



OPEN

The diagnostic yield, candidate genes, and pitfalls for a genetic study of intellectual disability in 118 middle eastern families

Ghalia Al-Kasbi¹, Fathiya Al-Murshedi^{1,2}, Adila Al-Kindi^{1,2}, Nadia Al-Hashimi³, Khalid Al-Thihli^{1,2}, Abeer Al-Saegh^{1,2}, Amna Al-Futaisi⁴, Watfa Al-Mamari⁴, Abdullah Al-Asmi⁵, Zandre Bruwer², Khalsa Al-Kharusi², Samiya Al-Rashdi¹, Fahad Zadjali⁶, Said Al-Yahyaee¹ & Almundher Al-Maawali^{1,2}✉

Global Developmental Delay/Intellectual disability (ID) is the term used to describe various disorders caused by abnormal brain development and characterized by impairments in cognition, communication, behavior, or motor skills. In the past few years, whole-exome sequencing (WES) has been proven to be a powerful, robust, and scalable approach for candidate gene discoveries in consanguineous populations. In this study, we recruited 215 patients affected with ID from 118 Middle Eastern families. Whole-exome sequencing was completed for 188 individuals. The average age at which WES was completed was 8.5 years. Pathogenic or likely pathogenic variants were detected in 32/118 families (27%). Variants of uncertain significance were seen in 33/118 families (28%). The candidate genes with a possible association with ID were detected in 32/118 (27%) with a total number of 64 affected individuals. These genes are novel, were previously reported in a single family, or cause strikingly different phenotypes with a different mode of inheritance. These genes included: *AATK*, *AP1G2*, *CAMSAP1*, *CCDC9B*, *CNTROB*, *DNAH14*, *DNAJB4*, *DRG1*, *DTNBP1*, *EDRF1*, *EEF1D*, *EXOC8*, *EXOSC4*, *FARSB*, *FBXO22*, *FILIP1*, *INPP4A*, *P2RX7*, *PRDM13*, *PTRHD1*, *SCN10A*, *SCYL2*, *SMG8*, *SUPV3L1*, *TACC2*, *THUMPD1*, *XPR1*, *ZFYVE28*. During the 5 years of the study and through gene matching databases, several of these genes have now been confirmed as causative of ID. In conclusion, understanding the causes of ID will help understand biological mechanisms, provide precise counseling for affected families, and aid in primary prevention.

Global Developmental Delay (GDD)/Intellectual disability (ID) represents a group of genetic, phenotypic, and clinically heterogeneous disorders that affect approximately 1% of children worldwide¹. Significant limitations define GDD or ID in both intellectual functioning and adaptive behavior that originates during brain development. Non-genetic causes such as infections, autoimmunity, and environmental factors are described, but the majority of such disorders have a genetic basis².

Hundreds of genes are thought to be involved in the etiology of ID³. The list of ID genes has expanded and according to the SysNDD database, there are now 2841 primary and candidate human ID genes⁴. In the last decade, advances in genetic technologies such as next-generation sequencing (NGS) have revolutionized clinical practice in medical genetics, aided clinical diagnosis, and proved to be very effective in discovering an ever-increasing number of ID-related genes. They have also enabled deciphering the ID's heterogeneous genetic mechanisms⁵.

Whole-Exome Sequencing (WES), as a clinical diagnostic test, has a success rate of about 30–40%⁶. The diagnostic yield of chromosomal microarray in children with no underlying cause of their ID is around 15 to 20%⁷. In

¹Department of Genetics, College of Medicine and Health Sciences, Sultan Qaboos University, Muscat, Oman. ²Genetic and Developmental Medicine Clinic, Sultan Qaboos University Hospital, Muscat, Oman. ³Department of Pediatrics, Royal Hospital, Ministry of Health, Muscat, Oman. ⁴Department of Child Health, College of Medicine and Health Sciences, Sultan Qaboos University, Muscat, Oman. ⁵Department of Medicine, College of Medicine and Health Sciences, Sultan Qaboos University, Muscat, Oman. ⁶Department of Biochemistry, College of Medicine and Health Sciences, Sultan Qaboos University, Muscat, Oman. ✉email: almaawali@squ.edu.om

trio-based WES done in groups of children with severe ID, the yield ranged from 13 to 35%⁸. In contrast, exome sequencing in samples from consanguineous families with various ID-associated phenotypes has produced a high yield. For example, in several studies of ID from the Middle East, NGS yield ranged from 37 to 90%, depending on the patients' cohorts, the study design, and the classification of the variants^{9–15}.

This study reports on the diagnostic yield and candidate genes of a genetic study of intellectual disability in 118 Omani families. The study period extended over five years and included 215 affected individuals. The diagnostic yield when considering both pathogenic and uncertain variants was 55%. Candidate genes with a possible association with ID phenotype seen in this cohort were detected in 32/118 (27%) with a total number of 64 affected individuals. Understanding ID causes will provide precise counseling for affected families and aid in primary prevention. Also, the costs of unnecessary investigations will be spared, with fewer diagnostic odysseys. The paper also discusses the pitfalls and challenges of candidate gene discovery in a consanguineous population.

Materials and methods

Human subjects. The Medical Research Ethics Committee approved the study of Sultan Qaboos University (SQU MREC#1362). Informed written consent was obtained from all participants or their guardians. All methods were performed in accordance with the relevant guidelines and regulations and in accordance with the declaration of Helsinki. The target patients included in this study presented with global developmental delay or intellectual disability, all assessed clinically by medical geneticists (detailed methods in SUPP_S1). Severe phenotypes causing death within the neonatal or early infantile period were also included as a neurological phenotype was evident, such as seizures, hypotonia, or brain malformations. Families with a likely autosomal recessive pattern of inheritance were selected. Patients with known molecular diagnoses at the time of recruitment were excluded. The study and exome data analysis were carried out over 5 years, between 2016 and 2021.

Whole exome sequencing and variant interpretation. Whole-exome sequencing analysis was performed for all affected individuals where samples were available. A detailed methodology is presented in the supplemental data (SUPP_S1). In brief, the method used hybrid capture technology (Agilent SureSelect Human All-exons-V6 or V7) for exome enrichment and capture. Illumina technology (HiSeq2500, HiSeq4000, or NovaSeq6000) of 150 bp paired-end, at 150–200X coverage, was used for sequencing. The reads were mapped against UCSC GRCh37/hg19 or GRCh38/hg38. Filtering and variant prioritization were analyzed using an in-house pipeline (SUPP_S1). Variant filtration was performed to keep only novel or rare variants ($\leq 1\%$). Public databases such as 1000 Genomes, Exome Variant Server, and GnomAD were used for alleles frequencies. For filtration of common variants against the Middle Eastern population, the Greater Middle East (GME) and variome database “al mena” that comprises data of 2497 samples was used^{16,17}. Our in-house population-specific exomes database, which contains data of 1564 WES, was also used. During the filtration process, the phenotype and mode of inheritance were both considered. Any potential variants identified after prioritization were further confirmed by Sanger sequencing. Members included in the segregation analysis ranged from 3 to 12 members of each family, depending on the DNA availability. Most of the segregation was performed for the parents and siblings alongside the index patient. When a definitive cause was not possible or when a candidate gene was considered, further analysis of copy number variants was performed on exome data using ExomeDepth¹⁸ (SUPP_S1).

Classification of variants was based on the published ACMG guideline^{19,20}. Pathogenic or likely pathogenic variants in known disease-causing genes which could be linked to the reported phenotypes of the affected patients were categorized as disease-causing variants. The second category was for the variants in known disease-causing genes that overlapped with the patient's phenotype, and these were considered possible disease-causing variants. These were rare and damaging variants of uncertain significance (VUS). Variants in candidate genes, which were predicted to be deleterious and found in genes not previously confirmed to be implicated in human disease, formed the third category. These genes are novel; they were previously reported in a single family or cause strikingly different phenotypes with different modes of inheritance. Supporting data for candidate genes included variants within a shared autozygosity area. The variant is of high or moderate impact; the population frequency supports genic intolerance to such variants, and in-silico prediction tools indicate a damaging effect. Data for gene function and network, gene expression, and animal models were also considered.

Results

Whole-exome sequence analysis was performed for 188 individuals representing 118 characterized families with a total number of 215 affected individuals. Of the 118 families included, 93 (78.8%) had a family history of one or more affected individuals in addition to the index patient, all with a similar phenotype. The age range of the affected individuals, at first clinical assessment, was from birth to 34 years old. The average age at which WES was completed was 8.5 years, and children below five years of age, represented 30% of the affected individuals when WES was completed. Males represented 57.2% of the studied group (123 Males:92 Females). The rate of consanguineous marriages within the included families was 91%. The affected patients exhibited diverse phenotypes, including global developmental delay, seizures, brain malformations, microcephaly, facial dysmorphism, and other systemic manifestations (Table 1 and Supplementary SUPP_Table1).

A total of 420 members' DNA samples were available for WES or Sanger sequencing for segregation analysis. These included healthy or affected members. However, DNA samples were not available for analysis in 22 out of the 215 affected individuals. Sanger sequencing was used to confirm the variant and phenotype-genotype segregation in all candidate variants. Only variants that were confirmed and segregated with the phenotype are reported.

Variants in previously known and described ID genes were seen in 65/118 families (55%). Following the ACMG guidelines of variant classification, pathogenic (P) or likely pathogenic (LP) variants were detected in 32/118 families (27%). Variants of uncertain significance were seen in 33/118 families (28%). The majority of

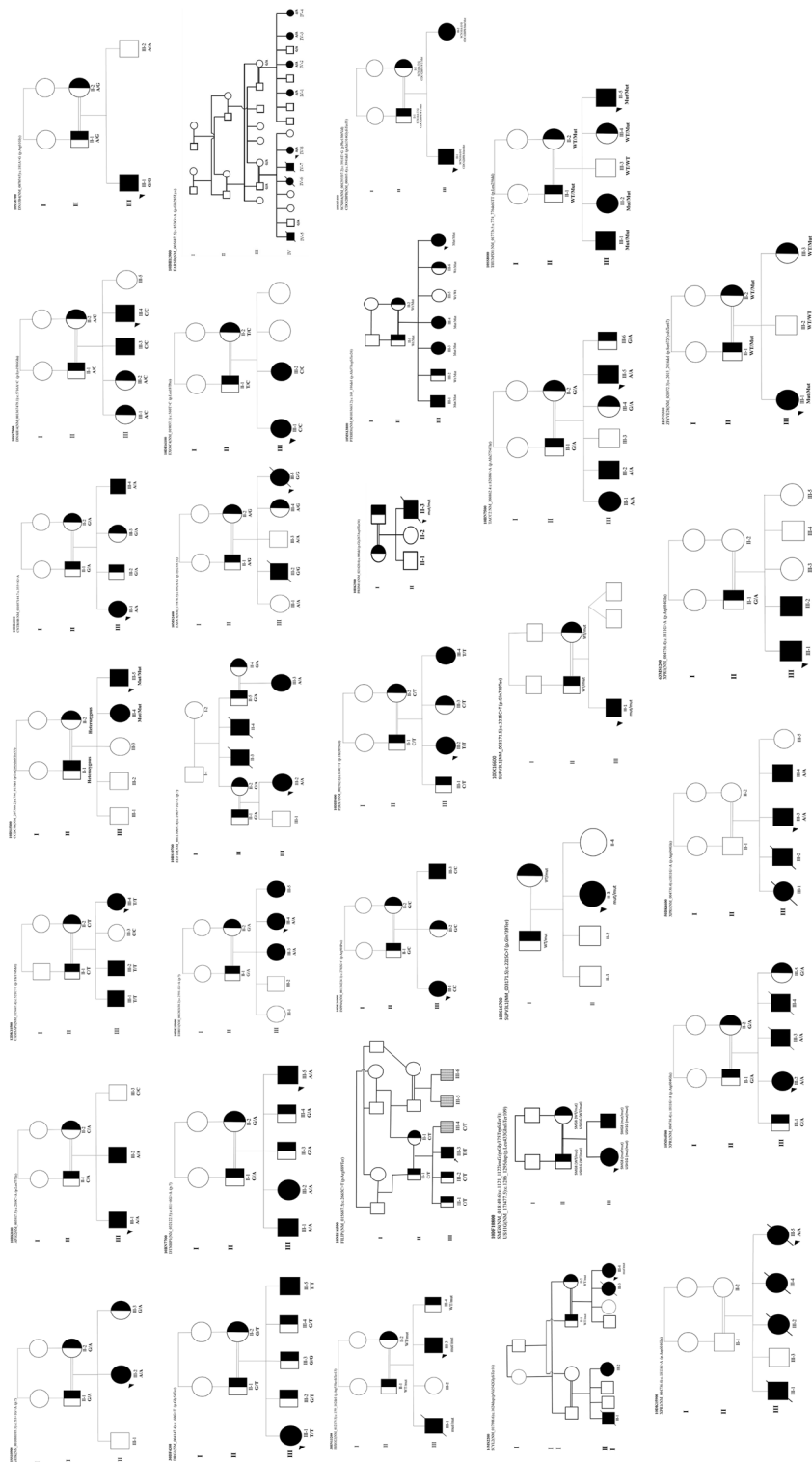


Figure 1. Pedigrees of families with candidate genes, showing the variants. Shaded symbols indicate affected individuals and arrows indicate the proband, zygosity for the variant is included.

Family-ID	Index's Age at exam test	Index's Gender	Other Affected	Phenotype	Gene-Approved	HGVs variant (hg19)	Zygosity	Consanguinity	avsnp150 rs	GnomAD frequency	Pathogenicity Scores *	Function & experimental models	Remarks
11SS1090	2 Years	Female	No	Refractory seizures and developmental delay	AATK	AATK(NM_001080395.3): c.533 + 1G > A (p.?)	homozygous	Yes	NA	Absent	NA	Apoptosis-associated tyrosine kinase (AATK) is a non-receptor type tyrosine kinase whose expression is up-regulated during the apoptosis and differentiation of 32D33 myeloblastic cells. AATK plays an essential role in mature neurons and neuronal differentiation ⁶⁴	No published genotype-phenotype association
10DK8100	12 Years	Male	Yes	Autistic features, intellectual disability, hyperactivity, and normal MRI brain	APIG2	APIG2(NM_003917.5): c.2269C > A (p.Leu757Ile)	homozygous	Yes	rs764435650	0.0000319	0/4	The APIG2 gene encodes gamma-2-adaptin, a protein very similar to gamma-adaptin (APIG1), a subunit of the AP1 adaptor complex that engages clathrin to the trans-Golgi network. It might have a role in protein sorting in late endosomes or multivesicular bodies ⁶⁵ . Zebrafish whole-mount in situ hybridization confirmed that y2 adaptin (encoded by APIG2) is detected in the developing central nervous system (CNS) and knockout in zebrafish embryos showed severe, lethal impairment of CNS development ⁶⁶ . The partner protein APIG1 was recently published to cause ID ⁶⁴	No published genotype-phenotype association
12DK11500	11 Years	Male	Yes	Facial dysmorphic features, severe neurodevelopmental delay, cortical visual impairment and seizures	CAMSAP1	CAMSAP1(NM_015447.4): c.521C > T (p.Thr174Met)	homozygous	Yes	rs1187560194	0.0000319	1/4	CAMSAP1 is a microtubule-organizing protein that binds to the minus end of non-centrosomal microtubules and regulates their dynamics and organization ⁶⁷ . It is involved in other processes, including neuron projection development and regulation of cell morphogenesis ⁶⁸	Gene Matched (understudy)
10BS15400	1 Year	Male	Yes	Severe intellectual disability, seizures, face dysmorphism, hypopigmented hair and skin	CCDC9B	CCDC9B(NM_207380.2): c.796_815del (p.Leu266AlafTer35)	homozygous	Yes	NA	Absent	NA	CCDC9B (C15orf52) is expressed in the brain. Mouse KO data indicated a neurological phenotype. It showed decreased exploration in new environments and decreased prepulse inhibition (MGI:2685199)	No published genotype-phenotype association
10DH6800	14 Years	Female	Yes	Short stature, microcephaly, cataract, abnormal teeth, obesity, hypertension, and insulin resistance	CNTROB	CNTROB(NM_001037144.7): c.355 + 1G > A (p.?)	homozygous	Yes	rs752440318	7.95E-06	NA	CNTROB is a centrosomal protein involved in centriole duplication and cytokinesis ⁶⁹ . It is expressed in the brain. It was observed that inhibiting centriole duplication causes cytokinesis defects. In mice, in hippocampal cells, centrin formed cytoplasmic dots. It enhanced microtubule formation outside as well as inside the centrosome	Gene Matched (understudy)
10S57900	7 Years	Male	Yes	Intellectual disability	DNAH14	DNAH14(NM_001367479.1): c.5716A > C (p.Lys1906Gln)	homozygous	Yes	rs868380191	Absent	4/4	DNAH14 encodes a microtubule-associated motor protein that contributes to centrosome integrity. DNAH14 was found to be a causative gene in patients with hydrocephalus. Since DNAH14 encodes axonemal dynein, which is found in motile cilia it is believed to affect cilia physiological function during hydrocephalus pathogenesis ⁷⁰ . Mutation in DNAH14 was also linked to human embryonic lethality ⁷¹	One study was published recently with no further confirmation ⁷² , and no associated OMIM phenotype entry
10SN8700	6 Years	Male	Yes	IUGR, recurrent pneumonia (lipoid), ventilator-dependent, inguinal hernia, thin corpus callosum and mild to moderate brain atrophy	DNAJB4	DNAJB4(NM_007034.5): c.181A > G (p.Arg61Gly)	homozygous	Yes	NA	Absent	4/4	DNAJB4 is a member of the highly conserved DNAJ/HSP40 protein family. It functions as a chaperone for E-cadherin, allowing appropriate folding, localization, and stability, implying a post-transcriptional level of control ⁷³ . DNAJB4 binds to the carboxyl tail of the human mu opioid receptor in the CNS ⁷⁴ . DNAJB4 was found to interact with SDIM1, which is thought to play a particular role in brain cell survival and/or receptor trafficking ⁷⁵	No published genotype-phenotype association
30DF4200	12 Years	Female	Yes	Intellectual disability, microcephaly, facial dysmorphism, micropthalmia and eczema	DRG1	DRG1(NM_004147.4): c.160G > T (p.Gly54Ter)	homozygous	Yes	NA	Absent	NA	DRG1 is a microtubule-binding protein that performs various microtubule-related functions ⁷⁶ . Based on its expression pattern, DRG1 may be involved in neuronal cell proliferation and/or differentiation ⁷⁷ . DRG1 transcript levels in the developing central nervous system of <i>Xenopus laevis</i> revealed a similar pattern of expression ⁷⁸	Gene Matched (understudy)
10BN7700	7 Years	Male	Yes	Normocephalic, global developmental delay, and spastic paraplegia	DTNBP1	DTNBP1(NM_032122.5): c.811 + 6G > A (p.?)	homozygous	Yes	NA	Absent	NA	Dysbindin protein is a significant component of the biogenesis of lysosome-related organelles complex-1 (BLOC-1), which controls protein trafficking in the lysosomal pathway ⁷⁹ . DTNBP1 has been identified to have a role in the development of schizophrenia by affecting brain development and neurotransmission activity ⁸⁰	Studies associated DTNBP1 variants with schizophrenia with no further functional confirmation ⁸¹
10DK15000	12 Years	Female	Yes	Intellectual disability	EDRF1	EDRF1(NM_001202438.2): c.2391 - 1G > A (p.?)	homozygous	Yes	NA	Absent	NA	EDRF1 is a transcription factor ⁸² ; it is highly expressed in the brain	No published genotype-phenotype association

Continued

Family-ID	Index's Age at exome test	Index's Gender	Other Affected	Phenotype	Gene-Approved	HGVs variant (hg19)	Zygosity	Consanguinity	avsnp150 rs	GnomAD frequency	Pathogenicity Scores *	Function & experimental models	Remarks
10BS10700	7 Years	Female	Yes	Global developmental delay, microcephaly, severe visual impairment, hypomyelination and hypoplastic corpus callosum	EEF1D	EEF1D(NM_001130053.4):c.1905 + 1G > A (p.?)	homozygous	Yes	rs1364060781	3.99E-06	NA	EEF1D gene encodes three different splice forms proteins in addition to a more recently described isoform called eEF1B δ , that functions during cellular stress as a transcription factor for heat shock element-containing genes ⁴¹ . Mutations in the EEF1D gene have been associated with neurodevelopmental abnormalities ⁴²	Gene Matched (understudy)
10MS2400	4 Years	Male	Yes	Severe acquired microcephaly, severe delay, seizures, optic nerve atrophy, and hypotonia	EXOC8	EXOC8(NM_175876.5):c.692A > G (p.Tyr231Cys)	homozygous	Yes	rs1558358137	Absent	4/4	EXOC8 is a subunit of the exocyst complex involved in docking exocytic vesicles with fusion sites on the plasma membrane. It is found to play an essential role in normal cortical development and its perturbation causes complex brain disorders. EXOC8 mutations (one family) are linked to neurodevelopmental disorders with microcephaly, seizures, and brain atrophy ⁶³	?Neurodevelopmental disorder with microcephaly, seizures, and brain atrophy (MIM# 619076). Two studies correlated EXOC8 variants with a neurodevelopmental phenotype ⁶⁴ , functional studies of the variant and studies of patient cells were not performed
10DF16100	14 Years	Female	Yes	Prenatal growth restrictions, failure to thrive, short stature, global developmental delay, intracerebral calcification and basal ganglia calcifications, and kidney failure	EXOSC4	EXOSC4(NM_019037.3):c.560 T > C (p.Leu187Pro)	homozygous	Yes	NA	Absent	2/4	EXOSC4 is a non-catalytic component of the RNA exosome complex with 3' > 5' exoribonuclease activity and involvement in numerous cellular RNA processing and degradation events ⁶⁵ . EXOSC4 is involved in the nucleic acid metabolic process and positive regulation of cell growth ⁶⁶ . Mutations in other 3' > 5' exoribonucleases enzymes were associated with a neurodevelopmental disorder ^{67,68}	Gene Matched (understudy)
10DH13900	28 Years	Female	Yes	Brain calcification, microcephaly, short stature and lung disease	FARSB	FARSB(NM_005687.5):c.853G > A (p.Glu285Lys)	homozygous	Yes	rs767956337	Absent	1/4	Cytoplasmic phenylalanine-tRNA synthetase is a heterodimer composed of a catalytic alpha subunit, FARSA, and a regulatory beta subunit, FARSB. FARSB is an aminoacyl-tRNA synthetase that charges tRNAs with specific amino acids. It is reported to be associated with neurodevelopmental disorders ^{69,70}	Published ⁶⁹
30BN13200	3 Years	Male	Yes	Symmetrical IUGR, clenched hands, hypertelorism, generalized hypotonia, duodenal atresia and annular pancreas	FRXO22	FRXO22(NM_012170.3):c.159_162del (p.Arg53SerfsTer13)	homozygous	Yes	NA	Absent	NA	FRXO22 is a Member of the F-box protein family, which is characterized by comprising 40-amino acid F-box domain. F-box proteins form complexes with other proteins and act as protein-ubiquitin ligases. FRXO22 plays a critical in carcinogenesis and may participate in the progression of parkinsonism ^{1,71}	Gene Matched (understudy)
10MS16500	1 Years	Male	Yes	Distal arthrogryposis with contractures of the knees and elbows, congenital club-foot, muscular hypotonia, and mild learning disability	FILIP1	FILIP1(NM_015687.5):c.2665C > T (p.Arg889Ter)	homozygous	Yes	rs775141616	1.19E-05	NA	FILIP1 regulates the migration of neocortical cells from the ventricular zone by acting via a filamin-AF-actin axis. It may cause filamin-A degradation ^{1,74}	Gene Matched (understudy)
10DK16000	8 Months	Female	Yes	Severe Global developmental delay and pontocerebellar hypoplasia	INPP4A	INPP4A(NM_001134224.1):c.2702G > C (p.Arg901Pro)	homozygous	Yes	NA	Absent	4/4	INPP4A catalyzes the hydrolysis of the 4-position phosphate of phosphatidylinositol 3,4-bisphosphate, as well as inositol 1,3,4-trisphosphate and inositol 1,4-bisphosphate ⁷⁵ . INPP4A-null mice developed neurodegeneration in the striatum and severe involuntary movement disorders (MGI:1,931,123)	One study identified a novel nonsense INPP4A mutation in a family with intellectual disability ⁷⁶ , with no further functional confirmation and no associated OMIM phenotype entry
10DH9400	11 Years	Female	Yes	Global developmental delay, seizures, and visual problems	P2RX7	P2RX7(NM_002562.6):c.614C > T (p.Thr205Met)	homozygous	Yes	rs140915863	0.000127	0/4	P2RX7 is an ATP receptor that functions as a ligand-gated ion channel. When its native ligand, ATP, is not present, it acts as a scavenger receptor, recognizing and engulfing apoptotic cells (Ivetic et al., 2005). It is responsible for the ATP-dependent lysis of macrophages ⁷⁷ . It can potentially participate in both rapid synaptic transmission and ATP-mediated lysis of antigen-presenting cells. P2RX7 was found to have roles in neuroinflammation and implications in Alzheimer's disease pathogenesis ⁷⁸	No published genotype-phenotype association
10DK3900	2 Years	Male	No	Central hypoventilation and apnea, seizures, cerebellar vermis hypoplasia and early death	PRDM13	PRDM13(NM_021620.4):c.800del (p.Gly267AspfsTer34)	homozygous	Yes	NA	Absent	NA	PRDM13 is implicated in transcriptional regulation. It is thought to be a vital component of a highly coordinated transcriptional network that determines the balance of inhibitory versus excitatory neurons in the dorsal spinal cord ⁷⁹	Published ⁸⁰
10MS13800	11 Years	Female	Yes	Global developmental delay, extrapyramidal features and spasticity	PTRHD1	PTRHD1(NM_001013663.2):c.169_196del (p.Ala57ArgfsTer26)	homozygous	Yes	rs553276736	0.000541	NA	Putative peptidyl-tRNA hydrolase with a PTH2 domain, implying that it works in the ubiquitin-proteasome pathway. It was suggested to be linked to early-onset parkinsonism and cognitive dysfunction ^{81,82}	Published ⁸¹

Continued

Family-ID	Index's Age at exam test	Index's Gender	Other Affected	Phenotype	Gene-Approved	HGVS variant (hg19)	Zygosity	Consanguinity	avsnp150 rs	GnomAD frequency	Pathogenicity Scores *	Function & experimental models	Remarks
10SS5400	4 Years	Male	Yes	Hypotonia, global developmental delay, and seizures. Parents are healthy	SCN10A; CDC42BPB	SCN10A(NM_001293307.2):c.3916 T > G (p.Phe1306Val);CDC42BPB(NM_006035.4):c.3941del (p.Ala1314GlyfsTer35)	homozygous; homozygous	Yes	NA; NA	Absent; Absent	4/4; NA	SCN10A gene encodes the alpha subunit of a voltage-gated sodium channel, an integral membrane glycoprotein responsible for the initial rising phase of action in most excitable cells ⁴⁵ . CDC42BPB encodes a Serine/threonine-protein kinase that is an important downstream effector of CDC42 and regulates cytoskeleton reorganization and cell migration. Recently loss-of-function variants in this gene were associated with a neurodevelopmental phenotype ⁴⁶	SCN10A: one study published as a cause of ID with no further confirmation and no associated OMIM entry ⁴⁶ ; other associated disease is an episodic pain syndrome; CDC42BPB: one study published only describing de novo mutations (autosomal dominant) with no further confirmation ⁴⁶ . OMIM phenotype: Chillon-Okur-Chung neurodevelopmental syndrome (MIM# 619841). No biallelic variants for CDC42BPB described
14MS2200	2 Years	Female	Yes	Arthrogryposis, global developmental delay, microcephaly, and optic disc pallor	SCYL2	SCYL2(NM_017988.6):c.1624dup (p.Val542GlyfsTer16)	homozygous	Yes	NA	Absent	NA	SCYL2 regulates the production of excitatory receptors at synapses and plays an essential role in regulating neuronal function, signaling, and brain development ⁴⁷ . Mutations in this gene have been linked to neurogenic arthrogryposis multiplex congenita-4 with corpus callosum agenesis (MIM# 618766)	Published ⁴⁸
10DF10800	5 Years	Male	Yes	Global developmental delay with severely impaired intellectual function and absent speech, additionally Usher syndrome phenotype	SMG8	SMG8(NM_018149.6):c.1121_1122insG (p.Gly375TrpfsTer3);USH1G(NM_173477.5):c.1286_1295dup (p.Leu453GlnfsTer109)	homozygous	Yes	NA; NA	Absent; Absent	NA; NA	SMG8 gene encodes a component of the cell's nonsense-mediated mRNA decay (NMD) machinery. SMG8 and SMG5 as partners regulate the kinase activity of SMG1	Published ⁴⁹
10BS16700	5 Years	Male	No	Severe intellectual disability, spastic quadriplegia, abnormal brain MRI, and retinal dystrophy	SUPV3L1	SUPV3L1(NM_003171.5):c.2215C > T (p.Gln739Ter)	homozygous	Yes	NA	Absent	NA	SUPV3L1 encodes a helicase that is localized in the mitochondria. It has been shown in vitro to possess both double-stranded RNA and DNA unwinding activity that is ATP-dependent	Gene Matched (understudy). One study was published recently with no further confirmation ⁵⁰
10DK16600	12 Years	Female	No	Intellectual disability, spastic quadriplegia, abnormal brain MRI, neuroregression, seizures, and optic neuropathy	SUPV3L1	SUPV3L1(NM_003171.5):c.2215C > T (p.Gln739Ter)	homozygous	Yes	NA	Absent	NA	SUPV3L1 encodes a helicase that is localized in the mitochondria. It has been shown in vitro to possess both double-stranded RNA and DNA unwinding activity that is ATP-dependent	Gene Matched (understudy). One study was published recently with no further confirmation ⁵⁰
10BN7500	13 Years	Male	Yes	Intellectual disability	TACC2	TACC2(NM_206862.4):c.8260G > A (p.Ala2754Thr)	homozygous	Yes	rs538968585	0.00002	0/4	TACC2 is involved in the microtubule-dependent coupling of the nucleus and the centrosome and in the processes that regulate centrosome-mediated interkinetic nuclear migration (INM) of neural progenitors ⁵¹ . TACC2 is a tumor suppressor and an oncogenic protein ⁵¹	No published genotype-phenotype association
10SS8000	7 Years	Male	Yes	Moderate to severe intellectual deficiency, behavioral abnormalities, facial dysmorphism, and ophthalmological abnormalities	THUMPDP1	THUMPDP1(NM_017736.5):c.774_776del (p.Leu258del)	homozygous	Yes	rs772419789	3.98E-06	NA	THUMPDP1 is highly expressed in the brain and it is a unique adaptor protein that regulates RNA acetylation interacts through interaction with NAT10 ⁵²	Published ⁵³
10DK15900	2 Months	Female	Yes	Variable signs of cardiomyopathy, neonatal pulmonary hypertension, hypophosphatemia, chronic lung disease, developmental delay, and brain basal ganglia calcification	XPR1	XPR1(NM_004736.4):c.1811G > A (p.Arg604Gln)	homozygous	Yes	NA	Absent	4/4	The XPR1 protein mediates phosphate export from the cell and binds inositol hexakisphosphate and related inositol polyphosphates, key intracellular signaling molecules. Mutations in XPR1 gene are associated with the dominant condition of idiopathic basal ganglia calcification-6 ⁵⁴ . No ID phenotype in this condition	The patients' phenotype is strikingly different from the reported phenotype, with no past ID phenotype described. No biallelic variants for XPR1 were described
10MS4500	4 Years	Female	Yes	Variable signs of cardiomyopathy, neonatal pulmonary hypertension, hypophosphatemia, chronic lung disease, developmental delay, and brain basal ganglia calcification	XPR1	XPR1(NM_004736.4):c.1811G > A (p.Arg604Gln)	homozygous	Yes	NA	Absent	4/4		
50DK4600	2 Years	Male	Yes	Variable signs of cardiomyopathy, neonatal pulmonary hypertension, hypophosphatemia, chronic lung disease, developmental delay, and brain basal ganglia calcification	XPR1	XPR1(NM_004736.4):c.1811G > A (p.Arg604Gln)	homozygous	Yes	NA	Absent	4/4		
63MS1200	3 Years	Male	Yes	Variable signs of cardiomyopathy, neonatal pulmonary hypertension, hypophosphatemia, chronic lung disease, developmental delay, and brain basal ganglia calcification	XPR1	XPR1(NM_004736.4):c.1811G > A (p.Arg604Gln)	homozygous	Yes	NA	Absent	4/4		

Continued

Family-ID	Index's Age at exome test	Index's Gender	Other Affected	Phenotype	Gene-Approved	HGVS variant (hg19)	Zygoty	Consanguinity	avsnp150 rs	GnomAD frequency	Pathogenicity Scores *	Function & experimental models	Remarks
22SN9300	6 Years	Female	Yes	Seizures and neurodegenerative disorder (? Leukodystrophy) in addition to also cystic fibrosis	ZFYVE28	ZFYVE28(NM_020972.3):c.2015_2016 del (p.Ser672CysfsTer67)	homozygous	Yes	NA	Absent	NA	ZFYVE28 has been found to operate as a negative regulator of epidermal growth factor receptor (EGFR) signaling. It promotes EGFR degradation in endosomes when not monoubiquitinated, thereby terminating EGFR signaling. The highest expression is seen in the brain at the cortex and cerebellum. These patients also have cystic fibrosis phenotype explained by CFTR pathogenic variants	No published genotype-phenotype association

Table 1. Candidate Genes. Several were published previously through gene matching. *MetaLR, MetaSVM, MetaRNN, REVEL.

these two groups showed homozygous variants (51/65; 78.5%). These variants are rare; they explain the disease manifestations, are predicted to be damaging, are confirmed by Sanger sequencing, and are segregated with the phenotype (All listed in SUPP_Table1).

Candidate genes with a possible association with ID phenotype seen in this cohort were detected in 32/118 (27%) with a total number of affected individuals of 64 (Table 1). These candidate genes were selected according to rarity and absence in homozygosity status in local control exomes or public databases. The impact of the variants is predicated damaging. The expression patterns or mouse models supported an association with neurological dysfunction. Importantly, Sanger sequencing confirmed segregation for all variants in candidate genes in up to 3 generations in the family pedigrees (Fig. 1). The total number of candidate genes identified was 28. Table 1 shows detailed findings for the candidate genes.

Twenty-one families (18%) with 40 affected individuals remain unsolved despite multiple molecular tests, including WES, chromosomal microarray, and Fragile X where applicable. There is a trend of milder phenotypes and usually non-syndromic intellectual disabilities in unresolved cases. However, we can not draw a firm conclusion because of the small numbers.

Discussion

Global Developmental Delay (GDD)/Intellectual disability (ID) represents a group of genetic, phenotypic, and clinically heterogenic disorders that affect approximately 1% of children worldwide. This study presents the results of 188 exome analyses representing 118 consanguineous Omani families. This cohort included 215 affected individuals with intellectual disabilities, including global developmental delay, seizures, brain malformations, microcephaly, facial dysmorphism, and other systemic manifestations. Overall 82% were found to have a possible explanation. Specifically, 55% had variants in previously described and known genes (P/LP or VUS) and 27% in possible candidate genes.

With the enrichment for consanguineous families (91%), it was not surprising that the majority (85.5%) of the overall variants in the three groups were homozygous. The consanguinity rate is high due to the preference to include families with autosomal recessive phenotypes and multiple affected individuals. This study detected pathogenic (P) or likely pathogenic (LP) variants in 32 families, making the diagnostic rate of the study 27%. The main goal of this study was to recruit unsolved cases. However, the diagnostic rate observed was higher than anticipated. This can be explained by frequent reanalysis of exome data during the last five years of the study, thus enabling newly published genes to be detected. Also, VUSs, including non-coding variants, were selected with further evidence of pathogenicity becoming available. Also, during the study's initial phase, some families did not have access to clinical exome sequencing and thus were channeled to the research exome.

In a large-scale exome sequencing study, Monies and colleagues¹⁵ reported the yield of exome sequencing on 2219 families from Saudi Arabia. The overall diagnostic yield of exome sequencing based on cases with confirmed pathogenic or likely pathogenic variants was 43.3%. However, if considering variants of unknown significance (VUS) that are in an established disease-related gene or candidate genes with compelling biological candidacy were considered, the yield rate would be 73%. The high throughput design of this study led to the discovery of 236 genes that have no established OMIM phenotypes and were proposed as candidate genes. The negative results (unsolved cases) accounted for 27% of the total.

The total number of candidate genes for intellectual disability identified in this study was 28. During the course of this study, Gene Matcher²¹ was used to provide further evidence of association with the phenotype. Through this study and in collaboration with the scientific community, several of these genes have been successfully confirmed to cause intellectual disability (Table 1). One interesting shared candidate variant within the *XPR1* gene was identified in multiple affected individuals from four apparently unrelated families (Table 1). These families come from different geographical areas of Oman. However, haplotype analysis using exome data indicated that they all shared the same haplotype (Data not shown). The *XPR1* protein functions to mediate phosphate export from the cell as well as binding inositol hexakisphosphate and related inositol polyphosphates, which are key intracellular signaling molecules²². Mutations in the *XPR1* gene are known to be associated with the dominant condition of idiopathic basal ganglia calcification-6; OMIM 616413²³. The earliest age of onset for this condition is in the third and fourth decade of life, with symptoms of cerebrovascular insufficiency associated with movement disorders, cognitive decline and psychiatric symptoms. Our patients' phenotype is strikingly different; we detected biallelic *XPR1* variants and apparently healthy parents. The phenotype included variable

Pitfalls of WES	Examples
Sequence related	Unmapped genome sequence, PCR amplification artifacts or capture target Structural variants include complex or Copy Number Variants Mitochondrial genome or epigenetic variants Mosaic or uniparental disomy variants Variants in GC rich regions, pseudogenes, or repetitive and homologous sequences
Annotation and prioritization	None-canonical (Alternative) transcripts, mini exons or gene definitions OMIM-database entries delay or lag Splicing and intronic variants Synonymous that affect splicing enhancer, suppressor or induce cryptic splicing site High allelic frequency in a population but pathogenic
Clinical and phenotype	Strikingly different phenotypes or different modes of inheritance from what was described previously Two or more genetic conditions within the same family Variability of the phenotypes Incomplete clinical reevaluation after negative exomes

Table 2. Pitfalls and challenges of exome analysis.

signs of neonatal pulmonary hypertension, cardiomyopathy, serum hypophosphatemia, chronic lung disease requiring oxygen therapy, severe developmental delay, and brain basal ganglia calcification. Further functional characterization for these variants has commenced.

Whole-genome sequencing (WGS) analysis can cover up to 98% of the human genome, whereas WES only covers about 95% of the coding regions and only 1–2% of the genome. WES has a lower cost per sample than WGS, a greater depth of coverage in target regions, lesser storage requirements, and easier data analysis²⁴. It is, however, worth highlighting the pitfalls and challenges that can occur when WES analysis is performed. WES is a high-throughput, complex technique with potential pitfalls at every step. These pitfalls and the consequential missing of the molecular diagnosis in exome sequencing and analysis is a recognized phenomenon²⁵. This phenomenon could be caused by various factors, including technological limitations in variant detection, a lack of enrolling additional family members, many variants identified for probands, or the causal variant being located outside of the coding regions²⁵. Pitfalls related to WES analysis can be categorized into three main groups (Table 2).

The first group of pitfalls consists of those which are sequence-related. Large rearrangements or complex structural variants are one example. Structural variants are genomic rearrangements larger than 50 bp in size, and they account for about 1% of the variation in human genomes²⁶. Complex structural variants have been shown to contribute to human genomic variation and to cause Mendelian disease²⁷. Unfortunately, these cannot be identified easily by WES. However, multiple pipelines for CNV analysis are available. In this study and using ExomeDepth, it was possible to detect CNVs in a homozygosity state as a cause of the phenotype (Families 11MS8800 and 10DH12500). Mitochondrial mutations are other causative factors that WES cannot detect. Other factors also include mosaicism, abnormal methylation, and uniparental disomy.

Other causes of missing variants in WES include decreased coverage, locus-specific features such as GC-rich regions, and sequencing biases^{28,29}. Difficulty in the alignment of indels (insertions/deletions) larger than 20–50 nucleotides long is one of the limitations, which is likely to be the reason for WES missing such variants³⁰. Incomplete human genome annotation and sequence might also affect the accuracy of variant mapping and annotation. For instance, an intronic variant could be located in an unannotated exon³¹. Another potential cause is the high sequence similarity between pseudogenes and their corresponding functional genes³².

The second group to consider is pitfalls due to annotation and prioritization errors. Annotation and prioritization steps are used in WES analyses to reduce thousands of variants to a few candidates. During the filtering process, all annotations except for the “canonical” transcripts (i.e., the longest transcript of a gene) are initially ignored. Remarkably, pathogenic variants can be missed if alternative transcripts are not fully considered³³. Splicing is thought to be involved in 15–30% of all inherited disease variants³⁴. Despite advances in exome capture methods or machine learning for detecting variants that affect splicing, accurate detections of deep intronic variants remain limited and this could be the reason for missing splicing variants. In our cohort, intronic variants that were likely to affect splicing were detected in 19/118 families (16.1%) of which 8 were in non-canonical splicing sites.

Databases like OMIM and HGMD are used to find gene-disease and variant–disease associations in the literature. Variants or genes listed in these databases would be flagged as potentially disease-causing³⁵. Nevertheless, a reason for an initial false-negative result is that such variants or disease databases have not been kept up to date. A recent study by Bruel and colleagues illustrated this issue in a study cohort of 313 individuals³⁶. Likewise, when considering deleterious variants, those with high population prevalence might be filtered out. Penetrance of the disease might be influenced by numerous factors, including other hidden rare variants, family history, inheritance, additional medical problems, and ethnic background.

Synonymous variants, which are known as ‘silent’ variants, represent almost 50% of the variant list identified by WES. Filtering out synonymous variants reduces the variants’ list because they are assumed to be benign. However, our increased knowledge about the relationship between genetic variants and disease has shown that synonymous variants play a significant role in human disease risk and other complex traits, including variants that affect splicing³⁷. Indeed, a recent study challenged the concept that synonymous variants are neutral. In their yeast study, Shen and colleagues showed a strong non-neutrality of most synonymous mutations³⁸. If this holds true for other genes and organisms, then numerous biological conclusions, including disease causation

about synonymous mutation, would require re-examination. An example in our cohort is the family 16SS2600, where GPT2 (p.Gly245Gly) was initially missed and flagged as a silent variant.

All labs encounter pitfalls related to clinical factors and phenotypes. As a result, a negative WES result must be interpreted in the context of the patient's clinical history to determine whether reevaluation or further testing is necessary. For example, within our cohort (family 10MS6600), two related families with multiple affected individuals were enrolled as having the same phenotype and WES analysis was initially negative. However, after a detailed clinical reevaluation and exome reanalysis, the results showed that two different diseases were possibly running in the family. Some of the affected members were indeed found to harbor a deep intronic variant in the *PGAP3*, which was recently reported to cause hyperphosphatasia with mental retardation syndrome type 4 (OMIM 615,716). Another example is family 10MS16500, demonstrating that multiple individuals carrying two or more different diseases can complicate the phenotype. Similarly, the family (10DF10800) with multiple affected individuals presented with developmental delay, congenital cataracts, and bilateral sensorineural hearing loss. WES analysis identified two different variants for two different syndromes in two genes, one of which is novel as the underlying cause³⁹. Another aspect is the mode of inheritance in which conditions are known to be autosomal dominant but manifest as autosomal recessive. Monies and colleagues reported many examples of genes or diseases inherited as both AD and AR¹⁵.

Conclusion

In conclusion, using WES to identify the novel causes of human disease has changed the research landscape of genetic and neurodevelopmental disorders. Although WES is comprehensive technology, its limitations must be considered when negative results are obtained. The pitfalls of WES can potentially reduce the effectiveness of this technique in biological and medical research as well as in clinical settings. Finally, it is worth emphasizing that identifying a likely candidate gene is often just the start of a long process to confirm the variant's pathogenicity.

Data availability

The authors confirm that the data supporting the findings of this study are available within the article and Supplementary material. Further derived data are available from the corresponding author upon reasonable request. The datasets generated and/or analysed during the current study are available in the [Clinvar] repository, [<https://www.ncbi.nlm.nih.gov/clinvar/>; submission number SCV002574702 to SCV002574743].

Received: 1 August 2022; Accepted: 7 October 2022

Published online: 07 November 2022

References

- Maulik, P. K., Mascarenhas, M. N., Mathers, C. D., Dua, T. & Saxena, S. Prevalence of intellectual disability: A meta-analysis of population-based studies. *Res. Dev. Disabil.* **32**, 419–436 (2011).
- Kvarnung, M. & Nordgren, A. Intellectual disability & rare disorders: A diagnostic challenge. *Adv. Exp. Med. Biol.* **1031**, 39–54 (2017).
- Van Bokhoven, H. Genetic and epigenetic networks in intellectual disabilities. *Annu. Rev. Genet.* **45**, 81–104 (2011).
- Kochinke, K. *et al.* Systematic phenomics analysis deconvolutes genes mutated in intellectual disability into biologically coherent modules. *Am. J. Hum. Genet.* **98**, 149–164 (2016).
- Heyne, H. O. *et al.* De novo variants in neurodevelopmental disorders with epilepsy. *Nat. Genet.* **50**, 1048–1053 (2018).
- Farwell, K. D. *et al.* Enhanced utility of family-centered diagnostic exome sequencing with inheritance model-based analysis: Results from 500 unselected families with undiagnosed genetic conditions. *Genet. Med.* **17**, 578–586 (2015).
- D'Arrigo, S. *et al.* The diagnostic yield of array comparative genomic hybridization is high regardless of severity of intellectual disability/developmental delay in children. *J. Child Neurol.* **31**, 691–699 (2016).
- Visser, L. E. L. M., Gilissen, C. & Veltman, J. A. Genetic studies in intellectual disability and related disorders. *Nat. Rev. Genet.* **17**, 9–18 (2016).
- Yavarna, T. *et al.* High diagnostic yield of clinical exome sequencing in Middle Eastern patients with Mendelian disorders. *Hum. Genet.* **134**, 967–980 (2015).
- Charng, W. L. *et al.* Exome sequencing in mostly consanguineous Arab families with neurologic disease provides a high potential molecular diagnosis rate. *BMC Med. Genomics* **9**, 1–14 (2016).
- Megahed, H. *et al.* Utility of whole exome sequencing for the early diagnosis of pediatric-onset cerebellar atrophy associated with developmental delay in an inbred population. *Orphanet J. Rare Dis.* **11**, 57 (2016).
- Anazi, S. *et al.* Clinical genomics expands the morbid genome of intellectual disability and offers a high diagnostic yield. *Mol. Psychiatry* **22**, 615–624 (2017).
- Reuter, M. S. *et al.* Diagnostic yield and novel candidate genes by exome sequencing in 152 consanguineous families with neurodevelopmental disorders. *JAMA Psychiat.* **74**, 293–299 (2017).
- Santos-Cortez, R. L. P. *et al.* Novel candidate genes and variants underlying autosomal recessive neurodevelopmental disorders with intellectual disability. *Hum. Genet.* **137**, 735–752 (2018).
- Monies, D. *et al.* Lessons learned from large-scale, first-tier clinical exome sequencing in a highly consanguineous population. *Am. J. Hum. Genet.* **104**, 1182–1201 (2019).
- Özçelik, T. & Onat, O. E. Genomic landscape of the Greater Middle East. *Nat. Genet.* **48**, 978–979 (2016).
- Koshy, R., Ranawat, A. & Scaria, V. Al mena: A comprehensive resource of human genetic variants integrating genomes and exomes from Arab, Middle Eastern and North African populations. *J. Hum. Genet.* **62**, 889–894 (2017).
- Plagnol, V. *et al.* A robust model for read count data in exome sequencing experiments and implications for copy number variant calling. *Bioinformatics* **28**, 2747–2754 (2012).
- Harrison, S. M., Biesecker, L. G. & Rehm, H. L. Overview of specifications to the ACMG/AMP variant interpretation guidelines. *Curr. Protoc. Hum. Genet.* **103**, e93 (2019).
- Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American college of medical genetics and genomics and the association for molecular pathology. *Genet. Med.* **17**, 405–424 (2015).
- Sobreira, N., Schiettecatte, F., Valle, D. & Hamosh, A. GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. *Hum. Mutat.* **36**, 928–930 (2015).

22. Yao, X. P. *et al.* Analysis of gene expression and functional characterization of XPR1: A pathogenic gene for primary familial brain calcification. *Cell Tissue Res.* **370**, 267–273 (2017).
23. Legati, A. *et al.* Mutations in XPR1 cause primary familial brain calcification associated with altered phosphate export. *Nat. Genet.* **47**, 579–581 (2015).
24. Alfares, A. *et al.* Whole-genome sequencing offers additional but limited clinical utility compared with reanalysis of whole-exome sequencing. *Genet. Med.* **20**, 1328–1333 (2018).
25. Corominas, J. *et al.* Clinical exome sequencing—Mistakes and caveats. *Hum. Mutat.* <https://doi.org/10.1002/humu.24360> (2022).
26. Tattini, L., D'Aurizio, R. & Magi, A. Detection of genomic structural variants from next-generation sequencing data. *Front. Bioeng. Biotechnol.* **3**, 1–8 (2015).
27. Sanchis-Juan, A. *et al.* Complex structural variants in Mendelian disorders: identification and breakpoint resolution using short- and long-read genome sequencing. *Genome Med.* **10**, 1–10 (2018).
28. Ross, M. G. *et al.* Characterizing and measuring bias in sequence data. *Genome Biol.* **14**, R51 (2013).
29. Lelieveld, S. H., Spielmann, M., Mundlos, S., Veltman, J. A. & Gilissen, C. Comparison of exome and genome sequencing technologies for the complete capture of protein-coding regions. *Hum. Mutat.* **36**, 815–822 (2015).
30. Pena, L. D. M. *et al.* Looking beyond the exome: A phenotype-first approach to molecular diagnostic resolution in rare and undiagnosed diseases. *Genet. Med.* **20**, 464–469 (2018).
31. Krebs, C. E. *et al.* The Sac1 domain of SYNJ1 identified mutated in a family with early-onset progressive Parkinsonism with generalized seizures. *Hum. Mutat.* **34**, 1200–1207 (2013).
32. Pink, R. C. *et al.* Pseudogenes: Pseudo-functional or key regulators in health and disease. *RNA* **17**, 792–798 (2011).
33. Bodian, D. L., Kothiyal, P. & Hauser, N. S. Pitfalls of clinical exome and gene panel testing: alternative transcripts. *Genet. Med.* **21**, 1240–1245 (2019).
34. Mertes, C. *et al.* Detection of aberrant splicing events in RNA-seq data using FRASER. *Nat. Commun.* **12**, 1–13 (2021).
35. Kesselheim, A., Ashton, E. & Bockenhauer, D. Potential and pitfalls in the genetic diagnosis of kidney diseases. *Clin. Kidney J.* **10**, 581–585 (2017).
36. Bruel, A. L. *et al.* Increased diagnostic and new genes identification outcome using research reanalysis of singleton exome sequencing. *Eur. J. Hum. Genet.* **27**, 1519–1531 (2019).
37. Sauna, Z. E. & Kimchi-Sarfaty, C. Understanding the contribution of synonymous mutations to human disease. *Nat. Rev. Genet.* **12**, 683–691 (2011).
38. Shen, X., Song, S., Li, C. & Zhang, J. Synonymous mutations in representative yeast genes are mostly strongly non-neutral. *Nature* **606**, 725–731 (2022).
39. Alzahrani, F. *et al.* Recessive, deleterious variants in SMG8 expand the role of nonsense-mediated decay in developmental disorders in humans. *Am. J. Hum. Genet.* **107**, 1178–1185 (2020).
40. Baker, S. J. *et al.* Characterization of an alternatively spliced AATYK mRNA: Expression pattern of AATYK in the brain and neuronal cells. *Oncogene* **20**, 1015–1021 (2001).
41. Barik, S. An intronic microRNA silences genes that are functionally antagonistic to its host gene. *Nucleic Acids Res.* **36**, 5232–5241 (2008).
42. Lewin, D. A. *et al.* Cloning, expression, and localization of a novel γ -adaptin-like molecule. *FEBS Lett.* **435**, 263–268 (1998).
43. Zizioli, D. *et al.* $\gamma 2$ and $\gamma 1AP-1$ complexes: Different essential functions and regulatory mechanisms in clathrin-dependent protein sorting. *Eur. J. Cell Biol.* **96**, 356–368 (2017).
44. Usmani, M. A. *et al.* De novo and bi-allelic variants in AP1G1 cause neurodevelopmental disorder with developmental delay, intellectual disability, and epilepsy. *Am. J. Hum. Genet.* **108**, 1330–1341 (2021).
45. Jiang, K. *et al.* Microtubule minus-end stabilization by polymerization-driven CAMSAP deposition. *Dev. Cell* **28**, 295–309 (2014).
46. King, M. D. A. *et al.* A conserved sequence in calmodulin regulated spectrin-associated protein 1 links its interaction with spectrin and calmodulin to neurite outgrowth. *J. Neurochem.* **128**, 391–402 (2014).
47. Zou, C. *et al.* Centrobin: A novel daughter centriole-associated protein that is required for centriole duplication. *J. Cell Biol.* **171**, 437–445 (2005).
48. Sakamoto, K. *et al.* Ependymal ciliary motion and their role in congenital hydrocephalus. *Child's Nerv. Syst.* **37**, 3355–3364 (2021).
49. Shamseldin, H. E. *et al.* Identification of embryonic lethal genes in humans by autozygosity mapping and exome sequencing in consanguineous families. *Genome Biol.* **16**, 1–7 (2015).
50. Li, J. *et al.* DNAH14 variants are associated with neurodevelopmental disorders. *Hum. Mutat.* **43**, 940–949 (2022).
51. Qiu, X. B., Shao, Y. M., Miao, S. & Wang, L. The diversity of the DnaJ/Hsp40 family, the crucial partners for Hsp70 chaperones. *Cell. Mol. Life Sci.* **63**, 2560–2570 (2006).
52. Ancevska-Taneva, N., Onoprishvili, I., Andria, M. L., Hiller, J. M. & Simon, E. J. A member of the heat shock protein 40 family, h1j1, binds to the carboxyl tail of the human μ opioid receptor. *Brain Res.* **1081**, 28–33 (2006).
53. Lei, J. X., Cassone, C. G., Luebbert, C. & Liu, Q. Y. A novel neuron-enriched protein SDIM1 is down regulated in Alzheimer's brains and attenuates cell death induced by DNAJB4 over-expression in neuro-progenitor cells. *Mol. Neurodegener.* **6**, 9 (2011).
54. Schellhaus, A. K. *et al.* Developmentally regulated GTP binding protein 1 (DRG1) controls microtubule dynamics. *Sci. Rep.* **7**, 9996 (2017).
55. Sazuka, T. *et al.* Expression of DRG during murine embryonic development. *Biochem. Biophys. Res. Commun.* **189**, 371–377 (1992).
56. Kumar, S., Iwao, M., Yamagishi, T., Noda, M. & Asashima, M. Expression of GTP-binding protein gene drg during *Xenopus laevis* development. *Int. J. Dev. Biol.* **37**, 539–546 (1993).
57. Tang, T. T. T. *et al.* Dysbindin regulates hippocampal LTP by controlling NMDA receptor surface expression. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 21395–21400 (2009).
58. Cheah, S. Y., Lawford, B. R., Young, R. M., Morris, C. P. & Voisey, J. Dysbindin (DTNBP1) variants are associated with hallucinations in schizophrenia. *Eur. Psychiatry* **30**, 486–491 (2015).
59. Maes, M. S. *et al.* Schizophrenia-associated gene dysbindin-1 and tardive dyskinesia. *Drug Dev. Res.* **82**, 678–684 (2021).
60. Wang, D., Li, Y. & Shen, B. A novel erythroid differentiation related gene EDRF1 upregulating globin gene expression in HEL cells. *Chin. Med. J. (Engl)* **115**, 1701–1705 (2002).
61. Kaitsuka, T. & Matsushita, M. Regulation of translation factor EEF1D gene function by alternative splicing. *Int. J. Mol. Sci.* **16**, 3970–3979 (2015).
62. McLachlan, F., Sires, A. M. & Abbott, C. M. The role of translation elongation factor eEF1 subunits in neurodevelopmental disorders. *Hum. Mutat.* **40**, 131–141 (2019).
63. Coulter, M. E. *et al.* Regulation of human cerebral cortical development by EXOC7 and EXOC8, components of the exocyst complex, and roles in neural progenitor cell proliferation and survival. *Genet. Med.* **22**, 1040–1050 (2020).
64. Ullah, A. *et al.* A novel nonsense variant in EXOC8 underlies a neurodevelopmental disorder. *Neurogenetics* **23**, 203–212 (2022).
65. Sato, R. *et al.* Novel biallelic mutations in the PNPT1 gene encoding a mitochondrial-RNA-import protein PNase cause delayed myelination. *Clin. Genet.* **93**, 242–247 (2018).
66. Van Dijk, E. L., Schilders, G. & Pruijn, G. J. M. Human cell growth requires a functional cytoplasmic exosome, which is involved in various mRNA decay pathways. *RNA* **13**, 1027–1035 (2007).
67. dos Santos, R. F. *et al.* Major 3'–5' Exoribonucleases in the metabolism of coding and non-coding RNA. *Prog. Mol. Biol. Transl. Sci.* **159**, 101–155 (2018).

68. Matilainen, S. *et al.* Defective mitochondrial RNA processing due to PNPT1 variants causes Leigh syndrome. *Hum. Mol. Genet.* **26**, 3352–3361 (2017).
69. Zadjali, F. *et al.* Homozygosity for FARSB mutation leads to Phe-tRNA synthetase-related disease of growth restriction, brain calcification, and interstitial lung disease. *Hum. Mutat.* **39**, 1355–1359 (2018).
70. Xu, Z. *et al.* Bi-allelic mutations in Phe-tRNA synthetase associated with a multi-system pulmonary disease support non-translational function. *Am. J. Hum. Genet.* **103**, 100–114 (2018).
71. He, Y. *et al.* Circular RNA circ_0006282 contributes to the progression of gastric cancer by sponging miR-155 to upregulate the expression of FBXO22. *Oncotargets Ther.* **13**, 1001–1010 (2020).
72. Zheng, X., Yu, S., Xue, Y. & Yan, F. FBXO22, ubiquitination degradation of PHLPP1, ameliorates rotenone induced neurotoxicity by activating AKT pathway. *Toxicol. Lett.* **350**, 1–9 (2021).
73. Nagano, T. *et al.* Filamin A-interacting protein (FILIP) regulates cortical cell migration out of the ventricular zone. *Nat. Cell Biol.* **4**, 495–501 (2002).
74. Kwon, M. *et al.* Functional characterization of filamin a interacting protein 1-like, a novel candidate for antivascular cancer therapy. *Cancer Res.* **68**, 7332–7341 (2008).
75. Ivetac, I. *et al.* The type Ia inositol Polyphosphate 4-Phosphatase generates and terminates phosphoinositide 3-kinase signals on endosomes and the plasma membrane. *Mol. Biol. Cell* **16**, 2218–2233 (2005).
76. Banihashemi, S., Tahmasebi-Birgani, M., Mohammadiasl, J. & Hajjari, M. Whole exome sequencing identified a novel nonsense INPP4A mutation in a family with intellectual disability. *Eur. J. Med. Genet.* **63**(4), 103846 (2020).
77. Gu, B. J. *et al.* A rare functional haplotype of the P2RX4 and P2RX7 genes leads to loss of innate phagocytosis and confers increased risk of age-related macular degeneration. *FASEB J.* **27**, 1479–1487 (2013).
78. Chen, Y. H., Lin, R. R. & Tao, Q. Q. The role of P2X7R in neuroinflammation and implications in Alzheimer's disease. *Life Sci.* **271**, 119187 (2021).
79. Chang, J. C. *et al.* Prdm13 mediates the balance of inhibitory and excitatory neurons in somatosensory circuits. *Dev. Cell* **25**, 182–195 (2013).
80. Coolen, M. *et al.* Recessive PRDM13 mutations cause fatal perinatal brainstem dysfunction with cerebellar hypoplasia and disrupt Purkinje cell differentiation. *Am. J. Hum. Genet.* **109**, 909–927 (2022).
81. Kuipers, D. J. S. *et al.* PTRHD1 Loss-of-function mutation in an african family with juvenile-onset Parkinsonism and intellectual disability. *Mov. Disord.* **33**, 1814–1819 (2018).
82. Khodadadi, H. *et al.* PTRHD1 (C2orf79) mutations lead to autosomal-recessive intellectual disability and parkinsonism. *Mov. Disord.* **32**, 287–291 (2017).
83. Al-Kasbi, G. *et al.* Biallelic PTRHD1 frameshift variants associated with intellectual disability, spasticity, and parkinsonism. *Mov. Disord. Clin. Pract.* **8**, 1253–1257 (2021).
84. Rabert, D. K. *et al.* A tetrodotoxin-resistant voltage-gated sodium channel from human dorsal root ganglia, hPN3/SCN10A. *Pain* **78**, 107–114 (1998).
85. Chilton, I. *et al.* De novo heterozygous missense and loss-of-function variants in CDC42BPB are associated with a neurodevelopmental phenotype. *Am. J. Med. Genet. Part A* **182**, 962–973 (2020).
86. Kambouris, M. *et al.* Biallelic SCN10A mutations in neuromuscular disease and epileptic encephalopathy. *Ann. Clin. Transl. Neurol.* **4**, 26–35 (2016).
87. Gingras, S. *et al.* SCYL2 protects CA3 pyramidal neurons from excitotoxicity during functional maturation of the mouse hippocampus. *J. Neurosci.* **35**, 10510–10522 (2015).
88. Seidahmed, M. Z. *et al.* Recessive mutations in SCYL2 cause a novel syndromic form of arthrogyposis in humans. *Hum. Genet.* **139**, (2020).
89. van Esveld, S. L. *et al.* Mitochondrial RNA processing defect caused by a SUPV3L1 mutation in two siblings with a novel neurodegenerative syndrome. *J. Inher. Metab. Dis.* **45**, 292–307 (2022).
90. Rutherford, E. L. *et al.* Xenopus TACC2 is a microtubule plus end-tracking protein that can promote microtubule polymerization during embryonic development. *Mol. Biol. Cell* **27**, 3013–3020 (2016).
91. Takayama, K. I. *et al.* TACC2 is an androgen-responsive cell cycle regulator promoting androgen-mediated and castration-resistant growth of prostate cancer. *Mol. Endocrinol.* **26**, 748–761 (2012).
92. Sharma, S. *et al.* Yeast Kre33 and human NAT10 are conserved 18S rRNA cytosine acetyltransferases that modify tRNAs assisted by the adaptor Tan1/THUMP1. *Nucleic Acids Res.* **43**, 2242–2258 (2015).
93. Broly, M. *et al.* THUMP1 bi-allelic variants cause loss of tRNA acetylation and a syndromic neurodevelopmental disorder. *Am. J. Hum. Genet.* **109**, 587–600 (2022).

Author contributions

(1) Research project: A. Conception, B. Organization, C. Execution; (2) Experiments: A. Design, B. Execution, C. Review and Critique; (3) Manuscript: A. Writing of the first draft, B. Review and Critique. (4) A. patient care, B. Clinical Phenotype characterization. G.K.:1A, 1B, 1C, 2A, 2B, 3A F.M., A.K., N.H., K.T., A.S., A.F., W.M., A.A.: 1A, 3A, 4A, 4B Z.B., K.K.: 4A, 4B SA: 2B, 2C, 3B F.Z.: 1A, 1B, 3C S.Y.: 1A, 1B, 1C, 2A, 2C, 3A, 3B A.M.: 1A, 1B, 1C, 2A, 2B, 2C, 3A, 3B. 4A, 4B .

Funding

This work was supported by Sultan Qaboos University; study code: SR/MED/GENT/16/01.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-22036-z>.

Correspondence and requests for materials should be addressed to A.A.-M.

Reprints and permissions information is available at www.nature.com/reprints.

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022