

CASE REPORT OPEN



Whole exome sequencing uncovered highly penetrant recessive mutations for a spectrum of rare genetic pediatric diseases in Bangladesh

Hosnara Akter^{1,2,14}, Mohammad Shahnoor Hossain^{3,14}, Nushrat Jahan Dity¹, Md. Atikur Rahaman¹, K. M. Furkan Uddin¹, Nasna Nassir⁴, Ghausia Begum⁴, Reem Abdel Hameid⁴, Muhammad Sougatul Islam⁵, Tahrima Arman Tusty³, Mohammad Basiruzzaman^{1,6}, Shaoli Sarkar^{1,6}, Mazharul Islam^{1,6}, Sharmin Jahan⁷, Elaine T. Lim⁸, Marc Woodbury-Smith^{9,10}, Dimitri James Stavropoulos¹¹, Darren D. O'Rielly¹², Bakhrum K. Berdeiv⁴, A. H. M. Nurun Nabi², Mohammed Nazmul Ahsan³, Stephen W. Scherer^{9,13} and Mohammed Uddin^{4✉}

Collectively, rare genetic diseases affect a significant number of individuals worldwide. In this study, we have conducted whole-exome sequencing (WES) and identified underlying pathogenic or likely pathogenic variants in five children with rare genetic diseases. We present evidence for disease-causing autosomal recessive variants in a range of disease-associated genes such as *DHH*-associated 46,XY gonadal dysgenesis (GD) or 46,XY sex reversal 7, *GNPTAB*-associated mucopolipidosis II alpha/beta (ML II), *BBS1*-associated Bardet–Biedl Syndrome (BBS), *SURF1*-associated Leigh Syndrome (LS) and *AP4B1*-associated spastic paraplegia-47 (SPG47) in unrelated affected members from Bangladesh. Our analysis pipeline detected three homozygous mutations, including a novel c. 863 G > C (p.Pro288Arg) variant in *DHH*, and two compound heterozygous variants, including two novel variants: c.2972dupT (p.Met991Ilefs*) in *GNPTAB* and c.229 G > C (p.Gly77Arg) in *SURF1*. All mutations were validated by Sanger sequencing. Collectively, this study adds to the genetic heterogeneity of rare genetic diseases and is the first report elucidating the genetic profile of (consanguineous and nonconsanguineous) rare genetic diseases in the Bangladesh population.

npj Genomic Medicine (2021)6:14; <https://doi.org/10.1038/s41525-021-00173-0>

INTRODUCTION

Rare genetic disorders, while individually uncommon, collectively impact significant numbers worldwide¹, with approximately 7,000 rare diseases estimated to affect an estimated 350 million people (European Organization for Rare Diseases; EURORDIS). These disorders encompass a spectrum of rare complex clinical manifestations^{2,3} that can be difficult to diagnose and pinpoint causation. Clinics in developing countries are not equipped with technologically advanced healthcare facilities for diagnosis and treatment. In Bangladesh, genetic tests are not widely used in a clinical setting and mostly applied for medical research purposes^{4,5}. The current standard of care does not include first-tier genetic testing in most developing nations. Limited use of targeted (single-gene) sequencing, low-resolution microarray, and karyotype technologies can identify a small fraction of the total population of patients with rare genetic disorders.

Whole exome sequencing (WES) is both an effective and efficient method to detect genetic defects underlying hereditary diseases^{6–9}. Diagnoses of rare genetic diseases are essential for improving the health care system of any developing country. Early disease diagnosis can lead to early and targeted therapeutic intervention and can avoid other unnecessary diagnostic investigations. In this study, we performed WES to investigate a series of

patients from Bangladesh with rare genetic diseases referred for clinical assessment. Currently, there is no standard of care for rare disorders in Bangladesh. Clinical presentation in these patients was complex¹⁰, prompting referral for genetic diagnosis. We report seven (7) extremely rare recessive variants that are potentially causal in a range of rare genetic diseases such as, *DHH*-associated 46,XY gonadal dysgenesis (GD) or 46,XY sex reversal 7, *GNPTAB*-associated mucopolipidosis II alpha/beta ((ML II), *BBS1*-associated Bardet–Biedl Syndrome (BBS), *SURF1*-associated Leigh Syndrome (LS) and *AP4B1*-associated spastic paraplegia-47 (SPG47) from a small cohort of five unrelated Bangladeshi patients.

RESULTS

After genomic analysis of the WES data, we have identified five (5) unrelated patients with extremely rare pediatric genetic diseases carrying autosomal recessive pathogenic variants in different genes. Specifically, WES identified three homozygous variants: c.863 G > C (p.Pro288Arg), c.1339 G > A (p.Ala447Thr), and c.1216 C > T (p.Arg406Ter) in *DHH*, *BBS1*, and *AP4B1* genes, respectively (Table 1: Patient 1, Patient 3 and Patient 5). Our analysis also identified two compound heterozygous variants c.3503_3504delTC (p.Leu1168Glnfs*) and c.2972dupT

¹Genetics and Genomic Medicine Centre, NeuroGen Children's Healthcare, Dhaka, Bangladesh. ²Department of Biochemistry and Molecular Biology, University of Dhaka, Dhaka, Bangladesh. ³Department of Genetic Engineering & Biotechnology, University of Dhaka, Dhaka, Bangladesh. ⁴College of Medicine, Mohammed Bin Rashid University of Medicine and Health Science, Dubai, UAE. ⁵BioTED, Dhaka, Bangladesh. ⁶Department of Child Neurology, NeuroGen Children's Healthcare, Dhaka, Bangladesh. ⁷Department of Endocrinology & Metabolism, Bangabandhu Sheikh Mujib Medical University, Dhaka, Bangladesh. ⁸Department of Genetics, Harvard Medical School, Boston, USA. ⁹The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Canada. ¹⁰Translational and Clinical Research Institute, Newcastle University, Newcastle upon Tyne, UK. ¹¹Genome Diagnostics, Department of Pediatric Laboratory Medicine, The Hospital for Sick Children, Toronto, Canada. ¹²Faculty of Medicine, Memorial University, St. John's, NL, Canada. ¹³McLaughlin Centre and Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada. ¹⁴These authors contributed equally: Hosnara Akter, Mohammad Shahnoor Hossain. ✉email: mohammed.uddin@mbru.ac.ae

Table 1. Summary of patients clinical information and list of clinically relevant mutations in a range of rare genetic diseases in Bangladesh.

Patient Id, Age, Consanguinity Status (Yes/No)	Clinical information	Supported disease on the basis of genotype and phenotype information	Gene name	RefSeq Id	Variant type	Variant	Zygosity	Clinical significance
1 (22 Y, No)	<ul style="list-style-type: none"> primary amenorrhea under development of breast tanner staging revealed breast stage B2 (Rt) & B3(Lt) pubic hair stage 4 with female pattern normal female external genitalia with no palpable gonads absence of müllerian derivatives testis in left inguinal region karyotyping showed 46XY 	46,XY gonadal dysgenesis (GD) or 46,XY sex reversal 7	DHH	NM_021044	Exon 3	Missense	c.863 G > C; p. Pro288Arg	Homozygous (Novel) Likely Pathogenic
2 (6 M, No)	<ul style="list-style-type: none"> respiratory distress transient tachypnea of newborn congenital heart disease (PDA 3 mm, small PFO) prominent philtrum coarse facis widely spaced nipple suspicion of joint contracture both knee and elbows 	Mucopolipidosis II alpha/beta (ML II)	GNPTAB	NM_024312	Exon 19	Frameshift Deletion	c.3503_3504delTC; p. Leu1168 Glnfs*	Heterozygous Pathogenic
3 (8 Y 4 M, Yes)	<ul style="list-style-type: none"> 8 episodes of focal seizure with impaired awareness impaired intelligence fatigue language delay constipation hypothyroidism polydactyly obesity strabismus 	Bardet-Biedl Syndrome (BBS)	GNPTAB	NM_024312	Exon 15	Frameshift Insertion	c.2972dupT; p. Met9911lefs*	Heterozygous (Novel) Likely Pathogenic
4 (4 Y, No)	<ul style="list-style-type: none"> cannot sit and walk with or without support slurred speech bilateral convergent squint low muscle tone lactic acidosis (8.9 mmol/L) bilateral cerebral and cerebellum leukodystrophy 	Leukodystrophy along Leigh syndrome	BBS1	NM_024649	Exon 13	Missense	c.1339 G > A; p. Ala 447Thr	Homozygous Likely Pathogenic
5 (15 M, Yes)	<ul style="list-style-type: none"> fever for 2 days with convulsion in the form of lip-smacking, rolling of eye ball lasted for more than 8 min walk few steps with support speech delay 	Spastic paraplegia-47 (SPG-47)	SURF1	NM_003172	Exon 3	Missense	c.229 G > C; p. Gly77Arg	Heterozygous (Novel) Likely Pathogenic
			SURF1	NM_003172	Exon 8	Frameshift deletion	c.792_793delAG; p. Arg264 Serfs*	Heterozygous Pathogenic
			AP4B1	NM_006594	Exon 8	Stopgain	c.1216 C > T p. Arg406Ter	Homozygous Likely pathogenic

Note: Y = Year; M = Month

(p.Met991Ilefs*) in *GNPTAB* (Table 1: Patient 2) as well as c.229 G > C(p.Gly77Arg) and c.792_793delAG(p.Arg264Serfs*) in *SURF1* (Table 1: Patient 4). All these mutations were further verified by Sanger sequencing (Figs. 1 and 2). Below are the results from clinical observation and the genetic mutations from WES.

Patient 1

A 22-year-old girl, presenting with primary amenorrhea and underdevelopment of sexual characteristics, was referred for genetic evaluation. Her antenatal and peripartum history was uneventful. She was born by normal delivery and declared as a girl by a local physician, and raised as a female. Her cognitive, language, and motor developmental milestones were all normal,

and her school performance was satisfactory in the mainstream setting. She was the third child of a non-consanguineous marriage, with no known family history of medical illness. At the age of 15 years, she failed to establish menses and consult with a local physician. Later, she was referred to NeuroGen Children's Healthcare for clinical exome tests by the clinician at the department of Endocrinology & Metabolism, Bangabandhu Sheikh Mujib Medical University (BSMMU), Dhaka, Bangladesh at the age of 22 years. At that examination, her height was 164 cm, BMI 17.1 kg/m², and arm span 186 cm. Tanner staging revealed breast stage B2(Rt) & B3(Lt), pubic hair stage 4 with female pattern, normal female external genitalia with no palpable gonads. Hormonal evaluation showed FSH 175.12 IU/L (normal range

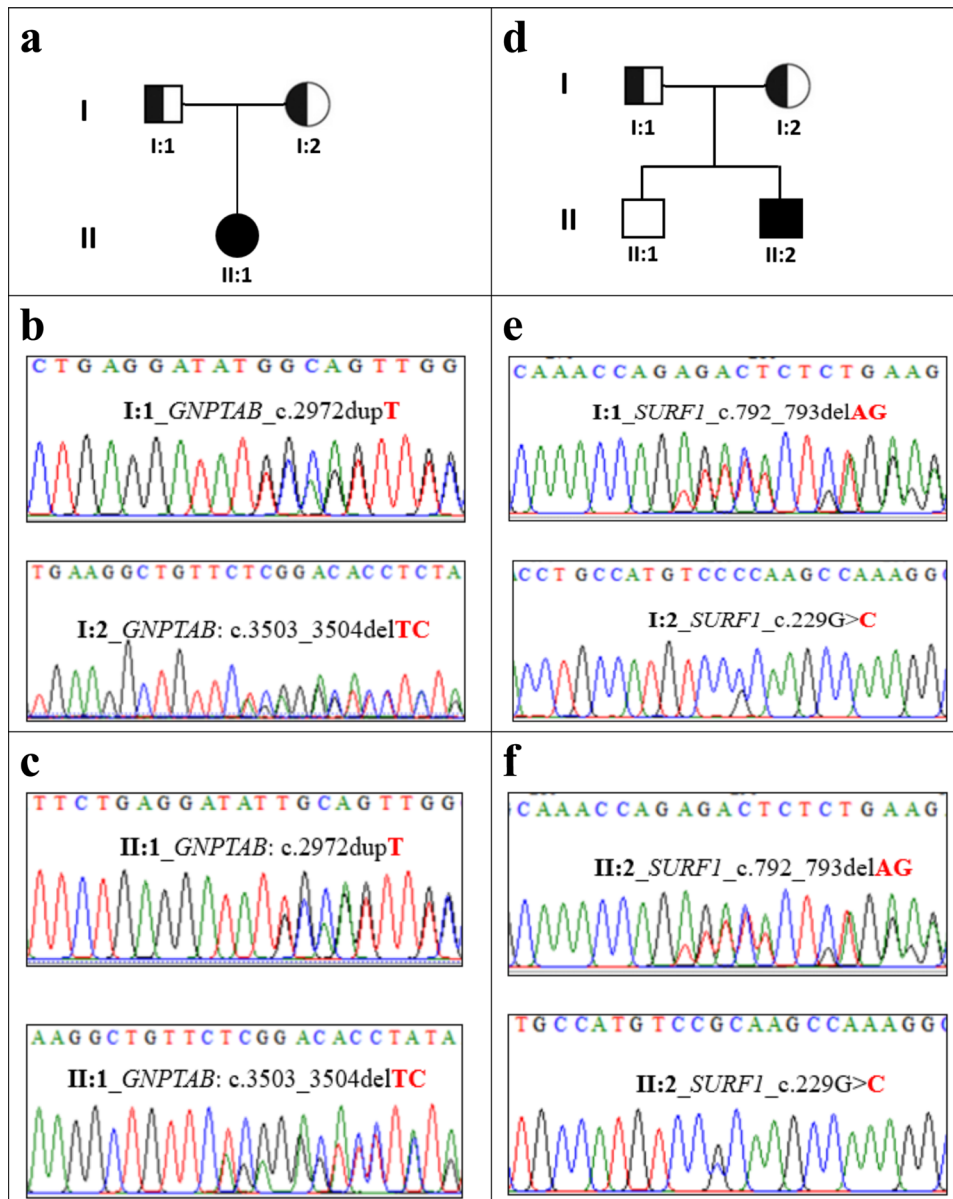


Fig. 1 Pedigree and Sanger validation of all compound heterozygous mutations in two unrelated non-consanguineous families (Patient 2 and 4, respectively). **a, d** Pedigree of two non-consanguineous families. Square indicates male and circle indicates female. The filled symbols represent affected individuals and partially filled symbols indicate carrier parents. Open symbols indicate healthy individual. **b** Chromatograms from Sanger sequencing of heterozygous *GNPTAB* mutations c.2972dupT and c.3503_3504delTC in father and mother, respectively. **c** Chromatograms from Sanger sequencing of compound heterozygous *GNPTAB* mutations c.2972dupT and c.3503_3504delTC in Patient-2. **e** Chromatograms from Sanger sequencing of heterozygous *SURF1* mutations c.792_793delAG and c.229 G > C in father and mother, respectively. **f** Chromatograms from Sanger sequencing of compound heterozygous *SURF1* mutations c.792_793delAG and c.229 G > C in patient-4. His elder brother (II:1) was apparently healthy and was not included in this study.

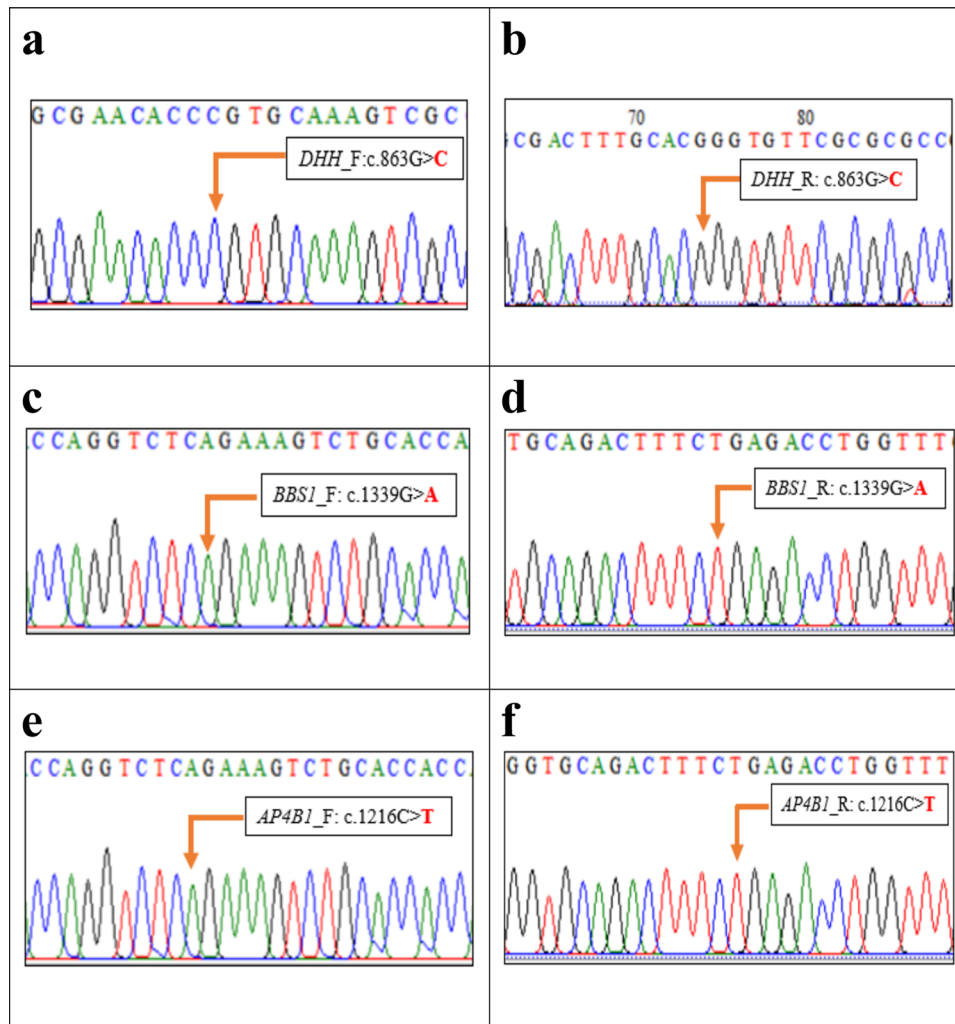


Fig. 2 Sanger validation of all homozygous mutations in *DHH*, *BBS1* and *AP4B1* genes. **a, b** Forward and reverse strand sequence of novel homozygous missense mutation c.863 G > C in *DHH* gene (Patient-1). **c, d** Forward and reverse strand sequence of homozygous missense mutation c.1339 G > A in *BBS1* gene (Patient-3). **e, f** Forward and reverse strand sequence of homozygous missense mutation c.1216 C > T in *AP4B1* gene (Patient-5).

for <33 years menstruating women: <7.0 IU/L), LH 51.13 IU/L (Women, follicular phase of menstrual cycle: 1.68 to 15 IU/L; Women, midcycle peak: 21.9–56.6 IU/L and Women, luteal phase: 0.61 to 16.3 IU/L), serum testosterone 92.8 ng/dl (Adult women normal range: 8–60 ng/dl and adult men normal range: 240–950 ng/dl), testosterone after HCG stimulation test 45 ng/dl, TSH 2.31 uIU/ml (Female at age 18–29: 0.4–2.34 mIU/L and Male at age 18–30: 0.5–4.15 mIU/L, FT4 1.58 ng/dl (Normal range: 0.7–1.9 ng/dl), estrogen 77.5 pg/ml (In premenopausal females, normal range: 30 to 400 pg/ml; In postmenopausal females, normal range: 0–30 pg/ml and In males, normal range: 10–50 pg/ml), ACTH 22.6 pg/ml (Normal range for a blood sample taken early in the morning are 9–52 pg/ml), cortisol 637 nmol/L. USG of the abdomen showed absence of Mullerian derivatives. MRI of the abdomen revealed a testis in the left inguinal region, and karyotyping showed 46XY.

After performing WES, a homozygous variant c.863 C > G (Table 1: Patient-1) in exon 3 of *DHH* gene was found that causes Proline to Arginine change at position 288 which belongs to *Dhh*-C functional domain (amino acids 199–396) of Desert Hedgehog protein. This variation was not observed in the gnomAD database (<http://gnomad.broadinstitute.org>) or the Human Gene Mutations Database (HGMD) (<http://www.hgmd.org>) or the 1000 Genomes Project and most likely reflects a novel mutation (see Fig. 2).

Multiple in silico analysis tools PolyPhen-2¹¹ and CADD_phred¹² predict that the novel variant c.863 C > G is damaging with high probability. This mutation led to a nonconservative amino acid substitution, changing a highly conserved Proline residue to Arginine. This gene is functionally shown to be relevant to the phenotype and homozygous mutations (missense/nonsense/frameshift) within this domain were classified as pathogenic for similar phenotypic patients in ClinVar database. This reported homozygous mutation can be classified as likely pathogenic in accordance with ACMG criteria¹³. For the remainder of patients, we have used similar criteria for variant interpretations.

Patient 2

Patient 2, a 6-month-old girl, was admitted to the hospital in respiratory distress, which had presented soon after birth. She was diagnosed with transient tachypnea of the newborn and was also noted to have congenital heart disease (PDA 3 mm, small PFO). She presented with craniofacial dysmorphism, including a prominent philtrum, coarse facial features, and widely spaced nipples. Her joint contractures of both knee and elbows were suspected to be the consequence of intrauterine growth restriction (IUGR). She was the firstborn of a non-consanguineous marriage and her paternal and maternal family members were healthy.

Following WES, a compound heterozygous mutation impacting *GNPTAB* was identified (c.3503_3504delTC [p.Leu1168Glnfs*], and c.2972dupT [p.Met991Ilefs*]) (see Table 1: Patient 2). The maternally inherited frameshift deletion, c.3503_3504delTC (Fig. 1), causes a change Leucine to Glutamine at position 1168 of *GNPTAB* protein and leads to a premature termination codon 5 amino acids downstream. The variant is well described in the literature^{14–17} and is the most common pathogenic variant for *GNPTAB*-related disorders in several populations, usually associated with a severe phenotype. Due to the potential impact of frameshift variants and the reported evidence, this variant is classified as pathogenic for *GNPTAB*-related disorders. The paternally inherited heterozygous frameshift insertion variant, c.2972dupT (Fig. 1), was identified in the exon 15 of *GNPTAB* that causes Methionine to Isoleucine change at position 991 of the protein and leads to a premature termination codon 25 amino acids downstream. This variation is not observed in gnomAD browser (<http://gnomad.broadinstitute.org>), the Human Gene Mutations Database (HGMD) (<http://www.hgmd.org>) or the 1000 Genomes Project, and most likely reflects a novel variant. These novel compound (frameshifts) variants can be classified as ‘likely pathogenic’ as they meet the likely pathogenic ACMG criteria¹³.

Patient 3

Patient 3, an 8 year-4-month-old boy, was referred to the pediatric neurology clinic due to focal seizures with loss of consciousness occurring 8 times within a 12 months period. His cognitive and language milestones were all delayed, and he also presented with constipation, and obesity. On examination, polydactyly and strabismus were noted, and initial workup revealed hypothyroidism. He was the second born of a consanguineous marriage and his older sister was healthy.

In this patient, a rare homozygous mutation impacting *BBS1* gene was identified using WES (c.1339 G > A [p.Ala447Thr]) that causes Alanine to Threonine change at position 447 of *BBS1* protein. This gene is a known cause of Bardet-Biedl Syndrome (BBS), an autosomal recessive and genetically heterogeneous ciliopathy characterized by retinitis pigmentosa, obesity, kidney dysfunction, polydactyly, behavioral dysfunction, and hypogonadism¹⁸. In very rare cases, seizure has also been found in patients of BBS^{19,20}. The variant c.1339 G > A falls at the last nucleotide of exon 13 of the *BBS1* coding sequence, which is part of the consensus splice site for this exon. Nucleotide substitutions within the consensus splice site are relatively common causes of aberrant splicing²¹. This variant was not found in the Genome Aggregation Database (gnomAD) as homozygous state. The heterozygous state of the variant is present in population databases at extremely rare frequencies (rs200116631, gnomAD frequency = 0.00003). Patient 3 meets the major clinical features of BBS with regard to hypothyroidism, polydactyly, obesity, strabismus, impaired intelligence in comparison to peers, and language delay. Therefore, this variant can be classified as ‘likely pathogenic’ in accordance with ACMG guideline. This variant was not screened in the parents’ DNA (Fig. 2).

Patient 4

A 4-year old boy presented to the hospital with parental concerns about loss of motor skills. Specifically, they described a one-month history of difficulty sitting, standing, and walking. According to his mother, he was able to stand, and then walk, albeit unsteadily, by 29 months. His language development followed a similar regressive pattern: he was using some simple one and two syllable words used in the appropriate context by 19 months but subsequently his speech became slurred and dysarthric at 29 months. On examination, when he first presented at 48 months, he was alert and responsive, with intact vision and hearing and a bilateral convergent squint. He had normal reflexes but low

muscle tone and bilateral weakness in all muscle groups. No abnormal or involuntary movements were noted. His OFC was 46 cm. Biochemical findings were random blood sugar of 4.7 mmol/L, serum ammonia and lactic acid level was 45 μmol/L and 8.9 mmol/L (Lactic acidosis) respectively. CT Scan of his brain identified bilateral cerebral and cerebellum leukodystrophy/ischemic change. He was the second child of a nonconsanguineous marriage and his elder brother was healthy.

A compound heterozygous variant in *SURF1* gene was identified using WES (c.229 G > C [p.Gly77Arg] and c.792_793delAG [p.Arg264-Serfs*27]). Mutations in this gene are a cause of Surf1 protein deficiency, a recessively inherited severe mitochondrial neurological disorder, and is the most frequent cause of Leigh syndrome (LS) associated with cytochrome c oxidase (COX, complex IV) deficiency^{22,23}. The maternally inherited heterozygous missense variant c.229 G > C [p.Gly77Arg] (Fig. 1) causes a Glycine to Arginine change at codon 77 of the *Surf1* protein. The heterozygous missense variant is located in the N-terminal domain (61–79 amino acid) of the *Surf1* protein that might disrupt the assembly of the *surf1* protein in the mitochondrial inner membrane. This variation was not observed in gnomAD database (<http://gnomad.broadinstitute.org>) or the Human Gene Mutations Database (HGMD) (<http://www.hgmd.org>) or the 1000 Genomes Project and likely reflects a novel mutation. In silico analysis tools SIFT_score²⁴, PolyPhen-2¹¹, and CADD_phred¹² predict that the variant is likely to be pathogenic (Table 1). This variant results in a nonconservative amino acid substitution, changing a highly conserved residue Glycine to Arginine. This novel variant is defined as likely pathogenic as it meets the criteria of ACMG likely pathogenic variant¹³. The p.Gly77Arg missense variant was in trans with another confirmed deleterious mutation, p.Arg264Serfs*. This paternally inherited frameshift deletion, c.792_793delAG (Fig. 1), causes a change Arginine to Serine at position 264 of the *Surf1* protein that leads to a premature termination codon 27 amino acids downstream (p.Lys291Ter) in the new reading frame. This variant has been reported previously in a patient with Leigh syndrome who also harbored a second *SURF1* variant²⁵ and has also been reported in a homozygous state in multiple individuals with Leigh syndrome^{22,26}. This gene functionally relevant to the phenotype and ClinVar²⁷ contains an entry for this variant as ‘pathogenic’ (Variation ID: 215238).

Patient 5

Patient 5, a 15-month old boy born to consanguineous healthy parents, was referred for genetic testing after being admitted to the hospital with an 8-min tonic-clonic seizure in the context of a 48-h fever. Prior to this, there had been no major concerns. He was delivered by an uneventful LSCS following a normal pregnancy and without any perinatal complications. His birth weight was 2.5 kg. He was able to hold up his head unsupported at 3 months, sit upright independently at 10 months, and by 15 months was able to walk a few steps with support. His speech was delayed, however, and he was only using one and two-syllable words such as “baba”, “mama”, and “dada”. On examination, at the time of first admission, he was conscious, yet irritable, and his vital signs were within the reference range. His muscle bulk was normal, as were muscle tone and reflexes. He had a plantar-extensor response. His EEG showed some non-specific background polymorphic activity. Brain MRI examination showed minimal atrophy in both frontoparietal and temporal lobes. Serum ammonia and lactic acid level were 60.0 μmol/L and 1.3 mmol/L respectively. He was the second-born, with his older female sibling, now age 6 years, previously diagnosed with psychomotor and language delay.

Using WES, a *AP4B1*-impacting homozygous variant c.1216 C > T (p.Arg406Ter) was identified that alters an amino acid from Arginine to stop codon, p.Arg406Ter. The *AP4B1* gene encodes a 739-amino acids beta 1 subunit of the heterotetrameric adaptor

protein (AP) complex²⁸. Abnormality of this gene leads to spastic paraplegia-47 (SPG47), intellectual disability, or an AP-4 deficiency syndrome^{29–31}. The variant c.1216 C > T is predicted to cause loss of normal protein function either through protein truncation or nonsense-mediated mRNA decay. It is a very rare allele; the allele frequency of this variant in gnomAD (<http://gnomad.broadinstitute.org>) is 0.00005567 (14/251462). These 14 alleles are present as heterozygotes and no homozygotes were observed in gnomAD. There is also no entry of this variant in 1000 g database, indicating it is not a common variant. ClinVar²⁷ has two submissions for this variant (Variation ID: 422147), which were listed as likely pathogenic. This variant was not screened in the parents' DNA (Fig. 2).

DISCUSSION

In this study, we performed WES in a cohort of five (5) unrelated patients with rare genetic diseases together with analysis of biological processes to explore the implementation of whole-exome sequencing for diagnosis of rare genetic diseases in Bangladeshi children. Bangladesh is a small and densely populated developing country with limited healthcare resources. As a result, there is limited availability of targeted genetic tests for diagnosing a broad range of genetic disorders. WES provides a comprehensive front-line genomic test to efficiently diagnosis a broad range of rare genetic disorders. The five patients described here had complex phenotypic manifestations. Therefore, it was difficult to do targeted genetic testing with diagnosis-specific customized panels. In all five (5) patients, the mutations described are very likely to be etiologically related to the clinical pictures described.

Patient 1 presented with primary amenorrhea and underdevelopment of secondary sexual characteristics with a 46XY karyotype (Table 1). From our genetic analysis, we identified a homozygous variant c.863 C > G (Table 1: Patient-1) in exon 3 of *DHH* that causes Proline to Arginine change at position 288 of Desert Hedgehog protein. *DHH* gene consists of 3 exons and is located on chromosome band 12q13.12³². It encodes a Desert Hedgehog signaling molecule that consists of 2 domains: a soluble amino-terminal "Hedge" domain (*DHh-N*: 23–198 amino acid) and a carboxyl-terminal "Hog" domain (*DHh-C*: 199 to 396 amino acid)³³. The Hedgehog pathway plays a critical role in developmental processes, embryogenesis, cell proliferation and tissue regeneration^{33–36}. *DHH* is important in gonadal differentiation^{37–39} as it is produced by Sertoli cells and is important for fetal Leydig cell differentiation⁴⁰. Once cleaved, *DHh-N* domain functions as a signaling molecule while the *DHh-C* participates in autoprocessing regulation³⁴. Disruption of the *DHh* auto-processing due to clinically relevant variants is associated with 46,XY gonadal dysgenesis (GD) or 46,XY sex reversal^{7,38,39,41}. We speculate that the p.Pro288Arg variant diminishes autoproteolysis activity of the *DHh-C* domain of *DHH* protein that lead to the clinical condition 46,XY gonadal dysgenesis (GD) or 46,XY sex reversal 7 in our patient (Table 1). Other clinical descriptions of mutations in this gene are also reported like polyneuropathy along with GD that is mostly associated with the complete disruption of *DHh* another domain *DHh-N*⁴². Patient 1 did not exhibit this feature as the variant is situated in the *DHh-C* domain.

Patient 2 presented with respiratory distress, congenital heart disease, craniofacial dysmorphism, and joint contractures (Table 1). A compound heterozygous mutation was identified (one novel heterozygous allele of frameshift insertion and another heterozygous allele of frameshift deletion) in exons 15 and 19 of *GNPTAB*, respectively (Table 1: Patient-2). The frameshift deletion, c.3503_3504delTC is well described in the literature^{14–17} and is the most common pathogenic variant for *GNPTAB*-related disorders in several populations, and is usually associated with a

severe phenotype. According to the Kudo et al. and Plante et al. this variant can cause a deleterious effect on *GNPTAB* gene resulting in Mucopolysaccharidosis II (MLII) phenotype in the probands. They concluded that the presence of a homozygous 3503_3504delTC variant is sufficient to produce MLII^{43,44}. Tappino et al. reported that its allele frequency is 51.1% (47/92) in their MLI and MLIII alpha/beta patients from diverse ethnic backgrounds (Italy, Argentina, Bangladesh, Bulgaria, and Hungary)¹⁶. In addition, this variant was found worldwide and resulted most probably from a unique founder molecular lesion^{14,44}. Functional studies demonstrate that the variant produces a truncated protein that is retained in the endoplasmic reticulum and not transported to the Golgi, and therefore does not form a mature subunit¹⁷. This variant has also been shown to result in absent or significantly decreased enzyme activity in patient fibroblasts^{43,45,46}.

In this patient, another heterozygous frameshift insertion variant, c.2972dupT, was identified in exon 15 of *GNPTAB* that causes Methionine to Isoleucine change at position 991 of protein and leads to a premature termination codon 25 amino acids downstream. We hypothesize that this novel truncating mutation, p.Met991Ilefs*, along with another truncating frameshift mutation, p.Leu1168Glnfs*, is sufficient to result in MLII. A single cytosine insertion at codon 1,049 has been identified in a homozygous state in a male ML II patient resulting in a frameshift and a premature termination (p.Gly1049Argfs*¹⁶)⁴⁷, and the insertion of the homologous mutation in the murine *Gnptab* results in mice displaying the clinical and biochemical features of ML II disease including the complete loss of GlcNAc-1-phosphotransferase activity⁴⁸. Mucopolysaccharidosis II-related pathogenic variants, p.Gly1049Argfs* and p.Leu1168Glnfs*, inhibit the exit from the endoplasmic reticulum and proteolytic cleavage of GlcNAc-1-phosphotransferase precursor protein (*GNPTAB*)¹⁷. Almost all patients with classical severe MLII have nonsense or frameshift mutations in *GNPTAB*^{43,44,47,49–51}. From the patient's genetic data, clinical conditions, and the above-mentioned review of literature, we confirm that patient 2 is likely to be affected by Mucopolysaccharidosis II alpha/beta (MLII) (Table 1).

Patient 3 was presented with focal seizures, delayed cognitive and language milestones, hypothyroidism, polydactyly, obesity, and strabismus (Table 1). On WES, a rare homozygous mutation (c.1339 G > A [p.Ala447Thr]) impacting *BBS1* gene was identified. At least 19 genes are associated with BBS⁵², with *BBS1* gene being the most frequently affected in BBS patients⁵³. *BBS1* maps to chromosome 11q13.1 and is composed of 17 exons⁵³, which encode a protein of 593 amino acids. The N-terminal part shares homology with other BBS-associated gene products (*BBS2* and *BBS7*). Mutations in *BBS1* gene are the most common cause of BBS, accounting for about one-quarter of all cases. The most common pathogenic sequence alterations in *BBS1* is the missense mutation p.M390R⁵³ and is found in about 20–30% of BBS cases^{54,55}. Schmid et al. found a homozygous variant c.479 G > A in the last nucleotide (splice site mutation) of exon 5 of *BBS1* that causes aberrant splicing of *BBS1* transcripts and co-segregated with the phenotype of retinitis pigmentosa⁵⁶. Although, the clinical presentation of this patient is consistent with Bardet–Biedl Syndrome (BBS), given the vast number of loci associated with BBS (*BBS1–BBS21*) combined with the fact that no customized BBS targeted gene panel exists in Bangladesh, a decision was therefore taken to undertake WES. From the WES data, we analyzed all causal genes for BBS, but did not find any clinically relevant variants in those genes except for the c.1339 G > A variant in *BBS1*. Patient 3 meets the major clinical features of BBS with regard to hypothyroidism, polydactyly, obesity, strabismus, impaired intelligence in comparison to peers, and language delay. Therefore, we define the variant as likely pathogenic (criteria described in patient 1) and assume that it is associated with BBS (Table 1). Our patient also has a history of seizures which is rare among BBS patients but has been previously reported in some BBS patients^{19,20}.

Patient 4 was presented with loss of cognitive and motor skills, lactic acidosis, and bilateral cerebral and cerebellum leukodystrophy (Table 1). On WES, a compound heterozygous variant in *SURF1* gene was identified (c.229 G > C [p.Gly77Arg] and c.792_793delAG [p.Arg264Serfs*]) (Table 1: Patient-4). The *SURF1* gene encodes a 300 amino acid long protein that contains two predicted transmembrane domains and acts as an important factor in cytochrome c oxidase (COX) assembly and maintenance. COX is the terminal component of the mitochondrial respiratory chain that mediates the transfer of electrons to oxygen^{57,58}. In multiple literatures, it was established that missense pathogenic variants in *SURF1* is associated with deficient cytochromecoxidase assembly in Leigh syndrome patients^{59,60}. *SURF1*-Associated Leigh Syndrome occurs due to null mutations accounts for 86% and missense mutations accounted for 14% of the disease alleles⁶¹. We hypothesize that this missense variant will produce an unstable *surf1* protein that will be readily degraded. In addition, p.Gly77Arg missense variant is in trans configuration with another definitive deleterious variant, p.Arg264Serfs*. Teraoka et al. identified a missense variant c.820 T > G along with the deleterious variant p.Arg264Serfs* in a Japanese patient with Leigh syndrome (256000). Most of the clinical conditions of this patient are consistent with the clinical symptoms present in our patient (patient-4)²⁵. CT Scan of the patient's brain found bilateral cerebral and cerebellum leukodystrophy or ischemic change. A similar pattern of leukodystrophy was also found in a child carrying a homozygous variant, p.Arg264Serfs*, in *SURF1*. Most of the clinical conditions of this patient are also consistent with the clinical symptoms present in our patient (Patient-4)⁶². From the patient's genotype (Patient-4), phenotype, and the above-mentioned review of literature, we hypothesize that the compound heterozygous variant p.Gly77Arg and p.Arg264Serfs* are associated with this patient's phenotype (Table 1).

Patient 5 presented with tonic-clonic seizures in the context of a 48 h fever. On WES, a homozygous nonsense variant c.1216 C > T in the exon 8 of *AP4B1* gene was identified that alters an amino acid from Arginine to the stop codon, p.Arg406Ter. *AP4B1* encodes a 739-amino acids beta 1 subunit of the heterotetrameric adaptor protein (AP) complex²⁸. AP-complex is ubiquitously expressed in neurons throughout the embryologic and postnatal developmental stages and interacts with delta2 and α -amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) glutamate receptors to selectively transport them from the trans-Golgi network to the postsynaptic somatodendritic domain⁶³. The vast majority of pathogenic variants identified in this gene are truncating variants, such as nonsense variants, frameshifts, and canonical splice sites which often disrupt gene function by leading to a complete absence of the transcript through nonsense-mediated decay⁶⁴. These mutations result in a non-functional AP-4 complex that leads to abnormal brain growth and development, which are associated with spastic paraplegia-47 (SPG47), intellectual disability, or AP-4 deficiency syndrome^{29–31}.

To date, twenty-two (22) pathogenic variants in *AP4B1* have been reported⁶⁴ which includes the c.1216 C > T variant found in our patient. This variant was reported in two unrelated *AP4B1*-associated SPG47 patients as compound heterozygotes³¹. Both were female and the disease was diagnosed at the age of 29 months and 36 months, respectively. The two patients had speech delay, infantile hypotonia, delayed motor development, developmental delay/intellectual disability, seizures, and lower extremity spasticity, but did not present with microcephaly. Our patient is clinically similar with regard to the seizures, absence of microcephaly, delayed speech and motor development. Spastic paraplegia was not seen in this patient, as this typically develops later in the course of the disease^{31,64}. As the patient's phenotype matches with other patients described in the literature, we hypothesize that the homozygous mutation, c.1216 C > T, is associated with Spastic Paraplegia-47 (Table 1).

Our findings demonstrate the potential utility of WES as a first-line diagnostic approach in patients lacking a clear differential diagnosis in a developing country with limited healthcare resources. This approach will facilitate the identification of rare genetic diseases, accelerate rare disease research and ultimately help implement precision medical care⁶⁵. We appreciate that genetic testing may not directly impact the course of the disorder, which, for neurodevelopmental disorders at least, is often chronic and without disorder-modifying treatments. However, there are a number of ways in which genetic testing can impact care pathways, which is true in Bangladesh to the same extent as other countries. Specifically, families generally seek genetic testing as it provides them with a sense of understanding regarding the origin of their child's difficulties. It also avoids the child going through unnecessary diagnostic investigations (diagnostic odyssey). It also opens up the opportunity of expert counseling for the family regarding prognosis and other possible medical complications, which also offers to possibility of screening and early intervention. We, therefore, believe that even in a country with limited health resources there are many benefits to diagnostic testing.

METHODS

Study cohort

From March 2019 to February 2020, 280 unrelated patients with neurodevelopmental disorders (NDDs), cancer, and rare genomic diseases were referred to NeuroGen Children's Healthcare from multiple tertiary hospitals within the country. Of the 280 cases, 52 patients were referred for targeted cancer panel sequencing by their physician⁴. The remaining patients were suspected of rare developmental disorders and were referred for WES. In each patient, the treating medical team noted a range of symptomatic manifestations, across different systems. The physicians collected phenotypic details, as summarized in Table 1. The children were not amenable to further formal clinical evaluation of development and cognition, and their families did not consent to further investigations due to the recent pandemic (COVID19).

Ethics statement

The study protocol was reviewed and approved by the Ethical Review Committee of the Faculty of Biological Sciences, University of Dhaka, and Institutional Review Board of Holy Family Red Crescent Medical College. Before participant enrollment, written informed consent including the use of peripheral blood and clinical data for research use and publication were obtained from the parents.

DNA extraction

Peripheral blood samples were collected from participants in an EDTA vacutainer tube. Genomic DNA was isolated using ReliaPrep™ Blood gDNA isolation kit (Promega, USA) according to manufacturer instructions. The concentration and quality of DNA was determined using NanoPhotometer C40 (Implan, Germany) (Supplementary Table 1). The quality of DNA was also assessed using 0.8% agarose gel electrophoresis.

Whole exome sequencing

Whole exome sequencing (WES) was conducted using the HiSeq sequencing platform. DNA libraries were prepared using the SureSelect XT library preparation kit and the SureSelect V7-Post Target Enrichment Kit. HiSeq uses a Burrows–Wheeler Aligner (BWA)⁶⁶ and the Genome Analysis ToolKit (GATK 4.0.11.0)⁶⁷ for converting raw sequence reads to Binary Alignment Map (BAM) and Variant Call Format (VCF) v4.1 files. The DNA sequence was assembled and aligned to reference gene sequences based on the human genome build GRCh38/UCSC hg38. ANNOVAR (2018Apr16 version) was used for functional annotation of the variants. For genomic annotations, GenomeArc (<http://genomearc.com/>), a custom clinical genetic annotation tool, was used to annotate gnomAD variant frequencies, clinvar pathogenic mutations, and pLI. Variant classification analysis was conducted based on the American College of Medical Genetics (ACMG) guidelines¹³ and Sanger sequencing was used to confirm the variants⁶⁸.

Variant validation

The recessive variants identified from WES were validated by Sanger sequencing using standard protocols. Validation primers (Supplementary Table 2) were designed using Primer 3 plus software, IDT, and UCSC Genome Browser. Primer and PCR conditions are included in the supplementary file (Supplementary Tables 2 and 3, respectively). The PCR products were run on 2.0% agarose gel and the products then purified using the Wizard® SV Gel and PCR Clean-Up System (Promega, USA) according to the manufacturer's instructions. Cycle sequencing was performed using purified PCR products as templates and BigDye® Terminator v3.1 (Applied Biosystem, USA). Bidirectional Sanger sequencing was performed using 3500 DNA Analyzer (Applied Biosystem, USA) to determine the sequence. Sanger sequencing data were analyzed using Sequence Scanner v2.0 (Applied Biosystem, USA).

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

DATA AVAILABILITY

Due to Bangladesh data privacy laws regarding patient data, both this and our patient consent does not permit the genomic or phenotypic data to be made publicly available. Genomic and phenotypic data used and/or analyzed during the current study can be shared for any collaborative research that involves NeuroGen Children's Healthcare. Please request this via the corresponding author

Received: 31 May 2020; Accepted: 6 January 2021;

Published online: 16 February 2021

REFERENCES

- Baird, P. A., Anderson, T. W., Newcombe, H. B. & Lowry, R. B. Genetic disorders in children and young adults: a population study. *Am. J. Hum. Genet.* **42**, 677 (1988).
- Uddin, K. F. et al. An ANKRD26 nonsense somatic mutation in a female with epidermodysplasia verruciformis (Tree Man Syndrome). *Clin. case Rep.* **6**, 1426 (2018).
- Guedenon, K. M. et al. Hutchinson-Gilford Progeria syndrome: report of the first Togolese case. *Am. J. Med. Genet. Part A* **182**, 1316–1320 (2020).
- Akter, H. et al. Novel mutations in actionable breast cancer genes by targeted sequencing in an ethnically homogenous cohort. *BMC Med. Genet.* **20**, 150 (2019).
- Rahman, M. M. et al. Gonadal mosaicism of large terminal de novo duplication and deletion in siblings with variable intellectual disability phenotypes. *Mol. Genet. Genom. Med.* **7**, e00954 (2019).
- Yu, F. et al. Compound heterozygous GNPTAB mutations cause mucopolipidosis II or III alpha/beta in two Chinese families. *Int. J. Clin. Exp. Pathol.* **12**, 2981 (2019).
- Neocleous, V. et al. 46, XY complete gonadal dysgenesis in a familial case with a rare mutation in the desert hedgehog (DHH) gene. *Hormones* **18**, 315–320 (2019).
- Uddin, M. et al. Brain-expressed exons under purifying selection are enriched for de novo mutations in autism spectrum disorder. *Nat. Genet.* **46**, 742 (2014).
- Tammimies, K. et al. Molecular diagnostic yield of chromosomal microarray analysis and whole-exome sequencing in children with autism spectrum disorder. *Jama* **314**, 895–903 (2015).
- Rahaman, M. A. et al. An exploration of physical and phenotypic characteristics of Bangladeshi children with autism spectrum disorder. *J. Autism Dev. Disorders* <https://doi.org/10.1007/s10803-020-04703-0> (2020).
- Adzhubei, I., Jordan, D. M. & Sunyaev, S. R. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr. Protoc. Hum. Genet.* **76**, 7.20. 21–27.20. 41 (2013).
- Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J. & Kircher, M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic acids Res.* **47**, D886–D894 (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**, 405–423 (2015).
- Coutinho, M. F. et al. Origin and spread of a common deletion causing mucopolipidosis type II: insights from patterns of haplotypic diversity. *Clin. Genet.* **80**, 273–280 (2011).
- Cathey, S. S. et al. Phenotype and genotype in mucopolipidoses II and III alpha/beta: a study of 61 probands. *J. Med. Genet.* **47**, 38–48 (2010).
- Tappino, B. et al. Molecular characterization of 22 novel UDP-N-acetylglucosamine-1-phosphate transferase α - and β -subunit (GNPTAB) gene mutations causing mucopolipidosis types IIa/ β and IIIa/ β in 46 patients. *Hum. Mutat.* **30**, E956–E973 (2009).
- De Pace, R. et al. Mucopolipidosis II-related mutations inhibit the exit from the endoplasmic reticulum and proteolytic cleavage of G I c NA c-1-phosphotransferase precursor protein (GNPTAB). *Hum. Mutat.* **35**, 368–376 (2014).
- Beales, P., Elcioglu, N., Woolf, A., Parker, D. & Flinter, F. New criteria for improved diagnosis of Bardet-Biedl syndrome: results of a population survey. *J. Med. Genet.* **36**, 437–446 (1999).
- Mugundhan, K., Thiruvartulvan, K. & Sivakumar, S. Bardet-Biedl syndrome with seizures—a rare association. *J. Assoc. Physician India* **61**, 484 (2013).
- Rathi, M. et al. Bardet-Biedl syndrome with end-stage kidney disease: a case report and review of literature. *Indian J. Nephrol.* **17**, 10 (2007).
- Buratti, E. et al. Aberrant 5' splice sites in human disease genes: mutation pattern, nucleotide structure and comparison of computational tools that predict their utilization. *Nucleic acids Res.* **35**, 4250–4263 (2007).
- Wedatilake, Y. et al. SURF1 deficiency: a multi-centre natural history study. *Orphanet J. rare Dis.* **8**, 96 (2013).
- Danis, D. et al. Mutations in SURF1 are important genetic causes of Leigh syndrome in Slovak patients. *Endocr. Regul.* **52**, 110–118 (2018).
- Ng, P. C. & Henikoff, S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic acids Res.* **31**, 3812–3814 (2003).
- Teraoka, M. et al. Two novel mutations of SURF1 in Leigh syndrome with cytochrome c oxidase deficiency. *Hum. Genet.* **105**, 560–563 (1999).
- Bindu, P. S. et al. Magnetic resonance imaging correlates of genetically characterized patients with mitochondrial disorders: a study from south India. *Mitochondrion* **25**, 6–16 (2015).
- Landrum, M. J. et al. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic acids Res.* **44**, D862–D868 (2016).
- Dell'Angelica, E. C., Mullins, C. & Bonifacino, J. S. AP-4, a novel protein complex related to clathrin adaptors. *J. Biol. Chem.* **274**, 7278–7285 (1999).
- Bauer, P. et al. Mutation in the AP4B1 gene cause hereditary spastic paraplegia type 47 (SPG47). *Neurogenetics* **13**, 73–76 (2012).
- Abdollahpour, H. et al. An AP4B1 frameshift mutation in siblings with intellectual disability and spastic tetraplegia further delineates the AP-4 deficiency syndrome. *Eur. J. Hum. Genet.* **23**, 256–259 (2015).
- Ebrahimi-Fakhari, D. et al. Clinical and genetic characterization of AP4B1-associated SPG47. *Am. J. Med. Genet. Part A* **176**, 311–318 (2018).
- Tate, G., Satoh, H., Endo, Y. & Mitsuya, T. Assignment1 of Desert Hedgehog (DHH) to human chromosome bands 12q12→ q13. 1 by in situ hybridization. *Cytogenetic Genome Res.* **88**, 93–94 (2000).
- Ingham, P. W., Nakano, Y. & Seger, C. Mechanisms and functions of Hedgehog signalling across the metazoa. *Nat. Rev. Genet.* **12**, 393–406 (2011).
- Sasai, N., Toriyama, M. & Kondo, T. Hedgehog signal and genetic disorders. *Front. Genet.* **10**, 1103 (2019).
- Lee, R. T. H., Zhao, Z. & Ingham, P. W. Hedgehog signalling. *Development* **143**, 367–372 (2016).
- Petrov, K., Wierbowski, B. M. & Salic, A. Sending and receiving hedgehog signals. *Ann. Rev. Cell Dev. Biol.* **33**, 145–168 (2017).
- Canto, P., Vilchis, F., Söderlund, D., Reyes, E. & Mendez, J. A heterozygous mutation in the desert hedgehog gene in patients with mixed gonadal dysgenesis. *Mol. Hum. Reprod.* **11**, 833–836 (2005).
- Canto, P., Söderlund, D., Reyes, E. & Mendez, J. Mutations in the desert hedgehog (DHH) gene in patients with 46, XY Comple. *pure gonadal dysgenesis. J. Clin. Endocrinol. Metab.* **89**, 4480–4483 (2004).
- Werner, R. et al. 46, XY gonadal dysgenesis due to a homozygous mutation in desert hedgehog (DHH) identified by exome sequencing. *J. Clin. Endocrinol. Metab.* **100**, E1022–E1029 (2015).
- Baldinotti, F. et al. Novel familial variant of the desert hedgehog gene: clinical findings in two Sisters with 46, XY gonadal dysgenesis or 46, XX karyotype and literature review. *Horm. Res. Paediatrics* **89**, 141–149 (2018).
- Tajouri, A. et al. In vitro functional characterization of the novel DHH mutations p.(Asn337Lysfs* 24) and p.(Glu212Lys) associated with gonadal dysgenesis. *Hum. Mutat.* **39**, 2097–2109 (2018).
- Umehara, F. et al. A novel mutation of desert hedgehog in a patient with 46, XY partial gonadal dysgenesis accompanied by minifascicular neuropathy. *Am. J. Hum. Genet.* **67**, 1302–1305 (2000).
- Kudo, M., Brem, M. S. & Canfield, W. M. Mucopolipidosis II (I-cell disease) and mucopolipidosis IIIA (classical pseudo-Hurler polydystrophy) are caused by mutations in the GlcNAc-6-phosphotransferase $\alpha\beta$ -subunits precursor gene. *Am. J. Hum. Genet.* **78**, 451–463 (2006).
- Plante, M. et al. Mucopolipidosis II: a single causal mutation in the N-acetylglucosamine-1-phosphotransferase gene (GNPTAB) in a French Canadian founder population. *Clin. Genet.* **73**, 236–244 (2008).

45. Encarnaç o, M. et al. Molecular analysis of the GNPTAB and GNPTG genes in 13 patients with mucopolipidosis type II or type III—identification of eight novel mutations. *Clin. Genet.* **76**, 76–84 (2009).
46. Velho, R. V. et al. Analyses of disease-related GNPTAB mutations define a novel GlcNAc-1-phosphotransferase interaction domain and an alternative site-1 protease cleavage site. *Hum. Mol. Genet.* **24**, 3497–3505 (2015).
47. Tiede, S. et al. Mucopolipidosis II is caused by mutations in GNPTA encoding the α/β GlcNAc-1-phosphotransferase. *Nat. Med.* **11**, 1109–1112 (2005).
48. Kollmann, K. et al. Lysosomal dysfunction causes neurodegeneration in mucopolipidosis II ‘knock-in’ mice. *Brain* **135**, 2661–2675 (2012).
49. Bargal, R. et al. When Mucopolipidosis III meets Mucopolipidosis II: GNPTA gene mutations in 24 patients. *Mol. Genet. Metab.* **88**, 359–363 (2006).
50. Paik, K. H. et al. Identification of mutations in the GNPTA (MGC4170) gene coding for GlcNAc-phosphotransferase α/β subunits in Korean patients with mucopolipidosis type II or type IIIA. *Hum. Mutat.* **26**, 308–314 (2005).
51. Tappino, B., Regis, S., Corsolini, F. & Filocamo, M. An Alu insertion in compound heterozygosity with a microduplication in GNPTAB gene underlies Mucopolipidosis II. *Mol. Genet. Metab.* **93**, 129–133 (2008).
52. Forsythe, E. & Beales, P. L. in *GeneReviews® [Internet]* (University of Washington, Seattle, 2015).
53. Myktyyn, K. et al. Identification of the gene (BBS1) most commonly involved in Bardet-Biedl syndrome, a complex human obesity syndrome. *Nat. Genet.* **31**, 435–438 (2002).
54. Myktyyn, K. et al. Evaluation of complex inheritance involving the most common Bardet-Biedl syndrome locus (BBS1). *Am. J. Hum. Genet.* **72**, 429–437 (2003).
55. Beales, P. L. et al. Genetic interaction of BBS1 mutations with alleles at other BBS loci can result in non-Mendelian Bardet-Biedl syndrome. *Am. J. Hum. Genet.* **72**, 1187–1199 (2003).
56. Schmid, F. et al. U1 snRNA-mediated gene therapeutic correction of splice defects caused by an exceptionally mild BBS mutation. *Hum. Mutat.* **32**, 815–824 (2011).
57. Echaniz-Laguna, A. et al. SURF1 deficiency causes demyelinating Charcot-Marie-Tooth disease. *Neurology* **81**, 1523–1530 (2013).
58. Zhu, Z. et al. SURF1, encoding a factor involved in the biogenesis of cytochrome c oxidase, is mutated in Leigh syndrome. *Nat. Genet.* **20**, 337–343 (1998).
59. Poyau, A. et al. Missense mutations in SURF1 associated with deficient cytochrome c oxidase assembly in Leigh syndrome patients. *Hum. Genet.* **106**, 194–205 (2000).
60. P equignot, M. O. et al. Mutations in the SURF1 gene associated with Leigh syndrome and cytochrome c oxidase deficiency. *Hum. Mutat.* **17**, 374–381 (2001).
61. Lee, I. C. et al. SURF1-associated Leigh syndrome: a case series and novel mutations. *Hum. Mutat.* **33**, 1192–1200 (2012).
62. Rahman, S., Brown, R. M., Chong, W. K., Wilson, C. J. & Brown, G. K. A SURF1 gene mutation presenting as isolated leukodystrophy. *Ann. Neurol.: Off. J. Am. Neurological Assoc. Child Neurol. Soc.* **49**, 797–800 (2001).
63. Verkerk, A. J. et al. Mutation in the AP4M1 gene provides a model for neuroaxonal injury in cerebral palsy. *Am. J. Hum. Genet.* **85**, 40–52 (2009).
64. Ruan, W.-C. et al. Novel variants in AP4B1 cause spastic tetraplegia, moderate psychomotor development delay and febrile seizures in a Chinese patient: a case report. *BMC Med. Genet.* **21**, 1–6 (2020).
65. Uddin, M., Wang, Y. & Woodbury-Smith, M. Artificial intelligence for precision medicine in neurodevelopmental disorders. *npj Digital Med.* **2**, 1–10 (2019).
66. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
67. McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
68. Sanger, F. & Coulson, A. R. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J. Mol. Biol.* **94**, 441–448 (1975).

ACKNOWLEDGEMENTS

We thank all the patients, parents, and participants of this study. Special thanks to GenomeArc Inc. for helping with additional clinical genomics annotation. This study is funded by College of Medicine, Mohammed Bin Rashid University of Medicine and Health Sciences, Al Jalila Foundation, and Genetics and Genomic Medicine Centre, NeuroGen Children’s Healthcare Research grant.

AUTHOR CONTRIBUTIONS

H.A. and M.S.H. contributed equally to this work. H.A., M.S.H., and M.U. conceived the idea of the project, and designed experiments. N.J.D., A.R., K.M.F.U., G.B, R.A.H., T.A.T., M.B., S.S., M.I., S.J., M.N.A. helped collecting clinical phenotypes, recruiting samples, and validation experiment. H.A., M.U., helped with variant interpretation and analysis. H.A., M.S.H., E.T.L., B.B., M.S.I., N.N., A.H.M.N.N., S.S., D.R., D.J.S., M.W., and M.U. contributed into writing the manuscript and critically reviewed the manuscript.

COMPETING INTERESTS

M.U. is on the Scientific Advisory Committees of GenomeArc Inc., and NeuroGen Children’s Healthcare. S.W.S. is on the Scientific Advisory Committees of Deep Genomics, Population Bio and is a Highly Cited Academic Advisor to the King Abdulaziz University. The remaining authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41525-021-00173-0>.

Correspondence and requests for materials should be addressed to M.U.

Reprints and permission information is available at <http://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 license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

  The Author(s) 2021