LETTER

Formation of new chromatin domains determines pathogenicity of genomic duplications

Martin Franke[1,2]*, Daniel M. Ibrahim[1,2,3]*, Guillaume Andrey[1], Wibke Schwarzer[4], Verena Heinrich[2,5], Robert Schöpflin[5], Katerina Kraft[1,2], Rieke Kempfer[1], Ivana Jerković[1,2], Wing–Lee Chan[2], Malte Spielmann[1,2], Bernd Timmermann[6], Lars Wittler[7], Ingo Kurth[8,9], Paola Cambiaso[10], Orsetta Zuffardi[11], Gunnar Houge[12], Lindsay Lambie[13], Francesco Brancati[14,15], Ana Pombo[3,16], Martin Vingron[5], Francois Spitz[4] & Stefan Mundlos[1,2,3,17]

Chromosome conformation capture methods have identified subchromosomal structures of higher-order chromatin interactions called topologically associated domains (TADs) that are separated from each other by boundary regions[1,2]. By subdividing the genome into discrete regulatory units, TADs restrict the contacts that enhancers establish with their target genes[3–5]. However, the mechanisms that underlie partitioning of the genome into TADs remain poorly understood. Here we show by chromosome conformation capture (capture Hi-C and 4C-seq methods) that genomic duplications in patient cells and genetically modified mice can result in the formation of new chromatin domains (neo-TADs) and that this process determines their molecular pathology. Duplications of non-coding DNA within the mouse *Sox9* TAD (intra-TAD) that cause female to male sex reversal in humans[6], showed increased contact of the duplicated regions within the TAD, but no change in the overall TAD structure. In contrast, overlapping duplications that extended over the next boundary into the neighbouring TAD (inter-TAD), resulted in the formation of a new chromatin domain (neo-TAD) that was isolated from the rest of the genome. As a consequence of this insulation, inter-TAD duplications had no phenotypic effect. However, incorporation of the next flanking gene, *Kcnj2*, in the neo-TAD resulted in ectopic contacts of *Kcnj2* with the duplicated part of the *Sox9* regulatory region, consecutive misexpression of *Kcnj2*, and a limb malformation phenotype. Our findings provide evidence that TADs are genomic regulatory units with a high degree of internal stability that can be sculptured by structural genomic variations. This process is important for the interpretation of copy number variations, as these variations are routinely detected in diagnostic tests for genetic disease and cancer. This finding also has relevance in an evolutionary setting because copy-number differences are thought to have a crucial role in the evolution of genome complexity.

SOX9 is a developmental transcription factor with important functions in chondrocyte differentiation and male sex determination[7]. The *SOX9* locus is a genomic region that has been linked to various human diseases with a broad range of phenotypes[6,8,9]. We investigated how genomic duplications in this region affect higher-order chromatin organization and, in particular, the formation of TADs. Hi-C, a technology to quantify chromatin contacts genome-wide, shows a compartmentalization of the human locus in two major TADs, one containing *SOX9* (referred to hereafter as *SOX9* TAD), the other containing the two potassium channels *KCNJ2* and *KCNJ16* (referred to hereafter as *KCNJ* TAD) (Fig. 1a)[2]. The large gene desert corresponding to the *SOX9* TAD has been shown to contain multiple regulatory elements and human-disease-related sites (Fig. 1a)[9]. Duplications including a region 0.5 megabases (Mb) upstream of *SOX9* (referred to as the RevSex region) lead to female-to-male sex reversal (Fig. 1a), indicating that the RevSex region contains a critical regulator of *SOX9* expression in the developing gonads[6,10]. Surprisingly, duplications that include the RevSex region but extend further upstream towards the neighbouring genes *KCNJ2* and *KCNJ16* have no effect on sexual development. Instead, they result in Cooks syndrome, a congenital limb malformation characterized by aplasia of nails and short digits[8]. We identified a family (mother and daughter) with a third type of duplication that includes the RevSex region and the entire gene desert upstream of *SOX9*, but not the *KCNJ2* and *KCNJ16* genes (Fig. 1a). In spite of complete overlap with the reported sex reversal duplications and large parts of the Cooks syndrome duplication, carriers of this variant are phenotypically normal. According to the Hi-C profile, all sex reversal associated duplications are located within the *SOX9* TAD, whereas the duplication with no phenotype and also the Cooks syndrome duplications extend into the neighbouring *KCNJ* TAD thereby spanning two TADs and their boundary. We refer two these two types of duplications as intra-TAD and inter-TAD, respectively.

To investigate the difference between the intra-TAD (sex reversal) and the inter-TAD (no phenotype) duplications, we performed 4C-seq of patient fibroblasts and compared them to controls. In control cells, the *SOX9* and the *KCNJ2* viewpoint showed an interaction profile restricted to their Hi-C predicted TADs (Fig. 1b). 4C-seq from a patient with sex reversal and a 150 kilobase (kb) intra-TAD duplication showed a slight increase with the duplicated part from *SOX9*, but an overall unchanged profile from both viewpoints (Fig. 1c). Similarly, no change in the configuration of the interaction profiles was observed with the larger inter-TAD (no phenotype) duplication (Fig. 1d). To identify the region specifically contacted by the duplicated part, we used the unique sequence created by the duplication breakpoints to perform allele-specific 4C-seq. In the intra-TAD duplication, we observed high interaction with the entire *SOX9* TAD (purple track, Fig. 1c). Similar results were obtained with fibroblasts from a second individual with a 470 kb intra-TAD duplication (Extended Data Fig. 1). By contrast, the interaction profile generated from the inter-TAD duplication breakpoint was restricted to the duplicated region, without contacting
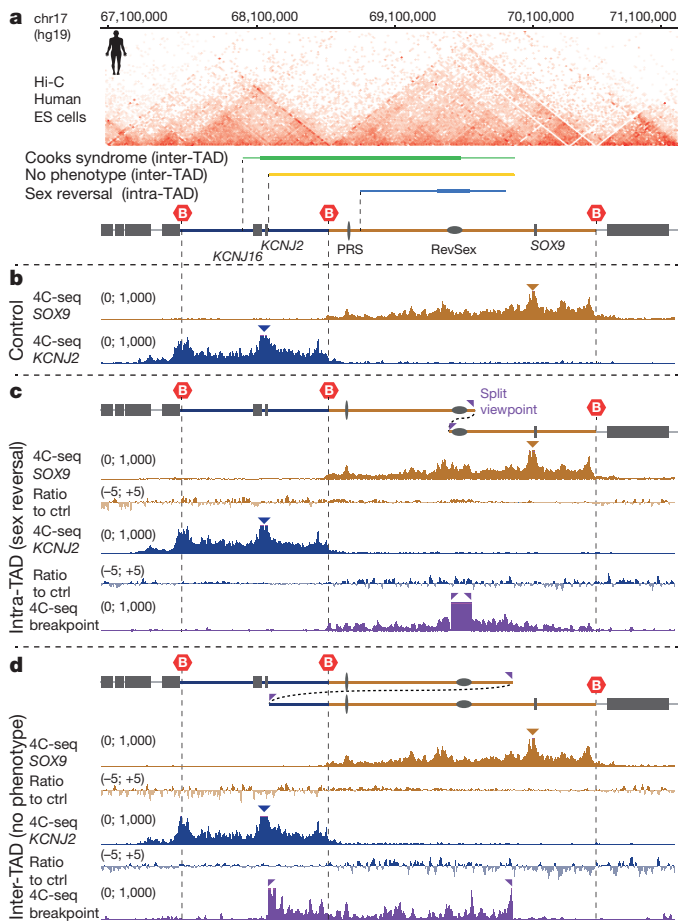
Figure 1 | 4C-seq from human fibroblasts with intra-TAD and inter-TAD duplications. a, *SOX9* locus with Hi-C interaction profile (H1 embryonic stem cells[2]). Schematic shows *KCNJ* TAD (blue), *SOX9* TAD (brown) and TAD boundaries (red hexagons). Ovals indicate regions associated with sex reversal (RevSex) or Pierre Robin syndrome (PRS), human duplications with corresponding phenotypes are indicated above (line thickness indicates maximum/minimum size). **b**, 4C-seq from control skin fibroblasts using *SOX9* and *KCNJ2* as viewpoints (triangles). **c**, 4C-seq from fibroblasts of an individual with sex reversal and intra-TAD duplication shows no change in overall interaction profile (below, log$_2$ ratio to control). 4C-seq from duplication breakpoint (below, in purple) shows increased interaction with the duplicated part and the rest of the *SOX9* TAD. **d**, 4C-seq from fibroblasts of individual with no phenotype and inter-TAD duplication. Note unchanged interaction profiles from *SOX9* and *KCNJ2* viewpoints. 4C-seq from duplication breakpoint shows interactions restricted to the duplicated region. All reads mapped to a wild-type genome (resulting in split viewpoint for duplication breakpoints). See Methods for sample collection.

the *SOX9*, *KCNJ2* or *KCNJ16* genes (purple track, Fig. 1d). To further investigate allele-specific interactions with the different viewpoints, we performed whole-genome sequencing of the patient sample with the inter-TAD duplication and determined if certain single nucleotide polymorphisms (SNPs) were contacted by one viewpoint and not the other. We identified several SNPs that were only contacted by the duplication specific viewpoint but not by the *SOX9* or *KCNJ2* viewpoints, suggesting that the duplicated region contacted primarily itself, thereby forming a separate domain (Extended Data Fig. 2 and Methods).

To investigate the effect of these duplications on TAD structure in more detail, we re-engineered the human duplications in mice and performed capture Hi-C (cHi-C) and 4C-seq of the extended *Sox9* locus from wild-type and mutant limb buds at embryonic day 12.5 (E12.5) (Fig. 2). We used CRISPR/Cas9 genome editing to create mice with a 400 kb intra-TAD duplication (Dup-S), equivalent to a

published patient with sex reversal[6]. Heterozygous Dup-S mice were phenotypically normal. This is probably explained by the finding that a known mouse gonadal enhancer, TESCO, resides outside of the duplication[11], and the human sex-reversal-associated regions could so far not be linked to *Sox9* regulation in mouse[10]. The larger 1.6 Mb inter-TAD duplication (Dup-L), equivalent to the individuals with no phenotype, was generated using the Cre/*loxP* system and *trans*-allelic recombination.

In agreement with the human data, our data show that the mouse locus is subdivided into a *Sox9* TAD, containing *Sox9* as the only gene, and a *Kcnj* TAD with *Kcnj2* and *Kcnj16*, separated by a boundary located in the gene desert (Fig. 2a). cHi-C from mice with the intra-TAD duplication (Dup-S) showed no change in the overall TAD structure (Fig. 2b). Subtraction of the wild-type profile from the Dup-S profile demonstrated increased interaction of the duplicated part with itself, but also with the rest of the *Sox9* TAD including *Sox9*. By contrast, cHi-C from the inter-TAD duplication (Dup-L) showed a new interaction domain covering the duplicated parts of the *Kcnj* and *Sox9* TADs (Fig. 2c). The subtraction profile showed a strong increase in interaction in comparison to wild type, but only within the duplicated region (thus excluding *Sox9*, *Kcnj2* and *Kcnj16*) and no changes in the adjacent TADs. This suggested that the new interaction domain was isolated from the neighbouring *Sox9* and *Kcnj* TADs. 4C-seq from outside of the duplication (*Sox9* and *Kcnj2*) confirmed the isolation of this domain by showing wild-type interaction profiles and no ectopic contacts (Extended Data Fig. 3). 4C-seq from the *lacZ* reporter located at the duplication breakpoint corroborated our other findings, showing interaction only with the duplicated region (Extended Data Fig. 3b). In agreement with the observed isolation of the duplication domain, Dup-L mice were phenotypically normal and fertile.

Thus, the two types of duplications differ fundamentally in their effect on chromatin organization in spite of the fact that they overlap to a large degree. The intra-TAD duplication (resulting in sex reversal) has no effect on the overall TAD structure, whereas the larger inter-TAD duplication (resulting in no phenotype in humans and mice) resulted in the generation of a new chromatin domain that we refer to as a 'neo-TAD'. The insulation of the neo-TAD prevents interaction with the neighbouring *Sox9*, *Kcnj2* or *Kcnj16* genes and explains the lack of pathology in humans and in mice that carry this type of duplication.

The presence of the *lacZ* reporter in the middle of the Dup-L duplication also allowed us to investigate the regulatory potential of the neo-TAD by analysing the *lacZ* expression pattern in Dup-L mutants and compare it to *lacZ* reporters inserted in the *Kcnj* and *Sox9* TADs (Sleeping Beauty (SB) transposons carrying *lacZ*; SB-Kcnj and SB-Sox9 alleles; Fig. 2d and Extended Data Fig. 4). SB-Sox9 gave a strong signal essentially recapitulating the endogenous *Sox9* expression, whereas SB-Kcnj showed a restricted signal in the developing jaw and nose. The LacZ staining obtained in Dup-L was very similar to that of SB-Sox9, suggesting that the regulatory sequence contained in the neo-TAD was functional and sufficient to recapitulate most of the *Sox9* expression domains, but not that of *Kcnj2*. This was further supported by the analysis of a deletion (Del-L allele, Extended Data Fig. 4) in which the corresponding region had been removed by *trans*-allelic recombination. LacZ staining in Del-L mice showed a complete loss of the *Sox9*, but a preservation of the *Kcnj2* expression domains (Fig. 2d). Thus, the regulatory region in the neo-TAD was able to drive expression in a tissue-specific *Sox9* pattern. As predicted by reporter assays and histone modification data, the region contains a large number of regulatory sites (Extended Data Fig. 5a) that, as shown by the *lacZ* reporter in Del-L mice, control the major part of *Sox9* expression. Interestingly, comparison of the interaction profiles of the *Kcnj2* and *Sox9* viewpoints in wild type with the profile from the *lacZ* viewpoint in Dup-L revealed a similar pattern and peak distribution of 4C-seq profiles, suggesting that the newly formed domain folded in a structured manner retaining its wild-type interaction characteristics (Extended Data Fig. 5).
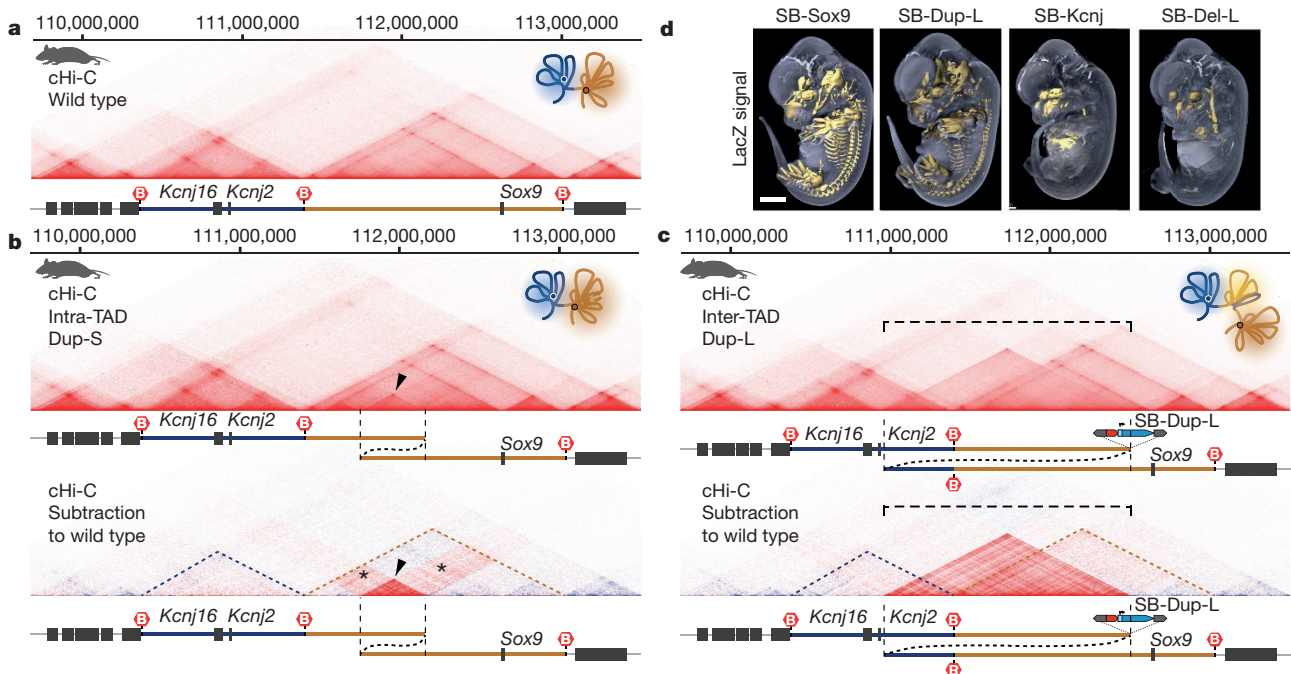
Figure 2 | Capture Hi-C from intra-TAD and inter-TAD duplications in mice. a, Capture Hi-C (cHi-C) of the *Sox9* locus from wild-type E12.5 limb buds showing separation in two TADs. Schematic on the top right depicts chromatin folding in *Kcnj* (blue) and *Sox9* (brown) TAD. **b**, cHi-C of intra-TAD duplication in Dup-S mice and subtraction map relative to wild type (below) shows an increase in interaction frequency of the duplicated region with itself (arrow) and the *Sox9* TAD (asterisks) **c**, cHi-C of inter-TAD duplication in Dup-L mice and subtraction map (below)

shows formation of a neo-TAD (brackets). *Kcnj* and *Sox9* TAD indicated by dashed lines. All coordinates chr11 (mm9). See Methods for sample collection. **d**, Optical projection tomography (OPT) of LacZ staining with insertion of *lacZ* reporter in *Sox9* TAD (SB-Sox9), at the duplication breakpoint (SB-Dup-L), in *Kcnj* TAD (SB-Kcnj), and in a deletion (SB-Del-L) corresponding to Dup-L (see Extended Data Fig. 4). Note recapitulation of *Sox9* pattern in Dup-L and loss in Del-L. Scale bar, 1 mm.

We hypothesized that an extension of the duplication towards the next flanking genes *Kcnj2* and *Kcnj16*, as reported in human Cooks syndrome[8], would result in incorporation of these genes in the neo-TAD. We generated this inter-TAD duplication (Dup-C) by *trans*-allelic recombination (Fig. 3). cHi-C of E12.5 Dup-C limb buds showed, like in the Dup-L mutant, a new chromatin domain corresponding to the

duplicated region (Fig. 3a). The subtraction showed a strong increase in interaction between the duplicated parts of the *Sox9* and *Kcnj* TADs including the *Kcnj2* and *Kcnj16* genes, but not with *Sox9*. Ectopic interaction of *Kcnj2* with parts of the *Sox9* TAD was confirmed by 4C-seq from mouse limb buds and by analysing fibroblasts from a Cooks syndrome patient (Extended Data Fig. 6). To investigate the effects
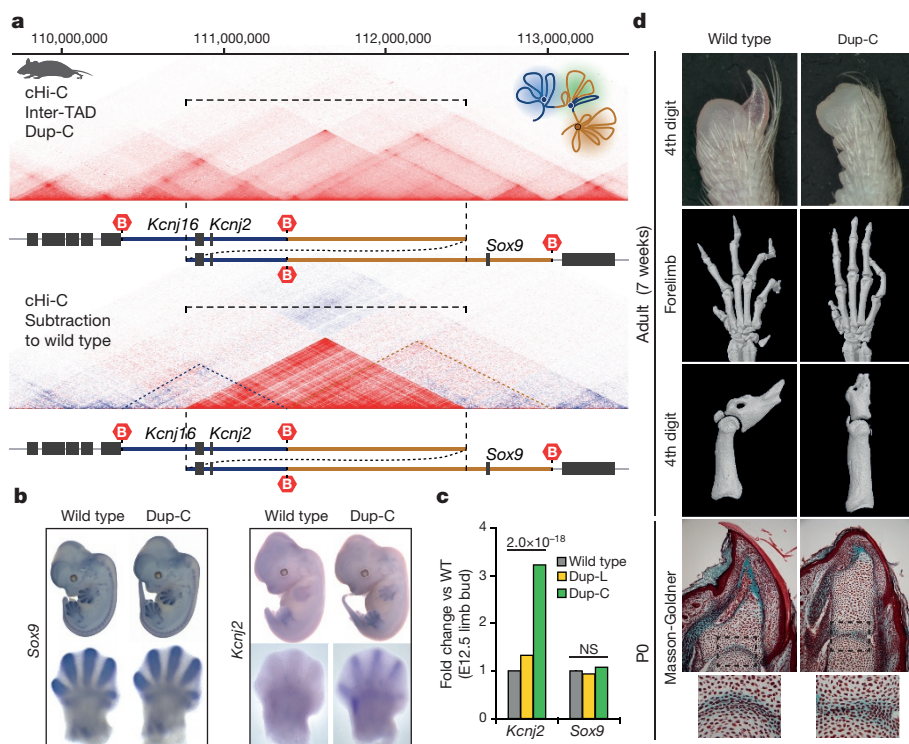

Figure 3 | Capture Hi-C, gene expression and phenotype in inter-TAD Cooks syndrome duplication. a, cHi-C and subtraction map relative to wild type (below) of inter-TAD duplication in Dup-C E12.5 limb buds showing formation of new interaction domain (neo-TAD, bracket) including *Kcnj2* and *Kcnj16*. *Kcnj* and *Sox9* TAD indicated by dashed lines. Coordinates chr11 (mm9). **b**, Whole mount *in situ* hybridization for *Sox9* and *Kcnj2* at E12.5. Note *Sox9*-like expression of *Kcnj2* in Dup-C digits, *n* = 5. **c**, RNA-seq from E12.5 limb buds shows upregulation of *Kcnj2* in Dup-C mice only (Benjamini–Hochberg adjusted *P* value, *n* = 2, see Methods). **d**, Phenotypic comparison of wild type and the Dup-C mutant. Note (top to bottom) absence of claw, abnormal shape and positioning of phalanges (micro-CT), hypoplasia of nail bed and abnormal distal phalanx (Masson–Goldner). Enlargement (dashed boxes) showing joint fusion in Dup-C mutants. Phenotype is fully penetrant (*n* = 30).

of these ectopic contacts on gene expression, we performed whole-mount *in situ* hybridization and found *Kcnj2* to be expressed in the digit anlagen, with a pattern similar to *Sox9* (Fig. 3b). RNA sequencing (RNA-seq) expression analysis of Dup-C limb buds at E12.5 and E17.5 confirmed the upregulation of *Kcnj2*, whereas other genes around the locus stayed unchanged, in particular *Sox9*, but also *Kcnj16* (Fig. 3c and Extended Data Fig. 7). Thus, the inclusion of *Kcnj2* in the neo-TAD resulted in its activation by regulatory elements that originally belonged to the *Sox9* TAD. In contrast to *Kcnj2*, *Kcnj16* was not responsive to ectopic activation. 4C-seq profiles in human fibroblasts with *KCNJ2* and *KCNJ16* as viewpoints showed slight differences in local intensity but generally the same size of ectopic interaction (Extended Data Fig. 6), indicating that a certain permissiveness or specificity must be present for promoter activation in this setting.

These new contacts and the associated misexpression of *Kcnj2* were accompanied by major phenotypic changes. Heterozygous Dup-C mice showed a limb malformation phenotype at birth closely resembling Cooks syndrome[8], highlighted by the absence or severe hypoplasia of all claws or nails (Fig. 3d). Micro-computed tomography (μCT) and histology demonstrated an abnormal shape and size of the distal phalanges, which were fixed in a straight position due to a malformed terminal phalanx and a partly fused interphalangeal joint. To rule out that the Cooks phenotype was merely produced by increased gene dosage of *Kcnj2,* we created intra-TAD duplications that included *Kcnj2* and *Kcnj16* (Dup-K_1 and Dup-K_2) (Extended Data Fig. 4). These mice were normal and had no digit phenotype. In addition, we created a second Cooks allele (Dup-C_2) that included *Kcnj2* in the duplication, but not *Kcnj16*. These mice showed the typical Cooks syndrome phenotype. Taken together, our data suggest misexpression of *Kcnj2* as the cause for Cooks syndrome. The *Drosophila melanogaster KCNJ2* homologue *Irk2*, an inwardly rectifying K$^+$ channel, has been shown to have additional functions in development via the *dpp* (bone morphogenic protein, BMP) pathway[12]. Mutations in components of the BMP pathway are a major cause of abnormalities in digits and joints[13], providing a possible connection to the observed pathology in Cooks syndrome.

The Dup-C duplication resulted in the inclusion of *Kcnj2* in the neo-TAD and its positioning next to *Sox9* regulatory domain without an intervening boundary. We hypothesized that a similar effect might be achieved by removing the boundary between the *Sox9* and *Kcnj* TADs. To test this hypothesis, we deleted using CRISPR/Cas9 the predicted boundary region (Bor), a small (18 kb) region containing conserved CTCF binding sites. Mice with this deletion (ΔBor) had no apparent phenotype. cHi-C analyses of homozygous ΔBor limb buds showed an increase of interaction but no fusion of the *Kcnj* and *Sox9* TADs (Extended Data Fig. 8a). This ectopic interaction resulted in the upregulation, but no site-specific misexpression of *Kcnj2* (Extended Data Fig. 8b). Thus, deletion of the boundary resulted in ectopic contacts, as previously reported by others *in vitro*[1,14], but the overall TAD structure remained unchanged. Similar results were obtained when deleting the boundaries in the Dup-L duplication (Extended Data Fig. 8). We observed increased interaction between the TADs, but overall the neo-TAD remained stable, indicating that the neo-TAD behaved like a 'regular' TAD. The importance of boundaries in restricting chromatin interactions was highlighted previously by deletions at the *Epha4* locus[3]. However, in these experiments large portions of the adjacent TADs were deleted together with the boundaries, thereby disrupting the overall TAD structure. Our present data indicate that other factors, such as additional CTCF sites and loops within TADs[14,15] contribute to TAD stability. The deletion of a boundary alone has therefore no major consequences, whereas larger deletions result in a re-organization of the locus enabling new contacts.

Duplications are generally thought to confer their phenotypic effect through an increase in gene dosage, but often the observed phenotype cannot be explained by alterations in gene dosage. Our data show how duplications can have different effects on higher-order chromatin structure, depending on their size and position. Duplications that are
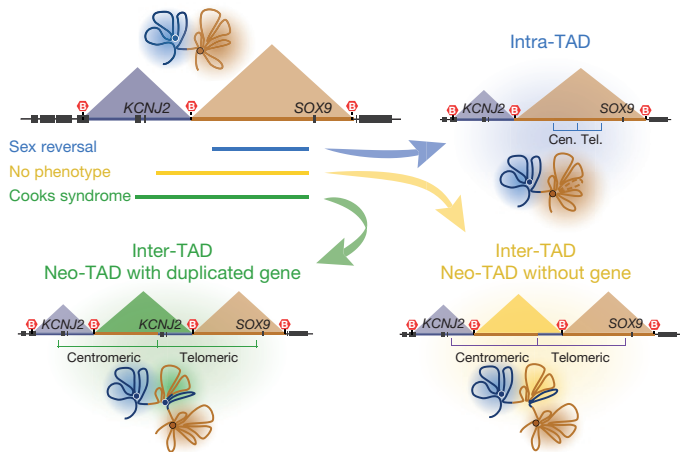

Figure 4 | Duplication-induced effects on chromatin organization and phenotype. Intra-TAD duplications (blue) do not change the overall TAD conformation but can result in abnormal gene regulation (sex reversal). The centromeric and telomeric parts of the tandem duplications are indicated. Inter-TAD duplications crossing TAD-boundaries (green and yellow) result in the formation of new chromatin domains (neo-TADs). Insulation from their neighbours results in neutralization of regulatory effects and normal phenotype. Incorporation of genes in the neo-TAD provides the duplicated gene with a novel regulatory landscape and can result in gene misexpression (Cooks syndrome).

confined to a TAD (intra-TAD) have no major effect on TAD structure but can result in increased interaction of duplicated regulatory elements with their target gene. In contrast, duplications that cover parts of two TADs and their boundary (inter-TAD) result in the formation of a new TAD that is insulated from its neighbours. We propose that these newly created domains should be called 'neo-TADs'. Genes that become incorporated in a neo-TAD can be activated by its regulatory elements, thereby eliciting pathogenic effects. Our findings also demonstrate that the genomic effects of structural variations cannot solely be explained by the rewiring of enhancer–promoter contacts[3]. Rather, our data show that TADs are robust and stable genomic units that can be rearranged and recombined to create new regulatory regions of the genome. The integrity of these units in relation to neighbouring TADs and genes determines their gene regulatory and thus pathogenic effect. Figure 4 shows a schematic of the proposed disease mechanism associated with TAD changes.

The concept presented here provides a framework to predict the phenotypic outcome of genomic variations that can be directly applied for the interpretation of copy number variations (CNVs) detected in diagnostic screens, routinely performed in patients with congenital malformations and/or intellectual disability[16], or for structural variations found in cancer[17]. Furthermore, the effects of genomic rearrangements described here are probably important for evolutionary mechanisms, as duplications are thought to be a major driver in evolution[18,19]. The process of gene neofunctionalization is thought to work through gene duplications and subsequent adaptation of one of the gene copies. Our data suggest a further mechanism in which the isolation of a newly formed TAD can result in a phenotypic change in the organism that is then directly subject to selective pressure, without affecting the parent copy of the gene. With variable shifting of TADs and recombination of regulatory activity with new target genes, an entire toolbox of possibilities for new gene functions can be acquired.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

1. Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485,** 381–385 (2012).

2. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485,** 376–380 (2012).
3. Lupiáñez, D. G. *et al.* Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161,** 1012–1025 (2015).
4. Symmons, O. *et al.* Functional and topological characteristics of mammalian regulatory domains. *Genome Res.* **24,** 390–400 (2014).
5. 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).
6. Benko, S. *et al.* Disruption of a long distance regulatory region upstream of *SOX9* in isolated disorders of sex development. *J. Med. Genet.* **48,** 825–830 (2011).
7. Wagner, T. *et al.* Autosomal sex reversal and campomelic dysplasia are caused by mutations in and around the SRY-related gene *SOX9*. *Cell* **79,** 1111–1120 (1994).
8. Kurth, I. *et al.* Duplications of noncoding elements 5′ of *SOX9* are associated with brachydactyly-anonychia. *Nature Genet.* **41,** 862–863 (2009).
9. Gordon, C. T. *et al.* Long-range regulation at the *SOX9* locus in development and disease. *J. Med. Genet.* **46,** 649–656 (2009).
10. Kim, G.-J. *et al.* Copy number variation of two separate regulatory regions upstream of *SOX9* causes isolated 46,XY or 46,XX disorder of sex development. *J. Med. Genet.* **52,** 240–247 (2015).
11. Sekido, R. & Lovell-Badge, R. Sex determination involves synergistic action of SRY and SF1 on a specific *Sox9* enhancer. *Nature* **453,** 930–934 (2008).
12. Dahal, G. R. *et al.* An inwardly rectifying K$^+$ channel is required for patterning. *Development* **139,** 3653–3664 (2012).
13. Mundlos, S. The brachydactylies: a molecular disease family. *Clin. Genet.* **76,** 123–136 (2009).
14. 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).
15. Guo, Y. *et al.* CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function. *Cell* **162,** 900–910 (2015).
16. Lupski, J. R. Genomic rearrangements and sporadic disease. *Nature Genet.* **39** (Suppl), S43–S47 (2007).
17. Stephens, P. J. *et al.* Complex landscapes of somatic rearrangement in human breast cancer genomes. *Nature* **462,** 1005–1010 (2009).
18. Ohno, S. *Evolution by Gene Duplication* (Springer, 1970).
19. Katju, V. & Bergthorsson, U. Copy-number changes in evolution: rates, fitness effects and adaptive significance. *Front. Genet.* **4,** 273 (2013).

METHODS

No statistical methods were used to predetermine sample size. There was no randomization of experiments, and investigators were not blinded during experiments and outcome assessment.

ES cell targeting and transgenic mouse strains. Embryonic stem (ES) cell culture was performed as described previously[20]. ES and feeder cells were tested for mycoplasma contamination using Mycoalert detection kit (Lonza, catalogue number LT07-118) and Mycoalert assay control set (Lonza, catalogue number LT07-518). A single *LoxP* site upstream of *Kcnj16* (mm9 chr11:110,772,110) was targeted in G4-ES cells (129/Sv × C57BL/6 F1 hybrid) using CRISPR/Cas9 and a single-stranded donor oligonucleotide (ssODN). ES cells were cultured on MEF feeder cells under standard conditions and then cotransfected (FuGene HD, Promega) with 8 μg of pX459 vector (Addgene) carrying the single guide RNA (cenKcnj) and 500 pM of a ssODN (40 bp homologous sequence flanking a *LoxP* site), sequences given in Supplementary Table 1. To delete the 18 kb TAD boundary region (Bor) in G4 ES cells and ES cells from Dup-L homozygous mice, two sgRNAs (cenΔBor, telΔBor) were cloned in pX459 vectors and cotransfected. The intra-TAD duplication was generated in G4 ES cells using two sgRNAs (cenDupS, telDupS). Embryos and live animals from ES cells were generated by diploid or tetraploid complementation[21]. Genotyping was performed by PCR analysis.

The SB-Kcnj and SB-Sox9 alleles contain an insertion of Sleeping Beauty (SB) transgene[22]. This transgene harbours a single *LoxP* site and a *lacZ* reporter gene, flanked by transposons. Both alleles were targeted using standard protocols for homologous recombination in E14 ES cells[23]. Primer sequences for amplifying homology sequences are provided in Supplementary Table 1. Positive ES cell clones were injected into donor blastocysts to generate chimaeras. The SB-Kcnj allele was further used for remobilization of the SB transgene, following the protocol in ref. 22, to generate new SB insertion sites at the locus. Duplications and corresponding deletions were generated as described previously[22]. All generated duplications/deletions and donor mouse strains are listed in Extended Data Fig. 4.

Mouse strains were maintained by crossing them with C57BL6/J mice. All animal procedures were conducted as approved by the local authorities (LAGeSo Berlin) under the license numbers G0368/08 and G0247/13.

ES cell generation of Dup-L mice. ES cells from Dup-L homozygous blastocysts were established using N2B27 medium supplemented with FGF/Erk and Gsk3 pathway inhibitors (2i) and LIF according to ref. 24. An established ES cell line was used for subsequent boundary deletion experiments using CRISPR/Cas9.

4C-seq. 4C-seq libraries were generated from microdissected mouse tissues or human fibroblasts as described previously[25]. The starting material for all 4C-seq libraries was 5×10^6 to 1×10^7 cells, which corresponds to limb buds from a pool of eight E12.5 embryos. All 4C-seq experiments were carried out in heterozygous animals. 4-bp cutters were used as primary and secondary restriction enzymes (Supplementary Table 2). For each viewpoint, a total of 1 to 1.6 μg DNA was amplified by PCR (Supplementary Table 2). All samples were sequenced with Illumina Hi-Seq technology according to standard protocols. 4C-seq experiments from all viewpoints were carried out in biological replicates in wild type and Dup-L mutants and tested for reproducibility (Pearson R for WT:*Sox9* = 0.87, *Kcnj2* = 0.90; Dup-L:*Sox9* = 0.94, *Kcnj2* = 0.94, *lacZ* = 0.89). A representative result is shown in the figures. Experiments from patient samples and Dup-C mutants were performed as singletons.

Bioinformatics. For 4C-seq data analysis, reads were pre-processed, mapped to a corresponding reference (GRCh37/hg19 or NCBI37/mm9) using BWA[26] and coverage normalized as reported previously[3]. The viewpoint and adjacent fragments 1.5 kb upstream and downstream were removed and a window of 10 fragments was chosen to normalize the data per million mapped reads (RPM). To compare interaction profiles of different samples, we obtained the $\log_2$-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.

SNP analysis of 4C-seq data. To indicate selective interactions in a patient with an inter-TAD duplication (no phenotype), raw reads of 4C-seq experiments with viewpoints in *SOX9*, *KCNJ2* and in the duplication breakpoint were mapped using BWA[26] without pre-processing and variants were called for all samples together using GATK (v3.4-46)[27] for the duplicated region (chr17:68195430–69981335, hg19). Bi-allelic SNP positions with a minimum coverage of 10 reads were selected. These were called homozygous for one allele contacted by the breakpoint viewpoint and homozygous for the other allele contacted by either the *SOX9* or *KCNJ2* viewpoint. Additionally, whole-genome sequencing was performed and alignment and variant calling was done in the same way. Allele frequencies of the WGS experiment were then compared to the allele frequencies of the 4C-seq experiments at the selected SNP positions.

SureSelect design. The library of SureSelect enrichment probes were designed over the genomic interval (mm9, chr11:109010000–114878000) using the SureDesign online tool of Agilent. Probes are covering the entire genomic region and were not designed specifically in proximity of DpnII sites. The probes covered 88% of the interval.

Capture Hi-C. cHi-C libraries were prepared from homozygous E12.5 limb buds (except Dup-S) as described previously[25]. In summary, 5×10^6 to 1×10^7 cells were used for crosslinking, cell lysis, DpnII digestion, ligation and de-crosslinking. Re-ligated products were then sheared using a Covaris sonicator (duty cycle: 10%, intensity:5, cycles per burst: 200, time: 6 cycles of 60 s each, set mode: frequency sweeping, temperature: 4 to 7 °C). Adaptors were added to the sheared DNA and amplified according to Agilent instructions for Illumina sequencing. The library was hybridized to the custom-designed sure-select beads and indexed for sequencing (50 to 100 bp paired-end) following Agilent instructions. Capture Hi-C experiments were performed as singletons. As an internal control, we compared the results from six experiments for regions outside of the region of interest (chr11:109,010,001–110,250,000 and chr11:113,100,001–114,870,000). The cHi-C maps of the internal control were highly correlated between the six samples (Spearman R: WT/Dup-L = 0.96; WT/Dup-C = 0.94; WT/Dup-S = 0.96; WT/ΔBor = 0.96; WT/Dup-LΔBor = 0.96; Dup-L/Dup-C = 0.95; Dup-L/Dup-S = 0.96; Dup-L/ΔBor = 0.96; Dup-L/Dup-LΔBor = 0.97; Dup-C/Dup-S = 0.94; Dup-C/ΔBor = 0.94; Dup-C/Dup-LΔBor = 0.95; Dup-S/ΔBor = 0.96; Dup-S/Dup-LΔBor = 0.96; ΔBor /Dup-LΔBor = 0.96) confirming the high reproducibility of the methodology.

Bioinformatics. Preprocessing and mapping of paired-end sequencing data, as well as filtering of mapped di-tags was performed with the HiCUP pipeline v.0.5.8 (ref. 28). The pipeline used Bowtie2 v.2.2.6 (ref. 29) for mapping short reads to reference genome (NCBI37/mm9). Filtered di-tags were further processed with Juicebox[5] command line tools to bin di-tags (10 kb bins) and to normalize the map by KR normalization. For this, only reads with a MAPQ ≥ 30 were considered. The DNA-capturing step enriches genomic region chr11:109,010,001–114,878,000 on mm9 leading to three different regimes in the cHi-C map: (i) enriched versus enriched, (ii) enriched versus non-enriched, and (iii) non-enriched versus non-enriched. For binning and normalization only di-tags in regime (i) were considered. Therefore di-tags were filtered for the enriched region and mm9 coordinates were shifted by −109,010,000 bp. For Juicebox a custom chromosome sizes file containing only the enriched region on chr11 (length 5,868,000 bp) was used. After binning and normalization, coordinates were shifted back to their original values. However, in general duplicated regions yield more signal compared to non-duplicated regions when mapped to the wildtype reference genome. The signal is flattened disproportionately in duplicated regions by a normalization procedure that balances the whole interaction matrix, such as KR normalization. Therefore, we used only raw count maps to calculate the differences between samples. Difference maps were generated based on raw count maps, scaled individually by dividing each value of the matrix by a factor (sum 'masked' triangle matrix/10⁶). To avoid copy number biases, the region spanning all tested duplications (chr11: 110,770,001–112,520,000) was not considered for the computation of the scaling factor. cHi-C maps of count values and difference maps were visualized with the WashU epigenome browser[30].

RNA-seq. E12.5 distal limbs were microdissected from wild-type and mutant embryos ($n = 2$) and immediately frozen in liquid nitrogen. From these RNA was isolated using TRIzol extraction and the RNeasy Mini Kit (QIAGEN). Samples were poly-A enriched and sequenced (paired-end 50 bp) using Illumina technology following standard protocols.

Bioinformatics: 50 bp paired-end reads were mapped to the mouse reference genome (mm9) using the STAR mapper[31] (splice junctions based on RefSeq; options: –alignIntronMin 20–alignIntronMax 500000–outFilterMultimapNmax 5–outFilterMismatchNmax 10–outFilterMismatchNoverLmax 0.1). Reads per gene were counted as described previously[3], and used for differential expression analysis with the DEseq2 package[32].

Phenotypic analysis. Phenotypic analysis for mutant mouse lines was carried out for at least three animals per analysis and developmental stage. Phenotypic analysis was carried out in heterozygous animals, for ΔBor in heterozygous and homozygous animals, and for Dup-LΔBor in homozygous animals. Penetrance of the phenotypes was investigated by analysing >30 offspring and considered fully penetrant if all animals were similarly affected.

Micro-computer tomography. Autopods of seven-week-old control and mutant mice ($n = 3$) were scanned using a Skyscan 1172 X-ray microtomography system (Brucker microCT, Belgium) at 5 μm resolution. 3D model reconstruction was done with the Skyscan image analysis software CT-Analyser and CT-volume (Brucker microCT, Belgium).

Whole-mount *in situ* hybridization and histology. Wild-type and mutant E12.5 embryos ($n = 5$) were subjected to whole mount *in situ* hybridization using standard procedures. *Sox9* and *Kcnj2* probes were generated by PCR amplification using mouse limb bud cDNA (Supplementary Table 1).

For Masson–Goldner staining, limbs of newborns (P0, $n = 2$) were first fixed in 4% PFA in PBS containing 0.5 M EDTA and then embedded in paraffin. Paraffin sections of 6 μM were then stained with the Merck Masson–Goldner staining kit (catalogue number 100485) and Meyers haematoxylin (catalogue number MHS80) according to the manufacturer's instructions.

LacZ staining and optical projection tomography. E12.5 mouse embryos ($n = 3$) were dissected in cold PBS, fixed in 4% PFA in PBS on ice for 30 min, washed twice with ice-cold PBS and 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[22]. After staining, embryos were washed in PBS and stored at 4 °C in 4% PFA in PBS.

For OPT scanning, stained embryos were embedded in 1% low-melt agarose and dehydrated over 1–2 days using methanol (2–3 methanol steps). Subsequently samples were cleared overnight in BABB (1 part benzyl alcohol, 2 parts benzyl benzoate). The samples were then scanned (802 frames) with a Bioptonics OPT 3001 M scanner with white light (exposure = 10–100 ms) for LacZ staining using GFP filter combination (exciter 425 nm/40 nm - emitter LP475 nm) and 20–380ms exposure for autofluorescence. 3D reconstructions were generated using Skyscan software and further analysed with Amira and Imaris software.
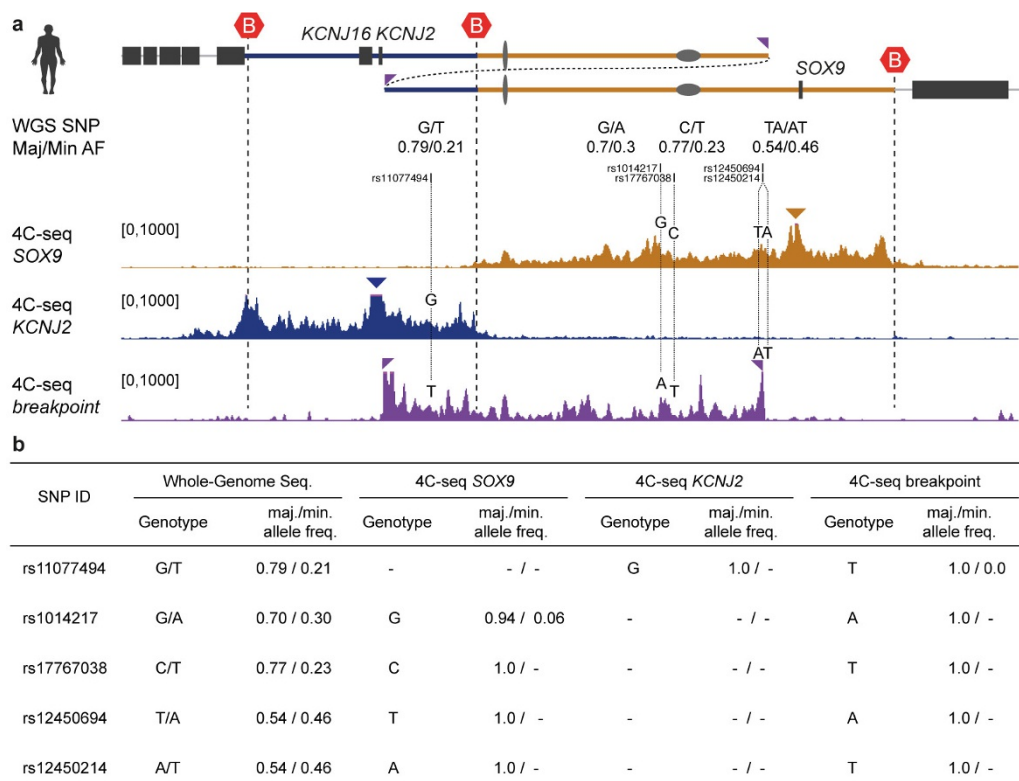
Human material. Skin biopsies from one sex reversal patient, the no phenotype individual, one Cooks syndrome patient and controls, as well as gonadal ('testicular') fibroblasts from one sex reversal individual were obtained by standard procedures. Fibroblasts were cultured in DMEM (Lonza) supplemented with 10% fetal calf serum (Gibco), 1% L-glutamine (Lonza) and 1% penicillin/streptomycin (Lonza). Written informed consent was obtained from all individuals studied that participated in this study. This study was approved by the Charité Universitätsmedizin Berlin ethics committee.

20. Kraft, K. et al. Deletions, inversions, duplications: engineering of structural variants using CRISPR/Cas in Mice. Cell Reports 10, 833–839 (2015).
21. Artus, J. & Hadjantonakis, A.-K. in Transgenic Mouse Methods and Protocols Vol. 693 Methods in Molecular Biology (eds Hofker, M. H. & van Deursen, J.) Ch. 3, 37–56 (Humana Press, 2011).
22. Ruf, S. et al. Large-scale analysis of the regulatory architecture of the mouse genome with a transposon-associated sensor. Nature Genet. 43, 379–386 (2011).
23. 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).
24. Nagy, K. & Nichols, J. in Advanced Protocols for Animal Transgenesis (eds Pease, S. & Saunders, T. L.) Ch. 18, 431–455 (Springer, 2011).
25. van de Werken, H. J. G. et al. in Methods in Enzymology Vol. 513 (eds Wu, C. & Allis, D. C.) 89–112 (Academic Press, 2012).
26. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25, 1754–1760 (2009).
27. McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303 (2010).
28. Wingett, S. et al. HiCUP: pipeline for mapping and processing Hi-C data. F1000Res. 4, 1310 (2015).
29. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359 (2012).
30. Zhou, X. et al. Exploring long-range genome interactions using the WashU Epigenome Browser. Nature Methods 10, 375–376 (2013).
31. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21 (2013).
32. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550 (2014).
33. Bhatia, S. et al. Functional assessment of disease-associated regulatory variants in vivo using a versatile dual colour transgenesis strategy in zebrafish. PLoS Genet. 11, e1005193 (2015).
34. Benko, S. et al. Highly conserved non-coding elements on either side of SOX9 associated with Pierre Robin sequence. Nature Genet. 41, 359–364 (2009).
35. 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).
36. Gordon, C. T. et al. Identification of novel craniofacial regulatory domains located far upstream of SOX9 and disrupted in Pierre Robin sequence. Hum. Mutat. 35, 1011–1020 (2014).
37. Yao, B. et al. The SOX9 upstream region prone to chromosomal aberrations causing campomelic dysplasia contains multiple cartilage enhancers. Nucleic Acids Res. 43, 5394–5408 (2015).
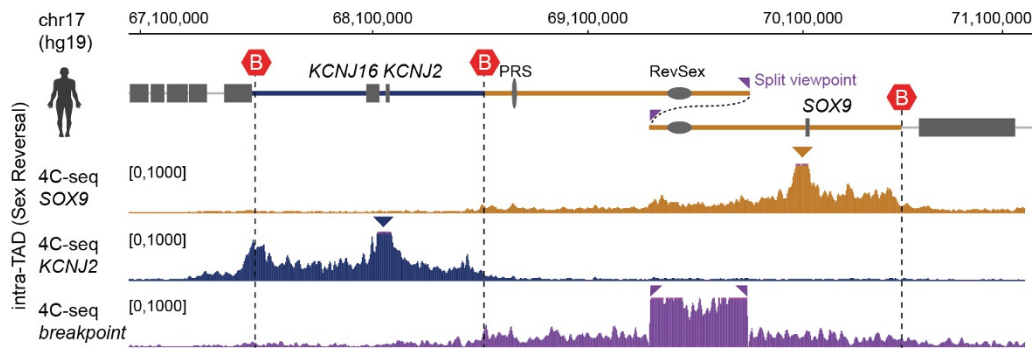
Extended Data Figure 1 | 4C-seq from patient fibroblasts with a 470 kb intra-TAD duplication (sex reversal). Schematic representation depicts TAD structure with *KCNJ* TAD (blue), *SOX9* TAD (brown) and TAD boundaries (red hexagons). Size and position of the 470 kb intra-TAD duplication in a patient with female-to-male sex reversal is indicated by the overlap. Patient fibroblasts were derived from a gonadal biopsy of ovo-testes. 4C-seq tracks below show interaction profiles from *SOX9* (brown), *KCNJ2* (blue) and breakpoint (purple) viewpoint. Note interaction profiles from *SOX9* and *KCNJ2* are restricted to their corresponding TADs. Unique viewpoint at the breakpoint shows that interactions are restricted to the *SOX9* TAD. Note all reads mapped to a wild-type genome resulting in split viewpoint for the duplication breakpoint.

a, Schematic of SNP-analysis in a patient carrying an inter-TAD
duplication (heterozygous) with no phenotype. SNP positions and
their allele frequency (AF) identified by whole-genome sequencing
from the patient are shown. Bottom, 4C-seq interaction profiles using

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

**Extended Data Figure 2 | Allele frequencies determined in 4C-seq
data indicate selective interactions in a patient with an inter-TAD
duplication (no phenotype).** SNP analysis related to Fig. 1d.

SOX9 (brown), *KCNJ2* (blue) and the duplication breakpoint (purple)
as viewpoints (triangles). SNPs contacted by individual viewpoints are
indicated. Note that variants contacted by the breakpoint viewpoint are
not contacted by *SOX9* and *KCNJ2* viewpoints, suggesting the formation
of an insulated interaction domain (purple). **b**, Summary of observed SNP
genotype and allele frequency from whole-genome sequencing data and
4C-seq viewpoints.
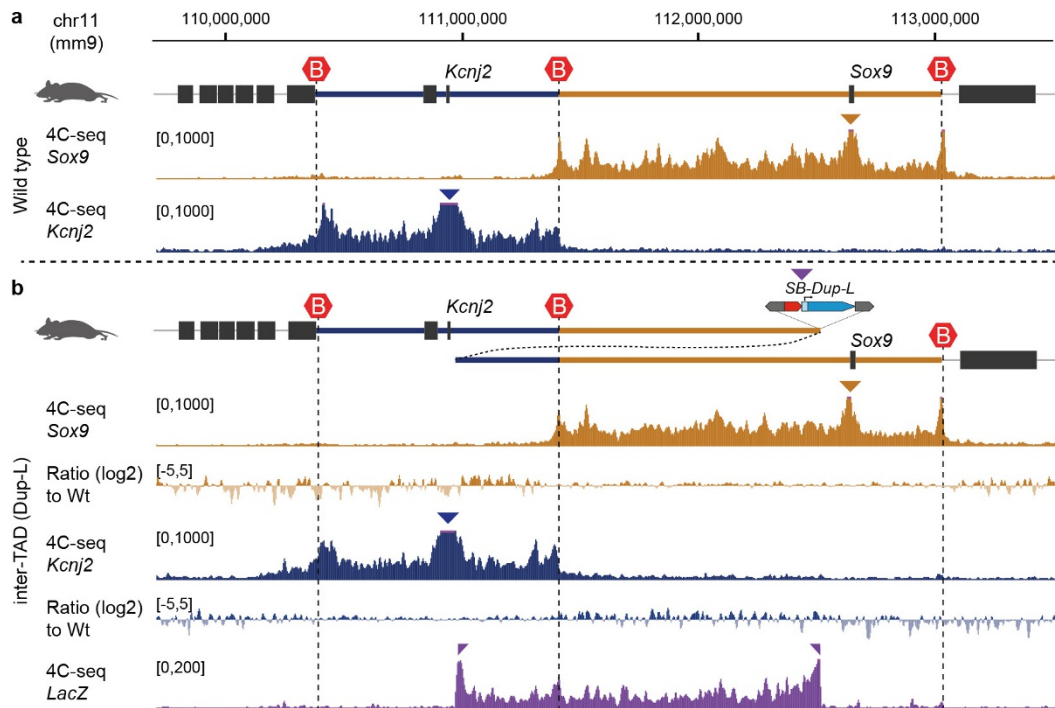
Extended Data Figure 3 | 4C-seq from an inter-TAD duplication in Dup-L mouse mutants. a, Schematic representation depicts TAD structure in wild type with *Kcnj* TAD (blue), *Sox9* TAD (brown) and TAD boundaries (red hexagons). Below, 4C-seq interaction profiles of viewpoints (triangles) in *Sox9* (brown) and *Kcnj2* (blue) from E12.5 wild-type limb buds. **b**, Schematic representation of Dup-L allele. Position of *lacZ* reporter at the duplication breakpoint is shown and duplication is indicated by overlap. Note that 4C-seq reads are mapped to the wild-type genome, which results in split viewpoint from *lacZ* viewpoint. 4C-seq profiles with viewpoint in *Sox9* (brown), *Kcnj2* (blue) and *lacZ* reporter (purple) in Dup-L are shown below. *Kcnj2* and *Sox9* profiles are unchanged, whereas the unique viewpoint in the *lacZ* reporter shows interactions that are restricted to the duplicated region, suggesting formation of a separate interaction domain.

Extended Data Figure 4 | Overview of mouse alleles used in this study. The extent of each duplicated and deleted region in mice are given as mm9 coordinates. Observed mouse phenotypes and human syndromes with the equivalent mutation are listed. For generation of alleles see Methods. SB, Sleeping Beauty transgene containing a single *loxP* site (red) and a *lacZ* reporter gene (blue) flanked by SB-transposons.
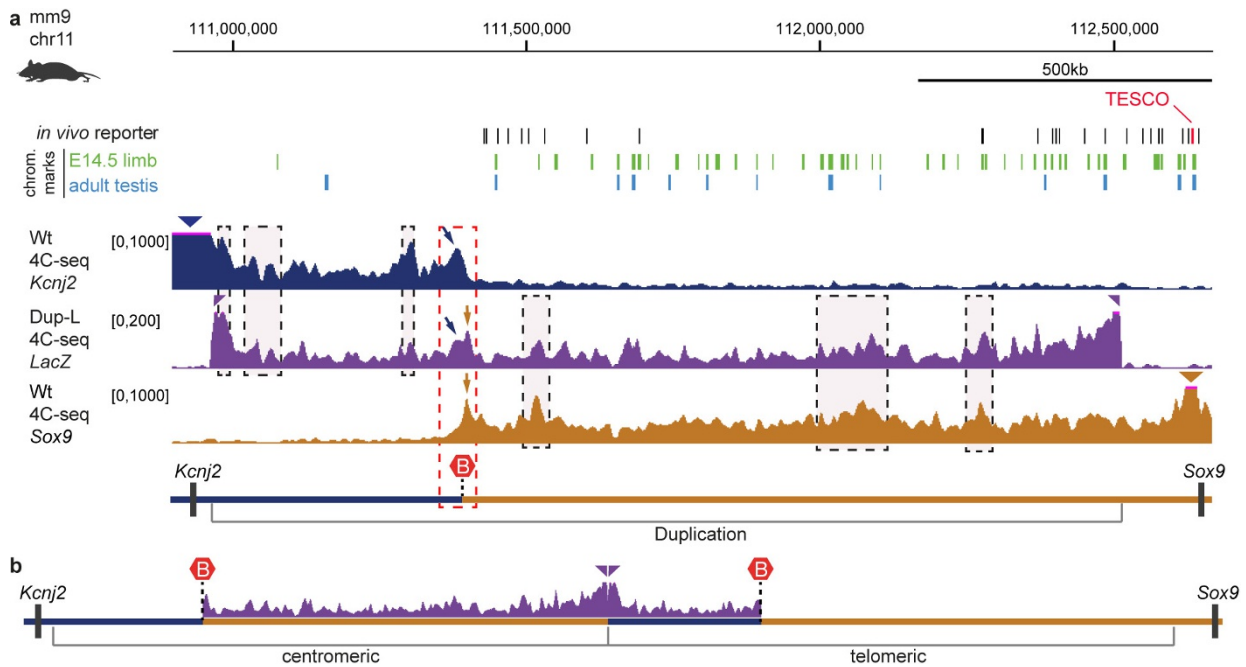
Extended Data Figure 5 | 4C-seq interaction profile of neo-TAD in Dup-L mutants resembles profiles from *Sox9* and *Kcnj2* viewpoints.
**a**, The position of published *in vivo* tested reporter constructs[11,33–37] and active chromatin marks (H3K4me1/H3K27ac positive) in *Sox9*-expressing tissue (mouse ENCODE project) are indicated above the 4C tracks. The only known enhancer driving testis-specific expression, TESCO, is indicated. 4C-seq interaction profiles from *Kcnj2* (blue) and *Sox9* (brown) in wild type and from the *lacZ* reporter gene (purple) in E12.5 Dup-L mutants are shown. Triangles indicate viewpoints. Note that 4C-seq reads are mapped to the wild-type genome resulting in split viewpoint from the *lacZ* viewpoint. The interaction profile from *lacZ* viewpoint located at the breakpoint of the Dup-L duplication shows a similar pattern and peak distribution (dashed boxes) as the *Sox9* and *Kcnj2* viewpoints. Note similarities of peak profiles at the boundary between *Kcnj* and *Sox9* TADs (arrows in red dashed box). **b**, Schematic of locus with an artificially duplicated region and corresponding 4C-seq profile from the *lacZ* viewpoint inside the neo-TAD in Dup-L mutants. The centromeric and telomeric part of the tandem duplication is indicated. The purple 4C-seq profile corresponds to the neo-TAD delimited by the duplicated TAD boundary.
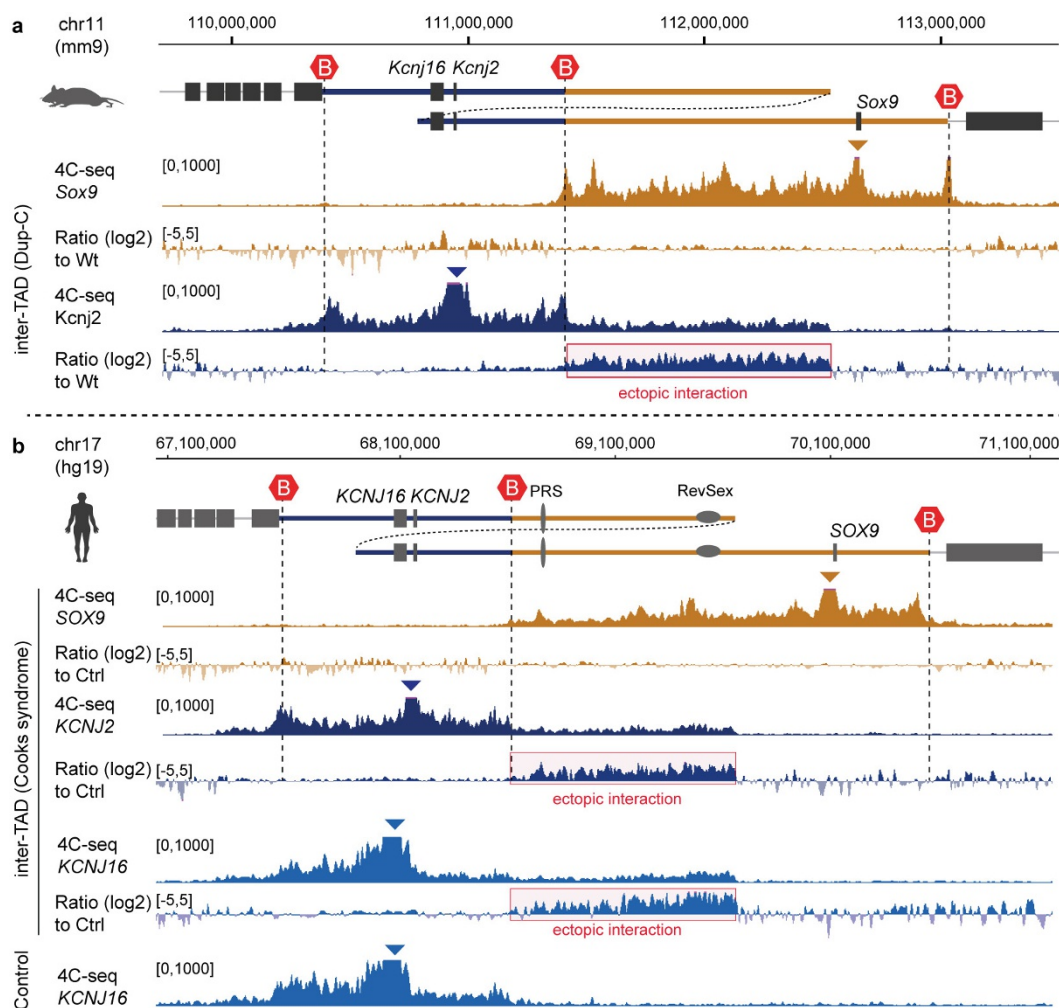
Extended Data Figure 6 | 4C-seq of inter-TAD Cooks syndrome duplications in mouse and human. Extent and position of the duplications is indicated by the overlap in the schematic. 4C-seq profiles with indicated viewpoints and ratio to control (below) **a**, *Sox9* (brown) and *Kcnj2* (blue) 4C-seq from Dup-C mutant limb buds at E12.5. **b**, 4C-seq from fibroblasts of an individual with Cooks syndrome and inter-TAD duplication from *SOX9* and *KCNJ2* viewpoints. Below, 4C-seq with viewpoint in *KCNJ16* (light blue) in patient and control fibroblasts and the ratio of patient to control. **a, b**, Incorporation of *Kcnj/KCNJ* genes in the neo-TAD results in ectopic contacts from the *Kcnj2* or *KCNJ2* and *KCNJ16* viewpoints with the duplicated part of the *Sox9* TAD, whereas interactions from the *Sox9/SOX9* viewpoint remain unchanged.
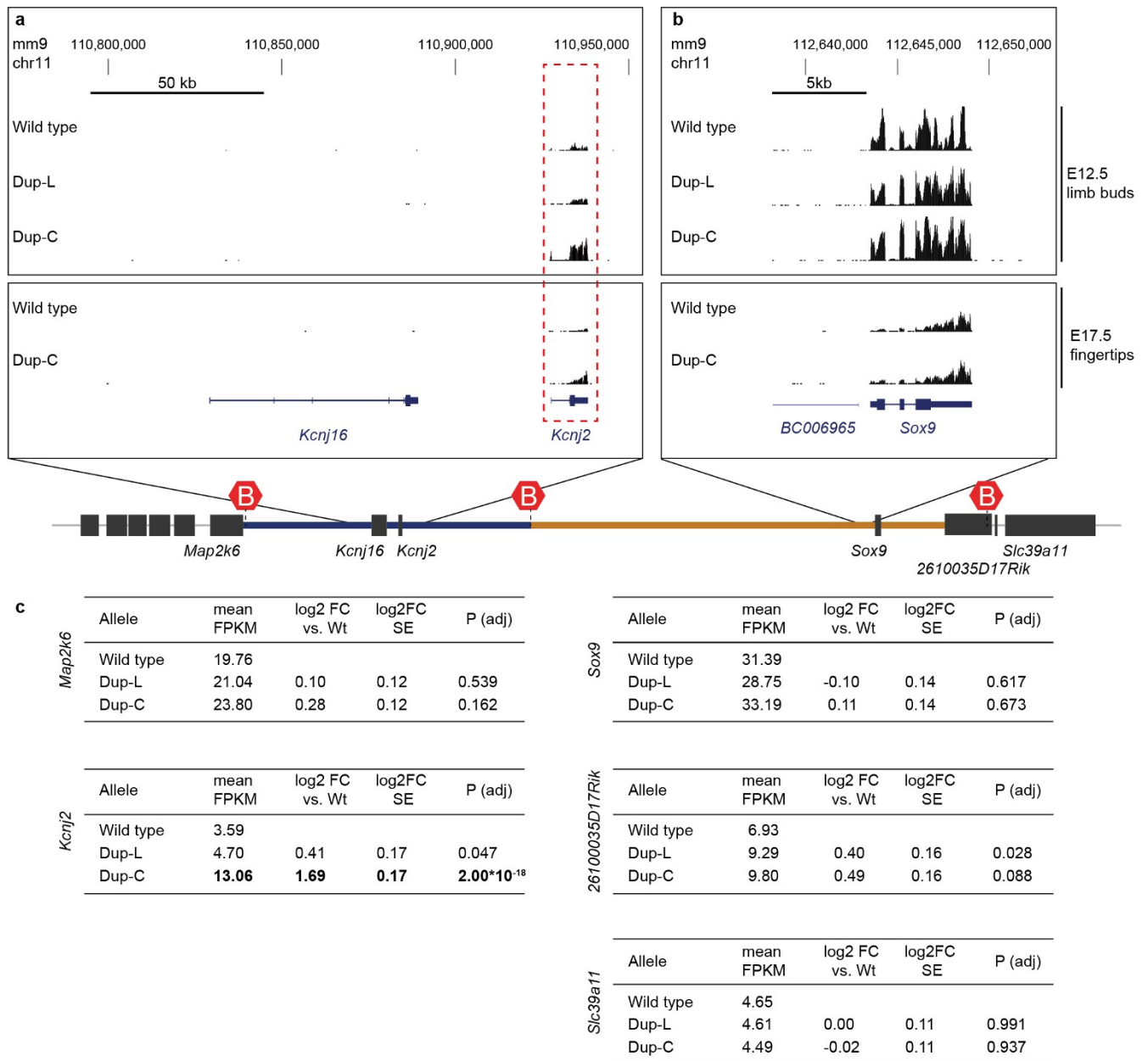
Extended Data Figure 7 | RNA-seq analysis of genes at the *Sox9, Kcnj2* and *Kcnj16* locus from mouse mutants used in this study. a, b, RNA-seq from wild-type and mutant E12.5 limb buds and E17.5 fingertips (lower two tracks) shows read profiles for *Kcnj16*, *Kcnj2* and *Sox9*. Note the absence of expression of *Kcnj16* in the examined tissue. **c**, Summary of expression values of genes at the *Sox9*, *Kcnj2* and *Kcnj16* locus from E12.5 limb buds. Significant expression changes are in bold. FPKM, fragments per kilobase of exon per million fragments mapped. Benjamini–Hochberg-adjusted *P* value, $n = 2$, cut-off $= 0.001$.
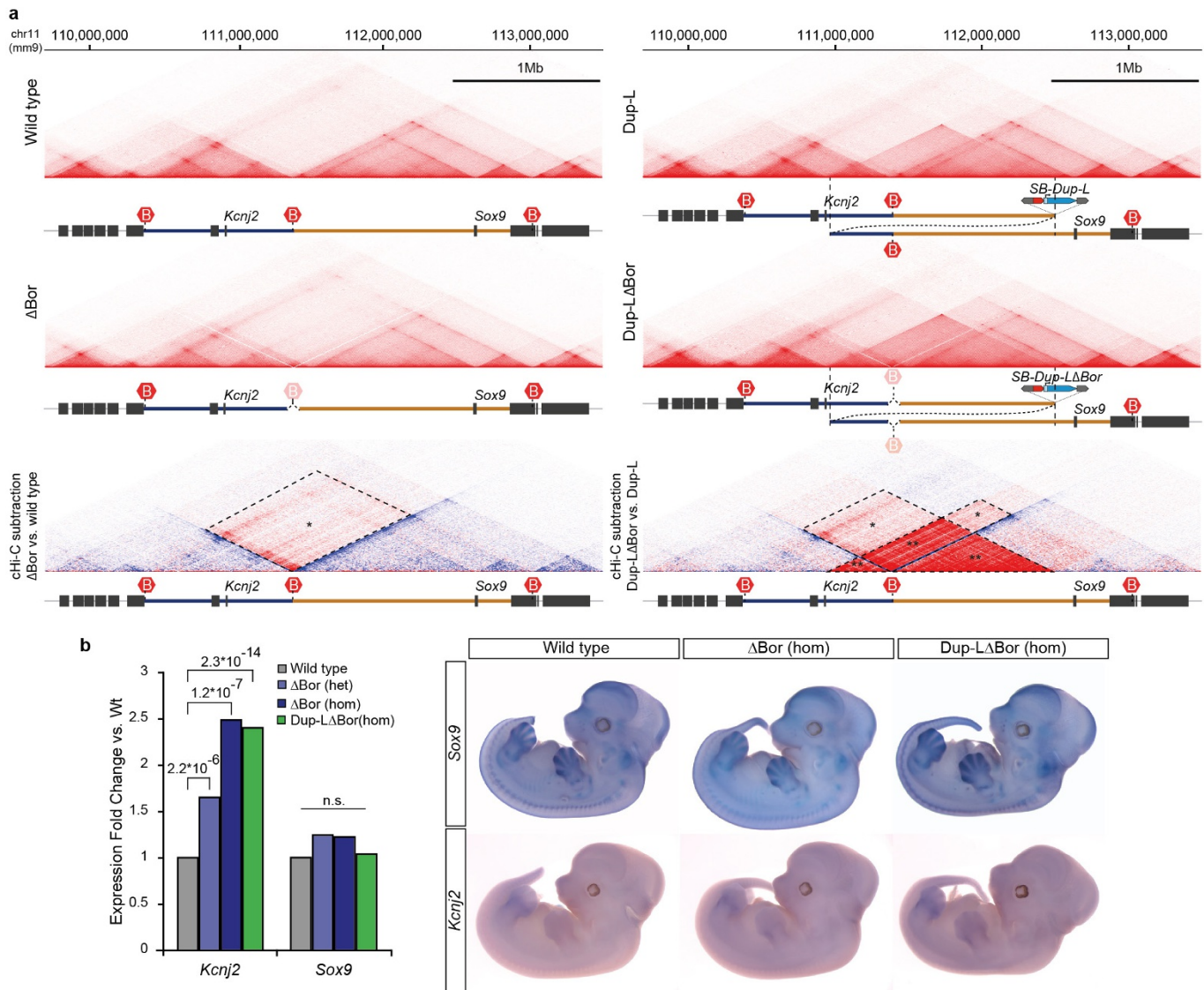
Extended Data Figure 8 | Deletion of the TAD boundary region in wild-type and Dup-L mutant mice. a, Left, cHi-C from wild type (top) and ΔBor (middle) E12.5 limb buds and subtraction map of ΔBor relative to wild type (bottom). Deletion of the boundary between the *Kcnj* and *Sox9* TADs leads to loss of insulation and ectopic interactions (*) between the TADs. Note overall TAD structure remains intact and ectopic interaction is restricted by remaining TAD boundaries. Right, cHi-C from Dup-L (top) and Dup-LΔBor (middle) E12.5 limb buds and subtraction map of Dup-LΔBor relative to Dup-L (bottom). Deletion of the two duplicated boundary regions flanking the neo-TAD, results in ectopic contacts of duplicated sequences with adjacent *Kcnj* and *Sox9* TADs (*) including *Sox9* and *Kcnj2*, as seen in ΔBor mutants. Loss of neo-TAD insulation results in increased interactions of the duplicated sequences (**). **b**, Expression analysis of *Sox9* and *Kcnj2* in wild-type, ΔBor and Dup-LΔBor embryos. RNA-seq expression analysis of mutant versus wild-type E12.5 limb buds. *Kcnj2* is upregulated in ΔBor (heterozygous and homozygous) and Dup-LΔBor (homozygous) limb buds (Benjamini-Hochberg adjusted *P* values, $n = 2$). Whole mount *in situ* hybridization shows no site-specific misexpression of *Kcnj2* in a *Sox9*-like pattern.