UNIVERSITA' DEGLI STUDI DI PAVIA

Dipartimento di Medicina Molecolare

Generation and characterization of structural variations at the *FGF8* locus

Giulia Cova

Dottorato di Ricerca in
Genetica, Biologia Molecolare e Cellulare
Ciclo XXX – A.A. 2014-2017

UNIVERSITA' DEGLI STUDI DI PAVIA

Dipartimento di Medicina Molecolare

Generation and characterization of structural variations at the *FGF8* locus

Giulia Cova

Supervised by Prof. Roberto Ciccone

Dottorato di Ricerca in
Genetica, Biologia Molecolare e Cellulare
Ciclo XXX – A.A. 2014-2017

To my extraordinary brother Riccardo

“Above all, don't fear difficult moments.
The best comes from them.”

Rita Levi-Montalcini

ABSTRACT

Chromosome conformation capture assays provided important insights into the 3D spatial organization of mammalian genomes. In particular, Hi-C technique allowed the identification of the so called topologically associated domains (TADs), specific structures that subdivide the genome into discreet genomic units, providing the physical proximity between CREs and their target genes, which is required for transcription. Different studies demonstrated that several human diseases can result from perturbations of the TAD structure as a consequence of structural variations, such as duplications, deletions and inversion. These structural rearrangements can lead indeed to disruption or alterations of proper contacts between a gene and its regulatory elements.

Several tandem duplications on chromosome 10q24 at the *FGF8* locus have been linked to the Split-Hand/Foot Malformation type 3 (SHFM3), a limb malformation affecting predominantly the central digits of hands and feet, in some cases also associated with intellectual disability, mental retardation, or craniofacial abnormalities, and whose precise molecular mechanism is still unknown. In this study CRISPR/Cas9 genome editing was used to reengineer SHFM3 duplications at the *FGF8* locus in mice, to try to understand the pathomechanism underlying this disease. Additionally, capture Hi-C (cHi-C) approach was used to investigate the spatial and functional organization of the *FGF8* locus in wild-type and mutant mice in order to evaluate the impact of these duplications on local chromatin architecture and on gene regulation. cHi-C analysis in wild-type mice revealed that the *Fgf8* locus is characterized by the presence of two TADs (referred to as *LBX1* TAD and *FGF8* TAD), each comprising specific genes and regulatory elements, whose interactions are constrained by TAD boundaries. Comparison of the SHFM3 duplications position relative to the TAD structure at the *FGF8* locus classified these duplications as inter-TAD duplications, as they encompassed the TAD boundary between the *LBX1* TAD and the *FGF8* TAD.

This type of duplications can lead to the formation of new TADs (neo-TADs) around the breakpoint of the duplication. Within these neo-TADs new ectopic interactions between regulatory elements and genes normally located in separate TADs can occur, resulting in aberrant gene expression and pathological phenotypes. Indeed, we observed that SHFM3 duplications were responsible for relocation of duplicated genes and regulatory elements to a new interactive context, leading to the onset of ectopic interactions between sequences brought in close proximity as a consequence of the rearrangement.

Despite the generated mutant mice did not completely recapitulate the SHFM3 phenotype and the whole project is not over yet, this study already offered important insights regarding gene regulation at the *FGF8* locus and the consequences that derive from perturbations of the local chromatin structure.

CONTENTS

1. INTRODUCTION

1.1 Gene regulation

The sequencing of the human genome by the Human Genome Project (HGP) estimated approximately 25,000 protein coding genes in our genome (International Human Genome Sequencing Consortium 2004), a lower number compared to the one found in corn (Schnable PS et al. 2009) and nearly twice the one in *Drosophila melanogaster* (Adams MD et al. 2000). Surprisingly, those 25,000 genes are encoded in only 2% of the genome, while the other 98% contain non-coding sequences, previously referred to as "junk DNA". However, international efforts to functionally annotate the genome revealed that the majority of non-coding sequences contain regulatory RNAs and regulatory elements that can control and instruct the expression of genes (The ENCODE Project Consortium 2012).

Precise and coordinated gene expression of thousands of genes is a fundamental task for multicellular organisms. During embryonic development, cells receive multiple inputs from signalling cascades, which requires an elaborate regulatory system able to process the huge number of signals in different cells of the embryo (Davidson EH et al. 2002). Especially developmental genes, involved in pattern formation and cell fate specification, are strictly regulated in a spatial and temporal fashion (Busser WB et al. 2008). Consequently, perturbations of their precisely adjusted expression patterns are often associated with congenital disease (Pfeifer D et al. 1999; Lettice LA et al. 2003) or cancer (Hatton BA et al. 2006).

The regulation of genes is achieved at multiple levels, ranging from the control of transcriptional initiation to mRNA processing and post-transcriptional modifications of the mRNA and proteins (Phillips T 2008). However, a major mechanism of regulation, especially for the majority of developmental genes, is the control at the genomic level via *cis*-regulatory elements (Howard ML and Davidson EH 2004).

1.1.1 Transcriptional control by *cis*-regulatory elements

In order to achieve organized gene expression patterns, multiple transcriptional regulatory elements that act in *cis* are required. These *cis*-regulatory elements (CREs) provide binding sites for *trans*-regulators, represented by general transcription factors (TFs) and tissue-specific TFs. The composition of recognition motifs in CREs and subsequent recruitment of co-factors can initiate or terminate transcription (Istrail S and Davidson EH 2004). In eukaryotes, transcription is initiated at the promoter by the recruitment of RNA Polymerase II (RNA Pol II), a DNA dependent RNA polymerase (Fuda NJ et al. 2009). Promoter and specific distal regulatory elements represent two distinct groups of CREs of protein-coding genes transcribed by RNA Pol II (Fuda NJ et al. 2009; Maston GA et al. 2006) (Figure 1).

Control elements at promoters are non-coding sequences located close to the protein-coding region of a gene. They are classically defined as the core promoter and the proximal promoter elements. The core promoter, located near the transcription start site (TSS), represents the anchor point for the basic transcriptional machinery and RNA polymerase II (Javahery R et al. 1994). The first eukaryotic core promoter motif identified was the TATA box (Goldberg ML, 1979). The proximal promoter represents instead a region a few hundred base pairs upstream of the core promoter characterized by the presence of multiple binding sites for transcriptional activators that can initiate transcription by release of PolI II from the promoter (Wu CH et al. 2003).

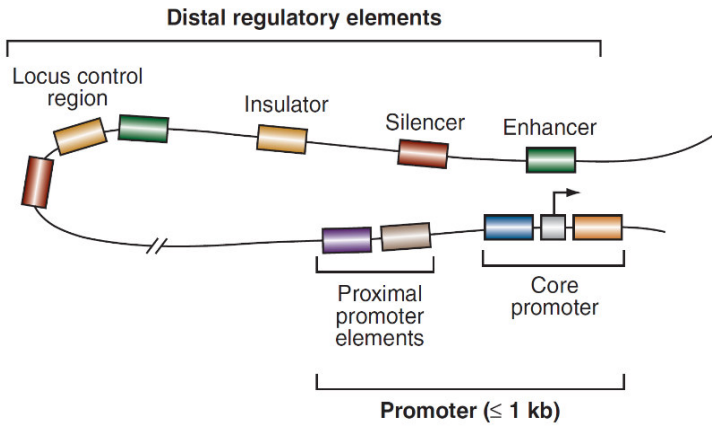


Figure 1. Schematic overview of cis-regulatory elements controlling gene expression. The promoter encompasses a region smaller than 1 kb and consists of two main components: the core promoter and the proximal promoter elements. Up to 1 Mb away from the promoter we can find the distal regulatory elements, including enhancers, silencers, insulators and locus control regions. These distal elements are able to contact the core promoter or the proximal promoter through DNA looping (Maston GA et al. 2006).

1.1.2 Distal *cis*-regulatory elements

A key feature of the vertebrate genome is the relative large distance of additional regulatory elements from the classical defined promoter elements. Distal regulatory elements can be located up to 1 Mb from the gene they control (Lettice LA et al. 2003). There are different classes of distal CREs depending on the factors binding and their effect on transcription. Enhancers promote activation of genes (Heintzman ND et al. 2007), while silencers are responsible for transcriptional repression, since they are bound by negative TFs called repressors (Ogbourne S and Antalis TM 1998). Insulators, also known as boundary elements, control the transcriptional activity of a gene preventing any kind of influences from regulatory elements of the neighbouring genes, therefore creating defined domains of expression (Recillas-Targa F et al., 2002).

1.1.2.1 Distal enhancers

Enhancers can be found several hundred kilobase pairs downstream or upstream of the transcription unit, in the intergenic regions and sometimes also in the introns of the neighbouring genes (Heintzman ND et al. 2007). Their function, concerning precise spatiotemporally and quantitatively gene activation, is totally independent of their orientation and the distance relative to the promoter (Zabidi MA et al. 2015; Bulger M and Groudine M 2011).

The physical contact between these distal elements and the core promoter or the proximal promoter occurs through a mechanism that involves the looping of DNA (Tolhuis B et al. 2002; Wang Q et al. 2005). In response to a signal, DNA-binding sites of enhancers are bound by pioneer TFs in condensed chromatin (Xu J et al. 2007), leading to the recruitment of ATP-dependent chromatin remodeling complex and histone-modifying enzymes. Histone-modifying enzymes can, for instance, acetylate, methylate, phosphorylate the histone components of the nucleosome, influencing transcriptional activity (Bannister AJ and Kouzarides T 2011), whereas ATP-dependent chromatin remodeling complex are able, through mobilization of the nucleosome, to increase the access to DNA sequences for more tissue- or stimulus-specific TFs (Hargreaves DC and Crabtree GR 2011; Clapier CR and Cairns BR 2009). These specific TFs can interact with cofactors, protein complexes without DNA-binding property involved in activation (coactivators) or repression (corepressors) of gene transcription; examples of coactivators are the Mediator complex, the histone acetyltransferase p300 and some general transcription factors (Heintzman ND et al. 2007; Maston GA et al. 2012; Lee TI and Young R 2013). In some cases, TFs bound to enhancers can interact also with each other forming stable complexes called “enhanceosomes”, also involved in activation (Pan Y and Russinov R 2011). The cascade of events described above brings to the recruitment of the preinitiation complex (PIC), which directs the RNA Pol II to the TSS. RNA Pol II binds to general transcriptions factors and the loop structure between the TSS and enhancers is stabilized by the Mediator complex and cohesin. RNA Pol II starts the transcription process from the TSS (Fromm G et al. 2013). Several human diseases and syndromes associated with mutations in transcriptional regulatory elements and/or in the transcriptional machinery have been reported (Kleinjan DA and van Heyningen V 2005; Lee TI and Young RI 2013).

Enhancers are the key regulatory elements of gene expression, that results from the sum of all the specific spatiotemporal activities of each single enhancer. The higher the complexity of expression patterns, the higher is the expected number of enhancers and the longer the

intergenic and intronic regions of a gene. This is especially true for the majority of developmental control genes which are characterized by large regulatory regions with numerous enhancers (Nelson CE et al. 2004). Recent studies also revealed a certain redundancy of these regulatory elements, showing the presence of more than one enhancer with overlapping activity (Perry MW et al. 2010) and of which one can be deleted without significant phenotypic effect (Cannavò E et al. 2016). These so called redundant enhancer, or "shadow" enhancer, are thought having an important role in the spatio-temporal precision and robustness of gene expression during the complex developmental processes (Perry MW et al. 2011; Perry MW et al. 2012; Dunipace L et al. 2011; Frankel N et al. 2010).

1.1.3 Identification of functional enhancers

Among all the CREs, enhancers are the best characterized because of their determining role in gene transcription and several strategies and tools have been developed for their identification.

Traditionally, candidate enhancers are predicted by **evolutionary conservation**. Wide-ranging vertebrate genome sequence comparisons revealed a strong conservation of many non-coding sequences that often contain CREs, therefore one common feature of CREs is the significant association with ultra-conserved elements and the high conservation of genomic location between species (Hardison RC 2000; Woolfe A et al. 2005). However, sequence conservation is not sufficient to predict all enhancers as there are also cases of transcription factor binding sites that are not conserved between species. For instance, comparison of tissue-specific enhancers has revealed varying evolutionary constraints in sequence conservation, depending on the tissue type (Blow MJ et al. 2010). To overcome these limitations other strategies have been developed for identification of enhancers.

A powerful advantage is given by **chromatin immunoprecipitation (ChIP)** techniques, which allow to detect enhancers using specific antibodies against the TFs binding these non-coding sequences. DNA-protein interactions are captured by cross-linking with formaldehyde, which creates covalent bonds between DNA and proteins bound to it. Chromatin immunoprecipitation followed by deep sequencing (ChIP-seq) identifies those genomic regions bound by a specific factor genome-wide (Johnson DS et al. 2007). ChIP antibodies can be directed not only against TFs, but also against protein-factors that are associated with active enhancers. P300, a histone acetyltransferase that co-bind with tissue-specific TFs, has been shown to be a powerful predictor of enhancers in different tissues (Visel A et al. 2009). Furthermore, antibodies against histone modifications representative of active enhancers, such as histone-3 lysine-27 acetylation (H3K27ac) and histone-3 lysine-4 mono-methylation (H3K4me1), are widely used to detect these CREs (Heintzman ND et al. 2007; Creygton MP et al. 2010). Transcription factors binding sites can be also identified through **DNase I hypersensitive site mapping**, a technique based on the fact that TF binding produces an open chromatin conformation, that is more accessible to the DNase I digestion (Crawford GE et al. 2003).

The function of candidate enhancers is usually tested in **reporter-gene assays**, which involves the cloning of the potential enhancer sequence into a plasmid upstream of a reporter gene with a minimal promoter, such as the green fluorescent protein (GFP), the β-galactosidase, or the luciferase gene (Figure 2A). The construct is subsequently transfected in cultured cells or injected into embryos of model animals and the activity of the reporter is measured throughout growth and development (Heintzman ND et al. 2007). Alternatively, reporter genes have been also used to detect the regulatory activity along chromosome or at selected loci (Ruf S et al. 2011; Uslu VV et al. 2014). Reporter constructs, consisting only of a reporter gene and the minimal promoter are integrated throughout the genome by transposition in order to investigate the regulatory activity directly from the endogenous genomic context (Figure 2B); the position of the integration of the reporter gene determines

a specific expression based on the activity of an enhancer or cluster of enhancers in the regulatory domain where the construct is integrated (Ruf S et al. 2011).

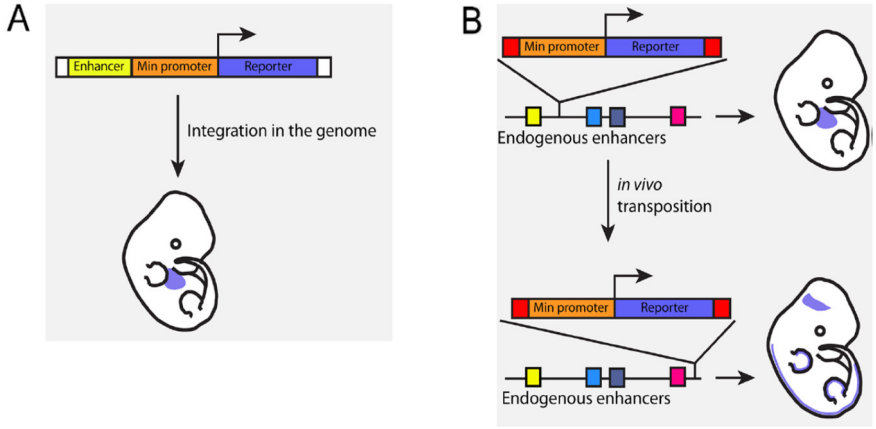


Figure 2. Reporter-gene assays. (**A**) The construct containing the candidate enhancer is randomly integrated in the genome and its activity is tested. (**B**) The construct does not contain a candidate enhancer, but only a reporter gene and the minimal promoter. The position of the integration defines a specific gene expression pattern based on the enhancers acting on the gene-reporter. The construct is flanked with transposon sites for integration throughout the genome (Kathiriya IS et al. 2015).

1.1.4 The role of 3D genome organization in gene regulation

Chromatin conformation capture (3C) assays combined with high-throughput sequencing provided important information regarding the structural organization of the genome. In particular they allowed the detection of genomic contacts, including long distance regulatory interactions as those between enhancers and promoters. The 3C approach is based on formaldehyde crosslinking of DNA-proteins contacts within the intact nucleus of a cell, followed by fragmentation of the chromatin with restriction enzymes in order to create free ends of the cross-linked DNA. The subsequent ligation creates chimeric DNA molecules of DNA sequences that are in close proximity (de Laat W and Dekker J 2012) (Figure 3A). Various methods have been developed to identify these ligation products and to determine their interaction frequency in the genome. The original developed **3C** method measures the interactions between a single pair of genomic loci (one vs. one) using polymerase chain reaction (PCR) (Dekker J et al. 2002) (Figure 3B). The **4C-seq** (circular chromosome conformation capture) involves a second round of digestion and ligation followed by inverse PCR with locus-specific primers and deep sequencing in order to detect all the possible interactions of a locus of interest and their frequency (one vs. all) (Zhao Z et al. 2006) (Figure 3B). Finally, the **Hi-C** method detects all the possible interactions throughout the entire genome (all vs. all) (Lieberman-Aiden E et al. 2009) (Figure 3B). Furthermore, **Capture-Hi-C**, a combination of Hi-C with the use of biotin-labelled probes complementary to a specific genomic region, uncover all the interactions occurring in the region of interest (Hughes JR et al. 2014; Mifsud B et al. 2015).

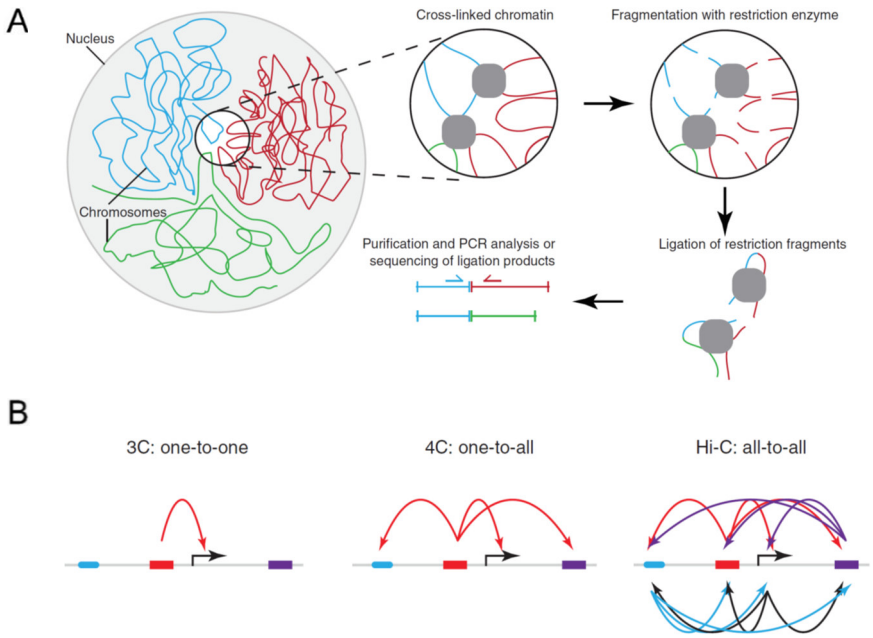


Figure 3. Chromatin conformation capture techniques. (**A**) This technology consists of chromatin-proteins crosslinking with formaldehyde, followed by digestion with restriction enzymes, re-ligation of chromatin fragments that are in close proximity, crosslinking removal and purification. PCRs or deep sequencing reactions are performed to detect ligation products. (**B**) Methods to determine one-to-one, one-to-all, or all-to-all chromatin interactions. Coloured rectangles represent *cis*-regulatory elements, black arrows indicate the TSSs, while the curved coloured arrows show the interactions obtained from each strategy (Wijchers PJ and de Laat W 2011).

Hi-C techniques revealed a non-random nuclear organization, which is strictly related to the transcriptional status and has an important role in gene regulation (Rao SSP et al. 2014). At the highest level of organization, interphase chromatin is organised in **chromosome territories**, one for each chromosome, that rarely interact with each other, as the Hi-C data showed a greater frequency of interactions between loci on the same chromosome compared to those located on different ones (Lieberman-Aiden E et al. 2009) (Figure 4A). At the mega-base scale chromosome territories are organized into two types of **compartments**, called **A** and **B**, determined by the transcriptional state. The A compartment is associated with euchromatin and active transcription, whereas B compartment corresponds to the gene poor, condensed and transcriptionally inactive heterochromatin (Rao SSP et al. 2014; Lieberman-Aiden E et al. 2009; Rowley MJ et al. 2017) (Figure 4B). The active A compartments are usually located in the nucleus and tend to interact preferentially with other type A domains, while the inactive B are predominant close to the nuclear envelope and tend also to interact with compartments of the same type (Lieberman-Aiden E et al. 2009). At the sub-megabase scale **topologically associated domains** (**TADs**) further subdivide the chromatin into domains with an average size of 800kb in mammals (Dixon JR et al. 2012) (Figure 4B). TADs are separated by TAD boundaries, and sequences inside TADs interact with each other more frequently than with sequences of neighbouring TADs (Dixon JR et al. 2012; Nora EP et al. 2012). The subdivision of chromatin into TADs, in which DNA sequences are in close proximity, as well as the fact that the majority of DNA contacts occur within TAD structures (Dixon JR et al. 2012), are of great importance for contacts between genes and CREs and are therefore crucial for gene regulation. Additionally, the chromatin organization and TAD structure have been shown to be conserved across cell types, human and mice (Dixon JR et al. 2012; Rao SSP et al. 2014). TAD boundaries also exhibit a strong enrichment of an architectural protein known as CCCTC-binding factor (CTCF) (Dixon JR et al. 2012). The interaction of two CTCF proteins, bound at different CTCF-binding sites with convergent motif orientation, is involved in chromatin loop formation, together with a ring

of cohesin that stabilizes the structure (Figure 4C) (Rao SSP et al. 2014; de Wit E et al. 2015; Gómez-Marín C et al. 2015). Additionally, cohesin is associated with the Mediator complex (Kagey MH et al. 2010), which brings enhancers and promoters in close proximity (Figure 4C), further supporting the idea of a correlation between chromatin loops and transcription (Rao SSP et al., 2014).

Structural variations such as inversions (Guo Y et al. 2015) or deletions (Sanborn A et al. 2015) of CTCF-binding sites result in loop alteration, leading to changes or reduction in intra-TAD contacts and increasing of inter-TAD interactions. Thus, compartmentalization into TADs provides a structural context for gene regulation by safeguarding promoters and corresponding CREs efficiently finding each other (Shen Y et al. 2012).

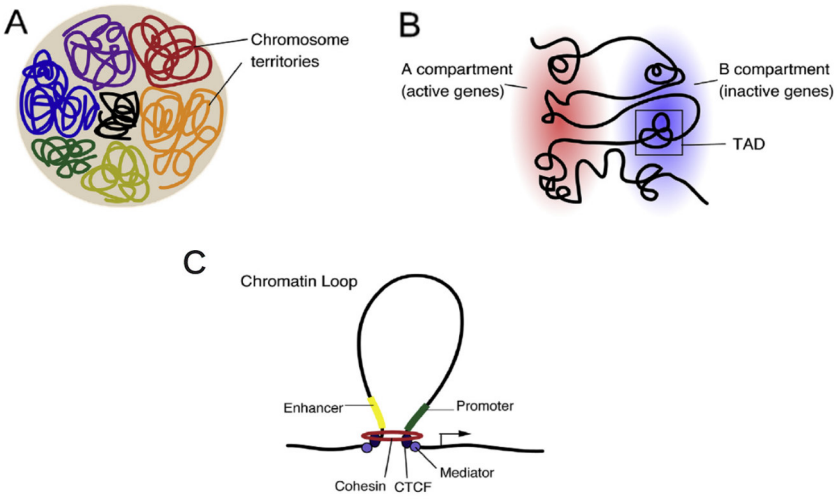


Figure 4. Hierarchical organization of chromatin. (**A**) The highest level of 3D architecture is represented by chromosome territories. (**B**) The transcriptionally active A compartment and the inactive B are the second level of organization, whereas at the sub-megabase scale we find structures called TADs. (**C**) Typical structure of a chromatin loop (Fujita Y and Yamashita T 2017).

1.2 Limb development: a model system for gene regulation

Limb formation represents an excellent system for studying and understanding developmental biology and also gene regulation, since its genesis results from a fine and well-coordinated regulatory network (Niswander L 2002; Williamson I et al. 2016). The limb starts as a small bud emerging from the lateral side of the body and composed of undifferentiated mesenchymal cells covered by a layer of ectoderm (Niswander L 2003). The three-dimensional development and growth occurs along three axes and is regulated by specific interactions and interconnections between three different signalling centres and their respective secreted molecules (Niswander L 2002) (Figure 5).

The **proximal-distal axis** is controlled by the apical ectodermal ridge (AER), localized along the distal limb bud tip at the interface between dorsal and ventral ectoderm (Casanova JC et al. 2011). The AER produces the fibroblast growth factors (FGFs), which are responsible for maintaining the mesenchymal cells in a proliferative state allowing the growth of the limb from the shoulder to the fingers (Niswander L 2002). FGFs signalling molecules are encoded by members of the *Fgf* gene family. Among the four *Fgf* genes (*Fgf4, Fgf8, Fgf9* and *Fgf17*) specifically expressed in the AER, *Fgf8* is the only one expressed from early AER formation. *Fgf4, Fgf9* and *Fgf17* are activated later in the posterior AER (Lewandoski M et al. 2000). It has been demonstrated that *Fgf4, Fgf9* and *Fgf17* individual loss of expression in the AER has no effect on limb development (Mariani FV et al. 2008), contrary to *Fgf8* loss alone (Lewandoski M et al. 2000; Moon AM and Capecchi MR 2000) or in combination with *Fgf4* (Sun X et al. 2002), underlining that *Fgf8* plays an important role in the proper limb outgrowth. *Fgf8* expression in the AER is induced by another gene of the *Fgf* family, *Fgf10*, which is expressed in the mesenchyme. The establishment of a positive feedback loop between the epithelial *Fgf8* and mesenchymal *Fgf10* is determinant to maintain the proliferation of mesenchymal cells and, consequently, limb outgrowth along the proximal-distal axis (Sekine K et al. 1999; Ohuchi H et al. 1997).

The **anterior-posterior axis** is mainly defined by the zone of polarizing activity (ZPA), which locates at the posterior limb bud mesenchyme and secrets the sonic hedgehog (SHH) signalling molecule, responsible for digits formation from V to II (Riddle RD et al. 1993). SHH diffuses across the limb bud generating a spatio-temporal gradient in terms of SHH concentration in different parts of the limb bud and time of exposure to its signalling. This specify the anterior-posterior identity and number of the digits from V to II. Digit I is SHH-independent (Harfe BD et al. 2004). *Shh* expression is activated and restricted to the posterior limb bud mesenchyme by several transcriptional regulators, including for instance HAND2 and GLI3 (Zeller R et al. 2009). Furthermore, SHH and FGFs mutually regulate each other as part of an important feedback loop, which involves also the bone morphogenic protein 4 (BMP4) and gremlin 1 (GREM1) and operates throughout the development of the limb (Bénazet JD et al. 2009).

The proximal-distal and the anterior-posterior axes are also related to skeleton formation, which results from mesenchymal cells condensation, differentiation into cartilage and, finally, into bone (Storm EE and Kingsley DM 1996). Three elements are specifically involved in the definition of the skeletal limb components: the stylopod defines humerus and femur, the zeugopod gives rise to radius/ulna and tibia/fibula, and the autopod generates the carpal/metacarpal, tarsal/metatarsal and phalanges bones (Chiang C et al. 2001; Zeller R et al. 2009) (Figure 6).

Finally, the **dorsal-ventral axis** is governed by several molecules, that promote the correct formation from the external to the internal side of the limb bud. WNT family member 7a (WNT7a) signalling from the dorsal ectoderm promotes the dorsal pattern formation inducing the activity of the transcription factor LMX1B in the dorsal mesenchyme. Transcription factor Engrailed1 (EN1), induced in the ventral ectoderm by bone morphogenic protein (BMP), determines instead the ventral pattern formation and, at the same time, inhibits *Wnt7a* expression in the ventral ectoderm (Ahn K et al. 2001).

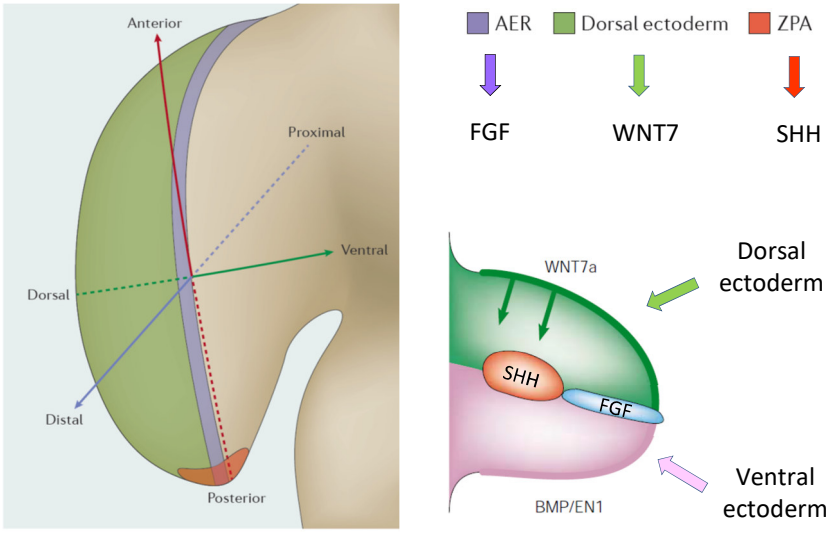


Figure 5. Limb formation axes. Three axes are involved in limb development: proximal-distal, anterior-posterior and dorsal-ventral, that are controlled by the apical ectodermal ridge, the zone of polarizing activity and the dorsal/ventral ectoderm signalling centers, respectively. Each area secrets key molecules involved in the regulation of the limb bud formation (adapted from Petit F et al. 2017 and Niswander L 2003).

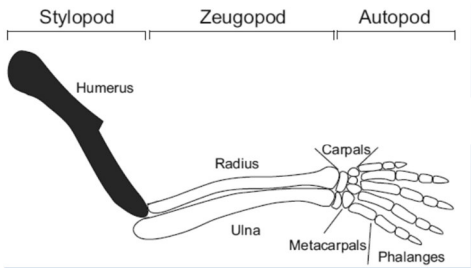


Figure 6. Skeletal components of the arm. Three specific segments give rise to the different elements of a arm skeleton. Humerus derives from the stylopod, radius and ulna from the zeugopod, and carpal, metacarpal and phalanges bones from the autopod. (Tanaka M 2016).

1.3 Structural variations and their impact on gene regulation

Structural variations (SVs) including duplications, deletions and inversions have been shown to contribute to human genetic diversity (Sebat J et al. 2004), consequently supporting evolution and adaptation of organisms to environmental changes. On the other hand, SVs have been linked to several human genetic disorders, affecting genes as well as parts of noncoding regions of the genome where multiple *cis*-regulatory elements are located (Mundlos S and Klopocki E 2011). Therefore, such SVs can lead to disruption or alterations of proper contacts between a gene and its regulatory elements. For instance, duplications encompassing *SHH* enhancer ZPA regulatory sequence (ZRS) have been linked to polysyndactyly or triphalangeal thumb-polysyndactyly syndrome (Klopocki E et al. 2008; Sun M et al. 2008). Additionally, SVs can relocate genes or enhancers in a different *cis*-regulatory environment leading to new spatio-temporal expression patterns, which can result in disease. This mechanism, through which a gene is positioned under the control of different enhancers, has been called “enhancer adoption” (Lettice LA et al. 2011).

The discovery of the three-dimensional organization of the genome further helped to understand and to study the impact of SVs on gene regulation. Different studies demonstrated that several human diseases can result from perturbations of the TAD structure as a consequence of SVs (Hnisz D et al. 2016; Lupiáñez DG et al. 2015; Franke M et al. 2016). In particular, deletions that overlap a TAD boundary are associated with ectopic interactions and gene misregulation, causing the activation of a proto-oncogene (Hnisz D et al. 2016), or limb malformations such as brachydactyly (Lupiáñez DG et al. 2015). Duplications that include part of two TADs and TAD boundary can lead to the formation of a new chromatin domain (neo-TAD) comprising genes and *cis*-regulatory elements normally located in separate TADs. This can promote ectopic interactions between these genes and *cis*-regulatory elements, which can have consequences on gene regulation and cause disease (Franke M et al. 2016; Weischenfeldt J et al 2017).

1.3.1 Structural variations at the *FGF8* locus are associated with Split-Hand/Foot Malformations type 3 (SHFM3)

Split-Hand/Foot Malformation (SHFM) is a clinically and genetically heterogeneous congenital limb defect affecting mainly the central rays of the autopod and presenting with clefts of the hands and the feet, often accompanied by syndactyly and aplasia or hypoplasia of the phalanges, metacarpals and metatarsals (Kano H et al. 2005). It occurs with an incidence of 1 in 8.500-25.000 newborns and accounts for 8-17% of all limb reduction defects (Elliott AM et al. 2006; Calzolari E et al. 1990; de Mollerat XJ et al. 2003). Seven chromosomal loci have been linked so far to different types of SHFM: SHFM1 on 7q21 (Crackower MA et al. 1996; Shamseldin HE et al. 2012), SHFM2 on Xq26 (Faiyaz ul Haque M et al. 1993), SHFM3 on 10q24 (Gurrieri F et al. 1996; de Mollerat XJ et al. 2003), SHFM4 on 3q27 (van Bokhoven H et al. 2001; Celli J et al. 1999), SHFM5 on 2q31 (Goodman FR et al. 2002), SHFM6 on 12q13 (Ugur SA and Tolun A 2008) and SHFM with long bone deficiency (SHFM/SHLD) on 17p13 (Klopocki E et al. 2012).

In particular, Split-Hand/Foot Malformation type 3 (SHFM3), which is inherited as an autosomal dominant trait (Evermann DB et al. 2006), has been associated with tandem duplications spanning a region of at least 500 kb on chromosome 10q24 (de Mollerat XJ et al. 2003; Kano H et al. 2005; Dimitrov BI et al. 2010). Patients usually present the typical absence of the central digits on hands and/or feet. Some cases are characterized by mental retardation, intellectual disability or craniofacial abnormalities, such as micrognathia, which can manifest concurrently, or not, with the limb phenotype (de Mollerat XJ et al. 2003; Elliot AM and Evans JA 2006; Dimitrov BI et al. 2010). These identified duplications span a region centromeric to *FGF8* and include *LBX1*, *BTRC*, *POLL*, *DPCD* and part of *FBXW4* (de Mollerat XJ et al. 2003; Kano H et al. 2005; Dimitrov BI et al. 2010). *LBX1* and *BTRC* are specifically involved in developmental processes. *LBX1* is expressed in the central nervous system and in the early myogenic cells of the limb bud (Brohmann H et al. 2000; Fernández-Jaén A et al.

2014; Jagla K wt al. 1995) and *Lbx1* knock-out mice show an extensive loss of limb muscles (Gross MK et al. 2000). *BTRC* is involved in the NF-kB (Nuclear Factor κB) signaling transduction pathway, which has been shown having an important role in limb development by stimulating *Shh* expression and repressing *Bmp4* in limb mesenchyme (Bushdid PB et al. 1998).

Although the SHFM phenotype is thought to result from a failure in maintaining the AER during the development of the autopod (Duijf PH et al. 2003), a specific gene and a precise molecular mechanism by which the duplications cause the SHFM3 and other phenotypic features in humans have not been found so far (Lyle R et al. 2006; Dimitrov BI et al. 2010; Kano H et al. 2005). An accepted model for the human SHFM3 is *Dactylaplasia* (*Dac*) in mice. The *Dac* limb phenotype, characterized by missing phalanges and reductions or fusions of metacarpals and metatarsals in each foot, is highly similar to SHFM3 human phenotype and the *Dac* alleles map at the region syntenic to the human duplications (Kano H et al. 2007). However, no duplications were detected in the *Dac* mice. Instead, two insertions of a retrotransposon element either within or upstream of the *Fbxw4* gene, a member of the F-box WD-40 gene family involved in ubiquitin-dependent degradation processes (Sidow A et al. 1999), were linked to the observed phenotype. However, the mechanism of how the insertions lead to the phenotype is still not known (Kano H et al. 2007; Friedli M et al. 2008). Furthermore, Kano H *et al*. (2005) suggested that the probability of *Fbxw4* involvement in SHFM3 is very low as, for instance, the gene is not even completely involved in the duplication.

Interestingly, a recent study in mice (Marinić M et al. 2013) uncovered a region of 220 kb between the *Btrc* gene and the region telomeric of *Fbxw4* characterized by the presence of multiple enhancers responsible for the proper spatio-temporal expression of *Fgf8*. At embryonic stage 10.5, *Fgf8* is expressed in the AER of the limb buds, the midbrain-hindbrain boundary, the forebrain commissural plate, the first brachial arch ectoderm, the nasal pit epithelium and the somites.

In particular, five of the identified enhancers are specific for *Fgf8* expression in the AER (Marinić M et al. 2013) and four out of these five are contained in the region synthenic to the human SHFM3 duplication. Therefore, rearrangements of these regulatory elements could be responsible for an altered spatio-temporal expression and eventually leading to the SHFM3 phenotype.

1.3.2 Engineering structural variations in the mouse genome

The mouse model is often used to study gene function and regulation and to investigate the pathomechanism of human diseases, as mice and humans are very similar in terms of development, physiological processes and genetics (Demetrius L 2005). Various genome editing techniques have been developed over the past decades to engineer targeted rearrangements in the mouse genome.

One *in vivo* method to generate SVs is known as TAMERE (Trans-Allelic targeted Meiotic Recombination), a site-specific recombination system derived from the *Escherichia coli* bacteriophage P1. It consists of a Cre recombinase enzyme and two short DNA sequences, named loxP sites, that are engineered to flank the target region. Cre recombinase enzyme is able to recognize the loxP sites, leading to excision of the target sequence and subsequently recombination (Hérault et al. 1998). Two mouse lines, each containing one loxP site either upstream or downstream of the target region, are required for this technique and, furthermore, a third mouse line carrying the Cre transgene. Then, triple-transgenic mice carrying the two different loxP sites and the *Cre* transgene can be obtained by multiple breedings. The Cre-recombinase is designed to be active in the germline, thus recombination events occur in gametes only. Therefore, triple-transgenic mice are crossed with wild-type to generate chromosomal rearrangement involving the targeted region in the offspring (Hérault et al. 1998).

Other approaches, such as the zinc-finger nucleases (ZNFs) and TALENs (Transcription Activator-Like Effector Nucleases), are mainly based on the generation of double strand breaks (DSBs), that are subsequently repaired by two cellular DNA repair mechanisms: the non-homologous end joining (NHEJ) and the homologous recombination (HR) (Peng Y et al. 2014). NHEJ is an error-prone mechanism, therefore it can result in deletions or insertions of a few bases at the cut site, while HR is more precise but requires the presence of a homologous repair template (Bibikova M et al. 2002). ZFNs are the result of the fusion of a zinc-finger DNA binding domain with the DNA-cleavage domain from the restriction endonuclease FokI. The zinc finger domains can be engineered to target a specific DNA sequence (Miller JC et al. 2007; Urnov FD et al. 2005). TALENS also derives from the fusion of a DNA-cleavage domain with, in this case, a TAL effector DNA binding domain from *Xanthomonas* bacteria, which can be modified to guide the nuclease to the sequence of interest that needs be rearranged (Miller JC et al. 2011; Christian M et al. 2010).

A revolutionary and extremely efficient alternative strategy to ZFNs and TALENs is the CRISPR/Cas9 genome editing. CRISPR-Cas (Clustered Regularly Interspaced Short Palindromic Repeats – CRISPR associated) systems are genetic hallmarks of adaptive immunity in bacteria, evolved to target and eliminate invading genetic elements such as viruses and plasmids (Makarova KS et al. 2011). These systems consist of a *cas* gene cassette and a CRISPR array that encodes a series of direct repeats, interspaced with short unique "spacer" sequences from foreign DNA, important for genetic memory (Jansen R et al. 2002; Kunin V et al. 2007; Bolotin A et al. 2005; Barrangou R et al. 2007). Spacer precursor foreign DNA sequences are referred to as "proto-spacers" (Mojica FJM 2009). After acquisition of foreign DNA, the following transcription and maturation of the CRISPR locus lead to the CRISPR RNAs (crRNAs) that form RNA-Cas nuclease complexes. These complexes are able to detect and eliminate any foreign DNA through target recognition by crRNA and removal by Cas nuclease cleavage (Makarova KS et al. 2011; Wang J et al. 2015).

There are three types of CRISPR-Cas system and each uses distinct molecular mechanisms for interference: type I and III use a large complex of Cas proteins for crRNA-guided targeting, whereas type II system requires only the Cas9 protein for RNA-guided DNA recognition and cleavage (Makarova KS et al. 2011), a property that proved to be extremely useful for genome engineering applications (Cong L et al. 2013; Wang H et al. 2013; Mali P et al. 2013).

In the *Streptococcus pyogenes* type II system, the Cas9 endonuclease is guided by two non-coding RNAs, the CRISPR RNA (crRNA) and the *trans*-activating crRNA (tracrRNA) (Anders C et al. 2014; Sternberg SH et al. 2014). Different adaptations of this system for the mammalian cells have been performed (Cong L et al. 2013). For instance, the two RNAs have been reengineered into a chimeric single-guide RNA (sgRNA) to optimize genome editing processes (Jinek M et al. 2012). The Cas9 enzyme has two active sites, each cleaving one strand of a target double-stranded DNA molecule demarcated by a Protospacer Adiacent Motif (PAM) sequence, which helps the recognition by the sgRNA complementary to the target DNA (Charpentier E and Doudna JA 2013). RNA-guided Cas9 activity creates site-specific double-stranded DNA breaks (DSBs), which can be repaired either by the NHEJ repair or HR. However, it has been observed that DSBs generated by Cas9 are usually repaired by NHEJ (Li J et al. 2015) and, when two sgRNAs are simultaneously used, small up to very large deletions, inversions and duplications can be efficiently obtained (Kraft K et al. 2015; Li J et al. 2015). Additionally, Kraft K et al. (2015) applied CRISPR/Cas9 technology to generate SVs directly in mouse embryonic stem cells. All these aspects, together with the fact that sgRNA sequences can be simply ordered as oligonucleotides and cloned into a vector with the *Cas9* transgene, underline the great advantages of CRISPR-Cas9 genome editing in terms of time and effort, offering a significant improvement for the study of SVs in human diseases.

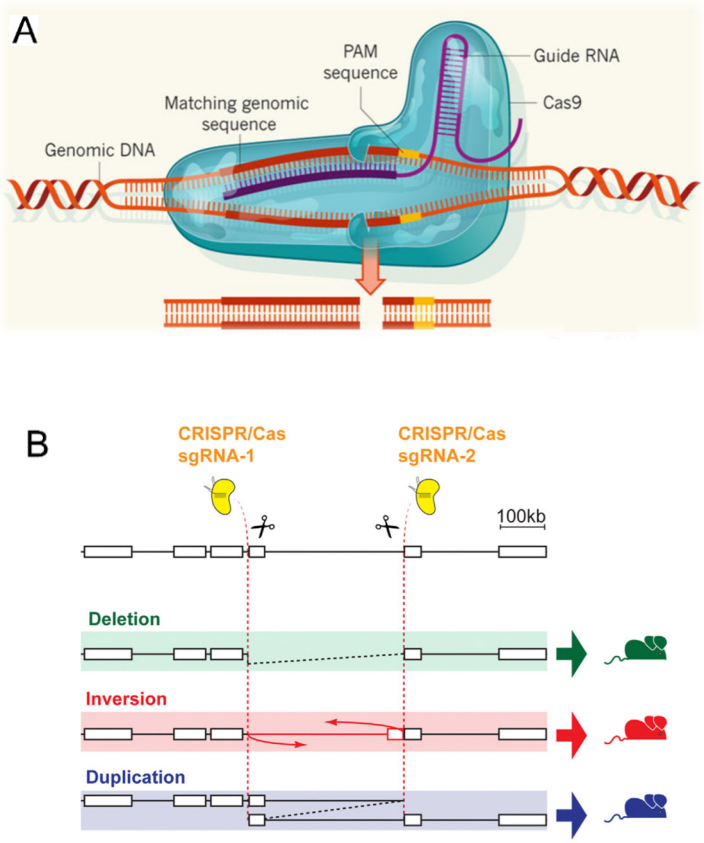


Figure 7. CRISPR-Cas9 genome editing. (**A**) The endonuclease Cas9, guided by a sgRNA, is able to create DSBs in a targeted genomic region demarcated by a PAM sequence. DSBs are then usually repaired by the NHEJ mechanism (Charpentier E and Doudna JA 2013). (**B**) Concurrent use of two sgRNAs can lead to the generation of small up to very large deletions, inversions and duplications (Kraft K et al. 2015).

2. AIMS OF THE RESEARCH

Fibroblast Growth Factor 8 (*FGF8*) gene codes for a signalling molecule that plays an important role during patterning, development, growth and organogenesis of many structures and organs throughout embryonic development like the kidney, brain and limb bud. To achieve correct development of these structures it is essential that *FGF8* is expressed in a timely and precise manner. The expression of this factor is regulated by an array of *cis*-regulatory elements that act as an integrated unit and are located in the genomic neighbourhood of the *FGF8* gene. This *FGF8* regulatory region with its numerous enhancers interspersed with other genes was conserved throughout the evolution hinting at the importance of the internal structure of this region for the *FGF8* regulation.

In this study we focus on overlapping tandem duplications encompassing this regulatory region and causing limb malformations in human patients. Most notably, duplications in this region are linked to the Split Hand/Foot Malformation 3 (SHFM3), which is characterized by absence of the central digits on hands and feet, variably associated with mental retardation, craniofacial findings and intellectual disability. So far, the molecular pathomechinsm through which these overlapping duplications cause the SHFM3 and other phenotypic features, as well as how they affect the organisation of the *FGF8* regulatory region, is still unknown. Therefore, we first aim to remodel these structural variations in mice to investigate the molecular pathomechanism underlying these malformations. Second, we aim to uncover the spatial and functional organization of the *FGF8* locus in mice in order to evaluate the impact of these duplications on local chromatin architecture of the *FGF8* locus and on gene regulation.

To investigate these questions this study utilized a combination of state-of-the-art genetics and genomics approaches like CRISPR/Cas9 genome editing, locus specific chromatin interaction mapping in high-resolution (cHi-C) and phenotypic functional analyses.

3. MATERIALS

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 J and Russel DW (2001) and Green MR and Sambrook J (2012).

Table 1: Buffers for Capture Hi-C.

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 for Whole Mount *In Situ* Hybridization

Buffer	Composition
Alkaline phosphatase	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 H2O
10x PBS (DEPC)	1.37 M NaCl, 27 mM KCl, 100 mM Na2HPO4, 20mM KH2PO4, adjust pH to 7.4 with HCl, in DEPC- H2O, autoclave
4% PFA/PBS	Dissolve 40 mg/ml PFA in 1x PBS (DEPC), heat to 55°C until PFA is dissolved, adjust pH to 7.4 with HCl
PBST	0.1 % TWEEN-20 in 1x PBS(DEPC)
RIPA buffer	Use DEPC treated reagents, 0.01 % SDS, 0.15 M NaCl, 0.01 % Nonidet-P40, 5 mg/ml deoxycholate, 1 mM EDTA pH 8.0, 50 mM Tris pH 8.0, in DEPC- H2O
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 H2O
5x MABT	0.5 M maleic acid pH 7.5, 0.75 M NaCl, 0.5 % TWEEN-20, in H2O
PBST/ levamisole	0.5 mg/ml tetramisole hydrochloride in PBST
Proteinase K buffer	20 mM Tris pH 7.0, 1 mM EDTA, in DEPC-H2O
20x SSC (DEPC)	3 M NaCl, 300 mM Na3Citrate x 2H2O, 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 Cell culture

Table 3: Cell culture ingredients.

Name	Supplier
100x Glutamin	Lonza BE17-605E
1x PBS	Lonza BE17-512F
100x Penicillin/Streptomycin	Lonza DE17-603
Beta-Mercaptoethanol	Sigma M-7522
Bicarbonate free Media	Gibco 52100
DMEM	Lonza BE12-733
DMSO	Sigma Aldrich D-2650
FCS for feeders	Biochrome
FCS for ESCs	PAN Biotech P122011
Fugene	Promega TM-238
Gelatine	Sigma G-1393
KO-DMEM	Gibco 10829-018
Non-essential Amino acids	Gibco 11140-35
LIF	Chemicon ESG1107
OptiMEM	Gibco
Trypsin	Gibco 25300-054

3.4 Kits

Standard procedures were conducted following manufacturer's instructions using kits in Table 4.

Table 4: Molecular biology kits used in this study.

Purpose	Kit	Supplier
Plasmid DNA purification	NucleoSpin Plasmid	Macherey-Nagel, Düren
Sanger Sequencing	BigDye Terminator v3.1	Applied Biosystems, Foster City, USA
qRT-PCR	GoTaq qPCR Master Mix	Promega

3.5 Enzymes

If not stated otherwise, restriction enzymes, T4 DNA ligase 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 pX459 vector (pSpCas9(BB)-2A-Puro, Addgene) was used for the CRISPR/Cas9 genome editing.

3.8 Primers

All primers were synthesized by Eurofins MWG Biotech AG (Ebersberg) and using HPSF (High Purity Salt Free) purification. All primer sequences are shown in 5' to 3' orientation in Table 5.

Table 5: Primers for single guides RNA generation and for guides RNA recognition in the pX459 vector.

Name	Sequence	Task
CR-1-L87-f1	**CACCG**TCTTCATCTGAGGATCAGCG	Guide RNA for Dup1 (centromeric breakpoint)
CR-1-L87-r1	**AAAC**CGCTGATCCTCAGATGAAGA**C**	Guide RNA for Dup1 (centromeric breakpoint)
CR-1-2-R77-f1	**CACCG**TTAAATTTCTCCAAGATAAG	Guide RNA for Dup1 and 2 (centromeric breakpoint)
CR-1-2-R77-r1	**AAAC**CTTATCTTGGAGAAATTTAA**C**	Guide RNA for Dup1 and 2 (centromeric breakpoint)
MS guide gFGF8_5pF	**CACCG**CTGACATTTAGCTGTGATCG	Guide RNA for Dup2 (centromeric breakpoint)
MS guide gFGF8_5pR	**AAAC**CGATCACAGCTAAATGTCAG**C**	Guide RNA for Dup2 (centromeric breakpoint)
CR-Seq-f1	GAAAGTAATAATTTCTTGGGTAGTTTGCAG	RNA guide in pX459
CR-Seq-r2	GCTCTAAAACAAAAAAGCACCGACTC	RNA guide in pX459

Table 6: Primer sequences for genotyping PCR of clones and mice and for qPCR

Name	Sequence	Task
genoCR-1-L87-f1	TCAGGCGATGTCCTTCCTAC	Genotyping PCR Dup1 (T1)
genoCR-1-L87-r1	CCTGAGGCCATGCTAAGGTA	Genotyping PCR Dup1 (T2)
genoCR-1-2-R77-f1	TTGATCAGGGGGTATGGTGT	Genotyping PCR Dup1 and 2 (T3)
genoCR-1-2-R77-r1	TCTGTTATTGGCGTGCAGTC	Genotyping PCR Dup1 and 2 (T4)
genoCR-2-MS-f1	ACTGCCCACATTTTGTAGCC	Genotyping PCR Dup2 (T1)
genoCR-2-MS-r1	CCTCCCAATCCTAGGGCTAC	Genotyping PCR Dup2 (T2)
qPCR-CR1-5flanking-f1	GGCCCTGATAGTCCCTCAAT	qPCR Dup1 centromeric out
qPCR-CR1-5flanking-r1	GACGGTCACCACATCTTCCT	qPCR Dup1 centromeric out
qPCR-CR1-5in-f1	CACAAATCGTTTTCTTTGTCCA	qPCR Dup1 centromeric in
qPCR-CR1-5in-r1	TTAGCATCATTGAACGACATCC	qPCR Dup1 centromeric in
qPCR-CR1-2-Ce1-f2	TCATAGCCGCCAGAATAAGG	qPCR Dup1 and 2 Center 1
qPCR-CR1-2-Ce1-r2	GCAGACTTTGGCCTGTTTGT	qPCR Dup1 and 2 Center 1
qPCR-CR1-2-Ce2-f2	TTCCCTGGGTGTTGTGTTGT	qPCR Dup1 and 2 Center 2
qPCR-CR1-2-Ce2-r2	TGGGGCTGCAGCTATAACTC	qPCR Dup1 and 2 Center 2
qPCR-CR1-2-Ce3-f2	CCATCCCAGAGGCATACACT	qPCR Dup1 and 2 Center 3
qPCR-CR1-2-Ce3-r2	CTACAGACAGACGGCAACCA	qPCR Dup1 and 2 Center 3
qPCR-CR1-2-3in-f1	GTCTGTCGCACAGGACACAG	qPCR Dup1 telomeric in
qPCR-CR1-2-3in-r1	AACTTCCGGAGAGCATTCCT	qPCR Dup1 telomeric in
qPCR-CR1-2-3flanking-f1	TCCAGCTCCAGACACTGAAC	qPCR Dup1 and 2 telomeric out
qPCR-CR1-2-3flanking-r1	CGAAGCCTGTGACACTTGTT	qPCR Dup1 and 2 telomeric out
qPCR-CR2-5flanking-f1	ATAACAATGACCGCCCTGAC	qPCR Dup2 telomeric out
qPCR-CR2-5flanking-r1	TGTGTTTCAGGTCCCTTTCC	qPCR Dup2 telomeric out
qPCR-CR2-5in-f1	GACCGGGGAGAGTCTAAGGT	qPCR Dup2 telomeric in
qPCR-CR2-5in-r1	CCTCCCAATCCTAGGGCTAC	qPCR Dup2 telomeric in
Control region 672	AGCTAGATTACCCTGAGTCCA	Control region
Control region 673	TTCAAGTAGGCTCGGTCACC	Control region

3.9 Instruments

Table 7: 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

3.10 Software and Internet resources

Table 8: Software and internet resources used in this study.

Name	Supplier/Web address	Application
AxioVision Rel.4.8	Zeiss	Microscopy, digital photography
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
CRISPR Design	http://www.genome-engineering.org/crispr/	RNA guides design
Ensembl Genome Browser	http://www.ensembl.org/index.html	Primer analysis
MGI – Mouse Genome Informatics	http://www.informatics.jax.org/marker/	Resource for mouse data
Bin Ren Hi-C data	http://yuelab.org/hi-c/database.php	Resource for mouse and human Hi-C data
ApE – A plasmid Editor	*en.bio-soft.net/plasmid/ApE.html*	Construct design and Sequence analysis

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 J and Russel DW (2001) and Green MR and Sambrook J (2012).

4.2 DNA isolation

4.2.1 Isolation of plasmid DNA

Isolation of plasmid DNA was performed with Nucleospin Plasmid kit (Macherey-Nagel) according to the instructions of the manufacturer.

4.2.2 Isolation of genomic DNA

DNA from embryonic material was isolated by conventional DNA precipitation procedures. Briefly, tissue was lysed in 300 µl Tissue-Lysation Buffer (17 mM Tris, pH 7.5; 17 mM EDTA; 170 mM NaCl, 0.85 % SDS) with freshly added 0.2 µg/ml proteinase K at 55°C for 1 to 12 hours under agitation in a Thermomixer. To remove any insoluble components, the lysate was centrifuged at full speed at room temperature for 10 min. Supernatant was transferred to new Eppendorf tube and mixed with 350 µl Isopropanol for precipitation. After centrifugation for 10min at full speed at room temperature the supernatant was removed and the DNA pellet was washed twice with 70% Ethanol. The DNA was air dried for 5 min, dissolved in water and stored at -20°C.

4.3 Cloning of single guide RNAs for CRISPR/Cas9 genome editing

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 PD 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 listed in Table 5. 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 Cas9 and the ampicillin and puromycin resistance genes. 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 *E.Coli* Top10 bacteria. Finally, the plasmid DNA was purified from 5 ml cultures and successful cloning was validated by Sanger sequencing using vector specific primers (CR-Seq-f1 and CR-Seq-r2 in Table 5).

4.4 Cell culture

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

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

4.4.2 Culturing feeder cells/ primary embryonic fibroblasts

Feeder cells were cultured in regular Dulbecco's Modified Eagle's Medium (DMEM) containing 4,500 mg/ml glucose, without sodium pyruvate (Lonza, #BE12-733F), supplemented with 10% regular fetal calf serum (FCS Superior, Biochrom, #S0615), 1x glutamine (100x, Lonza, #BE17-605E) and 1x penicillin/streptomycin (100x, Lonza, #DE17-603). Feeder cells were prepared from E13.5-14.5 CD1 (Rice MC and O'Brien S 1980) and DR4 mouse (Tucker KL 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) for mitotic inactivation and frozen in cryovials at a density of 2,5 x 10^6 cells/vial. The freezing medium consisted of regular feeder medium containing 20% FCS and 20% DMSO (Sigma, #D-2650).

4.4.3 Culturing mouse embryonic stem cells (mESCs)

G4 ES cells (129/Sv x C57BL/6 F1 hybrid, (George SHL et al. 2007)) were cultured onto feeder plates or wells. 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 x 10^4 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) containing 4,500 mg/ml glucose, supplemented 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). ES cell medium was changed every 24 hours and cells were frozen at a density of 1 x 10^6 cells/ vial in freezing medium, consisting of regular ES cell medium supplemented with 20 % FCS and 20% DMSO (Sigma, #D-2650).

4.4.4 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 x 10^6 G4 ES cells/well were plated onto the feeder cells. The following day, 8 µg of each pX459-Vector containing a guide RNA were transfected using FUGENE HD reagent (Promega, #E2311) according to manufacturer´s instructions. In particular, two pX459 sgRNAs were co-transfected for each desired mutation. After two days, ES cells were split onto four 6 cm dishes with puromycin resistant DR4 feeder cells with 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 manually 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 disaggregated and transferred to a new 96-well plate with regular ES cell medium and CD1 feeder cells. Cells were grown for two days, trypsinized and two thirds were frozen in 96-well plates containing ES cell freezing medium. The remaining cells were

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, clones with appropriate mutations were expanded and used for generating mouse embryos and live animals from ES cells by di- or tetraploid complementation (Artus J and Hadjantonakis AK 2011).

4.5 Genotyping of clones and mutant murine embryos

Clones and mutants were genotyped by standard PCR procedures using Taq and Pfu polymerases produced by Asita Stiege in the research group. Reagents were pipetted on ice into a 1.5 ml reaction tube and DNA was amplified in a thermocycler. PCR conditions for 25 µl reaction were the following: 2.5 µl 10x Taq buffer (750 mM Tris/ HCl pH 8.8; 200 mM (NH4)2 SO4; 0.1% Tween 20; 15 mM MgCl2), 0.25 µl dNTPs (12.5mM), 0.5µl forward primer (10 µM), 0.5 µl reverse primer (10 µM), 0.2 µl Taq enzyme, 0.05 µl Pfu enzyme, 1 µl template (~20 ng), 20 µl H2O bidest. The following PCR program was used: step 1: 95°C, 3 min; step 2: 95°C, 30 sec; step 3: 61°C, 45 sec; step 4: 72°C, 2 min; step 5: go to step 2 for 34x; step 6: 72°C, 7 min; step 7: 4°C, pause. Primer sequences for conventional genotyping are listed in Table 6. The PCR products were analysed on a 1% agarose gel.

4.5.1 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, 1 µl primerpairs (2.5 µM each) and 5µl gDNA (10ng). Each reaction was performed in triplicates. A standard curve for each primer pair was generated from gDNA that contained the target sequence. Each point corresponded to a different step of 1:5 dilution

series (1 – 0.2 – 0.04 – 0.008 – 0.0016). Only primers with an efficiency of >95% were used. Relative values for each target (primers within the duplicated region) were then normalized to the reference (primers outside the duplicated region). Primers are listed in Table 6.

4.6 Capture Hi-C

4.6.1 SureSelect design

The library of SureSelect enrichment probes was designed over the genomic interval chr19:44,440,000-46,400,000 (mm9) for mouse using the online tool of Agilent: SureDesign. Probes are covering the entire genomic region and were not designed specifically in proximity of *DpnII* sites.

4.6.2 Crosslinking and nuclei extraction

Embryonic tissue (limb buds) was prepared from E11.5 mouse embryos in 1x PBS. The tissue was dissociated by trypsin treatment for 10 min at 37°C, pipetting every 2 minutes to obtain a single-cell suspension. Treatment was stopped by adding 5x volume of 10% FCS/ PBS. To remove cell debris, the solution was filtered through 40 µm cell strainer (Corning, #352340). Cells were centrifuged at 1100 rpm for 5 min and the obtained pellet was then resuspended in 5 ml 10% FCS/ PBS. Crosslinking was initiated by adding 5 ml of freshly prepared 4% formaldehyde in 10% FCS/ PBS (final concentration 2%). 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 x 10^6 nuclei. After centrifugation at 2600 rpm for 2 min, nuclei were snap-frozen in liquid nitrogen. Samples were stored at -80°C until further processing for a maximum of 6 months.

4.6.3 Preparation of 3C library

The nuclei pellet was resuspended in 360 µl of water and mixed with 60 µl of 10x restriction buffer. 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 *DpnII* 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 decrosslinked 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 analysed 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 analysed on an agarose gel check for successful ligation.

The preparation of this ligation control was done as described for the digestion control. DNA of final 3C library was de-crosslinked by adding 30 µl of proteinase K (10 mg/ml) and incubating 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. DNA was precipitated by adding the following reagents 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 capture Hi-C.

4.6.4 Capture Hi-C library preparation and sequencing

3C libraries were sheared using a Covaris sonicator. 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 bp paired-end) following Agilent instructions. Samples were 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)

4.6.5 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 S et al. 2015).

The pipeline used Bowtie2 v2.2.6 (Langmead B and Salzberg SL 2012) for mapping short reads to reference genome (NCBI37/mm9). Filtered di-tags were further processed with Juicebox (Rao SSP 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 chr19:44,440,000-46,400,000 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, ditags were filtered for the enriched region and a custom chromosome file containing only the enriched region on chr19 was used for the Juicebox tool (Rao et al. 2014; Durand et al. 2016). Hi-C maps and subtraction maps were visualized with the HiC2-Viewer browser (licensed by Robert Schöpflin).

4.6.6 CTCF motif analysis

For analysis of CTCF motif orientation in ChIP-seq peaks (E14.5 mouse embryonic limbs for CTCF ChIP-seq from ENCODE/LICR), the FIMO algorithm of the MEME suite (Bailey TL 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).

4.7 Whole mount *In Situ* Hybridization (WISH)

Wild-type and mutant embryos at E11.5 were subjected to whole mount *in situ* hybridization using *Fgf8* and *Lbx1* probes to detect mRNA expression. 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.

All solutions used for whole-mount in situ hybridization are listed in Table 2. Buffers and solutions were treated with DEPC to inactivate RNase enzymes. 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) and stored at -20°C.

Prior to hybridization, embryos were rehydrated in 75%, 50% and 25% Methanol/PBST and washed twice with PBST. Subsequently, embryos were bleached in 6% hydrogen peroxide/PBST for 1 hour on ice and washed in PBST. Embryos were further treated with Proteinase K for 3 to 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 Blocking 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.

4.8 Histology

4.8.1 Skeletal preparations

Embryos at E18.5 were sacrificed and kept in water for 1 hour. Skin was removed manually after heat treatment for 1 min at 65°C. Animals were skinned, emptied of all the inner organs 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 24 hours. The animals were rinsed and post-fixed in 100% Ethanol overnight at room temperature. For initial clearing, animals were incubated for up to 24h in 0.2% KOH/ bidest H2O. Membranous bones were stained red using Alizarin Red staining solution (50 mg/l Alizarin Red (Sigma-Aldrich, #A5533) in 0.2% KOH/ bidest H2O). This staining was performed for up to 2 days with visual inspection of each sample until proper red staining was reached. Subsequently, remaining tissue was digested for up to 3 days with 0.2% KOH/bidest H2O. Clearing was stopped by placing preparations to increasing glycerin solutions (30%, 60%, and 80% glycerin/bidest H2O, for 24 hours each) and then stored in 80% glycerin.

5. RESULTS

5.1 Chromatin organization and structural variations at the *FGF8* locus

5.1.1 Using capture Hi-C to determine the TAD structure at the *Fgf8* locus

Fgf8 is located in a gene dense region that contains several enhancers working as an integrated unit, regulating *Fgf8* expression (Marinić M et al. 2013). Additionally to the *cis*-regulatory elements, chromatin organization has been shown to contribute to the establishment of preferential enhancer-gene promoter contacts. Such specific enhancer-promoter contacts are often constrained to the same unit of preferential interactions, the so-called topologically associated domain (TAD) (Dixon JR et al. 2012; Phillips-Cremins JE et al. 2013). In order to investigate the TAD structure and interactions at the *Fgf8* locus, capture Hi-C (cHi-C) was performed over a 2 Mb region (chr19:44,440,000-46,400,000, mm9) in E11.5 murine limb buds (forelimbs and hindlimbs collected together), where *Fgf8* is normally expressed. Using cHi-C, a high-resolution heat map was obtained, where each point in the triangular matrix represents the interaction between two DNA fragments over the linear locus (Figure 8a). The shades of red correspond to the frequency of interaction: the higher the interaction frequency, the darker the shade of red. Therefore, light red and white areas corresponded to low and absence of interactions, respectively.

The cHi-C map at the *Fgf8* locus revealed the presence of two TADs (black dashed lines in Figure 8a), that showed high intra-TAD interaction frequency and low or no contacts between the two TADs. One TAD, centromeric to the captured region and hereafter referred to as *Lbx1* TAD, contains *Lbx1* and *Btrc*. The second TAD, telomeric to the *Lbx1* TAD, encompasses *Poll, Dpcd, Fbxw4* and *Fgf8* itself (hereafter referred to as *Fgf8* TAD). Additionally, the boundary regions of both TADs exhibited high interaction frequencies, which appeared as red dots in the interaction matrix (black arrows in Figure 8a).

Interestingly, these interactions correlated with CTCF binding in the limb tissue at E11.5 (Figure 8b). CTCF, together with cohesin, is important for the formation of TAD structure (Rao SSP et al. 2014; Nora EP et al. 2017; Schwarzer W et al. 2017), while the Mediator complex facilitate long-range interactions between enhancers and promoters (Kagey MH et al. 2010). Furthermore, the directionality of the CTCF binding motifs has been shown to have an important role, as they are required to have convergent orientation for the proper formation of chromatin loops, meaning that one motif has to be in forward orientation and the other in a reverse orientation (Rao SSP et al., 2014; de Wit E et al. 2015; Guo Y et al. 2015). Since several CTCF-binding sites at TAD boundaries have been observed at the *Fgf8* locus (Figure 8b), their relative orientation was investigated (Bailey et al. 2009). This analysis uncovered that for each of the two TADs at least one couple of recognition motifs is present in a convergent orientation (Figure 8c).

Furthermore, the centromeric *Lbx1* TAD boundary and the telomeric *Fgf8* TAD boundary appeared to strongly contact each other (green arrow in Figure 8a), resulting in the formation of an additional loop encompassing the *Lbx1* and *Fgf8* TADs. However, the amount of interactions between the *Lbx1* and *Fgf8* TADs did not increase, as seen instead for each TAD individually. According to the loop extrusion model (Sanborn AL et al. 2015), strong CTCF binding sites can lead to the formation of further loops, but overlapping loops cannot coexist within the same cell. Rather, they represent an alternative folding characteristic of a subset of cells within the examined tissue (Sanborn AL et al. 2015; Giorgetti L et al. 2014), as it is probably the case for the overlapping loop observed between the centromeric *Lbx1* and telomeric *Fgf8* TAD boundaries.

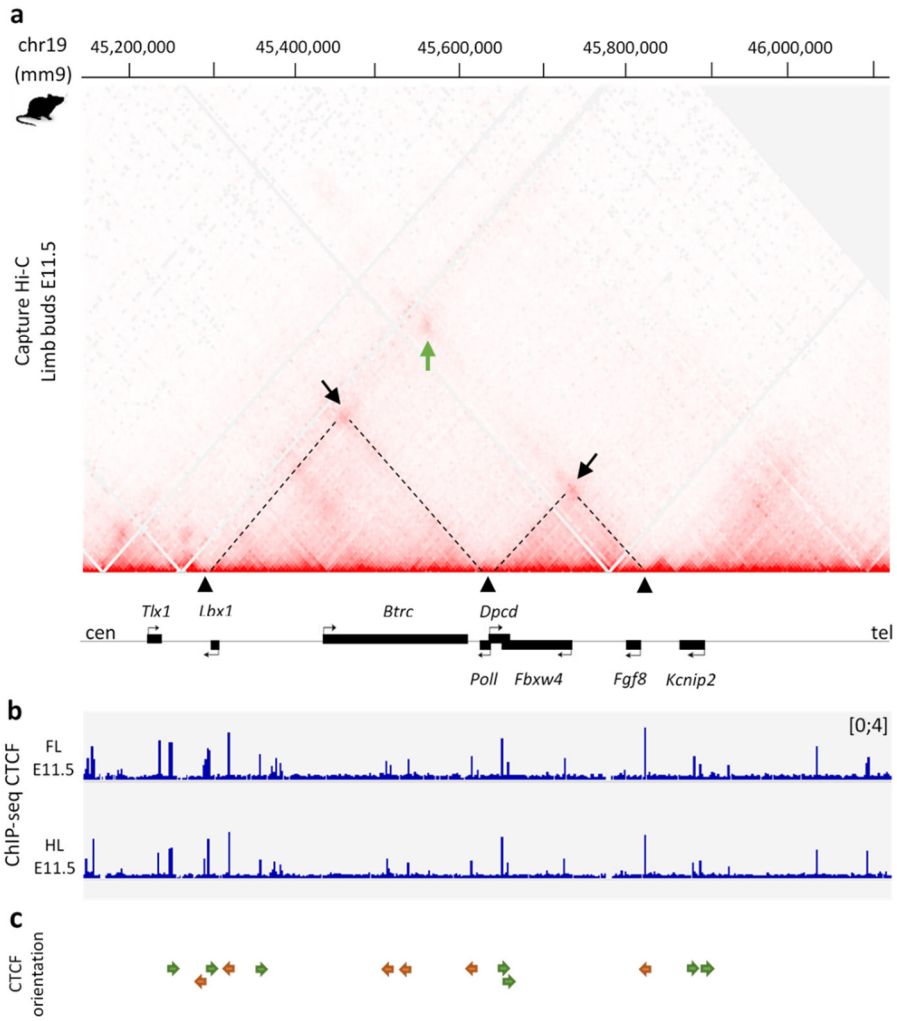

a
chr19
(mm9)
45,200,000
45,400,000
45,600,000
45,800,000
46,000,000
Capture Hi-C
Limb buds E11.5
Tlx1
Lhx1
Btrc
Dpcd
cen
tel
Poll
Fbxw4
Fgf8
Kcnip2
b
[0;4]
ChIP-seq CTCF
FL
E11.5
HL
E11.5
c
CTCF
orientation

Figure 8: Spatial organization at the *Fgf8* locus. (**a**) cHi-C interaction map over the *Fgf8* locus from wild-type E11.5 limb buds is shown. The interaction profile revealed the presence of two TADs (triangular matrices indicated with black dashed lines), which are characterized by a high level of interactions. The interactions are restricted by TAD boundaries (black arrowheads), which highly interact as shown by the red dots in the interaction matrix (black arrows). A schematic representation of the genes below the map indicate that the one TAD on the centromeric side includes *Lbx1* and *Btrc* genes (referred to as *Lbx1* TAD), while the TAD more telomeric encompasses *Poll, Dpcd, Fbxw4* and *Fgf8* genes (referred to as *Fgf8* TAD). Further contacts between the centromeric *Lbx1* and telomeric *Fgf8* TAD boundaries were also detected (green arrow). (**b**,**c**) ChIP-seq signals for CTCF from limbs at E11.5 (dataset produced by Ivana Jerković in the Mundlos laboratory) and orientation of CTCF-binding sites (green arrows indicate forward orientation, red arrows indicate reverse orientation). CTCF peaks and the convergent orientation of some binding sites correlate with the observed TAD boundary interactions in the cHi-C map.

5.1.2 TAD structure conservation between mouse and human

Previous studies demonstrated that TAD organization is conserved between species and across cell types (Dixon et al. 2012; Vietri Rudan M et al. 2015). In order to investigate the conservation of chromatin interactions at the *Fgf8* locus, a comparison of murine and human TAD structure was performed. Hi-C data derived from the human lymphoblastoid cell line GM12878 was obtained from publicly available database (http://yuelab.org/hi-c/database.php) and compared to the generated high-resolution cHi-C heat map of mouse limb buds at E11.5. The GM12878 dataset was chosen as it contains a sufficient number of pairwise interactions for comparing interaction matrices at 5 kb resolution. Despite differences in the genomic size between mouse and human, a high level of synteny between the two loci was observed (Figure 9). Furthermore, the TAD structure previously described in mouse, as well as the regions corresponding to TAD boundaries (arrowheads in Figure 9) and their interaction frequency (arrows in Figure 9), showed a similar structure and positions in human.

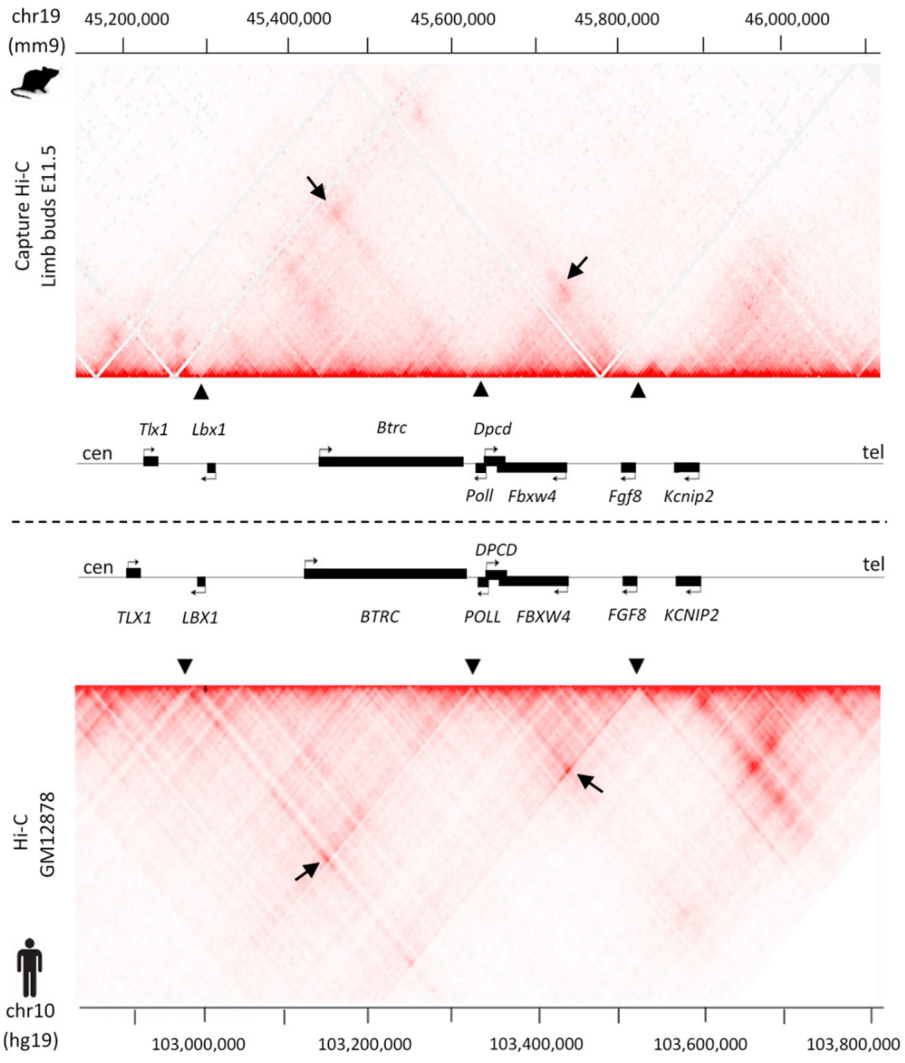

chr19
(mm9)
45,200,000
45,400,000
45,600,000
45,800,000
46,000,000
Capture Hi-C
Limb buds E11.5
Tlx1
Lbx1
Btrc
Dpcd
cen
tel
Poll
Fbxw4
Fgf8
Kcnip2
DPCD
cen
tel
TLX1
LBX1
BTRC
POLL
FBXW4
FGF8
KCNIP2
Hi-C
GM12878
chr10
(hg19)
103,000,000
103,200,000
103,400,000
103,600,000
103,800,000

Figure 9: TAD structure at the *Fgf8* locus is conserved between humans and mice. Comparison between the cHi-C heat map at the *Fgf8 l*ocus from E11.5 limb buds from wild-type mice (top) and the Hi-C heat map at the *FGF8* locus from human lymphoblastoid cell line GM12878 (bottom). A high degree of synteny between the two loci and conservation of the TAD structure was observed between humans and mice, as indicated by the conservation of the TAD boundaries (black arrowheads) and relative interaction frequency (black arrows).

5.1.3 Inter-TAD duplications at the *FGF8* locus

Several tandem duplications on chromosome 10q24 at the *FGF8* locus have been linked to the Split-Hand/Foot Malformation type 3 (SHFM3) by previous studies (de Mollerat XJ et al. 2003; Kano H et al. 2005; Dimitrov BI et al. 2010). SHFM3 is a limb malformation mostly characterized by absence of the central digits on hands and/or feet. Some patients, in concomitance or not with the limb phenotype, also exhibit mental retardation, intellectual disability and/or craniofacial abnormalities, such as micrognathia (de Mollerat XJ et al. 2003; Elliot AM and Evans JA 2006; Dimitrov BI et al. 2010).

Tandem duplications largely overlapping the *FGF8* locus were identified in several patients at Charitè Universitätmedizin Berlin and University of Pavia, characterized through array-CGH (Microarray-based Comparative Genomic Hybridization) and the breakpoints confirmed by sequencing. All duplications overlap to a large extent (green dashed lines in Figure 10) and all include the entire *LBX1*, *BTRC*, *POLL* and *DPCD* genes. In some cases, they also encompass *FBXW4* or part of it, and/or *TLX1*, but none cover *FGF8* (Figure 10). For all duplications, except one, patients exhibited the classical SHFM phenotype with aplasia of the central digits on hands and/or feet (yellow and black bars in Figure 10). The patient without classical split-hand/foot phenotype was characterized by mild intellectual disability only, despite the complete overlap with the duplications of the other patients (grey bar in Figure 10). Analysis of duplications position in the context of TAD structure organization revealed that the identified duplications are inter-TAD and all encompass the TAD boundary between *FGF8* and *LBX1* TADs (light blue arrowhead in Figure 10).

The majority of these inter-TAD duplications include regions of the *FGF8* TAD and almost the entire *LBX1* TAD (Figure 10). As shown by recent studies (Franke M et al. 2016; Weischenfeldt J et al 2017), inter-TAD duplications can be responsible for changes and reorganization of the chromatin structure, creating new chromatin domains (neo-TADs) that include genes and *cis*-regulatory elements normally located in different and separate TADs, therefore promoting their interactions. This can have consequences on gene regulation and cause disease (Franke M et al. 2016; Weischenfeldt J et al 2017). Thus, the here identified inter-TAD duplications could induce rearrangements of the TAD structure leading to the formation of a neo-TAD containing elements from both *LBX1* and *FGF8* TADs. Consequently, alteration of the proper gene regulation could eventually lead to the phenotype observed in SHFM3, whose precise mechanism is still unknown (Lyle R et al. 2006; Dimitrov BI et al. 2010). In order to understand the molecular mechanism underlining the identified SHFM3 malformations at the *FGF8* locus and to determine the impact of these structural variations on TAD structure and gene regulation, one of the duplications (yellow bar in Figure 10) was reengineered in mice. As the TAD structure at the *FGF8* locus is well conserved between mouse and human, reengineering duplications in mice is an adequate and compelling approach to study structural variations occurring in humans.

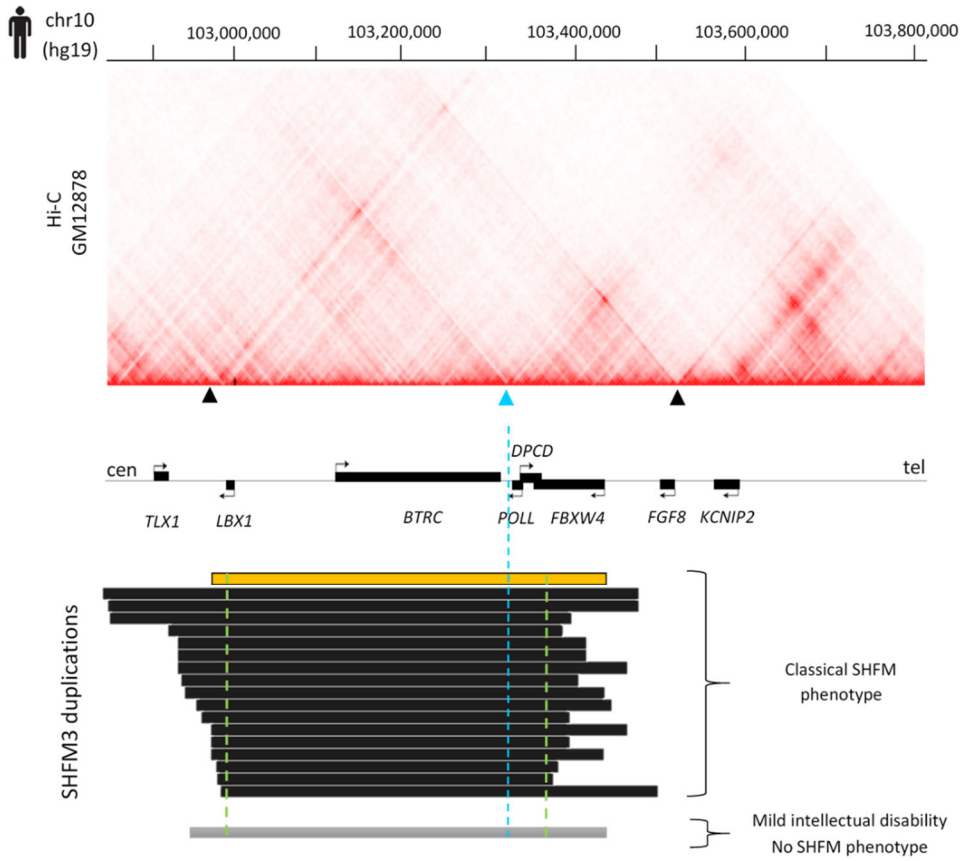


Figure 10: Duplications at the *FGF8* locus encompassing regions of the *LBX1* and *FGF8* TADs. Schematic representation of the identified SHFM3 duplications relative to the TAD structure at the locus. Arrowheads at the bottom of the heat map indicate TAD boundary positions, illustrating that duplications span parts of the *FGF8* TAD, *LBX1* TAD and, especially, the TAD boundary between *FGF8* and *LBX1* TADs (light blue arrowhead). The grey bar represents the duplication carried by the patient affected only by mild intellectual disability and no split-hand/foot phenotype. The yellow bar indicates the duplication reengineered in mice. The green dashed lines delimit the largest area of overlapping of duplications.

5.2 Generation of structural variations in mice using CRISPR/Cas9 genome editing

As mentioned in the previous section, one of the identified SHFM3 duplications (yellow bar in Figure 10), of approximately 485 kb and also spanning the largest overlap of all duplications (green dashed lines in Figure 10), was chosen to be reengineered in mice using CRISPR/Cas9 genome editing. Figure 11 illustrates the corresponding region of the duplication on murine chromosome 19 (yellow bar with black outlines in mm9). Due to the smaller size of the mouse genome, the duplication on chromosome 19 corresponds to a region of approximately 450 kb. To engineer this duplication two guide RNAs were designed to target the murine duplication breakpoints. One guide RNA targeted the centromeric breakpoint, which was almost identical to the human breakpoint, close to *Lbx1* gene (Figure 11). The second guide RNA targeted instead the telomeric breakpoint, which, however, was extended further telomeric, towards the *Fgf8* gene (Figure 11), to avoid the disruption of *Fbxw4* and *cis*-regulatory elements close to *Fbxw4*. Therefore, the engineered duplication, hereafter referred to as Dup1 (yellow bar in figure 11), resulted slightly larger (approximately 500 kb) than the original duplication. Furthermore, a second duplication, hereafter referred to as Dup2 (red bar in Figure 11), was generated to exclude *Lbx1* from the duplication, as one of the genes comprised in all the rearrangements and therefore potentially involved in the pathomechanism. Dup2 was generated using the same telomeric guide RNA as in Dup1, while a new centromeric guide RNA was designed to exclude *Lbx1* from the duplication (red scissor in Figure 11). Design, cloning and mouse embryonic stem cells (mESCs) transfection of the single guide RNAs were done as described previously in section 4.3 and 4.4.4, following the CRISPR/Cas9 genome editing protocol according to Kraft K *et al.* 2015. The guide RNAs used for genome editing are listed in Table 5.

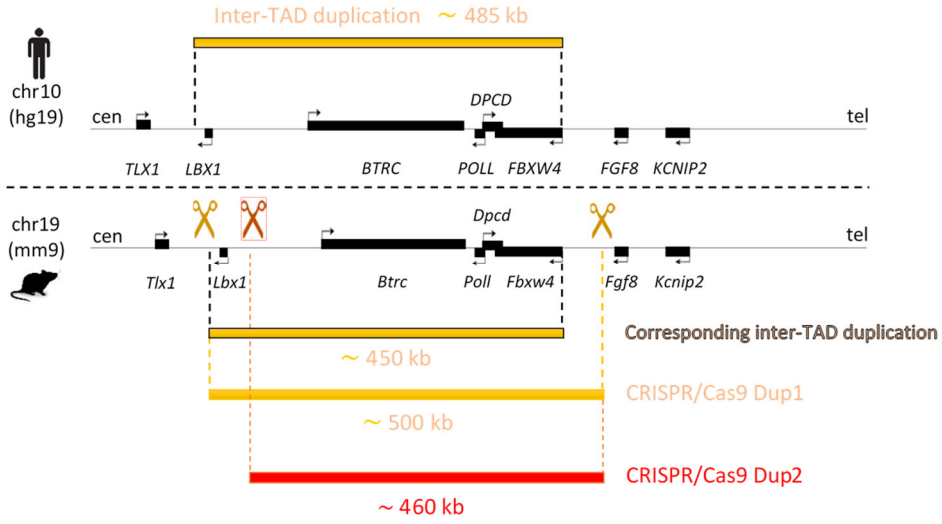


Figure 11: CRISPR/Cas9 genome editing at the *Fgf8* locus. Schematic representation of the SHFM3 duplication in the human genome and at the corresponding murine locus (yellow bars with black outlines at the top and at the bottom of the schematic locus representation, respectively). Yellow scissors represent the Cas9 enzymes for which two guide RNAs were designed to reengineer the corresponding human duplication breakpoints in mice (CRISPR/Cas9 Dup1 yellow bar). The same telomeric breakpoint was used also for Dup2 (CRISPR/Cas9 Dup2 red bar), while the centromeric breakpoint was shifted more telomeric compared to Dup1 to exclude *Lbx1* from the duplication Dup2 (red scissor).

Approximately 500 clones were collected for both Dup1 and Dup2. Several rearrangements, such as duplications, deletions and inversions were detected by conventional PCR using a combination of primers at the centromeric and telomeric breakpoints (Figure 12). PCR primers for genotyping are listed in Table 6.

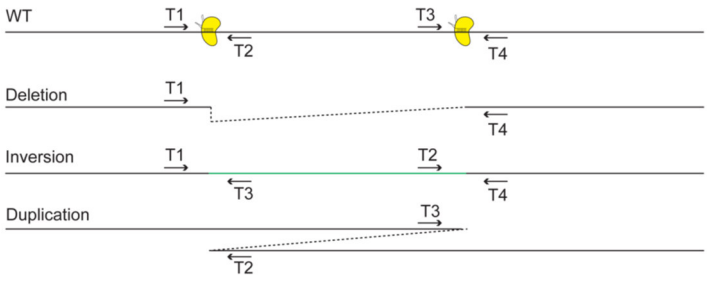


Figure 12: Genotyping PCR for CRISPR/Cas9 clones. A couple of primers is designed around each breakpoint in the wild-type sequence (T1+T2 and T3+T4). Then, using different combinations, deletions (T1+T4), inversion (T1+T3 and T2+T4) and duplications (T3+T2) can be detected. GuideRNA/Cas9 enzyme complexes are depicted in yellow (Kraft K et al. 2015).

Table 9 summarizes the total number of screened clones and the frequency of the detected rearrangements. Pure duplications, where pure indicates the presence only of one type of rearrangement, were found only in 1% of the clones for Dup1 and in 0.6% for Dup2. Clones carrying a pure deletion were 2% for Dup1 and 2.6% for Dup2, while the percentage of clones with pure inversion was 4% and 7.6%, respectively. Furthermore, a certain percentage of clones was characterized by multiple rearrangements. Specifically, 0.4% of Dup1 clones carried a duplication on one allele and a deletion or an inversion on the second allele. Or a deletion on the first allele and an inversion on the second allele. Similar or higher percentages were detected also among Dup2 clones.

	Coordinates (mm9)	Screened clones	Dup	Del	Inv	Dup/ Del	Dup/ Inv	Del/ Inv
Dup1	chr19:45,290,982-45,792,742	500	1%	2%	4%	0.4%	0.4%	0.4%
Dup2	chr19:45,332,272-45,792,742	500	0.6%	2%	7.6%	0.6%	0.4%	1%

Table 9: CRISPR/Cas9 results and frequency of several structural rearrangements. The genomic coordinates for Dup1 and Dup2 are indicated. The table depicts the total number of screened mESC clones and the percentage/frequency of different rearrangements obtained for each screen.

5.2.1 Validation of the copy number using Real-Time PCR

Dup1 and Dup2 clones positive for pure duplications based on the conventional PCR screening were further analysed with quantitative real-time PCR (qPCR) to confirm the genotype. To investigate the copy number, seven regions were selected for qPCR primer design (qPCR primers are listed in Table 6). In particular, five regions were chosen within the duplicated sequence (two closer to the centromeric and telomeric breakpoints, three more in the center). In addition, two flanking regions, one centromeric and one telomeric to the duplicated area, were selected as negative controls. Copy number analysis of Dup1 revealed no changes in the copy number compared to the wild-type, which was in contrast to the conventional PCR screening. This result suggested that the Dup1 ESC clones carried a deletion on one allele and a duplication on the other. In Figure 13a an exemplary qPCR of one of the clones is reported. Conversely, the analysis of Dup2 clones indicated changes in the copy number of the duplicated region with values up to 1,5 or more (Figure 13b illustrates one of the Dup2 clones analysed), suggesting that these clones carried the duplication in a heterozygous state and confirming the conventional genotyping.

Therefore, Dup2 ESC clones seemed to be characterized, based on both qualitative and quantitative PCR, by a pure duplication, while for Dup1 only clones carrying multiple rearrangements (duplication/deletion) were available. Thus, one of the Dup1 (duplication/deletion) and one of the Dup2 (pure duplication) clones were used for tetraploid complementation and embryos were then collected at different developmental stages to investigate changes in gene expression, to monitor the duplication-induced changes on the 3D organization and to investigate the phenotypic changes induced by these genomic rearrangements, particularly in the form of skeletal malformations.

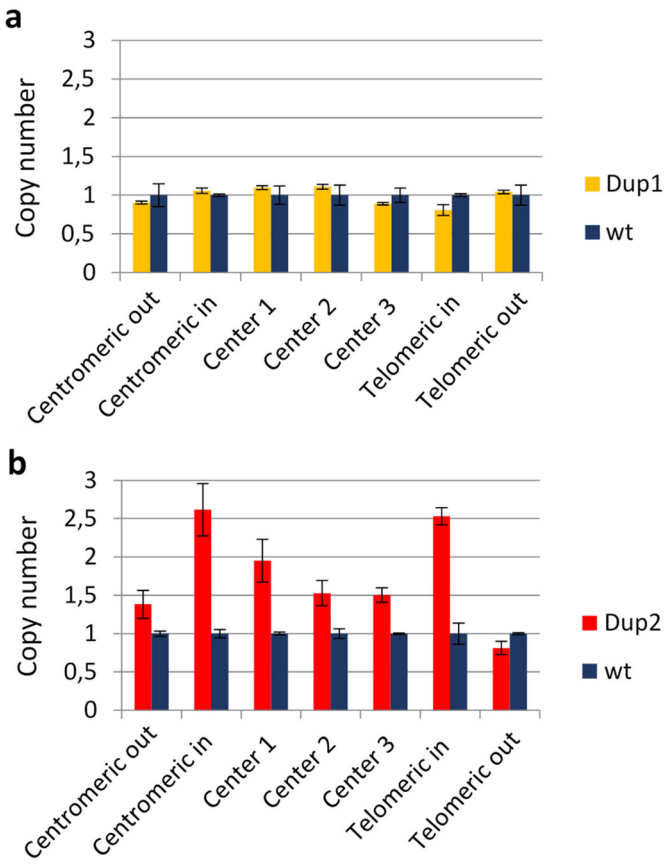


Figure 13: Copy number analysis of selected ESC clones. Clones positive for duplications were screened for copy number variation in order to confirm the genotyping observed in the conventional PCR analysis. Copy number detected in wt gDNA was set to 1. Seven regions were screened in total. Five within the duplicated area (Centromeric in and Telomeric in, closer to the centromeric and telomeric breakpoint, respectively. Center 1, Center 2 and Center3, localized more in the central part of the duplicated region) and two outside of the centromeric and telomeric breakpoints of the duplication (Centromeric out and Telomeric out) (**a**) Dup1 qPCR revealed the presence of a deletion on one allele and of a duplication on the other, as the bars of the graph regarding the duplicated area (Centromeric in, Center 1, Center 2, Center3 and Telomeric in) were all approximately at the same level as the wild-type bars, indicating no changes in the copy number. (**b**) Dup2 appeared to be a heterozygous duplication from the observation of the graph, as the bars relative to the duplicated area were characterized by values up to 1,5 or more.

5.3 Using Capture Hi-C to identify and characterize complex SVs

To study the effect of the inter-TAD duplications generated at the *Fgf8* locus, limb bud cells originating from Dup1 (duplication/deletion) and Dup2 (pure duplication) E11.5 embryos were processed for the cHi-C analysis. Generated interaction matrices were then compared to the wild-type map described earlier in section 5.1.1, as the cHi-C analysis in Dup1 and Dup2 embryos was performed over the same 2 Mb region (chr19:44,440,000-46,400,000, mm9). The comparison revealed major differences in the TAD structure for both Dup1 and Dup2.

The cHi-C heat map generated from Dup1 animals (Figure 14b) showed new interactions, as indicated by a strong red dot (black arrow in Figure 14b). Specifically, this new interaction represented a gain of contact between the breakpoint centromeric to *Lbx1* and the breakpoint centromeric to *Fgf8*. This result confirmed the duplication breakpoint position previously identified by the conventional PCR screening (T3+T2 in Figure 12).

To better visualize the effect of the rearrangement, a subtraction map was generated where the signal from the wild-type matrix was subtracted from the signal in the mutant matrix (Figure 14c). The subtraction analysis further revealed ectopic interactions of the regions involved in the duplication, i.e. parts of the *Lbx1* TAD, including *Lbx1*, and regions of the *Fgf8* TAD (red rectangular zone dense of interactions under the black arrow in Figure 14c and highlighted by dashed black lines). These observations confirmed that the mutant mice for Dup1 indeed carried a duplication, which recapitulates the inter-TAD duplication identified in patients. The duplication breakpoint and the pattern of ectopic chromatin interactions further suggest that a new chromatin domain (neo-TAD) was formed (Franke M et al. 2016). In this neo-TAD configuration *Lbx1* from *Lbx1* TAD is positioned closed to *cis*-regulatory region of *Fgf8* TAD. This regulatory region is located between *Btrc* and the area telomeric to *Fbxw4* (Marinić M et al. 2013) and, in the wild-type configuration, is usually separated from the *Lbx1* TAD by a TAD boundary (arrowhead in Figure 14a).

The cHi-C map additionally revealed the presence of a deletion allele, since the mutant Dup1 mice were generated from a ESC clone carrying a duplication and a deletion. This was evident by an additional breakpoint and a gain of interactions between the regions flanking the deletion breakpoint (arrowheads in Figure 14b). In the subtraction map this gain appeared as a dense red region (arrowheads in Figure 14c), indicating increased ectopic interactions of sequences surrounding the deletion. This gain of interactions was accompanied by a loss of contacts compared to wild-type chromatin configuration (intense blue areas depicted by asterisks in Figure 14c). For instance, the *Fgf8* gene showed a strong loss with sequences in the *Fgf8* TAD (blue area indicated by one asterisk in Figure 14c).

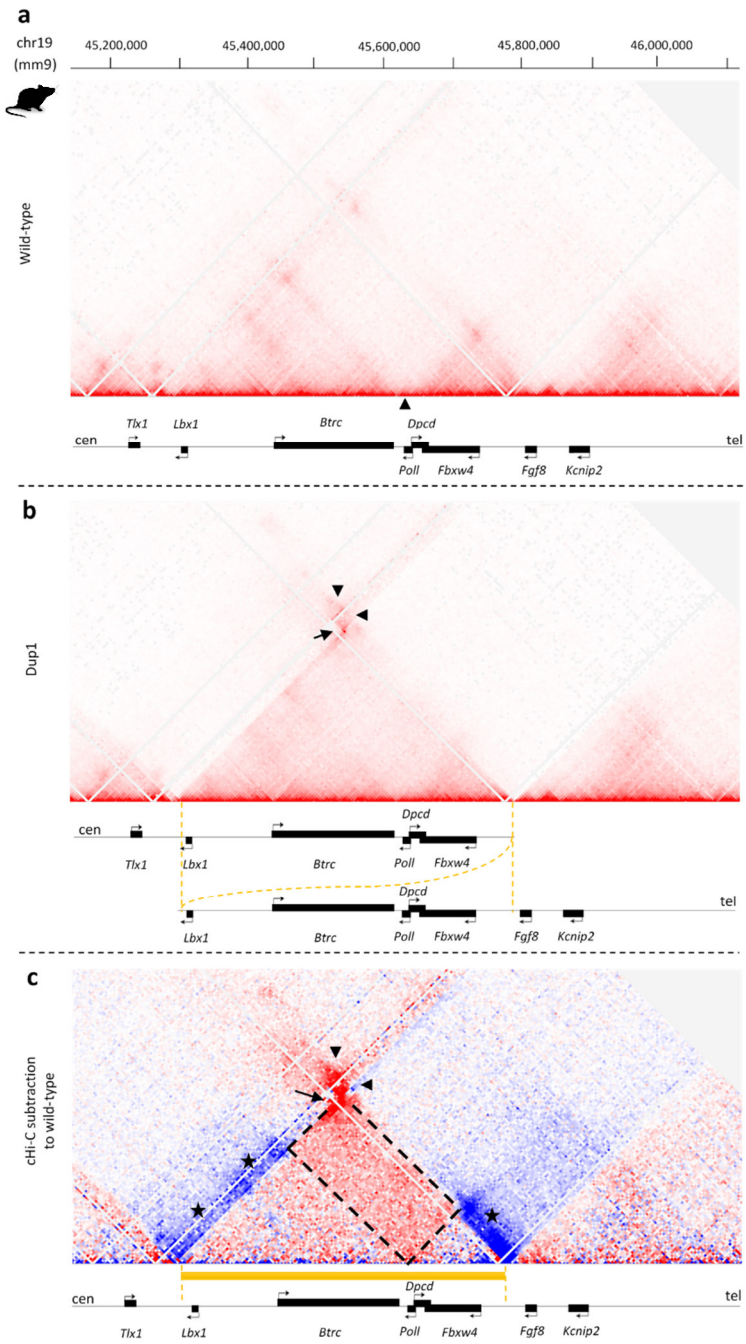

a
chr19
(mm9)
45,200,000
45,400,000
45,600,000
45,800,000
46,000,000
Wild-type
cen
Tlx1
Lbx1
Btrc
Dpcd
Poll
Fbxw4
Fgf8
Kcnip2
tel
b
Dup1
cen
Tlx1
Lbx1
Btrc
Dpcd
Poll
Fbxw4
Lbx1
Btrc
Dpcd
Poll
Fbxw4
Fgf8
Kcnip2
tel
c
cHi-C subtraction
to wild-type
cen
Tlx1
Lbx1
Btrc
Dpcd
Poll
Fbxw4
Fgf8
Kcnip2
tel

Figure 14: cHi-C from Dup1 mice and comparison to wild-type. (**a**) cHi-C from E11.5 limb buds of wild-type embryos. Black arrowhead indicates the TAD boundary between *Lbx1* and *Fgf8* TADs. (**b**) cHi-C from E11.5 limb buds of Dup1 mutants. Schematic of the allele is shown below the map and the duplication is indicated by an overlap. The dashed vertical yellow lines mark the duplication breakpoints in the linear genome. The black arrow highlights the ectopic interactions induced by the duplication breakpoint engineered by CRISPR/Cas9 genome editing. cHi-C also revealed the breakpoint and ectopic contacts of a corresponding deletion on the second allele (arrowheads). (**c**) Subtraction of wild-type cHi-C interactions from Dup1 mutants revealed ectopic interactions of regions of the *Lbx1* TAD (including *Lbx1*) and sequences of the Fgf8 TAD that are involved in the duplication. Ectopic interactions induced by the duplication are depicted by the red rectangular area (indicated by dashed black lines) below the duplication breakpoint (black arrow). The deletion on the second allele is characterized instead by ectopic contacts above the deletion breakpoint, i.e. ectopic contacts (arrowheads) between the regions flanking the deleted sequence. In addition, the deletion is accompanied by a loss of interactions (blue areas indicated by asterisks), e.g. loss of interactions between the *Fgf8* gene and regions in the *Fgf8* TAD. The yellow bar indicates the genomic region involved in the duplication and deletion.

For the Dup2 mutant mice, the cHi-C map clearly indicated a gain of interaction (arrow in Figure 15b), which arose from the duplication and showing, in the subtraction map, a pattern similar to the one observed for Dup1 (red rectangular zone dense of interactions under the black arrow in Figure 15c and highlighted by dashed black lines). However, the duplication allele in Dup2 mutant mice excludes *Lbx1* from the tandem duplication and therefore from a neo-TAD around the breakpoint. Unexpectedly and in stark contrast with the genotyping results, cHi-C analysis revealed an additional strong gain of interactions and breakpoint in Dup2 mice (arrowhead in Figure 15b). The subtraction map showed that a deletion of part of the sequence of the *Lbx1* TAD and *Fgf8* TAD (green arrowheads indicate the centromeric and telomeric deletion breakpoints in Figure 15c and 15b) is responsible for this gain in interactions as it brings in close proximity the flanking regions that are normally separated by a TAD boundary. Indeed, for instance, *Fgf8* showed a loss of interactions with the deleted region in the *Fgf8* TAD (asterisks in Figure 15c) and a gain of interactions with sequences of the remaining *Lbx1* TAD (red dense area indicated by an arrowhead in Figure 15c). This result led to the conclusion that the selected Dup2 ESC clone carried a deletion allele in addition to the duplication allele. Surprisingly, the deletion and the breakpoint did not correspond to the size and position of the duplication. Whereas the telomeric breakpoint was the same as in the duplication, the centromeric breakpoint was shifted towards the *Btrc* gene. This unexpected configuration of the deletion allele could explain why the deletion was not detected in the PCR screening and by the copy number analysis.

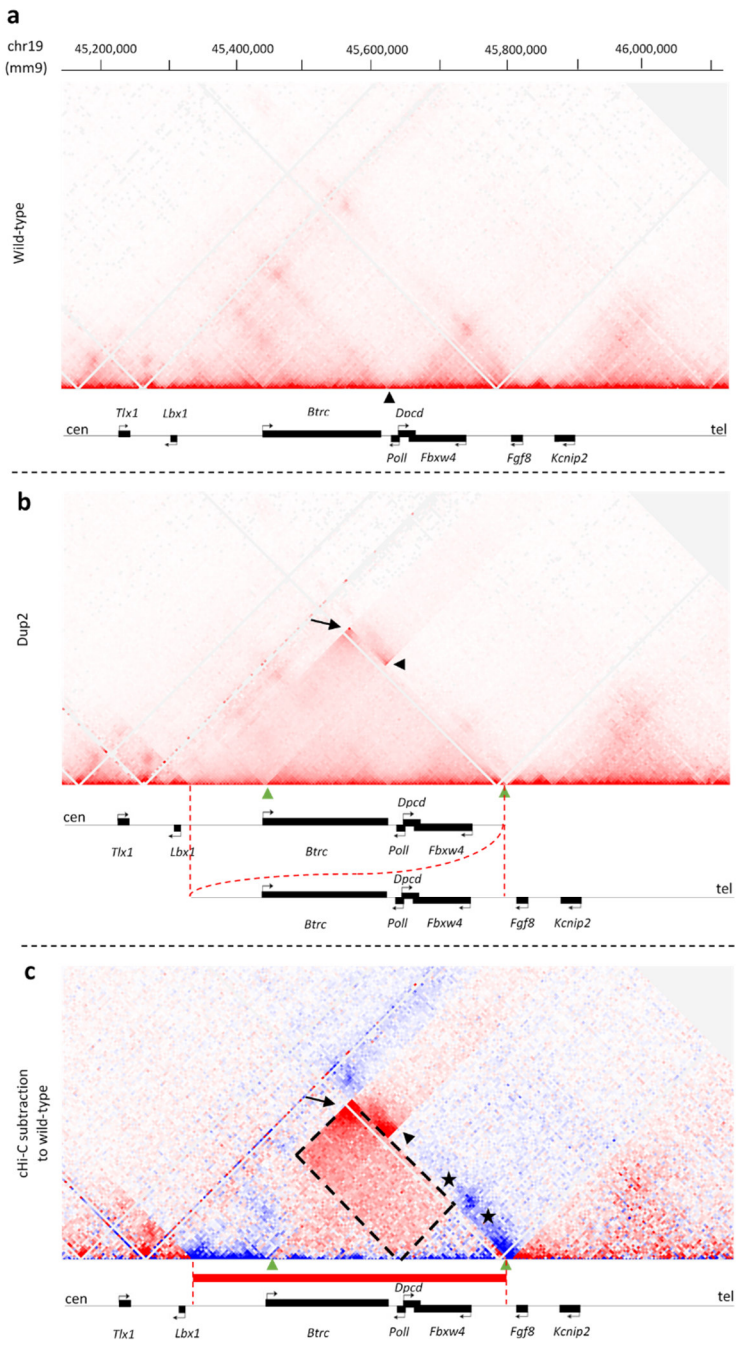

a
chr19
(mm9)
45,200,000
45,400,000
45,600,000
45,800,000
46,000,000
Wild-type
cen
Tlx1
Lbx1
Btrc
Dpcd
Poll
Fbxw4
Fgf8
Kcnip2
tel
b
Dup2
c
cHi-C subtraction
to wild-type

Figure 15: cHi-C from Dup2 mice and comparison to wt map. (**a**) cHi-C from E11.5 limb buds of wild-type embryos. Black arrowhead indicates the TAD boundary between *Lbx1* and *Fgf8* TADs. (**b**) cHi-C from E11.5 limb buds of Dup2. Schematic of the allele is shown below the map and the duplication is indicated by an overlap. The dashed vertical red lines mark the duplication breakpoints in the linear genome and is evident that *Lbx1* is excluded from the duplication. The black arrow highlights the ectopic interactions induced by the duplication breakpoint engineered by CRISPR/Cas9 genome editing. cHi-C also revealed an additional breakpoint and ectopic contacts resulting from a deletion (black arrowheads). Green arrowheads indicate the centromeric and telomeric breakpoints of the deletion. The telomeric breakpoint is the same as in the duplication, while the centromeric breakpoint is shifted towards *Btrc* gene. (**c**) Subtraction of wild-type cHi-C interactions from Dup2 mutants revealed ectopic interactions of regions of the *Lbx1* TAD (excluding *Lbx1*) and sequences of the Fgf8 TAD that are involved in the duplication. Ectopic interactions induced by the duplication are depicted by the red rectangular area (indicated by dashed black lines) below the duplication breakpoint (black arrow). The deletion on the second allele is characterized instead by ectopic contacts above the deletion breakpoint, i.e. ectopic interactions (black arrowhead) between the regions flanking the deleted sequence. In addition, the deletion is accompanied by a loss of interactions (blue areas indicated by asterisks), e.g. loss of interactions between the *Fgf8* gene and regions in the *Fgf8* TAD. Green arrowheads indicate the centromeric and telomeric breakpoints of the deletion. The red bar indicates the genomic region involved in the duplication.

5.4 Phenotypic analysis of structural variations at the *Fgf8* locus

5.4.1 Correlation of gene expression patterns and cHi-C identified genotypes

Although Dup1 and Dup2 mutant mice showed different structural rearrangements on both alleles, they were subjected to gene expression analysis. Murine embryos were collected at developmental stage E11.5 and the whole mount *in situ* hybridization was performed, to visualize the expression patterns of *Fgf8* and *Lbx1* in the whole embryo and to examine differential expression. In E11.5 wild-type embryos, *Fgf8* was primarily expressed in the AER, in the mandibular/maxillary region, in the midbrain-hindbrain boundary and in the nasal pit epithelium (arrowheads on *Fgf8* wild-type embryo in Figure 16b), while *Lbx1* resulted expressed in the early myogenic cells (arrowheads on *Lbx1* wild-type embryo in Figure 16a). Analysis of Dup1 and Dup2 embryos revealed some differences in the expression of *Fgf8* and *Lbx1* genes compared to wild-type embryos. These differences were in strong concordance with the genotype determined by the cHi-C analysis.

Dup1 mutant embryos exhibited a smaller body size and shorter forelimbs and hindlimbs at E11.5 compared to the wild-type (Dup1 E11.5 embryos in Figure 16a and b). *Lbx1* was ectopically expressed in the AER and in the mandibular/maxillary region, resembling *Fgf8* specific expression patterns (arrowheads on *Lbx1* Dup1 embryo and Dup1 limb bud in Figure 16a). This misexpression was in agreement with the observed ectopic interactions between the sequence where *Lbx1* is located and the regulatory region of the *Fgf8* TAD in the cHi-C analysis (red area dense of interactions highlighted by dashed black lines in the cHi-C subtraction map in Figure 16a). The duplication induced a new chromatin domain at the *Fgf8* locus, in which *Lbx1* adopted *Fgf8* specific *cis*-regulatory information. Consequently, Lbx1 misexpression likely arose from the duplication in Dup1 mutant mice.

Conversely, whole mount *in situ* hybridization of *Fgf8* revealed a reduced expression in the AER (arrowhead on *Fgf8* Dup1 limb bud in Figure 16b), which was most likely due to the deletion on the second allele. This was clearly visible in the cHi-C subtraction map as a loss of interactions, specifically the ones between *Fgf8* and its regulatory region in the *Fgf8* TAD (blue area highlighted by dashed black lines in the cHi-C subtraction map in Figure 16b). However, *Fgf8* expression in E11.5 Dup1 embryos was only reduced but not completely lost, as the duplication allele still retained a functional copy of the *Fgf8* TAD.

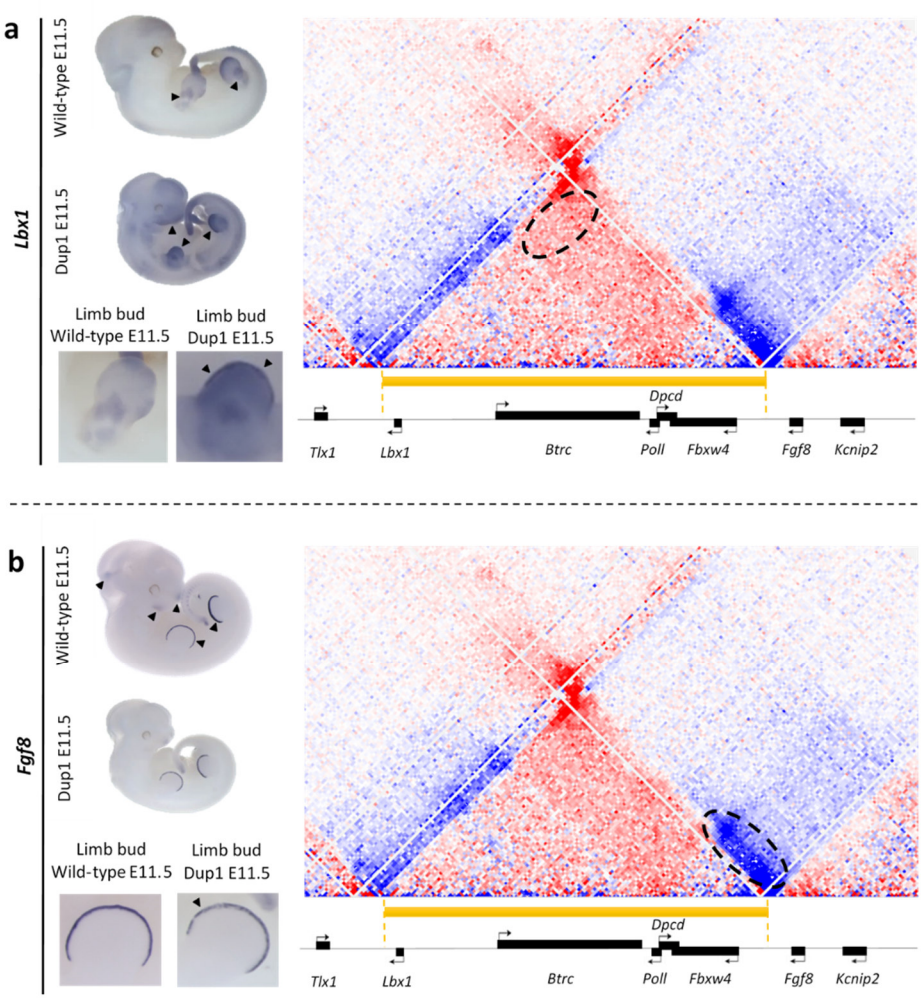


Figure 16: Whole mount *in situ* hybridization and correlation with cHi-C in Dup1. (**a**) Whole-mount *in situ* hybridization for *Lbx1* in E11.5 wild-type and Dup1 mutant embryos. The cHi-C subctraction map on the right correlates the observed ectopic contact with the observed misexpression of *Lbx1*. Mutant embryos showed a misexpression of *Lbx1* in the AER and in the mandibular/maxillary region, as a consequence of the gain of interactions with the *Fgf8* regulatory region observed in the cHi-C subtraction map (red area highlighted by dashed black lines). (**b**) Whole-mount *in situ* hybridization for *Fgf8* in E11.5 wild-type and Dup1 mutant embryos. The cHi-C subtraction map on the right correlates the observed loss of contacts with the reduced *Fgf8* expression observed in the AER. This reduction is a consequence of the deletion on the second allele, that involves the regulatory region of *Fgf8* and, therefore, leads to the loss of interactions between the regulatory region and *Fgf8* (blue area highlighted by dashed black lines).

In the case of Dup2 mutant mice, no significant changes in the expression of *Lbx1* were observed (*Lbx1* Dup2 embryo and Dup2 limb bud E11.5 in Figure 17a), which was in concordance with the absence of ectopic interaction of *Lbx1* in the cHi-C analysis (area with no interactions highlighted by dashed black lines in the cHi-C subtraction map in Figure 17a). *Lbx1*, indeed, was not involved in the duplication and, therefore, was excluded from the rearranged neo-TAD around the breakpoint (black arrow in Figure 17a).

The deletion detected in the cHi-C map of Dup2 mutant mice resulted in the misexpression of *Fgf8* in a *Lbx1* specific expression pattern (arrowheads on *Fgf8* Dup2 embryo and Dup2 limb bud E11.5 pictures in Figure 17b). The mutants clearly showed *Fgf8* expression in the myogenic cells of the limb bud, reflecting the gain of interactions between *Fgf8* and a potential *Lbx1* regulatory region of the remaining *Lbx1* TAD (red area dense of interactions highlighted by dashed black lines in the cHi-C subtraction map in Figure 17b).

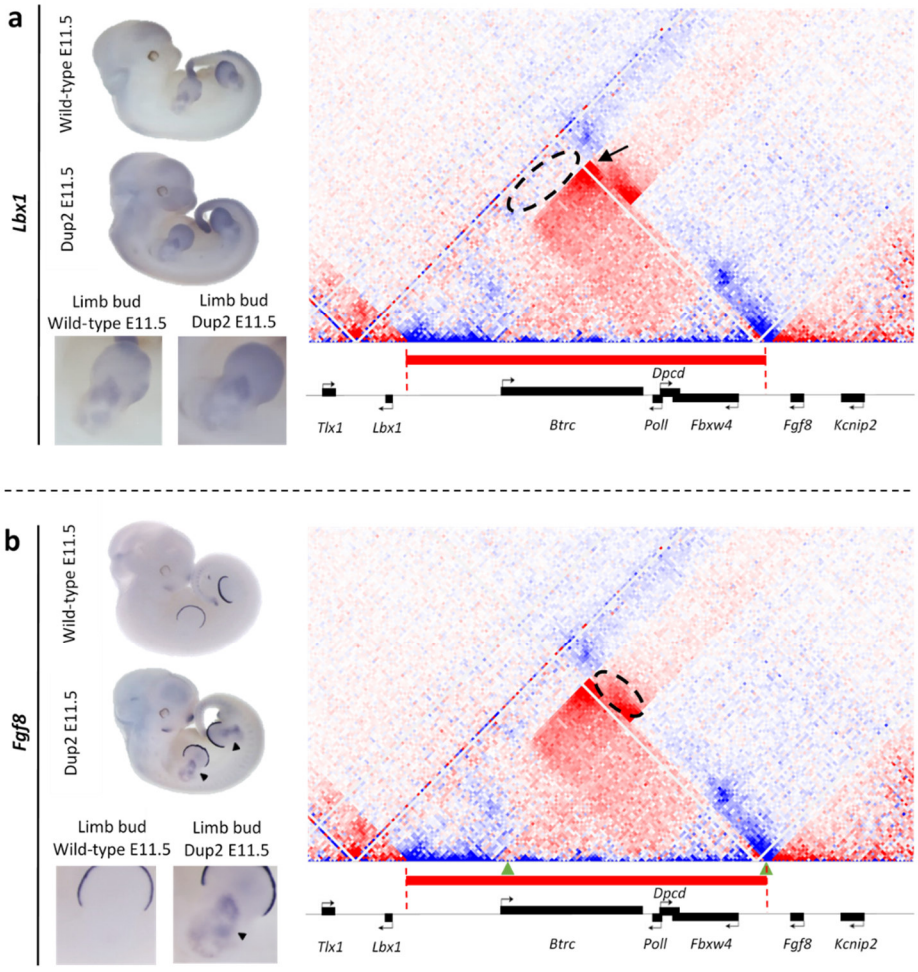

a
Lbx1
Wild-type E11.5
Dup2 E11.5
Limb bud
Wild-type E11.5
Limb bud
Dup2 E11.5
Dpcd
Tlx1
Lbx1
Btrc
Poll
Fbxw4
Fgf8
Kcnip2
b
Fgf8
Wild-type E11.5
Dup2 E11.5
Limb bud
Wild-type E11.5
Limb bud
Dup2 E11.5
Dpcd
Tlx1
Lbx1
Btrc
Poll
Fbxw4
Fgf8
Kcnip2

Figure 17: Whole mount *in situ* hybridization and correlation with cHi-C in Dup2. (**a**) Whole-mount *in situ* hybridization for *Lbx1* in E11.5 wild-type and Dup2 mutant embryos. The cHi-C subtraction map on the right correlates the absence of ectopic contacts (area with no gain or loss of interactions highlighted by dashed black lines) with no misexpression of *Lbx1*. This is due to the fact that *Lbx1* is not involved anymore in the duplication, as it was instead in Dup1. (**b**) Whole-mount *in situ* hybridization for *Fgf8* in E11.5 wild-type and Dup2 mutant embryos. The cHi-C subctraction map on the right correlates the observed ectopic contact with the observed misexpression of *Fgf8*. Mutant embryos showed a misexpression of *Fgf8* in the early myogenic cells, as a consequence of the gain of interactions with the remaining sequences of *Lbx1* TAD after the deletion, potentially containing *Lbx1* regulatory regions, observed in the cHi-C subtraction map (red area highlighted by dashed black lines). Green arrowheads indicate the centromeric and telomeric breakpoints of the deletion.

5.4.2 Phenotypic analysis of Dup1 and Dup2 mutant mice

Skeletal preparation and staining of embryos at E18.5 were performed to assess the structure of every bone for an overview of the possible phenotype related to the generated structural variations. Specifically, Dup1 embryos were characterized by microcephaly and micrognathia, shorter forelimbs due to the absence of the humerus and the radius in both right and left forelimbs, smaller hindlimbs due to the shorter tibia and fibula, and absence of the thumbs (Figure 18). Micrognathia is part of the array of SHFM3 related phenotypes. In the here generated Dup1 mutants micrognathia could be a consequence of *Lbx1* misexpression in the mandibular/maxillary area, observed in section 5.4.1, due to the inter-TAD duplication. However, despite the presence of the duplication on one allele in Dup1 mutant embryos, no split hand/foot phenotype was detected, as none of the central digits on hands and/or feet were affected. Missing thumbs, humerus and radius were instead a consequence of the deletion carried by the other allele and involving the regulatory region of *Fgf8*. Indeed, this observed phenotype resembled the *Fgf8* conditional knockout phenotype (Moon AM and Capecchi MR 2000; Lewandoski M et al. 2000). The phenotype of the Dup1 mutants, in particular the presence of the micrognathia, did not allow the generation of live animals. Dup2 embryos showed a normal skeletal phenotype compared to Dup1 embryos (Figure 18). However, an abnormal aspect was represented by a hole in the scapula, which was likely a consequence of the deletion carried by one allele. Dup2 live animals were generated, but they all died shortly after birth of unknown cause.

Therefore, the classical split-hand/foot phenotype observed in our patients could not be recapitulated in Dup1 mutants. However, the differences observed between the Dup1 and Dup2 embryos, in term both of phenotype and gene misexpression, offered new insights for the study of the pathomechanism of SHFM3 phenotype associated with duplications at the *FGF8* locus. Additionally, this brought us to a further experiment described in section 5.5.

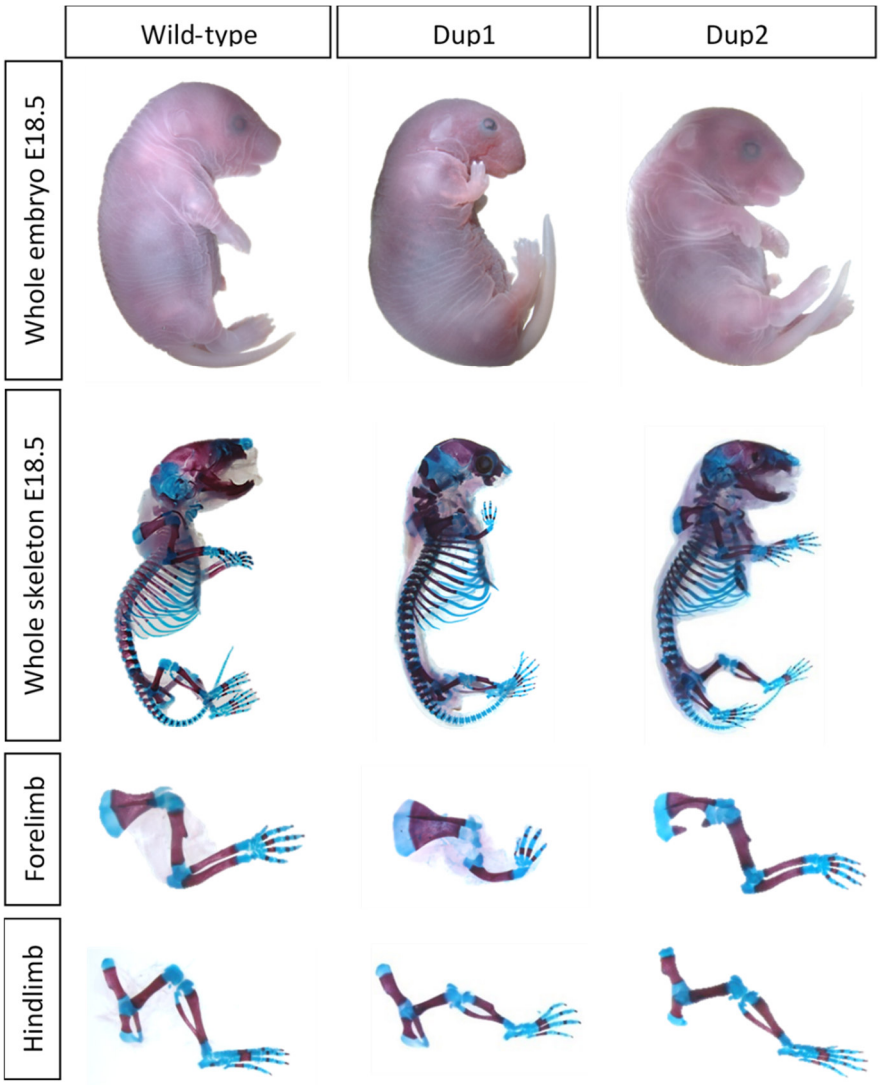

Wild-type
Dup1
Dup2
Whole embryo E18.5
Whole skeleton E18.5
Forelimb
Hindlimb

Figure 18: Skeletal phenotype of Dup1 and Dup2 mutants. Skeletal preparations at E18.5 are shown in comparison to the wild-type. Dup2 exhibit a largely normal phenotype, except for a hole in the scapula. Dup1 mutants, in contrast, showed a severe phenotype characterized by microcephaly, micrognathia, missing humerus and ulna, absence of the thumbs, and also shorter hindlimbs.

5.5 Inversion at the *Fgf8* locus to further study gene misexpression

The Dup1 tandem duplication positioned the *Lbx1* gene in close proximity to the *Fgf8* regulatory region, which was evident by the ectopic interactions between these two sequences. To further provide evidence that ectopic interaction of *Lbx1* with the *Fgf8* regulatory region is involved in *Lbx1* misexpression, we considered creating an inversion of the region between *Lbx1* and *Fgf8*. Indeed, this inversion would relocate the *Fgf8* regulatory region, normally positioned between *Btrc* and the area telomeric to *Fbxw4* (Marinić M et al. 2013), in close proximity to *Lbx1* gene and also away from *Fgf8*. Since the Dup2 CRISPR/Cas9 edited region overlaps the region between *Lbx1* and *Fgf8* (red bar in Figure 19), Dup2 ESC clones harbouring a pure inversion based on the conventional PCR screening (Table 9) were selected (Figure 19).

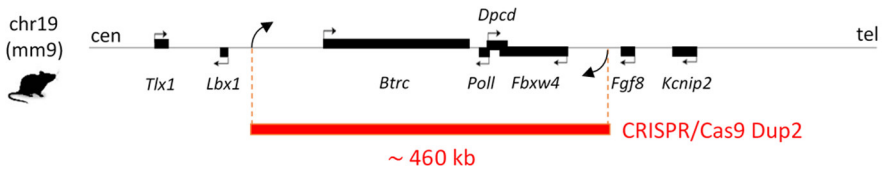
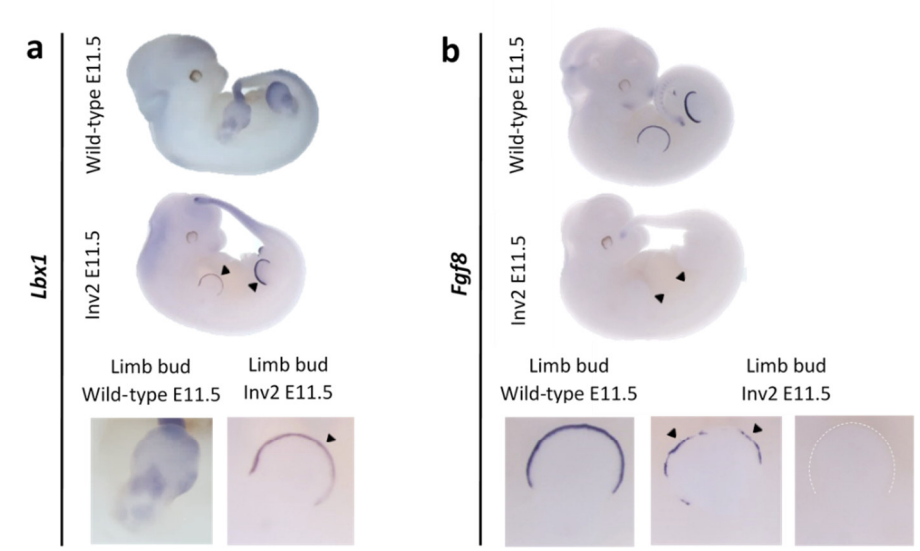


Figure 19: CRISPR/Cas9 inversion at the *Fgf8* locus. Schematic representation of the *Fgf8* locus and Dup2 inversion are shown. Dashed red lines and curve black arrows indicate the centromeric and telomeric breakpoints of the inversion.

Selected clones, hereafter referred to as Inv2 clones, were used for tetraploid complementation and embryos were collected at E11.5 to perform whole-mount *in situ* hybridization and at E18.5 for skeletal preparations. Whole-mount *in situ* hybridization analysis at E11.5 revealed a partial or complete loss of *Fgf8* expression and a strong gain of *Lbx1* expression in the AER (arrowheads on *Fgf8* Inv2 embryo and Inv2 limb buds in Figure 20b and arrowheads on *Lbx1* Inv2 embryo and Inv2 limb bud E11.5 in Figure 20a), as observed for the Dup1 mutants harbouring the duplication and the deletion allele. *Lbx1* misexpression in the AER clearly derived from the interaction of *Lbx1* with *Fgf8* regulatory elements. This confirmed the possibility for *Fgf8* regulatory elements to interact with *Lbx1* when they are in close proximity, which is exactly what occurs in Dup1 mutants as a consequence of the duplication and in Inv2 mutants when the sequence containing the regulatory elements is inverted.

Skeletal preparations of Inv2 mutants were characterized by microcephaly, absence of thumbs and ilium, partial formation of the scapula and malformation of the fibula (Figure 20c). Missing thumbs, underdeveloped scapula and leg bones were likely the consequence of *Fgf8* loss of expression. Furthermore, despite E11.5 Inv2 mutants did not show such a strong misexpression of *Lbx1* in the mandibular/maxillary region as in Dup1 embryos, E18.5 Inv2 skeletal preparations exhibited the presence of micrognathia, as in E18.5 Dup1 mutants. This further supported a possible correlation between ectopic interactions induced by the duplication and SHFM3-like phenotype.


a
Lbx1
Wild-type E11.5
Inv2 E11.5
Limb bud
Wild-type E11.5
Limb bud
Inv2 E11.5
b
Fgf8
Wild-type E11.5
Inv2 E11.5
Limb bud
Wild-type E11.5
Limb bud
Inv2 E11.5

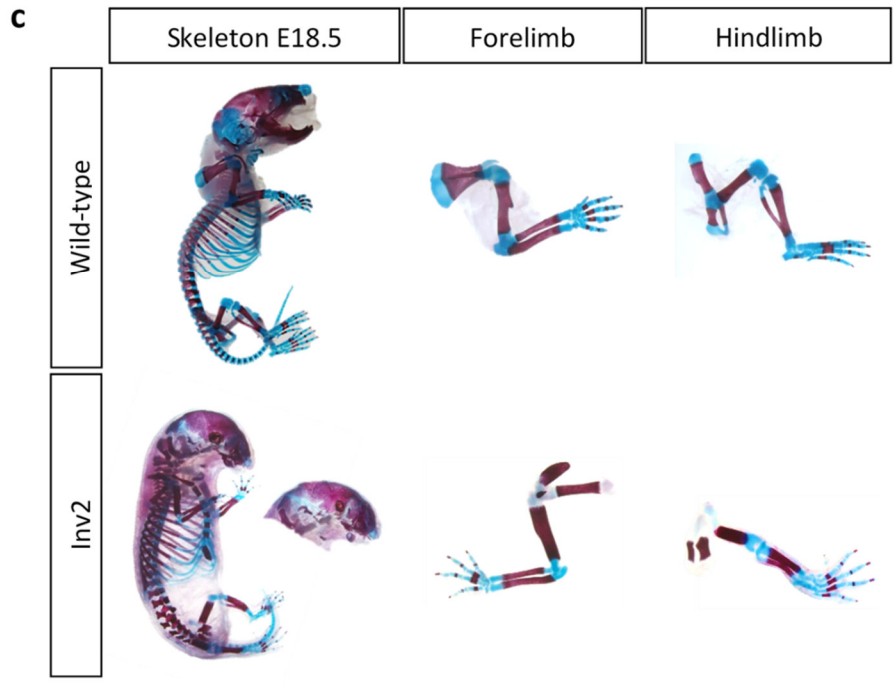

c
Skeleton E18.5
Forelimb
Hindlimb
Wild-type
Inv2

Figure 20: Whole mount *in situ* hybridization and skeletal phenotype of Inv2 mutants. (**a**) Whole-mount *in situ* hybridization for *Lbx1* in E11.5 wild-type and Inv2 mutant embryos. Mutant embryos show misexpression of *Lbx1* in the AER, due to a gain of interactions with the *Fgf8* regulatory region. (**b**) Whole-mount *in situ* hybridization for *Fgf8* in E11.5 wild-type and Inv2 mutant embryos. Mutant embryos exhibit a partial or complete loss of *Fgf8* in the AER, as the regulatory region of *Fgf8* is moved away from *Fgf8* itself with the inversion. (**c**) Skeletal preparations at E18.5 of Inv2 mutants and wild-type are shown. Inv2 embryos exhibit microcephaly, micrognathia, missing thumbs and ilium, underdeveloped scapula and deformed fibula.

6. DISCUSSION

6. Spatial and functional organization at the *FGF8* locus

Chromosome conformation capture assays provide important insights into the 3D spatial organization of mammalian genomes. In particular, Hi-C technique allows the identification of the so called topologically associated domains (TADs). TADs are specific structures that subdivide the genome into discreet genomic units, providing physical proximity between CREs and their target genes, which is required for transcription (Dixon JR et al. 2012; Deng W et al. 2012; Shen Y et al. 2012). Capture Hi-C (cHi-C) methodology detects all interactions occurring in a region of interest by combining Hi-C with the use of biotin-labelled probes complementary to a targeted genomic region (Hughes JR et al. 2014; Mifsud B et al. 2015). In this study, cHi-C was applied to investigate TAD structure at the *Fgf8* locus in limb buds of wild-type mice and mutant mice carrying structural variations generated by CRISPR/Cas9 genome editing.

6.1.1 Locus subdivision into two chromatin domains

cHi-C analysis in wild-type mice revealed that the gene dense region at the *Fgf8* locus is partitioned into two main TADs. *Lbx1* and *Btrc* genes are included in the *Lbx1* TAD located at the centromeric side of the captured region, while *Poll*, *Dpcd*, *Fbxw4* and *Fgf8* are positioned in the telomeric *Fgf8* TAD. Increased interactions were observed within each TAD, while contacts between them were low or completely absent. This goes along with previous studies demonstrating that sequences located within the same TAD contact each other with a much higher frequency than sequences of neighbouring TADs (Nora EP et al. 2012; Dixon JR et al. 2012).

TAD structure is determined largely by CTCF proteins together with cohesin (Rao SSP et al. 2014; Nora EP et al. 2017; Schwarzer W et al. 2017) and the Mediator complex, which facilitate long-range interactions between enhancers and promoters (Kagey MH et al. 2010). The high resolution of the Hi-C data showed high interaction frequencies between boundary regions of the *Lbx1* TAD and between boundary regions *Fgf8* TAD. Looping of TAD boundaries correlated with CTCF binding sites, whose sequence analysis revealed that loops are formed between CTCF sites with a convergent motif orientation. These observations are consistent with previous studies proposing that convergent motif orientation is required for loop formation (Rao SSP et al. 2014; de Wit E et al. 2015; Guo Y et al. 2015; Gómez-Marín C et al. 2015). Therefore, the strength of interactions between CTCF binding sites, associated with specific CTCF sites orientation, represent a good predictor of TAD structure. Indeed, CTCF depletion in mouse embryonic stem cells leads to disruption of the TAD structure with loss of proper insulation and increased interactions between neighbouring TADs (Nora EP et al. 2017).

Interestingly, across different mammals, CTCF binding sites and CTCF motif orientation are generally more conserved at TAD boundaries than within TADs (Vietri Rudan M et al. 2015). Additionally, between species and across cell types, TAD boundaries and TAD structures were shown to be highly conserved (Dixon JR et al. 2012; Rao SSP et al. 2014). In the present study, murine and human TAD structures were compared and, although there are differences in the genomic size, the spatial TAD organization was preserved between mouse and human. Especially, CTCF binding sites at the TAD boundaries regions demonstrated a high level of conservation. Therefore, this comparison suggests a conserved and similar gene regulation between mouse and human at the *FGF8* locus.

Furthermore, strong chromatin interactions were observed between the centromeric *Lbx1* TAD boundary and the telomeric *Fgf8* TAD boundary, forming an additional loop overlying the *Lbx1* and *Fgf8* TADs. However, this overlapping loop is not characterized by increased inter-TAD interactions between the *Lbx1* and *Fgf8* TAD. Considering the loop extrusion

model (Sanborn AL et al. 2015), overlapping loops can arise between strong CTCF binding sites, but only consecutive loop domains can occur simultaneously in the same cell. Therefore, overlapping loops represent alternative folding states within a cell population (Sanborn AL et al. 2015; Giorgetti L et al. 2014). Additionally, the here performed cHi-C experiments were carried out on the whole limb bud of embryos at E11.5. At this stage, limb bud contains multiple different cell types, i.e. cells of the AER or limb mesenchyme. The observed overlapping loop could, therefore, represent alternative folding patterns, which could be specific for a particular cell type. These aspects have to be carefully considered when interpreting cHi-C experiments that are based on measurements of chromatin interactions across a large cell population and tissues.

6.1.2 TAD structure organizes the *cis*-regulatory information required for *Fgf8* expression

Genes with complex expression patterns, such as developmental genes, are regulated by numerous enhancers (Cannavò E et al. 2016; Perry MW et al. 2010). Additionally, enhancers that regulate specific genes can also be somewhat redundant, which is thought to provide spatio-temporal precision and robustness to gene expression. This way, proper execution of regulatory processes during the course of development is ensured and, in case of mutations, this robust mechanism is supposed to buffer the mutation effect (Perry MW et al. 2011; Perry MW et al. 2012; Dunipace L et al. 2011; Frankel N et al. 2010).

A recent study identified a 220 kb region centromeric to *Fgf8* containing multiple regulatory elements necessary to drive *Fgf8* expression during embryonic development (Marinić M et al. 2013). In contrast to other developmental genes often located in gene desert regions, *Fgf8* is located in a gene dense genomic location. Its *cis*-regulatory elements are interspersed in the introns of the neighbouring genes, or in the regions between one neighbouring gene and another one, and many are located more than 90 kb away from the *Fgf8* promoter.

Using LacZ reporter gene assay the authors demonstrated that single enhancers showed specific, and in several cases overlapping, *Fgf8* expression patterns. Altogether these *cis*-regulatory elements recapitulate most of the known *Fgf8* expression domains. Interestingly, despite overlapping tissue-specific activities, they are all required to ensure precise control during development and they act as an integrated unit (Marinić M et al. 2013). These *cis*-regulatory elements are specifically located between the *Btrc* gene and the region telomeric to *Fbxw4*, which corresponds to the centromeric region of the here determined *Fgf8* TAD. Therefore, the observed Fgf8 TAD and loop ensure the proximity of the multiple *cis*-regulatory elements to the target promoter of *Fgf8* and the TAD boundaries prevent any contacts with other regulatory elements in neighbouring TADs, ensuring the correct gene regulation during development.

6.2 Duplications associated with SHFM3 at the *FGF8* locus are classified as inter-TAD duplications

Tandem duplications involving the regulatory region of the *FGF8* gene have been associated with the Split-Hand/Foot Malformation type 3 (SHFM3) (de Mollerat et al. 2003; Dimitrov BI et al. 2010). The typical feature of this limb malformation is the absence of one or more central digits on hands and feet. In some cases, this phenotype can manifest together with intellectual disability, mental retardation, or facial dysmorphisms. Interestingly, there are also some patients carrying the same duplications showing only the neuro-facial features (de Mollerat et al. 2003; Elliot AM and Evans JA 2006; Dimitrov BI et al. 2010).

Comparison of the duplications position of our patients relative to the TAD structure at the *FGF8* locus revealed that these duplications are classified as inter-TAD duplications. In particular they encompass parts of the *LBX1* TAD, all including *LBX1*, and parts of the *FGF8* TAD, all excluding *FGF8*. Recent studies reported that this type of duplication, encompassing a TAD boundary, can lead to the formation of new TADs (neo-TADs) around the breakpoint

of the duplication (Franke M et al. 2016; Weischenfeldt J et al. 2017). Within these neo-TADs new ectopic interactions between *cis*-regulatory elements and genes normally located in separate TADs can occur, resulting in aberrant gene expression and pathological phenotypes (Lupiañez DG et al. 2015; Franke M et al. 2016; Weischenfeldt J et al. 2017). Therefore, considering the configuration of the identified duplications at the *FGF8* locus, we hypothesized these inter-TAD duplications could interfere with the TAD structure at the locus, creating a new chromatin domain including genes and *cis*-regulatory elements from the *FGF8* and *LBX1* TADs.

6.2.1 Inter-TAD duplications are associated with gene misregulation

In this study CRISPR/Cas9 genome editing was used to create duplications at the *Fgf8* locus in mouse embryonic stem cells (mESCs). One aim of the thesis was to reengineer human SHFM3 inter-TAD duplications in mice (Dup1) in order to understand the molecular mechanism underlying this disease. Furthermore, an additional duplication (Dup2) was generated to exclude *Lbx1* from the rearrangement, as one of the genes potentially involved in the pathomechanism of SHFM3 at the locus.

When two double strand breaks (DSBs) simultaneously occur, each on one of the two homologous chromosomes, *trans*-allelic recombination between the two chromatids, involving the two cut sequences, can lead to the formation of a tandem duplication on one allele and a consequent deletion on the second allele (Lee HJ et al. 2012). Recent studies demonstrated that this type of rearrangements can be obtained in mice with CRISPR/Cas9 genome editing when using a pair of sgRNAs simultaneously (Li J et al. 2015; Kraft K et al. 2015). However, it has been shown that duplications can also arise without the deletion allele (Franke M et al. 2016), but the mechanism leading to pure duplications without a deletion on the second allele has not been clarified yet.

The screening of our ESC clones after CRISPR/Cas9 genome editing revealed only clones harboring the duplication together with the deletion allele. Despite the clones not carrying a pure duplication, the cHi-C and gene expression analysis revealed interesting insights into the possible effects of the rearrangement. cHi-C analysis in murine limb buds of Dup1 (duplication/deletion) showed ectopic interactions of *Lbx1* with regions of the *Fgf8* TAD. This is likely the consequence of the inter-TAD rearrangement, where the duplicated copy of *Lbx1* is relocated close to the regulatory region of *Fgf8*, with formation of a new chromatin domain that facilitates novel chromatin interactions. The observed ectopic interaction resulted in a misexpression of *Lbx1* in *Fgf8*-like expression patterns, specifically in the AER and in the mandibular/maxillary region, as observed in E11.5 Dup1 embryos. This misexpression was not observed in Dup2 mutants, which exclude *Lbx1* from the duplication. Additionally, the inter-TAD inversion involving the sequences between *Lbx1* and *Fgf8* (Inv2) also leaded to misexpression of *Lbx1* in an *Fgf8*-like expression pattern as in Dup1 mutants, in particular in the AER, as a consequence of the relocation of the regulatory region of *Fgf8* in close proximity to *Lbx1* promoter. Although the Dup1 engineered mice carried a duplication and deletion allele, observed novel interactions and expression patterns induced by the duplication could still be involved in the SHFM3 phenotype. Gain of contacts between the regulatory region of *Fgf8* and *Lbx1,* and consequently misexpression of *Lbx1,* confirmed that *Fgf8 cis*-regulatory can act on a new target gene.

Similar tandem duplications, involving the *Fgf8* regulatory region, have been shown to activate a GFP reporter in *Fgf8* expression domains, in particular in the AER (Marinić M et al. 2013). The authors could not detect the expression in the mandibular/maxillary process as they performed experiments in E10.5 mutant embryos and *Fgf8* expression in this area starts at E11.5 (Bachler M and Neubüser A 2001; Bei M and Maas R 1998).

6.2.2 Generated duplication only partially recapitulates the SHFM3 phenotype

The skeletal analysis of E18.5 murine embryos generated from tetraploid complementation of Dup1 (duplication/deletion) clone only partially recapitulated the SHFM3 phenotype, as they did not show any split hand/foot malformations, but the micrognathia. Micrognathia is part of the clinical presentation of SHFM3 and, despite the patients of this study did not show this specific phenotype, it was previously reported in the Decipher database (https://decipher.sanger.ac.uk/) in two patients with no split hand/foot phenotype, but carrying a duplication, part of which overlapping the duplicated region of our patients.

The micrognathia phenotype in Dup1 mutants could be derived from the misexpression of *Lbx1* in the mandibular/maxillary area. This phenotype was also recapitulated by E18.5 Inv2 embryos. *Lbx1* is normally expressed during neuronal and especially muscular development, where it ensures the correct migration of myogenic precursors along the rising limb (Brohmann H et al. 2000; Fernández-Jaén A et al. 2014; Jagla K wt al. 1995). In principle, *Lbx1* could be taken into consideration as potentially disease-causing gene, but possible misexpression of the other genes involved in the SHFM3 duplication has to be taken into consideration, as it could interfere as well with physiological processes during development and consequently affect morphology. For instance, *Btrc,* another gene with a role in development and involved in all the duplications, has not been tested yet for expression analysis in Dup1 and Dup2 mutants. In general, several cases of developmental genes misexpression in limb are known to affect patterning and/or morphology. For instance, when *PAX3* is misexpressed in the distal part of the autopod, patients present brachydactyly (Lupiáñez DG et al. 2015). Similarly, the misexpression of KCNJ2, a gene coding for a potassium channel (Tristani-Firouzi M et al. 2002), can also cause brachydactyly and nail aplasia (Franke M et al. 2016).

Furthermore, it is noteworthy that, in contrast to null mutations of genes or mutations involving single *cis*-regulatory elements (Dathe K et al. 2009; Lettice LA et al. 2011), the consequences of large SVs involving developmental genes and numerous regulatory elements are not so easy to interpret and to link to the phenotype observed in patients. However, three-dimensional chromatin structure helps to identify gene and regulatory regions that are potentially involved in the pathomechanisms.

Mice are commonly used as a model organism to study human disease because of the similarities in genetics, physiology and organ development. However, some inter-species differences, such as life span, morphometry and some aspects of the physiology (Demetrius L 2005), could account for genotype-to-phenotype divergences between mouse and humans. Several factors could explain why the Dup1 mutants did not display the split hand/foot phenotype. First, a potential dosage effect could be the reason why heterozygous duplications, causing the SHFM3 phenotype in humans, do not lead to the human-like phenotype in Dup1 mutants. It is possible that a higher dosage is required to obtain the SHFM3 phenotype in mice. However, we were not able to generate a homozygous duplication by breeding of the animals as the observed phenotype in Dup1 mutant embryos did not allow the generation of live animals.

Second, the presence of the deletion on the second allele could interfere with the duplication effects in mutants. Normally, in the case of a pure heterozygous SHFM3 duplication, the *Fgf8* TAD is maintained on both alleles and *Fgf8* expression is not disrupted. However, the deletion affecting one allele caused a phenotype resembling the *Fgf8* conditional knockout phenotype (Moon AM and Capecchi MR 2000; Lewandoski M et al. 2000), displaying no humerus, radius and toes, instead of a phenotype that should have been similar to that observed in heterozygous *Fgf8* loss of function (Moon AM and Capecchi MR 2000). The co-occurence of the *Lbx1* misexpression in the AER could be responsible for this aggravation of the deletion phenotype, but at the same time the deletion effects could

buffer the real phenotypic consequences of the duplication. In order to prove the real effects of the inter-TAD duplication, a pure duplication is clearly required.

6.3 Off-target effects and complex rearrangements induced by CRISPR/Cas9 genome editing

The efficiency of CRISPR/Cas9 approach is partly related to the accessibility of the Cas9 enzyme to the chromatin at the locus of interest (Kuscu C et al. 2014; Li J et al. 2015). Therefore, sgRNAs were designed to target the Cas9 enzyme to open chromatin regions that are characterized by DNase I hypersensitivity sites (Crawford GE et al. 2006). Moreover, the efficiency in generating SVs is largely determined by their size. In particular, an inverse relationship between the size and the frequency of the targeted rearrangement has been described previously (Canver MC et al. 2014). Additionally, deletions and inversions occur much more frequently than duplications (Li J et al. 2015). Deletions and inversions are immediately generated by NHEJ ligation at the breakpoints, whereas duplications require a translocation and re-ligation of sequences from different alleles (Lee HJ et al. 2012). Furthermore, NHEJ events are usually more frequent than HR, since the latter requires the presence of a repair template (Carroll D 2014; Jiang F and Doudna JA 2017). These observations correlated with the frequency of the detected rearrangements for Dup1 and Dup2 alleles. Due to large size of the targeted region, approximately 500 kb for Dup1 and 460 kb for Dup2, the total number of detected positive clones for the desired rearrangements was relatively low. Only 1% and 0.6% of the total number of ESC clones were represented by clones carrying pure duplications in Dup1 and Dup2, respectively (Table 9). Furthermore, ESC clones carrying deletions and especially inversions were more frequent than duplications (Table 9).

CRISPR/cas9 genome editing requires also precise genotyping because of the error prone DNA repair machinery, which often leads to unpredictable deletion at the targeted genomic site.

Therefore, clones were first screened with qualitative PCR that detect rearrangement-specific breakpoints. Subsequently, the breakpoints were confirmed using Sanger sequencing and quantitative RT-PCR was used to verify the genomic copy number. Quantitative analysis revealed that in several cases the CRISPR-targeted ESC clones were characterized by rearrangements on both alleles as detected by conventional PCR. All identified duplication alleles for Dup1 and Dup2 carried a deletion on the second allele. Therefore, only clones carrying more than one rearrangement were available for the generation of the mutant animals. Mouse mutants generated by tetraploid complementation using a clone characterized by different rearrangements can be breed with wild-type mice in order to promote the segregation and separation of the two mutated alleles. However, the phenotype of Dup1 (duplication/deletion) did not allow the generation of live animals, while mice generated from Dup2 died few days after birth for unknown causes.

Interestingly, the genotype of the heterozygous duplication and deletion allele in the Dup2 clone was only detected by cHi-C analysis and not with conventional methods described above. cHi-C identified all possible chromatin interactions and therefore all rearrangements at the targeted locus that occurred in the Dup2 mutants. This underlies the efficiency and specificity of the capture Hi-C approach to identify changes in chromatin. Furthermore, the deletion and duplication in Dup1 mutants had the same breakpoints, whereas the centromeric breakpoint of the deletion in Dup2 did not correspond to the breakpoint of the duplication. This observation underlines another important aspect that has to be taken into consideration when performing genome editing using CRISPR/Cas9. The difference in the breakpoint locations could potentially result from off-target effects induced by the CRISPR/Cas9 system as demonstrated in several studies. Mismatches between the guide RNA and the complementary target DNA can be tolerated depending on their position, quantity and distribution (Hsu PD et al. 2013; Cong L et al. 2013; Fu Y et al. 2013).

Several strategies have been developed to reduce the frequency of off-target effects. For instance, the CRISPR/Cas9 nickase uses a modified nuclease domain of the Cas9 enzyme to introduce a single-strand cut. Using the nickase activity, double strand breaks can be created using two Cas9 nickases, each guided by sgRNAs binding to adjacent sites on opposite DNA strands (Shen B et al. 2014; Ran FA et al. 2013; Mali P et al. 2013). Large SVs can be obtained using two pairs of Cas9 nickases, one working on the left breakpoint and the other on the right breakpoint of the desired rearrangement (Ran FA et al. 2013). As for the classical CRISPR/Cas9 approach, deletions, duplications and inversions can occur when the DSBs are repaired by NHEJ. However, the Cas9 nickase method does not prevent the co-occurrence of a deletion allele and a duplication allele.

Despite the incredible advantages of CRISPR/Cas9 approach for the generation of SVs, the combination of duplication/deletion on the two alleles in one clone and unpredictable breakpoint position represented a major challenge in this study. Alternatively, other methods could be used to generate duplications, such as the previously described TAMERE approach (Hérault Y et al. 1998) and ZNFs (Miller JC et al. 2007; Urnov FD et al. 2005). TAMERE is a precise method, but extremely time-consuming, as it requires cloning of targeting constructs and multiple breeding steps to generate duplications or deletions alleles. Therefore, CRISPR/Cas9 genome editing still represents the best choice to reengineer SVs in mice, as the time required to get mutant mice is not comparable to any other approach. Moreover, as already mentioned, in some cases CRISPR/Cas9 can directly generate pure duplications (Franke M et al. 2016). Additionally, in the absence of a lethal phenotype, mice carrying different rearrangements on the two alleles can still be breed with wild-type mice to get pure SVs. Furthermore, the screening of ESC targeted clones could be improved by using the powerful cHi-C approach. As shown in this study, Chi-C analysis allowed an exhaustive assess to SVs at the locus of interest detecting rearrangements not previously identified by conventional and quantitative PCR.

6.4 cHi-C and Hi-C as diagnostic tools to detect structural variations

In this study cHi-C represented a powerful tool to detect structural variations and therefore to determine the correct genotype from CRISPR/Cas9 engineered mutant mice. Structural rearrangements, such as deletions and duplications, were detected by prominent ectopic interactions at the breakpoint as well as by new contacts established between regions that are normally separated by TAD boundaries. These observations suggest the use of CHi-C and Hi-C for systematic screening of CRISPR/Cas9 targeted clones or in diagnostic settings to identify SVs. As already proposed in the previous section, cHi-C could be an excellent tool for the screening of mouse ESCs after the CRISPR/Cas9 targeting and before their use to generate mutant mice. This would allow the confirmation of the correct rearrangement of interest and the detection of additional rearrangements within the same ESC clone. The present study demonstrated the limit of qualitative and quantitative PCR screenings. Unpredictable rearrangements induced by the error prone CRISPR/Cas9 system or by the repair mechanisms of the cells cannot be detected by pre-set primer combinations. cHi-C does not require prior knowledge of the rearrangement at the locus. Structural variations, including deletions, duplications and inversions can be detected by ectopic interactions in the interaction matrix when compared to wild-type. Indeed, the here presented results for deletions and duplications demonstrated that the most prominent contacts originated from ectopic interactions at the breakpoint and from recombined regions. Therefore, a screening of potential ESC clones would not require high sequencing depth. Moreover, costs could be further reduced by the application of a multiplex approach to label individual samples prior the hybridization to the capture probes (Misfud B et al. 2015; Hughes JR et al. 2014; Gnirke A et al. 2009).

Furthermore, cHi-C and especially Hi-C could be applied in diagnostic settings. The implementation of these techniques for the screening of patient cells could help clinicians

and human geneticists to identify SVs and to determine the possible mechanisms through which specific SVs cause a certain pathological phenotype. The great advantage of cHi-C and Hi-C is the possibility to identify the breakpoints of a SV. Furthermore, each SV, especially large SVs, can be identified through a specific pattern of interactions, which is represented by strong ectopic interactions close to the breakpoint in the interaction matrix. These patterns reflect the impact of a specific rearrangement on the structure and on gene regulation (Figure 21). For instance, sequences flanking a deletion are brought in close proximity at the breakpoint, which results in ectopic interactions pointing upward relative to the breakpoint (Figure 21c). Ectopic interactions can be also gained as a consequence of a duplication. Tandem duplications can relocate duplicated genes and CREs to a new interactive context, which is represented by ectopic interactions pointing downward respect to the breakpoint (Figure 21b). Finally, inversions can also change the relative position of genes and CREs. In this case, the new interactions are detected left and right to the breakpoint (Figure 21d).

Therefore, cHi-C and Hi-C represent powerful approaches to detect SVs, which are frequently identified in human diseases (Burton JN et al. 2013; Rickman DS et al. 2012). Additionally, Hi-C is not restricted to copy-number variations such as deletions and duplications. Inversions, as wells also translocations, can be detected (Chakraborty A and Ay F 2017), which represents a great advantage over other methods, such as the array-CGH (Microarray-based Comparative Genomic Hybridization) (Theisen A 2008) that is routinely used to detect SVs. However, in contrast to array-CGH or whole genome sequencing (WGS), Hi-C approaches require patient cells and usually millions of cells for a single experiment. Nevertheless, TAD structures have been described as largely stable across cell types (Dixon JR et al. 2012; Rao SSP et al. 2014), suggesting that cHi-C and Hi-C analysis can be easily performed using, for instance, cultured fibroblasts from skin biopsies, as well as leukocytes isolated from blood samples (Franke M et al. 2016), limiting availability problems.

Additionally, improvements of the Hi-C technique leaded to a decrease in the number of cells required to generate high-resolution interaction maps (Bonev B et al. 2017).

Furthermore, Hi-C provides information not only about SVs, but also about the overall landscape of interactions occurring within each chromosome. As already mentioned, Hi-C allows the identification of both balanced and unbalanced chromosome rearrangements. Since inter-chromosomal interactions (*trans*-chromosomal contacts) are far less frequent than intra-chromosomal interactions (*cis*-chromosomal contacts), the probability to observe ectopic *trans*-chromosomal interactions is very high. Even a low sequencing coverage would allow to detect *trans*-chromosomal rearrangements, with a reduction of the sequencing costs compared to deep WGS (Harewood L et al. 2017; Belton JM et al. 2012). In summary, Hi-C provides a useful tool in diagnostic settings, that could complement existing technologies such as WGS to identify complex genomic rearrangements and to disentangle disease causing mechanisms. Duplications and deletions are generally thought to have an effect on gene dosage (Rice AM and McLysaght A 2017). However, this interpretation is not sufficient when other elements, such as the non-coding sequence of the genome, are involved and the phenotype cannot be explained as a consequence of an altered gene dosage (Kurth I et al. 2009). The discovery of the TAD structures, their role in gene regulation and their alterations due to SVs (Lupiañez DG et al. 2015; Franke M et al. 2016) offer clinicians and researchers the great opportunity to identify new disease-causing mechanisms and genes and also to uncover the regulatory elements and interactions controlling specific gene expression.

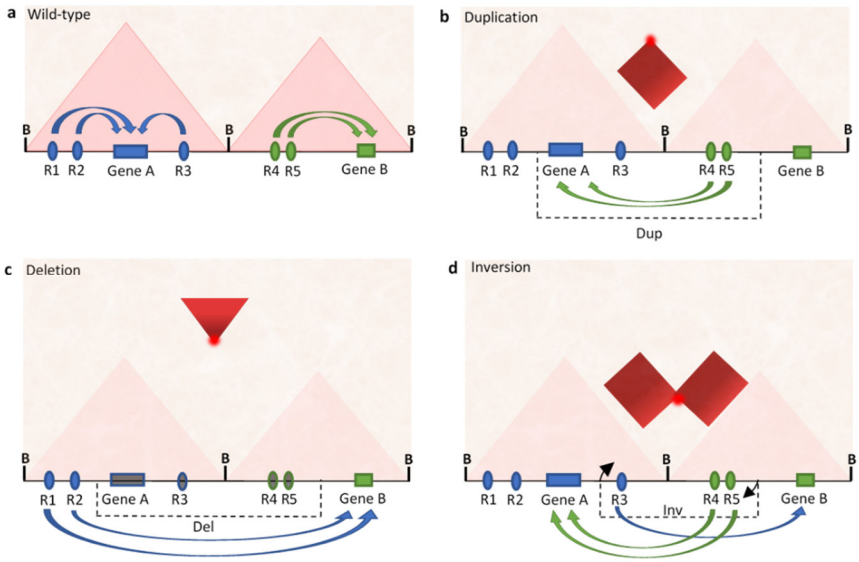


Figure 21: SVs induce characteristic ectopic interaction patterns in Hi-C data. (**a**) Wild-type configuration of a region characterized by two TADs with specific CRE-gene interactions. (**b**) Characteristic ectopic interaction patterns induced by tandem duplications spanning a TAD boundary (inter-TAD). The regulatory elements R4 and R5 are brought in close proximity to gene A, generating new interactions as illustrated by the red area and arrows. (**c**) Inter-TAD deletion of the region between the regulatory element R2 and gene B, brings these two elements, and also R1, in close proximity, defining new contacts pointing upwards from the breakpoint (red dot). (**d**) Inter-TAD inversion changes the location of regulatory elements R3, R4 and R5, which are normally positioned into two different and separate TADs. After inversion they are relocated close to and interact with different genes, as shown by ectopic interactions areas right and left of the breakpoint. (R = Regulatory element; B = boundary; Del = Deletion; Dup = Duplication; Inv = Inversion).

REFERENCES

Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF, George RA, Lewis SE, Richards S, Ashburner M, Henderson SN et al. The genome sequence of Drosophila melanogaster. *Science*. 2000; 287:2185-2195.

Ahn K, Mishina Y, Hanks MC, Behringer RR, Crenshaw EB. BMPR-IA signaling is required for the formation of the apical ectodermal ridge and dorsal-ventral patterning of the limb. *Development*. 2001; 128:4449-4461.

Anders C, Niewoehner O, Duerst A and Jinek M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature*. 2014; 513:569-573.

Artus J and Hadjantonakis AK. Transgenic Mouse Methods and Protocols. In MH Hofker and J van Deursen (Eds.) (pp. 37–56). *Humana Press*. 2011.

Bachler M and Neubüser A. Expression of members of the *Fgf* family and their receptors during midfacial development. *Mechanisms of Development*. 2001; 100:313-316.

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW and Noble WS. MEME SUITE: tool for motif discovery and searching. *Nucleic Acids Research*. 2009. 37:W202-208.

Bannister AJ and Kouzarides T. Regulation of chromatin by histone modifications. *Cell Research*. 2011. 21:381-395.

Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA and Horvath P. CRISPR provides acquired resistance against viruses in prokaryotes. *Science*. 2007; 315:1709-1712.

Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I and Zhao K. High-resolution profiling of histone methylations in the human genome. *Cell*. 2007; 129:823-837.

Behringer R, Anderson MD, Nagy KV and Nagy A. Manipulating the Mouse Embryo: A Laboratory Manual, Fourth Edition. Cold Spring Harbor Laboratory Press.

Bei M and Maas R. FGFs and BMP4 induce both *Msx1*-indipendent and *Msx1*-dependent signaling pathways in early tooth development. *Development*. 1998; 125:4325-4333.

Belton JM, McCord RP, Gibcus JH, Naumova N, Zhan Y, Dekker J. Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods*. 2012; 58:268-276.

Bénazet JD, Bischofberger M, Tiecke E, Gonçalves A, Martin JF, Zuniga A, Naef F and Zeller R. A self-regulatory system of interlinked signaling feedback loops controls mouse limb patterning. *Science*. 2009; 323:1050-1053.

Bibikova M, Golic M, Golic KG and Carroll D. Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases. *Genetics*. 2002; 161:1169-1175.

Blow MJ, McCulley DJ, Li Z, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F, Afzal V, Bristow J, Ren B, Black BL et al. ChIP-Seq identification of weakly conserved heart enhancers. *Nature Genetics*. 2010; 42:806-810.

Bolotin A, Quinquis B, Sorokin A and Ehrlich SD. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology*. 2005; 151:2551-2561.

Bonev B, Mendelson Cohen N, Szabo Q, Fritsch L, Papadopoulos GL, Lubling Y, Xu X, Lv X, Hugnot JP, Tanay A, Cavalli G. Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell*. 2017; 171:557-572.

Brohmann H, Jagla K and Birchmeier C. The role of *Lbx1* in migration of muscle precursor cells. *Development*. 2000; 127:437-445.

Bulger M and Groudine M. Functional and mechanistic diversity of distal transcription enhancers. *Cell*. 2011; 144:327-339.

Burton JN, Adey A, Patwardhan RP, Qiu R, Kitzman JO and Shendure J. Chromosome-scale scaffolding of *de novo* genome assemblies based on chromatin interactions. *Nature Biotechnology*. 2013; 31:1119-1125.

Bushdid PB, Brantley DM, Yull FE, Blaeuer GL, Hoffman LH, Niswander L and Kerr LD. Inhibition of NF-κB activity results in disruption of the apical ectodermal ridge and aberrant limb morphogenesis. *Nature*. 1998; 392:615-618.

Busser BW, Bulyk ML and Michelson AM. Toward a systems-level understanding of developmental regulatory networks. *Current Opinion in Genetics and Development*. 2008; 18:521-529.

Calzolari E, Manservigi D, Garani GP, Cocchi G, Magnani C and Milan M. Limb reduction defects in Emilia Romagna, Italy: epidemiological and genetic study in 173.109 consecutive births. *Journal of Medical Genetics*. 1990; 27:353-357.

Cannavò E, Khoueiry P, Garfield DA, Geeleher P, Zichner T, Gustafson EH, Ciglar L, Korbel JO and Furlong EM. Shadow enhancers are pervasive features of developmental regulatory networks. *Current Biology*. 2016; 26:38-51.

Canver MC, Bauer DE, Dass A, Yien YY, Chung J, Masuda T, Maeda T, Paw BH and Orkin SH. Characterization of genomic deletion efficiency mediated by clustered regularly interspaced palindromic repeats (CRISPR)/Cas9 nuclease system in mammalian cells. *The Journal of Biological Chemistry*. 2014; 289:21312-21324.

Carroll D. Genome engineering with targetable nucleases. *Annual Review of Biochemistry*. 2014; 83:409-439.

Casanova JC, Uribe V, Badia-Careaga C, Giovinazzo G, Torres M and Sanz-Ezquerro JJ. Apical ectodermal ridge morphogenesis in limb development is controlled by *Arid3b*-mediated regulation of cell movements. *Development*. 2011; 138:1195-1205.

Celli J, Duijf P, Hamel BC, Bamshad M, Kramer B, Smits AP, Newbury-Ecob R, Hennekam RC, Van Buggenhout G, van Haeringen A, Woods CG, van Essen AJ, de Waal R, Vriend G, Haber DA et al. Heterozygous germline mutations in the p53 homolog p63 are the cause of EEC syndrome. *Cell*. 1999; 99:143-153.

Chakraborty A and Ay F. Identification of copy number variations and translocations in cancer cells from Hi-C data. *Bioinformatics*. 2017; btx664.

Charpentier E and Doudna JA. Biotechnology: Rewriting a genome. *Nature*. 2013; 495:50-51.

Chiang C, Litingtung Y, Harris MP, Simandl BK, Li Y, Beachy PA and Fallon JF. Manifestation of the limb prepattern: limb development in the absence of sonic hedgehog function. *Developmental Biology*. 2001; 236:421-435.

Christian M, Cermak T, Doyle EL, Schmidt C, Zhang F, Hummel A, Bogdanove AJ, Voytas DF. Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics*. 2010; 186:757-761.

Clapier CR and Cairns BR. The biology of chromatin remodelling complexes. *Annual Review of Biochemistry*. 2009; 78:273-304.

Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA and Zhang F. Multiplex genome engineering using CRISPR/Cas systems. *Science*. 2013; 339:819-823.

Crackower MA, Scherer SW, Rommens JM, Hui CC, Poorkaj P, Soder S., Cobben JM, Hudgins L, Evans JP and Tsui LC. Characterization of the split hand/split foot malformation locus SHFM1 at 7q21.3-q22.1 and analysis of a candidate gene for its expression during limb development. *Human Molecular Genetics*. 1996; 5:571-579.

Crawford GE, Holt IE, Mullikin JC, Tai D, Blakesley R, Bouffard G, Young A, Masiello C, Green ED, Wolfsberg TG, Collins FS and National Institutes Of Health Intramural Sequencing Center. Identifying gene regulatory elements by genome-wide recovery of DNase hypersensitive sites. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*. 2003; 101:992-997.

Crawford GE, Holt IE, Whittle J, Webb BD, Tai D, Davis S, Margulies EH, Chen Y, Bernat JA, Ginsburg D, Zhou D, Luo S, Vasicek TJ, Daly MJ, Wolfsberg TG and Collins FS. Genome-wide mapping of DNase hypersensitive sites using massively parallel signature sequencing. *Genome Research*. 2006; 16:123-131.

Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, Boyer LA, Young RA, Jaenisch R. Histone H3K27ac separates active from poised enhancers and predict developmental state. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*. 2010; 107:21931-21936.

Dathe K, Kjaer KW, Brehm A, Meinecke P, Nürnberg P, Neto JC, Brunoni D, Tommerup N, Ott CE, Klopocki E, Seemann P and Mundlos S. Duplications involving a conserved regulatory element downstream of *BMP2* are associated with brachydactyly type A2. *American Journal of Human Genetics*. 2009; 84:483-92.

Davidson EH, Rast JP, Oliveri P, Ransick A, Calestani C, Yuh CH, Minokawa T, Amore G, Hinman V, Arenas-Mena C, Otim O, Brown CT, Livi CB, Lee PY, Revilla R et al. A genomic regulatory network for development. *Science*. 2002; 295: 1669-1678.

de Laat W and Dekker J. 3C-based technologies to study the shape of the genome. *Methods*. 2012; 58:189-191.

de Mollerat, Gurrieri F, Morgan CT, Sangiorgi E, Evermann DB, Gaspari P, Amiel J, Bamshad MJ, Lyle R, Blouin JL, Allanson JE, Le Marec B, Wilson M, Braverman NE, Radhakrishna U et al. A genomic rearrangement resulting in a tandem duplication is associated with split hand-split foot malformation 3 (SHFM3) at 10q24. *Human Molecular Genetics*. 2003; 12:1959-1971.

de Wit E, Vos ES, Holwerda SJ, Valdes-Quezada C, Verstegen MJ, Teunissen H, Splinter E, Wijchers PJ, Krijger PH and de Laat W. CTCF binding polarity determines chromatin looping. *Molecular Cell*. 2015; 60:676-684.

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

Dekker J. Gene regulation in the third dimension. *Science*. 2008; 319:1793-1794.

Demetrius L. Of mice and men. When it comes to studying ageing and the means to slow it down, mice are not just small humans. *EMBO Reports*. 2005; S39-44.

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

Dimitrov BI, de Ravel T, Van Driessche J, de Die-Smulders C, Toutain A, Vermeesch JR, Fryns JP, Devriendt K and Debeer P. Distal limb deficiencies, micrognathia syndrome, and syndromic forms of split hand foot malformation (SHFM) are caused by chromosome 10q genomic rearrangements. *Journal of Medical Genetics.* 2010; 47:103-111.

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

Duijf PH, van Bokhoven H and Brunner H. Pathogenesis of split-hand/split-foot malformation. *Human Molecular Genetic.s* 2003; R51-60.

Dunipace L, Ozdemir A and Stathopoulos A. Complex interactions between cis-regulatory modules in native conformation are critical for Drosophila snail expression. *Human Development.* 2011; 138:4075-4084.

Durand NC, Robinson JT, Shamin MS, Machol I, Mesirov JP, Lander E and Aiden EL. Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Systems.* 2016; 3:99-101.

Elliott AM, Reed MH, Chudley AE, Chodirker BN and Evans JA. Clinical and epidemiological findings in patients with central ray deficiency: split hand foot malformation (SHFM) in Manitoba, Canada. *American Journal of Human Genetics.* 2006; 140:1428-1439.

Elliott AM and Evans JA. Genotype-phenotype correlations in mapped split hand foot malformation (SHFM) patients. *American Journal of Human Genetics.* 2006; 140:1419-1427.

Faiyaz ul Haque M, Uhlhaas S, Knapp M, Schüler H, Friedl W, Ahmad M and Propping P. Mapping of the gene for X-chromosomal split-hand/split-foot anomaly to Xq26-q26.1. *Human Genetics.* 1993; 91:17-19.

Fernández-Jaén A, Suela J, Fernández-Mayoralas DM, Fernández-Perrone AL, Wotton KR, Dietrich S, Castellanos Mdel C, Cigudosa JC, Calleja-Pérez B and López-Martín S. Microduplication 10q24.31 in a Spanish girl with scoliosis and myopathy: the critical role of *LBX***.** *American Journal of Human Genetics.* 2006; 164A:2074-2078.

Franke M, Ibrahim DM, Andrey G, Schwarzer W, Heinrich V, Schöpflin R, Kraft K, Kempfer R, Jerković I, Chan WL, Spielmann M, Timmermann B, Wittler L, Kurth I, Cambiaso P et al. Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature.* 2016; 538:265-269.

Frankel N, Davis GK, Vargas D, Wang S, Payre F and Stern DL. Phenotypic robustness conferred by apparently redundant transcriptional enhancers. *Nature.* 2010; 466:490-493.

Friedli M, Nikolaev S, Lyle R, Arcangeli M, Duboule D, Spitz F and Antonarakis SE. Characterization of mouse Dactylaplasia mutations: a model for human ectrodactyly SHFM3**.** *Mammalian Genome.* 2008; 19:272-278.

Fromm G, Gilchrist DA and Adelman K. SnapShot: Transcription regulation: pausing. *Cell.* 2013; 153:930-930.

Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK and Sander JD. High frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature Biotechnology.* 2013; 31:822-826.

Fuda NJ, Ardehali MB and Lis JT. Defining mechanisms that regulate RNA polymerase II transcription in vivo. *Nature.* 2009; 461:186-192.

Fujita Y and Yamashita T. Spatial organization of genome architecture in neuronal development and disease. *Neurochemistry International.* 2017; (Article in press).

George SHL, Gertsenstein M, Vintrsten K, Korets-Smith E, Murphy J, Stevens ME et al. Developmental and adult phenotyping directly from mutant embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America (PNAS).* 2007; 104:4455-4460.

Giorgetti L, Galupa R, Nora EP, Piolot T, Lam F, Dekker J and Tiana G. Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. *Cell.* 2014; 157:950-963.

Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, Brockman W, Fennell T, Giannoukos G, Fisher S, Russ C, Gabriel S, Jaffe DB, Lander ES and Nusbaum C. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature Biotechnology.* 2009; 27:182-189.

Goldberg ML. Sequence analysis of *Drosophila* histone genes. Ph.D. dissertation. Stanford University, California, 1979.

Gómez-Marín C, Tena JJ, Acemel RD, López-Mayorga M, Naranjo S, de la Calle-Mustienes E, Maeso I, Beccari L, Aneas I, Vielmas E, Bovolenta P, Nobrega MA, Carvajal J and Gómez-Skarmeta JL. Evolutionary comparison reveals that diverging CTCF sites are signatures of ancestral topological associating domains borders. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*. 2015; 112:7542-7547.

Goodman FR, Majewski F, Collins AL and Scambler PJ. A 117-kb microdeletion removing *HOXD9-HOXD13* and *EVX2* causes synpolydactyly. *American Journal of Human Genetics*. 2002; 70:547-555.

Green MR and Sambrook J. Molecular Cloning - A Laboratory Manual (Fourth Edit). 2012.

Gross MK, Dottori M and Goulding M. *Lbx1* Specifies Somatosensory Association Interneurons in the Dorsal Spinal Cord. *Neuron*. 2002; 34:535-549.

Guo Y, Xu Q, Canzio D, Shou J, Li J, Gorkin DU, Jung I, Wu H, Zhai Y, Tang Y, Lu Y, Wu Y, Jia Z, Li W, Zhang MQ, Ren B, Krainer AR, Maniatis T and Wu Q. CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell.* 2015; 162:900-910.

Gurrieri F, Prinos P, Tackels D, Kilpatrick MW, Allanson J, Genuardi M, Vuckov A, Nanni L, Sangiorgi E, Garofalo G, Nunes ME, Neri G, Schwartz C and Tsipouras P. A split hand-split foot (SHFM3) gene is located at 10q24-->25. *American Journal of Human Genetics*. 1996; 62:427-436.

Hardison RC. Conserved noncoding sequences are reliable guides to regulatory elements. *Trends in Genetics*. 2000; 16:369-372.

Harewood L, Kishore K, Eldridge MD, Wingett S, Pearson D, Schoenfelder S, Collins VP and Fraser P. Hi-C as a tool for precise detection and characterisation of chromosomal rearrangements and copy number variation in human tumours. *Genome Biology*. 2017; 18:125.

Harfe BD, Scherz PJ, Nissim S, Tian H, McMahon AP and Tabin CJ. Evidence for an Expansion-Based Temporal Shh Gradient in Specifying Vertebrate Digit Identities. *Cell*. 2004; 118:517-528.

Hargreaves DC and Crabtree GR. ATP-dependent chromatin remodelling: genetics, genomics and mechanisms. *Cell Research*. 2011; 21:396-420.

Hatton BA, Knoepfler PS, Kenney AM, Rowitch DH, de Alborán IM, Olson JM and Eisenman RN. N-myc is an essential downstream effector of Shh signaling during both normal and neoplastic cerebellar growth. *Cancer Research*. 2006; 66:8655-8661.

Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu C, Ching KA, Wang W, Weng Z, Green RD, Crawford GE and Ren B. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature Genetics*. 2007; 39:311-318.

Hérault Y, Rassoulzadegan M, Cuzin F and Duboule D. Engineering chromosomes in mice through targeted meiotic recombination (TAMERE). *Nature Genetics*. 1998; 20:381-384.

Hnisz D, Weintraub AS, Day DS, Valton AL, Bak RO, Li CH, Goldmann J, Lajoie BR, Fan ZP, Sigova AA, Reddy J, Borges-Rivera D, Lee TI, Jaenisch R, Porteus MH, Dekker J and Young RA. Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science*. 2016; 351:1454-1458.

Howard ML and Davidson EH. *cis*-Regulatory control circuits in development. *Developmental Biology*. 2004; 271:109-118.

Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, Li Y, Fine EJ, Wu X, Shalem O, Cradick TJ, Marraffini LA, Bao G and Zhang F. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnology*. 2013; 31:827-832.

Hughes JR, Roberts N, McGowan S, Hay D, Giannoulatou E, Lynch M, De Gobbi M, Taylor S, Gibbons R and Higgs DR. Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nature Genetics*. 2014; 46:205-212.

International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature*. 2004; 431:931–945.

Istrail S and Davidson EH. Logic functions of the genomic cis-regulatory code. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*. 2005; 102:4954–4959.

Jagla K, Dollé P, Mattei MG, Jagla T, Schuhbaur B, Dretzen G, Bellard F and Bellard M. Mouse *Lbx1* and human *LBX1* define a novel mammalian homeobox gene family related to the Drosophila lady bird genes. *Mechanisms of Development.* 1995; 53:345–356.

Jansen R, Embden JD, Gaastra W and Schouls LM. Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular Microbiology.* 2002; 43:1565–1575.

Javahery R, Khachi A, Lo K, Zenzie-Gregory B and Smale ST. DNA sequence requirements for transcriptional initiator activity in mammalian cells. *Molecular Cell Biology* 1994; 14:117-127.

Jiang F and Doudna JA. CRISPR–Cas9 Structures and Mechanisms. *Annual Review of Biophysics*. 2017; 46:505–529.

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA and Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*. 2012; 337:816–821.

Johnson DS, Mortazavi A, Myers RM and Wold B. Genome-wide mapping of in vivo protein-DNA interactions. *Science*. 2007; 316:1497-1502.

Kagey MH, Newman JJ, Bilodeau S, Zhan Y, Orlando DA, van Berkum NL, Ebmeier CC, Goossens J, Rahl PB, Levine SS, Taatjes DJ, Dekker J and Young RA. Mediator and cohesin connect gene expression and chromatin architecture. *Nature*. 2010; 467:430-435.

Kano H, Kurosawa K, Horii E, Ikegawa S, Yoshikawa H, Kurahashi H and Toda T. Genomic rearrangement at 10q24 in non-syndromic split-hand/split-foot malformation. *Human Genetics*. 2005; 118:477-483.

Kano H, Kurahashi H and Toda T. Genetically regulated epigenetic transcriptional activation of retrotransposon insertion confers mouse dactylaplasia phenotype. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*. 2007; 104:19034-19039.

Kathiriya IS, Nora EP and Bruneau BG. Investigating the transcriptional control of cardiovascular development. *Circulation Research*. 2015; 116:700-714.

Kleinjan DA and van Heyningen V. Long-range control of gene expression: emerging mechanisms and disruption in disease. *American Journal of Human Genetics*. 2005; 76:8-32.

Klopocki E, Ott CE, Benatar N, Ullmann R, Mundlos S and Lehmann K. A microduplication of the long range SHH limb regulator (ZRS) is associated with triphalangeal thumb-polysyndactyly syndrome. *Journal of Medical Genetics*. 2008; 45:370-375.

Klopocki E and Mundlos S. Copy-number variations, noncoding sequences, and human phenotypes. *Annual Review of Genomics and Human Genetics*. 2011; 12:53-72.

Klopocki E, Lohan S, Doelken SC, Stricker S, Ockeloen CW, Soares Thiele de Aguiar R, Lezirovitz K, Mingroni Netto RC, Jamsheer A, Shah H, Kurth I, Habenicht R, Warman M, Devriendt K, Kordass U et al. Duplications of *BHLHA9* are associated with ectrodactyly and tibia hemimelia inherited in non-Mendelian fashion. *Journal of Medical Genetics*. 2012; 49:119-125.

Kraft K, Geuer S, Will AJ, Chan WL, Paliou C, Borschiwer M, Harabula I, Wittler L, Franke M, Ibrahim DM, Kragesteen BK, Spielmann M, Mundlos S, Lupiáñez DG and Andrey G. Deletions, Inversions, Duplications: Engineering of Structural Variants using CRISPR/Cas in Mice. *Cell Reports*. 2015; 10:833-839.

Kunin V, Sorek R and Hugenholtz P. Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome Biology*. 2007; 8:R61.

Kurth I, Klopocki E, Stricker S, van Oosterwijk J, Vanek S, Altmann J, Santos HG, van Harssel JJ, de Ravel T, Wilkie AO, Gal A and Mundlos S. Duplications of noncoding elements 5' of SOX9 are associated with brachydactyly-anonychia. *Nature Genetics*. 2009; 41:862-863.

Kuscu C, Arslan S, Singh R, Thorpe J and Adli M. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nature Biotechnology*. 2014; 32:677-683.

Langmead B and Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nature Methods*. 2012; 9:357-359.

Lee HJ, Kweon J, Kim E, Kim S and Kim JS. Targeted chromosomal duplications and inversions in the human genome using zinc finger nucleases. *Genome Research*. 2012; 22:539-548.

Lee TI and Young RI. Transcriptional regulation and its misregulation in disease. *Cell.* 2013; 152: 1237-1251.

Lettice LA, Heaney SJ, Purdie LA, Li L, de Beer P, Oostra BA, Goode D, Elgar G, Hill RE and de Graaff E. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Human Molecular Genetics.* 2003; 12: 1725-1735.

Lettice LA, Daniels S, Sweeney E, Venkataraman S, Devenney PS, Gautier P, Morrison H, Fantes J, Hill RE and FitzPatrick DR. Enhancer-adoption as a mechanism of human developmental disease. *Human Mutation.* 2011; 32: 1492-1499.

Lewandoski M, Sun X and Martin GR. Fgf8 signalling from the AER is essential for normal limb development. *Nature Genetics.* 2000; 26: 460-463.

Li J, Shou J, Guo Y, Tang Y, Wu Y, Jia Z, Zhai Y, Chen Z, Xu Q and Wu Q. Efficient inversions and duplications of mammalian regulatory DNA elements and gene clusters by CRISPR/Cas9. *Journal of Molecular Cell Biology.* 2015; 7: 284-298.

Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R, Bernstein B, Bender MA, Groudine M, Gnirke A et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science.* 2009; 326: 289-293.

Lyle R, Radhakrishna U, Blouin JL, Gagos S, Everman DB, Gehrig C, Delozier-Blanchet C, Solanki JV, Patel UC, Nath SK, Gurrieri F, Neri G, Schwartz CE and Antonarakis SE. Split-hand/split-foot malformation 3 (SHFM3) at 10q24, development of rapid diagnostic methods and gene expression from the region. *American Journal of Human Genetics Part A.* 2006; 140: 1384-1395.

Lupiáñez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, Horn D, Kayserili H, Opitz JM, Laxova R, Santos-Simarro F, Gilbert-Dussardier B, Wittler L, Borschiwer M, Haas SA, Osterwalder M et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell.* 2015; 161: 1012-1025.

Makarova KS, Aravind L, Wolf YI and Koonin EV. Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. *Biology Direct.* 2011; 6: 38.

Mali P, Aach J, Stranges PB, Esvelt KM, Moosburner M, Kosuri S, Yang L and Church GM. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature Biotechnology.* 2013; 31:833-838.

Mariani FV, Ahn CP and Martin GR. Genetic evidence that FGFs have an instructive role in limb proximal-distal patterning. *Nature.* 2008; 453:401-405.

Marinić M, Aktas T, Ruf S and Spitz F. An integrated holo-enhancer unit defines tissue and gene specificity of the Fgf8 regulatory landscape. *Developmental Cell.* 2013; 24:530-542.

Maston GA, Evans SK and Green MR. Transcriptional regulatory elements in the human genome. *Annual Review of Genomics and Human Genetics*. 2006; 7:29-59.

Maston GA, Landt SG, Snyder M and Green MR. Characterization of enhancer function from genome-wide analyses. *Annual Review of Genomics and Human Genetics*. 2012; 13:29-57.

Mifsud B, Tavares-Cadete F, Young AN, Sugar R, Schoenfelder S, Ferreira L, Wingett SW, Andrews S, Grey W, Ewels PA, Herman B, Happe S, Higgs A, LeProust E, Follows GA et al. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nature Genetics*. 2015; 47:598-606.

Miller JC, Holmes MC, Wang J, Guschin DY, Lee YL, Rupniewski I, Beausejour CM, Waite AJ, Wang NS, Kim KA, Gregory PD, Pabo CO and Rebar EJ. An improved zinc-finger nuclease architecture for highly specific genome editing. *Nature Biotechnology*. 2007; 25:778-785.

Miller JC, Tan S, Qiao G, Barlow KA, Wang J, Xia DF, Meng X, Paschon DE, Leung E, Hinkley SJ, Dulay GP, Hua KL, Ankoudinova I, Cost GJ, Urnov FD, Zhang HS, Holmes MC, Zhang L, Gregory PD and Rebar EJ. A TALE nuclease architecture for efficient genome editing. *Nature Biotechnology*. 2011; 29:143-148.

Mojica FJ, Díez-Villaseñor C, García-Martínez J and Almendros C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology*. 2009; 155:733-740.

Moon AM and Capecchi MR. Fgf8 is required for outgrowth and patterning of the limbs. *Nature Genetics*. 2000; 26:455-459.

Nelson CE, Hersh BM and Carroll SB. The regulatory content of intergenic DNA shapes genome architecture. *Genome Biology*. 2004; 5:R25.

Niswander L. Interplay between the molecular signals that control vertebrate limb development. *The International Journal of Developmental Biology*. 2002; 46:877-881.

Niswander L. Pattern formation: old models out on a limb. *Nature Reviews Genetics*. 2003; 4:133-143.

Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J, Sedat J, Gribnau J, Barillot E, Blüthgen N, Dekker J and Heard E. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*. 2012; 485:381-385.

Nora EP, Goloborodko A, Valton AL, Gibcus JH, Uebersohn A, Abdennur N, Dekker J, Mirny LA and Bruneau BG. Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell*. 2017; 169:930-944.

Ogbourne S and Antalis TM. Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes. *Biochemical Journal*. 1998; 331:1-14.

Ohuchi H, Nakagawa T, Yamamoto A, Araga A, Ohata T, Ishimaru Y, Yoshioka H, Kuwana T, Nohno T, Yamasaki M, Itoh N and Noji S. The mesenchymal factor, FGF10, initiates and maintains the outgrowth of the chick limb bud through interaction with FGF8, an apical ectodermal factor. *Development*. 1997; 124:2235-2244.

Pan Y and Nussinov R. The role of response elements organization in transcription factor selectivity: the IFN-β enhanceosome example. *PLoS Computational Biology* 2011; 8: e1002077.

Peng Y, Clark KJ, Campbell JM, Panetta MR, Guo Y and Ekker SC. Making designer mutants in model organisms. *Development.* 2014; 141:4042-4054.

Perry MW, Boettiger AN, Bothma JP and Levine M. Shadow enhancers foster robustness of Drosophila gastrulation. *Current Biology.* 2010; 20:1562-1567.

Perry MW, Boettiger AN and Levine M. Multiple enhancers ensure precision of gap gene-expression patterns in the Drosophila embryo. *Proceedings of the National Academy of Sciences of the United States of America (PNAS).* 2011; 108:13570-13575.

Perry MW, Bothma JP, Luu RD and Levine M. Precision of hunchback expression in the Drosophila embryo. *Current Biology.* 2012; 22:2247-2252.

Petit F, Sears KE, Ahituv N. Limb development: a paradigm of gene regulation. *Nature Reviews Genetics.* 2017; 18:245-258.

Pfeifer D, Kist R, Dewar K, Devon K, Lander ES, Birren B, Korniszewski L, Back E and Scherer G. Campomelic dysplasia translocation breakpoints are scattered over 1 Mb proximal to SOX9: evidence for an extended control region. *American Journal of Human Genetics.* 1999; 65:111-124.

Phillips T. Regulation of transcription and gene expression in eukaryotes. *Nature Education.* 2008; 1:199.

Phillips-Cremins JE, Sauria ME, Sanyal A, Gerasimova TI, Lajoie BR, Bell JS, Ong CT, Hookway TA, Guo C, Sun Y, Bland MJ, Wagstaff W, Dalton S, McDevitt TC, Sen R et al. Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell.* 2013; 153:1281-1295.

Ran FA, Hsu PD, Lin CY, Gootenberg JS, Konermann S, Trevino AE, Scott DA, Inoue A, Matoba S, Zhang Y and Zhang F. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell.* 2013; 154:1380-1389.

Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES and Aiden EL. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell.* 2014; 159:1665-1680.

Recillas-Targa F, Pikaart MJ, Burgess-Beusse B, Bells AC, Litt MD, West AG, Gaszner M and Felsenfeld G. Position-effect protection and enhancer blocking by the chicken β-globin insulator are separable activities. *Proceeding of the National Academy of Sciences of the United States of America (PNAS)*. 2002; 99:6883-6888.

Rice MC and O'Brien SJ. Genetic variance of laboratory outbred Swiss mice**.** *Nature*. 1980; 283:157-161.

Rickman DS, Soong TD, Moss B, Mosquera JM, Dlabal J, Terry S, MacDonald TY, Tripodi J, Bunting K, Najfeld V, Demichelis F, Melnick AM, Elemento O and Rubin MA. Dosage sensitivity is a major determinant of human copy number variant pathogenicity. *Proceeding of the National Academy of Sciences of the United States of America (PNAS)*. 2012; 109:9083-9088.

Riddle RD, Johnson RL, Laufer E and Tabin C. Sonic hedgehog mediates the polarizing activity of the ZPA**.** *Cell*. 1993; 75:1401-1416.

Robertson EJ. Teratocarcinomas and embryonic stem cells: A practical approach**.** *Trends in Genetics*. 1987; Vol.3. Elsevier.

Rowley MJ, Nichols MH, Lyu X, Ando-Kuri M, Rivera ISM, Hermetz K, Wang P, Ruan Y and Corces VG. Evolutionarily Conserved Principles Predict 3D Chromatin Organization**.** *Molecular Cell*. 2017; 67:837-852.

Ruf S, Symmons O, Uslu VV, Dolle D, Hot C, Ettwiller L and Spitz F. Large-scale analysis of the regulatory architecture of the mouse genome with a transposon-associated sensor**.** *Nature Genetics*. 2011; 43:379-386.

Sambrook J and Russel DW. Molecular Cloning - A Laboratory Manual (Third Edit). 2001.

Sanborn AL, Rao SS, Huang SC, Durand NC, Huntley MH, Jewett AI, Bochkov ID, Chinnappan D, Cutkosky A, Li J, Geeting KP, Gnirke A, Melnikov A, McKenna D, Stamenova EK et al. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proceeding of the National Academy of Sciences of the United States of America (PNAS)*. 2015; 112:E6456-6465.

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, Minx P, Reily AD, Courtney L, Kruchowski SS, Tomlinson C et al. The B73 maize genome: complexity, diversity, and dynamics. *Science*. 2009; 326:1112-1115.

Schwarzer W, Abdennur N, Goloborodko A, Pekowska A, Fudenberg G, Loe-Mie Y, Fonseca NA, Huber W, H Haering C, Mirny L and Spitz F. Two independent modes of chromatin organization revealed by cohesin removal**.** *Nature*. 2017; 551:51-56.

Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Månér S, Massa H, Walker M, Chi M, Navin N, Lucito R, Healy J, Hicks J, Ye K, Reiner A, Gilliam TC, Trask B, Patterson N, Zetterberg A and Wigler M. Large-scale copy number polymorphism in the human genome**.** *Science*. 2004; 305:525-528.

Sekine K, Ohuchi H, Fujiwara M, Yamasaki M, Yoshizawa T, Sato T, Yagishita N, Matsui D, Koga Y, Itoh N and Kato S. Fgf10 is essential for limb and lung formation**.** *Nature Genetics*. 1999; 21:138-141.

Shamseldin HE, Faden MA, Alashram W and Alkuraya FS. Identification of a novel DLX5 mutation in a family with autosomal recessive split hand and foot malformation**.** *Journal of Medical Genetics*. 2012; 49:16-20.

Shen Y, Yue F, McCleary DF, Ye Z, Edsall L, Kuan S, Wagner U, Dixon J, Lee L, Lobanenkov VV and Ren B. A map of the cis-regulatory sequences in the mouse genome**.** *Nature*. 2012; 488:116-120.

Sidow A, Bulotsky MS, Kerrebrock AW, Birren BW, Altshuler D, Jaenisch R, Johnson KR and Lander ES. A novel member of the F-box/WD40 gene family, encoding dactylin, is disrupted in the mouse dactylaplasia mutant. *Nature Genetics*. 1999; 23:104-107.

Storm EE and Kingsley DM. Joint patterning defects caused by single and double mutations in members of the bone morphogenetic protein (BMP) family. *Development*. 1996; 122:3969-3979.

Sun X, Mariani FV and Martin GR. Functions of FGF signalling from the apical ectodermal ridge in limb development. *Nature*. 2002; 41:501-508.

Sun M, Ma F, Zeng X, Liu Q, Zhao XL, Wu FX, Wu GP, Zhang ZF, Gu B, Zhao YF, Tian SH, Lin B, Kong XY, Zhang XL, Yang W, Lo WH and Zhang X. Triphalangeal thumb-polysyndactyly syndrome and syndactyly type IV are caused by genomic duplications involving the long range, limb-specific SHH enhancer. *Journal of Medical Genetics*. 2008; 45:589-595.

Tanaka M. Fins into limbs: Autopod acquisition and anterior elements reduction by modifying gene networks involving 5'Hox, Gli3, and Shh. *Developmental Biology*. 216; 413:1-7.

The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012; 489:57-74.

Theisen A. Microarray-based Comparative Genomic Hybridization (aCGH). *Nature Education*. 2008; 1:45.

Tolhuis B, Palstra RJ, Splinter E, Grosveld F and de Laat W. Looping and interaction between hypersensitive sites in the active beta-globin locus. *Molecular Cell*. 2002; 10:1453-1465.

Tristani-Firouzi M, Jensen JL, Donaldson MR, Sansone V, Meola G, Hahn A, Bendahhou S, Kwiecinski H, Fidzianska A, Plaster N, Fu YH, Ptacek LJ and Tawil R. Functional and clinical characterization of KCNJ2 mutations associated with LQT7 (Andersen syndrome). *The Journal of Clinical Investigation*. 2002; 110:381-388.

Tucker KL, Wang Y, Dausman and Jaenisch R. A transgenic mouse strain expressing four drug-selectable marker genes. *Nucleic Acids Research*. 1997; 25:3745-3746.

Ugur SA and Tolun A. Homozygous WNT10b mutation and complex inheritance in Split-Hand/Foot Malformation. *Human Molecular Genetics*. 2008; 17:2644-2653.

Urnov FD, Miller JC, Lee YL, Beausejour CM, Rock JM, Augustus S, Jamieson AC, Porteus MH, Gregory PD and Holmes MC. Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature*. 2005; 435:646-651.

Uslu VV, Petretich M, Ruf S, Langenfeld K, Fonseca NA, Marioni JC and Spitz F. Long-range enhancers regulating Myc expression are required for normal facial morphogenesis. *Nature Genetics*. 2014; 46:753-758.

van Bokhoven H, Hamel BC, Bamshad M, Sangiorgi E, Gurrieri F, Duijf PH, Vanmolkot KR, van Beusekom E, van Beersum SE, Celli J, Merkx GF, Tenconi R, Fryns JP, Verloes A, Newbury-Ecob RA et al. p63 Gene mutations in eec syndrome, limb-mammary syndrome, and isolated split hand-split foot malformation suggest a genotype-phenotype correlation. *The America Journal of Human Genetics*. 2001; 69:481-492.

Vietri Rudan M, Barrington C, Henderson S, Ernst C, Odom DT, Tanay A and Hadjur S. p63 Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell Reports*. 2015; 10:1297-1309.

Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F, Afzal V, Ren B, Rubin EM and Pennacchio LA. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature*. 2009; 457:854-858.

Wang H, Yang H, Shivalila CS, Dawlaty MM, Cheng AW, Zhang F and Jaenisch R. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell*. 2013; 153:910-918.

Wang J, Li J, Zhao H, Sheng G, Wang M, Yin M and Wang Y. Structural and Mechanistic Basis of PAM-Dependent Spacer Acquisition in CRISPR-Cas Systems. *Cell*. 2015 163:840-853.

Wang Q, Carroll JS and Brown M. Spatial and temporal recruitment of androgen receptor and its coactivators involves chromosomal looping and polymerase tracking. *Molecular Cell*. 2005; 19:631-642.

Wassarman PM and Soriano PM. Guide to techniques in mouse development. Preface.. *Methods in Enzymology*. 2010; 476, xix.

Weischenfeldt J, Dubash T, Drainas AP, Mardin BR, Chen Y, Stütz AM, Waszak SM, Bosco G, Halvorsen AR, Raeder B, Efthymiopoulos T, Erkek S, Siegl C, Brenner H, Brustugun OT et al. Pan-cancer analysis of somatic copy-number alterations implicates *IRS4* and *IGF2* in enhancer hijacking. *Nature Genetics*. 2017; 49:65-74.

Wijchers PJ and de Laat W. Genome organization influences partner selection for chromosomal rearrangements. *Trends in Genetics*. 2011; 27:63-71.

Williamson I, Lettice LA, Hill RE and Bickmore WA. Shh and ZRS enhancer colocalisation is specific to the zone of polarising activity. *Development*. 2016; 143:2994-3001.

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

Woolfe A, Goodson M, Goode DK, Snell P, McEwen GK, Vavouri T, Smith SF, North P, Callaway H, Kelly K, Walter K, Abnizova I, Gilks W, Edwards YJ, Cooke JE and Elgar G. Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Computational Biology*. 2005; 3:e7.

Wu CH, Yamaguchi Y, Benjamin LR, Horvat-Gordon M, Washinsky J, Enerly E, Larsson J, Lambertsson A, Handa H and Gilmour D. NELF and DSIF cause promoter proximal pausing on the hsp70 promoter in Drosophila. *Genes & Development*. 2003; 17:1402-1414.

Xu J, Pope SD, Jazirehi AR, Attema JL, Papathanasiou P, Watts JA, Zaret KS, Weissman IL and Smale ST. Pioneer factor interactions and unmethylated CpG dinucleotides mark silent tissue-specific enhancers in embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*. 2007; 104:12377-12382.

Zabidi MA, Arnold CD, Schernhuber K, Pagani M, Rath M, Frank O and Stark A. Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. *Nature*. 2015; 518:556-559.

Zeller R, López-Ríos J and Zuniga A. Vertebrate limb bud development: moving towards integrative analysis of organogenesis. *Nature Reviews Genetics*. 2009; 10:845-858.

Zhao Z, Tavoosidana G, Sjölinder M, Göndör A, Mariano P, Wang S, Kanduri C, Lezcano M, Sandhu KS, Singh U, Pant V, Tiwari V, Kurukuti S and Ohlsson R. Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nature Genetics*. 2006; 38:1341-1347.

ABBREVIATIONS

°C	degree Celsius
µ	micro (prefix)
3C	chromosome conformation capture
4C	circular chromosome conformation capture
AER	apical ectodermal ridge
ATP	adenosine triphosphate
bp	base pairs
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
DTT	dithiothreitol
dup	duplication
E	embryonic stage
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
ENCODE	Encyclopedia of DNA Elements
ES	embryonic stem
ESC	embryonic stem cell
FCS	fetal calf serum
FL	forelimb
g	gram
gDNA	genomic DNA
h	hour

het	heterozygous
hg	human genome
hom	homozygous
hybe	hybridization
kb	kilo base
l	liter
LIF	Leukemia Inhibitory Factor
m	milli (prefix)
M	molar
MABT	maleic acid buffer (containing) Tween 20
Mb	mega base
min	minute(s)
mm	*Mus musculus*
mol	moles
mRNA	messenger RNA
n	nano (prefix)
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PFA	paraformaldehyde
Pfu	Pyrococcus furiosus
qPCR	quantitative PCR
RNA	ribonucleic acid
rpm	rounds per minute
SSC	saline sodium citrate buffer
TAD	topologically associating domain
TAMERE	trans-allelic targeted meiotic recombination
Taq	Thermus aquaticus
tel	telomeric
TF	transcription factor
TSS	transcription start site
U	units
UCSC	University of California, Santa Cruz
Vol	volume
v/v	volume per volume
wt	wild type

PUBLICATIONS

LETTERS

nature genetics

Composition and dosage of a multipartite enhancer cluster control developmental expression of *Ihh* (Indian hedgehog)

Anja J Will[1,2], Giulia Cova[1,2], Marco Osterwalder[3], Wing-Lee Chan[1,2,4], Lars Wittler[5], Norbert Brieske[1], Verena Heinrich[6], Jean-Pierre de Villartay[7], Martin Vingron[6], Eva Klopocki[8], Axel Visel[3,9,10], Darío G Lupiáñez[1,2,4,11] & Stefan Mundlos[1,2,4,11]

Copy number variations (CNVs) often include noncoding sequences and putative enhancers, but how these rearrangements induce disease is poorly understood. Here we investigate CNVs involving the regulatory landscape of *IHH* (encoding Indian hedgehog), which cause multiple, highly localized phenotypes including craniosynostosis and synpolydactyly[1,2]. We show through transgenic reporter and genome-editing studies in mice that *Ihh* is regulated by a constellation of at least nine enhancers with individual tissue specificities in the digit anlagen, growth plates, skull sutures and fingertips. Consecutive deletions, resulting in growth defects of the skull and long bones, showed that these enhancers function in an additive manner. Duplications, in contrast, caused not only dose-dependent upregulation but also misexpression of *Ihh*, leading to abnormal phalanges, fusion of sutures and syndactyly. Thus, precise spatiotemporal control of developmental gene expression is achieved by complex multipartite enhancer ensembles. Alterations in the composition of such clusters can result in gene misexpression and disease.

Work by the Encyclopedia of DNA Elements (ENCODE) Consortium and others has helped to characterize a wide catalog of regulatory elements, also referred to as enhancers, that control developmental gene expression in many species[3–5]. One of the most intriguing characteristics of these elements is their tendency to arrange in clusters, displaying redundancy in reporter assays and similarities in transcription factor occupancy[6,7]. Previous studies in *Drosophila melanogaster* showed that the observed redundancy may provide the system with robustness and spatiotemporal precision[8–10]. However, how the complex patterns of gene expression during development are achieved and why this involves elements with apparently redundant or overlapping functions remain elusive. CNVs generally include noncoding regions of the genome and can thus interfere with the composition and dosage of regulatory elements, but the effects of such alterations are poorly understood.

We investigated the effects of deletions and duplications upstream of *IHH*, a master gene of skeletal development involved in chondrocyte differentiation, joint formation and osteoblast differentiation. Accordingly, *Ihh* inactivation in mice results in extreme shortening of bones, joint fusions and almost absent ossification, ultimately causing early lethality[11]. Interestingly, patients carrying duplications at this locus display completely different phenotypes, including craniosynostosis, syndactyly and polydactyly[1,2], indicating alternative pathomechanisms. To define the regulatory landscape of *Ihh*, we performed circular chromosome conformation capture and sequencing (4C–seq) in embryonic day (E) 14.5 developing limbs and compared our findings to published data sets[12]. Our data show that the *Ihh* promoter interacts preferentially with the third intron of the upstream neighboring gene *Nhej1* (**Fig. 1** and **Supplementary Fig. 1**), in a genomic region affected in all reported disease-associated duplications. The region contains multiple sites positive for H3K4me1 and H3K27ac (indicative of active enhancers) and binding sites for CTCF, an architectural protein involved in facilitating enhancer–promoter contact by looping. The convergent CTCF motif orientation observed across the locus might facilitate the interactions measured in the 4C–seq experiments (**Fig. 1** and **Supplementary Fig. 2**)[13–16].

In mice in which a *lacZ* reporter cassette (*Sleeping Beauty*)[17] was inserted to capture the regulatory capacity of the region, a pattern consistent with *Ihh* expression was observed, that is, activity in condensing digits, growth plates, fingertips and skull sutures. Using a combination of H3K27ac and H3K4me1 ChIP–seq signal in E14.5 limbs[18], evolutionary conservation[19] and our 4C–seq interaction profiles, we

[1]Max Planck Institute for Molecular Genetics, RG Development and Disease, Berlin, Germany. [2]Institute for Medical and Human Genetics, Charité–Universitätsmedizin Berlin, Berlin, Germany. [3]MS 84-171, Lawrence Berkeley National Laboratory, Berkeley, California, USA. [4]Berlin-Brandenburg Center for Regenerative Therapies (BCRT), Charité–Universitätsmedizin Berlin, Berlin, Germany. [5]Department of Developmental Genetics, Max Planck Institute for Molecular Genetics, Berlin, Germany. [6]Department of Computational Molecular Biology, Max Planck Institute for Molecular Genetics, Berlin, Germany. [7]Genome Dynamics in the Immune System Laboratory, INSERM, UMR 1163, Institut Imagine, Université Paris Descartes, Sorbonne Paris Cité, Paris, France. [8]Institute of Human Genetics, Biocentre, Julius Maximilians University Würzburg, Würzburg, Germany. [9]US Department of Energy Joint Genome Institute, Walnut Creek, California, USA. [10]School of Natural Sciences, University of California, Merced, California, USA. [11]These authors jointly directed this work. Correspondence should be addressed to S.M. (mundlos@molgen.mpg.de) or D.G.L. (lupianez@molgen.mpg.de).

Received 21 February; accepted 28 July; published online 28 August 2017; doi:10.1038/ng.3939

defined nine regions with enhancer potential and validated them in mouse transgenic enhancer activity assays[20] (**Fig. 1**). Embryos were analyzed at two time points, E14.5 and E17.5, to capture *Ihh* expression domains during digit development (fingertips and cartilage anlagen) and bone growth (skull sutures and growth plates), respectively. Five of the tested elements showed activity at both stages (**Fig. 1**), whereas additional elements were active only at E17.5 (**Supplementary Fig. 3**). We scored the activity of each element in the previously identified

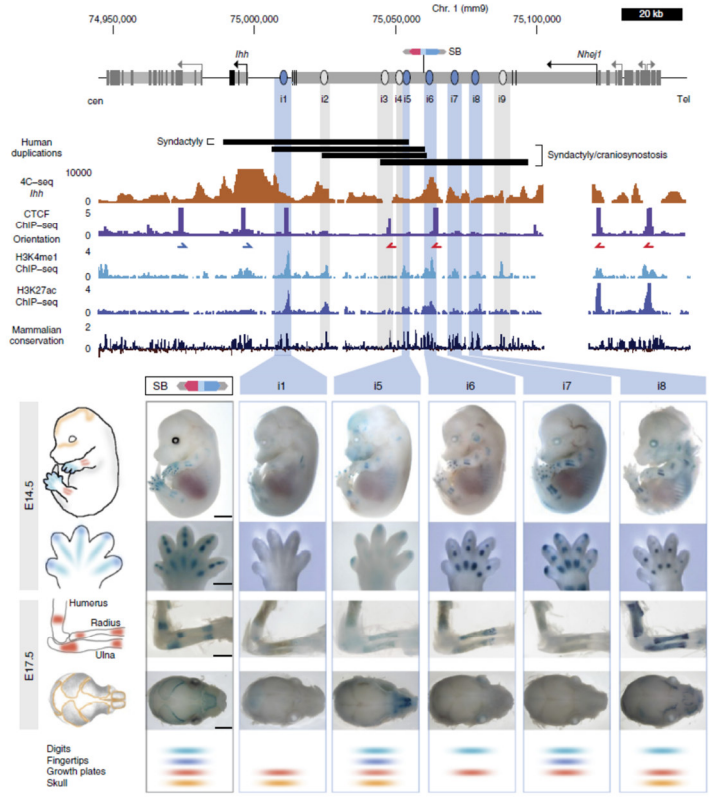


Figure 1 A cluster of enhancers interacts with the *Ihh* promoter during mouse development. Top, close-up view of the *Ihh* genomic region. Genes and their transcription start sites are indicated: black boxes, exons; gray boxes, introns. The position of the *lacZ* reporter insertion is shown (SB). Black bars indicate the size and position of previously described human duplications[1,2] converted to mouse genome coordinates. Findings from 4C–seq performed in E14.5 limbs using the *Ihh* promoter as the viewpoint are shown below. Note increased interactions with intron 3 of the adjacent *Nhej1* gene (see also **Supplementary Fig. 1**). The results of CTCF ChIP–seq performed in E14.5 limbs are shown (ENCODE)[3], where blue and red arrows indicate motif orientation. Additional tracks below show H3K4me1 and H3K27ac, as well as sequence conservation. This information was used to predict enhancers i1–i9, indicated by light blue and gray bars. Bottom, transgenic reporter assay (LacZ) of elements positive at E14.5 and E17.5 (marked in light blue; for each enhancer, an embryo and handplate at E14.5 and a dorsal view of a forelimb and a top view of the skull at E17.5 are shown). The regulatory activity of the region, as indicated by the activity of the inserted *lacZ* reporter (SB; black outline), is shown on the left. The lower panel shows scoring of each element for tissue specificity. Elements negative at E14.5 but with positive staining at E17.5 are marked in gray and shown in **Supplementary Figure 3**. Scale bars: 2,000 µm (embryos and skulls), 500 µm (handplates) and 1,000 µm (forelimbs).

regions (**Fig. 1**, **Supplementary Fig. 3** and **Supplementary Table 1**). This analysis highlighted the inherent complexity of the cluster, where almost every individual element displayed a unique pattern of activity. All elements gave a positive signal in growth plates, whereas other domains, like fingertips, were covered only by a small subset of enhancers (i5 and i7). This suggests that the enhancers in this cluster act in a modular fashion and that the degree of overlapping activity varies between tissues and developmental time points.

To evaluate the functionality of these elements, we deleted intron 3 of *Nhej1* (**Fig. 2**), which contains eight of the nine enhancers identified, using CRISVar[21]. *Nhej1* encodes a DNA repair protein essential for the non-homologous end-joining pathway, required for double-strand break repair. In humans, homozygous mutations in *NHEJ1* result in severe combined immunodeficiency (SCID) with microcephaly, growth retardation and sensitivity to ionizing radiation, reflecting a deficiency in DNA repair (MIM 611291)[22]. In contrast, *Nhej1*-knockout mice are viable and do not display any morphological phenotype[23,24]. µCT scans of *Nhej1*$^{-/-}$ skulls did not identify any abnormalities, indicating that *Nhej1* does not have a major role in skull and suture development (**Supplementary Fig. 4**). Mice homozygous for the *Nhej1* intronic deletion (Del(2–9)) displayed very short limbs, absent cortical bone and fused joints, as well as reduced skull ossification, very similar to the phenotypes observed upon *Ihh* inactivation[11]. Whereas *Nhej1* transcription levels remained basically unchanged (**Supplementary Fig. 5**), we observed a drastic reduction in *Ihh* mRNA levels in E13.5 limbs and E17.5 skulls (98% and 99% reduction, respectively), consistent with the observed phenotypes. Therefore, this genomic region contains most of the regulatory elements required for *Ihh* skeletal expression.

Next, we generated a series of specific deletions to assess the functional redundancy within this enhancer cluster (**Fig. 2**). Homozygous deletion of the enhancers located in the telomeric part of the intron (Del(4–9)) resulted in a lethal growth defect almost as severe as that observed with deletion of the entire intron, confirming that the most relevant enhancers are located in this telomeric region. Deletion of only the three central enhancers (Del(4–6)) reduced *Ihh* expression by approximately 70% in all tissues tested, whereas deletion of the three more telomeric enhancers (Del(7–9)) resulted in a 60% reduction in expression (**Fig. 2**). Both mutants were viable and phenotypically normal, but they showed a delay in skull ossification (**Fig. 2**) and a 10% reduction in bone length (**Supplementary Fig. 6**). All deletions except Del(7–9) resulted in loss of *Ihh* fingertip expression, indicating

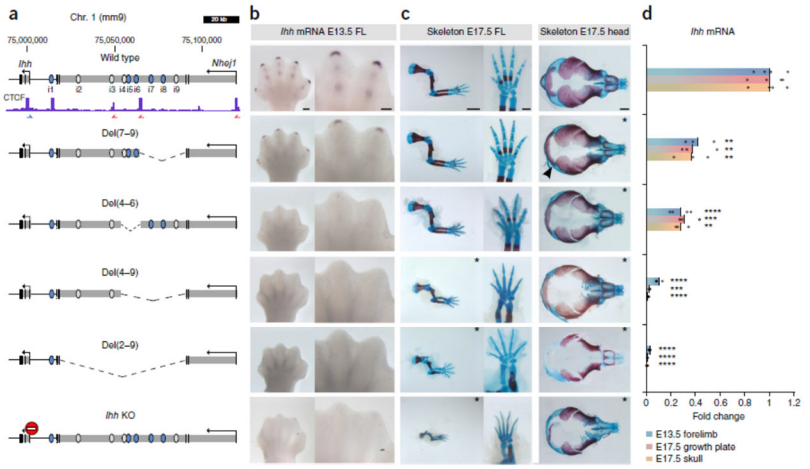


Figure 2 Deletions of regulatory elements highlight additive control of *Ihh* expression. (**a**) Deletions generated by CRISVar[21] at the *Ihh* locus. *Ihh* knockout (KO) is shown for comparison (stop sign). Findings from CTCF ChIP–seq performed in E14.5 limbs are shown (ENCODE)[3], where blue and red arrows indicate motif orientation. Each deleted chromosomal region is represented as a dashed line. Note that Del(4–9) and Del(7–9) delete only one intronic CTCF-binding site, maintaining another intact. (**b**) *In situ* hybridization shows *Ihh* expression in handplates (E13.5). Note expression in digit tips and condensing digits in wild-type embryo and loss of expression in all deletions encompassing enhancer i5. FL, forelimb. Scale bars, 200 µm (handplates). (**c**) Skeletal staining of forelimbs, autopods and skulls (E17.5). Mutants displaying abnormal phenotypes are indicated by asterisks. Both Del(2–9) and Del(4–9) result in massive reduction of limb size and reduced ossification similar to *Ihh* knockout, whereas Del(4–6) and Del(7–9) mice did not show noticeable limb abnormalities. All mutants studied displayed skull defects (delayed ossification), an effect that was less prominent in Del(7–9) mutants (arrowhead). Scale bars: 2,000 µm (forelimbs), 500 µm (autopods) and 1,000 µm (skulls). (**d**) *Ihh* qPCR analysis in E13.5 forelimb, E17.5 growth plate (elbow) and skull. Deletion of intron 3 of *Nhej1* encompassing enhancers i2–i9 results in almost complete loss of *Ihh* expression in all tissues, whereas smaller deletions partially reduce expression. Bars represent the mean of $n \geq 3$ different individuals (circles). Two-sided Student's *t* test, $^{**}P < 0.01$; $^{***}P < 0.001$; $^{****}P < 0.0001$.

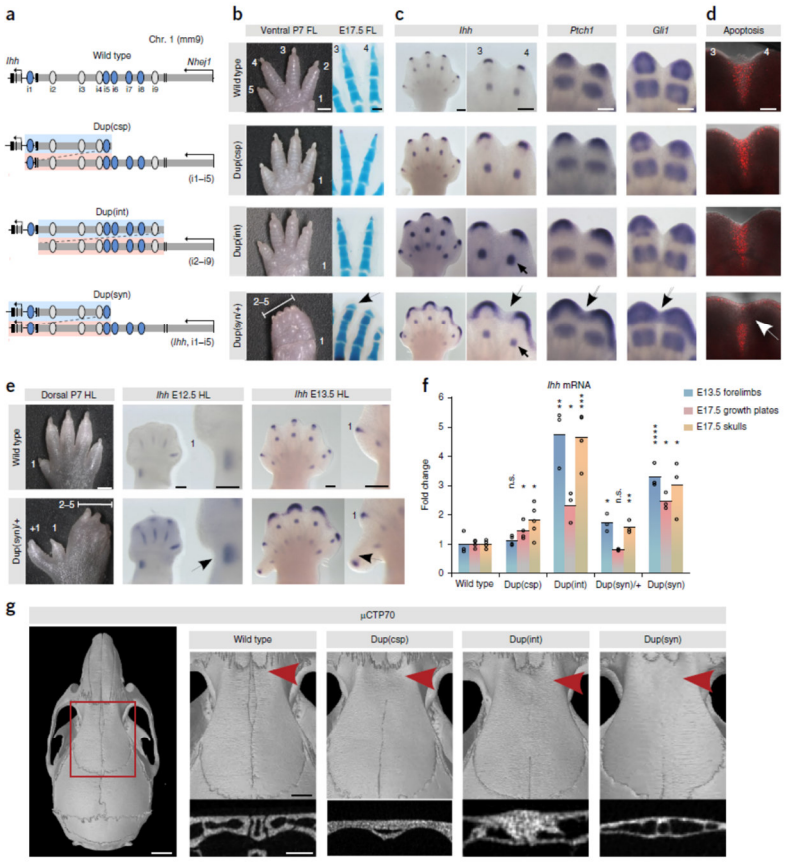


Figure 3 Duplications of enhancer elements result in *Ihh* over- and misexpression. (**a**) Duplications generated by CRISVar[21]. The duplicated fragments are shown in blue and pink. (**b**) Left, forelimb morphology (postnatal day (P) 7). Dup(int) and Dup(csp) mice (homozygous) are normal, but Dup(syn)/+ mice display 2/5 syndactyly. Right, skeletal staining at E17.5. Staining shows short and broad terminal phalanges in Dup(syn)/+ mice (arrow). Scale bars: 1,000 µm (P7) and 200 µm (E17.5 handplates). (**c**) *In situ* hybridization shows increased and broadened expression of *Ihh* and its downstream effectors *Ptch1* and *Gli1* in the mutants (Dup(csp) < Dup(int) < Dup(syn)/+). Expression domains in Dup(syn)/+ mice extend into distal interdigital space (arrows). Note the increased *Ihh* expression in digit condensations in Dup(int) as compared to Dup(syn)/+ mice (small arrow), also observed across the entire handplate. Scale bars, 200 µm. (**d**) Apoptosis in interdigital space (red signal). Note the lack of signal in the distal region in the Dup(syn)/+ mutant (arrow). Scale bar, 200 µm. (**e**) Hindlimb morphology of Dup(syn)/+ mice. Note the preaxial polydactyly and syndactyly 2/5. *In situ* hybridization shows increased *Ihh* expression (arrows) in the preaxial region (insets). HL, hindlimb. Scale bars: 1,000 µm (P7) and 200 µm (E12.5 and E13.5). (**f**) *Ihh* qPCR analysis. Duplications increase *Ihh* expression. High levels in Dup(csp) forelimbs (no phenotype) result from digit condensations, while moderate upregulation in Dup(syn)/+ forelimbs (syndactyly) derives from fingertips. Bars represent the mean of $n \geq 3$ different individuals (circles). Two-sided Student's *t* test, $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$; $^{****}P < 0.0001$; ns, not significant. (**g**) µCT skull analysis (P70). The red square indicates enlargement of the metopic suture region (right). Below, cross-sections of metopic sutures (red arrowheads). All mutants display complete suture fusion (maximum effect in Dup(int)). Scale bars: 2 mm (skull), 1 mm (enlargement), 0.5 µm (cross-section).

that element i5 acts as a major regulator for this region. These results demonstrate that *Ihh* expression is controlled by a cluster of redundant enhancers, which appear to act in an additive manner.

To understand the mechanisms underlying pathogenic duplications in the *IHH* locus, we duplicated the entire *Nhej1* intron (Dup(int)), equivalent to the sequence deleted in Del(2–9). In addition, we reengineered two of the previously described human duplications: Dup(csp), encompassing the region between enhancers i1 and i5 (reengineered human duplication causing craniosynostosis Philadelphia type[1,2]), and Dup(syn), which includes *Ihh* and the upstream region up to enhancer i5 (reengineered human duplication causing syndactyly Lueken type[2]) (**Fig. 3a**). Dup(int) and Dup(csp) mutants did not show gross morphological alterations in the heterozygous or homozygous state. In contrast, Dup(syn)/+ mice showed complete cutaneous syndactyly of digits 2/5 in fore- and hindlimbs (**Fig. 3b**), thus recapitulating the human phenotype.

Skeletal staining showed that the syndactyly of Dup(syn) mutants did not involve bony fusions. Digits and joints developed normally, but terminal phalanges were broad and short. *In situ* hybridization experiments in E13.5 limbs identified major changes in fingertips, where *Ihh* expression was not only increased but also broadened. These effects were weak in Dup(csp) mice, more pronounced in Dup(int) mice and most prominent in Dup(syn)/+ mice, in which *Ihh* expression extended into the distal interdigital space (**Fig. 3c**). Accordingly, the expression domains of the hedgehog downstream targets *Gli1* and *Ptch1* were broadened, and fusion of the normally separate domains was observed that was most pronounced in Dup(syn)/+ mutants. Except for *Bmp4* and *Nog*, we did not observe abnormalities in other genes or pathways involved in syndactyly and interdigital cell death (**Supplementary Fig. 7**), suggesting that hedgehog signaling alone is sufficient to induce this type of syndactyly. Next, we quantified interdigital apoptosis, which is required for digit separation[25]. Consistent with the observed phenotypes, we detected strong signal in the interdigital space in wild-type, Dup(csp) and Dup(int) embryos, but an absence of signal in the distal region in Dup(syn)/+ embryos (**Fig. 3d**). Thus, upregulation and misexpression of *Ihh* in fingertips beyond a certain threshold resulted in abnormalities of the distal phalanges, most likely by interfering with the phalanx-forming region[26], and syndactyly due to suppression of interdigital apoptosis.

In addition, Dup(syn) mutants displayed preaxial polydactyly on hindlimbs (50% penetrance; **Fig. 3e**). One major cause of polydactyly is ectopic activation of hedgehog signaling at the anterior developing limb bud[27,28]. Interestingly, Dup(syn)/+ embryos showed a prominent increase in *Ihh* expression in the distal zeugopod during hindlimb development starting at E12.5 (expression was absent at E10.5 and E11.5). As IHH is a potent diffusible morphogen, we hypothesize that the increased expression might interfere with the anterior–posterior hedgehog gradient. Thus, the observed phenotype seems to be the result of a loss of precision in spatiotemporal expression, indicating that, similar to the syndactyly, an increase in enhancer dosage can have site-specific effects.

Expression profiling by quantitative RT–PCR (qPCR) was used to quantify the effect of the duplications on gene expression (**Fig. 3f**). Whereas *Nhej1* and other nearby genes showed no alteration in expression (**Supplementary Fig. 5**), all mutants analyzed displayed increased *Ihh* expression in the skull and limbs, with the highest expression levels observed in Dup(int) mutants (up to fivefold upregulation). *In situ* hybridization of Dup(int) forelimb autopods (**Fig. 3c**) showed increased expression mainly in digits, whereas in Dup(syn) mutants the expression increase was most prominent in fingertips, consistent with the syndactyly observed. To investigate the effect of increased *Ihh* expression on skull development and suture formation, a detailed μCT analysis was performed (**Fig. 3g**). This analysis identified fusion of the metopic suture (craniosynostosis) in all mutants, but this phenotype was most pronounced in Dup(int) mice. The phenotypes observed in our mouse mutants (syndactyly, polydactyly and craniosynostosis) accurately recapitulate previous observations in human patients[1,2] (**Supplementary Fig. 8**). Thus, the induced changes in enhancer composition and dosage resulted in a disturbance of the levels and precision of gene expression, thereby causing abnormal development and disease. Interestingly, the observed phenotypes did not always correlate with the number of duplicated elements but appeared to be influenced by other factors such as the position of the duplication and the arrangement of individual elements relative to the cluster.

To investigate a possible effect of spatial configuration on the duplicated alleles, we performed 4C–seq experiments in E14.5 limbs (viewpoint at *Ihh*; **Fig. 4a**). In Dup(int)/+ mutants (with duplication of enhancers i2–i9), we observed increased interactions across the entire duplicated region. In contrast, Dup(syn)/+ mutants (with duplication of *Ihh* and enhancers i1–i5) only showed increased contact with the centromeric region of the enhancer cluster, suggesting that the centromeric *Ihh* copy created its own regulatory domain containing only the duplicated regulatory elements i1–i5 (**Fig. 4b**). The presence of two divergently oriented CTCF-binding sites near the promoter of the telomeric *Ihh* copy might explain this domain separation by limiting chromatin interaction beyond these elements. Moreover, the larger contact areas in Dup(int)/+ mutants correlate with the observed levels of *Ihh* upregulation as compared to Dup(syn)/+ mutants. As illustrated in **Figure 4c**, the syndactyly in Dup(syn)/+ mice is likely due to two types of interactions between the major fingertip enhancer i5 and the two copies of *Ihh*: one type involves long-range interactions and the other the presence of the i5 enhancer in direct proximity to *Ihh*. Together, these interactions result in localized upregulation of *Ihh* expression in the fingertips. Increased expression mediated by disconnection from a repressor element is unlikely, as none of the deletions studied resulted in any observable upregulation of *Ihh*. To further evaluate whether the observed limb phenotypes in the *Ihh*-containing duplication (Dup(syn)) merely corresponded to a gene-dosage effect, we crossed Dup(syn)/+ mice with $Ihh^{+/-}$ mice or with mice lacking the enhancer cluster (Del(2–9) mice). In both cases, double-heterozygous mice displayed the same syndactyly and polydactyly as was observed in Dup(syn)/+ mice (**Supplementary Fig. 9**), indicating that misexpression was due to the specific, partially duplicated regulatory landscape.

Our study shows that a multipartite enhancer ensemble regulates *Ihh* expression in fingertips, digit condensations, growth plates and skull sutures. The described functional redundancy appears to be a common phenomenon of these types of enhancers, as was recently shown for the α-globin and *Wap* super-enhancers[29,30]. At the *Ihh* locus, we observed a complex scenario, as not all enhancers displayed the same combination of expression domains, a phenomenon also described for the HoxD cluster and *Fgf8* (refs. 31,32). This modular nature and, in particular, correct dosage appear critical in conferring the required precision of gene expression. This is supported by our finding that an increase in enhancer number resulted in an increase in gene expression. However, this effect was site specific and dependent not only on enhancer number but also on enhancer position. CNVs, and in particular duplications, may affect this delicate balance, thereby causing over- and/or misexpression resulting in disease. The reported duplications do not interfere with topologically associating domain (TAD) boundaries, as reported at

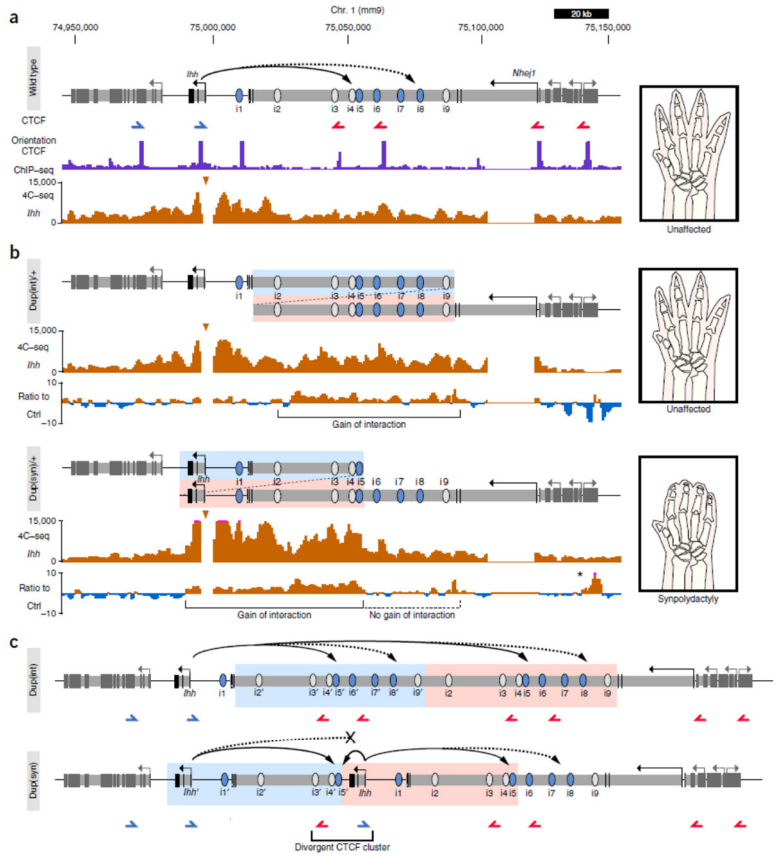


Figure 4 4C-seq identifies specific regulatory configurations in duplications. (**a**) Schematic of the wild-type *Ihh* locus. The continuous arrow indicates interaction between *Ihh* and enhancers i4–i6, while the discontinuous arrow indicates interaction of *Ihh* with i7–i9. CTCF ChIP-seq findings (E14.5 limbs; ENCODE)[3] are shown below, where blue and red arrows indicate motif orientation. 4C-seq results (viewpoint at the *Ihh* promoter) show the interaction profile in E14.5 handplates. A schematics of limb morphology is shown on the right. (**b**) Duplication of intron 3 of *Nhej1* (Dup(int); top) and a reengineered human duplication causing syndactyly (Dup(syn); bottom). Duplicated regions are shown in blue and pink. The 4C-seq profile (viewpoint at the *Ihh* promoter) and the ratio to wild-type E14.5 handplates control are shown for each duplication. Brackets indicate regions with gain of interaction. Note that there is no gain of interaction with the region containing enhancers i6–i9 (dashed bracket) in Dup(syn)/+ mice, indicating that the duplicated *Ihh* copy does not interact with this region. Observed phenotypes are schematically shown on the right. An asterisk indicates increased interactions with the *Cnppd1* and *Fam134a* genes, which do not have functional consequences (**Supplementary Fig. 5**). (**c**) Model of the regulatory interactions of the duplicated alleles. In Dup(int), *Ihh* can interact with the entire duplicated landscape, including two copies of the main digit enhancer (i5). In Dup(syn), both *Ihh* copies interact with a downstream copy of the i5 enhancer (long continuous arrows) but only the telomeric *Ihh* copy has access to i7–i9 (discontinuous arrows) because of the presence of the divergent CTCF cluster (bracket). Additionally, the duplicated i5 enhancer interacts with the telomeric *Ihh* copy because of its genomic proximity (short continuous arrow). Duplicated regions are shown in blue and pink.

the *Epha4* and *Sox9* loci[33,34], thus highlighting alternative mechanisms that should be considered when interpreting genomic duplications. Our study demonstrates the importance of analyzing regulatory elements in the complex setting of their native genomic environment, as reductionist approaches relying on reporter assays and deletions of individual enhancers insufficiently capture the multifaceted redundant and complementary functions of enhancer clusters.

URLs. FIMO, http://meme-suite.org/tools/fimo; JASPAR database, http://jaspar.binf.ku.dk/; Gene Expression Omnibus, https://www.ncbi.nlm.nih.gov/geo/.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank the sequencing core, transgenic unit and animal facilities of the Max Planck Institute for Molecular Genetics for technical assistance. This study was supported by a grant from the Deutsche Forschungsgemeinschaft to S.M. and E.K. S.M. was supported by the European Community's Seventh Framework Programme, grant agreement 602300 (SYBIL). M.O. was supported by a Swiss National Science Foundation (SNSF) fellowship. A.V. was supported by National Institutes of Health grants R01HG003988, U54HG006997, U01DE024427 and U01DE024427.

AUTHOR CONTRIBUTIONS

A.J.W., D.G.L. and S.M. conceived the study and designed the experiments. M.O. and A.V. performed LacZ experiments and analysis of individual enhancers, and A.J.W. performed analysis of *Sleeping Beauty* insertion. A.J.W. and L.W. generated transgenic mouse models. W.-L.C., J.P.d.V. and G.C. contributed to histological analysis. A.J.W., N.B. and G.C. performed qPCR, *in situ* hybridization and phenotype analysis. G.C., N.B. and W.-L.C. provided technical support. E.K., M.O. and A.V. contributed to scientific discussion. A.J.W. performed 4C–seq experiments, and V.H., M.V. and D.G.L. performed bioinformatic analysis. A.J.W., D.G.L. and S.M. wrote the manuscript with input from all authors.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Barroso, E. *et al.* Identification of the fourth duplication of upstream *IHH* regulatory elements, in a family with craniosynostosis Philadelphia type, helps to define the phenotypic characterization of these regulatory elements. *Am. J. Med. Genet. A.* **167A**, 902–906 (2015).
2. Klopocki, E. *et al.* Copy-number variations involving the *IHH* locus are associated with syndactyly and craniosynostosis. *Am. J. Hum. Genet.* **88**, 70–75 (2011).
3. ENCODE Project Consortium. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* **306**, 636–640 (2004).
4. Andersson, R. *et al.* An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455–461 (2014).
5. Bernstein, B.E. *et al.* The NIH Roadmap Epigenomics Mapping Consortium. *Nat. Biotechnol.* **28**, 1045–1048 (2010).
6. Hong, J.W., Hendrix, D.A. & Levine, M.S. Shadow enhancers as a source of evolutionary novelty. *Science* **321**, 1314 (2008).
7. Perry, M.W., Boettiger, A.N., Bothma, J.P. & Levine, M. Shadow enhancers foster robustness of *Drosophila* gastrulation. *Curr. Biol.* **20**, 1562–1567 (2010).
8. Dunipace, L., Ozdemir, A. & Stathopoulos, A. Complex interactions between *cis*-regulatory modules in native conformation are critical for *Drosophilasnail* expression. *Development* **138**, 4075–4084 (2011).
9. Cannavò, E. *et al.* Shadow enhancers are pervasive features of developmental regulatory networks. *Curr. Biol.* **26**, 38–51 (2016).
10. Perry, M.W., Bothma, J.P., Luu, R.D. & Levine, M. Precision of hunchback expression in the *Drosophila* embryo. *Curr. Biol.* **22**, 2247–2252 (2012).
11. St-Jacques, B., Hammerschmidt, M. & McMahon, A.P. Indian hedgehog signaling regulates proliferation and differentiation of chondrocytes and is essential for bone formation. *Genes Dev.* **13**, 2072–2086 (1999).
12. Andrey, G. *et al.* Characterization of hundreds of regulatory landscapes in developing limbs reveals two regimes of chromatin folding. *Genome Res.* **27**, 223–233 (2017).
13. Guo, Y. *et al.* CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function. *Cell* **162**, 900–910 (2015).
14. Rao, S.S. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
15. Sanborn, A.L. *et al.* Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl. Acad. Sci. USA* **112**, E6456–E6465 (2015).
16. Vietri Rudan, M. *et al.* Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell Reports* **10**, 1297–1309 (2015).
17. Ruf, S. *et al.* Large-scale analysis of the regulatory architecture of the mouse genome with a transposon-associated sensor. *Nat. Genet.* **43**, 379–386 (2011).
18. Creyghton, M.P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. USA* **107**, 21931–21936 (2010).
19. Pollard, K.S., Hubisz, M.J., Rosenbloom, K.R. & Siepel, A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* **20**, 110–121 (2010).
20. Visel, A., Minovitsky, S., Dubchak, I. & Pennacchio, L.A. VISTA Enhancer Browser—a database of tissue-specific human enhancers. *Nucleic Acids Res.* **35**, D88–D92 (2007).
21. Kraft, K. *et al.* Deletions, inversions, duplications: engineering of structural variants using CRISPR/Cas in mice. *Cell Rep.* http://dx.doi.org/10.1016/j.celrep.2015.01.016 (2015).
22. Buck, D. *et al.* Cernunnos, a novel nonhomologous end-joining factor, is mutated in human immunodeficiency with microcephaly. *Cell* **124**, 287–299 (2006).
23. Li, G. *et al.* Lymphocyte-specific compensation for XLF/cernunnos end-joining functions in V(D)J recombination. *Mol. Cell* **31**, 631–640 (2008).
24. Vera, G. *et al.* Cernunnos deficiency reduces thymocyte life span and alters the T cell repertoire in mice and humans. *Mol. Cell. Biol.* **33**, 701–711 (2013).
25. Hernández-Martínez, R., Castro-Obregón, S. & Covarrubias, L. Progressive interdigital cell death: regulation by the antagonistic interaction between fibroblast growth factor 8 and retinoic acid. *Development* **136**, 3669–3678 (2009).
26. Witte, F., Chan, D., Economides, A.N., Mundlos, S. & Stricker, S. Receptor tyrosine kinase-like orphan receptor 2 (ROR2) and Indian hedgehog regulate digit outgrowth mediated by the phalanx-forming region. *Proc. Natl. Acad. Sci. USA* **107**, 14211–14216 (2010).
27. Anderson, E., Peluso, S., Lettice, L.A. & Hill, R.E. Human limb abnormalities caused by disruption of hedgehog signaling. *Trends Genet.* **28**, 364–373 (2012).
28. Hill, R.E. How to make a zone of polarizing activity: insights into limb development via the abnormality preaxial polydactyly. *Dev. Growth Differ.* **49**, 439–448 (2007).
29. Hay, D. *et al.* Genetic dissection of the α-globin super-enhancer *in vivo*. *Nat. Genet.* **48**, 895–903 (2016).
30. Shin, H.Y. *et al.* Hierarchy within the mammary STAT5-driven Wap super-enhancer. *Nat. Genet.* **48**, 904–911 (2016).
31. Marinić, M., Aktas, T., Ruf, S. & Spitz, F. An integrated holo-enhancer unit defines tissue and gene specificity of the Fgf8 regulatory landscape. *Dev. Cell* **24**, 530–542 (2013).
32. Montavon, T. *et al.* A regulatory archipelago controls Hox genes transcription in digits. *Cell* **147**, 1132–1145 (2011).
33. Franke, M. *et al.* Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature* **538**, 265–269 (2016).
34. Lupiáñez, D.G. *et al.* Disruptions of topological chromatin domains cause pathogenic rewiring of gene–enhancer interactions. *Cell* **161**, 1012–1025 (2015).

ONLINE METHODS

Experimental design. No statistical methods were used to predetermine sample size. All experiments and analyses were performed using samples from at least three different animals and were repeated at least two times in the laboratory. Samples/animals were included or excluded according to genotype by PCR. Experiments were not randomized, and investigators were not blinded to allocation during experiments and outcome assessment.

ES cell targeting and transgenic mouse strains. Mouse embryonic stem (ES) cell culture was performed as described previously[21]. ES and feeder cells were tested for mycoplasma contamination using a Mycoalert detection kit (Lonza) and the Mycoalert assay control set (Lonza).

Duplications and deletions were generated in G4 ES cells (129/Sv × C57BL/6 F_1 hybrid) using CRISVar as described previously[21]. Target regions, sizes and guide sequences are listed in **Supplementary Table 2**. Embryos and live animals derived from ES cells were generated by diploid or tetraploid complementation[35]. Genotyping was performed by PCR analysis.

A *Sleeping Beauty* (SB) cassette[17] was inserted in G4 ES cells at the center of the third intron of the *Nhej1* gene (chr. 1: 75,060,87; mm9), by homologous recombination using standard protocols[36]. The *Sleeping Beauty* transgene carries a single *lacZ* reporter gene with a minimal human β-globin promoter and a neomycin-resistance cassette, flanked by transposable elements. Coordinates and primer sequences for amplifying homology sequences are provided in **Supplementary Table 3**. Positive ES cell clones were injected into donor blastocysts to generate chimeras. The neomycin-resistance cassette was removed by crossing chimaeric animals with a Flpe-deleter line. Genotyping was performed by PCR analysis.

Mouse strains were maintained by crossing the strains with C57BL/6J mice. All animal procedures were conducted as approved by the local authorities (LAGeSo Berlin) under license numbers G0368/08 and G0247/13.

***In vivo* enhancer validation.** Putative enhancer regions were amplified by PCR from mouse genomic DNA and cloned into a *Hsp68* promoter–*lacZ* reporter vector as previously described[20] (**Supplementary Table 4**). Transgenic embryos were generated and tested for LacZ reporter activity at E14.5 and E17.5. All animal work performed at the Lawrence Berkeley National Laboratory was reviewed and approved by the institutional Animal Welfare and Research Committee (AWRC). Sample sizes were selected empirically on the basis of our previous experience of performing transgenic mouse assays for >2,000 total putative enhancers. A summary of all transgenic mice can be found in **Supplementary Table 1**. As all transgenic mice were treated with identical experimental conditions, and as there were no groups of animals directly compared in this section of the study, randomization and experimenter blinding were unnecessary and were not performed.

Quantitative real-time PCR. Handplates (E13.5), forelimb and hindlimb growth plates (E17.5) and cranium (E17.5) were dissected from wild-type and mutant embryos ($n \geq 3$) in ice-cold PBS/DEPC and immediately frozen in liquid nitrogen. RNA isolation was performed using the RNeasy kit (Qiagen), and cDNA was transcribed using the TaqMan Reverse Transcription kit (Roche) according to the specifications of the manufacturer. qPCR was performed using SYBR Green (Qiagen) on an ABI Prism HT 7900 Real-Time Cycler. *GAPDH* was used as an internal control, and fold changes were calculated by relative quantification ($2^{-\Delta\Delta Ct}$). Primers are summarized in **Supplementary Table 5**.

4C–seq. 4C–seq libraries were generated from microdissected E14.5 mouse forelimb tissue (digits 2–5) as described previously[37]. The starting material for all 4C–seq libraries was 5×10^6 to 1×10^7 cells. All 4C–seq experiments were carried out in heterozygous animals, and results were compared to those in wild-type controls. 4-bp cutters were used as primary (Csp6I) and secondary (BfaI) restriction enzymes. A total of 1 to 1.6 µg of DNA was amplified by PCR (primer sequences in **Supplementary Table 6**). All samples were sequenced with Illumina HiSeq technology according to standard protocols. 4C–seq experiments were carried out in two biological replicates in wild-type, Dup(int) and Dup(syn)/+ mutants. A representative result is shown in **Figure 4**.

For 4C–seq data analysis, reads were preprocessed and mapped to a corresponding reference (mm9) using BWA-MEM[38]; coverage was normalized as reported previously[34]. The viewpoint and adjacent fragments 1.5 kb up- and downstream were removed, and a window of two fragments was chosen to normalize the data per million mapped reads (RPM). To compare the interaction profiles of different samples, we obtained the $\log_2$-transformed fold change for each window of normalized reads. To obtain ratios, duplicated regions were excluded for calculation of the scaling parameter used in RPM normalization. Code is available upon request.

CTCF motif orientation analysis. Orientation of the motifs within conserved CTCF peaks was obtained using FIMO (see URLs) with standard parameters[39]. The CTCF motif[40] was obtained from the JASPAR database (see URLs).

Phenotypic analysis. Phenotypic analysis for mutant mouse lines was carried out for at least three animals per analysis and developmental stage (E17.5, P7 and P70), in homo- and heterozygous animals. The penetrance of phenotypes was determined by analyzing $n > 20$ animals, and a genotype was considered fully penetrant if all mutants were similarly affected.

Microcomputer tomography. Skulls and autopods of control and mutant mice ($n > 3$) were scanned using a Skyscan 1172 X-ray microtomography system (Brucker microCT, Belgium) at 10 µm resolution. 3D model reconstruction and length measurements were performed with the Skyscan image analysis software CT-Analyser and CT-volume (Brucker microCT, Belgium). Cross-sections were performed at 10 µm resolution. Relative length was determined relative to wild-type controls.

Whole-mount *in situ* hybridization and skeletal preparations. Whole-mount *in situ* hybridization was performed in wild-type and mutant E13.5 embryos ($n = 4$) according to standard procedures. All probes were generated by PCR amplification using mouse limb bud cDNA. For skeletal preparations, wild-type and mutant E17.5 embryos ($n = 4$) were stained with Alcian blue/Alizarin red according to standard protocols.

LacZ staining. E14.5 and E17.5 mouse embryos ($n > 5$) were dissected in cold PBS, fixed in 4% paraformaldehyde (PFA)/PBS on ice for 30 min, washed twice with ice-cold PBS, washed once at room temperature (19–24 °C) and then stained overnight for β-galactosidase activity in a humid chamber at 37 °C as previously described[17]. After staining, embryos were washed in PBS and stored at 4 °C in 4% PFA/PBS.

Statistical analyses. Results are presented as the mean ± s.d. of at least three independent biological replicates ($n \geq 3$). Statistical differences between the means were examined by two-sided Student's *t* test. $P < 0.05$ was considered statistically significant. A prespecified effect size was not defined.

Code availability. Custom computer codes used to generate results reported in the manuscript will be made available upon request.

Data availability. Sequencing data are available from the Gene Expression Omnibus under accession GSE95062.

35. Artus, J. & Hadjantonakis, A.K. Generation of chimeras by aggregation of embryonic stem cells with diploid or tetraploid mouse embryos. *Methods Mol. Biol.* **693**, 37–56 (2011).
36. Hooper, M., Hardy, K., Handyside, A., Hunter, S. & Monk, M. HPRT-deficient (Lesch–Nyhan) mouse embryos derived from germline colonization by cultured cells. *Nature* **326**, 292–295 (1987).
37. van de Werken, H.J. *et al.* 4C technology: protocols and data analysis. *Methods Enzymol.* **513**, 89–112 (2012).
38. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
39. Grant, C.E., Bailey, T.L. & Noble, W.S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011).
40. Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823–837 (2007).

doi:10.1038/ng.3939

ACKNOWLEDGEMENTS

I would like to express my gratitude to Prof. Stefan Mundlos for giving me the great opportunity to join his amazing laboratory in Berlin, where I had the possibility to learn a lot, both from a scientific and personal point of view. I also want to say thank to Prof. Roberto Ciccone and Prof. Orsetta Zuffardi from the Italian side.

I am thoroughly and eternally grateful to Martin Franke, whose help and guidance have been so precious for my project and especially during the writing of the thesis. This thesis would not be as great as it is without him. Thank you for believing in me and supporting me always! It was an immense pleasure to work with you, both as scientist and person. You will be missed, I already know, but I wish you the most brilliant future you can have.

I am also extremely thankful to Christina Paliou, who gave me the strength and the calm at the same time. She saw me laugh and cry, she was next to me at every up or down, no matter the day or the hour. When such incredible people enter your life, you can just consider yourself a very lucky person, as they are both your teachers and your friends.

A huge thank goes also to my sweet Chiara Anania, whose support gave me even more strength. I consider our friendship as something very special, as I have the impression that we know each other since a lifetime. It's incredible how life gives us such wonderful people.

I am also thankful to Ivana Jerković, who found a bit of time to help during the writing of my thesis. You are a great person and your sweet hugs always made me feel at home.

I am grateful to Lila Allou, who never stopped to support me and who was always ready to tell me that I could make it.

I would also like to thank Alexandra Despang, one of the kindest people I know. She always helped me for everything and she made me feel very welcomed since the very first time I joined the Mundlos group. Thank you for being the wonderful person you are.

I would also like to express my gratitude to Darìo Lupiáñez, Fany Martinez Real, Bjørt Kragesteen, Salaheddine Ali, Arunima Murgai, Sophie Pöhle, Xiaoyan Wei, George Kotsaris, Asita Stiege, Ute Fischer, Norbert Brieske, to the rest of the present and past members of the Mundlos group and not only. I want to say thank for being my second great big family.

I am also extremely grateful to Magdalena Socha, whose friendship matters so much to me and I know that I can always rely on her. We shared so many adventures, events, celebrations, critical moments, always side by side.

Words will never be enough to thank my wonderful Mum, who has always a solution for every circumstance and especially the right words at the right time to give me the possibility to believe that nothing is impossible. Her love, support and her being proud of me are my greatest strength.

Words would not be enough also for my amazing Brother, who is always so proud of me. This thesis is dedicated to him and to his incredible strength in facing the hardest challenges of life. He gave me one more reason to believe that life can be incredible and that together we can face everything.

I am also so grateful to Francesco, once again by my side for this next scientific step. He is one of my biggest supporters, who sometimes believes in me more than what I do. He has been a very good and patient listener and I am so happy to share the joy of this thesis with him. Thank you, my love, for being always by my side.

A special thank goes also to my wonderful friend Amal, whose words and messages gave me always a reason to smile and to believe that I am so lucky to have a friend like her. Thanks for being part of my life in such amazing way.

I am also grateful to Anna, one of the biggest sources of motivation. Your strength was my strength.

Finally, I want to say thank to the dear Pia, to my friend Eleonora, to my father, my aunts, my grandmothers, to the rest of my family and to all the great people that always supported me.