Chapter 5

Autosomal Recessive Spinocerebellar Ataxia 7 (SCAR7) is Caused by Variants in *TPP1*, The Gene Involved in Classic Late-Infantile Neuronal Ceroid Lipofuscinosis 2 Disease (CLN2 Disease)

Yu Sun*, Rowida Almomani*, Guido Breedveld, Gijs W.E. Santen, Emmelien Aten, Dirk J. Lefeber, Jorrit I. Hoff, Esther Brusse, Frans W. Verheijen, Rob M. Verdijk, Marjolein Kriek, Ben Oostra, Martijn H. Breuning, Monique Losekoot, Johan T. den Dunnen, Bart P. van de Warrenburg, and Anneke J.A. Maat-Kievit

***The authors contributed equally to the work**

Hum Mutat. 2013; 34:706-13.

Abstract

Spinocerebellar ataxias are phenotypically, neuropathologically and genetically heterogeneous. The locus of autosomal recessive spinocerebellar ataxia type 7 (SCAR7) was previously linked to chromosome band 11p15. We have identified *TPP1* as the causative gene for SCAR7 by exome sequencing. A missense and a splice site variant in *TPP1*, cosegregating with the disease, were found in a previously described SCAR7 family and also in another patient with a SCAR7 phenotype. *TPP1*, encoding the tripeptidyl peptidase 1 enzyme, is known as the causative gene for neuronal ceroid lipofuscinosis disease 2 (CLN2). CLN2 is characterized by epilepsy, loss of vision, ataxia and a rapidly progressive course, leading to early death. SCAR7 patients showed ataxia and low activity of tripeptidyl peptidase 1, but no ophthalmologic abnormalities or epilepsy. Also, the slowly progressive evolution of the disease until old age and absence of ultra structural curvilinear profiles is different from the known CLN2 phenotypes.

Our findings now expand the phenotypes related to *TPP1*-variants to SCAR7. In spite of the limited sample size and measurements a putative genotype-phenotype correlation may be drawn: we hypothesize that loss of function variants abolishing TPP1 enzyme activity lead to CLN2, while variants that diminish TPP1 enzyme activity lead to SCAR7.

Introduction

Spinocerebellar ataxias are phenotypically, neuropathologically and genetically heterogeneous, with over 50 genes and loci associated with genetic forms of spinocerebellar ataxias (Matilla-Duenas, 2012; Vermeer, et al., 2011). The inheritance of the disease can be either autosomal dominant, autosomal recessive, X-linked or mitochondrial. Due to genetic heterogeneity of the the hereditary ataxias, it is time and money consuming to check all known genes by Sanger sequencing (Sailer and Houlden, 2012). Recently developed genomic techniques, such as exome sequencing that targets only the coding portion of the genome, offer an alternative strategy to rapidly sequence all genes in a comprehensive manner and its utility has been demonstrated in more diagnostic settings (Sailer, et al., 2012).

Breedveld *et al.* previously reported a unique Dutch family with a childhood onset, slowly progressive autosomal recessive spinocerebellar ataxia, referred to as SCAR7 (OMIM 609207) and distinguished from other recessive ataxia types (Suppl. Table S1) by locus, onset and/or clinical findings. A genome-wide linkage study mapped the causative gene on a 5.9 cM region on chromosome band 11p15, which contains more than 200 genes. No obvious candidate gene could be assigned, as genes for ataxia mostly have different functions and features (Breedveld, et al., 2004).

Here we report the results of exome sequencing in the Dutch family revealing disease-causing variants in the *TPP1* gene (OMIM 607998), encoding the lysosomal enzyme tripeptidyl peptidase 1.

Homozygous or compound heterozygous variants in *TPP1* usually lead to neuronal ceroid lipofuscinosis 2 disease (CLN2; OMIM 204500) (Williams and Mole, 2012), a neurodegenerative disorder generally characterized by onset at 2-4 years of age with seizures, ataxia and a progressive cognitive and motor dysfunction, and visual impairment later in the course of the disease, followed by death at the end of the first decade or beginning of the second (Santavuori, 1988; Williams, et al., 1999). Our findings expand the phenotypes of *TPP1* mutations (Kousi, et al., 2012) to SCAR7.

Materials & Methods

Patients

The clinical data of the original Dutch sib ship (pedigree family A, Figure 1A) have been reported previously (Breedveld, et al., 2004). In summary, patients of family A, suffer from a childhood-onset spinocerebellar ataxia with pyramidal signs and posterior column involvement and a postural tremor, without other (non-) neurological features. Neuroimaging shows atrophy of cerebellum, vermis, pons, and medulla oblongata.

Patient B-II.1 (Figure 1B), a 51 year-old woman reported an onset of symptoms with diplopia at age 18. Two years later, subtle gait changes occurred. At the age of 28 years, she was diagnosed with cerebellar atrophy. Symptoms have been very slowly progressive since then; she still walks unsupported, although with occasional falls. She volunteered some loss of dexterity, mild speech and swallowing difficulties, and urinary urgency. Family history was negative and there was no consanguinity known in the parents. On examination, we observed normal cognitive functions; square-wave jerks, jerky pursuit, and hypermetric saccades; cerebellar dysarthria; mild proximal leg muscle weakness; no extrapyramidal features; very mild gait and appendicular ataxia; clear hyperreflexia with ankle jerk clonus; equivocal plantar responses; and normal sensory examination. MRI showed diffuse cerebellar atrophy (Figure 2). Full ophthalmologic evaluation was completely normal. Negative or normal outcomes were obtained for molecular genetic testing of various SCA genes (1, 2, 3, 6, 7, 12, 13, 14, and 17), APTX, SETX, FXN, SACS, SPG7, and ANO10, as well as measurements of creatine kinase level, alpha-fetoprotein, vitamins, and acanthocytes, and lysosomal enzymes. However, increased activity of plasma chitotriosidase as a marker for lysosomal disorders (280 nmol/h/ml, reference <160) and decreased TPP1 activity was noted. The phenotype, TPP1 enzyme activity and *TPP1* mutations of these SCAR7 and other Dutch CLN2 patients (C-O) and relatives are described in Table 1.

All patients in this study provided informed consent for DNA studies, and for diagnostic procedures.

Exome sequencing

Genomic DNA from patients and relatives from family A was extracted from peripheral blood using the salt precipitation method (PUREGENE, QIAGEN). Exome sequencing was performed on patient A.III-2 using the SureSelect 50Mb exome capture kit (Agilent) following the manufacturer's protocol. The captured fragments were subsequently sequenced by Illumina HiSeq as previously described (Santen, et al., 2012), paired end mode. Read length is 100bp.

The raw Fastq files were aligned by bwa-0.5.9 (Li and Durbin, 2009). SAM/BAM files were manipulated by Samtools-0.1.10 (Li, et al., 2009) and Picard-1.57. Variations were called by GATK (McKenna, et al., 2010). The output vcf file was annotated by uploading to SeattleSeq 134 (http://snp.gs.washington.edu/SeattleSeqAnnotation134/). The responsible gene for autosomal recessive ataxia was mapped in a 5.9 cM linkage interval (4.5 Mb), so only variants in this region were considered as candidates.

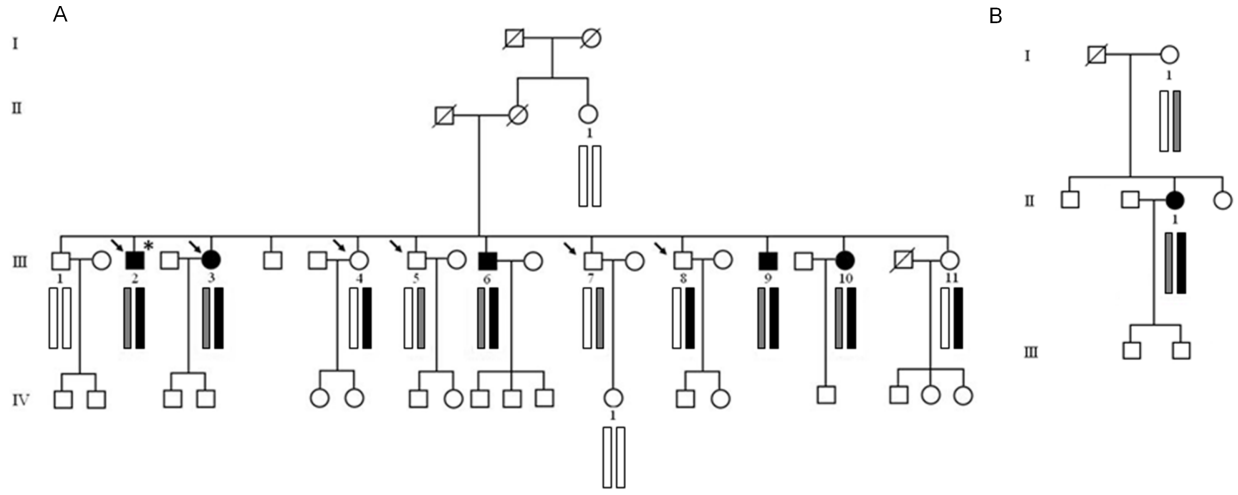


Figure 1. (A) Pedigree of family A and (B) family B with autosomal recessive spinocerebellar ataxia (SCAR7). Patients are marked with black symbols, unaffected relatives with open symbols. *TPP1* genotypes are shown below individuals (open bars indicate normal allele, black bars indicate alleles with c.509-1G>C, grey bars represent alleles with c.1397T>G). Genotype analysis shows co-segregation of variants with disease.

* =exome sequencing performed; → =RNA analysis performed.

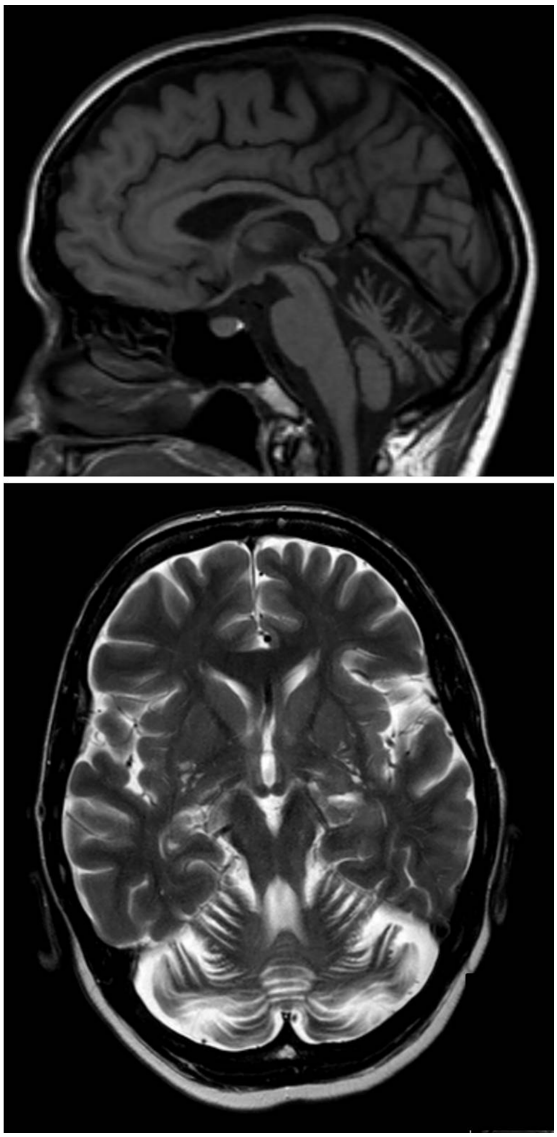


Figure 2. MRI of the brain of patient B. Sagittal T1-weighted (top) and transversal T2-weighted (bottom) show diffuse cerebellar atrophy.

Sanger sequencing

Sanger sequencing was applied to confirm the finding of exome sequencing and cosegregation of the variations in family A. PCR was performed by using Phire Hot Start II DNA polymerase (Finnzyme) following the official protocol. Primers used in PCR reactions are shown in Supp. Table S2. PCR products were first purified by QIAquick PCR purification kit (QIAGEN), subsequently mixed with 25pmol of either forward or reverse primers and sequenced.

For family B, the coding region and flanking intron sequences of the *TPP1* gene were examined by Sanger sequencing in a diagnostic setting, using standard procedures (The protocol and the primer sequences are available upon request).

RNA analysis

Leucocyte RNA from affected compound heterozygous patients (A.III-2, A.III-3), unaffected c.509-1G>C carriers (A.III-4, A.III-8), and unaffected c.1397T>G carriers (A.III-5, A.III-7) was isolated from blood using RNABEE following the official protocol. cDNA synthesis and RT-PCR was performed as previously described (Sun, et al., 2010). The primer sequences are listed in Supp. Table S2. The RT-PCR products were examined by 2% agarose gel, and followed by Sanger sequencing.

TPP1 enzyme activity assay

Enzyme activity of tripeptidyl peptidase 1 was assayed in leucocytes and fibroblasts of SCAR7 patients via the determination of fluorescent 7-amino-4 methylcoumarin, released from the substrate Ala-Ala-Phe 7-amido-4-methylcoumarin by incubation in cell homogenates as described previously (Van Diggelen, et al., 2001). Also TPP1 activity in leucoytes of SCAR7 carriers was measured. The TPP1 enzyme activity of B-II.1 was tested in Radboud University Nijmegen Medical Centre (normal range: 37-209 nmol/h/mg protein), while family A and other CLN2 samples (C-O) are analyzed in Erasmus Medical Center, with the normal range, 125-340 nmol/h/mg protein.

EM study

Fibroblasts from one of the patients of family A-III.2 and patient B-II.1 were fixed in glutaric-aldehyde, postfixated with osmium tetroxide and embedded in Epon (Hexion Specialty Chemicals, Inc, Danbury, Connecticut) and examined by electron microscopy.

Results

Exome sequencing reveals candidate variants

Exome sequencing was performed to target protein coding sequences in the human genome for potential disease-causing variants. An overview of the data obtained is listed in Supp. Table S3. Only variants inside the 5.9 cM linkage interval (from D11S4088 to D11S1331, genomic location: chr11.hg19:g.2,754,951-7,292,210) were analyzed. If the variant allele frequency in the NHLBI ESP exomes (http://evs.gs.washington.edu/EVS/) was larger than 5%, the variant was removed from the candidate list. We then selected stop-gain, stop-loss, missense, splice site, frameshift and in-frame coding indel variants and concordance with autosomal recessive inheritance (i.e. homozygous or compound heterozygous variants in one gene). Variants in three genes, *C11orf40*, *TPP1* and *DCHS1* (Table 2), fulfilled these criteria. However, *C11orf40* and *DCHS1,* located on the same allele, showed no co-segregation with the disease, they were excluded.. Thus, the only candidate gene left was *TPP1*, encoding the lysosomal serine protease with tripeptidyl-peptidase 1 activity. Two *TPP1* variants, a splice site variant, c.509-1G>C and a missense variant, c.1397T>G, p.(Val466Gly) (Figure 3A), co-segegrated with the disease (Figure

1A). Later, identical variants in the *TPP1* gene were found in a sporadic patient B, in whom *TPP1* variant analysis was requested because of the low TPP1 enzyme activity, obtained through the diagnostic work-up. Analysis of the *TPP1* closest homologs in other species (Figure 3B) showed that the Val466 residue is highly conserved during evolution, suggesting a functional role for this amino acid, and a deleterious effect for the predicted Val466Gly change (SIFT: deleterious, Polyphen: possibly damaging). Since both patients carried identical disease-causing variants we considered the possibility of the presence of founder alleles. Genealogical studies showed no close relation between the two families. Sanger sequencing of six closely linked variants (five of which are low frequent) covering the *OR56A3*, *TPP1* and *DCHS1* (Supp. Table S4) shows the presence of two haplotypes in patients of family A, one allele T-***C***-C-A-T, includes c.1397T>G and the other C-A-***G***-G-C, includes c.509-1G>C. They differ with the genotypes of patient B.II-1 at the two outer variants g.5968589C>T and g.6662466C>T, giving a maximal length of shared haplotype of 700 kb. The two families are not closely related, since the length is small. However, this does not exclude the variants derive from a common ancestor in the Dutch population. The results of molecular *TPP1* testing of patients from family A, patient B-II.1 and Dutch CLN2 patients (C-O) are summarized in Table 1.

Table 1. TPP1 enzyme activity, *TPP1* mutations and phenotype of patients from family A, B and other Dutch patients (C-O)

Patient	TPP1 activity leucocytes (nmol/h/m	TPP1 activity fibroblasts	TPP1 mutations	Predicted protein	Curvilinear profiles	Phenotype	Onset	Death yr/ (current age)	Affected relatives
A-III.2	16	7.2	c.509-1G>C/ c.1397T>G	splice defect/ p.Val466Gly	no	SCAR7	childhood	no/ (55)	4 sibs SCAR7
A-III.3	8	nd[1]	c.509-1G>C/ c.1397T>G	splice defect/ p.Val466Gly		SCAR7	childhood	no/ (59)	4 sibs, SCAR7
A-III.6	18	nd	c.509-1G>C/ c.1397T>G	splice defect/ p.Val466Gly		SCAR7	childhood	no/ (66)	4 sibs, SCAR7
A-III.9	Nd	nd	c.509-1G>C/ c.1397T>G	splice defect/ p.Val466Gly		SCAR7	childhood	68	4 sibs, SCAR7
A-III.10	32	nd	c.509-1G>C/ c.1397T>G	splice defect/ p.Val466Gly		SCAR7	childhood	no/ (73)	4 sibs, SCAR7
B-II.1	4.0-13.0[2]	nd	c.509-1G>C/ c.1397T>G	splice defect/ p.Val466Gly	no	SCAR7	18yr	no (51)	No
C	2.16	0	c.509 -1G>C/ c.509 -1G>C	splice defect/splice defect		CLN2	3 yr	no (13)	1 sib, CLN2
D	3.19	nd	nd	nd		CLN2	4.5 yr	23	No
E1	21.6	0.7-1.1	c.509 -1G>C/ c.509 -1G>C	splice defect/splice defect		CLN2	3.5 yr	11	1 sib, CLN2
E2	3.73	nd	c.509 -1G>C/ c.509 -1G>C	splice defect/splice defect		CLN2	3 yr	no (14)	1 sib, CLN2
F	3.88-9.04	nd	nd	nd		CLN2	4 yr	8	No
G	4.8- 5.14	0.4	c.509-1G>C/ c.622C>T	splice defect/p.Arg208X		CLN2	1.5 yr	no (11)	No
H	5.18	nd	nd	nd		CLN2	4 yr	no (6)	No
I	10.1-10.8	nd	c.622C>T/ c.1266G>C	p.Arg208X/p.Gln422His		CLN2	4 yr	no (9)	No
L	15.4	nd	nd	nd		CLN2	4 yr	12	No
M	24.6-32.3	0.3	c.509-1G>C/ c.622C>T	splice defect/p.Arg208X	yes	CLN2	3 yr	8	No
N	25.6-33.1	nd	c.509-1G>C/ c.622C>T	splice defect/p.Arg208X		CLN2	1.5 yr	no (7)	No
O1	27.7	nd	c.225A>G/ c.622C>T	splice defect/p.Arg208X	yes	CLN2	3 yr	10	1 sib, CLN2
O2	23.5	1.5	c.225A>G/ c.622C>T	splice defect/p.Arg208X	yes	CLN2	3 yr	no (17)	1 sib, CLN2

[1]nd = not done. [2] this sample was tested in Nijmegen; other samples in Rotterdam.

Table 2. The candidate variant list

Gene	Chromosome	Position	Reference base	Sample genotype	HGVS nomenclature	Function GVS
TPP1	11	6636430	A	A/C	NM_000391.3:c.1397T>G	Missense
TPP1	11	6638385	C	C/G	NM_000391.3:c.509-1G>C	splice-3
DCHS1	11	6645264	G	A/G	NM_003737.2:c.7643C>T	Missense
DCHS1	11	6662466	C	C/T	NM_003737.2:c.379G>A	Missense
C11orf40	11	4594558	-	-/G	NM_144663.1:c.286_287insC	Frameshift
C11orf40	11	4598956	C	C/T	NM_144663.1:c.95G>A	Nonsense

RNA test

The consequences of the splice site variant c.509-1G>C were studied by RT-PCR analysis of leucocyte RNA from patients in family A. Besides the expected normal fragment of 550 bp, a longer band (697 bp) was observed in the heterozygous carriers (Figure 3C). The bands were isolated from the agarose gel, and Sanger sequencing revealed that intron 5 was retained in the longer RNA fragment (r.[508_509ins508+1_509-1;509-1g>c]), indicating inactivation of the splice site by the c.509-1G>C change. The insertion of intron 5 caused a premature termination of translation 29 amino acids (p.Val170Glyfs*29).downstream of the splice site. Only one band was found for the RT-PCR around the c.1397T>G variant, indicating it had no effect on RNA processing (Figure 3D). The variant allele is therefore "active" and likely generating a p.Val466Gly missense variant at protein level.

TPP1 enzyme activity

TPP1 enzyme activity in leucocytes and fibroblasts of SCAR7 patients of family A, patient B-II.1 and several Dutch CLN2 patients (C-O) is described in Table 1. In all affected individuals, deficient activity of tripeptidyl peptidase 1 was found. There is however considerable overlap in enzyme activity in leucocytes from CLN2 and SCAR7 patients. For the affected individuals in family A, residual activity in leucocytes varied from 8 to 32 nmol/h/mg protein. The mean residual activity is 15% of the lowest control in family A and 10% in patient B-II.1 (4 nmol/h/mg protein). For CLN2 patients (C-O) the mean residual activity in leucocytes was 9% of the lowest control (2.16-23.5 nmol/h/mg protein), almost comparable with the SCAR7 patients. In fibroblasts however, this difference is more substantial with a residual enzyme activity of 0.4% of the lowest control in CLN2 patients and of 5% in patient A-III.2 (Table 1) from family A. The mean TPP1 activity in leucocytes of carriers with the splice site variant, c.509-1G>C was 132 nmol/h/mg protein and 139 nmol/h/mg protein in carriers with the missense variant, c.1397T>G, both as expected within the normal range (data not shown, but available upon request).

Electron microscopy

Electron microscopy of a skin biopsy tissue of one of the patients of family A (A-III.2) did not show the typical curvilinear profiles seen in patients with a typical CLN2 phenotype but some

granular osmiophilic deposits (GROD) and fingerprint profiles were noticed (Supp. Figure S1). Electron microscopy of a skin biopsy tissue of patient B-II.1 was normal; in particular there were no changes compatible with neuronal ceroid lipofuscinosis.

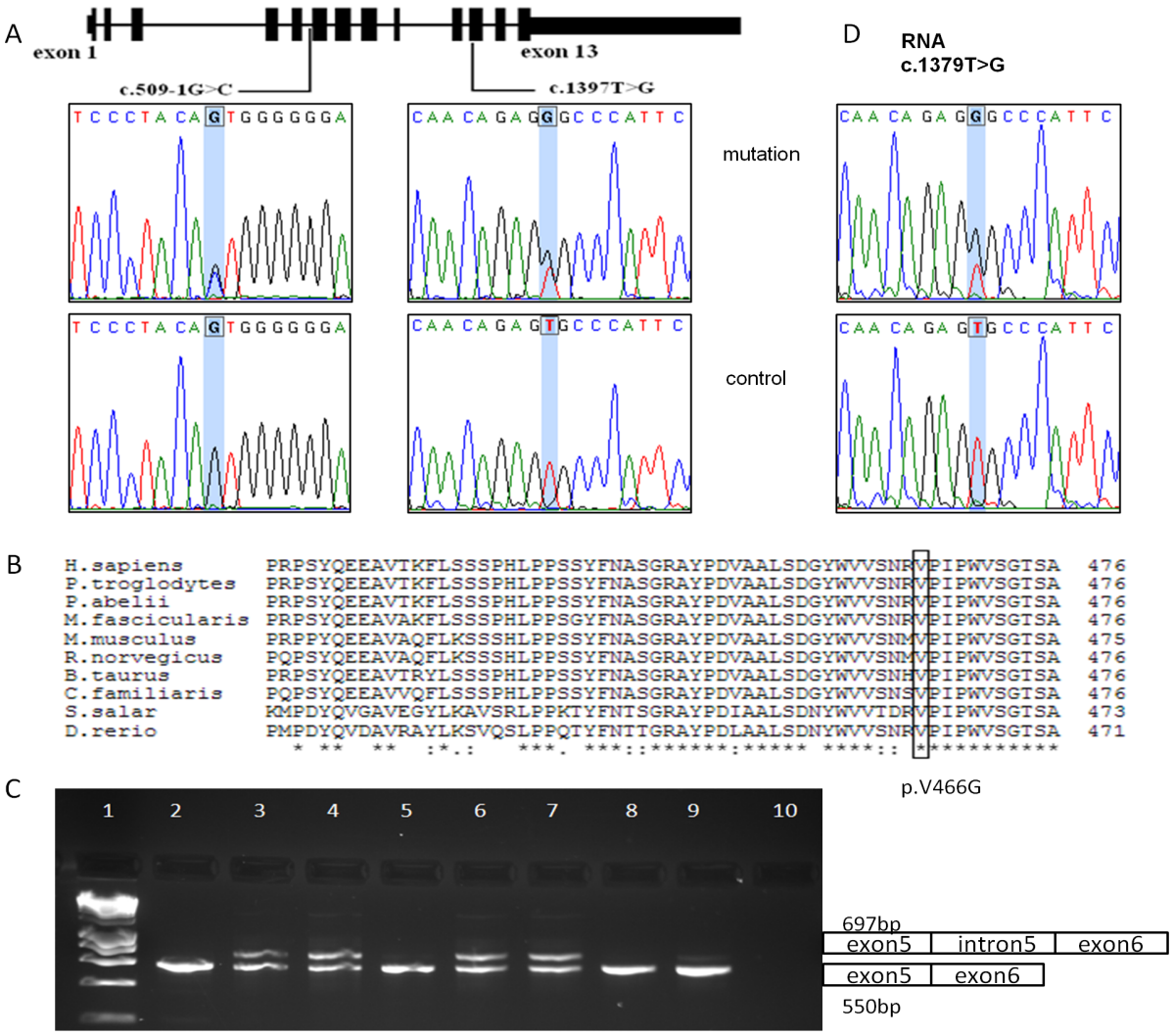


Figure 3. *TPP1* gene structure, variant analysis by Sanger sequencing and RT-PCR results. (A) *TPP1* gene structure with c.509-1G>C variant in exon 5 and c.1397T>G variant in exon 11 (top). Sanger sequencing confirmation of the two variants in patient sample compared to control (bottom). (B) Alignment of *TPP1* protein homologs. The amino acid Valine (V) at position 466, replaced by the missense variant in family A and patient B is highly conserved in species and highlighted in a rectangle (C) RT-PCR products from affected patients III-3, III-2 with compound heterozygous variants (c.509-1G>C, c.1397T>G) (lanes 3 and 6), un-affected III-4, III-8 (c.509-1G>C carriers) (lanes 7 and 4), unaffected III-5, III-7 (c.1397T>G carriers) (lanes 2 and 5), controls (lanes 8 and 9), and a blank (lane 10) analyzed in 2% agarose gel. Lane 1 size standard (Smartladder, Eurogentec). A longer band is observed in all heterozygous carriers of c.509-1G>C. (D) The Sanger sequencing result of c.1397T>G from RT-PCR, the variant is a true missense change.

Discussion

Exome sequencing in family A and Sanger sequencing, as part of a diagnostic workup in patient B-II.1, showed these compound heterozygous variants in *TPP1* as the cause of SCAR7. Defects in *TPP1* have previously been linked to CLN2. In the majority of cases, the age of onset of CLN2 is late infantile, between 2 to 4 years. It can also be infantile with onset before the age of 1 year (Ju, et al., 2002; Simonati, et al., 2000) or even juvenile, with disease onset between 6 and 10 years and a more protracted course (Bessa, et al., 2008; Elleder, et al., 2008; Hartikainen, et al., 1999; Kohan, et al., 2009; Sleat, et al., 1999; Wisniewski, et al., 1999). Developmental studies of TPP1 distribution in human brain and visceral organs, showed that the enzyme is not expressed in the developing neurons of the human fetus (Kida, et al., 2001; Kurachi, et al., 2001; Oka, et al., 1998). It appears in the neurons of the central nervous system at the age of 5 months and expression increases gradually to reach stable levels at the age of 3 years. This finding may explain why CLN2 and SCAR7 do not start in the early beginning of life.

The lipopigment pattern seen most often in CLN2 consists of curvilinear profiles, detectable by electron microscopy, in various cell types. There is a relationship between *TPP1* mutations, TPP1 activity, and curvilinear profiles (Mole, et al., 2005; Sleat, et al., 1999). In SCAR7 patients A.III-2 and B.II-1, however, the skin biopsy showed no curvilinear profiles. In CLN2 with a later onset and more protracted course, curvilinear profiles are not the only ultrastructural features found, also fingerprint profiles and GROD may appear (Wisniewski, et al., 1999). The skin biopsy of A.III-3 showed some GROD and fingerprint profiles, but no ultra structural features were found in patient B.II-1. It is suggested that there is a spectrum of ultra structural features in diseases caused by mutations in *TPP1,* ranging from curvilinear profiles in classic CLN2, mixed ultra structural features consisting of curvilinear- and fingerprint profiles and GROD in CLN2 with a late onset and protracted course, to only some GROD and fingerprint profiles or even absence of ultra structural features in SCAR7. Ultra structural findings show a correlation with the severity and course of the phenotype due to TPP1 deficiency, as was also shown before in mice (Sleat, et al., 2008).

The *TPP1* gene is composed of 13 exons. It encodes a member of the sedolisin family of serine proteases, tripeptidyl peptidase 1, mainly expressed in the lysosome and melanosome. The protease cleaves the N-terminal tripeptides from substrates, and it has a weak endopeptidase

activity. It is synthesized as a catalytically-inactive enzyme which is activated and auto-proteolyzed upon acidification. The TPP1 protein starts with a 19 amino acid signal peptide (Lin, et al., 2001; Sleat, et al., 1997) and a pro peptide of 176 amino acids, which will be removed in the mature form. The last part of the protein consists of the 368 amino acid tripeptidyl-peptidase 1 chain. The majority of the mutations in *TPP1* are located in the tripeptidyl-peptidase 1 domain, while only three mutations are localized in the propeptide domain, and none in signal peptide section. However, in the general population (data derived from the 1000 Genomes Project and GoNL project, including 500 unrelated Dutch individuals), the number of variations in propeptide domain and tripeptidyl-peptidase 1 chain is comparable (Supp. Table S5). This suggests that the propeptide section is more tolerant to variation, possibly due to the fact that this part of the protein is removed from the mature form and therefore may not have a significant effect on the function of the protein. The low variation in the signal peptide indicates the significance of that part of the protein. Without a recognizable signal peptide, the protein will not reach its destination nor will it be cleaved, and its function will probably be lost.

To facilitate genotype – phenotype studies, we examined the *TPP1* database, summarizing all variants published in the literature (http://www.ucl.ac.uk/ncl/cln2.shtml, date August 14, 2012). The mutations reported so far in relation to *TPP1*, including missense, nonsense, insertion, deletion, splice site, were scattered throughout the whole gene. Several mutations are recurrent, such as the stop codon p.(Arg208*), a splice site variant c.509-1G>C, which is present in patients of family A and B, and a Newfoundland founder variant p.(Gly284Val) (Ju, et al., 2002). The missense mutation found in patients of family A and patient B.II-1, c.1397T>G, Val466Gly, was not reported before, but showed conservation in evolution and is located in the peptidase region of the protein.

Genotype – phenotype relations in the neuronal ceroid lipofuscinoses have been reviewed and tested in Chinese hamster cells (Kousi, et al., 2012; Mole, et al., 2005; Walus, et al., 2010). A loss of TPP1 function will cause the CLN2-late infantile, which means the TPP1 enzyme activity will be extremely reduced or absent. By examining the variant spectrum of *TPP1*, some variants will evidently truncate the protein, while some missense and in-frame insertion or deletion variants are observed to impair enzyme function of the protein. We have summarized the *TPP1* genotype with at least one missense variant reported in the literature in Supp. Table S6, and link

the experimental data (Guhaniyogi, et al., 2009; Lin and Lobel, 2001; Pal, et al., 2009; Walus, et al., 2010) and prediction tools (Desmet, et al., 2009) to elucidate the potential effect of those mutations. The majority of variants clearly inactivate the gene by creating an early stop of translation by introducing either a nonsense, or a frameshift variant. Some mutations, like c.1266G>C and c.380G>A, are located near the splice sites, and may therefore alter splicing and disrupt the reading frame (Chao, et al., 2010; Sun, et al., 2010). Other variants are predicted to generate a new splice site by Human Splicing Finder (Desmet, et al., 2009) (http://www.umd.be/HSF/, date July 06, 2012). For those mutations, it is worthwhile to study RNA to verify the predicted truncating effect. Another category of variant are those located within the active site of the protein (c.827A>T (Kohan, et al., 2009), c.1424C>T (Sleat, et al., 1999)), so even a minor change might affect the function of the protein significantly.

In patients of family A and patient B-II.1, RNA analysis of the splice site variant c.509-1G>C, showed that retention of intron 5 in the reading frame generates a premature stop codon, leading to haplo-insufficiency through nonsense mediated decay. For the amino acid changes from Valine to Glycine, unlike other CLN2 "missense" mutations, prediction tools do not show that it will produce a severe splicing alteration (HSF), indicating the protein product translated from this allele might still work actively.

Although there is considerable overlap in mean enzyme activity in leucocytes from CLN2 (9%) and SCAR7 patients (10-15%), in fibroblasts it differs about a factor 10 (0.4% in CLN2 patients and 5% in SCAR7 patient A.III-2), although the number of patients studied is very small (Table 1). The low activity in blood and especially in fibroblasts give an indication of the overall TPP1 enzyme activity in the central nervous system. Sleat et al showed in CLN2 mutant mice that low TPP1 levels attenuated disease. Compound heterozygosity for a null allele and a presumed hypomorphic p.Arg447His missense variant resulted in a later onset and a protracted disease with survival into the third or fourth decade of life. Mice homozygous for this hypomorphic mutation, showed locomotor deficits at a later age, with a slower disease progression, compared to homozygous null allele mutated mice and compound heterozygote mice and also showed a greatly extended life span, approaching that of normal mice. The brains of these mice showed approximately 3% of normal TPP1 activity compared to homozygous null allele mutated mice expressing 0.2% of normal levels (Sleat, et al., 2008). A semantic data mining approach

comparing model organism and clinical phenotype data (Chen, et al., 2012) identified this homozygous TPP1 hypomorphic mouse as the 25th best match for SCAR7 out of the 25,141 mouse models annotated at MGI (Damian Smedley, personal communication). The mouse model shows the clinical features through the presence of ataxia, Purkinje cell degeneration, neurodegeneration, tremor, and audiogenic seizures, which highly resemble the SCAR7 phenotype.

We infer the following genotype – phenotype correlation: loss of function variants abolishing TPP1 enzyme activity lead to CLN2, while variants that diminish TPP1 enzyme activity lead to SCAR7 (Table 3). Therapeutic approaches, causing a small increase in TPP1 enzyme activity in brain, might change the course of the disease and extend the lifespan of CLN2 patients by pushing them towards a more SCAR7-like phenotype, but higher levels will be required to cure the disease. Further investigations are needed to confirm this hypothesis.

Table 3. The phenotypes and genotypes of patients with *TPP1* mutations

	CLN2, late infantile	CLN2, juvenile	SCAR7
General	very severely affected	less severely affected	mild phenotype and protracted course
Age of onset	2-4 years	10-Jun	childhood or teenage
Age of death	5-15 years	> 12-40 years	> 60 years
Clinical findings	seizures, dementia, visual loss, ataxia and cerebral atrophy	seizures, dementia, visual loss, ataxia and/or cerebral atrophy, protracted course	Cerebellar ataxia, pyramidal signs, deep sensory loss, cerebellar atrophy
TPP1 enzyme activity	extremely low or none	residual or very low	Residual
Ultrastructural features (EM)	curvilinear bodies	curvilinear bodies/GROD/fingerprint profiles	some GROD/fingerprint profiles/none
Alleles	null/null	null/partial affected	null/minor modification

To conclude, SCAR7 is caused by compound heterozygous variants in *TPP1*. The genetic background of cerebellar ataxias are even more heterogeneous than the neuronal ceroid lipofuscinoses with a still growing number of subtypes and we here add *TPP1* to the list of genes implicated in the autosomal recessive ataxias. The phenotype associated with *TPP1* variants is expanded now by an autosomal recessive form of slowly progressive cerebellar ataxia. Diagnostic work-up for unexplained spinocerebellar ataxias should thus include analysis of TPP1 enzyme activity, particularly if the family history or the age of onset is suggestive of an

autosomal recessive disorder. Other features that could suggest *TPP1* mutations, i.e. CNL2 features such as visual regression, epilepsy or curvilinear profiles in a skin biopsy, can be absent. This finding again illustrates the sometimes unexpected clinical spectrum of variants in known genes. We will encounter this phenomenon with increasing frequency using new techniques such as whole exome sequencing.

Acknowledgments

We would like to thank the patients for their kind participation and Leiden Genome Technology Center (LGTC), Laboratory for Diagnostic Genome Analysis (LDGA) and Department of Clinical Genetics, Erasmus Medical Center Rotterdam for technical support. Yu Sun was supported by China Scholarship Council.

References

Bessa C, Teixeira CA, Dias A, Alves M, Rocha S, Lacerda L, Loureiro L, Guimaraes A, Ribeiro MG. 2008. CLN2/TPP1 deficiency: the novel mutation IVS7-10A>G causes intron retention and is associated with a mild disease phenotype. Mol Genet Metab 93:66-73.

Breedveld GJ, van Wetten B, te Raa GD, Brusse E, van Swieten JC, Oostra BA, Maat-Kievit JA. 2004. A new locus for a childhood onset, slowly progressive autosomal recessive spinocerebellar ataxia maps to chromosome 11p15. J Med Genet 41:858-66.

Chao SC, Chen JS, Tsai CH, Lin JM, Lin YJ, Sun HS. 2010. Novel exon nucleotide substitution at the splice junction causes a neonatal Marfan syndrome. Clin Genet 77:453-63.

Chen CK, Mungall CJ, Gkoutos GV, Doelken SC, Kohler S, Ruef BJ, Smith C, Westerfield M, Robinson PN, Lewis SE and others. 2012. MouseFinder: Candidate disease genes from mouse phenotype data. Hum Mutat 33:858-66.

Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C. 2009. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic Acids Res 37:e67.

Elleder M, Dvorakova L, Stolnaja L, Vlaskova H, Hulkova H, Druga R, Poupetova H, Kostalova E, Mikulastik J. 2008. Atypical CLN2 with later onset and prolonged course: a neuropathologic study showing different sensitivity of neuronal subpopulations to TPP1 deficiency. Acta Neuropathol 116:119-24.

Guhaniyogi J, Sohar I, Das K, Stock AM, Lobel P. 2009. 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 284:3985-97.

Hartikainen JM, Ju W, Wisniewski KE, Moroziewicz DN, Kaczmarski AL, McLendon L, Zhong D, Suarez CT, Brown WT, Zhong N. 1999. Late infantile neuronal ceroid lipofuscinosis is due to splicing mutations in the CLN2 gene. Mol Genet Metab 67:162-8.

Ju W, Zhong R, Moore S, Moroziewicz D, Currie JR, Parfrey P, Brown WT, Zhong N. 2002. Identification of novel CLN2 mutations shows Canadian specific NCL2 alleles. J Med Genet 39:822-5.

Kida E, Golabek AA, Walus M, Wujek P, Kaczmarski W, Wisniewski KE. 2001. Distribution of tripeptidyl peptidase I in human tissues under normal and pathological conditions. J Neuropathol Exp Neurol 60:280-92.

Kohan R, Cismondi IA, Kremer RD, Muller VJ, Guelbert N, Anzolini VT, Fietz MJ, Ramirez AM, Halac IN. 2009. An integrated strategy for the diagnosis of neuronal ceroid lipofuscinosis types 1 (CLN1) and 2 (CLN2) in eleven Latin American patients. Clin Genet 76:372-82.

Kousi M, Lehesjoki AE, Mole SE. 2012. Update of the mutation spectrum and clinical correlations of over 360 mutations in eight genes that underlie the neuronal ceroid lipofuscinoses. Hum Mutat 33:42-63.

Kurachi Y, Oka A, Itoh M, Mizuguchi M, Hayashi M, Takashima S. 2001. Distribution and development of CLN2 protein, the late-infantile neuronal ceroid lipofuscinosis gene product. Acta Neuropathol 102:20-6.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754-60.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078-9.

Lin L, Lobel P. 2001. Expression and analysis of CLN2 variants in CHO cells: Q100R represents a polymorphism, and G389E and R447H represent loss-of-function mutations. Hum Mutat 18:165.

Lin L, Sohar I, Lackland H, Lobel P. 2001. The human CLN2 protein/tripeptidyl-peptidase I is a serine protease that autoactivates at acidic pH. J Biol Chem 276:2249-55.

Matilla-Duenas A. 2012. The Ever Expanding Spinocerebellar Ataxias. Editorial. Cerebellum

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M and others. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20:1297-303.

Mole SE, Williams RE, Goebel HH. 2005. Correlations between genotype, ultrastructural morphology and clinical phenotype in the neuronal ceroid lipofuscinoses. Neurogenetics 6:107-26.

Oka A, Kurachi Y, Mizuguchi M, Hayashi M, Takashima S. 1998. The expression of late infantile neuronal ceroid lipofuscinosis (CLN2) gene product in human brains. Neurosci Lett 257:113-5.

Pal A, Kraetzner R, Gruene T, Grapp M, Schreiber K, Gronborg M, Urlaub H, Becker S, Asif AR, Gartner J and others. 2009. Structure of tripeptidyl-peptidase I provides insight into the molecular basis of late infantile neuronal ceroid lipofuscinosis. J Biol Chem 284:3976-84.

Sailer A, Houlden H. 2012. Recent advances in the genetics of cerebellar ataxias. Curr Neurol Neurosci Rep 12:227-36.

Sailer A, Scholz SW, Gibbs JR, Tucci A, Johnson JO, Wood NW, Plagnol V, Hummerich H, Ding J, Hernandez D and others. 2012. Exome sequencing in an SCA14 family demonstrates its utility in diagnosing heterogeneous diseases. Neurology 79:127-31.

Santavuori P. 1988. Neuronal ceroid-lipofuscinoses in childhood. Brain Dev 10:80-3.

Santen GW, Aten E, Sun Y, Almomani R, Gilissen C, Nielsen M, Kant SG, Snoeck IN, Peeters EA, Hilhorst-Hofstee Y and others. 2012. Mutations in SWI/SNF chromatin remodeling complex gene ARID1B cause Coffin-Siris syndrome. Nat Genet 44:379-80.

Simonati A, Santorum E, Tessa A, Polo A, Simonetti F, Bernardina BD, Santorelli FM, Rizzuto N. 2000. A CLN2 gene nonsense mutation is associated with severe caudate atrophy and dystonia in LINCL. Neuropediatrics 31:199-201.

Sleat DE, Donnelly RJ, Lackland H, Liu CG, Sohar I, Pullarkat RK, Lobel P. 1997. Association of mutations in a lysosomal protein with classical late-infantile neuronal ceroid lipofuscinosis. Science 277:1802-5.

Sleat DE, El-Banna M, Sohar I, Kim KH, Dobrenis K, Walkley SU, Lobel P. 2008. Residual levels of tripeptidyl-peptidase I activity dramatically ameliorate disease in late-infantile neuronal ceroid lipofuscinosis. Mol Genet Metab 94:222-33.

Sleat DE, Gin RM, Sohar I, Wisniewski K, Sklower-Brooks S, Pullarkat RK, Palmer DN, Lerner TJ, Boustany RM, Uldall P and others. 1999. Mutational analysis of the defective protease in classic late-infantile neuronal ceroid lipofuscinosis, a neurodegenerative lysosomal storage disorder. Am J Hum Genet 64:1511-23.

Sun Y, Almomani R, Aten E, Celli J, van der Heijden J, Venselaar H, Robertson SP, Baroncini A, Franco B, Basel-Vanagaite L and others. 2010. Terminal osseous dysplasia is caused by a single recurrent mutation in the FLNA gene. Am J Hum Genet 87:146-53.

Van Diggelen OP, Keulemans JL, Kleijer WJ, Thobois S, Tilikete C, Voznyi YV. 2001. Pre- and postnatal enzyme analysis for infantile, late infantile and adult neuronal ceroid lipofuscinosis (CLN1 and CLN2). Eur J Paediatr Neurol 5 Suppl A:189-92.

Vermeer S, van de Warrenburg BP, Willemsen MA, Cluitmans M, Scheffer H, Kremer BP, Knoers NV. 2011. Autosomal recessive cerebellar ataxias: the current state of affairs. J Med Genet 48:651-9.

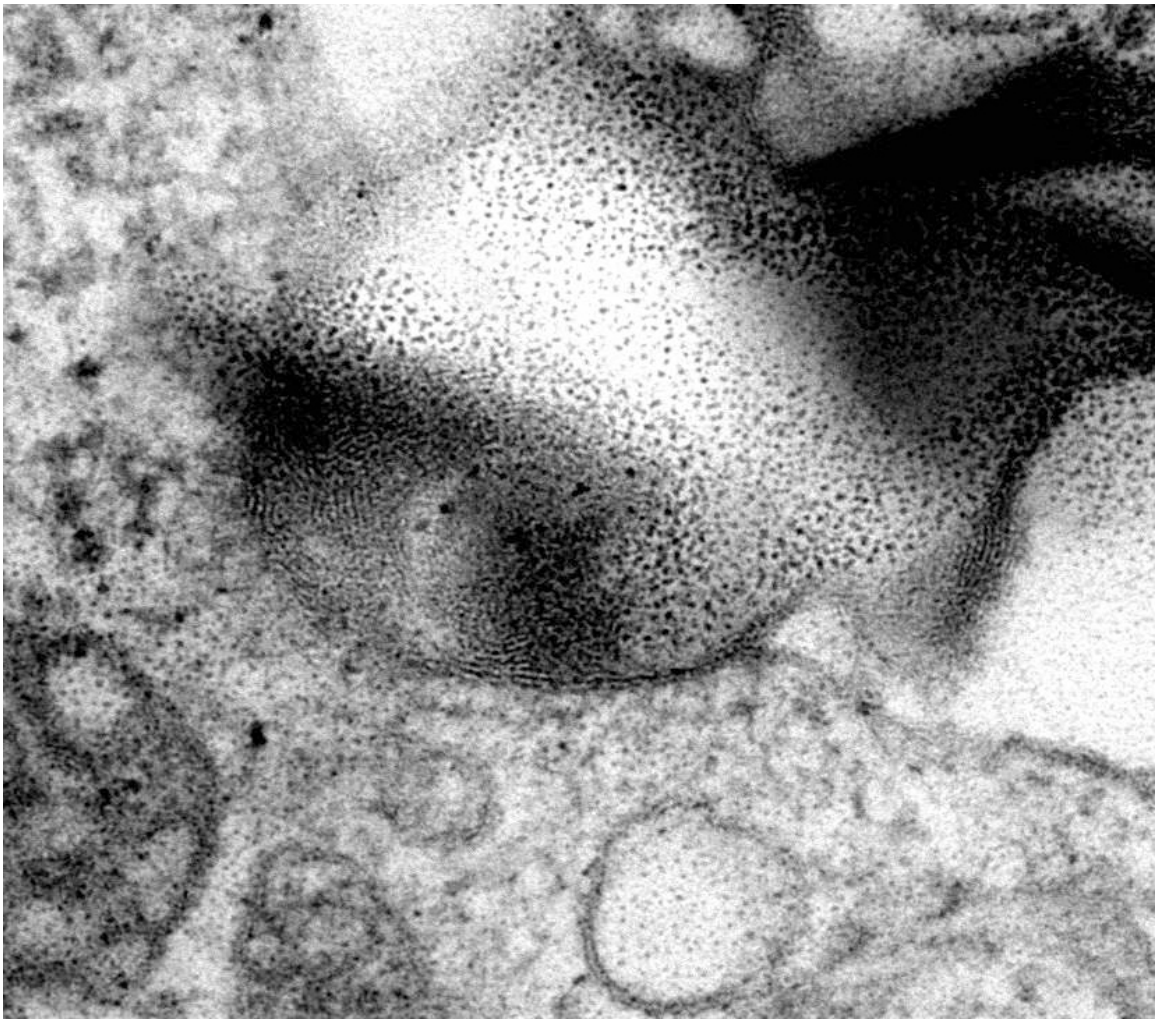
Walus M, Kida E, Golabek AA. 2010. Functional consequences and rescue potential of pathogenic missense mutations in tripeptidyl peptidase I. Hum Mutat 31:710-21.

Williams RE, Boyd S, Lake BD. 1999. Ultrastructural and electrophysiological correlation of the genotypes of NCL. Mol Genet Metab 66:398-400.

Williams RE, Mole SE. 2012. New nomenclature and classification scheme for the neuronal ceroid lipofuscinoses. Neurology 79:183-91.

Wisniewski KE, Kaczmarski A, Kida E, Connell F, Kaczmarski W, Michalewski MP, Moroziewicz DN, Zhong N. 1999. Reevaluation of neuronal ceroid lipofuscinoses: atypical juvenile onset may be the result of CLN2 mutations. Mol Genet Metab 66:248-52.

Supplementary data:



Supplementary Figure 1. Electron microscopy of a skin biopsy tissue of patient A.III-2 which shows granular osmiophilic deposits (GROD) and fingerprint profiles. The magnification of EM image was 20,000 times.

Supplementary. Table S1. SCAR types

SCAR	Locus	Gene	Clinical findings other than spinocerebellar ataxia
1	9q34.13	*SETX*	onset 10-25 yr, progressive, ocular apraxia, axonal neuropathy, tremor, pyramidal signs, elevated AFP
2	9q34-qter	*?*	congenital, ID, small head, cataracts, pyramidal signs, intention tremor, short stature
3	6p23-p21	*?*	early-onset, hearing impairment, optic atrophy
4	1p36	*?*	onset 3rd decade, progressive, pyramidal signs, myoclonic jerks, fasciculations, impaired joint position sense
5	15q25.3	*ZNF592*	Congenital, severe psychomotor retardation, short stature, pyramidal signs, microcephaly, optic atrophy, speech defect, abnormal osmiophilic pattern of skin vessels (CAMOS)
6	20q11-q13	*?*	onset in infancy, nonprogressive; delayed motor and speech development, no ID, hypotonia, pes planus
7	11p15	*TPP1*	childhood-onset, slowly progressive
8	6q25.1-q25.2	*SYNE1*	late-onset, slow progression
9	1q42.13	*ADCK3*	childhood onset, progressive, cerebellar atrophy, seizures, developmental delay, hyperlactatemia
10	3p22.1	*ANO10*	onset teenage-young adulthood, hyperreflexia, nystagmus, atrophy lower limbs with fasciculations tortuosity of the conjunctival vessels, ID, pes cavus
11	1q32.2	SYT14	onset 6th decade, progressive, ID
12	16q21-q23	?	onset early-childhood, generalized seizures, delayed psychomotor development, ID
13	6q24.3	GRM1	onset infancy, slowly progressive, ID with poor or absent speech, hyperreflexia, eye movement abnormalities

Supplementary Table S2. Primer list

Primer	Sequence (5'-3')	DNA/RNA
DCHS2_ex21_F	GTCAGCTGCAGCCACTGTTA	DNA
DCHS2_ex21_R	TGTGGCTGTGACTGAAGACC	DNA
DCHS2_ex2_F	GGTGCCAAAAGCTGTATGCT	DNA
DCHS2_ex2_R	TGCAGATTGATGAGGAGCAG	DNA
TPP1_ex11_F	AGGGGTTCTAGGTGCAAGGT	DNA
TPP1_ex11_R	CCAGGAACCTTTCCTCATCA	DNA
TPP1_ex5-6_F	TGTTATTGCTGGTGCCAGAG	DNA
TPP1_ex5-6_R	CAGGGATGCTCAGAGGTAGC	DNA
NOP56_ex11_F	AAGGAGTCCTCAGAGCACCA	DNA
NOP56_ex11_R	CCACTGTGAAACACGACCAC	DNA
TPP1_RNA_ex5_F2	GTCTCACCTTTGCCCTGAGA	RNA
TPP1_RNA_ex6_R2	AGGAACTGGGCACAGGCTT	RNA
TPP1_RNA_ex11_F	CTGATGGCTACTGGGTGGTC	RNA
TPP1_RNA_ex12-13_R	AGCCACGGGTTACATCAAAG	RNA

Supplementary Table S3. The overview of the exome sequencing

Patient	*A III-2*
total reads	77479996
Read 1	38739998
read length 1 (nt)	100
Read 2	38739998
read length 2 (nt)	100
aligned reads	75802233
PCT_aligned_reads	97.83%
properly paired	74211004
PCT_properly_paired	95.78%
PERCENT_DUPLICATION	34.65%
MEAN_BAIT_COVERAGE	48.930747
PCT_TARGET_BASES_10X	86.41%
PCT_SELECTED_BASES	82.89%
FOLD_ENRICHMENT	38.026774
ZERO_CVG_TARGETS_PCT	4.39%
FOLD_80_BASE_PENALTY	2.825506

Supplementary Table S4. The genotype of six variants around the *TPP1* mutations, and the inferred haplotypes in family A

Gene	OR56A3	TPP1	TPP1	TPP1	DCHS1	*DCHS1*
Variation in Transcript	c.13C>T	c.1542A>T	c.1397T>G	c.509-1G>C	c.7643C>T	c.379G>A
Genomic Location	g.5968589C>T	g.6636106T>A	g.6636430A>C	g.6638385C>G	g.6645264G>A	g.6662466C>T
A.III-2	C/T	T/A	A/C	C/G	G/A	C/T
A.III-3	C/T		A/C	C/G	G/A	C/T
A.III-6			A/C	C/G		
A.III-10			A/C	C/G		
A.III-9			A/C	C/G		
A.III-1	C/C		A/A	C/C	G/G	C/C
A.III-8	C/C		A/A	C/G	G/G	C/C
A.III-7	C/T		A/C	C/C	G/A	C/T
A.III-5	C/T		A/C	C/C	G/A	C/T
A.IV-1	C/C		A/A	C/C	G/G	C/C
A.III-11	C/C		A/A	C/G	G/G	C/C
A.II-1			A/A	C/C		
A.III-4			A/A	C/G		
B.II-1	C/C	T/A	A/C	C/G	G/A	C/C
B.I-1	C/C	T/A	A/C	C/C	G/A	C/C
Haplotype allele 1 in patients of family A	T		C	C	A	T
haplotype allele 2 in patients of family A	C		A	G	G	C

Supplementary Table S5. The variants found in normal populations

	1000 Genomes		GoNL	
	Nonsynonymous	Missense	Nonsynonymous	Missense
Signal peptide	1	0	0	1
Propeptide	4	6	0	3
Protease	4	6	5	2

Supplementary Table S6. Reported genotypes in *TPP1* with at least one missense, inframe change or intronic variant and information from literatures and prediction tools. When the phenotype is not available, the background is marked grey. Here CLN2=CLN2, late infantile, JNCL=CLN2, juvenile, INCL= CLN2, infantile. SD = splice donor site, SA = splice acceptor site, BS = splice branch site, HSF = predicted by Human Splicing Finder.

Missense Allele 1	protein domain	Allele 2	phenotype	Splicing effect (Experiment and prediction)	Functional Study by Walus 2010			Structure study by Pal 2008	Reference
					TPP1 activity[1]	Lysosomal transport	Half-lives of proenzymes[2]		
c.225A>G, p.(Gln75Gln)	Propeptide	c.509-1G>C	CLN2	Splice defect according to the article (prediction)					Sleat 1999
c.225A>G, p.(Gln75Gln)	Propeptide	c.1678_1679delCT, p.(Leu560Thrfs*47)	NA	Splice defect according to the article (prediction)					Kousi 2012
c.184T>A, p.(Ser62Thr)	Propeptide	?	NA	HSF- new BS					Kousi 2012
c.229G>A, p.(Gly77Arg)	Propeptide	c.509-1G>C	CLN2	Last nucleotide of exon 3	1	PA	4.5		Sleat 1999
c.229G>A, p.(Gly77Arg)	Propeptide	c.640C>T, p.(Gln214*)	CLN2	Last nucleotide of exon 3	1	PA	4.5		Kousi 2012
c.229G>A, p.(Gly77Arg)	Propeptide	c.790C>T, p.(Gln264*)	CLN2	Last nucleotide of exon 3	1	PA	4.5		Kousi 2012

c.229G>A, p.(Gly77Arg)	Propeptide	c.1062del, p.(Leu355Serfs*72)	CLN2	Last nucleotide of exon 3	1	PA	4.5		Kousi 2012
c.299A>G, p.(Gln100Arg)	Propeptide	c.1266+5G>A	NA	HSF-enhancer interrupted					Kousi 2012
c.1266+5G>A	tripeptidyl-peptidase 1 chain	c.299A>G, p.(Gln100Arg)	NA	HSF- broken +new SD					Kousi 2012
c.380G>A, p.(Arg127Gln)	Propeptide	c.380G>A, p.(Arg127Gln)	NA	Last nucleotide of exon 4	43.6	N	1.7	NA, surface exposure	Kousi 2012
c.380G>A, p.(Arg127Gln)	Propeptide	c.622C>T, p.(Arg208*)	CLN2	Last nucleotide of exon 4	43.6	N	1.7	NA, surface expo-sure	Steinfeld 2002
c.380G>A, p.(Arg127Gln)	Propeptide	c.509-1G>C	CLN2	Last nucleotide of exon 4	43.6	N	1.7	NA, surface exposure	Zhong 2000
c.380G>A, p.(Arg127Gln)	Propeptide	?	NA	Last nucleotide of exon 4	43.6	N	1.7	NA, surface exposure	Kousi 2012
c.381-17_-4del	Propeptide	c.229G>T, p.(Gly77*)	CLN2	HSF-SA Broken					Chang 2012
c.457T>C, p.(Ser153Pro)	Propeptide	?	NA, JNCL		NA	NA	NA		Cillaud 1999; Mole 2001
c.524G>A, p.(Arg175His)	Propeptide	?	NA	HSF-enhancer interrupted					Kousi 2012
c.605C>T, p.(Pro202Leu)	Tripeptidyl-peptidase 1 chain	?	NA		0	A	6.5, homodimer of proenzyme		Mole 2001
c.616C>T, p.(Arg206Cys)	Tripeptidyl-peptidase 1 chain	?	NA		0.7	A	2.3		Mole 2001
c.616C>T, p.(Arg206Cys)	Tripeptidyl-peptidase 1 chain	c.616C>T, p.(Arg206Cys)	CLN2		0.7	A	2.3		Tessa 2000
c.617G>A, p.(Arg206His)	Tripeptidyl-peptidase 1 chain	?	NA	HSF-SD Enhancer disrupted					Kousi 2012

c.625T>C, p.(Tyr209His)	Tripeptidyl-peptidase 1 chain	c.625T>C, p.(Tyr209His)	NA	HSF- new BS					Kousi 2012
c.646G>A, p.(Val216Met)	Tripeptidyl-peptidase 1 chain	c.1551+1G>A	CLN2	HSF-SD broken Enhancer disrupted	NA	NA	NA		Wang 2011
c.650G>T, p.(Gly217Val)	Tripeptidyl-peptidase 1 chain	c.640C>T, p.(Gln214*)	CLN2	HSF-SD broken	NA	NA	NA		Chang 2012
c.797G>A, p.(Arg266Gln)	Tripeptidyl-peptidase 1 chain	c.1015C>T, p.(Arg339Gln)	NA	HSF-new SA	NA	NA	NA		Kousi 2012
c.827A>T, p.(Asp276Val)	Tripeptidyl-peptidase 1 chain	c.827A>T, p.(Asp276Val)	CLN2		NA	NA	NA	Active site	Kohan 2009
c.827A>T, p.(Asp276Val)	Tripeptidyl-peptidase 1 chain	c.622C>T, p.(Arg208*)	CLN2		NA	NA	NA	Active site	Kohan 2009
c.829G>A, p.(Val277Met)	Tripeptidyl-peptidase 1 chain	?	CLN2		0	PA	6.8, homodimer of proenzyme	Might affect active site	Ju 2002
c.833A>C, p.(Gln278Pro)	Tripeptidyl-peptidase 1 chain	?	CLN2	HSF- SA broken disrupt an alpha-helix	NA	NA	NA	Might affect active site	Ju 2002
c.843G>T, p.(Met281Ile)	Tripeptidyl-peptidase 1 chain	?	NA	HSF-SD broken					Kousi 2012
c.851G>T, p.(Gly284Val)	Tripeptidyl-peptidase 1 chain	c.509-1G>C	CLN2	HSF-SD broken, enhancer disrupted	0.4	A	3.9, homodimer of proenzyme		Zhong 2000; Ju 2002
c.851G>T, p.(Gly284Val)	Tripeptidyl-peptidase 1 chain	c.622C>T, p.(Arg208*)	CLN2	HSF- SD broken	0.4	A	3.9, homodimer of proenzyme		Ju 2002
c.851G>T, p.(Gly284Val)	Tripeptidyl-peptidase 1 chain	c.851G>T, p.(Gly284Val)	CLN2	HSF- SD broken	0.4	A	3.9		Ju 2002
c.851G>T, p.(Gly284Val)	Tripeptidyl-peptidase 1 chain	?	CLN2	HSF- SD broken	0.4	A	3.9, homodimer of proenzyme		Ju 2002

c.851G>T, p.(Gly284Val)	Tripeptidyl-peptidase 1 chain	Uncharacterised 1bp deletion in exon 7	CLN2	HSF- SD broken	0.4	A	3.9, homodimer of proenzyme		Ju 2002
c.857A>G, p.(Asn286Ser)	Tripeptidyl-peptidase 1 chain	c.857A>G, p.(Asn286Ser)	CLN2	HSF-new SA	0	PA	3.1, homodimer of proenzyme	Loss N-linkd glycos-ylation, surface exposure	Steinfeld 2002
c.860T>A, p.(Ile287Asn)	Tripeptidyl-peptidase 1 chain	?	NA		0.1	PA	3.8, homodimer of proenzyme		Sleat 1999
c.887-10A>G	Tripeptidyl-peptidase 1 chain	c.196C>T, p.(Gln66*)/c.89+4A>G	JNCL, milder than CLN2	New SA by RT-PCR, 9 nt inserted between c.886_887insAAAATCCAG	NA	NA	NA	NA	Kohan 2009
c.887-10A>G, p.(Gly296delinsGluAsnProGly)	Tripeptidyl-peptidase 1 chain	c.887-10A>G, p.(Gly296delinsGluAsnProGly)	JNCL, much milder than CLN2	New SA by RT-PCR, 9 nt inserted between c.886_887insAAAATCCAG	NA	NA	NA	NA	Bessa 2008
c.887-18A>G, p.(Gly296delinsGlyLysLysLysAsnPoGly)	Tripeptidyl-peptidase 1 chain	c.509-1G>C	CLN2	Splice defect according to the article (prediction)					Sleat 1999
c.984_986del, p.(Asp328del)	Tripeptidyl-peptidase 1 chain	?	NA	splice site, according to the article HSF-enhancer dirupted					Kousi 2012
c.987_989delinsCTC, p.(Glu329_Asp330delinsAspSer)	Tripeptidyl-peptidase 1 chain	?	NA	HSF-SA broken	NA	NA	NA	NA	Kousi 2012
c.1015C>T, p.(Arg339Gln)	Tripeptidyl-peptidase 1 chain	c.797G>A, p.(Arg266Gln)	NA	HSF-new SA					Kousi 2012
c.797G>A, p.(Arg266Gln)	Tripeptidyl-peptidase 1	c.1015C>T, p.(Arg339Gln	NA	HSF-new SA					Kousi 2012

	chain	)							
c.1015C>T, p.(Arg339Gln)	Tripeptidyl-peptidase 1 chain	?	NA	HSF-new SA					Kousi 2012
c.1027G>A, p.(Glu343Lys)	Tripeptidyl-peptidase 1 chain	c.1027G>A, p.(Glu343Lys)	CLN2		0	A	ND		Sleat 1999
c.1057A>C, p.(Thr353Pro)	Tripeptidyl-peptidase 1 chain	c.509-1G>C	CLN2		NA	NA	NA		Steinfeld 2002
c.1064T>C, p.(Leu355Pro)	Tripeptidyl-peptidase 1 chain	?	NA		NA	NA	NA		Hofmann 2002, Kousi 2012
c.1093T>C, p.(Cys365Arg)	Tripeptidyl-peptidase 1 chain	c.1595dupA, p.(Glu534Profs*74)	CLN2	HSF-SD broken	0	PA	7.8	Loss of disulfide-bond, surface exposure	Sleat 1999
c.1094G>A, p.(Cys365Tyr)	Tripeptidyl-peptidase 1 chain	c.509-1G>C	CLN2	HSF-SD broken	NA	NA	NA	Loss of disulfide bond	Sleat 1999
c.1094G>A, p.(Cys365Tyr)	Tripeptidyl-peptidase 1 chain	c.1094G>A, p.(Cys365Tyr)	CLN2	HSF-SD broken	NA	NA	NA	Loss of disulfide bond	Sleat 1999
c.1094G>A, p.(Cys365Tyr)	Tripeptidyl-peptidase 1 chain	c.1361C>A, p(Ala454Glu)	CLN2	HSF-SD broken	NA	NA	NA	Loss of disulfide bond	Sleat 1999
c.1094G>A, p.(Cys365Tyr)	Tripeptidyl-peptidase 1 chain	c.1525C>T, p.(Gln509X)	NA	HSF-SD broken	NA	NA	NA		Kousi 2012
c.1094G>A, p.(Cys365Tyr)	Tripeptidyl-peptidase 1 chain	?	NA	HSF-SD broken	NA	NA	NA		Kousi 2012
c.1146C>G, p.(Ser382Arg)	Tripeptidyl-peptidase 1 chain	c.887-10A>G	NA	first nucleotide of exon 10					Kousi 2012
c.1154T>A, p.(Val385Asp)	Tripeptidyl-peptidase 1 chain	c.622C>T, p.(Arg208*)	CLN2	HSF-SD broken	NA	NA	NA		Sleat 1999

c.1166G>A, p.(Gly389Glu) + c.299A>G, p.(Gln100Arg)	Tripeptidyl-peptidase 1 chain	c.1166G>A, p.(Gly389Glu) + c.299A>G, p.(Gln100Arg)	CLN2	HSF-SD broken; two homozygous missense mutations	NA	NA	NA		Sleat 1999
c.1266G>C, p.(Gln422His)	Tripeptidyl-peptidase 1 chain	c.622C>T, p.(Arg208*)	CLN2, NA	Last nucleotide of exon 10	0	PA	5.1		Sleat 1999; Kousi 2012
c.1266G>C, p.(Gln422His)	Tripeptidyl-peptidase 1 chain	c.509-1G>C	CLN2	Last nucleotide of exon 10	0	PA	5.1		Sleat 1999; Steinfeld 2002
c.1266G>C, p.(Gln422His)	Tripeptidyl-peptidase 1 chain	c.1266G>C, p.(Gln422His)	CLN2	Last nucleotide of exon 10	0	PA	5.1		Sleat 1999
c.1269G>C, p.(Glu423Asp)	Tripeptidyl-peptidase 1 chain	?	CLN2	Third nucleotide of exon 11	NA	NA	NA		Sleat 2001
c.1284G>T, p.(Lys428Asn)	Tripeptidyl-peptidase 1 chain	c.509-1G>C	CLN2	HSF-SA broken	NA	NA	NA		Zhong 2000 , Ju 2002
c.1340G>A, p.(Arg447His)	Tripeptidyl-peptidase 1 chain	c.622C>T, p.(Arg208*)	JNCL		1.8	PA	7.1		Wisniewski 1999; Sleat 1999
c.1340G>A, p.(Arg447His)	Tripeptidyl-peptidase 1 chain	c.509-1G>C	JNCL		1.8	PA	7.1		Wisniewski 1999; Sleat 1999
c.1343C>T, p.(Ala448Val); c.1501G>T, p.(Gly501Cys)	Tripeptidyl-peptidase 1 chain	c.622C>T, p.(Arg208*)	NA						Kousi 2012
c.1358C>T, p.(Ala453Val)	Tripeptidyl-peptidase 1 chain	c.311T>A, p.(Leu104*)	CLN2	HSF-new SD	NA	NA	NA		Kohan 2009
c.1417G>A, p.(Gly473Arg)	Tripeptidyl-peptidase 1 chain	p.(Ser62Glyfs*25)	CLN2	HSF- new SA, SD broken	NA	NA	NA	Disturb catalytic activity	Lam 2001
c.1417G>A,	Tripeptidyl-peptidase 1	c.622C>T,	CLN2	HSF- new SA, SD	NA	NA	NA	Disturb catalytic	Zhong

p.(Gly473Arg)	chain	p.(Arg208*)		broken				activity	2000
c.1424C>T, p.(Ser475Leu)	Tripeptidyl-peptidase 1 chain	c.509-1G>C	CLN2, NA	Second last nucleotide of exon 11	0	N	1.2	Active site, surface exposure, impair processing	Sleat 1999; Kousi 2012
c.1424C>T, p.(Ser475Leu)	Tripeptidyl-peptidase 1 chain	c.622C>T, p.(Arg208*)	NA	Second last nucleotide of exon 11	0	N	1.2	Active site, surface exposure	Kousi 2012
c.1424C>T, p.(Ser475Leu)	Tripeptidyl-peptidase 1 chain	c.1424C>T, p.(Ser475Leu)	NA	Second last nucleotide of exon 11	0	N	1.2	Active site, surface exposure	Kousi 2012
c.1424C>T, p.(Ser475Leu)	Tripeptidyl-peptidase 1 chain	?	NA	Second last nucleotide of exon 11	0	N	1.2	Active site, surface exposure	Kousi 2012
c.1439T>G, p.(Val480Gly)	Tripeptidyl-peptidase 1 chain	c.622C>T, p.(Arg208*)	JNCL		NA	NA	NA		Elleder 2008
c.1439T>G, p.(Val480Gly)	Tripeptidyl-peptidase 1 chain	?	NA		NA	NA	NA		Elleder 2008
c.1442T>G, p.(Phe481Cys)	Tripeptidyl-peptidase 1 chain	c.851G>T, p.(Gly284Val)	INCL	HSF- new SD	NA	NA	NA		Ju 2002
c.1444G>C, p.(Gly482Arg)	Tripeptidyl-peptidase 1 chain	c.1444G>C, p.(Gly482Arg)	NA	HSF-SD broken	NA	NA	NA		Kousi 2009
c.1551+5_1551+6delinsTA	Tripeptidyl-peptidase 1 chain	c.622C>T, p.(Arg208*)	NA	Splice defect according to the article					Kousi 2012
c.1551+5_1551+6delinsTA	Tripeptidyl-peptidase 1 chain	c.1551+5_1551+6delinsTA	NA	Splice defect according to the article					Kousi 2012
c.1397T>G, p.(Val466Gly)	Tripeptidyl-peptidase 1 chain	c.509-1G>C	SCAR 7						This study

[1] % of wild type TPP1
[2] unit = hr, wild type TPP1 : 1.9 hr