Genome sequencing demonstrates high diagnostic yield in children with undiagnosed global developmental delay/intellectual disability: a prospective study

Yu Sun[1], Jing Peng[2], Desheng Liang[3], Xiantao Ye[4], Na Xu[5], Linlin Chen[4], Dan Yan[4], Huiwen Zhang[6], Bing Xiao[7], Wenjuan Qiu[1], yiping shen[8], Nan Pang[2], Yingdi Liu[9], Chen Liang[10], Zailong Qin[8], Jingsi Luo[8], Fei Chen[8], Jing-Min Wang[11], Zhixin Zhang[12], Haiyan Wei[13], Juan Du[14], Huifang Yan[15], Ruoyu Duan[15], Junyu Wang[15], Yu Zhang[15], Shixiu Liao[16], Kun Sun[1], Lingqian Wu[9], and Yongguo Yu[6]

[1]Xinhua Hospital Affiliated to Shanghai Jiaotong University School of Medicine
[2]Xiangya Hospital Central South University
[3]China Central South University
[4]Shanghai Jiaotong University School of Medicine Xinhua Hospital
[5]Department of Pediatric Endocrinology and Genetic Metabolism, Xinhua Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai Institute for Pediatric Research
[6]Department of Pediatric Endocrinology and Genetic Metabolism, Xinhua Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai Institute for Pediatric Research
[7]Center for Clinical Genetics, Xinhua Hospital Affiliated to Shanghai Jiao Tong University School of Medicine,
[8]Maternal and Child Health Hospital of Guangxi Zhuang Autonomous Region
[9]Central South University
[10]Jiangmen Maternity and Child Health Care Hospital
[11]Peking University
[12]China-Japan Friendship Hospital
[13]Department of Pediatric Endocrinology and Genetics, Children's Hospital Affiliated to Zhengzhou University
[14]Central South University Xiangya School of Medicine
[15]Peking University First Hospital
[16]Henan Provincial People's Hospital

September 25, 2021

Abstract

Genome sequencing(GS) has been applied in the diagnosis of global developmental delay(GDD)/intellectual disability(ID). However, the performance in those with inconclusive results from chromosomal microarray analysis(CMA) and exome sequencing(ES) is unknown. We recruited 100 pediatric GDD/ID patients from multiple sites in China from February 2018 to August 2020 for GS. Patients have received at least one genomic diagnostic test prior to enrollment. Reanalysis of CMA/ES data was

performed. The yield of GS was calculated and explanations for missed diagnoses by CMA/ES were investigated. Clinical utility was assessed by interviewing the parents by phone. The overall diagnostic yield of GS was 23%. Seven families could have been solved with reanalysis of ES data. 13 families were missed by previous CMA/ES due to improper method. Three remained unsolved after ES reanalysis due to allele dropout, complex variants missed by ES, and a CNV in untranslated regions. Follow-up of the diagnosed families revealed that nine families experienced changes in clinical management, including identification of targeted treatments, cessation of unnecessary treatment, and considerations for family planning. GS demonstrated high diagnostic yield and clinical utility in this cohort of undiagnosed GDD/ID patients, detecting a wide range of variant types of different sizes in a single workflow.

Genome sequencing demonstrates high diagnostic yield in children with undiagnosed global developmental delay/intellectual disability: a prospective study

Yu Sun[1,2,*]; Jing Peng[3,4,*]; Desheng Liang[5]; Xiantao Ye[1,2]; Na Xu[1,2]; Linlin Chen[1,2]; Dan Yan[1,2]; Huiwen Zhang[1,2]; Bing Xiao[1,2]; Wenjuan Qiu[1,2]; Yiping Shen[6]; Nan Pang[3,4]; Yingdi Liu[5]; Chen Liang[7]; Zailong Qin[6]; Jingsi Luo[6]; Fei Chen[6]; Jingmin Wang[8]; Zhixin Zhang[9]; Haiyan Wei[10]; Juan Du[11]; Huifang Yan[8]; Ruoyu Duan[8]; Junyu Wang[8]; Yu Zhang[8]; Shixiu Liao[12]; Kun Sun[13,#]; Lingqian Wu[5,#]; Yongguo Yu[1,2,#]

[1]Department of Pediatric Endocrinology and Genetic Metabolism, Xinhua Hospital, School of Medicine, Shanghai Jiao Tong University, 200092, Shanghai, China.

[2]Shanghai Institute for Pediatric Research, 200092, Shanghai, China.

[3] Department of Pediatrics, Xiangya Hospital, Central South University, 410008, Changsha, China.

[4] Hunan Intellectual and Developmental Disabilities Research Center, 410008, Changsha, China

[5]Center for Medical Genetics & Hunan Key Laboratory of Medical Genetics, School of Life Sciences, Central South University, 410078, Changsha, China

[6] Genetic and Metabolic Central Laboratory, Maternal and Child Health Hospital of Guangxi Zhuang Autonomous Region, 530003, Nanning, China.

[7] Medical Genetics Center, Jiangmen Maternity and Child Health Care Hospital, 529000, Jiangmen, China.

[8] Department of Pediatrics, Peking University First Hospital, 10034, Beijing, China.

[9] International Medical Services, China-Japan Friendship Hospital, 100029, Beijing, China.

[10] Department of Endocrinologic and Inherited Metabolic, Childen's Hospital affiliated to Zhengzhou University, 450018, Zhengzhou, China.

[11]Reproductive and Genetic Hospital of CITIC-Xiangya, 410078, Changsha, China.

[12] Provincial People's Hospital, Medical Genetics Institute of Henan Province, 450003, Zhengzhou, Henan Province, China.

[13] Center for Clinical Genetics, Xinhua Hospital, School of Medicine, Shanghai Jiao Tong University, 200092, Shanghai, China.

Yu Sun and Jing Peng should be considered joint first author. Kun Sun, Lingqian Wu and Yongguo Yu should be considered joint senior author.

Correspondence: Kun Sun, MD, PhD, Center for Clinical Genetics, Xinhua Hospital, School of Medicine, Shanghai Jiao Tong University, 200092, Shanghai, China (Tel: 86-21-25078480; Email: sunkun@xinhuamed.com.cn); Lingqian Wu, MD, PhD, Center for Medical Genetics & Hunan Key Laboratory of Medical Genetics, School of Life Sciences, Central South University, 410078, Changsha, China (Tel: 86-731-84805252 ; Email: wulingqiang@sklmg.edu.cn); Yongguo Yu, MD, PhD, Department of Pediatric Endocrinology and Genetic Metabolism, Shanghai Institute for Pediatric Research, Xinhua Hospital,

School of Medicine, Shanghai Jiao Tong University, 200092, Shanghai, China (Tel: 86-21-25076453; Email: yuyongguo@shsmu.edu.cn)

Funding information

Precision Medical Research of National Key Research and Development Program, Grant/Award Numbers: 2018YFC1002200, 2019YFC1005100, 2018YFC1002400

National Natural Science Foundation of China, Grant/Award Numbers: 81873724, 82070914, 81873671

Shanghai Shen Kang Hospital Development Center, Grant/Award Number: SHDC12017109

the Shanghai Science and Technology Commission, Grant/Award Number: 19140904500

Jiaotong University Cross Biomedical Engineering, Grant/Award Number: YG2017MS72

Foundation of Shanghai Municipal Health Commission, Grant/Award Numbers: shslczdzk05702

Municipal Education Commission-Gaofeng Clinical Medicine Grant Support, Grant/Award Numbers: 20191908

ABSTRACT

Genome sequencing(GS) has been applied in the diagnosis of global developmental delay(GDD)/intellectual disability(ID). However, the performance in those with inconclusive results from chromosomal microarray analysis(CMA) and exome sequencing(ES) is unknown. We recruited 100 pediatric GDD/ID patients from multiple sites in China from February 2018 to August 2020 for GS. Patients have received at least one genomic diagnostic test prior to enrollment. Reanalysis of CMA/ES data was performed. The yield of GS was calculated and explanations for missed diagnoses by CMA/ES were investigated. Clinical utility was assessed by interviewing the parents by phone. The overall diagnostic yield of GS was 23%. Seven families could have been solved with reanalysis of ES data. 13 families were missed by previous CMA/ES due to improper method. Three remained unsolved after ES reanalysis due to allele dropout, complex variants missed by ES, and a CNV in untranslated regions. Follow-up of the diagnosed families revealed that nine families experienced changes in clinical management, including identification of targeted treatments, cessation of unnecessary treatment, and considerations for family planning. GS demonstrated high diagnostic yield and clinical utility in this cohort of undiagnosed GDD/ID patients, detecting a wide range of variant types of different sizes in a single workflow.

TRAIL REGISTRATION

The study was registered in ClinicalTrials (No: NCT03424772).

KEYWORDS

genome sequencing, global developmental delay, intellectual disability, diagnostic yield, clinical utility

1 INTRODUCTION

Intellectual disability (ID), a condition characterized by significant limitations in general mental abilities and adaptive functioning that emerge during child development, is one of the most frequently reported impairments in children(Robert L. Schalock, 2021). The overall prevalence is around 1%(McGuire, Tian, Yeargin-Allsopp, Dowling, & Christensen, 2019). For younger patients in whom IQ scores are difficult to obtain, global developmental delay (GDD) is a condition assigned to patients who show significant delays in two or more developmental domains: fine/gross motor skills, speech/language, social/personal skills and daily living (Choo, Agarwal, How, & Yeleswarapu, 2019).

Early evaluation of children with suspected GDD/ID is valuable in determining whether intervention is required. Due to the high genotypic heterogeneity of GDD/ID, genomic diagnostic tests, such as chromosomal microarray analysis (CMA) and exome sequencing (ES) are frequently recommended to obtain a molecular genetic diagnosis. CMA focuses on copy number variations (CNV), and is recommended as the first-tier

method to diagnose ID or GDD by many associations(Miller et al., 2010; Subspecialty Group of Neurology & Project Expert Group of Childhood Neuropathy, 2018). The diagnostic yield of CMA in this population ranges from 12-20%(Miller et al., 2010; Moeschler, Shevell, & Committee on, 2014). Exome sequencing (ES) was initially used to detect small variants in regions of interest (ROI). A recent meta-analysis revealed that ES helped to solve 36% (range 8-90%) of cases in patients with suspected neurodevelopmental disorders (NDD), and has been recommended as a first-tier clinical diagnostic test for NDD(Srivastava et al., 2019). Periodic reanalysis of ES data can also provide additional diagnoses as new gene-disease associations are reported (Wenger, Guturu, Bernstein, & Bejerano, 2017; Xiao et al., 2018), and new algorithms are developed that enable variant calling beyond that of small variants(Fan et al., 2021; Sun et al., 2020).

Despite these recommendations, an important question remains: What is recommended for patients who remain undiagnosed after CMA, ES, or ES reanalysis? Studies have shown that many pathogenic variants are located outside of protein-coding regions of the genome which are not covered with ES or CMA. Genome sequencing (GS) can detect virtually all types of variation (e.g., chromosomal structural variation, trinucleotide repeats, mitochondrial variants) in the human genome without probe selection bias (Meynert, Ansari, FitzPatrick, & Taylor, 2014; Pang, Macdonald, Yuen, Hayes, & Scherer, 2014) and has been applied to clinical practice(French et al., 2019; Marshall et al., 2020; Wright, FitzPatrick, & Firth, 2018). For example, when used in the neonatal intensive care unit (NICU) setting, rapid GS demonstrates ultra-fast turn-around time, decreased rates of infant morbidity and reduced cost of hospitalization(Farnaes et al., 2018). Thus, adoption of GS has the potential to reduce or eliminate the diagnostic odyssey in patients with GDD/ID and have downstream effects on care management(Bowling et al., 2017; Gilissen et al., 2014; Zahir et al., 2017). Moreover, with the drop in sequencing price and development of data processing tools, GS could be offered as the first-tier test for select indications such as GDD/ID(Lindstrand et al., 2019; Lionel et al., 2018). Even only a small number of clinics would perform GS as the first test of choice currently. The American College of Medical Genetics and Genomics (ACMG) recommended GS as a first-tier or second-tier test for pediatric patients with congenital anomalies or ID together with ES in a recently published clinical guideline(Manickam et al., 2021).

Here, we present the results of a multi-center prospective study that evaluated the diagnostic and clinical utility of GS in children with suspected GDD/ID.

2 MATERIAL AND METHODS

2.1 Study participants

Study participants were recruited from 10 hospitals located in China from February 2018 to August 2020. Participants meeting the following inclusion criteria were enrolled: (i) less than 18 years of age with a clinical diagnosis of GDD/ID, (ii) brain MRI results available and accessible; (iii) normal karyotype, (iv) clinical and phenotypic information available (v) inconclusive results from at least one previous genomic test (CMA or ES), and (vi) raw data from CMA and/or ES available for reanalysis. The exclusion criteria were the following: (i) GDD/ID was determined to be caused by non-genetic factors (e.g., infection, trauma) and (ii) The patient was diagnosed by previous reanalysis.

Written consent was obtained for all patients. The Ethics Board of Xin Hua Hospital reviewed and approved the study protocol (XHEC-C-2018-002). The study adheres to the principles set out in the Declaration of Helsinki. Written consent was obtained for all patients. The study protocol was submitted to ClinicalTrials under the accession No: NCT03424772.

2.2 Study design

The overall workflow is illustrated in **Figure 1** . In total, 211 GDD/ID families were consecutively screened for eligibility from participating hospitals in China and105 patients met eligibility criteria. Parent-child trios/duos and affected family members (including siblings and grandparents) received GS when available. Five cases were later excluded due to the following reasons: (i) wrong samples (n=2), (ii) poor sequencing data (n=2), and (iii) variants were in the ES data but not detectable due to improper filtration (n=1). This

resulted in analysis of GS data from 100 patients, parents, and affected family members. Data analysis was performed to find clinically relevant variants and candidate variant lists were sent back to the referring clinician to evaluate the possible diagnosis. At the end of the study, the undiagnosed GS patients were analyzed again with updated database and gene annotation. The definitive diagnosis information was collected to calculate the yield. The pathogenic variant(s) were compared to CMA/ES data to determine reasons for missed diagnoses.

2.3 Genome sequencing

GS was performed by Berry Genomics (Beijing, China). The DNA of the affected families was isolated from peripheral blood using the QIAamp DNA Mini Kit (QIAGEN, Hilden, Germany). Library preparation was performed using TruSeq DNA PCR-Free Library Prep Kit (Illumina, San Diego, CA, USA) following the manufacturer's protocol. The prepared libraries were sequenced on a NovaSeq 6000 (Illumina) to generate 150bp paired end reads. The average depth was approximately 40X.

2.4 Data Analysis

Data analysis for all patients was performed at Xin Hua Hospital. The genome sequencing data analysis pipeline has been described previously(Sun et al., 2021). Small variants, CNVs, mitochondrial variants, loss of heterozygosity (LOH) and trinucleotide repeats were called from the sequencing result. For each family, small variants and CNVs were analyzed together under three possible inheritance patterns: autosomal dominant/de novo, autosomal recessive and X-linked. Patient phenotypes were compared with identified variants and their respective inheritance patterns. Phenotypic information was collected using Human Phenotype Ontology (HPO) terminology. The variant pathogenicity was evaluated according to ACMG guidelines(Richards et al., 2015; Riggs et al., 2020).

2.5 Variant validation

Clinically relevant small variants were validated by Sanger sequencing. The CNVs were confirmed by gap-polymerase chain reaction (PCR) and quantitative PCR (qPCR). The primer sequence is available on request. The LOH was validated by Multiplex Ligation-dependent Probe Amplification (MLPA) using SALSA MLPA Probemix ME028 Prader Willi/Angelman (MRC Holland, Amsterdam, the Netherlands).

2.6 Meta genome analysis

For sequencing files with low alignment percentage, the unaligned reads were collected using SAMtools(Li et al., 2009) and analyzed by Metagenomics Rapid Annotation using Subsystem Technology (MG-RAST)(Glass, Wilkening, Wilke, Antonopoulos, & Meyer, 2010).

2.7 Statistical analysis

A chi-square test with two degrees of freedom was applied to identify statistically significant differences. The p-value cutoff was 0.05.

2.8 Clinical utility

A survey was administered to parents via telephone during follow-up to assess the clinical utility of GS.

3 RESULTS

3.1 Cohort description

Samples from the patients and their families (n = 100) were collected from 10 hospitals located across China and included 44 females and 56 males. Of these, only one family member was affected in 89 families. The mean [interquarter range] age of symptom onset of the probands was 32 months [7, 38.75], while the mean [interquarter range] age at enrollment was 59 months [24.5, 87.75]. On average, it took more than two years to receive a diagnosis after symptom onset. Fourteen patients received inconclusive results from CMA and 62 received nondiagnostic ES. A total of 24 patients received both CMA and ES prior to enrollment (**Figure 2a**). Patients presented with phenotypes frequently associated with GDD/ID, such as abnormal brain MRI

(n=56), facial dysmorphism (n=30), and hypotonia (n=28). An overview of the families is summarized in **Table 1** . The detailed information of each family is listed in **Table S1** .

3.2 Diagnostic yield of the undiagnosed GDD/ID cohort

In this cohort, pathogenic variants were identified in 23 families with GS. Highly suspicious variants (i.e., variants reported in the literature but lack a well-established disease-causing relationship) were found in four families (**Table 2**). The identified pathogenic variations included small variants, CNVs and one LOH causing Angelman syndrome. No mitochondrial variants or abnormal expanded repeats were found in the study. The diagnostic yield of GS in CMA only cases was high (64.3%, 9/14), while in ES only families and CMA + ES families (12.9%, 8/62 and 25.0%, 6/24, respectively) the diagnostic yield of GS was significantly lower (p=0.000194, chi-square (degrees of freedom=2) = 17.0975).

We examined the pathogenic variants found by GS and compared them to the data from previous CMA/ES tests. Nine different scenarios were identified as plausible explanations for missed CMA/ES diagnoses (**Figure 2b**): 1. Patients received CMA only and the causative variants were too small to be detected (n=9); 2. The disease-causing genes were reported after study enrollment (n=5); 3. ES failed to detect pathogenic CNVs (n=4); 4. LOH was not called in ES reanalysis (n=1); 5. Improper annotation (n=1); 6. Patient did not manifest the clinical features when the previous tests were performed (n=1); 7. Mutant allele dropped out in ES (n=1,**Figure 3a**); 8. The complex variant was not captured well by ES (n=1, **Figure 3b**); and 9. A 3' untranslated region (3'UTR) deletion was not captured by ES (n=1).

Of the 23 positive cases, seven could have been solved by reanalyzing data from prior tests (30.4%): three had previous ES testing, and four were previously tested with both ES and CMA. Updates to analysis pipelines, annotation databases, and clinical follow-up were reasons for reanalysis success (**Figure 2b**). Reanalyzed data from CMA only cases provided no additional diagnoses.

3.3 Clinical implication of GS

Of the 23 cases with positive diagnoses, nine families experienced a change in clinical management and received information that impacted family planning. No changes to clinical care were noted in five families and clinical utility data was unavailable for six families.

Therapeutically, by knowing the molecular etiology of the patients, four families received potential targeted treatments. Patient iw138 in Family WGSI045XH was diagnosed with Pyruvate dehydrogenase E1-alpha deficiency (OMIM 312170). Her treatment was changed from levocarnitine to a ketogenic diet. Patient iw266 in Family WGSI086XH started methylene blue treatment after she was diagnosed by Methemoglobinemia (OMIM 250800) with compound heterozygous variants in *CYB5R3* gene. For diseases with currently no known effective treatment options, GS helped to point out the possible therapies. Family WGSI010XH, diagnosed by a newly reported mitochondrial disorder caused by biallelic *HPDL*variants(Sun et al., 2021), received "mito cocktail" treatment. Patient iw323 from Family WGSI104XH received omeprazole and acetylcysteine treatment by the local hospital as they found research on proton pump inhibitors as a potential therapy for *NGLY1* deficiency [PMID: 28512024]. After receiving the molecular diagnosis, two families (WGSI028XH and WGSI058XH) stopped the unnecessary levocarnitine medicine.

Five families reported that GS diagnosis results altered their family planning. Three families (WGSI010XH, WGSI015XH and WGSI036HR) welcomed healthy children with the help of preimplantation diagnosis and prenatal diagnosis. Two families (WGSI058XH and WGSI095XH) reported they were planning for the next pregnancy.

3.4 Detection of microorganism contamination

In one sample (iw052), only 90.5% reads could be aligned to the human reference genome, which was much lower than other samples (>98%) (**Figure S1a**). A review of unaligned reads revealed that 99.3% of them were enterobacterial in origin (**Figure S1b**), suggesting sample contamination or infection. We recollected the blood two months later and no enterobacteria were detected in the sequencing data.

4 DISCUSSION

GS offers comprehensive variant detection in coding and non-coding regions of the genome. Thus, GS may provide diagnostic benefit for conditions with high degrees of genetic and phenotypic heterogeneity, such as GDD/ID. In this study, pediatric patients with unexplained GDD/ID despite prior genomic testing (i.e., CMA and/or ES) received GS resulting in an overall diagnostic yield of 23% (23/100). When compared to previous testing, the reasons for missed diagnoses included small variants undetectable by CMA (9/23, 39.1%), CNVs missed by ES (4/23, 17.4%), and variants in new disease-causing genes (4/23, 17.4%).

For the families with previous CMA testing only, the yield after GS was high (9/14, 64.3%) and reanalysis of the CMA data did not provide any diagnoses. Though studies suggest that all previously reported genetic testing data should be periodically reinterpreted(SoRelle, Thodeson, Arnold, Gotway, & Park, 2019; Sun et al., 2020), the present data demonstrate that there is little benefit in reanalysis of CMA data. Rather, our data suggest that patients with negative CMA results should be offered GS to increase the likelihood of reaching a diagnosis.

Previous work has shown that periodic reanalysis of ES data would benefit some patients (Wenger et al., 2017; Xiao et al., 2018). Here, we show that 8.1% (7/86) of the patients with ES data could be solved by reanalysis. Although the study design included a reanalysis step, it was difficult to account for newly reported disease-causing genes that emerged during the time interval between study enrollment and study conclusion. Periodic reanalysis of ES has some economic advantages over GS(Koriath et al., 2021), however this is based on the assumption that the pathogenic variants are located in the coding region of the human genome (representing approximately 1% of the entire genome) and not in non-coding regions where ES is unable to interrogate.

Our data show that 15 cases could not be solved by reanalysis including seven that received previous ES. Due to technical limitations, CMA was unable to detect small variations, while ES could not detect CNVs when the probe information was unknown. This was observed in scenarios 1 and 3 (above). If many samples are captured by the same known exome probe set, new tools allow CNV detection by ES data(Fromer et al., 2012; Talevich, Shain, Botton, & Bastian, 2016). However, the samples in this study were obtained from different hospitals. Thus, we could not perform CNV analysis to avoid scenario 3. Furthermore, the algorithms were unable to detect single exon CNVs(Sun et al., 2020). In this study, iw098 of Family WGSI032XH carried an exon 1 deletion of *ARID1B*gene. Even if we had reanalyzed the CNV data, the family would remain undiagnosed. GS was able to overcome this limitation by detecting different variant types of different sizes in a single test.

ES uses probes to capture the ROI primarily in coding regions. For the WGSI078XH family, patient iw243 carried a deletion in the 3'UTR of*IGF2* and was diagnosed with Silver-Russell syndrome 3 (OMIM 616489). As there is no probe that covers the deleted region, it was not detected by ES. Moreover, probe capture is not perfect. The highly variable region is not well captured because the variant allele might have low affinity to the probe which could result in allele dropout (scenarios 7 and 8). Family WGSI072XH (**Figure 3a**) and iw028 of WGSI010XH represent examples of these scenarios (**Figure 3b**). GS simplifies this process by sequencing everything in the isolated DNA, thereby overcoming the aforementioned probe issues encountered with ES.

Taken together, GS demonstrated a diagnostic advantage over CMA and ES in large part due to their technical limitations. A bonus is that GS was able to detect DNA contamination or blood infection. With the ability to detect pathogenic variants in coding and noncoding regions of the genome and the steady decline in sequencing costs, GS has significant potential to eliminate or reduce the burden of a lengthy diagnostic odyssey in GDD/ID patients.

GS has shown great promise for identifying targeted treatments in NICU and pediatric intensive care unit (PICU) patients with its ability to achieve rapid diagnosis (French et al., 2019; Wang et al., 2020). Here, we show the clinical utility of GS in GDD/ID with 39.1% (9/23) of those diagnosed reporting changes to medical management. Potential treatments were identified for four families and two families stopped unnecessary

medications. Moreover, so far, at least five families have changed their reproductive plans, resulting in the birth of healthy children. For undiagnosed GDD/ID, the treatment is limited unless there is an underlying metabolic disorder, which is reflected in our study cohort. Although no effective therapies were found for some families, four families (WGSI023XH, WGSI028XH, WGSI036XH and WGSI052XH) reported that GS helped to relieve stress. Two of them (WGSI023XH and WGSI052XH) joined internet patient groups where they could communicate the disease courses with other families affected by the same disease. In total, 12 out of the 17 families who participated in the phone interview commented on the positive aspect brought by GS diagnosis.

Though GS might have advantages in structural variation (SV) detection, we did not identified patients with pathogenic SV in this study. This might be due to the limitation of short reads from next generation sequencing (NGS). For the patients remain undiagnosed, long read sequencing might help finding the causative variations. Furthermore, the families in this study were enrolled in a consecutive fashion and not at the same time. Thus, the time between receipt of the CMA/ES data and the study conclusion was relatively long. Thus, four cases were solved by reanalysis because of newly reported genes.

In conclusion, GS is a comprehensive method to detect different types of variants in genomic and mitochondrial DNA, which could reduce the burden of test selection faced by clinicians. This study suggests that GS has clinical advantages for undiagnosed GDD/ID patients even with reanalysis. Finally, because GS provides the most comprehensive level of coverage, it is possible to refer back to the original GS dataset to integrate newly discovered gene-disease associations and update clinical presentations that would further enhance the diagnostic yield.

ACKNOWLEDGEMENTS

We would like to thank the patients and their family members for participating in the study. The sequencing kits were supplied by Illumina. The genome sequencing was performed by Berry Genomics. The data were stored and analyzed in a cloud service GeneDock.

This work was funded by the Precision Medical Research of National Key Research and Development Program (2018YFC1002200, 2019YFC1005100 to Y. Yu, 2018YFC1002400 to K. Sun), National Natural Science Foundation of China (81873724 to Y. Sun, and 82070914 and 81873671 to Y. Yu), Shanghai Shen Kang Hospital Development Center (SHDC12017109 to Y. Yu), the Shanghai Science and Technology Commission (19140904500 to Y. Yu), Jiaotong University Cross Biomedical Engineering (YG2017MS72 to Y. Yu), and Foundation of Shanghai Municipal Health Commission (shslczdzk05702, to Y. Yu and Y. Sun), and Municipal Education Commission-Gaofeng Clinical Medicine Grant Support (20191908, to Y. Yu).

CONFLICT OF INTERESTS

The authors declare that they have no competing interests.

AUTHORS CONTRIBUTIONS

JP, DL, ZH, BX, WQ, YShen, NP, YL, CL, ZQ, JL, FC, Jingmin W, ZZ, HW, JD, HY, RD; Junyu W; YZ, SL, LW, and YY contributed to acquisition of the samples and data. YSun, NX and XY analyzed the data. XY, NX, LC, YD performed the experiments. YSun, KS, LW, YY conceptualized the work. YSun and YY drafted the manuscript. All author revised the manuscript.

DATA AVAILABILITY STATEMENT

All data relevant to the study are included in the article or uploaded as supplementary information. All clinically relevant variants had been submitted to LOVD database (https://databases.lovd.nl/) under the accession number: 00373535, 00373536, 00373656, 00373658, 00373659, 00373713, 00373714, 00373715, 00373716, 00373717, 00373718, 00373719, 00373720, 00373721, 00373722, 00373723, 00373725, 00373728, 00373729, 00373737, 00373738, 00373739, 00373740, 00373741, 00373742, 00373789, 00373790, 00373800, 00373801.

REFERENCES

Bowling, K. M., Thompson, M. L., Amaral, M. D., Finnila, C. R., Hiatt, S. M., Engel, K. L., . . . Cooper, G. M. (2017). Genomic diagnosis for children with intellectual disability and/or developmental delay. *Genome Med, 9* (1), 43. doi:10.1186/s13073-017-0433-1

Choo, Y. Y., Agarwal, P., How, C. H., & Yeleswarapu, S. P. (2019). Developmental delay: identification and management at primary care level. *Singapore Med J, 60* (3), 119-123. doi:10.11622/smedj.2019025

Fan, Y., Wang, L., Sun, Y., Xu, T., Gong, Z., Zhao, Q., . . . Yu, Y. (2021). Diagnostic yield of additional exome sequencing after the detection of long continuous stretches of homozygosity (LCSH) in SNP arrays. *J Hum Genet, 66* (4), 409-417. doi:10.1038/s10038-020-00854-1

Farnaes, L., Hildreth, A., Sweeney, N. M., Clark, M. M., Chowdhury, S., Nahas, S., . . . Kingsmore, S. F. (2018). Rapid whole-genome sequencing decreases infant morbidity and cost of hospitalization. *NPJ Genom Med, 3* , 10. doi:10.1038/s41525-018-0049-4

French, C. E., Delon, I., Dolling, H., Sanchis-Juan, A., Shamardina, O., Megy, K., . . . Raymond, F. L. (2019). Whole genome sequencing reveals that genetic conditions are frequent in intensively ill children. *Intensive Care Med, 45* (5), 627-636. doi:10.1007/s00134-019-05552-x

Fromer, M., Moran, J. L., Chambert, K., Banks, E., Bergen, S. E., Ruderfer, D. M., . . . Purcell, S. M. (2012). Discovery and statistical genotyping of copy-number variation from whole-exome sequencing depth. *Am J Hum Genet, 91* (4), 597-607. doi:10.1016/j.ajhg.2012.08.005

Gilissen, C., Hehir-Kwa, J. Y., Thung, D. T., van de Vorst, M., van Bon, B. W., Willemsen, M. H., . . . Veltman, J. A. (2014). Genome sequencing identifies major causes of severe intellectual disability. *Nature, 511* (7509), 344-347. doi:10.1038/nature13394

Glass, E. M., Wilkening, J., Wilke, A., Antonopoulos, D., & Meyer, F. (2010). Using the metagenomics RAST server (MG-RAST) for analyzing shotgun metagenomes. *Cold Spring Harb Protoc, 2010* (1), pdb prot5368. doi:10.1101/pdb.prot5368

Koriath, C. A. M., Kenny, J., Ryan, N. S., Rohrer, J. D., Schott, J. M., Houlden, H., . . . Mead, S. (2021). Genetic testing in dementia - utility and clinical strategies. *Nat Rev Neurol, 17* (1), 23-36. doi:10.1038/s41582-020-00416-1

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., . . . Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools.*Bioinformatics, 25* (16), 2078-2079. doi:10.1093/bioinformatics/btp352

Lindstrand, A., Eisfeldt, J., Pettersson, M., Carvalho, C. M. B., Kvarnung, M., Grigelioniene, G., . . . Nilsson, D. (2019). From cytogenetics to cytogenomics: whole-genome sequencing as a first-line test comprehensively captures the diverse spectrum of disease-causing genetic variation underlying intellectual disability. *Genome Med, 11* (1), 68. doi:10.1186/s13073-019-0675-1

Lionel, A. C., Costain, G., Monfared, N., Walker, S., Reuter, M. S., Hosseini, S. M., . . . Marshall, C. R. (2018). Improved diagnostic yield compared with targeted gene sequencing panels suggests a role for whole-genome sequencing as a first-tier genetic test. *Genet Med, 20* (4), 435-443. doi:10.1038/gim.2017.119

Manickam, K., McClain, M. R., Demmer, L. A., Biswas, S., Kearney, H. M., Malinowski, J., . . . Directors, A. B. o. (2021). Exome and genome sequencing for pediatric patients with congenital anomalies or intellectual disability: an evidence-based clinical guideline of the American College of Medical Genetics and Genomics (ACMG). *Genet Med* . doi:10.1038/s41436-021-01242-6

Marshall, C. R., Chowdhury, S., Taft, R. J., Lebo, M. S., Buchan, J. G., Harrison, S. M., . . . Medical Genome, I. (2020). Best practices for the analytical validation of clinical whole-genome sequencing intended for the diagnosis of germline disease. *NPJ Genom Med, 5* , 47. doi:10.1038/s41525-020-00154-9

McGuire, D. O., Tian, L. H., Yeargin-Allsopp, M., Dowling, N. F., & Christensen, D. L. (2019). Prevalence of cerebral palsy, intellectual disability, hearing loss, and blindness, National Health Interview Survey, 2009-2016. *Disabil Health J, 12* (3), 443-451. doi:10.1016/j.dhjo.2019.01.005

Meynert, A. M., Ansari, M., FitzPatrick, D. R., & Taylor, M. S. (2014). Variant detection sensitivity and biases in whole genome and exome sequencing. *BMC Bioinformatics, 15* , 247. doi:10.1186/1471-2105-15-247

Miller, D. T., Adam, M. P., Aradhya, S., Biesecker, L. G., Brothman, A. R., Carter, N. P., . . . Ledbetter, D. H. (2010). Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. *Am J Hum Genet, 86* (5), 749-764. doi:10.1016/j.ajhg.2010.04.006

Moeschler, J. B., Shevell, M., & Committee on, G. (2014). Comprehensive evaluation of the child with intellectual disability or global developmental delays.*Pediatrics, 134* (3), e903-918. doi:10.1542/peds.2014-1839

Pang, A. W., Macdonald, J. R., Yuen, R. K., Hayes, V. M., & Scherer, S. W. (2014). Performance of high-throughput sequencing for the discovery of genetic variation across the complete size spectrum. *G3 (Bethesda), 4* (1), 63-65. doi:10.1534/g3.113.008797

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., . . . Committee, A. L. Q. A. (2015). 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, 17* (5), 405-424. doi:10.1038/gim.2015.30

Riggs, E. R., Andersen, E. F., Cherry, A. M., Kantarci, S., Kearney, H., Patel, A., . . . Martin, C. L. (2020). Technical standards for the interpretation and reporting of constitutional copy-number variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen). *Genet Med, 22* (2), 245-257. doi:10.1038/s41436-019-0686-8

Robert L. Schalock, R. L., and Marc J. Tassé. (2021). *Intellectual Disability: Definition, Diagnosis, Classification, and Systems of Supports, 12th Edition* : Intellectual Disability: Definition, Diagnosis, Classification, and Systems of Supports, 12th Edition.

SoRelle, J. A., Thodeson, D. M., Arnold, S., Gotway, G., & Park, J. Y. (2019). Clinical Utility of Reinterpreting Previously Reported Genomic Epilepsy Test Results for Pediatric Patients. *JAMA Pediatr, 173* (1), e182302. doi:10.1001/jamapediatrics.2018.2302

Srivastava, S., Love-Nichols, J. A., Dies, K. A., Ledbetter, D. H., Martin, C. L., Chung, W. K., . . . Group, N. D. D. E. S. R. W. (2019). Meta-analysis and multidisciplinary consensus statement: exome sequencing is a first-tier clinical diagnostic test for individuals with neurodevelopmental disorders.*Genet Med, 21* (11), 2413-2421. doi:10.1038/s41436-019-0554-6

Subspecialty Group of Neurology, C. S. o. P. C. M. A., & Project Expert Group of Childhood Neuropathy, C. N. A. (2018). [Experts' consensus on the diagnostic strategies of etiology for intellectual disability or global developmental delay in children]. *Zhonghua Er Ke Za Zhi, 56* (11), 806-810. doi:10.3760/cma.j.issn.0578-1310.2018.11.003

Sun, Y., Wei, X., Fang, F., Shen, Y., Wei, H., Li, J., . . . Yu, Y. (2021). HPDL deficiency causes a neuromuscular disease by impairing the mitochondrial respiration.*J Genet Genomics* . doi:10.1016/j.jgg.2021.01.009

Sun, Y., Ye, X., Fan, Y., Wang, L., Luo, X., Liu, H., . . . Yu, Y. (2020). High Detection Rate of Copy Number Variations Using Capture Sequencing Data: A Retrospective Study.*Clin Chem, 66* (3), 455-462. doi:10.1093/clinchem/hvz033

Talevich, E., Shain, A. H., Botton, T., & Bastian, B. C. (2016). CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. *PLoS Comput Biol, 12* (4), e1004873. doi:10.1371/journal.pcbi.1004873

Wang, H., Lu, Y., Dong, X., Lu, G., Cheng, G., Qian, Y., . . . Zhou, W. (2020). Optimized trio genome sequencing (OTGS) as a first-tier genetic test in critically ill infants: practice in China. *Hum Genet, 139* (4), 473-482. doi:10.1007/s00439-019-02103-8

Wenger, A. M., Guturu, H., Bernstein, J. A., & Bejerano, G. (2017). Systematic reanalysis of clinical exome data yields additional diagnoses: implications for providers. *Genet Med, 19* (2), 209-214. doi:10.1038/gim.2016.88

Wright, C. F., FitzPatrick, D. R., & Firth, H. V. (2018). Paediatric genomics: diagnosing rare disease in children. *Nat Rev Genet, 19* (5), 253-268. doi:10.1038/nrg.2017.116

Xiao, B., Qiu, W., Ji, X., Liu, X., Huang, Z., Liu, H., . . . Sun, Y. (2018). Marked yield of re-evaluating phenotype and exome/target sequencing data in 33 individuals with intellectual disabilities. *Am J Med Genet A, 176* (1), 107-115. doi:10.1002/ajmg.a.38542

Zahir, F. R., Mwenifumbo, J. C., Chun, H. E., Lim, E. L., Van Karnebeek, C. D. M., Couse, M., . . . Marra, M. A. (2017). Comprehensive whole genome sequence analyses yields novel genetic and structural insights for Intellectual Disability. *BMC Genomics, 18* (1), 403. doi:10.1186/s12864-017-3671-0

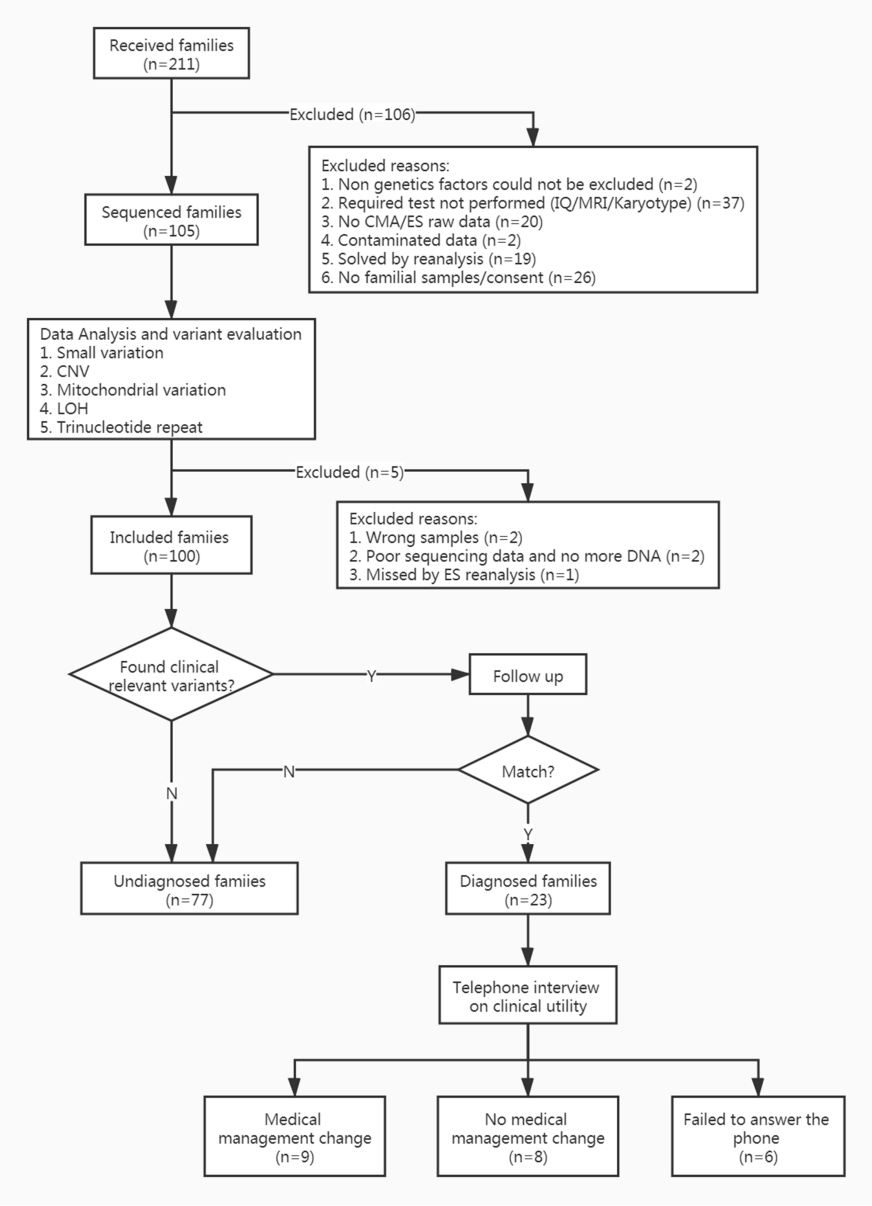
FIGURE LEGEND

Figure 1 . Study workflow.

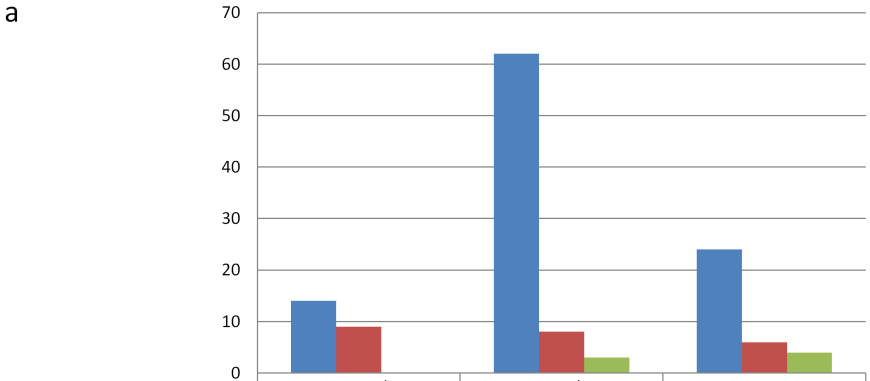
Figure 2 . The composition and yield of the cohort and result comparison with previous genomic tests. (a) Previous genomic tests received by study participants, including the diagnostic yield for each test and the number of families that could be solved by reanalysis. (b) Missed diagnosis scenarios. *WGSI010XH family harbored two complex variants in a new disease-causing gene and was counted in the "complex variants" group only.

Figure 3 . Variants detected by genome sequencing but not by exome sequencing. The bam files of the genome sequencing (upper panel) and exome sequencing (lower panel) of the same sample were visualized by Integrative Genomics Viewer. (a) iw225 of Family WGSI072XH carried a heterozygous variation NM_-152224.6:c.1811C>T which was dropped by exome sequencing. (b) iw028 of Family WGSI010XH harbored a complex heterozygous variant NM_032756.2:c.215_226delinsTGTACGGCCTGGAT. The variants would cause capture insufficiency, thus hampered variant calling.

Figure S1 . The contaminated sample iw052. (a) iw052 sample had lower alignment ratio compared to other sequencing samples. (b) 99.3% of the unaligned reads could be mapped to enterobacterial.

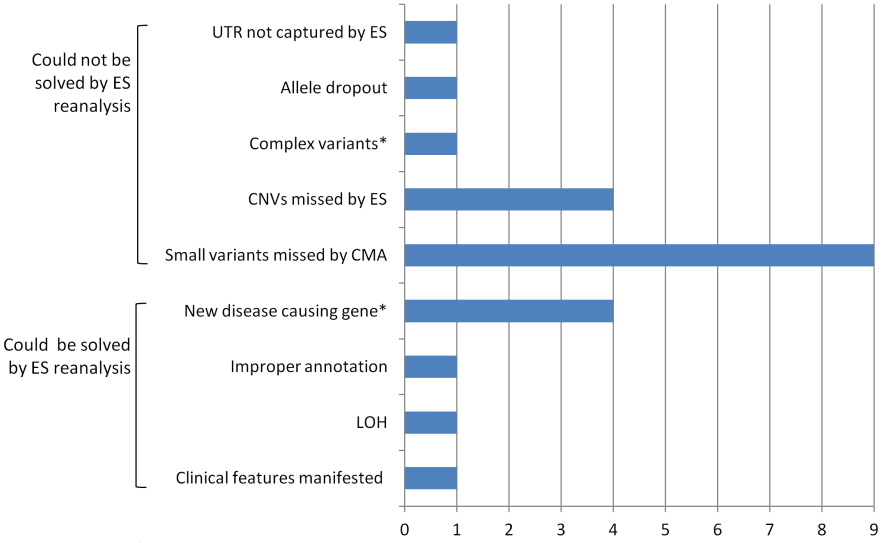


a

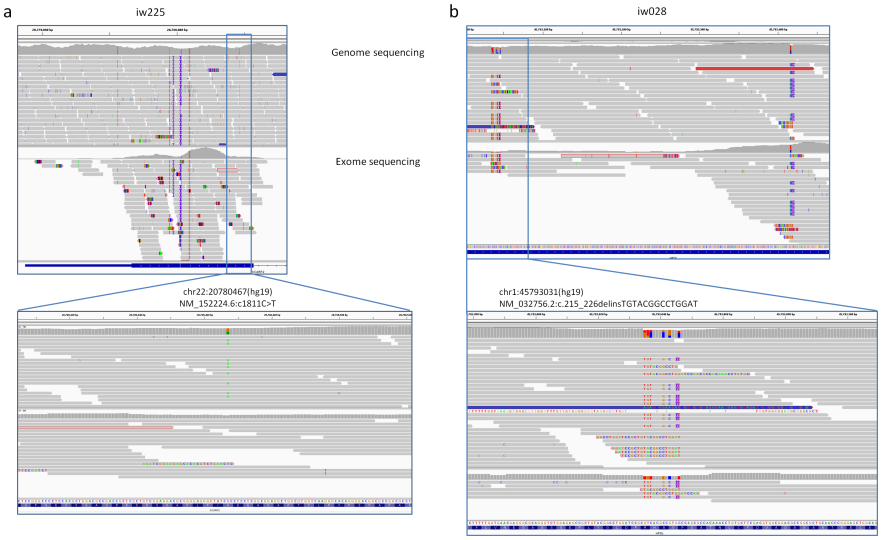


	CMA only	ES only	CMA + ES
Total	14	62	24
Diagnosed with GS	9	8	6
Diagnosed with ES reanalysis	0	3	4

b



* WGSI010XH familiy harbor two complex variants in a new disease causing gene, and was counted in complex variants group only



Hosted file

`Table.docx` available at https://authorea.com/users/333600/articles/538729-genome-sequencing-demonstrates-high-diagnostic-yield-in-children-with-undiagnosed-global-developmental-delay-intellectual-disability-a-prospective-study