OPEN ACCESS

Original research

Disruption of the topologically associated domain at Xp21.2 is related to 46,XY gonadal dysgenesis

Jakob A Meinel,[1] Verónica Yumiceba ,[2] Axel Künstner ,[3,4] Kristin Schultz,[2] Nathalie Kruse,[2] Frank J Kaiser,[5,6] Paul-Martin Holterhus,[7] Alexander Claviez ,[8] Olaf Hiort,[1] Hauke Busch ,[3,4] Malte Spielmann ,[2,9] Ralf Werner [1,10]

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/jmg-2022-108635).

For numbered affiliations see end of article.

Correspondence to
Dr Ralf Werner, Department of Pediatrics and Adolescent Medicine, Division of Pediatric Endocrinology and Diabetes, Universität zu Lübeck, Lübeck 23562, Germany; ralf.werner@uni-luebeck.de

JAM and VY contributed equally.

JAM and VY are joint first authors.

Received 29 April 2022
Accepted 25 August 2022
Published Online First 9 September 2022

To cite: Meinel JA, Yumiceba V, Künstner A, *et al*. *J Med Genet* 2023;**60**:469–476.

ABSTRACT

Background Duplications at the Xp21.2 locus have previously been linked to 46,XY gonadal dysgenesis (GD), which is thought to result from gene dosage effects of *NR0B1* (*DAX1*), but the exact disease mechanism remains unknown.

Methods Patients with 46,XY GD were analysed by whole genome sequencing. Identified structural variants were confirmed by array CGH and analysed by high-throughput chromosome conformation capture (Hi-C).

Results We identified two unrelated patients: one showing a complex rearrangement upstream of *NR0B1* and a second harbouring a 1.2 Mb triplication, including *NR0B1*. Whole genome sequencing and Hi-C analysis revealed the rewiring of a topological-associated domain (TAD) boundary close to *NR0B1* associated with neo-TAD formation and may cause enhancer hijacking and ectopic *NR0B1* expression. Modelling of previous Xp21.2 structural variations associated with isolated GD support our hypothesis and predict similar neo-TAD formation as well as TAD fusion.

Conclusion Here we present a general mechanism how deletions, duplications or inversions at the *NR0B1* locus can lead to partial or complete GD by disrupting the cognate TAD in the vicinity of *NR0B1*. This model not only allows better diagnosis of GD with copy number variations (CNVs) at Xp21.2, but also gives deeper insight on how spatiotemporal activation of developmental genes can be disrupted by reorganised TADs causing impairment of gonadal development.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Duplications at Xp21.2 have previously been linked to 46,XY gonadal dysgenesis (GD), and the aetiology has been attributed to enhanced *NR0B1* gene dosage.

⇒ The study presents the first individual with 46,XY GD harbouring a Xp21.2 duplication excluding NR0B1.

WHAT THIS STUDY ADDS

⇒ We provide a novel model of how duplications, deletions or inversions at Xp21.2 with or without *NR0B1* can lead to 46,XY GD by shuffling of topologically associating domains and enhancer hijacking.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ The findings of this study will improve future diagnostic practice in patients with 46,XY GD and help to elucidate the regulatory mechanisms of gonadal development induced by copy number variations.

INTRODUCTION

Mammalian sex determination is a time-dependent mechanism controlled by antagonising sex determination genes that direct the bipotential genital ridges to induce either ovarian or testicular development in any embryo. Gonadal dysgenesis (GD) as a diminished or absent reproductive system development occurs when there is an interference in timely expression of key genes and/or their appropriate gene expression levels are not reached. In the 46,XY embryo, GD leads to a difference/disorder of sex development (DSD) with a range of genital phenotypes from ambiguity of the external genitalia to female appearing genitalia, caused by the variable degrees of testicular failure. Even today, the genetic origin of GD remains elusive in a majority of up to 60% of cases.[1]

Thus far, genetic variants in more than 20 genes involved in sex determination have been described as monogenic causes of 46,XY GD. The genetic variants also include copy number variations (CNVs) of different loci, and the phenotype was explained by the altered allele number of candidate genes, that is, their gene dosage.[2] Among these, *NR0B1* (*nuclear receptor subfamily 0, group B, member 1*; also known as *DAX1*), is located within a 160 kb region termed dosage-sensitive sex reversal on chromosome Xp21.2.[3] During the last two decades, eight non-syndromic 46,XY GD patients with duplications at the Xp21.2 locus have been described encompassing *NR0B1*.[4–11] However, in only six of the cases, the approximate boundaries of the duplications were reported. In contrast, deletions or inactivating *NR0B1* mutations in 46,XY patients cause adrenal hypoplasia congenita (AHC).[12] Although individuals with AHC have a normal sexual development at birth, they present with hypogonadotropic hypogonadism at puberty (OMIM#300200).

In mice, *Nr0b1* expression starts in the genital ridge in both sexes and is synchronised with *Sry* (*sex determining region Y*) expression. However,

it is downregulated in the developing testis and persists in the ovary.[13 14] It has been shown that high exogenous *Nr0b1* expression in transgenic mice delays testis formation. Coexpression with a weak *Sry* allele even resulted in complete sex reversal,[15] resembling the phenotype of 46,XY GD patients with *NR0B1* duplications (OMIM #300018). This makes *NR0B1* the most plausible candidate gene for 46,XY GD in the Xp21.2 region.

The six previously published and well-characterised duplications at Xp21.2 associated with non-syndromic 46,XY GD harbour two copies of *NR0B1* indicating a gene dosage effect as the cause of GD. However, Smyk *et al*[16] reported a case of 46,XY GD carrying a Xp21.2 deletion upstream of *NR0B1* challenging the gene dosage hypothesis. This demands a novel explanation of the pathogenicity of structural variations (SVs) at this locus.

In recent years, the role of three-dimensional chromosome structure and their organisation in topologically associating domains (TADs) have described novel disease mechanisms.[17] TADs are mega-base, non-randomly organised regions that insulate local chromatin interaction between regulatory elements and their cognate promoters.[18] SVs disturbing TADs' boundaries were shown to alter the architecture and enhancer–promoter interactions within different domains leading to deleterious effects on development[17 19 20] and oncogenesis.[21 22] These recent data indicate that SVs should therefore be investigated for gene dosage effects and must be interpreted in the three-dimensional (3D) genomic context.

Here we present two new cases of 46,XY GD, with CNVs at Xp21.2 including and excluding *NR0B1*. Analysis of the size and orientation of the SVs by whole genome sequencing (WGS) and their effect on TAD structure by high throughput chromosome conformation capture (Hi-C) provides strong evidence for TAD disruption as a possible cause of GD in patients with Xp21.2 SVs. Our data indicate that TAD disruption and Neo-TAD formation with subsequent ectopic enhancer adoption are the most likely causes of GD in patients with SV at the Xp21.2 locus.

MATERIALS AND METHODS

Patient description

Patient 1 (P1)

First presentation in the University DSD Center in Kiel and Lübeck occurred during early adolescence due to primary amenorrhoea and pubertal delay. The girl had been treated for 15 months due to hyperprolactaemia and a left-sided gonadal tumour (6×11×11 cm) was surgically removed 3 months earlier. Histology revealed dysgerminoma. Tanner stages were B4 and P4-5. However, breast development started only few months before clinical tumour diagnosis. External genitalia were entirely female without clitoromegaly, and a uterus with tubular configuration was present. The karyotype was 46,XY. Hormonal evaluation revealed basal luetinizing hormone (LH) 53 IU/L (normal range for 46,XY control males, Tanner 4, 1.2–3.4 IU/L) and follicle stimulating hormone (FSH) 94.3 IU/L (normal range for 46,XY control males, Tanner 4, 3.0–5.2 IU/L) increasing to 200 IU/L (normal range for 46,XY control males, Tanner 4, 12.2–29.4 IU/L) and 128 IU/L (normal range for 46,XY control males, Tanner 4, 4.9–9.6 IU/L), respectively, 30 min following 60 µg/m^2 GnRH intravenously. Plasma oestradiol was 33 pmol/L (age-dependent reference interval 10.0–221.5 pmol/L for 46,XY control males and 10–507 pmol/L for 46,XX control females) (prepubertal for girls) and testosterone 4.26 nmol/L (age-dependent reference interval 0.1–17.6 nmol/L for 46,XY control males and 0.1–2.0 nmol/L for 46,XX control females) (midpubertal for males) (both determined by liquid chromatography tandem mass spectrometry (LC-MS/MS)).[23] Plasma anti-Mullerian hormone (AMH) was 5.1 pmol/L (age-dependent range for 46,XY control males with Tanner 4–5, 48±14 pmol /L (±SEM)[24]), which is extremely low. Therefore, the clinical diagnosis of GD was established, and the patient underwent whole genome sequencing to investigate the molecular genetic aetiology of her GD. Laparoscopic gonadectomy of the right side was performed and showed gonadoblastoma with focal transition into dysgerminoma. Five years after initial diagnosis, the patient remains in first remission.

Patient 2 (P2)

This patient was born at term after an uneventful pregnancy, and genital status was described as unequivocal female. The family history is unremarkable, with healthy siblings. Because of muscular hypotonia, a chromosomal analysis was initiated and revealed a 46,XY karyotype. On ultrasound, a small prepubertal uterus was seen, but the gonads could not be visualised. Hormone analysis in early childhood showed a prepubertal status with inhibin B below the threshold and AMH at 3.0 pmol/L (age-dependent range for 46,XY control males, 499±66 pmol /L (±SEM),[24]), which was considered low for her age, compatible with a clinical diagnosis of GD. A laparoscopy was performed, and the gonadal tissue was removed. The histology revealed a gonadoblastoma of the right gonad, while the left side was purely stromal tissue. Quantitative PCR was suggestive of a copy number gain at Xp21.2. The patient was seen several times and developed epilepsy in middle childhood.

Whole genome sequencing

WGS was performed to identify deleterious point mutations and indels as well as SVs and to fine map the breakpoints of the identified duplications and triplication at Xp21.2 in both patients. P1 was sequenced as a trio here and P2 as a singleton. Sequencing libraries were constructed from 1.0 µg DNA per sample using the Truseq Nano DNA HT Sample preparation Kit (Illumina, USA) following the recommendations of the manufacturer. Genomic DNA was randomly fragmented to a size of approximately 350 bp by Covaris cracker (Covaris, USA). DNA fragments were blunted, A-tailed and ligated with the full-length adapter for Illumina sequencing with further PCR amplification. Libraries were purified using AMPure XP (Beckman Coulter, USA), analysed for size distribution on an Agilent 2100 Bioanalyser (Agilent Technologies) and quantified by qPCR. Paired end sequencing was performed on Illumina HiSeq platforms (Illumina). Per sample more than 90 GB of raw data were obtained, resulting in an average genomic read depth of 30×.

Bioinformatics

Sequencing reads were mapped to human reference genome version GRCh37/hg19 using Burrows-Wheeler Aligner.[25] Resulting mapping files were screened for duplicated reads applying Picard tools *MarkDuplicates* version 1.111 (Picard: http://sourceforge.net/projects/picard/). Split-reads and discordant paired-end alignments were extracted using *SAMtools* V.0.1.18.[25] SNPs and InDels were called using *HaplotypeCaller* as implemented in Genome Analysis Toolkit V.3.8.0,[26] with standard parameters. Detection of SVs was performed using DELLY.[27] Variations were annotated using ANNOVAR.[28]

CNV validation

For validation of CNVs identified through WGS and qPCR, array-based comparative genomic hybridisation (aCGH) was

performed. DNA of P1, P2 and the mother of P2 was hybridised to an Agilent 180K aCGH (Agilent Technologies, Inc) and was either compared with a pooled sample of 10 normal males (P1 and P2) or 10 normal females (mother of P2). Data were analysed using *CytoGenomics* Software V.4.0.2.21 (Agilent). CNVs were analysed using the Database of Genomic Variants (DGV, Version CNV_DGV_hg19_v4, Toronto, Canada).

Preparation of Hi-C libraries

Lymphoblastoid cell lines (LCLs) were established by Epstein-Barr virus transformation of leucocytes from peripheral blood samples of P1 as well as of controls. Fibroblast cell lines were established from a skin biopsy of the mother of P2.

In situ Hi-C libraries were processed as described previously,[29] with minor modifications. Briefly, ~1 million cells were harvested, and genomic material was cross-linked with 2% of formaldehyde (PanReacAppliChem, A0877) in intact nucleus as above. gDNA from lysed cells was digested with a total of 200 U of *DpnII* enzyme (New England BioLabs (NEB), R0543) at 37°C, split in 2 intervals of 30 min each. The 5′ overhang of restricted fragments were filled in with biotin-14-dATP (Thermo Fisher Scientific, 19524016) plus dCTP, dGTP and dTTP (NEB, N0446) all at 0.3 mM. The resulting blunted-ends were ligated overnight at 4°C with 2000 U of T4 DNA ligase (NEB, M0202). DNA samples were reverse cross-linked with 25 µL Proteinase K (QIAGEN, 19131) and 1% SDS at 55°C for 30 min and 4-hour incubation with 0.5M NaCl (final concentration) at 68°C. DNA was subsequently purified by ethanol precipitation at 4°C. Hi-C libraries were prepared by shearing ~3 µg of DNA with Bioruptor Pico (Diagnode) to obtain fragments between 300 and 700 bp (12 cycles of 20s on, 60 s off each cycle). Sonicated and biotin-filled in fragments were pulled down using 150 µL Dynabeads MyOne Streptavidin T1 beads (Thermo Fisher Scientific, 65602). The DNA ends were repaired using 12 U of T4 DNA polymerase (NEB, M0203), 5U of Klenow fragment of DNA polymerase I (NEB, M0210) and 50 U of T4 Polynucleotide Kinase (NEB, M0201). Hi-C libraries were processed according to the NEBNext Multiple Oligos kit (E7335): adding first the adaptors and later the indexes through PCR amplification on beads (four to six cycles) using the NEBNext Ultra II Q5Master Mix (NEB, M0544). Double size selection (0.55X and 0.7, respectively) was carried out using Agencourt AMPure XPbeads (Beckman Coulter, A63881) to clean up the PCR products. Finally, Hi-C libraries were qPCR quantified by NEBNext Library Quant Kit (NEB, E7630) and sequenced (~200 million fragments) in a 150 bp paired-end run on a NextSeq2000 (Illumina).

Hi-C bioinformatic analysis

Interaction maps were generated using the HiC-Pro pipeline V.3.0.0 in parallel mode. The pipeline was set up with access to Bowtie2 V.2.3.5.1, Samtools V.1.9, R V.3.6.3 and Python V.3.7.6. Ligation sites for *DnpII* enzyme were generated with the HiC-Pro utils script digest genome using hg19 as reference.

Paired-end sequencing data were aligned with Bowtie2, parameters were HiC-Pro default, that is, very sensitive (modified seed length 30 for global alignment), score-min L, −0.6 to –0.2 and end-to-end. Reads were mapped to hg19. Singletons and reads with MAPQ <10 were discarded, and duplicates were removed. The resulting valid pairs were converted to Juicer format using the HiC-Pro utility script hicpro2juicebox, which uses Juicer Tools V.1.22.01. Juicer Tools was then used to add Knight and Ruiz (KR) matrix balancing. Since the normalisation assumes equal visibility of all loci and would distort the display of CNVs, we used raw interaction counts the locus of interest.

Genome-wide Hi-C maps were visualised using Juicebox (V.1.11.8). Intrachromosomal TADs in chromosome X were extracted using straw library function at a resolution of 10 kb and visualised as a heatmaps rotated 45°. Hi-C maps were compared with sex and cell-type matched controls. ChIP-seq ENCODE data for CTCF (human mammary fibroblast and GM12864; B-lymphoblastoid cell lines) were overlapped with heatmaps.

All chromosomal positions in this paper are according to GRCh37/hg19.

RESULTS

We generated whole genome data from two patients with GD and previously undescribed SVs at the Xp21.2 locus (figure 1). The SVs in both patients differed in size, complexity and their direct inclusion of *NR0B1*. Whereas P1 has a complex rearrangement of two duplications and two small deletions in the proximity of *NR0B1,* P2 carries a triplication including *NR0B1*. Copy number gains and losses were verified by aCGH (online supplemental figures S1 and S2) and WGS split reads at CNV borders were used to construct continuous breakpoint sequences. Thus, determining orientation and location of inserted copy number gains and losses (online supplemental figures S3 and S4). Breakpoint sequences were verified by Sanger sequencing (described in online supplemental). WGS also excluded any other known genetic cause of 46,XY GD in the patients. In both patients the SVs were maternally inherited.

P1 WGS revealed two major duplications and two small deletions at Xp21.2; one 389 kb duplication maps to a region downstream of *NR0B1* containing the *MAGEB* (*MAGE family member B*) genes 1–4 and a part of *IL1RAPL1* (*interleukin 1 receptor accessory protein like 1*). The second 447 kb duplication, containing *TASL* and *GK* and the 3′-part of *TAB3,* maps upstream of *NR0B1*. Between the two duplications two small deletions of 2.7 kb and 2.2 kb flank an inverted region of 1.2 kb (online supplemental figure S3). Both duplications are inserted upstream of *NR0B1*. Notably, the 447 kb duplication encompassing *TASL*, *GK* and the *TAB3* fragment is inserted proximal to *NR0B1* in an inverted position. The 389 kb duplication of *MAGEB1-4* and the *IL1RAPL1* fragment is inserted further upstream in the orientation of the reference sequence. WGS of both parents established that the SV was maternally inherited.

Investigation of other known DSD candidate genes in P1, revealed only one rare variant (MAF <0.01) of unknown significance in *ZFPM2 (zinc finger protein, FOG family member 2*) transmitted by the mother. *ZFPM2* variants are associated with abnormalities in testis determination; however, the patients SNV (dbSNP:rs202217256) has an average population frequency of 0.004 across all populations[30] and is listed as benign by ClinVar Miner (supplemental Table S1). Considering an incidence of 46,XY GD of around 1.5 per 100 000,[31] this SNV could be catalogued as a rare polymorphism, rather than a relevant pathogenic variant leading to GD.

P2 harbours a Xp21.2 triplication initially identified through qPCR copy number detection (online supplemental methods). In contrast to P1, this CNV includes the *NR0B1* gene. The 1.24 Mb triplication includes the genes *MAGEB1-4*, *NR0B1*, *TASL*, *GK*, *TAB3* and part of *IL1RAPL1*. The triplicated segments are arranged in tandem and are separated by a 49 bp insert (online supplemental figure S4).

Analysis of known DSD candidate genes revealed only SNVs reported as benign or likely benign in ClinVar. Except from a

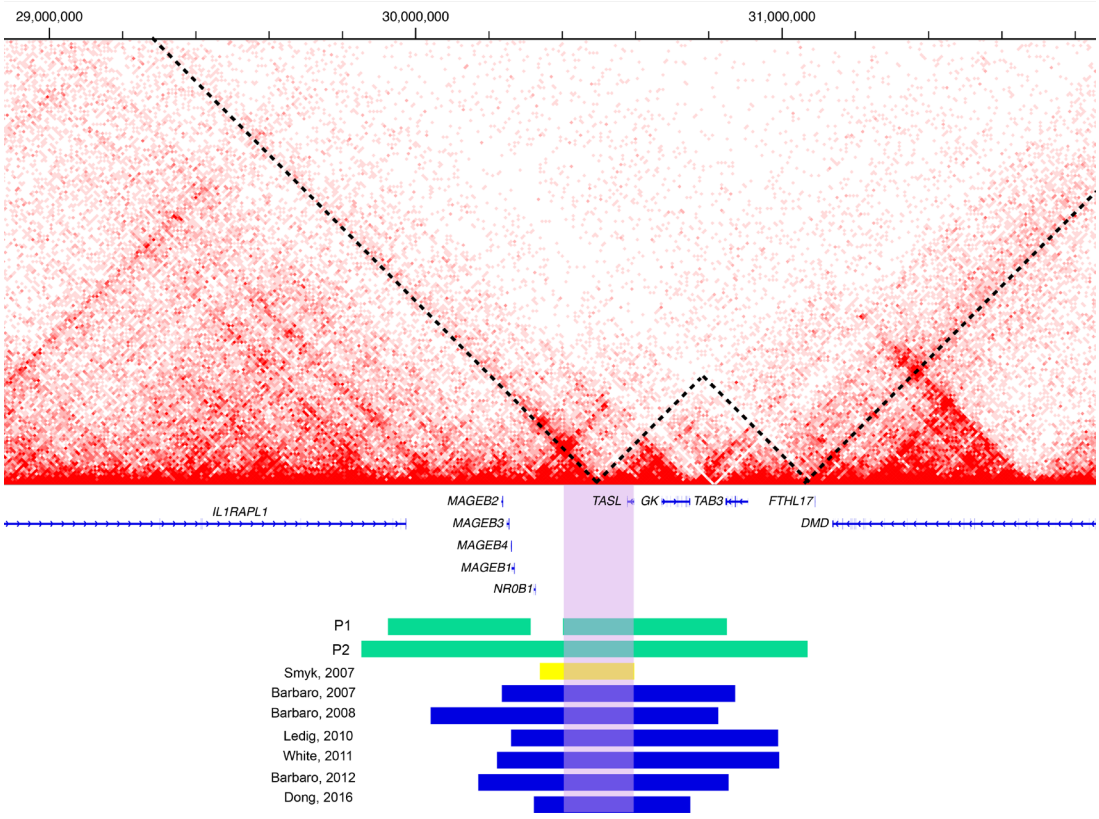

Figure 1 TAD structure at Xp21.2 and location of copy number variations (CNVs). Hi-C analysis identified a NR0B1 TAD containing IL1RAPL1, MAGEB1-4 and NR0B1, followed by a second TAD including TASL, GK and TAB3. A third TAD contains the FTHL17 and DMD genes (dashed lines). Below a schematic representation compares NR0B1 locus copy number variations in patients presented in this study with those previously described in the literature and associated with 46,XY gonadal dysgenesis. The green bars represent the duplications and the triplication described in the patients of this study (P1 – P2), and the blue bars represent duplications previously reported by other researchers.[4–10] All previously reported duplications include the NR0B1, TASL (CXorf21) and GK gene. The yellow bar marks the deletion reported by Smyk *et al*.[16] There is a mutual 195 kb region present in all Xp21.2 copy number variations associated with 46,XY GD, highlighted by the purple box. Hi-C, high throughput chromosome conformation capture; TAD, topological-associated domain.

missense variant (dbSNP: rs367855747) in *oestrogen receptor 2* (*ESR2*), with a clinical significance not discovered so far. However, recently homozygous and heterozygous *ESR2* mutations have been described in the context of 46,XY partial and complete GD.[32] The effect of this particular heterozygous SNV remains unclear, but in the context of widely agreed association of *NR0B1* copy number gains and 46,XY DSD, this remains a subordinate factor in the aetiology, if at all.

Analysis by qPCR, aCGH (online supplemental figure S5) and Sanger sequencing of the patient's mother identified her to carry a tandem duplication similar in size to P2s triplication. The mother agreed to donate fibroblasts, which were subsequently used for Hi-C analysis.

Hi-C analysis of P1 and P2

We performed Hi-C in patient fibroblast and LCLs to investigate the effects of the SVs on chromatin structure and 3D genome organisation. Hi-C maps of the *NR0B1 locus* revealed TADs delimited by its boundaries and insulating genes in chromatin domains (figure 1). The *NR0B1* TAD contains *IL1RAPL1*, *MAGEB1-4* and *NR0B1*. The contiguous TAD includes *TASL*, *GK* and TAB3. While the third TAD (centromeric direction) contains the *FTHL17* and *DMD* (figure 1). In P1 LCLs, Hi-C maps shows duplications as intense signal rising from the bottom and deletions as a loss of signal and white V-shape in between the duplications (figure 2). More importantly, the inverted fragment showed strong interactions between *NR0B1* and the region upstream of *TASL* indicating ectopic contacts with potential enhancer elements (marked in a circle in figure 2). The Hi-C data showed the formation of a novel chromatin domain (neo-TAD).

In figure 3, P2s mother's fibroblast Hi-C displayed a 1.2 Mb tandem duplication, explained by the intense interaction between the beginning and end of the duplicated region. The map clearly shows the formation of a neo-TAD where ectopic pathogenic interactions could take place, as reported in other cases.[33 34]

Form these Hi-C data, it became evident that both our cases show a minimal overlap of 195 kb (chrX: 30 401 819–30 596 386; GRCh37/hg19) harbouring a TAD boundary (chrX: 30 510 000–30 530 000; GRCh37/hg19 figure 1). The duplication of this boundary localises *NR0B1* in the vicinity of the genes and regulatory elements of the neighbouring TAD. To investigate the potential effect of all previously described SVs at the Xp 21.2 locus we modelled the effect on 3D genome architecture and TAD architecture. Remarkably, all previously reported duplications[4–9] (online supplemental figure S6), the deletion by Smyk *et al*,[16] and our two cases show an overlapping minimal critical region not including *NR0B1* but the TAD boundary next to it (figure 1). These data support the hypothesis that instead of gene dosage, enhancer hijacking could be the underlying causes of GD in patients with 46,XY karyotype and Xp21.2 SVs.

DISCUSSION

In our study, we report on the first 46,XY GD patient with Xp21.2 duplications, excluding *NR0B1*, the strongest candidate

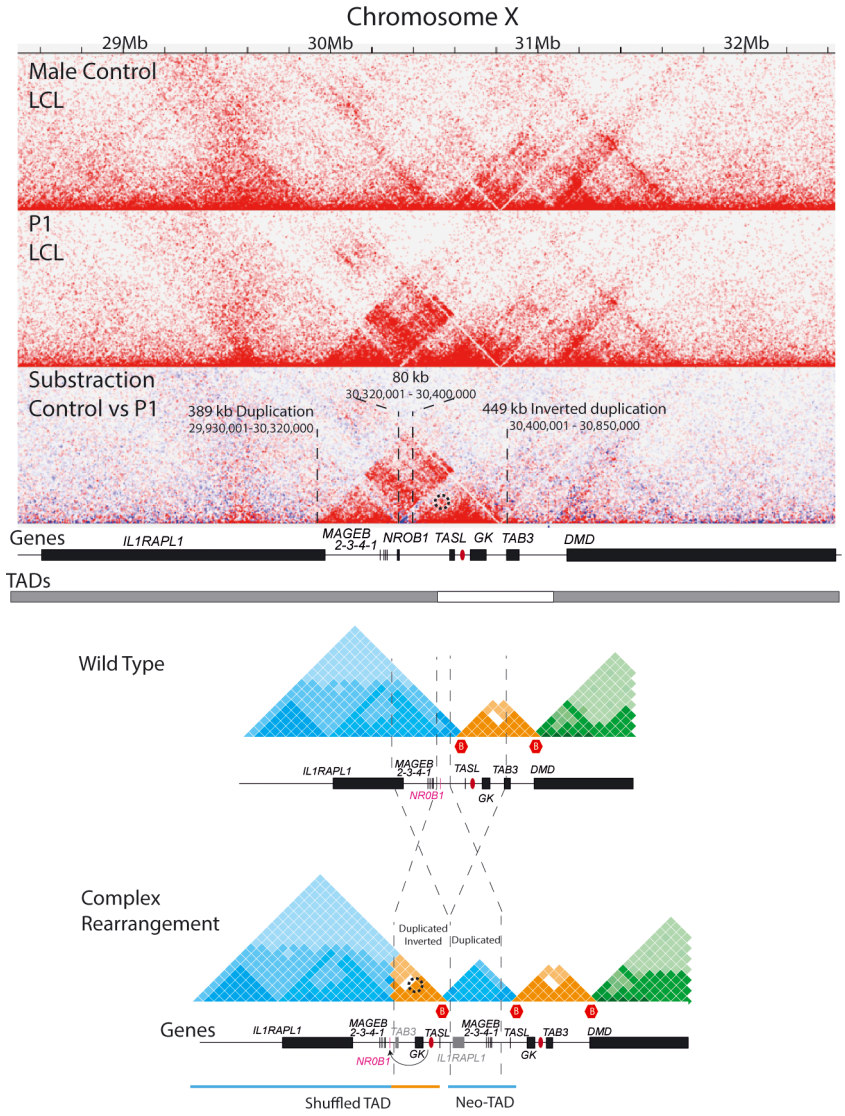

Figure 2 Hi-C analysis and representation of structural variation in P1. On the upper panel, the Hi-C maps from a male control LCLs and P1 patient's LCLs (10 kb resolution), shows the 3D architecture of the Xp 21.2 locus in absence and presence of SVs, respectively. The subtraction map from control versus P1 revealed the contact differences. Note the interaction between upstream NR0B1 and upstream TASL, marked with a dashed circumference. Genes (black boxes; hg19) are displayed at the bottom as well as the TADs organisation represented as grey-white bars on the track below. At the lower panel, there is a representation of the wild type TADs and the outcome after the complex rearrangement: one shuffled TAD and one neo-TAD. The genes drawn as grey squares were not duplicated completely (TAB3 and IL1RAPL). The potential enhancer is represented as a red oval, and its interaction with the gene is exhibited with an arrow. Hi-C, high throughput chromosome conformation capture; SVs, structural variations TAD, topological-associated domain.

gene related to the phenotype. This case presents a complex duplication and inversion at the Xp21.2 locus. Hi-C data show that the duplication and the inversion include a TAD boundary and lead to the formation of a neo-TAD and ectopic chromatin contact between *NR0B1* and its neighbouring domain including *TASL* and several predicted enhancer elements. Our data, together with the recent report of a deletion,[16] of this TAD boundary questioned that gene dosage effects of *NR0B1* alone are the underlying disease mechanism of this unique cause of GD. Instead, we propose that the duplications, the inversion and the deletion all lead to the rearrangement of a TAD boundary and result in enhancer hijacking between *NR0B1* and several predicted enhancer elements in the *TASL* TAD (online supplemental figure S7).

Our findings are in line with several recent studies showing that SVs can rewire the complex 3D chromatin architecture of a locus by deleting or repositioning regulatory elements and/or TAD boundaries, leading to ectopic enhancer–promoter interactions and ultimately pathogenic effects on development,[17 19 20 33 35–37] and oncogenesis.[21 22] To our knowledge, we performed the first Hi-C analysis of Xp21.2 SVs and were able to show how SVs altered TAD architectures when compared with normal healthy controls. In normal controls, *NR0B1* is in a separate TAD from neighbouring *TASL* and *GK* with their respective promoter and predicted enhancer regions (online supplemental figure S7). This TAD boundary is highly conserved across many different cell lines (online supplemental figure S8). Disrupting the TAD boundary can shuffle genes and their specific regulatory elements and potentially allowing enhancer adoption of the *TASL* and *GK* enhancers by *NR0B1*.

Enhancer hijacking of at the *NR0B1* locus could lead to upregulation or aberrant spatial-temporal expression of *NR0B1*, which in consequence would result in decreased SF1-mediated *SOX9* expression and impaired Sertoli cell differentiation and

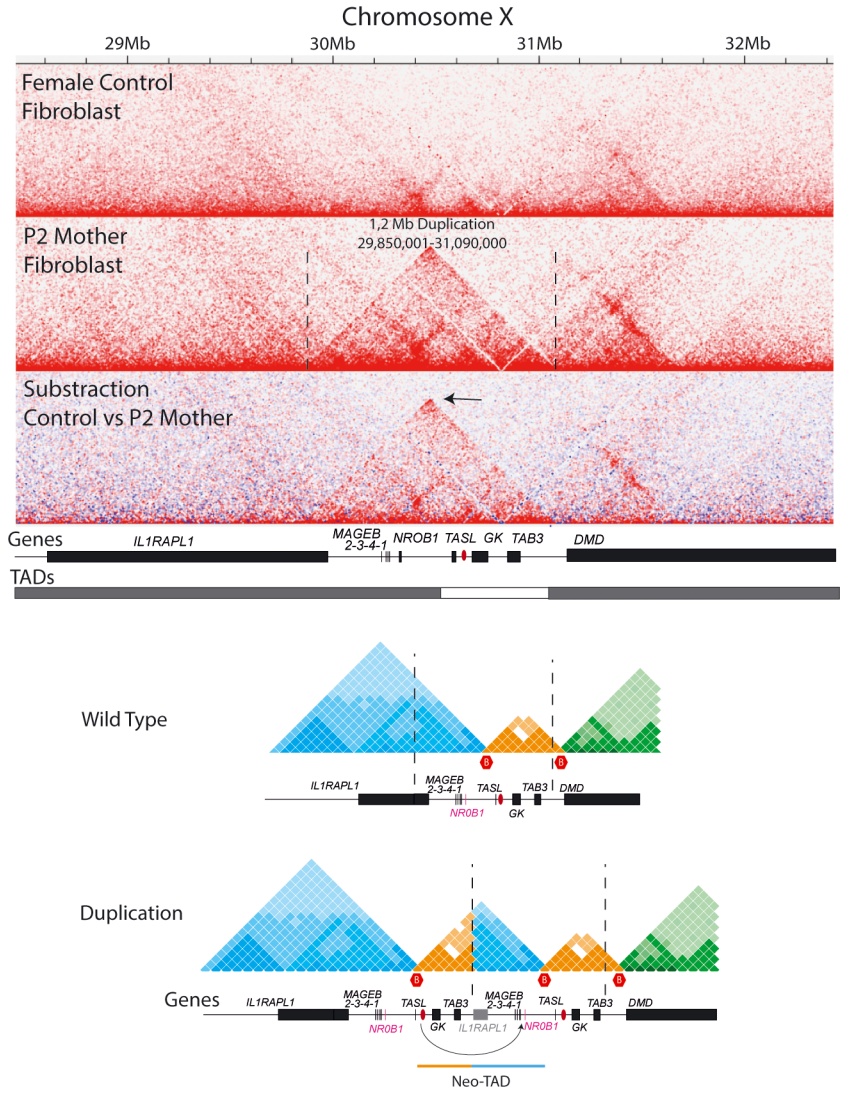

Figure 3 Hi-C analysis and representation of structural variation in mother of P2. Hi-C map from a female control and the P2 mother's fibroblasts with a 1.2 Mb duplication. The Hi-C subtraction between control and P2 mother indicates the increase in contacts within the duplication. The arrow in the subtraction map indicates the interaction between beginning and end of the duplication, which indicates this SV is in tandem. Below there are two tracks, the first displays the genes and the second the TADs organisation. In the schematic representation, its depicted that the duplication creates a neo-TAD containing: TASL, GK, TAB3, IL1RAPL1 partially, MAGEB1-4 and NR0B1. The potential enhancer (red oval) and non-cognate gene (NR0B1) interaction is shown with an arrow connection. Hi-C, high-throughput chromosome conformation capture; TADs, topological-associated domain.

hindering successful testis development.[38] This could be a shared disease mechanism among all recently published Xp21.2 CNVs,[4–7 9 16 39] in the context of GD as they share a common 195 kb overlap region crossing the TAD boundary causing the formation of neo and shuffled TADs.

Nevertheless, further research is needed to definitively identify the enhancer regions that may be hijacked by *NR0B1* after TAD disruption. Besides, in the selected region of TAD boundary, there are several CTCF binding sites, so it will be worth narrowing down the region for instance by inserting them in reverse orientation and evaluating TADs alterations. This is based in the knowledge that convergent CTCF sites are pairing at TAD boundaries. It would also be worth uncovering the exact genomic insertion of the other reported SVs at this locus, as well as more Hi-C data from GD patients to prove the truly unifying nature of this mechanism. Furthermore, it should be noted that this TAD disruption may also affect the *MAGEB1-4* genes, which are predominantly expressed in germ cells. A synergistic effect of *NR0B1* and *MAGEB1-4* dysregulation in the aetiology of 46,XY GD cannot be excluded.

In both patients, the SVs were maternally inherited, and P2 has healthy siblings. Maternal inheritance of an Xp21.2 duplication in GD is well known. No effects of reproductive development and function were reported so far. *NR0B1* duplications do not seem to impair ovarian function.[5 6 40 41]

Despite considerable advances in the understanding of sex development, the genetic aetiology of many 46,XY GD patients still remains unclear.[2 42] Our research emphasises that, besides unidentified genes in the gonadal developmental pathway, this may be due to neglected non-coding and/or regulatory elements, which explain the molecular basis of the GD phenotype. This highlights the relevance of identifying the position and orientation of SVs to deduce new enhancer-genes arrangements and improve our knowledge in genotype–phenotype relationships in this rare disease.

Our data further show the potential diagnostic power of whole genome sequencing to deliver detailed data on non-coding regions alterations, which allows accurate breakpoint identification of SVs through interrogation of individual reads. However, Hi-C has proven to be a sensitive tool for SVs detection in the clinical setting,[34] and in our case, aided in TAD recognition in the target locus and guide our prediction on the outcome of the SVs in a 3D context.

Author affiliations
[1]Department of Pediatrics and Adolescent Medicine, Division of Pediatric Endocrinology and Diabetes, Universität zu Lübeck, Lubeck, Germany
[2]Institute of Human Genetics, Universität zu Lübeck, Lübeck, Germany
[3]Group of Medical Systems Biology, Lübeck Institute of Experimental Dermatology, Universität zu Lübeck, Lübeck, Germany
[4]Institute for Cardiogenetics, Universität zu Lübeck, Lübeck, Germany
[5]Institute of Human Genetics, Universität Duisburg-Essen, Duisburg, Germany
[6]Essen Center for Rare Diseases (EZSE), University Hospital Essen, Essen, Germany
[7]University Medical Center for Pediatric Endocrinology and Diabetes, Department of Pediatrics and Adolescent Medicine I, Universitätsklinikum Schleswig-Holstein, Kiel, Germany
[8]Department of Pediatrics and Adolescent Medicine I, Division of Pediatric Oncology and Hematology, Universitätsklinikum Schleswig-Holstein, Kiel, Germany
[9]Partner Site Hamburg/Kiel/Lübeck, German Center for Cardiovascular Disease, Berlin, Germany
[10]Institute of Molecular Medicine, Universität zu Lübeck, Lübeck, Germany

Correction notice This article has been corrected since it was first published to indicate co-first authorship. The license has also been updated to CC BY.

Acknowledgements The authors are truly grateful to the patients and families of these patients who cooperated in this study. HB and AK acknowledge computational support from the Lübeck OmicsCluster. This work is part of the doctoral thesis of JAM.

Contributors Conceptualisation: RW; conceived model: JAM, VY, MS and RW; manuscript draft: JAM, VY, MS and RW; editing and review of manuscript: FJK, P-MH, AC, OH, HB, MS and RW; funding acquisition: OH and HB; experiments/data generation: JAM, VY and RW; bioinformatics: AK, HB, KS and NK; figures and visualisations: JAM, VY and RW; patient recruitment: PMH, OH and AC; guarantor: RW.

Funding This work was funded by financial support from Bundesministerium für Bildung und Forschung BMBF (01DQ17004) and Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy (EXC 22167-390884018).

Disclaimer The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and was approved by Ethics Committee of the Universitiy of Lübeck ID: AZ 08-081, Investigation of the molecular pathogenesis and pathophysiology of disorders of sex development. ID: AZ 17-219. Ethics Committee of the Christian-Albrechts-Universität zu Kiel ID: D 410/08. Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as supplementary information.

ORCID iDs
Verónica Yumiceba http://orcid.org/0000-0001-6998-7913
Axel Künstner http://orcid.org/0000-0003-0692-2105
Alexander Claviez http://orcid.org/0000-0001-7064-8556
Hauke Busch http://orcid.org/0000-0003-4763-4521
Malte Spielmann http://orcid.org/0000-0002-0583-4683
Ralf Werner http://orcid.org/0000-0003-3718-3595

REFERENCES

1 Eggers S, Sadedin S, van den Bergen JA, Robevska G, Ohnesorg T, Hewitt J, Lambeth L, Bouty A, Knarston IM, Tan TY, Cameron F, Werther G, Hutson J, O'Connell M, Grover SR, Heloury Y, Zacharin M, Bergman P, Kimber C, Brown J, Webb N, Hunter MF, Srinivasan S, Titmuss A, Verge CF, Mowat D, Smith G, Smith J, Ewans L, Shalhoub C, Crock P, Cowell C, Leong GM, Ono M, Lafferty AR, Huynh T, Visser U, Choong CS, McKenzie F, Pachter N, Thompson EM, Couper J, Baxendale A, Gecz J, Wheeler BJ, Jefferies C, MacKenzie K, Hofman P, Carter P, King RI, Krausz C, van Ravenswaaij-Arts CMA, Looijenga L, Drop S, Riedl S, Cools M, Dawson A, Juniarto AZ, Khadilkar V, Khadilkar A, Bhatia V, Dũng VC, Atta I, Raza J, Thi Diem Chi N, Hao TK, Harley V, Koopman P, Warne G, Faradz S, Oshlack A, Ayers KL, Sinclair AH. Disorders of sex development: insights from targeted gene sequencing of a large international patient cohort. *Genome Biol* 2016;17:243.
2 Audí L, Ahmed SF, Krone N, Cools M, McElreavey K, Holterhus PM, Greenfield A, Bashamboo A, Hiort O, Wudy SA, McGowan R, The EU COST Action. GENETICS IN ENDOCRINOLOGY: Approaches to molecular genetic diagnosis in the management of differences/disorders of sex development (DSD): position paper of EU COST Action BM 1303 'DSDnet'. *Eur J Endocrinol* 2018;179:R197–206.
3 Bardoni B, Zanaria E, Guioli S, Floridia G, Worley KC, Tonini G, Ferrante E, Chiumello G, McCabe ER, Fraccaro M. A dosage sensitive locus at chromosome Xp21 is involved in male to female sex reversal. *Nat Genet* 1994;7:497–501.
4 Barbaro M, Cicognani A, Balsamo A, Löfgren A, Baldazzi L, Wedell A, Oscarson M. Gene dosage imbalances in patients with 46,XY gonadal DSD detected by an in-house-designed synthetic probe set for multiplex ligation-dependent probe amplification analysis. *Clin Genet* 2008;73:453–64.
5 Barbaro M, Cook J, Lagerstedt-Robinson K, Wedell A. Multigeneration inheritance through fertile XX carriers of an NR0B1 (Dax1) locus duplication in a kindred of females with isolated XY gonadal dysgenesis. *Int J Endocrinol* 2012;2012:1–7.
6 Barbaro M, Oscarson M, Schoumans J, Staaf J, Ivarsson SA, Wedell A. Isolated 46,XY gonadal dysgenesis in two sisters caused by a Xp21.2 interstitial duplication containing the DAX1 gene. *J Clin Endocrinol Metab* 2007;92:3305–13.
7 Dong Y, Yi Y, Yao H, Yang Z, Hu H, Liu J, Gao C, Zhang M, Zhou L, Yi X, Liang Z. Targeted next-generation sequencing identification of mutations in patients with disorders of sex development. *BMC Med Genet* 2016;17:23.
8 Ledig S, Hiort O, Scherer G, Hoffmann M, Wolff G, Morlot S, Kuechler A, Wieacker P. Array-CGH analysis in patients with syndromic and non-syndromic XY gonadal dysgenesis: evaluation of array CGH as diagnostic tool and search for new candidate loci. *Human Reproduction* 2010;25:2637–46.
9 White S, Ohnesorg T, Notini A, Roeszler K, Hewitt J, Daggag H, Smith C, Turbitt E, Gustin S, van den Bergen J, Miles D, Western P, Arboleda V, Schumacher V, Gordon L, Bell K, Bengtsson H, Speed T, Hutson J, Warne G, Harley V, Koopman P, Vilain E, Sinclair A. Copy number variation in patients with disorders of sex development due to 46,XY gonadal dysgenesis. *PLoS One* 2011;6:e17793.
10 García-Acero M, Molina M, Moreno O, Ramirez A, Forero C, Céspedes C, Prieto JC, Pérez J, Suárez-Obando F, Rojas A. Gene dosage of DAX-1, determining in sexual differentiation: duplication of DAX-1 in two sisters with gonadal dysgenesis. *Mol Biol Rep* 2019;46:2971–8.
11 Qin S, Wang X, Li Y. [Genetic analysis of a 46,XY female with sex reversal due to duplication of NR0B1 gene]. *Zhonghua Yi Xue Yi Chuan Xue Za Zhi* 2018;35:804–7.
12 Suntharalingham JP, Buonocore F, Duncan AJ, Achermann JC. DAX-1 (NR0B1) and steroidogenic factor-1 (SF-1, NR5A1) in human disease. *Best Pract Res Clin Endocrinol Metab* 2015;29:607–19.
13 Swain A, Zanaria E, Hacker A, Lovell-Badge R, Camerino G. Mouse Dax1 expression is consistent with a role in sex determination as well as in adrenal and hypothalamus function. *Nat Genet* 1996;12:404–9.
14 Stévant I, Neirijnck Y, Borel C, Escoffier J, Smith LB, Antonarakis SE, Dermitzakis ET, Nef S. Deciphering cell lineage specification during male sex determination with single-cell RNA sequencing. *Cell Rep* 2018;22:1589–99.
15 Swain A, Narvaez V, Burgoyne P, Camerino G, Lovell-Badge R. Dax1 antagonizes Sry action in mammalian sex determination. *Nature* 1998;391:761–7.
16 Smyk M, Berg JS, Pursley A, Curtis FK, Fernandez BA, Bien-Willner GA, Lupski JR, Cheung SW, Stankiewicz P. Male-To-Female sex reversal associated with an approximately 250 kb deletion upstream of NR0B1 (DAX1). *Hum Genet* 2007;122:63–70.
17 Ibn-Salem J, Köhler S, Love MI, Chung H-R, Huang N, Hurles ME, Haendel M, Washington NL, Smedley D, Mungall CJ, Lewis SE, Ott C-E, Bauer S, Schofield PN, Mundlos S, Spielmann M, Robinson PN. Deletions of chromosomal regulatory boundaries are associated with congenital disease. *Genome Biol* 2014;15:423.

18 Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 2012;485:376–80.
19 Lettice LA, Daniels S, Sweeney E, Venkataraman S, Devenney PS, Gautier P, Morrison H, Fantes J, Hill RE, FitzPatrick DR. Enhancer-adoption as a mechanism of human developmental disease. *Hum Mutat* 2011;32:1492–9.
20 Lupiáñez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, Horn D, Kayserili H, Opitz JM, Laxova R, Santos-Simarro F, Gilbert-Dussardier B, Wittler L, Borschiwer M, Haas SA, Osterwalder M, Franke M, Timmermann B, Hecht J, Spielmann M, Visel A, Mundlos S. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* 2015;161:1012–25.
21 Hnisz D, Weintraub AS, Day DS, Valton A-L, Bak RO, Li CH, Goldmann J, Lajoie BR, Fan ZP, Sigova AA, Reddy J, Borges-Rivera D, Lee TI, Jaenisch R, Porteus MH, Dekker J, Young RA. Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science* 2016;351:1454–8.
22 Weischenfeldt J, Dubash T, Drainas AP, Mardin BR, Chen Y, Stütz AM, Waszak SM, Bosco G, Halvorsen AR, Raeder B, Efthymiopoulos T, Erkek S, Siegl C, Brenner H, Brustugun OT, Dieter SM, Northcott PA, Petersen I, Pfister SM, Schneider M, Solberg SK, Thunissen E, Weichert W, Zichner T, Thomas R, Peifer M, Helland A, Ball CR, Jechlinger M, Sotillo R, Glimm H, Korbel JO. Pan-Cancer analysis of somatic copy-number alterations implicates IRS4 and IGF2 in enhancer hijacking. *Nat Genet* 2017;49:65–74.
23 Kulle AE, Riepe FG, Melchior D, Hiort O, Holterhus PM. A novel ultrapressure liquid chromatography tandem mass spectrometry method for the simultaneous determination of androstenedione, testosterone, and dihydrotestosterone in pediatric blood samples: age- and sex-specific reference data. *J Clin Endocrinol Metab* 2010;95:2399–409.
24 Rey RA, Belville C, Nihoul-Fékété C, Michel-Calemard L, Forest MG, Lahlou N, Jaubert F, Mowszowicz I, David M, Saka N, Bouvattier C, Bertrand AM, Lecointre C, Soskin S, Cabrol S, Crosnier H, Léger J, Lortat-Jacob S, Nicolino M, Rabl W, Toledo SP, Baş F, Gompel A, Czernichow P, Josso N. Evaluation of gonadal function in 107 intersex patients by means of serum antimüllerian hormone measurement. *J Clin Endocrinol Metab* 1999;84:627–31.
25 Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;25:1754–60.
26 DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011;43:491–8.
27 Rausch T, Zichner T, Schlattl A, Stütz AM, Benes V, Korbel JO. DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* 2012;28:i333–9.
28 Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010;38:e164.
29 Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, Aiden EL. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 2014;159:1665–80.
30 Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, Tukiainen T, Birnbaum DP, Kosmicki JA, Duncan LE, Estrada K, Zhao F, Zou J, Pierce-Hoffman E, Berghout J, Cooper DN, Deflaux N, DePristo M, Do R, Flannick J, Fromer M, Gauthier L, Goldstein J, Gupta N, Howrigan D, Kiezun A, Kurki MI, Moonshine AL, Natarajan P, Orozco L, Peloso GM, Poplin R, Rivas MA, Ruano-Rubio V, Rose SA, Ruderfer DM, Shakir K, Stenson PD, Stevens C, Thomas BP, Tiao G, Tusie-Luna MT, Weisburd B, Won H-H, Yu D, Altshuler DM, Ardissino D, Boehnke M, Danesh J, Donnelly S, Elosua R, Florez JC, Gabriel SB, Getz G, Glatt SJ, Hultman CM, Kathiresan S, Laakso M, McCarroll S, McCarthy MI, McGovern D, McPherson R, Neale BM, Palotie A, Purcell SM, Saleheen D, Scharf JM, Sklar P, Sullivan PF, Tuomilehto J, Tsuang MT, Watkins HC, Wilson JG, Daly MJ, MacArthur DG, Exome Aggregation Consortium. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016;536:285–91.
31 Berglund A, Johannsen TH, Stochholm K, Viuff MH, Fedder J, Main KM, Gravholt CH. Incidence, Prevalence, Diagnostic Delay, and Clinical Presentation of Female 46,XY Disorders of Sex Development. *J Clin Endocrinol Metab* 2016;101:4532–40.
32 Baetens D, Güran T, Mendonca BB, Gomes NL, De Cauwer L, Peelman F, Verdin H, Vuylsteke M, Van der Linden M, Atay Z, Bereket A, de Krijger RR, Preter Kde, Domenice S, Turan S, Stoop H, Looijenga LH, De Bosscher K, Cools M, De Baere E, ESR2 STUDY GROUP. Biallelic and monoallelic ESR2 variants associated with 46,XY disorders of sex development. *Genet Med* 2018;20:717–27.
33 Franke M, Ibrahim DM, Andrey G, Schwarzer W, Heinrich V, Schöpflin R, Kraft K, Kempfer R, Jerković I, Chan W-L, Spielmann M, Timmermann B, Wittler L, Kurth I, Cambiaso P, Zuffardi O, Houge G, Lambie L, Brancati F, Pombo A, Vingron M, Spitz F, Mundlos S. Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature* 2016;538:265–9.
34 Melo US, Schöpflin R, Acuna-Hidalgo R, Mensah MA, Fischer-Zirnsak B, Holtgrewe M, Klever M-K, Türkmen S, Heinrich V, Pluym ID, Matoso E, Bernardo de Sousa S, Louro P, Hülsemann W, Cohen M, Dufke A, Latos-Bieleńska A, Vingron M, Kalscheuer V, Quintero-Rivera F, Spielmann M, Mundlos S. Hi-C identifies complex genomic rearrangements and TAD-Shuffling in developmental diseases. *Am J Hum Genet* 2020;106:872–84.
35 Kragesteen BK, Spielmann M, Paliou C, Heinrich V, Schöpflin R, Esposito A, Annunziatella C, Bianco S, Chiariello AM, Jerković I, Harabula I, Guckelberger P, Pechstein M, Wittler L, Chan W-L, Franke M, Lupiáñez DG, Kraft K, Timmermann B, Vingron M, Visel A, Nicodemi M, Mundlos S, Andrey G. Dynamic 3D chromatin architecture contributes to enhancer specificity and limb morphogenesis. *Nat Genet* 2018;50:1463–73.
36 Spielmann M, Lupiáñez DG, Mundlos S. Structural variation in the 3D genome. *Nat Rev Genet* 2018;19:453–67.
37 Kraft K, Magg A, Heinrich V, Riemenschneider C, Schöpflin R, Markowski J, Ibrahim DM, Acuna-Hidalgo R, Despang A, Andrey G, Wittler L, Timmermann B, Vingron M, Mundlos S. Serial genomic inversions induce tissue-specific architectural stripes, gene misexpression and congenital malformations. *Nat Cell Biol* 2019;21:305–10.
38 Chaboissier M-C, Kobayashi A, Vidal VIP, Lützkendorf S, van de Kant HJG, Wegner M, de Rooij DG, Behringer RR, Schedl A. Functional analysis of Sox8 and SOX9 during sex determination in the mouse. *Development* 2004;131:1891–901.
39 Ledig S, Hiort O, Scherer G, Hoffmann M, Wolff G, Morlot S, Kuechler A, Wieacker P. Array-CGH analysis in patients with syndromic and non-syndromic XY gonadal dysgenesis: evaluation of array CGH as diagnostic tool and search for new candidate loci. *Hum Reprod* 2010;25:2637–46.
40 Baumstark A, Barbi G, Djalali M, Geerkens C, Mitulla B, Mattfeldt T, de Almeida JC, Vargas FR, Llerena Júnior JC, Vogel W, Just W. Xp-duplications with and without sex reversal. *Hum Genet* 1996;97:79–86.
41 Barbaro M, Bens S, Haake A, Peter M, Brämswig J, Holterhus P-M, Lopez-Siguero JP, Menken U, Mix M, Sippell WG, Wedell A, Riepe FG. Multiplex ligation-dependent probe amplification analysis of the NR0B1(DAX1) locus enables explanation of phenotypic differences in patients with X-linked congenital adrenal hypoplasia. *Horm Res Paediatr* 2012;77:100–7.
42 Bashamboo A, Eozenou C, Rojo S, McElreavey K. Anomalies in human sex determination provide unique insights into the complex genetic interactions of early gonad development. *Clin Genet* 2017;91:143–56.

Online Supplemental

Methods

CNV screening

A qPCR approach was used to identifying further patients with CNVs at Xp21.2. 168 Patients with 46,XY DSD of unknown molecular genetic origin were studied. qPCRs were performed on a LightCycler 96 (Roche, Basel, Switzerland) with a CYBR-Green system using the Takyon No Rox SYBR MasterMix dTTP Blue (Eurogentec, Seraing, Belgium). Six primer pairs were designed to cover both the up- and downstream regions of *NR0B1* and located in the exons of *IL1RAPL1, MAGEB1, NR0B1, TASL* (also known as *CXorf21*), *GK* and *TAB3*. Primer validation consisted of standard curves for efficiency calculation (2.00 ± 0.10), melting curve analysis and gel electrophoresis of PCR products. Fertile males and females were included in each run as positive and negative controls for the x-chromosomal locus. *ZNF80* and *GPR15* were used as reference genes with primers published in RT-PrimerDB (ID 1021 and 1022, respectively) (Pattyn et al., 2003).

Breakpoint Sequencing

Primers at either end of the constructed breakpoint sequences were designed using Lasergene PrimerSelect (DNASTAR, Wisconsin, USA). Optimal annealing temperature was determined through gradient PCR and electrophoresis. Consequently, PCR setup was 35 cycles with 30 sec. at 95°C for denaturation, 30 sec. at primer specific temperature (Table S2, Supporting Information) for annealing and 1 min. extension at 72°C. Sanger Sequencing of amplicons was performed on a 3130 Genetic Analyzer (Applied Biosystems, Foster City, USA).

Meinel JA, *et al. J Med Genet* 2022;0:1–8. doi: 10.1136/jmg-2022-108635

Supporting Figures

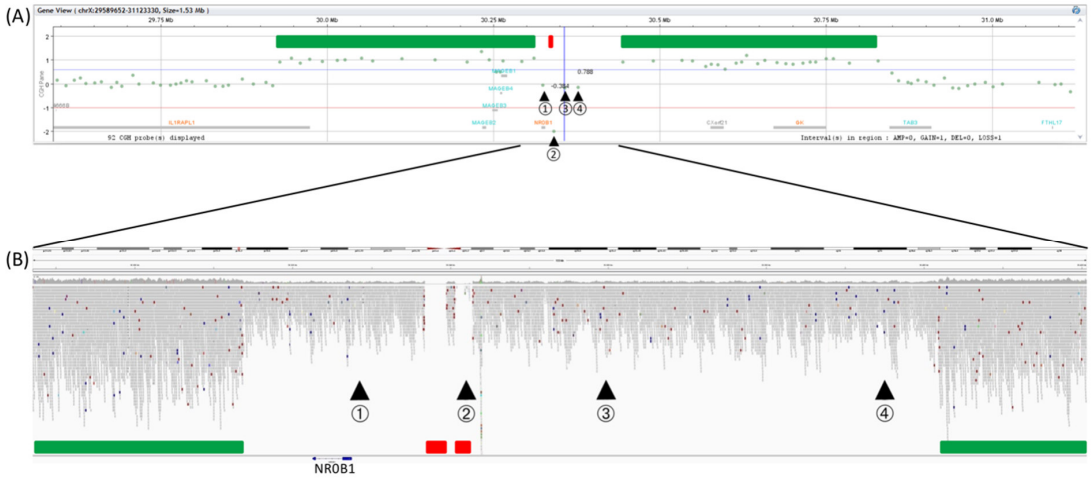

Figure S1. aCGH and GS comparison in P1 at *NR0B1* locus (A) aCGH results for P1. Green dots indicate probes used for hybridization. A pooled sample of 10 normal 46,XY males was used as control. Results show two separate duplications as illustrated by increased hybridization of probes and emphasized by green bars. To the left a 389kb duplication (chrX:29,924,420-30,313,761; GRCh37/hg19) was detected by increased hybridization of 17 probes. The second duplication was detected as 447kb of size (chrX: 30,401,819-30,848,988; GRCh37/hg19) indicated by 23 probes. The duplications are separated by four probes labelled 1 – 4. Probes 1, 3 and 4 hybridized only to a single copy. Probe 2 showed no hybridization. (B) Probes 1 to 4 have been compared to genome sequencing data. For visualisation of WGS data of P2 (region ChrX:30,292,562-30,423,695; GRCh37/hg19) the Integrative Genome Viewer (IGV) was used (Robinson et al., 2011). Increased number of reads (grey bars) at both sides of the viewing window signify duplication detected by WGS (green bars). Deletions are evident through the lack of reads in the two regions marked with red bars. Single hybridized CGH probes 1, 3 and 4 map to areas not duplicated according to WGS data. Probe 2 maps to a deletion and correspondingly did not hybridize. Note that the aCGH could be easily misinterpreted as one large duplication, if the different hybridisation of probes 1-4 is not recognized.

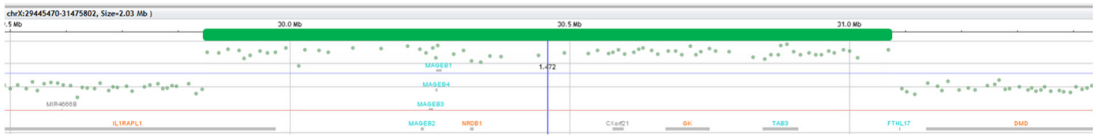

Figure S2. aCGH in P2 at *NR0B1* locus. Green dots indicate probes used for hybridization. A pooled sample of 10 normal 46,XY males was used as control. Results show a 1,2 Mb triplication (ChrX:29,851,537-31,069,736; GRCh37/hg19) emphasized by green bar. The genes *MAGEB1-4*, *NR0B1, TASL* (also known as *CXorf21*) and *TAB3* are fully triplicated whereas *IL1RAPL1* is only partially triplicated.

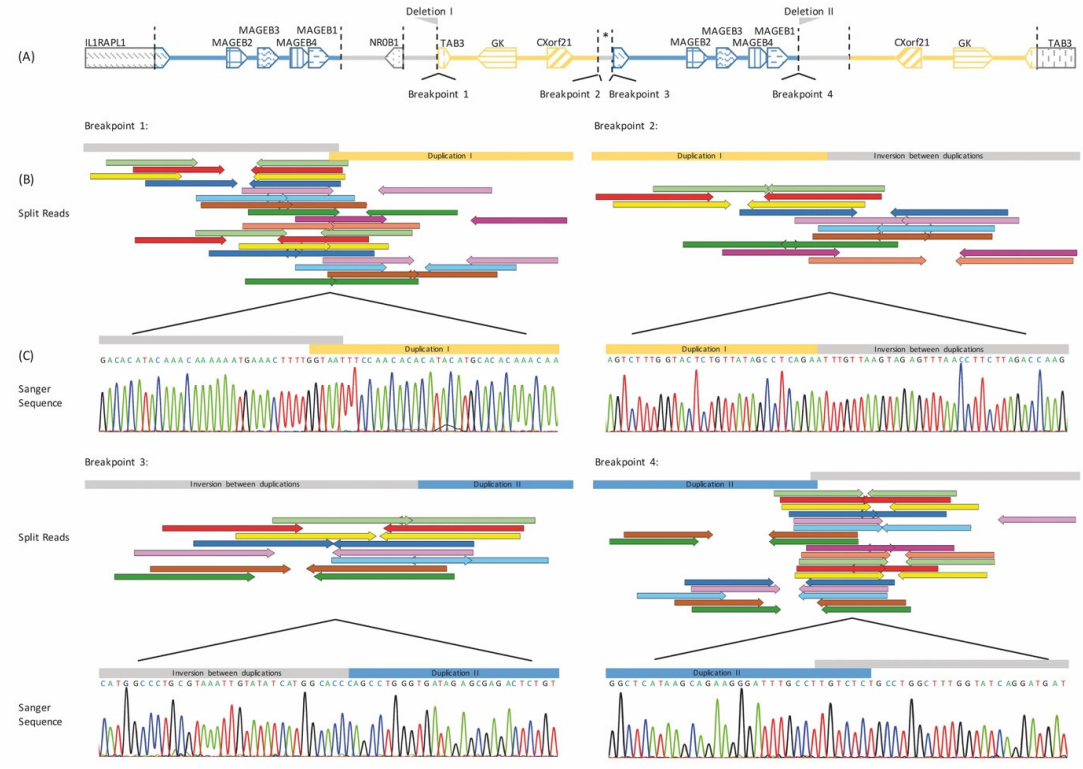

Figure S3. Structural Variation of Patient 1 (A) Overview of the structural variation upstream of the *NR0B1* gene. The genomic distances are not to scale. A 447kb duplication of *TASL* (*CXorf21)*, *GK* and part of *TAB3* (yellow) originally mapping to 30,401,819-30,848,988 is inserted in an inverted manner between breakpoint 1 (bp1; chrX:30,336,745; GRCh37/hg19) and bp2 (chrX: 30,340,605; GRCh37/hg19) 9.25kb upstream of the *NR0B1* reading frame. * marks a 1.2kb piece of reference sequence inverted between bp2 and bp3 separating the two large duplications. Both duplications are flanked by deletions indicated by the grey flags denoted *Deletion I & II*. A 389kb duplication of the *MAGEB* genes 1-4 and part of *IL2RAPL2* (blue) originally mapping to chrX: 29,924,420-30,313,761 (GRCh37/hg19) is inserted between breakpoint 3 (chrX: 30,339,452; GRCh37/hg19) and 4 (chrX: 30,342,785; GRCh37/hg19) in the same orientation as its reference sequence. (B) Shows position of extracted and aligned WGS split reads crossing the four breakpoints. Split reads mapped at varying distances apart for each breakpoint. Each pair was separated by at least >29kb. Sequences at either side of bp1 and bp4 showed homology, thus no exact definition of the breakpoint position was

possible, as depicted by the overlap of the grey and yellow or blue bars. (C) Shows verification of sequences deduced from WGS reads through Sanger sequencing.

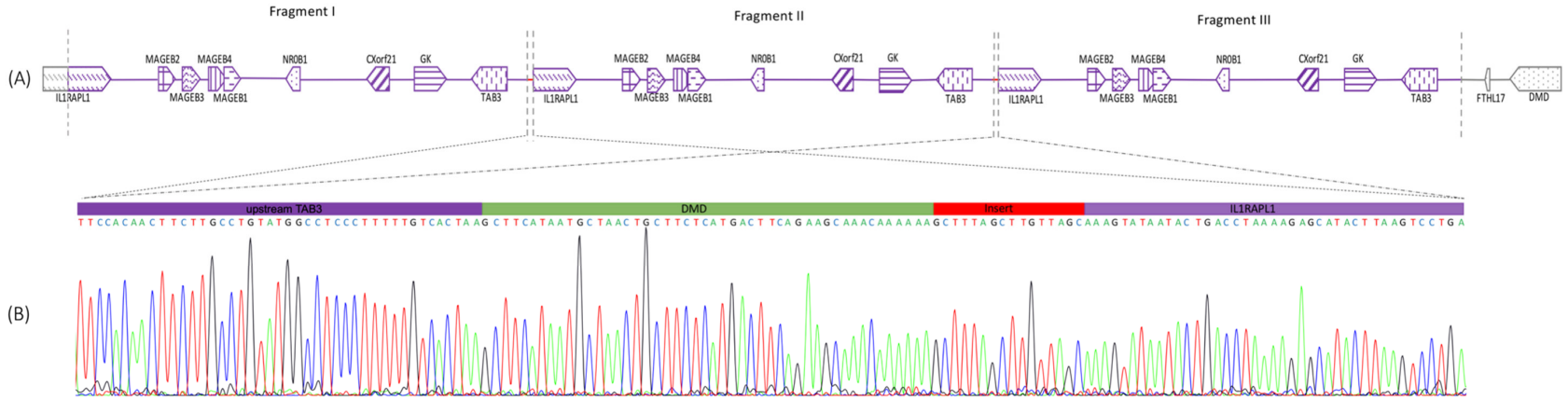

Figure S4. Structural Variation of Patient 2. (A) Overview of the structural variation at Xp21.2 locus. Genomic distances are not to scale. The chromosome segment is drawn with the distal chromosome arm to the left and centromere to the right. WGS showed the triplication originally mapped to chrX: 29,849,782 – 31,088,713 (GRCh37/hg19) and is arranged in a tandem manner. The triplicated fragments are connected by two identical 49 bp inserts at the breakpoints. This sequence originates from an intron of the *DMD* gene, located to the centromere of the SV event. *CXorf21* corresponds to *TASL* (B)Verification of sequences deduced from WGS reads through Sanger sequencing.

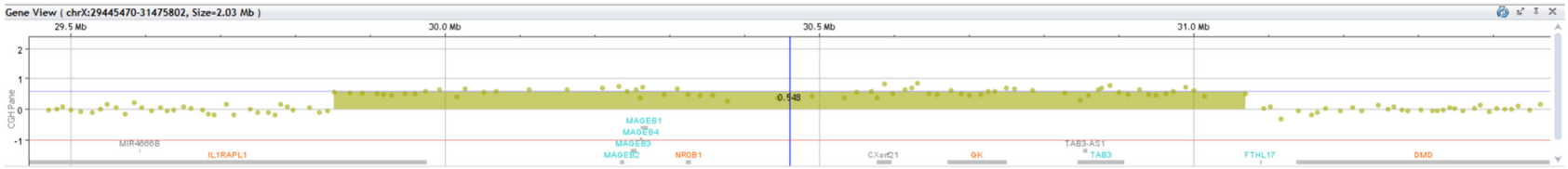

Figure S5. aCGH of mother of P2 at *NR0B1* locus. Green dots indicate probes used for hybridization. A pooled sample of 10 normal 46,XX females was used as control. Results show a 1,2 Mb duplication (ChrX:29,851,537-31,069,736; GRCh37/hg19) in a 46,XX karyotype. The size of the copy number gain corresponds to that of P2 i.e. the genes *MAGEB1-4*, *NR0B1, TASL (CXorf21)* and *TAB3* are fully duplicated whereas *IL1RAPL1* is only partially duplicated.

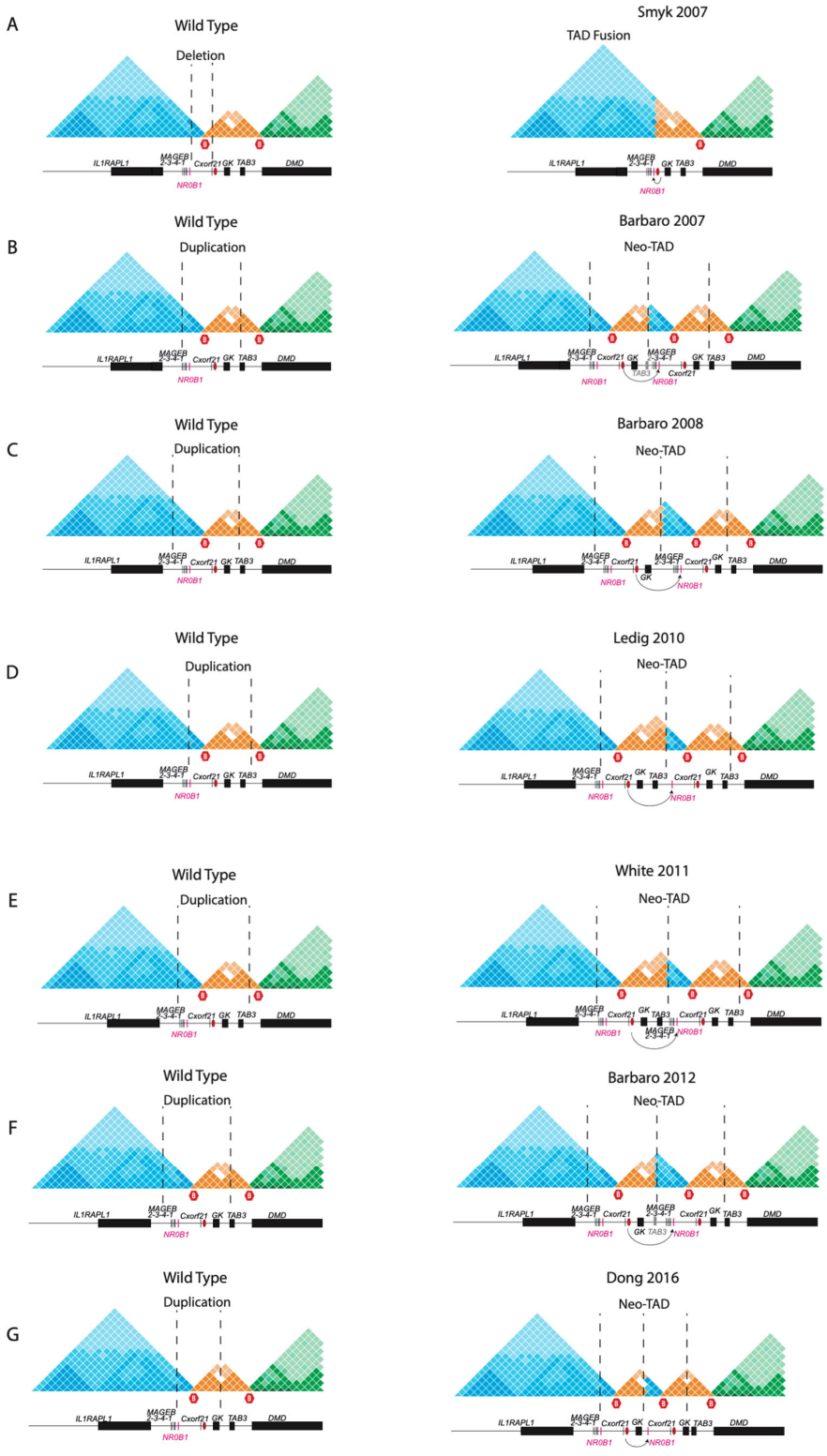
A
Wild Type
Deletion
Smyk 2007
TAD Fusion
B
Wild Type
Duplication
Barbaro 2007
Neo-TAD
C
Wild Type
Duplication
Barbaro 2008
Neo-TAD
D
Wild Type
Duplication
Ledig 2010
Neo-TAD
E
Wild Type
Duplication
White 2011
Neo-TAD
F
Wild Type
Duplication
Barbaro 2012
Neo-TAD
G
Wild Type
Duplication
Dong 2016
Neo-TAD
IL1RAPL1
MAGEB 2-3-4-1
Cxorf21
GK
TAB3
DMD
NR0B1

Figure S6. Proposed effects of inter-TAD deletions and duplications on Xp21.2 chromatin organisation.

Deletions (A) and duplications (B-F) at Xp21.2 are associated with gonadal dysgenesis and their chromatin landscape was modelled based on the TAD structure observed in our study. Since not for all duplications the exact breakpoints have been determined, tandem duplications have been assumed for all. The deletion (A) (Smyk et al., 2007) and all Inter-TAD duplications (B-F) (Barbaro et al., 2008; Barbaro et al., 2012; Barbaro et al., 2007; Dong et al., 2016; Ledig et al., 2010; White et al., 2011) containing *NR0B1* upstream boundary result in the incorporation of *NR0B1* into new chromatin domains (neo-TADs) with a novel regulatory landscape that can result in *NR0B1* upregulation. Red oval represents enhancers between *TASL* (*CXorf21)* and *GK*. Partially duplicated genes are in in light grey.

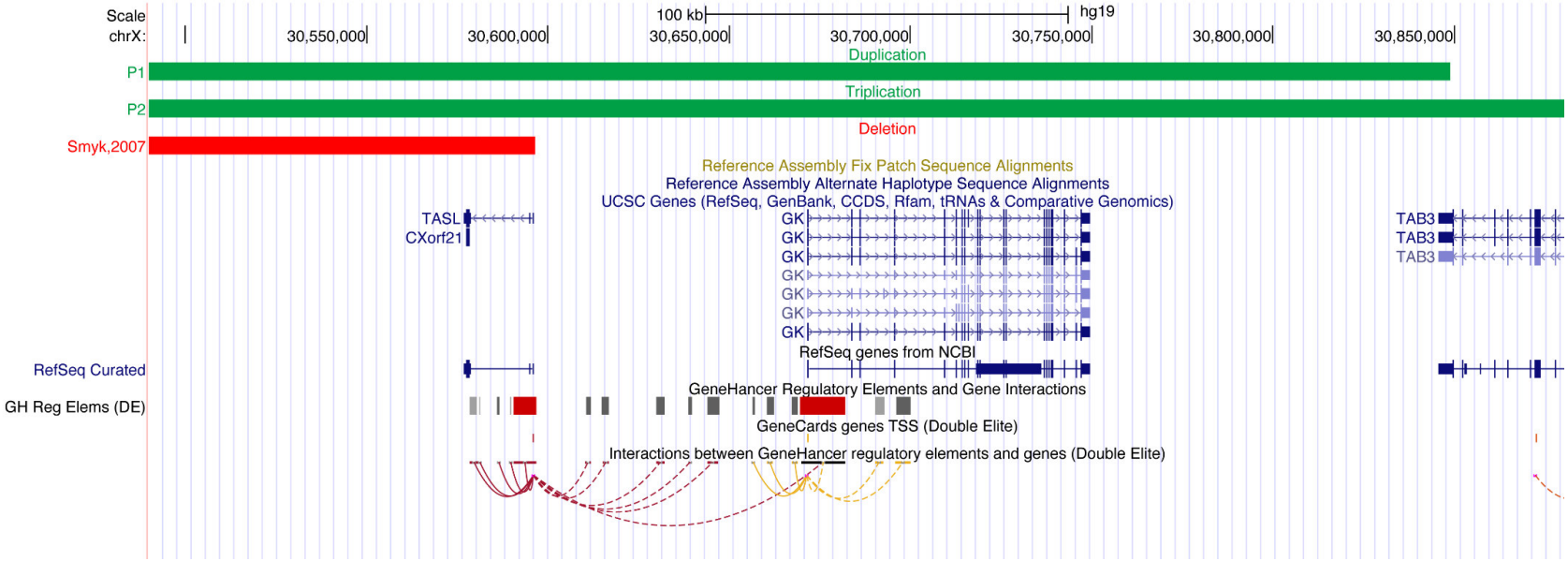

Figure S7. GeneHancer Promoters and Enhancers upstream of *NR0B1*. Data from the GeneHancer database (Fishilevich et al., 2017) was visualized in the UCSC genome browser (Kent et al., 2002). Promoters and enhancers are distinguished by red and grey colours respectively in the *GH Reg Elements (DE)* track. Colour intensity shows confidence score of the respective enhancers. According to Fishilevich et al. this is based on: "the number of supporting data sources, the number of unique TFBSs contained, and the overlap with ultra-conserved non-coding genomic elements".(Fishilevich et al., 2017) Arcs in this tack show proposed gene-enhancer interaction in the reference genome.

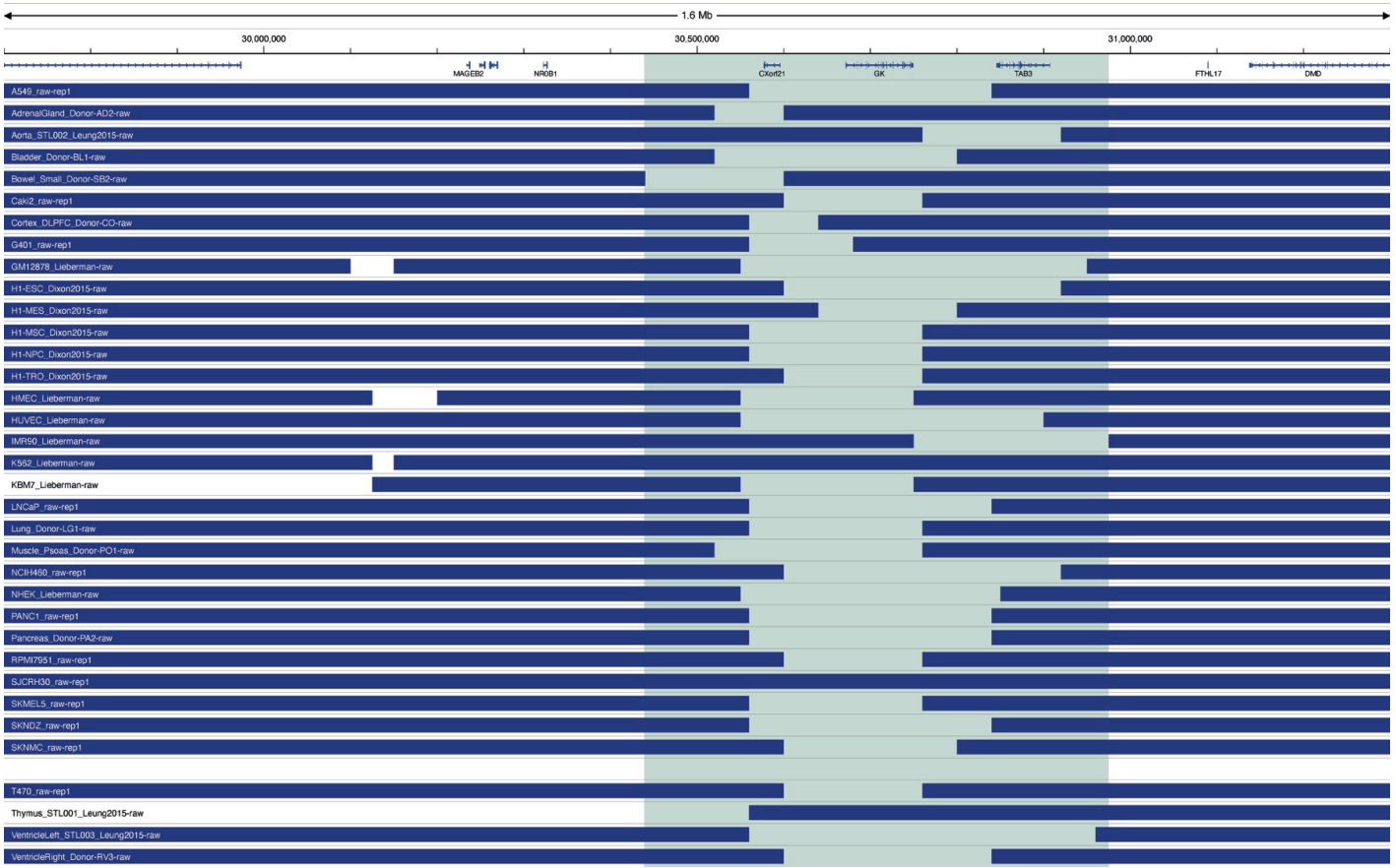

Figure S8: *NR0B1* TAD conformation of different human cell lines: Representation of TADs (rectangles) and their boundaries (empty spaces between rectangles) taken from 3D Genome Browser (Wang et al., 2018). The region encompassing all TAD boundaries consensus is shaded as light green.

Supporting Tables

Table S1. **Mutations in Genes Associated With 46,XY DSD Found in Patients 1 and 2.**

Patient	Gene	dbSNP ID	ExAc All (Lek et al., 2016)	ExAc European Non-Finish (Lek et al., 2016)	SIFT (Vaser et al., 2016)	Polyphen2 (Adzhubei et al., 2010)	MutationTaster (Schwarz et al., 2014)	ClinVar Miner (Henrie et al., 2018)	cDNA change	Protein Change
P1	ZFPM2	rs202217256	0.00366	0.00567	T (0.422)	P (0.914)	D (1.0)	benign	NM_012082.3: c.292G>A	NP_036214.2: p.Asp98Asn
P2	ESR2	rs367855747	0.0002	0.0004	D (0.021)	D (0.975)	D (0.999)	n.a.	NM_001214902.1: c. 1331G>A	NP_001201831.1: p.Ser444Asn

All three mutations are heterozygous missense SNVs.

SIFT: D = Deleterious, T = Tolerated; Polyphen2: B = benign, D = probably damaging, P = possibly damaging; MutationTaster: D = disease causing

References:

Adzhubei, I.A., S. Schmidt, L. Peshkin, V.E. Ramensky, A. Gerasimova, P. Bork, A.S. Kondrashov, and S.R. Sunyaev. 2010. A method and server for predicting damaging missense mutations. *Nat Methods*. 7:248-249.

Barbaro, M., A. Cicognani, A. Balsamo, A. Lofgren, L. Baldazzi, A. Wedell, and M. Oscarson. 2008. Gene dosage imbalances in patients with 46,XY gonadal DSD detected by an in-house-designed synthetic probe set for multiplex ligation-dependent probe amplification analysis. *Clinical genetics*. 73:453-464.

Barbaro, M., J. Cook, K. Lagerstedt-Robinson, and A. Wedell. 2012. Multigeneration Inheritance through Fertile XX Carriers of an NR0B1 (DAX1) Locus Duplication in a Kindred of Females with Isolated XY Gonadal Dysgenesis. *International journal of endocrinology*. 2012:504904.

Barbaro, M., M. Oscarson, J. Schoumans, J. Staaf, S.A. Ivarsson, and A. Wedell. 2007. Isolated 46,XY gonadal dysgenesis in two sisters caused by a Xp21.2 interstitial duplication containing the DAX1 gene. *The Journal of clinical endocrinology and metabolism*. 92:3305-3313.

Dong, Y., Y. Yi, H. Yao, Z. Yang, H. Hu, J. Liu, C. Gao, M. Zhang, L. Zhou, Asan, X. Yi, and Z. Liang. 2016. Targeted next-generation sequencing identification of mutations in patients with disorders of sex development. *BMC medical genetics*. 17:23.

Fishilevich, S., R. Nudel, N. Rappaport, R. Hadar, I. Plaschkes, T. Iny Stein, N. Rosen, A. Kohn, M. Twik, M. Safran, D. Lancet, and D. Cohen. 2017. GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database (Oxford)*. 2017.

Henrie, A., S.E. Hemphill, N. Ruiz-Schultz, B. Cushman, M.T. DiStefano, D. Azzariti, S.M. Harrison, H.L. Rehm, and K. Eilbeck. 2018. ClinVar Miner: Demonstrating utility of a Web-based tool for viewing and filtering ClinVar data. *Human mutation*. 39:1051-1060.

Kent, W.J., C.W. Sugnet, T.S. Furey, K.M. Roskin, T.H. Pringle, A.M. Zahler, and D. Haussler. 2002. The human genome browser at UCSC. *Genome research*. 12:996-1006.

Ledig, S., O. Hiort, G. Scherer, M. Hoffmann, G. Wolff, S. Morlot, A. Kuechler, and P. Wieacker. 2010. Array-CGH analysis in patients with syndromic and non-syndromic XY gonadal dysgenesis: evaluation of array CGH as diagnostic tool and search for new candidate loci. *Hum Reprod*. 25:2637-2646.

Lek, M., K.J. Karczewski, E.V. Minikel, K.E. Samocha, E. Banks, T. Fennell, A.H. O'Donnell-Luria, J.S. Ware, A.J. Hill, B.B. Cummings, T. Tukiainen, D.P. Birnbaum, J.A. Kosmicki, L.E. Duncan, K. Estrada, F. Zhao, J. Zou, E. Pierce-Hoffman, J. Berghout, D.N. Cooper, N. Deflaux, M. DePristo, R. Do, J. Flannick, M. Fromer, L. Gauthier, J. Goldstein, N. Gupta, D. Howrigan, A. Kiezun, M.I. Kurki, A.L. Moonshine, P. Natarajan, L. Orozco, G.M. Peloso, R. Poplin, M.A. Rivas, V. Ruano-Rubio, S.A. Rose, D.M. Ruderfer, K. Shakir, P.D. Stenson, C. Stevens, B.P. Thomas, G. Tiao, M.T. Tusie-Luna, B. Weisburd, H.H. Won, D. Yu, D.M. Altshuler, D. Ardissino, M. Boehnke, J. Danesh, S. Donnelly, R. Elosua, J.C. Florez, S.B. Gabriel, G. Getz, S.J. Glatt, C.M. Hultman, S. Kathiresan, M. Laakso, S. McCarroll, M.I. McCarthy, D. McGovern, R. McPherson, B.M. Neale, A. Palotie, S.M. Purcell, D. Saleheen, J.M. Scharf, P. Sklar, P.F. Sullivan, J. Tuomilehto, M.T. Tsuang, H.C. Watkins, J.G. Wilson, M.J. Daly, D.G. MacArthur, and Exome Aggregation Consortium. 2016. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 536:285-291.

Pattyn, F., F. Speleman, A. De Paepe, and J. Vandesompele. 2003. RTPrimerDB: the real-time PCR primer and probe database. *Nucleic Acids Res*. 31:122-123.

Robinson, J.T., H. Thorvaldsdottir, W. Winckler, M. Guttman, E.S. Lander, G. Getz, and J.P. Mesirov. 2011. Integrative genomics viewer. *Nature biotechnology*. 29:24-26.

Schwarz, J.M., D.N. Cooper, M. Schuelke, and D. Seelow. 2014. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Methods*. 11:361-362.

Smyk, M., J.S. Berg, A. Pursley, F.K. Curtis, B.A. Fernandez, G.A. Bien-Willner, J.R. Lupski, S.W. Cheung, and P. Stankiewicz. 2007. Male-to-female sex reversal associated with an approximately 250 kb deletion upstream of NR0B1 (DAX1). *Hum Genet*. 122:63-70.

Vaser, R., S. Adusumalli, S.N. Leng, M. Sikic, and P.C. Ng. 2016. SIFT missense predictions for genomes. *Nat Protoc*. 11:1-9.

Wang, Y., F. Song, B. Zhang, L. Zhang, J. Xu, D. Kuang, D. Li, M.N.K. Choudhary, Y. Li, M. Hu, R. Hardison, T. Wang, and F. Yue. 2018. The 3D Genome Browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions. *Genome biology*. 19:151.

White, S., T. Ohnesorg, A. Notini, K. Roeszler, J. Hewitt, H. Daggag, C. Smith, E. Turbitt, S. Gustin, J. van den Bergen, D. Miles, P. Western, V. Arboleda, V. Schumacher, L. Gordon, K. Bell, H. Bengtsson, T. Speed, J. Hutson, G. Warne, V. Harley, P. Koopman, E. Vilain, and A. Sinclair. 2011. Copy number variation in patients with disorders of sex development due to 46,XY gonadal dysgenesis. *PloS one*. 6:e17793.