A Dynamic Chromatin Architecture Modulates *Pitx1* Gene Regulation in Limb Development and Disease

Inaugural-Dissertation
to obtain the academic degree
Doctor rerum naturalium (Dr. rer. nat.)
submitted to the Department of Biology, Chemistry and Pharmacy
of Freie Universität Berlin

by

BJØRT KATRINARDÓTTIR KRAGESTEEN

from Tórshavn, The Faroe Islands

Berlin

2017

The present work was conducted from October 2013 to October 2017 at the Max Planck Institute for Molecular Genetics in Prof. Dr. Mundlos' research group.

1st Reviewer: Prof. Dr. Stefan Mundlos

2nd Reviewer: Prof. Dr. Sigmar Stricker

Date of defence: 29.01.2018

Table of contents

Table of contents

Table of contents

ACKNOWLEDGEMENTS

1. INTRODUCTION

1.1. DEVELOPMENTAL GENE REGULATION

A key feature of the vertebrate genome is the burst in pleiotropy, whereby genes perform multiple functions at different times and place during embryogenesis (Montavon and Duboule, 2012). Alongside pleiotropy emerged the necessity of tight regulation of promoters and cognate *cis*-regulatory elements scattered over hundreds of kilobases (kb). This regulatory information is encoded in several layers, from DNA-histone interactions to chromatin-chromatin interactions, and chromosome territories within the nucleus. These interactions and organisation has facilitated several control steps and a framework for the expansion of the diversity of lineage specific gene expression to build complex tissues such the embryonic limb (Andrey and Mundlos, 2017; Petit et al., 2017). In the next sections, I will introduce the regulation of transcription by *cis*-regulatory elements in the light of developmental regulatory landscapes.

1.1.1 Promoters

Promoters consist of core promoter elements and regulatory elements. Core promoters are defined as short stretches of DNA surrounding the transcriptional start site (TSS) 5' of a gene that are able to directly recruit RNA polymerase II and initiate transcription. Through the binding of general transcription factors (GTFs) to the core promoter, the preinitiation complex (PIC) is formed and guides the RNA polymerase II to the TSS of the gene and dictate the initiation of transcription (Thomas and Chiang, 2006; Zabidi and Stark, 2016). The core promoter alone cannot uphold efficient transcription on its own and requires the input of proximal and distal regulatory elements, such as enhancers. The activity of these regulatory elements is determined through the combinatorial binding of active and repressive DNA binding factors (transcription factors (TFs)) that in turn recruit co-factors and histone modifiers (Zabidi and Stark, 2016). Together they communicate with target core promoter to produce robust tissue specific or broad expression patterns, depending on the sequence specificity of the cognate enhancer-promoter pair (Nogales et al., 2017; Spitz and Furlong, 2012). Typically, tissue specific promoters are highly conserved, enriched in TATA box consensus sequences and CpG

islands, and embedded in large regulatory landscapes encompassing numerous *cis*-regulatory elements (Danino et al., 2015; Zabidi et al., 2015) (Figure 1A). These promoters bear characteristic marks of histone modification. For instance, the promoters of developmental genes primed for activity are marked by H3K27me3 Polycomb repressive mark and H3K4me3 activation mark in embryonic stem cells (ESCs) (Lesch et al., 2013). This bivalent mark of promoters enables fast activation and robust tissue specific expression during development (Bernstein et al., 2006; Ferrai et al., 2017). On the other hand, the promoters of ubiquitously expressed genes, also called housekeeping genes, are characterised by poor conservation, enrichment of Ohler motifs, and are typically regulated by proximal enhancers resulting in constitutive expression (Figure 1A) (Farre et al., 2007; Juven-Gershon and Kadonaga, 2010; Zabidi et al., 2015). The specificity of gene expression is thought to be encoded within the core promoters and enhancers that recruit tissue specific *trans*-acting factors mediating tissue specific communication between the two (Zabidi and Stark, 2016). In summary, developmental and housekeeping promoters form two kinds of regulatory landscapes that enable the tissue specific transcriptional output or broad expression across cells and tissues.

1.1.2 Silencers and insulators

The initiation and maintenance of the developmental program of the embryo requires the precise regulation of gene expression in time and space. The temporospatial regulation of developmental promoters thus requires additional regulators such as silencers, insulators, and enhancers to achieve tissue specific expression. Classically, silencers are defined as sequence specific stretches of DNA that bind TFs and result in repression of target enhancers and promoters (Figure 1A) (Maeda and Karch, 2011). At the *Grem1* locus, the regulatory element GRE1 functions both as an enhancer and as a silencer, exerting its activating or silencing function through the differential binding of GLI activators or repressors, respectively (Li et al., 2014). Thereby *cis*-regulatory elements can exhibit opposing roles enabling flexibility in the transcriptional output of target promoters.

Insulators are DNA elements that bind insulator proteins to block the effect of enhancers and silencers on target promoters (Figure 1B) (Bell et al., 1999; Gaszner and

Introduction

Felsenfeld, 2006). CCCTC-binding factor (CTCF) is a highly conserved and broadly expressed zinc-finger DNA binding factor that confers both transcriptional activation and insulation by binding to CTCF motifs (Ghirlando and Felsenfeld, 2016; Kim et al., 2015). In addition, CTCF is a key architectural protein thought to operate together with cohesin to facilitate physical loops between enhancers and cognate promoters (Bonev and Cavalli, 2016; Nora et al., 2017). Thus, silencers and insulators form building blocks used in gene regulation to fine-tune existing regulatory activity either to silence it or to insulate and promote cognate enhancer-promoter pairs.

1.1.3 Enhancers and transcription factors

Long range transcriptional enhancers are key players in promoting accurate and robust expression of developmental genes. Enhancers are gene regulatory DNA sequences that typically span a few hundred base pairs (bp) and characteristically act independent of their location, position, and orientation to activate target promoters in a cell type-, time-, or tissue-specific manner (Figure 1B) (Hu and Tee, 2017). Enhancer sequences are enriched in TF motifs that enable the differential binding of tissue specific TFs that in turn recruit co-factors and together dictate the communication of the enhancer with target promoters (Hu and Tee, 2017; Petit et al., 2017). Active enhancers can be recognised by their enrichment in H3K27ac deposited by histone acetylases (HATs) such as p300 (Visel et al., 2009).

TFs are proteins with a DNA binding domain typically recognising 6-12 bp long degenerate DNA sequences and bind with different affinities (Spitz and Furlong, 2012). Enhancers recruit activating and repressive TFs through motifs to form a distinctive binding combination enabling diverse temporal and spatial regulatory output that is communicated to target promoters via physical chromatin interactions (Figure 1B) (Krijger and de Laat, 2016; Spitz and Furlong, 2012). Moreover, multiple enhancers can regulate a single gene, thereby resulting in an immense combinatorial flexibility and gene expression patterns (Pennacchio et al., 2006; Spitz and Furlong, 2012). In this scenario, several enhancers can have unique or overlapping activities that in combination achieve an additive effect to convey pleiotropic expression patterns of developmental genes such as *Ihh* and genes of the *HoxD* cluster (Montavon et al., 2011; Will et al., 2017).

Introduction
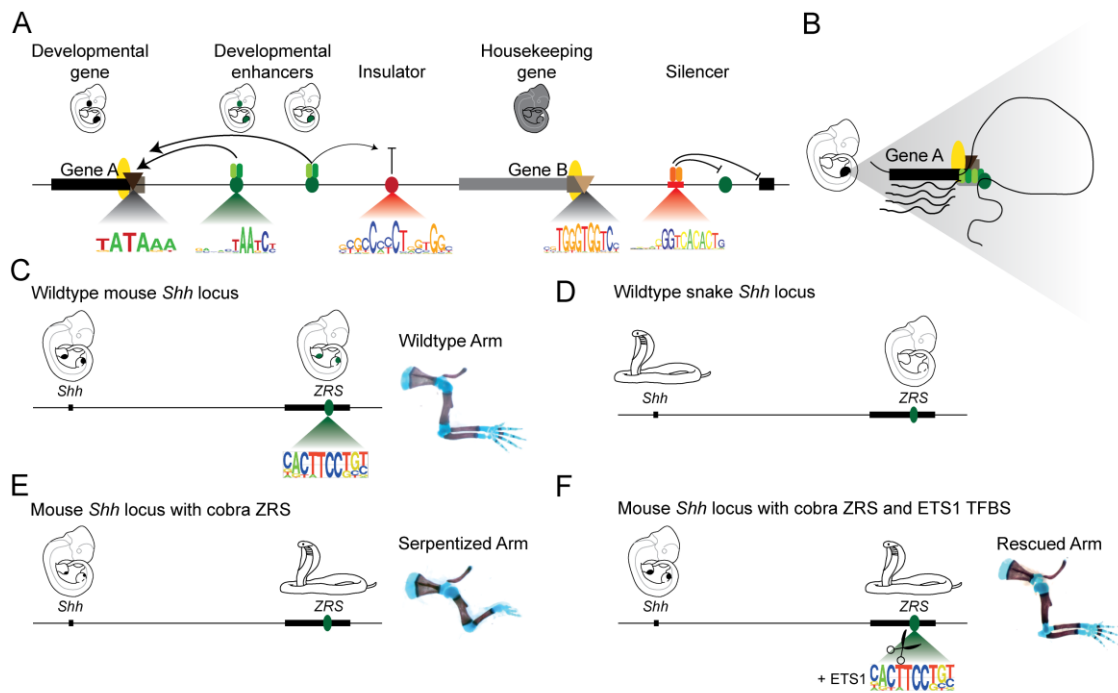
Figure 1 Overview of gene regulation

A. From left: Tissue specific developmental core promoters are enriched in TATA consensus sequences bound by GTFs (triangle) forming the PIC (square) and correct positioning of RNAP II (oval). They are regulated by distal developmental enhancers (green circles) bound by tissue specific TFs (green rounded rectangles). Insulators (red circle) block the promiscuous action of enhancers. Ubiquitously expressed housekeeping promoters are enriched in Ohler consensus sequences. Silencers (red rounded rectangle) bind tissue specific TFs to silence enhancers and promoters. **B.** Binding of TFs and mediator to developmental core promoters and distal enhancers mediates physical looping to enable proximity and full action of enhancers on promoters. **C.** Wildtype *Shh* locus in mouse where distal ZRS enhancer confers *Shh* expression in ZPA in the limb to ensure normal development the limb. **D.** In snakes, the ZRS is mutated and unable to activate *Shh*, resulting in the absence of limbs. **E.** Replacing the mouse ZRS sequence with the cobra ZRS results in severe reduction in *Shh* expression and truncation of the limb. **F.** The limb phenotype is rescued in mice when a single TFBS ETS1 is knock into the cobra enhancer in the mouse.

Mutations within enhancers have been shown to alter the combination of TFs binding or even abrogate their binding, modifying the transcriptional output of target promoters. This disruption or modification of developmental building blocks play a major role in disease and contribute to evolutionary changes in morphologies. An example of this is the regulation of *Shh* in the embryonic limb. *Shh* is expressed in the zone of polarising activity (ZPA) of the posterior limb and a single enhancer ZRS confers this tissue specific expression. Mutations disrupting the enhancer activity leads to a limb malformation in several vertebrate species including humans (Hill and Lettice, 2013; Lettice et al., 2003). For instance, snakes grow rudimentary limb buds at the embryonic stages, but the failure in its maintenance leads to their regression as a result of early arrest of *Shh* expression (Cohn and Tickle, 1999; Lawson et al., 2005). Replacing the mouse ZRS enhancer with the cobra ZRS enhancer results in strong reduction of *Shh*

expression and severe truncation of the limbs (Figure 1E) (Kvon et al., 2016). Strikingly, the phenotype is rescued when a single TF motif for transcriptional activator ETS1 is inserted into the cobra ZRS enhancer in the mouse, which form normal limbs (Kvon et al., 2016) (Figure 1F). Thus, in evolution, the sequence changes in non-coding regulatory regions enable the alteration of tissue morphologies and suggest that the loss of limb in snakes may have been positively selected for the improved locomotion in snakes, without risking the permanent change in the gene sequence and protein product (Kvon et al., 2016). However, the loss of limbs in snakes could also have been a result of a negative selection pressure in upstream molecular events or through genetic drift (Kvon et al., 2016).

1.1.4 Identification of enhancer elements

Identification of transcriptional enhancers is a key step in understanding the regulation of target genes. To classify putative enhancer elements, publically available genome wide data such as conservation, active enhancer mark H3K27ac, physical chromatin interactions, computational predictions of TF binding sites (TFBS), and local opening of chromatin, can be used (King et al., 2007; Visel et al., 2007). These methods have estimated the existence of millions of enhancer elements encoded in vertebrate genomes (Mouse et al., 2012). The enhancer activity on a region can be tested using transgenic assays, in which a reporter gene, such as *LacZ,* is under the control of a weak minimal promoter and is only activated if surrounding sequences encode a regulatory potential (Nobrega et al., 2003). Thousands of enhancers have been identified using reporter assays where the majority reflect the activity of surrounding genes (Visel et al., 2007). Nevertheless, several enhancers display activity inconsistent with that of the expression of nearby genes, and are termed "orphan enhancers" (Visel et al., 2007). An alternative way of probing the endogenous enhancer activity is by integrating regulatory sensors into an endogenous locus, whereby the *LacZ* cassette sense the surrounding regulatory information without affecting local promoters (Ruf et al., 2011; Symmons et al., 2016). Together, these techniques enable the identification and testing of enhancer sequences both exogenously (*LacZ* reporter assay) and endogenously (regulatory sensors).

Introduction

1.2. 3D-ORGANISATION OF THE VERTEBRATE GENOME

Tissue specific enhancers and promoters can be identified through the analysis of epigenetic marks and transcriptional profiles. However, understanding how the dynamic pairing of cognate enhancer and promoter occurs, the 3D organisation of the chromatin fiber needs to be considered. Various technologies of studying the 3D organisation of the genome have been developed, such as microscopy based methods assessing the radial positioning of regions of interest (Beagrie et al., 2017; Pombo and Dillon, 2015). However, microscopy methods are low throughput since single cells are used (Pombo and Dillon, 2015). In the last decade, chromosome conformation capture (3C)–based approaches have been developed that enable assessment of the average chromatin conformation across millions of cells shedding a new light on the shape and complexity of developmental regulatory landscapes (Krijger and de Laat, 2016; Montavon and Duboule, 2012).

1.2.1 Chromosome conformation capture

3C-technologies measure the proximity of contact frequency between chromatin regions in the nucleus (Dekker et al., 2002). A snapshot of a cell population average 3D genome architecture is obtained with the chemical fixing millions of nuclei with formaldehyde tht crosslinks DNA-DNA and DNA-proteins within 2 Å (Solomon and Varshavsky, 1985). The chromatin is subsequently fragmented using restriction enzymes, followed by in situ proximity ligation of ends producing chimeric DNA-DNA products (3C-library) (Figure 2A) (Dekker et al., 2002). Various approaches have been developed based on this principle, and in this work, circular 3C sequencing (4C-seq), Capture C (cC), and Capture Hi-C (cHi-C) are used. The contact frequencies of chimeric DNA fragments are represented using 1D-, 2D-heatmaps, or simulated 3D-polymer models based on Hi-C data (Krijger and de Laat, 2016). 4C-seq and Capture-C enable genome-wide measurement of chromatin contact frequencies from a chosen viewpoint (Hughes et al., 2014; van de Werken et al., 2012). In 4C-seq, the viewpoint is PCR amplified using primers hybridising in the restriction fragment of interest to amplify all ligation products, which are subsequently quantified with next-generation sequencing (NGS) (Figure 2B) (Chiariello et al., 2016; Krijger and de Laat, 2016). In Capture-C, targeted capture of viewpoint of interest from the 3C library is possible by using RNA

6

Introduction

oligos enriching i.e. promoter of interest, followed by NGS (Figure 2C) (Hughes et al., 2014). Hence hundreds of viewpoints can be produced from a single 3C library.
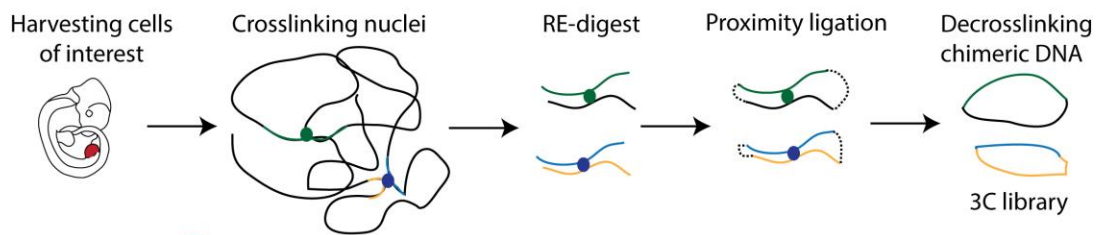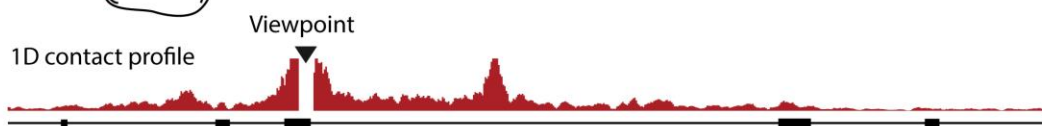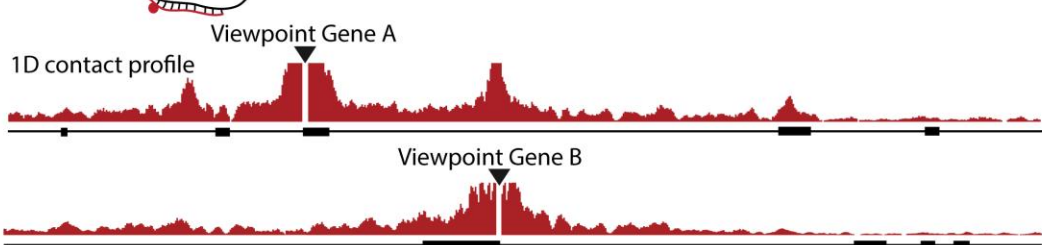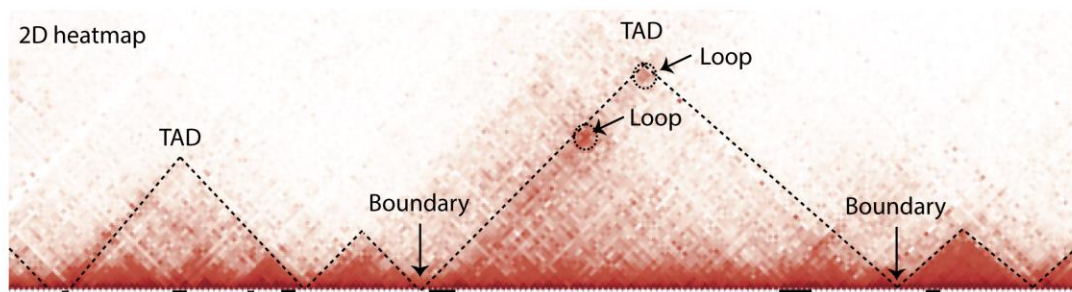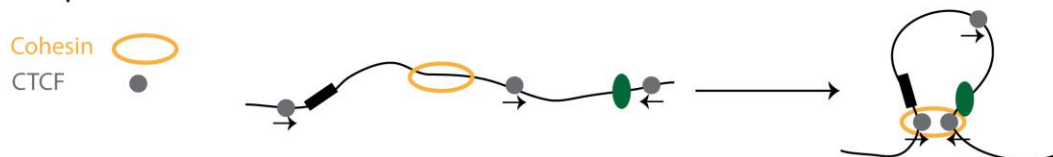
Figure 2 C-techniques and chromatin architecture

A. A 3C-library is produced by formaldehyde crosslinking of nuclei, followed by RE-digestion, proximity ligation, and decrosslinking of chimeric DNA fragments. **B.** In 4C-seq, viewpoint of interest is chosen and PCR amplified from 3C-library followed by NGS. Chromatin interactions are visualised using 1D maps. **C.** In Capture-C, viewpoint of interest is captured from 3C-library using RNA based probes followed by NGS. Chromatin interactions are visualised using 1D maps. **D.** In Capture-Hi-C millions of RNA oligoes are used to capture e.g. loci of interest from 3C library producing high-resolution complex 2D heatmaps. **E.** Loop extrusion model: cohesin is loaded onto the DNA and extruding it through its ring-like structure forming loops. When it reaches convergent CTCF bound regions, the extrusion is blocked forming a stable loop.

Introduction

Hi-C combines 3C with high throughput NGS enabling detection of genome wide interactions (van Berkum et al., 2010). A derivative of Hi-C is Capture Hi-C (cHi-C), where a genomic region of interest is captured using RNA oligos (Mifsud et al., 2015) (Figure 2D). Hi-C and cHi-C 2D-heatmaps illustrate the organisation of chromatin into domains that show preferential intra-domain interactions, separated by boundary regions that show a drop in inter-domain interactions (Dixon et al., 2012). The intensity of the colour scheme correlates with contact frequency between two genomic regions (Figure 2D). The domains and boundaries are computed using bioinformatical tools based on a change in chromosome interaction directionality and insulation score between neighbouring domains (Dixon et al., 2012). Together with microscopy studies in hundreds of cell types, tissues, and species, the C-technologies have built a solid ground of data demonstrating a highly organised and regulated folding of the genome within the nucleus (Beagrie et al., 2017; Bonev and Cavalli, 2016; Krijger and de Laat, 2016).

1.2.2 Structure of chromatin domains

Initial studies of chromosome folding using Hi-C revealed the segregation of chromosomes into median 850 kb sized self-associating chromosomal domains, named Topologically Associating Domains (TADs), whose boundary regions insulate TADs from each other (Dixon et al., 2012; Nora et al., 2012). TADs provide a framework that favour cognate enhancer-promoter contacts and insulates their regulatory activity from neighbouring TADs (Franke et al., 2016; Lupianez et al., 2015; Symmons et al., 2016). TADs and their boundaries have been demonstrated to be cell and tissue invariable, conserved throughout evolution, and correlating with replication domains and lamina associating domains, signifying their function as a basic scaffold of genome organisation (Acemel et al., 2017; Dixon et al., 2015). The increased resolution of chromatin interactions has revealed tissue specific dynamic formation of smaller domains within TADs, called subTADs and chromatin domains (Berlivet et al., 2013; Phillips-Cremins et al., 2013; Rao et al., 2014). Moreover, loop formation between TAD boundaries as well as enhancer and promoter pairs, is evident as small dots on top of two interactions on a Hi-C heatmap (Figure 2D) (Rao et al., 2014). In addition, TAD boundary regions are enriched in CTCF, tRNA genes, TSS, and housekeeping genes, suggesting that

architectural proteins and transcriptional activity play a role in the formation and maintenance of TADs and boundaries (Dixon et al., 2012). Thus, the discovery and characterisation of TADs has revolutionised the understanding of gene regulation, revealing the remarkable segregation of the genome into regulatory territories that shield and enable cognate enhancer and promoter pairs.

1.2.3 Architectural proteins

The nature of TAD formation and their causality is still being heavily investigated (Dixon et al., 2016). A model based on in silico modelling of Hi-C data suggests that chromatin loops are formed by loop extrusion factors that extrude the DNA until being physically blocked by architectural proteins (Fudenberg et al., 2016; Sanborn et al., 2015). The ring shaped cohesin complex has been suggested to translocate on the DNA to initiate extrusion of the chromatin resulting in the formation of chromatin loops (Figure 2E) (Fudenberg et al., 2016; Sanborn et al., 2015). The size of these loops is thought to be determined by CTCF bound regions which form "road blocks" stalling cohesin and thereby stabilising the complex (Figure 2E) (Fudenberg et al., 2016; Sanborn et al., 2015). CTCF exhibits a directionality, whereby the protein recognises the CTCF motifs and subsequently binds in distinct combinations of its zinc finger domains, and recruits cohesin through its C-terminal domain (Xiao et al., 2011). Thus, two CTCF bound motifs in forward-reverse orientation block the cohesin complex to stabilise the loop (Xiao et al., 2011). When CTCF sites are inverted or deleted, the loop expands until the next convergent orientation is located, resulting in rewiring of enhancer-promoter interactions (Guo et al., 2015; Lupianez et al., 2015). Thus, the loop extrusion model explains how chromatin loops are formed, and why not every bound CTCF site engages in long range interactions (Fudenberg et al., 2016; Sanborn et al., 2015).

The conditional degradation of CTCF in mESCs results in a dose-dependent insulation defect and loss of chromatin loops (Nora et al., 2017). Consequently, insulation between TADs is reduced and inter-TAD interactions increase (Nora et al., 2017). Conditional deletion of cohesin subunits in mouse liver or human colon cancer cell line result in a genome-wide disappearance of TADs and loops leading to disruption of long-range enhancer promoter interactions and global genome decompaction (Rao et al.,

Introduction

2017; Schwarzer, 2017). These studies support CTCF and cohesin as key architectural proteins in mediating chromatin loops and forming TADs.

Recent studies have shown the correlation between the changes in enhancer and transcriptionally activity, and thereby underlying functional chromatin, with the dynamic changes in chromatin conformation across development and cell differentiation (Andrey et al., 2013; Andrey et al., 2017; Freire-Pritchett et al., 2017). Specifically, the genome segregates functional chromatin into preferentially interacting active H3K27ac domains and Polycomb H3K27me3 repressive domains, both locally and genome wide forming chromatin domain and A/B compartments, respectively (Andrey et al., 2017; Lieberman-Aiden et al., 2009; Schoenfelder et al., 2015). From these studies, different mechanisms of genomes organisation have been put forward. The first is a cohesin-CTCF dependent formation of TADs and loops to pair long-range gene regulation (Schwarzer, 2017). The second is a cohesin-CTCF independent formation of chromatin architecture directed by the underlying epigenetic landscape permitting both local and global dynamic (Andrey et al., 2017; Bonev et al., 2017; Schoenfelder et al., 2015; Schwarzer, 2017).

1.2.4 Function of chromatin domains

The functional relevance of TADs has been examined both *in vitro* and *in vivo*. Disruption of TADs or deletion of a TAD boundaries result in reduced insulation and increased chromatin interaction between neighbouring TADs, with consequent ectopic gene activation, resulting in congenital malformations and cancer (Franke et al., 2016; Hnisz et al., 2016; Lupianez et al., 2015). These studies support the function of TADs as a regulatory territory, and when breached erroneous enhancer-promoter pairing is enabled. This is elegantly demonstrated at the *Shh* locus, where the gene and its ZRS enhancer are located within the same TAD, with a linear genetic distance of ~900 kb (Lettice et al., 2003). Integration of regulatory sensors at multiple sites within the *Shh* TAD recapitulate its expression pattern, while sensor integrations in adjacent TADs do not detect any *Shh* activity (Symmons et al., 2014) (Figure 3A). Interestingly, *Shh* and ZRS form a facultative loop in both mouse embryonic stem cells (mESCs) and anterior limbs buds where *Shh* is not expressed, though only the posterior cells express *Shh* (Figure 3B) (Anderson et al., 2014).

10

Introduction
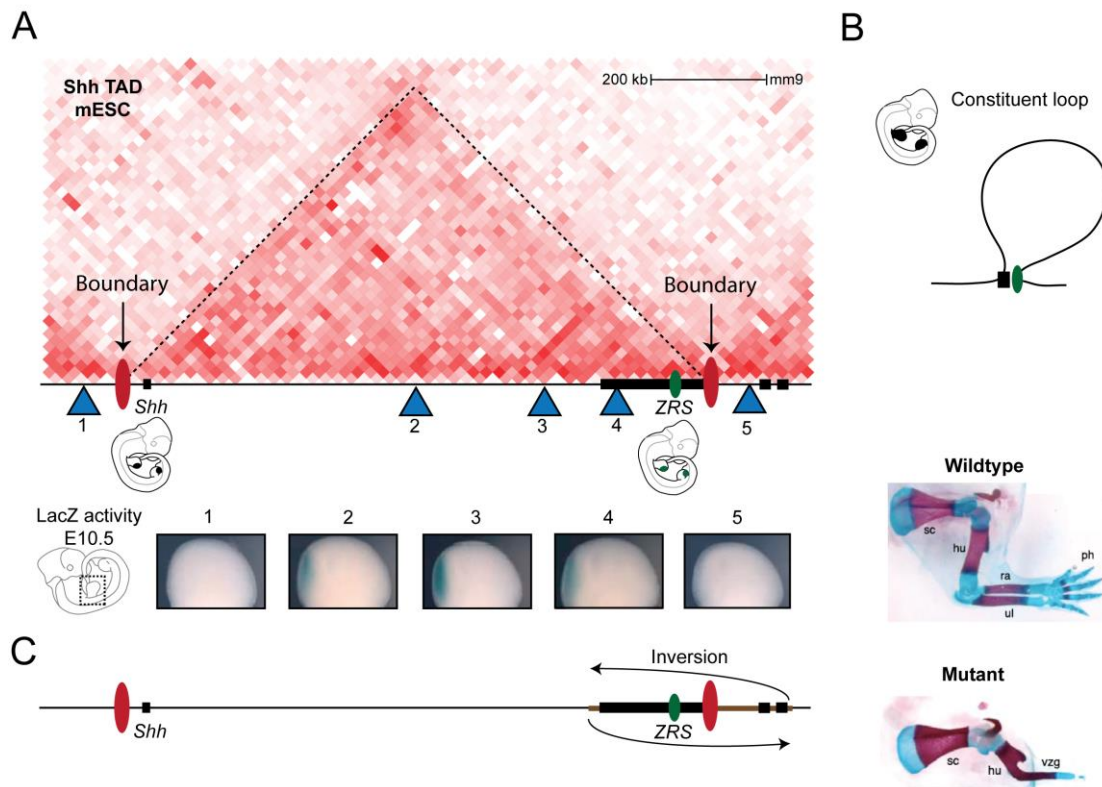
Figure 3 TADs mediate E-P communication

A. *Shh* and its enhancer ZRS are organised together in a 1Mb sized TAD. Insertion of *LacZ* sensors within the TAD reflects *Shh* activity, while insertion in neighbouring domains does not. **B.** FISH experiments in both anterior and posterior limb bud cells show that *Shh* and ZRS are in close proximity. **C.** Inversion of a fragment spanning *Shh* TAD and telomeric neighbouring TAD, including ZRS enhancer, results in a ZRS knock out phenotyping with consequent truncation of the limbs.

Inversions placing ZRS in the neighbouring telomeric TAD disrupts the chromatin interactions between ZRS and *Shh,* resulting in loss of expression (Symmons et al., 2016). Consequently, mice develop monodactyly and a severe truncation of the limb recapitulating the *ZRS/Shh* knock out phenotype (Figure 3C) (Symmons et al., 2016). This indicates that *Shh* TAD acts as a mould supporting ZRS to scan the regulatory landscape and pair with its cognate promoter. When being placed outside of this scaffold, localisation of its target promoter is prevented (Symmons et al., 2016).

The best studied *in vitro* example of dynamic enhancer-promoter interactions is at the *ß-globin* locus, which consists of three genes and a strong enhancer called the locus control region (LCR) (Deng et al., 2012; Deng et al., 2014; Levings and Bungert, 2002). The locus is key in regulating both foetal and adult haematopoiesis, and is enabled by the differential activation of *ß-globin* genes during developmental stages and adult time points (Levings and Bungert, 2002). To achieve this specific expression pattern in time and space, contacts between embryonic-type *ß-globin* promoters and LCR occur only in

primitive erythroid cells, whereas in definitive erythroid cells this interaction is lost and LCR establishes new contacts with the adult *ß-globin* promoters (Deng et al., 2012). In cells, where the ß-globin locus is inactive, physical interaction between the promoter and the enhancer is absent (Deng et al., 2012). Intriguingly, by forcing chromatin interaction between LCR and the embryonic ß-globin gene promoters using zinc finger nucleases in adult erythroid cells reactivates the silenced genes even in the absence of cell specific TFs (Deng et al., 2014). This allows the modulation of transcription by enhancer activity to form tissue specific active chromatin hubs in which the regulatory information together with local concentrations of RNA pol II and TFs lead to robust transcription (Kooren et al., 2007). This type of gene regulation allows a dynamic control of enhancer-promoter interactions, termed instructive, through the modulation of tissue specific binding of TFs (Beccari et al., 2016; Phillips-Cremins et al., 2013).

1.3 THE ROLE OF *PITX1* IN VERTEBRATE LIMB DEVELOPMENT AND DISEASE

The importance of the precise temporal spatial gene expression is particularly evident during morphogenesis, when undifferentiated cells and tissues commit to their terminal shapes and functions. Vertebrate limb development has long been used as a developmental model to understand the regulation of embryonic morphogenesis and organogenesis, due to its easily observable and experimentally modifiable tissue as well as viability of congenital malformations (Petit et al., 2017). The limb is subjected to complex patterning mechanisms and the signalling pathways giving rise to the molecular identity of the this tissue are well conserved throughout evolution, despite the vast morphological differences across different vertebrate species (Zuniga, 2015). This indicates continuous use of the same developmental program with species-specific adopted variations (Montavon and Duboule, 2012). *Pitx1* has been shown to be a key player in morphogenesis across evolution and is thus used in this study as a paradigm of limb development (Petit et al., 2017).

1.3.1 Limb development

The limbs emerge from the lateral plate mesoderm, whereby the epithelial cells of the somatopleure undergo epithelial-to-mesenchymal transition resulting in budding by the localised production of limb bud mesenchyme covered with an overlying ectoderm

Introduction

(Gros and Tabin, 2014). The limb buds proliferate and are patterned along three axes instructed by distinct signalling centres (Zeller et al., 2009). The proximal-distal axis stretching from shoulder/pelvis to digits is controlled by the apical ectodermal ridge (AER), where Fgf10 and Fgf8 are the key signalling molecules (Figure 4A). The AER instructs the underlying mesenchyme to proliferate supporting the outgrowth of the limb. The anterior-posterior axis extending from digits I to V is instructed by the zone of polarising activity (ZPA), where Shh is a crucial morphogen (Figure 4A). The dorsal-ventral axis, forming the back of the hand/feet to the palm, is regulated by several signalling molecules such as Wnt7a and Engrailed1 in the ectoderm (Figure 4A).

1.3.2 Limb specification

In early limb development, fore- and hindlimb buds, which ultimately develop into arms and legs, are morphologically indistinct. However, the limbs buds have divergent tissue specific molecular identities instructing the formation of distinct morphologies and functions of forelimbs and hindlimbs. Most developmental TFs and morphogenic factors involved in limb development are expressed in the same pattern in both tissues. The few TFs known to determine limb identity in tetrapods are Tbx4, Tbx5, Hox factors, and Pitx1. Tbx5 and Tbx4 are paralogues and expressed in the prospective forelimb and hindlimb mesenchyme, respectively (Gibson-Brown et al., 1996). Both TFs induce *Fgf10* expression in the early limb mesenchyme (Figure 4B) (Agarwal et al., 2003; Minguillon et al., 2005). Fgf10 in turn, induces *Fgf8* expression in the limb ectoderm forming a positive feedback loop that sustains the proliferative state of mesenchymal cells facilitating the outgrowth of the limb (Figure 4B) (Xu et al., 1998). Knock out of *Fgf10* in mice halts growth of the limb, though the limb bud formation is initiated, whereas *Tbx5* knock out mice do not initiate forelimb bud formation (Agarwal et al., 2003; Sekine et al., 1999). *Tbx4* knock out mice have severely reduced hindlimb buds (Naiche and Papaioannou, 2003). However, ectopic *Tbx5* in the hindlimb can rescue the *Tbx4* knock out phenotype and vice versa (Minguillon et al., 2005). Tbx4 and Tbx5 are thus important in the initiation and outgrowth of the limb bud, but do not confer the identity (Minguillon et al., 2005). *Pitx1*, on the other hand, is essential for the outgrowth and patterning of the hindlimb.

Introduction
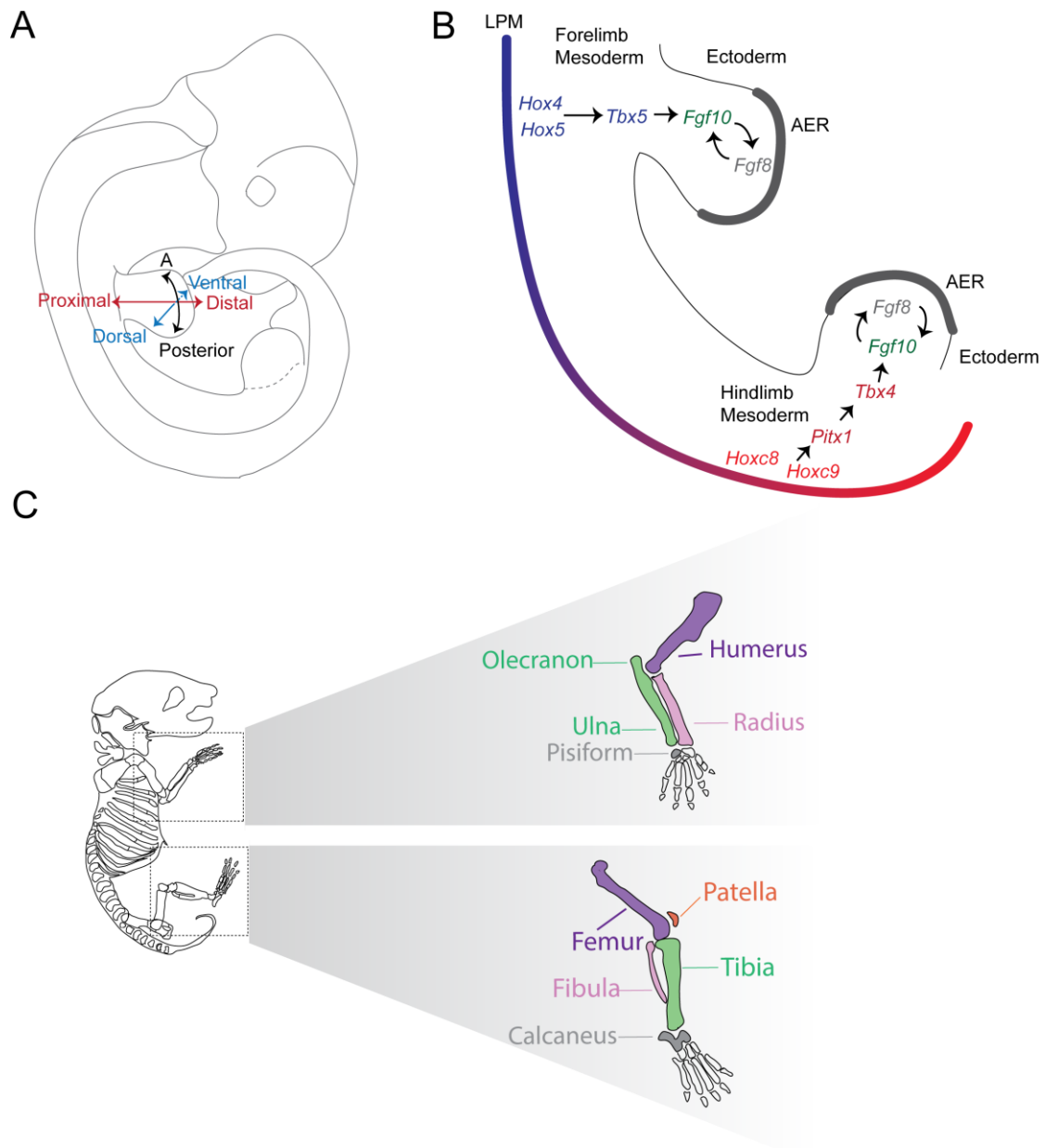
Figure 4 Limb development: signalling, specification, and formation of arms and legs

A. The three axes involved in limb development are proximal-distal, anterior-posterior, and dorsal-ventral. **B.** The specification of forelimbs and hindlimbs by A-P expression of *Hox* genes. Anterior *Hox* genes specify the formation of forelimb region and induce *Tbx5* expression, which in turn activates *Fgf10* and *Fgf8* expression. Posterior *Hoxc* genes induce *Pitx1* expression in assumptive hindlimb region, which activates *Tbx4* expression. **C.** Distinct morphologies in forelimb and hindlimbs.

Pitx1 knock out mice have reduced hindlimbs and show loss of hindlimb specific skeletal elements such as the patella (Lanctot et al., 1999; Szeto et al., 1999). Homeobox containing transcription factors (Hox) have long been known to set up the rostral-caudal axis of the embryo, and determine the relative positioning and downstream transcriptional program of the forelimb and hindlimbs. *Hox* genes are evolutionary highly conserved transcription factors containing a homeobox DNA binding domain

Introduction

(Akam, 1989). A characteristic feature of *Hox* genes is the colocalisation in the genome. This spatial organisation facilitates a collinear regulation rule that results in a spatiotemporal expression of the genes from 3' to 5', corresponding to the rostral-caudal expression pattern in the embryo (Mann et al., 2009). Deletions of *Hox* genes results in homeotic transformations of body segments and malformation in invertebrate and vertebrate species (Wellik and Capecchi, 2003). A famous example is the mutation in the *Antennapedia* gene in *Drosophila,* which results in homeotic antenna-to-legs transformation (Lewis, 1978; Wakimoto and Kaufman, 1981). 3' *HoxA* and *HoxD* genes specify the forelimb bud formation in the rostral part of the embryo by inducing *Tbx5* expression (Figure 4B). *In vitro* studies have identified an enhancer element within intron 2 of *Tbx5* containing *Hox* TFBS (Minguillon et al., 2012). Rostral Hox4 and Hox5 paralogues bind to this enhancer, specifically activating it in forelimb buds, whereas Hoxc8, Hoxc9, and Hoxc10 bind to the region in hindlimb buds, to repress the enhancer activity (Minguillon et al., 2012; Nishimoto et al., 2014). 5' *Hox* genes have been hypothesised to specify the hindlimb bud formation in the caudal part of the embryo by inducing *Pitx1* and *Tbx4* expression (Petit et al., 2017). Overexpression of *Hoxc9* into chick prospective forelimb LPM induces ectopic *Pitx1* expression, suggesting that *Hoxc9* is upstream of *Pitx1* regulation in hindlimb tissue (Nishimoto et al., 2014). At the *Tbx4* locus two enhancer elements, HLEA and HLEB, have been shown to synergistically act to activate the gene specifically in hindlimb buds (Menke et al., 2008). HLEA and HLEB contain TFBS for Pitx1 (Infante et al., 2013). Deletion of the enhancer elements result in reduction of the outgrowth of the hindlimb bud resulting in smaller bones (Menke et al., 2008). To date, no *Pitx1* regulatory elements have been identified in mammals.

1.3.3 Skeletal formation of arms and legs

As development proceeds, the molecularly divergent forelimb buds and hindlimb buds acquire recognisable arm and leg morphologies, respectively. Joint articulations, the shape and positioning of bones, muscles, and tendons form distinctive structures to support the specified function of the limb. Nonetheless, the forelimbs and hindlimb are homologous structures, whereby the following skeletal elements are homologs: humerus and femur, ulna and fibula, radius and tibia, olecranon and patella, and pisiform and calcaneus (DeLaurier et al., 2006). As of today, no vertebrate has been

Introduction

found where the fore- and hindlimbs look identical. In fact, they always form an opposing axis with respect to their flexion and rotation (DeLaurier et al., 2006). On the one hand, the elbow joint is formed by the distal head of the humerus articulating with the proximal ulna where the olecranon process extends proximally to form the characteristic joint, enabling forelimb specific flexion as well as supination (palm facing down) and pronation (palm facing up) (Figure 4C) (DeLaurier et al., 2006). On the other hand, the knee joint is formed by the articulation of the distal femur with the tibial plateau together with the patella enabling flexion and weight bearing (Figure 4C) (DeLaurier et al., 2006). However, the exact pathways leading to the development of the unique features of arms and legs are not fully understood.

1.3.4 *Pitx1* is a master regulator of morphogenesis

The specification of the hindlimb has been demonstrated to be conferred by *Pitx1*. Pitx1 is a bicoid related homeodomain transcription factor containing a homeobox DNA binding domain similar to the *Bicoid* gene in *Drosophila* (Lamonerie et al., 1996). Pitx1 binds TAATCC and related motifs (Berger et al., 2008). *In vivo* characterisation of Pitx1 binding using ChIP-seq in E11.5 hindlimbs demonstrates its broad association with H3K27ac marked active limb enhancers, such as the HLEA and HLEB enhancers regulating *Tbx4* (Infante et al., 2013). Knock out studies in mice show that Pitx1 plays a key role in both the outgrowth and patterning of the hindlimb (Lanctot et al., 1999; Marcil et al., 2003; Szeto et al., 1999). Specifically, skeletal preparation of these mice display severe shortening of the femur, loss of secondary cartilages of distal tibia and fibula, and a loss of the patella, thereby partially transforming the knee joint into an elbow like articulation (Figure 5A and 5B). Moreover, the fibula and tibia are the same size, resembling the homologous bones ulna and radius of the arm. In addition, the calcaneus is reduced in size and resembles the pisiform bone of the hand (Figure 5A and 5B). The ability of *Pitx1* to instruct hindlimb morphogenesis is demonstrated by the consequences of ectopic expression in the developing mouse and chick forelimbs (Figure 5B and 5C) (DeLaurier et al., 2006; Logan and Tabin, 1999). In the mouse, the *Prx* limb enhancer element, active in both fore- and hindlimbs, was used to express *Pitx1* ectopically in forelimbs (DeLaurier et al., 2006). *Prx-Pitx1* mice show a partial forelimb-to-hindlimb conversion characterized by the transformation and translocation

Introduction

of specific muscles, tendons, and bones of the forelimb acquiring a hindlimb-like morphology (DeLaurier et al., 2006).
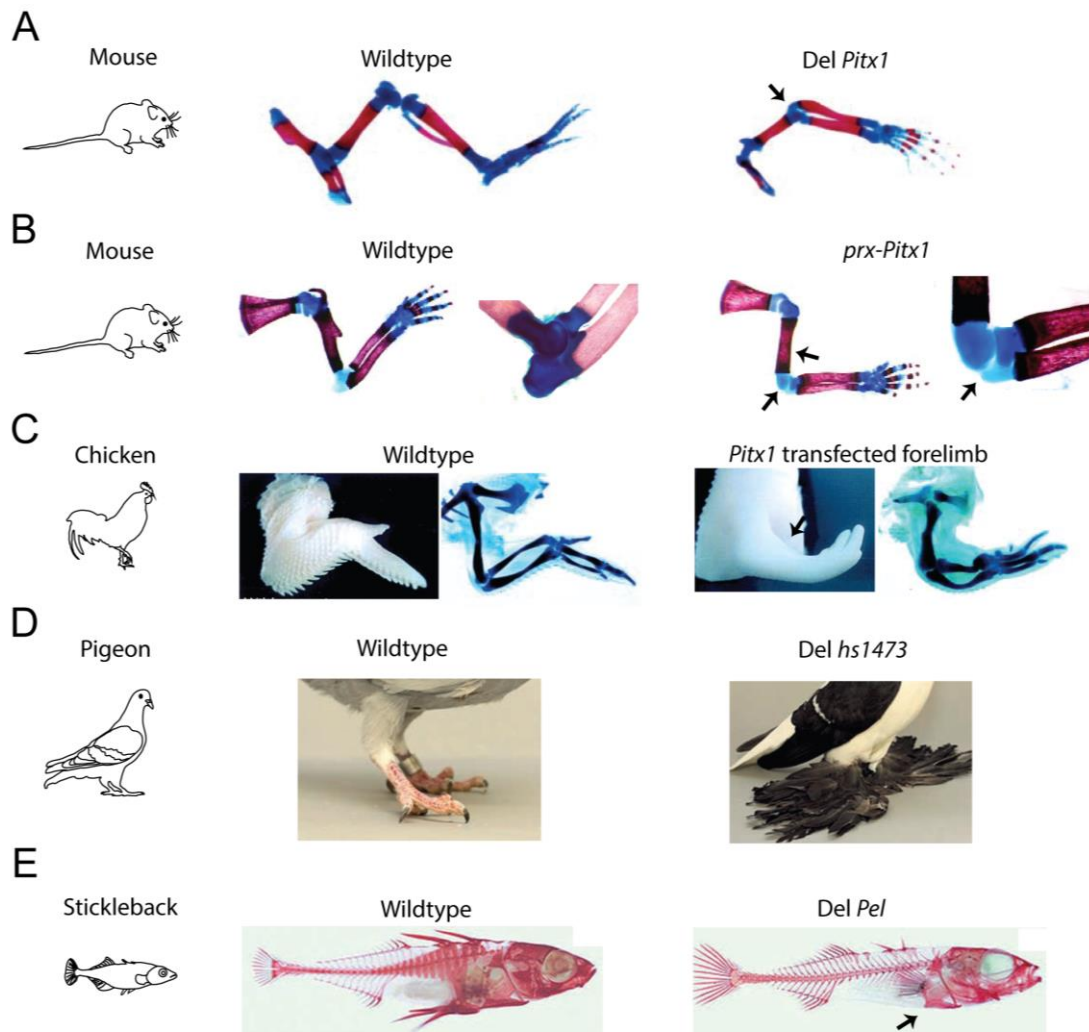
Figure 5 *Pitx1* is a master regulator of hindlimb morphologies across vertebrates

A. Knock out of *Pitx1* in mice results in a loss of patella, reduction of femur and tibia producing an elbow like joint (compare to 5B), and severe reduction of the olecranon. **B.** In mouse, misexpression of *Pitx1* in forelimb bud results in partial arm-to-leg transformation with a loss of deltoid crest and reduction of the olecranon. **C.** Ectopic expression of *Pitx1* in chicken forelimb bud results in partial wing-to-leg transformation. **D.** In certain pigeon breeds, an enhancer element of *Pitx1* is deleted producing feathered feet. **E.** Stickleback fish with a pelvic structure carry the *Pitx1* enhancer sequence for *Pel*. The enhancer is deleted in stickleback fish which have lost their pelvic structures.

This comprises the fusion of the pisiform and triquetral bones resembling the calcaneus, a reduction of the olecranon process, broadening of the humerus together with a loss of the lateral and medial epicondyles and resemble the femur, and an enlargement of the radial head like that of the tibial plateau (DeLaurier et al., 2006). Consequently, the elbow joint articulation is partially transformed into a knee-like articulation (Figure 5B). This causes an altered flexion of the forelimb with outwards

rotation, resembling the hindlimb flexion (DeLaurier et al., 2006). The phenotype is even more pronounced in homozygous animals, demonstrating a dosage effect of Pitx1 (DeLaurier et al., 2006). The ectopic expression of *Pitx1* in forelimb buds not only transforms skeletal structures; muscles and tendons show a hindlimb like patterning to integrate with the new skeletal identity and function (DeLaurier et al., 2006). These studies demonstrate that the role of *Pitx1* in conferring the molecular identity directing limb morphogenesis.

Pitx1 also plays a key role in morphogenesis across vertebrate species, as has been shown in in both feathered feet pigeons and pelvic free stickleback fish (Domyan et al., 2016; Shapiro et al., 2004). In pigeons, the deletion of the enhancer element hs1473 upstream of *Pitx1* results in reduced *Pitx1* expression in hindlimbs, concomitantly with the increased expression of *Tbx5* in the tissue, resulting in partial hindlimb-to-wing transformation (Figure 5D) (Domyan et al., 2016). In stickleback fish, *Pitx1* expression in pelvic region forms pelvic fins, which are homologous to hindlimbs in tetrapods (Yamanoue et al., 2010). Interestingly, certain stickleback species have lost their pelvic fins (Shapiro et al., 2004). Comparison of these distinct stickleback fish led to the discovery of an enhancer element *Pel* that specifically activates *Pitx1* specifically in the pelvic region (Chan et al., 2010). Repeated deletions of the *Pel* enhancer region in stickleback causes loss of *Pitx1* expression, and leads to elimination pelvic structures (Figure 5E) (Chan et al., 2010; Shapiro et al., 2004). Together, these naturally occurring mutations in the *Pitx1* cis-regulatory landscape confer to morphological changes across evolution.

1.3.5 Misregulation of *PITX1* causes human limb malformations

A range of mutations and structural variants encompassing *PITX1* and its putative regulatory landscape have been associated with limbs malformations in humans (Alvarado et al., 2011; Spielmann et al., 2012). Heterozygous mutations within *PITX1* coding regions resulting in haploinsufficiency causes isolated clubfeet in humans (Alvarado et al., 2011).

Introduction
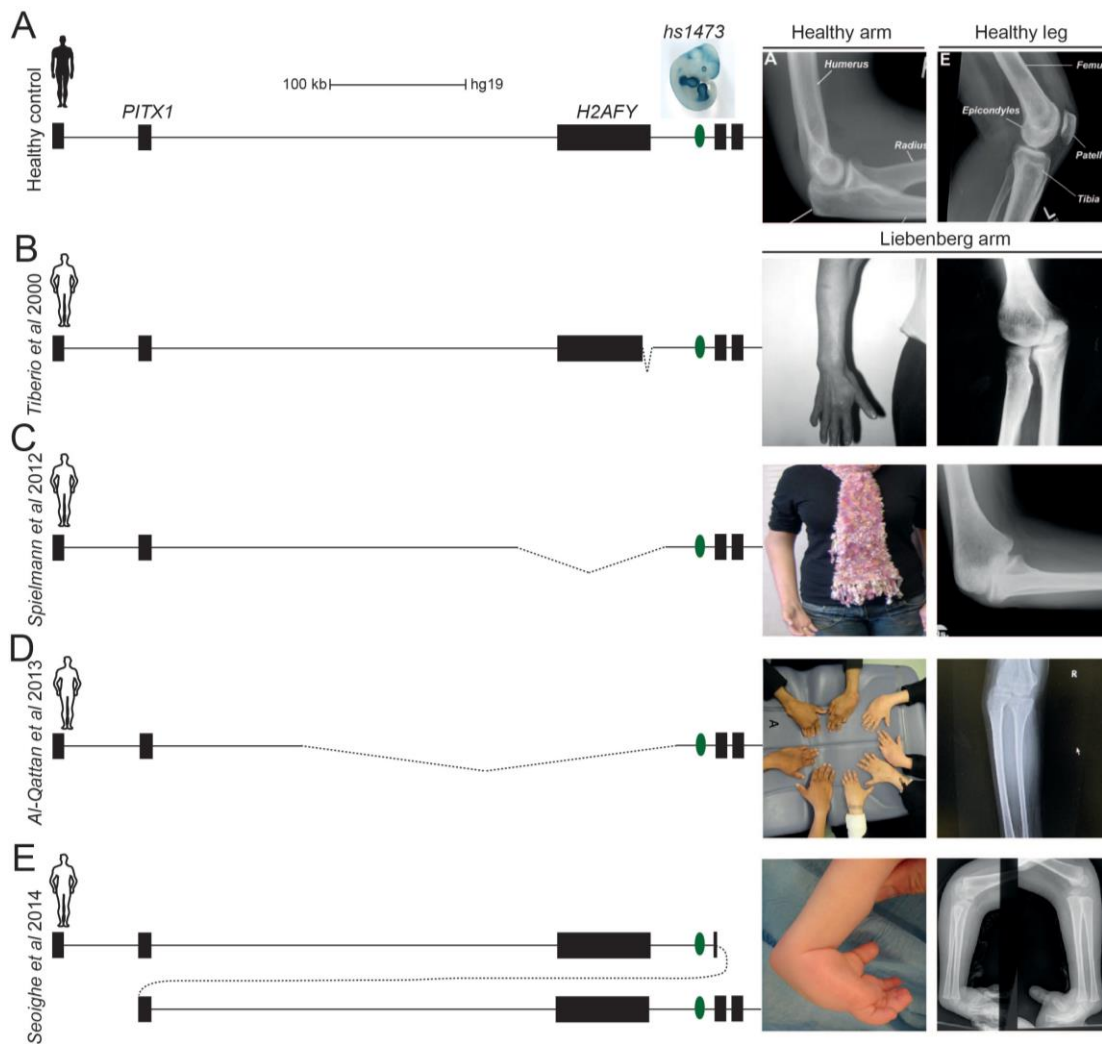
Figure 6 Liebenberg syndrome: structural variants and phenotypes

A. *PITX1* locus and x-ray of healthy elbow and knee joints. B. A 8.5 kb deletion removing the *H2AFY* promoter region, identified in this study, causes a mild form of Liebenberg syndrome (Tiberio et al., 2000). C. A 107 kb deletion encompassing the entire *H2AFY* gene results in the mediocre severity of Liebenberg syndrome D. A 275 kb deletion causes more severe Liebenberg syndrome (Al-Qattan et al., 2013). E. A 420 kb duplication encompassing *PITX1*, *H2AFY* and *hs1473*, placing *PITX1* immediately telomeric of *hs1473*, results in a very severe form of Liebenberg syndrome (Seoighe et al., 2014).

Clubfeet is characterised by the malalignment of the skeleton and joints of the foot and ankle. *Pitx1* haploinsufficient mice display an incomplete penetrance of clubfeet like morphology, whereby the mice show characteristic dragging of their hindlimbs (Alvarado et al., 2011). Structural variants spanning the *PITX1* putative regulatory landscape have been associated with autosomal dominant inherited Liebenberg syndrome (Al-Qattan et al., 2013; Spielmann et al., 2012). The syndrome is characterised by partial arm-to-leg transformation, and display a broadening of the distal humerus, similarly to the distal femur, the olecranon is hypoplastic, the radial head is flattened producing a tibial-like plateau, and the triquetrum and pisiform are

Introduction

fused and comparable to a primitive calcaneus (Mennen et al., 2014; Spielmann et al., 2012) (Figure 6). The structural variants have been suggested to bring the *pan*-limb enhancer *hs1473* into a closer vicinity of *PITX1*, resulting in ectopic *PITX1* expression in forelimbs and consequent partial homeotic transformation of the arm to a leg, such as that of *Prx-Pitx1* mice (DeLaurier et al., 2006; Spielmann et al., 2012). A range of deletions and a single duplication at the *PITX1* locus have been shown to cause Liebenberg syndrome, and a correlation between the altered distance between *hs1473* and *PITX1* and the severity of the phenotype is observed (Al-Qattan et al., 2013; Seoighe et al., 2014; Spielmann et al., 2012; Tiberio et al., 2000) (Figure 6 A-E).

1.3.6 *Pitx1* regulation during development

As described above, the function of *Pitx1* during hindlimb outgrowth and patterning has been extensively studied. However, little is known about its regulation. *Pitx1* is embedded in a relatively small gene desert compared to developmental loci such as *Shh* and *Sox9* (Figure 7). The telomeric side of the gene desert, spanning 240 kb, is flanked by the housekeeping gene *H2afy*, which is ubiquitously expressed and not involved in limb development, as shown in knock out mice (Figure 7) (Boulard et al., 2010; Changolkar et al., 2007).
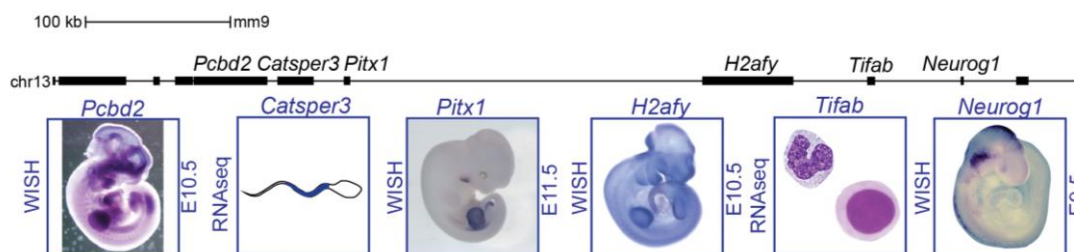
Figure 7 *Pitx1* locus in mice

From left: *Pcbd2* shows faint expression in limbs (WISH E10.5). *Catsper3* is expressed in the principal piece of sperm (RNA seq). *Pitx1* expression pattern (WISH E11.5). *H2afy* ubiquitous expression pattern (WISH E10.5). *Tifab* is expressed in monocytes and T lymphocytes (RNA seq). *Neurog1* expression pattern (WISH E9.5).

5' of *H2afy* is *Tifab* that is expressed in T lymphocytes and monocytes (Figure 7). Upstream of *Tifab* is *Neurog1*, a developmental gene expressed during neuronal development and not involved limb development (Hamada et al., 1999; Ma et al., 1998). Centromeric of *Pitx1* is *Catsper3,* a gene encoding a cation channel sperm associated protein and is expressed in the principal piece of the sperm (Lobley et al.,

Introduction

2003), and *Pcbd2* that encodes a Pterin-4 Alpha-Carbinolamine Dehydratase 2. This gene shows faint expression in limbs using WISH (Figure 7) (Gray et al., 2004).

2. AIMS OF STUDY

Pitx1 is a master regulator of hindlimb identity and its patterning during development has been extensively studied. How *Pitx1* achieves this highly specific gene expression pattern in mammals is unknown. Structural variants leading to Liebenberg syndrome, characterised by partial arm-to-leg transformation, have been associated with deletions 5' of *PITX1* and suggested to lead to its misregulation in developing forelimbs. The present work investigates regulation of *Pitx1*, as a test bed in mouse embryonic limb development, to understand its regulation and contribution to human disease.

Aim 1: Identification of *Pitx1* limb enhancers
Developmental transcription factors are often embedded in large gene deserts containing cognate enhancer elements and confer the tissue specific transcriptional output of promoters. To identify and functionally dissect these elements, *LacZ* reporter assays, CRISPR-Cas9 integration of *LacZ* sensors into the locus, and CRISPR-Cas9 genome engineering are deployed.

Aim 2: The *Pitx1* regulatory landscape in 3D
The communication between enhancers and their cognate promoters is mediated through physical interactions, and conventionally thought to be confined to a single topological associated domain (TAD). Structural variants in the human and mouse genome disrupting this TAD organisation, cause erroneous enhancer-promoter communication with consequent misexpression and disease. Central to this work is the investigation of the dynamics of the 3D architecture of the *Pitx1* locus by using chromosome conformation capture (4C-seq, Capture-C, and Capture Hi-C) analysis of mouse embryonic fore- and hindlimb buds at stages E10.5 and E11.5.

Aim 3: Structural variants at the *Pitx1* locus and disease
Finally, to understand the pathogenicity of structural variants at the *Pitx1* locus, CRISPR-Cas9 genome engineering is used in mice both dissect the chromatin architecture of the locus and to re-engineer structural variants identified in Liebenberg syndrome along with controls alleles to understand the underlying pathomechanism at the gene expression, chromatin configuration, and phenotypic level.

3. Materials

3.1 Instruments

INSTRUMENT	TYPE	SOURCE
Table Top centrifuge	5414D	Eppendorf
Cooling centrifuge	5417R	Eppendorf
Cooling centrifuge	Avanti J-E	Beckman-Coulter
Rotor	JLA16.250	Beckman-Coulter
Thermocycler	GeneAmp PCR System 2700, 2720, and 9700	Applied Biosystems
Real-Time Thermocycler	ABIPrism HAT 79000 RT	Applied Biosystems
Steromicroscope	MZ12	Leica
	Discovery V12	Zeiss
Camera	DFC420	Leica
Light source	KL1500 LCD	Leica
Photometer	Nanodrop 2000	Thermo Scientific
Transilluminator		Hertenstein
UV chamber		Hormuth-Vetter
Micro-CT	Skyscan 1172: X-ray microtomography system	Brucker microCT
Clean bench	HERASafe	Thermo Scientific
CO2-Incubator	HEPA Class 100	Thermo Scientific

Table 1 Instruments

3.2 Chemicals

All chemicals were obtained from Merck (Darmstadt), Roth (Karlsruhe), or Sigma-Aldrich (Hamburg, Seelze, Schnelldorf, and Steinheim), unless otherwise stated.

3.3 Buffers

Common buffers and solutions were prepared according to (Sambrook, 2001)

Materials

BUFFER	COMPOSITION
10xPBS -DEPC	1.37 M NaCl, 27 mM KCl, 100 mM Na2HPO4, 20mM KH2PO4, adjust pH to 7.4 with HCl, in DEPC- H2O, autoclave
4%PFA-PBS	Dissolve 40 mg/ml PFA in 1x PBS (DEPC), heat to 55°C until PFA is dissolved, adjust pH to 7.4 with HCl
Alkaline phosphatase buffer	0.02 M NaCl, 0.05 M MgCl$_2$, 0.1% Tween 20, 0.1 M Tris pH 9.5, and 0.05% levamisole/tetramisole in H$_2$O
Bleaching solution	6% H$_2$O$_2$/PBST
DEPC-H2O	0.1% DEPC in ddH2O
H1 hybridisation buffer	L1 with 0.1% tRNA and 0.05% heparin
H2 hybridisation buffer	hybridisation buffer 1 with 0.1% tRNA and 0.05% heparin and Dig-probe
L1 buffer	50% deionised formamide, 5x SSC, 1% SDS, 0.1% Tween 20 in DEPC; pH 4.5
L2 buffer	50% deionised formamide, 2x SSC pH 4.5, 0.1% Tween 20 in DEPC; pH 4.5
L3 buffer	2x SSC pH 4.5, 0.1% Tween 20 in DEPC; pH 4.5
Proteinase K Buffer	20 mM Tris pH 7.0, 1 mM EDTA, in DEPC-H2O
TBST 1	140mM NaCl, 2.7mM KCl, 25mM Tris-HCl, 1% Tween 20; pH 7.5
TBST 2	TBST with 0.1% Tween 20, and 0.05% levamisole/tetramisole
Blocking solution	TBST 1 with 2% calf-serum and 0.2% BSA
PBST	0.1 % TWEEN-20 in 1x PBS(DEPC)
RIPA buffer	Use DEPC treated reagents, 0.01 % SDS, 0.15 M NaCl, 0.01 % Nonidet-P40, 5 mg/ ml deoxycholate, 1 mM EDTA pH 8.0, 50 mM Tris pH 8.0, in DEPC- H2O
Rnase solution	0.1 M NaCl, 0.01 M Tris pH 7.5, 0.2% Tween 20, 100 µg/ml RNase A in H$_2$O

Table 2 Buffers and solutions for WISH

BUFFER	COMPOSITION
Lysis buffer	50 mM Tris pH7.5; 150 mM NaCl; 5 mM EDTA; 0.5 % Nonidet P-40; 1.15 % Triton X-100; 1x proteinase inhibitors (Roche, # 04693116001); prepare fresh and store on ice
37% Formaldehyde	0,555g PFA in 1050 µl 10 % FCS/PBS and 15 µl 1N NaOH, dissolve at 99 °C for ~10 min with vortexing every 2-3 min
10x ligation buffer	0.4 M Tris-HCl pH=7.8; 0.1 M MgCl2; 0.1 M DTT; 8,3 mM ATP

Table 3 Buffers and solutions used for chromosome conformation capture

3.4 Antibodies

ANTIBODIES	SOURCE	IDENTIFIER
Sheep Anti-Digoxigenin Fab fragments Antibody, AP Conjugated	Roche	#11093274910

Table 4 Antibodies used in this study

Materials

3.5 Enzymes

All enzymes used in this study including restriction enzymes, T4 DNA ligase, Phusion DNA-polymerase, and other DNA-modifying enzymes were obtained from Thermo Fisher Scientific, New England Biolabs, and Promega, unless otherwise stated. For standard genotyping, in house produced Taq and Pfu DNA polymerases was used (A.C. Stiege). RNase A (#R4875) and Proteinase K (#P2308) were purchased from Sigma-Aldrich.

3.6 Kits

KIT	TASK	SUPPLIER
BigDye Terminator v3.1 Sequencing Kit	Sanger Sequencing	Applied Biosystems
Expand Long Template PCR System	4C-viewpoint amplification	Roche
NucleoSpin Plasmid	Plasmid DNA purification	Macherey-Nagel
Nucleobond PC100 EF	Plasmid DNA purification	Macherey-Nagel
NucleoSpin Gel and PCR Clean-up	Gel and PCR clean up	Macherey-Nagel
PCR DIG Probe Synthesis Kit	Amplification and labelling of DIG-probe	Roche
RNAeasy Kit	RNA purification	Qiagen
GoTaq® qPCR Master Mix, 2X	qRT-PCR analysis	Promega
Superscript III Reverse Transcription Kit	Reverse Transcription reagents	Invitrogen
QIAquick PCR purification kit	PCR clean up	Qiagen

Table 5 Overview of commercial kits

3.7 Bacterial strains

All cloning steps were performed in *Eschericia coli* competent TOP10 strain (Invitrogen) and *Eschericia coli* HST08 Stellar competent strain (Clonetech).

3.8 Vectors

pTA-GFP (supplied by Dr. J. Hecht, MPIMG) was used to subclone PCR templates for whole mount in situ probes. For transgenic enhancer assays, pBluescript-LacZ vector was used to subclone NotI PCR amplified DNA regions. For targeted integration of LacZ-reporter constructs into the *ColA1* mouse locus, PgkATGfrt (pGK) targeting vector (Addgene) was used together with FLPe recombinase vector (Addgene) to mediate recombination. To identify the spatial regulatory information at the mouse *Pitx1* locus, LacZ-sensor vectors containing *LacZ* with *ß-globin* minimal promoter and homology arms, was used as a donor vector for CRISPR-Cas9 mediated knock in experiments. LacZ sensors 1-4 were ordered from https://www.vectorbuilder.com/. LacZ sensor 5 was

Materials

cloned using Infusion according to the manufacture's guidelines. PX459 vector (Addgene) was used to clone sgRNA for CRISPR-Cas9 targeting in mESCs. To test a silencing effect of mouse *H2afy* promoter, the pSilencer construct (Vokes et al., 2008) was used for random integrate in the mouse genome. All vectors are listed in Supplementary Table 1.

3.9 Primers

All primers were synthesised and HPSF purified by Eurofins MWG. The sequence is listed in 5' to 3' orientation.

NAME	SEQUENCE	FRAGMENT (MM9)
Pit	atatGCGGCCGCGATTCCATCCACGTTTGGAC atatGCGGCCGCAGGGGTCATCAGGAGACAGG	chr13:56,062,458-56,064,945
Con	atatGCGGCCGCGACCACAGGCCACCGTCTTG atatGCGGCCGCGGTGACAAATGGAGCATGGACA	chr13:56,167,110-56,172,034
Pen	atatGCGGCCGCGGTCCTTAGCTGGGATTTGG gtacGCGGCCGCAACAACCATACCTGCCTCTCC	chr13:56,260,781-56,265,406

Table 6 Primers used for cloning LacZ reporter constructs

In blue, the NotI overhang and template is shown.

NAME	SEQUENCE	FRAGMENT (MM9)
LacZ sensor 5	tgcttttttgtacaaacttgACTGCAATGCGCTCATTGT tgcttttttgtacaaacttgACTGCAATGCGCTCATTGT	Left homology arm chr13:56,256,519-56,257,254
	GCTGCAGGAATTCGATATCATAAGATGGAGGCTTGCCTGT AAGCGGAAGAGCGCCTTAATAGTGCTGGCTGTGTTCATTG	Right homology arm chr13:56,257,306-56,259,322
	CAAGTTTGTACAAAAAAGCAG CAACTTTTCTATACAAAGTTGGCGG	Vector backbone
H2afy-insulator	aatgctagcCATCTCTACTCACCCCAC aaggcgcgccATTGTTCTACCCTCCCTCTT	chr13:56,235,998-56,237,176

Table 7 Primers used for cloning LacZ sensor 5 and pSilencer-H2afy

In blue, the overhangs needed for infusion cloning is shown.
In red, the NotI overhang and template is shown

Materials

NAME	SEQUENCE	TASK
Pitx1 ^{del1} sgRNA left breakpoint	GAAACCGGGAGACGGTGATC	327 kb deletion of *Pitx1* 5' regulatory landscape
Pitx1 ^{del1} sgRNA right breakpoint	CGTGCTATCGAGGGACTAAT	
Pitx1 ^{+/-}	ATCAGCGTCGGACGATTCGC	*Pitx1* e1 frameshift mutation
Pitx1 ^{Pit} sgRNA left breakpoint	GCACGAAGTAATCCCGCCCC	8.5 kb deletion of *Pit* enhancer
Pitx1 ^{Pit} sgRNA right breakpoint	GGTGATCGTTGACGTTCAAA	
Pitx1 ^{Pen} sgRNA left breakpoint	GCCTAGTGGAGGCGCGGCTT	1.2 kb deletion of *Pen* enhancer
Pitx1 ^{Pen} sgRNA right breakpoint	CGTGCTATCGAGGGACTAAT	
Pitx1 ^{del2} sgRNA left breakpoint	GAAACCGGGAGACGGTGATC	236 kb deletion of *Pitx1* 5' gene desert
Pitx1 ^{del2} sgRNA right breakpoint	TTCACACCTGTACTCTAGTG	
Pitx1 ^{del3} sgRNA left breakpoint	TGCTAGGGACCTTACTGTAG	58 kb deletion of *H2afy* gene body, leaving the promoter
Pitx1 ^{del3} sgRNA right breakpoint	TAATGTATTAGAGGGGCGGC	
Pitx1 ^{del4} sgRNA left breakpoint	GAAACCGGGAGACGGTGATC	119 kb deletion of sequence between RA2 and RA3
Pitx1 ^{del4} sgRNA right breakpoint	TCGGATATAGGACGCCCAAA	
Pitx1 ^{del5} sgRNA left breakpoint	AACAAGATGATACCGCATGG	113 kb deletion of *H2afy* and *Pen*
Pitx1 ^{del5} sgRNA right breakpoint	CGTGCTATCGAGGGACTAAT	
Pitx1 ^{RA3} sgRNA left breakpoint	TCGGATATAGGACGCCCAAA	31.5 kb deletion of RA3
Pitx1 ^{RA3} sgRNA right breakpoint	GTATCAATACGCAGCCACTG	
Pitx1 ^{RA2} sgRNA left breakpoint	CCCCACCTAGGTTGGACTAC	16 kb deletion of *Pitx1* gene body/RA2
Pitx1 ^{RA2} sgRNA right breakpoint	CGCACTCTCGGGCGCCTGGT	
Pitx1 ^{delL} sgRNA left breakpoint	AACAAGATGATACCGCATGG	99 kb Liebenberg deletion
Pitx1 ^{delL} sgRNA right breakpoint	ACCCACCAGTACGATCGCTC	
Pitx1 ^{delP} sgRNA left breakpoint	TAATGTATTAGAGGGGCGGC	10 kb Liebenberg deletion
Pitx1 ^{delP} sgRNA right breakpoint	AGCTGACAAAAGTTAGCCCC	
Pitx1 ^{delSm} sgRNA left breakpoint	GAGGCCCGCGCGGCACTTAC	400 bp *H2afy* promoter deletion
Pitx1 ^{delSm} sgRNA right breakpoint	CGCGCGCAGACCGATGCCTG	
Pitx1 ^{inv1} sgRNA left breakpoint	AACAAGATGATACCGCATGG	113 kb inv *H2afy* and *Pen*
Pitx1 ^{inv1} sgRNA right breakpoint	ACCCACCAGTACGATCGCTC	
Pitx1 ^{inv2} sgRNA left breakpoint	AACAAGATGATACCGCATGG	99 kb inv H2afy
Pitx1 ^{inv2} sgRNA right breakpoint	CGTGCTATCGAGGGACTAAT	
Neurog1 ^{del} sgRNA left breakpoint	TGACAGCGACATCTAAACTG	45 kb deletion of *Neurog1* polycomb covered region
Neurog1 ^{del} sgRNA right breakpoint	GGCAGGAAGAGAGATTCACT	
Hoxc ^{del} sgRNA left breakpoint	GCTTAATTTGGCCGACGCAA	121 kb deletion of *Hoxc* cluster
Hoxc ^{del} sgRNA right breakpoint	GGAGCGTTCTTAAACCCGAT	
LacZ-sensor 1	GAAACCGGGAGACGGTGATC	sgRNA for LacZ-sensor 1 knock in
LacZ-sensor 2	GCACGAAGTAATCCCGCCCC	sgRNA for LacZ-sensor 2 knock in
LacZ-sensor 3	AGGAGTGAAAGTTATCAGCG	sgRNA for LacZ-sensor 3 knock in
LacZ-sensor 4	TTCACACCTGTACTCTAGTG	sgRNA for LacZ-sensor 4 knock in
LacZ-sensor 5	GAGAACAATGAGCGCATTGC	sgRNA for LacZ-sensor 5 knock in

Table 8 sgRNA used for CRISPR-Cas9 mediate genome editing and knock-in

29

Materials

NAME	SEQUENCE	TASK
Pitx1 ^{del1}	ACCGGGAGGAAAGCTACAAT TACATCTGCCTTTGAAGTGTTGGA	Genotyping del1
Pitx1 ^{+/-}	CGGACCCCTCTCTTCTAGCC GCGCACTGTCGTTTTCAGAC	Genotyping *Pitx1* e1 frameshift mutation
Pitx1 ^{Pit}	AGAGGATTGCCTGTGATGCTTT AACAGTGGCTTCAGACCTCCTC	Genotyping *Pit* enhancer del
Pitx1 ^{Pen}	GCCTAGTGGAGGCGCGGCTT CGTGCTATCGAGGGACTAAT	Genotyping *Pen* enhancer del
Pitx1 ^{del2}	ACCGGGAGGAAAGCTACAAT CCCTTGATGTCAACCTGGAT	Genotyping del2
Pitx1 ^{del3}	TGCCTCCAAGTTTCGATTAA TACAGCTTCCTTTCAGACTG	Genotyping del3
Pitx1 ^{del4}	ACCGGGAGGAAAGCTACAAT GAATCCACCCCTCCAGTCAA	Genotyping del4
Pitx1 ^{RA3}	ACTTGTCCTAGCCTCCTCCT AACAACGCACAGTAGGGAGA	Genotyping del RA3
Pitx1 ^{RA2}	TATCCAGGCAGTTTCACCCC GCTGTAGTTAAAGGAAGCTGGG	Genotyping del RA2
Pitx1 ^{delL}	GTCCTGTTAGTGGTGTGAAT ACACACCACAGTTTCCAATA	Genotyping 99 kb Liebenberg deletion
Pitx1 ^{delP}	TACAGCTTCCTTTCAGACTG TGTTTGCTACTATGAATCTGTC	Genotyping 10 kb Liebenberg deletion
Pitx1 ^{delSm}	GAGGCCCGCGCGGCACTTAC CGCGCGCAGACCGATGCCTG	Genotyping 400 bp *H2afy* promoter deletion
Pitx1 ^{inv1}	GTCCTGTTAGTGGTGTGAAT ATACAGGATTGCACAGTCTC	Genotyping 113 kb inv left break point
	CCTAAGAGGTCATGTGTCAC ATGTGAAATCAGGTAGGTGG	Genotyping 113 kb inv right break point
Pitx1 ^{inv2}	GTCCTGTTAGTGGTGTGAAT ATAGCCTCCTGTAGTAGCAA	Genotyping 99 kb inv left break point
	CTCACCAGCTATTGATCCAT ACACACCACAGTTTCCAATA	Genotyping 99 kb inv right break point
Pitx1 ^{del5}	GTCCTGTTAGTGGTGTGAAT ATGTGAAATCAGGTAGGTGG	Genotyping 113 kb del
Neurog1 ^{del}	TGACAGCGACATCTAAACTG GGCAGGAAGAGAGATTCACT	Genotyping 45 kb deletion of *Neurog1* polycomb covered
Hoxc ^{del}	GCTTAATTTGGCCGACGCAA GGAGCGTTCTTAAACCCGAT	Genotyping 121 kb deletion of *Hoxc* cluster
LacZ-sensor 1	ACAGGGACAGGGAAATGGTG CCGCTGTTTGGTCTGCTTTC	Genotyping left insertion break point of LacZ-sensor 1
	CTGCAAGGCGATTAAGTTGGG GGATGTAACGAACCGAACCG	Genotyping right insertion break point of LacZ-sensor 1
LacZ-sensor 2	GGGGAGAGGCTAGGGAAGAG CCGCTGTTTGGTCTGCTTTC	Genotyping left insertion break point of LacZ-sensor 2
	CTGCAAGGCGATTAAGTTGGG CAACCCCTGGAAGAGAGAGC	Genotyping right insertion break point of LacZ-sensor 2
LacZ-sensor 3	CCTTTCTCCTCCTGGCTTGG CCGCTGTTTGGTCTGCTTTC	Genotyping left insertion break point of LacZ-sensor 3
	CTGCAAGGCGATTAAGTTGGG GGCTCACGTTCCACACTGTA	Genotyping right insertion break point of LacZ-sensor 3
LacZ-sensor 4	GGTTTGCGAGGTGATGGAGA CCGCTGTTTGGTCTGCTTTC	Genotyping left insertion break point of LacZ-sensor 4
	CTGCAAGGCGATTAAGTTGGG GGGAAACAGTTGACGCACAC	Genotyping right insertion break point of LacZ-sensor 4
LacZ-sensor 5	CCTTGCTACCCACTCGTGTT CCGCTGTTTGGTCTGCTTTC	Genotyping left insertion break point of LacZ-sensor 5
	CTGCAAGGCGATTAAGTTGGG CTATGGCTGGTGGGTATGAC	Genotyping right insertion break point of LacZ-sensor 5

Table 9 Primers used for genotyping CRISPR-Cas9 mutants

Materials

VIEWPOINT	READ PRIMER	2nd PRIMER	GENOME	RE-STRATEGY	FRAGMENT
Pitx1 TSS 1	CCCCAAGAGAC TCAAGTAC	AACCTTAGGCC TAGGTCC	mm9	Csp6I/NlaIII	chr13:56,260,446-56,260,849
Pitx1 TSS 2	GGCTGCCAAAG ACCTGAT	GATGGGTGCCT TGTGAGAA	mm9	DpnII/NlaIII	chr13:55,933,525-55,934,217
H2afy TSS	GTATTTGCTAC ACAGGGATC	TTGCTTTTCCTC CACCAC	mm9	DpnII/NlaIII	chr13:56,234,079-56,234,392
Pen	TGAAGGGGGTA CTTTAGGAT	TTGGTACCACG GAGTAAGTG	mm9	DpnII/NlaIII	chr13:56,261,954-56,262,323

Table 10 Primers used as 4C-seq viewpoints

NAME	SEQUENCE
Pitx1	CCCGTGAACTGAATGTAGGG CTCTCCTGGCTCGACTCTTG
H2afy	TTTGGTTTCCGGTTTGATTG AATACGGACAACAGATCCAG

Table 11 Primers used for cloning WISH probes

3.10 Mouse lines

MUTATION	Alias	GENOMIC LOCATION
Pitx1 ^{fs}	Pitx1 frameA D7	chr13:55,830,781-55,830,811
Pitx1 ^{Pit}	Island-III del	chr13:56,057,065-56,065,575
Pitx1 ^{Pen}	Hs1473-del-B6	chr13:56,260,938-56,262,123
Pitx1 ^{del2}	Del-31	chr13:55,935,302-56,170,964
Pitx1 ^{delL}	H2afy 99 kb del	chr13:56,149,941-56,248,962
Pitx1 ^{del5}	H2afy 113 kb del H7	chr13:56,149,655-56,262,218
Pitx1 ^{Inv1}	H2afy 113 kb inv C2	chr13:56,149,464-56,262,157
Pitx1 ^{Inv2}	H2afy 99 kb inv	chr13:56,149,880-56,248,939
Pitx1 ^{delP}	H2afy 10 kb del D2	chr13:56,232,738-56,242,914
Pitx1 ^{delSm}	H2afy 355 bp del H10	chr13:56,236,684-56,237,072

Table 12 Overview of CRISPR-Cas9 generated mouse lines

3.11 Software and internet resources

SOURCE	SUPPLIER	APPLICATION
Adope Illustrator CS6	Adope	Making of figures
UCSC browser	http://genome.ucsc.edu/	Data visualisation, retrieval of genome sequences of different species
Ensembl	http://www.ensembl.org/index.html	Retrieval of cDNA and mRNA sequences, spotted gar genome assembly
NCBI	https://www.ncbi.nlm.nih.gov/	Retrieval of publications, alignment of sequences
Primer3	http://primer3.ut.ee/	Primer design
Netprimer	http://www.premierbiosoft.com/netprimer/	Primer design, in silico property testing of primers
SDS 2.1	Applied Biosystems	Analysis of qRT-PCR data
HiC2 viewer	Robert Schöpflin	Visualisation of chromatin conformation capture data
Ape	M. Wayne Davis	Plasmid editor
AxioVision Rel.4.8	Zeiss	Microscopy, digital photography

Table 13 Software and Internet resource

4. METHODS

4.1 MOLECULAR BIOLOGY METHODS

Standard laboratory molecular biological procedures, such as polymerase chain reaction (PCR), DNA cloning, transformation of competent E. coli, and gel electrophoresis was conducted according to Sambrook and Russels (Sambrook, 2001).

4.2 DNA ISOLATION

4.2.1 Plasmid DNA

Plasmid DNA isolation was performed using the kit Nucleospin (Macherey-Nagel), and in case of plasmid DNA isolation for transfection of mouse embryonic stem cells Nucleobond PC100 EF (Macherey-Nagel) was used. The plasmid was isolated according to the manufacturer's instructions.

4.2.2 Genomic DNA

DNA of postnatal mouse ear biopsies was extracted using extraction buffer QuickExtract^{TM} (Biozym). The biopsy was incubated with 50 μl of QuickExtract^{TM} for 20 minutes at 65°C, followed by an inactivation step for 2 minutes at 98°C. The DNA was genotyped using standard PCR.

For genotyping and qRT-PCR copy number analysis of embryonic material, the following DNA precipitation method was used: disruption of tissue by incubating samples in lysis buffer (17 mM Tris, pH 7.5; 17 mM EDTA; 170 mM NaCl, 0.85 % SDS complemented with fresh proteinase K (0.08 μg/μl final)) for 3-12 hours at 55°C. To the lysed tissue, 250 μl NaCl was added, followed by mixing of samples on a rocker for 10 minutes at room temperature (RT), and subsequently cooling the samples for 10 minute on ice. Next, the samples were centrifuged at 9000 RPM for 20 minutes at 4°C. The supernatant was transferred to a new tube containing 2x volume 100% ethanol and centrifuged at 13000 RPM for 30 minutes at 4°C. Finally, the DNA pellet was washed twice using 70% ethanol, dried at RT and dissolved in H_2O.

Methods

4.3 CLONING

Traditional cloning methods were used for cloning elements, except for the cloning of homology arms used for LacZ sensor 5, were Infusion was used according to manufacturer's guide lines.

4.3.1 LacZ Enhancer constructs

For enhancer testing in mice, candidate regions from mm9 were selected based on public tracks for conservation, H3K27ac ChIP-seq, DNase HS, and p300 ChIP-seq from 11.5 limbs (Andrey et al., 2017; Visel et al., 2009). To test a candidate enhancer region, the fragment was PCR amplified from genomic mouse DNA using Phusion®High-Fidelity DNA polymerase (New England Biolabs) and primers carrying NotI overhangs (Table 6). The fragment was subcloned into pBluescript-LacZ vector linearised with NotI. By using two different enzymes, the enhancer fragment together with the *LacZ* cassette was cloned into the pGK targeting vector containing frt sites (PgkATGfrt vector) carrying the ATG needed for the puromycin resistance of the C2 mouse embryonic cells, as well as a hygromycin selection cassette.

4.3.2 LacZ sensor constructs

The LacZ-sensor targeting constructs were synthesised by vector builder (https://www.vectorbuilder.com/). The vectors contain a *LacZ* reporter gene together with 50bp minimal *β-globin* promoter flanked by asymmetric homology arms of 750bp and 2 kb in size (Byrne et al., 2015). The sgRNA used for CRISPR-Cas9 mediated knock in, cuts within the first 50 bp of the short homology arm, and thus the PAM sequence was mutated in the 750bp homology arm. The homology arms of sensor 5 were amplified with primers (Table 7) containing overhangs overlapping the backbone (Table 7) of the LacZ-sensor and cloned using Infusion. A map of the constructs is shown in Supplementary Table 1.

4.3.3 pSilencer construct

The transgenic vector was obtained from the Vokes lab (Vokes et al., 2008). The vector consists of the *Mhox* promoter driving *LacZ* in the limb, with strongest activity in the middle wedge of the distal tip of E11.5 limb. *H2afy* promoter fragment was PCR

amplified using primers with NotI and SalI overhangs (Table 7) and together with pSilencer vector was digested using NotI and SalI followed by ligation. Ligation products were transformed into Top10 and plasmid purified from 5 ml cultures.

4.3.4 CRISPR-Cas9 sgRNA design and cloning

sgRNAs were designed using https://benchling.com/ and Feng Zhang's CRISPR-Cas online tool http://crispr.mit.edu:8079/. Off targets scores based on *in silico* quality and off-target predictions (Hsu et al., 2013) were checked, and sgRNA with a score higher than 80% (Feng Zhang) or 0.5 (Benchling) were chosen. Moreover, off targets were checked manually for hits in known limb regions, and excluded if necessary. Two complementary oligos containing the guide sequence and a BbsI recognition site overhang (oligo A: 5'caccgNNNNNNNNNN3´and oligo B aaacNNNNNNNNNN3') were annealed and phosphorylated (T4 Polynucleotide Kinase, Thermo Fisher Scientific, #EK0032) and cloned into the BbsI linearised PX459 vector (Addgene). PX459 contains a puromycin selection cassette, Cas9 endonuclease and chimeric tracrRNA. The PX459-sgRNA vector was transformed into Top10 *E. coli* and plasmid purified from 100 ml cultures. Sanger sequencing was used to validate the correct sgRNA using the following primer: ColR CACGCGCTAAAAACGGACTA. An overview of sgRNA used in this study is listed in Table 8.

4.4 CELL CULTURE AND MOUSE MODELS

The ES cell culture protocol used in the laboratory was established by Katerina Kraft and Malte Spielmann in cooperation with Heiner Schrewe and Lars Wittler (Department Developmental Genetics, Max Planck Institute for Molecular Genetics, Berlin) following standard procedures described in detail in (Behringer, 1994; Kraft et al., 2015; Nagy, 2011; Robertson, 1987; Wasserman, 2010).

4.4.1 Preparation and culturing of feeder cells

Regular Dulbecco's Modified Eagle's Medium (DMEM) containing 4.500 mg/ml glucose and without sodium pyrovate (Lonza #BE17-605E), supplemented with 10% foetal calf serum (FCS Superior, Biochrom #S0615), 1x glutamine (100x, Lonza #BE17-605E), and 1x penicillin/streptomycin (100x, Lonza #DE17-603) was used for culturing feeder cells.

Methods

Feeder cells were prepared from E13.5 CD1 mouse foetuses (Rice and O'Brien, 1980) and DR4 (puromycin/hygromycin resistant) (Tucker et al., 1997) mouse foetuses, and cultured for five passages before tested for mycoplasma contamination using Mycoalert detection kit (Lonza #LT07-118) and Mycoalert assay control set (Lonza #LT07-518). To inactivate the feeders, mitomycin C treatment was used (Sigma #M4287). Finally, cells were frozen at $2.5x10^6$ cells/vial. The freezing medium consisted of regular feeder medium with the addition of 20% FCS (FCS Superior, Biochrom #S0615) and 10% DMSO (Sigma #D-2650).

4.4.2 Culturing of mouse embryonic stem cells

C2 (C57B/6 and 129svJae F1 hybrid) and G4 (129/sv x C57BL/6 F1 hybrid) mouse embryonic stem cells (mESCs) were cultured on feeder cells in ESC medium consisting of Knockout Dulbecco's Modified Eagle's Medium containing 4.500 mg/ml glucose and with sodium pyruvate, supplemented with 15% FCS (PANSera ES #P30-2600, Lot 130407ES), 1x glutamine (100x, Lonza #BE17-605E), 1x penicillin/streptomycin (Lonza #DE17-603E), 1x non-essential amino acids (100x, Gibco #11140-35), 1x nucleosides (100x, Chemicon, #ES-0008D), 0.1 mM ß-mecaptoethanol (Gibco #3150-010), and 1000 U/ml LIF (Murine Leukaemia Inhibitory factor ESGRO^{TM} (10^7 U/ml, Chemicon #ESG1107)). Culturing surfaces were first coated with 0.1% gelatine (Sigma #G-1393) for 10 minutes and subsequently removed before plating feeder cells in a $3-4x10^4$ cell/cm^2 density. After 6 hours, a monolayer is formed and mESCs were seeded on top. The ESC medium was changed at least every 24 hours, split every 2-3 days, and frozen $1x10^6$ cells/vial in freezing medium consisting of ESC medium supplemented with 20% FCS (FCS Superior, Biochrom #S0615) and 10% DMSO (Sigma #D-2650).

4.4.3 Transfection of G4 ESCs with CRISPR-Cas9

CRISPR-Cas9 genome engineering in G4 mESCs was performed as follows: on day 1, CD1 feeder cells were seeded on a 6-well plate. On day 2, $0.4x10^6$ G4 ESCs were seeded into each well per transfection. On day 3, two hours prior to transfection ESC medium without pen/strep was added. For transfection, a DNA mix consisting of 8 µg of PX459-sgRNA vector was combined with 125 µl Optimem (Gibco #51985-026), and a transfection mix consisting of 25 µl FuGene HD agent (Promega #E2311), and 100 µl

Methods

OptiMEM (Gibco #51985-026), were combined and incubated at RT for 15 minutes before added dropwise onto the cells. To create structural variants (deletions or inversions) at the *Pitx1* locus, two sgRNAs were used in separate PX459 vectors and cotransfected using 8 μg each. For knock-ins, 8 μg of PX459-sgRNA vector and 4 μg of LacZ-sensor targeting vector were cotransfected. To create indel mutations, 8 μg of a single PX459-sgRNA vector was used. On day 4, three 6 cm dishes of DR4-puromycin resistant feeders were seeded for each transfection. On Day 5, targeted G4 cells were split onto three DR4 6 cm dishes and 48-hour selection initiated by adding puromycin (Sigma-Aldrich #P8833) to ESC medium (final concentration 2μg/ml). On day 6, selection was abrogated and recovery initiated by changing to ESC medium. Recovery period lasted around 4 days until clones were visible and large enough for picking. Clones were picked (circa 300 per deleltion/inversion/knock-in alleles. 100 clones picked for indels) in PBS (Lonza #BE17-512F) with pipette tips and transferred to U-bottom 96 well plates containing 1x trypsin-EDTA (Gibco #25300-054) and incubated for 10 minutes at 37°C. Single cells from each well were then resuspended and transferred into a 96-well plate containing CD1 feeders. After 3 days of culture, plates were split in triplicates, two for freezing, and one for growth and DNA harvesting. Cells were frozen using bicarbonate free DMEM (Gibco #52100) supplemented with 10 mM Hepes pH 7.2, 20% FCS (PANSera ES #P30-2600, Lot 130407ES), 10% DMSO (Sigma #D-2650), 1x glutamine (100x, Lonza #BE17-605E), 1x penicillin/streptomycin (Lonza #DE17-603E), 1x non-essential amino acids (100x, Gibco #11140-35), 1x nucleosides (100x, Chemicon, #ES-0008D), 0.1 mM ß-mecaptoethanol (Gibco #3150-010), and 1000 U/ml LIF (Murine Leukaemia Inhibitory factor ESGRO^{TM} (10^7 U/ml, Chemicon #ESG1107)). Cells were lysed in lysis buffer (10 mM Tris pH 7.5, 10 mM EDTA pH 8.0, 10 mM NaCl, 0.3% sacrosyl, 1mg/ml Proteinase K) over night. An overview of CRISPR-Cas9 mutants generated in this study is shown in Table 12.

4.4.4 Transfection and targeting of flippase mediated knock in of LacZ enhancers

The frt-flippase system was used as targeting strategy in C2 mESC , where the *ColA1* locus has been modified with a PGK neomycin selection cassette flanked by frt sites and an ATG-less, promoterless hygromycin cassette. The transfection procedure was as follows: On day 1, CD1 feeders were seeded onto a 6 well plate. On day 2, $0.8x10^6$ C2

cells were seeded onto each well per transfection. On day 3, prior to transfection, cells were washed twice with PBS and 1.75 ml ES medium with LIF and without penicillin/streptomycin was added. A DNA mix consisting of 9 µg of targeting construct, 3 µg FLP recombinase, 1 µl Lipofectamine LIPO-LTX Plus (ThermoFisher #15338100) with OptiMEM (Gibco #51985-026) up a final volume of 125 µl, was incubated first separately from the transfection mix consisting of 20 µl Lipofectamine (ThermoFisher #15338100) and 125 µl OptiMEM (Gibco #51985-026), and then combined after 5 minutes and incubated at RT for 30 minutes. The transfection-DNA mix was added drop-wise onto the C2 cells. DR4 feeders were seeded onto 6 cm dishes, three for each targeting construct. On day 4, each transfected C2 well was split onto three DR4 feeder dishes. On day 5, selection was initiated by adding ESC medium containing hygromycin B (Invtirogen # 10687010, final concentration 150 µg/ml), and selection continued until clones were ready to pick (between 4—12 clones) and transferred into a 96-well plate with CD1 feeders. After 3 days of culture, clones were grown until they reach an average of four million cells. Three vials were frozen, and DNA was harvested from the rest of the cells to confirm genotyping. An overview of tested enhancer candidate regions is listed in Table 6.

4.4.5 Screening of CRISPR-Cas9 targeted cells

Mouse mutants were genotyped by standard PCR procedures using Taq polymerase. The reagents were pipetted on ice into a 96 well plate and DNA was amplified in a thermocycler. PCR conditions for 20 µl reaction: 2 µl 10x Taq buffer (750 mM Tris/HCl pH 8.8; 200 mM $(NH_4)_2$ SO_4; 0.1% Tween 20; 15 mM $MgCl_2$), 0.5 µl dNTPs (10mM), 0.075µl forward primer (100 µM), 0.075 µl reverse primer (100 µM), 0.5 µl Taq enzyme, 1 µl template (~20 ng), 15.85 µl H2O. The following PCR program was used: step 1: 94°C, 4 min; step 2: 94°C, 30 sec; step 3: 60°C, 30 sec; step 4: 72°C, 45 sec; step 5: go to step 2 for 29x; step 6: 72°C, 7 min; step 7: 4°C, pause. The PCR products were analysed on a 1 % agarose gel. Primer sequences for conventional genotyping are listed in Table 9.

Methods

4.4.6 Copy number analysis of mutant clones

Copy number analysis from genomic DNA (gDNA) was performed with the Go-Taq (Promega) on an ABIPrism 7900 HT thermocycler. Primers were designed with the Primer3Plus online tool with a product size of 80-120 bp. The qRT-PCR reaction was set up in a 12 µl reaction on a 384-well plate with the following components: 6 µl of 2x SYBR mix, 1 µl primer (2.5 µM each) and 5µl DNA (10 ng). Each reaction was performed in triplicates. Relative Ct values for each target (primers within deleted region) were normalized to the reference region (primers outside deleted region).

4.4.7 Generation of transgenic mice using diploid and tetraploid aggregation

A frozen C2 or G4 mESC vial was seeded on CD1 feeders and grown for 2 days and used for the creation of embryos and live animals by diploid or tetraploid complementation (Artus and Hadjantonakis, 2011). Female mice of CD1 strain were used as foster mothers. Mouse strains were maintained by crossing them with C57BL6/J mice.

4.4.8 Mouse crossings

All animal procedures were in accordance with institutional, state, and government regulations (Berlin:LAGeSo G0247/13).

4.5 CHROMOSOME CONFORMATION CAPTURE

4.5.1 4C-seq

The protocol for 4C-seq described below was adapted from (van de Werken et al., 2012) and established in the laboratory by Martin Franke. Mouse tissues from E11.5 were collected and 3C-library and 4C-samples were processed according to this 4C-seq protocol. The 4C-seq profiles depicted here show the normalized read counts (RPM), representing the interaction frequency of a DNA fragment with a chosen viewpoint fragment.

4.5.1.1 Harvesting, fixation, and lysis of tissues

Embryonic tissues were dissected in PBS and a single cell suspension was made by adding prewarmed trypsin and shaking samples at 900 RPM for 10 min at 37°C. Trypsin was inactivated by adding PBS/10%FCS and cells homogenised using a 0.40 µm cell strainer followed by centrifugation at 1100 RPM for 5 minutes at RT. Supernatant was

Methods

discarded and pellet resuspended in 5 ml PBS/10%FCS with subsequent adding of 4% formaldehyde (FA – see Table 3) in PBS/10%FCS (final concentration 2% FA), and incubated for 10 min at RT while tumbling. Crosslinking was quenched by the addition of 1 ml 1.425 M glycine on ice, followed by immediate centrifugation at 1500 RPM for 8 minutes at 4°C. Supernatant was removed and pellet resuspended in 5 ml cold lysis buffer (see Table 3) and incubated on ice for at least 10 minutes. To control the fixation and cell lysis step, 3 µl of cell suspension was mixed with 3 µl Methyl Green-Pyronin staining on a glass slide with coverslip and incubated for 2 minutes before viewing the staining under a microscope. Fixed nuclei appear blue and lysed cell cytoplasm appears pink/purple. Next, the fixed and lysed cells were centrifuged at 2000 RPM for 5 minutes at 4°C. The pellet was resuspended in H_2O and transferred to 1.5 ml Eppendorf tube, followed by centrifugation at 2400 RPM for 2 minutes at 4°C, and final removal of supernatant and snap freezing of pellet. Pellets were stored at -80°C until used for the preparation of the 3C library.

4.5.1.2 Preparation of 3C-Library

Pellet was resuspended in 360 µl H_2O and 60 µl 10x restriction enzyme buffer (RE-buffer) (Fermentas). Samples were placed at 37°C shaking at 900 RPM, 15 µl of 10% SDS was added and sampled incubated for 1 hour to remove non-cross-linked proteins. To sequester SDS, 150 µl 10% Triton X-100 was added and samples were incubated shaking at 900 RPM for 1 hour at 37°C. As an undigested control, 5 µl aliquot of the sample was taken and stored at 4°C until further analysis. To the sample, 600 µl 1xRE-buffer was added together with 400 U RE and incubated shaking at 900 RPM for 4 hours at 37°C. After 4h, 200 U RE was added and samples were incubated shaking at 900 RPM overnight (o.n) at 37°C. The next morning, 200 U RE was added and samples incubated at 37°C shaking at 900 RPM for 4 hours. As a digested control, 5 µl aliquot from samples was taken. Digestion efficiency of the control undigested and digested aliquots were determined by adding 90 µl 10 mM Tris pH 7.5 and 2 µl RNaseA, and incubating them for 1 h at 37°C. Subsequently, 5 µl Proteinase K was added to the aliquots and incubated for 4 hours at 65°C. The DNA was precipitated by adding 100 µl phenol-chloroform and mixed vigorously, followed by centrifugation at 13200 RPM for 10 minutes at RT. The upper water phase was transferred to an Eppendorf tube and an aliquot of 20 µl analysed on a 1% agarose gel.

Methods

Before proceeding to the ligation step, the samples were incubated for 20 minutes at 65°C to heat inactivated RE. The samples were transferred to a 50 ml falcon and 700 µl 10x ligation buffer (table 3) and H_2O up to 7 ml was added. 50 U of T4-DNA ligase (ThermoFisher #EL0014) was added to the samples and incubated at 4°C tumbling o.n. To test for ligation efficiency, 100 µl aliquot was taken and subjected to 2 µl RNaseA for 1 hour at 37°C, followed by incubation with 5 µl proteinase K for 4 hours at 65°C. The DNA was precipitated by adding 100 µl phenol-chloroform to the aliquots, mixing vigorously, and centrifuging at 13200 RPM for 10 min at RT. The upper water phase was transferred to an Eppendorf tube and 20 µl analysed on a 1% gel. To de-crosslink the ligated samples, 30 µl Proteinase K was added and incubated at 65°C o.n. The next day 30 µl RNaseA was added and incubated for 45 minutes at 37°C. DNA was precipitated using 7 ml phenol-chloroform by mixing gently, followed by centrifugation at 3750 RPM for 15 minutes at RT. The upper water phase containing the chimeric DNA was transferred to a new 50 ml falcon tube, and the DNA was purified by adding 3 ml H_2O, 1.5 ml 2M NaAc pH 5.6, 7 µl glycogen (20mg/ml), and 35 ml 100% ethanol. The samples were gently mixed and placed at -80°C o.n. The next day, samples were centrifuged at 8350 g for 20 minutes at 4°C. The pellet was washed with 70% ethanol and centrifuged at 3300 g for 15 minutes at 4°C. The chromatin pellet was dried at RT and dissolved at 37°C in 150 µl 10 mM Tris pH 7.5.

4.5.1.3 Preparation of 4C-Library

To the 3C-sample, 295 µl H_2O, 50 µl 10xRE buffer, and 60 U second RE was added and incubated at 37°C o.n. To determine the digestion efficiency, 5 µl aliquot was taken and 5 µl 10 mM Tris pH 7.5 was added and analysed on a 0.6% gel. For the second ligation, RE was first heat inactivated by incubation at 65°C for 25 minutes, with subsequent transferral of samples to a 50 ml falcon adding 12.1 ml H2O, 1.4 10x ligation buffer, and 100 U T4-DNA ligase (ThermoFisher #EL0014), and incubated while rotating at 4°C o.n. To precipitate the DNA, 1.4 ml 2M NaAc pH 5.6, 1.75 µl glycogen (20 mg/ml), and 35 ml 100% ethanol was added, gently mixed, and stored at -80°C o.n. The samples were centrifuged at 8350g for 45 minutes at 4°C. The pellet was washed with 70% ethanol and centrifuged at 3300g for 15 minutes at 4°C. Next, the pellet was dried at RT and resuspended in 10 mM Tris pH 7.5 at 37°C. Finally, samples were purified using QIAquick PCR purification kit (Roche #28106) and 4C samples stored at -20°C.

Methods

4.5.1.4 PCR amplification of viewpoint and sequencing

4C samples were used as template for inverse PCR using primers, which has been designed according to the region of interest (viewpoint fragment)(Splinter et al., 2012; van de Werken et al., 2012), e.g. the transcription start site of a gene or enhancer sequence. Primer sequences for inverse PCR and corresponding digestion strategy used in this study are listed in Table 10. The read primer, next to the 1st RE-site and second primer, close to the 2nd RE-site, contained Illumina TrueSeq adapter sequences at their 5' end (read primer 1:CTACACGACGCTCTTCCGATCT; primer 2: CAGACGTGTGCTCTTCCGATCT). To control the primer and library quality, each 4C sample and primer pair was tested using serial template concentrations of 0.5, 1, 2 and 4 ng/ µl (final concentration) in a 25 µl PCR reaction. The Expand Long Template PCR System (Roche, #11759060001) was used for PCR amplification. All reagents were pipetted on ice with final concentrations of the following master mix: 1x buffer, 0.2 mM dNTPs (each), 1 µM primers, 0.075U/ µl DNA polymerase (Enzyme mix), 4C template in range of 0.5 ng/µl to 4 ng/µl, reaction filled up with water to desired volume. 4C samples were amplified in a thermocycler with the following program: 1. Initial denaturation, 94°C for 2 min; 2. Denaturation, 94°C for 15 sec; 3. Primer annealing, 55 °C for 1 min; 4 Elongation, 68°C for 3 min; 5, Go to step 2 with 29 additional cycles; 6 Final elongation, 68°C for 7 min; 8, End, 4°C. The PCR amplified viewpoint was analysed on a 1 % agarose gel. The highest template concentration, giving the broadest and most reproducible band pattern, was used for the final 4C PCR amplification. A total of 1.6 µg of each 4C sample was amplified in 50 µl PCR reactions, pooled, and purified using the QIAquick PCR purification kit (Roche #28106). PCR amplifications were verified on a 1% agarose gel. Samples were multiplexed and sequenced with Illumina Hi-Seq technology according to standard protocols. Sequencing was done by the sequencing core facility at the Max Planck Institute for Molecular Genetics (Bernd Timmermann) or at the Institute for Medical and Human Genetics, Charité Universitätsmedizin Berlin (Ulrike Krüger and Dr. Jochen Hecht). Up to 35 4C libraries were multiplexed, pooled, and sequenced on one lane of a Hi-Seq flow cell using 100 bp paired-end or 100 bp single-end sequencing kits. On average ten million reads per 4C library were obtained.

Methods

4.5.1.5 4C-seq data analysis

In-house pipeline for 4C data analysis and data processing was established in cooperation with Verena Heinrich (Department Computational Molecular Biology, Max Planck Institute for Molecular Genetics, Berlin). The primer sequences were clipped from short sequencing reads and quality assurance performed according to (van de Werken et al., 2012) with a customized Java program. The remaining reads were mapped to the reference sequence NCBI37/mm9 using BWA (Li and Durbin, 2009). Each 4C library fulfilled the following quality criteria: more than one million mapped reads and cis/overall ratio of mapped reads > 60%. BedGraph tracks for a normalized number of reads (reads per million, RPM) overlapping with fragments of 1st RE were created in a specified genomic range (chr13: for NCBI37/mm9) and smoothed over a specified number of fragments using customized Java programs. The viewpoint and adjacent fragments 1.5 kb up- and downstream were removed, and a window of ten fragments was used to normalize the data per million mapped reads (reads per million, RPM). BedGraph files of single 4C experiments were visualised as tracks in the UCSC browser.

4.5.2 Capture Hi-C

4.5.2.1 SureSelect design

The cHi-C SureSelect library was designed over the genomic interval (mm9, chr13: 53400000-57300000) using the SureDesign tool from Agilent.

4.5.2.2 Fixation and lysis of mouse embryonic tissues

Chromatin was prepared from homozygous E11.5 forelimb and hindlimb buds as described previously (van de Werken et al., 2012). cHi-C experiments were performed as quadruplets (WT) or duplicates (mutants). Per biological replicate, 5-6 pairs of limb buds (circa 3 $\times 10^6$ cells) were micro dissected in PBS at RT. A single cell suspension was obtained by incubating the tissue in 500 µl Trypsin-EDTA 0.05 % (Gibco #25300-054) and shaking at 900 RPM for 10 minutes at 37°C. The cells were resuspended and homogenised using a 0.40 µm cell strainer (Falcon) and diluted in 10% FCS/PBS. Cells were fixed by adding 650 µl 37% formaldehyde (Sigma-Aldrich #252549) with a final concentration of 2% while and left tumbling for 10 minutes at RT. Fixation was quenched using 1.425 M glycine on ice, and immediately centrifuged at 260 g for 8 minutes at 4°C. Supernatant was removed and the pellet resuspended in lysis buffer

Methods

(Table 3) and incubated on ice for 10 minutes. Removal of lysis buffer by centrifugation at 400 g for 5 minutes at 4°C, was followed by disposal of supernatant and snap-freezing and storage at -80°C.

4.5.2.3 Preparation of 3C-Library

Pellet was resuspended in 520 µl 1xDpnII buffer (Fermentas), and incubated with 7.4 µl 20% SDS shaking at 900 RPM for 1 hour at 37°C. Next, 75 µl 20% x-100 Triton was added and left shaking at 900 RPM for 1 hour at 37°C. A 15 µl aliquot was taken to use as a control for undigested chromatin (stored at -20°C). The chromatin was digested using 40 µl DpnII (10 U/µl) and incubated shaking at 900 RPM for 6 hours at 37°C. 40 µl of DpnII (10 U/µl) was added and samples were incubated shaking at 900 RPM 37°C o.n. On the third day, 20 µl DpnII (10 U/µl) was added to the samples shaking at 900 RPM for 5 hours at 37°C. DpnII was inactivated for 25 minutes at 65°C. A 50 µl aliquot was taken to test digestion efficiency (stored at -20°C). Next, the digested chromatin was diluted and re-ligated in 5.1 ml H_2O, 700 µl 10x ligation buffer (Table 3), and 5 µl 30 U/µl T4-ligase (ThermoFisher #EL0014) incubated for 4 hours in a 16°C water bath, shaking them manually 2-3 times. The ligated samples were incubated further 30 minutes at RT. The chimeric chromatin products and test aliquots were de-crosslinked over night by adding 30 µl and 5 µl Proteinase K, respectively, and incubated at 65°C o.n. On the fourth day, 30 µl or µl 5 µl of 10 mg/ml RNase was added to the samples and aliquots, respectively, and incubated for 45 minutes at 37°C. Next, chromatin was precipitated by adding 1 volume phenol-chloroform to the samples and aliquots, vigorously shaking them followed by centrifugation at 4000 RPM for 15 minutes at RT. The upper phase containing the chromatin was subsequently transferred to a new falcon. To the aliquots, 100% ethanol was added, and incubated for 30 minutes at -20°C, followed by centrifugation at 5000 RPM for 45 minutes at 4°C, washed with 70% ethanol, and resuspended in 20 µl 10 mM 10 mM Tris-HCl pH 7.5. To the samples, 7 ml H2O, 1 ml 3M NaAc pH 5.6, 7 µl glycogen (20 mg/ml, Thermo Fisher Scientific #R0561), and 35 ml 100% ethanol was added. The samples were frozen at -20°C for 3 hours. The chromatin was clarified by centrifugation at 5000 RPM for 45 minutes at 4°C. The chromatin pellet washed with 70% ethanol, and centrifuged at 5000 RPM for 15 minutes at 4°C. Finally, the 3C-library chromatin pellet was dried at RT and resuspended in 10 mM Tris pH 7.5.

Methods

To check the 3C-library, 3 µl was loaded on a 1% agarose gel together with the undigested and digested aliquots.

4.5.2.4 Preparation of 3C-library for sequencing

The 3C-library was sheared using a Covaris sonicator (duty cycle: 10%, intensity: 5, cycles per burst: 200, time: 6 cycles of 60 s each, set mode: frequency sweeping, temperature: 4 to 7 °C). Adaptors were added to the sheared DNA and amplified according to Agilent instructions for Illumina sequencing. The library was hybridized to the custom-designed sure-select beads and indexed for sequencing (50 to 100 bp paired-end) following Agilent instructions.

4.5.3 Hi-C Analysis

Preprocessing and mapping of paired-end sequencing data, as well as filtering of mapped di-tags was performed with the HiCUP pipeline v.0.5.8 (Wingett et al., 2015) (Nofill: 1, no size selection, Format: Sanger). The pipeline used Bowtie2 v.2.2.6 (Langmead and Salzberg, 2012) for mapping short reads to reference genome (NCBI37/mm9). Replicates were combined after mapping and filtering. Filtered di-tags were further processed with Juicer tools to bin di-tags (5 and 10 kb bins) and to normalize the map by Knight-Ruiz (KR) matrix balancing (Durand et al., 2016; Knight and Ruiz, 2013; Lieberman-Aiden et al., 2009). For this, only reads with a MAPQ ≥30 were considered. The DNA-capturing step enriches genomic region chr13:53,400,001-57,300,000 on mm9 leading to three different regimes in the cHi-C map: (i) enriched versus enriched, (ii) enriched versus non-enriched, and (iii) non-enriched versus non-enriched. For binning and normalization only di-tags in regime (i) were considered. Therefore di-tags were filtered for the enriched region and mm9 coordinates were shifted by 53,400,000 bp. For Juicer tools, a custom chromosome sizes file containing only the enriched region on chr13 (length 3900000 bp) was used. After binning and normalization, coordinates were shifted back to their original values. All maps were processed on the wildtype reference genome to work with the same genomic coordinates across all samples. To account for differences between maps in their distance dependent signal decay, maps were scaled pairwise by the sum of their sub-diagonals. Therefore, each sub-diagonal vector in one matrix is divided by its sum and multiplied by the average of the sums of both matrices. To normalize for sequencing

depth, each map was additionally converted to reads per million (RPM). To avoid copy number biases, a region spanning all tested deletions (chr13:55,730,001-56,500,000) was not considered for the computation of scaling factors, for the diagonal normalization as well as for the RPM normalization. cHi-C maps of count values and subtraction maps were visualized as heat maps truncating all values above the 99-th percentile for visualization purposes. The percentile was determined from absolute values within the same region used for normalisation.

4.5.3.1 Differential interactions

Differential interactions were determined from the pairwise subtraction of normalized maps. For diagonal and RPM normalization, as well as for the computation of P-values, individual regions spanning the corresponding deletions were excluded (see Table 14). To avoid artefacts from KR normalization due to low coverage, maps were analysed on 10 kb resolution and a single row with very low coverage (chr13:54,450,001-54,460,000) was excluded. To account for the distance dependence of the magnitude of differences, each difference value was subtracted by the mean and divided by the standard deviation of the corresponding sub-diagonal. For the computation of mean and variance per sub-diagonal, values above the 99.5-th percentile were not included. P-values for the z-transformed difference values were computed using a standard normal distribution and further corrected for multiple testing with false discovery rate (Benjamini and Hochberg, 1995). In histograms, the distributions of z-scores do not fit the normal distribution perfectly, but show a bell-like shape as well.

4.5.3.2 Virtual Capture-C profiles

To obtain more fine-grained interaction profiles, we generated virtual Capture-C-like profiles based on the same filtered BAM files also used for the cHi-C maps and defined several virtual viewpoints of 10 kb size. A read pair was considered in the profile, when one mate mapped to the defined viewpoint region and the other one outside of it. Reads were counted per restriction fragment and binned further to a regular 1kb grid. In case a fragment spanned more than one bin, the count value was distributed proportionally to the overlaps. Afterwards, the profiles were smoothed by averaging over a sliding window of 5 bins. For coverage normalisation the profiles were divided by sum of counts in the enriched region on chr13 and multiplied with 10^3. The region ± 5kb around the viewpoint as well as the regions in Table 1 were mutually excluded from the

Methods

computation of the scaling factor in pairwise comparisons. The profiles were generated with custom Java code using htsjdk v1.139 (https://samtools.github.io/htsjdk/).

GENOTYPE	EXCLUDED REGION (MM9)
Pitx1 ^{delL}	Chr13:56,140,001-56,260,000
Pitx1 $^{RA2-Pitx1}$	Chr13:55,920,001-55,950,000
Pitx1 ^{del4}	Chr13:55,930,001-56,060,000
Pitx1 ^{RA3}	Chr13: 56,050,001-56,090,000
Neurog1 ^{del}	Chr13:56,320,001-56,370,000

Table 14 Regions excluded in the analysis of differential interactions and virtual Capture-C profiles

4.5.4 3D Polymer Modelling

4.5.4.1 The SBS polymer model

To investigate the 3D structure of the *Pitx1* locus, the *String & Binders Switch* (SBS) model was employed (Nicodemi and Prisco, 2009), which is a polymer physics model (Chiariello et al., 2016). In the SBS a chromatin filament is represented as a self-avoiding chain of beads that include binding sites for diffusing molecular binders; the polymer folds spontaneously, under the laws of physics, as the binders can bridge and loop their cognate binding sites. In each of the studied cases, forelimb (FL) and hindlimb (HL) tissues, and forelimb *Pitx1* ^{inv1} inversion, the specific SBS models for the *Pitx1* locus were established by a Simulated Annealing (SA) optimization procedure that finds the minimal number of different types of binding sites (different colours, in our notation) and their arrangement along the chain returning the best agreement between the corresponding cHi-C data and the equilibrium pairwise contact map derived by the polymer model. Molecular Dynamics (MD) computer simulations are used to derive an ensemble of the model equilibrium 3D conformations.

4.5.4.2 Polymer 3D representation

A single representative configuration of the *Pitx1* locus in the globule state is shown for each different cell type; to better visualize the relative positions of *Pitx1* and its *RA's*, a coarse grained version of the simulated polymer is pictured. In Figure is shown the entire polymer model. In all cases, the coordinates of each bead are interpolated with a smooth third-order polynomial splice curve. Figures are produced with POV-RAY (Persistence of Vision Pty. Ltd. 2004).

Methods

4.6 ANALYSIS OF MOUSE PHENOTYPES

4.6.1 RNA extraction

To quantify mRNA levels in WT and mutant mice, E11.5 forelimb or hindlimb buds were microdissected in cold PBS and immediately snap frozen and stored at -80°C. To isolate RNA, 500 µl TRIzol (Invitrogen #15596026) was added to the tissue and homogenised using a pestle. Next, 100 µl chloroform was added to the samples and pulse vortexed for 15 seconds, followed by centrifugation at 10000 RPM for 15 minutes at 4°C. The supernatant was transferred into a new tube and mixed with 1 volume 70%, and loaded onto a RNeasy Mini Kit column (QIAGEN #74104) following the manufacturer's instructions. The RNA was eluted in 30 µl RNase free H_2O and concentration measured on a Nanodrop.

4.6.2 cDNA synthesis

cDNA was generated using the Superscript III^{TM} First-Strand Synthesis System (Invitrogen #18080051) using 300 ng of RNA for reverse transcription.

4.6.3 qRT-PCR

To quantify the relative abundance of transcripts, qRT-PCR analysis of 3-5 biological replicates with technical triplicates was done using the GoTaq® qPCR Master Mix (Promega #A6101). The double delta Ct method (Livak and Schmittgen, 2001) was used to calculate fold change between wildtype and mutant samples. Statistical significance was calculated using the student's t-test.

4.6.4 Whole-mount in situ hybridization

All solutions used for whole-mount in situ hybridization are listed in Table 2. Buffers and solutions were treated with DEPC to inactivate RNase enzymes.

4.6.4.1 Generation of DIG-labelled in situ probe

The *Pitx1* mRNA expression in E11.5 mouse embryos were assessed by whole mount in situ hybridisation (WISH) whereby digoxigenin-labeled *Pitx1* and *H2afy* antisense riboprobe (Table 11) was generated by PCR amplification using E11.5 mouse limb bud cDNA, and subsequently cloned into the pTA-GFP vector. Next, SP6 and T7 primers were used to PCR amplified template in the pTA-GFP vector. 200 ng of template mixed with

Methods

DIG RNA labelling mix (Roche #11277073910), transcription buffer and RNA polymerase SP6 or T7 (Roche #10999644001) was incubated for 2 hours at 37°C. DNaseI (Roche #04716728001) treatment was done for 15 minutes at 37°C. Next, the reaction was inhibited by adding 0.2 mM EDTA with subsequent addition of 76 µl H20-DEPC, 0,4 M LiCl and 300 µl chilled 100 % ethanol and frozen over night at -80°C. The next day, the samples were centrifuged at 13000 RPM for 20 minutes at 4°C. The RNA pellet was washed once with 75% ethanol and dissolved in H2O-DEPC.

4.6.4.2 Hybridisation

Whole embryos were fixed overnight in 4% PFA/PBS. The embryos were washed in PBST (0.1% Tween), and dehydrated stepwise in 25%, 50% and 75% methanol/PBST and finally stored at -20°C in 100% methanol. The WISH protocol was as follows: Day 1) Embryos were rehydrated on ice in reverse methanol/PBST steps and washed in PBST. Bleaching of embryos in 6% H_2O_2/PBST for 1 hour and followed by washing in PBST. Treatment of embryos with 10 µg/ml Proteinase K/PBST for 3 minutes, followed by incubation with glycine/PBST, washed in PBST and finally re-fixed for 20 minutes with 4% PFA/PBS with 0.2% glutaraldehyde and 0.1% Tween 20. After further washing steps with PBST, embryos were incubated at 68°C in L1 buffer (50% deionised formamide, 5x SSC, 1% SDS, 0.1% Tween 20 in DEPC; pH 4.5) for 10 minutes. Next, embryos were incubated for 2 hours at 68°C in hybridisation buffer 1 (L1 with 0.1% tRNA and 0.05% heparin). Subsequently, embryos were incubated in hybridisation buffer 2 (hybridisation buffer 1 with 0.1% tRNA and 0.05% heparin and 1:500 DIG-probe) at 68°C o.n.

4.6.4.3 Conjugation of alkaline phosphatase

Day 2) Removal of unbound probe was done through a series of washing steps 3x30 minutes each at 68°C: L1, L2 (50% deionised formamide, 2x SSC pH 4.5, 0.1% Tween 20 in DEPC; pH 4.5) and L3 (2x SSC pH 4.5, 0.1% Tween 20 in DEPC; pH 4.5). Successively, embryos were treated for 1 hour with RNase solution (0.1 M NaCl, 0.01 M Tris pH 7.5, 0.2% Tween 20, 100 µg/ml RNase A in H_2O), followed by washing in TBST 1 (140mM NaCl, 2.7mM KCl, 25mM Tris-HCl, 1% Tween 20; pH 7.5). Next, embryos were incubated in blocking solution (TBST 1 with 2% calf-serum and 0.2% BSA) for 2 hours at RT, followed by incubation in blocking solution containing 1:5000 Anti-Digoxigenin-AP (Roche #11093274910) at 4°C o.n.

Methods

4.6.4.4 Washing unbound antibody

Day 3) Removal of unbound antibody was done through a series of washing steps 8x30 minutes at RT with TBST 2 (TBST with 0.1% Tween 20, and 0.05% levamisole/tetramisole) with a final incubation step at 4°C o.n.

4.6.4.5 Staining of embryos

Day 4) Staining of the embryos began by washing at RT with alkaline phosphatate buffer (0.02 M NaCl, 0.05 M $MgCl_2$, 0.1% Tween 20, 0.1 M Tris-HCl, and 0.05% levamisole/tetramisole in H_2O) for 3x20 minutes, followed by staining with BM Purple AP Substrate (Roche #11442074001). Limb buds from at least three embryos were analysed from each mutant genotype. The stained limb buds were imaged using a Zeiss Discovery V.12 microscope and Leica DFC420 digital camera.

4.6.5 LacZ Reporter Staining

For whole mount in situ *LacZ* reporter staining (Lobe et al., 1999) E11.5 mouse embryos were dissected in cold PBS, fixed in 4% paraformaldehyde (PFA) in PBS on ice for 30 minutes, followed by 3x washing with LacZ buffer (2 mM $MgCl_2$, 0.01% sodium deoxycholate, 0.02% Nonidet-40 in PBS). The embryos were incubated in staining solution (0.5 mg/ml X-gal-DMSO, 5 mM potassium ferrocyanide, 5 mM potassium ferricyanide in LacZ buffer) at 37°C for a few hours or until the desired staining was obtained. After staining, embryos were washed in LacZ buffer, fixed in 4% PFA in PBS, and stored at 4°C.

4.6.6 Skeletal Preparation

E18.5 foetuses were processed and bone and cartilage stained as described previously (Mundlos, 2000). In short, foetuses were kept in H_2O for 1-2 hours at RT and heat shocked at 65°C for 1 minute. The skin was taken off and the abdominal and thoracic viscera were removed using forceps. The foetuses were then fixed in 100% ethanol overnight. The next day, the cartilage was stained overnight using alcian blue staining solution (150 mg/l alcian blue 8 GX in 80% ethanol and 20% acetic acid). On the third day, foetuses were rinsed and post-fixed in 100% ethanol overnight. On the fourth day, initial clearing was done by incubating the foetuses for 20 minutes in 1% KOH in H_2O, followed by alizarin red (50mg/l alizarin red S in 0.2% KOH) staining of bones o.n. On the

fifth day and onwards rinsing and clearing is done using low concentrations of KOH. The stained embryos were stored in 80% glycerol, and were dissected limbs and imaged using Zeiss Discovery V.12 microscope and Leica DFC420 digital camera.

4.6.7 Micro-computer Tomography

Limbs of control and mutant mice were wrapped in plastic film and scanned ex vivo using a Skyscan 1172 X-ray microtomography system (Brucker microCT, Belgium). The scanning was done with a 0.5mm Aluminium filter at 80kV and 120 µA at 10 µm resolution. 3D model reconstruction was done with the Skyscan image analysis software CT-Analyser and CT-volume (Brucker microCT, Belgium).

5. RESULTS

5.1 PITX1 CIS-REGULATORY LANDSCAPE IN EMBRYONIC LIMBS

Developmental genes are typically embedded in large gene deserts containing *cis*-regulatory elements enabling a complex spatiotemporal expression in several tissues and organs. This expression pattern is thought to be achieved by long-range physical interactions between cognate enhancers and promoters (Andrey and Mundlos, 2017). *PITX1* is a pleiotropic transcription factor patterning various developing tissues and organs in vertebrates. In the mouse, *Pitx1* is expressed in the first branchial arch and its derivatives, Radke´s pouch, stomodeum, genital bud, posterior lateral plate mesoderm (LPM), and hindlimb mesenchyme (Lanctot et al., 1997; Szeto et al., 1996). However, little is known about the transcriptional control of *Pitx1* in these tissues as no enhancers have been identified in mammals so far. In stickleback fish, a pelvic fin-specific enhancer (*Pel*) controls tissue specific expression of in pelvic fins, a structure homologous to hindlimb buds (Chan et al., 2010; Yamanoue et al., 2010). However, *Pel* is not conserved in mammals (Chan et al., 2010). To dissect the *Pitx1* regulatory landscape specifically in mouse embryonic limb buds, we produce and inspected chromatin interaction profile, epigenetics marks, and perform various functional transgenic assays.

5.1.1 Pitx1 cis-regulatory landscape contains a pan-limb cis-regulatory region

To predict putative *Pitx1* hindlimb regulatory elements, Capture-C using *Pitx1* promoter as a viewpoint and ChIP-seq dataset for the enhancer associated mark H3K27ac in hindlimb buds at embryonic stage E10.5 were examined (Andrey et al., 2017). *Pitx1* Capture-C profile reveals chromatin interactions spanning over a 420 kb-large region, specifically, 90 kb downstream including the *Pcbd2* gene body, and 330 kb upstream including the *H2afy* region (Figure 8A). Within the chromatin interaction profile, *Pitx1* forms five prominent chromatin loops enriched in H3K27ac, indicating a putative regulatory activity and thus termed regulatory anchors 1-5 (*RA1-5*, whereby *RA2* corresponds to *Pitx1* itself) (Figure 8A). To test for *in vivo* enhancer activity of regions *RA3-RA5,* the fragments of interest were cloned in front of a *LacZ* reporter and integrated at the *ColA1* locus in mouse embryonic stem cells (mESCs). mESCs were aggregated and embryos sacrificed at E11.5 for *LacZ* staining of whole embryos. *RA3* shows enhancer activity in Radke's pouch where *Pitx1* is expressed (Figure 8A).

Results

Surprisingly, both *RA4* and *RA5* show fore- and hindlimb bud enhancer activity (*pan*-limb), though *Pitx1* is never expressed in forelimb tissue (Figure 8A). Specifically, *RA4* shows activity in prospective cartilaginous condensations of the limb, thus termed *Con* enhancer. *RA5*, named *Pen*, for *pan*-limb enhancer, shows strong mesodermal activity in both limbs (Spielmann et al., 2012; Visel et al., 2007).
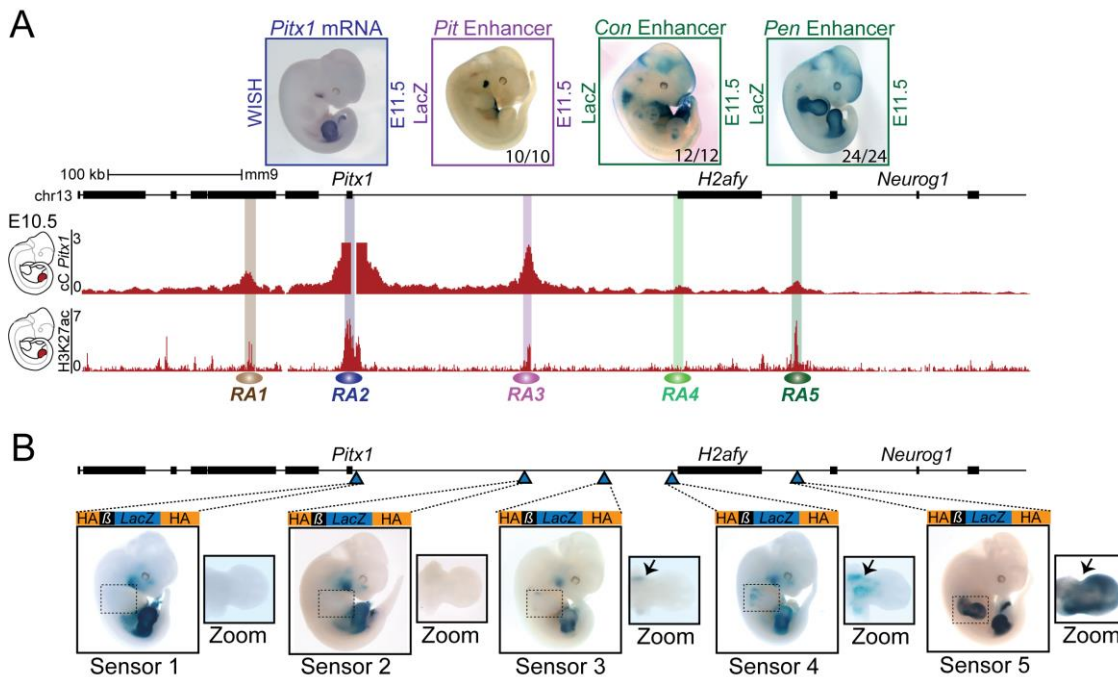
Figure 8 Identification of *cis*-regulatory elements at the *Pitx1* locus

A. Chromatin tracks show *Pitx1* Capture-C (Cc) interaction profile and ChIP-seq track for H3K27ac in E10.5 hindlimb buds. Putative regulatory regions, termed regulatory anchors are highlighted (*RA1-RA5*). From top left: WISH at E11.5 *Pitx1* mRNA expression, *LacZ* reporter assay of *Pit* (Radke's pouch), *Con* (*pan*-limb), and *Pen* (strong *pan*-limb). **B.** Sensors 1-5 reflect the endogenous regulatory activity at the *Pitx1* locus in E11.5 embryos. From left: Sensor 1 (strong *Pitx1* expression pattern), Sensor 2 (mediocre *Pitx1* expression pattern), Sensor 3 and Sensor 4 (weak *Pitx1* expression pattern and ectopic proximal forelimb activity), and Sensor 5 (strong forelimb and hindlimb). Note its similarity with *Pen*.

This unexpected *pan*-limb activity of *Con* and *Pen* might indicate that the *LacZ* reporter set-up, which is integrated at the exogenous *ColA1* locus, does not reflect the endogenous enhancer activity. Hence, at the *ColA1* locus, enhancers are tested in the absence of the endogenous regulatory information such as the surrounding epigenetic landscape and 3D chromatin architecture. To assay the regulatory potential at *RA2-5* within their native configuration, a *LacZ* reporter cassette containing a minimal *ß-globin* promoter was integrated into the mouse *Pitx1* locus. The naive *ß-globin* promoter is activated by surrounding regulatory information, and thus the whole construct is referred to as regulatory sensor (Ruf et al., 2011). Five positions at the *Pitx1* locus were chosen for targeted CRISPR-Cas9 mediated knock in of sensors: adjacent to *RA2, RA3,*

between *RA3* and *RA4, RA4,* and *RA5*. mESCs were aggregated and embryos sacrificed at E11.5 and tested for *LacZ* activity (Figure 8B). Sensor 1, which locates adjacently to *RA2/Pitx1*, strongly recapitulates the *Pitx1* endogenous expression in first branchial arch and its derivatives, in Radke's pouch, in the posterior LPM, and in hindlimb buds, but does not display forelimb activity (Figure 8B). Sensor 2, which locates adjacently to *RA3*, recapitulates endogenous *Pitx1* expression as well, though with a less strong staining compared to Sensor 1 (Figure 8B). Sensor 3 and Sensor 4 which locate adjacently to *RA3* and *RA4*, respectively, display a weaker *Pitx1* expression pattern, with reduced activity in the posterior LPM and distal limb bud. In this assay, both sensors display an additional activity in the proximal forelimb bud (Figure 8B). Remarkably, Sensor 5, which is integrated next to *RA5/Pen*, shows strong fore- and hindlimb staining, but no activity in the pituitary, first branchial arch, and posterior LPM. Thus, Sensor 5 corroborates *Pen* activity, as tested in the *LacZ* reporter assay (Figure 8A and 8B).

Thus, the *Pitx1* hindlimb regulatory landscape consist of five regulatory anchors, two of which show *in vivo pan*-limb regulatory potential. This puzzling finding raises the question whether these regulatory anchors form a functional part of *Pitx1* regulation in the limb.

5.1.2 *Pitx1* is regulated by *pan*-limb enhancer, *Pen*

To functionally dissect the *Pitx1* regulatory landscape, I generated an allelic series dissecting the *Pitx1* landscape (Kraft et al., 2015). Whole mount in situ hybridisation (WISH) and qRT-PCR analyses of homozygous mutant hindlimbs at E11.5 was used to detect putative changes in *Pitx1* gene expression. The first deletion in the allelic series removes the entire telomeric-interacting domain, upstream of *Pitx1*, including *Pen* (*Pitx1^{del1}*, 330 kb). *Pitx1$^{del1/del1}$* hindlimbs show a 95% decrease (20-fold) in *Pitx1* expression, which is evident in the WISH staining with low amounts of mRNA detectable (Figure 9A and Figure 22). Skeletal preparations of *Pitx1$^{del1/del1}$* E18.5 foetuses exhibit the characteristics of a *Pitx1* loss-of-function phenotype: the femur is shortened, the fibula is absent or severely reduced, the tibia is broader and shorter, the patella is lost, and the calcaneus is severely reduced (Figure 9B) (Lanctot et al., 1999; Marcil et al., 2003; Szeto et al., 1999). Thus, the 330 kb region upstream of *Pitx1*, including the *pan*-limb region, encompasses the regulatory elements critical for robust *Pitx1* hindlimb

expression. Accordingly, further analyses of *Pitx1* regulation were focused on the upstream telomeric region.
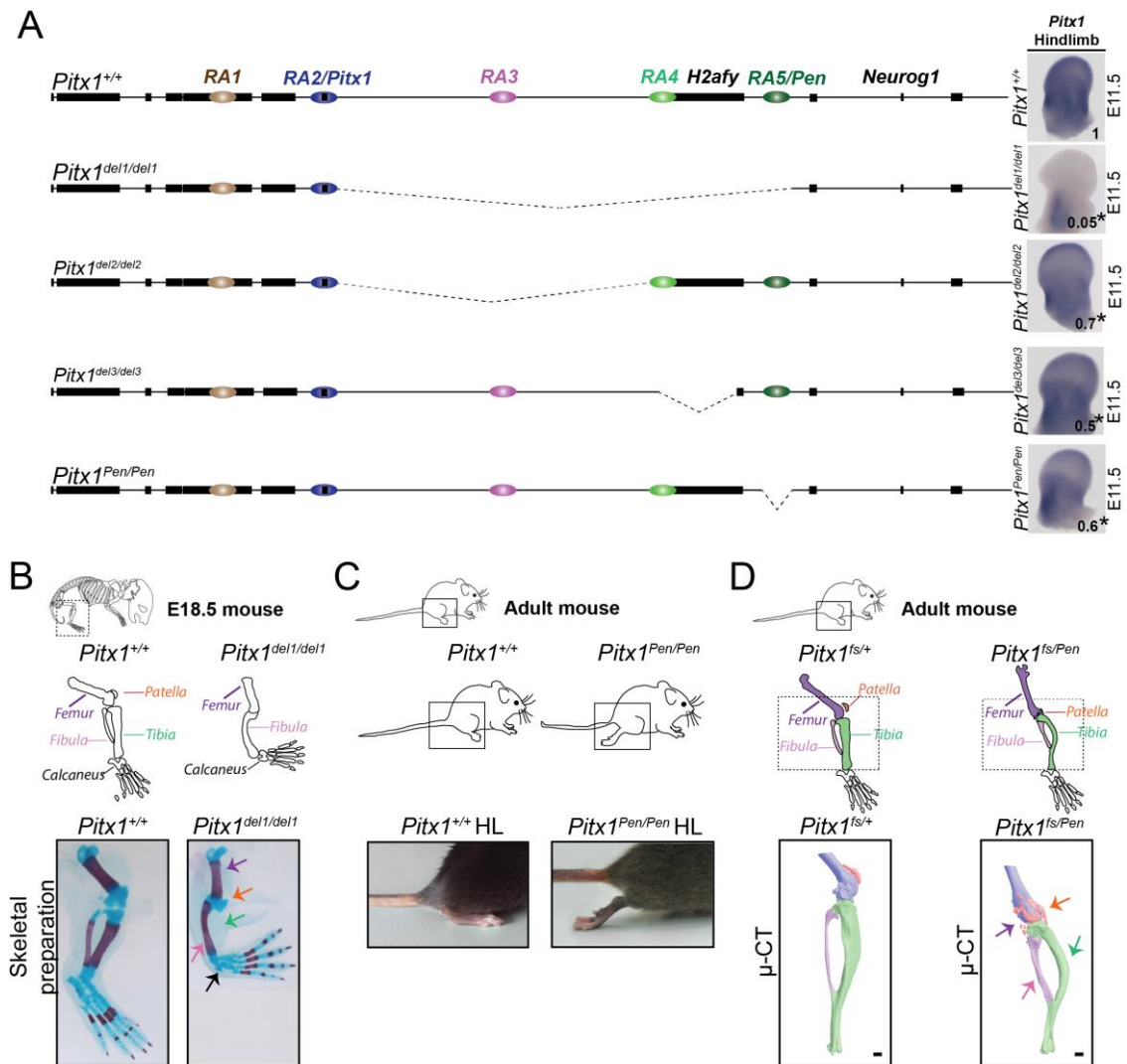
Figure 9 CRISPR-Cas9 dissection of the *Pitx1* hindlimb regulatory landscape

A. CRISPR-Cas9 allelic series and changes in *Pitx1* gene expression shown by WISH of E11.5 hindlimbs. The number indicates fold-change in *Pitx1* mRNA expression as quantified by qRT-PCR and asterisk designates significant changes (p> 0.05). From top: $Pitx1^{del1/del1}$ 330 kb deletion of *Pitx1* upstream chromatin interacting landscape results in 95% reduction of *Pitx1*. $Pitx1^{del2/del2}$ 235 kb deletion of gene desert causes 30% reduction of *Pitx1*. $Pitx1^{del3/del3}$ 58 kb del of *H2afy* gene body results in 50% reduction of *Pitx1*. $Pitx1^{Pen/Pen}$ removing the 1.2 kb *Pen* results in 40% reduction of *Pitx1* mRNA expression. **B.** Alcian blue and alizarin red skeletal staining of E18.5 $Pitx1^{del1/del1}$ foetus with a severe reduction of femur (purple arrow), fibula (pink arrow), calcaneus (black arrow), and a loss of fibula (green arrow) and patella (orange arrow). **C.** $Pitx1^{Pen/Pen}$ adult animals drag their hindlimb behind, characteristic of club feet. **D.** Micro-CT of adult control $Pitx1^{fs/+}$ with no skeletal phenotype, and compound heterozygous $Pitx1^{fs/Pen}$ mutant mice displaying hindlimb skeletal malformation whereby the rotation of the femur is altered (purple arrow), fibula is bowed (pink arrow), tibia broadened (green arrow), and patella is severely fragmented (orange arrow).

Results

To identify the exact location of key *Pitx1* hindlimb regulatory elements, I deleted smaller fragments within the critical 330 kb region. Removing the gene desert spanning from *Pitx1* to *H2afy* (*Pitx1$^{del2/del2}$*), results in a 30% reduction in *Pitx1* expression in hindlimbs tissue (Figure 9A and Figure 22B). Deletion of *RA4* and *H2afy* gene body (*Pitx1$^{del3/del3}$*) induces a 50% loss of *Pitx1* expression in hindlimb buds (Figure 9B and Figure 22C). Remarkably, eliminating the 1.2 kb *Pen* region (*Pitx1$^{Pen/Pen}$*) results in 40% reduction in *Pitx1* expression in hindlimb tissue (Figure 9B and Figure 22D). 6-week-old *Pitx1$^{Pen/Pen}$* mice do not show any obvious skeletal phenotype (not shown), but 10% of the mice develop unilateral clubfeet, observed by the dragging of their hindlimbs (Figure 9C). This is reminiscent of a partially penetrant clubfoot phenotype observed in 8.9% of *Pitx1* heterozygous knockout mice (Alvarado et al., 2011). Moreover, this phenotype is also observed in human patients with *PITX1* haploinsufficiency (Alvarado et al., 2011). To verify the functional regulatory link between *Pen* activity and *Pitx1* transcription, we crossed *Pitx1* loss of function mice (*Pitx1^{fs}*), which carry a frameshift mutation in *Pitx1* exon 1, and *Pitx1^{Pen}* mice. All *Pitx1$^{fs/Pen}$* compound heterozygous mice developed bilateral clubfeet, thus confirming the regulatory link between *Pen* and *Pitx1*. Moreover, micro-CT analysis of adult mice shows a skeletal malformation in characterised by a broader fibula broader, a reduced and rotated femur, which articulates with a deformed fibula. Strikingly, the identification of a patella in the micro-CT scan is challenging as it appeared severely fragmented (Figure 9D).

This series of deletions demonstrated that the *Pitx1* hindlimb regulatory landscape is confined to the upstream telomeric region, which includes the strong fore- and hindlimb enhancer *Pen*. This observation challenges the prevailing paradigm that enhancer activity reflects the tissue-specific expression of their target gene, suggesting the presence of additional regulatory layer.

5.1.3 Distinct chromatin and epigenetic profiles in forelimb and hindlimb tissues

To decipher how the *pan*-limb regulatory potential of the distal *Pitx1* regulatory landscape is translated into a hindlimb specific transcriptional output, we examined *Pitx1* Capture-C interaction profile between fore- and hindlimb tissues. Comparison of *Pitx1* Capture-C in E10.5 limbs, reveals striking differences in the chromatin interaction profile between the two tissues (Figure 10A) (Andrey et al., 2017). In forelimb tissue,

the *Pitx1* interactions with regulatory anchors is weaker compared to hindlimb tissue (Figure 10A). Instead, a forelimb specific distal interaction with the developmental neuronal gene *Neurog1*, which is inactive in limbs, is detected (Figure 10A). Subtraction of the forelimb and hindlimb Capture-C profiles further emphasise this difference: in hindlimbs, prominent chromatin loops are established with *RA1, RA3,* and *RA5/Pen*, whereas these are reduced in forelimb tissue, but instead one can observe the forelimb specific chromatin interaction with *Neurog1* (Figure 10A). Thus, *Pitx1* establishes fundamentally different chromatin interactions in hindlimbs with regards to forelimbs.
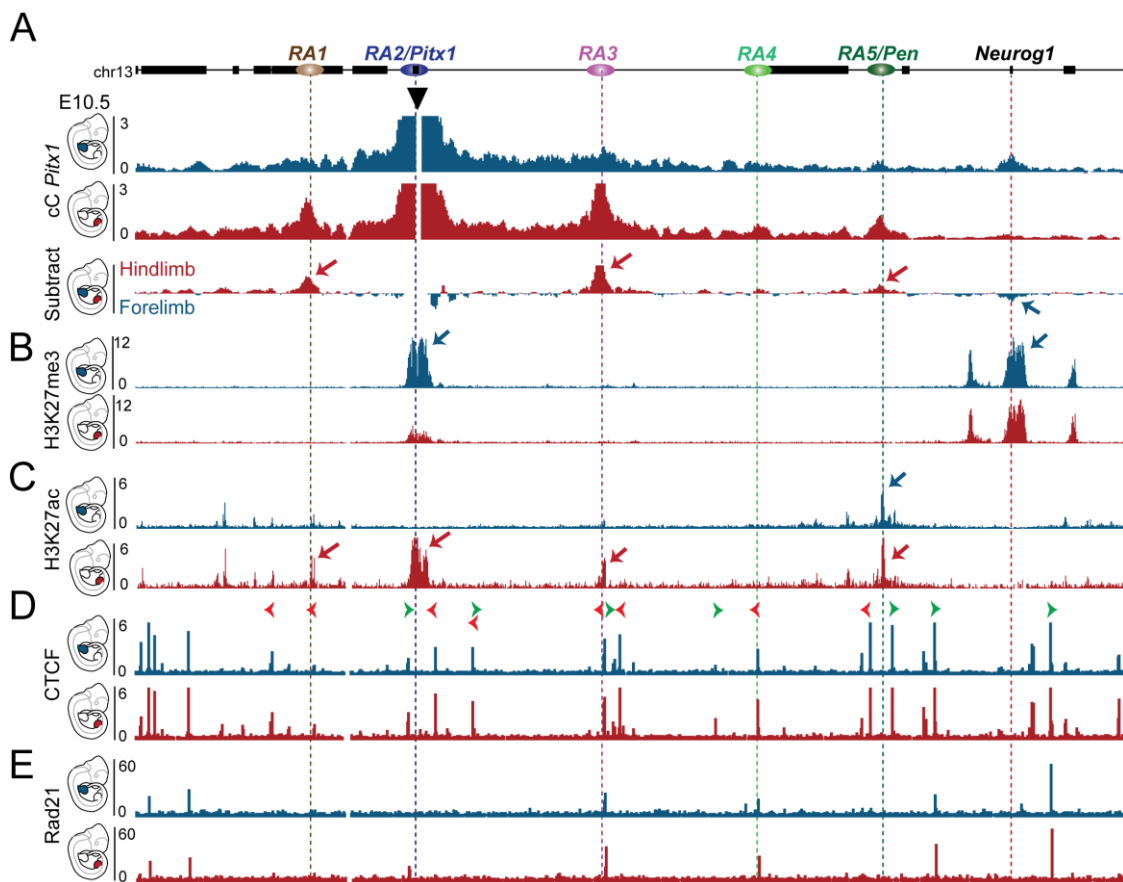
Figure 10 *Pitx1* chromatin landscape in embryonic forelimbs and hindlimbs

A. Capture-C (cC) *Pitx1* tracks in E10.5 forelimb (blue) and hindlimb (red) tissue, and the subtraction between the two tissues, with hallmarks of prominent hindlimb interactions indicated by red arrows and prominent forelimb interaction indicated by a blue arrow. **B.** ChIP-seq tracks of H3K27me3 in E10.5 forelimb (blue) and hindlimb (red) tissues. Polycomb mediated interaction between *Pitx1* and *Neurog1* in forelimb tissue is indicated by blue arrows. **C.** ChIP-seq tracks of H3K27ac in E10.5 forelimb (blue) and hindlimb (red) tissues. Chromatin interactions between active regions in hindlimb tissue are indicated by red arrows. Note the heavy acetylation of *pan*-limb region in both fore- and hindlimb tissue (blue and red arrows). **D.** ChIP-seq tracks of CTCF in E10.5 forelimb (blue) and hindlimb (red) tissues. CTCF motif orientation is indicated by green (+) and red (-) arrows. **E.** ChIP-seq tracks of Rad21 in E10.5 forelimb (blue) and hindlimb (red) tissue.

To investigate the nature of these chromatin interactions, E10.5 limb ChIP-seq tracks of H3K27ac, H3K27me3, CTCF and Rad21 (cohesin) were examined (Andrey et al., 2017).

The chromatin modifications do not change much between both tissues, except over the *Pitx1* gene body and at *RA3*. In forelimb tissues, characteristic of repressed developmental genes both *Pitx1* and *Neurog1* are covered in Polycomb-mediated chromatin mark H3K27me3 marking inactive genes (Figure 10B). Interestingly, these two Polycomb repressed regions, located 420 kb apart, overlap with the forelimb-specific chromatin interaction (Figure 10A). In hindlimb tissue, H3K27me3 is depleted from the *Pitx1* gene body and replaced by a counterpart mark H3K27ac, releasing its interaction with the H3K27me3-covered *Neurog1.* In turn, *Pitx1* establishes chromatin interactions with H3K27ac-enriched regulatory anchors (Figure 10C and 10A). Intriguingly, the *pan*-limb region is heavily enriched for H3K27ac in both forelimb and hindlimb, but in forelimbs, only a weak interaction occurs between *Pitx1* and this region.

The tissue-specific binding of chromatin loop extrusion factors CTCF and cohesin have been suggested to dictate the tissue specific changes in chromatin interactions at different loci (Rao et al., 2014). ChIP-seq tracks of CTCF and Rad21 (a cohesin complex subunit) in E10.5 forelimb and hindlimb buds show enriched binding at the *RA2* to *RA5/Pen*, however no differential binding between the two tissues is observed at the regulatory anchors (Figure 10D and 10E). This suggests a different mechanism promoting the differential chromatin interactions.

Together, the comparison of chromatin-associated features between fore- and hindlimbs unmasks two divergent states of chromatin interaction: one maintaining *Pitx1* in an inactive state, through H3K27me3-mediated Polycomb interaction with the *Neurog1* region in forelimb tissue, and another one promoting the contact between Pitx1 and other active H3K27ac-enriched regions in hindlimb tissues.

5.2 3D REGULATION OF *PITX1* IN EMBRYONIC LIMBS

Genome wide analysis of chromatin conformation by Hi-C has demonstrated the partition of the genome into topologically associating domains (TADs), which display higher internal chromatin interactions separated by regions of low interactions called boundaries (Dixon et al., 2012; Nora et al., 2012). TADs are conventionally thought to facilitate and confine tissue-specific enhancer-promoter interactions in a tissue-independent manner (Dixon et al., 2015; Dowen et al., 2014; Symmons et al., 2014;

Results

Vietri Rudan et al., 2015). In contrast, intra-TAD interactions often display dynamic changes during cell differentiation (Bonev et al., 2017; Phillips-Cremins et al., 2013). To investigate the dynamics in chromatin architecture of the *Pitx1* locus in more detail, two variants of chromosome conformation capture were used. Initially 4C-seq was deployed, followed by Capture-C (described above), both producing interaction profiles from a pre-defined viewpoint. Alternatively, to probe the *Pitx1* locus in a systematic way, capture Hi-C (cHi-C) was used to establish the interaction map of a 3Mb region surrounding the *Pitx1* locus, ultimately enabling 3D-modelling of the region using SBS polymer physics.

5.2.1 Comparison of Capture-C and 4C-seq in forelimb and hindlimb buds

First, 4C-seq using different viewpoints within the *Pitx1* regulatory landscape was compared to the available Capture-C results in E11.5 mouse forelimb and hindlimbs (Andrey et al., 2017) (Figure 11A). *Pitx1* Capture-C in E11.5 limb corroborates the E10.5 *Pitx1* interaction profile by showing similar profiles (Figure 10A and 11A). In contrast, 4C-seq analysis in E11.5 fore- and hindlimbs using two alternating *Pitx1* viewpoints (*Pitx1 TSS1* – more 5') and (*Pitx1 TSS2*– more 3'), only partially recapitulated the Capture-C results (Figure 11B). Specifically, *Pitx1 TSS1* 4C-seq profile in fore- and hindlimb show a similar interaction profile between the two tissues, where *Pitx1* forms prominent interactions with *RA3* and *RA4* (Figure 11B). Alternately, *Pitx1 TSS2* profile in forelimbs shows reduced chromatin interactions with *RA3* and *RA4*, and increased interactions with the *Neurog1* region in forelimb tissue (Figure 11C). In hindlimb tissue, prominent interactions with *RA3* and *RA4* are detected (Figure 11C). However, the two 4C-seq *Pitx1* profiles do not capture the same extent of chromatin interaction as Capture-C (Figure 11B and 11C). With 4C-seq, any viewpoint can be used in the same regulatory landscape without losing the reciprocal viewpoint information, whereas in the Capture-C, capturing a viewpoint excludes it from being captured from an alternative viewpoint (Hughes et al., 2014). Accordingly, I designed 4C-seq viewpoints in the *RA5/Pen* region (*Pen*) and *H2afy* promoter (*H2afy TSS*). From the *Pen* viewpoint, chromatin interactions are limited to the neighbouring genes *H2afy* and *Tifab* (Figure 11D). From *H2afy TSS* viewpoint, chromatin interactions are confined centromerically to *RA4* and telomerically up to the immediate neighbour gene *Tifab* (Figure 11E).
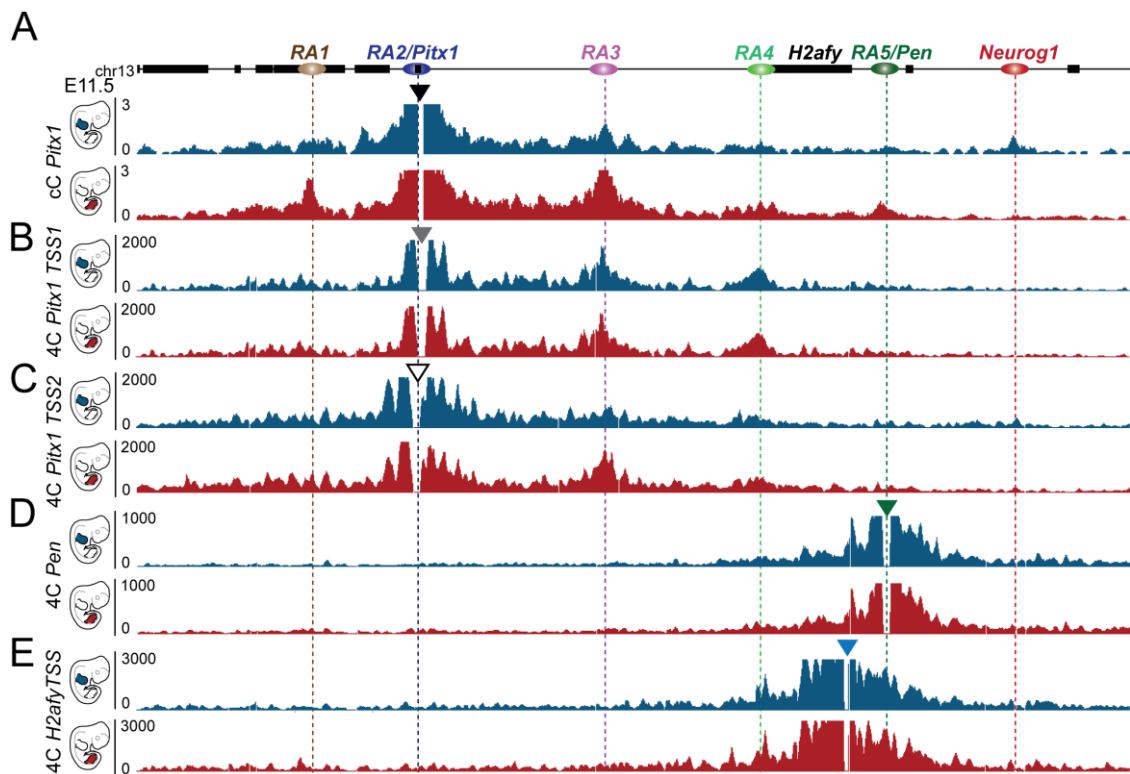
Figure 11 Comparison of Capture-C and 4C-seq in forelimb and hindlimbs

A. Capture-C (cC) *Pitx1* tracks in E11.5 forelimb (blue) and hindlimb (red) tissue. **B.** 4C-seq *Pitx1 TSS1* tracks in E11.5 forelimb (blue) and hindlimb (red) tissue. **C.** 4C-seq *Pitx1 TSS1* tracks in E11.5 forelimb (blue) and hindlimb (red) tissue. **D.** 4C-seq *Pen* tracks in E11.5 forelimb (blue) and hindlimb (red) tissue. **E.** 4C-seq *H2afy TSS* tracks in E11.5 forelimb (blue) and hindlimb (red) tissue.

Here, the use of 4C-seq did not provide substantial information on the extent of chromatin interactions at the *Pitx1* locus by using alternating viewpoints, which may be due to the selected restriction enzyme fragment used to capture interactions. 4C-seq is limited to the presence of restriction enzyme fragments with a minimum size of 350 bp in which specific primers can be designed (van de Werken et al., 2012). Furthermore, this often produces biases in PCR amplification of shorter versus longer fragments (Schmitt et al., 2016). Capture-C on the other hand uses RNA probes to pull down a region and can span several restriction enzyme sites increasing the size of the bait (Hughes et al., 2014). Thus, this may explain the difference in the observed difference between the two techniques.

5.2.2 cHi-C reveals a forelimb to hindlimb switch in chromatin architecture

To probe the *Pitx1* locus in a systematic way, we used cHi-C to produce complete interaction maps of a 3Mb region surrounding the *Pitx1* locus in mice. cHi-C was performed on both forelimb and hindlimb buds at E11.5. 400 million sequencing reads

Results

and 10 million uniquely mapped paired reads from four biological replicates were merged to produce high resolution interaction maps (5 kb binning). The resulting cHi-C heatmaps show the organisation of the 3Mb region into several TADs. Surprisingly, no TAD structure is observed at the *Pitx1* locus (Figure 12A). Indeed, no TAD boundaries separating the *Pitx1* regulatory landscape from upstream and downstream genes such as *Neurog1* are observed. Instead, the locus is organised into a handful of domains, with high intra-domain interactions, that are separated by the regulatory anchors, which in turn interact with one another, producing a structure that we refer to as a multi-anchor domain (MAD) (Figure 12B). A closer inspection of this type of organisation of chromatin architecture in the fore-and hindlimb tissues reveal dynamic changes in the MAD organisation. Specifically, in forelimbs the MAD is composed of four domains, with weak intra-domain interactions, and moderate interactions between *RA2/Pitx1* and *RA3* and *RA4*, as well as a prominent interaction between *RA2/Pitx1* and *Neurog1,* in agreement with the Capture-C results (Figure 12B ad 10A). On the contrary, the MAD organisation in hindlimb tissue comprises four sharply segregated domains and strong chromatin loops from *RA2/Pitx1* with *RA1, RA3, RA4*, and *RA5/Pen*, confirming the Capture-C results (Figure 12C and 10A). This forelimb-hindlimb difference is particularly evident in cHi-C maps subtraction, in which high intra-domain interaction and the strong gain of chromatin interactions between the regulatory anchors in hindlimb tissue are detected (Figure 12D).

To quantify the changes in frequency of chromatin interactions between the regulatory anchors in forelimb versus hindlimb tissues, we calculated the changes in chromatin interactions that are represented as pie charts in Figure 13A. In agreement with the subtraction of cHi-C maps, the strongest biases in chromatin interactions are found between *RA2/Pitx1* and *RA1, RA2/Pitx1* and *RA3, RA2/Pitx1* and *RA5/Pen* in the hindlimb*,* as well as *RA2/Pitx1* and *Neurog1* in the forelimb, showing up to 66% difference (Figure 13A). Moreover, the computational analysis of the insulation score of the region that detects the strength of intra-domain interactions (rise) and insulation between the domains (drop), further supports the dense compaction of the domains and the stronger insulation between them in hindlimb tissues (Figure 13A).
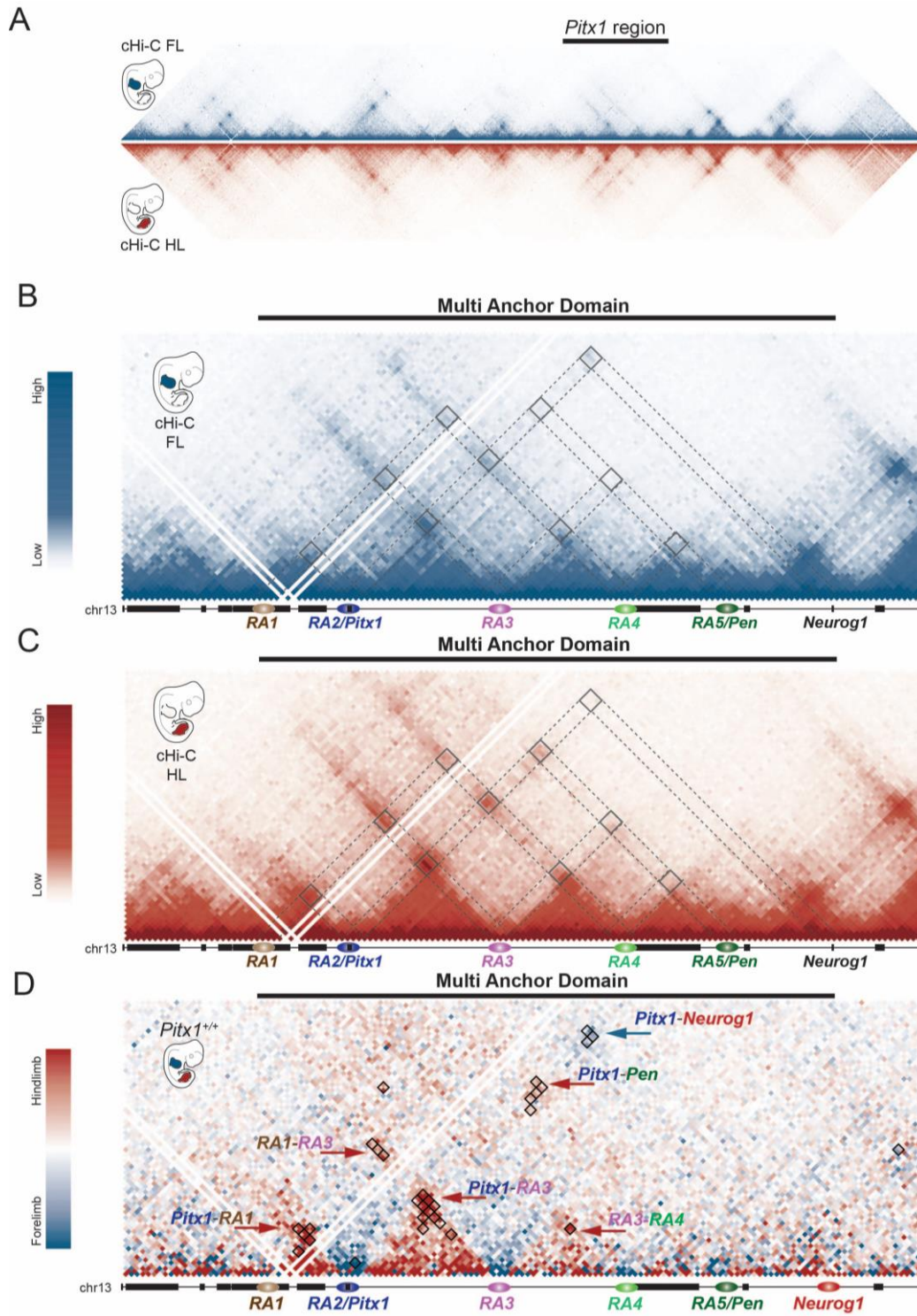
Figure 12 Capture-HiC reveals a switch in 3D chromatin architecture of forelimbs vs hindlimbs

A. cHi-C heatmap of forelimb (blue) and hindlimbs (red) illustrates the 3 Mb enriched region including *Pitx1* MAD and surrounding TADs. **B.** cHi-C zoom of forelimb *Pitx1* locus: the landscape is subdivided into five domains separated by RAs forming moderate to minimal loop formation between *RA2/Pitx1* with *RA1*, *RA3*, and *RA5/Pen*, but a prominent interaction between *RA2/Pitx1* with the functionally unrelated *Neurog1* gene. **C.** cHi-C zoom of hindlimb *Pitx1* locus: the region is subdivided into four distinct subdomains, where *RA2/Pitx1* establishes strong contacts with all the RAs including the distal *RA5/Pen* region, but not *Neurog1*. **D.** The fore- and hindlimb-specific biases in interactions between different RAs and the neighbouring *Neurog1* gene are particularly evident from subtraction between cHi-C heatmaps and black diamonds show the significant (FDR>0.05) differential interaction in chromatin contacts.

Results

An alternating way of visualising the insulation conferred by the regulatory anchors is by producing virtual Capture-C viewpoints (vC) whereby a desired 5 kb viewpoint along the *Pitx1* cHi-C region can be picked and its interaction profile displayed. Interestingly, depending on the location of the viewpoint, different interactions are observed e.g. vC-192 captures the intra-domain chromatin interaction between *RA1* and *RA2/Pitx1*, while vC-194 reflects the interactions between *RA2/Pitx1* with *RA1, RA3, RA4,* and *RA5/Pen*, and vC-195 reproduces the intra-domain interaction in between *RA2/Pitx1* and *RA3* (Figure 13B). This partially upholds the Capture-C and 4C-seq data, whereby depending on the chosen viewpoint, for example the *Pitx1* viewpoint (vC-193 and vC-194), all chromatin interactions between *Pitx1* and its regulatory anchors are captured. By moving the viewpoint inside a domain (vC-192 and vC-195), only the domain itself is captured (Figure 13B). Together, these results support the existence of a dynamic chromatin microarchitecture at the *Pitx1* locus.

To facilitate further interpretation of the cHi-C data beyond 2D-interaction maps, 3D-models using SBS polymer physics of the fore- and hindlimb cHi-C data were computed (Chiariello et al., 2016). The 3D-model of forelimb cHi-C data show the segregation of the *Pitx1* locus into two major chromatin hubs: one containing *RA2/Pitx1*, *RA3,* and *Neurog1*, physically separating it from the second hub containing *RA4* and *RA5/Pen* (Figure 13C). Of note is the observation that here *RA2*/*Pitx1* is buried within its chromatin hub together with *Neurog1*, while *RA5/Pen* sits at the outskirt of the adjacent hub, thereby preventing physical interactions between inactive *Pitx1* and active *pan*-limb region (Figure 13C). In contrast, 3D-modelling of hindlimb cHi-C data displays a segregation of the locus into three chromatin hubs: one containing *RA1*, the second containing *RA2/Pitx1* with *RA3*, and the third containing *RA4* together with *RA5/Pen* located at the border, while *Neurog1* is tucked away on the opposite side (Figure 13D). Here *Pitx1* is positioned at the verge of its hub enabling interactions with its active *pan*-limb region. Taken together, these results support the existence of a novel and dynamic chromatin architecture of the *Pitx1* locus that differs from the conventional understanding of TAD organisation. Whereas TADs and their boundaries are cell invariant and confine cognate enhancer-promoter pairs (Symmons et al., 2014), the here-described MAD consists of promoter and unspecific enhancer elements which are likely regulated by dynamic tissue specific interactions between the regulatory anchors.

Results
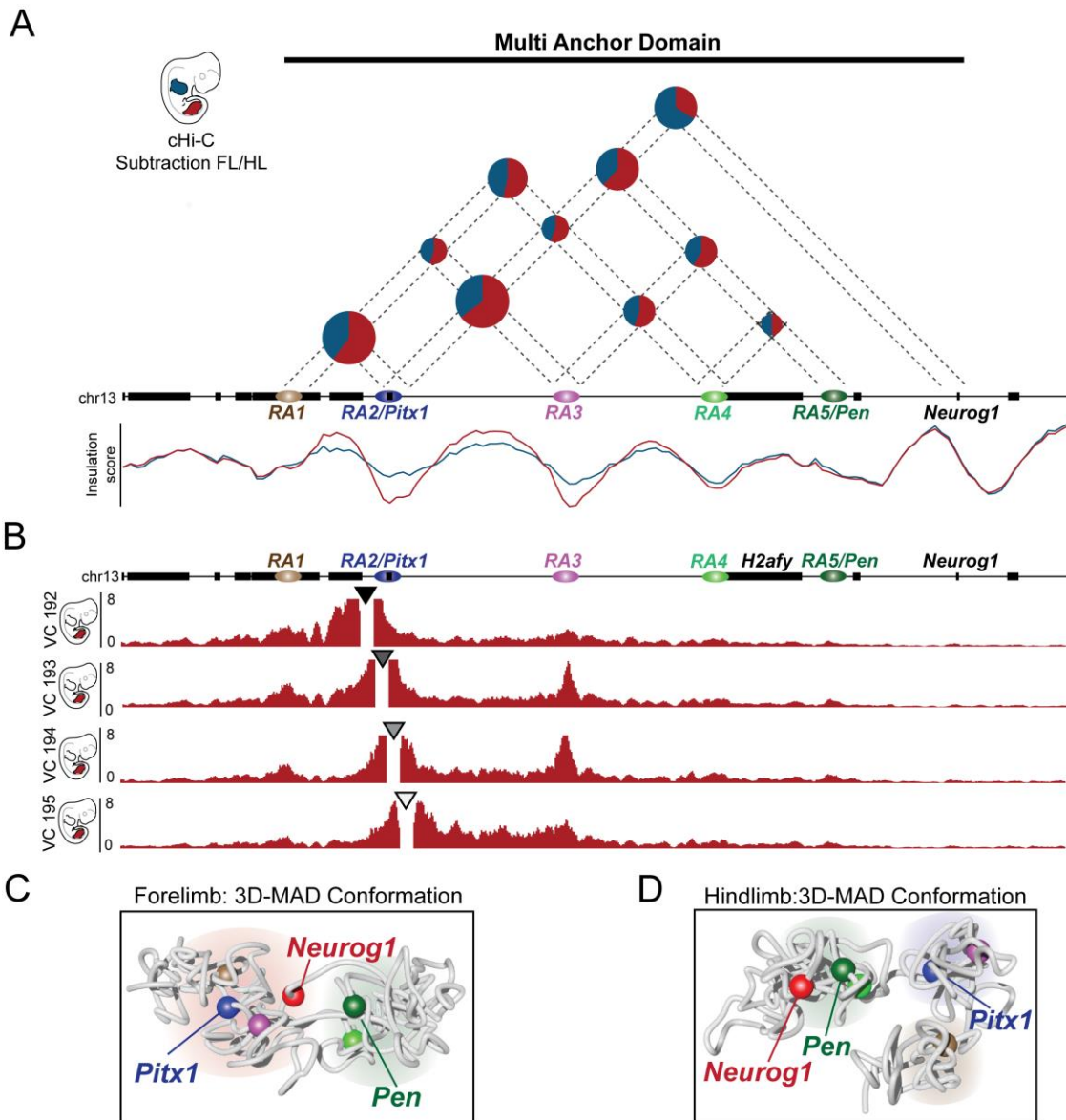
Figure 13 Tissue specific chromatin microarchitecture and 3D conformation of the *Pitx1* locus

A. Top: Quantification of differential contacts between forelimb (blue) and hindlimb (red) depicted as pie-charts. Bottom: Insulation score calculated from cHi-C data in forelimb (blue) and hindlimb (red) tissue. **B.** Four different vC around the *Pitx1* gene body and promoter reveal the microarchitecture of the *Pitx1* locus **C.** SBS polymer model of forelimb cHi-C data shows the predicted 3D conformation of the MAD into two chromatin hubs: one with *RA2/Pitx1, RA3,* and *Neurog1*, the other with *RA4* and *RA5/Pen*. **D.** SBS polymer model of hindlimb cHi-C data shows the predicted 3D conformations of the MAD into three chromatin hubs: one with *RA1* alone, second with *RA2/Pitx1* and *RA3,* and third with *RA4, RA5/Pen*, and *Neurog1*.

These dynamic changes in chromatin architecture correlate with loss of repressive H3K27me3 marks and establishment of active H3K27ac marks. Thus, the folding of the domains into a tissue-specific 3D-conformation plays an active role in altering the endogenous activity of *Pen*, which is critical for hindlimb versus forelimb identity, into a tissue specific transcriptional output of *Pitx1*.

5.3 PERTURBATION OF *PITX1* REGULATORY LANDSCAPE RESULTS IN LIMB MALFORMATIONS

Little is known about the mechanism underlying the establishment of chromatin interactions and TAD structures. Knock out studies of chromatin modifiers and architectural proteins such as PRC1, CTCF and cohesin have led to the proposition of at least two modes of folding the genome (Nora et al., 2017; Schoenfelder et al., 2015). One model proposes a CTCF and cohesin dependent mechanism, which has been suggested to be key players of the loop extrusion complex, whereby cohesin extrudes chromatin through its ring-like structure forming loops (Fudenberg et al., 2016; Sanborn et al., 2015). The size of the loops is determined by CTCF binding to the genome forming a "road-block" (Fudenberg et al., 2016; Sanborn et al., 2015). A second mechanism proposed to induce dynamic conformational changes of chromatin interactions, is the clustering of homotypic chromatin regions to form smaller (average 150 kb) and more dynamic chromatin domains (Bonev et al., 2017; Schwarzer, 2017). This principle has been shown to correlate across several species (Rowley et al., 2017). Here, the 3D conformation of the *Pitx1* locus plays an active role in modulating the tissue-specific interaction between *Pitx1* and its unspecific limb regulatory elements in hindlimb tissue, while maintaining it inactive in forelimb tissue. To identify which regions of the *Pitx1* regulatory landscape are key for its activation specifically in the hindlimb tissue and its silencing in forelimb tissue, cHi-C analysis of CRISPR-Cas9 engineered mice was performed in limbs buds at E11.5. Furthermore, reengineering of structural variants causing Liebenberg syndrome in mice was undertaken to study their effect on MAD organisation and gene regulation.

5.3.1 *Cis*-regulation of hindlimb chromatin folding of the *Pitx1* locus

Comparison between forelimb and hindlimb chromatin architecture shows that the regulatory anchors display the strongest bias in interactions, and thus might be directly involved in the generation of the hindlimb-specific chromatin conformation. Specifically, the chromatin loop formed between the two partners *RA2/Pitx1* and *RA3* undergoes the most striking differences (Figure 13A). To assess whether *RA3* is a key partner to establish a hindlimb architecture, we engineered a 32.5 kb deletion that removes the CTCF-bound and H3K27ac-enriched *RA3* region (*Pitx1$^{RA3/RA3}$*). cHi-C and derived vC

viewpoints in *Pitx1^RA3/RA3* hindlimb buds show a fusion of the two flanking chromatin domains, normally separated by *RA3* (Figure 14B and Figure 23A).
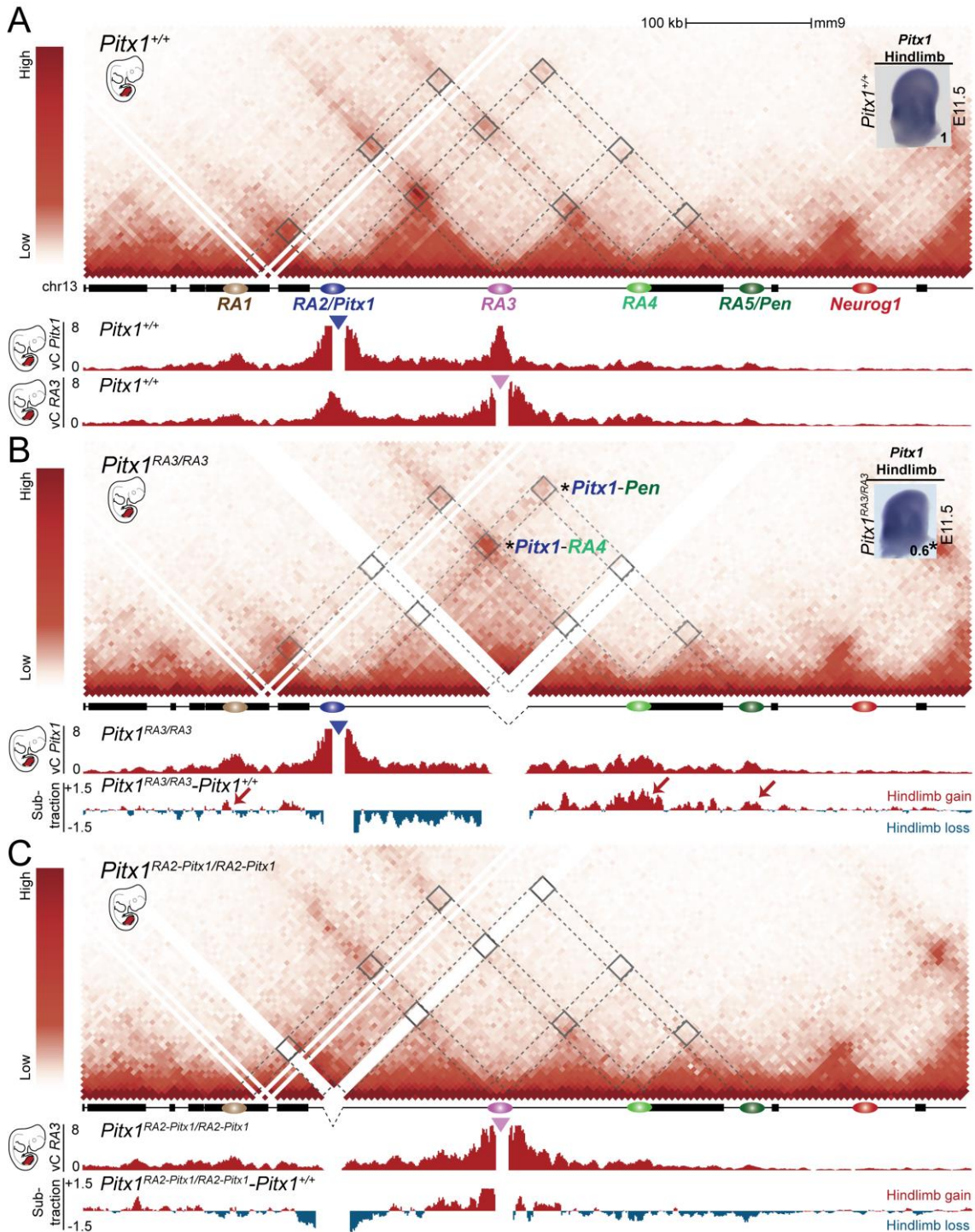
Figure 14 *Cis*-regulation of hindlimb chromatin folding of the *Pitx1* locus

A. cHi-C and vC using the *RA2/Pitx1* and *RA3* as viewpoints in wildtype hindlimb tissue at E11.5 are shown. **B.** cHi-C and derived vC *Pitx1* of *Pitx1^RA3/RA3* hindlimbs, with subtraction between wildtype and mutant hindlimbs at E11.5 in shown. Note the gain of chromatin interactions with *RA4* and *RA5/Pen* (red arrows and asterix). *Pitx1^RA3/RA3* hindlimbs show a 40% reduction in *Pitx1* expression as shown by WISH and qRT-PCR. **C.** cHi-C and derived vC *Pitx1* in *Pitx1^RA2-Pitx1/RA2-Pitx1* hindlimbs, with subtraction between wildtype and mutant hindlimbs, showing no significant of chromatin interactions with other RAs.

Results

Moreover, subtraction of the chromatin interactions between wildtype and mutant $Pitx1^{RA3/RA3}$ hindlimbs show a strong compensatory chromatin interaction between *RA2/Pitx1* with *RA4* and *RA5/pen* (Figure 14B and Figure 23A). Furthermore, the subtraction of cHi-C maps between $Pitx1^{RA3/RA3}$ mutant forelimbs and mutant hindlimbs demonstrates the preservation of a hindlimb-specific MAD organisation (Figure 23B). This proposes that the remaining regulatory anchors can rescue the loss of one anchor, an observation that fits with the loop-extrusion model whereby the removal of one "road-block" allows cohesin to extend the loop until stopped by the next convergent CTCF site (Fudenberg et al., 2016; Sanborn et al., 2015). Nevertheless, despite maintenance of the hindlimb MAD conformation, *Pitx1* expression is reduced by 40% (Figure 14B and Figure 23A). This loss of *Pitx1* expression might suggest that the regulatory anchors are important in conferring a spatiotemporal regulation of *Pitx1*. Alternatively, sequences within the 32.5 kb RA3 region might contain limb *cis*-regulatory elements.

RA2/Pitx1 is the central regulatory anchor of the MAD, as it interacts with all the other regulatory anchors, and thus might be the key brick in establishing the hindlimb architecture. To address this question, we engineered a 16 kb deletion removing the *Pitx1* gene body and its promoter region ($Pitx1^{RA2-Pitx1}$). cHi-C and derived vC in $Pitx1^{RA2-Pitx1/RA2-Pitx1}$ hindlimbs, show a loss of all *RA2/Pitx1*-mediated contacts (Figure 14C and Figure 23C). In contrast to $Pitx1^{RA3/RA3}$ hindlimbs, no compensatory chromatin interactions between the remaining regulatory anchors occur (Figure 14C and Figure 23C). In addition, comparison of the chromatin structure between $Pitx1^{RA2-Pitx1/RA2-Pitx1}$ mutant fore- and hindlimbs, does not reveal any differences (Figure 23D). In this case, the loop-extrusion model does not apply to the *Pitx1* locus. Taken together, these results demonstrate that sequences underlying *RA2/Pitx1* are central in initiating the hindlimb specific MAD organisation.

5.3.2 *Trans*-regulation of hindlimb chromatin folding of the *Pitx1* locus

To further identify the mechanism regulating *RA2/Pitx1* region, we probed the role of two putative *trans*-acting factors. *Pitx1* has been shown to be autoregulated in the pituitary gland (Goodyer et al., 2003), and to test whether PITX1 protein itself could regulate the structure of its own locus in hindlimbs, we engineered a PITX1 loss of

Results

function by inducing a frameshift mutation in exon 1 ($Pitx1^{fs}$). cHi-C maps of $Pitx1^{fs/fs}$ hindlimbs demonstrate a subtle decrease in the interaction between $RA2/Pitx1$ and $RA3$ (Figure 15A and Figure 24A). Subtraction of cHi-C maps between wildtype and mutant $Pitx1^{fs/fs}$ hindlimbs and derived $Pitx1$ vC, clearly display this decrease in interaction (Figure 15A and Figure 24A). This allele denotes a marginal contribution of PITX1 to the folding of its own locus in *trans*, and suggests the contribution of other upstream DNA binding factors.
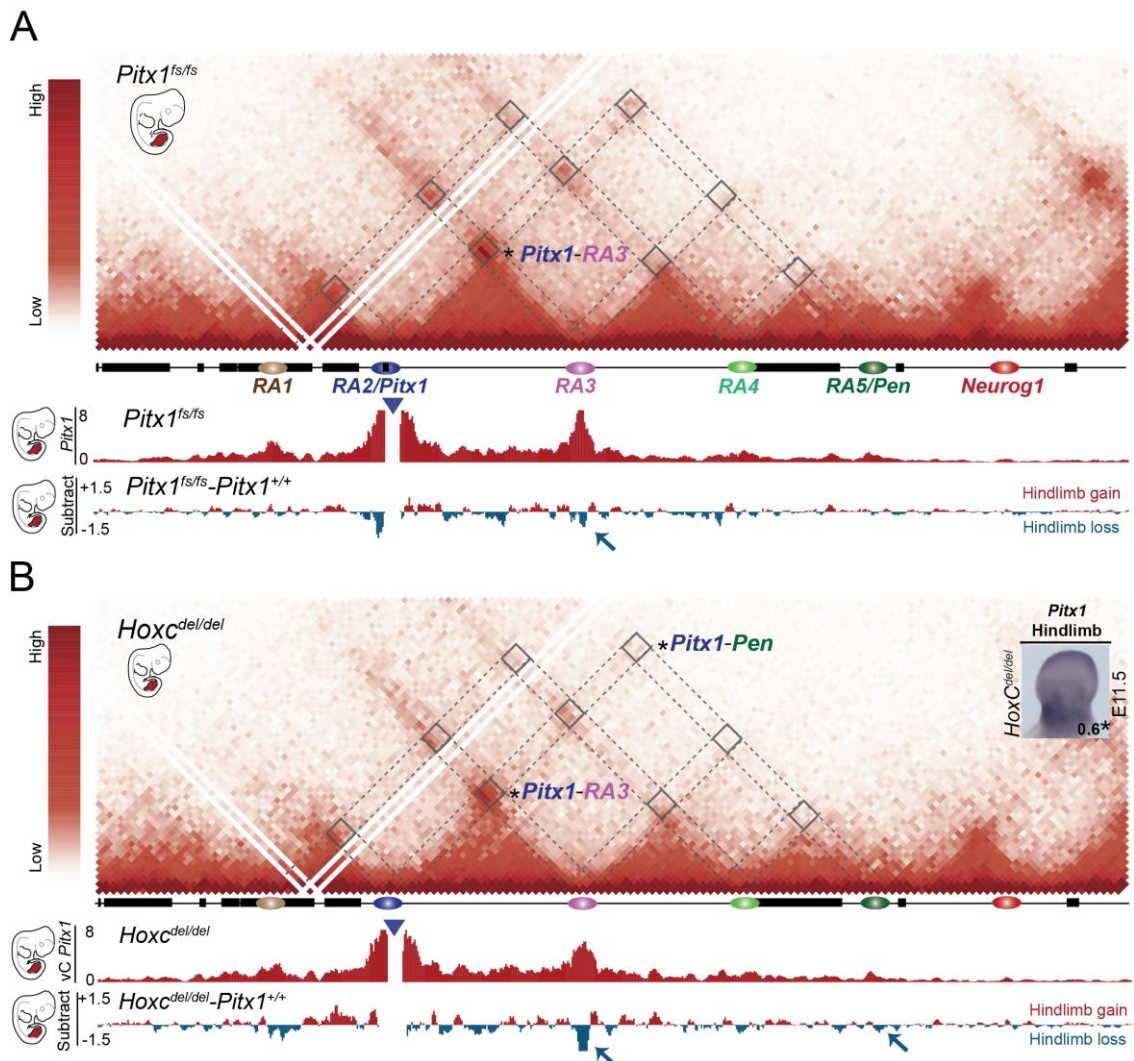
Figure 15 *Trans*-regulation of hindlimb chromatin folding of the *Pitx1* locus

A. cHi-C and vC *Pitx1* in $Pitx1^{fs/fs}$ hindlimbs and subtraction between wildtype and mutant hindlimbs at E11.5, show a significant loss chromatin interactions between *RA2/Pitx1* and *RA3* (blue arrow and asterisk). **B.** cHi-C and vC *Pitx1* in $Hoxc^{del/del}$ hindlimbs and subtraction between wildtype and mutant hindlimbs at E11.5, show a significant loss in chromatin interactions between *RA2/Pitx1* and *RA3*, and *RA2/Pitx1* and *RA5/Pen* (blue arrows and asterisk). $Hoxc^{del/del}$ hindlimbs show a 40% reduction in *Pitx1* expression, as detected with WISH and qRT-PCR.

Hoxc genes have been shown to be important transcriptional regulators of hindlimb differentiation (Logan and Tabin, 1999; Minguillon et al., 2012). To investigate the role

of *Hoxc* genes in mouse *Pitx1* regulation, a deletion encompassing the entire *HoxC* cluster (*HoxC^del*) was engineered. Strikingly, *HoxC^del/del* mice show a 40% reduction in *Pitx1* expression in hindlimbs (Figure 15B and Figure 24B). Concomitantly cHi-C and derived vC *Pitx1* interaction maps of mutant versus wildtype hindlimbs exhibit a significant loss of interaction between *RA2/Pitx1* and *RA3* as well as between *RA2/Pitx1* and *RA5/Pen* (Figure 15A and Figure 24B). Furthermore, ectopic expression of *HOXC9* in forelimb chick lateral place mesoderm has been shown to induce *Pitx1* expression in forelimb tissue (Nishimoto et al., 2014). To test whether *Hoxc9* can modulate the *Pitx1* locus in mouse forelimbs, it was ectopically expressed using the *Prx*-enhancer that is active in both forelimb and hindlimb buds (Martin et al., 1995). *Prx-Hoxc9* forelimb buds show a 5-fold increase in *Hoxc9* expression (Figure 24C). cHi-C and derived vC *Pitx1* of *Prx-Hoxc9* forelimbs show a marginal gain of chromatin interaction between *RA2/Pitx1* and *RA3* (Figure 16C and Figure 24C). However, this was not sufficient to induce ectopic *Pitx1* expression in forelimbs (Figure 16C and Figure 24C), perhaps due to the mild overexpression of *Hoxc9* in mouse forelimb tissue compared with chick. Thus, the HOXC9 protein has a negligible effect on the *Pitx1* chromatin architecture and transcription in forelimbs.

These results indicate that *RA2/Pitx1* is the key regulatory anchor, whereby HOXC transcription factors and its own protein product may bind to induce and maintain the hindlimb specific MAD. Nevertheless, *Pitx1* in hindlimb was not abolished by HOXC loss of function, suggesting multifactorial control to ensure robust tissue specific expression.

5.3.3 Switching *Pitx1* on in the forelimbs

In mESCs the *Pitx1* promoter is poised for activity and marked with H3K4me3 and H3K27me3 chromatin marks (Ferrai et al., 2017). As development proceeds, the promoter loses its H3K4me3 mark and stays inactivated in forelimbs buds (Andrey et al., 2017). To assess the possibility of altering the inactive state into an active one, the Polycomb repressive interaction between *RA2/Pitx1* and *Neurog1* was disrupted. A 45.5 kb region encompassing the entire H3K27me3 associated *Neurog1* region was deleted (*Neurog1^del*). However, cHi-C of *Neurog1^del/del* forelimbs shows no change in the chromatin interaction profile at the *Pitx1* locus (Figure 25A). Moreover, subtraction of derived *Pitx1* vC from wildtype and *Neurog1^del/del* forelimbs showed no compensatory

Results

cis-interactions with other nearby H3K27me3 enriched regions, and no upregulation of
Pitx1 transcription occurs (Figure 16B and Figure 25A). Thus, the loss of this repressive
Polycomb interaction is not sufficient to induce *Pitx1* transcription in forelimb tissue.
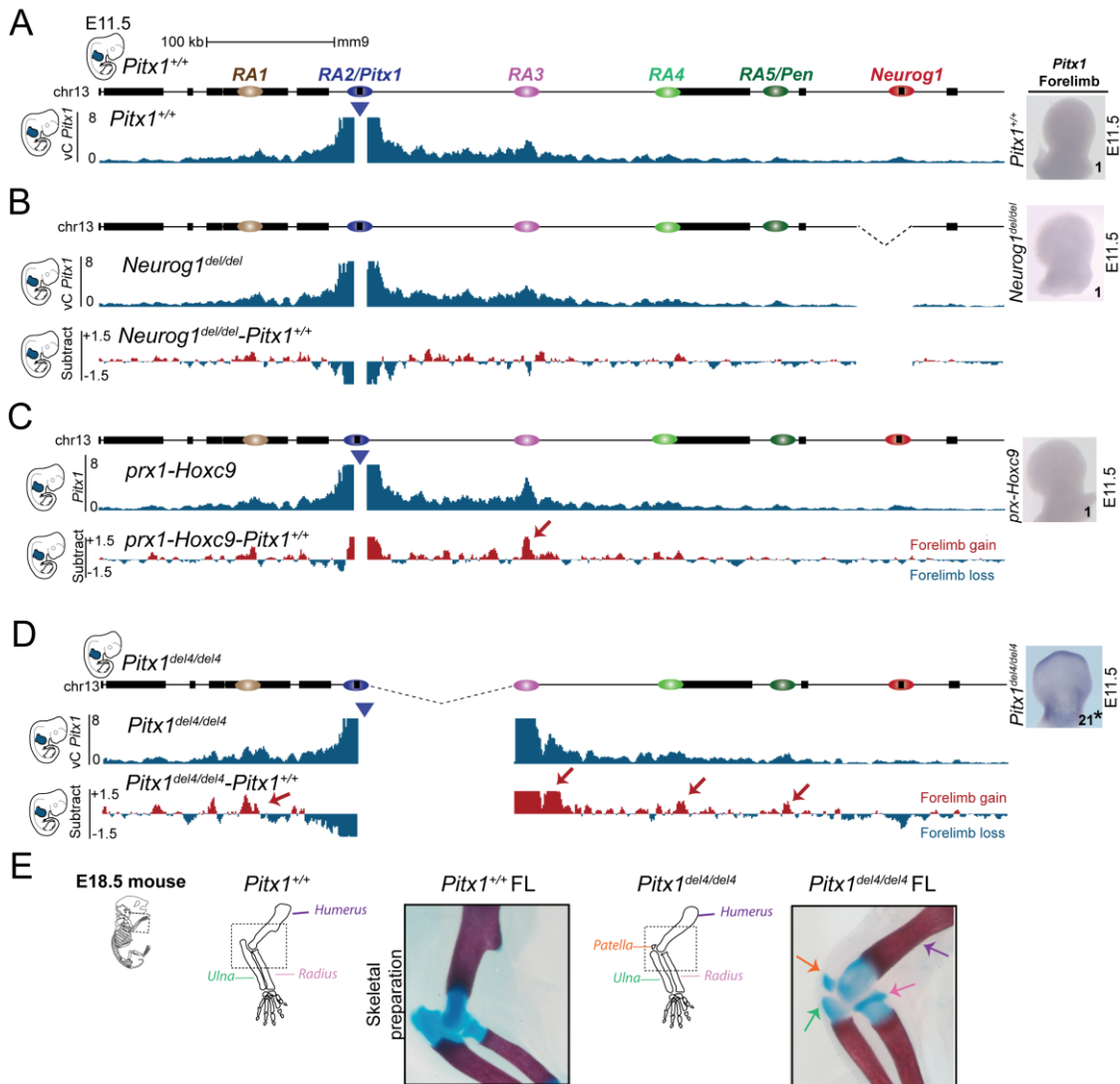
Figure 16 Perturbation of *Pitx1* regulation in forelimb tissue

A. vC *Pitx1* in wildtype E11.5 forelimb tissue and *Pitx1* mRNA analysis using WISH and qRT-PCR. **B.** vC *Pitx1* and subtraction between wildtype and *Neurog1^{del/del}* forelimbs at E11.5 do not show change in chromatin interactions. No change in *Pitx1* expression is detected, as shown by WISH and qRT-PCR. **C.** vC *Pitx1* and subtraction between wildtype and *prx-Hoxc9* forelimbs at E11.5 shows significant gain of chromatin interactions between *RA2/Pitx1* and *RA3* (red arrow). No changes in *Pitx1* expression are detected, as shown by WISH and qRT-PCR. **D.** vC *Pitx1* and subtraction between wildtype and *Pitx1^{del4/del4}* forelimbs at E11.5 show significant gain of chromatin interactions between *RA2/Pitx1* and *RA4, RA2/Pitx1* and *RA5/Pen,* and *RA2/Pitx1* and *RA1* (red arrow). 21-fold gain of ectopic *Pitx1* expression in forelimb tissue is detected by WISH and qRT-PCR. **E.** Alcian blue and alizarin red skeletal staining of E18.5 *Pitx1^{del4/del4}* forelimbs demonstrate a loss of the deltoid crest (purple arrow), broadening of the proximal radius (pink arrow), reduction of the olecranon (green arrow), and the appearance of an ectopic patella (orange arrow).

At the ß-globin locus, physical tethering between inactive embryonic genes and the
locus control region in adult erythrocytes induces inappropriate transcription of these

genes (Deng et al., 2012). An alternative way of activating *Pitx1* in forelimb tissue *in embryo*, is to induce the chromatin loop between *RA2/Pitx1* and *RA3*, thereby mimicking a main feature of the hindlimb chromatin architecture. To test this, we deleted the 122 kb region spanning between *RA2/Pitx1* and *RA3* (*Pitx1^{del4}*)*, without disrupting neither the promoter region nor *RA3*, thereby forcing proximity between the two regulatory anchors. Strikingly, cHi-C analysis of *Pitx1^{del4/del4}* demonstrates a transformation of the inactive forelimb MAD into an active hindlimb-like MAD. Specifically, prominent interactions between *RA2/Pitx1* and *RA1*, *RA4,* and RA5/*Pen*, are established, facilitating *Pitx1* interaction with its endogenous *pan*-limb region (Figure 16D and Figure 25B). This new forelimb chromatin configuration results in a 21-fold gain of *Pitx1* expression and a remarkable forelimb malformation, in which the olecranon is reduced, deltoid crest is lost, and some embryos even develop an ectopic patella, thereby resembling a knee-like joint (Figure 16E and Figure 25B).

Taken together, the inactive state of the *Pitx1* locus can be overturned in forelimb tissue by physically altering the chromatin architecture. Remarkably, the forced proximity between *RA2/Pitx1* and *RA3* does not only create increase the contact due to increased genetic proximity, but rather induces an active folding resembling the hindlimb MAD conformation and enables partial transformation of forelimb morphologies into hindlimb-like. Thus, the chromatin architecture plays an active role in regulating *Pitx1* expression.

5.3.3 Liebenberg syndrome is caused by *regulatory endoactivation* of *Pitx1* in forelimbs

Disruption of chromatin architecture by deleting or misplacing TAD boundaries has been demonstrated to cause pathogenic enhancer-promoter rewiring resulting in cancer and congenital malformations (Franke et al., 2016; Lupianez et al., 2015). The partial elbow-to-knee transformation in *Pitx1^{del4/del4}* animals shows remarkable similarities with the human Liebenberg syndrome, with a reduction of the olecranon and the appearance of an ectopic patella. Liebenberg patients carry deletions spanning the distal human *PITX1* regulatory landscape, thereby reducing the distance between *PITX1* and hs1473 enhancer element, which is equivalent to the *Pen* enhancer in mice. This has previously put forward the hypothesis of enhancer adoption as a cause of *PITX1* misexpression in forelimbs (Spielmann et al., 2012). To investigate the role of

Results

hs1473/*Pen* in driving the pathogenic misexpression of *Pitx1* in forelimb tissue, a series of deletion and inversions were engineered. An inversion placing *Pen* in the position of *RA4* (*Pitx1*$^{inv1/inv1}$), thereby increasing proximity between *Pen* and *Pitx1* by 112 kb, results in the activation of the *Pitx1* locus in forelimbs.
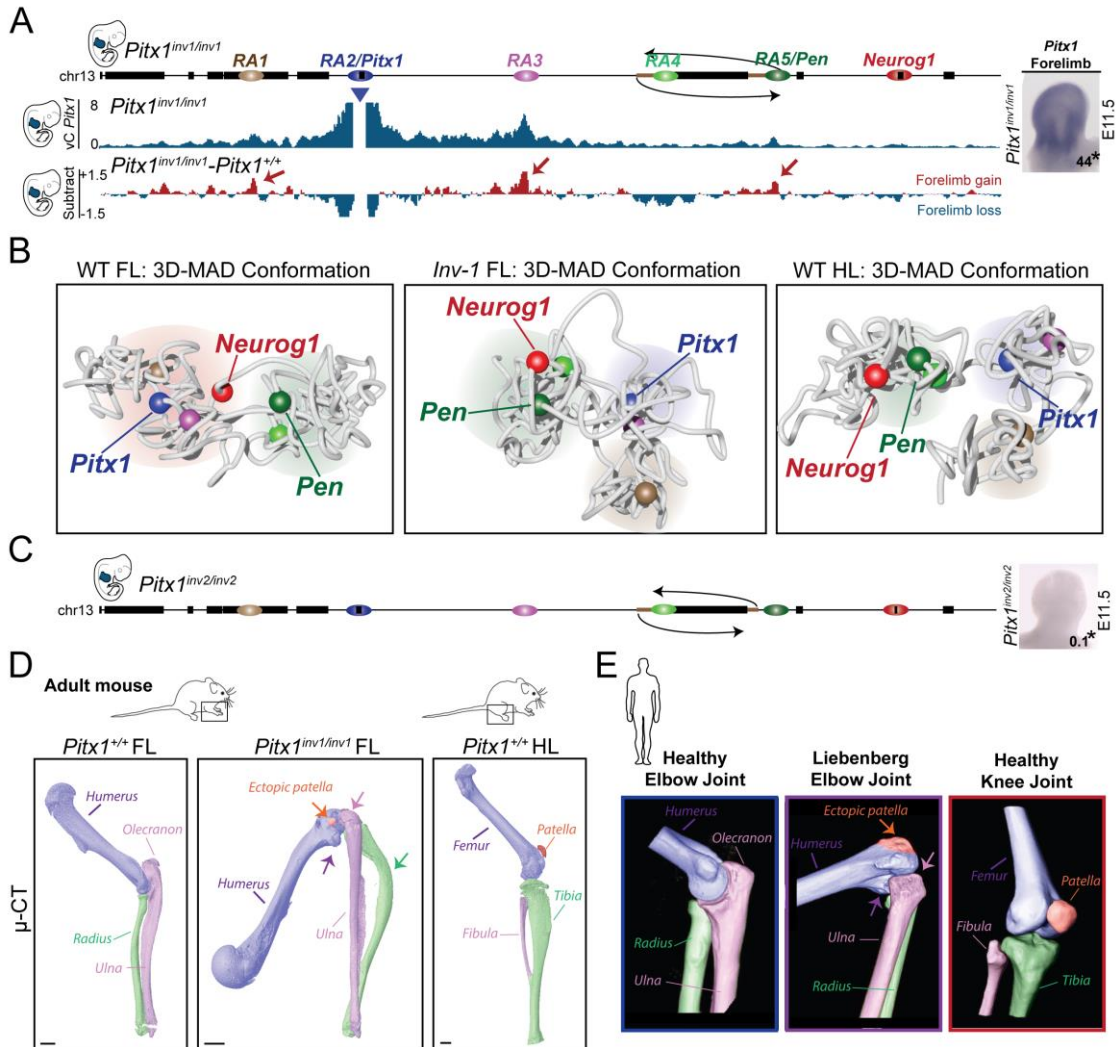
Figure 17 Structural variants result in activation of the *Pitx1* locus in forelimbs and limb malformation

A. vC *Pitx1* profile in E11.5 *Pitx1*$^{inv1/inv1}$ forelimb tissue and subtraction of wildtype and *Pitx1*$^{inv1/inv1}$ forelimbs, show gain of chromatin interactions between *RA2/Pitx1* with *RA1, RA3,* and *RA5/Pen,* indicated by red arrows. Right: WISH and qRT-PCR show a 44-fold upregulation of *Pitx1* expression in forelimb tissue. **B.** SBS polymer model of the *Pitx1* MAD conformation in *Pitx1*$^{inv1/inv1}$ forelimb tissue. Note how the mutant forelimb MAD (middle) loses the forelimb MAD organisation (left) and takes on a hindlimb like conformation (right). **C.** *Pitx1*$^{inv2/inv2}$ control allele inverting a 99 kb region containing *RA4* and *H2afy* and excluding *Pen*, show no ectopic *Pitx1* expression in forelimb tissue, as detected by WISH and qRT-PCR (right). **D.** Micro-CT of wildtype and mutant *Pitx1*$^{inv1/inv1}$ adult forelimbs and hindlimbs display partial arm-to-leg transformation. From left: a healthy forelimb, middle: mutant forelimbs showing distorted rotation of the elbow joint (purple arrow), a reduced olecranon of the ulna (pink arrow), deformed radius (green arrow), and the presence of an ectopic patella (orange arrow). **E.** 3D-CT scan of healthy human elbow joint (left), Liebenberg elbow joint (middle) exhibit a reduced olecranon (pink arrow), the presence of an ectopic patella fused to the humerus (orange arrow), and malformed distal radius (green arrow), thereby resembling a knee joint (right).

Results

Specifically, cHi-C heatmap and derived vC track of $Pitx1^{inv1/inv1}$ forelimbs display several hallmarks characteristic of a hindlimb MAD organisation with increased compaction and insulation of the domains. In addition, the subtraction of vC tracks demonstrate this strong gain in interaction between $Pitx1/RA2$ with $RA1$, $RA3$ and $RA5/Pen$ is induced. Moreover, the repressive interaction with $Neurog1$ is diminished (Figure 17A). SBS polymer modelling of the mutant forelimb cHi-C data demonstrates the 3D-folding of the locus in three chromatin hubs is strikingly similar to that of wildtype hindlimb (Figure 17B). This new chromatin architecture induces a 44-fold increase in $Pitx1$ expression (Figure 17A and Figure 25C). As a control, a 99 kb fragment including $RA4$ and $H2afy$, but excluding Pen, was inverted ($Pitx1^{inv2}$). $Pitx1^{inv2/inv2}$ do not show any ectopic expression of $Pitx1$ in forelimbs tissue (Figure 17C and Figure 25D), verifying that Pen causes the ectopic $Pitx1$ expression in $Pitx1^{inv1/inv1}$ forelimbs. The forelimbs of adult $Pitx1^{inv1/inv1}$ mice display a remarkable phenotype, in which the flexion and rotation of the forelimb resemble that of the hindlimb. Moreover, micro-CT of the forelimbs demonstrates a severe reduction of the olecranon, loss of the deltoid crest, bowing of the radius, and the presence of an ectopic patella (Figure 17D). This phenotype resembles the Liebenberg syndrome characteristics. Liebenberg individuals have an altered rotation and flexion of their arms. Moreover, 3D-CT scan of their elbow joint show a severely reduced olecranon, and an ectopic patella fused to the distal head of the humerus, whereby the joint articulation is transformed into a knee-like joint (Figure 18A) (Spielmann et al., 2012).

As a follow up, a human Liebenberg deletion, spanning 107 kb (Figure 17F), was re-engineered in mice, corresponding to a 99 kb region including both $RA4$ and $H2afy$ ($Pitx1^{delL1}$). cHi-C analysis and derived vC of $Pitx1^{delL1/delL1}$ forelimb buds show a gain of interaction between $RA2/Pitx1$ and $RA5/Pen$ (Figure 18B and Figure 26A). However, despite a 14-fold increase in $Pitx1$ expression in $Pitx1^{delL/+}$ forelimb buds, no skeletal phenotype was observed (Figure 17B and Figure 26A). Finally, to rescue the $Pitx1$ misexpression in forelimb tissue, a control allele, in which the Liebenberg deletion is extended by 13 kb to encompass Pen ($Pitx1^{del5}$), was engineered. $Pitx1^{del5/del5}$ E11.5 embryos do not display any ectopic $Pitx1$ expression in forelimb tissue (Figure 18C and Figure 26B), proving that Pen causes the pathogenic misexpression of $Pitx1$ in $Pitx1^{delL1/delL1}$ forelimb tissue.

Results
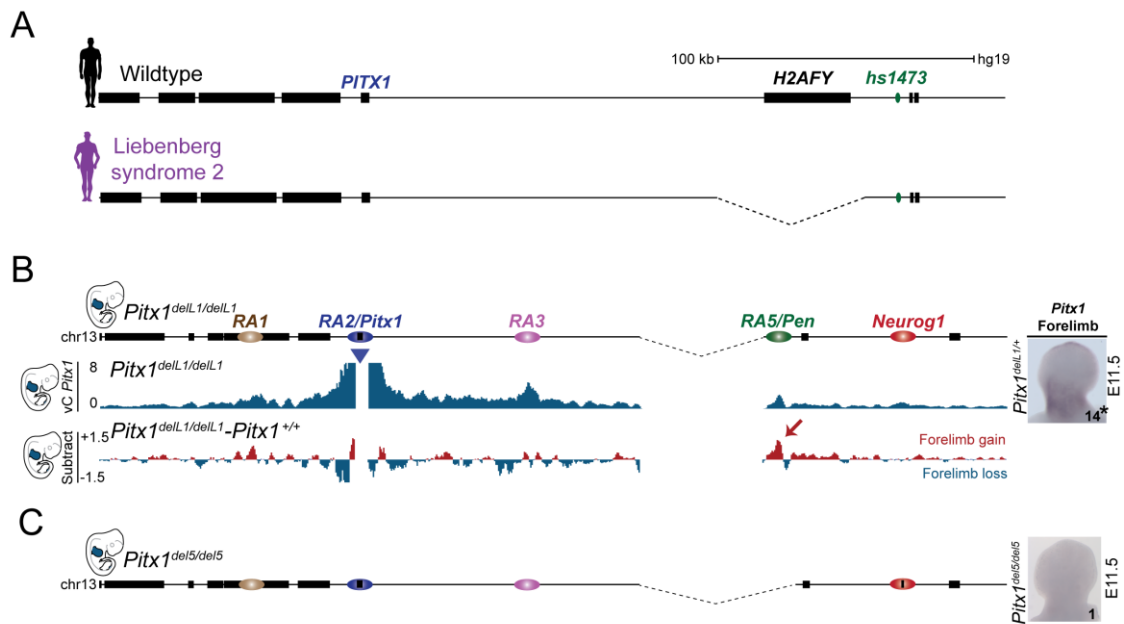
Figure 18 Liebenberg syndrome is caused by *Pitx1 regulatory endoactivation* in forelimbs

A. Liebenberg syndrome is caused by deletions in the *PITX1* regulatory landscape bringing *hs1473* element closer to *PITX1*. **B.** vC *Pitx1* in *Pitx1$^{delL1/delL1}$*, re-engineered Liebenberg deletion, E11.5 forelimbs show gain of interactions with *RA5/Pen*. Pitx1 expression is 14-fold increased as shown by WISH and qRT-PCR. **C.** Rescue allele, which extends the Liebenberg deletion by 13 kb to include *Pen,* eliminates ectopic *Pitx1* expression in forelimb tissue. WISH and qRT-PCR show absence of *Pitx1* expression in mutant forelimb tissue.

Together, the Liebenberg and control alleles demonstrate that structural variants at the *PITX1/Pitx1* loci perturbate the chromatin architecture in forelimb tissue, allowing the strong *pan*-limb enhancer hs1473*/Pen* to erroneously activate *PITX1/Pitx1*, in both humans and mice. We coined the term "*regulatory endoactivation*" to describe this novel pathomechanism in which a gene is ectopically activated by its endogenous enhancers, resulting in misexpression and disease.

5.3.4 *H2afy* promoter traps *Pen* enhancer activity in the forelimb

In the alleles described above, the perturbation of *H2afy* or placing *Pen* 3' of this housekeeping gene, seem to have a defining role for whether *Pen* can activate *Pitx1* in forelimbs. Interestingly, by whole genome sequencing, we identified a small 8.5 kb deletion removing the *H2AFY* promoter in an individual showing a mild form Liebenberg syndrome (Figure 19A and Figure 6B). Embryos with a re-engineered deletion (*Pitx1$^{delL2/delL2}$*), removing an equivalent 10 kb region of the *H2afy* promoter region, show a mild ectopic *Pitx1* expression in the proximal forelimb buds, quantified as 10-fold upregulation (Figure 19B and Figure 25C). Compared to *Pitx1$^{delL1/+}$* forelimbs, the

misexpression in $Pitx1^{delL2/delL2}$ was milder, indicating that the relative reduction in distance between *Pitx1* and *Pen* may determine the severity of misexpression. To investigate this, the entire *H2afy* gene body (58 kb), excluding the promoter region ($Pitx1^{del3/del3}$), was deleted. Strikingly, no ectopic expression of *Pitx1* was detected in $Pitx1^{del3/del3}$ forelimbs (Figure 19B and Figure 26D).
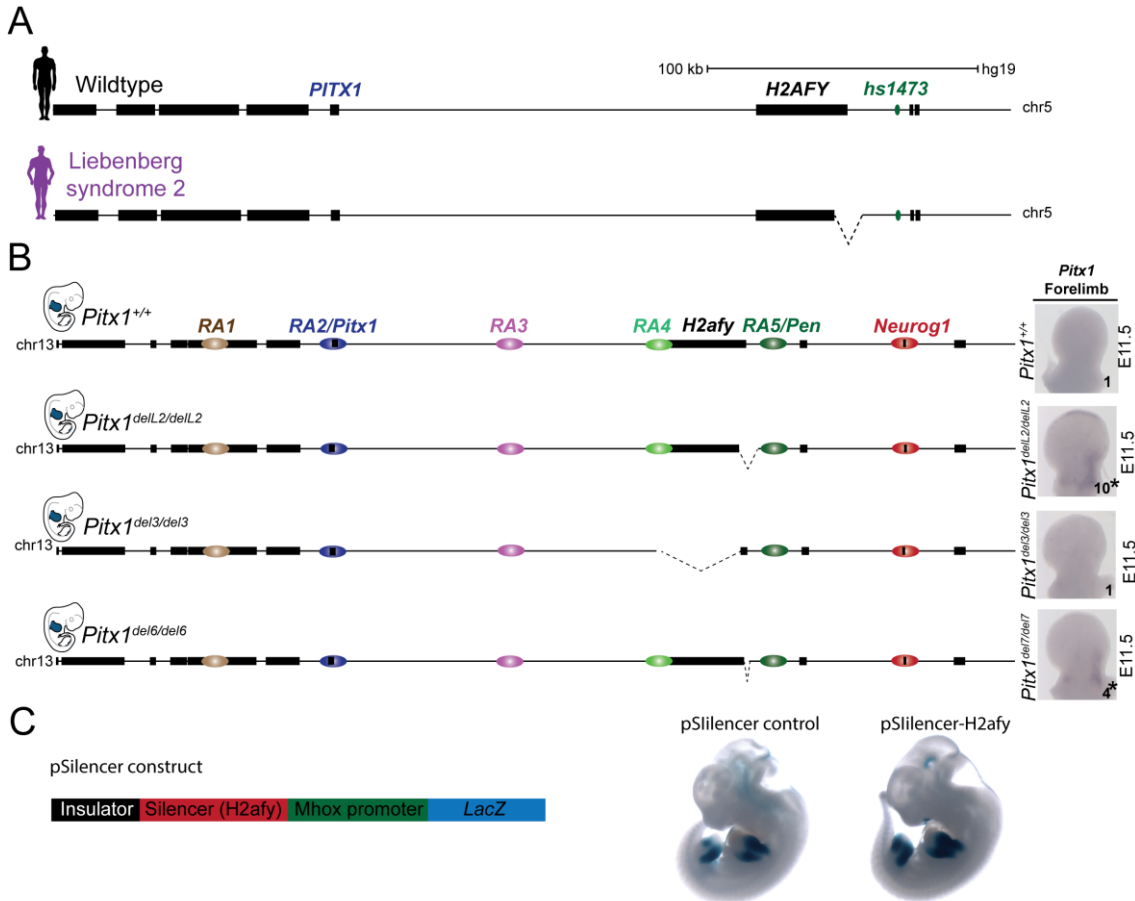
Figure 19 *H2afy* promoter acts as an enhancer trap in forelimb tissue

A. Smallest deletion known to cause Liebenberg syndrome in humans is 8.5 kb (Liebenberg syndrome 2). **B.** $Pitx1^{delL2/delL2}$ re-engineered Liebenberg syndrome 2 allele results in a 10-fold misexpression of *Pitx1* in forelimb tissue. $Pitx1^{del3/del3}$ E11.5 forelimbs do not show ectopic expression of *Pitx1* in forelimb tissue. $Pitx1^{del6/del6}$ mice carry a 400 bp deletion of the *H2afy* promoter region and show a 4-fold upregulation of *Pitx1* in forelimb tissue. **C.** pSilencer-H2afy construct testing if *H2afy* promoter can act as a silencer by silencing *Mhox* in limb buds. No reduction in *Mhox* induced *LacZ* activity is detected in pSilencer-H2afy mice (right) compared to control pSilencer mice (left).

This implies that the *H2afy* promoter region insulates *Pen* activity specifically in forelimb tissue. To narrow down this insulating effect, a 400 bp deletion encompassing a part of the *H2afy* promoter region was engineered ($Pitx1^{del6}$). At E11.5 $Pitx1^{del6/del6}$ mice displayed a faint proximal forelimb staining, with a 4-fold upregulation of *Pitx1* (Figure 19B and Figure 26E). This smaller deletion has a less severe misexpression of *Pitx1* in the forelimb tissue, indicating that the reminiscent *H2afy* region insulates most of the

Results

Pen activity. To address whether *H2afy* might be regulated by *Pen* enhancer, its expression in *Pitx1^{Pen/Pen}* limbs was assessed. *H2afy* is expressed at highly similar levels in wildtype forelimbs and hindlimbs, however in *Pitx1^{Pen/Pen}* forelimbs show a detectable reduction in *H2afy* expression than hindlimbs (Figure 26F). To test whether the *H2afy* promoter can function as a silencer, that is being able to block a promoter or enhancer element, 1 kb of its promoter region was tested in a silencer construct (pSilencer-H2afy). The construct contains a *Mhox* promoter which shows broad expression in the proximal limb, and thus if *H2afy* can silence the promoter its expression will be reduced. The construct was injected into zygotes and embryos with random integrations were analysed for *LacZ* activity at E11.5. Out of six transgenic embryos none show any repression of *LacZ* expression in the limbs (Figure 19C). Thus, in this experiment H2afy promoter did not exert a silencer activity. However, *H2afy* may have a specific effect on *Pen* enhancer, and this effect may be confined to the *Pitx1* locus specific 3D architecture. Alternatively, H2afy functions as an insulator within the genomic context by titrating the activity of *Pen*.

These data validate that Liebenberg syndrome is caused by the loss of the forelimb specific insulating effect of the *H2afy* promoter. Yet, the larger a portion is removed, the stronger the *Pitx1* misexpression. Strikingly, removing 58 kb of *H2afy* gene body leaving the promoter intact, did not lead to misexpression of *Pitx1* in forelimbs. Thus, the reduced distance between *Pen* and *Pitx1* does not induce the misexpression per say; rather the alteration in the 3D-architecture and loss of *H2afy* expression result in *Pitx1* misexpression and consequent forelimb malformation. Thus, not all sequences are equal at the *Pitx1* locus and the *H2afy* promoter seems to be there to promote the insulation between *Pitx1* and *Pen* in forelimbs.

6. DISCUSSION

6.1 *PITX1* REGULATORY LANDSCAPE IN EMBRYONIC HINDLIMBS

Normal embryonic development requires the precise temporospatial orchestration of gene regulation. Sequence specific enhancer elements embedded in regulatory landscapes instruct these tissue specific transcriptional activities to cognate promoters through long-range chromatin interactions (Deng et al., 2012; Montavon and Duboule, 2012). *Pitx1* is a known master regulator of hindlimb outgrowth and patterning, but its regulation in mammals is unknown. One of the main aims of this thesis was to identify the *Pitx1* regulatory landscape during embryonic hindlimb development. Here, the regulatory activity at the *Pitx1* locus was tested exogenously, using transgenic *LacZ* assays, and endogenously by integrating sensors into the locus. Moreover, CRISPR-Cas9 engineering was used to functionally dissect the *Pitx1* hindlimb regulatory landscape in mouse embryos.

6.1.1 *Pitx1* regulatory landscape contains a *pan*-limb region

The *Pitx1* regulatory landscape, as defined by the extent of chromatin interactions, spans over 420 kbs in hindlimb tissue. In particular, *Pitx1* establishes prominent chromatin loops with H3K27ac enriched regions, forming the regulatory anchors *RA1-RA5/Pen*. Surprisingly, transgenic reporter testing of the *RA4* and *RA5/Pen* regions identified fore- and hindlimb enhancer activity (*pan*-limb). This spurious activity was confirmed by the targeted integration of regulatory sensors into the endogenous locus. These findings deviate from the characterisation of regulatory landscape of genes such as *Shh* and *Ihh*, whereby transgenic assays and targeted integration of sensors solely reflect the endogenous expression of the target gene (Symmons et al., 2014; Will et al., 2017). The presence of unspecific limb regulatory activity at *Pitx1* locus, suggests the existence of additional layers of transcriptional regulation that can modulate the *RA4* and *RA5/Pen* enhancer activity.

6.1.2 *Pitx1* is regulated by *pan*-limb enhancer, *Pen*

Loss of PITX1 protein in hindlimbs affects both its outgrowth and patterning, evident by the absence of hindlimb-specific skeletal elements such as the patella (Lanctot et al., 1999; Szeto et al., 1999). Similarly, removal of the telomeric *Pitx1* regulatory landscape results in a hindlimb loss of function phenotype. This allele indicates that the enhancer elements regulating *Pitx1* in hindlimbs are contained within the 330kb telomeric region. Remarkably,

deleting the *Pen* enhancer element alone (*Pitx*[Pen]) leads to a partial *Pitx1* loss of function, resulting in a reduced penetrant clubfeet phenotype in adult mice. Similarly *Pitx1*[fs/Pen] loss-of-function mice develop clubfeet and remarkable skeletal malformation with a fragmented patella and altered rotation. Moreover, this phenotype is comparable to *Pitx1* haploinsufficiency in both mice and humans (Alvarado et al., 2011). Together, these alleles support the finding that *Pen* is needed for robust *Pitx1* expression in hindlimbs.

The finding raises the question why a key hindlimb patterning gene is regulated by a strong *pan*-limb enhancer, since detrimental effects may result from misregulation. Interestingly, the presence of enhancer elements with broader activities than that of the target gene have been described both at the *Tbx4* and *HoxD* locus (Menke et al., 2008; Montavon et al., 2011). The hindlimb-specific *Tbx4* gene is regulated by the HLEA enhancer element with fore- and hindlimb activity in transgenic assays (Menke et al., 2008). The transgenic mice were generated by pronuclear injection, which yields multiple insertions at different positions in the genome. This might lead to instability of the reporter gene and unspecific activity compared to an *in vivo* enhancer activity at the endogenous locus. However, our present results suggest that this forelimb HLEA activity might occur at its endogenous location and that it could be modulated by similar 3D chromatin dynamics. Similarly, the digit regulatory islands controlling the expression of posterior *Hoxd* genes in limbs do not strictly bear the same activity pattern as their target gene *Hoxd13* (Montavon et al., 2011). By using chromosome conformation capture analysis, tissue specific interactions between unspecific enhancers with target promoters was identified, thereby suggesting a similar modulation of enhancers activities (Andrey et al., 2013; Montavon et al., 2011). Indeed, in the autopod where *Hoxd13* is robustly expressed, all digit enhancers contact its promoter, whereas in the forearm, where *Hoxd13* is not expressed, only a few enhancers-promoter contacts are established, in a constitutive manner (Andrey et al., 2013; Montavon et al., 2011). This raises the possibility that the broad activity of unspecific enhancers contributes to gene expression robustness by interacting with their promoters in a tissue-specific fashion.

6.2 THE 3D-ORGANISATION OF *PITX1* REGULATORY LANDSCAPE

With the growing identification of long-range gene regulation, chromosome conformation capture techniques such as Hi-C, have demonstrated the organisation of these landscapes into topologically associating domains (TADs), separated by boundary elements, which creates a framework that mediates cognate enhancer-promoter pairs (Dixon et al., 2012;

Discussion

Lupianez et al., 2015; Symmons et al., 2016). However, little is known about the nature of TADs and how tissue specific enhancer-promoter interactions are established. The presence of a *pan*-limb enhancer, *Pen*, regulating *Pitx1* in the hindlimb, proposes its modulation through 3D-chromatin architecture. Here, the epigenetic landscape was compared between mouse embryonic fore- and hindlimbs to detect putative changes in functional chromatin. Moreover, the chromatin architecture of the *Pitx1* locus was investigated in fore- and hindlimbs using Capture Hi-C from which we derived 3D polymer models.

6.2.1 A dynamic chromatin conformation modulates *Pitx1* regulation in fore- and hindlimbs

Recent high resolution chromatin conformation data across various tissues and developmental time points have demonstrated dynamic changes in 3D-architecture that correlate with changes in functional chromatin modifications, as well as architectural proteins CTCF, cohesin, and mediator (Andrey et al., 2017; Freire-Pritchett et al., 2017; Phillips-Cremins et al., 2013; Rao et al., 2014). Consistent with these findings, a forelimb specific interaction between *RA2/Pitx1* and the unrelated developmental gene *Neurog1* is associated with H3K27me3 deposition on both inactive promoters. Accordingly, genome wide Polycomb mediated long-range interactions have been shown to occur between transcriptionally silenced developmental promoters (Denholtz et al., 2013; Schoenfelder et al., 2015). Polycomb mediated clustering of these transcriptionally silenced genes may sustain the continuous repression and thereby shape the inactive 3D-genome. However, removal of the H3K27me3 *Neurog1* region did not alter chromatin architecture at the *Pitx1* locus nor resulted in ectopic expression. This is consistent with the observation in PRC1 knock out mESCs, whereby the eradication H3K27me3 results in loss of long-range Polycomb mediated contacts at the *HoxD* locus, but does not cause major transcriptional changes (Schoenfelder et al., 2015). Thus, Polycomb-mediated interactions are likely a by-product of repression and dispensable for maintaining the inactive state of transcriptionally inactive genes.

On the other hand, H3K27ac enriched interactions between *RA2/Pitx1* with *RA1-RA5/Pen* are associated with an active conformation of the locus. Correspondingly, neuroectodermal promoters transitioning from inactive to transcriptional activation during the differentiation from embryonic stems cells into neuroectodermal cells, gain interactions with active enhancers in these cells (Freire-Pritchett et al., 2017). However, whether a genuine functional relation between chromatin interaction and H3K27ac marks exists remains to be

experimentally tested. One can speculate that histone acetylases depositing H3K27ac marks at enhancers and promoters, or H3K27ac chromatin readers, mediate chromatin looping between H3K27ac-enriched region, together with co-activators such as cohesin or Mediator complex. Alternatively, an affinity between homotypic H3K27ac regions could lead to formation of chromatin loops through phase separation (Hnisz et al., 2017). A recent report indeed supports this hypothesis. At the ß-globin locus, H3K27ac enriched chromatin interactions between active genes and locus control region (LCR) present in foetal erythrocytes form a transcriptionally active hub, but these are lost in adult cells and accompanied with physical separation and gene inactivation (Huang et al., 2017; Kooren et al., 2007).

6.2.2 *Pitx1* locus is organised into a Multi-Anchor-Domain (MAD)

In this study, high resolution Capture Hi-C demonstrates an alternate type of architecture organising the *Pitx1* locus, in which no clear TAD organisation defining the outer boundaries of the regulatory landscape is discernible. Opposing the conventional concept of a TAD organisation of the genome, the *Pitx1* locus is organized as a multi-anchor domain (MAD), dividing the locus into several distinct subdomains that interact with each other via their H3K27ac enriched regulatory anchors in the hindlimb. In contrast, in the forelimb, an alternative MAD configuration results in the loss of promoter-enhancer interactions and a gain in Polycomb mediated interactions. In addition, 3D polymer models of the locus, derived from cHi-C data, highlighted two fundamentally different configurations: one in which *Pitx1* is disconnected from its main regulators by physically separating them from their target gene (forelimb), silencing it together with Polycomb repressed *Neurog1,* and one in which *Pitx1* is embedded in a transcriptionally active pocket directly facing active *pan*-limb enhancer *Pen* (hindlimb). Interestingly, in human Hi-C data, a TAD is identified at the *PITX1* locus, with a weak boundary region whose location alters between *H2AFY* and *TIFIB* (Dixon et al., 2012). In mice, in contrast, no boundary regions are detected in the vicinity (Dixon et al., 2012). These Hi-C data are obtained from cells and tissues where *PITX1/Pitx1* is not expressed, and the resolution of data is low (20-50 kb). Furthermore, TADs and their boundaries are detected with computational algorithms that identify the transition in directionality of chromatin interactions, i.e. TADs and TAD boundaries can be called genome wide (Dixon et al., 2012; Schmitt et al., 2016). This is an arbitrary way of categorising the whole genome into regions

Discussion

without further in-depth analysis, and might suggest that the identification of a TAD at the *PITX1* locus in human is a correlation. Here, the high-resolution cHi-C reveal the organisation of the *Pitx1* locus in two alternating MAD structures in developing forelimb and hindlimbs tissue, where *Pitx1* is either transcriptionally inactive or robustly expressed, respectively. Moreover, unlike TADs, the MAD tissue dynamics actively pairs cognate *Pen* and *Pitx1* in hindlimb tissue, and separates them in forelimb tissue. In conclusion and in contrast to TADs, tissue-specific interactions, presence of small chromatin domains separated by regulatory anchors, interaction between the regulatory anchors as well as the absence of large chromatin domains encompassing the gene regulatory landscape are a key feature defining MADs.

6.2.3 Regulatory information is spatially segregated in MADs

The spatial organisation at the *Fgf8*, *Ihh,* and *Shh* loci, measured by integration of regulatory sensors reveal overlapping regulatory activities within the TAD, together reflecting the temporospatial expression pattern of local promoters (Marinic et al., 2013; Symmons et al., 2014; Will et al., 2017). Closer inspections of the activity of integrated sensors at the Pitx1 locus demonstrate a similar underlying segregation of the regulatory information. However, unlike that of the abovementioned loci, the regulatory information at the *Pitx1* is unspecific. Sensors located in the *Pitx1*-proximal telomeric region (between *RA2/Pitx1* and *RA3*) reproduce an endogenous *Pitx1* expression; insertions in the intermediate telomeric region (between *RA3* and *RA4*), show weaker *Pitx1* expression and a gain of proximal forelimb activity; insertions in the *Pitx1*-distal telomeric region (between *RA4* and *RA5*), solely captures the unspecific *pan*-limb activity of *Pen*. This demonstrates a partition of the tissue specific regulatory potential within the *Pitx1* regulatory landscape into subdomains demarcated by the regulatory anchors. Interestingly, the increased distance relative to the *Pitx1* promoter, correlates with increase in unspecific regulatory activity. The segmentation of the unspecific regulatory information might thus provide a tight control of the spatiotemporal regulation of *Pitx1* through the dynamic folding of the MAD.

6.3 Perturbation of MAD organisation results in *Pitx1* misregulation and limb malformation

Deprivation of CTCF or cohesin in cells and tissues leads to the disappearance of chromatin loops and TAD structure (Nora et al., 2017; Rao et al., 2017). While megabase large A/B compartments remain stable in these mutants, the loss of TADs reveal the existence of fine-scale compartments, with an average 150 kb in size that reflect the epigenetic landscape (Lieberman-Aiden et al., 2009; Nora et al., 2017; Rao et al., 2017). Thereby, epigenetic landscape seems to confer loop-extrusion independent attractive forces between homotypic regions both at the local scale and global scale (Schwarzer, 2017). At the various *Pitx1* regulatory anchors, CTCF and cohesin do not display differential binding in fore- and hindlimb buds. However, the CTCF and cohesin bound region seem to curb the active H3K27ac mark to contain it within the regulatory anchor region. Together, the data suggest a CTCF- and cohesin-independent mechanism conferring the hindlimb specific looping between *RA2/Pitx1* and its regulatory anchors. This active modulation of the MAD organisation of the *Pitx1* locus was further used as a test bed to dissect the underlying sequences and factors directing the MAD conformation to either activate *Pitx1* in hindlimb tissue or to inactivate it in forelimb tissue.

6.3.1 *RA2/Pitx1* controls tissue specific MAD conformation

CTCF and cohesin have been suggested to be key component of the loop extrusion model. This model proposes that the cohesin ring extrudes the DNA to form loops (Fudenberg et al., 2016; Sanborn et al., 2015). When the cohesin ring encounters convergent CTCF bound regions, the DNA extrusion is stalled and a stable loop is formed (Fudenberg et al., 2016; Sanborn et al., 2015). Importantly, the model seems to explain the mechanism of TAD formation (Fudenberg et al., 2016; Sanborn et al., 2015). Consistent with this model, the removal of *RA3* (*Pitx1^{RA3}*), bound by CTCF and cohesin results in loss of insulation between the two neighbouring chromatin domains, fusing them into a larger domain and resulting in compensatory interactions between *RA2/Pitx1* and *RA4*, both bound by CTCF and cohesin. In this mutant, the overall structure of the MAD maintains intact in hindlimb tissue. In contrast, the loop extrusion model does not hold true when the strongest anchor of the MAD, *RA2/Pitx1* itself (*Pitx1$^{RA2-Pitx1}$*), is removed. Here, no compensatory interactions between the remaining regulatory anchors is established and the hindlimb specific MAD is abolished. This

Discussion

finding further support a conceptual difference between MAD and TAD organisation. This insight into the MAD organisation thus suggests that the gene itself conducts the formation of the architecture with the help of *trans*-acting factors.

6.3.2 Hindlimb conformation is partly regulated by *Pitx1* and *Hoxc* genes

Autoregulation of developmental genes has been suggested to refine their tissue specific expression (Crews and Pearson, 2009). In line with this idea, *Pitx1* has been shown to be autoregulated in the pituitary gland (Goodyer et al., 2003). Here, we observed a mild reduction of the chromatin interaction between *RA2/Pitx1* and *RA3* in hindlimbs of *Pitx1* homozygous frame shift mutants. This result supports a role of *Pitx1* autoregulation in either the initiation or maintenance of this chromatin interaction. In this scenario, low levels of PITX1 protein might bind to regulatory anchors to instruct looping in concert with additional DNA binding factors that bring distal *pan*-limb enhancers into proximity to form an active MAD conformation. These factors have been suggested to be posterior *Hoxc* genes, whereby ectopic expression of *Hoxc9* can induce *Pitx1* in chick forelimbs (Nishimoto et al., 2014; Wellik and Capecchi, 2003). Remarkably, here the removal of the *HoxC* cluster (*HoxC^{del}*) resulted in a strong reduction in chromatin interaction both between *RA2/Pitx1* and *RA3,* as well as between *RA2/Pitx1* and *RA5/Pen* in hindlimbs. This was accommodated by a partial loss of *Pitx1* expression in hindlimbs, supporting a role for HOXC proteins upstream of *Pitx1* regulation. Moreover, human patients carrying deletions of 5' *HOXC* genes develop clubfeet, resembling the phenotype *PITX1* haploinsufficiency causes in both humans and mice (Alvarado et al., 2011). HOXC factors and PITX1 protein together with co-factors thus could cooperate to instruct the formation of chromatin loops at the locus. This observation hold true for other loci tested *in vitro*. The activation of the *β-globin* genes depends on a regulated erythrocyte-specific chromatin loop with the LCR, which also display enhancer activities in a precursor cell type, where the loop is not formed (Deng et al., 2012). *Trans*-acting regulatory factors KLF1, GATA1, and LDB1 bind the LCR and target genes in adult erythrocytes to instruct the loop formation, and removal of Gata1 blocks globin transcription (Deng et al., 2012; Deng et al., 2014). From these experiments one can speculate that the structure precedes the gene transcriptional activation. In support of this speculation, *Prx-Hox9* forelimbs show an alteration of the inactive forelimb MAD whereby the proximity between *RA2/Pitx1* and *RA3* is increased. However, this is not sufficient to induce *Pitx1* expression.

Discussion

6.3.3 Activation of *Pitx1* in forelimbs results in limb malformation

The inactivation of the active MAD structure in hindlimb tissue is challenging, since the genomic dissection perturbs the normal arrangement of the locus and cannot discern which factors may be at play. An alternative way of investigating the formation of a transcriptionally active MAD is by its activation in forelimbs. Using cell culture experiments, Deng and colleagues demonstrated that physical tethering of the LCR with developmentally silenced adult ß-globin genes in foetal erythrocytes ectopically activates the genes (Deng et al., 2014). Such experiments are challenging to perform *in embryo,* thus, as a proxy, an engineered physical proximity between *RA2/Pitx1* and *RA3* (*Pitx1^{del4}*) was induced. Strikingly, in forelimb tissue the inactive MAD is partially transformed into an active one. This new conformation induced a strong misexpression of *Pitx1* in forelimb tissue, with consequent limb malformation exhibiting a striking transformation of the elbow joint into a knee-like joint, resembling that of *Prx-Pitx1* mice (DeLaurier et al., 2006). This intriguing finding demonstrated that by forcing the *RA2/Pitx1* and *RA3* physical proximity, *Pitx1* can be inappropriately activated in forelimb tissue by its endogenous *pan*-limb enhancers, even in the absence of tissue specific TFs.

6.4 The Pathogenicity of Structural Variants at the *PITX1* locus

Deletions and duplications within in the *PITX1* regulatory landscape in humans have been associated with partial homeotic arm-to-leg transformation Liebenberg syndrome (Spielmann et al., 2012). The underlying pathomechanism has been suggested to be caused deletion of the TAD boundary at the *H2AFY* region, resulting in enhancer adoption whereby the *pan*-limb enhancer hs1473 (equivalent to *Pen*) ectopically activates *PITX1* in forelimbs (Spielmann et al., 2012). This type of *cis*-regulatory mutational mechanism has been demonstrated at developmental loci such as the *Epha4* locus, whereby the deletion of either TAD boundary results in rewiring *Epha4* enhancers, normally regulating the gene in the distal limb, with either *Ihh* or *Pax3*, resulting in their ectopic expression in the distal limb and consequent limb malformation (Lupianez et al., 2015). Here, we re-engineered human Liebenberg alleles in mice to determine the underlying pathomechanism as well as the concomitant changes in 3D genome architecture of the *Pitx1* locus.

Discussion

6.4.1 Liebenberg syndrome is caused by *regulatory endoactivation*

As discussed above, the fore- and hindlimb active enhancer *Pen* (hs1473 in humans) regulates *Pitx1* in hindlimb tissue. To test if structural variants enable *Pen* to erroneously activate *Pitx1* in forelimbs tissue, it was misplaced at several positions at the locus through inversion and deletion alleles. Remarkably, placing *Pen* in the position of *RA4* (*Pitx1^{inv1}*) resulted in a complete transformation of the inactive forelimb MAD into an active conformation. With this active forelimb MAD, followed the strong misexpression of *Pitx1* and consequent transformed forelimb-to-hindlimb morphologies. This homeotic phenotype, was the strongest forelimb-to-hindlimb transformation produced, and suggests that form instructs function i.e. the more accurate the recapitulation of the hindlimb-like active MAD conformation, the stronger the transformation of forelimb tissue morphology. The reengineering of a Liebenberg deletion resulted in a similar, but weaker forelimb-to-hindlimb transformation of chromatin architecture, and consequently weak ectopic expression of *Pitx1* in forelimb tissue. Interestingly, these mice only showed a molecular phenotype and no forelimb malformation. In contrast to established cases in which enhancers activate genes that are not their physiological targets, in the Liebenberg-associated *PITX1* alleles examined here, chromatin mis-folding permits the erroneous interaction between a gene and its endogenous, normally spatially sequestrated enhancer (Franke et al., 2016; Lettice et al., 2011; Lupianez et al., 2015; Weischenfeldt et al., 2017). To distinguish this process from previously described enhancer adoption scenarios, the term *"regulatory endoactivation"* has been coined to describe the underlying pathomechanism.

6.4.2 *H2afy* promoter region traps *Pen* activity in forelimbs

The Liebenberg chromosomal aberrations indicate that the relative distance between *PITX1/Pitx1* and hs1473/*Pen* induced by structural rearrangements affects the severity of the phenotype. This hypothesis is supported by the observation of Liebenberg patients whereby the closer hs1473 was misplaced in relation to *PITX1*, the more severe the phenotype is. For instance, a 420 kb duplication encompassing *PITX1*, *H2AFY* and *hs1473*, placing *PITX1* immediately telomeric of hs1473, results in a very severe form of Liebenberg syndrome (Seoighe et al., 2014), while a 8.5 kb removal of the *H2AFY* promoter identified in this study produces a mild phenotype (Tiberio et al., 2000). Remarkably, smaller deletions engineered in mice (up to 10 kb, *Pitx1$^{delL2/delL2}$* and *Pitx1$^{del6/del6}$*), spanning the *H2afy* promoter region induce

a weak *Pitx1* misexpression in forelimb tissue. Puzzlingly, deleting 58 kb of *H2afy* gene body, leaving the promoter intact, did not induce any ectopic expression in forelimb tissue. This demonstrates a distance independent ability of *Pen* to interact with its cognate promoter *Pitx1*, but only if *H2afy* promoter is removed. Housekeeping genes are often embedded in developmental loci, as is the case at the *Shh* locus (*Lmbr1*), *HoxD* locus (*Lnp and Mtx2*) and *Ihh* locus (*Nhej1*). However, the exact role of housekeeping genes at these loci is unknown. They have been suggested to buffer the regulatory activities of tissue specific genes by clustering them with other housekeeping genes (Brown et al., 2006). In this way, an open chromatin state of the developmental locus is maintained allowing quick activation of the promoter in the presence of tissue specific transcription factors. The insulating mechanism of *H2afy* might have evolved through alterations in enhancer-promoter specificity through slight alteration of the TF binding code in *Pen* region or the promoter region of *H2afy*. One may speculate that in forelimbs, TFs combinatory binding to *Pen* guides its activation of *H2afy*, whereas in hindlimbs, a slightly different combinatory binding of TFs dictates its activation of *Pitx1*. Nevertheless, in forelimbs, *Pen* has a strong affinity for the *Pitx1* promoter, since removal of 400 bp of *H2afy* promoter caused ectopic *Pitx1-Pen* communication in forelimbs.

6.4.3 Model of *Pitx1* regulation in limbs

Together with the above discussed results, I suggest the following *Pitx1* regulatory model (Figure 20). The *Pitx1* locus is poised for transcriptional onset in both prospective forelimb and hindlimb regions of the LPM, covered by the bivalent promoter marks H3K4me3 and H3K27me3 as well as poised RNA polymerase II.

Discussion
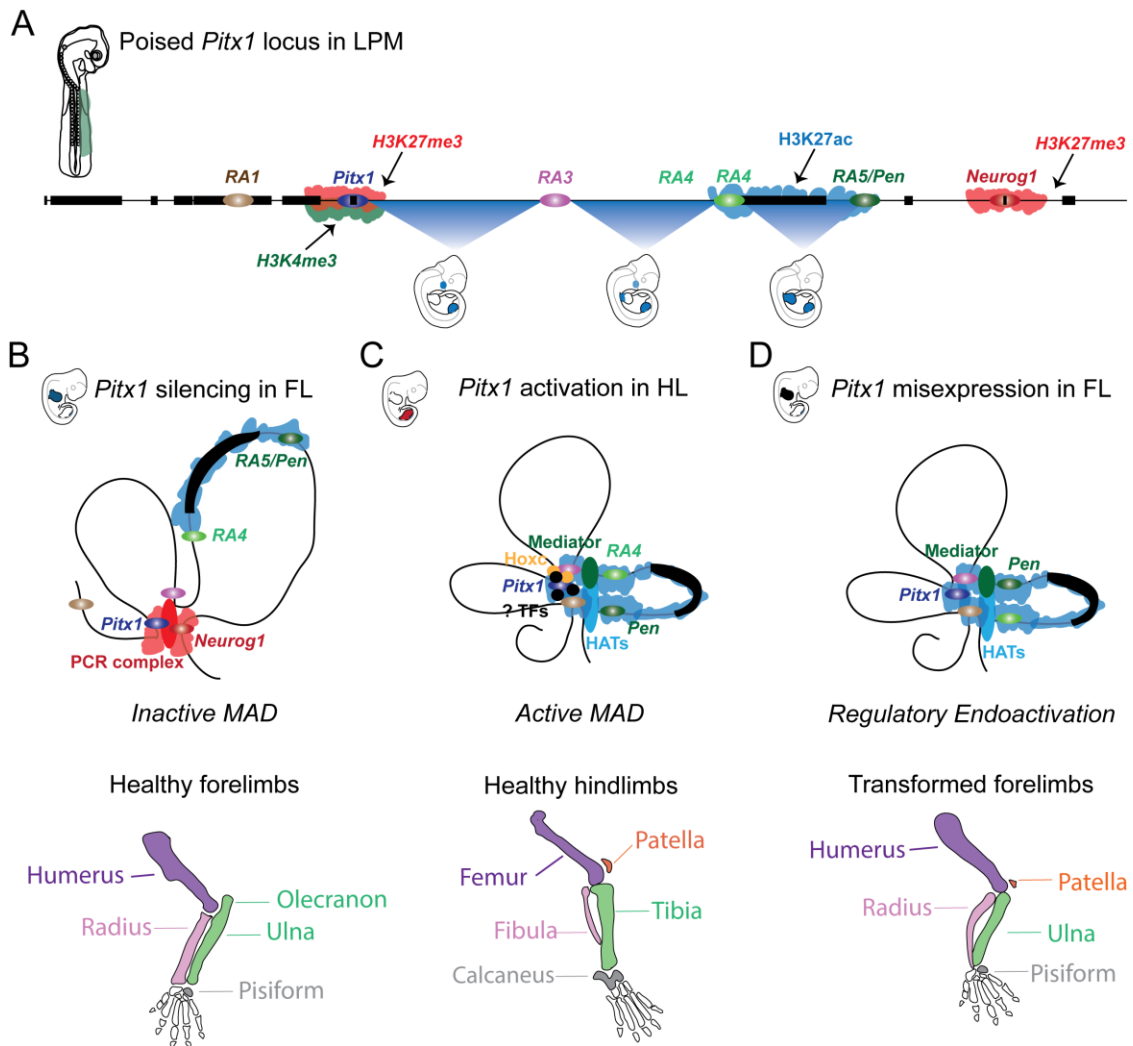
Figure 20 A regulatory model for *Pitx1* regulation in limb development and disease

A. *Pitx1* is poised for activity in LPM (green highlight on E8.75 embryo) with its bivalent promoter marks H3K27me3 and H3K4me3. The regulatory landscape contains both tissue specific and unspecific limb enhancers. **B.** The absence of additional TFs and coactivators in presumptive forelimb mesoderm renders *Pitx1* inactive reinforcing the H3K27me3 Polycomb mediated repression together with *Neurog1*. This ensures development of healthy arms. **C.** In presumptive hindlimb mesoderm, the presence of tissue specific TFs such as HoxC, bind to the *Pitx1* region and recruit HATs and Mediator complex to form a transcriptionally active pocket depositing H3K27ac marks. The affinity of H3K27ac marks together with DNA binding factors folds the *Pitx1* chromatin region into a transcriptionally active pocket result in robust expression and normal hindlimb development. **D.** Structural variants at the *Pitx1* locus misplacing the active *pan*-limb enhancer *Pen* results in activation and H3K27ac deposition of *Pitx1, RA1* and *RA3* and folding of the locus in the absence of tissue specific TFs, suggesting the active enhancer might recruit the HATs and direct the transcriptional machinery to induce robust *Pitx1* expression in forelimb tissue resulting in partial arm-to-leg transformation.

The presence of *pan*-limb TFs activates the *Pitx1 pan*-limb region by deposition of H3K27ac priming the locus for rapid and robust expression. Moreover, architectural proteins CTCF and cohesin are loaded at the regulatory anchors to assist in the stabilisation of interactions. In forelimbs, no further regulatory signal occurs, resulting in loss of H3K4me3 activation mark, rendering the *Pitx1* promoter inactive, and strengthening the repressive contact with *Neurog1*. In posterior LPM, on the other hand, tissue specific transcription factors such as

Discussion

HoxC TFs bind to the *RA2/Pitx1* and *RA3* regions and recruit HATs and Mediator complex. They, in turn, instruct chromatin loops between *RA2/Pitx1* and *RA3*, thereby compacting the locus, enabling interactions with the active *RA4* and *RA5/Pen*. Concomitantly, a transcriptionally active pocket is formed and a rapid and robust onset of *Pitx1* expression ensures the outgrowth and patterning of healthy hindlimbs. Structural variants perturbating this organisation, allow the ectopic activation of *Pitx1* by the *pan*-limb enhancers in the absence of posterior LPM TFs. The enhancers recruit histone modifying to replace repressive H3K27me3 marks with H3K27ac active marks, enabling RNA polymerase II to transcribe *Pitx1* in forelimb tissue. This *regulatory endoactivation* results in the transformation of forelimb morphologies into hindlimb-like.
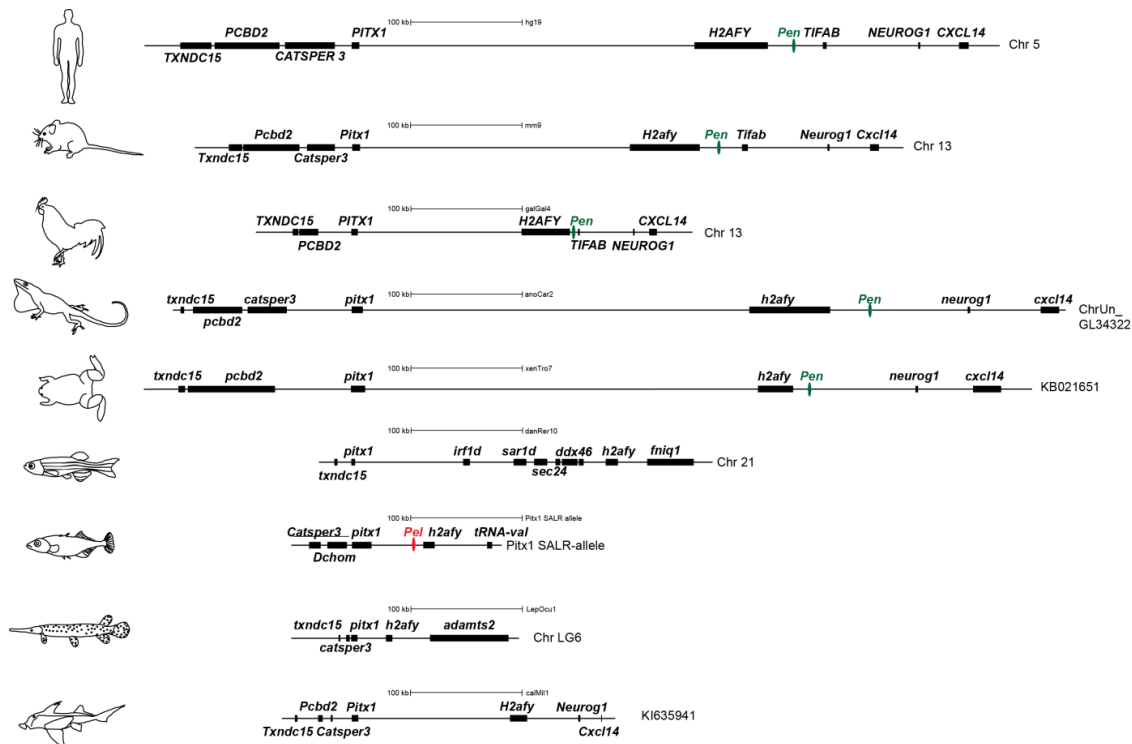
6.5 *PITX1* locus in vertebrate evolution

During evolution, developmental genes are often co-opted through the emergence of *cis*-regulatory elements or ultimately the mutation of cognate enhancers produce pleotropic expression patterns (de Laat and Duboule, 2013). The transformation of forelimb and hindlimb morphologies resulting from altered *Pitx1* regulation is found in several *phyla*. In fact, the synteny of the *Pitx1* locus is conserved in most vertebrates including the elephant shark. However, in certain teleost species, such as zebrafish that have undergone genome duplications, this synteny is broken (Figure 21). In the final step of this work, the role of *Pen* across vertebrate development is discussed.

6.3.1 *Pen* enhancer in vertebrate evolution

Marine stickleback fish have a prominent pelvic spine that has been suggested to protect them from predators. However, many fresh water stickleback fish have lost the pelvic spine, in the absence of predators and limited amounts of calcium in lakes (Reimchen, 1983). In these pelvic reduced stickleback species, deletions in the *Pitx1* regulatory landscape have been shown to remove the *Pel* enhancer, that normally drive *Pitx1* expression in pelvic fins (Chan et al., 2010). *Pel* is conserved in a few teleost fish, but not in other vertebrates (Chan et al., 2010).

Discussion

Figure 21 Evolutionary conservation of the *Pitx1* locus

Schematic representation of the *Pitx1* extended locus in several species, showing conserved synteny from humans across vertebrate species including the elephant shark. From top to bottom: Human (*Homo sapiens*), mouse (*Mus musculus*), chicken (*Gallus gallus*), lizard (*Anolis carolinensis*), frog (*Xenopus (Silurana) tropicalis),* zebrafish (*Danio rerio*), stickleback (*Gasterosteus aculeatus*), spotted gar (*Lepisosteus oculatus*), and elephant shark (*Callorhinchus milii*). *Pen* (in green) is found in tetrapods, but not in fish. The previous characterised pelvic enhancer, *Pel*, is displayed in red in Stickleback.

Interestingly, *Pel* is embedded in a genomic region, characterised by high number of repeats, which have been shown to be exceptionally flexible compared to the rest of the genome, being a hotspot for recurrent deletions of the enhancer in geographically distinct regions (Chan et al., 2010). Similarly, in Liebenberg syndrome, double strand breaks often occur in repeat rich region 5' of *H2AFY* and centromeric of hs1473 (Al-Qattan et al., 2013; Spielmann et al., 2012). Radiation induced double strand breaks accumulate in repeat rich regions and often result in deletions or error-prone repair by non-homolgous end-joining leading to cancer (Argueso et al., 2008; de Campos-Nebel et al., 2010; Durkin et al., 2008). Thus, the flexibility of the *Pitx1* locus might have undergone positive selection to facilitate expansion of *cis*-regulatory elements in the course of evolution (Chan et al., 2010). Interestingly, certain pigeon breeds, display a partial leg-to-wing homeotic transformation, evident from the appearance of leg feathers (Domyan et al., 2016). These pigeons carry deletions spanning the *Pen* enhancer region, with similar centromeric breakpoint in the fragile site 5' of *H2AFY*. In pigeons, this results in partial loss of *Pitx1* expression accompanied by an upregulation of the

Discussion

forelimb-specific gene *Tbx5* in the hindlimb (Boer et al., 2017; Domyan et al., 2016). Together, the modulation of *cis*-regulatory elements such as *Pen* in pigeons, mice and humans, as well as *Pel* in stickleback fish enables differential phenotypic output, as mutations in coding regions of *Pitx1* are lethal in mice and humans (Klopocki et al., 2012; Lanctot et al., 1999; Szeto et al., 1999).

As shown in this study, *Pen* is a very potent enhancer and thus its conservation might have been favoured during tetrapod evolution. It can be speculated that *Pen* arose in a tetrapod common ancestor, such as *Protopterus* and *Tiktaalik,* transitioning from teleost fish to land vertebrates, driving *Pitx1* expression in both forelimbs and hindlimbs to facilitate the specification limbs enabling weight bearing and locomotion needed to climb on land (Shubin et al., 2006). Later in tetrapod evolution, the development of a dynamic organisation of the *Pitx1* locus into a tissue specific MAD permitted the hindlimb specific use of *Pen*. This hypothesis is supported by the dynamic changes in domain structure observed at the *HoxD* locus. The *HoxD* locus is organised into a bipartite TAD separated by a flexible boundary that allows the adaptable use of the *cis*-regulatory elements from either the centromeric or telomeric gene regulatory deserts (Andrey et al., 2013). Accordingly, the rise of a 3D-chromatin organisation layer at the *HoxD* locus probably arose in tetrapods together with the emergence of the autopod (Acemel et al., 2016). Thus, this may be a general regulation of developmental genes, including *Pitx1*, rendering them flexible in the patterning of the embryo across animal evolution. However, not only TADs confer the additional layer of gene regulation, but here a previously undescribed MAD architecture provides an alternative mechanism of dynamic gene regulation.

6.5 Conclusion and outlook

Unlike the hindlimb specific expression pattern of *Pitx1*, a *pan*-limb enhancer, *Pen*, is here shown to be a key *Pitx1* limb enhancer. This challenges the prevailing definition and understanding of enhancers; enhancers have been defined as stretches of DNA to which tissue-specific TFs bind and dictate its activity on target genes via long-range interactions. *Pen* is active in both forelimbs and hindlimbs and needs to be modulated by the 3D-architecture into a hindlimb specific transcriptional output. This has implications for how enhancer reporter assays are interpreted. Many orphan enhancers, i.e. not reflecting the activity of nearby genes, have been identified genome wide (Visel et al., 2007). It is plausible that the

Discussion

unspecific activity of enhancers allows the robust and quick activation of target genes, which can be quickly modulated by the chromatin architecture. This type of instructive architecture without any delimiting boundary elements, deviates from TADs, whose intra-TAD structure is instructive, but its boundaries are facultative (Dixon et al., 2012 {Phillips-Cremins, 2013 #508)}. Thus, this instructive architecture of the *Pitx1* locus, here named MAD, plays an active role in specifying dynamic enhancer-promoter pairing and transcriptional output. Although MADs are distinct from TADs, their existence does not necessarily contradict each other. Rather, at loci in which very long range interactions between cognate enhancers and promoters occur, the TAD structure supports the pairing (Rao et al., 2017; Schwarzer, 2017). The MAD type of organisation identified in this study has not been described elsewhere; however, it is unlikely to be the only one. Inspection of high resolution 3D architecture genome wide and across tissues should be undertaken to identify other MAD genomic regions. The in-depth analysis of other loci is scarce in the literature. Thus, investigation of other smaller developmental loci that contain unspecific enhancer elements might unmask more MAD organised loci, such as that at the ß-globin locus (Huang et al., 2017). Moreover, 3D-polymer modelling of MADs show striking similarities to the folding of amino acids into proteins guided by the different properties of the amino acids (hydrophobic and hydrophilic). This allows the formation of an enzymatically active pocket, resulting in an environment with high concentration of transcriptional factors, resulting in robust expression. Thus, the 3D and 4D (time) visualisation of chromatin in cells and tissues will thus allow the understanding of how the genome is dynamically organised in the nucleus and how mutations lead to genome dysorganisation and disease.

The novel description and functional characterisation of MAD organisation has implication for human disease. Congenital limb malformations are commonly cause by structural variants at developmental loci affecting the non-coding *cis*-regulatory landscape (Flöttmann, 2017; Spielmann and Mundlos, 2013). In this work, we identify a novel type of *cis*-regulatory mutation, in which pathogenic misexpression of a gene is driven by its own unspecific endogenous enhancer, here termed "*regulatory endoactivation*". This delineation of this pathomecansim was only possible by combining transgenic assay, genetic re-engineering of mutants in mice, and cHi-C. Thus, this type of pathomechanism needs to be considered and a combination of tools used when analysis and interpreting structural variants leading to human disease.

Discussion

Last, but not least, this project left numerous outstanding questions unanswered. Does *Pitx1* have a hindlimb specific enhancer? When did the *Pen* enhancer evolve? How and when did the MAD organisation evolve at the *Pitx1* locus? What is the underlying nature of the instructive dynamic chromatin formation genome wide? Which *Pitx1* enhancers are important for early hindlimb outgrowth and which for patterning of the hindlimb specific structures? Are there other genes out there regulated by strong unspecific enhancers? Are there additional loci organised as a MAD? What are the additional upstream factors of *Pitx1*? What is the exact role of *H2afy* at the *Pitx1* locus? The following experiments may answer some of these questions. Continuous transgenic reporter testing of regions at *RA2/Pitx1* and sequences spanning up to *RA5/Pen* may lead to the identification of a hindlimb specific enhancer. Identification of additional *trans*-acting factors inducing and maintaining *Pitx1* regulation together with HOXC and PITX1 proteins could be assessed by systematic dissection of candidate upstream factors. RNA sequencing of anterior and posterior LPM might identify candidate upstream *Pitx1* TFs, and their role can be subsequently followed up using CRISPR-Cas9 genetic dissection. Furthermore, crossing of mutants, such as *Pitx1$^{fs/+}$* and *HoxC$^{del/+}$*, as well as between transgenic *LacZ* sensors mice with for instance *HoxC$^{del/+}$*, can delineate the contribution of the protein products in *trans* and the influence of enhancer elements on the regulatory activity of the locus. Furthermore, it remains to be tested whether H3K27ac is the underlying attraction force conveying dynamic chromatin interactions. The dead-CRISPR-Cas9 system coupled with p300 core unit facilitates the ectopic acetylation of the regulatory anchors in developing forelimb buds (Hilton et al., 2015). However, this remains challenging to perform *in embryo*. The role of housekeeping genes within developmental loci remains obscure. The specificity of the *H2afy* promoter could be tested by removing the promoter and replacing it by any other ubiquitously expressed promoter to examine whether it may contain sequence specific factors scavenging *Pen* specifically in forelimb tissue. Furthermore, identifying a pre-tetrapod species, such as lungfish (*Protopterus*), where *Pen* might have been active in both forelimb and hindlimb tissue, provides the unique opportunity to compare the organisation of the locus with that of mice. Quantification of *Pitx1* expression in pectoral and pelvic fins lungfish, stickleback, xenopus, and additional vertebrate specifies might provide an answer whether *Pitx1* is expressed in both appendages. If so, chromatin conformation capture of paired appendages in these species might provide a clue as to when the MAD

Discussion

arose at the *Pitx1* locus. Ultimately, such experiments will perhaps give us a regulatory insight into the rise of forelimbs and hindlimbs in pre-tetrapods.

Summary

7. SUMMARY

The tissue specific expression of developmental genes is encoded in enhancer elements often located hundreds of kb away from their cognate promoters. Physical chromatin interactions between these enhancers and target promoters are associated with active transcription and conventionally thought to be confined to topologically associating domains (TADs). However, little is known about the underlying nature and dynamics of this 3D-architecture during development and its perturbation in disease. In this work, the mouse embryonic limb bud was used as a paradigm to investigate the dynamics of gene regulation underlying the development of either arms or legs. *Pitx1* is the one of few transcription factors shown to be expressed exclusively in hindlimbs and not in forelimbs. Yet, its regulation in mammals continues to be largely unknown.

Here, we identified an unexpectedly complex regulatory basis of hindlimb-specific *Pitx1* expression that expands the current model of enhancer sequences as the sole determinants of tissue specificity. We demonstrate that *Pitx1* is regulated by the active fore- and hindlimb enhancer, *Pen*, that is required for normal expression of *Pitx1* in hindlimbs, but does not activate *Pitx1* expression in forelimbs. Investigation of the chromatin architecture of the *Pitx1* locus in both fore-and hindlimb buds using cHi-C and derived 3D-models, revealed a modular regulatory landscape that is not confined to a TAD structure. Instead, *Pitx1* is controlled by a Multi-Anchor Domain (MAD), which can assume distinct tissue-specific conformations. In the hindlimb, the locus forms an active MAD that enable *Pen* and *Pitx1* interactions, producing a transcriptionally active pocket. Intriguingly, an alternate forelimb-specific MAD conformation prevents the promiscuous activity of *Pen* by physically separating it from *Pitx1*. Disruption of this segregated forelimb chromatin conformation in engineered mice, as well as in human Liebenberg syndrome patients, results in forelimb *Pitx1* misexpression, and a partial transformation of forelimb morphology into a hindlimb-like. This work provides further understanding of gene regulation whereby unspecific enhancer activity is actively regulated by the dynamics in 3D-chromatin architecture, independent of TADs, to confer a tissue specific transcriptional output. Together our findings help build the groundwork for the interpretation of structural variants disrupting genome organisation, not only resulting in human disease, but also in the evolution of phenotypes in natural populations.

8. ZUSAMMENFASSUNG

Die gewebespezifische Expression von Entwicklungsgenen wird durch Enhancerelemente gesteuerte, welche häufig hunderte Kilobasen von ihren Zielpromotoren entfernt liegen können. Physische Chromatininteraktionen zwischen diesen Enhancern und ihren Zielpromotoren werden mit aktiver Transkription in Zusammenhang gebracht und ist in der Regel auf „topologically associating domains" (TADs) begrenzt. Allerdings ist noch wenig über die dieser 3D-Architektur in der Entwicklung zugrunde liegenden Natur und Dynamik bekannt und wie diese in Krankheiten gestört wird.

In dieser Arbeit wurde die embryonale Extremitätenentwicklung der Maus als Paradigma genutzt, um die komplexe Genregulation während der spezifischen Arm- und Beinentwicklung zu untersuchen. Pitx1 ist einer der wenigen Transkriptionsfaktoren, die ausschließlich in den hinteren Extremitäten, und nicht in den vorderen Extremitäten, exprimiert werden.

Dessen Regulation in Säugetieren ist jedoch bislang noch größtenteils unbekannt.

Hier, wurde eine unerwartet komplexe regulatorische Grundlage für die Beinentwicklung spezifische *Pitx1* Expression identifiziert. Diese erweitert das aktuelle Modell, welches Enhancersequenzen noch als alleinige Determinanten der Gewebespezifität benennt. Wir zeigen, dass *Pitx1* von einem in Vorder-und Hintergliedmaßen aktiven Enhancer, genannt *Pen*, reguliert wird. Dieser wird in den hinteren Extremitäten für die *Pitx1* Expression benötigt, führt aber in der regulären Entwicklung der Vordergliedmaßen zu keiner *Pitx1* Aktivierung.

Untersuchungen der Chromatinarchitektur, des *Pitx1* Lokus in Extremitätenknospen von sowohl Vorder-, als auch Hintergliedmaßen, mittels „capture Hi-C" und abgeleiteten 3D-Modellen, ergaben eine modulare regulatorische Landschaft, welche nicht auf eine TAD Struktur begrenzt ist. Stattdessen, konnte nachgewiesen werden, dass *Pitx1* von einer „Multi-Anchor Domain" (MAD), die Gewebespezifische Konformationen annehmen kann, gelenkt wird. In den Hintergliedmaßen formt der Lokus eine aktive MAD, die eine Interaktion zwischen *Pen* und *Pitx1* ermöglicht, wobei eine transkriptionell aktive Tasche erzeugt wird. Interessanterweise verhindert eine alternative Vordergliedmaßen-spezifische MAD Konformation unspezifische *Pen* Aktivitäten, indem *Pen* physisch von

Zusammenfassung

Pitx1 getrennt wird. Störungen dieser segregierten Chromatinkonformation, sowohl in genetisch manipulierten Mäusen, als auch in menschlichen Patienten mit Liebenberg Syndrom, führen zu einer Misexpression von *Pitx1* in den vorderen Extremitäten, wobei eine partielle Transformation der Vordergliedmaßenmorphologie in eine den Hintergliedmaßen ähnlichen Struktur entsteht. Diese Arbeit schafft neue Erkenntnisse über Genregulation, da nachgewiesen wurde, dass, unabhängig von TADs unspezifische Enhancer in einer aktiven Weise von den Dynamiken der 3D-Architektur des Chromatins reguliert werden können, um gewebespezifische transkriptionelle Aktivität zu ermöglichen. Gemeinsam helfen unsere Erkenntnisse Grundlagen für die Interpretation von Strukturvariationen, welche die 3D-Organisation des Genoms stören, und dadurch nicht nur menschliche Erkrankungen, sondern auch die Beeinflussung der Evolution von Phänotypen in natürlichen Populationen zur Folge haben können, zu bilden.

9. REFERENCES

Acemel, R.D., Maeso, I., and Gomez-Skarmeta, J.L. (2017). Topologically associated domains: a successful scaffold for the evolution of gene regulation in animals. Wiley Interdiscip Rev Dev Biol *6*.

Acemel, R.D., Tena, J.J., Irastorza-Azcarate, I., Marletaz, F., Gomez-Marin, C., de la Calle-Mustienes, E., Bertrand, S., Diaz, S.G., Aldea, D., Aury, J.M.*, et al.* (2016). A single three-dimensional chromatin compartment in amphioxus indicates a stepwise evolution of vertebrate Hox bimodal regulation. Nat Genet *48*, 336-341.

Agarwal, P., Wylie, J.N., Galceran, J., Arkhitko, O., Li, C., Deng, C., Grosschedl, R., and Bruneau, B.G. (2003). Tbx5 is essential for forelimb bud initiation following patterning of the limb field in the mouse embryo. Development *130*, 623-633.

Akam, M. (1989). Hox and HOM: homologous gene clusters in insects and vertebrates. Cell *57*, 347-349.

Al-Qattan, M.M., Al-Thunayan, A., Alabdulkareem, I., and Al Balwi, M. (2013). Liebenberg syndrome is caused by a deletion upstream to the PITX1 gene resulting in transformation of the upper limbs to reflect lower limb characteristics. Gene *524*, 65-71.

Alvarado, D.M., McCall, K., Aferol, H., Silva, M.J., Garbow, J.R., Spees, W.M., Patel, T., Siegel, M., Dobbs, M.B., and Gurnett, C.A. (2011). Pitx1 haploinsufficiency causes clubfoot in humans and a clubfoot-like phenotype in mice. Hum Mol Genet *20*, 3943-3952.

Anderson, E., Devenney, P.S., Hill, R.E., and Lettice, L.A. (2014). Mapping the Shh long-range regulatory domain. Development *141*, 3934-3943.

Andrey, G., Montavon, T., Mascrez, B., Gonzalez, F., Noordermeer, D., Leleu, M., Trono, D., Spitz, F., and Duboule, D. (2013). A switch between topological domains underlies HoxD genes collinearity in mouse limbs. Science *340*, 1234167.

Andrey, G., and Mundlos, S. (2017). The three-dimensional genome: regulating gene expression during pluripotency and development. Development *144*, 3646-3658.

Andrey, G., Schopflin, R., Jerkovic, I., Heinrich, V., Ibrahim, D.M., Paliou, C., Hochradel, M., Timmermann, B., Haas, S., Vingron, M.*, et al.* (2017). Characterization of hundreds of regulatory landscapes in developing limbs reveals two regimes of chromatin folding. Genome Res *27*, 223-233.

Argueso, J.L., Westmoreland, J., Mieczkowski, P.A., Gawel, M., Petes, T.D., and Resnick, M.A. (2008). Double-strand breaks associated with repetitive DNA can reshape the genome. Proc Natl Acad Sci U S A *105*, 11845-11850.

Artus, J., and Hadjantonakis, A.K. (2011). Generation of chimeras by aggregation of embryonic stem cells with diploid or tetraploid mouse embryos. Methods Mol Biol *693*, 37-56.

Beagrie, R.A., Scialdone, A., Schueler, M., Kraemer, D.C., Chotalia, M., Xie, S.Q., Barbieri, M., de Santiago, I., Lavitas, L.M., Branco, M.R.*, et al.* (2017). Complex multi-enhancer contacts captured by genome architecture mapping. Nature *543*, 519-524.

Beccari, L., Yakushiji-Kaminatsui, N., Woltering, J.M., Necsulea, A., Lonfat, N., Rodriguez-Carballo, E., Mascrez, B., Yamamoto, S., Kuroiwa, A., and Duboule, D. (2016). A role for HOX13 proteins in the regulatory switch between TADs at the HoxD locus. Genes Dev *30*, 1172-1186.

References

Behringer, R., Anderson, M. D., Nagy, K. V., & Nagy, A. (1994). Manipulating the Mouse Embryo: A Laboratory Manual, Vol Fourth Edition (Cold Spring Harbor Laboratory Press).

Bell, A.C., West, A.G., and Felsenfeld, G. (1999). The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. Cell *98*, 387-396.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. J Roy Stat Soc B Met *57*, 289-300.

Berger, M.F., Badis, G., Gehrke, A.R., Talukder, S., Philippakis, A.A., Pena-Castillo, L., Alleyne, T.M., Mnaimneh, S., Botvinnik, O.B., Chan, E.T.*, et al.* (2008). Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. Cell *133*, 1266-1276.

Berlivet, S., Paquette, D., Dumouchel, A., Langlais, D., Dostie, J., and Kmita, M. (2013). Clustering of tissue-specific sub-TADs accompanies the regulation of HoxA genes in developing limbs. PLoS Genet *9*, e1004018.

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

Boer, E.F., Van Hollebeke, H.F., and Shapiro, M.D. (2017). Genomic determinants of epidermal appendage patterning and structure in domestic birds. Dev Biol *429*, 409-419.

Bonev, B., and Cavalli, G. (2016). Organization and function of the 3D genome. Nat Rev Genet *17*, 772.

Bonev, B., Mendelson Cohen, N., Szabo, Q., Fritsch, L., Papadopoulos, G.L., Lubling, Y., Xu, X., Lv, X., Hugnot, J.P., Tanay, A.*, et al.* (2017). Multiscale 3D Genome Rewiring during Mouse Neural Development. Cell *171*, 557-572 e524.

Boulard, M., Storck, S., Cong, R., Pinto, R., Delage, H., and Bouvet, P. (2010). Histone variant macroH2A1 deletion in mice causes female-specific steatosis. Epigenetics Chromatin *3*, 8.

Brown, J.M., Leach, J., Reittie, J.E., Atzberger, A., Lee-Prudhoe, J., Wood, W.G., Higgs, D.R., Iborra, F.J., and Buckle, V.J. (2006). Coregulated human globin genes are frequently in spatial proximity when active. J Cell Biol *172*, 177-187.

Byrne, S.M., Ortiz, L., Mali, P., Aach, J., and Church, G.M. (2015). Multi-kilobase homozygous targeted gene replacement in human induced pluripotent stem cells. Nucleic Acids Res *43*, e21.

Chan, Y.F., Marks, M.E., Jones, F.C., Villarreal, G., Jr., Shapiro, M.D., Brady, S.D., Southwick, A.M., Absher, D.M., Grimwood, J., Schmutz, J.*, et al.* (2010). Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. Science *327*, 302-305.

Changolkar, L.N., Costanzi, C., Leu, N.A., Chen, D., McLaughlin, K.J., and Pehrson, J.R. (2007). Developmental changes in histone macroH2A1-mediated gene regulation. Mol Cell Biol *27*, 2758-2764.

Chiariello, A.M., Annunziatella, C., Bianco, S., Esposito, A., and Nicodemi, M. (2016). Polymer physics of chromosome large-scale 3D organisation. Sci Rep *6*, 29775.

Cohn, M.J., and Tickle, C. (1999). Developmental basis of limblessness and axial patterning in snakes. Nature *399*, 474-479.

Crews, S.T., and Pearson, J.C. (2009). Transcriptional autoregulation in development. Curr Biol *19*, R241-246.

References

Danino, Y.M., Even, D., Ideses, D., and Juven-Gershon, T. (2015). The core promoter: At the heart of gene expression. Biochim Biophys Acta *1849*, 1116-1131.

de Campos-Nebel, M., Larripa, I., and Gonzalez-Cid, M. (2010). Topoisomerase II-mediated DNA damage is differently repaired during the cell cycle by non-homologous end joining and homologous recombination. PLoS One *5*.

de Laat, W., and Duboule, D. (2013). Topology of mammalian developmental enhancers and their regulatory landscapes. Nature *502*, 499-506.

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. Science *295*, 1306-1311.

DeLaurier, A., Schweitzer, R., and Logan, M. (2006). Pitx1 determines the morphology of muscle, tendon, and bones of the hindlimb. Dev Biol *299*, 22-34.

Deng, W., Lee, J., Wang, H., Miller, J., Reik, A., Gregory, P.D., Dean, A., and Blobel, G.A. (2012). Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. Cell *149*, 1233-1244.

Deng, W., Rupon, J.W., Krivega, I., Breda, L., Motta, I., Jahn, K.S., Reik, A., Gregory, P.D., Rivella, S., Dean, A.*, et al.* (2014). Reactivation of developmentally silenced globin genes by forced chromatin looping. Cell *158*, 849-860.

Denholtz, M., Bonora, G., Chronis, C., Splinter, E., de Laat, W., Ernst, J., Pellegrini, M., and Plath, K. (2013). Long-range chromatin contacts in embryonic stem cells reveal a role for pluripotency factors and polycomb proteins in genome organization. Cell Stem Cell *13*, 602-616.

Dixon, J.R., Gorkin, D.U., and Ren, B. (2016). Chromatin Domains: The Unit of Chromosome Organization. Mol Cell *62*, 668-680.

Dixon, J.R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J.E., Lee, A.Y., Ye, Z., Kim, A., Rajagopal, N., Xie, W.*, et al.* (2015). Chromatin architecture reorganization during stem cell differentiation. Nature *518*, 331-336.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376-380.

Domyan, E.T., Kronenberg, Z., Infante, C.R., Vickrey, A.I., Stringham, S.A., Bruders, R., Guernsey, M.W., Park, S., Payne, J., Beckstead, R.B.*, et al.* (2016). Molecular shifts in limb identity underlie development of feathered feet in two domestic avian species. Elife *5*, e12115.

Dowen, J.M., Fan, Z.P., Hnisz, D., Ren, G., Abraham, B.J., Zhang, L.N., Weintraub, A.S., Schujiers, J., Lee, T.I., Zhao, K.*, et al.* (2014). Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. Cell *159*, 374-387.

Durand, N.C., Shamim, M.S., Machol, I., Rao, S.S., Huntley, M.H., Lander, E.S., and Aiden, E.L. (2016). Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. Cell Syst *3*, 95-98.

Durkin, S.G., Ragland, R.L., Arlt, M.F., Mulle, J.G., Warren, S.T., and Glover, T.W. (2008). Replication stress induces tumor-like microdeletions in FHIT/FRA3B. Proc Natl Acad Sci U S A *105*, 246-251.

Farre, D., Bellora, N., Mularoni, L., Messeguer, X., and Alba, M.M. (2007). Housekeeping genes tend to show reduced upstream sequence conservation. Genome Biol *8*, R140.

Ferrai, C., Torlai Triglia, E., Risner-Janiczek, J.R., Rito, T., Rackham, O.J., de Santiago, I., Kukalev, A., Nicodemi, M., Akalin, A., Li, M.*, et al.* (2017). RNA polymerase II primes

References

Polycomb-repressed developmental genes throughout terminal neuronal differentiation. Mol Syst Biol *13*, 946.

Flöttmann, R., Kragesteen, B.K., Geuer, S., Socha, M., Allou, L., Sowinska-Seidler, A., Bossquillon de Jarcy, L., Wagner, J., Jamsheer, A., Oehl-Jaschkowitz, B., Wittler, L., de Silva, D., Kurth, I., Maya, I., Santos-Simarro, F., Hülsemann, W., Klopocki, E., Mountford, R., Fryer, A., Borck, G., Horn, D., Lapunzina, P., Wilson, M., Mascrez, B., Duboule, D., Mundlos, S., Spielmann, M. (2017). Noncoding copy-number variations are associatedwith congenital limb malformation. Genetics in Medicine.

Franke, M., Ibrahim, D.M., Andrey, G., Schwarzer, W., Heinrich, V., Schopflin, R., Kraft, K., Kempfer, R., Jerkovic, I., Chan, W.L.*, et al.* (2016). Formation of new chromatin domains determines pathogenicity of genomic duplications. Nature *538*, 265-269.

Freire-Pritchett, P., Schoenfelder, S., Varnai, C., Wingett, S.W., Cairns, J., Collier, A.J., Garcia-Vilchez, R., Furlan-Magaril, M., Osborne, C.S., Fraser, P.*, et al.* (2017). Global reorganisation of cis-regulatory units upon lineage commitment of human embryonic stem cells. Elife *6*.

Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., and Mirny, L.A. (2016). Formation of Chromosomal Domains by Loop Extrusion. Cell Rep *15*, 2038-2049.

Gaszner, M., and Felsenfeld, G. (2006). Insulators: exploiting transcriptional and epigenetic mechanisms. Nat Rev Genet *7*, 703-713.

Ghirlando, R., and Felsenfeld, G. (2016). CTCF: making the right connections. Genes Dev *30*, 881-891.

Gibson-Brown, J.J., Agulnik, S.I., Chapman, D.L., Alexiou, M., Garvey, N., Silver, L.M., and Papaioannou, V.E. (1996). Evidence of a role for T-box genes in the evolution of limb morphogenesis and the specification of forelimb/hindlimb identity. Mech Dev *56*, 93-101.

Goodyer, C.G., Tremblay, J.J., Paradis, F.W., Marcil, A., Lanctot, C., Gauthier, Y., and Drouin, J. (2003). Pitx1 in vivo promoter activity and mechanisms of positive autoregulation. Neuroendocrinology *78*, 129-137.

Gray, P.A., Fu, H., Luo, P., Zhao, Q., Yu, J., Ferrari, A., Tenzen, T., Yuk, D.I., Tsung, E.F., Cai, Z.*, et al.* (2004). Mouse brain organization revealed through direct genome-scale TF expression analysis. Science *306*, 2255-2257.

Gros, J., and Tabin, C.J. (2014). Vertebrate limb bud formation is initiated by localized epithelial-to-mesenchymal transition. Science *343*, 1253-1256.

Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D.U., Jung, I., Wu, H., Zhai, Y., Tang, Y.*, et al.* (2015). CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. Cell *162*, 900-910.

Hamada, Y., Kadokawa, Y., Okabe, M., Ikawa, M., Coleman, J.R., and Tsujimoto, Y. (1999). Mutation in ankyrin repeats of the mouse Notch2 gene induces early embryonic lethality. Development *126*, 3415-3424.

Hill, R.E., and Lettice, L.A. (2013). Alterations to the remote control of Shh gene expression cause congenital abnormalities. Philos Trans R Soc Lond B Biol Sci *368*, 20120357.

Hilton, I.B., D'Ippolito, A.M., Vockley, C.M., Thakore, P.I., Crawford, G.E., Reddy, T.E., and Gersbach, C.A. (2015). Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. Nat Biotechnol *33*, 510-517.

References

Hnisz, D., Shrinivas, K., Young, R.A., Chakraborty, A.K., and Sharp, P.A. (2017). A Phase Separation Model for Transcriptional Control. Cell *169*, 13-23.

Hnisz, D., Weintraub, A.S., Day, D.S., Valton, A.L., Bak, R.O., Li, C.H., Goldmann, J., Lajoie, B.R., Fan, Z.P., Sigova, A.A.*, et al.* (2016). Activation of proto-oncogenes by disruption of chromosome neighborhoods. Science *351*, 1454-1458.

Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O.*, et al.* (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. Nat Biotechnol *31*, 827-832.

Hu, Z., and Tee, W.W. (2017). Enhancers and chromatin structures: regulatory hubs in gene expression and diseases. Biosci Rep *37*.

Huang, P., Keller, C.A., Giardine, B., Grevet, J.D., Davies, J.O.J., Hughes, J.R., Kurita, R., Nakamura, Y., Hardison, R.C., and Blobel, G.A. (2017). Comparative analysis of three-dimensional chromosomal architecture identifies a novel fetal hemoglobin regulatory element. Genes Dev *31*, 1704-1713.

Hughes, J.R., Roberts, N., McGowan, S., Hay, D., Giannoulatou, E., Lynch, M., De Gobbi, M., Taylor, S., Gibbons, R., and Higgs, D.R. (2014). Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. Nat Genet *46*, 205-212.

Infante, C.R., Park, S., Mihala, A.G., Kingsley, D.M., and Menke, D.B. (2013). Pitx1 broadly associates with limb enhancers and is enriched on hindlimb cis-regulatory elements. Dev Biol *374*, 234-244.

Juven-Gershon, T., and Kadonaga, J.T. (2010). Regulation of gene expression via the core promoter and the basal transcriptional machinery. Dev Biol *339*, 225-229.

Kim, S., Yu, N.K., and Kaang, B.K. (2015). CTCF as a multifunctional protein in genome regulation and gene expression. Exp Mol Med *47*, e166.

King, D.C., Taylor, J., Zhang, Y., Cheng, Y., Lawson, H.A., Martin, J., Regulation, E.g.f.T., Multispecies Sequence, A., Chiaromonte, F., Miller, W.*, et al.* (2007). Finding cis-regulatory elements using comparative genomics: some lessons from ENCODE data. Genome Res *17*, 775-786.

Klopocki, E., Kahler, C., Foulds, N., Shah, H., Joseph, B., Vogel, H., Luttgen, S., Bald, R., Besoke, R., Held, K.*, et al.* (2012). Deletions in PITX1 cause a spectrum of lower-limb malformations including mirror-image polydactyly. Eur J Hum Genet *20*, 705-708.

Knight, P.A., and Ruiz, D. (2013). A fast algorithm for matrix balancing. Ima J Numer Anal *33*, 1029-1047.

Kooren, J., Palstra, R.J., Klous, P., Splinter, E., von Lindern, M., Grosveld, F., and de Laat, W. (2007). Beta-globin active chromatin Hub formation in differentiating erythroid cells and in p45 NF-E2 knock-out mice. J Biol Chem *282*, 16544-16552.

Kraft, K., Geuer, S., Will, A.J., Chan, W.L., Paliou, C., Borschiwer, M., Harabula, I., Wittler, L., Franke, M., Ibrahim, D.M.*, et al.* (2015). Deletions, Inversions, Duplications: Engineering of Structural Variants using CRISPR/Cas in Mice. Cell Rep.

Krijger, P.H., and de Laat, W. (2016). Regulation of disease-associated gene expression in the 3D genome. Nat Rev Mol Cell Biol *17*, 771-782.

Kvon, E.Z., Kamneva, O.K., Melo, U.S., Barozzi, I., Osterwalder, M., Mannion, B.J., Tissieres, V., Pickle, C.S., Plajzer-Frick, I., Lee, E.A.*, et al.* (2016). Progressive Loss of Function in a Limb Enhancer during Snake Evolution. Cell *167*, 633-642 e611.

Lamonerie, T., Tremblay, J.J., Lanctot, C., Therrien, M., Gauthier, Y., and Drouin, J. (1996). Ptx1, a bicoid-related homeo box transcription factor involved in transcription of the pro-opiomelanocortin gene. Genes Dev *10*, 1284-1295.

References

Lanctot, C., Lamolet, B., and Drouin, J. (1997). The bicoid-related homeoprotein Ptx1 defines the most anterior domain of the embryo and differentiates posterior from anterior lateral mesoderm. Development *124*, 2807-2817.

Lanctot, C., Moreau, A., Chamberland, M., Tremblay, M.L., and Drouin, J. (1999). Hindlimb patterning and mandible development require the Ptx1 gene. Development *126*, 1805-1810.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*, 357-359.

Lawson, R., Slowinski, J.B., Crother, B.I., and Burbrink, F.T. (2005). Phylogeny of the Colubroidea (Serpentes): new evidence from mitochondrial and nuclear genes. Mol Phylogenet Evol *37*, 581-601.

Lesch, B.J., Dokshin, G.A., Young, R.A., McCarrey, J.R., and Page, D.C. (2013). A set of genes critical to development is epigenetically poised in mouse germ cells from fetal stages through completion of meiosis. Proc Natl Acad Sci U S A *110*, 16061-16066.

Lettice, L.A., Daniels, S., Sweeney, E., Venkataraman, S., Devenney, P.S., Gautier, P., Morrison, H., Fantes, J., Hill, R.E., and FitzPatrick, D.R. (2011). Enhancer-adoption as a mechanism of human developmental disease. Hum Mutat *32*, 1492-1499.

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.

Levings, P.P., and Bungert, J. (2002). The human beta-globin locus control region. Eur J Biochem *269*, 1589-1599.

Lewis, E.B. (1978). A gene complex controlling segmentation in Drosophila. Nature *276*, 565-570.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754-1760.

Li, Q., Lewandowski, J.P., Powell, M.B., Norrie, J.L., Cho, S.H., and Vokes, S.A. (2014). A Gli silencer is required for robust repression of gremlin in the vertebrate limb bud. Development *141*, 1906-1914.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O.*, et al.* (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science *326*, 289-293.

Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods *25*, 402-408.

Lobe, C.G., Koop, K.E., Kreppner, W., Lomeli, H., Gertsenstein, M., and Nagy, A. (1999). Z/AP, a double reporter for cre-mediated recombination. Dev Biol *208*, 281-292.

Lobley, A., Pierron, V., Reynolds, L., Allen, L., and Michalovich, D. (2003). Identification of human and mouse CatSper3 and CatSper4 genes: characterisation of a common interaction domain and evidence for expression in testis. Reprod Biol Endocrinol *1*, 53.

Logan, M., and Tabin, C.J. (1999). Role of Pitx1 upstream of Tbx4 in specification of hindlimb identity. Science *283*, 1736-1739.

Lupianez, D.G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J.M., Laxova, R.*, et al.* (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell *161*, 1012-1025.

References

Ma, Q., Chen, Z., del Barco Barrantes, I., de la Pompa, J.L., and Anderson, D.J. (1998). neurogenin1 is essential for the determination of neuronal precursors for proximal cranial sensory ganglia. Neuron *20*, 469-482.

Maeda, R.K., and Karch, F. (2011). Gene expression in time and space: additive vs hierarchical organization of cis-regulatory regions. Curr Opin Genet Dev *21*, 187-193.

Mann, R.S., Lelli, K.M., and Joshi, R. (2009). Hox specificity unique roles for cofactors and collaborators. Curr Top Dev Biol *88*, 63-101.

Marcil, A., Dumontier, E., Chamberland, M., Camper, S.A., and Drouin, J. (2003). Pitx1 and Pitx2 are required for development of hindlimb buds. Development *130*, 45-55.

Marinic, M., Aktas, T., Ruf, S., and Spitz, F. (2013). An integrated holo-enhancer unit defines tissue and gene specificity of the Fgf8 regulatory landscape. Dev Cell *24*, 530-542.

Martin, J.F., Bradley, A., and Olson, E.N. (1995). The paired-like homeo box gene MHox is required for early events of skeletogenesis in multiple lineages. Genes Dev *9*, 1237-1249.

Menke, D.B., Guenther, C., and Kingsley, D.M. (2008). Dual hindlimb control elements in the Tbx4 gene and region-specific control of bone size in vertebrate limbs. Development *135*, 2543-2553.

Mennen, U., Mundlos, S., and Spielmann, M. (2014). The Liebenberg syndrome: in depth analysis of the original family. J Hand Surg Eur Vol *39*, 919-925.

Mifsud, B., Tavares-Cadete, F., Young, A.N., Sugar, R., Schoenfelder, S., Ferreira, L., Wingett, S.W., Andrews, S., Grey, W., Ewels, P.A.*, et al.* (2015). Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. Nat Genet *47*, 598-606.

Minguillon, C., Del Buono, J., and Logan, M.P. (2005). Tbx5 and Tbx4 are not sufficient to determine limb-specific morphologies but have common roles in initiating limb outgrowth. Dev Cell *8*, 75-84.

Minguillon, C., Nishimoto, S., Wood, S., Vendrell, E., Gibson-Brown, J.J., and Logan, M.P. (2012). Hox genes regulate the onset of Tbx5 expression in the forelimb. Development *139*, 3180-3188.

Montavon, T., and Duboule, D. (2012). Landscapes and archipelagos: spatial organization of gene regulation in vertebrates. Trends Cell Biol *22*, 347-354.

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.

Mouse, E.C., Stamatoyannopoulos, J.A., Snyder, M., Hardison, R., Ren, B., Gingeras, T., Gilbert, D.M., Groudine, M., Bender, M., Kaul, R.*, et al.* (2012). An encyclopedia of mouse DNA elements (Mouse ENCODE). Genome Biol *13*, 418.

Mundlos, S. (2000). Skeletal morphogenesis. Methods Mol Biol *136*, 61-70.

Nagy, A., & Nichols, J. (2011). Derivation of murine ES cell lines (Heidelberg: Springer-Verlag).

Naiche, L.A., and Papaioannou, V.E. (2003). Loss of Tbx4 blocks hindlimb development and affects vascularization and fusion of the allantois. Development *130*, 2681-2693.

Nicodemi, M., and Prisco, A. (2009). Thermodynamic pathways to genome spatial organization in the cell nucleus. Biophys J *96*, 2168-2177.

Nishimoto, S., Minguillon, C., Wood, S., and Logan, M.P. (2014). A combination of activation and repression by a colinear Hox code controls forelimb-restricted expression of Tbx5 and reveals Hox protein specificity. PLoS Genet *10*, e1004245.

References

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

Nogales, E., Patel, A.B., and Louder, R.K. (2017). Towards a mechanistic understanding of core promoter recognition from cryo-EM studies of human TFIID. Curr Opin Struct Biol *47*, 60-66.

Nora, E.P., Goloborodko, A., Valton, A.L., Gibcus, J.H., Uebersohn, A., Abdennur, N., Dekker, J., Mirny, L.A., and Bruneau, B.G. (2017). Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. Cell *169*, 930-944 e922.

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

Pennacchio, L.A., Ahituv, N., Moses, A.M., Prabhakar, S., Nobrega, M.A., Shoukry, M., Minovitsky, S., Dubchak, I., Holt, A., Lewis, K.D.*, et al.* (2006). In vivo enhancer analysis of human conserved non-coding sequences. Nature *444*, 499-502.

Petit, F., Sears, K.E., and Ahituv, N. (2017). Limb development: a paradigm of gene regulation. Nat Rev Genet *18*, 245-258.

Phillips-Cremins, J.E., Sauria, M.E., Sanyal, A., Gerasimova, T.I., Lajoie, B.R., Bell, J.S., Ong, C.T., Hookway, T.A., Guo, C., Sun, Y.*, et al.* (2013). Architectural protein subclasses shape 3D organization of genomes during lineage commitment. Cell *153*, 1281-1295.

Pombo, A., and Dillon, N. (2015). Three-dimensional genome architecture: players and mechanisms. Nat Rev Mol Cell Biol *16*, 245-257.

Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S.*, et al.* (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665-1680.

Rao, S.S.P., Huang, S.C., Glenn St Hilaire, B., Engreitz, J.M., Perez, E.M., Kieffer-Kwon, K.R., Sanborn, A.L., Johnstone, S.E., Bascom, G.D., Bochkov, I.D.*, et al.* (2017). Cohesin Loss Eliminates All Loop Domains. Cell *171*, 305-320 e324.

Reimchen, T.E. (1983). Structural Relationships between Spines and Lateral Plates in Threespine Stickleback (Gasterosteus Aculeatus). Evolution *37*, 931-946.

Rice, M.C., and O'Brien, S.J. (1980). Genetic variance of laboratory outbred Swiss mice. Nature *283*, 157-161.

Robertson, E.J. (1987). Teratocarcinomas and embryonic stem cells: A practical approach. , Vol 3 (Elsevier).

Rowley, M.J., Nichols, M.H., Lyu, X., Ando-Kuri, M., Rivera, I.S.M., Hermetz, K., Wang, P., Ruan, Y., and Corces, V.G. (2017). Evolutionarily Conserved Principles Predict 3D Chromatin Organization. Mol Cell *67*, 837-852 e837.

Ruf, S., Symmons, O., Uslu, V.V., Dolle, D., Hot, C., Ettwiller, L., and Spitz, F. (2011). Large-scale analysis of the regulatory architecture of the mouse genome with a transposon-associated sensor. Nat Genet *43*, 379-386.

Sambrook, J., & Russel, D. W. (2001). Molecular Cloning - A Laboratory Manual Third edn (Cold Spring Harbor Laboratory Press.).

Sanborn, A.L., Rao, S.S., Huang, S.C., Durand, N.C., Huntley, M.H., Jewett, A.I., Bochkov, I.D., Chinnappan, D., Cutkosky, A., Li, J.*, et al.* (2015). Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. Proc Natl Acad Sci U S A *112*, E6456-6465.

References

Schmitt, A.D., Hu, M., and Ren, B. (2016). Genome-wide mapping and analysis of chromosome architecture. Nat Rev Mol Cell Biol *17*, 743-755.

Schoenfelder, S., Furlan-Magaril, M., Mifsud, B., Tavares-Cadete, F., Sugar, R., Javierre, B.M., Nagano, T., Katsman, Y., Sakthidevi, M., Wingett, S.W.*, et al.* (2015). The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. Genome Res *25*, 582-597.

Schwarzer, W., Abdennur, N., Goloborodko, A., Pekowska, A., Fudenberg, G., Loe-Mie, Y., Fonseca, N.A., Huber, W., Haering, C., Mirny, L., Spitz, F. (2017). Two independent modes of chromatin organization revealed by cohesin removal. Nature.

Sekine, K., Ohuchi, H., Fujiwara, M., Yamasaki, M., Yoshizawa, T., Sato, T., Yagishita, N., Matsui, D., Koga, Y., Itoh, N.*, et al.* (1999). Fgf10 is essential for limb and lung formation. Nat Genet *21*, 138-141.

Seoighe, D.M., Gadancheva, V., Regan, R., McDaid, J., Brenner, C., Ennis, S., Betts, D.R., Eadie, P.A., and Lynch, S.A. (2014). A chromosomal 5q31.1 gain involving PITX1 causes Liebenberg syndrome. Am J Med Genet A *164A*, 2958-2960.

Shapiro, M.D., Marks, M.E., Peichel, C.L., Blackman, B.K., Nereng, K.S., Jonsson, B., Schluter, D., and Kingsley, D.M. (2004). Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. Nature *428*, 717-723.

Shubin, N.H., Daeschler, E.B., and Jenkins, F.A., Jr. (2006). The pectoral fin of Tiktaalik roseae and the origin of the tetrapod limb. Nature *440*, 764-771.

Solomon, M.J., and Varshavsky, A. (1985). Formaldehyde-mediated DNA-protein crosslinking: a probe for in vivo chromatin structures. Proc Natl Acad Sci U S A *82*, 6470-6474.

Spielmann, M., Brancati, F., Krawitz, P.M., Robinson, P.N., Ibrahim, D.M., Franke, M., Hecht, J., Lohan, S., Dathe, K., Nardone, A.M.*, et al.* (2012). Homeotic arm-to-leg transformation associated with genomic rearrangements at the PITX1 locus. Am J Hum Genet *91*, 629-635.

Spielmann, M., and Mundlos, S. (2013). Structural variations, the regulatory landscape of the genome and their alteration in human disease. Bioessays *35*, 533-543.

Spitz, F., and Furlong, E.E. (2012). Transcription factors: from enhancer binding to developmental control. Nat Rev Genet *13*, 613-626.

Splinter, E., de Wit, E., van de Werken, H.J., Klous, P., and de Laat, W. (2012). Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: from fixation to computation. Methods *58*, 221-230.

Symmons, O., Pan, L., Remeseiro, S., Aktas, T., Klein, F., Huber, W., and Spitz, F. (2016). The Shh Topological Domain Facilitates the Action of Remote Enhancers by Reducing the Effects of Genomic Distances. Dev Cell *39*, 529-543.

Symmons, O., Uslu, V.V., Tsujimura, T., Ruf, S., Nassari, S., Schwarzer, W., Ettwiller, L., and Spitz, F. (2014). Functional and topological characteristics of mammalian regulatory domains. Genome Res *24*, 390-400.

Szeto, D.P., Rodriguez-Esteban, C., Ryan, A.K., O'Connell, S.M., Liu, F., Kioussi, C., Gleiberman, A.S., Izpisua-Belmonte, J.C., and Rosenfeld, M.G. (1999). Role of the Bicoid-related homeodomain factor Pitx1 in specifying hindlimb morphogenesis and pituitary development. Genes Dev *13*, 484-494.

Szeto, D.P., Ryan, A.K., O'Connell, S.M., and Rosenfeld, M.G. (1996). P-OTX: a PIT-1-interacting homeodomain factor expressed during anterior pituitary gland development. Proc Natl Acad Sci U S A *93*, 7706-7710.

References

Thomas, M.C., and Chiang, C.M. (2006). The general transcription machinery and general cofactors. Crit Rev Biochem Mol Biol *41*, 105-178.

Tiberio, G., Diglio, M.C., Graziani, M., Testa, F., and Giannotti, A. (2000). Liebenberg syndrome: brachydactyly with joint dysplasia (MIM 186550): a second family. J Med Genet *37*, 548-551.

Tucker, K.L., Wang, Y., Dausman, J., and Jaenisch, R. (1997). A transgenic mouse strain expressing four drug-selectable marker genes. Nucleic Acids Res *25*, 3745-3746.

van Berkum, N.L., Lieberman-Aiden, E., Williams, L., Imakaev, M., Gnirke, A., Mirny, L.A., Dekker, J., and Lander, E.S. (2010). Hi-C: a method to study the three-dimensional architecture of genomes. J Vis Exp.

van de Werken, H.J., Landan, G., Holwerda, S.J., Hoichman, M., Klous, P., Chachik, R., Splinter, E., Valdes-Quezada, C., Oz, Y., Bouwman, B.A.*, et al.* (2012). Robust 4C-seq data analysis to screen for regulatory DNA interactions. Nat Methods *9*, 969-972.

Vietri Rudan, M., Barrington, C., Henderson, S., Ernst, C., Odom, D.T., Tanay, A., and Hadjur, S. (2015). Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. Cell Rep *10*, 1297-1309.

Visel, A., Blow, M.J., Li, Z., Zhang, T., Akiyama, J.A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F.*, et al.* (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature *457*, 854-858.

Visel, A., Minovitsky, S., Dubchak, I., and Pennacchio, L.A. (2007). VISTA Enhancer Browser--a database of tissue-specific human enhancers. Nucleic Acids Res *35*, D88-92.

Vokes, S.A., Ji, H., Wong, W.H., and McMahon, A.P. (2008). A genome-scale analysis of the cis-regulatory circuitry underlying sonic hedgehog-mediated patterning of the mammalian limb. Genes Dev *22*, 2651-2663.

Wakimoto, B.T., and Kaufman, T.C. (1981). Analysis of larval segmentation in lethal genotypes associated with the antennapedia gene complex in Drosophila melanogaster. Dev Biol *81*, 51-64.

Wasserman, P.M., & Soriano, P. M. (2010). Guide to techniques in mouse development.

Weischenfeldt, J., Dubash, T., Drainas, A.P., Mardin, B.R., Chen, Y., Stutz, A.M., Waszak, S.M., Bosco, G., Halvorsen, A.R., Raeder, B.*, et al.* (2017). Pan-cancer analysis of somatic copy-number alterations implicates IRS4 and IGF2 in enhancer hijacking. Nat Genet *49*, 65-74.

Wellik, D.M., and Capecchi, M.R. (2003). Hox10 and Hox11 genes are required to globally pattern the mammalian skeleton. Science *301*, 363-367.

Will, A.J., Cova, G., Osterwalder, M., Chan, W.L., Wittler, L., Brieske, N., Heinrich, V., de Villartay, J.P., Vingron, M., Klopocki, E.*, et al.* (2017). Composition and dosage of a multipartite enhancer cluster control developmental expression of Ihh (Indian hedgehog). Nat Genet.

Wingett, S., Ewels, P., Furlan-Magaril, M., Nagano, T., Schoenfelder, S., Fraser, P., and Andrews, S. (2015). HiCUP: pipeline for mapping and processing Hi-C data. F1000Res *4*, 1310.

Xiao, T., Wallace, J., and Felsenfeld, G. (2011). Specific sites in the C terminus of CTCF interact with the SA2 subunit of the cohesin complex and are required for cohesin-dependent insulation activity. Mol Cell Biol *31*, 2174-2183.

Xu, X., Weinstein, M., Li, C., Naski, M., Cohen, R.I., Ornitz, D.M., Leder, P., and Deng, C. (1998). Fibroblast growth factor receptor 2 (FGFR2)-mediated reciprocal regulation

loop between FGF8 and FGF10 is essential for limb induction. Development *125*, 753-765.

Yamanoue, Y., Setiamarga, D.H., and Matsuura, K. (2010). Pelvic fins in teleosts: structure, function and evolution. J Fish Biol *77*, 1173-1208.

Zabidi, M.A., Arnold, C.D., Schernhuber, K., Pagani, M., Rath, M., Frank, O., and Stark, A. (2015). Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. Nature *518*, 556-559.

Zabidi, M.A., and Stark, A. (2016). Regulatory Enhancer-Core-Promoter Communication via Transcription Factors and Cofactors. Trends Genet *32*, 801-814.

Zeller, R., Lopez-Rios, J., and Zuniga, A. (2009). Vertebrate limb bud development: moving towards integrative analysis of organogenesis. Nat Rev Genet *10*, 845-858.

Zuniga, A. (2015). Next generation limb development and evolution: old questions, new perspectives. Development *142*, 3810-3820.

10. APPENDIX

10.1 Vector maps

PX459
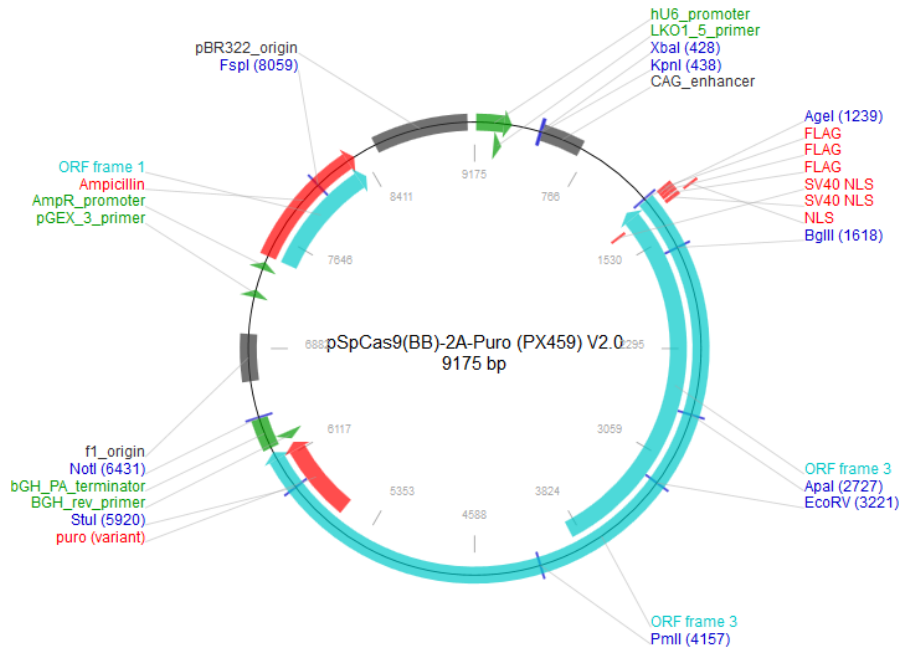
Flippase
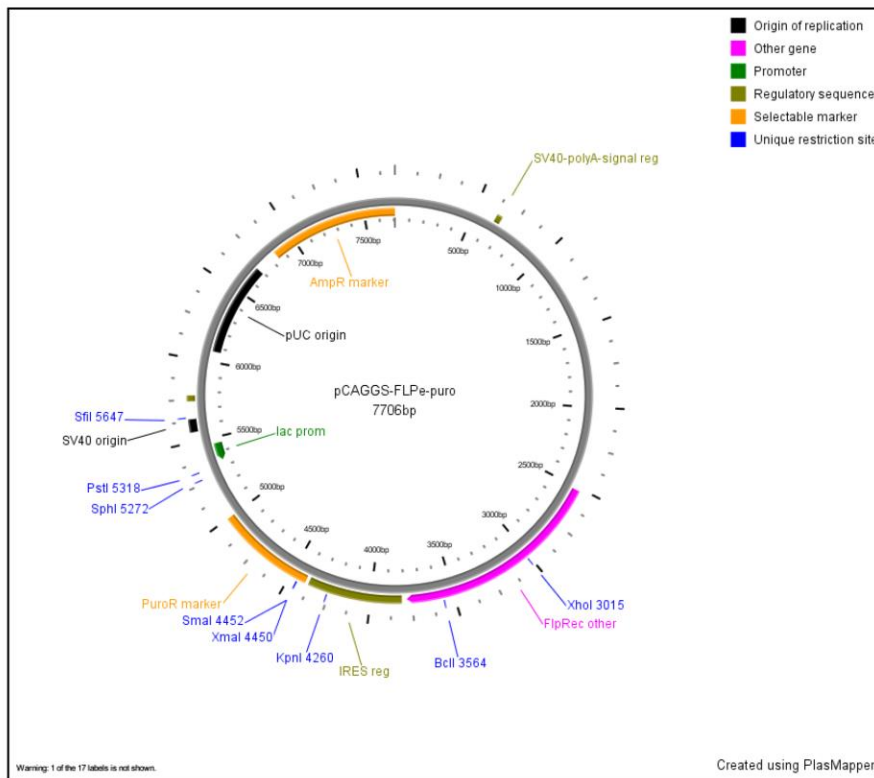
Appendix

Pgk-ATG-frt
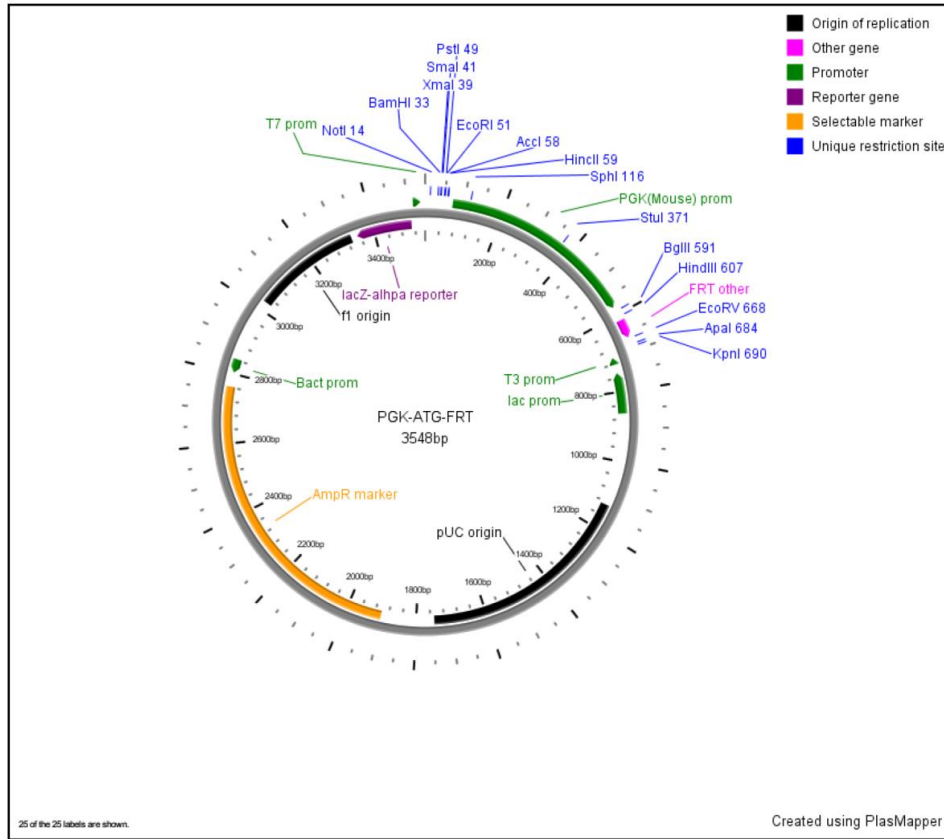
pBluescript-LacZ
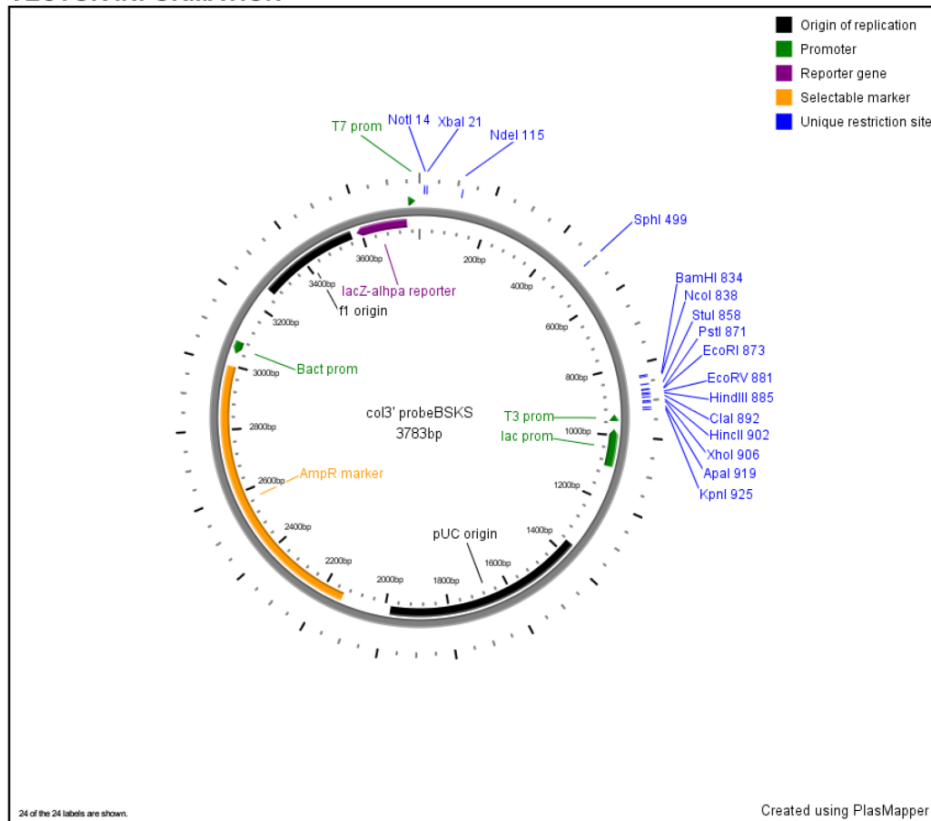
10.2 Supplementary figures
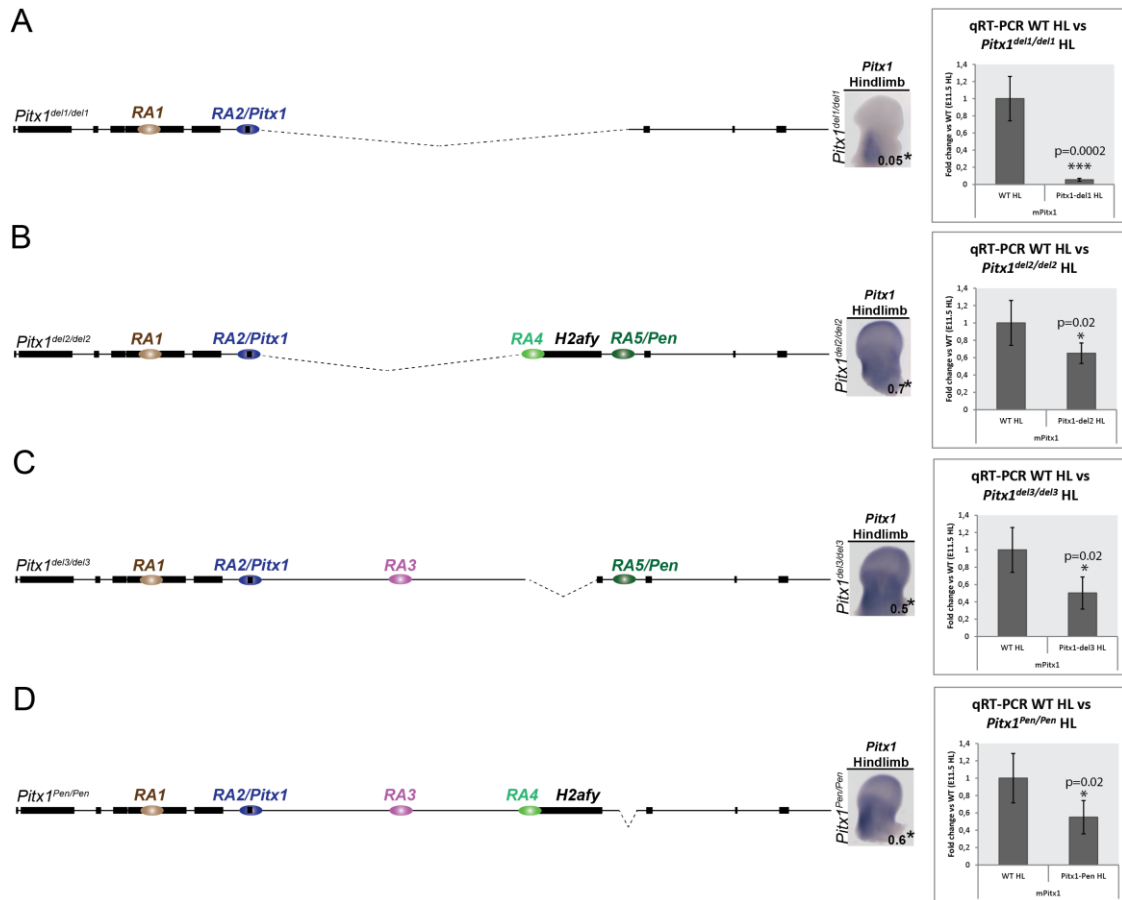
Figure 22 CRISPR-Cas9 alleles dissecting *Pitx1* regulatory landscape and quantification of gene expression

A. CRISVar deletion *Pitx1^{del1/del1}* deleting 330 kb upstream of *Pitx1*. Mutants show almost complete loss of *Pitx1* expression in E11.5 hindlimb as seen by WISH (picture) and qRT-PCR (p-value=0.0002, n=4). **B.** CRISVar deletion *Pitx1^{del2/del2}* deleting 234 kb of gene desert upstream of *Pitx1*. Mutants show 30% loss of *Pitx1* expression in E11.5 hindlimb as seen by WISH (picture) and qRT-PCR (p-value=0.02, n=4). **C.** CRISVar deletion *Pitx1^{del3/del3}* deleting 58 kb of *H2afy* gene body. Mutants show 50% loss of *Pitx1* expression in E11.5 hindlimb as seen by WISH (picture) and qRT-PCR (p-value=0.02, n=4). **D.** CRISVar deletion *Pitx1^{Pen/Pen}* deleting 1.2 kb *Pen* enhancer region. Mutants show 40% loss of *Pitx1* expression in E11.5 hindlimbs as seen by WISH (picture) and qRT-PCR (p-value=0.02, n=5).
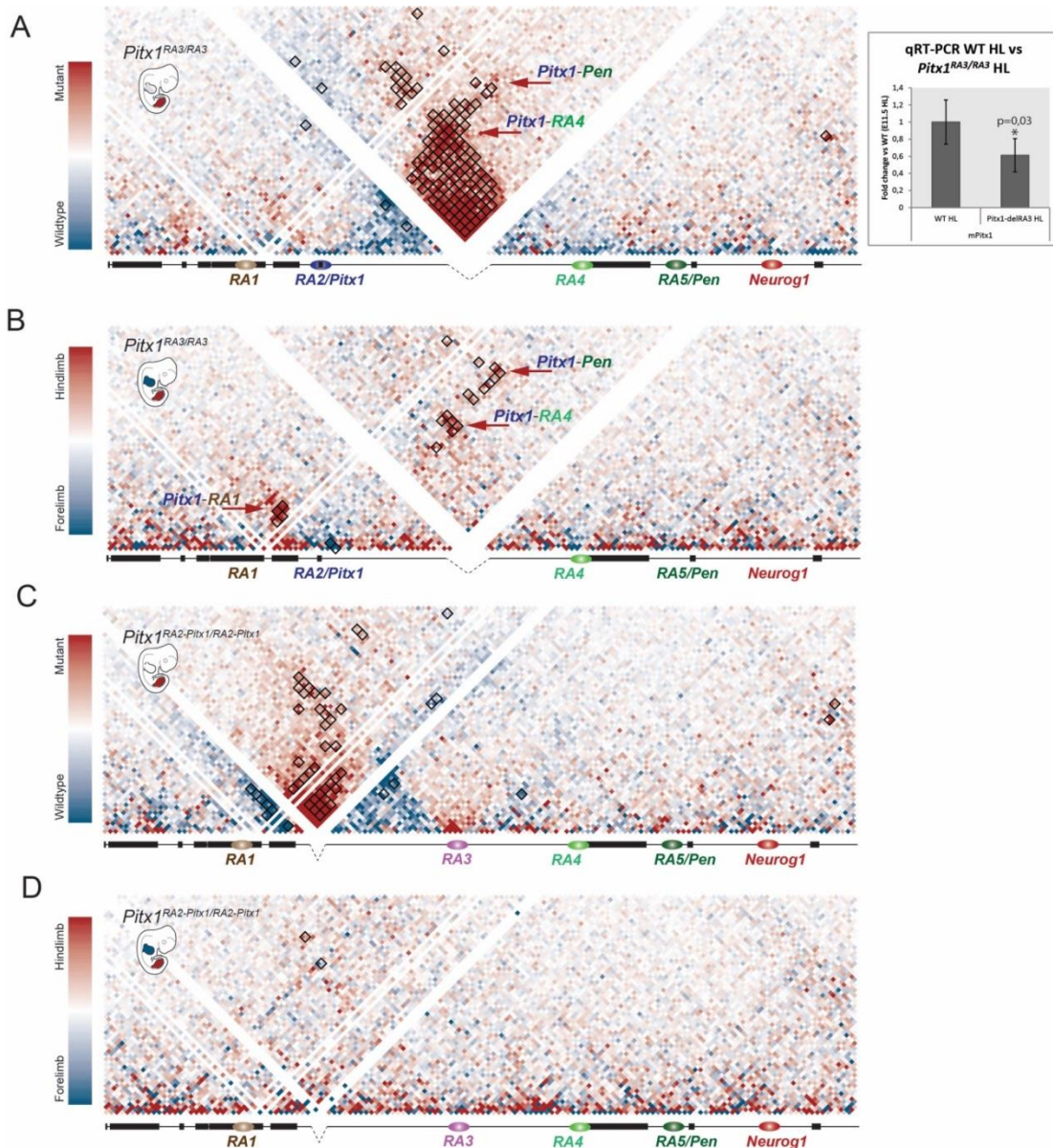
Figure 23 cHi-C subtraction of mutants disrupting *Pitx1* regulatory anchors

A. cHi-C subtraction between wildtype and *Pitx1*^{RA3/RA3} mutant hindlimb tissue at E11.5. Chromatin interactions more prevalent in mutant or wildtype hindlimbs tissues are shown in red and blue, respectively. Significant changes are highlighted in black boxes (FDR=0.05). Interactions significantly increased between regulatory anchors are indicated with red arrows (*Pitx1-RA4, Pitx1-RA5/Pen* interactions). *Pitx1*^{RA3/RA3} embryos display a 30% reduction of *Pitx1* expression in E11.5 hindlimbs as quantified by qRT-PCR (p=0.03, n=4). **B.** cHi-C subtraction between *Pitx1*^{RA3/RA3} mutant forelimb and *Pitx1*^{RA3/RA3} mutant hindlimb tissues at E11.5. Chromatin interactions more prevalent in mutant hindlimb or mutant forelimb tissues are shown in red or blue, respectively. Significant changes are highlighted in black boxes (FDR=0.05). Interactions significantly increased between regulatory anchors are indicated with red arrows (*Pitx1-RA1, Pitx1-RA4, Pitx1-RA5/Pen* interactions). **C.** cHi-C subtraction between wildtype and *Pitx1*^{RA2-Pitx1/RA2-Pitx1} mutant hindlimb tissues at E11.5. Chromatin interactions more prevalent in mutant or wildtype hindlimbs tissues are shown in red or blue, respectively. Significant changes are highlighted in black boxes (FDR=0.05). **D.** cHi-C subtraction between *Pitx1*^{RA2-Pitx1/RA2-Pitx1} mutant forelimb and *Pitx1*^{RA2-Pitx1/RA2-Pitx1} mutant hindlimb tissues at E11.5. Chromatin interactions more prevalent in mutant hindlimb or mutant forelimb tissues are shown in red or blue, respectively. Significant interactions are highlighted in black boxes (FDR=0.05).

114
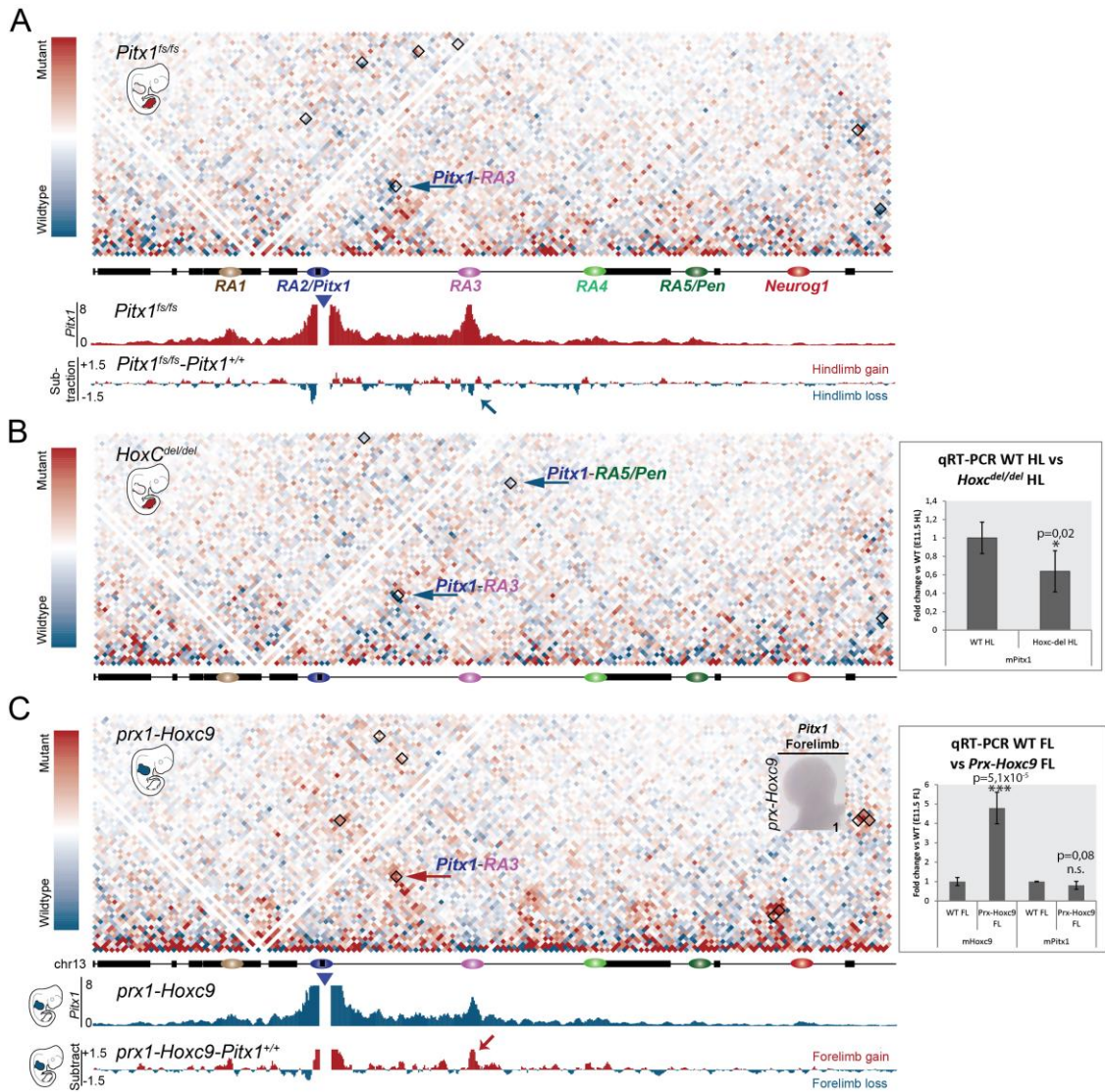
Figure 24 Trans-acting factors regulating *Pitx1* in hindlimbs

A. cHi-C subtraction between wildtype and *Pitx1^{fs/fs}* mutant hindlimb tissue at E11.5. Chromatin interactions more prevalent in mutant or wildtype hindlimbs tissues are shown in red or blue, respectively. Significant change in interactions are highlighted in black boxes (FDR=0.05). Interactions significantly reduced between regulatory anchors are indicated with a blue arrow (*Pitx1- RA3* interaction). Derived viewpoints from cHiC map, vC, are shown using the *Pitx1* viewpoint are shown in red. Below is the subtraction track between wildtype and mutant hindlimb tissues using the respective viewpoint. **B.** cHi-C subtraction between wildtype and *Hoxc^{del/del}* mutant hindlimb tissues at E11.5. Chromatin interactions more prevalent in mutant or wildtype hindlimbs tissues are shown in red or blue, respectively. Significant changes are highlighted in black boxes (FDR=0.05). Interactions significantly reduced between regulatory anchors are indicated with blue arrows (*Pitx1-RA3* and *Pitx1-RA5/Pen*). *Hoxc^{del/del}* embryos display a 40% reduction of *Pitx1* expression in E11.5 hindlimbs as quantified by qRT-PCR (p=0.02, n=4). **C.** cHi-C subtraction between wildtype and *prx-Hoxc9* mutant forelimb tissue at E11.5. Chromatin interactions more prevalent in mutant or wildtype forelimb tissues are shown in red or blue, respectively. Significant changes are highlighted in black boxes (FDR=0.05). The significantly increased chromatin interactions between regulatory anchors are indicated with a red arrow (*Pitx1-RA3*). vC using *Pitx1* viewpoint is shown in blue. Below is the subtraction track between wildtype and mutant forelimb tissue using the respective viewpoint. *prx-Hoxc9* E11.5 embryos display a 5-fold increase in *Hoxc9* expression, as quantified by qRT-PCR (p=5.1x10^{-5}, n=4). However, *Pitx1* is not induced in forelimb tissues as shown by WISH (picture) and quantified by qRT-PCR (p-value=n.s).
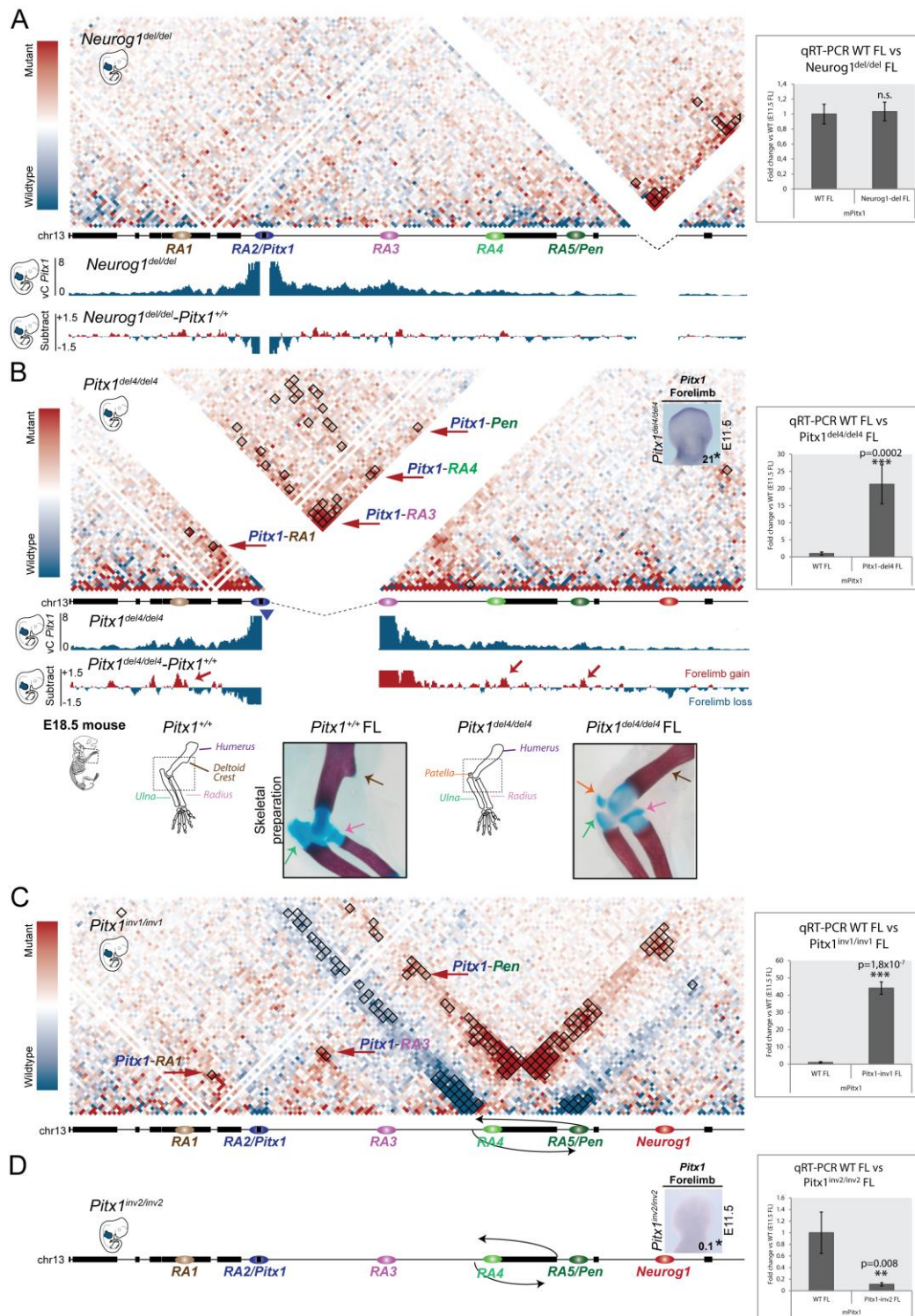
Figure 25 CRISPR-Cas9 induced structural variants cause activation of *Pitx1* in forelimbs

A. cHi-C subtraction between wildtype and *Neurog1del/del* mutant forelimb tissue at E11.5. Chromatin interactions more prevalent in mutant or wildtype forelimbs tissues are shown in red or blue, respectively. Significant changes are highlighted in black boxes (FDR=0.05). Derived vC from the *Pitx1* viewpoint are shown in blue. Below is the subtraction track between wildtype and mutant forelimb tissue using the respective viewpoint. *Neurog1del/del* embryos do not show changes in *Pitx1* expression in E11.5 forelimbs as seen in WISH (picture of forelimb) and quantified by qRT-PCR (p=n.s, n=6). **B.** cHi-C subtraction between wildtype and *Pitx1del4/del4* mutant forelimb tissue at E11.5. Chromatin interactions more prevalent in mutant or wildtype forelimbs tissues are shown in red and blue, respectively. Significant changes are highlighted in black boxes (FDR=0.05). The significantly increased chromatin interactions between regulatory anchors are indicated with red arrows (*Pitx1-RA3*, *Pitx1-RA4* and *Pitx1-*

Pen interactions). vC using *Pitx1* viewpoint is shown in blue. Below is the subtraction track between wildtype and mutant forelimb tissue using the same *Pitx1* viewpoint. *Pitx1*$^{del4/del4}$ E11.5 embryos show an ectopic expression of *Pitx1* in forelimb tissue and qRT-PCR shows a 21-fold increase in *Pitx1* expression in forelimb tissue (p=0.00002, n=4). Alcian blue and alizarin red skeletal staining of *Pitx1*$^{del4/del4}$ arms at E18.5 show a *Pitx1* gain of function related phenotype with a partial elbow-to-knee joint transformation: a loss of the deltoid crest (brown arrow), reduction of olecranon (green arrow), broadening of the radius (pink arrow), and appearance of an ectopic patella (orange arrow). **C.** cHi-C subtraction between wildtype and *Pitx1*$^{inv1/inv1}$ mutant forelimb tissue at E11.5. Chromatin interactions more prevalent in mutant or wildtype forelimbs tissues are shown in red or blue, respectively. Significant changes are highlighted in black boxes (FDR=0.05). Significantly increased Interactions between regulatory anchors are indicated with red arrows (*Pitx1-RA1*, *Pitx1-RA3*, *Pitx1-Pen* interaction). *Pitx1*$^{inv1/inv1}$ embryos display a 46-fold increase in *Pitx1* expression in E11.5 forelimbs as quantified by qRT-PCR (p= 1.8x10^{-7}, n=4). **D.** *Pitx1*$^{inv2/inv2}$ control allele: E11.5 embryos show gain of *Pitx1* expression in forelimb tissue as seen in WISH and quantified by qRT-PCR.
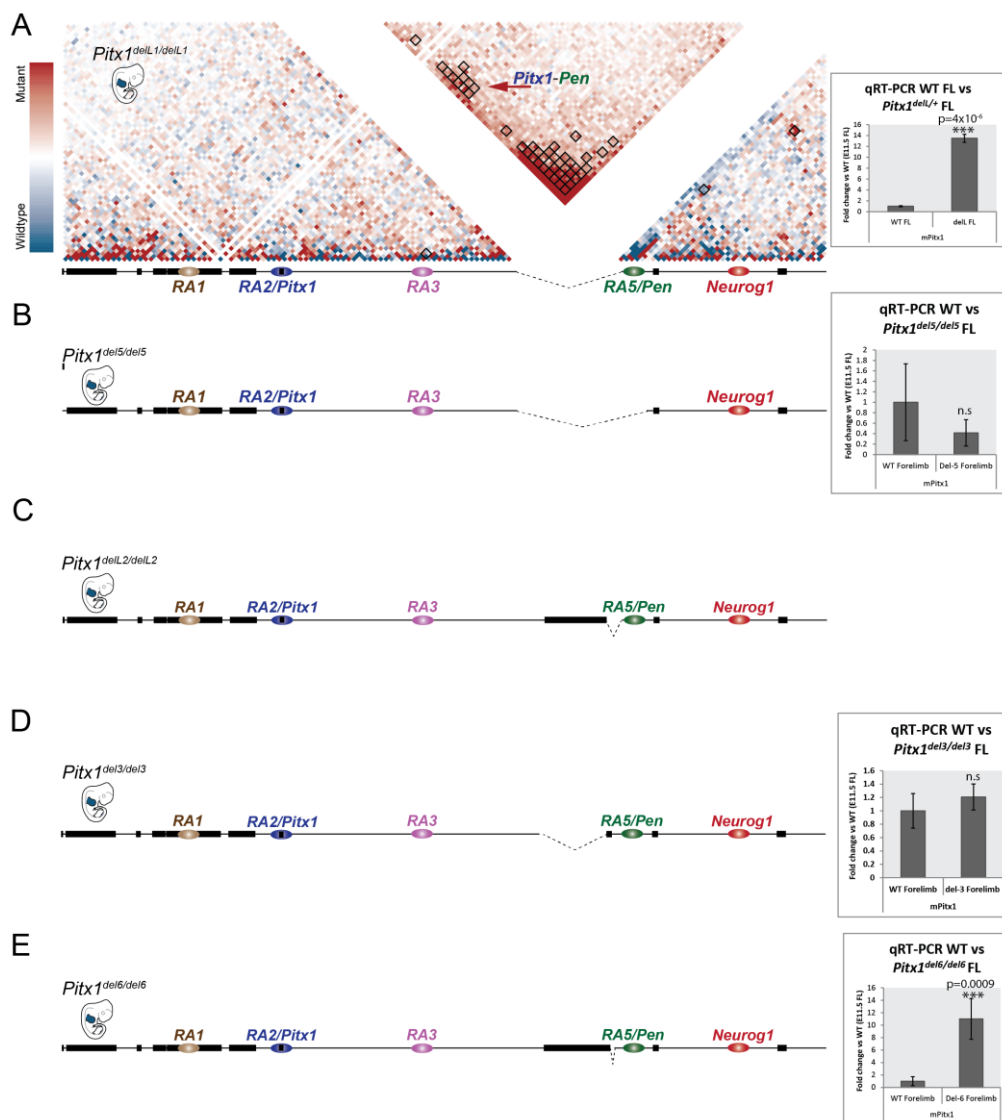
Figure 26 *H2afy* promoter insulates *Pen* activity in forelimbs

A. cHi-C subtraction between wildtype and *Pitx1*$^{delL1/delL1}$ mutant forelimb tissue, recapitulating Liebenberg deletion in mice, at E11.5 is shown. Chromatin interactions more prevalent in mutant or wildtype forelimbs tissues are shown in red or blue, respectively. Significant changes are highlighted in black boxes (FDR=0.05). On the right, qRT-PCR quantification of *Pitx1* expression in E11.5 mutant forelimb tissue (p-value=4x10^{-6}) is shown. **B.** *Pitx1*$^{del5/del5}$ qRT-PCR in mutant forelimbs. **C.** *Pitx1*$^{delL2/delL2}$ qRT-PCR in mutant forelimbs. **D.** *Pitx1*$^{del3/del3}$ qRT-PCR in mutant forelimbs. **E.** *Pitx1*$^{del6/del6}$ qRT-PCR in mutant forelimbs.

Appendix

10.3 List of Abbreviations

°C	degrees celcius
μ	micro (prefix)
3C	chromatin conformation capture
4C	circular chromatin conformation capture
AER	Apical ectodermal ridge
bp	base pair
cC	capture-c
cDNA	coding DNA
cHi-C	capture Hi-C
ChIP	Chromatin immoprecipitation
Chr	Chromosome
CTCF	CCCTC-binding factor
del	deletion
DEPC	diethylpyrocarbonate
DIG	digoxygenin
DMEM	Dulbeccos's modified eagle's medium
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide
E	embryonic stage
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
ENCODE	Encyclopedia of DNA Elements
ESC	embryonic stem cell
FL	forelimb
g	gram
GTF	general transcription factors
het	heterozygous
HL	hindlimb
hom	homozygous
kb	kilobases
l	liter
m	milli (prefix)
MAD	multi anchor domain
Mb	mega bases
mESCs	mouse embryonic stem cells
min	minutes
mm	*mus musculus*
mol	moles
mRNA	messenger RNA
n	nano (prefix)
o.n.	over night
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PFA	paraformaldehyde
pfu	*Pyrococcus furiosus*
PIC	preinitiation complex

Appendix

qRT-PCR	quantitative real time PCR
RNA	ribonucleic acid
RPM	rounds per minute
RT	room temperature
SSC	saline sodium citrate buffer
TAD	topologically associating domain
taq	*Thermus aquaticus*
TF	transcription factor
TFBS	transcription factor binding site
Tm	melting temperature
TSS	transcriptional start site
U	units
UCSC	University of California, Santa Cruz
vC	virtual capture-c
Vol	volume
v/v	volume per volume
w/v	weight per volume
wt	wild type
ZRS	zone of polarising activity

10.4 List of Tables

10.5 List of Figures

Appendix

11. DECLARATION OF INDEPENDENT WORK

I hereby declare the present work has been independently conceived and written, and that no technical aid has been used. I assure that this work or any part of it has not been submitted to, approved, or rejected by any other academic institution.

Berlin, 26.10.2017 --------------------------------

Bjørt Katrinardóttir Kragesteen

Scientific publications

12. SCIENTIFIC PUBLICATIONS

In submission

Kragesteen B.K., Spielmann M., Paliou C., Heinrich V., Schöpflin R., Esposito A., Annunziatella C., Bianco S., Chiariello A.M., Jerković I., Harabula I., Guckelberger P., Pechstein M., Wittler L., Chan W.L., Franke M., Lupiáñez D.G., Kraft K., Timmermann B., Vingron M., Visel A., Nicodemi M., Mundlos S. and Andrey G. Dynamic 3D Chromatin Architecture Determines Enhancer Specificity and Morphogenetic Identity in Limb Development.

Schreiber I., Ott C.E., **Kragesteen B.K.**, Schanze N., Lee C.T., Dörpholz G., Köhrle J., Mundlos S., Ruschke K. and Knaus P. BMPs as new insulin sensitizers: enhanced glucose uptake in mature 3T3-L1 adipocytes via PPARγ and GLUT4 upregulation.

Peer reviewed

Flöttmann R[*]., **Kragesteen B.K.**[*], Geuer S., Socha M., Allou L., Sowińska-Seidler A., Bosquillon de Jarcy L., Wagner J., Jamsheer A., Oehl-Jaschkowitz B., Wittler L., de Silva D., Kurth I., Maya I., Santos-Simarro F., Hülsemann W., Klopocki E., Mountford R., Fryer A., Borck G., Horn D., Lapunzina P., Wilson M., Mascrez B., Duboule D., Mundlos S. and Spielmann M. Noncoding copy-number variations are associated with congenital limb malformation. *Genet Med.* doi:10.1038/gim.2017.154 (2017)
[*] These authors contributed equally

Spielmann M., Hernandez-Miranda L.R., Ceccherini I., Weese-Mayer D.E., **Kragesteen B.K.**, Harabula I., Krawitz P., Birchmeier C., Leonard N., Mundlos S. Mutations in *MYO1H* cause a recessive form of central hypoventilation with autonomic dysfunction. *J Med Genet.* doi:10.1136/jmedgenet-2017-104765 (2017)

Kraft K., Geuer S., Will A.J., Chan W.L., Paliou C., Borschiwer M., Harabula I., Wittler L., Franke M., Ibrahim D.M., **Kragesteen B.K.**, Spielmann M., Mundlos S., Lupiáñez D.G., Andrey G. Deletions, Inversions, Duplications: Engineering of Structural Variants using CRISPR/Cas in Mice. *Cell Reports.* doi: 10.1016/j.celrep.2015.01.016. (2015)