Analytical validity and clinical utility of whole-genome sequencing for cytogenetically balanced chromosomal abnormalities in prenatal diagnosis: abridged secondary publication

BHY Chung *, ASY Kan, KYK Chan, W Yang, MHY Tang, CCY Mak, GKC Leung

KEY MESSAGES

1. Whole-genome sequencing was superior to conventional karyotyping in detection of banding in nine of 10 cases. Improvement in detection could be confirmed by orthogonal methods.
2. Gene disruption was identified in two cases and led to definitive diagnoses for two families with proband affected by X-linked epilepsy (disruption of PCDH19) or microcephalic osteodysplastic primordial dwarfism type II (MOPDII). In the third case, gene disruption was identified, but variant of uncertain significance was concluded.
3. Using both short read (Illumina) and long read (Nanopore & PacBio) sequencing data together with bioinformatics tool (WhatsHap) can detect additionally the phasing of mutations.
4. Understanding the genomic mechanism of gene disruption secondary to balanced chromosomal abnormalities can guide the genetic counselling. It allows accurate and personalised disease risk prediction.

Hong Kong Med J 2022;28(Suppl 1):S4-7
HMRF project number: 05162986

[1,2] **BHY Chung,** [2,3] **ASY Kan,** [2,3] **KYK Chan,** [1] **W Yang,** [2,3] **MHY Tang,** [1] **CCY Mak,** [1] **GKC Leung**

[1] *Department of Paediatrics and Adolescent Medicine, LKS Faculty of Medicine, The University of Hong Kong*
[2] *Department of Obstetrics and Gynaecology, Queen Mary Hospital, The University of Hong Kong*
[3] *Prenatal Diagnostic Laboratory, Department of Obstetrics and Gynaecology, Tsan Yuk Hospital*

* Principal applicant and corresponding author: bhychung@hku.hk

Introduction

Balanced chromosomal abnormalities (BCAs) are common genomic variations. Structural variants refer to a class of genomic variation at the chromosomal level, including copy number variations and balanced chromosomal rearrangements (eg, translocations and inversions), which involve changes in either localisation or orientation of a chromosomal segment without visible gain or loss of chromosomal material. BCA is a major contributor to human genetic variation. It occurs in both healthy and diseased individuals, affecting about one (0.2%) in every 500 newborns.[1] Although most cases are not associated with a pathological phenotype, these cases have an increased risk of developing multiple congenital anomalies, autism spectrum disorders, or intellectual disability.[1] In 370 000 women who underwent prenatal diagnosis by conventional cytogenetics, the risk for serious congenital anomalies was 6.1% for *de novo* reciprocal translocation, 3.7% for Robertsonian translocations, and 9.4% for inversions.[2] However, interpretation of BCAs by conventional cytogenetics is restricted by the resolution limit.

Conventional karyotyping is considered the standard for prenatal diagnosis of chromosomal disorders. It can detect numerical abnormalities as well as structural rearrangements (balanced or unbalanced) but is limited to a microscopic resolution of 3 to 10 Mb.[3] Whole-genome array comparative hybridisation can detect chromosomal imbalances at a resolution in the tens to hundreds of kilobases and can sometimes detect cryptic deletions and duplications associated with BCAs that are not detectable by karyotyping.[4]

Whole-genome sequencing (WGS) has emerged as a comprehensive diagnostic tool for the detection of a wide range of genomic changes. Paired-end sequencing refers to the massively parallel sequencing of both ends of the same DNA fragment by use of adaptors to generate short sequence reads. WGS performed using this method can localise BCA breakpoints at the nucleotide level. This can substantially improve the diagnostic resolution, compared with karyotyping and other cytogenetic methods.

Methods

Genomic DNA was extracted from the thawed cultured cells of the stored chorionic villi or amniotic fluid samples according to standard protocols. We followed the recommendations of the Laboratory Quality Assurance Committee of the American College of Medical Genetics and Genomics (ACMG) on the use of WGS for diagnostic purposes. We sequenced the 150-bp-paired ends of contiguous genomic DNA fragments of approximately 1 to 2 Kb. Sequencing using the Illumina HiSeq 2500

platform was performed by Macrogen or the Centre for Genomic Sciences, the University of Hong Kong.

We then interrogated the BCA breakpoints and determined whether they resulted in: (1) gene disruption (ie, whether the translocation breakpoint intercepts the open-reading frame of a gene that may result in termination of a gene transcript or translation of a fusion protein), (2) genomic imbalances (ie, cryptic deletion or duplication identified adjacent to the chromosome breakage), and (3) alteration of topologically associating domains (ie, long-range effects of cis-acting regulatory elements on protein-coding genes). Topologically associating domains are key elements of mammalian regulatory organisation and are conserved genomic conformations that partition the genome into megabase-sized compartments with frequent intra-domain regulation.

We retrospectively identified those with apparent *de novo* BCAs through the internal database of the Prenatal Diagnostic Laboratory, Tsan Yuk Hospital. Patients with multiple pregnancies, molar pregnancy, ectopic pregnancy, or complications associated with known environmental or teratogenic causes were excluded, as were those with isolated/ multiple major fetal structural abnormalities detected by ultrasonography. The presence of major structural abnormalities already affects the decision of termination of pregnancy. Ethics approval was obtained from the Institutional Review Board of the University of Hong Kong/ Hospital Authority Hong Kong West Cluster (UW 18-045).

Results

Ten patients with apparent *de novo* BCAs were identified: eight with reciprocal translocation and two with inversions (Table) Pair-ended short-read WGS identified the BCA and their breakpoints at single-nucleotide resolution. In five cases, extra DNA was available for validation using orthogonal methods (to confirm the WGS findings) including Sanger sequencing (n=4) and Nanopore sequencing and PacBio sequencing (n=1). The higher resolution of detecting breakpoints at the single-nucleotide level by WGS had improved detection of banding over karyotype in nine of ten cases and had the same banding detection as karyotype in one case.

WGS provided additional genomic information over conventional karyotype in genetic counselling in the following three cases.

Case 1 (subject 11F0523) was a product of intracytoplasmic sperm injection and showed a karyotype of 46,XX,t(18;19)(q12.2;q13.1)dn. Using WGS, we identified the breakpoints at chr18:29,652,147 and chr19:36,930,887, which were validated by Sanger sequencing (Fig 1). The translocation disrupted the gene *RNF125* at its last

TABLE. Comparison of the detection by Karyotype and WGS on 10 BCA cases

Case	Subject	Karyotype	Clinical indication for prenatal diagnosis	Banding (conventional vs whole-genome sequencing)	Breakpoints (GRCh37)	OMIM morbid gene	Orthogonal validation
1	11F0523	46,XX,t(18;19)(q12.2;q13.1)dn	Baby of intracytoplasmic sperm injection	18q12.2 vs 18.q12.1 19q13.1 vs 19q13.12	chr18:29,652,147 chr19:36,930,887	*RNF125*	Sanger
2	11F0621	46,XX,t(4;12)(q35;p13.1)dn	Maternal age	4q35 vs 4q35.2 12p13.1 vs 12p12.3	chr4:186,776,072 chr12:15,513,829	-	-
3	11F0959	46,X,inv(X)(p21q22.1)dn	Increased cardio-thoracic ratio	Xp21 vs Xp21.1 Xq22.1 vs Xq22.1	chrX:34,271,812 chrX:99,594,533	*PCDH19*	-
4	12F0839	46,XY,t(5;9)(q13;q32)dn	Bilateral prominent renal pelvis, borderline ventriculomegaly	5q13 vs 5q14.1 9q32 vs 9q32	chr5:78,899,784 chr9:117,899,085	-	-
5	12F1057	46,XY,t(1;4)(p21q21.1)dn	Maternal age	1p21 vs 1p21.2 4q21.1 vs 4q21.1	chr1:100,004,147 chr4:88,309,179	-	Sanger
6	13C0062	46,XX,t(8;11)(q22;q13)dn	Increased nuchal traslucency	8q22 vs 8q22.1 11q13 vs 11q13.1	chr8:97,086,794 chr11:65,540,889	-	Sanger
7	13C0266	46,XY,t(7;18;12)(q31;p11.3;q15)dn	Head shape roundish	7q31 vs 7q31.1 18p11.3 vs 18p11.31 12q15 vs 12q15	chr7:110493772 chr18:3766591 chr12:69794115	-	-
8	12F0050	46,XY,inv(21)q11.2q22.3)dn	Early-onset intrauterine growth retardation, mildly echogenic bowel	21q11.2 vs 21q11.2 21q22.3 vs 21q22.3	chr21:14,953,345 chr21:47,839,992	*PCNT*	PacBio, Nanopore
9	08F1641	46,XY,t(6;8)(p21.1;q24.1)dn	Echogenic focus in left ventricle, short femur, short humerus	6p21.1 vs 6p21.1 8q24.1 vs 8q24.21	chr6:45,841,060 chr8:126,545,471	-	Sanger
10	18T1192	46,XX,t(3;7,6)(q25;q36;q21)dn	Cleft lip	3q25 vs 3q25.32 6q21 vs 6q21 7q36 vs 7q36.3	chr3:157254970 chr6:106600067 chr7:158852085	-	-

exon. *RNF125* is associated with Tenorio syndrome (OMIM: 616260), which is an autosomal dominant disease characterised by overgrowth, macrocephaly, and intellectual disability syndrome. *RNF125* is a gene that is predicted to be tolerant to loss-of-function mutations (indicated by a low pLI score). In addition, all reported pathogenic variants in *RNF125* are missense mutations. According to the ACMG/AMP guideline, the variant was classified as a variant of uncertain significance. From the medical record, this pregnancy resulted in a healthy live-born with normal growth parameters.

Case 3 (subject 11F0959) was a female fetus with increased cardio-thoracic ratio who had a karyotype of 46,X,inv(X)(p21q22.1)dn. Using WGS, we identified the breakpoints at chrX:34,271,812 and chrX:99,594,533, which resulted in the disruption of exon 6 of *PCDH19*, a gene in which loss of function mutation was associated with X-linked early infantile epileptic encephalopathy (OMIM: 300088), which is known to affect mainly female. The carrier males are largely unaffected, except for minor psychiatric/behavioural abnormalities. According to the ACMG/AMP guideline for variant interpretation, the variant was classified as pathogenic. This pregnancy was terminated and therefore postnatal outcome was not assessed to compare with the molecular diagnosis. Nonetheless, we predicted that the baby would have a high chance of developing early-onset severe epilepsy causing encephalopathy, given the full penetrance of the condition.

Case 8 (subject 12F0050) was affected with severe intrauterine growth retardation (IUGR) of unknown aetiology during pregnancy. From prenatal diagnosis, a karyotype of 46,XY,inv(21)(q11.21q22.3)dn was identified, but this finding was not considered to be explaining the finding of IUGR. The couples decided to carry on the pregnancy, and the baby was delivered at 35 weeks. The baby showed features of IUGR, microcephalic, and mesomelic limb shortening. Microcephalic osteodysplastic primordial dwarfism type II (MOPDII) was suspected clinically. Initial NGS panel showed that there was one heterozygous frameshift mutation in *PCNT*:c.4633_4678delAGACAAGTGTTAAT p.(R1555Afs*6). Although the clinical presentations were compatible with MOPDII, the genetic diagnosis was incomplete because only one pathogenic variant was identified in an autosomal recessive condition. As *PCNT* was located on chromosome 22q11.21, we hypothesised that the *de novo* inversion might have disrupted the other *PCNT* allele not carrying the frameshift mutation. We identified one of the breakpoints of the inversion at chr21:47,839,992, which caused disruption of *PCNT*. However, the distance between the frameshift mutation and the inversion breakpoint is 20 kb apart, and thus we were unable to assess the phasing information of the two variants. Therefore, we decided to proceed to long-read sequencing using Nanopore and PacBio platforms to confirm the breakpoint of the inversion. Using a bioinformatic tool called WhatsHap, we combined the short- and long-read data to construct haplotypes that span across the frameshift mutation and the inversion breakpoint. The two variants were located in different haplotypes, indicating that they were located on different alleles (Fig 2). The molecular and clinical diagnosis of MOPDII was therefore substantiated. As the inversion was *de novo*, we informed the family of the low chance of recurrence in subsequent pregnancies.

Discussion

In contrast to karyotyping that detects BCA breakpoints at sub-band level, WGS can precisely detect breakpoints at single-nucleotide level and identify the disrupted genes. Combined with the ACMG/AMP 2015 guideline for variant interpretation in evaluating the pathogenicity of the genetic variants, we were able to provide a more

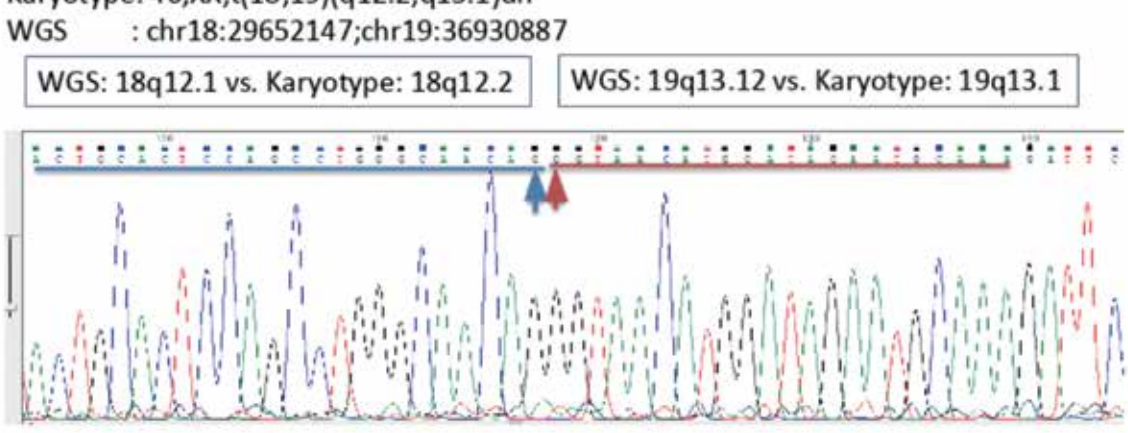
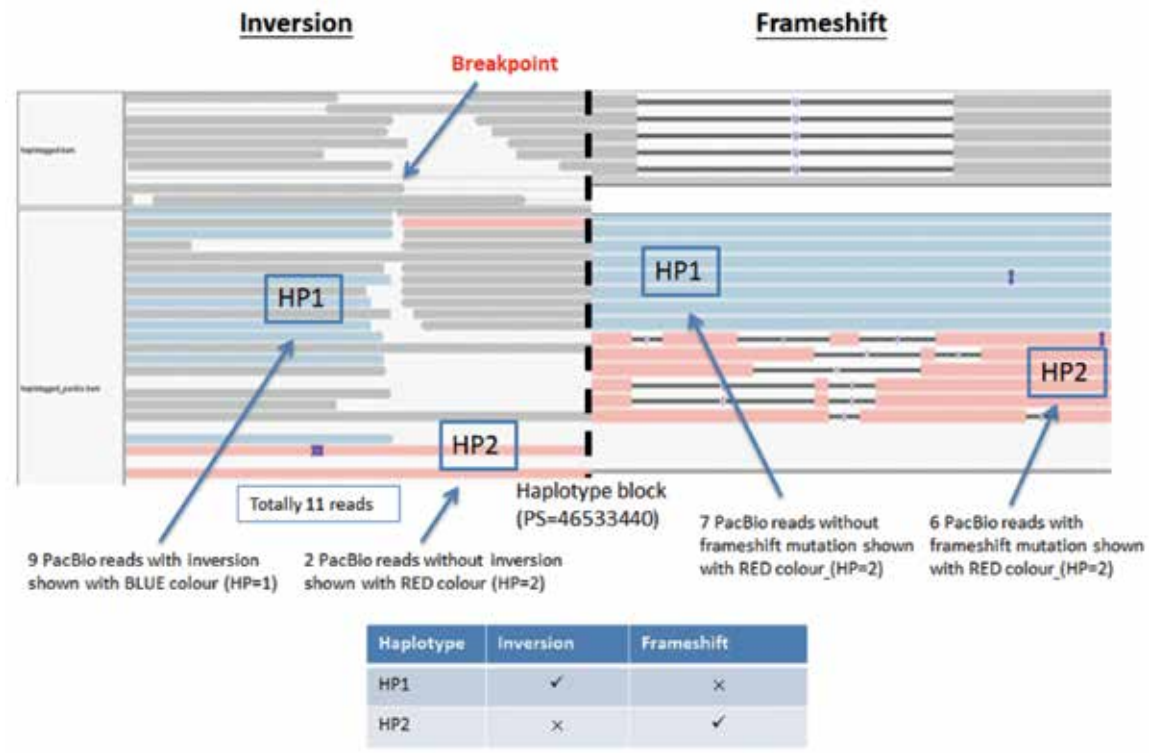


FIG 1. Sanger validation of case 1 (11F0523)



Haplotype	Inversion	Frameshift
HP1	✓	×
HP2	×	✓

FIG 2. Short read and long read of inversion on PCNT gene

precise and personalised genetic counselling on all ten cases, compared with the empiric risk of 6% to 9% recurrence risk. WGS showed improvement in detection of banding over karyotyping in nine of ten cases. This is consistent with a study stating "Revised the breakpoint localisation by at least one sub-band in 93% of subjects when compared to the karyotype interpretation".

We successfully set up the pipeline to analyse WGS data of both short-read (Illumina) and long-read sequencing data (Nanopore and PacBio) in the diagnosis of the case with MOPDII. With the limitation of comparatively large raw error rate (around 15%) of long-read sequencing (Nanopore and PacBio sequencing), the hybrid approach of using both short read and long-read sequencing is the most cost-effective and accurate method. The Illumina short-read platform provides high throughput and accurate sequencing of the single-nucleotide level, whereas the Nanopore and PacBio long-read platform provide easier structural variation and phasing detection. Illumina short-read WGS platform is good at detecting frameshift and structural variations but not good at detecting phasing, whereas Nanopore and PacBio long read is good at detecting structural variations but have limitations on detecting frameshift mutation and phasing owing to large raw error rate and not long enough reads at some cases. Combining both together with bioinformatics tools can effectively detect structural variations, frameshift mutation, and phasing.

Funding

This study was supported by the Health and Medical Research Fund, Food and Health Bureau, Hong Kong SAR Government (#05162986). The full report is available from the Health and Medical Research Fund website (https://rfs1.fhb.gov.hk/index.html).

Disclosure

The results of this research have been previously published in:

1. Yu MHC, Chau JFT, Au SLK, et al. Evaluating the clinical utility of genome sequencing for cytogenetically balanced chromosomal abnormalities in prenatal diagnosis. Front Genet 2021;11:620162.

References

1. Blake J, Riddell A, Theiss S, et al. Sequencing of a patient with balanced chromosome abnormalities and neurodevelopmental disease identifies disruption of multiple high risk loci by structural variation. PLoS One 2014;9:e90894.
2. Warburton D. De novo balanced chromosome rearrangements and extra marker chromosomes identified at prenatal diagnosis: clinical significance and distribution of breakpoints. Am J Hum Genet 1991;49:995-1013.
3. Redin C, Brand H, Collins RL, et al. The genomic landscape of balanced cytogenetic abnormalities associated with human congenital anomalies. Nat Genet 2017;49:36-45.
4. Martin M, Patterson M, Garg S, et al. WhatsHap: fast and accurate read-based phasing. bioRxiv 085050.
5. Cretu Stancu M, van Roosmalen MJ, Renkens I, et al. Mapping and phasing of structural variation in patient genomes using nanopore sequencing. Nat Commun 2017;8:1326.