Aus dem Institut für Medizinische Genetik und Humangenetik
der Medizinischen Fakultät Charité – Universitätsmedizin Berlin
und dem Max Planck Institut für Molekulare Genetik, Berlin

DISSERTATION

Effects of CTCF binding site deletions on genome architecture and gene expression in the *Epha4* locus

zur Erlangung des akademischen Grades
Doctor medicinae (Dr. med.)

vorgelegt der Medizinischen Fakultät
Charité – Universitätsmedizin Berlin

von

Niklas Gerhards

aus Berlin

Datum der Promotion: 25. November 2022

Index of abbreviations

Δ	deletion
3C	chromosome conformation capture
4C	circular chromosome conformation capture
ATP	adenosine triphosphate
B	boundary
Bid	bidest
Bp	base pair
Cas9	CRISPR associated protein 9
CDNA	complementary DNA
Cen	centromeric
CHiC	Capture HiC
ChIP	chromatin immunoprecipitation
Chr	chromosome
CNV	copy number variation
CRE	cis-regulatory element
CRISPR	clustered regulatory short interspaced palindromic repeats
CTCF	CCCTC-binding factor
DNA	deoxyribonucleic acid
DMSO	dimethyl sulfoxide
DNase	deoxyribonuclease
DSB	double-strand break
E	enhancer
E.Coli	escherichia coli
E11.5	embryonic stage 11.5
E17.5	embryonic stage 17.5
EDTA	ethylenediaminetetraacetic acid
EphA4	ephrin type-A receptor 4 gene
ESC	embryonic stem cell
F1/F2	forward primer 1/2
FA	formaldehyde
FBS	fetal bovine serum
FL	forelimb
FM	freezing media
GAPDH	glyceraldehyde 3-phosphate dehydrogenase
GRNA	guide RNA
H3K4me1	monomethylation on lysine 4 of histone 3
H3K4me3	trimethylation on lysine 4 of histone 3
H3K27ac	acetylation on the 27th lysine of histone 3
H3K27me	trimethylation on the 27th lysine of histone 3
Het	heterozygous
HL	hindlimb
Hom	homozygous
Kb	kilobase

LIF	leukemia inhibitory factor
Mb	megabase
NCBI	National Center for Biotechnology Information
NHEJ	non-homologous end joining
NM	normal media
ns	not significant
na	not applicable
P	promoter
PAM	protospacer adjacent motif
Pax3	Paired box 3
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PFA	paraformaldehyde
PNK	polynucleotide kinase
qPCR	quantitative PCR
R1/R2	reverse primer 1/2
RefSeq	reference sequence (database by NCBI)
RNA	ribonucleic acid
RNase	ribonuclease
RNAPII	RNA polymerase II
Rpm	rounds per minute
RT	room temperature
SDS	sodium dodecyl sulphate
Seq	sequencing
SHH/Shh	sonic hedgehog
SV	structural variation
TAD	topologically associating domain
TAE	Tris-acetate EDTA
Tel	telomeric
TF	transcription factor
TRE	trans-regulatory element
TSS	transcription start site
UCSC	University of California Santa Cruz (genome browser)
Wt	wildtype

Index of Figures

Index of Tables

Abstract

Enhancers and promoters are DNA sequences whose physical proximity, through chromatin looping, is a necessary condition to initiate gene transcription. Functional chromatin interactions are limited by topologically associating domains (TADs), structural regulatory units of the genome that constrain enhancer-promoter crosstalk. Recent studies have elucidated that CCCTC-binding factor (CTCF) regulates TAD and loop formation. However, it remains unclear whether CTCF is indispensable to mediate enhancer-promoter interactions and to achieve precise spatiotemporal patterns of gene expression during development.

To investigate this, we systematically deleted CTCF-binding motifs at the *Epha4* locus in mice using CRISPR-Cas9 and evaluated the effects on chromatin organization and gene expression *in vivo*. We focused on two CTCF sites associated to an enhancer-promoter interaction that occurs during mouse limb development. Analysis of genome architecture showed that CTCF binding site deletions induce a complete depletion of enhancer-promoter interactions. Analysis of *Epha4* expression levels revealed that while the enhancer-promoter loop is disrupted, up to 50% of physiological expression levels are retained.

Our results suggest that the baseline proximity generated by TADs can be sufficient to fail-safe expressional output during development. Previous research on TADs has stressed their insulating properties - highlighting their function of restricting enhancer-promoter contacts over TAD boundaries. Our results stress an additional function of TADs in supporting enhancer-promoter communication and in sustaining developmental gene expression.

Zusammenfassung

Enhancer und Promotoren sind DNA-Sequenzen, deren physische Nähe durch Chromatin-Schleifenbildung eine notwendige Voraussetzung für die Einleitung von Gentranskription ist. Solche Enhancer-Promoter Interaktionen werden durch topologisch assoziierende Domänen (TADs) eingeschränkt. Neuere Studien haben gezeigt, dass der CCCTC-Bindungsfaktor (CTCF) die TAD und Chromatin-Schleifenbildung reguliert. Es bleibt jedoch unklar, ob CTCF unverzichtbar ist, um Enhancer-Promoter-Interaktionen zu vermitteln und präzise räumlich-zeitliche Muster von Genexpression während der Entwicklung zu erzielen.

Um dies zu untersuchen, haben wir mit CRISPR-Cas9 systematisch CTCF-Bindungsmotive am *Epha4*-Locus bei Mäusen deletiert und die Auswirkungen auf Genexpression und Chromatin-Architektur *in vivo* ausgewertet. Wir konzentrierten uns auf zwei Bindungsmotive, die mit einer Enhancer-Promoter-Interaktion verbunden sind, welche während der Entwicklung der Extremitäten der Maus auftritt. Die Analyse der Genomarchitektur zeigte, dass die Deletionen der CTCF-Bindungsstelle eine vollständige Erschöpfung der Enhancer-Promoter-Interaktionen induzieren. Die Analyse der *Epha4*-Expressionsniveaus zeigte, dass während die Enhancer-Promoter-Schleife unterbrochen ist, bis zu 50% des physiologischen Expressionsniveaus erhalten bleiben.

Unsere Ergebnisse deuten darauf hin, dass die durch topologisch assoziierte Domänen (TADs) erzeugte Basisfrequenz an Enhancer-Promoter Kontakten ausreicht, um eine bestimmte Menge an expressivem Output zu erzielen. Frühere Forschungen über TADs haben hauptsächlich ihre isolierenden Eigenschaften betont - und ihre Funktion der Einschränkung von Enhancer-Promoter-Kontakten über TAD-Grenzen hinweg hervorgehoben. Unsere Ergebnisse unterstreichen eine zusätzliche Funktion von TADs bei der Unterstützung der Enhancer-Promotor-Kommunikation und bei der Aufrechterhaltung von Genexpression in der Entwicklung.

1. Introduction

1.1 Gene regulation

While most cells in a multicellular organism share the same genomic information encoded in deoxyribonucleic acid (DNA), it is the differential interpretation of this information, in a cell-specific manner, that allows organisms to evolve into their stunning complexity. The capacity to up-regulate and down-regulate expression of a gene in the correct tissue at the correct time, enables organisms to undergo cell specialization. Gene regulation can occur during every step of the process of a coding DNA sequence being converted into a functional protein (Figure 1-1). Transcriptional regulation is set at the very beginning of this process. As this work focused on the mechanisms of transcriptional regulation, the following shall give a broad overview of the key players involved.

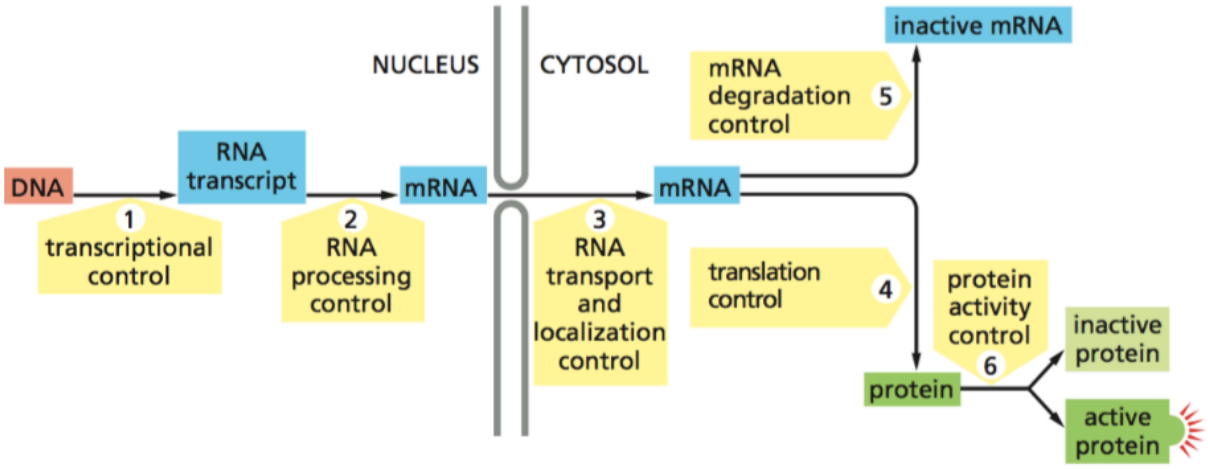

Figure 1-1: Levels of Gene Regulation (Molecular Cell Biology of the Cell, Alberts, pp373, 2016)

1.1.1 Cis-regulatory elements

The human genome is to over 98% made up of non-protein-coding sequences. The notion that these parts of the genome are mere "junk DNA" has been overcome, as their functional significance has been elucidated (1). Cis-regulatory elements (CREs) are non-coding DNA sequences involved in the transcriptional regulation of genes. They were termed cis, based on the latin prefix cis ("on this side"), because regulation and transcription occur within the same DNA molecule. They display a high degree of evolutionary conservation and mutations within their sequences are associated with disease (2), indicating their functional importance. Promotors, enhancers and insulators are all CREs, performing differentiated tasks in a complex network of transcriptional regulation. CREs possess DNA motifs,

functioning as binding sites for trans-regulatory elements (TREs) and it is the interplay of the two, which orchestrates the correct spatiotemporal expression of genes.

1.1.1.1 Promoters

Promoters are DNA sites where transcription is initiated. They are located upstream of their corresponding genes, flanking the 5‘ end of the transcription start site (TSS). Promoters are characterized for containing DNA binding sites for diverse transcription factors (TFs) (see section 1.1.2). Such binding sites may be general (e.g. TATA box) or promoter-specific. Further, they possess binding sites for RNA polymerase II (RNAPII) – the enzyme which in eukaryotes transcribes mRNA from the gene body. The binding of the transcription machinery, including RNAPII, initiates transcription and thus it is the affinity of the transcription machinery to the promoter, which determines the level of transcriptional output. Affinity to the promoter is regulated by TFs and distal enhancer elements. In many cases, the physical interaction of a distal enhancer with its target promoter (the mechanisms of which are discussed in section 1.2.3) is a necessary condition for the transcriptional machinery to be activated (3). In addition, the binding of different TFs may increase or reduce the affinity of RNAPII to the promoter. Therefore it is the interplay of these various mechanisms, which at the promoter is computed into a level of transcriptional output (4). Several promoters can compete for shared distal enhancers, resulting in alterations of gene expression levels and in some cases disease (5, 6).

1.1.1.2 Enhancers

Enhancers are DNA sequences that are capable of increasing the transcriptional response of genes. To achieve this, enhancers must physically interact with their target promoters(7). With few exceptions(8), enhancers exert their function on promoters located on the same chromosome and can therefore be considered as CREs. In most cases, they are referred to as distal regulatory elements, as the linear genomic distance separating them from their target promoters may involve several kilobases (Kbs), or even megabases (Mbs) in extreme cases (9, 10). However, in the current scientific discourse there is no clear definition of when an enhancer is distal or proximal.

The mechanisms directing an enhancer to its target promoter are still under active research. Enhancer-promoter interactions can be facilitated by orchestrated chromatin folding (see section 2.2), thus

establishing looping structures connecting enhancers with their target promoters in a direction independent manner. At the same time enhancers located in close linear genomic distance to promoters seem to be able to activate them, if not constrained by insulators, without coordinated chromatin folding, but simply based on the stochastic probability of interaction given by proximity (11) (7). Enhancers possess transcription factor binding sites and the binding of TFs drives enhancer activity. At promoters, enhancers initiate the binding of further TFs, leading to a decondensation of chromatin, allowing the transcription machinery to bind and thus initiate transcription (12).

Enhancers tend to be highly tissue specific. For example, in the *Sonic hedgehog* (*Shh*) locus several enhancers orchestrate the correct spatiotemporal expression of *Shh* in the brain, but only the singular ZRS enhancer drives *Shh* expression in the limb (10).

Genes that are key to cell specialization and cell identity typically associate with clusters of enhancers, generally referred to as super-enhancers (13). Individual enhancers within a super-enhancer tend to act upon associated promoters in an autonomous and additive form (14). However, also non-additive behavior of enhancers has been demonstrated: Enhancers can multiply each other's activity (15, 16), work together to repress ectopic expression in neighboring regions (17, 18) or operate together in hierarchical systems (19).

Another example of how enhancers can act non-additively is enhancer redundancy. Especially developmental genes regularly associate with both more proximal and more distal enhancers (sometimes termed shadow-enhancers) which exhibit overlapping or identical patterns of activity (20). Studies in *Drosophila* have demonstrated that redundant enhancers are essential in maintaining a stable transcriptional output under variable environmental conditions (17, 21) and may further play a role in fine-tuning gene expression patterns (22). Recently, a series of enhancer deletions in mice demonstrated the importance of redundancy in maintaining phenotypic stability (23). Even though these studies show the importance of enhancers for precise and robust gene expression, other studies have demonstrated that, at least in mammals, enhancers are less conserved and evolve faster than promoters (24). It is thus assumed that quick changes within enhancer-landscapes also enable evolutionary adaptation of gene-expression within shorter time-frames.

1.1.1.3 Insulators

Insulators are defined as CREs that insulate a gene from other genomic elements which can affect its

transcriptional output. To date, two molecular mechanisms have been identified by which insulators can exert their function. Firstly, insulators may act as barriers that can block the condensation and decondensation of chromatin (25). Secondly, and more importantly for this work, insulators can interfere with enhancer-promoter interactions, when located between such elements (25). In mammals, the only known insulator to date is the protein CTCF, which binds frequently across the genome (see section 2.2.4). Intriguingly, CTCF does not strictly function as an insulator, but also mediates proper 3D-chromatin folding. CTCF facilitates enhancer-promoter contacts just as much as it restricts them (26). The most established model explaining this dualistic function of CTCF is the model of loop extrusion (section 2.2.4).

1.1.2 Trans-regulatory elements

Trans-regulatory elements are encoded proteins involved in the regulation of transcription, often referred to as transcription factors. TFs bind to short, 6-12 base pairs (bp), highly conserved sequences of CREs and, as stated above, it is the interplay of these two classes of elements that orchestrates the spatiotemporal expression of genes with precision. TF binding motifs can be found on both enhancers and promoters. Enhancers often possess clusters of many TF binding sites, allowing for the binding of simultaneous TFs and thus an almost unlimited repertoire of combinatorial possibilities (27). The integrated information of a whole network of TFs may determine the output of a single enhancer, therefore enabling the high specificity of spatiotemporal activation of enhancers (28).

The binding of TFs to enhancers can be regulated at different levels (27). Firstly, the expression of TFs can be up- or downregulated, influencing the transcriptional response of enhancer elements. Secondly, the binding affinity of TFs to enhancers can be modulated by other TFs or by the level of chromatin accessibility, which in turn is also regulated by another class of TFs: pioneer factors(29). Pioneer factors bind to enhancers initially, reorganizing chromatin, increasing chromatin accessibility and allowing the recruitment of additional TFs that are required to drive enhancer activity.

1.1.3 Histone modifications and chromatin accessibility

The most basic unit of chromatin architecture in eukaryotic organisms is the nucleosome. It consists of approximately 147 bp of DNA wrapped around an octamer of histone proteins(30). Nucleosomes are set at the core of the compact packaging of DNA, allowing a linear strand of approximately 2m to

fit into a nucleus of only 10μm diameter. Nucleosome packaging is seen as a dynamic system of chromatin accessibility, inseparably linked with transcriptional regulation. The general theory behind this states that histones are competing with TFs for chromatin occupancy (27) - histones reduce chromatin accessibility by packaging it tightly, thus preventing TFs from binding. Indeed, there is evidence that the binding of TFs correlates well with nucleosome depleted regions (27). Further post-translational modifications of histones are crucial for this dynamic system of accessibility.

Specific histone modifications correlate highly with the function of the DNA associated with that same histone. Enhancers correlate with a monomethylation on lysine 4 of histone 3 (H3K4me1); promoters and actively transcribed gene bodies correlate with a trimethylation on lysine 4 of histone 3 (H3K4me3); active enhancers and active promoters show an increased frequency of acetylation on the 27th lysine of histone 3 (H3K27ac) (31, 32) and repressed regions are with increased frequency trimethylated on the 27th lysine of histone 3 (H3K27me3). Enhancers, in particular, exhibit highly variable patterns of histone modification during development and in adult tissues. This is indicative their central and dynamic role in cell specialization (33, 34). However, it remains unclear whether histone modifications are cause or consequence of specific enhancer function. Nevertheless, enhancer marks, as detected by chromatin immunoprecipitation techniques, constitute a powerful tool for the identification of enhancers and other CREs (35).

1.2 Genome architecture

As described above, nucleosomes are a core unit of DNA packaging and, at the same time, histone modifications are linked to DNA accessibility and cis-regulatory functions. Furthermore, enhancer activity is dependent on physical enhancer-promoter interactions. Thus, the mechanisms regulating and building genome architecture can be considered a fundamental aspect of transcriptional regulation.

1.2.1 Chromosome conformation capture methods

In 2002, Dekker et al. published a groundbreaking method termed chromosome conformation capture (3C) (36). Sparked by this initial work, the past 14 years have seen a rapid development of ever more sophisticated, 3C-derived methods (37). Chromosome conformation capture methods are employed to generate frequency and proximity-based interaction maps of DNA-DNA contacts. The approach is

common for all C- methods: nuclei from sample cells are fixed *in vivo*, with formaldehyde. Consequently, all protein-DNA and DNA-DNA interactions, present in a sample of cells at the time, are preserved by cross-linking. In order to evaluate with which frequency interactions occur, DNA is cut using a restriction enzyme and fragment ends are re-ligated, creating hybrid molecules of interacting genomic regions. At this point, the information about any 3D interactions is stored in a 2D library; the three-dimensionally interacting pair of DNA sequences is now aligned on a single-DNA strand.

To read out the information stored within this library, the frequency of ligation events of non-neighboring DNA sequences has to be determined. In the original 3C approach, this is done by designing two site specific primers, PCR amplifying and quantifying the result. Therefore, 3C only yields results about the frequency of interaction of two chosen viewpoints at a time. In combination with high-throughput sequencing, the original "one-to-one" approach was improved to allow for the generation of high resolution, high-throughput, genome-wide DNA-DNA contact maps. One variant, termed 4C-seq (38), allows for the evaluation of interactions of one genomic region (viewpoint) with all others ("one-to-all").

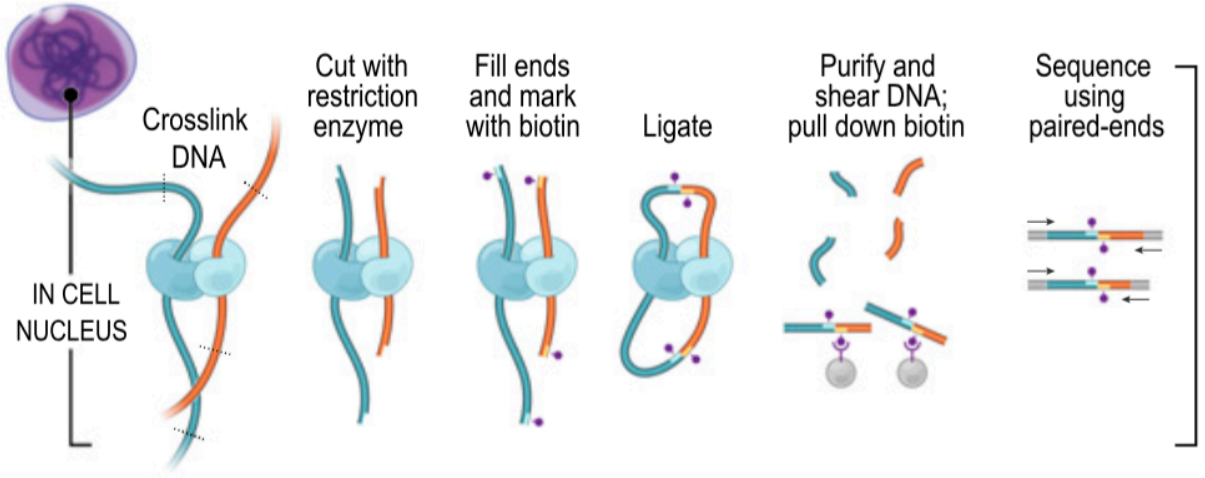

Figure 1-2: Schematic summary of the HiC protocol. (39)

A more powerful approach, denominated HiC (40), generates "all-to-all" contact maps, detecting any possible interaction happening in the cell nucleus (Figure 1-1). In HiC, fragment ends resulting from restriction enzyme cutting are biotin labeled. Then, DNA is purified, shared and biotin pull-down performed. By this, hybrid molecules containing information of interacting genomic fragments are

enriched in the generated libraries. After pulldown, adapters are added to either side of the ligation-junction, paired-end sequencing is performed and reads are mapped back to the genome as interactions.

A common limitation of HiC is the resolution of the detected interacting fragments, determined by the frequency of cutting sites of the restriction enzyme that was employed during library preparation. As every possible pairwise interaction is evaluated, sufficient coverage to confidently detect significant interactions comes at great sequencing efforts. Such efforts could be minimized by pooling genomic reads in genomic bins of various size (binning), but at the cost of decreasing the resolution of the detected interacting fragments. When Lieberman-Aiden et al. (40) first published their HiC protocol in 2009, they could not get resolution under 1Mb using approximately 10 million paired-end reads. There are two approaches to overcome the "low resolution" problem: Firstly, by increasing sequencing depth or, secondly, by an enrichment of fragments that cover a region of interest (41-43). The latter is used in the Capture HiC (CHiC) approach (applied in this work, see section 3.2.1). In CHiC, RNA baits are used to enrich for a genomic region of interest, thus allowing for an affordable, in depth sequencing of a selected region(44).

3C-derived methods have two major advantages and one disadvantage in comparison to traditional microscopy based works on genome architecture (37, 41). Firstly, they are considerably higher in resolution - Capture HiC can currently be performed at a resolution of up to 1kb. Secondly, 3C-derived methods provide the possibility of a systematic and genome-wide approach to chromatin architecture. They yield information not about a single locus or a single cell, but about the architectural conformation of entire genomes and of thousands of cells at a time. This strength is at the same time the greatest weakness of 3C methods; the generated results only ever represent an average of a whole population of cells at a selected time point and caution has to be taken when interpreting results in relation to dynamic systems and occurrences on a single-cell level. HiC data are visualized in contact matrixes as shown in Figure 1-2. The genomic region of interest is aligned along the x-axis. A point of interaction between two DNA sequences is represented as a pixel drawn in at a 45° angle from the two interaction regions. The number of reads (interactions) is represented by the color intensity of the pixel.

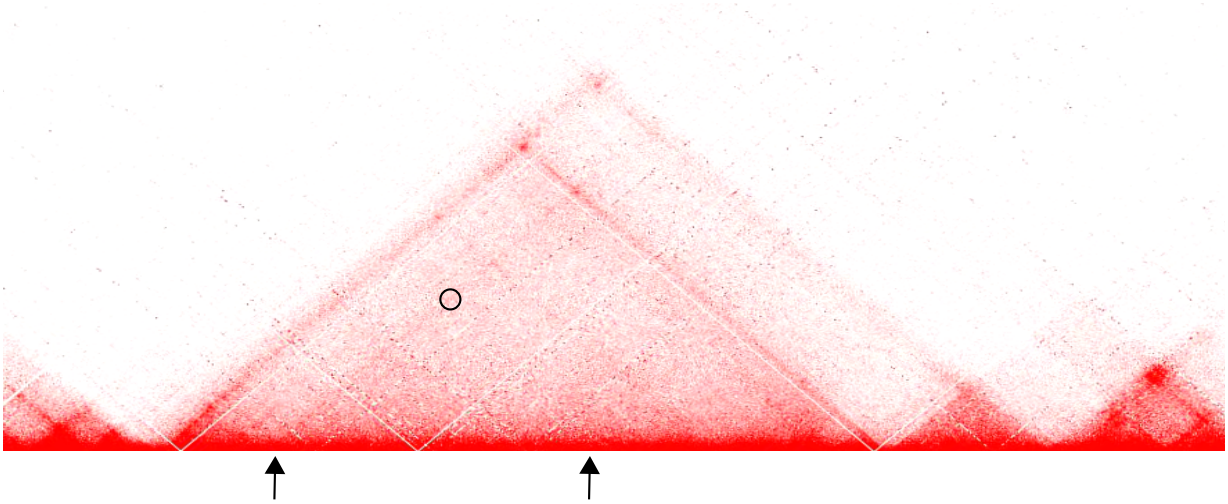

Figure 1-3: Contact matrix visualization of HiC reads. Exemplary contact matrix visualization of HiC reads. Each pixel represents the interaction of two 10kb regions. Consequently, the resolution of this data set is 10kb. The encircled area indicates a pixel as the visualized interaction of the two genomic regions marked by the arrows. Intensity of red correlates with intensity of interaction between genomic fragments.

1.2.2 Topologically associating domains (TADs)

An important finding of 3C-derived methods was the existence of Topologically Associating Domains (see Figure 1-4). In a randomly organized genome, it would be expected that interaction frequencies are dependent on the linear genomic distance from a viewpoint. However, HiC studies revealed a segmentation of genomes into megabase sized units - 800kb on average (45) - with higher frequency of contact than is to be stochastically expected in a randomly organized genome (46-48). This higher frequency of contact is the most basic characteristic of these self-associating domains termed TADs. Further, they are characterized by a steep drop-off of interaction between two separate units, which implies the existence of a physical boundary, that insulates two adjacent TADs. TADs are relatively invariant across different tissues (45, 49) and strongly conserved across species – indicating their evolutionary importance (45). However, more recent studies performed on single cells using both imaging and 3C-based methods have revealed that cell to cell variability of TADs is higher than previously thought (50-54).

The functional implications of TADs are still a matter of lively scientific discussion, however, there is thorough evidence for a functional role, highlighting the importance of TADs in transcriptional regulation. TADs can be seen as insulated units of cis-transcriptional regulation, within which long-range enhancer-promoter contacts are facilitated and interactions beyond TAD boundaries are

constrained (55). First evidence for this function stemmed from a study showing correlations between gene expression patterns of genes located within the same TAD (56). Further genome wide studies have shown that enhancer-promoters contacts are very frequently located within TADs (57-59) and enhancers' range of action correlate well with TAD boundaries (60). Functional studies have delineated the functional importance of TADs in promoting appropriate gene expression during development and in disease (47, 61-68). Particularly important for this thesis, is the work by Lupianez et al. 2015, in which structural variations (i.e., deletions, inversions, duplications) affecting boundary elements at the *Epha4* locus, causes ectopic enhancer-promoter rewiring, leading to aberrant gene expression and congenital disease. Following studies confirmed that proto-oncogenes can be activated through TAD breaking rearrangements driving the development of T-cell acute lymphoblastic leukemia, This finding highlights the implications of TADs in cancer (69).

In addition to their insulting function, there is also evidence that TADs can mediate enhancer-promoter contacts over long genomic distances as proposed by Symmons et al. 2016 (70). Using a series of chromosomal rearrangements in the *Shh* locus in mice, Symmons at al. demonstrate that rearrangements altering only intra-TAD distances do not affect gene expression. On the contrary, TAD-disrupting inversions affect enhancer-promoter contacts in a distant-dependent manner.

In summary, it is proposed that TADs serve the dualistic function of insulating enhancers and promoters from ectopic interactions while at the same time facilitating physiological enhancer-promoter contacts over large genomic distances.

1.2.3 Long-range interactions (loops)

As explained in section 2.1.1 it is generally assumed that an enhancer and its target promoter must physically interact to yield a transcriptional output, even if located at considerable linear genomic distances. Analyses of high-resolution Hi-C maps revealed the existence of chromatin loops, focal points displaying a high degree of interaction (42, 71). These loops often connect distal enhancers with promoters on an intra-TAD level (72), although they are also observed at the summit of TADs, connecting boundary regions (see Figure 1-4). Recent studies performing live measurement of enhancer-promoter proximity could show a close correlation between enhancer-promoter interaction and transcriptional output (3), although other live-cell imaging studies have disputed this claim (73,

74).

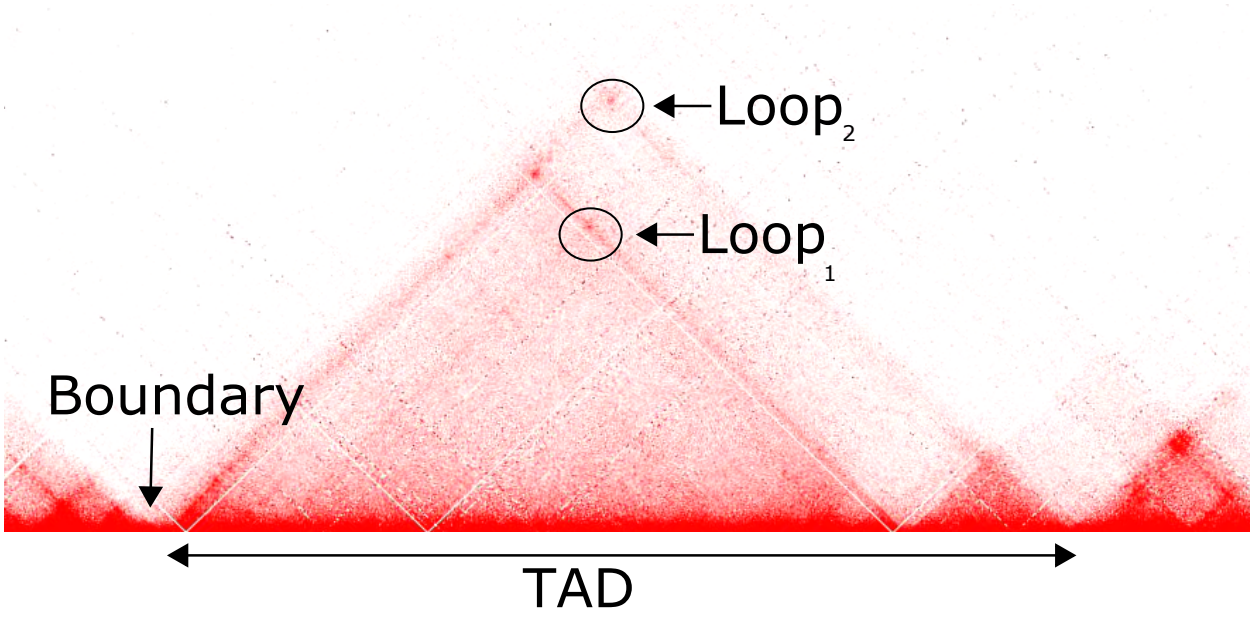

Figure 1-4: Exemplary HiC map of a TAD and long-range interactions (loops).

TAD shows self-associating properties (high contact frequency within TAD) and insulation at boundaries (steep contact frequency decline across boundary). Loop 1: Long-range chromatin interaction (displayed by higher contact frequency) of two regulatory units located within the same TAD. Loop 2: Long-range chromatin interaction (indicated by higher contact frequency) between TAD boundary elements.

When compared with TADs, these intra-TAD or sub-TAD interacting loops seem to represent a more dynamic system, being less conserved across both different tissues and species (42, 57, 75). The working model for incorporating the functions of such distinct genomic features is that TADs represent a more rigid, tissue invariable framework for transcriptional regulation by defining distinct units of interaction. Within TADs, a more dynamic system of tissue and time-point specific loops coordinates the spatiotemporal formation of enhancer-promoter contacts and thus regulations the transcriptional output. However, it remains unclear if TADs and loops are entirely different entities or merely the same phenomenon of looping observed on different levels of magnitude. Indeed, the mechanisms proposed to establish the formation of TADs and loops are largely identical (see section 2.2.4) (76). Further, looping structures may be present without there being any transcriptional output (67, 77). This highlights that physical enhancer-promoter contact is required but not necessarily sufficient to drive gene expression, as it might be influenced by many additional elements acting in trans.

1.2.4 CTCF and loop extrusion

Chromosome conformation capture methods provide valuable insight into 3D organization of genomes, but yield no information about the mechanisms driving the formation of the observed architecture. Numerous evidence position the CCCTC-binding factor (CTCF), a DNA binding protein comprised of 11 zinc finger units (78), as a crucial player in the formation of genome architecture. ChIP-Seq studies have shown a vast and ubiquitous distribution of CTCF binding across the genome (79) and a drastic enrichment of CTCF binding at TAD boundaries (46). The latter indicates the insulating function of CTCF (see section 2.1.1.3) which is further stressed by functional studies, showing that deletion or depletion of CTCF binding results in dramatic loss of TAD boundary function and increased inter-TAD interactions (61, 80-82). Seemingly contradictory to the insulating properties of CTCF, many studies suggest an enrichment of CTCF binding at enhancers and promoters and highlight CTCFs importance in facilitating formation of enhancer promoter contacts (61, 83, 84).

CTCF can dimerize in vivo, highly influenced by the orientation of its DNA-binding motif (83). Chromatin loops can generally be observed between CTCF binding motifs of convergent orientation (39). To date, the best model for incorporating this seemingly dualistic function of insulation and contact facilitation is the loop extrusion model (39, 76, 85). The computationally derived model proposes that loops are formed by the binding of two heterodimer Cohesin-CTCF complexes. The two binding units move along the chromatin strand in opposite directions, thus extruding a DNA loop behind them. Extrusion ceases when a heterodimer reaches a convergently orientated CTCF binding site, but continues otherwise. Therefore, loops would be preferentially formed between two CTCF binding sites of convergent orientation. This model of convergent CTCF site orientation is well supported both by genome wide correlation studies (42, 86, 87) and by functional studies (39, 88, 89).

1.3 The *EphrinA4* (*Epha4*) locus and its regulatory landscape

The *Epha4* gene encodes for a tyrosine kinase receptor of the family of Ephrin receptors and it is functionally implied mainly in the development of the nervous system (90). During development *Epha4* is expressed in the limb – in mice it facilitates correct motor-axonal pathfinding into the dorsal section of the developing limb bud. Homozygous *Epha4* knockout mice lack motor-neuronal innervation of the dorsal hindlimb, displaying peroneal muscular atrophy and absence of the peroneal

nerve (91).

1.3.1 Architecture and expression

The *Epha4* locus (Figure 1-2) is located on the long q-arm of chromosome 1. It is comprised of a large, over 2Mb TAD, containing only the single gene *Epha4* (Chr1:77.368.039-77.517.240). The centromeric side of the locus is characterized by a vast gene desert containing a few identified cis-regulatory elements. The most important of which - for this work at least - is a set of limb enhancers (Chr1:77.368.039-77.517.240). ChIP-Seq tracks for CTCF binding in E11.5 limb buds show the presence of a binding site adjacent to the described limb enhancer cluster, another near the *Epha4* promoter and the presence of a cluster of binding sites both at the centromeric and telomeric TAD boundary. These ChIP-Seq binding sites cover CTCF binding motifs with distinct orientations. The possible involvement of these CTCF binding sites in the spatial organization of the locus architecture is highlighted by the HiC data in Figure 1-2 which show: (1.) a loop between the convergently-orientated CTCF sites of the limb enhancer cluster and the *Epha4* promoter , (2.) a sharp drop off of contact frequency surpassing the centromeric and telomeric boundary elements, (3.) an increased contact frequency or loop between both boundary elements, the centromeric boundary element and the enhancer cluster and the centromeric boundary element and the *Epha4* promoter.

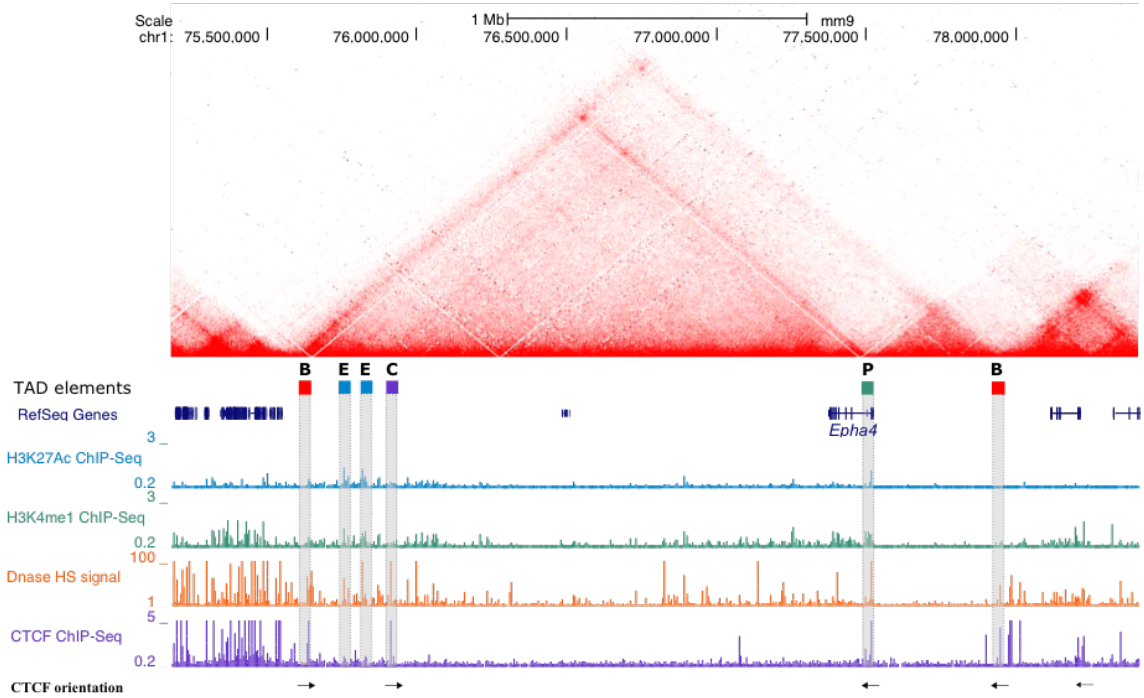

Figure 1-5: Characterization of the Epha4 locus. A RefSeq Gene track from UCSC browser, ChIP-Seq tracks for H3K27Ac (active enhancers), H3K4me1 (CREs in general), CTCF and a Dnase hypersensitivity signal (open chromatin mark) are aligned with the HiC map of the *Epha4* TAD in E11.5 mouse limb buds (provided by Dario Lupianez and Ivana Jerkovic). The orientation of CTCF sites is indicated by arrows below the CTCF ChIP-Seq track. Thus the crucial regulatory elements within the *Epha4* TAD are labeled: B = TAD boundary; E = distal limb enhancers; C = CTCF site flanking limb enhancer cluster; P = promoter of *Epha4* (CTCF site flanking the promoter not labeled but visible in ChIP-Seq CTCF track below).

1.3.2 The Paired Box 3 (*Pax*3) locus

On its telomeric side the *Epha4* TAD is flanked by a smaller TAD that contains the gene *Paired Box 3* (*Pax3*). *Pax3* encodes a TF involved in the development of numerous tissues including the nervous, muscular and cardiovascular system (92). Mutations affecting the *PAX*3 gene associate with several phenotypes (Waardenburg Syndrome) including limb malformations. ChIP-Seq tracks in E11.5 mouse limb buds indicate the presence of convergently-orientated CTCF binding sites at both boundaries of the Pax3 TAD, as well as a centromerically-orientated CTCF binding site upstream of the *Pax*3 promoter (Figure 1-2).

1.3.3 Pathogenic Structural Variations at the *Epha4* locus

Structural variations (SV) are generally defined as mutations larger than one kb. They can be categorized into balanced SVs, as translocations and inversions for which the copy number of the affected sequence is not altered, and unbalanced SVs, such as deletions, duplications and insertions which are consequently also referred to as copy number variations (CNVs). Structural variations are associated with disease (93) although in many cases the implied causality of this remains unknown (94). In 2015, Lupianez et al. created mouse models with large SVs in the *Epha4* locus that have been found to be associated with cases of familiar limb malformations in humans(61). Using these models, they determined the underlying causality between the described SVs and the corresponding limb malformations observed in these patients. In all studied cases, the SVs affected *Epha4* TAD boundaries and disrupted the spatial organization at the locus. The disruption of the insulating function of the TAD led to an ectopic rewiring of *Epha4* limb enhancers to genes that were physiologically located in an adjacent TAD. As a consequence, ectopic gene expression in developing limbs was observed, which resulted in pathological phenotypes that matched those observed in families with analogous mutations. The most important case for this thesis is from a patient affected by brachydactyly (shortening of fingers). The patient carried a 1.5 Mb deletion including parts of the *EPHA4* TAD, the gene itself, and the boundary that separates it from the adjacent *PAX*3 TAD (see

Figure 1-2). In the corresponding mouse model, the deletion (*Pax3* $^{DelB/DelB}$) caused an ectopic rewiring of the cluster of *Epha4* limb enhancers with the *Pax*3 promoter, leading to an ectopic expression of *Pax*3 in the limb (observed at stage E11.5). As a consequence, *Pax3* recapitulated the expression pattern of *Epha4* and caused the brachydactyly phenotype.

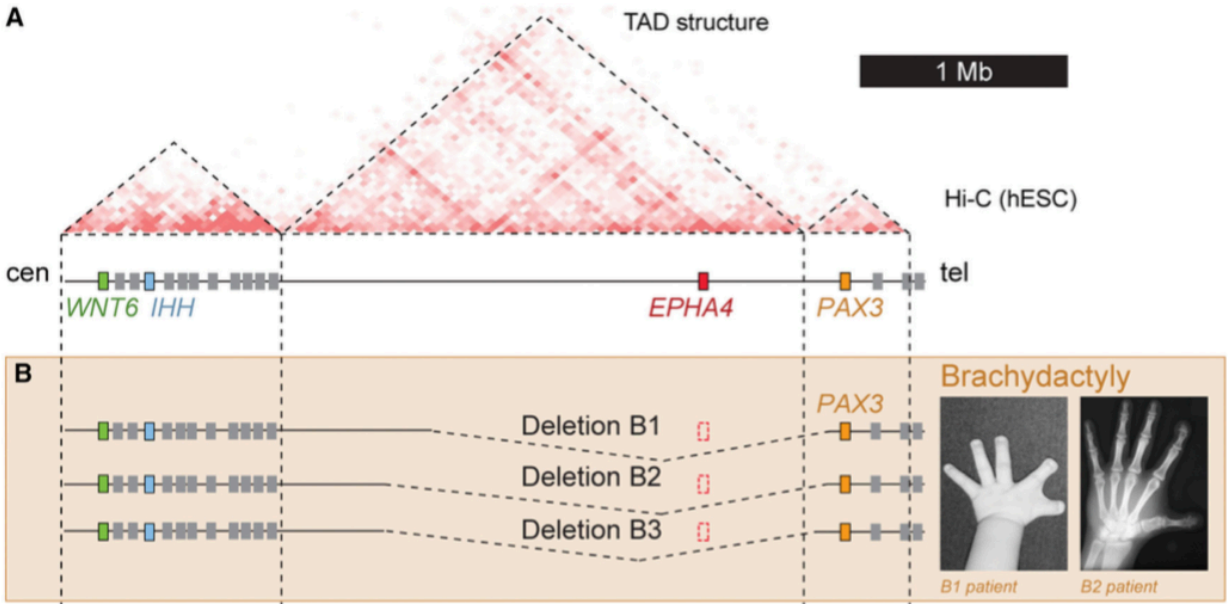

Figure 1-6: TAD disrupting pathogenic structural variation at the *Epha4* locus. A: Low resolution HiC map of the *Epha4* TAD and its surroundings. Gene names are aligned below. B: Three different deletions observed in unrelated families with brachydactyly phenotype are indicated and phenotype of two patients shown. (61)

1.4 Objective of this work

The function of TADs, as insulating units controlling gene expression, and the involvement of CTCF in TAD formation and function has been thoroughly investigated. However, the presence of CTCF is not exclusive of TAD boundaries, but also associated with intra-TAD loops. Such loops are generally associated with enhancers and promoters, suggesting a possible involvement of CTCF in mediating these functional contacts.

To investigate this, we used the CRISPR-Cas9 system to create homozygous deletions of CTCF binding sites in mice and test their functional consequences *in vivo*. First, we tested the integrity of the *Epha4* TAD by deleting three CTCF binding sites flanking loop anchor points, individually and in combination. Second, we studied the role of CTCF binding in the phenotypes observed on the *Pax3*$^{DelB/DelB}$ background, where a TAD boundary deletion leads to ectopic *Pax3* activation and

brachydactyly. For this purpose, we deleted CTCF-binding sites associated to the pathogenic interaction between *Epha4* enhancers and the *Pax3* gene. In both cases, we evaluated effects on genomic architecture and gene expression.

The results presented here facilitate our understanding of the general principles of 3D chromatin organization in mammals, as well as their functional implications during development and in disease.

2. Methods

2.1 Generating structural variations using CRISPR/Cas9

Structural variations in mouse embryonic stem cells (ESCs) were created using CRISPR-Cas9 (95). The experimental strategy is summarized in Figure 2-1.

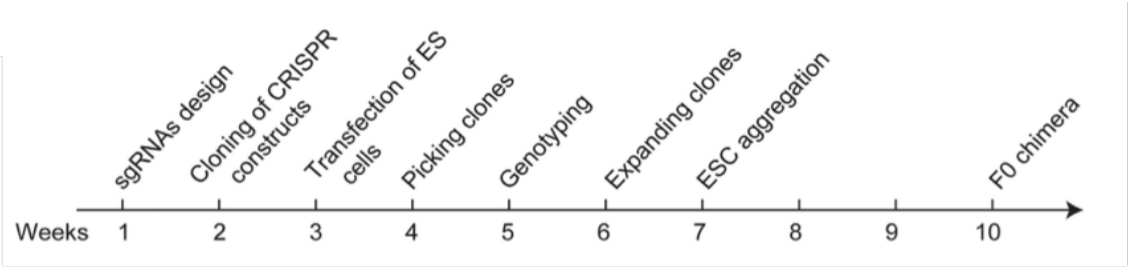

Figure 2-1: Time plan for creation of mouse models containing SVs using CRISPR-Cas9. The protocol spanning over 10 weeks is explained in detail over the following methods section from 2.1.1 – 2.1.5. (95)

2.1.1 Vector cloning

Target cells were transfected with a PX459 vector containing a U6 promoter for RNA polymerase binding, the coding sequence for the pSpCas9 enzyme (Cas9 enzyme from Streptococcus pyogenes), a puromycin resistance gene and two Bbs1 restriction enzyme sites, in between which the desired gRNAs desired was cloned.

2.1.1.1 Guide design

Single gRNAs were designed using the "design tool" on http://benchling.com. Guides were designed to have a length of 20 bp and low off-target potential, based on the scoring system of the design tool. Only guides with an off-target score above 50 and an on-target score above 60 were selected. Each gRNA was designed to have a G nucleotide at the 5' end of the gRNA, adjacent to the U6 promoter. If this nucleotide was not originally present in the sequence, the G was manually added to reach a total length of 21 bp. The complementary genomic sequence for each guide was flanked by the required downstream PAM sequence NGG. The corresponding sgRNAs oligo, as well as the corresponding reverse complementary were synthesized by an external company. Additional sequence overhangs "cacc" and "caaa" were added to the 5' end of the original oligo and to the 3' end of the complementary guide respectively. The gRNAs targeting the *Epha4* or *Pax3* promoters (*Epha4*ΔCTCF_P and *Pax*DelBΔCTCF_P, respectively) were designed to not interfere with promoter

function, based on ChIP-Seq tracks for H3K4me3 histone modifications in the developing limb in E11.5 mice. For *Epha4*ΔCTCF_B constructs, which sought to delete a cluster of CTCF sites spanning 2 kb within the centromeric boundary of the *Epha4* TAD, a double guide approach for CRISPR-Cas9 induced SVs was used - as presented in Kraft, Geuer et al. 2015.

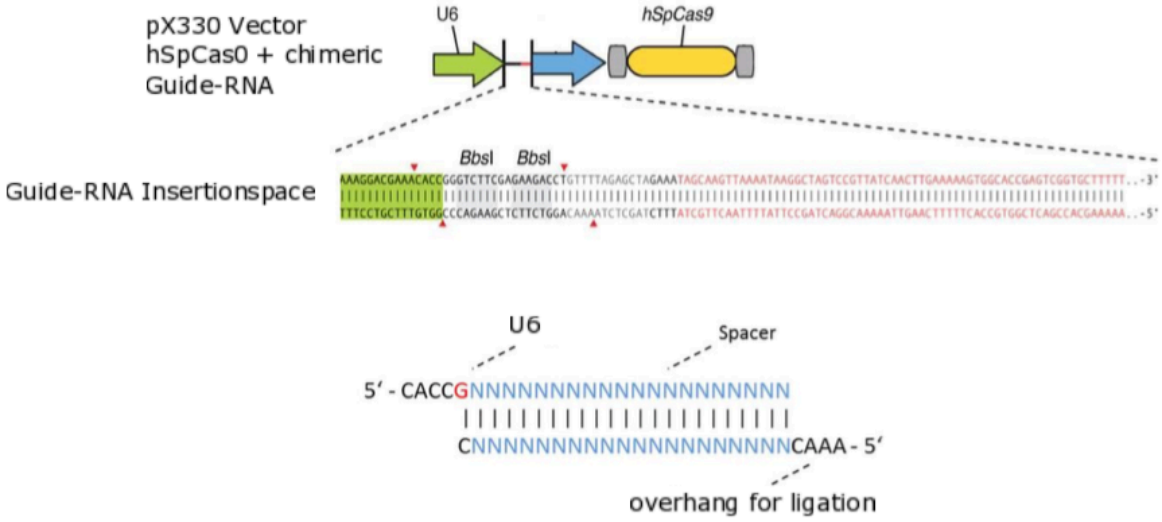

Figure 2-2: Cloning of a Guide-RNA Insert into a pX459 vector. The pX459 Vector contains DNA sequences for the hSpCas9 protein, the tracrRNA and Guide-RNA insertion site. The U6 promotor facilitates the binding of RNA-polymerase III and thus of transcription. The hSpCas9 sequence is flanked by the Nuclear localization signal, which ensures that it is transported into the cell nucleus. The Guide-RNA Insert possesses the 3‘ and 5‘ overhangs complementary to the overhangs remaining at the BsbI restriction site, which are required for ligation into the vector (modified after (96)).

2.1.1.2 Annealing

Guide-RNA oligonucleotides were diluted in bidistilled water (bid H2O) to generate stock solutions and stored at -20ºC, according to the instructions provided by the manufacturer. 10 μl of the forward strand and 10 μl of the reverse strand stock solutions were mixed with 10 μl of ligation buffer (10x) and 70 μl of bid H2O. The mixture was placed into a preheated metal block at 95°C and incubated for 15 min, followed by cool down at room temperature (RT) for 45 min.

2.1.1.3 Phosphorylation

5 µl of the annealed gRNA were mixed with 2 µl T4 Polynucleotide Kinase Reaction Buffer (10x) , 2 µl of ATO (10mM), 1 µl of T4 Polynucleotide Kinase (PNK) and 10 µl of H2O bid. The reaction mixture was mixed by pipetting, spun down and left to incubate at 37°C for 20 min. The mixture was placed at 75°C for 10 min, to inactivate the PNK enzyme.

2.1.1.4 Ligation

1 µl of the dephosphorylated PX459 vector (diluted 1:10), 2 µl of the phosphorylated oligonucleotides (approximately 2 ng of DNA in total), 2 µl of T4 DNA Ligase Buffer and 1 µl of T4 Ligase enzyme were eluted in 14 µl of H2O bid, mixed by pipetting and left to incubate at RT for two hours or 14°C overnight.

2.1.2 Cultivating plasmid in bacteria

2.1.2.1 Transformation

Top10 *Escherichia coli* cells were taken from -80°C and left to thaw gradually on ice. 100 µl of bacteria were added to 10 µl of ligation solution, mixed by pipetting and left to incubate for 30 min on ice. To increase the plasma membrane porosity the solution was heat shocked in a water bath at 42°C for 1 min. The solution was incubated further 5 min on ice. Than 900 µl of SOC media were added and the solution left to incubate for 60 min at 37°C. After 60 min the suspension was centrifuged for 2 min at 6000 rpm, 90% of the solution discarded, the remaining 10% mixed by pipetting and equally spread on a ampicillin treated agar plate. The plate was left to incubate over night at 37°C.

2.1.2.2 Colony PCR

Primers complementary to the backbone of the pX459 vector and to the integrated guide RNA were aliquoted in a standard PCR master-mix and placed on a 96well plate. Using a 10 µl pipet tip, single bacterial colonies were picked from the agar plate, plated and labeled on a further ampicillin-treated replica plate and subsequently placed into a well containing the PCR master-mix. By shaking the 96well plate slightly, it was ensured that sufficient bacterial template entered the PCR wells. The replica plate was left to incubate at 37°C overnight. The bands of colonies tested positive for the integrated vector were cut out for sequencing.

Table 1: Pipetting scheme for colony PCR

Reagent	Volume
Primer 1 (guide RNA)	0.4 µl
Primer 2 (KolR)	0.4 µl
dNTPs (1.25 mM)	0.5 µl
Taq	0.25 µl
Taq buffer (15mM)	2 µl
Bidistilled	16.45 µl
Total	20 µl

Table 2: Cycler program for colony PCR

Temp (C°)	Time	Cycles
96	5 min	
96	30 sec	
55	30 sec	25 x
72	2 min	
72	7 min	
4	∞	

2.1.2.3 Inoculation, glycerin stocks and plasmid isolation

Positively tested colonies were retrieved from the replica plate and inoculated in 2 ml Ampicillin + L-Media (1:1000) overnight. Bacteria were stored for further use at -80°C in the form of glycerin stocks: 0.5 ml L-Media with suspended bacteria were mixed with 1 ml 80% glycerin. After successful gRNA integration into the vector had been confirmed by sequencing (see section 2.1.4.2) positively tested colonies were retrieved from the glycerin stocks and inoculated in 50 ml Ampicillin + L-Media overnight. The following day plasmid DNA was extracted from the inoculated bacteria using the "NucleoSpin Plasmid" kit (Machinery-Nagel). The plasmid DNA was eluted in 50 µl H2O bid.

2.1.3 Mouse ESC-culture

2.1.3.1 General cell culture methods

6cm cell culture plates were covered with 0.1% gelatin and incubated at 37°C for 20 min. Afterwards, the gelatin was removed and the plates allowed to dry at RT for 5 min. Feeder cells were seeded on gelatinized plates, fed with ES media and maintained in a cell incubator at 37ºC, 5% CO2. On the next day, a frozen vial of mouse ESCs was thawed and cells were seeded onto the feeder plates. mESCs were fed daily with ES+LIF media (10.000:1).

When splitting or freezing was necessary, mESCs were previously fed with ES+LIF for 2 h. Afterwards, cells were washed 2 times with PBS and subsequently detached by adding 0.1% trypsin and incubating for 7 min at 37°C. After trypsin was inactivated by adding 2 volumes of ES media, cells were pelleted by centrifuging for 5 min at 1100 rpm and resuspended in the corresponding media. Cell quantities were determined using a Neubauer-Zählkammer and the following equation:

$$\textit{total cell number} = \textit{average of counted cells per quadrant} \times \textit{factor of dilution} \times 10^{4}\ \textit{(chamber factor)} \times \textit{total volume cell suspension}$$

2.1.3.2 Transfection

Mouse ESCs were transfected with a PX459 vector containing the gRNA for the different CRISPR-Cas9 constructs. Two days before transfection, $8x10^{5}$ CD1 feeder cells were plated on a 6cm gelatinized plate. The day before transfection $4x10^{5}$ ESCs were seeded on top of the feeder cells. After overnight incubation cells were washed with PBS twice and the standard media was changed to 1.75 ml ES+LIF without streptomycin, because the antibiotic interferes with the transfection process. FuGENE HD and OptiMEM were allowed to reach RT for 15 min. 100 μl of OptiMEM were mixed with 25 μl of FuGENE HD. Further 125 μl of OptiMEM were mixed with 8 μg of a plasmid. The 125 μl FuGENE-OptiMEM mix was added to the OptiMEM-DNA mix drop by drop and incubated at RT for 15 min. Finally, the total 250 μl of transfection solution were distributed equally over a 6cm dish with ESCs and incubated overnight at 37°C before changing media back to ES+LIF with streptomycin.

2.1.3.3 Selection

The day after transfection 3x 6cm dishes of puromycin resistant DR4 cells were seeded. Two days after transfection, the transfected cells were plated onto DR4 plates (1:3) and selected for 48 h by adding 2 μg of puromycin to 1ml of standard media. After selection, the media was changed back to

ES+LIF and resistant clones were grown for 4-6 days.

2.1.3.4 Clone picking

One day before picking, CD1 feeder cells were seeded out on 96well plates. On the following day, the 6cm plates with clones were washed twice and covered with 2 ml of PBS. Additional 96well plates with 30 μl of trypsin per well were prepared. Then, single clones were picked of the 6cm dishes using a 10 μl pipet and transferred to the trypsin plate. After incubating them for 10 min at 37°C, trypsin was deactivated by adding 120 μl ES+LIF, pipetted several times to achieve a single-cell suspension and transferred to the 96well plates covered with CD1 feeders. Cells were incubated at 37°C, 5%CO2 for 2-3 further days.

2.1.3.5 Split and freeze

Once cells were grown to a sufficient density, they were detached from the plates. Two thirds of the cells were frozen for further use (e.g. expansion and possible aggregation if the correct genotype was confirmed) and one third was left to grow for several more days to extract DNA for genotyping. Briefly, 30 μl of trypsin per well were added and incubation occurred at 37°C for 10 min. Normal media (NM= 20% FCS, 80% Bicarbonate free DMEM) and freezing media (FM= 20%FCS, 20% DMSO, 80% Bicarbonate free DMEM) were prepared. U-bottomed 96well freezing plates with 50 μl FM/well were prepared. Trypsin was deactivated by adding 120 μl of NM/well. Then, 50 μl of each well were transferred to corresponding wells on two separate 96well freezing plates. Freezing plates (replica plates) were packaged in styrofoam boxes and frozen at -80°C. 150 μl of ES-LIF were added to the remaining 50 μl of cells. Cells were incubated for further 2-3 days and DNA was extracted for genotyping.

2.1.3.6 Expanding and freezing of positive clones

DNA plates were screened by PCR to identify clones with the appropriate genotype (see section 2.1.4). Positive clones were subsequently confirmed by Sanger sequencing, retrieved from the frozen 96well replica plates and expanded to reach approximately $3x10^6$ cells. After expansion, three vials containing $8x10^5$ cells each were stored in cryotubes at -80°C or in liquid nitrogen for further use e.g. aggregation. For the freezing process, cells were suspended in 0.5 ml FM + 0.5ml NM.

2.1.4 Screening for positive clones

DNA was extracted from cells, PCR genotyped and finally deletions were verified by Sanger sequencing.

2.1.4.1 PCR based genotyping

2.1.4.1.1 Primer design

Primers were designed using the Primer3, NetPrimer and EnsembleBlat software. Two different approaches were applied for either the small CTCF motive deletions (1) or larger genomic deletions (kb-sized) incorporating several CTCF sites (2).

(1) For single CTCF motive deletions a forward and a reverse primer - centromerically and telomerically of the gRNA binding site - were designed (see Figure 2-3). The PCR product for clones carrying the required deletion thereby varied in size from the wild type (wt) PCR product - it was between 20-200 bp smaller. Deletion size was quantified by Sanger sequencing.

(2) For larger genomic deletions four primers were designed for every construct, allowing for the genotyping of deletions, inversions, duplications and wild type (see Figure 2-3) (95).

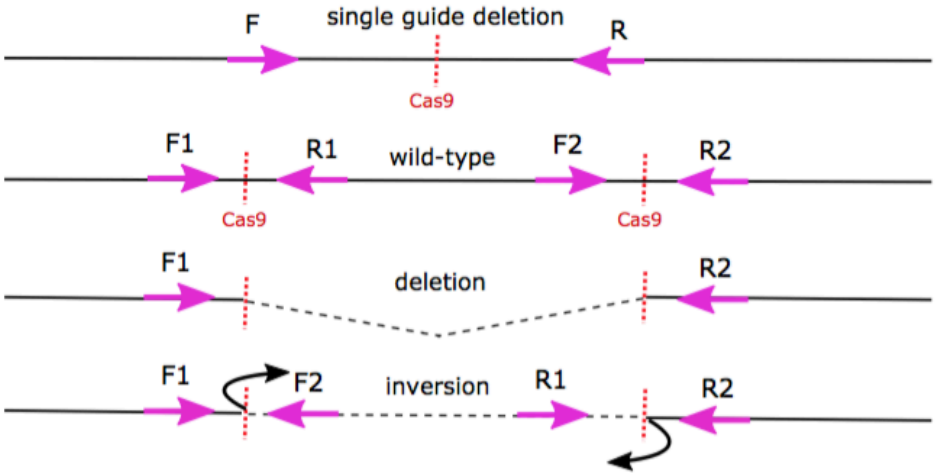

Figure 2-3: Genotyping approach - for small CTCF motif knockouts (20-200bp) and larger structural variations. Primers were designed to bind at least 250 bp away from the gRNA binding site (red dotted line) and with low off-target potential. Two primers were designed for the deletions created using the above described single gRNA approach. Four primers were designed for the genotyping of larger deletions and unwanted structural variations e.g. inversions and duplications (duplications not shown). A wild type PCR (using primer pairs F1/R1 and F2/R2) was always performed to verify homo- or heterozygosity.

2.1.4.1.2 DNA isolation and genotyping PCR

Cells in 96well DNA plates cells were washed twice with PBS. 50μl of lysis buffer + proteinase K (1:1000) was added per well. Lysis was performed overnight at 55°C. DNA was extracted using a "Mag Attract HMW DNA" kit and the DNA extraction program provided on the KingFisher robot provided by Biosoft. Subsequently, clones were PCR genotyped - the pipetting scheme and cycler program shown in Table 3 and 4 were applied. The temperature for the annealing step in the PCR cycler program was adjusted according to the annealing temperatures of the designed primers.

Table 3: Pipetting scheme for PCR based genotyping

Reagent	Volume
DNA (ca. 50 ng)	1-2 μl
10x PCR buffer	2 μl
dNTPs (1.25 mM)	0.5 μl
Forward Primer	0.4 μl
Reverse Primer	0.4 μl
Taq	0.24 μl
bidest	to 20 μl total

Table 4: Cycler Program for PCR based genotyping

^	Time	Cycles
96	5 min	
96	30 sec	
55	30 sec	35 x
72	1 min	
72	7 min	
4	∞	

2.1.4.1.3 Gel electrophoresis

PCR products were evaluated by gel electrophoresis on a 2% agarose gel. Again, two different approaches were used for singular CTCF motive deletions and kb-sized deletions: (1) Single CTCF motive deletions were detected by loading samples on a 2% agarose gel and allowing a clear separation from wildtype bands. Using this approach, clones with deletions >20 bp could be detected. „Deletion PCR bands" were consequently cut out of the gel, DNA was extracted and the deletion confirmed by Sanger sequencing. (2) Larger genomic deletions could be identified as only clones carrying deletions would yield a deletion band in the PCR (see Figure 3-2). Subsequently, samples with a deletion band were genotyped for inversions and duplications. In order to assess an homozygosity status, positive clones were subsequently screened for wild type alleles (see Figure 2-

3). To minimize a possible wild type signal from CD1 feeder cells, clones and CD1 cells were allowed to settle on a gelatinized plate for 30 min twice, before transferring them to another well. Deletion bands were cut from the agarose gel, DNA extracted using a "NucleoSpin Gel and PCR Clean-up" kit and the deletion verified or falsified by sequencing.

2.1.4.2 Sequencing

Deletions fragments were amplified by PCR and verified by Sanger sequencing. The cycler program and the sequencing protocol are summarized in table 5 and 6.

Table 5: Protocol for sequencing PCR

Reagent	Volume
DNA (ca. 50 ng)	2 μl
5x sequencing buffer	2 μl
BigDye terminator reaction mix	0.5 μl
Primer (10pM/μl)	0.4 μl
Bidistilled H2O	to 10 μl total

Table 6: Cycler Program for sequencing

Temp (C°)	Time	Cycles
96	1 min	
96	30 sec	
50	30 sec	25 x
60	4 min	
4	∞	

2.1.5 Tetraploid aggregations

Mutant embryos were generated from expanded clones via tetraploid aggregation(97). Aggregations were kindly performed by Karol Macura and Dr. Lars Wittler of the Herrmann Department at Max Planck Institute for Molecular Genetics (MPIMG). $8x10^5$ mutant cells were seeded out on a 6cm dish covered with CD1 feeders 2-3 days before aggregation.

2.2 Evaluating effects of genomic aberrations in distal limb E11.5 tissue

2.2.1 Preparation of embryos and treatment of animals

Caretaking of retransfered females was carried at the animal facility of Max Planck Institute for

Molecular Genetics under the supervision of Dr. Ludger Hartmann. All animal procedures were conducted as approved by the local authorities (LAGeSo Berlin) under license number G0247/13.

All experiments were performed in E11.5 distal limb tissue. Additional embryonic tissue was obtained for genotyping - tail tip tissue from embryos used for qPCR and HiC experiments, amnion tissue from embryos taken for RNA in situ hybridization. DNA was extracted using a QuickExtract solution, eluted to 50 ng/μl and genotyping performed as described in section 2.1.4.1.

2.2.2 qPCRs

Forelimb distal limb tissue was obtained from E11.5 embryos, snap-frozen in liquid nitrogen and stored at -80°C. To ensure a correct developmental stage, embryos were staged according to tail somite number and only those displaying 17-20 tail somites were used. RNA was isolated from samples using the "RNeasy Mini" kit from Qiagen by following the instructions provided by the manufacturer. The corresponding cDNA was generated by using the "Superscript" kit from Invitrogen and following the instructions provided by the manufacturer. The total amount of RNA used for cDNA synthesis was 500 ng. Real time PCRs were performed with SYBR Green technology. GAPDH was used as the reference gene for normalization. The generated data was analyzed using SDS (Applied Biosystems) and Excel.

2.2.3 RNA in situ hybridization

In situ hybridization was performed on E11.5 embryos using standard methodologies (98). This work was kindly done by Norbert Brieske of the group "Development and Disease" at the Max Planck Institute for Molecular Genetics. It was ensured that mutant embryos and wt control animals were at the same developmental stage by counting embryo somites.

2.2.4 Capture HiC

2.2.4.1 SureSelect design

SureSelect RNA enrichment probes were designed over the genomic interval chr1: 71,000,001-81,000,000 (mm9) using the SureDesign tool by Agilent. Probes were designed over the entire genomic region and not specifically proximal to DpnII sites, with coverage of 85%.

4.2.4.2 Capture HiC library preparation

Distal limb tissue from forelimb and hindlimb of E11.5 embryos was dissected and pooled in 10% FBS/PBS.

Disruption of tissue: FBS/PBS was aspirated and tissue resuspended in 1 ml of trypsin. Samples were incubated for 10 min at 37°C with pipetting every 2 min, to ensure a single-cell suspension.

Fixation of tissue: Trypsin was inactivated by adding 4 ml 10%FBS/PBS and the solution filtered with 40 μm cell strainer into a new falcon. The cell strainer was washed with further 5 ml of FBS/PBS and the solution centrifuged for 5 min at 1100 rpm. The supernatant was discarded, the pellet resuspended in 5 ml 10% FBS/PBS and 10 μl of sample taken for determining the quantity of cells. 5 ml of 4% formaldehyde (FA) solution in 10%FBS/PBS were added, the sample was mixed by inverting 5x and incubated on a rotating device at RT for 10 min. Fixation was stopped by transferring the sample to ice and by adding 1 ml of a 1.425M of glycine solution

Lysis: Samples were centrifuged at 4°C for 8 min, the supernatant removed, samples resuspended in 5ml cold lysis buffer and incubated on ice for 10 min. Lysis was validated by Methyl Green-Pyronin staining. Samples were centrifuged for 5 min at 2000 rpm at 4°C, the supernatant removed and sample pellets resuspended in 360 μl H_2O.

Digestion: 60 μl of 10X restriction buffer and 15 μl 10% SDS were added to the samples, which were then incubated in a thermoblock at 37°C for 1 h at 900 rpm. 150 μl of 10% Triton-X 100 were added and incubation continued for 1 h. 600 μl restriction buffer and 400U restriction enzymes were added and incubation performed for 4 h at 37°C at 900 rpm. Further 200U of restriction enzyme were added and incubation continued overnight. On the next day 200U restriction enzyme were added and samples incubated for further 4 h.

Ligation: Digestion efficiency was validated by gel electrophoresis. Afterwards, restriction enzymes were inactivated by incubating for 20 min at 65°C. Ligation was performed by adding 700 μl ligation buffer, 7 ml H_2O and 50U T4-DNA-Ligase and by incubating at 16°C overnight.

De-crosslinking: 30 μl of Proteinase K were added and samples were incubated overnight at 65°C. On the next day, 30 μl RNAse A were added and samples incubated for 45 min at 37°C. 7 ml phenol-chloroform were added, samples mixed and centrifuged for 15 min at 3750 rpm. The aqueous phase was subsequently transferred to new 50 ml tubes and 28 μl glycogen (5 mg/dl), 1.5 ml 2M NaAC (pH

5.6), 35 ml 100% ethanol and 7 ml H_2O were added. Samples were incubated at -80°C overnight. On the next day, samples were centrifuged at 4° for 20 min at 8350 rpm. After removing the supernatant, 10 ml of cold 70% ethanol were added. Samples were centrifuged at 4°C for 15 min at 3300rpm. The supernatant was removed and pellets allowed to dry at RT. Pellets were dissolved in 150 μl 10mM Tris pH 7.5 and stored at -20°C.

Capture: Samples were next sheared using a Covaris sonicator (duty cycle: 10%, intensity: 5, cycles per burst:200, time 6 cycles of 60 s, set mode: frequency sweeping, temperature 4° - 7°C). Adaptors were ligated to the sheared DNA and amplified according to instructions provided by Agilent. The library was hybridized to our designed SureSelect RNA probes and indexed for sequencing following Agilent instructions (100bp paired-end mode). The capture part of the CHiC protocol was kindly performed by Myriam Hochradel of the Max Planck Institute for Molecular Genetics sequencing facility.

4.2.4.3 Capture HiC analysis

Bioinformatic analysis was kindly performed by Robert Schöpflin of the Department of Computational Molecular Biology, Max Planck Institute for Molecular Genetics. Library sequencing was performed paired-end. The HiCUP pipeline v0.5.8 (99) (nofill: 1, format: Sanger, without di-tag length restriction) was used for mapping paired-end sequencing results and for filtering for unique and valid di-tags. The pipeline applied Bowtie2v2.2.6 (100), referencing to the genome mm9. For DelB mutants a customized genome file was applied. Juicebox command line tools (42, 101) were next used for binning and normalizing (KR normalization). Capture performance enriched for a genomic region of 10Mb, resulting in three different regimes of CHiC maps: (i) enriched vs. enriched, (ii) enriched vs. non-enriched and (iii) non-enriched vs. non-enriched. Only di-tags for regime (i) were considered for binning and normalization. Di-tags were filtered for the enriched genomic region and all coordinates were shifted by the beginning of the enriched interval. Accordingly, files in the Juicebox command line tool were set to only include the size of the enriched regions. The minimum MAPQ for importing was set to 30. After binning and normalizing, coordinates were returned to their original values.

3. Materials

3.1 Devices and software

Table 7: Devices

Device	Description	Manufacturer
Acio Cam	MRC5	Zeiss
Binocular	MZ 12	Leica
Centrifuge	Biofugepico	Thermo
	3417R	Eppendorf
	RC-Superspeed	DuPont Instruments Sorvall
	Megafuge 1,0 Heraeus	Thermo
Incubator	HEPA Class 100	Thermo
Gel documentation	Easy Win 32	HeroLab
Heating block	Ori-Block OV3	Techne
Neubauer cell chamber	0,0025 mm	Marienfeld
Petri dishes for aggregation	633102	Greiner
Pipets		Eppendorf
Pipetboy	acu	ISS Intgra Bioscience
Scales	PB303-S/PH	Mettler Toledo
Spectrometer	ND-100	NanoDrop
Thermal Cycler	2720	AB Applied Biosystems
UV-Chamber		HeroLab
Vortex	Microspin FV-2400	Lab4you
Water bath	2008 series	GFL

Table 8: Software programs

Program	Use of program
ApE Plasmid Editor	Analysis of Sequencing
Axio Vision 4.6	Photography
Benchling.com	CRISPR-Cas9/gRNA design
DNA Star Seqman	Analysis of Sequencing
Easy Win 32	Gel documentation after gel-electrophoresis
Endnote	Citations
Ensemble Blat	Analysis of off-target potential of primers
Inkscape	Image editing
NCBI	Database for research

NetPrimer	Primer design
Primer3	Primer design
UCSC	Database for gRNA design, primer design etc.

3.2 Reagents

Cell culture reagents: ES-Media was stored at 4° C for up to 5 days. For longer periods of time ES-Media was stored at -20°C. Freezing media was always freshly prepared.

Table 9: ES-Media (with and w/o LIF)

Component	Volume to 625 ml total (ratio)
KO DMEM	500ml (80%)
100x L-glutamine (200mM)	6.25ml (1%)
100x Non-essential amino acids	6.25ml (1%)
ß-Mercaptoethanol (10mM)	6.25ml (1%)
100x Nucleosides	6.25ml (1%)
Penicillin/streptomycin (10.000 U/ml)	6.25ml (1%)
Fetal calf serum	93.75ml (15%)
Leukemia Inhibitory Factor (LIF) 10^7 U/ml	62.5 μl (1:10^{5})

Table 10: Freezing media

Reagent	Volume (ratio)
F1 Freezing Media	
KO DMEM/bicarb DMEM	8ml (80%)
FCS	2ml (20%)
F2 Freezing Media	
KO DMEM/ bicarb DMEM	6ml (60%)
FCS	2ml (20%)
DMSO	2ml (20%)

Table 11: General lab-work reagents

Reagent	Composition
10x PCR reaction buffer	705mM Tris, HCl pH 8,8, 200mM $(NH_4)SO_4$, 15mM MgCL2
6x Agarose gel loading buffer	0,175 g Orange G, 15g Sucrose, 50ml H_2O

Proteinase K buffer	1M Tris (pH 7), 0,5M EDTA in DEPC H2O
PBS	2g KCl, 14,4g Na_2PO_4, 80g NaCl in 1l H2O DEPC
QuickExtract (DNA extraction)	17mM EDTA, 17mM Tris (pH 7,5), 170 mM NaCl, 0,85% SDS
TAE buffer	5mM sodium acetate, 0,04 M Tris, 1mM EDTA, pH8
Sequencing reagents	U6 primer: ACTATCATATGCTTACCGTAAC; ColR primer: CACGCGCTAAAAACGGACTA

Table 12: Capture HiC reagents

Reagent	Composition
Lysis buffer *(prepared freshly and stored on ice)*	500μl 1M Tris pH 7.5, 300μl 5M NaCl, 100μl 0.5M EDTA, 50μl NP-40 (0.0528g), 115μl Triton X-100 (0.106g), 400 μl 25X proteinase inhibitors, 8.353ml H_2O
Ligation buffer	0.4M Tris pH 7.8, 0.1M MgCl2, 0.1M DTT, 0.0083M ATP
37% formaldehyde	0.555g PFA in 1050μl 10% FCS/PBS, 15 μl NaOH
4% formaldehyde	594μl 37% formaldehyde, 4.9ml FCS/PBS

Table 13: Reagents for qPCRs

Kit	Function	Reference
RNeasy Mini	RNA isolation	Quiagen 74106
SuperScriptIII ds cDNA synthesis	cDNA synthesis	Invitrogen 11917-010
SYBR Green PCR Master Mix	expression analysis	Applied Biosystems 4367660

3.3 Oligonucleotides

Table 14: Guide-RNAs

Construct name	**CTCF motif**	**CTCF motif position**	**gRNA sequence**	**gRNA target**
*Epha4*ΔCTCF_E	GTGCCGACAGGGGGCGCCA	Chr1:75922486-75922504	AAAAACTCAGTGCCGACAGG	Chr1:75922477-75922496
*Epha4*ΔCTCF_P	TGGCCGGCAGGTGGCGCGG	Chr1:77513017-77513035	CATCTGGCCGGCAGGTGGCG	Chr1:77513020-77513039
*Epha4*ΔCTCF_B	TGGCCACCAGGTGGCGCCAT and TATTCTGCAGAGGGCATTAC	Chr1:75649269-75649287 and Chr1:75650312-75650331	GATATTAATGAGATACTTAG and GAGTTTAGTCAAAAGCTGAG	Chr1:75648868-75648887 and Chr1:75650949-75650968
*Pax*DelBΔCTCF_E	GTGCCGACAGGGGGCGCCA	Chr1:75922486-75922504	AAAAACTCAGTGCCGACAGG	Chr1: 75922477-75922496
*Pax*DelBΔCTCF_P	CTGCCGCCCGGCGGCGCAG	Chr1:78196958-78196976	GGTTGCTTGACTGCGCCGCC	Chr1: 78196948-78196967
*Pax*DelBΔCTCF_B	TGGCCACCAGGTGGCGCCAT and TATTCTGCAGAGGGCATTAC	Chr1:75649269-75649287 and Chr1:75650312-75650331	GATATTAATGAGATACTTAG and GAGTTTAGTCAAAAGCTGAG	Chr1:75648868-75648887 and Chr1:75650949-75650968

Table 15: Genotyping primers

Construct name	Forward primer	Reverse primer
*Epha4*ΔCTCF_E	GTGGGGTTCTTTCTCTTCGT	GGTGAGGCTGCTGTAGTGTT
*Epha4*ΔCTCF_P	TCCAAAAGATCCACCTTCCGT	CGTTCTCCATGACAACAACCA
*Epha4*ΔCTCF_B	F1: CCAGGGATGAAGTGGAAGTG	R1: AGCTCGCCTTGATGATGTGT
	F2: CCTTTTTGCTGATGGGATTG	R2: AGGGCGATAAAGGGATGTT
*Pax*DelBΔCTCF_E	GTGGGGTTCTTTCTCTTCGT	GGTGAGGCTGCTGTAGTGTT
*Pax*DelBΔCTCF_P	CACCGGTTGCTTGACTGCGCCGCC	AAACGGCGGCGCAGTCAAGCAACC
*Pax*DelBΔCTCF_B	F1: CCAGGGATGAAGTGGAAGTG	R1:AGCTCGCCTTGATGATGTGT

	F2: CCTTTTTGCTGATGGGATTG	R2: AGGGCGATAAAGGGATGTT

Table 16: qPCR primers

Analyzed gene	Forward primer	Reverse primer
Epha4	ATCAGCCGAAGACGGAGTAA	ATTCTCGAACTGCCTGGTTGG
Pax3	AACCCAAGCAGGTGACAACG	CTCAGGATGCGGCTGATAGA
GAPDH	GGCATTGCTCTCAATGACAA	TGTGAGGGAGATGCTCAGTG

3.4 Biological materials

Table 17: Biological materials

Materials	Description
ESCs	G4 ESCs were used for the creation of transgenic ESCs; they were stored in liquid nitrogen; originally provided by Andreas Nagy, Samuel Lunefeld Research Institute, Toronto Canada
Feeder cells	CD1 feeder cells and puromycin resistant DR4 feeder cells
Mice	All work with laboratory mice was done by the Animal Facility of the May Planck Institute for Molecular Genetics under the supervision of Dr. Ludger Hartmann, according to standards of ethics and procedures defined by local legislation through the LAGeSo Berlin. Bl6 mice were used for the breeding of mouse models. CD1 mice were used as foster animals for aggregations.
Bacteria	Top10 E. Coli bacteria were used for plasmid harnessing by transformation. They were stored in glycerol stocks at -80°C.

4. Results

4.1 Identification of target CTCF sites in wild type and *Pax*DelB ESC-lines

This work was performed on two separate backgrounds cell lines: Firstly, a wild type G4 ESC-line and secondly on the *Pax*DelB ESC-line created by Lupianez et al. 2015.

4.1.1 Identification of CTCF binding sites during limb development at the *Epha4* locus

The *Epha4* locus is characterized by a 2 Mb TAD that contains the *Epha4* gene (Chr1:77.368.039-77.517.240) and a set of limb enhancers (Chr1:77.368.039-77.517.240) located in the gene desert centromeric of that gene. Analysis of ChIP-Seq for CTCF in E11.5 mouse limb buds revealed the presence of CTCF binding sites adjacent to the set of limb enhancers, near the *Epha4* promoter and at the centromeric and telomeric TAD boundaries. HiC characterization of the *Epha4* TAD revealed:

(1.) insulating properties at both boundaries,

(2.) a loop between the CTCF sites flanking the *Epha4* promoter and the limb enhancer cluster,

(3.) a loop between the centromeric boundary and the telomeric boundary,

(4.) a loop between the centromeric boundary and the CTCF site flanking the *Epha4* promoter (see Figure 4-1).

DNA sequence analysis revealed the presence of CTCF binding motifs at the CTCF binding sites identified through ChIP-Seq, which were located at the base of the loops identified by CHiC. Furthermore, the orientation between looped CTCF binding motifs was convergent, thus in concordance with the model of loop extrusion (85).

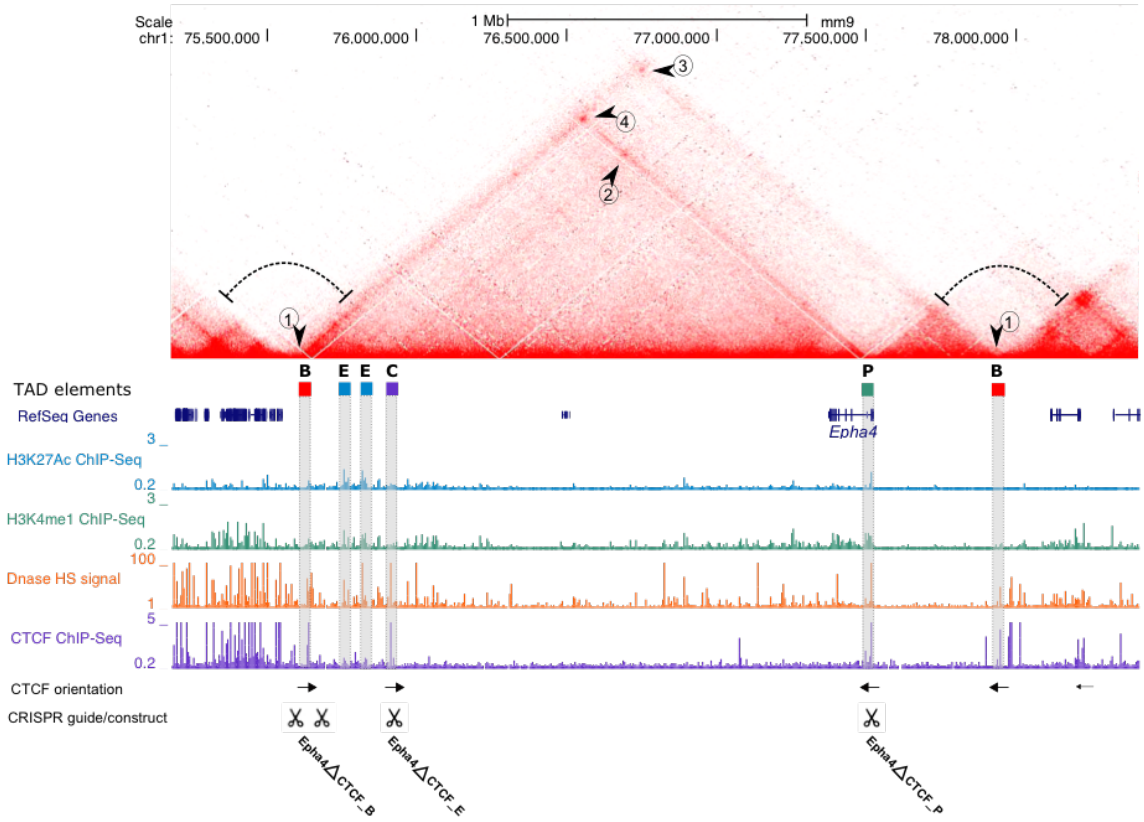

Figure 4-1: CRISPR-Cas9 *Epha4*ΔCTCF construct design. A RefSeq Gene track from UCSC browser, ChIP-Seq tracks for H3K27Ac (active enhancers), H3K4me1 (CREs in general), CTCF and a Dnase hypersensitivity signal (open chromatin mark) are aligned with the HiC map of the *Epha4* TAD in E11.5 mouse limb buds (provided by Dario Lupianez and Ivana Jerkovic). The orientation of CTCF sites is indicated by arrows below the CTCF ChIP-Seq track. Thus the crucial regulatory elements within the *Epha4* TAD are labeled: B = TAD boundary; E = distal limb enhancers; C = CTCF site flanking limb enhancer cluster; P = promoter of *Epha4* (CTCF site flanking the promoter not labeled but visible in ChIP-Seq CTCF track below). The HiC map shows (1) the insulating properties at both boundaries, (2) a loop between CTCF sites flanking P and E, (3) a loop between both boundaries, (4) a loop between the CTCF site flanking P and the centromeric boundary. The target regions of the designed guide RNAs are indicated by the scissors symbols above the construct names.

To analyze the transcriptional and regulatory effect of deleting CTCF binding sites at loop anchor points, guide RNAs were designed to target:

(1.) the CTCF motif flanking the limb enhancer cluster (construct name: *Epha4*ΔCTCF_E)

(2.) the CTCF motif flanking the *Epha4* promoter (construct name: *Epha4*ΔCTCF_P)

(3.) the CTCF cluster within the centromeric *Epha4* TAD boundary (construct name: *Epha4* ΔCTCF_B)

Identified functional CTCF motifs and corresponding gRNA sequences for the *Epha4*ΔCTCF constructs are summarized in Table 14 (see section 3.3.).

4.1.2 Identification of CTCF binding sites during limb development at the *Epha4-Pax*3 locus in *PaxDelB* mutants

In 2015 Lupianez et al. described a mouse model carrying a large homozygous deletion at the *Epha4* locus (*PaxDelB* mutants), for which a mESC line was generated. This deletion spans 1.7 Mb (Chr1:76,388,978-78,060,839), including the *Epha4* gene, *Epha4* promoter and telomeric TAD boundary. As a consequence, the remaining *Epha4* and *Pax*3 TADs are fused and ectopic loops are formed between:

(1.) the CTCF site flanking the *Pax*3 promoter and the CTCF site flanking the *Epha4* limb enhancers,

(2.) the CTCF site flanking the *Pax*3 promoter and the centromeric TAD boundary of the *Epha4* TAD,

(3.) the telomeric boundary of the *Pax*3 TAD and the CTCF site flanking the *Epha4* limb enhancers,

(4.) the telomeric boundary of the Pax3 TAD and the centromeric boundary of the *Epha4* TAD (see Figure 4-2).

The rearranged TAD results in ectopic expression of *Pax*3 (observed in E11.5 limb buds), which causes a brachydactyly phenotype (61).

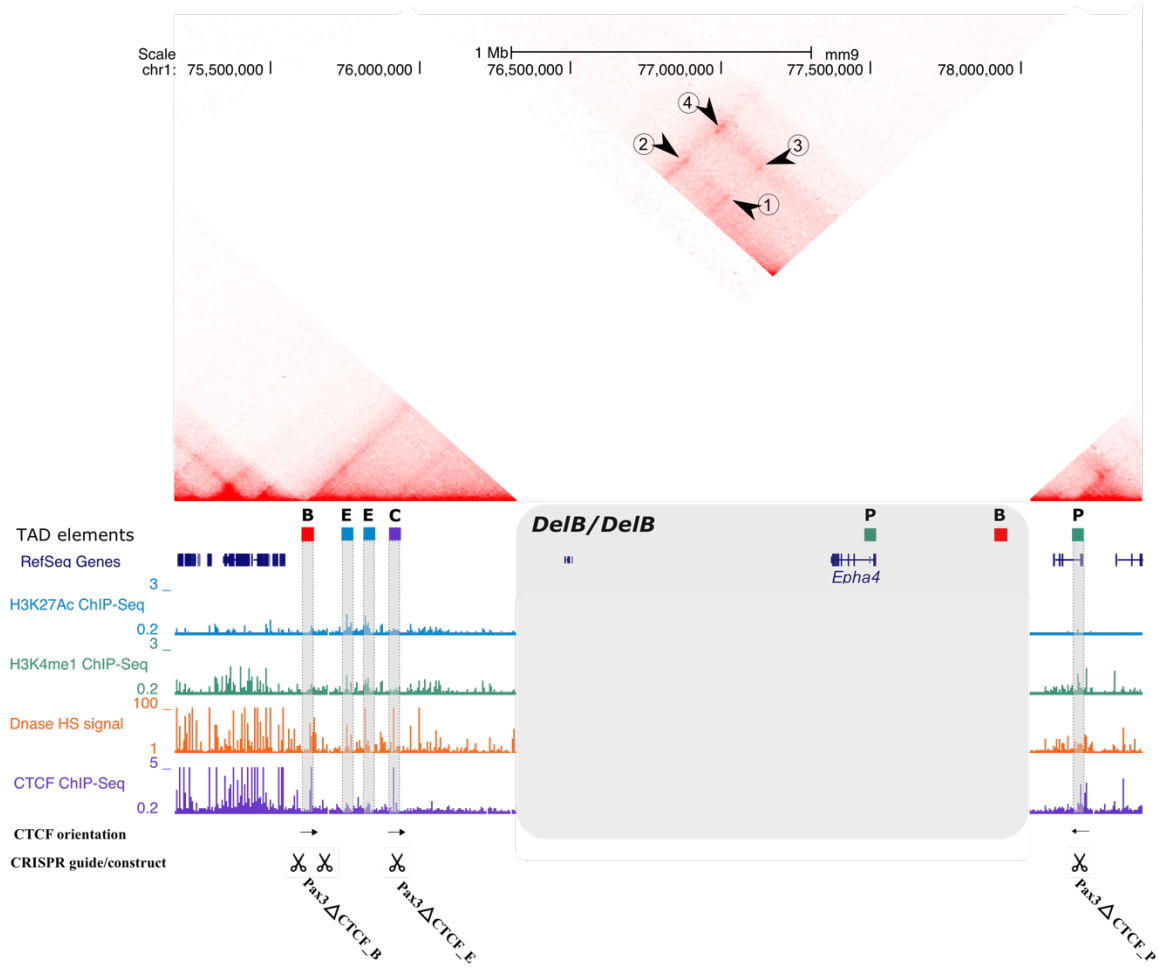

Figure 4-2: CRISPR-Cas9 *Pax*DelBΔCTCF construct design. A RefSeq Gene track from UCSC browser, ChIP-Seq tracks for H3K27Ac (active enhancers), H3K4me1 (CREs in general), CTCF and a Dnase hypersensitivity signal (open chromatin mark) are aligned with the HiC map of the fused *Epha4-Pax*3 locus stemming from E11.5 mouse limb buds from *Pax*DelB mutants (provided by Dario Lupianez and Ivana Jerkovic). The deleted section is indicated by the grey box labeled DelB/B. The orientation of CTCF sites is indicated by arrows below the CTCF ChIP-Seq track. The crucial regulatory elements within the *Epha4* TAD are labeled: B = TAD boundary; E = distal limb enhancers; C = CTCF site flanking limb enhancer cluster; P = promoter of *Epha4*/*Pax*3 (CTCF site flanking the promoters of *Epha4* and *Pax*3 are not labeled but visible in ChIP-Seq CTCF track below). The HiC map shows: (1) a loop between CTCF sites flanking P and E, (2) a loop between the CTCF sites flanking P and the centromeric *Epha4* boundary, (3) a loop between E and the telomeric *Pax3* boundary, (4) a loop between the centromeric Epha4 boundary and the telomeric Pax3 boundary. The targeted regions of the designed guide RNAs are indicated by the scissors symbols above the construct name.

To analyze the effect of deleting CTCF binding sites at loop anchors on the fused *Epha4-Pax*3 TAD and on ectopic *Pax*3 expression (for which we assume the CTCF mediated enhancer-promoter loop is critical), guide RNAs were designed to target:

(1.) the CTCF motif flanking the *Epha4* limb enhancers (construct name: *Pax*DelBΔCTCF_E)

(2.) the CTCF motif flanking the *Pax*3 promoter (construct name: *Pax*3DelBΔCTCF_P)

(3.) the CTCF cluster within the centromeric *Epha4* TAD boundary (construct name: *Pax*DelBΔCTCF_B)

Identified functional CTCF motifs and corresponding gRNA sequences for the *PaxDelB*ΔCTCF constructs are summarized in Table 14 (see section 3.3).

4.2 Sequential CRISPR-Cas9 retargeting generates combinations of CTCF-motif knockouts

A major advantage of the used CRISPR-Cas9 tool for genome engineering in ESCs is that clones can be targeted several time with varying constructs. We employed this strategy to delete both single and combined CTCF binding sites. The latter was achieved by retargeting a successfully genetically altered clone with an additional CRISPR-Cas9 construct. We generated the following combinations of CTCF knockouts:

(1.) *Epha4*ΔCTCF_P retargeted with *Epha4*ΔCTCF_E → *Epha4*ΔCTCF_P+E

(2.) *Pax*DelBΔCTCF_E retargeted with *Pax*DelBΔCTCF_P → *Pax*DelBΔCTCF_E+P

4.3 Identification of CTCF-motif deletions

CTCF binding site deletions were identified and characterized to base pair resolution using PCR based genotyping and confirmed by Sanger sequencing (see section 2.1.4). Clones were classified as ”positive” for the deletions (single CTCF motif deletions) if (1.) the entire CTCF binding motif had been deleted, (2.) the deletion was not larger than 200bp and (3.) the deletion did not overlap with active promoter regions (see section 2.1.1.1).

4.3.1 Detection of CTCF motif deletions by PCR

Homozygous deletions were successfully detected for all single gRNA constructs –*Epha4*ΔCTCF_E, *Epha4*ΔCTCF_P, *Epha4*ΔCTCF_P+E (shown in Figure 4-3) and for the constructs *Pax*DelBΔCTCF_E, *Pax*DelBΔCTCF_P and *Pax*DelBΔCTCF_E+P (shown in Figure 4-4).

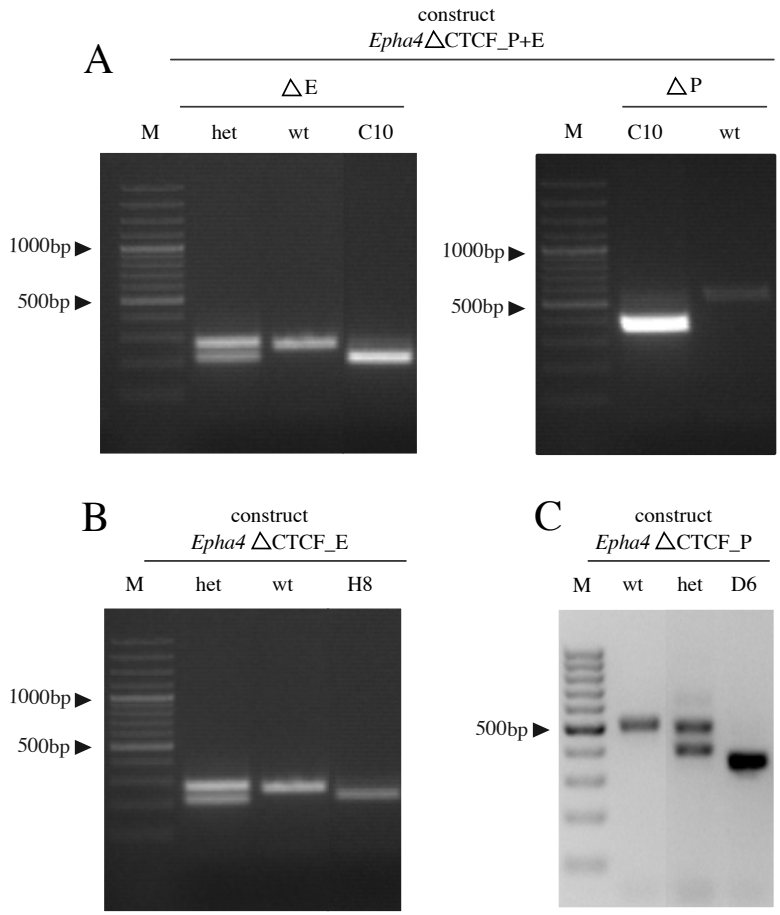

Figure 4-3: PCR based genotyping of single guide *Epha4*ΔCTCF constructs. A slight wt signal stemming from CD1 feeder cells is visible for some clones. Homozygosity of all clones was later confirmed by PCR-genotyping the embryos generated via tetraploid aggregation. DNA for wild type controls was taken from amnion tissue of E11.5 embryos. **(A)** M:100 bp+ gene ruler; C10: PCR products of the ESC clone carrying the homozygous deletions *Epha4*ΔCTCF_E and *Epha4*ΔCTCF_P; wt: wild type control; het: heterozygous control. **(B)** M: 100 bp+ gene ruler; H8: PCR product of the ESC clone carrying the homozygous deletion *Epha4*ΔCTCF_E; wt: wild type control; het: heterozygous control. **(C)** M: 100 bp gene ruler; D6: PCR product of the ESC clone carrying the homozygous deletion *Epha4*ΔCTCF_P; wt: wild type control; het: heterozygous control.

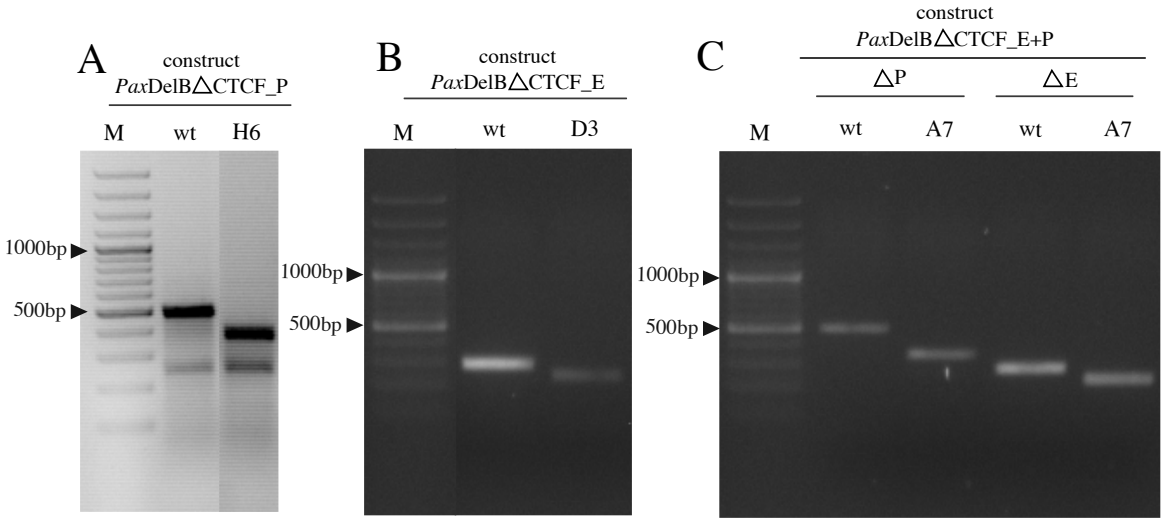

Figure 4-4: PCR based genotyping of single guide *Pax*DelBΔCTCF constructs. DNA for wild type controls was taken from amnion tissue of E11.5 embryos. **(A)** M: 100 bp+ gene ruler; H6: PCR product of the ESC clone carrying the homozygous deletion *Pax*DelBΔCTCF_P; wt: wild type control; the unspecific PCR band of 250 bp disappeared, when primer annealing temperature was increased to 57°C. **(B)** M: 100 bp gene ruler; D3: PCR product of the ESC clone carrying the homozygous deletion *Pax*DelBΔCTCF_E; wt: wild type control. **(C)** M: 100 bp+ gene ruler; PCR products of the ESC clone carrying the homozygous deletions *Pax*DelBΔCTCF_P and *Pax*DelBΔCTCF_E; wt: wild type controls.

4.3.2 Characterization of CTCF motif deletions to base pair resolution by sequencing

For all single gRNA constructs more than one "positive" clone (see 3.3) could be verified and characterized by Sanger sequencing. The following section only contains the sequencing results of those clones, which were after used for tetraploid aggregations.

4.3.2.1 Confirmation of CTCF motif deletions using *Epha4*ΔCTCF constructs

Homozygous clones carrying complete (entire sequence) CTCF motif deletions smaller than 200bp were verified by Sanger sequencing for all *Epha4*ΔCTCF constructs.

A total of 7 positive clones were identified for the *Epha4*ΔCTCF_E construct, 5 for the *Epha4*ΔCTCF_P construct and 3 for the *Epha4*ΔCTCF_E+P construct. It should be noted that out of all the clones selected and genotyped via PCR, only a select few were subsequently genotyped by sequencing. Therefore the number of actual positive clones out of the 300-500 picked clones might have been substantially higher. No conclusions can therefore be drawn from these results regarding

the efficiency of the CRISPR-Cas9 constructs in deleting the target sequences. Figure 4-5 shows the characterization of CTCF motif deletions (base pair resolution) in the three *Epha4*ΔCTCF clones, which were subsequently used for aggregations experiments.

Figure 4-5: Sanger sequencing results of *Epha4*ΔCTCF clones. The deleted sequences are shown in brackets and the precise genome coordinates of respective deletions are listed above. CTCF motifs are highlighted in green and gRNAs are shown aligned to the genome in red. The size of respective deletions varies from 53 bp (*Epha4*ΔCTCF_E_C10) over 62 bp (*Epha4*ΔCTCF_E_H8) to 155 bp (*Epha4*ΔCTCF_P_D6 and *Epha4*ΔCTCF_P_C10).

4.3.2.2 Confirmation of CTCF motif deletions using *Pax*DelBΔCTCF constructs

Homozygous clones carrying complete CTCF motif deletions smaller than 200 bp could be verified by sequencing for all *Pax*DelBΔCTCF constructs.

A total of 5 positive clones were identified for the *Pax*DelBΔCTCF_E construct, 4 for the *Pax*DelBΔCTCF_P construct and 2 for the *Pax*DelBΔCTCF_E+P construct. Note that, as for *Epha4*ΔCTCF clones, only a selected few were genotyped by sequencing.

Figure 4-6 shows the characterization of CTCF motif deletions (base pair resolution) in the three *Pax*DelBΔCTCF clones, which were subsequently used for aggregation experiments.

Figure 4-6: Sanger sequencing results of *Pax*DelBΔCTCF clones. The deleted sequences are shown in brackets and the precise genome coordinates of respective deletions are listed above. CTCF motifs are highlighted in green and gRNAs are shown aligned to the genome in red. The size of respective deletions varied from 49 bp (*Pax*DelBΔCTCF_E_D3 and *Pax*DelBΔCTCF_E_A7) over 114 bp (*Pax*DelBΔCTCF_P_H6) to 153 bp (*Pax*DelBΔCTCF_P_A7).

4.4 Generation of mutants carrying CTCF motif deletions via tetraploid aggregations

While expanding (see 2.1.3.6), clones were monitored for two characteristic features: the speed with which they grew and the morphology of the growing clones. After expansion of clones, the choice of which clone should be used for aggregation reflected a compromise of three desirable clone properties:

(1.) a preferably small CTCF motif deletion (so as to only impair CTCF binding and not delete further sequences which might be of functional importance)

(2.) a moderate speed of growth - neither those clones which had been very slow nor those which had been very fast in growing during expansion were used for aggregation

(3.) an undifferentiated morphology - clones with a round and smooth surface and clearly separated individual clones

Clones of the different constructs used for tetraploid aggregations are summarized in Table 19.

Table 18: Summary of aggregated clones

Construct	Number of positive clones	Aggregated clone	Homozygous/ heterozygous	Deletion size (bp)	Deletion localization
*Epha4*ΔCTCF_E	7	H8	Hom.	62	Chr.1: 75922469-75922530
*Epha4*ΔCTCF_P	5	D6	Hom.	155	Chr.1: 77512965-77513119
*Epha4*ΔCTCF_P+E	3	C10	Hom.	53 (E) + 155 (P)	Chr.1: 75922457-75922509, Chr.1: 77512965-77513119
*Pax*DelBΔCTCF_E	5	D3	Hom.	49	Chr.1: 75922468-75922516
*Pax*DelBΔCTCF_P	4	H6	Hom.	114	Chr.1: 78196867-78196980
*Pax*DelBΔCTCF_E+P	2	A7	Hom.	49 (E) + 153 (P)	Chr.1: 75922468-75922516 Chr.1: 78196902-78197054

E11.5 embryos were generated from aggregated clones to perform gene expression analyses (qPCR and in situ hybridization) as well as to study 3D chromatin organization (CHiC). To evaluate potential phenotypical effects of CTCF deletions, additional aggregations were performed on PaxDelBΔCTCF mutants and collected at E17.5. However, these experiments did not yield any mutant fetuses. Table 20 summarizes the clones used for aggregations.

Table 19: Summary of animals obtained from aggregations

Construct	Aggregated clone	Homozygous/ heterozygous	E11.5	E17.5
*Epha4*ΔCTCF_E	H8	Hom.	✓	-
*Epha4*ΔCTCF_P	D6	Hom.	✓	-
*Epha4*ΔCTCF_E+P	C10	Hom.	✓	-
*Pax*DelBΔCTCF_E	D3	Hom.	✓	-
*Pax*DelBΔCTCF_P	H6	Hom.	✓	x
PaxDelBΔCTCF_E+P	A7	Hom.	✓	x

Table 20: Summary of animals obtained by tetraploid aggregations of the different constructs. The table does not indicate the quantity of animals obtained, but merely whether or not it was at all possible to obtain animals at different stages through tetraploid aggregations. Successful aggregations are indicated by a ✓. Unsuccessful aggregations are indicated by an x. If no aggregation attempt was made, this is indicated by a - .

4.4 Gene expression analyses

The aim of this work was to analyze the effects of CTCF binding site deletions at loop anchor points on TAD organization and transcriptional output. For the latter those CTCF sites at loops associated with enhancer-promoter interactions were of particular interest. The transcriptional effects of CTCF motif deletions were analyzed quantitatively by qPCR and spatially by *in situ* hybridization.

4.4.1 Altered gene expression in mutants carrying deletions of CTCF binding sites at promoters

In wild type E11.5 embryos, *Epha4* is physiologically expressed in the distal portion of the developing mouse limb. We analyzed potential alterations of this pattern in three different homozygous *Epha4*ΔCTCF mutants (ΔCTCF_E, ΔCTCF_P and ΔCTCF_P+E) in developing limbs by qPCR (summarized in Figure 4-7).

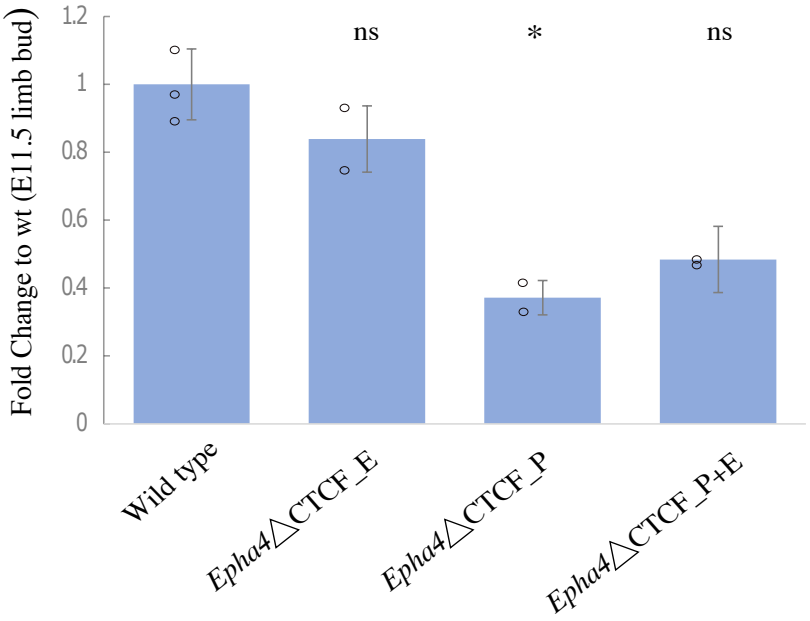

Figure 4-7: *Epha4* qPCR analysis in homozygous *Epha4*ΔCTCF mutants and wt. Results are normalized to GAPDH expression and wild type = 1. Bars represent the mean of replicas, arrow bars represent standard deviation and circles indicate individual biological replicas. Significance was analyzed using a one-sided, unpaired t-test (n=2-3 with 3 technical replicas each; P* <0.05, ns=not significant).

In mutants carrying a CTCF binding site deletion near the *Epha4* enhancers (*Epha4*ΔCTCF_E), *Epha4* expression is not changed when compared to wild type controls. However, in mutants carrying deletions affecting CTCF sites at the *Epha*4 promoter (*Epha4*ΔCTCF_P), we observed a significant reduction of around 50% of *Epha4* expression. A similar effect is also observed in double mutants (*Epha4*ΔCTCF_P+E). [1]

We next analyzed the effect of CTCF binding site deletions in a *Pax*DelB background. *Pax*DelB mutants carry a large deletion that results in the fusion of the *Epha4* and *Pax3* TADs. As a consequence, *Pax3*, which is lowly expressed in E11.5 wildtype animals, is misexpressed in the distal portion of the limb. This is the result of an ectopic interaction of the *Epha4* limb enhancer with the *Pax*3 promoter and causes a brachydactyly phenotype. We studied the effects of homozygous CTCF binding site deletions in limbs of *Pax*DelB mutants (ΔCTCF_E, ΔCTCF_P and ΔCTCF_E+P) by

[1] The results are not or not strongly significant. As they are strong in effect, the lack of significance is probably due to low case numbers (n=2). Further biological replicas are needed to prove significance the results.

qPCR. As control, results were compared with mutants of the original *Pax*DelB background (summarized in Figure 4-8).

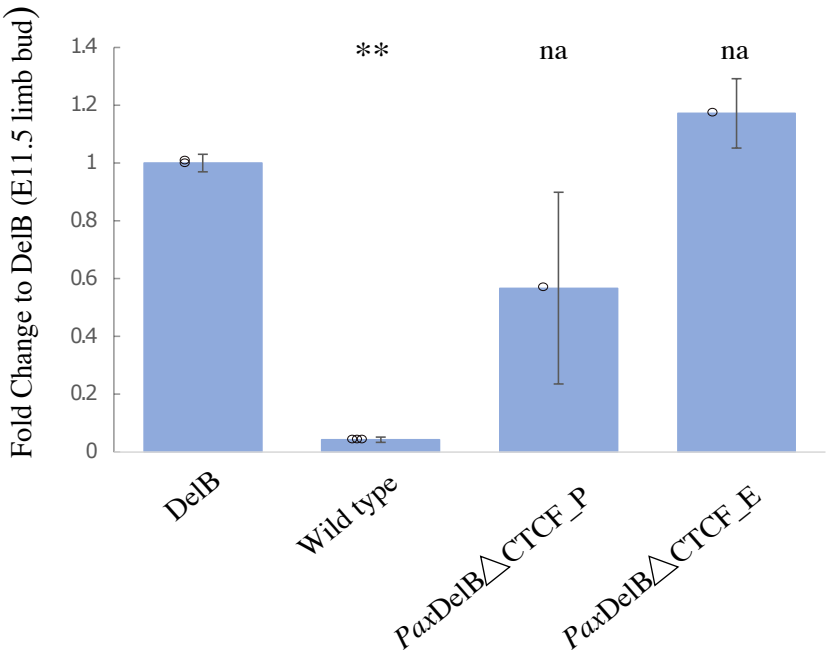

Figure 4-8: *Pax3* qPCR analysis in homozygous *Pax*DelBΔCTCF mutants and wt. Results are normalized to GAPDH expression and DelB = 1. Bars represent the mean of replicas, arrow bars represent standard deviation and circles indicate individual biological replicas. Significance was analyzed using a one-sided, unpaired t-test (n=1-3 with 3 technical replicas each; P** <0.0004, na=not applicable because n=1).

While this analysis could only be performed in one embryo per mutant, we observed similar results as for *Epha4*ΔCTCF animals: The deletion of a CTCF binding site at the enhancer cluster (*Pax*DelBΔCTCF_E) has no noticeable effect on gene expression. In contrast, deletions of CTCF binding sites near the Pax3 promoter showed a 50% reduction in gene expression. However, further biological replicas are needed to prove significance of these results.

In summary, we observed a loss of 50% of physiological gene expression in mutants carrying deletions of CTCF sites associated to promoters.

4.4.2 Spatial gene expression patterns are largely unaffected upon deletion of CTCF binding sites

As *Epha4* is prominently expressed in the distal portion of both fore and hindlimbs at E11.5, we sought

to evaluate if mutants carrying CTCF binding site deletions display alterations in this spatial expression pattern. To evaluate this, we performed in situ hybridization in E11.5 mutants. (As pointed out in the methods section this work was kindly done by Norbert Brieske.)

Analysis of in situ hybridizations (Figure 4-9) revealed that the expression pattern of *Epha4* is largely unaffected in mutants that carry deletions of CTCF sites associated to promoters or enhancers (*Epha4*ΔCTCF_P and *Epha4*ΔCTCF_E respectively). These results show that the reduced *Epha4* expression observed by qPCR in mutants carrying CTCF deletions at promoters is a quantitative, and not qualitative, reduction of gene expression.

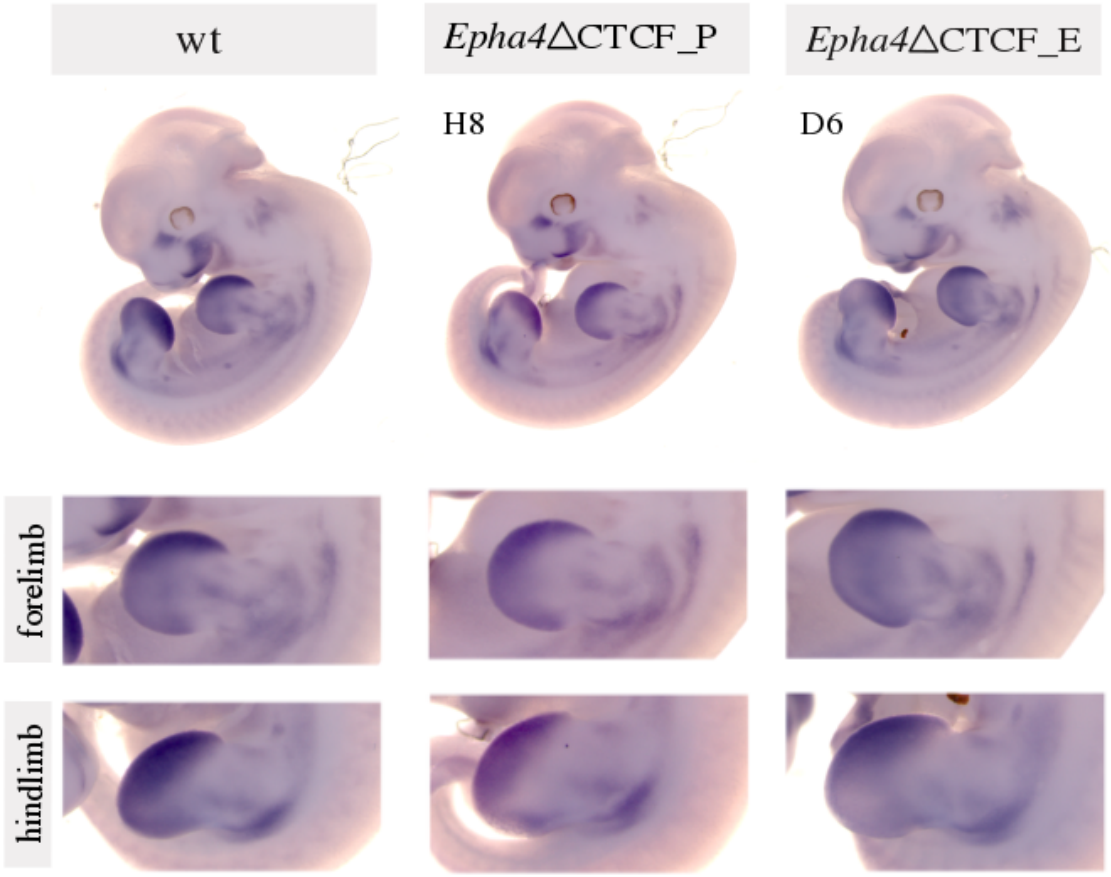

Figure 4-9: In situ hybridization for *Epha4* in *Epha4*ΔCTCF mutants. Whole-mount in situ hybridization for *Epha4* in E11.5 wild type, homozygous *Epha4*ΔCTCF_P and *Epha4*ΔCTCF_E embryos. The zoom-in highlights expression patterns in forelimb and hindlimb. *Epha4* expression patterns are largely unchanged in CTCF mutants.

Next, we evaluated the effect on the spatial expression pattern of *Pax3* in the CTCF mutants of the *Pax*DelB background. Although *Pax*3 is not expressed physiologically in the developing limb, *Pax*DelB mutants display ectopic *Pax3* expression that recapitulates the spatial pattern observed for *Epha4*. This results from *Pax3* being regulated by *Epha4* limb enhancers as a consequence of a large deletion that fuses the *Epha4* and *Pax3* TADs. Expression analysis by in situ hybridization at E11.5 (Figure 4-10) revealed that the expression pattern of *Pax3* is largely unaltered in mutants that carry deletions of CTCF sites associated to enhancers or promoters (*Pax*DelBΔCTCF_P, *Pax*DelBΔCTCF_E and *Pax*DelBΔCTCF_P+E). These results confirm that the reduced *Pax3* expression observed by qPCR in mutants carrying CTCF deletions at promoters originate from quantitative, and not qualitative, differences in gene expression.

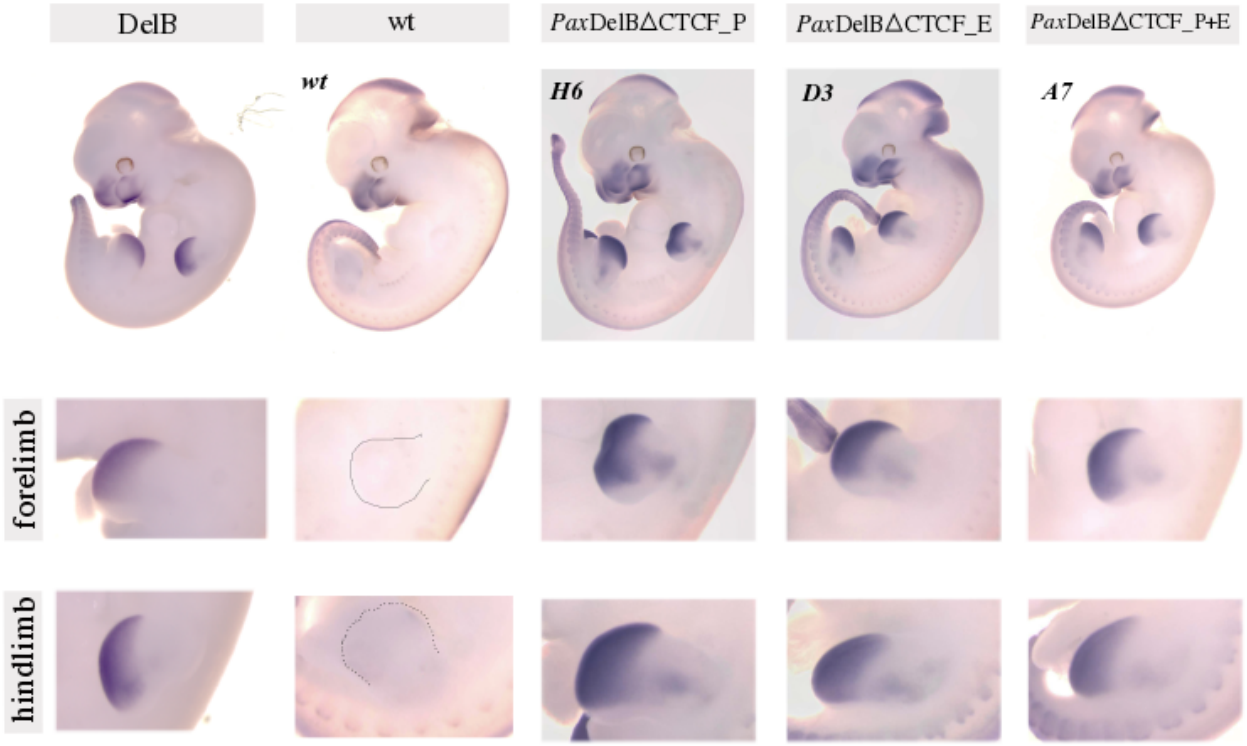

Figure 4-10: In situ hybridization for *Pax*3 in *Pax*DelBΔCTCF mutants. Whole-mount in situ hybridization for *Pax3* in E11.5 wt, homozygous DelB, *Pax*DelBΔCTCF_P, *Pax*DelBΔCTCF_E and *Pax*DelBΔCTCF_P+E embryos. The zoom-in highlights expression patterns in forelimb and hindlimb. *Pax3* expression patterns are largely unaltered in CTCF mutants.

4.5 Alterations in 3D chromatin organization upon deletion of promoter-associated CTCF sites

Due to the prominent gene expression changes in mutants that carry CTCF deletions at gene promoters, we next investigated potential alterations in 3D chromatin organization by CHiC. Specifically, we focused on the mutant where the CTCF binding site at the *Epha4* promoter is deleted (*Epha4*ΔCTCF_P), for which we performed CHiC experiments on distal limb tissue on day 11.5. Figure 4-11 shows a wt HiC map of the *Epha4* locus whereas Figure 4-12 shows the CHiC map generated from *Epha4*ΔCTCF_P mutants.

The analysis of the wt control map revealed characterizing architectural features of the *Epha4* locus:

(1.) the self-associating and insulating property of the *Epha4* TAD generated by the centromeric and telomeric boundary;

(2.) a CTCF-mediated chromatin loop (convergently oriented) between both boundary elements;

(3.) a CTCF-mediated chromatin loop (convergently oriented) connecting the *Epha4* enhancer and promoter regions (CTCF_P and CTCF_E, respectively);

(4.) a CTCF-mediated chromatin loop (convergently oriented) between the *Epha4* promoter region and the centromeric boundary;

(5.) a chromatin interaction trail that connects the *Epha4* promoter with the centromeric region of the *Epha4* TAD; and

(6.) no chromatin loop was detected between the CTCF binding site flanking the *Epha4* enhancer (CTCF_E) and the CTCF sites within the telomeric boundary.

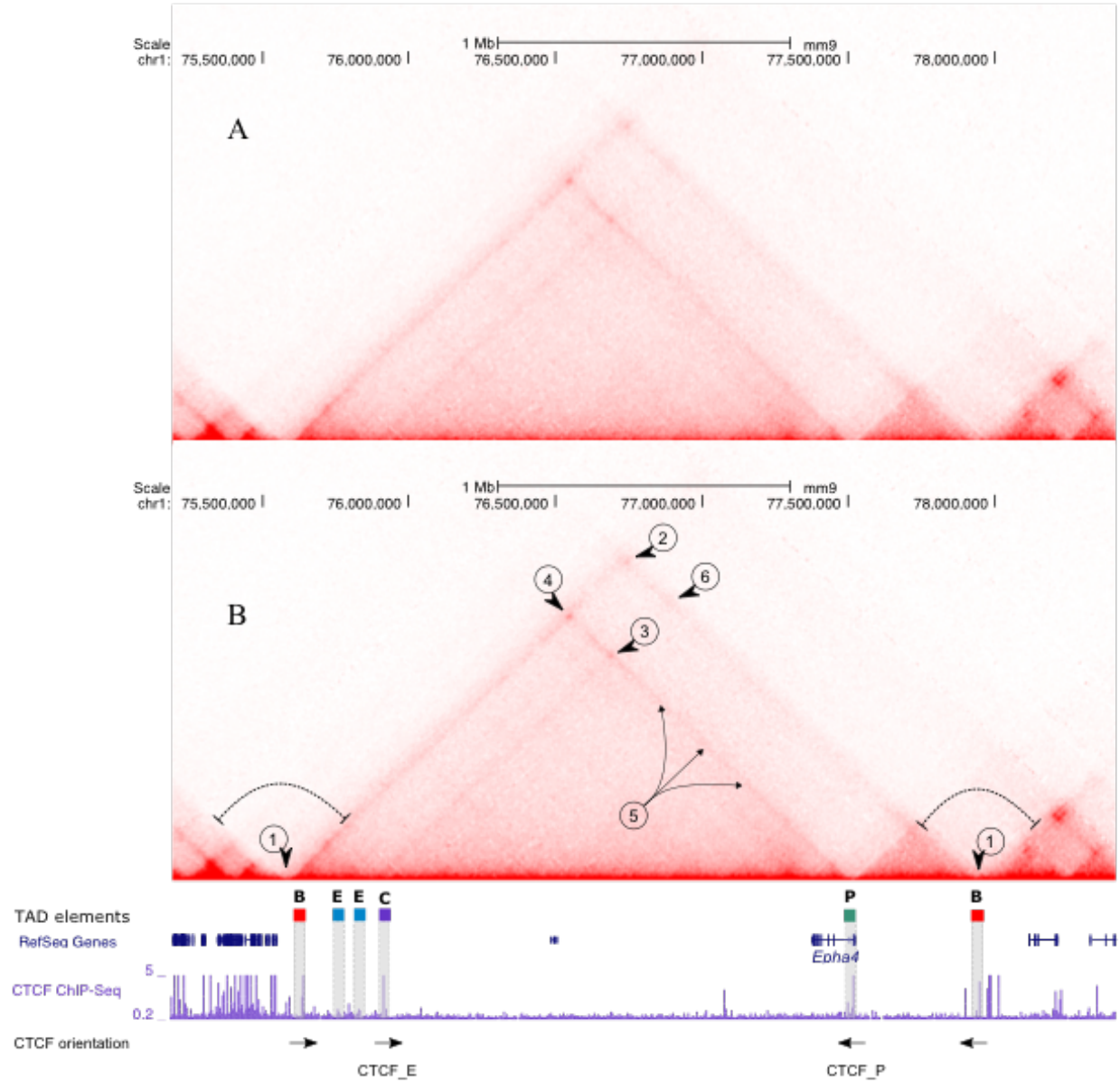

Figure 4-11: CHiC characterization of the *Epha4* locus in wt. CHiC maps from wt E11.5 distal limb samples. A RefSeq track from USCS browser and a ChIP-Seq track for CTCF are aligned with the CHiC map. Functionally important elements of the *Epha4* TAD are labeled below: B: boundary elements; E: limb enhancers; C: CTCF binding site adjacent to enhancers; P: promoter. **A:** uncommented CHiC map for better visualization. **B:** CHiC map with marked architectural characteristics: (1) insulating properties TAD boundaries; (2) CTCF-mediated chromatin loop between both boundary elements; (3) CTCF-mediated loop connecting the *Epha4* enhancer and promoter regions; (4) CTCF-mediated loop between the CTCF binding site flanking the *Epha4* promoter and the centromeric boundary; (5) chromatin interaction

trail between the *Epha4* promoter an the centromeric boundary; (6) lack of chromatin loop between the CTCF site flanking the *Epha4* enhancer and the telomeric boundary.

The analysis of CHiC mutant maps revealed specific differences in genomic architecture when compared to the wt control map:

(1.) the self-associating and insulating property of the *Epha4* TAD generated by the centromeric and telomeric boundary was unchanged;

(2.) the CTCF-mediated chromatin loop between both boundary elements was unaffected;

(3.) the CTCF-mediated chromatin loop connecting the *Epha4* enhancer and promoter regions was disrupted;

(4.) the CTCF-mediated chromatin loop between the *Epha4* promoter region and the centromeric boundary was disrupted;

(5.) the chromatin interaction trail from the *Epha4* promoter to the centromeric region of the *Epha4* TAD was lost; and

(6.) contact frequency between the CTCF site flanking the *Epha4* enhancer and the telomeric boundary is increased when compared to wild type – a new chromatin loop is formed in *Epha4*ΔCTCF_P mutants.

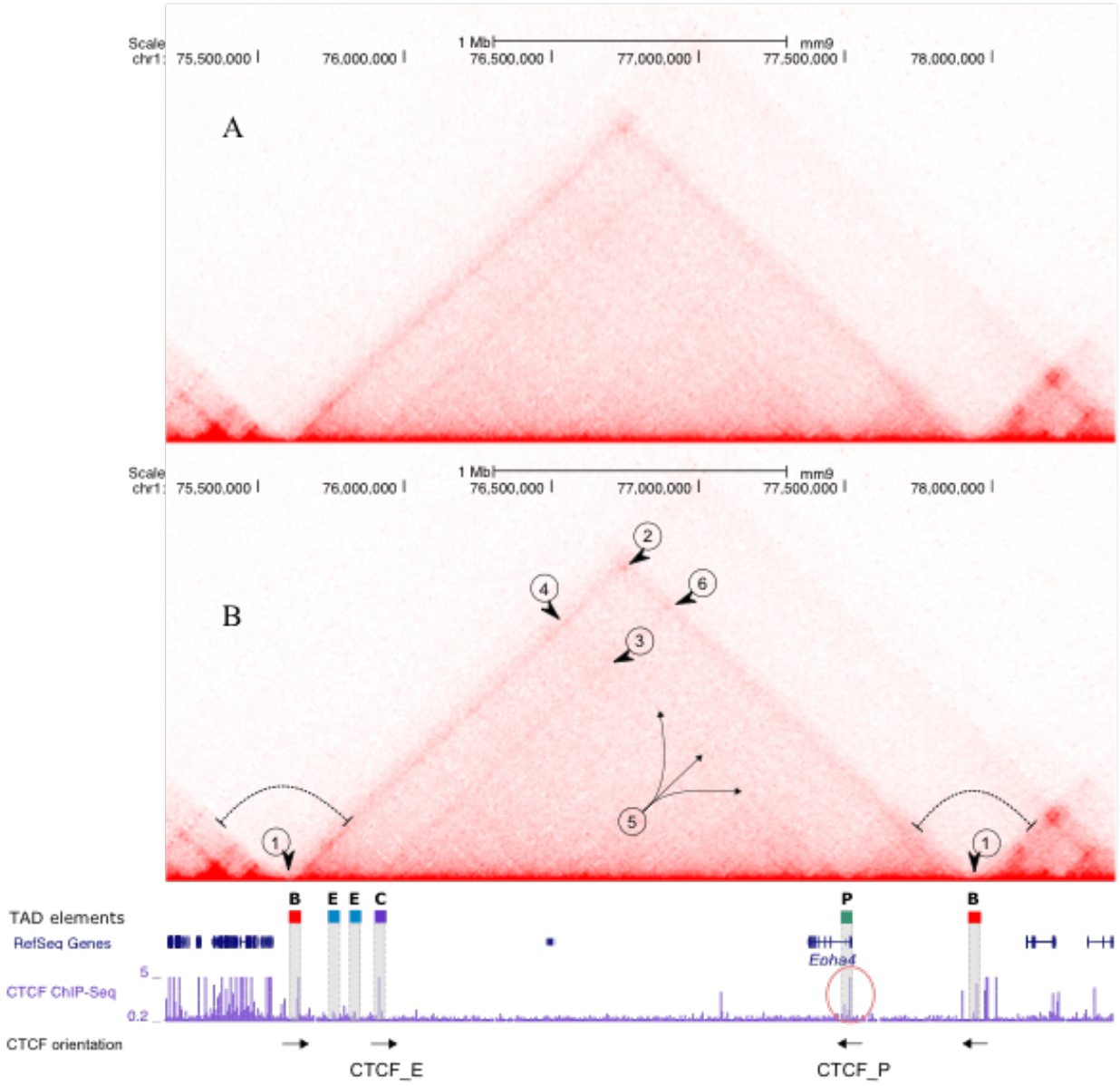

Figure 4-12: HiC characterization of the *Epha4* locus in *Epha4*ΔCTCF_P mutants. CHiC maps from *Epha4*ΔCTCF_P E11.5 distal limb samples. A RefSeq track from USCS browser and a ChIP-Seq track for CTCF are aligned with the CHiC map. Functionally important elements of the *Epha4* TAD are labeled below: B: boundary elements; E: limb enhancers; C: CTCF binding site adjacent to enhancers; P: promoter; the deleted CTCF site adjacent to the *Epha4* promoter is encircled in red. **A:** uncommented CHiC map for better visualization. **B:** CHiC map with marked architectural characteristics: (1) unchanged insulating properties of TAD boundaries; (2) unaffected CTCF-mediated chromatin loop between both boundary elements; (3) disruption of the CTCF-mediated loop connecting the *Epha4* enhancer and promoter regions; (4) disruption of the CTCF-mediated loop between the CTCF binding site flanking the *Epha4* promoter and the centromeric boundary; (5) loss of the chromatin interaction trail between the *Epha4* promoter an the centromeric boundary; (6) increased contact frequency between the CTCF site flanking the *Epha4* enhancer and the telomeric boundary.

5. Discussion

CTCF is a crucial player in the orchestration of genome architecture: Convergently orientated CTCF binding sites seem to be set at the very core of the formation of both TADs and long range contacts (loops) as indicated by an abundance of both functional and genome wide correlation studies (46, 61, 79, 80). The formulation of the loop extrusion model has provided a thorough theoretical, computationally derived model for explaining these results (39, 85). Meanwhile, the functional aspects of genome architecture (and thus of CTCF) for gene regulation have stepped into the limelight of scientific interest: Physical enhancer-promoter interactions seem to be a prerequisite for driving gene expression and the mechanism by which distal enhancer elements achieve proximity to their promoters is deemed to be CTCF mediated looping (102). The working model for explaining the functional implications chromatin looping has on gene regulation therefore is:

(1.) convergently orientated CTCF binding sites facilitate chromatin looping and

(2.) chromatin looping facilitates the interaction between promoters and distal enhancer elements to drive gene expression.

This works's primary aim was to experimentally challenge this working model by deleting CTCF binding sites. In the following sections, I summarize the effects CTCF binding site deletions had on gene expression and genome architecture (section 5.1), discuss how these results can be interpreted (5.2) and finally reevaluate the questions poised at the beginning of this work with the obtained results (section 5.3).

5.1. CTCF binding sites show distinct hierarchy in transcriptional control

The effects of different CTCF motif deletions on gene expression and genome architecture were highly similar in experiments performed on a wild type and the *Pax*DelB background: In both mouse models a deletion of the CTCF site flanking the promoter of the gene interacting with the *Epha4* limb enhancer cluster caused a 50% loss of expression (respectively, *Epha4* in wt and *Pax*3 in the DelB background). A similar degree of reduced expression was observed in mouse models that combined deletions of CTCF sites at promoters and near the *Epha4* limb enhancer cluster. However, mutant embryos carrying single CTCF deletions near the enhancer cluster did not show significant changes

in gene expression levels. Therefore, the 50% reduction of gene expression observed in the double CTCF knockout mutants (ΔCTCF_P+E) is probably the consequence of CTCF deletions at promoters (ΔCTCF_P).

To assess whether the observed changes in gene expression levels were the consequence of alterations in genome architecture caused by CTCF binding site deletions, CHiC experiments were performed on samples (E11.5 limb buds) taken from *Epha4*ΔCTCF_P embryos. The CHiC experiments revealed a prominent reorganization of the *Epha4* TAD with selective impairment of chromatin loops. Specifically, the loop associated to enhancer-promoter interaction was completely abolished. Furthermore, the chromatin loop between the CTCF site flanking the promoter and the centromeric boundary was also lost. In the absence of the CTCF_P site, however, a new chromatin loop between the CTCF site near the limb enhancer cluster and the telomeric boundary appeared.

So how can the presented results be interpreted? What possible flaws in the experimental setup must be considered? Are the results in line with the model of loop extrusion or do they contradict it? What questions arise from the results and what further steps would have to be taken to answer them?

5.2 Interpreting changes of gene expression and genome architecture in ΔCTCF mutants

As the results of experiments performed on wt or on DelB background were highly similar, in the following, results summarized under the name ΔCTCF_P contain the results from mutants carrying CTCF deletions associated to promoters (*Epha4*ΔCTCF_P and *Pax*DelBΔCTCF_P). Those summarized under ΔCTCF_E contain the results from mutants carrying CTCF deletions associated to the enhancer cluster (*Epha4*ΔCTCF_E and *Pax*DelBΔCTCF_E).

The two major questions arising from the results and the two major challenges in interpreting the results are the following:

(1.) Why is gene expression differentially affected upon CTCF deletions at promoters or enhancers (see 5.2.1)?

(2.) How is transcription still occurring in the absence of enhancer-promoter chromatin loops (see 5.2.2)?

To answer these questions it first has to be pointed out that it was not formally tested whether 3D chromatin organization is altered in mutants carrying CTCF deletions flanking the enhancers (*Epha4*ΔCTCF_E), as no CHiC experiment were performed in these mutants. It would be expected, however, that the deletion of the CTCF binding site would disrupt the enhancer-promoter chromatin loop to a comparable extend. If indeed the loop was disrupted and expression levels are not impaired, this would be an intriguing finding (possible explanations are discussed in section 5.2.2). In the following section, I will base the discussion on that assumption.

Further, it has to be pointed out, that there is no "waterproof" way of ensuring that by deleting CTCF binding sites associated to promoters, only the binding of CTCF and not promoter function itself is impaired. The reason for this is that as a functional unit promoters are still rather vaguely defined. In this work the region of active promoters was defined by ChIP-Seq tracks for H3K4me3 histone modifications in the developing limb in E11.5 mice. By keeping CTCF deletions at a minimal size and ensuring that deletions did not overlap with the H3K4me3 signal, it was sought to ensure that active promoters were not impaired. Still the results obtained from mutants carrying CTCF motif deletions flanking the promoter (ΔCTCF_P) have to be interpreted with the thought in mind, that promoter function impairment can at this point not be ruled out entirely but merely made unlikely.

5.2.1 Compensatory sub-TAD interactions might compensate the disruption of enhancer-promoter chromatin loops

Two alternative hypotheses might explain the gene expression differences between mutants carrying promoter or enhancer CTCF deletions. The first hypothesis considers to the possibility that other existing CTCF-mediated loops might compensate the loss of enhancer-promoter loops. In our *Epha4*ΔCTCF_P mutant we observe that not only the enhancer-promoter loop (marked as 3 in Figures 4.11 and 4.12) but also other interactions within the *Epha4* TAD are affected. Specifically, the chromatin loop that connects the promoter CTCF site and the centromeric boundary is also abolished (marked as 4 in Figures 4.11 and 4.12). Since the linear distance between the limb enhancer cluster

and the centromeric boundary is relatively small (150 kb vs the 2 Mb of the entire *Epha4* TAD), it is plausible that the boundary-promoter loop might compensate for the disruption of the enhancer-promoter loop. Since we do not expect this chromatin loop to be disrupted in ΔCTCF_E mutants, the drastic difference in gene expression between ΔCTCF_E and ΔCTCF_P mutants might be a result of this loop acting as a compensatory mechanism and buffering *Epha4* expression. To formally test this hypothesis, a deletion at the centromeric boundary should be induced by retargeting our ΔCTCF_E mutants and gene expression levels subsequently be analyzed.

The second hypothesis relates to the de novo appearance of chromatin loops in the ΔCTCF_E mutants. In these mutants the CTCF site at the *Epha4* promoter remains intact and could therefore engage in a new loop with previously unidentified enhancers. Indeed, ChIP-Seq tracks for CTCF in E11.5 limb buds reveal a CTCF site approximately 450 kb downstream of the *Epha4* promoter (see Figure 4.1). Furthermore, CHiC maps in wt and mutants limbs demonstrate that the CTCF site at the *Epha4* promoter facilitates interactions with the entire TAD (interaction trail, marked as 5 in Figure 4-11), potentially allowing *Epha4* to "scan" its entire regulatory landscape. ChIP-Seq tracks for H3K27ac show other regions displaying active enhancer marks within the *Epha4* TAD that might engage in functional interactions in the course of this scanning process (again see Figure 4.1). These elements might form a compensatory interaction that could maintain *Epha4* expression levels, similar to those observed in wt animals. To evaluate this hypothesis, CHiC experiments would have to be performed on the developing limb of ΔCTCF_E mutants.

However, we observe similar gene expression differences between ΔCTCF_P and ΔCTCF_E mutants in both the wt and the DelB background. This postulates the first hypothesis as more favorable, since *DelB* mutants lack a significant portion of the *Epha4* TAD and, therefore, most additional cis-regulatory elements.

5.2.2 Transcription still takes place in the absence of CTCF-mediated intra-TAD loops

The chromatin loop facilitating enhancer-promoter interactions is disrupted completely in ΔCTCF_P mutants, as shown in the CHiC experiments (marked as 3 in Figure 4-12). The loop connecting *Epha4* enhancers and the *Pax3* gene is also expected to be abolished in the deletions performed in the DelB

background. So how are expression levels maintained at least at 50%?

A plausible explanation relates to the hypotheses presented in section 5.2.1: In ΔCTCF_P mutants it seems possible, that a more proximal enhancer element compensates the loss of proximity between the distal limb-enhancer cluster and promoters. In ΔCTCF_P mutants this *de-novo* enhancer-promoter interaction should not be facilitated by a CTCF mediated loop, as the CTCF_P site is deleted. In accordance with this, HiC experiments performed on those animals show no new loop formations. It is, however, perfectly possible that a proximal enhancer element yields sufficient enhancer-promoter proximity to maintain 50% of gene expression without any looping occurring. Again there is the same strong objection to this argument already raised in section 4.2.2: In animals from the DelB background all more proximal enhancer elements possibly driving *Epha4* expression are deleted by the DelB mutation.

So how else could 50% of expression in ΔCTCF_P animals be maintained? Two further explanation seem possible: Firstly, in ΔCTCF_P mutants the CTCF binding site flanking the limb enhancer cluster is still present. When a CTCF-Cohesin complex binds to chromatin and begins extruding DNA, extrusion ceases at this CTCF site. Meanwhile, extrusion progresses towards the telomeric side to CTCF_E and allows the enhancer cluster to interact with genomic regions until reaching the next convergently orientated CTCF site (which in ΔCTCF_P mutants are located within the telomeric boundary), thus including the *Epha4* promoter. In the mutant CHiC maps (see figure 4-12), this interaction trail is visible as a line of increased contact frequency forming from the CTCF site CTCF_E up to the telomeric *Epha4* TAD boundary. Therefore, in ΔCTCF_P mutants a certain degree of interactions (higher than the baseline intra-TAD interaction) occur between the target promoter and distal limb enhancer cluster. However, since loop extrusion no longer ceases at the promoter flanking CTCF site, no enhancer-promoter loop is formed.

However, there is a strong evidence that challenges this line of argument: If the enhancer-promoter proximity generated by continued loop-extrusion from CTCF_E was responsible for maintaining expression in ΔCTCF_P animals, then the expression levels should be further reduced when both CTCF sites are simultaneously deleted. However, our double knockout models (ΔCTCF_P+E) demonstrate that the loss in expression is similar to single promoter CTCF deletions.

We favor the hypothesis that TADs themselves achieve a certain amount of interaction of any elements contained within (as shown in section 4.1). This basic TAD-interaction profile is generated by the boundary elements and is unaltered by the deletion of intra-TAD CTCF sites (see figure 4-12). The basic contact frequency in ΔCTCF_P animals might be sufficient to maintain basal levels of gene expression by achieving a low level of proximity between the distal limb enhancers and promoters. In order to address this question ΔCTCF_P mutants should be retargeted to subsequently delete the boundary elements of the *Epha4* TAD.

Several recent studies examining the effects of CTCF site deletions on chromatin architecture and gene expression, have also demonstrated relative robustness of gene expression to perturbation of enhancer-promoter loops and TAD architecture: By engineering a series of CTCF site deletions in the *Shh* locus Williamson et al. 2019 (103) demonstrate that while CTCF deletions alter TAD architecture and disrupt spatial proximity of *Shh* to its limb enhancer, the ZRS, *Shh* expression patterns and expression quantity remains largely unchanged. Similarly, Paliou et al. 2019 delete three CTCF sites flanking the ZRS enhancer. While distance between *Shh* and the ZRS is increased, expression of *Shh* is retained at 51%, suggesting that, although maximal expression levels of *Shh* depend on CTCF mediated enhancer-promoter proximity, other mechanisms contribute strongly to long-range enhancer-promoter interactions and robustness of *Shh* expression (84). Despang et al. 2019 show that combining deletions of intra-TAD and TAD boundary CTCF sites at the *Sox9-Kcnj2* locus results in complete fusion of neighboring TADs and disruption of normal chromatin architecture. Despite these dramatic alterations in chromatin architecture, only minor changes of gene expression and no phenotypic effects were observed (64). Interestingly, combining a boundary deletion with an intra-TAD inversion did result in ectopic misexpression and a pathological phenotype, indicating the importance of the orientation of TAD substructure in facilitating enhancer-promoter interactions(64).

5.3 Reevaluating starting questions

The experiments designed and performed in this work set out to tackle two major questions: Firstly, to investigate the widely accepted assumption that two convergently orientated CTCF sites are a necessity for loop formation and, thus, for transcription driven by distal enhancer elements. Secondly,

to evaluate if the pathological brachydactyly phenotype observed in the *Pax*3DelB mouse model can be reverted upon further alteration of 3D chromatin organization. In the following I will evaluate what implications the presented results have to these outgoing questions.

5.3.1 The deletion of a CTCF site associated to the *Pax3* promoter is embryonically lethal

We aimed to employ CRISPR-Cas9 genome engineering and our knowledge of 3D genomic architecture to reverse a complex pathological phenotype. The *Pax*DelB deletion removes a boundary element and rewires the *Epha4* limb enhancers towards the *Pax*3 promoter, causing an ectopic expression of *Pax*3 in the developing limb and a brachydactyly phenotype. This work sought to determine whether it is possible to unwire this ectopic interaction by deleting the CTCF motifs flanking the limb enhancer cluster and/or flanking the *Pax*3 promoter, thus preventing gene misexpression and the corresponding brachydactyly phenotype. This experiment substantially reduced half of *Pax*3 expression in the distal limb on day E11.5 in mutants carrying CTCF deletions associated to the promoter (*Pax*DelBΔCTCF_P and *Pax*DelBΔCTCF_P+E mutants). Therefore, it seemed possible that these animals would exhibit a reversal or weakening of the brachydactyly phenotype.

To assess the development of a brachydactyly phenotype, mice have to be at least 17.5 days old, for which we employed tetraploid aggregations of the corresponding mutant mESC clones. While embryos were viable and in reasonable numbers at E11.5, no animals could be retrieved from aggregations at 17.5 days or older. Therefore, it is likely that the deletion of the CTCF site associated to the *Pax3* promoter is lethal during the developmental window between the days 11.5 and 17.5.

*Pax*3 is a transcription factor which plays a crucial role in the differentiation of several biological structures, including tissues deriving from neural crest cells, cardiovascular tissue and skeletal muscle (92, 104). Therefore, the observed lethality might be explained by an impairment on the essential function of *Pax3* in other organs or tissues than the limbs.

5.3.2 Challenging the "CTCF dogma"

Convergently-oriented CTCF sites are thought to be a necessary condition for looping (42) and enhancer-promoter looping a necessary condition for distally controlled gene expression (83). In the following section, I will refer to this as the "CTCF dogma" and refer to the two parts of the dogma as the first and second premise. This represents a simplification of the scientific discourse about CTCF, loops and gene regulation but nonetheless it does not fail to grasp the general consensus of how they are linked.

The objective of this thesis was to challenge the CTCF dogma by studying limb development at the *Epha4* locus, CRISPR-Cas9 engineering CTCF motif deletions and, subsequently, analyzing changes in genome architecture and gene expression.

The *Epha4* locus is a suitable model locus for pursuing these questions because:

(1.) It contains a single gene, hence simplifying the analysis of gene expression data;

(2.) the 3D chromatin organization architecture and its influence in gene expression were already well characterized (61); and

(3.) the genomic extension of the *Epha4* TAD (2 Mb) and the distance between the limb enhancer cluster and the *Epha4* promoter allow to distinguish between proximity generated by looping or linear genomic distance.

While the *Epha4* locus is an intriguing model for discussing these questions, the presented results are of course only true for the *Epha4* locus and care should be taken when generalizing from it to other loci.

The first premise of the CTCF dogma was investigated by creating CTCF binding site deletions and evaluating their effect on chromatin interactions through CHiC experiments. The CHiC data obtained are perfectly in line with the first premise of the CTCF dogma: Deleting a convergently oriented CTCF sites (CTCF_P) disrupts its associated chromatin loops. Furthermore, the increased interaction

frequency with all centromerically-located sequences (interaction trail) up to centromeric TAD boundary is also disrupted. Additionally, we observed the formation of a novel loop between convergently-orientated CTCF pairs (CTCF_E and those within the telomeric TAD). In summary, the obtained data confirm the first premise of the CTCF dogma that convergently-oriented CTCF sites are required for loop formation.

This result is concordant with data from numerous recently published studies - including those by Despang et al., Paliou et al. and Williamson et al. discussed above.

The second premise of the CTCF dogma was investigated by evaluating whether changes in genome architecture showed correlations with changes in gene expression. Gene expression data were analyzed in all homozygous ΔCTCF mutants but, as stated above, genome architecture was only analyzed in the *Epha4*ΔCTCF_P mutants. Gene expression data showed that disrupting the enhancer-promoter loop within the *Epha4* TAD did not lead to a complete loss of gene expression. Instead, different CTCF motif deletions lead to variable changes or even lack of changes in quantitative expression output: deletions of CTCF at promoters decreased gene expression by half, whereas deletions at enhancers displayed no changes in expression levels. As discussed in section 5.2.1, in mutants carrying a CTCF deletion associated to the enhancer, full expression levels are likely retained by a compensatory sub-TAD interaction between the CTCF_P site and the centromeric boundary. In contrast, disrupting the enhancer-promoter loop by deleting a CTCF site associated to the promoter reduced expression by half. As discussed in section 5.2.2, our results suggest that in these mutants the base-line contact frequency within the *Epha4* TAD suffices to maintain half of physiological expression levels.

In the scientific discourse the primary function of TADs is thought to be insulation (45-48) : TADs are seen as insulating units, defining a cell type invariant framework. Inside TADs, a more dynamic system of cis-regulatory elements coordinated by sub-TAD loops is responsible for up- and downregulating gene expression in the correct spatiotemporal pattern. The here presented results put into question this this strict division of functions between TADs and loops. We show that TADs can, just as much as sub-TAD loops, facilitate enhancer-promoter proximity over vast genomic distances. The proximity generated by baseline-interaction seen in TADs is not as strong and lasting as that

generated by loops which can achieve a more targeted, precise and robust proximity of two sequences. This is highlighted by the CHiC maps shown in section 4.5 and at the same time it explains why baseline TAD interaction only has the capacity to preserve half of expression in animals carrying CTCF deletions associated to promoters (ΔCTCF_P and ΔCTCF_P+E). The here presented results propose a model in which TADs create the insulating framework within which sub-TAD loops may establish strong cis-regulatory contacts while, at the same time, TADs also preserve a certain degree enhancer-promoter interactions.

While most studies concerning TADs highlight their insulating properties, there is evidence from other studies pointing towards their distance neglecting properties: In Symmons, Pan et al. 2016 it is shown that a regulatory sensor is ubiquitously expressed when integrated along a 840 kb distance across the *Shh* locus. Expression levels do not correlate with the linear distance to the ZRS enhancer nor to the *Shh* gene. Instead, expression correlates well with the 4C interaction signal to the ZRS. This stresses the point that proximity with the ZRS enhancer facilitates expression. It also highlights the fact that the amount of proximity generated by baseline TAD interaction suffices to drive expression in the *Shh* locus without specific contributions of sub-TAD loops. Interestingly, the sequences showing increased interaction with the ZRS might not be detected according to the standard algorithms - especially because they do not correlate with CTCF anchor points. The studies by Williamson et al., Despang et al. and Paliou et al. discussed above yield similar results: Developmental gene expression can be unaltered by disruption of CTCF mediated enhancer-promoter loops (103) or diminished to a degree, which is not sufficient to cause pathological phenotypes (64, 84). Accordingly, besides enhancer-promoter loops, other mechanisms such as the baseline interaction within TADs must capable of fail-saving gene expression during development.

Our results poise the question whether TADs and loops are entirely different entities or merely the same phenomenon of loop-extrusion observed on different levels of magnitude. Evidence indicating that they are different architectural entities stems from studies showing that TADs are relatively invariant, whereas loops may vary more strongly between species and cell types (42, 57, 75). Furthermore, TAD boundaries exhibit strong insulating properties, while loop-anchor points do not (42). However, there is equally strong evidence for the similarities between loops and TADs: Firstly, the above referred arguments for distinguishing between TADs and loops point out their functional

differences but fail to point out that mechanistically they seem to be equally created by CTCF-mediated loop extrusion. Thus, the insulating properties of TAD boundaries might merely be the consequence of several CTCF binding sites, with divergent orientation, clustering within small genomic regions. Secondly, there is evidence that some TAD boundaries are not always entirely insulating (70). And thirdly, the here presented results suggest that loops might also possess insulating properties: In mutants carrying a deletion of the CTCF_P site, a new loop between CTCF_E and the telomeric boundary emerges (marked as 6 in Figure 4-12). This loop could previously have been prevented from forming by the insulating properties of CTCF_P. Concordantly with this, Despang et al. 2019 report that deletion of the *Sox9-Kcnj2* TAD boundary only results in TAD fusion, when intra-TAD CTCF sites were additionally deleted (64).

The results of this work support a model of gene regulation in which multiple cis-regulatory players may be actively involved or silent and eventually acting as buffering mechanisms. In that sense, transcription can be controlled by proximal or distal enhancers. CTCF mediated loops controlling distal enhancers can be constitutive or facultative. Equally enhancers can be active or silent. Enhancer-promoter proximity can be generated by high precision CTCF mediated enhancer-promoter loops or a lesser, more loose degree of proximity can be established by baseline TAD interaction.

How complex gene regulatory networks are and how many buffering systems might exist is still largely elusive. In this work we have entirely focused on cis-regulatory elements and excluded trans-regulatory elements that seem also to be highly relevant, as highlighted by recent studies. For example, the disruption of cis-regulatory networks by CTCF or Cohesin depletion only has a moderate effect on transcription (82, 87, 105, 106). Furthermore, higher order chromatin folding into A/B-compartments and chromatin state remain unaffected by CTCF or Cohesin depletion (82, 89, 106), suggesting that they are controlled by CTCF-independent factors (53, 107). These studies show that CTCFs role in chromosome folding is limited to a smaller genomic scale and that other factors – many acting in trans – might contribute to gene regulatory processes. However, the importance of TADs and loops is manifested by numerous studies at selected loci that show drastic effects on gene expression and phenotypes (19, 61, 63, 65, 68, 69). Recent advances on single cell and live imaging studies also suggest that TADs and loops are far less static structures than previously thought (51, 53, 54, 108). These studies postulate that TADs and loops are transient structures - building,

deconstructing and rebuilding in a timespan of minutes (109), and that transient enhancer-promoter interactions can be sufficient to achieve transcriptional output (108).

Together, these results highlight that gene regulation is highly dependent on the nature of the loci or the cell/tissue studied. For some loci enhancer-promoter proximity might require high-precision enhancer-promoter loops, whereas in others baseline TAD interaction suffices. In many, however, CTCF-mediated genome architecture seems to be nonessential in generating stable levels of gene expression. Overall, the mechanisms from which gene regulatory networks can draw are so numerous that they cannot be broken down to one principle. Instead it will be the tedious process of science, which will have to illuminate how gene regulation works at different loci, different tissues, different time points and different species individually – case-by-case one at a time.

The here presented work improves our understanding of the network regulating and fail-saving the expression of *Epha4* in the developing mouse limb. More globally, it highlights the numerous, vastly complex and intertwined principles of gene regulatory networks. The knowledge of these principles can be used to uncover the countless individual variations of gene regulatory networks in development and disease.

6. References

1. Maher B. ENCODE: The human encyclopaedia. Nature. 2012;489(7414):46-8.

2. Mathelier A, Shi W, Wasserman WW. Identification of altered cis-regulatory elements in human disease. Trends Genet. 2015;31(2):67-76.

3. Chen H, Levo M, Barinov L, Fujioka M, Jaynes JB, Gregor T. Dynamic interplay between enhancer-promoter topology and gene activity. Nat Genet. 2018;50(9):1296-303.

4. Haberle V, Lenhard B. Promoter architectures and developmental gene regulation. Semin Cell Dev Biol. 2016;57:11-23.

5. Fulco CP, Munschauer M, Anyoha R, Munson G, Grossman SR, Perez EM, Kane M, Cleary B, Lander ES, Engreitz JM. Systematic mapping of functional enhancer-promoter connections with CRISPR interference. Science. 2016;354(6313):769-73.

6. Cho SW, Xu J, Sun R, Mumbach MR, Carter AC, Chen YG, Yost KE, Kim J, He J, Nevins SA, Chin SF, Caldas C, Liu SJ, Horlbeck MA, Lim DA, Weissman JS, Curtis C, Chang HY. Promoter of lncRNA Gene PVT1 Is a Tumor-Suppressor DNA Boundary Element. Cell. 2018;173(6):1398-412 e22.

7. Long HK, Prescott SL, Wysocka J. Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. Cell. 2016;167(5):1170-87.

8. Monahan K, Horta A, Lomvardas S. LHX2- and LDB1-mediated trans interactions regulate olfactory receptor choice. Nature. 2019;565(7740):448-53.

9. Arnold CD, Gerlach D, Stelzer C, Boryn LM, Rath M, Stark A. Genome-wide quantitative enhancer activity maps identified by STARR-seq. Science. 2013;339(6123):1074-7.

10. Lettice LA, Heaney SJ, Purdie LA, Li L, de Beer P, Oostra BA, Goode D, Elgar G, Hill RE, de Graaff E. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. Hum Mol Genet. 2003;12(14):1725-35.

11. Bulger M, Groudine M. Functional and mechanistic diversity of distal transcription enhancers. Cell. 2011;144(3):327-39.

12. Maston GA, Evans SK, Green MR. Transcriptional regulatory elements in the human genome. Annu Rev Genomics Hum Genet. 2006;7:29-59.

13. Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-Andre V, Sigova AA, Hoke HA, Young RA. Super-enhancers in the control of cell identity and disease. Cell. 2013;155(4):934-47.

14. Hay D, Hughes JR, Babbs C, Davies JOJ, Graham BJ, Hanssen L, Kassouf MT, Marieke Oudelaar AM, Sharpe JA, Suciu MC, Telenius J, Williams R, Rode C, Li PS, Pennacchio LA, Sloane-Stanley JA, Ayyub H, Butler S, Sauka-Spengler T, Gibbons RJ, Smith AJH, Wood WG, Higgs DR. Genetic dissection of the alpha-globin super-enhancer in vivo. Nat Genet. 2016;48(8):895-903.

15. Yuh CH, Davidson EH. Modular cis-regulatory organization of Endo16, a gut-specific gene of the sea urchin embryo. Development. 1996;122(4):1069-82.

16. Stine ZE, McGaughey DM, Bessling SL, Li S, McCallion AS. Steroid hormone modulation of RET through two estrogen responsive enhancers in breast cancer. Hum Mol Genet. 2011;20(19):3746-56.

17. Perry MW, Boettiger AN, Levine M. Multiple enhancers ensure precision of gap gene-expression patterns in the Drosophila embryo. Proc Natl Acad Sci U S A. 2011;108(33):13570-5.

18. Marinic M, Aktas T, Ruf S, Spitz F. An integrated holo-enhancer unit defines tissue and gene specificity of the Fgf8 regulatory landscape. Dev Cell. 2013;24(5):530-42.

19. Maeda RK, Karch F. Gene expression in time and space: additive vs hierarchical organization of cis-regulatory regions. Curr Opin Genet Dev. 2011;21(2):187-93.

20. Barolo S. Shadow enhancers: frequently asked questions about distributed cis-regulatory information and enhancer redundancy. Bioessays. 2012;34(2):135-41.

21. Frankel N, Davis GK, Vargas D, Wang S, Payre F, Stern DL. Phenotypic robustness conferred by apparently redundant transcriptional enhancers. Nature. 2010;466(7305):490-3.

22. Dunipace L, Ozdemir A, Stathopoulos A. Complex interactions between cis-regulatory modules in native conformation are critical for Drosophila snail expression. Development. 2011;138(18):4075-84.

23. Osterwalder M, Barozzi I, Tissieres V, Fukuda-Yuzawa Y, Mannion BJ, Afzal SY, Lee EA, Zhu Y, Plajzer-Frick I, Pickle CS, Kato M, Garvin TH, Pham QT, Harrington AN, Akiyama JA, Afzal V, Lopez-Rios J, Dickel DE, Visel A, Pennacchio LA. Enhancer redundancy provides phenotypic robustness in mammalian development. Nature. 2018;554(7691):239-43.

24. Villar D, Berthelot C, Aldridge S, Rayner TF, Lukk M, Pignatelli M, Park TJ, Deaville R, Erichsen JT, Jasinska AJ, Turner JM, Bertelsen MF, Murchison EP, Flicek P, Odom DT. Enhancer evolution across 20 mammalian species. Cell. 2015;160(3):554-66.

25. Gaszner M, Felsenfeld G. Insulators: exploiting transcriptional and epigenetic mechanisms. Nat Rev Genet. 2006;7(9):703-13.

26. Phillips JE, Corces VG. CTCF: master weaver of the genome. Cell. 2009;137(7):1194-211.

27. Spitz F, Furlong EE. Transcription factors: from enhancer binding to developmental control. Nat Rev Genet. 2012;13(9):613-26.

28. Pan Y, Tsai CJ, Ma B, Nussinov R. Mechanisms of transcription factor selectivity. Trends Genet. 2010;26(2):75-83.

29. Iwafuchi-Doi M, Zaret KS. Cell fate control by pioneer transcription factors. Development. 2016;143(11):1833-7.

30. McGinty RK, Tan S. Nucleosome structure and function. Chem Rev. 2015;115(6):2255-73.

31. Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, Boyer LA, Young RA, Jaenisch R. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci U S A. 2010;107(50):21931-6.

32. Rada-Iglesias A, Wysocka J. Epigenomics of human embryonic stem cells and induced pluripotent stem cells: insights into pluripotency and implications for disease. Genome Med. 2011;3(6):36.

33. Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, Ye Z, Lee LK, Stuart RK, Ching CW, Ching KA, Antosiewicz-Bourget JE, Liu H, Zhang X, Green RD, Lobanenkov VV, Stewart R, Thomson JA, Crawford GE, Kellis M, Ren B. Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature. 2009;459(7243):108-12.

34. Bonn S, Zinzen RP, Girardot C, Gustafson EH, Perez-Gonzalez A, Delhomme N, Ghavi-Helm Y, Wilczynski B, Riddell A, Furlong EE. Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. Nat Genet. 2012;44(2):148-56.

35. Shlyueva D, Stampfel G, Stark A. Transcriptional enhancers: from properties to genome-wide predictions. Nat Rev Genet. 2014;15(4):272-86.

36. Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. Science. 2002;295(5558):1306-11.

37. Denker A, de Laat W. The second decade of 3C technologies: detailed insights into nuclear organization. Genes Dev. 2016;30(12):1357-82.

38. Zhao Z, Tavoosidana G, Sjolinder M, Gondor A, Mariano P, Wang S, Kanduri C, Lezcano M, Sandhu KS, Singh U, Pant V, Tiwari V, Kurukuti S, Ohlsson R. Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. Nat Genet. 2006;38(11):1341-7.

39. Sanborn AL, Rao SS, Huang SC, Durand NC, Huntley MH, Jewett AI, Bochkov ID, Chinnappan D, Cutkosky A, Li J, Geeting KP, Gnirke A, Melnikov A, McKenna D, Stamenova EK, Lander ES, Aiden EL. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. Proc Natl Acad Sci U S A. 2015;112(47):E6456-65.

40. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R, Bernstein B, Bender MA, Groudine M, Gnirke A, Stamatoyannopoulos J, Mirny LA, Lander ES, Dekker J. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science. 2009;326(5950):289-93.

41. de Wit E, de Laat W. A decade of 3C technologies: insights into nuclear organization. Genes Dev. 2012;26(1):11-24.

42. Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, Aiden EL. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell. 2014;159(7):1665-80.

43. Bonev B, Mendelson Cohen N, Szabo Q, Fritsch L, Papadopoulos GL, Lubling Y, Xu X, Lv X, Hugnot JP, Tanay A, Cavalli G. Multiscale 3D Genome Rewiring during Mouse Neural Development. Cell. 2017;171(3):557-72 e24.

44. Hughes JR, Roberts N, McGowan S, Hay D, Giannoulatou E, Lynch M, De Gobbi M, Taylor S, Gibbons R, Higgs DR. Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. Nat Genet. 2014;46(2):205-12.

45. Dixon JR, Gorkin DU, Ren B. Chromatin Domains: The Unit of Chromosome Organization. Mol Cell. 2016;62(5):668-80.

46. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature. 2012;485(7398):376-80.

47. Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J, Sedat J, Gribnau J, Barillot E, Bluthgen N, Dekker J, Heard E. Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature. 2012;485(7398):381-5.

48. Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, Hoichman M, Parrinello H, Tanay A, Cavalli G. Three-dimensional folding and functional organization principles of the Drosophila genome. Cell. 2012;148(3):458-72.

49. Vietri Rudan M, Barrington C, Henderson S, Ernst C, Odom DT, Tanay A, Hadjur S. Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. Cell Rep. 2015;10(8):1297-309.

50. Bintu B, Mateo LJ, Su JH, Sinnott-Armstrong NA, Parker M, Kinrot S, Yamaya K, Boettiger AN, Zhuang X. Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. Science. 2018;362(6413).

51. Finn EH, Pegoraro G, Brandao HB, Valton AL, Oomen ME, Dekker J, Mirny L, Misteli T. Extensive Heterogeneity and Intrinsic Variation in Spatial Genome Organization. Cell. 2019;176(6):1502-15 e10.

52. Finn EH, Misteli T. Molecular basis and biological function of variability in spatial genome organization. Science. 2019;365(6457).

53. Flyamer IM, Gassler J, Imakaev M, Brandao HB, Ulianov SV, Abdennur N, Razin SV, Mirny LA, Tachibana-Konwalski K. Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. Nature. 2017;544(7648):110-4.

54. Stevens TJ, Lando D, Basu S, Atkinson LP, Cao Y, Lee SF, Leeb M, Wohlfahrt KJ, Boucher W, O'Shaughnessy-Kirwan A, Cramard J, Faure AJ, Ralser M, Blanco E, Morey L, Sanso M, Palayret MGS, Lehner B, Di Croce L, Wutz A, Hendrich B, Klenerman D, Laue ED. 3D structures of individual mammalian genomes studied by single-cell Hi-C. Nature. 2017;544(7648):59-64.

55. Hnisz D, Day DS, Young RA. Insulated Neighborhoods: Structural and Functional Units of Mammalian Gene Control. Cell. 2016;167(5):1188-200.

56. Ruf S, Symmons O, Uslu VV, Dolle D, Hot C, Ettwiller L, Spitz F. Large-scale analysis of the regulatory architecture of the mouse genome with a transposon-associated sensor. Nat Genet. 2011;43(4):379-86.

57. Ji X, Dadon DB, Powell BE, Fan ZP, Borges-Rivera D, Shachar S, Weintraub AS, Hnisz D, Pegoraro G, Lee TI, Misteli T, Jaenisch R, Young RA. 3D Chromosome Regulatory Landscape of Human Pluripotent Cells. Cell Stem Cell. 2016;18(2):262-75.

58. Shen Y, Yue F, McCleary DF, Ye Z, Edsall L, Kuan S, Wagner U, Dixon J, Lee L, Lobanenkov VV, Ren B. A map of the cis-regulatory sequences in the mouse genome. Nature. 2012;488(7409):116-20.

59. Jin F, Li Y, Dixon JR, Selvaraj S, Ye Z, Lee AY, Yen CA, Schmitt AD, Espinoza CA, Ren B. A high-resolution map of the three-dimensional chromatin interactome in human cells. Nature. 2013;503(7475):290-4.

60. Symmons O, Uslu VV, Tsujimura T, Ruf S, Nassari S, Schwarzer W, Ettwiller L, Spitz F. Functional and topological characteristics of mammalian regulatory domains. Genome Res. 2014;24(3):390-400.

61. Lupianez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, Horn D, Kayserili H, Opitz JM, Laxova R, Santos-Simarro F, Gilbert-Dussardier B, Wittler L, Borschiwer M, Haas SA, Osterwalder M, Franke M, Timmermann B, Hecht J, Spielmann M, Visel A, Mundlos S. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell. 2015;161(5):1012-25.

62. Tsujimura T, Klein FA, Langenfeld K, Glaser J, Huber W, Spitz F. A discrete transition zone organizes the topological and regulatory autonomy of the adjacent tfap2c and bmp7 genes. PLoS Genet. 2015;11(1):e1004897.

63. Ibn-Salem J, Kohler S, Love MI, Chung HR, Huang N, Hurles ME, Haendel M, Washington NL, Smedley D, Mungall CJ, Lewis SE, Ott CE, Bauer S, Schofield PN, Mundlos S, Spielmann M, Robinson PN. Deletions of chromosomal regulatory boundaries are associated with congenital disease. Genome Biol. 2014;15(9):423.

64. Despang A, Schopflin R, Franke M, Ali S, Jerkovic I, Paliou C, Chan WL, Timmermann B, Wittler L, Vingron M, Mundlos S, Ibrahim DM. Functional dissection of the Sox9-Kcnj2 locus identifies nonessential and instructive roles of TAD architecture. Nat Genet. 2019;51(8):1263-71.

65. Franke M, Ibrahim DM, Andrey G, Schwarzer W, Heinrich V, Schopflin R, Kraft K, Kempfer R, Jerkovic I, Chan WL, Spielmann M, Timmermann B, Wittler L, Kurth I, Cambiaso P, Zuffardi O, Houge G, Lambie L, Brancati F, Pombo A, Vingron M, Spitz F, Mundlos S. Formation of new chromatin domains determines pathogenicity of genomic duplications. Nature. 2016;538(7624):265-9.

66. Hsu SC, Gilgenast TG, Bartman CR, Edwards CR, Stonestrom AJ, Huang P, Emerson DJ, Evans P, Werner MT, Keller CA, Giardine B, Hardison RC, Raj A, Phillips-Cremins JE, Blobel GA. The BET Protein BRD2 Cooperates with CTCF to Enforce Transcriptional and Architectural Boundaries. Mol Cell. 2017;66(1):102-16 e7.

67. Hanssen LLP, Kassouf MT, Oudelaar AM, Biggs D, Preece C, Downes DJ, Gosden M, Sharpe JA, Sloane-Stanley JA, Hughes JR, Davies B, Higgs DR. Tissue-specific CTCF-cohesin-mediated chromatin architecture delimits enhancer interactions and function in vivo. Nat Cell Biol. 2017;19(8):952-61.

68. Flavahan WA, Drier Y, Liau BB, Gillespie SM, Venteicher AS, Stemmer-Rachamimov AO, Suva ML, Bernstein BE. Insulator dysfunction and oncogene activation in IDH mutant gliomas. Nature. 2016;529(7584):110-4.

69. Hnisz D, Weintraub AS, Day DS, Valton AL, Bak RO, Li CH, Goldmann J, Lajoie BR, Fan ZP, Sigova AA, Reddy J, Borges-Rivera D, Lee TI, Jaenisch R, Porteus MH, Dekker J, Young RA. Activation of proto-oncogenes by disruption of chromosome neighborhoods. Science. 2016;351(6280):1454-8.

70. Symmons O, Pan L, Remeseiro S, Aktas T, Klein F, Huber W, Spitz F. The Shh Topological Domain Facilitates the Action of Remote Enhancers by Reducing the Effects of Genomic Distances. Dev Cell. 2016;39(5):529-43.

71. Sanyal A, Lajoie BR, Jain G, Dekker J. The long-range interaction landscape of gene promoters. Nature. 2012;489(7414):109-13.

72. Javierre BM, Burren OS, Wilder SP, Kreuzhuber R, Hill SM, Sewitz S, Cairns J, Wingett SW, Varnai C, Thiecke MJ, Burden F, Farrow S, Cutler AJ, Rehnstrom K, Downes K, Grassi L, Kostadima M, Freire-Pritchett P, Wang F, Consortium B, Stunnenberg HG, Todd JA, Zerbino DR, Stegle O, Ouwehand WH, Frontini M, Wallace C, Spivakov M, Fraser P. Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. Cell. 2016;167(5):1369-84 e19.

73. Alexander JM, Guan J, Li B, Maliskova L, Song M, Shen Y, Huang B, Lomvardas S, Weiner OD. Live-cell imaging reveals enhancer-dependent Sox2 transcription in the absence of enhancer proximity. Elife. 2019;8.

74. Benabdallah NS, Williamson I, Illingworth RS, Kane L, Boyle S, Sengupta D, Grimes GR, Therizols P, Bickmore WA. Decreased Enhancer-Promoter Proximity Accompanying Enhancer Activation. Mol Cell. 2019;76(3):473-84 e7.

75. Phillips-Cremins JE, Sauria ME, Sanyal A, Gerasimova TI, Lajoie BR, Bell JS, Ong CT, Hookway TA, Guo C, Sun Y, Bland MJ, Wagstaff W, Dalton S, McDevitt TC, Sen R, Dekker J, Taylor J, Corces VG. Architectural protein subclasses shape 3D organization of genomes during lineage commitment. Cell. 2013;153(6):1281-95.

76. Fudenberg G, Imakaev M, Lu C, Goloborodko A, Abdennur N, Mirny LA. Formation of Chromosomal Domains by Loop Extrusion. Cell Rep. 2016;15(9):2038-49.

77. Palstra RJ, Tolhuis B, Splinter E, Nijmeijer R, Grosveld F, de Laat W. The beta-globin nuclear compartment in development and erythroid differentiation. Nat Genet. 2003;35(2):190-4.

78. Filippova GN, Fagerlie S, Klenova EM, Myers C, Dehner Y, Goodwin G, Neiman PE, Collins SJ, Lobanenkov VV. An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. Mol Cell Biol. 1996;16(6):2802-13.

79. Chen X, Xu H, Yuan P, Fang F, Huss M, Vega VB, Wong E, Orlov YL, Zhang W, Jiang J, Loh YH, Yeo HC, Yeo ZX, Narang V, Govindarajan KR, Leong B, Shahab A, Ruan Y, Bourque G, Sung WK, Clarke ND, Wei CL, Ng HH. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell. 2008;133(6):1106-17.

80. Narendra V, Rocha PP, An D, Raviram R, Skok JA, Mazzoni EO, Reinberg D. CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. Science. 2015;347(6225):1017-21.

81. Zuin J, Dixon JR, van der Reijden MI, Ye Z, Kolovos P, Brouwer RW, van de Corput MP, van de Werken HJ, Knoch TA, van IWF, Grosveld FG, Ren B, Wendt KS. Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. Proc Natl Acad Sci U S A. 2014;111(3):996-1001.

82. Nora EP, Goloborodko A, Valton AL, Gibcus JH, Uebersohn A, Abdennur N, Dekker J, Mirny LA, Bruneau BG. Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. Cell. 2017;169(5):930-44 e22.

83. de Wit E, Vos ES, Holwerda SJ, Valdes-Quezada C, Verstegen MJ, Teunissen H, Splinter E, Wijchers PJ, Krijger PH, de Laat W. CTCF Binding Polarity Determines Chromatin Looping. Mol Cell. 2015;60(4):676-84.

84. Paliou C, Guckelberger P, Schopflin R, Heinrich V, Esposito A, Chiariello AM, Bianco S, Annunziatella C, Helmuth J, Haas S, Jerkovic I, Brieske N, Wittler L, Timmermann B, Nicodemi M, Vingron M, Mundlos S, Andrey G. Preformed chromatin topology assists transcriptional robustness of Shh during limb development. Proc Natl Acad Sci U S A. 2019;116(25):12390-9.

85. Dekker J, Mirny L. The 3D Genome as Moderator of Chromosomal Communication. Cell. 2016;164(6):1110-21.

86. Gomez-Marin C, Tena JJ, Acemel RD, Lopez-Mayorga M, Naranjo S, de la Calle-Mustienes E, Maeso I, Beccari L, Aneas I, Vielmas E, Bovolenta P, Nobrega MA, Carvajal J, Gomez-Skarmeta JL. Evolutionary comparison reveals that diverging CTCF sites are signatures of ancestral topological associating domains borders. Proc Natl Acad Sci U S A. 2015;112(24):7542-7.

87. Rao SSP, Huang SC, Glenn St Hilaire B, Engreitz JM, Perez EM, Kieffer-Kwon KR, Sanborn AL, Johnstone SE, Bascom GD, Bochkov ID, Huang X, Shamim MS, Shin J, Turner D, Ye Z, Omer AD, Robinson JT, Schlick T, Bernstein BE, Casellas R, Lander ES, Aiden EL. Cohesin Loss Eliminates All Loop Domains. Cell. 2017;171(2):305-20 e24.

88. Guo Y, Xu Q, Canzio D, Shou J, Li J, Gorkin DU, Jung I, Wu H, Zhai Y, Tang Y, Lu Y, Wu Y, Jia Z, Li W, Zhang MQ, Ren B, Krainer AR, Maniatis T, Wu Q. CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. Cell. 2015;162(4):900-10.

89. Haarhuis JHI, van der Weide RH, Blomen VA, Yanez-Cuna JO, Amendola M, van Ruiten MS, Krijger PHL, Teunissen H, Medema RH, van Steensel B, Brummelkamp TR, de Wit E, Rowland BD. The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension. Cell. 2017;169(4):693-707 e14.

90. Kullander K, Klein R. Mechanisms and functions of Eph and ephrin signalling. Nat Rev Mol Cell Biol. 2002;3(7):475-86.

91. Helmbacher F, Schneider-Maunoury S, Topilko P, Tiret L, Charnay P. Targeting of the EphA4 tyrosine kinase receptor affects dorsal/ventral pathfinding of limb motor axons. Development. 2000;127(15):3313-24.

92. Epstein JA. Pax3 and vertebrate development. Methods Mol Biol. 2000;137:459-70.

93. Spielmann M, Lupianez DG, Mundlos S. Structural variation in the 3D genome. Nat Rev Genet. 2018;19(7):453-67.

94. Stankiewicz P, Lupski JR. Structural variation in the human genome and its role in disease. Annu Rev Med. 2010;61:437-55.

95. Kraft K, Geuer S, Will AJ, Chan WL, Paliou C, Borschiwer M, Harabula I, Wittler L, Franke M, Ibrahim DM, Kragesteen BK, Spielmann M, Mundlos S, Lupianez DG, Andrey G. Deletions, Inversions, Duplications: Engineering of Structural Variants using CRISPR/Cas in Mice. Cell Rep. 2015.

96. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F. Multiplex genome engineering using CRISPR/Cas systems. Science. 2013;339(6121):819-23.

97. Artus J, Hadjantonakis AK. Generation of chimeras by aggregation of embryonic stem cells with diploid or tetraploid mouse embryos. Methods Mol Biol. 2011;693:37-56.

98. Witte F, Dokas J, Neuendorf F, Mundlos S, Stricker S. Comprehensive expression analysis of all Wnt genes and their major secreted antagonists during mouse limb development and cartilage differentiation. Gene Expr Patterns. 2009;9(4):215-23.

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

100. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357-9.

101. Durand NC, Shamim MS, Machol I, Rao SS, Huntley MH, Lander ES, Aiden EL. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. Cell Syst. 2016;3(1):95-8.

102. Deng W, Lee J, Wang H, Miller J, Reik A, Gregory PD, Dean A, Blobel GA. Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. Cell. 2012;149(6):1233-44.

103. Williamson I, Kane L, Devenney PS, Flyamer IM, Anderson E, Kilanowski F, Hill RE, Bickmore WA, Lettice LA. Developmentally regulated Shh expression is robust to TAD perturbations. Development. 2019;146(19).

104. Buckingham M, Relaix F. PAX3 and PAX7 as upstream regulators of myogenesis. Semin Cell Dev Biol. 2015;44:115-25.

105. Ghavi-Helm Y, Jankowski A, Meiers S, Viales RR, Korbel JO, Furlong EEM. Highly rearranged chromosomes reveal uncoupling between genome topology and gene expression. Nat Genet. 2019;51(8):1272-82.

106. Schwarzer W, Abdennur N, Goloborodko A, Pekowska A, Fudenberg G, Loe-Mie Y, Fonseca NA, Huber W, Haering CH, Mirny L, Spitz F. Two independent modes of chromatin organization revealed by cohesin removal. Nature. 2017;551(7678):51-6.

107. Nuebler J, Fudenberg G, Imakaev M, Abdennur N, Mirny LA. Chromatin organization by an interplay of loop extrusion and compartmental segregation. Proc Natl Acad Sci U S A. 2018;115(29):E6697-E706.

108. Fukaya T, Lim B, Levine M. Enhancer Control of Transcriptional Bursting. Cell. 2016;166(2):358-68.

109. Hansen AS, Cattoglio C, Darzacq X, Tjian R. Recent evidence that TADs and chromatin loops are dynamic structures. Nucleus. 2018;9(1):20-32.

7. Eidesstattliche Erklärung

„Ich, Niklas Gerhards, versichere an Eides statt durch meine eigenhändige Unterschrift, dass ich die vorgelegte Dissertation mit dem Thema: „Effects of CTCF binding site deletions on genome architecture and gene expression in the *Epha4* locus“ bzw. „Effekte von Deletionen von CTCF Bindungsstellen auf genomische Architektur und Genexpression im *Epha4* Lokus“ selbstständig und ohne nicht offengelegte Hilfe Dritter verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel genutzt habe.

Alle Stellen, die wörtlich oder dem Sinne nach auf Publikationen oder Vorträgen anderer Autoren/innen beruhen, sind als solche in korrekter Zitierung kenntlich gemacht. Die Abschnitte zu Methodik (insbesondere praktische Arbeiten, Laborbestimmungen, statistische Aufarbeitung) und Resultaten (insbesondere Abbildungen, Graphiken und Tabellen) werden von mir verantwortet.

Ich versichere ferner, dass ich die in Zusammenarbeit mit anderen Personen generierten Daten, Datenauswertungen und Schlussfolgerungen korrekt gekennzeichnet und meinen eigenen Beitrag sowie die Beiträge anderer Personen korrekt kenntlich gemacht habe (siehe Anteilserklärung). Texte oder Textteile, die gemeinsam mit anderen erstellt oder verwendet wurden, habe ich korrekt kenntlich gemacht.

Meine Anteile an etwaigen Publikationen zu dieser Dissertation entsprechen denen, die in der untenstehenden gemeinsamen Erklärung mit dem/der Erstbetreuer/in, angegeben sind. Für sämtliche im Rahmen der Dissertation entstandenen Publikationen wurden die Richtlinien des ICMJE (International Committee of Medical Journal Editors; www.icmje.og) zur Autorenschaft eingehalten. Ich erkläre ferner, dass ich mich zur Einhaltung der Satzung der Charité – Universitätsmedizin Berlin zur Sicherung Guter Wissenschaftlicher Praxis verpflichte.

Weiterhin versichere ich, dass ich diese Dissertation weder in gleicher noch in ähnlicher Form bereits an einer anderen Fakultät eingereicht habe.

Die Bedeutung dieser eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unwahren eidesstattlichen Versicherung (§§156, 161 des Strafgesetzbuches) sind mir bekannt und bewusst.“

Datum: Unterschrift:

8. Curriculum Vitae

Mein Lebenslauf wird aus datenschutzrechtlichen Gründen in der elektronischen Version meiner Arbeit nicht veröffentlicht.

9. Danksagung

Ich danke Prof. Dr. Stefan Mundlos für die Möglichkeit in seiner Arbeitsgruppe *„Development and Disease“* am Max Planck Institut für Molekulare Genetik promoviert haben zu können und für das stets offene Ohr und interessante Gespräche.

Besonders danke ich Dr. Dario Lupiañez für seine wissenschaftliche Betreuung, seine Anleitungen bei experimenteller Arbeit und seine geduldigen Korrekturen.

Ich bin der gesamten Forschungsgruppe dankbar für die herzliche Atmosphäre, die Hilfsbereitschaft und den Spaß, den ich mit euch geteilt habe. Asita Stiege, Ute Fischer und Norbert Brieske danke ich für ihre Arbeit, die das stille Rückgrat dieser Doktorarbeit bildet. Besonders möchte ich Katerina Kraft, Alexandra Despang und Sahaleddine Ali danken, mit denen ein Büro zu teilen, auch an schlechten Tagen gute Laune bereitet hat. Ich bedanke mich bei Andreas, Martin, Lila, Bjørt, Fany, Christina, Guillaume, Ivana, Giulia und Daniel.

Mein größter Dank gilt meinen Freund*Innen und meiner Familie, die mich durch diese Zeit begleitet haben.