Position effects at the *FGF8* locus are associated with femoral hypoplasia

Magdalena Socha,[1,15] Anna Sowińska-Seidler,[1,15] Uirá Souto Melo,[2,3] Bjørt K. Kragesteen,[2,4] Martin Franke,[2,5] Verena Heinrich,[6] Robert Schöpflin,[2,3,6] Inga Nagel,[7] Nicolas Gruchy,[8] Stefan Mundlos,[2,3,9] Varun K.A. Sreenivasan,[10] Cristina López,[11,12] Martin Vingron,[6] Ewelina Bukowska-Olech,[1] Malte Spielmann,[7,10,13,16,*] and Aleksander Jamsheer[1,14,16,*]

Summary

Copy-number variations (CNVs) are a common cause of congenital limb malformations and are interpreted primarily on the basis of their effect on gene dosage. However, recent studies show that CNVs also influence the 3D genome chromatin organization. The functional interpretation of whether a phenotype is the result of gene dosage or a regulatory position effect remains challenging. Here, we report on two unrelated families with individuals affected by bilateral hypoplasia of the femoral bones, both harboring *de novo* duplications on chromosome 10q24.32. The ~0.5 Mb duplications include *FGF8*, a key regulator of limb development and several limb enhancer elements. To functionally characterize these variants, we analyzed the local chromatin architecture in the affected individuals' cells and re-engineered the duplications in mice by using CRISPR-Cas9 genome editing. We found that the duplications were associated with ectopic chromatin contacts and increased *FGF8* expression. Transgenic mice carrying the heterozygous tandem duplication including *Fgf8* exhibited proximal shortening of the limbs, resembling the human phenotype. To evaluate whether the phenotype was a result of gene dosage, we generated another transgenic mice line, carrying the duplication on one allele and a concurrent *Fgf8* deletion on the other allele, as a control. Surprisingly, the same malformations were observed. Capture Hi-C experiments revealed ectopic interaction with the duplicated region and *Fgf8*, indicating a position effect. In summary, we show that duplications at the *FGF8* locus are associated with femoral hypoplasia and that the phenotype is most likely the result of position effects altering *FGF8* expression rather than gene dosage effects.

Introduction

Copy-number variations (CNVs), such as microdeletions and microduplications, are a common cause of congenital diseases.[1–3] The pathogenicity of CNVs is interpreted primarily on the basis of their effect on gene dosage[4–6] because many human genes are known to be dosage sensitive. For instance, several *de novo* or inherited deletions or duplications are known to result in disease phenotypes. However, recent studies show that the influence of CNVs is not only mediated by the direct perturbation of the gene dosage but also due to the change in the copy number of regulatory elements or the modification of the 3D genome architecture. The latter is thought to occur via the disruption of the higher-order chromatin organization, such as topologically associating domains (TADs), which are megabase-scale regulatory units insulated from each other by boundary elements.[7,8]

For congenital limb malformations, several CNVs affecting noncoding regulatory elements have recently been described.[3,9,10] However, the functional interpretation as to whether a phenotype is the result of the ensuing gene dosage or of an indirect, regulatory position effect remains challenging.

In the current study, we describe two unrelated families affected with isolated bilateral hypoplasia of the femoral and pelvic bones, a rare congenital anomaly with unknown genetic origin. All affected individuals carry duplications encompassing *FGF8* (MIM: 612702). We use the mouse limb as a model system to study the effect of the *FGF8* duplications on limb development and gene regulation. By combining genome editing in transgenic mice with chromatin and expression analysis, we identify that

[1]Department of Medical Genetics, Poznan University of Medical Sciences, Poznan 60-806, Poland; [2]Max Planck Institute for Molecular Genetics, RG Development and Disease, Berlin 14195, Germany; [3]Institute for Medical and Human Genetics, Charité Universitätsmedizin Berlin, Berlin 10117, Germany; [4]Department of Immunology, Weizmann Institute of Science, Rehovot 76100, Israel; [5]Centro Andaluz de Biología del Desarrollo, Consejo Superior de Investigaciones Científicas/Universidad Pablo de Olavide, Seville 41013, Spain; [6]Department of Computational Molecular Biology, Max Planck Institute for Molecular Genetics, Berlin 14195, Germany; [7]Institute of Human Genetics, Universitätsklinikum Schleswig Holstein Campus Kiel and Christian-Albrechts-Universität, Kiel 24105, Germany; [8]Department of Genetics, EA7450 BioTARGen, Normandie Université, UNICAEN, CHU de Caen Normandie, Caen 14000, France; [9]Berlin-Brandenburg Center for Regenerative Therapies, Charité Universitätsmedizin Berlin, Berlin 13353, Germany; [10]Institute of Human Genetics, Universitätsklinikum Schleswig Holstein Campus Lübeck and University of Lübeck, Lübeck 23562, Germany; [11]Institut d'Investigacions Biomèdiques August Pi i Sunyer, Ciber de Cáncer, Barcelona, Spain; [12]Centro de Investigación Biomédica en Red de Cáncer, Madrid, Spain; [13]Human Molecular Genomics Group, Max Planck Institute for Molecular Genetics, Berlin 14195, Germany; [14]Centers for Medical Genetics GENESIS, Poznan 60-529, Poland
[15]These authors contributed equally
[16]These authors contributed equally
*Correspondence: malte.spielmann@uksh.de (M.S.), jamsheer@wp.pl (A.J.)
https://doi.org/10.1016/j.ajhg.2021.08.001.

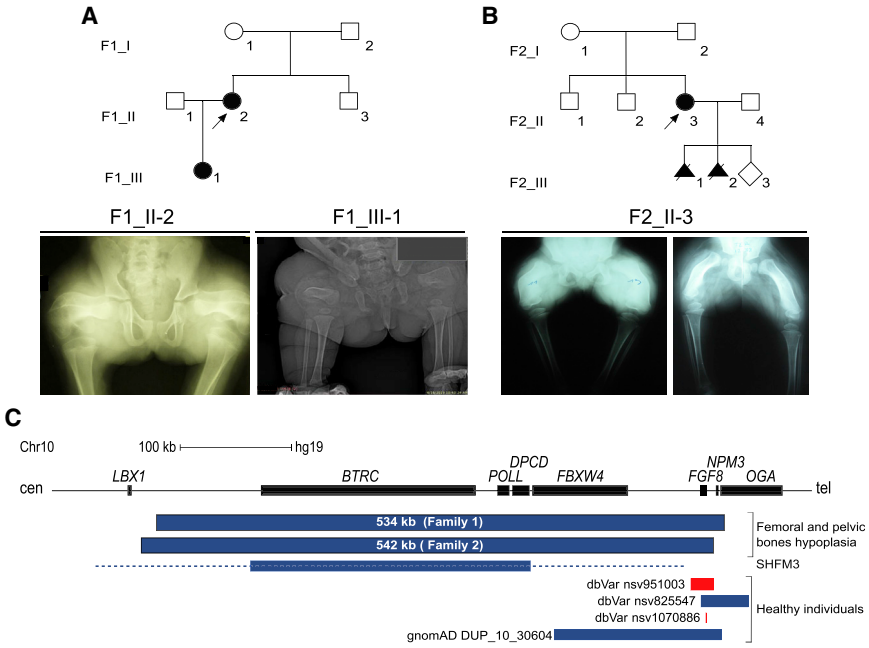


Figure 1. Two 10q24.32 duplications in unrelated families affected by isolated bilateral hypoplasia of the femoral and pelvic bones

(A and B) Pedigree of family 1 and radiographs of the proband F1_II-2 (left) at the age of 18 years and individual F1_III-1 at the age of 4 months, both of which show extremely short and hypoplastic femoral bones and moderate hypoplasia of the pelvic bone structures (A). Pedigree of family 2 and radiographs of the affected individual F2_II-3 at the age of 14 months (left) and 4 years (right) showing markedly hypoplastic, shortened, and bowed femoral bones as well as hypoplasia of the pelvic bones with narrow pelvis (B). The probands (F1_II-2 and F2_II-3) are indicated by black arrows.

(C) Mapping of the 10q24.32 duplications of the individuals described herein identified by array CGH, known duplications causative of split hand/foot malformation (SHFM3),[11] and CNVs in healthy individuals.[12–16] The dashed line represents the genomic region being duplicated in individuals with SHFM3. Duplications and deletions are colored in blue and red, respectively.

regulatory position effects at the *FGF8* locus cause femoral hypoplasia.

Subjects and methods

Subjects and ethics approval

The study was performed with the approval of the Review Board of the Poznan University of Medical Sciences and the CHU Caen ethics committee. All individuals were enrolled with written informed consent for participation in the study. The clinical evaluation included medical history interviews, a physical examination, and review of medical records. Blood samples or buccal swabs were obtained from each participating individual, and DNA was extracted via standard procedures.

To facilitate tracking of the members of family 1 and family 2, we numbered individuals by using prefixes F1 and F2, respectively. For pedigrees, see Figures 1A and 1B.

Paternity testing

The paternity of individual F1_I-2 was confirmed with a commercial Aneuploidy QF-PCR kit (Devyser Compact v.3, Devyser).

Microarray-based comparative genomic hybridization

For individual F1_II-2, microarray-based comparative genomic hybridization (array CGH) was performed by means of 1.4M NimbleGen oligonucleotide whole-genome array (Roche NimbleGen) according to the standard protocol provided by the manufacturer. The analysis was done with the Deva software (Roche NimbleGen) with the following settings: aberration algorithm, ADM-2; threshold, 6.0; window size, 0.2 Mb; filter, 5 probes; log2 ratio = 0.29. The profile of genomic aberrations was visualized in the SignalMap 1.9.0.05 software (NimbleGen Systems).

For individual F1_III-1 and family 2, array CGH was performed with the use of SurePrint G3 Human CGH Microarrays, 1×1M and 4×180K (Agilent, Agilent Technologies, Santa Clara, CA), following the manufacturer's protocol and analyzed with Agilent CytoGenomics 5.0.2.5 software (Agilent, Agilent Technologies, Santa Clara, CA). Analysis settings for Agilent arrays were as described above, except the filter, which was set to 3 probes.

CNVs were determined with thresholds of 0.4 for gains and −0.4 for losses. Commercial reference male or female genomic DNA (Promega or Agilent) was used for hybridization. Data (phenotypes and array CGH results) were submitted to the DECIPHER database; accession numbers are DECIPHER: 427511, 427699, and 427701.

Databases and *in silico* analysis

For the assessment of the CNVs, we used the following databases: DECIPHER, ClinVar, DGV, gnomAD SVs (v.2.1), and the VISTA Enhancer Browser.

Quantitative real-time polymerase chain reaction

We performed quantitative real-time polymerase chain reaction (real-time qPCR) to confirm the CNVs and to narrow down their genomic coordinates as well as to perform co-segregation analysis in the individuals with 10q24 duplication and their family members. We followed a previously described protocol[17] for individuals from family 1 and F2_III-1 and F2_III-2 from family 2. A different protocol[18] was used for testing individuals F2_I-1, F2_I-2, and F2_II-3. In short, we used a SYBR Green PCR Master Mix (Applied Biosystems) and ran all reactions in triplicate on a ViiA 7 or 7500 Real-Time thermal cycler (Applied Biosystems). Copy number analysis was performed with the comparative $2^{-\Delta\Delta CT}$ method with noncommercial healthy control DNA as a calibrator. Two genes were used for normalization—*ALB* or *PRKAB2*—and an internal sample control (sex determination) was performed by targeting of *F8* or *OFD1*. For primer sequences, see Table S1.

Breakpoint sequencing

We established the exact genomic coordinates of the aberrations by means of a series of long-range and nested-PCR reactions by using primers specific to the DNA fragment overlapping the 5′ and 3′ ends of the 10q24.32 duplications. For family 1, we performed the long-range PCR by following a previously described protocol[17] with the use of the Expand Long Template PCR System v.24 (Roche) for the target region amplification. For family 2, we utilized the Ranger Mix (Bioline) by following the producer's protocol. The obtained PCR products were sequenced with the Applied Biosystems PRISM 3700 DNA Analyzer with the BigDye Terminator v.3.1 Cycle Sequencing Kit (Applied Biosystems). See supplemental material and methods for nested-PCR conditions and Table S2 for primer sequences.

FGF8 and *BTRC* relative expression in skin fibroblasts derived from individual F1_II-2

Total RNA was extracted from skin fibroblasts derived from individual F1_II-2 and six controls with RNeasy Mini Kit (QIAGEN #74106) according to the manufacturer's protocol. 1 μg of the total RNA of each sample was reverse transcribed with the use of TaqMan Gold RT-PCR Kit (Applied Biosystems #N808-0233). The relative expression level of *FGF8* and *BTRC* (MIM: 603482) was established with the comparative $2^{-\Delta\Delta CT}$ method in which the individuals' target genes expression value was normalized to the *GAPDH* reference gene. We used the $2^{-\Delta\Delta CT}$ method to calculate fold change between the affected individual and the mean value of six controls. The fluorometric real-time qPCR assay was performed on ABI Prism HT 7900 Real-Time Cycler (Applied Biosystems) and SYBR Green PCR Master Mix (Applied Biosystems) on 384-well setup in a total volume of 12 μL. The reaction conditions were as previously described.[17] All reactions were carried out in triplicates. Statistical significance was calculated via one-tailed single-sample *Z* test. For primer sequences, see Table S3.

Chromosome conformation capture libraries preparation and sequencing

We generated circular chromosome conformation capture (4C) libraries from fibroblasts of individual F1_II-2 and three controls as described previously[19] by using 5×10^6 cells per sample. The 4C sequencing (4C-seq) experiment was designed at three viewpoints: *FGF8* transcription start site (TSS), *BTRC* TSS, and the *Fgf8 cis*-regulatory element (RE)—cloned enhancer 66 (CE66).[20] See Table S4 for primer sequences. For each viewpoint, the 4 bp cutters *Csp6I* and *DpnII* were used as primary and secondary restriction enzymes, respectively. A total of 1.6 μg of each 4C library was amplified by PCR. The amplified samples were sequenced with the Illumina HiSeq 4000 system according to standard procedures.

Capture Hi-C (cHi-C) was performed in wild-type (WT) and heterozygous embryonic day (E) 11.5 limb buds as described elsewhere.[21] Briefly, we fixed and lysed cells to obtain intact nuclei. Chromatin was digested with *DpnII*, and DNA ends were subsequently ligated. Purified 3C libraries were sheared to 300–600 bp fragments via Covaris (S220 focused-ultrasonicator). The DNA ends were repaired and prepared for Illumina sequencing according to instructions from Agilent. We hybridized libraries to custom-designed SureSelect (Agilent) capture probes to enrich chromatin interactions at the *Fgf8 locus* (chr19: 44,440,000–46,400,000, mm9). Samples were sequenced with Illumina HiSeq technology (50 bp paired-end) according to standard protocols.

Bioinformatic pipeline

4C-seq raw data were pre-processed as reported previously.[9] Sequencing reads were mapped to a reference genome (humanG1Kv37) with Burrows-Wheeler aligner,[22] resulting in over one million mapped reads with a *cis*/overall ratio ≥ 60%. 4C-seq contacts per fragment (1st RE) were normalized to reads per million (RPM) and further analyzed for the region chr10: 102,000,000–104,000,000 via customized scripts. Reads overlapping the viewpoint and adjacent fragments 1.5 kb upstream and downstream were excluded from further analyses. We applied a running window averaging of ten fragments to smooth the read counts per fragment and exported them to BedGraph format. To compare the interaction profiles for each viewpoint, we calculated log2 ratios between the signal from the individual and the average signal from the three control samples. The signal from the additional copy of the duplicated region was excluded from the log2 calculations. All 4C-seq experiments were run in biological duplicate.

In the preprocessing step of the cHi-C data, we truncated the reads in fastq files to 50 bp to obtain the same initial read length for both samples and subsequently processed them with the HiCUP pipeline v.0.6.1[23] (Nofill:1, no size restriction), which was set up with Bowtie2 v.2.3.4.1[24] and the reference genome mm9. We used Juicer tools v.1.19.02[25] to generate binned contact maps from valid and unique read pairs. For the generation of contact maps, only the enriched region chr19: 44,440,001–46,400,000 was considered and filtered for reads-pairs mapping to the defined region. The coordinates were shifted by the offset of the enriched region, and the data were imported with the Juicer tools "pre" command with a custom chrom.sizes file that contained only the size of the enriched region (1.96 Mb). Only read pairs with a mapping quality (MAPQ) ≥ 30 were considered. After exporting both the raw count maps and the KR-normalized[25–27] maps with 10 kb bin size with the Juicer tools "dump" command, the coordinates were shifted back to mm9 coordinates. The matrix-balancing approach of the Knight-Ruiz (KR) normalization affects the Hi-C signal intensity of duplications because it balances between all rows and columns. As a consequence, the signal in the duplicated region would be scaled down to fit to the loci with regular copy numbers.

In order to consider the duplicated region in the mutant sample explicitly, we created additional maps by applying the copy-number variation effects (LOIC)-normalization[28] (python package iced v.0.5.3[29]) to raw count maps. For the mutant sample, the number of alleles was set to three for all bins overlapping with the duplicated region (chr19: 45,332,291–45,826,577). For the WT sample, the number of alleles was assumed to be two throughout the entire genomic region for the normalization.

To compute subtraction maps, we first scaled each contact map according to its overall signal by multiplying each value by 10^6/overall sum. For the computation of this scaling factor, the duplicated region as well as the main diagonal were excluded for both samples. Subtraction maps were computed for raw-count maps. We displayed contact maps and subtraction maps as heatmaps in which we truncated the values above the 99th percentile to improve visualization.

Animal procedures

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

Embryonic stem cells (ESCs) culture and transgenic mouse strains

Transgenic mouse strains were generated by means of the CRISPR-Cas9 editing system according to previously published protocols.[30,31] To obtain transgenic alleles, we designed single guide RNAs (sgRNAs) by using the CRISPR Zhang laboratory design tool. See supplemental material and methods for the positive clone screening, Table S5 for the sgRNAs sequences, and Table S6 for genotyping primer sequences. Transgenic mice were generated by morula aggregation and tetraploid complementation according to previously published protocol.[32] CD1 female mice were used as foster mothers. ESCs will be available upon request. We performed four and two independent tetraploid aggregations of the modified ESC to obtain the *Fgf8*$^{+/dup}$ and *Fgf8*$^{dup/del}$ mutants, respectively. Two or three foster mothers from each aggregation were sacrificed at a desired time point, and eight to 12 embryos were obtained on average from each foster mother.

Whole-mount *in situ* hybridization

We assessed the *Fgf8* mRNA expression in E11.5 and E12.5 mouse embryos (n = 3 from two independent aggregations for the *Fgf8*$^{+/dup}$ mutant) by whole-mount *in situ* hybridization (WISH) by using a digoxigenin-labeled *Fgf8* antisense riboprobe transcribed from a cloned *Fgf8* probe (PCR DIG Probe Synthesis Kit, Roche) following a previously described protocol.[33] We used the *Fgf8* probe sequences published in Minina et al. 2005.[34] For imaging of the stained embryos, a Zeiss SteREO Discovery.V12 microscope with cold light source CL9000 and Leica DFC420 digital camera were used.

Skeletal preparations

Alizarin red/alcian blue staining for bone and cartilage was performed on WT and mutant E18.5 fetuses (n = 2 from two independent aggregations for each transgenic line) according to a previously published protocol.[35] Documentation of the skeletal preparations was done in either 25% or 30% glycerol after achieving full tissue digestion.

Results

Femoral and pelvic hypoplasia is associated with duplications at the *FGF8* locus

We studied two unrelated families with individuals presenting isolated bilateral hypoplasia of the femoral and pelvic bones, a phenotype with unknown underlying cause. In family 1 (Figure 1A), the two affected individuals (F1_II-2, female, and F1_III-1, daughter) exhibited isolated bilateral congenital lower limb malformation composed of extremely short and hypoplastic femoral bones and moderate hypoplasia of the pelvic bone structures. Both individuals had normal psychomotor development and no other associated abnormalities. In the second unrelated family (Figure 1B), the proband (F2_II-3, female) showed a similar phenotype, including bilateral femoral and pelvic hypoplasia, a narrow pelvis, and normal psychomotor development. In two of her pregnancies, prenatal ultrasound of the fetuses revealed developmental abnormalities consisting of short lower limbs (data not shown). The first fetus (F2_III-1, male) showed abnormal positioning of the feet and short bilateral femurs, and the second fetus (F2_III-2, female) had a club foot on the right and short bilateral femurs below the third percentile; these pregnancies were terminated in the 14th and 15th gestational weeks, respectively (Figures 1A and 1B).

We performed genome-wide copy number analysis by array CGH in probands from these two families. In family 1 (individual F1_II-2), we identified a ~519.8 kb interstitial duplication 10q24.32 (arr[GRCh37] 10q24.32(103,014,233_103,534,063)x3; [Figure S1]). An overlapping ~538.9 kb duplication 10q24.32 (arr[GRCh37] 10q24.32(103,003,677_103,542,575)x3) was identified in family 2 (individual F2_II-3) (Figure S1). Breakpoint sequencing revealed the true size and span of the aberrations: 533,943 kb (arr[GRCh37] 10q24.32(103,012,761_103,546,704)x3) in family 1 and 542,061 kb (arr[GRCh37] 10q24.32(103,001,852_103,543,913)x3) in family 2 (Figure 1C). Segregation analysis by real-time qPCR confirmed the duplications in the affected individuals and excluded them in unaffected parents and siblings, confirming *de novo* occurrence in the probands and co-segregation of the 10q24.32 duplications with the malformations (Figures S2A–S2C). The overlapping duplications harbor six protein-coding genes (*BTRC*, *POLL* [MIM: 606343], *DPCD* [MIM: 616467], *FBXW4* [MIM: 608071], *FGF8*, and *NPM3* [MIM: 606456]), and breakpoint sequencing revealed tandem orientation in both families (Figure S2D).

Comparative analysis revealed that the femoral hypoplasia duplications reported here almost completely overlap with 10q24.32 duplications that are known to associate with split hand/foot malformation (SHFM3 [MIM: 246560]), an entirely different limb malformation.[11] The only unique gene to the femoral hypoplasia duplications was *FGF8*, a key regulator of limb development that is known to direct proximo-distal growth[36–38] (Figure 1C). While initially gene dosage effects of *FGF8* seemed like a promising explanation for the femoral hypoplasia phenotype, we also identified several healthy individuals without skeletal abnormalities carrying smaller duplications that include *FGF8* in control cohorts[12–16] (Figure 1C). Therefore, pathomechanisms alternative to simple gene dosage were considered.

Ectopic chromatin interaction of *FGF8* in skin fibroblasts

Structural variants have been shown to rearrange the chromatin architecture of the genome, facilitating ectopic enhancer-promoter contacts to cause misexpression and disease.[9,21,39] Therefore, we performed circular chromosome conformation capture sequencing (4C-seq) analysis in skin fibroblasts of the affected individual F1_II-2 and three healthy control individuals. Recent data show that 4C techniques can effectively be used for studying of TAD structure in fibroblasts of affected individuals (Lupiáñez et al.[9] and Franke et al.[21]). We used the promoters of *FGF8* and *BTRC*, as well as the apical ectodermal ridge (AER) enhancer CE66,[20] as viewpoints to investigate the local chromatin interactions. The 4C-seq signal and

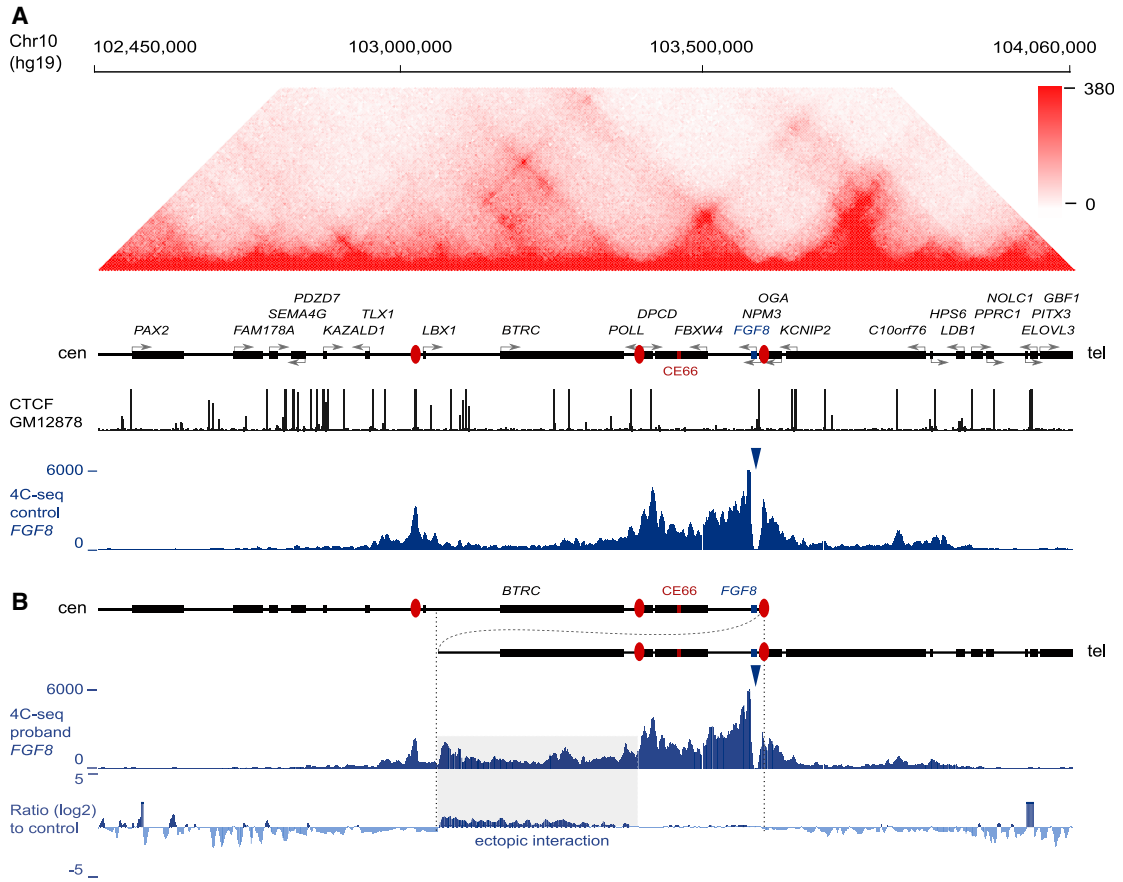


Figure 2. 4C-seq in control and F1_II-2 skin fibroblasts reveal altered 3D genome architecture at the *FGF8* locus
(A) Top: Hi-C map at the *FGF8* locus (Hg19; chr10: 102,450,000–104,060,000) derived from GM12878 cells.[27] Predicted TAD boundaries are indicated by red ovals. Middle: genes and their respective transcription start sites (TSSs); CTCF ChIP-seq tracks of the examined locus in the GM12878 cells.[40,41] Bottom: 4C-seq of control fibroblasts with the *FGF8* TSS as the viewpoint (indicated by triangle) representing the WT interactions at the *FGF8* locus.
(B) 4C-seq performed on fibroblasts of individual F1_II-2 showing interaction profiles from the *FGF8* TSS viewpoint (indicated by triangle). Top: schematic representation of the duplicated region. The duplication breakpoints are indicated by vertical dashed lines. Bottom: 4C-seq in individual F1_II-2 shows gain of ectopic interactions (indicated by blue rectangle) between *FGF8* and the *BTRC* TAD.

resulting domain structure in control cells corresponded to the TAD organization of the locus from published Hi-C data[27] (Figures 2A and S3), i.e., the *FGF8* TAD was insulated from the *BTRC* (and *LBX1* [MIM: 604255]) TAD by a strong boundary element and both chromatin domains show only marginal interactions with each other. In contrast, 4C analysis in the individual carrying the duplication (F1_II-2) revealed strong ectopic interaction between *FGF8* and the ~230 kb region within the neighboring *BTRC* TAD (Figure 2B). The CE66 enhancer element also showed a gain of interaction with the same region, confirming the results (Figure S3B). The reverse control experiment in which the *BTRC* promoter was used as a viewpoint also showed ectopic interaction with the *FGF8* TAD and the known *FGF8* regulatory landscape[20] (Figure S3B). Expression analysis in the proband's fibroblasts (F1_II-2) by real-time qPCR showed a 2.9-fold increased *FGF8* expression and 2.3-fold increased *BTRC* expression compared to seven healthy controls (Figure S4).

Thus, in the affected individual's fibroblasts, the duplication was associated with ectopic chromatin contacts and increased gene expression of the duplicated genes. However, it was still not possible to distinguish whether the phenotype was the result of gene dosage or position effects. Therefore, we decided to investigate the duplications in transgenic mice.

Position effects at the *Fgf8* locus cause shortening of the limbs

To investigate whether the 10q24.32 duplications identified in this study cause femoral hypoplasia, we generated a corresponding duplication in mice by using CRISPR-Cas9 genome editing. We used two sgRNAs to induce double-strand breaks at the orthologous sites in the mouse genome and screened mouse embryonic stem cells for duplications[30,31] (Figure 3). A clone carrying a heterozygous tandem duplication ($Fgf8^{+/dup}$) corresponding to the duplication of family 1 was used for tetraploid aggregation, and transgenic embryos were analyzed at E18.5 (Figure S5). The

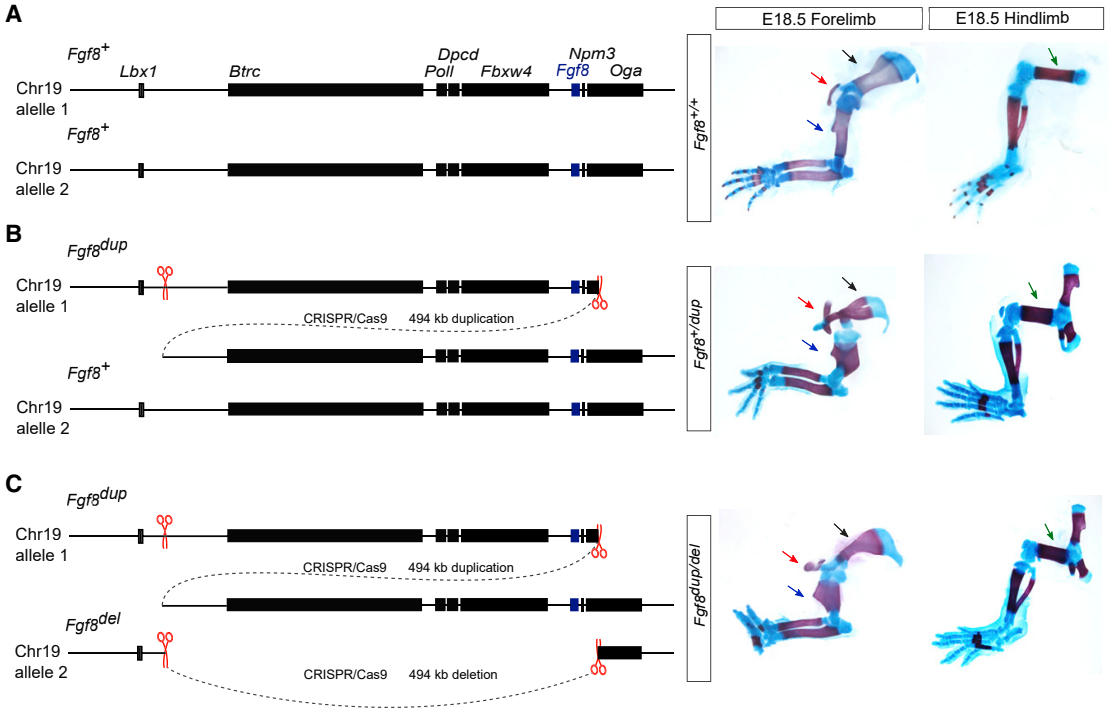


Figure 3. Position effects in *Fgf8*$^{+/dup}$ and *Fgf8*$^{dup/del}$ transgenic embryos cause shortening of the proximal limbs
(A–C) Left: the WT carries two copies of *Fgf8*, one on each allele (A). The *Fgf8*$^{+/dup}$ mutant carries three copies of *Fgf8*—two on the allele with a tandem duplication and one on the WT allele (B). In the *Fgf8*$^{dup/del}$ mutant, two copies of *Fgf8* are located on the duplication allele, while the other allele carries a corresponding deletion (C). The scissors mark the sgRNA binding/cutting sites. Right: skeletal preparations of the E18.5 WT (A) and mutant embryos (B and C). The *Fgf8*$^{+/dup}$ and *Fgf8*$^{dup/del}$ mutants show similar multiple skeletal abnormalities, including malformation of the scapulae (indicated by black arrows) and clavicles (red arrows); a severe shortening, widening, and malformation of the humeri (blue arrows); and flexed wrists as well as shortening and widening of the femoral bones (green arrows) (n = 2 skeletal stainings from two independent aggregations for each transgenic line). The cartilaginous and osseous tissues were dyed blue and red, respectively.

Fgf8$^{+/dup}$ transgenic mice showed skeletal abnormalities of both forelimbs and hindlimbs, including malformation of the scapulae and clavicles, severe shortening of the humeral and femoral bones, and abnormal wrists (Figure 3B). Although the mouse phenotype was severer than the human phenotype, the key feature of femur hypoplasia was remarkably similar. No feeding, behavioral, or other health-related problems were noticed. The skeletal malformation observed in the pups strongly limited their ability to move, and therefore, the mice were sacrificed at postnatal day 9.

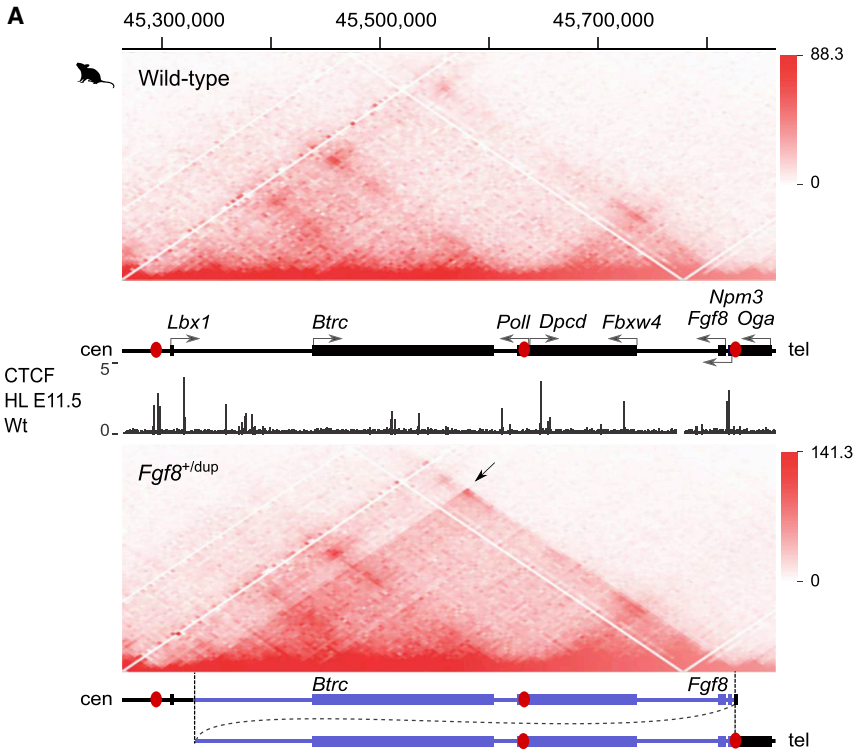
Next, we investigated whether the observed phenotype was the result of gene dosage or position effects resulting from the tandem duplication. Therefore, we generated a second transgenic mouse line (*Fgf8*$^{dup/del}$) carrying a tandem duplication encompassing *Fgf8* on one allele and a corresponding deletion on the other allele. Consequently, the gene dosage of *Fgf8* is comparable to the WT (Figure S5). The *Fgf8*$^{dup/del}$ mutant featured identical skeletal malformations of the forelimbs and hindlimbs, including the severe shortening of the humeral bones and shortening of the femoral bones, similar to the *Fgf8*$^{+/dup}$-mutant mice (Figure 3C).

Our data indicates that this phenotype is not simply a result of altered gene dosage but most likely due to position effects that change the regulatory landscape of the genomic region around *FGF8*.

Fgf8 duplications are associated with ectopic chromatin contacts

Finally, we aimed to study the effect of the duplication on the 3D genome architecture in the transgenic mice. Therefore, we performed cHi-C analysis in the WT and *Fgf8*$^{+/dup}$ limb buds at E11.5. Consistent with the human Hi-C[27] and our 4C-seq data, the WT mouse cHi-C map revealed the existence of two chromatin domains: the centromeric *Btrc* TAD and the telomeric *Fgf8* TAD. The cHi-C data in the *Fgf8*$^{+/dup}$ limb buds revealed a gain of chromatin interactions within the duplicated 10q24.32 region. In particular, ectopic interactions between the *Btrc* TAD and the *Fgf8* TAD were substantially elevated (Figure 4A).

To study the effect of ectopic chromatin interactions on *Fgf8* expression, we performed WISHs in the *Fgf8*$^{+/dup}$ -mutant and WT embryos. WISH at E11.5 and E12.5 revealed a gain of *Fgf8* expression in the AER, the central region of the autopod, the proximal part of the forelimbs, and in the AER of the hindlimbs in all tested embryos compared to WT mice (Figure 4B). A marked increase in *Fgf8* expression was observed in the hindlimb buds of the E12.5 mutant.



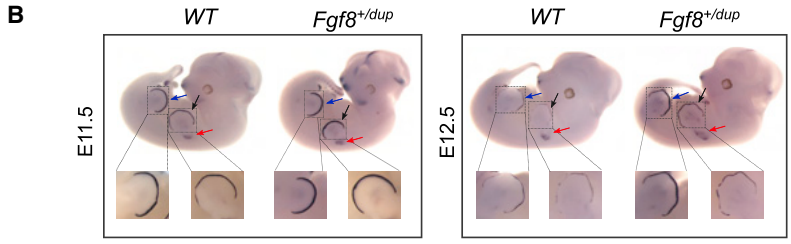


Figure 4. Genome architecture at the region of duplication and tissue-specific *Fgf8* expression in WT and *Fgf8*$^{+/dup}$ mice embryos

(A) cHi-C maps of WT and *Fgf8*$^{+/dup}$ E11.5 hindlimb buds. TAD boundaries at the examined locus are represented by red ovals. Below the upper cHi-C map: genes and their respective TSS; CTCF ChIP-seq tracks of the examined locus in the WT E11.5 hindlimb buds.[40,41] The black arrows in the *Fgf8*$^{+/dup}$ cHi-C map indicate ectopic interactions between the *Fgf8* TAD and the neighboring *Btrc* TAD. Breakpoints of the duplicated fragment are indicated by vertical dashed lines. Note that the neo-TAD[42] overlaps the duplicated region, shown below.

(B) WISH analysis revealed gain of the *Fgf8* gene expression in the AER and central region of the autopod in forelimb buds (black arrows) and the AER in hindlimb buds (blue arrows) in all tested embryos at the E11.5 and E12.5 compared to controls (n = 3 from two independent aggregations for each transgenic line and developmental stage). *Fgf8* overexpression was also noted in the proximal parts of the developing forelimbs (red arrows).

In conclusion, the 10q24.32 duplications of families 1 and 2 and the duplication in the *Fgf8*$^{+/dup}$ mice are associated with ectopic chromatin interactions that allow regulatory elements from the *BTRC*/*Btrc* TAD to ectopically interact with *FGF8*/*Fgf8*, leading to its altered expression. The limb malformations observed in these duplications were also conserved in the *Fgf8*$^{dup/del}$ mutant, where the gene dosage remained unchanged with respect to the WT.

Discussion

Here, we report on three individuals harboring duplications encompassing *FGF8* that result in a regulatory position effect associated with bilateral hypoplasia of the femoral and pelvic bones in humans and mice.

The duplications at the 10q24.32 locus associated with this condition overlap almost completely (except *FGF8*) with duplications implicated with split hand/foot malformation (SHFM3), a common limb defect.[3,11,43–45] Despite their similar sizes, the duplications result in very different phenotypes: while SHFM3 affects mainly hands and feet, the slightly larger duplications reported here cause femoral and pelvic hypoplasia with normal distal limbs. The only gene exclusive to the variants associated with femoral hypoplasia is *FGF8*, a key signaling factor for embryo patterning. *FGF8* is essential during the earliest stages of the limb bud development, promoting the formation and maintenance of the AER and limb elongation in the proximo-distal axis.[42,46] In the mouse, *Fgf8* expression has been shown to be precisely regulated by a set of noncoding regulatory elements located in the TAD between *Poll* and *Fgf8*.[20] The neighboring TAD harboring *LBX1* and *BTRC* is separated from *FGF8* by a boundary element. Our chromosome conformation data show that the tandem duplications reported here span across this boundary element, creating a different regulatory domain (neo-TAD[21]) containing *BTRC*, its regulators, and *FGF8*. This altered domain allows ectopic chromatin contacts between *FGF8* and enhancer elements located within *BTRC*, including its promoter, most likely resulting in altered *FGF8* expression during limb development. To rule out gene dosage effects of *FGF8*, we generated mouse models harboring the duplication and deleted *Fgf8* in the WT allele, obtaining another mouse line that only has two copies of *Fgf8*. Surprisingly, both mouse models showed limb malformations, indicating that it is the position effect that is the primary contributor to the phenotype and not gene dosage. WISH results showed a stronger *Fgf8* expression in the AER and the proximal parts of forelimb and hindlimb buds in all studied mutant embryos, confirming altered gene expression.

Our study has several limitations. First, although we could show that the duplications alter *Fgf8* expression at E11.5 and E12.5, it needs to be considered that the other genes in the duplicated fragments also contribute to the phenotype. Further in-depth studies of the mouse models at later development stages are necessary for investigation of the *FGF8* locus and its role in skeletal disease. Second, because of limited access to the affected individuals, we used only one proband-derived cell line to perform 4C-seq. Third, we cloned and sequenced the breakpoints of the duplications. However, more complex structural rearrangements at the locus might have been missed. Genome sequencing or long-read sequencing could help shed more light into the structure of the duplications. In summary, we show that duplications at the *FGF8* locus are associated with femoral hypoplasia and that the phenotype is most likely the result of position effects causing altered *FGF8* expression rather than gene dosage effects.

Data and code availability

The accession numbers for the aberrations identified during this study are available at DECIPHER: 427511, 427699, and 427701 (https://www.deciphergenomics.org/).

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.ajhg.2021.08.001.

Acknowledgments

We would like to thank both families for their collaboration and contribution to this project. We acknowledge the expert technical support from Asita Stiege and Norbert Brieske. We thank the Animal Facility of the Max Planck Institute for Molecular Genetics, Berlin, Germany, especially Karol Macura. M.S. and A.S.-S. were supported by the Polish National Science Centre (UMO-2016/23/N/NZ2/02362 to M.S. and UMO-2016/21/D/NZ5/00064 to A.S.-S.). A.S.-S. was also supported by the Polish National Science Centre scholarship for PhD students (UMO-2013/08/T/NZ2/00027). C.L. is supported by postdoctoral Beatriu de Pinós from Secretaria d'Universitats I Recerca del Departament d'Empresa i Coneixement de la Generalitat de Catalunya and by the Marie Sklodowska-Curie COFUND program from H2020 (2018-BP-00055). A.J. was supported by the Polish National Science Centre (UMO-2016/22/E/NZ5/00270) as well as the Polish National Centre for Research and Development (LIDER/008/431/L-4/12/NCBR/2013). M.S. is supported by grants from the Deutsche Forschungsgemeinschaft (DFG) (SP1532/3-1, SP1532/4-1, and SP1532/5-1), the Max Planck Foundation, and the Deutsches Zentrum für Luft- und Raumfahrt (DLR 01GM1925).

Declaration of interests

The authors declare no competing interests.

Received: February 26, 2021
Accepted: August 4, 2021
Published: August 24, 2021

Web resources

ClinVar, https://www.ncbi.nlm.nih.gov/clinvar/
dbVar, https://www.ncbi.nlm.nih.gov/dbvar/
DECIPHER, https://www.deciphergenomics.org/
DGV, http://dgv.tcag.ca/dgv/app/home
gnomAD Browser, https://gnomad.broadinstitute.org/
Hi-C project at Ren Lab, http://chromosome.sdsc.edu/mouse/hi-c/download.html
OMIM, https://www.omim.org/
Primer3, https://bioinfo.ut.ee/primer3-0.4.0/
UCSC Genome Browser, https://genome.ucsc.edu
VISTA Enhancer Browser, http://enhancer.lbl.gov
Zhang laboratory design tool, https://zlab.bio/guide-design-resources

References

1. Cooper, G.M., Coe, B.P., Girirajan, S., Rosenfeld, J.A., Vu, T.H., Baker, C., Williams, C., Stalker, H., Hamid, R., Hannig, V., et al. (2011). A copy number variation morbidity map of developmental delay. Nat. Genet. *43*, 838–846.
2. Coe, B.P., Witherspoon, K., Rosenfeld, J.A., van Bon, B.W., Vulto-van Silfhout, A.T., Bosco, P., Friend, K.L., Baker, C., Buono, S., Vissers, L.E., et al. (2014). Refining analyses of copy number variation identifies specific genes associated with developmental delay. Nat. Genet. *46*, 1063–1071.
3. Flöttmann, R., Kragesteen, B.K., Geuer, S., Socha, M., Allou, L., Sowińska-Seidler, A., Bosquillon de Jarcy, L., Wagner, J., Jamsheer, A., Oehl-Jaschkowitz, B., et al. (2018). Noncoding copy-number variations are associated with congenital limb malformation. Genet. Med. *20*, 599–607.
4. Riggs, E.R., Andersen, E.F., Cherry, A.M., Kantarci, S., Kearney, H., Patel, A., Raca, G., Ritter, D.I., South, S.T., Thorland, E.C., et al. (2020). Technical standards for the interpretation and reporting of constitutional copy-number variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen). Genet. Med. *22*, 245–257.
5. Yamasaki, M., Makino, T., Khor, S.S., Toyoda, H., Miyagawa, T., Liu, X., Kuwabara, H., Kano, Y., Shimada, T., Sugiyama, T., et al. (2020). Sensitivity to gene dosage and gene expression affects genes with copy number variants observed among neuropsychiatric diseases. BMC Med. Genomics *13*, 55.
6. Zhang, F., Gu, W., Hurles, M.E., and Lupski, J.R. (2009). Copy number variation in human health, disease, and evolution. Annu. Rev. Genomics Hum. Genet. *10*, 451–481.
7. Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376–380.
8. Nora, E.P., Dekker, J., and Heard, E. (2013). Segmental folding of chromosomes: a basis for structural and regulatory chromosomal neighborhoods? BioEssays *35*, 818–828.
9. Lupiáñez, D.G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J.M., Laxova, R., et al. (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell *161*, 1012–1025.
10. Kragesteen, B.K., Brancati, F., Digilio, M.C., Mundlos, S., and Spielmann, M. (2019). *H2AFY* promoter deletion causes

PITX1 endoactivation and Liebenberg syndrome. J. Med. Genet. *56*, 246–251.
11. Holder-Espinasse, M., Jamsheer, A., Escande, F., Andrieux, J., Petit, F., Sowinska-Seidler, A., Socha, M., Jakubiuk-Tomaszuk, A., Gerard, M., Mathieu-Dramard, M., et al. (2019). Duplication of 10q24 locus: broadening the clinical and radiological spectrum. Eur. J. Hum. Genet. *27*, 525–534.
12. Dogan, H., Can, H., and Otu, H.H. (2014). Whole genome sequence of a Turkish individual. PLoS ONE *9*, e85233.
13. Park, H., Kim, J.I., Ju, Y.S., Gokcumen, O., Mills, R.E., Kim, S., Lee, S., Suh, D., Hong, D., Kang, H.P., et al. (2010). Discovery of common Asian copy number variants using integrated high-resolution array CGH and massively parallel DNA sequencing. Nat. Genet. *42*, 400–405.
14. Thareja, G., John, S.E., Hebbar, P., Behbehani, K., Thanaraj, T.A., and Alsmadi, O. (2015). Sequence and analysis of a whole genome from Kuwaiti population subgroup of Persian ancestry. BMC Genomics *16*, 92.
15. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al.; Genome Aggregation Database Consortium (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. Nature *581*, 434–443.
16. MacDonald, J.R., Ziman, R., Yuen, R.K., Feuk, L., and Scherer, S.W. (2014). The Database of Genomic Variants: a curated collection of structural variation in the human genome. Nucleic Acids Res. *42*, D986–D992.
17. Sowińska-Seidler, A., Piwecka, M., Olech, E., Socha, M., Latos-Bieleńska, A., and Jamsheer, A. (2015). Hyperosmia, ectrodactyly, mild intellectual disability, and other defects in a male patient with an X-linked partial microduplication and overexpression of the *KAL1* gene. J. Appl. Genet. *56*, 177–184.
18. Yu, S., Kielt, M., Stegner, A.L., Kibiryeva, N., Bittel, D.C., and Cooley, L.D. (2009). Quantitative real-time polymerase chain reaction for the verification of genomic imbalances detected by microarray-based comparative genomic hybridization. Genet. Test. Mol. Biomarkers *13*, 751–760.
19. van de Werken, H.J., de Vree, P.J., Splinter, E., Holwerda, S.J., Klous, P., de Wit, E., and de Laat, W. (2012). 4C technology: protocols and data analysis. Methods Enzymol. *513*, 89–112.
20. Marinić, M., Aktas, T., Ruf, S., and Spitz, F. (2013). An integrated holo-enhancer unit defines tissue and gene specificity of the *Fgf8* regulatory landscape. Dev. Cell *24*, 530–542.
21. Franke, M., Ibrahim, D.M., Andrey, G., Schwarzer, W., Heinrich, V., Schöpflin, R., Kraft, K., Kempfer, R., Jerković, I., Chan, W.L., et al. (2016). Formation of new chromatin domains determines pathogenicity of genomic duplications. Nature *538*, 265–269.
22. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.
23. Wingett, S., Ewels, P., Furlan-Magaril, M., Nagano, T., Schoenfelder, S., Fraser, P., and Andrews, S. (2015). HiCUP: pipeline for mapping and processing Hi-C data. F1000Res. *4*, 1310.
24. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.
25. Durand, N.C., Shamim, M.S., Machol, I., Rao, S.S., Huntley, M.H., Lander, E.S., and Aiden, E.L. (2016). Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. Cell Syst. *3*, 95–98.
26. Knight, P.A., and Ruiz, D. (2012). A fast algorithm for matrix balancing. IMA J. Numer. Anal. *33*, 1029–1047.
27. Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., and Aiden, E.L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665–1680.
28. Servant, N., Varoquaux, N., Heard, E., Barillot, E., and Vert, J.P. (2018). Effective normalization for copy number variation in Hi-C data. BMC Bioinformatics *19*, 313.
29. Servant, N., Varoquaux, N., Lajoie, B.R., Viara, E., Chen, C.J., Vert, J.P., Heard, E., Dekker, J., and Barillot, E. (2015). HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome Biol. *16*, 259.
30. Kraft, K., Geuer, S., Will, A.J., Chan, W.L., Paliou, C., Borschiwer, M., Harabula, I., Wittler, L., Franke, M., Ibrahim, D.M., et al. (2015). Deletions, inversions, duplications: Engineering of structural variants using CRISPR/Cas in mice. Cell Rep. *10*, 833–839.
31. Andrey, G., and Spielmann, M. (2017). CRISPR/Cas9 Genome Editing in Embryonic Stem Cells. Methods Mol. Biol. *1468*, 221–234.
32. Artus, J., and Hadjantonakis, A.K. (2011). Generation of chimeras by aggregation of embryonic stem cells with diploid or tetraploid mouse embryos. Methods Mol. Biol. *693*, 37–56.
33. Kragesteen, B.K., Spielmann, M., Paliou, C., Heinrich, V., Schöpflin, R., Esposito, A., Annunziatella, C., Bianco, S., Chiariello, A.M., Jerković, I., et al. (2018). Dynamic 3D chromatin architecture contributes to enhancer specificity and limb morphogenesis. Nat. Genet. *50*, 1463–1473.
34. Minina, E., Schneider, S., Rosowski, M., Lauster, R., and Vortkamp, A. (2005). Expression of *Fgf* and *Tgfbeta* signaling related genes during embryonic endochondral ossification. Gene Expr. Patterns *6*, 102–109.
35. Mundlos, S. (2000). Skeletal morphogenesis. Methods Mol. Biol. *136*, 61–70.
36. Wolpert, L. (1969). Positional information and the spatial pattern of cellular differentiation. J. Theor. Biol. *25*, 1–47.
37. Niswander, L. (2003). Pattern formation: old models out on a limb. Nat. Rev. Genet. *4*, 133–143.
38. Zeller, R., López-Ríos, J., and Zuniga, A. (2009). Vertebrate limb bud development: moving towards integrative analysis of organogenesis. Nat. Rev. Genet. *10*, 845–858.
39. Ibrahim, D.M., and Mundlos, S. (2020). Three-dimensional chromatin in disease: What holds us together and what drives us apart? Curr. Opin. Cell Biol. *64*, 1–9.
40. ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.
41. Davis, C.A., Hitz, B.C., Sloan, C.A., Chan, E.T., Davidson, J.M., Gabdank, I., Hilton, J.A., Jain, K., Baymuradov, U.K., Narayanan, A.K., et al. (2018). The Encyclopedia of DNA elements (ENCODE): data portal update. Nucleic Acids Res. *46* (D1), D794–D801.
42. Bénazet, J.D., and Zeller, R. (2009). Vertebrate limb development: moving from classical morphogen gradients to an integrated 4-dimensional patterning system. Cold Spring Harb. Perspect. Biol. *1*, a001339.
43. Klopocki, E., Lohan, S., Doelken, S.C., Stricker, S., Ockeloen, C.W., Soares Thiele de Aguiar, R., Lezirovitz, K., Mingroni Netto, R.C., Jamsheer, A., Shah, H., et al. (2012). Duplications of *BHLHA9* are associated with ectrodactyly and tibia hemimelia inherited in non-Mendelian fashion. J. Med. Genet. *49*, 119–125.

44. Sowińska-Seidler, A., Socha, M., and Jamsheer, A. (2014). Split-hand/foot malformation - molecular cause and implications in genetic counseling. J. Appl. Genet. *55*, 105–115.
45. Tayebi, N., Jamsheer, A., Flöttmann, R., Sowinska-Seidler, A., Doelken, S.C., Oehl-Jaschkowitz, B., Hülsemann, W., Habenicht, R., Klopocki, E., Mundlos, S., and Spielmann, M. (2014). Deletions of exons with regulatory activity at the *DYNC1I1 locus* are associated with split-hand/split-foot malformation: array CGH screening of 134 unrelated families. Orphanet J. Rare Dis. *9*, 108.
46. Lewandoski, M., Sun, X., and Martin, G.R. (2000). *Fgf8* signalling from the AER is essential for normal limb development. Nat. Genet. *26*, 460–463.