The diagnostic yield of intellectual disability: Combined whole genome low-coverage sequencing and medical exome sequencing

CURRENT STATUS: UNDER REVIEW

Jun Wang
Affiliated Children's Hospital of Capital Institute of Pediatrics
jam1080800004@163.com Corresponding Author
ORCiD: 0000-0002-9385-7446

Yan Wang
Affiliated Children's Hospital of Capital Institute of Pediatrics

Liwen Wang
Affiliated Children's Hospital of Capital Institute of Pediatrics

Wangyang Chen
Kaiumph Medical Diagnostics Co, Ltd

DOI:
10.21203/rs.2.18388/v2

SUBJECT AREAS
Epigenetics & Genomics Neurology

KEYWORDS
Intellectual disability, copy number variation, single-nucleotide variations, detection rate, whole genome low-coverage sequencing, medical exome sequencing
Abstract

Background: Intellectual disability (ID) is a heterogeneous neurodevelopmental disorder with a complex genetic underpinning in its etiology. Chromosome microarray (CMA) is recommended as the first-tier diagnostic test for ID due to high detection rate of copy number variation (CNV).

Methods: To identify an appropriate clinical detection scheme for ID in Han Chinese patients, a whole genome low-coverage sequencing was performed as the first-tier diagnostic test, followed by medical exome sequencing as the second-tier diagnostic test for those patients with negative results of CNVs.

Results: A total of 19 pathogenic CNVs in 16/95 (16.84%) ID patients and 11 pathogenic single-nucleotide variations, including 7 novel mutations in 9/79 (11.39%) ID patients were identified on whom no pathogenic CNVs were discovered. The detection rate of CNVs in ID s with multiple congenital anomalies (MCA) subgroup was significantly higher than ID s with autism spectrum disorder and the other IDs subgroup. Also single-nucleotide variations showed a higher occurrence rate in the other IDs subgroup.

Conclusions: There were differences in the diagnostic yields of different variation types among the three ID subgroups. Our findings provided a new perspective on appropriate clinical detection scheme for different ID subgroups based on the statistical results of different ID subgroups. The application of whole genome low-coverage sequencing as the first-tier diagnostic test for ID with MCA subgroup and medical exome sequencing as the first-tier diagnostic test for other IDs subgroup were considered as an efficient clinical detection schemes.

Background

Intellectual disability (ID) is characterized by impairment of cognitive and social
adjustment due to brain damage or incomplete development, which accounted for a prevalence of about 1-3% in worldwide\textsuperscript{1-4}. ID is a complicated neurodevelopmental disorder that might be affected by a series of heterogeneous factors such as environmental, genetic, idiopathic, neonatal sequelae and other diseases\textsuperscript{2}. Many studies have been conducted to reveal the genetic etiology of ID and there are more than 600 genes and 130 rare copy number variations (CNVs) involved in cause of ID\textsuperscript{5-9}. However, the causes in up to 50% cases cannot be found due to complex etiological factor and high genetic heterogeneity\textsuperscript{7, 10-13}.

The well-known variant types of the causes of ID included chromosome number or structure abnormality, genome-wide microdeletion or microduplication and single gene defect. Chromosome microarray (CMA) has been recommended as the first-tier diagnostic test for ID and so, CNVs may have the highest positive rate with regard to ID\textsuperscript{14, 15}. Most of the previous studies conducted on genetic etiology of ID were based on CMA\textsuperscript{16-18}.

According to a previous study, chromosome abnormality and genome-wide microdeletion or microduplication accounted for 10%-20% of ID\textsuperscript{3, 19}. Although CMA has been recommended as the first-tier diagnostic test for ID, it is still limited due to insufficiency in detecting single gene defects.

With the rapid development and wide-use of next generation sequencing in the clinical diagnostic field, the novel technologies conducted genetic research on ID with a rapid approach to identify more causative CNVs and genes\textsuperscript{7, 20, 21}. Previous studies have shown that significant parts involved in ID patients, with negative results after CMA, and included a single gene defect. Single gene defect may account for about 10% of ID\textsuperscript{7}. As more and more novel or candidate genes were discovered, it could be even higher. A number of
novel ID candidate genes, such as ASH1L, MBOAT7 and TRIO, were identified by using next generation sequencing in patients who had negative results after CMA\(^{22-24}\).

Most of the genetic studies on intellectual disabilities were based on European populations. The ID genetic research studies on Chinese population were less. And majority of these studies focused on CNVs through CMA\(^{18}\). Investigating the distribution of CNVs, proportion of single gene defects, and evaluation of effects of different diagnostic platforms in Chinese ID populations provided the proof for selecting appropriate method of clinical genetic diagnosis in Chinese ID populations. We herein performed a whole genome low-coverage sequencing as the first-tier diagnostic test in 95 Chinese ID patients and medical exome sequencing was applied as the second-tier diagnostic test for those patients with negative results of CNVs to search for any additional single gene defects. Pathogenic CNVs and single gene mutations were identified by both clinical phenotype relevance and genetic interpretation. We compared the detection rates of CNVs and single gene defects in our study with the previous study. Furthermore, we subdivided ID patients into three different subgroups, ID with multiple congenital anomalies (MCA), ID with autism spectrum disorders (ASD) and other IDs. Compared with the detection rates of CNVs and single gene defects among the three subgroups provides a new vision on the choice of appropriate clinical diagnostic test method in different ID patients.

**Methods**

**Subjects and controls**

The patients were recruited from the Department of Neurology, Affiliated Children’s Hospital of the Capital Institute of Pediatrics between 2016.01 and 2018.12. Written informed consent was obtained from all parents Ethical No.SHERLL2013046. All enrolled 95 patients (30 females and 65 males) had clinical manifestations of ID (Supplemental
The patients were subdivided into three subgroups according to whether they had MCA or ASD except ID (ID with MCA, 52 patients; ID with ASD, 14 patients; and other IDs, 29 patients).

**Whole genome low-coverage sequencing**

Genomic DNA was extracted from the peripheral blood sample using QIAamp DNA Mini Kit (QIAGEN). The DNA was quantified by using Nanodrop 2000 (Thermal Fisher Scientific, DE). The DNA was sheared to size by approximately 300 bp with a Covaris S2 sonicator according to the Illumina TruSeq DNA protocol. The fragments were then end-repaired and A-tailed in preparation for ligation to adapters. The ligation product was amplified by PCR by using universal primers. The enriched libraries were sequenced on an Illumina HiSeq 2000 sequencer (Illumina, San Diego, CA, USA) for paired-end reads of 150 bp. High-quality paired-end reads were aligned to the NCBI human reference genome hg19 by using Short Oligonucleotide Analysis Package (SOAP) aligner software (SOAP2.21; soap.genomics.org.cn/soapsnp.html). CNV detection was performed according to a three-step method reported previously. 1~2X coverage was chosen and 3~6Gb are sequenced per patient (Supplemental Data Table S2). The resolution is about 100~200kb in this condition referred to previous literatures.

**The CNV Interpretation**

To evaluate the pathogenicity of CNVs, the criteria of ACMG guidelines for CNVs in 2011 were mainly referred. A CNV was evaluated as a pathogenic CNV when it complies with one of the following criteria: when 1) it overlaps with at least 50% of the critical region of known genomic disorders; 2) it contains known ID disease-causing genes and the variation
of dosage effect conform to the genetic pattern; and 3) the size of the CNV is larger than 1Mb and the region has been recorded in patients with ID in DECIPHER. CNV was evaluated as an uncertain clinical significant CNV when it was documented in <0.1% of the population and did not meet the criteria of pathogenic CNVs.

**Medical exome sequencing**

Genomic DNA was extracted from the peripheral blood by using QIAamp DNA Mini Kit (QIAGEN). The DNA was quantified with Nanodrop 2000 (Thermal Fisher Scientific, DE). A minimum of 3 ug DNA was used for the indexed Illumina libraries according to the manufacturer’s protocol. DNA fragments with sizes ranging from 350 bp to 450 bp and those including the adapter sequences were selected for DNA libraries. Next, over 4000 genes (Supplemental Data Table S1) associated with monogenic disorders were selected by a gene capture strategy by using a custom enrichment kit (IDT, Coralville, Iowa, USA). The enriched libraries were sequenced on an Illumina HiSeq XTen sequencer (Illumina, San Diego, CA, USA) for paired-end reads of 150 bp. 100~200X mean coverage was chosen and 4~6Gb are sequenced per patient (Supplemental Data Table S3). Following sequencing, the raw image files were processed by using Bcl2Fastq software (Bcl2Fastq 2.18.0.12, Illumina, Inc.) for base calling and raw data generation. Low-quality variations were filtered out by a quality score of ≥20. SOAP aligner software (SOAP2.21; soap.genomics.org.cn/soapsnp.html) was then used to align the clean reads with the reference human genome (hg19). Polymerase chain reaction (PCR) duplicates were removed by using the Picard program. Subsequently, single nucleotide polymorphisms (SNPs) were determined by using the SOAPsnp program, reads were realigned by using Burrows-Wheeler Aligner software 0.7.15, and the insertions and deletions (InDels) were detected by using Genome Analysis Toolkit software 3.7. The identified SNPs and InDels
were annotated by using the Exome-assistant program. The pathogenicity of SNPs and InDels was evaluated according to ACMG guidelines.

Results

Rare copy number variations

To explore whether the ID was caused by rare CNVs, a whole genome low-coverage sequencing was used as the first-tier diagnostic test in 95 patients diagnosed with ID. As many point variations were single nucleotide polymorphism, several genome-wide CNVs showed polymorphism\textsuperscript{29}. Therefore, the rare CNVs with population frequency <0.1% were mainly focused. A total of 312 rare CNVs were identified in 95 patients including 175 gains and 137 losses. The size of the rare CNVs in patients ranged from 102 Kb to 155 Mb. According to the criteria of evaluation of the pathogenicity of rare CNVs in the Methods section, 19 pathogenic CNVs were identified in 16 patients (Figure 1a, Table 1). Fourteen CNVs (8 losses and 6 gains) were associated with known genomic disorders including 2q31.1 microdeletion syndrome, 7q31 microdeletion syndrome, 7p duplication syndrome, 9p duplication syndrome, 10q deletion syndrome, Jacobsen syndrome, Prader-Willi and Angelman syndrome, Renal cysts and diabetes (RCAD), 17p11.2 duplication syndrome, Smith-Mageni syndrome, 2q37 deletion syndrome, and Klinefelter syndrome (Table 1, Patient 1, 3-16). We also identified a 4.43Mb duplication at 5q35 containing the dosage sensitivity gene NSD1 that contributed to Sotos syndrome (Table 1, Patient 2). Four CNVs were larger than 1Mb and the region has been recorded in patients with ID in DECIPHER included a 2.92Mb deletion at 2q24.1, a 6.36Mb duplication at 19q13.42-13.43, a 15.04Mb duplication at 20p12.1-p13 and a 11.66 Mb deletion at 9p23-24.3 (Table 1, Patient 1, 12, 13, 14).
Two concurrent pathogenic CNVs in one patient

Cytogenetic imbalances are the most frequently identified causes of ID/MCA. The diagnostic rate of submicroscopic terminal rearrangements was about 6%, with a range of 2% to 29% due to different resolutions of techniques, inclusion criteria and sample sizes\textsuperscript{30}. Meanwhile, a frequency of two or more chromosomal aberrations was estimated to be 2–4% by previous studies\textsuperscript{31}. Two concurrent pathogenic CNVs were identified by whole genome low-coverage sequencing in 4 of the 16 CNV-positive patients (Table 1, Patient 1, 12, 13, 14), accounting for 25% in CNV-positive patients and 4.2% in the total 95 patients, respectively. The rate was generally consistent with that reported by a previous study based on a large-scale cohort of ID. We identified a 2.06Mb deletion in 2q31.1 region in 1 patient with a concurrent 2.92Mb deletion in 2q24.1 region. Patient 14 has a 11.66Mb deletion in the 7p22.3-p31.2 region and a concurrent 14Mb duplication in 9p23-24.3 region. Two patients, 12 and 13, carried a deletion in the 2q37 region, with a concurrent 6.36Mb duplication at 19q13.41 and 15.04Mb duplication at 20p12.1, respectively (Figure 3a). Both large deletion and duplication in subtelomeric region were detected in patients 11, 12, and 13, suggesting it to be probably caused by subtelomeric rearrangement. These findings highlighted the importance of screening apparently “balanced” subtelomeric rearrangements inherited from a phenotypically normal parent in patients with ID.

Variable clinical phenotypes of CNVs

In patient 9, a 1.49Mb deletion was found at 17q12, and was known to be associated with renal cysts and diabetes (RCAD), (Figure 3d). The most consistent clinical feature of RCAD is the presence of renal cysts and most of the affected subjects also had early-onset diabetes. However, Patient 9 in our study showed no signs of renal abnormalities and
diabetes until 4 years and 7 months old. Patient 9 showed clinical manifestations of asophia, and brain dysplasia was revealed by MRI instead from the age of 2 years. Interestingly, the recurrent 17q12 deletion has also been documented with diverse range of phenotypes associated with neurodevelopment, such as autism spectrum disorder and attention-deficit hyperactivity disorder, except RCAD. The 1.49Mb deletion at 17q12 contains 14 protein-coding genes in addition to HNF1B, which was considered as the etiology of RCAD. This suggested that there might be other protein-coding genes at 17q12 that contributed for the neurodevelopmental disorder, such as LHX1 and ACACA, and were inferred as the genes involved in neurodevelopmental syndrome.

Both patients 12 and 13 carried a deletion in 2q37 region and presented facial dysmorphism and global developmental delay. Some other different clinical symptoms were still observed in 2 patients. In addition, a neonatal hypoglycemia was observed in patient 12 and a reproductive system abnormality was observed in patient 13 at 1 years and 4 months. The variable clinical phenotypes between patients 12 and 13 were probably due to different deletion sizes (7.92Mb vs 2.3Mb) at 2q37 (Figure 3c). The other explanation for these was the additional duplication regions identified in 2 patients, respectively, as both 19q13.42q13.43 and 20p12.1-13 duplication (Figure 3a, 3b) have been reported with ID and craniofacial dysmorphisms by few case reports.

**Single-nucleotide variations**

The medical exome sequencing of 79 patients showed negative results by whole genome low-coverage sequencing. Then our analysis on the genes that are previously associated with ID or related neurodevelopmental disorders was focused. We identified 11 predicted pathogenic variants in 9/79 patients (Table 2). These variants included de novo (n=5),
compounded heterozygous (n=2), homozygous (n=1) and hemizygous variants (n=1). No genes were found to be mutated in two families, suggesting a low incidence of hot spots underlying in ID in Han Chinese population. Compared with the phenotypes of 11 variants in genes that are known to cause ID, most of the patients reported similar phenotypic entity. However, some deviations do occur, which might result due to the clinical diversity or by the age of onset. The clinical features of patient 25 included learning disability, hyperactivity, epilepsy, congenital heart disease and moderate ID. Medical exome sequencing showed a novel homozygous variant in AP4M1, causing a replacement of serine by phenylalanine at position 9 of the protein. AP4M1 was reported to be the genetic cause of Spastic paraplegia 50, which is characterized by severe intellectual disability, spasticity, developmental delay, facial features and microcephaly on postnatal onset. Patient 25 did not present with spasticity or dystonia until 7 years but with congenital heart disease instead, and has not been reported in previous patients with AP4M1 mutations. Our findings might expand AP4M1 phenotypes to cardiac abnormality.

Novel mutations identified by medical exome sequencing

We identified 7 novel pathogenic or likely pathogenic single-nucleotide variations in 6/9 ID patients, which included HPRT1 c.419delG, AARS2 c.806G>A and c.374T>C, KIF1A c.1262A>C, STXBP1 c.536T>G, B4GALT7 c.319G>C and AP4M1 c.26C>T (Table 2, Patients 17, 18, 21, 23, 24, and 25). In patient 17, a hemizygous 1-bp deletion (419delG) that caused a frameshift was found at exon 6. The mutation resulted in premature termination of translation of HPRT mRNA (p.Gly140Alafs*26), which was not reported previously but a pathogenic missense amino acid change occurred at the same position. A novel compounded heterozygous mutation, c.806G>A and c.374T>C, was revealed in AARS2 gene
in patient 18 who presented with psychomotor retardation, hypotonia, cerebellar atrophy and white matter abnormal signal. Both the missense mutations substituted highly conserved amino acid residues in aminoacylation domain, and were predicted to be damaging by SIFT, Polyphen2 and Mutation Taster. Two novel de novo mutations, **KIF1A** c.1262A>C and **STXBP1** c.536T>G were identified in patients 21 and 23, respectively. In patient 24, a novel missense mutation c.319G>C caused heterozygous mutation of a novel compound with a reported pathogenic missense mutation at p.Leu205Pro in the **B4GALT7** gene. In addition, a homozygous mutation was also identified in **AP4M1**, which is a disease-caused gene of Spastic paraplegia 50, encoding for the m subunit of adaptor protein complex-4 in patient 25^39.

**Diagnostic yield and variation types in different ID subgroups**

The diagnostic yields informed in previous studies of large-scale cohort of IDs were mostly acquired by a mixed ID group that encompassed other phenotypes including MCA and ASD. Previous studies reported an increased diagnostic yield of CNVs in ID patients with MCA^41, 42. To investigate whether the diagnostic yields showed differences among different ID subgroups, ID patients were subdivided into 3 different subgroups, ID with MCA, ID with ASD and other IDs (ID without MCA and ASD). Diagnostic yield varied according to the ID subgroups. The highest diagnostic rate was observed in other ID subgroup (34.48%), followed by ID with MCA (26.92%). The lowest diagnostic rate was observed in ID with ASD subgroup (7.14%), (Figure 2b). Except diagnostic yield, the variation types were also varied by ID subgroups. In ID, the MCA CNVs accounted for 23.07% and single gene defects accounted for 3.85% (Figure 2b). In other IDs, CNVs accounted for 10.34% and single gene defects accounted for 24.14% (Figure 2b). The rate of ID CNVs in MCA
subgroup was significantly higher than the other two subgroups. Likewise, single gene defects contributed to a larger proportion in other ID subgroup. The varied diagnostic yields and variation types suggested that the clinical detection scheme of ID might be changed along with different ID subgroups that patients belong to. These result provided us a new perspective on the appropriate clinical detection scheme for different ID subgroups.

Discussion
A total of 19 pathogenic CNVs in 16.84% (16/95) of Han Chinese ID patients were found by whole genome low-coverage sequencing. This rate was mainly consistent with a previous study based on European and Chinese populations by CMA\textsuperscript{43-45}. The remaining subtle difference was likely to be due to other factors, such as the technology platform or the criteria of pathogenic CNVs. These results validated whole genome low-coverage sequencing as an alternative effective diagnostic method for genome-wide CNV detection in routine clinical application\textsuperscript{46}. In the clinical diagnosis of pathogenic CNVs, whole genome low-coverage sequencing approach shows equivalent effectiveness and advantages compared with CMA by recent studies. Meanwhile whole genome low-coverage sequencing is a more sensitive, operable, rapid and lower cost method rather than conventional CMA. The only limitation of whole genome low-coverage sequencing was unable to detect uniparental disomy compared with CMA. With the development of NGS, whole genome low-coverage sequencing will have a better application prospect compared with conventional CMA.

Due to relatively little realization of the CNVs when compared with single-nucleotide variations, the database of pathogenic CNVs and well-known genome-wide disorders still remained poor. Therefore, there were several CNVs judged as variants with uncertain
significance. Among CNVs that are associated with well-known genome-wide disorders in patients, a 1.49 Mb deletion at 17p11.2 locus was noticed which was reported by RCAD. The 17p11.2 deletion was also been found in a patient characterized by ID, cerebral dysplasia, asophia, big head circumference and slightly lower limb muscle tone. The patient who carried 17p11.2 deletion in our study also presented with some neurological symptoms such as asophia and brain dysplasia. These findings implied the correlation between the 17p11.2 region and neurodevelopmental disorders. Patient 12 and 13 both carried 2q37 deletion. They both had ID and facial dysmorphism, but also showed specific phenotypes themselves. The diversity in phenotypes might be explained by different boundaries of deletion regions encompassing different genes among different patients or additional modified variations in other areas such as epigenetics\textsuperscript{35, 47, 48}. This indicated that the study of pathogenic mechanisms of CNVs were more complicated than SNVs. The genetic contributors were yet to be revealed for many pathogenic CNVs and even for some well-known disorders such as 16p11.2 microdeletion syndrome.

In addition to the CNVs, we applied medical exome sequencing to screen single-nucleotide variations in patients where no pathogenic CNVs were detected. Diagnostic yield showed improvement as the resolution of cytogenetic testing in patients with developmental disabilities has been evolved. We identified 11 pathogenic mutations among 9 genes in 9 unrelated patients. No two patients carried the same disease-causing gene. This indicated a strong genetic heterogeneity and no hot spot genes for ID in Han Chinese population. The disease type caused by the genes varied from syndromic and non-syndromic ID to epileptic encephalopathy, leukoencephalopathy and spastic paraplegia, showing a strong clinical heterogeneity. This might explain the low detection rate (~10%) of whole exome sequencing for ID patients as previous study\textsuperscript{49}.
The patients with multiple congenital anomalies may have a higher detection rate of CNVs than other ID patients. Therefore, the ID patients were subdivided into 3 subgroups, including ID with MCA, ID with ASD and other IDs. Diagnostic yield and variation types were different in different ID subgroups. ID with MCA subgroup showed a higher detection rate of CNVs. ID with ASD showed the lowest detection rate in both CNVs and single-nucleotide variations. More single-nucleotide variations were identified in other ID subgroups. This suggested that the proportion of variation types were significantly different in different subgroups.

Seventy (73.69%) patients with ID have still not received a molecular diagnosis in our study. This was consistent with the results conducted by previous studies where a majority of ID cases remain undiagnosed after genetic detection due to different technology platforms such as microarray, ID panel and whole exome sequencing. Such a high negative rate might be caused by lots of factors. The singleton-approach was one the of the main reasons as many studies had showed trio-approach can further improve the detection rate of de-nove mutations in ID cohorts. The trio-approach can also break the limit of lacking parental genotypes. Other factors, caused the high negative rate, were the limitation in resolution of CNVs, the detection area of sequencing and the sensitivity of low frequency mosaic mutation. The resolution of CNVs through whole genome low-coverage sequencing is 100kb, which meant that the smaller CNVs cannot be detected such as single-exon and intra-exonic deletions. Medical exome sequencing was applied due to its cost-effective detection scheme for single-nucleotide variations. Meanwhile, the novel genes might be omitted and the mutations outside the coding regions remain to be further explored. Recent studies have revealed that the low frequency mosaic mutation might play a role partly in neurological and neuromuscular
disorders\textsuperscript{6, 50}.

Conclusions

Although there were many genetic studies that focused on uncovering the etiology of ID, the information of ID in Han Chinese population was still little. Most of the previous studies were based on CMA. We herein performed whole genome low-coverage sequencing as the first-tier diagnostic test and medical exome sequencing as the second-tier diagnostic test for those patients with negative results of CNVs. The detection rates of CNVs and SNVs were consistent with those of the previous studies at home and abroad. This detected scheme was considered appropriate and cost-effective in terms of the consequences. The significantly different proportions of variation types prompted us to select the appropriate detection scheme for different ID subgroups based on the results of this study. However, the sample size of this study is small, and the conclusion might be influenced by sample bias. This study should be replicated with a larger study to eliminate the influence of patient’s source selection. As a result, a larger and multicenter study should be conducted. Our study provided a new perspective regarding the study of clinical molecular diagnosis for ID patients in Han Chinese population.

Declarations

Ethics approval and consent to participate

This study was reviewed and approved by the ethical committee of Affiliated Children’s Hospital of Capital Institute of Pediatrics, China and Capital Characteristic Clinical Application Research (Z131107002213159). Informed consent was obtained from the patients, and for those patients under 18 years old, consent was obtained from the parents on their behalf.

Availability of data and materials
To ensure patient confidentiality, data containing potentially identifiable information was not shared. All data generated or analysed during this study, without identifiable information, is available from the corresponding author upon reasonable request.

**Competing interests**
The authors declare that they have no competing financial interests.

**Funding**
Funding for this project was from the Beijing Health System High Level Health Technical Personnel Training Plan [2015-3-084]. The funding organizations played no role in the design of the study, collection, analysis and interpretation of data and in writing the manuscript.

**Authors’ contributions**
Jun Wang and Yan Wang performed the analysis, prepared the figures and tables, wrote and reviewed the manuscript. WangYang Chen and Min Sheng performed the whole genome low-coverage sequencing and medical exome sequencing. Liwen Wang and Yan Wang recruited the patients and collected clinical information. All authors have read and approved the manuscript.

**Acknowledgements**
We acknowledge all the patients and families. We thank Liwen Wang (Department of Neurology, Affiliated Children’s Hospital of Capital Institute of Pediatrics) and WangYang Chen (Kaiumph Medical Diagnostics Co, Ltd) for constructive comments and discussion.

**Authors’ information**
Jun Wang¹, ¹Department of Neurology, Affiliated Children’s Hospital of Capital Institute of Pediatrics, Beijing 100020, China

Yan Wang¹, ¹Department of Neurology, Affiliated Children’s Hospital of Capital Institute of
Liwen Wang\textsuperscript{1,} \textsuperscript{1}Department of Neurology, Affiliated Children’s Hospital of Capital Institute of Pediatrics, Beijing 100020, China

Wang Yang Chen\textsuperscript{2,} \textsuperscript{2}Kaiumph Medical Diagnostics Co, Ltd, Beijing 100102, China

Min Sheng\textsuperscript{2,} \textsuperscript{2}Kaiumph Medical Diagnostics Co, Ltd, Beijing 100102, China

Jun Wang\textsuperscript{1*}, \textsuperscript{1}Department of Neurology, Affiliated Children’s Hospital of Capital Institute of Pediatrics, Beijing 100020, China; email: jam1080800004@163.com [tel: 13621142651

References

[1] Mehregan H, Najmabadi H, Kahrizi K. Genetic Studies in Intellectual Disability and Behavioral Impairment. Arch Iran Med. 2016;19:363-75.

[2] Karam SM, Barros AJ, Matijasevich A, Dos Santos IS, Anselmi L, Barros F, et al. Intellectual Disability in a Birth Cohort: Prevalence, Etiology, and Determinants at the Age of 4 Years. Public Health Genomics. 2016;19:290-7.

[3] Moeschler JB, Shevell M, Committee on G. Comprehensive evaluation of the child with intellectual disability or global developmental delays. Pediatrics. 2014;134:e903-18.

[4] Flore LA, Milunsky JM. Updates in the genetic evaluation of the child with global developmental delay or intellectual disability. Semin Pediatr Neurol. 2012;19:173-80.

[5] Ropers HH. Genetics of intellectual disability. Curr Opin Genet Dev. 2008;18:241-50.

[6] Gilissen C, Hehir-Kwa JY, Thung DT, van de Vorst M, van Bon BW, Willemsen MH, et al. Genome sequencing identifies major causes of severe intellectual disability. Nature. 2014;511:344-7.

[7] Grozeva D, Carss K, Spasic-Boskovic O, Tejada MI, Gecz J, Shaw M, et al. Targeted Next-Generation Sequencing Analysis of 1,000 Individuals with Intellectual Disability. Hum Mutat. 2015;36:1197-204.
[8] Chiurazzi P, Pirozzi F. Advances in understanding - genetic basis of intellectual disability. F1000Res. 2016;5.

[9] Martinez F, Caro-Llopis A, Rosello M, Oltra S, Mayo S, Monfort S, et al. High diagnostic yield of syndromic intellectual disability by targeted next-generation sequencing. J Med Genet. 2017;54:87-92.

[10] Rauch A, Hoyer J, Guth S, Zweier C, Kraus C, Becker C, et al. Diagnostic yield of various genetic approaches in patients with unexplained developmental delay or mental retardation. Am J Med Genet A. 2006;140:2063-74.

[11] Heslop P, Blair PS, Fleming P, Hoghton M, Marriott A, Russ L. The Confidential Inquiry into premature deaths of people with intellectual disabilities in the UK: a population-based study. Lancet. 2014;383:889-95.

[12] Vissers LE, Gilissen C, Veltman JA. Genetic studies in intellectual disability and related disorders. Nat Rev Genet. 2016;17:9-18.

[13] Mefford HC, Batshaw ML, Hoffman EP. Genomics, intellectual disability, and autism. N Engl J Med. 2012;366:733-43.

[14] Battaglia A, Doccini V, Bernardini L, Novelli A, Loddo S, Capalbo A, et al. Confirmation of chromosomal microarray as a first-tier clinical diagnostic test for individuals with developmental delay, intellectual disability, autism spectrum disorders and dysmorphic features. Eur J Paediatr Neurol. 2013;17:589-99.

[15] Zilina O, Teek R, Tammur P, Kuuse K, Yakoreva M, Vaidla E, et al. Chromosomal microarray analysis as a first-tier clinical diagnostic test: Estonian experience. Mol Genet Genomic Med. 2014;2:166-75.

[16] Chong WW, Lo IF, Lam ST, Wang CC, Luk HM, Leung TY, et al. Performance of chromosomal microarray for patients with intellectual disabilities/developmental delay, autism, and multiple congenital anomalies in a Chinese cohort. Mol Cytogenet. 2014;7:34.
[17] Capobianco S, Lava SA, Bianchetti MG, Martinet D, Belfiore M, Ramelli GP, et al. Chromosomal microarray among children with intellectual disability: a useful diagnostic tool for the clinical geneticist. Dev Med Child Neurol. 2014;56:290.

[18] Fan Y, Wu Y, Wang L, Wang Y, Gong Z, Qiu W, et al. Chromosomal microarray analysis in developmental delay and intellectual disability with comorbid conditions. BMC Med Genomics. 2018;11:49.

[19] Miller DT, Adam MP, Aradhya S, Biesecker LG, Brothman AR, Carter NP, et al. Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. Am J Hum Genet. 2010;86:749-64.

[20] Harripaul R, Vasli N, Mikhailov A, Rafiq MA, Mittal K, Windpassinger C, et al. Mapping autosomal recessive intellectual disability: combined microarray and exome sequencing identifies 26 novel candidate genes in 192 consanguineous families. Mol Psychiatry. 2018;23:973-84.

[21] Di Gregorio E, Riberi E, Belligni EF, Biamino E, Spielmann M, Ala U, et al. Copy number variants analysis in a cohort of isolated and syndromic developmental delay/intellectual disability reveals novel genomic disorders, position effects and candidate disease genes. Clin Genet. 2017;92:415-22.

[22] Pengelly RJ, Greville-Heygate S, Schmidt S, Seaby EG, Jabalameli MR, Mehta SG, et al. Mutations specific to the Rac-GEF domain of TRIO cause intellectual disability and microcephaly. J Med Genet. 2016;53:735-42.

[23] Johansen A, Rosti RO, Musaev D, Sticca E, Harripaul R, Zaki M, et al. Mutations in MBOAT7, Encoding Lysophosphatidylinositol Acyltransferase I, Lead to Intellectual Disability Accompanied by Epilepsy and Autistic Features. Am J Hum Genet. 2016;99:912-6.
[24] Stessman HA, Xiong B, Coe BP, Wang T, Hoekzema K, Fenckova M, et al. Targeted sequencing identifies 91 neurodevelopmental-disorder risk genes with autism and developmental-disability biases. Nat Genet. 2017;49:515-26.

[25] Li R, Yu C, Li Y, Lam TW, Yiu SM, Kristiansen K, et al. SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics. 2009;25:1966-7.

[26] Xie C, Tammi MT. CNV-seq, a new method to detect copy number variation using high-throughput sequencing. BMC Bioinformatics. 2009;10:80.

[27] Dong Z, Zhang J, Hu P, Chen H, Xu J, Tian Q, et al. Low-pass whole-genome sequencing in clinical cytogenetics: a validated approach. Genet Med. 2016;18:940-8.

[28] Kearney HM, Thorland EC, Brown KK, Quintero-Rivera F, South ST, Working Group of the American College of Medical Genetics Laboratory Quality Assurance C. American College of Medical Genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants. Genet Med. 2011;13:680-5.

[29] Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, et al. Global variation in copy number in the human genome. Nature. 2006;444:444-54.

[30] Santa Maria L, Faundes V, Curotto B, Morales P, Morales K, Aliaga S, et al. Comparison of two subtelomeric assays for the screening of chromosomal rearrangements: analysis of 383 patients, literature review and further recommendations. J Appl Genet. 2016;57:63-9.

[31] Kibe T, Mori Y, Okanishi T, Shimojima K, Yokochi K, Yamamoto T. Two concurrent chromosomal aberrations involving interstitial deletion in 1q24.2q25.2 and inverted duplication and deletion in 10q26 in a patient with stroke associated with antithrombin deficiency and a patent foramen ovale. Am J Med Genet A. 2011;155A:215-20.

[32] Mitchell E, Douglas A, Kjaegaard S, Callewaert B, Vanlander A, Janssens S, et al. Recurrent duplications of 17q12 associated with variable phenotypes. Am J Med Genet A. 2015;167A:3038-45.
[33] Moreno-De-Luca D, Consortium S, Mulle JG, Simons Simplex Collection Genetics C, Kaminsky EB, Sanders SJ, et al. Deletion 17q12 is a recurrent copy number variant that confers high risk of autism and schizophrenia. Am J Hum Genet. 2010;87:618-30.

[34] Bockenhauer D, Jaureguiberry G. HNF1B-associated clinical phenotypes: the kidney and beyond. Pediatr Nephrol. 2016;31:707-14.

[35] Clissold RL, Ashfield B, Burrage J, Hannon E, Bingham C, Mill J, et al. Genome-wide methylomic analysis in individuals with HNF1B intragenic mutation and 17q12 microdeletion. Clin Epigenetics. 2018;10:97.

[36] Nagamani SC, Erez A, Shen J, Li C, Roeder E, Cox S, et al. Clinical spectrum associated with recurrent genomic rearrangements in chromosome 17q12. Eur J Hum Genet. 2010;18:278-84.

[37] Hardies K, Weckhuysen S, Peeters E, Holmgren P, Van Esch H, De Jonghe P, et al. Duplications of 17q12 can cause familial fever-related epilepsy syndromes. Neurology. 2013;81:1434-40.

[38] Kwon SS, Kim J, Shin S, Lee ST, Lee KA, Choi JR. Deletion of 20p13 and Duplication of 20p13p12.3 in a Patient with Delayed Speech and Development. Ann Lab Med. 2018;38:77-9.

[39] Resta N, De Cosmo L, Susca FC, Capodiferro D, Nardone AM, Pastorivo D, et al. De novo unbalanced translocation leading to monosomy 9p24.3p24.1 and trisomy 19q13.42q13.43 characterized by microarray-based comparative genomic hybridization in a child with partial cortical dysplasia and craniofacial dysmorphisms without trigonocephaly. Am J Med Genet A. 2013;161A:632-6.

[40] Fujimori S, Tagaya T, Yamaoka N, Kamatani N, Akaoka I. Molecular analysis of hypoxanthine-guanine phosphoribosyltransferase deficiency in Japanese patients. Adv Exp Med Biol. 1991;309B:101-4.
[41] Vulto-van Silfhout AT, Hehir-Kwa JY, van Bon BW, Schuurs-Hoeijmakers JH, Meader S, Hellebrekers CJ, et al. Clinical significance of de novo and inherited copy-number variation. Hum Mutat. 2013;34:1679-87.

[42] Cooper GM, Coe BP, Girirajan S, Rosenfeld JA, Vu TH, Baker C, et al. A copy number variation morbidity map of developmental delay. Nat Genet. 2011;43:838-46.

[43] Vissers LE, de Vries BB, Osoegawa K, Janssen IM, Feuth T, Choy CO, et al. Array-based comparative genomic hybridization for the genomewide detection of submicroscopic chromosomal abnormalities. Am J Hum Genet. 2003;73:1261-70.

[44] Wagenstaller J, Spranger S, Lorenz-Depiereux B, Kazmierczak B, Nathrath M, Wahl D, et al. Copy-number variations measured by single-nucleotide-polymorphism oligonucleotide arrays in patients with mental retardation. Am J Hum Genet. 2007;81:768-79.

[45] Wang B, Ji T, Zhou X, Wang J, Wang X, Wang J, et al. CNV analysis in Chinese children of mental retardation highlights a sex differentiation in parental contribution to de novo and inherited mutational burdens. Sci Rep. 2016;6:25954.

[46] Dong Z, Jiang L, Yang C, Hu H, Wang X, Chen H, et al. A robust approach for blind detection of balanced chromosomal rearrangements with whole-genome low-coverage sequencing. Hum Mutat. 2014;35:625-36.

[47] Franklin TB, Mansuy IM. The involvement of epigenetic defects in mental retardation. Neurobiol Learn Mem. 2011;96:61-7.

[48] Kramer JM, van Bokhoven H. Genetic and epigenetic defects in mental retardation. Int J Biochem Cell Biol. 2009;41:96-107.

[49] Ropers HH. Genetics of early onset cognitive impairment. Annu Rev Genomics Hum Genet. 2010;11:161-87.

[50] Zahir FR, Mwenifumbo JC, Chun HE, Lim EL, Van Karnebeek CDM, Couse M, et al.
Comprehensive whole genome sequence analyses yields novel genetic and structural insights for Intellectual Disability. BMC Genomics. 2017;18:403.

Tables

Table 1. Pathogenic copy number variations identified by whole genome low-coverage sequencing.

| Patient | CNV variation | Region size | Known genomic disease |
|---------|---------------|-------------|-----------------------|
| 1       | del(2q24.1)(155.48Mb-158.4Mb)*1 del(2q31.1-q31.2)(176.16Mb-178.22Mb)*1 | 2.92Mb 2.06Mb | 2q31.1 microdeletion syndrome 2:15:lança 2:15:5 |
| 2       | Dup(5q35.2-q35.3)(175.74Mb-180.08Mb)*3 | 4.43Mb | Sotos syndrome(NDS1) 5:1 H |
| 3       | del(7q31.1-q31.33)(113.9Mb-125.52Mb)*1 | 11.62Mb | 7q31 microdeletion syndrome (? FOXP2, DOCK4, MET, and WNT2) 7:114 lang 7:112 t 7:1:7 |
| 4       | dup(7p14.3-p22.3)(0.1Mb-30.82Mb)*3 | 30.82Mb | 7p duplicationsyndrome 7:102: Constijsevere Narrow 7:503: Diaste Microg 7:109c develop ears, & 9:991c Aplasie Aplasie Aplasie Aplasie Diaste: Hypert 9:826f delay 10:12f from a 10:127 strabis impaim Sleep-t 10:125 abnor urachu rearrat 11:12- count set e |
| 5       | dup(9p22.2-p23)(10.06Mb-16.78Mb)*3 | 6.72Mb | 9p duplication syndrome 9:991c Aplasie Aplasie Aplasie Aplasie Diaste: Hypert 9:826f delay 10:12f from a 10:127 strabis impaim Sleep-t 10:125 abnor urachu rearrat 11:12- count set e |
| 6       | del(10q26.13-q26.3)(126.62Mb-135.52Mb) *1 | 8.9Mb | 10q deletion syndrome 10q deletion syndrome 10q deletion syndrome 10q deletion syndrome |
| 7       | del(11q24-q25)(124.55Mb-134.94Mb)*1 | 10.39Mb | Jacobsen syndrome 11:12- count set e |
| 8       | del(15q11.2-q13.2)(23.62Mb-30.38Mb)*1 | 6.76Mb | Prader-Willi and Angelman syndrome 15:2 |
|   | Chromosome | Start/End | Length | Description                                      |
|---|------------|-----------|--------|--------------------------------------------------|
| 9 | chr17: 34918482-36408851 | 1.49Mb | Renal cysts and diabetes (RCAD) |
| 10 | dup(17p11.2-p12)(14.76Mb-19.48Mb) | 4.72Mb | 17p11.2 duplication syndrome/CMT1A |
| 11 | del(17p11.2)(16.6Mb-20.5Mb) | 3.9Mb | Smith-Mageni syndrome |
| 12 | del(2q37.2-37.3)(235.16Mb–243.08Mb) | 7.92 Mb | 2q37 deletion syndrome |
|    | dup(19q13.42-q13.43)(52.76Mb–59.12Mb) | 6.36 Mb | |
| 13 | del(2q37.3)(240.70Mb-243Mb) | 2.30Mb | 2q37 deletion syndrome |
|    | dup(20p12.1-p13)(0.1Mb-15.14Mb) | 15.04Mb | |
| 14 | del(9p24.3-p23)(1Mb-12.66Mb) | 11.66Mb | 9p deletion syndrome |
|    | dup(7p22.3-p21.2)(1Mb-15Mb) | 14Mb | 7p duplication syndrome |
| 15 | 47XXY | X chromosome | Klinefelter syndrome |
| 16 | 47XXY | X chromosome | Klinefelter syndrome |

Table 2 Pathogenic gene mutations identified by medical exome sequencing.
| Patient | Gene     | Transcript | Nucleotide change | Amino acid change |
|---------|----------|------------|------------------|------------------|
| 17      | HPRT1    | NM_000194  | c.419delG        | p.Gly140Alafs*   |
| 18      | AARS2    | NM_020745  | c.806 G>A        | p.Gly269Asp      |
|         |          |            | c.374T>C         | p.Leu125Pro      |
| 19      | EEF1A2   | NM_001958  | c.796C>T         | p.Arg266Trp      |
| 20      | TCF4     | NM_001083962 | c.1153C >T    | p.Arg385Ter      |
| 21      | KIF1A    | NM_004321  | c.1262A >C       | p.His421Pro      |
| 22      | CACNA1A  | NM_001127221 | c.4991G >A    | p.Arg1664Glr     |
| 23      | STXBP1   | NM_003165  | c.536T>G         | p.Leu179Arg      |
| 24      | B4GALT7  | NM_007255  | c.319G>C         | p.Glu107Gln      |
|         |          |            | c.614T>C         | p.Leu205Pro      |
| 25      | AP4M1    | NM_004722  | c.26C>T          | p.Ser9Phe        |

**Figures**
Status of molecular diagnosis after CNV-seq and MES of 95 patients with intellectual disability. a) The detection rates of CNVs and SNVs. b) The proportion of different CNVs types. c) The inheritance patterns of diseases detected by MES.
Figure 2

Diagnostic yield in ID cohort (n=95) by subgroup distribution through whole genome low-coverage sequencing and medical exome sequencing.
Deletion

Patient 13

Deletion

Duplication
Figure 3
Map of CNVs associated with varying phenotypes of known genomic disorders. a) Two concurrent pathogenic CNVs, 2q37.1-q37.3 deletion and 19q13.42-q13.43 duplication, in patient 12. b) Two concurrent pathogenic CNVs, 2q37.3 deletion and 20p12.1-p13 duplication, in patient 13. c) Map of 2q37 deletions with included genes in 2 patients. and d) Map of 17q12 deletions with included genes in patient 9.

Supplementary Files
This is a list of supplementary files associated with the primary manuscript. Click to download.
Supplemental Data Table S4.xlsx
