J.Natn.Sci.Foundation Sri Lanka 2023 51 (4): 643 - 650
DOI: http://dx.doi.org/10.4038/jnsfsr.v51i4.11256

RESEARCH ARTICLE

Moleculalar Pathology

Whole exome sequencing coupled with *in silico* functional analysis identified *NID1* as a novel candidate gene causing neuro-psychiatric disorder in a Pakistani family

M Muzammal[1], S Ahmad[1], MZ Ali[1], S Fatima[1], S Abbas[1], J Khan[2], H Ullah[1], W Shah[3], L Duan[4*] and MA Khan[1*]

[1] *Gomal Centre of Biochemistry and Biotechnology, Gomal University Dera Ismail Khan, KP, Pakistan.*
[2] *Institute of Biological Sciences, Gomal University Dera Ismail Khan, KP, Pakistan.*
[3] *University of Science and Technology of China, Hefei, Anhui, China.*
[4] *Department of Reproduction Andrology, Sichuan Jinxin Xinan Woman & Childern Hospital, Chengdu, Sichuan, China.*

Submitted: 23 June 2022; Revised: 6 June 2023; Accepted: 27 June 2023

Abstract: Intellectual disability (ID) is a neuro-developmental condition that affects a person's cognitive ability and results in a learning defect. It affects 1–3% of the general population; however, the ratio may be expected to be more in a consanguineous population. Herein in the present study, we identified a nuclear family from Dera Ismail Khan City in Pakistan. Whole exome sequencing was performed to map the pathogenic variant. Protein structural modeling and interaction studies were carried out to validate the variant with disease-association. Molecular modeling of normal and mutated proteins was performed through I-TASSER and Chimera tools, while docking and interaction analysis was carried out using Cluspro. Clinical analysis of the patient determined mild intellectual disability and gait problem. Candidate gene analysis in this family found a homozygous missense mutation NM_002508:c.C2512T (p.Arg838Cys) in the 12th exon of *NID1* gene. Molecular modeling of wild-type and mutant NID1 proteins determined a significant effect on the protein's secondary and tertiary structure. Hence, based on the exome sequence analysis, *NID1* is proposed to be a strong novel candidate ID gene in this family. The genetic mapping of the present family led us to determine a novel candidate gene to be associated with intellectual disability. Linkage of additional ID families with genes would confirm its validity and strengthen our notion. Furthermore, expression studies and pathway analysis will help in exploring the biological mechanism of learning and memory.

Keywords: Exome sequencing, ID, NID1, novel candidate gene, Pakistani family.

INTRODUCTION

Intellectual disability (ID) is a neurological defect that occurred due to anomalies of the brain. Previously, ID was also known as mental retardation, learning defect or developmental delay (Ropers, 2010). Intellectual defect results in loss of social adaptive skills and usually occurs before the age of 18 (WHO, 1992). The worldwide prevalence of ID ranges from 1 to 3% (APA, 2013). Precisely, its prevalence is 1 to 2% in developed countries and 2 to 3% in less developed countries like Pakistan (Durkin, 2002; Maulik *et al*., 2011). However, its incidence is speculated to be high in those countries where the rate of cousin marriages is high (Salvador-Carulla *et al*., 2008). Based on the intelligence quotient (IQ) value, ID is characterized into mild (IQ range 55> - >69), moderate (IQ range 35> - >54), severe (IQ range 20> - >34) and profound (IQ range below 19) (Ahmed *et al*., 2021). Phenotypically, ID is categorized as syndromic and non-syndromic intellectual disability (NS-ID). In syndromic ID, the individual suffers from additional biological anomalies, while in NS-ID, the individual is merely suffering from

* Corresponding authors (675068375@qq.com; m.ahmad@gu.edu.pk; https://orcid.org/0009-0000-1105-2154; https://orcid.org/0000-0001-8591-0377)

learning defect, as in the case of autosomal recessive non-syndromic intellectual disability (ARNSID) (Bittles & Black, 2010). Much research has documented that approximately 10% of marriages worldwide are between blood relatives (*https://www.ncbi.nlm.nih.gov/omim/*), but its frequency is much higher in Central Asian regions especially Pakistan, India, and Iran. In Pakistan, the exact number of intellectually disabled children is still unknown due to unavailability of any proper survey. According to the census of 1998, Pakistan has 2.5% of people suffering from different kinds of disabilities. Until now, 75 genes/loci are enlisted in the Online Mendelian Inheritance in Man (OMIM) database (*https://www.ncbi.nlm.nih.gov/omim/*) involved in NS-ARID.

Almost 560 genes have been associated with intellectual disability (*https://www.omim.org*). All the modes of Mendelian heredity (*i.e.*, autosomal recessive, autosomal dominant, or X-linked) are associated with intellectual disabilities. Physiologically, these reported genes are involved in various cellular signaling cascades, inter-neuronal connectivity, neuronal proliferation, neuronal migration, and the extensive guideline of genetic/epigenetic transcription and translation (Muzammal *et al.*, 2022). Improvement in identifying genes accountable for the worldwide developmental delays and intellectual impairment has expanded our knowledge of the molecular mechanism in learning and memory, which is essential for understanding cognition and intelligence from a neurological perspective. Growing awareness of molecular pathology will assist in finding new pharmacological methodologies (Muzammal *et al*., 2022).

The current genetic study reported a Pakistani family to have the causative genetic factors of ARID. The current ID family was recruited from Dera Ismail Khan City of Pakistan. Exome sequence analysis in this family found a novel missense mutation NM_002508:c.C2512T (p.Arg838Cys) in the *NID1* gene.

MATERIALS AND METHODS

The current study was approved by the institutional Ethical Review Board of Gomal University, D.I. Khan, Pakistan, and supported by the Higher Education Commission (HEC) of Pakistan through an NPRU grant (project #5564/KPK/NRPU/R&D/HEC/2016). Venous blood samples were obtained from patients and normal individuals (including parents) of the affected family as per the standard protocols, and kept at 4^0C in EDTA tubes. and DNA was extracted using DNeasy Blood Kit (Qiagen).

Genetic analysis

The positional cloning approach was adopted for gene mapping, which involved whole exome sequencing (WES), homozygosity mapping (based on SNP data of WES), and candidate gene and/or variant prioritization. The detailed description of gene mapping and mutation analysis is outlined as follows.

Whole exome sequencing

Whole exome sequencing was done on NextSeq500 platform, using Nextera rapid capture exome kit for library construction, following the manufacturer's protocol (Illumina, San Diego, USA).

For candidate gene analyses, several online tools were employed, *e.g.*, Exomiser (*https://monarch-exomiser-web dev.monarchinitiative.org*), PhenIX (*http://compbio.charite.de/PhenIX/*), ENDEVOUR (*https://endeavour.esat.kuleuven.be*), and ToppGene (*https://www.toppgene.cchmc.org*).

Protein Secondary structure prediction

The secondary structure of the protein was predicted through PSI-blast based secondary structure Prediction (PSIPRED) tool (McGuffin *et al*., 2000). This method is used to examine the folding pattern of protein secondary structure.

Protein 3D Modeling and docking

I-TASSER online (Yang & Zhang, 2015) tool was used to design the 3D models of normal and mutant NID1 protein and its close interactor. Model with top confidence score (C-Score) was selected for further investigation (Muzammal *et al*., 2022). Chimera 1.13.1 (Goddard *et al*., 2007) was used to visualized the designed 3D models (Ali *et al*., 2022). Protein-protein docking was carried out with its close functional interactors through an online tool called Cluspro server (Kozakov *et al*., 2017).

RESULTS AND DISCUSSION

The current genetic study was conducted on a Saraiki origin consanguineous Pakistani family suffering from NS-ID, and *NID1* was identified as novel candidate gene involved in NS-ID.

The family was recruited from Dera Ismail Khan City. The nuclear family consisted of an affected father

(I:1), normal mother, and six children with one affected daughter (II: 3). The pedigree analysis apparently determined autosomal recessive mode of disease inheritance (Figure 1a). The patients were suffering from mild intellectual impairment with low IQ. The affected father also had some psychiatric problems and had attention deficit. The female patient (II: 3) had Gait issue, attention deficits, and kidney issues. The general phenotypes of patients were consistent with AR-NSID. Whole exome sequencing was selectively performed on patient II:3. The rest of the clinical details are described in Table 1.

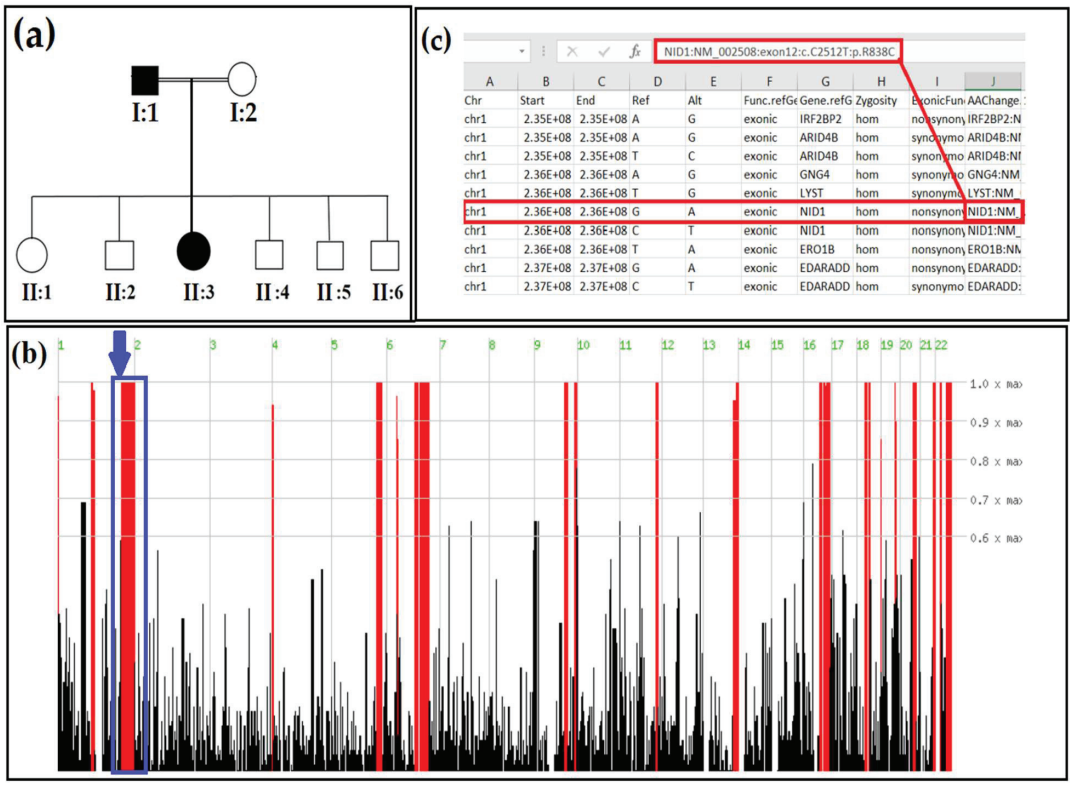

Figure 1: a) Pedigree of the family showing autosomal recessive mode of inheritance, (b) screenshot homozygosity mapper showing the homozygous region, largest region on chromosome is highlighted in the blue box. (c) Presence of candidate gene NID1 variant in the exome file.

Genetic analysis

Candidate gene analysis

Whole exome analysis found a total of 23,600 variants. Through homozygosity mapper, HBD analysis found a large significant homozygous region on chromosome 1 (chr1: 206579936-240371106), spanning a 33.7 Mb region on DNA. The subsequent exome analysis, with primary target on the large homozygous region coupled with screening of the top 10 candidate genes identified through different candidate gene analysis tools, identified homozygous single nucleotide substitution in NM_002508: c.C2512T:(p.Arg838Cys) in the 12th exon of the *NID1* gene (Figure 1 b & 1c). The identified variant (rs763658641) has a minor allele frequency of 0.000003 in ExAC and 0.000015 in the gnomAD exomes databases.

However, the clinical significance of this variant is not reported in the Clinvar database. ACMG-AMP classification categorized this variant as VUS (variant of unknown significance).

Table 1: Clinical features of patients (I:1 and II:3)

Phenotypes	II:3	I:1
Gender	Female	Male
Level of intellectual disability	Mild	Mild
Kidney stone	Yes	No
Muscular degeneration	Yes	No
Gait problem	Yes	No
Epileptic fits	No	No
Hearing problem	No	No
Dermal lesion	No	No
Psychiatric phenotypes	Late responsive, *i.e.*, attention deficit	Late responsive, *i.e.*, attention deficit
Jerking of limbs	Yes	No
Self-feeding	Yes	Yes
Microcephaly features	No	No
Macrocephaly features	No	No
Digit anomaly	No	No
Facial dysmorphism	No	No
Obesity	No	No
Shape and length of digits	Normal	Normal
Ophthalmic issues	No	No

In silico analysis

Secondary structure prediction of wild-type and mutant NID1 proteins

Secondary structure comparison of normal and mutant NID1 protein revealed different changes in the folding pattern (highlighted in red boxes). These changes were detected both in upstream and downstream regions from the site of mutation (highlighted in the purple box). These changes were present on the mutant protein's strand, helix, and coil. Changes observed in the mutant protein can be clearly seen in the Figure 2b.

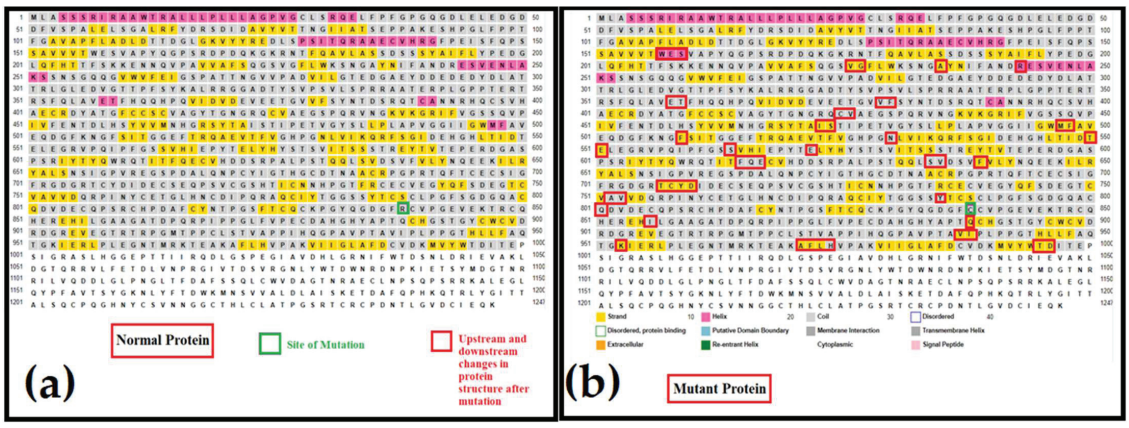

Figure 2: Secondary structure wild-type (a) and mutant NID1 protein (b) predicted through PSIPRED tools. Showing the site of mutation in green box, upstream and downstream changes after mutation are highlighted in red boxes.

Protein tertiary structure and interaction analysis

Superimposed 3D models of NID1 wild-type and mutant proteins determined a similarity index of 77.15% and different changes in folding of the 3D structure were observed (Figure 3). Protein-protein interaction of wild-type and mutant NID1 protein were done with their close functional interactor LAMC1 protein, which showed great changes in the interacting residues of all the wild-type and mutant proteins with their interactors (Figure 4). The mutant NID1 protein showed less interaction with its close interactor LAMC1 protein, *i.e.*, via 4 hydrogen bonds and 1 salt bridge only, while in the case of wild-type NID1 protein, its interaction with LAMC1 protein was through 17 hydrogen bonds and 2 salt bridges. Complete details of all the interacting residues of wild-type and mutant NID1 proteins with their close interactor are shown in Table 2.

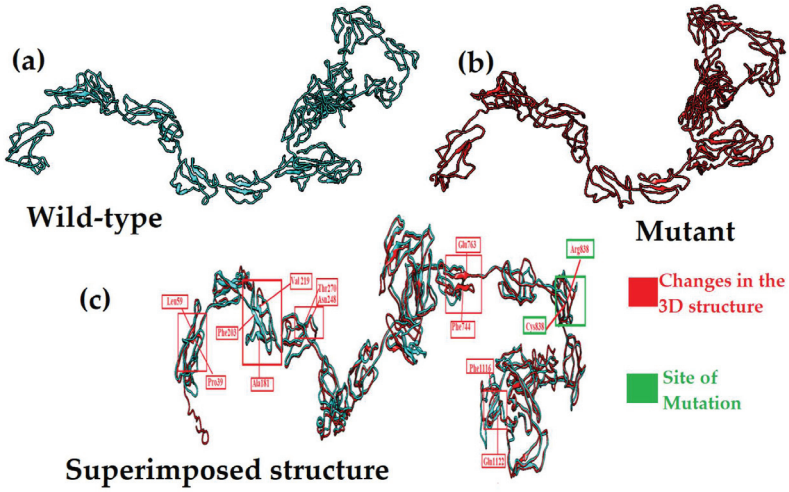

Figure 3: (a) 3D structure of wild-type NID1 protein; (b) 3D structure of mutant NID1 protein; (c) superimposed structure of wild-type and mutant NID1 protein.

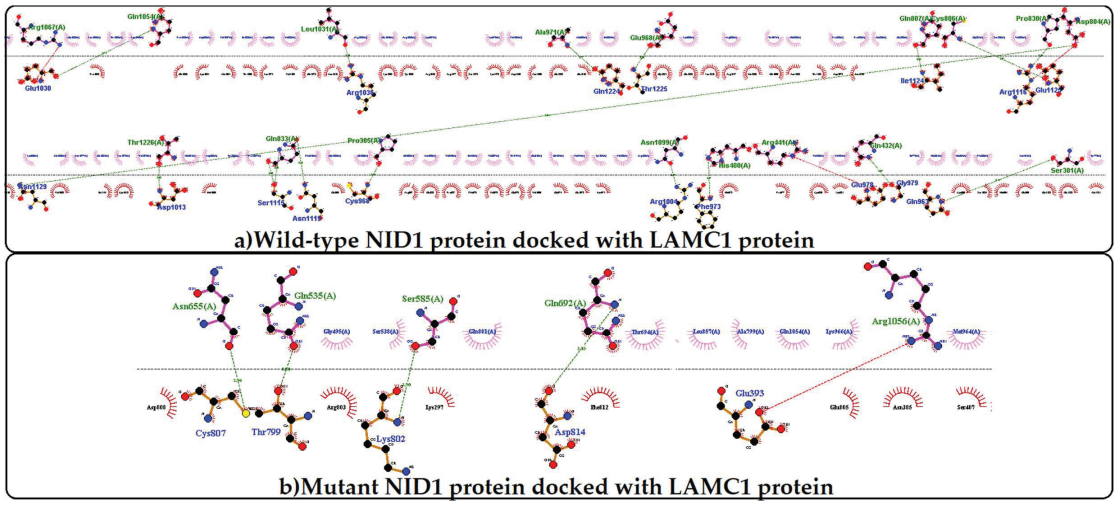

Figure 4: (a) Protein-protein interaction of wild-type NID1 and its close interactor LAMC1 protein; (b) protein-protein interaction of mutant NID1 and its close interactor LAMC1 protein

Table 2: Protein-Protein interaction of wild-type and Mutant NID1 proteins with LAMC1 protein

Protein	Interacting resides with LAMC1 protein	No and type of bonds
Wild-type NID1 protein	Glu1030, Arg1036, Gln1224, Thr1225, Ile1124, Arg1116, Glu1125, Asn1129, Asp1013, Ser1115, Asn119, Cys968, Arg1004, Phe973, Glu978, Gly979, Gln963	17 hydrogen bonds and 2 salt bridges
Mutant NID1 protein	Cys807, Thr799, Lys802, Asp814 and Glu393, and Tyr116	4 hydrogen bonds and 1 salt bridge

Discussion

The present family was recruited from Dera Ismail Khan city and ethnically it belonged to a Saraiki origin population. This family had 2 affected individuals, *i.e.*, an affected father (I: 1) and his daughter (II: 3). Both the patients were suffering from mild intellectual disability. The patient (II: 3) had some additional anomalies like attention deficits, a walking problem, and a kidney problem (Table 1). Molecular investigation in this family found a homozygous missense variant NM_002508: c.C2512T:(p.Arg838Cys) in the 12th exon of the *NID1* gene. Allele frequency of this variant in multiple population databases was found to be very low. The additional convincing evidence for *NID1* gene involvement included the presence of the variant in a large stretch of homozygous regions (Figure 1c), high expression of *NID1* gene in the brain, and strong evolutionary conservation of the substituted amino acid (p.Arg838) due to missense mutation (Figure 5). Not only p.Arg838 was conserved but the entire region was highly conserved throughout different species (Figure 5).

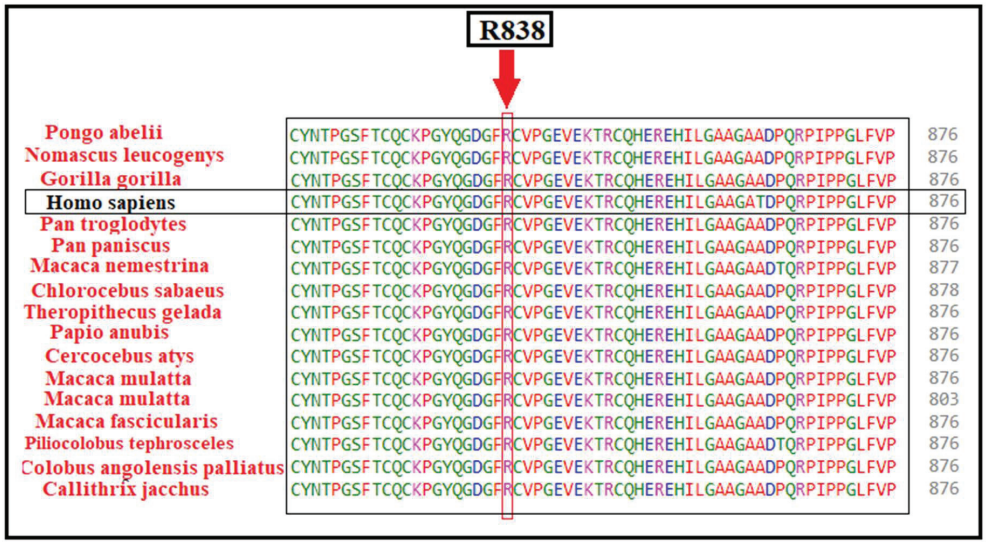

Figure 5: Conserved sequence alignment of substituted amino-acid. Position of substituted amino acid (R838) is given in red box.

Mattei *et al.* (1989) localized the human *NID1* gene to chromosome 1. Later in 1995, Zimmermann *et al.* (1995) described the *NID1* and documented that it consists of 20 exons with a genomic size of about 90kb. The NID1 protein consists of 1247 amino acids. Each exon encodes individual protein subdomains. The *NID1* gene codes a member of the nidogen family of basement membrane glycoproteins. The protein interacts with several other components of basement membranes, and may play a role in cell interactions with the extracellular matrix (Zimmermann *et al*., 1995). Jenkins *et al.* (1991) mapped the nidogen gene (Nid) to mouse chromosome 13 and

found a linkage to the beige (bg) mutation. The bg mutation is supposed to be the mouse homologue of the Chediak-Higashi syndrome (CHS). Major phenotypes of CHS comprise muscle stiffness, developmental delay, mental deficiency, and skin problems etc. (Jenkins *et al*., 1991).

Murshed *et al.* (2000) produced Nid1-deficient mice and noted that homozygous mice produced neither mRNA nor protein. Astonishingly, the homozygous mice were fertile and their basement membranes appeared normal, though Nid2 expression was much greater in certain basement membranes (Murshed *et al*., 2000). Darbo *et al.* (2013) described a large Vietnamese family with Dandy-Walker malformation with occipital cephalocele. They found a heterozygous mutation c.C1162T (p.Gln388*) in the *NID1* gene, which probably results in the complete truncation of the G2 and G3 regions of NID1 (Darbo *et al*., 2013).

Dong and his co-workers in 2002 targeted the disruption of the *NID1* in the mouse presented in their study and found a duplication of the entacin-1 locus. Homozygous mutants for the functional locus lacked NID1 mRNA and protein and often displayed seizure-like symptoms and loss of muscle control in the back legs (similar features were present in patient II:3) (Dong *et al*., 2002; Alazami *et al*., 2015). The behavior patterns suggested the presence of neurologic deficits in the CNS, thus providing genetic evidence linking entactin-1 to proper neuromuscular system functions. In homozygous mutants, structural alterations in the basement membranes were found only in selected tissues, *i.e.*, brain capillaries and the lens capsule. The structure of the basement membranes in other tissues were observed to be normal. These findings showed that the lost functions of NID1 protein results in pathologic alterations that are extremely tissue specific.

Based on the expression and animal studies, it can be showed that NID1 has significant linkage with intellectual functioning and cognition. Our patients did not disclose previously documented features, except ID, due to *NID1* gene mutation. Hence, to the best of our knowledge, *NID1* is a novel candidate gene to be involved in a new phenotype of non-syndromic ID.

CONCLUSION

The present genetic study reported a novel candidate *NID1* (c.C2512T; p.Arg838Cys) responsible for causing non-syndromic ID. Mutagenesis and cellular based expression studies on the novel identified *NID1* gene will further explore the molecular function of this gene. Linkage of additional ID families with this gene will strengthen the power of this study and assist in molecular diagnostics.

Acknowledgements

The authors are grateful to all intellectual disability family members for their voluntary participation in the present study, and the Higher Education Commission of Pakistan for supporting the present study through an NPRU grant (5564/KPK/NRPU/R&D/HEC/2016).

Data availability

The computational data are available upon request.

REFERENCES

Ahmed I., Muzammal M., Abbasi S.W., Khan M.A. & Mir A. (2021). Whole exome analysis in consanguineous Pakistani families determined ROR2 and RPTN as novel candidate genes to be involved in autosomal recessive non-syndromic intellectual disability. *International Journal of Computational Intelligence in Control* **13**(2): 163–171.

Alazami A.M., Patel N., Shamseldin H.E., Anazi S., Al-Dosari M.S., Alzahrani F. & Alkuraya F. S. (2015). Accelerating novel candidate gene discovery in neurogenetic disorders via whole-exome sequencing of prescreened multiplex consanguineous families. *Cell Reports* **10**(2): 148–161. DOI: https://doi.org/10.1016/j.celrep.2014.12.015

Ali M.Z. *et al.* (12 authors) (2022). In silico analysis identified putative pathogenic missense nsSNPs in human SLITRK1 gene. *Genes* **13**: 672. DOI: https://doi.org/10.3390/genes13040672

American Psychiatric Association (APA)(2013). *Diagnostic and Statistical Manual of Mental Disorders (DSM-5®)*. American Psychiatric Association, Washington DC, USA. DOI: https://doi.org/10.1176/appi.books.9780890425596

Bittles A.H. & Black M.L. (2010). Consanguinity, human evolution, and complex diseases. *Proceedings of the National Academy of Sciences* **107**(suppl 1): 1779–1786. DOI: https://doi.org/10.1073/pnas.0906079106

Darbro B.W., Mahajan V.B., Gakhar L., Skeie J.M., Campbell E., Wu S. & Bassuk A.G. (2013). Mutations in extracellular matrix genes NID1 and LAMC1 cause autosomal dominant D andy-Walker malformation and occipital cephaloceles. *Human Mutation* **34**(8): 1075–1079. DOI: https://doi.org/10.1002/humu.22351

Dong L., Chen Y., Lewis M., Hsieh J.C., Reing J., Chaillet J.R. & Chung A.E. (2002). Neurologic defects and selective disruption of basement membranes in mice lacking entactin-1/nidogen-1. *Laboratory Investigation* **82**(12): 1617–1630. DOI: https://doi.org/10.1097/01.LAB.0000042240.52093.0F

Durkin M. (2002). The epidemiology of developmental disabilities in low income countries. *Mental Retardation and Developmental Disabilities Research Reviews* **8**(3): 206–211.
DOI: https://doi.org/10.1002/mrdd.10039

Goddard T.D., Huang C.C. & Ferrin T.E. (2007). Visualizing density maps with UCSF Chimera. *Journal of Structural Biology* **1**(157): 281–287.
DOI: https://doi.org/10.1016/j.jsb.2006.06.010

http://compbio.charite.de/PhenIX/, accessed at 25 Jan 2021.

https://endeavour.esat.kuleuven.be, accessed at 20 Jan 2021.

https://monarch-exomiser-webdev.monarchinitiative.org, accessed at 16th Jan 2021.

https://www.ncbi.nlm.nih.gov/omim/, a Accessed at 16 May 2022.

https://www.toppgene.cchmc.org, a Accessed at 30 Jan 2021.

Jenkins N.A., Justice M.J., Gilbert D.J., Chu M.L. & Copeland N.G. (1991). Nidogen/entactin (Nid) maps to the proximal end of mouse chromosome 13 linked to beige (bg) and identifies a new region of homology between mouse and human chromosomes. *Genomics* **9**(2): 401–403.
DOI: https://doi.org/10.1016/0888-7543(91)90275-J

Kozakov D., Hall D.R., Xia B., Porter K.A., Padhorny D., Yueh C. & Vajda S. (2017). The ClusPro web server for protein–protein docking. *Nature Protocols* **12**(2): 255–278.
DOI: https://doi.org/10.1038/nprot.2016.169

Mattei M.G., Passage E., Weil D., Nagayoshi T., Knowlton R.G., Chu M.L. & Uitto J. (1989). Chromosomal mapping of human basement-membrane zone genes-laminin a chain At Locus 18p11. 31 and Nidogen At Locus 1q43. *In Cytogenetics and Cell Genetics* **51**(4): 1041–1041.

Maulik P.K., Mascarenhas M.N., Mathers C.D., Dua T. & Saxena S. (2011). Prevalence of intellectual disability: a meta-analysis of population-based studies. *Research in Developmental Disabilities* **32**(2): 419–436.
DOI: https://doi.org/10.1016/j.ridd.2010.12.018

McGuffin L.J., Bryson K. & Jones D.T. (2000). The PSIPRED protein structure prediction server. *Bioinformatics* **16**(4): 404–405.
DOI: https://doi.org/10.1093/bioinformatics/16.4.404

Murshed M., Smyth N., Miosge N., Karolat J., Krieg T., Paulsson M. & Nischt R. (2000). The absence of nidogen 1 does not affect murine basement membrane formation. *Molecular and Cellular Biology* **20**(18): 7007–7012.
DOI: https://doi.org/10.1128/MCB.20.18.7007-7012.2000

Muzammal M. *et al.* (11 authors) (2022). In silico analysis of the L-2-hydroxyglutarate dehydrogenase gene mutations and their biological impact on disease etiology. *Genes* **13**: 698.
DOI: https://doi.org/10.3390/genes13040698

Ropers H.H. (2010). Genetics of early onset cognitive impairment. *Annual Review of Genomics and Human Genetics* **11**: 161–187.
DOI: https://doi.org/10.1146/annurev-genom-082509-141640

Salvador-Carulla L., Rodríguez-Blázquez C. & Martorell A. (2008). Intellectual disability: an approach from the health sciences perspective. *Salud Pública de México* **50**(S2): 142–150.
DOI: https://doi.org/10.1590/S0036-36342008000800006

World Health Organization (WHO) (1992). The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines, volume 1. World Health Organization, Geneva, Switzerland.

Yang J. & Zhang Y. (2015). I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Research* **1**(43): 174–181.
DOI: https://doi.org/10.1093/nar/gkv342

Zimmermann K., Hoischen S., Hafner M. & Nischt R. (1995). Genomic sequences and structural organization of the human nidogen gene (NID). *Genomics* **27**(2): 245–250.
DOI: https://doi.org/10.1006/geno.1995.1038