# DISSECTION OF THE GENETIC BACKGROUND OF CHILDHOOD ONSET PROGRESSIVE MYOCLONIC EPILEPSIES

**Maria Kousi**

ACADEMIC DISSERTATION


*To be publicly discussed, with the permission of the Faculty of Medicine of the*

*University of Helsinki in Biomedicum Helsinki Lecture Hall 1, on Friday January 13th, at 12 noon*

Neuroscience Center and

Department of Medical Genetics,

University of Helsinki, and

Folkhälsan Institute of Genetics

Helsinki, 2011

***Supervised by***

Professor Anna-Elina Lehesjoki, M.D., Ph.D.
Folkhälsan Institute of Genetics and
Neuroscience Center
University of Helsinki
Helsinki, Finland

***Reviewed by***

Professor Mark Gardiner, MD, FRCPCH
Department of Paediatrics and Child Health
Royal Free and University College Medical School
University College London
London, UK

Docent Marjo Kestilä, Ph.D.
Department of Chronic Disease Prevention
Public Health Genomics
National Institute for Health and Welfare
Helsinki, Finland

***Official opponent***

Professor James R. Lupski, M.D., Ph.D.
Department of Molecular and Human Genetics
Baylor College of Medicine
Houston, TX, USA

ISBN 978-952-10-7560-5 (Paperback)

ISBN 978-952-10-7561-2 (PDF)

http://ethesis.helsinki.fi

Unigrafia Oy

Helsinki 2011

"The genes: our heritability, are the paper and pen we are born with.

The story we will write is up to us."

***Leena Peltonen-Palotie***

To my parents

# TABLE OF CONTENTS

# LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following original articles, which are referred to in the text by their Roman numerals. In addition, unpublished data are also presented.

I **Kousi M**, Lehesjoki AE, Mole SE. Update of the mutation spectrum and clinical correlations of over 360 mutations in eight genes that underlie the neuronal ceroid lipofuscinoses. ***Hum Mut*** doi: 10.1002/humu.21624 in press.

II **Kousi M**, Siintola E, Dvorakova L, Vlaskova H, Turnbull J, Topcu M, Yuksel D, Gokben S, Minassian B, Elleder M, Mole SE, Lehesjoki AE. Mutations in *CLN7/MFSD8* is a common cause of variant late infantile neuronal ceroid lipofuscinosis. ***Brain*** 132(Pt 3): 810-9; 2009

III Sharifi A*****, **Kousi M***, Sagné C, Bellenchi GC, Morel L, Darmon M, Hulkova H, Ruivo R, Debacker C, El Mestikawy S, Elleder M, Lehesjoki AE, Jalanko A, Gasnier B, Kyttälä A. Expression and lysosomal targeting of CLN7, a major facilitator superfamily transporter associated with variant late infantile neuronal ceroid lipofuscinosis. ***Hum Mol Genet*** 2010; 19(22):4497-514.

IV **Kousi M,** Anttila V, Schulz A, Calafato S, Coffey AJ, Jakkula E, Myllykangas L, Kalimo H, Topcu M, Gokben S, Yuksel D, Alehan F, Palotie A, Kopra O, Lehesjoki AE. Novel *KCTD7* mutations establish association to progressive myoclonus epilepsy. ***Submitted***

***** These authors contributed equally to this work.

Publication III has appeared previously in the doctoral thesis of Azita Sharifi (2010).

# ABBREVIATIONS

| | |
|---|---|
| aa | Amino acid |
| ANCL | Adult Neuronal Ceroid Lipofuscinosis |
| AP | Adaptor Protein |
| bp | Base Pair |
| cDNA | Complementary DNA |
| CLN1-10 | Ceroid Lipofuscinoses, Neuronal 1-10 disease |
| *CLN3, 5, 6, 8* | Ceroid Lipofuscinoses, Neuronal 3, 5, 6, 8 genes |
| cM | CentiMorgan |
| CL | Curvilinear |
| CNV | Copy Number Variation |
| CONCL | Congenital Neuronal Ceroid Lipofuscinosis |
| *CTSD* | Cathepsin D gene |
| DNA | Deoxyribonucleic Acid |
| EM | Electron Microscopy |
| EPM1-6 | Epilepsy, Progressive Myoclonic type 1-6 |
| EPMR | Progressive Epilepsy with Mental Retardation |
| ER | Endoplasmic Reticulum |
| ERGIC | ER-Golgi Intermediate Compartment |
| FCS | Fetal Calf Serum |
| FP | Fingerprint Profiles |
| GROD | Granular Osmiophilic Deposits |
| HA | Hemagglutinin |
| HGP | Human Genome Project |
| IBD | Identical By Descent |
| IBS | Identical By State |
| IF | Immunofluorescence |
| INAD | Infantile Neuroaxonal Dystrophy |

| | |
|---|---|
| INCL | Infantile Neuronal Ceroid Lipofuscinosis |
| Indel | Insertion/Deletion |
| JNCL | Juvenile Neuronal Ceroid Lipofuscinosis |
| kb | Kilobase |
| *KCTD7* | Potassium Channel Tetramerization Domain-Containing 7 gene |
| kDa | KiloDalton |
| LD | Linkage Disequilibrium |
| LINCL | Late Infantile Neuronal Ceroid Lipofuscinosis |
| Man6P | Mannose-6-Phosphate |
| Mb | Megabase |
| MFS | Major Facilitator Superfamily |
| *MFSD8* | Major Facilitator Superfamily Domain-containing 8 gene |
| MIM | Mendelian Inheritance in Man |
| MPR | Man6P Receptor |
| MRI | Magnetic Resonance Imaging |
| mRNA | Messenger RNA |
| NBIA | Neurodegeneration with Brain Iron Accumulation |
| NCBI | National Center for Biotechnology Information |
| NCL | Neuronal Ceroid Lipofuscinosis |
| OMIM | Online Mendelian Inheritance in Man |
| PBS | Phosphate-Buffered Saline |
| PCR | Polymerase Chain Reaction |
| PFA | Paraformaldehyde |
| *PLA2G6* | Phospholipase A2 Group VI |
| PM | Plasma Membrane |
| PME | Progressive Myoclonic Epilepsy |
| *PPT1* | Palmitoyl Protein Thioesterase 1 gene |
| RL | Rectilinear |
| RNA | Ribonucleic Acid |
| RT | Room Temperature |
| RT-PCR | Reverse Transcriptase-PCR |

| | |
|---|---|
| SAP | Sphingolipid Activator Protein |
| SCMAS | Subunit c of Mitochondrial ATP Synthase |
| SNP | Single Nucleotide Polymorphism |
| TGN | Trans-Golgi Network |
| TLC | TRAM-LAG1-CLN8 domain |
| TM | Transmembrane domain |
| *TPP1* | Tripeptidyl Peptidase I gene |
| *TXNDC6* | Thioredoxin Domain-Containing 6 gene |
| *USP19* | Ubiquitin Specific Peptidase 19 gene |

Only abbreviations that appear more than once and in more than one chapter are presented in this list.

# ABSTRACT

The progressive myoclonic epilepsies (PMEs) are a clinically and etiologically heterogeneous group of symptomatic epilepsies characterized by myoclonus, tonic-clonic seizures, psychomotor regression and ataxia. Different disorders have been classified as PMEs. Of these, the group of neuronal ceroid lipofuscinoses (NCLs) comprise an entity that has onset in childhood, being the most common cause of neurodegeneration in children. The primary aim of this thesis was to dissect the molecular genetic background of patients with childhood onset PME by studying candidate genes and attempting to identify novel PME-associated genes. Another specific aim was to study the primary protein properties of the most recently identified member of the NCL-causing proteins, MFSD8.

To dissect the genetic background of a cohort of Turkish patients with childhood onset PME, a screen of the NCL-associated genes *PPT1*, *TPP1*, *CLN3*, *CLN5*, *CLN6*, *MFSD8*, *CLN8* and *CTSD* was performed. Altogether 49 novel mutations were identified, which together with 56 mutations found by collaborators raised the total number of known NCL mutations to 364.

Fourteen of the novel mutations affect the recently identified *MFSD8* gene, which had originally been identified in a subset of mainly Turkish patients as the underlying cause of CLN7 disease. To investigate the distribution of *MFSD8* defects, a total of 211 patients of different ethnic origins were evaluated for mutations in the gene. Altogether 45 patients from nine different countries were provided with a CLN7 molecular diagnosis, denoting the wide geographical occurrence of *MFSD8* defects. The mutations are private with only one having been established by a founder-effect in the Roma population from the former Czechoslovakia. All mutations identified except one are associated with the typical clinical picture of variant late-infantile NCL.

To address the trafficking properties of MFSD8, lysosomal targeting of the protein was confirmed in both neuronal and non-neuronal cells. The major determinant for this lysosomal sorting was identified to be an N-terminal dileucine based signal (9-EQEPLL-14), recognized by heterotetrameric AP-1 adaptor proteins, suggesting that MFSD8 takes the direct trafficking pathway *en route* to the lysosomes. Expression studies revealed the neurons as the primary cell-type and the hippocampus and cerebellar granular cell layer as the predominant regions in which *MFSD8* is expressed.

To identify novel genes associated with childhood onset PME, a single nucleotide polymorphism (SNP) genomewide scan was performed in three small families and 18 sporadic patients followed by homozygosity mapping to determine the candidate loci. One of the families and a sporadic patient were positive for mutations in *PLA2G6*, a gene that had previously been shown to cause infantile neuroaxonal dystrophy. Application of next-generation sequencing of candidate regions in the remaining two families led to identification of a homozygous missense mutation in *USP19* for the first and *TXNDC6* for the second family. Analysis of the 18 sporadic cases mapped the best candidate interval in a 1.5 Mb region on chromosome 7q21. Screening of the positional candidate *KCTD7* revealed six mutations in seven unrelated families. All patients with mutations in *KCTD7* were reported to have early onset PME, rapid disease progression leading to dementia and no pathologic hallmarks. The identification of *KCTD7* mutations in nine patients and the clinical delineation of their phenotype establish *KCTD7* as a gene for early onset PME.

The findings presented in this thesis denote *MFSD8* and *KCTD7* as genes commonly associated with childhood onset symptomatic epilepsy. The disease-associated role of *TXNDC6* awaits verification through identification of additional mutations in patients with similar phenotypes. Completion of the genetic spectrum underlying childhood onset PMEs and understanding of the gene products' functions will comprise important steps towards understanding the underlying pathogenetic mechanisms, and will possibly shed light on the general processes of neurodegeneration and nervous system regulation, facilitating the diagnosis, classification and possibly treatment of the affected cases.

# 1. INTRODUCTION

Major advances have been made in the field of genetics during the past two decades. The landmark completion of the Human Genome Project published only a decade ago has paved the way to the design and use of revolutionary methods such as genomewide genotyping (Lander et al., 2001; Venter et al., 2001). In the 1990s identifying genes for Mendelian conditions required the recruitment of large pedigrees with rare and preferably highly penetrant Mendelian phenotypes for dominant conditions and large collections of informative families in recessively transmitted disorders. This is exemplified by the identification of *CHRNA4* as the first epilepsy-associated gene in a six-generation Australian kindred with 27 affected family members (Steinlein et al., 1995; Phillips et al., 1995). Today when whole exome/genome sequencing is possible having a large patient collection is no longer a prerequisite to a successful genetic study. Instead, as demonstrated in a study showing that mitochondrial alanyl-tRNA synthetase (mtAlaRS) is defective in hypertrophic mitochondrial cardiomyopathy, a single patient sample could suffice for initial gene discovery (Götz et al., 2011).

The advent of novel technologies has made possible the genotyping of the whole exome and/or genome of large sets of individuals producing an incredible amount of data. This facilitated the studies addressing the genetic basis of more common diseases and traits such as diabetes and height. Nevertheless, the initial confidence that whole genome data would speed up our understanding of the genetics of complex traits has not been confirmed. Moreover, only a small fraction of the identified genetic variation has been linked to function. Our understanding of the meaning of the enormous variation identified, how this relates to our history, biology and health remain scarce to date.

Given these limitations some researchers have proposed that the field of genetics should turn its focus again to the study of monogenic and rare phenotypes, utilizing previously poorly studied isolated populations, offspring of consanguineous marriages, or even very rare phenotypes encountered in very few sporadic cases (Antonarakis and Beckmann,

2006). Elucidation of the disease-associated mechanisms in such rare disorders could contribute in understanding the more complex polygenic disease phenotypes.

The PMEs, a heterogeneous group of symptomatic generalized epilepsies, comprise an example of monogenic disorders for which the genetic and functional backgrounds await further elucidation. Among the six clinical entities constituting the PMEs (reviewed in Ramachandran et al., 2009) the NCLs comprise the clinical entity that most commonly underlies neurodegeneration in children. Despite the description of ten clinical subtypes and nine NCL-causing genes (Noskova et al., 2011; reviewed in Mole et al., 2005), a significant number of patients exist for whom no genetic or molecular explanation of the disease could be provided. By recruiting inbred families and sporadic cases and applying next-generation sequencing it should be possible to identify the remaining NCL-associated genes in the near future. When more players participating in the affected pathways of such Mendelian disorders have been identified, the picture of more complex non-monogenic symptomatic epilepsies and other neurodegenerative disorders might start becoming clearer.

# 2. REVIEW OF THE LITERATURE

## 2.1 From chromosome discovery to the whole genome sequencing era

The fact that the individuals belonging to a species resemble their progenitors, their offspring, and each other suggests that the information dictating this resemblance is passed on from generation to generation. The first referrals of heredity date as far back as the 5th century BC with Hippocrates commenting on the sacred disease, epilepsy: "Its origin is hereditary, like that of other diseases; for if a phlegmatic person be born of a phlegmatic, and a bilious of a bilious, and a phthisical of a phthisical, and one having spleen disease of another having disease of the spleen, what is it to hinder it from happening that where the father and mother were subject to this disease certain of their offspring should be so affected too?". A strong emphasis was placed on inheritance by Charles Darwin, who argued that the passing on of the characteristics from the parents to the offsping in different combinations provides the genetic variability from which natural selection will determine who will thrive and who will perish. It was not, however, until 1865 that the inheritance of small discrete units was officially established by Gregor Mendel, who carried out his breeding experiments in pea plants, inaugurating the field of genetics.

As mentioned in the opening paragraph of the paper presenting the first draft of the human genome sequence, major discoveries have been made in the field of genetics in roughly each of the four quarters of the 20th century (Lander et al., 2001). Although the correct number of human chromosomes was not established until 1956 (Tjio and Levan, 1956), the first report that the chromosomes and the genes that reside on them comprise the units mediating heredity came by Thomas Hunt Morgan (Morgan et al., 1911). In the second quarter of the 20th century, the molecular structure of the deoxyribonucleic acid (DNA) was resolved to be a double helix consisting of anti-parallel strands (Watson and Crick, 1953). Following the discovery of DNA, the code in which the genetic information is encrypted was cracked by the discovery of the cellular mechanism of translation used to

convey triplets of nucleotides (codons) into proteins (Crick et al., 1961; Nirenberg and Matthaei, 1961). In the final quarter of the 20th century, the complete nucleotide sequence of the bacteriophage MS2-RNA was determined (Fiers et al., 1976), and the following year it became possible to sequence the DNA for the first time, leading to the report of the first sequence of the entire genome of another bacteriophage, Φ-X174 (Sanger et al., 1977).

The achievements of the past century paved the way to the genome era that we are currently experiencing, with the milestone of 1,000 identified disease associated genes being reached in 2000 (Antonarakis and McKusick, 2000). Since then, the progress made in disease gene identification has been exponential with a number of 2,665 genes causing or being associated with Mendelian disorders as of October 1st 2011 (http://omim.org/statistics/geneMap). The number of disorders with known molecular basis has reached 4,455, while the defective gene remains elusive in an additional 1,777 disorders (http://omim.org/statistics/geneMap).

## 2.2 The human genome project

The development of the human genome project (HGP) has comprised a major step towards understanding how the genes affect or contribute to human health. The HGP was inaugurated in 1990 funded by the U.S. Department of Energy and the National Institutes of Health (NIH), as a collaborative 15-year project aiming to provide a high-quality version of the human sequence and create a complete map of the human genome (Lander et al., 2001). Another major goal laid out by the HGP was to map and sequence the genome of other organisms such as the laboratory mouse *Mus musculus*, the fruitfly *Drosophila melanogaster*, and the human gut bacterium *Escherichia coli*. The first draft of the human genome sequence was published in the year 2001 by two independent groups, one being the HGP and the second the private biotechnology company Celera Genomics (Lander et al., 2001; Venter et al., 2001), while the complete sequence was released three years later (The International Human Genome Sequencing Consortium, 2004). Along

with the progress made in sequencing of the human DNA, the obtained knowledge has become publicly available in several online databases.

With the human genome being cracked, many of its secrets became known. The human "book of life" is composed of 3,280 million nucleotides (http://www.ensembl.org/Homo_sapiens/Info/StatsTable?db=core as of April 2011; Build 37; GRCh37.p3). Surprisingly about half of the human genome is estimated to be composed of various types of repeat sequences, such as transposon-derived repeats, inactive retroposed copies of cellular genes, segmental duplications and blocks of tandem repeats (Lander et al., 2001). Approximately 48% of the human genome that does not comprise repeats is non-coding and intergenic, and was originally termed 'junk DNA' since no role could be attributed to it. Recent cross-species comparisons have revealed strong conservation for these regions, suggesting plausible structural and functional roles for them (Birney et al., 2007). Only approximately 2% of the bases encode for proteins, with the present consensus predicting the existence of an approximate 20,500 genes. These findings highlight the C-value paradox, originally introduced in 1971 (Thomas, 1971), according to which the complexity of an organism is not directly proportional to its genome size. However, this can be explained today by the complex patterns of transcriptional regulation, production of alternative transcripts via messenger ribonucleic acid (mRNA) splicing, and post-translational modifications observed (Gerstein et al., 2007).

Following the completion of the HGP, more consortia have been formed aiming to further shed light on the secrets of the human genome and exploit the information found in the DNA in order to resolve several diseases that afflict mankind.

1. The HapMap Project (www.hapmap.org/) was launched in 2003 with the aim to map common patterns of DNA variation in individuals from four populations (the Central Europeans from Utah (CEU), the Han Chinese from Beijing (CHB), the Japanese from Tokyo (JPT), and the Yorubans from Ibadan, Nigeria (YRI)). In this three-phase effort 3.1 million common single-nucleotide polymorphisms (SNPs) were genotyped across the genome (The International HapMap Consortium, 2003, 2005, 2007). The rationale was to identify SNPs that are in linkage disequilibrium (LD) with each other, meaning that

they are not inherited independently, but rather comprise distinct haplotype blocks. From each such non-randomly associated block only one "tag SNP" was selected, with the idea that when genotyped it would give information about other adjacent SNPs with which it is in LD. In this way one would only need to genotype 200,000-600,000 tag SNPs across the genome rather than the whole 10 million SNPs identified today (The International HapMap Consortium, 2003).

2. The Encyclopedia of DNA elements (ENCODE) project (http://www.genome.gov/10005107) was created with the aim to provide a biologically informative image of the human genome by cataloguing the functional elements of the human genome (The ENCODE Consortium 2004).

3. The 1000 genomes project (www.1000genomes.org) was launched in 2008 as an attempt to provide a complete and detailed catalogue of human genetic variation. In this three-year effort, the genomes of 1000 individuals from different ethnic backgrounds will be sequenced, aiming to discover more than 95% of the common variation occurring in the human DNA. Already in the pilot phase of the project 15 million SNPs, 1 million short insertions and deletions, as well as 20,000 previously undescribed structural variants were reported (The 1000 Genomes Project Consortium, 2010).

4. An initiative to sequence 10,000 individuals from the United Kingdom (UK10K project; http://www.uk10k.org/) was undertaken in order to uncover many rare genetic variants that are important in human disease. Towards this the genomes of 4,000 individuals that have been studied for many diseases and traits over recent years will be completely sequenced, while another 6,000 people diagnosed with extreme obesity, neurodevelopmental disorders or other conditions will be screened only for the protein-coding parts of their genomes (http://www.sanger.ac.uk/about/press/2010/100624-uk10k.html).

## 2.3 Types of sequence variation

The sequence of any two humans is similar at a rate of 99.9%, differing at approximately only one base every 1,000 nucleotides (The International HapMap Consortium, 2005). The variation arises from mutations that occur spontaneously either due to mistakes that take place during DNA replication, or induced via exposure to chemicals and radiation. When these mutations occur in the germline they become transmissible from generation to generation and can be used to map specific traits that run in familial cases. The types of sequence variation that are usually associated with disease are:

1. SNPs. These occur via substitutions of single base pairs (bp). The most common single nucleotide changes encountered are caused by transitions where a purine is exchanged for a purine (A ↔ G) and a pyrimidine for a pyrimidine (T ↔ C). Less common are transversions where a purine is exchanged for a pyrimidine and vice versa (A/T ↔ G/C). When a SNP falls within the protein coding segments of a gene it can result in a silent amino acid (aa) substitution (the amino acid does not change), a missense change (the amino acid is exchanged for a new one), or a nonsense mutation (a termination codon is introduced usually resulting in a premature end to protein synthesis). When the change affects the exon-intron junction the splicing pattern of the transcript can be altered via silencing of splice enhancers or activation of cryptic splice sites, whereby the mutation is classified as splice affecting.
2. Insertions, deletions and insertion/deletion (indel) changes describe when bases are added, removed, or a number of bases is removed only to be substituted by another sequence. The occurrence of such changes within the coding region of a gene usually results in alterations of the reading frame of the coding sequence, called a frameshift. Only a very small fraction of insertions, deletions, or indels have no affect on the grouping of the three-base pair formed codons and thus do not result in a change of amino acids.
3. Repeats. There are two major types of repeat sequences in the human genome.

   *Interspersed repeats* are the first major class of repeats, which are usually found as single copies widely distributed across the genome. The interspersed

repeats usually derive from transposable elements and comprise the most common repeat type encountered in the human genome, constituting 45% of it (Lander et al., 2001). The most known transposons resulting in interspersed repeats are the long interspersed elements (LINEs) and the short interspersed elements (SINEs). The LINEs have an average size of 1 kilobases (kb) (Lander et al. 2001) and contain a promoter sequence at the 5' end which enables them to transcribe independently (Deininger et al., 2002). The SINEs on the other hand are shorter elements (<500 bp) which cannot be transcribed independently. Almost all SINEs are derived from transfer ribonucleic acid (tRNA) genes (Okada, 1991). The most abundant human SINEs are the Alu sequences (280 bp long) which are non-coding, and are flanked by *AluI* restriction sites. They comprise the only active SINE in the human genome and have been implicated in the causality of several diseases that occur through Alu-mediated recombination (Lander et al., 2001; Cordaux and Batzer, 2009).

*Single sequence repeats (SSRs)* or tandem repeats are the second class of repeats, and involve the repetition of a *k*-mer several times. On the basis of the repeat unit different SSRs can be recognized. *Satellite* DNA has repeat unit sizes of >100 bp and can span from 100 kb up to 1 megabase (Mb). Most human satellite DNA is found in the centromeres. *Minisatellite* DNA has repeat units of approximately 7-100 bp and can range in size from 1-20 kb. Finally short tandem repeats (STRs) or *microsatellites* have repeat units of 1-6 bp and can span regions of hundrends of bp (Fan and Chu, 2007; Ellegren, 2000). Based on the number of nucleotides being repeated the microsatellites can be mono-, di-, tri-, tetra-, penta- or hexanucleotide repeats. Both minisatellite and microsatellite DNA belong to the class of sequences called variable number tandem repeats (VNTRs), which is characterized by high mutation rates making them vary significantly between individuals and thus comprise unique markers for DNA fingerprinting (Jeffreys, 2005).

4. Copy number variations (CNVs) arising from large scale deletions, insertions, segmental duplications as well as other chromosomal abnormalities can result in

variation of the genome. The CNVs are defined as DNA segments that are 1 kb or larger in size and present at variable copy numbers in comparison with the reference genome (Conrad et al., 2010). In recent years CNVs have emerged as a very interesting class of human variation that cannot be overlooked when trying to identify the genetic defect underlying a disorder (Redon et al., 2006). For example CNVs have been shown to be enriched in patients with epilepsy, schizophrenia, autism and several other disorders (Helbig et al., 2009; Stefansson et al., 2008; Sebat et al., 2007; Mefford and Eichler, 2009)

## 2.4 Principles for identification of human disease-associated genes

Since the identification of phenylalanine hydroxylase as the protein underlying phenylketonuria (Jervis, 1953), many advances have taken place in the field of disease gene identification. During the past three decades a variety of methods have been developed which can be divided into two main categories: position-independent and position-dependent strategies.

### 2.4.1 Position-independent approaches

The position-independent approaches initially required as a prerequisite some information on the likely underlying defect, or the biological pathway affected. Today, with the advent of massive parallel sequencing, such information is no longer mandatory. The major position-independent approaches used have been:

1. *The candidate gene approach*. The basis and at the same time limitation of this approach is that it requires some prior knowledge of the biochemical defect involved. This prior knowledge can derive from expression array analyses, for example, where the gene or protein expression levels are compared between affected and healthy individuals (Hoh and Ott, 2004). One disease in which the defect was identified using this method was hemophilia A, in which a deficiency in the coagulation factor VIII was detected in the blood of patients (Gitschier et al.,

1984). Alternatively, genes belonging to the same protein family, bearing the same protein domains, or belonging to the same pathway can be considered as good candidates.

2. *The homologous gene approach*. In this approach the rationale is that when a gene is identified to be defective in an animal model with a manifesting phenotype and disease symptomatology similar to that presented by a group of patients, the human ortholog could be considered a good candidate. One example is the identification of the genetic defect causing the DFNA15 autosomal dominant form of progressive hearing loss (MIM_602459), which was first identified in mice. The animals were deaf when both alleles of the POU class 4 homeobox 3 gene (*Pou4f3*) were deleted (Erkman et al., 1996; Xiang et al., 1997). The human homolog (*POU4F3*) naturally became an excellent candidate for DFNA15, and indeed was shown to harbor heterozygous deletions in the affected family members screened (Vahava et al., 1998).
3. *The whole exome/genome re-sequencing approach*. Large-scale sequencing has only become available after the dramatic reduction of cost, and the development of the high-throughput technology. Despite its enormous potential one drawback of next-generation sequencing is the handling of the large volume of data produced. Successful applications of next-generation sequencing include the identification of the Trithorax-group histone methyltransferase (*MLL2*) in Kabuki syndrome (MIM_147920) (Ng et al., 2010) and of the DnaJ homolog, subfamily C, member 5 gene (*DNAJC5*) in autosomal dominant adult-onset NCL (MIM_162350) (Noskova et al., 2011).

### 2.4.2 Position-dependent approaches

The position-dependent approaches do not require any prior knowledge of the molecular pathogenesis of the underlying pathway, complex or defect. Thus no assumptions need to be made *a priori*, rather the defect is identified based primarily on its position in the genome. The position dependent approaches traditionally follow three steps (Figure 1).

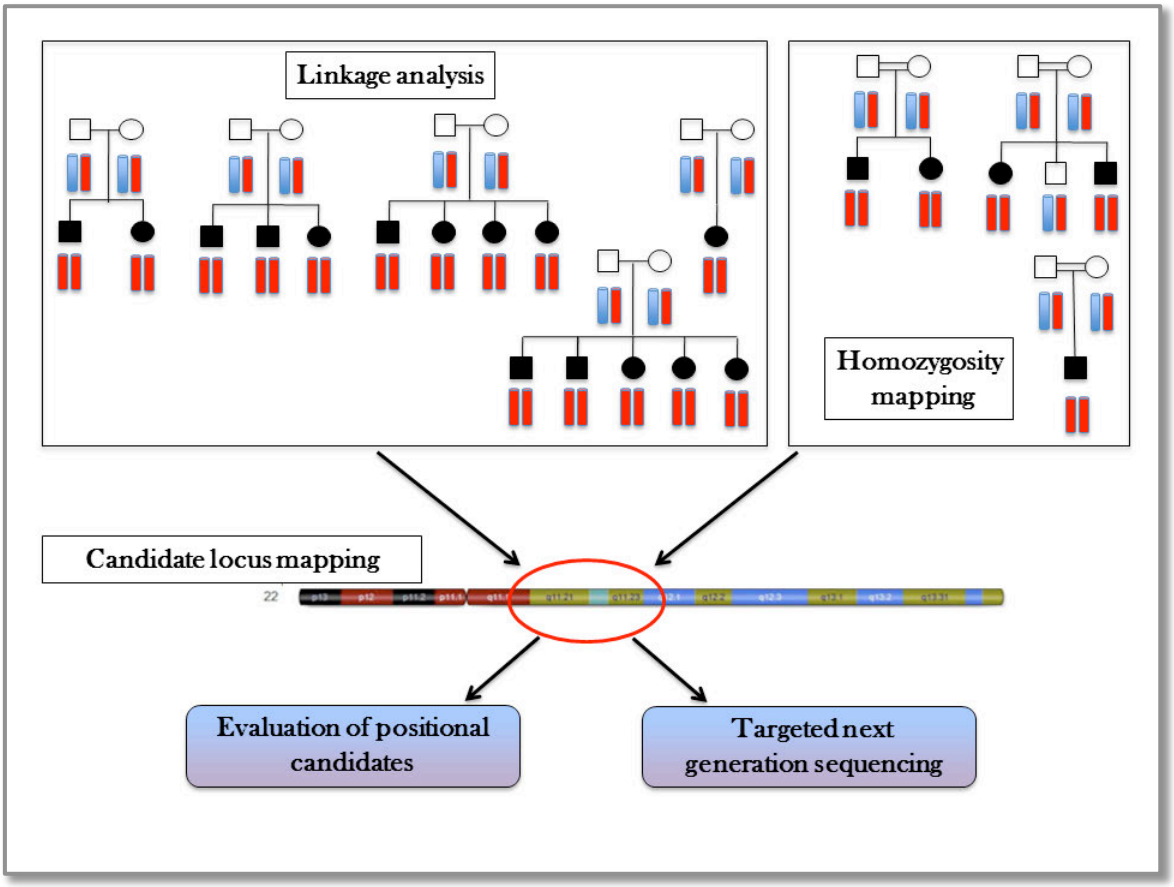


**Figure 1.** Schematic representation of the position-dependent gene identification approach. In the first step, the study material where a phenotype segregates with the disease genotype (shown as a red block in the pedigree representations) is collected. The non disease-associated chromosomal loci are shown as blue blocks. After inspection of the genotypes the chromosomal interval harboring the disease gene is mapped as tightly as possible (indicated with a red eclipse). Finally, the genes residing within the critical region are screened for mutations.

1. *Collection of the study material.* The first step of all the position-dependent approaches relies on the mapping of the critical interval in families and/or sporadic individuals, depending on the disease/study type. In linkage analysis a collection of families with the same trait or phenotype is investigated. Large pedigrees with many affected individuals are more informative in mapping of disease-associated loci. In homozygosity mapping consanguineous families and/or sporadic patients are investigated, with familial cases with more than one affected individuals providing increased power to the analysis. When multiple families are used in the same analysis (linkage or homozygosity mapping) it is of critical importance that their phenotype is carefully evaluated, so that misdiagnoses that could lead to false positive or false negative results can be avoided.
2. *Mapping of the candidate interval.* The second step is to define the candidate region in the evaluated family members as tightly as possible (Collins, 1992). Any gross chromosomal rearrangements or identification of deletions spanning tens or hundreds of kb can greatly facilitate the region mapping, as was demonstrated in the case of the cloning of the gene underlying Duchenne muscular dystrophy (MIM# 310200). Mapping of the candidate interval requires genotyping of the study material across the whole genome (genomewide analysis) with polymorphic markers. The most commonly used markers for genomewide analyses have been VNTRs (microsatellites and less frequently nowadays minisatellites) and SNPs. VNTRs are more informative than SNPs because of their higher mutation rate (approximately ~$10^{-3}$ to $10^{-4}$ per locus per generation; Ellegren, 2000). SNPs on the other hand, although less polymorphic (mutation rate of 2.5 x $10^{-8}$; Nachman and Crowell, 2000), are more common and evenly distributed across the genome. It has been estimated that in an analysis of genomewide scan data the power obtained from genotyping approximately 400 microsatellites would be equivalent to genotyping approximately 1,000 SNPs (Landegren et al., 1998). Genomewide genotyping data can either be used to perform linkage or homozygosity analysis depending on the study.

   Linkage analysis and LD to fine-map the region of interest: Linkage analysis is the method in which the markers that have been genotyped across the genome are

inspected to see whether they co-segregate with the disease phenotype giving an indication that they reside within the disease locus in the families evaluated. If a marker in one locus and the locus containing the disease gene are inherited together, they are said to be in genetic linkage. The co-inheritance of alleles across two loci is computed assuming different recombination fractions (θ) between them. When θ=0 the two loci have a 0% chance of being separated by recombination and are thus said to be linked. A θ value of 0.5 denotes that the loci are not linked and any value in the range 0.5> θ >0 represents positions with different probabilities of having a recombination event separating the two loci. Whether the two studied loci are likely to be truly linked and not co-inherited by simple chance is expressed as the base 10 logarithm of the likelihood of the loci being linked divided by the likelihood of the loci not being linked at a given θ, known as the logarithm of odds or LOD score (Z; Morton, 1995):

$$LOD(\theta) = \log_{10} \frac{(1-\theta)^N \times \theta^R}{0.5^{(N+R)}},$$

where N=number of non-recombinant offspring, and R=number of recombinant offspring.

Negative LOD score values argue against linkage (Z < -2.0), values of Z > 3.0 are suggestive of linkage, while values of 3.0 > Z > -2.0 are inconclusive.

With linkage studies most commonly resulting in the mapping of regions as long as several Mb, fine mapping is almost always necessary to narrow down the critical interval. Refining of a candidate region can be achieved either by mapping a denser set of molecular markers (such as SNPs, which are found at a rate of ≥1 per kb on average), or in founder populations by LD mapping. Two alleles in adjacent loci (e.g. a marker and a disease-causing mutation) are in LD when they are associated together more frequently than expected if the loci were segregating independently. This non-random association of markers occurs because the two loci are less likely to be separated by recombination. In LD fine-mapping the allele frequencies of markers within the critical interval are compared between affected

and control individuals. The set of marker alleles that are overrepresented in the patients versus the controls are indicative of close physical distance, being in non-random association and thus in LD with the disease mutation. LD fine-mapping has successfully been applied in many isolated populations such as the Finnish population for the positional mapping of, for example, diastrophic dysplasia (MIM_222600; Hästbacka et al., 1992), and the progressive myoclonic epilepsy type 1 (*EPM1*) gene for Unverricht-Lundborg disease (MIM_254800; Lehesjoki et al., 1993).

Homozygosity mapping: First introduced over two decades ago, homozygosity (or autozygosity) mapping represents a powerful tool for mapping rare genes causing recessive traits (Lander and Botstein, 1987). This approach looks for long chromososmal segments that are homozygous in affected individuals. Homozygosity can occur either when the alleles carried by an affected individual are identical by descent (IBD) over a locus, where the alleles are identical copies of the same ancestral allele, or when the alleles are identical by state (IBS) at a given locus, whereby the two allele-copies derive from different ancestral sources. Homozygosity mapping works by virtue of identifying haplotype signatures in IBD and for this reason it has been applied traditionally to families that show consanguinity. In such families a significant proportion of the offsprings' genome is identical, with children from a first-cousin marriage having on average 6% of their genome in a homozygous stage, from a double first-cousin marriage 12.5% and so on (Lander and Botstein, 1987). The degree of homozygosity encountered in such children by far exceeds the extent of homozygosity seen in a general population, resulting in enrichment of rare recessive diseases in the former (Broman and Webber, 1999). According to mathematical calculations inordinate numbers of families would be needed in order to have power to detect linkage in nuclear families with recessively inherited disorders (Wong et al., 1986). With homozygosity mapping, however, only ten unrelated affected children from consanguineous marriages are enough to map a recessive gene (Lander and Botstein, 1987), with no need to genotype all the intervening relatives (Carr et al., 2006). A single affected child of a first-cousin marriage is expected to contain the

equivalent total information as an outbred nuclear family with three affected children (Lander and Botstein, 1987). In such a child, approximately 20 segments exceeding 3 centiMorgans (cM) in length can be identified, with the longest segment usually harboring the disease locus (Woods et al., 2006). A clear pitfall arising from this observation is that in inbred children there is an excess of homozygosity and thus identifying the "true" disease locus might be challenging. A way to overcome this is by comparing the homozygous segments of unrelated individuals affected by the same disease and displaying similar phenotypes and prioritizing those over which several unrelated patients are simultaneously homozygous, which would suggest that homozygosity over such loci is less likely to have occurred by chance. Inspection of recombinations in each of these patients further facilitates in narrowing down the critical interval. Successful examples of disease gene identification by use of homozygosity mapping are the identification of the eighth NCL-causing gene, the major facilitator superfamily containing 8 (*MFSD8*), in a subset of mainly Turkish patients (Siintola et al., 2007), and the *RAB23* gene underlying Carpenter syndrome (Jenkins et al., 2007).

3. *Evaluation of positional candidate genes*. After the definition of the critical interval with as much resolution as possible, the genes lying within it are listed and prioritized for mutation screening (Collins, 1995). Prior to the HGP era extensive physical mapping and cloning were required. Today, after the incorporation of the HGP data in publicly available genome browsers, cataloguing of candidate genes has become very easy. Prioritization of the list of genes that map within the region of interest is achieved after having obtained *in silico* and comparative genomics information on their putative or demonstrated function and spatio-temporal expression. When the critical interval is too long and prioritization of the positional candidates is not possible, targeted next-generation sequencing of all the genes residing within the candidate locus, seems to be the most suitable approach. This method was successfully employed to screen 108 candidate genes residing within a previously identified linkage region on chromosome 9q34.3 in an affected individual with autosomal recessive prelingual nonsyndromic sensorineural hearing loss. Targeted re-sequencing of the critical 2.9 cM interval revealed a

nonsense mutation in *C9orf75*, designated taperin (*TPRN*), establishing it as the underlying disease-causing gene (Rehman et al., 2010).

### 2.4.3 Confirmation of a disease-causing gene

A gene identification study cannot be considered successful unless the disease-associated role of the genetic changes or of the gene has been established. There are several ways that can be used to shed light into the causality of the variants identified and thus into the resultant causality of the defective gene:

1. Segregation with the disease phenotype. Once sequence changes are identified, it should be confirmed that they segregate with the disease phenotype in the evaluated families. In dominant disorders with complete penetrance all the family members carrying one copy of the mutation should also manifest the phenotype in question. When the penetrance is incomplete, it is not mandatory that all the mutation carriers will have the disease phenotype. In this case, however, the penetrance in the study cohort should approximate the value described in the literature (whenever this is possible). In recessive disorders all affected individuals need to be homozygous for the mutation (or compound heterozygous when two different mutations run in the same family), and the unaffected parents obligate carriers.
2. Screening of control chromosomes. Towards establishing the pathogenic role of a sequence variant it is important to evaluate at least 200 preferably ethnically matched control chromosomes. If the variant is causing disease then it should not be detected in a homozygous state in the healthy population. If the change is identified in the control chromosomes then the frequency of the variant should not exceed the allele frequency expected for the population studied.
3. Identification of mutations in unrelated families. A major factor that can provide evidence in favor of the causality of a genetic defect is the identification of other families with defects in the same gene. If only a single family is found to carry defects in a given gene it is difficult to assess whether these changes represent true

disease-causing mutations, private polymorphisms or private modifiers of phenotype.

4. Impact of the change. Depending on the type of change identified, different conclusions can be drawn for the impact it might have on the gene product. Nonsense mutations can have a pronounced impact, causing the introduction of premature stop codons with the resultant proteins being truncated and erroneous, or targeted for nonsense-mediated decay by the cell's mRNA surveillance mechanism (Maquat, 2004; Chang et al., 2007a). Insertions, deletions and indels can either result in frameshifts or in in-frame removal or insertion of extra amino acid residues. The frameshift changes are predicted to cause the formation of aberrant proteins, either via introducing stop codons like the nonsense mutations, or via eliminating the existing stop codons. The effect of changes that affect the splicing is not always straightforward and sometimes further investigation is needed for their disease-association to be established. Whenever invariant splice site nucleotides at the donor and acceptor intronic splice junctions are affected, the normal splicing is abolished because cryptic splice sites, splice enhancers or splice silencers are activated, leading to an inability to define the exon boundaries resulting in exon skipping or intron retention and subsequently to abnormal proteins (Cartegni et al., 2002; Nakai and Sakamoto, 1994). However, changes that can affect the correct splicing of a transcript can occur beyond the consensus splice site positions. To be able to unequivocally test the impact of a sequence variation on splicing, the complementary DNA (cDNA) sequence must be evaluated with the reverse transcriptase polymerase chain reaction (RT-PCR) (Varon et al., 2003). Finally, the missense changes are the most difficult types of sequence variation to assess, since their impact on protein expression and function is not as straightforward. To deduce the possible outcome of a missense change the conservation of the affected residue must be evaluated. Typically, evolutionarily conserved amino acids represent critical residues for the protein function. A second method involves *in silico* analyses that in addition to conservation also compares the putative mutation effects on protein folding. Such analyses can be performed with freely available prediction programs like

PolyPhen 2 (http://genetics.bwh.harvard.edu/pph2/) and SNPs3D (http://www.snps3d.org/).

5. Functional evaluation. Only functional validation can result in the distinction between a disease-causing mutation and a benign polymorphism. Nevertheless, functional evaluation of the impact of a sequence variant is greatly dependent on what is already known about the protein. Assuming the protein function is known a simple test can be employed to assess the variant causality. For example if the encoded product is an enzyme, an enzymatic test measuring the activity levels of the mutant protein can be used. In cases where the protein function remains elusive primary functional properties can be assessed. These can involve subcellular localization studies to determine whether the protein localization is altered, or Western blot assays to evaluate changes in protein expression levels.
6. Disease modeling in cellular and/or *in vivo* systems. If the disease can be modeled in *in vitro* or *in vivo* systems, it provides a powerful tool to further study the impact of the changes identified. Several mouse models have been generated for a substantial number of human disorders thus far using many of the available techniques, such as knock-out models, where the gene is irreversibly silenced, or knock-in models, where specific allelic mutations are introduced in a given gene. A major drawback is that modeling a disease in another organism can be a very time-consuming process. Furthermore, it is not guaranteed that the models produced will recapitulate the key clinical features of the human disorder, given the different physiologies existing between humans and experimental animals. Alternatively, in animal models where the gene can be silenced with RNA interference (RNAi in mice or morpholinos in zebrafish), the causal effects of the mutations can be evaluated via injection of the mutant constructs, whereby restoration of the normal phenotype argues in favor of the mutant not being pathogenic, while failure to restore the normal phenotype provides evidence for the pathogenic role of the change (Lieschke and Currie, 2007).

### 2.4.4 Factors complicating disease gene identification

Several factors can influence the outcome of a linkage analysis. If such factors are not kept in consideration, linkage can be missed. One of the most important limiting characteristics of genetic diseases is *locus heterogeneity*, a phenomenon in which mutations in different genes residing in distinct loci can result in similar clinical phenotypes. In linkage studies of disorders for which locus heterogeneity exists in the collection of families studied, the locus or gene will be extremely hard to map. A second factor is *allelic heterogeneity*, which describes the case of different mutations in the same gene underlying significantly different clinical phenotypes. *Pleiotropy* resulting from concomitant mutations in different genes required to manifest the phenotype is a third factor that can also complicate genetic studies. *Variable expressivity* of a gene can greatly influence the outcome of an analysis since it can result in modification of a phenotype. Several phenomena can result in variable gene expressivity such as *epistasis*, the synergistic effect of nonallelic modifier genes, *gene-environment interactions*, *incomplete penetrance*, whereby individuals with the disease genotype can fail to manifest the disease, or *phenocopies*, where individuals who do not have genetic defects can mimic a particular phenotype. Finally, *mutations residing outside the studied region* such as in introns or in the regulatory elements of a gene, can greatly complicate disease gene identification.

## 2.5 Childhood-onset PMEs

Traditionally, three main categories of epilepsies are recognized on the basis of etiology: idiopathic, symptomatic, and cryptogenic. Idiopathic epilepsies are caused by a known or presumed genetic defect, and in these the seizures are the main or even only manifested symptom. In symptomatic epilepsies the epileptic events are only one of the symptoms associated with a defined neurological condition, which can be either acquired (brain tumor, hypoxic-ischemic episode) or hereditary (Capovilla et al. 2009). Cryptogenic epilepsies comprise all epilepsies with unidentifiable etiology.

A clinically definable group of symptomatic epilepsies with onset of symptoms in most cases in childhood comprises the progressive myoclonic epilepsies (PMEs; Berkovic et al., 1986). The PMEs are characterized by myoclonus, tonic-clonic seizures, progressive neurologic decline and in many cases premature death (Berkovic et al., 1986). A combination of clinical features such as age of onset or type of seizures manifested, together with pathologic examinations and molecular genetic analyses, have aided in the definition of six distinct main PME entities: Unverricht-Lundborg disease (ULD), Lafora disease, the NCLs, the sialidoses, myoclonic epilepsy with ragged red fibres (MERRF) and dentatorubraöpallidoluysian atrophy (DRPLA) (reviewed in Ramachandran et al., 2009). Age of onset in ULD, LD, sialidoses type II, and MERRF is typically in childhood, during the second decade of life, and patients survive at least until adulthood (Norio and Koskiniemi, 1979; Van Heycoptenhamm, 1963; Sasaki et al., 1983). In sialidoses type I the onset is in adulthood and survival is prolonged. Finally in the NCLs and in DRPLA, a wide spectrum of disease onset has been described with each disease entity having cases ranging from congenital to adulthood affection (reviewed in Mole et al., 2005; Noskova et al., 2011; Zupnac and Legros, 2004). The PMEs are autosomal recessive disorders, the only exceptions being the autosomal dominant DRPLA and MERRF (Table 1; Delgado-Escueta et al., 2001).

**Table 1.** Recessively inherited PMEs with onset in childhood

| Disorder | Subtype | MIM # | Defective gene | Protein | Onset | Storage material | Reference for gene identification |
|---|---|---|---|---|---|---|---|
| Unverricht-Lundborg disease | EPM1 | 601145 | *CTSB* | Cystatin B | 6-15 years | No storage material | Pennacchio et al., 1996 |
| Lafora Disease | EPM2A | 607566 | *EPM2A* | Lafora PTPase | 8-18 years | Periodic acid-Schiff-positive inclusions in various tissues | Minassian et al., 1998 |
| | EPM2B | 608072 | *NHLRC1* | NHL repeat containing protein 1 | Mean onset 12 years | | Chan et al., 2003 |
| Sialidosis | Type II | 608272 | *NEU1* | Sialidase-1 | 0 months-12 years | Lysosomal storage of sialidated glycopeptides and oligosaccharides | Bonten et al., 1996 |
| Neuronal ceroid lipofuscinoses (NCLs) | CLN1 disease | 256730 | *PPT1* | Palmitoyl-protein thioesterase 1 | 10-18 months | GROD | Vesa et al., 1995 |
| | CLN2 disease | 204500 | *TPP1* | Tripeptidyl-peptidase 1 | 2-4 years | CL, CL/FP, CL/GROD, or CL/FP/GROD | Sleat et al., 1997 |
| | CLN3 disease | 204200 | *CLN3* | CLN3 | 5-10 years | FP, CL, FP/CL, FP/GROD, FP/CL/RL, or FP/CL/GROD | The International Batten Disease Consortium, 1995 |
| | CLN5 disease | 256731 | *CLN5* | CLN5 | 4-7 years | FP, CL, FP/CL, FP/RL, FP/GROD, or FP/CL/RL/GROD | Savukoski et al., 1998 |
| | CLN6 disease | 601780 | *CLN6* | CLN6 | 3-8 years | FP/CL, or FP/GROD | Gao et al., 2002<br>Wheeler et al., 2002 |
| | CLN7 disease | 610951 | *MFSD8* | Major facilitator superfamily domain-containing protein 8 | 2-7 years | FP, FP/CL, FP/GROD, FP/RL/CL, or FP/CL/GROD | Siintola et al., 2007 |
| | CLN8 disease | 600143 | *CLN8* | CLN8 | 2-7 years | FP, CL, FP/CL, CL/GROD, or FP/CL/GROD | Ranta et al., 1999 |
| | CLN9 disease | 609055 | unknown | unknown | 4 years | GROD and CL | Gene remains unidentified |
| | CLN10 disease | 610127 | *CTSD* | Cathepsin D | At birth | GROD | Siintola et al., 2006<br>Steinfeld et al., 2006 |

**Abbreviations used:** GROD: granular osmiophilic deposits, CL: curvilinear bodies, FP: fingerprint profiles, RL: rectilinear bodies, n.d.: not defined

In addition to the six main PME entities, a number of less common subtypes have also been described (Berkovic et al., 1986). For example, a single homozygous mutation in prickle homolog 1 gene (*PRICKLE1*) was reported to underlie Unverricht-Lundborg type 1B (EPM1B; MIM_612437) in two families of Israeli and one family of Jordanian origin (Bassuk et al., 2008). A dosage effect was reported for the same gene since heterozygous changes in it have also been associated with epilepsy (Tao et al., 2011). Two heterozygous changes have also been identified in the homolog of *PRICKLE1*, *PRICKLE2*, as the underlying causes of a juvenile onset subtype termed EPM5 (MIM_613832; Tao et al., 2011). A single nonsense mutation (p.Arg99X) in the potassium channel tetramerization domain-containing 7 gene (*KCTD7*), identified in a Moroccan pedigree with three affected individuals, is associated with EPM3 (MIM_611725; van Bogaert et al., 2007). In EPM3 the disease onset is in late-infancy (16-24 months) and the patients become demented (van Bogaert et al., 2007). Action-myoclonus with or without renal-failure syndrome (AMRF), also known as EPM4 (MIM_254900), is another recessively inherited subtype of PME with onset ranging from juvenile age to adulthood (14-23 years). EPM4 is caused by mutations in the scavenger receptor class B, member 2 gene (*SCARB2*) and is characterized by the extraneuronal accumulation of irregularly shaped and autofluorescent pigmented granules in the brain (Berkovic et al., 2008; Dibbens et al., 2009). Finally, a homozygous missense change (p.Gly144Trp) identified in the Golgi SNAP receptor complex member 2 (*GOSR2*) in five individuals that apparently shared a common ancestor was reported to cause EPM6 (MIM_614018), with onset in early childhood and ataxia as the presenting symptom (Corbett et al., 2011).

## 2.6 Neuronal ceroid lipofuscinoses

The NCLs are one of the six major clinical entities comprising the group of PMEs (Berkovic et al., 1986). Although the NCLs are considered rare disorders, with incidence ranging between 1:12,500 and 1:100,000 (Santavuori, 1988; Rider and Rider, 1988), they comprise the most common cause of neurodegeneration in childhood. The hallmark that establishes a NCL diagnosis is the accumulation of autofluorescent material in the lysosomes of mainly neurons but also extraneural cells (Haltia, 2003; Mole et al., 2005). Today it remains an open question whether this intracellular accumulation is responsible for the extensive neuronal death taking place throughout disease progression. Most NCL subtypes are transmitted in a mendelian recessive manner. Nevertheless, descriptions of rare dominant cases also exist in the literature (reviewed by Mole et al., 2005). NCL-affected patients present with a variety of symptoms, including different types of epileptic seizures, regression in motor skills, visual failure, cognitive impairment and premature death. The course of the disease is continuously debilitating and patients eventually die prematurely (Haltia, 2003).

To date, ten different subtypes of human NCLs and nine genes underlying them have been identified (*PPT1*, *TPP1*, *CLN3*, *DNAJC5*, *CLN5*, *CLN6*, *MFSD8*, *CLN8*, and *CTSD*) (Tables 1 and 2; Noskova et al., 2011; reviewed by Mole et al., 2005). Classification of NCL patients into any of the different known subtypes has traditionally followed two systems: one based on the age of onset, and the second based on the morphology of the lipopigment material. Five main types are recognized according to age of onset: congenital (CONCL), infantile (INCL), late infantile (LINCL), juvenile (JNCL) and adult (ANCL) NCL (Mole et al., 2005). After almost two decades of genetic research, the identification of 364 mutations and the molecular genetic background resolution for several hundreds of patients, many atypical clinical cases have been described, rendering NCL classification a difficult task. Towards this, a new nomenclature has only recently been proposed which takes into consideration not only the clinical and neuropathological findings, but also the underlying genetic findings (study I).

**Table 2.** Genes and proteins associated with NCL disease

| Disease | Onset | Chromosomal position | Gene | Type of protein | Function | Compartment |
|---|---|---|---|---|---|---|
| CLN1 disease | Infancy | 1p32 | *PPT1* | Soluble | Thioesterase | Lysosomes |
| CLN2 disease | Late-infancy | 11p15 | *TPP1* | Soluble | Peptidase | Lysosomes |
| CLN3 disease | Juvenile | 16p12.1-11.2 | *CLN3* | Transmembrane (5TMs) | Not known | Lysosomes Endosomes |
| CLN4 disease | Adulthood | 20q13.33 | *DNAJC5* | Subunit of chaperone complex | Chaperone | Plasma membrane |
| CLN5 disease | Late-infancy | 13q21.1-q32 | *CLN5* | Soluble | Not known | Lysosomes |
| CLN6 disease | Late-infancy | 15q21-23 | *CLN6* | Transmembrane (7TMs) | Not known | ER |
| CLN7 disease | Late-infancy | 4q28.1-q28.2 | *MFSD8* | Transmembrane (12TMs) | Not known | Lysosomes |
| CLN8 disease | Late-infancy | 8q23 | *CLN8* | Transmembrane (5TMs) | Not known | ER and ERGIC |
| CLN9 disease | Juvenile | Not known | Not known | Not known | Not known | Not known |
| CLN10 disease | Congenital | 11p15.5 | *CTSD* | Soluble | Protease | Lysosomes |

**Abbreviations used:** ER: endoplasmic reticulum, ERGIC: ER-Golgi intermediate compartment

### 2.6.1 NCL subtypes with congenital onset: CLN10 disease

Congenital NCL is caused by mutations in the cathepsin D (*CTSD*) gene and is the earliest onset and most severe form among the NCL phenotypes (Siintola et al., 2006). To date only eleven cases of congenital NCL have been reported (Barohn et al., 1992; Brown et al., 1954; Humphreys, 1985; Garborg et al., 1987; Norman and Wood, 1941; Sandbank, 1968; Siintola et al., 2006; Fritchie et al., 2009). The onset is believed to be in embryonic life, with developmental delay starting from the $30^{th}$-$32^{nd}$ week of gestation

(Sandbank, 1968; Fritchie et al., 2009). Upon birth the patients show post-natal respiratory insufficiency, microcephally, reduced brain size, and status epilepticus. The degree of affection is so severe that patients die within hours to weeks after birth (Siintola et al., 2006). The autofluorescent storage bodies in CLN10 disease present granular osmiophilic deposit (GROD) morphology, while the main protein component of the inclusions is the sphingolipid activator protein (SAP) D (Humphreys et al., 1985; Garborg et al., 1987; Barohn et al., 1992).

The *CTSD* gene (NM_001909.3) was first identified in a naturally occurring ovine animal model for congenital NCL (Tyynelä et al., 2000). Subsequent identification of *CTSD* mutations in human patients has corroborated this gene as a member of the NCL family and the underlying cause of CLN10 disease (Siintola et al., 2006; Steinfeld et al., 2006; Fritchie et al., 2009). The protein encoded by *CTSD* is a lysosomal aspartic protease that belongs to the pepsin family (Metcalf and Fusek, 1993). Being synthesized as an inactive precursor proenzyme, CTSD is targeted to the endoplasmic reticulum (ER), where it is glycosylated and phosphorylated, via an N-terminal 20-aa target signal (Erickson et al., 1981). The modified proenzyme is then targeted to the lysosomes via either the mannose-6-phosphate (Man6P)-dependent, or Man6P-independent pathways (Zaidi et al., 2008). In its mature form, which occurs only in the acidic environment of the lysosomes, the enzyme is organized into a bi-lobed structure cleaving peptide bonds of polypeptide chains, mediating thus protein degradation, protease precursor form activation, or protease inhibitor inactivation (Rawlings and Barrett, 1995; Scarborough et al., 1994).

No common mutations exist in the CLN10 subtype, with only three missense and one nonsense mutations having been reported to affect the sequence of *CTSD* to date (http://www.ucl.ac.uk/ncl/catD.shtml). All *CTSD* mutations are associated with loss of enzymatic activity, while only one of them (p.Ser100Phe) abolishes normal protein localization causing retention of the mutant proteins in the ER (Fritchie et al., 2009; Steinfeld et al., 2006).

### 2.6.2 NCL subtypes with infantile onset: CLN1 disease

Mutations in PPT1 are usually associated with CLN1 disease, infantile. Children affected with this subtype have a disease onset at 6-18 months of age (Santavuori et al., 1973), although reduction in head growth can be observed even before 6 months of life, consistent with the disease beginning before the first symptoms become apparent (Santavuori et al., 2000). The presenting symptom is usually psychomotor decline during the second year of life, followed by seizures, ataxia, speech and visual failure, loss of ambulation by the age of 3 years, and death between 6-15 years of age (Santavuori et al., 1988). The accumulating inclusions in CLN1 disease, infantile, always form GRODs and contain SAP A and D (Das et al., 1998; Santavuori et al., 2000).

Mutations in *PPT1* (NM_000310.3) were first associated with CLN1 disease, infantile, in a group of mainly Finnish patients (Vesa et al., 1995). The enzyme encoded by *PPT1* is palmitoyl-protein thioesterase 1 (PPT1), a soluble fatty-acid hydrolase. A 25-aa signal peptide mediates lysosomal targeting of PPT1 (Schriner et al., 1996) via the Man6P-dependent or alternative pathways (Hellsten et al., 1996; Verkruyse et al., 1996; Lyly et al., 2007). Although the substrate on which PPT1 operates is unknown the enzyme catalyzes the removal of fatty acyl groups from modified cysteines of lipid-modified proteins *in vitro*, and fatty acylated proteins *in vivo* (Lu et al., 1996). In contrast to the lysosomal localization of PPT1 in non-neuronal cells in neurons it is axonally targeted, co-localizing with the synaptosomes and synaptic vesicles rather than with lysosomes (Ahtiainen et al., 2003; Lehtovirta et al., 2001; Heinonen et al., 2000).

To date 48 *PPT1* mutations (http://www.ucl.ac.uk/ncl/cln1.shtml) have been described in patients from 19 different countries. Three of the 48 known *PPT1* mutations are enriched in specific populations, with p.Thr75Pro and p.Arg122Trp being established by founder effects in Scotland and Finland, respectively (Munroe et al., 1998; Vesa et al., 1995) and p.Arg151X being detected worldwide (Das et al., 1998; Mitchison et al., 1998; Munroe et al., 1998; Mole et al., 2001). All described mutations result in varying levels of reduced enzyme activity (Mitchison et al., 1998; Das et al., 2001; Sleat et al., 2001; Simonati et al., 2009). Additionally, some missense mutations result in alteration of the

normal protein distribution because the Man6P receptors (MPRs) cannot recognize the mutant peptides which remain trapped in the ER (Das et al., 2001).

### 2.6.3 NCL subtypes with late infantile onset

#### *2.6.3.1 CLN2 disease*

Patients with classical CLN2 disease develop normally during the first year of life. Disease onset usually starts with seizures, which can be of any type between 2 and 4 years of age and is relatively fast followed by cognitive decline, loss of motor skills, myoclonous, ataxia, speech decline, and visual failure (Santavuori, 1988; Williams et al., 1999). Premature death can occur as early as 6 years of age or at the latest during the second decade of life (Mole et al., 2005). The ultrastructure of the autofluorescent material appears as curvilinear (CL) bodies, or a mixture of CL with fingerprint profiles (FP) and/or GRODs on electron microscopy (EM) analyses (Sleat et al., 1999; Hartikainen et al., 1999; Bessa et al., 2008; Elleder et al., 2008; Kohan et al., 2009). The protein material accumulating in the storage bodies consists of the subunit c of mitochondrial ATP synthase (SCMAS), and SAP A and D (Lake and Hall, 1993; Tyynelä et al.,1995).

CLN2 disease, classic late infantile, is caused by mutations in *TPP1* (NM_000391.3) which encodes for the lysosomal tripeptidyl-peptidase I (Sleat et al., 1997). Similar to PPT1 this pepstatin-insensitive carboxypeptidase is synthesized as a precursor polypeptide which becomes active in the lysosomal compartment (Sleat et al., 1997). Although it is known that lysosomal targeting of TPP1 is mediated by a 16-aa signal sequence the pathway(s) followed have not been specified (Steinfeld et al., 2004). TPP1 operates on small peptides undergoing degradation in the lysosomes and by removing tripeptides from their N-terminal domains at pH values of 4-4.5, and at lower pH levels (pH=3) it can act as an endopeptidase (Vines and Warburton 1998; Warburton and Bernardini 2000; Ezaki et al., 2000).

Altogether 66 *TPP1* affecting mutations have been reported in patients from 23 different countries (http://www.ucl.ac.uk/ncl/cln2.shtml). Of these, mutation p.Gly284Val is particularly predominant in the Canadian population, having been established by a founder effect (Ju et al., 2002). Two mutations (c.509-1G>C and p.Arg208X) have been encountered in high frequencies in patients worldwide (Zhong et al., 1998; Sleat et al., 1999). Most of the mutations identified are loss of function, causing reduced activity of the mutant enzymes. This, however, might not be the only way in which *TPPI* mutations result in disease manifestation, since some mutations have been shown to impair lysosomal trafficking by trapping the mutant peptides in the ER (Steinfeld et al., 2004).

#### *2.6.3.2 CLN5 disease*

CLN5 disease represents a variant subtype of the classical LINCL. Originally identified in a subset of Finnish patients it was also known as the Finnish variant NCL (Santavuori et al., 1982; Santavuori et al., 1993). Patients with CLN5 disease, late infantile variant, have disease onset at 4-7 years of age. After manifestation of motor clumsiness and concentration disturbances, which most commonly comprise the presenting symptoms, mental and motor decline, myoclonus, epilepsy, ataxia and visual deterioration ensue (Holmberg et al., 2000; Santavuori et al., 1982; Santavuori et al., 1991). Patients lose their ability to walk by the age of 10 years and die when they are 10-30 years old. On EM analysis the autofluorescent material appears either as FP, or a mixture of FP with rectilinear (RL) bodies, CL and rarely GRODs (Santavuori et al., 1982; Mole et al., 2005). The subunit accumulating in these granules is SCMAS, but can also be SAP A and D (Tyynelä et al., 1997).

The defective gene of this subtype is *CLN5* (NM_006493.2) which encodes a soluble lysosomal glycoprotein of unknown function (Savukoski et al., 1998; Schmiedt et al., 2010). Four different starting methionines (at positions p.1, p.30, p.50 and p.62) produce four different peptides of different lengths, the roles of which remains elusive (Isosomppi et al., 2002). Although the lysosomal signal peptide has not been defined proteolytic cleavage of the motif has been proposed to occur at p.96 (Schmiedt et al., 2010). Similar

to the sorting mechanism of PPT1 both the Man6P-dependent and alternative pathways can be used for the lysosomal trafficking of CLN5 (Sleat et al., 2009; Kollmann et al., 2005; Schmiedt et al., 2010).

Altogether 24 *CLN5* mutations (http://www.ucl.ac.uk/ncl/cln5.shtml) have been detected in patients from 19 different countries. Most *CLN5* mutations are private with only one, the nonsense p.Tyr392X, having been established due to a founder effect in the Finnish population (Savukoski et al., 1998). With the exception of a subset of four mutations that cause ER retention and subsequent degradation of the mutant proteins, the impact that the remaining 20 mutations might have is difficult to deduce while the protein's function remains unknown (Schmiedt et al., 2010; Lebrun et al., 2009).

#### *2.6.3.3 CLN6 disease*

CLN6 disease is the second variant subtype of LINCL, also known as the Costa Rican variant. The clinical presentation of patients positive for the CLN6 subtype is essentially similar to that of TPP1 positive patients. In CLN6 disease onset is slightly later that in CLN2 disease, at 3-8 years of age (Mole et al., 2005). The most common presenting symptoms are epileptic seizures and motor difficulties. Additional symptoms include mental regression, speech and visual impairment, myoclonus, and ataxia (Mole et al., 2005; Moore et al., 2008). Death occurs in the third decade of life (Pena et al., 2001). The ultrastructural granules accumulating in the lysosomes of CLN6 positive patients consist of a mixture of FP with CL and/or GRODs, and the protein in them consists of SCMAS (Sharp et al., 2003; Topcu et al., 2004; Elleder et al., 1997a; Mole et al., 2005).

The *CLN6* gene was simultaneously identified in a subset of Costa Rican and Venezuelan patients, and in the naturally occurring *nclf* mouse model (Gao et al., 2002; Wheeler et al., 2002). The encoded peptide comprises a putative transmembrane protein (TM) of unknown function (Gao et al., 2002; Wheeler et al., 2002). CLN6 localizes to the ER in both neuronal and non-neuronal cells by means of two ER retention signals, one in the first 49 N-terminal aa residues and the second in the segment comprising TM6-TM7 (Mole et al., 2004; Heine et al., 2007; Teixeira et al., 2006).

Although no evidence of a major founder effect exists for the 48 known *CLN6* mutations (http://www.ucl.ac.uk/ncl/cln6.shtml), three of them are common in certain populations: the nonsense p.Glu72X in Costa Rica, p.Ile154del in Portugal, and p.Val91GlufsX42 in Newfoundland (Wheeler et al., 2002; Gao et al., 2002; Teixeira et al., 2003; Moore et al., 2008). Mutations in *CLN6* do not have an impact on the normal distribution of the protein or its ability to dimerize (Mole et al., 2004; Kurze et al., 2010). Instead, they are postulated to exert their pathogenic effect on the stability and function of the mutant polypeptides, justified by the reduced rate of synthesis and stability of the CLN6 mutants compared to wild-type peptides (Kurze et al., 2010).

#### *2.6.3.4 CLN7 disease*

CLN7 disease comprises the third variant form of LINCL. The CLN7 disease typically starts between 2-6 years of age, with any of seizures, ataxia, visual failure, or developmental regression as the presenting symptoms (Topcu et al., 2004; Siintola et al., 2007). In advanced disease stages the CLN7 subtype clinically resembles CLN2 disease, though in earlier stages the two entities differ in the severity of seizures which are more aggravated in CLN7 patients (Topcu et al., 2004; Siintola et al., 2007). The majority of patients lose ambulation on average two years after disease onset, and can survive at maximum until the third decade of life (Topcu et al., 2004; Siintola et al., 2007). Defects in *MFSD8* are always associated with condensed FP (Siintola et al., 2007; Topcu et al., 2004). In addition to the FP findings on EM examination a mixture of FP with CL and/or GRODs can also be identified (Aiello et al., 2009).

The disease is caused by mutations in *MFSD8* (NM_152778.2; Siintola et al., 2007). Since *MFSD8* was originally identified in a subset of mainly Turkish patients, this subtype was originally referred to as the Turkish variant (Topcu et al., 2004; Siintola et al., 2007). The encoded protein is predicted to be transmembrane and anchored to the lysosomal membrane by 12 TMs (Siintola et al., 2007; Steenhius et al., 2010). Sequence homology analyses denote a predicted transporter function, as MFSD8 is a member of the major facilitator superfamily (MFS) of transporter proteins having an MFS domain (MFS_1) at positions p.42_477 and a sugar transporter domain (Sugar_tr) at positions p.72_147

(Siintola et al., 2007). Nevertheless, the substrate and precise role of MFSD8 remain unknown to date.

Of the 14 mutations detected in patients with CLN7 disease (http://www.ucl.ac.uk/ncl/cln7.shtml) almost all are private for the families in which they have been identified (Siintola et al., 2007; Aiello et al., 2009; Stogmann et al., 2009; Aldahmesh et al., 2009). Only the splice-site affecting mutation c.863+3_4insT has occurred on an almost identical genetic haplotype in three unrelated patients from Italy that most likely share a common ancestor (Aiello et al., 2009). Since none of the mutations detected results in changes of the lysosomal distribution the primary defects are likely to arise from impaired functional properties (Siintola et al., 2007).

#### *2.6.3.5 CLN8 disease*

CLN8 disease, EPMR (progressive epilepsy with mental retardation), otherwise also known as Northern epilepsy, is caused by mutations in *CLN8* (NM_018941.3; Ranta et al., 1999). The missense mutation p.Arg24Gly, representing a founder effect, was identified in 22 Finnish patients (Ranta et al., 1999; Hirvasniemi et al., 1994). In tandem with the findings in Finnish patients, a homozygous frameshift-causing mutation (c.267_268insC) identified in *Cln8* in the motor neuron degeneration mouse (*mnd*) established the gene's disease association (Ranta et al., 1999). CLN8 disease, EPMR was linked to the group of NCLs after the detection of CL bodies in the cells of affected individuals (Herva et al., 2000). Clinically, the patients have disease onset between 5-10 years of age, usually manifesting with tonic-clonic seizures and psychomotor decline, but do not develop myoclonus (Ranta et al., 1999).

The major phenotype associated with mutations in *CLN8* is CLN8 disease, late infantile variant, first described in a subset of Turkish patients (Mitchell et al., 2001; Topcu et al., 2004). The disease onset is around 2-7 years of age with seizures and/or motor skill decline (Ranta et al, 2004; Topcu et al., 2004), after which myoclonus, cognitive decline, speech impairment, ataxia, and visual failure ensue (Ranta et al, 2004; Topcu et al., 2004). The disease progresses very rapidly with patients developing the whole spectrum of

clinical symptoms within two years of disease onset (Topcu et al., 2004). The ultrastructural inclusions form condensed FP or a mixed pattern of FP with CL, RL and occasionally GRODs, in which the main accumulating protein component can be SCMAS or SAP A and D (Topcu et al., 2004; Cannelli et al., 2006; Palmer et al., 1989; Herva et al., 2000; Ezaki and Kominami, 2004).

CLN8 encodes a putative transmembrane protein with five TMs (Ranta et al., 1999). CLN8 is targeted to the ER and/or ER-Golgi intermediate compartment (ERGIC) of non-neuronal cells via a C-terminal ER retention signal (283-KKPR-286), which causes Golgi trapping when disrupted by mutations (Ranta et al., 1999; Lonka et al., 2000). In neurons the localization of CLN8 is somewhat different, with the protein being targeted very close to the plasma membrane (PM; Lonka et al., 2004). This differential localization between neuronal and non-neuronal cells is thought to arise from the existence of unknown vesicular transport mechanisms in the neuronal cells (Lonka et al., 2004). The TLC domain (extending over the region p.62-262) of CLN8 readily makes it a member of the TLC protein family (TRAM-LAG1-CLN8), which is postulated to have a role in sensing, biosynthesis, and metabolism of lipids or protection of proteins from proteolysis, (Winter and Ponting, 2002). Alternatively, CLN8 has been implied to predispose to cell survival, since it increases resistance of neurons to damage caused by toxic stimuli (Vantaggiato et al., 2009).

The sequence of *CLN8* is affected by 16 different mutations detected in patients from six different countries (http://www.ucl.ac.uk/ncl/cln8.shtml). Interestingly, only missense and deletion mutations have been detected in this gene. None of the missense mutations tested impaired the intracellular distribution of CLN8, suggesting that the primary defects result in disturbed functional properties rather than disturbed trafficking within ERGIC (Lonka et al., 2000; Lonka et al., 2004; Vantaggiato et al., 2009).

### 2.6.4 NCL subtypes with juvenile onset

#### *2.6.4.1 CLN3 disease*

Clinical onset in CLN3 disease, classic juvenile, is between 5 and 10 years of age. The hallmarks of CLN3 disease comprise progressive visual loss, which is usually the leading clinical symptom, and vacuolated lymphocytes in the peripheral blood as the main pathological finding (Santavuori, 1988; Zeman, 1976). Seizures follow the initial visual symptoms, while early in the second decade of life the children develop adaptive and motor disabilities leading to ataxia and motor decline (Santavuori, 1988). In the mid-to late stages of the disease patients develop dementia and behavioral abnormalities including hallucinations and schizophrenia-like behavior. Affected individuals die prematurely at an average age of 20-30 years (Järvelä et al., 1997). The autofluorescent material accumulating in the cells of patients with CLN3 disease, classic juvenile, usually form FP (Zeman, 1976; Santavuori, 1988). However, occasionally CL, or a mixture of FP, CL, RL and/or GRODs, can be detected upon EM analysis (Mole et al., 2005). The protein accumulating in the lipopigment inclusions is SCMAS (Palmer et al., 1992).

A collaborative effort identified *CLN3* (NM_000086.2) as the gene defective in CLN3 disease, classic juvenile (The International Batten Disease Consortium, 1995). The encoded peptide comprises a type III lysosomal transmembrane protein with six TM domains, and it is the most highly conserved of all the NCL proteins, having orthologous genes present from yeast to humans (The International Batten Disease Consortium, 1995; Phillips et al., 2005; Kyttälä et al., 2004). CLN3 is targeted to the membrane of the endosomal and lysosomal vesicles, travelling through the indirect trafficking pathway (Ezaki et al., 2003; Haskell et al., 2000; Järvelä et al., 1999; Järvelä et al., 1998; Mao et al., 2003). CLN3 sorting is mediated via two lysosomal targeting motifs, an unconventional signal found in the second cytoplasmic loop and a dileucine motif at the C-terminal cytosolic domain of the protein (Haskell et al., 2000; Kyttälä et al., 2004; Storch et al., 2007). In neurons, however, the distribution of CLN3 is in the synaptic vesicles (Haskell et al., 2000).

To date 46 different mutations have been associated with CLN3 disease, the majority of which produce abnormally short polypeptides (http://www.ucl.ac.uk/ncl/cln3.shtml). One of the reported mutations (c.460-280_677+382del967), commonly referred to as the 1.02 kb deletion, is found in 81% of CLN3 disease chromosomes (The International Batten Disease Consortium, 1995). Despite the implication of CLN3 in several different pathways such as autophagy, apoptosis, membrane trafficking and endocytic events, the precise function of CLN3 remains elusive, making it difficult to evaluate the impact of the mutations on the resultant peptides (Lane et al., 1996; Cao et al., 2006; Chang et al., 2007b; Uusi-Rauva et al., 2008; Luiro et al., 2004). Although some deletion mutations result in retention of the mutants in the ER, the missense mutations evaluated do not interfere with normal protein distribution, suggesting that the mutations are loss of function (Järvelä et al., 1999; Haskell et al., 2000).

#### *2.6.4.2 CLN9 disease*

CLN9 disease (MIM# 609055) represents a provisional assignment for a variant form of JNCL that remains to be identified. This subtype was described after exclusion of NCLs by means of genetic testing of the known genes and of other lysosomal storage disorders through enzyme level testing in two sisters from Serbia and two brothers from Germany. The clinical phenotype of both familial cases was similar to the CLN3 disease, classic juvenile, starting with visual failure and seizures. The disease onset is at 4 years of age and death occurs between ages 15 and 20 years (Schulz et al., 2004). Other clinical symptoms associated with this subtype involve cognitive decline, speech deterioration, neuropsychiatric symptoms, ataxia and loss of ambulation (Schulz et al., 2004). On EM analysis the characteristic NCL lipopigment inclusions detected are a combination of GRODs with CL and the protein component accumulating in them is SCMAS (Schulz et al., 2004).

Although the gene and protein underlying this subtype remain elusive, mass spectrometry analyses have revealed a possible link to sphingolipid metabolism which is perturbed in the patients' fibroblasts (Schulz et al., 2006).

### 2.6.5 NCL subtypes with adult onset: CLN4 disease

CLN4 disease can be divided into two overlapping and highly similar phenotypes: Kufs disease (MIM# 204300) and Parry disease (MIM# 162350). In both forms of ANCL the onset is usually in the third decade of life (ranging from 10-51 years of age). Although the disease progression is much slower compared to the other NCL subtypes, the patients die approximately 12 years after disease onset (Haltia, 2003). ANCL is not associated with visual decline. On biopsy examination the storage bodies identified form membrane-bound GRODs, FPs, CL and/or RL, enclosing SCMAS or SAP A and D proteins (Haltia, 2003).

Kufs disease shows a recessive mode of inheritance and can be further divided into two categories depending on the presented symptoms. Kufs type A is mainly characterized by PME, dementia, ataxia, and late pyramidal and extrapyramidal symptoms. Only recently, *CLN6* was identified as the underlying cause of Kufs type A, with nine different mutations in 10 unrelated families (Arsov et al., 2011). The majority of the mutations associated with the milder Kufs type A are missense. Mutations resulting in premature protein truncation were only identified in compound heterozygosity with "milder" missense mutations, suggesting that the resultant "milder" phenotype could arise from residual CLN6 function. Nevertheless, it is not until the function of CLN6 has been resolved that it can be tested whether the later onset and less severe Kufs type A is associated with mild alleles, or whether modifying alleles account for the phenotypic differences (Arsov et al., 2011). Type B on the other hand, usually starts with behavioral problems such as depression and progressive dementia, followed by ataxia, motor decline and extrapyramidal symptoms (Berkovic et al., 1988).

Parry disease is clinically similar to Kufs disease with the only difference being in the mode of inheritance, which in case of Parry disease is autosomal dominant (Goebel and Braak, 1989). Parry disease is characterized by genetic heterogeneity, with some clinical cases carrying mutations in *DNAJC5* and others remaining genetically unresolved (Noskova et al., 2011). Two distinct *DNAJC5* mutations (p.Leu116del and p.Leu115Arg) were identified in heterozygous state in five different families. The encoded cysteine-string protein α (SCPα) is a component of an enzymatically active chaperone complex tethered

to the synaptic vesicles and ensuring the correct folding of several proteins with synaptic function and relevance (Noskova et al., 2011). The *DNAJC5* genetic defects described affect conserved dileucine residues within the cysteine-string domain of the protein and are believed to exert their deleterious effects via mistargeting of the mutant proteins and via reducing their palmitoylation efficiency. The defective proteins in turn lead to depletion of SCPα, presynaptic dysfunction and inability to exhibit their neuroprotective role (Noskova et al., 2011).

### 2.6.6 Atypical phenotypes caused by mutations in known NCL genes

In addition to the typical phenotypes described in sections 2.6.1-2.6.5, each of the eight known NCL genes has also been reported to associate with atypical phenotypes (Table 3 and Figure 2). These can involve earlier or later disease onset, more protracted or more severe disease course, manifestation of some symptoms and not of others. This is exemplified in CLN6 disease, presented above, which can result in both a late infantile onset phenotype and a much less severe adult onset form. Additionally, many cases of inter- and intra-familial variation in disease severity have been described in the NCL disorders. How mutations in the same genes can result in strikingly different phenotypes has not been explained, but it is thought that additional factors such as unidentified genetic determinants in genes acting as modifiers of phenotype, or even perhaps environmental factors, may play a role in this phenomenon.

**Table 3.** Atypical phenotypes caused by the NCL genes

| Gene | Typical phenotype | Atypical phenotypes | Onset | Morphology of lipopigment inclusions |
|---|---|---|---|---|
| *PPT1* | CLN1 disease, classic infantile | CLN1 disease, late infantile | 1.5-3 years | GROD, FP |
| | | CLN1 disease, juvenile | 2-10 years | GROD, CL, RL, FP |
| | | CLN1 disease, adult | 20-38 years | GROD |
| *TPP1* | CLN2 disease, classic late infantile | CLN2 disease, infantile | 1-9 months | CL |
| | | CLN2 disease, juvenile | 6-10 years | CL, FP |
| *CLN3* | CLN3 disease, classic juvenile | CLN3 disease, infantile | 5 months | n.d. |
| | | CLN3 disease, protracted | 5-9 years | FP, CL, GROD |
| *CLN5* | CLN5 disease, late infantile variant | CLN5 disease, infantile | 4 months | FROD, FP |
| | | CLN5 disease, juvenile | 4-9 years | FP, CL |
| | | CLN5 disease, adult | 17 years | GROD, FP, CL, RL |
| *CLN6* | CLN6 disease, late infantile variant | CLN6 disease, adult or CLN6 disease, Kufs type A | 17-51 years | FP, GROD |
| *MFSD8* | CLN7 disease, late infantile variant | CLN7 disease, juvenile | 11 years | n.d. |
| *CLN8* | CLN8 disease, late infantile variant | CLN8 disease, EPMR | 5-10 years | n.d. |
| *CTSD* | CLN10 disease, congenital | CLN10 disease, late infantile | early school age | GROD |

**Abbreviations used:** n.d.: not defined, EPMR: progressive epilepsy with mental retardation

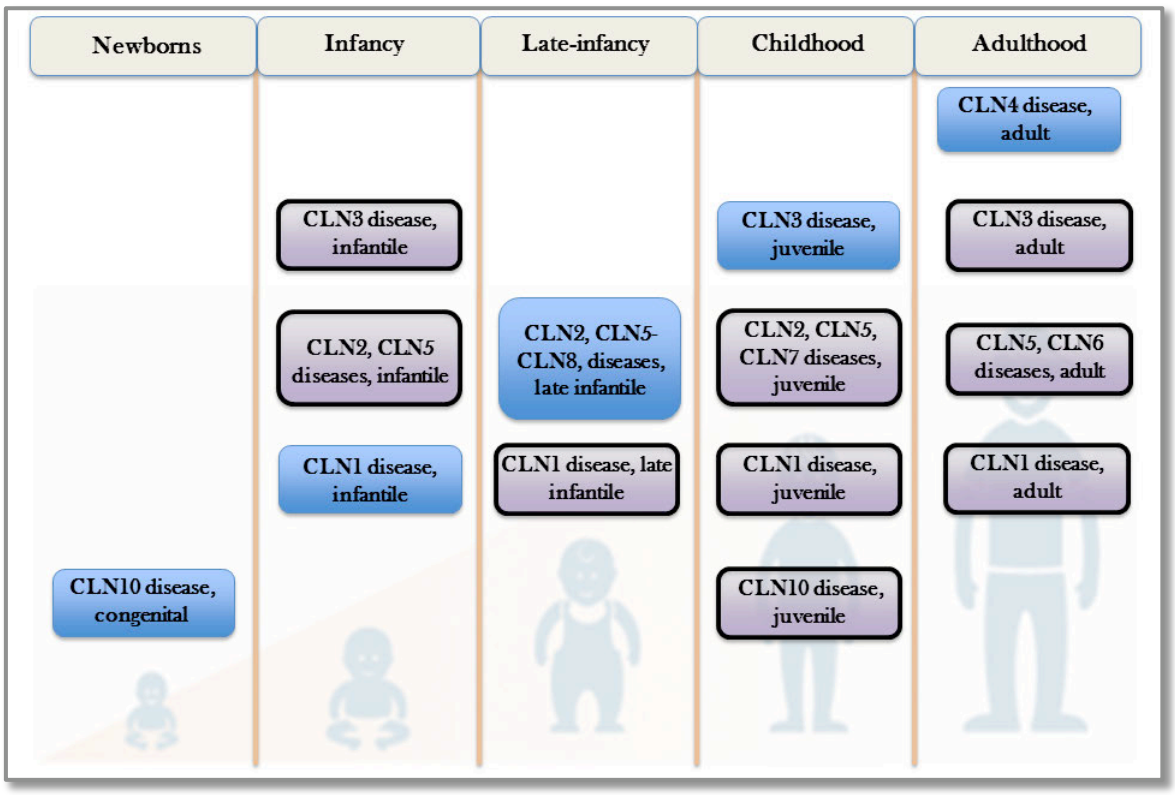


**Figure 2.** Differential diagnosis of NCLs. The classical phenotypes reproduced by mutations in each of the eight NCL-causing genes are shown in the blue boxes. The atypical phenotypes are shown in the light purple boxes. Disease subtypes are divided here according to the age of disease onset.

### 2.6.7 Animal models with NCL-like phenotypes

Animal models exist for all subtypes of NCL disorders except for CLN4 and CLN7 disease (http://www.ucl.ac.uk/ncl/animal.shtml). These are either spontaneously occurring or engineered and they have been described in organisms ranging from the single celled yeast to larger animal models such as sheep and dog models. Although the two mouse models developed for CLN1 disease recapitulate key features of the disorder (*Cln1*$^{-/-}$, Gupta et al., 2001; and *Ppt1*$^{\Delta ex4}$, Jalanko et al., 2005), the models developed in the nematode worm (Porter et al., 2005) and the fruit fly (Hickey et al., 2006) have not been as successful. In CLN2 disease only two animal models have been described, the generated knock-out mouse model (*Cln2*$^{-/-}$, Sleat et al., 2004) and a naturally occurring dog model (Awano et al., 2006a). Only engineered models exist for CLN3 disease, which comprise four mouse models which accurately replicate the NCL pathology (*Cln3*$^{-/-}$,

Mitchison et al., 1999; Katz et al., 1999; *Cln3*$^{\Delta ex7/8}$, Cotman et al., 2002; Eliason et al., 2007) a worm model, which does not exhibit the typical NCL characteristics (de Voer et al., 2005), and yeast models (*btn1*, Pearce and Sherman, 1997; Gachet et al., 2005). In CLN5 disease the engineered *Cln5*$^{-/-}$ mouse model (Kopra et al., 2004) and spontaneous models occurring in the Border Collie dog, the Borderdale sheep, and Devon cattle (Melville et al., 2005; Houweling et al., 2006; Frugier et al., 2008), reproduce phenotypes similar to that described in patients with this subtype. CLN6 disease is modeled in the naturally occurring *nclf* mouse, New Zealand South Hampshire, and Merino sheep, and the Australian Shepherd dog (Gao et al., 2002; Wheeler et al., 2002; Jolly et al., 1989; Tammen et al., 2006; Katz et al., 2011). A spontaneous mouse model also exists for CLN8 disease (*mnd*, Ranta et al., 1999). In addition to the *mnd* model, CLN8 disease is also modeled in the naturally occurring English Setter dog (Katz et al., 2005). Finally, several animal models exist for CLN10 disease, including the mouse *Ctsd*$^{-/-}$ (Saftig et al., 1995), and models in larger animals such as sheep and dogs (Tyynelä et al., 2000; Awano et al., 2006b).

### 2.6.8 Trafficking and localization of the NCL proteins

Four of the NCL associated proteins are soluble lysosomal enzymes while the remaining five transmembrane proteins reside in the lysosomes, ER, endosomal vesicles, or PM (Figure 3) (reviewed by Mole et al., 2005). The compartments where NCL proteins reside belong to the eukaryotic endomembrane system, of which there are three major pathways through which proteins traffic towards their final destination: the secretory pathway, the endocytic pathway and the lysosomal pathway. Which specific pathway a protein will follow in the cell depends on the targeting information carried on each newly synthesized peptide. Correct protein targeting is of primary importance for the function of the various organelles, as the properties that each of them displays are dependent on their protein composition.

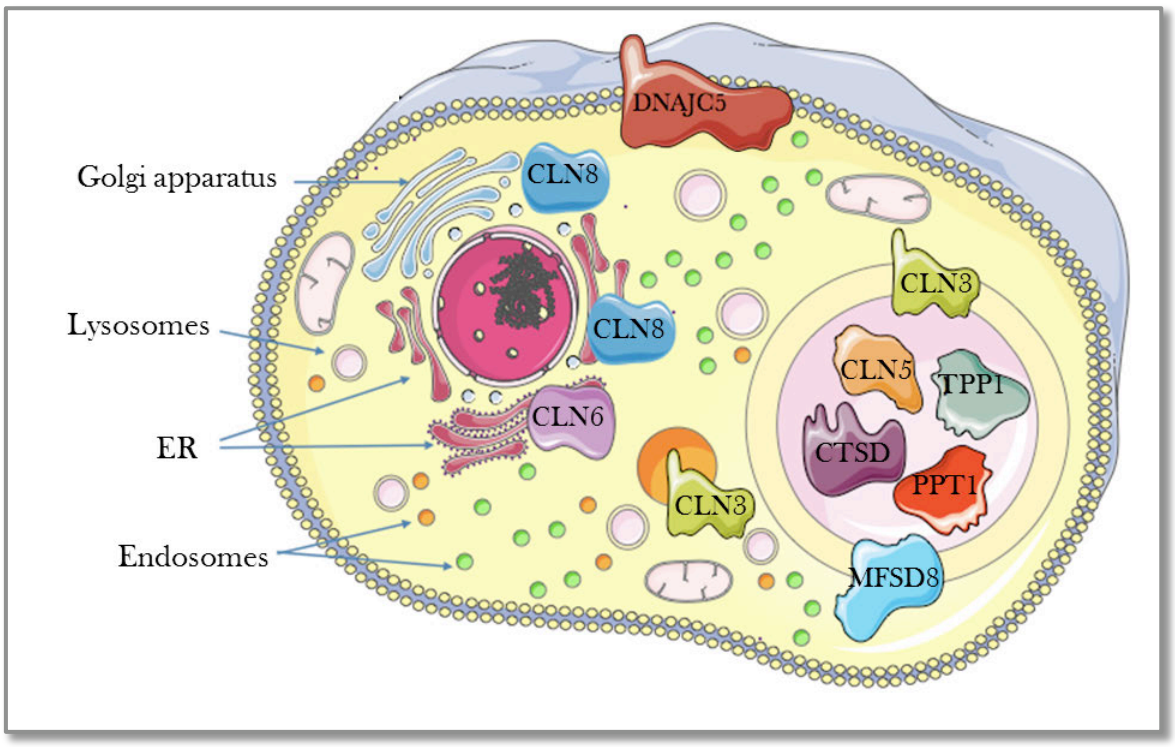


**Figure 3.** Localization of the NCL-associated proteins. The figure was produced using Servier Medical Art (http://www.servier.com/servier-medical-art/powerpoint-image-bank).

### *2.6.8.1 The secretory pathway*

For some proteins translation progresses and finishes in the cytosol where the proteins are eventually released. This does not mean, however, that these peptides will remain in the cytosol. If there are signals on their peptide sequences that mediate direction to the mitochondria or nucleus these are recognized by receptor proteins on the appropriate organelle.

For proteins that will enter the secretory pathway, the ribosomes that mediate their translation initially move freely in the cytoplasm and at a later point during protein synthesis become bound to the rough ER. This is mediated by an N-terminal ER signal peptide found in the newly synthesized polypeptide chain (Walter and Johnson, 1994). Within the ER lumen the proteins receive the necessary modifications in order to become functional, which involve the formation of disulfide bonds, glycosylation (Kornfeld and Kornfeld, 1985; Marth, 1996), proteolytic cleavage of propeptides and folding (Martoglio

and Dobberstein, 1998). At specific ER exit sites the proteins are packaged in transport vesicles composed of the coat protein II (COPII) (Barlowe, 1998) which subsequently fuse to form the ERGIC (Hauri and Schweizer, 1992; Saraste and Kuismanen, 1992). At the ERGIC the protein transportation can be bidirectional (Figure 4): there are the proteins in the COPII vesicles that move along the secretory pathway towards the Golgi compartment (anterograde transport) but there is also retrograde transportation of the ER recycling proteins packaged within the differently coated COPI vesicles that move from the Golgi towards the ER (Warren and Mellman, 1999; Aridor et al., 1995). Whether the proteins will follow the anterograde or retrograde route depends on signals existing in their peptide sequence (Kuehn and Schekman, 1997). If no such signals can be identified the peptides are secreted. After having arrived at the ERGIC the encapsulated proteins move to the *cis*-Golgi cisternae, from which they then move across the Golgi stack along membrane tubules (Thyberg and Moskalewski, 1999). Enzymes residing in the Golgi apparatus are responsible for the further modification of the travelling proteins, with the major type of modification being glycosylation. It is these modifications that provide the signals that determine the proteins' final destination in the cell. The final sorting of the proteins takes place in the *trans*-Golgi (TGN) cisterna, where proteins with the same signals are grouped together in order to be targeted to the same intracellular location. The TGN subsequently buds off into clathrin coated vesicles that migrate to their final destination, such as the PM, the lysosomes or the endosomes. The adaptor complex molecules that recognize the transport vesicles determine the route that will be followed towards the final destination (van Vliet et al., 2003). One option is targeting via the direct route, which involves transportation from the TGN to the early and late endosomes and finally the lysosomes. The second route is the indirect one, where the vesicles are targeted to the PM and from there enter the endocytic pathway that will lead them to the final destination.

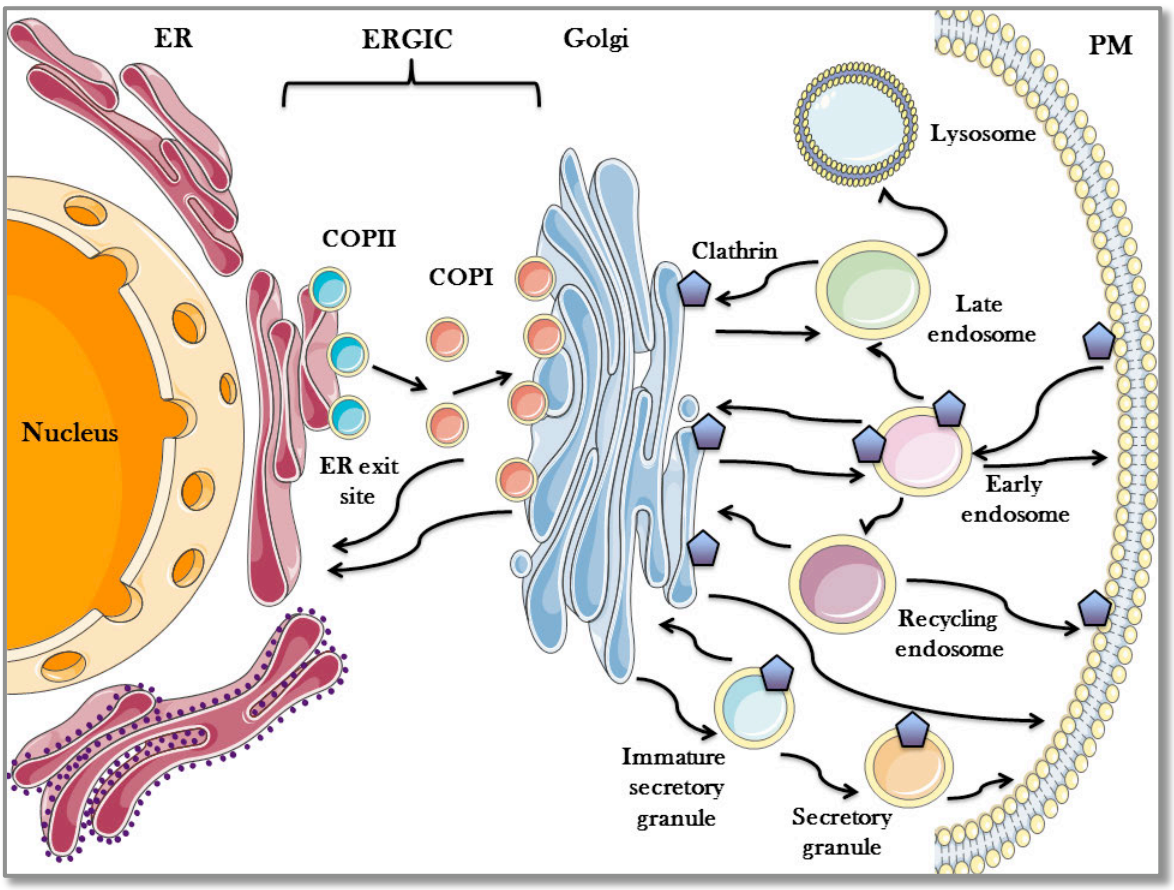


**Figure 4.** Schematic representation of the secretory and endocytic pathways. The COPI and COPII coated vesicles are shown as red and blue circles, respectively. The clathrin coated transport vesicles are shown as purple pentagons. The name of each organelle is given either in the upper part of the figure or beside the respective structures. Anterograde and retrograde transports are depicted with the black arrows. The figure was produced using Servier Medical Art (http://www.servier.com/servier-medical-art/powerpoint-image-bank).

### *2.6.8.2 The endocytic pathway*

The endocytic pathway is used for the uptake of dissolved solutes, macromolecules and other material from the extracellular space. The molecules to be internalized bind to membrane receptors found at specific coated sites along the PM. Following this interaction, clathrin-coated vesicles bud off from the PM and enter the endocytic pathway. The first stop are the early endosomes with which the transport vesicles fuse. The acidic pH (pH = 6.0-6.8) of the early endosomes causes the cargo and the receptors

to dissociate (Mellman et al., 1986). The receptors are then packaged in recycling vesicles that are transported back to the PM where they can be reused. The second stop is the late endosomes (pH = 5.5) where the even lower pH mediates the beginning of the breaking down of the material that has been internalized. This process is finalized and completed in the lysosomes (pH = 4.6), the third destination of the endocytic pathway. Delivery of the late endosomal contents to the lysosomes can be achieved in two ways. The first involves the fusion of the late endosomes with the lysosomes that results in the formation of a hybrid organelle, and the second the formation of a transient pore that connects the two organelles which remain intact in this case, a process called kiss-and-run fusion (Luzio et al., 2005). In addition to the digestion of exogenous macromolecules and material the endogenous pathway can also perform protein sorting functions. These involve the packaging of proteins in endosome-derived recycling or secretory vesicles that pinch off from the endosomes and are targeted either back to the PM or to the TGN where the endocytic and secretory pathways converge (Figure 4).

#### *2.6.8.3 Lysosomal targeting*

There are two major pathways through which the lysosomal proteins are transported to their target organelle. Soluble lysosomal hydrolases usually follow the Man6P-dependent pathway, while the lysosomal transmembrane proteins are more likely to follow the Man6P-independent pathway.

The Man6P-independent pathway: In transmembrane proteins destined to the lysosomes targeting is mediated by short, linear arrays of amino acid residues consisting of the sorting signals, usually situated at the cytosolic domains of the peptides (Bonifacino and Traub, 2003). Two major classes of lysosomal sorting signals are recognized today, known as "tyrosine-based" and "dileucine-based" motifs. The "tyrosine-based" signals conform to the consensus motif YXXΦ, in which Y is a tyrosine, X can be any amino acid and Φ is an amino acid with a bulky hydrophobic side chain (Canfield et al., 1991). The "dileucine-based" motifs can be either [DE]XXXL[IL] or DXXLL (Bonifacino and Traub, 2003). These sorting signals are recognized by clathrin adoptor protein (AP)

complexes that bind directly to the cytoplasmic tails containing these motifs, selecting the protein for inclusion into the appropriate transport vesicles (Pearse et al., 1988). AP complexes are heterotetrameric adaptor proteins (AP-1, AP-2, AP-3, and AP-4) and when in combination with the structural protein clathrin and various accessory factors form the clathrin coats (Bonifacino and Traub, 2003). AP-2 complexes are associated with transport vesicles that participate in endocytosis from the PM. The AP-1 complex has been localized to the TGN and mediates TGN to endosome transport (Reusch et al., 2002). The AP-3 complex on the other hand is implicated in endosome to lysosome, TGN to endosome or TGN to lysosome protein transport (Reusch et al., 2002).

The Man6P-dependent pathway: For the soluble lysosomal proteins targeting is mediated by specific modifications received while these hydrolases move across the Golgi apparatus. More specifically Man6P groups are added to the lysosomal hydrolases in a two step process involving the enzymes N-acetylglucosamine-1-phosphotransferase and N-acetylglucosamine-1-phosphodiester-α-N-acetylglucosaminidase (Little et al., 1987; Reitman and Kornfeld, 1981). Upon arrival to the TGN the proteins carrying Man6P moieties are recognized by MPRs. The MPRs can be either small (~46 kilodaltons (kDa) and dependent on divalent cations (cation-dependent; CD_MPRs) in order to efficiently recognize the Man6P groups, or bigger (~300 kDa) and cation-independent (CI_MPRs) (Hofllack and Kornfeld, 1985a; Hofllack and Kornfeld, 1985b). Once the MPRs have recognized the soluble lysosomal proteins the complexes are recognized by AP-1 complexes and golgi-localizing γ-adaptin ear homology domain ARF-binding proteins (GGAs), mediating packaging in transport vesicles coated by clathrin (le Borgne et al., 1996; Meyer et al., 2000; Puertollano et al., 2001). Subsequently the vesicles follow the direct intracellular route towards the lysosomes. Before delivery to the lysosomes the complexes pass from the endosomes, in the lumenal acidic pH of which the protein-receptor complexes dissociate, and the MPRs are transported back to the TGN.

# 3. AIMS OF THE STUDY

This thesis aimed to dissect the molecular genetic background underlying childhood onset PME in a cohort of 250 patients, with a particular interest in the NCL disease subtypes. Towards this, the specific aims were:

**Aim I:** To update the mutational spectrum underlying NCLs and provide clinically significant genotype-phenotype correlations for the different NCL subtypes.

**Aim II:** To evaluate the role of *MFSD8* as a disease-causing gene in patients from populations other than the Turkish population.

**Aim III:** To obtain insight into the primary functional properties of MFSD8 by characterizing its cellular and tissue expression pattern and by elucidating the sorting mechanisms that mediate lysosomal targeting.

**Aim IV:** To identify novel loci associated with childhood onset PME and identify the disease-causing gene(s).

# 4. MATERIALS AND METHODS

## 4.1 Patients and controls

Altogether 250 patients from 236 families were evaluated in this study (Table 4 and Figure 5). The majority of the subjects (185 patients from 178 families) were of Turkish origin. The study material was received from clinical collaborators either as blood or DNA samples. Samples from parents were available for 43 cases. A fibroblast cell line for RNA extraction was available from one of the patients and a control individual. All samples were collected after informed consent was obtained from the patients or their legal guardians, in accordance with the legislation governing local institutional review boards.

**Table 4.** List of patients screened for this thesis

| Ethnicity | Number of patients | Number of families |
|---|---|---|
| Albania | 1 | 1 |
| Bulgaria | 2 | 2 |
| Cook Islands | 1 | 1 |
| Croatia | 3 | 3 |
| Czech Republic | 6 | 6 |
| Denmark | 1 | 1 |
| Finland | 1 | 1 |
| Greece | 1 | 1 |
| India | 8 | 6 |
| Italy | 2 | 2 |
| Mexico | 1 | 1 |
| Morocco | 2 | 1 |
| New Zealand | 3 | 2 |
| Pakistan | 7 | 5 |
| Poland | 3 | 3 |
| Roma from the former Czechoslovakia | 17 | 16 |
| Sweden | 1 | 1 |
| The Netherlands | 1 | 1 |
| Turkish | 185 | 178 |
| UK | 3 | 3 |
| USA | 1 | 1 |

The majority of patients had disease onset between 1-4 years of age. All presented with epileptic seizures and psychomotor deterioration. EM analysis for the detection of storage material was available for 30 patients. A total of 26 samples (25 patients and one healthy family member) were genotyped in a SNP genomewide scan for identification of novel PME-associated genes (study IV; Figure 5). Selection of the families to be included in this study was based on consanguinity of the marriages and how informative the family structure was. Skin biopsies for EM analysis were available for three of these patients.

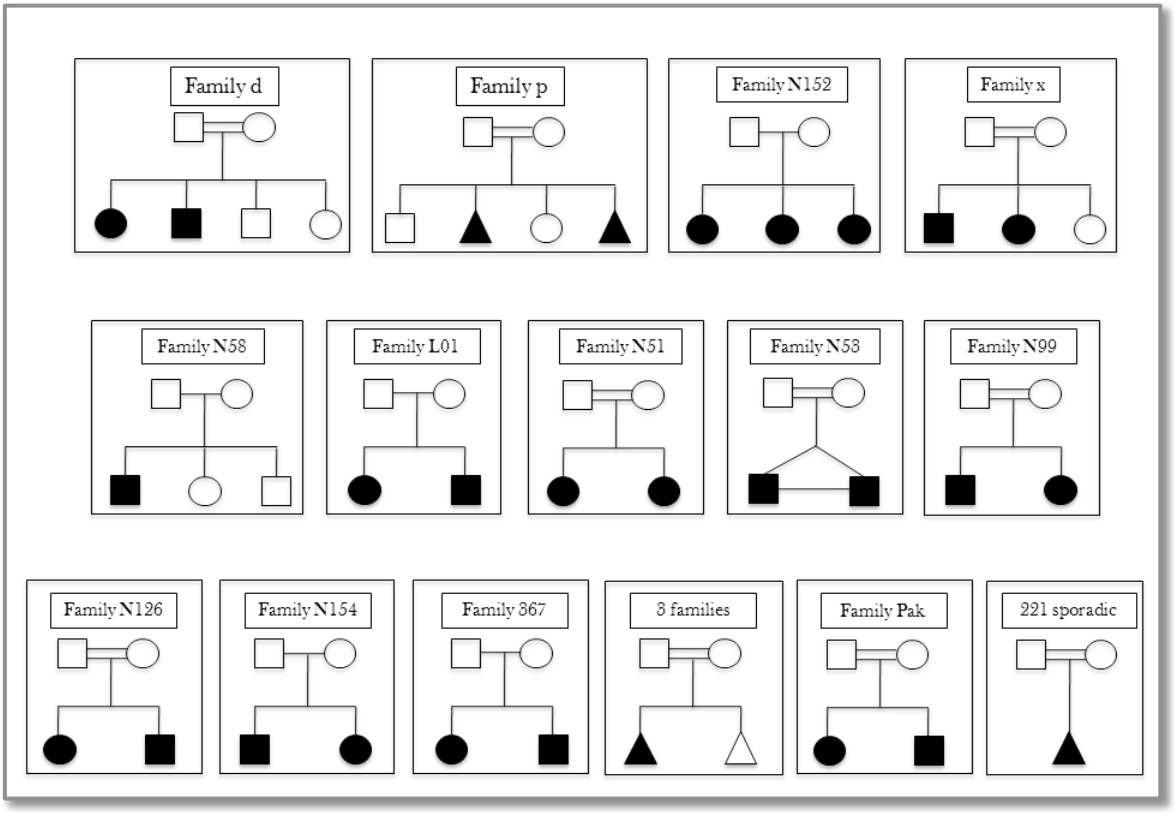


**Figure 5.** Families and sporadic patients studied throughout this thesis

Three sets of healthy control individuals were used in this study:

- Set 1: 160 Turkish individuals
- Set 2: 96 individuals from the Centre d' Etude du Polymorphisme Humain (CEPH: www.ceph.bio.fr).
- Set 3: 100 Indian control individuals.

For the RT-PCR analyses total RNA was extracted from a control fibroblast cell line.

## 4.2 Methods used

A list of the methods used in this thesis is given in Table 5. A short description of each of the techniques is also presented herein. Detailed descriptions of the protocols and methodology can be retrieved from the original publications.

### 4.2.1 Nucleic acids extraction and purification (I, II, IV)

Genomic DNA was extracted from the leukocytes of EDTA-blood samples (Puregene DNA Purification kit, Gentra Systems) according to the instructions provided by the manufacturer. Total RNA was extracted from fibroblast cells using the RNeasy mini kit (Qiagen). Concentrations of the extracted nucleic acids were determined using a spectrophotometer (GeneQuant Pro, Amersham Biosciences), or a NanoDrop ND-1000 (Thermo Fisher Scientific).

**Table 5.** List of methods used during this thesis

| Method | Original publication |
|---|---|
| Cell culture | II, III, IV |
| Co-immunoprecipitation | III |
| Confocal microscopy | III, IV |
| Deglycosylation assay | III |
| Direct sequencing analyses | I, II, III, IV |
| DNA extraction | I, II, IV |
| Genomewide SNP scan | IV |
| GST pulldown | III |
| Haplotype analysis | I, II, IV |
| Homozygosity mapping | IV |
| Immunofluorescence microscopy | II, III, IV |
| Immunofluorescence staining(s) | II, III, IV |
| Immunohistochemistry | IV |
| *In silico* analyses | I, II, III, IV |
| *In situ* hybridization | III |
| Linkage analysis | II, IV |
| Microsatellite genotyping | II, IV |
| Mutation analysis | I, II, III, IV |
| Polymerase chain reaction (PCR) | I, II, III, IV |
| Real-time PCR | III |
| Recombinant DNA techniques (molecular cloning) | II, III, IV |
| Reverse transcriptase PCR (RT-PCR) | II, III |
| RNA extraction from fibroblasts | II |
| Site-directed mutagenesis | II, III, IV |
| Targeted next-generation sequencing | Not published |
| Transient transfections | II, III, IV |
| Western blot analysis | III, IV |

### 4.2.2 PCR-based amplification and mutation analysis (I-IV)

The coding exons and exon-intron boundaries of the genes evaluated for mutations were PCR-amplified using standard techniques. PCR primers were designed with the ExonPrimer (http://ihg.gsf.de/ihg/ExonPrimer.html) and/or Primer3 (http://frodo.wi.mit.edu/primer3/) programs. The amplified PCR products were prepared with ExoSAP-IT (USB, Cleveland) and sequenced using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) and an ABI 3730 DNA Analyzer, or an ABI Prism 3100-Avant (Applied Biosystems). Sequence chromatograms were visualized with Sequencher 4.8 (Gene Codes Corporation).

In RT-PCR, extracted RNA was reverse transcribed using the M-MLV Reverse Transcriptase (Promega, Madison). For evaluation of the splicing patterns (study I) the cDNA portions of interest were amplified with exonic primers and sequenced. For real-time RT-PCR experiments cDNAs were labeled with Power SUBR Green PCR Master Mix (Applied Biosystems) and quantified using the comparative threshold cycle method.

### 4.2.3 Microsatellite genotyping (II)

Three or four fluorescently-labeled microsatellite markers were PCR-amplified around each of the *CLN1*, *CLN2*, *CLN3*, *CLN5*, *CLN6*, *CLN7*, *CLN8*, and *CLN10* loci (as described in Siintola et al., 2005). The results were analyzed with GENEMAPPER v4.0 software (Applied Biosystems). The CEPH 1347-2 sample was used as a reference for the allele sizes that were estimated. The haplotypes were constructed manually and homozygosity over each locus was evaluated.

### 4.2.4 Genomewide SNP scan and next-generation sequencing (IV and unpublished)

The genomewide analysis was performed at the Institute for Molecular Medicine Finland (FIMM) using the Human610-Quad DNA Analysis BeadChip SNP array (Illumina), featuring 550,000 SNPs per genotyped sample, as specified by manufacturer's

instructions. All patients had a genotype success rate greater that 95%, which was set as the threshold of acceptance and thus all were included in subsequent analyses.

Targeted next-generation sequencing was performed using the NimbleGen Sequence Capture array protocol, followed by subsequent Illumina re-sequencing of the coding and flanking intronic regions. Altogether eight chromosomal loci spanning a genomic region of 27,7 Mb were mapped. To screen the 484 genes and approximately 5,000 coding exons located within the candidate regions a total of 5,900,970 capture probes were custom designed according to the NimbleGen Sequence Capture Microarrays protocol (Roche). All designed probes were reviewed using the UCSC Genome browser (http://genome.ucsc.edu) to verify coverage of the regions of interest. Next-generation sequencing was performed at the Wellcome Trust Sanger Institute in Hixton, Cambridge.

### 4.2.5 Homozygosity mapping (IV)

The genomewide SNP data was analyzed for homozygosity mapping using the PLINK program (v1.06). Only windows of 500 kb of uninterrupted homozygosity were considered, where a maximum of 5 SNP calls were allowed to be missing and 2 to be heterozygous per sample. The patients were then grouped based on whether they shared homozygosity runs over the genome to determine the candidate loci. Illumina's Beadstudio-suite was used as a second method to visualize and confirm long homozygous regions in the patients genotyped.

### 4.2.6 Expression constructs and site-directed mutagenesis (II, III, IV)

Two tagged MFSD8 constructs were used for localization studies. The first comprised $^{HA}$MFSD8 (described by Siintola et al., 2007), and the second hCLN7, generated by PCR amplifying the full-length *MFSD8* and cloning it in the pEGFP-C2 expression vector C-terminally to the tag (Clontech). Similar to hCLN7, an EGFP tagged construct of mouse *Mfsd8* was created (mCLN7).

In study III two sets of CD8 chimeric proteins were constructed to address trafficking of MFSD8. First the N-terminal (p.1-40) and C-terminal (p.504-518) cytosolic tails of MFSD8 were amplified from a human brain cDNA library and cloned in a CD8-containing pBluescript vector (CD8-pBluescript), replacing the cytosolic segment of CD8, generating the CD8-CLN7Nterm and CD8-CLN7Cterm constructs, respectively, which were then subcloned into a pcDNA3.1(+) expression vector. For the CLN7Nterm-CLN3LIAA+$MX_9G$ chimera the traffic deficient full-length CLN3 (CLN3LIAA+ $MX_9G$; described by Kyttälä et al., 2004) was used as a PCR template to amplify the p.39-438 portion of CLN3 (CLN3dm-Δ1-38), which was cloned in the pcDNA3.1 plasmid. In a second cloning step the MFSD8 N-terminus (p.1-40) was PCR amplified and cloned in-frame at the N-terminus of CLN3dm-Δ1-38, giving rise to a chimera having the N-terminus of CLN3 replaced by that of MFSD8 (CLN7-CLN3dm-Δ1-38).

For the production of the GST fusion proteins the MFSD8 N-terminus (p.1-40) and the C-terminus (p.504-518) were cloned into a pGEX4T-1 vector (GST-CLN7Nterm and GST-CLN7Cterm respectively; Amersham Biosciences).

The complete open reading frame of *KCTD7* was cloned in-frame into the aminoterminal hemagglutinin (HA) tag containing pAHC expression vector (a pCI-neo derivative; kindly provided by Prof. Tomi Mäkelä, University of Helsinki, Finland) to create the $wt^{HA}$KCTD7 construct. A second construct was created by cloning the KCTD7 cDNA 5' to the HA-tag, generating the wtKCTD7$^{HA}$.

The mutations introduced in the template constructs described are presented in Table 6. All mutations were generated using the QuickChange site-directed mutagenesis kit (Stratagene), according to the manufacturer's instructions. All constructs were verified by sequence analysis.

**Table 6.** Mutant constructs used for cellular work

| Template construct | Mutation introduced | Mutant construct generated | Use | Study |
|---|---|---|---|---|
| wt$^{HA}$MFSD8 | p.Arg139His | $^{HA}$MFSD8p.R139H | Localization | I |
| wt$^{HA}$MFSD8 | p.Thr294Lys | $^{HA}$MFSD8p.T294K | Localization | I |
| wt$^{HA}$MFSD8 | p.Arg465Trp | $^{HA}$MFSD8p.R465T | Localization | I |
| wt$^{HA}$MFSD8 | p.Leu13Ala | $^{HA}$MFSD8p.L13A | Trafficking | II |
| wt$^{HA}$MFSD8 | p.Ile25Ala | $^{HA}$MFSD8p.I25A | Trafficking | II |
| wt$^{HA}$MFSD8 | p.Ile444Ala | $^{HA}$MFSD8p.I444A | Trafficking | II |
| wt$^{HA}$MFSD8 | p.Tyr503Ala | $^{HA}$MFSD8p.Y503A | Trafficking | II |
| wt$^{HA}$MFSD8 | p.Tyr513Ala | $^{HA}$MFSD8p.Y513A | Trafficking | II |
| $^{HA}$MFSD8p.L13A | p.Tyr513Ala | $^{HA}$MFSD8p.L13A+Y513A | Trafficking | II |
| $^{HA}$MFSD8p.L13A+Y513A | p.Ile25Ala | $^{HA}$MFSD8p.L13A+I25A+Y513A | Trafficking | II |
| $^{HA}$MFSD8p.L13A+Y513A | p.Ile444Ala | $^{HA}$MFSD8p.L13A+I444A+Y513A | Trafficking | II |
| $^{HA}$MFSD8p.L13A+Y513A | p.Tyr503Ala | $^{HA}$MFSD8p.L13A+Y503A+Y513A | Trafficking | II |
| hCLN7 | p.Gly52Arg | hCLN7p.G52R | Trafficking | II |
| hCLN7 | p.Tyr121Cys | hCLN7p.Y121C | Trafficking | II |
| hCLN7 | p.Ala157Pro | hCLN7p.A157P | Trafficking | II |
| hCLN7 | p.Pro412Leu | hCLN7p.P412L | Trafficking | II |
| hCLN7 | p.Pro447Leu | hCLN7p.P447L | Trafficking | II |
| hCLN7 | p.Leu13_Leu14 delinsAlaAla | hCLN7p.LL13/14AA | Trafficking | II |
| hCLN7 | p.Tyr503Ala | hCLN7p.Y503A | Trafficking | II |
| hCLN7 | p.Tyr513Ala | hCLN7p.Y513A | Trafficking | II |
| hCLN7p.LL13/14AA | p.Tyr503Ala | hCLN7p.LL13/14AA+Y503A | Trafficking | II |
| hCLN7p.LL13/14AA | p.Tyr513Ala | hCLN7p.LL13/14AA+Y513A | Trafficking | II |
| hCLN7 | p.Asn171Gln | hCLN7p.N171Q | Glycosylation | II |
| hCLN7 | p.Asn371Gln | hCLN7p.N371Q | Glycosylation | II |
| hCLN7 | p.Asn376Gln | hCLN7p.N376Q | Glycosylation | II |
| hCLN7p.N371Q | p.Asn376Gln | hCLN7p.N371Q+N376Q | Glycosylation | II |
| mCLN7 | p.Asn372Gln+ p.Asn377Gln | mCLN7p.N372Q+N377Q | Glycosylation | II |
| mCLN7p.N372Q+N377Q | p.Asn389Gln | mCLN7p.N372Q+N377Q+N389Q | Glycosylation | II |
| mCLN7 | p.Leu13_Leu14 delinsAlaAla | mCLN7p.LL13/14AA | Trafficking | II |
| CD8-CLN7Cterm | p.Tyr513Ala | CD8-CLN7Cterm_p.Y513A | Trafficking | II |
| CLN7- CLN3dm-Δ1-38 | p.Leu13Ala | CLN7p.L13A-CLN3dm-Δ1-38 | Trafficking | II |
| GST-CLN7Nterm | p.Leu13Ala | GST-CLN7Nterm_p.L13A | GST pull-down | II |
| wt$^{HA}$KCTD7 | p.Arg94Trp | $^{HA}$KCTD7p.R94W | Localization | IV |
| wt$^{HA}$KCTD7 | p.Asp115Tyr | $^{HA}$KCTD7p.D115Y | Localization | IV |
| wt$^{HA}$KCTD7 | p.Asn273Ile | $^{HA}$KCTD7p.N273I | Localization | IV |

### 4.2.7 Cell culture and transfections (II, III, IV)

Human cervical cancer cells (HeLa), african green monkey kidney cells (COS-1), and human embryonic kidney cells (HEK293) were cultured in Dulbecco's modified Eagle medium (DMEM; Gibco) and supplemented with 10% fetal calf serum (FCS), antibiotics and 1xGlutaMAX (Gibco). Mouse fibroblasts deficient for subunits of the AP-1 (Δμ1A) or AP3 adaptor protein complexes (AP3-βA) were cultivated in DMEM supplemented with 20% FCS, and antibiotics. BHK cells were cultured in Glasgow minimal essential medium (GMEM; Gibco), supplemented with 10% FCS, antibiotics, 1xGlutaMAX, and 5% tryptose phosphate broth (Sigma-Aldrich). Cells were trypsinized and plated on 6-well plates 24 hours (h) prior to transfection. The cells were transfected with 1 μg of wild-type or mutant constructs using either the FuGENE® 6 (Roche), or Lipofectamine 2000 (Invitrogen) transfection reagents, according to the manufacturer's guidelines. Whenever inhibition of protein production was required 50 μg/ml cyclohexamide (Chx) was added to the cells 2 h prior to fixation. The cells were fixed 18 h post transfection either with 4% paraformaldehyde (PFA) for 15 minutes (min) or with 100% ethanol (EtOH) for 5 min.

To obtain the cultures of embryonic hippocampal neurons the hippocampi of SV129-J (Jackson Laboratories) embryos were dissected at embryonic day 14 (E14) or 17-18 (E17-18). The tissue was trypsinized and dissociated mechanically with a Pasteur pipette and the cells plated in 6-well poly-d-lysine (Sigma)-coated plates containing complete Neurobasal medium supplemented with antibiotics, 1xB27 (Gibco), and glutamine.

### 4.2.8 Immunofluorescence stainings (IF) and microscopy (II, III, IV)

After fixation, the cells were blocked and permeabilized by incubating them in phosphate-buffered saline (PBS) supplemented with 0.2% saponin and 0.5% bovine serum albumin (BSA), or PBS supplemented with 0.1% Triton X-100 and 0.5% BSA, for 30 min at room temperature (RT). For IF primary antibody solutions were prepared in PBS (0.2% saponin/0.5% BSA or 0.1% Triton X-100/0.5% BSA). Fixed cells were incubated with the primary antibodies (Table 7) overnight or for 1 h at RT, after which excess antibodies were removed by washing the cells three times with PBS. Secondary antibodies (Table 7) diluted in PBS (0.2% saponin/0.5% BSA or 0.1% Triton X-

100/0.5% BSA) were subsequently applied to the cells for 1 h at RT. After washing off the secondary antibodies with three PBS washes, the coverslips were mounted on object glasses using Gel Mount (Sigma-Aldrich) or GluoromountG (Southern Biotechnology Associates).

To visualize the immunostainings an Axioplan 2 microscope, a confocal Zeiss LSM 510 Meta microscope, a Leica DMR confocal microscope, or a Nikon Eclipse TE-2000 microscope was used, each of which was equipped with a CCD camera.

### 4.2.9 Western blot (III, IV)

Samples for western blot analyses were prepared by lysing the cells in ice cold 1x lysis buffer supplemented with protease inhibitors. The cell lysates were collected after having scraped the cells, left to rotate at +4°C for 10 min, and centrifuged to separate the soluble proteins from the membranes and cell debris. To obtain the cell lysates from mouse cerebellum and liver the tissues were homogenized for 30 min in lysis buffer using a FastPrep (FP120) or a Beckman TLA 100.3 rotor, respectively. The protein concentrations were determined using the Bradford assay according to the instructions provided by the manufacturer. Equalized protein concentrations were prepared by adding 3x sample buffer containing a reducing agent (β-mercaptoethanol or dithiothreitol (DTT)) required for size separation of proteins and the denaturing detergent sodium dodecyl sulfate (SDS). Following sample preparation the proteins were heated at 95°C for 5 min and loaded on 14% SDS-PAGE gels. Separation of the bands was achieved by running the gel at 100 V for 2.5-3 h after which they were transferred onto a PVDF membrane (Millipore) using blotting buffer. The membranes were blocked overnight at 4°C and incubated with primary and secondary antibodies for 1 h at RT (Table 7). The antigens were detected using enhanced chemiluminescence (ECL) (Amersham Biosciences).

**Table 7.** Antibodies used in IF, western blot, and immunohistochemistry analyses

| Antibody | Organism | Dilution | Company / Reference | Target | Study |
|---|---|---|---|---|---|
| | | | *Immunofluorescence experiments* | | |
| H4A3 | mouse | 1:100 | Developmental Studies Hybridoma Bank | Lysosomes | I, II, IV |
| a-HA | rabbit | 1:500 | Santa Cruz Biotechnology | HA tag | I, II, IV |
| a-HA | mouse | 1:250 | Santa Cruz Biotechnology | | IV |
| a-CD8 | mouse | 1:200 | | Plasma membrane | II, IV |
| m385 | rabbit | 1:300 | Luiro et al., 2001 | CLN3 aa 242-258 | II |
| a-GFP | mouse | 1:1000 | Roche Molecular Biochemocals | GFP tag | II |
| a-AGA | rabbit | 1:400 | Halila et al., 1991 | Lysosomes | II |
| 1D4B | rat | 1:100 | Developmental Studies Hybridoma Bank | Lysosomes | II |
| 1D4B | rat | 1:500 | BD Biosciences | Lysosomes | II |
| a-KCTD7 | rabbit | 1:500 | Sigma-Aldrich | KCTD7 aa 2-51 | IV |
| a-CTSD | rabbit | 1:100 | Dako | Lysosomes | IV |
| EEA1 | mouse | 1:100 | BD Biosciences | Endosomes | IV |
| Giantin | rabbit | 1:1000 | BioSite | Golgi | IV |
| PDI | mouse | 1:50 | Stressgen | ER | IV |
| β-tubulin | mouse | 1:200 | Sigma-Aldrich | Cytoskeleton | IV |
| a-mouse Cy2 | donkey | 1:200 | Jackson ImmunoResearch | Secondary Ab | I, II |
| a-rabbit Cy3 | goat | 1:200 | Jackson ImmunoResearch | Secondary Ab | I, II |
| a-mouse Alexa 488 | | 1:200 | Molecular Probes | Secondary Ab | IV |
| a-rabbit Alexa 594 | | 1:200 | Molecular Probes | Secondary Ab | IV |
| | | | *Western blot* | | |
| AP-1γ | mouse | | BD Biosciences | AP-1 | II |
| AP-$2^{\alpha}$ | mouse | | BD Biosciences | AP-2 | II |
| δ-adaptin | mouse | | Peden et al., 2001 | AP-3 | II |
| a-KCTD7 | rabbit | 1:1000 | Sigma-Aldrich | KCTD7 aa 2-51 | IV |
| β-tubulin | mouse | 1:1000 | Sigma-Aldrich | Load control | IV |
| HRP-actin | | 1:1000 | Cell Signaling | Load control | IV |
| HRP a-mouse | | 1:3000 | Dako | Secondary Ab | IV |
| HRP a-rabbit | | 1:3000 | Dako | Secondary Ab | IV |
| | | | *Immunohistochemistry* | | |
| a-KCTD7 | rabbit | 1:500 | Sigma-Aldrich | KCTD7 aa 2-51 | IV |
| CDC47 | mouse | 1:100 | NeoMarkers | Neuronal cell progenitors | IV |
| GFAP | mouse | 1:100 | Dako | Astrocytes | IV |
| F4/80 | mouse | 1:50 | Serotec | Microglia | IV |
| NeuN | mouse | 1:100 | Chemicon | Neurons | IV |
| PV | mouse | 1:1000 | Swant | GABAergic neurons | IV |
| SYP | | 1:200 | Dako | Pre-synapses | IV |
| PSD-95 | | 1:100 | Transduction Laboratories | Post-synapses | IV |
| GABA | | 1:500 | Synaptic Systems | GABA transporter | IV |
| a-mouse Alexa 488 | | 1:200 | Molecular Probes | Secondary Ab | IV |
| a-rabbit Alexa 594 | | 1:200 | Molecular Probes | Secondary Ab | IV |

### 4.2.10 Co-immunoprecipitation assays (III)

For the GST pull-down experiments the recombinant GST fusion-tagged proteins described in section 4.2.6. were produced in *E. coli* BL21 bacteria. The bacterial pellets were lysed with STE buffer supplemented with lysozyme, sarcosyl and protease inhibitors, sonicated, and ultracentrifuged at 10,000 rpm for 10 min at 4°C and supplementated with 2% Triton X-100 to release the proteins. The fusion proteins were affinity-purified by binding to glutathione-Sepharose 4B beads overnight at 4°C (Amersham Biosciences), and subsequently incubated with a HeLa or mouse liver cell lysate for 4 h at 4°C.

In biotinylation experiments HEK293 cells were biotinylated by incubation with sulfo-NHS-SS-biotin (Pierce) twice for 20 min. PBS supplemented with 100 mM glycine was used to remove the unbound biotin after which the cells were lysed as described in section 4.2.9. The soluble proteins collected were affinity-purified by incubation with streptavidin-agarose beads (Fluka) for 2 h at 4°C.

Complexes of bead-bait proteins were eluted by adding 2x sample buffer, after which the samples were separated by 10% SDS-PAGE and analyzed by western blotting using appropriate antibodies (Table 7).

### 4.2.11 Deglycosylation assay (III)

Cell lysates were collected from HEK293 cells transfected with hCLN7p.N171Q, hCLN7p.N371Q, hCLN7p.N376Q, hCLN7p.N371Q+N376Q, mCLN7p.N372Q+N377Q, and mCLN7p.N372Q+N377Q+N389Q (Table 6). Each of the cell lysates was subsequently incubated overnight at 4°C with 1250 U peptide *N*-glycosidase F (PNGaseF) (New England Biolabs) to remove mannose N-glycans. After PNGaseF treatment the samples were separated in 10% SDS-PAGE gels and detected with antibodies targeted against the green fluorescent protein (GFP) tag of hCLN7 and mCLN7.

### 4.2.12 Immunohistochemistry (III, IV)

Paraffin embedded sagittal sections (5 µm) of P5, P7, P10, P14, 1 month, 2 month, and 4 month old SV129-J mice (Jackson laboratories) were deparaffinized by incubation in xylene and rehydrated by incubation in decreasing series of EtOH solutions. The tissue endogenous antigens were retrieved through 10 min heating at 95°C in 10 mM citrate buffer. Incubation of the sections with PBS supplemented with 1% FCS for 1 h at RT was used to block unspecific antibody binding. Primary antibody solutions were prepared in PBS (1% FCS) and left to incubate with the sections overnight at 4°C. Secondary antibodies were prepared in PBS (1% FCS) and applied for 1 h at RT. Prior to mounting with Gel Mount (Sigma-Aldrich), the sections were washed with $dH_2O$ to remove the PBS salts. The stainings were visualized with an Axioplan 2 microscope and the images were obtained using AxioVision 3.1 (Zeiss).

### 4.2.13 In situ hybridization (III)

Rat brains were removed and fixed in isopentane at -30°C. A cryostat was used to obtain sagittal and coronal sections at -20°C. The sections were postfixed with 3.7% PFA, dehydrated by incubation with increasing series of EtOH solutions and allowed to air dry. A mix of four [$^{35}$S]dATP labeled antisense oligonucleotides suspended in hybridization medium (Helios Biosciences, France) was applied to the sections overnight in a 42°C incubator. The next day the slides were washed, dried, and exposed to BAS-TR Fuji Imaging screens (Fuj Film Photo Co.) for 14 days. The screens were scanned using a BAS-5000 Fuji Bioimaging Analyzer.

### 4.2.14 In silico analyses (I-IV)

The genome browsers used to retrieve the positions and information on the genes and to determine the positional candidates within the candidate loci were the UCSC Human Genome Browser (http://genome.ucsc.edu/cgi-bin/hgGateway), the Ensembl Human Genome Browser (http://www.ensembl.org/Homo_sapiens/Info/Index), and the NCBI Gene Browser (http://www.ncbi.nlm.nih.gov/gene/).

Alignment of the primer oligos to the human genome to ensure product specificity was performed with UCSC Blat (http://genome.ucsc.edu/cgi-bin/hgBlat?command=start&org=Human&db=hg19&hgsid=206980099). Identification of protein paralogs and/or orthologs was done with NCBI protein-protein BLAST (http://www.ncbi.nlm.nih.gov/blast/Blast.cgi). Nucleotide or peptide sequences were aligned either with the MAFFT v5.8 programme (http://align.bmr.kyushu-u.ac.jp/mafft/online/server/) or with ClustalX (http://www.clustal.org/clustal2/). Alternative transcript variants were visualized with AceView (http://www.ncbi.nlm.nih.gov/).

Prediction programs PolyPhen (http://genetics.bwh.harvard.edu/pph/), PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/), SNPs3D (http://snps3d.org/), and SIFT (http://blocks.fhcrc.org/sift/SIFT.html) were used to evaluate the impact of the missense mutations identified on the resultant peptides.

Putative sorting signals, glycosylation sites, and phosphorylation sites were predicted by the programes SOSUI (http://bp.nuap.nagoya-u.ac.jp/sosui/sosui_submit.html), PSORT (http://psort.hgc.jp/) and PhosphoSitePlus (http://www.phosphosite.org/homeAction.do), respectively.

# 5. RESULTS AND DISCUSSION

## 5.1 Mutations in six NCL genes (I, II)

In order to provide molecular genetic diagnosis to the patients included in this study, each was screened for mutations in *PPT1*, *TPP1*, *CLN3*, *CLN5*, *CLN6*, *MFSD8*, *CLN8* and *CTSD*. This comprised an essential step to dissect the genetic defects in the patients with no mutations in the known NCL genes and identify novel childhood onset symptomatic epilepsy genes in study IV.

A total of 35 novel mutations were identified in *PPT1*, *TPP1*, *CLN3*, *CLN5*, *CLN6*, and *CLN8*. (Table 8). The 14 novel mutations detected in *MFSD8* alone are described in detail in section 5.2.1. (Table 9). An additional 56 mutations (described in detail in the original publication of study I), were contributed by collaborators. When these newly described mutations are combined with the 260 previously published mutations the total number of known NCL-causing mutations rises to a total of 364. The novel mutations are mostly evenly spaced across the coding and splice junction sequence of each of the genes evaluated, suggesting that no obvious mutation hot spots exist. The exception was *TPP1*, where the region spanning the first two coding exons was reported to be mutation free. It was hypothesized that since this region is not present in the mature peptide due to cleavage, any mutations occurring within it would not have a relevant impact on the polypeptide produced (Zhong et al., 2000). The splice-site affecting mutation c.17+1G>C in intron 1, identified in study II, together with mutations c.18-3C>G and c.37dupC identified by collaborators in intron 1 and exon 2, are contrary to the hypothesis of *TPP1* exons 1 and 2 being mutation-free. It is noteworthy that all three mutations are predicted to either have an effect on the downstream sequence producing frameshifts, or result in aberrant proteins that are likely to be targeted for nonsense-mediated decay. In light of these findings, it can be concluded that TPP1 missense mutations would possibly be tolerated and not cause the disease phenotype because these single amino acid changes would not be present in the mature peptide.

**Table 8.** Mutations identified in *PPT1*, *TPP1*, *CLN3*, *CLN5*, *CLN6*, and *CLN8* in studies I and II

| Position | Nucleotide change | Amino acid change | Number of patients | Ethnicity | Study |
|---|---|---|---|---|---|
| | | ***PPT1*** | | | |
| Exon 1 | c.114G>T | p.Trp38Cys | 3 | Turkey | I, II |
| Exon 4 | c.413C>T | p.Ser138Leu | 1 | Turkey | I |
| Exon 6 | c.538dupC | p.Leu180ProfsX9 | 1 | Turkey | I |
| Exon 6 | c.566C>G | p.Pro189Arg | 1 | Turkey | I |
| | | ***TPP1*** | | | |
| Intron 1 | c.17+1G>C | splice defect | 1 | Turkey | II |
| Exon 3 | c.184T>A | p.Ser62Thr | 1 | Turkey | I |
| Exon 5 | c.497dupA | p.His166GlnfsX22 | 1 | Turkey | I |
| Exon 10 | c.1204G>T | p.Glu402X | 2 | Turkey | I, II |
| Exon 11 | c.1444G>C | p.Gly482Arg | 1 | Turkey | II |
| Exon 12 | c.1497delT | p.Gly501AlafsX18 | 1 | Turkey | I |
| | | ***CLN3*** | | | |
| Intron 2 | c.126-1G>A | splice defect | 1 | Turkey | I |
| Exon 4 | c.233_234insG | p.Thr80AsnfsX12 | 1 | Turkey | I |
| Exon 14 | c.1067T>G | p.Leu356Arg | 1 | Turkey | unpublished |
| | | ***CLN5*** | | | |
| Exon 1 | c.61C>T | p.Pro21Ser | 1 | Turkey | I |
| Exon 1 | c.223T>C | p.Trp75Arg | 3 | Turkey | I |
| Exon 3 | c.524T>G | p.Leu175X | 1 | Turkey | I |
| Exon 3 | c.593T>C | p.Leu198Pro | 1 | Turkey | I |
| Exon 4 | c.726C>A | p.Asn242Lys | 2 | Turkey | I |
| Exon 4 | c.1026C>A | p.Tyr342X | 1 | Roma from the former Czechoslovakia | II |
| | | ***CLN6*** | | | |
| Exon 1 | c.34G>A | p.Ala12Thr | 1 | Turkey | I |
| Exon 1 | c.49G>A | p.Gly17Ser | 1 | Turkey | I |
| Exon 3 | c.270C>G | p.Asn90Lys | 1 | India | I |
| Exon 4 | c.476C>T | p.His157Arg | 1 | Turkey | II |
| Exon 5 | c.516T>A | p.Tyr172X | 1 | Turkey | I |
| Exon 6 | c.557T>C | p.Phe186Ser | 1 | Turkey | I |
| Exon 7 | c.775G>A | p.Gly259Ser | 3 | India | I |
| | | ***CLN8*** | | | |
| Exon 2 | c.209G>A | p.Arg70His | 1 | New Zealand | I |
| Exon 2 | c.227A>G | p.Gln76Arg | 1 | Turkey | I |
| Exon 2 | c.320T>G | p.Ile107Ser | 1 | Turkey | I |
| Exon 2 | c.374A>G | p.Asn125Ser | 1 | Turkey | I |
| Exon 2 | c.415C>T | p.His139Tyr | 1 | New Zealand | I |
| Exon 2 | c.470A>G | p.His157Arg | 1 | Turkey | II |
| Exon 3 | c.637_639delTGG | p.Trp213del | 1 | Turkey | I |
| Exon 3 | c.661G>A | p.Gly221Ser | 1 | Turkey | I |
| Exon 3 | c.806A>T | p.Glu269Val | 1 | Turkey | I |

The majority of the novel mutations are private, being identified in a single family. The latter has diagnostic implications since it emphasizes that the design of a diagnostic test

suitable for several patients within one or among several populations is not possible. Instead, molecular genetic testing involving sequencing of the complete region of each of the genes associated with the suspected subtype remains the only possibility to establish a clinical diagnosis.

## 5.2 Genetic and molecular characterization of *MFSD8*, the gene defective in CLN7 disease

Prior to this thesis, a homozygosity mapping approach in a group of mainly Turkish patients identified mutations in *MFSD8*, accounting for the CLN7 disease that is a variant form of LINCL (Siintola et al., 2007). The encoded peptide was reported to be a member of the MFS based on sequence homology analysis, and to be a putative transmembrane protein localizing to the lysosomes.

### 5.2.1 Mutations identified in MFSD8 (I, II, III)

Altogether 211 patients from 200 families were screened for mutations in *MFSD8*. Of these 146 were Turkish and 65 of various other ethnic origins. In 25 patients (nine Turkish and 16 Roma) prior haplotype analysis indicated homozygosity over the *MFSD8* locus, whereas the remaining patients were screened without prior knowledge of the haplotype status. A total of 14 novel mutations were identified in 36 patients (Table 9; Figure 6). Seven of the novel mutations comprise single nucleotide changes resulting in missense mutations (c.416G>A; p.Arg139His, c.479C>A; p.Thr160Asn, c.479C>T; p.Thr160Ile, c.881C>A; p.Thr294Lys, c.1393C>T; p.Arg465Trp, c.1394G>A; p.Arg465Gln, and c.1408A>G; p.Met470Val). One in-frame deletion/insertion mutation of two bp (c.468_469delinsCC) results in the in-frame substitution p.Ala157Pro. One frameshift-causing mutation involving the deletion of 16 bp (c.627_643delGTATACAACACCAGTTT; p.Met209IlefsX3) was identified in exon 7. The two mutations c.103C>T, and c.1420C>T introduce premature termination codons producing the residue changes p.Arg35X, and p.Gln474X, respectively. Finally, three mutations (c.754+1G>A, c.863+1G>C, and c.1103-2delA) affect splice sites. In study I

three more mutations (c.259C>T; p.Gln87X, c.554-1G>C, and c.1373C>A; p.Thr458Lys) were reported by collaborators.

**Table 9.** Mutations identified in *MFSD8* in studies I and II

| Position | Nucleotide change | Amino acid change | Number of patients | Origin | Study |
|---|---|---|---|---|---|
| Exon 3 | c.103C>T | p.Arg35X | 3 | Turkey, Cook Islands | I, II |
| Exon 5 | c.416G>A | p.Arg139His | 1 | India | I, II |
| Exon 6 | c.468_469delinsCC | p.Ala157Pro | 1 | Netherlands | II |
| Exon 6 | c.479C>A | p.Thr160Asn | 1 | Turkey | I |
| Exon 6 | c.479C>T | p.Thr160Ile | 1 | Cook Islands | I |
| Exon 7 | c.627_643delGTATACAACACCAGTTT | p.Met209IlefsX3 | 1 | Italy | II |
| Intron 8 | c.754+1G>A | splice defect | 2 | Turkey | I |
| *Intron 8* | *c.754+2T>A* | *splice defect* | 7 | Czech Republic, Turkey | I, II |
| Intron 9 | c.863+1G>C | splice defect | 1 | Turkey | II |
| Exon 10 | c.881C>A | p.Thr294Lys | 20 | Turkey, Roma from the former Czechoslovakia, Czech Republic, Spain | I, II |
| *Exon 10* | *c.929G>A* | *p.Gly310Asp* | 1 | Turkey | I |
| Intron 11 | c.1103-2delA | splice defect | 1 | Czech Republic | II |
| *Exon 12* | *c.1235C>T* | *p.Pro412Leu* | 1 | Mexico | I |
| Exon 13 | c.1393C>T | p.Arg465Trp | 1 | Albania/Greece | II |
| Exon 13 | c.1394G>A | p.Arg465Gln | 1 | Turkey | I |
| Exon 13 | c.1408A>G | p.Met470Val | 1 | Turkey | I |
| Exon 13 | c.1420C>T | p.Gln474X | 1 | Turkey | I |

**Note:** Mutations represented in italics have been originally reported by others; c.754+2T>A was first reported by Siintola et al., 2007; c.929G>A by Aiello et al., 2009, and c.1235C>T by Aldahmesh et al., 2009.

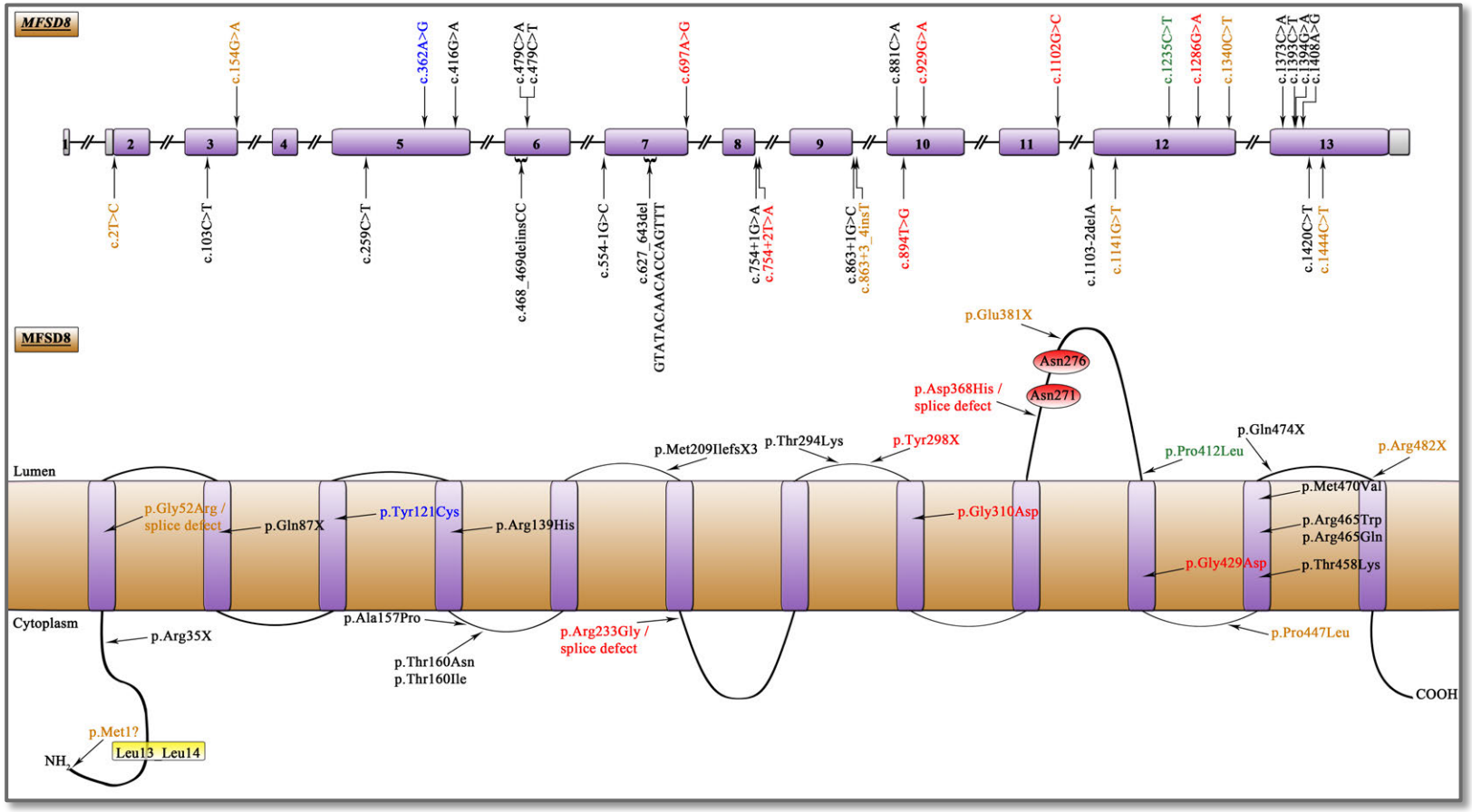


**Figure 6.** Spectrum of *MFSD8* mutations. In the schematic representation of the *MFSD8* gene in the upper half of the figure the purple boxes represent coding exons and are shown in scale. The grey boxes represent the untranslated regions and the black lines the introns and are not in scale. The missense mutations are shown in the above and other mutations in the below part of the gene. The putative structure of MFSD8 protein is shown in the lower half of the figure where the transmembrane domains are shown as purple boxes and the cytosolic and lysosomal facing domains as lines which are in scale. The brown rectangle depicts the lysosomal membrane. Mutations detected by this study are shown in black font, mutations reported by Siintola et al., 2007, are shown in red, those described by Aiello et al., 2009, Stogmann et al., 2009, and Aldahmesh et al., 2009, are shown in brown, blue and green, respectively. The residues at which N-glycosylation occurs are shown as red circles. The major lysosomal targeting motif is highlighted with a yellow square. Modified from Kousi et al., 2011.

In tandem with this study, eight more mutations were reported from three independent studies (Aiello et al., 2009; Stogmann et al., 2009; Aldahmesh et al., 2009), raising the total number of mutations in *MFSD8* to 31 (http://www.ucl.ac.uk/ncl/cln7.shtml). All mutations identified are evenly spread throughout the gene, revealing no obvious mutation hot spots (Figure 6).

In addition to the disease-causing mutations identified the deletion c.63-4delC in intron 2, which could not be unequivocally associated with the disease, was identified in heterozygous state in the Polish patient 450Pa and his father. The c.63-4delC change was not identified in 200 Turkish control chromosomes. Both the patient and his father had previously been found to carry another heterozygous missense change (c.1738G>A; p.Val580Met) in the chloride channel 6 (*CLCN6*) gene (Poet et al., 2006). Disruption of the *Clcn6* mouse ortholog causes a lysosomal storage condition characterized by the accumulation of lipofuscin and SCMAS (Poet et al., 2006), similar to NCLs. It was thus tempting to speculate the patient represented a case of digenic inheritance. However, the normal splicing pattern was maintained in the presence of c.63-4delC in RT-PCR analyses of RNA samples from the patient and a control individual, suggesting that the change is likely to be a rare private variant of benign or possibly disease-modifying role rather than the primary cause of the disease.

The nonsense and frameshift-causing mutations produce abnormally short polypeptides that are most likely targeted for nonsense-mediated decay. The true consequences of the splice affecting mutations cannot be unequivocally predicted due to the unavailability of RNA samples from the patients carrying them. It can be predicted nevertheless that they are likely to result in aberrant splicing, since they all affect invariable splice site residues, resulting either in aberrant proteins or peptides that are targeted for nonsense-mediated decay (Hentze and Kulozik, 1999). In the absence of a functional assay the impact that the missense mutations might have on MFSD8 is not clear. The fact that they all affect conserved residues in vertebrates, and that they are predicted to be probably or possibly damaging according to *in silico* prediction programs, are in favor of them being disease-associated. In localization studies none of the missense mutations identified from study II, and by Aiello et al., (p.Gly52Arg, p.Tyr121Cys, p.Arg139His, p.Ala157Pro,

p.Thr294Lys, p.Pro412Leu, p.Arg465Trp, and p.Pro447Leu) were found to interfere with normal distribution of MFSD8. This implies that these missense mutations are likely to induce disturbed functional properties rather than altered intracellular localization. Although the function of MFSD8 remains elusive, in light of these findings it is believed that the mutations causing CLN7 disease are loss of function.

The pathogenic role of the *MFSD8* mutations identified is further supported by the fact that they are absent from at least 200 Turkish or CEPH control chromosomes screened. Additionally, in all cases where parent samples were available the identified changes segregated with the disease phenotype.

### 5.2.2 Geographical distribution of MFSD8 defects (I, II)

CLN7 had long been thought to underlie the Turkish variant form of the disease. Identification of *MFSD8* as the gene underlying the CLN7 subtype in a group of Turkish and one Indian patients provided the first evidence suggesting that defects in this gene are not confined to a specific population alone (Siintola et al., 2007). The evaluation of 63 additional non-Turkish patients in studies I and II resulted in the description of a further 23 patients positive for *MFSD8* mutations. In tandem with this work, three studies reporting patients with *MFSD8* defects from Italy, Egypt and Saudi Arabia were published (Aiello et al., 2009; Stogmann et al., 2009; Aldahmesh et al., 2009). These studies, together with data from collaborators presented in study I, have collectively described *MFSD8* mutations in patients from 14 diverse ethnicities, establishing the worldwide occurrence of defects in this gene (Table 9). The worldwide distribution of defective alleles holds also for defects in most of the other known human NCL genes, except for *CTSD* and *DNAJC5* that have been studied in very few patients (study I; reviewed in Mole et al., 2005). One such example is the case of CLN8 disease EPMR, caused by the *CLN8* missense mutation p.Arg24Gly which was established by a founder effect in Finland and thought initially to be confined to the Finnish population (Ranta et al., 1999). Later genetic studies have identified defects in the same gene also in non-Finnish patients, such as Turkish and New Zealander (Table 8; Ranta et al., 2004;

http://www.ucl.ac.uk/ncl/cln8.shtml) confirming that none of the NCL-associated genes is population specific.

In one of the populations studied, the Roma patients originating from the former Czechoslovakia, the *MFSD8* missense mutation p.Thr294Lys was identified in homozygosity in 14/15 patients. Furthermore, haplotype analysis around the *CLN7* locus revealed a shared haplotype among all affected individuals, suggesting that a founder has established this mutation in the Roma population. A second common mutation is the splice-site affecting c.754+2T>A identified in seven out of nine Czech patients evaluated. Although haplotype analysis of polymorphic markers around the mutation locus was not performed it is likely that this mutation also represents a founder effect. Identification of these prevalent mutations in Roma and Czech patients is of crucial importance as this information can be used for the design of future diagnostic tests within these populations. Ultimately, diagnostic screening could help eliminate the incidence of the disease in these populations.

Prior to this study the group of Roma patients had been described to clinically resemble CLN6 disease (Elleder et al., 1997b). Linkage over the *CLN6* locus was confirmed with later haplotype analyses, in which it was further hypothesized that the Roma patients were likely to share a common ancestor with Indian patients, since the former had migrated from India in 1000AD (Sharp et al., 2003). The inability to detect *CLN6* mutations in the Roma patients, together with the detection of the *MFSD8* p.Thr294Lys in study II, imply that the reported linkage over the *CLN6* locus was probably due to the excess of homozygosity seen in such inbred populations. Subsequent genetic analyses in Indian patients have identified five private *CLN6* mutations in an equal number of families, suggesting that *CLN6* defects have not been established by a common ancestor in India (Sharp et al., 2003; Wheeler et al., 2002; Teixeira et al., 2003; Table 8). Moreover, the identification of two Indian patients with *MFSD8* mutations (Siintola et al., 2007; Table 9) provides further support that Indian LINCL was not established by a founder, but instead is genetically heterogeneous.

A second population characterized by genetic heterogeneity is the Turkish population. Originally, the Turkish variant had been proposed to be a distinct clinical entity, justified

by the lack of homozygosity across the then known NCL loci in a set of six consanguineous Turkish families (Wheeler et al., 1999). The identification of *CLN8* and *CLN6* mutations in subsets of Turkish patients provided the first evidence that the Turkish subtype is not caused by mutations in a single gene (Ranta et al., 2004; Siintola et al., 2005). Furthermore, in the study that identified *MFSD8* it became clear that unlike the Finnish subtype (caused by a single founder mutation in *CLN5*; Savukoski et al., 1998), a suspected founder effect in the Turkish population did not exist, justified by the identification of five different mutations in four families (Siintola et al., 2007). The report of 16 additional Turkish patients with private *MFSD8* mutations described here further highlights the genetic heterogeneity underlying NCL in this population (studies I, II). Today, defects in seven out of nine NCL-causing genes have been identified in Turkish patients, expanding our knowledge of the genetic heterogeneity characterizing this population (Tables 8, page 73 and 9, page 75). As the role of *DNAJC5* has not yet been studied in Turkish patients, it cannot be excluded that a fraction of NCL patients from this population will be reported to be positive for mutations in this gene in the future.

Despite the lack of detailed clinical descriptions, the broadly similar clinical phenotype of Turkish LINCL patients is caused by mutations in any of the *PPT1*, *TPP1*, *CLN3*, *CLN5*, *CLN6*, *MFSD8*, and *CLN8* genes (Tables 8 and 9). These findings confirm the fact that the classical and variant forms of the genetically heterogeneous LINCL subtype cannot be distinguished in terms of age of onset and/or clinical presentation. The clinical homogeneity of the LINCL subtypes in association with the genetic and allelic heterogeneity characterizing the Turkish population make prioritization of the genes to be screened during differential diagnosis of LINCL very difficult. Instead, the diagnostic procedure should first involve measurement of the enzyme activity levels of TPP1, which if found low denote TTP1 insufficiency and establish the CLN2 disease diagnosis. In the case of normal TPP1 activity levels all LINCL variant genes should be screened for mutations. In contrast to this, in cases where a founder effect has been identified, like the *CLN5* p.Tyr392X in Finnish patients (Savukoski et al., 1998) or the *CLN6* p.Glu72X in patients of Costa Rican origin (Gao et al., 2002; Wheeler et al., 2002), providing a molecular genetic diagnosis is facilitated through evaluation of these common mutations through DNA-based diagnostic tests.

It is noteworthy that in a patient originating from a remote island of the Pacific ocean compound heterozygosity was detected for the nonsense p.Arg35X and the missense p.Thr160Ile mutations (Table 9). Given the geographical isolation of the patient's place of origin, and the fact that the island's population derived from the few settlers that first inhabited it, a homozygous mutation due to IBD was anticipated, compatible with a local founder effect. Whether two distinct mutations have occurred in the same gene in an apparently limited initial gene pool because of random mutation or because of an unknown underlying mechanism is not known. A similar phenomenon has previously been described in NCL patients from Newfoundland (Moore et al., 2008), but also in other clinical disorders such as the Bardet-Biedl syndrome (Katsanis et al., 2001).

### 5.2.3 Clinical phenotype in CLN7 disease (I, II)

In CLN7 disease, variant late infantile, the majority of patients manifest the first symptoms between 2-7 years of age (Topcu et al., 2004; Siintola et al., 2007; Aiello et al., 2009; Stogmann et al., 2009; Aldahmesh et al., 2009; study II). The majority of MFSD8 positive patients were reported to manifest epilepsy or developmental regression as the most common presenting symptom (Aiello et al., 2009; Stogmann et al., 2009; study II). Nevertheless, patients with visual failure as the presenting symptom have also been described (Aiello et al., 2009; Aldahmesh et al., 2009; study II). Following the initial symptom, the patients develop rapidly the whole spectrum of disease symptoms, involving myoclonus, ataxia, psychomotor decline, and speech deterioration. Patients affected with CLN7 disease die at a mean age of 12 years with a disease duration ranging from 3 to 12 years, though patients surviving until the age of 18 years have also been reported (Aldahmesh et al., 2009; Stogmann et al., 2009; study II). Visual failure occurs at a mean age of 4.5 years. All patients for whom an electroretinogram (ERG) has been performed have been found to have abnormal findings. Magnetic resonance imaging (MRI) studies reveal cerebellar and cerebral atrophy in most patients imaged (Aiello et al., 2009; study II), although some had normal findings (Stogmann et al., 2009). Additionally, white matter signal changes are identified in many patients, especially in the periventricular regions.

The phenotype of patients with CLN7 disease is clinically indistinguishable from that of patients with other NCL subtypes and onset in late infancy. The only distinction that could be made is that CLN7 disease is characterized by a slightly later disease onset and a more severe seizure phenotype compared to the classic LINCL CLN2 disease (Topcu et al., 2004). Unequivocal classification of a patient suspected to be affected by LINCL into any of the five LINCL subtypes can only be achieved via molecular genetic testing and identification of the underlying disease causing mutations.

Similar to the other NCL subtypes with onset in late infancy, the ultrastructural material accumulating in the cells of patients with CLN7 disease present as a mixture of FPs with RL (Siintola et al., 2007). The FP/RL pattern is only occasionally associated with CLs (study II). This draws a line between the classic and variant forms of LINCL, whereby CLs predominate in the former, while FP and/or CL inclusions are the primary finding in the variant subtypes.

### 5.2.4 Atypical phenotypes caused by MFSD8 defects (II)

Only one case with mutations in *MFSD8* and an atypical phenotype has been detected to date. In a patient homozygous for the missense p.Ala157Pro of Dutch origin the disease started at 11 years of age with visual decline, resembling JNCL. After a silent period of more than 10 years the patient started manifesting motor difficulties at the age of 24 years and seizures at the age of 25 years. The patient became ataxic over a period of four years. Cognitive decline developed at the fourth decade of life, and the patient became wheelchair-bound only at age 39 years. Today the patient is still alive at the age of 41 years.

The mechanism through which the missense p.Ala157Pro can result in the manifestation of such a mild disease phenotype, as opposed to the typical CLN7 disease, variant late infantile phenotype, is not known. Similar to the other missense *MFSD8* mutations p.Ala157Pro does not alter the protein's intracellular distribution and thus is likely to have an impact on the functional properties of MFSD8. The milder phenotype could be

due to the mutation affecting a functionally non-critical residue, or to the amelioration of the clinical picture by modifier alleles residing in loci outside *CLN7*.

### 5.2.5 The MFSD8 protein: Glycosylation (III)

The putative transporter function of MFSD8, justified by sequence homology analyses that classify it as a member of the MFS transporter proteins (Siintola et al., 2007), distinguish MFSD8 from the other NCL-associated proteins which are either lysosomal enzymes or transmembrane proteins of unknown function. Elucidation of the precise function and substrate specificity of MFSD8 are crucial not only because the transporter substrates could comprise novel NCL candidates, but also because they might shed light on the mechanisms that lead to neurological, visual and psychomotor deterioration. Towards understanding of the protein function, primary protein properties such as glycosylation and trafficking were addressed in study III.

Evidence for glycosylation of MFSD8 came from western blot analyses in which hCLN7 treated with the PNGaseF endoglycosidase showed increased electrophoretic motility as opposed to the untreated protein. In human MFSD8 only positions p.Asn371, and p.Asn376 met all the requirements for possible *N*-glycosylation: they were exposed to the lumen and comprised *N*-glycosylation consensus sites (N-x-S/T, where N is asparagine, x can be any amino acid except for proline and S/T denotes that serine or threonine must be present in the third position). To address the role of these residues in MFSD8 glycosylation both asparagine residues were disrupted individually (p.Asn371Gln and p.Asn376Gln) or simultaneously (p.Asn371Gln+p.Asn376Gln; Table 6, page 65). Of the generated mutants only hCLN7p.Asn371Gln+p.Asn376Gln was insensitive to PNGaseF and travelled as a sharp band of ~70 kDa, a finding that was also confirmed by an independent study carried out in parallel with study III (Steenhuis et al., 2010). The molecular mass of the unglycosylated MFSD8 was smaller than the expected mass by 14 kDa, a discrepancy that was experimentally proven to not derive from C-terminal proteolytic processing, but rather by the high protein hydrophobicity that might have altered the electrophoretic mobility of the peptide (Steenhuis et al., 2010).

Disruption of the orthologous positions in the mouse peptide did not abolish sensitivity to PNGaseF. Only when a third residue was disrupted (p.Asn389Gln) was electrophoretic homogeneity achieved, suggesting that the Mfsd8 mouse is glycosylated at three positions (p.Asn372, p.Asn377, and p.Asn389), unlike its human ortholog which is *N*-glycosylated at two residues (p.Asn371 and p.Asn376).

### 5.2.6 Subcellular localization of MFSD8 (III)

In the study that originally identified MFSD8 the N-terminally HA-tagged protein ($^{HA}$MFSD8) was found to localize in the lysosomes of COS-1 and HeLa cells (Siintola et al., 2007). In study III additional tagged constructs were used to evaluate intracellular localization of MFSD8. Co-localization of EGFP-hCLN7 and EGFP-mCLN7 with LAMP1 confirmed the initial findings for lysosomal/endosomal targeting of MFSD8. Evidence that tagging MFSD8 does not interfere with protein distribution came from localization studies in which untagged mCLN7, detected with a peptide raised against a 22-aa peptide of the mouse ortholog sequence, also co-localized with lysosomal markers. Finally, the lysosomal/endosomal targeting of MFSD8 was confirmed in mouse hippocampal neurons where the endogenously expressed protein was detected, showing that protein overexpression did not affect protein targeting. The latter shows that unlike CLN8 that displays differential intracellular distribution in neuronal versus non-neuronal cells (Lonka et al., 2004), MFSD8 specifically exerts its functional role at the lysosomal membrane.

### 5.2.7 Sorting mechanisms of MFSD8 targeting (III)

Since the sorting signals mediating protein trafficking are typically found in the protein cytosolic tails, the CLN7-CLN3dm-Δ1-38 and CD8-CLN7-Cterm chimeras were constructed to address whether they contain relevant trafficking information. Expression of both chimeras resulted in re-direction of the CD8 reporter protein from the PM to punctate intracellular structures, suggesting that both tails of MFSD8 are likely to carry targeting information. Disruption of the major candidate signals (the N-terminal dileucine

motif 13-LL-14 and the C-terminal tyrosine-based motif 513-YGRI-516, generating the mutants CLN7p.L13A-CLN3dm-Δ1-38 and CD8-CLN7-Cterm_p.Y513A, respectively, abolished intracellular targeting and resulted in dramatic redirection of both chimeras to the PM. The chimera experiments showed that both termini of MFSD8 are likely to contain information necessary for lysosomal targeting and that 13-LL-14 and 513-YGRI-516 are the motifs that most likely contain this information.

To further evaluate the role of the two motifs revealed to contain targeting information by the chimera experiments, mutations were introduced to disrupt each of them in the full length $^{HA}$MFSD8. Of the two only disruption of 13-LL-14 ($^{HA}$MFSD8p.L13A) resulted in major but not complete misrouting of the protein to the PM. Contrary to the clear impact of 13-LL-14 disruption, mutations in the tyrosine-based 513-YGRI-516 motif ($^{HA}$MFSD8p. p.Y513A) did not have an impact on the distribution of $^{HA}$MFSD8, suggesting a minor if any role at all in targeting of MFSD8. Complete misdirection of MFSD8 to the PM was not achieved even when the two motifs were simultaneously disrupted ($^{HA}$MFSD8p.L13A+p.Y513A ), or when they were combined with mutations in other less-likely candidate motifs ($^{HA}$MFSD8p.L13A+p.I25A+p.Y513A, $^{HA}$MFSD8p.L13A+p.I444A+p.Y513A, and $^{HA}$MFSD8p.L13A+p.Y503A+p.Y513A). Biotinylation experiments measuring the amount of MFSD8 directed to the PM yielded the same results with the targeted-mutagenesis localization experiments, confirming the conclusion that although the N-terminal dileucine motif is the major lysosomal determinant for MFSD8, additional signals are required for complete PM redirection of the protein. Exclusion of all the putative candidate motifs in study III suggests that the MFSD8 lysosomal signals that remain unidentified are presumably unconventional. Existence of unconventional motifs could also explain the discrepancy between the chimera and full-length mutant protein experiments addressing the role of the C-terminal portion of MFSD8 in protein targeting.

To obtain an insight into the route that MFSD8 follows on its way to the lysosomes experiments blocking endocytosis by transfection with the AP180 mutant were performed. In these cells, which were unable to uptake fluorescent transferin, MFSD8 targeting was not affected, suggesting that MFSD8 does not need to be transported to the

PM prior to being delivered to the lysosomes but instead follows the direct trafficking route towards them. The same was suggested in immunoblotting experiments where the MFSD8 termini were expressed as GST-fusion proteins and used to pull-down AP complexes via incubation with HeLa cell lysates. In these experiments the major dileucine motif was found to specifically bind to AP-1, since the detected interaction was lost when the N-terminal dileucine motif was disrupted by mutations. To obtain more information on the APs mediating sorting of MFSD8 both wild-type ($wt^{HA}$MFSD8) and mutant constructs ($^{HA}$MFSD8p.L13A, $^{HA}$MFSD8p.Y513A, and $^{HA}$MFSD8p.L13A+p.Y513A) were expressed in AP-1 and AP-3 deficient cell lines. The fact that deficiency for either AP-1 or AP-3 alone does not alter the distribution of $wt^{HA}$MFSD8 suggests that the major dileucine motif is recognized by more than one AP. Furthermore, even when both p.L13A and p.Y513A mutations are expressed in AP-1 and AP-3 adaptors-deficient cell lines a pool of MFSD8 can still be correctly delivered to the lysosomes, supporting the idea that several motifs recognized by several adaptors must add up to form the complex repertoire of mechanisms underlying sorting of MFSD8.

The importance of the N-terminal dileucine motif was independently demonstrated by Steenhuis et al. (2010). The experimental set-up employed was very similar to the one used in study III, involving targeted-mutagenesis of the putative sorting motifs followed by localization studies with immunofluorescence microscopy (Steenhuis et al., 2010). Biotinylation experiments in both studies reported a minor sorting role for the C-terminal p.Y503 and p.Y513, justified by the increase of the portion of MFSD8 on the PM when the major dileucine signal and both tyrosines are simultaneously disrupted by mutations (Study III; Steenhuis et al., 2010). One major contradiction supported by two different experimental approaches in each study was reported. In Steenhuis et al. (2010) biotinylation experiments showed that 22% of wild-type MFSD8 resides at the PM at steady-state and that this fraction is increased when endocytosis is blocked, favoring the hypothesis that MFSD8 follows the indirect route to the lysosomes (Steenhuis et al., 2010). The same experimental approaches used in study III did not identify MFSD8 at the PM and showed that endocytosis blocking has no effect on lysosomal delivery of MFSD8, proposing that the direct route is followed by MFSD8. The latter is also supported by a third experimental procedure that identified the interaction between the

dileucine motif and AP-1 complexes (study III), which are known to operate in the endosomal / lysosomal pathway (Robinson and Bonifacino, 2001). Differences in the experimental set-up used could account for the discrepancies in the biotinylation and endocytosis blocking experiment results. Although the direct trafficking route is supported by three different approaches, additional experiments need to be performed in order to be able to unequivocally draw this conclusion.

### 5.2.8 Expression of MFSD8 in rodent brain (III)

To address the cellular and regional expression of *MFSD8* the mRNA levels were measured via real-time PCR analyses in different cell types and different brain regions of rats. On the cellular level the highest *MFSD8* mRNA levels were detected in the neurons, in which the protein was expressed 6- to 12-fold more than in astrocytes or microglia (Figure 7). Of the other NCL-associated genes used as controls, *PPT1* and *CLN3* were expressed in neurons at almost equal levels as in astrocytes and microglia, respectively. *MFSD8* is distinguished from *PPT1* and *CLN3*, however, because it is the only one with most prominent expression in neurons. *CTSD* was predominantly expressed in microglia and *TPP1* in astrocytes. The indirect evidence thus obtained from the transcript expression levels in the rat brain assigns a prominent role for *MFSD8* in neurons. It is noteworthy that although the genes evaluated affect different types of brain cells they all result in similar clinical phenotypes.

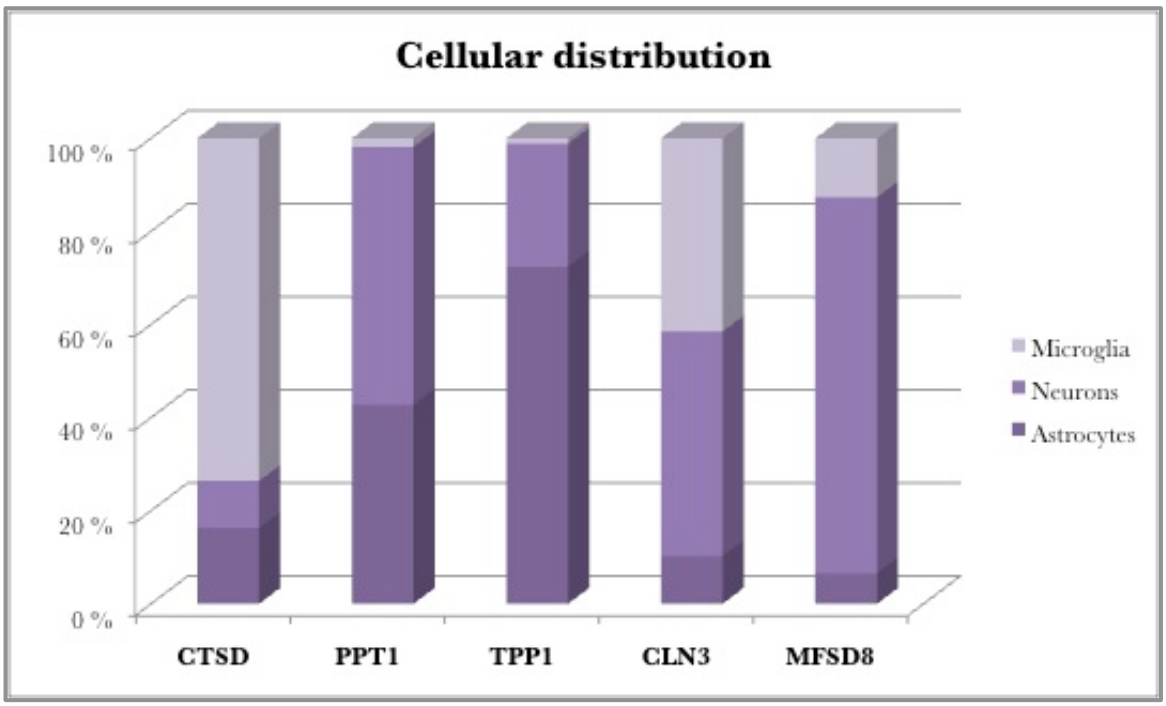


**Figure 7.** Expression levels of *CTSD*, *PPT1*, *TPP1*, *CLN3*, and *MFSD8* in neurons, astrocytes and microglial cells from rat brain.

In the absence of an antibody that could be used in immunohistochemical analyses to address the spatiotemporal expression of *MFSD8*, indirect evidence was sought from measuring the transcript expression levels in different rat brain regions. *MFSD8* was detected to be 5-fold more abundant in the hippocampus than in the other regions evaluated (Figure 8). Prominent hippocampal expression was also seen in the *in situ* hybridization experiments, which furthermore indicated the cerebellar granular layer and the deeper neocortex areas among the regions where *MFSD8* is mostly expressed. Interestingly, the expression levels of the transmembrane NCL proteins evaluated (*MFSD8* and *CLN3*) were more than 10-fold lower compared to the levels of expression detected for the soluble NCL mRNAs (*CTSD*, *PPT1*, and *TPP1*) (Figure 8). The significance of these differences in expression remains unknown.

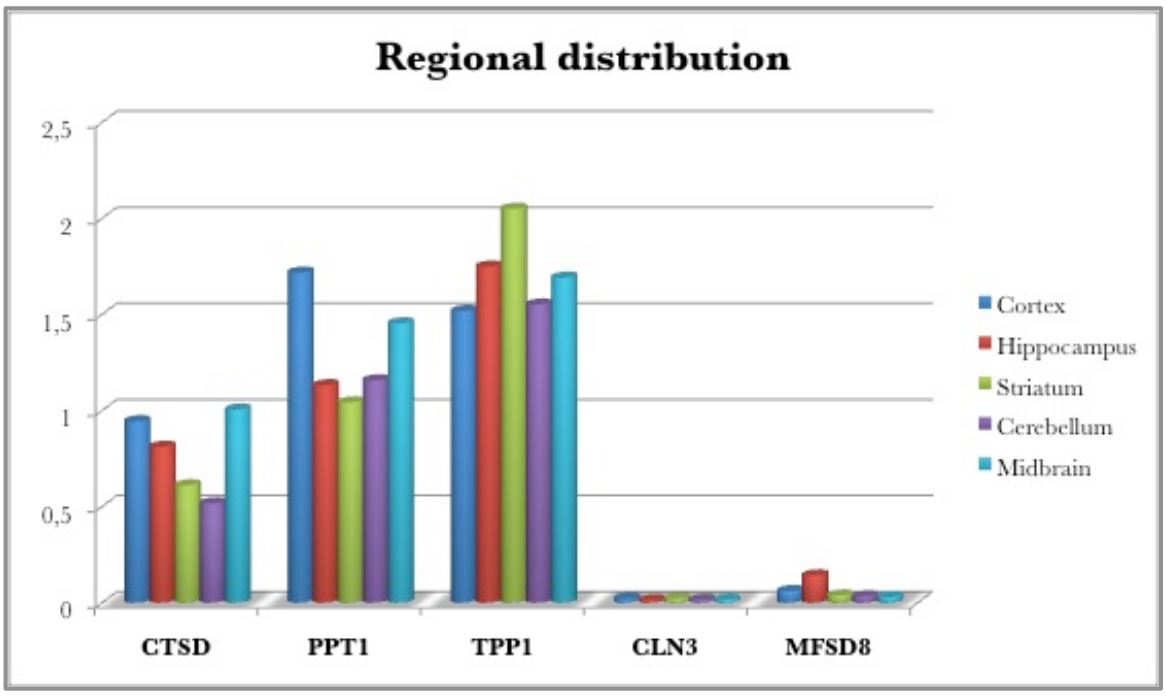


**Figure 8.** Expression levels of *CTSD*, *PPT1*, *TPP1*, *CLN3*, and *MFSD8* in the cortex, hippocampus, striatum, cerebellum and midbrain of the rat brain.

### 5.2.9 Neuropathological findings in postmortem human brains of MFSD8 positive patients (III)

Histological analyses of brains of MFSD8 positive patients, who died between ages 7-18 years, showed maximal accumulation of lipopigment inclusions in neurons, and to a lesser extent in oligodendrocytes or astrocytes. These findings are in agreement with the data obtained from rodent brains, denoting that MFSD8 is predominantly expressed in neurons and that neuronal populations such as those of the cerebellar granular cell layer die faster than other cell-types. This provides indirect evidence that MFSD8 has a central role in neurons, which are postulated to be the most vulnerable cell-type in MFSD8 deficiency. Furthermore, the accelerated degeneration of the cerebellar granule cell layer and the hippocampal pyramidal cell layer of patients, are in perfect agreement with the regional vulnerability seen in the *in situ* hybridization analysis in the rodent brain.

An interesting observation came from the comparison of storage material load in different cell types and survival of these cells. Purkinje cells in which storage was more prominent

survived longer than cerebellar granular cells in which the lipopigment accumulation was more discrete. This leads to the conclusion that either the Purkinje cells are more resistant to storage material accumulation or that it is not the accumulation of the storage material *per se* that induces the extensive neuronal cell death that takes place throughout the disease progression.

## 5.3 The hypothesis of NCL genes acting as modifiers of phenotype (unpublished)

While dissecting the NCL genetic background in the cohort of Turkish patients in study I, heterozygous changes (involving missense, nonsense and splice-site affecting) were identified in some of the patients (data not shown). The single sequence variations were either seen alone, or in combination with homozygous or heterozygous defects in other NCL genes. In the case where one of the heterozygous changes existed alone or in combination with a second heterozygous change in a second gene, the pattern of inheritance is not able to explain manifestation of the disease under a Mendelian model. A more likely hypothesis was that the second mutation comprised a change that could not be identified with the sequencing methods used, such as deletion, insertion, copy number variation or intronic change.

A second hypothesis would be that these mutations would not be disease-causing *per se*, but instead would increase the mutational load of a given NCL patient, modifying the phenotype. If the latter was true it could potentially contribute to the explanation of the pronounced inter- and intra-familial variability in disease manifestation described in several reports. It was therefore tempting to test the hypothesis of NCL loci acting as modifiers of the phenotype caused by mutations in other NCL loci.

To test this hypothesis 105 Turkish patients were sequenced for all the known human NCL genes except for *DNAJC5*, irrespective of the haplotype over the corresponding locus. Having established a complete genetic profile of each patient for the seven NCL genes, a total of 32 patients were detected to carry changes in at least two distinct NCL loci. In these cases it

**Table 10.** Locus interactions based on genetic observations

| Primary locus | Gene | Secondary locus |
|---|---|---|
| *CLN1* | *PPT1* | *CLN2, CLN5, CLN6, CLN7* |
| *CLN2* | *TPP1* | *CLN1, CLN3, CLN5, CLN7, CLN8* |
| *CLN3* | *CLN3* | *CLN2, CLN5, CLN7* |
| *CLN5* | *CLN5* | *CLN1, CLN2,CLN3, CLN5, CLN6, CLN7, CLN8* |
| *CLN6* | *CLN6* | *CLN1, CLN5, CLN7* |
| *CLN7* | *MFSD8* | *CLN1, CLN2, CLN3, CLN5, CLN6* |
| *CLN8* | *CLN8* | *CLN5* |

would be tempting to speculate that one NCL locus has an epistatic interaction with the second NCL locus. A list of all possible such interactions based on patient genotyping alone is given in Table 10. The locus interactions determined by the genetic findings are in agreement with the identified protein interactions as determined by experimental data (Figure 9). For example, homozygous or heterozygous changes in the *CLN1* locus can simultaneously be detected with homozygous or heterozygous changes in *CLN2*, *CLN5*, *CLN6*, or *CLN7* (Table 10). The same possible locus interactions are supported by results from the protein experiments whereby PPT1 can directly interact with CLN5 and CLN6, and indirectly (via CLN5) with TPP1 (Figure 9). No data on protein interactions exist yet for MFSD8.

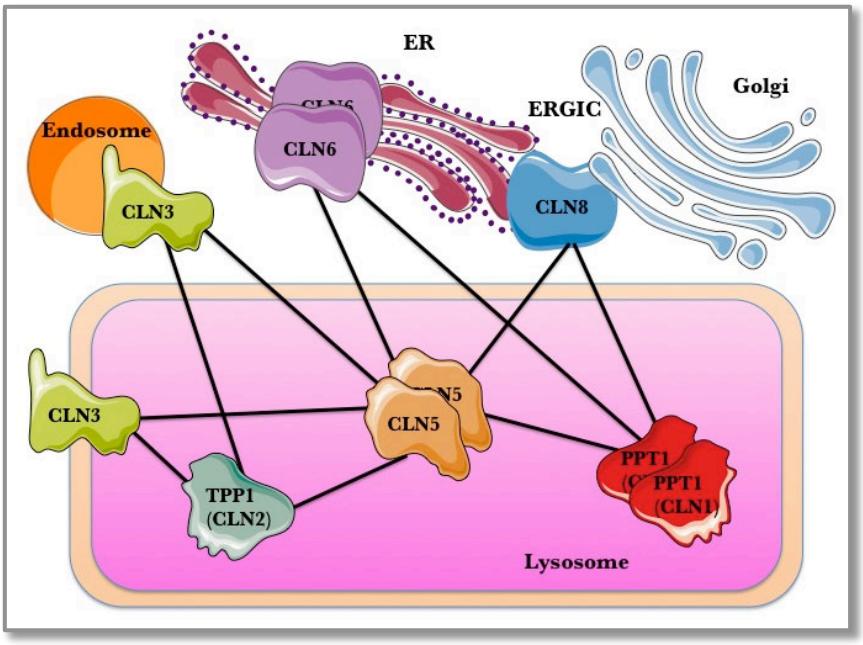


**Figure 9.** NCL protein interactions as determined by experimental evidence from protein and cellular work (von Schantz et al., 2008; Lyly et al., 2009; Uusi-Rauva et al., 2007; Luiro et al., 2006; unpublished observations by the group of Anu Jalanko). Duplication of the protein representation (PPT1, CLN5, and CLN6) indicate that the peptide of the given protein interacts with other peptides of the same protein. The figure was produced using Servier Medical Art (http://www.servier.com/servier-medical-art/powerpoint-image-bank).

In the modifying locus hypothesis, both the genetic and molecular approaches denote CLN5 to have a central role in the hypothetical NCL pathway. From a molecular standpoint CLN5 is known to interact with TPP1 (Vesa et al., 2002), and to be

transported together with PPT1 to the lysosomes via the Man6P-independent pathway (Lyly et al., 2009; Vesa et al., 2002). Additionally, CLN5 has confirmed interactions with TPP1 and CLN3 in the lysosomes, and with PPT1, CLN6, and CLN8, in the ER (Lyly et al., 2009). From a genetic point of view *CLN5* was the gene with most changes identified across its sequence. It is surprising how *CLN5*, which causes premature death and severe neurodegeneration when defective, can harbor so much variation. The disproportionate numbers of genetic variants identified and patients with *CLN5* molecular genetic diagnosis could be explained by the modifier hypothesis, whereby the *CLN5* locus contributes more modifying (60%) than disease-causing (10%) alleles (Figure 10).

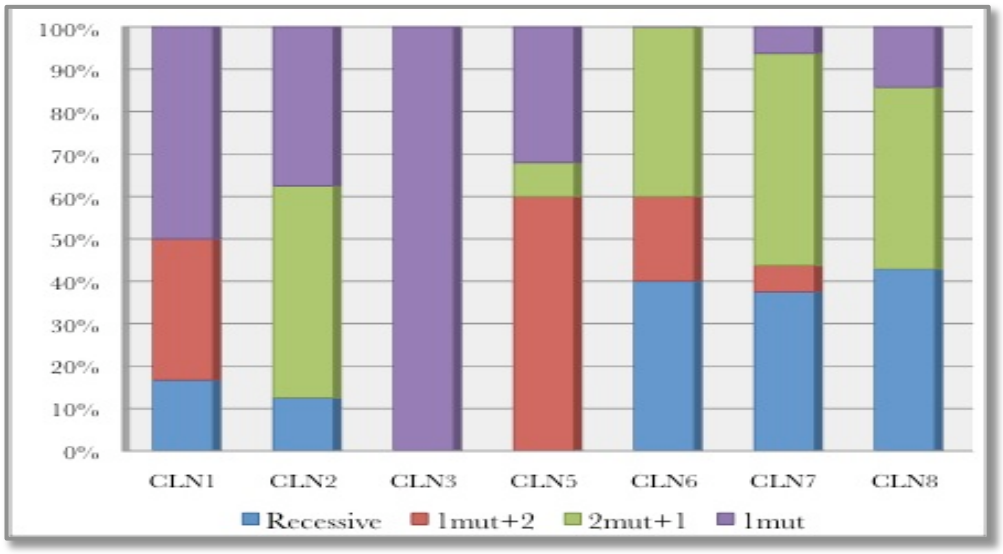


**Figure 10.** Percentage representation of the disease-causing (recessive and 2mut+1) and modifying alleles (1mut+2) across the seven NCL loci evaluated

The lack of comprehensive clinical data did not allow correlations to be made between specific sequence changes and manifestation or absence of specific clinical symptoms. Furthermore, with the sample size being so small it was not possible to draw statistically significant correlations between clinical timing or severity and combination of defects in particular NCL genes. It nevertheless remains tempting to test whether the NCL genes contribute modifying alleles. This hypothesis needs to be tested in animal models such as the *Cln1*$^{-/-}$ and *Cln5*$^{-/-}$ double knock-out mouse (under development by the Anu Jalanko lab).

## 5.4 Identification of novel childhood onset PME genes

A total of 135 patients of Turkish origin were excluded from all known NCL loci in study III. Of these three small families (families x, N36 and N51; Figure 5) and 19 singletons were analyzed in a genomewide SNP scan. Although data on consanguinity were not available for all patients genotyped homozygosity mapping was performed. The analysis concentrated first on the families, and subsequently on the singletons.

### 5.4.1 Identification of mutations in PLA2G6 (unpublished)

Homozygosity mapping in family N36 mapped three candidate loci, one in each of chromosomes 19, 21 and 22. In addition to the affected sibling of family N36 two more patients showed a long homozygosity run over the locus on chromosome 22 (N4603, and N3903). The critical interval on chromosome 22 was 10.2 Mb long (chr.22: 32.4-42.6 Mb) and harbored a total of 184 genes (Figure 11).

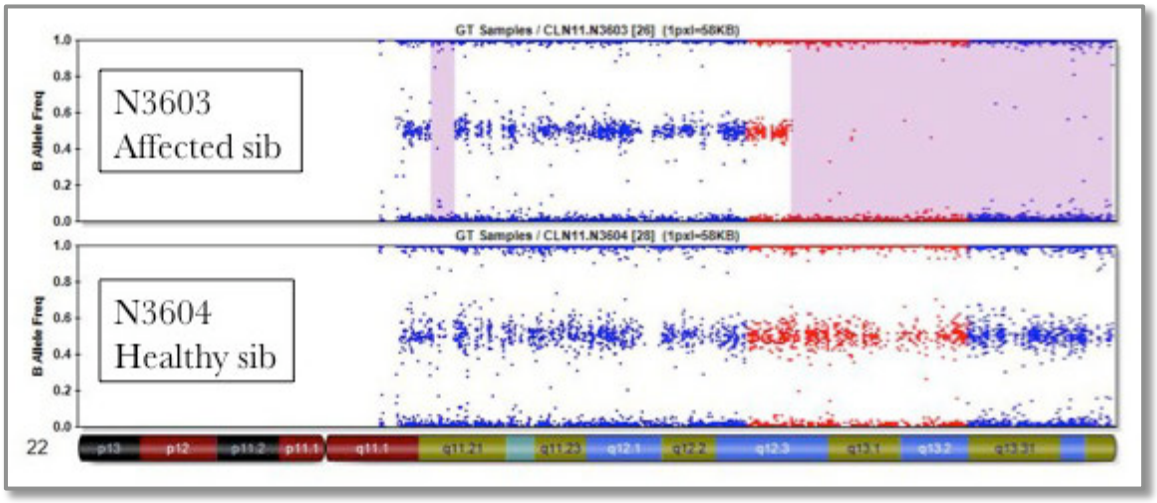


**Figure 11.** Homozygosity over the candidate locus on chromosome 22 for family N36. The genotype of each individual is represented as dots, whereby the dots distributed in the middle of each panel represent heterozygous calls, while the dots at the top or bottom of the panels represent homozygous variants. The homozygous segment in patient N3603 is highlighted in pink. The calls over the candidate locus are shown in red. The chromosomal view over which the genotypes are distributed is shown in the lower part of the figure.

The 184 positional candidate genes residing within the critical chromosomal interval were prioritized based on the putative function. After having performed sequencing analysis in 13 of the positional candidate genes (Table 11) two patients were found to carry mutations in *PLA2G6* (NM_001004426), a calcium-independent phospholipase A2 that catalyzes the release of fatty acids from phospholipids. Patient N3603 was homozygous and his healthy sister heterozygous for the nonsense variant c.1903C>T (p.Arg35X), while patient N3903 was homozygous for the in-frame deletion c.2065_2067delATC (p.Ile689del). The remaining patient linked over this locus (N4603) was not positive for mutations in *PLA2G6*. This could be explained by the fact that patient N4603, for whom 6% of the genome is in homozygosity, was homozygous over the region by chance.

**Table 11.** Genes screened from the candidate interval on chromosome 22.

| Gene name | Description | Subcellular localization | Disease association |
|---|---|---|---|
| *A4GALT* | alpha 1,4-galactosyltransferase | Golgi apparatus membrane | |
| *ADSL* | Adenylosuccinate lyase isoform a | | Adenylosuccinate deficiency |
| *GGA1* | Golgi associated, gamma adaptin ear containing | | |
| *KCNJ4* | Potassium-inwarding rectifying channel J4 | Membrane (multi-pass membrane protein) | |
| *KDELR3* | KDEL receptor 3 isoform b | | |
| *NAGA* | Alpha-N-acetylgalactosaminidase precursor | Lysosome | Schindler and Kanzaki disease (early onset neuroaxonal dystrophy) |
| *PACSIN2* | Protein kinase C and casein kinase substrate in | Cytoplasmic vesicle (by similarity) | |
| *PLA2G6* | Phospholipase A2 | | Infantile neuroaxonal dystrophy (INAD) and neurodegeneration with brain iron accumulation (NBIA) |
| *PSCD4* | Pleckstrin homology, Sec7 and coiled/coil | | |
| *SLC16A8* | Solute carrier family 16, member 8 | Cell-membrane (multi-pass membrane protein) | |
| *SLC25A17* | Solute carrier family 25 (mitochondrial carrier) | Peroxisome membrane (multi-pass membrane protein) | |
| *SYNGR1* | Synaptogyrin 1 isoform 1a | Membrane (multi-pass protein); melanosome; cell junction; synapse | |
| *TXN2* | Thioredoxin 2 precursor | Mitochondrion | |

Previously defects in *PLA2G6* had been associated with infantile neuroaxonal dystrophy 1 (INAD1 MIM# 256600) and neurodegeneration with brain iron accumulation (NBIA MIM# 610217; Morgan et al., 2006; Khateeb et al., 2006; Gregory et al., 2008). INAD is a progressive encephalopathy with onset at approximately 2 years of life. The disease is characterized by the accumulation of lipid storage in the brain, swelling and degeneration of the neuronal axons, as well as scattered spheroids in the central nervous system (Cowen and Olmstead, 1963). In NBIA the phenotype is very similar to INAD with the only difference that the former typically has a later disease onset and patients may survive until the third decade of life (Hortnagel et al., 2004). Also in NBIA there is iron accumulation in the basal ganglia, giving a sign on MRI that resembles the characteristic 'eye of the tiger'. Nevertheless, not all *PLA2G6* positive cases present the characteristic brain iron accumulation, and thus this gene comprises a good candidate not only for patients with neurodegeneration and brain iron accumulation, but also for patients with similar clinical findings but no iron accumulation in their brains (Paisan-Ruiz et al., 2010).

After the identification of mutations in *PLA2G6* a cohort of 16 patients with a putative INAD clinical diagnosis and a group of 107 Turkish patients with PME and no mutations in the NCL genes were evaluated for mutations in this gene. A further nine patients were found to carry *PLA2G6* mutations and were thus provided with a molecular diagnosis (Table 12). Of the 11 *PLA2G6* mutations identified eight are novel and three (p.Gly347Arg, p.Pro353Leu, and p.Arg635X) have been previously reported to cause INAD (Morgan et al., 2006). These findings raise the total number of mutations identified in *PLA2G6* to 60. With the identification of altogether 11 patients positive for mutations in *PLA2G6* it becomes apparent that the gene should be considered for mutation screening in differential diagnosis of patients with late infantile or childhood onset neurological deterioration.

**Table 12.** Mutations identified in the sequence of *PLA2G6*

| Position | Nucleotide change | Amino acid change | Patient identifier (status) | Study |
|---|---|---|---|---|
| Exon 5 | c.753_754insC | p.Asn252GlnfsX130 | M1203 (homozygous)<br>r3 (heterozygous) | IV |
| Exon 7 | c.902G>C | p.Arg301Pro | M1303 (homozygous) | IV |
| Exon 7 | c.1039G>A | p.Gly347Arg | M0903 (homozygous) | IV |
| Exon 7 | c.1058C>T | p.Pro353Leu | M0603 (homozygous) | IV |
| Exon 7 | c.1077G>A | Splice site affecting | M1103 (homozygous) | IV |
| Exon 10 | c.1408A>G | p.Met470Val | M1503 (homozygous) | IV |
| Exon 11 | c.1573G>A | p.Ala525Thr | M0703 (heterozygous) | IV |
| Exon 13 | c.1748T>C | p.Met583Thr | M0403 (heterozygous) | IV |
| Exon 14 | c.1903C>T | p.Arg635X | N3603 (homozygous) | IV |
| Exon 14 | c.1982C>T | p.Thr661Met | r3 (homozygous) | IV |
| Exon 15 | c.2065_2067delATC | p.Ile689del | N3903 (homozygous) | IV |

### 5.4.2 Identification of sequence variations in *USP19* and *TXNDC6* through targeted next-generation sequencing (unpublished)

Homozygosity analysis in the affected siblings of family N51 mapped four candidate loci (two distinct loci on chromosome 1, and one locus on each of chromosomes 3 and 16), harboring a total of 199 genes. In the affected sisters of family x four different loci (one on each of chromosomes 3, 8, 16, and 17) were determined, within which lay altogether 285 genes.

The exons and exon-intron junctions of the 484 positional candidate genes in both families were screened for mutations using the NimbleGen Sequence Capture array protocol followed by subsequent Illumina sequencing. Data analysis in the two families revealed one change segregating with the phenotype per family. The patients in family N51 were homozygous for the transition c.176G>A (p.Gly59Asp) in *USP19,* and the siblings in family x were homozygous for the transition c.433T>C (p.Tyr145His) in *TXNDC6* (Figure 12).

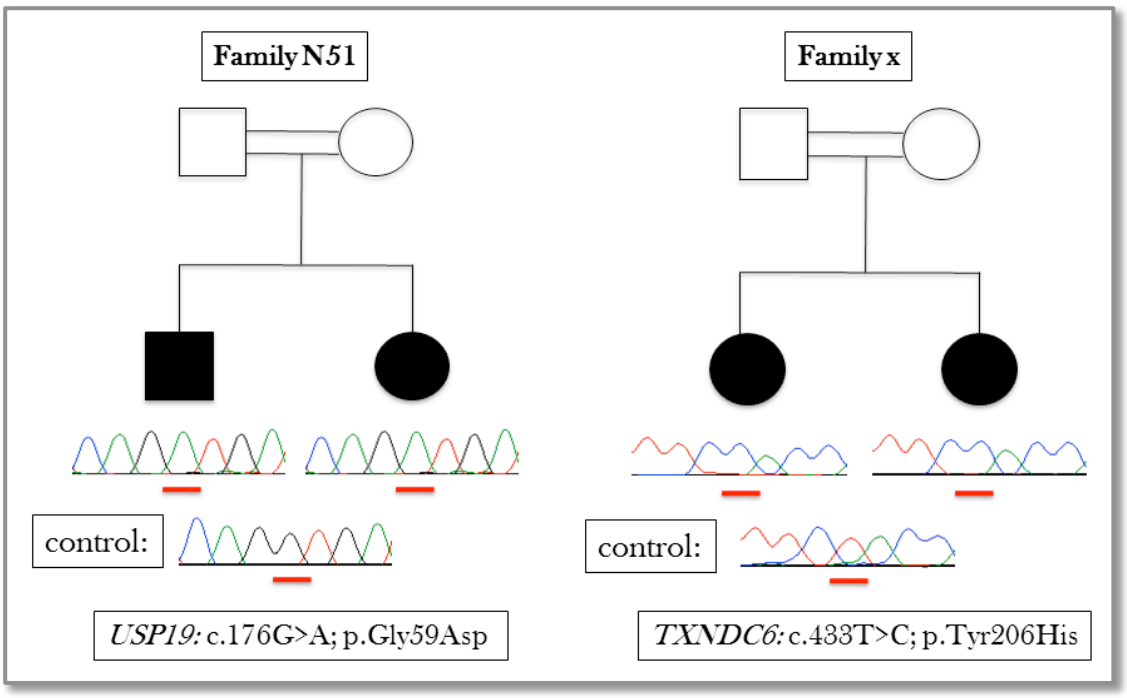


**Figure 12.** Schematic representation of the families N51 and x. The chromatogram showing the genetic defect identified in each of the families is given below the affected individuals. The sequence of a control individual is shown below the chromatograms showing the changes identified in patients. The base affected is underlined in red.

*USP19* (NM_006677.2) encodes for the ubiquitin specific peptidase 19. The encoded protein is a deubiquitinating enzyme that regulates the degradation of a variety of different proteins, and is also involved in the turnover of ER-associated degradation substrates (Hassink et al., 2009). The same change was identified in heterozygous state in two of the 107 PME patients that remained with no molecular genetic diagnosis, but a second change within *USP19* could not be identified. Collectively, the variant identified in family N51 was present in four patients of Turkish origin. Furthermore, the residue affected by the change is not conserved across vertebrates, which is most probably the reason why *in silico* analysis have predicted it to have a neutral impact.

*TXNDC6* (NM_178130.2) encodes for the thioredoxin domain containing protein 6. The function of TXNDC6 is not known. Sequence similarity analyses predict a putative cytoplasmic/cytoskeletal localization for the protein and a possible role in the regulation

of microtubule physiology. The p.Tyr145His change identified in patients of family x is highly conserved across vertebrates and was predicted to represent a probable damaging change. Screening of the panel of 107 PME patients that remained with no molecular genetic diagnosis did not reveal any changes in the sequence of *TXNDC6*.

Neither of the *USP19* or *TXNDC6* affecting changes was identified in 180 ethnically matched control chromosomes or in the release of the 1000 Genomes project, as of October 2011. This suggests that both variants are very rare. The variant identified in family x is likely to be private since it has only been detected in the family's affected sibs. The fact that *USP19* p.Gly59Asp was the only variant identified across the sequence of the gene in four different patients suggests that the most likely hypothesis is that this change represents a very rare population specific variant that is unlikely to be disease-causing, especially when considering that in two of the patients it represented the only identified change. It can of course not be excluded that in these two patients the second mutation could have been missed by the sequence screening methods used in this study, which cannot detect all deletions, insertions that might exist across the gene sequence or the changes in introns and regulatory elements. It can thus be concluded that although *USP19* is not likely to represent a disease-causing gene, the disease-association of either of the variants detected cannot be unequivocally established until a second family with mutations in the same genes has been identified. Thus, the role of *TXNDC6* and *USP19* as disease-associated genes awaits verification in further patients.

### 5.4.3 KCTD7 gene and protein

#### *5.4.3.1 Identification of KCTD7 in the sporadic patients (IV)*

Having completed the analysis in the families genotyped with the SNP-based genomewide scan array, homozygosity mapping analysis was subsequently applied to the 18 sporadic patients. The homozygous runs were determined and the 18 singletons were subsequently grouped together on the basis of overlapping homozygosity over the same loci. The candidate loci were prioritized based on how many patients shared homozygous

segments over them. Hence, a region on chromosomal position 7q11.21 over which eight patients were simultaneously homozygous (Figure 13) was considered the primary candidate locus. The homozygous segments over the region varied in length from a minimum of 1.9 Mb in patient k3 to the longest of 31.5 Mb in patient N4603. Evaluation of the homozygosity breakpoints helped refine the minimal critical interval shared by all eight patients to a 1.5 Mb region, with markers rs1812771 from patient k3 and rs4626481 from patient N4803 defining the breakpoints of the candidate region (chr.7: 64.563.264-66.111.815) (Figure 13).

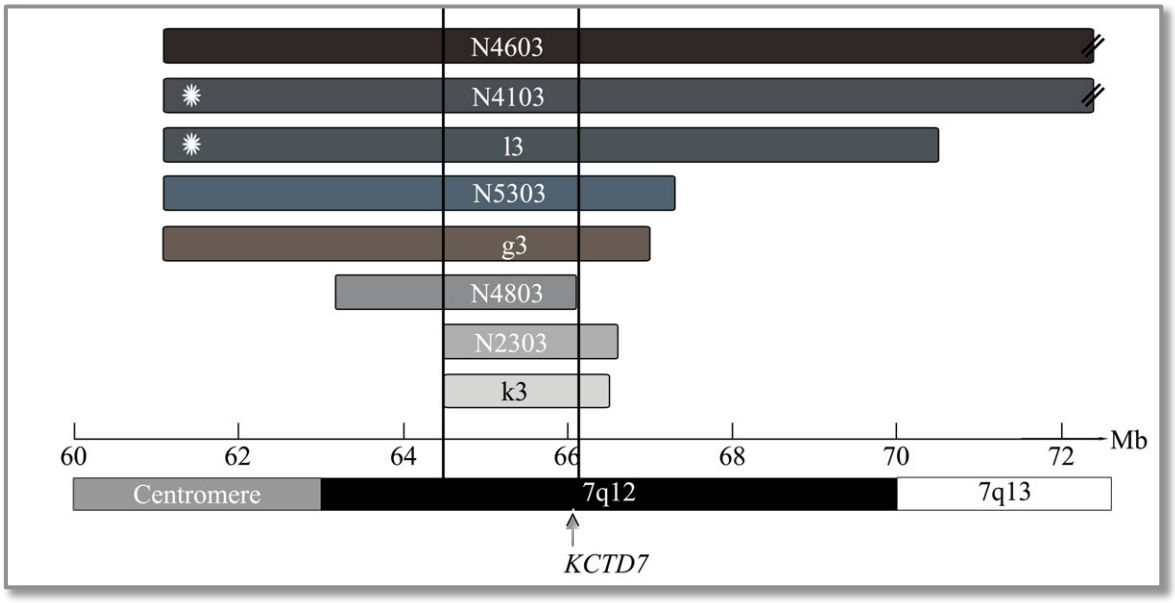


**Figure 13.** Candidate region on chromosome 7q11.1-21.2. Each bar represents a different patient's homozygous haplotype block over the candidate locus. The diagonal lines at the bar of patients N4103 and N4603 denote that the homozygous segment in these patients extends further than the indicated region. The physical and chromosomal positions are given at the lower part of the figure. The patients found to carry mutations in *KCTD7* are indicated with a white asterix at the left part of the homozygous segment. The black vertical bars show the homozygosity breakpoints at patients k3 and N4803, which defined the minimal critical interval to the 1.5 Mb region 64.563.264-66.111.815. The position of *KCTD7* is given with a black arrow at the bottom of the picture.

Altogether 11 genes resided within the 1.5 Mb candidate interval (Table 13). These were prioritized based on the available information on the known or putative function of the encoded protein, tissue expression and intracellular compartment localization.

**Table 13.** Genes residing within the critical interval on chromosome 7: base positions 64,563,264-66,111,815 (Build 37; GRCh37/hg19). The genes are listed according to their chromosomal position.

| Gene name | Description | Subcellular localization | Disease association |
|---|---|---|---|
| *VKORC1L1* | vitamin K epoxide reductase complex, subunit | Membrane; Multi-pass membrane protein (Potential) | No |
| *GUSB* | glucuronidase, beta | Lysosomes | Mucopolysaccharidosis type 7 |
| *ASL* | argininosuccinate lyase isoform 1 | n.a. | Arginosuccinicaciduria (ARGINSA) |
| *CRCP* | calcitonin gene-related peptide-receptor | Nucleus (By similarity). Cell membrane; Peripheral membrane protein; Cytoplasmic side (By similarity) | No |
| *TPST1* | tyrosylprotein sulfotransferase 1 | Golgi apparatus membrane; Single-pass type II membrane protein (By similarity) | No |
| *LOC285908* | hypothetical protein LOC285908 | n.a. | No |
| *KCTD7* | potassium channel tetramerisation domain | n.a. | Progressive myoclonic epilepsy type 3 (EPM3) |
| *RABGEF1* | RAB guanine nucleotide exchange factor (GEF) 1 | n.a. | No |
| *C7orf42* | hypothetical protein LOC55069 | Membrane; Multi-pass membrane protein (Potential) | No |
| *SBDS* | Shwachman-Bodian-Diamond syndrome protein | Cytoplasm (By similarity) | Schwachman-Diamond syndrome (SDS) |
| *TYW1* | radical S-adenosyl methionine and flavodoxin | n.a. | No |

The first gene to be screened for mutations in the eight patients sharing homozygosity over this region was *KCTD7* (NM_153022.3), encoding for the potassium channel tetramerization domain-containing 7 protein. Previously, *KCTD7* defects had been associated with PME type 3 in a Moroccan family with three affected offspring (van Bogaert et al., 2007) where the nonsense p.Arg99X in exon 2 had been found to segregate with the disease phenotype in the family. The three affected Moroccan patients presented with epilepsy as the initial disease symptom at a mean age of 19 months (range 16-24

months). Other clinical symptoms involved psychomotor delay, ataxia, myoclonus, and anarthria. After a series of metabolic tests, biopsy examination, EM analysis that was negative for the detection of storage material characteristic of NCLs, and molecular analyses to exclude other causes of PME, it was concluded that the patients were suffering from a novel form of PME, termed PME type 3 (EPM3; van Bogaert et al., 2007).

Screening of the *KCTD7* coding exons and exon-intron junctions revealed mutations in two of the patients that showed homozygosity over the *KCTD7* locus. Patient l3 was homozygous for the missense causing mutation c.280C>T (p.Arg94Trp) in exon 2 and patient N4103 was homozygous for the second missense change c.818A>T (p.Asn273Ile) in exon 4 (Figure 14). None of the remaining six patients (Figure 13) carried mutations in *KCTD7*. The latter could be explained by the inability to detect deletions or insertions in *KCTD7* or other types of mutations such as intronic or promoter region mutations with the screening methods used. The association of rearrangements of the region 7q11.23, located 6 Mb downstream of the region in which *KCTD7* resides, with neurodevelopmental syndromes such as the Williams-Beuren syndrome (Schubert, 2009), further supports the neurological relevance of this chromosomal segment and of some of the genes that reside within it. Fluorescence *In Situ* Hybridization (FISH) analyses or cDNA screening, had patient mRNA been available, would have been helpful in order to evaluate such changes and even to detect transcript variants that might have been missed during this analysis.

Following identification of mutations in *KCTD7* the second cohort of 107 Turkish patients and one Pakistani family with two affected siblings which remained with no molecular genetic diagnosis were screened. Of these, seven patients from five small families carried mutations in *KCTD7*. The siblings of the Pakistani family (Pak4 and Pak5) were homozygous for the missense c.322C>A (p.Leu108Met). Patient N2703 was compound heterozygous for the two missense-causing mutations c.343G>T (p.Asp115Tyr) and c.818A>T (p.Asn273Ile). Patients N3503 and N15103 were homozygous for the single base deleting mutation c.594delC, resulting in a frameshift and predicting a prematurely truncated protein by 16 amino acids (p.Ile199SerfsX74). Finally, the siblings of family N126 (N12604 and N12606) were homozygous for a three-base pair deletion

(c.861_863delATG), producing the peptide change p.Trp289X. With the six novel mutations reported in study IV the spectrum of *KCTD7* affecting changes has risen to seven (Figure 14) and the number of families positive for mutations in this gene to eight. No mutations were identified in the 22 patients with a confirmed NCL diagnosis justified by the identification of storage material upon EM examination.

The mutations p.Arg94Trp, p.Leu108Met, p.Asp115Tyr, and p.Trp289X are specific for the families in which they were identified. The fact that p.Asn273Ile is carried by both patients N2703 and N4103, and p.Ile199SerfsX74 by both N3503 and N15103, suggest either that each pair of patients shares a common ancestor, or that the affected residues represent relatively frequently mutable sites.

Different reasons that argue in favor of the identified *KCTD7* sequence variants having a disease-causing role exist. For the families in which parent samples were available the mutations segregated with the disease phenotype. None of the mutations was detected in at least 150 ethnically matched control chromosomes screened. KCTD7 has orthologues in several species across which it is highly conserved, and the majority of mutations identified affect conserved residues. Finally, different types of mutations have been detected in *KCTD7*, including missense, frameshift-causing, deletions and nonsense (this study; van Bogaert et al., 2007).

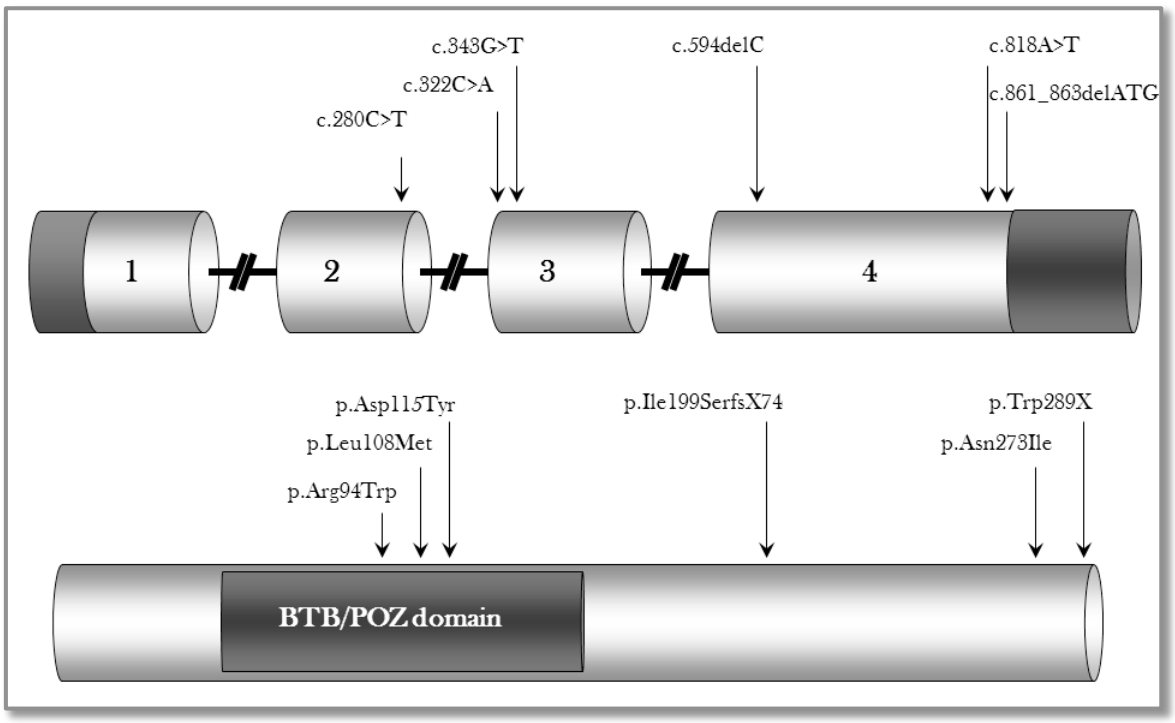


**Figure 14.** Schematic representation of the *KCTD7* gene (upper half of the figure) and protein (lower half) and relative position of the mutations identified. The exons are shown as grey cylinders and are in scale. In the gene representation the introns are indicated as black lines and the untranslated regions as non-numbered dark grey boxes. In the protein representation the full-length peptide is shown as a grey cylinder, and the BTB/POZ domain that it carries as a darker shaded box. The positions of the mutations are indicated by the arrows.

The effect of the frameshift-causing p.Ile199SerfsX74 is predicted to be deleterious since the mutation introduces 74 novel amino acids before terminating the mutant peptide prematurely by 16 amino acids. Determining the impact of the missense mutations, however, is not as straightforward. *In silico* prediction programs suggested a probably damaging role for p.Arg94Trp, p.Leu108Met and p.Asn273Ile, which affect highly conserved residues, and a benign impact on the protein level for p.Asp115Tyr, which could be a consequence of the non-conservation of the p.Asp115 across all vertebrates species evaluated (such as zebrafish and the puffer fish). Despite the conservation data and the *in silico* predictions, p.Asp115Tyr is believed to be disease-associated because it is absent from the control chromosomes evaluated and because it is detected in compound heterozygosity with the probably damaging p.Asn273Ile. Under a Mendelian recessive

model of inheritance in the case of patient N2703 (compound heterozygous for p.Asp115Tyr and p.Asn273Ile), p.Asp115Tyr needs to be disease-causing, and if not p.Asn273Ile needs to be combined with a second mutation that was not identified with the screening methods used, such as a genomic deletion.

The functional consequence of the in-frame deletion p.Trp298X is not easy to prove as the encoded mutant protein differs from the reference peptide only in that the former lacks the tryptophan residue prior to the termination codon (p.Trp289). A peptide with the same amino acid composition is also produced by the transcript variant 2 of *KCTD7* (NM_001167961.2), which is expressed in several cDNAs from different brain regions according to AceView (http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/av.cgi?db=human&term=KCTD7&submit=Go). Nevertheless, the DNA sequence through which two proteins with identical amino acid composition are ultimately produced differs between the two cases. In the case of family N126 the change detected on the DNA level was not identified in 150 Turkish control chromosomes, and segregated perfectly with the disease phenotype. Furthermore, the presence of p.Trp289X results in complete loss of the longer isoform produced by variant 1, a fact that could have a major effect on protein processing and function.

None of the identified changes affect residues that receive post-translational modifications such as sites of phosphorylation or glycosylation. Instead, prediction programs suggest a putative soluble protein with cytosolic distribution and no signal peptide along its sequence (SOSUI: http://bp.nuap.nagoya-u.ac.jp/sosui/sosui_submit.html). KCTD7 is predicted to be phosphorylated at p.Tyr131, p.Tyr162, p.Tyr163, p.Tyr200, and p.Tyr202 (PhosphoSitePlus http://www.phosphosite.org/homeAction.do) but not to have any *N*-glycosylation sites (PSORT: http://psort.hgc.jp/).

### *5.4.3.2 Subcellular localization of KCTD7 (IV)*

To address the intracellular distribution of KCTD7 constructs were generated whereby the protein was tagged N- or C-terminally with HA ($wt^{HA}$KCTD7 and wtKCTD7$^{HA}$). Overexpression of $wt^{HA}$KCTD7 in COS-1, BHK, and HeLa cells followed by immunofluorescence analysis revealed a wide distribution of KCTD7 across the cell, compatible with cytoplasmic localization. This localization was corroborated upon permeabilization of the PM of BHK cells with saponin, in which case a marked reduction of KCTD7 staining was observed in treated versus untreated cells. Additionally, KCTD7 immunostaining did not overlap with that of the cell structures evaluated (endosomes, lysosomes, Golgi, and PM).

To study the protein localization in a more natural context localization of endogenous KCTD7 was addressed in E14 mouse hippocampal neurons, in which similar to the overexpression experiments KCTD7 displayed a diffuse pattern of distribution across the cell soma. However, in the E14 neurons the protein was not confined only to the cell soma but was also transported to the varicosities of the neuronal axons and as far as the tips of the growing cones, suggesting an important role for KCTD7 across the whole cell. Similar to the experiments using tagged constructs, endogenous KCTD7 also does not seem to co-localize with any of the organelle markers tested (PDI, lysosomes, β-tubulin), with markers against pre- or post-synaptic vesicles (SYP and PSD-95, respectively) or with a vesicular GABA transporter (VGAT).

Taken together, the localization studies in different cell lines and cell types reveal that KCTD7 is a soluble cytosolic protein with expression throughout the cell. The latter has also been confirmed in an independent study whereby the localization of overexpressed KCTD7 was evaluated by immunofluorescence and confocal microscopy in COS7 cells (Azizieh et al., 2011). The concordant findings in the two independent studies in which the same experimental sets were used differing only in the antibodies utilized for detection of endogenous and tagged KCTD7, make it safe to conclude that KCTD7 is a cytosolic protein.

#### *5.4.3.3 Impact of KCTD7 mutations on the encoded protein (IV)*

To evaluate the impact of missense mutations on the encoded protein, the effect of three of them on intracellular localization and on protein expression was studied. The cytoplasmic distribution of KCTD7 was not affected in the presence of the missense p.Arg94Trp, p.Asp115Tyr, and p.Asn273Ile, representing patient mutations. In western blot analysis none of the three mutations seemed to alter the electrophoretic mobility of KCTD7 or its expression levels. These findings establish that the naturally occurring missense mutations do not interfere with protein localization and targeting or expression levels but are instead likely to exert their impact on the functional properties of KCTD7. The same can be anticipated for the frameshift p.Ile199SerfsX74 which introduces 74 novel amino acids after codon p.Ile199, significantly changing the produced protein composition which can also be targeted for nonsense-mediated decay. As for the in-frame deletion p.Trp289X, the mutation's impact is not easy to predict. Nevertheless, it can be hypothesized that in its presence the longest isoform of KCTD7 (isoform 1; UniProt accession number ID: Q96MP8-1) becomes absent, and the mode in which the mutant peptide is regulated and processed can be significantly different.

#### *5.4.3.4 Spatiotemporal expression of KCTD7 in the mouse brain (IV)*

To obtain insights into the spatiotemporal expression of KCTD7, different cell-types of the brain were assessed for expression of the protein in western blot analyses. Additionally, immunohistochemical sections from the brain of mice of different ages were used for detection of endogenous KCTD7. The protein was predominantly expressed in the neurons already at E14, but not in microglia and astrocyte cells. Immunohistochemistry analyses further supported these findings, showing a neuronal expression of KCTD7 throughout the brain with only few neuronal populations, such as the interneurons of the molecular layer of the cerebellum, being negative for KCTD7 immunostaining. The most prominent sites of KCTD7 expression were the cortex, the pyramidal and granular cell layers and the cerebellar Purkinje cells.

From a second standpoint, cerebellar protein lysates obtained from mice of different ages (P5, P14, 2 month, and 4 months old) showed that expression of KCTD7 is present already at P5 and it remains constant throughout the brain maturation. The same conclusion was supported by immunohistochemical evaluation of regional expression in P7, P10, P14, 2mo, and 4mo mouse brains, in which KCTD7 expression was constant. The most prominent sites of expression were the cortical neurons, in the granular and pyramidal cell layers of the hippocampus, as well as in the cerebellar Purkinje cells, in line with an independent study (Azizieh et al., 2011).

Taken together, the findings from these experiments denote the neuronal specificity of KCTD7 and its crucial role in both the developing and mature brain, suggesting an important role of the protein in neuronal function and survival.

The predominant role of KCTD7 in neurons was recently supported by a study reporting that when KCTD7 is overexpressed in neurons the cells become hyperpolarized and higher current values are required for action potential firing (Azizieh et al., 2011). Hence, KCTD7 deficiency is anticipated to lower the threshold of PM resting potential and promote the neuronal hyper-excitability that is a hallmark of epileptic events (Azizieh et al., 2011). Although it would be tempting to speculate that KCTD7 is involved in voltage-gated potassium ($K_v$) ion channel formation and function, justified by the homology to the $K_v$ T1 domain and the electrophysiology evidence, its cytoplasmic distribution does not support such a hypothesis. Instead it was shown that KCTD7 interacts with Cullin-3, a component of the E3 ubiquitin ligase (Azizieh et al., 2011), similar to another member of the KCTD family, KCTD5 (Dementieva et al., 2009). It was thus suggested that KCTD7 and KCTD5 might participate together with Cullin-3 in the formation of a functional E3 ubiquitin ligase that promotes proteasome-mediated degradation of proteins that indirectly affect ion channel function and conductance (Azizieh et al., 2011). The non-direct association of KCTD family members with ion channels was further shown in experiments where KCTD5 was explicitly shown to not be associated with any of the Kv4.2, Kv3.4, Kv2.1, or Kv1.2 channels (Dementieva et al., 2009). Finally, KCTD members can also be associated with many other phenotypes as in the case of KCTD8, KCTD12, and KCTD16 that act as auxiliary subunits of the $GABA_B$ receptor (Schwenk

et al., 2010; Bartoi et al., 2010), KCTD11 that acts as a tumor suppressor causing neuroblastoma when defective (Argenti et al., 2005), and KCTD15 that confers susceptibility to adult onset obesity (Elks et al., 2010).

To date the precise role of KCTD7 remains elusive. The protein's neuronal specificity is in favor of KCTD7 having a role in neuronal survival and protection, which could be displayed through maintenance of their proper electrophysiological properties. In order to provide answers to these open questions, the exact nature of the role of KCTD7 awaits to be addressed in future detailed functional studies.

#### *5.4.3.5 Clinical phenotype caused by mutations in KCTD7 (IV)*

The disease associated with mutations in *KCTD7* begins at a mean age of 19 months, with the youngest reported patient having disease onset at 10 months of age and the oldest at 3 years (van Bogaert et al., 2007; study IV). The presenting symptom was epilepsy in almost all patients. Only in patient N12604 was ataxia reported as the presenting symptom. All reported patients develop myoclonic seizures, although other seizure types such as generalized tonic-clonic, atonic and hypomotor seizures have also been reported (van Bogaert et al., 2007; study IV). The seizures have been reported to either fully or partially be controlled with combined polytherapy (study IV; van Bogaert et al., 2007). Three patients developed refractory seizures. The course of the disease is progressive, with the patients progressively losing their mental and motor skills, their speech, becoming unambulatory within 1-2 years after disease onset, and being unable to communicate with their environment. All metabolic tests performed were normal for all patients for whom this information was available (study IV; van Bogaert et al., 2007). The retinal examination of all reported patients was normal, although some have been reported to not be able to follow objects, which is thought to be caused by the mental regression. Neurodegeneration, as determined by MRI analysis revealing cerebellar and cerebral atrophy, is described in all *KCTD7* positive patients.

Clinically, KCTD7 disease greatly resembles the classic and variant LINCL subtypes of NCL. However, *KCTD7* does not belong to the NCL spectrum of genes. Many reasons

argue in favor of the latter conclusion. First, tissue biopsy examination of four patients (l3, N15103 and Pak4 in this study; van Bogaert et al., 2007) did not contain the characteristic NCL storage material which represents a hallmark pathological finding of NCLs. Second, none of the 22 confirmed NCL patients carried mutations in *KCTD7*. Finally, the *KCTD7* mutation positive patients did not have findings of retinal abnormality on opthalmological examination. These small clinical differences provide some guidance to physicians when providing differential diagnosis of patients with early childhood onset neurological deterioration. When patients have myoclonic seizures, normal retinal findings, normal biochemical test findings and remain negative for hallmark pathological findings, *KCTD7* screening should be considered. Nevertheless, clinical distinction among the neurodegenerative childhood disorders cannot be achieved on the basis of age of onset or symptomatology alone. This leaves molecular genetic testing for the causative genes as the only alternative towards providing an accurate diagnosis.

# 6. CONCLUSIONS AND FUTURE PROSPECTS

This thesis expanded the molecular genetic background of CLN7 disease, caused by mutations in *MFSD8*, by identifying the pathogenic mutations in 36 patients and by characterizing the majority of them also clinically. The identification of 36 patients from nine different ethnic backgrounds has not only established *MFSD8* as a relatively common cause of NCLs, and consequently childhood symptomatic epilepsy, but also emphasized the worldwide distribution of defects in this gene, which should be considered in patients with a vLINCL phenotype. Unfortunately, the efforts to provide tools that would facilitate diagnosis of CLN7 disease, have not revealed differences in disease onset and progression, in symptomatology, and/or in pathology. Instead, clinical diagnosis of such syndromes based on the clinical picture of the patient alone is very difficult today. Differential diagnosis is especially challenging in the clinically homogeneous and genetically heterogeneous LINCL group where molecular genetic testing remains the only alternative to provide an accurate diagnosis. Identification of common or population specific mutations can greatly facilitate the molecular diagnosis in groups such as the Roma patients from the former Czechoslovakia described here, allowing the families to benefit from prenatal testing and carrier status evaluation analyses.

The protein encoded by *MFSD8* was the first among the NCL proteins to be predicted to have a putative transporter function. The expression of MFSD8 is ubiquitous in the brain with the neurons being the primary cell-type expressing MFSD8. The lysosomal membrane is the intracellular site where MFSD8 is delivered and performs its transporter role. Identification of the major lysosomal determinant mediating this sorting has been of primary importance towards directing MFSD8 to the PM where the protein's function and properties, as well as its substrate specificity will be more easily addressed. Defects in this gene have been established to not interfere with the protein's normal distribution, suggesting that they instead affect the functional properties of the resultant peptides. Although light has been shed on the primary protein properties it is not until more experiments have been performed that the function of MFSD8 will be fully elucidated.

In this study a total of 85 patients with childhood onset PME have been provided with a molecular genetic diagnosis. Despite the phenotypic convergence seen among some of them, exemplified especially in the case of the vLINCL subtypes, several cases of phenotypic divergence have also been detected. Hence, many patients with mutations in the same gene and strikingly different phenotypes have been identified, such as the *MFSD8* positive Dutch patient with onset at 11 years and protracted disease course. It is now becoming clearer that although a fraction of such cases can be explained by the severity of the mutation (missense versus nonsense) and the domain it affects, some other cases, especially in affected siblings carrying the same disease-causing mutation(s), can only be explained by the contribution to the phenotype of genetic modifiers. One such tempting hypothesis is to evaluate the possibility of the NCL loci having epistatic interactions among them.

Several clinical entities exist today with phenotypes that clinically resemble NCLs but lack the confirmatory detection of autofluorescent material on EM examination. In clinical practice such patients present as PME and it is not until the genetic defects have been identified that they can be provided with an accurate diagnosis. Disorders that can be confounded with childhood onset NCL or PME can be NBIA and INAD, caused by mutations in *PLA2G6* and characterized, though not always, by high brain iron accumulation in the brain.

In the attempt to further dissect the molecular background of childhood onset PMEs, targeted next-generation sequencing identified a very rare homozygous change in the *TXNDC6* gene in one inbred family with two affected siblings. However, the inability to identify a second family with mutations in the same gene means that TXNDC6 cannot be unequivocally linked to disease causality. Instead its relevance to disease must await the evaluation of larger collections of patients with similar PME phenotypes. Unlike *TXNDC6*, the molecular basis of a distinct clinical entity, EPM type 3, has been characterized in seven families with mutations in *KCTD7*. Although the gene had previously been described as the underlying cause of EPM3 in a single family, the disease role of *KCTD7* is now well established having identified seven more families carrying six different defects in it. The clinical phenotype caused by mutations in *KCTD7* has onset in

early childhood, with the patients usually presenting with seizures and having a rapidly progressive course that leads to severe debilitation within 3 years of disease onset. Therefore, *KCTD7* should be added to the panel of genes evaluated in differential diagnosis of patients with early onset PMEs. From a functional standpoint, the prominent role of KCTD7 across the cytoplasmic matrix of neurons, together with the constant expression from the embryonic life until brain maturation and the fact that KCTD7 overexpression alters the neuronal excitability, suggest a unique role in nervous system development and possibly regulation for KCTD7.

Despite the progress made in identifying several PME-associated genes, several patients (130 in this study alone) remain without a molecular genetic diagnosis in clinical practice. The advent of the new technologies together with collaborative efforts in which the clinically thoroughly characterized patient samples are combined is likely to speed up disease gene identification in the years to come.

# 7. ACKNOWLEDGMENTS

The work performed in the context of this thesis was conducted in Folkhälsan Institute of Genetics, Neuroscience Center and Department of Medical Genetics of the University of Helsinki during the period 2007-2011. I would like to acknowledge the heads of the aforementioned institutes, Professors Anna-Elina Lehesjoki, Heikki Rauvala, and Irma Järvelä, for the excellent working facilities and resources.

This study was financed and supported by the Helsinki Biomedical Graduate School to whom I am deeply grateful for the PhD fellow position through the years 2008-2011. I am also thankful for the support I received from the Biomedicum Helsinki Foundation, the Emil Aaltonen Foundation, the Oskar Öflund Foundation, the Epilepsy Research Foundation, the Chancellor travel fund of the University of Helsinki, and the Orion Farmos Research Foundation.

I am mostly grateful to my supervisor, professor Anna-Elina Lehesjoki, who not only gave me the opportunity to work in her group before I had even completed my Bachelor degree, but has also been a patient and wise guide throughout my studies. I especially appreciate the patience with which she has faced all my worries as to whether the work done would ever shape up to become a complete thesis, but also for the personal advise and discussions that have been invaluable to me.

I would also like to express my gratitude to Docent Outi Kopra who although became involved in my work only during the last year of my studies has taught me so much on a scientific but also personal level. I wish I could have worked with Outi longer, because her enthusiasm and eagerness to find answers to questions are contagious, and because she is a teacher every student would be privileged to have.

I am grateful to Professor Aarno Palotie for accepting me as a visiting student to the Wellcome Trust Sanger Institute in Cambridge, UK. Aarno was always welcoming, kind and facing everything with a great sense of humor. His quality as a leader can also be appreciated by the excellent atmosphere in his group, which consists of inspired and

motivated people. Of them I am indebted to Verneri Anttila that has helped me with the analysis of the data. Without him I would now be far from done in my studies.

Professor Mark Gardiner and Docent Marjo Kestilä are warmly thanked for serving as the examiners of my thesis. Their comments and considerations were extremely valuable for improving the overall quality of the manuscript. Marjo has also been part of my thesis committee and together with Professor Irma Järvelä, have provided valuable support and guidance through the conduction of this thesis, always with a positive and motivating spirit. Irma is also thanked for accepting to serve as the custos of the dissertation.

I am indebted to Anna-Kaisa Anttonen for having taught me the basics of genetics upon my arrival in Folkhälsan and for always being there to answer any questions I was having, always with a smile on. Eija Siintola is also warmly thanked for her guidance through the NCL project, and the teaching in the lab. Aija Kyttälä and Anu Jalanko are also thanked for "hosting" me in the National Institute of Health and Welfare and introducing me to the field of molecular biology.

A special thanks is reserved for Jodie Painter for editing the language of this manuscript. Jodie is a colleague and friend that I have been missing since she left the lab to return to the "Down Under".

Isa Uski, Hanna Hellgren, and Paula Hakala are especially thanked for their expert technical assistance and the positive energy they have been transmitting to everyone around them in the lab. Aila Riikonen, Marjatta Valkama, Jaana Welin-Haapamäki, and Sinikka Lindh for helping with the practical matters. My roommates Saara, Anna, Kaisa and Hanna are thanked for all the discussions, and for making the everyday life look anything else but boring. Ann-Liz with her cheerful spirit and sunny wishes and hugs even in the darkest winter days has always been a source of positivity. Tarja, pH, and Jaakko have been the best company for a coffee break. I would also like to thank all the other members of the group and especially Maria Lehtinen, Mervi Kuronen, Vilma-Lotta Lehtokari, Inken Korber, Olesya Okuneva with whom I share very pleasant memories.

I have been lucky enough to have met exceptional people during my stay in Helsinki, that mean a lot to me. Harry and Teresa have been my "foster parents" in Finland, and

although I've been far away from home they made me feel I was not alone. I also thank Anna, Benediktos, Filippos, Emmanuel, Marianna, Helena and Jose, Tzina, and Efi for their friendship and all the blissful moments we shared. I am grateful to my "Greek and Italian support team" consisting of Anna, Dimitris, Akis, Aleksia, Voula, Spyros, Pavlos, Giannis, Artemis, Natassa, Axilleas, Andreas, Barbara and Edoardo for the support and love they have given me during my stay here.

I reserve the biggest thanks for my family that has always stood by my side, my brother Thanos, and my parents, Kalliopi and Michalis Kousis. My mother has been my biggest supporter and the first person to call and share my news. My father has always been my driving force, although from the skies, and the reason I kept pushing forward even at the most difficult times. Finally, I am most grateful to Maurizio that has been by my side through every day of this experience, making everything look easy. His support means a lot to me and I owe a big part of this work to him.

Helsinki, December 2011

Maria Kousi

# 8. REFERENCES

1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. Nature 2010; 467: 1061-73.

Ahtiainen L, Van Diggelen OP, Jalanko A, Kopra O. Palmitoyl protein thioesterase 1 is targeted to the axons in neurons. J Comp Neurol 2003; 455: 368-77.

Aiello C, Terracciano A, Simonati A, Discepoli G, Cannelli N, Claps D, et al. Mutations in MFSD8/CLN7 are a frequent cause of variant-late infantile neuronal ceroid lipofuscinosis. Hum Mutat 2009; 30: E530-40.

Aldahmesh MA, Al-Hassnan ZN, Aldosari M, Alkuraya FS. Neuronal ceroid lipofuscinosis caused by MFSD8 mutations: A common theme emerging. Neurogenetics 2009; 10: 307-11.

Antonarakis SE, McKusick VA. OMIM passes the 1,000-disease-gene mark. Nat Genet 2000; 25: 11.

Antonarakis SE, Beckmann JS. Mendelian disorders deserve more attantion. Nat Rev Genet 2006; 7: 277-282.

Argenti B, Gallo R, Di Marcotullio L, Ferretti E, Napolitano M, Canterini S, et al. Hedgehog antagonist REN(KCTD11) regulates proliferation and apoptosis of developing granule cell progenitors. J Neurosci 2005; 25: 8338-46.

Aridor M, Bannykh SI, Rowe T, Balch WE. Sequential coupling between COPII and COPI vesicle coats in endoplasmic reticulum to golgi transport. J Cell Biol 1995; 131: 875-93.

Arsov T, Smith KR, Damiano J, Franceschetti S, Canafoglia L, Bromhead CJ, et al. Kufs disease, the major adult form of neuronal ceroid lipofuscinosis, caused by mutations in CLN6. Am J Hum Genet 2011; 88: 566-73.

Awano T, Katz ML, O'Brien DP, Taylor JF, Evans J, Khan S, et al. A mutation in the cathepsin D gene (CTSD) in american bulldogs with neuronal ceroid lipofuscinosis. Mol Genet Metab 2006b; 87: 341-8.

Awano T, Katz ML, O'Brien DP, Sohar I, Lobel P, Coates JR, et al. A frame shift mutation in canine TPP1 (the ortholog of human CLN2) in a juvenile dachshund with neuronal ceroid lipofuscinosis. Mol Genet Metab 2006a; 89: 254-60.

Azizieh R, Orduz D, Van Bogaert P, Bouschet T, Rodriguez W, Schiffmann SN, et al. Progressive myoclonic epilepsy-associated gene KCTD7 is a regulator of potassium conductance in neurons. Mol Neurobiol 2011; 44: 111-121.

Barlowe C. COPII and selective export from the endoplasmic reticulum. Biochim Biophys Acta 1998; 1404: 67-76.

Barohn RJ, Dowd DC, Kagan-Hallet KS. Congenital ceroid-lipofuscinosis. Pediatr Neurol 1992; 8: 54-9.

Bartoi T, Rigbolt KT, Du D, Kohr G, Blagoev B, Kornau HC. GABAB receptor constituents revealed by tandem affinity purification from transgenic mice. J Biol Chem 2010; 285: 20625-33.

Bassuk AG, Wallace RH, Buhr A, Buller AR, Afawi Z, Shimojo M, et al. A homozygous mutation in human PRICKLE1 causes an autosomal-recessive progressive myoclonus epilepsy-ataxia syndrome. Am J Hum Genet 2008; 83: 572-81.

Berkovic SF, Andermann F, Carpenter S, Wolfe LS. Progressive myoclonus epilepsies: Specific causes and diagnosis. N Engl J Med 1986; 315: 296-305.

Berkovic SF, Dibbens LM, Oshlack A, Silver JD, Katerelos M, Vears DF, et al. Array-based gene discovery with three unrelated subjects shows SCARB2/LIMP-2 deficiency causes myoclonus epilepsy and glomerulosclerosis. Am J Hum Genet 2008; 82: 673-84.

Berkovic SF, Carpenter S, Andermann F, Andermann E, Wolfe LS. Kuf's disease: a critical reappraisal. Brain 1988; 111: 27-62.

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

Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Margulies EH, et al. Identification and analysis of functional elements in 1% of the human genome by ENCODE pilot project. Nature 2007; 447: 799-816.

Bonifacino JS, Traub LM. Signals for sorting of transmembrane proteins to endosomes and lysosomes. Annu Rev Biochem 2003; 72: 395-447.

Bonten E, van der Spoel A, Fornerod M, Grosveld G, d'Azzo A. Characterization of human lysosomal neuraminidase defines the molecular basis of the metabolic storage disorder sialidosis. Genes Dev 1996; 10: 3156-3169.

Broman KW, Weber JL. Long homozygous chromosomal segments in reference families from the centre d'etude du polymorphisme humain. Am J Hum Genet 1999; 65: 1493-500.

Brown NJ, Corner BD, Dodgson MC. A second case in the same family of congenital familial cerebral lipoidosis resembling amaurotic family idiocy. Arch Dis Child 1954; 29: 48-54.

Canfield WM, Johnson KF, Ye RD, Gregory W, Kornfeld S. Localization of the signal for rapid internalization of the bovine cation-independent mannose 6-phosphate/insulin-like growth factor-II receptor to amino acids 24-29 of the cytoplasmic tail. J Biol Chem 1991; 266: 5682-8.

Cannelli N, Cassandrini D, Bertini E, Striano P, Fusco L, Gaggero R, et al. Novel mutations in CLN8 in italian variant late infantile neuronal ceroid lipofuscinosis: Another genetic hit in the mediterranean. Neurogenetics 2006; 7: 111-7.

Cao Y, Espinola JA, Fossale E, Massey AC, Cuervo AM, MacDonald ME, et al. Autophagy is disrupted in a knock-in mouse model of juvenile neuronal ceroid lipofuscinosis. J Biol Chem 2006; 281: 20483-93.

Capovilla G, Berg AT, Cross JH, Moshe SL, Vigevano F, Wolf P, et al. Conceptual dichotomies in classifying epilepsies: Partial versus generalized and idiopathic versus symptomatic (april 18-20, 2008, monreale, italy). Epilepsia 2009.

Carr IM, Flintoff KJ, Taylor GR, Markham AF, Bonthron DT. Interactive visual analysis of SNP data for rapid autozygosity mapping in consanguineous families. Hum Mutat 2006; 27: 1041-6.

Cartegni L, Chew SL, Krainer AR. Listening to silence and understanding nonsense: Exonic mutations that affect splicing. Nat Rev Genet 2002; 3: 285-98.

Chan EM, Young EJ, Ianzano L, Munteanu I, Zhao X, Christopoulos CC, et al. Mutations in NHLRC1 cause progressive myoclonus epilepsy. Nat Genet 2003; 35: 125-7.

Chang JW, Choi H, Kim HJ, Jo DG, Jeon YJ, Noh JY, et al. Neuronal vulnerability of CLN3 deletion to calcium-induced cytotoxicity is mediated by calsenilin. Hum Mol Genet 2007b; 16: 317-26.

Chang YF, Imam JS, Wilkinson MF. The nonsense-mediated decay RNA surveillance pathway. Annu Rev Biochem 2007a; 76: 51-74.

Collins FS. Positional cloning moves from perditional to traditional. Nat Genet 1995; 9: 347-50.

Collins FS. Positional cloning: Let's not call it reverse anymore. Nat Genet 1992; 1: 3-6.

Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, et al. Origins and functional impact of copy number variation in the human genome. Nature 2010; 464: 704-12.

Corbett MA, Schwake M, Bahlo M, Dibbens LM, Lin M, Gandolfo LC, et al. A mutation in the golgi qb-SNARE gene GOSR2 causes progressive myoclonus epilepsy with early ataxia. Am J Hum Genet 2011; 88: 657-63.

Cordaux R, Batzer MA. The impact of retrotransposons on human genome evolution. Nat Rev Genet 2009; 10: 691-703.

Cotman SL, Vrbanac V, Lebel LA, Lee RL, Johnson KA, Donahue LR, et al. Cln3(Deltaex7/8) knock-in mice with the common JNCL mutation exhibit progressive neurologic disease that begins before birth. Hum Mol Genet 2002; 11: 2709-21.

Cowen D, Olmstead EV. Infantile neuroaxonal dystrophy. J Neuropathol Exp Neurol 1963; 22: 175-236.

Crick FH, Barnett L, Brenner S, Watts-Tobin RJ. General nature of the genetic code for proteins. Nature 1961; 192: 1227-32.

Das AK, Lu JY, Hofmann SL. Biochemical analysis of mutations in palmitoyl-protein thioesterase causing infantile and late-onset forms of neuronal ceroid lipofuscinosis. Hum Mol Genet 2001; 10: 1431-9.

Das AK, Becerra CH, Yi W, Lu JY, Siakotos AN, Wisniewski KE, et al. Molecular genetics of palmitoyl-protein thioesterase deficiency in the U.S. J Clin Invest 1998; 102: 361-70.

de Voer G, van der Bent P, Rodrigues AJ, van Ommen GJ, Peters DJ, Taschner PE. Deletion of the caenorhabditis elegans homologues of the CLN3 gene, involved in human juvenile neuronal ceroid lipofuscinosis, causes a mild progeric phenotype. J Inherit Metab Dis 2005; 28: 1065-80.

Deininger PL, Batzer MA. Mammalian retroelements. Genome Res 2002; 12: 1455-65.

Delgado-Escueta AV, Ganesh S, Yamakawa K. Advances in the genetics of progressive myoclonus epilepsy. Am J Med Genet 2001; 106: 129-38.

Dementieva IS, Tereshko V, McCrossan ZA, Solomaha E, Araki D, Xu C, et al. Pentameric assembly of potassium channel tetramerization domain-containing protein 5. J Mol Biol 2009; 387: 175-91.

Dibbens LM, Michelucci R, Gambardella A, Andermann F, Rubboli G, Bayly MA. SCARB2 mutations in progressive myoclonus epilepsy (PME) without renal failure. 2009; 66: 532-536.

Eliason SL, Stein CS, Mao Q, Tecedor L, Ding SL, Gaines DM, et al. A knock-in reporter model of batten disease. J Neurosci 2007; 27: 9826-34.

Elks CE, Loos RJ, Sharp SJ, Langenberg C, Ring SM, Timpson NJ, et al. Genetic markers of adult obesity risk are associated with greater early infancy weight gain and growth. PLoS Med 2010; 7: e1000284.

Elleder M, Sokolova J, Hrebicek M. Follow-up study of subunit c of mitochondrial ATP synthase (SCMAS) in batten disease and in unrelated lysosomal disorders. Acta Neuropathol 1997a; 93: 379-90.

Elleder M, Franc J, Kraus J, Nevsimalova S, Sixtova K, Zeman J. Neuronal ceroid lipofuscinosis in the czech republic: Analysis of 57 cases. report of the 'prague NCL group'. Eur J Paediatr Neurol 1997b; 1: 109-14.

Elleder M, Dvorakova L, Stolnaja L, Vlaskova H, Hulkova H, Druga R, et al. Atypical CLN2 with later onset and prolonged course: A neuropathologic study showing different sensitivity of neuronal subpopulations to TPP1 deficiency. Acta Neuropathol 2008; 116: 119-24.

Ellegren H. Microsatellite mutations in the germline: Implications for evolutionary inference. Trends Genet 2000; 16: 551-8.

ENCODE Project Consortium. The ENCODE (ENCyclopedia of DNA elements) project. Science 2004; 306: 636-40.

Erickson AH, Conner GE, Blobel G. Biosynthesis of a lysosomal enzyme. partial structure of two transient and functionally distinct NH2-terminal sequences in cathepsin D. J Biol Chem 1981; 256: 11224-31.

Erkman L, McEvilly RJ, Luo L, Ryan AK, Hooshmand F, O'Connell SM, et al. Role of transcription factors brn-3.1 and brn-3.2 in auditory and visual system development. Nature 1996; 381: 603-6.

Ezaki J, Kominami E. The intracellular location and function of proteins of neuronal ceroid lipofuscinoses. Brain Pathol 2004; 14: 77-85.

Ezaki J, Takeda-Ezaki M, Oda K, Kominami E. Characterization of endopeptidase activity of tripeptidyl peptidase-I/CLN2 protein which is deficient in classical late infantile neuronal ceroid lipofuscinosis. Biochem Biophys Res Commun 2000; 268: 904-8.

Ezaki J, Takeda-Ezaki M, Koike M, Ohsawa Y, Taka H, Mineki R, et al. Characterization of Cln3p, the gene product responsible for juvenile neuronal ceroid lipofuscinosis, as a lysosomal integral membrane glycoprotein. J Neurochem 2003; 87: 1296-308.

Fan H, Chu JY. A brief review of short tandem repeat mutation. Genomics Proteomics Bioinformatics 2007; 5: 7-14.

Fiers W, Contreras R, Duerinck F, Haegeman G, Iserentant D, Merregaert J, et al. Complete nucleotide sequence of bacteriophage MS2 RNA: Primary and secondary structure of the replicase gene. Nature 1976; 260: 500-7.

Fritchie K, Siintola E, Armao D, Lehesjoki AE, Marino T, Powell C, et al. Novel mutation and the first prenatal screening of cathepsin D deficiency (CLN10). Acta Neuropathol 2009; 117: 201-8.

Frugier T, Mitchell NL, Tammen I, Houweling PJ, Arthur DG, Kay GW, et al. A new large animal model of CLN5 neuronal ceroid lipofuscinosis in borderdale sheep is caused by a nucleotide substitution at a consensus splice site (c.571+1G>A) leading to excision of exon 3. Neurobiol Dis 2008; 29: 306-15.

Gachet Y, Codlin S, Hyams JS, Mole SE. btn1, the schizosaccharomyces pombe homologue of the human batten disease gene CLN3, regulates vacuole homeostasis. J Cell Sci 2005; 118: 5525-36.

Gao H, Boustany RM, Espinola JA, Cotman SL, Srinidhi L, Antonellis KA, et al. Mutations in a novel CLN6-encoded transmembrane protein cause variant neuronal ceroid lipofuscinosis in man and mouse. Am J Hum Genet 2002; 70: 324-35.

Garborg I, Torvik A, Hals J, Tangsrud SE, Lindemann R. Congenital neuronal ceroid lipofuscinosis. A case report. Acta Pathol Microbiol Immunol Scand A 1987; 95: 119-25.

Gerstein MB, Bruce C, Rozowsky JS, Zheng D, Du J, Korbel JO, et al. What is a gene, post-ENCODE? history and updated definition. Genome Res 2007; 17: 669-81.

Gitschier J, Wood WI, Goralka TM, Wion KL, Chen EY, Eaton DH, et al. Characterization of the human factor VIII gene. Nature 1984; 312: 326-30.

Goebel HH, Braak H. Adult neuronal ceroid-lipofuscinosis. Clin Neuropathol 1989; 8: 109-19.

Gotz A, Tyynismaa H, Euro L, Ellonen P, Hyotylainen T, Ojala T, et al. Exome sequencing identifies mitochondrial alanyl-tRNA synthetase mutations in infantile mitochondrial cardiomyopathy. Am J Hum Genet 2011; 88: 635-42.

Gregory A, Westaway SK, Holm IE, Kotzbauer PT, Hogarth P, Sonek S, et al. Neurodegeneration associated with genetic defects in phospholipase A(2). Neurology 2008; 71: 1402-9.

Gupta P, Soyombo AA, Atashband A, Wisniewski KE, Shelton JM, Richardson JA, et al. Disruption of PPT1 or PPT2 causes neuronal ceroid lipofuscinosis in knockout mice. Proc Natl Acad Sci U S A 2001; 98: 13566-71.

Haltia M. The neuronal ceroid-lipofuscinoses. J Neuropathol Exp Neurol 2003; 62: 1-13.

Hartikainen JM, Ju W, Wisniewski KE, Moroziewicz DN, Kaczmarski AL, McLendon L, et al. Late infantile neuronal ceroid lipofuscinosis is due to splicing mutations in the CLN2 gene. Mol Genet Metab 1999; 67: 162-8.

Haskell RE, Carr CJ, Pearce DA, Bennett MJ, Davidson BL. Batten disease: Evaluation of CLN3 mutations on protein localization and function. Hum Mol Genet 2000; 9: 735-44.

Hassink GC, Zhao B, Sompallae R, Altun M, Gastaldello S, Zinin NV, et al. The ER-resident ubiquitin-specific protease 19 participates in the UPR and rescues ERAD substrates. EMBO Rep 2009; 10: 755-61.

Hastbacka J, de la Chapelle A, Kaitila I, Sistonen P, Weaver A, Lander E. Linkage disequilibrium mapping in isolated founder populations: Diastrophic dysplasia in finland. Nat Genet 1992; 2: 204-11.

Hauri HP, Schweizer A. The endoplasmic reticulum-golgi intermediate compartment. Curr Opin Cell Biol 1992; 4: 600-8.

Heine C, Quitsch A, Storch S, Martin Y, Lonka L, Lehesjoki AE, et al. Topology and endoplasmic reticulum retention signals of the lysosomal storage disease-related membrane protein CLN6. Mol Membr Biol 2007; 24: 74-87.

Heinonen O, Kyttälä A, Lehmus E, Paunio T, Peltonen L, Jalanko A. Expression of palmitoyl protein thioesterase in neurons. Mol Genet Metab 2000; 69: 123-9.

Helbig HC, Mefford AJ, Sharp M, Guipponi M, Fichera A, Franke H, et al. 15q13.3 microdeletions increase risk of idiopathic generalized epilepsy. Nat Genet 2009; 41: 160–162.

Hellsten E, Vesa J, Olkkonen VM, Jalanko A, Peltonen L. Human palmitoyl protein thioesterase: Evidence for lysosomal targeting of the enzyme and disturbed cellular routing in infantile neuronal ceroid lipofuscinosis. EMBO J 1996; 15: 5240-5.

Hentze MW, Kulozik AE. A perfect message: RNA surveillance and nonsense-mediated decay. Cell 1999; 96: 307-10.

Herva R, Tyynela J, Hirvasniemi A, Syrjakallio-Ylitalo M, Haltia M. Northern epilepsy: A novel form of neuronal ceroid-lipofuscinosis. Brain Pathol 2000; 10: 215-22.

Hickey AJ, Chotkowski HL, Singh N, Ault JG, Korey CA, MacDonald ME, et al. Palmitoyl-protein thioesterase 1 deficiency in drosophila melanogaster causes accumulation of abnormal storage material and reduced life span. Genetics 2006; 172: 2379-90.

Hirvasniemi A, Lang H, Lehesjoki AE, Leisti J. Northern epilepsy syndrome: An inherited childhood onset epilepsy with associated mental deterioration. J Med Genet 1994; 31: 177-82.

Hoflack B, Kornfeld S. Purification and characterization of a cation-dependent mannose 6-phosphate receptor from murine P388D1 macrophages and bovine liver. J Biol Chem 1985a; 260: 12008-14.

Hoflack B, Kornfeld S. Lysosomal enzyme binding to mouse P388D1 macrophage membranes lacking the 215-kDa mannose 6-phosphate receptor: Evidence for the existence of a second mannose 6-phosphate receptor. Proc Natl Acad Sci U S A 1985b; 82: 4428-32.

Hoh J, Ott J. Genetic dissection of diseases: Design and methods. Curr Opin Genet Dev 2004; 14: 229-32.

Holmberg V, Lauronen L, Autti T, Santavuori P, Savukoski M, Uvebrant P, et al. Phenotype-genotype correlation in eight patients with finnish variant late infantile NCL (CLN5). Neurology 2000; 55: 579-81.

Hortnagel K, Nardocci N, Zorzi G, Garavaglia B, Botz E, Meitinger T, et al. Infantile neuroaxonal dystrophy and pantothenate-kinase-associated neurodegeneration: Locus heterogeneity. Neurology 2004; 63: 922-4.

Houweling PJ, Cavanagh JA, Palmer DN, Frugier T, Mitchell NL, Windsor PA, et al. Neuronal ceroid lipofuscinosis in devon cattle is caused by a single base duplication (c.662dupG) in the bovine CLN5 gene. Biochim Biophys Acta 2006; 1762: 890-7.

Humphreys S, Lake BD, Scholtz CL. Congenital amaurotic idiocy--a pathological, histochemical, biochemical and ultrastructural study. Neuropathol Appl Neurobiol 1985; 11: 475-84.

International HapMap Consortium. A haplotype map of the human genome. Nature 2005; 437: 1299-320.

International HapMap Consortium. The international HapMap project. Nature 2003; 426: 789-96.

International HapMap Consortium, Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, et al. A second generation human haplotype map of over 3.1 million SNPs. Nature 2007; 449: 851-61.

International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. Nature 2004; 431: 931-45.

Isosomppi J, Vesa J, Jalanko A, Peltonen L. Lysosomal localization of the neuronal ceroid lipofuscinosis CLN5 protein. Hum Mol Genet 2002; 11: 885-91.

Jalanko A, Vesa J, Manninen T, von Schantz C, Minye H, Fabritius AL, et al. Mice with Ppt1Deltaex4 mutation replicate the INCL phenotype and show an inflammation-associated loss of interneurons. Neurobiol Dis 2005; 18: 226-41.

Järvelä I, Lehtovirta M, Tikkanen R, Kyttälä A, Jalanko A. Defective intracellular transport of CLN3 is the molecular basis of batten disease (JNCL). Hum Mol Genet 1999; 8: 1091-8.

Järvelä I, Autti T, Lamminranta S, Aberg L, Raininko R, Santavuori P. Clinical and magnetic resonance imaging findings in batten disease: Analysis of the major mutation (1.02-kb deletion). Ann Neurol 1997; 42: 799-802.

Järvelä I, Sainio M, Rantamaki T, Olkkonen VM, Carpen O, Peltonen L, et al. Biosynthesis and intracellular targeting of the CLN3 protein defective in batten disease. Hum Mol Genet 1998; 7: 85-90.

Jeffreys AJ. Genetic fingerprinting. Nat Med 2005; 11: 1035-9.

Jenkins D, Seelow D, Jehee FS, Perlyn CA, Alonso LG, Bueno DF, et al. RAB23 mutations in Carpenter syndrome imply an unexpected role for hedgehog signaling in cranial-suture development and obesity. Am J Hum Genet 2007; 80: 1162-70.

Jervis GA. Phenylpyruvic oligophrenia deficiency of phenylalanine-oxidizing system. Proc Soc Exp Biol Med 1953; 82: 514-5.

Jolly RD, Shimada A, Dopfmer I, Slack PM, Birtles MJ, Palmer DN. Ceroid-lipofuscinosis (batten's disease): Pathogenesis and sequential neuropathological changes in the ovine model. Neuropathol Appl Neurobiol 1989; 15: 371-83.

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

Katsanis N, Lupski JR, Beales PL. Exploring the molecular basis of bardet-biedl syndrome. Hum Mol Genet 2001; 10: 2293-9.

Katz ML, Khan S, Awano T, Shahid SA, Siakotos AN, Johnson GS. A mutation in the CLN8 gene in english setter dogs with neuronal ceroid-lipofuscinosis. Biochem Biophys Res Commun 2005; 327: 541-7.

Katz ML, Shibuya H, Liu PC, Kaur S, Gao CL, Johnson GS. A mouse gene knockout model for juvenile ceroid-lipofuscinosis (batten disease). J Neurosci Res 1999; 57: 551-6.

Katz ML, Farias FH, Sanders DN, Zeng R, Khan S, Johnson GS, et al. A missense mutation in canine CLN6 in an australian shepherd with neuronal ceroid lipofuscinosis. J Biomed Biotechnol 2011; 2011: 198042.

Khateeb S, Flusser H, Ofir R, Shelef I, Narkis G, Vardi G, et al. PLA2G6 mutation underlies infantile neuroaxonal dystrophy. Am J Hum Genet 2006; 79: 942-8.

Kohan R, Cismondi IA, Kremer RD, Muller VJ, Guelbert N, Anzolini VT, et al. An integrated strategy for the diagnosis of neuronal ceroid lipofuscinosis types 1 (CLN1) and 2 (CLN2) in eleven latin american patients. Clin Genet 2009; 76: 372-82.

Kollmann K, Mutenda KE, Balleininger M, Eckermann E, von Figura K, Schmidt B, et al. Identification of novel lysosomal matrix proteins by proteome analysis. Proteomics 2005; 5: 3966-78.

Kopra O, Vesa J, von Schantz C, Manninen T, Minye H, Fabritius AL, et al. A mouse model for finnish variant late infantile neuronal ceroid lipofuscinosis, CLN5, reveals neuropathology associated with early aging. Hum Mol Genet 2004; 13: 2893-906.

Kornfeld R, Kornfeld S. Assembly of asparagine-linked oligosaccharides. Annu Rev Biochem 1985; 54: 631-64.

Kuehn MJ, Schekman R. COPII and secretory cargo capture into transport vesicles. Curr Opin Cell Biol 1997; 9: 477-83.

Kurze AK, Galliciotti G, Heine C, Mole SE, Quitsch A, Braulke T. Pathogenic mutations cause rapid degradation of lysosomal storage disease-related membrane protein CLN6. Hum Mutat 2010; 31: E1163-74.

Kyttälä A, Ihrke G, Vesa J, Schell MJ, Luzio JP. Two motifs target batten disease protein CLN3 to lysosomes in transfected nonneuronal and neuronal cells. Mol Biol Cell 2004; 15: 1313-23.

Lake BD, Hall NA. Immunolocalization studies of subunit c in late-infantile and juvenile batten disease. J Inherit Metab Dis 1993; 16: 263-6.

Landegren U, Nilsson M, Kwok PY. Reading bits of genetic information: Methods for single-nucleotide polymorphism analysis. Genome Res 1998; 8: 769-76.

Lander ES, Botstein D. Homozygosity mapping: A way to map human recessive traits with the DNA of inbred children. Science 1987; 236: 1567-70.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature 2001; 409: 860-921.

Lane SC, Jolly RD, Schmechel DE, Alroy J, Boustany RM. Apoptosis as the mechanism of neurodegeneration in batten's disease. J Neurochem 1996; 67: 677-83.

Le Borgne R, Griffiths G, Hoflack B. Mannose 6-phosphate receptors and ADP-ribosylation factors cooperate for high affinity interaction of the AP-1 golgi assembly proteins with membranes. J Biol Chem 1996; 271: 2162-70.

Lebrun AH, Storch S, Ruschendorf F, Schmiedt ML, Kyttälä A, Mole SE, et al. Retention of lysosomal protein CLN5 in the endoplasmic reticulum causes neuronal ceroid lipofuscinosis in asian sibship. Hum Mutat 2009; 30: E651-61.

Lehesjoki AE, Koskiniemi M, Norio R, Tirrito S, Sistonen P, Lander E, et al. Localization of the EPM1 gene for progressive myoclonus epilepsy on chromosome 21: Linkage disequilibrium allows high resolution mapping. Hum Mol Genet 1993; 2: 1229-34.

Lehtovirta M, Kyttälä A, Eskelinen EL, Hess M, Heinonen O, Jalanko A. Palmitoyl protein thioesterase (PPT) localizes into synaptosomes and synaptic vesicles in neurons: Implications for infantile neuronal ceroid lipofuscinosis (INCL). Hum Mol Genet 2001; 10: 69-75.

Lieschke GJ, Currie PD. Animal models of human disease: Zebrafish swim into view. Nat Rev Genet 2007; 8: 353-67.

Little L, Alcouloumre M, Drotar AM, Herman S, Robertson R, Yeh RY, et al. Properties of N-acetylglucosamine 1-phosphotransferase from human lymphoblasts. Biochem J 1987; 248: 151-9.

Lonka L, Kyttälä A, Ranta S, Jalanko A, Lehesjoki AE. The neuronal ceroid lipofuscinosis CLN8 membrane protein is a resident of the endoplasmic reticulum. Hum Mol Genet 2000; 9: 1691-7.

Lonka L, Salonen T, Siintola E, Kopra O, Lehesjoki AE, Jalanko A. Localization of wild-type and mutant neuronal ceroid lipofuscinosis CLN8 proteins in non-neuronal and neuronal cells. J Neurosci Res 2004; 76: 862-71.

Lu JY, Verkruyse LA, Hofmann SL. Lipid thioesters derived from acylated proteins accumulate in infantile neuronal ceroid lipofuscinosis: Correction of the defect in lymphoblasts by recombinant palmitoyl-protein thioesterase. Proc Natl Acad Sci U S A 1996; 93: 10046-50.

Luiro K, Yliannala K, Ahtiainen L, Maunu H, Järvelä I, Kyttälä A, et al. Interconnections of CLN3, Hook1 and rab proteins link batten disease to defects in the endocytic pathway. Hum Mol Genet 2004; 13: 3017-27.

Luiro K, Kopra O, Blom T, Gentile M, Mitchison HM, Hovatta I, et al. Batten disease (JNCL) is linked to disturbances in mitochondrial, cytoskeletal, and synaptic compartments. J Neurosci Res 2006; 84: 1124-38.

Luzio JP, Pryor PR, Gray SR, Gratian MJ, Piper RC, Bright NA. Membrane traffic to and from lysosomes. Biochem Soc Symp 2005; (72): 77-86.

Lyly A, von Schantz C, Heine C, Schmiedt ML, Sipila T, Jalanko A, et al. Novel interactions of CLN5 support molecular networking between neuronal ceroid lipofuscinosis proteins. BMC Cell Biol 2009; 10: 83.

Lyly A, von Schantz C, Salonen T, Kopra O, Saarela J, Jauhiainen M, et al. Glycosylation, transport, and complex formation of palmitoyl protein thioesterase 1 (PPT1)--distinct characteristics in neurons. BMC Cell Biol 2007; 8: 22.

Mao Q, Xia H, Davidson BL. Intracellular trafficking of CLN3, the protein underlying the childhood neurodegenerative disease, batten disease. FEBS Lett 2003; 555: 351-7.

Maquat LE. Nonsense-mediated mRNA decay: Splicing, translation and mRNP dynamics. Nat Rev Mol Cell Biol 2004; 5: 89-99.

Marth JD. Complexity in O-linked oligosaccharide biosynthesis engendered by multiple polypeptide N-acetylgalactosaminyltransferases. Glycobiology 1996; 6: 701-5.

Martoglio B, Dobberstein B. Signal sequences: More than just greasy peptides. Trends Cell Biol 1998; 8: 410-5.

Mefford HC, Eichler EE. Duplication hotspots, rare genomic disorders, and common disease. Curr Opin Genet Dev 2009; 19: 196-204.

Mellman I, Fuchs R, Helenius A. Acidification of the endocytic and exocytic pathways. Annu Rev Biochem 1986; 55: 663-700.

Melville SA, Wilson CL, Chiang CS, Studdert VP, Lingaas F, Wilton AN. A mutation in canine CLN5 causes neuronal ceroid lipofuscinosis in border collie dogs. Genomics 2005; 86: 287-94.

Metcalf P, Fusek M. Two crystal structures for cathepsin D: The lysosomal targeting signal and active site. EMBO J 1993; 12: 1293-302.

Meyer C, Zizioli D, Lausmann S, Eskelinen EL, Hamann J, Saftig P, et al. mu1A-adaptin-deficient mice: Lethality, loss of AP-1 binding and rerouting of mannose 6-phosphate receptors. EMBO J 2000; 19: 2193-203.

Minassian BA, Lee JR, Herbrick JA, Huizenga J, Soder S, Mungall AJ, et al. Mutations in a gene encoding a novel protein tyrosine phosphatase cause progressive myoclonus epilepsy. Nat Genet 1998; 20: 171-4.

Mitchell WA, Wheeler RB, Sharp JD, Bate SL, Gardiner RM, Ranta US, et al. Turkish variant late infantile neuronal ceroid lipofuscinosis (CLN7) may be allelic to CLN8. Eur J Paediatr Neurol 2001; 5 Suppl A: 21-7.

Mitchison HM, Bernard DJ, Greene ND, Cooper JD, Junaid MA, Pullarkat RK, et al. Targeted disruption of the Cln3 gene provides a mouse model for batten disease. the batten mouse model consortium [corrected. Neurobiol Dis 1999; 6: 321-34.

Mitchison HM, Hofmann SL, Becerra CH, Munroe PB, Lake BD, Crow YJ, et al. Mutations in the palmitoyl-protein thioesterase gene (PPT; CLN1) causing juvenile neuronal ceroid lipofuscinosis with granular osmiophilic deposits. Hum Mol Genet 1998; 7: 291-7.

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

Mole SE, Michaux G, Codlin S, Wheeler RB, Sharp JD, Cutler DF. CLN6, which is associated with a lysosomal storage disease, is an endoplasmic reticulum protein. Exp Cell Res 2004; 298: 399-406.

Mole SE, Zhong NA, Sarpong A, Logan WP, Hofmann S, Yi W, et al. New mutations in the neuronal ceroid lipofuscinosis genes. Eur J Paediatr Neurol 2001; 5 Suppl A: 7-10.

Moore SJ, Buckley DJ, MacMillan A, Marshall HD, Steele L, Ray PN, et al. The clinical and genetic epidemiology of neuronal ceroid lipofuscinosis in Newfoundland. Clin Genet 2008; 74: 213-22.

Morgan NV, Westaway SK, Morton JE, Gregory A, Gissen P, Sonek S, et al. PLA2G6, encoding a phospholipase A2, is mutated in neurodegenerative disorders with high brain iron. Nat Genet 2006; 38: 752-4.

Morgan TH. Chromosomes and associative inheritance. Science 1911; 34: 636-8.

Morton NE. LODs past and present. Genetics 1995; 140: 7-12.

Munroe PB, Greene ND, Leung KY, Mole SE, Gardiner RM, Mitchison HM, et al. Sharing of PPT mutations between distinct clinical forms of neuronal ceroid lipofuscinoses in patients from scotland. J Med Genet 1998; 35: 790.

Nachman MW, Crowell SL. Estimate of the mutation rate per nucleotide in humans. Genetics 2000; 156: 297-304.

Nakai K, Sakamoto H. Construction of a novel database containing aberrant splicing mutations of mammalian genes. Gene 1994; 141: 171-7.

Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, Gildersleeve HI, et al. Exome sequencing identifies MLL2 mutations as a cause of kabuki syndrome. Nat Genet 2010; 42: 790-3.

Nirenberg MW, Matthaei JH. The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides. Proc Natl Acad Sci U S A 1961; 47: 1588-602.

Norio R, Koskiniemi M. Progressive myoclonus epilepsy: Genetic and nosological aspects with special reference to 107 finnish patients. Clin Genet 1979; 15: 382-98.

Norman RM, Wood N. A congenital form of amaurotic family idiocy. J Neurol Psychiatr 1941; 4: 175-190.

Noskova L, Stranecky V, Hartmannova H, Pristoupilova A, Baresova V, Ivanek R, et al. Mutations in DNAJC5, encoding cysteine-string protein alpha, cause autosomal-dominant adult-onset neuronal ceroid lipofuscinosis. Am J Hum Genet 2011; 89: 241-52.

Okada N. SINEs: Short interspersed repeated elements of the eukaryotic genome. Trends Ecol Evol 1991; 6: 358-61.

Paisan-Ruiz C, Li A, Schneider SA, Holton JL, Johnson R, Kidd D, et al. Widespread lewy body and tau accumulation in childhood and adult onset dystonia-parkinsonism cases with PLA2G6 mutations. Neurobiol Aging 2010.

Palmer DN, Martinus RD, Cooper SM, Midwinter GG, Reid JC, Jolly RD. Ovine ceroid lipofuscinosis. the major lipopigment protein and the lipid-binding subunit of mitochondrial ATP synthase have the same NH2-terminal sequence. J Biol Chem 1989; 264: 5736-40.

Palmer DN, Fearnley IM, Walker JE, Hall NA, Lake BD, Wolfe LS, et al. Mitochondrial ATP synthase subunit c storage in the ceroid-lipofuscinoses (batten disease). Am J Med Genet 1992; 42: 561-7.

Pearce DA, Sherman F. Differential ubiquitin-dependent degradation of the yeast apo-cytochrome c isozymes. J Biol Chem 1997; 272: 31829-36.

Pearse BM. Receptors compete for adaptors found in plasma membrane coated pits. EMBO J 1988; 7: 3331-6.

Pena JA, Cardozo JJ, Montiel CM, Molina OM, Boustany R. Serial MRI findings in the costa rican variant of neuronal ceroid-lipofuscinosis. Pediatr Neurol 2001; 25: 78-80.

Pennacchio LA, Lehesjoki AE, Stone NE, Willour VL, Virtaneva K, Miao J, et al. Mutations in the gene encoding cystatin B in progressive myoclonus epilepsy (EPM1). Science 1996; 271: 1731-4.

Phillips HA, Scheffer IE, Berkovic SF, Hollway GE, Sutherland GR, Mulley JC. Localization of a gene for autosomal dominant nocturnal frontal lobe epilepsy to chromosome 20q 13.2. Nat Genet 1995; 10: 117-118.

Phillips SN, Benedict JW, Weimer JM, Pearce DA. CLN3, the protein associated with batten disease: Structure, function and localization. J Neurosci Res 2005; 79: 573-83.

Poet M, Kornak U, Schweizer M, Zdebik AA, Scheel O, Hoelter S, et al. Lysosomal storage disease upon disruption of the neuronal chloride transport protein ClC-6. Proc Natl Acad Sci U S A 2006; 103: 13854-9.

Porter MY, Turmaine M, Mole SE. Identification and characterization of caenorhabditis elegans palmitoyl protein thioesterase1. J Neurosci Res 2005; 79: 836-48.

Puertollano R, Aguilar RC, Gorshkova I, Crouch RJ, Bonifacino JS. Sorting of mannose 6-phosphate receptors mediated by the GGAs. Science 2001; 292: 1712-6.

Ramachandran N, Girard JM, Turnbull J, Minassian BA. The autosomal recessively inherited progressive myoclonus epilepsies and their genes. Epilepsia 2009; 50 Suppl 5: 29-36.

Ranta S, Topcu M, Tegelberg S, Tan H, Ustubutun A, Saatci I, et al. Variant late infantile neuronal ceroid lipofuscinosis in a subset of turkish patients is allelic to northern epilepsy. Hum Mutat 2004; 23: 300-5.

Ranta S, Zhang Y, Ross B, Lonka L, Takkunen E, Messer A, et al. The neuronal ceroid lipofuscinoses in human EPMR and mnd mutant mice are associated with mutations in CLN8. Nat Genet 1999; 23: 233-6.

Rawlings ND, Barrett AJ. Evolutionary families of metallopeptidases. Methods Enzymol 1995; 248: 183-228.

Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, et al. Global variation in copy number in the human genome. Nature 2006; 444: 444-54.

Rehman AU, Morell RJ, Belyantseva IA, Khan SY, Boger ET, Shahzad M, et al. Targeted capture and next-generation sequencing identifies C9orf75, encoding taperin, as the mutated gene in nonsyndromic deafness DFNB79. Am J Hum Genet 2010; 86: 378-88.

Reitman ML, Kornfeld S. Lysosomal enzyme targeting. N-acetylglucosaminylphosphotransferase selectively phosphorylates native lysosomal enzymes. J Biol Chem 1981; 256: 11977-80.

Reusch U, Bernhard O, Koszinowski U, Schu P. AP-1A and AP-3A lysosomal sorting functions. Traffic 2002; 3: 752-61.

Rider JA, Rider DL. Batten disease: Past, present, and future. Am J Med Genet Suppl 1988; 5: 21-6.

Robinson MS, Bonifacino JS. Adaptor-related proteins. Curr Opin Cell Biol 2001; 13: 444-453.

Saftig P, Hetman M, Schmahl W, Weber K, Heine L, Mossmann H, et al. Mice deficient for the lysosomal proteinase cathepsin D exhibit progressive atrophy of the intestinal mucosa and profound destruction of lymphoid cells. EMBO J 1995; 14: 3599-608.

Sandbank U. Congenital amaurotic idiocy. Pathol Eur 1968; 3: 226-9.

Sanger F, Air GM, Barrell BG, Brown NL, Coulson AR, Fiddes CA, et al. Nucleotide sequence of bacteriophage phi X174 DNA. Nature 1977; 265: 687-95.

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

Santavuori P, Rapola J, Sainio K, Raitta C. A variant of jansky-bielschowsky disease. Neuropediatrics 1982; 13: 135-41.

Santavuori P, Haltia M, Rapola J, Raitta C. Infantile type of so-called neuronal ceroid-lipofuscinosis. 1. A clinical study of 15 patients. J Neurol Sci 1973; 18: 257-67.

Santavuori P, Lauronen L, Kirveskari E, Aberg L, Sainio K, Autti T. Neuronal ceroid lipofuscinoses in childhood. Neurol Sci 2000; 21: S35-41.

Santavuori P, Rapola J, Raininko R, Autti T, Lappi M, Nuutila A, et al. Early juvenile neuronal ceroid-lipofuscinosis or variant jansky-bielschowsky disease: Diagnostic criteria and nomenclature. J Inherit Metab Dis 1993; 16: 230-2.

Santavuori P, Rapola J, Nuutila A, Raininko R, Lappi M, Launes J, et al. The spectrum of jansky-bielschowsky disease. Neuropediatrics 1991; 22: 92-6.

Saraste J, Kuismanen E. Pathways of protein sorting and membrane traffic between the rough endoplasmic reticulum and the golgi complex. Semin Cell Biol 1992; 3: 343-55.

Sasaki H, Kuzuhara S, Kanazawa I, Nakanishi T, Ogata T. Myoclonus, cerebellar disorder, neuropathy, mitochondrial myopathy, and ACTH deficiency. Neurology 1983; 33: 1288-93.

Savukoski M, Klockars T, Holmberg V, Santavuori P, Lander ES, Peltonen L. CLN5, a novel gene encoding a putative transmembrane protein mutated in finnish variant late infantile neuronal ceroid lipofuscinosis. Nat Genet 1998; 19: 286-8.

Scarborough PE, Dunn BM. Redesign of the substrate specificity of human cathepsin D: The dominant role of position 287 in the S2 subsite. Protein Eng 1994; 7: 495-502.

Schmiedt ML, Bessa C, Heine C, Gil Ribeiro M, Jalanko A, Kyttälä A. The neuronal ceroid lipofuscinosis protein CLN5: New insights into cellular maturation, transport and consequences of mutations. Hum Mutat 2010.

Schriner JE, Yi W, Hofmann SL. cDNA and genomic cloning of human palmitoyl-protein thioesterase (PPT), the enzyme defective in infantile neuronal ceroid lipofuscinosis. Genomics 1996; 34: 317-22.

Schulz A, Dhar S, Rylova S, Dbaibo G, Alroy J, Hagel C, et al. Impaired cell adhesion and apoptosis in a novel CLN9 batten disease variant. Ann Neurol 2004; 56: 342-50.

Schulz A, Mousallem T, Venkataramani M, Persaud-Sawin DA, Zucker A, Luberto C, et al. The CLN9 protein, a regulator of dihydroceramide synthase. J Biol Chem 2006; 281: 2784-94.

Schubert C. The genomic basis of the Williams-Beuren syndrome. Cell Mol Life Sci 2009; 66: 1178-1197.

Schwenk J, Metz M, Zolles G, Turecek R, Fritzius T, Bildl W, et al. Native GABA(B) receptors are heteromultimers with a family of auxiliary subunits. Nature 2010; 465: 231-5.

Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T, Yamrom B, et al. Strong Association of De Novo Copy Number Mutations with Autism. Science 2007; 316: 445-449.

Sharp JD, Wheeler RB, Parker KA, Gardiner RM, Williams RE, Mole SE. Spectrum of CLN6 mutations in variant late infantile neuronal ceroid lipofuscinosis. Hum Mutat 2003; 22: 35-42.

Siintola E, Topcu M, Kohlschutter A, Salonen T, Joensuu T, Anttonen AK, et al. Two novel CLN6 mutations in variant late-infantile neuronal ceroid lipofuscinosis patients of turkish origin. Clin Genet 2005; 68: 167-73.

Siintola E, Partanen S, Stromme P, Haapanen A, Haltia M, Maehlen J, et al. Cathepsin D deficiency underlies congenital human neuronal ceroid-lipofuscinosis. Brain 2006; 129: 1438-45.

Siintola E, Topcu M, Aula N, Lohi H, Minassian BA, Paterson AD, et al. The novel neuronal ceroid lipofuscinosis gene MFSD8 encodes a putative lysosomal transporter. Am J Hum Genet 2007; 81: 136-46.

Simonati A, Tessa A, Bernardina BD, Biancheri R, Veneselli E, Tozzi G, et al. Variant late infantile neuronal ceroid lipofuscinosis because of CLN1 mutations. Pediatr Neurol 2009; 40: 271-6.

Sleat DE, Sohar I, Gin RM, Lobel P. Aminoglycoside-mediated suppression of nonsense mutations in late infantile neuronal ceroid lipofuscinosis. Eur J Paediatr Neurol 2001; 5 Suppl A: 57-62.

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

Sleat DE, Ding L, Wang S, Zhao C, Wang Y, Xin W, et al. Mass spectrometry-based protein profiling to determine the cause of lysosomal storage diseases of unknown etiology. Mol Cell Proteomics 2009; 8: 1708-18.

Sleat DE, Gin RM, Sohar I, Wisniewski K, Sklower-Brooks S, Pullarkat RK, et al. Mutational analysis of the defective protease in classic late-infantile neuronal ceroid lipofuscinosis, a neurodegenerative lysosomal storage disorder. Am J Hum Genet 1999; 64: 1511-23.

Sleat DE, Wiseman JA, El-Banna M, Kim KH, Mao Q, Price S, et al. A mouse model of classical late-infantile neuronal ceroid lipofuscinosis based on targeted disruption of the CLN2 gene results in a loss of tripeptidyl-peptidase I activity and progressive neurodegeneration. J Neurosci 2004; 24: 9117-26.

Steenhuis P, Herder S, Gelis S, Braulke T, Storch S. Lysosomal targeting of the CLN7 membrane glycoprotein and transport via the plasma membrane require a dileucine motif. Traffic 2010; 11: 987-1000.

Stefansson H, Rujescu D, Cichon S, Pietiläinen OPH, Ingason A, Steinberg S, et al. Large recurrent microdeletions associated with schizophrenia. 2008; 455: 232-236.

Steinfeld R, Steinke HB, Isbrandt D, Kohlschutter A, Gartner J. Mutations in classical late infantile neuronal ceroid lipofuscinosis disrupt transport of tripeptidyl-peptidase I to lysosomes. Hum Mol Genet 2004; 13: 2483-91.

Steinfeld R, Reinhardt K, Schreiber K, Hillebrand M, Kraetzner R, Bruck W, et al. Cathepsin D deficiency is associated with a human neurodegenerative disorder. Am J Hum Genet 2006; 78: 988-98.

Steinlein OK, Mulley JC, Propping P, Wallace RH, Phillips HA, Sutherland GR, et al. A missense mutation in the neuronal nicotinic acetylcholine receptor alpha 4 subunit is associated with autosomal dominant nocturnal frontal lobe epilepsy. Nat Genet 1995; 11: 201-203.

Stogmann E, El Tawil S, Wagenstaller J, Gaber A, Edris S, Abdelhady A, et al. A novel mutation in the MFSD8 gene in late infantile neuronal ceroid lipofuscinosis. Neurogenetics 2009; 10: 73-7.

Storch S, Pohl S, Quitsch A, Falley K, Braulke T. C-terminal prenylation of the CLN3 membrane glycoprotein is required for efficient endosomal sorting to lysosomes. Traffic 2007; 8: 431-44.

Tammen I, Houweling PJ, Frugier T, Mitchell NL, Kay GW, Cavanagh JA, et al. A missense mutation (c.184C>T) in ovine CLN6 causes neuronal ceroid lipofuscinosis in merino sheep whereas affected south hampshire sheep have reduced levels of CLN6 mRNA. Biochim Biophys Acta 2006; 1762: 898-905.

Tao H, Manak JR, Sowers L, Mei X, Kiyonari H, Abe T, et al. Mutations in prickle orthologs cause seizures in flies, mice, and humans. Am J Hum Genet 2011; 88: 138-49.

Teixeira C, Guimaraes A, Bessa C, Ferreira MJ, Lopes L, Pinto E, et al. Clinicopathological and molecular characterization of neuronal ceroid lipofuscinosis in the portuguese population. J Neurol 2003; 250: 661-7.

Teixeira CA, Lin S, Mangas M, Quinta R, Bessa CJ, Ferreira C, et al. Gene expression profiling in vLINCL CLN6-deficient fibroblasts: Insights into pathobiology. Biochim Biophys Acta 2006; 1762: 637-46.

The international batten disease consortium.Isolation of a novel gene underlying batten disease, CLN3. Cell 1995; 82: 949-57.

Thomas CA,Jr. The genetic organization of chromosomes. Annu Rev Genet 1971; 5: 237-56.

Thyberg J, Moskalewski S. Role of microtubules in the organization of the golgi complex. Exp Cell Res 1999; 246: 263-79.

Tjio JH, Levan. The chromosome number of man. Hereditas 1956; 42: 1-6.

Topcu M, Tan H, Yalnizoglu D, Usubutun A, Saatci I, Aynaci M, et al. Evaluation of 36 patients from turkey with neuronal ceroid lipofuscinosis: Clinical, neurophysiological, neuroradiological and histopathologic studies. Turk J Pediatr 2004; 46: 1-10.

Tyynela J, Suopanki J, Baumann M, Haltia M. Sphingolipid activator proteins (SAPs) in neuronal ceroid lipofuscinoses (NCL). Neuropediatrics 1997; 28: 49-52.

Tyynela J, Baumann M, Henseler M, Sandhoff K, Haltia M. Sphingolipid activator proteins in the neuronal ceroid-lipofuscinoses: An immunological study. Acta Neuropathol 1995; 89: 391-8.

Tyynela J, Sohar I, Sleat DE, Gin RM, Donnelly RJ, Baumann M, et al. A mutation in the ovine cathepsin D gene causes a congenital lysosomal storage disease with profound neurodegeneration. EMBO J 2000; 19: 2786-92.

Uusi-Rauva K, Luiro K, Tanhuanpaa K, Kopra O, Martin-Vasallo P, Kyttälä A, et al. Novel interactions of CLN3 protein link batten disease to dysregulation of fodrin-na+, K+ ATPase complex. Exp Cell Res 2008; 314: 2895-905.

Vahava O, Morell R, Lynch ED, Weiss S, Kagan ME, Ahituv N, et al. Mutation in transcription factor POU4F3 associated with inherited progressive hearing loss in humans. Science 1998; 279: 1950-4.

Van Bogaert P, Azizieh R, Desir J, Aeby A, De Meirleir L, Laes JF, et al. Mutation of a potassium channel-related gene in progressive myoclonic epilepsy. Ann Neurol 2007; 61: 579-86.

van Vliet C, Thomas EC, Merino-Trigo A, Teasdale RD, Gleeson PA. Intracellular sorting and transport of proteins. Prog Biophys Mol Biol 2003; 83: 1-45.

Van Heycoptenhamm, DE Jager H. Progressive myoclonus epilepsy with lafora bodies. clinical-pathological features. Epilepsia 1963; 4: 95-119.

Vantaggiato C, Redaelli F, Falcone S, Perrotta C, Tonelli A, Bondioni S, et al. A novel CLN8 mutation in late-infantile-onset neuronal ceroid lipofuscinosis (LINCL) reveals aspects of CLN8 neurobiological function. Hum Mutat 2009; 30: 1104-16.

Varon R, Gooding R, Steglich C, Marns L, Tang H, Angelicheva D, et al. Partial deficiency of the C-terminal-domain phosphatase of RNA polymerase II is associated with congenital cataracts facial dysmorphism neuropathy syndrome. Nat Genet 2003; 35: 185-9.

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. Science 2001; 291: 1304-51.

Verkruyse LA, Hofmann SL. Lysosomal targeting of palmitoyl-protein thioesterase. J Biol Chem 1996; 271: 15831-6.

Vesa J, Chin MH, Oelgeschlager K, Isosomppi J, DellAngelica EC, Jalanko A, et al. Neuronal ceroid lipofuscinoses are connected at molecular level: Interaction of CLN5 protein with CLN2 and CLN3. Mol Biol Cell 2002; 13: 2410-20.

Vesa J, Hellsten E, Verkruyse LA, Camp LA, Rapola J, Santavuori P, et al. Mutations in the palmitoyl protein thioesterase gene causing infantile neuronal ceroid lipofuscinosis. Nature 1995; 376: 584-7.

Vines D, Warburton MJ. Purification and characterisation of a tripeptidyl aminopeptidase I from rat spleen. Biochim Biophys Acta 1998; 1384: 233-42.

von Schantz C, Saharinen J, Kopra O, Cooper JD, Gentile M, Hovatta I, et al. Brain gene expression profiles of Cln1 and Cln5 deficient mice unravels common molecular pathways underlying neuronal degeneration in NCL diseases. BMC Genomics 2008; 9: 146.

Walter P, Johnson AE. Signal sequence recognition and protein targeting to the endoplasmic reticulum membrane. Annu Rev Cell Biol 1994; 10: 87-119.

Warburton MJ, Bernardini F. Tripeptidyl-peptidase I deficiency in classical late-infantile neuronal ceroid lipofuscinosis brain tissue. evidence for defective peptidase rather than proteinase activity. J Inherit Metab Dis 2000; 23: 145-54.

Warren G, Mellman I. Bulk flow redux? Cell 1999; 98: 125-7.

Watson JD, Crick FH. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature 1953; 171: 737-8.

Wheeler RB, Sharp JD, Mitchell WA, Bate SL, Williams RE, Lake BD, et al. A new locus for variant late infantile neuroal ceroid lipofuscinosis-CLN7. Mol Genet Metab 1999; 66: 337-8.

Wheeler RB, Sharp JD, Schultz RA, Joslin JM, Williams RE, Mole SE. 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 2002; 70: 537-42.

Williams RE, Gottlob I, Lake BD, Goebel HH, Winchester BG, Wheeler RB. CLN2 classical late infantile NCL. In: Anonymous *The Neuronal Ceroid Lipofuscinosis (Batten Disease)*. Amsterdam: IOS Press; 1999. p. 37-54.

Winter E, Ponting CP. TRAM, LAG1 and CLN8: Members of a novel family of lipid-sensing domains? Trends Biochem Sci 2002; 27: 381-3.

Wong FL, Cantor RM, Rotter JI. Sample-size considerations and strategies for linkage analysis in autosomal recessive disorders. Am J Hum Genet 1986; 39: 25-37.

Woods CG, Cox J, Springell K, Hampshire DJ, Mohamed MD, McKibbin M, et al. Quantification of homozygosity in consanguineous individuals with autosomal recessive disease. Am J Hum Genet 2006; 78: 889-96.

Xiang M, Gan L, Li D, Chen ZY, Zhou L, O'Malley BW,Jr, et al. Essential role of POU-domain factor brn-3c in auditory and vestibular hair cell development. Proc Natl Acad Sci U S A 1997; 94: 9445-50.

Zaidi N, Maurer A, Nieke S, Kalbacher H. Cathepsin D: A cellular roadmap. Biochem Biophys Res Commun 2008; 376: 5-9.

Zeman W. Batten disease: Ocular features, differential diagnosis and diagnosis by enzyme analysis. Birth Defects Orig Artic Ser 1976; 12: 441-53.

Zhong N, Wisniewski KE, Hartikainen J, Ju W, Moroziewicz DN, McLendon L, et al. Two common mutations in the CLN2 gene underlie late infantile neuronal ceroid lipofuscinosis. Clin Genet 1998; 54: 234-8.

Zhong N, Moroziewicz DN, Ju W, Jurkiewicz A, Johnston L, Wisniewski KE, et al. Heterogeneity of late-infantile neuronal ceroid lipofuscinosis. Genet Med 2000; 2: 312-8.

Zupanc ML, Legros B. Progressive myoclonic epilepsy. Cerebellum 2004; 3: 156-71.