

Missense Variation in *TPP1* Gene causes Neuronal Ceroid Lipofuscinosis Type 2 in a Family from Jammu and Kashmir-India

Arshia Angural¹, Kalaiarasan Ponnusamy², Diksha Langeh^{1,#}, Mamta Kumari^{1,#}, Akshi Spolia^{1,#}, Ekta Rai¹, Ankush Sharma³, Kamal Kishore Pandita^{4,*}, Swarkar Sharma^{1,*}

¹Human Genetics Research Group, School of Biotechnology, Shri Mata Vaishno Devi University, Kakryal, Jammu and Kashmir, India

²School of Biotechnology, Jawaharlal Nehru University, New Delhi, India

³Senior Consultant, Department of Neurology, Shri Mata Vaishno Devi Narayana Superspeciality Hospital, Kakryal, Jammu and Kashmir, India

⁴Independent Researcher, Health Clinic, Swam Vihar, Muthi, Jammu, Jammu and Kashmir, India

[#]Equal contribution

*Corresponding authors:

Swarkar Sharma, email: swarkar.sharma@smvdu.ac.in, and

Kamal Kishore Pandita, email: panditakk69@gmail.com

ABSTRACT

We report diagnosis of Neuronal Ceroid Lipofuscinosis Type 2 (CLN2), a rare, hereditary neurodegenerative disease of childhood, in a four and a half year old girl, the first child of non-consanguineous parents with no family history. Despite extensive efforts by the parents, her clinical condition remained undiagnosed and without management, until recently. Our published “Bottom-up Approach”, based on comprehensive and multidisciplinary clinical, pathological, radiographical and genetic evaluations, played key role in diagnosis of the disease. Detailed analyses involving Next Generation Sequencing confirmed a missense variation NC_00011.10:g.6616374C>T (NP_000382.3:p.Arg339Gln; rs765380155) in exon 8 of *TPP1* gene. *In silico* analyses predicted it to be highly pathogenic. Further family screening (including her both unaffected parents and asymptomatic, one year old younger sister) of the identified variation through Sanger Sequencing, revealed a perfect autosomal recessive segregation in the family. This study is the first case report on classic CLN2 from Jammu and Kashmir-India. This study is also indicating the effectiveness of our “Bottom-up Approach” in understanding rare disorders in low resource regions and the importance of timely diagnosis. Like in the proband, had diagnosis been established a bit early, the family might have benefitted at least with reference to their second child through counselling programmes.

KEYWORDS

CLN2, epilepsy, Jammu and Kashmir, loss of ambulation, neuronal ceroid lipofuscinoses type 2, neuroregression, seizures, TPP1

INTRODUCTION

Neuronal ceroid lipofuscinoses (NCLs), also referred to Batten Disease, are a diverse group of hereditary, progressive neurodegenerative diseases which predominantly affect the children [1,2]. The NCLs share marked clinical abnormalities that are associated with neurodegeneration and its consequences including a combination of progressive visual impairment (leading to blindness), cerebellar ataxia, telencephalic manifestations including epileptic seizures and progressive dementia resulting in motor and mental deterioration, and poor prognosis with severe cases leading to premature death of the patients [2,3]. Morphological hallmarks of NCLs include selective neuronal degeneration primarily in the cerebral and cerebellar cortices and retina, which is associated with progressive cerebellar and cortical degeneration and secondary fiber tract atrophy [2,3]. Other pathognomonic feature of NCLs includes the abnormal intra-lysosomal accumulation of NCL-specific lipopigment residual bodies of heterogeneous origin in neurons and extra-neuronal tissues [1]. That is why, these group of neurodegenerative disease are categorized as “Lysosomal Storage Diseases (LSDs)” [2]. The characteristic heterogeneous lysosomal storage materials include auto-fluorescent ceroid-lipopigments and lipofuscin, “subunit c of mitochondrial ATP synthase (SCMAS)”, or “sphingolipid activator proteins (SAPs)” namely saponins A and D [2,4].

NCLs usually have an onset in childhood, occasionally as early as first few months of life, or adulthood. Based on the age-of-onset of clinical symptoms, NCLs have been broadly categorised into congenital, infantile, late-infantile, juvenile, adult and late-adult forms [2]. About thirteen NCL-associated proteins are known so far, each differing in their biological functions and intracellular localization which happens predominantly in the lysosomes (CLN1, CLN2, CLN3, CLN5, CLN7, CLN10, CLN11, CLN12 and CLN13), endoplasmic reticulum (CLN6 and CLN8) and cytosol (CLN4 and CLN14) [2]. Except CLN9, the genetic basis of rest of the NCL types is delineated [5]. Furthermore, NCLs can be subdivided according to a number of disease categories as mentioned in Table 1. NCLs are genetically heterogeneous with most of the clinical types following an autosomal recessive mode of inheritance, except one adult form [1,5]. NCL-associated variations are listed online in the NCL Mutation database (<http://www.ucl.ac.uk/ncl/>).

Table 1: A list of different categorization of NCLs [2,74].

S. No.	Category	Details	Types of NCL
1	Storage material in lysosomes	Subunit C of mitochondrial ATP synthase (SCMAS)	CLN2, CLN3, CLN5, CLN6, CLN8
		Sphingolipid activator proteins (saposins A and D)	CLN1
2	Vacuolations	observed in lymphocytes of the patients	CLN3
3	Lysosomal enzyme deficiencies	Palmitoyl protein thioesterase-1 (PPT1)	CLN1
		Tripeptidyl peptidase 1 (TPP1)	CLN2
		Cathepsin D (CTSD)	CLN10

		Cathepsin F (CTSF)	CLN13
4	Non-enzyme deficiencies	lysosomal proteins	CLN1, CLN2, CLN5, CLN10, CLN13
		integral transmembrane proteins	CLN3, CLN6, CLN7, CLN12
		proteins localized to endoplasmic reticulum	CLN6, CLN8
		proteins localized to compartments in the secretory pathways	CLN11
5	Histopathology	Granular osmophilic deposits (GROD)	CLN1, CLN4, CLN5, CLN8, CLN9, CLN10, CLN12, CLN14
		Curvilinear profile (CLP)	CLN2, CLN5, CLN6, CLN7, CLN8, CLN9
		Finger print profiles (FPP)	CLN3, CLN5, CLN6, CLN7, CLN8, CLN11, CLN13, CLN14
		Rectilinear complex (RLC)	CLN6, CLN7,
6	Inheritance	Autosomal recessive	CLN1, CLN2, CLN3, CLN5, CLN6, CLN7, CLN8, CLN9, CLN10, CLN11, CLN12, CLN13, CLN14
		Autosomal dominant	CLN4

Diagnosis of NCLs in children can rather be challenging during routine clinical set up, mainly due to their heterogeneous clinical presentations and underpinning genetic heterogeneity. Previously, preliminary diagnosis of NCLs used to rely on the results of electron microscopic studies of biopsied tissues (muscle, nerve, skin, conjunctiva) derived from the patients, enzyme activity assays, neuroimaging and electrophysiological investigations, the characteristic findings of which hold an evidence to the corresponding NCL. The ultra-structural microscopic evidence of NCLs include abnormal cytosomal morphologies, each specific for different forms of NCLs, observed in the cytoplasm in peripheral blood lymphocytes or skin biopsy of the patients [6,7]. Electroencephalography (EEG) findings seen among the NCL patients include variable slowing of the background activity, paroxysmal bursts of polyphasic epileptiform discharges, and an excessive photic response [8]. Other features observed in the NCL patients include abnormal electroretinography and visual evoked potential (VEP) responses and a giant somatosensory-evoked potential (SSEP) [8]. NCL-associated neuroradiological abnormalities include varying degrees of diffuse cortical atrophy (cerebral as well as cerebellar) and periventricular/capsular signal intensity alterations on T2-weighted images [9,10]. An actual confirmation of NCL is now established after genetic screening of the patient. Additionally, lack of facilities for NCL enzyme analysis and genetic tests in poor resource regions can cause either a delay or misdiagnosis in NCL patients, further depriving them of the available therapeutic interventions as well as genetic counselling and prenatal diagnosis in the suspected couples.

Epidemiological studies on the NCLs have reported on their worldwide distribution. Being the most common neurological storage disease in children, the overall prevalence of NCLs is around 2-4 in

100,000 individuals with an estimated rate of incidence of 1 in 100,000 births [6]. However, there is a geographical variation in the prevalence of NCLs. In the US, NCL prevalence is estimated to be 1 in 12,500 births or 300 to 350 new cases per year [11]. In some regions including Finland, Newfoundland and others, certain forms of NCL are relatively frequent [3,12,13]. However, there have been no epidemiological data on NCLs from India. Only a few NCL case reports from India have been published [8,14-18]. These studies were mainly based on enzymatic and clinicopathological studies and have indicated CLN2, also known as Late Infantile Neuronal Ceroid Lipofuscinosis (LINCL), as the most common NCL-type.

In this study, we report diagnosis of autosomal recessive, classic CLN2 (OMIM# 204500) in a four and a half year old girl from Jammu and Kashmir (J&K) – India, ascertained by employing our previously published “Bottom-up Approach” [19]. Using a targeted Next-generation Sequencing (NGS) approach, the proband was found to harbour a pathogenic homozygous missense variation NM_000391.4:c.1016G>A (rs765380155) in exon 8 of CLN2-associated *TPP1* gene, causing a surface amino acid substitution of Arginine at 339 position to Glutamine (NP_000382.3:p.Arg339Gln). Upon validation of the NGS results through PCR-based Sanger Sequencing, the same variation was confirmed in the proband and both of her parents were found to be carriers. It is pertinent to mention that this is the first NCL case report from J&K region – India. Despite the onset of early symptoms in the proband almost 2-3 years back between the age of 2.5-3 and its progression into a severe disease, the case had remained undiagnosed owing to limited resources until recently when investigated comprehensively using our multi-disciplinary Bottom-up Approach [19]. Due to lack of diagnosis and appropriate counselling programmes in the region, unfortunately, the family did not undergo genetic counselling earlier. It was disheartening to find out upon genetic screening of the family that the younger, year old asymptomatic kid is also harbouring the same variation in a homozygous manner. With new developments in place for the pharmacological treatment of symptoms presented in CLN2 paediatric cases as well as enzyme replacement therapy (ERT) and gene therapy [20-23], it is anticipated if interventions are provided at this juncture too, may help the family extensively and the affected kids (in particular) in leading a better life.

CLINICAL INVESTIGATIONS

A four and a half year old girl (II-1; Figure 1(a)), presented with status epilepticus including generalized myoclonic seizures, unsteady gait, loss of ambulation, clumsy hand movements and poor cognitive functions. She was delivered healthy at full-term to non-consanguineous parents belonging to same caste group, but with no family history of a similar disorder. As informed by her parents during clinical consultation, she was overall healthy during her early development till 3 years of age, after which her health declined progressively and later witnessed a global developmental delay. At the age of around two and a half, she developed signs of neuroregression which further progressed into cognitive and motor impairment (probably due to seizures), deterioration of speech and behavioural issues. At the time of clinical investigation, other features such as hydrocephalus, presence of pin-

head sized hyperpigmented macules over her face, dystonic limbs, hypertonia, and hyper-reflexia were observed. No signs of vision loss were observed on optic evaluation with a torch light; both eyes were symmetric, pupils of her both eyes were normal in size and displayed normal pupillary response towards torch-light (pupil constriction towards bright light stimuli). She was responsive to sounds, indicating an intact hearing ability. There was no history of birth complications (such as birth asphyxia, preterm delivery) or peri-natal complications. She has a younger, a year old sibling who is clinically fit so far.

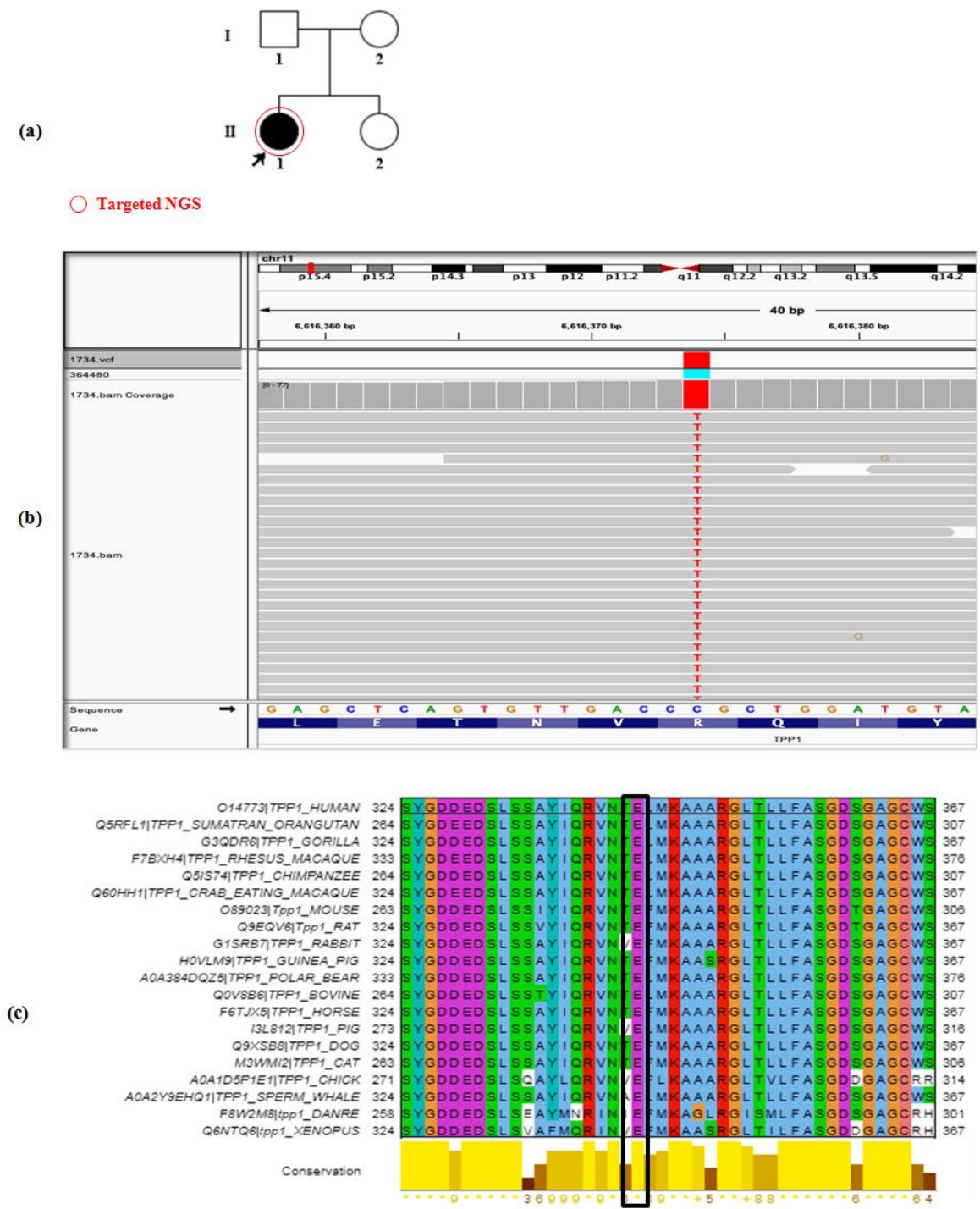


Figure 1: (a) Family pedigree representing clinical status of the proband (II-1) and her family during clinical presentation. Male is represented in square and females in circle (affected with solid representations). The proband (II-1) has been marked with a black arrow. The red circle indicates that

II-1 has undergone screening through targeted NGS. (b) An IGV snapshot of the II-1's NGS data indicating homozygous NC_000111.10:g.6616374C>T variant in exon 8 of *TPP1* gene. (c) A Jalview image showing a highly conserved TPP1 amino acid residue p.R339 aligned across selected vertebrates using ClustalO MSA.

All the clinical observations were highly suggestive of a neurodegenerative disease, most likely any of the inborn errors of metabolism, leukodystrophies, mitochondrial disorders or NCL. Further clinico-pathological investigations of the proband were conducted accordingly. The investigation of proband's blood histo-biochemical parameters was inconclusive, since no significant pathologies were detected. Her serum ceruloplasmin (39.20 mg/dL), ferritin (99.60 ng/mL), TSH (3.91 uIU/mL) and Vitamin D (25-OH) (51.8 ng/mL) levels were within biological reference range. Plasma Lactate levels (19.60 mg/dL) were normal. Creatinine (CPK) levels were normal (99.4 U/L). Her slit-lamp investigation and retinal-examination were also inconclusive; no optical pathologies were detected. Proband's MRI investigation conducted on Siemens 1.5T MR Scanner Magnetom Essenza indicated signs of generalized, mild cerebral and cerebellar atrophy, and prominent ventricular system and extra-axial cerebrospinal fluid (CSF) spaces (Supplementary Figure 1), which are non-age specific in her case. No intensity issues or blooming effect were observed in the cranial MRI. Electrophysiological studies conducted through EEG indicated the signs of asymmetric generalised epileptiform discharges or spikes during induced sleep (Supplementary Figure 2). Photoc stimulation (PS) was not conducted during the current EEG investigation. However, as per the clinical history provided by the proband's parents, a prior MRI (on a 3T-imager) as well as EEG investigation were conducted on the onset of symptoms when she was around 3 years old. That MRI investigation remained unremarkable as there were no clinically significant findings of intracranial and cervico-dorsal spine abnormality, whereas the EEG investigation indicated abnormal sleep patterns with signs of asymmetric generalised epileptiform spikes. Discharges of similar frequency were also observed on PS during the previous EEG. However, hyperventilation was not carried out during any EEG investigation of the proband. The results of these clinical investigations clearly indicated a history of progressive neurodegenerative disorder of early childhood.

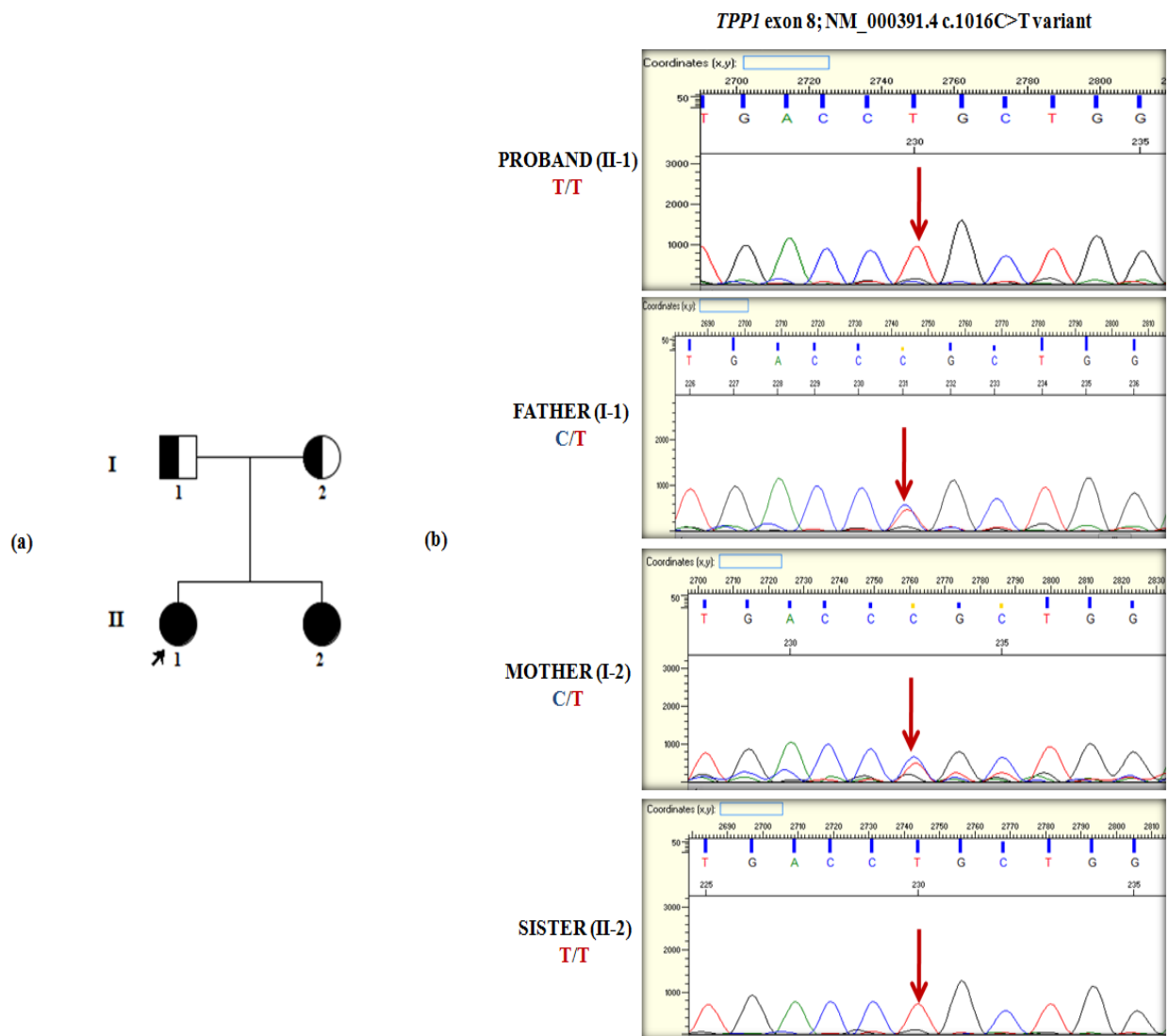


Figure 2: Autosomal recessive inheritance of NC_00011.10:g.6616374C>T *TPP1* variation; (a) The pedigree indicating autosomal recessive inheritance of the variation, and (b) Screenshots of the Sanger Sequencing electrophoregrams (forward strand) indicating the genotypes of the recruited participants for the variation.

GENETIC EVALUATION

Recruitment of Participants and Collection of Samples

Multiple rounds of clinical evaluation suggested preliminary clinical diagnosis of classic NCL2, which further indicated a need of thorough genetic evaluation including the evaluation of segregation of the disease in the family. The study has been approved under Genetics of rare diseases research work by the Institutional Ethical Review Board (IERB), SMVDU, J&K-India (SMVDU/IERB/18/67). For participation in the study and collection of the blood samples, signed informed consents were obtained from the proband's parents. About 1-2 mL of whole peripheral blood samples were drawn and collected in ethylenediaminetetraacetic acid (EDTA) vacutainers from the proband (II-1), her father (I-1), mother (I-2) and her sister (II-2) for the molecular investigations.

Extraction of DNA

DNA was extracted from peripheral blood lymphocytes of the collected blood samples using Xpress DNA Blood Mini Kit (Cat. No.: MG17BI-S50/250; MagGenome Technologies, India) by following the manufacturer's protocol. QC-analysis of the extracted DNA samples was carried out through gel electrophoresis and spectrophotometric methods.

Next Generation Sequencing and Raw Data Generation

A targeted gene sequencing approach for screening of II-1 had been carried out using the massive parallel sequencing (MPS) technology by Gene Lab, Surat – India, on an independent recommendation (list of targeted genes available in the Supplementary File). DNA libraries were constructed using Ion AmpliSeq™ Library Kit, Ion Xpress™ Barcode Adapters and Ion Library Equalizer™ Kit (ThermoFisher Scientific Inc., US). The templates were prepared for sequencing reactions using Ion OneTouch™ 2 System and Ion PI™ HiQ™ OT2200 Kit (ThermoFisher Scientific Inc., US). The sequencing was performed using an Ion Proton™ Sequencer, Ion PI™ Hi-Q™ Sequencing200 Kit and Ion PI™ Chip Kit (ThermoFisher Scientific Inc., US), and the calls were made using the Torrent Suite™ Software and Ion Reporter™ Software (ThermoFisher Scientific Inc., US).

NGS Data Analysis

The raw data was requested and retrieved from the company through the family and after obtaining informed consent of the parents, detailed data re-analysis was performed. For the alignment of sequencing reads generated from the proband's genomic sample, the human reference genome hg38 was used. About 27,426,174 reads were mapped to the reference genome. Variants were filtered for their stringencies including evolutionary conservation, minor allele frequencies and pathogenicity predictions based on the public genomic databases such as the GERP++, gnomAD, 1000 Genomes Browser, and others. Filters were applied to the NGS data for the identification of variants with a minor allele frequency of $\leq 5\%$, variants classified as disease-causing in public databases, and variants with a minor allele frequency of $\leq 1\%$ predicted to be loss-of-function variants. Genotype-phenotype correlations were then evaluated for each variant resulting from the filtering strategies. Only the variants with 30X or more coverage were considered for the clinical correlation. Non-synonymous and splice site variants found in the genes and relevant to the clinical symptoms were used for clinical interpretation. Silent variations that do not result in any change in amino acid in the coding region were not reported. Reanalyses of the NGS data for quality assurance was carried out by the pipeline developed by Key2Genes®, India and as previously described [24].

Validation of NGS findings using targeted PCR-based Sanger Sequencing

In order to validate the NGS findings and evaluate the inheritance pattern of identified variant in the recruited family, a targeted bi-directional PCR-based Sanger Sequencing (targeting the exonic region in which the variation was identified as well as its flanking exon-intron junctions) was performed

twice for each participant (sequence of forward and reverse primers, and details of PCR reaction and amplicon purification are available upon request). Sanger sequencing of the amplicons were facilitated by Biodroid® Innovations, India. Sequence electrophoregrams were visualized with Sequence Scanner Software 2 (Applied Biosystems, US).

COMPUTATIONAL ANALYSES

Multiple Sequence Alignment (MSA) analysis was performed through ClustalOmega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>), the results of which were further viewed based on default ClustalX colour scheme using Jalview tool [25,26]. Pathogenicity prediction of the identified variations were performed using some computational *in silico* tools based on algorithms developed to predict the effect of variations on protein structure and function. This included individual analyses using MutationAssessor (<http://mutationassessor.org/r3/>), Mutation Taster (<http://www.mutationtaster.org/>), Sorting Intolerant From Tolerant (SIFT; <https://sift.bii.a-star.edu.sg/>), PolyPhen-2 (<http://genetics.bwh.harvard.edu/pph2/>), and PoPMuSiC2.1 and SNPMuSiC tools of the dezyne web-suite (<http://www.dezyne.com/>), and the prediction results were further validated using Ensembl Variant Effect Predictor (VEP) tools [27-31].

Besides, 100 ns time-lag computational *in silico* molecular dynamics simulations (MDS) were performed for the wild-type as well as the variant protein models of the protein encoded by the indentified disease-associated gene. A complete protocol for the MDS has been provided in the Supplementary file. Using MDS data for the trajectories of wild-type and variant protein models, the Root Mean Square Deviation (RMSD), protein Radius of Gyration (Rg), Root Mean Square Fluctuation (RMSF) of alpha-carbon atoms (C α 's), Solvent Accessible Solvent Area (SASA) and hydrogen bond contacts were calculated. PyMol was used for *in silico* structural visualization of the protein models.

RESULTS

Detailed analyses of NGS data revealed that II-1 was harbouring a homozygous missense variation NC_00011.10:g.6616374C>T (NM_000391.4:c.1016G>A; rs765380155) in exon 8 of *TPP1* gene. An IGV screenshot for the same has been provided in Figure 1(b). It is a rare variation that has records in ClinVar (Accession Id: VCV000198725.2) and genomAD databases, but not in the 1000 Genomes database. The NM_000391.4:c.1016G>A variation has caused a surface amino acid substitution from Arginine (R) to Glutamine (Q) at 339 position (NP_000382.3:p.Arg339Gln (p.R339Q)) in TPP1 protein. MSA of human TPP1 (hTPP1; Uniprot ID: O14773) protein sequence against TPP1 sequence of other species using Clustal Omega online program and UCSC Genome Browser indicated that hTPP1 p.R339Q variation has occurred at a highly evolutionarily conserved residue position (R339) across vertebrates (Figure 1(c); Supplementary Figure 3)). Since, both R- and Q-residues have different physico-chemical properties, R339 residue being a surface amino acid and is evolutionary conserved in nature, there is a high likelihood that this amino acid substitution might have a plausible functional relevance in disease pathogenesis. Pathogenicity prediction performed using some *in silico*

tools confirmed the deleterious nature of hTPP1 p.R339Q variation and predicted it to be disease-causing. The individual predictions were – MutationAssessor (FI score = 2.925, functional impact = medium) (<http://mutationassessor.org/r3/>), Mutation Taster (p-score = 0.999997; prediction = disease causing), SIFT (score = 0; prediction = damaging), and PolyPhen2 (score = 1.000; prediction = probably damaging). These pathogenicity predictions were further validated using Ensembl VEP tools as likely pathogenic or pathogenic. Thermodynamic stability predictions conducted through PoPMuSiC2.1 and SNPMuSiC tools of the dezyne web-suite predicted hTPP1 p.R339Q variation to be deleterious and probably resulting in thermodynamic destabilization of the variant protein.

Since, loss-of-function *TPP1* gene variations result in a deficient lysosomal enzyme Tripeptidyl peptidase 1, associated with a rare, fatal, autosomal recessive neuronal disorder named CLN2, NGS findings and genotype-phenotype correlation confirmed the diagnosis of classic CLN2 in II-1, also indicated by a pathologically low enzymatic activity of TPP1 in the proband's blood sample (<0.49 $\mu\text{mol/L/h}$; pathological reference = <0.5 $\mu\text{mol/L/h}$). Additional analyses of Sanger sequencing data confirmed the presence of homozygous variation NM_000391.4:c.1016C>T in II-1 and its perfect autosomal recessive segregation in the family (Figure 2(a) and (b)). Both of her parents (I-1 and I-2) were found to be heterozygous carriers of the variation. This variation is already reported in limited CLN2 patients [32-34]. As the family had not known CLN2 as a genetic disease and did not undergo any genetic counselling programme, we proposed the parents for molecular screening of the variation in II-2. Unfortunately II-2, otherwise asymptomatic, was also found to be harbouring the same variation in a homozygous manner.

For understanding the biochemical features of p.R339Q hTPP1 variant protein and its plausible functional role in disease pathogenesis, comparative MDS of 100 ns duration were performed for wild-type R339 and variant Q339 hTPP1 protein models. Within 10 ns of the MDS, both R339 and Q339 protein models were found to be stable. After completion of MDS, the RMSD, Rg, RMSF, SASA and hydrogen bond contact were calculated. The RMSD values of the wild-type R339 (lowest: 0.000501 nm to highest: 0.200251 nm) protein model differed from that of the variant Q339 model (lowest: 0.000506 nm to highest: 0.195172 nm). The Rg for R339 and Q339 hTPP1 models were analyzed from a 100 ns simulation. The average values of Rg in R339 and Q339 were 2.32 nm and 2.33 nm, respectively. During MDS, it was observed that the molecular structure of variant Q339 became relatively less compact after 60 ns, whereas wild R339 depicted its characteristic structural compactness. The RMSF differences indicated that the dynamics of the R339 and Q339 hTPP1 models had four loop regions that displayed larger flexibility. The Q339 showed high fluctuation at position 192 with overall average flexibility of value 0.08 nm, whereas for R339 it was 0.07 nm. Amino acid position 369 of R339 model was found to be highly flexible in comparison to that of Q339 (which showed better flexibility at amino acid positions 93 and 192). The SASA calculations indicated the solvent accessible hydrophobic and hydrophilic surface contacts of both R339 and Q339 models. The average values of SASA showed that Q339 (215.96 nm^2) protein is slightly buried in comparison to Q339 (216.48 nm^2). The graphical plots for the Root Mean Square Deviation (RMSD),

protein Radius of Gyration (Rg), Root Mean Square Fluctuation (RMSF) and solvent accessible solvent area (SASA) analyses generated for hTPP1 R339 and Q339 models using MDS data have been represented in Supplementary Figures 5 (a), (b), (c), and (d), respectively. We also analyzed the inter- as well as intra-molecular hydrogen bond contacts of R339 and Q339 trajectories; the plots for the same have been provided in Supplementary Figures 6(a) and (b). R339 showed high number of intra-molecular hydrogen bond contacts (No.'s = 391) as compared to Q339 (No.'s = 381). Further, the residues involved in the catalytic activity of the hTPP1 protein, including E²⁷², D²⁷⁶, D³²⁷, D³⁶⁰ and S⁴⁷⁵, were subjected to intermolecular interaction analysis. It was observed that the variant Q339 hTPP1 model has slightly higher interactions within the catalytic residues. The total hydrogen bonds formed in all the conformation of the catalytic residues in the wild-type R339 and the variant Q339 hTPP1 models were 9953 and 10942, respectively (Supplementary Figure 6(b)). In some protein conformations, it was observed that three hydrogen bonds were established between the catalytic residues in Q339 model, whereas a maximum of two hydrogen bonds in R339 model. A hydrogen bond between E²⁷² and D²⁷⁶ residues with 2.5Å was observed in R339 model, whereas an additional hydrogen bond was observed between D²⁷⁶ and S⁴⁷⁵ residues with 2.0Å in Q339 (Supplementary Figure 7). In summary, overall analyses of MDS trajectories of both R339 and Q339 hTPP1 models revealed that the variant Q339 model relatively shows more dynamic flexibility and more hydrogen bond contacts than wild-type R339. All this could potentially result in the functional disruption of hTPP1 variant Q339.

DISCUSSION

Pathogenic changes in *TPP1* gene (also referred as *CLN2*; OMIM# 607998), mapped to cytogenetic location – 11p15.4, cause an autosomal recessive, lysosomal storage disorder named CLN2, also referred to as “LINCL”, “late infantile Batten disease”, or “Jansky-Bielschowsky disease” [35-38]. CLN2 is a rare progressive, fatal paediatric neurodegenerative disorder with typical onset of CNS and retinal manifestations between 2-4 years of age, progressing with worsening of clinical severity and leading to death by 6-15 years of life or later [37]. Clinical manifestation of CLN2 is usually heralded by partial, generalized tonic-clonic or secondarily generalized seizures, followed by ataxia, dementia, epilepsy, myoclonus, loss of vision (due to macular and retinal degeneration and optic atrophy) and speech, cerebellar, pyramidal and extra-pyramidal signs, and developmental regression, leading to death in the second decade of life [1,37,39]. The patients suffer loss of ambulation due to motor regression accompanied with progressive deterioration of speech and vision, resulting in inability to walk or sit unsupported and speak and blindness. The CLN2-associated neurohistopathological features include brain atrophy, neuronal loss and lysosomal accumulation of curvilinear bodies and auto-fluorescent storage material rich in SCMAS in neurons and other cell types in the CNS and peripheral tissues [40,41]. The accumulation of SCMAS in late-infantile and juvenile NCL has been described as an epiphenomenon and is not considered as disease-specific [42]. However, the underpinning biological mechanism for CLN2-associated neuronal and retinal degeneration has not yet determined. Despite CLN2-associated clinical heterogeneity/penetrance issues have not been

reported in the literature, yet a few number of atypical CLN2 cases with late age-of-onset and prolonged disease course are reported worldwide [43,44]. Although CLN2 is an ultra-rare disease with an estimated worldwide incidence of 1:100,000 live births, it is the most common form of NCL among paediatric cases [3,45].

A precise diagnosis of CLN2 is quite challenging during early stages of its precipitation and is often delayed until its significant progression in severity, leading to complex diagnostic odyssey characterized by several misdiagnoses in a clinical set up [23]. Delayed diagnosis or misdiagnosis of a disease usually results in disjointed care and treatment delay for the patients. Although brain imaging provides relatively non-specific findings, marked neuroradiological signs of cerebellar atrophy are observed at early disease stage in CLN2 patients [9]. However, the characteristic neuroelectrophysiological findings in EEG include a triad consisting of early presentation of a typical photo-paroxysmal response (PPR) known as “paroxysmal spike-wave response” against low frequency intermittent occipital photic stimulation (using flash rates at 1 to 2 Hz), grossly enhanced SSEPs and a late presentation of grossly enhanced cortical VEPs and diminished EEG [39,46,47]. For the confirmation of a CLN2 case suspected through differential diagnosis, the demonstration of decreased TPP1 enzyme activity (in leukocytes, fibroblasts or dried blood spots) through enzyme activity assays and molecular detection of *TPP1* (or *CLN2*) biallelic pathogenic variations have been regarded as the gold-standard diagnostic approaches [48].

The clinical management of CLN2 is complex, which primarily encompasses supportive and palliative care, pharmacological treatment of symptoms presented in CLN2 paediatric cases, enzyme replacement therapy (ERT) and gene therapy [20-23]. Antiepileptic drugs, including benzodiazepines (clobazam/clonazepam, lamotrigine, levetiracetum and valproate, are the common first-line therapeutics used for the management of epileptic as well as non-epileptic seizures and myoclonus [22]. A set of tailored inter-disciplinary therapies including physical, occupational, speech/orofacial myofunctional, nutritional, ophthalmologic, behavioural and complementary therapies are also sometimes sought for the management of CLN2-associated gait issues, pain, motor disturbances, retinal degeneration, sleep disorders, speech problems and behavioural issues in affected children as described in the literature [22]. For the treatment of symptoms in the proband, anti-epileptics including briviact (50 mg/day), rivotril (0.5 mg/day) and valparin oral solution (4 mL twice a day), and supplements including memocart oral solution (3 mL twice a day) and omilcal suspension (10mL/day) were clinically recommended. For the management of movement disorder, physiotherapy was recommended. Administration of a recently FDA- and EMA approved enzyme replacement therapy for CLN2, based on a recombinant proenzyme form of human TPP1 “cerliponase alfa”, [20] was also recommended, an early adoption of which could essentially aid in the long-term management of the disease as well as prevention of further disease progression in II-1 and prevention of precipitation of CLN2-associated clinical manifestations in her younger asymptomatic sibling (II-2).

More than 150 distinct CLN2-associated *TPP1* gene variations have been reported so far, information on which is available in the NCL Disease database (<https://www.ucl.ac.uk/ncl-disease/mutation-and->

[patient-database/mutation-and-patient-datasheets-human-ncl-genes/cln2-tpp1](#)). There is a marked allelic heterogeneity associated with CLN2. Different types of associated *TPPI* genetic changes include several missense, non-sense and splice-junction variations, and frame-shift changes including single-nucleotide insertions and small deletions [49]. Many of these changes have been reported to cause either reduced activity or inactivation of hTPP1 due to its disrupted folding, intracellular processing, and trafficking [4,50,51]. It has been estimated that nearly half of the reported CLN2 patients harbour homozygous changes in *TPPI* gene, while the remaining patients are compound heterozygous for different variations [34].

The most common CLN2-associated *TPPI* variations include a splice site variation c.509-1G>C and a nonsense variation c.622C>T (p.R208*), found in about 60% of the global CLN2 cases [49,52,53]. These variations are particularly common in CLN2 patients of North American and European descent [53]. The second most common *TPPI* variation in North America is c.851G>T (p.G284V) [53,54]. Besides, several CLN2-associated *TPPI* founder variations have also been reported from South American, European, middle-eastern, and South Asian countries [3,13,33,53]. Information on the *TPPI* variations previously reported in CLN2 patients of Indian origin have been provided in the Table 2 and are illustrated in the Supplementary Figure 4. These include p.Arg152Ser, p.Ser153Profs*19 and p.Tyr157* in exon 5, p.Arg206Cys and p.Arg208* in exon 6, p.Arg339Trp, p.Arg339Gln, p.Met345Leu in exon 8, p.Tyr459Ser in exon 11, and p.Phe516* in exon 12 [32,53]. Some patients have been found to harbour *TPPI* compound heterozygous variations including p.Asn286Ser and p.Ser62Argfs*19 in exon 7 and exon 3, respectively, while some harbouring p.F230Sfs*28 and p.Ile484Aspfs*7 in exon 7 and exon 12, respectively [32]. Few atypical CLN2 cases from India have been associated with some *TPPI* intronic variations, including c.89+1G>A in intron 2 and g.2023_2024insT in intron 4 [32,55].

Table 2: A list of *TPPI* variations identified in the Indian CLN2 patients and previously reported in the literature.

Zygosity	Variation	Location in <i>TPPI</i> Gene	Variation Type	Amino acid change	dbSNP Accession ID	Reference
Homozygous	c.89+1G>A	Intron 2	Splice site	NA	rs1855631590	[55]
	g.2023_2024insT	Intron 4	Insertion	NA	NA	[32]
	c.456G>C	Exon 5	Missense	p.Arg152Ser	rs869025274	[32]
	c.455_488del	Exon 5	Missense, Frameshift deletion	p.Ser153Profs	NA	[32]
	c.471C>A	Exon 5	Nonsense	p.Tyr157*	rs553522118	[32]
	c.616C>T	Exon 6	Missense	p.Arg206Cys	rs28940573	[32,53]
	c.622C>T	Exon 6	Nonsense	p.Arg208*	rs119455955	[32]
	c.1015C>T	Exon 8	Missense	p.Arg339Trp	rs750428882	[32]
	c.1016G>A	Exon 8	Missense	p.Arg339Gln	rs765380155	[32]; current study (parents are carriers)

	c.1376A>C	Exon 11	Missense	p.Tyr459Ser	rs864309505	[32]
	c.1547_1548delTT	Exon 12	Nonsense, Frameshift deletion	p.Phe516*fs	rs763961289	[32]
Heterozygous	c.1033A>C	Exon 8	Missense	p.Met345Leu	rs141482368	[32]
Compound Heterozygous	c.184delT;	Exon 3	Missense, Frameshift deletion	p.Ser62Argfs	rs1554902217	[32]
	c.857A>G	Exon 7	Missense	p.Asn286Ser	rs119455958	[32]
	c.689delT;	Exon 7	Missense, Frameshift deletion	p.Phe230Serfs	rs1554901898	[32]
	c.1449_1450insG	Exon 12	Missense, Frameshift insertion	p.Ile484Aspfs	rs1057516264	[32]
Abbreviations and Symbols: del = deletion; fs = frameshift; ins = insertion; NA = Not available; * = Termination The above variations refer to <i>TPP1</i> reference sequence numbers NM_000391.3, NG_008653.1 and NP_000382.3						

The NC_00011.10:g.6616374C>T identified in the current study is a rare variant and has already been reported among limited CLN2 cases in India and some other countries [32-34]. However, another *TPP1* variant NM_000391.4: c.1015C>T (rs750428882) resulting in substitution of R-amino acid with Tryptophan (W) at the same amino acid position (p.R339W) is also known in limited CLN2 patients across the globe [32,56]. It has also been implicated in a study that pathogenic changes in *TPP1* gene results in Spinocerebellar Ataxia 7 (SCAR7; OMIM# 609270), also known as atypical CLN2, which is characterized by residual TPP1 enzyme activity and neurological manifestations including ataxia and cerebellar atrophy but without seizures and retinal degeneration [27]. However in the present study, SCAR7 was ruled out in II-1 based on her overall classic clinical presentation.

In order to deduce the functional nature of NC_00011.10:g.6616374C>T (p.R339Q) variation, MDS for wild-type R339 and variant Q339 hTPP1 protein models were performed. On subjecting the MDS trajectories of these models to RMSD, Rg, RMSF and SASA analyses, the overall results indicated slight conformational changes in Q339 model rendering it relatively less compact in structure and more flexible in comparison to the wild-type R339. It is to be noted that conformational changes induced in a protein molecule by an amino acid change can usually impact its overall molecular dynamicity, function, its folding into a functional secondary structure, and functional interaction between its amino acid compositions and the surrounding environmental factors or molecules. All these can be simulated *in silico* through MDS indicating the RMSD, RMSF, Rg and SASA profiles of a protein model. SASA analysis of a protein molecule can be highly essential for the functional annotation of a disease-associated variant, in the sense that pathogenicity is highly associated with the buried property of a protein molecule and should always be considered while performing MDS for a potentially pathogenic protein variant [57]. Since in our case study, SASA analysis revealed hTPP1 Q339 variant to be slightly buried than R339, the variation can be considered as potentially pathogenic in nature. Intermolecular hydrogen bond analysis depicted more hydrogen bonds in the catalytic site of Q339 variant in comparison to the wild-type R339 hTPP1 protein model. Studies have

shown that Ser⁴⁷⁵ amino acid residue is the active site nucleophile, which along with Asp³⁶⁰ and Asp⁵¹⁷ residues are essential for the catalytic activity of TPP1 enzyme [58]. Besides, Glu²⁷², Asp²⁷⁶ and Ser⁴⁹⁵ residues in human TPP1 are also crucial to its catalytic activity [4]. The additional hydrogen bond formation in the catalytic site between D²⁷⁶ (Asp²⁷⁶) and S⁴⁷⁵ (Ser⁴⁷⁵) residues of the variant Q339 can be considered as contributory to its compromised biological function. Furthermore, the findings of our studies suggest that the Arg³³⁹ residue is also crucial for the enzymatic activity as well as overall molecular dynamics of the hTPP1 protein.

Studies have shown that newly synthesized misfolded proteins including lysosomal hydrolases or variants resulting in protein-misfolding undergo oligomerization in endoplasmic reticulum (ER) and are later sorted for ER-associated degradation (ERAD) which includes retro-translocation from ER to the cytoplasm for their ubiquitin/proteasomal degradation, resulting in their reduced half-life [51,59,60]. Overall MDS analyses of hTPP1 R339 and Q339 protein models in our study have revealed slight conformational discrepancies in Q339 which can likely cause a slight misfolding of the hTPP1 variant Q339, rendering it to be either functionally compromised or resulting in a faulty cellular processing and, subsequently, contributing pathogenically towards the disease manifestation in this case. However, further *in vivo* studies are required to prove the potential functional aspects (protein misfolding, low half-life, residual or no activity, excessive secretion to the extra-cellular space) of hTPP1 Q339 variant, which is beyond the current context of this study.

In general, *CLN2* gene encodes tripeptidyl-peptidase 1 (TPP1, EC 3.4.14.9), a biologically crucial and ubiquitously expressed lysosomal endopeptidase that belongs to a family of S53-type of serine-carboxyl proteinases ubiquitously expressed in mammalian tissues (including of human, rat and mouse) [41,61-63]. It has been demonstrated that the expression of TPP1 in human cerebral cortex usually increases with development and reaches the optimal levels after the age of 2 [64]. However, studies have demonstrated that fibroblast of *CLN2* patients harbouring pathogenic *TPP1* variations have less than 5% of TPP1 activity [65]. Similar findings were also observed through fluorimetric assay for TPP1 enzymatic activity in the proband. Some *TPP1* variations that result in atypical disease form probably do not completely abolish the function of the encoded TPP1 protein.

TPP1 is synthesized as a 563 amino acid containing inactive proenzyme or precursor known as “pro-TPP1” that when exposed to *in vivo* low pH conditions in the lysosome, gets auto-proteolytically processed into an active 368 C-terminal amino acids containing TPP1 [65-68]. During this maturation process, the initial signal sequence (19-amino acids) and further propeptide (176 amino acids) are removed from the pro-TPP1. Studies have shown that the autocatalytic processing of pro-TPP1 into active TPP1 is a Ca²⁺-dependent process [61]. The pro-TPP1 has an apparent Mr of 67 kDa, whereas that of the mature TPP1 is 46 kDa [67]. The amino-terminus of mature TPP1 begins at residue L196 [67]. The mature TPP1 contains a catalytic domain and five N-glycosylation sites [69]. The function of the active TPP1 protease is to sequentially cleave off tripeptides from the free unsubstituted amino-termini of oligopeptides during lysosomal protein catabolism [35,61,65]. Several *in vitro* studies have demonstrated that TPP1 degrades certain hormones and neurotransmitters such as Angiotension II and

III, β -amyloid, a Bcl2-interacting protein named Bid, cholecystokinin, glucagons, Neuromedin B, and Substance P, as well as SCMAS (that gets accumulated in most forms of NCL) [41,70-72]. However, all the *in vivo* substrates of active TPP1 are still unknown. Experimental studies have demonstrated that 3,4-dichloroisocoumarin and diisopropyl fluorophosphate inhibit the activity of TPP1 [67].

CONCLUSIONS

A four and a half year old girl presenting signs of neuroregression and an epileptiform disorder was diagnosed with CLN2 using a combinatorial approach named “Bottom-up Approach” [19]. Through a genotype-phenotype correlation based approach followed by sequencing data analyses and *in silico* predictions, the proband was diagnosed with CLN2 associated with a potential loss-of-function variation NC_00011.10:g.6616374C>T (rs765380155) in exon 8 of *TPP1* gene. The relevance of this study in the current context is that it would add information to the clinical literature and mutational landscape of CLN2-associated *TPP1* gene, particularly that of the population of J&K – India.

The authors highly recommend a comprehensive evaluation of paediatric cases suspected with CLN2 presenting with global development delay, seizures, loss of ambulation, ataxia and neuronal manifestations like brain atrophy and inclusion of CLN2-associated clinical guidelines to clinical specialists in their routine practice [73]. A combined clinical, pathological and molecular approach like “Bottom-up Approach” [19] for the precise delineation of rare diseases such as clinically suspected, paediatric NCL cases is highly recommended. Altogether this would ensure an early precise diagnosis of the disease (as depicted through this case report), thereby, facilitating an earlier clinical management or treatment to the symptomatic as well as even pre-symptomatic individuals using available interventional approaches. An early diagnosis would allow timely genetic counselling to be offered to the families of the patients, thereby, enabling genetic screening of the suspected families and their close-relatives as well as pre-implantation genetic diagnosis of the suspected couples or prenatal diagnosis through enzyme assay, identification of typical NCL inclusions through microscopy and genetic screening. In the current study, an immediate genetic counselling of the proband’s family had prompted genetic screening of her parents and younger sibling which has lead to the detection of familial segregation of the identified variation. It is to be mentioned that genetic screening has also lead to the delineation of CLN2 in proband’s younger sister, even before the manifestation of the disease. A timely detection have invoked consideration of the potential therapeutic measures so that future progression and precipitation of disease manifestations in the proband and her sister, respectively, could be prevented. A rigorous genetic counselling would also help in increasing the awareness about CLN2 and in providing information on disease prognosis to suspected CLN2 families. The knowledge obtained would also help in knowing the exact epidemiology of NCL particularly in J&K, to facilitate the framing of provisions for a better allocation of health (diagnostic and therapeutic) and financial resources for the care of patients and their families, as well as promoting research and development for understanding the underpinning molecular etiology of NCL and designing of personalized therapeutics.

LIST OF ABBREVIATIONS

Arginine (R)

Cerebrospinal fluid (CSF)

Electroencephalography (EEG)

Endoplasmic reticulum (ER)

ER-associated degradation (ERAD)

Ethylenediaminetetraacetic acid (EDTA)

Glutamine (Q)

Human Tripeptidyl peptidase 1 protein (hTPP1)

Institutional Ethical Review Board (IERB)

Jammu & Kashmir (J&K)

Late Infantile Neuronal Ceroid Lipofuscinosis (LINCL)

Lysosomal storage diseases (LSDs)

Massive parallel sequencing (MPS)

Molecular dynamics simulation (MDS)

Multiple Sequence Alignment (MSA)

Neuronal ceroid lipofuscinoses (NCLs)

Next-generation Sequencing (NGS)

Palmitoyl-protein thioesterase 1 (PPT1)

Photoparoxysmal response (PPR)

Photic stimulation (PS)

Radius of gyration (Rg)

Root mean square fluctuation (RMSF)

Root mean square deviation (RMSD)

Solvent accessible solvent area (SASA)

Somatosensory-evoked potential (SSEP)

Subunit c of mitochondrial ATP synthase (SCMAS)

Sphingolipid activator proteins (SAPs)

Tripeptidyl peptidase 1 (TPP1)

Visual evoked potential (VEP)

DECLARATIONS

Authors' Contributions

SS and KKP designed the work plan. SS, ER and AA carried out analyses of the sequencing data. AA prepared the manuscript. AA and KP performed computational *in silico* analyses. AS, KKP made and reviewed the clinical findings presented in the manuscript. DL, MK and AS have contributed equally in the experimental work. All the authors have critically reviewed the manuscript and contributed in its finalization.

Conflict of Interest

The authors declare no conflict of interest associated with the manuscript. For declaration purposes, SS is the founder and chief scientific advisor of Biodroid® Innovations Private Limited, India.

Ethics and Reporting

The study was approved by the IERB, SMVDU, J&K, India. Authors declare at manuscript submission that all relevant ethical guidelines have been followed, all necessary IRB and/or ethics committee approvals have been obtained, all necessary patient/participant consent has been obtained and the appropriate institutional forms have been archived.

Funding Statement

The authors have nothing to declare.

Acknowledgements

The authors would like to thank the proband and her family members for their participation in this study. A special acknowledgement to Dr. Parvinder Kumar, Assistant Professor, Institute of Human Genetics, University of Jammu, J&K-India and Dr. Sushil Razdan, a renowned Neurologist based in J&K-India for providing their critical reviews on the manuscript. Authors are highly grateful to Dr. Ankit Mahajan (SERB N-PDF, HGRG, SMVDU) for facilitating the acquisition of radiographical MR-images, Dr. Varun Sharma and Dr. Indu Sharma (both V.S. and I.S. are researchers at Ancient DNA Laboratory, Birbal Sahni Institute of Paleosciences, Lucknow, UP – India and members of HGRG, SMVDU) for their critical reviews, and Mr. Hemender Singh and Ms. Shipra Kowra for their experimental support. We would also like to acknowledge Gene Lab, Surat – India for facilitating the

NGS studies, Key2Genes® for their NGS pipeline and Biodriod® Innovations Private Limited, India for facilitating sequencing of the samples.

REFERENCES

1. Mole, S.E.; Williams, R.E.; Goebel, H.H. Correlations between genotype, ultrastructural morphology and clinical phenotype in the neuronal ceroid lipofuscinoses. *Neurogenetics* **2005**, *6*, 107-126, doi:10.1007/s10048-005-0218-3.
2. Carcel-Trullols, J.; Kovacs, A.D.; Pearce, D.A. Cell biology of the NCL proteins: What they do and don't do. *Biochim Biophys Acta* **2015**, *1852*, 2242-2255, doi:10.1016/j.bbadis.2015.04.027.
3. Santorelli, F.M.; Garavaglia, B.; Cardona, F.; Nardocci, N.; Bernardina, B.D.; Sartori, S.; Suppiej, A.; Bertini, E.; Claps, D.; Battini, R.; et al. Molecular epidemiology of childhood neuronal ceroid-lipofuscinosis in Italy. *Orphanet J Rare Dis* **2013**, *8*, 19, doi:10.1186/1750-1172-8-19.
4. Mole, S.E. The genetic spectrum of human neuronal ceroid-lipofuscinoses. *Brain Pathol* **2004**, *14*, 70-76, doi:10.1111/j.1750-3639.2004.tb00500.x.
5. Mole, S.E.; Cotman, S.L. Genetics of the neuronal ceroid lipofuscinoses (Batten disease). *Biochim Biophys Acta* **2015**, *1852*, 2237-2241, doi:10.1016/j.bbadis.2015.05.011.
6. Simonati, A.; Rizzuto, N. Neuronal ceroid lipofuscinoses: pathological features of bioptic specimens from 28 patients. *Neurol Sci* **2000**, *21*, S63-70, doi:10.1007/s100720070042.
7. Goebel, H.H.; Sharp, J.D. The neuronal ceroid-lipofuscinoses. Recent advances. *Brain Pathol* **1998**, *8*, 151-162, doi:10.1111/j.1750-3639.1998.tb00142.x.
8. Jadav, R.H.; Sinha, S.; Yasha, T.C.; Aravinda, H.; Gayathri, N.; Rao, S.; Bindu, P.S.; Satishchandra, P. Clinical, electrophysiological, imaging, and ultrastructural description in 68 patients with neuronal ceroid lipofuscinoses and its subtypes. *Pediatr Neurol* **2014**, *50*, 85-95, doi:10.1016/j.pediatrneurol.2013.08.008.
9. Santavuori, P.; Vanhanen, S.L.; Autti, T. Clinical and neuroradiological diagnostic aspects of neuronal ceroid lipofuscinoses disorders. *Eur J Paediatr Neurol* **2001**, *5 Suppl A*, 157-161, doi:10.1053/ejpn.2000.0454.
10. De Volder, A.G.; Cirelli, S.; de Barse, T.; Brucher, J.M.; Bol, A.; Michel, C.; Goffinet, A.M. Neuronal ceroid-lipofuscinosis: preferential metabolic alterations in thalamus and posterior association cortex demonstrated by PET. *J Neurol Neurosurg Psychiatry* **1990**, *53*, 1063-1067, doi:10.1136/jnnp.53.12.1063.
11. Rider, J.A.; Rider, D.L. Thirty years of Batten disease research: present status and future goals. *Mol Genet Metab* **1999**, *66*, 231-233, doi:10.1006/mgme.1999.2827.
12. Uvebrant, P.; Hagberg, B. Neuronal ceroid lipofuscinoses in Scandinavia. Epidemiology and clinical pictures. *Neuropediatrics* **1997**, *28*, 6-8, doi:10.1055/s-2007-973654.
13. Moore, S.J.; Buckley, D.J.; MacMillan, A.; Marshall, H.D.; Steele, L.; Ray, P.N.; Nawaz, Z.; Baskin, B.; Frecker, M.; Carr, S.M.; et al. The clinical and genetic epidemiology of neuronal ceroid lipofuscinosis in Newfoundland. *Clin Genet* **2008**, *74*, 213-222, doi:10.1111/j.1399-0004.2008.01054.x.
14. Kamate, M.; Prashanth, G.P.; Hattiholi, V. Clinico-investigative profile of infantile and late-infantile neuronal ceroid lipofuscinoses. *Neurol India* **2012**, *60*, 316-320, doi:10.4103/0028-3886.98524.
15. Sinha, S.; Satishchandra, P.; Santosh, V.; Gayatri, N.; Shankar, S.K. Neuronal ceroid lipofuscinosis: a clinicopathological study. *Seizure* **2004**, *13*, 235-240, doi:10.1016/S1059-1311(03)00163-8.

16. Nadkarni, S.; Deshpande, D.H.; Mondkar, V.P.; Bharucha, E.P. Neuronal ceroid lipofuscinosis: clinical and histochemical observations in 2 cases. *J Neurol Sci* **1979**, *43*, 395-404, doi:10.1016/0022-510x(79)90018-2.
17. Gulati, S.; Maheshwari, R.; Kabra, M.; Verma, I.C.; Kalra, V. Juvenile neuronal ceroid lipofuscinosis. *Indian J Pediatr* **2000**, *67*, 689-691, doi:10.1007/BF02762187.
18. Verma, R.; Raut, T.P.; Tiwari, N.; Malhotra, K.P.; Hussain, N.; Malhotra, H.S. Late infantile neuronal ceroid lipofuscinosis: A case report with review of literature. *Ann Indian Acad Neurol* **2013**, *16*, 282-285, doi:10.4103/0972-2327.112500.
19. Angural, A.; Spolia, A.; Mahajan, A.; Verma, V.; Sharma, A.; Kumar, P.; Dhar, M.K.; Pandita, K.K.; Rai, E.; Sharma, S. Review: Understanding Rare Genetic Diseases in Low Resource Regions Like Jammu and Kashmir - India. *Front Genet* **2020**, *11*, 415, doi:10.3389/fgene.2020.00415.
20. Markham, A. Cerliponase Alfa: First Global Approval. *Drugs* **2017**, *77*, 1247-1249, doi:10.1007/s40265-017-0771-8.
21. Sondhi, D.; Kaminsky, S.M.; Hackett, N.R.; Pagovich, O.E.; Rosenberg, J.B.; De, B.P.; Chen, A.; Van de Graaf, B.; Mezey, J.G.; Mammen, G.W.; et al. Slowing late infantile Batten disease by direct brain parenchymal administration of a rh.10 adeno-associated virus expressing CLN2. *Sci Transl Med* **2020**, *12*, doi:10.1126/scitranslmed.abb5413.
22. Williams, R.E.; Adams, H.R.; Blohm, M.; Cohen-Pfeffer, J.L.; de Los Reyes, E.; Denecke, J.; Drago, K.; Fairhurst, C.; Frazier, M.; Guelbert, N.; et al. Management Strategies for CLN2 Disease. *Pediatr Neurol* **2017**, *69*, 102-112, doi:10.1016/j.pediatrneurol.2017.01.034.
23. Mazurkiewicz-Beldzinska, M.; Del Toro, M.; Haliloglu, G.; Huidekoper, H.H.; Kravljanc, R.; Muhlhausen, C.; Andersen, B.N.; Prpic, I.; Striano, P.; Auvin, S. Managing CLN2 disease: a treatable neurodegenerative condition among other treatable early childhood epilepsies. *Expert Rev Neurother* **2021**, 1-8, doi:10.1080/14737175.2021.1885374.
24. Rai, E.; Mahajan, A.; Kumar, P.; Angural, A.; Dhar, M.K.; Razdan, S.; Thangaraj, K.; Wise, C.A.; Ikegawa, S.; Pandita, K.K.; et al. Whole Exome Screening Identifies Novel and Recurrent WISP3 Mutations Causing Progressive Pseudorheumatoid Dysplasia in Jammu and Kashmir-India. *Sci Rep* **2016**, *6*, 27684, doi:10.1038/srep27684.
25. Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T.J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Soding, J.; et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **2011**, *7*, 539, doi:10.1038/msb.2011.75.
26. Waterhouse, A.M.; Procter, J.B.; Martin, D.M.; Clamp, M.; Barton, G.J. Jalview Version 2-a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **2009**, *25*, 1189-1191, doi:10.1093/bioinformatics/btp033.
27. Adzhubei, I.; Jordan, D.M.; Sunyaev, S.R. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet* **2013**, Chapter 7, Unit7 20, doi:10.1002/0471142905.hg0720s76.
28. Kumar, P.; Henikoff, S.; Ng, P.C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* **2009**, *4*, 1073-1081, doi:10.1038/nprot.2009.86.
29. Schwarz, J.M.; Cooper, D.N.; Schuelke, M.; Seelow, D. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Methods* **2014**, *11*, 361-362, doi:10.1038/nmeth.2890.
30. McLaren, W.; Gil, L.; Hunt, S.E.; Riat, H.S.; Ritchie, G.R.; Thormann, A.; Flicek, P.; Cunningham, F. The Ensembl Variant Effect Predictor. *Genome Biol* **2016**, *17*, 122, doi:10.1186/s13059-016-0974-4.

31. Gonnelli, G.; Rooman, M.; Dehouck, Y. Structure-based mutant stability predictions on proteins of unknown structure. *J Biotechnol* **2012**, *161*, 287-293, doi:10.1016/j.jbiotec.2012.06.020.
32. Sheth, J.; Mistri, M.; Bhavsar, R.; Pancholi, D.; Kamate, M.; Gupta, N.; Kabra, M.; Mehta, S.; Nampoothiri, S.; Thakker, A.; et al. Batten disease: biochemical and molecular characterization revealing novel PPT1 and TPP1 gene mutations in Indian patients. *BMC Neurol* **2018**, *18*, 203, doi:10.1186/s12883-018-1206-1.
33. Kohan, R.; Carabelos, M.N.; Xin, W.; Sims, K.; Guelbert, N.; Cismondi, I.A.; Pons, P.; Alonso, G.I.; Troncoso, M.; Witting, S.; et al. Neuronal ceroid lipofuscinosis type CLN2: a new rationale for the construction of phenotypic subgroups based on a survey of 25 cases in South America. *Gene* **2013**, *516*, 114-121, doi:10.1016/j.gene.2012.12.058.
34. Kousi, M.; Lehesjoki, A.E.; Mole, S.E. Update of the mutation spectrum and clinical correlations of over 360 mutations in eight genes that underlie the neuronal ceroid lipofuscinoses. *Hum Mutat* **2012**, *33*, 42-63, doi:10.1002/humu.21624.
35. Golabek, A.A.; Kida, E.; Walus, M.; Wujek, P.; Mehta, P.; Wisniewski, K.E. Biosynthesis, glycosylation, and enzymatic processing in vivo of human tripeptidyl-peptidase I. *J Biol Chem* **2003**, *278*, 7135-7145, doi:10.1074/jbc.M211872200.
36. Sleat, D.E.; Donnelly, R.J.; Lackland, H.; Liu, C.G.; Sohar, I.; Pullarkat, R.K.; Lobel, P. Association of mutations in a lysosomal protein with classical late-infantile neuronal ceroid lipofuscinosis. *Science* **1997**, *277*, 1802-1805, doi:10.1126/science.277.5333.1802.
37. Santavuori, P. Neuronal ceroid-lipofuscinoses in childhood. *Brain Dev* **1988**, *10*, 80-83, doi:10.1016/s0387-7604(88)80075-5.
38. Sharp, J.D.; Wheeler, R.B.; Lake, B.D.; Savukoski, M.; Jarvela, I.E.; Peltonen, L.; Gardiner, R.M.; Williams, R.E. Loci for classical and a variant late infantile neuronal ceroid lipofuscinosis map to chromosomes 11p15 and 15q21-23. *Hum Mol Genet* **1997**, *6*, 591-595, doi:10.1093/hmg/6.4.591.
39. Haltia, M. The neuronal ceroid-lipofuscinoses. *J Neuropathol Exp Neurol* **2003**, *62*, 1-13, doi:10.1093/jnen/62.1.1.
40. Palmer, D.N.; Fearnley, I.M.; Walker, J.E.; Hall, N.A.; Lake, B.D.; Wolfe, L.S.; Haltia, M.; Martinus, R.D.; Jolly, R.D. Mitochondrial ATP synthase subunit c storage in the ceroid-lipofuscinoses (Batten disease). *Am J Med Genet* **1992**, *42*, 561-567, doi:10.1002/ajmg.1320420428.
41. Koike, M.; Shibata, M.; Ohsawa, Y.; Kametaka, S.; Waguri, S.; Kominami, E.; Uchiyama, Y. The expression of tripeptidyl peptidase I in various tissues of rats and mice. *Arch Histol Cytol* **2002**, *65*, 219-232, doi:10.1679/aohc.65.219.
42. Tyynela, J.; Palmer, D.N.; Baumann, M.; Haltia, M. Storage of saposins A and D in infantile neuronal ceroid-lipofuscinosis. *FEBS Lett* **1993**, *330*, 8-12, doi:10.1016/0014-5793(93)80908-d.
43. Elleder, M.; Dvorakova, L.; Stolnaja, L.; Vlaskova, H.; Hulkova, H.; Druga, R.; Poupetova, H.; Kostalova, E.; Mikulastik, J. Atypical CLN2 with later onset and prolonged course: a neuropathologic study showing different sensitivity of neuronal subpopulations to TPP1 deficiency. *Acta Neuropathol* **2008**, *116*, 119-124, doi:10.1007/s00401-008-0349-3.
44. Wisniewski, K.E.; Kaczmarek, A.; Kida, E.; Connell, F.; Kaczmarek, W.; Michalewski, M.P.; Moroziewicz, D.N.; Zhong, N. Reevaluation of neuronal ceroid lipofuscinoses: atypical juvenile onset may be the result of CLN2 mutations. *Mol Genet Metab* **1999**, *66*, 248-252, doi:10.1006/mgme.1999.2814.
45. Geraets, R.D.; Koh, S.; Hastings, M.L.; Kielian, T.; Pearce, D.A.; Weimer, J.M. Moving towards effective therapeutic strategies for Neuronal Ceroid Lipofuscinosis. *Orphanet J Rare Dis* **2016**, *11*, 40, doi:10.1186/s13023-016-0414-2.

46. Specchio, N.; Bellusci, M.; Pietrafusa, N.; Trivisano, M.; de Palma, L.; Vigeveno, F. Photosensitivity is an early marker of neuronal ceroid lipofuscinosis type 2 disease. *Epilepsia* **2017**, *58*, 1380-1388, doi:10.1111/epi.13820.
47. Albert, D.V.; Yin, H.; De Los Reyes, E.C.; Vidaurre, J. Unique Characteristics of the Photoparoxysmal Response in Patients With Neuronal Ceroid Lipofuscinosis Type 2: Can EEG Be a Biomarker? *J Child Neurol* **2016**, *31*, 1475-1482, doi:10.1177/0883073816658659.
48. Fietz, M.; AlSayed, M.; Burke, D.; Cohen-Pfeffer, J.; Cooper, J.D.; Dvorakova, L.; Giugliani, R.; Izzo, E.; Jahnova, H.; Lukacs, Z.; et al. Diagnosis of neuronal ceroid lipofuscinosis type 2 (CLN2 disease): Expert recommendations for early detection and laboratory diagnosis. *Mol Genet Metab* **2016**, *119*, 160-167, doi:10.1016/j.ymgme.2016.07.011.
49. Sleat, D.E.; Gin, R.M.; Sohar, I.; Wisniewski, K.; Sklower-Brooks, S.; Pullarkat, R.K.; Palmer, D.N.; Lerner, T.J.; Boustany, R.M.; Uldall, P.; et al. Mutational analysis of the defective protease in classic late-infantile neuronal ceroid lipofuscinosis, a neurodegenerative lysosomal storage disorder. *Am J Hum Genet* **1999**, *64*, 1511-1523, doi:10.1086/302427.
50. Steinfeld, R.; Steinke, H.B.; Isbrandt, D.; Kohlschutter, A.; Gartner, J. Mutations in classical late infantile neuronal ceroid lipofuscinosis disrupt transport of tripeptidyl-peptidase I to lysosomes. *Hum Mol Genet* **2004**, *13*, 2483-2491, doi:10.1093/hmg/ddh264.
51. Walus, M.; Kida, E.; Golabek, A.A. Functional consequences and rescue potential of pathogenic missense mutations in tripeptidyl peptidase I. *Hum Mutat* **2010**, *31*, 710-721, doi:10.1002/humu.21251.
52. Zhong, N.; Wisniewski, K.E.; Hartikainen, J.; Ju, W.; Moroziewicz, D.N.; McLendon, L.; Sklower Brooks, S.S.; Brown, W.T. Two common mutations in the CLN2 gene underlie late infantile neuronal ceroid lipofuscinosis. *Clin Genet* **1998**, *54*, 234-238, doi:10.1111/j.1399-0004.1998.tb04291.x.
53. Gardner, E.; Bailey, M.; Schulz, A.; Aristorena, M.; Miller, N.; Mole, S.E. Mutation update: Review of TPP1 gene variants associated with neuronal ceroid lipofuscinosis CLN2 disease. *Hum Mutat* **2019**, *40*, 1924-1938, doi:10.1002/humu.23860.
54. Warriar, V.; Vieira, M.; Mole, S.E. Genetic basis and phenotypic correlations of the neuronal ceroid lipofuscinosis. *Biochim Biophys Acta* **2013**, *1832*, 1827-1830, doi:10.1016/j.bbadis.2013.03.017.
55. Saini, A.G.; Sankhyani, N.; Singhi, P. Chorea in Late-Infantile Neuronal Ceroid Lipofuscinosis: An Atypical Presentation. *Pediatr Neurol* **2016**, *60*, 75-78, doi:10.1016/j.pediatrneurol.2016.02.015.
56. Ohba, C.; Osaka, H.; Iai, M.; Yamashita, S.; Suzuki, Y.; Aida, N.; Shimozaawa, N.; Takamura, A.; Doi, H.; Tomita-Katsumoto, A.; et al. Diagnostic utility of whole exome sequencing in patients showing cerebellar and/or vermis atrophy in childhood. *Neurogenetics* **2013**, *14*, 225-232, doi:10.1007/s10048-013-0375-8.
57. Savojardo, C.; Manfredi, M.; Martelli, P.L.; Casadio, R. Solvent Accessibility of Residues Undergoing Pathogenic Variations in Humans: From Protein Structures to Protein Sequences. *Front Mol Biosci* **2020**, *7*, 626363, doi:10.3389/fmolb.2020.626363.
58. Lin, L.; Sohar, I.; Lackland, H.; Lobel, P. The human CLN2 protein/tripeptidyl-peptidase I is a serine protease that autoactivates at acidic pH. *J Biol Chem* **2001**, *276*, 2249-2255, doi:10.1074/jbc.M008562200.
59. Nakatsukasa, K.; Brodsky, J.L. The recognition and retrotranslocation of misfolded proteins from the endoplasmic reticulum. *Traffic* **2008**, *9*, 861-870, doi:10.1111/j.1600-0854.2008.00729.x.
60. Brooks, D.A. Protein processing: a role in the pathophysiology of genetic disease. *FEBS Lett* **1997**, *409*, 115-120, doi:10.1016/s0014-5793(97)00423-7.

61. Kuizon, S.; DiMaiuta, K.; Walus, M.; Jenkins, E.C., Jr.; Kuizon, M.; Kida, E.; Golabek, A.A.; Espinoza, D.O.; Pullarkat, R.K.; Junaid, M.A. A critical tryptophan and Ca²⁺ in activation and catalysis of TPPI, the enzyme deficient in classic late-infantile neuronal ceroid lipofuscinosis. *PLoS One* **2010**, *5*, e11929, doi:10.1371/journal.pone.0011929.
62. Rawlings, N.D.; Barrett, A.J. Tripeptidyl-peptidase I is apparently the CLN2 protein absent in classical late-infantile neuronal ceroid lipofuscinosis. *Biochim Biophys Acta* **1999**, *1429*, 496-500, doi:10.1016/s0167-4838(98)00238-6.
63. Kida, E.; Golabek, A.A.; Walus, M.; Wujek, P.; Kaczmarek, W.; Wisniewski, K.E. Distribution of tripeptidyl peptidase I in human tissues under normal and pathological conditions. *J Neuropathol Exp Neurol* **2001**, *60*, 280-292, doi:10.1093/jnen/60.3.280.
64. Kurachi, Y.; Oka, A.; Itoh, M.; Mizuguchi, M.; Hayashi, M.; Takashima, S. Distribution and development of CLN2 protein, the late-infantile neuronal ceroid lipofuscinosis gene product. *Acta Neuropathol* **2001**, *102*, 20-26, doi:10.1007/s004010000321.
65. Vines, D.J.; Warburton, M.J. Classical late infantile neuronal ceroid lipofuscinosis fibroblasts are deficient in lysosomal tripeptidyl peptidase I 1. *FEBS Letters* **1999**, *443*, 131-135, doi:[https://doi.org/10.1016/S0014-5793\(98\)01683-4](https://doi.org/10.1016/S0014-5793(98)01683-4).
66. Guhaniyogi, J.; Sohar, I.; Das, K.; Stock, A.M.; Lobel, P. Crystal structure and autoactivation pathway of the precursor form of human tripeptidyl-peptidase 1, the enzyme deficient in late infantile ceroid lipofuscinosis. *J Biol Chem* **2009**, *284*, 3985-3997, doi:10.1074/jbc.M806943200.
67. Lin, L.; Lobel, P. Expression and analysis of CLN2 variants in CHO cells: Q100R represents a polymorphism, and G389E and R447H represent loss-of-function mutations. *Hum Mutat* **2001**, *18*, 165, doi:10.1002/humu.1170.
68. Ezaki, J.; Tanida, I.; Kanehagi, N.; Kominami, E. A lysosomal proteinase, the late infantile neuronal ceroid lipofuscinosis gene (CLN2) product, is essential for degradation of a hydrophobic protein, the subunit c of ATP synthase. *J Neurochem* **1999**, *72*, 2573-2582, doi:10.1046/j.1471-4159.1999.0722573.x.
69. Pal, A.; Kraetzner, R.; Gruene, T.; Grapp, M.; Schreiber, K.; Gronborg, M.; Urlaub, H.; Becker, S.; Asif, A.R.; Gartner, J.; et al. Structure of tripeptidyl-peptidase I provides insight into the molecular basis of late infantile neuronal ceroid lipofuscinosis. *J Biol Chem* **2009**, *284*, 3976-3984, doi:10.1074/jbc.M806947200.
70. Ezaki, J.; Takeda-Ezaki, M.; Kominami, E. Tripeptidyl peptidase I, the late infantile neuronal ceroid lipofuscinosis gene product, initiates the lysosomal degradation of subunit c of ATP synthase. *J Biochem* **2000**, *128*, 509-516, doi:10.1093/oxfordjournals.jbchem.a022781.
71. Junaid, M.A.; Wu, G.; Pullarkat, R.K. Purification and characterization of bovine brain lysosomal pepstatin-insensitive proteinase, the gene product deficient in the human late-infantile neuronal ceroid lipofuscinosis. *J Neurochem* **2000**, *74*, 287-294, doi:10.1046/j.1471-4159.2000.0740287.x.
72. Golabek, A.A.; Kida, E. Tripeptidyl-peptidase I in health and disease. *Biol Chem* **2006**, *387*, 1091-1099, doi:10.1515/BC.2006.135.
73. Mole, S.E.; Schulz, A.; Badoe, E.; Berkovic, S.F.; de Los Reyes, E.C.; Dulz, S.; Gissen, P.; Guelbert, N.; Lourenco, C.M.; Mason, H.L.; et al. Guidelines on the diagnosis, clinical assessments, treatment and management for CLN2 disease patients. *Orphanet J Rare Dis* **2021**, *16*, 185, doi:10.1186/s13023-021-01813-5.
74. Nita, D.A.; Mole, S.E.; Minassian, B.A. Neuronal ceroid lipofuscinoses. *Epileptic Disord* **2016**, *18*, 73-88, doi:10.1684/epd.2016.0844.