

Deep phenotyping and whole-exome sequencing improved the diagnosis yield for nuclear pedigrees with neurodevelopmental disorders

Qingqing Wang¹, Xia Tang¹, ke yang², Xiaodong Huo¹, Hui Zhang¹, Keyue Ding¹, and Shixiu Liao¹

¹Henan Provincial People's Hospital

²Henan Provincial People's Hospital

July 16, 2020

Abstract

Neurodevelopmental disorders, a group of early-onset neurological disorders with significantly clinical and genetic heterogeneity, remain a diagnostic odyssey for clinical genetic evaluation. In a total of 45 parent-child trios/quads with these disorders that was 'not yet diagnosed' by the traditional testing methods, we assessed the diagnostic yield by the combined use of standardized phenotypes and whole-exome sequencing data. Using a standardized vocabulary of phenotypic abnormalities from Human Phenotype Ontology (HPO), we performed deep phenotyping for these pedigrees to characterize multiple clinical features that was extracted from Chinese electronic medical records (EMRs). By matching HPO terms with known human diseases or cross-species comparison, together with whole-exome sequencing data, we prioritized candidate mutations/genes that underlies these pedigrees. We obtained a diagnostic yield of 49% (22 out of 45) with probably or possibly genetic diagnosis, of which the compound heterozygosity and de novo mutations accounted for the half of the diagnosis. Of note, the pedigrees with probable or possible diagnosis accompanied with a greater number of phenotypes implicated in non-nervous systems. The combined use of deep phenotyping and whole-exome sequencing provide implications for etiological evaluation for neurodevelopmental disorders in the clinical setting.

Qingqing Wang¹, Xia Tang¹, Ke Yang, Xiaodong Huo, Hui Zhang, Keyue Ding*, and Shixiu Liao*

Medical Genetic Institute of Henan Province, Henan Provincial People's Hospital, Henan Key Laboratory of Genetic Diseases and Functional Genomics, National Health Commission Key Laboratory of Birth Defect Prevention, Henan Provincial People's Hospital of Henan University, People's Hospital of Zhengzhou University, Zhengzhou, Henan Province 450003, P.R. China

¹ Contributed equally

* Corresponding authors

Keyue Ding, PhD, E-mail: ding.keyue@igenetics.org.cn

Shixiu Liao, MD, E-mail: ychslshx@henu.edu.cn

Abstract

Neurodevelopmental disorders, a group of early-onset neurological disorders with significantly clinical and genetic heterogeneity, remain a diagnostic odyssey for clinical genetic evaluation. In a total of 45 parent-

child trios/quads with these disorders that was ‘not yet diagnosed’ by the traditional testing methods, we assessed the diagnostic yield by the combined use of standardized phenotypes and whole-exome sequencing data. Using a standardized vocabulary of phenotypic abnormalities from Human Phenotype Ontology (HPO), we performed deep phenotyping for these pedigrees to characterize multiple clinical features that was extracted from Chinese electronic medical records (EMRs). By matching HPO terms with known human diseases or cross-species comparison, together with whole-exome sequencing data, we prioritized candidate mutations/genes that underlies these pedigrees. We obtained a diagnostic yield of 49% (22 out of 45) with probably or possibly genetic diagnosis, of which the compound heterozygosity and *de novo* mutations accounted for the half of the diagnosis. Of note, the pedigrees with probable or possible diagnosis accompanied with a greater number of phenotypes implicated in non-nervous systems. The combined use of deep phenotyping and whole-exome sequencing provide implications for etiological evaluation for neurodevelopmental disorders in the clinical setting.

Keywords: Neurodevelopmental disorders, Human Phenotype Ontology, Whole-exome sequencing, Diagnosis

Introduction

Neurodevelopmental disorders - a group of early onset neurological disorder - affect more than 3% of children [1]. According to the DSM-5 [2], it can be classified into disorders of intellectual disability, communication, autism spectrum, attention-deficit hyperactivity, specific learning, motor, and others. Neurodevelopmental disorders are common reasons of referrals to genetic counselor [3] and there remains great challenges in genetic evaluation due to the heterogeneous clinical presentation [4].

Clinical laboratory investigations for neurodevelopmental disorders include neuroimaging, metabolic screening, traditional genetic testing (e.g., karyotype, chromosomal microarray analysis, or gene panel sequencing), and invasive tests [5]. However, >50% of patients of neurodevelopmental disorders did not receive an etiologic diagnosis [6,7]. Recently, the application of whole-exome or -genome sequencing for its diagnosis has been assessed. A large family-based study ($n = 4,293$) showed that approximately 42% of patients with development disorders harbored *de novo* pathogenic mutations [8]. Specifically, a diagnosis rate of 36%-48% was obtained for patients with neurodevelopmental disorders [9-13]. Furthermore, the diagnostic yield can be increased using an improved analysis pipeline, e.g., it was increased from 27% to 40% by re-analyzing 1,133 families with developmental disorders [14], and a 15.4% of additional diagnosis for 416 children with congenital anomalies or mental retardations was achieved [15]. However, in a clinical setting, there remains great interest in implementing novel approaches for increasing the diagnostic yield for ‘not yet diagnosed’ patients.

Deep phenotyping - the precise and comprehensive analysis of phenotypic abnormalities - aims to provide the best clinical care for each patient according to disease stratification [16]. The human phenotype ontology (HPO), a standard vocabulary for describing the phenotypic abnormalities in human disease, provides the most comprehensive resources for deep phenotyping [16,17]. Using the standardized human phenotype ontology, several tools (e.g., *Phenolyzer*, *Phenomiser*, and *Exomiser*) have been developed for clinical and genetic diagnosis. *Phenolyzer* (i.e., phenotype-based gene analyzer) discovers genes implicated in diseases based on prior phenotype or disease information [18]; and *Exomiser* prioritizes disease-associated genes/mutations by analyzing whole-exome sequencing data with its matched phenotypes [19,20]. To the best of our knowledge, incorporating deep phenotyping with whole-exome sequencing for the assessment of the diagnostic yield for neurodevelopment disorders in family-based study remains limited.

Here, we performed the phenotype-driven diagnosis for ‘not yet diagnosed’ pedigrees with neurodevelopment disorders. Deep phenotyping for heterogenous clinical features, together with whole-exome sequencing, prioritized potentially pathogenic genes/mutations underlying these families using phenotype matching algorithms.

Materials and Methods

The conceptual framework for the phenotype-driven diagnosis for nuclear pedigrees with neurodevelopmental disorders was composed of deep phenotyping, whole-exome sequencing, variant filtering, and phenotype-matching based prioritization (*Fig 1*).

The recruitment of pedigree with neurodevelopmental disorders

We recruited nuclear pedigrees (i.e., parent-offspring) with chief complaints of ‘developmental delay’, ‘intellectual disability or ‘seizure’ (*Table S1*), who had genetic counseling visit our medical genetic institute. We excluded patients with known etiologies based on traditional genetic testing or metabolic screening (*Table S2*).

Deep phenotyping and phenotype standardization

Various clinical notes including medical history, laboratory tests and/or radiologic reports, were collected from the Chinese electronic medical records at the Henan Provincial People’s Hospital. Clinical features of symptoms, signs, laboratory and radiologic tests were extracted manually for each proband (i.e., deep phenotyping), as present in our previous studies [21,22]. The extracted phenotypes in Chinese were then standardized by searching for HPO terms in Chinese Human Phenotypic Ontology browser (<http://www.chinahpo.org>), and one of the most matched HPO term was selected if multiple terms were noted (*Table S3*). The packages of ‘ontologyIndex’ and ‘hpoPlot’ in *R* were used to perform HPO-based analyses [23].

Whole-exome sequencing

Genomic DNA was extracted from the peripheral blood lymphocytes using the QIAamp DNA Blood Mini kit (Qiagen, Hilden Germany), and was then fragmented into 250-300bp by sonicator (Covaris, Woburn MA). The sequencing library was constructed using SureSelect Human All Exon V6 kit (Agilent, Santa Clara CA) and whole-exome sequencing was performed on HiSeq Xten sequencing platform (Illumina, San Diego CA) at the Beijing Genomics Center (Shenzhen, China). Sequencing data with paired-end length of 150 bp were obtained from 45 nuclear pedigrees (including 55 affected cases and 97 individuals without phenotypes), with an approximately mean sequencing depth of 30×.

Data analysis

Whole-exome sequencing reads were aligned to the hg19 reference genome with BWA [24], applied GATK (v.4.1) for indel realignment, duplicate removal and base score quality recalibration, and single nucleotide variant (SNV) and small insertion and deletion (indel) across all individuals in a family was identified using standard hard filtering according to GATK Best Practice Guide [25].

Following variant calling, we annotated and classified mutations into pathogenic, likely pathogenic, variants with uncertain significance (VUS), likely benign and benign according to the ACMG/AMP guideline [26], as implemented in InterVar [27]. The minor allele frequency (MAF) was ascertained from the 1000 Genome Project (1000G) [28] or the Genome Aggregation Database (gnomAD, v.2.0.2) [29]. Databases of ClinVar [30], OMIM (omim.org) and HGMD [31] were used to identify the known pathogenic variants implicated in the neurodevelopmental disorders.

To identify potentially causal genes/mutations underlying the affected pedigree and thus make a genetic diagnosis, we used *Exomiser*[32] to prioritize mutations by integrating the calculation of phenotype similarity between the known human diseases and mouse models (i.e., cross-species comparisons) with evaluation of mutations according to pathogenicity, MAF (< 0.1%), and mode of inheritance. A detailed protocol for conducting such analyses was provided recently [33]. We also applied *Phenolyzer* [18] to discover genes implicated in neurodevelopmental disorders using HPO terms alone, leveraging prior biological knowledge and phenotype information.

Sanger sequencing

The identified potentially causal mutations were validated using Sanger sequencing. PCR products were sequenced in bidirection using ABI 3700 sequencer, and were analyzed using SnapGene Viewer (<https://www.snapgene.com/snapgene-viewer/>).

Results

The recruited pedigrees

During 2019-01 to 2019-06, we recruited a total of 45 ‘not yet diagnosed’ pedigrees with neurodevelopmental disorders, who were referred to the outpatient clinic at our institute (*Table 1*). The ratio of male: female of the proband was approximately 2:1, and all probands had age at symptom onset < 8 years. The majority of the pedigrees were referred for prenatal genetic counseling but five affected adults came for clinical genetic evaluation. These pedigrees included 31 parent-child trios, 12 parent-child quads, one parent-child quin, and one family with a second-degree relative. Consanguinity was not documented for the parents of the proband.

HPO encoded phenotypes recapitulating significant clinical heterogeneity for neurodevelopmental disorders

Deep phenotyping compiled the clinical features extracted from the Chinese electronic medical records and ascertained a total of 121 HPO terms via standardization (*Table S3* and *Fig S1*). An ontology plot showed the annotated HPO terms (i.e., as nodes indicated) as subgraphs of the full ontology, where a lineage represented a system hierarchy (*Fig 2A*), for example, two branches of ‘nervous system physiology’ and ‘nervous system morphology’ were under the lineage of ‘phenotypic abnormality of nervous system’. The plot also showed that the population frequency of HPO terms differed significantly among branches. Of note, phenotypes ‘*is-a*’ relation with ‘HP:0001249 (intellectual disability)’ or ‘HP:00012758 (neurodevelopmental delay)’ showed relatively higher frequencies.

Overall, more than half of HPO terms were associated with nervous system ($n = 66$, 55%), whereas the remaining terms were noted to be implicated in multiple non-nervous systems (*Fig 2B*). Per pedigree, a median number of eight HPO terms were annotated; only one pedigree present with ‘seizure’ and one had at most 16 phenotypes (*Fig 2C*). Cumulatively, nearly all families ($n = 44$) present with more than two phenotypes that such a multi-morbidity has important clinical implications [34]. We also noted that the phenotypes presenting in the affected individuals from the same pedigree varied, in part due to incomplete penetrance or at a later symptom onset (*Table S4*). One example in undiagnosed pedigree (UDP) #7 showed that the proband (p701) exhibited ‘HP:0001250 (seizures)’, ‘HP:0001270 (motor development delay)’ and ‘HP:0000750 (language development delay)’ at six months of age, whereas his sibling (p702) developed only ‘seizures’ at four years of age.

The most frequent phenotypes associated with nervous system were ‘HP:0000750 (delayed speech and language development)’, ‘HP:0001270 (motor delay)’, ‘HP:0001249 (intellectual disability)’, ‘HP:0012434 (delayed social development)’, and ‘HP:0001263 (global developmental delay)’ (*Fig 2D*). However, approximately half of these phenotypes were singleton ($n = 35$, 53%) or doubleton ($n = 6$, 9%). According to DSM-5, phenotypes under the lineage of ‘nervous system physiology’ were grouped each of which showing dominant phenotypes (e.g., ‘HP:0000750’ in communication disorder, and ‘HP:0001270’ in motor disorder), whereas phenotypes in attention deficit/hyperactivity disorder, autism spectrum disorder and specific learning disorder were less present. Phenotypes associated with non-nervous system were likely to be present (*Fig 2E*), indicative of syndromic features in a proportion of pedigrees ($n = 29$, 64%). The vast majority of these phenotypes were singleton or doubleton but ‘HP:0001252 (muscular hypotonia)’, ‘HP:0012389 (appendicular hypotonia)’ and ‘HP:0003808 (abnormal muscle tone)’ had a frequency of 12.5%, 10%, and 7%, respectively.

An increased diagnostic yield by incorporating HPO encoded phenotypes and whole-exome sequencing

We filtered the SNVs and indels by removing common variants (MAF > 0.01 in gnomAD database), and

then evaluating the remaining variants based on the predicted pathogenicity. Given the mode of inheritance, variants co-segregated with the pedigree were selected. For example, homozygous or compound heterozygous mutations were required under a mode of autosomal recessive inheritance. We assigned phenotypic score for genes based on the comparison with known human diseases or animal models with mutations in orthologues, and obtained the final ranking as the sum of the individual scores [32]. We then assigned the prioritized variants according to the following criteria: 1) pathogenic variant (PV): a variant presented in HGMD [31], Clinvar [30] or classified to be ‘pathogenic’ based on ACMG guidelines with matched phenotypes to neurodevelopmental disorders; 2) likely pathogenic variant (LPV): a non-HGMD or non-Clinvar variant, but was classified to be ‘pathogenic’ based on ACMG guidelines in the gene for which previously reported patients have matched phenotypes to neurodevelopmental disorders; and 3) variant of unknown significance (VUS): the variant that do not fulfill the above criteria but the corresponding gene has matched phenotypes to neurodevelopmental disorders. Thus, we classified genetic diagnosis into three groups as: 1) probable diagnosis: a PV or LPV identified in a gene relevant to phenotypes in the patient; (2) possible diagnosis: a VUS variant(s) identified in a gene relevant to phenotypes in the patient or the gene was prioritized by walking the interactome [32]; and (3) undiagnosed: no disease-associated variant(s) detected.

According to the ontology plot (*Fig 2A*), we hypothesized that the replacement of a given phenotype term (e.g., a term located at the low-level of the ontology) with its ancestral term may affect phenotype-matching. One striking example of UDP #9 highlighted such effects on prioritization. Deep phenotyping originally characterized ‘HP:0000253 (progressive microcephaly)’ in the proband (p901), which was descended from its ancestral term of ‘HP:0000252 (microcephaly)’. When ‘HP:0000252’ was used, the compound heterozygous mutations in *TSNE2* were prioritized with a Phenotype score of 0.878 and Exomiser score of 0.993, significantly greater than ‘HP:0000253’ was used (0.000 and 0.015, respectively). Thus, we replaced 35 terms with their corresponding ancestral terms in 27 pedigrees (i.e., one for 19 pedigrees, and two for eight pedigrees) (*Table S5*).

Overall, we achieved 13 probable and nine possible diagnosis from a total of 45 pedigrees (*Fig 3A*), leading to a diagnostic yield of 49%. Of the diagnosed pedigrees, nine were inherited in an autosomal dominant (AD) manner, four in an autosomal recessive (AR) with compound heterozygous mutations, eight in X-linked recessive (XR), and one in X-linked dominant (XD). In addition to the compound heterozygous mutations, the *de novo* mutation accounted for seven pedigrees in an AD manner (*Fig 3B*). A detailed annotation for these mutations, including population frequency, ACMG-guided clinical classification, and associated clinical syndromes were provided in *Table 2*. As expected, the Phenotype (*Fig 3C*) and Exomiser (*Fig 3D*) score were slight increased when the original phenotype term was replaced with its corresponding ancestral term, as noted obviously in UDP #9.

To assess the use of HPO-encoded phenotype alone in clinical diagnosis, we used *Phenolyzer* [18] for identifying the associated clinical syndromes underlying the pedigrees and its corresponding causal genes. The rank of the prioritized genes identified in *Exomiser* was compared with the rank of the genes seeded from *Phenolyzer*. Our findings indicated that, incorporating pathogenic mutations significantly increased the ranking for the prioritized genes, whereas a large set of seed genes generated by *Phenolyzer* created more difficulty in the prioritization process (Wilcoxon test, $p = 6 \times 10^{-5}$) (*Fig 3E*).

We finally investigated whether the phenotypic structure differed between the ‘diagnosed’ and ‘undiagnosed’ pedigrees, as shown in the landscape of HPO-encoded phenotypes for all pedigrees (*Fig S1*). Of note, the number of phenotypes associated with non-nervous system in the diagnosed pedigrees ($n = 22$) was significantly greater than that in the undiagnosed pedigrees ($n = 23$) (Wilcoxon test, $p = 9 \times 10^{-4}$), suggesting an increased power for the diagnosis of pedigrees with syndromic features (*Fig 3F*).

Case examples

For the purpose of illustration, we summarized the analyses for five pedigrees accompanied with various phenotypes (*Fig 4A*). The confirmation of Sanger sequencing for mutations prioritized for the remaining 17

pedigrees was shown in *Fig S2* .

UDP #1. The pedigree illustrated the differential diagnosis for clinical genetic evaluation according to the prioritized mutations (*Fig 4B*). The proband (p102, a 14-year-old boy) was delivered by Cesarean section at full term accompanied with hypoxia (Apgar score unknown) in 2005. At 18 months of age, he was able to walk with support; and he could not speak and has a lower self-reported intelligence quotient when compared with children at the same age. He was diagnosed with ‘ischemic hypoxic encephalopathy’ at a local hospital and received rehabilitation but the response to therapy was limited. His sibling (p101, a 9-year-old boy) had similar clinical features. Both siblings present with short-stature (127 cm and 117 cm, respectively), multiple facial scars, wind ears, strabismus, and multiple alopecia areate scattered on head. Multiple coffee skin spots on back, right cryptorchid, horseshoe valgus on both feet, and knee-jerk hyperactivity, were recorded in medical history. A novel pathogenic splicing mutation (NM_004187.5:c.2517.2622del) in *KDM5C* and a hemizygous VUS mutation (NM_021120.4:c.1334G>A, p.Arg445His) in *DLG3* were initially prioritized. *DLG3* is associated with mental retardation, X-linked 90 (OMIM:300850), but their healthy uncle harbored the same mutation. *KDM5C* is associated with X-linked intellectual disability (XLID) (OMIM:300534), and the splicing-site mutation was co-segregated with the phenotypes. Both siblings present with typical features of XLID, and we therefore inferred that the splicing-site mutation in *KDM5C* was causal.

UDP #21. The pedigree demonstrated that whole-exome sequencing in parent-child quad may increase the power to identify the causal mutation (*Fig 4C*). Both affected siblings were referred to a genetic counselor due to intellectual disability, accompanied with macrocephaly, postnatal overgrowth, delayed speech and language development, motor deterioration, and Chiari malformation. The proband (p2101) was delivered by Cesarean section at full term without hypoxia. At three years of age, his head circumference was 54 cm (>3 s.d.), and received a decompression surgery for subamygdala hernia in cerebellum six months later. Her sibling (p2102) had a similar medical history and present with delays in language development and motor skills, poor coordination and social adaptability at 19 months of age. She received the decompression surgery at two years of age, and developed intermittent seizures thereafter. She had a head circumference of 54 cm (>3 s.d.) at 28 months of age. Currently, she is able to speak and walk with support. Both siblings have facial abnormalities including prominent forehead, long face, short philtrum, dental malocclusion, everted lower lip vermilion, narrow mouth with open-mouth appearance, and down slanted palpebral fissures. A heterozygous nonsense mutation (NM_002501.4:c.935G>A, p.Trp312*) in *NFIX* was prioritized, which was associated with Sotos syndrome 2 (OMIM:614753) or Marshall-Smith syndrome (OMIM:602535). Sanger sequencing confirmed the same mutation in the mother but with a low mutation fraction, and the mechanism for the potentially gonadal chimeras needs to be further investigated.

UDP #10. The family identified a *de novo* mutation in the well-known gene associated with Rett syndrome (*Fig 4D*). The naturally delivered proband (p1001) developed torticollis at five months of age and got worse seven months later. She was able to sit but fail to crawl at eight months of age, and was able to stand at 14 months of age. At 18 months of age, she gradually showed hand involuntary movements, short attention span, and limited language ability. Neuropsychological examination showed that her intellectual development, social adaptability, and motor development was equivalent to eight months, nine months, and 11 months, respectively. At 33 months of age, she presented a global developmental delay with a head circumference of 45 cm, a height of 90 cm, and a weight of 11 kg. She showed tongue extension and astigmatism in the left eye. Our analysis identified a *de novoknown* nonsense mutation (rs61749721, NM_004992.3:c.763C>T, p.Arg255*) in *MECP2* - a well-known gene causing X-linked dominant Rett syndrome (OMIM:312750).

UDP #9. The pedigree showed that the identification of the pathogenic mutation may benefit for prenatal genetic diagnosis and thus reduce risks of birth defects (*Fig 4E*). The proband (p901) was delivered at full term but showed microcephaly at birth (< 3.2 s.d., compared to age matched and normal standards). He got feeding difficulty and intermittent opisthotonus progressively. At 12 months of age, he developed an encephalopathy with refractory tonic myoclonic epilepsy, failed to fix or follow vision, and present with limbic hypertonia. At eight years of age, his microcephaly had deteriorated considerably (<5.45 s.d.), and

was not able to speak, and he could not stand alone and present muscle atrophy in lower limbs. MRI scan could not be performed due to un-cooperation. A compound heterozygosity for two pathogenic mutations in *TSEN2*(NM025265.4:c.904G>A, p.Glu302Lys and NM025265.4:c.1354C>A, p.Arg452*;) was prioritized, associated with pontocerebellar hypoplasia type 2B (OMIM: 612389), which is characterized by abnormalities in the cerebellum and brainstem with a progressive microcephaly combined with extrapyramidal dyskinesia and epilepsy. The mother came for prenatal diagnosis in her 20th week of pregnancy. An ultrasound imaging showed that the head circumference of the fetus (p902) was less than the normal gestational age, and further genetic testing identified the same compound heterozygous mutations in *TSEN2* .

UDP #30. This pedigree showed that the identified pathogenic mutation may accelerate our understanding of the pleiotropic effect of the causal gene (*Fig 4F*). The proband (a 2-year-old girl) was born naturally but developed epileptic seizures 10 days later with a frequency of 2 to 30 times per day, lasting for 5 seconds to 5 minutes for each occurrence, occasionally with continuous attacking and shortness of breath. She showed developmental delays at nine months of age, e.g., unable to raise his head, with a low-pitched crying, unwilling to laugh, abnormal gaze, without ocular pursuit, and pupillary light reflex disappeared. MRI scan showed a C-shaped spine, thinning of the corpus callosum on back, and dysplastic white matter. An electroencephalogram showed a periodic eruption-suppression wave, and visual evoked potentials indicated severe abnormalities in bilateral visual pathways. An echocardiography reported an atrial septal defect (type II). Family history revealed that one of her siblings had similar symptoms and died three months after giving birth. A compound heterozygosity of a frameshift mutation (NM_016373.3:c.854dupA, p.N285fs*10; pathogenic) and a missense mutation (NM_016373.3:c.1063G>C, p.G355R; VUS) in *WWOX* was identified, derived from maternally and paternally, respectively. Homozygous or compound heterozygous mutations in *WWOX* - a tumor suppressor gene - cause spino-cerebellar ataxia autosomal recessive, type 12 (OMIM:614322) or epileptic encephalopathy, early infantile, 28 (EIEE28, OMIM:616211). A pedigree with EIEE28 reported to have atrial septal defect may be explained by the *WWOX* homozygous deletion or the compound heterozygous mutations in *HSPG2* [35].

Discussion

In the present study, we assessed the diagnostic yield for ‘not yet diagnosed’ pedigrees with neurodevelopment disorders in a clinical setting through the combined use of HPO-based deep phenotyping, whole-exome sequencing, and phenotype-matching algorithms (*Fig 1*). Clinical whole-exome sequencing has improved our ability to discover genetic causes for various rare undiagnosed disorders although the currently diagnostic rate is 25%~40% [11,36]. Hurdles remain for the establishment of genotype-phenotype correlation in the context of genetic evaluation [36,37]. In the clinical practice, the use of HPO in annotating phenotypic information remains unexplored [38]. Here, we compiled multiple clinic features from Chinese electronic medical records using HPO-based deep phenotyping (*Fig 2*). Together with the filtered pathogenic mutations, we obtained an improved diagnostic yield of 49%, and identified a large proportion of *de novo* and compound heterozygosity mutations underlying these pedigrees (*Fig 3* and *Table 2*).

The current study refined and extended the mutational and phenotypic spectrum of the known neurodevelopmental disorders. We reported, for the first time, at least 13 novel pathogenic mutations in the known genes associated with neurodevelopmental disorders, providing a basis for further functional validation and/or the elucidation of molecular mechanism. In addition, we characterized a cardiac phenotype (i.e., atrial septal defect) in both affected siblings in *UDP #30* diagnosed with EIEE28 with the compound heterozygous mutations in *WWOX* . Expanding the mutational spectrum in the causal genes and phenotypic spectrum of neurodevelopmental disorders could offer accurate and reliable genetic counseling and prenatal diagnosis to the patients, and thereby minimize the birth of affected individuals [39].

Our results provide take-home messages for health professionals in the clinical management of patients with neurodevelopment disorders. The annotated phenotypes recapitulate syndromic features of neurodevelopmental disorders, which cannot be neglected in clinical diagnosis and genetics evaluation. Phenotypes

associated with non-nervous system may have an increased power to make a genetic diagnosis, and precision phenotyping enables revealing causal genes with pleiotropic effects [40]. Furthermore, return of actionable genetic findings for genetic counselor will accelerate prenatal genetic diagnosis, and control the risk/burden for birth defects [41,42].

There are several explanations for the remaining 23 undiagnosed pedigrees, including five child-parent quads. First, deep phenotyping based on electronic medical records cannot extract all phenotypes related to patients, such as phenotypes associated with non-nervous system. Second, the selection of phenotype according to the hierarchical ontology has an obvious effect on phenotype-matching (e.g., UDP #9). An algorithm by iterating all corresponding ancestral terms may enable identifying the most matched disorder. Finally, whole-exome sequencing cannot capture genetic aberrations outside the exon regions, and a sequencing depth of approximately 30× has limited power to accurately identify copy number variants [43,44]. Therefore, the use of whole-genome sequencing for clinical diagnosis for rare undiagnosed pedigree is urgently needed.

Conclusions

The combined use of deep phenotyping and whole-exome sequencing increased the diagnostic yield for nuclear pedigrees with neurodevelopmental disorders. Our analysis may provide an avenue for shortening the diagnostic odyssey for such rare undiagnosed disease in the clinical setting, and an economical and practical approach will be widely applied for their evaluation of genetic etiology.

Ethics, consent and permissions

The study was approved by the Institutional Research Board (IRB) at the Henan Provincial People's Hospital, and all participants or their guardians signed the informed consent.

Consent to publish

We have confirmed that we have obtained consent to publish from the participant (or their guardians).

Availability of data and materials

The datasets supporting the conclusions of this article are included within the article and its additional files.

Acknowledgements

This work was supported by National Natural Science Foundation of China (No. 81650010, S.L.), the Science and Technology Cooperation Project of Henan Province (No.182106000058, S.L.), and a grant from National Health Commission Key Laboratory of Birth Defects Prevention (ZD201907, K.D.).

Competing interests

The authors disclose no conflicts of interest.

Authors contributions

Q.W., K.Y., X.H., and S.L. recruited pedigrees, Q.W. conducted deep phenotyping with guidance from K.D., X.T. and K.D. performed computation and data analysis, Q.W. established genotype-phenotype correlation with input from H.Z., K.D., and S.L., X.T. prepared the graphics and tables with input from Q.W., K.D. and Q.W. wrote the manuscript with input from X.T. and S.L., and K.D. and S.L. conceived and designed the project.

Web resources

CHPO: <http://www.chinahpo.org/>

HPO: <https://hpo.jax.org/app/>

GATK: <https://gatk.broadinstitute.org/hc/en-us>

gnomAD: <https://gnomad.broadinstitute.org/>

Exomiser: <https://www.sanger.ac.uk/tool/exomiser/>

Phenolyzer: <http://phenolyzer.wglab.org/>

InterVar: <https://github.com/WGLab/InterVar>

Clinvar: <https://www.ncbi.nlm.nih.gov/clinvar/>

OMIM: <https://omim.org/>

HGMD: <http://www.hgmd.cf.ac.uk/ac/index.php>

References

1. Bellman M, Byrne O, Sege R. Developmental assessment of children. *The BMJ*. 2013;346:e8687–7.
2. Regier DA, Kuhl EA, Kupfer DJ. The DSM-5: Classification and criteria changes. *World Psychiatry*. 2013;12:92–8.
3. Gahl WA, Markello TC, Toro C, Fajardo KF, Sincan M, Gill F, et al. The national institutes of health undiagnosed diseases program: insights into rare diseases. *Genetics in Medicine*. 2012;14:51–9.
4. Soden SE, Saunders CJ, Willig LK, Farrow EG, Smith LD, Petrikin JE, et al. Effectiveness of exome and genome sequencing guided by acuity of illness for diagnosis of neurodevelopmental disorders. *Science Translational Medicine*. 2014;6:265ra168–8.
5. Shashi V, McConkie-Rosell A, Rosell B, Schoch K, Vellore K, McDonald M, et al. The utility of the traditional medical genetics diagnostic evaluation in the context of next-generation sequencing for undiagnosed genetic disorders. *Genetics in Medicine*. 2014;16:176–82.
6. 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. *American Journal of Human Genetics*. 2010;86:749–64.
7. 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. *European Journal of Paediatric Neurology*. 2013;17:589–99.
8. Deciphering Developmental Disorders Study. Prevalence and architecture of de novo mutations in developmental disorders. *Nature*. 2017;542:433–8.
9. Srivastava S, Cohen JS, Vernon H, Barañano K, McClellan R, Jamal L, et al. Clinical whole exome sequencing in child neurology practice. *Annals of Neurology*. 2014;76:473–83.
10. Tetreault M, Bareke E, Nadaf J, Alirezaie N, Majewski J. Whole-exome sequencing as a diagnostic tool: current challenges and future opportunities. *Expert Review of Molecular Diagnostics*. 2015;15:749–60.
11. Thevenon J, Duffourd Y, Masurel-Paulet A, Lefebvre M, Feillet F, El Chehadé-Djebbar S, et al. Diagnostic odyssey in severe neurodevelopmental disorders: toward clinical whole-exome sequencing as a first-line diagnostic test. *Clinical Genetics*. 2016;89:700–7.

12. Nolan D, Carlson M. Whole exome sequencing in pediatric neurology patients. *Journal of Child Neurology*. 2016;31:887–94.
13. Evers C, Staufner C, Granzow M, Paramasivam N, Hinderhofer K, Kaufmann L, et al. Impact of clinical exomes in neurodevelopmental and neurometabolic disorders. *Molecular Genetics and Metabolism*. 2017;121:297–307.
14. Wright CF, McRae JF, Clayton S, Gallone G, Aitken S, Fitzgerald TW, et al. Making new genetic diagnoses with old data: iterative reanalysis and reporting from genome-wide data in 1,133 families with developmental disorders. *Genetics in Medicine*. 2018;20:1216–23.
15. Nambot S, Thevenon J, Kuentz P, Duffourd Y, Tisserant E, Bruel A-L, et al. Clinical whole-exome sequencing for the diagnosis of rare disorders with congenital anomalies and/or intellectual disability: substantial interest of prospective annual reanalysis. *Genetics in Medicine*. 2018;20:645–54.
16. Robinson PN. Deep phenotyping for precision medicine. *Human Mutation*. 2012;33:777–80.
17. Köhler S, Carmody L, Vasilevsky N, Jacobsen JOB, Danis D, Gouridine J-P, et al. Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources. *Nucleic Acids Research*. 2019;47:D1018–27.
18. Yang H, Robinson PN, Wang K. Phenolyzer: phenotype- based prioritization of candidate genes for human diseases. *Nature Methods*. 2015;1–6.
19. Smedley D, Jacobsen JOB, Jäger M, Köhler S, Holtgrewe M, Schubach M, et al. Next-generation diagnostics and disease-gene discovery with the Exomiser. *Nature Protocols*. 2015;10:2004–15.
20. Snijders Blok L, Hiatt SM, Bowling KM, Prokop JW, Engel KL, Cochran JN, et al. De novo mutations in MED13, a component of the Mediator complex, are associated with a novel neurodevelopmental disorder. *Human Genetics*. 2018;137:375–88.
21. Chen L, Song L, Shao Y, Li D, Ding K. Using natural language processing to extract clinically useful information from Chinese electronic medical records. *International Journal of Medical Informatics*. 2019;124:6–12.
22. Tang X, Chen W, Zeng Z, Ding K, Zhou Z. An ontology-based classification of Ebstein’s anomaly and its implications in clinical adverse outcomes. *International Journal of Cardiology*. 2020; <https://doi.org/10.1016/j.ijcard.2020.04.073>
23. Greene D, Richardson S, Turro E. ontologyX: a suite of R packages for working with ontological data. *Bioinformatics*. 2017;33:1104–6.
24. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25:1754–60.
25. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Angel G del, Levy-Moonshine A, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Current Protocols in Bioinformatics*. 2013;43:11.10.1–11.10.33.
26. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine*. 2015;17:405–23.
27. Li Q, Wang K. InterVar: Clinical interpretation of genetic variants by the 2015 ACMG-AMP guidelines. *Am J Hum Genet*. 2017;100:267–80.
28. The 1000 Genomes Project Consortium, Boerwinkle E, Doddapaneni H, Han Y, Korchina V, Kovar C, et al. A global reference for human genetic variation. *Nature*. 2015;526:68–74.
29. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016;536:285–91.

30. Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Research*. 2013;42:D980–5.
31. Stenson PD, Mort M, Ball EV, Shaw K, Phillips AD, Cooper DN. The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. *Human Genetics*. 2013;133:1–9.
32. Robinson PN, Köhler S, Oellrich A, Sanger Mouse Genetics Project, Wang K, Mungall CJ, et al. Improved exome prioritization of disease genes through cross-species phenotype comparison. *Genome Research*. 2014;24:340–8.
33. Köhler S, Øien NC, Buske OJ, Groza T, Jacobsen JOB, McNamara C, et al. Encoding clinical data with the Human Phenotype Ontology for computational differential diagnostics. *Current Protocols in Human Genetics*. 2019;103:e92.
34. Barnett K, Mercer SW, Norbury M, Watt G, Wyke S, Guthrie B. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. *Lancet*. 2012;380:37–43.
35. Davids M, Markello T, Wolfe LA, Chepa-Lotrea X, Tiffit CJ, Gahl WA, et al. Early infantile-onset epileptic encephalopathy 28 due to a homozygous microdeletion involving the WWOX gene in a region of uniparental disomy. *Human Mutation*. 2018;40:42–7.
36. Trinh J, Kandaswamy KK, Werber M, Weiss MER, Oprea G, Kishore S, et al. Novel pathogenic variants and multiple molecular diagnoses in neurodevelopmental disorders. *Journal of Neurodevelopmental Disorders*. 2019;11:11–6.
37. O'Donnell-Luria AH, Miller DT. A Clinician's perspective on clinical exome sequencing. *Human Genetics*. 2016;135:643–54.
38. Aitken S, Firth HV, McRae J, Halachev M, Kini U, Parker MJ, et al. Finding diagnostically useful patterns in quantitative phenotypic data. *American Journal of Human Genetics*. 2019;105:933–46.
39. Katsanis SH, Katsanis N. Molecular genetic testing and the future of clinical genomics. *Nature Reviews Genetics*. 2013;14:415–26.
40. Wang W, Corominas R, Lin GN. De novo Mutations from whole exome sequencing in neurodevelopmental and psychiatric disorders: from discovery to application. *Frontiers in Genetics*. 2019;10:258.
41. Kullo IJ, Olson J, Fan X, Jose M, Safarova M, Radecki Breitkopf C, et al. The return of actionable variants empirical (RAVE) study, a mayo clinic genomic medicine implementation study: design and initial results. *Mayo Clinic Proceedings*. 2018;93:1600–10.
42. Bick D, Jones M, Taylor SL, Taft RJ, Belmont J. Case for genome sequencing in infants and children with rare, undiagnosed or genetic diseases. *Journal of Medical Genetics*. 2019;56:783–91.
43. Zhang L, Bai W, Yuan N, Du Z. Comprehensively benchmarking applications for detecting copy number variation. *PLoS Computational Biology*. 2019;15:e1007069.
44. Hwang MY, Moon S, Heo L, Kim YJ, Oh JH, Kim Y-J, et al. Combinatorial approach to estimate copy number genotype using whole-exome sequencing data. *Genomics*. 2015;105:145–9.

Tables

Table 1. Clinical characteristics of the recruited pedigrees.

Table 2. Probably and possibly genetic diagnosis for nuclear pedigrees with neurodevelopmental disorders.

Figure legends

Fig 1. - A framework of clinical genetic evaluation for ‘not yet diagnosed’ nuclear pedigrees with neurodevelopmental disorders.

Fig 2. - Deep phenotyping characterized phenotypes for pedigrees with neurodevelopmental disorders, which were standardized using HPO terms. (A). An ontology plot indicated the relationship of phenotypes. The color represents the frequency of a term in the HPO database (dark green: high; and yellow: low). The solid circle represents the abnormal phenotype present in the recruited pedigrees; (B). The distribution of HPO terms grouped by nervous and non-nervous systems; (C). The distribution of HPO terms per pedigree; (D). The distribution of HPO terms associated with nervous system. The terms under the lineage of nervous system physiology were grouped using DSM-5; and (E). The distribution of HPO terms associated with non-nervous system.

Fig 3. - Genetic diagnosis for nuclear pedigrees with neurodevelopmental disorders. (A). The diagnosis of 22 pedigrees and their inheritance patterns; (B). The distribution of compound heterozygosity and *de novo* mutations; (C-D). An increased *Phenotype* and *Exomiser* scores when a given term was replaced with its corresponding ancestral term according to the ontology; (E). The comparison of the rank of the prioritized genes identified in *Exomiser* and *phenolyzer* ; and (F). The distribution of HPO terms in diagnosed and undiagnosed pedigrees. ***, $p < 0.001$; NS, not significant.

Fig 4. - Case examples. (A). A landscape of HPO terms; and (B-F). Sanger sequencing confirmed the genetic diagnoses for five pedigrees.

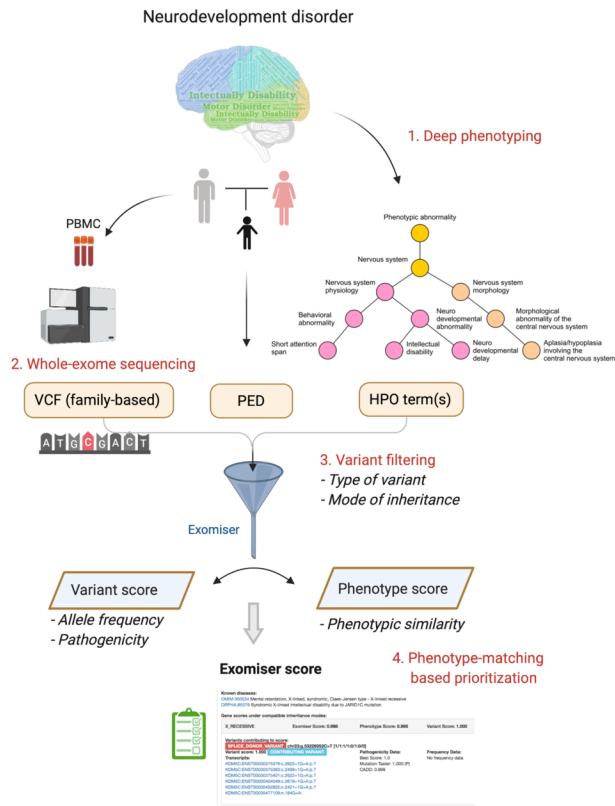


Fig. 1.

Hosted file

image2.emf available at <https://authorea.com/users/342907/articles/469687-deep-phenotyping-and-whole-exome-sequencing-improved-the-diagnosis-yield-for-nuclear-pedigrees-with-neurodevelopmental-disorders>

Fig. 2.

Hosted file

image3.emf available at <https://authorea.com/users/342907/articles/469687-deep-phenotyping-and-whole-exome-sequencing-improved-the-diagnosis-yield-for-nuclear-pedigrees-with-neurodevelopmental-disorders>

Fig. 3.

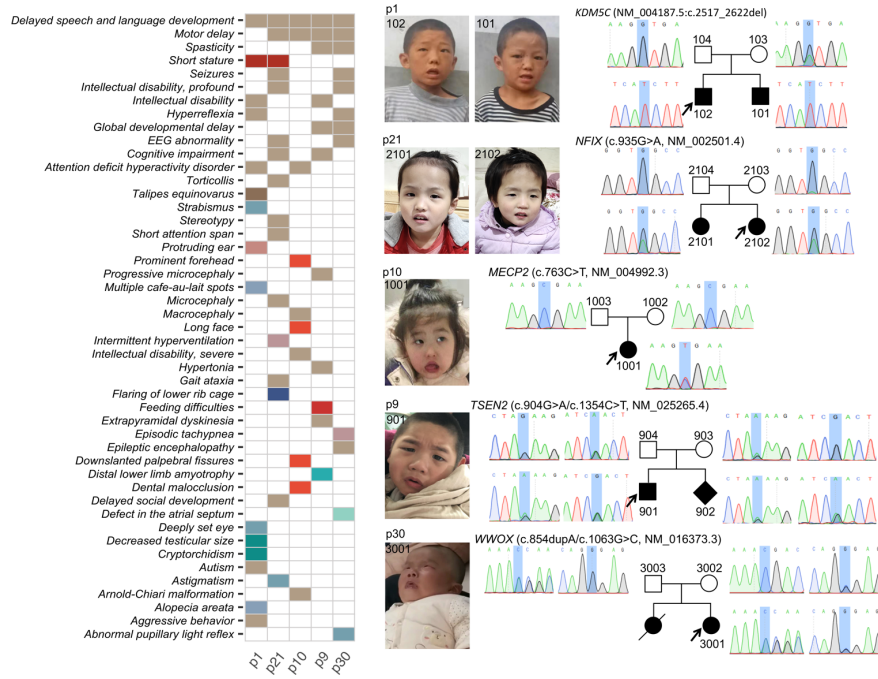


Fig. 4.

Hosted file

Table1.docx available at <https://authorea.com/users/342907/articles/469687-deep-phenotyping-and-whole-exome-sequencing-improved-the-diagnosis-yield-for-nuclear-pedigrees-with-neurodevelopmental-disorders>

Hosted file

Table2_0714.docx available at <https://authorea.com/users/342907/articles/469687-deep-phenotyping-and-whole-exome-sequencing-improved-the-diagnosis-yield-for-nuclear-pedigrees-with-neurodevelopmental-disorders>