaries of three separately regulated neighboring loci. *The EMBO*, *20*(9), 2224-2235.

Lizarraga, G., Ferrari, D., Kalinowski, M., Ohuchi, H., Noji, S., Kosher, R. A. and Dealy, C. N. (1999). FGFR2 signaling in normal and limbless chick limb buds. *Dev. Genet.*, *25*, 331-338.

Lodder, E. M., Eussen, B. H., van Hassel, D. A. C. M., et al. (2009). Implication of long-distance regulation of the HOXA cluster in a patient with postaxial polydactyly, *Chromosome Research*, *17*, 737-744. doi:10.1007/s10577-009-9059-5

Lorente, M. del M., Marcos-Gutierrez, C., Perez, C., Schoorlemmer, J., Ramirez, A., Magin, T., & Vidal, M. (2000). Loss-and gain-of-function mutations show a polycomb group function for Ring1A in mice. *Development*, *127*, 5093-5100.

Loots, G. G., Kneissel, M., Keller, H., Baptist, M., Chang, J., Collette, N., M., Ovcharenko, D., Plajzer-Frick, I., Rubin, E. M. (2005) Genomic deletion of a long-range bone enhancer misregulates sclerostin in Van Buchem disease. *Genome Research*, *15*, 928-935.

Lu, X., Wontakal, S. N., Emelyanov, A. V., Morcillo, P., Konev, A. Y., Fyodorov, D. V., & Skoultchi, A. I. (2009). Linker histone H1 is essential for Drosophila development, the establishment of pericentric heterochromatin, and a normal polytene chromosome structure. *Genes & development*, *23*(4), 452-65.

Lynch, M. D., Smith, A. J. H., Gobbi, M. D., Flenley, M., Hughes, J. R., Vernimmen, D., Ayyub, H., et al. (2011). An interspecies analysis reveals a key role for unmethylated CpG dinucleotides in vertebrate Polycomb complex recruitment. *The EMBO Journal*, *31*(2), 317-329.

Ma, M. K.-W., Heath, C., Hair, A., & West, A. G. (2011). Histone crosstalk directed by H2B ubiquitination is required for chromatin boundary integrity. *PLoS genetics*, *7*(7), e1002175.

Mahy, N. L., Perry, P. E., & Bickmore, W. A. (2002). Gene density and transcription influence the localization of chromatin outside of chromosome territories detectable by FISH. *The Journal of cell biology*, *159*(5), 753-63.

Mahy, N. L., Perry, P. E., Gilchrist, S., Baldock, R. a, & Bickmore, W. A. (2002). Spatial organization of active and inactive genes and noncoding DNA within chromosome territories. *The Journal of cell biology*, *157*(4), 579-89.

Mak, W., Nesterova, T., Napoles, M. D., Appanah, R., Yamanaka, S., Otte, A. P. & Brockdorff, N. (2004). Reactivation of the paternal X chromosome in early mouse embryos. *Science*, *303*, 666-669.. doi:10.1126/science.1092674

Mao, J., Mcglinn, E., Huang, P., Tabin, C. J., & Mcmahon, A. P. (2009). Fgf-Dependent Etv4/5 Activity Is Required for Posterior Restriction of Sonic hedgehog and Promoting Outgrowth of the Vertebrate Limb. *Developmental Cell*, *16*(4), 600-606.

Marchetti, M. (2003). Differential expression of the Drosophila BX-C in polytene chromosomes in cells of larval fat bodies: a cytological approach to identifying in vivo targets of the homeotic Ubx, Abd-A and Abd-B proteins. *Development*, *130*(16), 3683-3689.

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*, 401-405

Martens, J., O'Sullivan, R., & Braunschweig, U., Opravil, S., Radolf, M., Steinlein, P. & Jenuwein, T. (2005). The profile of repeat-associated histone lysine methylation states in the mouse epigenome. *The EMBO*, *24*(4), 800-812.

Maston, G. a, Evans, S. K., & Green, M. R. (2006). Transcriptional regulatory elements in the human genome. *Annual review of genomics and human genetics*, *7*, 29-59.

McFadden, D. G., Charité, J., Richardson, J. a, Srivastava, D., Firulli, a B., & Olson, E. N. (2000). A GATA-dependent right ventricular enhancer controls dHAND transcription in the developing heart. *Development*, *127*(24), 5331-41.

McGrane, M. M. (2007). Vitamin A regulation of gene expression: molecular mechanism of a prototype gene. *The Journal of nutritional biochemistry*, *18*(8), 497-508.

McLean CY, Reno PL, Pollen AA, Bassan AI, Capellini TD, Guenther C, Indjeian VB, Lim X, Menke DB, Schaar BT, Wenger AM, Bejerano G, Kingsley DM. (2011). Human-specific loss of regulatory DNA and the evolution of human-specific traits. Nature. *471*, 216-219.

Mendenhall, E., Koche, R., Truong, T., Zhou, V., Biju Issac, B., Chi, A. S., Ku, M., & Bernstein, B. E. (2010). GC-rich sequence elements recruit PRC2 in mammalian ES cells. *PLoS genetics*, *6*(12), e1001244.

Mercader, N., Selleri, L., Criado, L. M., Pallares, P., Parras, C., Cleary, M. L., & Torres, M. (2009). Ectopic Meis1 expression in the mouse limb bud alters P-D patterning in a Pbx1-independent manner. *The International journal of developmental biology*, *53*(8-10), 1483-94.

Merika, M., Williams, A., Chen, G., Collins, T., & Thanos, D. (1998). Recruitment of CBP/p300 by the IFN enhanceosome is required for synergistic activation of transcription. *Molecular cell*, *1*, 277-287.

Meyer, C. A., Krum, S. A., Rhodes, D. R., Liu, X. S., & Brown, M. (2009). Coactivator Function Defines the Active Estrogen Receptor Alpha Cistrome. *Molecular and Cellular Biology*, *29*(12), 3413-3423.

Mikkelsen, T., Ku, M., Jaffe, D., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*, *448*(August), 553-560.

Milne, T., Briggs, S., Brock, H., Martin, M., Gibbs, D., Allis, C. D., & Hess, J. L. (2002). MLL Targets SET Domain Methyltransferase Activity to Hox Gene Promoters. *Molecular cell*, *10*, 1107-1117.

Mishiro, T., Ishihara, K., Hino, S., Tsutsumi, S., Aburatani, H., Shirahige, K., Kinoshita, Y., et al. (2009). Architectural roles of multiple chromatin insulators at the human apolipoprotein gene cluster. *The EMBO journal*, *28*(9), 1234-45.

Mizzen, C., Yang, X., Kokubo, T., & Brownell, J. (1996). The TAF (II) 250 subunit of TFIID has histone acetyltransferase activity. *Cell*, *87*, 1261-1270.

Mohn, F., Weber, M., Rebhan, M., Roloff, T., Richter, J., Stadler, M. B., Bibel, M. & Schubeler, D. (2008). Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Molecular cell*, *30*, 755-766.

Montavon, T., Garrec, J.-françois L., Kerszberg, M., & Duboule, D. (2008). Modeling Hox gene regulation in digits : reverse collinearity and the molecular origin of thumbness. *Genes & Development*, *22*, 346-359.

Montavon, T., Soshnikova, N., Joye, E., Thevenet, L., Splinter, E., de Laat, W., Spitz, F., et al. (2011). A Regulatory Archipelago Controls Hox Genes Transcription in Digits. *Cell*, *147*, 1132-1145.

Morey, C., Kress, C., & Bickmore, W. A. (2009). Lack of bystander activation shows that localization exterior to chromosome territories is not sufficient to up-regulate gene expression. *Genome research*, *19*(7), 1184-94.

Morey, C., Silva, N. R. D., Perry, P., & Bickmore, W. A. (2007). Nuclear reorganisation and chromatin decondensation are conserved , but distinct , mechanisms linked to Hox gene activation. *Development*, *134*, 909-919.

Morey, L., Pascual, G., Cozzuto, L., Roma, G., Wutz, A., Benitah, S. A., & Di Croce, L. (2012). Nonoverlapping Functions of the Polycomb Group Cbx Family of Proteins in Embryonic Stem Cells. *Cell Stem Cell*, *10*(1), 47-62.

Morikawa, Y., Dai, Y. S., Hao, J., Bonin, C., Hwang, S., Cserjesi P. (2005) The basic helix-loop-helix factor Hand2 regulates autonomic nervous system development. *Dev. Dyn., 234*, 613-621.

Nativio, R., Wendt, K. S., Ito, Y., Huddleston, J. E., Uribe-Lewis, S., Woodfine, K., Krueger, C., et al. (2009). Cohesin is required for higher-order chromatin conformation at the imprinted IGF2-H19 locus. *PLoS genetics*, *5*(11), e1000739.

Nelson, C. E., Morgan, B. A., Burke, A. C., Laufer, E., Dimambro, E., Murtaugh, L. C., Gonzales, E., et al. (1996). Analysis of Hox gene expression in the chick limb bud, *1466*, 1449-1466.

Niswander, L., Jeffrey, S., Martin, G. R., Tickle, C. (1994) A positive feedback loop coordinates growth and patterning in the vertebrate limb, *Nature*, *371*, 609-612

Nobrega, M. A., Ovcharenko, I., Afzal, V., Rubin, E. M. (2003) Scanning human gene deserts for long-range elements. *Science*, *302*, 413.

Noonan, J. P., & Mccallion, A. S. (2010). Genomics of Long-Range Regulatory Elements. *Annu. Rev. Genomics Hum. Genet.*, *11*, 1-23

Noordermeer, D., Branco, M. R., Splinter, E., Klous, P., van Ijcken, W., Swagemakers, S., Koutsourakis, M., et al. (2008). Transcription and chromatin organization of a housekeeping gene cluster containing an integrated beta-globin locus control region. *PLoS genetics*, *4*(3), e1000016.

Noordermeer, D., & de Laat, W. (2008). Joining the loops: beta-globin gene regulation. *IUBMB life*, *60*(12), 824-33.

Nora, E. P., Lajoie, B. R., Schulz, E. G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*, *485*, 381-385.

Ogryzko, V., Schiltz, R., Russanova, V., & Howard, B. (1996). The transcriptional coactivators p300 and CBP are histone acetyltransferases. *Cell*, *87*, 953-959.

Ogura, T., Alvarez, I. S., Vogel, A., Rodríguez, C., Evans, R. M. and Izpisúa-Belmonte, J. C. (1996). Evidence that Shh cooperates with a retinoic acid inducible co-factor to establish ZPA-like activity. *Development*, *122*, 537-542.

Oikawa, T., & Yamada, T. (2003) Molecular biology of the Ets family of transcription factors. *Gene*, *303*, 11-34.

Okada, Y., Nobori, H., Shimizu, M., Watanabe, M., Yonekura, M., Nakai, T., Kamikawa, Y., et al. (2011). Multiple ETS family proteins regulate PF4 gene expression by binding to the same ETS binding site. *PloS one*, *6*(9), e24837.

Okamoto, I., Otte, A. P., Allis, C. D., Reinberg, D., & Heard, E. (2004). Epigenetic dynamics of imprinted X inactivation during early mouse development. *Science*, *303*(5658), 644-649.

Oki, M., Valenzuela, L., Chiba, T., Ito, T. & Kamakaka, R. T. (2004). Barrier proteins remodel and modify chromatin to restrict silenced domains. *Molecular and cellular*, *24*(5), 1956-1967.

O'Carroll, D., Erhardt, S., Pagani, M., Barton, S. C., Surani, M. A. & Jenuwein, T. (2001). The Polycomb-Group GeneEzh2 Is Required for Early Mouse Development. *Molecular and cellular biology*, *21*(13), 4330-4336.

Ovcharenko, I., Loots, G. G., Nobrega, M. A., Hardison, R., C., Miller, W. & Stubbs, L. (2005) Evolution and functional classification of vertebrate gene deserts. *Genome Research*, *15*, 137-145.

Pan, G., Tian, S., Nie, J., Yang, C., Ruotti, V., Wei, H., Jonsdottir, G. A., Stewart, R. & Thomson, J. A. (2007). Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells. *Cell Stem Cell*, *1*, 299-312.

Pasini, D., Bracken, A., Jensen, M., Denchi, E., & Helin, K. (2004). Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. *The EMBO*, *23*(20), 4061-4071.

Pearson, J. C., Lemons, D., & McGinnis, W. (2005). Modulating Hox gene functions during animal body patterning. *Nature reviews. Genetics*, *6*(12), 893-904.

Peckham, H., Thurman, R., Fu, Y., Stamatoyannopoulos, J. A., Noble, W. S., Struhl, K. & Weng, Z. (2007). Nucleosome positioning signals in genomic DNA. *Genome Research*, *17*, 1170-1177. doi:10.1101/gr.6101007.dependent

Pekowska, A., Benoukraf, T., Zacarias-Cabeza, Belhocine, J. M., Frederic Koch, F., Holota, H., Imbert, J., Andrau, J., Ferrier, P., Spicuglia, S. (2011). H3K4 tri-methylation provides an epigenetic signature of active enhancers. *The EMBO Journal*, *30*(20), 4198-4210.

Pietersen, A. M., & van Lohuizen, M. (2008). Stem cell regulation by polycomb repressors: postponing commitment. *Current opinion in cell biology*, *20*(2), 201-7.

Pirrotta, V. (1997). Chromatin-silencing mechanisms in Drosophila maintain patterns of gene expression. *Trends in Genetics*, *13*(8), 314-318.

Pomerantz, M. M., Ahmadiyeh, N., Jia, L., Herman, P., Verzi, M. P., et al. (2009) The 8q24 cancer risk variant rs6983267 shows long-range interaction with *MYC* in colorectal cancer. *Nature Genetics*, *41*, 882-884.

Prabhakar, S., Visel, A., Akiyama, J. a, Shoukry, M., Lewis, K. D., Holt, A., Plajzer-Frick, I., et al. (2008). Human-specific gain of function in a developmental enhancer. *Science*, *321*(5894), 1346-50.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S., Flynn, R. A. & Wysocka, J. (2010). A unique chromatin signature uncovers early developmental enhancers in humans. *Nature*, *470*(7333), 279-283.

Ragoczy, T., Bender, M. a, Telling, A., Byron, R., & Groudine, M. (2006). The locus control region is required for association of the murine beta-globin locus with engaged transcription factories during erythroid maturation. *Genes & development*, *20*(11), 1447-57.

Ragoczy, T., Telling, A., Sawado, T., Groudine, M., & Kosak, S. T. (2003). A genetic analysis of chromosome territory looping: diverse roles for distal regulatory elements. *Chromosome Research* , *11*(5), 513-25.

Reddy, K. L., Zullo, J. M., Bertolino, E., & Singh, H. (2008). Transcriptional repression mediated by repositioning of genes to the nuclear lamina. *Nature*, *452*(7184), 243-7.

Rebeiz, M., Pool, J. E., Kassner, V. A., Aquadro, C. F., Carrol, S. B. (2009). Stepwise modification of a modular enhancer underlies adaptation in a *Drosophila* population. Science. *326*, 1663-1667.

Richardson, L., Venkataraman, S., Stevenson, P., Yang, Y., Burton, N., Rao, J., Fisher, M., Baldock, R.A., Davidson, D.R., and Christiansen, J.H. (2010). EMAGE mouse embryo spatial gene expression database: 2010 update. *Nucleic Acids Res.*, *38* (Database issue), D703–D709.

Riddle, R. D., Johnson, R. L., Laufer, E., Tabin, C. (1993) Sonic hedgehog mediates the polarizing activity of the ZPA. *Cell*, *75*, 1401-1416.

Ristevski, S., Tam, P.P., Hertzog, P.J., and Kola, I. (2002). Ets2 is expressed during morphogenesis of the somite and limb in the mouse embryo. *Mech. Dev.*, *116*, 165–168.

Ringrose, L., & Paro, R. (2007). Polycomb/Trithorax response elements and epigenetic memory of cell identity. *Development*, *134*(2), 223-32.

Rinn, J., Kertesz, M., Wang, J., Squazzo, S., Xu, X., Samantha A. Brugmann, S. A., Goodnough, L. H., Helms, J. A., Farnham, P. J., Eran Segal, E. & Chang, H. Y. (2007). Functional Demarcation of Active and Silent Chromatin Domains in Human HOX Loci by Noncoding RNAs. *Cell*, *129*, 1311-1323.

Ristevski, S., Leary, D. A. O., Thornell, A. P., Owen, M. J., Kola, I., & Hertzog, P. J. (2004). The ETS Transcription Factor GABPα Is Essential for Early Embryogenesis. *Molecular and Cellular Biology*, *24*(13), 5844-5849.

Robert,B. & Y.Lallemand. (2006) Anteroposterior patterning in the limb and digit specification: contribution of mouse genetics. *Dev. Dyn*,. *235*, 2337-2352.

Robertson, A., Bilenky, M., Tam, A., & Zhao, Y., et al. (2008). Genome-wide relationship between histone H3 lysine 4 mono-and tri-methylation and transcription factor binding. *Genome Research*, *18*, 1906-1917.

Rodriguez, P., Bonte, E., Krijgsveld, J., Kolodziej, K. E., Guyot, B., Heck, A. J. R., Vyas, P., et al. (2005). GATA-1 forms distinct activating and repressive complexes in erythroid cells. *The EMBO journal*, *24*(13), 2354-66.

Roguev, A., Schaft, D., Shevchenko, A., Pijnappel, W. W. W. P., Wilm, M., Aasland, R. & Stewart, A. F. (2001). The Saccharomyces cerevisiae Set1 complex includes an Ash2 homologue and methylates histone 3 lysine 4. *The EMBO*, *20*(24), 3-9.

Roh, T., Cuddapah, S., Cui, K., & Zhao, K. (2006). The genomic landscape of histone modifications in human T cells. *PNAS*, *103*(43), 15782-15787.

Ruest, L. B., Dager, M., Yanagisawa, H., Charité, J., Hammer, R. E., Olson, E. N., Yanagisawa, M., Clouthier, D. E. (2003) dHAND-Cre transgenic mice reveal specific potential functions of dHAND during craniofacial development. *Dev. Biol.*, *257*, 263-277.

Sachs, R. K., van den Engh, G., Trask, B., Yokota, H., & Hearst, J. E. (1995). A random-walk/giant-loop model for interphase chromosomes. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(7), 2710-4.

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*(10), 1665-74.

Sagai T, Masuya H, Tamura M, Shimizu K, Yada Y, Wakana S, Gondo Y, Noda T, Shiroishi T. (2004) Phylogenetic conservation of a limb-specific, cis-acting regulator of Sonic hedgehog (Shh). *Mamm Genome*. *15*(1), 23-34.

Sagai, T., Hosoya, M., Mizushina, Y., Tamura, M., & Shiroishi, T. (2005). Elimination of a long-range cis-regulatory module causes complete loss of limb-specific Shh expression and truncation of the mouse limb. *Development*, *132*(4), 797-803.

Santangelo, A. M., de Souza, F. S. J., Franchini, L. F., Bumaschny, V. F., Low, M. J., Rubinstein, M. (2007). Ancient exaptation of a CORE-SINE retrotransposon into a highly conserved mammalian neuronal enhancer of the proopiomelanocortin gene. *PLoS Genetics.*, *3*, 1813-1826.

Sasak H, Yamaoka T, Ohuchi H, Yasue A, Nohno T, Kawano H, Kato S, Itakura M, Nagayama M, Noji S. (2002). Identification of cis-elements regulating expression of Fgf10 during limb development. *Int J Dev Biol*., *46*(7), 963-7.

Sasaki, T., Nishihara, H., Hirakawa, M., Fujimura, K., Tanaka, M., Kokubo, N., Kimura-yoshida, C., et al. (2008). Possible involvement of SINEs in mammalian-specific brain formation. *PNAS*, *105*(11), 4220-4225.

Saunders, J. W., Gasseling, M. T. (1968) Ectodermal-mesenchymal interactions in the origin of limb symmetry. As read in Hill, R. E. (2007) How to make a zone of polarizing activity: Insights into limb development via the abnormality preaxial polydactyly. *Develop. Growth Differ*., *49*, 439-448

Schachterle, W., Rojas, A., Xu, S.-M., & Black, B. L. (2012). ETS-dependent regulation of a distal Gata4 cardiac enhancer. *Developmental biology*, *361*(2), 439-49.

Schneider, I., Aneas, I., & Gehrke, A., Dahn, R. D., Nobregab, M. A. & Shubin, N. H. (2011). Appendage expression driven by the Hoxd Global Control Region is an ancient gnathostome feature. *PNAS*, *108*, 12782-12786.

Schoeftner, S., Sengupta, A., Kubicek, S., Mechtler, K., Spahn, L., Koseki, H., Jenuwein, T., et al. (2006). Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing. *The EMBO*, *25*(13), 3110-3122.

Schreiber, et al. (1989) *NAR*, *17* (5), 6419.

Schuettengruber, B., & Cavalli, G. (2009). Recruitment of polycomb group complexes and their role in the dynamic regulation of cell fate choice. *Development*, *136*(21), 3531-42.

Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B., & Cavalli, G. (2007). Genome regulation by polycomb and trithorax proteins. *Cell*, *128*(4), 735-45.

Schwartz, Y. B., Kahn, T. G., Nix, D. A., Li, X., Bourgon, R., Biggin, M. & Pirrotta, V. (2006) Genome-wide analysis of Polycomb targets in *Drosophila melanogaster* . *Nature Genetics*, *38*, 700-705.

Semenza, G. L., Delgrosso, K., Poncz, M., Malladi, P., Schwartz, E., & Surrey, S. (1984). The silent carrier allele: beta thalassemia without a mutation in the beta-globin gene or its immediate flanking regions. *Cell*, *39*(1), 123-8.

Semenza, G., Nejfelt, M., Chi, S., & Antonarakis, S. E. (1991). Hypoxia-inducible nuclear factors bind to an enhancer element located 3'to the human erythropoietin gene. *PNAS*, *88*, 5680-5684.

Semerci, C.N., Demirkan, F., Ozdemir, M., Biskin, E., Akin, B., Bagci, H., & Akarsu, N.A. (2009). Homozygous feature of isolated triphalangeal thumbpreaxial polydactyly linked to 7q36: no phenotypic difference between homozygotes and heterozygotes. *Clin. Genet.*, *76*, 85–90.

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., et al. (2012). Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell*, *148*(3), 458-72.

Sharrocks, A.D. (2001). The ETS-domain transcription factor family. *Nat. Rev. Mol. Cell Biol.*, *2*, 827–837.

Shen, X., Liu, Y., Hsu, Y., Fujiwara, Y., Kim, J., Mao, X., Yuan, G., et al. (2008). EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency. *Molecular cell*, *32*(4), 491-502.

Shi, Y., & Whetstine, J. R. (2007). Dynamic regulation of histone lysine methylation by demethylases. *Molecular cell*, *25*(1), 1-14.

Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., de Wit, E., van Steensel, B., et al. (2006). Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nature genetics*, *38*(11), 1348-54.

Sing, A., Pannell, D., Karaiskakis, A., Sturgeon, K., Djabali, M., Ellis, J., Lipshitz, H. D., et al. (2009). A vertebrate Polycomb response element governs segmentation of the posterior hindbrain. *Cell*, *138*(5), 885-97.

Smyth, G. K., Michaud, J. and Scott, H. S. (2005). Use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics, 21*, 2067-2075.

Soshnikova, N., & Duboule, D. (2009). Epigenetic temporal control of mouse Hox genes in vivo. *Science*, *324*(5932), 1320-3.

Santangelo, A. M., Souza, S. J. D., Franchini, L. F., Viviana F. Bumaschny, V. F., Malcolm J. Low, M. J. & Rubinstein, M. (2007). Ancient Exaptation of a CORE-SINE Retroposon into a Highly Conserved Mammalian Neuronal Enhancer of the Proopiomelanocortin Gene, *Plos Genetics*, *3*(10), 1813-1826.

Spitz, F., Gonzalez, F., Duboule, D. (2003). A Global Control Region Defines a Chromosomal Regulatory Landscape Containing the HoxD Cluster. *Cell*, *113*, 405-417.

Spitz, F., Herkenne, C., Morris, M. A, & Duboule, D. (2005). Inversion-induced disruption of the Hoxd cluster leads to the partition of regulatory landscapes. *Nature genetics*, *37*(8), 889-93.

Spitz, F., Montavon, T., Monso-Hinard, C., Morris, M., Ventruto, M.-L., Antonarakis, S., Ventruto, V., Duboule, D. (2002). A t(2;8) balanced translocation with breakpoints near the human HOXD complex causes mesomelic dysplasia and vertebral defects. *Genomics*, *79*(4), 493-8.

Srinivasan, L., & Atchison, M. L. (2004). YY1 DNA binding and PcG recruitment requires CtBP. *Genes & development*, *18*(21), 2596-601.

Srivastava, D., Cserjesi, P. & Olsen, E. N. (1995) A subclass of bHLH proteins required for cardiac morphogenesis. *Science*, *270*, 1995-1999.

Stock, J., Giadrossi, S., Casanova, M., Brookes, E., Vidal, M., Koseki, H., Brockdorff, N., Amanda G. Fisher, A. G. & Pombo, A. (2007). Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. *Nature Cell Biology*, *9*(12), 1428-1435.

Sun, X., Mariani, F. V., Martin, G. R. (2002) Functions of FGF signalling from the apical ectodermal ridge in limb development. *Nature*, *418*, 501-508

Surface, L. E., Thornton, S. R., & Boyer, L. A. (2010). Polycomb group proteins set the stage for early lineage commitment. *Cell stem cell*, *7*(3), 288-98.

Tarchini, B., & Duboule, D. (2006). Control of Hoxd Genes' Collinearity during Early Limb Development. *Developmental Cell*, *10*, 93-103.

Tarchini, B., Duboule, D., Kmita, M. (2006) Regulatory constraints in the evolution of the tetrapod limb anterior-posterior polarity. *Nature*, *443*, 985-988

Tavares, L., Dimitrova, E., Oxley, D., Webster, J., Poot, R., Demmers, J., Bezstarosti, K., et al. (2012). RYBP-PRC1 Complexes Mediate H2A Ubiquitylation at Polycomb Target Sites Independently of PRC2 and H3K27me3. *Cell*, *148*, 664-678.

te Welscher P, Zuniga A, Kuijper S, Drenth T, Goedemans HJ, Meijlink F, Zeller R. (2002) Progression of vertebrate limb development through SHH-mediated counteraction of GLI3. *Science*, *298*, 827-30.

Tickle, C. (2006). Making digit patterns in the vertebrate limb. *Nat. Rev. Mol. Cell Biol. 7*, 45–53.

Tie, F., Banerjee, R., Stratton, C., Prasad-Sinha, J., Stepanik, V., Andrei, Z., Diaz, M. O., et al. (2009). CBP-mediated acetylation of histone H3 lysine 27 antagonizes Drosophila Polycomb silencing. *Development*, *3141*, 3131-3141.

Toedling, J., & Huber, W. (2008). Analyzing ChIP-chip data using bioconductor. *PLoS computational biology*, *4*(11), e1000227.

Tolhuis, B., Blom, M., Kerkhoven, R. M., Pagie, L., Teunissen, H., Nieuwland, M., Simonis, M., et al. (2011). Interactions among Polycomb domains are guided by chromosome architecture. *PLoS genetics*, *7*(3), e1001343.

Towers, M., Mahood, R., Yin, Y., and Tickle, C. (2008). Integration of growth and specification in chick wing digit-patterning. *Nature*, *452*, 882–886.

Towers, M., & Tickle, C. (2009). Growing models of vertebrate limb development. *Development*, *136*, 179–190.

Tschopp, P. and Duboule, D. (2011). A regulatory "landscape effect" over the HoxD cluster. *Dev. Biol.*, *351*, 288-296.

Tusher, V.G., Tibshirani, R., & Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. USA*, *98*, 5116–5121.

Tuupanen, S., Turunen, M., Lehtonen, R., Hallikas, O., Vanharanta, S., et al. (2009) The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signalling. *Nature Genetics*, *41*, 885-890.

Vakoc, C. R., Letting, D. L., Gheldof, N., Sawado, T., Bender, M. a, Groudine, M., Weiss, M. J., et al. (2005). Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. *Molecular cell*, *17*(3), 453-62.

Valouev A, Johnson DS, Sundquist A, Medina C, Anton E, et al. 2008. Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data. *Nat. Methods*, *5*, 829–34

van den Engh, G., Sachs, R., & Trask, B. J. (1992). Estimating genomic distance from DNA sequence location in cell nuclei by a random walk model. *Science*, *257*(5075), 1410-2.

Vernimmen, D., Lynch, M., Gobbi, M. D., Garrick, D., Sharpe, J. A., Sloane-Stanley, J. A., Andrew J.H. Smith, A. J. H. & Higgs, D. R. (2011). Polycomb eviction as a new distant enhancer function. *Genes & Development*, *25*, 1583-1588.

Vickerman, L., Neufeld, S., & Cobb, J. (2011). Shox2 function couples neural, muscular and skeletal development in the proximal forelimb. *Developmental biology*, *350*(2), 323-36.

Visel, A., Blow, M. J., Li, Z., Zhang, T., Akiyama, J. A., Holt, A., Plajzar-Frick, I., et al. (2009) ChIP-seq accurately predicts tissue-specific activity of enhancers, *Nature,457*, 854-58.

Visser, M., Kayser, M., & Palstra, R.-J. (2012). HERC2 rs12913832 modulates human pigmentation by attenuating chromatin-loop formation between a long-range enhancer and the OCA2 promoter. *Genome research*, *22*(3), 446-55.

Vokes, S. A, Ji, H., Wong, W. H., & McMahon, A. P. (2008). A genome-scale analysis of the cis-regulatory circuitry underlying sonic hedgehog-mediated patterning of the mammalian limb. *Genes & development*, *22*(19), 2651-63.

Volpi, E. V., Chevret, E., Jones, T., Vatcheva, R., Williamson, J., Beck, S., Campbell, R. D., et al. (2000). Large-scale chromatin organization of the major histocompatibility complex and other regions of human chromosome 6 and its response to interferon in interphase nuclei. *Journal of cell science*, *113*, 1565-76.

Voncken, J., Roelen, B., Roefs, M., de Vries, S., Verhoeven, E., Marino, S., Deschamps, J., et al. (2003). Rnf2 (Ring1b) deficiency causes gastrulation arrest and cell cycle inhibition. *PNAS*, *100*, 2468-2473.

Wakimoto, B. T., & Hearn, M. G. (1990). The effects of chromosome rearrangements on the expression of heterochromatic genes in chromosome 2L of Drosophila melanogaster. *Genetics*, *125*, 141-154.

Wang, H., Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R., & Zhang, Y. (2004). Role of histone H2A ubiquitination in Polycomb silencing. *Nature*, *431*, 873-878.

Wang, J., Kumar, R. M., Biggs, V. J., Lee, H., Chen, Y., Kagey, M. H., Richard A. Young, R. A. & Abate-Shen, C. (2011). The Msx1 Homeoprotein Recruits Polycomb to the Nuclear Periphery during Development. *Developmental Cell*, *21*, 1-14.

Wang, L., Brown, J. L., Cao, R., Zhang, Y., Kassis, J. A., Jones, R. S., Hill, C., et al. (2004). Hierarchical Recruitment of Polycomb Group Silencing Complexes. *Molecular cell*, *14*, 637-646.

Wang, P., Lin, C., Smith, E. R., Guo, H., Sanderson, B. W., Wu, M., Gogol, M., et al. (2009). Global Analysis of H3K4 Methylation Defines MLL Family Member Targets and Points to a Role for

MLL1-Mediated H3K4 Methylation in the Regulation of Transcriptional Initiation by RNA Polymerase II. *Molecular and cellular biology*, *29*(22), 6074-6085.

Wang, Z., Zang, C., Cui, K., Schones, D. E., Barski, A., Peng, W., & Zhao, K. (2009). Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. *Cell*, *138*(5), 1019-31.

Wang, Z., Zang, C., Rosenfeld, J. a, Schones, D. E., Barski, A., Cuddapah, S., Cui, K., et al. (2008). Combinatorial patterns of histone acetylations and methylations in the human genome. *Nature genetics*, *40*(7), 897-903.

Wagner, A. (1999). Genes regulated cooperatively by one or more transcription factors and their identification in whole eukaryotic genomes. *Bioinformatics*, *15*, 776–784.

Wasserman, N. F., Aneas, I., Nobrega, M. A. (2010). An 8q24 gene desert variant associated with prostrate cancer risk confers differential *in vivo* activity to a *MYC* enhancer. *Genome Research*, *20*, 1191-1197.

Wederell, E., Bilenky, M., Cullum, Thiessen, R. N., Dagpinar, M., et al. (2008). Global analysis of in vivo Foxa2-binding sites in mouse adult liver using massively parallel sequencing. *Nucleic acids*, *36*(14), 4549-4564.

Wei, G., Srinivasan, R., Cantemir-Stone, C. Z., Sharma, S. M., Santhanam, R., Weinstein, M., Muthusamy, N., et al. (2009). Ets1 and Ets2 are required for endothelial cell survival during embryonic angiogenesis. *Blood*, *114*(5), 1123-30.

Werner, T., Haqmmer, A., Wahlbuhl, M., Bosl, M. R. & Wegner, M. (2007) Multiple conserved regulatory elements with overlapping functions determine Sox10 expression in mouse embryogenesis, *Nucleic Acids Research*, *35*, 6526-6538

Whyte, W. a, Bilodeau, S., Orlando, D. a, Hoke, H. a, Frampton, G. M., Foster, C. T., Cowley, S. M., et al. (2012). Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. *Nature*, *482*(7384), 221-5.

Williamson, I., Hill, R. E. and Bickmore, W. A. (2011). Enhancers: from developmental genetics to the genetics of common human disease. *Dev. Cell*, *21*, 17-19.

Wit, E. D., & Greil, F. (2005). Genome-wide HP1 binding in Drosophila: developmental plasticity and genomic targeting signals. *Genome research*, *1*, 1265-1273.

Woolfe, A., Goodson, M., Goode, D., & Snell, P. (2004). Highly conserved non-coding sequences are associated with vertebrate development. *PLoS biology*, *3*(1).

Woo, C. J., Kharchenko, P. V., Daheron, L., Park, P. J., & Kingston, R. E. (2010). A region of the human HOXD cluster that confers polycomb-group responsiveness. *Cell*, *140*(1), 99-110.

Wood, A., Shukla, A., Schneider, J., Lee, J., Stanton, J. D., et al. (2007). Ctk complex-mediated regulation of histone methylation by COMPASS. *Molecular and Cellular Biology*, *27*(2), 709-720.

Wutz, A. (2011). Gene silencing in X-chromosome inactivation: advances in understanding facultative heterochromatin formation. *Nature reviews. Genetics*, *12*(8), 542-53.

Wyngaarden, L. A., Delgado-olguin, P., Su, I.-hsin, Bruneau, B. G., & Hopyan, S. (2011). Ezh2 regulates anteroposterior axis specification and proximodistal axis elongation in the developing limb. *Development*, *138*, 3759-3767.

Xi, H., Shulha, H. P., Lin, J. M., Vales, T. R., Fu, Y., Bodine, D. M., McKay, R. D. G., et al. (2007). Identification and characterization of cell type-specific and ubiquitous chromatin regulatory structures in the human genome. *PLoS genetics*, *3*(8), e136.

Xiang, Y., Zhu, Z., Han, G., Ye, X., Xu, B., Peng, Z., Ma, Y., Yu, Y., Lin, H., Chen, A. P. & Chen, C. D. (2007). JARID1B is a histone H3 lysine 4 demethylase up-regulated in prostate cancer. *PNAS*, *104*, 19226-19231.

Yaffe, E., & Tanay, A. (2011). Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture. *Nature genetics*, *43*(11), 1059-65.

Yamagishi, T., et al. (2007) *Evx-Hoxd13* intergenic region restricts enhancer association to the *Hoxd13* promoter. *PL^{O}S ONE*, *1*, e175, 1-8.

Yanagisawa, H., Clouthier, D. E., Richardson, J. A., Jeroen Charité, J. & Olson, E. N. (2003). Targeted deletion of a branchial arch-specific enhancer reveals a role of dHAND in craniofacial development. *Development*, *130*(6), 1069-1078.

Yang, Y., Drossopoulou, G., Chuang, P.T., Duprez, D., Marti, E., Bumcrot, D., Vargesson, N., Clarke, J., Niswander, L., McMahon, A., and Tickle, C. (1997). Relationship between dose, distance and time in Sonic Hedgehog-mediated regulation of anteroposterior polarity in the chick limb. *Development*, *124*, 4393–4404.

Yokota, H., Singer, M. J., Engh, G. J. V. D., & Trask, B. J. (1997). Regional differences in the compaction of chromatin in human G 0 /G 1 interphase nuclei. *Genome Research*, *5*, 157-166.

Yuan, G., Liu, JS. (2008). Genomic sequence is highly predictive of local nucleosome depletion. *PLoS computational biology*, *4*(1), 0164-0174.

Yusufzai, T. M., Tagami, H., Nakatani, Y., & Felsenfeld, G. (2004). CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species. *Molecular cell*, *13*(2), 291-8.

Zakany, J. and Duboule, D. (2007). The role of Hox genes during vertebrate limb development. *Curr. Opin. Genet. Dev.*, *17*, 359-366.

Zakany, J., Kmita, M., Duboule, D. (2004) A dual role for Hox genes in limb anteriorposterior asymmetry. *Science*, *304*, 1669-1672

Zentner, G. E., Tesar, P. J., & Scacheri, P. C. (2011). Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. *Genome research*, *21*(8), 1273-83.

Zhang, Z., Verheyden, J.M., Hassell, J.A., and Sun, X. (2009). FGF-regulated Etv genes are essential for repressing Shh expression in mouse limb buds. *Dev. Cell*, *16*, 607–613.

Zhao, J., Ohsumi, T. K., Kung, J. T., Ogawa, Y., Grau, D. J., Sarma, K., Song, J. J., et al. (2010). Genome-wide identification of polycomb-associated RNAs by RIP-seq. *Molecular cell*, *40*(6), 939-53.

Zhao, J., Sun, B. K., Erwin, J. a, Song, J.-J., & Lee, J. T. (2008). Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science*, *322*(5902), 750-6.

Zhou, V. W., Goren, A., & Bernstein, B. E. (2010). Charting histone modifications and the functional organization of mammalian genomes. *Nature reviews. Genetics*, *12*(1), 7-18.

Zhu, J., Nakamura, E., Nguyen, M.T., Bao, X., Akiyama, H., and Mackem, S. (2008). Uncoupling Sonic hedgehog control of pattern and expansion of the developing limb bud. *Dev. Cell*, *14*, 624–632.

Zuniga, A., Michos, O., Spitz, F., Haramis, A.-P. G., Panman, L., Galli, A., Vintersten, K., et al. (2004). Mouse limb deformity mutations disrupt a global control region within the large regulatory landscape required for Gremlin expression. *Genes & development*, *18*(13), 1553-64.

Zuniga, 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

Zúñiga, a, & Zeller, R. (1999). Gli3 (Xt) and formin (ld) participate in the positioning of the polarising region and control of posterior limb-bud identity. *Development*, *126*(1), 13-21.

Appendix A

Genomic loci covered by the tiling array

Genomic regions covered by the NimbleGen tiling array. Coordinates taken from the Mouse NCBI37 (mm9) build (UCSC Genome Browser July 2007).

Locus	Base pairs	Coordinates
HoxA	818,585	chr6:51,781,416-52,600,000
HoxB	755,001	chr11:95,845,224-96,600,224
HoxC	702,398	chr15:102,400,000-103,102,397
HoxD	668,990	chr2:74,241,352-74,910,341
Shh	1,221,001	chr5:28,489,000-29,710,000
Hand2	900,000	chr8:59,050,000-59,951,000
Grem1	990,000	chr2:112,862,000-113,856,000
Fgf10	115,623	chr13:119,484,571-119,600,193
Hba	237,547	chr11:32,087,454-32,325,000
Hbb	197,639	chr7:110,861,362-111,059,000
Kcnq1	394,001	chr7:150,256,000-150,650,000
Dlk1	136, 468	chr12:110,693,533-110,830,000
Gnas	69,619	chr2:174,105,382-174,175,000
H19	137,587	chr7:149,730,414-149,868,000
Mest	225,001	chr6:30,685,000-30,910,000
Pax6	35,000	chr2:105,505,001-105,540,000
Brachyury	150, 001	chr17:8,554,691-8,704,691
Gsc	50, 001	chr12:105,687,160-105,737,160
Hhex	40,001	chr19:37,489,346-37,529,346
Nestin	28,077	chr3:87,765,656-87,793,732
Lin28	27,001	chr4:133,553,000-133,580,000
Pou5f1	9,401	chr17:35,640,600-35,650,000
Sox2	23,606	chr3:34,540,712-34,564,317
Nanog	10,001	chr6:122,656,000-122,666,000
Klf4	13,200	chr4:55,536,078-55,549,277
c-myc	15,063	chr15:61,811,875-61,826,937
mirlet7-d	14,164	chr13:48,630,830-48,644,993
Gata1	17,112	chrX:7,532,108-7,549,219
Gata4	13,904	chr14:63,855,878-63,869,781
Cdx2	12,700	chr5:148,110,571-148,123,270
Asl1	6,000	chr10:86,953,114-86,959,113
Nkx2-9	6,000	chr12:57,709,722-57,715,721
Myf5	6,262	chr10:106,919,672-106,925,933
Olig2	7,038	chr16:91,223,137-91,230,174
Magea3	8,801	chrX:151,379,200-151,388,000
Tex19.1	8,770	chr11:121,004,657-121,013,426
Dnmt3b	16,799	chr2:153,470,702-153,487,500
Rpl19	6,164	chr11:97,886,000-97,892,163
Actb	5,414	chr5:143,663,892-143,669,305
Aard	8,426	chr15:51,869,406-51,877,831
Adssl1	31,878	chr12:113,852,987-113,884,864
Ankhd1	6,329	chr18:36,715,293-36,721,621
Aurkb	9,030	chr11:68,856,737-68,865,766
Brd3	41,055	chr2:27,296,495:27,337,549
Chchd7	10,373	chr4:3,862,488-3,872,860
Chmp4b	56,637	chr2:154,473,323-154,529,959
Crabp1	12,453	chr9:54,610,540-54,622,992
Ctnnal1	81,381	chr4:56,810,244-56,891,624
Ddx6	48,314	chr9:44,404,928-44,453,241
Emilin2	88,185	chr17:71,586,819-71,675,003
Eno1	8,210	chr4:149,609,995-149,618,204
Fus	21,830	chr7:135,107,354-135,129,183
Gypa	25,112	chr8:83,013,758-83,038,869
Hebp1	46,046	chr6:135,079,862-135,125,907

Hemgn	15,293	chr4:46,404,312-46,419,604
Hnrpa1	8,303	chr15:103,068,649-103,076,951
Jarid1b	8,076	chr1:136,453,744-136,461,819
Ly6a	6,777	chr15:74,823,425-74,830,201
Ly6c1	5,724	chr15:74,874,491-74,880,214
Mapk1ip1l	37,167	chr14:47,911,795-47,948,961
Mlkl	39,156	chr8:113,829,174-113,868,329
Npm3	5,490	chr19:45,820,805-45,826,294
Nrn1	13,244	chr13:36,815,287-36,828,530
Pcdha	1,012,030	chr18:36,996,988-38,009,017
Pdlim2	20,183	chr14:70,560,661-70,580,843
Ppm1b	5,838	chr17:85,352,959-85,358,796
Ppp1cc	5,664	chr5:122,604,605-122,610,268
Pramel4	15,290	chr4:143,646,480-143,661,769
Prf1	9,716	chr10:60,757,991-60,767,706
Psip1	36,768	chr4:83,097,979-83,134,746
Shmt2	11,919	chr10:126,950,487-126,962,405
Sif1	13,637	chr16:18,262,780-18,276,416
Slc7a3	9,198	chrX:98,273,946-98,283,143
Snx17	12,596	chr5:31,492,177-31,504,772
Srebf2	86,538	chr15:81,963,276-82,049,813
Tcfap2a	27,278	chr13:40,807,558-40,834,835
Tle1	4,134	chr4:71,858,971-71,863,104
Tnik	6,790	chr3:28,159,250-28,166,039
Xlr4a	12,245	chrX:70,317,643-70,329,887
Zbtb2	31,535	chr10:5,953,177-5,984,711
Xist	34,292	chrX:100,649,996-100,684,287

Appendix B

Published papers

Development 139, 3157-3167 (2012) doi:10.1242/dev.081174

Anterior-posterior differences in HoxD chromatin topology in limb development

Iain Williamson, Ragnhild Eskeland*, Laura A. Lettice, Alison E. Hill, Shelagh Boyle, Graeme R. Grimes, Robert E. Hill‡ and Wendy A. Bickmore‡

SUMMARY

A late phase of HoxD activation is crucial for the patterning and growth of distal structures across the anterior-posterior (A-P) limb axis of mammals. Polycomb complexes and chromatin compaction have been shown to regulate Hox loci along the main body axis in embryonic development, but the extent to which they have a role in limb-specific HoxD expression, an evolutionary adaptation defined by the activity of distal enhancer elements that drive expression of 5′ Hoxd genes, has yet to be fully elucidated. We reveal two levels of chromatin topology that differentiate distal limb A-P HoxD activity. Using both immortalised cell lines derived from posterior and anterior regions of distal E10.5 mouse limb buds, and analysis in E10.5 dissected limb buds themselves, we show that there is a loss of polycomb-catalysed H3K27me3 histone modification and a chromatin decompaction over HoxD in the distal posterior limb compared with anterior. Moreover, we show that the global control region (GCR) long-range enhancer spatially colocalises with the 5′ HoxD genomic region specifically in the distal posterior limb. This is consistent with the formation of a chromatin loop between 5′ HoxD and the GCR regulatory module at the time and place of distal limb bud development when the GCR participates in initiating Hoxd gene quantitative collinearity and *Hoxd13* expression. This is the first example of A-P differences in chromatin compaction and chromatin looping in the development of the mammalian secondary body axis (limb).

KEY WORDS: Autopod, Enhancer, Polycomb, Mouse

INTRODUCTION

Regulated Hox gene expression is important in anterior-posterior (A-P) patterning of the primary embryonic axis, with Hox genes being first activated at gastrulation (Wyngaarden et al., 2011; Deschamps and van Nes, 2005). The HoxD cluster has also been co-opted more recently in evolution into regulating the growth and patterning of the limb and digits.

Polycomb repressive complexes (PRC1 and PRC2) are required to maintain Hox genes in a silent compact chromatin state in embryonic stem (ES) cells (Boyer et al., 2006; Lee et al., 2006; Endoh et al., 2008; Stock et al., 2007; Eskeland et al., 2010). Whereas roles of polycomb at Hox loci are well established in early embryonic development and in differentiation along the primary embryonic axis (Voncken et al., 2003; Faust et al., 1998; Chambeyron et al., 2005; Soshnikova and Duboule, 2009), whether polycomb-mediated chromatin changes are involved in Hox regulation in the secondary body axis is unclear.

Two phases of Hoxd expression are important in limb development and patterning (Tarchini and Duboule, 2006; Zakany and Duboule, 2007). The first phase results in expression of 3′ Hoxd genes (*Hoxd1-9*) earlier than the 5′ genes. This restricts 5′ HoxD expression to the posterior side of the distal limb bud and is required for limb outgrowth, proximal limb development, limb A-P polarity and the posterior expression of sonic hedgehog (*Shh*). This 3′-5′ temporal and spatial collinearity is reminiscent of regulation in the main embryonic axis, which is accompanied by progressive loss of histone H3 lysine 27 tri-methylation (H3K27me3) catalysed by PRC2 (Soshnikova and Duboule, 2009). However, it might also require as yet undefined regulatory elements 3′ of HoxD (Spitz et al., 2005).

A later phase [embryonic day (E) 10.5] of Hoxd expression in the distal limb is required for digit morphogenesis (Spitz et al., 2003). This is characterised by ‘quantitative collinearity’ in which expression of the most 5′ gene, *Hoxd13*, is initially strongest in the posterior distal mesenchyme, with progressively less strong expression of *Hoxd12* to *Hoxd10*. The *Hoxd13* expression domain then spreads anteriorly, so that by E12.5 it is the only Hoxd gene for which expression is robust enough to be detectable on the most anterior side (Montavon et al., 2008). This is driven by enhancer elements including a ~40 kb global control region (GCR) located 180 kb 5′ (centromeric) of *Hoxd13* beyond *Evx2* and *Lnp*, and the Prox enhancer located between *Evx2* and *Lnp* (Fig. 4A) (Spitz et al., 2003; Tschopp and Duboule, 2011). It is not clear whether late phase HoxD regulation involves polycomb-mediated repression.

Here, we use immortalised mesenchymal cells derived from either the anterior (A) or the posterior (P) distal limb of E10.5 embryos, which show high levels of *Hoxd13* expression in the posterior-derived cells, to show that the ancestral role of polycomb in regulating the HoxD cluster appears to be maintained during distal limb development. There is a loss of H3K27me3 and a decompaction of higher order chromatin structure over HoxD in the distal posterior, compared with the anterior, limb cells. This is confirmed by analysis in dissected limb buds. Furthermore, we show a spatial colocalisation of the GCR and 5′ HoxD that is restricted to the distal posterior limb, consistent with the notion of physical association between this enhancer and its target genes. This is the first demonstration of differential chromatin compaction and enhancer-gene colocalisation across the A-P axis of the developing limb.

MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Crewe Road, Edinburgh EH4 2XU, UK.

*Present address: Department of Molecular Biosciences, University of Oslo, N-0316 Oslo, Norway
‡Authors for correspondence (Bob.Hill@igmm.ed.ac.uk; Wendy.Bickmore@igmm.ed.ac.uk)

Accepted 8 June 2012

MATERIALS AND METHODS

Immortomouse cell lines

The posterior third and the anterior two thirds of distal forelimb buds were dissected from E10.5 mouse embryos from an Immortomouse (H-2kb-tsA58) × CD1 cross (Jat et al., 1991). After washing in PBS, limb buds were treated using trypsin (0.2 g/l)/Versene for 15-20 minutes, and dispersed gently. Cells were plated into 6-well plates in DMEM, 10% foetal calf serum, 20 ng/ml γ-Interferon (Peprotech) and grown at 33°C, the permissive temperature for the temperature-sensitive T antigen. A1/P1 and A2/P2 cell lines pairs were derived from separate litters.

RNA expression

RNA was prepared using Trizol reagent (Invitrogen) according to the manufacturer's protocol [for limb buds, dissected anterior and posterior tissue was dissociated into single cells in Trizol using a syringe with a 25G (0.5 mm) needle (BD Microlance)], followed by phenol:chloroform extraction and digestion with 2U DNaseI (Ambion) for 30 minutes at 37°C. cDNA was made using a First Strand Synthesis Kit (Roche) and amplified by PCR or real-time qRT-PCR using specific primers (supplementary material Table S1).

For expression arrays, 400 ng of RNA from A2/P2 cells were amplified using the Illumina TotalPrep RNA Amplification Kit (Ambion). Amplified, biotinylated cRNAs (1.5 μg) were labelled and hybridised to Illumina MouseRef6 Gene Expression beadchip arrays [Gene Expression Omnibus (GEO) accession platform number GPL6887, expression data GSE38370]. Data were analysed in R using limma (Smyth et al., 2005) and beadarray (Dunning et al., 2007) bioconductor packages. Probes with detection *P*-values less than 0.01 were removed from further analysis. Signals were quantile normalised to remove technical variation, and differential expression was assessed using limma's lmFit, eBayes and topTable function. *P*-values were corrected for multiple testing (Benjamini and Hochberg, 1995).

The 5944 genes with significantly differential A:P expression (adjusted $P<0.05$) were ranked by *P*-value and searched for enriched gene ontology (GO) terms using the Gene Ontology enrichment analysis and visualisation tool (GOrilla) (Eden et al., 2009). To avoid false positives $P<10^{-7}$ was set as the cut off for terms to be considered, based on the Bonferroni correction.

Acid extraction of histones and western blot analysis

Nuclei were isolated from 3-6×10^6 cells and histones acid-extracted and analysed as described (Eskeland et al., 2010).

Chromatin immunoprecipitation

Native chromatin immunoprecipitation (nChIP) from cell lines was performed as previously described (Eskeland et al., 2010), except that 40-46 Boehringer units of MNase (Worthington) was used to digest 1-6×10^8 cells.

For tissue, distal anterior and posterior forelimb buds were dissected from 50-55 E10.5 embryos. Owing to the lower cell numbers (~1×10^6), nChIP was performed with the following modifications: cells were digested with 8-9 Boehringer units of MNase (Worthington). Released chromatin (10-30 μg) was incubated with 3-5 μg prebound (to Protein A or G magnetic beads, Invitrogen) H3K27me3 antibody (Millipore) in the presence of 25 μg BSA for 3 hours at 4°C.

For Ring1B ChIP, 0.5-3×10^7 anterior and posterior limb tissue cells dissected from the distal forelimb buds of 70-75 E10.5 embryos were fixed with 1% formaldehyde (25°C, 10 minutes) and stopped with 0.125 M glycine. Sonication was as described (Eskeland et al., 2010). Released chromatin (30-50 μg) was incubated with 1 μg of prebound (Protein A magnetic beads, Invitrogen) Ring1B antibody (MBL, D139-3) or mouse IgG (Santa Cruz, sc-2025).

ChIP analysis

Relative quantification of ChIP by qPCR was as previously described (Eskeland et al., 2010) using primers described in supplementary material Table S2.

Input or ChIP DNA (10 ng) was amplified using the WGA2 Whole Genome Amplification Kit (Sigma) and labelled with Cy3 or Cy5 by random priming according to the NimbleGen ChIP-chip protocol (Roche). Two or three biological replicates, with dye swaps, were hybridised for 20 hours to a custom 3×720 K mouse tiling array (NimbleGen, Roche) containing 179,493 unique probes from different genomic regions, with each probe represented by four replicates (GEO accession platform number GPL14936, array data GSE38526). Arrays were washed according to the manufacturer's protocol and scanned on a NimbleGen MS 200 Microarray scanner (Roche) using 100% laser power and 2 μm resolution. Raw signal intensities were quantified from TIFF images using MS 200 Data Collection software.

Microarray data were analysed and ChIP enrichments determined as described by Pradeepa et al. (Pradeepa et al., 2012). To determine the significance of the difference in H3K27me3 enrichment over HoxD in anterior versus posterior limb tissue cells, the median log enrichment value of probes covering the locus or specific locus sub-regions (5′, 3′; promoters, genes, intergenic regions) were calculated. The statistical significance of any A:P differences were assessed using the non-parametric Mann-Whitney U-test. Mean log enrichment values were used to compare individual genes and promoter regions within HoxD, and adjacent genes and statistical significance was assessed using the two-sample Student's *t*-test.

Fluorescence in situ hybridisation (FISH)

2D FISH was as described previously (Chambeyron and Bickmore, 2004; Eskeland et al., 2010). Fosmid clones (supplementary material Table S3) were prepared and labelled as previously described (Morey et al., 2007). Between 80 and 120 ng of biotin- and digoxigenin-labelled probe were used per slide, with 8-12 μg of mouse Cot-1 DNA (Invitrogen) and 10 μg salmon sperm DNA.

For 3D FISH, E10.5-11 embryos from CD1 mice were collected, fixed, embedded, sectioned and processed as previously described (Morey et al., 2007), except that sections were cut at 6 μm.

Image analysis

For 2D FISH, slides were analysed as described (Morey et al., 2007) except that the Chroma #83000 triple band pass filter set (Chroma Technology Corporation, Rockingham, VT, USA) and a motorised filter wheel (Prior Scientific Instruments, Cambridge, UK) were used.

For 3D analysis of tissue sections, slides were imaged with a Hamamatsu Orca AG CCD camera (Hamamatsu Photonics, Welwyn Garden City, UK), Zeiss Axioplan II fluorescence microscope with Plan-neofluar or Plan apochromat objectives, a Lumen 200W metal halide light source (Prior Scientific Instruments) and Chroma #89014ET single excitation and emission filters (Chroma Technology Corporation) with the excitation and emission filters installed in Prior motorised filter wheels. A piezoelectrically driven objective mount (PIFOC model P-721, Physik Instrumente, Karlsruhe, Germany) was used to control movement in the *z* dimension. Hardware control, image capture and analysis were performed using Volocity (PerkinElmer, Waltham, MA, USA). Images were deconvolved using a calculated point spread function with the constrained iterative algorithm of Volocity. Image analysis was carried out using the Quantitation module.

RESULTS

Cell lines derived from the mesenchyme of E10.5 limb buds

Activity of limb enhancers in the HoxD GCR first becomes apparent in posterior mesenchyme cells of the distal forelimb bud at E10.5, followed later by anterior extension of this expression zone across the distal limb (Spitz et al., 2003). Therefore, we restricted our analysis to E10.5 limb buds to capture the initiation of this second wave of HoxD regulation.

Analysing chromatin structure in the small number of cells dissected from mouse embryos is challenging (Soshnikova and Duboule, 2009). However, conditionally immortalised cell lines

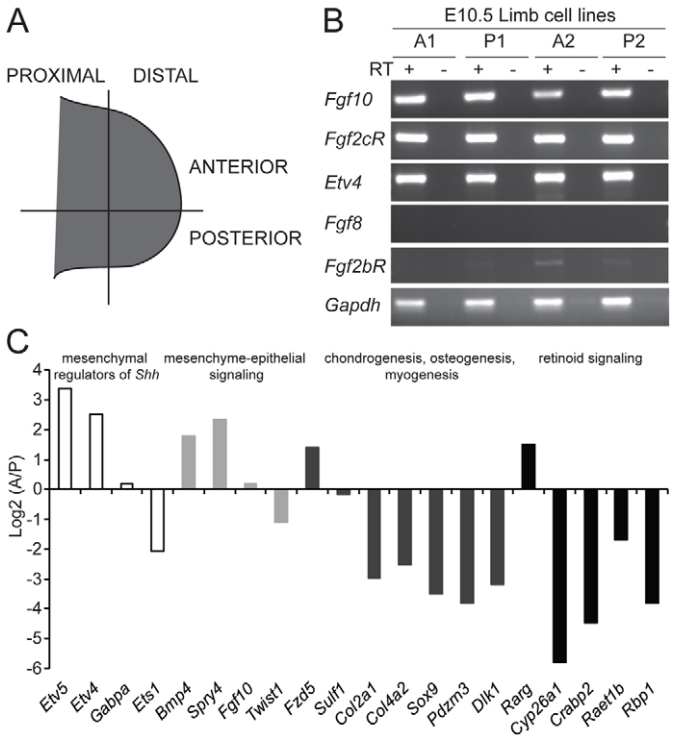

Fig. 1. Characterisation of cell lines from distal posterior and anterior mouse forelimb. (**A**) Schematic of E10.5 forelimb bud showing the dissection into anterior (A) and posterior (P) distal regions. (**B**) RT-PCR to detect the expression of mesenchymal (*Fgf10*, *Fgf2cR*, *Etv4*) and epithelial (*Fgf8*, *Fgf2cR*) markers in immortomouse cell lines derived from the anterior (A) or posterior (P). The two cell line pairs derive from embryos from different litters. Primer sequences are indicated in supplementary material Table S1. –, negative control lacking reverse transcriptase. (**C**) Log2 A/P expression from microarray analysis of A2/P2 cells for selected genes categorised according to their function.

can be derived from the transgenic 'immortomouse', which expresses temperature-sensitive SV40 T antigen (Jat et al., 1991). Such cell lines appear to retain many biological properties of the cells from which they were derived, including gene expression patterns and response to signalling pathways (Kohn et al., 2010). We derived two sets of cell lines from dissected E10.5 forelimb buds to represent the most posterior third of the distal limb (cell lines P1 and P2) or the anterior two-thirds (A1 and A2) (Fig. 1A).

The embryonic limb bud consists of two main cell types: mesenchyme and the surface ectoderm, which at the most distal margin forms the apical ectodermal ridge (AER). The morphology of the cell lines indicated that they were likely to be mesenchymal in origin and RT-PCR analysis confirmed this (Fig. 1B). *Fgf10* is expressed in limb bud mesenchyme and signals to the Fgfr2b receptor expressed in the AER to induce expression of *Fgf8* there. By contrast, *Fgfr2c* is expressed in both the mesoderm and ectoderm of the developing limb bud (Lizarraga et al., 1999; Duboc and Logan, 2011). The detection of mRNAs from *Fgf10* and *Fgfr2c*, but not *Fgf8* or *Fgfr2b* in immortomouse-derived limb bud cell lines thus indicates their derivation from the limb mesenchyme (Fig. 1B). The origin of these cells from the distal, rather than proximal, margin of the limb is supported by the expression of *Etv4* (Mao et al., 2009). Analysis by expression microarrays provided further evidence for this. Genes with proximally restricted expression domains, such as *Tbx15*, *Cart1* (*Alx1* – Mouse Genome Informatics), *Emx2*, *Pax1* (chondrogenesis), *Pax3* (myogenesis) (Kuijper et al., 2005), *Vcan*, *Ebf2* (neurogenesis) (Krawchuk and Kania, 2008) and *Shox* (chondrogenesis, myogenesis, neurogenesis) (Vickerman et al., 2011), could not be detected (data not shown). We conclude that the four cell lines retain at least some of the specific developmental functions expected of their limb origin (Robert and Lallemand, 2006; Hill, 2007).

Expression microarrays provided further insight into the differences between the anterior- and posterior-derived cell lines (A2 and P2) (GEO accession number GSE38370). For example, levels of *Etv4* and *Etv5* mRNAs (known to be expressed in the distal limb mesenchyme and important for the posterior restriction of Shh expression) (Mao et al., 2009) were higher in the anterior cells than in posterior cells, whereas expression of other ETS factors that define the spatial boundary of *Shh* were higher in the posterior cells (*Ets1*) or similar (*Gabpa*) throughout the distal limb bud (Lettice et al., 2012) (Fig. 1C). Genes involved in mesenchymal-epithelial signalling and distal limb patterning were upregulated in anterior (*Bmp4*, *Spry4*) and posterior (*Twist1*) cells or expressed evenly across the A-P axis (*Fgf10*). Among the genes with the highest ratio of expression in the posterior cells were those involved in retinoid signalling, and in chondrogenesis, osteogenesis or myogenesis (Fig. 1C).

Gene ontology analysis (GOrilla) indicated that the most significantly different A/P expression levels were for genes enriched for GO Biological Process terms ($P<10^{-9}$) such as A-P axis and pattern specification, anatomical structure morphogenesis and embryonic skeletal system morphogenesis – categories that reflect the cell lines' origins from mesenchymal tissue of the A-P axis of a developing appendage (supplementary material Fig. S1).

Cell lines reflect differential 5′ Hoxd gene expression in distal posterior limb bud

Spatial expression domains of Hoxd genes in E10.5 limb buds are well characterised (Spitz et al., 2003; Spitz et al., 2005; Tarchini and Duboule, 2006; Zakany and Duboule, 2007). We used RT-PCR and qRT-PCR on both sets of cell lines and on tissue dissected from E10.5 anterior and posterior forelimb buds to determine how relevant the limb cell lines are for analysis of HoxD spatial regulation.

Hoxd1 mRNA was not detected in any of the cell lines and only very faintly in anterior limb tissue (Fig. 2A). *Hoxd3* expression was detected at generally low levels (Fig. 2A) that are similar between the anterior and posterior limb tissue and the corresponding cell lines, except for even lower levels in P2 cells (Fig. 2B). There was no significant A/P difference in *Hoxd3* expression in either cell line pairs or limb tissue (Fig. 2C). Conversely, at the 5′ end of HoxD, *Hoxd13* expression levels were 17-fold higher in posterior limb and limb-derived cell lines than in anterior equivalents (Fig. 2C). A2/ P2 cell lines showed expression levels similar to those in anterior and posterior tissue, but levels in A1/P1 were both proportionally lower (Fig. 2B). We observed slightly higher *Hoxd12* expression in the posterior limb tissue than in the anterior, and higher *Hoxd11* in P1 compared with A1 cells, but their expression levels of were both too low for reliable quantification. *Hoxd10* expression showed a high A/P difference in both cell line pairs; the apparent large difference (Fig. 2C) measured in the A1/P1 cell lines is due to the very low expression levels in A1. *Hoxd8* expression showed an A/P differential (threefold higher in P2 than in A2) in the second cell line pair only due to its increased expression in P2 cells, expression

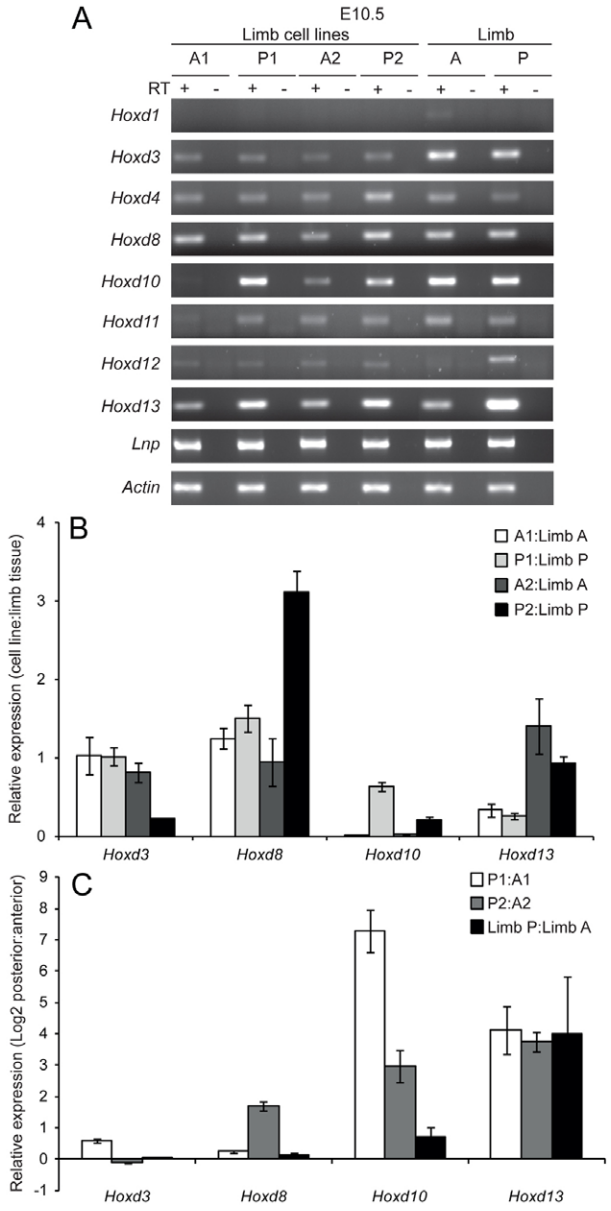

Fig. 2. Hoxd expression in distal anterior and posterior limb cell lines and mouse tissue. (**A**) RT-PCR to detect the expression of 3′ (*Hoxd1*, *Hoxd3*, *Hoxd4*, *Hoxd8*) and 5′ (*Hoxd10*, *Hoxd11*, *Hoxd12*, *Hoxd13*) Hoxd genes and the adjacent *Lnp* in both sets of cell lines (A1/P1, A2/P2) and in E10.5 distal forelimb tissue (A and P). Primer sequences are indicated in supplementary material Table S1. (**B**,**C**) Quantitative (q)RT-PCR to compare expression levels of 3′ (*Hoxd3*, *Hoxd8*) and 5′ (*Hoxd10*, *Hoxd13*) Hoxd genes in both sets of cell lines and in anterior or posterior distal forelimb tissue. (B) Expression in each cell line compared with the corresponding limb tissue. (C) Log2 P/A expression for both sets of cell lines (white and grey bars) and in distal forelimb tissue (black bars).

levels in A2, A1 and P1 being similar to limb tissue (Fig. 2B,C). No A/P difference in *Hoxd8* expression was detected in the E10.5 limb tissue. This suggests that both cell line pairs capture cells that have activated 5′ Hoxd genes specifically in the posterior compartment at the start of the second wave of HoxD activation in the distal limb and that this activation has a greater extension 3′ in the 2P cell line (to *Hoxd8*) than it does in 1P cells (*Hoxd10*). Gene activation extends more 3′ in both posterior-derived cell lines than in the limb tissue. This could be due to the outgrowth of cells that have more extensive gene activation than the average in the posterior limb or, more likely, that the regulatory mechanisms driving progressive 5′-3′ HoxD activation continue operating for some time after tissue dissection and during cell immortalisation.

Beyond *Hoxd13*, *Evx2* expression was not be detected in cell lines (data not shown) suggesting that the cells might originate from outside of the small *Evx2* expression domain at the extreme distal margin of the E10.5 forelimb bud (Tarchini and Duboule, 2006). *Shh* expression was also not detected, suggesting that it is not required for maintenance of the second wave of HoxD activation, at least in cell lines, as has been previously proposed (Harfe et al., 2004).

Although both immortomouse-derived distal limb cell line pairs show an A/P difference in 5′ HoxD activity, the A2/P2 pair show a more extensive domain of activation (*Hoxd13* to *Hoxd8*) and so were chosen for further study.

Loss of H3K27me3 over the HoxD cluster in posterior cells

The PRC2 polycomb complex is fundamental to regulation of Hox gene clusters during ES cell differentiation and early embryogenesis and is responsible for blanketing HoxD with H3K27me3 (Boyer et al., 2006; Lee et al., 2006). Polarised (3′-5′) loss of H3K27me3 accompanies 3′ Hoxd gene activation during ES cell differentiation (Eskeland et al., 2010) and in the tail bud of the primary axis during early embryogenesis (Soshnikova and Duboule, 2009). PRC2 function in the limb is required for cell survival and for proximodistal elongation of the limb (Wyngaarden et al., 2011), but whether differential polycomb-mediated chromatin changes are involved in regulating A-P Hoxd gene expression during the late phase of distal limb patterning is unclear.

Immunoblotting shows similar H3K27me3 levels globally in A2 and P2 cells (Fig. 3A). Using native ChIP (nChIP) combined with custom high-density tiling arrays covering multiple regions of the mouse genome, including Hox loci (Eskeland et al., 2010), we determined the H3K27me3 profile of the A2/P2 distal forelimb-derived cell lines (Fig. 3B) (GEO accession number GSE38526).

H3K27me3 is pervasive over HoxD in anterior cells, with both the 5′ and 3′ (*Hoxd1-d4*) genes densely covered (Fig. 3B). The dip in H3K27me3 between *Hoxd4* and *Hoxd8*, is similar to that seen in ES cells (Eskeland et al., 2010). Whereas H3K27me3 still blankets *Hoxd1-d3* in posterior (P2) cells, it is largely absent over the 5′ genes (Fig. 3B,C, upper two tracks). Beyond *Hoxd13*, H3K27me3 covers *Evx2* in both cell lines. In contrast to ES cells, in which both HoxB and HoxD loci are both blanketed (Eskeland et al., 2010), H3K27me3 is largely absent from HoxB in A2 and P2 cell lines, with only *Hoxb13* enriched, emphasising the particular regulation of HoxD across the A-P limb axis. As controls, the *Pax6* polycomb target was blanketed with H3K27me3 in both anterior and posterior, whereas H3K27me3 was absent from the widely expressed *Brd3*.

Of those arrayed probes significantly enriched in H3K27me3 in anterior cells (log2 H3K27me3/input≥1) (Fig. 3E, upper pie chart) only 8% are from HoxD. Yet 93% of probes with an at least twofold (log2≥1) A/P difference in H3K27me3 enrichment are from this locus and 88% are from 5′ HoxD (Fig. 3E, lower pie chart). By contrast, only 2% of probes with at least twofold A/P H3K27me3 difference are from HoxA. Therefore, extensive A/P differences in H3K27me3 are specific to HoxD in these cell lines. Quantitative PCR (qPCR) confirmed the lower H3K27me3 levels in P2 versus A2

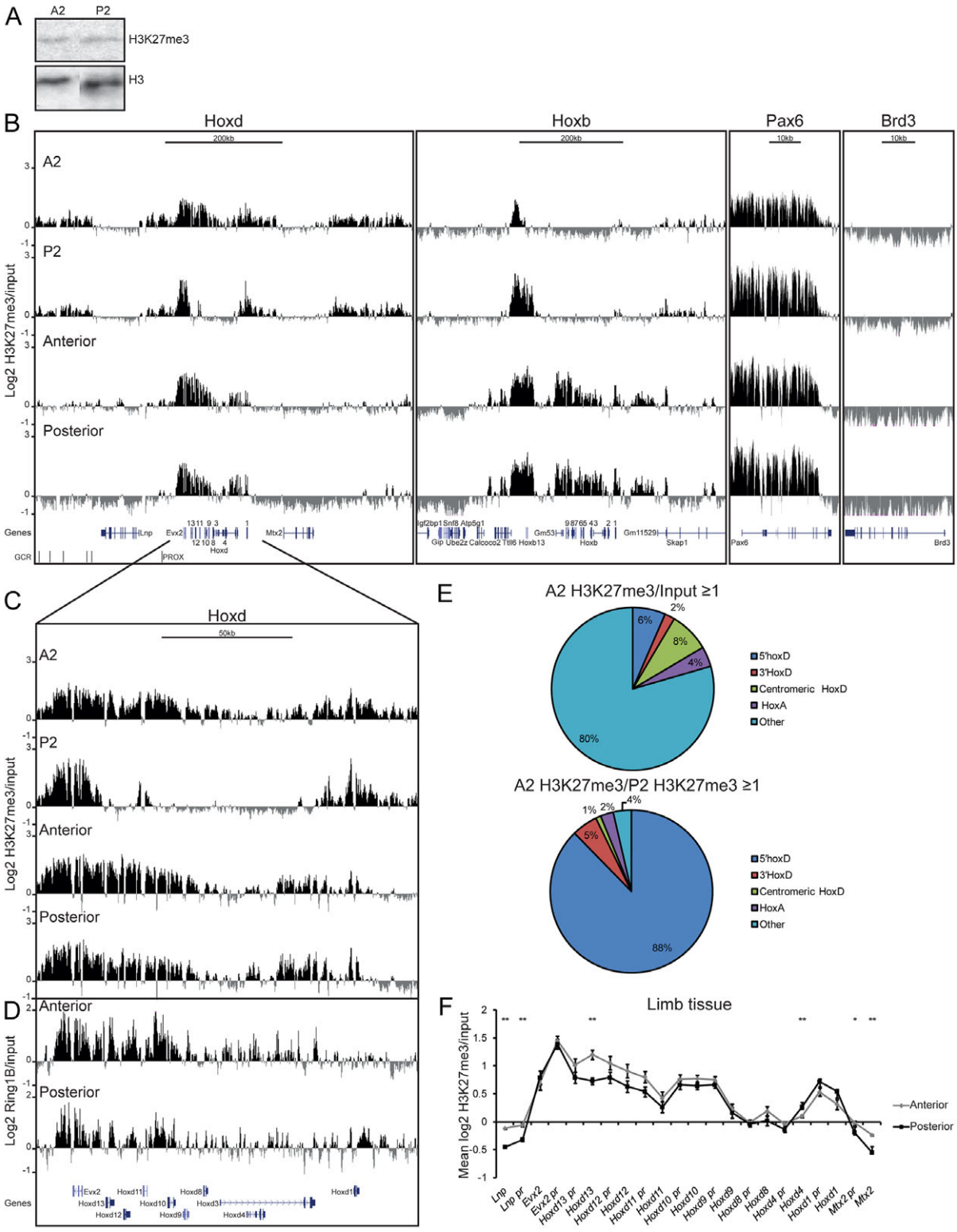

Fig. 3. H3K27me3 and Ring1B distribution in E10.5 limb cell lines and mouse forelimb tissue. (**A**) Western blot of H3K27me3 in A2 and P2 cells. Levels of H3 are shown for comparison. (**B**) Mean log2 H3K27me3/input at HoxD, HoxB, *Pax6* and *Brd3* loci in A2 and P2 cell lines (top two rows, *n*=2 biological replicates) and for anterior and posterior forelimb tissue (bottom two rows, *n*=4 – 2 biological and 2 technical – replicates). (**C**) As in B, but with an expanded view of the HoxD cluster. (**D**) Mean log2 Ring1B/input at the HoxD region for anterior and posterior forelimb tissue. *n*=2 biological replicates. (**E**) Pie charts showing the genomic distribution of different probes categories enriched for: (top) H3K27me3 in A2 cells (log2 H3K27me3/input≥1) versus (bottom) the proportion with an A2/P2 difference of log2≥1. (**F**) Mean (± s.e.m.) log2 H3K27me3/input at HoxD and neighbouring genes and promoters in distal forelimb anterior and posterior tissue. Average log2 values were calculated from each individual probe value that covered the genomic locations. The statistical significance of A:P differences in H3K27me3 enrichment over each gene and promoter were examined by two-sample *t*-test (**P*<0.01, ***P*<0.0001).

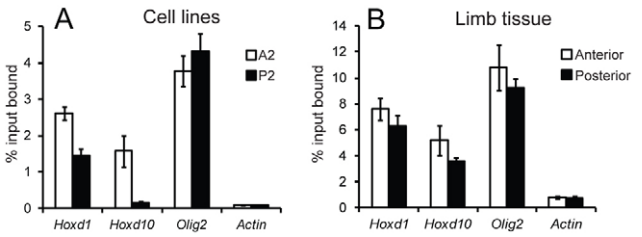

Fig. 4. Confirmation of A/P differences in H3K27me3 enrichment. (**A**) qPCR analysis of ChIP for H3K27me3 at *Hoxd1*, *Hoxd10*, *Olig2* and *Actb* (Actin) promoters in A2 (white) and P2 (black) cells. Enrichment is shown as mean percent input bound ± s.e.m. over three biological replicates. (**B**) As in A, but from E10.5 distal anterior (white) and posterior (black) forelimb tissue. *n*=2 biological replicates.

cells at *Hoxd10* (Fig. 4A), with a less dramatic decrease at *Hoxd1*. There was no significant A-P difference in H3K27me3 at *Olig2* (positive control) or *Actb* (β-actin, negative control).

The limb cell lines are apparently a relatively homogenous mesenchymal cell population. By contrast, more heterogeneous cell populations are inevitably present in dissected limb tissue, where AER- and differentiating mesenchyme-derived cells expressing different cohorts of Hoxd genes might mask differences in H3K27me3 levels between the anterior and posterior mesenchyme. Nevertheless, we analysed H3K27me3 in E10.5 anterior and posterior dissected distal forelimbs (Fig. 3B,C, lower two tracks) by nChIP. An H3K27me3 block covered the 3′ end of HoxD (*d1-d4*) in both limb regions. Both anterior and posterior limb samples also show a second block of modification over the 5′ end of HoxD, from *Hoxd9* through to *Evx2*, but this was at a significantly higher level in anterior compared with the posterior region (*Hoxd13-Hoxd10*; $P<0.0001$) (Fig. 3F; supplementary material Fig. S2A). Whereas in the anterior distal limb H3K27me3 remains high from *Evx2* to beyond *Hoxd12*, in the posterior H3K27me3 declines from the *Evx2-Hoxd13* intergenic region up to *Hoxd11* (Fig. 3C,F). This A-P difference in H3K27me3 levels in limb tissue is specific to 5′ HoxD (supplementary material Fig. S2B), being significantly greater ($P<0.0001$) than that for all Hox loci combined. The A-P difference for the 3′ HoxD region is not significant ($P=0.57$). H3K27me3 was more pervasive over the HoxB locus in cells dissected from the anterior or posterior limb tissues than in the cell lines (Fig. 3B), probably reflecting the heterogeneity of the former cell populations, and, unlike for HoxD, there were no differences between H3K27me3 at HoxB in anterior versus posterior limb.

Lower H3K27me3 levels at the *Hoxd10* promoter in the posterior versus anterior limb tissue cell populations was confirmed by qPCR, whereas levels at the *Hoxd1* promoter were similar in both distal limb samples (Fig. 4B). These data are consistent with a role for polycomb-mediated repression in regulating A-P differences in Hoxd expression during the patterning of the distal limb.

PRC1 levels are reduced over HoxD in distal posterior cells

The PRC1 complex recognises and binds to H3K27me3-modified chromatin to bring about chromatin compaction and gene repression. The ChIP profile of the PRC1 component Ring1B (Rnf2 – Mouse Genome Informatics) correlates with that of H3K27me3 in E10.5 anterior and posterior dissected limb bud tissues at HoxD (Fig. 3C,D), HoxB and *Pax6* (supplementary material Fig. S2C). A block of Ring1B extends from *Evx2* to *Hoxd9* in distal anterior limb, whereas in posterior tissue Ring1B coverage is more sparse over *Hoxd13-11*. The A/P difference in Ring1B levels was significant throughout 5′ HoxD (promoters, genes and intergenic; $P<0.0001$) (supplementary material Fig. S2D) and was greater than at all the Hox loci combined ($P<0.0001$) (supplementary material Fig. S2E).

A-P differences in chromatin compaction at HoxD in the distal limb bud

PRC1 brings about a chromatin compaction at target loci that is detectable by measuring the spatial separation of fluorescence in situ hybridisation (FISH) signals from closely apposed probe pairs (Eskeland et al., 2010) (Fig. 5A). Using these approaches, a decompaction of HoxD is seen as Hox genes are activated during ES cell differentiation and in the tail-bud of the embryo (Morey et al., 2007). We first used 2D FISH to assay chromatin compaction (d^2) across HoxD in limb bud cell lines. Any difference in inter-probe distances due to variation in nuclear size between cell lines was also assessed by normalising d^2 to the nuclear radius (r^2). However, in practice the same conclusions were reached when considering interprobe separation without normalisation to nuclear size (supplementary material Table S4).

Chromatin across HoxD (*Hoxd3-Hoxd13*) was significantly less compact in both distal posterior cell lines compared with the anterior cell lines ($P=0.0002$ and $P=0.03$ for A1/P1 and A2/P2, respectively) (Fig. 5B; supplementary material Table S4). This was restricted to the region with differential H3K27me3, as there is no significant A-P difference in compaction of the GCR-*Lnp* region 5′ of HoxD where similarly low levels of H3K27me3 are seen between A and P cell lines (Fig. 3B). Nor was there differential compaction at a control region around *Pax6* (*Rcn-Rpl10*) on the same chromosome (MMU2) as HoxD, and which contains genes with no known role in limb development and no differential expression between A and P cell lines.

We sought confirmation of differential HoxD compaction by 3D FISH in E10.5-11.0 embryo tissue sections cut through the anterior and posterior regions of the forelimb (Fig. 6A). Here, not only were we able to compare the distal anterior and posterior regions, but also more proximal limb regions and indeed the flank region adjacent to the forelimb bud tissue where Hoxd genes are not expressed. As observed in the cell lines, HoxD chromatin was significantly less compact (d^2) at the distal posterior forelimb bud compared with the distal anterior ($P=0.0008$) but also compared with other limb and non-limb regions analysed (proximal posterior, $P=0.02$; proximal anterior, $P=0.0002$; flank, $P=0.0002$) (Fig. 6B; supplementary material Table S5). There was no significant difference in HoxD chromatin compaction between any of the other limb regions or even between these other limb regions and the flank tissue. Likewise, no significant differences in chromatin compaction could be identified between any of the tissue regions, including the distal forelimb bud, at the GCR-*Lnp* and the *Rcn-Rpl10* control regions (Fig. 6B; supplementary material Table S5).

Analysis of the frequency of hybridisation signals separated by defined distance (d) intervals also demonstrates the less compact chromatin conformation of the HoxD locus compared with the adjacent GCR-*Lnp* region. The proportion of colocalised signals (<200 nm) was less for *Hoxd3-Hoxd13* than both GCR-*Lnp* (with a similar genomic distance) and the *Rcn-Rpl10* control region, which has a larger genomic distance separating the two probes (supplementary material Fig. S3). Conversely, the proportion of signal pairs >400 nm apart was greater for *Hoxd3-Hoxd13* than for the other two regions. The subpopulation of HoxD probe pairs separated by >600 nm

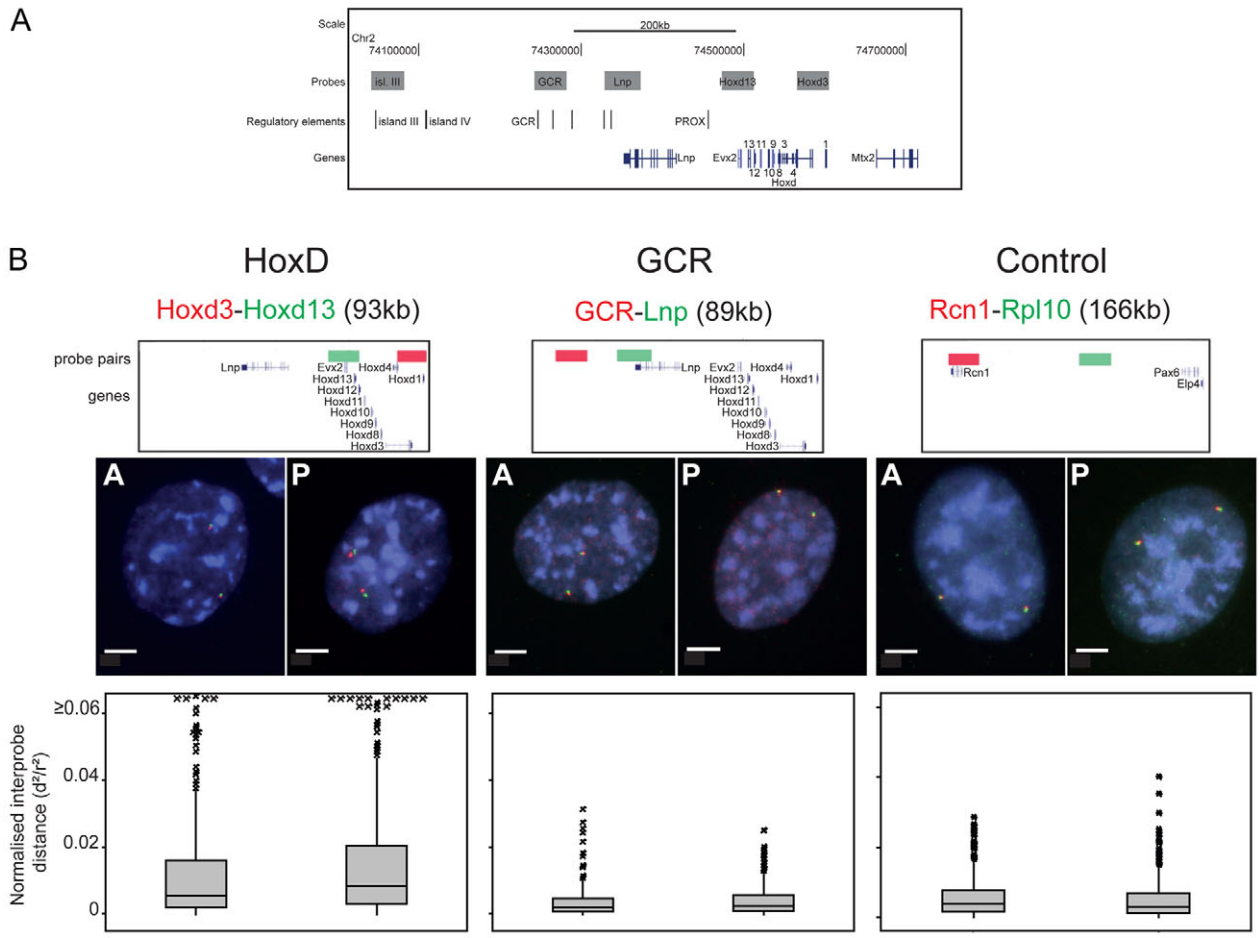

Fig. 5. Chromatin decompaction at HoxD in distal posterior limb cells. (**A**) Schematic of the genomic region around HoxD. The grey boxes above depict the probes used for FISH. Regulatory elements including the GCR and the PROX enhancer are also indicated. (**B**) 2D FISH with probe pairs at HoxD, the region centromeric to HoxD (GCR-*Lnp*) and a control region on MMU2, in A1 and P1 nuclei counterstained with DAPI (blue). Scale bars: 5 µm. Probe positions are shown above the images. Box plots show the distribution of interprobe distances (d^2) normalised to nuclear radius (r^2) for A1 and P1 cells. Shaded boxes show the median and interquartile range of data; crosses indicate outliers. *n*=~400 for HoxD, *n*=~300 for GCR, *n*=~300 for control. The statistical significance of differences were examined by Mann-Whitney U-tests (supplementary material Table S4).

(~25%) was significantly greater in the distal posterior forelimb compared with the rest of the limb regions and the adjacent flank (distal anterior, P=0.0001; proximal posterior, P=0.0001; proximal anterior, P=0.001; flank, P<0.0001) (supplementary material Fig. S3 and Table S6). We conclude that chromatin unfolding at HoxD accompanies the start of the second wave of Hoxd gene activation during limb development.

Nuclear colocalisation of the GCR and 5′ HoxD in distal posterior cells

Developmental stage E10.5 marks the start of the later phase of distal limb development, characterised by 'quantitative collinearity' in which strong expression of *Hoxd13* is initiated in the distal margin of the posterior mesenchyme. This is driven by enhancer elements located 5′ of HoxD, the best characterised of which is the GCR ~180 kb centromeric of *Hoxd13* (Fig. 3B, Fig. 5A) (Montavon et al., 2008).

One possible mechanism for long-range cis-regulation invokes the spatial colocalisation of the enhancer and target promoter (Williamson et al., 2011). Indeed, modelling of gene expression changes that occur as a consequence of 5′ HoxD deletions led to the suggestion that the first step in 5′ Hoxd gene activation in the distal posterior limb bud might be a long-range interaction (looping) between regulatory sequences such as GCR and the *Hoxd13* region (Montavon et al., 2008). To test this model, we analysed the hybridisation signals for the GCR and *Hoxd13* regions and determined the proportion that colocalised (d<200nm). Multiple areas within the E11.0 limb bud were analysed (Fig. 7A), including those where late phase gene activation initiates (distal posterior); where it is poised for later activation (distal anterior); where the early phase of 3′-5′ *Hoxd* gene activation, which does not depend on the GCR, has occurred (proximal limb); and then in the control flank mesoderm.

In contrast to the *Hoxd3-d13* region, average GCR-*Hoxd13* interprobe distances (d^2) were significantly smaller in the posterior distal limb bud compared with distal anterior (P=0.04) and also compared with the flank (P<0.0001), the proximal anterior (P=0.02) and proximal posterior (P=0.002) (Fig. 7B; supplementary material Table S7). This was not seen for distances between GCR and a probe at the 3′ end of *Lnp* (Fig. 6B) indicating that there is not a simple compaction of the whole region 5′ of *Hoxd13*. Moreover, the proportion of alleles for which GCR and *Hoxd13* probe signals were spatially colocalised (d<200nm) was far higher in posterior distal limb (>30%) than in all other areas

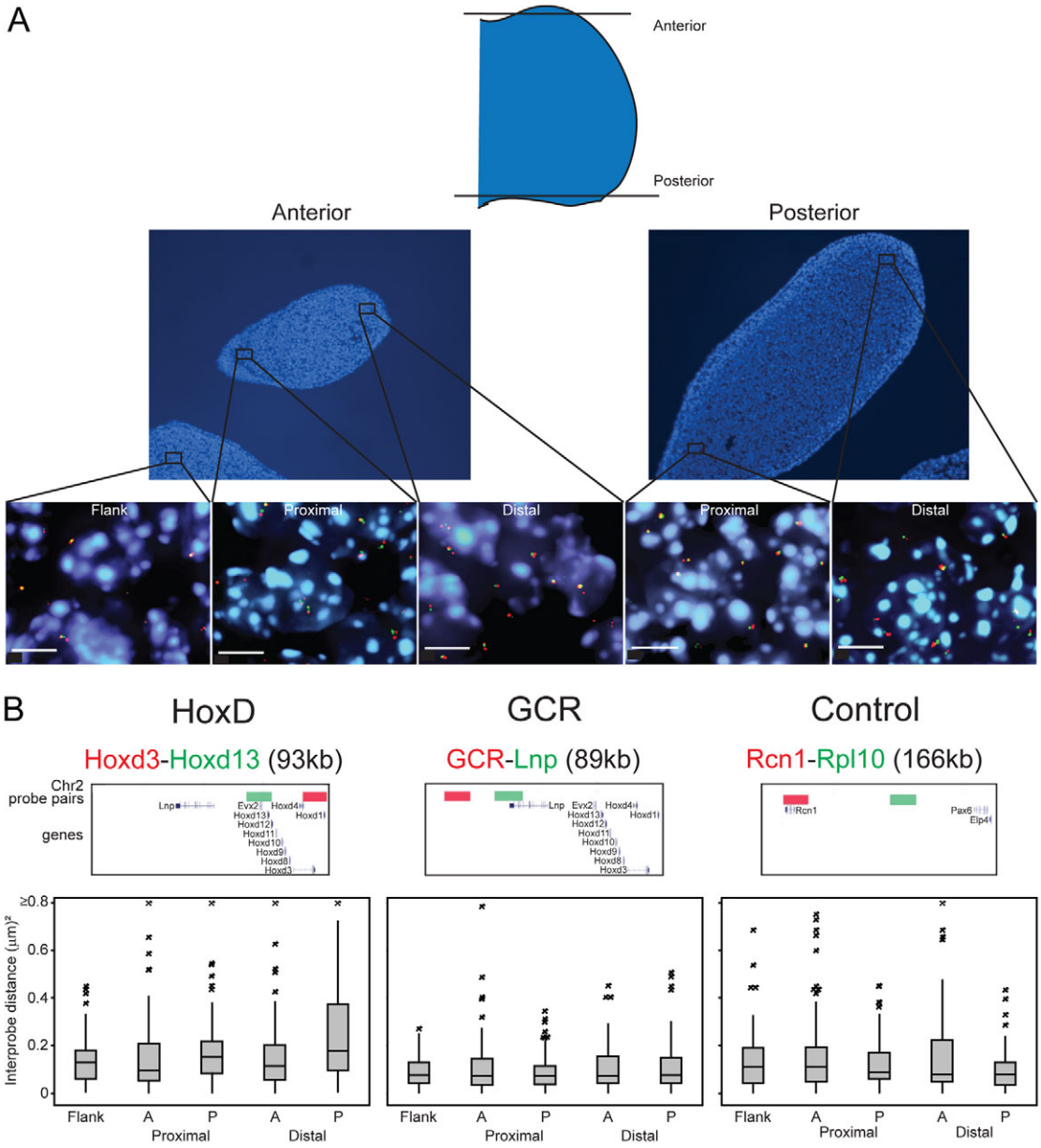

Fig. 6. Decompaction of HoxD specific to the distal posterior region of E11 mouse forelimbs. (**A**) FISH on anterior and posterior tissue sections. Schematic above indicates the position and plane of the sections. Below are examples of nuclei from each of the limb regions and the adjacent flank, hybridised with HoxD probe pairs. Scale bars: 5 μm. (**B**) Box plots show the distribution of interprobe distances (d^2) at the HoxD, GCR and control loci for the proximal and distal anterior and posterior forelimb bud and the adjacent flank. *n*=100 for HoxD, *n*>80 for GCR, *n*>80 for control. Probe positions as indicated in Fig. 5A. Statistical analysis is shown in supplementary material Table S5.

tested (Fig. 7C; supplementary material Fig. S4A and Table S8). These data are consistent with the formation of a chromatin loop between *Hoxd13* and the GCR regulatory module at the time and place of distal limb bud development when the GCR participates in initiating HoxD collinearity and *Hoxd13* expression.

Recently, 4C analysis was used to identify sequences that could be captured together with *Hoxd13* by cross-linking in tissue from across the distal limb at E12.5, a later stage of limb development than that studied here (Montavon et al., 2011). This identified multiple new potential regulatory regions in the extensive gene desert centromeric of HoxD and located far beyond the GCR. It was suggested that these elements might simultaneously interact with each other and with 5′ HoxD in the distal limb to form a compact regulatory hub. One of these new elements, island III, is ~200 kb centromeric of the GCR. We measured the physical separation between the island III region and the GCR – a genomic distance equivalent to that separating GCR from *Hoxd13* (~200 kb) – and also between the island III region and *Hoxd13* (Fig. 5A) in E11.0 limb buds. We did not detect any significantly increased frequency of colocalisation between island III and *Hoxd13* or between island III and GCR in posterior distal limb, compared with other forelimb regions (Fig. 7C; supplementary material Fig. S4B,C and Table S8). However, average interprobe distances (d^2) between 5′ HoxD and island III are similar to those between 5′ HoxD-GCR and GCR-island III despite the former being double the genomic distance, and the compaction of the island III-Hoxd13

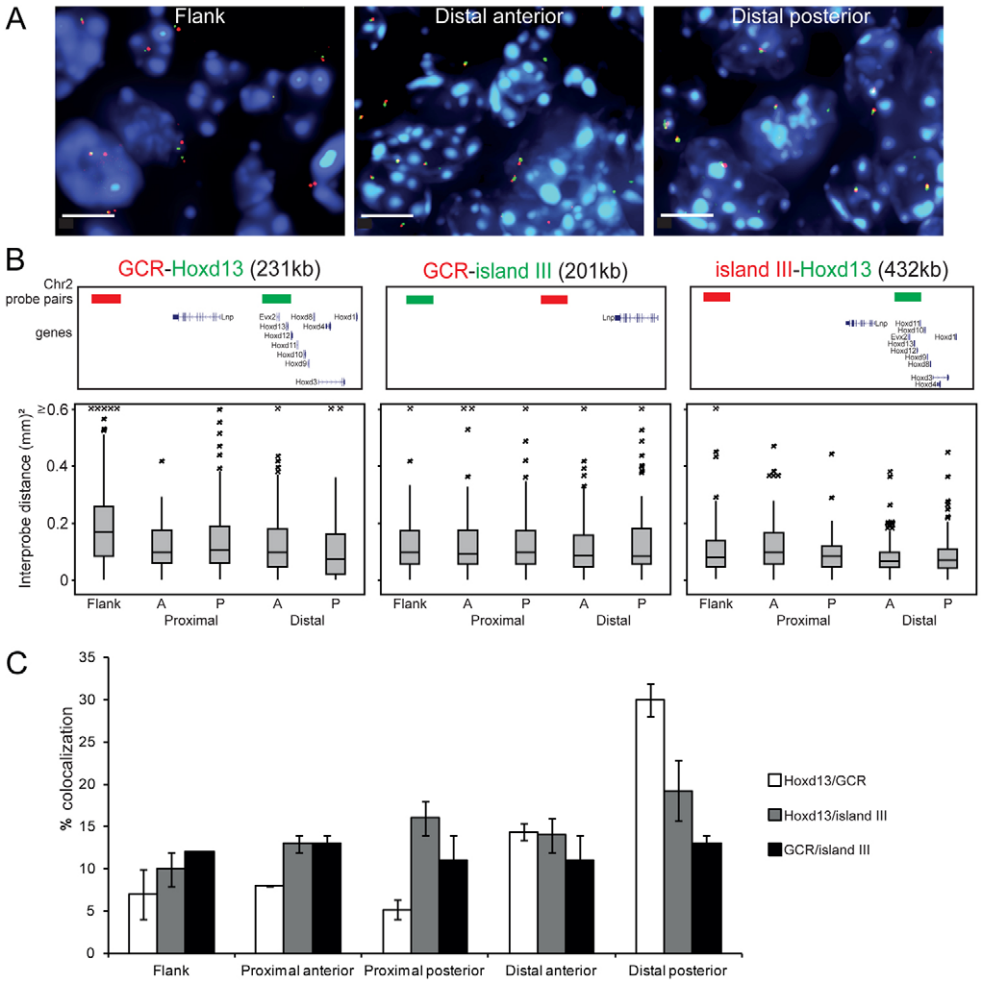

Fig. 7. GCR–5′ HoxD colocalisation at the E11 distal posterior mouse forelimb bud. (**A**) Nuclei from distal limb and flank tissue sections after FISH with GCR and 5′ HoxD probe pairs. Scale bars: 5 μm. (**B**) FISH with probe pairs: GCR–5′ HoxD, GCR–island III and 5′ HoxD–island III at distal and proximal forelimb bud and flank regions. $n \geq 100$ for each. Distribution of interprobe distances as in Fig. 6B and statistical analysis shown in supplementary material Table S7. (**C**) Percentage colocalised probe pairs (d<200 nm) for each of the genomic regions assayed in the distal anterior, distal posterior, proximal anterior, proximal posterior forelimb bud and the flank. n=100. Error bars indicate s.e.m. obtained from two different tissue sections. The statistical significance of the differences in colocalisation between the limb regions and the flank were examined by Fisher's exact test (supplementary material Table S8).

region is higher in distal limb than in the proximal region (Fig. 7; supplementary material Table S7). These data suggest that the entire ~400 kb genomic region from the 5′ end of HoxD into the centromeric gene desert is configured in a relatively compact, yet flexible, conformation.

DISCUSSION

Differential 5′ Hoxd gene expression across the limb A-P axis in mouse E10.5 embryos

We have shown that immortalised cell lines derived from the anterior versus posterior E10.5 distal limb sustain global expression patterns consistent with their mesenchymal origin from the limb during A-P axis specification (Fig. 1). Moreover, they maintain some of the differential Hoxd gene expression patterns of the temporal developmental stage from which they are derived (Fig. 2). Limb-specific secondary activation of HoxD initiates in the distal posterior forelimb bud at ~E10.5 in the mouse. We confirmed that 5′ *Hoxd* gene expression is significantly higher in distal posterior E10.5 cells relative to distal anterior (Fig. 2), commensurate with the spatial and temporal aspects of late phase collinearity (Spitz et al., 2003). This A-P difference is lost by E11.5-12.5 as the 5′ genes are activated in more distal anterior regions, but with *Hoxd13* still being the most strongly expressed (supplementary material Fig. S5) (Montavon et al., 2008).

A model for late phase HoxD activation in the distal limb proposes a reciprocal activation pathway involving restriction of earlier 5′ *Hoxd* expression to the posterior margin of the limb bud that then ensures posterior-only *Shh* activation, which in turn effects the initiation of secondary 5′ *Hoxd* expression in the distal posterior limb bud (Harfe et al., 2004; Robert and Lallemand, 2006). We did not detect *Shh* expression in either (P1 or P2) posterior-derived cell lines, which nonetheless maintain upregulated 5′ *Hoxd* gene expression, especially of *Hoxd13* (Fig. 2). However, late phase Hoxd gene expression still occurs in $Shh^{-/-}$ $Gli3^{-/-}$ mice, and is only reduced in $Shh^{-/-}$ mice (Litingtung et al., 2002). *Shh* expression prevents *Gli3* repression of 5′ Hoxd genes, both by repressing *Gli3* expression posteriorly and by antagonising Gli3 processing to Gli3R (repressor form). Gli3R has been suggested to be the repressor of 5′ Hoxd gene expression in anterior limb (Tarchini and Duboule, 2006). It is likely that posterior *Shh* expression in the embryonic limb, up to the point of cell immortalisation, was sufficient to activate late phase HoxD expression. The subsequent strong maintenance of *Hoxd13* expression in the posterior limb-derived cell lines suggests that *Shh* is not required for maintaining HoxD activity. Lack of *Shh* expression in cell lines could be due to the loss of interaction between the mesenchyme and the overlying AER (Laufer et al., 1994; Ogura et al., 1996).

Posteriorly restricted Hoxd expression is accompanied by chromatin decompaction and reduced polycomb binding

We identified differential H3K27me3 over 5′ Hoxd genes between anterior- and posterior-derived limb cell lines. In anterior cells, there is extensive H3K27me3 over *Hoxd13* towards *Hoxd10* that is

absent from posterior cells (Fig. 3). That this contributes to the repression of 5′ Hoxd genes in the anterior distal limb is consistent with the ectopic anterior expression of *Hoxd11* and *Hoxd13* in E10.5 limb buds after ablation of the PRC2 component Ezh2 (Wyngaarden et al., 2011). As Hoxd genes have been transcriptionally active during the earlier phase of limb development, we cannot determine whether there is active H3K27me3 removal from 5′ HoxD in cells of the posterior distal E10.5 limb, or re-establishment of H3K27me3 in anterior distal cells. The presence of H3K27me3 over 3′ HoxD in both anterior and posterior cell lines and in limb tissue probably represents a re-imposition of this histone mark after the initial phase of Hoxd gene expression in early, proximal limb bud development (Fig. 3).

Late phase 5′ Hoxd gene expression in the posterior distal limb bud is also accompanied by reduced Ring1B binding (Fig. 3; supplementary material Fig. S2). Consistent with a requirement for PRC1 to maintain Hox clusters in a silent, compact chromatin state in ES cells (Eskeland et al., 2010), we found differential chromatin compaction at HoxD both between the anterior- and posterior-derived limb cell lines and in limb tissue (Figs 5, 6), with a greater degree of decompaction in cells from the posterior distal region than in cells measured in any other region of the limb bud at this stage of development (Fig. 6). Our data suggest that dynamic polycomb-mediated control of HoxD expression and higher-order chromatin structure, previously described as occurring along the main A-P body axis (Morey et al., 2007; Soshnikova and Duboule, 2009) is recapitulated during limb-specific Hoxd gene expression.

Enhancer-promoter looping restricted to the posterior distal region of the limb bud

We observed a closer average proximity, and higher absolute colocalisation frequency, of the GCR to *Hoxd13* specifically at the distal posterior limb bud (Fig. 7; supplementary material Fig. S4A), consistent with the formation of a chromatin loop between enhancers of the GCR and the 5′ HoxD at the time and place of distal limb bud development when the GCR is involved in initiating HoxD quantitative collinearity and *Hoxd13* expression (Spitz et al., 2003; Montavon et al., 2008). Our analysis at single cell resolution is consistent with average interaction frequencies captured between the 5′ HoxD region and GCR by 4C analysis in the limb bud (Montavon et al., 2011), although there was no analysis of anterior versus posterior regions of the limb in that study. By contrast, we did not find significantly increased colocalisation of 5′ HoxD and a more distal regulatory region (island III) (Fig. 7C; supplementary material Fig. S4C) previously identified by 4C analysis. However, that study was conducted at a later developmental stage (E12.5) and any colocalisation might therefore occur later than E11. Our analysis does indicate that the ~400 kb region extending out from the 5′ HoxD locus into the adjacent gene desert is in a generally compact chromatin conformation in all the limb regions tested (Fig. 7A,B; supplementary material Table S7). Thus, the entire regulatory domain may be configured to minimise the nuclear search space for enhancer-gene interactions to then occur at the correct time and place. Further analyses over the entire temporal stage of late phase HoxD activity (E10.5-12.5) and across the A-P distal axis should clarify the spatial and temporal specificity of the multiple regulatory elements. We could not assess the colocalisation of the Prox regulatory element (Fig. 5A) as it is too close to *Hoxd13* for us to resolve by FISH.

This article demonstrates that complex higher-order chromatin dynamics is implicated in the regulation of Hoxd gene expression during distal limb development, with two different processes occurring: chromatin decompaction and looping. It also establishes immortomouse-derived cell lines as a tractable model system to investigate chromatin states in relation to spatiotemporally regulated gene expression in development. This is the first demonstration of differential chromatin compaction and enhancer-gene colocalisation across the A-P axis of the developing limb, as a previous study analysing the spatial colocalisation of *Shh* and its limb enhancer reported similar colocalisation in both posterior (*Shh*-expressing) and anterior (no *Shh* expression) limb regions (Amano et al., 2009).

Acknowledgements
We thank the Welcome Trust Clinical Research Facility (Western General Hospital, Edinburgh) for Illumina expression arrays.

Funding
This work was supported by the Medical Research Council, UK. Deposited in PMC for release after 6 months.

Competing interests statement
The authors declare no competing financial interests.

Supplementary material
Supplementary material available online at
http://dev.biologists.org/lookup/suppl/doi:10.1242/dev.081174/-/DC1

References

Amano, T., Sagai, T., Tanabe, H., Mizushina, Y., Nakazawa, H. and Shiroishi, T. (2009). Chromosomal dynamics at the Shh locus: limb bud-specific differential regulation of competence and active transcription. *Dev. Cell* **16**, 47-57.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. B* **57**, 289-300.

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. et al. (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nat. Cell Biol.* **441**, 349-353.

Chambeyron, S. and Bickmore, W. A. (2004). Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription. *Genes Dev.* **18**, 1119-1130.

Chambeyron, S., Da, Silva, N. R., Lawson, K. A. and Bickmore, W. A. (2005). Nuclear re-organisation of the Hoxb complex during mouse embryonic development. *Development* **132**, 2215-2223.

Condie, B. G. and Capecchi, M. R. (1993). Mice homozygous for a targeted disruption of Hoxd-3 (Hox-4.1) exhibit anterior transformations of the first and second cervical vertebrae, the atlas and the axis. *Development* **119**, 579-595.

Deschamps, J. and van Nes, J. (2005). Developmental regulation of the Hox genes during axial morphogenesis in the mouse. *Development* **132**, 2931-2942.

Duboc, V. and Logan, M. P. O. (2011). Regulation of limb bud initiation and limb-type morphology. *Dev. Dyn.* **240**, 1017-1027.

Dunning, M. J., Smith, M. L., Ritchie, M. E. and Tavaré, S. (2007). beadarray: R classes and methods for Illumina bead-based data. *Bioinformatics* **23**, 2183-2184.

Eden, E., Navon, R., Steinfeld, I., Lipson, D. and Yakhini, Z. (2009). GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* **10**, 48.

Endoh, M., Endo, T. A., Endoh, T., Fujimura, Y.-I., Ohara, O., Toyoda, T., Otte, A. P., Okano, M., Brockdorff, N., Vidal, M. et al. (2008). Polycomb group proteins Ring1A/B are functionally linked to the core transcriptional regulatory circuitry to maintain ES cell identity. *Development* **135**, 1513-1524.

Eskeland, R., Leeb, M., Grimes, G. R., Kress, C., Boyle, S., Sproul, D., Gilbert, N., Fan, Y., Skoultchi, A. I., Wutz, A. et al. (2010). Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination. *Mol. Cell* **38**, 452-464.

Faust, C., Lawson, K. A., Schork, N. J., Thiel, B. and Magnuson, T. (1998). The Polycomb-group gene eed is required for normal morphogenetic movements during gastrulation in the mouse embryo. *Development* **125**, 4495-4506.

Harfe, B. D., Scherz, P. J., Nissim, S., Tian, H., McMahon, A. P. and Tabin, C. J. (2004). Evidence for an expansion-based temporal Shh gradient in specifying vertebrate digit identities. *Cell* **118**, 517-528.

Hill, R. E. (2007). How to make a zone of polarizing activity: insights into limb development via the abnormality preaxial polydactyly. *Dev. Growth Differ.* **49**, 439-448.

Jat, P. S., Noble, M. D., Ataliotis, P., Tanaka, Y., Yannoutsos, N., Larsen, L. and Kioussis, D. (1991). Direct derivation of conditionally immortal cell lines

from an H-2Kb-tsA58 transgenic mouse. *Proc. Natl. Acad. Sci. USA* **88**, 5096-5100.

Kohn, E. A., Du, Z., Sato, M., Van Schyndle, C. M., Welsh, M. A., Yang, Y.-A., Stuelten, C. H., Tang, B., Ju, W., Bottinger, E. P. et al. (2010). A novel approach for the generation of genetically modified mammary epithelial cell cultures yields new insights into TGFβ signaling in the mammary gland. *Breast Cancer Res.* **12**, R83.

Krawchuk, D. and Kania, A. (2008). Identification of genes controlled by LMX1B in the developing mouse limb bud. *Dev. Dyn.* **237**, 1183-1192.

Kuijper, S., Beverdam, A., Kroon, C., Brouwer, A., Candille, S., Barsh, G. and Meijlink, F. (2005). Genetics of shoulder girdle formation: roles of Tbx15 and aristaless-like genes. *Development* **132**, 1601-1610.

Laufer, E., Nelson, C. E., Johnson, R. L., Morgan, B. A. and Tabin, C. (1994). Sonic hedgehog and Fgf-4 act through a signaling cascade and feedback loop to integrate growth and patterning of the developing limb bud. *Cell* **79**, 993-1003.

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.-I. et al. (2006). Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* **125**, 301-313.

Lettice, L. A, Williamson, I., Wiltshire, J. H., Peluso, S., Devenney, P. S., Hill, A. E., Essafi, A., Hagman, J., Mort, R., Grimes, G. et al. (2012). Opposing functions of the ETS factor family define Shh spatial expression in limb buds and underlie polydactyly. *Dev. Cell* **22**, 459-467.

Litingtung, Y., Dahn, R. D., Li, Y., Fallon, J. F. and Chiang, C. (2002). Shh and Gli3 are dispensable for limb skeleton formation but regulate digit number and identity. *Nature* **418**, 979-983.

Lizarraga, G., Ferrari, D., Kalinowski, M., Ohuchi, H., Noji, S., Kosher, R. A. and Dealy, C. N. (1999). FGFR2 signaling in normal and limbless chick limb buds. *Dev. Genet.* **25**, 331-338.

Mao, J., McGlinn, E., Huang, P., Tabin, C. J. and McMahon, A. P. (2009). Fgf-dependent Etv4/5 activity is required for posterior restriction of Sonic Hedgehog and promoting outgrowth of the vertebrate limb. *Dev. Cell* **16**, 600-606.

Mizusawa, N., Hasegawa, T., Ohigashi, I., Tanaka-Kosugi, C., Harada, N., Itakura, M. and Yoshimoto, K. (2004). Differentiation phenotypes of pancreatic islet beta- and alpha-cells are closely related with homeotic genes and a group of differentially expressed genes. *Gene* **331**, 53-63.

Montavon, T., Le Garrec, J.-F., Kerszberg, M. and Duboule, D. (2008). Modeling Hox gene regulation in digits: reverse collinearity and the molecular origin of thumbness. *Genes Dev.* **22**, 346-359.

Montavon, T., Soshnikova, N., Mascrez, B., Joye, E., Thevenet, L., Splinter, E., de Laat, W., Spitz, F. and Duboule, D. (2011). A regulatory archipelago controls hox genes transcription in digits. *Cell* **147**, 1132-1145.

Morey, C., Da Silva, N. R., Perry, P. and Bickmore, W. A. (2007). Nuclear reorganisation and chromatin decondensation are conserved, but distinct, mechanisms linked to Hox gene activation. *Development* **134**, 909-919.

Morey, C., Kress, C. and Bickmore, W. A. (2009). Lack of bystander activation shows that localization exterior to chromosome territories is not sufficient to up-regulate gene expression. *Genome Res.* **19**, 1184-1194.

Ogura, T., Alvarez, I. S., Vogel, A., Rodríguez, C., Evans, R. M. and Izpisúa-Belmonte, J. C. (1996). Evidence that Shh cooperates with a retinoic acid inducible co-factor to establish ZPA-like activity. *Development* **122**, 537-542.

Pradeepa, M. M., Sutherland, H. G., Ule, J., Grimes, G. R. and Bickmore, W. A. (2012). Psip1/Ledgf p52 binds methylated histone H3K36 and splicing factors and contributes to the regulation of alternative splicing. *PLoS Genet*. **8**, e1002717.

Robert, B. and Lallemand, Y. (2006). Anteroposterior patterning in the limb and digit specification: contribution of mouse genetics. *Dev. Dyn.* **235**, 2337-2352.

Smyth, G. K., Michaud, J. and Scott, H. S. (2005). Use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics* **21**, 2067-2075.

Soshnikova, N. and Duboule, D. (2009). Epigenetic temporal control of mouse Hox genes in vivo. *Science* **324**, 1320-1323.

Spitz, F., Gonzalez, F. and Duboule, D. (2003). A global control region defines a chromosomal regulatory landscape containing the HoxD cluster. *Cell* **113**, 405-417.

Spitz, F., Herkenne, C., Morris, M. A. and Duboule, D. (2005). Inversion-induced disruption of the Hoxd cluster leads to the partition of regulatory landscapes. *Nat. Genet.* **37**, 889-893.

Stock, J. K., Giadrossi, S., Casanova, M., Brookes, E., Vidal, M., Koseki, H., Brockdorff, N., Fisher, A. G. and Pombo, A. (2007). Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. *Nat. Cell Biol.* **9**, 1428-1435.

Tarchini, B. and Duboule, D. (2006). Control of Hoxd genes' collinearity during early limb development. *Dev. Cell* **10**, 93-103.

Tschopp, P. and Duboule, D. (2011). A regulatory "landscape effect" over the HoxD cluster. *Dev. Biol.* **351**, 288-296.

Vickerman, L., Neufeld, S. and Cobb, J. (2011). Shox2 function couples neural, muscular and skeletal development in the proximal forelimb. *Dev. Biol.* **350**, 323-336.

Voncken, J. W., Roelen, B. A. J., Roefs, M., de Vries, S., Verhoeven, E., Marino, S., Deschamps, J. and van Lohuizen, M. (2003). Rnf2 (Ring1b) deficiency causes gastrulation arrest and cell cycle inhibition. *Proc. Natl. Acad. Sci. USA* **100**, 2468-2473.

Williamson, I., Hill, R. E. and Bickmore, W. A. (2011). Enhancers: from developmental genetics to the genetics of common human disease. *Dev. Cell* **21**, 17-19.

Wyngaarden, L. A., Delgado-Olguin, P., Su, I.-H., Bruneau, B. G. and Hopyan, S. (2011). Ezh2 regulates anteroposterior axis specification and proximodistal axis elongation in the developing limb. *Development* **138**, 3759-3767.

Zakany, J. and Duboule, D. (2007). The role of Hox genes during vertebrate limb development. *Curr. Opin. Genet. Dev.* **17**, 359-366.

Developmental Cell
Short Article

Opposing Functions of the ETS Factor Family Define *Shh* Spatial Expression in Limb Buds and Underlie Polydactyly

Laura A. Lettice,[1] Iain Williamson,[1] John H. Wiltshire,[1] Silvia Peluso,[1] Paul S. Devenney,[1] Alison E. Hill,[1] Abdelkader Essafi,[1] James Hagman,[2] Richard Mort,[1] Graeme Grimes,[1] Carlo L. DeAngelis,[1] and Robert E. Hill[1,*]
[1]MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh EH4 2XU, UK
[2]Integrated Department of Immunology, National Jewish Health, Denver, CO 80206, USA
*Correspondence: bob.hill@hgu.mrc.ac.uk
DOI 10.1016/j.devcel.2011.12.010

SUMMARY

Sonic hedgehog (*Shh*) expression during limb development is crucial for specifying the identity and number of digits. The spatial pattern of *Shh* expression is restricted to a region called the zone of polarizing activity (ZPA), and this expression is controlled from a long distance by the *cis*-regulator ZRS. Here, members of two groups of ETS transcription factors are shown to act directly at the ZRS mediating a differential effect on *Shh,* defining its spatial expression pattern. Occupancy at multiple GABPα/ETS1 sites regulates the position of the ZPA boundary, whereas ETV4/ETV5 binding restricts expression outside the ZPA. The ETS gene family is therefore attributed with specifying the boundaries of the classical ZPA. Two point mutations within the ZRS change the profile of ETS binding and activate *Shh* expression at an ectopic site in the limb bud. These molecular changes define a pathogenetic mechanism that leads to preaxial polydactyly (PPD).

INTRODUCTION

The zone of polarizing activity (ZPA) was experimentally defined as the region located at the posterior margin of the developing limb bud that determines digit number and identity (Hill, 2007; Tickle, 2006). The polarizing activities attributed to the ZPA are mediated by sonic hedgehog (SHH) which is postulated to act as a diffusible morphogen. A number of models have been proposed to explain *Shh* activity (Towers and Tickle, 2009) and most recently, SHH was shown to act as both a morphogen and a mitogen to coordinate digit formation by integrating growth with digit specification during limb development (Towers et al., 2008; Zhu et al., 2008). The restricted spatial expression within the ZPA is an essential element of all the proposed models of *Shh* function (Ahn and Joyner, 2004; Harfe et al., 2004; Yang et al., 1997).

A critical step in understanding the complexity of *Shh* expression was the identification of the *cis*-regulatory element called the ZRS for *Z*PA regulatory *s*equence (Lettice et al., 2002, 2003) (also called MFCS1 [Sagai et al., 2005]). The ZRS comprises ~800 bp of a multispecies conserved sequence and encodes most, if not all, of the information that regulates the spatiotemporal pattern of *Shh* expression in the developing limb bud. The ZRS is the paradigm for long-range gene regulation, operating over ~1 Mb to regulate *Shh* (Lettice et al., 2003; Sagai et al., 2005). Single point mutations in the human ZRS are found in patients presenting with a range of limb skeletal malformations. These include preaxial polydactyly type 2 (PPD2), triphalangeal thumb polysyndactyly (TPTPS), syndactyly type IV (SD4), and Werner's mesomelic syndrome (WMS), collectively referred to as ZRS-associated syndromes (Lettice et al., 2003, 2008; Farooq et al., 2010; Furniss et al., 2008; Gurnett et al., 2007; Semerci et al., 2009; Wieczorek et al., 2010). The point mutations function to generate ectopic expression at the anterior margin of the limb bud (Furniss et al., 2008; Lettice et al., 2008), which is the underlying cause of PPD.

Here, members of the large group of ETS transcription factors (Sharrocks, 2001) are shown to play distinct roles in the spatial pattern of *Shh*. Occupancy at multiple ETS sites, which bind the factors GABPα and ETS1, regulates the position of the *Shh* expression boundary in the limb, thus defining the ZPA. Multiple binding of ETV4 and ETV5 at the ZRS, in contrast, represses ectopic *Shh* expression outside the ZPA. In addition, we show that two PPD mutations disrupt the balance in number of ETS binding sites derepressing expression in the anterior limb bud.

RESULTS

Multiple ETS Sites Identified in the ZRS

The notion of "homotypic clustering" (Gotea et al., 2010; Lifanov et al., 2003; Wagner, 1999) suggests that *cis*-regulators contain multiple sites for crucial regulatory factors. Analysis of the ZRS identified a 7 bp motif (AGGAA$^{G}/_{A}$T) that is repeated five times (Figure 1A), with each repetition located in a highly conserved position (Figure S1A available online). This purine-rich sequence is contained within the consensus for the 8 bp ETS1 binding motif ($^{C}/_{G}/_{A}$AGGAA$^{G}/_{A}$T) found in putative distal regulators of genes in T cells (Hollenhorst et al., 2009). None of the numerous point mutations in the ZRS that cause limb deformities fall within these conserved motifs; however, mutations identified in two families with PPD2, Family A & C (AC) (Gurnett et al., 2007) and an Australian family (AUS) (E. de Graaff, personal

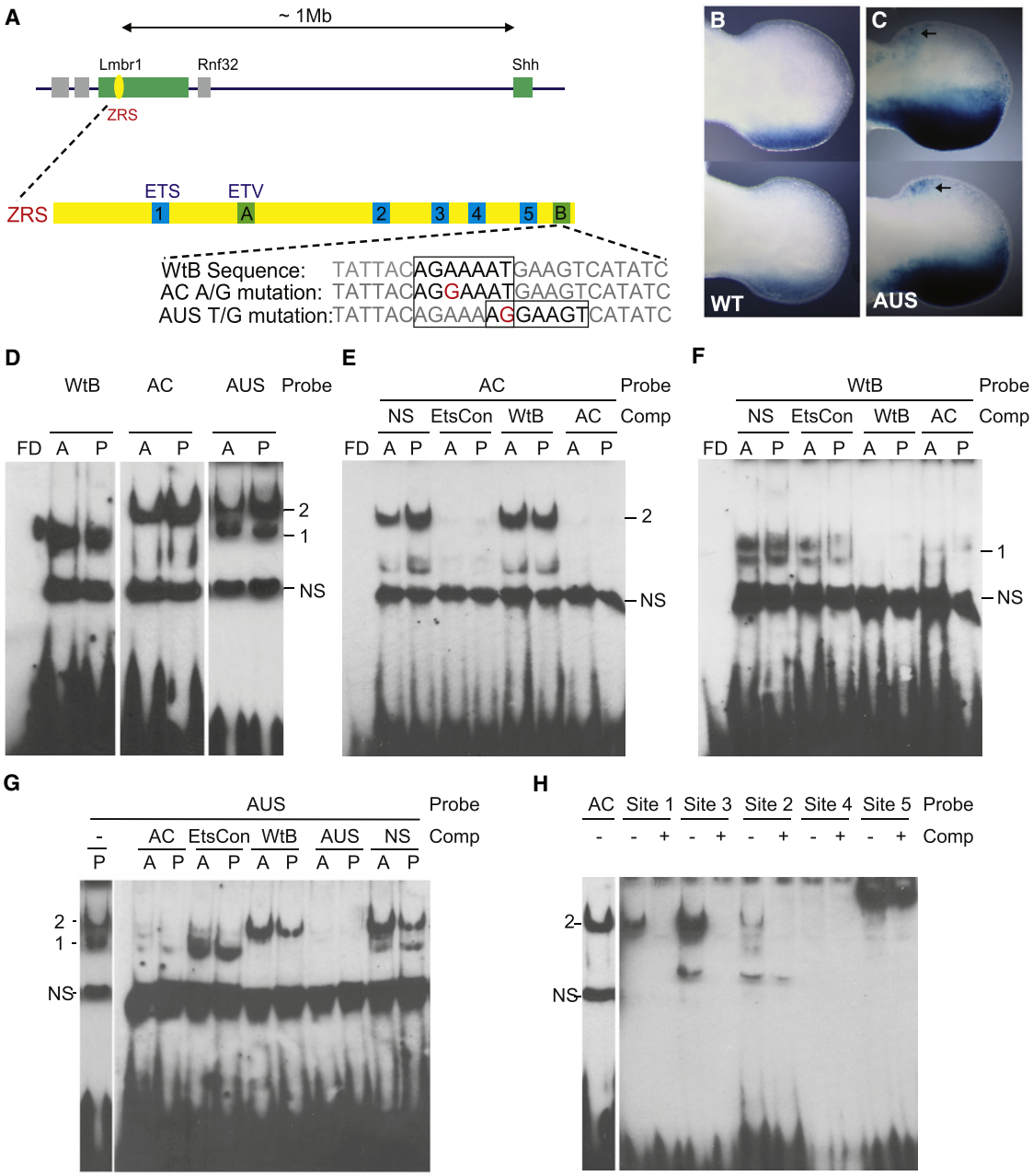

Figure 1. Point Mutations Alter *Shh* Expression and Protein Binding Profiles

(A) Schematic showing the ZRS (yellow box), which resides within intron 5 of LMBR1, 1 Mb for the *Shh* gene. The positions of the ETS sites 1–5 and ETV sites A and B identified within the ZRS are marked by blue and green boxes, respectively. The sequences around the mutations identified in families with PPD (Family AC and AUS) are shown.

(B and C) Limbs from transgenic animals carrying wild-type (B) and mutant (C) ZRS reporter constructs (forelimb buds are shown on top and hindlimb buds below) demonstrate that the AUS mutation results in expansion of the posterior expression (compare to B) and ectopic staining in the anterior mesenchyme (arrows).

(D–H) EMSA analysis of nuclear extracts from anterior (A) and posterior (P) halves of E11.5 limb buds.

(D) Nuclear extract was incubated with ds-oligos containing the WT, AC, or AUS sequence. The WT sequence produced a specific band (1); the AC point mutation resulted in a higher migrating band (2); and the AUS mutation produced a combination of WT and AC binding; bands 1 and 2. A nonspecific (NS) band was observed for all ds-oligos.

(E–G) EMSA using the AC ds-oligo (E), WT ds-oligo (F), and AUS ds-oligo (G), and using an unlabeled NS sequence, ETS consensus sequence (EtsCon), WT, or AC oligonucleotide as their competitors.

(H) Comparison by EMSA of the binding for the wild-type ZRS sites 1–5, showing a greater extent of binding to the AC mutant site and sites 1 and 3. The unlabelled AC oligonucleotide (lanes labeled +) specifically competes for band 2.

communication), convert the surrounding sequence to additional ETS motifs (Figure 1A).

Transgenic analysis using a construct containing the ZRS drives expression of the LacZ reporter gene in the expected posterior position in the limb (Furniss et al., 2008; Lettice et al., 2003, 2008) (Figure 1B). Addition of the AUS point mutation to the ZRS was sufficient to drive the ectopic expression at the anterior margin of E11.5 limb buds (Figure 1C, arrows). (The number of transgenic embryos is reported in Table S1.) The mutation also caused an overall increase in the width of posterior expression such that the boundary extended deeper into the middle of the limb (Figure 1C).

Differential Binding at the Mutant and Wild-Type Sites of the ZRS

The protein-binding profile associated with the 7 bp motif was analyzed by electrophoretic mobility shift assays (EMSAs) using biotin-labeled double-strand oligonucleotide (ds-oligos) probes and nuclear extracts from E11.5 embryonic limbs. Initially, the AUS and AC mutant sites were analyzed using a series of 24 bp ds-oligos (Figure 1A) spanning either the wild-type sequence (WtB) or the mutant sequences. The WtB sequence probe produced a single specific band (band1 in Figure 1D), which was replaced by a higher migrating band in the presence of the AC mutation (band 2 in Figure 1D). In comparison, the AUS mutant probe exhibited a combination of both the WtB and AC mutant band-shift patterns. Specificity of binding was confirmed using unlabelled competitors for either the WtB or the AC mutant sequence (Figures 1E–1G). An unrelated sequence (Fisher et al., 1992) that contained the 8 bp ETS1 CAGGAAGT motif (designated EtsCon) competed for the upper band detected by both the AC (Figure 1E) and AUS (Figure S1B) probes; whereas, the WtB probe competed but with an appreciably lower affinity (Figure 1F). In contrast, competition with the unlabelled WtB sequence did not affect the AC banding pattern (Figure 1E) but did disrupt binding of the lower band detected with the AUS sequence (Figure 1G). These data are consistent with the two point mutations binding similar proteins (Figure 1A) at closely apposed positions. The two mutations affect wild-type protein binding differently; whereas the AC mutation causes replacement of wild-type binding, the AUS mutation allows binding of either protein.

The five endogenous AGGAA$^{G}/_{A}$T sites within the 800bp ZRS (Figure S1A) were also analyzed. ds-Oligo probes for sites 1, 2 and 3 each generated a band that migrated to the same position as that for the AC probe (Figure 1H) and showed specificity for binding by competition with the unlabelled AC probe. Site 4 probe did not detectably bind a protein while site 5 bound an unidentified, higher migrating band which was not competed with the AC sequence.

Identification of the Factors that Bind the ZRS In Vivo

A number of ETS family members are expressed in the early-stage limb and are distributed along the distal mesenchyme. *Ets1* and *Ets2* (Ristevski et al., 2002) (Figures 2A and 2B) are expressed in the posterior mesenchyme overlapping the *Shh* domain in the limb bud at E10.5 and by E11.5 extend distally, incorporating the anterior margin (Figure 2A). *Etv4* and *Etv5* (also called *Pea3* and *Erm*, respectively) (Figure 2C and 2D) (Mao et al., 2009; Zhang et al., 2009) are expressed in the mesenchyme along the entire distal edge. *Gabpα* (EMAGE:2836) (Ristevski et al., 2004) is reportedly more highly expressed in the distal mesenchyme in the posterior margin of the limb and *Elf1* (EMAGE:1462) is expressed throughout the entire limb mesenchyme (Richardson et al., 2010).

To investigate the binding of candidate ETS proteins to the ZRS sequences, we used ETS antibodies specific for ETS1, ETS2, GABPα, ELF1, ETV4, and ETV5. Western blot analysis of limb extracts (Figure 2E) showed that these proteins are produced in both anterior and posterior halves of the limb bud and are enriched in the nuclear extracts. EMSA analysis showed that the anti-ETV4 (αETV4 in Figure 2F) antibody depleted the binding of the wild-type protein responsible for the WtB banding pattern, whereas the anti-ETS1 antibody (αETS1 in Figure 2F) depleted the protein binding to the upper band of the AC sequence. The other antibodies did not detectably affect the banding pattern (data not shown).

The sequence that ETV4 binds encompasses the noncanonical site AGAAAAT (referred to as ETV4 site B) (Figures 1A and S2A) (Xin et al., 1992). A second ETV4 binding site found to contain the AGAAA sequence (referred to as site A) (Figures 1A and S1) is the site of the previously published PPD2 mutation Belg2 (Lettice et al., 2003). The Belg2 mutation, which converts the sequence to AGGAA, was previously shown to drive ectopic expression in transgenic assays (Lettice et al., 2008). EMSA demonstrated that ETV4 binds the Belg2 wild-type ds-oligo (WtA), as confirmed by depletion with the ETV4 antibody (αETV4 in Figure 2G). Analysis of the mutant Belg2 sequence, however, showed binding to an additional unidentified factor (band 2 in Figure 2G).

To establish that ETS factors bind directly to the ZRS in the limb, we performed chromatin immunoprecipitation (ChIP) in nuclear extracts from whole autopods of E11.5 embryonic limbs with the series of antibodies above. GABPα and ELF1 were of particular interest, given that recent analysis of ETS1 binding in vivo showed co-occupation by these factors at a substantial subset of sites (Hollenhorst et al., 2007, 2009). In our initial screen of the ChIP, neither the ETS2 nor the ELF1 antibody showed any appreciable binding to the ZRS using qPCR (data not shown). Further analysis using high-density tiling microarrays showed the binding of both ETS1 and GABPα to the ZRS region (Figure 2H). The binding profiles suggested that GABPα occupied the whole region of the ZRS, whereas ETS1 binding overlapped but was skewed toward the 3′ end of the ZRS. Accordingly, the two sites at the 3′ end (sites 4 and 5) contain the sequence AGGAA**A**T (Figure S1A), while the remaining sites contain AGGAA**G**T. The *Shh* gene (Figure 2H), including the promoter, and other control regions showed no binding to any of these ETS factors. ELF1 was used as a negative control and showed no binding (Figure 2H) at the ZRS, while appreciable binding was detected at a clustering of AGGAAA sites located near the 5′ end of an uncharacterized SH3-containing gene (3′Hba, Figure 2H).

ETV4 and ETV5 act to repress *Shh* expression in the posterior margin of the limb bud to ensure that no ectopic production occurs (Mao et al., 2009; Zhang et al., 2009). In limb extracts, both the anti-ETV4 and the anti-ETV5 antibodies detect binding to the ZRS region (Figure 2H). ETV4 binding appeared as two

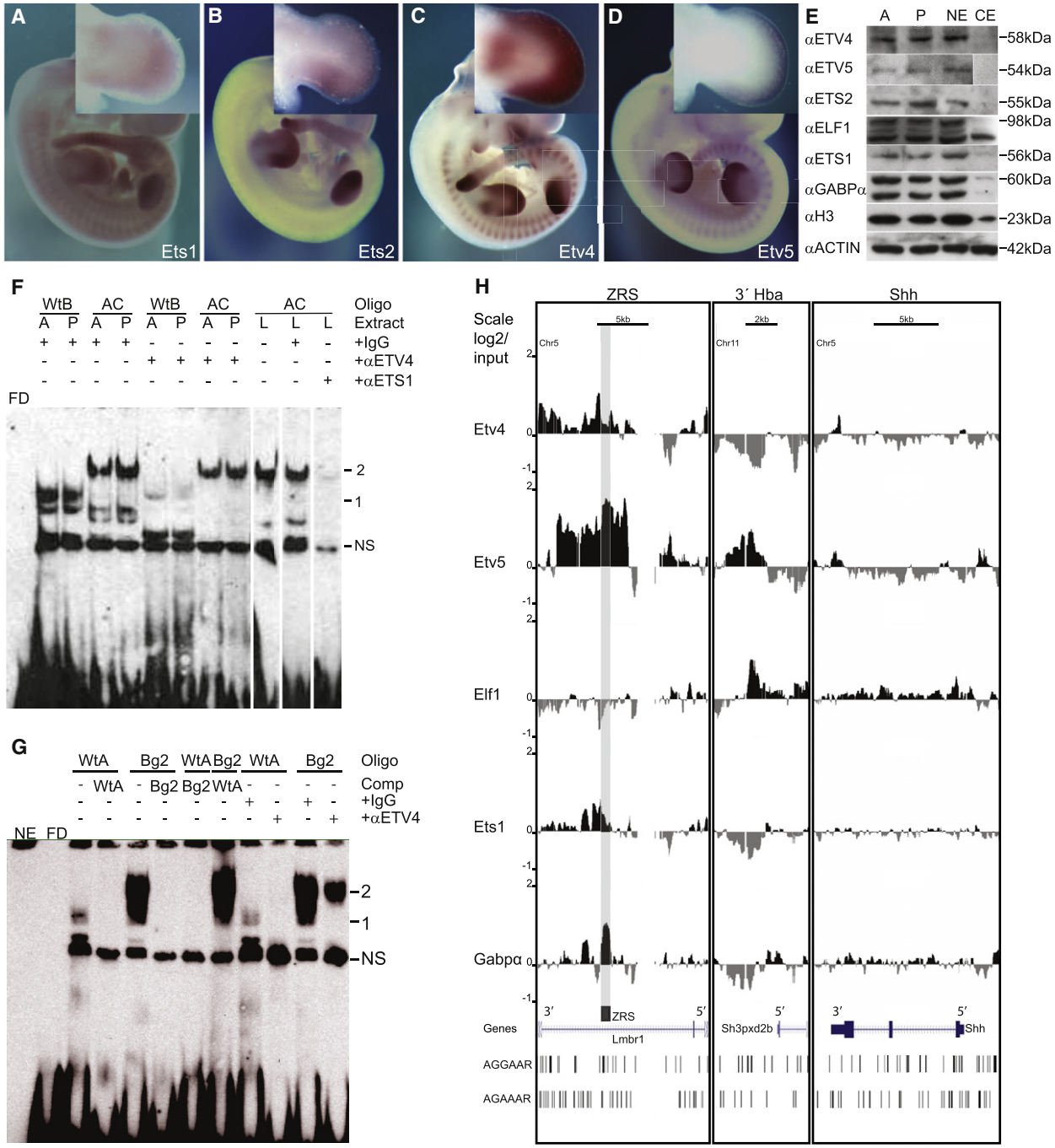

Figure 2. ETS Factors Are Expressed in the Limb and Bind to the ZRS

(A–D) Whole-mount in situ hybridization analysis for *Ets1* (A), *Ets2* (B), *Etv4* (C), and *Etv5* (D) are shown in E11.5 embryos and limb buds.

(E) Western blot analysis using antibodies raised against ETS factors, designated αETV4, αETV5, αETS2, αELF1, αETS1, and αGABPα, and against histone H3 (αH3) and actin (αactin), with nuclear extracts from the anterior (A) and posterior (P) halves of the limb buds (E11.5). Also shown is a comparison between limb nuclear extracts (NE) and cytoplasmic extracts (CE). αH3 and αactin were used as loading controls.

(F) EMSA shows WtB and AC ds-oligo binding in nuclear extracts depleted for ETV4 or ETS1 using specific antibodies (αEtv4 and αEts1). (IgG was used as a nonspecific control.) Extracts from anterior (A) or posterior (P) halves or whole limbs (L) from E11.5 limb buds were used. Band 1 observed with the WtB probe was specifically depleted by the addition of αETV4 antibody, while Band shift 2 observed with the AC probe was specifically depleted by the addition of αETS1 antibody.

(G) EMSAs were conducted with ds-oligos containing the sequence for the wild-type ETV4 site A (WtA) or the Belg2 mutation (Bg). WtA ds-oligo shows a specific band (1) while that for Bg sequence gives an additional higher migrating band (2). The anti-ETV4 antibody depletes Band 1 observed with WtA and Bg probes (nonspecific IgG used as control).

peaks in the region of the ZRS, which reflects the location of the two predicted ETV4 sites. Surprisingly, ETV5 occupies a broad region of >5 kb with a peak of binding over the whole of the ZRS (Figure 2H). ETV5 binding is highly specific for the ZRS region, as shown by the tiling microarray over the whole 1 Mb of the *Shh* locus (Figure S2B). ETV5 binds the ZRS directly; the broad region may reflect the high density of AGAAA$^{G}/_{A}$ sites available at nearby sites (Figure 2H).

Ets1 is expressed not only in the posterior mesenchyme but also along the anterior margin overlapping the ectopic domain driven by the AUS mutation (Figure 1C). To determine whether ETS1 binds to the ZRS at both the anterior and the posterior halves, we dissected limb buds at E11.5 for ChIP. qPCR showed significant binding of ETS1 at the ZRS (Figure S2C) in both the anterior and posterior halves, suggesting that the ZRS is open for factor interactions on the quiescent as well as the active side of the limb.

Ectopic *Shh* Expression Occurs by Two Mechanisms

We investigated two possible explanations for the ectopic expression driven by the AUS point mutation. First, by displacing the ETV4/ETV5 repressor at site B (Figure 1A), the binding of GABPα/ETS1 to the mutant site may lead to ectopic activation. To investigate this possibility, we generated two different mutations that disrupt site B. Neither a terminal deletion of 44 bp (tDel, Figure 3A), which removes the entire site, nor replacement of three base pairs within the ETV4 binding site (AGA**AAA**T→AGA**GCG**T) (-ETVB, Figure 3B) caused ectopic expression in transgenic embryos. In fact, expression levels decreased but importantly, the spatial boundary appears unaffected. Thus we tested a second possibility, specifically that the AUS mutation, by creating an additional AGGAAGT site, may directly drive ectopic activity. An extra AGGAAGT site was added to the ZRS. A putative neutral position, a variable stretch of Ts that differ in mammalian species (human has six Ts, mouse has eight, and rat has 22), was selected as the site for the insertion (Figure 3E). The modified ZRS generated ectopic expression activity (Figure 3G), whereas the control (insertion of seven Ts [Figure 3E]) showed no detectable ectopic expression (Figure 3F). These data show that an additional GABPα/ETS1 site is sufficient, on its own, to generate ectopic expression.

These analyses, as a consequence, raised questions about the direct role ETV4/ETV5 plays in restricting ectopic expression of *Shh*. To investigate the regulatory role of the ETV4/ETV5 proteins further, the three-base-pair replacement in site B, discussed above, was made in site A. Similar to the case with loss of only site B, no ectopic expression was observed with loss of site A (Figure 3C). The double mutation of both sites A and B, however, resulted in ectopic expression along the distal, anterior margin of the limb (Figure 3D). The simultaneous removal created a loss of ETV function and confirms that this ETV subfamily acts directly at the ZRS to restrict the expression pattern of *Shh* to the posterior margin.

The Endogenous GABPα/ETS1 Sites Define the Boundary of *Shh* Expression

To investigate the consequences of having multiple, clustered ETS sites, the five endogenous sites in the ZRS were systematically inactivated (**AGG**AA$^{G}/_{A}$T → **CTT**AA$^{G}/_{A}$T) and evaluated by the transgenic assay (Figures 3H–3Q). For each transgenic, the relative domain widths of the limb buds were measured as indicated in Figure 3H and plotted. The mean domain widths of fore- and hindlimbs combined for each experimental group were then compared. The mean size of the expression domain was found to be highly significantly related to the combination of ETS sites that were disrupted (analysis of variance [ANOVA] $p < 0.0001$). The results of subsequent pairwise comparisons (Tukey's HSD test) are shown in Table S2.

Disruption of both sites 1 and 3 (the two sites that showed highest affinity for ETS1 binding in vitro) (Figure 3L) resulted in a significant reduction (p values in Table S2) in reporter-gene expression as compared to the wild-type construct (Figure 3I), whereas singly (Figure 3J), neither site detectably changed limb-bud expression (Figures 3I–3K). Disruption of additional sites resulted in no further reduction in the expression domain (Figure 3H). The activity levels of sites 1 and 3 were further analyzed. The sole presence of either one of these sites (Figures 3O and 3P) established boundaries of expression approaching that of the wild-type construct, whereas the presence of both 1 and 3 together (Figure 3Q) generated a wide expression domain that was indistinguishable. Sites 2, 4, and 5 (Figure 3L) showed some activity, as in the presence of either site 1 or 3, producing an expression domain similar to that of the wild-type (Figure 3H).

The addition of the extra AUS mutant site caused a further upregulation in transgenic embryos, with expression extending deeper into the center of the limb (Figure 3H). The construct carrying the AUS mutation but lacking site 1 showed little change (Figure 3R), whereas, again, removing sites 1 and 3 caused a retraction of the expression boundary (Figure 3S) but only to the extent of that generated by the wild-type construct (Figure 3H). Further removal, deleting sites 1, 2 and 3 together (Figure 3T), or all five sites, caused little further change in expression, suggesting that this single high-affinity site is sufficient for generating the wild-type transgenic expression. With the inclusion of the AUS mutation, ectopic anterior expression occurs in the limbs (Figure 1C); however, in the absence of both sites 1 and 3 the number of limbs with ectopic expression decreases from 83% to 33%, and in the absence of all five sites (Table S1), no ectopic expression was detected, suggesting that the wild-type ETS sites assist in the ectopic expression. In addition, an earlier stage of limb development was examined (E10.5) using constructs containing the wild-type ZRS, and with site 1 and sites 1 and 3 disrupted and with the AUS mutation added. The relative size differences of the expression domains were unchanged; thus, no temporal differences were apparent (data not shown).

(H) ChIP using antibodies to five different ETS factors (ETV4, ETV5, Elf1, ETS1, and GABPα) analyzed by hybridizing to tiling microarrays. Summary is presented using three different genomic regions, the y axis is Log_2 for each ChIP/input DNA and the x axis represents a segment of DNA from the microarray. The DNA region containing the ZRS is highlighted by the gray shading. As controls, the whole of the *Shh* coding region plus promoter (Shh) and the region downstream of the α-globin locus (3′Hba) are shown. Scale bars are shown at the top and the positions of potential ETS1/GABPα (AGGAA$^{G}/_{A}$) and ETV4/ETV5 (AGAAA$^{G}/_{A}$) binding sites at the bottom of each panel.

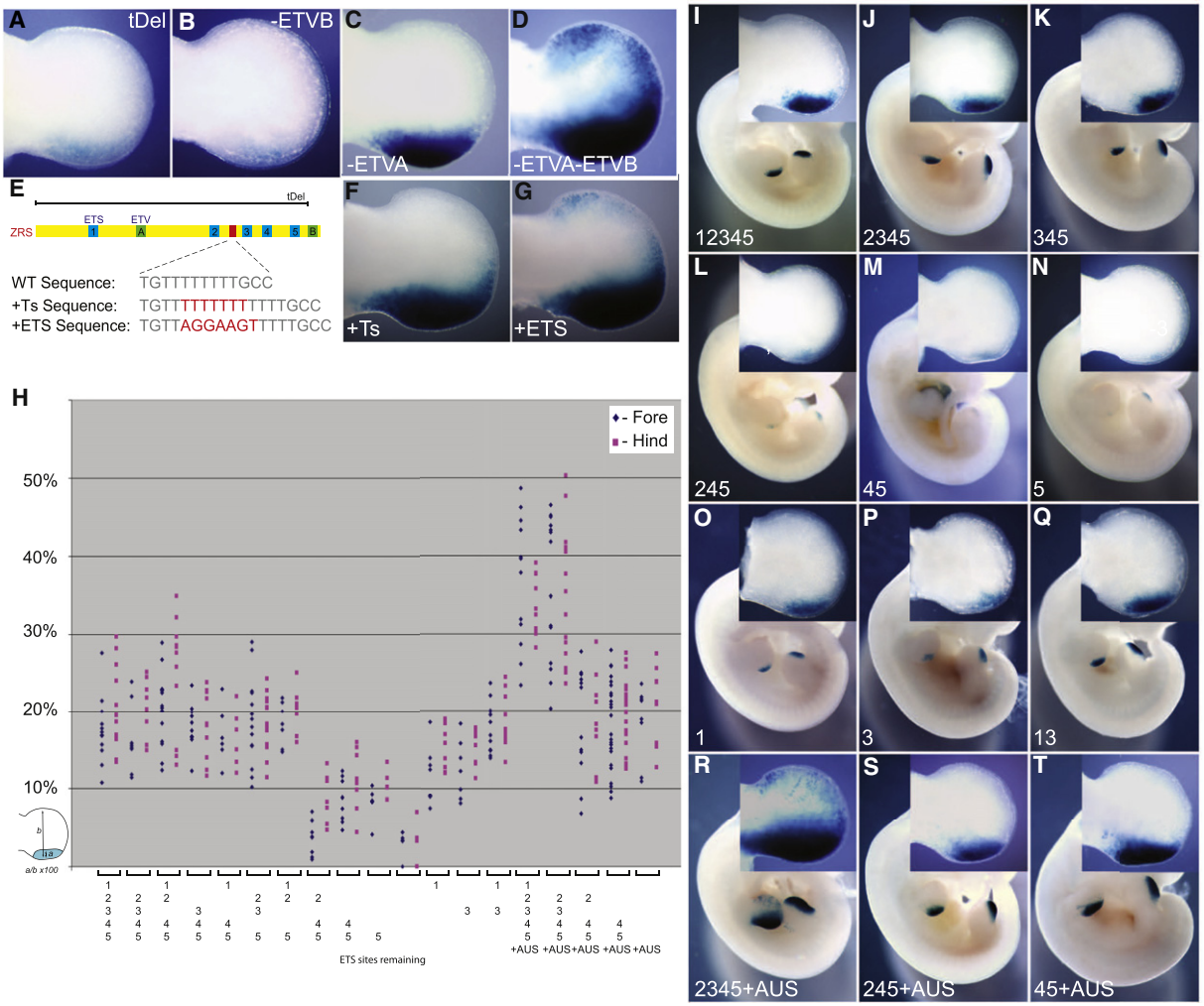

Figure 3. Transgenic Analysis of Embryos Carrying Mutant ZRS Sequences

(A–D) Limbs from transgenic embryos carrying the following mutant ZRS sequences: the 44 bp terminal deletion (tDel) (A), the 3 bp change in the ETV4 Site B (-ETVB) (B), and Site A (-ETVA) (C). Disruption of both sites in combination (D) results in ectopic expression in the anterior of the limb.

(E–G) Position of the run of Ts within the ZRS and the changes added are shown in red. Expression due to these changes is shown by comparison of the addition of seven Ts (F) and the extra ETS1/GABPα site (AGGAAGT) (G). Ectopic expression is detected in (G).

(H) Graphical representation of the expression pattern resulting from mutations within the endogenous ETS sites. Expression pattern is the ratio of the width of the expression domain divided by the width of the limb, expressed as a percentage (see p values in Table S2).

(I–Q) Examples of transgenic embryos for the ETS mutations analyzed are shown. The ETS sites remaining are depicted in the lower-left-hand corner of each figure.

(R–T) Transgenic embryos that represent the addition of the AUS mutation in combination with ETS site mutations are shown. A close-up of a forelimb for each is shown in the insets.

DISCUSSION

Asymmetric expression is essential for SHH morphogen activity in the developing limb. The ETS gene family plays a central role in the *Shh* spatial pattern, both positioning the expression of Shh at the posterior margin of the limb and repressing ectopic expression at the anterior margin. The limitation of the cell-free approach to identify the full spectrum of ETS family members was highlighted by the in vivo ChIP analysis, which identified additional binding factors. The co-occupancy identified for ETS1 and GABPα at the ZRS corresponds to the genome-wide occupancy analysis of Hollenhorst et al. (2009), which shows that the vast majority of GABPα sites colocalize with those that bind ETS1 sites. In addition, the ChIP approach showed that both ETV4 and ETV5 act directly at the ZRS. Even with the resolution afforded by ChIP in combination with tiling microarrays, it is difficult to determine whether GABPα and ETS1 or ETV4 and ETV5 are competing for the same sites; however, the differences in profiles do suggest that there are site preferences.

Redundancy within this large group of ETS factors complicates both the biochemical and genetic analysis of gene function of individual family members. However, sequential inactivation of the ETS sites provided an approach to investigate the

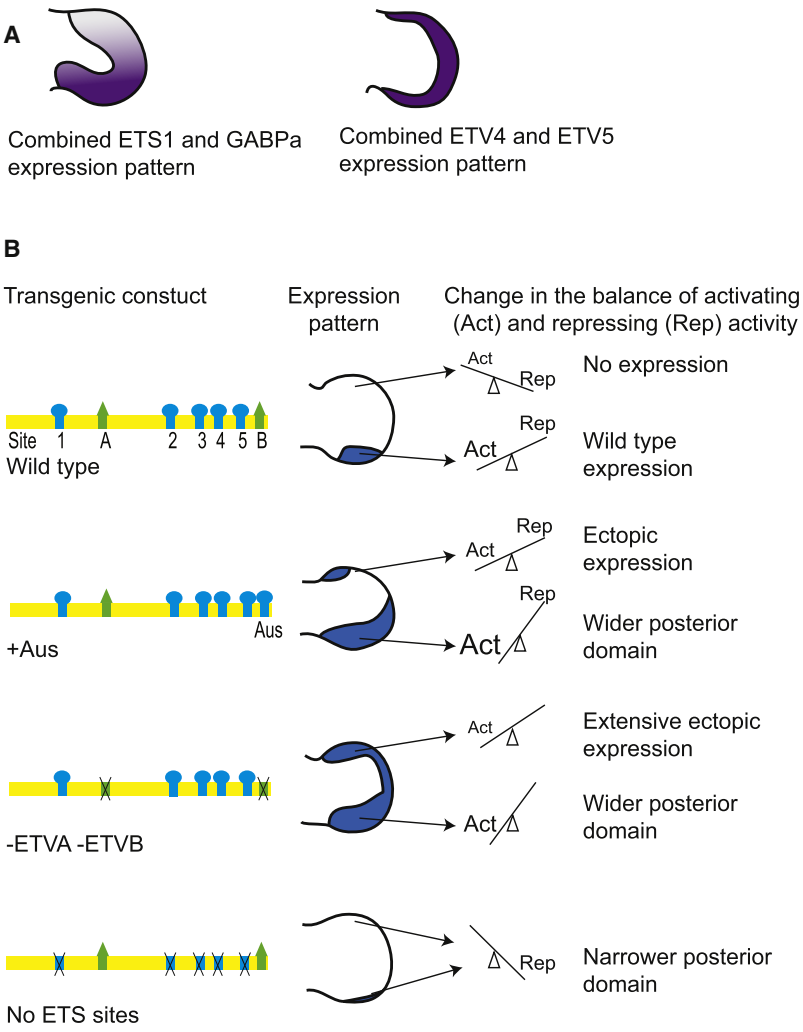

Figure 4. A Model Representing the Fine Balance of ETS Factor Binding and Their Effects on *Shh* Expression

(A) Representation of the expression patterns of the activating ETS factors (ETS1 and GABPα) and the repressing ETS factors (ETV4 and ETV5).

(B) A summary diagram of how four of the transgenic constructs are proposed to interact with the available ETS factors in the limb, with the expression pattern observed for each construct shown in the middle. The change in the balance between activating and repressing activity represented on the right shows the relative balance in the anterior and posterior margins of the limb. The size of the lettering represents the relative amounts of the activating and repressing activities.

role these factors play. ETS1/GABPα directs the position of the *Shh* expression boundary delineating the experimentally defined ZPA. Our data suggest that *Shh* expression is most substantially affected by either site 1 or 3, and it appears that the two sites act cumulatively to achieve the wild-type position of the spatial boundary. Clearly, other sites participate given that, for example, the boundary produced by site 1 is augmented by the presence of sites 2, 4, and 5. The five binding sites do not participate equivalently, and affinity of binding, as detected in the EMSA, is probably part of the explanation. The combinatorial nature of these sites is further supported by the addition of the AUS mutant site in the ZRS, which causes extension of the expression boundary further into the middle of the limb bud. This mutant binding site is the most active of the sites analyzed and, on its own, is capable of driving expression to the approximate boundary of that generated by the wild-type construct. ETS1 and GABPα are expressed at highest levels in the posterior domain of the limb. We suggest that the levels of ETS1/GABPα and the multiple binding sites act in concert to establish a regulatory balance at the ZRS to, first, adjust the position of the *Shh* boundary and, second, in combination with the repressive activity of ETV4/ETV5, to restrict expression to the ZRS (see summary in Figure 4).

Two ETV4/ETV5 binding sites were identified in the ZRS. In transgenics, a single ETV binding site is sufficient to repress ectopic expression; the loss of both sites results in the loss of repressor activity and as a consequence activation of ectopic expression. Previously, expression of both ETV4 and ETV5 in the distal mesenchyme of the limb bud was shown to be maintained by FGF signaling emanating from the AER (Mao et al., 2009; Zhang et al., 2009). FGF is known to be essential for limb outgrowth and maintenance of *Shh* expression. We show that ETV4/ETV5 binding links FGF signaling directly to regulation at the ZRS, showing an unexpected role for FGF, acting through these factors, to repress expression at the anterior margin of the limb.

The AUS mutation has an additional pathogenetic activity, which is to drive ectopic expression in the developing limb bud. We show that the wild-type sites contribute to the ectopic expression but are dependent on the additional activity provided by the extra binding site. Previous data suggest that the ZRS is primed for activity in both the anterior and posterior margins of the limb bud (Amano et al., 2009). In accord, we show that the ZRS is open and fully capable of binding to activating factors such as ETS1 in the anterior domain of the limb bud. As a result, a new, mutant site such as that produced by the AUS mutation would be capable of binding ETS factors at both the normal and the ectopic domains of expression. ETV4/ETV5 is crucial for ensuring that at the primed ZRS, ectopic anterior expression does not occur during limb development (Figure 4). The addition of an extra single high-affinity ETS binding site (as with the AUS mutation) apparently overrides ETV4/ETV5 repression, causing the loss of *Shh* spatial restriction. These molecular events lead to the ectopic expression of *Shh* that underlies the preaxial polydactyly caused by the AC and AUS mutations.

The high conservation throughout the ~770 bp of the ZRS suggests that there is scope for binding a complex mixture of factors. The ETS factor binding sites are most likely functioning along with the binding of other factors at the ZRS. In combination,

these proteins would endow the ZRS with the properties that would not only delineate the boundary but also dictate precise temporal activity. In addition, multiple ETS sites along with other factors may encode an activity that is sufficiently robust to enable long-range recognition and activation of the *Shh* promoter.

EXPERIMENTAL PROCEDURES

Materials

The antibodies used were: IgG (Santa Cruz, sc-2025), ETS1 (Maier et al., 2003), ETS2 (a kind gift from R. M. Roberts), ETV4 (Abcam, ab860902), ETV5 (Abcam, ab102010), GABPα (Santa Cruz, sc-22810), and ELF1 (Santa Cruz, sc-631).

Transgenic Assay and In Situ Hybridization

Transgenic embryos were made and stained in accordance with standard techniques (Lettice et al., 2003), and assembly of mutant ZRS constructs is described in the Supplemental Information. Whole-mount in situ hybridizations were described previously (Hecksher-Sørensen et al., 1998). The *Etv4*, *Etv5*, and *Ets2* probes were transcribed from EST cDNA clones (Geneservice), whereas the *Ets1* probe was generated by RT-PCR and cloned into pZero (Invitrogen). Primers used to amplify Ets1 were 5′-GGAGCATCTAGAGAT CCTGC-3′ and 5′-CAGCCATCTCCTGTCCAGC-3′.

Measuring the Depth of ZRS Staining and Statistical Analysis

Measurement of the extent of expression in each transgenic in the ZRS shown in Figure 3H was measured in Photoshop and calculated as a percentage of the width of the limb bud (to correct for stage differences between the embryos), as shown in Figure 3H.

Statistical comparisons were performed using the statistical package R (http://www.r-project.org/) (R Development Core Team, 2008). For these comparisons, the hind- and forelimb data for each injected construct were combined and a one-way analysis of variance (ANOVA) was used to compare the mean values between each of the 19 groups. The result was highly statistically significant. Therefore, further pairwise post-hoc tests were performed to compare the individual groups using Tukey's HSD (honestly significant difference) test. The table of significance values is found in Table S2.

Electrophoretic Mobility Shift Assays

Nuclear extracts were prepared directly from embryonic limb tissue (E11.5). EMSA analysis is described in detail in the Supplemental Information.

Chromatin Immunoprecipitation and Tiling Microarrays

Cells from dissected E11.5 limbs were fixed with 1% formaldehyde (25°C, 10 min) and stopped with 0.125 M glycine. Crosslinked ChIP was performed as described (Stock et al., 2007). In brief, the nuclei were sonicated using a Diagenode Bioruptor (Leige, full power 30 s on, 30 s off, in an ice bath for 50 min) to produce fragments of <300 bp. Chromatin (350 μg) was incubated with 5 μg prebound (to Protein A or G magnetic beads, Invitrogen) IgG (Santa Cruz, sc-2025) or antibodies raised to ETS1 (Maier et al., 2003), ETV4 (Abcam, ab860902), ETV5 (Abcam, ab102010), GABPα (Santa Cruz, sc-22810), or ELF1 (Santa Cruz, sc-631) in the presence of 50 μg of BSA, washed, and eluted. Reverse crosslinked DNA was purified with Proteinase K (Glenaxxon) and QIAGEN PCR purification kit. ChIP DNA and input DNA were amplified (WGA2 kit, Sigma), labeled, and hybridized according to the manufacturer's protocol to a 3 × 720,000 probe custom microarray containing specific tiled regions encompassing 8.2 megabases of the mouse genome (Nimblegen). The array platform number is GPL14936 and the GEO accession number for the ChIP data is GSE33997.

Microarray data were analyzed in R/Bioconductor (http://genomebiology.com/2004/5/10/R80) with the Epigenome (PROT43) protocol (http://www.epigenome-noe.net/WWW/researchtools/protocol.php?protid = 43) with the following parameters. The mean signal intensity of the four replicate probes on each array was taken. Loess normalization was used within arrays to correct for the dye bias, and scale normalization was used within the replicates group to control interarray variability. The log enrichment for each group was calculated by subtracting the mean of log2 input intensities from the mean of log2 enriched intensities. Probes were tested for significant enrichment using the significance analysis of microarrays (SAM) technique (Tusher et al., 2001), and the local false discovery rate based on the SAM statistic was calculated using the Locfdr R package (Efron, 2007). A false discovery rate of 0.05 was used as the significance cutoff. The median value of each probe was then calculated from a five-probe rolling window to overcome outliers with values that are very different from their neighboring probes.

SUPPLEMENTAL INFORMATION

Supplemental Information includes two figures, two tables, and Supplemental Experimental Procedures and can be found with this article online at doi:10.1016/j.devcel.2011.12.010.

ACKNOWLEDGMENTS

We thank the staff at the Evans Building and especially Emma Murdoch for expert technical assistance. Also, we thank Prof. N. Hastie for critically reading the manuscript. This work was supported by an MRC core grant, and J.H. was supported by NIH R01 grant AI054661.

Received: August 24, 2010
Revised: September 20, 2011
Accepted: December 15, 2011
Published online: February 13, 2012

REFERENCES

Ahn, S., and Joyner, A.L. (2004). Dynamic changes in the response of cells to positive hedgehog signaling during mouse limb patterning. Cell *118*, 505–516.

Amano, T., Sagai, T., Tanabe, H., Mizushina, Y., Nakazawa, H., and Shiroishi, T. (2009). Chromosomal dynamics at the Shh locus: limb bud-specific differential regulation of competence and active transcription. Dev. Cell *16*, 47–57.

Efron, B. (2007). Correlation and large-scale simultaneous significance testing. J. Am. Stat. Assoc. *102*, 93–103.

Farooq, M., Troelsen, J.T., Boyd, M., Eiberg, H., Hansen, L., Hussain, M.S., Rehman, S., Azhar, A., Ali, A., Bakhtiar, S.M., et al. (2010). Preaxial polydactyly/triphalangeal thumb is associated with changed transcription factor-binding affinity in a family with a novel point mutation in the long-range cis-regulatory element ZRS. Eur. J. Hum. Genet. *18*, 733–736.

Fisher, R.J., Koizumi, S., Kondoh, A., Mariano, J.M., Mavrothalassitis, G., Bhat, N.K., and Papas, T.S. (1992). Human ETS1 oncoprotein. Purification, isoforms, -SH modification, and DNA sequence-specific binding. J. Biol. Chem. *267*, 17957–17965.

Furniss, D., Lettice, L.A., Taylor, I.B., Critchley, P.S., Giele, H., Hill, R.E., and Wilkie, A.O. (2008). A variant in the sonic hedgehog regulatory sequence (ZRS) is associated with triphalangeal thumb and deregulates expression in the developing limb. Hum. Mol. Genet. *17*, 2417–2423.

Gotea, V., Visel, A., Westlund, J.M., Nobrega, M.A., Pennacchio, L.A., and Ovcharenko, I. (2010). Homotypic clusters of transcription factor binding sites are a key component of human promoters and enhancers. Genome Res. *20*, 565–577.

Gurnett, C.A., Bowcock, A.M., Dietz, F.R., Morcuende, J.A., Murray, J.C., and Dobbs, M.B. (2007). Two novel point mutations in the long-range SHH enhancer in three families with triphalangeal thumb and preaxial polydactyly. Am. J. Med. Genet. A. *143*, 27–32.

Harfe, B.D., Scherz, P.J., Nissim, S., Tian, H., McMahon, A.P., and Tabin, C.J. (2004). Evidence for an expansion-based temporal Shh gradient in specifying vertebrate digit identities. Cell *118*, 517–528.

Hecksher-Sørensen, J., Hill, R.E., and Lettice, L. (1998). Double labeling for whole-mount in situ hybridization in mouse. Biotechniques *24*, 914–916, 918.

Hill, R.E. (2007). How to make a zone of polarizing activity: insights into limb development via the abnormality preaxial polydactyly. Dev. Growth Differ. *49*, 439–448.

Hollenhorst, P.C., Shah, A.A., Hopkins, C., and Graves, B.J. (2007). Genome-wide analyses reveal properties of redundant and specific promoter occupancy within the ETS gene family. Genes Dev. *21*, 1882–1894.

Hollenhorst, P.C., Chandler, K.J., Poulsen, R.L., Johnson, W.E., Speck, N.A., and Graves, B.J. (2009). DNA specificity determinants associate with distinct transcription factor functions. PLoS Genet. *5*, e1000778.

Lettice, L.A., Horikoshi, T., Heaney, S.J., van Baren, M.J., van der Linde, H.C., Breedveld, G.J., Joosse, M., Akarsu, N., Oostra, B.A., Endo, N., et al. (2002). Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly. Proc. Natl. Acad. Sci. USA *99*, 7548–7553.

Lettice, L.A., Heaney, S.J., Purdie, L.A., Li, L., de Beer, P., Oostra, B.A., Goode, D., Elgar, G., Hill, R.E., and de Graaff, E. (2003). A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. Hum. Mol. Genet. *12*, 1725–1735.

Lettice, L.A., Hill, A.E., Devenney, P.S., and 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*, 978–985.

Lifanov, A.P., Makeev, V.J., Nazina, A.G., and Papatsenko, D.A. (2003). Homotypic regulatory clusters in Drosophila. Genome Res. *13*, 579–588.

Maier, H., Colbert, J., Fitzsimmons, D., Clark, D.R., and Hagman, J. (2003). Activation of the early B-cell-specific mb-1 (Ig-alpha) gene by Pax-5 is dependent on an unmethylated Ets binding site. Mol. Cell. Biol. *23*, 1946–1960.

Mao, J., McGlinn, E., Huang, P., Tabin, C.J., and McMahon, A.P. (2009). Fgf-dependent Etv4/5 activity is required for posterior restriction of Sonic Hedgehog and promoting outgrowth of the vertebrate limb. Dev. Cell *16*, 600–606.

R Development Core Team. (2008). R: A Language and Environment for Statistical Computing (Vienna, Austria: R Foundation for Statistical Computing).

Richardson, L., Venkataraman, S., Stevenson, P., Yang, Y., Burton, N., Rao, J., Fisher, M., Baldock, R.A., Davidson, D.R., and Christiansen, J.H. (2010). EMAGE mouse embryo spatial gene expression database: 2010 update. Nucleic Acids Res. *38* (Database issue), D703–D709.

Ristevski, S., Tam, P.P., Hertzog, P.J., and Kola, I. (2002). Ets2 is expressed during morphogenesis of the somite and limb in the mouse embryo. Mech. Dev. *116*, 165–168.

Ristevski, S., O'Leary, D.A., Thornell, A.P., Owen, M.J., Kola, I., and Hertzog, P.J. (2004). The ETS transcription factor GABPalpha is essential for early embryogenesis. Mol. Cell. Biol. *24*, 5844–5849.

Sagai, T., Hosoya, M., Mizushina, Y., Tamura, M., and Shiroishi, T. (2005). Elimination of a long-range cis-regulatory module causes complete loss of limb-specific Shh expression and truncation of the mouse limb. Development *132*, 797–803.

Semerci, C.N., Demirkan, F., Ozdemir, M., Biskin, E., Akin, B., Bagci, H., and Akarsu, N.A. (2009). Homozygous feature of isolated triphalangeal thumb-preaxial polydactyly linked to 7q36: no phenotypic difference between homozygotes and heterozygotes. Clin. Genet. *76*, 85–90.

Sharrocks, A.D. (2001). The ETS-domain transcription factor family. Nat. Rev. Mol. Cell Biol. *2*, 827–837.

Stock, J.K., Giadrossi, S., Casanova, M., Brookes, E., Vidal, M., Koseki, H., Brockdorff, N., Fisher, A.G., and Pombo, A. (2007). Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. Nat. Cell Biol. *9*, 1428–1435.

Tickle, C. (2006). Making digit patterns in the vertebrate limb. Nat. Rev. Mol. Cell Biol. *7*, 45–53.

Towers, M., and Tickle, C. (2009). Growing models of vertebrate limb development. Development *136*, 179–190.

Towers, M., Mahood, R., Yin, Y., and Tickle, C. (2008). Integration of growth and specification in chick wing digit-patterning. Nature *452*, 882–886.

Tusher, V.G., Tibshirani, R., and Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. Proc. Natl. Acad. Sci. USA *98*, 5116–5121.

Wagner, A. (1999). Genes regulated cooperatively by one or more transcription factors and their identification in whole eukaryotic genomes. Bioinformatics *15*, 776–784.

Wieczorek, D., Pawlik, B., Li, Y., Akarsu, N.A., Caliebe, A., May, K.J., Schweiger, B., Vargas, F.R., Balci, S., Gillessen-Kaesbach, G., and 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*, 81–89.

Xin, J.H., Cowie, A., Lachance, P., and Hassell, J.A. (1992). Molecular cloning and characterization of PEA3, a new member of the Ets oncogene family that is differentially expressed in mouse embryonic cells. Genes Dev. *6*, 481–496.

Yang, Y., Drossopoulou, G., Chuang, P.T., Duprez, D., Marti, E., Bumcrot, D., Vargesson, N., Clarke, J., Niswander, L., McMahon, A., and Tickle, C. (1997). Relationship between dose, distance and time in Sonic Hedgehog-mediated regulation of anteroposterior polarity in the chick limb. Development *124*, 4393–4404.

Zhang, Z., Verheyden, J.M., Hassell, J.A., and Sun, X. (2009). FGF-regulated Etv genes are essential for repressing Shh expression in mouse limb buds. Dev. Cell *16*, 607–613.

Zhu, J., Nakamura, E., Nguyen, M.T., Bao, X., Akiyama, H., and Mackem, S. (2008). Uncoupling Sonic hedgehog control of pattern and expansion of the developing limb bud. Dev. Cell *14*, 624–632.

Developmental Cell
Forum

Enhancers: From Developmental Genetics to the Genetics of Common Human Disease

Iain Williamson,[1] Robert E. Hill,[1] and Wendy A. Bickmore[1,*]
[1]MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, Crewe Road, Edinburgh EH4 2XU, UK
*Correspondence: w.bickmore@hgu.mrc.ac.uk
DOI 10.1016/j.devcel.2011.06.008

In mammals, long-range gene regulation became apparent through simple Mendelian disease genetics in human and developmental genetics in the mouse. Can the insights into gene control, provided by the study of these enhancers, help us understand the functional significance of sequence variation associated with common/complex human disease and quantitative traits?

Enhancers are *cis*-regulatory DNA sequences capable of inducing transcription from a distant promoter and can be located either upstream or downstream of the activated gene. A plethora of such candidate *cis*-regulatory regions are being identified through their chromatin signatures, including DNase hypersensitivity, histone variants (H2A.Z), histone modifications (histone H3 lysine 4 mono- or dimethylation, histone H3 lysine 27 acetylation), bound cofactors (p300, mediator), and even RNA polymerase II itself. Functional studies are still in their infancy, but tissue-specific enhancer activity for a number of noncoding sequences has been demonstrated using transgenic assays. Sequence conservation across species is also a guide to the genomic location of some, but not all, enhancers (Blow et al., 2010).

The Role of Distal Enhancers in Evolution and Development

How did such *cis*-regulatory sequences evolve? Whereas simple elements that bind few factors may have arisen de novo by small base changes, it seems unlikely that complex enhancers could have arisen this way. Possible routes are the exaptation of retroposons to form clusters of protein-binding motifs at new genomic locations (Santangelo et al., 2007; Sasaki et al., 2008) or the acquisition of novel regulatory functions by promoters of archaic genes whose coding sequences have been subsequently lost. Existing enhancers may also evolve to acquire new functions in development.

What are the evolutionary consequences of mutation and/or deletion of existing enhancers? Whereas alteration of coding regions may have quite blunt effects on protein function, alterations at enhancers have the potential to subtly adjust levels of gene expression—a situation required for the evolution of quantitative traits and response to environmental stresses. Moreover, phenotypic effects of mutating enhancers can be confined to discrete tissues or developmental stages and so could allow a preexisting gene to be harnessed to a new developmental scenario (Rebeiz et al., 2009). For instance, neural-specific activation of *Lunapark* and *Evx2*, situated between the HoxD cluster and the global control region (GCR), is regulated by enhancer elements within the GCR. Co-option of the GCR to then facilitate *HoxD13-10* expression during limb development may have been due to the accrual of mutations to produce an element able to function in the developing limb (Gonzalez et al., 2007). A limb-specific enhancer of *Prx1*—which promotes bone elongation in the forelimb—is responsible for higher transcription levels and more extensive expression domains at key limb bud locations and embryonic stages in the fruit bat (Cretekos et al., 2008). Replacing the endogenous mouse enhancer with the bat equivalent produced mouse embryos with longer forelimbs but no morphological effects in other organs where *Prx1* is expressed. Similarly, in transgenic mouse assays, human-specific point mutations within a highly conserved noncoding element (HACNS1) confer a limb expression pattern, including the presumptive anterior wrist and proximal thumb, that appears human-specific compared to that conferred by orthologous sequences from nonhuman primates (Noonan and McCallion, 2010). Human-specific deletions, of mostly noncoding sequences that are highly conserved between chimpanzees and other mammals, include regions near the androgen receptor locus and the neural-specific *GADD45G* (McLean et al., 2011). These contain tissue-specific enhancers whose loss correlates with human-specific traits: loss of secondary sexual characteristics associated with androgen signaling and the expansion of specific brain regions.

The Genomic Context of Enhancers

A sizeable proportion of enhancers are situated tens to hundreds (even thousands) of kilobases from their target genes, often in large gene-poor regions. Indeed, many genes with tight spatial and/or temporal expression domains tend to be located adjacent to such gene deserts. The absence of other neighboring genes could help ensure regulatory specificity between enhancer and promoter. An example of problems that can arise from the promiscuous action of enhancers placed within gene-rich domains is illustrated by the ectopic activation of neighboring genes when the β-globin locus control region (LCR) is integrated into a gene-rich domain of housekeeping genes (Noordermeer et al., 2008).

However, there must be specificity between some enhancers and promoters. The sonic hedgehog gene (*Shh*), which codes for a morphogen directing cell fate in several tissues during organogenesis, is adjacent to a gene desert containing at least three enhancers that are active only within *Shh*-expressing tissues of the embryonic brain. However, *Shh* expression during limb development is regulated by another enhancer, the *z*one of polarizing activity *r*egulatory *s*equence (ZRS),

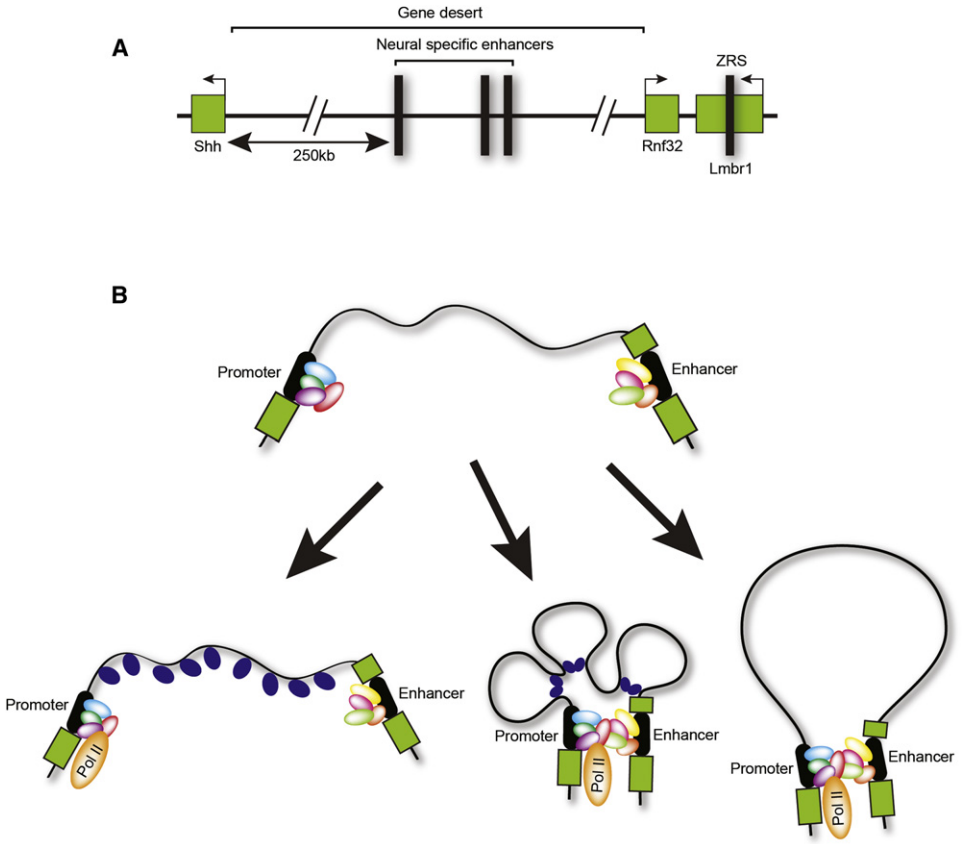

Figure 1. The Sonic Hedgehog Regulatory Region and Possible Mechanisms of Enhancer Activity
(A) Linear representation of genes (green boxes) at the *Shh* locus showing the location of long-range neural-specific and limb-specific enhancers (black boxes). (B) Models of enhancer-promoter communication. Factors bound to the enhancer and promoter recruit chromatin-modifying factors that: reorganize the intervening chromatin, with the factors themselves being the signal for gene activation (left), enable direct enhancer-promoter interaction by the formation of chromatin "mini-loops" (middle), or allow for spatial colocalization of the enhancer and promoter, with looping out of the intervening chromosomal region (right).

situated about 1 megabase away, beyond the neural enhancers and within an intron of another gene (Lettice et al., 2003) (Figure 1A). The ZRS bypasses both the gene in which it is located and a neighboring gene in order to activate only *Shh*.

How Do Long-Range Enhancers Activate Target Genes?

Enhancers are clusters of transcription factor (TF) binding sites. There have been three, not entirely mutually exclusive, mechanisms proposed for enhancer-gene communication (Figure 1B). In the first, sequence-specific factors bound to the enhancer and/or promoter recruit further factors—perhaps even including RNA polymerase—to reorganize chromatin throughout the intervening region as the signal for gene activation (Bulger and Groudine, 2010). Due to the large genomic distances involved, and even intervening promoters, this is unlikely to be a feasible mechanism for very long-range control, such as that at *Shh*. The second mechanism also involves reorganizing the intervening chromatin, but to allow for direct enhancer-promoter interaction, possibly through a series of "mini" chromatin loops.

The third mechanism invokes the spatial colocalization in the nucleus of the enhancer and promoter, which can then directly interact with each other if the necessary factors are bound, looping out the intervening region. An attractive feature of this model is that it ensures that an enhancer can activate the expression of only one gene at a time. However, enhancer-promoter mediated looping may be insufficient for gene activation. Nuclear colocalization of the ZRS and *Shh* is observed in both expressing and nonexpressing halves of the developing limb bud, but on the posterior side, the active *Shh* locus is additionally extruded from its chromosome territory (Amano et al., 2009). A similar looping out from chromosome territories is induced by an ectopic β-globin LCR (Noordermeer et al., 2008).

Does there need to be direct physical interaction between enhancer and promoter for activation of gene expression? Clustered TF binding sites at enhancers might simply act to reduce the effective nuclear search space of these proteins for a promoter by diffusion/nonspecific binding. Enhancer-promoter specificity would then rely upon shared high-affinity binding sites for the same set of factors. Moreover, enhancers themselves often bind RNA polymerase II (De Santa et al., 2010) and produce short noncoding (nc) RNAs (Kim et al., 2010). Whether all such ncRNAs have a function in vivo is still to be determined.

Enhancers and Disease

Mutations associated with β-thalassaemia, an inherited disorder caused by altered expression of globin genes, were what first led to the identification of the β-globin LCR. Deletions that affect an enhancer's function, or translocations that separate enhancers from their cognate promoters, are also the cause of several simple Mendelian human genetic diseases (Noonan & McCallion, 2010). Point mutations in the *Shh* ZRS in preaxial polydactyly result in a gain of function of enhancer activity and ectopic activation of *Shh* expression in the anterior limb bud (Lettice et al., 2003).

However, almost half of single nucleotide polymorphisms (SNPs) that show statistical associations with common/complex human disease and quantitative traits in genome-wide association studies (GWAS) are within noncoding regions and gene deserts and thus potentially involve enhancers (Noonan and McCallion, 2010). For example, a SNP within an intronic enhancer of *RET* confers a greater risk of Hirschsprung disease than the major allele. SNPs in a gene desert 330 kb upstream of the *c-myc* protooncogene are associated with greater risk of several cancers (Wasserman et al., 2010). Why might so much of the genetic variation apparently associated with complex human disease be in enhancers? Whereas point mutations in protein coding regions can completely ablate gene function and so produce strong and penetrant phenotypes, point mutations in an enhancer may alter binding affinity for a specific TF and so result in only subtle changes in the level, time, or place of gene expression. The resulting phenotype could therefore be more modest and subject to interaction with other genes and with the environment.

Future Prospects

Our current lack of mechanistic understanding about how enhancers regulate temporal and spatial patterns of expression from the correct target gene is impeding exploitation of the investments made in GWAS and the functional analysis of human disease. What needs to be done to better understand how enhancers work? The dearth of knowledge on enhancer biochemistry needs to be addressed in order to gain a more complete understanding of how enhancers integrate and process signaling information that promotes transcription. Enhancer elements function as nucleoprotein complexes, with multiple associated sequence-specific and nonspecific binding proteins. What are these proteins and how do they act together? There could be a high level of cooperative and coordinate action between the bound TFs to give unitary outputs to the transcriptional machinery, or they may instead form multiple functional units that each independently regulate gene expression. In the former case, mutation of individual factor binding sites may ablate enhancer function, whereas in the latter case they would be less detrimental to enhancer function.

The precise chronological sequence of chromatin events at enhancers is also to be elucidated. Do direct physical enhancer-promoter interactions only occur once factors are bound to both elements, or does the interaction itself contribute to the loading of TFs or transcriptional machinery to the promoter? Current assays for long-range physical interactions—fluorescence in situ hybridization (FISH) and chromosome conformation capture (3C) and its derivatives—only give snapshots or cell population-averaged information, respectively. There is a need for dynamic assays of chromosome organization to capture long-range interactions as they occur in real time. There is also a need for these techniques to be coupled to methods for quantifying levels of gene expression in single cells at defined times and places in differentiation and development.

REFERENCES

Amano, T., Sagai, T., Tanabe, H., Mizushina, Y., Nakazawa, H., and Shiroishi, T. (2009). Dev. Cell *16*, 47–57.

Blow, M.J., McCulley, D.J., Li, Z., Zhang, T., Akiyama, J.A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., et al. (2010). Nat. Genet. *42*, 806–810.

Bulger, M., and Groudine, M. (2010). Dev. Biol. *339*, 250–257.

Cretekos, C.J., Wang, Y., Green, E.D., Martin, J.F., Rasweiler, J.J., 4th, and Behringer, R.R.; NISC Comparative Sequencing Program. (2008). Genes Dev. *22*, 141–151.

De Santa, F., Barozzi, I., Mietton, F., Ghisletti, S., Polletti, S., Tusi, B.K., Muller, H., Ragoussis, J., Wei, C.L., and Natoli, G. (2010). PLoS Biol. *8*, e1000384.

Gonzalez, F., Duboule, D., and Spitz, F. (2007). Dev. Biol. *306*, 847–859.

Kim, T.K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., et al. (2010). Nature *465*, 182–187.

Lettice, L.A., Heaney, S.J., Purdie, L.A., Li, L., de Beer, P., Oostra, B.A., Goode, D., Elgar, G., Hill, R.E., and de Graaff, E. (2003). Hum. Mol. Genet. *12*, 1725–1735.

McLean, C.Y., Reno, P.L., Pollen, A.A., Bassan, A.I., Capellini, T.D., Guenther, C., Indjeian, V.B., Lim, X., Menke, D.B., Schaar, B.T., et al. (2011). Nature *471*, 216–219.

Noonan, J.P., and McCallion, A.S. (2010). Annu. Rev. Genomics Hum. Genet. *11*, 1–23.

Noordermeer, D., Branco, M.R., Splinter, E., Klous, P., van Ijcken, W., Swagemakers, S., Koutsourakis, M., van der Spek, P.A., Pombo, A., and de Laat, W. (2008). PLoS Genet. *4*, e1000016.

Rebeiz, M., Pool, J.E., Kassner, V.A., Aquadro, C.F., and Carroll, S.B. (2009). Science *326*, 1663–1667.

Santangelo, A.M., de Souza, F.S.J., Franchini, L.F., Bumaschny, V.F., Low, M.J., and Rubinstein, M. (2007). PLoS Genet. *3*, 1813–1826.

Sasaki, T., Nishihara, H., Hirakawa, M., Fujimura, K., Tanaka, M., Kokubo, N., Kimura-Yoshida, C., Matsuo, I., Sumiyama, K., Saitou, N., et al. (2008). Proc. Natl. Acad. Sci. USA *105*, 4220–4225.

Wasserman, N.F., Aneas, I., and Nobrega, M.A. (2010). Genome Res. *20*, 1191–1197.