Hereditas

RESEARCH

Open Access

CrossMark

Long-read sequencing identified a causal structural variant in an exome-negative case and enabled preimplantation genetic diagnosis

Hefan Miao[1,2†], Jiapeng Zhou[3†], Qi Yang[3], Fan Liang[3], Depeng Wang[3], Na Ma[1,2], Bodi Gao[1,2], Juan Du[1,2], Ge Lin[1,2], Kai Wang[4*] and Qianjun Zhang[1,2*]

Abstract

Background: For a proportion of individuals judged clinically to have a recessive Mendelian disease, only one heterozygous pathogenic variant can be found from clinical whole exome sequencing (WES), posing a challenge to genetic diagnosis and genetic counseling. One possible reason is the limited ability to detect disease causal structural variants (SVs) from short reads sequencing technologies. Long reads sequencing can produce longer reads (typically 1000 bp or longer), therefore offering greatly improved ability to detect SVs that may be missed by short-read sequencing.

Results: Here we describe a case study, where WES identified only one heterozygous pathogenic variant for an individual suspected to have glycogen storage disease type Ia (GSD-Ia), which is an autosomal recessive disease caused by bi-allelic mutations in the *G6PC* gene. Through Nanopore long-read whole-genome sequencing, we identified a 7.1 kb deletion covering two exons on the other allele, suggesting that complex structural variants (SVs) may explain a fraction of cases when the second pathogenic allele is missing from WES on recessive diseases. Both breakpoints of the deletion are within Alu elements, and we designed Sanger sequencing and quantitative PCR assays based on the breakpoints for preimplantation genetic diagnosis (PGD) for the family planning on another child. Four embryos were obtained after in vitro fertilization (IVF), and an embryo without deletion in *G6PC* was transplanted after PGD and was confirmed by prenatal diagnosis, postnatal diagnosis, and subsequent lack of disease symptoms after birth.

Conclusions: In summary, we present one of the first examples of using long-read sequencing to identify causal yet complex SVs in exome-negative patients, which subsequently enabled successful personalized PGD.

Keywords: Whole-exome sequencing, WES, Structural variants, Long-read sequencing, G6PC, PGD, GSD-Ia

Background

Whole exome sequencing (WES) is now widely used in genetic testing on patients who are suspected or have been clinically demonstrated to have genetic disorders. However, a large proportion (~ 60–70%) of patients judged clinically to have a Mendelian disease receive negative results on WES with current Illumina short-read sequencing technology [1–5]. Compared to WES, the use of whole-genome sequencing (WGS) does not appear to significantly improve the diagnostic yield [6, 7] or present much economic advantage [8]. Therefore, WES/WGS-negative cases pose a significant challenge to the clinical diagnosis of genetic diseases. Several reasons may explain the lack of positive findings, such as the inefficiency of template DNA capture, the biases in sequencing coverage, the failure to call causal variants from data, the inability to catalog all functional variants (especially non-coding variants), the incorrect clinical interpretation of genetic

* Correspondence: wangk@email.chop.edu; zhangqianjun@csu.edu.cn
†Hefan Miao and Jiapeng Zhou contributed equally to this work.
[4]Raymond G. Perelman Center for Cellular and Molecular Therapeutics, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA
[1]Institute of Reproductive and Stem Cell Engineering, Central South University, Changsha 410078, Hunan, China
Full list of author information is available at the end of the article

variants, the possibility of complicated oligogenic disease in certain patients and the possibility of disease causal mechanisms due to somatic or epigenetic origin. Among these reasons, the limited ability to interrogate repeat elements such as tandem repeats [9] and structural variants (SVs) [10] may play important roles. Clinical microarrays such as SNP arrays or array-CGH can detect relatively large deletions and duplications [11–14], but have difficulty to reliably identify small (< 10 kb) exonic deletions or complex SVs in clinical settings [15]. Furthermore, several published studies demonstrated examples where disease causal SVs were missed by short-read WES/WGS [16], or that certain classes of disease causal repeats failed to be identified by WES/WGS [17]. In particular, conventional short-reads sequencing approaches have been reported to lack sensitivity, exhibit very high false positive rate, and misinterpret complex or nested SVs [18].

Long-read sequencing technologies, such as the 10X Genomics linked-read sequencing, the Oxford Nanopore Technologies (ONT) and PacBio single molecule real-time (SMRT) sequencing, offer complementary strengths to traditional WES/WGS based on short-read sequencing. Long-read sequencing can produce read length (typically 1000 bp or longer) that is far higher than the 100-150 bp produced by short-read sequencing, therefore allowing for the resolution of breakpoints of complex SVs [19] or the detection of long tandem repeats [20, 21]. In particular, recent de novo human genome assemblies via long-read sequencing have revealed tens of thousands of SVs per genome, several times more than previously observed via WGS, suggesting an underestimation of the extent and complexity of SVs [22–25].

In the present study, we applied a long-read whole-genome sequencing to yield the genetic diagnosis of glycogen storage disease type Ia (GSD-Ia) in a patient whose causal variants were unsolved by Sanger sequencing and WES. Glycogen storage disease type I (GSD-I) is a group of autosomal recessive metabolic diseases caused by defects in the glucose-6-phosphatase (G6Pase) complex, with an overall incidence of approximately 1:20,000–40,000 cases per live birth [26, 27]. The most common form, GSD-Ia, represents more than 80% of GSD-I cases [28]. Mutations in the *G6PC* gene have been found to be the cause of this disease. The *G6PC* comprises of five exons on chromosome 17q21, and encodes a 35 kD monomeric protein, the G6Pase catalytic subunit, which plays a role in the endoplasmic reticulum [29, 30]. Through Sanger sequencing and WES, we were able to identify one deleterious mutation in the proband, yet we suspected the presence of a complex SV due to the observation of Mendelian inconsistency in the family. As a result of long-read sequencing, we made a positive diagnosis of GSD-Ia on the patient and accurately identified the breakpoints of a causal SV in the other allele of the G6PC gene, which further guided

genetic counseling in the family and enabled a successful preimplantation genetic diagnosis (PGD) for in vitro fertilization (IVF) on the family.

Methods
Patient characteristics
The study was approved by the CITIC-Xiangya Reproductive and Genetics Hospital, Central South University. The collection and use of tissues followed procedures that are in accordance with ethical standards as formulated in the Helsinki Declaration, and informed consent was obtained from the study participants. The proband was a 12 year old boy with hepatosplenomegaly and growth retardation, and was diagnosed with possible GSD-Ia in Xiangya Hospital of Central South University in 2017 (Table 1). The parents of the proband came to Xiangya Hospital to seek help to have another child via IVF, but genetic testing by Sanger sequencing and whole-exome sequencing (WES) failed to identify a definitive genetic cause of the disease in the family. Indeed, a homozygous mutation of c.326G > A (p.C109Y) in exon 2 of the G6PC gene was identified via WES and was interpreted to be likely pathogenic. However, Sanger sequencing showed that the mother is a carrier, yet the father does not carry the mutation, therefore complicating genetic counseling on the family and subsequent design of PGD on the proposed IVF.

Long-read sequencing by Oxford Nanopore technology
Due to the presence of Mendelian inconsistency on the c.326G > A mutation, and that sperm sequencing on the father was inconclusive, we suspected that a SV may have encompassed the exon but with breakpoints in non-exonic regions, evading detection by whole-exome sequencing. Due to the complex genomic architecture around this gene, we decided to sequence the patient by low-coverage long-read sequencing on the Oxford

Table 1 Biochemical indicators of the proband suggest a probable diagnosis of GSD-Ia

Parameter	Tested value	Reference range	Result
Glucose (mmol/L)	0.17↓	3.6–6.1	Abnormal
CO2 (mmol/L)	7.1↓	19–33	Abnormal
Sodium (mmol/L)	132.7	135–153	
Chlorinum (mmol/L)	89.2↓	96–108	Abnormal
Calcium (mmol/L)	2.65	2–2.6	
Total bile acid (μmol/L)	16.7↑	0–12	Abnormal
ALT (IU/L)	257.4↑	7–56	Abnormal
AST (IU/L)	357.6↑	0–40	Abnormal
Triglyceride (mmol/L)	5.74↑	0.52–1.56	Abnormal
HDL (mmol/L)	2.01↑	0.88–1.76	Abnormal

Abbreviations: *ALT* alanine aminotransferase, *AST* aspartate aminotransferase, *HDL* high-density lipoproteins

Nanopore sequencing platform. Genomic DNA was extracted and large insert-size libraries were created according to the manufacturer recommended protocols (Oxford Nanopore, UK). Five μg genomic DNA was sheared to ~ 5-25 kb fragments using Megaruptor (Diagenode, B06010002), size selected (10-30 kb) with a Blue-Pippin (Sage Science, MA) to ensure removal of 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 according to protocol recommendations was performed using the Ultra II End Prep module (NEB, E7546L). At last, the purified dA tailed sample, blunt/TA ligase master mix (#M0367, NEB), tethered 1D adapter mix using SQK-LSK108 were incubated and purified. Libraries were sequenced on R9.4 flowcells using GridION X5. Four GridION flowcells generated 2,251,269 base-called reads containing 35,595,548,336 bases with an average read length of 16,579 bp. We used NGMLR [31] to align the long reads to the human reference genome (GRCh37). Structural variations (SVs) were called by Sniffles [31] and single nucleotide variants (SNVs) were called by SAMtools [32] for comparison to WES results. Ribbon [33] and IGV [34] were used to visualize the alignment results and manually validate possible SV calls.

Sanger sequencing and quantitative real-time PCR

Sanger sequencing and quantitative real-time PCR (RT-PCR) were conducted to validate the long-read sequencing result in the family. Genomic DNA was extracted from peripheral leukocytes using QIAamp® DNA Blood Mini Kit (Qiagen, Germany). The exons of the G6PC gene as well as the potential breakpoint junction were amplified using PCR primers (Table 2). Each PCR reaction was performed in a total volume of 40 μl containing 20 μl of GoTaq® 2 × Green Master Mix, 0.8 μl of 10 mM/l mixture of forward and reverse primers, 1.5 μl of 80 ng/μl genomic DNA template, and 17.6 μl of nuclease-free water. The PCR was conducted under the following cycling conditions: 95 °C for 5 min, 35 cycles of denaturation at 95 °C for 30 s, annealing at 58 °C for 30 s, and elongation at 72 °C for 30 s followed by a final elongation of 5 mins, and PCR products were sequenced using Sanger sequencing. The exons of the G6PC gene were also quantitatively tested using quantitative RT-PCR (Table 2). Data were analyzed using the △△CT method.

We further designed an easy assay to detect the deletion reliably by PCR, and performed nucleic acid gel electrophoresis and used the ß-Globin gene as an internal control to validate this assay. The amplification primers for the ß-globin gene are: Primer ß-F: 5'-TGAG TCTATGGGACGCTTGA-3′ and Primer ß-R: 5'-ATCC

Table 2 Primers for Sanger sequencing. No.1–10: Exon primers; No.11–20: Quantitative real-time primers; No.21–26: Breakpoint primers

No.	Primer	Position	Sequence (5′-3′)
1	G6PC-1F	Exon1	CACCACCAAGCCTGGAATAAC
2	G6PC-1R		CAGACATTGCGAGAGCGAATG
3	G6PC-2F	Exon2	GCATTCATTCAGTAACCC
4	G6PC-2R		AGACAGAAGCTGAGTGGA
5	G6PC-3F	Exon3	CACCTTTACTCCATTCTCTTTC
6	G6PC-3R		GTGCCACAACTCTTAATCAGCG
7	G6PC-4F	Exon4	CACTGAGAGCACCTAAGTTTGC
8	G6PC-4R		CTGATTACACACAGGATGTGG
9	G6PC-5F	Exon5	CATGTCACCCACTCCTCCAAAC
10	G6PC-5R		GTCACTTGCTCCAAATACCAGTG
11	G6PC-1ForF	5′-Flanking introns	TTTCACAGTCCTCCGTGACC
12	G6PC-1ForR		AGGGCTTCTATATCTTGAGCTTTC
13	G6PC-1QF	Exon1	TCCAGTCAACACATTACCTCCA
14	G6PC-1QR		TAAAGACGAGGTTGAGCCAGTC
15	G6PC-2inF	Intron2–3	AAGTTGGGACAAGGGAATCAGA
16	G6PC-2inR		CATTCTTAATTCCTCTACCCTGAGA
17	G6PC-4QF	Exon4	GCTGAAGGATCTGCACCTGT
18	G6PC-4QR		AGGGAGTCAGATCAGCCCAT
19	G6PC-5QF	Exon5	CAGCTTCGCCATCGGATTTT
20	G6PC-5QR		ACAATAGAGCTGAGGCGGAA
21	G6PC-D1F		GTGGGGAAAATGCCTGAGGA
22	G6PC-D2F		TTTTCACCCTTGGGAGCCTG
23	G6PC-D3F		GGTCACCCTGTCCCACTAGA
24	G6PC-D4F		CTCACCTGTTTTCCCACGGA
25	G6PC-D5F		GGGAGGAGACTCCAGGTCAT
26	G6PC-comR	Intron2–3	CTTTCCAGTCTGTGCCTCCAT

AGCCTTATCCCAACC-3′. The primers designed to amplify the sequence near the breakpoint were G6PC-DEL-F: 5'-GAGTTAGAAGGAGATGGCGGG-3′ and G6 PC-DEL-R: 5'-GGCCTATCCTACATATTAATAGTT-3′ which generates a target fragment of 418 bp.

Assessment of pathogenicity of the mutations

We analyzed the WES data through the variant filtering pipelines implemented in the ANNOVAR software [35], by focusing on coding variants and by removing common variants observed in the gnomAD database. We identified a novel and homozygous missense mutation (c.326G > A; p.C109Y) in the G6PC gene. The novel missense mutation has not been recorded in the HGMD version 2017.4 (http://www.hgmd.cf.ac.uk/ac/index.php) or in the dbSNP (https://www.ncbi.nlm.nih.gov/snp/) database. According to the ACMG-AMP 2015 Standards and Guidelines [36], and facilitated by the InterVar

software tool [37], we analyzed the pathogenicity of the novel mutation and determined that it is a likely pathogenic mutation responsible for the disease manifestation. No other candidate genes were found in our WES analysis that may explain the observed phenotypes of the proband.

## In vitro fertilization (IVF) and pre-implementation genetic diagnosis (PGD)

The parents of the proband had genetic counseling at the Reproductive & Genetic Hospital of CITIC-Xiangya, and proceeded with in vitro fertilization (IVF). A total of four oocytes were retrieved; all were in metaphase II, and all of them were inseminated (day 0) by intracytoplasmic sperm injection. The embryos were cultured to blastocyst stage, and 2–3 zona pellucida cells were used to detect mutations in the G6PC gene, and the embryos were also scored according to the Istanbul consensus [38] (Additional file 1: Table S1). PGD was performed using assays designed for detecting both the missense mutation and the exonic deletion in G6PC. The best embryo scored by the Istanbul consensus [38] was selected for implementation after PGD.

Results

Clinical examination

We were presented with a 12 year-old boy with hepatosplenomegaly and growth retardation at the Xiangya Hospital of Central South University, Hunan, China in 2017 (Fig. 1a). The clinical features include a rounded doll's face, fatty cheeks and protuberant abdomen (Fig. 1b). Based on examination on the proband's skeletal development by X-rays, his hip and wrist showed osteoporosis (Fig. 1c, d). In the upper liver, right intercostal midline sixth intercostal, liver rib length is 51 mm, thickness is 37 mm. Maximum oblique diameter of right liver is 159 mm, suggesting severe liver enlargement (Fig. 1d). Spleen is swollen to a thickness of 34 mm.

Additional biochemical assays were performed on the proband (Table 1). His fasting blood glucose value was 0.17 mmol/l, which had reached a dangerously low value. Levels of cholesterol, triglycerides and chlorinum were abnormal. Investigation of liver function showed elevated aminotransferases (AST, ALT) and other biochemical abnormalities. Serum copper and ceruloplasmin were at normal levels. Altogether, these clinical and biochemical
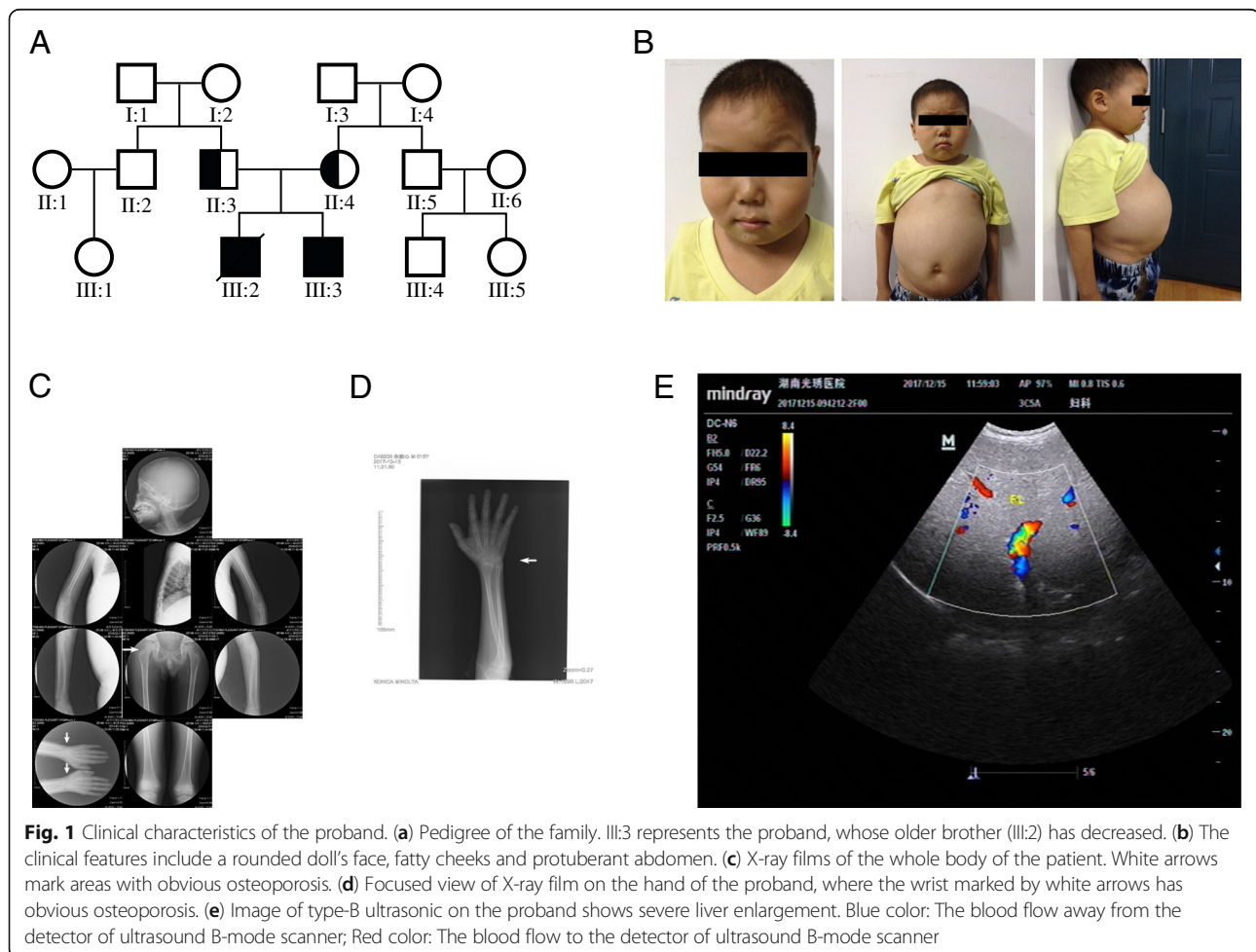
Fig. 1 Clinical characteristics of the proband. (**a**) Pedigree of the family. III:3 represents the proband, whose older brother (III:2) has decreased. (**b**) The clinical features include a rounded doll's face, fatty cheeks and protuberant abdomen. (**c**) X-ray films of the whole body of the patient. White arrows mark areas with obvious osteoporosis. (**d**) Focused view of X-ray film on the hand of the proband, where the wrist marked by white arrows has obvious osteoporosis. (**e**) Image of type-B ultrasonic on the proband shows severe liver enlargement. Blue color: The blood flow away from the detector of ultrasound B-mode scanner; Red color: The blood flow to the detector of ultrasound B-mode scanner

examinations indicated that the boy is likely to be affected with GSD-Ia.

The parents were non-consanguineous, and neither has any symptom of GSD-Ia (Fig. 1a). They came to Xiangya Hospital to seek help to obtain a genetic diagnosis and plan to have another child by in vitro fertilization (IVF). Given that GSD-Ia is a recessive disease, we hypothesized that the proband carries bi-allelic *G6PC* mutations inherited from the father and mother, respectively. Yielding a confirmed genetic diagnosis and determining the exact disease causal variants are necessary to perform preimplantation genetic diagnosis (PGD) from in vitro fertilization (IVF).
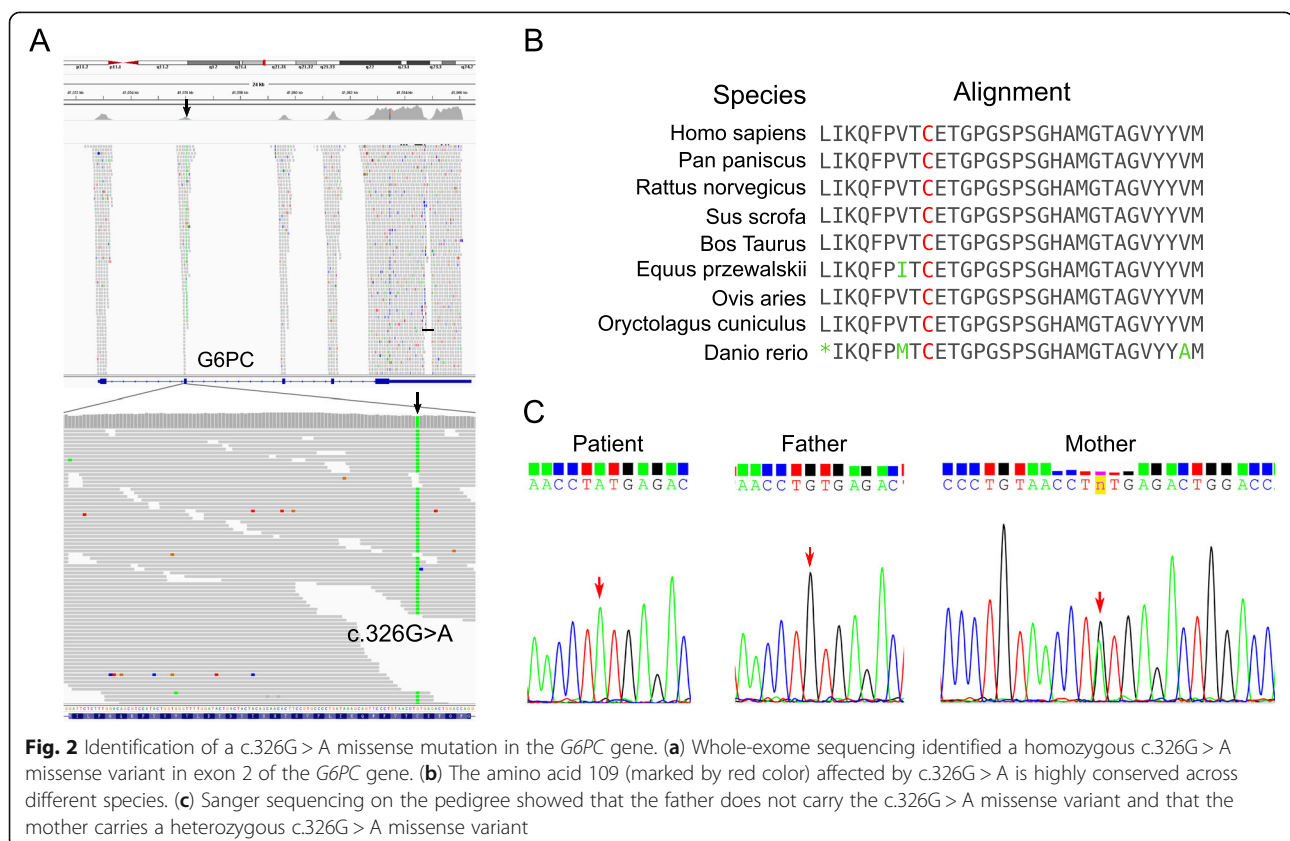
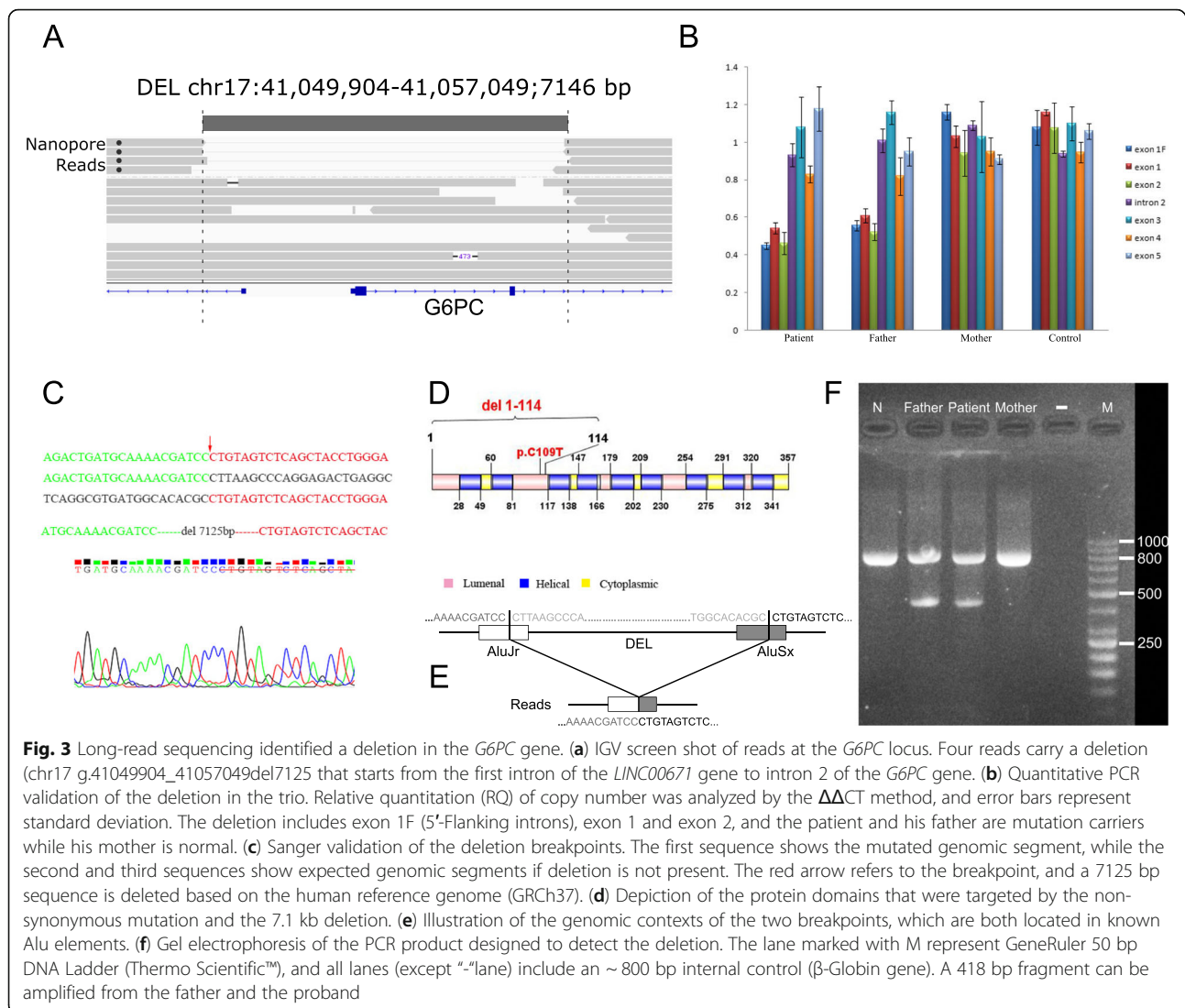### Whole-exome and sanger sequencing identified one pathogenic variant

To confirm the clinical diagnosis, we conducted clinical exome sequencing on the proband. A novel missense mutation of the *G6PC* gene (c.326G > A) was identified by WES (Fig. 2a), which appears to be homozygous and affects a highly conserved position in the protein sequence (Fig. 2b). Bioinformatics analysis by InterVar [37] and manual examination of the ACMG-AMP 2015 guidelines [36] determined the mutation to be likely pathogenic. We further validated the mutation by Sanger sequencing on all family members (Fig. 2c). However, the mutation was not detected in the father and was present in a heterozygous state in the mother (Fig. 2c). In order to examine whether germline mosaicism is present in the father, the DNA of the father's sperm was sequenced, but the results were largely inconclusive as a small peak for A allele and an even smaller peak for C allele is present at the c.326 position (Additional file 1: Figure S1).

Long-read sequencing identified a structural variant in the other allele

To evaluate whether a structural variant is present in the proband that masks the c.326G > A mutation as homozygous, we carried out long-read whole-genome sequencing on the proband using the Oxford Nanopore technology. We generated 2,251,269 base-called reads containing 35,595,548,336 bases (~12X whole-genome coverage) with an average read length of 16,579 bp. Using the long-read sequencing data, a novel deletion on chr17 g.41049904_41057049del7146 (GRCh37) was detected in one allele of *G6PC* (Fig. 3a), and the known heterozygous point mutation c.326G > A was detected in the other allele. This deletion was supported by four reads, though with slightly discordant breakpoints due to possible alignment errors. The deletion completely covers the first two exons of the *G6PC* gene, thus resulting in loss of function.

Fig. 2 Identification of a c.326G > A missense mutation in the *G6PC* gene. (**a**) Whole-exome sequencing identified a homozygous c.326G > A missense variant in exon 2 of the *G6PC* gene. (**b**) The amino acid 109 (marked by red color) affected by c.326G > A is highly conserved across different species. (**c**) Sanger sequencing on the pedigree showed that the father does not carry the c.326G > A missense variant and that the mother carries a heterozygous c.326G > A missense variant

Fig. 3 Long-read sequencing identified a deletion in the *G6PC* gene. (**a**) IGV screen shot of reads at the *G6PC* locus. Four reads carry a deletion (chr17 g.41049904_41057049del7125 that starts from the first intron of the *LINC00671* gene to intron 2 of the *G6PC* gene. (**b**) Quantitative PCR validation of the deletion in the trio. Relative quantitation (RQ) of copy number was analyzed by the ΔΔCT method, and error bars represent standard deviation. The deletion includes exon 1F (5′-Flanking introns), exon 1 and exon 2, and the patient and his father are mutation carriers while his mother is normal. (**c**) Sanger validation of the deletion breakpoints. The first sequence shows the mutated genomic segment, while the second and third sequences show expected genomic segments if deletion is not present. The red arrow refers to the breakpoint, and a 7125 bp sequence is deleted based on the human reference genome (GRCh37). (**d**) Depiction of the protein domains that were targeted by the non-synonymous mutation and the 7.1 kb deletion. (**e**) Illustration of the genomic contexts of the two breakpoints, which are both located in known Alu elements. (**f**) Gel electrophoresis of the PCR product designed to detect the deletion. The lane marked with M represent GeneRuler 50 bp DNA Ladder (Thermo Scientific™), and all lanes (except "-"lane) include an ~ 800 bp internal control (β-Globin gene). A 418 bp fragment can be amplified from the father and the proband

By quantitative RT-PCR, we estimated the copy number of each exon of the *G6PC* gene from the proband and his parents (Fig. 3b). The copy numbers of deletion exon 1F (5′-Flanking region), exon 1 and exon 2 were only about half of the value of the control group, which indicated that the patient and his father both carry a heterozygous deletion.

To further clarify the location of the breakpoint of this large deletion, we designed six breakpoint primers (Table 1) to amplify and sequence the suspected breakpoint region. The result confirmed our prediction and found the breakpoints precisely at chr17:41049879 and chr17:41057003 (Fig. 3c). These breakpoints were only 20-50 bp different from the predictions from the long-read sequencing data. This large deletion (chr17 g.41049879_41057003del7125) was thus 7125 bp in length and contains the 5′ regulatory sequence as well as exon 1, intron 1, exon 2 and partial

intron 2 of the *G6PC* gene (Fig. 3d). Motif analysis was performed and found that the 5′ breakpoint was located in AluJr and the 3′ breakpoint in AluSx (Fig. 3e). The Alu family members have a sequence similarity of over 87% and cover 11% of the human genome [39], and Alu-Alu recombination usually produces a fragment deletion via Alu recombination-mediated deletion (ARMD) [40, 41]. Note that WES missed this deletion and judged the heterozygous missense mutation as homozygous, since we were unable to observe any obvious coverage differences between the several exons in the gene from WES (Fig. 2a). Therefore, the patient inherited the missense mutation (c.326G > A) from his mother and inherited the deletion mutation (chr17 g.41049879_41057003del7125) from his father. We further designed a PCR-based assay to easily differentiate samples with and without deletion using a set of primers

(G6PC-Del-F/G6PC-Del-R), which generates a 418 bp target fragments in individuals carrying the deletion (Fig. 3f).

Preimplantation genetic diagnosis and implantation of the embryo

In order to help the family to plan for another child, a reproductive intervention was carried out by in vitro fertilization with preimplantation genetic diagnosis (PGD). To avoid the allelic drop-out (ADO), microsatellite markers D17S760 (location 5′ of G6PC, – 0.9 M), DS17S793 (location 5′ of G6PC, – 0.7 M) and DS17S951 (location 3′ of G6PC, 0.8 M) in linkage with the breakpoint junction and the point mutations were tested. Four embryos developed to blastocysts, and were collected and biopsied. The deletion mutation (chr17 g.41049879_41057003del7125) was not detected in four embryos, and we ruled out the possibility of allelic drop-out via linkage analysis by microsatellite markers. However, all the four embryos were identified to be carriers of the missense mutation (c.326G > A). Data on the STR sites D17S760, DS17S793 and DS17S951 also indicated that the embryo inherited the maternal risk chromosome (Additional file 1: Table S2). Furthermore, three alleles of D17S760 were found in embryo No. 2, suggesting that it may be partial trisomy in chromosome 17. Considering the state of the four embryos comprehensively, embryo No. 1 was implanted (Additional file 1: Table S1).

After the implantation of the embryo, the patient's mother succeeded in pregnancy and came to our hospital for a review at the 19+ week of pregnancy. We obtained amniotic fluid cells of the fetus to extract DNA. The genetic testing confirmed that the fetus is a carrier of the missense mutation (c.326G > A), and does not inherit the deleterious deletion. The newborn was revisited in our hospital in December 2017. The baby had a fasting blood glucose level of 5.5 mmol/L and her B-ultrasonogram showed that the liver and kidneys were normal (Additional file 1: Figure S2), confirming that she was not affected with GSD-Ia.

Discussion

In the current study, we performed genetic diagnosis on an affected subject with suspected GSD-Ia through clinical whole-exome sequencing and long-read whole-genome sequencing. The proband was misidentified as a homozygote for the c.326G > A mutation by WES, but later confirmed to be a compound heterozygous carrier of the c.326G > A mutation and a 7.1 kb deletion spanning this point mutation. The missense mutation and the deletion were inherited from the mother and father, respectively. Therefore, through combined exome sequencing and long-read whole-genome sequencing, we yielded a definitive genetic diagnosis on

the proband, and used this information to design assays to enable successful personalized preimplantation genetic diagnosis following IVF.

After we identified the two causal mutations, we also retrospectively examined the exome sequencing data to understand why the deletion was not found previously. All exons in the G6PC gene were covered well in the exome data: the mean depth for exon 1, exon 2, exon 3, exon 4 and exon 5 were 123, 74, 119, 190 and 259, respectively. However, due to the large variability of coverage between exons, and due to the presence of two intronic breakpoints, we did not determine that a deletion covering exon 1 and exon 2 was present from the short-read sequencing data. If we had performed qPCR assays on each exon after WES, we could have identified the deletion that covers exon 1 and 2; nevertheless, we still cannot find the exact breakpoint from WES or qPCR data, yet knowing the exact breakpoint will be important for the purpose of PGD.

The human G6PC gene is a single-copy gene that contains five exons and spans 12.5 kb of DNA on chromosome 17q21 [42]. The G6PC gene encodes G6Pase which is a 357 amino acid protein anchored to the endoplasmic reticulum (ER) membrane with nine transmembrane domains [43]. The amino-terminus of the protein lies in the ER lumen with the enzymatic active site and the carboxyl-terminus in the cellular cytoplasm [26, 44]. Based on the predicted structure on UniProt (http://www.uniprot.org/uniprot/P35575), the novel point mutation (c.326G > A; p.C109Y) detected in this study is located in the lumen of the ER and the deletion fragment (chr17:g.41049879_41057003del7125) contains at least five transmembrane domains. Since the deletion spans the starting codon of the protein, the allele carrying the deletion should not to be transcribed or translated.

To date, differential diagnosis of GSD generally relies on the molecular analysis and has replaced the traditional liver biopsy [28, 44]. Detection and analysis of suspicious disease-causing mutations have become a powerful tool for differential diagnosis of GSD and can guide the implementation of PGD and the personalized treatment even further. WES has gradually become a powerful means by which clinicians and scientists can detect the underlying cause of various genetic diseases. However, a major shortcoming of WES is uneven coverage of sequence reads over the capture regions, contributing to many low coverage regions, which hinders accurate variant calling [45]. This challenge did not affect our study per se, since we were able to identify the c.326G > A mutation from WES accurately, but the uneven coverage from exome sequencing prevented us from finding the deletion covering two exons. Nevertheless, given that this mutation is a very rare mutation (not documented in public databases) and that there is

no known consanguinity in the family, it is unlikely that the proband inherits both alleles from parents. Initially we suspected that the father carries germline mosaicism, yet sperm sequencing on the father did not fully resolve the question (Additional file 1: Figure S1). A low coverage long-read whole-genome sequencing resolved this issue. A large deletion (initially designated as chr17:g.41049 904_41057049del7146 based on alignment) was detected in the proband and his father. Then we conducted the Sanger sequencing and the quantitative RT-PCR to validate this result and refine the exact position of the breakpoint junction (chr17:g.41049879_41057003del7125). The slight inconsistence of the deletion breakpoints between initial long-read sequencing and subsequent Sanger sequencing were likely due to the higher error rate of long-read sequencing and the imperfect alignment of the reads. Long-read sequencing can identify complex SVs effectively, thus compensated for the shortcomings of WES and avoided a misdiagnosis and potential failure of PGD.

Conclusion

In summary, we present one of the first examples of using long-read sequencing to identify causal yet complex structural variants (SVs) in exome-negative patients, which subsequently enabled successful personalized preimplantation genetic diagnosis. Our study suggests that long-read sequencing offers a means to discover overlooked genetic variation in patients undiagnosed or misdiagnosed by short-read sequencing, and may potentially improve diagnostic yields in clinical settings, especially when only one pathogenic mutation is found in an affected individual suspected to carry a recessive disease.

Additional file

Additional file 1: Table S1. Characteristics of four embryos after in vitro fertilization. The four embryos were scored according to the Istanbul consensus. Embryo No.1 was transplanted. **Figure S1.** Sanger sequencing on the blood samples (left panel) and sperm samples (right panel) of the father, to determine whether germline mosaicism on c.326G > A is present. However, the results were inconclusive, and indeed a small peak of A allele and an even smaller peak for C allele is present at the c.326 position. **Figure S2**. Image of type-B ultrasonic of liver on the newborn baby. B-ultrasonogram showed that the liver and kidneys were normal. SP: Spleen; LK: Left kidney. (DOCX 372 kb)

Abbreviations
ACMG: The American College of Medical Genetics and Genomics; G6PC: glucose-6-phosphatase; GSD: Glycogen storage disease; GSD-Ia: Glycogen storage disease type Ia; HGMD: Human gene mutation database; IVF: In-vitro fertilization; PCR: Polymerase chain reaction; PGD: Preimplantation genetic diagnosis; WES: Whole exome sequencing

Availability of data and materials
The following additional data are available with the online version of this paper. Additional file 1: Tables S1 and Additional file 1: Figures S1-S2 are available online. All raw long-read sequencing data, short-read sequencing data, Sanger sequencing data and real-time PCR data in the genomic region of interest are available through appropriate institutional MTA. The detailed instructions on the software tools used in the analysis are also available to facilitate reproduce the results presented in the current study.

Authors' contributions
HM and JZ lead the experimental design, facilitated data analysis and wrote the manuscript. QY, FL generated Nanopore sequencing data and performed data analysis. NM, BG, JD, GL performed clinical assessment of the patients, generated short-read sequencing data, performed preimplantation genetic diagnosis and advised on data interpretation. KW and QZ conceived and guided the execution of the study, and revised the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate
The research including human subjects, human material, human data, has been performed in accordance with the Declaration of Helsinki and was approved by the ethics committee of Reproductive and Genetic Hospital of CITIC-Xiangya.

Consent for publication
The research contains individual person's data (including detailed clinical phenotypes and images) and we have obtained consent from parents of the children.

Competing interests
J.Z., Q.Y., F.L., D.W. are employees and K.W. is consultant of Grandomics Biosciences. All the other authors have no competing interests.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
[1]Institute of Reproductive and Stem Cell Engineering, Central South University, Changsha 410078, Hunan, China. [2]Reproductive and Genetic Hospital of CITIC-Xiangya, Changsha 410078, Hunan, China. [3]GrandOmics Biosciences, Beijing 102206, China. [4]Raymond G. Perelman Center for Cellular and Molecular Therapeutics, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA.

References
1. Posey JE, Rosenfeld JA, James RA, Bainbridge M, Niu Z, Wang X, Dhar S, Wiszniewski W, Akdemir ZH, Gambin T, et al. Molecular diagnostic experience of whole-exome sequencing in adult patients. Genet Med. 2016; 18(7):678–85.
2. Yang Y, Muzny DM, Reid JG, Bainbridge MN, Willis A, Ward PA, Braxton A, Beuten J, Xia F, Niu Z, et al. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. N Engl J Med. 2013;369(16):1502–11.
3. Retterer K, Juusola J, Cho MT, Vitazka P, Millan F, Gibellini F, Vertino-Bell A, Smaoui N, Neidich J, Monaghan KG, et al. Clinical application of whole-exome sequencing across clinical indications. Genet Med. 2016;18(7):696–704.
4. Trujillano D, Bertoli-Avella AM, Kumar Kandaswamy K, Weiss ME, Koster J, Marais A, Paknia O, Schroder R, Garcia-Aznar JM, Werber M, et al. Clinical exome sequencing: results from 2819 samples reflecting 1000 families. Eur J Hum Genet. 2017;25(2):176–82.
5. Lee H, Deignan JL, Dorrani N, Strom SP, Kantarci S, Quintero-Rivera F, Das K, Toy T, Harry B, Yourshaw M, et al. Clinical exome sequencing for genetic identification of rare Mendelian disorders. Jama. 2014;312(18):1880–7.

6. Gilissen C, Hehir-Kwa JY, Thung DT, van de Vorst M, van Bon BW, Willemsen MH, Kwint M, Janssen IM, Hoischen A, Schenck A et al: Genome sequencing identifies major causes of severe intellectual disability. Nature 2014, 511(7509):344–347.

7. Willig LK, Petrikin JE, Smith LD, Saunders CJ, Thiffault I, Miller NA, Soden SE, Cakici JA, Herd SM, Twist G, et al. Whole-genome sequencing for identification of Mendelian disorders in critically ill infants: a retrospective analysis of diagnostic and clinical findings. Lancet Respir Med. 2015;3(5):377–87.

8. Schwarze K, Buchanan J, Taylor JC, Wordsworth S. Are whole-exome and whole-genome sequencing approaches cost-effective? Genet Med: A systematic review of the literature; 2018.

9. Hannan AJ. Tandem repeats mediating genetic plasticity in health and disease. Nat Rev Genet. 2018;19(5):286–98.

10. Ashley EA. Towards precision medicine. Nat Rev Genet. 2016;17(9):507–22.

11. Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SFA, Hakonarson H, Bucan M. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res. 2007;17:1665–74.

12. Colella S, Yau C, Taylor JM, Mirza G, Butler H, Clouston P, Bassett AS, Seller A, Holmes CC, Ragoussis J. QuantiSNP: an objective Bayes hidden-Markov model to detect and accurately map copy number variation using SNP genotyping data. Nucleic Acids Res. 2007;35(6):2013–25.

13. Venkatraman ES, Olshen AB. A faster circular binary segmentation algorithm for the analysis of array CGH data. Bioinformatics. 2007;23(6):657–63.

14. Olshen AB, Venkatraman ES, Lucito R, Wigler M. Circular binary segmentation for the analysis of array-based DNA copy number data. Biostatistics. 2004;5(4):557–72.

15. Alkan C, Coe BP, Eichler EE. Genome structural variation discovery and genotyping. Nat Rev Genet. 2011;12(5):363–76.

16. Merker JD, Wenger AM, Sneddon T, Grove M, Zappala Z, Fresard L, Waggott D, Utiramerur S, Hou Y, Smith KS, et al. Long-read genome sequencing identifies causal structural variation in a Mendelian disease. Genet Med. 2017.

17. Kirby A, Gnirke A, Jaffe DB, Baresova V, Pochet N, Blumenstiel B, Ye C, Aird D, Stevens C, Robinson JT, et al. Mutations causing medullary cystic kidney disease type 1 lie in a large VNTR in MUC1 missed by massively parallel sequencing. Nat Genet. 2013;45(3):299–303.

18. Sedlazeck FJ, Rescheneder P, Smolka M, Fang H, Nattestad M, von Haeseler A, Schatz MC: Accurate detection of complex structural variations using single-molecule sequencing. Nat Methods 2018, 15(6):461–468.

19. Sanchis-Juan A, Stephens J, French CE, Gleadall N, Mégy K, Penkett C, Stirrups K, Delon I, Dewhurst E, Dolling H, et al. Complex structural variants resolved by short-read and long-read whole genome sequencing in Mendelian disorders. BioRxiv. 2018. https://doi.org/10.1101/281683.

20. Liu Q, Zhang P, Wang D, Gu W, Wang K. Interrogating the "unsequenceable" genomic trinucleotide repeat disorders by long-read sequencing. Genome Med. 2017;9(1):65.

21. Loomis EW, Eid JS, Peluso P, Yin J, Hickey L, Rank D, McCalmon S, Hagerman RJ, Tassone F, Hagerman PJ. Sequencing the unsequenceable: expanded CGG-repeat alleles of the fragile X gene. Genome Res. 2013;23(1):121–8.

22. Shi L, Guo Y, Dong C, Huddleston J, Yang H, Han X, Fu A, Li Q, Li N, Gong S, et al. Long-read sequencing and de novo assembly of a Chinese genome. Nat Commun. 2016;7:12065.

23. Seo JS, Rhie A, Kim J, Lee S, Sohn MH, Kim CU, Hastie A, Cao H, Yun JY, Kim J, et al. De novo assembly and phasing of a Korean human genome. Nature. 2016;538(7624):243–7.

24. Jain M, Koren S, Miga KH, Quick J, Rand AC, Sasani TA, Tyson JR, Beggs AD, Dilthey AT, Fiddes IT, et al. Nanopore sequencing and assembly of a human genome with ultra-long reads. Nat Biotechnol. 2018;36(4):338–45.

25. Pendleton M, Sebra R, Pang AW, Ummat A, Franzen O, Rausch T, Stutz AM, Stedman W, Anantharaman T, Hastie A, et al. Assembly and diploid architecture of an individual human genome via single-molecule technologies. Nat Methods. 2015;12(8):780–6.

26. Chou JY, Mansfield BC. Mutations in the glucose-6-phosphatase-alpha (G6PC) gene that cause type Ia glycogen storage disease. Hum Mutat. 2008; 29(7):921–30.

27. Zheng BX, Lin Q, Li M, Jin Y. Three novel mutations of the G6PC gene identified in Chinese patients with glycogen storage disease type Ia. Eur J Pediatr. 2015;174(1):59–63.

28. Kishnani PS, Austin SL, Abdenur JE, Arn P, Bali DS, Boney A, Chung WK, Dagli AI, Dale D, Koeberl D, et al. Diagnosis and management of glycogen storage disease type I: a practice guideline of the American College of Medical Genetics and Genomics. Genetics in medicine : official journal of the American College of Medical Genetics. 2014;16(11):e1.

29. Matern D, Seydewitz HH, Bali D, Lang C, Chen YT. Glycogen storage disease type I: diagnosis and phenotype/genotype correlation. Eur J Pediatr. 2002; 161(Suppl 1):S10–9.

30. Chou JY, Jun HS, Mansfield BC. Type I glycogen storage diseases: disorders of the glucose-6-phosphatase/glucose-6-phosphate transporter complexes. J Inherit Metab Dis. 2015;38(3):511–9.

31. Sedlazeck FJ, Rescheneder P, Smolka M, Fang H, Nattestad M, von Haeseler A, Schatz MC. Accurate detection of complex structural variations using single molecule sequencing. Nat Methods. 2018;15(6):461–468.

32. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. Genome project data processing S: the sequence alignment/map format and SAMtools. Bioinformatics. 2009;25(16):2078–9.

33. Maria, Nattestad C-SC, Schatz MC. Ribbon: Visualizing complex genome alignments and structural variation. bioRxiv. 2016. https://doi.org/10.1101/082123.

34. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 2013;14(2):178–92.

35. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16):e164.

36. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015;17(5):405–24.

37. Li Q, Wang K. InterVar: clinical interpretation of genetic variants by the 2015 ACMG-AMP guidelines. Am J Hum Genet. 2017;100(2):267–80.

38. Alpha Scientists in Reproductive Medicine and ESHRE Special Interest Group of Embryology. The Istanbul consensus workshop on embryo assessment: proceedings of an expert meeting. Hum Reprod. 2011;26(6):1270–83.

39. Ade C, Roy-Engel AM, Deininger PL. Alu elements: an intrinsic source of human genome instability. Current opinion in virology. 2013;3(6):639–45.

40. Sen SK, Han K, Wang J, Lee J, Wang H, Callinan PA, Dyer M, Cordaux R, Liang P, Batzer MA. Human genomic deletions mediated by recombination between Alu elements. Am J Hum Genet. 2006;79(1):41–53.

41. Kim S, Cho CS, Han K, Lee J. Structural variation of Alu element and human disease. Genomics Inform. 2016;14(3):70–7.

42. Lei KJ, Shelly LL, Pan CJ, Sidbury JB, Chou JY. Mutations in the glucose-6-phosphatase gene that cause glycogen storage disease type 1a. Science. 1993;262(5133):580–3.

43. Chou JY, Jun HS, Mansfield BC. Glycogen storage disease type I and G6Pase-beta deficiency: etiology and therapy. Nat Rev Endocrinol. 2010; 6(12):676–88.

44. Gu LL, Li XH, Han Y, Zhang DH, Gong QM, Zhang XX. A novel homozygous no-stop mutation in G6PC gene from a Chinese patient with glycogen storage disease type Ia. Gene. 2014;536(2):362–5.

45. Wang Q, Shashikant CS, Jensen M, Altman NS, Girirajan S. Novel metrics to measure coverage in whole exome sequencing datasets reveal local and global non-uniformity. Sci Rep. 2017;7(1):885.