Expanding the genetic heterogeneity of intellectual disability

Shams Anazi*[1], Sateesh Maddirevula*[1], Yasmine T Asi[2], Saud Alsahli[1], Amal Alhashem[3], Hanan E. Shamseldin[1], Fatema AlZahrani[1], Nisha Patel[1], Niema Ibrahim[1], Firdous M. Abdulwahab[1], Mais Hashem[1], Nadia Alhashmi[4], Fathiya Al Murshedi[4], Ahmad Alshaer[12], Ahmed Rumayyan[5,6], Saeed Al Tala[7], Wesam Kurdi[9], Abdulaziz Alsaman[17], Ali Alasmari[17], Mohammed M Saleh[17], Hisham Alkuraya[10], Mustafa A Salih[11], Hesham Aldhalaan[12], Tawfeg Ben-Omran[13], Fatima Al Musafri[13], Rehab Ali[13], Jehan Suleiman[14], Brahim Tabarki[3], Ayman W El-Hattab[15], Caleb Bupp[18], Majid Alfadhel[19], Nada Al-Tassan[1,16], Dorota Monies[1,16], Stefan Arold[20], Mohamed Abouelhoda[1,16], Tammaryn Lashley[2], Eissa Faqeih[17], Fowzan S Alkuraya[1,3,16,21,18]

*These authors have contributed equally

[1]Department of Genetics, King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia.

[2]Queen Square Brain Bank for Neurological Disorders, Department of Molecular Neuroscience, UCL Institute of Neurology, University College London, London, UK.

[3]Department of Pediatrics, Prince Sultan Military Medical City, Riyadh, Saudi Arabia.

[4]Department of Genetics, College of Medicine, Sultan Qaboos University, Sultanate of Oman.

[5]King Saud bin Abdulaziz University for Health Sciences, Riyadh, Saudi Arabia.

[6]Neurology Division, Department of Pediatrics, King Abdulaziz Medical City, Riyadh, Saudi Arabia.

[7]Armed Forces Hospital Khamis Mushayt, Department of Pediatrics and Genetic Unit, Riyadh, Saudi Arabia.

[9]Department of Obstetrics and Gynecology, King Faisal Specialist Hospital, Riyadh, Saudi Arabia

[10]Department of Ophthalmology, Specialized Medical Center Hospital, Riyadh, Saudi Arabia.

[11]Division of Pediatric Neurology, Department of Pediatrics, King Khalid University Hospital and College of Medicine, King Saud University, Riyadh, Saudi Arabia.

[12]Pediatric Neurology, King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia.

[13]Clinical and Metabolic Genetics, Department of Pediatrics, Hamad Medical Corporation, Qatar.

[14]Department of Pediatrics, Division of Neurology, Tawam Hospital, Al Ain, United Arab Emirates.

[15]Division of Clinical Genetics and Metabolic Disorders, Department of Pediatrics, Tawam Hospital, Al-Ain, United Arab Emirates.

[16]Saudi Human Genome Program, King Abdulaziz City for Science and Technology, Riyadh, Saudi Arabia.

[17]Department of Pediatric Subspecialties, Children's Hospital, King Fahad Medical City, Riyadh, Saudi Arabia.

[18]Spectrum Health Genetics, Grand Rapids, MI, USA

[19]King Abdullah International Medical Research Centre, King Saud bin Abdulaziz University for Health Sciences, Genetics Division, Department of Pediatrics, King Abdulaziz Medical City, Riyadh, Saudi Arabia

[20]King Abdullah University of Science and Technology (KAUST), Computational Bioscience Research Center (CBRC), Division of Biological and Environmental Sciences and Engineering (BESE), Thuwal, 23955-6900, Saudi Arabia

[21]Department of Anatomy and Cell Biology, College of Medicine, Alfaisal University, Riyadh, Saudi Arabia

Authors declare no conflict of interest

Corresponding author: falkuraya@kfshrc.edu.sa

ABSTRACT

Intellectual disability (ID) is a common morbid condition with a wide range of etiologies. The list of monogenic forms of ID has increased rapidly in recent years thanks to the implementation of genomic sequencing techniques. In this study, we describe the phenotypic and genetic findings of 67 families (104 patients) all with novel ID-related variants. In addition to established ID genes, including ones for which we describe unusual mutational mechanism, some of these variants represent the first confirmatory disease-gene links following previous reports (*TRAK1*, *GTF3C3*, *SPTBN4* and *NKX6-2*), some of which were based on single families. Furthermore, we describe novel variants in 14 genes that we propose as novel candidates (*ANKHD1*, *ASTN2*, *ATP13A1*, *FMO4*, *MADD*, *MFSD11*, *NCKAP1*, *NFASC*, *PCDHGA10*, *PPP1R21*, *SLC12A2*, *SLK*, *STK32C* and *ZFAT*). We highlight *MADD* as a particularly compelling candidate in which we identified biallelic likely deleterious variants in two ID families. We also highlight *NCKAP1* as another compelling candidate in a large family with autosomal dominant mild intellectual disability that fully segregates with a heterozygous truncating variant. The candidacy of *NCKAP1* is further supported by its biological function, and our demonstration of relevant expression in human brain. Our study expands the locus and allelic heterogeneity of ID and demonstrates the power of positional mapping to reveal unusual mutational mechanisms.

INTRODUCTION

Intellectual disability (ID) is a common morbidity that affects 1-3% of the population [1; 2]. The resulting functional impairment varies depending on the severity of ID, which ranges from mild to severe as quantified by the intelligence quotient (IQ) score. The etiology of ID is highly heterogeneous although genetic forms are increasingly recognized as a major etiological category. *De novo* genetic and genomic variants account for at least 50% of severe ID in outbred populations [3]. In contrast, the majority of ID in highly inbred populations is caused by recessive mutations [4; 5].

The recent expansion in the use of genomic sequencing techniques (primarily whole-exome sequencing) has resulted in a rapid expansion of the genetic determinants of ID. In the case of dominant causes, very large sequencing projects have concluded that the nearly all major ID genes have likely been identified [6]. This appears to be in striking contrast to recessive ID where recent large sequencing projects have revealed very little overlap in their lists of novel candidate genes, which suggests that a significant proportion of recessive ID genes have yet to be identified [4; 7-9].

Consanguinity loops are not only helpful in facilitating the discovery of novel recessive disease genes, including ID genes, but they also facilitate the occurrence of homozygous truncating (knockout) alleles in genes previously reported as candidates, which strongly corroborates those tentative disease-gene links [10]. Additionally, consanguinity can unmask recessive mutations in genes that are only known to cause diseases in an autosomal dominant fashion [11]. This can lead to a remarkably different phenotype from the previously reported dominant one [5; 12-14]. Finally, certain classes of pathogenic mutations may evade detection/interpretation by the typical genomic sequencing and only come to focus with the help of positional mapping aided by

consanguinity [15]. In this study, we attempted to exploit the above advantages of consanguinity to expand the morbid genome of ID.

MATERIALS AND METHODS

Human Subjects

All patients with documented ID or significant component of cognitive impairment in the setting of developmental delay in a young child were eligible. Only patients in whom we identified variants that potentially explain their phenotype were included. Those with potentially causal variants that are not novel were excluded. Available family members were recruited for segregation analysis as appropriate. A written informed consent was obtained from all subjects prior to participation in accordance with a KFSHRC IRB-approved research protocol (RAC# 2121053).

Mutation Identification

All subjects with negative family history had molecular karyotyping performed as described before[4; 16]. Only those in whom normal results were obtained were included in subsequent analysis. Specifically, a previously described multigene panel (599 genes) were used to screen for mutations in known ID genes [17]. If negative, we proceeded with autozygome analysis and whole exome sequencing (WES) as described before [17]. Some patients were directly tested by clinical WES. Variant classification was according to the recent ACMG guidelines [18]. For PM2 score, we used gnomAD as well as an in-house ethnically matched database of 2,363 exomes and 1,607 neuro gene panels. We only included cases in whom pathogenic or likely pathogenic variants were identified. The only exceptions are those in whom variants in candidate genes

were identified. These were included for the purpose of proposing novel candidate genes. Candidacy of these novel genes was based on several factors as described before [4; 16]. Sanger confirmation was performed for all reported variants and segregation with the phenotype was confirmed whenever samples from relatives were available.

RTPCR and Molecular Cloning

RNA isolation and gene-specific RTPCR were as described [19]. PCR amplified product was subjected to molecular cloning into sequencing vector pGEM-T Easy (Promega) followed by sequencing with SP6/T7 polymerase primers.

Tissue Collection

Samples were collected from brains donated to the Queen Square Brain Bank for Neurological Disorders, Institute of Neurology, University College London, UK. The donations were made according to ethically approved protocols, and tissue at Queen Square Brain Bank is stored under a license issued by the Human Tissue Authority (No. 12198).

Computational structural analysis of mutants

Sequences were retrieved from the Uniprot database. SwissModel and RaptorX [20; 21] were used to produce homology models. RaptorX was used for prediction of secondary structure and protein disorder. QUARK [22] was used for ab initio structural modeling. Models were manually inspected, and mutations evaluated, using the Pymol program (pymol.org).

Immunohistochemistry

Immunohistochemistry Formalin-fixed paraffin-embedded tissue blocks from the frontal cortex, hippocampus and cerebellum of normal adult brains were used. Immunohistochemistry for

NCKAP1 (1:750; Novus Biologicals, Littleton, CO) was performed using a standard avidin/biotin technique with chromogen diaminobenzidine. Briefly, eight-micrometer-thick sections were placed in methanol/H_2O_2 solution to neutralize endogenous peroxidase activity. After pressure cooking in citrate buffer (0.1M) at pH 6.0, sections were placed in 10% non-fat milk to reduce non-specific binding, followed by incubation in the primary antibody overnight at 4^0C. After washes in PBS, the sections were incubated with biotinylated secondary antibody (Vector Laboratories, Burlingame, CA) antibody, then washed off and incubated in the avidin-biotin complex solution. The immunoreaction was visualized by treating the sections with diaminobenzidine/H_2O_2 solution and then counterstaining with Mayer's hematoxylin.

RESULTS

We report 104 patients representing 67 families with novel alleles in known, tentative or novel candidate disease genes. A list of these variants and the corresponding clinical phenotypes is provided in Table S1; detailed clinical description is provided in Table S2.

Expanding the Allelic Heterogeneity of Known ID Genes

Among the 67 families we describe in this study, 47 had potentially causal variants in genes with established link to ID. With one exception, none of these alleles was a common founder mutation, consistent with the notion that the overwhelming majority of these had been identified and reported by us in previous work [5]. The sole exception was a novel variant in *SPG20* in four apparently unrelated families who all shared the core features of ID and growth hormone deficiency of Troyer syndrome (Figure 1A). At least one of these families had had "negative" investigation by exome sequencing. In hindsight, this variant was most likely missed at the stage

of interpretation because it is consistently predicted to be non-deleterious at the protein level by different *in silico* tools. However, by running "SpliceAid" tool, we found that this variant predicts loss of the important exonic splicing suppressor. Indeed, RTPCR followed by cloning confirmed that the normal transcript is largely replaced by an aberrant transcript with 25 bps deletion (Figure 1C-D). Reassuringly, autozygome analysis confirms that three of these families (the fourth was not available for genotyping) map to the *SPG20* locus (Figure 1B). Another unusual mutational mechanism was the finding of a homozygous apparently loss of function mutation in *CDH15*, a gene that had only been reported in the context of autosomal dominant ID [23]. The phenotype we observed in that family is similarly non-syndromic ID (Table S2).

Confirming the Candidacy of Previously Reported Candidate ID Genes

A single missense variant in *GTF3C3* was very recently reported as a novel candidate cause of ID [8]. In 15DG0315, we observed a homozygous splicing (+3) variant (NM_012086.4:c.1382+3A>G), which we confirmed by RTPCR to result in skipping of exon10 and part of exon11 (Figure S1B,C). However, we note that our patient has profound secondary microcephaly with a characteristic facial appearance (Figure S1A). It is possible that the more severe nature of our variant compared to that by Reuter et al led to the more severe phenotype. Future cases will be required to further delineate this novel syndrome. Similarly, a single nonsense variant in *SPTBN4* was reported very recently to cause ID associated with congenital myopathy, deafness, and neuropathy [24]. In 16DG1625, the phenotype we observed in the context of a different homozygous truncating variant was a predominantly central nervous system phenotype in the form of global developmental delay and diffuse T2 hyper-intense signal abnormality predominantly in the subcortical white matter.

In addition to the above genes that had been reported on the basis of single variants, we also identified variants in genes reported only once but with multiple variants. These include the very recently reported *TRAK1* and *NKX6-2* [25; 26].

Identification of Novel Candidate ID Genes

We propose 14 genes (in 15 families) not previously linked to any Mendelian disease in humans as potential ID genes. Justification for their selection is given in Table 1 along with detailed phenotypic features. Two such genes deserve a special mention. Biallelic variants were identified in *MADD* in two families. One Saudi family was homozygous for NM_003682.3:c.2930T>G:p.(Val977Gly) and were able to identify a second family from USA that is compound heterozygous for NM_003682.3:c.593G>A:p.(Arg198His) and NM_003682.3:c.979C>T:p.(Arg327*)). *MADD* is a regulator of neurotransmitter release and mouse model exhibits severe neuronal defects with early lethality [27; 28]. 3D modeling of these missense variants in *MADD* supports the pathogenicity of identified variants by *in silico* analysis (Figure 2). Arg198 and Arg327 are located in the DENN domain. The structure of the MADD DENN domain can be inferred based on similarity to the crystal structure of the human DENND1B DENN domain, bound to the Rab GTPase (PDB 3tw8) [29] (Figure 2A). For structure modeling, we deleted the linker regions between the DENN domains, based on DENN domain sequence alignments and predictions of the secondary structure elements and protein disorder. This truncated MADD DENN domain has a 25 % sequence identity to the DENND1B DENN domain (which also has loop deletions to allow crystallization). Arg198 is located in the 3D homology model on the outside of Alpha-helix H2 of the so-called longin module of the N-terminal half (Figure 2A) [29]. This longin domain forms part of the GTPase interaction site in DENND1B. The GTPase binding site is well conserved in MADD, however the H2 region is

distant from the GTPase site (Figure 2A). Given that Arg198 is surface exposed, a histidine in position 198 would not lead to steric clashes, but the shortening of the side chain and the loss of a positive charge might affect intra- or intermolecular interactions. The precise effect of the Arg198His mutation cannot be predicted, because of a lack of structural and functional knowledge. The Arg327* truncation clearly demolishes the second half of the DENN domain, including the GTPase binding site (Figure 2A). The resulting protein is expected to be highly instable and to have lost most, if not all of its functions (including GTPase binding). Leu977Arg (corresponding to Leu1040Arg in the canonical isoform1) is located in the C-terminal region of MADD. Structure and function are unknown, but this region is predicted to be stably folded into a ~110 residue helical domain. Ab initio structure predictions of this region suggest that Leu977 is located centrally on a hydrophobic surface of an Alpha-helix (Figure 2B). This positioning is compatible with either Leu977 forming part of the hydrophobic core of the protein domain, or, alternatively, forming part of a hydrophobic interaction surface. In both cases, the mutation Leu977Arg is expected to be highly disruptive, either severely destabilizing the 3D fold or the interaction.

NCKAP1 was identified based on a large multigenerational family segregating non-syndromic mild ID (Figure 4). Genomewide filtering of the exomic variants in the index revealed a heterozygous truncating variant in *NCKAP1* as the most likely candidate. Previous studies have suggested expression of *NCKAP1* in human brain based on Northern blot analysis [30]. To confirm, we conducted immunohistochemical analysis on normal adult human brain sections using a commercially available antibody (Novus Biologicals, Littleton, CO) following a standard avidin/biotin technique with chromogen diaminobenzidine. Our analysis revealed that NCKAP1 is evident in cells ofvarious brain regions including Purkinje cells and dentate nucleus of the

cerebellum, CA4 region and dentate gyrus of the hippocampus, and in frontal grey and white (Figure 4A-F). We also note that ExAC (Exome Aggregation Consortium) lists a probability of loss of function intolerance (pLI) score of 1.00 indicating that this gene is extremely constrained against haploinsufficiency in the human genome.

DISCUSSION

ID is one of the most common indications for genomic testing [5]. The extreme clinical and genetic heterogeneity of this condition was key to the professional recommendation of performing molecular karyotyping as a first-tier test in these patients [31]. Although no similar recommendation has been made with respect to other genomic tests (panels, WES and whole-genome sequencing), our experience and that of others strongly support their implementation as first-tier tests in those with positive family history and in parallel with molecular karyotyping in those with negative family history [4].

The aim of this study is not to show the yield of genomic tests in the setting of ID. Rather, we set out to specifically share with the clinical and molecular genetics community a large number of carefully annotated variants that potentially cause ID. In the era of expanding use of public databases for the interpretation of variants, it is critical that individual efforts to annotate medically relevant variants are made available on timely basis [32]. In addition, our confirmation of the candidacy of genes with previously reported tentative link to disease is an important contribution to the publicly-funded efforts of improving the rigor with which disease-gene links are established for the purpose of clinical testing [33].

For the candidate genes that we report for the first time in this study, we acknowledge that the disease-gene link is tentative since these are based on single families, with the exception of *MADD*. However, sharing of these candidates can be very helpful in facilitating post-publication matchmaking and future confirmation as we have demonstrated for several genes in this study [34]. Reassuringly, the strict selection criteria we apply by our Mendelian Genomics Program in this study has proven helpful in the past in enriching for genes that are likely to be confirmed by subsequent reports. For example, of the 33 novel candidate genes we proposed in 2015, we note that 12 have now been reported to harbor deleterious variants that caused similar phenotypes in additional patients [16].

Although we emphasized the candidacy of *NCKAP1* and *MADD*, we note that the other 12 novel candidate genes have also been selected on the basis of relevant biological data as shown in Table 1. For example, *SLC12A2* encodes a Na-K-Cl cotransporter that is highly expressed in developing cortical neurons, is necessary for NGF-induced neurite outgrowth, regulates hippocampal neuronal development and has been implicated recently in schizophrenia risk [35-38]. *ANKHD1* regulated the development of neuroprogenitors [39]. *MFSD11* encodes a putative solute carrier that is abundant in developing brain and has been implicated in regulating energy homeostasis [40]. *NFASC* encodes neurofascin, which has been found to be necessary for the formation of paranodal axo-glial junctions, which are in turn necessary for the propagation of nerve impulses [41]. *SLK* (STE20-LIKE PROTEIN KINASE) gene encodes a novel kinase that governs early cell signaling pathways like cell division, and the mouse model appears to recapitulate the human phenotype [42].

In summary, we present a number of novel variants that we hope will contribute to the global endeavor of improving the medical annotation of the human genome for the benefit of families with various forms of intellectual disability.

ACKNOWLEDGEMENT

We thank the study families for their enthusiastic participation. This work was supported in part by King Salman Center for Disability Research (FSA). We acknowledge the support of the Saudi Human Genome Program and the Sequencing and Genotyping Core Facilities at KFSRHC. The research by STA and MAA reported in this publication was supported by funding from King Abdullah University of Science and Technology (KAUST).

REFERENCES

1. Bhasin, T.K., Brocksen, S., Avchen, R.N., and Braun, K.V.N. (2006). Prevalence of four developmental disabilities among children aged 8 years: Metropolitan Atlanta Developmental Disabilities Surveillance Program, 1996 and 2000.(US Department of Health and Human Services, Centers for Disease Control and Prevention).
2. Leonard, H., and Wen, X. (2002). The epidemiology of mental retardation: challenges and opportunities in the new millennium. Mental retardation and developmental disabilities research reviews 8, 117-134.
3. Gilissen, C., Hehir-Kwa, J.Y., Thung, D.T., van de Vorst, M., van Bon, B.W., Willemsen, M.H., Kwint, M., Janssen, I.M., Hoischen, A., and Schenck, A. (2014). Genome sequencing identifies major causes of severe intellectual disability. Nature.
4. Anazi, S., Maddirevula, S., Faqeih, E., Alsedairy, H., Alzahrani, F., Shamseldin, H., Patel, N., Hashem, M., Ibrahim, N., and Abdulwahab, F. (2016). Clinical genomics expands the morbid genome of intellectual disability and offers a high diagnostic yield. Molecular Psychiatry.
5. Monies, D., Abouelhoda, M., AlSayed, M., Alhassnan, Z., Alotaibi, M., Kayyali, H., Al-Owain, M., Shah, A., Rahbeeni, Z., and Al-Muhaizea, M.A. (2017). The landscape of genetic diseases in Saudi Arabia based on the first 1000 diagnostic panels and exomes. Human Genetics, 1-19.
6. Study, T.D.D.D. (2015). Large-scale discovery of novel genetic causes of developmental disorders. Nature 519, 223-228.
7. Harripaul, R., Vasli, N., Mikhailov, A., Rafiq, M.A., Mittal, K., Windpassinger, C., Sheikh, T., Noor, A., Mahmood, H., and Downey, S. (2017). Mapping autosomal recessive intellectual disability: combined microarray and exome sequencing identifies 26 novel candidate genes in 192 consanguineous families. Molecular Psychiatry.
8. Reuter, M.S., Tawamie, H., Buchert, R., Gebril, O.H., Froukh, T., Thiel, C., Uebe, S., Ekici, A.B., Krumbiegel, M., and Zweier, C. (2017). Diagnostic yield and novel candidate genes by exome sequencing in 152 consanguineous families with neurodevelopmental disorders. JAMA psychiatry 74, 293-299.
9. Riazuddin, S., Hussain, M., Razzaq, A., Iqbal, Z., Shahzad, M., Polla, D., Song, Y., van Beusekom, E., Khan, A., and Tomas-Roca, L. (2016). Exome sequencing of Pakistani consanguineous families identifies 30 novel candidate genes for recessive intellectual disability. Molecular psychiatry.
10. Amos, J.S., Huang, L., Thevenon, J., Kariminedjad, A., Beaulieu, C.L., Masurel-Paulet, A., Najmabadi, H., Fattahi, Z., Beheshtian, M., and Tonekaboni, S.H. (2017). Autosomal recessive mutations in THOC6 cause intellectual disability: syndrome delineation requiring forward and reverse phenotyping. Clinical genetics 91, 92-99.
11. Monies, D., Maddirevula, S., Kurdi, W., Alanazy, M.H., Alkhalidi, H., Al-Owain, M., Sulaiman, R.A., Faqeih, E., Goljan, E., and Ibrahim, N. (2017). Autozygosity reveals recessive mutations and novel mechanisms in dominant genes: implications in variant interpretation. Genetics in Medicine.
12. Tabarki, B., AlMajhad, N., AlHashem, A., Shaheen, R., and Alkuraya, F.S. (2016). Homozygous KCNMA1 mutation as a cause of cerebellar atrophy, developmental delay and seizures. Human genetics 135, 1295-1298.
13. Aldahmesh, M.A., Khan, A.O., Mohamed, J., and Alkuraya, F.S. (2011). Novel recessive BFSP2 and PITX3 mutations: insights into mutational mechanisms from consanguineous populations. Genetics in Medicine 13, 978-981.
14. Patel, N., Faqeih, E., Anazi, S., Alfawareh, M., Wakil, S.M., Colak, D., and Alkuraya, F.S. (2015). A novel APC mutation defines a second locus for Cenani–Lenz syndrome. Journal of medical genetics, jmedgenet-2014-102850.

15. Shamseldin, H.E., Maddirevula, S., Faqeih, E., Ibrahim, N., Hashem, M., Shaheen, R., and Alkuraya, F.S. (2016). Increasing the sensitivity of clinical exome sequencing through improved filtration strategy. Genetics in Medicine.
16. Alazami, A.M., Patel, N., Shamseldin, H.E., Anazi, S., Al-Dosari, M.S., Alzahrani, F., Hijazi, H., Alshammari, M., Aldahmesh, M.A., and Salih, M.A. (2015). Accelerating novel candidate gene discovery in neurogenetic disorders via whole-exome sequencing of prescreened multiplex consanguineous families. Cell reports 10, 148-161.
17. Group, S.M. (2015). Comprehensive gene panels provide advantages over clinical exome sequencing for Mendelian diseases. Genome biology 16, 1-14.
18. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., and Spector, E. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genetics in Medicine 17, 405-423.
19. Aldahmesh, M.A., Mohammed, J.Y., Al-Hazzaa, S., and Alkuraya, F.S. (2012). Homozygous null mutation in ODZ3 causes microphthalmia in humans. Genetics in Medicine 14, 900-904.
20. Arnold, K., Bordoli, L., Kopp, J., and Schwede, T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics 22, 195-201.
21. Källberg, M., Margaryan, G., Wang, S., Ma, J., and Xu, J. (2014). RaptorX server: a resource for template-based protein structure modeling. Protein Structure Prediction, 17-27.
22. Xu, D., and Zhang, Y. (2012). Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field. Proteins: Structure, Function, and Bioinformatics 80, 1715-1735.
23. Bhalla, K., Luo, Y., Buchan, T., Beachem, M.A., Guzauskas, G.F., Ladd, S., Bratcher, S.J., Schroer, R.J., Balsamo, J., and DuPont, B.R. (2008). Alterations in CDH15 and KIRREL3 in patients with mild to severe intellectual disability. The American Journal of Human Genetics 83, 703-713.
24. Knierim, E., Gill, E., Seifert, F., Morales-Gonzalez, S., Unudurthi, S.D., Hund, T.J., Stenzel, W., and Schuelke, M. (2017). A recessive mutation in beta-IV-spectrin (SPTBN4) associates with congenital myopathy, neuropathy, and central deafness. Human Genetics, 1-8.
25. Chelban, V., Patel, N., Vandrovcova, J., Zanetti, M.N., Lynch, D.S., Ryten, M., Botía, J.A., Bello, O., Tribollet, E., and Efthymiou, S. (2017). Mutations in NKX6-2 Cause Progressive Spastic Ataxia and Hypomyelination. The American Journal of Human Genetics 100, 969-977.
26. Barel, O., Malicdan, C.V., Ben-Zeev, B., Kandel, J., Pri-Chen, H., Stephen, J., Castro, I.G., Metz, J., Atawa, O., and Moshkovitz, S. (2017). Deleterious variants in TRAK1 disrupt mitochondrial movement and cause fatal encephalopathy. Brain 140, 568-581.
27. Del Villar, K., and Miller, C.A. (2004). Down-regulation of DENN/MADD, a TNF receptor binding protein, correlates with neuronal cell death in Alzheimer's disease brain and hippocampal neurons. Proceedings of the National Academy of Sciences 101, 4210-4215.
28. Tanaka, M., Miyoshi, J., Ishizaki, H., Togawa, A., Ohnishi, K., Endo, K., Matsubara, K., Mizoguchi, A., Nagano, T., and Sato, M. (2001). Role of Rab3 GDP/GTP exchange protein in synaptic vesicle trafficking at the mouse neuromuscular junction. Molecular biology of the cell 12, 1421-1430.
29. Wu, X., Bradley, M.J., Cai, Y., Kümmel, D., Enrique, M., Barr, F.A., and Reinisch, K.M. (2011). Insights regarding guanine nucleotide exchange from the structure of a DENN-domain protein complexed with its Rab GTPase substrate. Proceedings of the National Academy of Sciences 108, 18672-18677.
30. Suzuki, T., Nishiyama, K., Yamamoto, A., Inazawa, J., Iwaki, T., Yamada, T., Kanazawa, I., and Sakaki, Y. (2000). Molecular cloning of a novel apoptosis-related gene, human Nap1 (NCKAP1), and its possible relation to Alzheimer disease. Genomics 63, 246-254.

31. Miller, D.T., Adam, M.P., Aradhya, S., Biesecker, L.G., Brothman, A.R., Carter, N.P., Church, D.M., Crolla, J.A., Eichler, E.E., and Epstein, C.J. (2010). Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. The American Journal of Human Genetics 86, 749-764.
32. Abouelhoda, M., Faquih, T., El-Kalioby, M., and Alkuraya, F.S. (2016). Revisiting the morbid genome of Mendelian disorders. Genome Biology 17, 235.
33. Strande, N.T., Riggs, E.R., Buchanan, A.H., Ceyhan-Birsoy, O., DiStefano, M., Dwight, S.S., Goldstein, J., Ghosh, R., Seifert, B.A., and Sneddon, T.P. (2017). Evaluating the clinical validity of gene-disease associations: an evidence-based framework developed by the Clinical Genome Resource. The American Journal of Human Genetics.
34. Boycott, K.M., Rath, A., Chong, J.X., Hartley, T., Alkuraya, F.S., Baynam, G., Brookes, A.J., Brudno, M., Carracedo, A., and den Dunnen, J.T. (2017). International cooperation to enable the diagnosis of all rare genetic diseases. The American Journal of Human Genetics 100, 695-705.
35. Dzhala, V.I., Talos, D.M., Sdrulla, D.A., Brumback, A.C., Mathews, G.C., Benke, T.A., Delpire, E., Jensen, F.E., and Staley, K.J. (2005). NKCC1 transporter facilitates seizures in the developing brain. Nature medicine 11, 1205-1213.
36. Nakajima, K.-i., Miyazaki, H., Niisato, N., and Marunaka, Y. (2007). Essential role of NKCC1 in NGF-induced neurite outgrowth. Biochemical and biophysical research communications 359, 604-610.
37. Callicott, J.H., Feighery, E.L., Mattay, V.S., White, M.G., Chen, Q., Baranger, D.A., Berman, K.F., Lu, B., Song, H., and Ming, G.-l. (2013). DISC1 and SLC12A2 interaction affects human hippocampal function and connectivity. The Journal of clinical investigation 123, 2961-2964.
38. Merner, N.D., Mercado, A., Khanna, A.R., Hodgkinson, A., Bruat, V., Awadalla, P., Gamba, G., Rouleau, G.A., and Kahle, K.T. (2016). Gain-of-function missense variant in SLC12A2, encoding the bumetanide-sensitive NKCC1 cotransporter, identified in human schizophrenia. Journal of psychiatric research 77, 22-26.
39. Hermann, R. (2014). Regulation of Neural Progenitor Proliferation by ANKHD1.
40. Perland, E., Lekholm, E., Eriksson, M.M., Bagchi, S., Arapi, V., and Fredriksson, R. (2016). The putative SLC transporters Mfsd5 and Mfsd11 are abundantly expressed in the mouse brain and have a potential role in energy homeostasis. PloS one 11, e0156912.
41. Thaxton, C., Pillai, A.M., Pribisko, A.L., Labasque, M., Dupree, J.L., Faivre-Sarrailh, C., and Bhat, M.A. (2010). In vivo deletion of immunoglobulin domains 5 and 6 in neurofascin (Nfasc) reveals domain-specific requirements in myelinated axons. Journal of Neuroscience 30, 4868-4876.
42. Machicoane, M., de Frutos, C.A., Fink, J., Rocancourt, M., Lombardi, Y., Garel, S., Piel, M., and Echard, A. (2014). SLK-dependent activation of ERMs controls LGN–NuMA localization and spindle orientation. J Cell Biol 205, 791-799.

FIGURE LEGENDS

Figure 1. (A) Pedigrees and facial images of 4 families with *SPG20* mutation. Note the lack of major facial dysmorphism. (B) Agileidiogramapper shows shared haplotype on Chr13. (C) RTPCR showing the aberrant transcript in *SPG20*. (D) Diagram representing location of deletion In between exon3 and 4. Chromatogram showing deletion of 25 bps in *SPG20* (NM_001142294.1:c. 988A>G: r.986-1010 del). Pt: patient; N: normal control.

Figure 2 (A) Secondary structure representation of the homology model of the tripartite MADD DENN domain (residues 1 to 605, with truncated loop regions), based on the structure of the DENN domain from DENND1B (PDB 3tw8). The MADD DENN domain is colored gray up to residue 326, and orange thereafter, to highlight the region absent in the Arg327* mutant. R198 is shown in green. The GTPase (yellow) was taken from the superimposed structure of DENND1B bound to Rab (3tw8). (B) Illustration of an ab initio 3D structure prediction of the MADD domain containing Leu977 (L1040 in isoform1). The Leu977 side chain is highlighted in green. Hydrophobic residues present on the same helix are colored in orange.

Figure 3. (A) Pedigree showing multiple affected members. -/- denotes wild type and +/- denotes heterozygous mutation in *NCKAP1*. (B) Facial features of affected girl showing mild hypertelorism. (C) Chromatogram of *NCKAP1* heterozygous missense mutation. (D) Genomic organization of *NCKAP1* and mutation in in exon 32.

Figure 4. NCKAP1 expression in normal adult human brain. NCKAP1 protein expression is evident in various regions of the normal human adult brain as shown by NCKAP1 immunohistochemistry. The protein is present in both neuronal and glial cells as shown in the Perkinjie cells (A, arrows) and dentate nucleus (B) of the cerebellum, CA4 region (C) and

dentate gyrus (D) of the hippocampus, and frontal grey (E) and white matter (F). *Scale bar: A,B: 50μm; C-F: 50μm.*

Figure S1. A) Clinical features of patient with *GTF3C3*-related ID showing facial asymmetry, bilateral temporal narrowing, epicanthal folds, upslanting palpebral fissures, bulbous nose, and full cheeks. (C) Chromatogram showing skipping of exone10 and partial part of exon11.

Table:1. Novel candidate genes identified in this study. Homo: homozygous change; Het: heterozygous change.

Family	**Clinical synopsis (HPO Terms)**	**Novel candidate gene and variant**	**Zygosity**	**Supporting evidences**
HC03426016	Failure to thrive; Speech delay and language development; Microcephaly; Short stature; Anteriorly placed anus; Abnormal facial shape; Intrauterine growth retardation	*ANKHD1* (NM_017747.2:c.7365dup:p.(His2456Serfs*13))	De novo	Involved in cell survival, cell-cycle regulation, ion channel, cell survival, cell signaling, and protein–protein interactions and apoptosis (PMID:16098192), segregated within family, pLI score of 1.00.
15DG0307	Abnormal facial shape; hypospadias; chordee; Global developmental delay; Depressed nasal bridge; Frontal bossing; Abnormality of the frontal hairline; Microtia; anteverted nares; café au lait spot	*ASTN2* (NM_014010.4:c.892G>C:p.(Asp298His))	Homo	Associated with neurodegenerative disease and hippocampal volume (PMID:25410587 and 28098162), segregated within family.
13DG1545	Intellectual disability; Attention deficit hyperactivity disorder; Recurrent respiratory infections; Downslanted palpebral fissures; Prominent nose; Hyperplasia of the maxilla; Abnormality of the fingernails	*ATP13A1* (NM_020410.2:c.1045G>A:p.(Glu349Lys))	Homo	Protects from iron-induced neuro cytotoxicity (PMID: 25912790), segregated within family.
13DG1202	Abnormal CNS myelination; Peripheral axonal neuropathy; Abnormality of the foot; Scoliosis; Drooling; Ulnar claw; Microcephaly;	*FMO4* (NM_002022.1:c.83C>A:p.(Pro28His))	Homo	Identified by positional mapping and segregated within family.

	Areflexia; Intellectual disability			
72960215, 1506605420	Speech delay and language development; Autism; Poor eye contact	*MADD* (NM_003682.3:c.593G>A:p.(Arg198His) and NM_003682.3:c.979C>T:p.(Arg327*))	Compound Het	Critical regulator of neurotransmitter release in synapse and of neuronal viability (PMID: 11359932 and 15007167)
2616102	Global developmental delay; Failure to thrive; Poor eye contact	*MADD* (NM_003682.3:c.2930T>G:p.(Val977Gly))	Homo	Critical regulator of neurotransmitter release in synapse and neuronal viability (PMID: 11359932 and 15007167)
15DG2492	Intellectual disability; Short stature; Abnormal facial shape; Abnormal heart morphology; Abnormality of the genitourinary system	*MFSD11* (NM_001242532.1:c.143G>C:p.(Gly48Ala))	Homo	Expressed in mouse brain particularly in excitatory and inhibitory neurons (PMID: 27272503) and segregated within family.
12DG1370	Tip-toe gait; Intellectual disability; ; Speech delay and language development; Attention deficit hyperactivity disorder; Hypertelorism; Macrocephaly; Tall stature	*NCKAP1* (NM_205842.2:c.3298G>T:p.(Glu1100*))	Hetero	Expressed in the hippocampus and cerebral cortex in mouse brain and associated with Alzheimer's disease (AD) pathology (PMID: 11418237). Our expression studies of NCKAP1 on human brain clearly show abundant expression in human brain and segregated with in family members. pLI score 1.00.
14DG0056	Global developmental delay;Congenital laryngeal stridor; Hypertonia; Neonatal respiratory distress; Abnormal facial shape; Small anterior fontanelle; Recurrent respiratory infections; Hypertelorism; Hypotonia; Wide nasal bridge; Micrognathia; Glossoptosis; Hyperextensibility of the finger joints; 11 pairs of ribs	*NFASC* (NM_001160331.1:c.1109G>C:p.(Arg370Pro))	Homo	Critical for Ranvier node maintenance and Myelination of axon Function (PMID: 28217083) and segregated within family.

14DG1188	Intellectual disability; Abnormal facial shape; Bicuspid aortic valve; Cleft upper lip; Inguinal hernia; Recurrent otitis media; Microcephaly; Brachycephaly; Anteverted ears; Bulbous nose; Long philtrum; Thin upper lip vermilion; Downturned corners of mouth; Synophrys; Pes planus; Arachnoid cyst	*PCDHGA10* (NM_018913.2:c.823G>A:p.(Glu275Lys))	Homo	Cadherin family genes regulates neuronal network in the brain (PubMed: 10380929) and segregated within family.
N010	Severe global developmental delay; Neonatal respiratory distress; Dilatation of lateral ventricles; Hypotonia; Generalized muscle weakness; Recurrent respiratory infections; Abnormal facial shape; Wide nasal bridge; Upslanted palpebral fissures; Coarse facial features; Generalized hirsutism; Low-set, posteriorly rotated ears; Thick lower lip vermilion; High, narrow palate; Hepatomegaly; Myopia; Rotatory Nystagmus; Areflexia; Abnormal CNS myelination; Cavum septum pellucidum; Enlarged cisterna magna	*PPP1R21* (NM_001193475.1:c.2056C>T:p.(Gln686*))	Homo	Segregated within family.
SFH-871000	Global developmental delay;	*SLK (*NM_014720.1:c.1414G>T:p.(Glu472*))	Homo	Insertional mutagenesis of *SLK* gene in mouse

	Failure to thrive; Recurrent respiratory infections; Hydrocephalus; Cutis laxa; Joint hyperextensibility; Hepatomegaly; Hypotonia; Generalized muscle weakness; Strabismus; Oculomotor apraxia; Ventriculomegaly			shows developmental defects along with neuronal and skeletal defects (Mouse Genome Informatics (MGI))
KFMC-435029086	Global developmental delay; Failure to thrive; Central hypotonia; Microcephaly;Neonatal respiratory distress; Recurrent aspiration pneumonia; Osteopenia	*SLC12A2* (NM_001046.2:c.2617-2A>G)	Homo	Involved in hippocampal neuronal development (PMID: 23921125) and segregated within family.
10DG0720	Global developmental delay; Seizures; Microcephaly; hypertonia; Hypocalcemic seizures; Hypomagnesemia; strabismus; Pes planus; Hypocalcemia; Brachycephaly	*STK32C (*NM_173575.2:c.451G>C:p.(Val151Leu))	Homo	Associated with depression PMID: 24929637 and segregated within family.
15DG2661	Global developmental delay; Hypotonia; Abnormal facial shape; Hypotonia; Hip dysplasia; Failure to thrive; Intrauterine growth retardation; Enlarged cisterna magna; Polycythemia; Coarse facial features; Hypertrichosis; Thick eyebrow; Upslanted palpebral fissure; High	*ZFAT* (NM_020863.3 :c.1199G>A:p.(Arg400Gln))	Homo	Peripheral T cell homeostasis and immune development (PMID: 22828507) and segregated within family.

	forehead; Strabismus; Prominent nose; Thin vermilion border; Prominent nasal tip; High, narrow palate; Arachnodactyly; Pectus excavatum; Generalized amyotrophy; Overlapping toe; osteopenia; 11 pairs of ribs			