Title
Hi-C Identifies Complex Genomic Rearrangements and TAD-Shuffling in Developmental Diseases

Journal
American Journal of Human Genetics, 106(6)

ISSN
0002-9297

Authors
Melo, Uirá Souto
Schöpflin, Robert
Acuna-Hidalgo, Rocio
et al.

Publication Date
2020-06-01

DOI
10.1016/j.ajhg.2020.04.016

Peer reviewed

Hi-C Identifies Complex Genomic Rearrangements and TAD-Shuffling in Developmental Diseases

Uirá Souto Melo,[1,2,18] Robert Schöpflin,[1,2,18] Rocio Acuna-Hidalgo,[1,2,18] Martin Atta Mensah,[2] Björn Fischer-Zirnsak,[1,2] Manuel Holtgrewe,[2,3] Marius-Konstantin Klever,[1,2] Seval Türkmen,[2] Verena Heinrich,[4] Ilina Datkhaeva Pluym,[5] Eunice Matoso,[6,7] Sérgio Bernardo de Sousa,[6] Pedro Louro,[6,8,9] Wiebke Hülsemann,[10] Monika Cohen,[11] Andreas Dufke,[12] Anna Latos-Bieleńska,[13,14] Martin Vingron,[4] Vera Kalscheuer,[1] Fabiola Quintero-Rivera,[15] Malte Spielmann,[16,17,*] and Stefan Mundlos[1,2,*]

Genome-wide analysis methods, such as array comparative genomic hybridization (CGH) and whole-genome sequencing (WGS), have greatly advanced the identification of structural variants (SVs) in the human genome. However, even with standard high-throughput sequencing techniques, complex rearrangements with multiple breakpoints are often difficult to resolve, and predicting their effects on gene expression and phenotype remains a challenge. Here, we address these problems by using high-throughput chromosome conformation capture (Hi-C) generated from cultured cells of nine individuals with developmental disorders (DDs). Three individuals had previously been identified as harboring duplications at the SOX9 locus and six had been identified with translocations. Hi-C resolved the positions of the duplications and was instructive in interpreting their distinct pathogenic effects, including the formation of new topologically associating domains (neo-TADs). Hi-C was very sensitive in detecting translocations, and it revealed previously unrecognized complex rearrangements at the breakpoints. In several cases, we observed the formation of fused-TADs promoting ectopic enhancer-promoter interactions that were likely to be involved in the disease pathology. In summary, we show that Hi-C is a sensible method for the detection of complex SVs in a clinical setting. The results help interpret the possible pathogenic effects of the SVs in individuals with DDs.

Introduction

Over the last decade, the development and refinement of genomic technologies has paved the way for a major transformation regarding genotype-phenotype correlations in the field of human genetics.[1] Genome-wide methods such as microarrays and high-throughput sequencing techniques have improved the identification of genetic alterations involved in disease, increasing our understanding of genetic disorders.[2–5] Genetic variation occurs at various levels ranging from single-nucleotide variations (SNVs) and small indels to larger structural variants (SVs). SVs are a large class of genome alterations with a size >50 bp; this class includes deletions, duplications, inversions, insertions, and translocations.[6] SVs can be detected by various genetic and/or genomic screening tools that each hold certain advantages and disadvantages in regard to detection sensitivity and specificity. For instance, while array comparative genomic hybridization (CGH) can reliably detect large copy number changes, it does not reveal their precise genomic position and fails to detect copy-number-neutral variants, such as inversions and translocations. Short-read-based whole-genome sequencing (WGS) can in principle detect all SV types genome-wide, but its resolution and specificity is limited in repetitive regions of the genome. Furthermore, the detection of balanced SVs relies on the presence of chimeric reads spanning the breakpoint. Thus, detecting and resolving the layout of complex genomic rearrangements (CGRs) with multiple breakpoints can be challenging with WGS.[7,8]

Aside from the detection of SVs, their clinical interpretation is a largely unsolved problem, especially when they do not disrupt protein coding parts of the genome. Recent work has shown that SVs can disrupt the complex three-dimensional (3D) architecture of the human genome, causing position effects and thereby contributing to developmental disorders (DDs).[9] The development of high-throughput chromosome conformation capture (Hi-C)[10,11]

[1]Max Planck Institute for Molecular Genetics, RG Development and Disease, 13353 Berlin, Germany; [2]Institute for Medical Genetics and Human Genetics, Charité Universitätsmedizin Berlin, 13353 Berlin, Germany; [3]Berlin Institute of Health (BIH), Core Unit Bioinformatics, 10117 Berlin, Germany; [4]Max Planck Institute for Molecular Genetics, Department of Computational Molecular Biology, 13353 Berlin, Germany; [5]Department of Obstetrics and Gynecology, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA 90095, USA; [6]Medical Genetics Unit, Centro Hospitalar e Universitário de Coimbra, 3000-075 Coimbra, Portugal; [7]Center of Investigation on Environment Genetics and Oncobiology (iCBR-CIMAGO), Faculty of Medicine, University of Coimbra, 3000-548 Coimbra, Portugal; [8]Familial Risk Clinic, Instituto Português de Oncologia de Lisboa Francisco Gentil, 1099-023 Lisboa, Portugal; [9]Faculty of Health Sciences, Universidade da Beira Interior, 6201-001 Covilhã, Portugal; [10]Handchirurgie Kinderkrankenhaus Wilhelmstift, 22149 Hamburg, Germany; [11]kbo-Kinderzentrum München, 81377 München, Germany; [12]Institut für Medizinische Genetik und Angewandte Genomik, 72076 Tübingen, Germany; [13]Department of Medical Genetics, University of Medical Sciences in Poznan, 60-806 Poznan, Poland; [14]Centers for Medical Genetics GENESIS, Grudzieniec st, 60-601 Poznan, Poland; [15]Department of Pathology and Laboratory Medicine, UCLA Clinical Genomics Center, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA 90095, USA; [16]Max Planck Institute for Molecular Genetics, Human Molecular Genomics Group, 13353 Berlin, Germany; [17]Institut für Humangenetik Lübeck, Universität zu Lübeck, 23538 Lübeck, Germany
[18]These authors contributed equally to this work
*Correspondence: spielman@molgen.mpg.de (M.S.), mundlos@molgen.mpg.de (S.M.)
https://doi.org/10.1016/j.ajhg.2020.04.016.

revealed that chromatin interactions occur preferentially within defined and stable regions of the genome; these regions are known as topologically associating domains (TADs).[12,13] TADs are domains in the genome which are separated by insulating proteins (e.g., CTCF) and which build a framework for contacts of regulatory elements and genes. SVs can disrupt the TAD architecture and rewire the regulatory landscape of genes, which can lead to gene misregulation and cause DDs.[14–16] Therefore, genomic alterations leading to TAD disruption, abnormal chromatin interactions, and subsequent misregulation of gene expression are emerging as largely unexplored mechanisms involved in genetic disorders.[17]

Here we applied Hi-C to samples from nine individuals with DDs. For all of these cases, a candidate SV related to each individual had previously been detected via karyotyping and/or array CGH. Hi-C can detect and help resolve complex SVs in clinically available tissues and, in some cases, a plethora of CGRs was detected which had been overlooked with conventional molecular tools. Furthermore, we evaluated the changes of the TAD landscapes associated with the SVs in order to predict their disease-causing potential. Based on the changes of the 3D chromatin structure, we suggest putative disease-associated genes for several cases, and we derive hypotheses of the disease mechanism on a genetic level. This demonstrates that Hi-C can be a powerful tool for the identification and clinical interpretation of SVs in the context of DDs, helping improve the understanding of the mechanisms through which SVs can perturb development.

Material and Methods

Subjects and Ethics Approval

The study was performed with the approval of the Charité Ethics Committee, Berlin, Germany. All subjects, their parents, or their legal guardians provided informed and written consent for participation in this study. The clinical evaluation included medical history interviews, a physical examination, and review of medical records. Blood samples were obtained from each participating individual except for one fetus, and DNA was extracted using standard procedures. Additionally, skin biopsies were obtained from two individuals and amniocytes from one fetus.

Cytogenetic and Genomic Screening

Samples from individuals with DD (n = 7) were submitted for chromosome banding using the trypsin-Giemsa technique (Table 1). The detected chromosomal translocations were further confirmed through the use of fluorescence *in situ* hybridization (FISH) on metaphases. Copy number analysis on individuals with DD (n = 4) was carried out via array CGH using a whole-genome 1 M oligonucleotide array (Agilent; Table 1 and Table S1), or 2.6M Affymetrix CytoScan HD (Affymetrix). 1 M arrays were analyzed through the use of Feature Extraction v9.5.3.1 and CGH Analytics v3.4.40 or Cytogenomics v2.5.8.11 software, respectively (Agilent). The analysis settings were as follows: aberration algorithm—ADM-2; threshold—6.0; window size—0.2 Mb; filter—five probes, log2 ratio = 0.29. The 2.6M array data was analyzed using Chromosome Analysis Suite 4.0 (ChAS) version 3.3.0.139 (r10838). Copy number variants (CNVs) detected via array CGH were confirmed using qPCR.

Chromosomal analysis was performed on metaphases obtained from 72 h phytohemagglutinin (PHA) stimulated peripheral blood lymphocyte cultures according to standard procedures. Analysis of GTG-banded chromosomes was done at a resolution of 700 bands per haploid genome according to the International System for Human Cytogenomic Nomenclature (ISCN) 2016. Confirmatory FISH was performed on metaphases obtained from 72 h PHA stimulated peripheral blood lymphocyte cultures according to standard procedures.

WGS was also performed in blood from individuals DD1, DD3, and DD4; in lymphoblastoid cell lines (LCLs) from DD5–DD9; and in fetal amniocytes from individual DD2 in order to validate the SVs' breakpoints (Table 1 and Tables S1 and S2). Sequencing was performed using Macrogen on Illumina HiSeq X machines with Illumina TruSeq PCR-free chemistry. After quality control, reads were aligned to the GRCh37 sequence (hs37d5.fa) with BWA-MEM,[18] duplicates were masked using SAMBLASTER,[19] and the reads were sorted and converted to BAM files using Samtools.[20] SVs were detected using Delly,[21] PopDel,[22] and ERDS.[23] Phenotype based exome analysis was performed using Exomiser[24] and MutationDistiller.[25] All rare (MAF < 0.01) heterozygous loss-of-function variants are shown in Table S3.

Cell Culture

Fibroblast cell lines were established from skin biopsies of two individuals with DD and one healthy control by following a standard procedure. Fibroblasts were cultured in Dulbecco's Modified Eagle Medium (DMEM; Thermo Fisher Scientific) supplemented with 15% fetal bovine serum (FBS; Thermo Fisher Scientific), 1% L-glutamine (Thermo Fisher Scientific), and 1% penicillin-streptomycin (Thermo Fisher Scientific). Amniocytes from one individual with DD were cultured in AmnioMAX II Complete Medium (Thermo Fisher Scientific). LCLs were established by Epstein-Barr virus (EBV) transformation of leucocytes from peripheral blood samples of individuals with DD (n = 6) and of one control. LCLs were cultured in Roswell Park Memorial Institute (RPMI) medium (Thermo Fisher Scientific) with 15% fetal calf serum (FCS) and 1% penicillin-streptomycin. Fibroblasts were grown to confluence prior to fixation for the preparation of the Hi-C libraries, whereas LCLs and amniocytes were cultured until the plateau phase of growth was reached.

Preparation of Hi-C Libraries

Hi-C libraries were processed as described in the previously published *in situ* protocol.[11] In brief, ~1 million cells were fixed in 2% formaldehyde, lysed, and digested overnight with *DpnII* enzyme (New England BioLabs, R0543). Digested DNA ends were marked with biotin-14-dATP (Thermo Fisher Scientific, 19524016) and ligated overnight using T4 DNA ligase (New England BioLabs, M0202). Formaldehyde crosslinking was reversed by incubation in 5 M NaCl for 2 h at 68°C, followed by ethanol precipitation. Covaris (S-Series 220) was used to shear the DNA to fragments of 300–600 bp for library preparation, and biotin-filled DNA fragments were pulled down using Dynabeads MyOne Streptavidin T1 beads (Thermo Fisher Scientific, 65602). The DNA ends were subsequently repaired using T4 DNA polymerase and the Klenow fragment of DNA polymerase I (New England BioLabs, M0203 and M0210) and phosphorylated with T4 Polynucleotide Kinase NK (New England BioLabs, M0201). The DNA was further

Table 1. Overview of Our Cohort with Nine Individuals Presenting with DDs and Their Respective SVs

Cytogenetics/Cytogenomics Results Prior to This Study									WGS		Hi-C	
Subjects	Detection Tool	Structural Variant				ClinGen CNVPathogenicity Calculator		Case SolvedPrior to Hi-C?	-		-	
-	-	ISCN 2016 Nomenclature	Type	Size	Inheritance	Prediction	Score	-	Detectedthe SV?*	Cell Type	SV Is Visible?	Case Solved Using Hi-C?
DD1	array CGH	arr[GRCh37] 17q24.3 (67956481_ 70087077)x3	duplication	2.1 Mb	*de novo*	Pathogenic	2.35	yes	yes	fibroblasts	yes	NA
DD2	array CGH	arr[GRCh37] 17q24.3 (67442273_ 70559193)x3	duplication	3.1 Mb	*de novo*	likely pathogenic	0.9	no	no	amniocytes	yes	no
DD3	array CGH	arr[GRCh37] 17q24.3q25.1 (68620187_ 71083594)x3	duplication	2.4 Mb	NR	VUS	−0.3	no	no	fibroblasts	yes	yes
DD4	karyotyping	46,XX,t(5;18) (q31.1;q12.3)	reciprocal translocation	NA	*de novo*	NA	NA	yes	yes	LCLs	yes	NA
DD5	karyotyping	46,XY,t(14;20) (q12;q13.2)	reciprocal translocation	NA	*de novo*	NA	NA	no	no	LCLs	yes	yes
DD6	karyotyping	46,XX,t(3;5) (q24;q14.3)	reciprocal translocation	NA	NR	NA	NA	no	yes**	LCLs	yes	yes
DD7	karyotyping	46,XY,t(2;4) (p12;q34)	reciprocal translocation	NA	*de novo*	NA	NA	no	yes	LCLs	yes	no
DD8	karyotyping	46,XY,t(7;8) (p12;q22.1)	reciprocal translocation	NA	NR	NA	NA	no	yes	LCLs	yes	no
DD9	karyotyping	46,XX,t(2;7) (q33.1;p21)	reciprocal translocation	NA	*de novo*	NA	NA	no	yes	LCLs	yes	no

Legend: DDs—developmental disorders; SV—structural variant; CNV—copy number variant; Hi-C—high-throughput chromosome conformation capture; CGH—comparative genomic hybridization; ISCN—International System for Human Cytogenomic Nomenclature; LCLs—lymphoblastoid cell line; WGS—whole-genome sequencing; NR—not reported; NA—Not applicable; *SV callers; **One or more breakpoint(s) missed by callers.

prepared for sequencing by ligating adaptors to the DNA fragments, using the NEBNext Multiplex Oligos for Illumina kit (New England BioLabs, E7335 and E7500). Indexes were added via PCR amplification (4–8 cycles) using the NEBNext Ultra II Q5 Master Mix (New England BioLabs, M0544). PCR purification and size selection were carried out using Agencourt AMPure XP beads (Beckman Coulter, A63881). Libraries were deep sequenced (~370 million fragments) in a 75 bp paired-end run on a HiSeq4000 (Illumina). For each individual, the Hi-C library was created by pooling a total of four technical replicates generated from two different cell cultures in order to ensure higher complexity of the sequencing library.

Hi-C Bioinformatics Analysis

Paired-end sequencing data were processed using the Juicer pipeline v1.5.6, CPU version.[26] The pipeline was set up with BWA v0.7.17[27] for aligning short reads with BWA-MEM to the reference genome hg19. Alternative haplotypes were removed from hg19, and the sequence of the Epstein-Barr virus (NC_007605.1) was added. Replicates were merged by combining the output files of the Juicer pipeline containing the information from filtered and deduplicated read-pairs. Juicer Tools[26] was used to create raw count maps, as well as maps normalized with Knight and Ruiz (KR) matrix balancing;[11,26,28] we inspected both. In addition to generating Hi-C maps binned to a regular grid, we also generated restriction-fragment-based Hi-C maps for further inspection. For the generation of Hi-C maps, we used read-pairs with mapping quality (MAPQ) ≥ 30. However, for spotting genomic rearrangements, it can be helpful to also generate and inspect Hi-C maps with lower, more permissive MAPQ thresholds. In some cases, such as copy number variations, the KR normalization can be disadvantageous because the algorithm assumes equal visibility of all loci,[11] and that is not the case when the number of copies differs locally. In case of duplications, the algorithm would tend to reduce the signal of the duplicated region to compensate for the additional allele. Due to this, we directly used raw count maps for the display of duplication events. This has the disadvantage that locus-specific biases, which are addressed by the KR normalization, are still contained in the map, but has the advantage that the copy number variation becomes more clearly visible (see also Supplemental Notes).

In order to generate Hi-C maps with reduced sequencing depth, we extracted the same number of reads for each replicate from the FASTQ files, such that the number of extracted read-pairs summed to 1 M, 10 M, 50 M, and 100 M, respectively. These subsampled FASTQ files were processed as described above.

Genome-wide Hi-C maps, inter-chromosomal Hi-C maps, and symmetric intra-chromosomal Hi-C maps were visualized using

Juicebox (Desktop version 1.8.8).[29] For specific loci of interest, intra-chromosomal Hi-C maps were visualized as a heatmap of the upper triangle matrix rotated by 45 degrees. For this type of visualization, very high values were truncated to improve the display of smaller count values. We used ChIP-seq ENCODE data for CTCF (AG09303; human toe fibroblast from an apparently healthy adult), H3K27ac, H3K4me1, and H3K4me3.[30]

Criteria for Selecting Candidate Genes Related to the Individual's Phenotype

Chromosomal alterations can move genes or regulatory elements into new regulatory landscapes. When an SV disrupts the integrity of a TAD, new hybrid domains may emerge between the TAD boundaries adjacent to the breakpoints, fusing together parts from different regulatory landscapes. Depending on the nature of the SVs, these hybrid domains emerge as neo-TADs (duplication), fused-TADs (deletion), or shuffled-TADs (inversion/translocation).[9] By these events, new enhancer-promoter interactions (EPIs) as well as loss of EPIs are possible. Based on the gain or loss of EPIs, we applied the following two criteria to select candidate genes putatively linked to the disease, depending on the individual's phenotype: (A) the gene was already associated with a neurodevelopmental phenotype or limb phenotype, described in Online Mendelian Inheritance in Man (OMIM) or (B) a phenotype similar to the individuals' symptoms was reported from an animal model (*Mus musculus*) as retrieved from specific databases (Mouse Genomic Informatics and Mouse Phenotype). We selected candidate genes related to the individual's phenotype when fulfilling both criteria listed above.

We applied the new ClinGen CNV Pathogenicity Calculator to evaluate the pathogenicity of all duplication cases (individuals DD1–DD3). The scores and criteria were the following:

- For DD1:
  - 1A. Contains protein-coding or other known functionally important elements (assigned points: 0)
  - 2A. Complete overlap; the TS gene or minimal critical region is fully contained within the observed copy number gain (assigned points: 1)
  - 3A. 0–34 genes (assigned points: 0)
  - 4A. … the reported phenotype is highly specific and relatively unique to the gene or genomic region (assigned points: 0.9)
  - 5A. Use appropriate category from *de novo* scoring section in Section 4 (assigned points: 0.45)
  - Prediction: pathogenic (total score: 2.35)
- For DD2:
  - 1A. Contains protein-coding or other known functionally important elements (assigned points: 0)
  - 2H. HI gene fully contained within observed copy number gain (assigned points: 0)
  - 3A. 0–34 genes (assigned points: 0)
  - 4A. … the reported phenotype is highly specific and relatively unique to the gene or genomic region. Confirmed *de novo*: 0.45 points (assigned points: 0.45)
  - Observed copy number gain is *de novo*
  - 5A. Use appropriate category from *de novo* scoring section in Section 4. (assigned points: 0.45)
  - Prediction: likely pathogenic (total score: 0.9).
- For DD3:
  - 1A. Contains protein-coding or other known functionally important elements (assigned points: 0)
  - 2H. HI gene fully contained within observed copy number gain
  - 3A. 0–34 genes (assigned points: 0)
  - 4D. … the reported phenotype is *not* consistent with the gene/genomic region or not consistent in general (assigned points: −3)
  - 5A. Use appropriate category from *de novo* scoring section in Section 4
  - Prediction: variant of uncertain significance (total score: −0.3).

Real-Time Quantitative (qPCR and RT-qPCR) Analyses

qPCR was performed to validate small duplications detected by using Hi-C in DNA from individual DD3. We designed two primer pairs as calibrators outside the duplicated region and three located within each duplication. RT-qPCR was performed to check *MEF2C* expression in DD6. RNA was extracted from LCLs (n = 5) by using RNeasy Mini Kit (QIAGEN). Total RNA (1 μg/μL) was reverse-transcribed into cDNA with oligo(dT) primers by using SuperScript IV First-strand Synthesis System (Thermo Fisher). qPCR and RT-qPCR were performed using the PowerUp SYBR® Green Master Mix (Thermo Fisher) and subjected to QuantStudio 6 Flex Real-Time PCR System, 384-well (Applied Biosystems). Copy number and gene expression were calculated using the $2^{-\Delta\Delta CT}$ method.[31] Each experiment was performed once with three technical triplicates per sample.

Results

Cohort of Individuals with DD and Detection of SVs

We generated genomic data from a cohort of nine individuals with DD. Based on initial diagnostic array CGH or karyotype results, at least one candidate SV had been detected in each individual (Table 1). The detected SVs were mostly *de novo* events with high impact on genomic structure and were therefore likely pathogenic, although the molecular mechanism of disease in most of the cases was unclear (seven out of nine). Clinical aspects and their respective SVs are described in detail in Table S1. Three out of nine affected individuals had microduplications (> 2 Mb) involving 17q24, and six had reciprocal chromosomal translocations. The SV breakpoints of all cases were validated using WGS. We performed Hi-C in all nine samples using the available clinical cells for each case (e.g., fibroblasts, amniocytes, or LCLs). Samples were sequenced for, on average, 370 million reads; and ~160-250 (~55%) million reads passed the quality control within the Juicer pipeline per sample (Figure S1).

Hi-C Reveals the Complex Nature of Duplications at the SOX9 Locus

We investigated three individuals (DD1, DD2, and DD3) with microduplications at the SOX9 locus on chromosome 17 (17q24.3). DD1 was an individual with Cooks syndrome (MIM: 106995), a condition previously described by us as being caused by microduplications containing enhancers upstream of the SOX9 locus but not the *SOX9* gene itself, and the neighboring *KCNJ2* gene (Table 1).[15,32] The microduplication in DD2 (3.1 Mb) encompassed *SOX9*, *KCNJ2*, and *KCNJ16*, as well as parts of *MAP2K6*. This

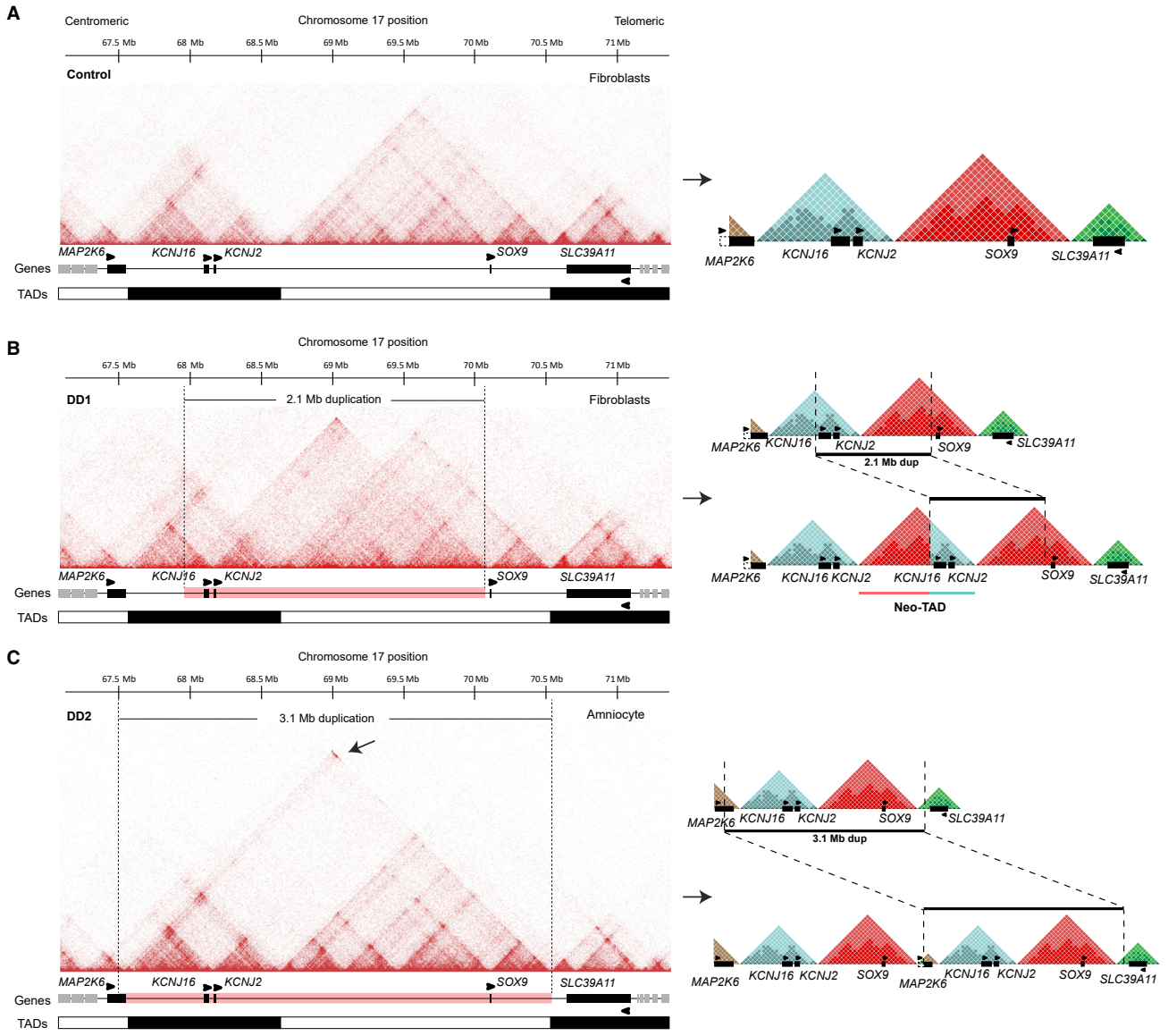

Figure 1. Hi-C Reveals Disruption of 3D Chromatin Folding Resulting from SVs at the 17q24 Region

(A) Hi-C map of a control sample (skin fibroblast; 10 kb resolution; raw count map) from an unaffected individual, showing the 3D landscape of the 17q24 region. TADs are represented by white and black bars on the track below. On the right: schematic representation of TAD structure on the 17q24 region (*MAP2K6* TAD in brown; *KCNJs* TAD in light blue; *SOX9* TAD in red; *SLC39A11* TAD in green).

(B) Hi-C map from fibroblasts of DD1 with a 2.1 Mb tandem duplication (chr17:67,958,880–70,085,143; GRCh37/hg19; highlighted in pink in the gene track below the respective map). Note that *SOX9* is not included in the duplication. The duplication induces the formation of a neo-TAD containing *KCNJ2* and *KCNJ16* and known regulatory elements of the *SOX9* locus, leading to ectopic expression of *KCNJ2*.

(C) Hi-C map from DD2 (amniocytes) showing a 3.1 Mb tandem duplication (highlighted in pink) on 17q24.3 (chr17:67,441,705–70,564,888), probably associated with the individual's phenotype. Note that the TSS of *MAP2K6* and *SLC39A11* are not included in the duplication. The underlying SV enables contacts between genomic regions at the margins of the duplication chr17:67,440,000–67,560,000 and chr17:70,545,000–70,565,000 (highlighted by an arrow), neither region containing ORFs, while the TADs of *SOX9* and *KCNJ* genes are copied as a whole.

individual had a severe heart defect, skeletal anomalies, and hydrops fetalis. The duplication (2.4 Mb) in individual DD3 is shifted to the telomeric side, including *SOX9* and the next flanking gene *SLC39A11*, but not *KCNJ2*. Individual DD3 had psychomotor developmental delay, microcephaly, and mild intellectual disability. Thus, in spite of their partially overlapping duplications, the outcome was very different, resulting in distinct phenotypes. We employed Hi-C in these individuals to investigate the effects of the rearrangements on chromatin structure in order to better understand their very different effects on development.

First, we performed Hi-C in fibroblasts from a healthy individual to serve as a control (Control; Figure 1A). At the locus of the *SOX9* gene, several well-defined TADs are visible in the Hi-C map. The TAD containing *SOX9* is

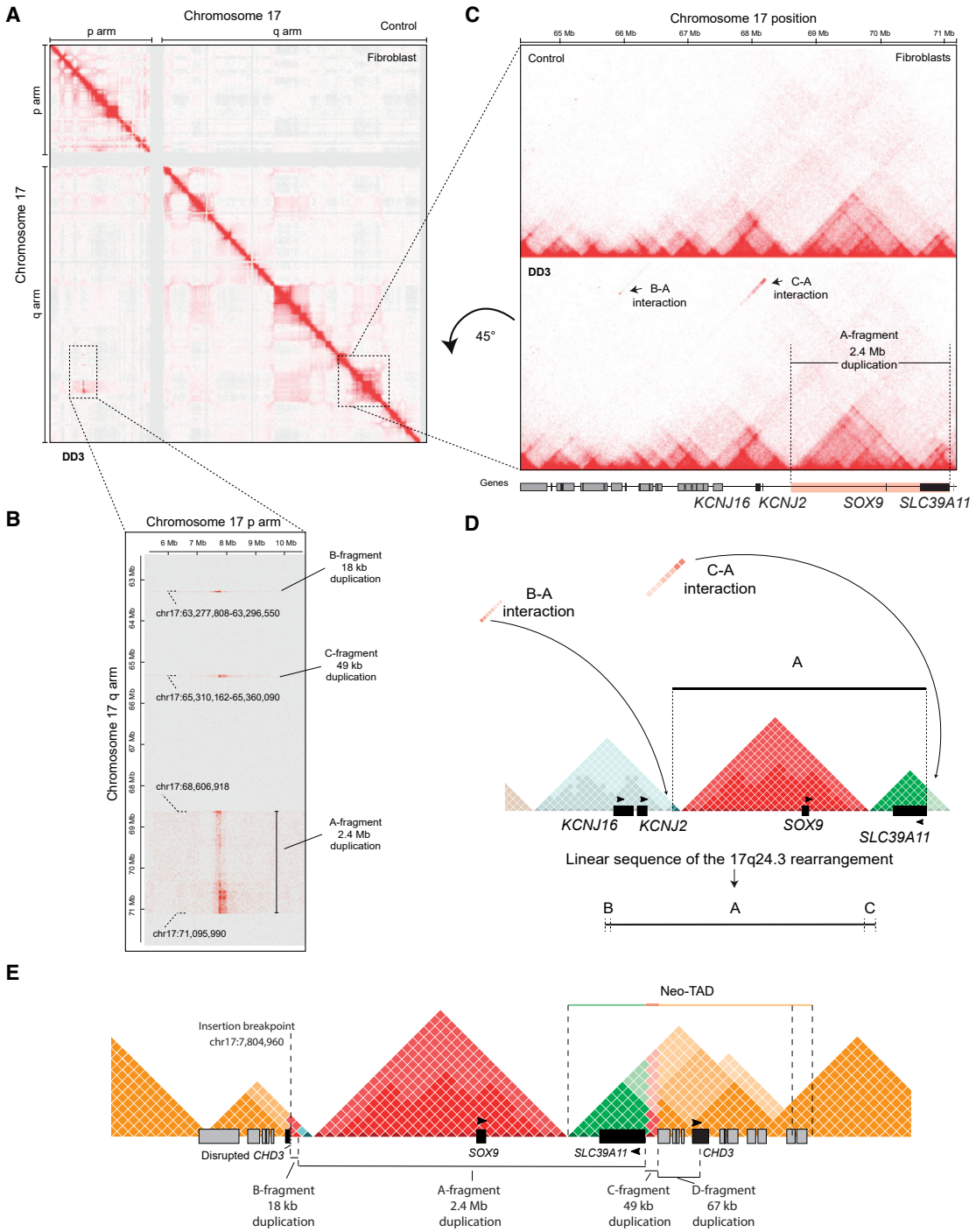

Figure 2. Hi-C Reveals a Complex Rearrangement Composed of Several SVs in an Individual with DD

(A) Hi-C map from DD3 fibroblasts reveals a complex rearrangement involving p- and q-arms of chromosome 17 (500 kb resolution; raw count map).

(B) Zoom-in on the ectopic signal of chr17 showing the large 2.4 Mb duplication (A-fragment), plus two other duplications (18 kb, B-fragment; and 49 kb, C-fragment). Note: Based on the orientation of the gradient of the ectopic Hi-C signal (fades away in horizontal direction, along the p-arm), we could infer that the three duplications from the q-arm are inserted on the p-arm (insertion breakpoint: chr17:7,804,960; for details see Figure S3).

(C) Hi-C cis-map of the 17q24 region (control and DD3) showing the interaction of the small duplications (B- and C-fragments) with the large one (A-fragment). Note that the *SLC39A11* ORF is part of the duplication (25 kb resolution; raw count map).

(D) Schematic view of the 17q24 rearrangement containing the B-A-C rearranged fragments. The 18 kb duplication is inserted upstream of the 2.4 Mb duplication, in front of the SOX9 TAD boundary (B-A interaction), and the 49 kb duplication (C-fragment) is inserted downstream of the large duplication, within the SLC39A11 neo-TAD (C-A interaction).

(legend continued on next page)

delimited by boundaries insulating it from the neighboring TADs and contains no further genes. The TAD flanking the *SOX9* TAD on the left (centromeric direction) contains the potassium channel genes *KCNJ2* and *KCNJ16*, while the TAD flanking the *SOX9* TAD on the right (telomeric direction) contains the gene *SLC39A11*.

The individual with Cooks syndrome (DD1) presented with hypoplasia and aplasia of nails and distal phalanges, similar to the previously described cases,[28] and a 2.1 Mb heterozygous duplication at the 17q24.3 locus had been detected initially using array CGH (Figure S1C and Table 1). Hi-C performed in fibroblasts showed a gain of new chromatin interactions at the SOX9 locus not comprising *SOX9* but including *KCNJ2* and *KCNJ16* (Figure 1B). Hi-C showed a strong interaction between the beginning and end of the duplicated region; this indicates that the 2.1 Mb duplication was in tandem (Figure 1B). The data show the emergence of a new chromatin domain (neo-TAD), which is visible as a new chromatin interaction superimposed on the wild-type SOX9 and KCNJs TADs (boxed region in Figure 1B). The newly formed TAD contains copies of *KCNJ2* and *KCNJ16*, as well as regulatory elements of the SOX9 TAD, which can now ectopically interact with each other (Figure 1B, right panel). As shown previously, the *SOX9* enhancers in the neo-TAD are capable of driving *KCNJ2* misexpression in a SOX9-like fashion in the developing limb, leading to hypoplasia and/or aplasia of nails and distal phalanges.[15]

Individual DD2 was a fetus with severe heart defects, skeletal anomalies, and hydrops fetalis (Table 1 and Table S1). A diagnostic array CGH revealed an ~3.1 Mb duplication (Figure S1D and Table 1), which included the complete SOX9 TAD and the neighboring KCNJ2/16 TADs, as well as parts of *MAP2K6* but not the transcription start site (TSS). Hi-C generated from amniocytes revealed that the duplication was in tandem (Figure 1C). A new chromatin interaction (neo-TAD) occurred between the centromeric margin of the duplication (estimated coordinates, chr17:67,440,000–67,560,000; ~120 kb) and the telomeric margin (estimated coordinates, chr17:70,545,000–70,565,000; 20 kb), but neither region contains genes, except for the remaining truncated region of *MAP2K6* (arrow on Figure 1C). The SOX9 and KCNJs TADs genes were copied as a whole including their boundaries, thus preserving their integrity. The effect of this 3.1 Mb microduplication is thus increased gene dosage without disruption of any regulatory landscape. Based on a gene dosage sensitivity tool (ClinGen), pure dosage effects of this gene are not expected to result in significant phenotypes. Inward-rectifier potassium channels (IRK) such as *KCNJ2* are important regulators of resting membrane potential and cell excitability. Dominant-negative mutations in *KCNJ2* result in Andersen syndrome (MIM: 170390), a condition characterized by periodic paralysis, cardiac arrhythmias, and dysmorphic features.[33] In addition, autosomal dominant atrial fibrillation (MIM: 613980) has been associated with activating mutations in *KCNJ2*.[34] Duplications have not been reported. It is possible that an increased gene dosage of *KCNJ2* leads to fetal cardiac arrhythmias and thus to hydrops. However, this is unlikely given the absence of arrhythmias in all the ultrasounds; this also would not explain the heart malformation. We conclude that while Hi-C provided more information on the 3D configuration of the genome of individual DD2, it still does not fully explain the clinical findings observed in this case.

In the third case with duplication at the SOX9 locus (DD3), array CGH detected an ~2.4 Mb duplication (Figure S2A) that included *SOX9* and the flanking *SLC39A11* gene, a presumed cellular zinc transporter. G-banded karyotype detected additional material on the short arm of chromosome 17 (46,XX,add(17)(p13.1); Figure S2B), which was later confirmed via FISH to be an insertion of the 2.4 Mb duplication (Figure S2C). The resulting rearrangement is described as 46,XX,add(17)(p13.1).ish der(17)ins(17)(p13.1) (RP11-84E24+).arr[GRCh37] 17q24.3q25.1(68620187_71083594)x3. Hi-C generated from DD3 fibroblasts revealed a gain of a new chromatin contact on the cis-map of chromosome 17 (Figure 2B) resulting from physical interaction between the 2.4 Mb fragment from the q-arm (here named A-fragment) with the p-arm of the same chromosome (Figure 2B). Hi-C also detected two small duplications from the 17q24.3 locus of 18 kb (named B-fragment) and 49 kb (named C-fragment), all validated through the use of WGS and qPCR (Figure 2C; Figure S2D). The 2.4 Mb duplication was poorly visible in the DD3 Hi-C map (Figure 2D), but it became more apparent after we subtracted a map from a reference sample (Figure S2E). Hi-C revealed that the 18 and 49 kb duplications were inserted up- (B-A interaction) and downstream (C-A interaction) of the 2.4 Mb duplication, and this was further validated using WGS (Figure S3). The linear sequence of the rearranged fragment is B-A-C (Figure 2C and 2D).

Hi-C indicated that the B-A-C fragment was inserted in 17p13 (Figure 2B). The region on 17p13 is gene dense with small chromatin domains without clear boundaries (Figures S4). The inserted fragment from the long arm disrupted the *CHD3* open reading frame (ORF) (Figure S3A). However, further analysis revealed a third duplication of 67 kb (D-fragment; validated by qPCR, Figure S4), containing four genes including a complete copy of *CHD3* (Figure S5). To unravel the complex 3D landscape of the nested rearrangement, we produced a schematic linear map of the region (Figure 2E). This reconstructed map suggests a TAD-fusion of the remaining SLC39A11 TAD (green) with the duplicated C-fragment (red) plus

(E) Schematic representation of the 3D chromatin landscape on the der(17p13) region. Hi-C maps helped to unveil the derivative 3D landscape and the order of fragments in a linear sequence, which consists of a neo-TAD containing *SLC39A11*, *CHD3*, and several other genes.

the 17p13 region (including the D-fragment; yellow), permitting ectopic interaction of several genes and enhancers (Figure S6). Lastly, Hi-C also revealed that the duplicated *CHD3* loses contact with its wild-type downstream region, which may contain regulatory elements essential for its spatio-temporal expression (Figure S4B). *CHD3* is associated with a known DD (Snijders Blok-Campeau syndrome; MIM: 618205),[35] and this might explain part of the individual's symptoms. However, due to the complex nature of these CGRs, other alterations to gene regulation are likely to have an effect on the DD3 phenotype.

In summary, Hi-C of duplications at the SOX9 locus detected by array-CGH revealed a higher degree of rearrangement complexity than was previously thought. Hi-C maps help reconstruct the new genomic architecture and interpret the molecular mechanism involved in the pathogenesis of these phenotypes.

Hi-C Is Highly Effective in Translocation Detection and the Reconstruction of Complex Genomic Breakpoints

Inversions and translocations are copy-number-neutral SVs that can disrupt chromatin domains by rearranging enhancer elements with respect to their cognate target genes, potentially resulting in a regulatory loss and/or gain of function. Here, we sought to investigate whether Hi-C could identify complex SVs in clinical samples harboring chromosomal translocations previously detected through the use of karyotyping. Hi-C maps were generated from LCLs of six individuals with DD carrying reciprocal chromosomal translocations (Table 1). Hi-C readily shows the translocations in all cases (Figure 3). We also asked if the detected SVs (from DD1 to DD9) could be observed in the Hi-C maps when using lower sequencing depth. After reducing the number of initial paired-end reads for all maps (from 100 M down to 1 M sequenced fragments), we could visualize the tandem duplications with 50 M, the intra-chromosomal insertions with 10 M, and, surprisingly, the reciprocal translocations with only 1 M sequenced fragments (Figure S7). We also compared the breakpoints estimated from binned Hi-C data and the breakpoints provided by WGS, yielding on average a distance of 3.5 kb (143 bp to 8.6 kb) between both approaches (Table S2). When switching from binned Hi-C maps to restriction fragment based Hi-C maps, the estimation of the breakpoint may be further improved (Table S2). Interestingly, the WGS SV callers Delly, PopDel, and ERDS failed to detect the 3.1 Mb microduplication from DD2, the 2.4 Mb and the 49 kb duplications from DD3, one reciprocal translocation (DD5), and one breakpoint in the CGR of DD6 (Table 1).

We selected two cases in the series of translocations to illustrate the Hi-C results. An individual presenting with dysmorphic features and severe intellectual disability (DD6) harbored a reciprocal translocation 46,XX,t(3;5)(q24;q14.3) previously detected through the use of karyotyping. Hi-C from DD6 LCLs confirmed the translocation (Figure 4A and B) but revealed a much more complex structure of the derivative chromosomes 3 and 5. For simplification, we named the four fragments on chromosome 5 "A to D" based on their breakpoint location. The Hi-C trans-map indicated a direct contact of fragment B with chr3. The most likely explanation is that this fragment is inserted in reverse orientation after fragment C (Figure 4C). At the region around the breakpoint on chr5 (Figure 4D), we observed the CGRs shown in the trans-map (Figure 4B). The estimated breakpoint chr5:88,045 kb disrupts the *MEF2C* ORF. The fragment B (light gray), containing *MEF2C* and several known enhancers, was inserted downstream of fragment C, however, it was inserted in an inverted orientation (arrow on Figure 4D and 4E). Thus, the shuffled-TAD contains only a disrupted *MEF2C*. Indeed, RT-qPCR of DD6 LCLs revealed reduced expression (50%) of *MEF2C* compared to controls (Figure 4F). Haploinsufficiency of *MEF2C* causes severe ID, epilepsy, and/or cerebral malformations (MIM: 613443), and this makes it a likely candidate for the patient's phenotype.[36]

In a second case, the molecular mechanism of DD5 was solved with the help of Hi-C. In brief, the translocation in DD5 disrupts the FOXG1 TAD on 14q12, separating this gene from several of its known brain enhancers (Figure S8).[37,38] *FOXG1* haploinsufficiency is associated with the congenital variant of Rett syndrome (MIM: 613454), and several non-coding SVs that disrupt *FOXG1* downstream enhancers have been identified in individuals with Rett syndrome-like phenotypes.[37,38] Therefore, loss of interactions between *FOXG1* and its regulatory elements caused by the reciprocal translocation putatively causes a regulatory loss of function, thus explaining the phenotype of DD5.

The remaining three unsolved translocation cases, individuals with intellectual disability and development delay, are described in Figures S9 and S10. Although the exact molecular mechanisms were not completely solved in these cases, the translocation breakpoints are located inside TADs, potentially disrupting wild-type EPI (i.e., loss-of-function) or/and creating enhancer adoption by TAD reshuffling (i.e., gain of function). In two cases (DD7 and DD9), genes involved in neurological disorders and highly expressed in the brain (e.g., *SATB2* and *CTNNA2*) are likely to be involved in the patients' phenotypes. In summary, Hi-C effectively maps reciprocal translocations and helps identify likely pathogenic mechanisms associated with the phenotypes in two to four (out of five) unsolved translocation cases.

Discussion

High throughput technologies are revolutionizing the field of human genetics, and the introduction of WGS in the clinic as a one-test-for-all promises to replace all other molecular technologies in the near future.[1–5] Although WGS has been shown to outperform exome sequencing for SNV detection,[39] substantial challenges remain for the

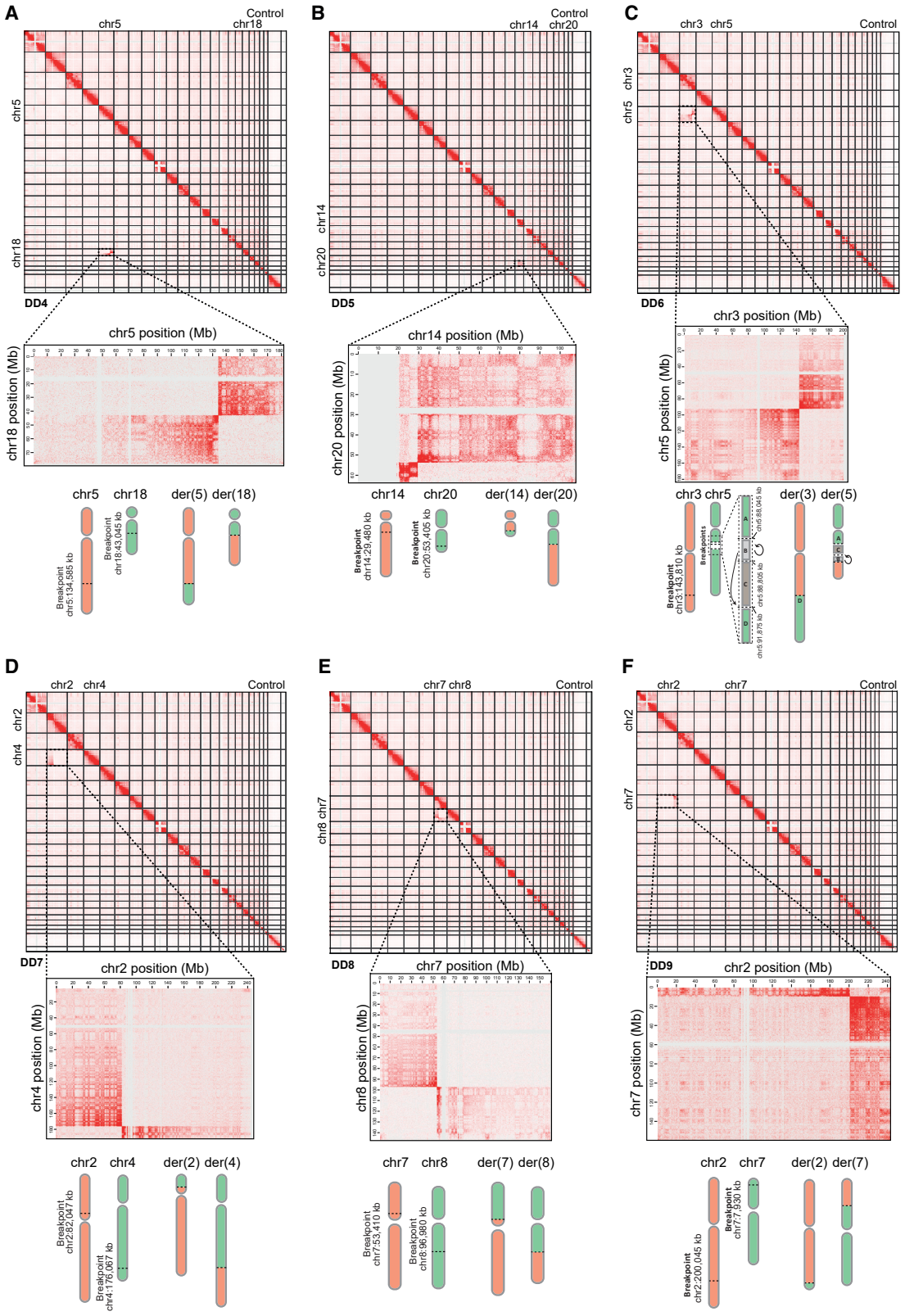

Figure 3. Reciprocal Translocations Are Readily Detected by Hi-C
(A–F) First panel: Hi-C maps of LCLs from individuals DD4–DD9 detect the reciprocal translocations in each case. Second panel: Reciprocal translocations are visible as bow-tie-like patterns in the Hi-C trans-maps (250 kb resolution). Third panel: Schematic representation of the derivative chromosomes based on the pattern of the signal intensity from the trans-map.

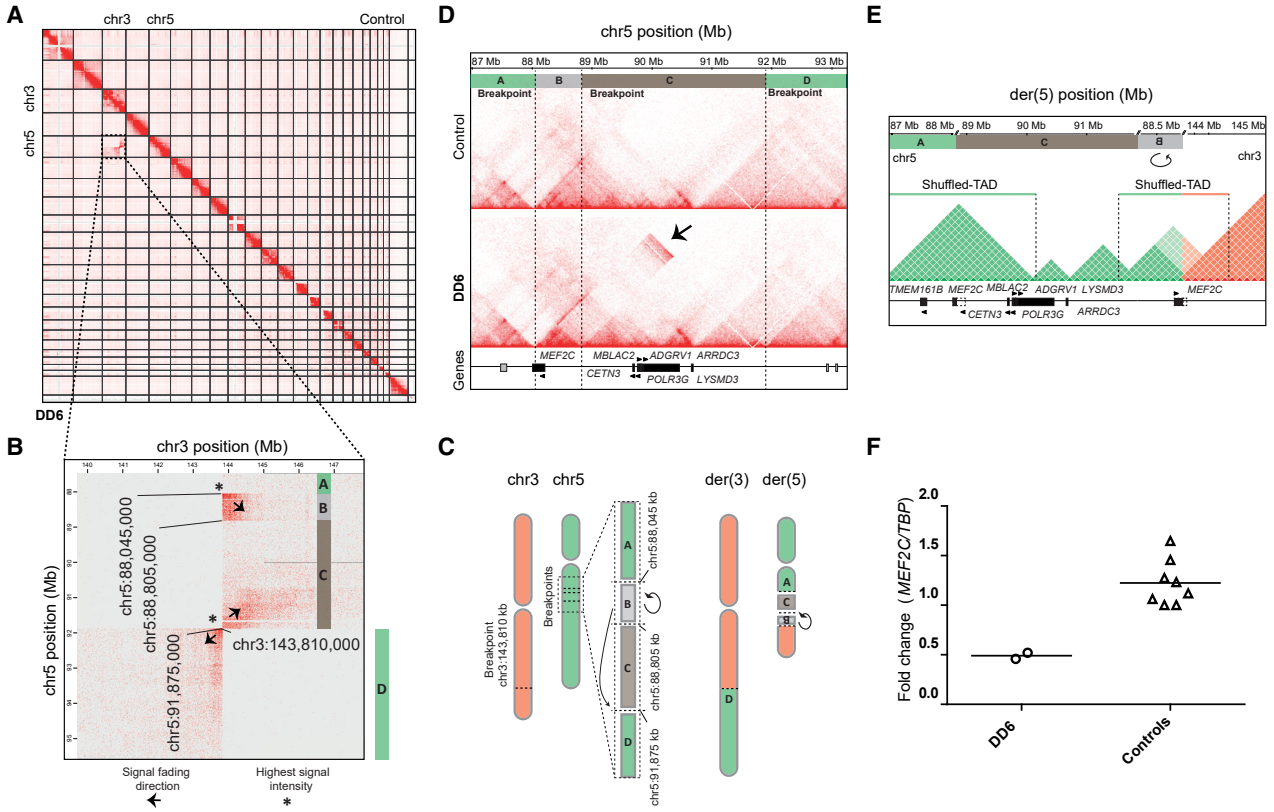

Figure 4. Hi-C Maps Showing a Reciprocal Translocation t(3;5) with a CGR Changing the 3D Landscape of the Derivative Chromosomes

(A) Hi-C map from LCLs of individual DD6 confirmed the t(3;5).

(B) Hi-C detected one breakpoint on chr3 and three on chr5 (chr5:88,045 kb; chr5:88,805 kb; and chr5:91,875 kb). Based on the gradient of the Hi-C signal observed in the trans-map, we could infer that fragment B is inserted downstream of fragment C, also inversely oriented and fused to chr3 (250 kb resolution; raw count map).

(C) Schematics of the derivative chromosomes based on the breakpoints and fragment orientations inferred from Hi-C bin signal from Figure 4B. The der(5) harbors a CGR composed of an insertion and inversion.

(D) The cis-map of 5q14.3 (25 kb resolution, KR-normalized). Three breakpoints on 5q14.3 can be observed and are represented by four fragments on chr5 (red—A, light gray—B, brown—C, and red—D). Note the inversion of the fragment B on the cis-map of DD6 (arrow).

(E) Schematic representation of der(5) showing two novel shuffled-TADs, one containing several genes within the same domain and the other devoid of an ORF.

(F) RT-qPCR showing half the expression of *MEF2C* in DD6 compared to controls.

detection and interpretation of complex SVs with multiple breakpoints. These SVs, including duplications, deletions, translocations, insertions, and inversions, have the potential to disrupt higher-order chromatin organization, thereby rewiring the complex 3D chromatin organization of a locus.[9] This may lead to the repositioning of TAD boundaries and/or the relocation of enhancer elements into other regulatory landscapes, causing gene misexpression and in some cases a deleterious phenotype.[14,15,40,41] Here, we set out to use Hi-C for the identification and interpretation of SVs in clinical samples. As proof of concept, we therefore performed Hi-C in nine samples from individuals with DD, samples with known large SVs detected by commonly used diagnostic cytogenetic and/or cytogenomic tools, and we interrogated whether Hi-C would be useful to detect, resolve, and interpret these variants. Indeed, Hi-C could show all previously known SVs plus a number of additional chromosomal rearrangements, revealing a high degree of complexity in some cases that had remained undetected with conventional methods. Due to these structural changes, the Hi-C maps enabled us to reconstruct the altered 3D genomic architecture. Based on this, we were able to identify regions in which the formation of aberrant regulatory interactions presumably contributed to the phenotypes of the screened individuals (five out of nine).

Our data show that Hi-C is a powerful tool for detecting SVs in a clinical setting. Compared to clinical array CGH, Hi-C has better resolution (3–5 kb in Hi-C compared to around 30–50 kb in array CGH), plus Hi-C can also identify copy number neutral variants, such as inversions and translocations, and add positional information to CNVs.[11,17] Here, for instance, we identified duplications in tandem (DD1 and DD2), duplications and an insertion (DD3), and a translocation with an inversion (DD6). Hi-C also emerges as a versatile tool that can detect large SVs through the use of lower sequencing resources, whereby we were able to spot reciprocal translocations in

reduced sequencing-depth maps down to 1 M fragments (Figure S7). It is noteworthy that the WGS SV callers used in this study failed to detect three out of nine main SVs in our cohort. Although Hi-C has been used before to identify large SVs in primary brain tumor samples[42] and B cells lymphomas[43] and in combination with optical mapping and WGS in other cancer cell lines,[17] the potential of Hi-C for the investigation of the genetic causes of DDs has not been evaluated yet. Our data indicate that Hi-C has a large diagnostic potential. Low-coverage Hi-C could be used in patient cells as an alternative to array CGH or as a secondary tool for excluding balanced SVs instead of classical karyotyping. Hi-C could also be extremely useful in the field of cancer cytogenetics[17,42,43] and in precision medicine.[44]

Our study also has several limitations. First, the limited resolution in our binned Hi-C maps of around 5 kb, the high noise level in Hi-C, and the effects of binning short reads from ligated restriction fragments to a regular grid allow for only a rough estimate of breakpoints and do not show specific enhancer promoter contacts in the newly formed TADs. Improvements in algorithms and tools for data analysis may solve these limitations,[8] e.g., analyzing restriction-fragment-based Hi-C maps could improve the estimation of the breakpoint (Table S2). Second, the computational analysis for Hi-C is still not as user-friendly as with current array CGH and WGS tools. Third, better algorithms should be developed and tested to automatize the detection, in a genome-wide perspective, of all SVs and ectopic Hi-C contacts in samples from clinical individuals.[8,17,45,46]

In summary, our study illustrates the importance of considering not only the linear sequence of the genome, but also the 3D chromatin structure when interpreting the impact of SVs. We show that Hi-C is a powerful tool useful for detecting and resolving the structure of complex chromosomal rearrangements in DDs. If developed further, Hi-C may become a method of choice for efficient SV detection in cases with suspected genetic causes. To gain an even deeper understanding of the complex nature of SVs, we propose combining Hi-C with WGS, long read sequencing, and/or optical mapping.[17] The combination of these powerful tools will allow for a more precise determination of cryptic rearrangements and enables more accurate predictions of phenotypic consequences.

Data and Code Availability

The accession numbers for the CNVs coordinates reported in this manuscript are in DECIPHER: 412419, 412420, 412421.

Supplemental Data

Supplemental Data can be found online at https://doi.org/10.1016/j.ajhg.2020.04.016.

Acknowledgments

We would like to thank the individuals and families for their collaboration and contribution to this project. Technical assistance: We thank Susanne Rothe, Vanessa Suckow, and Celina São-José for their excellent work. We thank Angela Maria Vianna-Morgante for assistance in reviewing the cytogenetics nomenclature. Funding: M.S. and S.M. are supported by grants from the Deutsche Forschungsgemeinschaft (DFG) (SP1532/3-1, SP1532/4-1, SP1532/5-1, and MU 880/16-1) and the Max Planck Foundation. U.S.M. is a fellow of the Capes-Alexander von Humboldt Foundation.

Declaration of Interests

Rocio Acuna-Hidalgo is a founder, shareholder, and full-time employee of Nostos Genomics.

Conflict of Interest

The other authors declare no competing interests.

Received: October 8, 2019
Accepted: April 29, 2020
Published: May 28, 2020

Web Resources

ClinGen CNV Pathogenicity Calculator, http://cnvcalc.clinicalgenome.org/cnvcalc/cnv-gain
DECIPHER, https://decipher.sanger.ac.uk/
Mouse Genome Informatics, http://www.informatics.jax.org/
Mouse Phenotype, https://www.mousephenotype.org/
Online Mendelian Inheritance in Man (OMIM), https://www.omim.org/

References

1. Goodwin, S., McPherson, J.D., and McCombie, W.R. (2016). Coming of age: ten years of next-generation sequencing technologies. Nat. Rev. Genet. *17*, 333–351.
2. Vissers, L.E.L.M., van Ravenswaaij, C.M.A., Admiraal, R., Hurst, J.A., de Vries, B.B.A., Janssen, I.M., van der Vliet, W.A., Huys, E.H.L.P.G., de Jong, P.J., Hamel, B.C.J., et al. (2004). Mutations in a new member of the chromodomain gene family cause CHARGE syndrome. Nat. Genet. *36*, 955–957.
3. Ng, S.B., Buckingham, K.J., Lee, C., Bigham, A.W., Tabor, H.K., Dent, K.M., Huff, C.D., Shannon, P.T., Jabs, E.W., Nickerson, D.A., et al. (2010). Exome sequencing identifies the cause of a mendelian disorder. Nat. Genet. *42*, 30–35.
4. Bamshad, M.J., Ng, S.B., Bigham, A.W., Tabor, H.K., Emond, M.J., Nickerson, D.A., and Shendure, J. (2011). Exome sequencing as a tool for Mendelian disease gene discovery. Nat. Rev. Genet. *12*, 745–755.
5. Chong, J.X., Buckingham, K.J., Jhangiani, S.N., Boehm, C., Sobreira, N., Smith, J.D., Harrell, T.M., McMillin, M.J., Wiszniewski, W., Gambin, T., et al.; Centers for Mendelian Genomics (2015). The Genetic Basis of Mendelian Phenotypes: Discoveries, Challenges, and Opportunities. Am. J. Hum. Genet. *97*, 199–215.

6. Feuk, L., Carson, A.R., and Scherer, S.W. (2006). Structural variation in the human genome. Nat. Rev. Genet. *7*, 85–97.
7. Sanchis-Juan, A., Stephens, J., French, C.E., Gleadall, N., Mégy, K., Penkett, C., Shamardina, O., Stirrups, K., Delon, I., Dewhurst, E., et al. (2018). Complex structural variants in Mendelian disorders: identification and breakpoint resolution using short- and long-read genome sequencing. Genome Med. *10*, 95.
8. Wang, S., Lee, S., Chu, C., Jain, D., Nelson, G., Walsh, J.M., et al. (2020). HiNT: a computational method for detecting copy number variations and translocations from Hi-C data. Genome Biology *21* ((1):73), In press. https://doi.org/10.1186/s13059-020-01986-5.
9. Spielmann, M., Lupiáñez, D.G., and Mundlos, S. (2018). Structural variation in the 3D genome. Nat. Rev. Genet. *19*, 453–467.
10. Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science *326*, 289–293.
11. Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., and Aiden, E.L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665–1680.
12. Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376–380.
13. Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature *485*, 381–385.
14. Lupiáñez, D.G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J.M., Laxova, R., et al. (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell *161*, 1012–1025.
15. Franke, M., Ibrahim, D.M., Andrey, G., Schwarzer, W., Heinrich, V., Schöpflin, R., Kraft, K., Kempfer, R., Jerković, I., Chan, W.-L., et al. (2016). Formation of new chromatin domains determines pathogenicity of genomic duplications. Nature *538*, 265–269.
16. Flöttmann, R., Kragesteen, B.K., Geuer, S., Socha, M., Allou, L., Sowińska-Seidler, A., Bosquillon de Jarcy, L., Wagner, J., Jamsheer, A., Oehl-Jaschkowitz, B., et al. (2018). Noncoding copy-number variations are associated with congenital limb malformation. Genet. Med. *20*, 599–607.
17. Dixon, J.R., Xu, J., Dileep, V., Zhan, Y., Song, F., Le, V.T., Yardımcı, G.G., Chakraborty, A., Bann, D.V., Wang, Y., et al. (2018). Integrative detection and analysis of structural variation in cancer genomes. Nat. Genet. *50*, 1388–1398.
18. Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv, arXiv:1303.3997.
19. Faust, G.G., and Hall, I.M. (2014). SAMBLASTER: fast duplicate marking and structural variant read extraction. Bioinformatics *30*, 2503–2505.
20. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R.; and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.
21. Rausch, T., Zichner, T., Schlattl, A., Stütz, A.M., Benes, V., and Korbel, J.O. (2012). DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics *28*, i333–i339.
22. Roskosch, S., Jónsson, H., Björnsson, E., Beyter, D., Eggertsson, H.P., Sulem, P., Stefánsson, K., Halldórsson, B.V., and Kehr, B. (2019). PopDel identifies medium-size deletions jointly in tens of thousands of genomes. bioRxiv. https://doi.org/10.1101/740225.
23. Zhu, M., Need, A.C., Han, Y., Ge, D., Maia, J.M., Zhu, Q., Heinzen, E.L., Cirulli, E.T., Pelak, K., He, M., et al. (2012). Using ERDS to infer copy-number variants in high-coverage genomes. Am. J. Hum. Genet. *91*, 408–421.
24. Smedley, D., Jacobsen, J.O., Jäger, M., Köhler, S., Holtgrewe, M., Schubach, M., Siragusa, E., Zemojtel, T., Buske, O.J., Washington, N.L., et al. (2015). Next-generation diagnostics and disease-gene discovery with the Exomiser. Nat. Protoc. *10*, 2,004–2,015.
25. Hombach, D., Schuelke, M., Knierim, E., Ehmke, N., Schwarz, J.M., Fischer-Zirnsak, B., and Seelow, D. (2019). MutationDistiller: user-driven identification of pathogenic DNA variants. Nucleic Acids Res. *47* (W1), W114–W120.
26. Durand, N.C., Shamim, M.S., Machol, I., Rao, S.S.P., Huntley, M.H., Lander, E.S., and Aiden, E.L. (2016a). Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. Cell Syst. *3*, 95–98.
27. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.
28. Knight, P.A., and Ruiz, D. (2013). A fast algorithm for matrix balancing. IMA J. Numer. Anal. *33*, 1029–1047.
29. Durand, N.C., Robinson, J.T., Shamim, M.S., Machol, I., Mesirov, J.P., Lander, E.S., and Aiden, E.L. (2016b). Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. Cell Syst. *3*, 99–101.
30. ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.
31. Schmittgen, T.D., and Livak, K.J. (2008). Analyzing real-time PCR data by the comparative C(T) method. Nat. Protoc. *3*, 1101–1108.
32. Kurth, I., Klopocki, E., Stricker, S., van Oosterwijk, J., Vanek, S., Altmann, J., Santos, H.G., van Harssel, J.J., de Ravel, T., Wilkie, A.O., et al. (2009). Duplications of noncoding elements 5′ of SOX9 are associated with brachydactyly-anonychia. Nat. Genet. *41*, 862–863.
33. Plaster, N.M., Tawil, R., Tristani-Firouzi, M., Canún, S., Bendahhou, S., Tsunoda, A., Donaldson, M.R., Iannaccone, S.T., Brunt, E., Barohn, R., et al. (2001). Mutations in Kir2.1 cause the developmental and episodic electrical phenotypes of Andersen's syndrome. Cell *105*, 511–519.
34. Xia, M., Jin, Q., Bendahhou, S., He, Y., Larroque, M.-M., Chen, Y., Zhou, Q., Yang, Y., Liu, Y., Liu, B., et al. (2005). A Kir2.1 gain-of-function mutation underlies familial atrial fibrillation. Biochem. Biophys. Res. Commun. *332*, 1012–1019.
35. Snijders Blok, L., Rousseau, J., Twist, J., Ehresmann, S., Takaku, M., Venselaar, H., Rodan, L.H., Nowak, C.B., Douglas, J., Swoboda, K.J., et al.; DDD study (2018). CHD3 helicase domain mutations cause a neurodevelopmental syndrome with macrocephaly and impaired speech and language. Nat. Commun. *9*, 4619.

36. Redin, C., Brand, H., Collins, R.L., Kammin, T., Mitchell, E., Hodge, J.C., Hanscom, C., Pillalamarri, V., Seabra, C.M., Abbott, M.A., et al. (2017). The genomic landscape of balanced cytogenetic abnormalities associated with human congenital anomalies. Nat. Genet. *49*, 36–45.
37. Allou, L., Lambert, L., Amsallem, D., Bieth, E., Edery, P., Destrée, A., Rivier, F., Amor, D., Thompson, E., Nicholl, J., et al. (2012). 14q12 and severe Rett-like phenotypes: new clinical insights and physical mapping of FOXG1-regulatory elements. Eur. J. Hum. Genet. *20*, 1216–1223.
38. Ibn-Salem, J., Köhler, S., Love, M.I., Chung, H.R., Huang, N., Hurles, M.E., Haendel, M., Washington, N.L., Smedley, D., Mungall, C.J., et al. (2014). Deletions of chromosomal regulatory boundaries are associated with congenital disease. Genome Biol. *15*, 423.
39. Lelieveld, S.H., Spielmann, M., Mundlos, S., Veltman, J.A., and Gilissen, C. (2015). Comparison of exome and genome sequencing technologies for the complete capture of protein-coding regions. Hum. Mutat. *36*, 815–822.
40. Kragesteen, B.K., Spielmann, M., Paliou, C., Heinrich, V., Schöpflin, R., Esposito, A., Annunziatella, C., Bianco, S., Chiariello, A.M., Jerković, I., et al. (2018). Dynamic 3D chromatin architecture contributes to enhancer specificity and limb morphogenesis. Nat. Genet. *50*, 1463–1473.
41. Kraft, K., Magg, A., Heinrich, V., Riemenschneider, C., Schöpflin, R., Markowski, J., Ibrahim, D.M., Acuna-Hidalgo, R., Despang, A., Andrey, G., et al. (2019). Serial genomic inversions induce tissue-specific architectural stripes, gene misexpression and congenital malformations. Nat. Cell Biol. *21*, 305–310.
42. Harewood, L., Kishore, K., Eldridge, M.D., Wingett, S., Pearson, D., Schoenfelder, S., Collins, V.P., and Fraser, P. (2017). Hi-C as a tool for precise detection and characterisation of chromosomal rearrangements and copy number variation in human tumours. Genome Biol. *18*, 125.
43. Díaz, N., Kruse, K., Erdmann, T., Staiger, A.M., Ott, G., Lenz, G., and Vaquerizas, J.M. (2018). Chromatin conformation analysis of primary patient tissue using a low input Hi-C method. Nat. Commun. *9*, 4938.
44. Li, Y., Tao, T., Du, L., and Zhu, X. (2020). Three-dimensional genome: developmental technologies and applications in precision medicine. J. Hum. Genet. *65*, 497–511.
45. Chakraborty, A., and Ay, F. (2018). Identification of copy number variations and translocations in cancer cells from Hi-C data. Bioinformatics *34*, 338–345.
46. Stansfield, J.C., Cresswell, K.G., Vladimirov, V.I., and Dozmorov, M.G. (2018). HiCcompare: an R-package for joint normalization and comparison of HI-C datasets. BMC Bioinformatics *19*, 279.