The Role of Topologically Associating Domains for Developmental Gene Regulation - A Systematic Functional Analysis at the *Sox9* and *Shh* Loci

Inaugural-Dissertation

to obtain the academic degree

Doctor rerum naturalium (Dr. rer. nat.)

submitted to the Department of Biology, Chemistry, Pharmacy

of Freie Universität Berlin

by

Alexandra Friederike Despang

from Würzburg

Berlin 2021

The present work was conducted from April 2015 to January 2021 at the Max Planck Institute for Molecular Genetics in Prof. Dr. Mundlos' research group.

1st Reviewer: Prof. Dr. Stefan Mundlos
2nd Reviewer: Prof. Dr. Sigmar Stricker

Date of defense: 03.05.2021

“In other words,
just because TADs don’t do everything for gene regulation,
doesn’t mean they do nothing.”

Daniel M. Ibrahim, 2019

TABLE OF CONTENT

1. INTRODUCTION

1.1 Developmental Gene Regulation in Metazoans

Despite the vast morphological diversity from nematodes to vertebrates, all animals share a basic set of genes orchestrating developmental processes. The precise regulation of these genes in time and space is critical for proper embryogenesis and body plan establishment. A variety of mechanisms evolved at a cellular, molecular and genomic level to achieve such sophisticated transcriptional control required to coordinate one cell becoming a multicellular organism.

The decryption of the human genetic code in 2001 revealed that only 2-3% of the mammalian genome contains protein-coding regions, with the remaining 97% have no immediately clear function (Lander et al., 2001). However, recent advances have demonstrated non-coding "junk" DNA harbors cis-regulatory elements (CREs) which are critical for gene regulation. Supporting this, the number of genes within different lineages has remained comparatively stable while CREs have expanded, diversified and altered during millions of years of evolution. Hence, it seems CREs are critical for the development of complex gene regulation patterns with changes in their activity driving altered target expression and evolutionary novelty. In the last decades, increasing evidence of how CREs facilitate such complex gene expression emerged. However, which features facilitate their communication remains poorly understood.

To better characterize which factors determine the emergence of complex expression patterns, in this study, we focused on the well-studied gene regulation in embryonic limb development in mice and utilized it as an *in vivo* model system. In the following sections I will elaborate on the common and state-of-the-art knowledge within the field of transcriptional gene regulation.

1.1.1 The developing limb as a model system for gene regulation

Body axis patterning, cell fate specification, cell proliferation and cell migration are essential for embryogenesis and demand coordinated control in space and time. During limb organogenesis and morphogenesis these processes are orchestrated by a network of signaling pathways consisting of a largely self-regulatory signaling system (Bénazet et al., 2009). Generally, the master key regulators initiating interlinked pathways were well-characterized in the last decades, thereby making the limb bud an ideal and easy-to-assess model system to study mechanisms of gene regulation (Zeller et al., 2009).

Key regulators during limb development

Specification of positional information is crucial during limb development. The limb bud emerges from the flank mesenchyme of the lateral plate mesoderm during mouse embryogenesis at E9.5

days post coitum (dpc). In general, the limb bud develops along three major axes, defined by expression of specific factors in two main signaling centers, the apical ectodermal ridge (AER) and zone of polarizing activity (ZPA) (Fig. 1-1 A and C). For example, the proximal-distal (PD) axis defines limb outgrowth from the trunk of the body to the limb digits. Following outgrowth-initiation by *Fibroblast growth factor 10* (*Fgf10*), progression and definition of the proximal-distal axis is maintained by *Fgf8*-signalling secreted from the AER (Fig. 1-1 C). Dorso-ventral patterning, from the back of the hand to the hand palm, is controlled by the signaling molecules of the *Wnt* family and *Engrailed-1* (*En1*) (Loomis et al., 1996). Finally, *Sonic Hedgehog* (*Shh*) is secreted from cells located in the ZPA at the posterior margin of the limb bud, thereby defining anterior-posterior patterning and growth of digits I-V (Riddle et al., 1993; Tickle, 1981) (Fig. 1-1 C). When expressed, the *Shh* morphogen diffuses in a gradient-like manner along the posterior-anterior axis to define the number and specifying the identity of digits to be formed (Fig. 1-1C) (Lettice et al., 2012, 2017; Osterwalder et al., 2014). Consequently, *Shh* expression is critical for ensuring most of the higher-order mammals pentadactyl and so initiate digits in the fore- and hindlimb with only the thumb developing independently of *Shh* expression. *Shh* achieves this through two major processes. Specifically, in an early patterning phase *Shh* first regulates digit identity and then, in a second phase, promotes digit outgrowth (Zhu et al., 2008). To restrict *Shh* signaling, its pre-activator, the *zinc finger-protein glioma-associated oncogene 3* (*Gli3*), is constitutively converted into its repressive and inactive form *Gli3R* at the anterior margin of the limb bud, opposite of the ZPA (Fig. 1-1 C) (Zúñiga and Zeller, 1999). By the counter play and feedback-loop of *Gli3* and *Shh* itself a morphogen gradient can be formed (Fig. 1-1 C). Taken together, *Shh*-mediated signaling in the ZPA is critical for proper posterior-anterior patterning and outgrowth of digits.

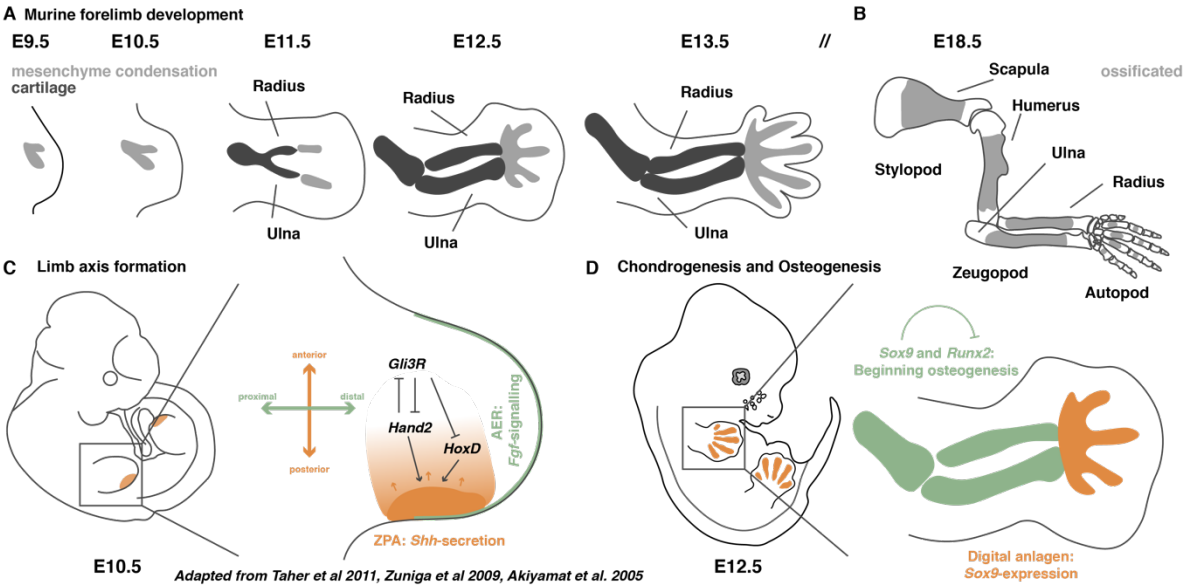

Figure 1-1 Murine forelimb bud development (A) Depiction of mesenchymal condensation (light grey) and beginning chondrogenesis (dark grey) during embryonic stages E9.5-E13.5 dpc in the forelimb. (B) Embryonic limb skeleton at E18.5. Partially ossificated elements are depicted in dark grey. The limb subdivides into stylopod (Humerus), zeugopod (Radius and Ulna) and autopod (wrist and digits). (C) E10.5 mouse embryo with a magnification of the developing forelimb. The main axes and their key regulators/signaling pathways during patterning and axis formation are depicted in green (proximal-distal, PD) and orange (anterior-posterior, AP). *Shh* is secreted from cells localized in the ZPA, initiated by *Hand2* and *HoxD*, and diffuses along the AP-axis, restricted by *Gli3R* counteracting from the anterior site. The proximal-distal outgrowth and patterning is determined by Fgf-signaling. (D) Chondrogenesis and osteogenesis at embryonic stage E12.5. Whereas *Sox9*, initiated by *BMP*-signaling, is expressed in the digital anlagen to induce chondrogenesis (orange), the stylo- and zeugopod already expressing *Runx2* (green), thus initiating osteogenesis in the according skeletal structures. The figure was adapted from *(Akiyama‡ et al., 2005; Taher et al., 2011; Zeller et al., 2009)*

Apart from axis determination by the AER and ZPA other essential limb features are establishing in parallel. As the established spatial axes pattern and polarize the limb, a fraction of cells is aggregates into mesenchymal condensates and builds the progenitors of the later skeleton (Fig. 1-1 A). The limb skeleton subdivides into the stylopod consisting of the humerus, the zeugopod containing the radius and ulna, and the autopod comprising the wrist and digits (Fig. 1-1 B). The *SRY-box transcription factor 9* (*Sox9*) is crucial to the process of differentiation and cartilage development. *Bone morphogenetic proteins* (*BMPs*) induce the expression of *Sox9* already during main axis development which in turn initiates chondrogenesis in committed condensed mesenchymal cells (Duprez et al., 1996). Sox9's significance for long bone development out of limb cartilage anlagen was shown by a conditional *Sox9* KnockOut (KO). In prechondrogenic limb mesenchyme cells *Sox9*-loss results in the absence of cartilage and bone in mice (Akiyama et al., 2002; Duprez et al., 1996). After generating a cartilage skeleton, differentiated chondrocytes begin drastically increasing their volume and turn into hypertrophic chondrocytes which, at later developmental stages, progress to osteoblasts through *Runt-related transcription factor 2* (*Runx2*) expression (Fig. 1-1 D) (Akiyama‡ et al., 2005; Hecht et al., 2008; Stricker et al., 2002). During the process of endochondral ossification, *Sox9* initiates the secretion of extracellular matrix (ECM)-genes such as collagens e.g. *Col2A1*, forming a matrix which later calcifies into bone (Bell et al., 1997). Thus, like *Shh*, *Sox9* forms part of an essential regulatory network that defines limb bud development and so requires precise spatiotemporal regulation.

Considering the importance of these genes, perturbations in their expression mostly have a direct phenotypic readout observable in the limb. This predestines the limb as an easily accessible system to study gene regulation in a developmental context *in vivo*, where expression alterations of one of these key genes translate into abnormal limb development (Zuniga et al., 2012). For example, point mutations and large genomic rearrangements affecting *Shh* limb bud expression

strongly affect the formation of digits and are implicated in congenital malformations (see Figure 1-2). Indeed, ectopic misexpression at the anterior margin of the limb bud induces Poly- and/or Syndactyly defined by additional or fused digits (Fig. 1-2 B). In contrast, a complete loss of *Shh* expression leads to only one remaining rudimentary *Shh*-independent digit (Monodactyly), no handplate and fusion of stylo- and autopod (Fig. 1-2 C) (Klopocki and Mundlos, 2011). Moreover, mutations associated with and around the *Sox9*-gene can be lethal. Sox9-haploinsufficiency has been demonstrated for human and mice, which developed severe chondrodysplasia (Bi et al., 2001; Wagner et al., 1994).

Collectively, these features make the limb bud perfectly suited to study the regulation of the interlinked master regulators. In the next sections I will elaborate on the known and unknown features of mammalian gene regulation on a transcriptional level.

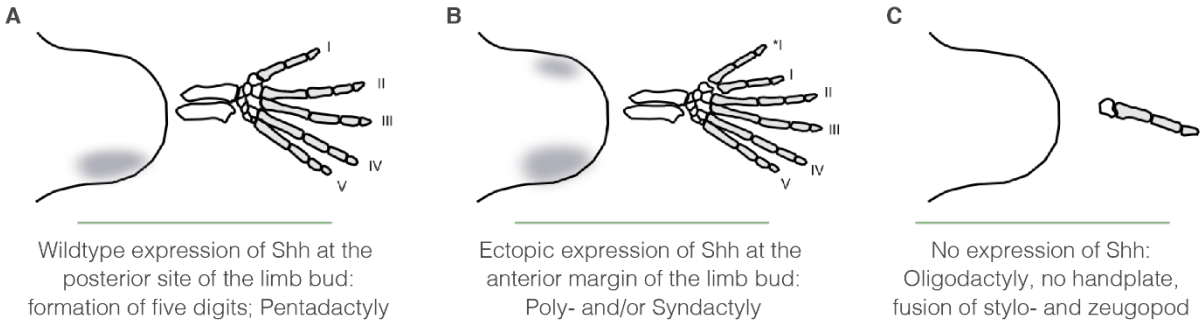

Figure 1-2 Alterations of *Shh* expression levels are causative for limb malformations. (A) Normal wildtype expression of *Shh* at the posterior margin of the developing limb bud at murine embryonic stage E10.5, resulting in pentadactyly (digit I-V). (B) Misexpression at the anterior site of the limb bud potentially leads to the formation of one or more additional digits (Polydactyly) and/or fusion of digits (Syndactyly). (C) Complete absence of *Shh* causes Monodactyly, absence of the handplate and a fusion of the stylo- and zeugopod along with severe reduction of the general limb size. The figure was adapted from Zuniga et al. 2011.

1.1.2 The non-coding genome - Cis-regulatory elements (CREs)

For a long time it was unclear how spatiotemporal expression of genes is regulated. Today, we know that complex gene regulation is controlled by cis-regulatory elements (CREs) which can be separated into distinct classes. Historically, CREs have been generally placed into two classes, promoters and enhancers, that each distinctly drive target gene expression in an orchestrated organization. However, CREs can also be repressive elements like silencers and insulators. For example, insulators are DNA elements binding insulating proteins to block the effect of an enhancer when positioned between an enhancer and its target promoter (Gaszner and Felsenfeld, 2006; Maeda and Karch, 2011). The best characterized insulator-binding protein is the 11-zinc finger protein CCCTC-binding factor (CTCF), which has been associated with repressive activity. However, the biochemical mechanisms of this repressive activity are only recently beginning to

be understood (see section 1.2) (Bell et al., 1999; Klenova et al., 2001). Repressive elements increasingly moved into the attention of research (Jayavelu et al., 2020). However, in this study the focus lies on the most studied CRE classes, namely the activating elements enhancers and promoters.

1.1.3 Promoters and Enhancers

Complex gene expression is regulated by CREs. Among them, the activating elements enhancers and promoter are the most studied. Conventionally promoters and enhancers are defined as stretches of DNA varying in size from a few base pairs (bp) to thousands of bps and have the ability to activate transcription. Both CREs are generally highly conserved within species and can be found dispersed throughout the genome. A common feature of enhancers and promoters is their abundance of transcription factor binding sites that facilitate transcription. Nevertheless, promoters display some unique characteristics.

Promoters - Mammalian promoters consist of core and proximal components (Fig. 1-3 A). The core promoters are comprised of general TFs and, by this, assist in the formation of the pre-initiations complex (PIC) of RNA-Polymerase II (RNA PolII) within 50-100 bps of the enclosed transcription start site (TSS) (Haberle and Stark, 2018; Hampsey, 1998). They commonly have a low basal activity, general motifs like a TATA-box, frequently contain CpG-islands (70%) (Deaton and Bird, 2011; Saxonov et al., 2006), and bear the potential repression or activation by other more distal CREs. By contrast, proximal promoters are located prior to the transcription start site (TSS) and serve as a platform for binding of tissue-specific TFs. Functionally, it is unclear how the proximal promoter communicates with the core and how it confers to gene activation. One possibility is its communication with activating regulatory sequences like enhancers to ultimately release assembled RNA PolII complexes to initiate active transcription (reviewed in (Haberle and Stark, 2018). So far there is no coherent characterization of proximal promoters, although the depiction as core promoter-neighboring enhancers seems reasonable (see below).

Enhancers - While promoters are located directly upstream of the TSS, the location of enhancers is highly variable. The earliest described transcription-enhancing DNA sequence was discovered in 1981 in the laboratories of Schaffner and Chambon (Banerji et al., 1981; Moreau et al., 1981). A 72bp sequence from the SV40 virus was found to significantly upregulate episomal transcription of the rabbit beta-globin gene *in vitro*. Later, it became more evident that this kind of sequence, termed an enhancer, can have cell-type and developmental-stage specific activities. However, due to their more varied positions and absence of a TSS, a common definition of enhancers is pending. Rather, they are defined by their function, thus, the ability to activate a genes' expression.

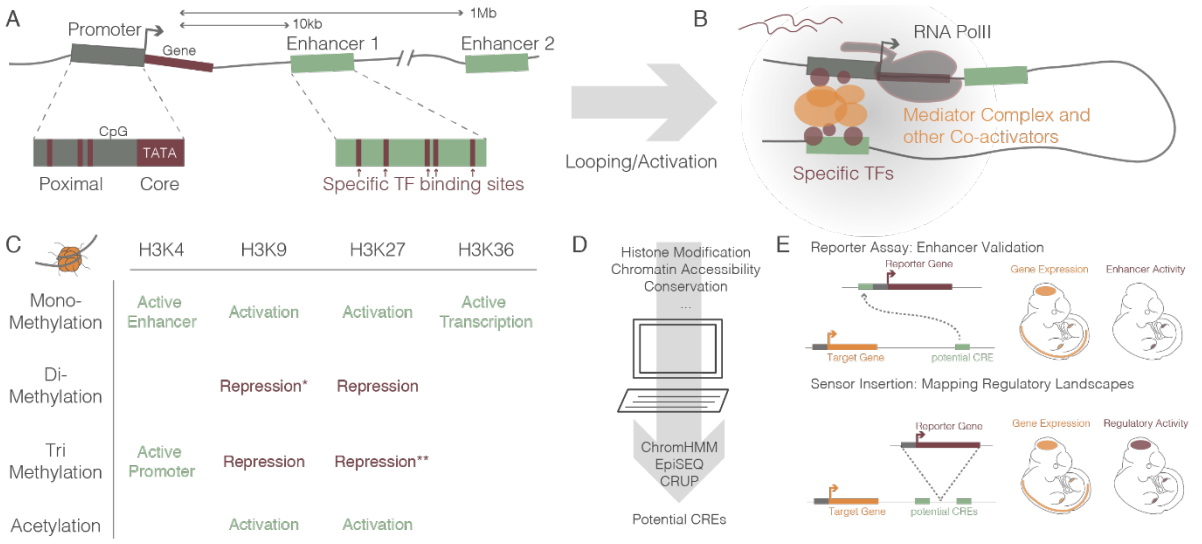

Figure 1-3 Cis-Regulatory Elements: Function, Identification and Validation. (A) Promoter (grey) with proximal and core and adjacent or distal enhancers (green). Both CREs carry TF binding sites (red). (B) Looping of promoter and enhancer mediated by TFs (red circles) binding both elements, mediator complex and other co-activators like p300 (orange) recruiting RNA PolII for active transcription (light red). Highly concentrated numerous complexes of this kind are also described as phase-separating membraneless condensates or active hubs. (C) Summary of the most common type of histone modifications used for CRE-identification and classification of activity state. *LAD marker, **Polycomb marker. (D) Computational process of data integration in algorithms for CRE-identification. (E) Schematics of reporter systems. On top, enhancer validation and below sensor insertions to map a regulatory landscape. In both cases a reporter gene is representing the activity of the sequence to be tested in an exogenous locus or endogenously.

1.1.4 Identification of CREs

The identification of CREs predominantly relies on sequence conservation between species and the epigenome. Conserved non-coding sequences are short stretches of DNA (200-500 bp) that are preserved over millions of years of evolution (Maeso et al., 2013; Vavouri and Lehner, 2009). Many of these have been shown to be tissue-specific enhancers in transgenic reporter assays (Pennacchio et al., 2006). Thus, conservation can be employed to identify enhancers. However, determining in which cell-types or tissues a candidate element is active in remains challenging. Chromatin modifications modeling the epigenome are considered to identify the state of activity of a regulatory region in a certain tissue or cell type. The majority of the DNA fiber is packed into nucleosomes, where 147 bp of DNA are wrapped around a histone-octamer complex. Different post-translational modifications of histone proteins, particularly on their N-terminal domains, are characteristic for a certain functional state of chromatin (summarized in Fig. 1-3 C) (Visel et al., 2009). For example, active promoters and enhancers are often acetylated at lysine 27 of histone H3 (H3K27ac), whereas tri-methylation at the same position (H3K27me3) supports repression by the polycomb machinery (Kouzarides, 2007; Schwartz and Pirrotta, 2007). Nevertheless,

enhancers and promoters can be distinguished through differentiation methylation of lysine 4 on histone H3. Specifically, active promoters display enriched H3K4me3 and H3K4me1, while enhancers possess only elevated H3K4me1 (Bernstein and Kellis, 2005). Beyond histone modifications, active regulatory elements also display distinct accessibilities. Open chromatin regions are typically assessed with DNaseI hypersensitivity sequencing or Assay for Transposase-Accessible Chromatin (ATAC)-seq. Thus, a combination of histone modifications, chromatin accessibility and sequence conservation can often be used to identify candidate enhancers. Accordingly, several computational tools have been developed which integrate complex information into specific algorithms to systematically annotate regulatory regions (Fig. 1-3 D) (ChromHMM, CRUP) (Ernst and Kellis, 2017; Ramisch et al., 2019).

By definition, enhancers are characterized functionally and not by chromatin state. Computational annotation of potential regulatory elements based on epigenetic marks tend to give an idea of activity in a certain tissue. Yet, the called regions need to be validated. The evaluation of enhancers is typically done by transgenic reporter assays of two types. In ectopic assays an inert minimal promoter is fused to a lacZ reporter gene and a candidate element at an ectopic genomic location. Subsequent reporter expression in cells or embryos is then used to reveal the candidate enhancer's pattern of activity (Fig. 1-3 E). Alternatively, minimal promoter-lacZ reports can be integrated into any location of interest to map the regulatory information available at that position. Consequently, such regulatory assays test not only an individual element at a random non-endogenous position but rather measure the regulatory activity of all CREs in the vicinity of the insertion site (reviewed in (Kvon, 2015; Symmons and Spitz, 2013). Large-scale functional transgenics combined with high-throughput evaluation like the VISTA enhancer data base are testing highly conserved non-coding elements of humans *in vivo* by generating reporter-mouse embryos (Visel et al., 2007). Such higher throughput assays have revealed that enhancers display diverse and highly precise spatiotemporal activities that can act combinatorically to generate complex gene expression patterns. While housekeeping genes are mostly depleted from regulatory landscapes harboring distally acting elements, developmental or tissue specific genes are controlled by an average of five enhancers, which can appear in clusters (super enhancers, SEs) and have largely overlapping but not necessarily interchangeable activity (Hnisz et al., 2013; Will et al., 2017).

Thus, thousands of putative enhancers in various tissues and cell-types contribute tremendously to the diverse spatiotemporal activity and complex regulation of genes (Ruf et al., 2011; Symmons et al., 2014; Visel et al., 2007). However, this also shows that 97% of the genome is dispersed with enhancers.

1.1.5 Long-range regulation by distal enhancers

Promoters and enhancers are crucial for the regulation of genes which can be mapped and validated by various techniques. However, how an enhancer induces transcription remains largely unclear. As enhancers are thought to function in a position- and orientation-independent manner to activate their target genes (Bulger and Groudine, 2011a) and can be located very distal to their cognate promoter (Long et al., 2020), physical proximity is required to initiate transcription.

Long-range regulation frequently occurs for developmental genes, with linear distances ranging from 1-1.45 Mb (Lettice et al., 2003; Long et al., 2020). The common idea of how a promoter and a distal enhancer find each other in the nuclear space to activate a gene's expression is by the looping of the DNA fiber (Kagey et al., 2010). Chromatin loops which facilitate long-range regulation of distal enhancers and their target promoters, are frequently observed in regulatory landscapes of genes. Different models have been proposed to facilitate looping between a promoter and an enhancer which in turn leads to gene activation (Furlong and Levine, 2018). The current prevalent model for enhancer function is based on homotypic interactions between tissue-specific TFs bound at both elements, thereby bridging their interaction and recruiting the transcription machinery (Fig. 1-3 B). Like proximal promoter regions, enhancers comprise specific TF binding motifs (see above). These specific transacting factors integrate complex information into active transcription at their target genes (Mitchell and Tjian, 1989; Spitz and Furlong, 2012). By the recognition and binding of individual DNA motifs situated in promoters and enhancers, TFs recruit the co-activators including Mediator and p300 that are necessary for transcription apparatus recruitment (Bulger and Groudine, 2011b; Malik and Roeder, 2005).

Despite serving as a binding platform for TFs, liquid-liquid phase separation has recently been proposed as a possible mechanism for TF-mediated gene activation. Here, transcription is induced by the formation of active hubs with high concentrations of the transcription apparatus containing TFs, mediator and RNA PolII (Cho et al., 2018; Hnisz et al., 2017; Sabari et al., 2018). Formation of membrane-less condensates within the nucleus containing TFs, co-activators and RNA PolII, have been described for the pluripotency factor OCT4 in human induced pluripotent stem cells, resulting in gene activation (Boija et al., 2018). Moreover, alterations of the phase-separating capacity of the intrinsically disordered regions (IDR) of the HoxD13-protein affects transcriptional regulation *in vivo* and has been implicated in congenital malformation (Basu et al., 2020).

In summary, CREs in combination with TFs are essential for establishing the complex and precise gene expression patterns in time and space for cell fate specification and development in multicellular organisms. Yet, physically, how long-range regulation is established and an enhancer finds and regulates its target promoter remains elusive. As they can be located very distal from one and another, the identification which enhancer regulates which gene is another challenge.

1.2 The 3-dimensional genome

Numerous studies demonstrated the requirement of physical proximity of a given enhancer toward its target promoter for gene activation (Chen et al., 2018; Deng et al., 2014; Paliou et al., 2019; Williamson et al., 2016). Yet, how this proximity is controlled mechanistically remains unknown. Significantly, increasing evidence from the last decades connects the 3D-organization of the chromatin fiber with transcriptional regulation, facilitating the functional assignment of CREs to specific target genes. Multiple methods exist to study genome organization, which all provide complementary, yet distinct insights into the 3D architecture. In general, bulk-techniques like proximity ligation assays complement methods on the single cell level like FISH (Huang et al., 2020). In the following sections, I will elaborate on the current state of knowledge.

Within the nucleus the DNA fiber is highly organized. Already in the 1960's electron microscopy demonstrated the existence of dark spots within the nucleus. Later it became evident that these spots are inactive regions of the DNA, termed heterochromatin positioned at the nuclear periphery, which spatially separates from active euchromatin in the center of the nucleus. Moreover, the discovery of chromosome territories in the late 1980 revealed that chromosomes are preferentially locating in a non-random fashion at distinct nuclear areas (Lichter et al., 1988; Pinkel et al., 1988) (Fig. 1-4 A). Since then, the influence of spatial positioning and organization on gene regulation remained in the focus of research (reviewed in Cremer and Cremer, 2001; Haaf and Schmid, 1991).

Technological advances, particularly with the development of Chromosome Conformation Capture (3C)-based techniques, have significantly increasing insights into the organization of the chromatin fiber on a locus-level, helping to associate enhancers with target promoters (Dekker et al., 2002).

1.2.1 CCC-based techniques reveal several organizational layers of the 3D-genome

3C-techniques create a snapshot of any two genomic loci in the nucleus based on reversible crosslinking of DNA-DNA or DNA-protein contacts that are in close 3D proximity. Through this, a genome-wide map of frequent chromatin interactions can be generated. Various methods have been developed on the principle of proximity ligation (reviewed in (Dekker et al., 2013). In brief, physically proximal fragments are crosslinked, subjected to digestion and finally re-ligated (Fig. 1-4 D). This creates circular hybrid molecules whose frequency of abundance translates into interaction frequency, analyzed by quantitative real time-PCR (one vs. one, 3C). The combination of 3C with deep sequencing led to the further development of various proximity ligation based-methods. 4C-seq is based on an additional PCR-enrichment of a genomic region of interest and so creates a one-versus-all interaction map for a specific viewpoint following sequencing.

CaptureC includes additional enrichment of multiple viewpoints of one library (e.g. several promoters) by (RNA) probes followed by deep sequencing (many vs. all). Finally, sequencing of all obtained ligation fragments generates a genome-wide interaction map (all vs. all, HiC). Capture HiC is a combined derivative of HiC and CaptureC which creates high-resolution interaction frequency maps of an enriched genomic region of interest (up to 5Mb) without sequencing all obtained fragments.

Compartments

Several genome-wide studies using HiC revealed numerous inter- and intrachromosomal organizational layers of interphase chromosomes. Generally, higher-order chromatin organization like compartmentalization segregates chromatin according to its functional state into active A- and repressed B-compartments (Fig. 1-4 B). This separation is likely based on homotypic interaction depending on the transcriptional level, epigenetic state, compaction (eu- vs. heterochromatin), replication timing and association with the nuclear envelope (Lamina-associated domains; LADs) (Lieberman-Aiden et al., 2009).

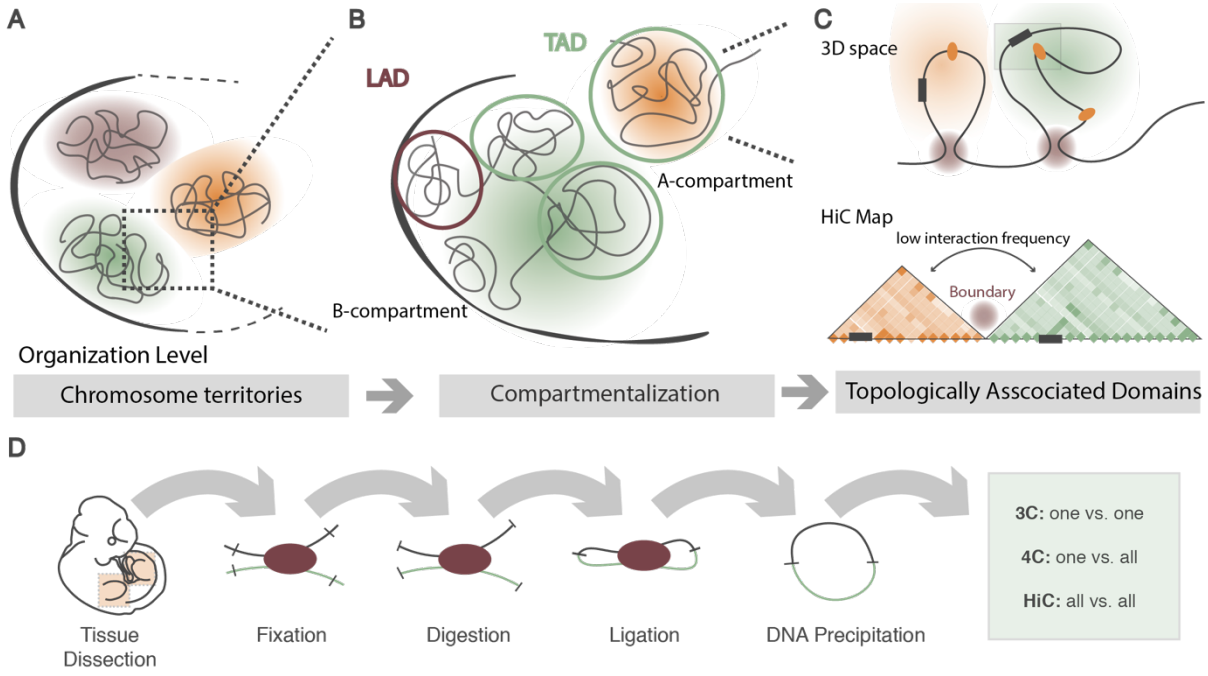

Figure 1-4 Organizational layers of the 3D-genome. (A) Within the nucleus chromosomes locate at distinct areas called chromosome territories (B) Active and repressed regions on chromosomes separate into A- and B-compartments. Furthermore, the genome is partitioned into TADs (green circles) which largely overlap with LADs (red circle) and non-LADs. (C) Detailed view of two exemplary TADs (orange and green) separated by a boundary (red). Genes are depicted with black boxes, enhancers with orange ovals Below corresponding HiC interaction frequency map on a linear scale (D) Basic principles and workflow of chromosome conformation capture (3C)-based techniques: After tissue dissection the cells are fixed, digested with a restriction enzyme, re-ligated and the DNA

precipitated. Further processing of the created library depends on the applied technique. The figure was adapted from *(Dekker et al., 2013; Robson et al., 2019)*.

Topologically Associating Domains (TADs)

Later studies discovered the partitioning of the chromosomes along their length into sub-megabase units called topologically associating domains (TADs, Fig. 1-4 C) (Dixon et al., 2012; Nora et al., 2012). Interestingly, the architectural units largely overlap with the described regulatory landscapes of developmental genes, as demonstrated by sensor insertions (Ruf et al., 2011; Symmons et al., 2014). In these studies, the reporters recapitulated the gene-specific regulatory information until inserted into a demarcated neighboring TAD. Thus, TADs assist in the association of an enhancer with a promoter.

With an average size of 850kb, TADs are defined by their preferential self-interaction and are highly conserved between species and cell-types (Dixon et al., 2012; Nora et al., 2012). Importantly, neighboring TADs are insulated by boundaries, enriched in directly DNA binding CTCF and the structural maintenance of chromosomes (SMC)-complex cohesin (Dixon et al., 2012; Nasmyth and Haering, 2009; Nora et al., 2012; Rao et al., 2014). How or if CTCF establishes and maintains TADs, or even what TADs physically represent, remained a black box for some time. Currently, the most prevalent mechanism is the ATP-dependent loop extrusion model, strongly supported by computational polymer modelling (Fudenberg et al., 2016; Nuebler et al., 2018). Here, the formation of chromatin loops likely depends on the CTCF site orientation towards each other. Convergently orientated binding sites form loops and appear enriched in divergent orientation at TAD boundaries (de Wit et al., 2015; Guo et al., 2015; Sanborn et al., 2015). After loading onto the DNA by Nipbl, cohesin progressively extrudes the chromatin fiber bidirectionally. For CTCF motifs oriented in direction of active extrusion, the cohesin complex is thought to bypass until stopped by a CTCF bound toward it (Fig. 1-5) (Fudenberg et al., 2016; Nuebler et al., 2018). As this process is highly dynamic, cohesin constantly dissociates from chromatin through the releasing factor Wapl (Fig. 1-5) (Busslinger et al., 2017; Haarhuis et al., 2017; Wutz et al., 2017). Importantly, many predictions have now been tested experimentally. Yet direct evidence for the proposed mechanism *in vivo* is methodologically challenging, as both architectural proteins are critical during cell cycle. However, experiments conducted in mouse zygotes demonstrated cohesin (SscI) dependency of TADs and links TAD formation and size with the linear density of the releasing factor Wapl (Gassler et al., 2017). Furthermore, imaging of Xenopus egg extracts showed symmetric extrusion by cohesin during interphase (Golfier et al., 2020). A bidirectional loop extrusion of the DNA is additionally supported by single molecule imaging of a homodimer formed by human cohesin (Kim et al., 2019). Moreover, depletion experiments of either CTCF- or cohesin, circumventing deleterious cell-cycle effects,

demonstrated a genome-wide TAD-loss *in vitro*, stressing their relevance for TAD formation (Nora et al., 2017; Rao et al., 2017; Schwarzer et al., 2017).

Taken together, due to the high correlation with gene regulatory landscapes, TADs are believed to provide an insulated microenvironment for enhancer-promoter communication, shaped by CTCF and cohesin. Yet, the functional relation between TADs and gene regulations remains puzzling.

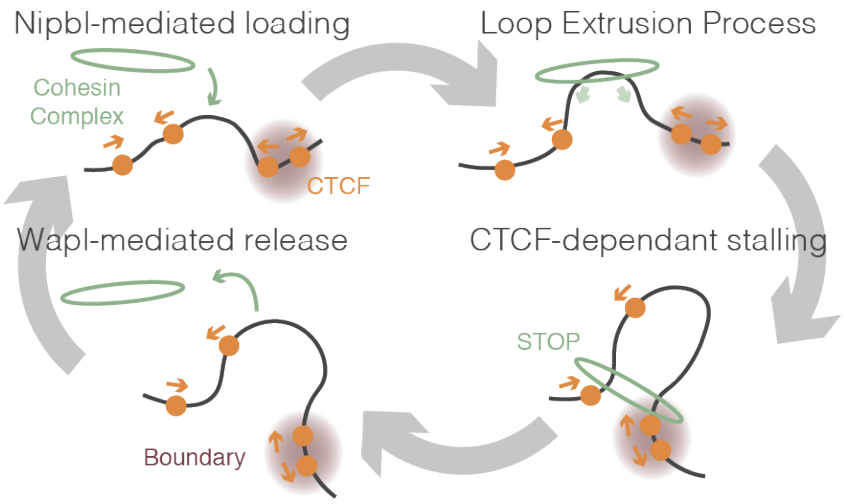

Figure 1-5 Loop extrusion model. After Nipbl-mediated loading onto the DNA, the cohesin complex starts to extrude in a bidirectional fashion (B), till stalled by CTCF orientated towards the direction of extrusion (C). CTCF sites directed in orientation with extrusion can be bypassed by the cohesin complex.

1.2.2 Dynamics of intra-TAD interactions: CTCF in shaping promoter-enhancer contacts

TADs are defined by HiC and seem to represent insulated domains, which can seemingly function as regulatory units due to sensor insertion studies. Although the restricting function of TADs in setting limits to regulatory activity seems clear, how the internal TAD structure could facilitate enhancer-promoter communication is less understood. De Laat and Duboule proposed a permissive model, where enhancers and promoters remain in an invariant state of proximity by the 3D architecture of the genome until activated by tissue-specific TFs. In contrast, the instructive model describes de-novo interactions or looping of the chromatin fiber, resulting in the immediate active transcription of a gene (Laat and Duboule, 2013). Yet, the direct influence of CTCF in both scenarios on gene activation remains controversial. Matching these models, two modes of contacts have been observed in various studies.

Indeed, interactions within TADs are highly dynamic during cell-fate specification and development, but also within a cell population. As HiC data reflects batch analysis of a pool of fixed nuclei, microscopy and single cell studies have shown that TADs are highly variable. Within a homogenous population of cells, high-resolution FISH and single cell-HiC captured the dynamics of loop interaction within domains (Bintu et al., 2018; Stevens et al., 2017; Szabo et al., 2020). Furthermore, high-resolution HiC following the differentiation from mESC to NPCs demonstrated cell-type specific dynamics of TAD formation and intra-TAD interactions (Bonev et

al., 2017). These TAD transitions during cell-fate specification and, moreover, immediate transcriptional responses upon external stimuli were supported by similar results obtained in transdifferentiating of immune cells (Stik et al., 2020). Thus, TADs are indeed highly dynamic, yet, their impact on enhancer-promoter contacts remains debated.

High-throughput techniques like HiC, microC and HiChIP have revealed numerous genome-wide enhancer-promoter loops at various loci (Hsieh et al., 2016; Krietenstein et al., 2020; Mumbach et al., 2016). Matching the permissive model, *Shh* and its limb enhancer ZRS are maintained in close proximity by two pairs of convergent CTCF sites at the edges of the *Shh* TAD. Paliou et al. demonstrated that disruption of this preformed loop reduces *Shh* expression by 50% in embryonic limb buds, indicating that tissue-invariant structures are important to sustain full regulatory function in some cases (Paliou et al., 2019). Besides the frequent co-occupancy at TAD boundaries, CTCF and cohesin are found within TAD substructures, forming mostly invariant chromatin loops (Bonev et al., 2017; Dixon et al., 2012; Nora et al., 2012; Phillips-Cremins et al., 2013; Rao et al., 2014). CTCF sites within TADs are thought to support promoter enhancer interactions. More precisely, together with cohesin, CTCF sites proximal to a promoter likely facilitate its sampling of the regulatory environment for a suitable enhancer. However, tissue-specific interactions which are CTCF-dependent have been described to be rather constitutive long-range interactions in invariant sub-domains than dynamic de novo-contacts (Phillips-Cremins et al., 2013). It appears that newly established, and thus dynamic enhancer-promoter contacts of a certain tissue during a differentiation process, do not necessarily rely on CTCF to find their cognate equivalent. Rather it seems they establish de novo due to tissue-specific TF and/or the polycomb interaction network (Bonev et al., 2017). Accordingly, the characterization of hundreds of developmental gene landscapes by CaptureC in the developing mouse limb bud demonstrated strong association of tissue invariant loops with CTCF, whereas dynamic interactions during development form somewhat independently and hinge more on the epigenetic state (Andrey et al., 2016). Recent data confirmed these results and revealed CTCF-dependent and -independent promoter-enhancer contacts during neural differentiation, where the former seem to promote long-range interactions with promoter-proximal CTCF sites serving as an anchor to reel in potential regulatory elements for activation (Kubo et al., 2021).

The dynamics of lineage-specific CTCF-dependent loops require regulating mechanisms. Active transcription and DNA-methylation have been described to control CTCF-mediated promoter-enhancer contacts of genes. The deletion of the RNA-binding domain (RBDi) of CTCF led to a marked genome-wide decrease in strength of CTCF-dependent loops in mESCs, identifying mainly two regimes of CTCF-anchored loops, RNA-dependent and RNA-independent (Hansen et al., 2019). This potentially explains cell-type specific domain formation during development, where active transcription stabilizes long-range contacts. Furthermore, DNA-hypermethylation in

IDH mutant gliomas led to reduced CTCF-binding and consequently a loss of insulation between TADs. Specifically, the lost insulation allows an enhancer to ectopically activate an oncogene, resulting in cancer (Flavahan et al., 2016).

Taken together, CTCF has been demonstrated in the formation of various invariant and dynamic chromatin loops, matching the permissive and the instructive model (Laat and Duboule, 2013). Yet, the minority of dynamic promoter-enhancer contacts can be explained by CTCF.

1.2.3 Chromatin loop-formation beyond CTCF

Numerous chromatin loops are CTCF-mediated, yet, other CTCF-independent contacts are formed within TADs. Other architectural proteins apart from CTCF have been implicated in loop formation. Direct evidence links the zinc finger protein Yin Yang 1 (YY1) to cell type- specific looping. Identified at the base of enhancer-promoter loops during neural lineage commitment, the knockdown of the gene specifically disrupts the previously observed tissue-specific interactions (Beagan et al., 2017). Moreover, YY1-HiChIP experiments demonstrated strong association with enhancers and promoters and their corresponding interaction (Mumbach et al., 2016). Additionally, artificial tethering of YY1 to a mutated DNA binding site rescued interactions lost prior, as seen by 4C (Weintraub et al., 2017). Another lineage-specific TF at the beta-globin locus, the LDB1 complex, likely mediates looping via self-interaction. Artificial tethering of LDB1 to the gene promoter forced loop formation independently of the mediator or cohesin complex. However, the complex was also described to loop directly with promoter-proximal CTCF binding sites (Lee et al., 2017). Furthermore, developmental genes and their regulatory sequences in a repressive epigenetic state are highly enriched with the polycomb repressive complex (PRC). The silencing of genes by PRC cooperates with and influences chromosome conformation. By the formation of chromatin-associated multi-protein complexes depositing H3K27me3 marks, repressed elements establish long range contacts, thereby shaping the chromatin structure on a sub-megabase scale. A classic example for this kind of interaction is the transition from repressed *Hox* domains into transcriptionally active TADs (reviewed in (Entrevan et al., 2016).

In addition, transcription itself was described as a chromatin remodeler, dynamically changing genome architecture. Rapid changes and remodeling of TADs in line with compartment transitions from repressed B to active A were associated with active transcription. These dynamics were observed after influenza A infection in macrophages, where elongated RNA PolII disrupts cohesin-association with CTCF and therefore locus-decompaction, enabling TF binding (Heinz et al., 2018). In general, active transcription seems to be fundamental for dynamic transcriptional control itself. Since the discovery of actively transcribed enhancers resulting in so-termed “enhancer RNAs” (eRNAs) (Kim et al., 2010; Santa et al., 2010) several studies revealed their functional implication in lineage-specific looping, chromatin modifications and thus

transcriptional regulation (Arnold et al., 2020). Further studies relate actively transcribed retroviral elements (RE) like ERVs with variant cell-type specific boundaries. Additionally REs are concurrent with the emergence of new boundaries during evolution, where jumping viruses give rise to new domains (Zhang et al., 2019).

Taken together, dynamic contacts of enhancer and promoters in individual cells have many layers of regulation, ranging from structures mediated by architectural proteins, homotypic interactions of histone modifications or TF to transcription itself.

1.2.4 Emerging influence of chromatin architecture on gene regulation

Based on our understanding of TADs, functionally, they facilitate spatial proximity of enhancers with their target promoter and thus provide a regulatory framework for genes. In line with this, TADs largely overlap with regulatory landscapes of developmental genes (Ruf et al., 2011; Symmons et al., 2014). Accordingly, growing numbers of TAD alterations have been described to be causal for pathogenic phenotypes. Summarized in Figure 1-6, several genomic rearrangements including large-scale duplications, deletions or inversions have been associated with congenital malformations and cancer (Spielmann et al., 2018). For example, a tandem duplication of 1.7 Mb at the *Sox9/Kcnj2*-locus including a TAD boundary generates a so-called neo-TAD (Fig. 1-6 D and Fig. 1-8). Within this newly emerged domain the *Sox9* regulatory landscape is fused with a duplicated copy of the gene *Kcnj2*. Misexpression of the potassium channel in a *Sox9*-like pattern in the developing limb bud results in Cooks-syndrome, characterized by aplasia of nails and brachydactyly, defined by missing middle, but elongated terminal and proximal, phalanges (Cooks et al., 1985; Franke et al., 2016). Furthermore, multiple megabase-scaled inter-TAD spanning structural variants including a TAD boundary at the *WNT6/IHH/EPHA4/PAX6*-locus led to a rewiring and activation of enhancers with non-target promoter, resulting in limb malformations (Lupiáñez et al., 2015). Moreover, this process of enhancer “adoption” or “hijacking” caused by the disruption of domains by structural variants has been described in cancer (Hnisz et al., 2016; Weischenfeldt et al., 2017).

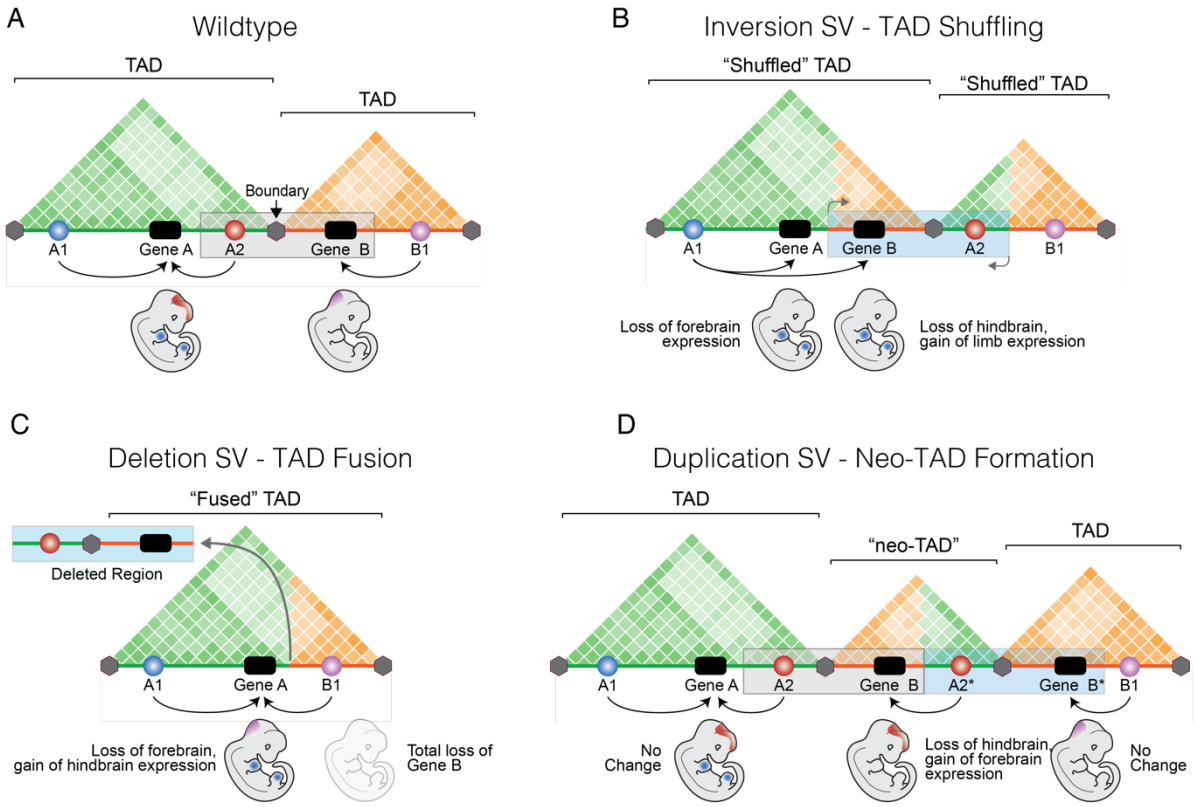

Figure 1-6 Structural Variants on a megabase-scale. TAD shuffling can be causative for gene misexpression or loss of expression by rewiring of enhancers with new promoters or isolation from their cognate genes. (A) Wildtype TADs of Gene A and Gene B with their corresponding enhancers and their activity patterns, depicted in the embryos below. The two neighboring domains are separated via a boundary. Genes: black squares; enhancers: blue, red and violet circles; Boundary: grey hexagon. (B) Inter-TAD spanning Inversion including a boundary leads to rewiring of enhancer A1 with Gene B and isolation of enhancers A2 and B1. (C) Deletion results in TAD fusion of both domains, resulting in a loss of Gene B and enhancer A1. Furthermore, enhancer B1 rewires with Gene A. (D) Tandem duplication including a boundary leads to the formation of a neo-TAD. Additional to wildtype expression Gene B gets ectopically expressed in the expression pattern of Gene A driven by enhancer A2 (The figure was adapted from *(Robson et al., 2019)*

Hence, TADs are believed to define coherent neighborhoods of developmental genes and provide an architectural framework for complex regulatory landscapes of genes. However, the direct influence of TADs on gene regulation remains poorly understood. Several experimental approaches aimed to unravel the significance of the architectural units for gene regulation. Connecting transcriptional control with nuclear organization genome-wide remains challenging due to side-effects during cell cycle directly influencing the results, as both CTCF and the cohesin complex are part of the cell division apparatus for holding together sister chromatids during division.

Genome-wide depletion experiments highly support the involvement of the two architectural proteins in TAD formation by loop extrusion (Nora et al., 2017; Rao et al., 2017; Schwarzer et al.,

2017). Using an auxin-inducible degron system, Nora et al. could circumvent the deleterious cell cycle effects of a CTCF knock out and elegantly demonstrated a loss of TADs after CTCF depletion in mESCs (Nora et al., 2017). Although TADs vanished, compartmentalization remained unchanged. Intriguingly, about 20% of analyzed boundaries were not affected. However, the immediate effects on gene regulation after one day of depletion and TAD-loss were surprisingly low (370 mis-regulated genes) and a further increase of mis-regulation (ten-fold) at later points in time is most likely due to cell proliferation defects. Remarkably, 80% of downregulated genes lost a CTCF site near their TSS. Moreover, tissue-specific cohesin deletion in mouse liver cells and a human cancer cell line exhibited similar results (Rao et al., 2017; Schwarzer et al., 2017). While TADs disappeared, compartmentalization became even more pronounced in the cells. Similar to the CTCF-depletion experiment, no significant transcriptional changes could be detected, probably because of non-dividing terminally differentiated hepatocytes and thus lack of mitotic effects, as hypothesized earlier (Schwarzer et al., 2017).

However, the results outlined above stand in contrast to described alterations of TADs, summarized as structural variants, which potentially lead to gene misexpression by the rewiring of enhancer and promoter contacts, ultimately resulting in ectopic activation (Melo et al., 2020; Spielmann et al., 2018; Weischenfeldt et al., 2017). The large overlap of the genes' regulatory landscapes with TADs could explain why TAD shuffling possibly connects wrong enhancers with non-target promoters or isolates regulatory information from its cognate gene.

1.3 Deciphering TAD function on developmental gene regulation

The relation between TADs and gene regulation remains puzzling. Several studies link alterations with severe effects *in vivo*, where the loss of TADs has minor effects *in vitro*. As the discussion about the influence of TADs remains controversial, here, we aimed to challenge the contradictory results of previous studies through investigating the loci of two well-studied developmental genes.

1.3.1 Editing the regulatory genome

Genetic targeting and the introduction of mutations of interest to study specific function of regulatory elements or genome architecture have been an elaborate and time-consuming process. Since the discovery of CRISPR/Cas9 (Clustered Regularly Interspaced Short Palindromic Repeats; CRISPR-associated) genome editing became significantly easier (reviewed in (Doudna and Charpentier, 2014). In essence, the powerful technique takes advantage of the primitive bacterial immune response to viruses, by guiding the endonuclease Cas9 via single guide RNAs to unique genomic positions in order to create efficiently targeted double strand breaks (DSBs) (Fig. 1-7 A). During the cell cycle, the targeted cell mainly repairs the DSB with two alternative pathways: the error-prone non-homologous end joining (NHEJ) occurs from G1- to M-

phase during cell cycle and connects the two loose ends of the double-helix randomly, thereby creating so-called InDel-mutations of a couple of basepairs. In comparison, homology-directed repair (HDR) utilizes the sister chromatid as a repair template and, avoids possible mutations through this (Fig. 1-7 B). Applying an external template like Oligonucleotides or plasmids containing a sequence of interest additionally exploits the repair mechanism for the integration of a specific DNA-sequence.

In vivo editing of embryonic stem cells or blastocytes can be easily applied to generate transgenic mouse models relatively fast. Utilizing one or two single guide RNAs (sgRNAs) in combination with Cas9 creates a powerful tool to generate Indels or structural variants (SV) on a megabase-scale (CRISVar, Fig. 1-7 C) (Kraft et al., 2015). Furthermore, the technique facilitates efficiently targeted insertions based on homology directed repair (HDR) ranging from a single nucleotide exchange using oligonucleotides to insertions of several kilobase in size and undergoes constitutive progress (Andrey and Spielmann, 2016; Byrne et al., 2015)

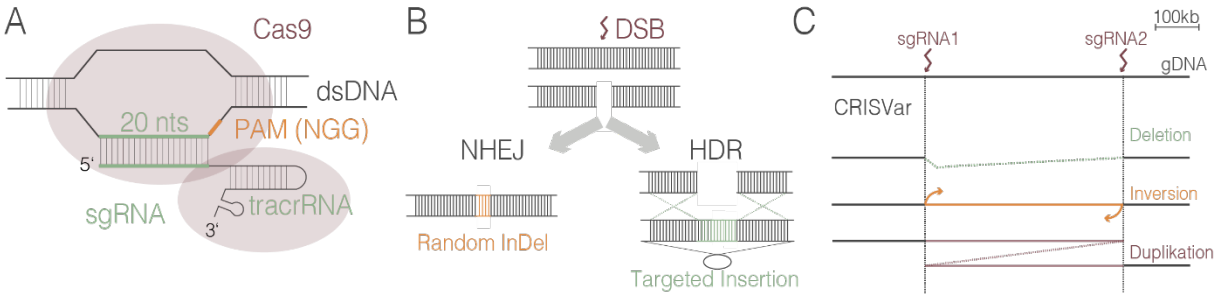

Figure 1-7 CRISPR/Cas9 mediated genome editing. (A) CRISPR/Cas9. Red: Cas9-enzyme, Orange: PAM-sequence consisting of the nucleotides NGG. Green: 20 nucleotide-long locus-specific guide RNA-sequence followed by the tracrRNA (B) Repair mechanisms. NHEJ: error-prone Non-homologous end-joining. HDR: Homology-directed Repair using either the homologous sister chromatid or an externally applied template as a matrix for the repair. (C) Principle of CRISVar. Utilizing two sgRNAs in combination with Cas9 induces two DBS breaks which can result due to error-prone repair mechanism in distinct SVs.

Here, we wanted to utilize CRISPR/Cas9-based genome editing, to create targeted mutations at the *Sox9/Kcnj2*- and *Shh*-locus. By generation of transgenic mouse lines we wanted to systematically assess the effect of TADs on developmental gene regulation at the *Sox9/Kcnj2*- and *Shh*-locus.

1.3.2 Spatial organization at the *Sox9/Kcnj2*-locus

During development the transcription factor *Sox9* plays a crucial role in sex determination and chondrogenesis (see above). The gene itself is situated in a vast gene desert, harboring the regulatory information for its precise spatiotemporal expression patterns in several tissues. cHiC of the locus revealed its spatial organization into two TADs separated by a TAD boundary (Franke

et al., 2016). The self-interacting capacity of both domains was further confirmed by 4C-experiments with viewpoints in either of the genes' promoters in various tissues (Franke, 2017). Particularly interesting are the distinct expression patterns of *Sox9* and its neighboring gene *Kcnj2* in the developing limb bud in the mouse at E12.5 (Fig. 1-8 A). However, we know from previous studies at the locus that the gene *Kcnj2* generally is responsive to *Sox9*-enhancers. A TAD-spanning duplication including the TAD boundary results in the formation of a neo-TAD (Fig. 1-8 B), causing pathogenic misexpression of *Kcnj2* causing a limb malformation termed Cooks-syndrome (see above) (Franke et al., 2016). Although in wildtype some *Sox9*-regulatory enhancers are in closer proximity to *Kcnj2*, their activity is confined to *Sox9*. Yet, deletion of the separating TAD boundary has no significant effect on gene expression (Franke et al., 2016). This raises the question of the functional importance of TADs and their boundaries for gene regulation.

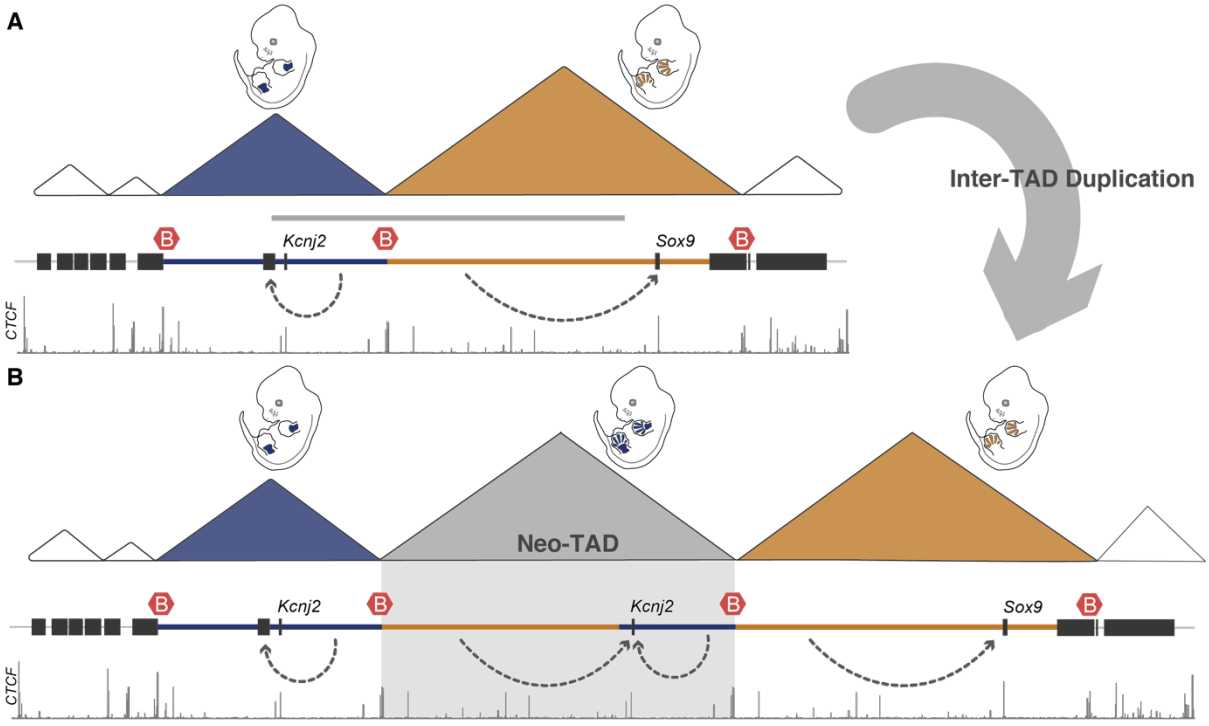

Figure 1-8 Spatial Organization of the *Sox9*/*Kcnj2*-locus (A) The locus consists of two TADs harboring *Kcnj2* and *Sox9*. Generally, both domains are separated by a TAD boundary (red hexagon) that restricts regulatory activity to the corresponding TAD. *Sox9* is active in the digital anlagen of E12.5 limb buds (orange) where *Kcnj2* is solely expressed in the proximal part (blue). A TAD-spanning duplication including the TAD boundary (grey bar) leads to the formation of a neo-TAD (B). In the neo-TAD (highlighted in grey), huge parts of the regulatory domain of *Sox9* (orange) are fused to a duplicated copy of *Kcnj2*. By this, the potassium channel becomes misexpressed in a *Sox9*-like pattern resulting in limb malformation called Cooks-syndrome.

1.3.3 Spatial Organization at the *Shh*- locus

Shh is a gene involved in a plethora of developmental processes. Aside from its well-studied function in the limb, *Shh* expression is regulated by multiple enhancers dispersed throughout its regulatory landscape in a vast gene desert correlating with its TAD, as shown by cHiC and 5C-experiments (Fig. 1-9 A) (Paliou et al., 2019; Williamson et al., 2019) More than a dozen enhancers drive *Shh* activity in a variety of tissues, crucial to the development of structures of the epithelial linings and the central nervous system (CNS) (Anderson et al., 2014; Sagai et al., 2009, 2019; Seo et al., 2018). The diversity of the complex spatiotemporal expression patterns is summarized in figure 1-9 B and C. A prominent example for long-range interaction is the regulation of the *Shh* and its limb-specific enhancer, the zone of polarizing activity regulatory sequence (ZRS). The gene promoter and the ZRS are situated almost 1 Mb apart at the edges of the *Shh* TAD (Fig. 1-9 A and B). This distal ZRS-position is conserved throughout species and is the only enhancer driving *Shh* expression in the developing limb (Lettice et al., 2003; Sagai et al., 2004, 2005). Enhancers are thought to function orientation- and position-independently. Systematic studies of enhancers which do not rely on artificial assays, however, are pending. Regulatory sensors inserted throughout the domain recapitulate the limb-specific signal but are deprived of it once located in the adjacent TAD of the developmental gene *Engrailed-2* (*En2*) (Symmons et al., 2014). Yet, the strength of the reporter signal differs at the various positions within the domain, raising the question of whether enhancers fully function at different genomic locations and to which extent the chromatin context at a given position influences its activity within a TAD.

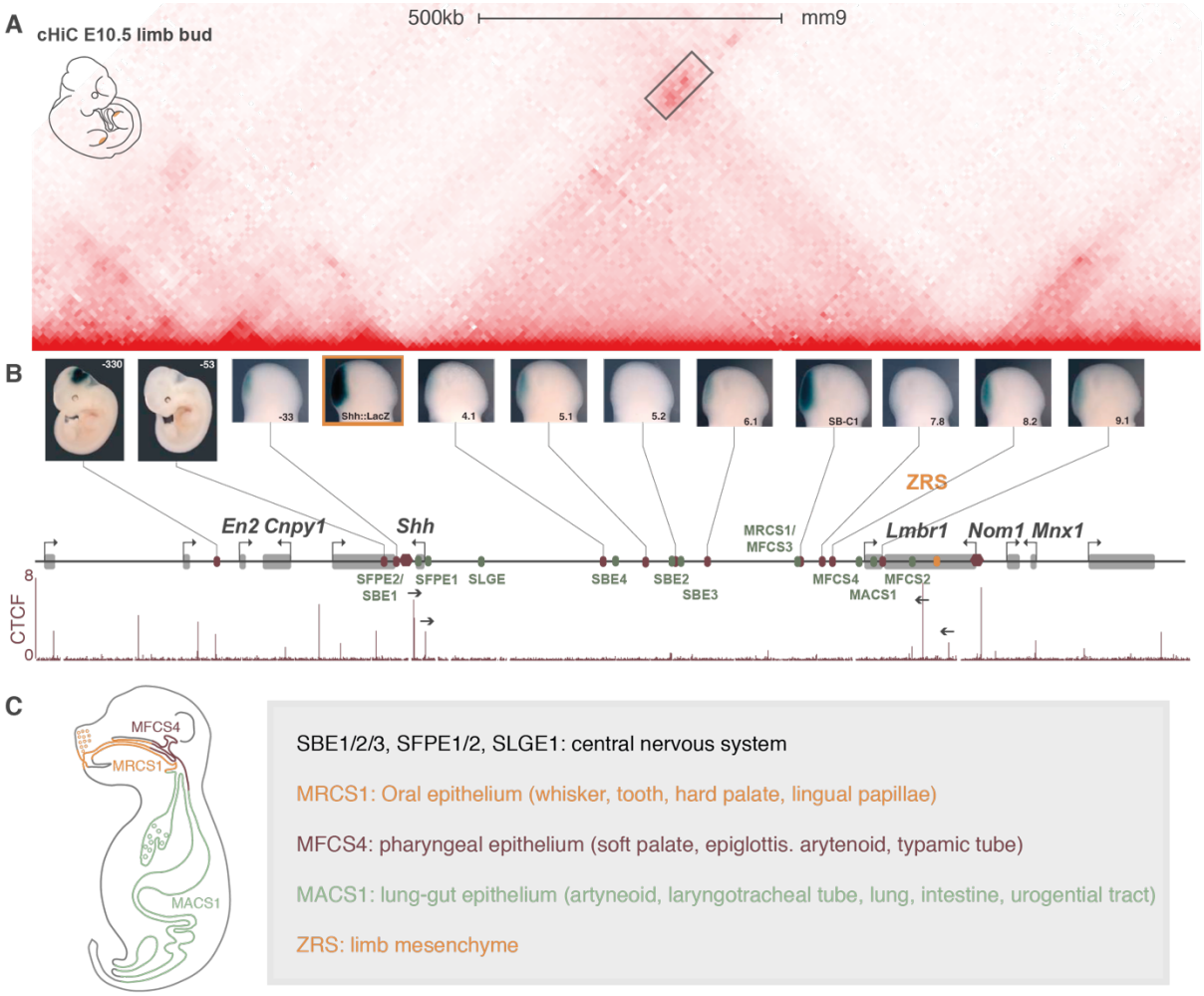

Figure 1-9 The *Shh* regulatory landscape highly correlates with the 3D conformation of the locus. (A) cHiC of murine E10.5 limb buds. Grey box highlights the preformed structure shaped by the two pairs of CTCF-site flanking either the *Shh* promoter or the limb-specific enhancer ZRS, directed in a convergent orientation toward each other (The figure was adapted from Paliou et al.2019). (B) LacZ sensor insertions throughout the *Shh*-TAD are taking up the ZPA-signal. Once positioned outside the domain in the neighboring TAD the reporter signal is lost in the limb bud and recapitulates the expression pattern of *En2* in the developing midbrain. Below: CTCF-ChIP-seq of murine E10.5 limb buds. Orientation of CTCF sites are indicated by arrowheads. (C) Expression domains of *Shh* in the epithelial linings of the developing mouse embryo besides the limb bud and the central nervous system and corresponding characterized enhancers depicted in (B) as green ovals: Oral epithelium (orange), Pharyngeal epithelium (red), lung-gut epithelium (green). Light grey boxes: genes; green ovals: *Shh* enhancers; orange oval: ZRS, Grey box: Tissue-invariant CTCF-mediated chromatin loop. The figure was adapted from Symmons et al. 2014; Sagai et al. 2009).

2. AIM OF THIS STUDY

Sox9 and *Shh* are both key developmental genes which are particularly important during limb development. Their complex spatiotemporal expression patterns are controlled by numerous CREs embedded in gene deserts contained in their respective TADs. Despite the correlation between 3D structure and gene regulatory landscapes, particularly at these loci, precisely how TADs affect transcriptional control remains puzzling. On the one hand, SVs at the *Sox9*, *Shh*, and other loci have been shown to cause gene misexpression and disease in patients and mouse models. On the other hand, genome-wide loss of TADs has surprisingly mild effects on gene regulation in cell culture experiments.

Here, we wanted to test the influence of the 3D chromatin structure for gene regulation during embryonic development. The easy accessibility of the limb and the rich knowledge of the gene regulatory networks controlling its development makes the limb an excellent model. By assessing the influence of TADs in two parts, we addressed two distinct questions.

2.1 PART 1: Functional dissection of the *Sox9/Kcnj2*-TAD

In the first part, we focused on the targeted deconstruction of the TADs at the *Sox9/Kcnj2*-locus. This approach is designed to disentangle the function of the TAD boundary and TAD substructure and their individual contribution to TAD formation and gene regulation.

The systematic targeting of the TAD boundary and major intra-TAD CTCF bindings sites shows their combinatorial role in TAD formation and their (limited) importance for gene regulation. In addition, a series of SVs shows how these gain-of-function mutations actively restructure the TADs at the locus and induce gene misexpression.

2.2 PART 2: Enhancer Shuffling at the *Shh*-locus

The second part investigates the relevance of TADs for the functionality of an individual enhancer. Enhancers were originally defined to be able to regulate a gene regardless of its position or orientation. Here, we took advantage of *Shh* expression in the limb bud, which is controlled by a single non-redundant enhancer, the ZRS. In a series of mutant mice the ZRS is repositioned within and outside the *Shh*-TAD to identify the relevance of the genomic position for accurate gene activation and regulation within regulatory domains. By interpreting the ability of the repositioned ZRS to activate *Shh* at its new position, we show how TADs restrict enhancer activity, but are not the only factor in this process.

3. MATERIAL

3.1 Instruments

All instruments used in this study are listed below in table 3-1.

Table 3-1 Instruments used in this study.

Instrument	Type	Source
Table Top centrifuge	5414D	Eppendorf
Cooling centrifuge	5417R	Eppendorf
Cooling centrifuge	Avanti J-E	Beckman-Coulter
Cooling centrifuge Rotor	JLA16.250	Beckman-Coulter
Thermocycler	GeneAmp PCR System 2700, 2720, and 9700	Applied Biosystems
Real-Time Thermocycler	ABIPrism HAT 79000 RT	Applied Biosystems
Steromicroscope	MZ12 Discovery V12	Zeiss
Camera	DFC420	Leica
Light Source	KL1500 LCD	Leica
Photometer	Nanodrop 2000	Thermo Scientific
Clean Bench	HERASafe	Thermo Scientific
CO2-Incubator	HEPA Class 100	Thermo Scientific
Tranilluminator		Hertenstein
Micro-CT	Skyscan 1172	Brucker microC

3.2 Chemicals

All chemicals were obtained from Merck (Darmstadt), Roth (Karlsruhe) or Sigma-Aldrich (Hamburg, Seelze, Schnelldorf and Steinheim) in analytical grade quality if not stated differently.

3.3 Buffers

All common buffers and solutions were prepared according to Green&Sambrook (Green and Sambrook, 2012) and are listed in table 3-2 and table 3-3.

Table 3-2 Buffers and Solutions for WISH.

Buffer	Composition
10xPBS -DEPC	1.37 M NaCl, 27 mM KCl, 100 mM Na2HPO4, 20mM KH2PO4, adjust pH to 7.4 with HCl, in DEPC- H2O, autoclave
4%PFA-PBS	Dissolve 40 mg/ml PFA in 1x PBS (DEPC), heat to 55°C until PFA is dissolved, adjust pH to 7.4 with HCl
Alkaline phosphatase buffer	0.02 M NaCl, 0.05 M $MgCl_2$, 0.1% Tween 20, 0.1 M Tris pH 9.5, and 0.05% levamisole/tetramisole in H_2O
Bleaching solution	6% H_2O_2/PBST
DEPC-H_2O	0.1% DEPC in ddH_2O
H1 hybridisation buffer	L1 with 0.1% tRNA and 0.05% heparin

H2 hybridisation buffer	hybridisation buffer 1 with 0.1% tRNA and 0.05% heparin and Dig- probe
L1 buffer	50% deionised formamide, 5x SSC, 1% SDS, 0.1% Tween 20 in DEPC; pH 4.5
L2 buffer	50% deionised formamide, 2x SSC pH 4.5, 0.1% Tween 20 in DEPC; pH 4.5
L3 buffer	2x SSC pH 4.5, 0.1% Tween 20 in DEPC; pH 4.5
Proteinase K Buffer	20 mM Tris pH 7.0, 1 mM EDTA, in DEPC-H2O
TBST 1	140mM NaCl, 2.7mM KCl, 25mM Tris-HCl, 1% Tween 20; pH 7.5
TBST 2	TBST with 0.1% Tween 20, and 0.05% levamisole/tetramisole
Blocking solution	TBST 1 with 2% calf-serum and 0.2% BSA
PBST	0.1 % TWEEN-20 in 1x PBS(DEPC)
RIPA buffer	Use DEPC treated reagents, 0.01 % SDS, 0.15 M NaCl, 0.01 % Nonidet-P40, 5 mg/ml deoxycholate, 1 mM EDTA pH 8.0, 50 mM Tris pH 8.0, in DEPC- H2O
Rnase solution	0.1 M NaCl, 0.01 M Tris pH 7.5, 0.2% Tween 20, 100 µg/ml RNase A in H_2O

Table 3-3 Buffers and solutions used for cHiC and ChIP-seq.

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

3.4 Antibodies and Enzymes

All antibodies and enzymes used in this study are listed in table 3-4.

Table 3-4 Antibodies and enzymes.

Enzyme	Task	Supplier
Restriction Enzymes: SalI, NotI, AatII, PvuI, BbsI, DpnI	Cloning	Applied Biosystems
Restriction Enzyme: DpnII	cHiC-Digest	New England Biolabs
Fusion Polymerase	PCR amplification	
T4 Ligase	Ligation	Applied Biosytems
Proteinase K	Protein degradation	Sigma-Aldrich
RNase A	RNA digestion	Sigma-Aldrich
DNaseI	DNA digestion	Qiagen
Antibodies		
Anti-CTCF	CTCF ChIP-seq	Active Motif (#613111)
Sheep Anti-Digoxigenin Fab fragments Antibody, AP Conjugate	WISH	Roche (#11093274910)
Anti-H3K27ac	H3K27ac ChIP-seq	Diagenode (#C15410174)

3.5 Kits

All molecular biology kits used in this study are listed in table 3-5 below.

Table 3-5 Molecular biology kits used in this study

Kit	Task	Supplier
BigDye Terminator v3.1 Sequencing Kit	Sanger Sequencing	Applied Biosystems
NucleoSpin Plasmid	Plasmid DNA purification	Macherey&Nagel
Nucleobond PC100 EF	Plasmid DNA purification	Macherey&Nagel
PCR&Gel Clean up		Qiagen
QIAquick PCR purification kiT	PCR clean up	Qiagen
RNeasy Kit	RNA purification	Qiagen
SYBR green master mix	qRT-PCR analysis	
NEBuilder® HiFi DNA Assembly Master Mix	Gibson Assembly/Cloning	New England Biolabs

3.6 Plasmids and bacterial strains

All plasmids used are listed in table 3.6.. Vector maps can be found in the appendix (Fig 10-1, 10-2, 10-3 and 10-4). For transformation and propagation of plasmids the chemically competent Escherichia coli (e.coli) TOP10 strain was used (Laboratory internal preparation of Asita C. Stiege) .

Table 3-6 Plasmids used in this study.

Plasmid	Task	Supplier
pTA-GFP	Cloning of WISH-probes	Asita C. Stiege
pKan	Cloning of targeting constructs	Daniel M. Ibrahim
px459 (pSpCas9(BB)-2A-Puro)	Cloning of sgRNAs	Addgene

3.7 Synthetic DNA

All fragments have been ordered from genewiz (genewiz.com, FragmentGENE) with compatible overhangs for Gibson cloning into the pKan backbone. According sequences are listed in the Appendix (10.3) (ExtraTAD1 telHR, ExtraTAD2, IntraTAD1).

3.8 Primers

All primers have been synthesized from IDT Eurofins MWG. Sequences are displayed in 5' to 3' orientation and can be found in tables 10-1 to 10-5 in the appendix.

3.9 Mouse Lines

Table 3-7 CRISPR/Cas9 generated mouse lines

Name of transgenic mouse line	Alias	Genomic location (mm9)
IntraTAD1	Shh mZRS KI	chr5:28,797,519-28,797,539

IntraTAD2	GD mZRS KI	chr5:29,013,316-29,013,335
IntraTAD3	SBE2 mZRS KI	chr5:29,201,903-29,201,922
ExtraTAD1	En2 mZRS KI	chr5:28,517,130-28,517,149
*ExtraTAD2**	Nom1/Mnx1 mZRS KI	chr5:29,747,481-29,747,500
ΔBor	-	chr11:111,383,849-111,402,212
ΔBor*C1*	-	chr11:111,514,370-111,514,389
ΔBor*C1-2*	-	chr11:111,514,370-111,514,389 chr11:111,765,593-111,765,612
ΔBor*C1-4*	DeltaBor C1-C4	chr11:111,514,370-111,514,389 chr11:111,765,593-111,765,612 chr11:111,964,757-111,964,776 chr11:112,070,905-112,070,926
Δ*CTCF*	DeltaBor C1-C4+K_C	chr11:111,307,279-111,307,298 chr11:111,514,370-111,514,389 chr11:111,765,593-111,765,612 chr11:111,964,757-111,964,776 chr11:112,070,905-112,070,926
InvC	Sox9 Inv91-B1	chr11:112514677-112514719 chr11:111383920-111383995
Inv-Intra	Sox9 Inv91-B3	chr11:112514719-112514758 chr11:112514773-112514821
*Bor-KockIn**	KcBor KI	chr11:112514690-112514710

*Only tetraploid aggregations, no line establishment

4. METHODS

4.1 Molecular biological methods

Molecular biological methods such as Polymerase Chain Reaction (PCR), DNA cloning, transformation of competent E. coli and agarose gel electrophoresis were performed according to standard procedures from Green and Sambrook (Green and Sambrook, 2012).

4.1.1 DNA Isolation

Isolation of Plasmid DNA

Plasmid DNA was isolated according to the manufacturer's instruction using the kit Nucleospin plasmid (Macherey&Nagel). For transfection of mouse embryonic stem cells plasmid DNA was isolated with the Nucleobond PC100 EF (Macherey&Nagel).

Isolation of genomic DNA

Genomic DNA of mESC or embryonic tissue was extracted by conventional DNA precipitation method. Samples were incubated in 500 µl lysis buffer (17 mM Tris, pH 7.5; 17 mM EDTA; 170 mM NaCl, 0.85 % SDS) supplemented with proteinase K (final concentration: 0.08 mg/ml) for at least 3h at 55°C for tissue disruption. 250 µl 5M NaCl were added, followed by incubation for 20 minutes at 55°C while shaking and, subsequently, incubation on ice for 10 minutes. Afterwards, samples were centrifuged at 9000 rpm for 20 minutes at 4°C. 500 µl supernatant was mixed with 0.6x volumes of Isopropanol and centrifuged for 30 minutes at maximum speed (>13000 rpm) at 4°C. Precipitated DNA-pellets were washed with 70% ethanol, dried and dissolved in 50-150 µl water.

Biopsies of postnatal mouse ear biopsies were incubated with 50µl of QuickExtract™ (Biozym) for 20 minutes at 65°C, followed by inactivation for 2 minutes at 98°C. 2µl of DNA were used for standard genotyping PCRs.

4.1.2 Cloning of single guide RNAs for CRISPR/Cas9

For design of single guide RNAs (sgRNAs) the CRISPR Design Tool by Feng Zhang lab (http://www.genome-engineering.org/crispr/) or the guide RNA selection tool CRISPOR (http://crispor.tefor.net/; (Haeussler et al., 2016) were utilized. A forward and a reverse complement reverse oligonucleotide were ordered with IDT, containing the 20bp-recognition sequence and a BbsI-specific overhang (Forward Oligo: 5'-caccgN^{20}; Reverse Oligo: 5'-aaacN^{20}). Complementary oligos were annealed and cloned into the pX459 vector (pSpCas9(BB)-2A-Puro, Addgene) containing the chimeric tracrRNA, Cas9 and the puromycin and ampicillin resistance

gene. Annealed Oligos were ligated into the BbsI-digested vector using T4-Ligase and transformed into chemically competent Top10 bacteria (produced laboratory internal by Asita Carola Stiege). Plasmid DNA was purified and successful integration was confirmed with Sanger Sequencing.

4.1.3 Cloning of targeting constructs for insertions (Boundary, Enhancer, lacZ sensor)

Targeting constructs were designed with asymmetric homology regions (HR), ranging between 500-700 bp and 1000-2000 kb (Byrne et al., 2015). PAM sites were either mutated or disrupted by the fragment which should be inserted. The size of fragments for insertions varied between 1-6 kb. All targeting constructs were designed on the mm9 genome assembly using the UCSC genome browser (http://genome-euro.ucsc.edu). Fragments with a maximum size of 3 kb were synthesized with GENEWIZ® (FragmenGENEs, dsDNA fragments) already containing 40 bp complementary overhangs for Gibson Cloning (Gibson et al., 2008) into the pKan plasmid (produced laboratory internal by Daniel M. Ibrahim) containing a Kanamycin resistance cassette gene or in the TelCD40 plasmid (produced laboratory internal by Martin Franke) containing an Ampicillin resistance gene using the NEBuilder® HiFi DNA Assembly Master Mix (NEB, #E2621). For cloning of lacZ Sensor constructs the β-Globin minimal promoter was fused to a lacZ reporter gene (Ruf et al., 2011). An overview containing cloning primers and all sequences of targeting constructs can be found in the appendix.

4.2 Cell Culture

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

4.2.1 Preparation and culturing of feeder cells (mouse embryonic fibroblasts)

E13.5 and E14.5 CD1 and DR4 (puromycin/hygromycin/geneticin resistant) mouse fetuses were used for preperation of feeder cells. Mouse Embryonic Fibroblasts (MEFs) were cultured in regular Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% regular fetal calf serum (FCS Superior, Biochrom, #S0615), 1x Glutamine (100x Lonza, #BE17-605E) and 1x penicillin/streptomycin (Lonza, #DE17-603) and expanded until passage 5. MEFs were inactivated with Mitomycin C (Sigma, #M-0503)-supplemented Medium for 2 hours, then washed with 3x PBS and frozen in freezing medium containing 10% DMSO (Sigma, #D-2650) with a densitiy of $2.5x10^6$ cells/cryovial.

4.2.2 Culturing of mouse embryonic stem cells (mESCs)

The mouse embryonic stem cell (mESC)-strain G4 with a hybrid genetic background of 129Sv x C57BL/6 was used in all experiments (Georg et al 2007??). Generally, G4-cells were seeded on top of a CD1 feeder-layer, previously plated (at least 6h before) on gelatinized dishes or wells (15 minutes, 37°C, 7.5% CO2, 0.01% gelatin, Sigma, #G-1393) in a density of 3.5 x 10^4 cells/cm^2. mESCs were grown in Knockout DMEM 4,500 mg/ml glucose with sodium pyruvate (Gibco, #10829-018) containing 15% FCS (PANSera ES, #P30-2600), 10mM Glutamine (100x Lonza, #BE17-605E) and 1x penicillin/streptomycin (Lonza, #DE17-603), 1x non-essential amino acids (100x Gibco, #11140-35), 1x nucleosides (100x, Chemicon, ,ES-008D), 0.1 mM beta-Mercaptoethanol (Gibco, #3150-010) and 1000 U/ml LIF (Murine Leukemia Inhibitory Factor ESGRP™, 10^7 U/ml, Chemicon, #ESG1107) at 37°C, 7.5% CO2. Medium was changed every 24 hours, split every 2-3 days and cells were frozen in a density of 1x 10^6 cells/cryovial in ESC medium containing 20% FCS and 20% DMSO (Sigma, #D-2650).

4.2.3 Transfection and CRISPR/Cas9-based Genome Editing of mESCs

Genome Editing of G4 mESCs was performed as follows: CD1 feeders were plated on 6cm-dishes (day 1). The day after, 4 x 10^6 G4-mESCs were seeded on top oft the feeder cells and medium was exchanged with ESC medium without penicillin/streptamycin 2 hours before transfection the next day. pX459-sgRNA vector(s) were either mixed solely or in combination with a targeting vector (KnockIns) in a total volume of 125 µl OptiMEM (Gibco, #51985-026) as a DNA-Mix. An additional 100 µl OptiMEM was supplemented with 25 µl of FuGENE HD agent (Promega, #E2311) in a Transfection Mix. Both mixes were combined, vortexed vigorously and incubated at RT for 15 minutes before being added dropwise onto the mESCs (day 2-3). To create structural variants two different pX459-sgRNAs were transfected (4 ug each). For targeting CTCF-binding sites via creation of Indel-mutations only one pX459-sgRNA was used (8 ug). KnockIns were performed with a combination of 8 ug pX459-sgRNA and 4 ug of targeting construct. After 16-18 hours of transfection medium was changed to regular ESC medium and puromycin-resistant DR4 feeders were plated on 3x 6cm-dishes (day 4). 48 hours after transfection G4-mESCs were split on 3x 6cm DR4 feeder cells and 48 hour-selection was initiated by supplementing ESC medium with puromycin (final concentration: 2 ug/ml, Sigma-Aldrich, #P8833) (day 5-7). Selection was abrogated by medium change to regular ESC medium and cells were cultured for recovery until ESC clones were visible and large enough for picking (day 8-11). Single mESC-colonies (16-30 cells) were picked in PBS with a pipette tip and transferred into U-bottom 96well-plates containing 0.2% Trypsin-EDTA (Gibco, #25300-054) and incubated 10 minutes at 37°C. After resuspension, single cell solutions were transferred into 96well-plates containing CD1 feeder cells

and cultured for additional 3 days. Plates were then split into triplicates. Two replicates were frozen using U-bottom 96-well plates in bicarbonate-free DMEM (Gibco, #52100) containing 10 mM HEPES pH 7.2, 20% FCS (PANSera ES, #P30-2600) and 10% DMSO (Sigma, #D-2650). One plate was left in culture for growth and harvesting of genomic DNA for genotyping (see 4.2.6). An overview of generated mutant mESC lines in this study is given in table 3-7.

4.2.4 Screening and validation of targeted mESCs

mESCs were lysed in lysis buffer (17 mM Tris, pH 7.5; 17 mM EDTA; 170 mM NaCl, 0.85 % SDS) supplemented with proteinase K (final concentration: 0.08 mg/ml) overnight at 55°C and extracted with DNA magnetic beads (MagAttract®, Qiagen) using the Thermo Fisher KingFisher™ Flex robot. To 8 µl of beads 50 µl of cell lysate and 32 µl of 85% ethanol were added. The ESC_gDNA_85_GADMI-protocol in the BindIt 3.1 Software was conducted by the extraction robot ((i)Binding of Beads (ii) 2x Washing with 85% EtOH, final concentration: 70% (iii) Drying of Beads (iv) Elution in H_2O).

Standard PCR procedures utilizing Taq-polymerase produced laboratory internal by Asita Carola Stiege were used for genotyping of the picked mESC clones. Generally, reagents were pipetted on ice in a 96well-plate and PCR products were amplified in a thermocycler and analyzed on 1-2% agarose gels. PCR conditions for a 25 µl reaction: 2.5 µl 10x Taq buffer (750 mM TRIS/HCl pH 8.0; 200 mM $(NH_4)_2SO_4$; 0.1% Tween 20; 15 mM $MgCl_2$), 0.1 µl dNTPs (12.5 mM), 0.1 µl forward primer (100 µM), 0.1 µl reverse primer (100 µM), 0.5 µl Taq enzyme, 2.5 µl template (circa 30 ng gDNA) and 19.2 µl bidest H_2O. The following program was used: step 1: 95°C, 5 minutes; step 2: 95°C, 30 seconds; step 3: 60°C, 30 seconds; step 4: 72°C, 45 seconds; step 5: go to step 2 for 34x; step 6: 72°C, 5 minutes; step 7: 4°C, pause.

A list of Primer sequences for conventional genotyping used in this study is given in table xy in the appendix.

4.2.5 Copy Number Analysis of mutant mESC clones via qRT-PCR

Copy Number Analysis from genomic DNA (gDNA) was conducted using the SYBR Green Mastermix (Applied Biosystems) on an ABIPrism Quant7 thermocycler. The Primer3Plus online tool was used for Primer Design with an average product size of 80-120 bp. The qRT-PCR reaction was set up in 384well-plate as follows: 6 µl SYBR Green Mix, 2 µl PrimerMix (1.5 µM each) and 4 µl gDNA (60 ng). All reactions were performed as triplicates. Standard series for each primer pair was generated from control gDNA containing the target sequence in 1:5 dilution steps (1; 0.2; 0.04; 0.008; 0.0016). For normalization relative cT-ratios of the tested region (primerpair within Insertion, Deletion, Inversion) to a region outside the mutated region were created and compared to wildtype controls.

4.2.6 Generation and Genotyping of mutant mouse lines (Aggregation, Crossings)

Mutant G4 mESCs were seeded in a densitiy of 2 to 4x10^5 cells onto CD1 feeder cells. Mutant embryos or live animals were generated by diploid or tetraploid complementation (Artus and Hadjantonakis, 2010). CD1 female mice were used as foster mothers.

Mouse strains were maintained by crossing them with C57BL6/J mice. All animal procedures were in accordance with institutional, state, and government regulations (Berlin: LAGeSo G0176/19, G0247/13).

Standard PCR procedures utilizing Taq-polymerase produced laboratory internal by Asita Carola Stiege were used for genotyping of mouse mutants. Generally, reagents were pipetted on ice in a 96well-plate and PCR products were amplified in a thermocycler and analyzed on 1-2% agarose gels. PCR conditions for a 25 µl reaction: 2.5 µl 10x Taq buffer (750 mM TRIS/HCl pH 8.0; 200 mM $(NH_4)_2SO_4$; 0.1% Tween 20; 15 mM $MgCl_2$), 0.1 µl dNTPs (12.5 mM), 0.1 µl forward primer (100 µM), 0.1 µl reverse primer (100 µM), 0.5 µl Taq enzyme, 2.5 µl template (circa 30 ng gDNA) and 19.2 µl bidest H_2O. The following program was used: step 1: 95°C, 5 minutes; step 2: 95°C, 30 seconds; step 3: 60°C, 30 seconds; step 4: 72°C, 45 seconds; step 5: go to step 2 for 34x; step 6: 72°C, 5 minutes; step 7: 4°C, pause.

4.3 Chromosome Conformation Capture(3C)-Technology

The Capture Hi-C protocol described below was established by Martin Franke in the laboratory and is adapted from (Werken et al., 2012)

4.3.1 CaptureHiC

Fixation and nuclei extraction

Embryonic tissues were dissected in PBS. After adding prewarmed Trypsin for 10 minutes at 37°C, the reaction was stopped by adding 10% FCS/PBS and single cell suspensions were homogenized using a 0.40 µm cell strainer. After centrifugation at 1100 rpm for 5 minutes, pellets were resuspended in 5 ml 10%FCS/PBS and subsequently fixed by adding 4% Formaldehyd (FA) in 10%FCS/PBS for 10 minutes at RT while tumbling (final concentration: 2% FA). Crosslinking was quenched by adding 1 ml of 1.425 M glycine on ice and immediate centrifugation for 8 minutes at 1500 rpm at 4°C. The cells were lysed and nuclei extracted by resuspension in 5 ml freshly prepared lysis buffer (see table 3-3) on ice for at least 10 minutes. For confirmation of lysis 3 µl of nuclei suspension were mixed with 3 µl of methyl green pyronin staining solution (Waldeck, Pappenheim, #2C-186). The number of nuclei was determined using a standard counter chamber (Neubauer) and cells were pelleted by centrifugation at 2000 rpm for 5 minutes at 4°C. After

washing once with PBS, cells were aliquoted (2-5 x10^6 nuclei) and after another centrifugation step at 2600 rpm for 2 minutes at 4°C snap-frozen in liquid nitrogen and stored at -80°C.

Capture HiC sample and library preparation

Nuclei pellets were resuspended in 360 µl H_2O bidest and supplemented with 60 µl DpnII 10x restriction buffer (NEB). Samples were incubated at 37°C in a thermomixer at 900 rpm. After adding 15 µl of 10% SDS for an hour, the remaining SDS was separated from the reaction by adding 150 µl of 10% Triton X-100 for 1 hour. A 15 µl aliquot was taken as a control for undigested chromatin. A restriction digest was started by adding 600 µl of 1x restriction buffer and 400 Units of DpnII, followed by adding additional 200 Units after 4 hours and overnight restriction at 37°C. The next day, another 200 Units of DpnII were added and incubated for 4 hours. After taking a digestion control aliquot (30 µl), the digest was stopped by heat inactivation of the enzyme for 20 minutes at 64°C and transferred to a 50 ml Falcon tube. The volume was adjusted to 7 ml with H_2O, supplemented with 1x Ligation buffer and 50 units of T4 DNA Ligase (Thermo Fisher Scientific, #EL0013), followed by incubation for 4 hours at 16°C and 30 minutes at RT. A 100 µl ligation control aliquot was taken before de-crosslinking by adding 30 µl of proteinase K (10 mg/ml) and overnight incubation at 65°C.

The next day, 30 µl of RNaseA (10 mg/ml) was added to the sample and incubated for 30 minutes at 37°C. Then, DNA was extracted by adding 7 ml of phenol-chloroform and mixed vigorously by inverting the tube. The water phase was separated via centrifugation at 3750 rpm for 15 minutes at RT and transferred to a fresh falcon tube. To precipitate the DNA, 7 ml H_2O, 1 ml 3M NaAc pH 5.3, 140 µg glycogen and 35 ml of 100% cold ethanol were added and the sample incubated at -20°C overnight, followed by centrifugation for 60 minutes at 6000 rpm at 4°C. The sample was washed once with 30 ml 70% ethanol, air dried and dissolved with 150 µl 10 mM TRIS/HCl pH 7.5. Samples were stored at -20°C.

Before proceeding to sequencing the efficiency of the library was tested. Chromatin from control aliquots obtained during the sample preparation were de-crosslinked overnight at 65°C after adding 5 µl proteinase K (10 mg/ml), followed by addition of 2 µl RNaseA (10 mg/ml) and incubation for 30 minutes at 37°C. DNA was extracted by adding 100 µl of phenol-chloroform, mixed and centrifuged for 15 minutes at maximum speed at RT. Supernatant was transferred into fresh tubes and 20 µl of the cleaned up DNA was loaded on a 1% agarose gel.

4.3.2 SureSelect Design

SureSelect enrichment probes were designed using the SureDesign online tool from Agilent (https://earray.chem.agilent.com/suredesign/). At the *Sox9*-locus, probes covered the genomic

interval chr11:109,010,000-114,878,000 (mm9, 3 Mb). At the *Shh*-locus, probes covered the genomic interval chr5:27,800,001-30,600,000 (mm9, 2.8 Mb).
Library samples were measured using Qbit, sheared using a Covaris sonicator (Duty cycle: 10%, Intensity: 5, Cycles per Burst: 200, 6 cycles of 60 seconds, Set Mode: Frequency Sweeping) and adaptors were ligated for amplification, following the Agilent instructions for Illumina Sequencing. Next, libraries were enriched with the according custom-designed SureSelect enrichment probes. After indexing, samples were sequenced on the NextSeq500 or HiSeq2500 (Illumina) with a depth of 200 million readpairs (100 bp paired-end).

4.3.3 Capture HiC Data Analysis

Bioinformatic analysis was performed in cooperation with Robert Schöpflin (Department Computational Molecular Biology, Max Planck Institute for Molecular Genetics, Berlin). Processing, mapping and filtering of mapped paired-end sequencing data was performed using the HiCUP pipeline v0.6.1 (Wingett et al., 2015). For mapping short reads to the reference genome (NCBI37/mm9) the pipeline used Bowtie2 v.2.2.6 .(Langmead and Salzberg, 2012) Further, the Juicer tool was used to process filtered di-tags to bin di-tags (5 and 10 kb bins) and for Knight-Ruiz (KR) matrix balancing normalization (Durand et al., 2016; Knight and Ruiz, 2012; Lieberman-Aiden et al., 2009). Only reads with a MAPQ ≥30 were considered. The DNA-capturing step enriches the genomic interval chr11:109,010,000-114,878,000 (mm9, 3 Mb) for *Sox9*. The *Shh*-locus probes covered the genomic interval chr5:27,800,001-30,600,000 (mm9, 2.8 Mb). This leads to three different regimes in the cHiC-map: (a) enriched versus enriched, (b) enriched versus non-enriched, and (c) non-enriched versus non-enriched. Only di-tags in regime (b) were considered for binning and normalization. Di-tags were filtered for the enriched regions and mm9 coordinates were shifted by -109,010,000 bp (*Sox9*) or -27,800,000 bp (*Shh*).

4.3.4 CTCF Motif Analysis with FIMO (Meme Suite)

CTCF motif orientation in ChIP-seq peaks (CTCF ChIP-seq from E14.5 limb buds from ENCODE/LICR) was analyzed with the FIMO algorithm of the MEME suite (Bailey et al. 2009) with default parameters. The position weight matrix (PWM) from (Jerković et al., 2016) was used as a CTCF motif matrix. Genomic regions underlying a CTCF peak of 100-200 bp were analyzed.

4.4 Chromatin Immunoprecipitaion (ChIP)-Sequencing

After microdissection of E10.5 or E12.5 embryonic limb buds, the tissue was digested with Trypsin–EDTA 0.05% (Gibco) for 10min at 37°C. The reaction was stopped by mixing with 10% FCS/PBS and subsequently filtered with a 40-µm cell strainer (Falcon). The single cell suspension was fixed in 1% paraformaldehyde (PFA)/10% FCS/PBS for 10 min at room temperature,

tumbling. After stopping of fixation by adding Glycine and subsequent centrifugation, cells were lysed in Lysis buffer (50mM Tris, pH 7.5; 150mM NaCl; 5mM EDTA; 0.5% NP-40; 1.15% Triton X-100; protease inhibitors (Roche)) for 10 min on ice. Nuclei were resuspended in sonication buffer (10mM Tris–HCl, pH 8.0; 100mM NaCl; 1mM EDTA; 0.5mM EGTA; 0,1% Na-deoxycholate; 0.5% N-lauroylsarcosine; protease inhibitors (Roche complete)). Chromatin was sheared using a Bioruptor until reaching a fragment size of 200–500base pairs (bp). After clarification of lysates, protein–DNA complexes were immunoprecipitated overnight at 4°C with the respective antibody. A total of 10–15µg (histone) or 25 µg (transcription factor) chromatin was used for each replicate ChIP. Antibodies: H3K27ac (C15410174; Diagenode) and CTCF (Active Motif: 613111). ChIP-seq was performed according to (Andrey et al., 2016).

4.4.1 ChIP-seq data analysis

Nextera adapters were used for preparation of sequencing libraries. Samples were sequenced on the NextSeq500 or HiSeq2500 (Illumina) 50-100bp single-end (SE) read. Mapping was conducted with bowtie (v2.2.6) (Langmead and Salzberg, 2012) to the mm9 reference genome and filtered for mapping quality MAPQ≥10. Duplicates were removed with samtools rmdup (v1.8). Reads were extended to 300bp (H3K27ac) or 200bp (CTCF) and scaled to r.p.m. (10^6 per number of unique reads) using bedtools genomecov v2.27.1 for generation of coverage tracks.

4.5 RNA expression analysis

For quantification of gene expression levels, E10.5 or E13.5 limb buds from littermates (wt/het/hom) of mutant mice were microdissected in cold PBS and immediately snap frozen in liquid nitrogen. Samples were stored at -80°C.

4.5.1 RNA extraction

To isolate RNA from embryonic tissue samples were homogenized in 350 µl RLT buffer supplemented with β-Mercaptoethanol using a 1 ml syringe and a size 20 cannula (G 27 x 3/4"" / ø 0,40 x 20 mm). After centrifugation at full speed for 3 minutes, supernatant was transferred into a fresh tube and mixed with 350 µl of 70% ethanol. The sample was loaded onto a RNeasy Mini Kit Column (Qiagen, #74104) and proceeded following the manufacturer's instructions including an on column- DNase digest (Qiagen, # 79254).

4.5.2 cDNA synthesis

For cDNA synthesis 300-500 ng of extracted RNA was used, utilizing the Superscript III™ First-Strand Synthesis System (Invitrogen, #18080051) for reverse transcription.

4.5.3 Expression analysis with quantitative RT-qPCR

Quantification of relative transcript abundance was performed for 2-6 biological replicates in technical triplicates using the SYBR green mix (Applied Biosystems). Samples were internally normalized by taking the ratio to housekeeping genes (*Gapdh* or *Rps9*). Mutant expression was compared to wildtype littermates (set as 1).

4.6 Whole Mount In Situ Hybridization (WISH)

Buffers and solutions used for whole mount in situ hybridization were treated with DEPC to inactivate RNase enzymes and are listed in table 3-2.

DIG-labeled Probe generation

mRNA expression of the genes *Shh, Mnx1, En2, Sox9* and *Kcnj2* was assessed by whole mount in situ hybridization (WISH) in E10.5 and E12.5 mouse embryos. Digoxigenin (DIG)-labeled antisense riboprobes (table 10-5) were generated by PCR amplification utilizing E10.5 or E12.5 mouse limb bud cDNA and subsequently cloned into the pTA-GFP vector. For PCR amplification of the template the SP6 and T7 primer were used. Then, 200 ng of PCR template were mixed with DIG RNA labelling mix (Roche, #11277073910), transcription buffer and RNA polymerase SP6 or T7 (Roche, #10999644001) and incubated for 2 hours at 37°C. After DNaseI treatment (Roche, #04716728001) for 15 minutes at 37°C the reaction was stopped by adding 0.2 mM EDTA/H_20-DEPC (pH 8.0), 0.4 M LiCl and 3 volumes chilled 100% ethanol and subsequently precipitated over night at -80 °C. After centrifugation at 13000 rpm for 20 minutes at 4°C and one washing step with 70% ethanol, the RNA pellet/probe was dissolved in H_20-DEPC.

Preparation, Hybridization and Staining of embryos for WISH

Wildtype and Mutant embryos were dissected in PBS-DEPC and fixed overnight in 4% PFA/PBS-DEPC at 4°C. The next day, embryos were washed twice in PBS supplemented with 0.1% Tween (PBST) for 30 minutes and subsequently dehydrated in a methanol/PBST series (25%, 50%, 75%, 10 minutes each at 4 °C) and stored in 100% methanol at -20 °C.

For WISH embryos were rehydrated at 4°C in a reverse methanol/PBST series (75%, 50%, 25%) and washed in PBST. Bleaching of embryos in 6% H2O2/PBST for 1 hour was followed by washing with PBST and Proteinase K treatment (10 ug/ml; E10.5: 3 minutes; E12.5: 5 minutes). After additional washing with PBST/glycine, PBST and RIPA buffer, embryos were fixed for 20 minutes in 4% PFA/0.2% glutaraldehyd/PBST. Embryos were washed in PBST, PBST/L1 buffer (1:1) and incubated at 68°C for 10 minutes in L1 buffer. Prior to hybridization DIG-labeled RNA probes were diluted 1:100 in H1 buffer. Denaturation of probes at 80°C for 5 minutes was performed before being added to the embryos for hybridization over night at 68°C. The next day,

unbound probes were removed by washing three times with L1 buffer at 68°C for 30 minutes followed by washing three times with L2 buffer (30 minutes, 68°C) and L3 buffer (15 minutes, 68°C). After cooling down to RT, embryos were treated twice with RNase solution containing RNaseA (100 ug/ml) for 30 minutes at 37°C and subsequently washed three times with TBST 1 (1% Tween20) for 5 minutes at RT. Then, embryos were incubated in blocking solution for 2 hours at RT while incubating the DIG-antibodies 1:5000 in blocking solution on a rotating wheel at 4°C. Embryos and DIG-labeled probes were incubated overnight on a shaker at 4°C. The next day, unbound antibodies were removed via a series of washing steps (5x 5 minutes TBST2, 8x 30 minutes TBST2 at RT) and left in TBST2 solution over night at 4°C with shaking.

After washing three times with Alkaline phosphatase buffer for 20 minutes at RT, embryos were stained with BM Purple AP Substrate (Roche, #1442074) for at least 1hour at RT. When staining was completed, embryos were washed with PBST, fixed with 4%PFA/PBS/0.2% glutaraldehyd/5mM EDTA and stored at 4°C.

Mutant embryo expression patterns were compared to wildtype littermates (if possible). The stained embryos were imaged with a Zeiss Discovery V.12 microscope and a Leica DFC420 digital camera.

4.7 LacZ Sensor Staining

LacZ reporter stainings were conducted according to the protocol of LacZ reporter stainings from Lobe et al. 1999. Embryos were dissected in PBS and fixed in 4% PFA/PBS at 4°C for 20-30 minutes. After washing three times with LacZ washing buffer (2mM MgCl2, 0.01% Na-deoxycholat, 0.02% Nonidet-40 in PBS) the embryos were incubated in LacZ washing buffer supplemented with X-Gal DMSO (0.5 mg/ml), 5mM potassium ferrocyanide, 5mM potassium ferricynide at 37°C for a few hours or overnight. After the desired staining was obtained, embryos were washed with LacZ Washing Buffer and stored in 4%PFA/PBS at 4°C.

4.8 Skeletal Preparations

E18.5 embryos or P0 animals were decapitated and kept in water over night at RT. Next, animals were heat shocked at 70°C for 40 seconds and skinned, disemboweled and incubated for at least 3 days in 100% ethanol. After washing them with acetone for 1 hour, cartilage was stained with alcian blue staining solution (150 mg/L Alcian Blue 8GX, Sigma-Aldrich, #A5268, 80% ethanol, 20% acetic acid) for 24-48 hours at RT. After overnight fixation in 100% ethanol, animals were stepwise rehydrated with a ethanol series (75%, 50%, 25%, H2O at least 8 hours each). Next, the animals were initially cleared via incubation in 1% KOH/bidest H2O for 30-60 minutes. Cartilage was stained using Alizarin Red (50 mg/l, Sigma-Aldrich, #A5533) in 0.2% KOH/bidest

H2O for up to 48 hours and stopped after visual inspection. Remaining tissue was digested with KOH/bidest H2O concentrations ranging from 0.1%-1% and subsequently stopped by adding increasing glycerin solutions (30%, 60%, 80% glycerin/bidest H_2O, 24 hours each). Preparations were stored in 80% glycerin at RT.

4.9 Micro-computer tomography (microCT)

Micro-computer tomography (microCT) analysis was performed in cooperation with Wing Lee Chan (Charité-Universitätsmedizin Berlin, Institut für Medizinische Genetik und Humangenetik, Berlin). For microCT of autopods of adult mice the a SkyScan 1172 X-ray microtomography system (Brucker microCT, Belgium) was used at 5-µm resolution. Reconstruction of 3D models was conducted with the SkyScan image analysis software CT-Analyser and CT-volume (Brucker microCT, Belgium).

5. RESULTS

5.1 The *Kcnj2/Sox9*-locus as a model system

Recent genome-wide depletion experiments of CTCF or subunits of the cohesin complex demonstrated that the immediate effect on gene expression upon the resulting loss of TADs is relatively mild (Nora et al., 2017; Rao et al., 2017; Schwarzer et al., 2017). However, these results contradict the described potential pathogenic effect of structural variants (SV) on gene regulation, which is caused by an altered chromatin architecture (Spielmann et al., 2018; Weischenfeldt et al., 2017). Here, we wanted to dissect the role of CTCF and TAD architecture for gene regulation in a developmental *in vivo* setting in mice and correlate changes in chromatin architecture with gene expression, expression patterns, and potential mouse phenotypes (Fig. 5-1). We did so by applying CRISPR/Cas9 to mouse embryonic stem cells (mESCs). We generated a series of mutant mESC lines from which we derived transgenic mouse lines. Then, we used embryos from these lines to analyze gene expression and 3D chromatin structure in the developing mouse limb.

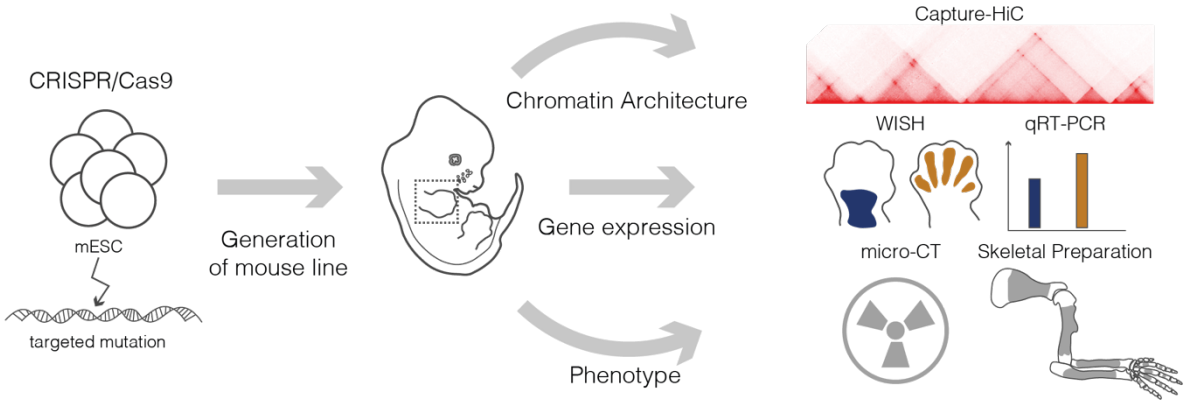

Figure 5-1 Experimental set-up and strategy. CRISPR/Cas9 genome editing of mESC was followed by generation of mouse lines via diploid or tetraploid complementation. Chromatin architecture of mutant E12.5 limb buds by cHiC was correlated with changes in gene expression assessed by qPCR and WISH and potential phenotypes with micro-CT and skeletal preparations.

The *Sox9/Kcnj2*-locus consists of two well-defined TADs harboring the regulatory information required for the accurate and precisely timed expression of their target genes. The *Sox9* gene is located on the telomeric side of its TAD of almost 2 Mb in size, situated in a vast gene desert. In this stretch of DNA, many described and potential CREs in the form of enhancers are embedded, regulating and driving the complex spatiotemporal *Sox9* expression pattern (Fig. 5-2 A). In this study, we were particularly interested in the regulation of *Sox9* in the developing cartilage anlagen of the limb bud at embryonic stage E12.5 (Fig. 5-2 B). The neighboring genes *Kcnj2* and *Kcnj16* are potassium channels located in the adjacent centromeric TAD, whereas solely the first one is weakly expressed at E12.5 in the distal zeugopod (Fig. 5-2 B). Regulatory sensors (Ruf et al., 2011) integrated in various positions in either of the two TADs are capable of recapitulating the

regulatory activity of the two genes vastly, although not completely (Franke, 2017). The *Kcnj2*-TAD is defined by overlapping regulatory patterns in the maxilla, forebrain and proximal limb, varying in strength at the different positions. At some integration sites in the *Sox9*-TAD, the sensor strongly recapitulates the limb signal, while at other positions it is barely detectable. In addition to the limb, signal is visible in the scapula, nasal septum and capsule of the E12.5 embryos. However, the signal is limited to the corresponding domain, restricted by the TAD boundary (Fig. 5-2 B).

Ectopic interactions between the two domains are restricted by a separating TAD boundary, as demonstrated by 4C-experiments (Franke et al., 2016). 3D-chromatin contacts within both domains are shaped by CTCF sites located at the TAD boundary (Bor) but also within the TAD. In particular, the *Sox9*-TAD harbors four major binding sites (referred to as C1-C4) and an additional site at the promoter, forming an internal domain-substructure (Fig. 5-2 B). Equally, within the *Kcnj2*-TAD several CTCF sites are located near the promoter as well as between the *Kcnj2* gene and the TAD boundary (C_{Kc}). It is known that *Kcnj2* misexpression in a *Sox9*-like pattern results in Cooks syndrome in adult animals, as described previously (Franke et al., 2016). As shown in the interaction frequency maps from cHiC and CTCF ChIP-Seq from mouse limb buds at embryonic stage E12.5, the boundary separating the two domains consists of two pairs of CTCF binding sites in divergent orientation (Fig 5-2 B, magnification). Within the *Sox9*-TAD, the substructure with several loops is linked to at least four additional CTCF-binding sites, described previously. Deleting the boundary separating these two regulatory landscapes, however, does not result in gene mis-regulation (Franke et al., 2016).

In this study we first wanted to experimentally assess the effect of chromatin structure on gene regulation by gradual deletion of CTCF binding sites. Secondly, we wanted to test the effect of structural variants on gene regulation, and, on the basis of SVs, investigate the function and influence of TAD boundary and substructure separately.

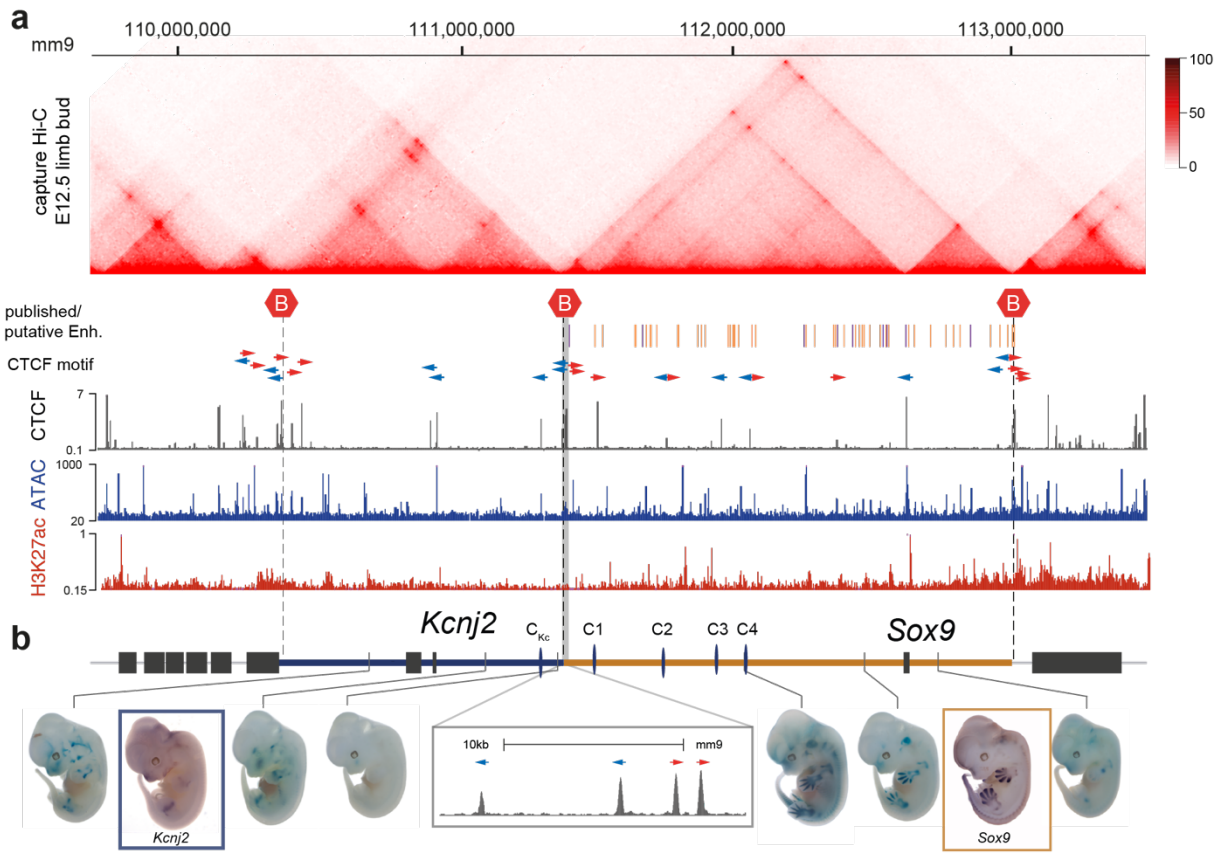

Figure 5-2: The *Sox9/Kcnj2*-locus consists of two TADs. (A) cHiC from E12.5 mouse limb buds (Franke et al. 2016) demonstrates the subdivision of the locus into the centromeric *Kcnj2*- and telomeric *Sox9*-TAD separated by a TAD boundary. Published and putative regulatory regions of the *Sox9*-TAD are depicted with purple and orange bars, respectively. Below are corresponding ChIP-seq tracks for CTCF and H3K27Ac and ATAC-seq (data from Andrey et al. 2016). (B) Mapping of the regulatory information of the locus with lacZ-sensors recapitulating regulatory activity of a specific location. Expression patterns are shown by WISH for comparison (Blue Box: *Kcnj2*; Orange Box: *Sox9*). Grey box: Magnification of the CTCF ChIP-seq track of the TAD boundary region, depicting the convergent orientation in 15kb window. Grey bars: genes. Arrow heads: CTCF site orientation. Red hexagons: TAD boundary. Major CTCF sites labeled as C1, C2, C3, C4 and C_{Kc} *(Despang et al., 2019)*

5.1.1 Deciphering the role of CTCF in shaping elaborate chromatin domains exemplarily at the *Sox9/Kcnj2*-locus

To elucidate the role of CTCF in shaping 3D chromatin domains we utilized a CRISPR/Cas9-based series of CTCF-site deletions in addition to the boundary deletion. To analyze the chromatin structure of the locus we conducted Capture HiC from E12.5 limb buds to map all chromatin interactions over a 5.9 Mb region at the extended *Sox9/Kcnj2*-locus (chr11:109,010,000-114,878,000, mm9).

As mentioned before, deletion of the TAD boundary caused only a moderate increase in inter-TAD contacts (Data from (Franke et al., 2016). In order to test if and how the intra-TAD CTCF

sites contribute to the formation of TADs, we subsequently targeted the intra-TAD CTCF binding sites by a stepwise generation of homozygous Indel-mutations within the consensus CTCF binding site. In addition to the TAD boundary, we deleted the C1-CTCF binding site, and then sequentially the C2, C3-C4 and finally, the C_{Kc}-CTCF binding site located between the *Kcnj2*-promoter and the TAD boundary (*ΔCTCF*) (Fig. 5-3).

5.1.2 CTCF-deletion series results in gradual TAD-fusion of neighboring domains

cHiC of wildtype E12.5 mouse embryonic limb buds distinctly shows the separation at the locus into two TADs. Virtual 4C from either, the *Kcnj2* or the *Sox9* promoter, clearly demonstrates the interactions stopping at the TAD boundary (Fig. 5-3 A). Deletion of the boundary *(ΔBor*) only results in a slight increase of inter-TAD contacts measured by interaction score (Fig. 5-4 A), almost indistinguishable from wildtype. Using the boundary deletion mutant as a basis, the additional removal of the first C1-CTCF binding site *(ΔBorC1*) in the *Sox9*-substructure led to a marked increase in inter-TAD contacts between the neighboring *Kcnj2*- and *Sox9*-TAD, as shown by cHiC. An additional increase of contacts could be observed by the sequential deletion of the C2-binding site *(ΔBorC1-C2*), successively expanding after deletion of all four binding sites (*ΔBorC1-C4*) (Fig. 5-3 C). This was accompanied by a stepwise increase in inter-TAD interactions as measured by the interaction score, concomitant with decreased fractions of intra-TAD contacts of the *Sox9*- and *Kcnj2*-domain (Fig. 5-4 A). Remarkably, deletion of the C_{Kc}-CTCF binding site *(ΔCTCF*) and thus eliminating all bound CTCF sites separating the *Kcnj2*- and *Sox9* promoter, led to complete TAD fusion (Fig. 5-3 C).

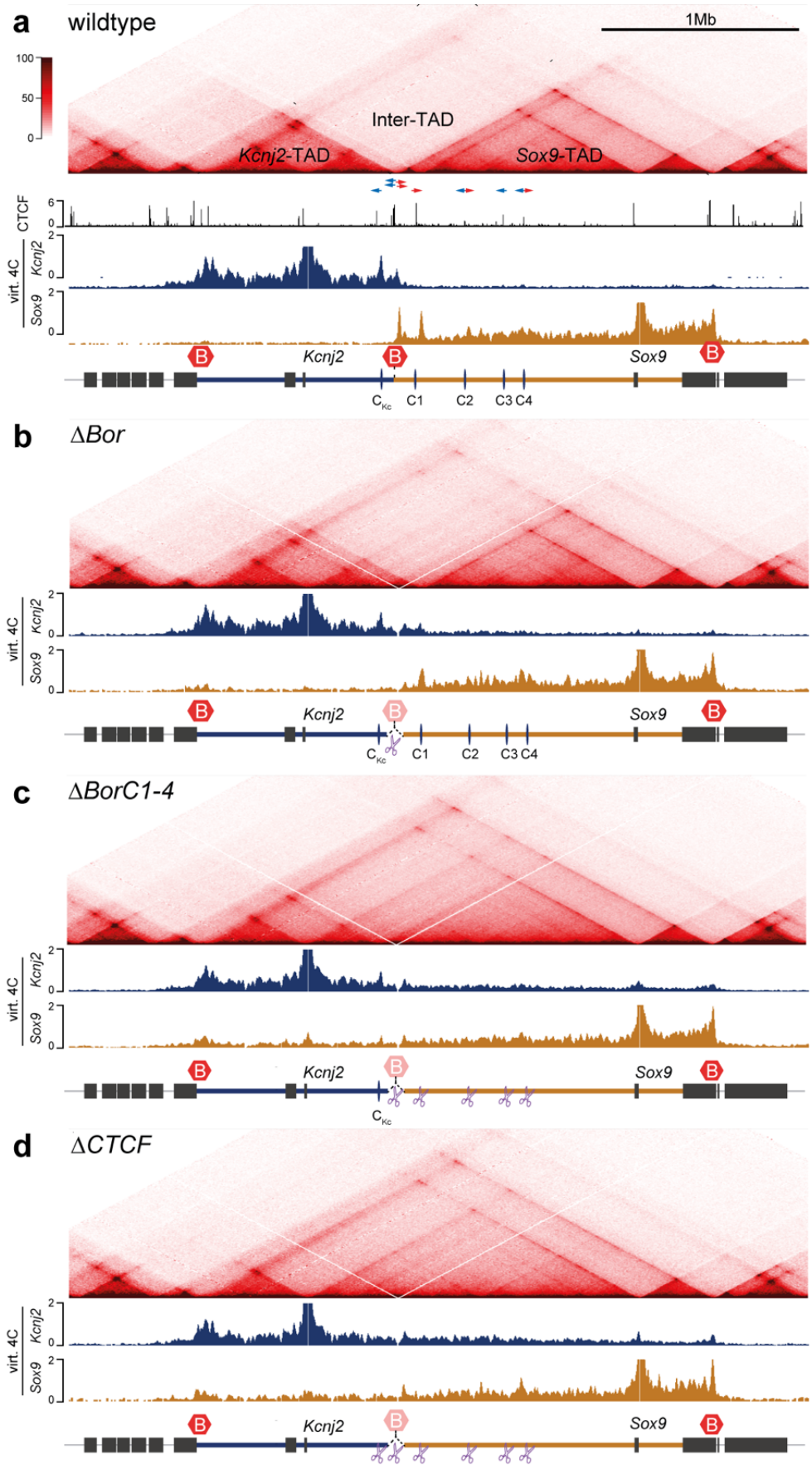

Figure 5-3: Gradual TAD fusion upon CTCF-deletion. (A) cHiC from E12.5 wildtype limb buds with corresponding CTCF ChIP-seq track and virtual 4C from either the *Kcnj2* (blue) or the *Sox9* promoter (orange) as a viewpoint.

Arrowheads indicate CTCF-site orientation. (B)*ΔBor*: Deletion of the boundary separating the two domains leaves the overall TAD configuration at the locus largely unchanged (C) *ΔBorC1-C4*: Additional removal on top of the boundary deletion of the C1, C2, C3 and C4 CTCF binding sites within the *Sox9*-TAD causes fusion of the *Kcnj2*- and *Sox9*-domain. (D) *ΔCTCF*: Deletion of all major CTCF sites separating the *Kcnj2* and *Sox9* promoter (C1, C2, C3, C4 and C_{Kc}) leads to further fusion of the two TADs.

In order to assess possible changes in gene expression upon gradual TAD fusion, *Sox9* and *Kcnj2* transcript abundance was measured by qRT-PCR from E13.5 limb buds and expression patterns were visualized by WISH in E12.5 embryos. In all CTCF-deletion alleles a significant 2-fold increase in *Kcnj2* expression could be measured. *Sox9* expression, however, was unchanged except for a mild downregulation to ~85% in the *ΔCTCF* mutant (Fig. 5-4 B). We used whole-mount in situ hybridization (WISH) to visualize possible changes in gene expression patterns in embryos at E12.5. In the *ΔBor, ΔBorC1, ΔBorC1-C2* alleles no *Sox9*-like misexpression of *Kcnj2* could be detected using WISH. *Kcnj2* became slightly misexpressed in a *Sox9*-like pattern only upon deletion of the C1 to C4 sites (*ΔBorC1-C4*), and was more pronounced in the *ΔCTCF* allele. WISH for *Sox9* expression in all alleles was indistinguishable from wildtype. Surprisingly, in all CTCF-deletion alleles (*ΔBor, ΔBorC1, ΔBorC1-C2, ΔBorC1-C4* and *ΔCTCF*) the effects on gene expression were relatively mild (Fig. 5-4 C). The *Kcnj2* up-regulation, although significant, seemed to be rather an unspecific de-repression, as no strong changes in expression patterns were detected by WISH.

Importantly, the adult mice were unaffected, could breed to homozygosity and did not display any symptoms of a Cooks-like phenotype, as analyzed by microCT (appendix, Fig. 10-5 C), supporting the observations of only small changes in gene expression.

Thus, massive reconstruction of the locus' architecture led to only subtle changes in gene expression, which were almost exclusively quantitative, as expression patterns barely changed.

This suggests that a loss of insulation between the *Sox9*- and *Kcnj2*-TAD upon CTCF-deletion is not causing directed rewiring of *Sox9* enhancers with the *Kcnj2* promoter resulting in pathogenic gene misexpression. Rather it leads to a leakage of activity towards the wrong target gene.

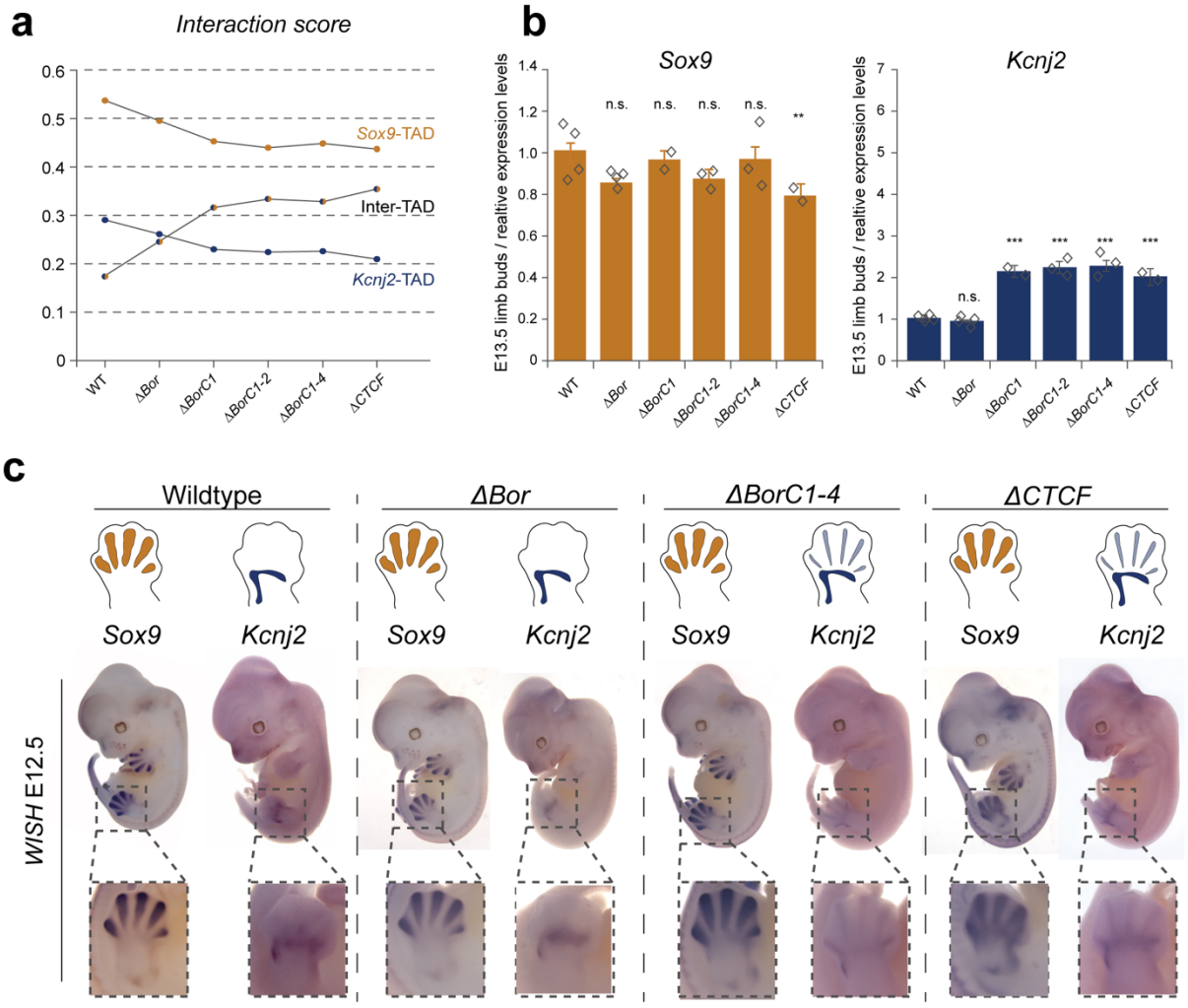

Figure 5-4 Changes in gene expression upon gradual CTCF-deletion and TAD fusion. (A) Changes in interaction score by successive removal of CTCF sites. The more CTCF sites are removed, the more inter-TAD interaction can be observed concomitant with a decrease in intra-TAD interactions. (B) Gene expression changes assessed by RT-qPCR. Relative gene expression levels of *Sox9* and *Kcnj2* in E13.5 limb buds normalized to *Gapdh* expression (wildtype = 1). Bars indicate mean expression, error bars standard deviation. Diamonds indicate individual replicates. One-sided, unpaired t-test was conducted to test significance in comparison to wildtype (**p<0.01; ***p<0.001, ns=not significant). N=2-4. (C) Changes in gene expression patterns by WISH of E12.5 embryos (n=3). Schematic on top summarizes observed differences for individual mutants. Magnification depicts a detailed view of the hindlimbs. *Sox9* is expressed in the digital anlagen. *Kcnj2* is expressed in the distal zeugopod. Note only low misexpression of *Kcnj2* in the *ΔBorC1-4* and *ΔCTCF* mutants.

5.1.3 Structural Variants cause pathogenic rewiring of *Sox9* enhancers with the *Kcnj2* promoter

The unexpected results of the CTCF-deletion series raised the question how structural variants bear the potential to cause pathogenic misexpression of non-target genes, as we found that it cannot be the mere loss of insulation at the *Sox9/Kcnj2*-locus. In principle, *Kcnj2* is able to

respond to *Sox9* regulatory information upon formation of a so-called neo-TAD, as we know from duplications at this locus (Fig. 1-8) (Franke et al., 2016). Based on this and other studies, we predicted that a 1.1 Mb inter-TAD spanning inversion (*InvC*) including the majority of the *Sox9* regulatory landscape and the TAD boundary should result in redirecting *Sox9* regulatory activity towards *Kcnj2*. Using CRISPR/Cas9 we created mice carrying such an inversion.

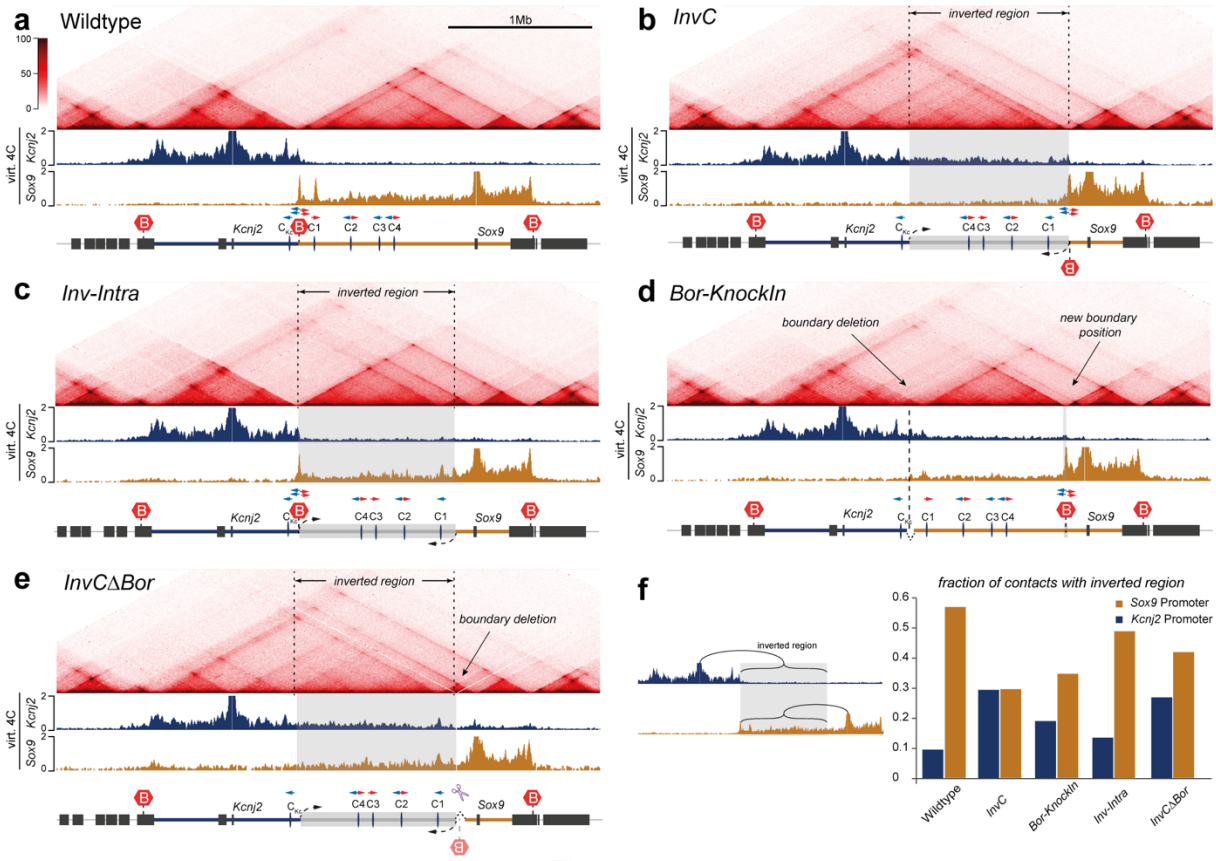

Figure 5-5 Reorganization of TAD architecture after introduction of structural variants. (A) cHiC of E12.5 limb buds from wildtype with corresponding virtual 4C from either the *Kcnj2* (blue) or the *Sox9* promoter (orange) as a viewpoint. (B) Inter-TAD inversion including the boundary (*InvC*) leads to reorganization of the two TADs. The regulatory part of the *Sox9* is fused with the *Kcnj2*-TAD whereas *Sox9* remains in a much smaller domain. (C) Intra-TAD inversion (*Inv-Intra*) only inverts the regulatory domain of *Sox9* leaving the boundary at its wildtype position. Despite reorganization of CTCF-anchored loops, no striking domain reorganization can be observed. (D) Targeted insertion of the TAD boundary (*Bor-KnockIn*) restricts *Sox9* from contacting its regulatory landscape and creates three domains. (E) Deletion of the repositioned TAD boundary in the *InvC* background leads to TAD fusion and rescues interactions of the *Sox9* promoter and its substructure. (F) Fractions of contacts from the *Kcnj2* (blue)- or *Sox9* promoter (orange) with the inverted region. Arrowheads: CTCF site orientation. Grey boxes: Inverted Region.

cHiC from E12.5 limb buds carrying the inversion (*InvC*) showed a comprehensive rearrangement of the 3D chromatin architecture of the locus. Concomitantly, the inversion isolated *Sox9* from the inverted centromeric part of the TAD, which was now completely fused with the *Kcnj2*-TAD

(Fig. 5-5 B). The fraction of contacts from the *Sox9* promoter to the inverted part of the substructure dramatically decreased, whereas *Kcnj2* promoter contacts increased accordingly, as measured by virtual 4C (Fig. 5-5 F). As a consequence of the fused TADs a significant gene misexpression could be detected using qRT-PCR and WISH. *Kcnj2* was fivefold upregulated and expressed in a *Sox9*-like pattern (WISH) (Fig. 5-6 B and C). However, *Sox9*, isolated from its regulatory domain, was downregulated to 50% of wildtype-levels, leading to a loss-of-function (LOF)-phenotype comparable to a heterozygous *Sox9*-knockout described in Bi et al. 2001. Homozygous *InvC*-animals died perinatally with a cleft palate and recapitulated other features observed upon heterozygous *Sox9*-loss, like hypoplasia and bending of many cartilage-derived skeletal structures like the long bones (Bi et al., 2001). Homozygous E18.5 embryos displayed shortened long bones, short snout and delayed ossification (appendix, Fig. 10-6 C). Importantly, heterozygous adult animals were viable but showed a Cooks-like syndrome, in accordance with the *Kcnj2* expression in a *Sox9*-pattern. They displayed malformed terminal phalanges with high penetrance, including loss of the dorsal flexion, sesamoid bones, and claw-shaped form of the terminal phalanx in the 4th digit shown by 3D-uCT (appendix fig. 10-6 B).

Taken together, we could demonstrate that structural variants (SVs) are capable of massively reorganizing chromatin architecture which in turn leads to developmental gene mis-regulation, ultimately resulting in pathogenic phenotypes.

5.1.4 Untangling the role of the boundary and the substructure on the basis of the inter-TAD-Inversion (*InvC*)

The phenotype observed in the *InvC* mutant described previously revealed the power of SVs to cause rewiring of CREs. However, it remained unclear whether this was due to the repositioned TAD boundary, the re-orientated regulatory landscape of *Sox9*, or a combination of both. We decided to create three additional alleles to decipher the function of the different elements.

To investigate the effect of the inverted substructure alone we generated a 18kb smaller intra-TAD inversion (*Inv-Intra*), which carries the identical breakpoint on the telomeric side but excluding the TAD boundary on the centromeric side. Although CTCF sites within the substructure were re-orientated towards the *Kcnj2*-TAD forming stronger loops with the TAD boundary, *Sox9* was still able to contact all its endogenous enhancers (Fig. 5-5 C). In contrast to the *InvC* allele, the ~1 Mb intra-TAD inversion led to no changes in gene expression (qRT-PCR) or patterns (WISH) compared to wildtype (Fig. 5-6 B and C). Accordingly, the animals were fertile, bred to homozygosity and displayed none of the described phenotypes (appendix fig. 10-6 D).

In a second allele, we took advantage of the boundary deletion background mESC line (*ΔBor*) to insert a 6.3 kb construct carrying all four CTCF sites at the exact telomeric breakpoint of the inversions (*Bor-KnockIn*, appendix fig. 10-6 D). As clearly visible in the cHiC map from

homozygous E12.5 limb buds, the boundary insertion split the *Sox9*-TAD in two smaller domains. *Sox9* remained isolated in a smaller telomeric TAD, whereas the centromeric part of the *Sox9*-TAD formed a second domain between the C1 site and the repositioned boundary (Fig. 5-5 D). In contrast, the *Kcnj2*-TAD did not fuse with the centromeric *Sox9*-domain, as seen in the *InvC* allele (Fig. 5-5 B). Instead the regulatory part of the *Sox9*-TAD forms a new, isolated domain by itself, barely interacting with the *Kcnj2* promoter, analyzed with virtual 4C (Fig. 5-5 F). Gene expression analysis using qRT-PCR and WISH showed, based on the cHiC data, expected expression changes. Whereas *Sox9* expression was reduced to 70%, *Kcnj2* was only slightly upregulated by twofold in the mutant (Fig. 5-6 B and C). As a consequence of the expression changes, homozygous animal died perinatally from a cleft palate, and analysis of E18.5 embryos showed a strong *Sox9*-LOF phenotype, yet milder than the one observed in homozygous *InvC* animals (appendix fig. 10-6 D). Importantly, the animals did not display any features of a Cooks-related phenotype at this developmental stage (comparable to the *ΔCTCF*-mutant).

These results demonstrated clearly that a TAD boundary consisting of divergently orientated CTCF-sites can serve as a very potent insulator. Even after relocation from its endogenous position, it was capable of isolating the *Sox9* promoter from the majority of its regulatory activity.

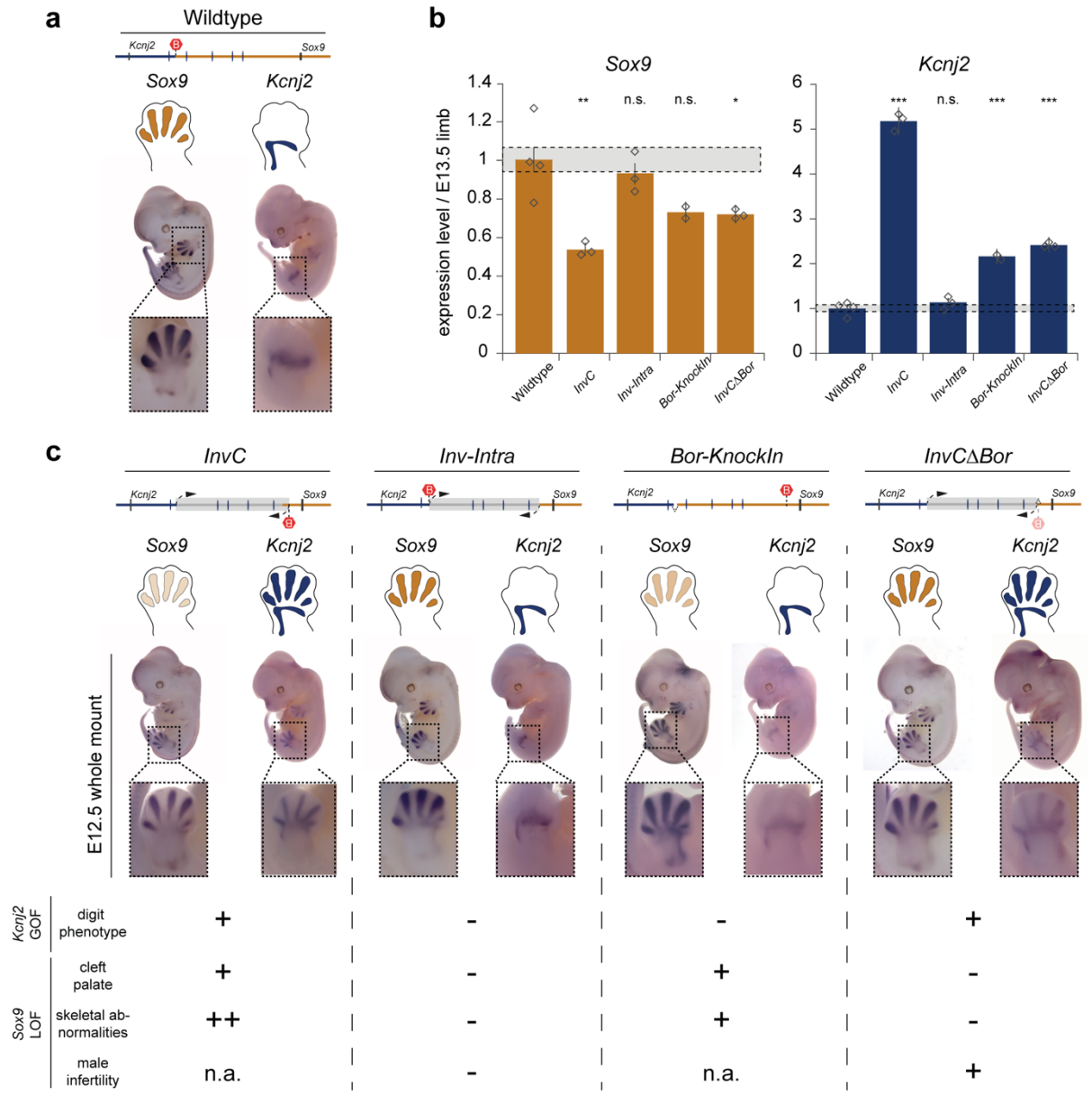

Figure 5-6 Redirection of regulatory information by structural variants results in pathogenic *Kcnj2*-gain-of-function and *Sox9*- loss-of-function phenotypes. (A) Gene expression patterns of *Sox9* and *Kcnj2* in Wildtype (n=7) (B) Gene expression changes assessed by RT-qPCR. Relative gene expression levels of *Sox9* and *Kcnj2* in E13.5 limb buds normalized to *Gapdh* expression (wildtype = 1). Bars indicate mean expression, error bars standard deviation. Diamonds indicate individual replicates. One-sided, unpaired t-test was conducted to test significance in comparison to wildtype (*p<0.05; **p<0.01; ***p<0.001, ns=not significant). n=2-4. (C) Changes in gene expression patterns by WISH of E12.5 embryos (n=3). Schematic on top summarizes rearrangements of individual mutants and observed expression patterns. Magnification below depicts a detailed view of the hindlimbs. *Sox9* is expressed in the digital anlagen. *Kcnj2* is expressed in the distal zeugopod. Note strong misexpression of *Kcnj2* in the *InvC* and I*nvCΔBor* mutants.

The *Inv-Intra* and *Bor-KnockIn* alleles demonstrated that neither of the individual structural variants exclusively is sufficient to cause significant and pathogenic gene misexpression, contrary to both combined. We hypothesized that redirecting regulatory activity by an inverted

substructure alone was not sufficient to cause *Kcnj2*-misexpression since the boundary is blocking interactions and the enhancers tend to have an intrinsic target promoter affinity. To analyze how an inverted substructure/regulatory landscape behaves without the isolating activity of a boundary nearby separating the two domains, we deleted the boundary in the *InvC* allele.
Indeed, by deletion of the boundary in the *InvC* allele (*InvCΔBor*) we could re-establish interactions of the *Sox9* promoter with the centromeric part of its TAD (Fig. 5-5 E). On the other hand, contacts of the *Kcnj2*-TAD with the regulatory domain of *Sox9* were still present, although slightly reduced compared to *InvC* (Fig. 5-5 F). Again, as shown by cHiC, we could observe a complete fusion of the two TADs. In accordance, expression of *Sox9* was restored to 70% of wildtype levels measured by RT-qPCR (Fig. 5-6 B), rescuing the observed *Sox9*-LOF in the *InvC*. Remarkably, *Kcnj2* expression was upregulated to 2.5-fold and the directed rewiring of *Sox9*-enhancers with the *Kcnj2* promoter led to a pronounced *Kcnj2* misexpression in a *Sox9*-like pattern (WISH) (Fig. 5-6 B and C). Adult animals displayed a Cooks-like phenotype in the limb (appendix fig. 10-6 B). Although heterozygous males and females and homozygous females bred normally, homozygous males were infertile, suggesting some *Sox9*-LOF effects due to its role in sex determination and gonad development (Bi et al., 2001; Wagner et al., 1994).
In summary we were able to demonstrate that TAD boundaries act as potent insulators regardless of their position within the domain, whereas TAD substructure orientation is sufficient to redirect regulatory activity towards non-target (responsive) promoters overcoming intrinsic promoter-enhancer affinity. Compared to the relatively mild effects on gene regulation by TAD fusion through CTCF-deletions, domain fusion by TAD shuffling with SVs led to pathogenic rewiring of enhancers with new promoters.

5.2 Evaluating the impact of genomic position on enhancer function by repositioning of a non-redundant limb enhancer at the *Shh*-locus

The results at the *Sox9/Kcnj2*-locus demonstrated that TADs serve as regulatory scaffolds where regulatory information is transmitted throughout the domain and restricted by TAD boundaries. Inversion of the regulatory domain alone did not lead to a decreased target gene expression (*Inv-Intra*). In conclusion, enhancers at the *Sox9/Kcnj2*-locus can function regardless of their position, in line with the generally accepted paradigm of position-independency. However, *Sox9* is regulated by partially redundant enhancers situated in the vast gene desert (Fig. 5-2 A). Therefore, possible changes in individual enhancer activity are not feasible to assess. Thus, we turned to the well-studied regulation of *Shh* during limb bud development by only one single highly-conserved enhancer, the ZRS (Fig. 5-7 B). Deletion of the enhancer entirely abrogates *Shh* expression in the limb bud (Sagai et al., 2005), thus makes it an ideal system to study individual enhancer activity. Regulatory sensor insertions within and outside the *Shh*-TAD recapitulated the ZRS-driven ZPA-expression pattern in the posterior margin of the limb bud vastly, although not completely (Fig. 5-7 A) (Symmons et al., 2014). As these experiment rely on an artificial system and do not reflect properties of the endogenous enhancer nor promoter, we decided to investigate this *in vivo* in mice. To evaluate the impact of the genomic position within and outside the *Shh*-TAD on the functionality of the endogenous enhancer, we applied CRISPR/Cas9 to relocate the ZRS and correlated the new position with enhancer activity assessed by qRT-PCR, WISH and possible phenotypes with skeletal preparations as a direct readout.

5.2.1 Enhancer shuffling at the *Shh*-locus by targeted integrations

In order to assess enhancer functionality in the context of genomic position, we first generated an ESC line carrying a homozygous deletion of the ZRS position (*ΔZRS*). Homozygosity was verified with copy number analysis by qRT-PCR of the mESC clone and the expected single digit oligodactyly was confirmed by generation of E18.5 embryos from tetraploid aggregation (Sagai et al., 2005). The deletion was followed by the reintegration of the enhancer applying targeted CRISPR/Cas9-based KnockIns in the *ΔZRS*-mESC line at two positions located in the adjacent TADs (*ExtraTAD1* and *ExtraTAD2*) and three positions within the domain with increasing distance to the *Shh* promoter (*IntraTAD1-3*; 10 kb, 218 kb, 406 kb). The ZRS sequence was defined according to Sagai et al. including flanking regions of a 100 bps at each side and sequence accuracy was confirmed by Sanger-Sequencing (Fig. 5-7 B, grey arrow bar). From these five different cell lines we generated transgenic mouse lines to then evaluate the effect of the enhancer shuffling by examining transcript abundance, expression pattern, and possible phenotypes by skeletal preparations at E18.5. Every experiment was conducted from limb buds at embryonic stage E10.5.

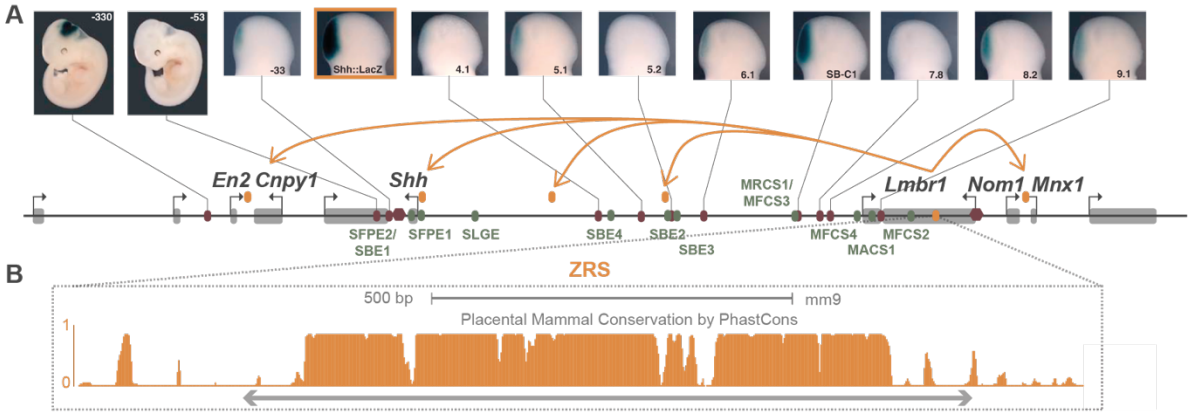

Figure 5-7 ZRS-Repositioning at the *Shh*-locus. (A) Regulatory sensors recapitulate the *Shh* expression pattern in the developing limb at E10.5 in the ZPA vastly but not completely with differing strength *(Symmons et al., 2014)*. Yellow arrows and ovals depict new positions for ZRS-relocation. (B) The ZRS is highly conserved throughout species, as shown by PhastCons. The grey arrow represents the PCR-amplified ZRS-sequence for repositioning.

5.2.2 Relocation outside the *Shh*-TAD causes ZRS-loss-of-function

Relocating the ZRS outside its endogenous TAD in either the neighboring centromeric *En2*-TAD (272 kb of *Shh* promoter, *ExtraTAD1*) or between the genes *Nucleolar protein with MIF4G domain 1 (*Nom1) and *Motor neuron and pancreas homeobox 1* (Mnx1) telomeric of the *Shh* domain (980 kb from *Shh* promoter, *ExtraTAD2*), resulted in a loss-of-function (LOF)-phenotype analogous to the described ZRS knockout by Sagai et al. 2005 (Fig. 5-8 A and B). *Shh* expression was not measurable by qRT-PCR and no expression could be detected by WISH at developmental stage E10.5 in comparison to wildtype (Fig. 5-8C). Animals at E18.5 displayed the typical single digit oligodactyly like animals carrying the *ΔZRS* allele, with loss of the radial bone and a general size-reduction and fusion of the stylo- and zeugopod (Fig. 5-8 D). However, no other organs were affected and we were able to maintain *ExtraTAD1* as a heterozygous mouse line displaying a wildtype phenotype.

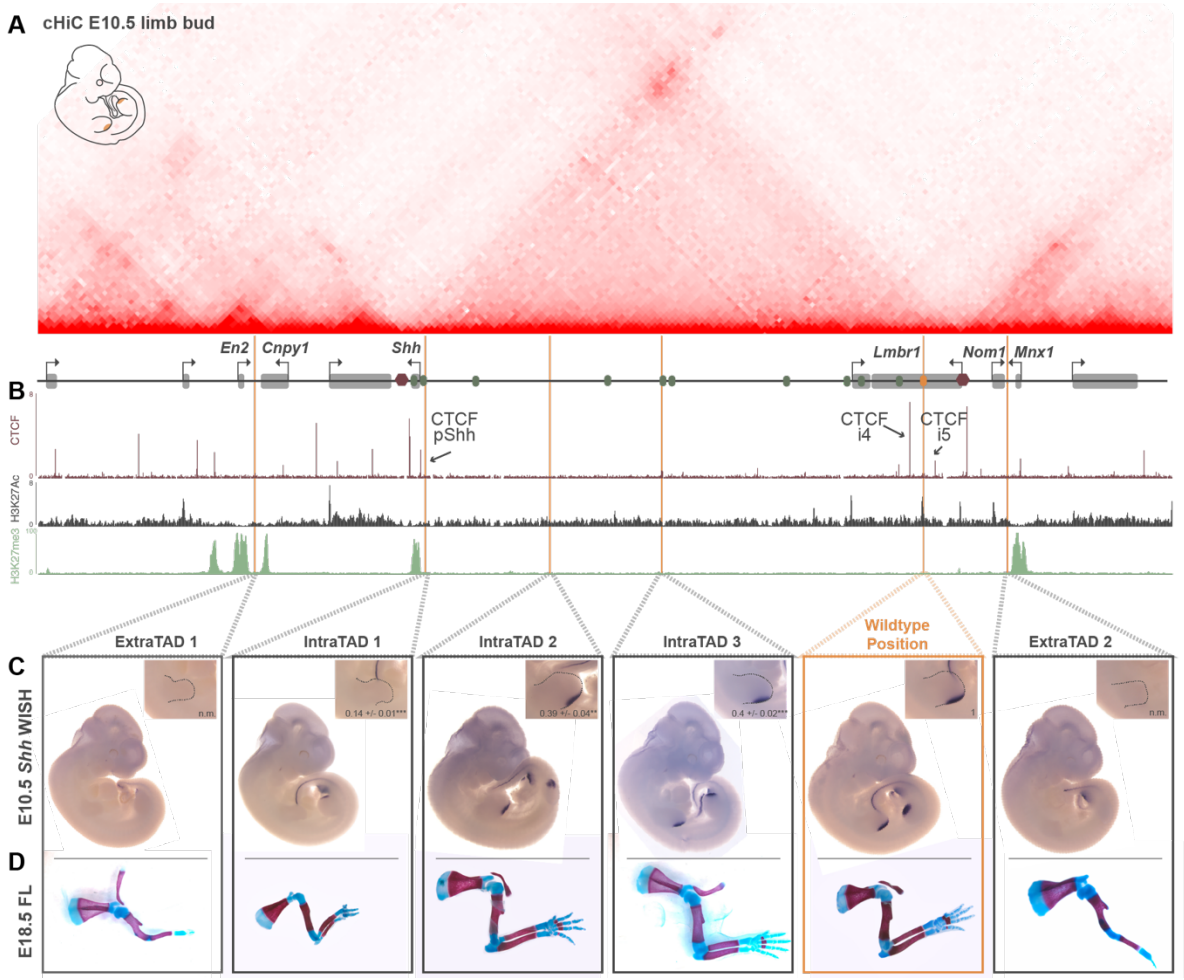

Figure 5-8 ZRS-Repositioning decreases enhancer functionality. (A) Spatial organization of the *Shh*-locus identified by cHiC of wildtype E10.5 limb buds. Schematics below show genes (light grey bars) and enhancers driving *Shh*-expression in different tissues (green ovals) and the limb-specific enhancer ZRS (yellow oval). TAD boundaries are indicated with red hexagons. ZRS-relocation sites are depicted with vertical lines in yellow (data from *(Paliou et al., 2019)*. (B) Chromatin properties identified by ChIP-seq from equivalent tissue of CTCF (top panel, red) and the histone modifications H3K27ac (dark grey) and H3K27me3 (green) (data from *(Andrey et al., 2016)*. Arrows point out CTCF sites telomeric of the *Shh* promoter (pShh) and flanking of the ZRS (i4 and i5). (C) Gene expression patterns (WISH) and quantification of *Shh* expression levels of different *Extra-* and *IntraTAD* mutants (dark grey rectangles) compared to wildtype (yellow rectangle). Magnification of forelimbs are depicted in the upper right of each rectangle together with gene expression quantification data. Relative gene expression levels of *Shh* in E10.5 limb buds were assessed with qRT-PCR and normalized to *Rps9* expression (wildtype = 1, n.m.: not measurable). (D) skeletal preparations of E18.5 forelimbs with alcian blue and alizarin red (Blue: bone. Red: cartilage). (n=2-8)

5.2.3 *IntraTAD* positions result in decreased ZRS activity

Next, we examined differences in ZRS-enhancer activity in the *IntraTAD* mutants. As mentioned earlier, regulatory sensor integrations within the *Shh*-TAD recapitulate the ZRS activity throughout the domain (Symmons et al., 2014). However, different integration sites display inconsistent strength of reporter gene expression. For example, position 4.1 compared to the 5.1 integration

site shows substantial differences in the intensity of the reporter (fig 5-7 B A). Keeping this in mind, the three *IntraTAD* locations were chosen due to their different chromatin properties within the domain, which might explain the variability in reporter gene expression. All locations differ in distance to target promoter, chromatin accessibility, CTCF-site abundance, histone modifications and proximity to other *Shh* enhancers. To quantify changes of *Shh*-transcript abundance we performed qRT-PCR. In addition, possible alterations in gene expression patterns were assessed by WISH.

The *IntraTAD3* position is the most distal of the three relocation sites. In this allele, the ZRS is integrated 400 kb away from the *Shh* promoter, centromeric to the *Shh* brain enhancer 2 (SBE2). The midbrain enhancer is active at the same developmental stage, harboring a potential CTCF-binding site approximately 8 kb telomeric, orientated towards the *Shh* promoter (Fig. 5-9 H). In comparison, at the *IntraTAD2*-position the ZRS is relocated 218 kb upstream from its target promoter in a "chromatin desert" without any histone marks, ATAC-seq peaks, adjacent *Shh*-enhancers or potential CTCF-binding sites in the limb bud at E10.5 (Fig. 5-9 G).

Regardless of the larger distance towards the *Shh* promoter, in embryos carrying the *IntraTAD2* (218kb) or *IntraTAD3* (400kb) allele similar levels of *Shh* expression were measurable in homozygosity compared to wildtype littermates. No differences in expression pattern or strength were visible in WISH from E10.5 embryos (Fig. 5-8 C). However, we found significantly reduced activity of the repositioned enhancer in both alleles (*IntraTAD2* and *IntraTAD3*) assessed with qRT-PCR. *Shh* expression was decreased to 40% of wildtype-levels in the *IntraTAD2* allele (**; p=0,02, n=2-3, one-sided, unpaired t-test). Surprisingly, *IntraTAD3* displayed an equal reduction with 39% of wildtype expression (***; p=0.0003, one-sided, unpaired t-test, n=6-8) (Fig 5-8 C). Homozygous embryos at E18.5 of both mutants did not display any skeletal abnormalities and appeared indistinguishable from wildtype-littermates (Fig. 5-8 D).

The *IntraTAD1* is the position closest to the *Shh* promoter, located 10 kb telomeric in a region depleted of any histone modifications or ATAC-seq peaks in the limb bud (Fig. 5-9 F). Moreover, the ZRS is relocated in the center of the *Shh*-floorplate enhancer (SFPE1), inactive at E10.5 in the limb. Surprisingly, only at the *IntraTAD1* position directly upstream of the *Shh* promoter a size-reduction of the forelimb at E10.5 was noticeable (Fig. 5-8 C). In the forelimb transcript abundance was decreased to 14% of wildtype expression levels (***; p=0.0009, one-sided, unpaired t-test; n=3) compared to the hindlimb with 20% (***, p=0.0003, one-sided, unpaired t-test, n=3). To exclude a forelimb-specific phenotype we analyzed the ratios of forelimb- to hindlimb-expression of *Shh* in comparison to wildtype expression. Equal ratios were calculated for *IntraTAD1* mutants (0.68) and wildtype littermates (0.70). The E10.5 embryos displayed a severe reduction in *Shh*-expression visible in WISH compared to wildtype littermates (Fig. 5-8 C). As a consequence of the drastic reduction in expression, homozygous embryos at E18.5 developed a pronounced

oligodactyly with only 2 digits in the forelimb and 4-5 digits in the hindlimb, visible in the skeletal preparations (Fig. 5-8 D).

The surprising result of the *IntraTAD* mutant series raised the question of why the ZRS is not functioning to its full extent at the diverse positions. As we could not correlate expression level at the different relocation sites with distance to the *Shh* promoter, we took a closer look at the varying chromatin properties of the *IntraTAD* positions.

5.2.4 Generation of a sensitized background at *IntraTAD*-positions to trigger possible phenotypes

As there were similar changes in gene expression at some positions and subtle differences might not be detectable in homozygosity, we decided to mate the animals with a heterozygous *ΔZRS* mouse line carrying only a single-copy of the enhancer. Thus, we created a sensitized background to assess possible differences in enhancer function at the *IntraTAD* positions.

We were able to confirm the reduced *Shh* expression in all *IntraTAD* alleles by qRT-PCR. In hemizygous embryos (*ΔZRS/IntraTAD*) of the *IntraTAD2* and *IntraTAD3* positions, a similar reduction of 33% of wildtype-levels in the *IntraTAD2* mutant (*; p=0.0011, one-sided, unpaired t-test; n=2) and 25% in the *IntraTAD3* (***; p=0.0002, one-sided, unpaired t-test; n=3) could be detected by qRT-PCR (Fig. 5-9 E). However, the relative expression changes between both *IntraTAD* positions were not significant (n.s.; p=0.22, one-sided, unpaired t-test) (Fig. 5-9 E). Moreover, no changes in *Shh* expression patterns (WISH) were visible in a hemizygous state in E10.5 embryos (Fig. 5-9 C and D). Hemizygous embryos at E18.5 did not display any skeletal abnormalities and appeared indistinguishable from wildtype (Fig. 5-9 C and D). Again, the comparable *Shh* expression levels were remarkable, as *IntraTAD2* is located in a region depleted of histone marks, ATAC-seq peaks, *Shh* enhancers or potential CTCF sites. In contrast, *IntraTAD3* is neighboring a known midbrain enhancer (SBE2) active at the same developmental stage. Furthermore, a potential CTCF binding site in convergent orientation towards the *Shh* promoter is located only 8 kb telomeric to the insertion site (Fig 5-9 H).

In line with our first set of experiments, we found an enhanced phenotype in hemizygous *IntraTAD1* animals, showing an overall reduction of *Shh* expression to 13% of wildtype-levels. In comparison to wildtype, E10.5 embryos of the *IntraTAD1* mutant carrying only one repositioned copy of the ZRS showed a clearly visible decrease of *Shh* expression specific to the ZPA, visualized with WISH. Moreover, the forelimb was noticeably reduced in size. Accordingly, hemizygous embryos at E18.5 displayed a more pronounced skeletal phenotype. In addition to the oligodactyly, the forelimb was characterized by additional loss of the radial bone and fusion of stylo- and zeugopod (Fig. 5-9 B).

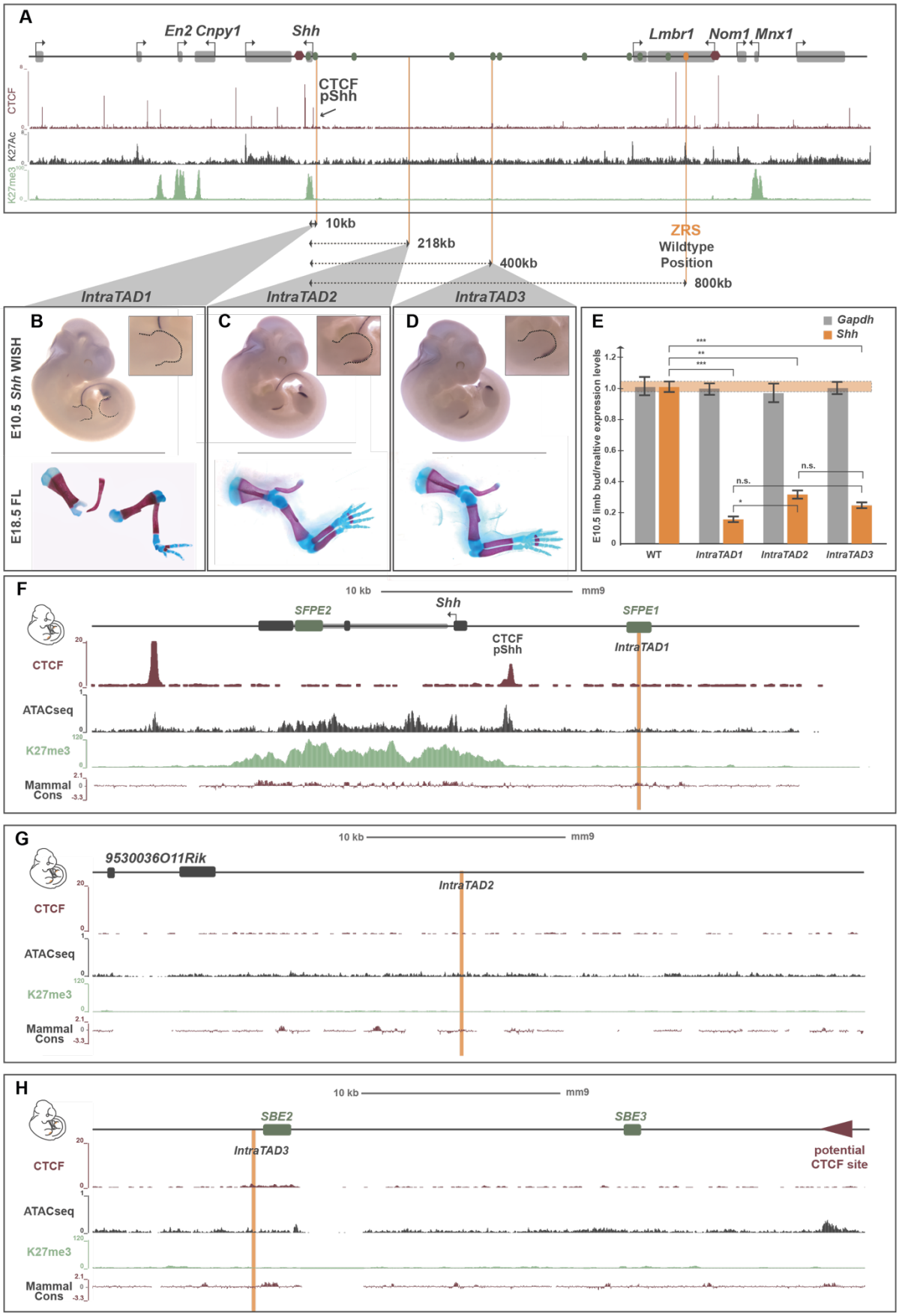

Figure 5-9 Chromatin features at *IntraTAD* positions influence decreased ZRS activity. (A) Schematics show genes (light grey bars) and enhancers driving *Shh* expression in different tissues (green ovals) and the limb-specific

enhancer ZRS (yellow oval). TAD boundaries are indicated with red hexagons. ZRS-relocation sites are depicted with vertical lines in yellow. Dashed lines indicate distance of insertion sites to the *Shh* promoter for each mutant. Chromatin properties identified by ChIP-seq from equivalent tissue of CTCF (top panel, red) and the histone modifications H3K27ac (dark grey) and H3K27me3 (green) (data from *(Andrey et al., 2016)* (B)-(D) Gene expression patterns (WISH) of hemizygous *IntraTAD* mutants. Magnification of forelimbs are depicted in the upper right of each rectangle. Below are skeletal preparations of E18.5 forelimbs with alcian blue and alizarin red (blue: bone. red: cartilage). (E) Relative gene expression levels of *Shh* in E10.5 limb buds were assessed with qRT-PCR and normalized to *Rps9* expression (wildtype = 1, n.m.: not measurable). *Gapdh* was normalized as a control. Bars indicate mean expression, error bars standard deviation. One-sided, unpaired t-test were conducted to test significance in comparison to wildtype (*p<0.05; **p<0.01; ***p<0.001, ns=not significant). n=2-4. (F)-(H) Magnifications of different *IntraTAD* ZRS-relocation sites and chromatin properties. CTCF (top panel, red), ATAC-seq (dark grey), H3K27me3 (green) and Mammal Conservation from PhastCons (red) (ATAC-seq data from *(Paliou et al., 2019)*. Potential CTCF binding sites are indicated by a red arrow. *Shh*-enhancers are depicted with green bars. Insertion sites are represented by yellow lines (n=2-4).

Taken together, relocating the ZRS within its endogenous TAD led to decreased enhancer activity in varying degree at all *IntraTAD* positions. As we could not correlate expression changes with linear distance to the target promoter, we reasoned that rather the chromatin properties present at the new positions seem to have an influence on full function of the enhancer. For the *IntraTAD2* and *3* positions no significant differences in *Shh* expression were detectable, in contrast to the *IntraTAD1* allele. Although located directly upstream of the *Shh* promoter at the *IntraTAD1* position (10 kb), the ZRS activity might be impeded by the flanking inactive floor plate enhancer (SFPE1). Moreover, CTCF binding between the enhancer and the target promoter might influence the activity, suggesting an insulating function (Fig. 5-9 F).

5.2.5 ZRS shows selectivity towards non-endogenous promoters

As described above, the ZRS was unable to induce *Shh* expression when repositioned outside of its endogenous TAD, as it was deprived of the ability to activate the *Shh* promoter. However, several studies describe enhancer-adopted regulation of non-target promoters upon TAD-reorganization by e.g. SVs, leading to ectopic expression. In order to test if the relocation of the ZRS into the adjacent TADs activates the newly neighboring genes in a *Shh* like pattern, we performed WISH and quantified transcript abundance with qRT-PCR of the *ExtraTAD* mutants.
Only at the *ExtraTAD1* position could we detect ectopic expression of *En2*, normally expressed in the developing midbrain at E10.5 and completely absent in the developing limb bud. In this allele the ZRS is located 18 kb telomeric of *En2* and 11.5 kb centromeric of *Cnpy1* (Fig. 5-10 A). *En2* was upregulated 5-fold compared to wildtype littermates as shown by qRT-PCR (***, p=0.0008, one-sided, unpaired t-test, n=3) *(*Fig. 5-10 B). Using WISH we found that *En2* was

misexpressed in a *Shh*-like pattern at the posterior margin of the limb bud in the ZPA (Fig. 5-10 B, grey arrow). Accordingly, *Cnpy1* located downstream of *En2* was upregulated 5.5-fold (**, p=0.003, one-sided, unpaired t-test, n=3) (Fig. 5-10 B). The misexpression of the two genes in the limb bud of *ExtraTAD1* heterozygous animals did not result in any skeletal phenotypes.

At the *ExtraTAD2* position the ZRS is located 23 kb centromeric to the gene *Mnx1* and 19.5 kb telomeric to the gene *Nom1* (Fig. 5-10 A). In wildtype, *Mnx1* is moderately expressed in a small fraction of cells partially overlapping the ZPA in the limb bud at E10.5. *Nom1* is a housekeeping gene, constitutively expressed in several tissues. To our surprise, we could not detect any increase or ectopic expression of the two genes by qRT-PCR or WISH (Fig. 5-10 C).

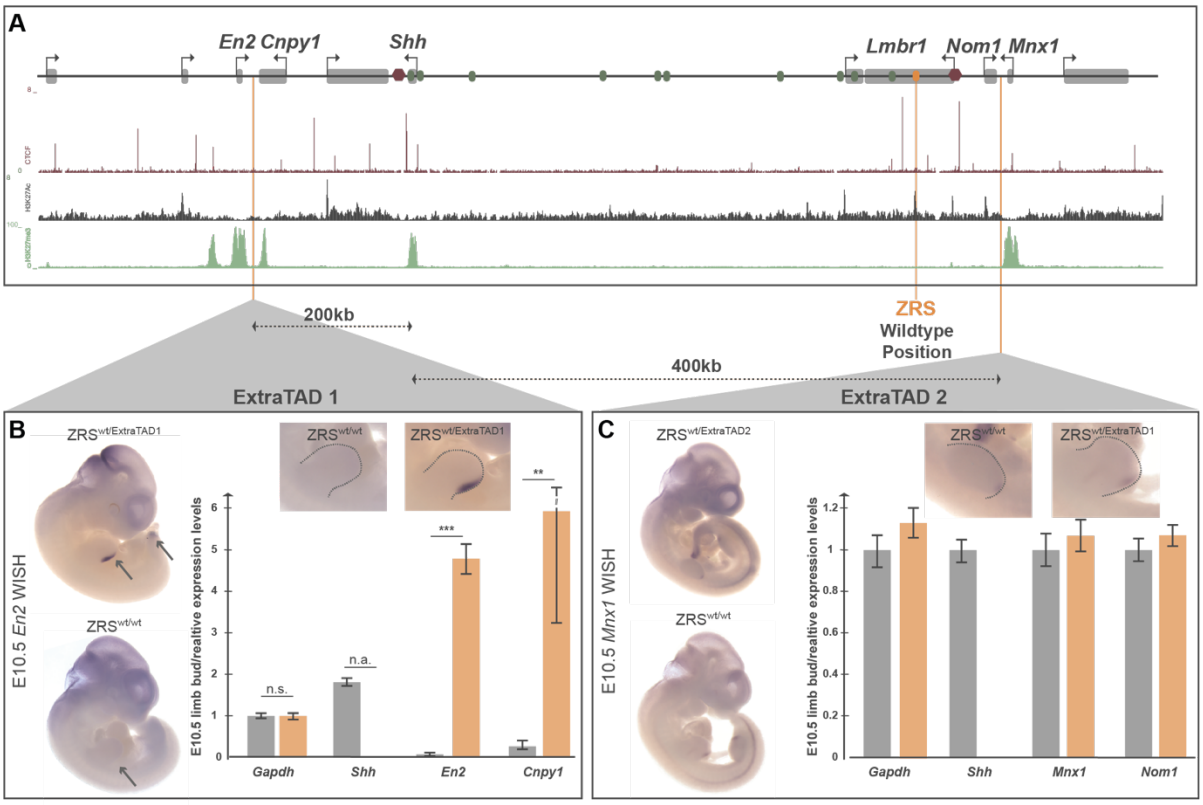

Figure 5-10 ZRS shows selectivity towards non-target promoters upon relocation in the adjacent TADs. (A) Schematics show genes (light grey bars) and enhancers driving *Shh* expression in different tissues (green ovals) and the limb-specific enhancer ZRS (yellow oval). TAD boundaries are indicated with red hexagons. ZRS-relocation sites are depicted with vertical lines in yellow. (B) Chromatin properties identified by ChIP-seq from equivalent tissue of CTCF (top panel, red) and the histone modifications H3K27ac (dark grey) and H3K27me3 (green) (data from Andrey et al. 2016). (B) *En2* expression patterns (WISH) of *ExtraTAD1* mutant compared to wildtype littermate. Ectopic expression of *En2* in the ZPA in the *ExtraTAD* mutant is highlighted with grey arrows. Magnification of forelimbs are depicted on top of the diagram. Relative gene expression levels of *Shh*, *En2* and *Cnpy1* in E10.5 limb buds were assessed with qRT-PCR and normalized to *Rps9* expression. *Gapdh* was normalized as an internal control. Bars indicate mean expression, error bars standard deviation. One-sided, unpaired t-test were conducted to test significance in comparison to wildtype (*p<0.05; **p<0.01; ***p<0.001, ns=not significant, n.a.=not aplicable). (C)

Mnx1-gene expression patterns (WISH) of *ExtraTAD2* mutant compared to wildtype of equal embryonic stage. No ectopic expression of *Mnx1* could be detected. Magnification of forelimbs show *Mnx1*-wildtype expression in the limb partially overlapping the ZPA. Relative gene expression levels of *Shh*, *Mnx1* and *Nom1* in E10.5 limb buds were assessed with qRT-PCR and normalized to *Rps9* expression (wildtype = 1, n.m.: not measurable). *Gapdh* was normalized as a control. Bars indicate mean expression, error bars standard deviation. One-sided, unpaired t-test were conducted to test significance in comparison to wildtype (*p<0.05; **p<0.01; ***p<0.001, n.s.=not significant, n.a.=not applicable, n=3-4).

In summary, in the *ExtraTAD* alleles distinct promoters showed different responsiveness towards the newly positioned ZRS. This was remarkable, as enhancers are thought to promiscuously activate any gene in their vicinity. We were not able to explain our results with differing histone modifications at the different promoters at E10.5 in the developing limb bud (Fig. 5-10 A). Moreover, *En2* and *Mnx1* are both developmental genes like *Shh* and very likely have a similar promoter type. We reasoned that the ZRS shows some sort of selectivity towards non-endogenous promoters. However, the underlying mechanisms are yet to be solved.

6. DISCUSSION

A plethora of cellular processes require accurate gene regulation. Particularly during embryonic development, spatiotemporal control of gene expression is crucial for the constitutive patterning, polarization and cell fate specification during lineage commitment. Such sophisticated regulation is accomplished by CREs. In the nuclear space, the 3D units referred to as TADs are believed to facilitate proximity of enhancers to their cognate promoter and restrict inappropriate contacts. Yet the influence of TADs as architectural microenvironments on gene regulation remains controversial. In this work I aimed to analyze this ambiguity in detail specifically at the domains of the *Sox9*/*Kcnj2*-locus and the *Shh*-locus. The distinct expression patterns of the corresponding genes together with the associated phenotypes suit both loci perfectly for this approach. By dissecting the role of TAD boundaries, intra-TAD CTCF sites, and orientation of TAD substructures for TAD formation, I was able to correlate changes in the spatial organization of the loci with gene expression analysis and phenotypes. This revealed important features of genome architecture with respect to transcriptional output i*n vivo* during limb bud development, highlighting that the mere loss of insulation is not sufficient for pathogenic misexpression. However, massive rearrangements by reorganizing functional units of TADs identifies instructive roles.

In the second part, I was able to analyze enhancer activity dependent on its location within and outside its TAD at the *Shh*-locus. This is the first study assessing enhancer function at differing genomic locations during development *in vivo*. Our results challenge the accepted paradigm of position-independent enhancer function.

6.1 Functional dissection of TADs at the *Sox9*/*Kcnj2*-locus

Functionally, TADs are thought of as scaffolds for regulatory landscapes of genes, facilitating frequent interactions within and restricting ectopic contacts outside of a domain. Long-range contacts by formation of chromatin loops frequently occur in complex gene regulatory landscapes of developmental genes, highly correlating with the CTCF-mediated structure. However, the functional relationship between TADs and transcriptional control remains elusive. As CTCF depletion in mESC demonstrated a loss of TADs genome wide with, surprisingly, only mild effects on gene regulation, we were curious how sequential deletion of distinct CTCF binding sites would impact gene regulation at a single locus *in vivo* during limb bud development.

The *Sox9*/*Kcnj2*-locus consists of two TADs separated by a TAD boundary. In particular, the *Sox9*-TAD spans a genomic region of 1.7 Mb, whereas the adjacent smaller TAD of the potassium

channel *Kcnj2* and *Kcnj16* covers roughly 1 Mb (Franke, 2017). Showing the characteristic features of two pairs of divergently oriented CTCF binding sites, the TAD boundary forms chromatin loops not only with the distal neighboring boundaries, but also with CTCF sites located within the TAD substructures, as demonstrated by cHiC (C1-C4, C_{Kc}). We utilized capture HiC at the extended *Sox9*/*Kcnj2*-locus in E12.5 limb buds to analyze spatial reorganization upon gradual CTCF-deletion with regards to transcriptional control.

6.1.1 TAD boundaries act combinatorial with intra-TAD CTCF sites and redundantly on spatial separation

Previous work at the locus from Franke et al. revealed that deleting the TAD boundary has mild consequences on the spatial separation of the locus. Concomitantly, no significant changes in *Sox9* or *Kcnj2* expression could be detected (Franke et al., 2016). However, both TADs are shaped by additional CTCF sites located within the substructures, mediating the overall TAD structures, visible in cHiC. Taking advantage of the CRISPR/Cas9 technique we systematically targeted CTCF binding sites within the *Sox9*/*Kcnj2*-TADs. By sequential CTCF site removal we created a deletion-series that demonstrated gradual fusion of the *Kcnj2*- with the *Sox9*-TAD. However, not all intra-TAD CTCF binding sites contributed equally to the insulation. Most of the TAD fusion occurred after deletion of the boundary in combination with the C1-binding site (*ΔBorC1*). Surprisingly, only a moderate increase could be detected after additional deletions. Intriguingly, we could not correlate the differing insulation potential to binding strength or motif orientation of the individual CTCF binding sites. The sequential deletion of the TAD boundary in combination with intra-TAD CTCF sites is necessary for comprehensive TAD fusion and thus, the loss of insulation. This indicates that intra-TAD CTCF sites act together with the TAD boundary in a redundant fashion. Hence, TADs seemingly are built in a cooperative action which facilitates strength and robustness to the overall domain architecture.

Similar to our results, resistance of TADs was reported at the *HoxD* gene cluster. The TAD subdivides into two enhancer-rich sub-regions centromeric and telomeric to the cluster, each of them controlling distinct gene expression patterns in the limb. Therefore, the *HoxD* cluster itself is considered a boundary. However, only a deletion spanning the entire HoxD cluster and two adjacent genes resulted in a TAD fusion, similar to the one we report here. Yet, both enhancer clusters continued to work independently (Rodríguez-Carballo et al., 2017).

TAD boundaries formed by CTCF, however, do not exclusively contribute to TAD formation. A CTCF-deletion at the *HoxA* cluster implicated the involvement of Polycomb-mediated spatial hubs (Narendra et al., 2016). Furthermore, removing the *Xist*/*Tsix* boundary by deletion of 58kb surrounding a CTCF site led to domain-fusion, indicating other mechanisms of TAD formation (Nora et al., 2012).

However, the majority of altered spatial separation by TADs were results of an affected CTCF-dependent insulation. In different pathogenic deletions at the *EphA4*-locus the presence of a TAD boundary determines whether two TADs merge. The resulting misexpression upon TAD fusion of *Pax3* is, however, the result of a deletion that removes not only the boundary but, in addition, the majority of the *Epha4* TAD. In comparison to the CTCF-deletion series presented here, this is the effect of an alteration of the overall configuration of the locus (Lupiáñez et al., 2015).
In summary, our deletion series demonstrates the importance of intra-TAD CTCF sites for TAD formation through buffered and redundant mechanisms, and highlights the function of boundaries as strong insulators at the edges of TADs.

6.1.2 Structural fusion of TADs is not a functional fusion

We could circumvent the cytotoxic effect of CTCF-loss by systematically targeting CTCF binding sites of an individual locus. By this, we were able to address the role of CTCF *in vivo* during development. Surprisingly, we did not detect any major changes in gene expression levels nor pattern. *Sox9* is regulated by a large quantity of putative enhancers located in the gene desert. Despite the structural fusion of the domains, the enhancers were presumably still able to contact and regulate their target promoters, indicating a mechanism independent of CTCF.
This suggests that CTCF is not essential for individual enhancer function and, thus, the observed structural fusion of TADs is not a functional fusion. This interpretation is in agreement with the mild changes in gene expression observed upon CTCF-depletion in mESCs (Nora et al., 2017). Furthermore this suggests the presence of additional mechanisms conferring to promoter-enhancer communication independent of a CTCF-mediated architecture.

Such sophisticated mechanisms could be mediated by homotypic interactions of TFs. Biochemically, the formation of transcriptionally active chromatin in membrane-less condensates might promote the bridging of distal enhancers with their cognate promoter (Boija et al., 2018). Phase-separating condensates with high concentration of the transcription apparatus have been described recently as a mechanisms of super-enhancer function (Sabari et al., 2018). The recruitment of RNA PolII and other co-activators by the phase-separating capacity of TFs into transcriptionally active condensates would not be disturbed by CTCF-removal at the *Sox9/Kcnj2*-locus. In this light, gene regulation by TFs would not be affected by the overall 3D configuration of the locus. Here, TADs serve as a framework for robustness and precision, supporting TF-mediated enhancer-promoter interactions by bringing them into spatial proximity, without being essential to establish them.
Another important aspect involves the loop extrusion process. CTCF-deletion *per se* does not affect the recruitment of the extruding factor cohesin on the DNA. Therefore, cohesin complexes

can still facilitate enhancer-promoter contacts. The lack of stalling elements usually set by CTCF bound to the DNA, might broaden the limits of these interactions upon CTCF-deletion. Sequential removal of the TAD boundary and additional intra-TAD CTCF sites therefore increased the contact frequency of the *Kcnj2* promoter with *Sox9*-regulatory enhancers, but not with the *Sox9*-promoter, as measured by virtual 4C. We could only detect a 2-fold upregulation of *Kcnj2* in all alleles, and only upon deletion of each CTFC binding sites separating the two promoters was a misexpression of *Kcnj2* in a *Sox9*-like pattern evident. Interestingly, interaction frequency of the *Kcnj2* promoter and the *Sox9*-regulatory landscape of the *ΔCTCF* was similar to the one observed in the *InvC*, yet it was not sufficient to cause any phenotype. This suggests that contact frequency cannot be directly translated into regulatory activity. Variances in gene regulation might be caused by promoter competition for the same enhancers in the *ΔCTCF*, whereas interactions in the inversion are limited by the repositioned boundary, isolating *Sox9*. Assessing this mechanism *in vivo*, however, is unfeasible due to the lethality of *Sox9* promoter deletions.

Another important architectural protein which has been implemented in facilitating promoter-enhancer contacts, is YY1. HiChIP data revealed numerous contacts mediated by YY1 (Beagan et al., 2017; Weintraub et al., 2017). Thus, the nearly unchanged levels in gene expression could be due to a mechanism involving YY1, facilitating specific interactions of the *Sox9*-enhancers and their cognate promoter, independent from CTCF.

6.1.3 De-repression and activity spreading as a consequence of absent insulation

Another interpretation of the mild *Kcnj2* misexpression in the CTCF-deletion series could be as follows. *Sox9* is only active in the digital anlagen of E12.5 limb buds, whereas *Kcnj2* is expressed in the proximal part at the same developmental stage. Repressed genes comprise of a different chromatin signature compared to active ones and are often associated with the nuclear lamina in so-called Lamina-associated domains, frequently overlapping with TADs (Guelen et al., 2008; Steensel and Belmont, 2017). In the case of the *Sox9*/*Kcnj2*-locus in the limb bud the *Kcnj2*-TAD is associated with the Lamina whereas the *Sox9*-TAD is not (Ringel et al., unpublished data, personal communication). By removing CTCF and by these insulating elements activity is no longer delimited by spatial partitioning anymore, as proposed by Nora et al. (Nora et al., 2012). CTCF has been found to be enriched at domain boundaries of facultative heterochromatin characterized by H3K27me3-marks (Cuddapah et al., 2009). At the *HoxA*-locus removal of two CTCF binding sites which subdivide the locus into two functional chromatin domains depicted with different histone modifications, led to decreased K27me3-levels associated with an up-regulation of the gene (Narendra et al., 2015). Thus, the mild upregulation of *Kcnj2* might be the effect of a de-repression by activity spreading and release from the nuclear lamina. Additionally, transcription has been described as an active remodeller in compartment switching (Heinz et al.,

2018). We cannot exclude active transcription of enhancing elements in the *Sox9*-TAD (Arnold et al., 2020; Li et al., 2016). As RNA PolII elongation is not delimited by a competing TAD boundary in the *ΔCTCF* allele, locus decompaction of the *Kcnj2*-TAD from a repressed B- into an active A-compartment might be a consequence, resulting in deregulated activation of the *Kcnj2*-gene. Therefore, the TAD boundary in cooperation with intra-TAD CTCF binding sites might not only restrict interaction of enhancers with non-target promoters, but might also maintain the integrity of the state of activity within and between domains.

Despite this, the loss of TADs and the concomitant de-insulation at the *Sox9*/*Kcnj2*-locus had no phenotypic consequences, indicating that there were no substantial effects on *Sox9* or *Kcnj2* regulation throughout development. Presumably, *Sox9* regulatory enhancers have an intrinsic affinity towards their target promoter, providing an additional layer of gene regulation.

We reason that the changes in gene expression accompanied by the loss of CTCF-mediated insulation were induced by leakage of regulatory activity towards the wrong promoter in the neighboring TAD. *Sox9* and *Kcnj2* only had a mild de-regulation. Hence, TADs and their boundaries at the *Sox9*/*Kcnj2*-locus are not essential for developmental gene expression, but, they confer precision and robustness. In other scenarios, however, insulation might be essential and such spreading of regulatory activity can result in disease phenotypes (Narendra et al., 2016).

Our findings untangle the idea of TAD structures as fundamental basis of gene regulation and rather hint towards to distinct regulatory layers: (i) The spatial separation by TADs provides an architectural framework for (ii) promoter-enhancer interaction which can function independently of the former. Both layers stabilize each other but are not inherently related. However, the results obtained in this study relating the 3D chromatin organization with gene regulation concentrates on one particular locus. Other TADs might differ in terms of their structure and its function, emphasizing the importance of comprehensive in-depth analysis of individual loci, as genome-wide approaches are not feasible.

6.1.4 Structural Variants rearrange TADs

The results of the CTCF-deletion series stand in conflict to the pathogenic effect of altered TADs in structural variants. In a second series we generated inversions and knock-ins to reveal the relevance of rearranged TADs on gene regulation. By focusing on the TAD boundary and the TAD substructure separately, we could decipher the effect of both independent from one another.

In the *InvC* allele we created a TAD-spanning inversion, inverting the *Sox9* regulatory substructure together with the repositioning of the TAD boundary. By the relocation of a strong insulator, *Sox9* was isolated from its enhancers in a significantly smaller remaining domain resulting in a loss-of-function phenotype similar to a heterozygous *Sox9* KO (Bi et al., 2001). Simultaneously the re-orientated substructure together with the *Sox9* regulatory domain fuses with the *Kcnj2*-TAD by

forming new loops with the inverted CTCF binding sites. Rewiring the *Kcnj2* promoter with the *Sox9* enhancers leads to gain-of-function and pathogenic phenotypes. The misexpression is a consequence of the combination of both structural rearrangements. In this scenario, the intrinsic affinity of *Sox9* enhancers towards their target promoter is blocked by the repositioned TAD boundary. Simultaneously, the reorientation of the TAD substructure towards *Kcnj2* is sufficient to connect them with the wrong target promoter, thereby causing *Kcnj2* mis-regulation. Similar results were obtained by an inversion of the bipartite spatial organization of TADs at the X-inactivation center in mESCs. A 40kb-spanning inversion including the TAD boundary of the *Xist*/*Tsix* transcriptional unit changed the preferential interaction profiles of the replaced promoters, causing aberrant activation of *Xist* (Bemmel et al., 2019).

Thus, the rearrangement of TAD structures can induce misexpression and, in case of the *Sox9*/*Kcnj2*-locus, phenotypes. Our findings indicate that TAD boundaries and TAD substructures function together, but the latter cannot override the insulating function of a strong boundary (see *Inv-Intra* and *Bor-KnockIn*). The inversion of the substructure, however, is obligatory to redirect the regulatory activity to obtain pathogenic misexpression, as seen in our mutant series. Again, this emphasizes that the importance of rearranged 3D chromatin structure for gene misexpression, e.g. in disease-causing SVs, needs to be considered independently from its role for “normal” gene regulation.

However, not all TAD rearrangements result in gene misexpression. For example, Drosophila balancer chromosome are highly rearranged. Several structural variants like inversions, duplications and deletions are found, reorganizing domains by shuffling and fusions all over the genome. Remarkably, the majority of genes in rearranged TADs remained unaffected in the developing embryo (Ghavi-Helm et al., 2019). Yet subtle changes in gene expression were hard to detect, as no specific tissue was analyzed but the whole embryo, diluting out potentially mis-regulated genes. In addition, selection pressure is exerted on balancer chromosomes. Strong misexpression of developmental genes may have lethal consequences. Regardless, the Drosophila genome organization differs from that of mammals and seemingly depends on chromatin state domains rather than on precise loops anchored by CTCF and cohesin (Matthews and White, 2019). A fundamentally different organization makes a direct comparison of the impact of TAD structure on gene regulation challenging.

The mild effects on gene regulation upon CTCF or cohesin depletion oppose the drastic effects of TAD reorganization in pathogenic structural variants. The results obtained here can explain this discrepancy. We found a high redundancy of CTCF sites maintaining TAD structures. Because of this sustained scaffolding, smaller structural variants are likely to be tolerated. Based on our findings, gene mis-regulation needs an actively reorganized chromatin structure, where changes in contact are mostly an activity redirection towards a non-target gene. Duplications or inversions

including TAD boundaries together with adjacent CTCF binding sites reconnect regulatory regions. TAD boundary repositioning, however, separates a gene from its enhancers and can result in a loss-of-expression. Recent studies revealed the influence of the surrounding chromatin on the isolating potential of domain boundaries (Zhang et al., 2020), stressing the context-dependency of TAD function at each individual locus.

Therefore, gene misexpression induced by structural variants is not generated by the simple removal of barriers or single enhancer–promoter rewiring. Rather, it is the consequence of a comprehensive rearrangement by connecting larger regulatory structures with novel target genes through CTCF-mediated loops.

6.2 Enhancer Shuffling at the *Shh*-locus

The results at the *Sox9/Kcnj2*-locus demonstrated the ability of structural variants to cause misexpression, largely due to repositioned CTCF binding sites of the TAD substructure and boundary, forming new chromatin loops. However, it remains unclear how and to what extend TADs impact individual enhancer function. As shown with our CTCF-deletion series, the mere loss of TADs does not lead to severe gene deregulation. This is in line with genome-wide TAD-loss upon depletion of CTCF, causing only mild changes in the regulation of a handful of genes *in vitro* (Nora et al., 2017). This also suggests that TADs are not absolutely necessary for enhancer function. Rather, TADs promote a framework for their function, although enhancers can work without TADs. Traditionally, enhancer function and activity is believed to be position-independent and enhancers can promiscuously activate any gene in their vicinity (Bulger and Groudine, 2011b). This generally accepted paradigm, however, largely relies on artificial assays, where engineered reporters do not investigate enhancer function dependence on its chromatin environment. Hence, if TADs have no impact on transcriptional control at a given locus an enhancer should be able to regulate a target gene regardless of its position, as proposed in the earliest enhancer definition.

To test this hypothesis, we took advantage of the well-studied *Shh*-locus. We repositioned the ZRS enhancer to various locations within and outside its TAD. By this, we were able to assess the role of genomic position relative to TAD structure for enhancer-driven activation of its cognate promoter in the developing limb bud *in vivo*.

6.2.1 Enhancer activity is restricted via TAD-boundaries

We observed that the limb-enhancer ZRS is deprived of its ability to activate the *Shh* promoter if positioned outside the endogenous *Shh*-TAD. This impaired activity is most likely constricted via TAD boundaries. Despite the closer genomic distance to the *Shh* promoter in the *ExtraTAD1* allele, the ZRS failed to activate its target gene. Similar to the effects of a separating TAD boundary

at the *Sox9/Kcnj2*-locus, activity remained restricted to the neighboring domain of *En2*, underlining the functional relevance of insulation maintained by TADs. A similar loss-of-function of the enhancer could be observed in the *ExtraTAD2* mutant. Again, interactions of the ZRS with the *Shh* promoter were presumably blocked by an interfering TAD boundary located between the *Shh*-TAD and the adjacent *Nom1/Mnx1*-domain, maintained by numerous CTCF binding sites.
Several studies, including this one, highlighted the power of CTCF-mediated TAD boundaries (Despang et al., 2019; Zhang et al., 2020). Boundary-spanning SVs have been shown in several cases to be associated with ectopic expression or isolation of genes from their regulatory elements (Despang et al., 2019; Franke et al., 2016; Lupiáñez et al., 2015; Weischenfeldt et al., 2017). Moreover, ectopic activation of proto-oncogenes has been described as a consequence of TAD boundary disruption (Hnisz et al., 2016), emphasizing the importance of TAD boundary integrity for appropriate gene regulation.

6.2.2 Relocation within the *Shh*-TAD impedes ZRS enhancer functionality

TADs might provide microenvironments mediating regulatory activity from enhancers towards their target promoter. Transmission of regulatory information is supposed to be present throughout a domain, as shown by regulatory sensors integrated in several developmental landscapes (Franke, 2017; Ruf et al., 2011; Symmons et al., 2014). However, the regulatory information recapitulated by these sensors varies throughout the TADs, which suggests that not all positions within the domain are equally receptive to regulatory information. Here, we found in our *IntraTAD* series that relocation of the ZRS within its endogenous domain impedes its functionality. We could show that any relocation of the ZRS within the *Shh*-TAD influenced the transcriptional output and that this could not be explained by linear distance. However, not all generated mutant mice displayed limb malformations. The range of *Shh* expression levels which cause a phenotype is highly variable ranging from 10-50% of wildtype levels (Krebs et al., 2003; Lettice et al., 2014; Paliou et al., 2019). This shows that gene regulation and phenotypic effects of mis-regulation might be buffered, which is advantageous in context of disease, development and evolution, as it confers robustness.

6.2.3 Effects of the surrounding chromatin for ZRS function at the *IntraTAD* positions

Integration of the ZRS at the *IntraTAD2* and *3* positions revealed similar levels of activity (40% of wt), sufficient to not cause skeletal abnormalities. The nearby chromatin features of the two positions differ with respect to chromatin accessibility, histone modifications, *Shh* enhancer abundance and potential CTCF binding sites in their vicinity. However, this was not reflected in expression levels, nor the distance to the target promoter, since both relocation sites are approximately 200 kb apart. Importantly, a comparable decrease of enhancer-promoter distance

by intra-TAD deletions did not affect *Shh* expression levels (Symmons et al., 2016). These deletions, however, did not change the overall chromatin environment of the ZRS, in contrast to the *IntraTAD* positions.

We chose the *IntraTAD3* position due to the presence of a neighboring *Shh* brain enhancer (SBE2). In the developing brain and other structures of the central nervous system (CNS), *Shh* is under the control of several partially redundant enhancers (Fig. 1-9 C). In general, we know that the brain enhancer is able to regulate *Shh* expression at the same developmental stage at E10.5, although in a different tissue. The *IntraTAD3* position, however, displayed similar expression levels and therefore enhancer activity, as the *IntraTAD2* position which was deficient of nearby *Shh* enhancers. The absence of advantageous effects on transcriptional output might be explained by fundamentally different mechanisms of gene activation in the limb compared to the CNS. 3D FISH at the *Shh* locus during NPC differentiation demonstrated an increase of physical distance of the *Shh* promoter and the SBE enhancers upon activation (Benabdallah et al., 2019). These results are counterintuitive and disagree with the prevalent model of activation relying on physical proximity, which has been shown for the ZRS and *Shh* (Williamson et al., 2019) (see below). Similar observations were obtained by live-cell imaging of mESC, where no enhanced spatial proximity upon activation between *Sox2* and its regulating region could be detected (Alexander et al., 2019). The proximity-based *Shh* activation of the ZRS in the limb (see below) is in stark contrast to the activation-mode of the brain regulatory elements, conferring increased spatial separation (Benabdallah et al., 2019; Williamson et al., 2019).

At the *Shh*-locus different mechanisms of enhancer-mediated gene activation might be present for the identical promoter in the same TAD. Both mechanisms co-exist but cannot compensate for each other or be exchanged, emphasizing the diversity of spatiotemporal gene regulation. An interesting experiment would be the relocation of the SBE2 enhancer to the ZRS-wildtype position to assess whether this activation mechanism functions in a position-independent manner.

6.2.4 Chromatin topology affects enhancer function at the *Shh*-locus

There were no significant differences in *Shh* expression levels between the *IntraTAD2* and *3* alleles. Yet both were significantly decreased compared to wildtype. Despite the general activity loss, no significant differences in enhancer functionality relating to chromatin features or distance to target promoter between both individual relocation sites were obvious, hinting towards a more universal origin of the decreased expression.

A possible explanation of the decreased transcriptional output upon repositioning is the presence of a preformed complex. At its endogenous site, the ZRS is located in a tissue-invariant loop with the *Shh* promoter (Paliou et al., 2019; Williamson et al., 2016, 2019) The stable topology is maintained by two pairs of convergently oriented CTCF binding sites flanking the *Shh* promoter

and the ZRS. By stalling of cohesin complexes, the enhancer and the promoter are kept in place for immediate transcriptional activation. Repositioning the ZRS away from those CTCF-sites could result in less frequent contacts with the *Shh* promoter. Moreover, the proximity of the ZRS towards *Shh* has been shown to be critical for the gene's activation. Within the preformed structure the promoter and enhancer stay already invariantly in close vicinity ranging from 200-400nm distance. Upon transcriptional initiation in cells specific to the ZPA both elements are compacted further to less than 200 nm (Williamson et al., 2016). As a result of repositioning the ZRS elsewhere in the TAD, the contact frequency with the *Shh* promoter might be decreased, possibly originating from "looping" away the enhancer (Figure 6-1 A and B). However, the ZRS is still able to find its cognate promoter, although to a lesser degree.

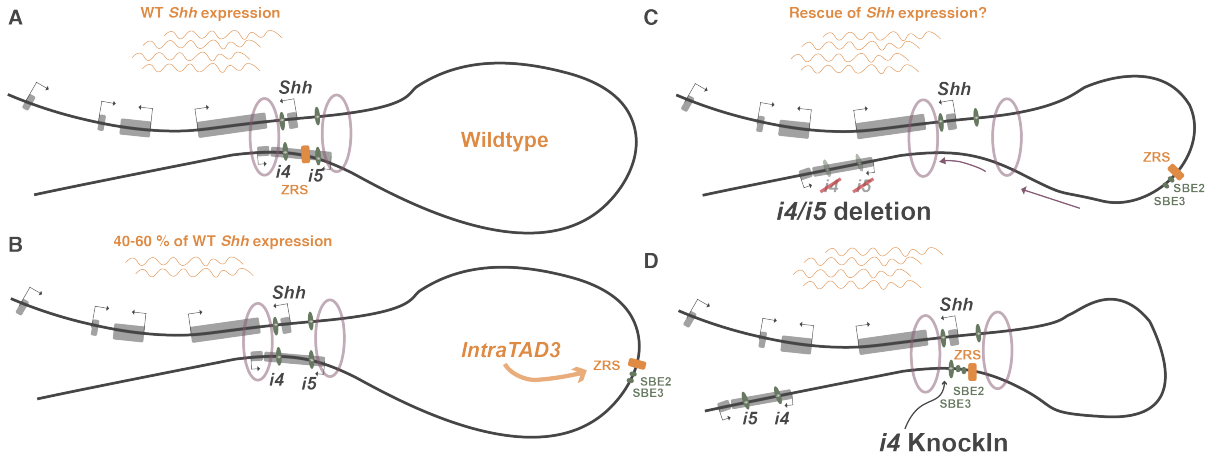

Figure 6-1 Decreased ZRS activity of *IntraTAD* positions as a consequence of a preformed chromatin topology

.(A) In the wildtype position the ZRS is in close proximity to the *Shh* promoter, kept in place by cohesin complexes stalled at the two pairs of flanking CTCF binding sites. (B) At the *IntraTAD3* (C) Destruction of the preformed complex allows cohesin complexes to further extrude the DNA fiber and reels in the ZRS more frequently. (D) Knock in of i4-CTCF binding site stabilizes contact frequency by chromatin loop formation with the convergently orientated ones flanking the *Shh* promoter. Green ovals: CTCF binding site; Purple rings: cohesin complex; orange box: ZRS.

Despite of operating to a lesser degree, the regulation at the *Shh*-locus still works after ZRS-repositioning to distinct locations within the domain, possibly mediated through an intrinsic promoter-enhancer affinity. The expression-loss of 60% at the *IntraTAD2* and *3* positions cannot be explained by the mere distance to target promoter, nor properties of the surrounding chromatin. Interestingly, removal of the preformed complex results in a similar decrease of transcriptional output (50%) (Paliou et al., 2019). The preset chromatin topology has been hypothesized to confer robust and precise gene expression (Paliou et al., 2019). Therefore, the effect of removing the CTCF-mediated invariant proximity compared to the relocation in a CTCF-

depleted area within the same domain here, might have similar consequences (50% versus 40% expression).

In this scenario, breaking up the preset topology should partially re-establish prior lost contact frequencies. Deletion of the CTCF sites flanking the ZRS in wildtype (referred to as i4 and i5) should be able to rescue the loss of expression, as there is no "looping" out of the ZRS at the new positions (Fig. 6-1C). However, loss of the whole complex could have an even stronger effect on *Shh* regulation. Another strategy of testing this hypothesis would be the targeted knock in of the i4-CTCF bindings site telomeric of the integration site, simultaneously avoiding secondary effects (Fig. 6-1 D). By this, new chromatin loops could be formed with the convergently oriented binding sites flanking the *Shh* promoter and by this re-establish the effect of a preformed complex. Preformed complexes for enhancer loops have been described to remain stable during development (Ghavi-Helm et al., 2014). Perturbations of such structures by disruption of the architectural loops themselves (Paliou et al., 2019) or, as here, taking out an enhancer from its endogenous position, might affect the cis-integrity of regulatory elements. Relocated from its evolutionary optimized setting for enhancer functionality, the residual activity is sufficient to control *Shh* expression in the limb bud, yet remains impeded.

6.2.5 ZRS requires long-range regulation for *Shh* activation

Proximity of the ZRS to the *Shh* promoter has been described as crucial for ZRS functionality (Williamson et al., 2016). Remarkably, the enhancer displayed the most significant functionality-loss at the proximal position. To our surprise, *IntraTAD1* positioned 10 kb telomeric, and closest to the *Shh* promoter, displayed the most severe reduction of expression concomitant with a loss-of-function phenotype. This effect of decreased ZRS activity might be due to insulation of the *Shh* promoter and the enhancer by a CTCF binding site telomeric to the gene body. Here, the CTCF sites would not facilitate, but insulate enhancer-promoter contacts. Williamson et al. showed that deletion of this 5' CTCF binding site (CTCF2) in mESC caused a general weakening of the centromere *Shh*-TAD boundary. As a consequence, the deletion resulted in decreased interactions of the *Shh* promoter with the rest of its TAD. Despite subtle changes, FISH revealed that the proximity of the ZRS to the promoter was comparable to wildtype and no significant changes in *Shh* expression were detectable (Williamson et al., 2019). These results underline the strength of the *Shh*-flanking CTCF site in forming a TAD boundary and its importance for intra-TAD interactions.

In this light, our results suggest alternative long- and short-range effects of this promoter-proximal CTCF site. In the scenario of the *IntraTAD1* allele it rather seems that the 5' CTCF binding site blocks the interaction of the ZRS at the new position with the *Shh* promoter. Here, CTCF does not facilitate enhancer-promoter interactions but presumably insulates by preventing the ZRS

from contacting the *Shh* promoter (Burgess-Beusse et al., 2002). Even though the insulator function led to the discovery of CTCF (Bell et al., 1999), only a minority of CTCF bindings sites is known to act as insulators, yet, several examples exist in the literature (Braccioli and Wit, 2019). Rescue of *Shh* expression, even partially, by a deletion of the promoter-flanking CTCF binding site in the *IntraTAD1* allele would support our hypothesis.

Furthermore, the ZRS is inserted into a *Shh*-floor plate enhancer 2 (SFPE2), which it disrupts. SFPE2 is inactive in the developing limb bud. Although in wildtype E10.5 limb buds SFPE2 lacks repressive histone modifications at the *IntraTAD1* integration site, we cannot exclude spreading of another kind of repression like DNA-methylation by the surrounding inactive enhancer, thus preventing ZRS-function (Smith and Meissner, 2013). It is therefore compelling to test if a slightly shifted integration outside the SFPE2-enhancer would result in a similar phenotype.

6.2.6 Enhancer activity beyond chromatin architecture

The results demonstrate a decreased enhancer-functionality once relocated within the endogenous domain. This suggests that alternative mechanisms, beyond CTCF mediated chromatin architecture, confer enhancer functionality. One such mechanism could be the transcription of enhancers (Arnold et al., 2020). In Wildtype, the ZRS is situated within intron 5 of the housekeeping gene *Lmbr1*, constitutively transcribed in the majority of tissues. A possible explanation for the impeded activity could be the loss of active transcription at the integration site, as deletion of the *Lmbr1* promoter led to a moderate but significant decrease of *Shh* expression in the developing limb bud at E10.5 (Paliou et al., 2019). Several examples in the literature have shown that active transcription of an enhancer can be crucial for its function by correlating RNA PolII pausing and elongation with enhancer activity (Henriques et al., 2018). Moreover, depletion of an elongation factor in murine B-cells disrupted the functional interaction of a super-enhancer and its target gene, and, thus, gene expression (Fitz et al., 2020). In general, constitutive transcription at housekeeping genes is thought to keep the transcriptional apparatus already in place and on standby. Upon activation a rapid interaction and activation of the target promoter could be achieved by the phase separating capacity of RNA PolII and TF bound to the according regulatory elements. The absence of active transcription running constitutively over the ZRS by enhancer-repositioning could be less favorable to form such active hubs (Furlong and Levine, 2018), affecting gene regulation.

How the ZRS activates the *Shh* promoter independent from a preformed contact remains elusive. The residual 40% of *Shh* expression upon relocation in the *IntraTAD* 2 and *3* could possibly be explained by an intrinsic enhancer affinity towards its target promoter. Specific TFs like *Ets* and *HoxD*, which are known to bind the ZRS, have been implicated in bridging of contacts and, in case of the latter, formation of phase separating condensates, mechanism which support

enhancer-promoter interaction (Basu et al., 2020; Lettice et al., 2012, 2017). A loss of the motifs during evolution is responsible for the secondary deprivation of limbs in snakes. However, reintegration of a single *Ets*-motif is capable of rescue and development of limbs, underlining the relevance of these motifs for proper enhancer function beyond chromatin architecture (Kvon et al., 2016; Leal and Cohn, 2016). Thus, TF-mediated activation of a target gene by an enhancer in a developmental context seems to be an important layer of regulation despite CTCF-mediated chromatin architecture. Furthermore, as described previously, TFs have been implicated in the formation of transcription-driving condensates by liquid-liquid-phase separation (Hnisz et al., 2017). It is possible that a high concentration of the transcription apparatus is still nucleated by TF bound to the ZRS, thus leading to gene activation.

The highly conserved location of ZRS throughout species within the constitutively transcribed housekeeping gene *Lmbr1* assumes a high relevance of the position in ensuring complete enhancer function. An intriguing experiment would be a targeted integration of the *Lmbr1*-promoter next to the *IntraTAD* relocation sites to then asses a possible rescue of enhancer activity and thus transcriptional output.

6.2.7 Selectivity of enhancer-driven promoter activation

The relocation of the ZRS in the neighboring TADs of *En2* (*ExtraTAD1*) and *Nom1/Mnx1* (*ExtraTAD2*) led to a loss of *Shh* expression. In addition, we could observe ectopic expression of *En2* and *Cnpy1* in the developing limb bud but not of *Mnx1* nor *Nom1*. Ectopic gene activation has been described in several studies linking the process of enhancer hijacking with gene misexpression. Notably, not all genes react similarly to structural variants (Kraft et al., 2019). This is in contrast to the generally accepted idea that enhancers act promiscuously, meaning that they will activate any promoter in their vicinity. Instead, it appears to be a kind of enhancer-promoter specificity. We could observe a similar promoter selectivity at the *Sox9/Kcnj2*-locus. In the *InvC*-allele the expression of the gene directly neighboring *Kcnj2*, *Kcnj16*, was not affected by the reoriented *Sox9*-enhancers.

In a previous study, we could show that polycomb repressed genes, marked by high levels of H3K27me3, are generally more responsive to ectopic activation (Kraft et al., 2019). The histone modifications for all four genes (*En2*, *Cnpy1*, *Nom1* and *Mnx1*) at the two *ExtraTAD* positions, however, cannot explain the selectivity of the ZRS towards different non-target promoter. Numerous reports show various mechanisms including promoter type (Zabidi et al., 2015), proximity to target promoter, chromatin accessibility and histone modification, which influence the transcriptional response of promoters to enhancers (Kraft et al., 2019); yet, the specificity which restricts the response of a promoter to a specific enhancer remains abstract. Several studies intend to answer this open question by high-throughput genome-wide assays, aiming for

a better understanding of gene regulation and promoter-enhancer-specificity (Trauernicht et al., 2019). Repositioning of promoters usually located in LADs identified distinct classes of promoters by displaying strong variation in reporter expression depending on the integration site. In this context, the sensitivity to a repressive environment and the ability to overcome this determines promoter strength (Leemans et al., 2019), underlining the importance of genomic position for individual promoter function. However, this is not including the required TF composition of a cell type or lack of additional enhancers crucial for promoter activation.

Apart from chromatin context, the intrinsic affinity of core promoters to specific co-activators can dictate transcriptional output, as tested by high-throughput promoter activity assays in drosophila S2 and human HCT116 cells (Haberle et al., 2019). The diversity of core promoters provides an explanation as to why specific promoters reply to a certain set of enhancers.

How these compatibilities transfer to developmental gene regulation *in vivo*, however, needs to be explored and opens an exciting new field of research.

6.3 Conclusion and Outlook

My analysis revealed that enhancer function is restricted via TAD-boundaries and its position within a TAD highly impacts the precision and robustness of enhancer-driven gene activation: The results show that (i) the enhancer needs to be located within its TAD and (ii) that the surrounding chromatin structure and CTCF scaffold might modify enhancer activity. Finally, (iii) by analyzing the activation of the ZRS-neighboring genes upon relocation outside the *Shh*-TAD, enhancers do not act promiscuously but show some sort of selectivity towards non-target promoters.

However, in combination with the results at the *Sox9*/*Kcnj2*-locus, a highly complex picture of gene regulation emerges. Removing the ZRS from its chromatin context showed more drastic effects on the regulation of the cognate gene compared to the SVs at the *Sox9*/*Kcnj2*-locus (see above) (Despang et al., 2019). Inversion of the regulatory substructure and, by this, changing orientation and distance of enhancers to the *Sox9* promoter in the *Inv-Intra* mutant had no effect on *Sox9* expression, despite the formation of new CTCF-mediated chromatin loops (Despang et al., 2019). In contrast, it has been shown that the dynamic chromatin configuration during limb development at the *Pitx1*-locus in mice determines its hindlimb-specific regulation by a pan-enhancer (Kragesteen et al., 2018). Furthermore, inversions of a sub-domain at the *HoxD* cluster including regulatory elements demonstrated the importance of chromatin topology for temporal expression over spatial distribution of the transcripts (Rodríguez-Carballo et al., 2020).

It is therefore possible that the impact of CTCF and the CTCF-mediated structure of a locus have differing effects on transcriptional control and require individual analysis. Our results hint towards distinct roles of CTCF at the *Shh*-locus. With our series of *IntraTAD* integrations we could demonstrate: (i) CTCF as a mediator of proximity to confer evolutionary optimized robustness of

expression compared to (ii) CTCFs function as an insulator at the *IntraTAD1* site, blocking promoter-enhancer interactions. At other loci, however, the influence might be less critical on transcriptional control.

The results achieved by this study mirror the currently conflicting results and discussion in the field about the influence of TADs on gene regulation. The role of CTCF in TAD formation and the generation of TAD boundaries seems clear, where the influence of TAD substructures on enhancer function is far less understood. While we found moderate effect of a lost insulation and depletion of the CTCF-mediated scaffold within the domain at the *Sox9/Kcnj2*-locus, repositioning the ZRS from its loop-anchored topology results in sufficient but decreased expression.

However, the results are not contradictory to the largely accepted idea of TADs as an evolutionary optimized microenvironments facilitating promoter-enhancer contacts to confer robustness and precision to developmental gene regulation. Rather, they demonstrate different effects at different loci.

Moreover, our analysis at the *Shh*-locus demonstrated that enhancers are not functioning at any integration site equally compared to their endogenous position. In addition, at positions outside the *Shh*-TAD the ZRS failed to promiscuously activate any non-target gene in its vicinity, questioning the common paradigm of enhancer function. However, further studies need to verify the obtained results at other loci, where the impact of genomic position might be less critical on transcriptional output.

7. ABSTRACT

Precise spatiotemporal gene expression during embryonic developmental is controlled by cis-regulatory elements (CREs) such as enhancers and promoters. Their physical chromatin proximity is correlated with active transcription and thought to be restricted to topologically associated domains (TADs) that help establish interactions between CREs and limit inappropriate contacts. Accordingly, TADs frequently overlap with gene regulatory landscapes, in which are contained diverse enhancers that transmit their activity across the domain towards their target promoter. Large structural variants reorganizing TADs were shown to cause gene misexpression and disease thereby linking gene regulation to chromatin structure. Recently, several studies revealed controversial results questioning the importance of TADs for transcriptional control. Acute depletion of CTCF and other architectural proteins *in vitro* led to loss of TAD structures with surprisingly modest effects on gene expression. However, the cytotoxicity of such depletion assays hindered analysis of more complex gene regulatory scenarios and their effect during development. This study specifically addresses the connection between TADs and developmental gene regulation through two projects using the murine limb as a model system.

First, we took advantage of the *Sox9/Kcnj2*-locus that is subdivided into two adjacent TADs with distinct expression patterns of *Sox9* and *Kcnj2*. The systematic deletion of individual CTCF binding sites at the TAD boundary and within the TAD resulted in gradual fusion of the neighboring domains without major effects on gene expression. TAD rearrangement by TAD-spanning inversions and repositioning of the boundary, however, redirected the regulatory activity and resulted in pathogenic gene misexpression. Thus, TAD structures may not be essential for developmental gene regulation, yet CTCF-dependent rearrangement of TADs can lead to the redirection of enhancer–promoter contacts and gene misexpression.

In the second project, we studied how enhancer position relative to its TAD influences the function of an individual enhancer at the *Shh*-locus. Therefore, we repositioned the *Shh*-limb enhancer ZRS to five alternative locations inside and outside of its TAD. As expected, the enhancer lost all function in the positions outside of the *Shh*-TAD. Interestingly, the new positions inside the TAD also displayed decreased enhancer activity, albeit to varying degrees. Further analysis suggests that CTCF likely functions in some positions as a facilitator of enhancer-promoter contacts, while insulating short-range contacts in others. Ultimately, the ZRS is only able to ectopically activate some genes if repositioned to novel TADs, displaying strong enhancer-promoter selectivity.

In summary, the results demonstrate that TADs provide robustness and precision to gene regulation, guiding enhancer-promoter interaction without being essential. The findings in this work build a basis for future studies aiming to understand enhancer-promoter interaction and can help in contextualizing potential disease-causing mutations disrupting TADs.

8. ZUSAMMENFASSUNG

Die präzise räumliche und zeitliche Expression von Genen während der Embryonalentwicklung wird durch cis-regulatorische Elemente (CREs) wie Enhancern und Promotoren kontrolliert, deren physikalische Nähe mit aktiver Transkription korreliert wird. Dabei wird angenommen, dass diese Interaktionen auf topologisch assoziierende Domänen (TADs) beschränkt sind, welche helfen funktionale Kontakte zwischen CREs zu etablieren und inadäquate einzugrenzen. Dementsprechend überschneiden sich TADs häufig mit genregulatorischen Landschaften, in denen Enhancer-Aktivität innerhalb der gesamten Domäne in Richtung eines Zielpromoters übertragen wird. Es konnte gezeigt werden, dass große strukturelle Varianten TADs reorganisieren können, was zu einer Fehlexpression von Genen und Krankheiten führt und so Genregulation mit Chromatinstruktur in Verbindung bringt. Jedoch zeigten jüngere Studien kontroverse Ergebnisse, welche die Bedeutung von TADs für die Transkriptionskontrolle in Frage stellen. So führte die akute Depletion von CTCF und anderen Architekturproteinen *in vitro* zum Verlust von TAD-Strukturen, hatte aber überraschenderweise nur geringe Auswirkungen auf die Genexpression. Die Zytotoxizität dieser Experimente schränkt jedoch die Analyse komplexerer genregulatorischer Szenarien ein. Diese Studie befasst sich speziell mit dem Zusammenhang zwischen TADs und entwicklungsbedingter Genregulation in zwei Projekten, welche die Gliedmaßenentwicklung in Mäusen als Modellsystem verwenden.

Zunächst nutzten wir den *Sox9/Kcnj2*-Lokus, der in zwei benachbarte TADs mit unterschiedlichen Expressionsmustern von *Sox9* und *Kcnj2* unterteilt ist. Die systematische Deletion einzelner CTCF-Bindungsstellen an der TAD-Grenze und innerhalb der TADs führte zu einer schrittweisen Fusion der benachbarten Domänen ohne größere Auswirkungen auf die Genexpression. Eine TAD-Umstrukturierung durch TAD-übergreifende Inversionen und eine Neupositionierung der TAD-Grenze lenkte jedoch die regulatorische Aktivität um und führte zu einer pathogenen Gen-Fehl-Expression. TAD-Strukturen scheinen also nicht essentiell für die entwicklungsbedingte Genregulation zu sein, jedoch kann eine CTCF-abhängige Neuanordnung von TADs zu einer Umlenkung der Enhancer-Promotor-Kontakte und zu Gen-Misexpression führen.

Im zweiten Teil untersuchten wir den Einfluss der genomische Position eines einzelnen Enhancers relativ zu seiner TAD am *Shh*-Lokus. Dazu wurde der *Shh*-Gliedmaßen-Enhancer ZRS an fünf alternative Positionen innerhalb und außerhalb seiner TAD neu positioniert. Wie erwartet, verlor der Enhancer jegliche Funktion an den Positionen außerhalb der *Shh*-TAD. Interessanterweise zeigten auch die neuen Positionen innerhalb der TAD einen Verlust der Enhancer-Aktivität, wenn auch in unterschiedlichem Ausmaß. Die Analyse letzterer deutet darauf hin, dass CTCF

wahrscheinlich an einigen Positionen als Vermittler von Enhancer-Promotor-Kontakten fungiert, während es an anderen Positionen Kontakte mit kurzer Reichweite isoliert. Darüber hinaus wurden durch die ZRS nur einige Gene ektopisch aktiviert, wenn sie in benachbarte TADs positioniert wurde, was eine starke Enhancer-Promotor-Selektivität zeigt.

Zusammenfassend zeigen die Ergebnisse dieser Arbeit, dass TADs Robustheit und Präzision für die Genregulation vermitteln, indem sie helfen Enhancer-Promotor-Interaktion zu etablieren, ohne dafür essentiell zu sein. Dies bildet eine Grundlage für zukünftige Studien, die darauf abzielen, Enhancer-Promoter-Interaktionen zu verstehen, welche helfen könnten potenziell krankheitsverursachende Mutationen die TADs betreffen zu interpretieren.

9. REFERENCES

Akiyama, H., Chaboissier, M.-C., Martin, J.F., Schedl, A., and Crombrugghe, B. de (2002). The transcription factor Sox9 has essential roles in successive steps of the chondrocyte differentiation pathway and is required for expression of Sox5 and Sox6. Gene Dev *16*, 2813–2828.

Akiyama‡, H., Kim‡, J.-E., Nakashima, K., Balmes, G., Iwai, N., Deng, J.M., Zhang, Z., Martin, J.F., Behringer, R.R., Nakamura, T., et al. (2005). Osteo-chondroprogenitor cells are derived from Sox9 expressing precursors. P Natl Acad Sci Usa *102*, 14665–14670.

Alexander, J.M., Guan, J., Li, B., Maliskova, L., Song, M., Shen, Y., Huang, B., Lomvardas, S., and Weiner, O.D. (2019). Live-cell imaging reveals enhancer-dependent Sox2 transcription in the absence of enhancer proximity. Elife *8*, e41769.

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

Andrey, G., and Spielmann, M. (2016). Enhancer RNAs, Methods and Protocols. Methods Mol Biology Clifton N J *1468*, 221–234.

Andrey, G., Schöpflin, R., Jerković, I., Heinrich, V., Ibrahim, D.M., Paliou, C., Hochradel, M., Timmermann, B., Haas, S., Vingron, M., et al. (2016). Characterization of hundreds of regulatory landscapes in developing limbs reveals two regimes of chromatin folding. Genome Res *27*, 223–233.

Arnold, P.R., Wells, A.D., and Li, X.C. (2020). Diversity and Emerging Roles of Enhancer RNA in Regulation of Gene Expression and Cell Fate. Frontiers Cell Dev Biology *7*, 377.

Artus, J., and Hadjantonakis, A.-K. (2010). Transgenic Mouse Methods and Protocols. Methods Mol Biology Clifton N J *693*, 37–56.

Banerji, J., Rusconi, S., and Schaffner, W. (1981). Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. Cell *27*, 299–308.

Basu, S., Mackowiak, S.D., Niskanen, H., Knezevic, D., Asimi, V., Grosswendt, S., Geertsema, H., Ali, S., Jerković, I., Ewers, H., et al. (2020). Unblending of Transcriptional Condensates in Human Repeat Expansion Disease. Cell *181*, 1062-1079.e30.

Beagan, J.A., Duong, M.T., Titus, K.R., Zhou, L., Cao, Z., Ma, J., Lachanski, C.V., Gillis, D.R., and Phillips-Cremins, J.E. (2017). YY1 and CTCF orchestrate a 3D chromatin looping switch during early neural lineage commitment. Genome Res *27*, 1139–1152.

Behringer, B.R., Texas, U. of, Gertsenstein, M.D.A.C.C.M., Phenogenomics, T.C. for, Nagy, T.C.K.V., Institute, S.L.R., Hospital, M.S., Nagy, T.A., Institute, S.L.R., Hospital, M.S., et al. (1994). Manipulating the Mouse Embryo: A Laboratory Manual, Fourth Edition (Cold Spring Harbor Laboratory Press).

Bell, A.C., West, A.G., and Felsenfeld, G. (1999). The Protein CTCF Is Required for the Enhancer Blocking Activity of Vertebrate Insulators. Cell *98*, 387–396.

Bell, D.M., Leung, K.K.H., Wheatley, S.C., Ng, L.J., Zhou, S., Ling, K.W., Sham, M.H., Koopman, P., Tam, P.P.L., and Cheah, K.S.E. (1997). SOX9 directly regulates the type-II collagen gene. Nat Genet *16*, 174–178.

Bemmel, J.G. van, Galupa, R., Gard, C., Servant, N., Picard, C., Davies, J., Szempruch, A.J., Zhan, Y., Żylicz, J.J., Nora, E.P., et al. (2019). The bipartite TAD organization of the X-inactivation center ensures opposing developmental regulation of Tsix and Xist. Nat Genet *51*, 1024–1034.

Benabdallah, N.S., Williamson, I., Illingworth, R.S., Kane, L., Boyle, S., Sengupta, D., Grimes, G.R., Therizols, P., and Bickmore, W.A. (2019). Decreased Enhancer-Promoter Proximity Accompanying Enhancer Activation. Mol Cell *76*, 473-484.e7.

Bénazet, J.-D., Bischofberger, M., Tiecke, E., Gonçalves, A., Martin, J.F., Zuniga, A., Naef, F., and Zeller, R. (2009). A Self-Regulatory System of Interlinked Signaling Feedback Loops Controls Mouse Limb Patterning. Science *323*, 1050–1053.

Bernstein, B.E., and Kellis, M. (2005). Large-scale discovery and validation of functional elements in the human genome. Genome Biol *6*, 312.

Bi, W., Huang, W., Whitworth, D.J., Deng, J.M., Zhang, Z., Behringer, R.R., and Crombrugghe, B. de (2001). Haploinsufficiency of Sox9 results in defective cartilage primordia and premature skeletal mineralization. Proc National Acad Sci *98*, 6698–6703.

Bintu, B., Mateo, L.J., Su, J.-H., Sinnott-Armstrong, N.A., Parker, M., Kinrot, S., Yamaya, K., Boettiger, A.N., and Zhuang, X. (2018). Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. Science *362*, eaau1783.

Boija, A., Klein, I.A., Sabari, B.R., Dall'Agnese, A., Coffey, E.L., Zamudio, A.V., Li, C.H., Shrinivas, K., Manteiga, J.C., Hannett, N.M., et al. (2018). Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. Cell *175*, 1842-1855.e16.

Bonev, B., Cohen, N.M., Szabo, Q., Fritsch, L., Papadopoulos, G.L., Lubling, Y., Xu, X., Lv, X., Hugnot, J.-P., Tanay, A., et al. (2017). Multiscale 3D Genome Rewiring during Mouse Neural Development. Cell *171*, 557-572.e24.

Braccioli, L., and Wit, E. de (2019). CTCF: a Swiss-army knife for genome organization and transcription regulation. Essays Biochem *63*, 157–165.

Bulger, M., and Groudine, M. (2011). Functional and Mechanistic Diversity of Distal Transcription Enhancers. Cell *144*, 825.

Burgess-Beusse, B., Farrell, C., Gaszner, M., Litt, M., Mutskov, V., Recillas-Targa, F., Simpson, M., West, A., and Felsenfeld, G. (2002). The insulation of genes from external enhancers and silencing chromatin. Proc National Acad Sci *99*, 16433–16437.

Busslinger, G.A., Stocsits, R.R., Lelij, P. van der, Axelsson, E., Tedeschi, A., Galjart, N., and Peters, J.-M. (2017). Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl. Nature *544*, 503–507.

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

Chen, H., Levo, M., Barinov, L., Fujioka, M., Jaynes, J.B., and Gregor, T. (2018). Dynamic interplay between enhancer–promoter topology and gene activity. Nat Genet *50*, 1296–1303.

Cho, W.-K., Spille, J.-H., Hecht, M., Lee, C., Li, C., Grube, V., and Cisse, I.I. (2018). Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. Science *361*, eaar4199.

Cooks, R.G., Hertz, M., Katznelson, M.B.M., and Goodman, R.M. (1985). A new nail dysplasia syndrome with onychonychia and absence and/or hypoplasia of distal phalanges. Clin Genet *27*, 85–91.

Cremer, T., and Cremer, C. (2001). Chromosome territories, nuclear architecture and gene regulation in mammalian cells. Nat Rev Genet *2*, 292–301.

Cuddapah, S., Jothi, R., Schones, D.E., Roh, T.-Y., Cui, K., and Zhao, K. (2009). Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. Genome Res *19*, 24–32.

Deaton, A.M., and Bird, A. (2011). CpG islands and the regulation of transcription. Gene Dev *25*, 1010–1022.

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing Chromosome Conformation. Science *295*, 1306–1311.

Dekker, J., Marti-Renom, M.A., and Mirny, L.A. (2013). Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. Nat Rev Genet *14*, 390–403.

Deng, W., Rupon, J.W., Krivega, I., Breda, L., Motta, I., Jahn, K.S., Reik, A., Gregory, P.D., Rivella, S., Dean, A., et al. (2014). Reactivation of Developmentally Silenced Globin Genes by Forced Chromatin Looping. Cell *158*, 849–860.

Despang, A., Schöpflin, R., Franke, M., Ali, S., Jerković, I., Paliou, C., Chan, W.-L., Timmermann, B., Wittler, L., Vingron, M., et al. (2019). Functional dissection of the Sox9-Kcnj2 locus identifies nonessential and instructive roles of TAD architecture. Nat Genet *51*, 1263–1271.

de Wit, E., Vos, E.S.M., Holwerda, S.J.B., Valdes-Quezada, C., Verstegen, M.J.A.M., Teunissen, H., Splinter, E., Wijchers, P.J., Krijger, P.H.L., and de Laat, W. (2015). CTCF Binding Polarity Determines Chromatin Looping. Mol Cell *60*, 676–684.

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

Doudna, J.A., and Charpentier, E. (2014). The new frontier of genome engineering with CRISPR-Cas9. Science *346*, 1258096.

Duprez, D.M., Coltey, M., Amthor, H., Brickell, P.M., and Tickle, C. (1996). Bone Morphogenetic Protein-2 (BMP-2) Inhibits Muscle Development and Promotes Cartilage Formation in Chick Limb Bud Cultures. Dev Biol *174*, 448–452.

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

Entrevan, M., Schuettengruber, B., and Cavalli, G. (2016). Regulation of Genome Architecture and Function by Polycomb Proteins. Trends Cell Biol *26*, 511–525.

Ernst, J., and Kellis, M. (2017). Chromatin-state discovery and genome annotation with ChromHMM. Nat Protoc *12*, 2478–2492.

Fitz, J., Neumann, T., Steininger, M., Wiedemann, E.-M., Garcia, A.C., Athanasiadis, A., Schoeberl, U.E., and Pavri, R. (2020). Spt5-mediated enhancer transcription directly couples enhancer activation with physical promoter interaction. Nat Genet *52*, 505–515.

Flavahan, W.A., Drier, Y., Liau, B.B., Gillespie, S.M., Venteicher, A.S., Stemmer-Rachamimov, A.O., Suvà, M.L., and Bernstein, B.E. (2016). Insulator dysfunction and oncogene activation in IDH mutant gliomas. Nature *529*, 110–114.

Franke, M. (2017). The Role of Higher-Order Chromatin Organization at the SOX9 Locus in Gene Regulation and Disease.

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

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

Furlong, E.E.M., and Levine, M. (2018). Developmental enhancers and chromosome topology. Science *361*, 1341–1345.

Gassler, J., Brandão, H.B., Imakaev, M., Flyamer, I.M., Ladstätter, S., Bickmore, W.A., Peters, J., Mirny, L.A., and Tachibana, K. (2017). A mechanism of cohesin-dependent loop extrusion organizes zygotic genome architecture. Embo J *36*, 3600–3618.

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

Ghavi-Helm, Y., Klein, F.A., Pakozdi, T., Ciglar, L., Noordermeer, D., Huber, W., and Furlong, E.E.M. (2014). Enhancer loops appear stable during development and are associated with paused polymerase. Nature *512*, 96–100.

Ghavi-Helm, Y., Jankowski, A., Meiers, S., Viales, R.R., Korbel, J.O., and Furlong, E.E.M. (2019). Highly rearranged chromosomes reveal uncoupling between genome topology and gene expression. Nat Genet *51*, 1272–1282.

Gibson, D.G., Benders, G.A., Axelrod, K.C., Zaveri, J., Algire, M.A., Moodie, M., Montague, M.G., Venter, J.C., Smith, H.O., and Hutchison, C.A. (2008). One-step assembly in yeast of 25

overlapping DNA fragments to form a complete synthetic Mycoplasma genitalium genome. Proc National Acad Sci *105*, 20404–20409.

Golfier, S., Quail, T., Kimura, H., and Brugués, J. (2020). Cohesin and condensin extrude DNA loops in a cell-cycle dependent manner. Elife *9*, e53885.

Green, M.R., and Sambrook, J. (2012). Molecular Cloning: A loboratory manual (Cold Spring Harbor Laboratory Press).

Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M.B., Talhout, W., Eussen, B.H., Klein, A. de, Wessels, L., Laat, W. de, et al. (2008). Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. Nature *453*, 948–951.

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

Haaf, T., and Schmid, M. (1991). Chromosome topology in mammalian interphase nuclei. Exp Cell Res *192*, 325–332.

Haarhuis, J.H.I., Weide, R.H. van der, Blomen, V.A., Yáñez-Cuna, J.O., Amendola, M., Ruiten, M.S. van, Krijger, P.H.L., Teunissen, H., Medema, R.H., Steensel, B. van, et al. (2017). The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension. Cell *169*, 693-707.e14.

Haberle, V., and Stark, A. (2018). Eukaryotic core promoters and the functional basis of transcription initiation. Nat Rev Mol Cell Bio *19*, 621–637.

Haberle, V., Arnold, C.D., Pagani, M., Rath, M., Schernhuber, K., and Stark, A. (2019). Transcriptional cofactors display specificity for distinct types of core promoters. Nature *570*, 122–126.

Haeussler, M., Schönig, K., Eckert, H., Eschstruth, A., Mianné, J., Renaud, J.-B., Schneider-Maunoury, S., Shkumatava, A., Teboul, L., Kent, J., et al. (2016). Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. Genome Biol *17*, 148.

Hampsey, M. (1998). Molecular genetics of the RNA polymerase II general transcriptional machinery. Microbiol Mol Biology Rev Mmbr *62*, 465–503.

Hansen, A.S., Hsieh, T.-H.S., Cattoglio, C., Pustova, I., Saldaña-Meyer, R., Reinberg, D., Darzacq, X., and Tjian, R. (2019). Distinct Classes of Chromatin Loops Revealed by Deletion of an RNA-Binding Region in CTCF. Mol Cell *76*, 395-411.e13.

Hecht, J., Stricker, S., Wiecha, U., Stiege, A., Panopoulou, G., Podsiadlowski, L., Poustka, A.J., Dieterich, C., Ehrich, S., Suvorova, J., et al. (2008). Evolution of a Core Gene Network for Skeletogenesis in Chordates. Plos Genet *4*, e1000025.

Heinz, S., Texari, L., Hayes, M.G.B., Urbanowski, M., Chang, M.W., Givarkes, N., Rialdi, A., White, K.M., Albrecht, R.A., Pache, L., et al. (2018). Transcription Elongation Can Affect Genome 3D Structure. Cell *174*, 1522-1536.e22.

Henriques, T., Scruggs, B.S., Inouye, M.O., Muse, G.W., Williams, L.H., Burkholder, A.B., Lavender, C.A., Fargo, D.C., and Adelman, K. (2018). Widespread transcriptional pausing and elongation control at enhancers. Gene Dev *32*, 26–41.

Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-André, V., Sigova, A.A., Hoke, H.A., and Young, R.A. (2013). Super-Enhancers in the Control of Cell Identity and Disease. Cell *155*, 934–947.

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

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

Hsieh, T.-H.S., Fudenberg, G., Goloborodko, A., and Rando, O.J. (2016). Micro-C XL: assaying chromosome conformation from the nucleosome to the entire genome. Nat Methods *13*, 1009–1011.

Huang, Y., Neijts, R., and Laat, W. (2020). How chromosome topologies get their shape: views from proximity ligation and microscopy methods. Febs Lett *594*, 3439–3449.

Jayavelu, N.D., Jajodia, A., Mishra, A., and Hawkins, R.D. (2020). Candidate silencer elements for the human and mouse genomes. Nat Commun *11*, 1061.

Jerković, I., Ibrahim, D.M., Andrey, G., Haas, S., Hansen, P., Janetzki, C., Navarrete, I.G., Robinson, P.N., Hecht, J., and Mundlos, S. (2016). Genome-wide binding of posterior HOXA/D transcription factors reveals subgrouping and association with CTCF. Biorxiv 073593.

Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., Berkum, N.L. van, Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S., et al. (2010). Mediator and cohesin connect gene expression and chromatin architecture. Nature *467*, 430–435.

Kim, T.-K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., et al. (2010). Widespread transcription at neuronal activity-regulated enhancers. Nature *465*, 182–187.

Kim, Y., Shi, Z., Zhang, H., Finkelstein, I.J., and Yu, H. (2019). Human cohesin compacts DNA by loop extrusion. Science *366*, 1345–1349.

Klenova, E.M., Chernukhin, I.V., El-Kady, A., Lee, R.E., Pugacheva, E.M., Loukinov, D.I., Goodwin, G.H., Delgado, D., Filippova, G.N., León, J., et al. (2001). Functional Phosphorylation Sites in the C-Terminal Region of the Multivalent Multifunctional Transcriptional Factor CTCF. Mol Cell Biol *21*, 2221–2234.

Klopocki, E., and Mundlos, S. (2011). Copy-Number Variations, Noncoding Sequences, and Human Phenotypes. Annu Rev Genom Hum G *12*, 53–72.

Knight, P.A., and Ruiz, D. (2012). A fast algorithm for matrix balancing. Ima J Numer Anal *33*, 1029–1047.

Kouzarides, T. (2007). Chromatin Modifications and Their Function. Cell *128*, 693–705.

Kraft, K., Geuer, S., Will, A.J., Chan, W.L., Paliou, C., Borschiwer, M., Harabula, I., Wittler, L., Franke, M., Ibrahim, D.M., et al. (2015). Deletions, Inversions, Duplications: Engineering of Structural Variants using CRISPR/Cas in Mice. Cell Reports *10*, 833–839.

Kraft, K., Magg, A., Heinrich, V., Riemenschneider, C., Schöpflin, R., Markowski, J., Ibrahim, D.M., Acuna-Hidalgo, R., Despang, A., Andrey, G., et al. (2019). Serial genomic inversions induce tissue-specific architectural stripes, gene misexpression and congenital malformations. Nat Cell Biol *21*, 305–310.

Kragesteen, B.K., Spielmann, M., Paliou, C., Heinrich, V., Schöpflin, R., Esposito, A., Annunziatella, C., Bianco, S., Chiariello, A.M., Jerković, I., et al. (2018). Dynamic 3D chromatin architecture contributes to enhancer specificity and limb morphogenesis. Nat Genet *50*, 1463–1473.

Krebs, O., Schreiner, C.M., Scott, W.J., Bell, S.M., Robbins, D.J., Goetz, J.A., Alt, H., Hawes, N., Wolf, E., and Favor, J. (2003). Replicated anterior zeugopod (raz): a polydactylous mouse mutant with lowered Shh signaling in the limb bud. Development *130*, 6037–6047.

Krietenstein, N., Abraham, S., Venev, S.V., Abdennur, N., Gibcus, J., Hsieh, T.-H.S., Parsi, K.M., Yang, L., Maehr, R., Mirny, L.A., et al. (2020). Ultrastructural Details of Mammalian Chromosome Architecture. Mol Cell *78*, 554-565.e7.

Kubo, N., Ishii, H., Xiong, X., Bianco, S., Meitinger, F., Hu, R., Hocker, J.D., Conte, M., Gorkin, D., Yu, M., et al. (2021). Promoter-proximal CTCF binding promotes distal enhancer-dependent gene activation. Nat Struct Mol Biol 1–10.

Kvon, E.Z. (2015). Using transgenic reporter assays to functionally characterize enhancers in animals. Genomics *106*, 185–192.

Kvon, E.Z., Kamneva, O.K., Melo, U.S., Barozzi, I., Osterwalder, M., Mannion, B.J., Tissières, V., Pickle, C.S., Plajzer-Frick, I., Lee, E.A., et al. (2016). Progressive Loss of Function in a Limb Enhancer during Snake Evolution. Cell *167*, 633-642.e11.

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

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860–921.

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

Leal, F., and Cohn, M.J. (2016). Loss and Re-emergence of Legs in Snakes by Modular Evolution of Sonic hedgehog and HOXD Enhancers. Curr Biol *26*, 2966–2973.

Lee, J., Krivega, I., Dale, R.K., and Dean, A. (2017). The LDB1 Complex Co-opts CTCF for Erythroid Lineage-Specific Long-Range Enhancer Interactions. Cell Reports *19*, 2490–2502.

Leemans, C., Zwalm, M.C.H. van der, Brueckner, L., Comoglio, F., Schaik, T. van, Pagie, L., Arensbergen, J. van, and Steensel, B. van (2019). Promoter-Intrinsic and Local Chromatin Features Determine Gene Repression in LADs. Cell *177*, 852-864.e14.

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

Lettice, L.A., Williamson, I., Wiltshire, J.H., Peluso, S., Devenney, P.S., Hill, A.E., Essafi, A., Hagman, J., Mort, R., Grimes, G., et al. (2012). Opposing Functions of the ETS Factor Family Define Shh Spatial Expression in Limb Buds and Underlie Polydactyly. Dev Cell *22*, 459–467.

Lettice, L.A., Williamson, I., Devenney, P.S., Kilanowski, F., Dorin, J., and Hill, R.E. (2014). Development of five digits is controlled by a bipartite long-range cis-regulator. Development *141*, 1715–1725.

Lettice, L.A., Devenney, P., Angelis, C.D., and Hill, R.E. (2017). The Conserved Sonic Hedgehog Limb Enhancer Consists of Discrete Functional Elements that Regulate Precise Spatial Expression. Cell Reports *20*, 1396–1408.

Li, W., Notani, D., and Rosenfeld, M.G. (2016). Enhancers as non-coding RNA transcription units: recent insights and future perspectives. Nat Rev Genet *17*, 207–223.

Lichter, P., Cremer, T., Borden, J., Manuelidis, L., and Ward, D.C. (1988). Delineation of individual human chromosomes in metaphase and interphase cells by in situ suppression hybridization using recombinant DNA libraries. Hum Genet *80*, 224–234.

Lieberman-Aiden, E., Berkum, N.L. van, Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. Science *326*, 289–293.

Long, H.K., Osterwalder, M., Welsh, I.C., Hansen, K., Davies, J.O.J., Liu, Y.E., Koska, M., Adams, A.T., Aho, R., Arora, N., et al. (2020). Loss of Extreme Long-Range Enhancers in Human Neural Crest Drives a Craniofacial Disorder. Cell Stem Cell.

Loomis, C.A., Harris, E., Michaud, J., Wurst, W., Hanks, M., and Joyner, A.L. (1996). The mouse Engrailed-1 gene and ventral limb patterning. Nature *382*, 360–363.

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

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

Maeso, I., Irimia, M., Tena, J.J., Casares, F., and Gómez-Skarmeta, J.L. (2013). Deep conservation of cis-regulatory elements in metazoans. Philosophical Transactions Royal Soc B Biological Sci *368*, 20130020.

Malik, S., and Roeder, R.G. (2005). Dynamic regulation of pol II transcription by the mammalian Mediator complex. Trends Biochem Sci *30*, 256–263.

Matthews, N.E., and White, R. (2019). Chromatin Architecture in the Fly: Living without CTCF/Cohesin Loop Extrusion? Bioessays *41*, 1900048.

Melo, U.S., Schöpflin, R., Acuna-Hidalgo, R., Mensah, M.A., Fischer-Zirnsak, B., Holtgrewe, M., Klever, M.-K., Türkmen, S., Heinrich, V., Pluym, I.D., et al. (2020). Hi-C Identifies Complex Genomic Rearrangements and TAD-Shuffling in Developmental Diseases. Am J Hum Genetics *106*, 872–884.

Mitchell, P., and Tjian, R. (1989). Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science *245*, 371–378.

Moreau, P., Hen, R., Wasylyk, B., Everett, R., Gaub, M.P., and Chambon, P. (1981). The SV40 72 base repair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. Nucleic Acids Res *9*, 6047–6068.

Mumbach, M.R., Rubin, A.J., Flynn, R.A., Dai, C., Khavari, P.A., Greenleaf, W.J., and Chang, H.Y. (2016). HiChIP: efficient and sensitive analysis of protein-directed genome architecture. Nat Methods *13*, 919–922.

Nagy, and Nichols, J. (2011). Derivation of murine ES cell lines. In S. Pease & T. Saunders (Eds.), Advanced Protocols for Animal Transgenesis. (Heidelberg: Springer-Verlag), pp. 431–455.

Nagy, A., Nagy, K., and Gertsenstein, M. (2010). Chapter 8 Production of Mouse Chimeras by Aggregating Pluripotent Stem Cells with Embryos. Methods Enzymol *476*, 123–149.

Narendra, V., Rocha, P.P., An, D., Raviram, R., Skok, J.A., Mazzoni, E.O., and Reinberg, D. (2015). CTCF establishes discrete functional chromatin domains at the *Hox* clusters during differentiation. Science *347*, 1017–1021.

Narendra, V., Bulajić, M., Dekker, J., Mazzoni, E.O., and Reinberg, D. (2016). CTCF-mediated topological boundaries during development foster appropriate gene regulation. Gene Dev *30*, 2657–2662.

Nasmyth, K., and Haering, C.H. (2009). Cohesin: Its Roles and Mechanisms. Genetics *43*, 525–558.

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

Nora, E.P., Goloborodko, A., Valton, A.-L., Gibcus, J.H., Uebersohn, A., Abdennur, N., Dekker, J., Mirny, L.A., and Bruneau, B.G. (2017). Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. Cell *169*, 930-944.e22.

Nuebler, J., Fudenberg, G., Imakaev, M., Abdennur, N., and Mirny, L. (2018). Chromatin Organization by an Interplay of Loop Extrusion and Compartmental Segregation. Biophys J *114*, 30a.

Osterwalder, M., Speziale, D., Shoukry, M., Mohan, R., Ivanek, R., Kohler, M., Beisel, C., Wen, X., Scales, S.J., Christoffels, V.M., et al. (2014). HAND2 targets define a network of transcriptional regulators that compartmentalize the early limb bud mesenchyme. Dev Cell *31*, 345–357.

Paliou, C., Guckelberger, P., Schöpflin, R., Heinrich, V., Esposito, A., Chiariello, A.M., Bianco, S., Annunziatella, C., Helmuth, J., Haas, S., et al. (2019). Preformed chromatin topology assists

transcriptional robustness of Shh during limb development. Proc National Acad Sci *116*, 12390–12399.

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

Phillips-Cremins, J.E., Sauria, M.E.G., Sanyal, A., Gerasimova, T.I., Lajoie, B.R., Bell, J.S.K., Ong, C.-T., Hookway, T.A., Guo, C., Sun, Y., et al. (2013). Architectural Protein Subclasses Shape 3D Organization of Genomes during Lineage Commitment. Cell *153*, 1281–1295.

Pinkel, D., Landegent, J., Collins, C., Fuscoe, J., Segraves, R., Lucas, J., and Gray, J. (1988). Fluorescence in situ hybridization with human chromosome-specific libraries: detection of trisomy 21 and translocations of chromosome 4. Proc National Acad Sci *85*, 9138–9142.

Ramisch, A., Heinrich, V., Glaser, L.V., Fuchs, A., Yang, X., Benner, P., Schöpflin, R., Li, N., Kinkley, S., Römer-Hillmann, A., et al. (2019). CRUP: a comprehensive framework to predict condition-specific regulatory units. Genome Biol *20*, 227.

Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., et al. (2014). A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. Cell *159*, 1665–1680.

Rao, S.S.P., Huang, S.-C., Hilaire, B.G.S., Engreitz, J.M., Perez, E.M., Kieffer-Kwon, K.-R., Sanborn, A.L., Johnstone, S.E., Bascom, G.D., Bochkov, I.D., et al. (2017). Cohesin Loss Eliminates All Loop Domains. Cell *171*, 305-320.e24.

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

Robertson, E.J. (1987). Teratocarcinomas and embryonic stem cells: A practical approach. Trends in Genetics (Trends In Genetics).

Robson, M.I., Ringel, A.R., and Mundlos, S. (2019). Regulatory Landscaping: How Enhancer-Promoter Communication Is Sculpted in 3D. Mol Cell *74*, 1110–1122.

Rodríguez-Carballo, E., Lopez-Delisle, L., Zhan, Y., Fabre, P.J., Beccari, L., El-Idrissi, I., Huynh, T.H.N., Ozadam, H., Dekker, J., and Duboule, D. (2017). The HoxD cluster is a dynamic and resilient TAD boundary controlling the segregation of antagonistic regulatory landscapes. Gene Dev *31*, 2264–2281.

Rodríguez-Carballo, E., Lopez-Delisle, L., Willemin, A., Beccari, L., Gitto, S., Mascrez, B., and Duboule, D. (2020). Chromatin topology and the timing of enhancer function at the HoxD locus. Proc National Acad Sci 202015083.

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

Sabari, B.R., Dall'Agnese, A., Boija, A., Klein, I.A., Coffey, E.L., Shrinivas, K., Abraham, B.J., Hannett, N.M., Zamudio, A.V., Manteiga, J.C., et al. (2018). Coactivator condensation at super-enhancers links phase separation and gene control. Science *361*, eaar3958.

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

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

Sagai, T., Amano, T., Tamura, M., Mizushina, Y., Sumiyama, K., and Shiroishi, T. (2009). A cluster of three long-range enhancers directs regional Shh expression in the epithelial linings. Development *136*, 1665–1674.

Sagai, T., Amano, T., Maeno, A., Ajima, R., and Shiroishi, T. (2019). SHH signaling mediated by a prechordal and brain enhancer controls forebrain organization. Proc National Acad Sci *116*, 23636–23642.

Sanborn, A.L., Rao, S.S.P., Huang, S.-C., Durand, N.C., Huntley, M.H., Jewett, A.I., Bochkov, I.D., Chinnappan, D., Cutkosky, A., Li, J., et al. (2015). Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. Proc National Acad Sci *112*, E6456–E6465.

Santa, F.D., Barozzi, I., Mietton, F., Ghisletti, S., Polletti, S., Tusi, B.K., Muller, H., Ragoussis, J., Wei, C.-L., and Natoli, G. (2010). A Large Fraction of Extragenic RNA Pol II Transcription Sites Overlap Enhancers. Plos Biol *8*, e1000384.

Saxonov, S., Berg, P., and Brutlag, D.L. (2006). A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. P Natl Acad Sci Usa *103*, 1412–1417.

Schwartz, Y.B., and Pirrotta, V. (2007). Polycomb silencing mechanisms and the management of genomic programmes. Nat Rev Genet *8*, 9–22.

Schwarzer, W., Abdennur, N., Goloborodko, A., Pekowska, A., Fudenberg, G., Loe-Mie, Y., Fonseca, N.A., Huber, W., Haering, C.H., Mirny, L., et al. (2017). Two independent modes of chromatin organization revealed by cohesin removal. Nature *551*, 51–56.

Seo, H., Amano, T., Seki, R., Sagai, T., Kim, J., Cho, S.W., and Shiroishi, T. (2018). Upstream Enhancer Elements of Shh Regulate Oral and Dental Patterning. J Dent Res *97*, 1055–1063.

Smith, Z.D., and Meissner, A. (2013). DNA methylation: roles in mammalian development. Nat Rev Genet *14*, 204–220.

Spielmann, M., Lupiáñez, D.G., and Mundlos, S. (2018). Structural variation in the 3D genome. Nat Rev Genet *19*, 453–467.

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

Steensel, B. van, and Belmont, A.S. (2017). Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. Cell *169*, 780–791.

Stevens, T.J., Lando, D., Basu, S., Atkinson, L.P., Cao, Y., Lee, S.F., Leeb, M., Wohlfahrt, K.J., Boucher, W., O'Shaughnessy-Kirwan, A., et al. (2017). 3D structures of individual mammalian genomes studied by single-cell Hi-C. Nature *544*, 59–64.

Stik, G., Vidal, E., Barrero, M., Cuartero, S., Vila-Casadesús, M., Mendieta-Esteban, J., Tian, T.V., Choi, J., Berenguer, C., Abad, A., et al. (2020). CTCF is dispensable for immune cell transdifferentiation but facilitates an acute inflammatory response. Nat Genet *52*, 655–661.

Stricker, S., Fundele, R., Vortkamp, A., and Mundlos, S. (2002). Role of Runx Genes in Chondrocyte Differentiation. Dev Biol *245*, 95–108.

Symmons, O., and Spitz, F. (2013). From remote enhancers to gene regulation: charting the genome's regulatory landscapes. Philosophical Transactions Royal Soc B Biological Sci *368*, 20120358.

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

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

Szabo, Q., Donjon, A., Jerković, I., Papadopoulos, G.L., Cheutin, T., Bonev, B., Nora, E.P., Bruneau, B.G., Bantignies, F., and Cavalli, G. (2020). Regulation of single-cell genome organization into TADs and chromatin nanodomains. Nat Genet 1–7.

Taher, L., Collette, N.M., Murugesh, D., Maxwell, E., Ovcharenko, I., and Loots, G.G. (2011). Global Gene Expression Analysis of Murine Limb Development. Plos One *6*, e28358.

Tickle, C. (1981). The number of polarizing region cells required to specify additional digits in the developing chick wing. Nature *289*, 295–298.

Trauernicht, M., Martinez-Ara, M., and Steensel, B. van (2019). Deciphering Gene Regulation Using Massively Parallel Reporter Assays. Trends Biochem Sci *45*, 90–91.

Vavouri, T., and Lehner, B. (2009). Conserved noncoding elements and the evolution of animal body plans. Bioessays *31*, 727–735.

Visel, A., Minovitsky, S., Dubchak, I., and Pennacchio, L.A. (2007). VISTA Enhancer Browser—a database of tissue-specific human enhancers. Nucleic Acids Res *35*, D88–D92.

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

Wagner, T., Wirth, J., Meyer, J., Zabel, B., Held, M., Zimmer, J., Pasantes, J., Bricarelli, F.D., Keutel, J., Hustert, E., et al. (1994). Autosomal sex reversal and campomelic dysplasia are caused by mutations in and around the SRY-related gene SOX9. Cell *79*, 1111–1120.

Wassarman, P.M., and Soriano, P.M. (2010). Guide to Techniques in Mouse Development, Part A: Mice, Embryos, and Cells, 2nd Edition. (Science Direct), pp. 1–482.

Weintraub, A.S., Li, C.H., Zamudio, A.V., Sigova, A.A., Hannett, N.M., Day, D.S., Abraham, B.J., Cohen, M.A., Nabet, B., Buckley, D.L., et al. (2017). YY1 Is a Structural Regulator of Enhancer-Promoter Loops. Cell *171*, 1573-1588.e28.

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

Werken, H.J.G. van de, Vree, P.J.P. de, Splinter, E., Holwerda, S.J.B., Klous, P., Wit, E. de, and Laat, W. de (2012). Chapter Four 4C Technology: Protocols and Data Analysis. Methods Enzymol *513*, 89–112.

Will, A.J., Cova, G., Osterwalder, M., Chan, W.-L., Wittler, L., Brieske, N., Heinrich, V., Villartay, J.-P. de, Vingron, M., Klopocki, E., et al. (2017). Composition and dosage of a multipartite enhancer cluster control developmental expression of Ihh (Indian hedgehog). Nat Genet *49*, 1539–1545.

Williamson, I., Lettice, L.A., Hill, R.E., and Bickmore, W.A. (2016). Shh and ZRS enhancer colocalisation is specific to the zone of polarising activity. Development *143*, 2994–3001.

Williamson, I., Kane, L., Devenney, P.S., Flyamer, I.M., Anderson, E., Kilanowski, F., Hill, R.E., Bickmore, W.A., and Lettice, L.A. (2019). Developmentally regulated Shh expression is robust to TAD perturbations. Development *146*, dev179523.

Wutz, G., Várnai, C., Nagasaka, K., Cisneros, D.A., Stocsits, R.R., Tang, W., Schoenfelder, S., Jessberger, G., Muhar, M., Hossain, M.J., et al. (2017). Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. Embo J *36*, 3573–3599.

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

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

Zhang, D., Huang, P., Sharma, M., Keller, C.A., Giardine, B., Zhang, H., Gilgenast, T.G., Phillips-Cremins, J.E., Hardison, R.C., and Blobel, G.A. (2020). Alteration of genome folding via contact domain boundary insertion. Nat Genet *52*, 1076–1087.

Zhang, Y., Li, T., Preissl, S., Amaral, M.L., Grinstein, J.D., Farah, E.N., Destici, E., Qiu, Y., Hu, R., Lee, A.Y., et al. (2019). Transcriptionally active HERV-H retrotransposons demarcate topologically associating domains in human pluripotent stem cells. Nat Genet *51*, 1380–1388.

Zhu, J., Nakamura, E., Nguyen, M.-T., Bao, X., Akiyama, H., and Mackem, S. (2008). Uncoupling Sonic Hedgehog Control of Pattern and Expansion of the Developing Limb Bud. Dev Cell *14*, 624–632.

Zúñiga, A., and Zeller, R. (1999). Gli3 (Xt) and formin (ld) participate in the positioning of the polarising region and control of posterior limb-bud identity. Dev Camb Engl *126*, 13–21.

Zuniga, A., Zeller, R., and Probst, S. (2012). The molecular basis of human congenital limb malformations. Wiley Interdiscip Rev Dev Biology *1*, 803–822.

10. APPENDIX

10.1 Primers

Table 10-1: Primers used for cloning of targeting constructs

Name	Sequence	Task
mZRS fwd	CGATgacgtcGGCCGATCCTGTCCTTACTA	Amplification of ZRS
mZRS rev	CGATcgatcgTGAAAGTGGGGAAAATATCTCAC	Amplification of ZRS
IntraTAD1 cenHR fwd	GATCgtcgacGCCAGGAGTGGCCATAGAAT	Amplification of centromeric Homology region for IntraTAD1
IntraTAD1 cenHR rev	CGATgacgtcCCATTTTGTTCTCTGGGGTGA	Amplification of centromeric Homology region for IntraTAD1
IntraTAD1 telHR fwd	CGATcgatcgCAGAGCAAAGGACTGGTAGTAAGC	Amplification of telomeric Homology region for IntraTAD1
IntraTAD1 telHR rev	GATCgcggccgcTCTCTGGGGAAGACAGTCCA	Amplification of telomeric Homology region for IntraTAD1
IntraTAD1 PAM Mutagenesis fwd	tccagggacccATTagcatccagaa	Mutation of PAM
IntraTAD1 PAM Mutagenesis rev	ttctggatgctAATgggtccctgga	Mutation of PAM
IntraTAD2 cenHR fwd	GATCgtcgacCCTGCTCCACCAATGTAAGC	Amplification of centromeric Homology region for IntraTAD2
IntraTAD2 cenHR rev	CGATgacgtcATGGGAAATGGCGAGTGC	Amplification of centromeric Homology region for IntraTAD2
IntraTAD2 telHR fwd	CGATcgatcgCATAATTGTCCCTGCTGCCT	Amplification of telomeric Homology region for IntraTAD2
IntraTAD2 telHR rev	GATCgcggccgcCTACCCAACACCTGAAGGACC	Amplification of telomeric Homology region for IntraTAD2
IntraTAD2 PAM Mutagenesis fwd	cctgtggcatcTTAgagaatcgaat	Mutation of PAM
IntraTAD2 PAM Mutagenesis rev	attcgattctcTAAgatgccacagg	Mutation of PAM
IntraTAD3 cenHR fwd	CGATgtcgacCCAGCAACCTCCACTGATCT	Amplification of centromeric Homology region for IntraTAD3
IntraTAD3 cenHR rev	CGATgacgtcGAGGGCTTTAATGTGGGTGC	Amplification of centromeric Homology region for IntraTAD3
IntraTAD3 telHR fwd	CGATcgatcgATTAAAGCCCTCCAGCGTCT	Amplification of telomeric Homology region for IntraTAD3
IntraTAD3 telHR rev	CGATgcggccgcAAGCTGAGCCCACATTCTGT	Amplification of telomeric Homology region for IntraTAD3
IntraTAD3 PAM Mutagenesis fwd	CTCTCCTTTCCtaTTTAACCAGGTCT	Mutation of PAM
IntraTAD3 PAM Mutagenesis rev	AGACCTGGTTAAAtaGGAAAGGAGAG	Mutation of PAM

KcBor C1 fwd	gatcccgcggATTGAGAGCTGTGGGAGGTG	Amplification of CTCF site 1 of the Sox9/Kcnj2 TAD boundary for pKan
KcBor C1 rev	gatcctcgagGGGGAGCACTGGATATAGCA	Amplification of CTCF site 1 of the Sox9/Kcnj2 TAD boundary for pKan
KcBor C2-C4 Gibson fwd	ATCCAGTGCTCCCCCACACACGTACGGGCAAATAA	Amplification of CTCF sites 2-4 of the Sox9/Kcnj2 TAD boundary for pKan-C1
KcBor C2-C4 Gibson rev	GACACACAGCCTCATGAAGCCCTTTATAGTCACAG	Amplification of CTCF sites 2-4 of the Sox9/Kcnj2 TAD boundary for pKan-C1

Table 10-2 Primers used for qRT-PCR.

Gene expression analysis	
Gapdh fwd	TCAAGAAGGTGGTGAAGCAG
Gapdh rev	ACCACCCTGTTGCTGTAGCC
Shh fwd	ACCCCGACATCATATTTAAGGA
Shh rev	TTAACTTGTCTTTGCACCTCTGA
Rps90 fwd	GACCAGGAGCTAAAGTTGATTGGA
Rps90 rev	TCTTGGCCAGGGTAAACTTGA
En2 fwd	TATTCTGACCGGCCTTCTTC
En2 rev	TGGTCTGAAACTCAGCCTTG
Cnpy1 fwd*	GCGAATGAACGATTACCAG
Cnpy1 rev*	AATAAGTTCGAATATCTCATCTTC
Mnx1 fwd	CATTTCATTCGGCGGTTC
Mnx1 rev	ACCCAAGCGTTTTGAGGTG
Nom1 fwd	TGGGACTCGGGGAAATAAAG
Nom1 rev	AGGAGGGATGTACTTTTCACCAC
Sox9 fwd	ACTCCCCACATTCCTCCTCC
Sox9 rev	CAGCTTGCACGTCGGTTTTG
Kcnj2 fwd	cttggttccccaatgatgtt
Kcnj2 rev	ttttatggggtggctggtta
Copy Number Analysis	
mZRS fwd	CAATGAACGCTCATGGAGTC
mZRS rev	CAGATGACTTTTCCCCTCAGTG
KcBor KI fwd	CACCCAGAATATTGCCTACAACT
KcBor KI rev	AGAACACCCAAAACAAATGAAAG

*(Anderson et al., 2014)

Table 10-3 Primers used for cloning of sgRNAs.

Name	Sequence
Enhancer Insertions	
IntraTAD1 sgRNA fwd	caccgAATAATATGTCCAGGGACCC
IntraTAD1 sgRNA rev	aaacGGGTCCCTGGACATATTATTc
IntraTAD2 sgRNA fwd	caccGCGAGTGCCATTCGATTCTC
IntraTAD2 sgRNA rev	aaacGAGAATCGAATGGCACTCGC

IntraTAD3 sgRNA fwd	caccGGGGTAGGAGACCTGGTTAA
IntraTAD3 sgRNA rev	aaacTTAACCAGGTCTCCTACCCC
ExtraTAD1 sgRNA fwd	caccgAACATACCTTTAGCGGTCCA
ExtraTAD1 sgRNA rev	aaacTGGACCGCTAAAGGTATGTTc
ExtraTAD2 sgRNA fwd	caccgATGACCTCGTTCACGTGCCA
ExtraTAD2 sgRNA rev	aaacTGGCACGTGAACGAGGTCATc
ΔZRS cen sgRNA fwd	caccgCACACGGATACTAAGACGGG
ΔZRS cen sgRNA rev	aaacCCCGTCTTAGTATCCGTGTGc
ΔZRS tel sgRNA fwd	caccgCTAAGAAACCTCACTACACG
ΔZRS tel sgRNA rev	aaacCGTGTAGTGAGGTTTCTTAGc
CTCF-deletion series	
ΔBor cen sgRNA fwd	caccGATCATTTTAGGTAACGACCC
ΔBor cen sgRNA rev	aaacGGGTCGTTACCTAAAATGATC
ΔBor tel sgRNA fwd	caccGATTTAGCGTCCCCTAGCATA
ΔBor tel sgRNA rev	aaacTATGCTAGGGGACGCTAAATC
ΔBor C1 sgRNA fwd	caccgTGGATTCCAAAAGAGGGCAG
ΔBor C1 sgRNA rev	aaacCTGCCCTCTTTTGGAATCCAc
ΔBor C2 sgRNA fwd	caccgTGTAAGTGGGCATTGCCACC
ΔBor C2 sgRNA rev	aaacGGTGGCAATGCCCACTTACAc
ΔBor C3 sgRNA fwd	caccgTACTGACCTCTAGTGGTTGG
ΔBor C3 sgRNA rev	aaacCCAACCACTAGAGGTCAGTAc
ΔBor C4 sgRNA fwd	caccgAGAGTCACTGCGCCCTCTAG
ΔBor C4 sgRNA rev	aaacCTAGAGGGCGCAGTGACTCTC
ΔK_C sgRNA fwd	caccgATGTGATCTCTTTCGCCCTC
ΔK_C sgRNA rev	aaacGAGGGCGAAAGAGATCACATc
Structural Variants	
InvC cenBP sgRNA fwd	caccGATCATTTTAGGTAACGACCC
InvC cenBP sgRNA rev	aaacGGGTCGTTACCTAAAATGATC
InvC telBP sgRNA fwd	caccGAAGCAAATACGTGAGTCTAC
InvC telBP sgRNA rev	aaacGTAGACTCACGTATTTGCTTC
InvIntra cenBP sgRNA fwd	caccGATTTAGCGTCCCCTAGCATA
InvIntra cenBP sgRNA rev	aaacTATGCTAGGGGACGCTAAATC
InvC ΔBor cen sgRNA fwd	caccgATCCCGAAATTAAACTGCCC
InvC ΔBor cen sgRNA rev	aaacGGGCAGTTTAATTTCGGGATc
InvC ΔBor tel sgRNA fwd	caccgCTTCTTAACAAAAACCGAAC
InvC ΔBor tel sgRNA rev	aaacGTTCGGTTTTTGTTAAGAAGc
KcBorKI sgRNA fwd	caccgTTCTTAACAGCCCTGTTGCA
KcBorKI sgRNA rev	aaacTGCAACAGGGCTGTTAAGAAc

Table 10-4 Primers used for genotyping of mESC and transgenic mouse lines.

Name	Sequence	Task
Enhancer Insertions		
mZRS rev	CAGATGACTTTTCCCCTCAGTG	Genotyping ZRS insertions
mZRS fwd	CAATGAACGCTCATGGAGTC	Genotyping ZRS insertions
IntraTAD1 GT cen fwd	CCAATCACCCAGTTAAGTTTCA	Genotyping of centromeric ZRS insertion site at IntraTAD1

IntraTAD1 GT tel rev	GGGGCAAGAAAGTTCTGCTA	Genotyping of telomeric ZRS insertion site at IntraTAD1
IntraTAD1 BP fwd	GCTTTGCTGACTCCCTTGAT	Genotyping of IntraTAD1 insertion site
IntraTAD1 BP rev	TTTGCTCTACACCCCAAACC	Genotyping of IntraTAD1 insertion site
IntraTAD2 GT cen fwd	GCTCATGATGACCCAAAGAGTC	Genotyping of centromeric ZRS insertion site at IntraTAD2
IntraTAD2 GT tel rev	TGGCTGTAATGAGCAAACCAA	Genotyping of telomeric ZRS insertion site at IntraTAD2
IntraTAD2 BP fwd	GCTTTGAATTTCATCCCTTCC	Genotyping of IntraTAD2 insertion site
IntraTAD2 BP rev	CTGCACCTCATCTGTGAAGCA	Genotyping of IntraTAD2 insertion site
IntraTAD3 GT cen fwd	GAAGTGTGGCAGCTTGTCCT	Genotyping of centromeric ZRS insertion site at IntraTAD3
IntraTAD3 GT tel rev	TCTGGCTGGGGTTTATTTTG	Genotyping of telomeric ZRS insertion site at IntraTAD3
IntraTAD3 BP fwd	TCCAAATTGCAACCACCTAAACA	Genotyping of IntraTAD3 insertion site
IntraTAD3 BP rev	GGGCACGATTTGCTCTCTTCA	Genotyping of IntraTAD3 insertion site
ExtraTAD1 GT cen fwd	CACTTCACTGACTGAGCCATCTC	Genotyping of centromeric ZRS insertion site at ExtraTAD1
ExtraTAD1 GT tel rev	CTCAGCTCTGGGCATCCTGT	Genotyping of telomeric ZRS insertion site at ExtraTAD1
ExtraTAD1 BP fwd	GTTATGGTGCCCGTGAGACT	Genotyping of ExtraTAD1 insertion site
ExtraTAD1 BP rev	CCATTGGTTTGACCGTTTCT	Genotyping of ExtraTAD1 insertion site
ExtraTAD2 GT cen fwd	GCAGCACCATTCTTCACTCTG	Genotyping of centromeric ZRS insertion site at ExtraTAD2
ExtraTAD2 GT tel rev	GACCAAAGCCTACTCTGAGATG	Genotyping of telomeric ZRS insertion site at ExtraTAD2
ExtraTAD2 BP fwd	CCATAGTCAATAGCACATGAGTGA	Genotyping of ExtraTAD2 insertion site
ExtraTAD2 BP rev	AGGATGAAACCAGGCAGATG	Genotyping of ExtraTAD2 insertion site
ΔZRS cenBP fwd	TTGGGAGCATGGTAATTAAAAG	Genotyping of ZRS deletion
ΔZRS cenBP rev	CCACCAACAAAACCAGACAG	Genotyping of ZRS deletion
ΔZRS tel BP fwd	GGCTAATGAGCATGTTGATCG	Genotyping of ZRS deletion
ΔZRS tel BP rev	ATTGCTCTGACTTGGGAAGC	Genotyping of ZRS deletion
CTCF-deletion series		
ΔBor C1 fwd	GACACTTGTCTAACCACCTATCCTC	Genotyping of C1-CTCF site deletion in the Sox9-TAD
ΔBor C1 rev	TCGCCACATAAAATTCTCTTATT	Genotyping of C1-CTCF site deletion in the Sox9-TAD
ΔBor C2 fwd	ATGCCTTTGGCTATGTTTGC	Genotyping of C2-CTCF site deletion in the Sox9-TAD

ΔBor C2 rev	TTCACACCGTCAGGTGATGT	Genotyping of C2-CTCF site deletion in the Sox9-TAD
ΔBor C3 fwd	ttggagggaggagagggaag	Genotyping of C3-CTCF site deletion in the Sox9-TAD
ΔBor C3 rev	agatggtgtagagcagaagttgt	Genotyping of C3-CTCF site deletion in the Sox9-TAD
ΔBor C4 fwd	TGGGATACAAGGGCCAGAGA	Genotyping of C4-CTCF site deletion in the Sox9-TAD
ΔBor C4 rev	TCCAGGTTACCCCTCCACAA	Genotyping of C4-CTCF site deletion in the Sox9-TAD
ΔBor K_C fwd	CATGCTGTGTTTAGCATGGAA	Genotyping of Kc-CTCF site deletion in the Kcnj2-TAD
ΔBor K_C rev	AACTTCCTGGCCCTGGTG	Genotyping of Kc-CTCF site deletion in the Kcnj2-TAD
Structural Variants		
InvC cenBP fwd	GACCCTGGGCAAGTGAATAA	Genotyping centromeric breakpoint of InvC
InvC cenBP rev	ACAATGGATGGTGCAGAGTTAGG	Genotyping centromeric breakpoint of InvC
InvC telBP fwd	TGGGCACAGGCAGTTATGTA	Genotyping telomeric breakpoint of InvC/InvIntra
InvC telBP rev	GGCAGACACTGTTTTTCAAGAGG	Genotyping telomeric breakpoint of InvC/InvIntra
InvIntra cenBP fwd	tggagaagccagtcttttgg	Genotyping centromeric breakpoint InvIntra
InvIntra cenBP rev	ACAATGGATGGTGCAGAGTTAGG	Genotyping centromeric breakpoint InvIntra
InvC ΔBor cen fwd	TTGTGGCAGGGAGAAGACAT	Genotyping boundary deletion
InvC ΔBor tel rev	TGGAAATGTAGTGGGAGGGA	Genotyping boundary deletion
KcBorKI cen fwd	CTGGGTTTATATGCTGCTGGAGAG	Genotyping boundary insertion
KcBorKI tel rev	GACAGCACCTCCCACAGCTC	Genotyping boundary insertion

Table 10-5 Primers used for cloning of WISH probes.

Name	Sequence
Shh fwd	
Shh rev	
En2 fwd*	AAAGGGGACTGTTTAGGGTTTC
En2 rev*	GAAGATGATTCCAACTCGCTCT
Mnx1 fwd	GACGAGGATGATGAAGAAGAGG
Mnx1 rev	GCCACTCCTAGAAAGGGTAGGT
Sox9 fwd	CGGGCGAGCACTCTGG
Sox9 rev	TGGGGCTCAGCTGCTCC
Kcnj2 fwd	ATATGACTGGCTGATTCCGTCT
Kcnj2 rev	CTCAACACTGACGTCTTAACGTT

*from Eurexpress (http://www.eurexpress.org/ee/)

10.2 Plasmid maps

Figure 10-1 pSpCas(BB)-2A-Puro (px459) v2.0 vector map

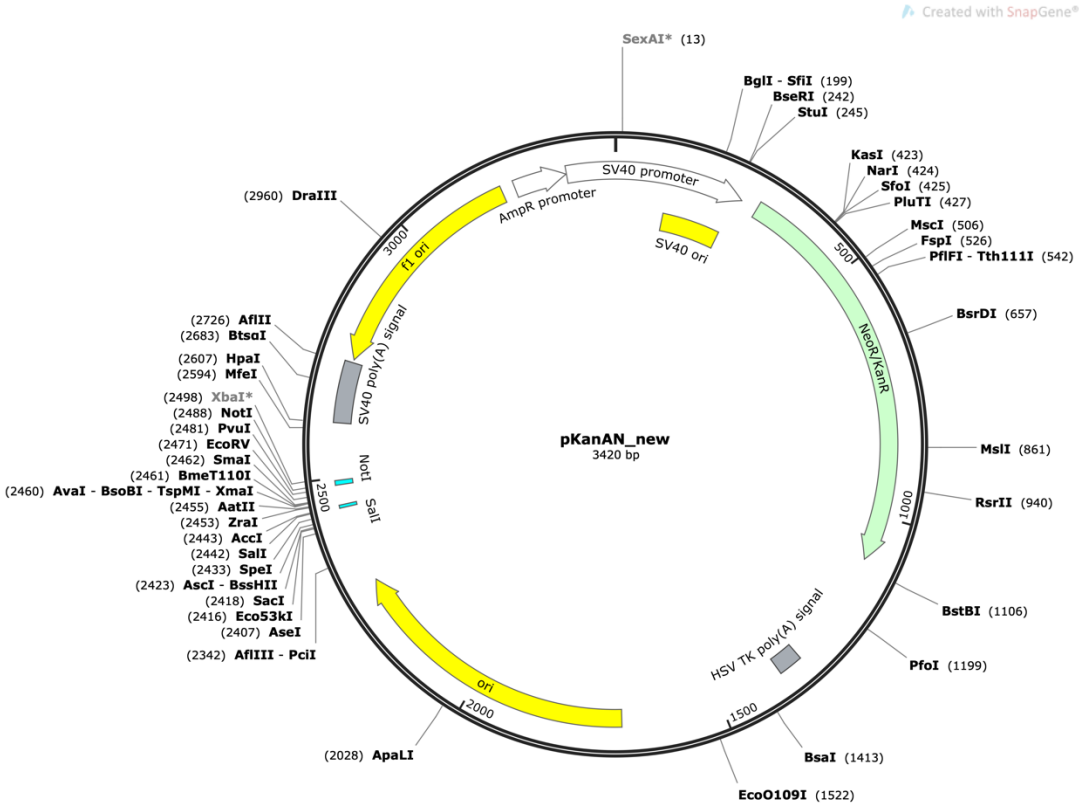

Figure 10-2 pKan Backbone for cloning of targeting constructs

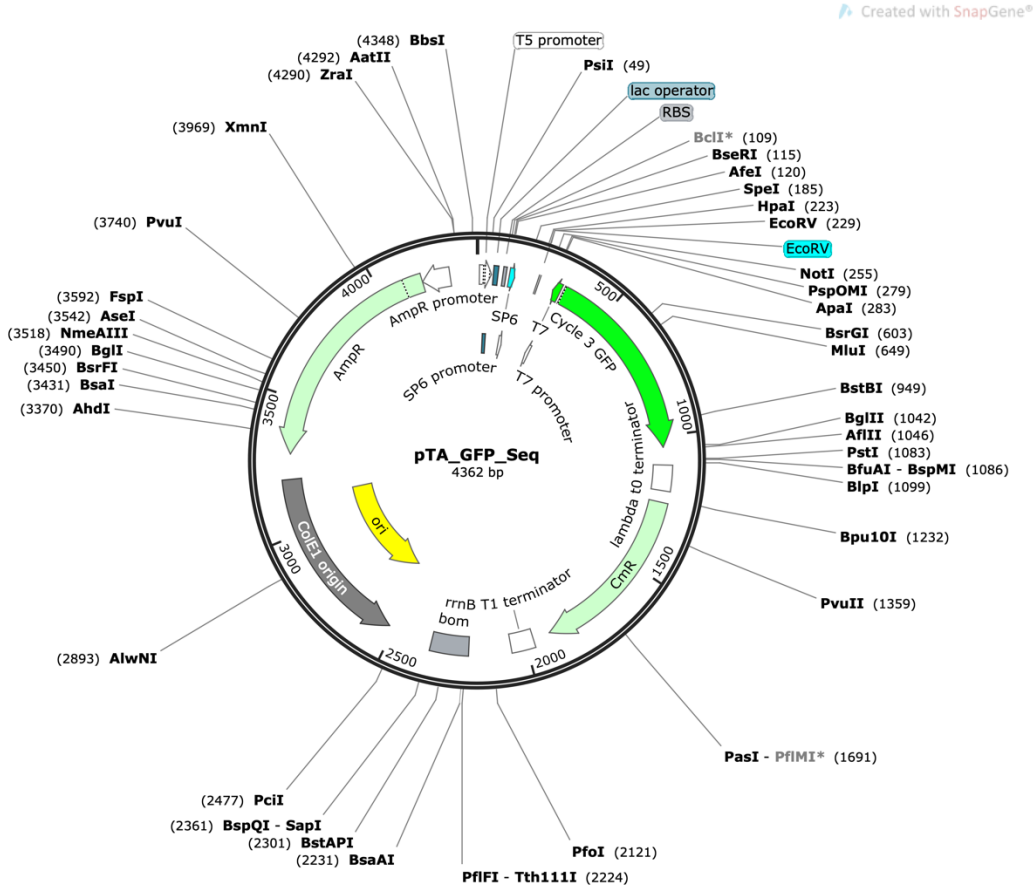

Figure 10-3 pTA-GFP vector map utilized for subcloning of fragments and WISH probes.

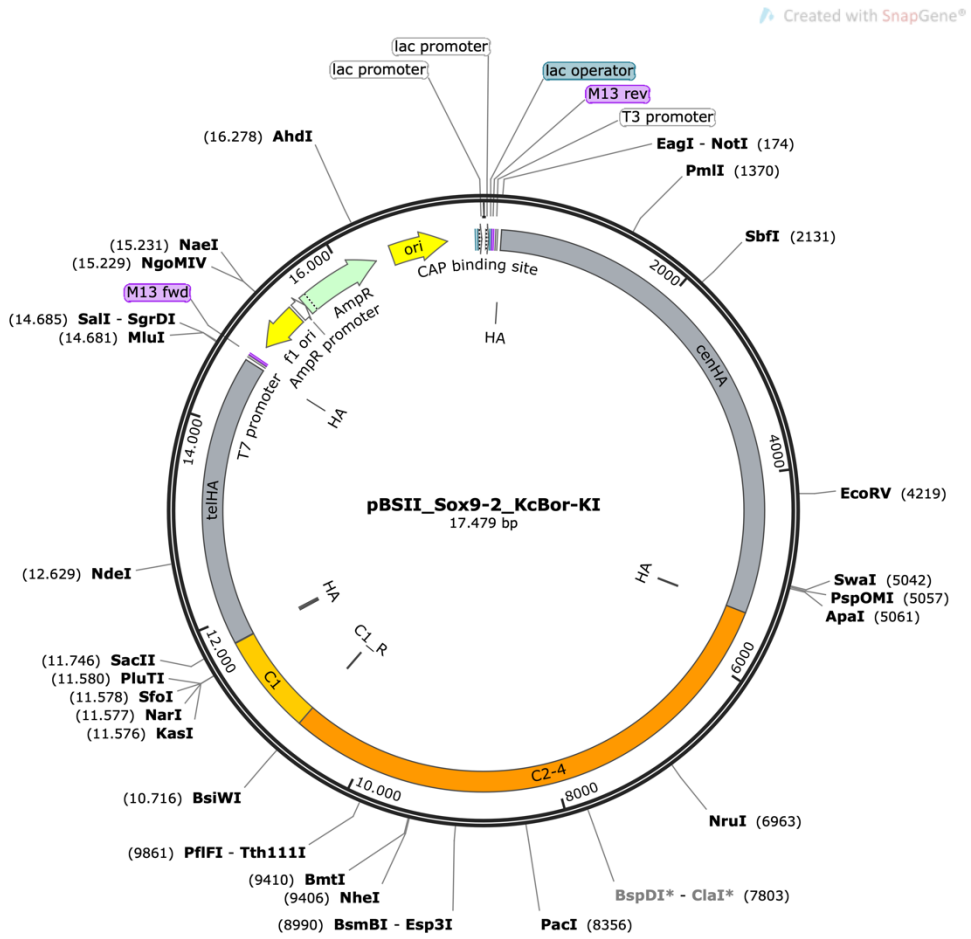

Figure 10-4 pBSII-Sox9-KcBor-KnockIn.

10.3 Synthetic DNA sequences

10.3.1 ExtraTAD1 telHR

TGCCTTACCTCTTTGCACCTctagtggggctcacacccagtcccaggaacccttttctagaagccttttagcctcc
catttcttgctgctttcccctcattgtgcctactcttgccccctccttctccagccttagatggagtgtctcca
ttcccttctctcctttttgtttccacagcgatctctacactgtttgctcacttaaagcaatctctgtctttcacct
tagccattctggccactgcacagagcccacagccatgtttccgtgagtgccagatgcctgaagttagccatggca
accataggcaaggaaggggctcagcctctgcctcagcttccctctgtagggaagggcaacaaggagcggcttgg
ggctcccaggagacaaaggaagaatatttcagcgaccaaactgagtccaggtactgagactgtaaagctggacag
acccaattggagctagcagtgtgggtaacacagatgccagaaactgatgtctgaatgaaacagttagcttaatcc
tgggcatatgggaaagtgctggcctctgcccatgggtgtaacttagcctttctgattagactgctgcccaggcct
tctctgcactgatgtcatagccttcccctaccctgcaactgctcttctgagcctcactgtccccaccgagcatgc
aggcaaaggggacagtccagcactgattcctgtgagagcagatgcagctacttccagagcctccctttctcttt
ggccagtggtatttcccctgttaaagtcacactcggcacagaagatgcagccatgcatgcaggcctagcagccat
ccaacactggctcttttcagggttctgcacagcatccatgcacagcttgctcctccttgaggacttcttaaccgt
aagtactgacagtggtgtcttagcagagagtgggaaaggggccctgagcttttaactccacatccccacagattc
tgcctccaccccactccatctgggccctcaatctctcaaccagccaggactgtggctggaaaaacctgctcatac
agagccctgaacaccaaatgccattcttgctcattgccagggtgcccagtgccccatgctgttatccaagcaaat
aaaaaggggtttctggtcacagcctctccctacactgagaagcaggacatagccaaagggccctgctctgtgtca
tgtttccaaattcagccaccatgctttgtcccagctcttccctcccaggatcctgctgtgactgcctctttctca
ccctccttccatctgctactatggtccccagaaacaggcctctgtcctaacccagcacagcctgtggctctggat
gtcacccaaagaagacattctccattacctttggtcatctggtattccagtccaaatgactctccttaagtgcat

ggcgccctctcctcaactcacagccatttgtacaatgcagaccctcaccatggtctcaggagcttgccgggacag
aggagggtccccaaactctgccataacgatgtcgtaaggtccaggtccctggtgagaagatcccaaacaggactg
aatacacttagaatctcagagaagttccagtcagttgctacacccctgcatgctttgcttaagtacaagaattcc
ttaggtccctctggagactgggccctgcttccacactagtggaacagagcagaggcagactccaccccaaggagc
tgacacagagaaaccacaaggctggtgttatggtgcccgtgagactccctacccagaccaaggaactgacctgag
gctttggggctcttattaacacagaccatagccaggatccttctcaggtctgtttggcctggttctggaggggaa
gcagcccaatacctctctgctagagtctgagaaccagagatggtgcagagctaggGAGTGGATAAGGGTTATCAG
AAGG

10.3.2 ExtraTAD2

5‘CTCTCCCACAACCTTGGAGCTGTCTACACTCACCATATGGCAGCTTGCTTCCTCTCTCTCTTGCCTACCATGGCTAACTGG
GCCTACCATGGTTAACTAGGTCAGGATGGACCCCAAGGCACCCAACTTCAAGCTTCCAGAATTTCACCAGCCAGGGAATGTCT
CACAGAGCCTTAGTCAAGGATTCCAGAGAATGGGTGGAAAGGAAGCGTATGGCAGGAACTATACAAAGAGACATTTCACAAAG
TAGTGAGCCATGGGTCAGAGGATAACAAGACAGAATACAGAACTCATCTGCTGGTGGCATCTGGTAAGTCAGTCTAGGAGCTG
TGGATTCCTAGGTGTGCCAGCTCCACCCCACCCTCCAGGTTCTCCAGATCACTAAGGGCCCCAGGTGTGATGAGAGGTGCAAG
CTTGGATACCTGTATGAGCACAGCTGGAAGGTGGGGCAGACATGCTCACTGACCATTGCTCCACCTCAGACTAAGGTCAGGGA
TACCCAGAGTTCTATGTCCTAAATGTACTATGGAGCCGATCTGTGAGCGTATGACGTCGGCCGATCCTGTCCTTACTATTTCT
GGAAAGAAACCAGTGCTCCCTAGTGGGGGAGAGCAGAGAGTTCTGATTAATACATGACTACATCTTTTTCTTGCAGGTGTTGG
GAGAATCAAATTAACACATAGCAACAGTTAGTGAGATATGACTTCATTTTCTGTAATAAATGCCAAGATCAAAACATGCCCAA
GTCAAATTTCCTTGCAAGGTTCCCAGCAGGAGCCTTCCTCTTGCCTGTGATTTCCTTTCACCCAGCAGAACCAGACTGAGTTT
GCGCATTTGCCACTAACACTAAGCAGCACTTCCTGAATCGCTCATTTCCAACAATTTATGGATCATCAGTGGCAAAAAAAACA
AGCAAAAATAATGAAAGAATCCAATGAACGCTCATGGAGTCCCAGGCTGGACTTCCTACTCTCTTCTCTTTAAGATGGAGGCC
TGAGACAAATTAGCCACTGAGGGGAAAAGTCATCTGGTCATAAAATACAGTACAAGGTCACTTTTATGTAAGTTTGCCAAAAG
CGACATAAACCAGGACAATCTCAAACTGTGACACAGGATAGAAACATATTAAAACGATCTTAGTTCCTCCTCTATTGTGCTGT
CATGTTGCTTGGCTTTATGGACAGCCTGAGCCCAGCAAAGTGCTTCGCCTCCACCTGGTCAGTGCAACTAATTAAGAGAGGAA
GTGACCTCAGATTTTGTGCATTTTACTTTTATTATGAAAGTACTTTGTTTGAAGTTACAGGAGCTAACCTCTTAATCCTATAG
ATCATGTGTGAGGTTCTGGACACTCTAGGATGGCTGGATGGTTTGGATAATTGGTCAGGATGTTGCTGTCATTATGTTAAGTT
TTATGCCAGGACTTCAAATCAAAGATAGGCATTAAAGTAAAGGGCAATATAAATCTGCTACCAGAAAAAGTCCTGGGCTTTAT
GAATTAATCTGTCGTCATATTTACATTACACTTACAAAACAATCATATGGTGAGATATTTTCCCCACTTTCACGATCGGCTGC
CTGCTGCTCTCTCACACCCCTGCAGCTCTTGTGAAGTTGGCGGTCCGTTTTCCAACACAGCAGCCTGACTGTTAGCACTTCCC
AAGATTATCCTACAAGTCAGCCCTGAGGCCAGACGTGACTTCCAGCGTCCCCTGCAGCAACTTTCTTTTGCTGAAATCACATG
CGCTCCTGAAGAGATGAGATGGTAAGACTTGGACTCTCAGGGTGATGTGGCCTCTGCTTAGCAGCCACATTCCCTTGGAGGTT
GCTGGAATGATCTCTGTTTTGTTCTCTGGAAAACTCTGGTTTTCCTCCACTGGCTGCATCACAGTGTAGCAATGGATTAGGAA
AGTGCAAAGTGGGTAAGTATATAGCCACCTAGCATCCCATTCACAAAGAGCATCACTGCAGGAGGCCAGTGGCAGTACCTTCC
CAGGGCAGCACCCTATGACTTGATGACTTCCCACAAGGCCTTCCTTTTAAATGCTTTACAGGCCCCACACGGCCATCCTGGGG
CCCAGCTCCCAACACAAACCTTGGAAGCAAACCACACCACTGCAGTAATGGACACTGCTCAGAATTTGGAGGCTAAGAAGAGG
GCATGTGGAGAAAGGCTAGTTTGGTGTAAGGCACACTGCATTTTCTCAGACCTGAGAGATCTGTGGTGGTCAGGACCAGGCAG
ACATCACCACTCTACCTCTGGAAGGAGCCTGTGGACTTCCAGCATCTAGGATCCCTGGAGAAGAATAAAGTCAGGAGGAAAGA
GTGCTATGAGTGAACTTCTGGACAGCTAAACATAGACCACAGTTGTCCAATCAAGTGAGCCACATGCTGGGAGGTCAGATTCC
GGGATGGGAGGATTTGGCTATTGCTCCCCAGATCCTCCTTAGCACCCTGCCCAGAAAAAAACAAAACATGTCTGTTGGAATGC
TTGGGCTATGAGTCCATGTGACCACAGAGGGATTAAAGATCCATGTGAAGGTAGCCCTGACAGGCAGGGGTACTTGGCCTTCC
CTCCCCTTTATGCCTTGTGTCATGGGTTACAGCACAGTGATAGTAGGGACTTGGCTGTTTGGAGGTAAAGCTCAAGAGAAGGG
TACATTTATCTTGAGTAGAGCAGAATAATAGTGAAATATTTATGTTTGCATGAGATGGTCAAGGGTGACATTGGGTCCAGTGT
GACAAGGTCCAGGTTGAAGGGTGATCTCAGGGCTGTGGCTCTTGTACACGCAAAAGACAACCGTATAGCATGTGCCCCACTGG
AAGGCAACTTTGCCCATGCTGAGCAGCTGTGACTTTGGACAAGTCCAGAGAAGGACATGGCAGCTGCCAACCTTCCCTGGGTC
CCCAGCTCATGTTTGCAAC-3‘

10.3.3 IntraTAD1

GTTATTAATGAGAGCTCATGGCGCGCCATGACTAGTATCGGAAATGTGGCCGAATTCTTAGCTTAGTCTCCGCTTTAGGGGCC
CAGGGAGAGCCATGGTTCATTCCTTTTCTCTGCCTCGGAAGGTGGGACCTCAAGTGAGCTGTGCACTAATTTTGAACTTAATT
CTCCTAACCTAAGACTTGAACTTCCAGGAGGGCCCAGTGTTCAGGTGCTGTCCAGCACTCACTCAAAGTCATTTCCTTCCCCA
CTAAGGACTAACATGGACTGACATTCTGAGAGGGGGCTGGCTGTTCCTACCAGCAGCCTGTCCACAGCAAGGGGAGACTCCTC
ATAGCCTGTTCCCCAGAGAGTGCCTTGGTGCCATTTTTCCACAGATGGGAGGGCTGAGAAGCCTGGATGGTGTCTCACTGGGG
GAGGGCAGGATGCCAAGAGAGACCCAAAGGTATGAGAAGGAGCCAAACTCCACTTCTCCCCTGGCCTTTTCAGTGTGAGGTGG
CTTGCCCCAAAGTACACAGACCCTCAGATGAATCCAGCTTTCTGTTTGAGTAGCACCCAGACAAAGCCTGCCCTCGGGGATGC
TCCTCTGCTCCCTCCCACCATGGTAACTCTGGCATTCACTGGACCATGGGAGCAGGttAAGTATGCGTGTGATTGTCTCGATT
CGGCCGATCCTGTCCTTACTAtttctggaaagaaaccagtgctccctagtgggggagagcagagagttctgattaatacatga
ctacatctttttcttgcaggtgttgggagaatcaaattaacacatagcaacagttagtgagatatgacttcattttctgtaat
aaatgccaagatcaaaacatgcccaagtcaaatttccttgcaaggttcccagcaggagccttcctcttgcctgtgatttcctt
tcacccagcagaaccagactgagtttgcgcatttgccactaacactaagcagcacttcctgaatcgctcatttccaacaattt
atggatcatcagtggcaaaaaaaacaagcaaaaataatgaaagaatccaatgaacgctcatggagtcccaggctggacttcct
actctcttctctttaagatggaggcctgagacaaattagccactgaggggaaaagtcatctggtcataaaatacagtacaagg
tcacttttatgtaagtttgccaaaagcgacataaaccaggacaatctcaaactgtgacacaggatagaaacatattaaaacga

tcttagttcctcctctattgtgctgtcatgttgcttggctttatggacagcctgagcccagcaaagtgcttcgcctccacctg
gtcagtgcaactaattaagagaggaagtgacctcagattttgtgcattttacttttattatgaaagtactttgtttgaagtta
caggagctaacctcttaatcctatagatcatgtgtgaggttctggacactctaggatggctggatggtttggataattggtca
ggatgttgctgtcattatgttaagttttatgccaggacttcaaatcaaagataggcattaaagtaaagggcaatataaatctg
ctaccagaaaaagtcctgggctttatgaattaatctgtcgtcatatttacattacacttacaaaacaatcatatgGTGAGATA
TTTTCCCCACTTTCATCTTGTTTCTTCACATTGGCAGACAGTGTTTGAATGTTGTGTACATGTTTTTCTTTCTCATCCTGACA
TATTTATACCCTAGTTTTGTTTTGTTTTATTTGGCCCGGTGGAGAAGGGATGTGTGTGGGAGGTGACAGAGTGACAGGGACAC
CCTTTGAATTAGATTTGTTGTGAATGCCCATCTAAGGATCTTAGAATTTGATCACGGAAGGAAGGAACTTGGAAGCATTCTTA
AAACTGCGTGCCAATGTAATTGCACCACATTCTGATCCAGCTATTAATGAATCAAATGAAGCAGGTATAAAGGCTTCACATAC
GCAGGCAACAGTTCCATAAAGGAGGAAGTAAAACACGTAATTAGGCAAATCAAATAAGGCACGGAAACAGCTTTCAAATATTT
AGTCACATGCCATTTCCCATTAGTTTACTGTGTGCTTAAATGAGCCTGTGGTTTTATGGCCTGGTTAAACGAGAGGCAACTGC
AGACACATGGTTTTTAAAGGTTGCTTCTTCACACCTCAGGGTGTTGAGCAGGGCCAGTTATACCTAATAGAAGGGGAGAACAG
GAATAGGACTTCCCAAAGATGGCCTGGAGATTGCCAGGGTCACATGAACATCAGTTCATGAGGCTGCAGCTAGAGGGGTCTCT
GGGGGGTACACCCCCATCCCATTTCCACAAGGTGGATGGTCTTTTGAACAATTACAGAACTTTTGTTCAAATCTGCTGCAATG
TGGCAACTATTGGTCAAGCCATTGGTCGCCTTTCTTTGTATAAATAATCAATTCTAATCAGTAATTGCATCTGGGCAACAGCT
CAACTCTGCTGTCATCTTTTGGCTTCCAGCACCGTGACTCTGAGGGCAGCAATACACTTTACTATAACGGTACTGAAGGCCAG
GCAGCATTTATTCTTATGTTTAACTGTCTGGTTAACAGAGTGTCATTTCAGTTCCATCTCCACCGTATGTAAAAGAAAGGTAG
GTAAGAGGCAGAATCTTTAGAATATCCACTATATGAGCTCCCAGAGAGGAGAAGGACCATCTCTCTTAGTTTTCATAGCTGCA
TCCTGGGATTGTTGAAGAGATTTCTTGGGCAAGTACCTTTTAAAATCAAtctctctttctctctctctctctcCTTGTTTGAT
ACAGAACAGAATGGAGGACTGAGGGATAAAAATAGCTATTAAGTCAAAGATAAACTCAAAACCCTCTGCAAAGAGCTTTCAGC
TTATCGACTCTAGATCATAATCAGCCATACCACATTTGTAG

10.4 Supplementary Figures

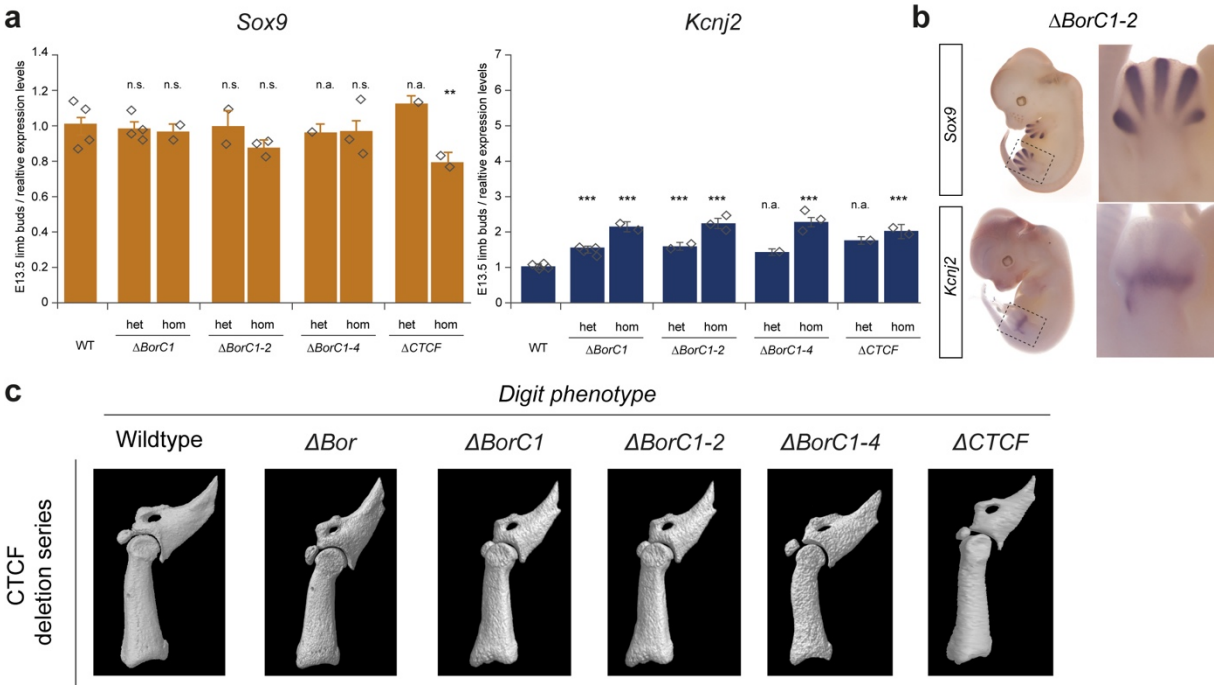

Figure 10-5 Subtle changes in gene expression and absence of phenotypes in the CTCF-deletion series. (A) Gene expression changes assessed by RT-qPCR. Relative gene expression levels of *Sox9* and *Kcnj2* in E13.5 limb buds normalized to *Gapdh*-expression (wildtype = 1). Bars indicate mean expression, error bars standard deviation. Diamonds indicate individual replicates. One-sided, unpaired t-test was conducted to test significance in comparison to wildtype (*p<0.05; **p<0.01; ***p<0.001, ns=not significant). n=2-4. (B) Changes in gene expression patterns by WISH of E12.5 embryos (n=3). S Magnification depicts a detailed view of the hindlimbs. *Sox9* is expressed in the digital anlagen. *Kcnj2* is expressed in the distal zeugopod. Note strong misexpression of *Kcnj2* in the *ΔBorC1-2* mutants. (C) 3D microcomputed tomography scan of terminal phalanges of adult animals from adult animals of the CTCF-deletion series and wildtype (n=2-4, 7-12 weeks).

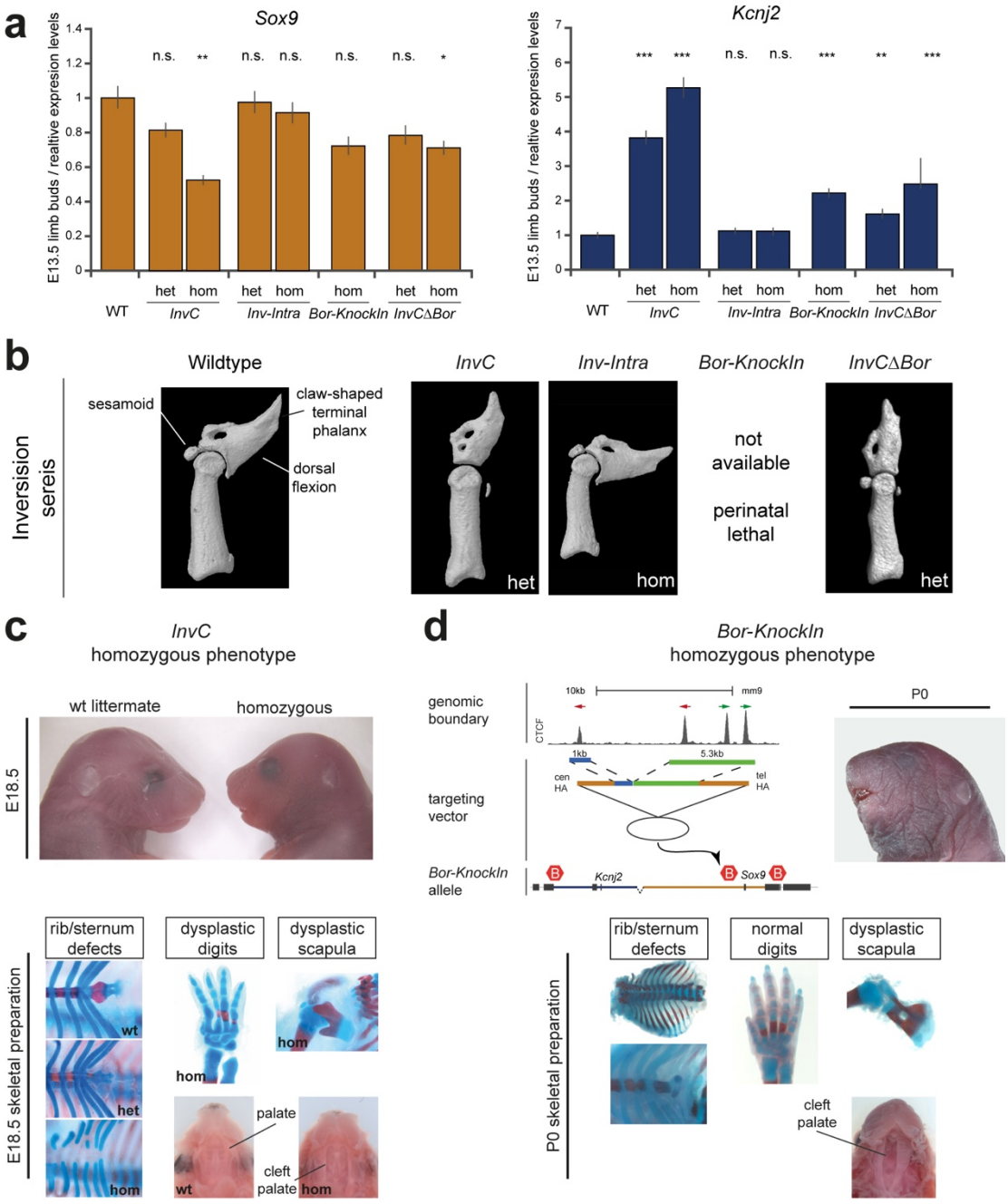

Figure 10-6 Sox9-LOF and Kcnj2-GOF phenotypes upon redirection of interactions in the SVs series. (A) Gene expression changes assessed by RT-qPCR. Relative gene expression levels of *Sox9* and *Kcnj2* in E13.5 limb buds normalized to *Gapdh*-expression (wildtype = 1). Bars indicate mean expression, error bars standard deviation. Diamonds indicate individual replicates. One-sided, unpaired t-test was conducted to test significance in comparison to wildtype (*p<0.05; **p<0.01; ***p<0.001, ns=not significant). n=2-4. (B) 3D microcomputed tomography scan of terminal phalanges of adult animals from adult animals of the SV series and wildtype (n=2-4, 7-12 weeks). *InvC* and *InvCΔBor* animals display loss of dorsal flexion, sesamoid bones and claw-shaped form of the terminal phalange. (C) Phenotypes of homozygous *InvC* E18.5 embryos compared to wt-littermates. Lateral view of the head on top. Note, shortening of the snout and micrognathy in the mutant embryo. Below, skeletal

preparations shows distinct defects characteristic for a *Sox9*-LOF. (D) Phenotypes of homozygous *Bor-KnockIn* E18.5 embryos derived from tetraploid aggregation. Top left, design of the targeting construct containing all four CTCF bindings sites of the TAD boundary (C1-C4, 6.3kb, HR: 2 and 4kb). Top rigt, lateral view of homozygous P0 newborn with a short snout and micrognathy. Below, skeletal preparations shows distinct defects characteristic for a *Sox9*-LOF. Overall, *Bor-KnockIn* animals displayed a milder *Sox9*-LOF phenotype compared to homozygous *InvC* animals and no phenotype in the digits.

10.5 List of Tables

10.6 List of Figures

10.7 List of Abbreviations

°C	degrees celsius
µ	micro (prefix)
∞	infinite
3C	chromatin conformation capture
3D	Three-dimensional
4C	circular chromatin conformation capture
AER	Apical ectodermal ridge
bp	basepair
cDNA	complementary DNA
cen	centromeric
cHiC	Capture HiC
ChIP	Chromatin immunoprecipitation
Chr	Chromosome
CTCF	CCCTC-binding factor
del	deletion
DEPC	diethylpyrocarbonate
DIG	digoxygen

DMEM	Dulbecco's modified eagle's medium
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide
DSB	Double Strand Break
E	Embryonic stage
E.coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
ENCODE	Encyclopedia of DNA elements
ESC	embryonic stem cell
EtOH	Ethanol
FL	forelimb
g	Gram
h	Hour
het	heterozygous
HL	hindlimb
hom	homozygous
kb	kilobases
Kcnj2/16	*potassium inwardly rectifying channel subfamily J member 2/16*
KI	KnockIn
KO	KnockOut
l	liter
LIF	Leukemia Inhibiting Factor
m	Milli (prefix)
Mb	Megabase
mESC	Mouse embryonic stem cell
MeOH	Methanol
min	minute
mm	*Mus musculus*
mol	Moles
mRNA	Messenger RNA
n	Nano (prefix)
ON	Over night
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PFA	paraformaldehyd
PIC	Preinitiation complex
qRT-PCR	Quantitative real-time PCR
RNA	Ribonucleic acid
rpm	Rounds per minute
RT	Room temperature
s	second
sgRNA	Single guide RNA
Shh	*Sonic Hedgehog*
Sox9	*SRY-box transcription factor 9*

SSC	Saline sodium citrate buffer
TAD	Topologically associating domain
taq	*Thermus aquatius*
tel	telomeric
TF	Transcription factor
Tm	Melting temperature
TSS	Transcriptional start site
U	Units
UCSC	University of California, Santa Cruz
Vol	Volume
WISH	Whole Mount In-Situ Hybridization
w/v	Weight per volume
wt	Wildtype
ZPA	Zone of polarizing activity
ZRS	ZPA regulatory sequence

11. DANKSAGUNG

An aller erster Stelle möchte ich mich bei Prof. Dr. Stefan Mundlos bedanken für die Möglichkeit Teil eines so hilfsbereiten und dynamischen Teams zu sein, in welchem ich meine Doktorarbeit anfertigen konnte, die Diskussionen, Hilfe und die richtigen Fragen an der richtigen Stelle. Mein besonderer Dank gilt auch Prof. Dr. Sigmar Stricker für die Übernahme des Zweitgutachtens.

Daniel - Du hast mir vor allen Dingen beigebracht meinen eigenen Kopf zu benutzen und Dinge nicht einfach nur auszuführen, sondern zu verstehen, kritisch zu hinterfragen und eigene Ideen zu entwickeln. Dafür bin ich dir sehr dankbar und ich finde wir sind ein klasse Team. Meine Arbeit, ja schon von vornherein ein Aprilscherz. ChIP kann ich immer noch nicht richtig mappen, weil ich mir keine Notizen machen darf. Und meine Triplikate bei qPCRs SIND einfach sauberer als deine. Danke für alles, die wichtigen Tritte an den richtigen Stellen, deine ehrliche Meinung, Inspiration, Motivation und ich freue mich schon auf Song 6 der Band.

Besondere Dank geht an Asita- du hast mich direkt an die Hand genommen als ich angefangen habe und bist immer eine tolle Unterstützung! Ute, danke für deine freundliche, liebe Art und jegliche Hilfe. Norbert, du hast selbst die schlimmsten Aufnahmen in hübsche Fotos verwandelt- Danke! Nicht zu vergessen, einen riesen Dank an die beste Studentin der Welt: Cinzia. Durch meine Arbeit habe ich nicht nur Kollegen, sondern auch vor allen Dingen fantastische Freunde gewonnen. Danke Alessa-ohne dich wäre Alexa einfach unvollständig; Bjørt, für dein inspirierendes und herzliches Naturell; Mike, da fehlen mir *in general* die Worte; und Sala, für deinen steten Beistand in Team-Brahim und darüber hinaus.

Danke, danke, danke an die ganze AG Mundlos. Ich wünsche mir für jeden eine so soziale und freundliche Gruppe. Danke für alle Diskussionen, Hilfe und noch viel mehr die zahlreichen schönen Momente: Alicia, Andreas, Blanka, Carola, Cesar, Chiara, Christina, Fany, Felix, Fiona, Friederike, Giulia, Guillaume, Henrike, Ivana, Jana, Jessy, Josh, Juliane, Katerina, Konrad, Lila, Magdalena, Malte, Martin, Masha, Mikie, Mira, Patricia, Pedro, Philine, Robert, Rocío, Tobi, Uirá, Vanessa, Vera und Verena.

Besonders Dank gilt meinen Freunden: Grüße und Küsse gehen raus an das Kleeblatt, die Sektschwestern, das Blumenküble, Team Tropical und die Geflügelsalamis. Ihr seid die allerbesten die ich mir wünschen kann. Nur mit euch kann ich rheinisches Lebensgefühl im kalten Osten erleben. Und Georg: das ist mein persönlicher Bester!

Ich danke meiner Familie und besonders Mama und Papa. Ich denke ihr wisst wie unfassbar verbunden ich euch bin und ich freue mich einfach wenn ihr euch freut!

Yes the picture's changing, every moment; And your destination, you don't know it. (Roxy Music)

12. DECLARATION OF INDEPENDENT WORK

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

Berlin, 30.01.2021

_____

Alexandra Despang

13. SCIENTIFIC PUBLICATIONS

Despang, A., Schöpflin, R., Franke, M., Ali, S., Jerković, I., Paliou, C., Chan, W., Timmermann, B., Wittler, L., Vingron, M., Mundlos, S. & Ibrahim, D.M. et al. Functional dissection of the Sox9–Kcnj2 locus identifies nonessential and instructive roles of TAD architecture. Nat Genet 51, 1263–1271 (2019). https://doi.org/10.1038/s41588-019-0466-z

Kraft, K., Magg, A., Heinrich, V., Riemenschneider, C., Schöpflin, R., Markowski, J., Ibrahim, D.M., Acura-Hidalgo, R., **Despang, A**., Andrey, G., Wittler, L., Timmermann, B., Vingron, M. & Mundlos, S. / Serial genomic inversions induce tissue-specific architectural stripes, gene misexpression and congenital malformations. Nat Cell Biol 21, 305–310 (2019). https://doi.org/10.1038/s41556-019-0273-x