



## OPEN ACCESS

## EDITED BY

Youri I. Pavlov,  
University of Nebraska Medical Center,  
United States

## REVIEWED BY

Jinbo Zhu,  
University of Cambridge,  
United Kingdom  
Weiyang Li,  
Jining Medical University, China  
Jing Zhao,  
Beijing Genomics Institute (BGI), China

## \*CORRESPONDENCE

Xin-Rui Wang,  
wanxiru@sjtu.edu.cn

## SPECIALTY SECTION

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

RECEIVED 02 June 2022

ACCEPTED 11 October 2022

PUBLISHED 31 October 2022

## CITATION

Lei Y-Q, Xu L-P, Cao H and Wang X-R  
(2022), A method of large DNA fragment  
enrichment for nanopore sequencing in  
region 22q11.2.  
*Front. Genet.* 13:959883.  
doi: 10.3389/fgene.2022.959883

## COPYRIGHT

© 2022 Lei, Xu, Cao and Wang. 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.

# A method of large DNA fragment enrichment for nanopore sequencing in region 22q11.2

Yu-Qing Lei<sup>1,2,3</sup>, Liang-Pu Xu<sup>1,4</sup>, Hua Cao<sup>1,2,3</sup> and  
Xin-Rui Wang<sup>1,2,5\*</sup>

<sup>1</sup>Fujian Maternity and Child Health Hospital, College of Clinical Medicine for Obstetrics and Gynecology and Pediatrics, Fujian Medical University, Fuzhou, China, <sup>2</sup>NHC Key Laboratory of Technical Evaluation of Fertility Regulation for Non-human Primate (Fujian Maternity and Child Health Hospital), Fuzhou, China, <sup>3</sup>Department of Cardiac Surgery, Fujian Children's Hospital (Fujian Branch of Shanghai Children's Medical Center), College of Clinical Medicine for Obstetrics and Gynecology and Pediatrics, Fujian Medical University, Fuzhou, China, <sup>4</sup>Medical Genetic Diagnosis and Therapy Center, Fujian Maternity and Child Health Hospital, College of Clinical Medicine for Obstetrics and Gynecology and Pediatrics, Fujian Medical University, Fuzhou, China, <sup>5</sup>Medical Research Center, Fujian Maternity and Child Health Hospital, College of Clinical Medicine for Obstetrics and Gynecology and Pediatrics, Fujian Medical University, Fuzhou, China

**Background:** 22q11.2 deletion syndrome (22q11.2DS) is a disorder caused when a small part of chromosome 22 is missing. Diagnosis is currently established by the identification of a heterozygous deletion at chromosome 22q11.2 through chromosomal microarray analysis or other genomic analyses. However, more accurate identification of the breakpoint contributes to a clearer understanding of the 22q11.2 deletion syndrome.

**Methods:** In this study, we present a feasible nanopore sequencing method of 22q11.2 deletion. This DNA enrichment method—region-specific amplification (RSA)—is able to analyze the 22q11.2 deletion by specific amplification of an approximately 1-Mb region where the breakpoint might exist. RSA introduces universal primers into the target region DNA by a Y-shaped adaptor ligation and a single primer extension. The enriched products, completed by amplification with universal primers, are then processed by standard ONT ligation sequencing protocols.

**Results:** RSA is able to deliver adequate coverage (>98%) and comparable long reads (average length >1 Kb) throughout the 22q11.2 region. The long nanopore sequencing reads, derived from three umbilical cord blood samples, have facilitated the identification of the breakpoint of the 22q11.2 deletion, as well as by Sanger sequencing.

**Conclusion:** The Oxford Nanopore MinION sequencer can use RSA to sequence the target region 22q11.2; this method could also be used for other hard-to-sequence parts of the genome.

**Abbreviations:** RSA, region-specific amplification.

## KEYWORDS

region-specific amplification, nanopore sequencing, 22q11.2 deletion syndrome, genomic analyses, target region enrichment method

## Introduction

22q11.2 deletion syndrome (22q11.2DS) is a disorder caused when a small part of chromosome 22 is missing. Medical problems associated with 22q11.2 deletion syndrome commonly include heart defects, poor immune system function, a cleft palate, complications related to low levels of calcium in the blood, and delayed development with behavioral and emotional problems (McDonald-McGinn et al., 1999; McDonald-McGinn et al., 2015). The diagnosis of 22q11.2DS is established by the identification of a heterozygous deletion at chromosome 22q11.2 using chromosomal microarray analysis or other genomic analyses (McDonald-McGinn et al., 1999; McDonald-McGinn et al., 2015). None of these methods can accurately assess the breakpoint, thereby making them also unable to accurately assess the involved genes.

The Oxford Nanopore MinION sequencer, known as a “third-generation” sequencing platform, is capable of producing significantly longer read lengths and analyzing traditionally problematic sequence regions with high GC content (Mikheyev and Tin, 2014). The longer read lengths have been shown to be highly advantageous in identifying structural variants and haplotype phasing within complex genomic loci (Ritz et al., 2014; Wang et al., 2015).

Targeted sequencing enables researchers to enrich loci of interest, thereby reducing sequencing costs and labor, and facilitating high coverage data for genomic regions of interest. For next-generation sequencing, the most common strategy is amplification or hybridization capture (Kozarewa et al., 2015). However, none of these take advantage of new long-read sequencing technologies such as nanopore because long-fragment amplicons with relatively large bias and hybridization capture have yet to be fully optimized for long fragments. Long-fragment PCR and CATCH-seq3, in which regions of interest are excised by dual Cas9 cleavage and then enriched by size selection, have been described for target enrichment with nanopore sequencing (Gabrieli et al., 2018; Gilpatrick et al., 2020); however, these methods are not suitable for the enrichment of larger regions (>1 Mb).

Our DNA enrichment method, region-specific amplification (RSA), addresses this unmet need by semi-nested long DNA fragment amplification of 1–1.5 Kbp in length. The first step of RSA is adaptor ligation by the NGS library construction kit. The targeted genomic DNA segments are then extended by 5'-end thiophosphorylated primers. The

majority of the genomic DNA was cleared using the T7 exonuclease. To improve the specificity, the targeted genomic DNA segments are then extended by the inner primers, with the universal sequence at the 5'-end being extended for a second time. Thus, the DNA fragments of the target region are successfully connected with a universal sequence. After limited-step amplification using universal primers, the enriched products are then processed by standard ONT ligation sequencing protocols (Figure 1).

Here, we demonstrate the utility of RSA for the targeted sequencing of the region 22q11.2 by the Oxford Nanopore MinION sequencer, which may also be applied to other complex regions of the genome.

## Methods

### Sample source

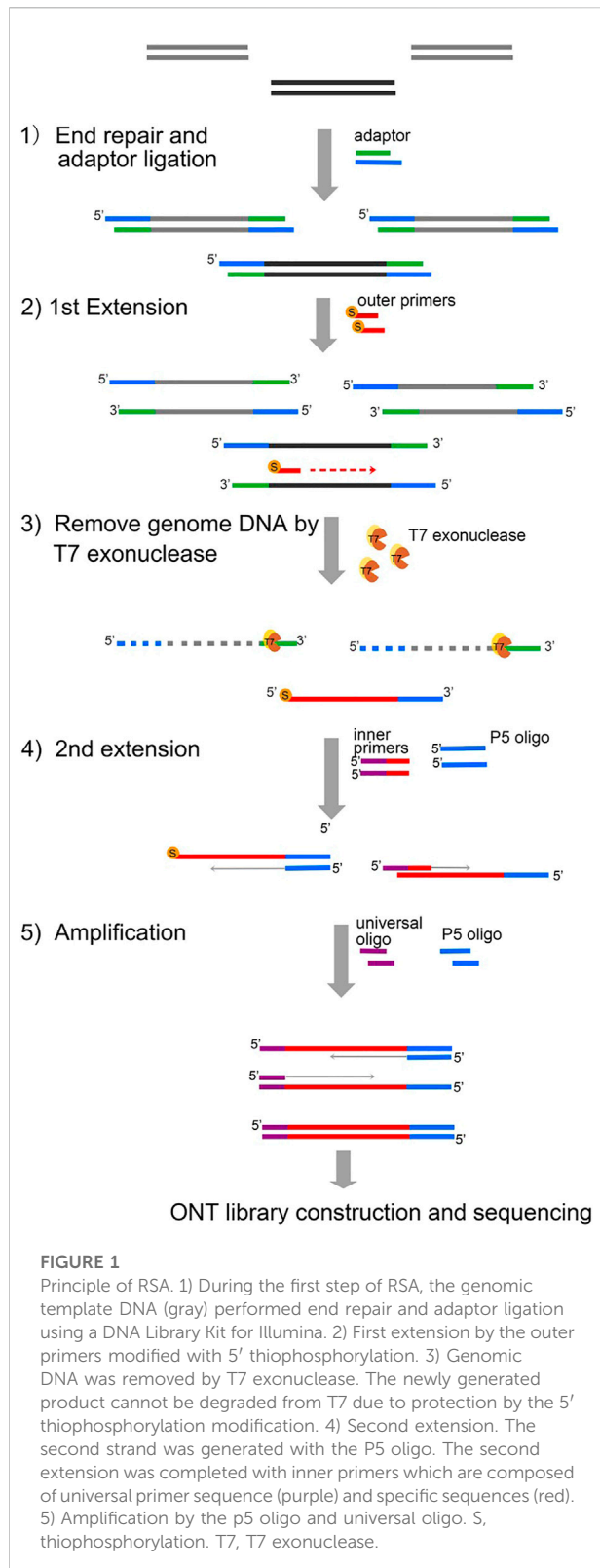
The samples used in this study were collected at Fujian Maternity and Child Health Hospital between January and December 2019. The cord blood cells were collected from patients with a gestational age of 24–35 weeks. The 22q11 microdeletion was confirmed by the Affymetrix CytoScan 750K Array with umbilical cord blood cells from pregnant women who had been diagnosed with cardiac malformation by routine prenatal ultrasound. The SNP-array results were performed by the pathology department. Written informed consent for the study using these samples was obtained from the patients and approved by the Ethics Committee of Fujian Maternal and Child Health Hospital.

### Genomic DNA preparation from umbilical cord blood cells

DNA extraction was performed with the TIAnmp Blood DNA Kit (TIANGEN, Cat. no. 4992207) according to the manufacturer's protocol, using a 200- $\mu$ L umbilical cord blood sample for each reaction.

### RSA primer design

The primer design software Primer3 (Untergasser et al., 2012) was applied to design the primers, which were spaced



1–1.5 Kbp apart. When they were settled, the homology between the selected primers and the rest of the genome was checked with BLAT (Kent, 2002). In this experiment,

819 primers were designed to target the entire 1 Mbp of the 22q11 at an average spacing of ≈1 Kbp. The first two bases of all the first extension primers (outer primers) were modified by thiophosphorylation. All the second extension primers (inner primers) had a common sequence at the 5'-end (see Additional File 1: Supplementary Table S1 for the list of primers and supporting information). The primers were synthesized by Sangon Biotech (Shanghai) and were then combined (in water) to an equimolar ratio of all primers.

## Region-specific amplification

### End repair and adaptor ligation

End repair and adaptor ligation were performed using the QuarPrep Ultra DNA Library Kit for Illumina (Dynasty gene, Cat. no. L1001A) using 200 ng genomic DNA. This step was modified from the manufacturer's protocol: the input DNA was not fragmented, and the final amplification step was skipped.

### Primer extension and genomic DNA removal

For each 20 μL annealing reaction, the ligation products were combined with 5 μM first extension primer mixture (outer primers), 1× NEB CutSmart Buffer, 0.4 mM dNTPs, and DNase-free water. The mixed annealing reaction was performed in a sterile DNase-free microfuge tube. The DNA/primer was denatured at 95°C for 5 min, briefly spined, and quickly placed on ice. After adding eight units of Bst 2.0 WarmStart DNA polymerase (NEB, Cat. no. M0538), the samples were incubated at 60°C for 10 min and 80°C for 20 min. Ten units of T7 exonuclease (NEB, Cat. no. M0263) were then added and incubated at 25°C for 30 min. The targeted DNA was purified by 1× AMPure XP beads.

### Semi-nested extension

For each 20 μL annealing reaction, the ligation products were combined with 5 μM second extension primer mixture (inner primers), 5 μM P5 oligo, 1× NEB CutSmart Buffer, 0.4 mM dNTPs, and DNase-free water. The mixed annealing reaction was performed in a sterile DNase-free microfuge tube. The DNA/primer was denatured at 95°C for 5 min, briefly spined, and quickly placed on ice. After adding eight units of Bst 2.0 WarmStart DNA polymerase (NEB, Cat. no. M0538), the samples were incubated at 60°C for 20 min and at 80°C for 20 min. The targeted DNA was purified by using 1× AMPure XP beads.

### Amplification

The primer extension products were amplified with the introduced sequences. For each 25 μL amplification reaction,

TABLE 1 Sequencing results.

	E3359	E3534	E3199
Reads	47,966	48,019	64,270
Reads (>Q10)	27,295(56.9%)	26,153(54.5%)	35,789(55.7%)
Total bases	55,341,275	57,602,066	76,462,240
Coverage rate in targeted region (depth>1)	98.3%	98.4%	98.5%
Coverage rate in targeted region (depth>5)	96.4%	95.7%	96.8%
Mapped reads to whole genome (%)	98.6%	98.7%	98.8%
Mapped reads to targeted region (%)	43.2%	43.7%	31.4%
Average depth in targeted region	51.4	53.5	50.5
Average length pg region	1154	1,200	1,190

the primer extension products were combined with 1× KAPA HiFi HotStart ReadyMix (Roche Sequencing Store, Cat. no. KK2602), 0.4 μM each of P5 oligo, and universal oligo. The thermocycling protocol was 10 min at 94°C for the initial denaturation, followed by 25 cycles of 30 s at 94°C, 30 s at 60°C, 5 min at 72°C, and a final extension for 5 min at 72°C. The amplification products were purified using 1× AMPure XP beads.

## ONT library construction and sequencing

The ONT library was constructed according to the PCR Barcoding Kit protocol (Oxford Nanopore, Cat. no. SQK-PBK004) using 200 ng amplification products. Sequencing was performed using a R9.4 flow cell on a MinION device (Oxford Nanopore) with the MinKNOW.

## Sanger sequencing

To validate the variant calls, genomic DNA was applied as a template. PCR primers for each variant were designed manually *via* the primer-design software Primer3. The full list of PCR primers is provided in Additional File 2: [Supplementary Table S2](#). To set up each PCR reaction, 50 ng of gDNA DNA was combined with 1×KAPA HiFi HotStart ReadyMix (Roche Sequencing Store, Cat. no. KK2602), 0.4 μM each of forward and reverse primers, and water. The thermocycling protocol is 10 min at 94°C for the initial denaturation, followed by 35 cycles of 30 s at 94°C, 30 s at 55°C, 60 s at 72°C, and a final extension for 5 min at 72°C. The amplification was validated by 1% agarose gel electrophoresis. Each amplicon was then purified with the QIAquick Gel Extraction Kit (Qiagen, Cat. no. 28704). Sanger sequencing was performed at Sangon Biotech (Shanghai) using the same primers for PCR amplification.

## Sequence alignment and variant detection

Sequencing data were generated by a R9.4 flow cell on a MinION device (Oxford Nanopore) with MinKNOW. Basecalling was performed with Guppy (v.4.2.2). The reads were trimmed using Porechop (<https://github.com/rrwick/Porechop>) and were then aligned to the reference genome (Hg19) using Minimap2 (Li, 2018) with parameters "-a -x map-ont --MD." After alignment, structure variant calling was fulfilled using Sniffles (Sedlazeck et al., 2018) (v1.0.12, <https://github.com/fritzsedlazeck/Sniffles>) with parameters "--max\_distance 4000000 --minmapping\_qual 10."

## Results

### Statistics for sequence data

Raw data (fastq) files have been made publicly available and are accessible through the NCBI SRA website (SRA accession: SRP326421). In all three samples, more than 30% of the reads are mapped in the target region, ensuring that sufficient valid data could be generated. A majority of the targeted region maintains high coverage depth, and the coverage rate of 5× or above is more than 95%. The average read length is more than 1 kB, and longer reads are presented—beneficial for detecting SV variation (Table 1; Figure 2D).

### Distribution of reads

Most of the primer binding regions have a high depth, indicating that the primers hybridized to a specific region during annealing, and extension then occurred as expected (Figure 2B). Statistical analysis was conducted on the relative depth of the relative primer positions; it was found that the

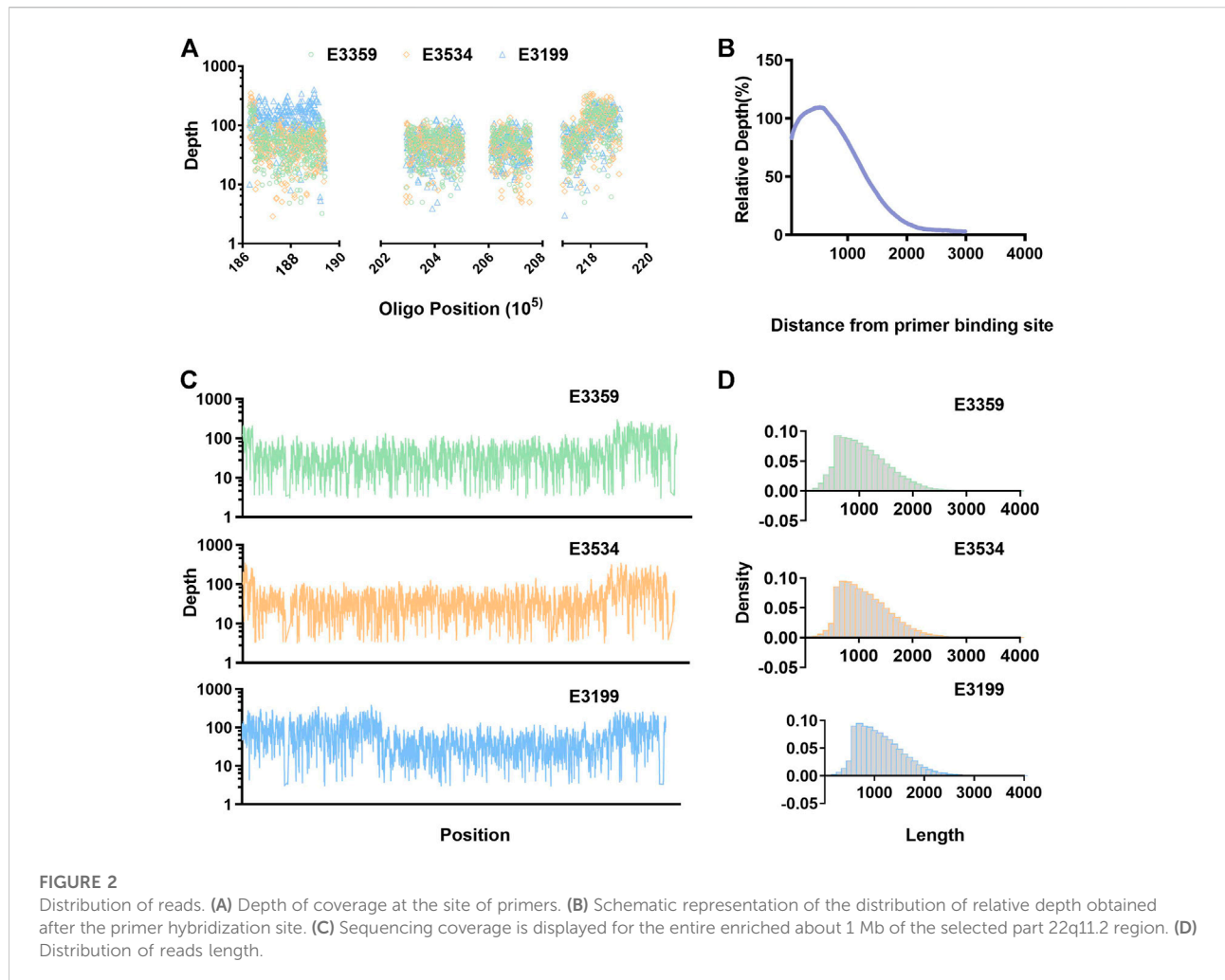


TABLE 2 Results of structural variation analysis.

Sample	ONT sequencing	SNP array
E3359	22:18650021-21769660del	arr[hg19] 22q11.21(18,648,855–21,800,471)x1
E3534	22:18653575-21767736del	arr[hg19] 22q11.21(18,648,855–21,800,471)x1
E3199	22:18919574-21788492del	arr[hg19] 22q11.21(18,916,842–21,800,741)x1

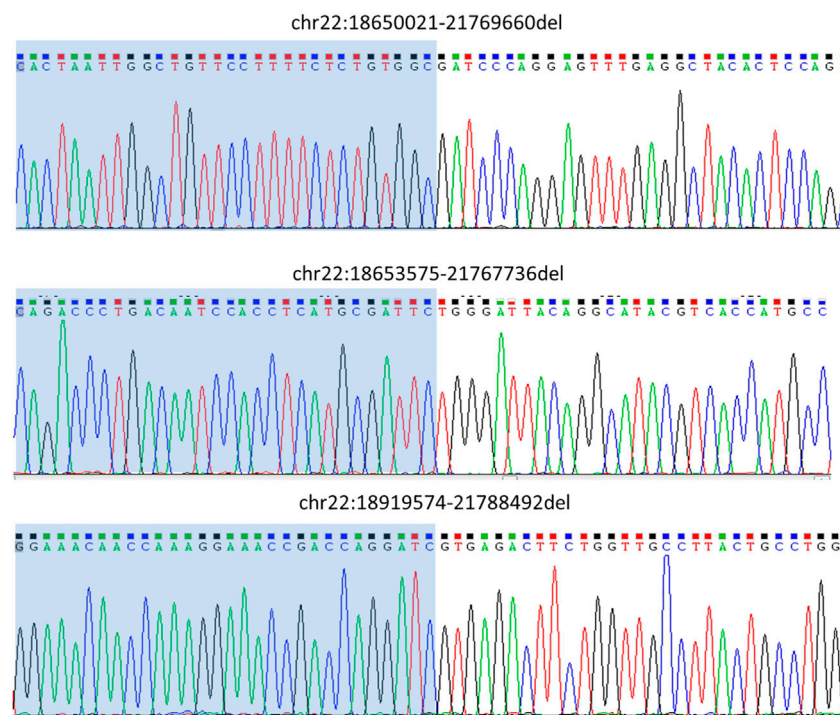
distribution was basically semi-normal. The relative depth decreased further away from the primer hybridization site. Interestingly, we found that the relative depth was highest at about 400 bp from the primer binding site, rather than at the nearest location (Figure 2A). This may be due to short DNA molecules having been filtered in the experiment and part of the amplified product being degraded or broken at the ends.

The average length of reads was more than 1 kb, so we chose primer spacing of 1–1.5 Kb to achieve high coverage of the target region. This was confirmed by the coverage and depth analysis (Figure 2C, Supplementary Table S3), although this may not be the best spacing.

### Variant analysis and Sanger validation

The sequencing results were then analyzed using Sniffles to detect structural variations. We got breakpoints for all three samples. The results were consistent with the SNP-array results. We then designed primers (Table 2) at both ends of the breakpoints to perform PCR and Sanger sequencing. The Sanger sequencing results supported the analysis and Sniffles analysis results (Figure 3, Supplementary Figure S1–S3, Supplementary Table S4). This indicates that the RSA can effectively enrich the structural variation areas.



**FIGURE 3**

Variant validation by Sanger sequencing. Primers that were designed at both ends of the breakpoints were to perform PCR and Sanger sequencing. The Sanger sequencing results supported the Sniffles analysis results.

## Discussion

Current target-region enrichment methods are mostly based on DNA hybridization to capture an oligonucleotide (Gnirke et al., 2009) or multiplexed PCR amplification reaction (Tewhey et al., 2009). These methods, which were previously applied in NGS, can only produce relatively short (less than 1,000 base pair) sequencing templates. As a result, they cannot complete the sequencing of the large-sized fragments in one read and then comprehensively characterize complex genomic loci, such as extended repeats, or resolve sections of unexpected sequence that have been rearranged within the targeted region.

Although the cost of nanopore sequencing is higher than that of NGS, it can produce significantly longer read lengths and walk through traditionally difficult sequence regions which cannot be effectively resolved by NGS. An effective enrichment method suitable for the nanopore sequencing platform would greatly extend the platform's application range. Johannes et al. (Dapprich et al., 2016) supplied a target region enrichment method that can successfully obtain a large-length template; however, this method is very time-consuming and the yield of products is so low that it cannot be applied to the third-generation sequencing

platform. Here, we present a method of target-region enrichment in which the adaptor ligation products are extended with two round nested-primers and then PCR with the newly introduced universal primers. This target-region enrichment method can produce large-sized products. Our experimental data reveal that the average length of data was more than 1 KB, with some read lengths reaching 7–8K. This is of great advantage for the sequencing of complex regions and structural variation searching. As the last step of this method is PCR, it is possible to obtain enough DNA to meet the needs of ONT ligation sequencing.

The spacing parameter in this study was not optimized in primer design but only set at a value of 1 KB, based on subjective judgment, so that the final number of primers would be relatively large. In fact, this parameter may be worth optimizing to help control the synthesis cost of primers and interference between them. Antholigo, which was adopted as the primer design method by Johannes et al. (Dapprich et al., 2016), was actually what we wanted to use at the beginning. However, due to its lowest spacing of 5 KB, we were concerned that the amplification efficiency of such a large segment of DNA would be low, so we did not use this software. However, it might be possible to use the software directly if amplification was performed using polymerases that

specialize in long DNA-fragment PCR. Finally, according to the design idea of Antholigo, we designed one primer per 1 KB with Primer3, and then evaluated the dimer between primers. If there was a serious dimer, we eliminated and redesigned it.

The RSA method provided in this paper could not only be applied to the 22q11 region involved in this study but also extended to enrichment in other larger regions. The RSA method can capture long DNA segments from genomic DNA samples, which may include entire adjacent target regions—including exon, intron, and intergenic regions. It is thus a feasible nanopore sequencing method for those complex regions, such as large segment insertions, deletions, ectopics, and other structural variations.

## Conclusion

We have described a feasible nanopore sequencing method for the 22q11.2 deletion, which may be extended to the enrichment in other larger regions. RSA was able to deliver adequate coverage (>98%) and adequate long reads (average length >1 Kb) throughout the 22q11.2 region. The long ONT sequencing reads, derived from three umbilical cord blood samples, have facilitated the breakpoint finding of the 22q11.2 deletion, which was validated by Sanger sequencing.

## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/[Supplementary Material](#).

## Ethics statement

The studies involving human participants were reviewed and approved by the Ethics Committee of Fujian Maternity and Child Health Hospital. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

## References

- Dapprich, J., Ferriola, D., Mackiewicz, K., Clark, P. M., Rappaport, E., D'Arcy, M., et al. (2016). The next generation of target capture technologies - large DNA fragment enrichment and sequencing determines regional genomic variation of high complexity. *BMC Genomics* 17, 486. doi:10.1186/s12864-016-2836-6
- Gabrieli, T., Sharim, H., Fridman, D., Arbib, N., Michaeli, Y., and Ebenstein, Y. (2018). Selective nanopore sequencing of human BRCA1 by Cas9-assisted targeting of chromosome segments (CATCH). *Nucleic Acids Res.* 46 (14), e87. doi:10.1093/nar/gky411
- Gilpatrick, T., Lee, I., Graham, J. E., Raimondeau, E., Bowen, R., Heron, A., et al. (2020). Targeted nanopore sequencing with Cas9-guided adapter ligation. *Nat. Biotechnol.* 38 (4), 433–438. doi:10.1038/s41587-020-0407-5

## Author contributions

XW conceptualized, designed, analyzed data, and supervised the study. YL designed the study, and drafted and edited manuscript. LX and HC recruited and collected samples. All authors reviewed the manuscript and agreed the manuscript to be submitted.

## Funding

This study was supported by the grants from the Natural Science Foundation of Fujian Province, China (No. 2020J01344); the Key Project on the Integration of Industry, Education and Research Collaborative Innovation of Fujian Province (No. 2021YZ034011); and the Key Project on Science and Technology Program of Fujian Health Commission (No. 2021ZD01002). National Natural Science Foundation of China (grant number 82000166).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2022.959883/full#supplementary-material>

- Gnirke, A., Melnikov, A., Maguire, J., Rogov, P., LeProust, E. M., Brockman, W., et al. (2009). Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat. Biotechnol.* 27 (2), 182–189. doi:10.1038/nbt.1523
- Kent, W. J. (2002). BLAT--the BLAST-like alignment tool. *Genome Res.* 12 (4), 656–664. doi:10.1101/gr.229202
- Kozarewa, I., Armisen, J., Gardner, A. F., Slatko, B. E., and Hendrickson, C. L. (2015). Overview of target enrichment strategies. *Curr. Protoc. Mol. Biol.* 112, 1–7. doi:10.1002/0471142727.mb0721s112
- Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* 34 (18), 3094–3100. doi:10.1093/bioinformatics/bty191

- McDonald-McGinn, D. M., Hain, H. S., Emanuel, B. S., and Zackai, E. H. (1999). "22q11.2 deletion syndrome," in *GeneReviews® [internet]*. M. P. Adam, H. H. Ardinger, R. A. Pagon, S. E. Wallace, L. J. H. Bean, G. Mirzaa, et al. (Seattle (WA): University of Washington, Seattle), 1993–2022.
- McDonald-McGinn, D. M., Sullivan, K. E., Marino, B., Philip, N., Swillen, A., Vorstman, J. A., et al. (2015). 22q11.2 deletion syndrome. *Nat. Rev. Dis. Prim.* 1, 15071. doi:10.1038/nrdp.2015.71
- Mikheyev, A. S., and Tin, M. M. (2014). A first look at the Oxford Nanopore MinION sequencer. *Mol. Ecol. Resour.* 14 (6), 1097–1102. doi:10.1111/1755-0998.12324
- Ritz, A., Bashir, A., Sindi, S., Hsu, D., Hajirasouliha, I., and Raphael, B. J. (2014). Characterization of structural variants with single molecule and hybrid sequencing approaches. *Bioinformatics* 30 (24), 3458–3466. doi:10.1093/bioinformatics/btu714
- 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 (6), 461–468. doi:10.1038/s41592-018-0001-7
- Tewhey, R., Warner, J. B., Nakano, M., Libby, B., Medkova, M., David, P. H., et al. (2009). Microdroplet-based PCR enrichment for large-scale targeted sequencing. *Nat. Biotechnol.* 27 (11), 1025–1031. doi:10.1038/nbt.1583
- Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., et al. (2012). Primer3--new capabilities and interfaces. *Nucleic Acids Res.* 40 (15), e115. doi:10.1093/nar/gks596
- Wang, M., Beck, C. R., English, A. C., Meng, Q., Buhay, C., Han, Y., et al. (2015). PacBio-LITS: A large-insert targeted sequencing method for characterization of human disease-associated chromosomal structural variations. *BMC Genomics* 16 (1), 214. doi:10.1186/s12864-015-1370-2