The Role of Higher-Order Chromatin Organization at the *SOX9* Locus in Gene Regulation and Disease

Dissertation

zur Erlangung des akademischen Grades des
Doktors der Naturwissenschaften (Dr. rer. nat.)

eingereicht im Fachbereich Biologie, Chemie, Pharmazie
der Freien Universität Berlin

vorgelegt von
Martin Franke
aus Weimar

Berlin
2017

Die vorliegende Arbeit wurde im Zeitraum Oktober 2010 bis Februar 2017 am Max-Planck-Institut für molekulare Genetik in der Arbeitsgruppe von Prof. Dr. Mundlos angefertigt.

Erstgutachter: Prof. Dr. Stefan Mundlos

Zweitgutachter: Prof. Dr. Sigmar Stricker

Eingereicht am: 21.03.2017
Tag der Disputation: 07.06.2017

„Zum Entdecken gehört Glück, zum Erfinden Geist,
und beide können beides nicht entbehren.“

Johann Wolfgang von Goethe (1749 – 1832);
Goethe's sämmtliche Werke,
Tétot Frères, Paris 1836, Band 4,
Zur Naturwissenschaft

Table of Contents

1 Introduction

1.1 Gene regulation

Development of multi-cellular organisms requires an orchestrated interplay of fundamental processes at the cellular, molecular and genomic level. The communication between these levels via signaling pathways and transcription factors ultimately defines the way genetic information encoded in effector genes is transformed into building and maintaining an organism. Unlike ubiquitously expressed genes, most developmental genes involved in pattern formation and cell fate specification exhibit a precisely adjusted expression in a time-, space- and cell-type specific manner. Hence, perturbations in the exact coordination of gene expression often lead to developmental defects and diseases in humans (Kleinjan & Van Heyningen 2005). Several mechanisms, such as post-transcriptional modification of proteins, chromatin modifications and *cis*-regulatory networks, can modify the level of transcription and account for complex spatio-temporal gene expression patterns in vertebrates (Levine & Tjian 2003). The following sections will pay a particular attention to *cis*-regulatory elements, which represent a major basis of transcriptional gene regulation.

1.1.1 Transcriptional control by cis-regulatory elements

Simplified, gene transcription requires the recruitment of a DNA-depended RNA polymerase to the gene promoter upstream of the gene and a switch, which turns on or turns off transcription. Transcriptional control in bacteria fulfills this fundamental view of gene transcription. Bacterial promoters contain specific binding sites to recruit RNA polymerase and additional operator sequences, DNA motifs, close to the recruitment sites that bind various regulatory proteins. Depending on environmental conditions, activator or repressor proteins are present in the cell and are able to bind their particular motif and stimulate or terminate transcription. This dual control by *cis*-regulatory elements (CREs) was first described with the *lac* operon in *Escherichia coli* (Jacob & Monod 1961; Malan & McClure 1984). However, a single promoter in multi-cellular and higher eukaryotes requires more than one turn-on or turn-off state and additional CREs to modulate the rate of transcription for several cell types. The integration of different cell- or tissue-specific factors at a single promoter therefore necessitates a more complex and combinatorial control of transcription. From this perspective, the promoter region alone is simply too small to accommodate complex *cis*-regulatory modules (Riethoven, 2010). Therefore the increase of organismal complexity is accompanied by an increase of CRE complexity in promoters on the one hand and of CREs that are physically separated from promoters on the other.

Interestingly, the physical separation of CREs from their target core promoter region can comprise several thousand base pairs in higher metazoans (Birney et al., 2007; Kvon et al., 2014). Based on their functional effect on gene transcription, they have been classified into enhancers or repressors (Banerji et al., 1981; Moreau et al., 1981; Ogbourne & Antalis 1998). Enhancer and repressor CREs provide additional binding sites for tissue- and sequence-specific transcription factors (TFs) to modulate transcription by regulating the rate of formation and binding of the basic transcriptional machinery at the promoter or attenuate quantitative levels of transcription. Among these types of CREs, enhancers are the best characterized because of their determining influence on gene transcription, which will be discussed in detail herein. Another important class of CREs are genomic elements involved in insulation and shaping of the chromatin architecture, which is important to physically linking the additional information provided by enhancers to their target gene promoters (Ong & Corces 2009; Bulger & Groudine 2011).

1.1.2 CREs of promoters

In contrast to prokaryotes, promoters in multi-cellular eukaryotes are broader on the whole and exhibit greater diversity. Core promoter elements, which recruit the basic transcriptional machinery and RNA polymerase, are composed of a number of elements around the transcriptional start site, like the TATA box and Initiator elements (Maston, Evans, & Green 2006). Additional CREs located a few hundred base pairs upstream of core promoter elements have been historically defined as proximal promoter elements. These regions bind general transcriptional activators that can act synergistically with the transcription machinery or bind specific TFs to initiate transcription in a cell-type-specific manner (McKnight & Kingsbury 1982; Tsai et al., 1996). Importantly, besides their traditional function in regulating transcription, proximal promoter regions have been postulated to serve as tethering elements for more distantly located CREs (Calhoun et al., 2002; Akbari et al., 2008; Riethoven 2010). For instance, the binding of specific TFs and LIM domain binding protein 1 (LDB1) at proximal promoter elements at the *beta-globin* promoter recruit additional transcriptional cofactors from distally located enhancers. This in turn influences the assembly and initiation of the basic transcriptional machinery and governs the processivity of RNA polymerase through phosphorylation (Deng et al., 2012). Another level of transcriptional regulation at promoters is mediated by chromatin compaction. Many promoters are enriched for stretches of cytosine and guanine nucleotides (CpG-islands) (Saxonov, Berg, & Brutlag 2006). The methylation of CpG dinucleotides is involved in chromatin compaction and therefore in repressing transcription by decreasing the ability of TFs to bind to their recognition motifs (Deaton & Bird 2011). In summary, promoters of metazoan eukaryotes, especially mammals, are diverse in organization and composition of core promoter elements and thus

contribute to the regulatory specificity of a gene. In addition, proximal promoter elements provide further promoter specificity for gene expression as they can serve as tethering elements and determine the regulatory input from other, more distally located CREs (Calhoun et al., 2002; Maston et al., 2006; Deng et al., 2012).

1.1.3 Distal enhancers

Distal enhancers have a similar function to the historically defined proximal promoter elements but the terminological distinction between them is vague in the literature. This is illustrated by the fact that in many cases, proximal promoter elements and distal enhancers share the same set of TFs (Maston et al., 2006; Perry et al., 2010; Deng et al., 2012). However, distal enhancers act over long genomic distances of thousands of nucleotides from the promoter they regulate. They are commonly located in intergenic regions, or introns, of the genes they regulate or in neighboring genes (Lettice et al., 2003; Birney et al., 2007). Enhancers were initially reported 35 years ago in the context of remote viral sequences that could increase the expression of human genes (Banerji et al., 1981; Moreau et al., 1981). Since their discovery, enhancers have been functionally defined by their ability to activate transcription in transgenic reporter gene assays *in vitro* and *in vivo* (Arnold et al., 2013; Kvon et al., 2014), independent of their location or orientation relative to the tested promoter. The sequence of enhancers is typically a few hundred nucleotides in length and contains several binding sites for TFs. The composition of TF-binding sites determines the transcriptional output of a gene in response to developmental and physiological signals in a certain cell type (Spitz & Furlong 2012).

The importance of distal enhancers in gene regulation was recognized early through identified deletions or chromosomal translocations that did not disrupt the gene itself but resulted in gene loss-of-function phenotypes in patients. For example, deletions or translocations associated with beta-thalassemia remove an enhancer 60 kb upstream of the *beta-globin* gene and leads to failure to transcriptionally activate the *cis*-linked globin genes in erythroid cells (Kioussis et al., 1983; Forrester et al., 1990). Transgenic mouse studies finally established the enhancer as the major determinant region that activates gene expression at the locus, also referred to as the locus control region (LCR) (Magram et al., 1985; Grosveld et al., 1987).

Human mutations and transgenic studies have also provided clear evidence for distal enhancers contributing to distinct site- and stage-specific gene expression patterns during development. Sonic hedgehog (SHH), for instance, is expressed in the notochord, floor plate and anterior limb during development and mutations in the *Shh* gene are associated with severe malformations in humans, including defects of the forebrain (holoprosencephaly) (Roessler et al., 1996). However, mutations in TF-binding sites in a single distal enhancer, located 1 Mb upstream in the intron of a

neighboring gene, causes isolated polydactyly in humans and mice as a consequence of misexpression of *Shh* in the anterior limb bud (Lettice et al., 2003; Sagai et al., 2005).

In contrast, other reports have demonstrated that proximal promoter elements and *cis*-linked enhancers show partially or completely redundant expression patterns. Studies in *Drosophila* have indicated that redundant enhancers are crucial for the transcriptional robustness of a gene in response to varying environmental conditions during the course of development. Under normal conditions, the deletion of redundant enhancers has no consequences on gene expression and development whereas under stress conditions, such as elevated temperatures, enhancer loss causes sporadic loss of gene expression along with developmental defects (Perry et al., 2010; Frankel et al., 2010).

These aforementioned examples illustrate the widespread function of distal enhancers in activating gene expression or ensuring precise and robust gene expression. In addition, distal enhancers endow their target gene with additional TF-binding sites for incorporation with complex gene regulatory networks. Hence, combinatorial binding of TFs to distal enhancers facilitates accurate spatio-temporal gene regulation and expression patterns during development. Given their importance in gene regulation, one can assume that distal enhancers underlie a certain evolutionary constraint. However, the comparison of sequence conservation of proximal promoters and distal enhancers among mammals uncovered that distal enhancers are less conserved and evolved faster than the promoter sequences (Villar et al., 2015). As a result, rapid changes in remote enhancers of the genome are thought to serve as a source for a quick evolutionary adaption to acquire new gene expression patterns and functions (Shapiro et al., 2004; Chan et al., 2010).

1.1.4 Identification of enhancers

Genome-wide association studies have shown that more than 80 % of risk-associated variants lay outside of coding sequences, further emphasizing the significance of distal CREs in non-coding regions in human diseases (Manolio et al., 2009). Provided their importance in gene regulation and disease etiology, multiple efforts have been dedicated to identifying and functionally characterizing distal enhancers in the human genome. Enhancer prediction, based on conservation, has been used to select candidate enhancers and test them for reporter activity in transgenic mouse assays (Visel et al., 2008). However, sequence conservation alone is not sufficiently predictive as the evolutionary constraint of enhancers can vary depending on the tissue type (Blow et al., 2010).

New strategies and technological advances in genome-wide approaches employing chromatin immuno-precipitation (ChIP-seq) with antibodies against TFs and chromatin marks, mostly post-

translational histone modifications, have enabled easier screening for tissue-specific enhancers. TFs usually co-bind enhancers with p300-CBP activators, which are components of the transcriptional machinery (Vo & Goodman 2001). ChIP-seq utilizing p300 has been shown to be a powerful predictor of identify enhancers, even without knowing the specific TFs bound to them (Visel et al., 2009). Furthermore, several chromatin marks have been observed associated or enriched with active enhancers. H3K4 mono-methylation is linked to poised or active enhancers and H3K27 acetylation is correlated with active cis-regulatory regions (Heintzman et al. 2007; Creyghton et al. 2010).

The Encyclopedia of DNA Elements (ENCODE) project has systematically analyzed the human genome by mapping regions of transcription, chromatin accessibility, transcription factor binding and histone modifications (Dunham et al., 2012). Less than 3 % of the human genome accounts for protein-coding regions but the data suggests that the vast majority (80.4 %) of the genome features biochemical functions. Functional annotation provides further insight into gene regulation and demonstrates that many CREs are physically associated with one another as well as with gene expression (Dunham et al., 2012). These approaches and studies from individual loci have identified the widespread distribution of CREs across the mammalian genome, localized hundreds of kilo bases away from the genes they control. This raises the question of how CREs and enhancers identify their target genes to make certain the precise transcriptional activation and the establishment of complex spatio-temporal expression patterns during development.

1.1.5 Chromatin organization guides enhancer-promoter contacts

Promoter activation by distal enhancers requires contact between their DNA sequences, so that bridging of TFs and co-activators with the transcriptional machinery occurs (Banerji et al., 1981). Distal enhancers are brought into close physical proximity to the promoter via DNA looping (Dekker 2008; Bulger & Groudine 2011; Deng et al., 2012). This type of enhancer-promoter communication is evidenced by two technologies: 1) microscopy coupled to fluorescence *in situ* hybridization (FISH) with fluorescently-labeled DNA probes (DNA-FISH) for locating the relative position of loci within the nucleus (Langer-Safer et al., 1982; reviewed in Fraser et al. 2015) and 2) biochemical approaches, such as nuclear ligation assays, also referred to as chromosome conformation capture (3C), which directly measures the physical contact of DNA sequences in the nucleus (Cullen et al., 1993; Dekker et al., 2002). Recent progress in coupling 3C assays with high-throughput sequencing has provided new means to explore long distance regulatory interactions between enhancers and promoters on a genome-wide scale (reviewed in de Wit & de Laat 2012).

3C methods employ formaldehyde fixation to cross-link contacts between DNA and proteins, thereby producing a snapshot of physically linked chromatin within the intact nucleus (Figure 1a).

The chromatin is then fragmented with restriction enzymes that produce free ends of the cross-linked DNA. The ligation of these free ends generates a single DNA molecule of interacting DNA fragments, the so-called "ligation fragments". After reversal of formaldehyde cross-linking, the frequency of ligation fragments is determined within a cell population to yield a direct measure of the physical proximity within the nucleus. Various methods have been developed to detect ligation fragments that answer different questions regarding chromatin contacts (Figure 1). 3C assesses the interaction frequency between two previously defined DNA fragments by polymerase chain reaction (PCR) (one vs. one) (Job Dekker et al., 2002). Circular 3C followed by high-throughput sequencing (4C-seq) utilizes a PCR enrichment step for a predefined fragment in the genome (viewpoint) to discern all possible ligation partners and their frequencies (one vs. all) (Zhao et al., 2006). Finally, Hi-C identifies chromatin contacts in a genome-wide fashion by sequencing all ligation fragments at the same time (all vs. all) (Lieberman-Aiden et al., 2009).

Genome-wide analysis in mammals using Hi-C has revealed the structural compartmentalization of genomes. At the multi-mega base scale, the genome is subdivided into the so-called "A and B compartments", correlating generally with eu- and heterochromatin (Lieberman-Aiden et al., 2009). Subsequent Hi-C assays with increased sequencing depth to gain a higher resolution in the Hi-C data showed chromatin domains at the sub-mega base scale, termed topologically associating domains (TADs). TADs have a median size of 800 kb and they are separated by TAD boundaries. Sequences located within the same TAD contact each other with a much higher incidence than sequences of neighboring TADs (Dixon et al., 2012; Nora et al., 2012) (Figure 1b and Figure 1c). Importantly, the structural organization of the genome and the close proximity of DNA sequences within TADs are of great importance for contacts between genes and the corresponding CREs and are therefore crucial for gene regulation. The analysis of chromatin contacts further uncovered that long-range contacts are widespread throughout the genome. Most active genes involved in cell-specific functions are engaged in long-range interactions with an average of five enhancers, whereas highly expressed housekeeping genes do not have specific distal interactions (Dunham et al., 2012). Moreover, TADs significantly overlap with enhancer-promoter pairs that are co-regulated in a tissue-specific manner (Shen et al., 2012; Jin et al., 2013). Thus, compartmentalization into TADs provides a structural context for gene regulation by safeguarding promoters and corresponding CREs efficiently finding each other. Interestingly, previous studies using random insertions of a naive reporter gene indicated there was a broad distribution of enhancer activities within the genome and uncovered the regulatory domain structure of the mouse genome (Ruf et al., 2011; Tsujimura et al., 2015). Remarkably, TADs overlap to a large degree with these regulatory domains and the range of action of enhancers is limited by TAD boundaries. Further, enhancers can freely interact and promiscuously activate

genes within the spatially constrained TADs as the reporter gene responds to cis-regulatory information at any position and throughout extended regions (Symmons et al., 2014; Spitz 2016).

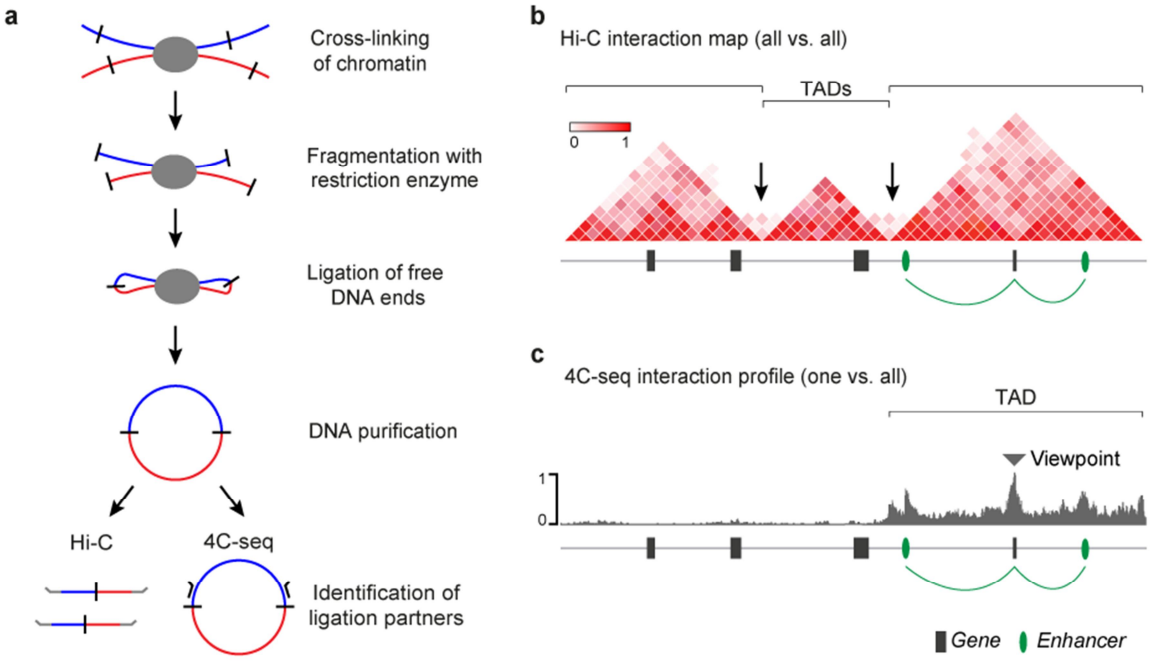


Figure 1: Hi-C identifies chromatin segmentation into TADs. a, Schematic of basic principle of 3C technology, including cross-linking, fragmentation and ligation of chromatin. After DNA purification, ligation partners are determined with different approaches. Hi-C technology includes biotin-labelling of restriction ends, DNA sharing and biotin pulldown after DNA purification (not shown). Subsequently, all ligation partners are detected by next-generation sequencing. 4C-seq makes use of an inverse PCR enrichment step to identify all ligation partners with a region of interest. The figure was adapted from Dekker et al. (2013). b, Schematic of Hi-C interaction map for a region of interest containing genes (black boxes) and enhancers (green ovals). Interaction frequencies (from 0 to 1) for each point in the region are marked in the triangular matrix. Genome-wide analysis discerned segmentation of the genome into TADs. TADs represent domains of frequently interacting chromatin that are separated by TAD boundaries (arrows). This image was adapted from Ong & Corces (2014). c, Schematic of a 4C-interaction profile with frequently interacting chromatin within a region of interest termed a viewpoint (triangle). Note that the 4C-interaction profile corresponds to the Hi-C predicted TAD with a high internal interaction frequency compared to regions outside the TAD. TADs provide a structural context for CREs by ensuring that enhancers and promoters efficiently find each other.

TAD positioning and boundaries are largely stable structures across cell types and species, which is consistent with their proposed role in genome organization and gene regulation (Dixon et al., 2012; Dixon et al., 2015; Vietri Rudan et al., 2015). However, conformational changes within TADs, including sub-TAD structures and specific enhancer-promoter interactions, are highly dynamic and likely account for tissue- and cell-type-specific gene activation and regulation during development (Montavon et al., 2011; Rao et al., 2014; Dixon et al., 2015). The information that is encoded within the three-dimensional chromatin structure is of great importance for understanding the function of our genome and gene regulation. Of note is how chromatin organization is established and how it contributes to gene regulation at the molecular level continue to be key questions in molecular biology.

1.1.6 Cohesin and CTCF as facilitators of long-range interactions

Cell-type-specific gene expression involves long-range chromatin interactions between enhancers and promoters that bind cell-type-specific TFs. These unique enhancer-promoter contacts are established and stabilized by protein-protein interactions. Large protein complexes, including transcriptional coactivators, such as Mediator, have been shown to physically and functionally connect enhancers and promoters. In addition, Mediator complex can form complexes via the architectural protein, cohesin, a key component of chromatin (Kagey et al., 2010). Cohesin is a large multi-protein complex able to accommodate two chromatin fibers within its ring-like core subunit. It mediates, for example, sister chromatid cohesion, which is essential for proper segregation of chromosomes during mitosis and meiosis (reviewed in Nasmyth & Haering 2009). ChIP-seq experiments in various cell types suggested that cohesin, Mediator- and tissue-specific TFs co-localize at active enhancers and promoters, providing evidence that cohesin and Mediator are usually cell-type-specific for enhancer-promoter interactions (Kagey et al., 2010; Schmidt et al., 2010; Faure et al., 2012). The depletion of cohesin abrogates these interactions and leads to gene misregulation (Zuin et al., 2014).

Of interest is that DNA looping does not strictly include the physical interaction between enhancers and promoters. Long-range chromatin interactions are also established between DNA elements that appear to be involved in the architectural organization of chromatin. For instance, promoters themselves are engaged in contacts with elements that bind the architectural protein, CCCTC-binding factor (CTCF), or CTCF together with cohesin. In contrast to the highly dynamic tissue-specific interactions, these particular interactions are stable across cell types and time points during development (Andrey et al., 2016). CTCF, a zinc-finger DNA-binding protein, was initially identified as a nuclear factor involved in transcriptional regulation, binding at proximal promoter elements and distal CREs in vertebrates (Lobanenkov et al., 1990; Baniahmad et al., 1990). Later, CREs that bind CTCF were described as insulator elements as they had the ability to block enhancer-promoter communication during *in vitro* reporter assays (Bell, West, & Felsenfeld 1999). In addition, CTCF is enriched at the transitions between active and repressive chromatin domains and can serve as a barrier insulator between these chromatin states (Cuddapah et al., 2009). The deletion of CTCF-binding sites within the *HoxA* locus results in the expansion of active chromatin into the repressive domain (Narendra et al., 2015).

However, growing evidence indicates that CTCF is a key component in the topological organization of chromatin. CTCF binds DNA with high affinity at several tens of thousands of binding sites within the genome and this usually coincides with binding of the architectural protein complex cohesin (Heger et al., 2012; Wang et al., 2012). As well, TAD boundaries have been found to be associated with constitutively bound and cell-type invariant CTCF-binding sites

(Dixon et al., 2012; Jin et al., 2013). Further, CTCF proteins can form homo- or multi-dimers, which are invoked in a model where chromosomal contacts are established through the combination of CTCF-DNA and CTCF-protein interactions (Pant et al., 2004; Dixon et al., 2012).

Finally, high resolution Hi-C maps led to the discovery of an additional feature of CTCF in chromatin folding. Strong interactions of two distantly located loci are visible as focal points in Hi-C maps and are what are referred to as DNA loops (Figure 2). The majority of CTCF-associated DNA loops at TAD boundaries and other long-range interactions within TADs are believed to depend on CTCF motif orientation and binding strength. Strong DNA loops between CTCF sites are formed between CTCF motifs in a convergent orientation (Gómez-Marín et al., 2015; Rao et al., 2014).

The observation that DNA loops correlate with the binding strength of CTCF and a preferred CTCF motif orientation led to the model of DNA-loop extrusion (Fudenberg et al., 2016; Sanborn et al., 2015). Polymer modeling suggests that cohesin complexes bind to chromatin and extrude the chromatin fiber to form a chromatin loop. The process of extrusion finally stops when the cohesin reaches CTCF proteins, which are bound at distally located motifs in a head-to-head configuration. The process of extrusion is likely to be dynamic, whereby the cohesin complex repeatedly extrudes chromatin, terminates at CTCF and falls off. This loop extrusion model explains the strong interactions seen between CTCF-associated TAD boundaries as contacts between convergent CTCF sites that are transiently stabilized. Moreover, it illustrates the high internal interaction frequency within TADs as the dynamic process of extrusion increases the probability of chromatin contacts on the extruded chromatin. Additionally, deletion experiments have demonstrated that single CTCF sites are required for DNA looping (Splinter et al., 2006; Sanborn et al., 2015). Related is that complete CTCF depletion in embryonic stem cells (ESCs) disrupts the TAD structure, which elevates the number of contacts between insulated TADs, resulting in gene misexpression (Nora et al., 2016). On the contrary, cohesin depletion leads to loss of interactions within TADs but does not affect the overall TAD organization (Zuin et al., 2014).

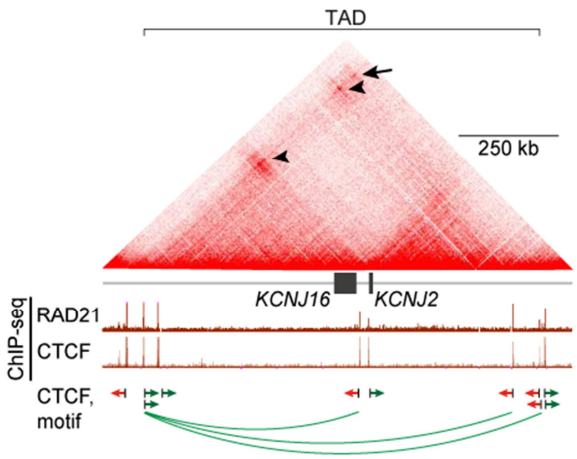


Figure 2: CTCF-mediated loop formation of chromatin. High-resolution Hi-C data from a human lymphocyte cell line (GM12878) at the *KCNJ2* and *KCNJ16* locus (Rao et al., 2014). ChIP-seq signal for cohesin (RAD21 subunit) and CTCF as well as CTCF motif orientation (red and green arrows) are shown. Data was retrieved from ENCODE and visualized in the UCSC genome browser. Note that CTCF and cohesin binding overlaps and that CTCF sites with convergent orientation are engaged in DNA loops (green lines). DNA loops, which appear as focal red dots in Hi-C data, are established between TAD boundaries (arrow) and CTCF sites within TADs (arrowhead).

Taken together, long-range DNA looping involves the action of architectural proteins, such as CTCF and cohesin. CTCF-mediated DNA looping, such as CTCF-CTCF loops found at TAD boundaries, contributes to TAD organization, which spatially constrains the range of enhancer-promoter contacts and is stable across different cell types. On the other hand, tissue-specific enhancer-promoter contacts involve a complex interplay between TFs, Mediator and cohesin proteins that act within a shorter range and in the pre-established context of TADs. Although these different proteins differentially contribute to chromatin organization, they play a crucial role in proper gene regulation. However, the dynamics of tissue-specific interactions within TADs and the hieratical relationship between dynamic and stable long-range interactions remain poorly understood.

1.2 The role of the transcription factor, SOX9, in development and disease

SOX9 was first discovered as the disease-causing gene in patients with campomelic dysplasia (CD) (Foster et al., 1994; Wagner et al., 1994). CD is a haploinsufficiency disorder that is characterized by an abnormal development of endochondral bone and weakened cartilage. Patients are born with bowing of long bones, especially of the lower legs, and other skeletal abnormalities, like hypoplastic scapulae and a small thoracic cage. A common feature in CD patients is the Pierre-Robin sequence (PRS), a condition with quite distinct facial features, e. g., a cleft palate, a setback tongue (glossoptosis) and a smaller lower jaw (micrognathia). Furthermore, underdeveloped cartilage of the upper respiratory tract induces breathing problems and is often life-threatening in infants. Interestingly, two-thirds of male CD patients are also diagnosed with disorders of sexual development (DSD), including complete male-to-female sex reversal or a combination of female and male reproductive systems (Wagner et al., 1994; Mansour et al., 1995). These observed phenotypes of abnormal cartilage and testis development reflect two functions of SOX9 in its regulation during development.

1.2.1 Molecular characteristics of SOX9

SOX9 belongs to a protein family of transcriptional regulators that contain a high-mobility group domain (HMG). The HMG domain is required for binding a DNA consensus motif at promoters or distal CREs (enhancers) in the genome. This DNA-binding domain was first described for the SRY protein (sex-determining region on the Y chromosome) and proteins that share high homology in the HMG domain are referred to as SOX (SRY-Box) proteins (Sinclair et al., 1990; Gubbay et al., 1990). SOX proteins can be further divided into subgroups based on the structural homology outside the HMG domain. SOX9, together with SOX8 and SOX10, belong to the SOXE subgroup

and share significant homologies in their self-dimerization and transactivation domains (Wegner 1999). These structural similarities result in functional redundancy depending on the temporal expression and amount of SOXE group proteins in a certain tissue. For instance, the double knock-out of *Sox9* and *Sox10* in oligodendrocytes causes apoptosis whereas individual knock-outs elicit normal development (Finzsch et al., 2008). Moreover, SOX proteins exhibit remarkable cross-talk with other transcription factors. SOX proteins can form partner complexes either with homologous or heterologous SOX proteins or TFs from other protein families based on binding site distribution at their target gene promoters or enhancers. Therefore, SOX proteins can act as transcriptional activators as well as repressors, depending on partner-protein expression (Kamachi, Uchikawa & Kondoh 2000).

1.2.2 SOX9 during chondrogenesis

SOX9 can be considered the key regulator of chondrogenesis. This became evident with the finding that haploinsufficiency results in chondrodysplasia in humans and mice (Wagner et al., 1994; Bi et al., 2001). Heterozygous *Sox9* knockout mice recapitulate most of the skeletal abnormalities observed in campomelic dysplasia syndrome. Mice die perinatally with cleft palate, hypoplasia and bending of many skeletal structures which derive from cartilage (Bi et al., 2001).

Cartilage prefigures the future elements of long bones that undergo endochondral ossification. Endochondral bone formation commences with the condensation of undifferentiated mesenchymal cells that subsequently differentiate into chondrocytes (Figure 3a). Chondrocytes secrete collagens and aggrecans into the extracellular matrix (ECM), forming the cartilage primordium that is surrounded by flattened cells, the perichondrium (Figure 3b). Following differentiation, chondrocytes in the cartilage primordium undergo rapid proliferation to form the growth plate, consisting of clonal cell stacks that drive the linear growth of skeletal elements. In parallel, centrally located chondrocytes undergo maturation whereby they become hypertrophic and secret type X collagen into the cartilaginous matrix (Figure 3c). These hypertrophic chondrocytes finally undergo apoptosis and chondroclasts degrade the mineralized cartilage via matrix metalloproteinase 9 (MMP9) (Vu et al., 1998). Simultaneously, perichondral cells differentiate into osteoblasts forming the cortical bone. The degradation of hypertrophic chondrocytes is accompanied by the invasion of blood vessels and osteoblasts into the inner cavity which finally leads to the primary ossification center at the bone shaft. Proliferation and differentiation of chondrocytes continues at the articular ends and results in the distal displacement of the two growth plates, which enables longitudinal bone growth (Hunziker, 1988) (Figure 3d). Postnatally, two additional secondary ossification centers emerge at the end of the bone to form the final bone element (reviewed in Mundlos & Horn 2014 and Long & Ornitz 2013)

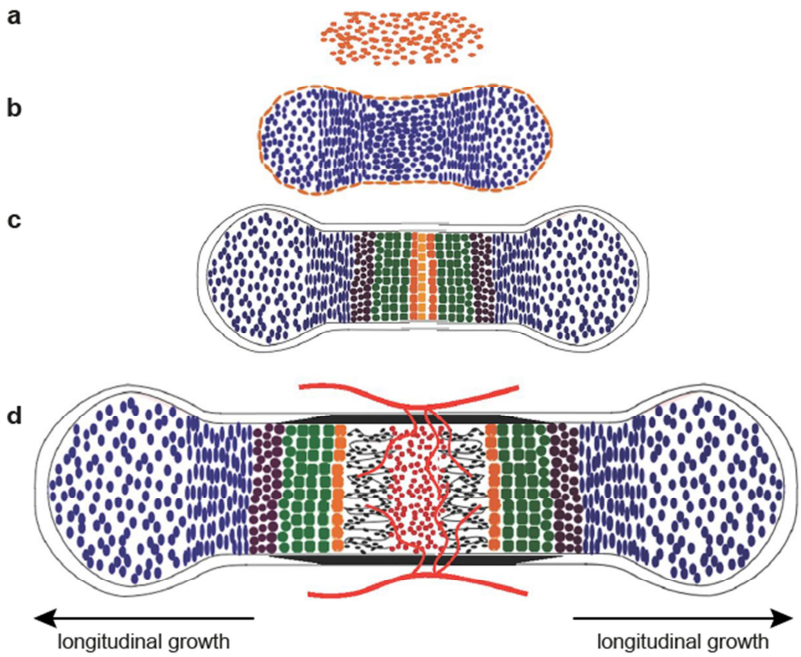


Figure 3: Development of endochondral bone. a, Condensation of undifferentiated mesenchymal cells (orange). b, Cells in the center differentiate into chondrocytes (blue), expressing the transcription factor, Sox9, to form the cartilage primordium. Flattened cells in the periphery form the perichondrium (orange). c, Cells in the center of the cartilage primordium further undergo maturation through prehypertrophy (purple), hypertrophy (green) and terminal hypertrophy (orange). Chondrocytes in the periphery continue to proliferate and form stacked chondrocytes within the growth plate. d, Terminal hypertrophic cells undergo apoptosis, which is accompanied by invasion of blood vessels (red lines) and osteoblasts from the perichondrium. Osteoblasts form the bone collar (black) and the inner bone marrow cavity is populated by trabecular bone (black) and hematopoietic cells (red). Longitudinal growth from the remaining growth plates is indicated. Figure adapted from Long & Ornitz (2013).

SOX9 is the earliest known nuclear factor and is already expressed in mesenchymal condensations prior to chondrogenesis, and is necessary for subsequent differentiation into chondrocytes. The conditional knockout of *Sox9* in prechondrogenic limb mesenchyme prevents chondrogenesis, resulting in a complete absence of limb cartilage and bone in mice (Akiyama, 2002). In addition to this early function, *Sox9* expression is maintained in chondrocytes and is necessary for their proliferation and survival (Wright et al., 1995; Akiyama, 2002). In proliferating chondrocytes, SOX9 induces the expression of key ECM genes for chondrocytes, e.g., type-II collagen encoded by *Col2a1* ((Bell et al., 1997). SOX9 homo-dimers bind directly to an enhancer, located in the intron of *Col2a1* and cooperative binding of an additional partner complex, consisting of SOX5 and SOX6, activates gene expression (Ikeda et al., 2004; Lefebvre et al., 1997). In contrast to its role as a trans-activating factor, SOX9 can directly repress target genes, such as *COL10A1*, that are required for further maturation of chondrocytes into hypertrophic chondrocytes (Leung et al., 2011). Additionally, SOX9 can block the activity of the osteogenic master regulator, Runt-related transcription factor 2 (RUNX2) (Stricker et al., 2002; Zhou et al., 2006). During later stages of chondrocyte maturation, *SOX9* expression is switched off, permitting the expression of

hypertrophic markers, like *COL10A1,* vascularization and bone marrow formation (Hattori et al., 2010).

SOX9 exhibits a complex spatio-temporal expression pattern during endochondral bone development, and this necessitates tight control at the transcriptional level and intergration of multiple signaling pathways. Bone morphogenic proteins (BMPs) positively regulate the initiation of chondrogenesis. Misexpression of BMPs in the chicken limb bud induces ectopic chondrogenesis (Duprez et al., 1996; Zou et al., 1997), and, conversely, perturbation of BMP signaling by dominant-negative BMP receptors inhibits the formation of cartilage (Kawakami et al., 1996). WNT and FGF signaling have been shown to modulate chondrogenesis, as well. Fibroblast growth factors (FGFs), secreted from the apical ectodermal ridge (AER), are primarily needed for limb bud outgrowth but are also necessary for the cell fate decisions of mesenchymal cells to engage in chondrogenesis (Berge et al., 2008). On the other hand, WNT signals from the limb bud ectoderm can block *Sox9* expression and therefore restrict cartilage formation to the core of the limb bud mesenchyme (Day et al., 2005; Rudnicki & Brown 1997).

Signaling pathways function via intracellular mediator proteins, such as beta-catenin for WNT signaling, and by the action of TFs that regulate *SOX9* expression on the genomic level. However, the control of *SOX9* expression via TFs at their respective enhancer elements is not well understood. Identification of chondrogenic-specific enhancers would further provide a better understanding of how signaling pathways incorporate their repressive or activating function at the genomic level. *In vivo* reporter assays using *LacZ* expression have demonstrated that the mouse *Sox9* promoter is sufficient to drive expression in the spinal cord and hindbrain during embryonic development but not in cartilage progenitors. This suggests the importance of the additional remote enhancer in *Sox9* regulation (Bagheri-Fam et al., 2006). For instance, an enhancer element, approximately 70 kb upstream of *Sox9*, is sufficient to drive reporter expression in most *Sox9*-positive somatic tissues, including the cartilage primordium. Although the enhancer is not absolutely required for endogenous *Sox9* expression, analysis shows that the enhancer activity in the cartilage primordium depends on the activity of SOX9 homodimers, indicative of a SOX9 auto-regulatory mechanism during cartilage development (Mead et al., 2013).

1.2.3 SOX9 during sex determination

A well-studied function of *SOX9* is the determination of the mammalian bipotential gonad to a testicular fate. Haploinsufficiency in humans causes partial or complete male-to-female sex reversal, whereas mice lacking one intact allele of *Sox9* exhibit normal sexual development (Wagner et al. 1994; Bi et al. 2001). However, the conditional knockout of both alleles in the

bipotential gonad in mice induces complete male-to-female sex reversal (Barrionuevo, 2005). Furthermore, *SOX9* gene duplication has been linked to testis development in female patients and Odsex mouse mutants that constitutively express *Sox9* in the gonad show XX sex reversal (Huang et al., 1999; Bishop et al., 2000). SOX9 is thus sufficient to induce testis development in the absence of SRY, and that the usual requirement of SRY for testis development can be bypassed by SOX9.

The mammalian testes and ovaries arise during early embryonic development from a sexually indifferent primordial structure, the genital ridge. It is constituted by a regional thickening of the coelomic epithelium on the ventro-medial surface of the embryonic kidney, the mesonephros. During genital ridge outgrowth, cells from the coelomic epithelium proliferate and expand. Concurrently, the underlying mesenchyme enlarges through proliferation of mesenchymal cells and by cell recruitment from the mesonephros and coelomic epithelium (Svingen & Koopman 2013). In mice, primordial germ cells migrate from the hindgut of the embryo between embryonic stages E8.0 to E10.5 and colonize the genital ridge, later becoming spermatocytes or oocytes (Anderson et al., 2000). Within the mesenchyme, somatic cells subsequently differentiate into sertoli cells in males or granulose cells in females, the supporting cell lineages of the gonads.

The bipotential genital ridge is established in mice at E10.5 by the steroid factor 1 (SF1) and the Wilms tumor-suppressor 1 (WT1) (Luo, Ikeda, & Parker 1994; Kreidberg et al., 1993). Notably, both factors initiate a low expression of *Sox9* before sex determination and the onset of *Sry* in males (Wilhelm 2002; Kidokoro et al., 2005). In the presence of the Y chromosome, the supporting cells of the genital ridge begin to express *Sry*. Subsequently, *Sox9* is upregulated in these cells, which depend on the presence of SRY and SF1 (Sekido et al., 2004). At later stages, the expression of *Sox9* is maintained independent of SRY by a SOX9 auto-regulatory mechanism and a positive feed-forward loop via FGF9 signaling (Sekido & Lovell-Badge 2008; Kim et al., 2006). With this, SOX9 directly activates genes required for further differentiation of sertoli cells and male gonad development. For example, SOX9 directly binds as a homo-dimer at the promoter of the prostaglandin D synthase (*Ptgds*) (Wilhelm et al., 2007) or indirectly by protein-protein interactions via SF1 to an enhancer upstream of the anti-Müllerian hormone (*Amh*) promoter (De Santa Barbara et al., 1998). Contrarily, the absence of SRY in XX supporting cells of the genital ridge results in accumulation of ovarian specific signals and repression of SOX9. WNT4 and R-spondin-1 (RSPO1) are expressed at early stages and both are ligands of the canonical WNT/beta-catenin signaling pathway that promotes the ovarian fate. Furthermore, the accumulation of beta-catenin can block testis development by repressing SOX9 activity (Maatouk et al., 2008).

These male and female signaling pathways activate sex-specific genes that primarily modulate *SOX9* expression levels. The key to inducing the male specific pathway is the upregulation of *SOX9*

above a certain threshold at the right time point (Sekido & Lovell-Badge 2009). *In vivo* reporter assays in mice led to the identification of an enhancer element that integrates the molecular upstream events in the developing male gonad into transcriptional activity of *Sox9*. A well-characterized testis-specific enhancer (*Tes*), 14 kb upstream of *Sox9*, recapitulates the *Sox9* expression pattern in transgenic reporter assays. The *Tes* enhancer contains several binding sites for SF1 and SRY that synergistically regulate enhancer activity as well as binding sites for SOX9 itself. *Tes* activity depends on endogenous SOX9, indicating that SOX9 regulates itself to maintain a high expression level even in the absence of SRY (Sekido & Lovell-Badge 2008). A core 1.4 kb region in this enhancer (*Tes* core element, *Tesco*) is crucial for enhancer function and is well-conserved in mammals. However, the human *TES* sequence is not able to drive transgenic reporter expression in the developing male gonad in mice, the functional conservation of *Tes* therefore being questionable (Sekido & Lovell-Badge 2009). As well, no mutations associated with disorders of sexual development have been identified in the human *TES* sequence (Georg et al., 2010). These observations suggest the existence of other regulatory elements in humans controlling testis expression of *SOX9*.

1.2.4 SOX9 function beyond development

SOX9 functioning is not only restricted to cartilage development and sex determination. The gene has a complex spatio-temporal expression pattern in various other tissues and organ systems during development, i.e., the heart, kidney, pancreas and central nervous system (Wagner et al., 1994; Ng et al., 1997). *Sox9* knockout mice also showed hypoplastic endocardial cushions that comprise the primordia of the valves and septa in the heart. In this context, loss of *Sox9* in the heart inhibits the migration of endothelial cells (Akiyama et al., 2004). Further, genetic ablation studies suggest a functional role of SOX9 for homeostasis in adult organs and adult stem or progenitor cells. For example, *Sox9* is expressed in epithelia hair follicle stem cells and is heavily involved in proliferation and maintenance of the hair stem cell niche. Conditional knockout of *Sox9* in the skin of mice results in complete lack of external hair (Vidal et al., 2005). A Similar function has been observed in the intestinal epithelium. Here, *Sox9* is expressed in progenitor cells and in Paneth cells of the epithelium, and its inactivation inhibits differentiation and leads to dysplasia of the local stem cell compartment (crypts) of the intestinal epithelium (Bastide et al., 2007).

1.3 Structural variations in the human genome

The DNA sequence of genomes is constantly changing, which contributes to phenotypic variation and consequently allows organisms to evolve and adapt to changing environmental conditions. In humans, variations have been detected and described at several levels that range from single nucleotide polymorphisms (SNPs) up to cytogenetically recognizable segments with extreme size, which affect parts of chromosomes (Jacobs et al. 1992; The International HapMap Consortium 2005). However, the phenotypic effects of genomic variations represent a broad spectrum that can range from no phenotypic consequences to adaptive traits or disease-causing variations (Perry et al. 2007; Pinto et al. 2010).

The development of genome-wide approaches, such as comparative genome hybridization (CGH) and whole-genome sequencing (WGS) through next-generation sequencing (NGS), are commonly employed to detect structural variations (SVs) in diagnostic settings (Redon et al. 2006; Campbell et al. 2008). SVs of intermediate size, ranging from kb to Mb scale, including deletions, duplications and inversions of DNA segments, as well as translocations, have been recognized to contribute substantially to natural human genetic variations (Sebat et al., 2004). For instance, deletions and duplications (termed ”copy number variations”; CNVs) are believed to account for 5 to 10 % of the overall variability in the human genome (Zarrei et al., 2015).

On the other hand, SVs have been linked to an ever-rising number of human genetic disorders, including intellectual disabilities, congenital malformations and cancer (Campbell et al., 2008; Kurth et al., 2009; Craddock et al., 2010). A major challenge for clinical geneticists is to distinguish pathogenic SVs from alterations that represent natural variations among human genomes. The molecular mechanisms whereby SVs can foster phenotypic diversity or induce pathological phenotypes are varied and in many cases, can have multiple effects. One of the most recognized disease mechanisms for deletions and duplications is the alteration of the copy number of a gene that is sensitive to its gene dosage (Lupski et al., 1992). However, this gene-centered view is incomplete and does not apply for all genes as duplications and deletions often comprise multiple genes (Klopocki et al., 2012). Furthermore, genome-wide diagnostic analysis led to the discovery of a previously unknown class of SVs connected to human disorders, affecting genes as well as large parts of non-coding regions of the genome (Klopocki & Mundlos 2011). Interestingly, large genomic deletions or duplications can be associated with pathological phenotypes that are very different from those associated with null mutations of genes themselves (Kurth et al., 2009; Lupiáñez et al., 2015). Importantly, the continuous detection of SVs requires a fundamental knowledge of their potential pathological effects.

1.3.1 Structural variations interfere with gene regulation and TAD structure

Understanding long-range enhancer-promoter interactions has provided researchers with the proper framework for interpreting SVs. 'Enhanceropathies', induced, for example, by deletions of distant CREs, can affect target gene expression and illustrate their functional importance for gene regulation (Smith & Shilatifard 2014). Another disease mechanism, which is potentially brought on by all types of SVs, is positional effects. Deletions and translocations, for instance, can lead to the deletion or disconnection of enhancer-gene contacts but also, place an enhancer or gene in a new *cis*-regulatory environment. Through this, genes can adopt new enhancers and therefore acquire novel spatio-temporal expression patterns, which can result in pathological phenotypes (Lettice et al., 2011). These observations emphasize that SVs interfere with gene regulation and that modifications to the non-coding part of the genome has a great influence on phenotypic diversity.

Importantly, the identification of the three-dimensional genomic organization opened another avenue for studying SVs and their impact on gene regulation. TADs form domains along chromosomes and build a regulatory backbone within the genome. SVs can reshuffle this organization and disrupt TAD structure. Deletions of TAD boundaries can modify the domain structure by allowing enhancers from one TAD to ectopically activate genes in another TAD. Genomic deletions overlapping with TAD boundaries, can explain a substantial fraction of this type of copy-number variation and associated human disease (Ibn-Salem et al., 2014). Furthermore, the relevance of TAD integrity for gene regulation and human disease has been shown at the *EPHA4* locus, where deletions and inversions cause disruption or repositioning of TAD boundaries (Lupiáñez et al., 2015). At this locus, a cluster of limb enhancers is normally located within the same TAD as its target gene, *EPHA4*. The deletion of large parts of the *EPHA4* TAD and the boundary to the neighboring TAD leads to ectopic interactions of the *EPHA4* limb enhancers and *PAX3* of the neighboring TAD. These ectopic interactions result in misexpression of *PAX3* and a limb-associated phenotype observed in human patients. However, this adoption in gene expression only takes place if the TAD boundary is disrupted by the deletion. Similarly, inversions encompassing a TAD boundary of the *EPHA4* TAD and the same set of limb enhancers misplaces them in front of the *WNT6* gene in the neighboring TAD, causing misexpression and a skeletal-limb phenotype, known as F-syndrome.

Medical clarification of SVs is challenging based on the multiple effects they can have on the gene copy number and/or gene regulation. The implementation of the TAD structure into such interpretations has been shown to better predict the pathogenic effects of SVs and the identification of the underlying disease-causing gene. Further understanding the mechanisms by

which SVs modify and shape the structural organization of the genome will aid in determining the effect of SVs on gene regulation.

1.3.2 Structural variations at the SOX9 locus interfere with SOX9 cis-regulation

SOX9 has versatile functions and exhibits a complex tissue-specific expression pattern during development. The pattern is believed to be controlled by enhancers located around the gene, with each regulating a subset of the *SOX9* expression domains (Bagheri-Fam et al., 2006). Interestingly, the locus surrounding the *SOX9* gene on chromosome 17 in humans is characterized by a large gene desert. This region of 2 Mb upstream and 0.5 Mb downstream of *SOX9* lacks any protein-coding gene. However, SVs mapped to this gene desert have been associated with partial or complete disease-mimicking loss-of-function *SOX9* mutations (Figure 4a). Translocation breakpoints mapped in a cluster between 50 to 375 kb upstream of *SOX9* are linked to campomelic dysplasia and sex reversal in females (Pfeifer et al., 1999). A similar loss-of-function phenotype has been described for a large 1.5 Mb deletion upstream of *SOX9*, which features all aspects of campomelic dysplasia, including sex reversal in male patients (Pop et al., 2004). As a consequence of the similarities of these phenotypes to mutations disrupting the *SOX9* gene, it is likely that translocations result in the disconnection and deletions in the loss of important enhancer elements that regulate *SOX9* expression (Klopocki & Mundlos 2011). Moreover, the significance of individual enhancers was demonstrated in patients with isolated PRS that carried either a 75 kb deletion or a point mutation in a highly conserved element located 1.44 Mb upstream of *SOX9* (Figure 4a). Through mouse reporter assays, it was found that this region is able to drive tissue-specific expression in the developing jaw and that the point mutation interferes with binding of a potential transcriptional activator. Translocation breakpoints that map to a cluster between this enhancer and *SOX9* have been linked with PRS, as well (Benko et al., 2009). These deletions and translocations show that *SOX9* regulation is influenced by a complex *cis*-regulatory landscape containing multiple enhancer elements in the gene desert upstream of *SOX9*.

An interesting scenario involving SVs is the overlapping and recurrent tandem duplications that map to the gene desert upstream of *SOX9*. Although these duplications cover the same genomic region, they lead to diverse phenotypic outcomes. Genomic duplications within the gene desert as along with duplications of the *SOX9* gene itself result in female-to-male sex reversal, potentially indicating a dosage effect of duplicated CREs in this region (Benko et al., 2011; Kim et al., 2015) (Figure 4a). In contrast, duplications that completely overlap with these duplications but extend further centromerically towards neighboring genes, *KCNJ2* and *KCNJ16*, have no effect on sexual

development though cause Cooks syndrome, a congenital limb malformation characterized by partial or complete aplasia of nails and brachydactyly of all finger and toes (Figure 4). The brachydactyly phenotype is typified by missing middle phalanges and elongated terminal as well as proximal phalanges (Kurth et al., 2009). Although these duplications include *KCNJ2* and *KCNJ16* at the centromeric end, these genes have so far been excluded from involvement in the described limb phenotype. *SOX9*, on the other hand, has been shown to be strongly expressed in the anlagen of developing claws in mouse embryos and that it has a vital part in chondrogenesis and skeletal growth. Therefore, alteration of the copy number of enhancers, induced by duplications, has been suggested to induce misexpression or overexpression of *SOX9* in a stage- and/or site-specific fashion (Kurth et al., 2009).

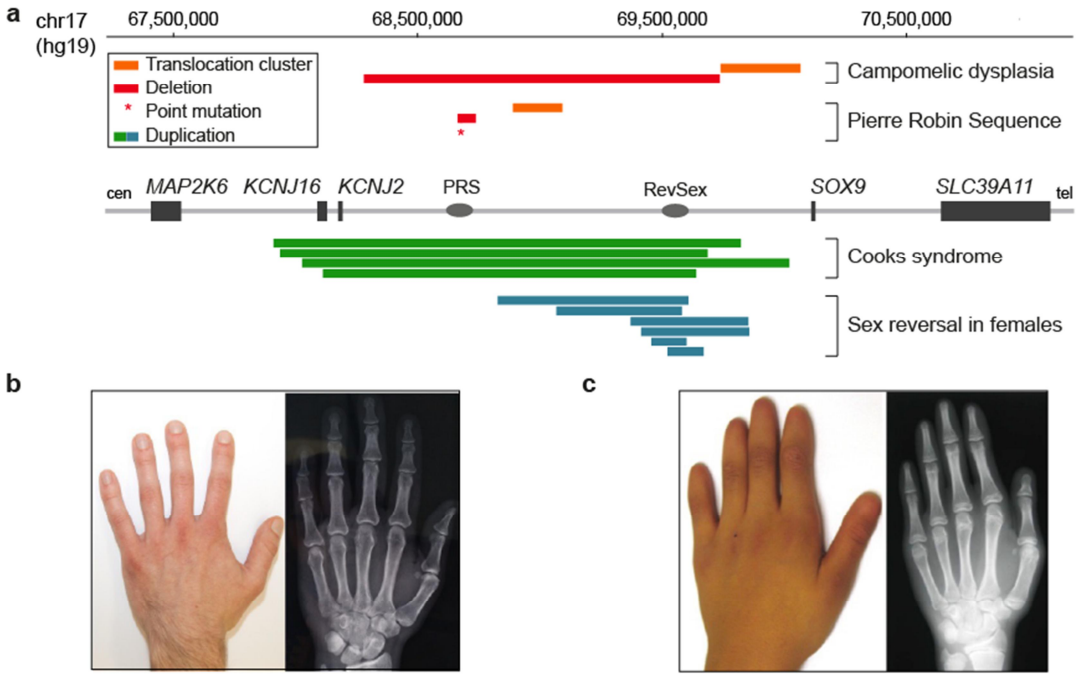


Figure 4: Structural variations and associated disease at the *SOX9* locus. a, Structural variations mapping to the 2 Mb gene desert downstream of *KCNJ2* and upstream of *SOX9*. Deletion in red (Pop et al., 2004) and translocations in orange (Pfeifer et al., 1999) proximal to *SOX9* result in campomelic dysplasia or PRS when located more distally to *SOX9* (Benko et al., 2009), mimicking partial or complete loss-of-function phenotypes of *SOX9*. Overlapping duplications (green and blue) are associated with either Cooks syndrome (Kurth et al., 2009) or female-to-male sex reversal (Benko et al. 2011; Kim et al. 2015). PRS enhancers (Benko et al., 2009) and RevSex linked to sex-reversal duplications (Benko et al., 2011; Kim et al., 2015) are indicated as ovals in the linear genomic view. Schematic adapted from (Klopocki & Mundlos 2011). b and c, images and radiographs of hands of healthy control (b) and a patient with Cooks syndrome (c). Cooks syndrome is associated with brachydactyly and nail aplasia. Clinical photographs of the Cooks syndrome patient was provided by Ingo Kurth, and that of healthy control hand were provided by the author; control radiograph was supplied by Daniel M. Ibrahim.

Nevertheless, the phenomenon of overlapping duplications, which contain a similar set of CREs, leading to completely different phenotypes cannot be elucidated with the current knowledge. All Cooks syndrome duplications include the minimal critical region (RevSex) that is described for all

sex-reversal duplications (Benko et al., 2009; Kim et al., 2015). How duplications can result in isolated sex reversal on the one hand and isolated skeletal phenotypes without sex reversal on the other is not understood. It is conceivable that the duplication of CREs affect the expression of *SOX9* or genes in the vicinity of the genomic alteration, thus resulting in modified gene expression. Establishing the different pathogenic mechanisms of these duplications at the *SOX9* locus will further provide a better understanding of how SVs influence gene regulation.

1.3.3 Engineering structural variations in the mouse genome

The mouse is the classical model system to study mammalian development and gene function. Gene knockouts in the mouse genome through targeted mutations provide researchers with the necessary tools to investigate the effect of gene mutations observed in human diseases (Hall, Limaye, & Kulkarni 2009). The standard gene-centered view of our genome led to the emergence of international consortia that provide specific gene knock-out alleles to the wide research community (Skarnes et al., 2011). On the contrary, investigation of SVs and their effects on gene regulation have been extremely limited, in part based on the difficulty of approaching such large intervals with existing targeting strategies. Studies on SVs originally relied on spontaneous mutations or mutations induced by forward-genetic screening using irradiation (Lyon et al., 1996). However, recent advances in alternative strategies and new technologies have resulted in the development of efficient reverse-genetic engineering approaches to bring about targeted SVs observed in human disorders (Wu et al., 2007; Ruf et al., 2011; Kraft et al., 2015).

1.3.4 In vivo Cre/loxP recombination and transposition

Trans-allelic targeted meiotic recombination (TAMERE) is an *in vivo* method for engineering mice with structural variations, such as tandem duplications and deletions (Hérault et al., 1998). The system relies on Cre-mediated recombination between *loxP* sites located on the two parental chromosomes. Cre recombinase induces an unequal recombination of the region between the *loxP* sites, resulting in one chromosome carrying a duplication and the other the corresponding deletion (Figure 5a). The technique requires the targeting of the *loxP* sites to the desired genomic position followed by two breeding steps to obtain triple-transgenic mice with the two *loxP* site in a trans configuration and a Cre-driver under control hypoxanthine-guanine-phosphoribosyl-transferase (*Hprt)* (Tang et al., 2002). The Cre recombinase expressed from the *Hprt::Cre* transgene is active in the germline of transgenic animals, hence offspring are produced with the desired duplication or deletion when mated with wild-type mice (Ruf et al., 2011) (Figure 5b).

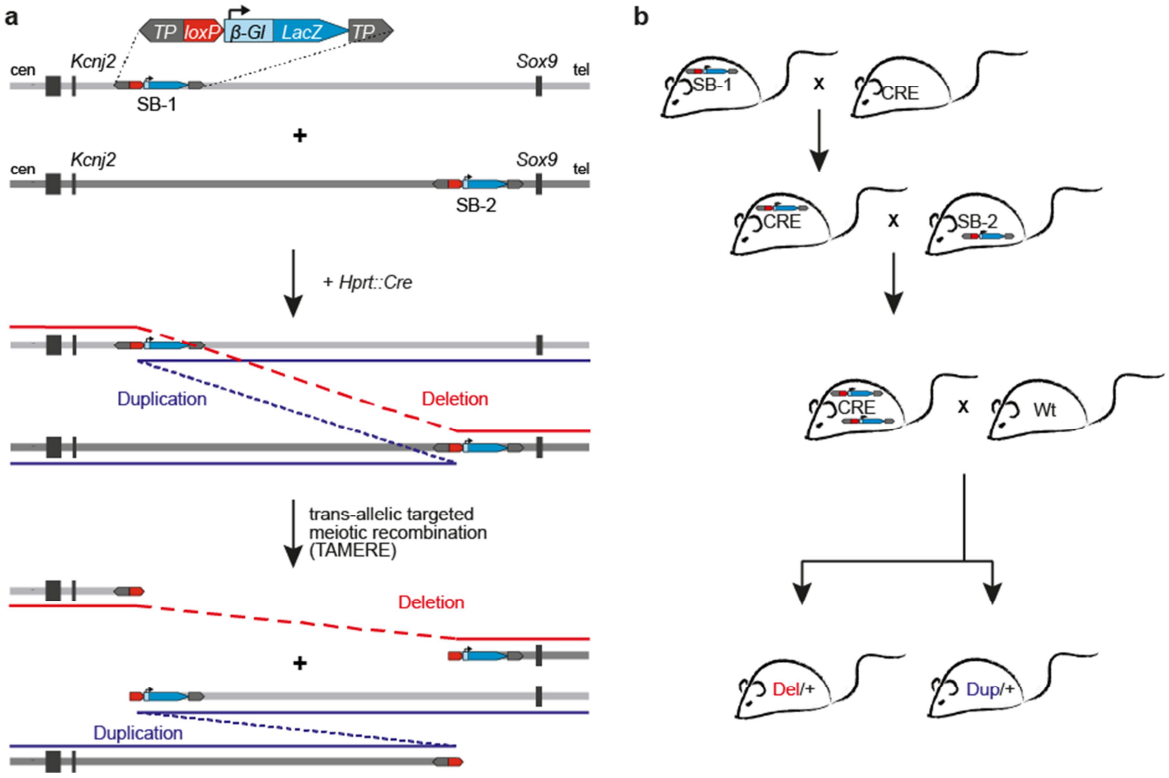


Figure 5: *In vivo* Cre/*loxP* recombineering for generation of duplications and deletions. a, Close-up view of the Sleeping Beauty (SB) transgene containing a *LacZ* reporter gene under the control of the human *beta-globin* promoter (blue), *loxP* site (red) and inverted repeats of SB transposon (grey). Trans-allelic recombination of two *loxP* sites on the parental chromosome is mediated by a Cre recombinase under the control of the *Hprt* promoter. The resulting unequal recombination between homolog chromosomes produces a deletion and duplication of the region between the *loxP* site. b, Breeding strategy to obtain triple-transgenic mice with the desired chromosomal rearrangement.

The engineering of serial duplications or deletions at a particular locus would necessitate several targeting steps to introduce an additional *loxP* site. A more efficient chromosomal tagging and engineering strategy was developed by combining *loxP* integrations with a SB-based transposition system (Ruf et al., 2011). The system described in Ruf et al. consists of a transgene containing a *loxP* site and a *LacZ* reporter gene that are both flanked by SB-transposons (Figure 6a). Existing mouse lines carrying the SB transgene are then used to generate new insertions *in vivo*. The preference of the SB transposon to integrate close to its previous location permits the generation of several mouse lines carrying a SB insertion within the locus of interest. To initiate remobilization, mice carrying an SB insertion are crossed with those carrying a transgene containing the hyperactive SB transposase under the control of the *protamin-1* promoter (*Prm1::HSB16*). The transposase is expressed exclusively in the male germline and mobilizes the SB transposon to new locations in a “cut and paste” manner. Mating of transgenic males with female wild-type mice will therefore produce a series of new SB insertions, each available for a series of new duplications or deletions (Ruf et al., 2011) (Figure 6b). Moreover, the system grants the ability to detect regulatory activity along chromosomes. The *LacZ* reporter gene is governed by the minimal promoter region of the human *beta-globin* promoter, which shows no activity

alone but responds as a sensor of regulatory activity that is triggered at the insertion point. The analysis of multiple SB insertions in the region of interest thereby allows the identification of genomic regions under enhancer control *in vivo* (Uslu et al., 2014; Symmons et al., 2014).

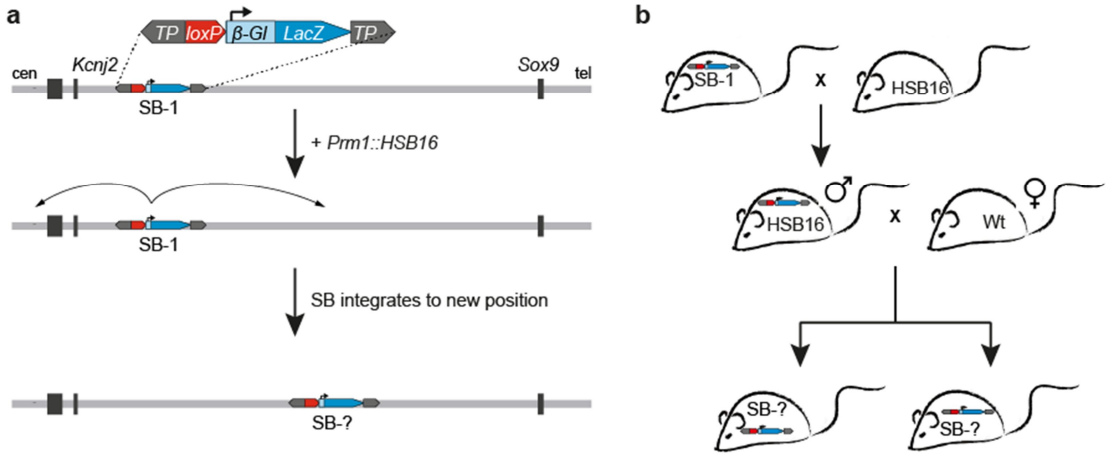


Figure 6: Transposon-mediated *in vivo* transposition. a, SB insertions jump to new genomic positions mediated by the activity of the hyperactive SB transposase and inverted repeats of the SB transposons flank the transgene (*LacZ* reporter and *loxP* site). b, Breeding strategy to obtain new SB insertions *in vivo*. Mice with new insertions are available to produce new genomic rearrangement by *in vivo* Cre/*loxP* recombination. The *LacZ* reporter responds to the regulatory activity on the integration site.

1.3.5 CRISPR/Cas9-mediated genome editing

The discovery of the CRISPR/Cas9 system and its application in various vertebrate and mammalian model systems dramatically revolutionized the field of genomic editing (Xiao et al., 2013; Cho et al., 2013; Kraft et al., 2015). It was first discovered in the bacterial genome, consisting of clustered regularly interspaced short palindromic repeats (CRIPSR) and Cas9 nucleases (Ishino et al., 1987; Jansen et al., 2002; Garneau et al., 2010). Bacteria use this system to protect their genome upon bacteriophage infection (Barrangou et al., 2007). Sequence-specific RNAs form a complex with Cas9 nuclease to destroy bacteriophage DNA. Ultimately, the recent adaption and application of the system to mammalian systems has supplied researchers with new possibilities for inducing targeted mutations in genomes.

As described earlier, the induction of SVs by conventional targeting procedures is time-consuming - it relies on designing targeting constructs and multiple breeding steps. CRISPR/Cas9 technology, however, enables the generation of SVs directly in mouse ESCs (Kraft et al., 2015). Using two sgRNAs at two distal genomic sites causes double-strand breaks that are mediated by the Cas9 nuclease (Figure 7). The error-prone repair mechanism of non-homologous end joining (NHEJ) subsequently carries out the rearrangement of the intermediate DNA fragment by direct ligation of DNA fragments (reviewed in Sfeir & Symington 2015). Thus, SVs of different sizes and types, like deletions, inversions and duplications, can easily and efficiently be obtained in ESCs (Kraft et

al., 2015). Single ESC clones carrying the desired SVs can be identified and used to produce chimeric mice or for retargeting to introduce additional mutations.

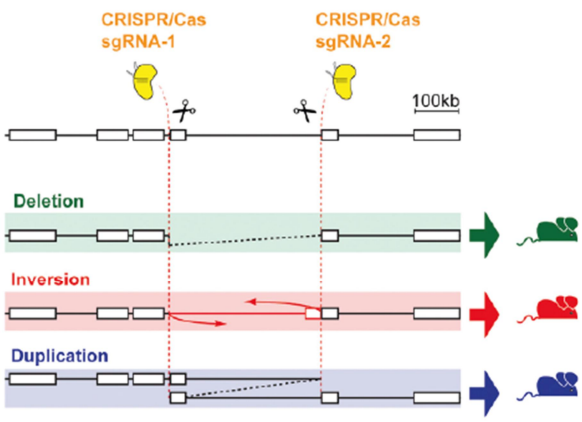


Figure 7: Engineering SVs using CRISPR/Cas9 technology. Two sgRNAs guide a Cas9 nuclease (yellow) specific to two distal genomic positions in mouse ESCs. Induced double-strand breaks are repaired by NHEJ that induces genomic rearrangement of the intermediate fragment. ESC clones with the desired SVs, such as deletions, inversions or duplications, are selected and utilized to generate transgenic mice. Figure taken from Kraft et al., (2015).

2 Aim of study

SOX9 is a key developmental regulatory TF embedded in a large gene desert populated by numerous described CREs that control its complex spatio-temporal expression pattern. Interestingly, overlapping tandem duplications mapping to the gene desert upstream of *SOX9* have been linked to either Cooks syndrome or female-to-male sex reversal. However, little is known about the mechanisms leading to these diverse phenotypic effects. The present research seeks to identify the disease mechanism of Cooks syndrome and sex reversal associated with duplications at the *SOX9* locus.

The physical interaction of CREs with their target promoter is crucial for gene regulation and TADs have been demonstrated to provide the mammalian genome structural context in which these interactions occur. SVs, like duplications, can interfere with the TAD organization and consequently gene regulation. Therefore, and central to this work, is the characterization of the spatial and functional organization at the *SOX9* locus in the human and mouse genome. To achieve this, chromatin interactions were first investigated with 3C approaches (4C-seq and capture Hi-C) to define TAD organization at the *SOX9* locus in different human cell types and mouse tissues at E12.5. Furthermore, the regulatory activity at the locus was analyzed *in vivo* using the SB transposition system. Several mouse lines carrying SB insertions with a regulatory sensor (*LacZ* reporter) were generated by genome targeting or *in vivo* transposition to investigate the regulatory domain structure. Whole mount LACZ staining at E12.5 was used to assess the regulatory potential at the locus and was then compared to the spatial organization.

Finally, to understand the impact of duplications on gene regulation and TAD organization, the corresponding duplications for Cooks syndrome and sex reversal were remodeled in the mouse genome. Tandem duplications at the locus were generated with SB insertions for *in vivo* Cre/*loxP* recombination or applying CRISPR/Cas9 technology. Mouse limb buds at E12.5 as well as patient cells with Cooks syndrome or sex reversal were analyzed for aberrant chromatin interactions and changes in gene expression. The mouse models were further evaluated for morphological changes during development at E17.5 and P0.

3 Material

3.1 Chemicals

Unless stated otherwise, chemicals were obtained from Merck (Darmstadt), Roth (Karlsruhe) or Sigma-Aldrich (Hamburg, Seelze, Schnelldorf and Steinheim) in analytical grade quality.

3.2 Buffers

Common buffers and solutions were prepared according to Sambrook et al. 2001.

Table 1: Buffers for chromosome conformation capture.

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

Table 2: Buffers and reagents for southern blotting.

Buffer	Composition
Denaturation solution	0.5 M NaOH; 1.5 M NaCl
Neutralization solution	0.5 M Tris-HCl pH 7.5; 1.5 M NaCl
20x SSC	3 M NaCl; 300 mM sodium citrate; pH adjusted to 7.0 with HCl
Church buffer (NaPi)	1M Na2HPO4; pH adjusted to 7.2 with H3PO4
10x DIG1	1M Maleic Acid; 1.5 M NaCl; pH adjusted to 7.5 with NaOH pellets
Blocking Reagent	10 % w/v of Roche blocking reagent powder (Luminescent Detection Kit, Roche, #11363514910) in 1x DIG1
Church washing buffer	40 mM Church buffer; 1 % SDS
Church hybridization buffer	500 mM Church buffer; 7 % SDS; 1 mM EDTA (pH 8.0)
DIG 3	0.1 M Tris pH9.5; 0.1 M NaCl

Table 3: Buffer for whole mount *in situ* hybridization.

Buffer	Composition
Alkaline phosphatase buffer	0.1 M NaCl, 50 mM MgCl2, 0.1 % TWEEN-20, 0.1 M Tris pH 9.5, 0.5 mg/ ml tetramisole hydrochloride, in H_2O
10x PBS (DEPC)	1.37 M NaCl, 27 mM KCl, 100 mM Na2HPO4, 20mM KH2PO4, adjust pH to 7.4 with HCl, in DEPC- H_2O, 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
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- H_2O
Hybe buffer	2.5 $x10^{-5}$ % v/v formamide, 0.0025 mg/ ml heparin, 0.005% TWEEN-20, in 0.25x SSC (DEPC)
SSC/FA/T	50 % v/v formamide, 0.1 % TWEEN-20, in 1x SSC (DEPC)
RNase solution	0.5 M NaCl, 0.1 % TWEEN-20, in H_2O
5x MABT	0.5 M maleic acid pH 7.5, 0.75 M NaCl, 0.5 % TWEEN-20, in H_2O
PBST/ levamisole	0.5 mg/ ml tetramisole hydrochloride in PBST
Proteinase K buffer	20 mM Tris pH 7.0, 1 mM EDTA, in DEPC-H_2O
20x SSC (DEPC)	3 M NaCl, 300 mM Na3Citrate x $2H_2O$, adjust pH to 4.5 with 1M citric acid, 0.1 % DEPC, incubate o.n. at 37°C, autoclave
DEPC-H20	0.1 % DEPC, incubate o.n. at 37°C, autoclave

3.3 Antibodies

The following antibodies were used:

Anti-Digoxigenin conjugated to alkaline phosphatase (Roche, #11093274910)

3.4 Kits

Standard procedures, such as DNA or RNA purification, cDNA-synthesis, DNA-sequencing, protein concentration measurement were conducted following manufacturer's instructions using listed kits.

Table 4: Molecular biology kits used in this study.

Task	Kit	Supplier
Plasmid DNA purification	NucleoSpin Plasmid	Macherey-Nagel, Düren
Plasmid DNA purification	Nucleobond PC100	Macherey-Nagel, Düren
Plasmid DNA purification	Nucleobond PC100 EF	Macherey-Nagel, Düren
Gel and PCR clean-up	NucleoSpin Gel and PCR Clean-up	Macherey-Nagel, Düren
RNA-purification	RNeasy Mini Kit	Qiagen, Hilden
cDNA-synthesis	TaqMan Reverse Transcription Reagents	Applied Biosystems, Foster City, USA
Sanger-Sequencing	BigDye Terminator v3.1 Sequencing Kit	Applied Biosystems, Foster City, USA
PCR clean up	QIAquick PCR Purification Kit	Qiagen, Hilden

3.5 Enzymes

If not stated otherwise, restriction enzymes, T4 DNA ligase, Phusion DNA-polymerase and other DNA-modifying enzymes were obtained from Thermo Fisher Scientific (St. Leon-Roth), New England Biolabs (Frankfurt) and Promega (Mannheim). Taq and Pfu DNA polymerases for standard genotyping PCR were produced in house (A.C. Stiege). RNase A (Cat.-No. R4875) and Proteinase K (Cat.-No. P2308) were purchased from Sigma-Aldrich.

3.6 Bacterial strains

General cloning steps were performed in the *E. coli* Top10 (Invitrogen) strain.

3.7 Vectors

The vector pTA-GFP (supplied by Dr. J. Hecht, MPI for molecular genetics) was used for sub-cloning PCR products and *in situ* probes. pBluescript II SK (-) vector was modified for cloning targeting constructs of SB transgenes.

3.8 Primers

All primers were synthesized by MWG Biotech AG (Ebersberg) and HPSF purified. All primer sequences are shown in 5' to 3' orientation.

Table 5: Oligonucleotides for cloning the SB-1 and SB-2 targeting constructs.

Name	Sequence	Task
LinkpBSKS1	CTAGGCGCGCCATCGTCGACCTGATCGATCTACCGCGGTAGCCCGGGAGACTCGAGACTGCTAGCTCAGCGGCCGCACTGGCGCGCCTGAGCT	Custom multiple cloning site
LinkpBSKS2	CAGGCGCGCCAGTGCGGCCGCTGAGCTAGCAGTCTCGAGTCTCCCGGGCTACCGCGGTAGATCGATCAGGTCGACGATGGCGCGCCTAGGTAC	Custom multiple cloning site
SalI_SB-1_Long_F	GTGAGTCGACGGAATATAAAGGAAAGCTGGATGG	Amplification of SB-1 homology region
SacII_SB-1_Long_R	ACTACCGCGGCGGGCTATTGCTATAAGACTTGTG	Amplification of SB-1 homology region
XhoI_SB-1_ShortF	TGATCTCGAGCACTTCTCTTTCACACTTCTTTGAGC	Amplification of SB-1 homology region
NotI_SB-1_ShortR	ATACGCGGCCGCCTTGGCTGTAAAGAATCATGTCTAGC	Amplification of SB-1 homology region
SalI_SB-2_ShortF	CTATGTCGACGCGTATTGATGCTTTGAGAATGTG	Amplification of SB-2 homology region
SacII_SB-2_ShortR	ACGTCCGCGGTGGGAGACTTGCCTTAGAACAAC	Amplification of SB-2 homology region
XhoI_SB-2_LongF	TAGTCTCGAGGAAATATCAGTGTCCCTCCCTGAG	Amplification of SB-2 homology region
NotI_SB-2_LongR	CTGAGCGGCCGCAGGTTCTATTGCATTACTTGGTGAC	Amplification of SB-2 homology region

Table 6: Oligonucleotides for single *loxP* targeting and primer for single guide RNA generation.

Name	Sequence	Task
cenKcnj_*loxP*	TGATTTTGCGCTGTGATTTTGAACGTAATGA CCGTTGATATCATAACTTCGTATAGCATACAT TATACGAAGTTATTGTGAGGAAAGCGGGTT ATCTGTGAAGCCTCAGT	Single-stranded donor Oligonucleotide (ssODN) for targeting *loxP* site
cenKcnj_sg1_F	caccGAACGTAATGACCGTTTGTG	Guide RNA for *loxP* targeting
cenKcnj_sg1_R	aaacCACAAACGGTCATTACGTTC	Guide RNA for *loxP* targeting
cenΔBor_F	caccGATCATTTTAGGTAACGACCC	Guide RNA for cenΔBor
cenΔBor_R	aaacGGGTCGTTACCTAAAATGATC	Guide RNA for cenΔBor
telΔBor_F	caccGATTTAGCGTCCCCTAGCATA	Guide RNA for telΔBor
telΔBor_R	aaacTATGCTAGGGGACGCTAAATC	Guide RNA for telΔBor
cen_Dup-S_F	caccGTGCTGAAGTTGAACGATGCG	Guide RNA for cen_Dup-S
cen_Dup-S_R	aaacCGCATCGTTCAACTTCAGCAC	Guide RNA for cen_Dup-S
tel_Dup-S_F	caccGTTAGAAATCCTTGTCCCAAC	Guide RNA for tel_Dup-S
tel_Dup-S_R	aaacGTTGGGACAAGGATTTCTAAC	Guide RNA for tel_Dup-S

Table 7: Primers for long-range PCR and southern blot probes for the SB-1 and SB-2 alleles.

Name	Sequence	Task
103_LacZ-5DS-R	ttgaggggacgacgacagtatc	Long-range PCR 5' of SB-1 (5.95 kb)
121_cenIgArm_F	ggacacttctatccctcgctgac	Long-range PCR 5' of SB-1 (5.95 kb)
95_PGK451-F	tccagactgccttgggaaaagc	Long-range PCR 3' of SB-1 (3.1 kb)
97_cenShArm-Rb	gcagtggggattgacccttttattctc	Long-range PCR 3' of SB-1 (3.1 kb)
122_tellgArm_R	acagcatcaggacaacccacaac	Long-range PCR 3' of SB-2 (5.7 kb)
95_PGK451-F	tccagactgccttgggaaaagc	Long-range PCR 3' of SB-2 (5.7 kb)
101_telshArm-Fb	ctgggtttatatgctgctggagag	Long-range PCR 5' of SB-2 (3.9 kb)
103_LacZ-5DS-R	ttgaggggacgacgacagtatc	Long-range PCR 5' of SB-2 (3.9 kb)
132_CenSon5_3F	tccagtctatggtggggttcc	3' probe for SB-1, EcoRV
133_CenSon5_3R	gtatgcaacagaaccaggaagtgtc	3' probe for SB-1, EcoRV
126_CenSon2_5F	gtctctgctccattttgtccctac	5' probe for SB-1, BamHI
165_CenSon2_bR	ccattcccaaaacatctgctg	5' probe for SB-1, BamHI
NeoF	ttttgtcaagaccgacctgtcc	Internal probe for SB-1, EcoRV
NeoR	ccacagtcgatgaatccagaaaag	Internal probe for SB-1, EcoRV
134_TelSon1_5F	aggaaggccaagttaggtgctatc	5' probe for SB-2, HindIII
135_TelSon1_5R	aggggataagaaagggtctcagg	5' probe for SB-2, HindIII
172_TelSon4_3bF	catttctggtttatgtaaatcccttc	3' probe for SB-2, BglII
173_TelSon4_3bR	taagtgagtaatccaggctactttcc	3' probe for SB-2, BglII
199_Son13F	aggacagtcgtttgccgtct	Internal probe for SB-2, SacI
200_Son13R	tcggtcagacgattcattgg	Internal probe for SB-2, SacI

Table 8: Primer sequences for genotyping mouse lines.

Name	Sequence	Task
p317-L1	CTGGAATTGTGATACAGTGAATTATAAGTG	Mapping of SB insertion site
p318-L2	CTTGTGTCATGCACAAAGTAGATGTCC	Mapping of SB insertion site
p319-L3	AAGTAGATGTCCTAACTGACTTGC	Mapping of SB insertion site
p320-R1	CTTCTGACCCACTGGGAATGTGATG	Mapping of SB insertion site
p321-R2	GTGGTGATCCTAACTGACCTAAGAC	Mapping of SB insertion site
p322-R3	TCCTAACTGACCTAAGACAGG	Mapping of SB insertion site
p323-KmonP	GTACGAGAATCGCTGTCCT(7xN)CTCAG	Mapping of SB insertion site
p324-KmonP	GTACGAGAATCGCTGTCCT(7xN)TCCTG	Mapping of SB insertion site
p325-KmonP	GTACGAGAATCGCTGTCCT	Mapping of SB insertion site
p214	GCCATGTGATATTCTGAATCTGTCTC	Genotyping SB-1 (5'*lacZ* + p318)
P215	AGAGCTCTACAGTGCATCAGAAATTG	Genotyping SB-1 (3'*lacZ* + p321)
p472	CCCATGTTGGGATAGGCTTT	Genotyping SB-2 (5'*lacZ* + p318)
p473	TTGTGGCAGGGAGAAGACAT	Genotyping SB-2 (3'*lacZ* + p321)
p636	TTTTCAAGAACTGGGTATGACCTG	Genotyping SB-13 (5'*lacZ* + p318)
p635	ATGCCATGAAGATGCGTGAG	Genotyping SB-13 (3'*lacZ* + p321)
p668	AAGGACCAAGCAGAGGGAAA	Genotyping SB-16 (5'*lacZ* + p318)
p667	CCCAAAGTCTTTCATTTTGTGAAG	Genotyping SB-16 (3'*lacZ* + p321)
p663	ATTGCGTGACACAGATGAACC	Genotyping SB-17 (5'*lacZ* + p318)
p664	AGGGGGACTTAAGGGAAAAGAA	Genotyping SB-17 (3'*lacZ* + p321)
p689	CAGATTTCATGGGCTCAAGG	Genotyping SB-18 (5'*lacZ* + p318)
p690	TCAAAGGAGGTGGATGACTCC	Genotyping SB-18 (3'*lacZ* + p321)
p696	CTGACTGTTCCTGTCTTGTTTCC	Genotyping SB-20 (5'*lacZ* + p318)
p695	TGGGTTTGGTATCTATGTGCAA	Genotyping SB-20 (3'*lacZ* + p321)
p692	TGGGGCAGAGTAAGGAATGA	Genotyping SB-22 (5'*lacZ* + p318)
p691	AACCACTGAGTCGTCGCTACA	Genotyping SB-22 (3'*lacZ* + p321)
p700	CCTGGGTTCAAGGATCATCA	Genotyping SB-23 (5'*lacZ* + p318)
p699	CACTCCCTTTCCTCGCTTCT	Genotyping SB-23 (3'*lacZ* + p321)
p720	AGTTCAATTGCAGGCCGATT	Genotyping SB-24 (5'*lacZ* + p318)
p721	TTGGCTTCCTAACGGAGGAG	Genotyping SB-24 (3'*lacZ* + p321)
p744	CCACTGGAACGCTGATGTTT	Genotyping SB-25 (5'*lacZ* + p318)
p745	GGGGCACTTTGACAAGTGAG	Genotyping SB-25 (3'*lacZ* + p321)
p577	AGAAAACCAACGTGAAGCCC	Genotyping *loxP* (Kcnj2/16)
p579	TGTCCCTGGGAAGTCCATTC	Genotyping *loxP* (Kcnj2/16
p472	CCCATGTTGGGATAGGCTTT	Genotyping Dup-L (5'*lacZ* + p318)
p215	AGAGCTCTACAGTGCATCAGAAATTG	Genotyping Dup-L (3'*lacZ* + p321)
p215	AGAGCTCTACAGTGCATCAGAAATTG	Genotyping Dup-LΔLacZ
p217	ACAATGGATGGTGCAGAGTTAGG	Genotyping Dup-LΔLacZ
p214	GCCATGTGATATTCTGAATCTGTCTC	Genotyping Del-L (5'*lacZ* + p318)
p473	TTGTGGCAGGGAGAAGACAT	Genotyping Del-L (3'*lacZ* + p321)
p472	CCCATGTTGGGATAGGCTTT	Genotyping Dup-C
p579	TGTCCCTGGGAAGTCCATTC	Genotyping Dup-C
p472	CCCATGTTGGGATAGGCTTT	Genotyping Dup-C2 (5'*lacZ* + p318)
p664	AGGGGGACTTAAGGGAAAAGAA	Genotyping Dup-C2 (3'*lacZ* + p321)
p720	AGTTCAATTGCAGGCCGATT	Genotyping Dup-K1
p214	GCCATGTGATATTCTGAATCTGTCTC	Genotyping Dup-K2
p720	AGTTCAATTGCAGGCCGATT	Genotyping Dup-K3 (5'*lacZ* + p318)
p835	CAAAATTAACCGTGTCCTCAAA	Genotyping Dup-S
p541	CTCCAGCGAAATGGAATGAG	Genotyping ΔBor
p548	CCTTCCCTCCTGAATTCTCTG	Genotyping ΔBor

Table 9: 4C-seq primers for inverse PCR and digestion strategy used for 4C library preparation.

Viewpoint	Read primer 1	Primer 2	Ref. genome	RE (1^{st}/2^{nd})	Fragment
Sox9 TSS	CCACGAAAGTGTGTTTATATTCT	TGGCTATTGTTTGTAAATCTCTT	mm9	DpnII/ BfaI	chr11:112,641,406 -112,644,076
Sox9 TSS	GAGAACACCTTCCCCAAG	CTAGGACCCTCCACAACC	mm9	DpnII/ Csp6I	chr11:112,643,068 -112,644,076
Kcnj2 TSS	CGGTGGTGGAGAGAAAGA	TCACCAAACAACCTCTACAAA	mm9	DpnII/ BfaI	chr11:110,926,035 -110,926,698
LacZ	TCCGACTTCAACTGTAGGG	GCAGAGCCATCTATTGCTTA	NA	DpnII/ BfaI	unique sequence, *β-globin promoter:LacZ*
Sox9 tel	TTTGACTCCAAGGACCAGA	TGTGTAAAGGCGGACAGA	mm9	DpnII/ BfaI	chr11:113,027,139 -113,027,634
Sox9 cen	AACTGGGTATAGTGTGGAGAGA	GGACATCAATGGAACATAGC	mm9	DpnII/ BfaI	chr11:111,445,674 -111,446,487
BOR (Kcnj/Sox9)	AATGCTGTGGACATCCTG	TCCGTCAAGTTCCATGTG	mm9	DpnII/ BfaI	chr11:111,393,725 -111,394,348
Kcnj tel	TCTGTGCCAAATGAACTATTCT	AGCACCATCTGTCCTCCTAT	mm9	DpnII/ BfaI	chr11:111,312,122 -111,313,114
Kcnj cen	AAGAAGAGGGGCTCAAAAT	CTGATTGGCATGGGATAGT	mm9	DpnII/ BfaI	chr11:110,399,546 -110,400,841
SOX9 TSS	TCCAAGTGTGTAAGTTTGTCG	AATCTCCTGGACCCCTTC	hg19	Csp6I/ NlaIII	chr17:70,116,889 -70,117,913
KCNJ2 TSS	TTCAGTGAACGTATTTGTTGAA	GAGCGTTTGATGTTTTACCA	hg19	Csp6I/ NlaIII	chr17:68,162,418 -68,164,266
KCNJ16	TGATTACGTCACCACCTCTG	AGCCCTTGCCTGCTATTAT	hg19	Csp6I/ NlaIII	chr17:68,071,287-68,072,497
Breakpoint (150 kb intra-TAD dup)	ATGACCTTGGCAACAATTTA	AGGCTGAGTCAGGAGAATTT	hg19	Csp6I/ DpnII	unique fragment between chr17:69,669,737 and chr17:69,521,930
Breakpoint (470 kb intra-TAD dup)	TCATCCTCCTTAGAGACTGAGA	GACTAGGGCAAGGCTTACAT	hg19	DpnII/ Csp6I	unique fragment between chr17:69404464 and chr17: 69872848
Breakpoint (1.8 Mb inter-TAD dup)	TTGGTTGTTTTGATGAGGAT	GGGCATGATTTCATTCTTT	hg19	DpnII/ Csp6I	unique fragment between chr17:69981277 and chr17:68195564

Table 10: Primer sequences for amplification of *In situ* probes.

Name	Sequence	Task
490Sox9-ISH_F	CGGGCGAGCACTCTGG	*In situ* probe for *Sox9*
491Sox9-ISH_R	TGGGGCTCAGCTGCTCC	*In situ* probe for *Sox9*
492Kcnj2-ISH_F	ATATGACTGGCTGATTCCGTCT	*In situ* probe for *Kcnj2*
493Kcnj2-ISH_R	CTCAACACTGACGTCTTAACGTT	*In situ* probe for *Kcnj2*
494Kcnj16-ISH_F	TTATTGCATATACCTGTGTCTCCT	*In situ* probe for *Kcnj16*
495Kcnj16-ISH_R	CACAGTGGGATTATGAGCAC	*In situ* probe for *Kcnj16*

3.9 Mouse lines

Table 11: Mouse lines used or generated in this study.

Line name	Type of allele	Source/ origin	Location (mm9)/ *loxP* orientation
C57BL/6	Wt inbred strain	Mouse facility, MPI Berlin	-
CD1 IGS	Wt outbred strain	Mouse facility, MPI Berlin	-
Hprt:Cre	Cre recombinase	Spitz lab (Tang et al., 2002)	-
Prm1:HSB16-A	SB transposase	Spitz lab (Ruf et al., 2011)	-
hACTB:FLPe	Flp recombinase	F. Spitz (Rodríguez et al., 2000)	-
SB-1 (SB-Kcnj)	SB-transgene	This study, targeted	chr11:110,959,589/ +
SB-2 (SB-Sox9)	SB-transgene	This study, targeted	chr11:112,514,692/ +
SB-13	SB-transgene	This study, transposition	chr11:113,156,746/ -
SB-16	SB-transgene	This study, transposition	chr11:111,061,524/ -
SB-17	SB-transgene	This study, transposition	chr11:110,838,025/ +
SB-18	SB-transgene	This study, transposition	chr11:112,738,699/ +
SB-20	SB-transgene	This study, transposition	chr11:110,741,676/ -
SB-22	SB-transgene	This study, transposition	chr11:111,150,850/ -
SB-23	SB-transgene	This study, transposition	chr11:112,512,145/ -
SB-24	SB-transgene	This study, transposition	chr11:111,373,724/ +
SB-25	SB-transgene	This study, transposition	chr11:112,480,568/ +
loxP (Kcnj2/16)	Single *loxP* site	This study, targeted	chr11:110,772,110/ +

Table 12: Generated mouse alleles carrying duplications and deletions.

Line name	Donor alleles	*LacZ* reporter	Rearrangement/ genomic location (mm9)
Dup-L (Sox9Dup)	SB-1 and SB-2	yes	Duplication/ chr11:110,959,589-112,514,692 (1,56 Mb)
Dup-LΔLacZ (Sox9Dup-LacZ)	Dup-L, *LacZ* removed by transposition	no	Duplication/ chr11:110,959,589-112,514,692 (1,56 Mb)
Dup-LΔBor	Dup-L; CRISPR/Cas9-system	yes	Duplication/ chr11:110,959,589-112,514,692 (1,56 Mb); Deletion/ chr11:111,383,859-111,402,200
Del-L (Sox9Del)	SB-1 and SB-2	yes	Deletion/ chr11:110,955,589-112,514,692 (1,56 Mb)
Dup-C (Sox9Dup/Kcnj2/16)	*loxP* and SB-2	no	Duplication/ chr11:110,772,110-112,514,692 (1,74 Mb)
Dup-C2 (Sox9dup/Kcnj2)	SB-17 and SB-2	yes	Duplication/ chr11:110,838,025-112,514,692 (1,68 Mb)
Dup-K1 (Dup(*loxP*/SB-24))	*loxP* and SB-24	no	Duplication/ chr11:110,772,110-111,373,724 (601 kb)
Dup-K2 (Dup(*loxP*/SB-1)	*loxP* and SB-1	no	Duplication/ chr11:110,772,110-110,959,589 (187 kb)
Dup-K3 (Dup(SB-1/SB-24)	SB-1 and SB-24	yes	Duplication/ chr11:110,959,589-111,373,724 (414 kb)
Dup-S	CRISPR/Cas9-system	no	Duplication/ chr11:111,752,617-112,172,936 (420 kb)
ΔBor	CRISPR/Cas9-system	no	Deletion/ chr11:111,383,859-111,402,200 (18,3 kb)

3.10 Human material

Table 13: Human fibroblasts, lymphoblastoid cells (Lcl) and white blood cells used for 4C-seq.

Sample	Mutation/ Syndrome	Cell type	Genomic coordinates (hg19)	Provided by
PL/ normal Phenotype	46,XX, inter-TAD dup, 1.78 Mb	Skin fibroblasts	chr17:68195430-69981335	Francesco Brancati (IRCCS, Rome, Italy)
RN/ Cooks syndrome	46, XY, inter-TAD dup, 1.77 Mb	Skin fibroblasts	chr17: 67907000-69682000	Lindsay Lambie (Uni. Witwatersrand, South Africa)
1039-12/ DSD	46, XX, intra-TAD dup, 468 kb	Gonadal fibroblasts	chr17:69404081-69872909	Orsetta Zuffardi (Uni. of Pavia, Italy)
VP09-0039/ DSD-2	46,XX, intra-TAD dup, 148 kb	Skin fibroblasts	chr17:69521863-69670036	Gunnar Houge (S. Benko et al., 2011)
VP09-0036	46, XY, intra-TAD dup, 148 kb	Skin fibroblasts	chr17:69521863-69670036	Gunnar Houge (S. Benko et al., 2011)
Control	46, XY	Skin fibroblast	none	Lab member
III.1. Control	46, XX	Lcl and white blood cells	none	Ingo Kurth, (Kurth et al., 2009),family 1
III.2. Cooks syndrome	46, XY, inter-TAD dup, 1.96 Mb	Lcl and white blood cells	chr17:68053486-70019664	Ingo Kurth, (Kurth et al., 2009),family 1
III.3. Control	46, XY	Lcl and white blood cells	none	Ingo Kurth, (Kurth et al., 2009),family 1
III.4.Cooks syndrome	46, XY, inter-TAD dup, 1.96 Mb	Lcl and white blood cells	chr17:68053486-70019664	Ingo Kurth, (Kurth et al., 2009),family 1

3.11 Instruments

Table 14: Instruments.

Instrument	Type/ Supplier
Table Top centrifuge	5414D/ Eppendorf, Hamburg
Cooling centrifuge	5417R/ Eppendorf, Hamburg
Cooling centrifuge	Avanti J-E/ Beckman-Coulter, Palo Alto, USA
Rotor	JLA16.250/ Beckman-Coulter, Palo Alto, USA
Thermocycler	GeneAmp PCR System 2700, 2720 and 9700/ Applied Biosystems, Foster City, USA
Real-time Thermocycler	ABIPrism 7900 HT/ Applied Biosystems, Foster City, USA
Stereomicroscope	MZ12/ Leica, Bensheim
Camera	Axiocam MRc5/ Zeiss, Göttingen
Light source	KL1500 LCD/ Leica, Bensheim
Photometer	NanoDrop 2000/ Thermo Scientific, Wilmington, USA
Biochemical Analysis	Agilent Technologies, Böblingen

3.12 Software and internet resources

Table 15: Software and internet resources used in this study.

Name	Supplier/ web address	Application
ImageJ	National Institute of Health (NIH)	Image processing
AxioVision Rel.4.8	Zeiss	Microscopy, digital photography
Adobe Photoshop CS6	Adobe	Composition of figures
Vector NTI	Invitrogen	Depiction and modification of DNA sequences
SDS 2.2.1	Applied Biosystems	Analysis of qPCR data
UCSC browser	http://genome.ucsc.edu/	Data visualization
Primer3	http://biotools.umassmed.edu/bioapps/primer3_www.cgi	Primer design
NetPrimer	http://www.premierbiosoft.com/netprimer	Primer design
Bin Ren Hi-C data	http://yuelab.org/hi-c/database.php	Resource for mouse and human Hi-C data
IGV browser	https://www.broadinstitute.org/igv/	Visualization of 4C-seq, ChIP-seq data
WashU browser	http://epigenomegateway.wustl.edu/browser/	Visualization of Hi-C data
Juicebox	http://www.aidenlab.org/juicebox/	Visualization of Hi-C data; resource for human Hi-C data
Mendeley	https://www.mendeley.com/	Bibliography management

4 Methods

4.1 Molecular biological methods

Standard molecular biological procedures, such as cloning of DNA fragments, transformation of chemically competent E. coli, gel electrophoresis were conducted according to Sambrook and Russels (Sambrook & Russel, 2001).

4.2 DNA isolation

4.2.1 Isolation of plasmid DNA

Isolation of plasmid DNA was performed with Nucleospin or Nucleobond PC100 kits (Macherey-Nagel) according to the instructions of the manufacturer. Plasmid DNA for transfection of eukaryotic cells was isolated with the Nucleobond PC100 EF kit.

4.2.2 Isolation of genomic DNA

DNA of mouse biopsies from postnatal stages was isolated with QuickExtract™ DNA extraction solution (Biozym). 50 µl of the QuickExtract was added to the sample, incubated for 20 minutes at 65°C and subsequently, the solution was inactivated for 2 min at 98°C. Extracted DNA solution was used for standard genotyping PCR.

DNA from embryonic material e.g. amnion or DNA for copy number analysis via qRT-PCR was isolated by conventional DNA precipitation procedures. Briefly, tissue was lysed in 500 µl lysis buffer (17 mM Tris, pH 7.5; 17 mM EDTA; 170 mM NaCl, 0.85 % SDS) with freshly added proteinase K (0.08 µg/ µl final) at 55°C for 3 to 12 hours. After adding 250 µl 5M NaCl, samples were incubated for 10 minutes at room temperature while shaking, then for 10 minutes on ice and centrifuged for 20 minutes at 4°C at 9000 rpm. 500 µl of supernatant was mixed with 2x volume of 100 % ethanol and centrifuged for 25 minutes at 13000 rpm at 4°C. DNA pellet was washed once with 70 % ethanol, dried and dissolved in water.

4.3 Cloning of targeting constructs for SB-1 and SB-2 insertions

The SB-1 and SB-2 alleles contained an insertion of the Sleeping Beauty (SB) transgene. This transgene harbours a single *loxP* site and a *LacZ* reporter gene, flanked by SB-transposons (Ruf et al., 2011). Detailed description of the transgene is available at the TRACER database from the Spitz lab (http://www.ebi.ac.uk/panda-srv/tracer/). Both alleles were targeted using standard protocols for homologous recombination in E14-ES cells (Hooper et al., 1987).

The targeting constructs were cloned into the pBluescript II SK (-) vector. For this purpose, the original multiple cloning site (MCS) between the restriction sites *KpnI* and *SacI* was replaced by a custom MCS which was generated by two annealed oligos (Table 5). Subsequently, the SB-transgene (Ruf et al., 2011), provided by Francois Spitz (EMBL, Heidelberg) was cloned into the modified pBluescript II SK (-) using *SacII* and *XhoI* (Figure 8). Next, homology regions of 5 kb and 3 kb flanking the genomic integration site (SB-1, chr11:110,959,589; SB-2, chr11:112,514,692; mm9) were amplified by PCR from gDNA of E14 ES cells. Primers were designed with overhangs of the required restriction site (Table 5). The generated PCR products with A-overhangs were subcloned into the pTA-GFP vector and sequenced (Sanger) in order to confirm the sequence. Homology regions were subsequently cloned 5' and 3' of the SB-transgene within the modified pBluescript II SK (-) vector (Figure 8). Midi-preparation of bacteria culture was performed with endotoxin-free Nucleobond PC100 EF (Macherey-Nagel). The targeting construct was then linearized using the *AscI* restriction enzyme, separated on an agarose gel and purified. The purified product was then used for transfection in E14 mouse embryonic stem cells (see section 4.5.1.4).

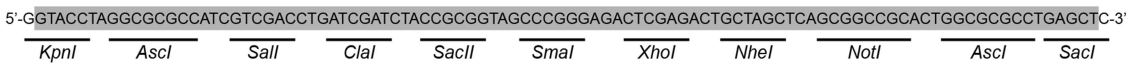


Figure 8: Cloning strategy for SB-1 and SB-2 targeting constructs. Custom multiple cloning site of the modified pBluescript II SK (-) vector is shown in grey. Black lines depict restriction sites used for cloning the SB-transgene (*SacII, XhoI*), and homology regions 5' (*SalI, SacII*) and 3' (*XhoI, NotI*) of the SB-transgene. *AscI* sites were used to linearize the targeting construct for ES cells transfection.

4.4 Cloning of single guide RNAs for CRISPR/Cas9

Single guide RNAs were designed using the CRISPR Design Tool by Feng Zhang lab (http://www.genome-engineering.org/crispr/), which implements *In silico* quality tests and off-target predictions (Hsu et al., 2013). Only guide RNAs with a quality score above 70 % and an off-target potential in exonic regions below 1 were chosen. Guide sequences used in this study are given in Table 6. Two oligonucleotides containing the guide specific sequence and BbsI recognition site overhangs (Oligo 1: 5'-caccgNNNNNNNNNN-3'; Oligo 2, reverse complement: 5'-aaacNNNNNNNNNNc-3') were annealed, phosphorylated (T4 Polynucleotide Kinase, Thermo Fisher Scientific, #EK0032) and cloned into the pX459 vector (pSpCas9(BB)-2A-Puro, Addgene). The vector contains the chimeric tracrRNA, the Cas9 and puromycin resistance gene. The vector was digested with the BbsI restriction enzyme, dephoshorylated (FastAP Thermosensitive Alkaline Phosphatase, ThermoFisher Scientific, #EF0654) and purified. The chosen guide RNAs were cloned into the linearized pX459 vector and transformed into chemical competent Top10 bacteria.

Finally, the plasmid DNA was purified from 5 ml cultures and successful cloning was validated by Sanger sequencing using a vector specific primer (ColR: CACGCGCTAAAAACGGACTA).

4.5 Cell culture

4.5.1 Culturing and manipulation of mouse embryonic stem cells (mESCs)

The ES cell culture protocol in the laboratory was established by Katerina Kraft and 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, Anderson, Nagy, & Nagy, 1994), (Wassarman & Soriano, 2010), (Robertson, 1987) and (Kraft et al., 2015).

4.5.1.1 Culturing feeder cells/ primary embryonic fibroblasts

Feeder cells were cultured in regular Dulbecco's Modified Eagle's Medium (DMEM) 4,500 mg/ ml glucose, without sodium pyruvate (Lonza, #BE12-733F), containing 10 % regular fetal calf serum (FCS Superior, Biochrom, #S0615), 1x glutamine (100x, Lonza, #BE17-605E) and 1x penicillin/ streptomycin (100x, Lonza, #DE17-603). The preparation of feeder cells was conducted in our laboratory from E13.5 to E14.5 CD1 (M. C. Rice & O'Brien, 1980) and DR4 mice (Tucker, Wang, Dausman, & Jaenisch, 1997). Fibroblasts were expanded until passage 5 and tested for Mycoplasma contamination using the Mycoalert detection kit (Lonza, #LT07-118) and the Mycoalert assay control set (Lonza, #LT07-518). The feeder cells were treated with mitomycin C (Sigma, #M-4287 or M-0503) and frozen in cryovials at a density of 3 x10^6 cells/ vial. The freezing medium consisted of regular feeder medium containing 20 % FCS and 20 % DMSO (Sigma, #D-2650).

4.5.1.2 Culturing mouse embryonic stem cells (ES cells)

ES cells, i.e. G4 cells (129/Sv x C57BL/6 F1 hybrid, (George et al., 2007)) were cultured on plates or wells gelatinized and covered with feeder cells. In detail, the culturing dishes or wells were coated with 0.1 % gelatin (Sigma, #G-1393). After 30 min incubation at 37°C the gelatin was aspirated and feeder cells were plated at a density of 3-4 x10^4 cells/ cm^2. After at least 6 hours, ES cells were seeded on top of the feeder layer and grown in Knockout Dulbecco's Modified Eagle's Medium (DMEM) 4,500 mg/ml glucose, with sodium pyruvate (Gibco, #10829-018) containing 15 % FCS (PANSera ES, #P30-2600, Lot 130407ES), 1x glutamine (100x, Lonza, #BE17-605E), 1x penicillin/ streptomycin (100x, Lonza, #DE17-603), 1x non-essential amino acids (100x, Gibco, #11140-35), 1x nucleosides (100x, Chemicon, #ES-008D), 0.1 mM β-mercaptoethanol (Gibco, #3150-010) and 1000 U/ ml LIF (Murine Leukemia Inhibitory Factor ESGRO™ (10^7 U/ ml, Chemicon, #ESG1107).

The ES cell medium was changed every 24 hours and cells were frozen at a density of 1 x10^6 cells/ vial in freezing medium, consisting of regular ES cell medium supplemented with 20 % FCS and 20 % DMSO (Sigma, #D-2650).

4.5.1.3 Transfection of G4 ES cells for CRIPR/Cas9 induced genome editing

CD1 feeder cells were seeded out on 6-well plates. The next day, 0.35 x10^6 G4 ES cells/ well were plated onto thefeeder cells. The following day, 8 µg of each pX459-Vector containing a guide RNA were transfectedusing FUGENE HD reagent (Promega, #E2311) according to manufacturer´s instructions. For targeting a single *loxP* site upstream of *Kcnj16* (chr11:110,772,110, mm9), G4ES cells were co-transfected with 8 µg pX459-Vector (Addgene) carrying the sgRNA (cenKcnj) and 500 pM of a ssODN (single-stranded donor oligonucleotide, containing 40 bp homologous sequence flanking a *loxP* site). For deleting the 18 kb TAD boundary fragment (ΔBor, chr11:111,383,849-111,402,212, mm9) in G4 ES cells and in ES cells from Dup-L homozygous mice, two pX459 sgRNAs (cenΔBor, telΔBor) were co-transfected. The Dup-S duplication was generated in G4 ES cells using also two sgRNAs (cenDupS, telDupS). Sequences of the sgRNAs are given in Table 6. After two days, ES cells were split and seeded on three 6 cm dishes with puromycin resistant DR4 feeder cells in a medium containing 2 µg/ ml Puromycin (Sigma-Aldrich, #P8833). After two days of antibiotic selection, the medium was replaced by regular ES cell medium. Cells were grown until single clones could be visible. These clones were picked in PBS (Lonza, #BE17-512F) with sterile pipette tips and transferred to U-bottom 96-well plates, containing 1x trypsin-EDTA (Gibco, #25300-054). After 12 min incubation at 37°C, clones were re-suspended and transferred to a new 96-well plate with regular ES cell medium and CD1 feeder cells. The cells were grown for two days, trypsinized and two thirds were frozen in 96-well plates containing ES cell freezing medium. The remaining one third was further grown for DNA isolation. Confluent cells were lysed in lysis buffer (10 mM Tris pH 7.5, 10 mM EDTA pH 8.0, 10 mM NaCl, 0.3 % Sacrosyl, 1 mg/ ml ProteinaseK) and processed for DNA precipitation. After genotyping by PCR, positive clones were expanded and used for generating mouse embryos and live animals from ES cells by di- or tetraploid complementation (Artus & Hadjantonakis, 2011).

4.5.1.4 Transfection and targeting of SB-1 and SB-2 in E14 ES cells

The SB-1 and SB-2 alleles were targeted in E14 ES cells (Hooper et al., 1987) using standard protocols for homologous recombination. The ES cell culture and generation of chimeric animals was carried out in the laboratory of Francois Spitz (EMBL, Heidelberg). For each targeting construct, 1 x10^7 ES cells were suspended in Embryo Max electroporation buffer (Millipore, #ES-003-D) and mixed with 15 µg (1 µg/ kb) of linearized vector in a total volume of 600 µl. After 5 min incubation at room temperature, the solution was transferred to a Gene Pulser cuvette (BioRad,

#165-2088). Cells were electroporated at 240 V, 500 µF using Gene Pulser device (Biorad) and seeded in different densities (35, 10 and 5 % of initial cell number) on 10 cm dishes with feeder cells. After two days ES cell were grown for additional 24 hours in medium containing 200 µg/ ml Geneticin (G418, Gibco, #11811023). The selection was continued for additional five days with medium supplemented with 250 µg/ ml Geneticin. Single clones were picked, expanded and frozen as described in section 4.5.1.3. DNA from single clones was genotyped by PCR and the correct homologous recombination of candidate clones was verified by southern blot analysis, which is described in section 4.6. Positive ES cell clones were injected into donor blastocysts to generate chimeras. Blastocyst injection was performed in the transgenic facility at the EMBL, Heidelberg and chimeric animals were transferred to the animal facility at the Max Planck Institute for Molecular Genetics (Berlin).

4.5.2 ES cell derivation from Dup-L mice

ES cells from Dup-L homozygous blastocysts were established in cooperation with Lars Wittler (Department Developmental Genetics, Max Planck Institute for Molecular Genetics, Berlin) using N2B27 Medium supplemented with FGF/ Erk, Gsk3 pathway inhibitors (2i) and LIF according to Nagy and Nichols (Nagy & Nichols, 2011). ES cell lines were confirmed by PCR genotyping. A male ES cell line was established and used for subsequent deletion experiments using CRISPR/Cas9.

4.5.3 Culturing human-derived cells

Human-derived fibroblasts from skin biopsies were obtained from collaboration with clinical partners. A written informed consent was obtained from all studied individuals to participate in this study. Table 13 lists all patient-derived cells used in this study. Fibroblasts and lymphoblastoid cells were cultured in DMEM (4,5 g/ L glucose, without pyruvate; Gibco, #11960-044) supplemented with 15 % FCS (Biochrome, #S0615), 1x Penicillin/ Streptomycin (Biochrome, #A2213) and 1x L-glutamin (Biochrome, #K0283). Sub-confluent cultures were split or frozen in culture medium containing 20 % FCS and 20 % DMSO.

Nucleated blood cells, such as leukocytes and lymphocytes from peripheral blood samples were isolated at the Institute for Medical and Human Genetics atthe Charité Universitätsmedizin in Berlin. Blood samples from patients and controls were diluted with 1x vol. of RPMI medium (Gibco, #11875093) and gently mixed by inverting the samples. In the next step, 5 ml Bicoll (Biochrome, #L6715) per 6 ml sample were carefully added on top of the sample. After centrifugation at 400g for 30 min (without centrifuge brake), the serum phase was removed and discarded. The remaining upper phase, the buffy coat containing the lymphocytes, was isolated, resuspended in 12 ml RPMI medium and centrifuged (with centrifuge break) for 12 min at 400g.

After a second washing step with RPMI medium and centrifugation, the cell pellet was subjected to crosslinking and nuclei extraction for circular chromosome conformation assay (section 4.10). In average, 8 x10^5 cells per 1 ml blood sample were isolated.

4.6 Screening and southern blot analysis of the targeted SB-1 and SB-2 alleles in ESCs

ES cell clones for the SB-1 and SB-2 allele (400 clones each) were pre-screened by long-range PCR using the Expand Long Template PCR System (Roche, #11759060001). To do so, two primer pairs covering the homologous region 5' and 3' of the SB-transgene were used. In each primer set, one primer is located within the SB-transgene and the second primer in the genomic region outside the region of homology (see Table 7 for primer sequences). Seven clones for SB-1 and eight clones for SB-2 with the correct PCR product size of the 5' and 3' homology region could be identified. The correct recombination of these clones was further confirmed by southern blot analysis. The results are shown in the Appendix Figure 35 and primer sequences for probe generation are listed in Table 7. An internal probe, hybridizing to the transgene was used to verify single recombination events. In addition, two DNA probes (5' and 3' external probes), hybridizing outside the region of homology were used to verify homologous recombination on both sides of the transgene. The clones 157 (for SB-1) and 181 (for SB-2) were used to generate the corresponding mouse lines.

4.6.1 DNA probe generation

Complementary DNA probes with a length of 300-400 bp and a GC content of 40 %, located in non-repetitive regions, were amplified from gDNA or vector DNA using the primers listed in Table 7. Labeling was done using PCR DIG Probe Labeling Kit from Roche (Cat. No.: 11636090910) following manufacturer's instructions. The labeled PCR products were separated on an agarose gel and the expected band was purified using NucleoSpin Gel and PCR clean-up kit. Isolated and labeled probes were eluted in 30µl water.

4.6.2 Southern blot

All buffers used for Southern blot analysis are listed in Table 2. 10 µg genomic DNA (gDNA) from pre-selected clones was digested using restriction enzymes to yield a fragment of known size. Restriction enzymes and fragment sizes are given in Figure 35 in the Appendix. Digested gDNA and a DIG-labeled molecular weight marker (Roche, # 11218590910) were separated on a 1 % agarose gel. The gDNA on the gel was denatured by incubating the gel in denaturation solution for 30 min at room temperature with gentle agitation. To neutralize the denaturing salts, the gel was incubated twice in neutralization solution for 20 min at room temperature. Subsequently,

separated gDNA was transferred to a nylon membrane by capillary blotting. An extension block was placed in a tray which was then filled with 10x SSC. Two sheets of Whatman paper, matching the exact size of the gel were placed on the extension block, followed by the upturned gel. The nylon membrane (GE healthcare, positively charged Nylon Transfer membrane, Cat No: RPN203B) matching the size of the gel was placed on top, followed by two additional layers of Whatman paper. The entire tray and blot were sealed with saran foil. A hole matching the exact size of the blot was cut out with a scalpel. A batch of paper towels and a weight were placed on top to provide capillary force from the gel to the membrane. The gDNA containing membrane was washed with 50 mM NaPi for 5 min under gentle agitation at room temperature. Subsequently, the membrane was baked at 80°C for 2 hours to fix the gDNA on the membrane, which was stored at 4°C until probe hybridization.

4.6.3 Probe hybridization and chemiluminescent assay

The nylon membrane was rinsed in bidistilled water and incubated in 25 ml Church hybridization solution in a glass bottle at 65°C for 30 min on a rolling device. 4µl of DIG-labeled probe was diluted in Church hybridization solution and placed in boiling water for 10 min to denature the probe. The hybridization solution in the bottle was replaced by the probe containing solution and probe hybridization was performed at 65°C overnight. Unbound probe was removed by washing with pre-heated Church washing buffer at 65°C, followed by an additional washing step for 10 min at room temperature with gentle agitation. To prepare the membrane for blocking, a 5-min washing step in 1x DIG1+0.3 % Tween was performed. The membrane was sealed in a plastic bag containing blocking reagent and incubated for 30 min at room temperature under vigorous agitation. After replacing the blocking solution by blocking reagent containing alkaline phosphatase-coupled anti-DIG antibody 1:20.000 (Anti-Digoxigenin-AP Fab fragments from sheep, Roche #11093274910), the membrane was incubated again for 30 min at room temperature under vigorous agitation. Next, the membrane was placed in a tray and washed with 1x DIG1+0.3 % Tween for 20 min with shaking to remove unbound antibodies. An additional washing step in DIG3 for 5 min at room temperature removed remaining detergents. Lastly, the membrane was placed in a plastic envelope, incubated with 1 ml of CDPstar (Roche, CDPstar ready-to-use #12041677001) and covered by the plastic envelope. Chemiluminescent signal was detected by exposing the membrane to a film and subsequent film processing.

4.7 Mouse crossings

All animal procedures carried out at the Max Planck Institute for Molecular Genetics were in accordance with institutional, state, and governmental regulations (Berlin: LAGeSo). All mouse lines used or generated in this study are listed in Table 11. Generated mouse lines were back-crossed to the *C57BL/6* background. Germline transmission was obtained from different chimeric animals for the SB-1, SB-2 and Kcnj_*loxP* allele. The FRT-flanked selection cassette from the SB-1 and SB-2 alleles was removed by breeding with hACTB-FLPe (Rodríguez et al., 2000) mice, leaving only the SB-transgene at the integration site.

4.7.1 Generation of new SB-insertion alleles by in vivo transposition

The SB-transgene was re-mobilized from the mouse lines SB-1 and SB-2 and new insertions were generated as described in (Ruf et al., 2011). Both, SB-1 and SB-2 animals were crossed with the *Prm1::HSB16* mouse line. Double transgenic males ("seed males") were then crossed to female *C57BL/6* mice. The transposase from the *Prm1::HSB16* allele initiated the remobilization of the SB-transgene in the sperm of double transgenic males (e.g. SB-1; Prm1::HSB16). New SB-insertions were generated by mating "seed males" to wild-type females. The offspring was subsequently screened for new insertions as described in section 4.8.

4.7.2 Generation of genomic rearrangements by in vivo recombineering

Cre/*loxP* recombination was used to generate genomic rearrangements, such as deletions and duplications at the *Sox9* locus. Two *loxP* sites with the same orientation but located on different alleles combined with a Cre recombinase induced trans-allelic recombination *in vivo,* as described in (Hérault et al., 1998). All donor mouse alleles with *loxP* sites and the resulting mouse mutants with deletions or duplications are listed in Table 11 and Table 12. As an example, to generate Dup-L mutants, mice carrying the SB-1 allele were crossed with mice of the Hprt:Cre line and the resulting double transgenic animals were crossed to the SB-2 mouse line. This led to triple transgenic animals that have the SB-1 and SB-2 allele in trans-configuration and the Hprt:Cre transgene. Hprt:Cre is a strong Cre driver which induces recombination of the two *loxP* site during meiosis (Tang et al., 2002). The triple transgenic animals generate gametes with a duplication or deletion of the interjacent region between the *loxP* sites. Triple transgenic animals were crossed to C57BL/6 mice and the offspring was screened for duplications or deletions, which occurred in 10 to 20 % per litter (Ruf et al., 2011). The screening was conducted using PCR primers that amplify unique sequences at the breakpoint (Table 8). The obtained founder animals were crossed with C57BL/6 animals to establish a new mouse line.

4.8 Mapping of SB-insertion sites

To screen for remobilization events from the primary SB-insertions SB-1 and SB-2, the offsprings of the corresponding "seed males" were prescreened by conventional genotyping PCR. The presence of the SB-insertion (confirmed by a primer located internally of the SB-transgene) and absence of locus specific SB insertions (primers for SB-1 or SB-2) indicated a remobilization event. These candidates were further screened for their exact integration sites in the genome. The protocol for mapping SB-insertion sites was provided by Orsolya Symmons (EMBL, Heidelberg) and is also available on the webpage of TRACER Database (http://www.ebi.ac.uk/panda-srv/tracer/). The method is based on an asymmetric PCR approach. In a first PCR, specific primers (L1 or R1) pointing outward of the SB-transgene and a set of random primers were used (Figure 9a). The random primers contained a 5' adaptor sequence (KmonP), followed by a stretch of seven random nucleotides and a 'seed sequence' (Table 8). The 'seed sequence' has been designed to have a match every 200 to 300 bp approximately in the mouse genome. The first ten cycles of PCR reaction favor the linear amplification from the transgene specific primer (L1 and R1) by high annealing temperatures. In the following cycles, the PCR asymmetry is sustained by repeated high-to-low annealing temperatures. PCRs were set up using Phusion High-Fidelity DNA Polymerase (Thermo Scientific, # F-530S or F-530L).

Condition for 1st PCR round: 25µl reaction: 1x HF buffer, 0.2 mM dNTPs, 0.2 µM L1 or R1, 0.5 µM random primers (KmonP-N7-ctcag and KmonP-N7-tcctg each), 0.5 Units Phusion polymerase, 20 ng DNA template. The following PCR program was used: step 1: 95°C, 3 min; step 2: 94°C, 20 sec; step 3: 63°C, 45 sec; step 4: 72°C, 3 min; step 5: go to step 2 for 9x; step 6: 94°C, 20 sec; step 7: 63°C, 45 sec; step 8: 72°C, 3 min; step 9: 94°C, 20 sec; step 10: 63°C, 45 sec; step 11: 72°C, 3 min; step 12: 94°C, 20 sec; step 13: 44°C, 1 min; step 14: 72°C, 3 min; step 15: go to step 6 for 12x; step 16: 72°C, 7 min; step 17: 16°C, pause.

In the 2nd PCR reaction a 1:100 dilution from the 1st PCR was used as template. Primers L2 or R2 and KmonP were used to obtain flanking sequences from both sides of the SB-transgene integration site. KmonP anneals to an adapter sequence contained in the random primers (Figure 9b).

Condition for 2st PCR round: 25µl reaction: 1x HF buffer, 0.2 mM dNTPs, 0.2 µM L2 or R2, 0.2 µM KmonP, 0.5 Units Phusion polymerase, 1:100 dilution of 1st PCR as template. The following PCR program was used: step 1: 95°C, 4 min; step 2:94°C, 20 s; step 3: 63°C, 45 s; step 4: 72°C, 3 min; step 5: 94°C, 20 s; step 6: 63°C, 45 s; step 7: 72°C, 3 min; step 8: 94°C, 20 s; step 9: 53°C, 45 s; step 10: 72°C, 3 min; step 11: go to step 2 for 9x; step 12:94°C, 20 s; step 13: 63°C, 45 s; step 14: 72°C, 3 min; step 15: 94°C, 20 s; step 16: 53°C, 45 s; step 17: 72°C, 3 min; step 18: 72°C, 7 min; step 19: go to step 12 for 8x; step 20: 16°C, pause.

After the 2nd PCR, a 5 µl aliquot was analyzed on a 1 % agarose gel for quality control. Successful PCR amplification resulted in different banding patterns larger than 300 bp. The samples were then enzymatically purified using 1 µl FastAP (1 U/ µl, Thermo Fisher Scientific, #EF0654) and 0.5 µl ExoI (20 U/ µl, Thermo Fisher Scientific, #EN0581) per 5 µl of PCR product. Purified PCR products were sequenced by Sanger sequencing using the primer L3 or R3 (Figure 9c). Sequences contained about 80 to 90 bp of the SB-transgene specific sequence. The sequence that followed the transgene specific sequence (5'TTCCGACTTCAACTGTA'3) was used for BLAT search against the mouse genome (UCSC genome browser). The mapped integration site for each new SB mouse line was confirmed by conventional genotyping PCR with locus specific primers (Table 8).

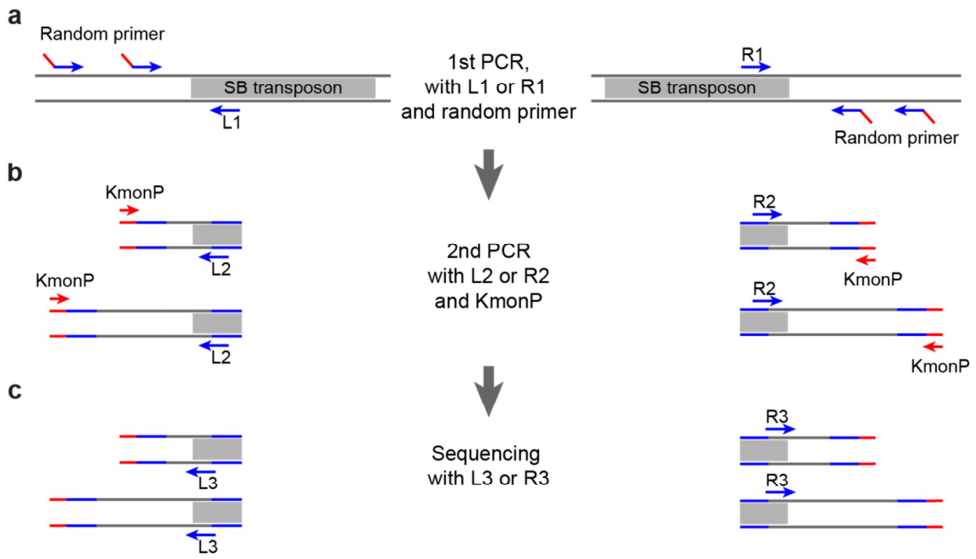


Figure 9: PCR workflow for mapping new SB-insertion sites. a, 1st round of PCR with SB-transgene specific primer L1 or R1 and random primers that map approximately every 200 to 300 bp in the mouse genome. b, 2nd round of PCR with template from 1st PCR, nested SB-transgene-specific primer L2 or R2 and KmonP primer, which binds to adapter sequence of all random primers. c, PCR products are sequenced using nested SB-transgene specific primer L3 or R3 to obtain flanking genomic sequence left and right of the SB-transgene.

4.9 Genotyping of mutant mice

Mouse mutants were genotyped by standard PCR procedures using Taq polymerase produced by Asita Stiege in the research group. The reagents were pipetted on ice into a 0.2 ml reaction tube and DNA was amplified in a thermocycler. PCR conditions for 25 µl reaction: 2.5 µl 10x Taq buffer (750 mM Tris/ HCl pH 8.8; 200 mM $(NH_4)_2$ SO_4; 0.1% Tween 20; 15 mM $MgCl_2$), 0.5 µl dNTPs (12.5 mM), 0.5µl forward primer (10 µM), 0.5 µl reverse primer (10 µM), 0.5 µl Taq enzyme, 1 µl template (~20 ng), 19,5 µl H2O bidest. The following PCR program was used: step 1: 95°C, 3 min; step 2: 95°C, 30 sec; step 3: 60°C, 45 sec; step 4: 72°C, 45 sec; step 5: go to step 2 for 29x; step 6: 72°C, 7 min; step 7: 4°C, pause. Primer sequences for conventional genotyping are listed in Table 8. The PCR products were analyzed on a 1.2 % agarose gel.

4.9.1 Quantitative Real-Time PCR (qRT-PCR)

Copy number analysis from genomic DNA (gDNA) was performed with the SYBR Green I chemistry (Promega) on an ABIPrism 7900 HT thermocycler. Primers were designed with the Primer3Plus online tool with an average product size of 100 bp. The qRT-PCR reaction was set up in a 12 µl reaction on a 384-well plate with the following components: 6 µl of 2x SYBR mix, 2 µl primer (1.5 µM each) and 4µl gDNA (4 ng). Each reaction was performed in triplicates. A standard curve for each primer pair was generated from gDNA that contained the target sequence and was diluted in 1:5 steps (1 – 0.2 – 0.04 – 0.008 – 0.0016). Relative values (starting quantities) for each target (primers within duplicated or deleted region) were then normalized to the reference (primers outside duplicated or deleted region). The genomic coordinates for duplicated or deleted regions are given in Table 12.

4.10 Circular chromosome conformation capture (4C-seq)

The protocol for 4C library preparation described here was adapted from (van de Werken, de Vree, et al., 2012). Mouse tissues from E12.5 or cultured cells were collected and 4C libraries were processed by the established 4C-seq protocol. According to the quality measurements of the 4C protocol by (van de Werken, Landan, et al., 2012), only samples that contained more than 1 million mappable reads after sequencing and that had most of these reads mapping in *cis* (>60 %) relative to the viewpoint were used for analysis. The 4C-seq profiles displayed in this study show the normalized read counts, which represent the interaction frequency of a DNA fragment with a chosen viewpoint fragment.

4.10.1 Crosslinking and nuclei extraction

Embryonic tissue (limb bud, forebrain or liver) was prepared from E12.5 in 1x PBS. The tissue was disaggregated by trypsin treatment for 10 min at 37°C, while disrupting the tissue by pipetting every 2 minutes. Treatment was stopped by adding 5x volume of 10 % FCS/ PBS. To obtain a single cell suspension from trypsinized tissue or patient cells, the solution was filtered through 40 µm cell strainer (Corning, #352340). Cells were centrifuged at 1100 rpm for 5 min and the obtained pellets were then resuspended in 5 ml 10 % FCS/ PBS. Crosslinking was initiated by adding 5 ml of freshly prepared 4 % formaldehyde in 10 % FCS/ PBS. After incubation for 10 min at room temperature while rotating, the crosslinking reaction was stopped by adding 1 ml 1.425M glycine on ice. Another centrifugation step at 1500 rpm and 4°C for 8 min was followed by resuspension in 5 ml freshly prepared, cold lysis buffer and incubation for at least 10 min on ice. To confirm cell lysis, a 3 µl aliquot was mixed on a microscope slide with 3µl of methyl green pyronin staining solution (Waldeck, Pappenheim, #2C-186), which stains pink the cytoplasm and blue the nuclei indicating complete lysis. The number of nuclei was determined using a standard counter chamber for cell culture. Nuclei were pelleted by centrifugation at 2000 rpm at 4°C, washed with 1x PBS and aliquoted in tubes with 2.5-5 x10^6 nuclei. After centrifugation at 2600 rpm for 2 min, nuclei were snap-frozen in liquid nitrogen. Samples can be stored at -80°C for up to 6 months.

4.10.2 Preparation of 3C library

The nuclei pellet was resuspended in 360 µl of water and mixed with 60 µl of 10x restriction buffer. The digestion strategies (1st and 2nd restriction enzyme) used for 3C and 4C library preparation are listed in Table 9. Samples were placed at 37°C in a thermomixer at 900 rpm. Next, 15 µl of 10 % SDS was added and incubated for one hour with occasional pipetting to dissolve the nuclei aggregates. Remaining SDS was separated from the solution by adding 150µl of 10 % Triton X-100. After one hour of incubation, 600 µl of 1x restriction buffer and 400 units of restriction enzyme were added. Additional 200 units of restriction enzyme were added after four hours and again, after overnight incubation at 37°C with shaking at 900 rpm. Meanwhile, a digestion control was tested by agarose gel electrophoresis. For this, a 5 µl aliquot was mixed with 90 µl 10mM Tris pH 7.5 and 2 µl RNase A (10 mg/ ml) and incubated for one hour at 37°C. Chromatin was de-crosslinked by adding 5 µl proteinase K (10 mg/ ml) and incubation at 65°C for four hours. The DNA was then extracted by adding 100 µl phenol-chloroform. Samples were mixed by inverting the tubes and centrifuged for 10 min at 13200 rpm at room temperature. The upper water phase was transferred into a new tube and analyzed on a 1 % agarose gel. Restriction enzyme was heat-inactivated according to manufacturer's instructions. The samples were transferred to 50 ml Falcon tubes and 700 µl 10x ligation buffer was added. The volume was filled up to 7 ml with

water and 50 units of T4 DNA ligase (Thermo Fisher Scientific, #EL0013) were added. The ligation mix was incubated overnight at 16°C. A 100 µl aliquot of de-crosslinked DNA was analyzed on an agarose gel to monitor successful ligation. The preparation of this ligation control was done as described for the digestion control. DNA of final 3C library was de-crosslinked and precipitated. 30 µl of proteinase K (10 mg/ ml) was added to the ligation and incubated overnight at 65°C. Then, 30 µl RNase A (10 mg/ ml) was added and the sample was incubated for 45 min at 37°C. The DNA was then extracted by adding 7 ml phenol-chloroform. The solution was mixed by inverting the tube and the water phase was separated by centrifugation at 3750 rpm for 15 min at room temperature. The following reagents were added to the water phase: 7 ml water 1.5 ml 2M NaAc pH 5.6, 140 µg glycogen, 35 ml 100 % ethanol. All reagents were mixed and placed at -80°C, until the sample was completely frozen. The sample was thawed and centrifuged for 20 min at 8350 g and 4°C. DNA pellet was washed with 30 ml cold 70 % ethanol and centrifuged 15 min at 3300 g at 4°C. Dried pellet was dissolved in 150 µl 10mM Tris pH 7.5. The 3C library was subsequently used for 4C library preparation or capture Hi-C.

4.10.3 Preparation of 4C library

150 µl of a 3C sample was mixed with 295 µl water, 50 µl of 10x restriction buffer, 60 units of the second restriction enzyme and incubated overnight at 37°C. Successful digestion was tested using a 5 µl aliquot of the sample on a 1 % agarose gel. After digestion the enzyme was heat-inactivated by incubation following manufacturer's instructions. The sample was transferred to a 50 ml Falcon tube and 12.1 ml water, 1.4 ml 10x ligation buffer, and 100 units T4 DNA ligase (Thermo Fisher Scientific, #EL0013) were added. Ligation was carried out overnight at 16°C and the ligation was tested on a 1 % agarose gel. After confirming DNA ligation the DNA of the final 4C library was precipitated and purified. 1.4 ml 2M NaAC pH 5.6, 35 µg glycogen and 35 ml 100 % ethanol were added to the sample and placed at -80°C; DNA was pelleted by centrifugation at 8350 g for 45 min at 4°C. The pellet was then washed with 30 ml of 70 % ethanol and centrifuged at 3300 g for 15 min at 4°C. The dry pellet was dissolved in 300 µl 10mM Tris pH 7.5 and the final purification step was carried out using the QIAquick PCR purification kit (Roche, #28106) with one column per 1 $x10^6$ cell input. The concentration was measured on a Nanodrop spectrophotometer and an additional column purification was repeated for samples with A260/ A230 ratios higher than 3.0. The purified 4C libraries were then stored at -20°C.

4.10.4 Inverse PCR and sequencing of 4C samples

4C libraries were used as template for inverse PCR using primers, which were designed according to the region of interest (viewpoint fragment), e.g. the transcription start site of a gene. Primer sequences for inverse PCR and corresponding digestion strategy used in this study are listed in Table 9. Primers were designed as described in (Erik Splinter, de Wit, van de Werken, Klous, & de Laat, 2012). The read primer 1, next to the 1st restriction site and primer 2, close to the 2nd restriction site contained Illumina TrueSeq adapter sequences at their 5' end (read primer 1: CTACACGACGCTCTTCCGATCT; primer 2: CAGACGTGTGCTCTTCCGATCT). Each 4C template and primer pair was tested for amplification cycles to be applied. For this purpose, serial template concentrations of 0.5, 1, 2 and 4 ng/ µl (final concentration) were tested in a 25 µl PCR reaction and analyzed on a 1 % agarose gel. The highest template concentration, which gave a reproducible band pattern and still belonged to the linear phase of the PCR reaction, was selected for the final 4C PCR amplification. A total of 1.6 µg of each 4C library was amplified in 50 µl PCR reactions, pooled and purified using the QIAquick PCR purification kit (Roche, #28106). One column per five PCR reactions was used for purification. Correct PCR amplification was verified on a 1 % agarose gel. Samples were multiplexed and sequenced with Illumina Hi-Seq technology according to standard protocols. Sequencing was done by the sequencing core facility at the Max Planck Institute for Molecular Genetics (Bernd Timmermann) or at the Institute for Medical and Human Genetics, Charité Universitätsmedizin Berlin (Ulrike Krüger and Jochen Hecht). Up to 35 4C libraries were multiplexed, pooled and sequenced on one lane of a Hi-Seq flow cell using 100 bp paired-end or 100 bp single-end sequencing kits. On average ten million reads per 4C library were obtained.

PCR conditions for inverse PCR: For amplification of 4C libraries, the Expand Long Template PCR System (Roche, #11759060001) was used. All reagents were pipetted on ice. The PCR conditions with final concentrations are the following: 1x Buffer, 0.2 mM dNTPs (each), 1 µM primers, 0.075 U/ µl DNA polymerase (Enzyme mix), 4C template in range of 0.5 ng/ µl to 4 ng/ µl, reaction filled up with water to desired volume. 4C samples were amplified in a thermocycler with the following program: 1. Initial denaturation, 94°C for 2 min; 2. Denaturation, 94°C for 15 sec; 3. Primer annealing, 55 °C for 1 min; 4 Elongation, 68°C for 3 min; 5, Go to step 2 with 29 additional cycles; 6 Final elongation, 68°C for 7 min; 8, End, 4°C.

4.10.5 4C-seq data analysis

An in-house pipeline for 4C data analysis was established in cooperation with Verena Heinrich (Department Computational Molecular Biology, Max Planck Institute for Molecular Genetics, Berlin). Processing of data was done by Verena Heinrich. The primer sequences were clipped from short sequencing reads and quality assurance was done as suggested by (van de Werken, Landan, et al., 2012) with a customized Java program and the remaining reads were mapped to the reference sequences GRCh37/hg19 or NCBI37/mm9 using BWA (Li & Durbin, 2009) . Each 4C library fulfilled the following quality criteria: more than one million mapped reads and *cis*/ overall ratio of mapped reads >= 60%. BedGraph tracks for a normalized number of reads (reads per million, RPM) that overlapped with fragments of first restriction enzyme were created in a specified genomic range (chr11:108,500,000-115,000,000 for NCBI37/mm9 and chr17:64,999,999-72,000,000 for GRCh37/hg19) and smoothed over a specified number of fragments using customized Java programs. The viewpoint and adjacent fragments 1.5 kb up- and downstream were removed and a window of ten fragments was chosen to normalize the data per million mapped reads (reads per million, RPM). BedGraph files of single 4C experiments were visualized as tracks in the UCSC browser. 4C tracks used in Figures of this study visualize the following genomic coordinates (chr11:109,689,384-113,519,619 for NCBI37/mm9 and chr17:66,967,346-71,165,753 for GRCh37/hg19). For comparing interaction profiles of different samples, the log2 fold change (ratio) was obtained for each window of normalized reads. Scaling differences during the normalization induced by the extra copy in mutants were corrected by excluding the duplicated part from the mutant as well as from the wild-type allele. To test the reproducibility of 4C-seq experiments, biological and technical replicates were compared for the *Sox9*, *Kcnj2* and *LacZ* viewpoint. Biological replicates represent the independent preparation of tissues and 4C-libraries whereas technical replicates used the same 4C library but independent 4C inverse PCRs. The correlation (Pearson) indicated high reproducibility between biological replicates, ranging from 0.79 to 0.9 (Figure 34). Comparison of technical replicates showed an even higher reproducibility (from 0.89 to 0.94) indicating that tissue preparation and fixation contribute slightly to variation in independent experiments. Two biological replicates and the corresponding ratio are represented in Figure 34a. In all other Figures a representative result is shown.

4.11 Capture Hi-C

4.11.1 SureSelect design

The library of SureSelect enrichment probes were designed over the genomic interval chr11:109,010,000-114,878,000 (mm9) for mouse and chr17:66,966,822-71,540,661 (hg19) for human samples using the online tool of Agilent: SureDesign. Probes are covering the entire genomic region and were not designed specifically in proximity of *DpnII* sites. The probes covered 88 % of the interval.

4.11.2 Capture Hi-C library preparation

3C libraries were prepared as described in section 4.10.1 and 4.10.2, i.e. crosslinking, cell lysis, *DpnII* digestion, ligation, and de-crosslinking. Re-ligated products were then sheared using a Covaris sonicator (Duty cycle: 10 %, Intensity: 5, Cycles per Burst: 200, Time: 6 cycles of 60 sec each, Set Mode: Frequency sweeping, Temperature: 4° to 7°C). Adaptors were added to the sheared DNA and amplified according to Agilent instructions for Illumina sequencing. The library was hybridized to the custom-designed SureSelect beads and indexed for sequencing (50 or 100 bp paired-end) following Agilent instructions.

4.11.3 Capture Hi-C data analysis

Data analysis was performed in cooperation with Robert Schöpflin (Department Computational Molecular Biology, Max Planck Institute for Molecular Genetics, Berlin). Pre-processing and mapping of paired end sequencing data, as well as filtering of mapped paired-end di-tags was performed with the HiCUP pipeline v0.5.8 (Wingett et al., 2015). The pipeline used Bowtie2 v2.2.6 (Langmead & Salzberg, 2012) for mapping short reads to reference genome (NCBI37/mm9). Filtered di-tags were further processed with Juicebox (Rao et al., 2014) command line tools to bin di-tags (10 kb bins) for normalizing by KR normalization. For this, only reads with a quality score, MAPQ >= 30 were considered. The DNA-capturing step enriches the genomic region chr11:109,010,001-114,878,000 on mm9 leading to three different regimes in the cHi-C map: (i) enriched vs. enriched, (ii) enriched vs. non-enriched, and (iii) non-enriched vs. non-enriched regions. For binning and normalization, only di-tags in regime (i) were considered. Therefore, di-tags were filtered for the enriched region and mm9 coordinates were shifted by -109,010,000 bp. A custom chromosome file containing only the enriched region on chr11 (length 5,868,000 bp) was used for the Juicebox tool (Rao et al., 2014; Durand et al., 2016). After binning and normalization, coordinates were shifted back to their original values. However, duplicated regions generally yielded a higher signal compared to non-duplicated ones, when mapped to the wild-

type reference genome. To avoid a flattening of the signal in duplicated regions by the normalization procedure which balances the whole interaction matrix, such as KR normalization, we used only raw count maps. Subtraction maps which were scaled individually by dividing each value of the matrix by a factor (sum 'masked' triangle matrix/ 10^6), were generated based on raw count maps. The region spanning all tested duplications (chr11:110,772,110-112,514,692) was not considered for the computation of the scaling factor. cHi-C maps of count values and subtraction maps were visualized with the WashU epigenome browser (X. Zhou et al., 2013).

4.11.4 CTCF motif analysis

For analysis of CTCF motif orientation in ChIP-seq peaks (E14.5 limbs for CTCF from ENCODE/LICR), the FIMO algorithm of the MEME suite (Bailey et al. 2009) was used with default parameters. The genomic region of 100-200 bp underlying a CTCF peak was used as input sequence. The CTCF motif matrix used as input corresponds to the position weight matrix (PWM) of Barski et al., 2007, downloaded from the Jaspar database (http://jaspar.genereg.net).

4.12 Whole genome sequencing of patient samples

To identify the duplication breakpoint and to screen for SNPs in the duplicated sequence, the genome of a patient displaying an inter-TAD duplication but normal phenotype was sequenced using 100 bp paired-end sequencing on an Illumina HiSeq system. Sequencing was carried out at the Institute for Medical and Human Genetics, Charité Universitätsmedizin Berlin (Ulrike Krüger) and bioinformatic analysis was performed by Peter Krawitz (Institute for Medical and Human Genetics, Charité Universitätsmedizin Berlin). Raw sequence data yielded 126 Gb of sequences that corresponded to 40x mean coverage of the genome. Sequence reads were mapped to the human reference GRCh37.3 using BWA (Li & Durbin, 2009). The region of interest was then analyzed for duplications using a structural variation discovery method, called DELLY, that integrates short insert paired-ends, long-range mate-pairs and split-read alignments to accurately delineate genomic rearrangements at single-nucleotide resolution (Rausch et al., 2012). The tandem duplication was identified and the breakpoint was mapped at the following coordinate positions, chr17:68,195,430 and chr17:69,981,335 (GRCh37/hg19). The breakpoint contained additional micro-duplication of 11 bp (ATGTGTTCAAA). The breakpoint was subsequently confirmed by conventional PCR and Sanger sequencing of the PCR product, using the following primers: 766BP_F1: GGGATCCTCATCAAAACAACC and 767BP_R1: GAATGAAATCATGCCCTTGG.

4.12.1 SNP analysis of 4C-seq data derived from patient cells

SNP analysis was performed in cooperation with Verena Heinrich (Department Computational Molecular Biology, Max Planck Institute for Molecular Genetics, Berlin). To identify selective interactions in a patient with an inter-TAD duplication (normal phenotype), raw reads of 4C-seq experiments with viewpoints in *SOX9*, *KCNJ2* and in the duplication breakpoint were mapped using BWA (Li & Durbin, 2009) without pre-processing. Variants were called for all samples together using GATK (v3.4-46; McKenna et al. 2010) for the duplicated region (chr17:68195430-69981335, hg19). Bi-allelic SNP positions with a minimum coverage of 10 reads were selected. They were called homozygous for one allele, when contacted by the breakpoint viewpoint and homozygous for the other allele, when contacted by either the *SOX9* or *KCNJ2* viewpoint. Additionally, whole genome sequencing was performed and alignment and variant calling was done in the same way. Allele frequencies of the WGS experiment were then compared to the allele frequencies of the 4C-seq experiments at the selected SNP positions (Figure 38 in the Appendix).

4.13 RNA expression analysis

4.13.1 RNA extraction

E12.5 distal limbs were micro-dissected from wild-type and mutant embryos and immediately frozen in liquid N_2 and stored at -80°C. 500 µl of Trizol (Ambion RNA Trizol, #15596018) was added in each sample and the tissue was disrupted using a pestle (VWR disposable pestle, #431-0094) and pestle motor (Kontes). Subsequently, 100 µl of chloroform was added and sample was mixed by pulse vortexing for 15 seconds. After centrifugation (10000 rpm, 4 °C), the upper aqueous phase (~200 µl) was transferred to a new 1.5 ml tube and mixed with 1 volume of 70 % ethanol by pipetting. The solution was loaded onto a column of the RNeasy Mini Kit (QIAGEN) and the following steps, including on column DNase digestion, correspond to manufacturer's instructions. After eluting the isolated RNA in 30 µl RNase-free water, the concentration was measured on the Nanodrop.

4.13.2 RNA sequencing

Extracted and purified RNA was poly-A enriched and sequenced (paired-end 50 bp) using Illumina technology following standard protocols. 50 bp paired-end reads were mapped to the mouse reference genome (mm9) using the STAR mapper (Dobin et al., 2013) (splice junctions based on RefSeq; options: --alignIntronMin 20 --alignIntronMax 500000 --outFilterMultimapNmax 5 --outFilterMismatchNmax 10 --outFilterMismatchNoverLmax 0.1). Uniquely mapped read pairs

were assembled and merged using the Cufflinks package, version 2.2.1 with default settings (Trapnell et al., 2012). Transcripts were assembled via Cufflinks to reference gene annotations from Ensembl and the resulting assemblies were merged using Cuffmerge. Differential gene expression analysis was done with the DEseq2 package (Love, Huber, & Anders, 2014).

4.14 RNA in situ Hybridization

Wild-type and mutant embryos at E12.5 were subjected to whole mount *in situ* hybridization using *Sox9* and *Kcnj2* probes to detect mRNA expression throughout development. After labelling the RNA probe complementary to the target mRNA with Digoxigenin, tissue-specific mRNA expression was visualized by a DIG-specific antibody, coupled to a reporter enzyme.

4.14.1 Generation of DIG-labelled probes

The probe sequence was generated by PCR amplification from cDNA of mouse limb bud. Primer sequences for probe amplification are listed in Table 10. The probe was amplified with 3'-A overhangs, size-selected on agarose gel and purified. Subsequently, it was cloned into the pTA-GFP vector, containing 3'-T overhangs and transformed into Top 10 chemo-competent cells. Vector DNA from positive clones was isolated and individual probes were PCR-amplified from the vector using Sp6 and T7 primers. The PCR product was sequenced by Sanger sequencing. Sp6 or T7 RNA polymerase was chosen for *in vitro* transcription from the purified PCR product based on the orientation of the probe integration in the vector. For in vitro transcription, the DIG RNA labelling kit (Roche, # 11175025910) was used according to manufacturer's instructions. The reaction was stopped with 0.2 M EDTA/ DEPC- H_2O (pH 8.0) and RNA was cleaned by LiCl precipitation. The probe was recovered in 100µl DEPC-H20 and stored at -80°C.

4.14.2 Whole-mount in situ Hybridization (WISH)

All solutions used for whole-mount in situ hybridization are listed in Table 3. Buffers and solutions were treated with DEPC to inactivate RNase enzymes. Embryos were dissected in 1x PBS and fixed overnight in 4 % PFA/ PBS at 4°C. Fixed embryos were washed twice with PBST and dehydrated in increasing serial Methanol dilutions in PBST (25 %, 50 %, 75 % Methanol/ PBST, 2x 100 % Methanol, 10 min each) and stored at -20°C. Prior to hybridization, embryos were rehydrated in 75 %, 50 % and 25 % Methanol/ PBST and washed twice with PBST, each step 10 min. After, embryos were bleached in 6 % hydrogen peroxide/ PBST for 1 hour on ice and washed in PBST. According to the embryonic stage, the embryos were further treated with Proteinase K for 3 to 5 min (E11.5 for 3 min, E12.5 for 5 min), washed with PBST/ glycine, PBST and RIPA buffer, and fixed for 20 min in 4 % PFA/ 0.2 % glutaraldehyde in PBS/ 0.1 % Tween 20. Embryos were washed in

PBST, hybe buffer and incubated in hybe buffer at 65°C for 1.5 hours. The RNA probe was diluted 1:100 with hybe buffer, and 100µg/ ml tRNA was added. The probe was denatured at 85°C for 5 min prior to hybridization and then added to the embryos for hybridization overnight at 65°C. The following day all unbound probe was removed by washing twice with hybe buffer at 65°C for 30 min. After cooling down to room temperature, the embryos were washed in 1:1 hybe buffer/ RNase solution for 5 min. They were then incubated twice for 30 min in RNase solution containing 100 µg/ ml RNaseA and moved to 1:1 RNase solution: SSC/ FA/ T for 5 min. In the next step, the embryos were incubated in SSC/ FA/ T for 2x 5 min, 3x 10 min and 6x 20 min at 65°C, cooled down to room temperature and washed in 1:1 SSC/ FA/ T: 1x MABT and twice in 1x MABT for 10 min. The embryos were then incubated for 1 h in 10 % Boehringer's Bocking Reagent in 1x MABT prior to antibody incubation. The antibody anti-Digoxygenin-AP (Roche, # 11093274910) was diluted 1:5000 in blocking solution, added to the embryos and incubated at 4°C overnight. Unbound antibodies were removed by washing with PBST/ levamisole 3x 5 min, 8x 30 min and an overnight incubation at 4°C. Embryos were washed 3x 30 min in alkaline phosphatase buffer and antibody detection was carried out in BM Purple AP-substrate (Roche, #1442074) until a clear signal appeared. Embryos were then washed twice in alkaline phosphatase buffer and fixed in 4 % PFA/ PBS/ 0,2 % glutaraldehyde and 5mM EDTA and stored at 4°C.

4.15 LACZ stainings and optical projection tomography (OPT)

E12.5 mouse embryos were dissected in cold 1x PBS, fixed with fixation solution (1 % formaldehyde, 0.2 % glutaraldehyde, 0.02 % NP-40 in 1x PBS) on ice for 50 min, washed twice in ice-cold PBS and once at room temperature. Fixed embryos were stained overnight for β-galactosidase activity in staining solution (1 mg/ ml X-Gal (Stock solution: 40 mg/ ml DMSO; 5 mM K3Fe(CN)6; 5 mM K4fe(CN)6; 2 mM $MgCl_2$ in 1x PBS), in a humid chamber at 37°C. After staining, embryos were washed in 1x PBS, postfixed in 4 % PFA in PBS and stored at 4°C.

OPT scans were performed by Wibke Schwarzer (EMBL, Heidelberg) with technical support from Felix Bestvater at the DKFZ, Heidelberg. For OPT-Scanning, stained embryos were embedded in 1 % low melt agarose and dehydrated over 1-2 days using methanol (2-3 steps). Subsequently, samples were cleared overnight in BABB solution (1 volume of Benzyl Alcohol, 2 volumes of Benzyl Benzoate). The samples were scanned (802 frames) with a Bioptonics OPT 3001M Scanner under white light (exposure=10-100ms) for LacZ-staining using GFP filter combination (exciter 425nm/ 40nm, emitter 475nm LP) and 320-380ms exposure for autofluorescence. 3D reconstructions were generated using Skyscan software and further analyzed with Amira and Imaris software.

4.16 Histology

4.16.1 Masson-Goldner staining

For Masson-Goldner staining, limbs of newborns (P0) were first fixed in 4 % PFA/ PBS containing 0.5 M EDTA and then embedded in Paraffin. Paraffin sections of 6 µM thickness were then stained with the Merck Masson-Goldner staining kit (Cat-No. 100485) and Meyers Hematoxylin (Cat-No. MHS80) according to manufacturer's instructions.

4.16.2 Skeletal preparations

Embryos at E17.5 or animals at P0 were decapitated and kept in water for 1 hour. After heat shock for 1 min at 65°C, the animals were skinned, disembowelled and incubated overnight in 100 % Ethanol at room temperature. The cartilage was stained blue by incubating the animals in alcian blue staining solution (150 mg/ l Alcian Blue 8GX (Sigma-Aldrich, #A5268), dissolved in 80 % ethanol/ 20 % acetic acid) for up to 20 hours. The animals were rinsed and post-fixed in 100 % ethanol overnight at room temperature. For initial clearing, animals were incubated for 30 min to 1 hour in 1 % KOH/ bidest H_2O. Membranous bone was stained red, using alizarin red staining solution (50 mg/l Alizarin Red (Sigma-Aldrich,# A5533) in 0.2 % KOH/ biddest H20). Staining was performed for up to 2 days with visual inspection of each specimen for proper red staining. Subsequently, remaining tissue was digested for up to 3 days with lowering KOH concentration from 1 % to 0.2 % KOH/ bidest H_2O by adjusting clearing times. Clearing was stopped by placing preparations to increasing glycerin solutions (30 %, 60 %, and 80 % glycerin/ bidest H_2O, for 24 hours each) and then stored in 80 % glycerin.

4.17 Micro-computer tomography

Micro-computer tomography (µCT) was performed in cooperation with Wing-Lee Chan (Institute for Medical and Human Genetics, Charité Universitätsmedizin Berlin). Autopods of seven week old wildtype and mutant mice were scanned using a Skyscan 1172 X-ray microtomography system (Brucker microCT, Belgium) at 5 µm resolution. 3D model reconstruction was done with the Skyscan image analysis software CT-Analyser and CT-volume (Brucker microCT, Belgium).

5 Results

5.1 TAD structure and dynamic chromatin interactions at the Sox9 locus

The genome-wide analysis of chromatin conformation by Hi-C identified TADs. TADs subdivide the genome into large genomic units with higher chromatin interactions within which enhancers and promoters can interact (Dixon et al., 2012; Nora et al., 2012).

As stated before, the *SOX9* locus is characterized by a large gene desert with a total size of 2.5 Mb in humans and 2 Mb in mice. Two genes, *KCNJ2* and *KCNJ16*, that encode for potassium channels are located at the centromeric end, while *SOX9* resides at the telomeric end. These genes are expressed in different tissues, and *Sox9* especially exhibits a complex spatio-temporal expression pattern during development in mice. This is believed to be controlled by CREs located in the gene desert. To investigate chromatin interactions and the overall TAD structure at the *Sox9* locus in mice with high resolution, capture Hi-C (cHi-C) was performed. In addition, circular 3C followed by high-throughput sequencing (4C-seq) was established and carried out to map tissue-specific interactions at the locus.

5.1.1 Mapping the TAD structure at the Sox9 locus

Capture Hi-C was employed to map all interactions over a 5.9 Mb region at the *Sox9* locus (chr11:109,010,000-114,878,000, mm9) in limb buds at E12.5. At this developmental stage, both genes, *Kcnj2* and *Sox9*, are expressed. The interaction map in Figure 10a depicts a triangular interaction matrix, where each point reflects the interaction between two DNA fragments over the linear locus. The different shades of red indicate the interaction frequency between two DNA fragments. Dark red corresponds to high interaction frequencies, while light red or white indicates less or no interactions at all. The interaction matrix revealed that the gene desert upstream of *Sox9* and the locus itself was subdivided into two large TADs (arrowheads in Figure 10a). These TADs had high interaction frequencies within but low or no interaction frequencies between them. The same TAD structure was previously shown by Hi-C experiments in mouse ES cells (Dixon et al., 2012). One TAD in the centromeric portion of the locus contains the *Kcnj2* and *Kcnj16* genes that encode the respective potassium channel proteins and it extends up to 0.53 Mb centromerically and 0.46 Mb telomerically to *Kcnj2*. The second more telomeric TAD possesses only the *Sox9* gene and extends 1.25 Mb centromerically and 0.38 Mb telomerically to *Sox9*. The two TADs are referred to as *Kcnj* TAD (blue) and *Sox9* TAD (ocher). The high frequency of the internal interactions within each TAD suggested a folding into two globular structures in the three-dimensional nuclear space that were separated from each other by TAD boundaries (model

in top right of Figure 10a). The boundaries of each TAD exhibited high interaction frequencies that appeared as red dots in the interaction matrix, indicating DNA looping between TAD boundaries (arrowheads in Figure 10a). Moreover, the cHi-C data also uncovered several internal structures in each TAD. Both TADs were further subdivided into sub-TADs with high frequencies of interactions and additional loops. The *Kcnj* TAD, for example, was further subdivided into centromeric and telomeric parts with the border being between the *Kcnj16* and *Kcnj2* genes.

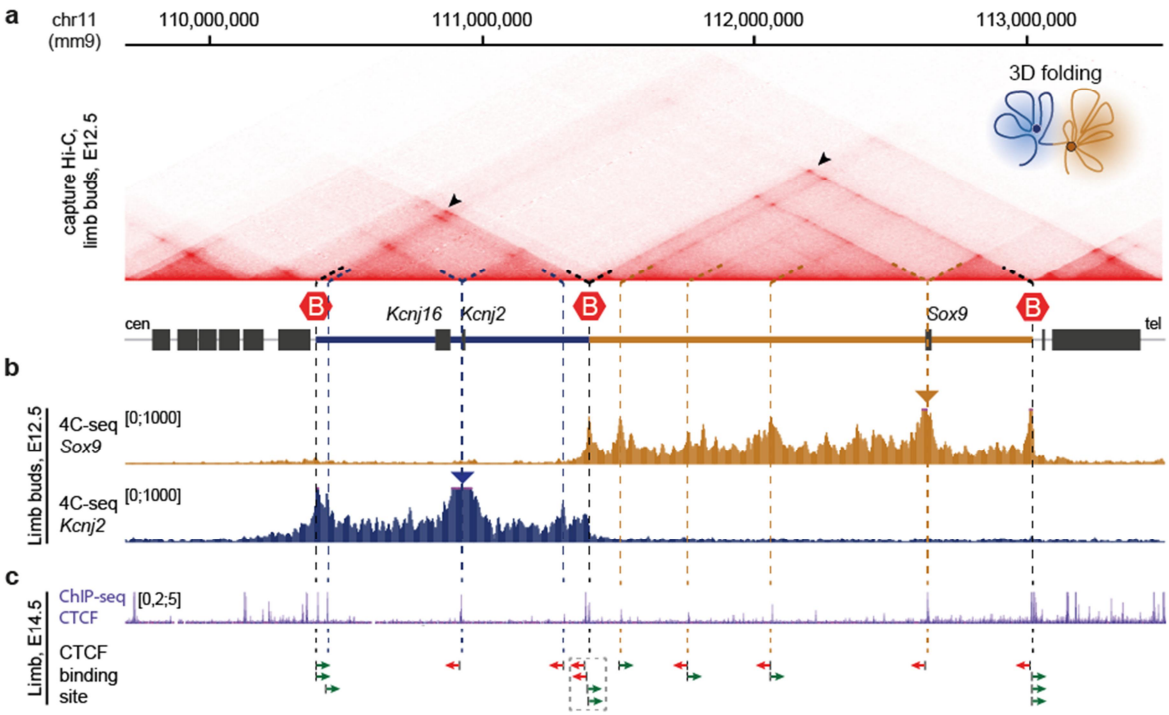


Figure 10: Spatial organization at the *Sox9* locus identified by cHi-C and 4C-seq. a, CHi-C interaction profiles over the extended *Sox9* locus from wild-type E12.5 limb buds showing separation of the locus into two TADs (triangular matrix below arrowheads). The schematic below indicates *Kcnj* TAD (blue), *Sox9* TAD (brown) and TAD boundaries (red hexagons). Frequent interactions within each TAD are restricted by TAD boundaries. The schematic at the top right visualizes a model of three-dimensional chromatin folding of *Kcnj* and *Sox9* TADs. b, 4C-seq from wild-type E12.5 limb buds with viewpoints (triangle) in the promoter of *Sox9* and *Kcnj2*. Frequent interactions from each promoter are confined to the respective TADs, recapitulating TAD structure at the locus. The 4C-seq signal drops off at the TAD boundaries from a high interaction frequency within TADs to almost nothing more than background signals outside them. C, ChIP-seq signal for CTCF from limbs at E14.5 (ENCODE/LICR) and orientation of CTCF-binding sites CTCF binding. Dotted vertical lines depict strong interactions (looping of chromatin) from 4C-seq profiles (peaks) and cHi-C (red dots). Looping of chromatin is associated with CTFC binding and convergent orientation of CTCF-binding sites, e.g., loop formation between TAD boundaries (arrowheads in a). Dotted grey box indicates cluster of CTCF-sites at the TAD boundary between *Kcnj* and *Sox9* TAD.

In this study, 4C-seq was used as a complementary method to cHi-C. 4C-seq from E12.5 limb bud cells was performed and compared to cHi-C data. In contrast to cHi-C, 4C-seq identifies all interacting DNA fragments with a selected DNA fragment, referred to as a viewpoint. The *Kcnj2* and *Sox9* promoters were used as viewpoints to map chromatin interactions in the *Kcnj* and *Sox9* TADs, respectively. The 4C-seq profiles in Figure 10b display the normalized read counts, which signify the interaction frequency of a DNA fragment with the viewpoint fragment. The two profiles showed typical characteristics for 4C-seq experiments. Regions near the viewpoint showed higher read counts that resulted in a broad peak on both sides of it. This high interaction frequency declines along with the increase of the linear distance to the viewpoint. However, each profile exhibited a broad interaction domain with consistently high interaction frequency corresponding to the high internal interactions of the *Kcnj* and *Sox9* TADs as seen with cHi-C. The restriction of the 4C interactions to each TAD from both viewpoints was striking. The physical separation between the two TADs by the boundary was clearly visible by the abrupt drop in 4C interaction frequency. The 4C-seq signal from the viewpoint diminished to an overall background signal outside each TAD. Although cHi-C and 4C-seq differ in their process for quantifying interactions and 4C-seq is limited to one viewpoint, several features of chromatin interactions were captured with both techniques. Specifically, they map the general TAD organization at the locus. In addition, several prominent peaks were captured in the 4C profiles, e.g., between the *Sox9* viewpoint and boundaries of the *Sox9* TAD, which corresponded to a subset of DNA loops detected in cHi-C (vertical dotted lines in Figure 10a andFigure 10b).

Interestingly, the overlap of the cHi-C and 4C-seq data with ChIP-seq data for CTCF from limb tissue at E14.5 (ENCODE/LICR) showed that interactions, captured as focal points in cHi-C and as peaks in 4C-seq, correlated with CTCF binding (Figure 10c). CTCF binding has previously been shown to be a structural component for chromatin interactions and TAD integrity (Dixon et al., 2012; Sanborn et al., 2015). CTCF facilitates long-range interactions by two recognition motifs with a convergent orientation, meaning that one interacting motif is in forward orientation and the other in a reverse orientation (Rao et al., 2014; Sanborn et al., 2015). An approximately 200 bp sequence underlying ChIP-seq peaks engaged in chromatin looping at the locus was further investigated for CTCF-binding sites and their orientation (Bailey et al., 2009). The majority of CTCF-binding sites between the observed loops featured a convergent motif orientation (Figure 10c). For example, the TAD boundary that separated the *Kcnj* and *Sox9* TADs was typified by a cluster of four CTCF peaks. Within the cluster, CTCF-binding sites exhibited a convergent orientation with their interaction partners and three out of the four were highly conserved among mammals (Figure 11).

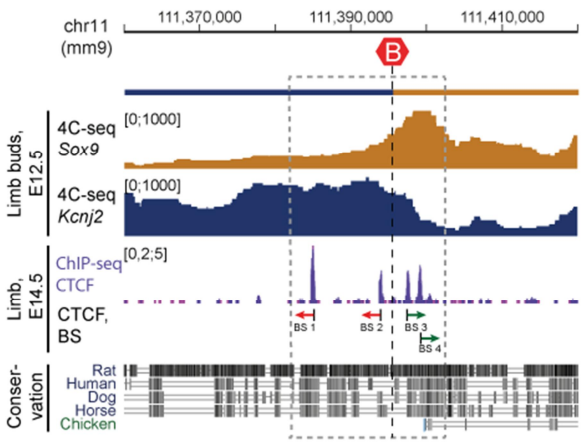


Figure 11: TAD boundary between *Kcnj* and *Sox9* TADs is associated with cluster of CTCF-binding sites. Close-up view of the TAD boundary that separates the *Kcnj* and *Sox9* TADs (related to Figure 10), showing 4C-seq from *Sox9* and *Kcnj2* viewpoints. Below, CTCF binding (ChIP-seq from limbs at E14.5, ENCODE/LICR), orientation of CTCF-binding sites and conservation among vertebrates are shown. Strong interactions from each 4C-seq viewpoint are associated with CTCF binding and convergent orientation of binding sites. Note the signal drop-off of 4C-seq profiles in the middle of the CTCF cluster. Grey dashed box indicates 18 kb genomic region, containing CTCF cluster.

Overall, cHi-C demonstrated that the locus was spatially separated into the *Kcnj* TAD, containing *Kcnj2* and *Kcnj16,* and the *Sox9* TAD, featuring the actual *Sox9* TF. Further, the high resolution of the cHi-C data depicted prominent interactions, e.g., loop formation of TAD boundaries and further compartmentalization of both TADs into sub-TADs that were associated with chromatin looping. The observed chromatin loops overlapped with CTCF binding. Although restricted to one viewpoint, 4C-seq recapitulated the TAD structure at the locus and overlapped with cHi-C observed looping in high resolution.

5.1.2 Using 4C-seq to map tissue-specific chromatin interactions

To learn more about the dynamics of chromatin interactions at the locus, 4C-seq from two additional tissues (forebrain and liver) was performed with the *Sox9* promoter as the viewpoint. The resulting interaction profiles were compared to that of limb bud cells (Figure 12a). It was seen that *Sox9* is strongly expressed in the ventricular zone of the forebrain but absent in liver cells. Cells from limb buds and forebrain, which express *Sox9,* exhibited a high and comparable interaction profile within the *Sox9* TAD. Despite these similarities, tissue-specific changes in the interaction profiles were discerned. Forebrain cells had higher interaction frequencies telomerically to *Sox9* and lower frequencies with regions centromeric to *Sox9* versus limb bud cells. The observed tissue-specific interactions correlated with the presence of potentially active CREs in the region, as demonstrated by the histone modification H3K27 acetylation (H3K27ac) in Figure 12a (ChIP-seq from ENCODE, E14.5 from limbs, brain and liver). Limb cells featured high ChIP-seq signals distributed over the entire *Sox9* TAD, whereas forebrain cells had lower signal centromerically to *Sox9* but higher signal telomerically to *Sox9*. These tissue-specific contacts suggested that the *Sox9* promoter requires close proximity to the active CREs for tissue-specific gene expression. This was in line with the 4C-interaction profile from liver cells that did not express *Sox9*. Compared to limb bud cells, the interaction frequency was reduced over the whole of the *Sox9* TAD. The reduced frequencies were in correlation with the low or almost complete depletion of H3K27ac signal within the *Sox9* TAD. Nevertheless, the interaction profile from liver

cells resembled that of cells that expressed *Sox9,* such as a high interaction frequency within the *Sox9* TAD and a signal drop-off at the TAD boundaries, especially towards the telomeric TAD boundary. In addition, cell-type invariant interactions that correlated with the presence of CTCF binding were observed in all three tissues (arrowheads in Figure 12).

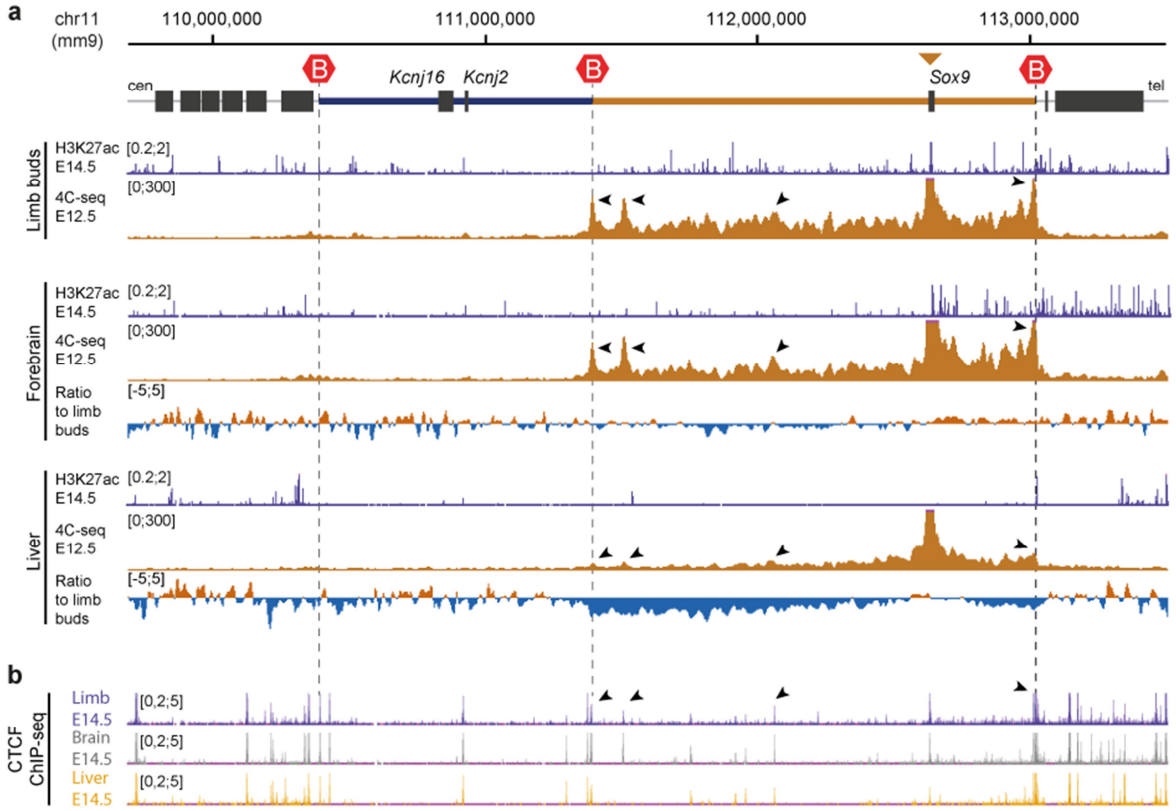


Figure 12: Tissue-specific interactions captured by 4C-seq. a, Schematic of the locus featuring the *Kcnj* TAD (blue) and *Sox9* TAD (brown). Red hexagons and dotted lines indicate TAD boundaries of the *Kcnj* and *Sox9* TADs. Below, 4C-seq profiles using *Sox9* promoter as viewpoint from E12.5 limb buds, forebrain and liver cells. Log2 ratio depicts changes in interaction frequency of either forebrain or liver compared to limb bud cells. ChIP-seq signal for H3K27ac histone modification from equivalent tissues at E14.5 from ENCODE/LICR are shown above each 4C-seq profile. Note that for tissues expressing *Sox9* (limb bud and forebrain), high interaction frequencies within the *Sox9* TAD and the *Sox9* promoter preferentially contact regions enriched for H3K27ac. In contrast, liver cells, negative for *Sox9* expression and with low H3K27ac coverage, exhibit overall reduced interaction frequencies within the contact domain. Cell-type invariant peaks (arrowhead) in the 4C-seq interaction profiles could be detected in all tested tissues. b, CTCF ChIP-seq track from liver (adult, 8 weeks), whole brain and limb at E14.5 from ENCODE/LICR. Cell-type invariant interaction peaks are enriched for regions with CTCF binding.

The analysis of tissue-specific interactions showed that the 4C interactions within the *Sox9* TAD were correlated with the transcriptional status of *Sox9*. Tissues that expressed *Sox9* preferentially interacted with regions in the *Sox9* TAD that contained active CREs. Despite these specific interactions, tissue-invariant interactions and the *Sox9* TAD structure were captured in all tested tissues independent of the expression status of *Sox9* or the presence of active CREs. The observed cell-type invariant interactions were associated with CTCF binding, corroborating the role of CTCF in establishing the three-dimensional organization of chromatin, independent of the transcriptional status of *Sox9* at the locus (arrowheads in Figure 12b).

5.2 The regulatory organization of the Kcnj and Sox9 TADs

Compartmentalization into TADs suggests that genomic regions within TADs are under the influence of the same *cis*-regulatory information, potentially defining a regulatory domain structure, as previously shown in a genome-wide fashion (Symmons et al., 2014). To systematically investigate the regulatory potential at different positions within the *Kcnj* and *Sox9* TADs, the SB transposition system was employed (Ruf et al., 2011). It consists of a SB transgene containing a *loxP* site and a *LacZ* reporter gene both flanked by SB transposons. The *LacZ* reporter gene is controlled by a neutral *beta-globin* promoter that has no activity on its own but senses the regulatory influence from endogenous CREs at integration sites (Ruf et al., 2011). Creating multiple mouse alleles with a single insertion of the SB transgene allows determining the range of regulatory activity within the *Kcnj* and *Sox9* TADs during mouse development. Additionally, the system serves as a tool to distribute *loxP* sites at the locus used for Cre/*loxP*-mediated recombination to induce genomic rearrangements *in vivo* (Hérault et al., 1998).

5.2.1 Generation of multiple SB insertions

To create mouse lines with an insertion of the SB transgene at the locus, two mouse alleles, SB-1 and SB-2, were generated by homologous recombination in E14 mouse ESCs. The correct recombination and integration of the transgene into the genome was confirmed by southern blot analysis (see Figure 35 in the Appendix). The SB-1 allele carried the SB transgene in the *Kcnj* TAD, approximately 30 kb telomerically to *Kcnj2,* and the SB-2 carried the SB-transgene 128 kb centromerically to *Sox9* (Figure 13a).

The advantage of the SB transposon system is the simultaneous generation of additional SB insertions at the locus *in vivo* (Ruf et al., 2011). The established mouse lines, SB-1 and SB-2, were utilized to remobilize the SB transgene to new locations in the genome by a simple breeding strategy. To do so, each mouse line was crossed with a line carrying the highly active HSB16 SB transposase (*Prm1::HSB16A*). The transposase initiated the transposition of the SB transgene in a ''cut and paste'' manner during spermatogenesis of the double transgenic mice. Double transgenic males were then crossed with wild-type females and the offspring was screened for remobilization events. The SB transgene remobilized with similar frequencies from the SB-1 and SB-2 allele, that is 14,5 % (51 out of 351 screened animals) for SB-1 and 11,7 % (46 out of 393 screened animals) for SB-2 (Figure 13b). More than 85 % of new integration sites could be mapped onto the genome with the majority occurring on a different chromosome (inter-chromosomal – 76,5 % for SB-1 and 63 % for SB-2). However, because of the bias of the SB transposition system for reintegrating close to the starting point from the SB-1 and SB-2 insertions, new insertions in the *Kcnj* and *Sox9* TADs were recovered. Nine and five animals with

the SB-1 and SB-2 alleles, respectively, could be obtained with an insertion within 2 Mb of the starting point. In total, 11 mouse lines with a single SB insertion at the locus were generated and maintained (Figure 13c). The exact genomic location and orientation of the SB insertions are listed in Table 11. The mouse lines were further used for the analysis of their *LacZ* reporter activity at the integration site and for targeted genomic rearrangements using the Cre/*loxP* recombination system.

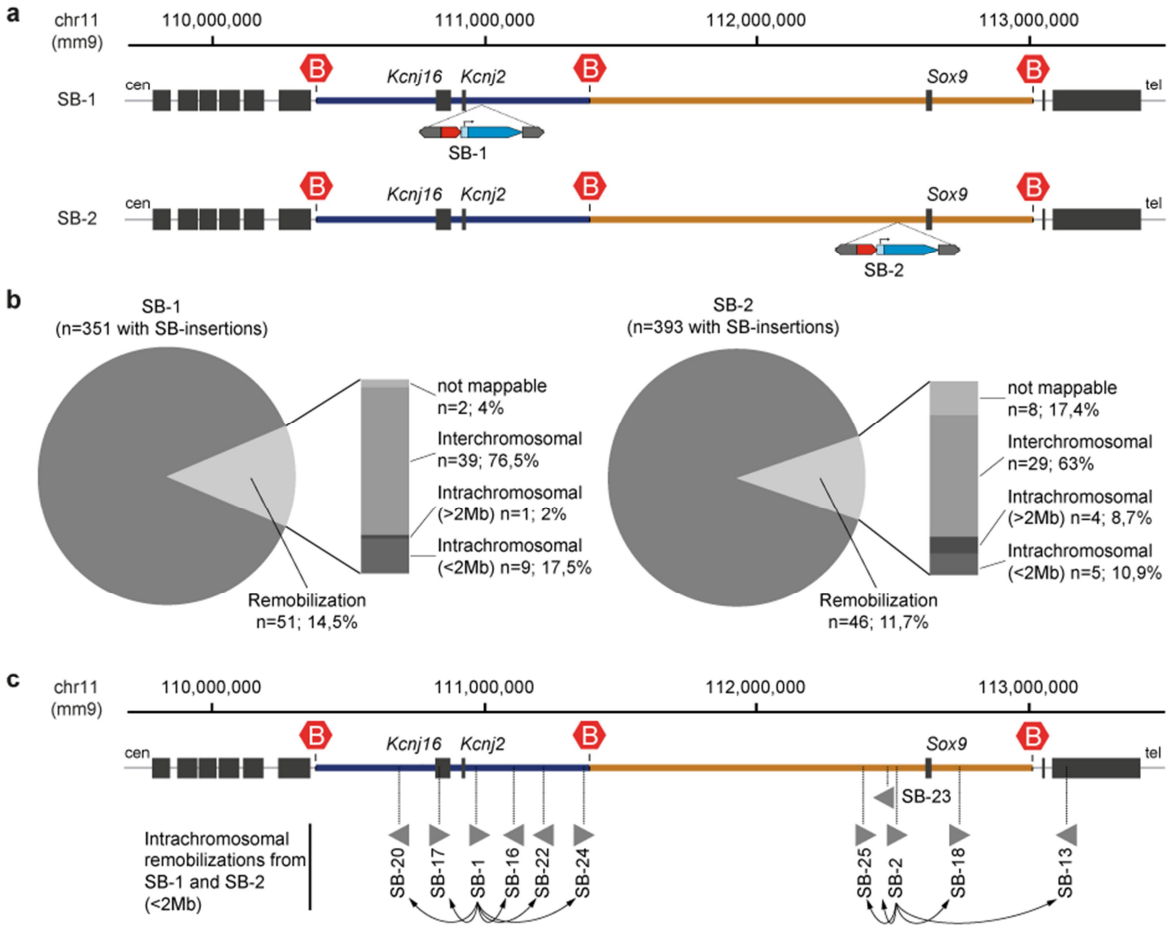


Figure 13: Generation of multiple SB insertions covering the *Kcnj* and *Sox9* TADs. a, Schematic representation of the two targeted mouse alleles, SB-1 and SB-2, depicting the position of the SB transgene (red, *loxP* site with orientation; blue, *LacZ* reporter driven by *beta-globin* promoter; grey, SB transposons). b, Remobilization efficiency from SB-1 (left) and SB-2 (right). The pie charts depict the number of remobilization events (14,5 % for SB-1 and 11,7 % for SB-2) among all mice carrying the SB transgene. The bar charts portray the number of new integration sites within the genome relative to the initial integration site, i.e., not mappable locations, mapped on a different chromosome (inter-chromosomal) or on the same chromosome (intra-chromosomal) within or greater than 2 Mb of its starting point. C, Summary of recovered SB insertions at the *Sox9* locus shown as triangles with insertion number. Each triangle represents an individual mouse line. The orientation of the triangle indicates the orientation of the *loxP* site in the SB transgene. Arrows signify the origin of the new insertions of the remobilization screen from the SB-1 and SB-2 alleles.

5.2.2 The Kcnj and Sox9 TADs consist of α complex regulatory domain structure

To determine the regulatory potential within the *Kcnj* and *Sox9* TADs, all recovered mouse lines carrying a SB transgene insertion were analyzed for *LacZ* reporter activity and compared to endogenous expression of *Kcnj2*, *Kncj16* and *Sox9* (Figure 14). To do so, mice were mated with wild-type animals and embryos were fixed and stained for LACZ activity at the E12.5 developmental stage. It is at this stage that *Kcnj2* and *Sox9* exhibited a complex and specific expression pattern – *Kcnj2* was expressed in the maxilla and limb mesenchyme whereas *Sox9* showed strong expression in the cartilage condensations of the skeleton. Contrarily, *Kcnj16* showed no specific expression pattern at this developmental stage (Figure 14b and 14c).

Six SB insertions located within the *Kcnj* TAD showed consistent *LacZ* expression in the maxilla and limb mesenchyme resembling the expression pattern of *Kcnj2* (Figure 14b and 14d). Expression in the maxilla was observed over a genomic region of 630 kb from the insertion SB-20 to SB-24, defining the maxillary regulatory domain. The limb regulatory domain covered a region of 410 kb from SB-20 to SB-22. Despite the large overlap, regional differences in expression were detected for the insertions SB-20 and SB-17. Both insertions were located centromerically to *Kcnj2* and showed additional expression in the head of the embryo, indicating a subdivision of regulatory activity within the *Kcnj* TAD. Of interest was that *Kcnj16,* embedded in this complex regulatory domain structure with expression domains in the maxilla and the limb, did not respond to this *cis*-regulatory influence. In contrast, the SB-17 insertion, which was located in the intron of *Kcnj16*, less than 20 kb downstream of the transcriptional start site, did react to the regulatory information. These differences in promoter activation suggested that the promoter of *Kcnj16* was repressed or not primed for transcription in these tissues. Indeed, *Kcnj16* is expressed in the glomeruli of the kidney and in the floor plate of the hindbrain and midbrain at E14.5 (Diez-Roux et al., 2011, Eurexpress 2016). No specific expression domains were detected in the kidney by LACZ activity or by whole-mount *in situ* hybridization for *Kcnj16* mRNA at the E12.5 stage.

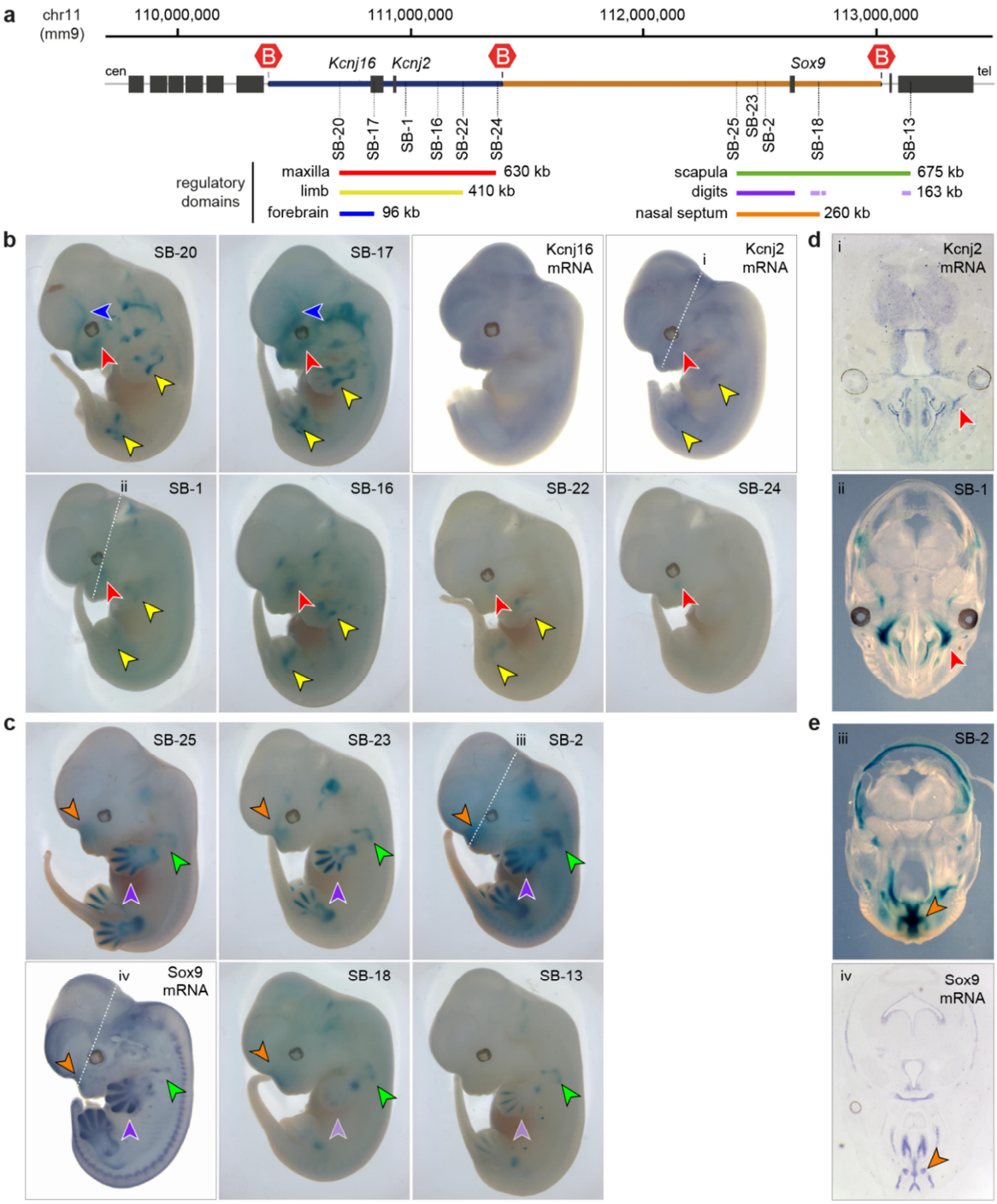


Figure 14: *LacZ* reporter activity of SB insertion lines reveals complex regulatory domain structures. a, Schematic representation of the locus depicting *Kcnj* (blue) and *Sox9* (brown), TAD structure and position of SB insertions (dotted lines) used for *LacZ* expression analysis. The overlapping expression patterns of adjacent insertions and genes at the locus (shown in b and c) define broad overlapping regulatory domains indicated by colored bars, i.e., expression in the maxilla (red), limb mesenchyme (yellow) and forebrain (blue) within the *Kcnj* TAD; expression in the nasal septum and the capsule (orange), digits (blue) and scapula (green) within the *Sox9* TAD. The extent of the regulatory domains is indicated in kilo bases (kb). b and c portray whole-mount LACZ staining in E12.5 embryos of SB mouse lines and whole-mount mRNA *in situ* hybridization for *Kcnj16*, *Kcnj2* and *Sox9* in wild-type embryos. Arrows highlight expression domains in the embryo and the color code corresponds to the regulatory domains as described before. Section planes are indicated by white dotted lines and the corresponding sections are shown in d and e. d and e, Sections of E13.5 embryos. d highlights the expression pattern in the maxilla (red arrow) of *Kcnj2* mRNA (i) and LACZ signal for SB-1 (ii) and e highlights the expression in the cartilage primordium of the nasal septum and capsule of *Sox9* mRNA (iii) and LACZ signal for SB-2 (iv).

A similarly complex regulatory domain structure was observed in the *Sox9* TAD. Four insertions, SB-25 to SB-18, recapitulated *Sox9* expression in the developing skeletal system. Consistent *LacZ* expression was detected in the cartilage condensations of the nasal septum, scapula and digits over a region of 260 kb (Figure 14a andFigure 14c). Digit expression of the SB-18 insertion, however, was much weaker than for SB insertions centromeric of *Sox9*. Surprisingly, the SB-13 insertion, located outside the *Sox9* TAD in the *Slc39a11* gene, exhibited similar expression as the insertions within the *Sox9* TAD, suggesting a broader extent of regulatory domains into the neighboring TAD. The scapula expression pattern was recapitulated to a high degree, whereas digit expression was weaker and differed in terms of pattern. However, the exact location of the CREs driving regulatory activity could not be determined with this assay.

5.2.3 Regulatory domains overlap with TAD structure at the Sox9 locus

In the next step, the regulatory domain structure was compared to the TAD structure. First, 4C-seq experiments were carried out to test the influence of SB insertions on TAD structure and to examine with which regions the *LacZ* promoter interacted. Therefore, limb buds from E12.5 heterozygous SB-2 mice were used and viewpoints were designed at the *Sox9* and *LacZ* promoter (Figure 15). The comparison of the *Sox9* viewpoint in SB-2 and in wild-type mice showed that the integration of the SB transgene did not affect the *Sox9* interaction profile. Intriguingly, the *LacZ* viewpoint further demonstrated that the SB transgene was completely incorporated into the endogenous chromatin structure. The *LacZ* promoter represents a unique sequence in the mouse genome and, therefore, the viewpoint demonstrated specific interactions that were restricted to the allele carrying the SB transgene. As described for *Sox9*, the *LacZ* promoter interacted with the region confined to the whole *Sox9* TAD and the range of interactions were limited by the TAD boundaries. Moreover, both viewpoints exhibited a similar interaction pattern although located at different positions within the TAD. They interacted with the same regions and produced similar peak distributions. Nonetheless, there were local differences between the viewpoints, e.g., the *LacZ* promoter interacted preferentially with the centromeric part of the *Sox9* TAD upstream of *Sox9*.

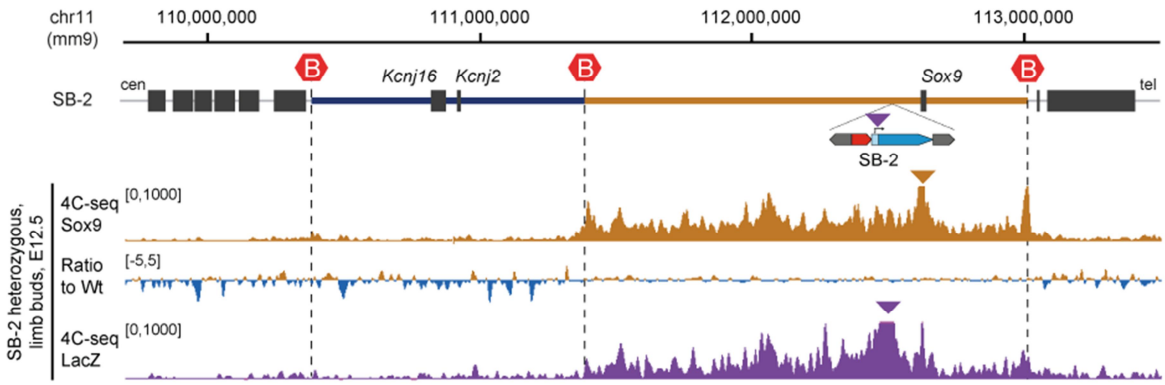


Figure 15: SB insertions are integrated into the TAD structure. Schematic representation of the SB-2 allele depicting the position of the SB transgene (red, *loxP* site with orientation; blue, *LacZ* reporter driven by *beta-globin* promoter; grey, SB transposons) upstream of *Sox9* and within the *Sox9* TAD (brown). 4C-seq data is from E12.5 limb buds of mice heterozygous for SB-2. The 4C profile from the *Sox9* viewpoint and corresponding log2 ratio to the wild-type mice shows that the transgene does not affect the *Sox9* TAD structure. The viewpoint from the *LacZ* promoter (purple) reveals that the transgene is incorporated into the pre-existing *Sox9* TAD structure.

4C-seq with *LacZ* as a viewpoint demonstrated that the SB transgene in SB-2 mice was embedded in the existing chromatin interactions of the *Sox9* TAD. To relate the regulatory domain structure with the spatial organization, the position of all SB insertions and their regulatory activity was compared to cHi-C data (Figure 16). This uncovered an existing overlap of the TAD and the regulatory domain structure. For instance, almost all SB insertions within the *Kcnj* TAD and the *Kcnj2* gene shared expression in the limb and the maxilla. Of interest was that the subdomain structure of the *Kcnj* TAD with two small sub-TADs correlated with the differences in the *LacZ* expression observed by the SB insertions (arrows in Figure 16a). The insertions, SB-20 and SB-17, located in the centromeric sub-TAD, were also expressed in the forebrain. However, this forebrain expression was not detected for *Kcnj2* or in embryos with SB insertions in the telomeric sub-TADs. This suggested that the subdivision into TADs and smaller sub-TADs probably restricted the activity of the CREs to a certain region and, therefore, defined the range of their activity.

A similar spatial and functional subdivision was detected in the *Sox9* TAD. LACZ staining of SB insertions around *Sox9* showed *Sox9*-like expression patterns, e.g., in the scapula, digits and nasal septum. Expression in the scapula and the nasal septum was detected in all lines with SB insertions in the *Sox9* TAD, whereas strong digit expression was observed only for *Sox9* and SB insertions in the centromeric *Sox9* sub-TAD (arrows in Figure 16a). 4C-seq from the *LacZ* promoter confirmed that the SB-2 insertion preferentially contacted the centromeric sub-TAD in the limb bud (Figure 15), which contained the majority of limb-specific CREs as previously shown by H3K27 acetylation (Figure 12).

Taken together, the complex regulatory domain structures reflected by *LacZ* expression analysis of several SB insertions indicated a functional subdivision at the locus. It was noteworthy that the comparison of the TAD structure with the regulatory domain structure exhibited that the

functional subdivision overlapped with the spatial compartmentalization of the *Kcnj* and *Sox9* TADs. Genomic regions with high interaction frequencies, such as TADs and sub-TADs, shared the same regulatory domains. The analysis demonstrated that the *Sox9* locus was spatially and functionally subdivided and underscored the function of TADs as genomic units that guide regulatory information to their target genes.

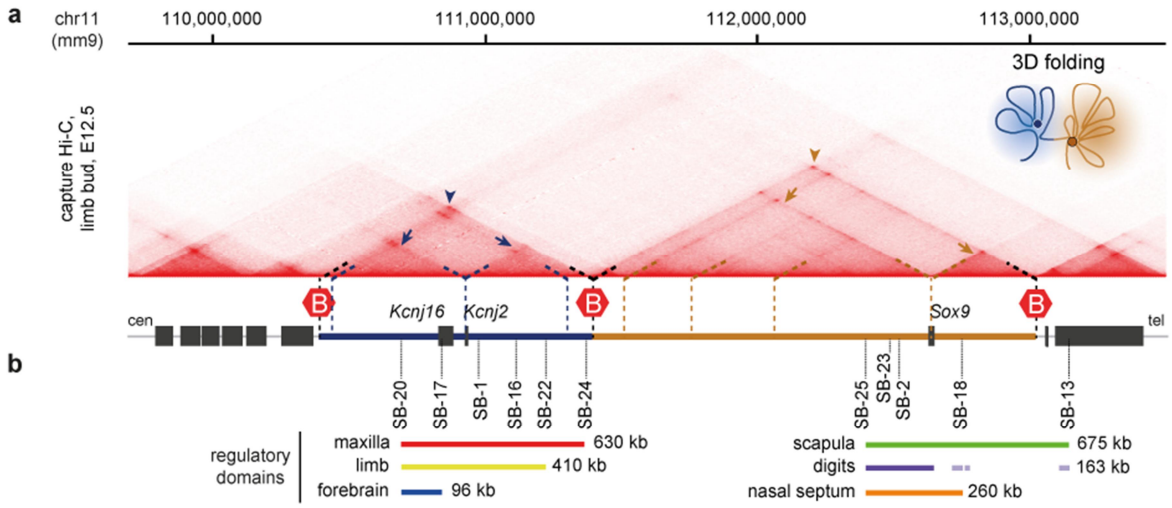


Figure 16: Regulatory domain structure overlaps with TAD structure. a, cHi-C of the *Sox9* locus from wild-type E12.5 limb buds with schematic representation of the *Sox9* locus below depicting *Kcnj* TAD (blue) and *Sox9* (brown) TAD and TAD boundaries (red hexagons). Arrows indicate main TADs and sub-TADs structure. The schematic at the top right indicates three-dimensional chromatin folding of the locus. b, Vertical lines indicate position of SB insertions that were used for *LacZ* expression analysis. Detected regulatory domains and their range in kilo bases (kb) are indicated by horizontal lines. Note that the spatial separation of the locus into TADs (arrowheads) and sub-TADs (arrows) overlaps with observed regulatory domain structure, e.g., expression in the maxilla (red), limb mesenchyme (yellow), which is confined to the entire *Kcnj* TAD, and additional forebrain expression (blue) restricted to the centromeric *Kcnj* sub-TAD.

5.2.4 Conservation of TAD structure between mouse and human

According to previous studies, TADs appear as genomic structural units that are stable across species and cell types (Dixon et al., 2012; Lupiáñez et al., 2015). Although the genomic size between human and mouse varies, the *SOX9*/*Sox9* locus shows a high level of synteny. To determine the human TAD structure in high resolution, cHi-C from fibroblasts was performed and compared to data obtained from mouse limb buds at E12.5 (Figure 17). The comparison of the cHi-C interaction maps along the locus revealed that the subdivision into the *KCNJ* and *SOX9* TADs was highly conserved. Furthermore, the internal compartmentalization of the two TADs into sub-TADs and position of the TAD boundary regions were captured in a similar way. As shown previously, the CTFC sites associated with the TAD boundary between the *Kcnj* and *Sox9* TADs was highly conserved between vertebrates (Figure 11 in Section 5.1). The conservation of the TAD structure suggests that genes at the locus undergo a similar regulation, making the mouse an appropriate model to study structural variations observed in humans.

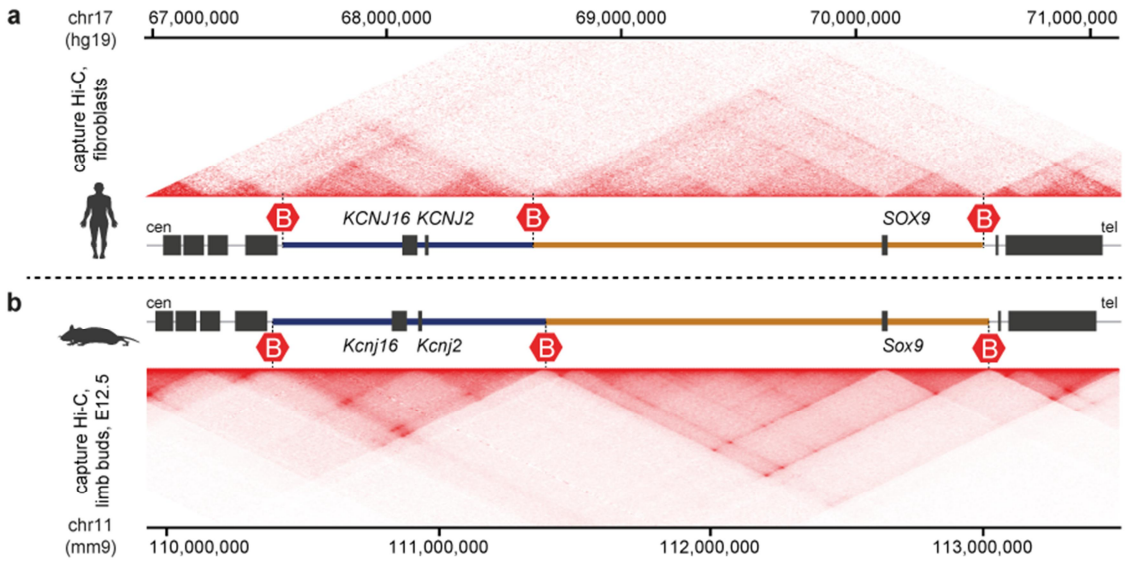


Figure 17: TAD structure at the *Sox9* locus is conserved between humans and mice. Comparison of cHi-C heat maps at the *SOX9*/*Sox9* locus from (a) human fibroblasts and (b) E12.5 limb buds from wild-type mice. Schematic representations depict the *KCNJ*/*Kcnj* TADs (blue) and *SOX9*/*Sox9* (brown) TADs as well as the TAD boundaries (red hexagons). The synteny and TAD structure at the locus are highly conserved between humans and mice. Note that genomes are displayed at different scales to ensure there is fitting the TAD structure for the sake of comparison.

5.3 Position of duplications relative to TAD structure determines their pathogenic effect

Tandem duplications that map to the gene desert upstream of *SOX9* can lead to diverse phenotypic outcomes in humans (Figure 18a), just as discussed earlier on. Large duplications covering almost the entire gene desert upstream of *SOX9* and the neighboring *KCNJ2* and *KCNJ16* genes cause Cooks syndrome, whereas smaller duplications upstream of *SOX9* are associated with female-to-male sex reversal (Benko et al., 2011; Kurth et al., 2009). It is worth noting that by searching for naturally occurring copy number variants using the Decipher database (http://decipher.sanger.ac.uk, Wellcome Trust, Sanger Centre), a third duplication type was detected. This variant was of intermediate size and duplicated the whole gene desert upstream of *SOX9* as in Cooks syndrome-associated duplications, but excluding *KCNJ2* and *KCNJ16*. However, the identified duplication carrier, a mother and daughter, displayed neither a skeletal phenotype nor abnormalities of sexual development. The breakpoint of the latter duplication was confirmed by Sanger sequencing and Table 13 summarizes the position and size of all human duplications.

When comparing the duplication positions to the TAD structure, they can be further classified into two duplication types (Figure 18a). Intra-TAD duplications encompass sequences within the *SOX9* TAD as seen in sex-reversal-associated duplications. In contrast, inter-TAD duplications cover genomic sequences of the *KCNJ* and *SOX9* TADs and their intersecting TAD boundary, as observed for the larger duplications associated with Cooks syndrome or normal phenotype.

Human duplications were re-engineered in mice to investigate the effect of these two duplication types on TAD structure and gene regulation in more detail (Figure 18b). CRISPR/Cas9 genome editing was applied to generate the smaller 400 kb intra-TAD duplication (Dup-S), equivalent to the duplication associated with sex reversal in humans. Two guide RNAs targeting the duplication breakpoints were employed as per the previously described methodology in Section 4.5.1.3 according to Kraft et al., (2015). The larger inter-TAD duplications, equivalent to the Cooks syndrome and normal phenotype, were produced by in vivo recombination using the Cre/loxP system as discussed in Section 4.7.2. A detailed list of the generated mouse mutant alleles from this study can be found in Table 12. For an overview of the generated alleles, a schematic representation is located in Figure 36 within the Appendix.

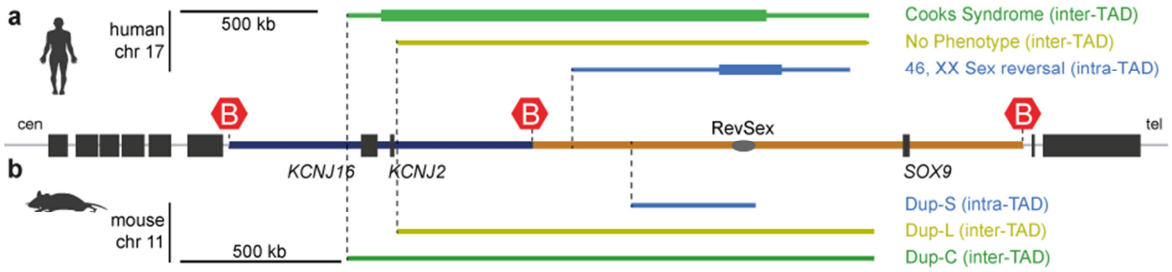


Figure 18: Classification of genomic duplications in inter- and intra-TAD duplications. a, Schematic depicting the human *SOX9* locus, *Kcnj* TAD (blue), *Sox9* (brown) TAD and TAD boundaries (red hexagons). Oval indicates region associated with sex reversal (RevSex) in humans. Duplications associated with three different phenotypes in humans, i.e., Cooks syndrome (green), normal phenotype (yellow) or sex reversal (blue) are indicated above (line thickness reflects maximum/minimum size). Although all duplications overlap to a large extent, including the RevSex region, they are linked to different phenotypes. According to the TAD structure, the duplications can be categorized as duplications that duplicate sequences within the *SOX9* TAD (intra-TAD duplication; XX sex reversal) only and those that additionally duplicate sequences from the *KCNJ* TAD and the TAD boundary (inter-TAD duplication; Cooks syndrome and normal phenotype). b, Re-engineered human duplications in mice, i.e., intra-TAD duplication, Dup-S (equivalent to sex reversal, blue) and two inter-TAD duplications, Dup-L (equivalent to normal phenotype, yellow) and Dup-C (equivalent to Cooks syndrome, green). Note that the position and size of the generated duplications are shown relative to the human genome.

5.3.1 The effect of intra- and inter-TAD duplications on TAD structure

In contrast to humans, heterozygous Dup-S mice with the intra-TAD duplication were phenotypically indifferent, fertile and did not display abnormal sexual development. Heterozygous and homozygous Dup-L mice carrying the large 1.6 Mb inter-TAD duplication were phenotypically normal as human carriers with this duplication (normal phenotype). Both mouse mutants were next employed to study the effect of intra- and inter-TAD duplications on TAD configuration at the *Sox9* locus. To do so, limb bud cells from heterozygous Dup-S and homozygous Dup-L mice at E12.5 were subjected to cHi-C analysis. The resulting interaction matrices were compared to wild-type mice and each interaction matrix was normalized to the total read count and mapped to the standard reference genome (Figure 19).

Both the intra- and inter-TAD duplications featured the wild-type TAD configuration with the *Kcnj* and *Sox9* TADs. Structures, such as loops between TAD boundaries and internal loops, were captured in a highly reproducible manner. However, differences from the wild-type interactions could be detected. Dup-S mice carrying the intra-TAD duplication showed increased chromatin interactions, which appeared as a triangular structure of the duplicated region within the *Sox9* TAD (arrow in Figure 19b). Dup-L mice with the inter-TAD duplication exhibited ectopic interactions of the duplicated sequence between the *Kcnj* and *Sox9* TAD (arrow in Figure 19c).

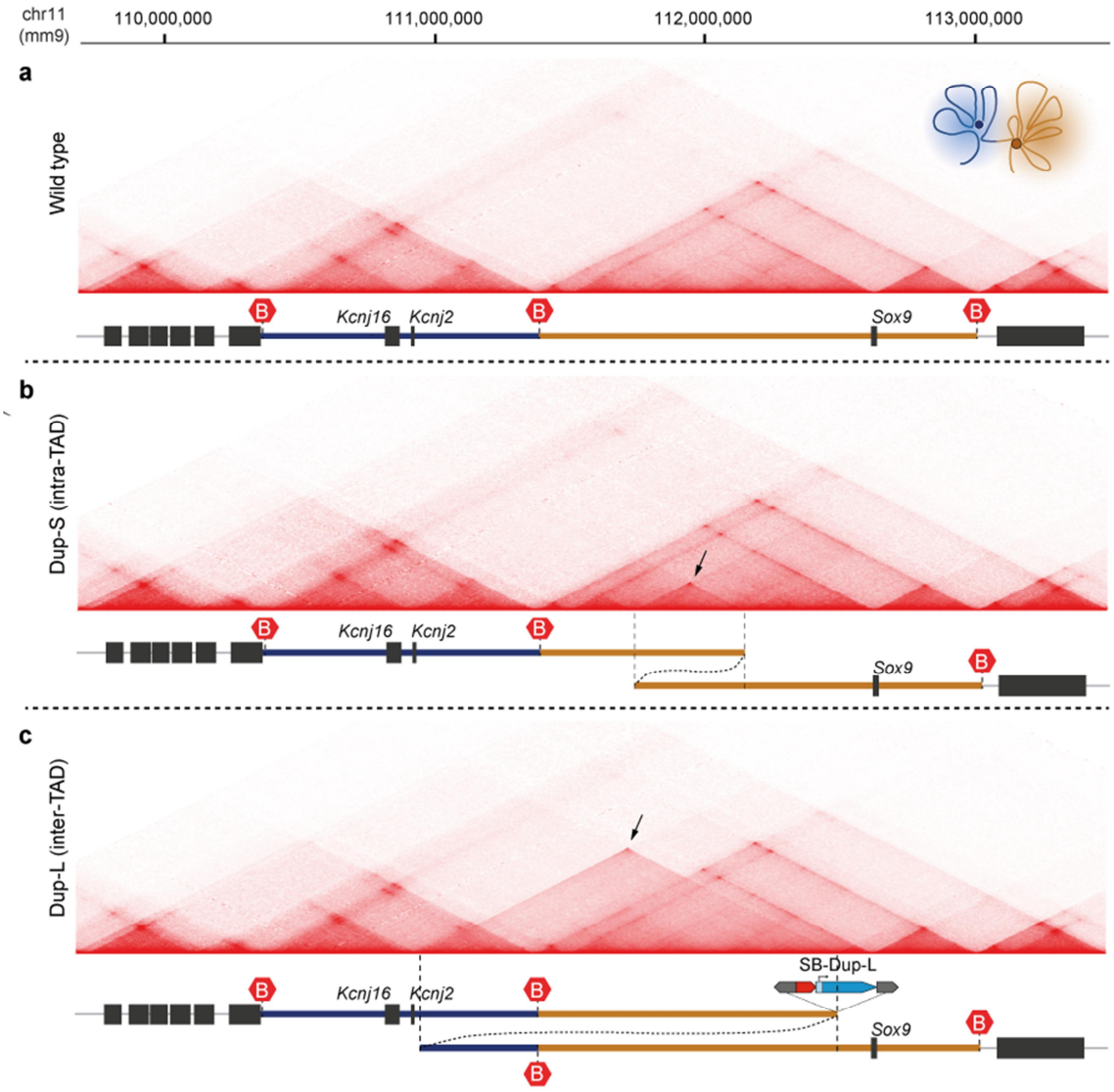


Figure 19: cHi-C from intra- and inter-TAD duplications in mice. cHi-C from E12.5 limb buds of wild-type (a), heterozygous Dup-S (b) and homozygous Dup-L mice (c). Corresponding schematic of the allele is shown below, depicting *Kcnj* TAD (blue), *Sox9* (brown) TAD and TAD boundaries (red hexagons). Duplications are indicated by overlap in schematic representation. Dotted vertical lines mark the duplication breakpoint in the linear genome. Arrow in b indicates increased interaction induced by intra-TAD duplication and arrow in c shows additional interactions in the inter-TAD duplication.

To better visualize the differences between the two duplication types, the wild-type interaction matrix was subtracted from the mutants (Figure 20). To correct for scaling differences during the normalization process, introduced by the additional copy in the mutants, the duplicated sequence was excluded from the mutant and wild-type samples. The results indicated major differences between intra- and inter-TAD duplications with respect to TAD topology at the locus. Interestingly, the subtraction map of the Dup-S mice (intra-TAD duplication) showed that the duplication led to a gain of contacts with sequences of the entire *Sox9* TAD, including the *Sox9* gene itself (asterisks in Figure 20a). This additional signal appeared as a wide stripe of increased interactions, which was restricted to the *Sox9* TAD. The rise in the number interactions with the *Sox9* TAD ended at the TAD boundaries and no differences were discerned beyond them. These

results suggest that the extra copy was implemented into the *Sox9* TAD and was therefore able to interact with *Sox9*. On the contrary, the inter-TAD duplication of Dup-L mice resulted in ectopic chromatin interactions covering just the duplicated portion of the *Kcnj* and *Sox9* TADs (arrows in Figure 20b). Importantly, regions outside the duplicated part did not reflect increased contacts as seen for the intra-TAD duplication in Dup-S mice. This indicated the formation of a separate chromatin domain in which the duplicated sequence was insulated from sequences in the *Kcnj* and *Sox9* TADs. The sequence rearrangement induced by tandem duplication further suggested that this new chromatin domain, referred to as a neo-TAD, was located between the *Kcnj2* and *Sox9* TADs. The schematic representations in Figure 20 summarize the potential effect of intra- and inter-TAD duplications on chromatin folding at the locus.

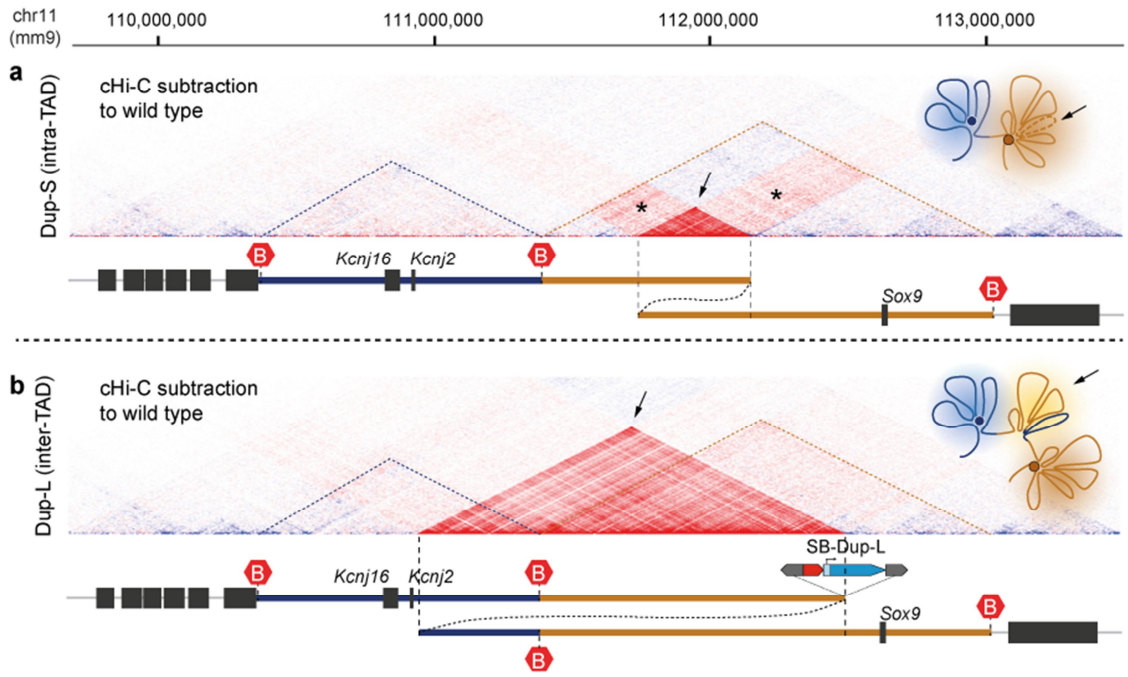


Figure 20: Effect of inter- and intra-TAD duplications on genome architecture. Subtraction of wild-type cHi-C interactions from Dup-S (a) and Dup-L (b) mutants revealed the effect of intra- and inter-TAD duplication on TAD configuration. Corresponding schematic of the alleles is shown below depicting the *Kcnj* TAD (blue), *Sox9* (brown) TAD and TAD boundaries (red hexagons). Duplications are indicated by overlap in schematic representation. Dotted vertical lines mark the duplication breakpoint in the linear genome. The schematic at the top right reflects three-dimensional chromatin folding in the mutants. a, intra-TAD duplication leads to an increase in contact frequency of the duplicated region with itself (arrow) and, importantly, within the *Sox9* TAD (asterisks), including *Sox9*. This suggests that the duplication is incorporated within the *Sox9* TAD (see schematic on top right) b, the inter-TAD duplication demonstrates increased and ectopic interactions only for the duplicated sequence, suggestive of insulation between the duplicated TAD boundary and formation of a new chromatin domain (neo-TAD) between the *Kcnj* and *Sox9* TADs (see schematic on top right).

5.3.2 Inter-TAD duplications result in the formation of a neo-TAD and therefore insulation of the duplicated sequence

To confirm the formation of a new interaction domain and to show selective interactions that appear only in the neo-TAD, allele specific 4C experiments of limb buds from Dup-L mice at E12.5 were performed. The Dup-L allele originated from mice carrying the SB transgene with a *LacZ* reporter at the breakpoints. After *in vivo* recombination, the SB transgene with the *LacZ* reporter remained at the breakpoint of the duplication. Hence, the *LacZ* reporter gene represented a unique sequence in the Dup-L allele, which was located within the neo-TAD and between the duplicated boundaries (Figure 21a). Taking advantage of this unique sequence, 4C-seq experiments from three different viewpoints, i.e., *Kcnj2* in the *Kcnj* TAD, *LacZ* in the neo-TAD and *Sox9* in the *Sox9* TAD, could be designed. 4C-seq profiles from the *Sox9* and *Kcnj2* viewpoints in Dup-L mice exhibited normal interaction patterns. The corresponding log2 ratios to wild type indicated no increased interactions or ectopic signals for the duplicated region (Figure 21a). Furthermore, the interactions within the *Kcnj* and *Sox9* TADs were limited by the pre-existing boundaries from these viewpoints. Therefore, the three-dimensional conformation of the *Sox9* and *Kcnj* TADs surrounding the neo-TAD were not impacted by the duplication. The resulting 4C track from the *LacZ* viewpoint (Figure 21a, purple 4C track) featured an unusual interaction profile with a domain that had frequent interactions being flanked by two broad peaks. This can be understood through the mapping of the 4C-seq data to the wild-type reference genome, which resulted in a split viewpoint from the *LacZ* sequence at the duplication breakpoint. Consequently, the two observed peaks corresponded to high frequency interactions at the viewpoint, appearing as two breakpoints in the linear genome. Yet, the 4C-seq profile from the *LacZ* viewpoint finally exhibited the additional sequences induced by the duplication in Dup-L mice. The *LacZ* sequence interacted exclusively with the duplicated sequence confined by the neo-TAD and appeared as a third interaction domain in the 4C profile. Moreover, the absence of contacts from the *LacZ* viewpoint with sequences outside the duplication, e.g., *Kcnj2* or *Sox9*, confirmed that the neighboring *Kcnj* and *Sox9* TADs were unaffected. The observed domain structure consisting of the *Kcnj*, *Sox9* and neo-TAD, and the lack of cross-interactions between them, further reflected the spatial isolation of the duplicated sequence within the neo-TAD.

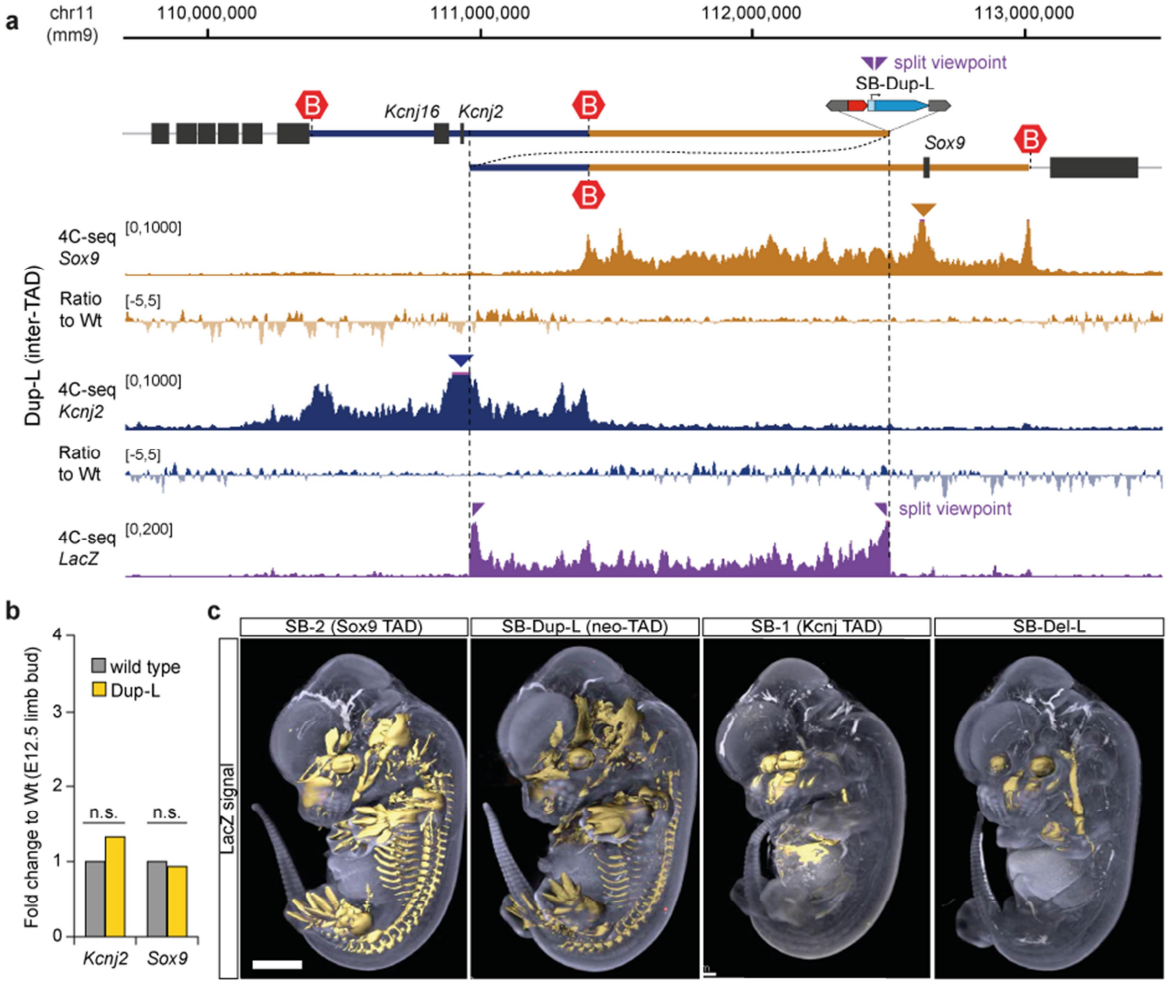


Figure 21: The neo-TAD is spatially and functionally insulated from neighboring TADs. a, Schematic representation of the Dup-L allele depicting the *Kcnj* TAD (blue), *Sox9* (brown) TAD and TAD boundaries (red hexagons). Duplication is indicated by overlap in the schematic representation. Dotted vertical lines mark the duplication breakpoint in the linear genome. Below, the 4C-seq profiles from the *Sox9* and *Kcnj2* viewpoints with corresponding log2 ratio relative to wild type are presented. *Sox9* and *Kcnj2* interaction profiles are unchanged relative to wild type and, therefore, the *Sox9* and *Kcnj* TADs are not affected by inter-TAD duplications. Below this, the unique viewpoint within the neo-TAD, provided by *LacZ,* showing the interactions restricted to the duplicated sequence (purple track). Note that all 4C-seq reads are mapped to the linear genome, which results in a split viewpoint from that of *LacZ* at the duplication breakpoint. b, Expression analysis (expression fold change to wild type) by RNA-seq from E12.5 limb buds showing no changes in expression, confirming insulation of duplicated sequence within the neo-TAD (Benjamini-Hochberg adjusted *P* value, n=2). See Figure 37 in the Appendix for summary of RNA-seq read profiles and expression values. c, OPT analysis of LacZ staining from SB-2 (*Sox9* TAD), Dup-L (neo-TAD), SB-1 (*Kcnj* TAD) and Del-L mice mutants. The *LacZ* insertion within the neo-TAD (SB-Dup-L) recapitulates *Sox9*-like expression as seen for the *LacZ* insertion in SB-2 within the *Sox9* TAD. Note auto fluorescent OPT signals in eyes and blood vessels.

This spatial insulation of the neo-TAD was in agreement with the normal phenotype of heterozygous and homozygous Dup-L mice mutants. RNA-seq from E12.5 limb buds and differential expression analysis to wild type yielded no changes in expression for genes flanking the neo-TAD, i.e., *Kcnj2* and *Sox9* (Figure 21b). To establish whether the duplicated sequence within the neo-TAD was functional and had any regulatory potential with regards to gene expression, *LacZ* expression analysis in Dup-L mice mutants was carried out and compared to the *LacZ* insertions of the SB-1 and SB-2 mouse lines (Figure 21c). To visualize the LACZ signal in detail,

embryos at E12.5 were subjected to optical projection tomography (OPT). As already reported in Section 5.2.2, SB-2 located in the *Sox9* TAD recapitulated most of the endogenous *Sox9* expression pattern, with a particularly strong signal in the cartilage condensation of the developing skeletal system. Contrastingly, SB-1 in the *Kcnj* TAD showed a strong signal in the developing jaw and nose. The LACZ signal from Dup-L mouse mutants was similar to that obtained from SB-2 (*Sox9* TAD), but not SB-1 (*Kcnj* TAD). This suggested that the duplicated sequence within the neo-TAD contained large proportions of *Sox9 cis*-regulatory sequences and that the genomic region was sufficient for driving *LacZ* expression in a *Sox9*-like pattern.

This observation was further supported by a deletion (Del-L) produced by recombination. The deletion corresponded in size and position to the duplication of the Dup-L allele (Figure 36 in the Appendix). LACZ staining in Del-L mice revealed the loss of *Sox9 cis*-regulatory information in the developing skeletal system, whereas the regulatory information of the *Kcnj* TAD, in the lower jaws and nose, was maintained. The loss of *Sox9* CREs in Del-L mice was accompanied by a *Sox9* loss-of-function phenotype that resembled one that was acampomelic. Heterozygous Del-L mice died after birth and displayed hypoplastic scapulae, a missing pair of ribs and a small thoracic cage (data not shown). The loss of *cis*-regulatory information and the lethal phenotype in heterozygous mice highlighted the importance of non-coding sequences in the centromeric part of the *Sox9* TAD (data not shown).

Interestingly, a closer look at the 4C interaction profiles from viewpoints in the *Kcnj*, *Sox9* TAD and the neo-TAD revealed pronounced similarities in three-dimensional folding (Figure 22a). Although the viewpoints from *Sox9*, *Kcnj2* and the *LacZ* promoter differ with respect to their position relative to the duplicated sequence, they contacted the sequence in patterns akin to each other. Figure 22 compares the 4C interaction profiles from the *Sox9* and *Kcnj2* viewpoints in wild-type mice with the profile of the *LacZ* viewpoint in the Dup-L allele mice. Similarities in peak distribution of the duplicated sequence within the neo-TAD and within the *Kcnj* or *Sox9* TADs are indicated by gray dashed boxes. The *LacZ* promoter was evidently able to contact potential CREs in the neo-TAD in the same way as *Sox9* did so in the *Sox9* TAD and *Kcnj2* in the *Kcnj* TAD. This probably allowed the *LacZ* reporter to be expressed in a *Sox9*-like pattern. The duplicated TAD boundary, which flanked the neo-TAD, was contacted in a similar way to the *Kcnj* and *Sox9* TADs (red box), suggesting that the neo-TAD was folded in a structured manner and retained its wild-type interaction characteristics. For better visualization of the neo-TAD, the 4C-seq profile of the *LacZ* viewpoint was rearranged and relocated to an artificially duplicated region (Figure 22b). The linear view illustrates the interactions from the *LacZ* viewpoint inside the neo-TAD and between the duplicated TAD boundaries.

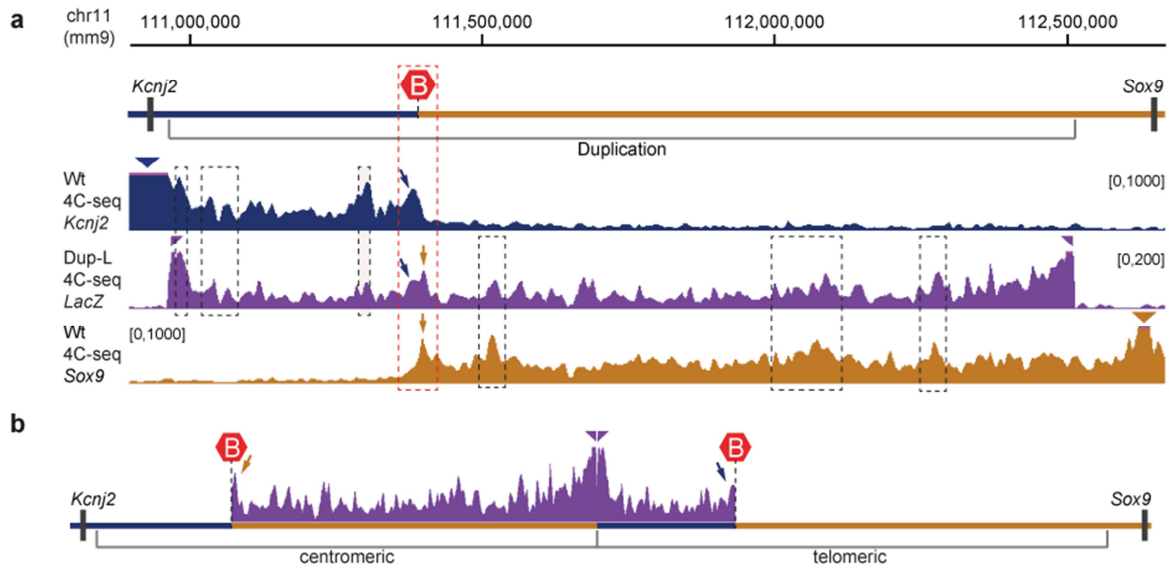


Figure 22: Duplicated sequence within neo-TAD retains its folding properties. a, Schematic shows close-up view of the duplicated region in Dup-L mice depicting the *Kcnj* TAD (blue), *Sox9* (brown) TAD and the boundary (red hexagons) between TADs. Below, there is the comparison of 4C-seq interaction profiles from limb buds at E12.5 with the viewpoint of the *LacZ* promoter (middle, purple) in Dup-L mice and viewpoints in *Kcnj2* (top, blue) and *Sox9* (bottom, brown) from wild type. Dashed boxes indicate similarities in peak distribution and patterns between neo-, *Kcnj* and *Sox9* TADs. The duplicated sequence within the neo-TAD retains folding properties as the copies in the *Kcnj* and *Sox9* TADs (dashed boxes). The duplicated TAD boundary flanking the neo-TAD has a peak profile that merges the peak profiles from the *Kcnj* and *Sox9* TADs (arrows in red dashed box). b, The 4C profile from the *LacZ* viewpoint is plotted on an artificially duplicated region, visualizing the linear order of the duplicated sequence within the neo-TAD. The extent of the tandem duplication is reflected by a centromeric and telomeric part. Note that the 4C profile and viewpoint of the *LacZ* reporter at the breakpoint are in their correct linear order and the duplicated TAD boundaries flank the neo-TAD-specific 4C profile.

Taking these findings together, allele-specific 4C-seq in Dup-L mutants confirmed the formation of a neo-TAD induced by inter-TAD duplications. 4C-seq from the unique *LacZ* viewpoint within the neo-TAD exclusively contacted the duplicated sequence and no sequences outside the duplication. The sequence within the neo-TAD was therefore spatially insulated by the duplicated boundaries of the neo-TAD, thereby preventing interactions with flanking genes and neighboring TADs. The analysis of Dup-L mutants further provided evidence that the duplicated sequence in the neo-TAD folded in a similar pattern as in the neighboring TADs. These particular contacts with CREs of the *LacZ* sequence led to activated *LacZ* expression in a *Sox9*-like expression pattern. Overall, the neo-TAD appeared as a new chromatin structure that was spatially and functionally isolated from its surrounding genomic environment.

5.3.3 Cooks duplications incorporate Kcnj16 and Kcnj2 within the neo-TAD

In humans, duplications associated with normal phenotype or Cooks syndrome belong to the class of inter-TAD duplications that induce the formation of a neo-TAD. However, all Cooks syndrome duplications extend further centromerically towards the next flanking genes, *Kcnj2* and *Kcnj16*. These larger duplications would result in the incorporation of both genes into the neo-TAD and potentially to a misregulation of the gene copies driven by CREs therein. To test this hypothesis, the equivalent Cooks-associated duplication was generated in mice (Dup-C) by trans-allelic recombination (see Figure 36 in the Appendix and Table 12 for detailed genomic positions). In Dup-C mice, the duplication had the same telomeric breakpoint as Dup-L mice but extended 180 kb further centromerically (Figure 23).

To demonstrate the formation of the neo-TAD in Dup-C mouse mutants, cHi-C and 4C-seq from E12.5 limb buds were performed. Subtracting the cHi-C interaction matrix from that of wild-type mice resulted in the appearance of an additional interaction domain corresponding to the duplicated region as was the case for Dup-L mouse mutants (Figure 23a). This confirmed that a neo-TAD was created that included *Kcnj2* and *Kcnj16*. Incorporation of the genes into the neo-TAD was further validated by 4C-seq. The *Kcnj2* viewpoint showed additional ectopic interactions with parts of the *Sox9* regulatory region and the ratio to wild type uncovered the extent of the neo-TAD-specific interactions (Figure 23b). As seen in the Dup-L mice before, no changes in the configuration of the interaction profile was observed from the *Sox9* viewpoint.

The copy of *Kcnj2*, integrated in the neo-TAD, now interacted freely with the duplicated *Sox9* regulatory sequences. The findings demonstrated that by variable shifting the size of inter-TAD duplications, genes can be included within a neo-TAD. Genes that were incorporated in the neo-TAD, i.e., the *LacZ* reporter in Dup-L or *Kcnj16* and *Kcnj2* in Dup-C mice mutants, were placed in a new chromatin environment where they ectopically interacted with new *cis*-regulatory sequences.

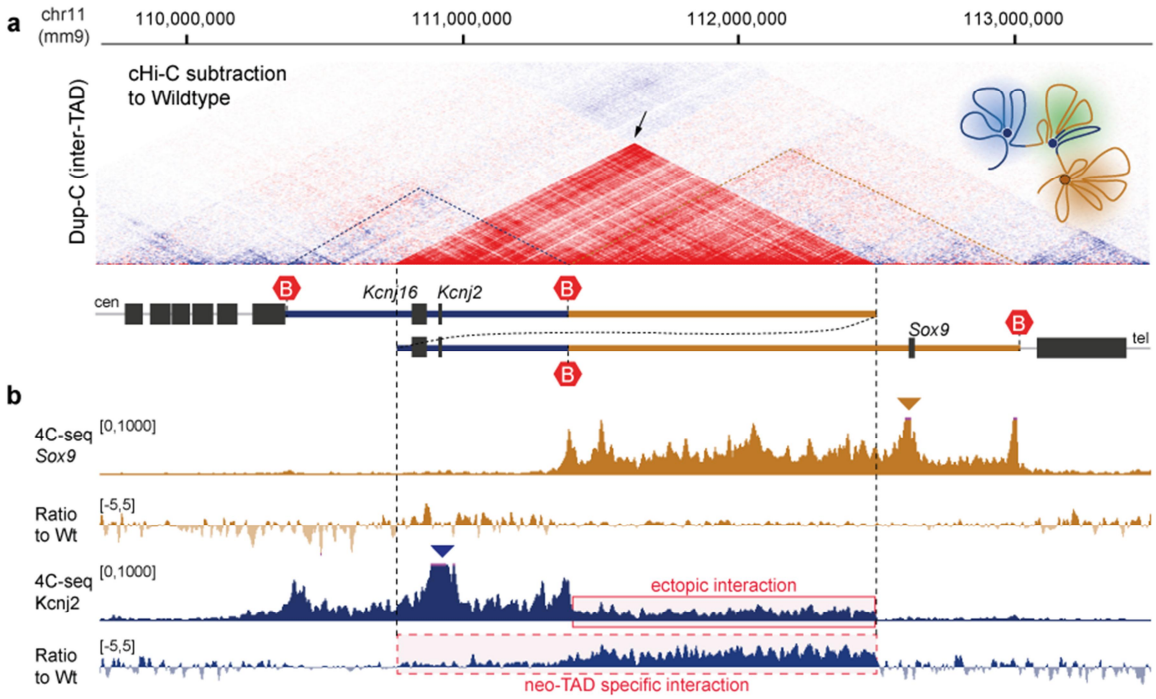


Figure 23: Cooks-like duplication results in incorporation of *Kcnj2* and *Kcnj16* within the neo-TAD. a, cHi-C subtraction of wild-type interactions from Dup-C mice mutant limb buds at E12.5. The schematic below depicts the *Kcnj* TAD (blue), *Sox9* (brown) TAD and TAD boundaries (red hexagons). The duplication is indicated by overlap in the schematic representation. Dotted vertical lines mark the duplication breakpoint in the linear genome. Note that the telomeric breakpoint is the same as in Dup-L mice but the centromeric breakpoint is shifted towards *Kcnj2* and *Kcnj16*. As shown with Dup-L mice, the inter-TAD duplication exhibits increased and ectopic interactions only of the duplicated sequence, suggesting formation of a neo-TAD but including *Kcnj2* and *Kcnj16*. The schematic at the top right indicates three-dimensional chromatin folding in Dup-C mutants. b, 4C-seq profiles from *Sox9* (brown) and *Kcnj2* (blue) viewpoints with the corresponding log2 ratio to wild type from E12.5 limb buds. The *Sox9* interaction profile is unchanged compared to wild type. The 4C-seq profile from *Kcnj2* shows ectopic contacts with the duplicated part of the *Sox9* TAD. The ratio to wild type indicates neo-TAD-specific interactions of the duplicated *Kcnj2* gene within the neo-TAD, confirming the results from cHi-C.

5.3.4 Misregulation of Kcnj2 results in Cooks syndrome

To evaluate the impact of the ectopic contacts of *Kcnj16* and *Kcnj2* on gene expression observed in heterozygous Dup-C mutant mice, RNA-seq from E12.5 limb buds was performed (Figure 24a). The differential expression analysis revealed a three-fold upregulation of *Kcnj2*, while other genes remained at the same levels (Figure 37 in the Appendix). Whole-mount *in situ* hybridization at E12.5 demonstrated that the upregulation of *Kcnj2* was caused by misexpression in the cartilaginous anlagen of the digits, resembling the expression pattern of *Sox9* (Figure 24b). The observed upregulation and misexpression of *Kcnj2* was maintained during development. RNA expression analysis by RNA-seq in the digit tips at E17.5 verified an altered expression of *Kcnj2* (see Figure 37 in the Appendix). Thus, the incorporation of one copy of *Kcnj2* into the neo-TAD led to its activation by *Sox9* regulatory elements. The *LacZ* reporter in the neo-TAD of Dup-L mice resembled most of the complex *Sox9* expression pattern. However, whole-mount *in situ* hybridization in Dup-C exhibited only misexpression of *Kcnj2* in the digit anlagen. This indicated

that the promoters of the *LacZ* reporter and the *Kcnj2* gene responded differently to the regulatory information within the neo-TAD. These differences in promoter activation were also observed for *Kcnj16*, which was also integrated into the neo-TAD in Dup-C mice. In contrast to *Kcnj2*, *Kcnj16* was not responsive to any regulatory activity.

Misexpression of *Kcnj2* in Dup-C mouse mutants resulted in a phenotype resembling Cooks syndrome in humans. At birth, heterozygous and homozygous Dup-C mouse mutants displayed an abnormal phenotype with affected distal phalanges of the fore- and hind limbs. Severe hypoplasia of all claws was seen (Figure 24c). Furthermore, high resolution micro-computed tomography (µCT) of adult mice mutants showed an abnormal shape and size of the distal and middle phalanges. The distal phalanges were smaller whereas the middle phalanges were elongated compared to the wild-type animals. Additionally, both distal and middle phalanges were fixed in a straight position. In contrast, the dorsal angle between the distal and middle phalange was almost 90 degrees in wild-type mice. Masson-Goldner staining of sections of new born mice suggested that the interphalangeal joint between the distal and middle phalanx was malformed, underdeveloped and partly fused, which could have led to the fixation of the distal phalanges in a straight position during development. Apart from the absence of claws and the abnormal development of the distal phalanges, the mice had no other abnormalities and behaved normally.

Additionally, another inter-TAD duplication (Dup-C2) with a different centromeric breakpoint located between *Kcnj16* and *Kcnj2* was generated. *Kcnj16* was thus excluded from the duplication and the neo-TAD (Figure 36 in the Appendix). These mice exhibited an identical phenotype to the Dup-C mice mutants, further supporting the notion that *Kcnj2* misexpression alone causes the phenotype. Additional intra-TAD duplications were generated by recombination to exclude the possibility that the Cooks-like phenotype was simply produced by the elevated copy number of the *Kcnj2* gene itself or by regulatory sequences of the *Kcnj2* TAD (Figure 36 in the Appendix). For example, the Dup-K1 allele duplicated the *Kcnj* genes and their telomeric regulatory region, whereas Dup-K2 duplicated exclusively the *Kcnj* genes and Dup-K3 exclusively the telomeric regulatory region within the *Kcnj* TAD. All three mutant mice lines had a normal phenotype. In summary, the combination of genes and regulatory sequences contained within the neo-TAD determines duplication pathogenesis. Inter-TAD duplications that are responsible for positioning of the *Kcnj2* and *Sox9 cis*-regulatory sequences within the neo-TAD induce a Cooks-like phenotype in mice.

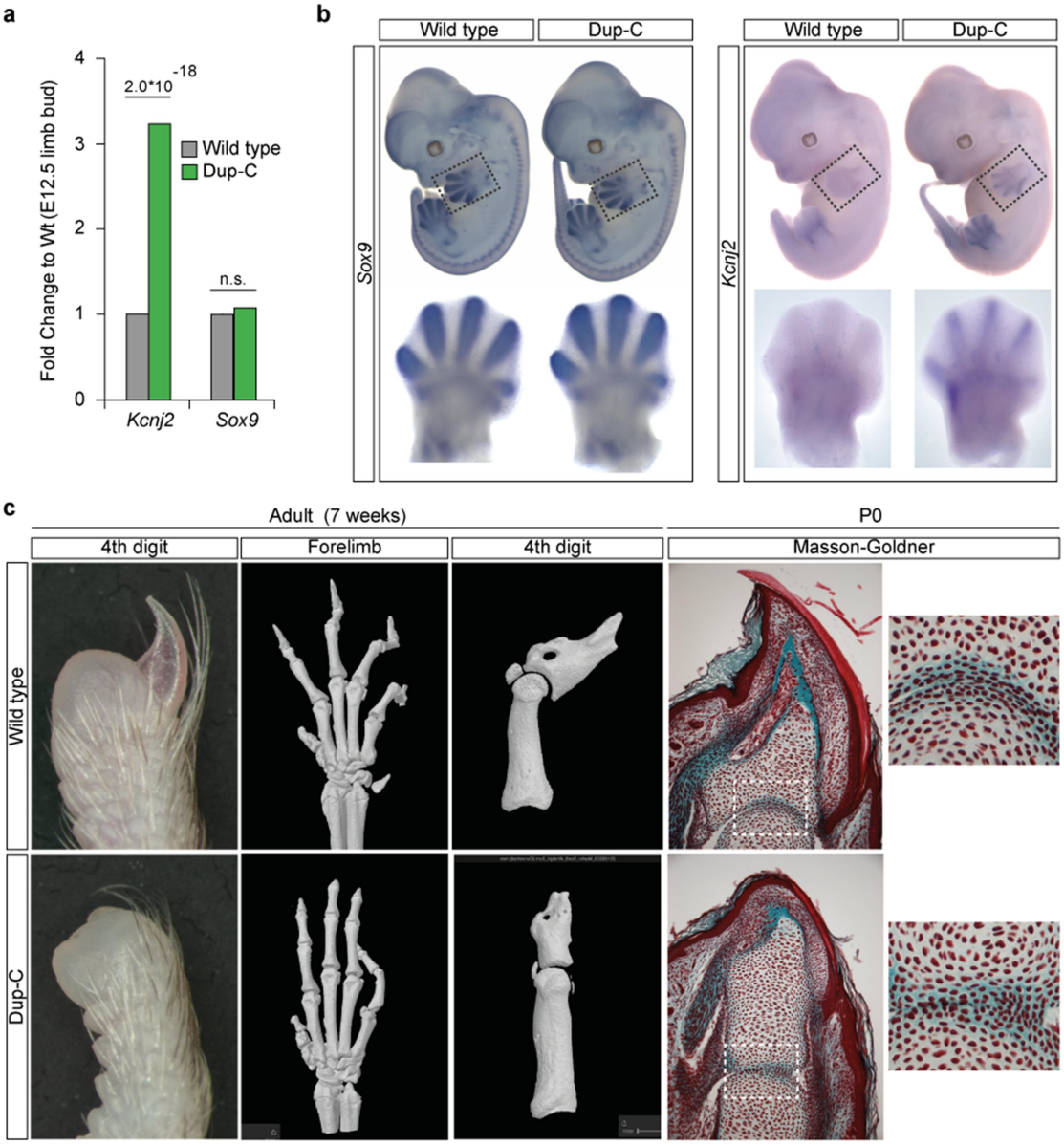


Figure 24: Neo-TAD drives misregulation of *Kcnj2*, which results in a Cooks-like phenotype in Dup-C mice. a, RNA-seq from E12.5 limb buds in wild-type and heterozygous Dup-C mice mutants; expression fold change to wild type shows upregulation of *Kcnj2* (Benjamini-Hochberg adjusted *P* value, n=2). See Figure 37 in the Appendix for summary of RNA-seq read profiles and expression values. b, Whole-mount *in situ* hybridization for *Sox9* and *Kcnj2* in E12.5 wild-type and heterozygous Dup-C embryos. Note the expression of *Sox9* in the digit anlagen and expression of *Kcnj2* in the limb mesenchyme (enlargement) in wild-type embryos. Dup-C embryos misexpress *Kcnj2* in the digits with a *Sox9*-like pattern. c, Phenotypic analysis of wild-type (upper panel) and heterozygous Dup-C mutants (lower panel). Note that the (left-to-right) side view of 4th digit shows the absence of claws; µCT analysis of autopod, middle and distal phalanges of 4th digit showing the abnormal position and shape of the phalanges; Masson-Goldner staining indicates hypoplasia of the nail bed and abnormal shape of distal phalanx; enlargement (of white dashed boxes) shows joint fusion in mutants.

5.4 TAD boundaries insulate TADs and the neo-TAD

The studied inter-TAD duplications herein induced the formation of a new TAD located between the duplicated TAD boundary. This raised the question of whether the TAD boundary was required for the separation of TADs at the *Sox9* locus and the formation and spatial insulation of the neo-TAD depended on it. As shown by the 4C-seq experiments, the boundary between the *Kcnj* and *Sox9* TADs showed a sharp decline in interactions from the *Sox9* as well as from the *Kcnj2* viewpoint in a 18 kb region (described previously in Figure 11). Within this region, three conserved CTCF sites were present that had opposite directionality. To assess the role of this boundary region (Bor), CRISPR/Cas9 genome editing was applied to delete the 18 kb region in wild-type G4 ESCs (dashed grey box in Figure 11). An ESC clone with a homozygous deletion was identified and used for diploid and tetraploid complementation to generate mouse mutants (ΔBor). Homozygous mouse mutants had a normal phenotype.

To investigate the effect of the boundary deletion on TAD architecture, limb buds at E12.5 from homozygous mice were subjected to cHi-C experiments. The resulting cHi-C interaction map revealed that the overall subdivision and structure of the *Kcnj* and *Sox9* TADs were largely preserved upon boundary deletion. The sub-TAD structure within each TAD and the chromatin loops between other CTCF sites were not affected (arrowheads in Figure 25a andFigure 25b). This suggested that the sub-TAD structure and inter-TAD chromatin loops contributed to the formation and stability of the TADs. However, the deletion of the CTCF cluster at the TAD boundary induced loss of specific chromatin loops with the telomeric boundary of the *Sox9* TAD and the centromeric boundary of the *Kcnj* TAD. This loss of chromatin loops was accompanied by the formation of new interactions absent in the wild type. Novel interactions were established between regions associated with CTCF binding and CTCF motifs in a convergent orientation (see Figure 10 for CTCF motif orientation). For instance, the remaining TAD boundaries or the centromeric boundary of the *Kcnj* TAD and a CTCF-bound region close to *Sox9* were engaged in new contacts (arrows in Figure 25b). Thus, the loss of the CTCF-interaction partners at the boundary created, at the same time, new interactions between partners across the two TADs. These novel contacts across TADs were not just restricted to chromatin loops. Moreover, the subtraction of the wild-type interaction matrix demonstrated that ectopic contacts between the *Kcnj* and *Sox9* TADs were established over the whole extent of both TADs (dotted area marked by asterisk in Figure 25c). Interestingly, these ectopic contacts were limited by the remaining TAD boundaries of the *Kcnj* and *Sox9* TAD, affecting only the two TADs. The deletion in the wild type suggested that the TAD boundary was not necessary for the formation of the TADs, but clearly showed that it was required for insulation and complete separation of the genomic sequences within each TAD.

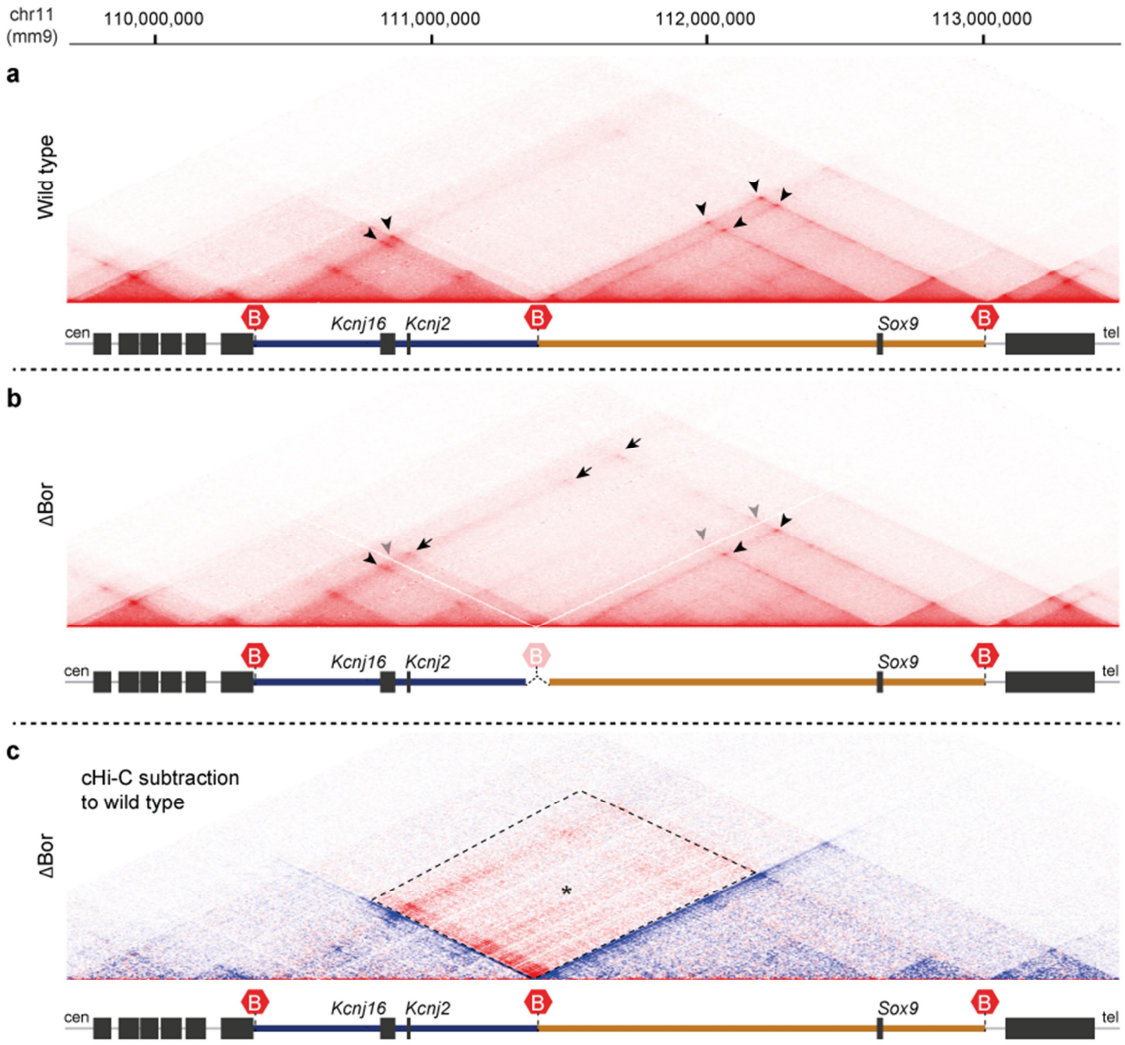


Figure 25: The boundary separates TADs from each other. CHi-C of E12.5 limb buds from (a) wild-type and (b) homozygous ΔBor mutants and (c) subtraction map of ΔBor relative to wild type. The schematic below each interaction matrix shows the locus of the *Kcnj* TAD (blue), *Sox9* TAD (brown) and TAD boundaries. Deletion of the TAD boundary between the *Kcnj* and *Sox9* TADs leads to loss of chromatin loops (compare arrowheads in a and grey arrowheads in b) and to ectopic contacts (arrows in b and dotted area, marked by * in c) and, therefore, loss of insulation between TADs. Note that the ectopic contacts include *Kcnj2* and *Sox9* and that they are restricted by the remaining TAD boundaries. However, the overall TAD structure, especially the sub-TAD structure with associated chromatin loops (black arrowheads in a and b) remained the same.

To disrupt the formation of the neo-TAD and to demonstrate that the duplicated sequence within the neo-TAD was indeed insulated, the two copies flanking the neo-TAD in the Dup-L allele were deleted. To do so, ESCs from homozygous Dup-L mouse mutants were established and treated with CRISPR/Cas9 guide RNAs directed to the 18 kb boundary region. An ESC clone (Dup-LΔBor), homozygous for the duplication and with all four boundary sites deleted, was produced and utilized for ESC aggregation. The ESC clone failed to produce live animals with embryos being reabsorbed in the uterus at the E13.5 stage. Blastocyst injections with the same clone failed as well, suggesting that early embryonic fatality was probably because of ESC quality or the genetic background of the ESC line. The established cell line was recovered from mice which were

originally generated from E14 ESCs and additionally back-crossed to *C57BL/6* for four generations. The mixed background could potentially interfere with the performance of the ESCs during later stages of embryonic development.

However, limb buds of Dup-LΔBor embryos at E12.5 from tetraploid aggregations could be harvested and processed for cHi-C and RNA-seq analysis. The obtained cHi-C data confirmed the results of the boundary deletion in the wild type (ΔBor). The TAD structure in the Dup-LΔBor allele, consisting of the neo-, *Kcnj* and *Sox9* TADs, was still visible upon boundary deletion. (Figure 26a and Figure 26b). Most importantly, the Dup-LΔBor allele highlighted the spatial insulation of the duplicated sequence within the neo-TAD. The subtraction map of Dup-LΔBor relative to Dup-L exhibited greater interaction frequencies of the duplicated sequence with itself (** in Figure 26c), indicative of the sequence within the neo-TAD no longer being insulated. Furthermore, the boundary deletion resulted in ectopic interactions with sequences outside the duplication (* in Figure 26c). The ectopic contacts seen resembled the results from the ΔBor deletion in the wild type, where new contacts between TADs were established.

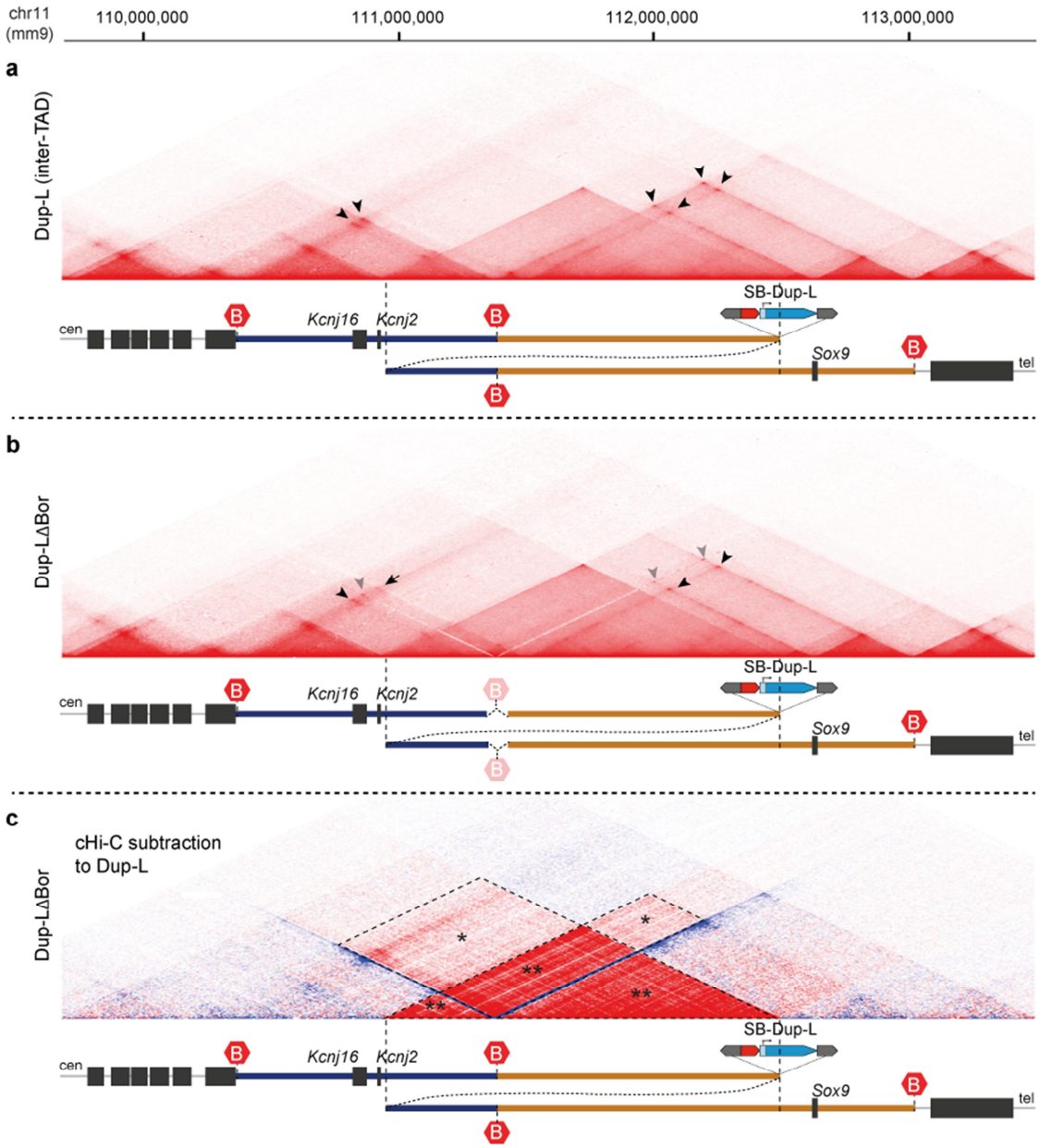


Figure 26: The neo-TAD is insulated by TAD boundaries as regular TADs. CHi-C of E12.5 limb buds from (a) homozygous Dup-L and (b) Dup-L ΔBor mutants and the (c) subtraction map of Dup-L ΔBor relative to Dup-L. The corresponding schematic of the alleles is depicted below featuring the *Kcnj* TAD (blue), *Sox9* (brown) TAD and TAD boundaries (red hexagons). Duplications are indicated by overlapping. Deletion of the duplicated TAD boundary flanking the neo-TAD results in loss of chromatin loops (compare arrowheads in a and grey arrowheads in b), to novel chromatin loops (arrow in b) and to ectopic interactions between TADs (* in c). The boundary deletion in the Dup-L allele further demonstrates that the duplicated sequence within the neo-TAD is no longer insulated and forms more contacts with their copy in the neighboring TADs (** in c). Note that the overall TAD structure remains the same as seen in ΔBor mutants.

The loss of the TAD boundary in the ΔBor and Dup-LΔBor allele resulted in ectopic contacts between TADs. To assess the effect on gene expression in both mutants, RNA-seq from limb buds at E12.5 was performed. Differential expression analysis showed that *Kcnj2* was upregulated by two-and-a-half-fold in homozygous mutants, whereas other genes, i.e., *Sox9*, remained unchanged (Figure 27a). A similar fold change of *Kcnj2* was observed in Dup-C mice. However, this upregulation did not result in any site-specific misexpression of *Kcnj2* as seen with *in situ* hybridization (Figure 27b). The obtained signal was diffused without a specific pattern in the digits, which could probably explain the absence of a Cooks-like phenotype in ΔBor mice.

The deletion experiments in wild-type and Dup-L allele highlighted the role of TAD boundaries and other CTCF sites as a structural component of TADs. The boundary deletion did not interfere with the general formation of TADs, suggesting that other factors contributed to the spatial organization of the locus. Interactions between other CTCF sites that formed internal loops and the sub-TAD structure within the *Kcnj* or *Sox9* TAD were not altered upon boundary deletion. Their preservation likely contributed to TAD formation and stability. All the same, the findings with both alleles (ΔBor and Dup-LΔBor) showed that the boundary was required for insulation and complete separation of TADs. Most prominently, the results of the Dup-LΔBor allele demonstrated that the neo-TAD was spatially separated and behaved like 'regular' TADs. Therefore, the duplication of a boundary by inter-TAD duplication was necessary to fully isolate the *cis*-regulatory information and gene content within the neo-TAD from neighboring TADs. The lack of insulation by the boundary deletion resulted in ectopic contacts between TADs and modified gene expression. However, *Kcnj2* exhibited no site-specific misexpression as it was induced by the neo-TAD in the Dup-C mutants. The neo-TAD provided a spatial environment akin to 'regular' TADs where genes and CREs interact and orchestrate a specific spatio-temporal expression pattern. The overall unchanged TAD structure upon boundary deletion probably preserved the wild-type interactions within the TADs and the observed ectopic contacts were not sufficient to drive a site- and stage-specific misexpression of *Kcnj2*.

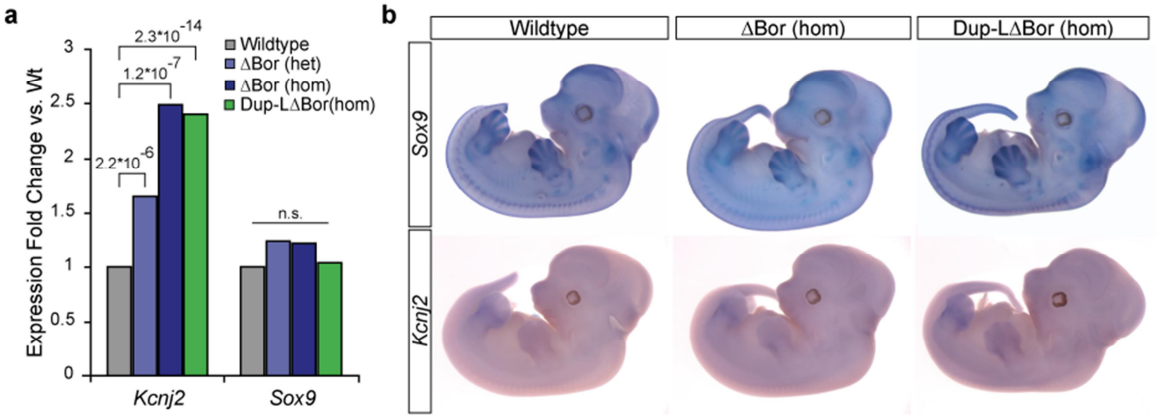


Figure 27: Loss of TAD boundary leads to altered gene expression. a, RNA-seq from E12.5 limb buds in wild-type, heterozygous and homozygous ΔBor and homozygous Dup-L ΔBor mutants. Expression fold change to wild type shows upregulation of *Kcnj2* in all mutants with a TAD boundary deletion (ΔBor) (Benjamini-Hochberg adjusted *P* value, n=2). See Figure 37 in the Appendix for summary of RNA-seq read profiles and expression values. b, Whole-mount *in situ* hybridization of *Sox9* and *Kcnj2* in E12.5 wild-type, homozygous ΔBor and Dup-L ΔBor embryos. Upregulation of *Kcnj2* did not result in site-specific misregulation as seen in the digits for Dup-C mouse mutants.

5.5 Analysis of intra- and inter-TAD duplications in patient cells

The genetic experiments performed in mice in this work demonstrated the differential effects of intra- and inter-TAD duplications on TAD structure and gene regulation. Next, patient-derived cells were employed to confirm these observations in humans and assess whether patient-derived material was suitable for the diagnosis and interpretation of intra- and inter-TAD duplication-induced changes in TAD configuration.

4C-seq was applied to patient-derived skin fibroblasts and the results compared to that for control fibroblasts (Figure 28). 4C-seq in control cells with viewpoints in the promoters of *KCNJ2* and *SOX9* recapitulated the cHi-C-predicted TAD structure of the *KCNJ* and *SOX9* TADs (Figure 28a). Cells from a patient with sex reversal and a 150 kb intra-TAD duplication showed no changes in the overall interaction profiles from both viewpoints (Figure 28b). The log2 ratio to control cells revealed a slight increase of signal of the *SOX9* promoter with the duplicated region containing the RevSex region (arrow in Figure 28b). However, the increased interaction frequency in the log2 ratio was very weak and poorly distinguished from the background signal between patient and control samples, rendering it difficult to identify the duplicated region. To overcome this and visualize all potential interaction partners of the duplicated sequence, an allele-specific 4C-seq approach was established. The sequence at the duplication breakpoint created a unique restriction fragment in the genome that was used to design a viewpoint for 4C-seq. The interaction profile from the breakpoint viewpoint showed that the duplication interacted with all sequences within the *SOX9* TAD and that the interactions were constrained by the TAD boundaries. These results validated that the extra copy induced by inter-TAD duplications was able to freely interact with other genomic regions within the *SOX9* TAD region (Figure 28b).

The same 4C-seq experiments were performed from a fibroblast culture of a patient with an inter-TAD duplication displaying no phenotypic changes (Figure 28c). The *KCNJ2* and *SOX9* viewpoints, which were located outside the duplication, exhibited no changes in interaction profiles in the respective *KCNJ* and *SOX9* TADs. The breakpoint viewpoint showed an additional interaction domain restricted to the duplicated sequence, recapitulating the results from the Dup-L mouse allele. The inter-TAD duplication led to the formation of a new chromatin domain (neo-TAD) that isolated the duplicated sequence from genes in the neighboring TADs and, consequently, did not elicit changes in patients carrying it.

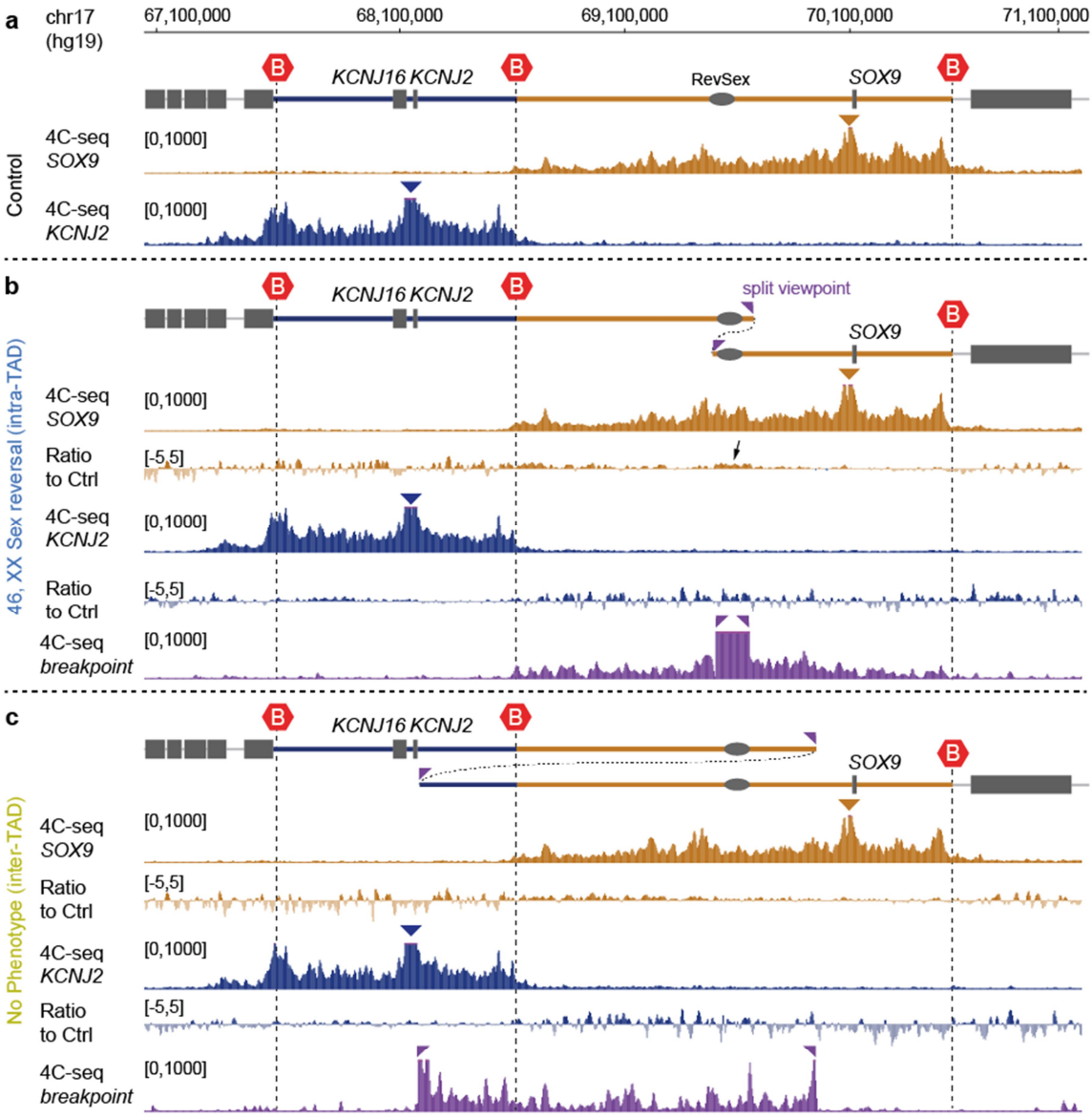


Figure 28: Analysis of intra-and inter-TAD duplications using 4C-seq in patient-derived skin fibroblasts. a, 4C-seq analysis from control skin fibroblasts with viewpoints in *KCNJ2* (blue) and *SOX9* (brown). 4C-seq interaction profiles recapitulate cHi-C-predicted TAD structure at the locus. b, 4C-seq from fibroblasts from a patient with a 150 kb intra-TAD duplication and sex reversal. The overall 4C-seq interaction profiles are unchanged. However, a slight increase from the *SOX9* viewpoint with the duplicated region was detected (arrow). 4C-seq from the duplication breakpoint (purple track) uncover the effect of the intra-TAD duplication. The duplicated region interacts with all regions within the *SOX9* TAD. c, 4C-seq from fibroblasts of a patient with an inter-TAD duplication and no phenotypic changes. The duplicated region was not detected from the viewpoints (*KCNJ2* and *SOX9*) outside the duplication interval. The breakpoint viewpoint (purple track), however, has an interaction profile that is restricted to the duplication, confirming the formation of a neo-TAD in human cells.

Naturally occurring tandem duplications can arise from unequal cross-over events of homolog chromosomes during meiosis, which introduces allele differences in the two copies of the duplication (Ohno, 1970; Zhang et al., 2003). This raised the possibility of leveraging allele differences, such as SNPs, to distinguish between the two copies to further investigate allele-specific interactions inside and outside the neo-TAD. 4C-seq data from patients with the inter-TAD duplication (normal phenotype) were used to determine selective interactions from the *KNCJ2*,

SOX9 and breakpoint viewpoints within the duplication interval (Figure 29). To do so, SNPs from a single 4C-seq experiment were identified and compared to SNP positions of all three 4C-seq experiments. Only SNPs that were homozygous for one variant from the breakpoint viewpoint and homozygous for another variant contacted by either the *KCNJ2* or *SOX9* viewpoints were considered. Several SNPs were determined that interacted exclusively with the breakpoint viewpoint but not with *KCNJ2* or *SOX9*. Additionally, WGS of the patient genome was conducted to verify the detected allele variants. The selected SNP positions from 4C-seq experiments were present with an average allele frequency of 2 to 1 (major/minor allele frequency), suggesting the existence of three alleles. The allele variants that interacted with the breakpoint viewpoint represented the minor allele frequency, confirming that the duplicated allele within the neo-TAD was present only once in the genome. On the other hand, the allele variants contacted by the *KCNJ2* or *SOX9* viewpoint were represented twice in the genome (homozygous SNP). A summary of SNP genotypes and allele frequencies is given in Figure 38 in the Appendix. The observed allele-specific interactions confirmed that the two copies of the duplicated sequences were physically separated from each other. The *KCNJ* and *SOX9* TADs retained their normal interactions and exhibited no cross-interactions with the sequences within the neo-TAD.

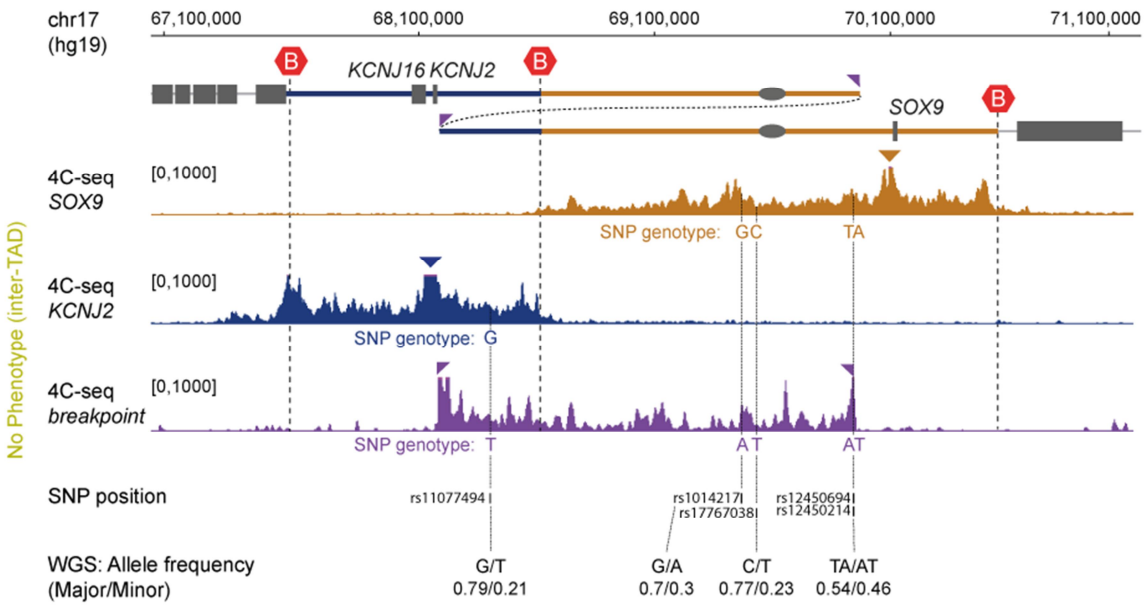


Figure 29: SNP analysis with 4C-seq data from patient cells confirmed spatial insulation of the neo-TAD. 4C-seq data from fibroblasts of a patient with an inter-TAD duplication and no phenotypic changes. Interaction profiles from *SOX9* (brown), *KCNJ2* (blue) and duplication breakpoint viewpoints (purple) are presented. SNP genotypes in the duplication interval that are contacted by the respective viewpoint are indicated below each track. Only SNP positions that are homozygous for one variant from the breakpoint viewpoint and homozygous for another variant from the *KCNJ2* or *SOX9* viewpoints are shown. The allele frequency from WGS confirms the observed SNP genotypes. Note that SNPs contacted by the duplication breakpoint are not contacted by the *KCNJ2* and *SOX9* viewpoints, suggesting the spatial insulation of the neo-TAD. See also Figure 38 in the Appendix for a summary of SNP genotypes and allele frequencies.

The 4C-seq analysis of patient-derived fibroblasts reproduced the results obtained from mice experiments. It also demonstrated that patient cells can be used to directly identify aberrant changes in chromatin architecture. However, applying 4C-seq with viewpoints in *KCNJ2* and *SOX9* only was insufficient to discern the effect of intra- and inter-TAD duplications on TAD structure. The increased interaction frequency of the extra copy induced by the intra-TAD duplication, for example, was barely detected from the *SOX9* viewpoint. The design of an allele-specific viewpoint for 4C-seq at the duplication breakpoint provided a way to uncover the effect of the two duplication types. Hence, all interaction partners could be identified. While intra-TAD duplications showed interactions limited to their respective TAD, the interactions of the inter-TAD duplication were restricted to the neo-TAD. Additionally, SNP analysis assisted in determining allele-specific interactions within the neo-TAD, further supporting the observation of a new spatially isolated chromatin compartment within the genome.

5.5.1 4C-seq analysis of different patient-derived cell types

The comparison of 4C-seq interaction profiles in different mouse tissues revealed that interactions within the *Sox9* TAD, which are associated with CTCF binding, are cell-type invariant and independent of *Sox9* expression. This observation raised the possibility of detecting the TAD structure and changes in TAD configuration induced by duplications in any patient-derived cell. Common sources of actively proliferating patient-derived cells are fibroblast cultures from skin biopsies or lymphoblastoid cell lines isolated from blood samples. However, skin fibroblast cultures are not always available and obtaining skin biopsies requires a minimally invasive procedure for patients. Furthermore, to generate lymphoblastoid cell lines, which are established by infecting peripheral blood lymphocytes with Epstein-Barr virus (EBV), requires special laboratory safety levels (Neitzel, 1986). An alternative source of patient cells are white blood cells (leukocytes) that are directly isolated from total blood samples of patients. These isolates possess diverse differentiated and nucleated cell types, e.g., 60 % neutrophils, 30 % lymphocytes and 5 % monocytes (Wheater, Burkitt, & Daniels 1979).

Skin fibroblasts, lymphoblastoid cells and white blood cells from healthy individuals were used to test their suitability for detecting chromatin interactions confined to TAD structure at the *SOX9* locus. First, the expression of *KCNJ2* and *SOX9* was assessed using the human genotype-tissue expression (GTEx) database, which provided expression data for a large sample set for the examined cell types (GTEx Consortium, 2015). *SOX9* was expressed in skin fibroblasts and lymphoblastoid cells but not in white blood cells, whereas *KCNJ2* exhibited the opposite expression levels (Figure 30).

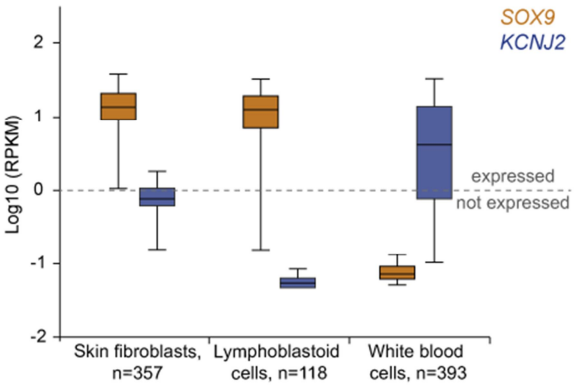


Figure 30: Expression levels of *SOX9* and *KCNJ2* in human-derived cells. Expression data are taken from the GTEx Consortium (2015). Shown are expression values of *SOX9* (brown) and *KCNJ2* (blue) in skin fibroblasts, lymphoblastoid cells and white blood cells. Box plots exhibit median RPKM (reads per kilo base of transcript per million mapped reads) and the 25th and 75th percentiles at a log10 scale. *SOX9* is expressed in skin fibroblasts (RPKM of 12.8) and lymphoblastoid cells (RPKM of 12.18) but not in white blood cells (RPKM of 0.07). *KCNJ2* is not expressed in skin fibroblast (RPKM of 0.78) or lymphoblastoid cells (RPKM of 0.06), but is in white blood cells (RPKM of 4.28).

Thereafter, chromatin interactions were captured in all three cell types using 4C-seq with viewpoints in the promoters of *SOX9* and *KCNJ2*. As described in the previous section, the interaction profiles of skin fibroblasts recapitulated the cHi-C predicted TAD structure (Figure 31a). Overall, lymphoblastoid cells had a reduced interaction frequency in the *KCNJ* TAD and a strong reduction in the centromeric region of the *SOX9* TAD compared to skin fibroblasts (Figure 31b). Yet, the interactions within each TAD were still more frequent than the background interactions outside the TADs, which allowed for detection of chromatin interactions restricted by TAD boundaries at the locus. In contrast, the centromeric region of the *SOX9* TAD demonstrated increased interactions compared to fibroblasts, suggesting cell-type-specific interactions in this genomic region within lymphoblastoid cells.

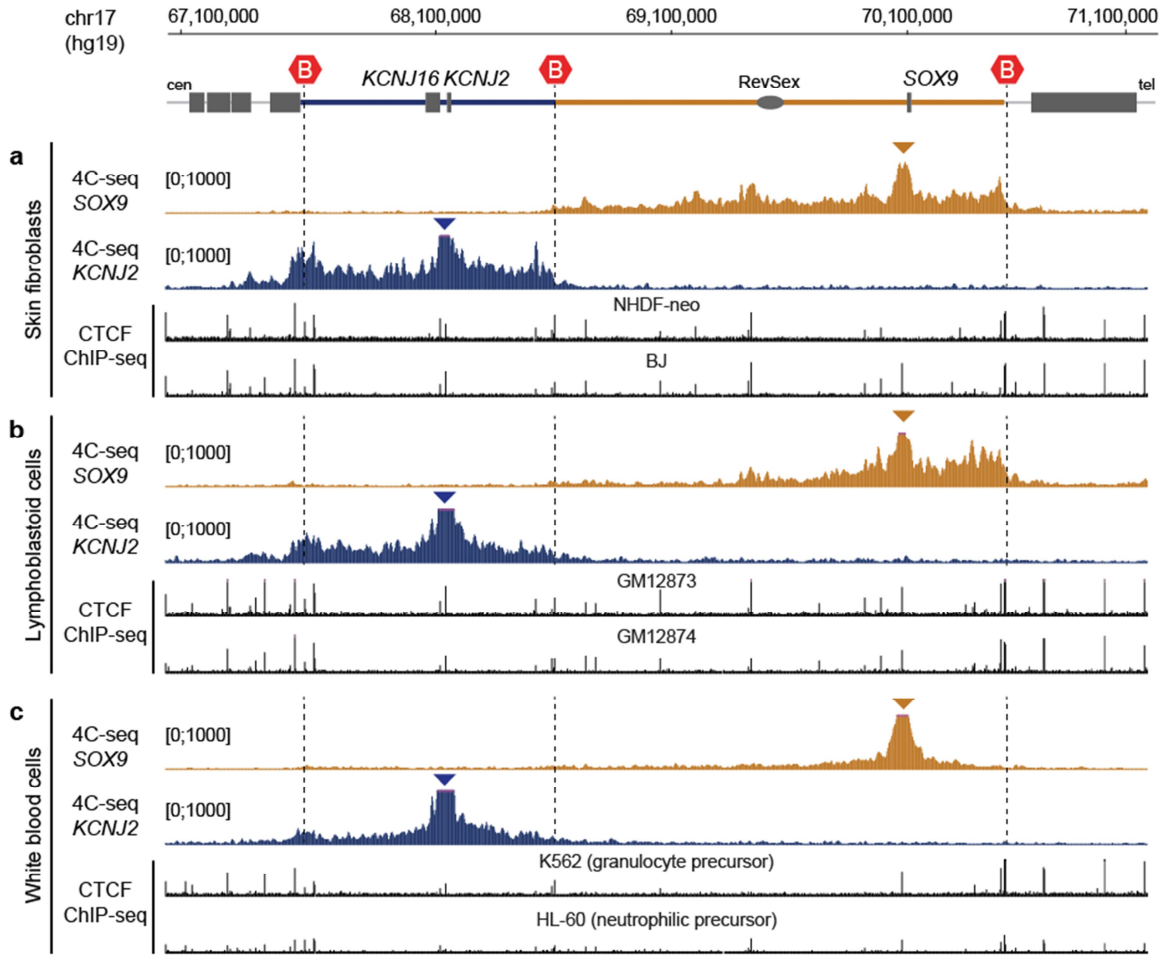


Figure 31: 4C-seq analysis in different human-derived cell types. 4C-seq analysis in skin fibroblasts (a), lymphoblastoid cells (b) and white blood cells (c) with viewpoints in *KCNJ2* (blue) and *SOX9* (brown). ChIP-seq signal from equivalent cell types from ENCODE/U and ENCODE/UW are shown below the 4C tracks. 4C-seq interaction profiles in skin fibroblasts and lyphoblastoid cells recapitulate the cHi-C-predicted TAD structure at the locus. White blood cells had an overall reduced interaction frequency from both viewpoints and lack the TAD restricted interaction profile from the *SOX9* viewpoint. Note that loss of TAD structure correlates with less binding of CTCF in white blood cells.

Surprisingly, chromatin interactions within the *KCNJ* and *SOX9* TADs in white blood cells were reduced to nearly background compared to the other cell types (Figure 31c). The 4C-seq profile from the *KCNJ2* viewpoint showed frequent interactions within the *KCNJ* TAD but the signal drop-off at the boundary towards the *SOX9* TAD was not discerned. Interactions from the *SOX9* viewpoint rapidly declined with linear distance to an overall background signal. Neither interactions within the TAD structure nor prominent interactions close to the viewpoint or with the TAD boundaries were observed. A probable basis for the overall reduced interaction frequency may be the storage conditions of the blood samples. White blood cells were isolated and prepared for 4C libraries 12 hours after the blood sample was taken. To show that the observation was specific to white blood cells, different time points and storage conditions were tested. 4C libraries from fresh blood samples were compared to blood samples stored for 24 and 48 hours at 4 °C and room temperature, respectively. No changes in the interaction profiles from

the *SOX9* viewpoint were observed with the different storage conditions, indicating that the observed 4C-seq profile was specific for white blood samples (data not shown).

SOX9 was not expressed in white blood cells, which corresponded with the lack of interactions seen. However, the absence of cell-type invariant interactions was not observed in the other tested cell types or tissues in this study. *KCNJ2* was only expressed in white blood cells, but the captured interaction frequency within the *KCNJ* TAD was more robust in skin fibroblasts and lymphoblastoid cells, which did not express *KCNJ2*. This observation was contradictory and negatively correlated with the expression of *KCNJ2*. Notably, the reduction of interactions within the *KCNJ* TAD and absence of any interactions within the *SOX9* TAD corresponded with the diminished binding of CTCF protein. Comparing CTCF ChIP-seq signals for equivalent cell lines from the ENCODE project revealed that skin fibroblasts and lymphoblastoid cell lines harbored strong CTCF signals, whereas precursor cells for different white blood cell types had reduced CTCF-binding signals (Figure 31). Therefore, the decreased interaction frequency and loss of compartmentalization into TADs at the locus depended on CTCF binding rather than on the transcriptional status of genes.

In summary, the results demonstrated the potential of human-derived cells for garnering insights into the configuration of TADs. Regardless of the variations within TADs, the overall configuration at the locus was detected in different cell types, independent of the gene expression levels. Chromatin interactions specific to the *KCNJ* and *SOX9* TADs were detected in skin fibroblasts and lymphoblastoid cells, although *KCNJ2* was not expressed. Interestingly, the comparison further suggested CTCF binding as an important structural element for TAD integrity and revealed that the abundance of CTCF binding was a strong predictor for determining TAD structure. The TAD structure at the locus was only observed in cell types that exhibited a large abundance of CTCF binding. White blood cells, with only weak CTCF binding, showed a strong reduction of interactions in the *KCNJ* TAD, although *KCNJ2* was expressed in these cells. Moreover, the 4C-seq interaction profile from the *SOX9* viewpoint demonstrated a complete absence of a spatially constrained *SOX9* TAD. Therefore, only white blood cells were not appropriate for accurately detecting the TAD configuration at the *SOX9* locus.

Finally, different cell types from patients with an inter-TAD duplication associated with Cooks syndrome were analyzed with the same 4C-seq approach (Figure 32). The 4C-seq profiles from skin fibroblasts and lymphoblastoid cells confirmed the same as those results from mice where one copy of *KCNJ2* was incorporated into the neo-TAD. The ratios relative to the respective controls showed no changes from the *SOX9* viewpoint, whereas the *KCNJ2* viewpoint ectopically contacted the duplicated portion of the *SOX9* TAD (Figure 32a and Figure 32b). This pattern of ectopic interaction was also observed in the 4C-seq profile from white blood cells. However, the absence of a spatially constrained TAD structure in these cells resulted in a gradual decline in the ectopic interaction signal. The telomeric region (arrowhead in Figure 32c), which was closer to the viewpoint, exhibited a higher interaction frequency than the centromeric region (arrow in Figure 32c). This was in contrast to cells that possessed an intact TAD structure at the locus. The ectopic signal, which was confined to the neo-TAD, had a constant interaction frequency over the whole region (Figure 32a and Figure 32b). The observed gradual loss in the 4C-seq interaction profile reflected the lack of compartmentalization in the neo-TAD as observed for the *SOX9* TAD in white blood cells. Although the formation of the neo-TAD was observed in skin fibroblasts and lymphoblastoid cells in patients with Cooks syndrome, the influence on gene regulation could not be addressed. The ectopic interactions seen resulted in an upregulation of *Kcnj2* in mice driven by regulatory elements within the neo-TAD. However, expression analysis in different patient cells was inconclusive as the variability in gene expression of the individual control and patient samples was very high (data not shown).

In conclusion, the formation of a neo-TAD induced by inter-TAD duplications was observed in cells with an intact TAD structure. Skin fibroblasts and lymphoblastoid cells from patients with Cooks syndrome recapitulated the properties of the neo-TAD, which were observed in the mice experiments to a great extent. Although the ectopic interactions from the *KCNJ2* viewpoint were also detected in white blood cells, the interaction profile had different properties. The captured linear decay in the interaction signal reflected a linear distance effect of the duplicated sequence at the locus. The interaction profile from the *KCNJ2* viewpoint, the absence of strong CTCF binding and the TAD structure all suggested that no spatially constrained neo-TAD was formed in these cells. The 4C-seq analysis in human cells demonstrated that patient-derived cells were adequate to describe changes in chromatin architecture induced by intra- and inter-TAD duplications. However, the effect on gene regulation could not be addressed based on high variability between samples.

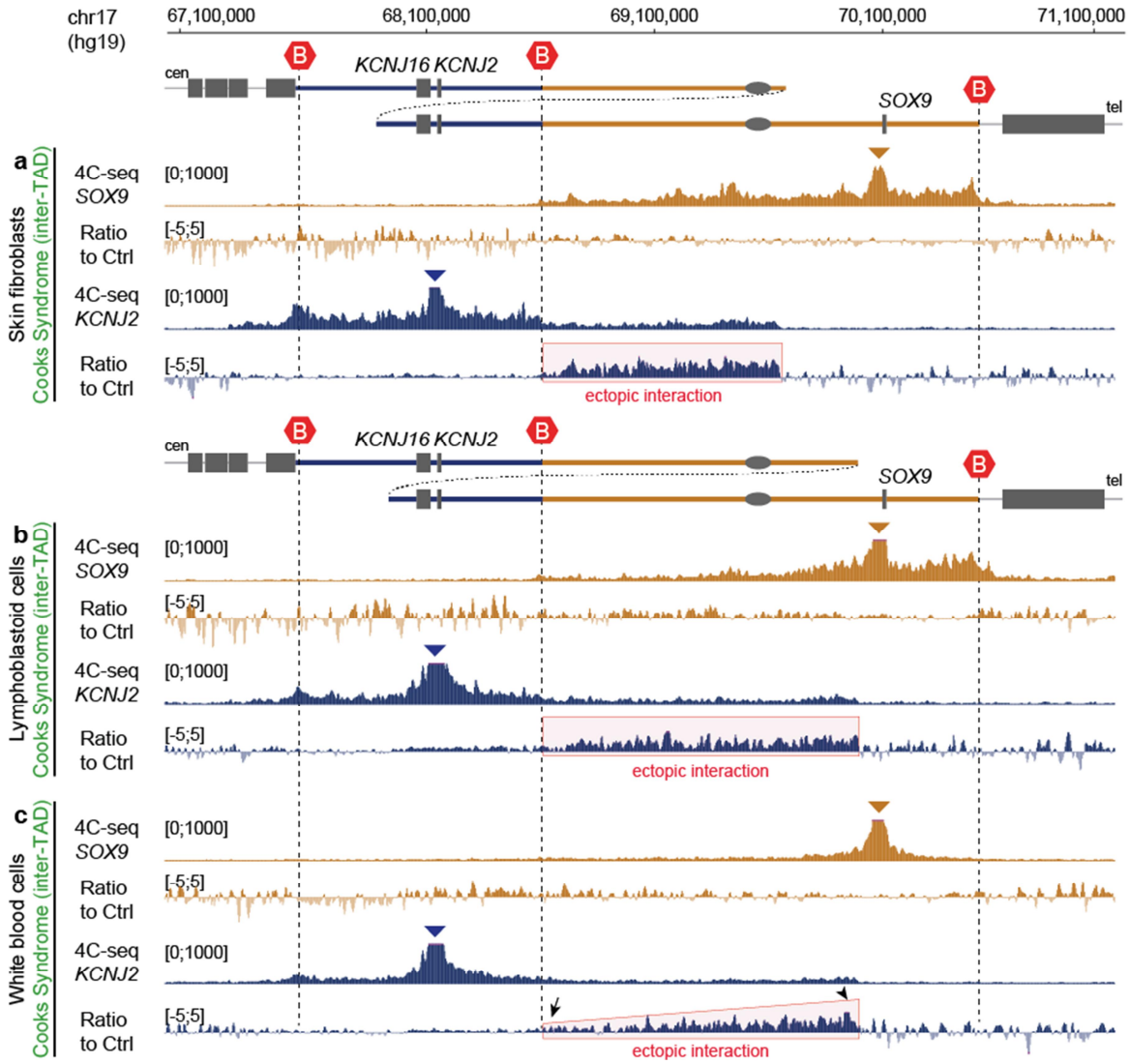


Figure 32: 4C-seq analysis in different patient-derived cell types with Cooks syndrome. 4C-seq analysis in skin fibroblasts (a), lymphoblastoid cells (b) and white blood cells (c) from patients with an inter-TAD duplication and Cooks syndrome. The lymphoblastoid cell line and the white blood cells originated from the same patient. 4C-seq profiles show viewpoints in *KCNJ2* (blue) and *SOX9* (brown) with corresponding log2 ratio to healthy controls. *SOX9* interaction profiles exhibited no changes compared to controls. The 4C-seq profiles from *KCNJ2* show ectopic contacts with the duplicated portion of the *SOX9* TAD, indicating incorporation of one copy of *KCNJ2* within the neo-TAD. Note that white blood cells showed a gradual decline from the *SOX9* viewpoint, suggesting the absence of *SOX9* TAD structure. The ectopic signal in the log2 ratio from the *KCNJ2* viewpoint had a similar decline with high interaction frequency close to the viewpoint (arrowhead) and low frequency more distally (arrow), suggesting the lack of compartmentalization in the neo-TAD.

6 Discussion

6.1 The spatial and functional organization at the SOX9 locus

Genome-wide approaches, such as Hi-C, have previously uncovered the spatial organization of the mammalian genome into TADs. TADs provide a structural context for genes and their CREs, as being physically proximal to remotely located enhancers with their target gene is required for transcription (Dixon et al. 2012; Deng et al. 2012). One of the aims of this thesis was to characterize the large gene desert upstream of *SOX9* to understand the involvement of the non-coding sequences in gene regulation. Here, 4C-seq and cHi-C were applied to investigate the TAD structure at the *SOX9* locus at high resolution in various human cell types and mouse tissues. In addition, the analysis of the spatial organization was combined with the expression analysis of several *LacZ* reporter insertions at the locus to investigate the activity of CREs and the presence of a regulatory domain structure at the locus (Ruf et al., 2011).

6.1.1 Subdivision of the SOX9 locus into two large chromatin domains

By taking advantage of the most recent advances in 3C technologies, cHi-C analysis revealed that the *SOX9* locus is partitioned into two large TADs. These TADs subdivide the large gene desert, forming a centromeric TAD, containing *KCNJ16* and *KCNJ2,* and a telomeric TAD, featuring *SOX9*. The identified spatial organization at the locus corresponds to the TAD structure that had been previously identified by genome-wide Hi-C (Dixon et al., 2012). These findings were consistent with the 4C-seq interaction profiles from viewpoints within each TAD, which accurately recapitulate the TAD structure. The 4C-seq profiles from *KCNJ2* and *SOX9* showed a broad region of consistently high interaction within the corresponding TAD, which is restricted by TAD boundaries.

TADs and their TAD boundaries are stable genomic units that are conserved across species (Dixon et al. 2012; Rao et al. 2014). In the present study, the TAD structure was compared in different mouse tissues as well as human cell types, which displayed a high degree of conservation between them. Although the investigated genomic region in humans is larger in size, the spatial TAD organization was conserved. The position of the TAD boundary between the *KCNJ* and *SOX9* TADs is determined by the presence of a cluster of CTCF-binding sites, whereby three of these sites are conserved between human and mice and constitutively bound by CTCF proteins. Comparative analysis of Hi-C data and CTCF binding in four mammals has demonstrated that the modular organization into TADs is robustly conserved in syntenic regions with highly conserved CTCF-binding sites at TAD boundaries (Vietri Rudan et al., 2015). These similarities in chromatin interactions and TAD structure indicate a conserved and similar regulation of gene expression

between the human and mouse *SOX9* locus, suggesting evolutionary constraints to maintain this spatial organization.

6.1.2 CTCF-mediated chromatin looping determines TAD structure

The high resolution of the generated interaction data set by cHi-C and 4C-seq herein from mice allowed the identification of additional features of chromatin interactions at the *Sox9* locus. The *Kcnj* and *Sox9* TADs showed additional domains of high chromatin interactions, which further compartmentalize the two TADs into smaller sub-TADs. Furthermore, both TADs and the internal sub-TADs were associated with chromatin loops, which were detected as focal points in the cHi-C and prominent interaction peaks in 4C-seq. The detected chromatin loops were all associated with CTCF binding and the analysis of the underlying sequence revealed that these loops were predominantly formed between CTCF sites with a convergent motif orientation. These findings are consistent with previous reports from high resolution Hi-C data and with the proposed model of TAD formation in which loop domains are formed between distantly located CTCF sites with a convergent motif orientation through extrusion (Rao et al., 2014; Sanborn et al., 2015; Fudenberg et al., 2016).

4C-seq analysis using the *Sox9* promoter as a viewpoint uncovered that these CTCF-mediated interactions are stable across different mouse tissues at E12.5 and are thus independent of *Sox9* transcription. *Sox9* is expressed in cells of the forebrain and limb bud but not in the liver. The 4C profile from liver cells exhibited an overall reduced interaction frequency throughout the *Sox9* TAD when compared to tissues expressing *Sox9*, indicating the relationship of chromatin interactions and gene expression. However, CTCF-associated interactions were maintained and represented the most prominent interactions, which were sufficient for delineating the *Sox9* TAD structure in liver cells. For instance, a CTCF-binding site 2 kb upstream of the *Sox9* promoter was engaged in CTCF-mediated interactions with the TAD boundaries and other CTCF-binding sites within the *Sox9* TAD. These results are consistent with the paradigm that TADs are stable genomic units that can be detected in different cell types, independent of the expression level of genes (Dixon et al., 2012; Lupiáñez et al., 2015).

The 4C-seq analysis in different human cell types with viewpoints in *KCNJ2* and *SOX9* further illustrated the role of CTCF in the integrity of the TAD structure. High interaction frequencies were detected throughout the *KCNJ* and *SOX9* TADs in skin fibroblasts and lymphoblastoid cells, although *KCNJ2* was not expressed. Both cells types had high signals for CTCF binding. In contrast, white blood cells showed a strong reduction of CTCF binding in both TADs. Interactions within the *KCNJ* TAD were strongly diminished although *KCNJ2* was expressed in these cells. This effect was even more pronounced within the *SOX9* TAD. Remarkably, the 4C-seq interaction profile from the

SOX9 viewpoint showed a complete absence of any interaction peak and no spatially constrained TAD in white blood cells. The comparison highlights the role of CTCF as a structural component for TAD formation and suggests that the presence and strength of CTCF binding is a robust predictor of TAD structure. This observation corroborated the results from recent studies demonstrating that the partial or complete depletion of CTCF protein leads to a loss of the segmental organization into TADs and the looping between CTCF sites(Nora et al., 2016; Zuin et al., 2014).

6.1.3 The dynamic nature of intra-TAD chromatin interactions

The comparison of chromatin interactions in different mouse tissues identified cell-type invariant interactions connected with CTCF binding that determined the overall *Sox9* TAD structure. However, the analysis also revealed an additional subset of highly dynamic and tissue-specific chromatin interactions from the *Sox9* promoter within its TAD. The *Sox9* promoter in limb bud cells had a preference for contacting the centromeric part of the *Sox9* TAD whereas forebrain cells preferentially contacted the telomeric part. These regional preferences correlated with the presence of H3K27 acetylation, a histone modification associated with active *cis*-regulatory regions (Creyghton et al., 2010). In contrast, cells from the whole liver, lacking H3K27 acetylation within the *Sox9* TAD, had an overall reduction in interactions. These dynamic chromatin interactions with potential enhancers suggest a complex *cis*-regulatory landscape within the *Sox9* TAD, which presumably contributes to control of the complex spatio-temporal expression pattern of *Sox9*.

Of note was that the observed tissue-specific enhancer-promoter interactions were independent of CTCF binding at the locus, in agreement with a recent investigation comparing 4C-seq profiles from several hundred promoters (Andrey et al., 2016). This study revealed the existence of two distinct regimes of chromatin interactions by comparing the interaction profiles across different tissues. One regime represents stable interactions present in all examined tissues and mainly associated with CTCF binding, while the second represents highly tissue-specific interactions strongly associated with chromatin marks typically observed for CREs. Mediator and cohesin have been demonstrated to cooperate in order to facilitate long-range interactions between enhancer-promoter pairs (Kagey et al., 2010). Mediator/cohesin usually bridge specific enhancer-promoter contacts at shorter ranges than CTCF/cohesin-mediated interactions (Phillips-Cremins et al., 2013). 4C-seq analysis has demonstrated that the *Sox9* promoter is engaged in both stable CTCF-mediated interactions and dynamic enhancer-promoter contacts. The observation of a stable *Sox9* TAD structure, independent of *Sox9* expression levels and the presence of active CREs, as seen in the mouse liver, suggest that specific enhancer-promoter interactions are formed in the context of pre-established CTCF-mediated contacts. Similar observations of a preset chromatin

conformation have been made at the *Shh* locus. Posterior limb bud expression of *Shh* is regulated by the *ZRS* (zone of polarizing activity) enhancer, located one mega base away from its target gene. Chromatin looping of the enhancer to *Shh* has been observed even in the anterior limb bud where *Shh* is not expressed (Amano et al., 2009). Future work, deleting CTCF-mediated looping to the *Sox9* promoter and comparing chromatin interactions in different tissues, will help to dissect the relationship between these two types of interactions.

6.1.4 Regulatory information is incorporated in TADs and sub-TADs

The analysis of the spatial organization at the *Sox9* locus was combined with the analysis of the regulatory activity using SB insertions. Taking advantage of the SB transposition system, multiple SB insertions were recovered located in the *Kcnj* and *Sox9* TADs. The SB insertions are coupled to a *LacZ* reporter gene controlled by a neutral promoter, which has no activity on its own but senses the activity of CREs at the integration site (Ruf et al., 2011). Thus, *LacZ* expression can be used as an indicator of potential regulatory activity of the genomic region containing the SB insertion.

The analysis of single SB insertions revealed highly tissue-specific expression patterns of the *LacZ* reporter at E12.5. Strikingly, SB insertions located within the *Kcnj* or *Sox9* TADs reproduced the endogenous gene expression patterns of *Kcnj2* or *Sox9* to a large degree, defining a regulatory domain structure that overlapped with the TAD structure at the locus. This approach was used previously to explore the regulatory domain structure at a genomic scale (Symmons et al., 2014). The results at the *Sox9* locus are consistent with the widespread presence of regulatory domains in the genome and that domains with similar regulatory activity overlap with TADs. TADs form spatially constrained domains with a high internal interaction frequency of chromatin, increasing the likelihood of genomic regions interacting with each other. Genome-wide approaches and studies from single loci have indicated that genes and enhancers that share the same interaction domain are highly correlated in terms of expression, as the physical proximity of enhancer-promoter pairs is an important mechanism for gene regulation (Arner et al., 2015; Deng et al., 2012; Lupiáñez et al., 2015).

Of interest was that the high density of SB insertions at the locus also revealed regional differences of *LacZ* reporter activity within the *Kcnj* and *Sox9* TAD. The comparison of regulatory domains with the high resolution cHi-C data from limb buds at E12.5 demonstrated that these regional differences overlap with the sub-TAD structure within each TAD. This is illustrated by SB insertions located in the sub-TAD centromeric of *Sox9*. All SB insertions in this region had strong *LacZ* expression in the digit anlagen at E12.5, recapitulating endogenous *Sox9* expression, whereas SB insertions in the telomeric sub-TAD showed only a weak signal in a subset of cells in

the digit anlagen. 4C-seq from limb buds further revealed that the centromeric sub-TAD is predominantly contacted by the *Sox9* and *LacZ* promoter of the SB-2 insertion that exhibited strong digit expression. Moreover, both 4C-seq profiles were similar with respect to patterns of interactions with regions in the centromeric sub-TAD, although both genes have different genomic positions. This observation implicates the *LacZ* reporter being incorporated into the sub-TAD and sharing the same regulatory sequences as the endogenous *Sox9* promoter. As described in the previous section, the centromeric sub-TAD was enriched for several regions decorated with H3K27 acetylation, indicating limb-specific enhancers. These findings suggest that chromatin domains can restrict the regulatory information encoded by enhancers on the one hand and that information for tissue-specific expression patterns at the *Sox9* locus is guided by fine-scale chromatin structures, such as sub-TADs, on the other.

The results also showed that enhancers act promiscuously on promoters located within chromatin domains, as *Sox9* and *LacZ* reporters shared the same enhancers and expression domains despite being located at different positions within the centromeric sub-TAD. The widespread activity of enhancer action within TADs supports a model in which enhancers act dynamically in defined chromatin compartments to activate gene expression (Symmons et al., 2014). Another study that measured the frequency and strength of transcription in *Drosophila* demonstrated that one enhancer can simultaneously activate the transcription of two promoters (Fukaya al., 2016). These observations support a model of regulatory neighborhoods constrained by chromatin domains or TADs, thus facilitating the spatial proximity of enhancers and promoters. This can create a local concentration of transcriptional co-activators and cofactors at gene promoters, which might result in the coordinated and simultaneous activation of different target genes by enhancers (Fukaya et al., 2016; Muerdter & Stark 2016).

The combination of chromatin interaction data and *LacZ* expression obtained at the *SOX9* locus underscore the role of TADs and sub-TADs as structural and functional units for gene regulation. Whereas the overall TAD structure at the locus remains stable across tissues, the observed differences in reporter activity along the *Sox9* TAD, together with the tissue-specific interactions by 4C-seq, imply dynamic changes within the TAD. The formation of sub-TADs may be vital for the spatio-temporal regulation of *Sox9.* However, the analysis of cell-type specific interactions at the *Sox9* locus is limited in this study to 4C-seq experiments in limb buds and forebrain and cHi-C in limb buds only. Further confirmatory work evaluating the TAD and sub-TAD structure with cHi-C in other *Sox9* expressing tissues will help elucidate the extent of tissue-specific changes and identify genomic regions within the *Sox9* TAD responsible for tissue-specific expression. The combination of these data with enhancer specific characteristics in the respective tissues, such as histone modifications or chromatin accessibility, serves as a powerful assay to identify novel candidate

enhancers for *Sox9* (Buenrostro et al., 2013; Creyghton et al., 2010). This would be particularly interesting for *Sox9* as the gene might require precise and dose-dependent regulation to fulfil its pleiotropic functions during development.

6.1.5 CTCF-mediated looping at TAD boundaries insulates regulatory and gene content within TADs

The spatial and functional compartmentalization into TADs at the *Sox9* locus seems largely determined by the presence of CTCF-mediated looping. CTCF-mediated loops were detected between the TAD boundaries of the *Kcnj* and *Sox9* TADs and between CTCF-binding sites associated with sub-TADs. The deletion of the TAD boundary that normally separates the *Kcnj* and *Sox9* TADs yielded important insights into the role TAD boundaries in insulating chromatin interactions within TADs and the role of additional CTCF-binding sites for TAD integrity and formation.

The *Kcnj2*/*Sox9* TAD boundary is characterized by a cluster of four CTCF-binding sites in a region of 18 kb in mice. Two CTCF sites face, with their motif towards, the *Kcnj* TAD and the other two sites towards the *Sox9* TAD. CHi-C further revealed that these binding sites are engaged in chromatin loops with other CTCF sites with a convergent CTCF motif within the adjacent TADs and TAD boundaries, respectively. The deletion of the cluster in mice (ΔBor) demonstrated that the TAD boundary was necessary for the spatial and functional separation of TADs. The subtraction of the wild-type interactions exhibited that new contacts across the *Kcnj* and *Sox9* TAD were formed, which resulted in gene upregulation of *Kcnj2*. These findings largely agree with previous findings at other loci. Deletion of a boundary at the *Xist* locus, for instance, led to a partial fusion of two adjacent TADs and ectopic gene activation (Nora et al., 2012). Further, *in vivo* studies at the *EPHA4* locus showed that the presence or absence of boundaries associated with CTCF-binding sites determined ectopic cross-TAD interactions and consequently affected gene regulation (Lupiáñez et al., 2015).

CHi-C in limb buds of ΔBor mice further exhibited that the loss of the CTCF-binding sites at the TAD boundary resulted in a redirection of chromatin loops. As an example, a new chromatin loop was established between the centromeric boundary of the *Kcnj* TAD and the telomeric boundary of the *Sox9* TAD. Significantly, the observed ectopic contacts throughout the *Kcnj* and *Sox9* TADs were restricted by this new chromatin loop, implicating the formation of a new domain. Furthermore, this domain was associated with CTCF-binding sites with convergent CTCF motifs, providing evidence for the hypothesis that convergent motif orientation is necessary for loop formation (Fudenberg et al., 2016; Rao et al., 2014). Multiple studies have demonstrated that deletions or inversions of single CTCF motifs can redirect chromatin loops and change the

topology of the associated chromatin domains. The inversion of a single CTCF-binding site within an enhancer reconfigures the chromatin loop of the enhancer to another target promoter at the *protocadherin* locus (Guo et al., 2015). At a more global level, Sandborn et al. showed through multiple CTCF-binding site deletions at different loci how new loop domains are formed between remaining CTCF-binding sites with convergent CTCF motifs (Sanborn et al., 2015).

Despite the formation of novel loops between the previously isolated *Kcnj2* and *Sox9* TADs, the boundary deletion did not disrupt the overall structure of the *Kcnj* and *Sox9* TADs. In particular, chromatin loops and the sub-TAD structure within both TADs were not affected and their preservation was probably sufficient for TAD formation. The overall unchanged sub-TAD structure largely maintained the chromatin interactions observed in wild-type animals, which could explain the limited effect of the observed ectopic contacts on gene regulation. In limb buds, *Kcnj2* was upregulated whereas *Sox9* expression remained unaltered. Moreover, RNA *in situ* hybridization at E12.5 elicited no site-specific misexpression of *Kcnj2* in limb buds and heterozygous or homozygous ΔBor mice had a normal limb phenotype.

It is noteworthy that the cHi-C experiments performed here were based on measurements of chromatin contacts across a large population of cells in the limb bud, an aspect that should be carefully considered. The loop extrusion model and experimental CTCF-binding site deletions presented in Sandborn et al. provide important implications for the interpretation of population-based Hi-C data. The authors showed that consecutive loop domains, as seen, for example, for the two sub-TADs upstream and downstream of *Sox9*, can occur simultaneously in the same cell. In contrast, overlapping loop domains cannot be present at the same time and therefore represent alternative states of chromatin interactions in different cells (Giorgetti et al., 2014; Sanborn et al., 2015). For instance, the TAD boundaries of the *Kcnj* and *Sox9* TADs in wild-type animals formed loop domains that overlapped with underlying loop domains or the sub-TADs, therefore representing differential states at the locus. The new interactions between the remaining TAD boundaries in ΔBor mice represented loop domains that overlapped with the preserved loop domains or sub-TAD structure, as well. It is therefore conceivable that the observed ectopic interactions with *Kcnj2* and *Sox9* reflect just a small subpopulation of cells which is not adequate to substantially interfere with gene regulation in the limb bud.

Supporting this notion, Lupiáñez et al. reported that large deletions that included sub-TADs and the boundary of the *EPHA4* TAD led to strong ectopic contacts of *EPHA4* enhancers with *PAX3* in the neighboring TAD and, consequently, gene misregulation and a pathological phenotype (Lupiáñez et al., 2015). Larger deletions at the TAD boundary that also included other CTCF-binding sites engaged in loop domains could potentially eliminate the sub-TADs at the *Sox9* locus. This in turn could force the state of chromatin interactions towards a large and fused TAD,

comprising parts of the *Kcnj* and *Sox9* TADs that are delineated by the remaining TAD boundaries. Alternatively, deleting additional CTCF-mediated loops in the ΔBor allele could induce a more dramatic redirection of chromatin interactions between both TADs, which could result in a complete fusion of TADs at the *Sox9* locus. These serial deletions would only interfere with CTCF-mediated looping and may yield further mechanistic insights into the role of three-dimensional architecture of mammalian genomes in gene regulation.

Although the TAD boundary at the *Sox9* locus is not necessary for the formation of the *Kcnj* and *Sox9* TAD, it is required to sufficiently insulate the *cis*-regulatory and gene content within the TADs. The results further suggest that the insulating function at the boundary is mainly determined by CTCF-mediated looping. Therefore, the presence of the CTCF cluster with two pairs of divergently orientated CTCF sites in the mouse genome might ensure robust insulation by creating overlapping loop domains with the corresponding neighboring TAD boundaries. Interestingly, across different mammals, the CTCF motif orientation at TAD boundaries is generally more conserved than within TADs (Vietri Rudan et al., 2015). Three of the four CTCF-binding sites in mice are conserved in humans, which keeps the divergent motif orientation of the cluster at the boundary. The boundary might therefore act as a robust insulator between TADs while interactions within TADs can vary across different species. Furthermore, a system of stable overlapping loop domains at TAD boundaries would allow dynamic stage- and tissue-specific changes in underlying loop domains or sub-TADs without changing the overall TAD configuration at the locus.

6.2 The effect of intra- and inter-TAD duplications on chromatin organization and gene regulation

Several tandem duplications, mapping to the gene desert upstream of *SOX9*, have been reported that have diverse phenotypic effects in humans. Small duplications upstream of SOX9 induce isolated sex reversal in female patients, whereas larger duplications that extend further centromerically, including the genes *KCNJ2* and *KCNJ16,* are associated with Cooks syndrome (Benko et al., 2011; Kurth et al., 2009). Additionally, a third duplication type of intermediate size that duplicates exclusively the entire gene desert upstream of *SOX9* was identified. Despite complete overlap with the sex-reversal duplications and a partial overlap with Cooks syndrome duplications, the carriers showed no abnormal phenotype with respect to sexual and skeletal development. A major goal of the present study was to elucidate the pathogenic effects of these duplications to explain the diverse phenotypic effects.

The comparison of the duplications position relative to the TAD structure at the *SOX9* locus uncovered a classification into intra- and inter-TAD duplications. All sex-reversal duplications increased the copy number of sequences within the *SOX9* TAD and they are referred to as intra-TAD duplications. On the contrary, inter-TAD duplications, including Cooks syndrome or normal phenotype duplications, comprised regions of the two neighboring *SOX9* and *KCNJ* TADs and the interjacent TAD boundary. Previous studies have described SVs potentially interfering with the TAD structure, leading to aberrant interaction and gene misregulation (Lupiáñez et al., 2015; Tsujimura et al., 2015). Here, cells from patients with sex reversal, Cooks syndrome and no phenotypic changes as well as equivalent mice models were employed to analyze the effect of the duplications on TAD structure and gene regulation at the *SOX9* locus.

6.2.1 The effect of intra-TAD duplications is restricted to the SOX9 TAD

Capture Hi-C analysis from limb buds at E12.5 from heterozygous mice carrying a 420 kb intra-TAD duplication (Dup-S) revealed that the extra copy resulted in an increase of interactions within the *Sox9* TAD. This increase was observed for all genomic sequences within the *Sox9* TAD, including *Sox9*. Importantly, the gain of interactions was restricted by the *Sox9* TAD boundaries, implying that the duplicated part is incorporated into the *Sox9* TAD without affecting the overall TAD configuration or other genes at the locus. The observation in mice was confirmed by 4C-seq analysis in patient fibroblasts with a 150 kb intra-TAD duplication and sex reversal. The *SOX9* viewpoint had increased interactions with the extra copy and an allele-specific 4C-seq approach with a viewpoint at the duplication's breakpoint established that contacts determined by the duplicated part were limited to sequences of the *SOX9* TAD. All intra-TAD duplications associated with sex reversal in humans contained the RevSex region (Kim et al., 2015). The analysis

demonstrated that this potential regulator for female-to-male sex reversal is “free” to interact with *SOX9*.

6.2.2 The impact of intra-TAD duplications on SOX9 regulation

The level of *SOX9* expression during gonad development is imperative for the determination of testicular fate (Foster et al., 1994; Wagner et al., 1994). Elevated expression of *SOX9*, for example, caused by gene duplications or constitutive activation of *SOX9* can result in female-to-male sex reversal in humans and mice (Bishop et al., 2000; Huang et al., 1999). Similarly, upregulation of *SOX9* induced by intra-TAD duplications, including the RevSex region, have been discussed as the cause for female-to-male sex reversal. Several models have been proposed for how duplications can result in elevated *SOX9* expression. One model hypotheses that the increase of the copy number of the RevSex region could potentially raise the site- and stage-specific expression of *SOX9* in the developing gonad, leading to testis differentiation independent of *SRY* (Benko et al. 2011; Kim et al. 2015). Before the onset of *Sry* expression, the bipotential gonad in both sexes expresses *Sox9* at low levels, which is controlled by SF1 and WT1 at this early stage (Kidokoro et al., 2005; Wilhelm, 2002). An additional copy with potential binding sites for these nuclear factors could increase the basal expression levels above a threshold required for testis induction (Kim et al., 2015). Another model suggests that the RevSex region is characterized by active chromatin marks in males but inactive ones in females. The duplication could induce a position effect, leading to an escape of the extra copy into an active state in females (Benko et al., 2011; Lybæk et al., 2014).

The results obtained herein suggest *SOX9* being the causative gene for sex reversal for the described human duplications, as the inclusion of the RevSex region results in a gain of contacts with *SOX9* in humans and mice. Given the importance of physical proximity of enhancer-promoter pairs for gene regulation, the observed rise in number of interactions could potentially lead to an increase in *SOX9* expression. This mechanism has recently been suggested for recurrent enhancer duplications in cancer (Zhang et al., 2016). However, the consequences for gene expression have not been addressed in this study. Future work is needed to determine the effect of intra-TAD duplications on *SOX9* expression. A first step towards this is to perform expression analysis on gonads at different stages during sex determination in animals of the generated intra-TAD duplication (Dup-S).

In contrast to humans, the intra-duplication in Dup-S mice mutants did not result in obvious abnormal sexual development or infertility in heterozygous females. The absence of a phenotype could potentially arise from different threshold levels of *Sox9* expression required for testis specification in human and mice. This becomes evident from loss-of-function alleles.

Heterozygous null mutations of *Sox9* do not lead to male-to-female sex reversal in mice, which contrasts with the haploinsufficiency in humans (Bi et al., 2001; Wagner et al., 1994). In addition, a recent work demonstrated that a reduction to 25 % of the normal *Sox9* expression level in mice is required to induce partial male-to-female sex reversal, resulting in ovo-testis (Gonen et al., 2017). Similarly, inter-species differences in threshold levels of *Sox9* to induce female-to-male sex reversal could account for the absence of a phenotype.

However, if the Dup-S allele is an appropriate model to study a potential *cis* regulatory effect of intra-TAD duplications and sex reversal in the mouse, it is unclear as of yet. Several lines of evidence point to a different *cis*-regulation of *Sox9* during testis development between human and mice. The well-known testis-specific enhancer, *Tes*, in mice, located 20 kb upstream of *Sox9* is not included in the duplication of Dup-S mice (Sekido & Lovell-Badge 2008). Moreover, no mutations in the human *TES* region have so far been described and the human *TES* sequence is not able to drive reporter expression in the developing gonad in mice (Sekido & Lovell-Badge 2009; Georg et al., 2010). Furthermore, the human RevSex region has so far not been linked to *Sox9* regulation in mouse gonad specification (Kim et al., 2015). The analysis of a molecular phenotype and histological inspection of hetero- and homozygous Dup-S mice in the gonads could aid in answering the question of whether the RevSex region contains potential enhancers for mouse testis development and how intra-TAD duplications interfere with gene regulation. Of note is that it was recently confirmed that *Tes* is a crucial testis enhancer for *Sox9* expression in mice gonads. However, the *Tes* enhancer contributes only 25 % to the normal expression levels of *Sox9* in the developing testis, indicating the existence of additional enhancers at the *Sox9* locus that act redundantly (Gonen et al., 2017).

6.2.3 Neo-TADs insulate the extra copy induced by inter-TAD duplications

Inter-TAD duplications at the *SOX9* locus in humans are phenotypically different from intra-TAD duplications as they are not associated with sex reversal. The analysis of chromatin interactions showed that the effect of inter-TAD duplications on TAD topology at the *Sox9* locus is fundamentally different to the observations from intra-TAD duplications. The additional copy, caused by inter-TAD duplication, is spatially isolated from the *Kcnj* and *Sox9* TAD and consequently from *Kcnj2* and *Sox9*. CHi-C analysis from Dup-L mice mutants revealed ectopic interactions between the duplicated parts of the *Kcnj* and *Sox9* TADs. Importantly, the additional contacts are exclusively confined to the duplicated region and genomic sequences outside of the duplicated region are not affected. The results from this study led to the concept of the formation of a new chromatin domain (neo-TAD) comprising the duplicated region between the duplicated TAD boundaries. This new interaction domain forms between the *Kcnj* and *Sox9* TADs. The term

neo-TAD used herein was chosen as this new chromatin domain features typical properties of "regular" TADs.

First, TADs represent genomic units with spatially constrained internal interactions and it is rare there are interactions beyond their respective TAD boundaries (Dixon et al., 2012). Allele-specific 4C-seq with viewpoints specific for sequences located within the neo-TAD (*LacZ* at the duplication breakpoint) identified interactions confined to the neo-TAD. The interaction profiles from the *LacZ* viewpoint in Dup-L mice revealed an additional interaction domain that was not captured from the *Kcnj2* or *Sox9* viewpoints in the neighboring TADs. A similar 4C-seq approach in combination with SNP analysis in patient cells with inter-TAD duplication uncovered selective interactions that were specific to the neo-, *KCNJ* or *SOX9* TADs. Furthermore, this new TAD configuration and neo-TAD formation at the *SOX9* locus was observed in mutant mice limb buds as well as different human patient cell types, suggesting an evolutionary conservation of the mechanisms leading to its formation (Dixon et al., 2012; Lupiáñez et al., 2015).

Second, as discussed in the previous section (6.1.5), the duplicated TAD boundary that separates under wild-type conditions the *Kcnj* and *Sox9* TADs has similar functions in the neo-TAD. The deletion of the duplicated TAD boundary in Dup-LΔBor mice resulted in ectopic contacts across the neo-, *Kcnj* and *Sox9* TADs and altered the gene expression of *Kcnj2* in the adjacent TAD, recapitulating the results obtained with the ΔBor mutants. The experiment confirmed that the neo-TAD is a separate chromatin domain which requires TAD boundaries to fully insulate the duplicated portion within the neo-TAD from its neighboring TADs.

Third, TADs represent functional units that retain regulatory activities, contributing to the establishment of specific gene expression patterns (Symmons et al., 2014; Tsujimura et al., 2015). *LacZ* expression analysis in Dup-L mutants demonstrated that the genomic sequence within the neo-TAD is able to drive tissue-specific expression patterns resembling endogenous *Sox9* expression. The neo-TAD contains large parts of the *cis*-regulatory domain upstream of *Sox9*. Comparison of the interaction profiles with viewpoints in *Sox9* and *Kcnj2* in the wild-type as well as that of *LacZ* in Dup-L mice revealed that sequences in the neo-TAD retain their wild-type specific interaction patterns. The similar folding pattern within the neo-TAD presumably allows the *LacZ* promoter to be activated in a *Sox9*-like expression pattern.

The experiments in mice and humans demonstrated that neo-TADs are spatially constrained chromatin domains that isolate the duplicated region from their genomic neighboring TADs. Inter-TAD duplications in patients that exhibit normal phenotype result in a physical separation of additional *cis*-regulatory sequences, including the potential regulator, RevSex. The insulation within the neo-TAD accounts for the absence of a *SOX9*-related phenotype, such as female-to-male sex reversal described for intra-TAD duplications in humans.

6.2.4 Regulatory and gene content within neo-TADs determine pathogenicity of inter-TAD duplications

Inter-TAD duplications associated with Cooks syndrome or normal phenotypes in humans overlap almost completely with regions in the *SOX9* TAD. However, Cooks syndrome-associated duplications extend further centromerically into the *KCNJ* TAD, including *KCNJ2* and *KCNJ16* (Kurth et al., 2009). CHi-C and 4C-seq analysis of the equivalent duplication in mice mutants (Dup-C) showed that the duplicated genes, *Kcnj2* and *Kcnj16*, are incorporated into a neo-TAD. Similarly, 4C-seq from patient cells with a viewpoint in *KCNJ2* confirmed incorporation through showing ectopic interactions with the *SOX9* duplicated region. As with the *LacZ* reporter in Dup-L mutants, the copy of *Kcnj2* in Dup-C mice adopted the *Sox9*-specific *cis*-regulatory information retained within the neo-TAD, leading to misregulation of *Kcnj2* in a *Sox9*-like expression pattern in the digit anlagen.

Of particular interest was that despite the *LacZ* reporter in Dup-L mice and the genes, *Kcnj2* and *Kcnj16*, in Dup-C mice being exposed to the same *cis*-regulatory information in the neo-TAD, they responded differentially. The *LacZ* reporter had a *Sox9*-like expression in the digit anlagen and other skeletal structures, recapitulating endogenous expression of *Sox9*. Meanwhile, *Kcnj2* misexpression was only detected in the digits and *Kcnj16* did not respond to any *cis*-regulatory input. The three genes have different positions within the neo-TAD relative to the duplication breakpoint, which could induce local differences in chromatin interaction and therefore result in differential gene activation. However, 4C-seq interaction profiles from the *LacZ* reporter and *Kcnj2* generally featured the same extent of ectopic interactions within the neo-TAD. Furthermore, *LacZ* expression analysis of different SB insertions at the locus demonstrated that the regulatory information within TADs has a widespread activation potential that can act promiscuously on genes.

Other mechanisms that determine enhancer-promoter specificity, such as the biochemical compatibility between enhancers and promoters, could explain the expression differences between genes integrated into the neo-TAD. Genome-wide screening has shown that enhancers show distinct preferences for particular promoter types (Arnold et al., 2013; Zabidi et al., 2015). Furthermore, enhancer-promoter specificity can be mediated by unique core promoter motifs and additional proximal promoter elements that can regulate the regulatory input of a gene (Butler & Kadonaga, 2001; Deng et al., 2012). The promoter of the *LacZ* reporter has been previously selected for its ability to react to widespread regulatory activity and previous studies have shown that the promoter is active in multiple tissues during development (Morgan et al., 1996; Ruf et al., 2011). In contrast, *Kcnj2* and *Kcnj16* respond differentially to the regulatory information in their endogenous genomic environment. Indeed, the analysis of SB insertions in

the *Kcnj* TAD showed that *Kcnj2* and *Kcnj16* are embedded in a complex regulatory domain structure with expression domains in the limb and maxilla. *Kcnj2* possesses the ability to respond to the regulatory information whereas *Kcnj16* is not expressed in these tissues. Interestingly, as shown by RNA *in situ* hybridization, is that the limb is the only tissue in which *Kcnj2* is misregulated in Dup-C mice, indicating that the promoter of *Kcnj2* is compatible and permissive to activation in this tissue.

In addition, the tissue-specific chromatin compositions at the promoter of *Kcnj2* and *Kcnj16* could prevent gene activation. The repressive chromatin mark, H3K27 tri-methylation (H3K27me3) at promoters and the gene body, established by Polycomb complex activity, has been demonstrated to be associated with gene repression (Young et al., 2011). The copies of *Kcnj2* and *Kncj16* in the neo-TAD largely preserve their endogenous genomic context around the gene, which may be sufficient for retaining their chromatin composition and therefore, exemplify a different response to regulatory input in the neo-TAD.

6.2.5 *KCNJ2* as a disease-causing gene in Cooks syndrome

Misregulation of *SOX9* has been previously proposed as a cause of Cooks syndrome (Kurth et al., 2009). Instead, this study shows that only misexpression of *Kcnj2* in a *Sox9*-like expression pattern resulted in a limb phenotype at birth, recapitulating the human Cooks syndrome to a large extent. Additional mouse alleles in this study strongly support misregulation of *Kcnj2* in the limb as the disease-driving mechanism. The inter-TAD duplication in Dup-C2 mice excluded *Kcnj16* from being duplicated, with incorporation of *Kcnj2* alone in the neo-TAD. This configuration was adequate for inducing the Cooks-like phenotype in heterozygous Dup-C2 mice. Furthermore, two intra-TAD duplications in the *Kcnj* TAD were generated that negated the possibility Cooks syndrome was merely produced by a copy number increase of *Kcnj2*. Heterozygous mice (Dup-K1 and Dup-K2), duplicating *Kcnj2* and *Kcnj16*, exhibited a normal phenotype.

Intriguingly, the observed digit and joint defects in heterozygous mice with Cooks-like phenotype and the brachydactyly phenotype in Cooks patients is reminiscent of other brachydactyly syndromes that have been linked to alterations in the BMP pathway (Stricker & Mundlos, 2011). BMPs belong to the transforming growth factor-beta (TGF-beta) superfamily that positively regulates chondrogenesis (Duprez et al., 1996; Pizette & Niswander 2000). BMPs are expressed in the perichondrium and interdigital mesenchyme. Binding of the ligand to the BMP receptor, expressed in condensing chondrocytes, leads to phosphorylation and activation of intracellular signaling molecules, including SMAD proteins. Phosphorylated SMADs associate with other co-factors, translocate into the nucleus and activate target gene expression, including *SOX9* in chondrocytes (Stricker & Mundlos 2011). The formation of joints requires the inhibition of BMP signaling, and mutations in BMP antagonists, such as NOGGIN, result in joint fusion in mice and abnormal joint formation in humans (Brunet et al., 1998; Warman et al., 1999). However, how misexpression of *KCNJ2* in chondrocytes could potentially interfere with BMP signaling is not documented in humans or mice.

KCNJ2 belongs to the family of inward rectifying potassium channels and dominant negative mutations, suppressing the channel function brings about Anderson-Tawil syndrome in humans (Plaster et al., 2001; Tristani-Firouzi et al., 2002). Among other channelopathies, Anderson-Tawil syndrome is unique as it combines a cardiac muscle phenotype and morphological abnormalities, including low-set ears, cleft palate, micrognathia and clinodactyly (Andersen et al., 1971; Tawil et al., 1994;Tristani-Firouzi et al., 2002). Homozygous knockout mice for *Kcnj2* also exhibit cleft palate and digit defects, demonstrating that mutations in the *KCNJ2* potassium channel can lead to developmental defects (Dahal et al., 2012; Zaritsky et al., 2000). Whereas the electrophysiological consequences of *KCNJ2* mutation in the muscle are well-characterized in humans and mice, the mechanisms leading to developmental defects are unclear (Tristani-Firouzi

et al., 2002; Zaritsky et al., 2000). It is noteworthy that the knockout or overexpression of a dominant negative mutation of the *KCNJ2* homolog, *Irk2*, in *Drosophila* has been shown to cause wing patterning defects similar to those found when signaling through the BMP homolog, *Dpp* (Decapentaplegic), is disrupted (Dahal et al., 2012). Dahal et al. showed that reduced Irk2 levels reduce Dpp signaling in the wing disc. They found diminished phosphorylation of the SMAD homolog, Mad, and decreased levels of Dpp target gene expression, providing a possible link for how intracellular potassium concentrations can modulate Dpp signaling (Dahal et al., 2012). It is thus conceivable that the specific misexpression of *KCNJ2* in a *SOX9*-like expression pattern in chondrocytes disrupts BMP signaling during digit development, creating the unique Cooks phenotype. Overexpression of *KCNJ2* could increase the intracellular levels of potassium, which potentially promote BMP signaling, resulting in the underdeveloped and partly fused joints as seen in the Dup-C mice mutants with a Cooks-like phenotype. However, future work is necessary to understand the effects of *KCNJ2* overexpression in the digits. The generated Dup-C mice mutants provide a model for examining the contribution of potassium channel activity to developmental signaling, such as that of the BMP pathway.

RNA expression analysis also demonstrated an upregulation of *Kcnj2* in Dup-C at E17.5 in the tips of the digits in Dup-C mice. This upregulation is likely from misregulation of *Kcnj2* in a *Sox9*-like expression pattern in the developing nail organ. *Sox9* is expressed at E17.5 in the nail field, the earliest recognizable nail structure at the distal tips of the digits in mice and humans (Krahl & Sellheyer 2010; Kurth et al., 2009; Zaias 1963). During later stages, *SOX9* expression becomes confined to the nail bed below the nail plate (Krahl & Sellheyer 2010). Although the role of *SOX9* during nail development is not completely clear, the early onset of its expression makes it probable that the protein is critical for nail development (Krahl & Sellheyer 2010). *SOX9* could have similar functions in the nail organ as in hair follicles – they are closely related structures that derive from the primitive epidermis (Baran, Dawber & Haneke 2005). Tissue-specific knockouts have already demonstrated that *Sox9* is required for differentiation and development of hair follicles and for the specification of hair follicle stem cells (Nowak et al., 2008; Vidal et al., 2005). It is therefore possible that *SOX9* regulates nail stem cell specification during early embryogenesis, as well (Krahl & Sellheyer 2010). However, how *KCNJ2* misexpression in *SOX9* expressing cells in the nail field could lead to nail hypoplasia or aplasia needs to be investigated further.

The histological examination of the digit tips at P0 in Dup-C mice mutants showed that the nail plate and underlying nail bed is formed but underdeveloped and severely reduced in size, indicating possible proliferation defects during early development. Potassium channels have been implicated in the control of cell proliferation in multiple cell lineages during development and in cancer cells (reviewed in Urrego et al. 2014). Intra-cellular potassium concentrations can

modulate membrane potential and different potassium channels exhibit variations in expression and activity during the cell cycle (Arcangeli et al., 1995; Pardo et al., 1998). Moreover, potassium channels are the driving force for calcium to enter the cell. Calcium, in turn, is an important mediator of intracellular signaling in many physiological processes, including the control of proliferation (Lallet-Daher et al., 2009; Lee, Sayeed, & Wurster 1993). It is hence likely that changes in the potassium exchange induced by misexpression of any channel would affect cell physiological properties and cell cycle control (Urrego et al., 2014). Interestingly, two recent studies provided further evidence that the balance of potassium channel activity is critical during tissue development, including that of the nail organ. Gain-of-function mutations in *KCNH1*, a voltage-gated potassium channel, are associated with Temple-Baraitser and Zimmermann-Laband syndromes (Kortüm et al., 2015; Simons et al., 2014). Both are multi-systemic disorders, characterized by defects in the central nervous system with epileptic seizures, intellectual disabilities and several developmental defects, including hypoplasia or aplasia of nails (Laband, Habib & Humphreys 1964; Temple & Baraitser 1991). The mutations lead to overactivity of KCNH1 and, consequently, in changes of the membrane potential *in vitro*. Furthermore, the observed shift in membrane potential impedes other voltage-gated sodium and calcium channels, which could clarify the wide phenotypic spectrum seen in both syndromes (Kortüm et al., 2015; Simons et al., 2014). Specific misexpression of *KCNJ2* in the nail field could therefore shift the membrane potential and alter the intra-cellular calcium concentration. With this, the resulting proliferation defects during early development or at later stages in stem cells could account for the hypoplasia or aplasia of nails in Cooks syndrome.

6.2.6 Formation of neo-TADs explains diverse phenotypic effects of overlapping tandem duplications at the SOX9 locus

Genome-wide diagnostic screenings frequently led to identification of genomic SVs, ranging from the kilo- to mega base scale, which can include genes as well as large portions of non-coding sequences (Klopocki & Mundlos 2011). It is often the case that prediction of their pathogenic effects is a challenging task for clinicians and human geneticists as there is more than one possible mechanism by which SVs can disrupt gene function and result in a certain phenotype. The potential pathogenicity of duplications or deletions, for example, is generally assessed by their effect on gene dosage (Rice & McLysaght 2017). However, this explanation is sometimes insufficient, as either the observed pathological phenotype cannot be understood through alterations in gene dosage or only non-coding sequence of the genome is involved (Benko et al., 2011; Kurth et al., 2009). The knowledge of the compartmentalization of the mammalian genome into TADs supplies researchers and clinicians with new strategies to study the consequences of SVs on gene regulation and to identify the underlying disease-causing mechanisms (Lupiáñez et al., 2015).

The analysis of overlapping tandem duplications at the *SOX9* locus by 4C-seq and cHi-C demonstrated the different effects duplications can have on chromatin organization and gene regulation in humans and mice. The findings involve a previously undescribed mechanism, whereby tandem duplications lead to the formation of neo-TADs. In particular, the present study described three scenarios of how overlapping duplications lead to diverse phenotypic outcomes (Figure 33). First, intra-TAD duplications can result in increased interactions of the extra copy with target genes in the TAD they belong to without changing the overall TAD configuration at the locus. The increased contacts of *SOX9* with CREs, including the RevSex region in the duplicated sequence, potentially causes female-to-male sex reversal in humans. Second, inter-TAD duplications that cover parts of two TADs and the intersecting TAD boundary induce the formation of neo-TADs. Neo-TADs form a separate chromatin domain that shields the duplicated region from neighboring TADs and genes. The insulator effect explains the absence of a pathogenic phenotype. The last scenario involves neo-TADs relevant to disease depending on their gene content. Genes that are integrated into the neo-TAD as a result of larger inter-TAD duplications adopt the neo-TAD specific *cis*-regulatory information. The combination of duplicated *KCNJ2* with *cis*-regulatory sequences normally regulating *SOX9* in the neo-TAD brings about *KCNJ2* misexpression and Cooks syndrome.

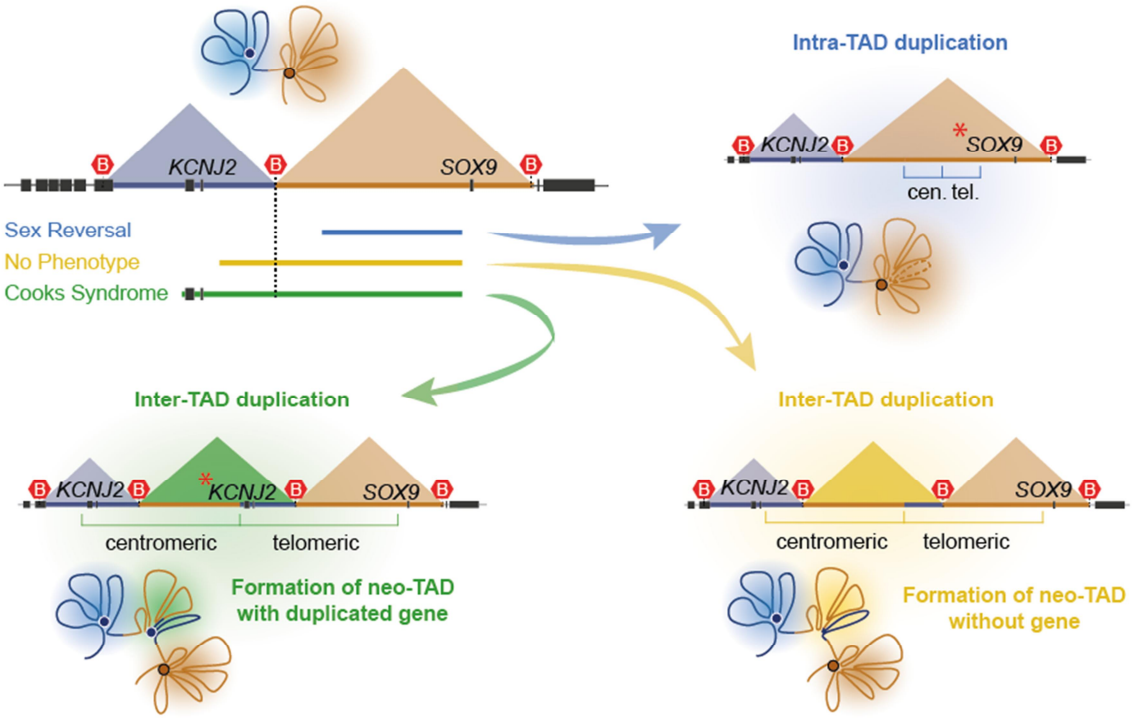


Figure 33: Disease mechanisms induced by intra- and inter-TAD duplications at the *SOX9* locus. Schematic (upper left) illustrating TAD configuration at the locus into *KCNJ* and *SOX9* TADs under normal conditions. Below, the size and position of overlapping tandem duplications associated either with sex reversal, no phenotypic changes or Cooks syndrome (Benko et al., 2011; Kurth et al., 2009). The duplication position relative to the TAD structure determines their pathogenic impact. The effect of intra-TAD duplications (blue) is restricted to the *SOX9* TAD. Inter-TAD duplications (yellow and green), intersecting with a TAD boundary, result in the formation of new chromatin domains (neo-TADs) between the *KCNJ* and *SOX9* TADs. The neo-TAD insulates the duplicated sequences and therefore neutralizes gene regulatory effects (yellow). Genes incorporated into the neo-TAD adopt its regulatory potential and can result in gene misexpression (green, Cooks syndrome). Asterisks indicate genes affected by the duplication. The centromeric and telomeric portions of the tandem duplications are indicated. Figure adapted from Franke et al. (2016).

These findings can be similarly applicable to SVs at other genomic loci and are not only restricted to developmental diseases. As stated earlier on, Lupiañez et al. reported that at the *EPHA4* locus, there are inter-TAD duplications linked with limb malformation, known as F-syndrome. This malformation results from *WNT6* misexpression in the limb induced by ectopic contacts of *WNT6* with limb-specific enhancers that are normally separated by a TAD boundary (Lupiáñez et al., 2015). The result resembles the observed ectopic contacts of *KCNJ2* in the neo-TAD. In addition, a recent study identified recurrent inter-TAD duplications in cancer. The authors described the formation of a *de novo* contact domain encompassing a super-enhancer that is normally inaccessible to *IGF2* gene regulation. This new contact domain bears largely the same features as that described here with regards to neo-TADs and demonstrates that they can drive massive overexpression of *IGF2* in somatic cancer (Weischenfeldt et al., 2016).

It is critical to consider the TAD structure and the potential chromatin interaction changes for the identification of disease-causing genes. *KCNJ2,* for instance, has previously been ruled out as a

cause of Cooks syndrome (Kurth et al., 2009). *KCNJ2* function is not sensitive to dosage and until now, the limb phenotype in Cooks syndrome has not been associated with *KCNJ2* mutations (Zaritsky et al., 2000). However, the implementation of chromatin interaction data in this study led to the discovery of *KCNJ2* as the underlying disease-causing gene for Cooks syndrome. The comparison of duplications' sizes and positions in the context of TAD structure and TAD boundary position at a particular locus will assist clinicians and researchers in better predicting the pathogenic effects of duplications. As shown in the work presented here, the overall TAD structure and CTCF binding at TAD boundaries is largely preserved across different cell types. Hi-C or CTCF ChIP-seq data for multiple human and mouse cell lines are publically available and can be directly utilized to study potential changes in TAD structure (Birney et al., 2007; Dixon et al., 2012; Jin et al., 2013; Rao et al., 2014). As well, the effect of intra- and inter-TAD duplications that has been investigated here can easily be tested in diagnostic settings. Importantly, this study demonstrates a cost efficient 4C-seq approach in patient cells using the duplication breakpoint as a unique viewpoint in the genome. The approach allows determining the extent of ectopic or increased interactions between non-coding sequences and genes that are potentially involved in pathogenicity. Ultimately, the present study offers a framework for how to interpret SVs, such as tandem duplications, in the context of gene regulation and disease.

7 Summary

Chromosome conformation capture methods have identified sub-mega base structures of higher-order chromatin interactions in vertebrates called topologically associating domains (TADs) that are separated from each other by TAD boundaries. TADs subdivide the genome into discrete chromatin domains that direct the contacts enhancers can establish with their target genes. However, the mechanisms that underlie the partitioning of the genome into TADs and how this might be disturbed by structural variations (SVs), such as duplications, remain poorly understood.

Here we show by capture Hi-C and 4C-seq that the *SOX9* locus is spatially subdivided in two TADs in humans and mice, one comprising *SOX9* (*SOX9* TAD), the other the neighboring genes *KCNJ2* and *KCNJ16* (*KCNJ* TAD). *LacZ* expression analysis of reporter insertions in both TADs revealed that this spatial organization confines the *cis*-regulatory activity at the locus. Furthermore, we show in patient samples and genetically modified mice that tandem duplications in this region can interfere with TAD structure, resulting in the formation of new chromatin domains (neo-TADs). Neo-TADs are functionally and spatially separated from their neighboring *KCNJ* and *SOX9* TADs and thereby determine the duplications' pathogenesis. In humans, partially overlapping duplications upstream of *SOX9* can cause female-to-male sex reversal, can have no phenotypic effect, or result in a limb malformation called Cooks syndrome. Duplications that occur within the *SOX9* TAD but do not extend over the TAD boundary (intra-TAD), cause sex reversal. We observe that the duplicated *cis*-regulatory region showed increased contacts with the target gene *SOX9* and only within the *SOX9* TAD. In contrast, duplications expanding over the TAD boundary into the neighboring *KCNJ* TAD (inter-TAD), result in the formation of neo-TADs. We show that the duplicated *SOX9* regulatory region is insulated inside the neo-TAD and has no effect on gene regulation and no phenotypic consequence. In the Cooks syndrome associated duplications, however, incorporation of the next flanking gene, *KCNJ2*, in the neo-TAD leads to consecutive misexpression of *KCNJ2* in a *SOX9*-like expression pattern. As a consequence, these mice have a limb malformation phenocopying Cooks syndrome. Further, we validate the insulating function of TAD boundaries in regular TADs and in the neo-TAD. Loss of boundaries leads to a partial loss of insulation and ectopic interactions across TADs, which is accompanied by misregulation of genes in the adjacent TADs.

Besides giving important insights into the relationship of chromatin architecture and gene regulation, the results provide a framework for the interpretation of SVs in respect to disease. SVs are routinely detected in diagnostic tests for genetic diseases and cancer and their assessment according to changes in the TAD structure and enhancer-promoter interactions is critical for the identification of potential disease causing genes.

8 Zusammenfassung

Drei-dimensionale Interaktionen der DNA bzw. des Chromatins im Wirbeltiergenom folgen einem spezifischen Muster, welches entscheidenden Einfluss auf die genaue Genexpression hat. Messungen von Chromatin-Interaktionen mittels “Chromosome-Conformation-Capture“ (3C) haben zur Entdeckung von sogenannten “Topologically associating domains“ (TADs) geführt, die eine Größe von mehreren 100 Kilobasen erreichen können. Bereiche innerhalb einer TAD sind durch eine hohe Interaktionsfrequenz gekennzeichnet und benachbarte TADs sind von Grenzbereichen, sogenannten “Boundary-Regionen“, voneinander isoliert. TADs werden als regulatorische Einheiten im Genom betrachtet, da sie regulatorische Elemente zu ihren Zielgenen führen. Allerdings sind die Mechanismen, die zur Bildung von TADs führen und welche Rolle diese Mechanismen bei strukturellen Veränderungen im Genom spielen, unzureichend verstanden.

In der vorliegenden Arbeit wurden die Chromatin-Interaktionen am *SOX9* Lokus mit 3C Methoden (capture Hi-C und 4C-seq) im Menschen und in der Maus untersucht. Der Lokus ist in zwei TADs unterteilt, wobei eine TAD (*SOX9*-TAD) das *SOX9* Gen und die benachbarte TAD (*KCNJ*-TAD) die zwei Gene *KCNJ2* und *KCNJ16* enthält. Die Expressionsanalyse von *LacZ*-Reporter Insertionen in beiden TADs zeigte, dass die regulative Aktivität durch die TAD-Struktur begrenzt wird. Darüber hinaus ergab die Analyse von Patientenzellen und genetisch veränderten Mäusen, dass Tandem-Duplikationen die TAD-Struktur verändern können. Duplikationen können zur Bildung neuer TADs (neoTADs) führen. Neo-TADs sind funktional und räumlich von benachbarten TADs getrennt und können zur Bildung neuer regulatorischer Einheiten führen. Ihre Entstehung kann verschiedene Krankheitsbilder erklären, die durch überlappende Duplikationen beim Menschen beobachtet werden. Je nach Lage der Duplikation relativ zu *SOX9* sind diese entweder ohne phänotypische Ausprägung, verursachen eine Umkehrung des Geschlechts von weiblich zu männlich oder führen zu einer Gliedmaßenfehlbildung, dem Cooks Syndrom. Duplikationen, die Regionen innerhalb der *SOX9*-TAD betreffen (intra-TAD), verursachen Geschlechtsumwandlung. Der Effekt dieser Duplikation ist auf die *SOX9*-TAD beschränkt, wobei eine erhöhte Interaktionsfrequenz von *SOX9* mit der duplizierten Region gemessen werden kann. Im Gegensatz dazu führen größere Duplikationen, die Bereiche der *SOX9*- und *KCNJ*-TAD umfassen und somit eine Boundary-Region beinhalten (inter-TAD), zur Entstehung einer neo-TAD. Der duplizierte regulatorische Bereich ist in der neo-TAD isoliert und hat dadurch keinen Einfluss auf die Genexpression und keine phänotypische Konsequenz. Größere inter-TAD Duplikationen beinhalten *KCNJ2* und führen zu Cooks Syndrom. Die entstehende neo-TAD beinhaltet eine Kopie von *KCNJ2*, die mit regulatorischen Bereichen in der neo-TAD interagiert. Dies hat zur Folge, dass *KCNJ2* ein *SOX9*-ähnliches Expressionsmuster in den Gliedmaßen annimmt, was zu einem Cooks-ähnlichen Phänotyp in der Maus führt.

Die hier präsentierten Ergebnisse geben wichtige Einblicke in das Zusammenspiel von Chromatin-Interaktionen und Generegulation und bilden einen konzeptionellen Rahmen zur Interpretation von strukturellen Variationen im Genom. Duplikationen werden als häufige Ursache bei Erbkrankheiten und Krebs identifiziert und durch die Einbeziehung der TAD-Struktur können wesentlich genauere Aussagen über die Auswirkungen dieser Mutationen getroffen werden.

9 References

Akbari, O. S., Bae, E., Johnsen, H., Villaluz, A., Wong, D., & Drewell, R. A. (2008). A novel promoter-tethering element regulates enhancer-driven gene expression at the bithorax complex in the Drosophila embryo. *Development (Cambridge, England)*, *135*(1), 123–31.

Akiyama, H. (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. *Genes & Development*, *16*(21), 2813–2828.

Akiyama, H., Chaboissier, M.-C., Behringer, R. R., Rowitch, D. H., Schedl, A., Epstein, J. A., & de Crombrugghe, B. (2004). Essential role of Sox9 in the pathway that controls formation of cardiac valves and septa. *Proceedings of the National Academy of Sciences*, *101*(17), 6502–6507.

Amano, T., Sagai, T., Tanabe, H., Mizushina, Y., Nakazawa, H., & Shiroishi, T. (2009). Chromosomal dynamics at the Shh locus: limb bud-specific differential regulation of competence and active transcription. *Developmental Cell*, *16*(1), 47–57.

Andersen, E. D., Krasilnikoff, P. A., & Overvad, H. (1971). Intermittent muscular weakness, extrasystoles, and multiple developmental anomalies. A new syndrome? *Acta Paediatrica Scandinavica*, *60*(5), 559–64.

Anderson, R., Copeland, T. K., Schöler, H., Heasman, J., & Wylie, C. (2000). The onset of germ cell migration in the mouse embryo. *Mechanisms of Development*, *91*(1–2), 61–8.

Andrey, G., Schöpflin, R., Jerković, I., Heinrich, V., Ibrahim, D. M., Paliou, C., … Mundlos, S. (2016). Characterization of hundreds of regulatory landscapes in developing limbs reveals two regimes of chromatin folding. *Genome Research*, gr.213066.116.

Arcangeli, A., Bianchi, L., Becchetti, A., Faravelli, L., Coronnello, M., Mini, E., … Wanke, E. (1995). A novel inward-rectifying K+ current with a cell-cycle dependence governs the resting potential of mammalian neuroblastoma cells. *The Journal of Physiology*, 455–71.

Arner, E., Daub, C. O., Vitting-Seerup, K., Andersson, R., Lilje, B., Drablos, F., … Hayashizaki, Y. (2015). Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. *Science*, *347*(6225), 1010–1014.

Arnold, C. D., Gerlach, D., Stelzer, C., Boryn, L. M., Rath, M., & Stark, A. (2013). Genome-Wide Quantitative Enhancer Activity Maps Identified by STARR-seq. *Science*, *339*(6123), 1074–1077.

Artus, J., & Hadjantonakis, A.-K. (2011). Transgenic Mouse Methods and Protocols. In M. H. Hofker & J. van Deursen (Eds.) (pp. 37–56). Humana Press, 2011.

Bagheri-Fam, S., Barrionuevo, F., Dohrmann, U., Günther, T., Schüle, R., Kemler, R., … Scherer, G. (2006). Long-range upstream and downstream enhancers control distinct subsets of the complex spatiotemporal Sox9 expression pattern. *Developmental Biology*, *291*(2), 382–397.

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., … Noble, W. S. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research*, *37*(Web Server issue), W202-8.

Banerji, J., Rusconi, S., & Schaffner, W. (1981). Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell*, *27*(2 Pt 1), 299–308.

Baniahmad, A., Steiner, C., Köhne, A. C., & Renkawitz, R. (1990). Modular structure of a chicken lysozyme silencer: involvement of an unusual thyroid hormone receptor binding site. *Cell*, *61*(3), 505–14.

Baran, R., Dawber, R. P. R., & Haneke, E. (2005). Hair and nail relationship. *Skinmed*, *4*(1), 18–23.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., … Horvath, P. (2007). CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science*, *315*(5819), 1709–1712.

Barrionuevo, F. (2005). Homozygous Inactivation of Sox9 Causes Complete XY Sex Reversal in Mice. *Biology of Reproduction*, *74*(1), 195–201.

Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Schones, D. E., Wang, Z., … Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. *Cell*, *129*(4), 823–37.

Bastide, P., Darido, C., Pannequin, J., Kist, R., Robine, S., Marty-Double, C., … Jay, P. (2007). Sox9 regulates cell proliferation and is required for Paneth cell differentiation in the intestinal epithelium. *The Journal of Cell Biology*, *178*(4), 635–648.

Behringer, R., Anderson, M. D., Nagy, K. V., & Nagy, A. (1994). *Manipulating the Mouse Embryo: A Laboratory Manual, Fourth Edition* (4th Editio). Cold Spring Harbor Laboratory Press.

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

Bell, D. M., Leung, K. K. H., Wheatley, S. C., Ng, L. J., Zhou, S., Wing Ling, K., … Cheah, K. S. E. (1997). SOX9 directly regulates the type-II collagen gene. *Nature Genetics*, *16*(2), 174–178.

Benko, S., Fantes, J. A., Amiel, J., Kleinjan, D.-J., Thomas, S., Ramsay, J., … Lyonnet, S. (2009). Highly conserved non-coding elements on either side of SOX9 associated with Pierre Robin sequence. *Nature Genetics*, *41*(3), 359–64.

Benko, S., Gordon, C. T., Mallet, D., Sreenivasan, R., Thauvin-Robinet, C., Brendehaug, A., … Lyonnet, S. (2011). Disruption of a long distance regulatory region upstream of SOX9 in isolated disorders of sex development. *Journal of Medical Genetics*, *48*(12), 825–30.

Bi, W., Huang, W., Whitworth, D. J., Deng, J. M., Zhang, Z., Behringer, R. R., & de Crombrugghe, B. (2001a). Haploinsufficiency of Sox9 results in defective cartilage primordia and premature skeletal mineralization. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(12), 6698–703.

Bi, W., Huang, W., Whitworth, D. J., Deng, J. M., Zhang, Z., Behringer, R. R., & de Crombrugghe, B. (2001b). Haploinsufficiency of Sox9 results in defective cartilage primordia and premature skeletal mineralization. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(12), 6698–703.

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

Bishop, C. E., Whitworth, D. J., Qin, Y., Agoulnik, A. I., Agoulnik, I. U., Harrison, W. R., … Overbeek, P. A. (2000). A transgenic insertion upstream of sox9 is associated with dominant XX sex reversal in the mouse. *Nature Genetics*, *26*(4), 490–4.

Blow, M. J., McCulley, D. J., Li, Z., Zhang, T., Akiyama, J. A., Holt, A., … Pennacchio, L. A. (2010). ChIP-Seq identification of weakly conserved heart enhancers. *Nature Genetics*, *42*(9), 806–10.

Brunet, L. J., McMahon, J. A., McMahon, A. P., & Harland, R. M. (1998). Noggin, cartilage morphogenesis, and joint formation in the mammalian skeleton. *Science (New York, N.Y.)*, *280*(5368), 1455–7.

Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y., & Greenleaf, W. J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature Methods*, *10*(12), 1213–8.

Bulger, M., & Groudine, M. (2011). Functional and mechanistic diversity of distal transcription enhancers. *Cell*, *144*(3), 327–39.

Butler, J. E., & Kadonaga, J. T. (2001). Enhancer-promoter specificity mediated by DPE or TATA core promoter motifs. *Genes & Development*, *15*(19), 2515–9.

Calhoun, V. C., Stathopoulos, A., & Levine, M. (2002). Promoter-proximal tethering elements regulate enhancer-promoter specificity in the Drosophila Antennapedia complex. *Proceedings of the National Academy of Sciences*, *99*(14), 9243–9247.

Campbell, P. J., Stephens, P. J., Pleasance, E. D., O'Meara, S., Li, H., Santarius, T., … Futreal, P. A. (2008). Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nature Genetics*, *40*(6), 722–729.

Chan, Y. F., Marks, M. E., Jones, F. C., Villarreal, G., Shapiro, M. D., Brady, S. D., … Kingsley, D. M. (2010). Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. *Science (New York, N.Y.)*, *327*(5963), 302–5.

Cho, S. W., Kim, S., Kim, J. M., & Kim, J.-S. (2013). Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nature Biotechnology*, *31*(3), 230–232.

Cox, J. J., Willatt, L., Homfray, T., & Woods, C. G. (2011). A SOX9 duplication and familial 46,XX developmental testicular disorder. *The New England Journal of Medicine*, *364*(1), 91–3.

Craddock, N., Hurles, M. E., Cardin, N., Pearson, R. D., Plagnol, V., Robson, S., … Donnelly, P. (2010). Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature*, *464*(7289), 713–720.

Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., … Jaenisch, R. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences*, *107*(50), 21931–21936.

Cuddapah, S., Jothi, R., Schones, D. E., Roh, T.-Y., Cui, K., & Zhao, K. (2009). Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Research*, *19*(1), 24–32.

Cullen, K. E., Kladde, M. P., & Seyfred, M. A. (1993). Interaction between transcription regulatory regions of prolactin chromatin. *Science (New York, N.Y.)*, *261*(5118), 203–6.

Dahal, G. R., Rawson, J., Gassaway, B., Kwok, B., Tong, Y., Ptacek, L. J., & Bates, E. (2012). An inwardly rectifying K+ channel is required for patterning. *Development*, *139*(19), 3653–3664.

Day, T. F., Guo, X., Garrett-Beal, L., & Yang, Y. (2005). Wnt/β-Catenin Signaling in Mesenchymal Progenitors Controls Osteoblast and Chondrocyte Differentiation during Vertebrate Skeletogenesis. *Developmental Cell*, *8*(5), 739–750.

De Santa Barbara, P., Bonneaud, N., Boizet, B., Desclozeaux, M., Moniot, B., Sudbeck, P., … Berta, P. (1998). Direct interaction of SRY-related protein SOX9 and steroidogenic factor 1 regulates transcription of the human anti-Müllerian hormone gene. *Molecular and Cellular Biology*, *18*(11), 6653–65.

Deaton, A. M., & Bird, A. (2011). CpG islands and the regulation of transcription. *Genes & Development*, *25*(10), 1010–22.

de Wit, E., & de Laat, W. (2012). A decade of 3C technologies: insights into nuclear organization. *Genes & Development*, *26*(1), 11–24.

Dekker, J. (2008). Gene Regulation in the Third Dimension. *Science*, *319*(5871), 1793–1794.

Dekker, J., Marti-Renom, M. A., & Mirny, L. A. (2013). Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nature Reviews Genetics*, *14*(6), 390–403.

Dekker, J., Rippe, K., Dekker, M., & Kleckner, N. (2002). Capturing chromosome conformation. *Science (New York, N.Y.)*, *295*(5558), 1306–11.

Deng, W., Lee, J., Wang, H., Miller, J., Reik, A., Gregory, P. D., … Blobel, G. A. (2012). Controlling Long-Range Genomic Interactions at a Native Locus by Targeted Tethering of a Looping Factor. *Cell*, *149*(6), 1233–1244. 1

Diez-Roux, G., Banfi, S., Sultan, M., Geffers, L., Anand, S., Rozado, D., … Ballabio, A. (2011). A high-resolution anatomical atlas of the transcriptome in the mouse embryo. *PLoS Biology*, *9*(1), e1000582.

Dixon, J. R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J. E., Lee, A. Y., … Ren, B. (2015). Chromatin architecture reorganization during stem cell differentiation. *Nature*, *518*(7539), 331–336.

Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., … Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, *485*(7398), 376–80.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., … Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)*, *29*(1), 15–21.

Dunham, I., Kundaje, A., Aldred, S. F., Collins, P. J., Davis, C. A., Doyle, F., … Birney, E. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, *489*(7414), 57–74.

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

Durand, N. C., Robinson, J. T., Shamim, M. S., Machol, I., Mesirov, J. P., Lander, E. S., & Aiden, E. L. (2016). Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Systems*, *3*(1), 99–101.

Faure, A. J., Schmidt, D., Watt, S., Schwalie, P. C., Wilson, M. D., Xu, H., … Flicek, P. (2012). Cohesin regulates tissue-specific expression by stabilizing highly occupied cis-regulatory modules. *Genome Research*, *22*(11), 2163–2175.

Finzsch, M., Stolt, C. C., Lommes, P., & Wegner, M. (2008). Sox9 and Sox10 influence survival and migration of oligodendrocyte precursors in the spinal cord by regulating PDGF receptor expression. *Development*, *135*(4), 637–646.

Forrester, W. C., Epner, E., Driscoll, M. C., Enver, T., Brice, M., Papayannopoulou, T., & Groudine, M. (1990). A deletion of the human beta-globin locus activation region causes a major alteration in chromatin structure and replication across the entire beta-globin locus. *Genes & Development*, *4*(10), 1637–49.

Foster, J. W., Dominguez-Steglich, M. A., Guioli, S., Kwok, C., Weller, P. A., Stevanović, M., … Schafer, A. J. (1994). Campomelic dysplasia and autosomal sex reversal caused by mutations in an SRY-related gene. *Nature*, *372*(6506), 525–530.

Franke, M., Ibrahim, D. M., Andrey, G., Schwarzer, W., Heinrich, V., Schöpflin, R., … Mundlos, S. (2016). Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature*, *538*(7624), 265–269.

Frankel, N., Davis, G. K., Vargas, D., Wang, S., Payre, F., & Stern, D. L. (2010). Phenotypic robustness conferred by apparently redundant transcriptional enhancers. *Nature*, *466*(7305), 490–3.

Fraser, J., Williamson, I., Bickmore, W. A., & Dostie, J. (2015). An Overview of Genome Organization and How We Got There: from FISH to Hi-C. *Microbiology and Molecular Biology Reviews : MMBR*, *79*(3), 347–72.

Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., & Mirny, L. A. (2016). Formation of Chromosomal Domains by Loop Extrusion. *Cell Reports*, *15*(9), 2038–49.

Fukaya, T., Lim, B., & Levine, M. (2016). Enhancer Control of Transcriptional Bursting. *Cell*, *166*(2), 358–368.

Garneau, J. E., Dupuis, M.-È., Villion, M., Romero, D. A., Barrangou, R., Boyaval, P., … Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*, *468*(7320), 67–71.

Georg, I., Bagheri-Fam, S., Knower, K. C., Wieacker, P., Scherer, G., & Harley, V. R. (2010a). Mutations of the SRY-responsive enhancer of SOX9 are uncommon in XY gonadal dysgenesis. *Sexual Development : Genetics, Molecular Biology, Evolution, Endocrinology, Embryology, and Pathology of Sex Determination and Differentiation*, *4*(6), 321–5.

Georg, I., Bagheri-Fam, S., Knower, K. C., Wieacker, P., Scherer, G., & Harley, V. R. (2010b). Mutations of the SRY-responsive enhancer of SOX9 are uncommon in XY gonadal dysgenesis. *Sexual Development : Genetics, Molecular Biology, Evolution, Endocrinology, Embryology, and Pathology of Sex Determination and Differentiation*, *4*(6), 321–5.

George, S. H. L., Gertsenstein, M., Vintersten, K., Korets-Smith, E., Murphy, J., Stevens, M. E., … Nagy, A. (2007). Developmental and adult phenotyping directly from mutant embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(11), 4455–60.

Giorgetti, L., Galupa, R., Nora, E. P., Piolot, T., Lam, F., Dekker, J., … Heard, E. (2014). Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. *Cell*, *157*(4), 950–963.

Gómez-Marín, C., Tena, J. J., Acemel, R. D., López-Mayorga, M., Naranjo, S., de la Calle-Mustienes, E., … Gómez-Skarmeta, J. L. (2015). Evolutionary comparison reveals that diverging CTCF sites are signatures of ancestral topological associating domains borders. *Proceedings of the National Academy of Sciences*, *112*(24), 7542–7547.

Gonen, N., Quinn, A., O'Neill, H. C., Koopman, P., & Lovell-Badge, R. (2017). Normal Levels of Sox9 Expression in the Developing Mouse Testis Depend on the TES/TESCO Enhancer, but This Does Not Act Alone. *PLOS Genetics*, *13*(1), e1006520.

Grosveld, F., van Assendelft, G. B., Greaves, D. R., Kollias, G., Alter, B., Goff, S., … Skoultchi, A. (1987). Position-independent, high-level expression of the human β-globin gene in transgenic mice. *Cell*, *51*(6), 975–985.

GTEx Consortium. (2015). Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science (New York, N.Y.)*, *348*(6235), 648–60.

Gubbay, J., Koopman, P., Collignon, J., Burgoyne, P., & Lovell-Badge, R. (1990). Normal structure and expression of Zfy genes in XY female mice mutant in Tdy. *Development (Cambridge, England)*, *109*(3), 647–53.

Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D. U., … Wu, Q. (2015). CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell*, *162*(4), 900–910.

Hall, B., Limaye, A., & Kulkarni, A. B. (2009). Overview: generation of gene knockout mice. *Current Protocols in Cell Biology*, *Chapter 19*, Unit 19.12 19.12.1-17.

Hattori, T., Muller, C., Gebhard, S., Bauer, E., Pausch, F., Schlund, B., … von der Mark, K. (2010). SOX9 is a major negative regulator of cartilage vascularization, bone marrow formation and endochondral ossification. *Development*, *137*(6), 901–911.

Heger, P., Marin, B., Bartkuhn, M., Schierenberg, E., & Wiehe, T. (2012). The chromatin insulator CTCF and the emergence of metazoan diversity. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(43), 17507–12.

Heintzman, N. D., Stuart, R. K., Hon, G., Fu, Y., Ching, C. W., Hawkins, R. D., … Ren, B. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature Genetics*, *39*(3), 311–318.

Hérault, Y., Rassoulzadegan, M., Cuzin, F., & Duboule, D. (1998). Engineering chromosomes in mice through targeted meiotic recombination (TAMERE). *Nature Genetics*, *20*(4), 381–4.

Hooper, M., Hardy, K., Handyside, A., Hunter, S., & Monk, M. (1987). HPRT-deficient (Lesch-Nyhan) mouse embryos derived from germline colonization by cultured cells. *Nature*, *326*(6110), 292–5.

Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., … Zhang, F. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnology*, *31*(9), 827–32.

Huang, B., Wang, S., Ning, Y., Lamb, A. N., & Bartley, J. (1999). Autosomal XX sex reversal caused by duplication of SOX9. *American Journal of Medical Genetics*, *87*(4), 349–53.

Hunziker, E. B. (1988). Growth Plate Structure and Function. *Pathology and Immunopathology Research*, *7*(1–2), 9–13.

Ibn-Salem, J., Köhler, S., Love, M. I., Chung, H.-R., Huang, N., Hurles, M. E., … Robinson, P. N. (2014). Deletions of chromosomal regulatory boundaries are associated with congenital disease. *Genome Biology*, *15*(9), 423.

Ikeda, T., Kamekura, S., Mabuchi, A., Kou, I., Seki, S., Takato, T., … Chung, U. (2004). The combination of SOX5, SOX6, and SOX9 (the SOX trio) provides signals sufficient for induction of permanent cartilage. *Arthritis & Rheumatism*, *50*(11), 3561–3573.

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., & Nakata, A. (1987). Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *Journal of Bacteriology*, *169*(12), 5429–33.

Jacob, F., & Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology*, *3*, 318–56.

Jacobs, P. A., Browne, C., Gregson, N., Joyce, C., & White, H. (1992). Estimates of the frequency of chromosome abnormalities detectable in unselected newborns using moderate levels of banding. *Journal of Medical Genetics*, *29*(2), 103–8.

Jansen, R., Embden, J. D. A. van, Gaastra, W., & Schouls, L. M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular Microbiology*, *43*(6), 1565–75. R

Jin, F., Li, Y., Dixon, J. R., Selvaraj, S., Ye, Z., Lee, A. Y., … Ren, B. (2013). A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature*, *503*(7475), 290–4.

Kagey, M. H., Newman, J. J., Bilodeau, S., Zhan, Y., Orlando, D. A., van Berkum, N. L., … Young, R. A. (2010). Mediator and cohesin connect gene expression and chromatin architecture. *Nature*, *467*(7314), 430–435.

Kamachi, Y., Uchikawa, M., & Kondoh, H. (2000). Pairing SOX off: with partners in the regulation of embryonic development. *Trends in Genetics : TIG*, *16*(4), 182–7.

Kawakami, Y., Ishikawa, T., Shimabara, M., Tanda, N., Enomoto-Iwamoto, M., Iwamoto, M., … Nohno, T. (1996). BMP signaling during bone pattern determination in the developing limb. *Development (Cambridge, England)*, *122*(11), 3557–66.

Kidokoro, T., Matoba, S., Hiramatsu, R., Fujisawa, M., Kanai-Azuma, M., Taya, C., … Yonekawa, H. (2005). Influence on spatiotemporal patterns of a male-specific Sox9 activation by ectopic Sry expression during early phases of testis differentiation in mice. *Developmental Biology*, *278*(2), 511–525.

Kim, G.-J., Sock, E., Buchberger, A., Just, W., Denzer, F., Hoepffner, W., … Scherer, G. (2015). Copy number variation of two separate regulatory regions upstream of SOX9 causes isolated 46,XY or 46,XX disorder of sex development. *Journal of Medical Genetics*, *52*(4), 240–7.

Kim, Y., Kobayashi, A., Sekido, R., DiNapoli, L., Brennan, J., Chaboissier, M.-C., … Capel, B. (2006). Fgf9 and Wnt4 Act as Antagonistic Signals to Regulate Mammalian Sex Determination. *PLoS Biology*, *4*(6), e187.

Kioussis, D., Vanin, E., DeLange, T., Flavell, R. A., & Grosveld, F. G. (1983). Beta-globin gene inactivation by DNA translocation in gamma beta-thalassaemia. *Nature*, *306*(5944), 662–6.

Kleinjan, D. A., & Van Heyningen, V. (2005). Long-Range Control of Gene Expression: Emerging Mechanisms and Disruption in Disease. *Am. J. Hum. Genet*, *76*(1), 8–32.

Klopocki, E., Lohan, S., Doelken, S. C., Stricker, S., Ockeloen, C. W., Soares Thiele de Aguiar, R., … Mundlos, S. (2012). Duplications of BHLHA9 are associated with ectrodactyly and tibia hemimelia inherited in non-Mendelian fashion. *Journal of Medical Genetics*, *49*(2), 119–25.

Klopocki, E., & Mundlos, S. (2011). Copy-Number Variations, Noncoding Sequences, and Human Phenotypes. *Annual Review of Genomics and Human Genetics*.

Kortüm, F., Caputo, V., Bauer, C. K., Stella, L., Ciolfi, A., Alawi, M., … Kutsche, K. (2015). Mutations in KCNH1 and ATP6V1B2 cause Zimmermann-Laband syndrome. *Nature Genetics*, *47*(6), 661–667.

Kraft, K., Geuer, S., Will, A. J., Chan, W. L., Paliou, C., Borschiwer, M., … Andrey, G. (2015). Deletions, Inversions, Duplications: Engineering of Structural Variants using CRISPR/Cas in Mice. *Cell Reports*.

Krahl, D., & Sellheyer, K. (2010). Sox9, more than a marker of the outer root sheath: spatiotemporal expression pattern during human cutaneous embryogenesis. *Journal of Cutaneous Pathology*, *37*(3), 350–356.

Kreidberg, J. A., Sariola, H., Loring, J. M., Maeda, M., Pelletier, J., Housman, D., & Jaenisch, R. (1993). WT-1 is required for early kidney development. *Cell*, *74*(4), 679–91.

Kurth, I., Klopocki, E., Stricker, S., van Oosterwijk, J., Vanek, S., Altmann, J., … Mundlos, S. (2009). Duplications of noncoding elements 5' of SOX9 are associated with brachydactyly-anonychia. *Nature Genetics*, *41*(8), 862–3.

Kvon, E. Z., Kazmar, T., Stampfel, G., Yáñez-Cuna, J. O., Pagani, M., Schernhuber, K., … Stark, A. (2014). Genome-scale functional characterization of Drosophila developmental enhancers in vivo. *Nature*, *512*(7512), 91.

Laband, P. F., Habib, G., & Humphreys, G. S. (1964). HEREDITARY GINGIVAL FIBROMATOSIS. REPORT OF AN AFFECTED FAMILY WITH ASSOCIATED SPLENOMEGALY AND SKELETAL AND SOFT-TISSUE ABNORMALITIES. *Oral Surgery, Oral Medicine, and Oral Pathology*, *17*, 339–51.

Lallet-Daher, H., Roudbaraki, M., Bavencoffe, A., Mariot, P., Gackière, F., Bidaux, G., … Prevarskaya, N. (2009). Intermediate-conductance Ca2+-activated K+ channels (IKCa1) regulate human prostate cancer cell proliferation through a close control of calcium entry. *Oncogene*, *28*(15), 1792–1806.

Langer-Safer, P. R., Levine, M., & Ward, D. C. (1982). Immunological method for mapping genes on Drosophila polytene chromosomes. *Proceedings of the National Academy of Sciences of the United States of America*, *79*(14), 4381–5.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–9.

Lee, Y. S., Sayeed, M. M., & Wurster, R. D. (1993). Inhibition of cell growth by K+ channel modulators is due to interference with agonist-induced Ca2+ release. *Cellular Signalling*, *5*(6), 803–9.

Lefebvre, V., Huang, W., Harley, V. R., Goodfellow, P. N., & de Crombrugghe, B. (1997). SOX9 is a potent activator of the chondrocyte-specific enhancer of the pro alpha1(II) collagen gene. *Molecular and Cellular Biology*, *17*(4), 2336–46.

Lettice, L. A., Daniels, S., Sweeney, E., Venkataraman, S., Devenney, P. S., Gautier, P., … FitzPatrick, D. R. (2011). Enhancer-adoption as a mechanism of human developmental disease. *Human Mutation*, *32*(12), 1492–1499.

Lettice, L. A., Heaney, S. J. H., Purdie, L. A., Li, L., de Beer, P., Oostra, B. A., … de Graaff, E. (2003). A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Human Molecular Genetics*, *12*(14), 1725–35.

Leung, V. Y. L., Gao, B., Leung, K. K. H., Melhado, I. G., Wynn, S. L., Au, T. Y. K., … Cheah, K. S. E. (2011). SOX9 Governs Differentiation Stage-Specific Gene Expression in Growth Plate Chondrocytes via Direct Concomitant Transactivation and Repression. *PLoS Genetics*, *7*(11), e1002356.

Levine, M., & Tjian, R. (2003). Transcription regulation and animal diversity. *Nature*, *424*(6945), 147–151.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, *25*(14), 1754–60.

Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., … Dekker, J. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science (New York, N.Y.)*, *326*(5950), 289–93.

Lobanenkov, V. V, Nicolas, R. H., Adler, V. V, Paterson, H., Klenova, E. M., Polotskaja, A. V, & Goodwin, G. H. (1990). A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5'-flanking sequence of the chicken c-myc gene. *Oncogene*, *5*(12), 1743–53.

Long, F., & Ornitz, D. M. (2013). Development of the Endochondral Skeleton. *Cold Spring Harbor Perspectives in Biology*, *5*(1), a008334–a008334.

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550.

Luo, X., Ikeda, Y., & Parker, K. L. (1994). A cell-specific nuclear receptor is essential for adrenal and gonadal development and sexual differentiation. *Cell*, *77*(4), 481–90.

Lupiáñez, D. G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., … Mundlos, S. (2015). Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. *Cell*, *161*(5), 1012–1025.

Lupski, J. R., Wise, C. A., Kuwano, A., Pentao, L., Parke, J. T., Glaze, D. G., … Patel, P. I. (1992). Gene dosage is a mechanism for Charcot-Marie-Tooth disease type 1A. *Nature Genetics*, *1*(1), 29–33.

Lybæk, H., de Bruijn, D., den Engelsman-van Dijk, A. H. A., Vanichkina, D., Nepal, C., Brendehaug, A., & Houge, G. (2014). RevSex duplication-induced and sex-related differences in the SOX9 regulatory region chromatin landscape in human fibroblasts. *Epigenetics*, *9*(3), 416–427.

Lyon, M. F., Quinney, R., Glenister, P. H., Kerscher, S., Guillot, P., & Boyd, Y. (1996). Doublefoot: a new mouse mutant affecting development of limbs and head. *Genetical Research*, *68*(3), 221–31.

Maatouk, D. M., DiNapoli, L., Alvers, A., Parker, K. L., Taketo, M. M., & Capel, B. (2008). Stabilization of -catenin in XY gonads causes male-to-female sex-reversal. *Human Molecular Genetics*, *17*(19), 2949–2955.

Mackay, G. E., & West, J. D. (2005). Fate of tetraploid cells in 4n<-->2n chimeric mouse blastocysts. *Mechanisms of Development*, *122*(12), 1266–81.

Magram, J., Chada, K., & Costantini, F. (1985). Developmental regulation of a cloned adult β-globin gene in transgenic mice. *Nature*, *315*(6017), 338–340.

Malan, T. P., & McClure, W. R. (1984). Dual promoter control of the Escherichia coli lactose operon. *Cell*, *39*(1), 173–80.

Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., … Visscher, P. M. (2009). Finding the missing heritability of complex diseases. *Nature*, *461*(7265), 747–753.

Mansour, S., Hall, C. M., Pembrey, M. E., & Young, I. D. (1995). A clinical and genetic study of campomelic dysplasia. *Journal of Medical Genetics*, *32*(6), 415–20.

Maston, G. A., Evans, S. K., & Green, M. R. (2006). Transcriptional Regulatory Elements in the Human Genome. *Annual Review of Genomics and Human Genetics*, *7*(1), 29–59.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., … DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–303.

McKnight, S. L., & Kingsbury, R. (1982). Transcriptional control signals of a eukaryotic protein-coding gene. *Science (New York, N.Y.)*, *217*(4557), 316–24.

Mead, T. J., Wang, Q., Bhattaram, P., Dy, P., Afelik, S., Jensen, J., & Lefebvre, V. (2013). A far-upstream (-70 kb) enhancer mediates Sox9 auto-regulation in somatic tissues during development and adult regeneration. *Nucleic Acids Research*, *41*(8), 4459–4469.

Montavon, T., Soshnikova, N., Mascrez, B., Joye, E., Thevenet, L., Splinter, E., … Duboule, D. (2011). A Regulatory Archipelago Controls Hox Genes Transcription in Digits. *Cell*, *147*(5), 1132–1145.

Moreau, P., Hen, R., Wasylyk, B., Everett, R., Gaub, M. P., & 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 Research*, *9*(22), 6047–68.

Morgan, B. A., Conlon, F. L., Manzanares, M., Millar, J. B., Kanuga, N., Sharpe, J., … Sedgwick, S. G. (1996). Transposon tools for recombinant DNA manipulation: characterization of transcriptional regulators from yeast, Xenopus, and mouse. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(7), 2801–6.

Muerdter, F., & Stark, A. (2016). Gene Regulation: Activation through Space. *Current Biology*, *26*(19), R895–R898.

Mundlos, S., & Horn, D. (2014). Development of the Limbs. In *Limb Malformations* (pp. 2–9). Berlin, Heidelberg: Springer Berlin Heidelberg.

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

Narendra, V., Rocha, P. P., An, D., Raviram, R., Skok, J. A., Mazzoni, E. O., & Reinberg, D. (2015). CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. *Science (New York, N.Y.)*, *347*(6225), 1017–21.

Nasmyth, K., & Haering, C. H. (2009). Cohesin: its roles and mechanisms. *Annual Review of Genetics*, *43*(1), 525–58.

Neitzel, H. (1986). A routine method for the establishment of permanent growing lymphoblastoid cell lines. *Human Genetics*, *73*(4), 320–6.

Ng, L.-J., Wheatley, S., Muscat, G. E. ., Conway-Campbell, J., Bowles, J., Wright, E., … Koopman, P. (1997). SOX9 Binds DNA, Activates Transcription, and Coexpresses with Type II Collagen during Chondrogenesis in the Mouse. *Developmental Biology*, *183*(1), 108–121.

Nora, E. P., Goloborodko, A., Valton, A.-L., Gibcus, J., Uebersohn, A., Abdennur, N., … Bruneau, B. (2016). Targeted degradation of CTCF decouples local insulation of chromosome domains from higher-order genomic compartmentalization. *bioRxiv*.

Nora, E. P., Lajoie, B. R., Schulz, E. G., Giorgetti, L., Okamoto, I., Servant, N., … Heard, E. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*, *485*(7398), 381–5.

Nowak, J. A., Polak, L., Pasolli, H. A., & Fuchs, E. (2008). Hair Follicle Stem Cells Are Specified and Function in Early Skin Morphogenesis. *Cell Stem Cell*, *3*(1), 33–43.

Ogbourne, S., & Antalis, T. M. (1998). Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes. *The Biochemical Journal*, (Pt 1), 1–14.

Ohno, S. (1970). *Evolution by Gene Duplication*. Berlin, Heidelberg: Springer Berlin Heidelberg.

Ong, C.-T., & Corces, V. G. (2009). Insulators as mediators of intra- and inter-chromosomal interactions: a common evolutionary theme. *Journal of Biology*, *8*(8), 73.

Ong, C.-T., & Corces, V. G. (2014). CTCF: an architectural protein bridging genome topology and function. *Nature Reviews Genetics*, *15*(4), 234–246.

Pant, V., Kurukuti, S., Pugacheva, E., Shamsuddin, S., Mariano, P., Renkawitz, R., … Ohlsson, R. (2004). Mutation of a single CTCF target site within the H19 imprinting control region leads to loss of Igf2 imprinting and complex patterns of de novo methylation upon maternal inheritance. *Molecular and Cellular Biology*, *24*(8), 3497–504.

Pardo, L. A., Brüggemann, A., Camacho, J., & Stühmer, W. (1998). Cell cycle-related changes in the conducting properties of r-eag K+ channels. *The Journal of Cell Biology*, *143*(3), 767–75.

Perry, G. H., Dominy, N. J., Claw, K. G., Lee, A. S., Fiegler, H., Redon, R., … Stone, A. C. (2007). Diet and the evolution of human amylase gene copy number variation. *Nature Genetics*, *39*(10), 1256–1260.

Perry, M. W., Boettiger, A. N., Bothma, J. P., & Levine, M. (2010). Shadow enhancers foster robustness of Drosophila gastrulation. *Current Biology : CB*, *20*(17), 1562–7.

Pfeifer, D., Kist, R., Dewar, K., Devon, K., Lander, E. S., Birren, B., … Scherer, G. (1999). Campomelic dysplasia translocation breakpoints are scattered over 1 Mb proximal to SOX9: evidence for an extended control region. *American Journal of Human Genetics*, *65*(1), 111–24.

Phillips-Cremins, J. E., Sauria, M. E. G., Sanyal, A., Gerasimova, T. I., Lajoie, B. R., Bell, J. S. K., … Corces, V. G. (2013). Architectural Protein Subclasses Shape 3D Organization of Genomes during Lineage Commitment. *Cell*, *153*(6), 1281–1295.

Pinto, D., Pagnamenta, A. T., Klei, L., Anney, R., Merico, D., Regan, R., … Betancur, C. (2010). Functional impact of global rare copy number variation in autism spectrum disorders. *Nature*, *466*(7304), 368–372.

Pizette, S., & Niswander, L. (2000). BMPs Are Required at Two Steps of Limb Chondrogenesis: Formation of Prechondrogenic Condensations and Their Differentiation into Chondrocytes. *Developmental Biology*, *219*(2), 237–249.

Plaster, N. M., Tawil, R., Tristani-Firouzi, M., Canún, S., Bendahhou, S., Tsunoda, A., … Ptácek, L. J. (2001). Mutations in Kir2.1 cause the developmental and episodic electrical phenotypes of Andersen's syndrome. *Cell*, *105*(4), 511–9.

Pop, R., Conz, C., Lindenberg, K. S., Blesson, S., Schmalenberger, B., Briault, S., … Scherer, G. (2004). Screening of the 1 Mb SOX9 5' control region by array CGH identifies a large deletion in a case of campomelic dysplasia with XY sex reversal. *Journal of Medical Genetics*, *41*(4), e47.

Rao, S. S. P., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., … Aiden, E. L. (2014). A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell*, *159*(7), 1665–80.

Rausch, T., Zichner, T., Schlattl, A., Stutz, A. M., Benes, V., & Korbel, J. O. (2012). DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics*, *28*(18), i333–i339.

Redon, R., Ishikawa, S., Fitch, K. R., Feuk, L., Perry, G. H., Andrews, T. D., … Hurles, M. E. (2006). Global variation in copy number in the human genome. *Nature*, *444*(7118), 444–454.

Rice, A. M., & McLysaght, A. (2017). Dosage sensitivity is a major determinant of human copy number variant pathogenicity. *Nature Communications*, *8*, 14366.

Rice, M. C., & O'Brien, S. J. (1980). Genetic variance of laboratory outbred Swiss mice. *Nature*, *283*(5743), 157–61.

Riethoven, J.-J. M. (2010). Regulatory Regions in DNA: Promoters, Enhancers, Silencers, and Insulators. In *Methods in molecular biology (Clifton, N.J.)* (Vol. 674, pp. 33–42).

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

Rodríguez, C. I., Buchholz, F., Galloway, J., Sequerra, R., Kasper, J., Ayala, R., … Dymecki, S. M. (2000). High-efficiency deleter mice show that FLPe is an alternative to Cre-loxP. *Nature Genetics*, *25*(2), 139–40.

Roessler, E., Belloni, E., Gaudenz, K., Jay, P., Berta, P., Scherer, S. W., … Muenke, M. (1996). Mutations in the human Sonic Hedgehog gene cause holoprosencephaly. *Nature Genetics*, *14*(3), 357–360.

Rudnicki, J. A., & Brown, A. M. C. (1997). Inhibition of Chondrogenesis byWntGene Expressionin Vivoandin Vitro. *Developmental Biology*, *185*(1), 104–118.

Ruf, S., Symmons, O., Uslu, V. V., Dolle, D., Hot, C., Ettwiller, L., & Spitz, F. (2011). Large-scale analysis of the regulatory architecture of the mouse genome with a transposon-associated sensor. *Nature Genetics*, *43*(4), 379–86.

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

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

Sanborn, A. L., Rao, S. S. P., Huang, S.-C., Durand, N. C., Huntley, M. H., Jewett, A. I., … Aiden, E. L. (2015). Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(47), E6456-65.

Saxonov, S., Berg, P., & Brutlag, D. L. (2006). A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(5), 1412–7.

Schmidt, D., Schwalie, P. C., Ross-Innes, C. S., Hurtado, A., Brown, G. D., Carroll, J. S., … Odom, D. T. (2010). A CTCF-independent role for cohesin in tissue-specific transcription. *Genome Research*, *20*(5), 578–588.

Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., … Wigler, M. (2004). Large-Scale Copy Number Polymorphism in the Human Genome. *Science*, *305*(5683), 525–528.

Sekido, R., Bar, I., Narváez, V., Penny, G., & Lovell-Badge, R. (2004). SOX9 is up-regulated by the transient expression of SRY specifically in Sertoli cell precursors. *Developmental Biology*, *274*(2), 271–279.

Sekido, R., & Lovell-Badge, R. (2008). Sex determination involves synergistic action of SRY and SF1 on a specific Sox9 enhancer. *Nature*, *453*(7197), 930–934.

Sekido, R., & Lovell-Badge, R. (2009). Sex determination and SRY: down to a wink and a nudge? *Trends in Genetics*, *25*(1), 19–29.

Sfeir, A., & Symington, L. S. (2015). Microhomology-Mediated End Joining: A Back-up Survival Mechanism or Dedicated Pathway? *Trends in Biochemical Sciences*, *40*(11), 701–14.

Shapiro, M. D., Marks, M. E., Peichel, C. L., Blackman, B. K., Nereng, K. S., Jónsson, B., … Kingsley, D. M. (2004). Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature*, *428*(6984), 717–723.

Shen, Y., Yue, F., McCleary, D. F., Ye, Z., Edsall, L., Kuan, S., … Ren, B. (2012). A map of the cis-regulatory sequences in the mouse genome. *Nature*, *488*(7409), 116–120.

Simons, C., Rash, L. D., Crawford, J., Ma, L., Cristofori-Armstrong, B., Miller, D., … Taft, R. J. (2014). Mutations in the voltage-gated potassium channel gene KCNH1 cause Temple-Baraitser syndrome and epilepsy. *Nature Genetics*, *47*(1), 73–77.

Sinclair, A. H., Berta, P., Palmer, M. S., Hawkins, J. R., Griffiths, B. L., Smith, M. J., … Goodfellow, P. N. (1990). A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature*, *346*(6281), 240–244.

Smith, E., & Shilatifard, A. (2014). Enhancer biology and enhanceropathies. *Nature Structural & Molecular Biology*, *21*(3), 210–219.

Spitz, F. (2016). Gene regulation at a distance: From remote enhancers to 3D regulatory ensembles. *Seminars in Cell & Developmental Biology*, *57*, 57–67.

Spitz, F., & Furlong, E. E. M. (2012). Transcription factors: from enhancer binding to developmental control. *Nature Reviews. Genetics*, *13*(9), 613–26.

Splinter, E., de Wit, E., van de Werken, H. J. G., Klous, P., & de Laat, W. (2012). Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: from fixation to computation. *Methods (San Diego, Calif.)*, *58*(3), 221–30.

Splinter, E., Heath, H., Kooren, J., Palstra, R.-J., Klous, P., Grosveld, F., … de Laat, W. (2006). CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes & Development*, *20*(17), 2349–2354.

Stricker, S., Fundele, R., Vortkamp, A., & Mundlos, S. (2002). Role of Runx Genes in Chondrocyte Differentiation. *Developmental Biology*, *245*(1), 95–108.

Stricker, S., & Mundlos, S. (2011). Mechanisms of digit formation: Human malformation syndromes tell the story. *Developmental Dynamics*, *240*(5), 990–1004.

Svingen, T., & Koopman, P. (2013). Building the mammalian testis: origins, differentiation, and assembly of the component cell populations. *Genes & Development*, *27*(22), 2409–2426.

Symmons, O., Uslu, V. V., Tsujimura, T., Ruf, S., Nassari, S., Schwarzer, W., … Spitz, F. (2014). Functional and topological characteristics of mammalian regulatory domains. *Genome Research*, *24*(3), 390–400.

Tang, S.-H. E., Silva, F. J., Tsark, W. M. K., & Mann, J. R. (2002). A Cre/loxP-deleter transgenic line in mouse strain 129S1/SvImJ. *Genesis (New York, N.Y. : 2000)*, *32*(3), 199–202.

Tawil, R., Ptacek, L. J., Pavlakis, S. G., DeVivo, D. C., Penn, A. S., Özdemir, C., & Griggs, R. C. (1994). Andersen's syndrome: Potassium-sensitive periodic paralysis, ventricular ectopy, and dysmorphic features. *Annals of Neurology*, *35*(3), 326–330.

Temple, I. K., & Baraitser, M. (1991). Severe mental retardation and absent nails of Hallux and Pollex. *American Journal of Medical Genetics*, *41*(2), 173–175.

ten Berge, D., Brugmann, S. A., Helms, J. A., & Nusse, R. (2008). Wnt and FGF signals interact to coordinate growth with cell fate specification during limb development. *Development*, *135*(19), 3247–3257.

The International HapMap Consortium, T. I. H. (2005). A haplotype map of the human genome. *Nature*, *437*(7063), 1299–1320.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., … Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols*, *7*(3), 562–78.

Tristani-Firouzi, M., Jensen, J. L., Donaldson, M. R., Sansone, V., Meola, G., Hahn, A., … Tawil, R. (2002). Functional and clinical characterization of KCNJ2 mutations associated with LQT7 (Andersen syndrome). *The Journal of Clinical Investigation*, *110*(3), 381–8.

Tsai, E. Y., Yie, J., Thanos, D., & Goldfeld, A. E. (1996). Cell-type-specific regulation of the human tumor necrosis factor alpha gene in B cells and T cells by NFATp and ATF-2/JUN. *Molecular and Cellular Biology*, *16*(10), 5232–44.

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

Tucker, K. L., Wang, Y., Dausman, J., & Jaenisch, R. (1997). A transgenic mouse strain expressing four drug-selectable marker genes. *Nucleic Acids Research*, *25*(18), 3745–6.

Urrego, D., Tomczak, A. P., Zahed, F., Stühmer, W., & Pardo, L. A. (2014). Potassium channels in cell cycle and cell proliferation. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *369*(1638), 20130094.

Uslu, V. V., Petretich, M., Ruf, S., Langenfeld, K., Fonseca, N. A., Marioni, J. C., & Spitz, F. (2014). Long-range enhancers regulating Myc expression are required for normal facial morphogenesis. *Nature Genetics*, *46*(7), 753–8.

van de Werken, H. J. G., de Vree, P. J. P., Splinter, E., Holwerda, S. J. B., Klous, P., de Wit, E., & de Laat, W. (2012). 4C technology: protocols and data analysis. *Methods in Enzymology*, *513*, 89–112.

van de Werken, H. J. G., Landan, G., Holwerda, S. J. B., Hoichman, M., Klous, P., Chachik, R., … de Laat, W. (2012). Robust 4C-seq data analysis to screen for regulatory DNA interactions. *Nature Methods*, *9*(10), 969–72.

Vidal, V. P. I., Chaboissier, M.-C., Lützkendorf, S., Cotsarelis, G., Mill, P., Hui, C.-C., … Schedl, A. (2005). Sox9 Is Essential for Outer Root Sheath Differentiation and the Formation of the Hair Stem Cell Compartment. *Current Biology*, *15*(15), 1340–1351.

Vietri Rudan, M., Barrington, C., Henderson, S., Ernst, C., Odom, D. T., Tanay, A., & Hadjur, S. (2015). Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell Reports*, *10*(8), 1297–309.

Villar, D., Berthelot, C., Aldridge, S., Rayner, T. F., Lukk, M., Pignatelli, M., … Odom, D. T. (2015). Enhancer Evolution across 20 Mammalian Species. *Cell*, *160*(3), 554–566.

Visel, A., Blow, M. J., Li, Z., Zhang, T., Akiyama, J. A., Holt, A., … Pennacchio, L. A. (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature*, *457*(7231), 854–8.

Visel, A., Prabhakar, S., Akiyama, J. A., Shoukry, M., Lewis, K. D., Holt, A., … Pennacchio, L. A. (2008). Ultraconservation identifies a small subset of extremely constrained developmental enhancers. *Nature Genetics*, *40*(2), 158–60.

Vo, N., & Goodman, R. H. (2001). CREB-binding Protein and p300 in Transcriptional Regulation. *Journal of Biological Chemistry*, *276*(17), 13505–13508.

Vu, T. H., Shipley, J. M., Bergers, G., Berger, J. E., Helms, J. A., Hanahan, D., … Werb, Z. (1998). MMP-9/gelatinase B is a key regulator of growth plate angiogenesis and apoptosis of hypertrophic chondrocytes. *Cell*, *93*(3), 411–22.

Wagner, T., Wirth, J., Meyer, J., Zabel, B., Held, M., Zimmer, J., … Scherer, G. (1994). Autosomal sex reversal and campomelic dysplasia are caused by mutations in and around the SRY-related gene SOX9. *Cell*, *79*(6), 1111–20.

Wang, H., Maurano, M. T., Qu, H., Varley, K. E., Gertz, J., Pauli, F., … Stamatoyannopoulos, J. A. (2012). Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Research*, *22*(9), 1680–8.

Warman, M. L., Gong, Y., Krakow, D., Marcelino, J., Wilkin, D., Chitayat, D., … Francomano, C. A. (1999). Heterozygous mutations in the gene encoding noggin affect human joint morphogenesis. *Nature Genetics*, *21*(3), 302–304.

Wassarman, P. M., & Soriano, P. M. (2010). Guide to techniques in mouse development. Preface. *Methods in Enzymology*, *476*, xix.

Wegner, M. (1999). From head to toes: the multiple facets of Sox proteins. *Nucleic Acids Research*, *27*(6), 1409–20.

Weischenfeldt, J., Dubash, T., Drainas, A. P., Mardin, B. R., Chen, Y., Stütz, A. M., … Korbel, J. O. (2016). Pan-cancer analysis of somatic copy-number alterations implicates IRS4 and IGF2 in enhancer hijacking. *Nature Genetics*, *49*(1), 65–74.

Wheater, P. R., Burkitt, H. G., & Daniels, V. G. (1979). *Functional histology : a text and colour atlas*. Churchill Livingstone.

Wilhelm, D. (2002). The Wilms tumor suppressor WT1 regulates early gonad development by activation of Sf1. *Genes & Development*, *16*(14), 1839–1851.

Wilhelm, D., Hiramatsu, R., Mizusaki, H., Widjaja, L., Combes, A. N., Kanai, Y., & Koopman, P. (2007). SOX9 Regulates Prostaglandin D Synthase Gene Transcription in Vivo to Ensure Testis Development. *Journal of Biological Chemistry*, *282*(14), 10553–10560.

Wingett, S., Ewels, P., Furlan-Magaril, M., Nagano, T., Schoenfelder, S., Fraser, P., & Andrews, S. (2015). HiCUP: pipeline for mapping and processing Hi-C data. *F1000Research*, *4*, 1310.

Wright, E., Hargrave, M. R., Christiansen, J., Cooper, L., Kun, J., Evans, T., … Koopman, P. (1995). The Sry-related gene Sox9 is expressed during chondrogenesis in mouse embryos. *Nature Genetics*, *9*(1), 15–20.

Wu, S., Ying, G., Wu, Q., & Capecchi, M. R. (2007). Toward simpler and faster genome-wide mutagenesis in mice. *Nature Genetics*, *39*(7), 922–30.

Xiao, A., Wang, Z., Hu, Y., Wu, Y., Luo, Z., Yang, Z., … Zhang, B. (2013). Chromosomal deletions and inversions mediated by TALENs and CRISPR/Cas in zebrafish. *Nucleic Acids Research*, *41*(14), e141–e141.

Young, M. D., Willson, T. A., Wakefield, M. J., Trounson, E., Hilton, D. J., Blewitt, M. E., … Majewski, I. J. (2011). ChIP-seq analysis reveals distinct H3K27me3 profiles that correlate with transcriptional activity. *Nucleic Acids Research*, *39*(17), 7415–27.

Zabidi, M. A., Arnold, C. D., Schernhuber, K., Pagani, M., Rath, M., Frank, O., & Stark, A. (2015). Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. *Nature*, *518*(7540), 556–9.

Zaias, N. (1963). Embryology of the Human Nail. *Archives of Dermatology*, *87*(1), 37.

Zaritsky, J. J., Eckman, D. M., Wellman, G. C., Nelson, M. T., & Schwarz, T. L. (2000). Targeted disruption of Kir2.1 and Kir2.2 genes reveals the essential role of the inwardly rectifying K(+) current in K(+)-mediated vasodilation. *Circulation Research*, *87*(2), 160–6.

Zarrei, M., MacDonald, J. R., Merico, D., & Scherer, S. W. (2015). A copy number variation map of the human genome. *Nature Reviews Genetics*, *16*(3), 172–183.

Zhang, J., Bridges, C. B., Stephens, S. G., Ohno, S., Nei, M., Ohno, S., … al., et. (2003). Evolution by gene duplication: an update. *Trends in Ecology & Evolution*, *18*(6), 292–298.

Zhang, X., Choi, P. S., Francis, J. M., Imielinski, M., Watanabe, H., Cherniack, A. D., & Meyerson, M. (2016). Identification of focally amplified lineage-specific super-enhancers in human epithelial cancers. *Nat Genet*, *48*(2), 176–182.

Zhao, Z., Tavoosidana, G., Sjölinder, M., Göndör, A., Mariano, P., Wang, S., … Ohlsson, R. (2006). Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nature Genetics*, *38*(11), 1341–7.

Zhou, G., Zheng, Q., Engin, F., Munivez, E., Chen, Y., Sebald, E., … Lee, B. (2006). Dominance of SOX9 function over RUNX2 during skeletogenesis. *Proceedings of the National Academy of Sciences*, *103*(50), 19004–19009.

Zhou, X., Lowdon, R. F., Li, D., Lawson, H. A., Madden, P. A. F., Costello, J. F., & Wang, T. (2013). Exploring long-range genome interactions using the WashU Epigenome Browser. *Nature Methods*, *10*(5), 375–6.

Zou, H., Wieser, R., Massagué, J., & Niswander, L. (1997). Distinct roles of type I bone morphogenetic protein receptors in the formation and differentiation of cartilage. *Genes & Development*, *11*(17), 2191–203.

Zuin, J., Dixon, J. R., van der Reijden, M. I. J. A., Ye, Z., Kolovos, P., Brouwer, R. W. W., … Wendt, K. S. (2014). Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proceedings of the National Academy of Sciences*, *111*(3), 996–1001.

10 Appendix

10.1 Supplementary figures

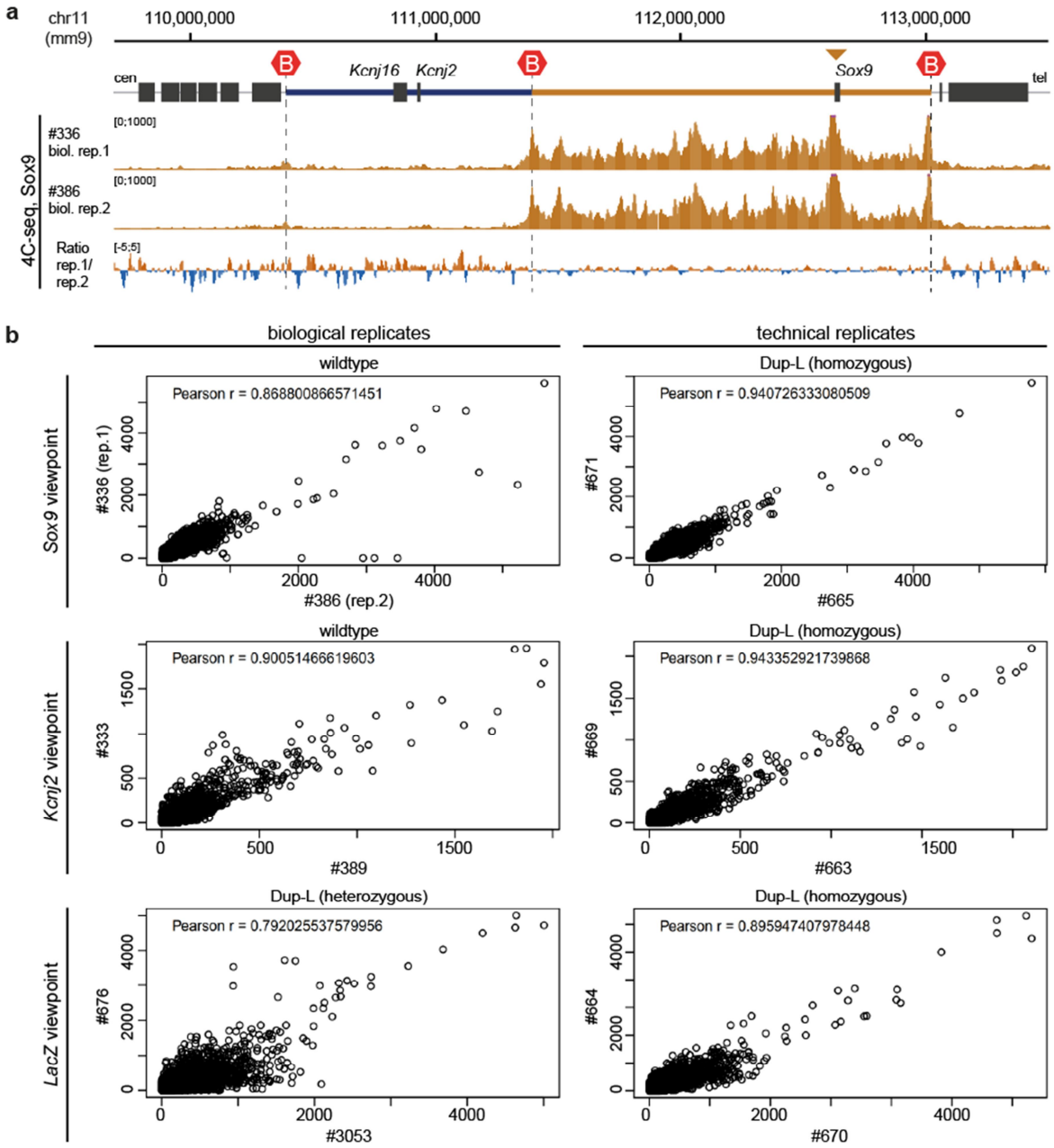


Figure 34: Comparison of biological and technical 4C-seq replicates. A, Schematic of wild-type *Sox9* locus, depicting TAD boundaries (red hexagons), *Kcnj* (blue) and *Sox9* (brown) TAD. Below, 4C-seq profiles from two biological replicates from wild-type limb buds (E12.5) with viewpoint (triangle) in *Sox9* promoter. The corresponding log2 ratio shows no major differences in the 4C-seq profile within the *Sox9* TAD. The alternating pattern of high and low interaction frequencies between the samples reflects the low base line signal and low interaction frequency outside of *Sox9* TAD. B, Pearson correlation of biological (left) and technical replicates (right) with viewpoints in *Sox9*, *Kcnj2* and *LacZ* promoter from E12.5 limb buds. Biological replicates were generated from different 4C libraries whereas technical replicates were generated from the same 4C library but different 4C-inverse PCRs. Note, technical replicates show slightly higher correlation than biological replicates, indicating 4C library preparation mainly contribute to variations.

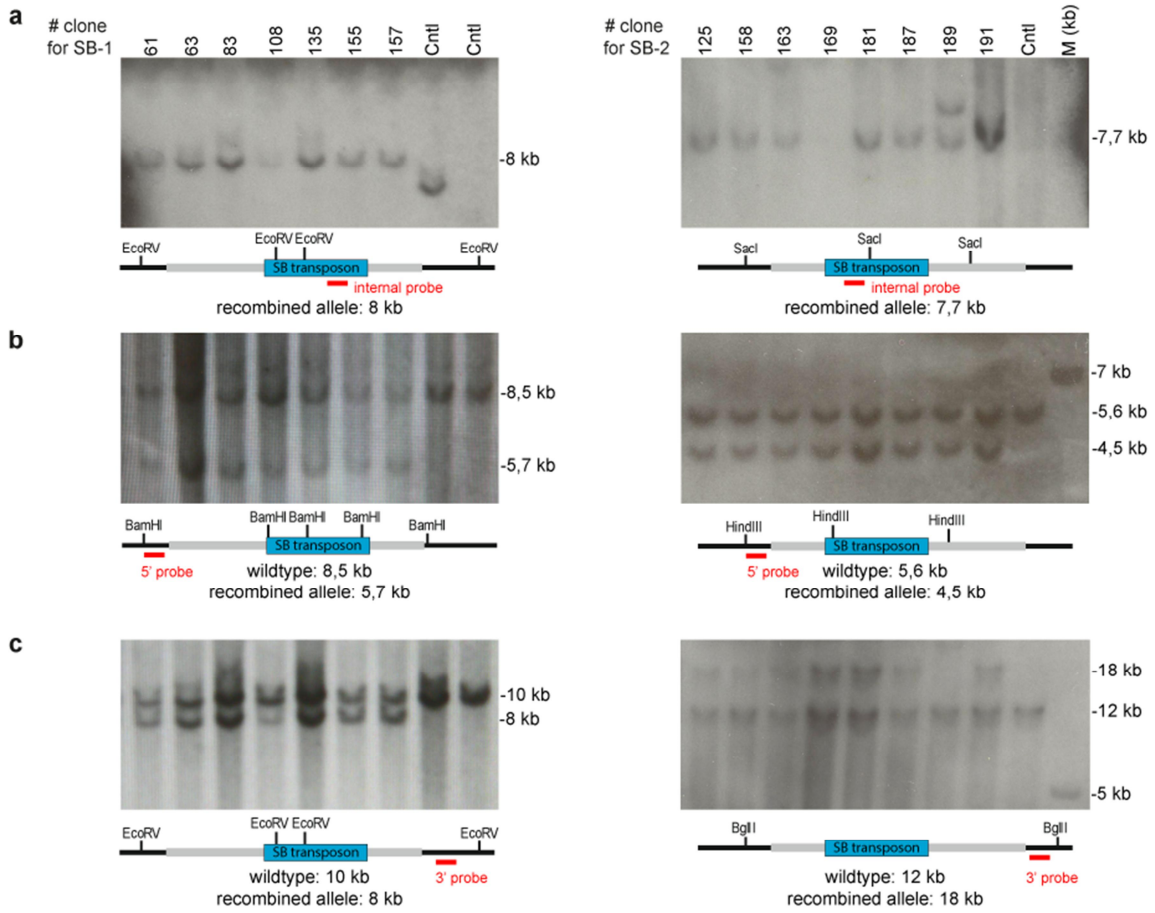


Figure 35: Southern blot analysis SB-1 and SB-2 targeted alleles in ESCs. Southern blot analyses of selected ESC clones for SB-1 and SB-2 alleles. Below each blot, schematic of the SB transgene (blue) with regions of homology (grey) and surrounding non-homology regions (black) at the integration site is shown. Restriction enzymes and DNA probes (red) used for the analysis of restriction fragment polymorphisms are indicated. Homologous recombination was verified by an internal probe hybridizing to the transgene (a), 5' external probe (b) and 3' external probe (c) hybridizing to genomic regions outside the cloned homology region. Sizes of restriction fragments for the wild-type and recombined alleles are indicated.

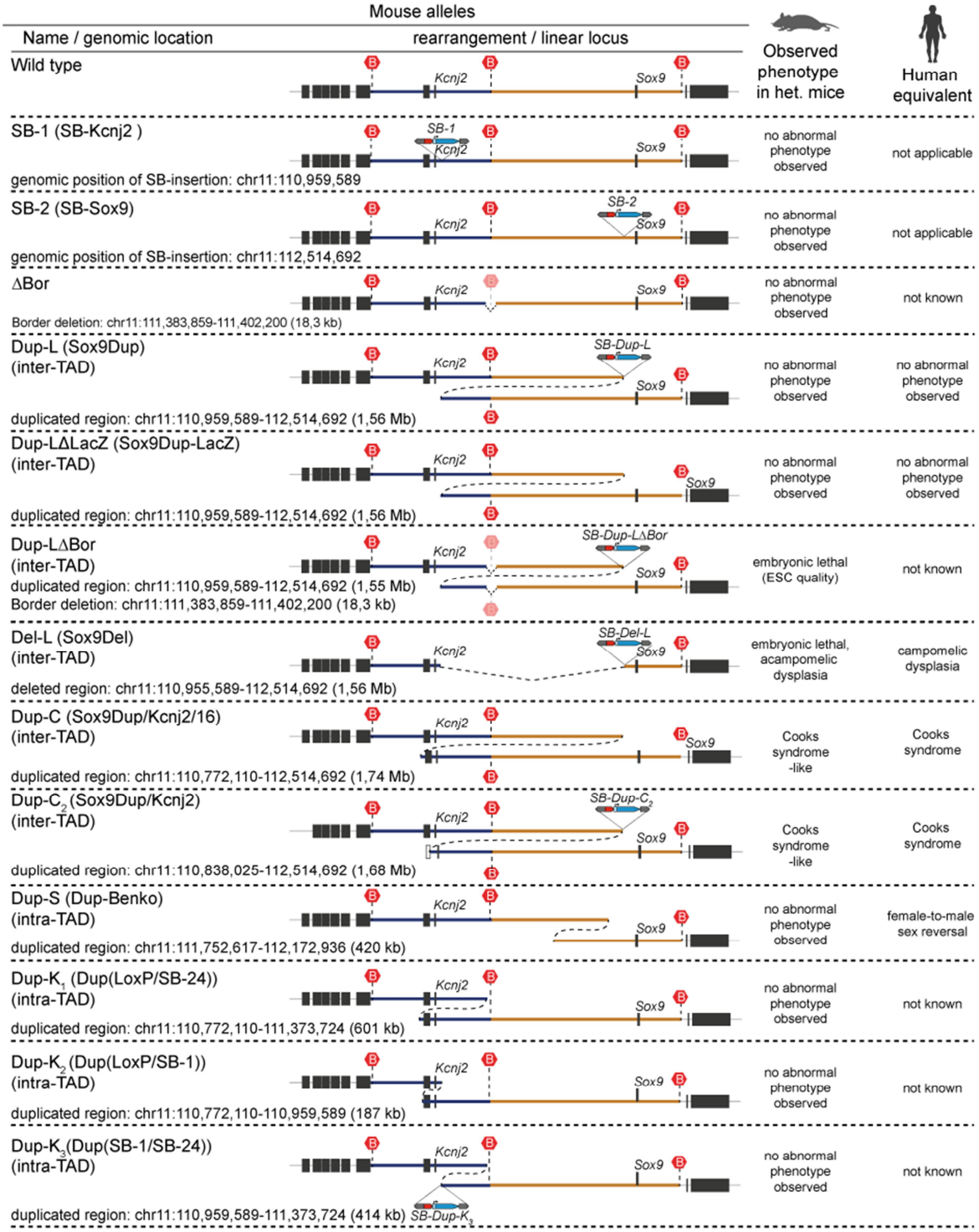
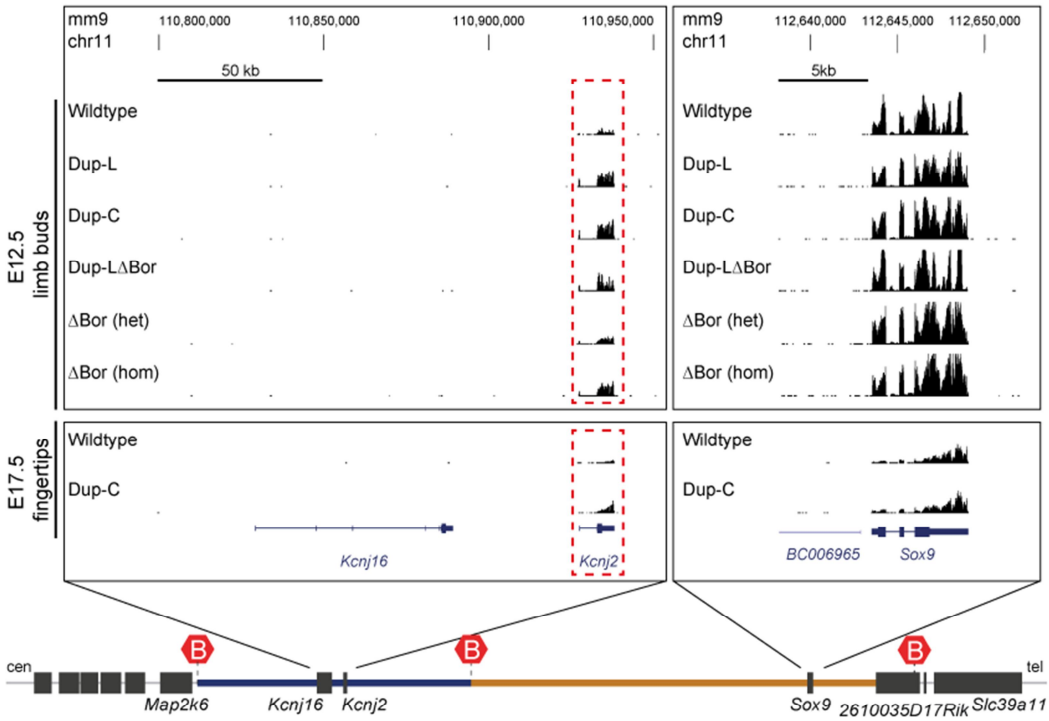


Figure 36: Mouse alleles generated in this study. Representations of mouse alleles depict TAD boundaries (red hexagons), *Kcnj* (blue) and *Sox9* (brown) TAD structure. The extent of the duplicated or deleted region is given as genomic coordinates (mm9) and is visualized by overlap for duplications and dotted lines for deletions. The position of the SB transgene at the breakpoint is visualized, depicting *loxP* site with orientation (red); *LacZ* reporter driven by *beta-globin* promoter (blue) and SB (Sleeping Beauty) transposons (grey). The observed mouse phenotypes and equivalent syndromes in humans are listed on the right side. For generation of mouse alleles see Section 4.7.2. Table 11 and Table 12 provide further information about donor mouse strains used for trans-allelic recombination.

a



b

Map2k6

Allele	mean FPKM	log2 FC vs. Wt	log2FC SE	P (adj)
Wildtype	19.76			
Dup-L	21.04	0.10	0.12	0.539
Dup-C	23.80	0.28	0.12	0.162
ΔBor (het)	22.00	0.16	0.12	0.369
ΔBor (hom)	23.93	0.32	0.15	0.125
Dup-LΔBor	21.07	0.12	0.10	0.488

Sox9

Allele	mean FPKM	log2 FC vs. Wt	log2FC SE	P (adj)
Wildtype	31.39			
Dup-L	28.75	-0.10	0.14	0.617
Dup-C	33.19	0.11	0.14	0.673
ΔBor (het)	39.35	0.32	0.14	0.102
ΔBor (hom)	37.66	0.29	0.17	0.214
Dup-LΔBor	31.91	0.06	0.12	0.816

Kcnj2

Allele	mean FPKM	log2 FC vs. Wt	log2FC SE	P (adj)
Wildtype	3.59			
Dup-L	4.70	0.41	0.17	0.047
Dup-C	**13.06**	**1.69**	**0.17**	**$2.00*10^{-18}$**
ΔBor (het)	**6.05**	**0.72**	**0.17**	**$2.23*10^{-03}$**
ΔBor (hom)	**9.84**	**1.31**	**0.20**	**$1.19*10^{-07}$**
Dup-LΔBor	**9.09**	**1.26**	**0.15**	**$2.34*10^{-14}$**

2610035D17Rik

Allele	mean FPKM	log2 FC vs. Wt	log2FC SE	P (adj)
Wildtype	6.93			
Dup-L	9.29	0.40	0.16	0.028
Dup-C	9.80	0.49	0.16	0.088
ΔBor (het)	6.05	-0.14	0.16	0.576
ΔBor (hom)	7.11	0.12	0.19	0.663
Dup-LΔBor	9.95	0.53	0.13	0.003

Slc39a11

Allele	mean FPKM	log2 FC vs. Wt	log2FC SE	P (adj)
Wildtype	4.65			
Dup-L	4.61	0.00	0.11	0.991
Dup-C	4.49	-0.02	0.11	0.937
ΔBor (het)	5.10	0.14	0.11	0.376
ΔBor (hom)	4.50	0.02	0.14	0.925
Dup-LΔBor	4.06	-0.16	0.09	0.253

Figure 37: Summary of RNA-seq read profiles and expression values from mouse mutants used in this study. a, Read profiles for *Kcnj16*, *Kcnj2,* and *Sox9* from RNA-seq analysis in wild-type and mouse mutant E12.5 limb buds and E17.5 finger tips (lower two tracks). b, Summary of expression values for transcripts indicated in the schematic at the extended *Sox9* locus from E12.5 limb buds. Note that *Kcnj2* is the only gene showing altered expression in Dup-C, ΔBor, and Dup-LΔBor mice. Significant expression changes are indicated in bold. FPKM, fragments per kilo base of exon per million fragments mapped. Benjamin-Hochberg-adjusted *P* value, n=2, cut-off=0.001.

SNP ID	Whole-Genome Seq.		4C-seq *SOX9*		4C-seq *KCNJ2*		4C-seq breakpoint	
	Genotype	maj./min. allele freq.	Genotype	maj./min. allele freq.	Genotype	maj./min. allele freq.	Genotype	maj./min. allele freq.
rs11077494	G/T	0.79 / 0.21	-	- / -	G	1.0 / -	T	1.0 / 0.0
rs1014217	G/A	0.70 / 0.30	G	0.94 / 0.06	-	- / -	A	1.0 / -
rs17767038	C/T	0.77 / 0.23	C	1.0 / -	-	- / -	T	1.0 / -
rs12450694	T/A	0.54 / 0.46	T	1.0 / -	-	- / -	A	1.0 / -
rs12450214	A/T	0.54 / 0.46	A	1.0 / -	-	- / -	T	1.0 / -

Figure 38: Allele frequencies determined in 4C-seq and whole-genome sequencing in a patient with inter-TAD duplication (normal phenotype). Values are related to Figure 29. Identification number (ID) of SNPs identified in the duplication. Table lists SNP genotype together with major and minor allele frequency from whole-genome sequencing (WGS) and 4C-seq with viewpoints in *SOX9*, *KCNJ2,* and at the duplication breakpoint located within the neo-TAD. Note that SNPs in WGS data are captured with 2 to 1 of major and minor frequency, indicating the presence of three alleles. The breakpoint viewpoint only captures the genotype with the minor frequency, whereas the *SOX9* and *KCNJ2* the genotype with the major frequency.

10.2 List of abbreviations

°C	degree Celsius
µ	micro (prefix)
3C	chromosome conformation capture
4C	circular chromosome conformation capture
AER	apical ectodermal ridge
bp	base pairs
CD	campomelic dysplasia
cDNA	coding DNA
cen	centromeric
cHi-C	capture Hi-C
ChIP	chromatin immunoprecipitation
chr	chromosome
CRE	*cis*-regulatory element
CTCF	CCCTC-binding factor
del	deletion
DEPC	diethylpyrocarbonate
DIG	digoxygenin
DMEM	Dulbeccos's modified eagle's medium
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide
dup	duplication
E	embryonic stage
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
ENCODE	Encyclopedia of DNA Elements
ES	embryonic stem
ESC	embryonic stem cell
FL	forelimb
g	gram
gDNA	genomic DNA
h	hour
het	heterozygous
hg	human genome
hom	homozygous
hybe	hybridization
kb	kilo base
l	liter
m	mili (prefix)
M	molar
Mb	mega base
min	minute(s)
mm	*Mus musculus*
mol	moles
mRNA	messenger RNA
n	nano (prefix)
P	postnatal day
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PFA	paraformaldehyde
PFM	enlarged parietal foramina

Pfu	Pyrococcus furiosus
qPCR	quantitative PCR
RNA	ribonucleic acid
rpm	rounds per minute
RT	reverse transcription
SB	Sleeping Beauty
SNP	single nucleotide polymorphism
SRY	Sex determining region of Y-Gen
SSC	saline sodium citrate buffer
TAD	topologically associating domain
TAMERE	trans-allelic targeted meiotic recombination
Taq	Thermus aquaticus
tel	telomeric
TF	transcription factor
Tm	melting temperature
TSS	transcription start site
U	units
UCSC	University of California, Santa Cruz
Vol	volume
v/v	volume per volume
w/v	weight per volume
WGS	whole-genome sequencing
wt	wild type

10.3 List of tables

10.4 List of figures

11 Danksagung

Mein Dank gilt in erster Linie Prof. Dr. Stefan Mundlos für die Möglichkeit meine Doktorarbeit in seiner Forschungsgruppe Entwicklung und Krankheiten am MPI für molekulare Genetik anfertigen zu können. Des Weiteren danke ich ihm für seine hervorragende wissenschaftliche Betreuung, die vielen Diskussionen und vor allem die stete Unterstützung. Bei Prof. Dr. Sigmar Stricker möchte ich mich für die Übernahme des Zweitgutachtens und die zahlreichen Tipps und Tricks in Sachen der Mausgenetik bedanken.

Bevor ich mich bei einzelnen Mitgliedern der Forschungsgruppe bedanke, möchte ich meine Wertschätzung der gesamten Gruppe aussprechen. Die bemerkenswerte Hilfsbereitschaft untereinander und die produktive, aber dennoch überaus lockere und lebendige Arbeitsatmosphäre haben wesentlich zum Erfolg dieser Arbeit beigetragen. Es macht einfach Spaß in einem solchen Team zu arbeiten.

Besonderer Dank gilt Daniel Ibrahim für die großartige Zusammenarbeit am Sox9 Projekt. Mit gemeinsamer Kraft, vielen Diskussionen mit Stefan und "sportlichen" Kaffeekonsum haben wir es tatsächlich geschafft die Story in "N" zu publizieren; Danke! Hervorheben möchte ich die Arbeit von Asita Stiege, Norbert Brieske, Nicole Rösener und Ute Fischer. Sie unterstützen uns in allen Bereichen unserer Forschung und ohne sie würde einfach mal Garnichts funktionieren. Ich danke Rieke Kempfer und Korina Driva, den besten Studenten die ich je betreuen durfte. Euer Engagement war nicht nur eine Hilfe, vielmehr eine Quelle steter Motivation für mich. Dario Lupiáñez danke ich für die Zusammenarbeit und Diskussionen in allen Lebensbereichen. Wir haben von Anfang an gemeinsam an Problemlösungen gearbeitet und sind mit den Herausforderungen in der Wissenschaft unaufhörlich gewachsen. Ich danke Ivana, Robert, Fany, Katerina, Guillaume, Bjørt, Verena, Pedro, Alex, Malte, Sala, Niclas, Mike, Lila, Hendrikje, Saniye für die schöne Zeit am MPI und den vielen Spaß außerhalb der Arbeit. Bei Christina Paliou möchte ich mich im Besonderen für die Hilfe und Diskussionen, vor allem in den wirklich wichtigen Bereichen des Lebens bedanken. Zudem für ihre Geduld mit mir und dem Korrekturlesen bedanken. Francois Spitz und seinem Team danke ich für die schöne Zeit am EMBL in Heidelberg. Im Besonderen Wibke Schwarzer und Sandra Ruf.

Meinen Freunden, von nah und fern danke ich für die doch so dringenden Abwechslungen im Leben und für euer Verständnis, wenn ich doch wieder länger abwesend oder gestresst war. Saksia Wutke danke ich für die gemeinsame Zeit in Berlin und für den stetigen Beweis, dass es sich letztendlich auszahlt unbequeme Wege zu gehen. Zu guter Letzt möchte ich mich bei meiner Familie, meinen Eltern und meinem Bruder für den Rückhalt und die Unterstützung während der gesamten Zeit bedanken.

12 Selbstständigkeitserklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbst angefertigt und keine anderen, als die hier angegebenen Hilfsmittel verwendet habe. Ich versichere, dass ich diese Arbeit weder in dieser noch in einer anderen Form bei einer anderen Prüfungsbehörde eingereicht habe.

Berlin, den 17.03.2017

Martin Franke

13 Scientific publications

Franke M, Ibrahim DM, Andrey G, Schwarzer W, Heinrich V, Schöpflin R, Kraft K, Kempfer R, Jerković I, Chan WL, Spielmann M, Bernd Timmermann, Wittler L, Kurth I, Cambiaso P, Zuffardi O, Houge G, Lambie L, Brancati F, Pombo A, Vingron M, Spitz F, Stefan Mundlos: *Formation of new chromatin domains determines pathogenicity of genomic duplications*. Nature 10/2016; 538(7624). DOI:10.1038/nature19800

Beate Hoppe, Stefan Pietsch, Martin Franke, Sven Engel, Marco Groth, Matthias Platzer, Christoph Englert: *MiR-21 is required for efficient kidney regeneration in fish*. BMC Developmental Biology 11/2015; 15(1). DOI:10.1186/s12861-015-0089-2

Darío G Lupiáñez, Katerina Kraft, Verena Heinrich, Peter Krawitz, Francesco Brancati, Eva Klopocki, Denise Horn, Hülya Kayserili, John M Opitz, Renata Laxova, Fernando Santos-Simarro, Brigitte Gilbert-Dussardier, Lars Wittler, Marina Borschiwer, Stefan A Haas, Marco Osterwalder, Martin Franke, Bernd Timmermann, Jochen Hecht, Malte Spielmann, Axel Visel, Stefan Mundlos: *Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions*. Cell 05/2015; 161(5). DOI:10.1016/j.cell.2015.04.004

Katerina Kraft, Sinje Geuer, Anja J. Will, Wing Lee Chan, Christina Paliou, Marina Borschiwer, Izabela Harabula, Lars Wittler, Martin Franke, Daniel M. Ibrahim, Bjørt K. Kragesteen, Malte Spielmann, Stefan Mundlos, Darío G. Lupiáñez, Guillaume Andrey: *Deletions, Inversions, Duplications: Engineering of Structural Variants using CRISPR/Cas in Mice*. Cell Reports 02/2015; 10(5). DOI:10.1016/j.celrep.2015.01.016

Malte Spielmann, Francesco Brancati, Peter M Krawitz, Peter N Robinson, Daniel M Ibrahim, Martin Franke, Jochen Hecht, Silke Lohan, Katarina Dathe, Anna Maria Nardone, Paola Ferrari, Antonio Landi, Lars Wittler, Bernd Timmermann, Danny Chan, Ulrich Mennen, Eva Klopocki, Stefan Mundlos: *Homeotic Arm-to-Leg Transformation Associated with Genomic Rearrangements at the PITX1 Locus*. The American Journal of Human Genetics 09/2012; 91(4). DOI:10.1016/j.ajhg.2012.08.014

Franziska Mussbach, Martin Franke, Ansgar Zoch, Buerk Schaefer, Siegmund Reissmann: *Transduction of Peptides and Proteins Into Live Cells by Cell Penetrating Peptides*. Journal of Cellular Biochemistry 12/2011; 112(12). DOI:10.1002/jcb.23313