ORIGINAL ARTICLE

Whole genome paired-end sequencing elucidates functional and phenotypic consequences of balanced chromosomal rearrangement in patients with developmental disorders

Caroline Schluth-Bolard,[1,2] Flavie Diguet,[1,2] Nicolas Chatron,[1,2]
Pierre-Antoine Rollat-Farnier,[1] Claire Bardel,[3] Alexandra Afenjar,[4,5] Florence Amblard,[6]
Jeanne Amiel,[7] Sophie Blesson,[8] Patrick Callier,[9] Yline Capri,[10] Patrick Collignon,[11]
Marie-Pierre Cordier,[1] Christine Coubes,[12] Benedicte Demeer,[13]
Annabelle Chaussenot,[14] Florence Demurger,[15] Françoise Devillard,[6]
Martine Doco-Fenzy,[16] Céline Dupont,[10] Jean-Michel Dupont,[17] Sophie Dupuis-Girod,[1]
Laurence Faivre,[18] Brigitte Gilbert-Dussardier,[19] Anne-Marie Guerrot,[20] Marine Houlier,[7]
Bertrand Isidor,[21] Sylvie Jaillard,[22] Géraldine Joly-Hélas,[23] Valérie Kremer,[24]
Didier Lacombe,[25] Cédric Le Caignec,[21] Aziza Lebbar,[17] Marine Lebrun,[26]
Gaetan Lesca,[1,2] James Lespinasse,[27] Jonathan Levy,[10] Valérie Malan,[28]
Michele Mathieu-Dramard,[13] Julie Masson,[1,2] Alice Masurel-Paulet,[18] Cyril Mignot,[29]
Chantal Missirian,[30] Fanny Morice-Picard,[25] Sébastien Moutton,[25]
Gwenaël Nadeau,[27,31] Céline Pebrel-Richard,[32] Sylvie Odent,[15,33]
Véronique Paquis-Flucklinger,[14] Laurent Pasquier,[15] Nicole Philip,[34] Morgane Plutino,[14]
Linda Pons,[1,2] Marie-France Portnoï,[4] Fabienne Prieur,[26] Jacques Puechberty,[12]
Audrey Putoux,[1,2] Marlène Rio,[7] Caroline Rooryck-Thambo,[25] Massimiliano Rossi,[1,2]
Catherine Sarret,[35] Véronique Satre,[6,36] Jean-Pierre Siffroi,[4] Marianne Till,[1]
Renaud Touraine,[26] Annick Toutain,[8] Jérome Toutain,[25] Stéphanie Valence,[5,37]
Alain Verloes,[10] Sandra Whalen,[4] Patrick Edery,[1,2] Anne-Claude Tabet,[10]
Damien Sanlaville[1,2]

Check for updates

ABSTRACT

Background Balanced chromosomal rearrangements associated with abnormal phenotype are rare events, but may be challenging for genetic counselling, since molecular characterisation of breakpoints is not performed routinely. We used next-generation sequencing to characterise breakpoints of balanced chromosomal rearrangements at the molecular level in patients with intellectual disability and/or congenital anomalies.

Methods Breakpoints were characterised by a paired-end low depth whole genome sequencing (WGS) strategy and validated by Sanger sequencing. Expression study of disrupted and neighbouring genes was performed by RT-qPCR from blood or lymphoblastoid cell line RNA.

Results Among the 55 patients included (41 reciprocal translocations, 4 inversions, 2 insertions and 8 complex chromosomal rearrangements), we were able to detect 89% of chromosomal rearrangements (49/55). Molecular signatures at the breakpoints suggested that DNA breaks arose randomly and that there was no major influence of repeated elements. Non-homologous end-joining appeared as the main mechanism of repair (55% of rearrangements). A diagnosis could be established in 22/49 patients (44.8%), 15 by gene disruption (*KANSL1*, *FOXP1*, *SPRED1*, *TLK2*, *MBD5*, *DMD*, *AUTS2*, *MEIS2*, *MEF2C*, *NRXN1*, *NFIX*, *SYNGAP1*, *GHR*, *ZMIZ1*) and 7 by position effect (*DLX5*, *MEF2C*, *BCL11B*, *SATB2*, *ZMIZ1*). In addition, 16 new candidate genes were identified. Systematic gene expression studies further supported these results. We also showed the contribution of topologically associated domain maps to WGS data interpretation.

Conclusion Paired-end WGS is a valid strategy and may be used for structural variation characterisation in a clinical setting.

INTRODUCTION

Structural variations (SV) are rearrangements of chromosome architecture that may be benign or pathogenic. They were first identified on the basis of karyotype in 0.7% of the population.[1] Two groups are distinguished: unbalanced chromosomal rearrangements, such as deletions or duplications,

BMJ

generally associated with an altered phenotype, and apparently balanced chromosomal rearrangements (ABCR), such as reciprocal translocations, inversions and insertions. The latter category is characterised by the absence of gain or loss of genetic material and has usually no phenotypic consequence for the carrier, except reproductive issues such as infertility or miscarriages.[1] However, it has been estimated that up to 27% of these ABCR may be associated with an abnormal phenotype.[2 3] Until recently, the pathophysiology was poorly understood and these cases have been for a long time a challenge for genetic counselling. The development of molecular cytogenetic techniques, such as FISH[4] and chromosomal microarray (CMA),[5] contributed to decipher the underlying aetiologies of abnormal phenotypes in ABCR. In approximately 40% of cases, pathogenic CNV were uncovered.[6] In some other cases, the phenotype could be related to gene disruption[7] or to position effect,[8] a mechanism in which the modification of the chromatin environment alters gene expression. However, the precise characterisation of ABCR breakpoints is time-consuming and rarely proposed in diagnostic settings.

More recently, a further step has been taken by the use of next-generation sequencing that has proved to be a rapid and efficient method to detect SV, including CNVs and balanced SVs.[9] It was used to characterise the breakpoints of ABCR in patients with developmental disorders.[10–16] By allowing the refinement of SV breakpoints at base pair level, these studies shed light on some mechanistic and pathophysiological aspects. In particular, they showed the unexpected complexity of ABCR,[11] the role of repeated elements[13 14] and the predominance of non-homologous end-joining (NHEJ).[11–13 17] They also assessed their impact on genome architecture, especially on topologically associated domains (TADs)[12–14] and their involvement in patient phenotype.[13 14 16] However, studies remain scarce and often included small number of cases except for few of them.[2 11–14 16] In addition, these studies used mainly mate-pair or similar strategies[11–14 16] that were developed for SV detection and are not available in all laboratories. Paired-end sequencing is most widely used in genetics laboratories and may have other applications.[18] We previously demonstrated that paired-end whole genome sequencing (WGS) was also powerful for SV detection.[19] In the study presented herein, we applied this strategy to one of the largest cohorts of patients with intellectual disability and/or multiple congenital anomalies (ID/MCA) associated with ABCR and no pathogenic CNV detected by CMA in order to characterise ABCR breakpoints and study their mechanisms as well as their functional and phenotypic consequences.

MATERIALS AND METHODS

Patients

Fifty-five patients were recruited between April 2015 and February 2017 among 21 French clinical genetics centres as part of the ANI project (ABCR characterisation by next-generation sequencing in patients with ID/MCA, ClinicalTrials NCT02451761). Inclusion criteria were the following: i) abnormal phenotype including intellectual disability and/or congenital anomalies; ii) presence of ABCR on standard karyotype; iii) no pathogenic CNVs detected by CMA (according to French CMA Guidelines V.3.1); iv) ABCR de novo or inherited from a parent presenting the same phenotype. All patients or their parents gave written informed consent for this study, which was conducted with respect to the recommendations of the Helsinki Declaration. All patients had a detailed clinical examination by a trained geneticist.

Whole genome sequencing

Genomic DNA was extracted from EDTA-blood samples with the QIAmp DNA Blood Midikit (Qiagen, Vento, The Netherlands) according to the manufacturer's instructions. Genomic DNA libraries of 350 bp fragments were prepared following the Illumina TruSeq DNA PCR-free protocol (Illumina, San Diego, California, USA) with 3 µg DNA. DNA libraries were sequenced on an Illumina NextSeq 500 as paired-end 101 bp reads using the High Output (300 cycles) NextSeq500 kit, with two patients per flow-cell yielding a mean sequencing depth between 5.53X and 17.91X. For patient MD/0110, carrying a mosaic translocation, a complete flow cell was used with a 24.9X mean sequencing depth. Image analysis and base calling were performed using Illumina Real-Time Analysis Pipeline 2 and bcl2fastq with default parameters. For each sample, an alignment of the reads against the hg19 version of the human genome was done using BWA-MEM V.0.7.10.[20] The reads were then sorted using Samtools V.1.3.1,[21] and the duplicates removed by PicardTools V.1.138. Then, SV were detected using BreakDancer V.1.4.5.[22] The generated outputs were then annotated using an in-house programme, Svagga (Rollat-Farnier et al, https://gitlab.inria.fr/NGS/svagga). For each sample, we used the other samples of the ANI project as reference, in order to remove recurrent variants. A distance of 500 nucleotides was used as the maximal distance for Svagga to consider two breakpoints as identical. Integrative Genomics Viewer V.2.3[23] was used for the SV visualisation and validation. In case of failed variant identification, or discrepancy with the corresponding karyotype, the same pipeline was used but with an alignment against the hg38 version of the human genome.

PCR amplification of junction fragment and Sanger sequencing

Each breakpoint identified by WGS was confirmed by PCR and Sanger sequencing. Junction fragments were amplified using the Taq DNA Core kit 10 (MP Biomedicals, Solon, Ohio, USA), from patient DNA. A DNA from an individual who was not a carrier of chromosomal rearrangement was included as negative control; amplification with the primer pair for the *ATP1A3* gene (exons 7–8) was used as positive control. Specific PCR products corresponding to the junction fragments were sequenced by the Sanger method. Breakpoint sequences were aligned to the reference genome hg19 using BLAT tool (UCSC) in order to determine the breakpoint coordinates. For some complex rearrangements, FISH on lymphocyte metaphase spreads was performed according to standard procedures.

CNV validation by qPCR

CNVs associated with the chromosomal rearrangement were confirmed by real-time qPCR. It was performed according to the manufacturer's recommendations with QuantiTect SYBR Green PCR kit (Qiagen, Courtaboeuf, France) on a Light Cycler 2000 (Roche Applied Science, Indianapolis, Indiana, USA) using specific primers amplifying a unique sequence within the CNV and *ADORA2B* as reference gene.

Expression studies

Blood RNA was extracted from Paxgene samples according to the manufacturer's instruction (Qiagen). Lymphoblastoid cell lines (LCL) RNA was extracted using the RNeasy plus mini kit (Qiagen). RT was performed using 500 ng RNA with the Superscript II RT kit (Invitrogen, Life Technologies, Carlsbad,

California, USA). Real-time PCR was performed using 1/20 dilutions of cDNAs with the QuantiTect SYBR Green PCR kit (Qiagen) on a LightCycler 2000 (Roche Applied Science) in triplicate. Primers were designed for each disrupted gene and genes adjacent to the breakpoints; *ACTB* was used as reference gene. Samples included patient and two to four sex-matched and tissue-matched control cDNAs, as well as RT-negative RT products. Data were analysed by relative quantification according to the 2-ΔΔCt method.[24]

For fusion transcript amplification, primer pairs were chosen on each side of the putative mRNA junction. The fusion was then amplified by PCR from cDNA with Taq DNA Core kit 10 (MP Biomedicals).

All primer sequences are available on request.

Determination of breakpoints characteristics

ABCR were classified according to the number of breakpoints assessed by WGS in simple rearrangements (2 break rearrangements), complex rearrangements (3 to 10 breakpoints) and chromoanagenesis (>10 breakpoints with clustering of the breakpoints).[11] Disruption of gene, TADs and repeated elements were established according to UCSC (UCSC genes track), 3D Genome Browser (GM12878 cells[25]) and RepeatMasker, respectively. Molecular signatures included deletion, duplication, microhomology defined as a series of nucleotides (<70) that were identical at the junctions of the two genomic segments that contributed to the rearrangement[26] and templated insertions defined as insertions originating from nearby segments that contributed to the rearrangement.[27] The mechanisms were defined at the junction level according to the

following criteria. Non-homologous end-joining was defined by the presence of blunt ends, small deletions or duplications, microhomology not exceeding 4 nucleotides, insertions of <10 nucleotides.[28] Microhomology-mediated end-joining was considered in junction showing microhomology, deletions of >10 nucleotides and templated insertion of >10 nucleotides.[28] Replicative mechanisms (including microhomology-mediated break-induced replication [MMBIR] and fork stalling and template switching [FoSTeS]) were considered in junction showing microhomology, templated insertions and possible gains or losses of nucleotides.[26] Pathogenicity of ABCR was assessed according to the criteria used by Redin *et al*.[13] For position effect, genes located in the same TAD as the breakpoint were taken into account.

Statistical analyses

Statistical analyses were performed using the R software V.3.3.3.[29] The proportion of breakpoints disrupting genes, located in TADs or in repeated elements, were compared with reference proportions calculated from the following databases: UCSC genes, 3D Genome Browser (GM12878 cells[25]) and RepeatMasker using a standard one sample test to compare proportions. The distribution of breakpoints in the different classes of repeated element was compared with the proportion of the reference genome in the different classes of repeated elements using a X^2 'goodness-of-fit' test. Five classes were studied: SINE, LINE, LTR, DNA and satellite+segmental duplications. The correlation between the expression level and the absolute value of the distance to the breakpoint was estimated and tested with the R package rmcorr.[30] The R package
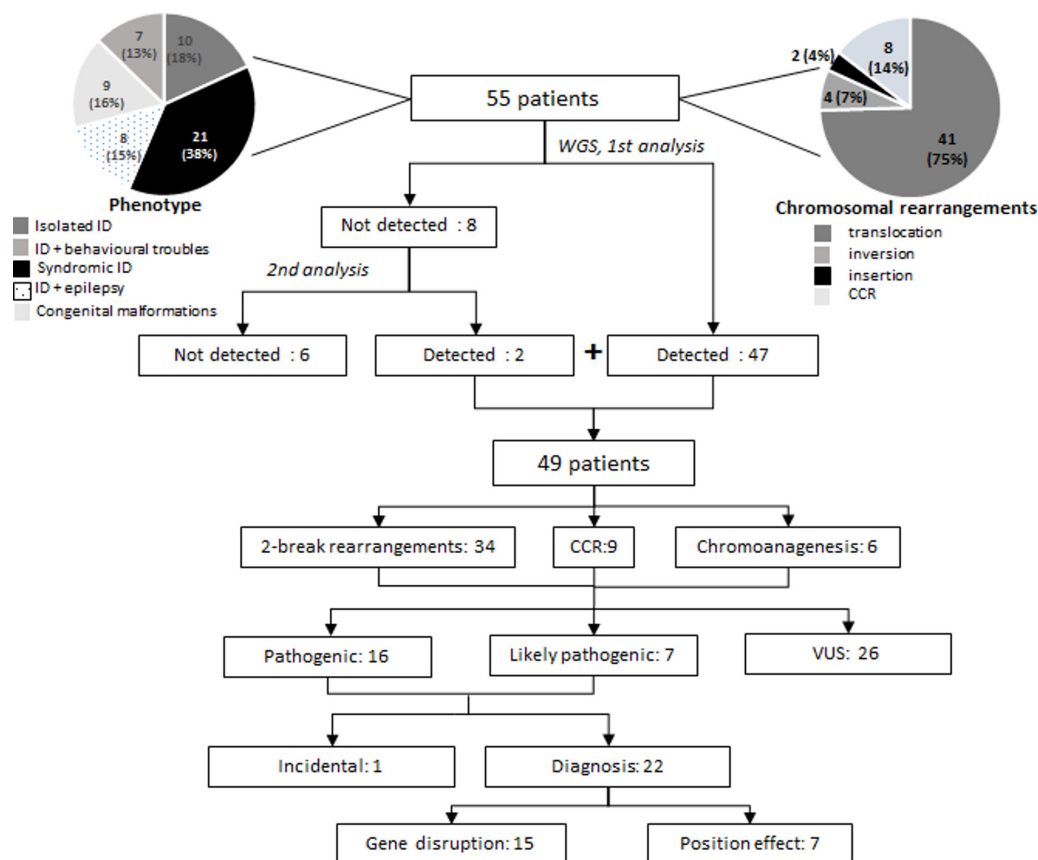

Figure 1 Study flow chart. ID, intellectual disability; CCR, complex chromosomal rearrangement; VUS, variants of unknown significance; WGS, whole genome sequencing.

nlme was used to compare the mean expression level of the genes located in the same TAD as the breakpoint, in an adjacent TAD or in an inter-TAD region.

RESULTS
Population characteristics
In the present study, 55 patients were included; there were 23 males and 32 females (sex ratio 0.7). They were aged between 1 and 44 years (mean 13.1 years, SD 9.6). They all presented with an abnormal karyotype (figure 1). In 49 patients, ABCR arose de novo; in 6 patients, ABCR was familial and cosegregated with the phenotype. Eighty-seven per cent of the patients (48/55) presented with isolated or syndromic intellectual disability. Seven patients (13%) had congenital malformations without intellectual disability. Details of patients' phenotype and karyotype are available in online supplementary table S1 and figure S1.

Yield of paired-end whole genome sequencing for breakpoint characterisation
In order to characterise the breakpoint sequences of these rearrangements, a paired-end WGS strategy was applied. It allowed the detection of the rearrangements in 49/55 patients (89%)

(figure 1, online supplementary data S1 and table S2). Chromosomal rearrangements were detected with the standard pipeline in 46 patients, including a mosaic translocation (MD/0110) and rearrangements involving alpha-satellite sequences (MD/0104) or segmental duplications (JP/0107, IL/1901). A second analysis (including hg38 alignment and/or focused analysis on breakpoint region defined by karyotype) was required to detect rearrangements in three additional patients (CM/0103, VD/2401, JE/1401). For two patients carrying complex SV (NM/0201, KT/1403), WGS was not able to resolve the entire complexity of the breakpoints. In six patients, WGS did not detect the chromosomal rearrangements (EB/0501, MG/1001, ML/1402, VL/1102, LD/0108, PS/0502). In all these problematic cases, at least one chromosomal breakpoint, as defined by karyotype, involved highly repetitive sequences or sequence gaps: short arms of acrocentric chromosomes (CM/0103, MG/1001, ML/1402, KT/1403), alpha-satellite regions (PS/0502), constitutive heterochromatin (EB/0501), subtelomeric regions (VL/1102, LD/0108) or segmental duplications (VD/2401, JE/1401, NM/0201).

In 9/49 patients, additional CNVs (loss and/or gain) were detected at the breakpoint. The CNV size ranged from 2 to
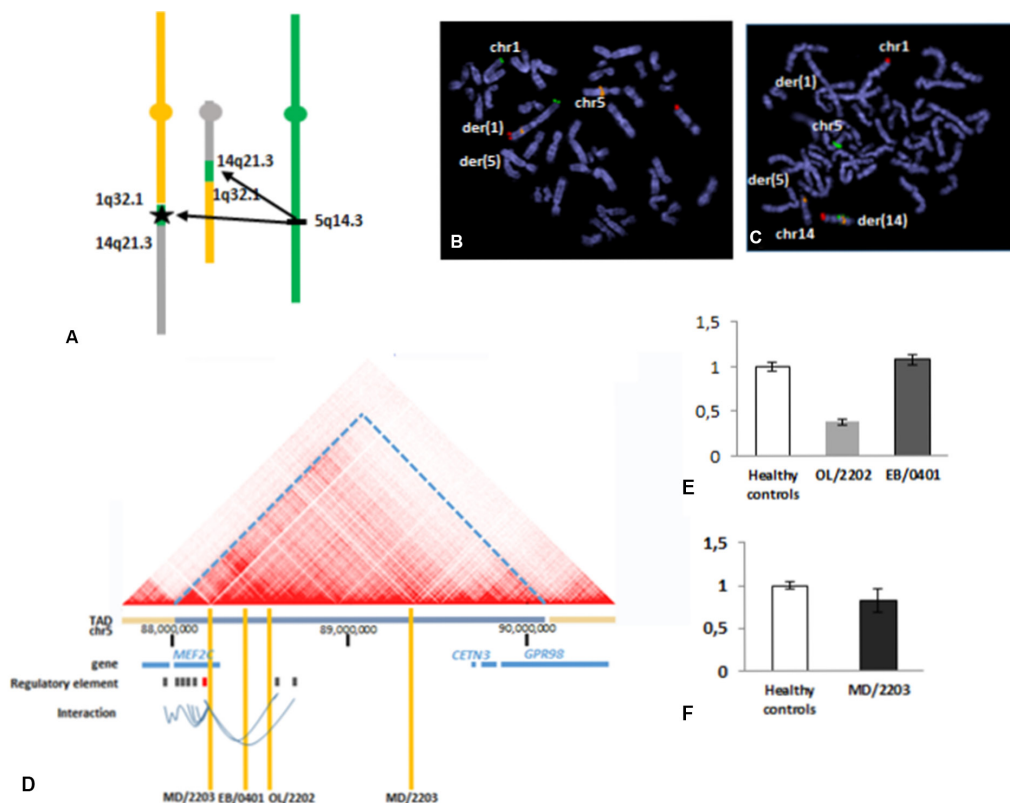

Figure 2 Characterisation of a reciprocal translocation t(1;14)(q32;q22) uncovered additional complexity that accounted for the patient's phenotype (OL/2202). (A) Schematic representation of the rearrangement according to WGS result: insertion of chromosomal fragments of 153 and 736 kb from the 5q14.3 region to the breakpoint of derivative 1 and to the breakpoint of derivative 14, respectively. The black star indicates the localisation of *MEF2C*. Confirmation of the insertions by FISH on metaphase spread. (B) Insertion of 5q14.3 region in the derivative 1: RP11-484D1(5q14.3) (rhodamine) (BlueFish), 1pter (FITC) (Cytocell), 14qter (Texas Red) (Cytocell). (C) insertion of 5q14.3 region in the derivative 14: RP11-109H16 (5q14.3) (rhodamine) (RainbowFish), RP11-120I18 (14q12) (FITC), 1qter (Texas Red) (Cytocell). (D) Three-dimensional (3D)-genome map at the 5q14.3 locus derived from Hi-C data of GM12878[25] (10 kb resolution). TAD, topologically associated domain (3D-genome browser); regulatory elements for *MEF2C*: defined according to GeneHancer, enhancer in grey, promoter in red; interactions between regulatory element and genes according to GeneHancer. Yellow bars indicate patients' breakpoint in 5q14.3 locus. For patients OL/2202 and EB/0401, breakpoints disrupt interaction between enhancer and *MEF2C*. For patient MD/2203, proximal breakpoint disrupt *MEF2C* gene whereas the distal breakpoint has no consequence on regulatory elements. (E) Expression of *MEF2C* in blood cells for patients OL/2202 and EB/0401 compared with healthy controls. (F) Expression of *MEF2C* in lymphoblastoid cell line for patient MD/2203 compared with healthy controls.

Table 1 Genomic features disrupted by chromosomal breakpoints

Disrupted elements	Breakpoints			
	Simple (n=68)	CCR (n=45)	Chromoanagenesis (n=105)	Total 218
Gene	34	21	40	95 (43.5%)
TAD	60	39	93	192 (88%)
Repeated elements	30	26	54	110 (50.4%)
SINE	7	7	13	27
LINE	11	12	20	43
LTR	4	6	14	24
DNA	5	1	4	10
Satellite	1	0	1	2
Segmental duplication	2	0	2	4

Repeated elements were defined according to Repeat Masker. TAD were defined according to GM12878 map.[25]

CCR, complex chromosomal rearrangements; TAD, topologically associated domain.

Table 2 Molecular signatures at junction sequences

Molecular signatures	Junctions			
	Simple (n=68)	CCR (n=44)	Chromoanagenesis (n=104)	Total 216
Deletion (1–26 355 bp)	48	30	52	130 (60.1%)
1–10 bp	33	14	28	75
11–1000 bp	10	12	18	40
>1000 bp	5	4	6	15
Gain (1–71 339 bp)	14	7	33	54 (25%)
1–10 bp	13	7	32	52
11–1000 bp	0	0	1	1
>1000 bp	1	0	0	1
Insertion (1–72 bp)	14	12	18	44 (20.3%)
Templated	11	5	111	27
Random	3	7	7	17
Microhomology				
0–1 bp	33	28	64	125
2–4 bp	27	15	32	74
≥5 bp	8	1	8	17

CCR, complex chromosomal rearrangements.

71 kb, which is smaller than the resolution of CMA in diagnostic setting. They were all confirmed by real-time qPCR (online supplementary table S3).

Moreover, WGS showed an additional level of complexity of SVs compared with karyotype. WGS detected 218 breakpoints whereas 119 breakpoints were expected from karyotype (+83.2%). Fifteen patients presented a complex SV (at least three breakpoints) after WGS (figure 1). For 14 of them, the level of complexity was unexpected (28.5% patients). The various degrees of complexity included a small inverted fragment at the breakpoints, cryptic insertions and chromoanagenesis events (figure 2 and online supplementary figure S2-S3). Overall, the karyotype breakpoints were revised in 29/49 patients (59%).

Genomic features at the breakpoints

Different genomic features involved in SVs were assessed, including genes, TADs and repeated elements (table 1). Ninety-five of the 218 breakpoints disrupted a gene (43.5%). This proportion did not significantly differ from the proportion of genes in the reference genome hg19 (p=0.8374). Eighty-three different genes were disrupted, one gene being disrupted in two different patients (*MBD5*) and nine genes being disrupted by multiple breakpoints in the same patient (*BC015590, CAMK1D, CCDC3, MEF2C, KANSL1, TENM2, CELF2, TTC23, PKN2, GBP3*). A mean of 1.6 genes were disrupted per patient (0–7 genes per patient). TAD disruption was observed in 192/218 breakpoints (88%). This was not statistically different from the TAD proportion in the genome (p=0.2155). Seven TADs were disrupted in a recurrent manner, in at least two patients (chr2:148675000_149850000, chr5:88000000_90100000, chr10:79600000_82050000, chr14:46950000_49825000, chr14:94100000_94675000, chr14:97425000_99900000, chr15:35800000_38200000, hg19). Repeated elements, were involved in 110/218 breakpoints (50.4%), which was not significantly different from the proportion of repeated sequences in the genome (p=0.4227). Similarly, the distribution of breakpoints among the different families of repeated elements did not significantly differ from their distribution in the genome (p=0.607).

Thus, all these results suggest that DNA breaks at the origin of chromosomal rearrangements arose randomly and that there was no major influence of DNA architecture or repeated sequences.

Molecular signatures and mechanism of ABCR

We also studied the molecular signatures at junction sequences, including deletion, gain, insertion and microhomology, in order to infer the underlying mechanism at the origin of the rearrangements (table 2). Among the 49 rearrangements, 57.1% (28/49) were probably due to NHEJ and 10.2% (5/49) were likely to be the result of microhomology-mediated mechanisms, including replicative mechanisms, such as FoSTeS and MMBIR,[31] and microhomology-mediated end-joining (MMEJ).[32] It was not possible to attribute 20.4% of cases (10/49) to a specific mechanism. Interestingly, 12.2% of rearrangements (6/49) may have combined both NHEJ and microhomology-mediated mechanisms (online supplementary table S4).

Phenotypic consequences of chromosomal rearrangements

We address the question of the contribution of ABCR to the phenotype of patients. To assess the pathogenicity of ABCR, we used the classification proposed by Redin *et al*.[13] Sixteen ABCR were considered as pathogenic, 7 as likely pathogenic and 26 as variants of unknown significance (VUS) (table 3). Pathogenic and likely pathogenic ABCR accounted for the phenotype in 22 patients (44.8%). In patient DM/0109, disruption of *PRDM16* was considered as a secondary finding as it did not account for his phenotype but may be responsible for left ventricular non-compaction (MIM615373). Fifteen patients with pathogenic ABCR presented disruption of well-known disease genes (*KANSL1, FOXP1, SPRED1, MEIS2, MBD5, DMD, MEF2C, NRXN1, NFIX, AUTS2, SYNGAP1, GHR*) or more recently described gene (*TLK2, ZMIZ1*).[33 34] In 11 of these patients, expression studies could be performed and supported further the role of the putative causative gene in 7 patients (IS/0101, CL/2001, SL/0105, MH/1103, JP/0107, MD/2203, JB/1404) (see online supplementary figure S1). In patient CG/0106 showing an X-autosome translocation disrupting *DMD*, X-inactivation study showed a biased inactivation profile, supporting the role of *DMD* disruption in the phenotype, as previously described.[35] For patients showing likely pathogenic position effect, the breakpoint lied within the same TAD as the putative causal gene (*BCL11B, MEF2C, DLX5, SATB2, ZMIZ1*) separated the gene from its regulatory element (figure 2), resulting in regulatory

Table 3 Apparently balanced chromosomal rearrangements with a phenotypic impact in the patients

Patient	Karyotype	Phenotype summary	Mechanism	Gene	OMIM phenotype	Classification
IS/0101	t(9;17)(p13;q21)	ID, growth retardation, FD	Disruption	KANSL1	Koolen de Vries syndrome (610 443)	Pathogenic
IL/1901	der(3)inv(3)(p13;p22)inv(3)(p12;q26.3)	ID, speech delay, cleft palate	Disruption	FOXP1	MR with language impairment (613 670)	Pathogenic
VJ/0601	t(4;14)(p15;q32.1)	ID, speech delay, ASD, FD	PE	BCL11B	IDD and T-cell abnormalities (618 092)	Likely pathogenic
CS/1002	ins(15;12) (q15 or q21;q24.1q24.2)	ID, macrocephaly, FD, café-au-lait spots	Disruption	SPRED1	Legius syndrome (611 431)	Pathogenic
CL/2001	t(11;15)(p14;q14)	ID, absent speech, epilepsy, FD, CHD, larynx malposition,	Disruption	MEIS2	Cleft palate, cardiac defect, MR (600 987)	Pathogenic
SL/0105	inv(2)(p12q22)	ID, microcephaly, behavioural troubles, epilepsy, ataxia	Disruption	MBD5	MR (156 200)	Pathogenic
EB/0401	ins(5)(q15q23.3q34)	ID, growth retardation, microcephaly, epilepsy	PE	MEF2C	MR (613 443)	Likely pathogenic
CG/0106	t(X;13;15)(p21;q22;q22),t(6;11)(q21;q24)	ID, absent speech, attention deficit, muscular dystrophy	Disruption	DMD	Duchenne muscular dystrophy (310 200)	Pathogenic
MH/1103	inv(2)(q22;q32.3)	ID, speech delay, epilepsy	Disruption	MBD5	MR (156 200)	Pathogenic
JP/0107	t(4;17)(q27;q23)	Overweight, learning difficulties, attention disorder, FD	Disruption	TLK2	MR (618 050)	Pathogenic
CS/1902	t(4;14)(q31.1;q32.2)	Learning difficulties, speech delay, psychiatric disorders, FD	PE	BCL11B	IDD and T-cell abnormalities (618 092)	Likely pathogenic
OL/2202	t(1;14)(q32;q22)	ID, absent speech, stereotypy, epileptic encephalopathy	PE	MEF2C	MR (613 443)	Likely pathogenic
VD/2401	inv(7)(q22q34)	Mandibulofacial dysostosis, toes syndactyly, hearing loss	PE	DLX5	Split hand/foot malformation 1 (183 600)	Likely pathogenic
MD/2203	t(1;3;5)(p21;p24;q?34)	ID, absent speech, ASD, epilepsy, FD	Disruption	MEF2C	MR (613 443)	Pathogenic
DM/0109	t(1;19)(p36;q13.1)	ID, learning difficulties, dyslexia, behavioural troubles	Disruption	PRDM16	Left ventricular non compaction (615 373)	Pathogenic
BA/2303	t(2;8)(p16.3;p11.23)	ID, behavioural troubles, cryptorchidism	Disruption	NRXN1	Autism (209 850)	Pathogenic
JE/1401	t(X;10)(q?27;q?23)	ID, behavioural troubles, FD, obesity	PE	ZMIZ1	Syndromic neurodevelopmental disorder	Likely pathogenic
EM/2403	t(4;19)(q33;p13.2)	ID, overgrowth, macrocephaly, speech delay, FD	Disruption	NFIX	Sotos syndrome 2 (614 753)	Pathogenic
CF/2304	t(2;15;8;10)(q32.3;q26;q21.3;q23)	ID, speech delay	PE	SATB2	Glass syndrome (612 313)	Likely pathogenic
MC/1803	t(7;10)(q21.1;q26.1)	ID, microcephaly, absent speech, FD	Disruption	AUTS2	MR (615 834)	Pathogenic
JB/1404	t(10;12)(q22.2;q24.3)	ID, microcephaly, stereotypy	Disruption	ZMIZ1	Syndromic neurodevelopmental disorder	Pathogenic
JJ/0402	t(3;6)(q21.3;p21.33)	ID, hypotonia, FD, behavioural troubles	Disruption	SYNGAP1	MR (612 621)	Pathogenic
BM/1201	t(5;8)(q23;q24)t(5;11)(p12;p11)	ID, growth retardation, speech delay	Disruption	GHR	Partial GH deficiency (604 271)	Pathogenic

ASD, autistic spectrum disorder; CHD, congenital heart defect; FD, facial dysmorphism; GH, growth hormone; ID, intellectual disability; IDD, intellectual developmental disorder; MR, mental retardation; PE, position effect.

loss of function. It was further supported by expression studies showing decreased gene expression in five patients (VJ/0601, CS/1902, OL/2202, VD/2401, JE/1401) (online supplementary data S1). It is of note that for 6/23 patients the pathogenic breakpoint accounting for the phenotype was part of additional breakpoint complexity only visible after WGS (IL/1901, CS/1002, MD/2203, OL/2202, EB/0401, CF/2304).

We also identified four potential in-frame fusion transcripts in three patients: *M1AP-MBD5* (SL/0105), *CCDC3-CCSER1* (TS/1101), *MECOM-DDX24* and *DDX24-MECOM* (OP/0701) (online supplementary figure S4). Three of them could not be amplified in blood or LCL samples due to tissue-specific expression pattern (*M1AP-MBD5*, *CCDC3-CCSER1* and *MECOM-DDX24*). *M1AP* is mainly expressed in testis, *CCDC3* is expressed in arterial tissues and *MECOM* is not expressed in blood (GTEx). We were able to amplify the *DDX24-MECOM* fusion. This fusion transcript retained almost all the coding sequence of *MECOM* (exons 2–17).

Identification of candidate genes in developmental disorders

We then searched to identify new candidate genes for developmental disorders. First, we looked among the disrupted genes, those that showed a high probability of loss of function intolerance (>0.85) suggesting they may be responsible for a phenotype by haploinsufficiency. Fourteen genes not previously reported in developmental disorders were identified (table 4). Then we observed that in 8/21 patients with pathogenic and probably pathogenic ABCR, the causal breakpoint involved one of the seven recurrently disrupted TADs defined above. We looked at the genes of the other recurrently disrupted TADs and were able to identify two more candidate genes (*DDX24*, *MDGA2*)

Table 4 Candidate genes identified through gene disruption or TAD disruption

Gene	OMIM	Putative function	pLI	Patients	Phenotype summary	Gene classification
MAGI	602 625	Guanylate kinase, cellular interaction	1.00	IL/1901	ID, speech delay, cleft palate	VUS
FAM19A1	617 495	Cerebral chemokine	0.86	IL/1901	ID, speech delay, cleft palate	VUS
PPARGC1A	605 417	Energy metabolism regulation	1.00	VJ/0601	ID, speech delay, ASD, FD	VUS
RBM27	–	Unknown	1.00	EB/0401	ID, growth retardation, microcephaly, epilepsy	VUS
TENM2	610 119	Unknown	1.00	EB/0401	ID, growth retardation, microcephaly, epilepsy	VUS
CNTN4	607 280	Contactin, CNS development	1.00	TS/1101	ID, behavioural troubles, FD	VUS
CELF2	602 538	RNA splicing	1.00	TS/1101	ID, behavioural troubles, FD	VUS
CAMK1D	607 957	Serine-threonine kinase	0.99	TS/1101	ID, behavioural troubles, FD	VUS
DCAF5	603 812	Mitochondrial protein	1.00	AG/1003	FD, clubfeet, Crohn's disease	VUS
UNC79	616 884	Sodium channel subunit	1.00	EC/2302	ID, psychiatric troubles, FD	VUS
RBM26	-	Unknown	1.00	BM/1603	ID, FD CHD,	VUS
SEMA6A	605 885	CNS development	1.00	BM/1201	ID, growth retardation, speech delay	VUS
PKN2	602 549	Unknown	1.00	MD/2203	ID, absent speech, ASD, epilepsy, FD	VUS
STXBP5L	609 381	Exocytosis inhibition	1.00	CM/0103	ID, epilepsy	VUS
DDX24	606 181	RNA helicase	0.22	OP/0701; EC/2302	ID, behavioural troubles	VUS
MDGA2	611 128	Synaptic regulation	0.99	OL/2202; KT/1403	ID, absent/delayed speech	VUS

ASD, autistic spectrum disorder; CHD, congenital heart defect; CNS, central nervous system; FD, facial dysmorphism; ID, intellectual disability; pLI, probability of loss of function intolerance; TAD, topologically associated domain; VUS, variants of unknown significance.

(table 4). All these genes are predicted to be involved in different cellular processes and all showed cerebral expression (GTEx).

Role of genome architecture in gene regulation

We then explored the general impact of breakpoints on the regulation of genes. In particular, we looked for an effect of the distance from the breakpoints[36] and of genomic architecture (TADs) on gene expression level. For this purpose, we systematically explored the expression in blood or LCL of disrupted genes and genes at either side of the breakpoints. One hundred eighty-nine genes in 47 patients were studied. Regarding disrupted genes, relative mRNA ratios ranged from 0.239 to 1.6 in blood and from 0.187 to 2.007 in LCL. For genes in the vicinity of the breakpoints, relative mRNA ratios ranged from 0.258 to 6.821 in blood and from 0.644 to 1.996 in LCLs. No significant correlation was found between the mRNA ratio and the absolute value of the distance from the breakpoint in blood (repeated measure correlation coefficient $r=-0.187$, $p=0.097$) and in LCL (repeated measures correlation coefficient $r=-0.0637$, $p=0.77$). There was also no significant difference between the mean mRNA ratio for genes located in the same TAD as the breakpoint, genes located in an adjacent TAD or genes in inter-TAD regions (repeated measures analysis of variance, $p=0.7581$; online supplementary figure S5).

DISCUSSION

In the present work, we developed a paired-end WGS approach to characterise ABCRs and demonstrated its efficiency in a clinical setting.

WGS has already proved to be powerful for breakpoint mapping of chromosomal rearrangements.[10 11 13 14 16 19] Most of the studies used a mate-pair library or similar approaches, which are more cost-effective and more robust.[11 13 14 37] These approaches are based on long insert sequencing and require less sequencing to achieve a better sequence coverage. They are also expected to overcome, at least in part, difficulties for mapping in repeated sequences and gaps. In the present study, the chromosomal rearrangement detection rate was similar to previous studies using mate-pair sequencing.[13] The paired-end strategy was able to detect breakpoints in highly repetitive regions such as centromeric regions, segmental duplications and short arms of acrocentric chromosomes. Working with the latest version of reference genome, hg38, was helpful to minimise gaps.[17] We were also able to characterise the breakpoints of a mosaic SV. Mosaicism detection from next-generation sequencing data has scarcely been considered in balanced SV and need to be further addressed.

The present study provided a diagnosis in nearly half of patients, stopping a long diagnostic odyssey. For these patients and their family, it improved medical management and allowed an informed genetic counselling. For a subset of patients, WGS was the only way to achieve this result, as the pathogenic breakpoint could not be inferred from karyotype alone. Although we opted for a low depth WGS in this study, which is sufficient for SV detection, the analysis strategy we developed would be easily applicable to classical 30X depth WGS. An advantage of the 30X paired-end WGS over mate-pair is that it provides the possibility of an overall approach combining SV, including balanced SVs and CNVs, and SNV detection. A recent study found that WGS strategy combining SNV and CNV analysis could reach a diagnostic yield of up to 62% in patients with severe ID.[18] It was also pointed out that phenotype could result from variation on multiple loci in 4.9% of patients.[38] Regarding patients with ABCR and abnormal phenotype, an extrapolation combining analysis of CNVs (40%), SV (40% of patients with no CNVs) and SNV (based on a yield of 39%)[18] could reach a diagnostic yield of 78%. Concerning most patients with ID/MCA, karyotype is no longer performed systematically and CMA, used as first-tier diagnostic method, is not able to detect balanced rearrangements. It has been demonstrated that WGS could provide unbiased detection of previously unknown ABCR.[39] In the future, as WGS could be considered as the initial test in ID/MCA, pipelines should include balanced SV detection as it would be the unique way to determine gene disruption. It is of note however that we were faced with a secondary finding in a patient for which the disrupted gene caused a disorder unrelated to his phenotype.

This observation underlines the importance of complete pretest information including the possibility of secondary findings even if the analysis is focused on breakpoint mapping.[40]

In the 22 patients with a diagnosis, the pathophysiology mechanism was either gene disruption or position effect, both supposed to result in haploinsufficiency. In one patient, we identified a fusion transcript *DDX24-MECOM* that allowed preservation of almost all *MECOM* coding sequence. Chimeric genes have been rarely described in constitutional rearrangements.[41 42] Their pathogenicity is not well established and could result from a new function, expression deregulation or dominant negative effect. In the present case, the phenotype of patient OP/0701 was not relevant with the phenotype of radioulnar synostosis with amegakaryocytic thrombocytopaenia (MIM616738) described in *MECOM* mutations. It cannot be excluded that deregulation of *MECOM* expression under the control of *DDX24* promoter may contribute to the phenotype.

The organisation of the genome is not linear, and domains of preferential chromatin interactions, called TADs, have been recently mapped across the genome.[25] Their disruption by SV has been shown to be involved in altered phenotypes.[43] In the present study, TADs map help support the pathogenicity of breakpoints in seven patients. Two TADs, one involving *MEF2C* and the other *SATB2*, have already been described as recurrently disrupted in patients with ID.[13] In the future, TAD maps may be integrated into tools for breakpoint annotation. This will be all the more possible as TAD mapping algorithms will improve[44] and in silico modelling tools will become available.[45] Nevertheless, it has to be noted that no systematic correlation between TAD disruption and altered gene expression could be demonstrated in the present study. Although expression studies were limited by the tissues available, blood and LCL, this result may highlight fine intra-TAD regulations, dependent on tissue and developmental stages.[46]

Regarding ABCR mechanisms, the distribution of the breakpoints suggests that non-recurrent rearrangements occur randomly; in particular, we did not show any enrichment for repeated elements. This may be explained by the fact that the present cohort did not include any patient with large CNV, more prone to be associated with repeated elements.[13]Second, some rearrangements hold molecular signatures suggesting that they may result from the combination of multiple repair mechanisms, in particular, combination of NHEJ and MMEJ. MMEJ, also called alternative NHEJ, is a double-strand break repair mechanism. It is associated with microhomology and larger deletion than NHEJ and may be activated when canonical NHEJ pathway is deficient.[32] This observation needs to be further validated in a larger cohort.

In conclusion, the present study demonstrated the relevance of paired-end WGS for ABCR breakpoint characterisation and its contribution to diagnosis with a yield of 44.8%. This analysis should be systematically proposed to patients harbouring ABCR with abnormal phenotype and SV analysis should be included in WGS pipeline in clinical setting.

Author affiliations
[1]Service de Génétique, Hospices Civils de Lyon, Bron, France
[2]INSERM U1028, CNRS UMR5292, UCBL1, GENDEV Team, Neurosciences Research Center of Lyon, Bron, France
[3]Cellule bioinformatique de la plateforme NGS, Hospices Civils de Lyon, CNRS, Laboratoire de Biométrie et Biologie Evolutive UMR5558, Lyon 1 University, Bron, France
[4]Département de génétique et embryologie médicale, Centre de référence des déficiences intellectuelles de causes rares, AP-HP, Hôpital Armand Trousseau, Paris, France
[5]GRC n°19, pathologies Congénitales du Cervelet-LeucoDystrophies, AP-HP, Hôpital Armand Trousseau, Sorbonne Université, Paris, France
[6]Laboratoire de Génétique Chromosomique, Hôpital Couple Enfant, CHU Grenoble, Grenoble, France
[7]Service de Génétique Médicale, Hôpital Necker-Enfants Malades, Paris, France
[8]Service de Génétique, CHRU de Tours, Tours, France
[9]Laboratoire de Cytogénétique, CHU Dijon, Dijon, France
[10]Département de Génétique, Hôpital Robert Debré, Paris, France
[11]Service de Génétique Médicale, CHI Toulon, Toulon, France
[12]Service de Génétique, Hôpital Arnaud de Villeneuve, Montpellier, France
[13]Centre d'activité de génétique clinique, CLAD nord de France, CHU Amiens, Amiens, France
[14]Service de Génétique Médicale, CHU Nice, Nice, France
[15]Service de Génétique Clinique, CHU Rennes, Rennes, France
[16]Service de Génétique, EA3801, SFR CAP SANTE, CHU Reims, Reims, France
[17]Laboratoire de Cytogénétique Constitutionnelle, APHP-HUPC site Cochin, Paris, France
[18]Centre de référence anomalies du développement et syndromes malformatifs, FHU TRANSLAD et équipe GAD INSERM UMR1231, CHU Dijon-Bourgogne et Université de Bourgogne-Franche Comté, Dijon, France
[19]Service de Génétique, EA3808, Université de Poitiers, CHU de Poitiers, Poitiers, France
[20]Unité de Génétique Clinique, CHU de Rouen, Rouen, France
[21]Service de Génétique Médicale, CHU-Nantes, Nantes, France
[22]Laboratoire de Cytogénétique et de Biologie Cellulaire, CHU Pontchaillou, Rennes, France
[23]Laboratoire de Cytogénétique, CHU de Rouen, Rouen, France
[24]Laboratoire de Cytogénétique, CHU Strasbourg, Strasbourg, France
[25]Service de Génétique Médicale, Hôpital Pellegrin, Université de Bordeaux, MRGM INSERM U1211, CHU Bordeaux, Bordeaux, France
[26]Service de Génétique Clinique, Chromosomique et Moléculaire, CHU Hôpital Nord, Saint-Etienne, France
[27]Laboratoire de Génétique Chromosomique, CH Général, Chambéry, France
[28]Service de Cytogénétique, Hôpital Necker Enfants Malades, Paris, France
[29]Département de Génétique; Centre de Référence Déficience Intellectuelle de Causes Rares, Groupe Hospitalier Pitié-Salpêtrière, APHP, Paris, France
[30]Laboratoire de Génétique Chromosomique, Département de Génétique Médicale, AP-HM, Marseille, France
[31]Service de Cytogénétique, CH Valence, Valence, France
[32]Service de Cytogénétique Médicale, Hôpital Estaing, CHU Clermont-Ferrand, Clermont-Ferrand, France
[33]CNRS, IGDR (Institut de Génétique et Développement de Rennes) UMR 6290, Université de Rennes, Rennes, France
[34]Département de Génétique Médicale, Unité de Génétique Clinique, AP-HM, Marseille, France
[35]Service de Génétique Médicale, Hôpital Estaing, CHU Clermont-Ferrand, Clermont-Ferrand, France
[36]Equipe Génétique, Epigénétique et Thérapies de l'Infertilité, IAB, INSERM 1209, CNRS UMR5309, Grenoble, France
[37]Service de Neurologie Pédiatrique, Hôpital Armand Trousseau, APHP, GHUEP, Paris, France

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

REFERENCES

1 Gardner R, Sutherland G, Shaffer L. Gametogenesis and conception, pregnancy loss and infertility. In: Hall J, Harper PS, Hudgkins L, Eichler E, Epstein CJ, *et al*. eds. *Chromosome abnormalities and genetic counseling*. 4th edn. New York: Oxford University Press, 2018:377–400.

2 Halgren C, Nielsen NM, Nazaryan-Petersen L, Silahtaroglu A, Collins RL, Lowther C, Kjaergaard S, Frisch M, Kirchhoff M, Brøndum-Nielsen K, Lind-Thomsen A, Mang Y, El-Schich Z, Boring CA, Mehrjouy MM, Jensen PKA, Fagerberg C, Krogh LN, Hansen J, Bryndorf T, Hansen C, Talkowski ME, Bak M, Tommerup N, Bache I. Risks and recommendations in prenatally detected de novo balanced chromosomal rearrangements from assessment of long-term outcomes. *Am J Hum Genet* 2018;102:1090–103.

3 Warburton D. De novo balanced chromosome rearrangements and extra marker chromosomes identified at prenatal diagnosis: clinical significance and distribution of breakpoints. *Am J Hum Genet* 1991;49:995–1013.

4 Fantes JA, Boland E, Ramsay J, Donnai D, Splitt M, Goodship JA, Stewart H, Whiteford M, Gautier P, Harewood L, Holloway V, Sharkey F, Maher E, van Heyningen V, Clayton-Smith J, Fitzpatrick DR, Black GC. FISH mapping of de novo apparently balanced chromosome rearrangements identifies characteristics associated with phenotypic abnormality. *Am J Hum Genet* 2008;82:916–26.

5 Gribble SM, Prigmore E, Burford DC, Porter KM, Ng BL, Douglas EJ, Fiegler H, Carr P, Kalaitzopoulos D, Clegg S, Sandstrom R, Temple IK, Youings SA, Thomas NS, Dennis NR, Jacobs PA, Crolla JA, Carter NP. The complex nature of constitutional de novo apparently balanced translocations in patients presenting with abnormal phenotypes. *J Med Genet* 2005;42:8–16.

6 Schluth-Bolard C, Delobel B, Sanlaville D, Boute O, Cuisset JM, Sukno S, Labalme A, Duban-Bedu B, Plessis G, Jaillard S, Dubourg C, Henry C, Lucas J, Odent S, Pasquier L, Copin H, Latour P, Cordier MP, Nadeau G, Till M, Edery P, Andrieux J. Cryptic genomic imbalances in de novo and inherited apparently balanced chromosomal rearrangements: array CGH study of 47 unrelated cases. *Eur J Med Genet* 2009;52:291–6.

7 Kalscheuer VM, Tao J, Donnelly A, Hollway G, Schwinger E, Kübart S, Menzel C, Hoeltzenbein M, Tommerup N, Eyre H, Harbord M, Haan E, Sutherland GR, Ropers HH, Gécz J. Disruption of the serine/threonine kinase 9 gene causes severe X-linked infantile spasms and mental retardation. *Am J Hum Genet* 2003;72:1401–11.

8 Kleinjan DA, Lettice LA. Long-range gene control and genetic disease. *Adv Genet* 2008;61:339–88.

9 Korbel JO, Urban AE, Affourtit JP, Godwin B, Grubert F, Simons JF, Kim PM, Palejev D, Carriero NJ, Du L, Taillon BE, Chen Z, Tanzer A, Saunders AC, Chi J, Yang F, Carter NP, Hurles ME, Weissman SM, Harkins TT, Gerstein MB, Egholm M, Snyder M. Paired-end mapping reveals extensive structural variation in the human genome. *Science* 2007;318:420–6.

10 Chen W, Ullmann R, Langnick C, Menzel C, Wotschofsky Z, Hu H, Döring A, Hu Y, Kang H, Tzschach A, Hoeltzenbein M, Neitzel H, Markus S, Wiedersberg E, Kistner G, van Ravenswaaij-Arts CM, Kleefstra T, Kalscheuer VM, Ropers HH. Breakpoint analysis of balanced chromosome rearrangements by next-generation paired-end sequencing. *Eur J Hum Genet* 2010;18:539–43.

11 Chiang C, Jacobsen JC, Ernst C, Hanscom C, Heilbut A, Blumenthal I, Mills RE, Kirby A, Lindgren AM, Rudiger SR, McLaughlan CJ, Bawden CS, Reid SJ, Faull RL, Snell RG, Hall IM, Shen Y, Ohsumi TK, Borowsky ML, Daly MJ, Lee C, Morton CC, MacDonald ME, Gusella JF, Talkowski ME. Complex reorganization and predominant non-homologous repair following chromosomal breakage in karyotypically balanced germline rearrangements and transgenic integration. *Nat Genet* 2012;44:390–7.

12 Nilsson D, Pettersson M, Gustavsson P, Förster A, Hofmeister W, Wincent J, Zachariadis V, Anderlid BM, Nordgren A, Mäkitie O, Wirta V, Käller M, Vezzi F, Lupski JR, Nordenskjöld M, Lundberg ES, Carvalho CMB, Lindstrand A. Whole-genome sequencing of cytogenetically balanced chromosome translocations identifies potentially pathological gene disruptions and highlights the importance of microhomology in the mechanism of formation. *Hum Mutat* 2017;38:180–92.

13 Redin C, Brand H, Collins RL, Kammin T, Mitchell E, Hodge JC, Hanscom C, Pillalamarri V, Seabra CM, Abbott MA, Abdul-Rahman OA, Aberg E, Adley R, Alcaraz-Estrada SL, Alkuraya FS, An Y, Anderson MA, Antolik C, Anyane-Yeboa K, Atkin JF, Bartell T, Bernstein JA, Beyer E, Blumenthal I, Bongers EM, Brilstra EH, Brown CW, Brüggenwirth HT, Callewaert B, Chiang C, Corning A, D'Hooge M, de Vries BB, Earl DL, Ferguson HL, Fisher H, FitzPatrick DR, Gerrol S, Giachino D, Glessner JT, Gliem T, Grady M, Graham BH, Griffis C, Gripp KW, Gropman AL, Hanson-Kahn A, Harris DJ, Hayden MA, Hill R, Hochstenbach R, Hoffman JD, Hopkin RJ, Hubshman MW, Innes AM, Irons M, Irving M, Jacobsen JC, Janssens S, Jewett T, Johnson JP, Jongmans MC, Kahler SG, Koolen DA, Korzelius J, Kroisel PM, Lacassie Y, Lawless W, Lemyre E, Leppig K, Levin AV, Li H, Li H, Liao EC, Lim C, Lose EJ, Lucente D, Macera MJ, Manavalan P, Mandrile G, Marcelis CL, Margolin L, Mason T, Masser-Frye D, McClellan MW, Mendoza CJ, Menten B, Middelkamp S, Mikami LR, Moe E, Mohammed S, Mononen T, Mortenson

ME, Moya G, Nieuwint AW, Ordulu Z, Parkash S, Pauker SP, Pereira S, Perrin D, Phelan K, Aguilar RE, Poddighe PJ, Pregno G, Raskin S, Reis L, Rhead W, Rita D, Renkens I, Roelens F, Ruliera J, Rump P, Schilit SL, Shaheen R, Sparkes R, Spiegel E, Stevens B, Stone MR, Tagoe J, Thakuria JV, van Bon BW, van de Kamp J, van Der Burgt I, van Essen T, van Ravenswaaij-Arts CM, van Roosmalen MJ, Vergult S, Volker-Touw CM, Warburton DP, Waterman MJ, Wiley S, Wilson A, Yerena-de Vega MC, Zori RT, Levy B, Brunner HG, de Leeuw N, Kloosterman WP, Thorland EC, Morton CC, Gusella JF, Talkowski ME. The genomic landscape of balanced cytogenetic abnormalities associated with human congenital anomalies. *Nat Genet* 2017;49:36–45.

14 Talkowski ME, Rosenfeld JA, Blumenthal I, Pillalamarri V, Chiang C, Heilbut A, Ernst C, Hanscom C, Rossin E, Lindgren AM, Pereira S, Ruderfer D, Kirby A, Ripke S, Harris DJ, Lee JH, Ha K, Kim HG, Solomon BD, Gropman AL, Lucente D, Sims K, Ohsumi TK, Borowsky ML, Loranger S, Quade B, Lage K, Miles J, Wu BL, Shen Y, Neale B, Shaffer LG, Daly MJ, Morton CC, Gusella JF. Sequencing chromosomal abnormalities reveals neurodevelopmental loci that confer risk across diagnostic boundaries. *Cell* 2012;149:525–37.

15 Utami KH, Hillmer AM, Aksoy I, Chew EG, Teo AS, Zhang Z, Lee CW, Chen PJ, Seng CC, Ariyaratne PN, Rouam SL, Soo LS, Yousoof S, Prokudin I, Peters G, Collins F, Wilson M, Kakakios A, Haddad G, Menuet A, Perche O, Tay SK, Sung KW, Ruan X, Ruan Y, Liu ET, Briault S, Jamieson RV, Davila S, Cacheux V. Detection of chromosomal breakpoints in patients with developmental delay and speech disorders. *PLoS One* 2014;9:e90852.

16 Vergult S, Van Binsbergen E, Sante T, Nowak S, Vanakker O, Claes K, Poppe B, Van der Aa N, van Roosmalen MJ, Duran K, Tavakoli-Yaraki M, Swinkels M, van den Boogaard MJ, van Haelst M, Roelens F, Speleman F, Cuppen E, Mortier G, Kloosterman WP, Menten B. Mate pair sequencing for the detection of chromosomal aberrations in patients with intellectual disability and congenital malformations. *Eur J Hum Genet* 2014;22:652–9.

17 Guo Y, Dai Y, Yu H, Zhao S, Samuels DC, Shyr Y. Improvements and impacts of GRCh38 human reference on high throughput sequencing data analysis. *Genomics* 2017;109:83–90.

18 Gilissen C, Hehir-Kwa JY, Thung DT, van de Vorst M, van Bon BW, Willemsen MH, Kwint M, Janssen IM, Hoischen A, Schenck A, Leach R, Klein R, Tearle R, Bo T, Pfundt R, Yntema HG, de Vries BB, Kleefstra T, Brunner HG, Vissers LE, Veltman JA. Genome sequencing identifies major causes of severe intellectual disability. *Nature* 2014;511:344–7.

19 Schluth-Bolard C, Labalme A, Cordier MP, Till M, Nadeau G, Tevissen H, Lesca G, Boutry-Kryza N, Rossignol S, Rocas D, Dubruc E, Edery P, Sanlaville D. Breakpoint mapping by next generation sequencing reveals causative gene disruption in patients carrying apparently balanced chromosome rearrangements with intellectual deficiency and/or congenital malformations. *J Med Genet* 2013;50:144–50.

20 Li H. *Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM*. arXiv: Cornell University Library, 2013:arXiv: 1303.3997v2.

21 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009;25:2078–9.

22 Chen K, Wallis JW, McLellan MD, Larson DE, Kalicki JM, Pohl CS, McGrath SD, Wendl MC, Zhang Q, Locke DP, Shi X, Fulton RS, Ley TJ, Wilson RK, Ding L, Mardis ER. BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nat Methods* 2009;6:677–81.

23 Thorvaldsdottir H, Robinson JT, Mesirov JP, Viewer IG. IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* 2013;14:178–92.

24 Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 2001;25:402–8.

25 Rao SS, 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.

26 Ottaviani D, LeCain M, Sheer D. The role of microhomology in genomic structural variation. *Trends Genet* 2014;30:85–94.

27 Carvalho CM, Lupski JR. Mechanisms underlying structural variant formation in genomic disorders. *Nat Rev Genet* 2016;17:224–38.

28 Chang HHY, Pannunzio NR, Adachi N, Lieber MR. Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nat Rev Mol Cell Biol* 2017;18:495–506.

29 R-Core-Team. *Pinheiro J. R: A language and environment for statistical computing., R Foundation for Statistical Computing ed*. Vienna, Austria: R Foundation for Statistical Computing, 2017.

30 Bakdash JZ, Marusich LR, Correlation RM*Front Psychol* 2017;8:456.

31 Zhang F, Khajavi M, Connolly AM, Towne CF, Batish SD, Lupski JR. The DNA replication FoSTeS/MMBIR mechanism can generate genomic, genic and exonic complex rearrangements in humans. *Nat Genet* 2009;41:849–53.

32 Wang H, Xu X. Microhomology-mediated end joining: new players join the team. *Cell Biosci* 2017;7:6.

33 Carapito R, Ivanova EL, Morlon A, Meng L, Molitor A, Erdmann E, Kieffer B, Pichot A, Naegely L, Kolmer A, Paul N, Hanauer A, Tran Mau-Them F, Jean-Marçais N, Hiatt SM, Cooper GM, Tvrdik T, Muir AM, Dimartino C, Chopra M, Amiel J, Gordon CT, Dutreux F, Garde A, Thauvin-Robinet C, Wang X, Leduc MS, Phillips M, Crawford HP, Kukolich MK, Hunt D, Harrison V, Kharbanda S, Smigiel R, Gold N, Hung CY, Viskochil DH, Dugan SL, Bayrak-Toydemir P, Joly-Helas G, Guerrot AM, Schluth-Bolard

C, Rio M, Wentzensen IM, McWalter K, Schnur RE, Lewis AM, Lalani SR, Mensah-Bonsu N, Céraline J, Sun Z, Ploski R, Bacino CA, Mefford HC, Faivre L, Bodamer O, Chelly J, Isidor B, Bahram S. Deciphering Developmental Disorders StudyUniversity of Washington Center for Mendelian Genomics. ZMIZ1 Variants Cause a Syndromic Neurodevelopmental Disorder. *Am J Hum Genet* 2019;104:319–30.

34 Reijnders MRF, Miller KA, Alvi M, Goos JAC, Lees MM, de Burca A, Henderson A, Kraus A, Mikat B, de Vries BBA, Isidor B, Kerr B, Marcelis C, Schluth-Bolard C, Deshpande C, Ruivenkamp CAL, Wieczorek D, Baralle D, Blair EM, Engels H, Lüdecke HJ, Eason J, Santen GWE, Clayton-Smith J, Chandler K, Tatton-Brown K, Payne K, Helbig K, Radtke K, Nugent KM, Cremer K, Strom TM, Bird LM, Sinnema M, Bitner-Glindzicz M, van Dooren MF, Alders M, Koopmans M, Brick L, Kozenko M, Harline ML, Klaassens M, Steinraths M, Cooper NS, Edery P, Yap P, Terhal PA, van der Spek PJ, Lakeman P, Taylor RL, Littlejohn RO, Pfundt R, Mercimek-Andrews S, Stegmann APA, Kant SG, McLean S, Joss S, Swagemakers SMA, Douzgou S, Wall SA, Küry S, Calpena E, Koelling N, McGowan SJ, Twigg SRF, Mathijssen IMJ, Nellaker C, Brunner HG, Wilkie AOM. Deciphering Developmental Disorders Study. De novo and inherited loss-of-function variants in TLK2: Clinical and genotype-phenotype evaluation of a distinct neurodevelopmental disorder. *Am J Hum Genet* 2018;102:1195–203.

35 Kenwrick S, Patterson M, Speer A, Fischbeck K, Davies K. Molecular analysis of the Duchenne muscular dystrophy region using pulsed field gel electrophoresis. *Cell* 1987;48:351–7.

36 Merla G, Howald C, Henrichsen CN, Lyle R, Wyss C, Zabot MT, Antonarakis SE, Reymond A. Submicroscopic deletion in patients with Williams-Beuren syndrome influences expression levels of the nonhemizygous flanking genes. *Am J Hum Genet* 2006;79:332–41.

37 Liang WS, Aldrich J, Tembe W, Kurdoglu A, Cherni I, Phillips L, Reiman R, Baker A, Weiss GJ, Carpten JD, Craig DW. Long insert whole genome sequencing for copy number variant and translocation detection. *Nucleic Acids Res* 2014;42:e8.

38 Posey JE, Harel T, Liu P, Rosenfeld JA, James RA, Coban Akdemir ZH, Walkiewicz M, Bi W, Xiao R, Ding Y, Xia F, Beaudet AL, Muzny DM, Gibbs RA, Boerwinkle E, Eng CM, Sutton VR, Shaw CA, Plon SE, Yang Y, Lupski JR. Resolution of disease phenotypes resulting from multilocus genomic variation. *N Engl J Med* 2017;376:21–31.

39 Dong Z, Wang H, Chen H, Jiang H, Yuan J, Yang Z, Wang WJ, Xu F, Guo X, Cao Y, Zhu Z, Geng C, Cheung WC, Kwok YK, Yang H, Leung TY, Morton CC, Cheung SW, Choy KW. Identification of balanced chromosomal rearrangements previously unknown among participants in the 1000 Genomes Project: implications for interpretation of structural variation in genomes and the future of clinical cytogenetics. *Genet Med* 2018;20:697–707.

40 Rehm HL, Bale SJ, Bayrak-Toydemir P, Berg JS, Brown KK, Deignan JL, Friez MJ, Funke BH, Hegde MR, Lyon E. Working Group of the American College of Medical Genetics and Genomics Laboratory Quality Assurance Commitee. ACMG clinical laboratory standards for next-generation sequencing. *Genet Med* 2013;15:733–47.

41 Backx L, Seuntjens E, Devriendt K, Vermeesch J, Van Esch H. A balanced translocation t(6;14)(q25.3;q13.2) leading to reciprocal fusion transcripts in a patient with intellectual disability and agenesis of corpus callosum. *Cytogenet Genome Res* 2011;132:135–43.

42 Córdova-Fletes C, Domínguez MG, Delint-Ramirez I, Martínez-Rodríguez HG, Rivas-Estilla AM, Barros-Núñez P, Ortiz-López R, Neira VA. A de novo t(10;19)(q22.3;q13.33) leads to ZMIZ1/PRR12 reciprocal fusion transcripts in a girl with intellectual disability and neuropsychiatric alterations. *Neurogenetics* 2015;16:287–98.

43 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.

44 Forcato M, Nicoletti C, Pal K, Livi CM, Ferrari F, Bicciato S. Comparison of computational methods for Hi-C data analysis. *Nat Methods* 2017;14:679–85.

45 Bianco S, Lupiáñez DG, Chiariello AM, Annunziatella C, Kraft K, Schöpflin R, Wittler L, Andrey G, Vingron M, Pombo A, Mundlos S, Nicodemi M. Polymer physics predicts the effects of structural variants on chromatin architecture. *Nat Genet* 2018;50:662–7.

46 Phillips-Cremins JE, Sauria ME, Sanyal A, Gerasimova TI, Lajoie BR, Bell JS, Ong CT, Hookway TA, Guo C, Sun Y, Bland MJ, Wagstaff W, Dalton S, McDevitt TC, Sen R, Dekker J, Taylor J, Corces VG. Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* 2013;153:1281–95.