ARTICLE

The Novel Neuronal Ceroid Lipofuscinosis Gene *MFSD8* Encodes a Putative Lysosomal Transporter

Eija Siintola, Meral Topcu, Nina Aula, Hannes Lohi, Berge A. Minassian, Andrew D. Paterson, Xiao-Qing Liu, Callum Wilson, Ulla Lahtinen, Anna-Kaisa Anttonen, and Anna-Elina Lehesjoki

The late-infantile–onset forms are the most genetically heterogeneous group among the autosomal recessively inherited neurodegenerative disorders, the neuronal ceroid lipofuscinoses (NCLs). The Turkish variant was initially considered to be a distinct genetic entity, with clinical presentation similar to that of other forms of late-infantile–onset NCL (LINCL), including age at onset from 2 to 7 years, epileptic seizures, psychomotor deterioration, myoclonus, loss of vision, and premature death. However, Turkish variant LINCL was recently found to be genetically heterogeneous, because mutations in two genes, *CLN6* and *CLN8,* were identified to underlie the disease phenotype in a subset of patients. After a genomewide scan with single-nucleotide–polymorphism markers and homozygosity mapping in nine Turkish families and one Indian family, not linked to any of the known NCL loci, we mapped a novel variant LINCL locus to chromosome 4q28.1-q28.2 in five families. We identified six different mutations in the *MFSD8* gene (previously denoted "*MGC33302*"), which encodes a novel polytopic 518–amino acid membrane protein that belongs to the major facilitator superfamily of transporter proteins. *MFSD8* is expressed ubiquitously, with several alternatively spliced variants. Like the majority of the previously identified NCL proteins, MFSD8 localizes mainly to the lysosomal compartment. However, the function of MFSD8 remains to be elucidated. Analysis of the genome-scan data suggests the existence of at least three more genes in the remaining five families, further corroborating the great genetic heterogeneity of LINCLs.

The neuronal ceroid lipofuscinoses (NCLs) are a group of autosomal recessive neurodegenerative lysosomal storage disorders characterized by the accumulation of autofluorescent storage material in many cell types, including neurons.[1] Clinically, they present with a variable age at onset, epileptic seizures, progressive psychomotor decline, visual failure, and premature death.[2] To date, 10 different forms of human NCLs and seven causative genes (*PPT1/CLN1* [MIM 600722], *TPP1/CLN2* [MIM 607998], *CLN3* [MIM 607042], *CLN5* [MIM 608102], *CLN6* [MIM 606725], *CLN8* [MIM 607837], and *CTSD/CLN10* [MIM 116840]) have been identified.[3–11] The loci and genes for adult-onset NCL (CLN4 [MIM 204300]),[12] the Turkish variant (CLN7) late-infantile–onset NCL (LINCL),[13] and the recently identified CLN9-variant LINCL (MIM 609055)[14] have remained undetected. Within the NCLs, the late-infantile–onset group is genetically the most heterogeneous, with causative mutations found in *CLN1, CLN2, CLN5, CLN6,* and *CLN8*.[5,6,8,9,15–17]

The phenotype of approximately half of the Turkish patients with NCL is classified as variant LINCL (vLINCL), with clinical presentation similar to the other forms of vLINCL.[18] The age at onset ranges from 2 to 7 years, and, in most patients, onset has been considered to be somewhat later than in the classical LINCL (CLN2 [MIM 204500]).[18] The initial symptom is most commonly epileptic seizures, but motor, visual, and speech impairment, as well as developmental regression with ataxia, may also be presenting symptoms. The course of seizures is usually more severe than in classical LINCL. With time, progressive cognitive and motor deterioration, myoclonus, personality changes, and loss of vision develop, leading to premature death.[18,19] On electron microscopic analysis, Turkish patients with vLINCL have shown condensed fingerprint profiles in circulating lymphocytes.[18] Initially, Turkish vLINCL was considered a distinct clinical and genetic entity,[13] but it has turned out to be genetically heterogeneous. We recently reported four mutations in the *CLN8* gene and two mutations in the *CLN6* gene accounting for the disease in a subset of Turkish patients with vLINCL.[16,20] However, in most patients, the phenotype is not linked to any of the known NCL genes.

We here describe the identification and initial characterization of a novel gene, major facilitator superfamily (MFS) domain containing 8 (*MFSD8*), for Turkish vLINCL. The gene was identified after homozygosity mapping with a panel of 10 families, for which we had excluded, by haplotype analysis, all known human NCL loci as well as two loci homologous for genes causing NCL-like phenotypes in animal models (*CLCN3* [MIM 600580][21,22] and *CLCN7* [MIM 602727][23]).

From the Folkhälsan Institute of Genetics (E.S.; N.A.; H.L.; U.L.; A.-K.A.; A.-E.L.), Neuroscience Center (E.S.; N.A.; U.L.; A.-K.A.; A.-E.L.), and Department of Medical Genetics (E.S.; H.L.; A.-K.A.), University of Helsinki, Helsinki, Finland; Department of Pediatrics, Faculty of Medicine, Section of Child Neurology, Hacettepe University, Ankara, Turkey (M.T.); Genetics and Genome Biology, Hospital for Sick Children (H.L.; B.A.M.; A.D.P.; X.-Q.L.), and Department of Public Health Sciences, University of Toronto (A.D.P.), Toronto, Canada; and Starship Children's Hospital, Auckland, New Zealand (C.W.)

Received February 26, 2007; accepted for publication April 13, 2007; electronically published May 14, 2007.

Address for correspondence and reprints: Anna-Elina Lehesjoki, Folkhälsan Institute of Genetics, Biomedicum Helsinki, P.O. Box 63 (Haartmaninkatu 8), 00014 University of Helsinki, Helsinki, Finland. E-mail: anna-elina.lehesjoki@helsinki.fi

Am. J. Hum. Genet. 2007;81:136–146. © 2007 by The American Society of Human Genetics. All rights reserved. 0002-9297/2007/8101-0013$15.00
DOI: 10.1086/518902

Material and Methods

Patients, Controls, and Samples

Nine Turkish families with vLINCL (a, b, c, d, e, f, g, h, and l), one family of Turkish-Georgian origin (k), and one family of Indian origin (j), altogether comprising 13 patients, 21 parents, and five unaffected siblings, were analyzed. All families except one (j) were consanguineous. The clinical phenotypes of eight patients were described elsewhere (patients 17, 18, 22, 24, 25, 27, 28, and 29 in the study by Topcu et al.[18]). The corresponding patient codes in the present study are e3, a3, f3, b3, c3, d4, d3, and g3, respectively.

Patient e5 (the brother of e3) presented at age 3.5 years with a history of delayed speech development, ataxia since age 2.5 years, and stereotyped hand movements. Brain magnetic resonance imaging (MRI) showed cerebellar and cerebral atrophy and increased signal intensity in white matter. At age 5 years, he had prominent myoclonic seizures and rare nocturnal seizures that responded well to treatment with levetiracetam and intravenous immunoglobulin. Electroencephalogram (EEG) showed slow background activity and multifocal epileptic discharges. Eye ground examination showed retinopathy and optic atrophy. At age 6 years, he could not walk without support. For the past 6 mo, he has been wheelchair bound and unable to speak.

Patient h3 presented at age 5.5 years with a history of delayed speech development, as well as ataxia that had been worsening over the previous 6 mo, frequent falling, and sleep disorders since age 3.5 years. On examination at 5.5 years, she had ataxic gait with abnormal cerebellar tests. She was able to speak slowly. Brain MRI showed cerebellar atrophy. She is taking sodium valproate for myoclonic seizures. She has tested negative for *MECP2* (MIM 300005) mutations and for cerebral folate deficiency. No vacuoles were found on analysis of peripheral blood smear.

Patient j3, who was of Indian origin, presented at age 3 years with a history of mild generalized developmental delay. Over the previous 3 mo, her gait had become increasingly ataxic, and, on examination, she walked with a wide-based gait. At age 4 years, she developed frequent clonic seizures that responded well to treatment with sodium valproate. A wide range of metabolic investigations, including enzymologic analysis of PPT1 and TPP1 and molecular analysis of *CLN3* and *CLN6,* were normal. MRI of the brain showed mild cerebellar atrophy and evidence of delayed white matter maturation. A rectal biopsy showed neuronal curvilinear bodies. She had progressive neurological decline and, by age 6 years, was wheelchair bound and had limited communication, significant visual impairment, and frequent breakthrough seizures. At age 7 years, she developed treatment-resistant status epilepticus. She died just before her 8th birthday.

Patient k3, who was of Turkish-Georgian origin, presented at age 8 years with a history of delayed psychomotor development (she walked at age 2.5 years, spoke at age 3 years, and did not learn to read). She showed cognitive deterioration, occasional agitation (with a few hospitalizations in a psychiatric unit), loss of vision, and walking difficulties. At age 22 years, she was agitated, she had axial akinesia, hypertonia, palilalia, and blindness. She developed problems with swallowing. Peripheral blood lymphocytes showed vacuoles, and a skin biopsy revealed vacuoles with fingerprint profiles.

Patient l3 presented at age 2 years with a 6-wk history of ataxia and myoclonic and atonic seizures that responded well to treatment with sodium valproate. Psychomotor development, retinal examination, and MRI were normal, whereas EEG showed occipital spikes. An axillary skin biopsy showed fingerprint profiles.

Genomic DNA was extracted using standard techniques from EDTA blood samples collected after informed consent was obtained. The control panel for the identified *MFSD8* mutations consisted of 212 Turkish and 92 Centre d'Etude du Polymorphisme Humain (CEPH) chromosomes. RNA for the analysis of a splice-site mutation was extracted from peripheral blood samples and for the analysis of the alternative splicing pattern of *MFSD8,* from a control fibroblast cell line. An institutional review board of the Helsinki University Central Hospital approved the study.

Candidate-Locus and Genomewide SNP-Scan Analyses

The previously known NCL loci (*CLN1, CLN2, CLN3, CLN5, CLN6, CLN8, CLN10, CLCN3,* and *CLCN7*) were excluded by microsatellite analysis, as described elsewhere.[20] Four microsatellite markers (*D4S427, D4S2975, D4S2938,* and *D4S429*) spanning ~6.5 cM (according to the Marshfield Genetic Map in the Mammalian Genotyping Service) were genotyped around the novel *CLN7* locus.

Ten families with vLINCL (a, b, c, d, e, f, g, j, l, and k) were included in the genomewide SNP scan. One affected individual was genotyped from each family, except from one pedigree (d) that had four individuals (two affected and two unaffected siblings) genotyped. The SNP scan was performed using Affymetrix's GeneChip Human Mapping 50K Xba 240 arrays in accordance with the manufacturer's instructions. SNPs were merged with the Rutgers map[24] by their physical locations (NCBI build 35). Interval-specific interpolation was performed. From a total of 58,960 SNPs, the ones without chromosome numbers and physical location (466) or with zero heterozygosity values or male heterozygotes on chromosome X (3,904) were removed. Because the sample size in this data set was too small to calculate SNP heterozygosity values, the values from the Affymetrix white sample were used. Since linkage disequilibrium between markers can cause false-positive results when both parents are not genotyped,[25] we selected markers for linkage analysis instead of using all of them. We did this by grouping the remaining 54,590 SNPs into clusters, with at least 0.1 cM between the clusters. One SNP with the largest heterozygosity value was chosen from each cluster, resulting in the use of 8,704 SNPs. The computer program Merlin (version 1.0.1) was used for multipoint homozygosity mapping.[26] The disease-allele frequency was set at 0.0001, with penetrances of 0, 0, and 1 for an autosomal recessive model.

Gene Identification and Mutation Analysis

The gene content of the ~19.5-Mb genomic region between *rs2051785* and *rs984369* on chromosome 4q28.1-q28.2 was examined with the NCBI Map Viewer, build 35.1. Positional candidate genes were chosen on the basis of the known or putative functions of the encoded proteins. The exons and the exon-intron boundaries of the candidate genes were screened for mutations by genomic sequencing. PCR conditions and primer sequences are available from the authors on request. Sequencing of the purified PCR products was performed with an ABI 3730 DNA Analyzer (Applied Biosystems). Controls were screened for the identified mutations by genomic sequencing.

Northern Blot and RT-PCR

The tissue expression of *MFSD8* was analyzed with Multiple Tissue Northern Blots (Human, Human Brain II, and Human Brain V [BD Biosciences Clontech]) with a 559-bp (c.36_594) probe that was PCR amplified from *MFSD8* cDNA (I.M.A.G.E Consortium cDNA clone ID 5272664[27] and RZPD clone IRATp970E0532D). The probe was [^{32}P]-labeled with the Rediprime II Random Prime Labeling System (Amersham Biosciences) and was hybridized in accordance with the manufacturer's protocol (BD Biosciences Clontech).

For RT-PCR, RNA isolated from blood or from cultured fibroblasts was reverse transcribed using M-MLV Reverse Transcriptase (Promega). The consequence of the c.754+2T→A mutation was analyzed using primers from exons 6 and 11, and the alternative splicing of *MFSD8* in fibroblasts was analyzed using several exonic primer pairs that covered the entire coding region, as well as primers specific for various exon-lacking transcripts. The PCR products were extracted from agarose gels and were sequenced.

Construction of Expression Plasmids and Site-Directed Mutagenesis

The PCR-amplified ORF of the human *MFSD8* cDNA was cloned in frame into the aminoterminal hemagglutinin (HA) tag containing pAHC expression vector (kindly provided by Prof. T. Mäkelä, University of Helsinki, Finland), a derivative of pCIneo (Promega) (^{HA}MFSD8 construct), and into the pcDNA3.1(+) expression vector (Invitrogen), into which a carboxylterminal HA tag was generated (MFSD8^{HA} construct) by site-directed mutagenesis by use of the QuikChange Site-Directed Mutagenesis Kit (Stratagene). Two nucleotide changes, c.929G→A (p.Gly310Asp) and c.1286G→A (p.Gly429Asp), corresponding to patient missense mutations, were introduced into the ^{HA}MFSD8 construct by site-directed mutagenesis. All constructs were verified by sequencing.

In Vitro Translation

In vitro translation was performed with the TnT Quick Coupled Transcription/Translation System (Promega) and the Canine Pancreatic Microsomal Membranes (Promega) by use of the ^{HA}MFSD8 and MFSD8^{HA} constructs as templates. The [^{35}S]-labeled translation products were analyzed on 12% SDS-PAGE gels and were visualized by autoradiography.

Cell Culture, Transfections, and Immunofluorescence Analysis

COS-1 and HeLa cells (American Type Culture Collection) were cultivated in Dulbecco's modified Eagle's medium (BioWhittaker) supplemented with 10% and 5% fetal calf serum (PromoCell), respectively, penicillin, streptomycin, and 1 × GlutaMAX (Gibco). The cells (1.5–2 × 10^5 cells) were plated onto 6-well plates on coverslips 1 d before transfection and were transfected with 2 μg of wild-type or missense mutation containing ^{HA}MFSD8 or wild-type MFSD8^{HA} constructs per well by use of FuGENE 6 Transfection Reagent (Roche). At 18 h after transfection, cells were incubated with cycloheximide for 2 h, were fixed with 4% paraformaldehyde, and were permeabilized either with 0.1% Triton X-100 in PBS or with 0.2% saponin in PBS supplemented with 0.5% BSA. Overexpressed proteins were detected using mouse monoclonal or rabbit polyclonal anti-HA antibodies (clones 16B12 [Covance Research Products] and Y-11 [Santa Cruz Biotechnology], respectively). Antibodies used as organelle markers were mouse monoclonal anti-early endosome antigen 1 (EEA1 [BD Biosciences]), rabbit polyclonal anti-cathepsin D (CTSD [DAKO]), rabbit polyclonal anti-giantin (BioSite), mouse monoclonal anti–lysosomal–associated membrane protein 1 (Lamp-1, H4A3 [developed by J. T. August and J. E. K. Hildreth and obtained from the Developmental Studies Hybridoma Bank, developed under the auspices of the National Institute of Child Health and Human Development and maintained by The University of Iowa, Department of Biological Sciences, Iowa City]), mouse monoclonal anti-lysobisphosphatidic acid (LBPA, 6C4 [kindly provided by Professor Jean Gruenberg, University of Geneva, Switzerland]), mouse monoclonal anti–mannose-6-phosphate receptor 46 (MPR46 [kindly provided by Professor Kurt von Figura, Göttingen, Germany]), and mouse monoclonal anti–protein-disulfide isomerase (PDI [Stressgen]). Cy2- or Cy3-conjugated secondary antibodies (Jackson ImmunoResearch) were used to visualize the primary antibodies. The cells were viewed using an Axioplan 2 microscope and were photographed with AxioVision 3.1 (Zeiss).

Bioinformatics

Alternatively spliced transcripts were identified using NCBI AceView. The topology of MFSD8 was predicted using HMMTOP, TMHMM, and TMpred prediction programs. Protein domains were searched for in the Pfam 20.0 database. Proteins homologous and similar to MFSD8 were searched for by NCBI protein-protein BLAST. The peptide sequences were aligned using MAFFT version 5.8.

Results

Genomewide SNP Scan and Homozygosity Mapping

Analysis of genomewide SNP-scan data in 10 families with vLINCL—for which we had excluded all known NCL loci either elsewhere[20] or in this study (data not shown) (see the "Material and Methods" section)—revealed three regions, on chromosomes 4, 8, and 15, with heterozygosity LOD (HLOD) scores >2 (see the tab-delimited ASCII file, which can be imported into a spreadsheet, of data set 1 [online only]). Six families (a, b, c, e, f, and j) contributed to the highest HLOD score of 3.39 at SNP *rs348085* on chromosome 4q28.1-q28.2, where an ~19.5-Mb region (from *rs7657655* to *rs10518621*) of overlapping homozygosity was detected in all but one of these families. In family a, the homozygous region was <1 Mb, which was considered to be too short to reflect a region homozygous by descent, considering the close consanguinity. The haplotypes across the region were different in each family (data not shown). Four families (d, e, g, and l) contributed to the HLOD score of 2.48 on chromosome 15 (at SNP *rs2956147*). Four families (b, c, e, and l) contributed to the HLOD score of 2.05 on chromosome 8 (at SNP *rs2432960*). We decided to first focus on the ~19.5-Mb region on chromosome 4 because it had the strongest evidence for linkage. Genotyping four microsatellite markers (*D4S427, D4S2975, D4S2938,* and *D4S429*) in patients in families b, c, e, f, and j and in an additional family (h) did not narrow the region but revealed that the patient in family h also is homozygous for these markers.

Identification of Mutations in MFSD8 *Underlying vLINCL*

Within the critical region of ~19.5 Mb, we identified at least 90 known or putative genes. After excluding two of the genes (*TRAM1L1* and *TRPC3* [MIM 602345]) by sequencing, we identified six homozygous mutations (table 1 and fig. 1) in *MFSD8* (a HUGO Gene Nomenclature Committee–approved symbol; previously denoted "*MGC33302*"; GenBank accession number NM_152778), by screening the 12 protein-coding exons (exons 2–13) of the gene. All six mutations cosegregated with the disease phenotype in the respective families and were not found in 212 Turkish and 92 CEPH control chromosomes. Patient j3 was homozygous for nonsense mutation c.894T→G in exon 10, which created a premature stop codon (p.Tyr298X) and was predicted to truncate the protein by 221 aa. Patients c3 and b3 were homozygous for missense mutations c.929G→A (p.Gly310Asp) in exon 10 and c.1286G→A (p.Gly429Asp) in exon 12, respectively, which affect amino acids that are conserved across vertebrates (fig. 2). In two patients, we identified homozygous nucleotide changes at the exon-intron junctions that may either change an amino acid or affect the splicing of the transcript: an A→G transition at the second-to-last nucleotide of exon 7 (c.697A→G; p.Arg233Gly) in patient f3, and a transversion of G→C at the last nucleotide of exon 11 (c.1102G→C; p.Asp368His) in patient h3. Both Arg233 and Asp368 are conserved in vertebrates (fig. 2). The only intronic mutation detected (c.754+2T→A) was identified in intron 8 in two affected siblings, e3 and e5. An RT-PCR analysis performed from an RNA sample of patient e5 with primers from exons 6 and 11 revealed an altered pattern of PCR products (fig. 1*B*), implying aberrant splicing of the mutant transcript. No mutations in *MFSD8* were identified in patients that lacked homozygosity over the *CLN7* locus (families a, d, g, k, and l).

MFSD8 *mRNA Expression and Alternatively Spliced Variants*

In northern-blot analysis of *MFSD8* expression, an ~5-kb transcript, expressed at very low levels, was detected in all tissues analyzed (fig. 3). In all brain regions and in lung, it was the only transcript seen. In other tissues, transcripts that ranged in size from ~1 to ~3 kb and/or ~6 kb were also present.

Both northern-blot and EST database analyses support the ubiquitous expression of the main *MFSD8* transcript, including exons 1–13, whereas the alternatively spliced variants seem to have more-limited tissue distribution (NCBI AceView). In particular, four alternative, partial variants were present in EST databases: one lacking exon 2 (BI553701), one lacking exon 7 (BX341359), one lacking exons 7 and 8 (BG434527), and one lacking exon 11 (represented by three sequences: BX341358, BG542082, and BU618424). None of these variants produce full-length in-frame transcripts. Translation of the variant lacking exon 2 is predicted to start at methionine 46 at exon 3 (p.Met46). The variant lacking both exons 7 and 8 produces a transcript with an in-frame deletion of 67 aa, whereas the others lead to frameshifts and premature stop codons. We verified these and identified several other alternatively spliced transcripts by RT-PCR analyses (data not shown).

MFSD8 Protein

MFSD8 is predicted to encode a 518-aa protein of ~58 kDa with 12 predicted transmembrane domains (fig. 4). Pfam analysis of the MFSD8 amino acid sequence showed that it contains an MFS domain (MFS_1) at amino acid positions p.42_477 and a sugar (and other) transporter domain (Sugar_tr) at amino acid positions p.72_147. In an in vitro translation assay, HA-tagged MFSD8 proteins were detected as ~60-kDa bands on SDS-PAGE (data not shown), in agreement with the calculated molecular weight of the protein. In spite of several attempts under various conditions, we were not able to detect soluble overexpressed HA-MFSD8 in transfected cell lysates on western blot with the use of HA-antibody, which is likely to be a result of the predicted excessive hydrophobicity of the protein, poor transfer of the protein, or partial proteolysis of the HA tag.

MFSD8 appears to be evolutionarily conserved, since a BLAST search returned several homologues for MFSD8 in different species. In each vertebrate species, MFSD8 has a single orthologue. For instance, mouse and zebrafish orthologues are ~82% and ~61% identical to the human protein, respectively (fig. 2). In invertebrates, several weakly similar proteins (~18%–35% of amino acids are

Table 1. *MFSD8* Mutations Identified in the Present Study

Nucleotide Change	Amino Acid Change or Predicted Consequence	Exon or Intron	Patient Code in Present Study (Code in Study by Topcu et al.[18])	Origin
c.697A→G	p.Arg233Gly or splicing effect	Exon 7	f3 (22)	Turkey
c.754+2T→A	Altered splicing	Intron 8	e3 (17) and e5	Turkey
c.894T→G	p.Tyr298X	Exon 10	j3	India
c.929G→A	p.Gly310Asp	Exon 10	c3 (25)	Turkey
c.1102G→C	p.Asp368His or splicing effect	Exon 11	h3	Turkey
c.1286G→A	p.Gly429Asp	Exon 12	b3 (24)	Turkey

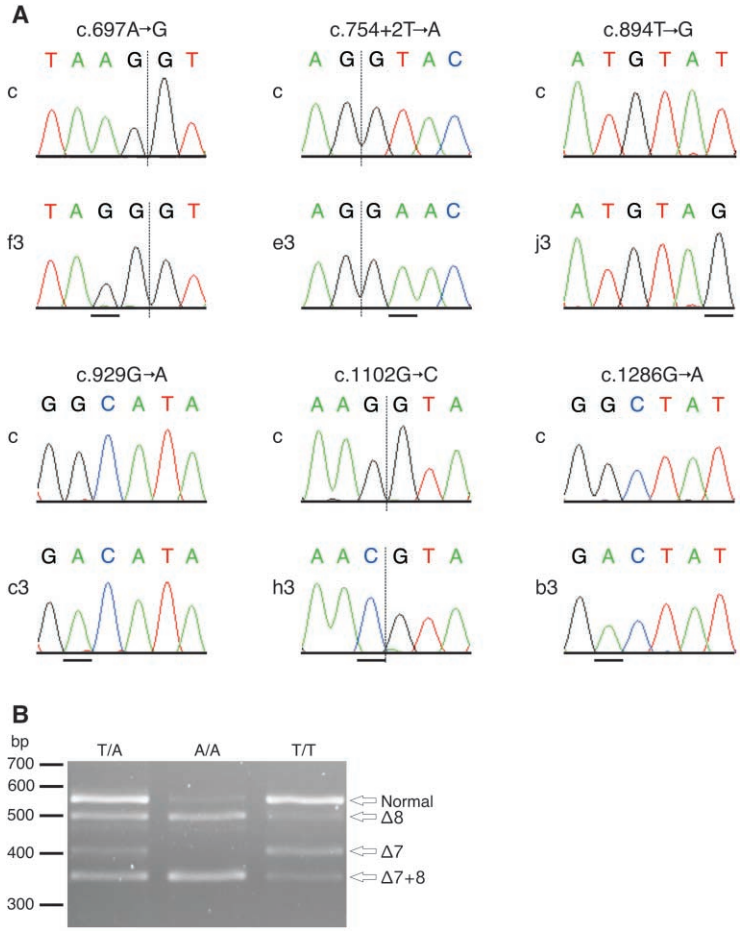

Figure 1. *MFSD8*-associated mutations. *A,* Sequence chromatograms showing homozygous *MFSD8* mutations in six patients with vLINCL (f3, e3, j3, c3, h3, and b3) and the respective control (c) sequences. The mutated bases are underlined, and the vertical lines indicate exon-intron boundaries. *B,* Aberrant splicing of the c.754+2T→A mutant transcript. The agarose gel shows the mRNA splicing pattern in patient e5 (A/A), who is homozygous for c.754+2T→A, in his heterozygous carrier mother (T/A), and in a control individual (T/T), analyzed by RT-PCR with primers from exons 6 and 11. In the sample from patient e5, an altered pattern of PCR products is seen, with almost complete lack of the normal transcript (~550 bp) containing exons 7–10, a complete lack of an alternatively spliced variant lacking exon 7 (Δ7) (~400 bp), and increased expression of two alternatively spliced variants: one lacking exon 8 (Δ8) (~500 bp) and one lacking exons 7 and 8 (Δ7+8) (~350 bp). In addition, a faint band of ~480 bp, the sequence of which remains unknown, is visible in the samples from patient e5 and his mother. The positions of the 100-bp ladder-size marker are shown on the left.

identical to those in human MFSD8) were identified: 2 in *Drosophila melanogaster,* 18 in *Caenorhabditis elegans,* and 4 in *Saccharomyces cerevisiae*.

Intracellular Localization of MFSD8

To study the subcellular localization of MFSD8, we transiently overexpressed the ^{HA}MFSD8 construct in COS-1 cells and, by using HA antibodies in immunofluorescence analysis, detected the protein in punctate structures in the cytoplasm (fig. 5*A,* 5*D,* and 5*G*). This staining pattern was confirmed by overexpressing MFSD8^{HA} in COS-1 cells and also by transfecting HeLa cells with both constructs (data not shown). A variety of organelle markers were used to identify the MFSD8-positive punctate cytoplasmic structures in COS-1 cells. The HA-MFSD8 protein displayed the strongest overlap with lysosomal markers Lamp-1 (fig. 5*A*–5*C*), CTSD, and LBPA (data not shown). No overlap was observed with early endosomal protein EEA1 (fig. 5*D*–5*F*) or with markers used to visualize the compartments of the early secretory pathway, including endoplasmic reticulum (ER) resident PDI, Golgi protein giantin (data not shown), and trans-Golgi network protein MPR46 (fig. 5*G*–5*I*). Patient missense mutations p.Gly310Asp and p.Gly429Asp, which were introduced into the wild-type MFSD8, did not interfere with the lysosomal localization of the HA-MFSD8 protein in transiently transfected COS-1 cells (fig. 5*J*–5*O*).

Discussion

In this article, we describe the mapping, identification, and initial characterization of a novel NCL gene. Because

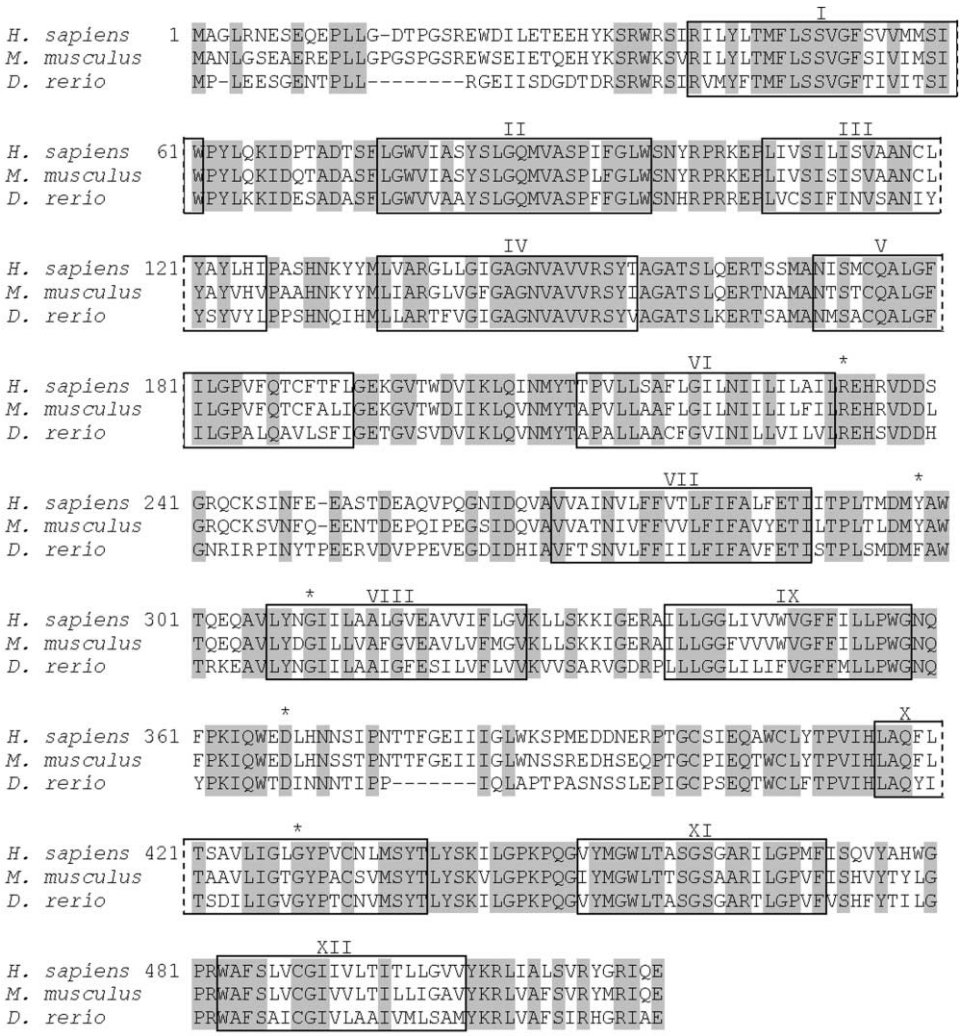

Figure 2. Conservation of the MFSD8 amino acid sequence among human (*Homo sapiens*), mouse (*Mus musculus*), and zebrafish (*Danio rerio*). Amino acid sequences were aligned using MAFFT version 5.8. The amino acid numbering is indicated for the human polypeptide. Identical amino acids are indicated with gray shading. Sites of exonic disease-causing mutations are marked with an asterisk (*): p.Arg233Gly/splice, p.Tyr298X, p.Gly310Asp, p.Asp368His/splice, and p.Gly429Asp. The predicted 12 transmembrane domains are indicated with boxes and Roman numerals.

the identification of this gene was mainly based on Turkish families with vLINCL, we denoted the linked locus as *CLN7,* the long-sought locus for Turkish vLINCL. Given that all families except one were consanguineous, we assumed that the chromosomal region harboring the disease gene would be identical by decent in most families. We therefore used a SNP-based homozygosity mapping approach[28] with 10 families, to localize the disease gene. In line with our assumption, we identified a region on chromosome 4 with statistically significant linkage, with five families showing large overlapping regions of homozygosity, reflecting identity by descent. However, the haplotypes were different between families, suggesting the occurrence of a private mutation in each family.

Consequently, we identified six nucleotide changes in *MFSD8,* one in each of the six families. These are likely to be disease-causing, because they altered the predicted amino acid composition of the MFSD8 protein. It is not known whether the only nonsense mutation identified (p.Tyr298X) results in a truncated protein that is nonfunctional or leads to degradation of the mRNA through nonsense-mediated decay.[29] The two identified missense mutations (p.Gly310Asp and p.Gly429Asp) change the highly conserved glycines that reside in the predicted 8th and 10th transmembrane domains, respectively, to aspartic acids (figs. 2 and 4). These changes do not, however, alter the localization of the protein, as shown by immunofluorescence staining of the overexpressed mutant HA-tagged proteins. This suggests that the primary defect of these mutants is more likely the result of the functional properties of the mutated protein than the result of altered localization. Two of the mutations (c.697A→G and

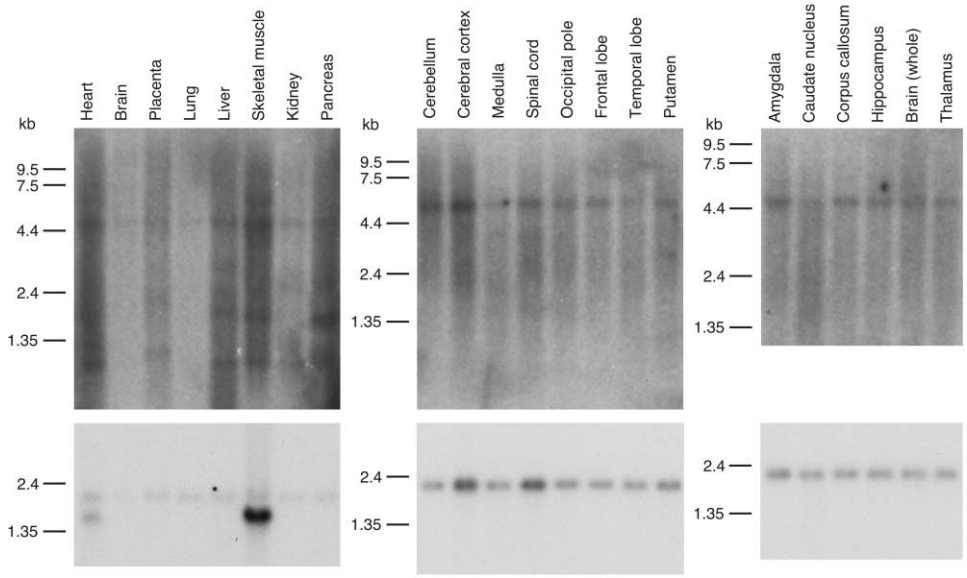

Figure 3. Northern-blot analysis of *MFSD8* expression. An ~5-kb *MFSD8* transcript (*upper panels*) was detected in all tissues analyzed. In all brain regions and in lung, it is the only transcript seen. In other tissues, transcripts that ranged in size from ~1 to ~3 kb and/or ~6 kb were also present. *Lower panels,* Hybridization with human β-actin cDNA as a control for RNA loading. The positions of size markers are shown on the left of each blot.

c.1102G→C) may change highly conserved amino acids (p.Arg233Gly and p.Asp368His, respectively) and may result in changes in the protein structure and/or function. Alternatively, they may affect the splicing of the transcript, because they change sequences at the exon-intron junctions in the 3′ ends of exons 7 and 11 and thus may lead to the production of abnormal mRNAs and/or proteins.[30] Because of the unavailability of RNA samples from patients with these mutations, we were not able to test the outcome of these changes at the mRNA level. However, RNA was available from the patient with the intronic c.754+2T→A mutation that affects the donor-splice site of intron 8. In RT-PCR analysis, we showed that the mutation results in an altered splicing pattern. This is expected, since a nucleotide change at the nearly invariant second thymine at the 5′ end of an intron usually leads to inaccurate recognition of exon-intron boundaries and to disturbed splicing.[30]

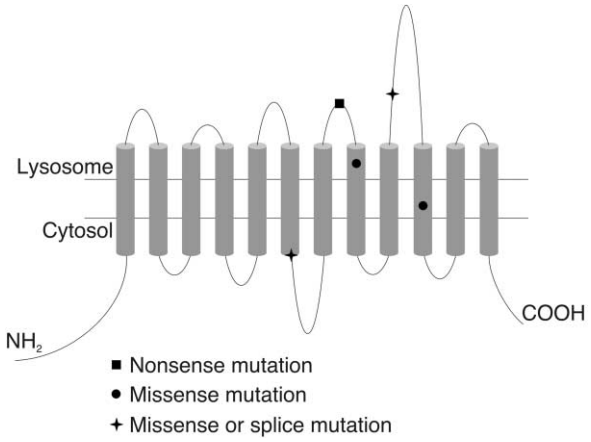

Figure 4. Schematic structure of the MFSD8 protein, with positions of patient mutations that affect exonic sequences. MFSD8 is predicted to contain 12 transmembrane domains, according to the prediction programs HMMTOP, TMHMM, and TMpred. The site of the nonsense mutation (p.Tyr298X) is indicated with a square, the site of the missense mutations (p.Gly310Asp and p.Gly429Asp) with circles, and the site of mutations that either change amino acids or affect splicing (p.Arg233Gly/splice and p.Asp368His/splice) with crosses.

MFSD8 is ubiquitously expressed, with the main transcript of ~5 kb being the only one present in the brain. In addition, the EST and RT-PCR data suggest that *MFSD8* shows a complex pattern of alternative splicing. The functional relevance of the alternative transcripts seen in northern-blot analysis and the less abundant transcripts, represented only by one or few ESTs, is not known, but some of them may reflect a posttranscriptional mechanism for regulation of *MFSD8* expression.[31] MFSD8 is predicted to be a polytopic integral membrane protein with 12 membrane-spanning domains (fig. 4). In immunofluorescence analysis, MFSD8 was shown to colocalize with lysosomal markers, implying that it is a novel integral lysosomal membrane protein. The majority of the previously identified NCL proteins show lysosomal localization—*CTSD, PPT1,* and *TPP1* encode soluble lysosomal enzymes,[4,5,32–35] whereas *CLN3* and *CLN5* code for lyso-

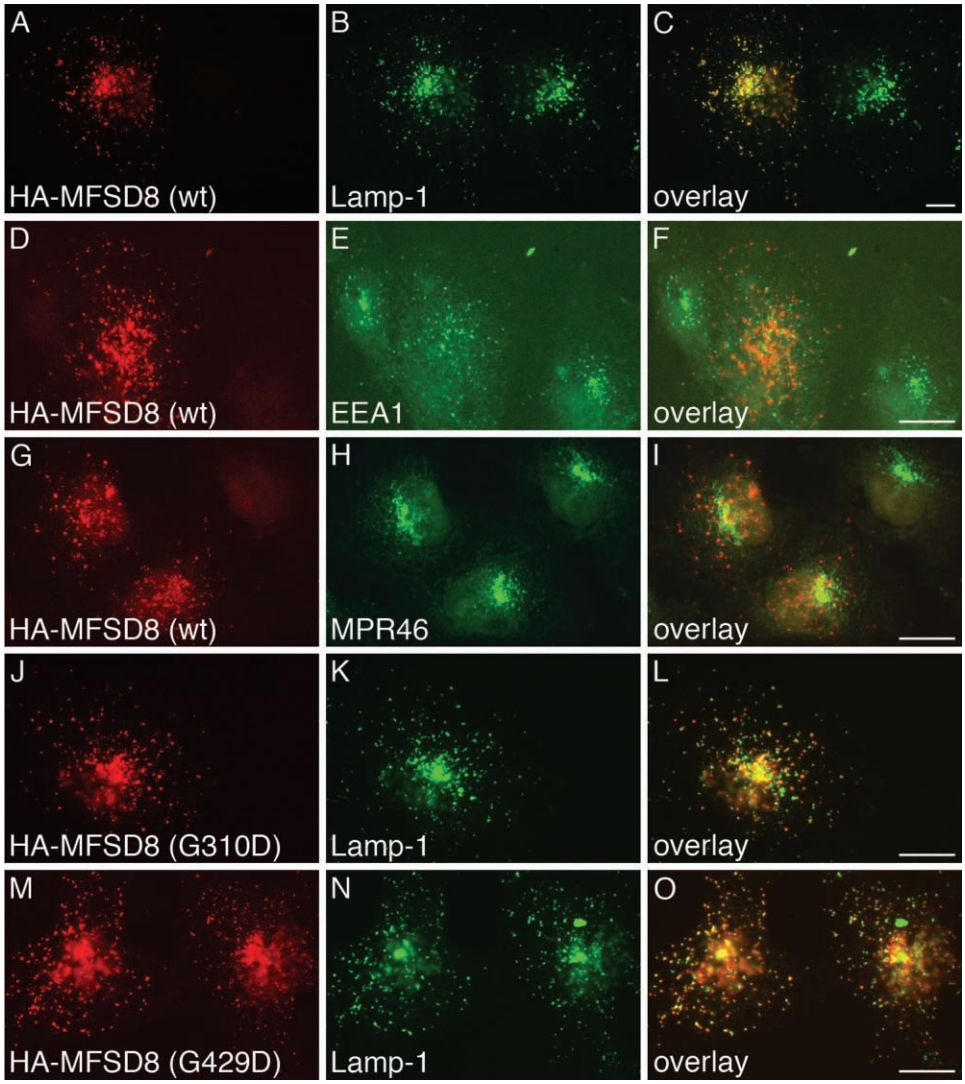

Figure 5. Lysosomal localization of both wild-type and mutant MFSD8 proteins. Subcellular localization of HA-tagged wild-type (wt) and mutant (G310D and G429D) MFSD8 was seen in transiently transfected COS-1 cells. *A, D,* and *G,* Distribution of the HA-MFSD8 protein by use of an antibody against HA (*red*). *B,* Lysosomal staining pattern by use of an anti–Lamp-1 antibody (*green*). *C,* Overlay of the HA-MFSD8 and Lamp-1 stainings. *E,* Early endosomal staining pattern by use of an anti-EEA1 antibody (*green*). *F,* Overlay of the HA-MFSD8 and EEA1 stainings. *H,* Trans-Golgi-network staining pattern by use of an anti-MPR46 antibody (*green*). *I,* Overlay of the HA-MFSD8 and MPR46 stainings. *J* and *M,* Distribution of the G310D mutant (*J*) and the G429D mutant (*M*) HA-MFSD8 proteins by use of an antibody against HA (*red*). *K* and *N,* Lysosomal staining pattern by use of an anti–Lamp-1 antibody (*green*). *L* and *O,* Overlays of the mutant HA-MFSD8 and Lamp-1 stainings. Yellow indicates overlap of wild-type or mutant HA-MFSD8 and the subcellular markers. Scale bars (10 μm) are shown in panels C, F, I, L, and O.

somal proteins of unknown function.[3,6,36,37] On the contrary, two NCL proteins localize either exclusively to the ER (CLN6)[38,39] or to both the ER and the ER-Golgi intermediate compartment (CLN8).[40]

On the basis of homology analyses, MFSD8 belongs to the MFS of proteins. The MFS is a very large family of proteins present ubiquitously in all classes of organisms.[41] These proteins are single-polypeptide carriers that are able to transport small solutes by using chemiosmotic ion gradients.[41] Many of these proteins span the membrane 12 times, as does MFSD8. The MFS proteins are classified into several phylogenetic subfamilies, and this classification correlates with the function, each subfamily transporting a specific class of substrates, including sugars, drugs, inorganic and organic cations, various metabolites, etc.[41] In humans, the most similar proteins to MFSD8 include MFS domain containing 9 (MFSD9 or MGC11332) and tetracycline transporter–like protein (TETRAN), which are predicted to have tetracycline-transporter activities; solute carrier family 18 member 2 (SLC18A2, also denoted "VMAT2" [MIM 193001]), which translocates monoamines from cytosol to synaptic vesicles,[42] and SLC22A18

(also denoted "ORCTL2" [MIM 602631]), which acts in transport of organic cations.[43] MFSD8 is conserved in vertebrates with a single identifiable orthologue, but similar proteins are identified also in invertebrates. In *S. cerevisiae,* the most similar proteins to MFSD8 are Jen1p, Git1p, Ykr105cp, and Ycr023cp, which all belong to the MFS. Jen1p is a functional lactate transporter,[44,45] whereas Git1p mediates uptake of glycerophosphoinositol and glycerophosphocholine.[46,47] In *D. melanogaster,* the most similar proteins to MFSD8 are CG8596-PA, which has an unknown function, and a tetracycline-resistance protein CG5760-PA, both belonging to the MFS (FlyBase). In *C. elegans,* there are several proteins similar to MFSD8, most of which belong to the MFS and many of which have protein domains involved in transport processes (WormBase). Being a member of the MFS, MFSD8 is likely to function as a transporter. However, the substrate specificity of MFSD8 remains to be elucidated.

When first described, Turkish vLINCL was considered a distinct clinical and genetic entity.[13] However, it is now evident that, even if the clinical phenotype is quite uniform, Turkish vLINCL is genetically very heterogeneous, with mutations identified already in three genes, *CLN6,*[20] *CLN8,*[16] and *MFSD8*. In the remaining five Turkish families—for which we excluded the currently known NCL genes, including *CLCN6* (MIM 602726) and *CTSF* (MIM 603539), homologues of which were recently reported to cause NCL-like phenotypes in mice[48,49]—the SNP-scan data imply the presence of at least three more underlying genes. This further corroborates the great heterogeneity of Turkish vLINCL. Identification and characterization of the *MFSD8* gene and further dissection of the genetic background of Turkish vLINCL will be important steps toward complete understanding of the genetic spectrum of the NCLs and, additionally, toward understanding their underlying molecular mechanisms.

Acknowledgments

We thank the families with NCL, for their participation, and Drs. Christine Tranchant and Füsun Alehan, for data on patients k3 and l3, respectively. We also thank Paula Hakala, Mairi Kuris, Mervi Kuronen, Ahmed Mohamed, Hanna Olanne, Eva Reinmaa, and Teija-Tuulia Toivonen, for technical assistance; Jaana Tyynelä, for helpful advice; and Anu Jalanko, for marker antibodies. This study was supported by the Center of Excellence in Complex Disease Genetics of the Academy of Finland, the Folkhälsan Research Foundation, the Sigrid Juselius Foundation, and the European Commission (project LSHM-CT-2003-50305), to A.-E.L., and Genome Canada through the Ontario Genomics Institute, Sigrid Juselius, and Emil Aaltonen Foundations (to H.L.). E.S. and A.-K.A. are fellows of the Helsinki Biomedical Graduate School. A.D.P. holds a Canada Research Chair in the Genetics of Complex Diseases. B.A.M. holds a Canada Research Chair in Pediatric Neurogenetics.

Web Resources

The accession number and URLs for data presented herein are as follows:

BLAST, http://www.ncbi.nlm.nih.gov/BLAST/ (for NCBI protein-protein BLAST)
FlyBase, http://flybase.bio.indiana.edu/
GenBank, http://www.ncbi.nlm.nih.gov/Genbank/ (for *MFSD8* [accession number NM_152778])
HMMTOP, http://www.enzim.hu/hmmtop/
MAFFT version 5.8, http://align.bmr.kyushu-u.ac.jp/mafft/online/server/
Mammalian Genotyping Service, http://research.marshfieldclinic.org/genetics/
NCBI AceView, http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/
NCBI Map Viewer, http://www.ncbi.nlm.nih.gov/mapview/ (for build 35.1)
Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/ (for *PPT1/CLN1, TPP1/CLN2, CLN3, CLN5, CLN6, CLN8, CTSD/CLN10,* CLN4, CLN9, CLN2, *CLCN3, CLCN7, MECP2, TRPC3,* VMAT2, ORCTL2, *CLCN6,* and *CTSF*)
Pfam, http://pfam.janelia.org/
TMHMM, http://www.cbs.dtu.dk/services/TMHMM-2.0/
TMpred, http://www.ch.embnet.org/software/TMPRED_form.html
WormBase, http://www.wormbase.org/

References

1. Santavuori P (1988) Neuronal ceroid-lipofuscinoses in childhood. Brain Dev 10:80–83
2. Haltia M (2003) The neuronal ceroid-lipofuscinoses. J Neuropathol Exp Neurol 62:1–13
3. The International Batten Disease Consortium (1995) Isolation of a novel gene underlying Batten disease, CLN3. Cell 82:949–957
4. Vesa J, Hellsten E, Verkruyse LA, Camp LA, Rapola J, Santavuori P, Hofmann SL, Peltonen L (1995) Mutations in the palmitoyl protein thioesterase gene causing infantile neuronal ceroid lipofuscinosis. Nature 376:584–587
5. 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 277:1802–1805
6. Savukoski M, Klockars T, Holmberg V, Santavuori P, Lander ES, Peltonen L (1998) *CLN5,* a novel gene encoding a putative transmembrane protein mutated in Finnish variant late infantile neuronal ceroid lipofuscinosis. Nat Genet 19:286–288
7. Ranta S, Zhang Y, Ross B, Lonka L, Takkunen E, Messer A, Sharp J, Wheeler R, Kusumi K, Mole S, et al (1999) The neuronal ceroid lipofuscinoses in human EPMR and *mnd* mutant mice are associated with mutations in *CLN8*. Nat Genet 23: 233–236
8. Gao H, Boustany RM, Espinola JA, Cotman SL, Srinidhi L, Antonellis KA, Gillis T, Qin X, Liu S, Donahue LR, et al (2002) Mutations in a novel CLN6-encoded transmembrane protein cause variant neuronal ceroid lipofuscinosis in man and mouse. Am J Hum Genet 70:324–335
9. Wheeler RB, Sharp JD, Schultz RA, Joslin JM, Williams RE, Mole SE (2002) The gene mutated in variant late-infantile neuronal ceroid lipofuscinosis (*CLN6*) and in *nclf* mutant mice encodes a novel predicted transmembrane protein. Am J Hum Genet 70:537–542
10. Siintola E, Partanen S, Stromme P, Haapanen A, Haltia M, Maehlen J, Lehesjoki AE, Tyynela J (2006) Cathepsin D de-

ficiency underlies congenital human neuronal ceroid-lipofuscinosis. Brain 129:1438–1445
11. Steinfeld R, Reinhardt K, Schreiber K, Hillebrand M, Kraetzner R, Bruck W, Saftig P, Gartner J (2006) Cathepsin D deficiency is associated with a human neurodegenerative disorder. Am J Hum Genet 78:988–998
12. Berkovic SF, Carpenter S, Andermann F, Andermann E, Wolfe LS (1988) Kufs' disease: a critical reappraisal. Brain 111:27–62
13. Wheeler RB, Sharp JD, Mitchell WA, Bate SL, Williams RE, Lake BD, Gardiner RM (1999) A new locus for variant late infantile neuronal ceroid lipofuscinosis—CLN7. Mol Genet Metab 66:337–338
14. Schulz A, Dhar S, Rylova S, Dbaibo G, Alroy J, Hagel C, Artacho I, Kohlschutter A, Lin S, Boustany RM (2004) Impaired cell adhesion and apoptosis in a novel CLN9 Batten disease variant. Ann Neurol 56:342–350
15. Das AK, Becerra CH, Yi W, Lu JY, Siakotos AN, Wisniewski KE, Hofmann SL (1998) Molecular genetics of palmitoyl-protein thioesterase deficiency in the U.S. J Clin Invest 102:361–370
16. Ranta S, Topcu M, Tegelberg S, Tan H, Ustubutun A, Saatci I, Dufke A, Enders H, Pohl K, Alembik Y, et al (2004) Variant late infantile neuronal ceroid lipofuscinosis in a subset of Turkish patients is allelic to Northern epilepsy. Hum Mutat 23:300–305
17. Cannelli N, Cassandrini D, Bertini E, Striano P, Fusco L, Gaggero R, Specchio N, Biancheri R, Vigevano F, Bruno C, et al (2006) Novel mutations in CLN8 in Italian variant late infantile neuronal ceroid lipofuscinosis: another genetic hit in the Mediterranean. Neurogenetics 7:111–117
18. Topcu M, Tan H, Yalnizoglu D, Usubutun A, Saatci I, Aynaci M, Anlar B, Topaloglu H, Turanli G, Kose G, et al (2004) Evaluation of 36 patients from Turkey with neuronal ceroid lipofuscinosis: clinical, neurophysiological, neuroradiological and histopathologic studies. Turk J Pediatr 46:1–10
19. Williams RE, Topcu M, Lake BD, Mitchell W, Mole SE (1999) CLN7 Turkish variant late infantile NCL. In: Goebel HH, Mole SE, Lake BD (eds) The neuronal ceroid lipofuscinosis (Batten disease). IOS Press, Amsterdam, pp 114–116
20. Siintola E, Topcu M, Kohlschutter A, Salonen T, Joensuu T, Anttonen AK, Lehesjoki AE (2005) Two novel CLN6 mutations in variant late-infantile neuronal ceroid lipofuscinosis patients of Turkish origin. Clin Genet 68:167–173
21. Stobrawa SM, Breiderhoff T, Takamori S, Engel D, Schweizer M, Zdebik AA, Bosl MR, Ruether K, Jahn H, Draguhn A, et al (2001) Disruption of ClC-3, a chloride channel expressed on synaptic vesicles, leads to a loss of the hippocampus. Neuron 29:185–196
22. Yoshikawa M, Uchida S, Ezaki J, Rai T, Hayama A, Kobayashi K, Kida Y, Noda M, Koike M, Uchiyama Y, et al (2002) CLC-3 deficiency leads to phenotypes similar to human neuronal ceroid lipofuscinosis. Genes Cells 7:597–605
23. Kasper D, Planells-Cases R, Fuhrmann JC, Scheel O, Zeitz O, Ruether K, Schmitt A, Poet M, Steinfeld R, Schweizer M, et al (2005) Loss of the chloride channel ClC-7 leads to lysosomal storage disease and neurodegeneration. EMBO J 24: 1079–1091
24. Kong X, Murphy K, Raj T, He C, White PS, Matise TC (2004) A combined linkage-physical map of the human genome. Am J Hum Genet 75:1143–1148
25. Huang Q, Shete S, Amos CI (2004) Ignoring linkage disequilibrium among tightly linked markers induces false-positive evidence of linkage for affected sib pair analysis. Am J Hum Genet 75:1106–1112
26. Abecasis GR, Cherny SS, Cookson WO, Cardon LR (2002) Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet 30:97–101
27. Lennon G, Auffray C, Polymeropoulos M, Soares MB (1996) The I.M.A.G.E. Consortium: an integrated molecular analysis of genomes and their expression. Genomics 33:151–152
28. Lander ES, Botstein D (1987) Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. Science 236:1567–1570
29. Hentze MW, Kulozik AE (1999) A perfect message: RNA surveillance and nonsense-mediated decay. Cell 96:307–310
30. Cartegni L, Chew SL, Krainer AR (2002) Listening to silence and understanding nonsense: exonic mutations that affect splicing. Nat Rev Genet 3:285–298
31. Blencowe BJ (2006) Alternative splicing: new insights from global analyses. Cell 126:37–47
32. Press EM, Porter RR, Cebra J (1960) The isolation and properties of a proteolytic enzyme, cathepsin D, from bovine spleen. Biochem J 74:501–514
33. Rawlings ND, Barrett AJ (1995) Families of aspartic peptidases, and those of unknown catalytic mechanism. Methods Enzymol 248:105–120
34. Hellsten E, Vesa J, Olkkonen VM, Jalanko A, Peltonen L (1996) Human palmitoyl protein thioesterase: evidence for lysosomal targeting of the enzyme and disturbed cellular routing in infantile neuronal ceroid lipofuscinosis. EMBO J 15:5240–5245
35. Verkruyse LA, Hofmann SL (1996) Lysosomal targeting of palmitoyl-protein thioesterase. J Biol Chem 271:15831–15836
36. Jarvela I, Sainio M, Rantamaki T, Olkkonen VM, Carpen O, Peltonen L, Jalanko A (1998) Biosynthesis and intracellular targeting of the CLN3 protein defective in Batten disease. Hum Mol Genet 7:85–90
37. Isosomppi J, Vesa J, Jalanko A, Peltonen L (2002) Lysosomal localization of the neuronal ceroid lipofuscinosis CLN5 protein. Hum Mol Genet 11:885–891
38. Heine C, Koch B, Storch S, Kohlschutter A, Palmer DN, Braulke T (2004) Defective endoplasmic reticulum-resident membrane protein CLN6 affects lysosomal degradation of endocytosed arylsulfatase A. J Biol Chem 279:22347–22352
39. Mole SE, Michaux G, Codlin S, Wheeler RB, Sharp JD, Cutler DF (2004) CLN6, which is associated with a lysosomal storage disease, is an endoplasmic reticulum protein. Exp Cell Res 298:399–406
40. Lonka L, Kyttala A, Ranta S, Jalanko A, Lehesjoki AE (2000) The neuronal ceroid lipofuscinosis CLN8 membrane protein is a resident of the endoplasmic reticulum. Hum Mol Genet 9:1691–1697
41. Pao SS, Paulsen IT, Saier MH Jr (1998) Major facilitator superfamily. Microbiol Mol Biol Rev 62:1–34
42. Liu Y, Edwards RH (1997) The role of vesicular transport proteins in synaptic transmission and neural degeneration. Annu Rev Neurosci 20:125–156
43. Reece M, Prawitt D, Landers J, Kast C, Gros P, Housman D, Zabel BU, Pelletier J (1998) Functional characterization of ORCTL2—an organic cation transporter expressed in the renal proximal tubules. FEBS Lett 433:245–250
44. Casal M, Paiva S, Andrade RP, Gancedo C, Leao C (1999) The

lactate-proton symport of *Saccharomyces cerevisiae* is encoded by *JEN1*. J Bacteriol 181:2620–2623
45. Soares-Silva I, Schuller D, Andrade RP, Baltazar F, Cassio F, Casal M (2003) Functional expression of the lactate permease Jen1p of *Saccharomyces cerevisiae* in *Pichia pastoris*. Biochem J 376:781–787
46. Patton-Vogt JL, Henry SA (1998) *GIT1,* a gene encoding a novel transporter for glycerophosphoinositol in *Saccharomyces cerevisiae*. Genetics 149:1707–1715
47. Fisher E, Almaguer C, Holic R, Griac P, Patton-Vogt J (2005) Glycerophosphocholine-dependent growth requires Gde1p (YPL110c) and Git1p in *Saccharomyces cerevisiae*. J Biol Chem 280:36110–36117
48. Poet M, Kornak U, Schweizer M, Zdebik AA, Scheel O, Hoelter S, Wurst W, Schmitt A, Fuhrmann JC, Planells-Cases R, et al (2006) Lysosomal storage disease upon disruption of the neuronal chloride transport protein ClC-6. Proc Natl Acad Sci USA 103:13854–13859
49. Tang CH, Lee JW, Galvez MG, Robillard L, Mole SE, Chapman HA (2006) Murine cathepsin F deficiency causes neuronal lipofuscinosis and late-onset neurological disease. Mol Cell Biol 26:2309–2316