SUPPLEMENTARY INFORMATION

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

Arshia Angural[1], Kalaiarasan Ponnusamy[2], Diksha Langeh[1,#], Mamta Kumari[1,#], Akshi Spolia[1,#], Ekta Rai[1], Ankush Sharma[3], Kamal Kishore Pandita[4,*], Swarkar Sharma[1,*]

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

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

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

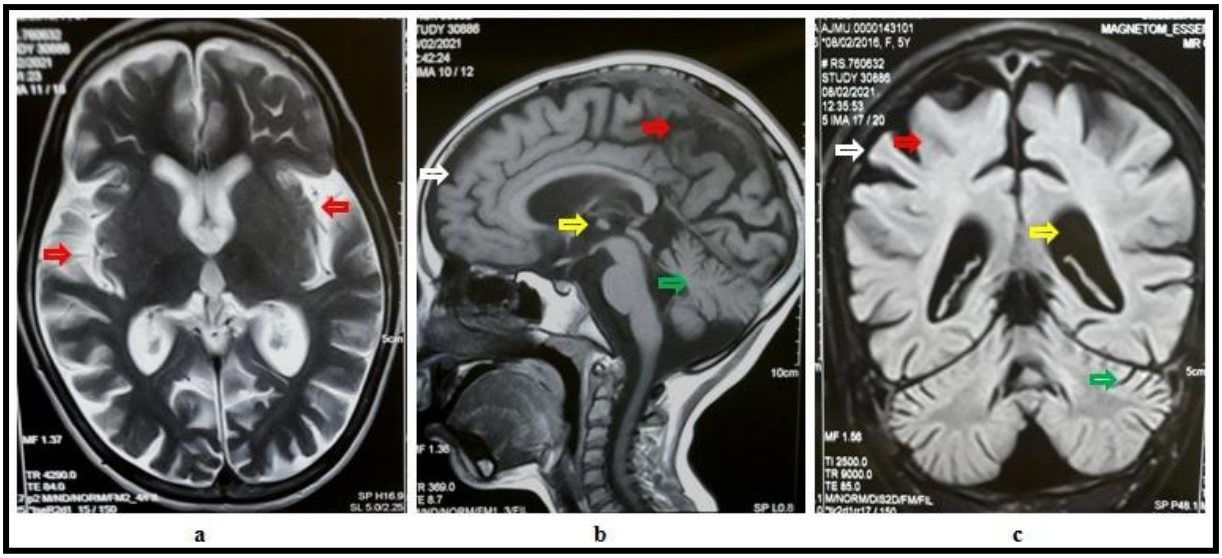
[4]Independent Researcher, Health Clinic, Swam Vihar, Muthi, Jammu, Jammu and Kashmir, India

[#]Equal contribution

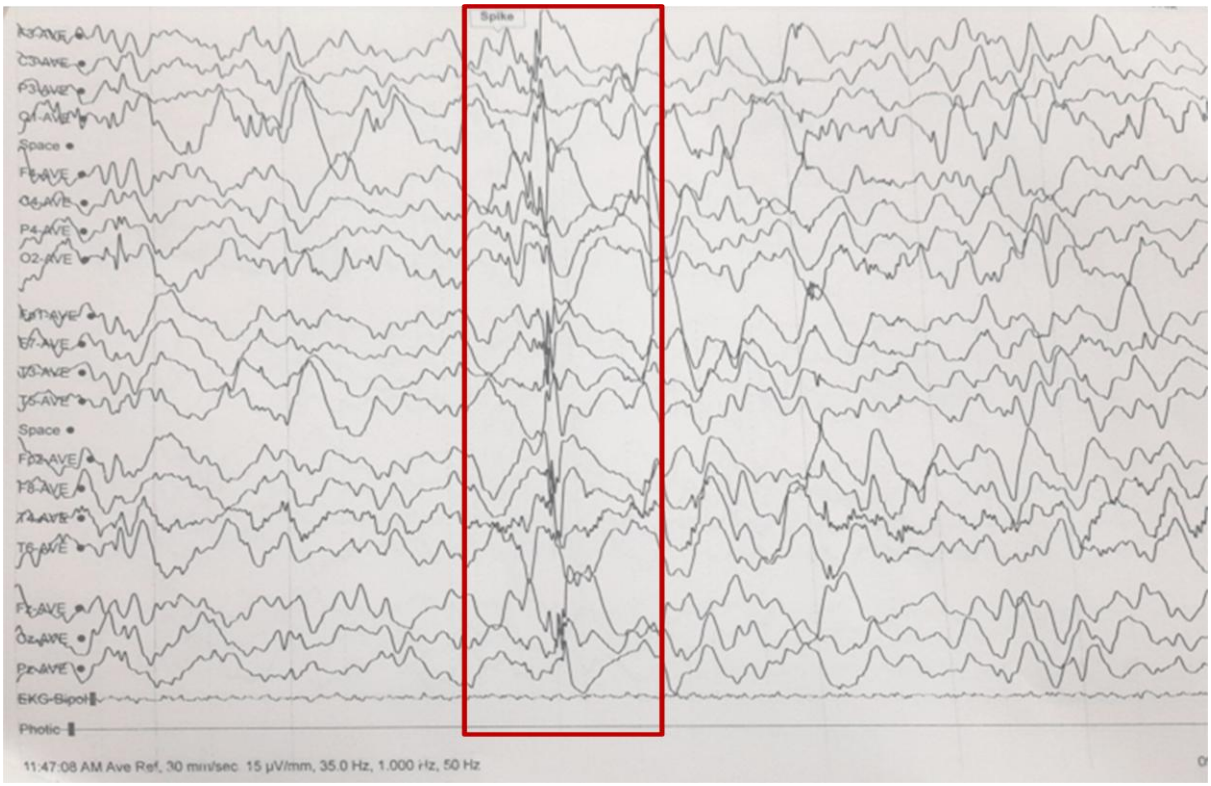
*Corresponding authors:
Swarkar Sharma, email: swarkar.sharma@smvdu.ac.in, and
Kamal Kishore Pandita, email: panditakk69@gmail.com

ABSTRACT

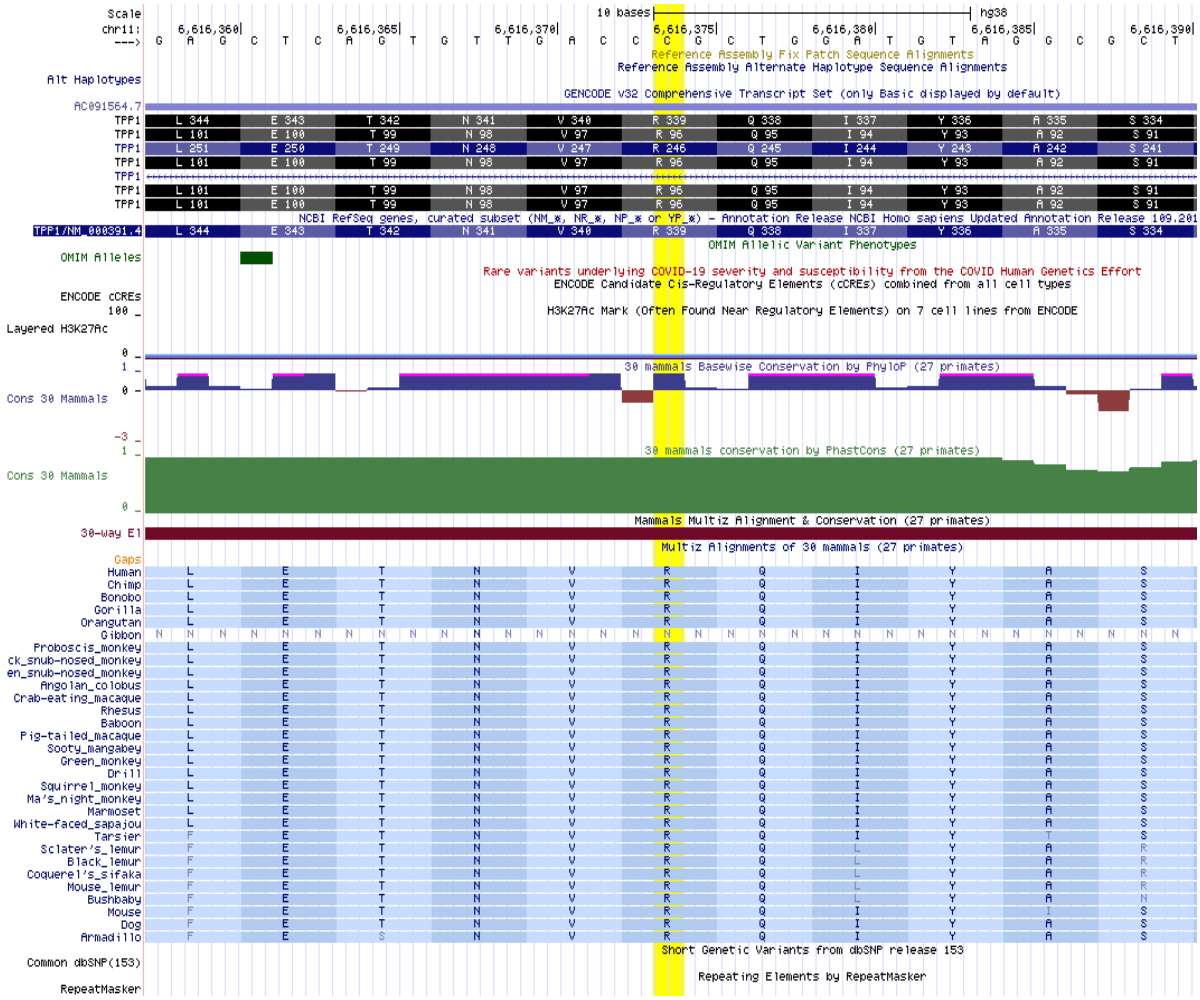
We report diagnosis of Neuronal Ceroid Lipofuscinosis Type 2 (CLN2), a rare, hereditary neurodegenerative disease of childhood, in a four and a half year old girl, the first child of non-consanguineous parents with no family history. Despite extensive efforts by the parents, her clinical condition remained 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.



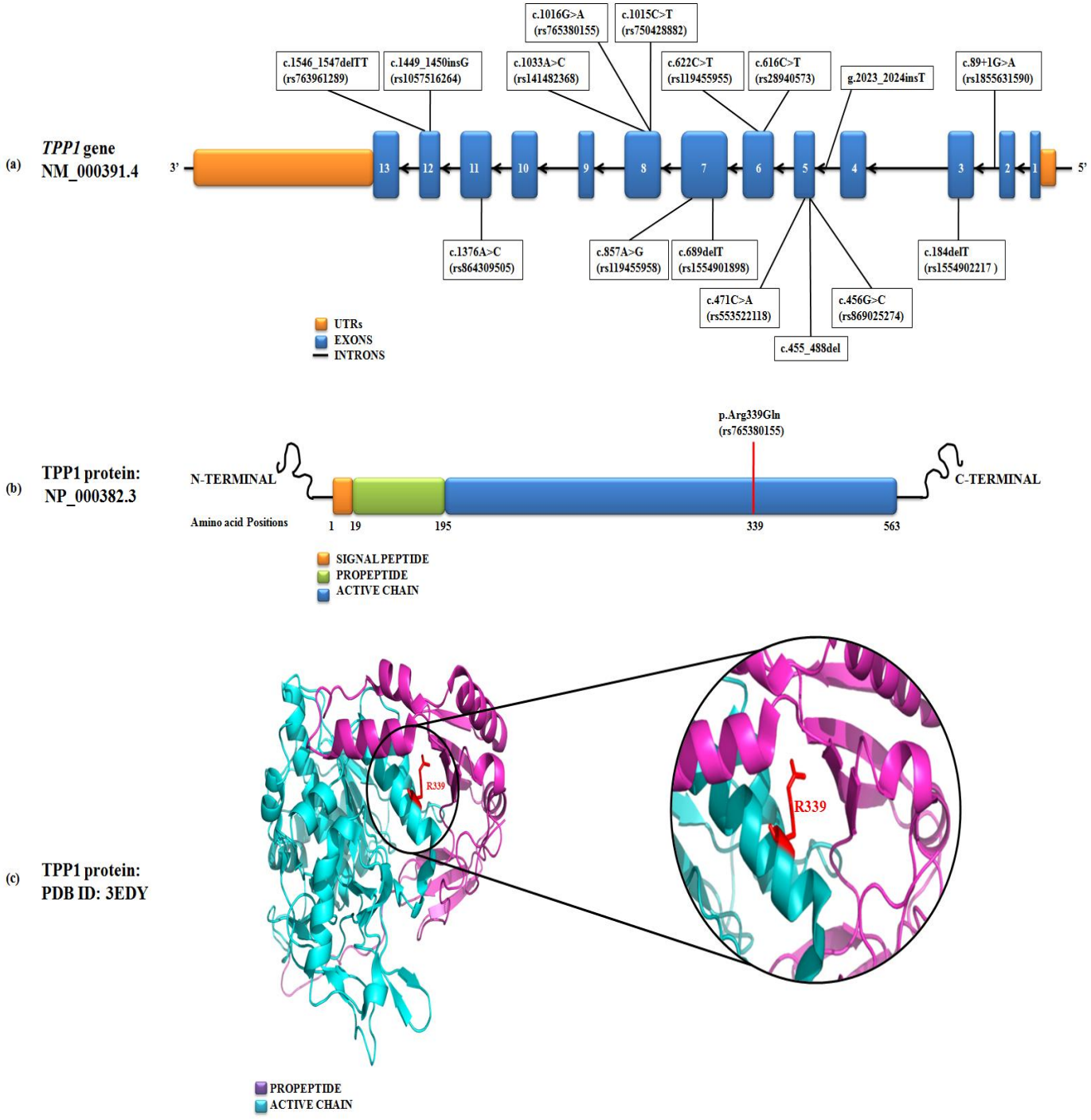
Supplementary Figure 1: (a) T2-weighted axial, (b) T2-weighted sagittal, and (c) FLAIR coronal sections of 4.5 year old proband's MRI revealing signs of generalized cerebral atrophy (red arrows), cerebellar atrophy (green arrows), and prominent ventricular system (yellow arrows) and CSF spaces (white arrows).



Supplementary Figure 2: A montage screenshot of the current EEG investigation of the 4.5 year old proband showing slow background activity with generalized epileptiform multifocal discharges and absence of sleep spindles during sleep record. Red rectangular bar on the montage is indicating epileptiform multifocal spikes.



Supplementary Figure 3: A UCSC screenshot indicating conserved nature of wild-type allele (highlighted in yellow) of NC_00011.10:g.6616374C>T variation across 100 vertebrate species.



Supplementary Figure 4: (a) An illustration depicting the genomic locations of *TPP1* gene variations reported previously in the Indian LINCL patients, (b) An illustration depicting the genomic location of the *TPP1* p.R339Q variation reported in the current study, and (c) A PyMol visualization of human, wild-type TPP1 protein (PDB ID: 3EDY) indicating the location of R339 residue (see the outset image).

GENES TARGETED FOR NGS

The following genes associated with clinical information were screened:

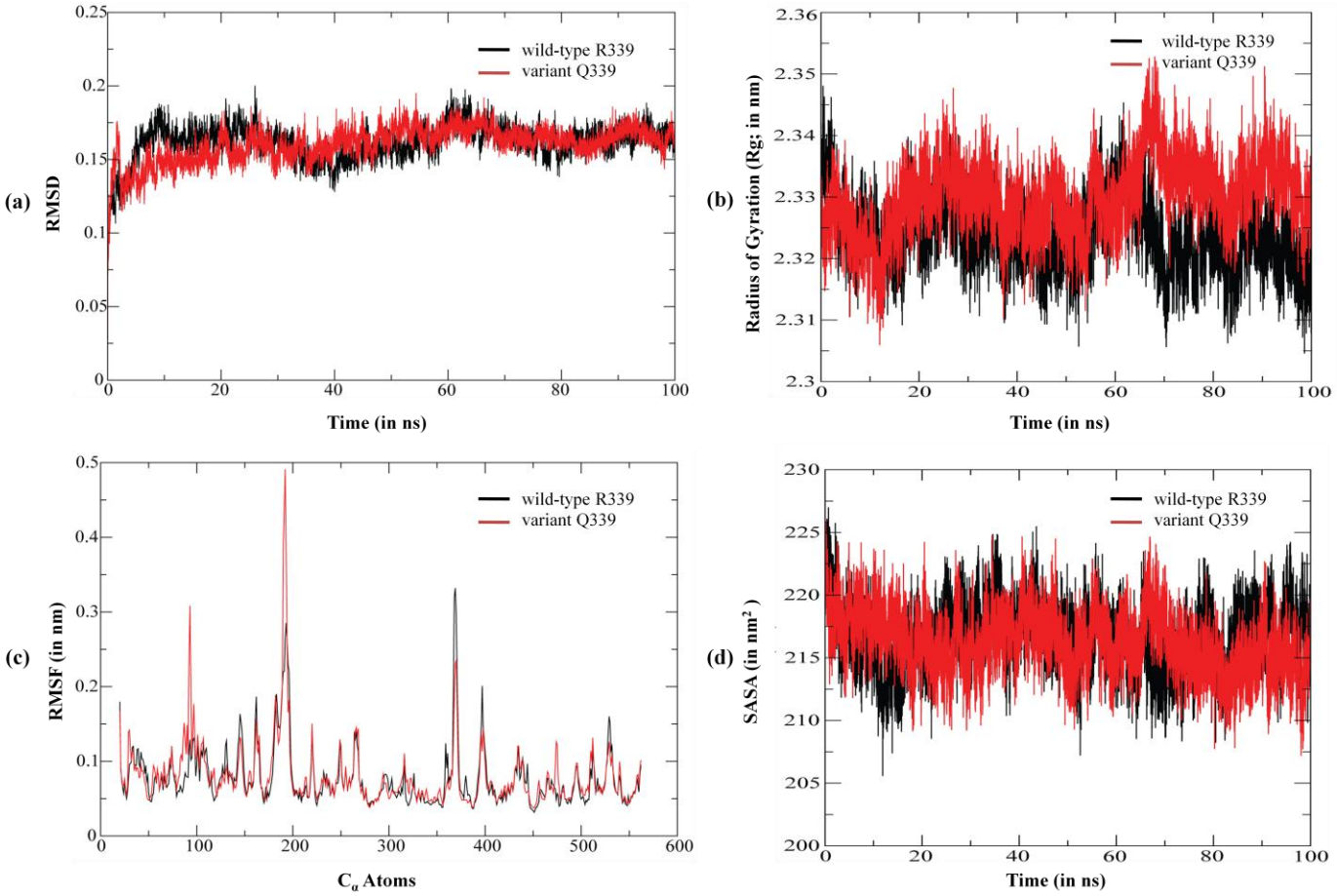
CLN3, CLN8, CLN5, CLN6, TPP1, CTSD, KCTD7, MFSD8, PPT1, DNAJC5, GRN, ADSL, ALDH5A1, ATP1A3, ATRX, CACNA1A, CASK, CDKL5, CHD2, CHRNA2, CHRNA4, CHRNA7, CHRNB2, CLCN4, CNTNAP2, CSTB, DDX3X, DEPDC5, DYRK1A, EEF1A2, EHMT1, EPM2A, FOLR1, FOXG1, GABRA1, GABRB2, GABRB3, GABRG2, GAMT, GATM, GNAO1, GOSR2, GRIN1, GRIN2A, IQSEC2, KANSL1, KCNA2, KCNC1, KCNT1, KCNMA1, KDM6A, KIAA2022, LGI1, MAGI2, MBD5, MECP2, MEF2C, NALCN, NGLY1, NHLRC1, NPRL3, NRXN1, PACS1, PCDH19, PIGN, PNKP, POLG, PPP2R5D, PURA, SCN2A, SCN1B, SCN1A, SLC19A3, SLC2A1, SLC6A1, SLC9A6, SLC6A8, SMC1A, SPATA5, STX1B, SYNGAP1, TBC1D24, TCF4, UBE3A, WDR45, ZEB2, AGRN, SCNN1D, TAS1R3, MIB2, GNB1, GABRD, PEX10, PLCH2, MMEL1, TP73, CLSTN1, NMNAT1, KIF1B, CORT, CASZ1, TARDBP, MTOR, MTHFR, CLCN6, NPPB, MFN2, TNFRSF1B, PRDM2, EFHD2, ATP13A2, UBR4, PINK1, KIF17, ECE1, C1QA, EPHB2, HNRNPR, ID3, STMN1, LIN28A, NR0B2, TRNP1, SLC9A1, OPRD1, SDC3, HDAC1, HMGB4, KIAA0319L, ADPRHL2, EPHA10, MACF1, MFSD2A, COL9A2, RIMS3, ST3GAL3, CCDC23, PTCH2, SLC6A9, TOE1, PRDX1, MAST2, POMGNT1, TAL1, STIL, ELAVL4, FAF1, RNF11, NRD1, SCP2, CPT2, LRP8, PCSK9, DAB1, JUN, DOCK7, DNAJC6, LEPR, SGIP1, PTGER3, NEGR1, LHX8, FAM73A, DDAH1, KIAA1107, PTBP2, DBT, GPSM2, WDR47, TAF13, SORT1, AMIGO1, AMPD2, ALX3, SLC6A17, PTPN22, NGF, NHLH2, GDAP2, FAM72B, SRGAP2C, FAM72C, NBPF23, NBPF15, SRGAP2B, FAM72D, PEX11B, HYDIN2, SV2A, LINGO4, S100A5, CHTOP, HAX1, KCNN3, EFNA3, EFNA1, GBA, HCN3, RIT1, ARHGEF2, LMNA, SEMA4A, MIR9-1, MEF2D, BCAN, NTRK1, TAGLN2, KCNJ9, PEX19, MPZ, FCGR2A, FCGR2B, ATF6, NOS1AP, RGS4, LMX1A, CD247, MYOC, MIR3120, ASTN1, TOR1AIP1, CACNA1E, GLUL, TRMT1L, PTGS2, PLA2G4A, KCNT2, ASPM, KIF14, NAV1, GPR37L1, SYT2, CHI3L1, KISS1, LRRN2, NFASC, CNTN2, KLHDC8A, FAM72A, SRGAP2, CAMK1G, HSD11B1, KCNH1, SLC30A1, DTL, NENF, VASH2, SLC30A10, MARK1, PYCR2, PARP1, PSEN2, ARF1, ARV1, DISC1, DISC2, FMN2, KMO, AKT3, ZBTB18, COX20, TMEM18, SNTG2, MYT1L, KIDINS220, MYCN, VSNL1, GDF7, ATAD2B

PROTOCOL FOR MOLECULAR DYNAMICS SIMULATIONS

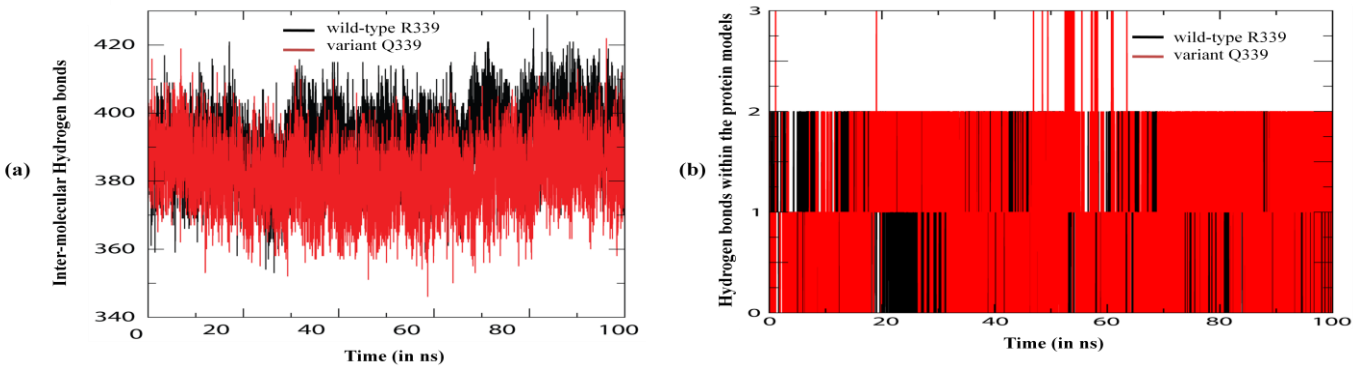
In order to understand the functional relevance of the missense *TPP1* p.R339Q variation, *in silico* molecular dynamics simulations (MDS) for the wild-type R339 (PDB ID:3EDY) as well as the variant Q339 human TPP1 (hTPP1) protein models were performed for 100 ns time-lag using GROMACS 5.1 package using OPLS force field [1]. The protein structures were immersed in to cubic box with periodic boundary conditions, dimensions ensuring that any atom of the protein was at least 10Å away from the wall of the box. The box is filled with small point charge (SPC) water molecule to solvate the protein and five Cl counter ions were added to neutralize the total charge of the system. The energy minimization is carried out using steepest descent method until it converges with F_{max} no greater than 1000KJ mol^{-1} nm^{-1}. Equilibration of the R339 and Q339hTPP1 models is performed for 1000ps in NVT and NPT ensemble. During equilibration, the coupling constants were set to 0.1 ps and 2.0 ps for temperature and pressure, respectively. The V-rescale temperature coupling and Parrinello-Rehman pressure coupling was used to keep the system in a stable environment (300K, 1 bar). The coulomb and Van dar Waals interactions were truncated at 10 Å, whereas long range electrostatic interactions used Particle Mesh Ewald (PME) algorithm [2]. All the covalent bonds were constrained using Linear Constraint Solver (LINCS) algorithm [3]. The production run of wild as well as mutant human TPP1 protein models was carried out for 100 ns and all the trajectories were stored for every 2 ps for further analysis.

The structural analyses of molecular dynamics trajectories of both wild-type R339 and variant Q339 hTPP1 protein models were carried out using GROMACS in-built tools [4]. Both the protein models (R339 and Q339) were subjected to analyses including root mean square deviation (RMSD) of the backbone of protein models, radius of gyration (Rg) of protein models, and root mean square fluctuation (RMSF) of alpha-carbon atoms (C_α's) using "*gmx rms*", "*gmx gyrate*" and "*gmx rmsf*" commands, respectively. The index was created for the catalytic residues and the intermolecular hydrogen bond interactions were calculated for both wild and variant MD trajectories using "*gmx hbond*" command considering donor-acceptor distance (< 3.6 Å) and donor-hydrogen-acceptor angle (> 90° degree). To understand the solvent accessible surface area of both protein models, SASA was also performed using GROMACS "*gmx sasa*" command. The graphs for RMSD, Rg, RMSF, hydrogen-bond and SASA analyses were generated using GRaphing, Advanced Computation and Exploration (GRACE) tool. PyMol was used for *in silico* structural visualization of the human TPP1 (hTPP1) R339 and Q339 protein models.

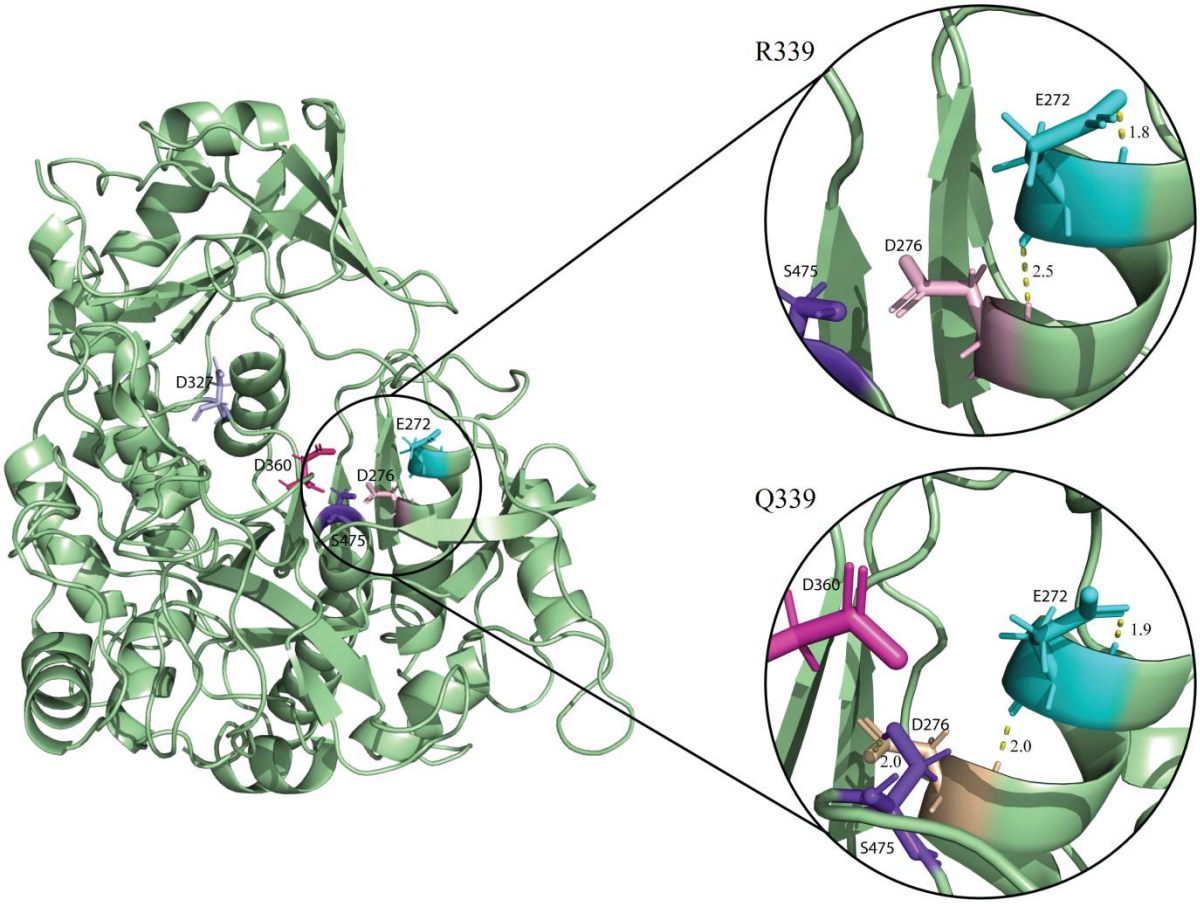
PLOTS FOR MOLECULAR DYNAMICS SIMULATIONS (MDS)



Supplementary Figure 5: MDS-plots for wild-type R339 (PDB ID: 3EDY) and variant Q339 protein models representing comparative: (a) RMSD of protein backbone, (b) Rgof protein, (c) RMSF of Cα's,and (d) Solvent Accessible Surface Area of protein models.



Supplementary Figure 6: MDS-plots for wild-type R339 (PDB ID: 3EDY) and variant Q339 protein models representing comparative number of: (a) inter-molecular Hydrogen bond contacts between protein models and solvent, and (b) Hydrogen bonds between the amino acid residues of protein models.



Supplementary Figure 7: An illustration for the hTPP1 protein model (PDB ID: 3EDY) depicting E^{272}, D^{276}, D^{327}, D^{360} and S^{475} residues in the catalytic site and hydrogen bond between these residues in R339 and Q339 hTPP1 protein models (outset).

REFERENCES

1. Abraham, M.J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J.C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1*, 19-25.
2. Essmann, U.; Perera, L.; Berkowitz, M.L.; Darden, T.; Lee, H.; Pedersen, L.G. A smooth particle mesh Ewald method. *J of Chem Phys* **1995**, *103*, 8577-8593, doi:https://doi.org/10.1063/1.470117.
3. Hess, B.; Bekker, H.; Berendsen, H.J.; Fraaije, J.G. LINCS: a linear constraint solver for molecular simulations. *J Comput Chem* **1997**, *18*, 1463-1472.
4. Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A.E.; Berendsen, H.J. GROMACS: fast, flexible, and free. *J Comput Chem* **2005**, *26*, 1701-1718, doi:10.1002/jcc.20291.