Helsinki University Biomedical Dissertations No. 107

IDENTIFICATION OF TWO NOVEL HUMAN NEURONAL CEROID LIPOFUSCINOSIS GENES

Eija Siintola

Folkhälsan Institute of Genetics

Department of Medical Genetics
and
Neuroscience Center
University of Helsinki

ACADEMIC DISSERTATION

To be publicly discussed, with the permission of the Medical Faculty of the University of Helsinki, in the Folkhälsan lecture hall Arenan, Topeliuksenkatu 20, Helsinki, on May 16th 2008 at 12 noon.

Helsinki 2008

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 Molecular Medicine
National Public Health Institute
Helsinki, Finland

Official opponent

Professor Kari Majamaa, M.D., Ph.D.
Institute of Clinical Medicine
University of Turku
Turku, Finland

ISSN 1457-8433
ISBN 978-952-10-4670-4 (paperback)
ISBN 978-952-10-4671-1 (PDF)
http://ethesis.helsinki.fi
Helsinki University Print
Helsinki 2008

To my family

TABLE OF CONTENTS

Table of contents

LIST OF ORIGINAL PUBLICATIONS

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 In addition, some unpublished results are presented.

I Siintola E*, Partanen S*, Strömme P, Haapanen A, Haltia M, Maehlen J, Lehesjoki A-E, Tyynelä J (2006) Cathepsin D deficiency underlies congenital human neuronal ceroid-lipofuscinosis. *Brain* 129:1438-45.

II Siintola E, Topcu M, Kohlschütter A, Salonen T, Joensuu T, Anttonen A-K, Lehesjoki A-E (2005) Two novel *CLN6* mutations in variant late-infantile neuronal ceroid lipofuscinosis patients of Turkish origin. *Clin Genet* 68:167-73.

III Siintola E, Topcu M, Aula N, Lohi H, Minassian BA, Paterson AD, Liu X-Q, Wilson C, Lahtinen U, Anttonen A-K, Lehesjoki A-E (2007) The novel neuronal ceroid lipofuscinosis gene *MFSD8* encodes a putative lysosomal transporter. *Am J Hum Genet* 81:136-46.

*These two authors contributed equally to this work.

Publication I also appears in the thesis of Sanna Partanen (2006).

ABBREVIATIONS

ABC	ATP-binding cassette
ANCL	adult neuronal ceroid lipofuscinosis
AP	adaptor protein
ATP	adenosine triphosphate
bp	base pair(s)
BHK cells	baby hamster kidney cells
BLAST	basic local alignment search tool
c.	coding DNA reference sequence position
cathD	*Drosophila melanogaster* cathepsin D gene
cDNA	complementary DNA
CEPH	Centre d'Etude du Polymorphisme Humain
CLCN3,6,7/CLCN3,6,7	human chloride channel 3, 6, and 7 gene or locus/protein
Clcn3,6,7/CLCN3,6,7	mouse chloride channel 3, 6, and 7 gene or locus/protein
CLN1-10/CLN1-10	human, canine, or bovine CLN1-10 gene or locus/protein or disease
Cln1-10/CLN1-10	mouse CLN1-10 gene or locus/protein
cM	centiMorgan(s)
COS-1 cells	African green monkey kidney cells
CTSB,D,F,L/CTSB,D,F,L	human, sheep, and canine cathepsin B, D, F, and L gene or locus/protein
Ctsb,d,f,l/CTSB,D,F,L	mouse cathepsin B, D, F, and L gene or locus/protein
del	deletion
DNA	deoxyribonucleic acid
dup	duplication
EDTA	ethylenediamine tetra acetic acid
EPMR	progressive epilepsy with mental retardation
ER	endoplasmic reticulum
ERGIC	ER–Golgi intermediate compartment
EST	expressed sequence tag
GROD	granular osmiophilic deposit
HA	hemagglutinin
HLOD	heterogeneity LOD
INCL	infantile neuronal ceroid lipofuscinosis
JNCL	juvenile neuronal ceroid lipofuscinosis
kb	kilobase(s)
kDa	kilodalton(s)
LAMP1,2/LAMP1,2	lysosomal-associated membrane protein 1 and 2 gene/protein
LD	linkage disequilibrium

LINCL	late-infantile neuronal ceroid lipofuscinosis
LOD	logarithm of odds
LSD	lysosomal storage disorder
M6P	mannose 6-phosphate
Mb	megabase(s)
MFS	major facilitator superfamily
MFSD8,9/MFSD8,9	major facilitator superfamily domain containing 8 and 9 gene or locus/protein
MPR	mannose 6-phosphate receptor
mRNA	messenger RNA
NCBI	National Center for Biotechnology Information
NCL	neuronal ceroid lipofuscinosis
NMD	nonsense-mediated mRNA decay
OMIM	Online Mendelian Inheritance in Man
p.	protein reference sequence position
PCR	polymerase chain reaction
PPT1,2/PPT1,2	human palmitoyl protein thioesterase 1 and 2 gene or locus/protein
Ppt1,2/PPT1,2	mouse palmitoyl protein thioesterase 1 and 2 gene or locus/protein
RNA	ribonucleic acid
RT-PCR	reverse transcriptase PCR
SLC	solute carrier
SNP	single nucleotide polymorphism
TGN	*trans*-Golgi network
θ	recombination fraction
TPP1/TPP1	human or canine tripeptidyl peptidase I gene or locus/protein
Tpp1/TPP1	mouse tripeptidyl peptidase I gene or locus/protein
vLINCL	variant late-infantile neuronal ceroid lipofuscinosis
VNTR	variable number tandem repeat

Only the abbreviations appearing more than once in the text are listed here.

ABSTRACT

The neuronal ceroid lipofuscinoses (NCLs) are a group of mostly autosomal recessively inherited neurodegenerative disorders. Of NCLs, congenital NCL is the earliest-onset and the most aggressive form whereas Turkish variant late-infantile NCL (vLINCL) belongs to the heterogeneous group of late-infantile onset NCLs. The aim of this thesis was to characterize the molecular genetic bases of these, previously genetically undetermined, NCL forms.

In order to define the molecular genetic background of congenital NCL, a candidate gene approach was undertaken. Previously, a mutation in the cathepsin D (*CTSD*) gene was shown to cause congenital NCL in sheep. Based on the close resemblance of the clinical phenotypes between sheep and human patients with congenital NCL, *CTSD* was considered as a potential candidate gene in humans as well. When screened for mutations by sequencing, a homozygous nucleotide duplication creating a premature stop codon was identified in *CTSD* in one family with congenital NCL. While *in vitro* the transiently overexpressed mutant protein was stable although truncated and inactive, the absence of CTSD staining in brain tissue samples of patients indicated degradation of the mutant CTSD *in vivo*. A lack of CTSD staining was detected also in another, unrelated family with congenital NCL but the presence of *CTSD* mutation(s) could not be confirmed. These results imply that CTSD deficiency underlies congenital NCL.

While initially Turkish vLINCL was considered a distinct genetic entity (CLN7), mutations in the *CLN8* gene were later reported to account for the disease in a subset of Turkish patients with vLINCL. To further dissect the genetic basis of Turkish vLINCL a candidate gene approach was first undertaken in 13 mainly consanguineous, Turkish vLINCL families. All known NCL loci were screened for homozygosity by haplotype analysis of microsatellite markers, and if homozygosity of marker alleles was detected the genes were sequenced from genomic DNA of the respective patients. Two novel, family-specific homozygous mutations were identified in the *CLN6* gene. In the remaining families, all known human NCL loci as well as the loci underlying NCL-like phenotypes in animal models were excluded.

To further characterize the genetic background of Turkish vLINCL and specifically, to identify novel gene(s) underlying vLINCL, a genomewide single nucleotide polymorphism scan, homozygosity mapping, and positional candidate gene sequencing were performed in ten mainly consanguineous, Turkish vLINCL families. On chromosome 4q28.1-q28.2, a novel major facilitator superfamily domain containing 8 (*MFSD8*) gene with six family-specific homozygous mutations in vLINCL patients was identified. By northern blot, *in silico*, and RT-PCR analyses, the *MFSD8*

transcript was shown to be ubiquitously expressed with a complex pattern of alternative splicing. MFSD8 is predicted to be a transmembrane protein with 12 membrane-spanning domains. It belongs to the major facilitator superfamily of transporter proteins. In immunofluorescence analysis, transiently overexpressed MFSD8 was shown to colocalize with lysosomal markers. These results suggest that MFSD8 is a novel lysosomal integral membrane transporter protein, the cellular function of which remains to be elucidated. Identification of *MFSD8* further emphasizes the genetic heterogeneity of Turkish vLINCL as well as the genetic heterogeneity of NCLs in general. In families where no *MFSD8* mutations were detected, additional NCL-causing genes remain to be identified.

The identification of mutations in *CTSD* and *MFSD8* increases the number of known human NCL-causing genes from six to eight, and is an important step towards the complete understanding of the genetic spectrum underlying NCLs. In addition, it is a starting point for dissecting the molecular mechanisms behind the associated NCLs and contributes to the challenging task of understanding the molecular pathology underlying the group of NCL disorders.

INTRODUCTION

The neuronal ceroid lipofuscinoses (NCLs) are mostly autosomal recessively inherited neurodegenerative disorders of which at least ten forms are thought to exist (CLN1-CLN9 and congenital NCL) (Haltia 2003). Prior to this thesis, six human NCL-causing genes (*PPT1*, *TPP1*, *CLN3*, *CLN5*, *CLN6*, and *CLN8*) had been identified (Mole 2004). The molecular genetic basis has, however, remained undetermined in a great number of patients with various NCL phenotypes. Identification of the genes responsible for these phenotypes is of critical importance for revealing the complete molecular genetic background of NCLs. This, in turn, is crucial for understanding the disease mechanisms and the molecular pathology underlying the group of NCL disorders.

The identification of the genes underlying the major NCL forms was in most cases based on relatively large family material suitable for traditional linkage analyses. Moreover, on many occasions an existing founder effect eased the process. The search for new genes underlying the rarer NCL forms may be more challenging due to the limited and genetically heterogeneous family material. Many of these families are, however, consanguineous, and thus, application of a homozygosity mapping approach (Lander and Botstein 1987) will increase the probability of the identification of novel disease genes.

The process of identification of disease genes has undergone tremendous change since the completion of the Human Genome Project in 2003 and the consequent availability of genome sequences (International Human Genome Sequencing Consortium 2004). This, accompanied by the development of new technologies for marker genotyping and data analysis, provides new tools for the identification of genes underlying genetically inherited diseases, including NCLs.

This thesis describes the characterization of the molecular genetic basis of two non-common NCL forms, congenital NCL and Turkish variant late-infantile NCL (vLINCL). Since in both diseases the available family material was small, approaches alternative to traditional linkage analysis were used for the identification of the disease genes. In congenital NCL, a candidate gene approach was applied since the disease-causing gene was known in the corresponding ovine disease (Tyynelä *et al.* 2000). In Turkish vLINCL, where the majority of the families were consanguineous, a central method leading to gene identification was homozygosity mapping.

REVIEW OF THE LITERATURE

1. Identification of disease genes

The process of identifying disease genes has changed dramatically during recent years due to the completion of the Human Genome Project and the advancement of technological platforms (Antonarakis and Beckmann 2006). The consequent availability of human (and other species) genome sequences as well as high-throughput assays for marker genotyping have enabled the bypassing or acceleration of some very laborious steps in the gene identification process. Although the number of newly identified genes underlying monogenic disorders has declined during the present compared with the previous decade, the study of Mendelian disorders will continue to be fundamental in elucidating gene functions as well as in understanding both normal and pathological pathways (Antonarakis and Beckmann 2006). Additionally, it will provide clues to unravel the susceptibility alleles for polygenic, complex phenotypes.

1.1. The Human Genome Project

The Human Genome Project, initiated in 1990, was an international, collaborative effort whose central goals were to determine the nucleotide sequence of the human genome and to identify all human genes. The International Human Genome Sequencing Consortium released a draft sequence in 2000 (Lander *et al.* 2001), simultaneously with a private company, Celera Genomics (Venter *et al.* 2001). The complete sequence, published in 2003, covered ~99% of the euchromatic genome but still contained some gaps (International Human Genome Sequencing Consortium 2004). In addition to the human genome, genome sequences of hundreds of other organisms including more than 70 eukaryotic species have been produced (www.ebi.ac.uk/genomes/index.html, April 2008). In 2007, the first complete diploid genome sequence of a single human individual was published (Levy *et al.* 2007).

The sequence information is freely available in online public databases and provides biomedical researchers with invaluable tools to explore various features of the genomes. Analysis of the sequence data for the extraction of its full information will continue. Of the estimated 20 000-25 000 human protein-coding genes, some are presumably still to be identified and many more to be annotated and characterized in order to understand their roles and functions in health and disease (International Human Genome Sequencing Consortium 2004). The regulatory elements and other non-coding parts of the genome need to be analyzed. Comparative analyses between

genomes of different species will identify functional elements of genomes, and enhance our comprehension of the evolution and diversity of species. Characterization of genetic sequence variations will be important in understanding their association with phenotypic differences and diseases (International Human Genome Sequencing Consortium 2004).

1.2. Approaches in disease gene identification

The choice of approach for disease gene identification depends on many things, such as the availability and properties of the family material, and the level of understanding of the biochemical basis of the disease. In many approaches, careful delineation of the disease phenotype and the collection of families with accurate diagnoses are of critical importance at the initial stage.

Disease gene identification based on genomic location, an approach referred to as positional cloning, has been widely used in the research of monogenic diseases (Collins 1992, Collins 1995). The process proceeds in successive steps of which the first is to define the chromosomal position of the disease gene by genomewide marker genotyping (section 1.3) and linkage analysis (section 1.4) in families in which the responsible gene is segregating. Next, the identified candidate region needs to be characterized. Before the availability of the genomic sequence data this was carried out by genetic and physical fine mapping that required a lot of laboratory work. Nowadays the characteristics of the candidate region can be examined from physical maps available in online databases. Haplotype analyses and linkage disequilibrium (LD) mapping (section 1.4) are utilized to decrease the length of the critical region as much as possible. Finally, all genes in the candidate region are identified (section 1.6) and screened for mutations (section 1.7) in order to find out which gene underlies the disease. After the first successful gene identification by this method in X-linked chronic granulomatous disease (Royer-Pokora *et al.* 1986), it has been employed in determining the genetic background of a number of diseases. The sequence information provided by the Human Genome Project has considerably eased and accelerated the process of positional cloning, especially of the characterization of candidate regions and the identification of candidate genes. The critical candidate regions may, however, be very large and contain up to hundreds of genes. An application of positional cloning, positional candidate gene cloning, aims at reducing the number of candidate genes in a candidate region by selecting the most likely candidates with the aid of functional information (section 1.6) (Collins 1995). This method has been successfully used in the cloning of many disease genes, for instance of the rhodopsin gene underlying retinitis pigmentosa (Dryja *et al.* 1990). However, deciding on which are the best candidate genes is sometimes difficult. Moreover, the genes ultimately identified sometimes turn out to encode protein

products without known function or with a function not obviously related to the respective disease phenotype. The selection of these as functional candidate genes would have been unlikely.

The identification of disease genes without information of their chromosomal position, functional cloning, has been used in certain instances when the disease has been biochemically well defined (Collins 1992). Prior to the availability of genome sequences this approach was hampered by the difficulty of identifying the disease gene even if the defective protein was known. Functional cloning was used, for instance, in the identification of the phenylalanine hydroxylase gene underlying phenylketonuria (Robson *et al.* 1982). Nowadays, if the protein(s) potentially involved in disease pathogenesis can be deduced, the genes encoding them (candidate genes, section 1.6) can usually be found directly from the online databases and tested for linkage and/or screened for mutations in a process designated as candidate gene cloning.

1.3. Polymorphic markers

Individual variations between human genomes are exploited in human gene mapping. The majority of the human sequence variation is attributable to single nucleotide polymorphisms (SNPs), whereas the rest is caused by insertions or deletions of nucleotides, repeat length polymorphisms, and rearrangements (Sachidanandam *et al.* 2001). These genetic variants can be used as genetic markers in linkage, LD, and haplotype analyses (section 1.4). In order to utilize these markers in genetic analyses there should be sufficient amounts of allelic variation between individuals to enable the analysis of adequate numbers of informative meioses within pedigrees.

The first method suitable for genomewide analysis of polymorphic deoxyribonucleic acid (DNA) markers was the restriction fragment length polymorphism (RFLP) method (Botstein *et al.* 1980). This method was able to detect SNPs (see below) as well as variable number tandem repeats (VNTRs), in particular minisatellites. After the development of the polymerase chain reaction (PCR) technique, microsatellites (or short tandem repeats, STRs), another class of VNTRs, have become widely used genetic markers. They are usually di-, tri-, or tetranucleotide repeats that are very polymorphic, showing high levels of allelic variation in the number of repeat units (Gray *et al.* 2000), and thus are very informative. They are widely distributed throughout the human genome, occurring approximately once per 2 kilobases (kb) of genomic DNA (Lander *et al.* 2001). The high variability of microsatellites is due to their high mutation rate (in humans 10^{-5}-10^{-3} nucleotides per cell division) explained mainly by slipped strand mispairing during

DNA replication (Fan and Chu 2007). Microsatellites can nowadays be analyzed by PCR-based methods at a relatively large scale (Dearlove 2002).

In addition, SNPs are at present commonly used in a broad spectrum of human genetic analyses. They are less polymorphic compared to microsatellites, with generally only two alleles, but are more stable due to lower mutation rates (Gray *et al.* 2000). The density of SNPs is very high, with an estimated more than 10 million SNPs in the human genome (Kruglyak and Nickerson 2001). The current (April 2008) Single Nucleotide Polymorphism database (dbSNP) build 128 in the National Center for Biotechnology Information (NCBI) database contains more than 4.9 million SNPs (www.ncbi.nlm.nih.gov/sites/entrez?db=snp). SNPs can be genotyped at very large scales using various assays including microarray-based methods involving allele-specific hybridization (Syvänen 2005), as in the method utilized in this thesis (Matsuzaki *et al.* 2004).

Along with the analysis of genomewide sequence data, it has become evident that large-scale copy-number variants (CNVs) account for a substantial proportion of variation between human genomes (Iafrate *et al.* 2004, Sebat *et al.* 2004, Redon *et al.* 2006). Obviously, they will have an impact on human genetic studies but their usefulness as polymorphic markers is still unclear.

1.4. Linkage, linkage disequilibrium, and haplotype analyses

Linkage analysis is used for the detection of the chromosomal location of a disease-causing gene. The idea is to find out if two loci cosegregate more often than they should if they were not physically close to each other on the same chromosome (Terwilliger and Ott 1994). The closer the one locus (e.g. a polymorphic marker) is to the other locus (e.g. a locus harbouring the disease gene), the more rarely are they separated by recombinations (Terwilliger and Ott 1994). The proportion of recombinations observed, a recombination fraction (θ), is used as a measure of genetic distance between two loci (Terwilliger and Ott 1994). For unlinked loci $\theta = 0.5$, whereas for linked loci $\theta < 0.5$, corresponding to observing recombinations in less than 50% of meioses (Terwilliger and Ott 1994). The genetic linkage is measured as a likelihood ratio of two hypotheses: two loci are linked at a certain θ compared to them being unlinked (Morton 1955). This likelihood is expressed as a logarithm of odds (LOD) score (Z). A LOD score higher than 3 (odds ratio 1000:1) is generally considered as significant evidence for linkage, whereas lower than -2 (odds ratio 1:100) is considered as proof of exclusion (Terwilliger and Ott 1994). The chromosomal localization of the disease gene can be inferred since the approximate distance between the two loci is equal to the value of θ at which the LOD score is highest (Terwilliger and Ott 1994). Locus heterogeneity can be taken into account in

calculating LOD scores with results expressed as heterogeneity LOD (HLOD) scores. In two-point linkage analysis the LOD scores are calculated individually at single marker loci, whereas in multipoint linkage analysis allelic data from several loci is combined in the calculations. Parametric linkage analysis requires information on the specific disease model, including information of mode of inheritance, penetrances, and allele frequencies. In nonparametric linkage analyses the disease model does not have to be known. Linkage analyses are conducted by computer programs, such as programs of the LINKAGE package (Lathrop and Lalouel 1984), GENEHUNTER (Kruglyak *et al.* 1996), and Merlin (Abecasis *et al.* 2002).

The candidate regions identified by linkage analyses are often broad and contain a large number of genes. LD and haplotype analyses are useful tools in fine mapping these regions, especially in isolated populations where the major founder mutations originate from single ancestors (de la Chapelle and Wright 1998, Peltonen *et al.* 1999). LD refers to the co-inheritance of particular marker alleles at loci close to each other with a frequency greater than expected from random segregation. A certain set of marker alleles, a haplotype, surrounding the disease-causing mutation is inherited along with the ancestral mutation. The extent of LD decreases in each generation, as the ancestral haplotype is disrupted by historical recombinations in successive generations, and consequently, the length of the disease-associated haplotype tends to decrease (de la Chapelle and Wright 1998, Peltonen *et al.* 2000a). In the case of locus and allelic homogeneity, comparison of patient haplotypes in order to identify a major shared haplotype restricted by individual historical recombinations can considerably narrow down the candidate region (Peltonen *et al.* 2000a). The degree of LD varies across the genome and is suggested to be structured into discrete sequence blocks separated by hotspots of recombination (Daly *et al.* 2001, Jeffreys *et al.* 2001). LD has been utilized in mapping of genes underlying several diseases, including many belonging to the Finnish Disease Heritage (Peltonen *et al.* 1999), such as diastrophic dysplasia (Hästbacka *et al.* 1992). LD and haplotype analyses are also exploited in homozygosity mapping (Lander and Botstein 1987) (section 1.5). In addition, LD is utilized in association analyses in dissecting the genetic components of complex traits (Morton 2005).

1.5. Homozygosity mapping

Homozygosity (or autozygosity) mapping is a powerful strategy for disease gene identification in consanguineous families suffering from recessively inherited diseases (Lander and Botstein 1987). In such an inbred child the marker alleles on the region surrounding the disease locus are almost always homozygous over several centiMorgans (cM) (Lander and Botstein 1987). An ancestral founder chromosome segment is passed from both parents to the affected child and is said to be

homozygous or identical by descent (IBD). A genomic region that is consistently homozygous by descent in all patients is most likely to harbour the disease-causing gene (Lander and Botstein 1987). Segments homozygous by descent are expected to be observed also on some other genomic regions in each child of a consanguineous family, with a higher frequency the closer the relationship of the parents, and, in theory, ~6% of the genome of a child of first cousins is expected to be homozygous by descent (Lander and Botstein 1987). In addition, the segments of homozygosity are longer the closer the relationship, and the average size of a homozygous segment is 20 cM in a child of first cousins (Woods *et al.* 2006). The extensive homozygosity, in turn, results in an increased incidence of autosomal recessive diseases (Woods *et al.* 2006). In populations with long traditions of consanguineous marriages, e.g. in the Middle East (Bittles 2001), prolonged parental inbreeding has been reported to result in increased overall levels of homozygosity (Woods *et al.* 2006). By homozygosity mapping disease gene loci can be successfully identified even in very small families without genotyping all intervening relatives (Carr *et al.* 2006). Homozygosity mapping can also be applied in mapping gene(s) for a disease with a heterogeneous genetic basis by a strategy of simultaneously searching for several loci, at least one of which is homozygous by descent in most of the patients (Lander and Botstein 1987). Homozygosity can be exploited also in fine mapping the candidate region by searching for a shared, overlapping region of homozygosity between the patients, and thus possibly decreasing the length of the critical region (Lander and Botstein 1987). Homozygosity mapping has been applied in the identification of many disease genes, one recent example being the tripartite motif-containing 32 (*TRIM32*) gene in Bardet-Biedl syndrome (Chiang *et al.* 2006). The homozygosity mapping approach also has some potential pitfalls. There may be unexpected genetic heterogeneity present within families resulting in loss of shared homozygosity of marker alleles flanking the disease locus (Miano *et al.* 2000). In addition, regions homozygous by descent unrelated to the disease locus may be identified by chance, and an underestimated extent of inbreeding may result in spuriously high LOD scores (Miano *et al.* 2000).

1.6. Candidate genes

Candidate gene identification is an important step towards disease gene identification, both in positional and non-positional cloning approaches. It is relatively straightforward today with the availability of human (and many other species) genome sequences and consequently, the availability of information on physical locations and sequences of many genes.

In positional cloning approaches, all known or putative genes in a candidate locus are identified from databases using genome browsers, such as NCBI (www.ncbi.nlm.nih.gov/), Ensembl (www.ensembl.org), or Santa Cruz

(genome.ucsc.edu/) browsers on the internet. Despite the existence of almost complete human genomic sequence information, all genes are not yet identified and catalogued, and thus these browsers should not be entirely relied upon. Additional genes may be revealed by gene prediction programs (Brent and Guigo 2004).

Functional candidate genes both in positional (in positional candidate gene cloning) and non-positional (in functional and candidate gene cloning) approaches are chosen based on prior knowledge of the pathology and biochemical basis of the disease. The candidates can be chosen based on the known or putative function or expression pattern of the encoded proteins or their homologs either in humans (paralogs) or in other species (orthologs). Good candidates may also be genes that are homologous or related to human or animal genes where mutations cause similar phenotypes. Additionally, candidate genes may encode interaction partners of a protein defective in a disease related to the one under study (Antonarakis and Beckmann 2006). Differences in gene expression between cells or tissues from affected and control individuals detected by genomewide expression profiling using microarray-based techniques may also suggest candidate genes (Antonarakis and Beckmann 2006).

1.7. Mutation analysis

Candidate genes have to be individually tested to determine whether or not they are responsible for the disease phenotype. To gain confirmative evidence for the pathogenic role of a particular gene, several criteria have to be fulfilled. Mutation screening (see below) in patients has to reveal one or preferably more sequence variants that segregate with the disease in the respective families according to the predicted mode of inheritance. In order to differentiate disease-causing sequence variants from neutral polymorphisms, an adequate number of control chromosomes representing the respective population have to be screened (Collins and Schwartz 2002). In autosomal recessive disorders it is anticipated that the disease-causing alteration is not seen in homozygous form in unaffected individuals. In addition, the nature of the sequence variants and their consequences are assessed (see below). The ultimate proof of the pathogenic role of a gene would be obtained by a functional test, such as restoration of the normal phenotype *in vitro*, or by generation of an animal model for the disease. Unfortunately these may be time consuming and not always immediately feasible, or may not give definite proof for the pathogenic role of the gene.

While a variety of methods for mutation screening have traditionally been employed, nowadays a commonly used method due to its reduced costs is direct sequencing. It can be performed following the PCR amplification of exons and exon-

intron boundaries of the candidate genes from the genomic DNA, or of exons from the complementary DNA (cDNA), of patients, carriers, and control individuals. However, even this method sometimes fails to detect mutations. Moreover, some mutations are difficult to find, especially in intronic or regulatory regions, or in the case of large chromosomal rearrangements. Other methods used relatively frequently at present include, for example, denaturing high-performance liquid chromatography (DHPLC) (Xiao and Oefner 2001) and multiplex ligation-dependent probe amplification (MLPA) (Schouten *et al.* 2002).

Various types of sequence alterations can be identified, and the assessment of their pathogenic role differs accordingly. In recessively inherited diseases, the mutations are usually expected to inactivate the gene product (loss of function mutations). The easiest to evaluate are probably nonsense mutations that introduce premature stop codons into the coding regions. If the nonsense mutations occur more than ~50 nucleotides upstream from the last exon-exon junction they usually lead to the degradation of the transcript by nonsense-mediated messenger ribonucleic acid (mRNA) decay (NMD) (Maquat 2004). Alternatively they may produce truncated protein products. Splicing mutations can be evaluated according to their effects on splicing of the corresponding transcripts by reverse transcriptase PCR (RT-PCR). Mutations affecting the consensus sequences at splice donor or splice acceptor sites or at splice branch sites may render them less effective and abolish the splicing partially or completely (Cartegni *et al.* 2002). In addition, cryptic splice-sites can be activated or splicing enhancers and silencers altered. These may lead to complete or partial exon skipping or intron retention, as well as to changes in the ratios of different splice variants. Deletions, insertions, and duplications are assessed based on their predicted effects on amino acid sequences. These, as well as splicing mutations, can introduce frameshifts and premature stop codons with consequences as described above. If the reading frame is maintained amino acid(s) can be deleted or inserted. Missense mutations are sometimes difficult to differentiate from rare neutral polymorphisms. A missense change is more likely to be pathogenic if the affected amino acid is conserved among homologous proteins in different species and/or in human. Differences in the chemical nature of the side chain (acidic vs. basic, polar vs. nonpolar) between the original and the substituting amino acid suggest a pathogenic role for a variant. A sequence variant in a part of the gene that codes for a functionally important domain in the corresponding protein is likely to be pathogenic.

1.8. Functional analyses

To understand the molecular pathology of the disease, a substantial amount of research is usually needed after the underlying gene has been identified. Revealing the normal biochemical and cellular function of the gene product will not only help in understanding the disease mechanism but will also give new information on biological functions and metabolic pathways in general. Many hints of the gene function can be obtained by analyzing the gene and/or protein sequence using bioinformatic sequence analysis tools and by exploring various databases on the internet. While a large number of these exist, only some are discussed below as examples. These tools are not always very effective and reliable, and different programs and/or databases may sometimes provide contradictory results. At least some laboratory research is usually needed to figure out the functions of the genes comprehensively.

Many physical and chemical characteristics of proteins can be predicted using various programs on the internet. For instance, discrimination between soluble and membrane proteins as well as assessment of the membrane protein topology can be achieved using several programs (e.g. TMHMM, www.cbs.dtu.dk/services/TMHMM-2.0/) (Krogh *et al.* 2001). A search for homologous or related sequences with known or predicted function, for example using different basic local alignment search tool (BLAST) (Altschul *et al.* 1990) programs through the NCBI web pages (www.ncbi.nlm.nih.gov/BLAST/), may be worthwhile since homologous genes have common evolutionary ancestors and are likely to have related functions. In addition, sequence alignments may identify conserved protein domains that have functional roles. Protein domain families have been collected in various databases, such as Pfam (available e.g. from pfam.sanger.ac.uk) (Sonnhammer *et al.* 1997).

The experimental approaches for the characterization of gene function are various and vary widely in the time frame and costs at which they are possible to carry out. The simplest experiments include the analysis of spatial and temporal gene expression patterns by northern blot and RT-PCR analyses as well as by ribonucleic acid (RNA) *in situ* hybridization in different cell lines and/or tissues collected from animals at different developmental stages. The intracellular localization of the wild-type and mutant proteins is also relatively easy to study in an overexpression system in cell cultures.

For the more demanding experiments aiming at the deep understanding of protein function and disease mechanisms, the limit is set only by the imagination. However, application of many of these approaches requires production of an antibody specific

for the protein in question. The disease may be modeled and the function of the protein studied in cell cultures and in various animal models, especially in mutant mice generated by gene targeting (Muller 1999). Inhibition of endogeneous gene expression by RNA interference (RNAi) (Fire *et al.* 1998) may also be useful for these purposes. Genomewide expression profiling using microarray techniques (Lockhart *et al.* 1996) may also be used. Analysis of interaction partners of a given protein by various methods may reveal the protein complex or cellular pathway it is involved in (Berggard *et al.* 2007).

2. Neuronal ceroid lipofuscinoses

The NCLs are worldwide occurring progressive encephalopathies that are, as a group, considered to be the most common childhood progressive hereditary neurodegenerative disorders (Rider and Rider 1988, Mole *et al.* 2005). The pathological features characteristic to all NCLs are the accumulation of autofluorescent ceroid and lipofuscin-like lipopigments in both neuronal and extraneuronal tissues, brain atrophy, and degeneration of neurons (Goebel 1997). NCLs are inherited in an autosomal recessive manner, with the exception of some rare autosomal dominant adult-onset forms (Peltonen *et al.* 2000b). Although the age of onset in NCLs varies from newborn to adult, the clinical manifestations are generally similar in all forms, including progressive psychomotor decline, epileptic seizures, loss of vision, and ultimately, premature death (Haltia 2003). At least ten forms of NCLs are thought to exist (CLN1-CLN9 and congenital NCL) (Haltia 2003) (Table 1), and before this thesis, six genes (*PPT1, TPP1, CLN3, CLN5, CLN6*, and *CLN8*) underlying them had been identified (Mole 2004) (Table 2). NCLs belong to the larger group of lysosomal storage disorders (section 3.4) (Jeyakumar *et al.* 2005). The storage bodies in the cells of NCL patients show variable ultrastructural characteristics that correlate with the form and the underlying gene, and typically have the appearance of granular osmiophilic deposits (GRODs), or curvilinear, rectilinear, or fingerprint profiles (Elleder *et al.* 1999) (Table 1). There are currently no effective treatment for NCLs (Hobert and Dawson 2006).

Table 1. Classification, phenotypes, and storage material in NCLs.

Gene/ putative gene	Major (and minor) phenotypes	Main accumulated protein	Ultrastructural phenotype
PPT1 (CLN1)	Infantile (late-infantile, juvenile, adult) NCL	SAPs A and D	GROD
TPP1 (CLN2)	Late-infantile (infantile, juvenile, protracted) NCL	subunit c	CL
CLN3	Juvenile (atypical, protracted, delayed) NCL	subunit c	FP
CLN4	Adult NCL: Kufs / Parry diseases	subunit c / SAP D	RL, CL, FP / GROD
CLN5	Late-infantile (atypical, delayed) NCL	subunit c	RL, CL, FP
CLN6	Late-infantile (atypical, protracted) NCL	subunit c	RL, CL, FP
CLN7	Late-infantile NCL	n.d.	RL, CL, FP
CLN8	EPMR: juvenile, protracted / late-infantile NCL	subunit c / n.d.	CL-like, granular / FP, CL, GROD-like
CLN9	Juvenile NCL	subunit c	GROD, FP, CL
n.d.	Congenital NCL	n.d.	GROD

Abbreviations: n.d. = not determined, EPMR = progressive epilepsy with mental retardation, SAP = sphingolipid activator protein (saposin), subunit c = subunit c of mitochondrial adenosine triphosphate synthase, GROD = granular osmiophilic deposits, CL = curvilinear profiles, FP = fingerprint profiles, RL = rectilinear profiles

Table 2. Currently known human NCL genes and proteins.

Gene	Chromosomal location	Protein	Cellular localization
PPT1 (CLN1)	1p34.2	Palmitoyl protein thioesterase 1 (PPT1)	Lysosomes, presynaptic regions, synaptosomes, synaptic vesicles
TPP1 (CLN2)	11p15.4	Tripeptidyl peptidase I (TPP1)	Lysosomes
CLN3	16p11.2	CLN3, unknown function	Lysosomes, early endosomes, presynaptic regions (not synaptic vesicles)
CLN5	13q22.3	CLN5, unknown function	Lysosomes, ER, neuronal extensions
CLN6	15q23	CLN6, unknown function	ER
CLN8	8p23.3	CLN8, unknown function	ER and ERGIC

Abbreviations: ER = endoplasmic reticulum, ERGIC = ER-Golgi intermediate compartment; chromosomal locations according to the Ensembl genome browser (www.ensembl.org)

2.1. Classification of NCLs

Originally, NCLs were classified into three childhood and two adult forms: infantile NCL (INCL; Haltia-Santavuori disease; CLN1), late-infantile NCL (LINCL; Jansky-Bielschowsky disease; CLN2), juvenile NCL (JNCL; Spielmeyer-(Vogt)-Sjögren or Batten disease; CLN3), and two forms of adult NCLs (ANCL; Kufs and Parry diseases;

CLN4) (Rider and Rider 1988). These are distinguished by differences in the ages of onset, slight variations in the clinical features and their order of appearance, and differences in neuropathological findings including ultrastructure of the storage material (Table 1). Following more detailed clinical characterization of these phenotypes as well as the progress in molecular genetic and biochemical studies that facilitated the definitive diagnoses, a still growing number of additional forms have been recognized. The late-infantile group has been revealed to be especially heterogeneous with several variant forms identified: CLN5 (Finnish vLINCL), CLN6, CLN7 (Turkish vLINCL), and CLN8 (Haltia 2003). The juvenile-onset group shows also some heterogeneity: Northern epilepsy (progressive epilepsy with mental retardation, EPMR; CLN8, allelic to CLN8-deficient vLINCL) and CLN9-deficient NCL can be considered as belonging to this category. In addition, a rare congenital NCL form occurring in newborns expands the age-range of NCL classification even further. After identification of NCL-causing genes, this classification has become somewhat out-of-date since almost all of these genes underlie atypical, protracted, less severe, and/or delayed forms of the diseases in addition to the classical phenotypes used in the original classification. All above mentioned NCL forms are discussed in more detail and with appropriate references in sections 2.2 and 2.3.

2.2. NCLs with known molecular genetic basis

2.2.1. CLN1

The *PPT1* gene, assigned by linkage analysis to the short arm of chromosome 1 in Finnish families with INCL (Järvelä *et al.* 1991), was identified by the positional candidate gene cloning method as a gene encoding palmitoyl protein thioesterase 1 (PPT1) (Vesa *et al.* 1995). *In vitro*, PPT1 removes palmitate residues from *S*-acylated proteins (Camp and Hofmann 1993), whereas the *in vivo* substrates remain unidentified. While in non-neuronal cells PPT1 is a soluble lysosomal enzyme (Hellsten *et al.* 1996, Verkruyse and Hofmann 1996), in neurons it is also present in presynaptic regions of axons and specifically in synaptosomes and synaptic vesicles, suggesting an extralysosomal function in the brain (Heinonen *et al.* 2000, Lehtovirta *et al.* 2001, Ahtiainen *et al.* 2003). The crystal structure of PPT1 has been determined, providing a structural basis for the genotype-phenotype correlations (Bellizzi *et al.* 2000). The exact physiological function of PPT1 is still poorly understood.

Approximately half of the mutations identified in the *PPT1* gene thus far lead to the most common, the earliest-onset, and the most severe form of CLN1, INCL, that

is enriched in the Finnish population with an incidence of 1:20 000 (Santavuori *et al.* 2000). Affected children are healthy until the age of 6-18 months after which the symptoms including rapid psychomotor deterioration, hypotonia, ataxia, visual failure, microcephaly, myoclonus, and epilepsy develop (Santavuori *et al.* 1973, Santavuori *et al.* 1974). The disease progresses rapidly and death occurs at ~10 years of age (Santavuori *et al.* 2000). In addition to INCL, *PPT1* mutations cause a variety of other clinical phenotypes, with the age of onset varying up to adulthood (Vesa *et al.* 1995, Das *et al.* 1998, Mitchison *et al.* 1998, van Diggelen *et al.* 2001). The unifying feature in all PPT1 deficiencies is the characteristic granular storage material with GRODs in patient cells (Haltia *et al.* 1973, Das *et al.* 1998, Mitchison *et al.* 1998, van Diggelen *et al.* 2001). At least in INCL the major portion of the accumulated proteins in the storage bodies constitutes of sphingolipid activator proteins (saposins) A and D (Tyynelä *et al.* 1993).

More than 40 *PPT1* mutations have been identified (NCL Mutation Database, www.ucl.ac.uk/ncl/mutation). Some common mutations exist, of which the most common (c.364A>T, p.Arg122Trp), associated with INCL, is enriched in the Finnish population due to a founder effect (Vesa *et al.* 1995). There is some evidence of genotype-phenotype correlation. Mutations predicted to result in loss of mRNA and/or protein or in severely truncated proteins (nonsense or frameshift-causing mutations) are usually associated with INCL. Other mutations (e.g. missense mutations) that lead to INCL generally occur near the enzymatically active site and affect catalysis, substrate binding, conformation, or stability of PPT1 more dramatically than those that lead to later-onset phenotypes and occur in more peripheral sites (Bellizzi *et al.* 2000, Das *et al.* 2001). In addition, INCL-causing mutations lead to lack of PPT1 enzymatic activity both in overexpression systems and in patient cells whereas some residual activity is detected with mutations associated with later-onset diseases (Vesa *et al.* 1995, Das *et al.* 1998, Das *et al.* 2001, van Diggelen *et al.* 2001, Lyly *et al.* 2007). Some of the mutations have also been shown to disturb the intracellular routing of the protein in overexpression systems, and the severity of the defect correlates to some extent with the severity of the resulting phenotype (Hellsten *et al.* 1996, Das *et al.* 2001, Salonen *et al.* 2001, Lyly *et al.* 2007).

No naturally occurring animal models for PPT1 deficiency exist but two mouse models replicating INCL phenotype have been generated by targeted disruption of *Ppt1* (Gupta *et al.* 2001, Jalanko *et al.* 2005).

2.2.2. CLN2

A homozygosity mapping approach was applied to localize the *TPP1* gene to chromosome 11p15 in families with classical LINCL originating from several countries

(Sharp *et al.* 1997). The gene, encoding a pepstatin-insensitive carboxyl protease, was identified by a biochemical strategy as a lysosomal, mannose-6-phosphorylated protein missing from the brain samples of patients with classical LINCL (Sleat *et al.* 1997). Subsequently, the protein was recognized as tripeptidyl peptidase I (TPP1), a serine-carboxyl proteinase that removes tripeptides from the N-termini of polypeptides (Rawlings and Barrett 1999, Tomkinson 1999, Vines and Warburton 1999, Lin *et al.* 2001, Wlodawer *et al.* 2001). Although several substrates have been suggested based on *in vitro* experiments (reviewed in (Kyttälä *et al.* 2006)), the natural substrates and cellular function of TPP1 are not currently known.

Most of the mutations identified in the *TPP1* gene lead to classical LINCL with the age of onset between two and four years (Williams *et al.* 1999). The symptoms include seizures, ataxia, myoclonus, developmental regression, psychomotor deterioration, and visual failure. Patients become chair-bound between four and six years of age, and death occurs in middle childhood (Williams *et al.* 1999, Williams *et al.* 2006). The storage bodies in patient cells are most commonly curvilinear in their ultrastructural appearance (Williams *et al.* 1999), and have subunit c of mitochondrial adenosine triphosphate (ATP) synthase as the main protein component (Hall *et al.* 1991, Palmer *et al.* 1992). In addition to classical LINCL, mutations in *TPP1* cause juvenile-onset or protracted diseases (Hartikainen *et al.* 1999, Sleat *et al.* 1999, Wisniewski *et al.* 1999, Steinfeld *et al.* 2002) as well as infantile-onset disease (Ju *et al.* 2002).

To date, over 50 *TPP1* mutations have been identified (NCL Mutation Database, www.ucl.ac.uk/ncl/mutation). Two of the mutations, IVS5-1G>C that affects the splicing of the transcript, and a nonsense mutation c.622C>T (p.Arg208X), are especially common. No clear genotype-phenotype correlation has been established, and most of the mutations, irrespective of the mutation type and the resulting phenotype, have been reported to lead to a deficiency in TPP1 enzyme activity (Sleat *et al.* 1999, Wisniewski *et al.* 1999, Steinfeld *et al.* 2004). In addition, some missense mutations have been shown to disturb the intracellular trafficking of the protein in an overexpression system (Steinfeld *et al.* 2004).

A mutation in the canine *TPP1* gene has been identified as the cause for an NCL-like phenotype occurring in Dachshund dogs (Awano *et al.* 2006b). In addition, a mouse model for TPP1 deficiency has been generated by targeted disruption of *Tpp1* (Sleat *et al.* 2004).

2.2.3. CLN3

The *CLN3* gene was first localized by linkage analysis and fine mapping to 16p11.2-p12.1 in families with JNCL originating from several countries (Eiberg *et al.* 1989, Mitchison *et al.* 1994), and subsequently identified by positional cloning (The International Batten Disease Consortium 1995). *CLN3* codes for a novel membrane protein with most likely six transmembrane domains (The International Batten Disease Consortium 1995, Janes *et al.* 1996, Kyttälä *et al.* 2004). The subcellular localization of CLN3 depends on the cell type. In extraneural cells it is targeted primarily to the lysosomal compartment (Järvelä *et al.* 1998, Järvelä *et al.* 1999, Kyttälä *et al.* 2004). In neurons, additional localization to early endosomes and to synaptic regions excluding synaptic vesicles suggests extralysosomal roles in neuronal cells (Luiro *et al.* 2001, Kyttälä *et al.* 2004). The function of CLN3 is still not clear but possible roles in the maintenance of lysosomal pH homeostasis, arginine transport, membrane trafficking, and in preventing apoptosis have been suggested (reviewed by (Kyttälä *et al.* 2006)).

The most common phenotype caused by *CLN3* mutations is JNCL that manifests at the age of four to seven years with rapidly progressing visual failure (Hofman *et al.* 1999). Other symptoms include progressive mental and motor deterioration, and seizures. Patients become non-ambulatory usually at 15-28 years of age, and die in the second or third decade of life (Santavuori *et al.* 2000). Globally, JNCL is the most common form of the NCLs, and, moreover, it is especially common in the Finnish population with an incidence of 1:21 000 (Santavuori *et al.* 2000). Special findings in patients with *CLN3* mutations are the occurrence of vacuolated lymphocytes as well as fingerprint profiles as the characteristic ultrastructural feature of the storage material (Hofman *et al.* 1999). The main protein component in storage material is subunit c of mitochondrial ATP synthase (Hall *et al.* 1991, Palmer *et al.* 1992). *CLN3* mutations have also been associated with atypical, protracted, or delayed forms of the disease (Järvelä *et al.* 1997, Munroe *et al.* 1997, Åberg *et al.* 1998, Wisniewski *et al.* 1998, Lauronen *et al.* 1999).

More than 40 mutations in *CLN3* have been identified thus far (NCL Mutation Database, www.ucl.ac.uk/ncl/mutation). The most widespread of these, occurring in almost all patients in at least heterozygous form, is a 1-kb deletion (of exons 7-8, c.461-677del) that produces a frameshift and a premature stop codon (The International Batten Disease Consortium 1995, Munroe *et al.* 1997). Recently, it has been suggested that rather than the protein being degraded, truncated mutant proteins are produced that retain some CLN3 function (Kitzmuller *et al.* 2008). Nevertheless, the intracellular trafficking of the mutant protein has been shown to be blocked in overexpression systems (Järvelä *et al.* 1999). In most patients the homozygous 1-kb deletion is associated with the JNCL phenotype whereas in

combination with other mutations it may lead to atypical phenotypes (Järvelä *et al.* 1997, Munroe *et al.* 1997, Åberg *et al.* 1998, Wisniewski *et al.* 1998, Lauronen *et al.* 1999). In contrast to the 1-kb deletion, many of the missense mutations have been reported to have no effect on the subcellular localization of the protein in overexpression systems (Järvelä *et al.* 1999, Haskell *et al.* 2000). Some missense mutations that are associated with milder phenotypes restore the function of the CLN3 homolog battenin (Btn1p, Yhc3p) in Btn1p-deficient yeast cells (Haskell *et al.* 2000).

No natural model organisms for CLN3 deficiency exist but several have been artificially generated. These include two knock-out mice (Katz *et al.* 1999, Mitchison *et al.* 1999), one knock-in mouse (*Cln3*$^{\Delta ex7\text{-}8}$, (Cotman *et al.* 2002)), and one β-galactosidase reporter mouse (Eliason *et al.* 2007) models. In addition, yeast (*Saccharomyces cerevisiae* (Pearce and Sherman 1997) and *Schizosaccharomyces pombe* (Gachet *et al.* 2005)), and *Caenorhabditis elegans* (de Voer *et al.* 2005) models for CLN3 deficiency have been produced.

2.2.4. CLN5

The *CLN5* gene was localized by linkage analysis to chromosome 13q22 in Finnish families with vLINCL (Savukoski *et al.* 1994, Klockars *et al.* 1996), and subsequently identified by positional cloning as a gene encoding a novel glycoprotein with four isoforms of different lengths resulting from the use of alternative initiator methionines (Savukoski *et al.* 1998, Isosomppi *et al.* 2002, Vesa *et al.* 2002). CLN5 may exist in both soluble and membrane-associated forms, and in non-neuronal cells it is targeted to lysosomes (Isosomppi *et al.* 2002, Vesa *et al.* 2002, Holmberg *et al.* 2004, Bessa *et al.* 2006). In neurons, CLN5 localizes to cell soma, to lysosomes and endoplasmic reticulum (ER), and to neuronal extensions (Holmberg *et al.* 2004). CLN5 has been shown to interact with TPP1 and CLN3 proteins *in vitro* (Vesa *et al.* 2002). The function of CLN5 is currently unknown.

CLN5 disease was initially identified in Finnish patients, and has therefore been denoted as Finnish vLINCL (Santavuori *et al.* 1982). While clinically the disease resembles classical LINCL (CLN2), the age of onset is somewhat later, between four and seven years (Santavuori *et al.* 1982, Santavuori *et al.* 1991). In addition, the disease shows a slower clinical course. Patients lose the ability to walk at ~10 years of age, while death usually occurs between the ages of 14 and 32 years (Santavuori *et al.* 1982, Santavuori *et al.* 1991, Santavuori *et al.* 1999). The ultrastructure of the storage bodies in patient cells includes features of rectilinear, curvilinear, and fingerprint profiles (Santavuori *et al.* 1982, Tyynelä *et al.* 1997). The major protein component in storage material is subunit c of mitochondrial ATP synthase (Tyynelä *et*

al. 1997). Occasionally, *CLN5* mutations are associated with atypical and/or later-onset disease phenotypes (Pineda-Trujillo *et al.* 2005, Cannelli *et al.* 2007).

Although at least 13 mutations in *CLN5* have been reported to underlie the disease (NCL Mutation Database, www.ucl.ac.uk/ncl/mutation), one of these, the Finnish founder mutation c.1175delAT (p.Tyr392X), is identified in the majority of the families (Savukoski *et al.* 1998). Contradictory results of its effect on trafficking of the mutant protein have been reported in overexpression systems: it either arrests the protein in the Golgi or has no effect (Isosomppi *et al.* 2002, Vesa *et al.* 2002). Some of the other mutations have been shown not to interfere with the subcellular localization of the mutant proteins (Vesa *et al.* 2002). Mutations in *CLN5* have also been shown to disrupt the interaction of CLN5 with TPP1 but not with CLN3 (Vesa *et al.* 2002). There seems to be no obvious genotype-phenotype correlation since, except for a few atypical cases (Pineda-Trujillo *et al.* 2005, Cannelli *et al.* 2007), the phenotype is rather uniform irrespective of the underlying mutation ((Holmberg *et al.* 2000), NCL Mutation Database, www.ucl.ac.uk/ncl/mutation).

Two natural animal models for CLN5 deficiency exist: *CLN5* mutations cause NCL-like phenotypes in Border collie dogs (Melville *et al.* 2005) and in Devon cattle (Houweling *et al.* 2006). In addition, a CLN5 mouse model has been generated by targeted disruption of *Cln5* (Kopra *et al.* 2004).

2.2.5. CLN6

The *CLN6* gene was localized to chromosome 15q21-23 using a homozygosity mapping strategy in two consanguineous families with vLINCL originating from the Indian subcontinent (Sharp *et al.* 1997). The gene was subsequently identified by positional cloning as a gene coding for a novel ER-resident transmembrane protein with seven membrane-spanning domains (Gao *et al.* 2002, Wheeler *et al.* 2002, Heine *et al.* 2004, Mole *et al.* 2004, Heine *et al.* 2007). The function of CLN6 is unknown although its defects have been reported to result in lysosomal dysfunction (Heine *et al.* 2004).

Virtually all mutations in the *CLN6* gene lead to a vLINCL phenotype that, apart from the later-onset at three to eight years of age and the slower progression, is clinically similar to classical LINCL (CLN2) (Mole *et al.* 2005). The ultrastructure of the storage bodies is comprised of fingerprint bodies as well as curvilinear and rectilinear profiles whereas the main protein component seems to be subunit c of mitochondrial ATP synthase (Elleder *et al.* 1997, Mole *et al.* 2005).

In all, 27 mutations have been identified in the *CLN6* gene (NCL Mutation Database, www.ucl.ac.uk/ncl/mutation). There is no major founder mutation in *CLN6* and most of the mutations are family-specific (Sharp *et al.* 2003). However, two mutations are more common than others: c.214G>T (p.Glu72X) identified in Costa Rican patients and c.460_462delATC (p.Ile154del) in Portuguese patients (Gao *et al.* 2002, Wheeler *et al.* 2002). Mutations in *CLN6* have also been detected in patients originating, for example, from other Mediterranean countries and from the Indian subcontinent (Wheeler *et al.* 2002, Sharp *et al.* 2003, Teixeira *et al.* 2003). Irrespective of the underlying *CLN6* mutation, the vLINCL phenotype is clinically broadly uniform with only a few patients reported to show atypical or protracted disease course ((Sharp *et al.* 2003), NCL Mutation Database, www.ucl.ac.uk/ncl/mutation). When studying the effects of mutations on protein levels, CLN6 protein was shown to be absent from fibroblasts of two patients with homozygous *CLN6* mutations introducing premature stop codons (c.316dupC and p.Glu72X) (Mole *et al.* 2004). In addition, some of the mutations affecting single amino acids were shown to have no effect on the ER localization of the mutant CLN6 proteins in an overexpression system (Mole *et al.* 2004).

Naturally occurring mutations in *CLN6* orthologs have been identified in the neuronal ceroid lipofuscinosis (*nclf*) mouse (Bronson *et al.* 1998, Gao *et al.* 2002, Wheeler *et al.* 2002) and in Merino sheep (Tammen *et al.* 2006) that show NCL-like phenotypes.

2.2.6. CLN8

The *CLN8* gene, localized by linkage analysis to chromosome 8p23 in Finnish families with EPMR (Tahvanainen *et al.* 1994), and subsequently isolated by positional cloning, is predicted to encode a membrane protein with several (4-7) transmembrane domains (Ranta *et al.* 1999, Lonka 2004). In extraneural cells, ER-resident CLN8 protein has been suggested to recycle between the ER and ER–Golgi intermediate compartment (ERGIC) (Lonka *et al.* 2000), whereas in neurons it is localized mainly to the ER and possibly additional locations outside the ER (Lonka *et al.* 2004). The function of CLN8 is unknown but it belongs to the TRAM-Lag1p-CLN8 (TLC) family of proteins, members of which are suggested to have roles in biosynthesis, metabolism, transport, and sensing of lipids (Winter and Ponting 2002).

Northern epilepsy, in which the *CLN8* gene was first identified, is present exclusively in Finnish patients (Ranta *et al.* 1999). This mildest form of the childhood-onset NCLs presents with epilepsy usually at the age of 5–10 years, after which the other symptoms, progressive mental deterioration, and motor and behavioural problems, follow (Hirvasniemi *et al.* 1994). The patients may survive until 50-60

years of age (Hirvasniemi *et al.* 1995). The storage material in EPMR patient cells consists mostly of subunit c of mitochondrial ATP synthase and shows patterns resembling curvilinear, and to a lesser extent, granular ultrastructure (Herva *et al.* 2000). Although EPMR was for a long time the only phenotype associated with mutations in the *CLN8* gene, *CLN8* mutations have now been identified in patients with more severe vLINCL phenotypes originating from Turkey, Italy, and Israel (Mitchell *et al.* 2001, Ranta *et al.* 2004, Topcu *et al.* 2004b, Cannelli *et al.* 2006, Zelnik *et al.* 2007). The disease in these patients is rather uniform and closely resembles the other vLINCLs with an age of onset between two and seven years. Ultrastructural examination of the storage material revealed fingerprint profiles and/or curvilinear bodies, and occasionally GROD-like deposits (Mitchell *et al.* 2001, Ranta *et al.* 2004, Topcu *et al.* 2004b, Cannelli *et al.* 2006, Zelnik *et al.* 2007).

Eleven mutations have now been identified in *CLN8* (NCL Mutation Database, www.ucl.ac.uk/ncl/mutation). The most common mutation c.70C>G (p.Arg24Gly) has been identified in homozygous form in all but one EPMR patient (Ranta *et al.* 1999, Siintola *et al.* 2006). This and two of the missense mutations present in Turkish patients (p.Arg204Cys and p.Trp263Cys) have been shown to have no effect on the subcellular localization of mutant CLN8 in overexpression systems (Lonka *et al.* 2000, Lonka *et al.* 2004). Among *CLN8*-associated vLINCL, no genotype-phenotype correlation can be detected.

Two naturally occurring animal models for CLN8 deficiency have been described: NCL-like phenotypes in motor neuron degeneration (*mnd*) mice and in English Setter dogs are caused by mutations in the corresponding *CLN8* orthologs (Ranta *et al.* 1999, Katz *et al.* 2005).

2.3. NCLs with unknown molecular genetic basis

2.3.1. CLN4

A heterogeneous group of ANCLs are the mildest forms of the NCLs, with ages of onset ranging from 11 to 50 years (Berkovic *et al.* 1988). The symptoms, depending on the form of the disease, may include dementia, myoclonus, epilepsy, ataxia, motor disturbances, late pyramidal and extrapyramidal symptoms, and behavioural changes, whereas no visual failure occurs (Berkovic *et al.* 1988, Martin *et al.* 1999). The diseases progress slowly and lead to death on average 12.5 years after onset (Berkovic *et al.* 1988). Ultrastructural examinations of the storage material have revealed variable patterns with granular, fingerprint, curvilinear, and rectilinear

profiles detected in various combinations in different tissues (Martin *et al.* 1999). Accumulation of subunit c of mitochondrial ATP synthase and/or saposin D in storage material has been shown at least in some patients (Hall *et al.* 1991, Nijssen *et al.* 2003).

ANCL is most commonly inherited in an autosomal recessive manner (Kufs disease) but some cases with autosomal dominant inheritance (Parry disease) have also been reported (Berkovic *et al.* 1988, Martin *et al.* 1999, Nijssen *et al.* 2002). While the putative gene locus *CLN4* has been assigned to ANCL, its molecular genetic background is most probably genetically heterogeneous and remains to be elucidated. Given that most of the NCL-causing genes have additionally been associated with later-onset and/or atypical forms, they can be considered candidate genes for ANCL forms. Most notably, *PPT1* mutations have been identified in adult-onset NCL with GRODs (van Diggelen *et al.* 2001).

2.3.2. CLN7

The *CLN7* locus was first designated in 1999 as the one harbouring the causative gene for vLINCL present in Turkish patients (Wheeler *et al.* 1999). This Turkish vLINCL was considered a distinct genetic entity since the other NCL loci known at that time (*CLN1/PPT1, CLN2/TPP1, CLN3, CLN5,* and *CLN6*) were excluded by homozygosity analyses (Wheeler *et al.* 1999). Subsequent homozygosity mapping in consanguineous Turkish vLINCL families identified a region of shared homozygosity of marker alleles on chromosome 8p23 (Mitchell *et al.* 2001) that contained yet another NCL-causing gene, *CLN8*, that had meanwhile been identified (Ranta *et al.* 1999). After extending the family panel with additional, mostly consanguineous Turkish families with vLINCL (described by (Topcu *et al.* 2004b)), the disease-causing mutations were finally identified in *CLN8* in a subset of Turkish patients with vLINCL, thus excluding these patients from the CLN7 entity (Ranta *et al.* 2004). However, the remaining patients, still considered to represent a distinct genetic entity (CLN7), were lacking a molecular genetic explanation for their disease.

The clinical course of the disease in these remaining Turkish patients with vLINCL is broadly similar to the other vLINCLs caused by mutations in *CLN5, CLN6*, and *CLN8*. The disease presents at the age of two to seven years, most commonly with seizures that show more a severe course compared to classical LINCL (CLN2) (Mitchell *et al.* 2001, Topcu *et al.* 2004b). The other initial symptoms may be motor, visual, and speech impairment (Topcu *et al.* 2004b). The disease progresses rapidly, with the additional symptoms including myoclonus, mental regression, blindness, ataxia, and personality disorders. Most patients have been reported to become unambulatory a few years after disease onset (Mitchell *et al.* 2001, Topcu *et al.* 2004b), and to die

prematurely (Topcu M, personal communication). Electron microscopic examinations revealed solid fingerprint profiles, curvilinear bodies, and/or rectilinear bodies in the storage material of patient cells (Fig. 1) ((Mitchell *et al.* 2001, Topcu *et al.* 2004b), Elleder M, personal communication). Electroencephalogram (EEG), electroretinogram (ERG), and visual evoked potentials (VEP) recordings have been reported to be abnormal (Topcu *et al.* 2004b). On magnetic resonance imaging (MRI), patients showed atrophy of the cerebellum and more mildly of cerebrum, and in most cases, also brainstem involvement (Topcu *et al.* 2004b).

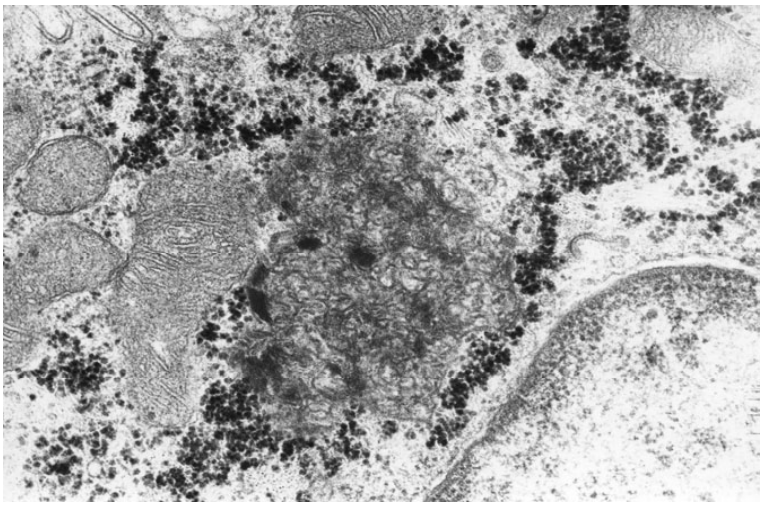

Figure 1. An electron micrograph from an epithelial cell of the skin eccrine gland of a patient with CLN7. The ultrastructural pattern consists of irregular curvilinear or rectilinear profiles in a mixture with fingerprints. Magnification is 38 000x. The picture is a courtesy of Prof. Milan Elleder (Institute of Inherited Metabolic Disorders, Charles University, Prague, Czech Republic).

2.3.3. CLN9

The most recently described NCL form is CLN9-deficient NCL (Schulz *et al.* 2004). It resembles JNCL (CLN3) clinically but is differentiated from it and from most other NCLs by a distinctive gene expression pattern and phenotype of CLN9-deficient cells (Schulz *et al.* 2004). Storage material in patient cells showed GRODs as well as fingerprint and curvilinear patterns of ultrastructure and was in one patient shown to be immunoreactive for subunit c of mitochondrial ATP synthase (Schulz *et al.* 2004). The molecular genetic basis of CLN9 is currently unknown and involvement of all known NCL-causing genes have been excluded either by sequence analysis (*CLN3, CLN5, CLN6*, and *CLN8*), or by enzymatic tests (*PPT1* and *TPP1*) (Schulz *et al.* 2004). However, transfection of CLN9-deficient cells with *CLN8* partially corrects the phenotype of these cells (Schulz *et al.* 2006). *CLN9* has been suggested to encode a regulator of dihydroceramide synthase (Schulz *et al.* 2006).

2.3.4. Congenital NCL

Congenital NCL, the earliest-onset and the most aggressive form of the human NCLs, is very rare with only a few cases described (Norman and Wood 1941, Brown *et al.* 1954, Sandbank 1968, Humphreys *et al.* 1985, Garborg *et al.* 1987, Barohn *et al.* 1992). Clinically the disease presents at birth with microcephaly, respiratory problems, rigidity, and epilepsy. Patients die within hours to weeks after birth. Upon autopsies, extreme atrophy of the brains of the patients was observed, and the brains were extremely small and firm. In addition, severe loss of neurons in the cerebral and cerebellar cortex accompanied by extensive gliosis was detected. The autofluorescent storage bodies showed granular ultrastructure (GRODs) (Humphreys *et al.* 1985, Garborg *et al.* 1987, Barohn *et al.* 1992). Based on the occurrence of the disease in more than one child of healthy parents in two families, a recessive mode of inheritance has been proposed (Brown *et al.* 1954, Sandbank 1968). However, prior to this thesis the specific molecular genetic defect was unknown.

2.4. Genes underlying NCL-like phenotypes in animals

NCL-like phenotypes have been identified in many animals, including cow, dog, cat, sheep, goat, and mouse (Lingaas *et al.* 1999). In some of these the underlying genetic defect is known. In some cases, the corresponding human ortholog is associated with a human NCL (as discussed in section 2.2), whereas for others no NCL-causing mutations have been identified in the orthologous human gene. In many NCL-like diseases of animals, the underlying genetic defect remains to be identified.

Genes whose defects cause NCL-like phenotypes in animals but not in humans include cathepsin D (*CTSD*) that is defective in naturally occurring congenital ovine NCL (Tyynelä *et al.* 2000), and in an NCL-like phenotype described in American Bulldogs (Awano *et al.* 2006a). Disruption of cathepsin D (*Ctsd/cathD*) by gene targeting in mouse (Saftig *et al.* 1995, Koike *et al.* 2000) and in *Drosophila melanogaster* (Myllykangas *et al.* 2005) also leads to NCL-like phenotypes. CTSD and its deficiencies will be discussed in more detail in section 3.2.1. In addition, artificially generated mouse models for CTSF deficiency, as well as for chloride channel 3 (CLCN3), 6 (CLCN6), and 7 (CLCN7) deficiencies, exhibit NCL-like phenotypes (Kornak *et al.* 2001, Yoshikawa *et al.* 2002, Kasper *et al.* 2005, Poet *et al.* 2006, Tang *et al.* 2006). Moreover, phenotypes of mice deficient for a PPT1-homolog, PPT2 (Gupta *et al.* 2001, Gupta *et al.* 2003), for two cathepsins, CTSB and CTSL (Felbor *et al.* 2002, Koike *et al.* 2005), and for an ancillary β-subunit of CLCN7, osteopetrosis associated transmembrane protein 1 (OSTM1) (Lange *et al.* 2006), have been reported to resemble NCLs.

3. Lysosomes

Lysosomes, discovered by Christian de Duve and his colleagues more than 50 years ago (De Duve *et al.* 1955), are intracellular, acidic, membrane-enclosed, digestive organelles. They degrade and recycle material received from the secretory, endocytic, autophagic, and phagocytic membrane-trafficking routes (Luzio *et al.* 2007). They are present in most animal cells, and are heterogeneous in size and morphology. Lysosomes contain at least 50-60 soluble hydrolytic enzymes that degrade various macromolecules (Journet *et al.* 2002, Kollmann *et al.* 2005, Sleat *et al.* 2005). The acidic pH optimum of the lysosomal enzymes, so-called acid hydrolases, is achieved by vacuolar H^+ -ATPases that pump protons into the lysosomal lumen and maintain the pH at 4.5-5 (Mellman *et al.* 1986). Lysosomes are surrounded by a single membrane that preserves their integrity and protects the other parts of the cell from the harmful effects of the digestive enzymes (Eskelinen *et al.* 2003). The membrane has a complex protein composition with up to 215 membrane proteins identified (Bagshaw *et al.* 2005, Schröder *et al.* 2007). Besides the roles in turnover of endogenous and exogenous macromolecules, lysosomes have other functions as well, for instance as a secretory compartment in some cell types (Blott and Griffiths 2002). Defects in many of the lysosomal enzymes and membrane proteins as well as proteins involved in their trafficking and in lysosome biogenesis lead to lysosomal storage disorders (LSDs) (Futerman and van Meer 2004).

3.1. Targeting of lysosomal proteins

Most of the newly synthesized soluble lysosomal proteins receive mannose 6-phosphate (M6P) modifications to their *N*-linked oligosaccharides in the *cis*-Golgi by an enzymatic reaction series (reviewed by (Kornfeld and Mellman 1989)). In the *trans*-Golgi network (TGN) these proteins are selected for lysosomal transport as M6P residues bind to M6P receptors (MPRs), either to cation-dependent (~46 kilodaltons (kDa); CD-MPR; MPR46) or cation-independent (~300 kDa; CI-MPR; MPR300) MPRs (Ghosh *et al.* 2003). In order to be transported to the endosomal compartments, the MPR-attached proteins are packaged to clathrin-coated, adaptor protein-1 (AP-1) containing vesicles that form with the aid of GGA (Golgi-localized, γ-ear-containing, adenosine diphosphate ribosylation factor-binding protein) adaptors (Doray *et al.* 2002). In the acidic pH of the endosome lumen the proteins dissociate from the receptors and are transported to the lysosomes while the MPRs are recycled back to the TGN (Ghosh *et al.* 2003). The site of the digestive action of the hydrolases is the hybrid organelle that results from the fusion of late endosome and lysosome by a

transient (kissing event) and/or a permanent fusion of these organelles (Bright *et al.* 2005). Finally, lysosomes are reformed by a maturation process that involves condensation of the lumenal contents and removal of some membrane components (Pryor *et al.* 2000, Bright *et al.* 2005). Lysosomes can thus be seen as storage granules for lysosomal enzymes. To reach lysosomes soluble lysosomal proteins may also use alternative, MPR-independent pathways that may be protein and cell type specific, as shown in a study of mice deficient for both MPRs (Dittmer *et al.* 1999).

Newly synthesized lysosomal membrane proteins are transported to lysosomes independently of MPRs either by an indirect route through plasma membrane or by direct intracellular routes. The signals for the lysosomal targeting of membrane proteins lie most commonly in their cytoplasmic domains (Bonifacino and Traub 2003). The most common are signals with tyrosine-based (Peters *et al.* 1990, Williams and Fukuda 1990) and dileucine-based (Letourneur and Klausner 1992) consensus sequences. The targeting signals are recognized by cytosolic adaptor protein (AP) complexes of which AP-3 has been proposed to be mainly responsible for lysosomal sorting acting both in TGN and early endosomes (Ihrke *et al.* 2004). Membrane proteins may use several pathways to reach late endosomes/lysosomes: directly from TGN (Rous *et al.* 2002) or through early endosomes (Peden *et al.* 2004), or indirectly via plasma membrane and early endosomes (Ihrke *et al.* 2004). Use of these pathways may depend on specific protein and cell types.

3.2. Soluble lysosomal proteins

The majority of the soluble lysosomal proteins are hydrolytic enzymes including proteases, glycosidases, lipases, nucleases, phosphatases, and sulfatases (Journet *et al.* 2002, Kollmann *et al.* 2005, Sleat *et al.* 2005). Lysosomal proteases, such as cathepsins, are according to their active site amino acids divided into three subgroups: cysteine, aspartic, and serine proteases (Brix 2005). Among soluble lysosomal proteins involved in protein degradation are PPT1 and TPP1, defects in which underlie CLN1 and CLN2 diseases, respectively, which were reviewed in sections 2.2.1 and 2.2.2.

3.2.1. Cathepsin D

Cathepsin D is a lysosomal enzyme that belongs to the pepsin family of aspartic proteases (Press *et al.* 1960, Rawlings and Barrett 1995). Human CTSD is encoded by nine exons of the *CTSD* gene on chromosome 11p15.5 (Faust *et al.* 1985, Augereau *et al.* 1988, Redecker *et al.* 1991). Synthesis and maturation of CTSD to an active enzyme is a complex process involving trimming during several proteolytic processes

during the transport to lysosomes. CTSD is synthesized as a larger inactive preproenzyme that is co-translationally translocated into ER with the aid of an N-terminal 20 amino acid signal sequence, a pre-sequence, that is cleaved off by ER signal peptidases (Erickson and Blobel 1979, Hasilik and Neufeld 1980b, Erickson *et al.* 1981). The resulting proCTSD (53 kDa) is transported to lysosomes either through MPR-dependent or MPR-independent pathways (Hasilik and Neufeld 1980a, Kornfeld and Mellman 1989, Rijnboutt *et al.* 1991, Glickman and Kornfeld 1993, Dittmer *et al.* 1999). ProCTSD is converted to a single chain intermediate form of 43 kDa by cleavage of the N-terminal 44 amino acid prosequence (propeptide) that functions as an activation peptide keeping the enzyme inactive before it is cleaved (Erickson *et al.* 1981, Gieselmann *et al.* 1983). In several species, including human, the single-chain form is processed further to a two-chain form consisting of two polypeptide chains linked together noncovalently by disulfide bonds: a light chain of 14 kDa (N-terminal part) and heavy chain of 31 kDa (C-terminal part) (Huang *et al.* 1979, Hasilik and Neufeld 1980b, Metcalf and Fusek 1993). The mechanisms of these proteolytic cleavages are not well characterized but they appear to depend on cysteine proteases (Gieselmann *et al.* 1985, Samarel *et al.* 1989).

The structure of CTSD is bilobed with an active site cleft located in the groove in between the two domains (Metcalf and Fusek 1993). The catalytic site contains two aspartic acids crucial for the enzymatic activity, one on each CTSD chain (Faust *et al.* 1985, Metcalf and Fusek 1993). CTSD is an endopeptidase with a pH optimum of 3-4 that prefers peptide bonds flanked by hydrophobic amino acid residues (Press *et al.* 1960, Rawlings and Barrett 1995). An important role of CTSD in the turnover of lysosomal proteins is most likely dependent on limited proteolysis rather than on bulk protein degradation (Saftig *et al.* 1995). CTSD has been implicated in processing of cell and tissue specific substrates, such as antigens (Rodriguez and Diment 1992, Mohamadzadeh *et al.* 2004) and hormones (Diment *et al.* 1989), as well as in apoptosis (Liaudet-Coopman *et al.* 2006). CTSD has been associated with several diseases, including NCLs (see below), Alzheimer's disease (Cataldo *et al.* 1995, Papassotiropoulos *et al.* 1999), and cancer (Fusek and Vetvicka 2005, Liaudet-Coopman *et al.* 2006).

3.2.1.1. Cathepsin D deficiency in animals

Both naturally occurring and artificially generated CTSD defects cause NCL-like phenotypes in animals, as mentioned in section 2.4. Naturally occurring congenital ovine NCL (CONCL) was recognized in White Swedish Landrace sheep (Järplid and Haltia 1993). The lambs were severely affected at birth: they were trembling, weak, and unable to stand, and died within a few days (Järplid and Haltia 1993). The brains of the affected lambs were very small with thin cortices (Tyynelä *et al.* 2000). Loss of

neurons was more severe in the cerebrum and hippocampus than in the cerebellum. Non-neuronal tissues were unaffected. The autofluorescent storage material, present both in neuronal and non-neuronal tissues, showed GRODs upon electron microscopic examination (Järplid and Haltia 1993), and contained saposins A and D (Tyynelä *et al.* 2000). The disease was inherited in an autosomal recessive manner, and the underlying genetic defect was identified in the *CTSD* gene: a homozygous missense mutation G->A changes an active site aspartic acid to an asparagine (amino acid 269 in sheep CTSD corresponding to p.295 in human CTSD) (Järplid and Haltia 1993, Tyynelä *et al.* 2000). The resulting protein product was shown to be enzymatically inactive, but stable and normally processed (Tyynelä *et al.* 2000).

Another naturally occurring CTSD deficiency has recently been recognized in American Bulldogs (Evans *et al.* 2005, Awano *et al.* 2006a). The NCL-like disease, observed in young adult dogs (approximately one to three years old), is mild compared to the disease in sheep (Evans *et al.* 2005). The clinical features include hypermetria, ataxia, slowly progressing psychomotor deterioration, and premature death before the age of seven years (Evans *et al.* 2005). The autofluorescent storage material has an atypical ultrastructure for an NCL consisting of round uniformly staining inclusions embedded within granular matrixes (Evans *et al.* 2005, Awano *et al.* 2006a). The disease is inherited in an autosomal recessive manner (Evans *et al.* 2005), and was shown to be caused by a homozygous mutation in the canine *CTSD* gene (c.597G>A) resulting in a methionine to isoleucine change at position 199 (p.Met199Ile) (Awano *et al.* 2006a). The CTSD activity in the brains of the affected dogs was relatively high, 36% of that detected in control dogs (Awano *et al.* 2006a).

CTSD-deficient mice, generated by targeted disruption of the *Ctsd* gene, are normal at birth and for the first two weeks of their life (Saftig *et al.* 1995). After that, they develop progressive atrophy of the intestinal mucosa and a massive loss of lymphoid cells of the spleen and thymus, and die at the age of 25-27 days (Saftig *et al.* 1995). *Ctsd*-/- mice were recognized to have an NCL-like phenotype that includes seizures and blindness after ~20 days of age (Koike *et al.* 2000). The brains of these mice are moderately atrophied with particularly evident neuronal loss in cerebral cortex and thalamus (Haapanen *et al.* 2007, Partanen *et al.* 2008). Furthermore, the autofluorescent storage bodies detected in their tissues have characteristic features of GRODs and fingerprint profiles upon electron microscopical investigation, and show accumulation of subunit c of mitochondrial ATP synthase (Koike *et al.* 2000). CTSD activity is undetectable in the tissues of these mice (Saftig *et al.* 1995).

Another artificially generated animal model for CTSD deficiency was created by inactivating *cathD* in *D. melanogaster* (Myllykangas *et al.* 2005). The mutant flies develop normally, are viable and fertile, and have a normal lifespan. Yet they exhibit

an NCL-like phenotype since their tissues, especially the brains, progressively accumulate autofluorescent storage material that is granular in ultrastructural appearance and closely resembles GRODs. Moreover, they show modest age-related neurodegeneration in the central nervous system (Myllykangas *et al.* 2005).

3.3. Lysosomal membrane proteins

The membrane surrounding the lysosomes has many important functions conducted largely by associated proteins. It preserves the integrity of lysosomes and by sequestering the lysosomal hydrolases it protects the other cellular compartments from their harmful effects (Eskelinen *et al.* 2003). The lysosomal membrane has to permit traffic through it: the entry of the compounds that will be degraded and the exit of the digestion products (Verheijen and Mancini 2005). These are achieved by regulating the fusion and fission processes with other vacuoles, by specific transporters, and by passive diffusion. The membrane also maintains the acidity of the intralysosomal environment as vacuolar H^+ -ATPases pump protons into the lumen (Forgac 2007). Interactions with other cellular compartments and structures such as the cytoskeleton are also mediated by the membrane (Winchester 2001). Some of the lysosomal membrane proteins are briefly reviewed here as examples.

Lysosomal membrane transporter proteins are solute carriers, pumps, and channels, each of which has a high specificity for groups of amino acids, carbohydrates, nucleosides, inorganic ions, or vitamins (Verheijen and Mancini 2005). Of these various transporters, only a few are defined at the molecular level. Two amino acid transporters have been identified. The cystine transporter, cystinosin, is an integral membrane protein with a highly glycosylated large luminal part and seven predicted membrane-spanning domains that carries cystine out of the lysosome (Town *et al.* 1998, Kalatzis *et al.* 2001). Mutations in the respective gene underlie one of the LSDs, cystinosis (Town *et al.* 1998). Solute carrier family 36 member 1 (SLC36A1; previously known as lysosomal amino acid transporter 1, LYAAT-1) actively exports neutral amino acids from lysosomes (Sagne *et al.* 2001). A relatively well characterized example of monosaccharide transporters is sialin, an integral membrane protein containing 12 transmembrane domains, that transports sialic acid and other acidic monosaccharides out of the lysosomes (Verheijen *et al.* 1999, Morin *et al.* 2004). Sialin is defective in free sialic acid storage disorders, Salla disease and infantile sialic acid storage disease, that belong to the LSDs (Verheijen *et al.* 1999). Sialin is a member of the major facilitator superfamily (MFS) of transporter proteins. Three ATP-binding cassette (ABC) transporters, ABCA2, ABCA5, and ABCB9, have been shown to localize to lysosomes (Zhang *et al.* 2000, Vulevic *et al.* 2001, Kubo *et al.* 2005), and of these ABCA2 and ABCB9 are suggested to have roles in trafficking and/or metabolism of lipids (Davis *et al.* 2004, Sakai *et al.* 2007), and transport of

peptides (Wolters *et al.* 2005), respectively. CLN3, a lysosomal membrane protein defective in CLN3 disease, was reviewed in section 2.2.3.

Lysosomal vacuolar H^+ -ATPase, maintaining the acidic pH of the lysosomal lumen, is a membrane-associated complex consisting of 14 different subunits organized into V_1 and V_0 domains that hydrolyze ATP and translocate the protons across the membrane, respectively (Arai *et al.* 1993, Forgac 2007). Mutations in the T cell immune regulator 1 (*TCIRG1*) gene encoding one of the integral membrane subunits (a3) of vacuolar H+ -ATPase cause infantile malignant osteopetrosis (Frattini *et al.* 2000). In addition, defects in a lysosomal membrane protein CLCN7 underlie osteopetrosis (Kornak *et al.* 2001). An artificially generated mouse model deficient for this chloride channel displays additionally an NCL-like phenotype, as discussed in section 2.4.

The most abundant lysosomal membrane proteins are lysosomal-associated membrane proteins (LAMPs) and lysosomal integral membrane proteins (LIMPs) that represent more than 50% of the total number of membrane proteins of late endosomes and lysosomes (Eskelinen *et al.* 2003). LAMP1 and LAMP2 are highly homologous membrane proteins with single transmembrane, short cytoplasmic, and large N-terminal, highly glycosylated luminal domains (Chen *et al.* 1985, Lewis *et al.* 1985, Fambrough *et al.* 1988, Fukuda *et al.* 1988). LAMP2 has been shown to be involved in autophagy (Tanaka *et al.* 2000), in selective uptake of cytosolic proteins for degradation in lysosomes (Cuervo and Dice 1996), and in major histocompatibility complex (MHC) class II presentation of cytoplasmic antigens (Zhou *et al.* 2005). Mutations in *LAMP2* lead to an LSD, Danon disease (Nishino *et al.* 2000). The cellular function of LAMP1 remains unknown but it has been suggested to partially overlap the function of LAMP2 as mice deficient for either of these proteins are viable and fertile but double-knock-out mice show embryonic lethality (Eskelinen *et al.* 2004). Recently, it was shown that these proteins are essential in phagosome maturation (Huynh *et al.* 2007).

3.4. Lysosomal storage disorders

Defects in biogenesis and function of lysosomes cause a variety of diseases. More than forty of these, collectively referred to as lysosomal storage disorders, are currently known (Futerman and van Meer 2004). While the majority of these are caused by defects in lysosomal enzymes, defects in their activators, in transporters of the digestion products, or in proteins involved in endosomal/lysosomal vesicular trafficking can also be the underlying causes (Greiner-Tollersrud and Berg 2005). These defects lead to lysosomal accumulation of material, either directly of the unhydrolyzed substrate of a defective enzyme, or indirectly of some other material

due to lysosomal dysfunction. Mutations in the genes encoding lysosomal proteins may have various consequences, for example, in the case of soluble hydrolases they can affect the synthesis, folding, activity, or targeting of the mutant proteins (Vellodi 2005).

The variable effects of lysosomal dysfunction are reflected by the wide spectrum of clinical phenotypes of LSDs. However, some common characteristics apply to many of them. All LSDs are monogenic and most of them are autosomal recessively inherited (Vellodi 2005). Almost all of them are phenotypically heterogeneous but usually no obvious genotype-phenotype correlations can be observed (Futerman and van Meer 2004). LSDs most commonly affect infants and young children, and are progressive in nature (Jeyakumar *et al.* 2005, Vellodi 2005). They are usually multisystemic, and in most there is neurological involvement that manifests with symptoms including developmental delay, ataxia, seizures, blindness, abnormal ocular movements, spasticity, motor problems, and psychiatric disease (Jeyakumar *et al.* 2005, Vellodi 2005). As a group the prevalence of LSDs has been determined as ~1 per 8000 live births (Meikle *et al.* 1999, Poorthuis *et al.* 1999). LSDs are commonly classified according to the nature of either the defective protein or the accumulated substrate. Based mainly on the latter criteria they are classified into mucopolysaccharidoses, sphingolipidoses, oligosaccharidoses and glycoproteinoses, lipidoses, diseases caused by defects in membrane proteins, and other LSDs (Futerman and van Meer 2004). Some examples of LSDs were mentioned in the context of the review of lysosomal proteins (sections 3.2 and 3.3). The two diseases studied in this thesis, congenital NCL and Turkish vLINCL (CLN7), are classified as LSDs along with other NCLs.

The various biochemical and cellular pathways affected by lysosomal storage are relatively poorly characterized (Futerman and van Meer 2004). The defects in these pathways, in turn, are responsible for the diverse tissue pathology and symptoms of LSDs. While the existing cures for LSDs, such as bone marrow transplantation, enzyme replacement, and substrate reduction therapies, are currently few and rather inefficient, the delineation of these pathways may be crucial in developing new treatments for these diseases (Futerman and van Meer 2004).

AIMS OF THE STUDY

1. To characterize the molecular genetic background of congenital NCL, and
2. To characterize the molecular genetic background of Turkish vLINCL.

MATERIALS AND METHODS

1. Patients and controls

1.1. Patients with congenital NCL

Four patients with congenital NCL from two families were included in the study. The patients manifested with microcephaly, respiratory insufficiency, and epilepsy at birth, and died within hours to days after birth (Humphreys *et al.* 1985, Garborg *et al.* 1987). The storage material in the patient brain tissue samples showed granular ultrastructure (GRODs). In family A, that was consanguineous and of Pakistani descent, three of the seven children were affected. From two affected boys (patients 1 and 2), born in 1985 and 1989, respectively, only paraffin-embedded brain tissue samples were available, and from one affected boy (patient 3), born in 1999, and from their healthy father, only ethylenediamine tetra acetic acid (EDTA) -blood derived DNA was available. No samples from the other family members were available. The mother of family B was British whereas the origin of the father is not known. From the affected boy (patient 4), born in 1982, only a paraffin-embedded brain tissue sample was available. Samples from the parents were not available. Two of the patients (patients 1 (Garborg *et al.* 1987) and 4 (Humphreys *et al.* 1985)) had been described earlier.

1.2. Patients with variant late-infantile NCL

Altogether 16 patients with vLINCL from 13 families were included in the study (Table 3, Fig. 2). The disease onset was at the age of two to seven years and the clinical symptoms included seizures, psychomotor deterioration, visual failure, myoclonus, ataxia, and personality disorders (Topcu *et al.* 2004b). In electron microscopic examinations, either fingerprint profiles or curvilinear bodies were seen in the patient cells (lymphocytes, or cells present in skin or rectal biopsies). Eleven of the families were of Turkish origin, one (k) of Turkish-Georgian origin, and one (j) of Indian origin. In the beginning, all families were reported to be consanguineous but later, one of them (j) was reported to be non-consanguineous. In eight families, *CLN8* was previously excluded by the lack of homozygosity in haplotype analysis (Ranta *et al.* 2004). EDTA-blood derived DNA samples were available from 16 patients, 25 parents, and five unaffected siblings (indicated by coding in Fig. 2). In two families (family 4: patient 9124, and family e: individuals e1, e2, e4, and e5), fibroblast cells and

peripheral blood, respectively, were obtained for the analysis of splicing mutations in *CLN6* and *MFSD8* from mRNA.

Table 3. vLINCL patients included in this study.

Family (patient) code in II	Family (patient) code in III	Patient code in Topcu *et al*. 2004	Country of origin	Consan-guinity	Sample type	*CLN8* excluded in Ranta *et al*. 2004
1 (17, -)	e (e3, e5)	17, -	Turkey	yes	DNA, RNA (from blood) (e5)	yes
2 (29)	g (g3)	29	Turkey	yes	DNA	yes
3 (20, 20S)	-	20, -	Turkey	yes	DNA	yes
4 (9124)	-		Turkey	yes	DNA, RNA (from fibroblasts)	-
-	a (a3)	18	Turkey	yes	DNA	yes
-	b (b3)	24	Turkey	yes	DNA	yes
-	c (c3)	25	Turkey	yes	DNA	yes
-	d (d3, d4)	28, 27	Turkey	yes	DNA	yes
-	f (f3)	22	Turkey	yes	DNA	yes
-	h (h3)	-	Turkey	yes	DNA	-
-	j (j3)	-	India	no	DNA	-
-	k (k3)	-	Turkey/ Georgia	yes	DNA	-
-	l (l3)	-	Turkey	yes	DNA	-

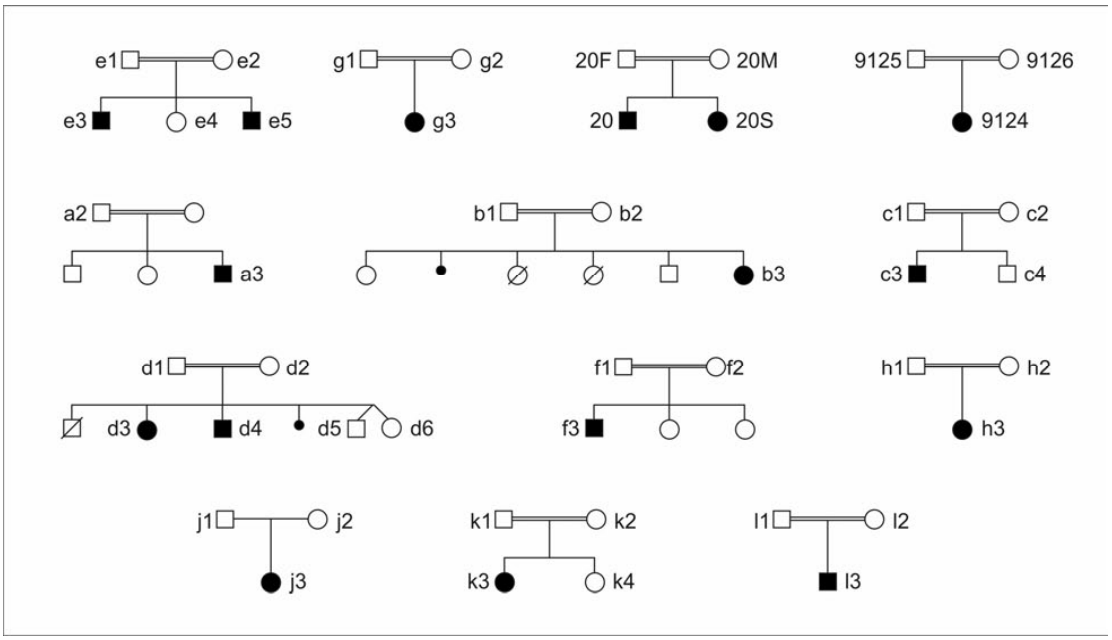

Figure 2. Families with vLINCL included in this study. The patients are indicated with filled symbols. All families are of Turkish origin except for family j which is of Indian origin and family k which is of Turkish-Georgian origin. The DNA samples were available from individuals indicated by coding.

1.3. Control samples

The controls for screening for the *CTSD* mutation consisted of 550 Caucasian (Centre d'Etude du Polymorphisme Humain (CEPH), and Finnish) chromosomes, for the *CLN6* mutations of 119 Turkish chromosomes, and for the *MFSD8* mutations of 212 Turkish and 92 CEPH chromosomes. The control RNAs for RT-PCR analyses were extracted from fibroblast cells and peripheral blood obtained from control individuals.

1.4. Ethical aspects

All studies were approved by an institutional review board of the Helsinki University Central Hospital. The samples were collected after informed consent was obtained.

2. Methods

The methods used in the original articles included in this study are summarized in Table 4.

Table 4. Methods used in this study. The original publications in which the methods were used are indicated with Roman numerals.

Method	Original publication
Agarose gel electrophoresis	I, II, III
Cell culture	I, III
Cathepsin D activity assay	I
DNA extraction	I, II, III
DNA sequencing	I, II, III
Database and computer analysis	I, II, III
Genomewide SNP scan	III
Haplotype analysis	II, III
Immunofluorescence microscopy	III
Immunofluorescence staining	III
Immunohistochemistry	I
In vitro translation	III
Microsatellite analysis	II, III
Mutation analysis	I, II, III
Northern blot analysis	III
Polymerase chain reaction (PCR)	I, II, III
Recombinant DNA techniques (cloning)	I, II, III
Reverse transcriptase PCR (RT-PCR)	II, III
RNA extraction	II, III
Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)	I, III
Site-directed mutagenesis	I, III
Transient transfections	I, III
Western blot analysis	I

RESULTS AND DISCUSSION

1. Cathepsin D deficiency in congenital human NCL (I)

Before this study, eight cases of congenital human NCL had been reported (Norman and Wood 1941, Brown *et al.* 1954, Sandbank 1968, Humphreys *et al.* 1985, Garborg *et al.* 1987, Barohn *et al.* 1992). The molecular genetic background of this very rare and aggressive disease, however, had remained undetermined. Previously, a mutation affecting one of the active site aspartic acids in the ovine *CTSD* gene had been shown to underlie congenital NCL in sheep (Tyynelä *et al.* 2000). On the basis of the close resemblance of the clinical phenotypes as well as neuropathological and ultrastructural findings between congenital NCLs in sheep and human, *CTSD* was considered as a potential candidate gene in human as well and its contribution to the disease pathogenesis of congenital human NCL was explored.

1.1. Mutations in the *CTSD* gene underlie congenital NCL

In order to screen *CTSD* for mutations, its nine exons and respective exon-intron boundaries were sequenced from the genomic DNA of a patient of Pakistani origin (patient 3) whereby a homozygous nucleotide duplication, c.764dupA, was identified in exon 6 (I, Fig. 1B). The father of the patient was a heterozygous carrier of the alteration, whereas a DNA sample was unavailable from the mother. The change was not found in control chromosomes. The alteration is likely to be a disease-causing mutation as it creates a premature stop codon (TAC>TAAC) at position 255 (p.Tyr255X). If a mutant protein, truncated by 158 amino acids, was produced and stable, it would lack the active site aspartic acid residue at position 295 in the CTSD polypeptide (Faust *et al.* 1985, Metcalf and Fusek 1993). Most likely, however, the abnormal mRNA is degraded by NMD since the mutation occurs far upstream of the last exon-exon junction (Maquat 2004). Further evidence for the degradation of the mutant CTSD either at the mRNA or protein level is provided by the lack of CTSD immunostaining in the brain samples from the two affected siblings of patient 3 (patients 1 and 2) (see section 1.3). From the old paraffin-embedded brain tissue samples from these two patients and from one unrelated patient (patient 4) we were not able to extract good quality DNA despite several attempts. Thus, the presence of *CTSD* mutations in these samples could not be confirmed. In the affected siblings of patient 3, however, the c.764dupA mutation is most likely present and underlies the disease but the nature of the mutation(s) in the unrelated patient 4 remains unknown.

In addition to the c.764dupA mutation, a nucleotide change c.845G>A (p.Gly282Arg) was identified in *CTSD* in homozygous form in patient 3 and in heterozygous form in his father. Since this alteration is located at the 3' side of the c.764dupA duplication, the changed amino acid is most probably not translated, and is unlikely to be the change underlying the disease. Moreover, the affected amino acid (p.Gly282) is not conserved among species.

1.2. Truncation and inactivation of mutant CTSD in BHK cells

In order to study the consequence of the c.764dupA (p.Tyr255X) mutation at the protein level *in vitro*, wild-type and c.764dupA mutant *CTSD* cDNAs were transiently expressed in baby hamster kidney (BHK) cells. As a control, an enzymatically inactive c.883G>A mutant, corresponding to the mutation causing congenital ovine NCL (Tyynelä *et al.* 2000, Partanen *et al.* 2003), was used. In western blot analysis, the p.Tyr255X mutant CTSD appeared truncated but stable, since it was detected as a single band of ~27 kDa corresponding to the calculated size of the prematurely truncated protein (I, Fig. 2B). On the contrary, wild-type CTSD appeared as two protein bands in the blots, reflecting the proteolytically processed form of the enzyme (~43 kDa, single chain active polypeptide, and ~31 kDa, mature heavy chain polypeptide) (Huang *et al.* 1979, Hasilik and Neufeld 1980b, Erickson *et al.* 1981). The mutant protein produced from the construct with the c.883G>A mutation was normally processed but had increased electrophoretic mobility, as reported earlier (Partanen *et al.* 2003).

In CTSD activity assays from cell lysates, the p.Tyr255X mutant CTSD, predicted to lack one of the active site aspartic acids, was inactive as expected, showing activity values similar to the non-transfected controls and cells transfected with the inactive c.883G>A mutant construct (I, Fig. 2A). The lack of CTSD activity will be discussed in more detail in section 1.5.

1.3. CTSD deficiency in the brains of patients with congenital NCL

To analyze the brains of patients with congenital NCL for the presence of CTSD, paraffin-embedded brain tissue samples from the two affected siblings of patient 3 (patients 1 and 2) were immunohistochemically stained with a CTSD antibody. In neurons and microglial cells no immunoreactivity for CTSD was detected (I, Fig. 5A). However, sometimes a weak, diffuse, and probably unspecific CTSD staining was detected in hypertrophic astrocytes (I, Fig. 5B). The sample from control brain showed a typical lysosomal CTSD staining pattern (I, Fig. 5C). These findings suggest the degradation of the p.Tyr255X mutant CTSD at mRNA or protein levels *in vivo*

despite the fact that *in vitro* the truncated mutant protein was shown to be stable (section 1.2). Further, the lack of CTSD staining was confirmed in the brain tissue sample of an unrelated patient (patient 4) with congenital NCL. On the basis of the absence of CTSD in these three patients as well as the nearly identical clinical and neuropathological findings in all patients, CTSD deficiency is suggested to cause the disease in all four patients.

1.4. Neuropathological findings in the brains of patients with congenital NCL

The neuropathological features of patients 1, 2, and 4 from which the brain samples were available were studied. For patients 1 and 4 some data had been reported earlier (Humphreys *et al.* 1985, Garborg *et al.* 1987). The normal structure of the cerebral and cerebellar cortices was destroyed, showing extensive neuronal loss. In the cerebrum, neurons were disorganized, while in the cerebellum Purkinje cells and inner granule cells were lost. Almost no axons or myelin were present in white matter. The glial activation in the brains of the patients was detected by immunohistochemical staining. Hypertrophic astrocytes, abundant throughout the brain tissue, and especially in white matter, contained dense cores staining intensely with glial fibrillary acidic protein (GFAP) antibody (I, Fig. 3A-C), whereas activated microglia, present especially in the deeper layers of grey matter of cortices, were strongly immunoreactive for CD68 (I, Fig. 3D-F). These findings indicate that the neuropathological changes were extremely severe and uniform in all of these patients.

The storage deposits detected in the patient tissues stained with an antibody for saposin D (I, Fig. 4A), but not for subunit c of mitochondrial ATP synthase (I, Fig. 4B). This feature divides human NCL forms into two categories: the group where the major protein component of the storage bodies is subunit c of mitochondrial ATP synthase (most forms of NCLs, including CLN2, CLN3, CLN5, CLN6, CLN8, and CLN9 (Hall *et al.* 1991, Palmer *et al.* 1992, Elleder *et al.* 1997, Tyynelä *et al.* 1997, Herva *et al.* 2000, Schulz *et al.* 2004)), and the group where it is saposin A and/or D (CLN1 and Parry disease (Tyynelä *et al.* 1993, Nijssen *et al.* 2003)). The present results indicate that congenital human NCL is a member of the latter group. Of CTSD-deficient animals, the composition of the storage material has been studied in sheep and mice. In sheep, the major protein components of storage material are saposins A and D (Tyynelä *et al.* 2000) but, interestingly, in mice accumulation of subunit c of mitochondrial ATP synthase has been reported (Koike *et al.* 2000). This variation, however, may be due to the detection method used (Tyynelä J, personal communication).

1.5. CTSD deficiency – a novel form of NCL, CLN10

Our results provide the first molecular genetic explanation for congenital human NCL. While in one patient the disease-causing mutation was demonstrated by sequencing from genomic DNA, the lack of CTSD staining in tissue samples from his two affected siblings indicates that the same mutation is the most likely underlying cause for their disease as well. In the unrelated patient, the absence of CTSD staining implies that also here CTSD deficiency underlies the disease, but the nature of the disease-causing mutation(s) remains unknown. Based on nearly identical phenotypes and pathological features of these patients and the other reported cases of congenital NCL, CTSD deficiency may underlie all cases of congenital NCL, and should be considered as a possible diagnosis in microcephalic newborns suffering from seizures at birth.

In tandem with this study, a novel, autosomal recessively inherited NCL-like disorder with CTSD deficiency was described in one German patient (Steinfeld *et al.* 2006). The patient presented with visual disturbances and ataxia at early school-age which were followed by progressive psychomotor decline. Atrophy of cerebrum and cerebellum were observed. In electron microscopic analysis heterogeneously appearing GRODs were identified in Schwann cells derived from skin biopsy material. In the *CTSD* gene compound heterozygosity of two missense mutations, c.685T>A (p.Phe229Ile), and c.1149G>C (p.Trp383Cys), was identified. These mutations were shown to lead to markedly reduced CTSD activity (6.7% of that detected in controls), and decreased amounts of CTSD in patient fibroblasts. In overexpression systems, maturation of the p.Phe229Ile mutant was slightly delayed whereas processing and intracellular trafficking of the p.Trp383Cys mutant was severely disturbed (Steinfeld *et al.* 2006).

These two human CTSD deficiencies, the later-onset NCL described by Steinfeld *et al*. (2006) and congenital NCL described by us, have subsequently been combined under the disease entity of CLN10 (Online Mendelian Inheritance in Man (OMIM) database, http://www.ncbi.nlm.nih.gov/entrez/dispomim.cgi?id=610127). Despite having defects in the same gene (Table 5), the diseases are different in many respects. While congenital NCL manifests at birth, the later-onset NCL with CTSD deficiency has a more late-infantile or juvenile-onset range. Moreover, differences in the severity and progression of the diseases are great, as congenital NCL is very aggressive and leads to death within hours to weeks, while the patient with later-onset NCL with CTSD deficiency suffers from milder symptoms and has survived to at least the age of 17 years (Steinfeld *et al.* 2006).

The difference in the severity of the CTSD deficiency disorders may be due to the different levels of residual CTSD activity. CTSD activity is completely absent in severe CTSD deficiency phenotypes: in congenital human and ovine NCLs (I and (Tyynelä *et*

al. 2000)) as well as in severe NCL-like phenotype of *Ctsd*-/- mice (Koike *et al.* 2000). Significant residual CTSD activity, reflecting partial CTSD inactivation by missense mutations in the patient with later-onset human NCL (Steinfeld *et al.* 2006) and in American bulldogs with NCL (Awano *et al.* 2006a), is associated with relatively mild phenotypes. This implies that CTSD activity is critical for early survival and development in these vertebrates.

Table 5. All *CTSD* mutations identified in human patients to date.

Nucleotide change	Amino acid change or predicted consequence	Exon	Origin	Phenotype	Reference
c.685T>A	p.Phe229Ile	Exon 5	Germany	Later-onset NCL-like disorder	Steinfeld *et al*. 2006
c.764dupA	p.Tyr255X	Exon 6	Pakistan	Congenital NCL	I
c.1149G>C	p.Trp383Cys	Exon 9	Germany	Later-onset NCL-like disorder	Steinfeld *et al*. 2006

Although different in many respects, there are also similarities between the two human forms of CTSD deficiency. In both diseases the storage material in patient cells showed granular ultrastructure (GRODs) (Humphreys *et al.* 1985, Garborg *et al.* 1987, Steinfeld *et al.* 2006). GRODs or GROD-like structures are also observed in other CTSD deficiencies: in sheep (Järplid and Haltia 1993), in mice (Koike *et al.* 2000), and in *D. melanogaster* (Myllykangas *et al.* 2005). Conversely, in American bulldogs deficient for CTSD the storage material has an atypical ultrastructure with round uniformly staining inclusions embedded within granular matrixes (Evans *et al.* 2005, Awano *et al.* 2006a). Thus, while there is some variation in the ultrastructure of the storage material between different CTSD deficiencies, most seem to resemble GRODs.

In addition to the presence of saposin A and/or D as the major protein component of the storage material in both congenital NCL and CLN1, the presence of GRODs relates CLN10 to CLN1 with a similar ultrastructure of the storage material (Haltia *et al.* 1973, Humphreys *et al.* 1985, Garborg *et al.* 1987, Das *et al.* 1998, Mitchison *et al.* 1998, van Diggelen *et al.* 2001, Steinfeld *et al.* 2006). This has implications for the diagnostics of patients with NCL phenotypes with GRODs. Since both *CTSD* and *CLN1* mutations have been shown to underlie NCLs with a wide range of onset ages and variable disease progressions ((Vesa *et al.* 1995, Das *et al.* 1998, Mitchison *et al.* 1998, van Diggelen *et al.* 2001, Steinfeld *et al.* 2006) and I), both genes should be considered as diagnostic alternatives in patients of all ages with NCL phenotypes with GRODs.

2. Molecular genetic background of Turkish vLINCL (II, III, unpublished)

Variant late-infantile NCL in Turkish patients was initially considered a distinct clinical and genetic entity, CLN7 (Wheeler *et al.* 1999). However, mutations in the *CLN8* gene were later reported to cause the disease in a subset of Turkish patients with vLINCL (Ranta *et al.* 2004). Subsequently, a number of patients were still lacking a molecular genetic explanation for their disease. To further dissect the molecular genetic background of vLINCL in the remaining Turkish patients, a candidate gene approach was first undertaken to explore the contribution of the previously known NCL genes to the disease pathogenesis, and then a genomewide homozygosity mapping approach was applied with the aim to identify novel gene(s) underlying the disease.

2.1. Analysis of known NCL genes in Turkish patients with vLINCL

2.1.1. Exclusion of known NCL genes in Turkish families with vLINCL (II, III)

Since it is clear that defects in most of the known NCL genes cause variable phenotypes (Mole *et al.* 2005), all human and animal genes known to underlie NCLs or NCL-like phenotypes were considered as candidate genes for Turkish vLINCL. All these NCL gene loci (*CLN1/PPT1, CLN2/TPP1, CLN3, CLN5, CLN6, CLN8, CTSD, CLCN3,* and *CLCN7*) were genotyped using three or four fluorescently labelled microsatellite markers flanking each locus in altogether 11 Turkish, one Turkish-Georgian, and one Indian family with vLINCL. Because at that time these families were all reported to be consanguineous, the haplotypes of marker alleles over these loci were analyzed for homozygosity, as the genomic region over a locus harbouring a defective gene is anticipated to be homozygous by descent in consanguineous families (Lander and Botstein 1987).

All above mentioned NCL loci were excluded in nine families based on the lack of homozygosity. Three or four adjacent markers flanking the *CLN3* and *CLN6* loci showed allele homozygosity in patients from two families each (families 1=e and 2=g for *CLN3*, and families 3 and 4 for *CLN6*) (II, Fig. 1). Haplotypes were different over the respective loci in each family. In addition, one of the patients homozygous for marker alleles over *CLN6* (patient 9124 in family 4) showed homozygosity also over *CLN1/PPT1*. However, since *PPT1* had previously been excluded by PPT1 enzymatic analysis, sequencing of the *PPT1* gene was not carried out. The identification of

homozygous regions by chance in a child of a consanguineous family is not unexpected, since for example in a child of first cousins homozygosity by descent can be anticipated at ~6% of all loci (Lander and Botstein 1987). When sequenced from genomic DNA of patients homozygous over *CLN3* (e3 and g3) no mutations were identified in the 15 exons or the exon-intron boundaries of *CLN3,* excluding it with a high probability as the underlying gene. Instead, in two families (families 3 and 4) two novel homozygous mutations in the *CLN6* gene were identified (section 2.1.2).

The families in which all NCL loci were excluded either by the lack of homozygosity in haplotype analysis alone (nine families: a, b, c, d, f, h, j, k, and l) or in combination with genomic sequencing of the *CLN3* gene (families e and g) were hypothesized to represent a distinct genetic entity, the "true" Turkish vLINCL (CLN7), the genetic background of which, at that time, remained to be defined. We continued the study of these 11 families by a genomewide SNP scan and homozygosity mapping in order to identify the underlying locus and gene, *CLN7* (section 2.2).

2.1.2. Identification of two novel CLN6 *mutations in Turkish patients with vLINCL (II)*

The *CLN6* gene was screened for mutations by sequencing its seven exons and the respective exon-intron boundaries from the genomic DNA of three patients (patients 20 and 20S in family 3, and patient 9124 in family 4) that were found to be homozygous for marker alleles over the *CLN6* locus. Two novel homozygous sequence alterations were identified (II, Fig. 2) that are likely to be disease-causing mutations, as they predict altered structure of the CLN6 protein. A C-to-G transition (c.663C>G) in exon 6 was identified in two affected siblings (20 and 20S) in family 3, whereas a G-to-T transversion (c.542+5G>T) in the fifth nucleotide of intron 5 was identified in the only patient (9124) in family 4. Both mutations cosegregated with the disease phenotype in the respective families, and moreover, were not found in control chromosomes, indicating that they are not likely to be polymorphisms.

The c.663C>G alteration is a nonsense mutation that creates a premature stop codon at tyrosine 221 (p.Tyr221X) in the CLN6 polypeptide and predicts a truncation of the protein by 91 amino acids. The resulting abnormal protein is likely to be non-functional or degraded. Since the mutation occurs only three nucleotides from the 3' end of the penultimate exon 6 and thus not more than 50 nucleotides upstream from the last exon-exon junction, degradation of mRNA by NMD is less likely (Maquat 2004). Truncation of CLN6 might affect the subcellular localization as well as the homodimerization of the protein since these have been reported to be influenced by the C-terminal part of the protein (Heine *et al.* 2007). Tyrosine 221, a highly conserved amino acid predicted to lie within a transmembrane segment, has earlier

been reported to be mutated (c.662A>C, p.Tyr221Ser) in one allele of a patient originating from Argentina (Sharp *et al.* 2003).

The c.542+5G>T alteration affecting the donor splice site sequence of intron 5 was shown to alter the splicing of the mRNA by RT-PCR analysis from RNA extracted from fibroblasts of patient 9124 (II, Fig. 3). No fragments of the expected size (303 base pairs, bp), corresponding to the correctly spliced *CLN6* transcript, were observed. Instead, five other fragments, with sizes ranging from 250 to 400 bp corresponding to abnormal splice products, were detected. Sequence analysis was technically possible from three of the fragments revealing skipping of exon 5 in all and inclusion of intron 4 sequences in two of the fragments. In the two latter fragments this indicates the use of cryptic donor splice sites (c.486+26 and c.486+119) in intron 4 in combination with the acceptor splice site of intron 5. At the amino acid level, the mutation predicts the absence of a functional CLN6 protein since two of the mRNA variants contain frameshifts and premature stop codons while one encodes a product with an in-frame replacement of exon 5 encoded amino acids with nine new amino acids encoded from intron 4 sequences. The altered splicing resulting from nucleotide changes at splice sites is not unexpected, as mutations affecting splice sites may either inactivate them, or render them less effective allowing the splicing machinery to use other, naturally weaker splice sites (Cartegni *et al.* 2002).

The current number of *CLN6* mutations is 27 (NCL Mutation Database, www.ucl.ac.uk/ncl/mutation). The two mutations identified in this study were the first and only ones identified in the *CLN6* gene in patients of Turkish origin thus far. This study further supports the hypothesis that *CLN6* is a highly mutable gene, based on the occurrence of mutations in several small inbred populations or families around the world, e.g. in the Mediterranean region, in India, and in Costa Rica (Gao *et al.* 2002, Wheeler *et al.* 2002, Sharp *et al.* 2003, Teixeira *et al.* 2003). The clinical phenotypes of the vLINCL patients with CLN6 defects are indistinguishable between patients of Turkish and other origins.

2.2. Identification and characterization of the novel *MFSD8* gene underlying vLINCL (III, unpublished)

2.2.1. Identification of the CLN7 *locus*

All families with vLINCL included in this study were in the beginning of the study reported to be consanguineous, and therefore a SNP-based homozygosity mapping approach was utilized in the search for novel gene(s) underlying vLINCL. The idea behind this strategy is to search for the chromosomal region with homozygous marker

alleles shared among patients in as many families as possible (Lander and Botstein 1987). This segment would, with a high probability, be inherited from a common ancestor of the parents in each family, and therefore be identical by descent, and possibly contain the disease-causing gene.

The genomewide SNP scan was performed in patients from eight Turkish (a, b, c, d, e=1, f, g=2, and l), one Turkish-Georgian (k), and one Indian (j) families with vLINCL in which all known NCL loci had previously been excluded (section 2.1.1). The homozygosity mapping revealed three regions with HLOD scores over 2. The strongest evidence for linkage with a statistically significant HLOD score of 3.39 was observed on chromosome 4 at SNP *rs348085* (Fig. 3) where six families (a, b, c, e, f, and j) contributed to this value. The other two linkage peaks were on chromosomes 8 and 15, where four families at both loci (b, c, e, and l, and d, e, g, and l) contributed to the HLOD scores of 2.05 and 2.48, respectively.

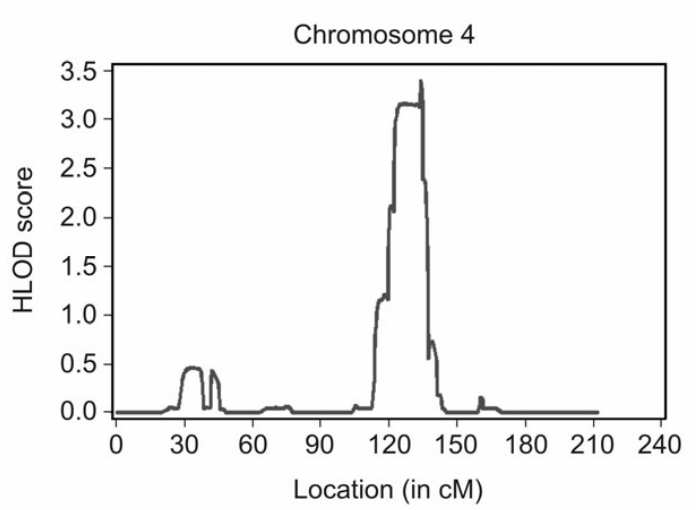

Figure 3. Results of the linkage analysis on chromosome 4. Multipoint linkage analysis by the computer program Merlin in 10 families with vLINCL gave the maximum HLOD score of 3.39 at 134.18 cM, at SNP *rs348085*.

Over the best candidate locus on chromosome 4q26-q28.3, patients in all except one of the six families shared a ~19.5 megabase (Mb) region (from *rs7657655* to *rs10518621*) of homozygous marker alleles (Fig. 4). In family a, the homozygous region spanned less than 1 Mb, and was considered too short to be homozygous by descent, considering the close consanguinity in the family, and therefore unlikely to harbour the disease-causing gene in the respective family. Analysis of microsatellite

markers (*D4S427, D4S2975, D4S2938,* and *D4S429*) in this ~19.5 Mb candidate region in families b, c, e, f, and j did not narrow down the region. Analysis of an additional patient (h3) not included in the genomewide SNP scan also revealed homozygosity of marker alleles over the candidate region but did not restrict it either (Fig. 4). The identification of a large homozygous region shared in the majority (six out of eleven) of the families was consistent with our hypothesis that the chromosomal segment harbouring the disease gene would be identical by descent in most of the families. The haplotypes over this region were different in all of these six families (b, c, e, f, h, and j) suggesting the existence of family-specific mutations in the disease-causing gene.

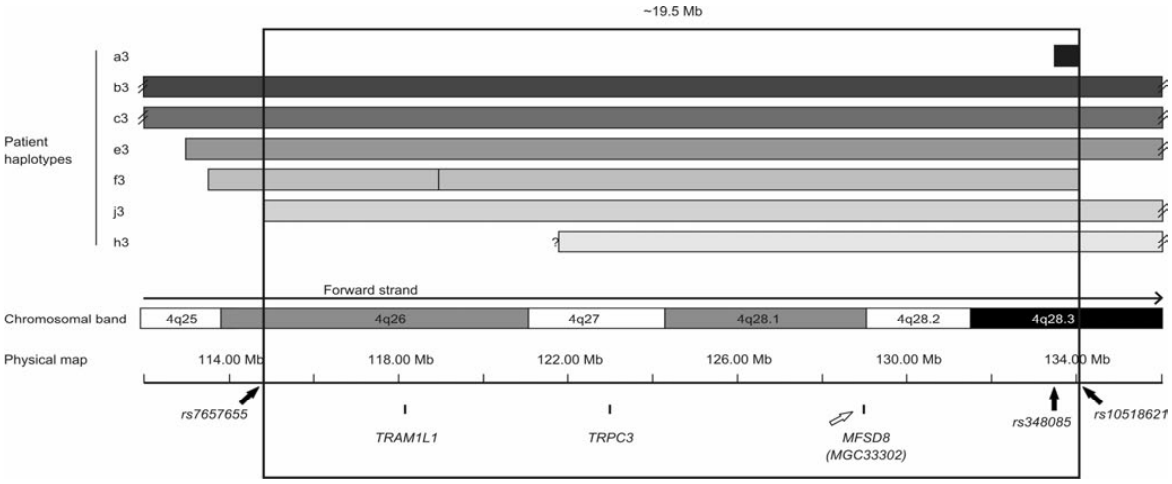

Figure 4. The *CLN7* candidate region on chromosome 4q26-q28.3. The critical region of ~19.5 Mb is shown within the box. The position of the region is shown both in relation to chromosomal bands and on the physical map. The homozygosity of marker alleles in each patient (a3, b3, c3, e3, f3, j3, and h3) is shown by differently shaded gray bars, indicating the different haplotypes detected in these patients. The diagonal lines at the ends of the bars indicate that the respective homozygous haplotypes extend beyond the region shown in the figure. In patient f3 the vertical line marks one heterozygous SNP within the otherwise homozygous haplotype. Patient h3 was genotyped only for microsatellite markers (*D4S427, D4S2975, D4S2938,* and *D4S429*), and thus the analysis did not cover the whole region and the lack of information is indicated by a question mark. The sequenced candidate genes, *TRAM1L1, TRPC3*, and *MFSD8 (MGC33302)*, shown below the physical map with black bars are not to scale. *MFSD8* is indicated with a white arrow, and the SNPs (*rs7657655* and *rs10518621*) restricting the candidate region and the SNP (*rs348085*) with the highest HLOD score in the linkage analysis with black arrows. The picture has been modified from the Ensembl genome browser view of the region.

2.2.2. Identification of mutations in MFSD8

Within the best candidate region spanning ~19.5 Mb on chromosome 4q26-q28.3 (Fig. 4), more than 90 known or putative genes were identified from the NCBI database. The positional candidate genes were chosen from these based on the known or predicted functions of the encoded proteins and sequenced from genomic DNA of patients showing homozygosity over the region (in families b, c, e, f, h, and j). After the exclusion of two genes (translocation associated membrane protein 1 - like 1, *TRAM1L1*, and transient receptor potential cation channel, subfamily C, member 3, *TRPC3*), six nucleotide changes were identified in these six families in the hypothetical gene *MGC33302* (GenBank accession number NM_152778) by screening its 12 protein-coding exons (exons 2-13) (Fig. 5; III, Table 1 and Fig. 1A). All these sequence variants were homozygous and family-specific, as was expected on the basis of the haplotype analysis (section 2.2.1). Furthermore, all identified alterations cosegregated with the disease phenotype in the respective families. Subsequently, this gene was named as major facilitator superfamily domain containing 8 (*MFSD8*). However, as the gene identification was mainly based on Turkish families, the corresponding locus is denoted as *CLN7*, the symbol initially assigned for Turkish vLINCL.

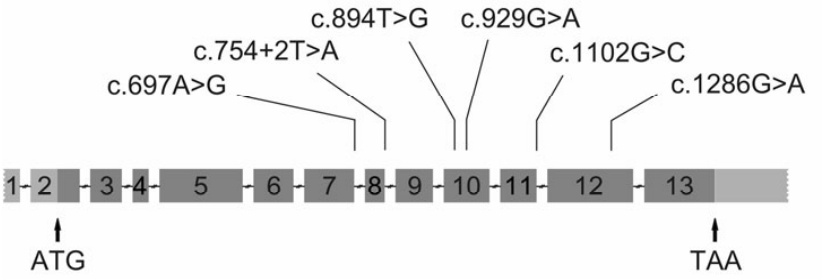

Figure 5. Genomic structure of the *MFSD8* gene and positions of vLINCL-associated mutations. Within the 13 exons (boxes 1–13) the coding regions are shown in darker gray while the untranslated regions are shown in lighter gray. Introns are not to scale. Positions of the vLINCL-associated mutations are indicated with lines above the gene and the initiation (ATG) and stop (TAA) codons are indicated by arrows below the gene.

All identified nucleotide changes are predicted to be disease-causing mutations, since they change the predicted amino acid composition of the MFSD8 protein, and moreover, were not found in control chromosomes. In patient j3, a nonsense mutation, c.894T>G, was identified in exon 10. It creates a premature stop codon at position 298 (p.Tyr298X) predicting a protein truncated by 221 amino acids which, if translated, would probably be non-functional or degraded. More likely, however, is that the mRNA is degraded by NMD since the mutation occurs far upstream from the last exon-exon junction (Maquat 2004). In two patients (c3 and b3) missense mutations, c.929G>A (p.Gly310Asp) and c.1286G>A (p.Gly429Asp), were identified in exons 10 and 12, respectively. The affected amino acids are conserved across vertebrates, and according to the hydrophobicity predictions (section 2.2.5), lie within the 8th and 10th transmembrane domains of MFSD8, respectively (III, Figs. 2 and 4). These changes, when introduced into the hemagglutinin (HA) tag and wild-type *MFSD8* containing construct, did not interfere with the lysosomal localization of the HA-MFSD8 protein (section 2.2.7) (III, Fig. 5J-O). This suggests a defect in MFSD8 function rather than in its trafficking. Nucleotide changes (c.697A>G and c.1102G>C) identified in two patients (f3 and h3, respectively) are missense mutations (p.Arg233Gly and p.Asp368His, respectively) affecting highly conserved amino acids, and may thus result in changes in functional and/or structural properties of the protein. Alternatively they may alter the splicing of the transcript, since the affected nucleotides reside at the exon-intron boundaries in the 3' ends of exons 7 and 11, respectively, and lead to production of abnormal mRNAs and/or proteins. Because samples suitable for RNA extraction were not available from these two patients, the consequences of these mutations could not be assessed at the mRNA level.

The only intronic *MFSD8* mutation detected was c.754+2T>A at the donor splice site of intron 8, identified in two affected children in family e (patients e3 and e5). RT-PCR analysis from patient (e5) lymphoblastoid RNA with primers from exons 6 and 11 revealed an altered pattern of splicing products (III, Fig. 1B). The fragment corresponding to the normal transcript (~550 bp), containing exons 7 to 10, was almost completely absent. There was also a total lack of an alternatively spliced variant lacking exon 7 (~400 bp), and increased expression of two alternatively spliced variants, one without exon 8 (~500 bp) and one without both exons 7 and 8 (~350 bp). An additional weak band of ~480 bp, the sequence of which remained unknown, was detected. The first two variants are predicted to contain frameshifts and premature stop codons while the one lacking both exons 7 and 8 is predicted to encode a product with an in-frame deletion of amino acids encoded from the respective exons. Nonetheless, the c.754+2T>A sequence variant is likely to be disease-causing since it leads to almost complete lack of the full-length, in-frame transcript. The detection of an altered splicing pattern was not unexpected since nucleotide changes at the nearly invariant second thymine at the 5' end of an intron

usually lead to imprecise recognition of exon-intron junctions and to disturbed splicing (Cartegni *et al.* 2002).

As anticipated, no mutations were identified in the *MFSD8* gene in patients lacking homozygosity over the *CLN7* locus (in families a, d, g, l, and k). In addition to the exclusion of all known NCL loci in these families (section 2.1.1), they were, by analysis of the genomewide SNP data, excluded for candidate genes *CTSF* and *CLCN6*, homologs of which are defective in mice with NCL-like phenotypes (Poet *et al.* 2006, Tang *et al.* 2006). These results indicate that additional, novel NCL-causing gene(s) underlie the disease in these families. In fact, when the genomewide scan data were reanalysed in these families, two extensive regions of homozygosity were identified: on chromosomes 15 (11.7 Mb shared in families g, and l), and 22 (10.8 Mb in family d). The highest HLOD score calculated in these five families at chromosome 15 was 2.57 (at SNP *rs10519740*) while at chromosome 22 it was 1.04 (at SNP *rs133710*). However, for family d only, the highest LOD score at chromosome 22 was 2.05 (at SNPs *rs2746967* and *rs1534882*). Since not all families showed homozygosity over these two regions, the results imply the possible presence of at least three disease-causing genes in these five families. Genotyping of additional families with vLINCL in these candidate regions as well as on a genomewide scale followed by linkage analysis and/or homozygosity mapping will pave the way for the identification of novel genes underlying vLINCL.

2.2.3. MFSD8 *as a novel NCL gene*

Although Turkish vLINCL was initially considered a distinct clinical and genetic entity (Wheeler *et al.* 1999), recent studies ((Ranta *et al.* 2004), II, III) indicate that Turkish vLINCL, despite the relative homogeneity of the phenotype, is genetically very heterogeneous. Mutations have already been identified in three genes, *CLN6*, *CLN8*, and *MFSD8* ((Ranta *et al.* 2004), II, III). The existence of additional genes underlying vLINCL in Turkish patients indicated by this study further corroborates the genetic heterogeneity of Turkish vLINCL. This should be taken into consideration when diagnosing vLINCL patients of Turkish descent.

Further studies have led to the identification of several *MFSD8* mutations in patients of various ethnic origins (unpublished observations by the group). This implies that although the identification of *MFSD8* was based mainly on Turkish patients, CLN7 is not confined to the Turkish population, and moreover, *MFSD8* seems to be a relatively common gene underlying vLINCL in other populations as well. Mapping and identification of the disease gene in a certain population followed by the detection of mutations in other populations is not an uncommon phenomenon within the NCL field. The *CLN5* and *CLN8* genes were initially mapped and identified in

patients originating mainly (*CLN5*) or entirely (*CLN8*) from the Finnish population (Savukoski *et al.* 1994, Tahvanainen *et al.* 1994, Savukoski *et al.* 1998, Ranta *et al.* 1999). Later, however, mutations in both genes have also been detected in patients from other populations (Ranta *et al.* 2004, Pineda-Trujillo *et al.* 2005, Bessa *et al.* 2006, Cannelli *et al.* 2006, Cannelli *et al.* 2007, Zelnik *et al.* 2007), thus proving that these diseases are not restricted to the Finnish population.

Within the NCLs, the late-infantile onset group is genetically the most heterogeneous with mutations identified in *PPT1, TPP1, CLN5, CLN6, MFSD8*, and *CLN8* ((Sleat *et al.* 1997, Das *et al.* 1998, Savukoski *et al.* 1998, Gao *et al.* 2002, Wheeler *et al.* 2002, Ranta *et al.* 2004, Cannelli *et al.* 2006), III). Differential diagnosis, especially in patients with vLINCL, is often difficult due to the similar phenotypes and ultrastructural features within this group. Moreover, the defects in certain NCL-causing genes are not confined to specific populations (NCL Mutation Database, www.ucl.ac.uk/ncl/mutation), and thus the disease-causing gene cannot be deduced from the ethnic origin of the patient. For these reasons, the ability to provide a molecular genetic diagnosis for many patients nowadays is of particularly great value. However, the diagnostics remain challenging because of the great locus and allelic heterogeneity within LINCLs.

The mapping and subsequent identification of the *MFSD8* gene in this study is a good example of a successful application of the homozygosity mapping approach (Lander and Botstein 1987). With our relatively limited family material it was possible to identify the disease-causing locus and gene by utilizing consanguineous families. Since consanguineous marriages are relatively common in the Turkish population (Bittles 2001), there are additional examples of the identification of disease genes by the homozygosity mapping approach in consanguineous Turkish families (Topcu *et al.* 2004a, Collin *et al.* 2008). Further, mapping genes in patients originating from Turkey and other countries with high consanguinity rates will most probably be beneficial in identifying additional human disease genes, especially in rare inherited diseases. In respect to the family j that was reported to be non-consanguineous, the presence of an extensive region of homozygosity in the affected child's genome suggests that the parents may yet share a common ancestor. Finally, although the genetic background in our family material turned out to be heterogeneous, the defective gene was shared in the majority of the families which was sufficient to facilitate the identification of the locus and the gene. Since the genetic background of the genetically undefined vLINCL forms is, based on this study and on other, unpublished observations, heterogeneous and moreover, the family material is often small, the homozygosity mapping approach utilizing consanguineous families may be ideal in attempts to identify the novel disease-causing genes in these NCL forms as well.

2.2.4. MFSD8 *mRNA expression and alternatively spliced variants*

To initially characterize the function of the *MFSD8* gene, its expression was first analyzed by northern blot, *in silico*, and RT-PCR analyses. The ubiquitous expression of *MFSD8* was supported both by northern blot analyses from human tissues and *in silico* by expressed sequence tag (EST) database searches. In human tissue northern blot analyses the main transcript of ~5 kb was detected to be expressed at very low levels in all tissues analyzed (III, Fig. 3). In all regions of the brain and in lung, it was the only transcript detected. In addition, several transcripts ranging from ~1 to ~3 kb were seen in heart, placenta, liver, skeletal muscle, kidney, and pancreas. In heart and skeletal muscle, transcripts of ~6 kb were also detected.

MFSD8 shows a complex pattern of alternative splicing, as suggested by the EST and RT-PCR data. The majority of the ESTs identified in the databases included all exons 1-13. Some ESTs, however, represented four alternatively spliced variants with more limited tissue distribution: one lacking exon 2 (BI553701), one lacking exon 7 (BX341359), one lacking exons 7 and 8 (BG434527), and one lacking exon 11 (represented by three ESTs: BX341358, BG542082, and BU618424). In addition to these, several other alternatively spliced partial variants were identified by RT-PCR analyses covering the open reading frame of the gene (Table 6). None of the variants are predicted to produce full-length, in-frame transcripts. Observing the alternatively spliced variants of *MFSD8* mRNA was not unexpected since the majority of human genes are estimated to be alternatively spliced (Ben-Dov *et al.* 2008). However, the functional relevance of these variants as well as the variants seen in northern blot analyses is not clear. Alternative splicing has been suggested to be important in generating protein diversity as well as in regulating gene expression (Stamm *et al.* 2005). It may result in the production of proteins with different amino acid compositions, protein isoforms, that may or may not be biologically relevant, or of mRNAs with different properties important, for example, for their stability or translation. Alternatively, it may result in the production of mRNAs with premature stop codons that serve as targets for NMD and thus lead to their elimination (Lewis *et al.* 2003, Hillman *et al.* 2004).

Table 6. Alternatively spliced variants of *MFSD8* mRNA detected in this study.

Partial transcript detected by RT-PCR analyses	Predicted consequence (difference to NM_152778)	Corresponding EST
Lacking exon 2	Translation initiation codon at p.Met46	BI553701
Lacking exon 6	In-frame deletion of 38 aa	-
Lacking exons 6 and 7	Frameshift and premature stop codon	-
Lacking exons 6, 7, and 8	In-frame deletion of 105 aa	-
Lacking exon 7	Frameshift and premature stop codon	BX341359
Lacking exons 7 and 8	In-frame deletion of 67 aa	BG434527
Lacking exons 7, 8, and 11	In-frame deletion of 67 aa (by lack of exons 7 and 8), frameshift and premature stop codon (by lack of exon 11)	-
Lacking exons 7 and 11	Frameshift and premature stop codon (by lack of exon 7)	-
Lacking exon 8	Frameshift and premature stop codon	-
Lacking exon 11	Frameshift and premature stop codon	BG542082, BU618424, BX341358

Abbreviations: RT-PCR = reverse transcriptase polymerase chain reaction, aa = amino acid(s), EST = expressed sequence tag

2.2.5. MFSD8 protein

The 1554 bp open reading frame of *MFSD8* encodes a 518 amino acid protein with a predicted molecular weight of ~58 kDa. In agreement with this was the detection of the HA-tagged MFSD8 proteins as ~60 kDa bands in sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) following an *in vitro* translation assay. Several attempts to detect the soluble overexpressed HA-MFSD8 in transfected cell lysates on western blots with HA antibodies were unsuccessful, presumably due to the hydrophobicity of the protein, problems in its transfer, or proteolytic degradation of the HA tag.

MFSD8 is predicted to be a polytopic integral membrane protein with 12 transmembrane domains (III, Fig. 4). It contains a MFS domain (MFS_1) at amino acid positions p.42_477, and a sugar (and other) transporter domain (Sugar_tr) at amino acid positions p.72_147, as detected in the Pfam analysis of the MFSD8 amino acid sequence. This suggests that MFSD8 is a novel member of the major facilitator superfamily, a very large family of secondary active transporters, present ubiquitously from bacteria to eukaryotes (Pao *et al.* 1998, Lemieux 2007). Proteins belonging to MFS are single-polypeptide carriers for various small solutes, including sugars, drugs, inorganic and organic cations, and various metabolites (Pao *et al.* 1998). The structural architecture of MFS proteins is suggested to be similar, and like MFSD8,

most are predicted to have 12 membrane-spanning domains (Lemieux 2007). The MFS proteins are divided into phylogenetic subfamilies, members of which are predicted to have similar functions and substrate specificity (Pao *et al.* 1998). MFSD8, as a member of MFS, is likely to function as a transporter, but its substrate specificity and exact cellular function are currently unknown.

2.2.6. MFSD8 homologs

MFSD8 is evolutionary conserved with several homologs in different species identified by BLAST searches. In vertebrates, MFSD8 has single homologs in each species, whereas in invertebrates, usually several weakly similar proteins are identifiable. In *S. cerevisiae*, the most similar proteins to MFSD8 whose functions are known are MFS proteins involved in the transport of lactate (Jen1p), and glycerophosphoinositol and -choline (Git1p) (Patton-Vogt and Henry 1998, Casal *et al.* 1999, Soares-Silva *et al.* 2003, Fisher *et al.* 2005). In humans, the most similar protein to MFSD8 is MFSD9 (previously known as MGC11332) with an unknown function. *MFSD9* has recently been identified as a candidate gene showing evidence for association with type 2 diabetes in a genomewide association scan (Rampersaud *et al.* 2007). The other similar human proteins are or are predicted to be involved in the transport of tetracycline (tetracycline transporter-like protein, TETRAN), monoamines (SLC18A2 or VMAT2), and organic cations (SLC22A18 or ORCTL2) (Liu and Edwards 1997, Reece *et al.* 1998). Some human diseases have been associated with genes encoding MFS proteins. Interestingly, vLINCL shares clinical features, e.g. encephalopathy, with some of these diseases. Among these, according to MimMiner (www.cmbi.ru.nl/MimMiner/), that searches for similarities between phenotypes based on the OMIM database (van Driel *et al.* 2006), Salla disease most closely resembles vLINCL. Salla disease and an allelic disease, infantile sialic acid storage disease, are caused by mutations in *SLC17A5* encoding a lysosomal integral membrane protein sialin which exports sialic acid and other acidic monosaccharides out of lysosomes (Verheijen *et al.* 1999, Morin *et al.* 2004).

2.2.7. Intracellular localization of MFSD8

In order to study the subcellular localization of MFSD8, N- and C-terminally HA-tagged MFSD8 proteins were transiently overexpressed in African green monkey kidney (COS-1) and HeLa cells. In immunofluorescence analyses of the cells using HA antibodies MFSD8 was detected in punctate structures in the cytoplasm (for COS-1 cells with N-terminally HA-tagged MFSD8: III, Fig. 5A, 5D, and 5G). A variety of organelle markers were used to identify these structures in COS-1 cells. The HA-MFSD8 showed the strongest overlap with lysosomal markers LAMP1 (III, Fig. 5A-C),

CTSD, and lysobisphosphatidic acid (LBPA). No overlap was observed with early endosomal protein early endosome antigen 1 (EEA1) (III, Fig. 5D-F) or with markers for the earlier compartments of the secretory pathway, including ER resident protein-disulfide isomerase (PDI), Golgi protein giantin, and MPR46 (III, Fig. 5G-I). These results imply that MFSD8 localizes to lysosomes similarly to the majority of the previously identified NCL proteins (CTSD, PPT1, TPP1, CLN3, and CLN5) (Press *et al.* 1960, Rawlings and Barrett 1995, Hellsten *et al.* 1996, Verkruyse and Hofmann 1996, Järvelä *et al.* 1998, Isosomppi *et al.* 2002). Based on the analyses carried out in this study, the MFSD8 protein is proposed to be a novel integral membrane lysosomal protein. The exact cellular function of this novel putative lysosomal transporter remains to be elucidated.

CONCLUSIONS AND FUTURE PROSPECTS

This thesis describes the identification of two novel human NCL-causing genes, *CTSD* and *MFSD8*, which increases the number of known human NCL genes from six to eight, and thus substantially contributes to the understanding of the complete molecular genetic spectrum of NCLs.

We provided the first molecular genetic explanation for congenital human NCL by identification of a disease-causing mutation in the *CTSD* gene in one family. CTSD deficiency was shown to underlie the disease also in another family, and it may underlie the disease in all cases of congenital NCL. Thus, *CTSD* should be studied as a candidate gene in families with this disease.

The molecular genetic basis of vLINCL in the Turkish population was partially resolved in this study by detection of mutations in one previously known NCL gene, *CLN6*, and most importantly, by the mapping, identification, and characterization of a novel gene, *MFSD8* (*CLN7*), underlying the disease. MFSD8 is a novel putative lysosomal integral membrane protein which, as a member of the major facilitator superfamily, is predicted to function as a transporter. In addition, this study indicates the existence of novel genes underlying vLINCL in Turkish families. These results further emphasize the genetic heterogeneity of Turkish vLINCL as well as the genetic heterogeneity of NCLs in general. Moreover, they raise the expectations of the identification of novel NCL genes in the near future.

For an individual family with congenital NCL or vLINCL, the most important effect of this study is definitely the new availability of a molecular genetic diagnosis for the patients as well as the opportunity for carrier and prenatal screening within the family. The diagnostics may still, however, be difficult because of the great locus and allelic heterogeneity within NCLs. In addition, some NCL-causing genes are yet to be identified, hence providing an exact genetic diagnosis may be challenging.

In the long run, the identification of mutations in *CTSD* and *MFSD8* is a starting point for dissecting the molecular mechanisms behind the associated disorders. As *MFSD8* is a novel gene, functional studies are of special importance. Resolving the substrate specificity of this putative transporter will be crucial in unraveling its cellular function and the disease mechanism in the associated vLINCL. Cell and animal models may be utilized in these studies. Whether *MFSD8* is the underlying disease-causing gene in some of the existing, naturally occurring animal models for NCL remains to be explored.

This thesis underlines the importance of the research on the molecular genetic background of rare NCL forms. The information that becomes available immediately upon gene identification itself, along with the biochemical and functional studies on the encoded proteins and on the disease mechanisms in associated disorders, contributes to the challenging task of understanding the complete picture of the molecular pathology underlying the group of NCL disorders. This will be critical in the development of preventive or curative therapies for patients with NCLs.

ACKNOWLEDGEMENTS

This study was carried out at the Folkhälsan Institute of Genetics, and University of Helsinki, Department of Medical Genetics, and Neuroscience Center, during 2004-2008. I warmly thank the former and current heads of these institutes, Professors Anna-Elina Lehesjoki, Leena Palotie, Kristiina Aittomäki, Päivi Peltomäki, and Heikki Rauvala for providing excellent research facilities. I am most grateful for the opportunity to have been a student of the Helsinki Biomedical Graduate School (from 2006).

The financial support of the Folkhälsan Research Foundation, the Sigrid Jusélius Foundation, the Academy of Finland (Centre of Excellence Programmes 2000-2005 and 2006-2011), the European Commission (project LSHM-CT-2003-50305), the Emil Aaltonen Foundation, the University of Helsinki Funds, and the Biomedicum Helsinki Foundation are gratefully acknowledged.

I wish to express my warmest gratitude to my excellent supervisor Professor Anna-Elina Lehesjoki for the opportunity to work in her group. I am grateful to her for guiding me in the world of medical genetics and through this project, and for her positive attitude and encouragement along the way.

The official pre-examiners of my thesis Professor Mark Gardiner and Docent Marjo Kestilä are warmly acknowledged for thorough reviews and valuable comments. Marjo, as well as Jouni Vesa, my thesis committee members, are warmly thanked for the evaluation of my work and for their positive attitude and interest towards my work. Jouni is also thanked for the real-life mentoring in OC.

I am deeply grateful to all co-authors and collaborators of this study for their most valuable contribution. I express my deepest appreciation to Petter Strömme, Jan Maehlen, Meral Topcu, Alfried Kohlschütter, and Callum Wilson for their clinical expertise and for providing the patient samples that have been most valuable for these projects. In particular, the contribution of Meral Topcu has been invaluable in the Turkish vLINCL project. Hannes Lohi, Berge Minassian, Andrew Paterson, and Xiao-Qing Liu are thanked for the smooth and efficient collaboration in the SNP scan of the *CLN7* gene search.

I warmly thank Jaana Vesterinen for giving me the opportunity to collaborate with her group in the congenital NCL project. I am grateful for all inspiring discussions and for all advice and support in this and other projects. Jaana is also thanked for critically reading parts of my thesis manuscript. I wish to thank Sanna Partanen for excellent collaboration, for inspirational conversations, and for her friendship. Matti Haltia, the “father of the congenital NCL project”, and Aleksi Haapanen are thanked for fruitful collaboration.

I am deeply grateful to Anna-Kaisa Anttonen for sharing her expertise in medical genetics with me, and for her time and help in these projects. Special thanks to Anna-Kaisa also for patiently answering my hundreds of questions about the thesis preparation and dissertation. I would like to thank Tarja Joensuu and Tarja Salonen for providing their expert advice always with most helpful attitudes. I owe my warmest thanks to Nina Aula for choosing the correct gene in the “*CLN7* lotto”. Nina is also thanked for her positive attitude, and for advice, help, and support in all kinds of

things, especially during the last few months. Ulla Lahtinen is warmly thanked for her excellent cell biological advice.

The other former and present CLN8/NCL group members who I have been very lucky to work with, Liina Lonka, Teija Toivonen, Mervi Kuronen, Outi Kopra, and Maria Kousi, are all warmly thanked for the co-work and support during my thesis project. Liina is especially thanked for introducing me to the NCL world, and Maria for energetically continuing the NCL genetics project.

Teija Toivonen is warmly thanked for her excellent technical assistance. I am grateful to Paula Hakala, Mairi Kuris, and Ahmed Mohamed for their skilful help in the lab. I thank sincerely Aila Riikonen, Jaana Welin-Haapamäki, Stephan Keskinen, Marjatta Valkama, and Solveig Halonen for kindly helping me with so many practical matters during these years. Sinikka Lindh is acknowledged for organizing the patient samples.

Jodie Painter is thanked for carefully revising the language of my thesis, and for her interest and positive attitude towards my work.

I am thankful to all former and present colleagues at Folkhälsan and Biomedicum for their helpful attitude, and the good company in and out of the lab. Especially, my former and current room mates, Kirsi, Liina, Tarja S, Juha K, Anne S, Liisa, Nina A, Anne P, Sanna, and Otto, are thanked for the positive working environment, for interesting discussions, and for their help in the most diverse matters, from solutions to computer problems to the advice in life. My fellow PhD students during these years, Liina, Juha K, Kati, Jukka, Riikka, Kirsi, Anna-Kaisa, Saara, Anne S, Vilma, Mervi, Anna, Maria S, Hanna, Jaakko, Milja, Jeanette, Minttu, Maria K, and Otto, are thanked for sharing the good and bad days of graduate student life, and for the excellent company in the lab and on congress trips.

All my dear friends are thanked for being there, and for all the enjoyable moments spent together. With all my heart I would like to thank my parents Marjut and Antti, and my sister Anna and my brother Asko for their continuous love, help, and support. I thank my dear Tomi for his love and care, and especially for supporting me in finishing this project.

Finally, I wish to express my warmest gratitude to the NCL patients and their families for their kind participation in this research.

Helsinki, April 2008

REFERENCES

Abecasis GR, Cherny SS, Cookson WO, Cardon LR (2002) Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* 30:97-101.

Åberg L, Järvelä I, Rapola J, Autti T, Kirveskari E, Lappi M, Sipilä L, Santavuori P (1998) Atypical juvenile neuronal ceroid lipofuscinosis with granular osmiophilic deposit-like inclusions in the autonomic nerve cells of the gut wall. *Acta Neuropathol (Berl)* 95:306-312.

Ahtiainen L, Van Diggelen OP, Jalanko A, Kopra O (2003) Palmitoyl protein thioesterase 1 is targeted to the axons in neurons. *J Comp Neurol* 455:368-377.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403-410.

Antonarakis SE, Beckmann JS (2006) Mendelian disorders deserve more attention. *Nat Rev Genet* 7:277-282.

Arai K, Shimaya A, Hiratani N, Ohkuma S (1993) Purification and characterization of lysosomal H(+)-ATPase. An anion-sensitive v-type H(+)-ATPase from rat liver lysosomes. *J Biol Chem* 268:5649-5660.

Augereau P, Garcia M, Mattei MG, Cavailles V, Depadova F, Derocq D, Capony F, Ferrara P, Rochefort H (1988) Cloning and sequencing of the 52K cathepsin D complementary deoxyribonucleic acid of MCF7 breast cancer cells and mapping on chromosome 11. *Mol Endocrinol* 2:186-192.

Awano T, Katz ML, O'Brien D P, Taylor JF, Evans J, Khan S, Sohar I, Lobel P, Johnson GS (2006a) A mutation in the cathepsin D gene (CTSD) in American Bulldogs with neuronal ceroid lipofuscinosis. *Mol Genet Metab* 87:341-348.

Awano T, Katz ML, O'Brien DP, Sohar I, Lobel P, Coates JR, Khan S, Johnson GC, Giger U, Johnson GS (2006b) A frame shift mutation in canine TPP1 (the ortholog of human CLN2) in a juvenile Dachshund with neuronal ceroid lipofuscinosis. *Mol Genet Metab* 89:254-260.

Bagshaw RD, Mahuran DJ, Callahan JW (2005) A proteomic analysis of lysosomal integral membrane proteins reveals the diverse composition of the organelle. *Mol Cell Proteomics* 4:133-143.

Barohn RJ, Dowd DC, Kagan-Hallet KS (1992) Congenital ceroid-lipofuscinosis. *Pediatr Neurol* 8:54-59.

Bellizzi JJ, 3rd, Widom J, Kemp C, Lu JY, Das AK, Hofmann SL, Clardy J (2000) The crystal structure of palmitoyl protein thioesterase 1 and the molecular basis of infantile neuronal ceroid lipofuscinosis. *Proc Natl Acad Sci U S A* 97:4573-4578.

Ben-Dov C, Hartmann B, Lundgren J, Valcarcel J (2008) Genome-wide analysis of alternative pre-mRNA splicing. *J Biol Chem* 283:1229-1233.

Berggard T, Linse S, James P (2007) Methods for the detection and analysis of protein-protein interactions. *Proteomics* 7:2833-2842.

Berkovic SF, Carpenter S, Andermann F, Andermann E, Wolfe LS (1988) Kufs' disease: a critical reappraisal. *Brain* 111 (Pt 1):27-62.

Bessa C, Teixeira CA, Mangas M, Dias A, Sa Miranda MC, Guimaraes A, Ferreira JC, Canas N, Cabral P, Ribeiro MG (2006) Two novel CLN5 mutations in a Portuguese patient with vLINCL: insights into molecular mechanisms of CLN5 deficiency. *Mol Genet Metab* 89:245-253.

Bittles A (2001) Consanguinity and its relevance to clinical genetics. *Clin Genet* 60:89-98.

Blott EJ, Griffiths GM (2002) Secretory lysosomes. *Nat Rev Mol Cell Biol* 3:122-131.

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

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet* 32:314-331.

Brent MR, Guigo R (2004) Recent advances in gene structure prediction. *Curr Opin Struct Biol* 14:264-272.

Bright NA, Gratian MJ, Luzio JP (2005) Endocytic delivery to lysosomes mediated by concurrent fusion and kissing events in living cells. *Curr Biol* 15:360-365.

Brix K (2005) Lysosomal proteases. In: Saftig P (ed) *Lysosomes*. Landes Bioscience / Eurekah.com, Georgetown, Texas, USA, pp 50-59.

Bronson RT, Donahue LR, Johnson KR, Tanner A, Lane PW, Faust JR (1998) Neuronal ceroid lipofuscinosis (nclf), a new disorder of the mouse linked to chromosome 9. *Am J Med Genet* 77:289-297.

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

Camp LA, Hofmann SL (1993) Purification and properties of a palmitoyl-protein thioesterase that cleaves palmitate from H-Ras. *J Biol Chem* 268:22566-22574.

Cannelli N, Cassandrini D, Bertini E, Striano P, Fusco L, Gaggero R, Specchio N, Biancheri R, Vigevano F, Bruno C, Simonati A, Zara F, Santorelli FM (2006) Novel mutations in CLN8 in Italian variant late infantile neuronal ceroid lipofuscinosis: Another genetic hit in the Mediterranean. *Neurogenetics* 7:111-117.

Cannelli N, Nardocci N, Cassandrini D, Morbin M, Aiello C, Bugiani M, Criscuolo L, Zara F, Striano P, Granata T, Bertini E, Simonati A, Santorelli FM (2007) Revelation of a novel CLN5 mutation in early juvenile neuronal ceroid lipofuscinosis. *Neuropediatrics* 38:46-49.

Carr IM, Flintoff KJ, Taylor GR, Markham AF, Bonthron DT (2006) Interactive visual analysis of SNP data for rapid autozygosity mapping in consanguineous families. *Hum Mutat* 27:1041-1046.

Cartegni L, Chew SL, Krainer AR (2002) Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat Rev Genet* 3:285-298.

Casal M, Paiva S, Andrade RP, Gancedo C, Leao C (1999) The lactate-proton symport of Saccharomyces cerevisiae is encoded by JEN1. *J Bacteriol* 181:2620-2623.

Cataldo AM, Barnett JL, Berman SA, Li J, Quarless S, Bursztajn S, Lippa C, Nixon RA (1995) Gene expression and cellular content of cathepsin D in Alzheimer's disease brain: evidence for early up-regulation of the endosomal-lysosomal system. *Neuron* 14:671-680.

Chen JW, Murphy TL, Willingham MC, Pastan I, August JT (1985) Identification of two lysosomal membrane glycoproteins. *J Cell Biol* 101:85-95.

Chiang AP, Beck JS, Yen HJ, Tayeh MK, Scheetz TE, Swiderski RE, Nishimura DY, Braun TA, Kim KY, Huang J, Elbedour K, Carmi R, Slusarski DC, Casavant TL, Stone EM, Sheffield VC (2006) Homozygosity mapping with SNP arrays identifies TRIM32, an E3 ubiquitin ligase, as a Bardet-Biedl syndrome gene (BBS11). *Proc Natl Acad Sci U S A* 103:6287-6292.

Collin RW, Kalay E, Tariq M, Peters T, van der Zwaag B, Venselaar H, Oostrik J, Lee K, Ahmed ZM, Caylan R, Li Y, Spierenburg HA, Eyupoglu E, Heister A, Riazuddin S, Bahat E, Ansar M, Arslan S, Wollnik B, Brunner HG, Cremers CW, Karaguzel A, Ahmad W, Cremers FP, Vriend G, Friedman TB, Riazuddin S, Leal SM, Kremer H (2008) Mutations of ESRRB encoding estrogen-related receptor beta cause autosomal-recessive nonsyndromic hearing impairment DFNB35. *Am J Hum Genet* 82:125-138.

Collins FS (1992) Positional cloning: let's not call it reverse anymore. *Nat Genet* 1:3-6.

Collins FS (1995) Positional cloning moves from perditional to traditional. *Nat Genet* 9:347-350.

Collins JS, Schwartz CE (2002) Detecting polymorphisms and mutations in candidate genes. *Am J Hum Genet* 71:1251-1252.

Cotman SL, Vrbanac V, Lebel LA, Lee RL, Johnson KA, Donahue LR, Teed AM, Antonellis K, Bronson RT, Lerner TJ, MacDonald ME (2002) Cln3(Deltaex7/8) knock-in mice with the common JNCL mutation exhibit progressive neurologic disease that begins before birth. *Hum Mol Genet* 11:2709-2721.

Cuervo AM, Dice JF (1996) A receptor for the selective uptake and degradation of proteins by lysosomes. *Science* 273:501-503.

Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES (2001) High-resolution haplotype structure in the human genome. *Nat Genet* 29:229-232.

Das AK, Becerra CH, Yi W, Lu JY, Siakotos AN, Wisniewski KE, Hofmann SL (1998) Molecular genetics of palmitoyl-protein thioesterase deficiency in the U.S. *J Clin Invest* 102:361-370.

Das AK, Lu JY, Hofmann SL (2001) Biochemical analysis of mutations in palmitoyl-protein thioesterase causing infantile and late-onset forms of neuronal ceroid lipofuscinosis. *Hum Mol Genet* 10:1431-1439.

Davis W, Jr., Boyd JT, Ile KE, Tew KD (2004) Human ATP-binding cassette transporter-2 (ABCA2) positively regulates low-density lipoprotein receptor expression and negatively regulates cholesterol esterification in Chinese hamster ovary cells. *Biochim Biophys Acta* 1683:89-100.

De Duve C, Pressman BC, Gianetto R, Wattiaux R, Appelmans F (1955) Tissue fractionation studies. 6. Intracellular distribution patterns of enzymes in rat-liver tissue. *Biochem J* 60:604-617.

de la Chapelle A, Wright FA (1998) Linkage disequilibrium mapping in isolated populations: the example of Finland revisited. *Proc Natl Acad Sci U S A* 95:12416-12423.

de Voer G, van der Bent P, Rodrigues AJ, van Ommen GJ, Peters DJ, Taschner PE (2005) 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* 28:1065-1080.

Dearlove AM (2002) High throughput genotyping technologies. *Brief Funct Genomic Proteomic* 1:139-150.

Diment S, Martin KJ, Stahl PD (1989) Cleavage of parathyroid hormone in macrophage endosomes illustrates a novel pathway for intracellular processing of proteins. *J Biol Chem* 264:13403-13406.

Dittmer F, Ulbrich EJ, Hafner A, Schmahl W, Meister T, Pohlmann R, von Figura K (1999) Alternative mechanisms for trafficking of lysosomal enzymes in mannose 6-phosphate receptor-deficient mice are cell type-specific. *J Cell Sci* 112 (Pt 10):1591-1597.

Doray B, Ghosh P, Griffith J, Geuze HJ, Kornfeld S (2002) Cooperation of GGAs and AP-1 in packaging MPRs at the trans-Golgi network. *Science* 297:1700-1703.

Dryja TP, McGee TL, Reichel E, Hahn LB, Cowley GS, Yandell DW, Sandberg MA, Berson EL (1990) A point mutation of the rhodopsin gene in one form of retinitis pigmentosa. *Nature* 343:364-366.

Eiberg H, Gardiner RM, Mohr J (1989) Batten disease (Spielmeyer-Sjogren disease) and haptoglobins (HP): indication of linkage and assignment to chr. 16. *Clin Genet* 36:217-218.

Eliason SL, Stein CS, Mao Q, Tecedor L, Ding SL, Gaines DM, Davidson BL (2007) A knock-in reporter model of Batten disease. *J Neurosci* 27:9826-9834.

Elleder M, Lake BD, Goebel HH, Rapola J, Haltia M, Carpenter S (1999) Definitions of the Ultrastructural Patterns found in NCL. In: Goebel HH, Mole SE, Lake BD (eds) *The Neuronal Ceroid Lipofuscinosis (Batten Disease)*. IOS Press, Amsterdam, pp 5-15.

Elleder M, Sokolova J, Hrebicek M (1997) Follow-up study of subunit c of mitochondrial ATP synthase (SCMAS) in Batten disease and in unrelated lysosomal disorders. *Acta Neuropathol* 93:379-390.

Erickson AH, Blobel G (1979) Early events in the biosynthesis of the lysosomal enzyme cathepsin D. *J Biol Chem* 254:11771-11774.

Erickson AH, Conner GE, Blobel G (1981) Biosynthesis of a lysosomal enzyme. Partial structure of two transient and functionally distinct NH2-terminal sequences in cathepsin D. *J Biol Chem* 256:11224-11231.

Eskelinen EL, Schmidt CK, Neu S, Willenborg M, Fuertes G, Salvador N, Tanaka Y, Lullmann-Rauch R, Hartmann D, Heeren J, von Figura K, Knecht E, Saftig P (2004) Disturbed cholesterol traffic but normal proteolytic function in LAMP-1/LAMP-2 double-deficient fibroblasts. *Mol Biol Cell* 15:3132-3145.

Eskelinen EL, Tanaka Y, Saftig P (2003) At the acidic edge: emerging functions for lysosomal membrane proteins. *Trends Cell Biol* 13:137-145.

Evans J, Katz ML, Levesque D, Shelton GD, de Lahunta A, O'Brien D (2005) A variant form of neuronal ceroid lipofuscinosis in American bulldogs. *J Vet Intern Med* 19:44-51.

Fambrough DM, Takeyasu K, Lippincott-Schwarz J, Siegel NR (1988) Structure of LEP100, a glycoprotein that shuttles between lysosomes and the plasma membrane, deduced from the nucleotide sequence of the encoding cDNA. *J Cell Biol* 106:61-67.

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

Faust PL, Kornfeld S, Chirgwin JM (1985) Cloning and sequence analysis of cDNA for human cathepsin D. *Proc Natl Acad Sci U S A* 82:4910-4914.

Felbor U, Kessler B, Mothes W, Goebel HH, Ploegh HL, Bronson RT, Olsen BR (2002) Neuronal loss and brain atrophy in mice lacking cathepsins B and L. *Proc Natl Acad Sci U S A* 99:7883-7888.

Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC (1998) Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* 391:806-811.

Fisher E, Almaguer C, Holic R, Griac P, Patton-Vogt J (2005) Glycerophosphocholine-dependent growth requires Gde1p (YPL110c) and Git1p in Saccharomyces cerevisiae. *J Biol Chem* 280:36110-36117.

Forgac M (2007) Vacuolar ATPases: rotary proton pumps in physiology and pathophysiology. *Nat Rev Mol Cell Biol* 8:917-929.

Frattini A, Orchard PJ, Sobacchi C, Giliani S, Abinun M, Mattsson JP, Keeling DJ, Andersson AK, Wallbrandt P, Zecca L, Notarangelo LD, Vezzoni P, Villa A (2000) Defects in TCIRG1 subunit of the vacuolar proton pump are responsible for a subset of human autosomal recessive osteopetrosis. *Nat Genet* 25:343-346.

Fukuda M, Viitala J, Matteson J, Carlsson SR (1988) Cloning of cDNAs encoding human lysosomal membrane glycoproteins, h-lamp-1 and h-lamp-2. Comparison of their deduced amino acid sequences. *J Biol Chem* 263:18920-18928.

Fusek M, Vetvicka V (2005) Dual role of cathepsin D: ligand and protease. *Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub* 149:43-50.

Futerman AH, van Meer G (2004) The cell biology of lysosomal storage disorders. *Nat Rev Mol Cell Biol* 5:554-565.

Gachet Y, Codlin S, Hyams JS, Mole SE (2005) btn1, the Schizosaccharomyces pombe homologue of the human Batten disease gene CLN3, regulates vacuole homeostasis. *J Cell Sci* 118:5525-5536.

Gao H, Boustany RM, Espinola JA, Cotman SL, Srinidhi L, Antonellis KA, Gillis T, Qin X, Liu S, Donahue LR, Bronson RT, Faust JR, Stout D, Haines JL, Lerner TJ, MacDonald ME (2002) Mutations in a novel CLN6-encoded transmembrane protein cause variant neuronal ceroid lipofuscinosis in man and mouse. *Am J Hum Genet* 70:324-335.

Garborg I, Torvik A, Hals J, Tangsrud SE, Lindemann R (1987) Congenital neuronal ceroid lipofuscinosis. A case report. *Acta Pathol Microbiol Immunol Scand [A]* 95:119-125.

Ghosh P, Dahms NM, Kornfeld S (2003) Mannose 6-phosphate receptors: new twists in the tale. *Nat Rev Mol Cell Biol* 4:202-212.

Gieselmann V, Hasilik A, von Figura K (1985) Processing of human cathepsin D in lysosomes in vitro. *J Biol Chem* 260:3215-3220.

Gieselmann V, Pohlmann R, Hasilik A, Von Figura K (1983) Biosynthesis and transport of cathepsin D in cultured human fibroblasts. *J Cell Biol* 97:1-5.

Glickman JN, Kornfeld S (1993) Mannose 6-phosphate-independent targeting of lysosomal enzymes in I-cell disease B lymphoblasts. *J Cell Biol* 123:99-108.

Goebel HH (1997) Morphologic diagnosis in neuronal ceroid lipofuscinosis. *Neuropediatrics* 28:67-69.

Gray IC, Campbell DA, Spurr NK (2000) Single nucleotide polymorphisms as tools in human genetics. *Hum Mol Genet* 9:2403-2408.

Greiner-Tollersrud OK, Berg T (2005) Lysosomal storage disorders. In: Saftig P (ed) *Lysosomes*. Landes Bioscience / Eurekah.com, Georgetown, Texas, USA, pp 60-73.

Gupta P, Soyombo AA, Atashband A, Wisniewski KE, Shelton JM, Richardson JA, Hammer RE, Hofmann SL (2001) Disruption of PPT1 or PPT2 causes neuronal ceroid lipofuscinosis in knockout mice. *Proc Natl Acad Sci U S A* 98:13566-13571.

Gupta P, Soyombo AA, Shelton JM, Wilkofsky IG, Wisniewski KE, Richardson JA, Hofmann SL (2003) Disruption of PPT2 in mice causes an unusual lysosomal storage disorder with neurovisceral features. *Proc Natl Acad Sci U S A* 100:12325-12330.

Haapanen A, Ramadan UA, Autti T, Joensuu R, Tyynela J (2007) In vivo MRI reveals the dynamics of pathological changes in the brains of cathepsin D-deficient mice and correlates changes in manganese-enhanced MRI with microglial activation. *Magn Reson Imaging* 25:1024-1031.

Hall NA, Lake BD, Dewji NN, Patrick AD (1991) Lysosomal storage of subunit c of mitochondrial ATP synthase in Batten's disease (ceroid-lipofuscinosis). *Biochem J* 275 (Pt 1):269-272.

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

Haltia M, Rapola J, Santavuori P, Keränen A (1973) Infantile type of so-called neuronal ceroid-lipofuscinosis. 2. Morphological and biochemical studies. *J Neurol Sci* 18:269-285.

Hartikainen JM, Ju W, Wisniewski KE, Moroziewicz DN, Kaczmarski AL, McLendon L, Zhong D, Suarez CT, Brown WT, Zhong N (1999) Late infantile neuronal ceroid lipofuscinosis is due to splicing mutations in the CLN2 gene. *Mol Genet Metab* 67:162-168.

Hasilik A, Neufeld EF (1980a) Biosynthesis of lysosomal enzymes in fibroblasts. Phosphorylation of mannose residues. *J Biol Chem* 255:4946-4950.

Hasilik A, Neufeld EF (1980b) Biosynthesis of lysosomal enzymes in fibroblasts. Synthesis as precursors of higher molecular weight. *J Biol Chem* 255:4937-4945.

Haskell RE, Carr CJ, Pearce DA, Bennett MJ, Davidson BL (2000) Batten disease: evaluation of CLN3 mutations on protein localization and function. *Hum Mol Genet* 9:735-744.

Hästbacka J, de la Chapelle A, Kaitila I, Sistonen P, Weaver A, Lander E (1992) Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland. *Nat Genet* 2:204-211.

Heine C, Koch B, Storch S, Kohlschutter A, Palmer DN, Braulke T (2004) Defective endoplasmic reticulum-resident membrane protein CLN6 affects lysosomal degradation of endocytosed arylsulfatase A. *J Biol Chem* 279:22347-22352.

Heine C, Quitsch A, Storch S, Martin Y, Lonka L, Lehesjoki AE, Mole SE, Braulke T (2007) Topology and endoplasmic reticulum retention signals of the lysosomal storage disease-related membrane protein CLN6. *Mol Membr Biol* 24:74-87.

Heinonen O, Kyttälä A, Lehmus E, Paunio T, Peltonen L, Jalanko A (2000) Expression of palmitoyl protein thioesterase in neurons. *Mol Genet Metab* 69:123-129.

Hellsten E, Vesa J, Olkkonen VM, Jalanko A, Peltonen L (1996) Human palmitoyl protein thioesterase: evidence for lysosomal targeting of the enzyme and disturbed cellular routing in infantile neuronal ceroid lipofuscinosis. *Embo J* 15:5240-5245.

Herva R, Tyynelä J, Hirvasniemi A, Syrjäkallio-Ylitalo M, Haltia M (2000) Northern epilepsy: a novel form of neuronal ceroid-lipofuscinosis. *Brain Pathol* 10:215-222.

Hillman RT, Green RE, Brenner SE (2004) An unappreciated role for RNA surveillance. *Genome Biol* 5:R8.

Hirvasniemi A, Herrala P, Leisti J (1995) Northern epilepsy syndrome: clinical course and the effect of medication on seizures. *Epilepsia* 36:792-797.

Hirvasniemi A, Lang H, Lehesjoki AE, Leisti J (1994) Northern epilepsy syndrome: an inherited childhood onset epilepsy with associated mental deterioration. *J Med Genet* 31:177-182.

Hobert JA, Dawson G (2006) Neuronal ceroid lipofuscinoses therapeutic strategies: past, present and future. *Biochim Biophys Acta* 1762:945-953.

Hofman I, Kohlschutter A, Santavuori P, Gottlob I, Goebel HH, Lake BD, Schutgens RBH, Greene NDE, Leung K-Y, Mitchison HM, Munroe PB, Taschner PEM (1999) CLN3 Juvenile NCL. In: Goebel HH, Mole SE, Lake BD (eds) *The Neuronal Ceroid Lipofuscinosis (Batten Disease)*. IOS Press, Amsterdam, pp 55-76.

Holmberg V, Jalanko A, Isosomppi J, Fabritius AL, Peltonen L, Kopra O (2004) The mouse ortholog of the neuronal ceroid lipofuscinosis CLN5 gene encodes a soluble lysosomal glycoprotein expressed in the developing brain. *Neurobiol Dis* 16:29-40.

Holmberg V, Lauronen L, Autti T, Santavuori P, Savukoski M, Uvebrant P, Hofman I, Peltonen L, Järvelä I (2000) Phenotype-genotype correlation in eight patients with Finnish variant late infantile NCL (CLN5). *Neurology* 55:579-581.

Houweling PJ, Cavanagh JA, Palmer DN, Frugier T, Mitchell NL, Windsor PA, Raadsma HW, Tammen I (2006) Neuronal ceroid lipofuscinosis in Devon cattle is caused by a single base duplication (c.662dupG) in the bovine CLN5 gene. *Biochim Biophys Acta* 1762:890-897.

Huang JS, Huang SS, Tang J (1979) Cathepsin D isozymes from porcine spleens. Large scale purification and polypeptide chain arrangements. *J Biol Chem* 254:11405-11417.

Humphreys S, Lake BD, Scholtz CL (1985) Congenital amaurotic idiocy--a pathological, histochemical, biochemical and ultrastructural study. *Neuropathol Appl Neurobiol* 11:475-484.

Huynh KK, Eskelinen EL, Scott CC, Malevanets A, Saftig P, Grinstein S (2007) LAMP proteins are required for fusion of lysosomes with phagosomes. *Embo J* 26:313-324.

Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C (2004) Detection of large-scale variation in the human genome. *Nat Genet* 36:949-951.

Ihrke G, Kyttala A, Russell MR, Rous BA, Luzio JP (2004) Differential use of two AP-3-mediated pathways by lysosomal membrane proteins. *Traffic* 5:946-962.

International Human Genome Sequencing Consortium (2004) Finishing the euchromatic sequence of the human genome. *Nature* 431:931-945.

Isosomppi J, Vesa J, Jalanko A, Peltonen L (2002) Lysosomal localization of the neuronal ceroid lipofuscinosis CLN5 protein. *Hum Mol Genet* 11:885-891.

Jalanko A, Vesa J, Manninen T, von Schantz C, Minye H, Fabritius AL, Salonen T, Rapola J, Gentile M, Kopra O, Peltonen L (2005) Mice with Ppt1Deltaex4 mutation replicate the INCL phenotype and show an inflammation-associated loss of interneurons. *Neurobiol Dis* 18:226-241.

Janes RW, Munroe PB, Mitchison HM, Gardiner RM, Mole SE, Wallace BA (1996) A model for Batten disease protein CLN3: functional implications from homology and mutations. *FEBS Lett* 399:75-77.

Järplid B, Haltia M (1993) An animal model of the infantile type of neuronal ceroid-lipofuscinosis. *J Inherit Metab Dis* 16:274-277.

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

Järvelä I, Lehtovirta M, Tikkanen R, Kyttälä A, Jalanko A (1999) Defective intracellular transport of CLN3 is the molecular basis of Batten disease (JNCL). *Hum Mol Genet* 8:1091-1098.

Järvelä I, Sainio M, Rantamäki T, Olkkonen VM, Carpen O, Peltonen L, Jalanko A (1998) Biosynthesis and intracellular targeting of the CLN3 protein defective in Batten disease. *Hum Mol Genet* 7:85-90.

Järvelä I, Schleutker J, Haataja L, Santavuori P, Puhakka L, Manninen T, Palotie A, Sandkuijl LA, Renlund M, White R, Aula P, Peltonen L (1991) Infantile form of neuronal ceroid lipofuscinosis (CLN1) maps to the short arm of chromosome 1. *Genomics* 9:170-173.

Jeffreys AJ, Kauppi L, Neumann R (2001) Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. *Nat Genet* 29:217-222.

Jeyakumar M, Dwek RA, Butters TD, Platt FM (2005) Storage solutions: treating lysosomal disorders of the brain. *Nat Rev Neurosci* 6:713-725.

Journet A, Chapel A, Kieffer S, Roux F, Garin J (2002) Proteomic analysis of human lysosomes: application to monocytic and breast cancer cells. *Proteomics* 2:1026-1040.

Ju W, Zhong R, Moore S, Moroziewicz D, Currie JR, Parfrey P, Brown WT, Zhong N (2002) Identification of novel CLN2 mutations shows Canadian specific NCL2 alleles. *J Med Genet* 39:822-825.

Kalatzis V, Cherqui S, Antignac C, Gasnier B (2001) Cystinosin, the protein defective in cystinosis, is a H(+)-driven lysosomal cystine transporter. *Embo J* 20:5940-5949.

Kasper D, Planells-Cases R, Fuhrmann JC, Scheel O, Zeitz O, Ruether K, Schmitt A, Poet M, Steinfeld R, Schweizer M, Kornak U, Jentsch TJ (2005) Loss of the chloride channel ClC-7 leads to lysosomal storage disease and neurodegeneration. *Embo J* 24:1079-1091.

Katz ML, Khan S, Awano T, Shahid SA, Siakotos AN, Johnson GS (2005) A mutation in the CLN8 gene in English Setter dogs with neuronal ceroid-lipofuscinosis. *Biochem Biophys Res Commun* 327:541-547.

Katz ML, Shibuya H, Liu PC, Kaur S, Gao CL, Johnson GS (1999) A mouse gene knockout model for juvenile ceroid-lipofuscinosis (Batten disease). *J Neurosci Res* 57:551-556.

Kitzmuller C, Haines RL, Codlin S, Cutler DF, Mole SE (2008) A function retained by the common mutant CLN3 protein is responsible for the late onset of juvenile neuronal ceroid lipofuscinosis. *Hum Mol Genet* 17:303-312.

Klockars T, Savukoski M, Isosomppi J, Laan M, Järvelä I, Petrukhin K, Palotie A, Peltonen L (1996) Efficient construction of a physical map by fiber-FISH of the CLN5 region:

refined assignment and long-range contig covering the critical region on 13q22. *Genomics* 35:71-78.

Koike M, Nakanishi H, Saftig P, Ezaki J, Isahara K, Ohsawa Y, Schulz-Schaeffer W, Watanabe T, Waguri S, Kametaka S, Shibata M, Yamamoto K, Kominami E, Peters C, von Figura K, Uchiyama Y (2000) Cathepsin D deficiency induces lysosomal storage with ceroid lipofuscin in mouse CNS neurons. *J Neurosci* 20:6898-6906.

Koike M, Shibata M, Waguri S, Yoshimura K, Tanida I, Kominami E, Gotow T, Peters C, von Figura K, Mizushima N, Saftig P, Uchiyama Y (2005) Participation of autophagy in storage of lysosomes in neurons from mouse models of neuronal ceroid-lipofuscinoses (Batten disease). *Am J Pathol* 167:1713-1728.

Kollmann K, Mutenda KE, Balleininger M, Eckermann E, von Figura K, Schmidt B, Lubke T (2005) Identification of novel lysosomal matrix proteins by proteome analysis. *Proteomics* 5:3966-3978.

Kopra O, Vesa J, von Schantz C, Manninen T, Minye H, Fabritius AL, Rapola J, van Diggelen OP, Saarela J, Jalanko A, Peltonen L (2004) A mouse model for Finnish variant late infantile neuronal ceroid lipofuscinosis, CLN5, reveals neuropathology associated with early aging. *Hum Mol Genet* 13:2893-2906.

Kornak U, Kasper D, Bosl MR, Kaiser E, Schweizer M, Schulz A, Friedrich W, Delling G, Jentsch TJ (2001) Loss of the ClC-7 chloride channel leads to osteopetrosis in mice and man. *Cell* 104:205-215.

Kornfeld S, Mellman I (1989) The biogenesis of lysosomes. *Annu Rev Cell Biol* 5:483-525.

Krogh A, Larsson B, von Heijne G, Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 305:567-580.

Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES (1996) Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet* 58:1347-1363.

Kruglyak L, Nickerson DA (2001) Variation is the spice of life. *Nat Genet* 27:234-236.

Kubo Y, Sekiya S, Ohigashi M, Takenaka C, Tamura K, Nada S, Nishi T, Yamamoto A, Yamaguchi A (2005) ABCA5 resides in lysosomes, and ABCA5 knockout mice develop lysosomal disease-like symptoms. *Mol Cell Biol* 25:4138-4149.

Kyttälä A, Ihrke G, Vesa J, Schell MJ, Luzio JP (2004) Two motifs target Batten disease protein CLN3 to lysosomes in transfected nonneuronal and neuronal cells. *Mol Biol Cell* 15:1313-1323.

Kyttälä A, Lahtinen U, Braulke T, Hofmann SL (2006) Functional biology of the neuronal ceroid lipofuscinoses (NCL) proteins. *Biochim Biophys Acta* 1762:920-933.

Lander ES, Botstein D (1987) Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. *Science* 236:1567-1570.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ,

Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blocker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ (2001) Initial sequencing and analysis of the human genome. *Nature* 409:860-921.

Lange PF, Wartosch L, Jentsch TJ, Fuhrmann JC (2006) ClC-7 requires Ostm1 as a beta-subunit to support bone resorption and lysosomal function. *Nature* 440:220-223.

Lathrop GM, Lalouel JM (1984) Easy calculations of lod scores and genetic risks on small computers. *Am J Hum Genet* 36:460-465.

Lauronen L, Munroe PB, Järvelä I, Autti T, Mitchison HM, O'Rawe AM, Gardiner RM, Mole SE, Puranen J, Häkkinen AM, Kirveskari E, Santavuori P (1999) Delayed classic and protracted phenotypes of compound heterozygous juvenile neuronal ceroid lipofuscinosis. *Neurology* 52:360-365.

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

Lemieux MJ (2007) Eukaryotic major facilitator superfamily transporter modeling based on the prokaryotic GlpT crystal structure. *Mol Membr Biol* 24:333-341.

Letourneur F, Klausner RD (1992) A novel di-leucine motif and a tyrosine-based motif independently mediate lysosomal targeting and endocytosis of CD3 chains. *Cell* 69:1143-1157.

Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, Axelrod N, Huang J, Kirkness EF, Denisov G, Lin Y, MacDonald JR, Pang AW, Shago M, Stockwell TB, Tsiamouri A, Bafna V, Bansal V, Kravitz SA, Busam DA, Beeson KY, McIntosh TC, Remington KA, Abril JF, Gill J, Borman J, Rogers YH, Frazier ME, Scherer SW, Strausberg RL, Venter JC (2007) The diploid genome sequence of an individual human. *PLoS Biol* 5:e254.

Lewis BP, Green RE, Brenner SE (2003) Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proc Natl Acad Sci U S A* 100:189-192.

Lewis V, Green SA, Marsh M, Vihko P, Helenius A, Mellman I (1985) Glycoproteins of the lysosomal membrane. *J Cell Biol* 100:1839-1847.

Liaudet-Coopman E, Beaujouin M, Derocq D, Garcia M, Glondu-Lassis M, Laurent-Matha V, Prebois C, Rochefort H, Vignon F (2006) Cathepsin D: newly discovered functions of a long-standing aspartic protease in cancer and apoptosis. *Cancer Lett* 237:167-179.

Lin L, Sohar I, Lackland H, Lobel P (2001) The human CLN2 protein/tripeptidyl-peptidase I is a serine protease that autoactivates at acidic pH. *J Biol Chem* 276:2249-2255.

Lingaas F, Mitchison M, Mole SE, Koppang N, Goebel HH, Lake BD (1999) Animal Models of NCL. In: Goebel HH, Mole SE, Lake BD (eds) *The Neuronal Ceroid Lipofuscinosis (Batten Disease)*. IOS Press, Amsterdam, pp 152-167.

Liu Y, Edwards RH (1997) The role of vesicular transport proteins in synaptic transmission and neural degeneration. *Annu Rev Neurosci* 20:125-156.

Lockhart DJ, Dong H, Byrne MC, Follettie MT, Gallo MV, Chee MS, Mittmann M, Wang C, Kobayashi M, Horton H, Brown EL (1996) Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat Biotechnol* 14:1675-1680.

Lonka L (2004) *Neuronal ceroid lipofuscinosis CLN8 - from gene to protein*. University of Helsinki, Helsinki, Finland.

Lonka L, Kyttälä A, Ranta S, Jalanko A, Lehesjoki AE (2000) The neuronal ceroid lipofuscinosis CLN8 membrane protein is a resident of the endoplasmic reticulum. *Hum Mol Genet* 9:1691-1697.

Lonka L, Salonen T, Siintola E, Kopra O, Lehesjoki AE, Jalanko A (2004) Localization of wild-type and mutant neuronal ceroid lipofuscinosis CLN8 proteins in non-neuronal and neuronal cells. *J Neurosci Res* 76:862-871.

Luiro K, Kopra O, Lehtovirta M, Jalanko A (2001) CLN3 protein is targeted to neuronal synapses but excluded from synaptic vesicles: new clues to Batten disease. *Hum Mol Genet* 10:2123-2131.

Luzio JP, Pryor PR, Bright NA (2007) Lysosomes: fusion and function. *Nat Rev Mol Cell Biol* 8:622-632.

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

Maquat LE (2004) Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics. *Nat Rev Mol Cell Biol* 5:89-99.

Martin JJ, Gottlob I, Goebel HH, Mole SE (1999) CLN4 Adult NCL. In: Goebel HH, Mole SE, Lake BD (eds) *The Neuronal Ceroid Lipofuscinosis (Batten Disease)*. IOS Press, Amsterdam, pp 77-90.

Matsuzaki H, Dong S, Loi H, Di X, Liu G, Hubbell E, Law J, Berntsen T, Chadha M, Hui H, Yang G, Kennedy GC, Webster TA, Cawley S, Walsh PS, Jones KW, Fodor SP, Mei R (2004) Genotyping over 100,000 SNPs on a pair of oligonucleotide arrays. *Nat Methods* 1:109-111.

Meikle PJ, Hopwood JJ, Clague AE, Carey WF (1999) Prevalence of lysosomal storage disorders. *Jama* 281:249-254.

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

Melville SA, Wilson CL, Chiang CS, Studdert VP, Lingaas F, Wilton AN (2005) A mutation in canine CLN5 causes neuronal ceroid lipofuscinosis in Border collie dogs. *Genomics* 86:287-294.

Metcalf P, Fusek M (1993) Two crystal structures for cathepsin D: the lysosomal targeting signal and active site. *Embo J* 12:1293-1302.

Miano MG, Jacobson SG, Carothers A, Hanson I, Teague P, Lovell J, Cideciyan AV, Haider N, Stone EM, Sheffield VC, Wright AF (2000) Pitfalls in homozygosity mapping. *Am J Hum Genet* 67:1348-1351.

Mitchell WA, Wheeler RB, Sharp JD, Bate SL, Gardiner RM, Ranta US, Lonka L, Williams RE, Lehesjoki AE, Mole SE (2001) Turkish variant late infantile neuronal ceroid lipofuscinosis (CLN7) may be allelic to CLN8. *Eur J Paediatr Neurol* 5 Suppl A:21-27.

Mitchison HM, Bernard DJ, Greene ND, Cooper JD, Junaid MA, Pullarkat RK, de Vos N, Breuning MH, Owens JW, Mobley WC, Gardiner RM, Lake BD, Taschner PE, Nussbaum RL (1999) Targeted disruption of the Cln3 gene provides a mouse model for Batten disease. The Batten Mouse Model Consortium [corrected]. *Neurobiol Dis* 6:321-334.

Mitchison HM, Hofmann SL, Becerra CH, Munroe PB, Lake BD, Crow YJ, Stephenson JB, Williams RE, Hofman IL, Taschner PE, Martin JJ, Philippart M, Andermann E, Andermann F, Mole SE, Gardiner RM, O'Rawe AM (1998) Mutations in the palmitoyl-protein thioesterase gene (PPT; CLN1) causing juvenile neuronal ceroid lipofuscinosis with granular osmiophilic deposits. *Hum Mol Genet* 7:291-297.

Mitchison HM, Taschner PE, O'Rawe AM, de Vos N, Phillips HA, Thompson AD, Kozman HM, Haines JL, Schlumpf K, D'Arigo K, Boustany RM, Callen DF, Breuning MH, Gardiner RM, Mole SE, Lerner TJ (1994) Genetic mapping of the Batten disease locus (CLN3) to the interval D16S288-D16S383 by analysis of haplotypes and allelic association. *Genomics* 22:465-468.

Mohamadzadeh M, Mohamadzadeh H, Brammer M, Sestak K, Luftig RB (2004) Identification of proteases employed by dendritic cells in the processing of protein purified derivative (PPD). *J Immune Based Ther Vaccines* 2:8.

Mole SE (2004) The genetic spectrum of human neuronal ceroid-lipofuscinoses. *Brain Pathol* 14:70-76.

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

Mole SE, Williams RE, Goebel HH (2005) Correlations between genotype, ultrastructural morphology and clinical phenotype in the neuronal ceroid lipofuscinoses. *Neurogenetics* 6:107-126.

Morin P, Sagne C, Gasnier B (2004) Functional characterization of wild-type and mutant human sialin. *Embo J* 23:4560-4570.

Morton NE (1955) Sequential tests for the detection of linkage. *Am J Hum Genet* 7:277-318.

Morton NE (2005) Linkage disequilibrium maps and association mapping. *J Clin Invest* 115:1425-1430.

Muller U (1999) Ten years of gene targeting: targeted mouse mutants, from vector design to phenotype analysis. *Mech Dev* 82:3-21.

Munroe PB, Mitchison HM, O'Rawe AM, Anderson JW, Boustany RM, Lerner TJ, Taschner PE, de Vos N, Breuning MH, Gardiner RM, Mole SE (1997) Spectrum of mutations in the Batten disease gene, CLN3. *Am J Hum Genet* 61:310-316.

Myllykangas L, Tyynelä J, Page-McCaw A, Rubin GM, Haltia MJ, Feany MB (2005) Cathepsin D-deficient Drosophila recapitulate the key features of neuronal ceroid lipofuscinoses. *Neurobiol Dis* 19:194-199.

Nijssen PC, Brusse E, Leyten AC, Martin JJ, Teepen JL, Roos RA (2002) Autosomal dominant adult neuronal ceroid lipofuscinosis: parkinsonism due to both striatal and nigral dysfunction. *Mov Disord* 17:482-487.

Nijssen PC, Ceuterick C, van Diggelen OP, Elleder M, Martin JJ, Teepen JL, Tyynela J, Roos RA (2003) Autosomal dominant adult neuronal ceroid lipofuscinosis: a novel form of NCL with granular osmiophilic deposits without palmitoyl protein thioesterase 1 deficiency. *Brain Pathol* 13:574-581.

Nishino I, Fu J, Tanji K, Yamada T, Shimojo S, Koori T, Mora M, Riggs JE, Oh SJ, Koga Y, Sue CM, Yamamoto A, Murakami N, Shanske S, Byrne E, Bonilla E, Nonaka I, DiMauro S, Hirano M (2000) Primary LAMP-2 deficiency causes X-linked vacuolar cardiomyopathy and myopathy (Danon disease). *Nature* 406:906-910.

Norman RM, Wood N (1941) A congenital form of amaurotic family idiocy. *Journal of Neurology and Psychiatry* 4:175-190.

Palmer DN, Fearnley IM, Walker JE, Hall NA, Lake BD, Wolfe LS, Haltia M, Martinus RD, Jolly RD (1992) Mitochondrial ATP synthase subunit c storage in the ceroid-lipofuscinoses (Batten disease). *Am J Med Genet* 42:561-567.

Pao SS, Paulsen IT, Saier MH, Jr. (1998) Major facilitator superfamily. *Microbiol Mol Biol Rev* 62:1-34.

Papassotiropoulos A, Bagli M, Feder O, Jessen F, Maier W, Rao ML, Ludwig M, Schwab SG, Heun R (1999) Genetic polymorphism of cathepsin D is strongly associated with the risk for developing sporadic Alzheimer's disease. *Neurosci Lett* 262:171-174.

Partanen S, Haapanen A, Kielar C, Pontikis C, Alexander N, Inkinen T, Saftig P, Gillingwater TH, Cooper JD, Tyynela J (2008) Synaptic changes in the thalamocortical system of cathepsin D-deficient mice: a model of human congenital neuronal ceroid-lipofuscinosis. *J Neuropathol Exp Neurol* 67:16-29.

Partanen S, Storch S, Loffler HG, Hasilik A, Tyynelä J, Braulke T (2003) A replacement of the active-site aspartic acid residue 293 in mouse cathepsin D affects its intracellular stability, processing and transport in HEK-293 cells. *Biochem J* 369:55-62.

Patton-Vogt JL, Henry SA (1998) GIT1, a gene encoding a novel transporter for glycerophosphoinositol in Saccharomyces cerevisiae. *Genetics* 149:1707-1715.

Pearce DA, Sherman F (1997) BTN1, a yeast gene corresponding to the human gene responsible for Batten's disease, is not essential for viability, mitochondrial function, or degradation of mitochondrial ATP synthase. *Yeast* 13:691-697.

Peden AA, Oorschot V, Hesser BA, Austin CD, Scheller RH, Klumperman J (2004) Localization of the AP-3 adaptor complex defines a novel endosomal exit site for lysosomal membrane proteins. *J Cell Biol* 164:1065-1076.

Peltonen L, Jalanko A, Varilo T (1999) Molecular genetics of the Finnish disease heritage. *Hum Mol Genet* 8:1913-1923.

Peltonen L, Palotie A, Lange K (2000a) Use of population isolates for mapping complex traits. *Nat Rev Genet* 1:182-190.

Peltonen L, Savukoski M, Vesa J (2000b) Genetics of the neuronal ceroid lipofuscinoses. *Curr Opin Genet Dev* 10:299-305.

Peters C, Braun M, Weber B, Wendland M, Schmidt B, Pohlmann R, Waheed A, von Figura K (1990) Targeting of a lysosomal membrane protein: a tyrosine-containing endocytosis signal in the cytoplasmic tail of lysosomal acid phosphatase is necessary and sufficient for targeting to lysosomes. *Embo J* 9:3497-3506.

Pineda-Trujillo N, Cornejo W, Carrizosa J, Wheeler RB, Munera S, Valencia A, Agudelo-Arango J, Cogollo A, Anderson G, Bedoya G, Mole SE, Ruiz-Linares A (2005) A CLN5 mutation causing an atypical neuronal ceroid lipofuscinosis of juvenile onset. *Neurology* 64:740-742.

Poet M, Kornak U, Schweizer M, Zdebik AA, Scheel O, Hoelter S, Wurst W, Schmitt A, Fuhrmann JC, Planells-Cases R, Mole SE, Hubner CA, Jentsch TJ (2006) Lysosomal storage disease upon disruption of the neuronal chloride transport protein ClC-6. *Proc Natl Acad Sci U S A* 103:13854-13859.

Poorthuis BJ, Wevers RA, Kleijer WJ, Groener JE, de Jong JG, van Weely S, Niezen-Koning KE, van Diggelen OP (1999) The frequency of lysosomal storage diseases in The Netherlands. *Hum Genet* 105:151-156.

Press EM, Porter RR, Cebra J (1960) The isolation and properties of a proteolytic enzyme, cathepsin D, from bovine spleen. *Biochem J* 74:501-514.

Pryor PR, Mullock BM, Bright NA, Gray SR, Luzio JP (2000) The role of intraorganellar Ca(2+) in late endosome-lysosome heterotypic fusion and in the reformation of lysosomes from hybrid organelles. *J Cell Biol* 149:1053-1062.

Rampersaud E, Damcott CM, Fu M, Shen H, McArdle P, Shi X, Shelton J, Yin J, Chang YP, Ott SH, Zhang L, Zhao Y, Mitchell BD, O'Connell J, Shuldiner AR (2007) Identification of novel candidate genes for type 2 diabetes from a genome-wide association scan in the Old Order Amish: evidence for replication from diabetes-related quantitative traits and from independent populations. *Diabetes* 56:3053-3062.

Ranta S, Topcu M, Tegelberg S, Tan H, Ustubutun A, Saatci I, Dufke A, Enders H, Pohl K, Alembik Y, Mitchell WA, Mole SE, Lehesjoki AE (2004) Variant late infantile neuronal ceroid lipofuscinosis in a subset of Turkish patients is allelic to Northern epilepsy. *Hum Mutat* 23:300-305.

Ranta S, Zhang Y, Ross B, Lonka L, Takkunen E, Messer A, Sharp J, Wheeler R, Kusumi K, Mole S, Liu W, Soares MB, Bonaldo MF, Hirvasniemi A, de la Chapelle A, Gilliam TC, Lehesjoki AE (1999) The neuronal ceroid lipofuscinoses in human EPMR and mnd mutant mice are associated with mutations in CLN8. *Nat Genet* 23:233-236.

Rawlings ND, Barrett AJ (1995) Families of aspartic peptidases, and those of unknown catalytic mechanism. *Methods Enzymol* 248:105-120.

Rawlings ND, Barrett AJ (1999) Tripeptidyl-peptidase I is apparently the CLN2 protein absent in classical late-infantile neuronal ceroid lipofuscinosis. *Biochim Biophys Acta* 1429:496-500.

Redecker B, Heckendorf B, Grosch HW, Mersmann G, Hasilik A (1991) Molecular organization of the human cathepsin D gene. *DNA Cell Biol* 10:423-431.

Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, Cho EK, Dallaire S, Freeman JL, Gonzalez JR, Gratacos M, Huang J, Kalaitzopoulos D, Komura D, MacDonald JR, Marshall CR, Mei R, Montgomery L, Nishimura K, Okamura K, Shen F, Somerville MJ, Tchinda J, Valsesia A, Woodwark C, Yang F, Zhang J, Zerjal T, Zhang J, Armengol L, Conrad DF, Estivill X, Tyler-Smith C, Carter NP, Aburatani H, Lee C, Jones KW, Scherer SW, Hurles ME (2006) Global variation in copy number in the human genome. *Nature* 444:444-454.

Reece M, Prawitt D, Landers J, Kast C, Gros P, Housman D, Zabel BU, Pelletier J (1998) Functional characterization of ORCTL2--an organic cation transporter expressed in the renal proximal tubules. *FEBS Lett* 433:245-250.

Rider JA, Rider DL (1988) Batten disease: past, present, and future. *Am J Med Genet Suppl* 5:21-26.

Rijnboutt S, Kal AJ, Geuze HJ, Aerts H, Strous GJ (1991) Mannose 6-phosphate-independent targeting of cathepsin D to lysosomes in HepG2 cells. *J Biol Chem* 266:23586-23592.

Robson KJ, Chandra T, MacGillivray RT, Woo SL (1982) Polysome immunoprecipitation of phenylalanine hydroxylase mRNA from rat liver and cloning of its cDNA. *Proc Natl Acad Sci U S A* 79:4701-4705.

Rodriguez GM, Diment S (1992) Role of cathepsin D in antigen presentation of ovalbumin. *J Immunol* 149:2894-2898.

Rous BA, Reaves BJ, Ihrke G, Briggs JA, Gray SR, Stephens DJ, Banting G, Luzio JP (2002) Role of adaptor complex AP-3 in targeting wild-type and mutated CD63 to lysosomes. *Mol Biol Cell* 13:1071-1082.

Royer-Pokora B, Kunkel LM, Monaco AP, Goff SC, Newburger PE, Baehner RL, Cole FS, Curnutte JT, Orkin SH (1986) Cloning the gene for an inherited human disorder--chronic granulomatous disease--on the basis of its chromosomal location. *Nature* 322:32-38.

Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, Mullikin JC, Mortimore BJ, Willey DL, Hunt SE, Cole CG, Coggill PC, Rice CM, Ning Z, Rogers J, Bentley DR, Kwok PY, Mardis ER, Yeh RT, Schultz B, Cook L, Davenport R, Dante M,

Fulton L, Hillier L, Waterston RH, McPherson JD, Gilman B, Schaffner S, Van Etten WJ, Reich D, Higgins J, Daly MJ, Blumenstiel B, Baldwin J, Stange-Thomann N, Zody MC, Linton L, Lander ES, Altshuler D (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 409:928-933.

Saftig P, Hetman M, Schmahl W, Weber K, Heine L, Mossmann H, Koster A, Hess B, Evers M, von Figura K, et al. (1995) Mice deficient for the lysosomal proteinase cathepsin D exhibit progressive atrophy of the intestinal mucosa and profound destruction of lymphoid cells. *Embo J* 14:3599-3608.

Sagne C, Agulhon C, Ravassard P, Darmon M, Hamon M, El Mestikawy S, Gasnier B, Giros B (2001) Identification and characterization of a lysosomal transporter for small neutral amino acids. *Proc Natl Acad Sci U S A* 98:7206-7211.

Sakai H, Tanaka Y, Tanaka M, Ban N, Yamada K, Matsumura Y, Watanabe D, Sasaki M, Kita T, Inagaki N (2007) ABCA2 deficiency results in abnormal sphingolipid metabolism in mouse brain. *J Biol Chem* 282:19692-19699.

Salonen T, Heinonen-Kopra O, Vesa J, Jalanko A (2001) Neuronal trafficking of palmitoyl protein thioesterase provides an excellent model to study the effects of different mutations which cause infantile neuronal ceroid lipofuscinocis. *Mol Cell Neurosci* 18:131-140.

Samarel AM, Ferguson AG, Decker RS, Lesch M (1989) Effects of cysteine protease inhibitors on rabbit cathepsin D maturation. *Am J Physiol* 257:C1069-1079.

Sandbank U (1968) Congenital amaurotic idiocy. *Pathol Eur* 3:226-229.

Santavuori P, Haltia M, Rapola J (1974) Infantile type of so-called neuronal ceroid-lipofuscinosis. *Dev Med Child Neurol* 16:644-653.

Santavuori P, Haltia M, Rapola J, Raitta C (1973) Infantile type of so-called neuronal ceroid-lipofuscinosis. 1. A clinical study of 15 patients. *J Neurol Sci* 18:257-267.

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

Santavuori P, Rapola J, Haltia M, Tyynelä J, Peltonen L, Mole SE (1999) CLN5 Finnish Variant Late Infantile NCL. In: Goebel HH, Mole SE, Lake BD (eds) *The Neuronal Ceroid Lipofuscinosis (Batten Disease)*. IOS Press, Amsterdam, pp 91-101.

Santavuori P, Rapola J, Nuutila A, Raininko R, Lappi M, Launes J, Herva R, Sainio K (1991) The spectrum of Jansky-Bielschowsky disease. *Neuropediatrics* 22:92-96.

Santavuori P, Rapola J, Sainio K, Raitta C (1982) A variant of Jansky-Bielschowsky disease. *Neuropediatrics* 13:135-141.

Savukoski M, Kestilä M, Williams R, Järvelä I, Sharp J, Harris J, Santavuori P, Gardiner M, Peltonen L (1994) Defined chromosomal assignment of CLN5 demonstrates that at least four genetic loci are involved in the pathogenesis of human ceroid lipofuscinoses. *Am J Hum Genet* 55:695-701.

Savukoski M, Klockars T, Holmberg V, Santavuori P, Lander ES, Peltonen L (1998) CLN5, a novel gene encoding a putative transmembrane protein mutated in Finnish variant late infantile neuronal ceroid lipofuscinosis. *Nat Genet* 19:286-288.

Schouten JP, McElgunn CJ, Waaijer R, Zwijnenburg D, Diepvens F, Pals G (2002) Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Res* 30:e57.

Schröder B, Wrocklage C, Pan C, Jäger R, Kösters B, Schäfer H, Elsässer HP, Mann M, Hasilik A (2007) Integral and associated lysosomal membrane proteins. *Traffic* 8:1676-1686.

Schulz A, Dhar S, Rylova S, Dbaibo G, Alroy J, Hagel C, Artacho I, Kohlschutter A, Lin S, Boustany RM (2004) Impaired cell adhesion and apoptosis in a novel CLN9 Batten disease variant. *Ann Neurol* 56:342-350.

Schulz A, Mousallem T, Venkataramani M, Persaud-Sawin DA, Zucker A, Luberto C, Bielawska A, Bielawski J, Holthuis JC, Jazwinski SM, Kozhaya L, Dbaibo GS, Boustany RM (2006) The CLN9 Protein, a Regulator of Dihydroceramide Synthase. *J Biol Chem* 281:2784-2794.

Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Maner S, Massa H, Walker M, Chi M, Navin N, Lucito R, Healy J, Hicks J, Ye K, Reiner A, Gilliam TC, Trask B, Patterson N, Zetterberg A, Wigler M (2004) Large-scale copy number polymorphism in the human genome. *Science* 305:525-528.

Sharp JD, Wheeler RB, Lake BD, Savukoski M, Järvelä IE, Peltonen L, Gardiner RM, Williams RE (1997) Loci for classical and a variant late infantile neuronal ceroid lipofuscinosis map to chromosomes 11p15 and 15q21-23. *Hum Mol Genet* 6:591-595.

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

Siintola E, Lehesjoki AE, Mole SE (2006) Molecular genetics of the NCLs -- status and perspectives. *Biochim Biophys Acta* 1762:857-864.

Sleat DE, Donnelly RJ, Lackland H, Liu CG, Sohar I, Pullarkat RK, Lobel P (1997) Association of mutations in a lysosomal protein with classical late-infantile neuronal ceroid lipofuscinosis. *Science* 277:1802-1805.

Sleat DE, Gin RM, Sohar I, Wisniewski K, Sklower-Brooks S, Pullarkat RK, Palmer DN, Lerner TJ, Boustany RM, Uldall P, Siakotos AN, Donnelly RJ, Lobel P (1999) Mutational analysis of the defective protease in classic late-infantile neuronal ceroid lipofuscinosis, a neurodegenerative lysosomal storage disorder. *Am J Hum Genet* 64:1511-1523.

Sleat DE, Lackland H, Wang Y, Sohar I, Xiao G, Li H, Lobel P (2005) The human brain mannose 6-phosphate glycoproteome: a complex mixture composed of multiple isoforms of many soluble lysosomal proteins. *Proteomics* 5:1520-1532.

Sleat DE, Wiseman JA, El-Banna M, Kim KH, Mao Q, Price S, Macauley SL, Sidman RL, Shen MM, Zhao Q, Passini MA, Davidson BL, Stewart GR, Lobel P (2004) 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* 24:9117-9126.

Soares-Silva I, Schuller D, Andrade RP, Baltazar F, Cassio F, Casal M (2003) Functional expression of the lactate permease Jen1p of Saccharomyces cerevisiae in Pichia pastoris. *Biochem J* 376:781-787.

Sonnhammer EL, Eddy SR, Durbin R (1997) Pfam: a comprehensive database of protein domain families based on seed alignments. *Proteins* 28:405-420.

Stamm S, Ben-Ari S, Rafalska I, Tang Y, Zhang Z, Toiber D, Thanaraj TA, Soreq H (2005) Function of alternative splicing. *Gene* 344:1-20.

Steinfeld R, Heim P, von Gregory H, Meyer K, Ullrich K, Goebel HH, Kohlschutter A (2002) Late infantile neuronal ceroid lipofuscinosis: quantitative description of the clinical course in patients with CLN2 mutations. *Am J Med Genet* 112:347-354.

Steinfeld R, Reinhardt K, Schreiber K, Hillebrand M, Kraetzner R, Bruck W, Saftig P, Gartner J (2006) Cathepsin d deficiency is associated with a human neurodegenerative disorder. *Am J Hum Genet* 78:988-998.

Steinfeld R, Steinke HB, Isbrandt D, Kohlschutter A, Gartner J (2004) Mutations in classical late infantile neuronal ceroid lipofuscinosis disrupt transport of tripeptidyl-peptidase I to lysosomes. *Hum Mol Genet* 13:2483-2491.

Syvänen AC (2005) Toward genome-wide SNP genotyping. *Nat Genet* 37 Suppl:S5-10.

Tahvanainen E, Ranta S, Hirvasniemi A, Karila E, Leisti J, Sistonen P, Weissenbach J, Lehesjoki AE, de la Chapelle A (1994) The gene for a recessively inherited human childhood

progressive epilepsy with mental retardation maps to the distal short arm of chromosome 8. *Proc Natl Acad Sci U S A* 91:7267-7270.

Tammen I, Houweling PJ, Frugier T, Mitchell NL, Kay GW, Cavanagh JA, Cook RW, Raadsma HW, Palmer DN (2006) 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* 1762:898-905.

Tanaka Y, Guhde G, Suter A, Eskelinen EL, Hartmann D, Lullmann-Rauch R, Janssen PM, Blanz J, von Figura K, Saftig P (2000) Accumulation of autophagic vacuoles and cardiomyopathy in LAMP-2-deficient mice. *Nature* 406:902-906.

Tang CH, Lee JW, Galvez MG, Robillard L, Mole SE, Chapman HA (2006) Murine cathepsin f deficiency causes neuronal lipofuscinosis and late-onset neurological disease. *Mol Cell Biol* 26:2309-2316.

Teixeira CA, Espinola J, Huo L, Kohlschutter J, Persaud Sawin DA, Minassian B, Bessa CJ, Guimaraes A, Stephan DA, Sa Miranda MC, MacDonald ME, Ribeiro MG, Boustany RM (2003) Novel mutations in the CLN6 gene causing a variant late infantile neuronal ceroid lipofuscinosis. *Hum Mutat* 21:502-508.

Terwilliger JD, Ott J (1994) *Handbook of Human Genetic Linkage*. The Johns Hopkins University Press, Baltimore, Maryland, USA.

The International Batten Disease Consortium (1995) Isolation of a novel gene underlying Batten disease, CLN3. *Cell* 82:949-957.

Tomkinson B (1999) Tripeptidyl peptidases: enzymes that count. *Trends Biochem Sci* 24:355-359.

Topcu M, Jobard F, Halliez S, Coskun T, Yalcinkayal C, Gerceker FO, Wanders RJ, Prud'homme JF, Lathrop M, Ozguc M, Fischer J (2004a) L-2-Hydroxyglutaric aciduria: identification of a mutant gene C14orf160, localized on chromosome 14q22.1. *Hum Mol Genet* 13:2803-2811.

Topcu M, Tan H, Yalnizoglu D, Usubutun A, Saatci I, Aynaci M, Anlar B, Topaloglu H, Turanli G, Kose G, Aysun S (2004b) Evaluation of 36 patients from Turkey with neuronal ceroid lipofuscinosis: clinical, neurophysiological, neuroradiological and histopathologic studies. *Turk J Pediatr* 46:1-10.

Town M, Jean G, Cherqui S, Attard M, Forestier L, Whitmore SA, Callen DF, Gribouval O, Broyer M, Bates GP, van't Hoff W, Antignac C (1998) A novel gene encoding an integral membrane protein is mutated in nephropathic cystinosis. *Nat Genet* 18:319-324.

Tyynelä J, Palmer DN, Baumann M, Haltia M (1993) Storage of saposins A and D in infantile neuronal ceroid-lipofuscinosis. *FEBS Lett* 330:8-12.

Tyynelä J, Sohar I, Sleat DE, Gin RM, Donnelly RJ, Baumann M, Haltia M, Lobel P (2000) A mutation in the ovine cathepsin D gene causes a congenital lysosomal storage disease with profound neurodegeneration. *Embo J* 19:2786-2792.

Tyynelä J, Suopanki J, Santavuori P, Baumann M, Haltia M (1997) Variant late infantile neuronal ceroid-lipofuscinosis: pathology and biochemistry. *J Neuropathol Exp Neurol* 56:369-375.

van Diggelen OP, Thobois S, Tilikete C, Zabot MT, Keulemans JL, van Bunderen PA, Taschner PE, Losekoot M, Voznyi YV (2001) Adult neuronal ceroid lipofuscinosis with palmitoyl-protein thioesterase deficiency: first adult-onset patients of a childhood disease. *Ann Neurol* 50:269-272.

van Driel MA, Bruggeman J, Vriend G, Brunner HG, Leunissen JA (2006) A text-mining analysis of the human phenome. *Eur J Hum Genet* 14:535-542.

Vellodi A (2005) Lysosomal storage disorders. *Br J Haematol* 128:413-431.

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q,

Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, Di Francesco V, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji RR, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Wang Z, Wang A, Wang X, Wang J, Wei M, Wides R, Xiao C, Yan C, Yao A, Ye J, Zhan M, Zhang W, Zhang H, Zhao Q, Zheng L, Zhong F, Zhong W, Zhu S, Zhao S, Gilbert D, Baumhueter S, Spier G, Carter C, Cravchik A, Woodage T, Ali F, An H, Awe A, Baldwin D, Baden H, Barnstead M, Barrow I, Beeson K, Busam D, Carver A, Center A, Cheng ML, Curry L, Danaher S, Davenport L, Desilets R, Dietz S, Dodson K, Doup L, Ferriera S, Garg N, Gluecksmann A, Hart B, Haynes J, Haynes C, Heiner C, Hladun S, Hostin D, Houck J, Howland T, Ibegwam C, Johnson J, Kalush F, Kline L, Koduru S, Love A, Mann F, May D, McCawley S, McIntosh T, McMullen I, Moy M, Moy L, Murphy B, Nelson K, Pfannkoch C, Pratts E, Puri V, Qureshi H, Reardon M, Rodriguez R, Rogers YH, Romblad D, Ruhfel B, Scott R, Sitter C, Smallwood M, Stewart E, Strong R, Suh E, Thomas R, Tint NN, Tse S, Vech C, Wang G, Wetter J, Williams S, Williams M, Windsor S, Winn-Deen E, Wolfe K, Zaveri J, Zaveri K, Abril JF, Guigo R, Campbell MJ, Sjolander KV, Karlak B, Kejariwal A, Mi H, Lazareva B, Hatton T, Narechania A, Diemer K, Muruganujan A, Guo N, Sato S, Bafna V, Istrail S, Lippert R, Schwartz R, Walenz B, Yooseph S, Allen D, Basu A, Baxendale J, Blick L, Caminha M, Carnes-Stine J, Caulk P, Chiang YH, Coyne M, Dahlke C, Mays A, Dombroski M, Donnelly M, Ely D, Esparham S, Fosler C, Gire H, Glanowski S, Glasser K, Glodek A, Gorokhov M, Graham K, Gropman B, Harris M, Heil J, Henderson S, Hoover J, Jennings D, Jordan C, Jordan J, Kasha J, Kagan L, Kraft C, Levitsky A, Lewis M, Liu X, Lopez J, Ma D, Majoros W, McDaniel J, Murphy S, Newman M, Nguyen T, Nguyen N, Nodell M, Pan S, Peck J, Peterson M, Rowe W, Sanders R, Scott J, Simpson M, Smith T, Sprague A, Stockwell T, Turner R, Venter E, Wang M, Wen M, Wu D, Wu M, Xia A, Zandieh A, Zhu X (2001) The sequence of the human genome. *Science* 291:1304-1351.

Verheijen FW, Mancini GSM (2005) Lysosomal transporters and associated diseases. In: Saftig P (ed) *Lysosomes*. Landes Bioscience / Eurekah.com, Georgetown, Texas, USA, pp 74-81.

Verheijen FW, Verbeek E, Aula N, Beerens CE, Havelaar AC, Joosse M, Peltonen L, Aula P, Galjaard H, van der Spek PJ, Mancini GM (1999) A new gene, encoding an anion transporter, is mutated in sialic acid storage diseases. *Nat Genet* 23:462-465.

Verkruyse LA, Hofmann SL (1996) Lysosomal targeting of palmitoyl-protein thioesterase. *J Biol Chem* 271:15831-15836.

Vesa J, Chin MH, Oelgeschlager K, Isosomppi J, DellAngelica EC, Jalanko A, Peltonen L (2002) Neuronal ceroid lipofuscinoses are connected at molecular level: interaction of CLN5 protein with CLN2 and CLN3. *Mol Biol Cell* 13:2410-2420.

Vesa J, Hellsten E, Verkruyse LA, Camp LA, Rapola J, Santavuori P, Hofmann SL, Peltonen L (1995) Mutations in the palmitoyl protein thioesterase gene causing infantile neuronal ceroid lipofuscinosis. *Nature* 376:584-587.

Vines DJ, Warburton MJ (1999) Classical late infantile neuronal ceroid lipofuscinosis fibroblasts are deficient in lysosomal tripeptidyl peptidase I. *FEBS Lett* 443:131-135.

Vulevic B, Chen Z, Boyd JT, Davis W, Jr., Walsh ES, Belinsky MG, Tew KD (2001) Cloning and characterization of human adenosine 5'-triphosphate-binding cassette, sub-family A, transporter 2 (ABCA2). *Cancer Res* 61:3339-3347.

Wheeler RB, Sharp JD, Mitchell WA, Bate SL, Williams RE, Lake BD, Gardiner RM (1999) A new locus for variant late infantile neuronal ceroid lipofuscinosis-CLN7. *Mol Genet Metab* 66:337-338.

Wheeler RB, Sharp JD, Schultz RA, Joslin JM, Williams RE, Mole SE (2002) The gene mutated in variant late-infantile neuronal ceroid lipofuscinosis (CLN6) and in nclf mutant mice encodes a novel predicted transmembrane protein. *Am J Hum Genet* 70:537-542.

Williams MA, Fukuda M (1990) Accumulation of membrane glycoproteins in lysosomes requires a tyrosine residue at a particular position in the cytoplasmic tail. *J Cell Biol* 111:955-966.

Williams RE, Aberg L, Autti T, Goebel HH, Kohlschutter A, Lonnqvist T (2006) Diagnosis of the neuronal ceroid lipofuscinoses: an update. *Biochim Biophys Acta* 1762:865-872.

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

Winchester BG (2001) Lysosomal membrane proteins. *Eur J Paediatr Neurol* 5 Suppl A:11-19.

Winter E, Ponting CP (2002) TRAM, LAG1 and CLN8: members of a novel family of lipid-sensing domains? *Trends Biochem Sci* 27:381-383.

Wisniewski KE, Kaczmarski A, Kida E, Connell F, Kaczmarski W, Michalewski MP, Moroziewicz DN, Zhong N (1999) Reevaluation of neuronal ceroid lipofuscinoses: atypical juvenile onset may be the result of CLN2 mutations. *Mol Genet Metab* 66:248-252.

Wisniewski KE, Zhong N, Kaczmarski W, Kaczmarski A, Sklower-Brooks S, Brown WT (1998) Studies of atypical JNCL suggest overlapping with other NCL forms. *Pediatr Neurol* 18:36-40.

Wlodawer A, Li M, Dauter Z, Gustchina A, Uchida K, Oyama H, Dunn BM, Oda K (2001) Carboxyl proteinase from Pseudomonas defines a novel family of subtilisin-like enzymes. *Nat Struct Biol* 8:442-446.

Wolters JC, Abele R, Tampe R (2005) Selective and ATP-dependent translocation of peptides by the homodimeric ATP binding cassette transporter TAP-like (ABCB9). *J Biol Chem* 280:23631-23636.

Woods CG, Cox J, Springell K, Hampshire DJ, Mohamed MD, McKibbin M, Stern R, Raymond FL, Sandford R, Malik Sharif S, Karbani G, Ahmed M, Bond J, Clayton D, Inglehearn CF (2006) Quantification of homozygosity in consanguineous individuals with autosomal recessive disease. *Am J Hum Genet* 78:889-896.

Xiao W, Oefner PJ (2001) Denaturing high-performance liquid chromatography: A review. *Hum Mutat* 17:439-474.

Yoshikawa M, Uchida S, Ezaki J, Rai T, Hayama A, Kobayashi K, Kida Y, Noda M, Koike M, Uchiyama Y, Marumo F, Kominami E, Sasaki S (2002) CLC-3 deficiency leads to phenotypes similar to human neuronal ceroid lipofuscinosis. *Genes Cells* 7:597-605.

Zelnik N, Mahajna M, Iancu TC, Sharony R, Zeigler M (2007) A novel mutation of the CLN8 gene: is there a Mediterranean phenotype? *Pediatr Neurol* 36:411-413.

Zhang F, Zhang W, Liu L, Fisher CL, Hui D, Childs S, Dorovini-Zis K, Ling V (2000) Characterization of ABCB9, an ATP binding cassette protein associated with lysosomes. *J Biol Chem* 275:23287-23294.

Zhou D, Li P, Lin Y, Lott JM, Hislop AD, Canaday DH, Brutkiewicz RR, Blum JS (2005) Lamp-2a facilitates MHC class II presentation of cytoplasmic antigens. *Immunity* 22:571-581.