

Chromothripsis as a mechanism driving complex *de novo* structural rearrangements in the germline[†]

Wigard P. Kloosterman¹, Victor Guryev², Mark van Roosmalen¹, Karen J. Duran¹, Ewart de Bruijn², Saskia C.M. Bakker³, Tom Letteboer¹, Bernadette van Nesselrooij¹, Ron Hochstenbach¹, Martin Poot¹ and Edwin Cuppen^{1,2,*}

¹Department of Medical Genetics, University Medical Center Utrecht, Universiteitsweg 100, 3584 CG Utrecht, The Netherlands, ²Hubrecht Institute and University Medical Center Utrecht, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands and ³Department of Pediatrics, Gelre Hospital, PO Box 9014, 7300 DS Apeldoorn, The Netherlands

Received February 8, 2011; Revised February 8, 2011; Accepted February 18, 2011

A variety of mutational mechanisms shape the dynamic architecture of human genomes and occasionally result in congenital defects and disease. Here, we used genome-wide long mate-pair sequencing to systematically screen for inherited and *de novo* structural variation in a trio including a child with severe congenital abnormalities. We identified 4321 inherited structural variants and 17 *de novo* rearrangements. We characterized the *de novo* structural changes to the base-pair level revealing a complex series of balanced inter- and intra-chromosomal rearrangements consisting of 12 breakpoints involving chromosomes 1, 4 and 10. Detailed inspection of breakpoint regions indicated that a series of simultaneous double-stranded DNA breaks caused local shattering of chromosomes. Fusion of the resulting chromosomal fragments involved non-homologous end joining, since junction points displayed limited or no homology and small insertions and deletions. The pattern of random joining of chromosomal fragments that we observe here strongly resembles the somatic rearrangement patterns—termed chromothripsis—that have recently been described in deranged cancer cells. We conclude that a similar mechanism may also drive the formation of *de novo* structural variation in the germline.

INTRODUCTION

Complex chromosomal rearrangements having at least three breakpoints involving more than a single chromosome are associated with a variety of clinical phenotypes, including mental retardation and complex congenital abnormalities (1,2). Detailed delineation of rearrangements requires the application of advanced molecular cytogenetic techniques, such as fluorescence *in situ* hybridization, karyotyping and array-based copy number analysis (3–7). Recently, massively parallel mate-pair and paired-end sequencing was shown to be a powerful methodology to characterize breakpoints for previously identified (*de novo*) balanced rearrangements at nucleotide resolution (8–11). Paired-end sequencing generally involves the bidirectional sequencing of genomic libraries with short inserts (up to ~500 bp), whereas mate-pair libraries

are based on larger insert sizes, which allows for much higher genomic coverage per sequenced clone and gives the ability to span repetitive or low-complexity elements that are commonly driving genomic instability events (12). A major advantage of massively parallel paired-end and mate-pair sequencing over array-based technology is the possibility to detect breakpoints of balanced and unbalanced structural variation at high resolution (11), making these technologies also well suited for obtaining mechanistic insight into the origin of structural changes.

Several mechanisms have been proposed that could account for the complexity of *de novo* genomic rearrangements (13). Non-homologous end joining (NHEJ) can explain the formation of reciprocal translocations following the occurrence of double-stranded chromosomal breaks (14,15), although recent studies provide evidence that translocations occur

*To whom correspondence should be addressed. Tel: +31 887568901 or +31 302121969; Fax: +31 887568479; Email: e.cuppen@hubrecht.eu

[†]Mate-pair sequencing data are available at our website (<http://fedor21.hubrecht.eu/trio>) and the EBI ENA (European Nucleotide Archive) under accession numbers (submission pending). All detected structural variation is available from the supplementary material online.

more frequently in mammalian cells when NHEJ components like Ku70 and the ligase Xrcc4 are missing, implying that alternative end-joining (alt-EJ) mechanisms could mediate translocation formation (16,17). Modifications of free DNA ends at sites of double-stranded breaks are frequently observed and deletions are most commonly found (16,18). Other hallmarks of DNA junctions formed by end-joining processes are microhomology and inserted sequences (16).

Recently, a novel replication-based model called fork stalling and template switching (FoSTeS) was proposed to explain the complexity and microhomologies at breakpoints of non-recurrent duplications and deletions associated with genomic disorders (19). The FoSTeS model is related to another model called microhomology-mediated break-induced replication (MMBIR) (19,20). MMBIR is based on the mechanism of break-induced replication (21) and is proposed to account for a variety of complex rearrangements, mainly involving duplications and deletions (22).

To reconstruct the full complement of molecular events and mechanisms that lead to the formation of complex rearrangements, it is critical to detect and characterize all *de novo* breakpoints involved. We set out to characterize balanced and unbalanced human structural variation, including inherited and *de novo* events, at high resolution by performing next-generation mate-pair sequencing of a family trio consisting of a child with congenital defects and both his parents. On the basis of our mate-pair data, we predicted 25 *de novo* rearrangements and confirmed 17 of these by polymerase chain reaction (PCR) and capillary sequencing. Twelve of the rearrangements clustered in small regions on chromosomes 1, 4 and 10. Nucleotide resolution analysis of junction regions shows that a series of simultaneous double-stranded DNA breaks has triggered the formation of these *de novo* rearrangements, similar to a mechanism, termed chromothripsis, which has recently been proposed for deranged cancer cells (23). Our data provide evidence that local shattering of chromosomes followed by NHEJ may drive formation of complex constitutional rearrangements involved in congenital defects.

RESULTS

Structural variation detection by mate-pair sequencing of a family trio

We analyzed a male patient with a complex congenital phenotype, involving severe psychomotor retardation, speech delay, hypertelorism and kyphoscoliosis, but no further symptoms that may relate the condition of this patient to a known syndrome.

To identify a putative genetic basis of the phenotypic defects in the patient, we performed standard karyotyping with a resolution of 550 chromosomal bands of the patient and his parents. This analysis revealed a constitutional *de novo* complex chromosomal rearrangement involving chromosomes 1, 4 and 10 (Fig. 1). No imbalances within the translocation breakpoint regions were detected by genome-wide segmental aneuploidy profiling using HumanHap 300K BeadChip SNP arrays (data not shown).

To obtain more detailed insights into the structural genomic variations in this patient and into the possible genetic cause(s)

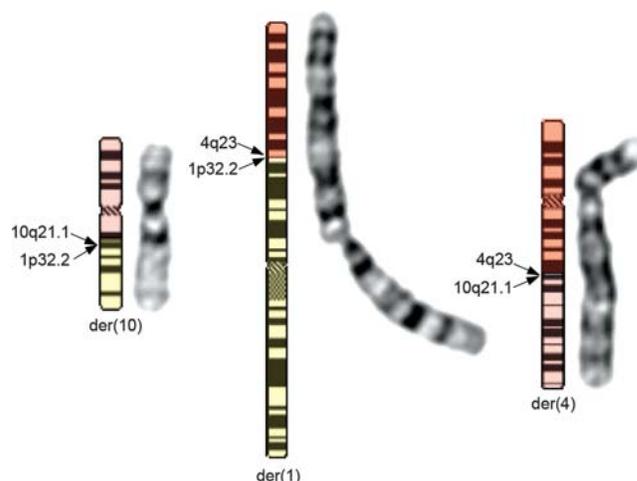


Figure 1. Partial karyotype of the patient. Karyotyping revealed a complex chromosomal rearrangement involving chromosomes 1, 4 and 10 (46,XY,t(1;10;4)(p32.2;q21.1;q23)dn). Karyotypes of the patient and his parents were ascertained in stimulated peripheral blood lymphocytes at a 550-band level according to ICSN2009 and following standard procedures.

Table 1. Mate-pair sequencing statistics

Library	Total mapped	Local ^a	Coverage	5 clones ^b	Mean clone size
Father	51.7 M	35.9 M	11.2×	96.90%	2429 ± 276
Mother	91.1 M	23.1 M	8.7×	91.60%	2354 ± 397
Child	52.7 M	26.8 M	10×	96.35%	2501 ± 263

^aLocal mate-pairs were determined based on a clone size smaller than 100 kb.

^bPercentage of genome covered by at least five clones.

of the clinical phenotype, we decided to perform genome-wide mate-pair sequence analysis of the patient and his parents. For each sample between 23.1 and 35.9 million concordantly and unambiguously mapping mate-pair clones with an average insert size of ~2.5 kb were generated resulting in 8.7–11.2× average clone coverage per haploid genome (Table 1; Supplementary Material, Fig. S1A).

Mate-pairs were split into local (mate-pair span < 100 kb) and remote (mate-pair span > 100 kb) categories (Table 1). The remote and anomalously mapping local mate-pairs were clustered to predict intrachromosomal rearrangements smaller than 100 kb (deletions, insertions, inversions and tandem-duplications), interchromosomal rearrangements and intrachromosomal rearrangements larger than 100 kb (12,24).

To gain more insight into the validity of our predictions and to experimentally determine a cut-off for the number of independent mate-pair clones that should minimally support a prediction of a structural rearrangement (when compared with the human reference genome), we used PCR-based capillary sequencing to verify a set of 66 predicted deletions in the father, mother and child ranging in size between ~2.5 and 80 kb (Supplementary Material, Table S1). We detected the (approximate) breakpoints of 39 out of 66 deletions and 37 of these were supported by at least 5 mate-pairs in at least 1 of the 3 libraries. For 31 out of 39 confirmed deletions, we could identify a sequence read overlapping the breakpoint.

Table 2. *De novo* breakpoints on chromosomes 1, 10 and 4

Id	chr1	ori	co1	chr2	ori	co2	Sequence in between breakpoints	Chromosomal pieces
De novo 1	4	H	105,025,700	10	H	57,519,913	4-ctcttggccttccaccatg-10 insertion	4:105,026,363–105,025,700 -> 10:57,519,913–57,520,041
De novo 2	4	T	105,745,783	4	H	105,035,150	at microhomology	4:105,745,506–105,745,783 -> 4:105,035,150–105,036,708
De novo 3	4	T	105,036,708	10	H	55,793,182	at microhomology	4:105,035,150–105,036,708 -> 10:55,793,182–55,793,321
De novo 4	4	T	102,287,791	10	H	57,523,805	4-actg-10 insertion	4:102,287,386–102,287,791 -> 10:57,523,805–57,524,597
De novo 5	10	H	57,521,088	4	H	105,028,395	10-agttataagt-4 insertion	10:57,521,337–57,521,088 -> 4:105,028,395–105,029,770
De novo 6	4	H	105,745,828	4	T	105,029,770	T microhomology	4:105,028,395–105,029,770 -> 4:105,745,828–105,745,982
De novo 7	10	T	55,793,180	4	H	104,738,996	aaat microhomology	10:55,792,170–55,793,180 -> 4:104,738,996–104,739,533
De novo 8	10	T	57,523,787	1	H	50,761,470	10-ttctggat-1 insertion; tggat is microduplication	10:57,523,465–57,523,787 -> 1:50,761,470–50,761,546
De novo 9	1	T	50,761,463	4	H	102,287,798	agg microhomology	1:50,761,351–50,761,463 -> 4:102,287,798–102,287,883
De novo 10	4	H	105,036,735	10	H	57,519,913	at microhomology	4:105,036,875–105,036,735 -> 10:57,519,913–57,521,100
De novo 11	4	T	105,028,400	10	T	57,521,100	4-aggcaaaagaa-10 insertion/microduplication	4:105,028,023–105,028,400 -> 10: 57,521,100–57,519,913
De novo 12	4	T	104,738,136	10	T	57,519,917	4-ct-10 insertion	4:104,737,918–104,738,136 -> 10:57,519,917–57,519,782

The breakpoint coordinates and orientation of chromosomal pieces for *de novo* structural changes. The data are based on capillary sequencing reads across the breakpoints. In many cases, either microhomology between the two fused chromosomal fragments was observed or a random or duplicated sequence was inserted at the breakpoint. H, head; T, tail.

We observed that 27 of these deletions have microhomology of 1–30 bp at their breakpoints when compared with the reference genome, whereas 4 have an inserted sequence of 6–86 bp, in line with recent reports (11,25–28).

In addition to this experimental approach to determine a cut-off for structural variation detection, we also estimated the distribution of the coverage of local mate-pair clones throughout the genome. To this end, we calculated the clone coverage for 5000 genomic DNA segments larger than 100 kb and with less than 500 bp of continuous repeat-masked DNA sequence. We found that 91.6–96.9% of all bases in these regions are covered by at least five independent mate-pairs in each of the three libraries (Table 1; Supplementary Material, Fig. S1B). Based on a cut-off of minimally five mate-pairs in at least one of the libraries, we predict 1764 intra-chromosomal rearrangements smaller than 100 kb (1015 deletions ranging from 1 to 80 kb, 209 inversions, 178 tandem duplications, 362 insertions ranging between 1.5 and 2.5 kb) and 2582 inter- and intrachromosomal rearrangement breakpoints involving distant (>100 kb apart) chromosomal fragments (Supplementary Material, Tables S2–S6 and Fig. S2A). Between 59 and 82% of these predicted rearrangements were found in all three libraries, indicating the presence of population specific haplotypes in the trio studied (29,30) (Supplementary Material, Fig. S2B).

Resolving a complex pattern of *de novo* chromosomal rearrangements at nucleotide resolution

We predicted 25 *de novo* structural rearrangements by filtering the predicted structural changes for variants that were uniquely found in the library of the patient and supported by at least five independent mate-pair clones (Supplementary Material, Table S7). To verify these predictions, we performed PCR-based capillary sequencing on DNA from the child and the parents. We obtained patient-specific PCR products for 15 predicted *de novo* structural changes. Surprisingly, 10 of

these 15 confirmed *de novo* variations occurred in the junction regions on chromosomes 1, 4 and 10 as indicated by the karyogram (Table 2). To identify additional *de novo* breakpoints, we filtered our predictions for rearrangements on chromosomes 1, 4 and 10 that were supported by only four mate-pairs in the child. This resulted in the prediction of two *de novo* rearrangements involving chromosomes 4 and 10 (Supplementary Material, Table S7; Table 2), both of which could be confirmed by capillary sequencing in the patient.

In total, we resolved 12 *de novo* junction points in confined areas of 2–3 Mb on chromosomes 1, 4 and 10, resulting in a series of chromosomal rearrangements that are far more complex than indicated by conventional cytogenetic analysis alone (Fig. 2; Table 2).

We reconstructed the rearranged chromosomes based on these 12 *de novo* junction points (Fig. 3). For the der(1) and the der(4) chromosomes, we could simply follow the junction points (links) as shown in Figure 2B to generate a complete picture of the rearranged chromosomes. This shows that der(4) is a simple translocation chromosome involving one junction between chromosome arm 4p fused tail-to-head to chromosome arm 10q. The der(1) chromosome appeared more complex with an insertion of a segment from chromosome 10 and a fragment from chromosome 4 in between the 4q and 1q chromosomal arms. Both reconstructed chromosomes matched with the der(1) and der(4) chromosomes as shown by the karyotype (Fig. 1). The der(10) chromosome shows the highest complexity and we show that the translocation breakpoint area for der(10) is highly complex involving eight rearranged chromosomal fragments derived from chromosomes 4 and 10 and ranging in size from 1188 bp to 2.4 Mbp. In conclusion, our mate-pair data enabled a reconstruction of three rearranged chromosomes with all fusion points described to the nucleotide precise.

In addition to the *de novo* breakpoints on chromosomes 1, 4 and 10, we predicted five *de novo* rearrangements elsewhere in the genome for which the two tags of the mate-pairs

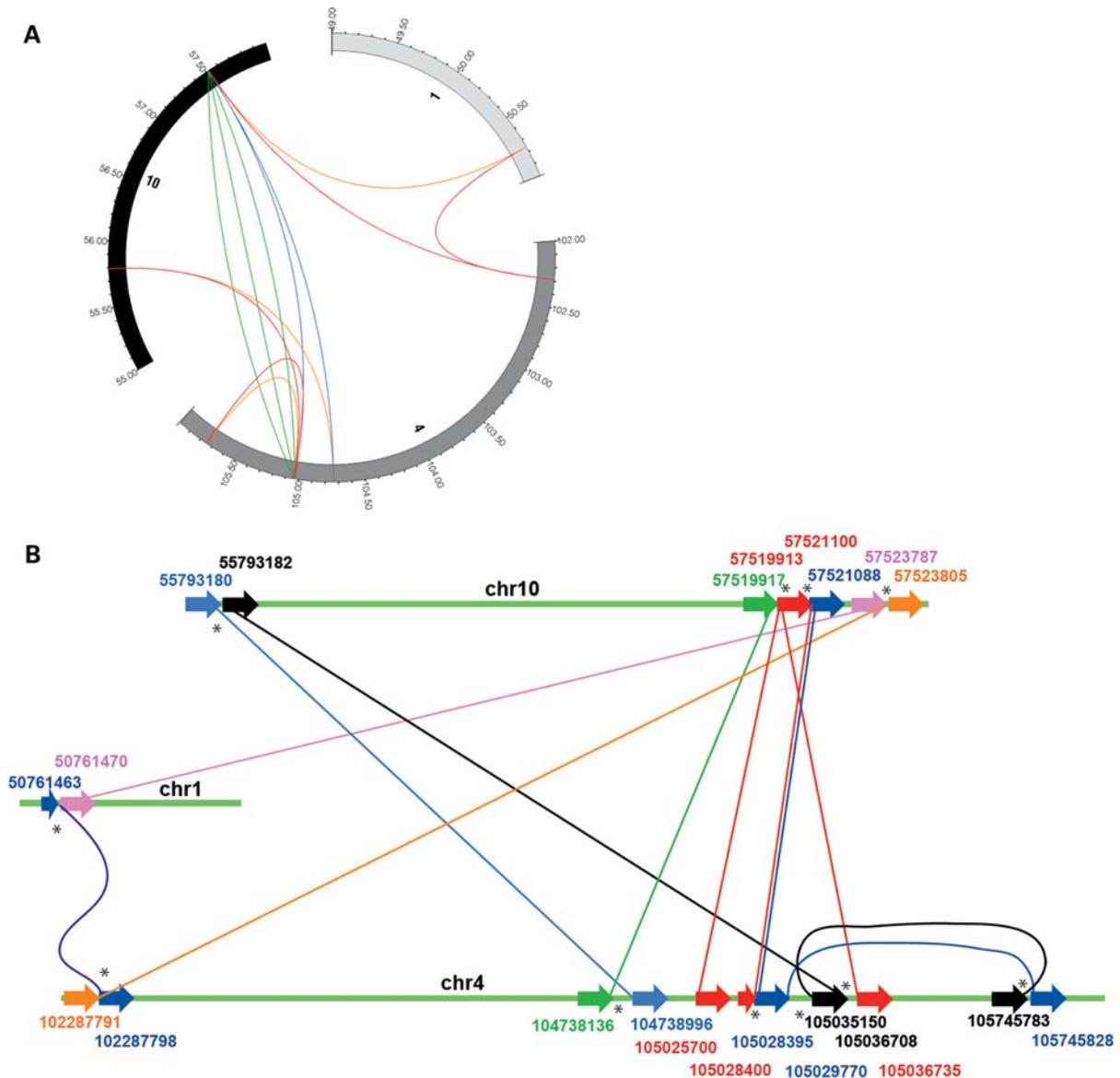


Figure 2. Overview of *de novo* rearrangements on chromosomes 1, 4 and 10. (A) Circos plot of 12 *de novo* breakpoints on chromosomes 1, 4 and 10. Chromosomal coordinates are indicated in mega base-pair. All breakpoints are located within a few mega base-pair of each other. Line coloring indicates orientation of fusion ends. Blue is head-head, green is tail-tail, red is tail-head, orange is head-tail (low chromosome number to high chromosome number). (B) Schematic picture showing chromosomal segments from chromosomes 1, 4 and 10. Colored lines connecting two chromosomal locations visualize the junction points. Interconnected fragments have the same color and the coordinates at which fragments are connected are indicated. Arrowheads at the end of each fragment indicate the orientation of chromosomal fragments. Some breakpoints are indicative of loss or gain of genomic sequence. For example, two remote genomic pieces have been connected to the head side of chromosome 10 at coordinate 57,519,913, indicating a duplication event. Similarly, fusion points were found on chromosome 4 at coordinates 105,036,708 (tail side) and 105,036,735 (head side), indicating loss of the sequence in between (see also Supplemental material, Table S9 for a complete list). Asterisks (*) indicate a double-stranded DNA break and paired links.

supporting these events map to homologous segmental duplications. We obtained patient-specific PCR products for these five rearrangements (rem8, rem10, rem14, rem17 and anti3) (Supplementary Material, Fig. S3 and Table S7). For three of these rearrangements, we also obtained PCR products of distinct sizes for the DNA of the father and the mother. Alignment of the capillary sequencing reads derived from the patient-specific PCR fragments to the human reference genome did not unambiguously lead to the identification of a breakpoint, because the sequence reads aligned with similar E-score to different homologous segmental duplicated regions.

Joining of shattered chromosomal fragments leads to formation of complex *de novo* structural rearrangements in one event

Next, we attempted to identify potential mechanisms that could account for the complex *de novo* rearrangements on chromosomes 1, 4 and 10. Careful examination of all fusion points on chromosomes 1, 4 and 10 shows that the links (lines in Fig. 2B) between different remote chromosomal fragments occur in pairs, with the paired links being in correct orientation with respect to each other; i.e. if one half of a

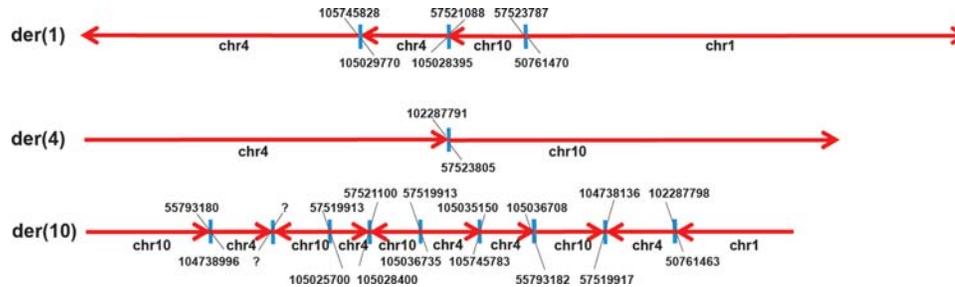


Figure 3. Reconstruction of derivative chromosomes. The der(1), der(4) and der(10) chromosomes were reconstructed based on the coordinates that we obtained for all *de novo* breakpoints. Arrows indicate the orientation of chromosomal segments. For the der(1) and der(4) chromosomes, we could generate a complete view of the fused chromosomal segments. However, we could not fully reconstruct the der(10) chromosome, since we did not find any evidence in our mate-pair data for the presumed 13th fusion point that is marked with question marks. Chromosomes and chromosomal fragments have not been drawn to scale.

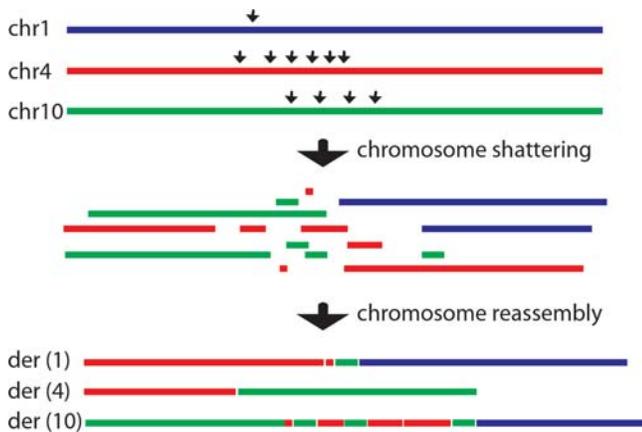


Figure 4. Model explaining the molecular events that lead to formation of complex rearrangements described in this paper. In a first step, double-stranded breaks (arrow heads) are introduced due to a locally acting exogenous or endogenous stimulus (e.g. ionizing radiation, reactive oxygen species, inadvertent topoisomerase activity). This leads to local shattering of chromosomes. Subsequently, a non-homologous end-joining process reassembles the chromosomal pieces. Joining of fragments may be driven by microhomology or occurs randomly based on fragments that are in close physical proximity. The resulting fusion points in the newly formed derivative chromosomes form concordantly oriented pairs on the reference genome, identical to what we observed for our patient (Fig. 2).

pair has its head side connected to another chromosomal fragment, the other half has its tail side connected to another chromosomal fragment. This highly interrelated pattern of paired links is indicative of a series of simultaneous double-stranded DNA breaks resulting in free chromosomal ends joined to other free chromosomal ends, which, in turn, also resulted from double-stranded breaks and follow the same orientation rules when connected to yet other chromosomal fragments. In simple form, a similar mechanism accounts for reciprocal translocations where two double-stranded breaks result in four free chromosomal ends which all accidentally join to an incorrect partner forming two derivative chromosomes (31,32). We here show that the same mechanism of chromosome breakage and reassembly may also underlie complex rearrangements.

To further substantiate our observations, we categorized the exact sequences at the junction points (Table 2; Supplementary Material, Fig. S4) (22,33,34). For six fusion points, we found an insertion of 2–20 bp between the two fused

chromosomal regions. The inserted sequence for one fusion point (*de novo* 11) forms a small duplication derived from the flanking segment from chromosome 4, indicating that synthesis of DNA may have been involved. Such short insertions are frequently observed for translocation junctions (16,35). For five fusion points, we found microhomology of up to four bases and for one breakpoint we observed blunt fusion (9,10,16,35). Combining position information of adjacent (paired) links indicates that sequences are frequently deleted (up to 5379 bp) at the breakpoints (Supplementary material, Table S9; Fig. 2B). Typically nucleolytic resection from the free ends that result from a double-stranded break involves up to 14 bases (14), but larger losses have been observed in *in vitro* non-homologous end-joining assays in mammalian cells (16). Loss of DNA may also result from dropout of a chromosomal fragment that arises from the shattering process (23).

All together, we observed microhomology, complete absence of homology, microduplications, small insertions and deletions at the junction points between different chromosomal fragments, indicating a non-homologous mechanism of break repair, such as NHEJ or alternative end joining (14,16,31). The concordant pattern of paired links indicates that double-stranded DNA breaks caused simultaneous and locally restricted shattering of chromosomes 1, 4 and 10, resulting in chromosomal fragments that were randomly joined together by non-homologous end-joining mechanisms to form complex derivative chromosomes (Figs 3 and 4). This process resembles a phenomenon, termed chromothripsis, which was recently proposed to explain complex somatic rearrangements in deranged cancer cells (23). Our results show that complex constitutional rearrangements may arise from a similar cascade of chromosome breakage and joining.

Since all 12 confirmed *de novo* structural events in the child were confined to a few mega base on chromosomes 1, 4 and 10, we explored whether the breakpoint regions displayed any distinct sequence characteristics that may promote chromosome breakage. Overall, GC content of breakpoint flanks was within the normal distribution for the respective chromosomes (data not shown). However, we found a slightly increased incidence of known repetitive elements at the 24 breakpoint coordinates (Fisher exact, $P = 3.43 \times 10^{-3}$; Supplementary Material, Table S8). However, it is unclear how this could lead to the observed complex patterns of double-stranded DNA breaks.

To determine whether the cluster of *de novo* breakpoints occurred between paternal or maternal chromosomes, we searched for informative polymorphic positions in the regions flanking the breakpoints. For one of the regions (4:105,028,395–105,029,770), we found a single nucleotide polymorphism (rs4699104) for which capillary sequencing showed that the mother carried only the G allele and the father carried the A allele. The G allele was found in the healthy chromosome of the patient, while the A allele was found in the breakpoint regions (Supplementary Material, Fig. S5). This suggests that the rearrangement has taken place in the chromosome of the father. This is in agreement with previous studies showing that paternal chromosomes are frequently a source of chromosomal rearrangements (6,36,37).

DISCUSSION

Here, we reported mate-pair sequence analysis of a family trio to identify genome-wide inherited and *de novo* structural variation at high resolution. We used long mate-pair libraries (2.5 kb insert size) as opposed to paired-end libraries, which are easier to generate than mate-pair libraries, but the insert sizes are limited to ~500 bp. Therefore, with paired-end libraries, it is more difficult to bridge repetitive regions, which are often hypersensitive sites for structural changes (11,27). In addition, reliable detection of structural variation requires sufficient physical genomic coverage, which can more easily be reached by long mate-pair libraries.

The long mate-pair sequencing approach used in this study proved to be very powerful to detect complex *de novo* non-recurring chromosomal rearrangements in a patient with multiple congenital abnormalities and mental retardation, and it allowed us to characterize the *de novo* fusion points at the base-pair level. Similar efforts to fine map *de novo* rearrangement breakpoints were limited to analysis of selected chromosomes and known translocation breakpoints (9,10). We demonstrate that systematic categorization of *de novo* structural variation is possible using mate-pair sequencing in a trio setup, irrespective of prior knowledge on the chromosomal positions of rearrangements in the subject of study. By comparing mate-pair data from the affected child with those from both parents, we were able to base our *de novo* structural variation predictions on the haplotype structures of the parents. Without parental information, identification of *de novo* events will be nearly impossible as structural variation within a population is very large (26,38), as is also emphasized by this study.

Based on our mate-pair analysis, the cluster of *de novo* chromosomal rearrangements on chromosomes 1, 4 and 10 appeared much more complex than expected from the karyotyping and segmental aneuploidy analysis, both of which are standard tools in genome diagnostics. There are two explanations for this: (i) our mate-pair data indicate that the vast majority of the rearrangements are balanced and (ii) the rearranged segments involve relatively small chromosomal segments (1188 bp to 2.4 Mb). Although we did observe small imbalances (deletions and duplications) associated with breakpoints regions, these are generally smaller than 50 bp.

We expect that next-generation (mate-pair) sequencing approaches for the detection of *de novo* structural variation in patients will uncover a much higher complexity of structural rearrangements than is currently seen and will lead to identification of structural changes in patients for whom a genetic diagnosis was thus far unattainable.

In total, we identified 12 fusion points between regions on chromosomes 1, 4 and 10. All of these were clustering in small genomic regions of up to 3.5 Mb in size. We found that the fusions (links) between chromosomes occur in concordantly oriented pairs, similarly as the two fusion points that form a reciprocal translocation after the occurrence of two double-stranded DNA breaks on two different chromosomes (32,39). Based on this highly concordant and locally restricted pattern of paired links, we conclude that the complex rearrangements we describe here have been triggered by a cascade of simultaneous double-stranded DNA breaks. It is unclear what caused the double-stranded DNA breaks, but an environmental stimulus, such as free radicals or ionizing radiation, may be involved (14,32). We regard it unlikely that the rearrangements resulted from independent events during consecutive cell divisions, since all fusions are tightly interrelated and locally restricted. The pattern of rearrangements that we observed here is reminiscent of a phenomenon, termed chromothripsis, that was very recently reported to underlie somatic rearrangements in deranged tumor cells (23). Our data show that this mechanism may also be responsible for complex constitutional rearrangements (Fig. 4).

It appears unlikely that FoSTeS/MMBIR has been driving this complex chromosomal rearrangement. To our knowledge, there is no good explanation on how a replication-based mechanism may account for such a concordant pattern of paired fusion points as observed here. Therefore, we hypothesize that non-recurrent rearrangements that constitute unique and complex disease phenotypes may be triggered entirely by accidental double-stranded DNA breaks induced by an environmental or endogenous factor. The resulting shattered chromosomal fragments are subsequently joined by non-homologous non-replicative mechanisms (14,20,22,40).

We likely missed one fusion point involved in the rearrangements on chromosomes 1, 4 and 10, since one of the chromosomal fragments involved has a loose end (Fig. 2B). Possibly, the fusion point involves a small chromosomal fragment and was therefore missed by our mate-pair analysis. For three fusion points, we identified the presence of small duplicated sequences (Supplementary Material, Table S9). Double-stranded DNA breaks may involve staggered nicks and subsequent fill-in of the unpaired sequences can result in small duplications (39). However, we also identified a larger duplicated segment (>129 bp) based on two segments from chromosome 4 that were linked to the same segment on chromosome 10 (10:57,519,913–57,520,041). Such long duplications have been described before for translocation breakpoints and in this case more complex multistep non-homologous mechanism may be involved in break repair (39).

One breakpoint that was identified in this study by mate-pair sequencing and not by standard diagnostics disrupts the *PCDH15* gene. In addition, three rearranged segments contain protein-coding genes. We have no proof that the *de novo* rearrangements are in general pathogenic, but in

light of the complexity of the phenotype, we regard it likely that at least some of the observed rearrangements may have contributed to the clinical phenotype.

Besides the complex rearrangements involving chromosomes 1, 4 and 10, we detected five apparently *de novo* rearrangements that appeared not related to the structural changes on chromosomes 1, 4 and 10. It turned out that the ends of the mate-pair clones supporting these five rearrangements mapped to homologous segmental duplications. The rearrangements likely originated independently from the rearrangements on chromosomes 1, 4 and 10 and may have occurred through non-allelic homologous recombination, similarly as the recurring rearrangements that underlie genomic disorders (29,41). Future studies should address the *de novo* rates of pathogenic and non-pathogenic structural changes in families containing healthy individuals to assess the structural plasticity of human genomes.

MATERIALS AND METHODS

Patient material

The Medical Ethics Committee (METC) of the University Medical Centre Utrecht, The Netherlands approved the genetic analysis of DNA from the patient and the parents. DNA samples were previously acquired as part of a series of routine diagnostic screenings of the patient in our genome diagnostics department. We obtained informed consent from the patient's parents to perform genetic analysis of DNA from the patient and the parents and to publish the observed findings.

Karyotyping and segmental aneuploidy profiling

Karyotypes of the patient and his parents were ascertained in stimulated peripheral blood lymphocytes at a 450-band level according to standard procedures. For genome-wide segmental aneuploidy profiling, Infinium HumanHap300 Genotyping BeadChip have been obtained from Illumina, Inc. (San Diego, CA, USA) and used as described (42).

Preparation of mate-pair libraries and SOLiD sequencing

Mate-paired libraries were generated from 35 to 91 μ g DNA isolated from peripheral blood samples. Mate-pair library preparation was essentially as described in the SOLiDv3.5 library preparation manual (Applied Biosystems). We only performed one DNA size selection directly after CAP adaptor ligation to select genomic fragments between 2 and 3 kb. Libraries were cloned and 384 clones per library were picked for capillary sequencing to assess the presence of adaptors, insert sizes and chimerism. To this end, mate-pair inserts were extracted from the capillary reads and blasted to the human reference genome. Mate-pairs with a tag distance larger than 100 kb were regarded as chimeric molecules. We sequenced 2×50 bp mates for each library. The library of the father was sequenced on two quadrants on a SOLiD V3.5 instrument, and the mother and child libraries were sequenced on three quadrants each.

Bioinformatic analysis of mate-pair reads

The F3 and R3 mate-pair tags were mapped independently to the human reference genome (GRCh37/hg19) using BWA software V0.5.0 with the following settings: `-c -l 25 -k 2 -n 10 (43)`. Mate-pair tags with unambiguous mapping were combined and split into local (<100 kb) and remote (>100 kb) mate-pair sets. Local mate-pairs were further split into mate-pairs with normal orientation of the tags relative to each other, mate-pairs with inverted tags and mate-pairs with everted tags (12).

Deletions were called from local mate-pairs with correct orientation and with a mate-pair span larger than 4000 bp and insertions were called from correctly oriented local mate-pairs with a mate-pair span smaller than 1100 bp. Tandem duplications were called from local mate-pairs with everted orientation and inversions were called from local mate-pairs with inverted orientation. Mate-pairs were clustered based on overlapping mate-pairs with a maximal tag distance of 5000 for deletions, 1000 for insertions, 5000 for tandem duplications and 5000 for inversions. The remote mate-pairs were clustered independent of the relative orientation of the mate-pair tags (maximal tag distance of 5000). The orientation of the different mate-pair tags in a cluster relative to each other is indicated by H (or h for the minus strand) when the tag has its 'head' side (the side that points towards the start of the chromosome) opposed to the pairing tag and T (or t for the minus strand) when a tag has its 'tail' side (the side that points towards the end of the chromosome) opposed to the pairing tag. We generated raw files with clusters for local deletions, local insertions, local inversion and local tandem duplications and remote rearrangements based on a minimum of two mate-pairs per rearrangement in total. These were used for subsequent filtering steps.

Coverage calculations

To estimate the genome-wide coverage in non-repetitive areas of the genome, we generated a set of genomic segments with less than 500 bp of continuous repeat-masked DNA sequence. From this set, we extracted the largest 5000 genomic segments, and coordinates of these were used to count the number of local and correctly oriented mate-pair tags that overlap any genomic position within these regions. From these calculations, we derived a coverage distribution profile that was used to set a cut-off for the detection of structural rearrangements. Average coverage was determined by calculating the sum of local mate-pairs spans and dividing this by $2 \times$ the haploid genome size.

Capillary sequencing of breakpoints and analysis of sequence reads

Primers for breakpoints of structural variants were designed using primer3 software avoiding repetitive elements if possible and according to the orientations that were indicated by the mate-pair tags. For *de novo* breakpoints, we designed nested PCR primers, while for the set of 66 (inherited) deletions, we performed single PCR reactions. PCR was performed with both Taq polymerase (Invitrogen) and Phusion

polymerase (Finnzymes) with elongation times of 1 or 2 min for Taq polymerase and 1 min for Phusion polymerase. PCR products were purified from gel when needed. Sequencing reads were aligned to the human reference genome (GRCh37) using BLAST and BLAT software. Hits were analyzed manually to define the exact breakpoint and breakpoint characteristics.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

We thank D. Lindhout for critical reading the manuscript.

Conflict of Interest statement. None declared.

FUNDING

This work was made possible by primary funding from the Department of Medical Genetics to E.C.

REFERENCES

- Batanian, J.R. and Esvara, M.S. (1998) De novo apparently balanced complex chromosome rearrangement (CCR) involving chromosomes 4, 18, and 21 in a girl with mental retardation: report and review. *Am. J. Med. Genet.*, **78**, 44–51.
- Lee, N.C., Chen, M., Ma, G.C., Lee, D.J., Wang, T.J., Ke, Y.Y., Chien, Y.H. and Hwu, W.L. (2010) Complex rearrangements between chromosomes 6, 10, and 11 with multiple deletions at breakpoints. *Am. J. Med. Genet. A*, **152A**, 2327–2334.
- De Gregori, M., Ciccone, R., Magini, P., Pramparo, T., Gimelli, S., Messa, J., Novara, F., Vetro, A., Rossi, E., Maraschio, P. *et al.* (2007) Cryptic deletions are a common finding in "balanced" reciprocal and complex chromosome rearrangements: a study of 59 patients. *J. Med. Genet.*, **44**, 750–762.
- Baptista, J., Mercer, C., Prigmore, E., Gribble, S.M., Carter, N.P., Maloney, V., Thomas, N.S., Jacobs, P.A. and Crolla, J.A. (2008) Breakpoint mapping and array CGH in translocations: comparison of a phenotypically normal and an abnormal cohort. *Am. J. Hum. Genet.*, **82**, 927–936.
- Gribble, S.M., Prigmore, E., Burford, D.C., Porter, K.M., Ng, B.L., Douglas, E.J., Fiegler, H., Carr, P., Kalaitzopoulos, D., Clegg, S. *et al.* (2005) The complex nature of constitutional de novo apparently balanced translocations in patients presenting with abnormal phenotypes. *J. Med. Genet.*, **42**, 8–16.
- Poot, M., van't Slot, R., Leupert, R., Beyer, V., Passarge, E. and Haaf, T. (2009) Three de novo losses and one insertion within a pericentric inversion of chromosome 6 in a patient with complete absence of expressive speech and reduced pain perception. *Eur. J. Med. Genet.*, **52**, 27–30.
- Houge, G., Liehr, T., Schoumans, J., Ness, G.O., Solland, K., Starke, H., Claussen, U., Stromme, P., Akre, B. and Vermeulen, S. (2003) Ten years follow up of a boy with a complex chromosomal rearrangement: going from a >5–15-breakpoint CCR. *Am. J. Med. Genet. A*, **118A**, 235–240.
- Chen, W., Kalscheuer, V., Tzschach, A., Menzel, C., Ullmann, R., Schulz, M.H., Erdogan, F., Li, N., Kijas, Z., Arkesteijn, G. *et al.* (2008) Mapping translocation breakpoints by next-generation sequencing. *Genome Res.*, **18**, 1143–1149.
- Chen, W., Ullmann, R., Langnick, C., Menzel, C., Wotschovsky, Z., Hu, H., Doring, A., Hu, Y., Kang, H., Tzschach, A. *et al.* (2010) Breakpoint analysis of balanced chromosome rearrangements by next-generation paired-end sequencing. *Eur. J. Hum. Genet.*, **18**, 539–543.
- Slade, I., Stephens, P., Douglas, J., Barker, K., Stebbings, L., Abbaszadeh, F., Pritchard-Jones, K., Cole, R., Pizer, B., Stiller, C. *et al.* (2010) Constitutional translocation breakpoint mapping by genome-wide paired-end sequencing identifies HACE1 as a putative Wilms tumour susceptibility gene. *J. Med. Genet.*, **47**, 342–347.
- Korbel, J.O., Urban, A.E., Affourtit, J.P., Godwin, B., Grubert, F., Simons, J.F., Kim, P.M., Palejev, D., Carriero, N.J., Du, L. *et al.* (2007) Paired-end mapping reveals extensive structural variation in the human genome. *Science*, **318**, 420–426.
- Medvedev, P., Stanciu, M. and Brudno, M. (2009) Computational methods for discovering structural variation with next-generation sequencing. *Nat. Methods*, **6**, S13–S20.
- Hastings, P.J., Lupski, J.R., Rosenberg, S.M. and Ira, G. (2009) Mechanisms of change in gene copy number. *Nat. Rev. Genet.*, **10**, 551–564.
- Lieber, M.R. (2010) The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu. Rev. Biochem.*, **79**, 181–211.
- McVey, M. and Lee, S.E. (2008) MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. *Trends Genet.*, **24**, 529–538.
- Simsek, D. and Jasin, M. (2010) Alternative end-joining is suppressed by the canonical NHEJ component Xrcc4-ligase IV during chromosomal translocation formation. *Nat. Struct. Mol. Biol.*, **17**, 410–416.
- Boboila, C., Jankovic, M., Yan, C.T., Wang, J.H., Wesemann, D.R., Zhang, T., Fazeli, A., Feldman, L., Nussenzweig, A., Nussenzweig, M. *et al.* (2010) Alternative end-joining catalyzes robust IgH locus deletions and translocations in the combined absence of ligase 4 and Ku70. *Proc. Natl Acad. Sci. USA*, **107**, 3034–3039.
- Weinstock, D.M., Elliott, B. and Jasin, M. (2006) A model of oncogenic rearrangements: differences between chromosomal translocation mechanisms and simple double-strand break repair. *Blood*, **107**, 777–780.
- Lee, J.A., Carvalho, C.M. and Lupski, J.R. (2007) A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell*, **131**, 1235–1247.
- Hastings, P.J., Ira, G. and Lupski, J.R. (2009) A microhomology-mediated break-induced replication model for the origin of human copy number variation. *PLoS Genet.*, **5**, e1000327.
- Smith, C.E., Llorente, B. and Symington, L.S. (2007) Template switching during break-induced replication. *Nature*, **447**, 102–105.
- Zhang, F., Khajavi, M., Connolly, A.M., Towne, C.F., Batish, S.D. and Lupski, J.R. (2009) The DNA replication FoSTeS/MMBIR mechanism can generate genomic, genic and exonic complex rearrangements in humans. *Nat. Genet.*, **41**, 849–853.
- Stephens, P.J., Greenman, C.D., Fu, B., Yang, F., Bignell, G.R., Mudie, L.J., Pleasance, E.D., Lau, K.W., Beare, D., Stebbings, L.A. *et al.* (2011) Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell*, **144**, 27–40.
- Zeitouni, B., Boeva, V., Janoueix-Lerosey, I., Loeillet, S., Legoix-ne, P., Nicolas, A., Delattre, O. and Barillot, E. (2010) SVDetect: a tool to identify genomic structural variations from paired-end and mate-pair sequencing data. *Bioinformatics*, **26**, 1895–1896.
- Conrad, D.F., Bird, C., Blackburne, B., Lindsay, S., Mamanova, L., Lee, C., Turner, D.J. and Hurles, M.E. (2010) Mutation spectrum revealed by breakpoint sequencing of human germline CNVs. *Nat. Genet.*, **42**, 385–391.
- Kidd, J.M., Cooper, G.M., Donahue, W.F., Hayden, H.S., Sampas, N., Graves, T., Hansen, N., Teague, B., Alkan, C., Antonacci, F. *et al.* (2008) Mapping and sequencing of structural variation from eight human genomes. *Nature*, **453**, 56–64.
- Vissers, L.E., Bhatt, S.S., Janssen, I.M., Xia, Z., Lalani, S.R., Pfundt, R., Derwinska, K., de Vries, B.B., Gilissen, C., Hoischen, A. *et al.* (2009) Rare pathogenic microdeletions and tandem duplications are microhomology-mediated and stimulated by local genomic architecture. *Hum. Mol. Genet.*, **18**, 3579–3593.
- Kidd, J.M., Graves, T., Newman, T.L., Fulton, R., Hayden, H.S., Malig, M., Kallicki, J., Kaul, R., Wilson, R.K. and Eichler, E.E. (2010) A human genome structural variation sequencing resource reveals insights into mutational mechanisms. *Cell*, **143**, 837–847.
- Antonacci, F., Kidd, J.M., Marques-Bonet, T., Teague, B., Ventura, M., Girirajan, S., Alkan, C., Campbell, C.D., Vives, L., Malig, M. *et al.* (2010) A large and complex structural polymorphism at 16p12.1 underlies microdeletion disease risk. *Nat. Genet.*, **42**, 745–750.

30. Kidd, J.M., Sampas, N., Antonacci, F., Graves, T., Fulton, R., Hayden, H.S., Alkan, C., Malig, M., Ventura, M., Giannuzzi, G. *et al.* (2010) Characterization of missing human genome sequences and copy-number polymorphic insertions. *Nat. Methods*, **7**, 365–371.
31. Lieber, M.R. (2010) NHEJ and its backup pathways in chromosomal translocations. *Nat. Struct. Mol. Biol.*, **17**, 393–395.
32. Tsai, A.G. and Lieber, M.R. (2010) Mechanisms of chromosomal rearrangement in the human genome. *BMC Genomics*, **11**(Suppl. 1), S1.
33. Beck, C.R., Collier, P., Macfarlane, C., Malig, M., Kidd, J.M., Eichler, E.E., Badge, R.M. and Moran, J.V. (2010) LINE-1 retrotransposition activity in human genomes. *Cell*, **141**, 1159–1170.
34. Iskow, R.C., McCabe, M.T., Mills, R.E., Torene, S., Pittard, W.S., Neuwald, A.F., Van Meir, E.G., Vertino, P.M. and Devine, S.E. (2010) Natural mutagenesis of human genomes by endogenous retrotransposons. *Cell*, **141**, 1253–1261.
35. Welzel, N., Le, T., Marculescu, R., Mitterbauer, G., Chott, A., Pott, C., Kneba, M., Du, M.Q., Kusec, R., Drach, J. *et al.* (2001) Templated nucleotide addition and immunoglobulin JH-gene utilization in t(11;14) junctions: implications for the mechanism of translocation and the origin of mantle cell lymphoma. *Cancer Res.*, **61**, 1629–1636.
36. Kaiser-Rogers, K.A., Rao, K.W., Michaelis, R.C., Lese, C.M. and Powell, C.M. (2000) Usefulness and limitations of FISH to characterize partially cryptic complex chromosome rearrangements. *Am. J. Med. Genet.*, **95**, 28–35.
37. Grossmann, V., Hockner, M., Karmous-Benailly, H., Liang, D., Puttinger, R., Quadrelli, R., Rothlisberger, B., Huber, A., Wu, L., Spreiz, A. *et al.* (2010) Parental origin of apparently balanced de novo complex chromosomal rearrangements investigated by microdissection, whole genome amplification, and microsatellite-mediated haplotype analysis. *Clin. Genet.*, **78**, 548–553.
38. Durbin, R.M., Abecasis, G.R., Altshuler, D.L., Auton, A., Brooks, L.D., Gibbs, R.A., Hurles, M.E. and McVean, G.A. (2010) A map of human genome variation from population-scale sequencing. *Nature*, **467**, 1061–1073.
39. Gajecka, M., Gentles, A.J., Tsai, A., Chitayat, D., Mackay, K.L., Glotzbach, C.D., Lieber, M.R. and Shaffer, L.G. (2008) Unexpected complexity at breakpoint junctions in phenotypically normal individuals and mechanisms involved in generating balanced translocations t(1;22)(p36;q13). *Genome Res.*, **18**, 1733–1742.
40. Zhang, F., Seeman, P., Liu, P., Weterman, M.A., Gonzaga-Jauregui, C., Towne, C.F., Batish, S.D., De Vriendt, E., De Jonghe, P., Rautenstrauss, B. *et al.* (2010) Mechanisms for nonrecurrent genomic rearrangements associated with CMT1A or HNPP: rare CNVs as a cause for missing heritability. *Am. J. Hum. Genet.*, **86**, 892–903.
41. Turner, D.J., Miretti, M., Rajan, D., Fiegler, H., Carter, N.P., Blayney, M.L., Beck, S. and Hurles, M.E. (2008) Germline rates of de novo meiotic deletions and duplications causing several genomic disorders. *Nat. Genet.*, **40**, 90–95.
42. Poot, M., Beyer, V., Schwaab, I., Damatova, N., Van't Slot, R., Prothero, J., Holder, S.E. and Haaf, T. (2010) Disruption of CNTNAP2 and additional structural genome changes in a boy with speech delay and autism spectrum disorder. *Neurogenetics*, **11**, 81–89.
43. Li, H. and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754–1760.