



Location of Balanced Chromosome-Translocation Breakpoints by Long-Read Sequencing on the Oxford Nanopore Platform

OPEN ACCESS

Liang Hu^{1,2,3,4†}, Fan Liang^{5†}, Dehua Cheng^{1,4}, Zhiyuan Zhang⁵, Guoliang Yu⁵, Jianjun Zha⁵, Yang Wang⁵, Qi Xia⁵, Daoli Yuan⁵, Yueqiu Tan^{1,2,3,4}, Depeng Wang⁵, Yu Liang^{5*} and Ge Lin^{1,2,3,4*}

Edited by:

Chuan-Le Xiao,
Sun Yat-sen University, China

Reviewed by:

Serap Yalin,
Mersin University, Turkey
Prashanth N. Suravajhala,
Birla Institute of Scientific Research,
India

*Correspondence:

Yu Liang
liangyu@grandomics.com
Ge Lin
linggf@hotmail.com

†These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Genomic Assay Technology,
a section of the journal
Frontiers in Genetics

Received: 03 September 2019

Accepted: 29 November 2019

Published: 14 January 2020

Citation:

Hu L, Liang F, Cheng D, Zhang Z,
Yu G, Zha J, Wang Y, Xia Q, Yuan D,
Tan Y, Wang D, Liang Y and Lin G
(2020) Location of Balanced
Chromosome-Translocation
Breakpoints by Long-Read
Sequencing on the Oxford
Nanopore Platform.
Front. Genet. 10:1313.
doi: 10.3389/fgene.2019.01313

¹ Institute of Reproduction and Stem Cell Engineering, School of Basic Medical Science, Central South University, Changsha, China, ² Department of Genetics, Reproductive and Genetic Hospital of CITIC-Xiangya, Changsha, China, ³ Key Laboratory of Reproductive and Stem Cell Engineering, National Health and Family Planning Commission, Changsha, China, ⁴ Department of Research, National Engineering Research Center of Human Stem Cells, Changsha, China, ⁵ GrandOmics Biosciences, Beijing, China

Genomic structural variants, including translocations, inversions, insertions, deletions, and duplications, are challenging to be reliably detected by traditional genomic technologies. In particular, balanced translocations and inversions can neither be identified by microarrays since they do not alter chromosome copy numbers, nor by short-read sequencing because of the unmappability of short reads against repetitive genomic regions. The precise localization of breakpoints is vital for exploring genetic causes in patients with balanced translocations or inversions. Long-read sequencing techniques may detect these structural variants in a more direct, efficient, and accurate manner. Here, we performed whole-genome, long-read sequencing using the Oxford Nanopore GridION sequencer to detect breakpoints in six balanced chromosome translocation carriers and one inversion carrier. The results showed that all the breakpoints were consistent with the karyotype results with only ~10× coverage. Polymerase chain reaction (PCR) and Sanger sequencing confirmed 8 out of 14 breakpoints; however, other breakpoint loci were slightly missed since they were either in highly repetitive regions or pericentromeric regions. Some of the breakpoints interrupted normal gene structure, and in other cases, micro-deletions/insertions were found just next to the breakpoints. We also detected haplotypes around the breakpoint regions. Our results suggest that long-read, whole-genome sequencing is an ideal strategy for precisely localizing translocation breakpoints and providing haplotype information, which is essential for medical genetics and preimplantation genetic testing.

Keywords: Oxford Nanopore Technologies, structural variants, balanced translocation, long-read sequencing, preimplantation genetic testing

INTRODUCTION

Structural variants (SVs), including translocations, inversions, deletions, and duplications, account for genetic disorders through damaging or changing functions of vital genes (Feuk et al., 2006; Conrad et al., 2010; Stankiewicz and Lupski, 2010; Collins et al., 2017). In particular, balanced chromosome translocation is caused by the interchange of chromosomal segments, whereas inversions occur inside a single chromosome by self-breakage and rearrangement. In most cases, balanced translocation/inversion has no immediate observable phenotype because the overall gene-copy number remains unchanged and all genes are expressed as normal. However, in a few cases, translocations/inversions have been reported to be associated with various diseases (Imaizumi et al., 2002; Aplan, 2006; Rizzolio et al., 2006; Fantes et al., 2008; Vandeweyer and Kooy, 2009; Sandberg and Meloni-Ehrig, 2010; Mikelsaar et al., 2012; Utami et al., 2014). Nevertheless, in most of these cases, we can only speculate that the translocations and inversions damage normal gene expression or function as the precise breakpoints remain unknown.

Karyotyping, fluorescence *in situ* hybridization (FISH), and Southern blot are the traditional approaches for detecting translocations/inversions at the chromosome level. Karyotype analysis is the most widely used and cost-efficient method at present; however, it can only discover breakpoints at the chromosome level, which usually contains dozens or even hundreds of genes (Pasquier et al., 2016). Precisely designed FISH and Southern blot for specific cases can localize the breakpoints at a single gene level; however, results obtained with these strategies can also not be used for generalization. In addition, these techniques cannot accurately retrieve the sequences of breakpoints, and it is difficult to determine the specific impact of the chromosome translocation on the gene structure (Schluth-Bolard et al., 2013). With the development of sequencing technology, next-generation sequencing (NGS) serves as a new method for translocation detection and breakpoint analysis (Abel and Duncavage, 2013; Schluth-Bolard et al., 2013; Dong et al., 2014; Utami et al., 2014). Translocation detection by NGS usually uses the mate-pair strategy according to the coordinate, strand, and orientation of pair-end reads due to the disadvantage of producing short read lengths (Yao et al., 2012). Moreover, when breakpoints are located in complex repetitive regions with low mapping rate, it is difficult to accurately detect their location when using NGS.

Nanopore sequencing, a single-molecule long-read sequencing technology, was first independently proposed by Deamer, Branton, and Church (Pennisi, 2012), and rapid and great improvements in this technology, as well as bioinformatics tools, have made it a state-of-the-art approach for clinical testing, overcoming the limitations of short-read sequencing. However, it has a relatively high error rate, which currently hinders its application in detecting single-nucleotide substitutions and small frameshift mutations (Tsiatis et al., 2010) under low-coverage conditions. Notwithstanding, its generation of long-read lengths (> 10 kilobases on average) would greatly improve SV detection

regardless of whether or not the SVs are located in repetitive regions and enable the discovery of translocation breakpoints.

Long reads are especially helpful in resolving breakpoints in repetitive genomic regions with transposable elements. Transposable elements, including DNA transposons and retrotransposons, are major contributors to genomic instability. Endogenous retroviruses, long-interspersed elements (LINEs), and short-interspersed elements (SINEs) are classified as retrotransposons. Alu elements, one type of SINEs, represent the most widely scattered retrotransposons in primate genomes, accounting for 10% of the human genome (Szmulewicz et al., 1998). Genomic rearrangements induced by Alu insertions account for approximately 0.1% of human diseases, and genomic deletions mediated by Alu transpositions are responsible for approximately 0.3% of human genetic disorders (Callinan et al., 2005; Sen et al., 2006; Hancks and Kazazian, 2012).

Long reads are also useful for resolving haplotypes between translocations and nearby SNPs or indels, which are of particular importance in preimplantation genetic diagnosis (PGD). Due to the presence of allelic drop-out when assaying single cells in PGD, markers along a very long stretch of DNA can indicate whether the chromosome carries a translocation in an embryo. This method, known as preimplantation genetic haplotyping, is a simple, efficient, and widely used method for identifying and distinguishing all translocation forms in cleavage-stage embryos before implantation (Zhang et al., 2017). Informative haplotypes are usually generated by polymorphic markers that cover two megabases up- and down-stream around breakpoints.

Balanced translocation occurs in approximately 0.2% of the human population and 2.2% in patients with a history of recurrent miscarriages or repeated *in vitro* fertilization failure (Ogilvie CM and Scriven, 2001; Alfarawati et al., 2011). In somatic cells, chromosomes with balanced translocation can undergo normal mitosis and genomic replication. However, during meiosis, chromosomes carrying balanced translocations are prone to abnormal segregation, leading to a variety of unbalanced translocations (up to approximately 70%), which are derivatives with duplication and deletion of terminal sequences on either side of the breakpoint (Scriven, 1998; Munne, 2005). Thus, parents carrying chromosomes with balanced translocation are confronted with common problems, including inability to conceive, multiple miscarriages, and giving birth to children with a chromosomal disease syndrome (Suzumori and Sugiura-Ogasawara, 2010). These couples commonly seek help from assisted reproduction technology (ART) and PGD, which can identify balanced euploid embryos for intrauterine transplantation and subsequent development into a healthy infant (Munne, 2005; Fischer et al., 2010). Hence, the precise location of translocation breakpoints is of great importance for increasing the success rates of ART, considering the economic and psychological burdens to families.

In this study, we demonstrated the ability of Oxford Nanopore sequencing to detect translocations and localize their breakpoints, which were initially detected by conventional karyotyping. Fourteen breakpoints from seven carriers were

identified successfully. We also obtained haplotype information near the breakpoint regions, facilitating single-cell sequencing in PGD. Our results indicate that low-coverage, whole-genome sequencing is an ideal method for precisely localizing translocation breakpoints, which may be widely applied in SV detection, therapeutic monitoring, ART, and PGD.

MATERIALS AND METHODS

Samples

The study was approved by the Institutional Review Board of the CITIC-Xiangya Reproductive and Genetics Hospital, and written informed consent was obtained from all participants. A total of seven patients, including three with long-standing infertility, were recruited at the CITIC-Xiangya Reproductive and Genetics Hospital. Among them, six balanced translocations and one inversion were previously identified by karyotyping. The mean maternal age was 30.4 years (21–34 years), indicating a moderate risk of incidental aneuploidy. This study included three female carriers and four male carriers. DNA was extracted using the FineMag Blood DNA Kit (GENFINE BIOTECH), according to the manufacturer's instructions.

Library Preparation and Sequencing

Genomic DNA (5 µg) was sheared to ~5–25-kilobase fragments using Megaruptor[®] 2 (Diagenode, B06010002) and was then size-selected (10–30 kilobases) with a BluePippin device (Sage Science, MA) to remove the small DNA fragments. Subsequently, genomic libraries were prepared using the Ligation Sequencing 1D Kit (SQK-LSK108, Oxford Nanopore, UK). End-repair and dA-tailing of DNA fragments were performed using the Ultra II End Prep module (New England Biolabs, E7546L), according to manufacturer's recommended protocols. Finally, the purified dA-tailed sample was incubated with blunt/TA ligase master mix (#M0367, NEB), tethered with 1D adapter mix from the SQK-LSK108 Kit (Oxford Nanopore Technologies), and purified. The resulting library was sequenced on R9.4 flow cells using GridION X5.

SV Analysis

The raw sequencing data were in FAST5 format and converted to FASTQ format using the MINKNOW local basecaller. SVs were called using a pipeline that combines NGMLR-sniffles and LAST-NanoSV. Briefly, long reads were aligned to the human reference genome (hg19) using NGMLR (Sedlazeck et al., 2018) (version 0.2.6) with “-x ont” argument and LAST (version 912) separately, then SV calling was performed with sniffles (version 1.0.6) using “-report_BND -ignore_sd -q 0 -genotype -n 10 -t 20 -l 50 -s 1” and NanoSV (Cretu Stancu et al., 2017) with “-c 1” arguments. To improve the sensitivity of translocation calling, a custom Python script was developed to obtain all split reads that mapped to different chromosomes. In addition, alignment information related to the identity, mapping quality, matching location, and matching length was retained. Integrative Genomics Viewer (IGV) (Robinson et al., 2011) and Ribbon

(Nattestad et al., 2016) were used for visual examinations of translocations in target regions. Inversions were detected by combining the results of sniffles and NanoSV.

Breakpoint Verification

We designed PCR primers to detect the translocation breakpoints for each sample. Primer3Plus (<http://primer3plus.com/>) was used for primer design. The sequences of all primers used in this study are provided in **Table S1**. PCR was performed using 2× Taq Plus Master Mix polymerase (P211-01/02/03, Vazyme), and the products were electrophoresed on a 1.0% agarose gel and sequenced by Sanger sequencing on an ABI3730XL sequencer (Applied Biosystems).

Haplotype Analysis

MarginPhase is a method that uses a hidden Markov model to segment long reads into haplotypes (Ebler et al., 2019). After identifying candidate SVs using the combined pipeline described above, we obtained 2 megabases of sequence data in both upstream and downstream of the breakpoint. To identify mutations, SNPs/indels were first called using SAMtools mpileup and bcftools. Finally, we generated haplotype calls using MarginPhase.

Copy Number Variant (CNV) Analysis

CNV analysis was performed by Xcavator, a software package for CNV identification using short and long reads from whole-genome sequencing experiments (Magi et al., 2017). During the sequencing process, each read was randomly and independently sequenced, and the copy number of any genomic region could be estimated by counting the number of reads (read count) aligned to consecutive and non-overlapping windows of the genome. Given the low sequencing coverage (0–10×), we selected a 10 kb window size with no control mode.

RESULTS

Chromosomal Analysis of Carriers With Balanced or Inversion Translocations

We recruited seven carriers with translocations for the study from CITIC-Xiangya Reproductive and Genetics Hospital (**Table 1**). These subjects had either long-standing infertility, a history of recurrent miscarriages, or children with chromosome-related syndromes. Approximately 5 ml of blood was obtained from each carrier, 2 ml of which was mixed with peripheral blood culture medium and cultured in an incubator at 37°C. After 72 h, chromosome specimens were prepared and subjected to a G-banding karyotype analysis by standard protocols, according to the International System for Human Cytogenetic Nomenclature. The results revealed that six carriers had reciprocal balanced translocations and one carrier had an inversion translocation (**Figure S1**). We performed whole-genome, long-read sequencing analysis on all subjects to find the precise coordinate of breakpoints. Based on the karyotyping results, we chose different

TABLE 1 | The list of subjects analyzed in the current study and the details on the inferred breakpoints.

Sample	Karyotype	Depth (X)	No. of mapped sequencing reads	No. of mapped sequencing bases	Coverage rate (%)	No. of spanning breakpoints reads	Breakpoint position (GRCh37)	Disrupted gene (breakpoint)
DM17A2236	46,XY,t(6;8)(q25;q22)	11.32	2,262,314	32,111,789,470	91.85	11	6:167281717 8:113696089	Intergenic region <i>CSMD3</i>
DM17A2237	46,XX,t(18;21)(q11;q11)	10.31	2,316,017	29,746,593,714	93.44	11	18:28685658 21:29073597	<i>DSCAS</i> Intergenic region
DM17A2246	46,XX,t(8;22)(q24;q11)	9.87	1,931,784	28,742,307,402	94.34	6	8:125495366 22:20326956~20327048	<i>RNF139</i> Intergenic region
DM17A2247	46,XY,t(11;22)(q23;q11)	9.98	2,024,838	29,361,507,192	95.30	5	11:116683166 22:20326993	Intergenic region Intergenic region
DM17A2248	46,XX,inv(11)(q11q21)	10.94	2,498,061	32,758,847,457	96.96	10	11:58265643 11:100448937	Intergenic region Intergenic region
DM17A2249	46,XY,t(2;18)(p13;q23)	10.26	1,790,385	29,628,601,968	93.52	11	2:80320441 18:66637011	<i>CTNNA2</i> <i>CCDC102B</i>
DM17A2250	46,XX,t(3;9)(p13;p13)	13.54	3,150,533	39,494,541,253	94.43	7	3:90504855 9:44216447	Centromere region Intergenic region

analytical strategies and tools to analyze the translocation breakpoints in the next step.

DNA Extraction and Sequencing With the GridION X5 Instrument

For all subjects, genomic DNA was sheared to 10–20-kilobase fragments, and DNA libraries were prepared and sequenced using standard protocols on the Oxford Nanopore GridION X5 sequencer. For all samples, the mean and median read identity to the reference genome was mostly higher than 85% (Figure 1A). We obtained 32–44 gigabases of sequence data for each sample, with a mean read length of 12.3–16.3 kb and a depth of 9.87–13.54× (Figure 1B). These results suggested that we obtained high-quality sequencing data to facilitate downstream analysis. After sequencing, all reads generated for each sample were aligned to the human reference genome (hg19) and used for subsequent downstream data analysis. The detailed results are summarized in Table S2 and Table S3.

Translocation Detection and Breakpoint Characterization

We analyzed the long-read sequencing data obtained with the Oxford Nanopore platform to detect breakpoints in six individuals with balanced translocations and one individual with an inversion, using a custom bioinformatics pipeline that incorporated several existing tools (Figure 1C). This bioinformatics pipeline identified potential breakpoints from the alignment data. We successfully discovered 14 breakpoints in the seven carriers, and the breakpoint locations were consistent with the karyotyping results. For each breakpoint, around 10 reads were covered, as illustrated in Figures 2A, B. Detailed information regarding the breakpoints and sequencing data quality for the seven samples is summarized in Figure S2, Figure S3, and Table S2.

Checking these breakpoints in the UCSC Genome Browser, we found breakpoints inside introns of genes *CSMD3*, *AK129567*, *AK302545*, *RNF139*, and *CCDC102B* in samples DM17A2236, DM17A2246, DM17A2247, and DM17A2249

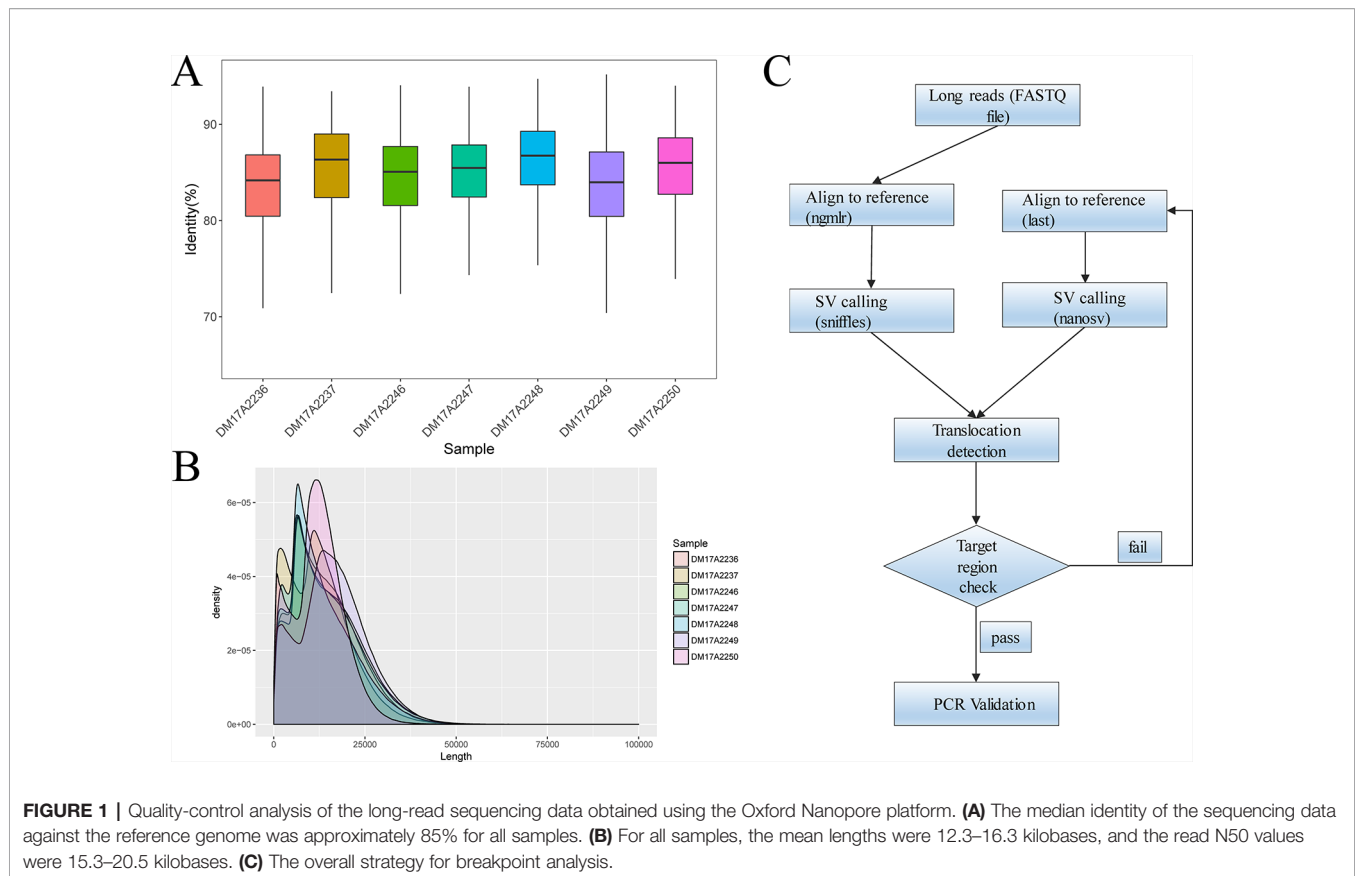
(Table 1). Therefore, these breakpoints disrupted the gene structures, causing the exchange of chromosomal segments, thereby impairing gene function since a portion of a gene in one chromosome is moved to another chromosome. However, there was no obvious impact on the phenotype of the carriers from whom the above four samples were obtained, except for primary infertility. We also found that the aligned sequence of DM17A2246 was located at 22q11.21 with a 79 bp deletion (chr22:20656022–20656100). DM17A2247 had a 33 kb gap (chr22:206326985–20656120). These results indicate that micro deletions/insertions often occur in conjunction with translocations/inversions, even though the underlying mechanism remains unknown. Furthermore, clusters of low-copy repeats (LCRs) occur in 22q11.21 of DM17A2246, which suggests the possible mechanism of translocation occurrence.

We found that in sample DM17A2237, the breakpoint at chr18:28685658 occurred in an *AluY* element; yet, in sample DM17A2250, the breakpoint at chr9:44216447 occurred in an *AluSx3* element. In sample DM17A2249, we found the breakpoint in the *LIPA4* region, which is a LINE element.

Interestingly, sample DM17A2250 was found to have a karyotype of 46, XX, t(3;9) (p13;p13), whose coordinates are chr3:90,490,057–90,504,855 and chr9:44, 225,822, respectively. The breakpoint on chromosome 3 was very close to the acrocentric centromere. Parts of all the long reads that support the breakpoint in chromosome 9 were mapped to an alpha satellite near the gap caused by the centromere. Due to the gap region of the reference genome (hg19), the position of the breakpoint was imprecise. However, these long reads provide strong evidence that the breakpoint in the centromere region is consistent with the karyotyping results. All these observations show the ability of long reads for breakpoint detection in such low complexity genome regions.

Inversion Detection and Breakpoint Characterization

Similar to balanced translocations, inversions do not change the chromosome copy number, and they are difficult to detect using



conventional short-read sequencing technology, although they have vital functional consequences in medical genetics (Puig et al., 2015). Here, we successfully detected an inversion occurring in carrier DM17A2248 at chr11:58,255,398–58,293,470 and chr11:100,430,372–100,461,378 (**Figure S2**). After verification by PCR and Sanger sequencing, the breakpoints were finally identified as chr11:58,265,643 and chr11:100,448,937, respectively, consistent with the karyotyping results. Our results demonstrated an example where long-read sequencing was capable of accurately resolving complex breakpoints for inversions.

Breakpoint Validation by Sanger Sequencing

To further validate the exact translocation breakpoints and neighboring SNPs, PCR and Sanger sequencing were performed to extract the breakpoint sequences at the level of single bases. For translocations, we successfully identified breakpoints in samples DM17A2236, DM17A2237, DM17A2248, and DM17A2249 by Sanger sequencing, but not in samples DM17A2246, DM17A2247, and DM17A2250 (**Figure 2C** and **Figure S4**). Because the approximate breakpoints in samples DM17A2246 and DM17A2247 were located in highly repetitive regions and the breakpoint in sample DM17A2250 was near a centromere, it was challenging to obtain a PCR product for these breakpoints, despite multiple attempts. Nevertheless, it is worth noting that for sample

DM17A2247, we successfully obtained the target PCR bands from the normal chromosome (without translocations), but no band was found reflecting rearranged chromosomes (**Figure S5**), suggesting that a deletion or larger insertion near the breakpoints may have broken the binding sites of our primers. The results above further suggest the power of long-read sequencing in detecting the precise locations of translocation breakpoints, whereas karyotype analysis can only provide crude results at the megabase level. Therefore, long-read sequencing may be a more precise tool for detecting translocation breakpoints and may complement or validate karyotyping results in clinical-diagnostic settings.

Haplotype Detection

Haplotype identification of chromosomes is of great importance to PGD, such that adjacent SNP information can be used to predict the presence or absence of balanced translocations in single-cell assays. Here, we performed haplotype analysis by using the breakpoints as precise markers. Through these markers, we successfully found informative SNPs near the breakpoint regions, which enabled differentiation of the chromosomal regions involved in the translocation (and the corresponding normal homologous chromosomes) in sample DM17A2237 at a low-level sequencing coverage (10×) (**Figure 3**). Haplotypes can help distinguish between embryos with balanced translocations and structurally normal chromosomes through PGD analysis in cases where the spouse

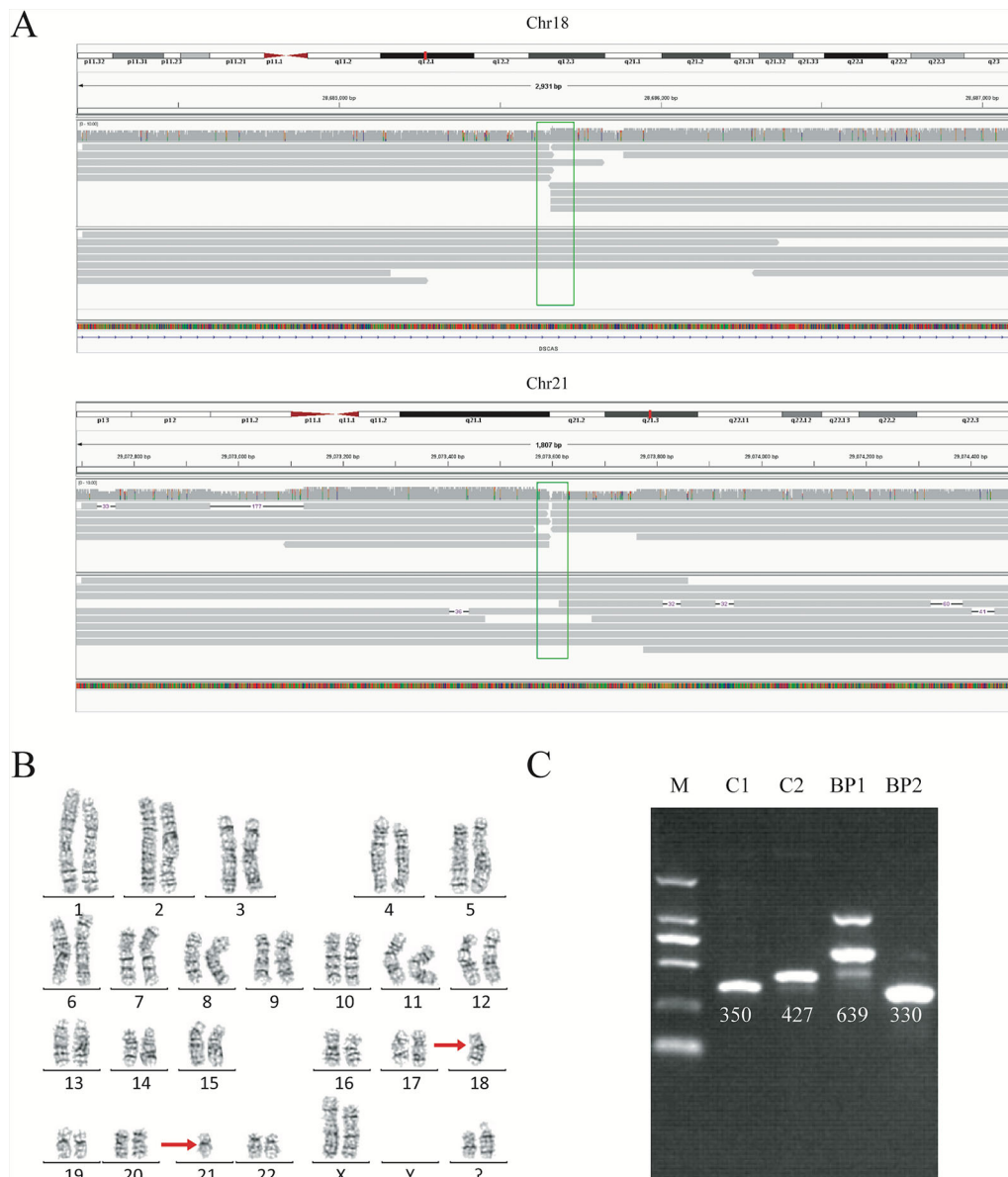


FIGURE 2 | Balanced translocation by sequencing and karyotyping for subject DM17A2237. **(A)** Read mapping of the breakpoints for the balanced translocations. DNA fragments were compared to the reference human genome (GRCh37/hg19), and the breakpoints were shown in IGV. Twenty reads adjacent to the breakpoint were obtained. **(B)** Karyotype of carrier DM17A2237. Karyotype analysis was determined from G-banding analysis, following a standard protocol. The karyotype result revealed an approximate region where the breakpoint occurred. **(C)** PCR analysis and Sanger sequencing for validating the breakpoints. An ethidium bromide-stained agarose gel was showing the presence of two new bands created by the rearrangement of chromosomal segments at breakpoints (BP1 and BP2). BP, breakpoint; C, control; M, marker. Primer information is available in **Table S1**.

of a carrier has a normal karyotype. These results demonstrate that it is possible to determine haplotypes by low-coverage long-read sequencing.

Exploratory Analysis of CNVs by Low-Coverage Long-Read Sequencing

CNV is an essential type of SVs, and the identification of CNVs is also useful for clinical diagnoses. Using Xcavator with a 100-kb

window size, no CNVs beyond 1 Mb were found in all the samples (**Table S4**). Since our study focused on translocations that were already identified by karyotyping, we did not perform a more detailed analysis for CNVs. However, these results and simulations demonstrate that even with low-coverage data, long-read sequencing still can detect a large number of potential CNVs and may be used to validate candidate CNVs that are detected by other platforms such as SNP arrays.

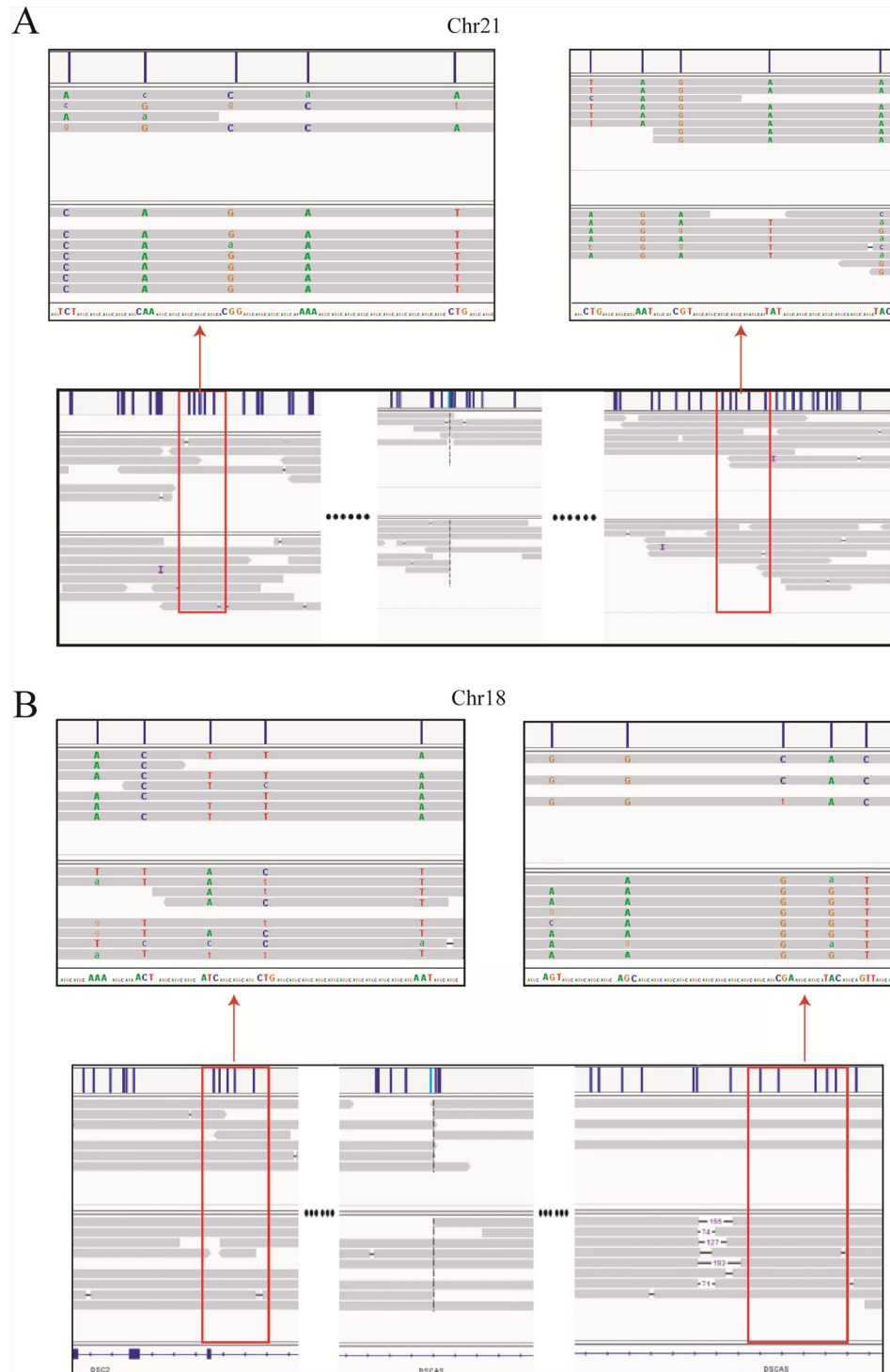


FIGURE 3 | Long-read sequencing enabled haplotype detection around the translocation breakpoints in sample DM17A2237. Using the breakpoints as anchoring markers, we obtained 2-megabase sequences on either side of the breakpoints. Through SNP calling and the MarginPhase tool, we phased the haplotypes around the breakpoints in chr18 (**A**) and chr21 (**B**). Reads around breakpoints were shown in IGV (bottom panel) and regions in red box were enlarged (top panel). Capital letters represent accurate sequencing information, whereas lowercase letters represent fuzzy base information.

DISCUSSION

Currently, karyotype analysis is the most widely used technology for clinically diagnosing chromosomal translocations (Comas et al., 2010). However, karyotype analysis is a low-resolution method that cannot identify exact breakpoints, which are often required for a better understanding of how translocations impact genes and phenotypes. NGS technology enables high-resolution and high-throughput analysis (Abel and Duncavage, 2013; Schluth-Bolard et al., 2013). However, because it generates short read lengths, paired-end or mate-pair libraries with large DNA inserts (usually larger than 2 kb) are always used for SV detection, as larger DNA insert sizes have been shown to be more advantageous in terms of SV detection in complicated DNA sequences, such as repetitive regions or large genomic rearrangements. Moreover, larger DNA insert size libraries also provide higher physical coverage with minimum sequencing efforts than smaller insert sizes (Yao et al., 2012; Van Heesch et al., 2013). Nanopore technology yields longer reads than NGS. In this study, reads longer than 100 kb were detected in each library, and we could obtain not only the two ends of the template generated by NGS, but also the entire DNA sequence. Thus, we believe that nanopore is a more powerful tool for translocation and other SV detection.

In this study, we analyzed genomic variations in seven patients with long-term reproductive disorders. All seven patients carried chromosomal translocations in their genomes, with six having reciprocal balanced translocations and one having an inversion. We successfully identified and sequenced every breakpoint in these seven carriers by long-read sequencing. All 14 breakpoints identified by long-read sequencing were consistent with their corresponding karyotype results. Moreover, we found that the breakpoints in four carriers (DM17A2246, DM17A2249, DM17A2237, and DM17A2250) occurred in repetitive regions; the breakpoints in DM17A2246 were located in the LCR region, those in DM17A2249 occurred in LINE, and those in DM17A2237 and DM17A2250 occurred in Alu. This finding provides strong evidence that long-read sequencing shows flexibility in sequence preferences, even if the breakpoints appear in highly repetitive and complex regions.

Furthermore, PCR analysis of samples DM17A2249 and DM17A2248 showed clear target bands for the wild-type copies at the breakpoint sites but failed to generate any band for one or both breakpoints in the homologous chromosomes carrying the translocations. Reciprocal chromosome translocations are often accompanied with some additional rearrangements, such as deletions and duplications, involving only a few base pairs or up to millions of bases. As previously reported, almost 50% of balanced translocations show large deletions and duplications at the breakpoint junction (De Gregori et al., 2007; Howarth et al., 2011). The failure in breakpoint identification by PCR in samples DM17A2249 and DM17A2248 may be due to the existence of this kind of rearrangement, where a deletion leads to loss of PCR primer-binding site(s) or a large insertion makes the PCR product too long to be amplified.

In conclusion, by taking advantage of long reads, low-coverage whole-genome sequencing could be a more efficient and powerful

tool for analyzing chromosomal translocations than traditional methods such as FISH and NGS. By comparing karyotyping and Sanger sequencing results, we confirmed that nanopore sequencing exhibits high resolution and accuracy. We believe that long-read sequencing may play a more important role in chromosomal-translocation analysis and breakpoint detection in the future, as well as offer valuable insights for assisting the genetic diagnosis of reproduction and preimplantation.

AUTHOR'S NOTE

This manuscript has been released as a Pre-Print at <https://www.biorxiv.org/content/141953/419531v1> (Liang et al., 2018).

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the NCBI (PRJNA559962).

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Reproductive and Genetic Hospital of CITIC-Xiangya. The patients/participants provided their written informed consent to participate in this study. The research, including human subjects, human data and material, has been performed in accordance with the Declaration of Helsinki.

AUTHOR CONTRIBUTIONS

GL, YL, and DW designed the research. LH, FL, ZZ, and DY wrote the manuscript. DC, GY, JZ, YW, QX, and YT executed the research. DW and GL directed the discussion of the manuscript. All authors approved the final manuscript.

FUNDING

This study was partially supported by the National Key R&D Program of China (2018YFC1003100) and National Natural Science Foundation of China (81571497) Merck Serono China Research Fund for Fertility Experts.

ACKNOWLEDGMENTS

The authors want to thank patients who participated in this study to evaluate novel genomic approaches for improving

genetic diagnosis of balanced translocations and inversions. We also thank the genetic counselors and clinical geneticists who interviewed the patients and collected DNA samples. We also thank Dr. Kai Wang for his guidance on structural variations analysis.

REFERENCES

- Abel, H. J., and Duncavage, E. J. (2013). Detection of structural DNA variation from next generation sequencing data: a review of informatic approaches. *Cancer Genet.* 206, 432–440. doi: 10.1016/j.cancergen.2013.11.002
- Alfarawati, S., Fragouli, E., Colls, P., and Wells, D. (2011). First births after preimplantation genetic diagnosis of structural chromosome abnormalities using comparative genomic hybridization and microarray analysis. *Hum. Reprod.* 26, 1560–1574. doi: 10.1093/humrep/der068
- Aplan, P. D. (2006). Causes of oncogenic chromosomal translocation. *Trends Genet.* 22, 46–55. doi: 10.1016/j.tig.2005.10.002
- Callinan, P. A., Wang, J., Herke, S. W., Garber, R. K., Liang, P., and Batzer, M. A. (2005). Alu retrotransposition-mediated deletion. *J. Mol. Biol.* 348, 791–800. doi: 10.1016/j.jmb.2005.02.043
- Collins, R. L., Brand, H., Redin, C. E., Hanscom, C., Antolik, C., Stone, M. R., et al. (2017). Defining the diverse spectrum of inversions, complex structural variation, and chromothripsis in the morbid human genome. *Genome Biol.* 18, 36. doi: 10.1186/s13059-017-1158-610
- Comas, C., Echevarria, M., Carrera, M., and Serra, B. (2010). Rapid aneuploidy testing versus traditional karyotyping in amniocentesis for certain referral indications. *J. Matern. Fetal Neonatal Med.* 23, 949–955. doi: 10.3109/14767050903334893
- Conrad, D. F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y., et al. (2010). Origins and functional impact of copy number variation in the human genome. *Nature* 464, 704–712. doi: 10.1038/nature08516
- Cretu Stancu, M., van Roosmalen, M. J., Renkens, I., Nieboer, M. M., Middelkamp, S., de Ligt, J., et al. (2017). Mapping and phasing of structural variation in patient genomes using nanopore sequencing. *Nat. Commun.* 8, 1326. doi: 10.1038/s41467-017-01343-4
- De Gregori, M., Ciccone, R., Magini, P., Pramparo, T., Gimelli, S., Messa, J., 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. doi: 10.1136/jmg.2007.052787
- Dong, Z., Jiang, L., Yang, C., Hu, H., Wang, X., Chen, H., et al. (2014). A robust approach for blind detection of balanced chromosomal rearrangements with whole-genome low-coverage sequencing. *Hum. Mutat.* 35, 625–636. doi: 10.1002/humu.22541
- Ebler, J., Haukness, M., Pesout, T., Marschall, T., and Paten, B. (2019). Haplotype-aware genotyping from noisy long reads. *Genome Biol.* 3;20 (1), 116. doi: 10.1186/s13059-019-1709-0
- Fantes, J. A., Bolland, E., Ramsay, J., Donnai, D., Splitt, M., Goodship, J. A., et al. (2008). FISH mapping of *de novo* apparently balanced chromosome rearrangements identifies characteristics associated with phenotypic abnormality. *Am. J. Hum. Genet.* 82, 916–926. doi: 10.1016/j.ajhg.2008.02.007
- Feuk, L., Carson, A. R., and Scherer, S. W. (2006). Structural variation in the human genome. *Nat. Rev. Genet.* 7, 85–97. doi: 10.1038/nrg1767
- Fischer, J., Colls, P., Escudero, T., and Munne, S. (2010). Preimplantation genetic diagnosis (PGD) improves pregnancy outcome for translocation carriers with a history of recurrent losses. *Fertil. Steril.* 94, 283–289. doi: 10.1016/j.fertnstert.2009.02.060
- Hancks, D. C., and Kazazian, H. H. Jr. (2012). Active human retrotransposons: variation and disease. *Curr. Opin. Genet. Dev.* 22, 191–203. doi: 10.1016/j.cde.2012.02.006
- Howarth, K. D., Pole, J. C., Beavis, J. C., Batty, E. M., Newman, S., Bignell, G. R., et al. (2011). Large duplications at reciprocal translocation breakpoints that might be the counterpart of large deletions and could arise from stalled replication bubbles. *Genome Res.* 21, 525–534. doi: 141953/gr.114116.110
- Imaizumi, K., Kimura, J., Matsuo, M., Kurosawa, K., Masuno, M., Niikawa, N., et al. (2002). Sotos syndrome associated with a *de novo* balanced reciprocal translocation t(5;8)(q35;q24.1). *Am. J. Med. Genet.* 107, 58–60. doi: 10.1002/ajmg.10080
- Liang, H., Fan, L., De, C., Zhi, Z., Guo, Y., Jian, Z., et al. (2018). Localization of balanced chromosome translocation breakpoints by long-read sequencing on the Oxford Nanopore platform. *bioRxiv* 419531. doi: 141953/419531
- Magi, A., Pippucci, T., and Sidore, C. (2017). XCAVATOR: accurate detection and genotyping of copy number variants from second and third generation whole-genome sequencing experiments. *BMC Genomics* 18 (1), 747. doi: 10.1186/s12864-017-4137-0
- Mikelsaar, R., Nelis, M., Kurg, A., Zilina, O., Korrovits, P., Ratsep, R., et al. (2012). Balanced reciprocal translocation t(5;13)(q33;q12) and 9q31.1 microduplication in a man suffering from infertility and pollinosis. *J. Appl. Genet.* 53, 93–97. doi: 10.1007/s13353-011-0078-5
- Munne, S. (2005). Analysis of chromosome segregation during preimplantation genetic diagnosis in both male and female translocation heterozygotes. *Cytogenet. Genome Res.* 111, 305–309. doi: 10.1159/000086904
- Nattestad, M., Chin, C. S., and Schatz, M. C. (2016). Ribbon: Visualizing complex genome alignments and structural variation. *BioRxiv*. doi: 141953/082123
- Ogilvie, C. M., Braude, P., and Scriven, P. N. (2001). Successful pregnancy outcomes after preimplantation genetic diagnosis (PGD) for carriers of chromosome translocations. *Hum. Fertil.* 4 (3), 168–171. doi: 10.1080/1464727012000199252
- Pasquier, L., Fradin, M., Cherot, E., Martin-Coignard, D., Colin, E., Journal, H., et al. (2016). Karyotype is not dead (yet)! *Eur. J. Med. Genet.* 59, 11–15. doi: 10.1016/j.ejmg.2015.11.016
- Pennisi, E. (2012). Genome sequencing. Search for pore-fection. *Science* 336, 534–537. doi: 10.1126/science.336.6081.534
- Puig, M., Casillas, S., Villatoro, S., and Caceres, M. (2015). Human inversions and their functional consequences. *Brief Funct. Genomics* 14, 369–379. doi: 10.1093/bfpg/elv020
- Rizzolio, F., Bione, S., Sala, C., Goegan, M., Gentile, M., Gregato, G., et al. (2006). Chromosomal rearrangements in Xq and premature ovarian failure: mapping of 25 new cases and review of the literature. *Hum. Reprod.* 21, 1477–1483. doi: 10.1093/humrep/dei495
- Robinson, J. T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., et al. (2011). Integrative genomics viewer. *Nat. Biotechnol.* 29, 24–26. doi: 10.1038/nbt1754
- Sandberg, A. A., and Meloni-Ehrig, A. M. (2010). Cytogenetics and genetics of human cancer: methods and accomplishments. *Cancer Genet. Cytogenet.* 203, 102–126. doi: 10.1016/j.cancergen.2010.10.004
- Schluth-Bolard, C., Labalme, A., Cordier, M. P., Till, M., Nadeau, G., Tevissen, H., et al. (2013). 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.* 50, 144–150. doi: 10.1136/jmedgenet-2012-101351
- Scriven, P. N. (1998). Communicating chromosome rearrangements and their outcomes using simple computer-generated color ideograms. *Genet. Test.* 2, 71–74. doi: 10.1089/gte.1998.2.71
- Sedlazeck, F. J., Rescheneder, P., Smolka, M., Fang, H., Nattestad, M., von Haeseler, A., et al. (2018). Accurate detection of complex structural variations using single-molecule sequencing. *Nat. Methods* 15, 461–468. doi: 10.1038/s41592-018-0001-7
- Sen, S. K., Han, K., Wang, J., Lee, J., Wang, H., Callinan, P. A., et al. (2006). Human genomic deletions mediated by recombination between Alu elements. *Am. J. Hum. Genet.* 79, 41–53. doi: 10.1086/504600
- Stankiewicz, P., and Lupski, J. R. (2010). Structural variation in the human genome and its role in disease. *Annu. Rev. Med.* 61, 437–455. doi: 10.1146/annurev-med-100708-204735

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.01313/full#supplementary-material>

- Suzumori, N., and Sugiura-Ogasawara, M. (2010). Genetic factors as a cause of miscarriage. *Curr. Med. Chem.* 17, 3431–3437. doi: 10.2174/092986710793176302
- Szmulewicz, M. N., Novick, G. E., and Herrera, R. J. (1998). Effects of Alu insertions on gene function. *Electrophoresis* 19, 1260–1264. doi: 10.1002/elps.1150190806
- Tsiatis, A. C., Norris-Kirby, A., Rich, R. G., Hafez, M. J., Gocke, C. D., Eshleman, J. R., et al. (2010). Comparison of Sanger sequencing, pyrosequencing, and melting curve analysis for the detection of KRAS mutations: diagnostic and clinical implications. *J. Mol. Diagn.* 12, 425–432. doi: 10.2353/jmoldx.2010.090188
- Utami, K. H., Hillmer, A. M., Aksoy, I., Chew, E. G., Teo, A. S., Zhang, Z., et al. (2014). Detection of chromosomal breakpoints in patients with developmental delay and speech disorders. *PLoS One* 9, e90852. doi: 10.1371/journal.pone.0090852
- Van Heesch, S., Kloosterman, W. P., Lansu, N., Ruzius, F. P., Levandowsky, E., et al. (2013). Improving mammalian genome scaffolding using large insert mate-pair next-generation sequencing. *BMC Genomics* 14, 257. doi: 10.1186/1471-2164-14-257
- Vandeweyer, G., and Kooy, R. F. (2009). Balanced translocations in mental retardation. *Hum. Genet.* 126, 133–147. doi: 10.1007/s00439-009-0661-6
- Yao, F., Ariyaratne, P. N., Hillmer, A. M., Lee, W. H., Li, G., et al. (2012). Long span DNA paired-end-tag (DNA-PET) sequencing strategy for the interrogation of genomic structural mutations and fusion-point-guided reconstruction of amplicons. *PLoS One* 7, e46152. doi: 10.1371/journal.pone.0046152
- Zhang, S., Lei, C., Wu, J., Zhou, J., Sun, H., Fu, J., et al. (2017). The establishment and application of preimplantation genetic haplotyping in embryo diagnosis for reciprocal and Robertsonian translocation carriers. *BMC Med. Genomics* 10, 60. doi: 10.1186/s12920-017-0294-x

Conflict of Interest: LH, DC, YT and GL are employees of Reproductive and Genetic Hospital of CITIC–Xiangya. ZZ, FL, YW, QX, YL, DY, GY, JZ, and DW are employees of Grandomics Biosciences.

Copyright © 2020 Hu, Liang, Cheng, Zhang, Yu, Zha, Wang, Xia, Yuan, Tan, Wang, Liang and Lin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.