Kansanterveyslaitos
Folkhälsoinstitutet
National Public Health Institute

Publications of the National Public Health Institute A 2 / 2004

Ville Holmberg

CLN5 - FROM MUTATION TO DEFECTIVE PROTEIN AND CLINICAL PHENOTYPE

Department of Molecular Medicine,
National Public Health Institute, Helsinki, Finland and
Department of Medical Genetics,
University of Helsinki, Finland and
Hospital for Children and Adolescents
University of Helsinki, Finland

Helsinki 2004

CLN5 - FROM MUTATION TO DEFECTIVE PROTEIN AND CLINICAL PHENOTYPE

Ville Holmberg

Department of Molecular Medicine
National Public Health Institute
Helsinki, Finland

Department of Medical Genetics
University of Helsinki

Hospital for Children and Adolescents
University of Helsinki

Academic Dissertation

To be publicly discussed with the permission of the Medical Faculty of the University of Helsinki, in Lecture Hall 2 of Biomedicum Helsinki, Haartmaninkatu 8, on March 12th 2004, at 12 noon.

Helsinki 2004

Supervisor
Professor Leena Peltonen-Palotie
National Public Health Institute and University of Helsinki

Reviewers
Professor Helena Kääriäinen
University of Turku

Professor Kari Majamaa
University of Oulu

Opponent
Professor Mark Gardiner
University College London

Julkaisija - Utgivare - Publisher

Kansanterveyslaitos (KTL)
Mannerheimintie 166
00300 Helsinki
Puh. vaihde (09) 474 41, telefax (09) 4744 8408

Folkhälsoinstitutet
Mannerheimvägen 166
00300 Helsingfors
Tel. växel (09) 474 41, telefax (09) 4744 8408

National Public Health Institute
Mannerheimintie 166
FIN-00300 Helsinki, Finland
Telephone +358 9 474 41, telefax +358 9 4744 8408

Publications of the National Public Health Institute, KTL A2/2004
Copyright National Public Health Institute

ISBN 951-740-409-3
ISSN 0359-3584
ISBN 951-740-410-7 (PDF version)
ISSN 1458-6290 (PDF version)
http://ethesis.helsinki.fi

Helsinki University Biomedical Dissertations No. 50, ISSN 1457-8433

Cosmoprint Oy
Helsinki 2004

"Everyone is responsible to everyone for everything"
- Fyodor Dostoevsky

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.

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

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

III Holmberg, V, Jalanko, A, Isosomppi, J, Fabritius, AL, Peltonen, L and Kopra, O (2004) The mouse ortholog of the neuronal ceroid lipofuscinosis *CLN5* gene encodes a soluble lysosomal glycoprotein expressed in the developing brain. Neurobiology of Disease. In press.

Publication I has previously appeared in the theses of Tuomas Klockars (1998) and Minna Savukoski (1999).

ABBREVIATIONS

aa	amino acid
CL	curvilinear profiles
CLN5	ceroid lipofuscinosis, neuronal 5
cM	centi Morgan
CNS	central nervous system
COS-1	African green monkey kidney cells
CT	computed tomography
CVS	chorionic villus sample
EPMR	epilepsy with mental retardation
ERG	electroretinogram
EST	expressed sequence tag
FDH	Finnish Disease Heritage
FISH	fluorescence in situ hybridization
FP	fingerprint bodies
fVEP	flash visual evoked potential
Gb	gigabase
GROD	granular osmiophilic deposits
HeLa	Human cancer cell line
HGP	Human Genome Project
JBVD	variant Jansky-Bielschowsky disease
kb	kilobase
kDa	kilo Dalton
LD	linkage disequilibrium
Mb	megabase
MEG	magnetoencephalography
MIM	Mendelian Inheritance in Man, index number
MRI	magnetic resonance imaging
OMIM	Online Mendelian Inheritance in Man
ORF	open reading frame
PAC	P1 derived artificial chromosome
PCR	polymerase chain reaction
PKU	phenylketonuria
PPT	palmitoyl protein thioesterase
RL	rectilinear profiles
SEP	somatosensory evoked potential
VEP	visual evoked potential
vLINCL	variant late infantile neuronal ceroid lipofuscinosis
$vLINCL_{Fin}$	Finnish variant late infantile neuronal ceroid lipofuscinosis

In addition, the standard abbreviations for nucleotides and amino acids are used. The names of genes are written in italics and the names of proteins in regular letters. Human gene and protein names are capitalized, mouse gene and protein names are written in lower case letters.

ABSTRACT

The neuronal ceroid lipofuscinoses (NCLs) are the most common group of neurodegenerative disorders in childhood, with an incidence of 7-8 per 100,000 live births. They are recessively inherited lysosomal storage diseases caused by mutations in six different genes. A few years of normal development is followed by progressive motor, mental and visual decline, confinement to bed and severe disability, further complicated by frequent epileptic seizures and ultimately premature death. One of the subtypes, CLN5, also known as the Finnish variant form of late infantile neuronal ceroid lipofuscinoses, is caused by mutations in the *CLN5* gene, identified in this study by positional cloning. *CLN5* encodes a lysosomal glycoprotein whose function is so far unknown. The CLN5 disease is enriched in the isolated Finnish population and so far 29 patients have been diagnosed in Finland, one in Sweden, and one in the Netherlands. Four different mutations have been identified and the most common one is found in 88% of the Finnish disease chromosomes. The carrier frequency of this major Finnish mutation is 1:24-1:44 in the high-risk area in Western Finland, where clustering of the ancestors' birthplaces has been demonstrated. When analyzing eight patients with different mutations, no correlation could be observed between the genotype and the clinical phenotype. Thus it appears likely that each of the four mutations severely disturbs the normal function of the protein, and results in functional null alleles.

We have isolated the orthologous mouse *Cln5* gene and analyzed its spatiotemporal expression in the central nervous system by *in situ* hybridization and immunohistochemistry. *Cln5* was expressed throughout the embryonic brain and the expression steadily increased during development. In the adult mouse brain, prominent expression was detected in the cerebellar Purkinje cells, in the cerebral neurons, and in the hippocampal interneurons. These findings are well in line with the brain imaging and pathological findings in CLN5 patients, the hallmark being the severe atrophy of cerebellum. In vitro expression of *Cln5* in COS-1, HeLa and neuronal cells indicated that mouse Cln5 is a soluble lysosomal glycoprotein, closely resembling human CLN5.

REVIEW OF THE LITERATURE

1. HUMAN GENETICS

1.1 The Finnish Disease Heritage

According to current knowledge, Finland has been inhabited without interruption since the end of the latest glacial period approximately 10,000 years ago. The dual theory of the population history of Finland, supported by analyses of Y chromosome haplotypes, assumes an early migratory wave of eastern Uralic speakers some 4000 years ago, followed by more continuous immigration of small groups from the South and West during several thousand years (Jutikkala and Pirinen, 1996; Kittles et al., 1998; Norio, 2003b). The majority of the genetic profile of the current Finnish population originates from Western Europe, suggesting that language replacement occurred when the "genetic founders" arrived in the area where the "Finnish speakers" had already stabilized an agricultural lifestyle (Sajantila and Pääbo, 1995). Both maternally inherited mitochondrial sequences and paternally inherited Y chromosome haplotypes show an exceptional low level of genetic diversity of the Finns as compared with other European populations (Sajantila et al., 1996). The reason for the relative isolation of the Finnish population is mostly geographical and geopolitical, explained by its position between Sweden and Russia, two powers with distinct cultures, languages, and religions.

Around year 1500 AD the population of Finland was approximately 250,000 and was heavily concentrated along the coastline (early settlement; Figure 1). However, an internal migration movement began in the 16th century from a small southeastern area (southern Savo) to the western, central and finally eastern and northern parts of the country (Figure 1). During the regime of the Swedish king Gustavus Wasa in 1523-1560, inhabitation of the wilderness close to the Russian border was highly favored. Reasons for the internal immigration were also the need of new land for cultivation and the avoidance of increasing taxation by the Crown. This late settlement, aided by the founder effect and by genetic drift, shows clusters of regional isolates with a very homogeneous genetic background and an overrepresentation of certain rare alleles. (Norio, 2003a)

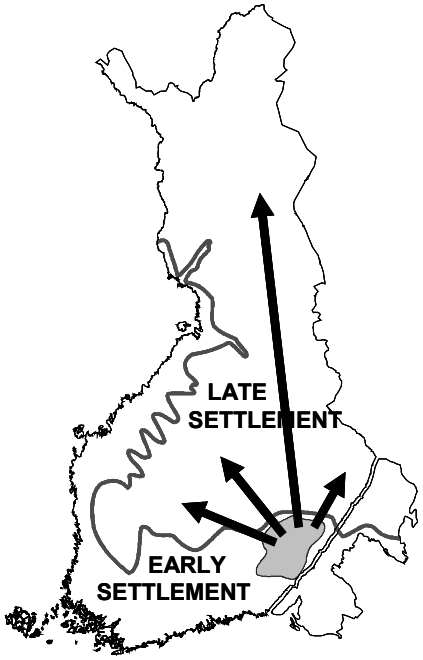

Figure 1. The internal migration movement of the 16th century resulting in regional subisolates. Modified from Peltonen et al., 1999.

The Finnish Disease Heritage (FDH) comprises at least 36 monogenic diseases: 32 autosomal recessive, two autosomal dominant, and two X-chromosomal (Norio, 2003c). The disease spectrum extends to all branches of medicine but is most visible in paediatrics. A study of 2151 samples from different parts of Finland showed that, on average, one in three individuals was a carrier of one of the 31 most common mutations (Pastinen et al., 2001). The total incidence of the FDH disorders is 1:1,000, meaning approximately 60 new cases every year in Finland per 60,000 annual births.

The FDH has its origin in the special population history of Finland. The small number of original founders, followed by isolation and rapid expansion, formed a unique gene pool. The frequency of some rare diseases became overrepresented, whereas some disease alleles that are relatively common in other countries became very rare. One of the most striking example is phenylketonuria (PKU), with an incidence of less than 1:100,000 in Finland compared to about 1:10,000 in most of Northern Europe (Guldberg et al., 1995). In Finnish subisolates, the small number of founders inevitably resulted in consanguineous marriages, although the relationships were typically remote and unknown to the individuals concerned. This random inbreeding increased the local incidence of rare recessive disorders and in some Finnish diseases a strong regional clustering of cases can still be observed. A typical feature of the FDH disorders is that one disease mutation is found in the majority of the Finnish disease chromosomes, for many diseases in more than 90% of the chromosomes (Peltonen et al., 1999).

Of the 36 Finnish disorders, the genetic loci have been established for 33 and the defective gene characterized for 27 (Norio, 2003c). This has been possible thanks to the peculiar population history, the reliable and well conserved population records kept by the church, the smooth collaboration between clinicians and researchers, and especially the positive attitude of the patients and their families towards research. The FDH has proved to be valuable for developing the strategies for identification of disease genes, and several novel genes and pathological mechanisms have been characterized. For the monogenic Finnish diseases, interest now focuses more and more on the proteins and their functions. However, the knowledge accumulated can now also be used for studying hereditary components of common multifactorial diseases in isolated populations (Peltonen et al., 2000; Varilo et al., 2000; Laitinen et al., 2001).

Four of the neuronal ceroid lipofuscinoses belong to the Finnish disease heritage. CLN5 and CLN8 are typical rare Finnish diseases with only 20-30 patients identified in Finland, while they are almost non-existent elswhere. CLN1 (INCL) is strongly enriched in Finland with about 170 pateints diagnosed here, compared to about 200 outside Finland. Approximately 200 CLN3 (JNCL) patients have been recognized in Finland, but the incidence (1:20.000) is almost as high in several other populations. (Norio, 2003c)

1.2 The Human Genome Project

The sequencing of the human genome has been a huge step towards understanding human evolution, the causation of disease, and the interplay between the environment and heredity in defining the human condition (Lander et al., 2001; Venter et al., 2001). The fundaments for decoding the mechanisms of inheritance were laid in the middle of the 19th century when Charles Darwin presented his theory of natural selection (Darwin, 1859) and Gregor Mendel established the laws of inheritance, based on experiments in plant hybridization (Mendel, 1866).

Milestones along the way have been the recognition of DNA as the hereditary material (Avery et al., 1944), the determination of its structure (Franklin and Gosling, 1953; Watson and Crick, 1953), the cracking of the genetic code (Nirenberg, 1963), the development of recombinant DNA technologies (Jackson et al., 1972; Cohen et al., 1973), and the establishment of powerful methods for DNA sequencing (Sanger and Coulson, 1975; Maxam and Gilbert, 1977; Smith et al., 1986) and amplification (Saiki et al., 1988). Thereby, the stage was set for the official international Human Genome Project to begin in 1990.

The goals of the international Human Genome Project (HGP) were to identify all the genes in human DNA, to determine the sequences of the 3 billion base pairs of the

human genome, to store this information in public databases, to improve tools for data analysis, and to address the ethical, legal, and social issues (ELSI) that may arise from the project. In April 2003, the Human Genome Project celebrated the completion of the full human genome sequence and Francis Collins announced that all of the initial objectives of the HGP had been achieved at least two years ahead of expectation. (Collins et al., 2003)

The concrete results of the HGP are the production of a genetic map and the sequencing of the human genome, the sequencing of several million expressed sequence tags (ESTs) representing transcripts of the vast majority of human genes, and keeping all this data available in public databases. During the project, several technological improvements were accomplished making PCR, sequencing and genotyping more effective through automation and miniaturization. New methods, including capillary-based sequencing, analytical methods for single-nucleotide polymorphism (SNPs) detection, and microarrays for transcript profiling have also been developed. The progress in bioinformatics and biocomputing has been invaluable for analyzing the rapidly expanding amount of information produced by high-throughput genome-wide strategies.

With all these resources now available, we have entered the genomic era and the potential of the HGP could be used to improve human health and well-being (Collins et al., 2003). Challenges for the future are to understand the geography and function of the genome, to characterize metabolic pathways and protein interactions, to monitor the spatiotemporal expression pattern of the human genes, and to develop genome-based approaches for diagnostics and predictions of individual drug response. Detailed characterization of the heritable variation in the human genome is needed to increase the understanding of traits and diseases with more complex origins, involving the interplay between multiple genetic and environmental factors (Peltonen and McKusick, 2001; Gabriel et al., 2002).

1.3 Identification of Human Disease Genes

Once the role of genetic factors in the aetiology of a disease has been established, there are several different approaches to identifying the defective genes. Functional cloning refers to the isolation of a disease gene based on knowledge of the biochemical defect underlying the disease, i.e. by purification of the defective protein and characterizing its amino acid sequence. The candidate gene approach also requires information about the pathogenesis to enable sequencing of those genes most likely to be causing a certain disease. However, experience has shown that the prediction of responsible genes based on the phenotype is very difficult. Thus, the development of map-based positional cloning was a breakthrough for medical genetics. With this strategy, disease genes can be cloned solely on the basis of their chromosomal location, and no

knowledge of their function is required. The positional candidate approach, a combination of the last two strategies, has proved to be useful as well. It is based on mutation analyses of the most promising genes in a well-defined chromosomal location, when the critical region is too big to allow sequencing of all the genes in the region.

Positional cloning has been a time-consuming but reliable method for identification of the genes underlying monogenic disorders. It includes several steps, beginning with the collection of blood samples from the families and ending with the detection of the mutations (Figure 2). The technological development and the accomplishments of the HGP during the last decade have drastically reduced the time needed for positional cloning of disease genes.

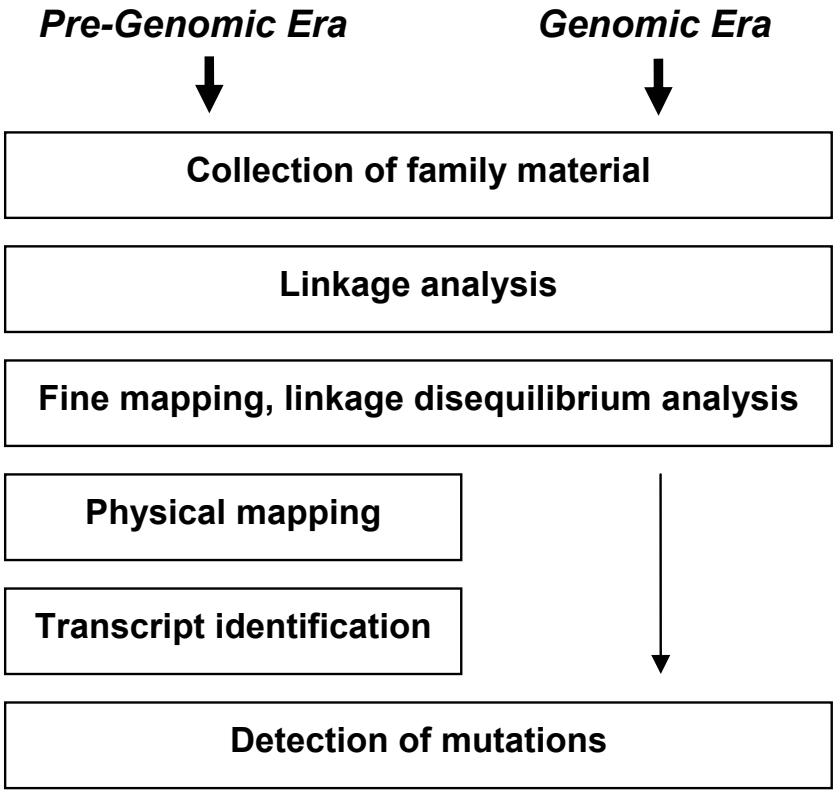

Figure 2. The different steps of the positional cloning strategy in the Pre-Genomic Era and in the current Genomic Era. Physical mapping and transcript identification is no longer needed as the Human Genome Project has made this information available in public databases.

The first step in positional cloning is the assignment of the locus of a disease gene to a specific chromosomal region. After collecting a sufficient family material carrying the inherited disease, the family members are genotyped with a set of polymorphic markers covering the whole human genome in order to be able to establish linkage

between the phenotype and a certain genetic locus. The development of dense DNA marker maps, sophisticated statistical methods and linkage analysis software have been essential for making positional cloning a useful tool for mapping the genes that are defective in human diseases (Weissenbach et al., 1992).

In isolated populations, linkage disequilibrium (LD) and haplotype sharing can be used for restriction of the chromosomal region defined by linkage analysis. It is based on the assumption that the patients share the same haplotypes in markers close to the mutation, as one single disease-causing allele has been inherited from a common ancestor. LD and the recombinations observed in the disease haplotypes often allow finer genetic mapping of the locus of interest (Hästbacka et al., 1992; de la Chapelle, 1993).

Previously, physical mapping followed by transcript identification were the steps, which could easily take several years (Figure 2). Howerver, thanks to the Human Genome Project, these steps can now be replaced in a few seconds by a single database search on the net (Genome Browser: http://genome.ucsc.edu/). Physical maps were constructed by assembling contigs of i.e. YAC and PAC clones over the critical chromosomal region. Subsequently, the coding sequences of the PAC DNA were identified by cDNA library screenings and database searches for expressed sequence tags (ESTs).

When the critical region has been restricted to contain only about ten genes or less, mutations can be detected by PCR amplification (Saiki et al., 1988) of patient and control DNA followed by automated fluorescent sequencing (Sanger and Coulson, 1975; Smith et al., 1986). Currently, cycle sequencing is commonly performed using BigDye terminator chemistry and ABI Prism DNA analyzers (Applied Biosystems).

1.4 Functional and Comparative Genomics

The genomes are dynamic structures, continually subjected to modifications by the forces of evolution. Extant animal, plant, and microbial species are the result of hundreds of millions of years of trials and errors for acclimatization to the terrestrial environment. A correct interpretation of the message written in the human genome requires a parallel understanding of the similarities and differences of genomes across species. Comparison of DNA sequences from different species has emerged as a powerful tool for identification of the functional elements in the genome. Characterization of the mechanisms of mutational processes is essential, as mutations both drive evolution and cause inherited diseases.

The first genome of a self-replicating free-living organism to be sequenced was the 1.8 million base pairs genome of the bacterium *Haemophilus influenzae* (Fleischmann et

al., 1995). The nematode *Caenorhabditis elegans* was the first animal and the first multicelluar organism whose genome was sequenced (The C. elegans Sequencing Consortium, 1998). This 1 mm worm consists of 959 somatic cells and a total number of approximately 20,000 genes. It reproduces sexually and has a nervous system that includes a simple brain. Thus it has proved useful for studies in developmental biology and neurobiology.

The common fruit fly, *Drosophila melanogaster*, has been an important model organism for experimental genetics and much knowledge about the fundamental mechanisms of genetic variation and the inheritance of specific traits has been generated by *Drosophila* research. Sequencing of the *Drosophila* genome revealed approximately 13,600 genes (Adams et al., 2000). A survey of 269 sequenced human disease genes showed that 177 of them had orthologs in the *Drosophila* genome. This emphasized the potential of comparative genomics and the role of *Drosophila* as a model organism for studies on human diseases.

The most important model organism for biomedical research, however, is the laboratory mouse, *Mus musculus*. The sequence of the mouse genome provides a key informational tool for understanding the human genome and human diseases (Waterston et al., 2002). Its unique advantages include a century of genetic studies, a profound knowledge of mouse anatomy and physiology, numerous inbred strains, hundreds of spontaneous disease models and developed techniques for engineering of the genome through transgenic, knockout and knockin strategies.

In the roughly 75 million years of evolution since the divergence of the human and mouse lineages, the process of evolution has altered the genomes unevenly so that functionally important elements can be recognized based on their higher degree of conservation (Waterston et al., 2002). By comparing the extent of sequence conservation, the proportion of the mammalian genome that is under evolutionary selection can be estimated to be about 5%. Such a high proportion cannot be explained by protein-coding sequences alone, implying that the genome contains many additional features such as untranslated regions, regulatory elements, non-protein-coding genes, and chromosomal structural elements. The mouse genome is about 14% smaller than the human genome (2.5 Gb compared with 2.9 Gb) and over 90% of the mouse and human genomes can be partitioned into corresponding regions of conserved synteny. The mouse and human genomes each seem to contain about 30,000 protein-coding genes. Approximately 80% of the human genes seem to have one single identifiable ortholog in the mouse genome. On the other hand, the proportion of human genes without any detected homologous gene in the mouse is less than 1%.

The number of genomes undergoing sequencing is increasing as the techniques are becoming faster and more reliable. So far draft sequences are available for hundreds of

viruses, tens of bacteria, and several plants, fungi, and animals. The number of genes in different species has been a surprise for many - the human has only twice as many genes as the worm or fly, and notably less than the rice with approximately 50,000 genes. However, the human genes are more complex, with more alternative splicing generating a larger number of protein products. The mechanisms of regulting gene expression may also be more sophisticated in humans than in many other species. (Lander et al., 2001)

2. NEURONAL CEROID LIPOFUSCINOSES

2.1 Historical Background

In 1896 Bernard Sachs used the term amaurotic familial idiocy (AFI) to describe a progressive neurological disorder with infant onset, psychomotor deterioration, blindness, and early death (Sachs, 1896). Fifteen years earlier Warren Tay had noticed that a retinal red spot was associated to an identical disorder, which thereby became known as Tay-Sachs disease. In the beginning of the 20th century neurologists, neuropathologists and ophthalmologists from different countries described AFI-like disorders in several families with clinical and pathological features clearly distinct from the classical Tay-Sachs disease (Batten, 1903; Spielmeyer, 1905; Vogt, 1905; Jansky, 1908; Bielschowsky, 1913). Despite the broad clinical variation, this spectrum of disorders was considered to belong to the same etiopathological process until the 1960s. Eventually, the storage material in Tay-Sachs disease was characterized as GM_2-ganglioside (Svennerholm, 1962), whereas accumulation of material resembling ceroid and lipofuscin was found in Batten-Spielmeyer-Vogt disease (Zeman and Donahue, 1963). Thereby it became evident that the amaurotic family idiocy disorders could be divided into two main categories, gangliosidoses and ceroid lipofuscinoses, and the term neuronal ceroid lipofuscinosis was introduced (Zeman and Dyken, 1969). More detailed descriptions of the history of the NCL disorders have been presented in several review articles (Haltia, 1975; Rapola, 1993; Goebel et al., 1999).

2.2 Classification and Nomenclature

The neuronal ceroid lipofuscinoses (NCLs) are progressive encephalopathies with autosomal recessive inheritance, characterized by neural and extraneural accumulation of ceroid and lipofuscin-like storage cytosomes (Santavuori, 1988). Collectively, they are the most common cause of childood neurodegeneration, with an average incidence of approximately 7-8 per 100,000 live births (Rider and Rider, 1988). The main clinical features are progressive mental retardation, ataxia, myoclonus, visual failure, and epileptic seizures (Santavuori et al., 2000). Based on the age at onset, the NCL disorders were originally divided into four main types: infantile, late infantile, juvenile and adult (Table 1). Since then, three variant forms of the late infantile NCL have been described. Atypical forms of juvenile and adult NCL have also been found, and Northern epilepsy has been shown to belong to the NCL group.

In addition to the clinical observations, the diagnosis has traditionally been based on electron microscopy (EM) analyses of biopsy specimens and blood samples (Goebel et al., 1999; Goebel, 2000; Tyynelä and Suopanki, 2000). The ultrastructural patterns of the lysosomal storage materials found in NCL are granular osmiophilic deposits, curvilinear profiles, fingerprint bodies, and rectilinear profiles (Table 1).

Table 1. Classification of the NCL diseases based on the clinical course and the histopathological findings. GROD, granular osmiophilic deposits; CL, curvilinear profiles; FP, fingerprint bodies; RL, rectilinear profiles; EPMR, epilepsy with mental retardation. *Turkish variant late infantile NCL was originally given the locus name *CLN7*, but in some families mutations have recently been detected in the *CLN8* gene. (See chapter 2.3 for references.)

Clinical onset	Disease	Ultrastructure	Gene
8-18 months	Infantile NCL	GROD	*CLN1*
2-4 years	Classic late infantile NCL	CL	*CLN2*
2-7 years	Finnish variant late infantile NCL	FP+CL+RL	*CLN5*
2-8 years	Czech variant late infantile NCL	FP+CL+RL	*CLN6*
1-6 years	Turkish variant late infantile NCL	FP+CL+RL	*CLN7/8**
4-8 years	Juvenile NCL	FP+CL+RL	*CLN3*
6-10 years	Juvenile NCL with GROD	GROD	*CLN1*
5-10 years	Northern epilepsy / EPMR	CL+GROD	*CLN8*
11-50 years	Adult NCL	mixed	*CLN4?*

Table 2. Classification of NCL diseases based on the defective genes and proteins. (See chapter 2.3 for references.)

Defective gene	Defective protein	Clinical phenotype
CLN1	**PPT1** (lysosomal enzyme)	- infantile - later onset up to adulthood
CLN2	**TPP1** (lysosomal enzyme)	- late infantile - later onset up to juvenile
CLN3	**CLN3** (transmembrane protein)	- juvenile
CLN5	**CLN5** (lysosomal protein)	- late infantile (Finnish variant)
CLN6	**CLN6** (transmembrane protein)	- late infantile (Czech variant)
CLN8	**CLN8** (transmembrane protein)	- Northern epilepsy / EPMR - late infantile (Turkish variant)

In the last decade major advances have occurred in the understanding of the molecular pathogenesis behind the NCL disorders (Gardiner, 2000). Mutations in six different genes have been detected and the corresponding defective proteins have been identified and analyzed. This has made possible a clear and exact classification of the NCL disorders, based on the defective genes and proteins (Table 2). In this thesis I will use this classification and consequently refer to the diseases according to their defective

genes (CLN1, CLN2, etc.). In future, when the pathological mechanisms are characterized in detail, it might be more suitable to name the diseases according to the functional defect of the proteins (PPT1 deficiency, TPP1 deficiency, etc.). However, as different mutations in the same gene can cause a broad spectrum of phenotypes, the clinical course cannot be predicted from the defective gene alone. Subgroups of each NCL disease can be defined according to the phenotypes that the different mutations cause.

The medical literature has a strong history of using eponyms as names of diseases. The term Batten disease is still often used, sometimes for CLN3, and sometimes for the whole group of NCL diseases. The eponyms previously used for the different forms of NCLs are mentioned in chapter 2.3.

2.3 Defective Genes and Proteins

CLN1 / PPT1

Infantile neuronal ceroid lipofuscinosis (CLN1; INCL; Santavuori-Haltia disease; MIM 256730) is caused by mutations in the *CLN1* gene encoding a lysosomal enzyme, palmitoyl protein thioesterase 1 (PPT1) (Vesa et al., 1995; Hellsten et al., 1996; Verkruyse and Hofmann, 1996). The gene was identified by positional cloning strategies after linkage to chromosome 1p32 had been demonstrated (Järvelä, 1991). PPT1 removes palmitate groups from proteins *in vitro* (Camp and Hofmann, 1993; Camp et al., 1994). In COS-1 cells PPT1 is targeted to the lysosomes through the mannose-6-phosphate pathway. In neurons, however, PPT1 is located in synaptosomes, synaptic vesicles and axons (Heinonen et al., 2000a; Lehtovirta et al., 2001; Suopanki et al., 2002; Ahtiainen et al., 2003). The PPT1 protein is highly conserved in mammals, and the mouse Ppt1 protein shows processing and intracellular localization similar to its human counterpart (Salonen et al., 1998). A subtype of CLN1 with juvenile onset, but with the typical granular osmiophilic deposits (GROD), is caused by "milder" missense mutations on the *CLN1* gene (Mitchison et al., 1998). The proteins of the storage material in CLN1 consist mainly of sphingolipid activator proteins (SAPs), detected only in small amounts in other NCLs (Tyynelä et al., 1993). A second enzyme, PPT2, shares 18% amino acid identity with PPT1. Like PPT1, PPT2 is a lysosomal thioesterase that hydrolyzes the long-chain fatty acyl coenzyme A, but with a substrate specificity that is distinct from that of PPT1 (Soyombo and Hofmann, 1997). The disruption of either *Ppt1* or *Ppt2* causes neuronal ceroid lipofuscinosis in knockout mice (Gupta et al., 2001). In Finland about 160 patients have been diagnosed, compared to about 200 elsewhere (Norio, 2003c). 98% of the Finnish disease chromosomes carry the same missense mutation, although 41 different mutations in the CLN1 gene has been detected so far.

CLN2 / TPP1

Classical late infantile neuronal ceroid lipofuscinosis (CLN2; LINCL; Jansky-Bielschowsky disease; MIM 204500) is caused by deficiency of a lysosomal serine protease, tripeptidyl peptidase I (TPP1) (Sleat et al., 1997; Liu et al., 1998; Sleat et al., 1999). The defective protein was identified using a mannose 6-phosphate modification of newly synthesized lysosomal enzymes as an affinity marker (Sleat et al., 1997). When comparing labeled protein samples, a single protein absent in LINCL patients was detected. TPP1 cleaves tripeptides from the N-terminus of small peptides undergoing lysosomal degradation (Vines and Warburton, 1999). The TPP1 proenzyme is converted into the mature form of ~48 kDa in the lysosomes (Golabek et al., 2002). The pH optimum of the active enzyme is 3.5 (Sohar et al., 1999). The mouse Cln2 protein consists of 562 amino acids with 88% identity to its human ortholog (Katz et al., 1999a). The nature of the physiological substrates and the range and specificity of the enzyme remain unclear.

CLN3

The locus for Juvenile neuronal ceroid lipofuscinosis (CLN3; JNCL; Batten disease; Spielmeyer-Vogt-Sjögren disease; MIM 204200) was initially assigned to chromosome 16 by linkage analyses (Eiberg et al., 1989; Gardiner et al., 1990). The *CLN3* gene was identified in 1995 and encodes a transmembrane protein, which may have a role in the regulation of the vacuolar pH (International Batten Disease Consortium, 1995; Pearce et al., 1999). In addition to lysosomes (Järvelä et al., 1998; Golabek et al., 1999), several other intracellular localizations have been suggested (Katz et al., 1997; Kremmidiotis et al., 1999; Margraf et al., 1999). In the neurons, the protein is transported along the neuronal extensions and targeted to neuronal synapses (Järvelä et al., 1999; Luiro et al., 2001). Three mouse models of CLN3 have been created, showing progressive neurodegeneration and premature death (Katz et al., 1999b; Mitchison et al., 1999; Cotman et al., 2002). Interestingly, an autoantibody to glutamic acid decarboxylase (GAD65) has been detected in the serum of *Cln3* knock-out mice, but not in normal mice (Chattopadhyay et al., 2002). The autoantibody to GAD65 was also present in the sera of all 20 CLN3 patients studied, in contrast to controls. Thus, an autoimmune response to GAD65 is proposed to contribute to a preferential loss of GABAergic neurons associated with juvenile neuronal ceroid lipofuscinosis. About 200 CLN3 patients have been recognized in Finland, but the incidence (1:20.000) is almost as high in several other populations (Norio, 2003c). In the CLN3 gene 31 disease causing mutations have been detected.

CLN4

Adult onset forms of neuronal ceroid lipofuscinosis (CLN4; ANCL; Kufs' disease, Parry disease; MIM 204300) have been described with both a recessive and a dominant pattern of inheritance. Although they represent a heterogeneous spectrum of clinical phenotypes and histopathological findings (Berkovic et al., 1988; Nijssen et al., 2003), the hypothetical genetic locus of the adult NCL has been named *CLN4*. Insufficient family material and lack of distinct diagnostic criteria has made a genome-wide linkage analysis approach difficult. So far, no locus or gene associated with adult NCL has been reported. The varying phenotypes might suggest that the adult-onset NCL cases are caused by "milder" mutations in the known CLN genes, causing later onset and protracted phenotypes, as has already been reported for CLN1 and CLN2. Consequently, a specific *CLN4* locus may not be found.

CLN5

Finnish variant late infantile neuronal ceroid lipofuscinosis (CLN5, $vLINCL_{Fin}$, MIM 256731) is described in detail in chapter 3 of the review of the literature as well as in the result and discussion parts of this thesis.

CLN6

Variant late infantile neuronal ceroid lipofuscinosis (CLN6; vLINCL; MIM 606725) is caused by mutations in a predicted transmembrane protein (Gao et al., 2002; Wheeler et al., 2002). Preliminary data suggests localization to the endoplasmic reticulum (Mole et al., 2003). The defective *CLN6* gene was identified by positional cloning strategies, after mapping to chromosome 15q21-23 (Sharp et al., 1997; Sharp et al., 1999; Sharp et al., 2001). A total of 16 mutations have been reported and minor differences in the pattern of disease symptom evolution were identified (Sharp et al., 2003). In the naturally occurring *nclf* mouse, with an NCL-like phenotype, a 1 bp insertion was detected in the orthologous *Cln6* gene (Gao et al., 2002; Wheeler et al., 2002).

CLN7

Turkish variant late infantile neuronal ceroid lipofuscinosis (CLN7, Turkish variant LINCL) was originally given the locus name *CLN7*. In some of the Turkish families, mutations have recently been detected in the *CLN8* gene (Ranta et al., 2003). Whether the remaining families represent a novel vLINCL locus, *CLN7*, or are caused by mutations in *CLN2*, *CLN5* or *CLN6* remains to be clarified.

CLN8

Mutations in the *CLN8* gene cause two different clinical phenotypes: Progressive epilepsy with mental retardation (EPMR; Northern epilepsy; MIM 600143) and Turkish variant late infantile neuronal ceroid lipofuscinosis (former CLN7). The *CLN8* gene was initially linked to chromosome 8 and subsequently identified by sequencing the transcripts on the restricted region with DNA of EPMR patients (Tahvanainen et al., 1994; Ranta et al., 1999). Recently, mutations in the *CLN8* gene have also been detected in Turkish families with late infantile NCL (Ranta et al., 2003). The naturally occurring motor neuron degeneration (*mnd*) mouse was observed to harbour mutations in the orthologous *Cln8* gene (Ranta et al., 1999). CLN8 is an endoplasmic reticulum (ER) resident multi-pass transmembrane protein that recycles between ER and the ER-golgi intermediate compartment (ERGIC) (Lonka et al., 2000). CLN8 has been suggested to belong to a novel eukaryotic family of lipid-sensing domains including two other ER-resident proteins: the human TRAM and the *Saccharomyces cerevisiae* Lag1 (Winter and Ponting, 2002).

2.4 Animal Models

The identification of the genes underlying the NCL disorders has facilitated the development of mouse models of the NCLs (Table 3). Following the cloning of the mouse orthologs, mouse models have been produced of CLN1, CLN2, CLN3 and CLN5. Also two naturally occurring mouse disorders with NCL-like symptoms and pathological findings, *nclf* and *mnd*, have been confirmed to be caused by mutations in *Cln* genes. In addition, three other knockout mice with NCL-like phenotypes have been characterized (*Ppt2, Clcn3* and *Ctsd*). However, no NCL patient has shown any association to the human orthologs of these three genes.

Ppt1 null mutant mice exhibit an NCL-like phenotype with widespread accumulation of autofluorescent lipopigments, a shortened lifespan, progressive motor abnormalities and spontaneous myoclonic seizures (Gupta et al., 2001). These mice also exhibit profound loss of subpopulations of inhibitory interneurons in the cortex and hippocampus, with the persisting neurons exhibiting a range of dendritic abnormalities and loss of synaptic spines (Cooper, 2003). Based on the findings in the knockout mice, Ppt1 appears to play a role in the maintenance of certain neuronal populations rather than during their development. Inactivation of *Ppt2*, another lysosomal thioesterase, produces a milder phenotype, suggesting that Ppt2 plays a role in the brain not carried out by Ppt1 (Gupta et al., 2001). A new mouse model for Ppt1 has been produced, using a knock-in strategy to disrupt exon 4, harbouring the major Finnish mutation (unpublished data). These mice showed progressive motor abnormalities and died prematurely at the ages of 6-9 months. The pathological

findings, includings GRODs and autofluorecent storage materials, are the same as in the CLN1 patients.

A mouse model for CLN2 has been produced, and preliminary observations suggest that these mice have reduced TPP1 activity in the brain in addition to progressive neurological symptoms (Sleat et al., 2003).

Table 3. Mouse models that are orthologous to human NCL types or show pathological similarities, such as lysosomal deposition of autofluorescent storage materials and neurodegeneration. The exons targeted for mutagenesis using knock-in strategies are marked by delta (Δ).

Human gene	Mouse model	Key references
CLN1	*Ppt1*	Gupta et al., 2001
	Ppt1 (Δex4)	Unpublished data
CLN2	*Cln2*	Sleat et al., 2003
CLN3	*Cln3*	Mitchison et al., 1999
	Cln3	Katz et al., 1999
	Cln3 (Δex7/8)	Cotman et al., 2002
CLN5	*Cln5*	Unpublished data
CLN6	*nclf*	Bronson et al., 1998
CLN8	*mnd*	Bronson et al., 1993
PPT2	*Ppt2*	Gupta et al., 2001
CLC-3	*Clcn3*	Yoshikawa et al., 2002
Cathepsin D	*Ctsd*	Koike et al., 2000

Two *Cln3* null mutant models have been produced by targeted disruption of the *Cln3* gene (Katz et al., 1999b; Mitchison et al., 1999). Intracellular accumulation of autofluorescent material containing subunit C of mitochondrial ATP synthase was observed, as in the CLN3 patients. Mutant animals also showed neuropathological abnormalities, with loss of certain cortical interneurons and hypertrophy of many interneuron populations in the hippocampus (Mitchison et al., 1999). Retinal manifestations included accumulation of autofluorescent material and moderate retinal degradation, but no significant abnormality was observed in the vision function (Seigel et al., 2002).

A third mouse model has been made, using a knock-in strategy to mimic the ~1 kb deletion in the *CLN3* gene present in the majority of JNCL patients (Cotman et al., 2002). These $Cln3^{\Delta ex7/8}$ mice develop earlier and more pronounced NCL-like symptoms than the previous null mutant models, and detectable pathological findings are evident even prenatally. As the harmful impact of the mutation on the CNS did not

correlate with the membrane deposition, the CLN3 protein activity might be essential for the survival of neurons.

The *Cln5* knockout mice developed normally and were fertile (unpublished data). However, at the age of 3-4 months they developed seizures and their vision started to decline. Similarly to the CLN5 patients, the TPP1 activity was significantly increased in the brain, kidney and liver of the *Cln5*$^{-/-}$ mice. At autopsy, the mice showed cerebellar atrophy and electron microscopy revealed an ultrastructure with fingerprint and curvilinear profiles, analogous to human CLN5 tissues.

Two naturally occurring murine forms of NCL, the *mnd* (motor neuron degeneration) and *nclf* (neuronal ceroid lipofuscinosis) mice, have been identified on the basis of the typical intraneuronal inclusions of autofluorescent material, immunoreactive with antibodies to subunit c of mitochondrial ATP synthase (Bronson et al., 1993; Bronson et al., 1998). These mice exhibit almost identical phenotypes, with progressive retinal atrophy and paralysis at around 9 months of age. Thus it was not surprising that the underlying mutations were found in two mouse genes, *Cln6* and *Cln8*, orthologous to human genes mutated in two different forms of late-infantile NCL (Gao et al., 2002; Wheeler et al., 2002; Ranta et al., 2003). In the *mnd* (*Cln8*) mouse, pronounced early accumulations of autofluorescent lipopigments has been detected in subpopulations of GABAergic neurons, including interneurons in the cortex and hippocampus (Cooper et al., 1999). Treatment with insulin-like growth factor-1 partially restored the interneuronal number and reduced hypertrophy in some subregions. The *mnd* mouse has been shown to manifest behavioral abnormalities, including increased activity, hightened aggression and poor memory, prior to the onset of gross motor symptoms (Bolivar et al., 2002).

Cathepsin D deficient mice have been shown to manifest seizures and become blind near the terminal stage (Saftig et al., 1995). The *Ctsd*$^{-/-}$ mice are born normally but die on postnatal day 26 ± 1 because of massive intestinal necrosis, thromboembolia, and lymphopenia (Koike et al., 2000). Subunit C of the mitochondrial ATP synthase was found to accumulate in the lysosomes of neurons, and the activity of TPP1 was significantly increased in the brain. These findings suggest that Cathepsin D deficient mice show a new form of NCL. Surprisingly, chloride channel-3 (*Clcn3*) null mutant mice also exhibit an aggressive NCL-like phenotype, which closely resembles that of *Ctsd*$^{-/-}$ (Yoshikawa et al., 2002). Furthermore, an elevation in the endosomal pH of the *Clcn3*$^{-/-}$ mice was detected. Among other roles, chloride channels act to regulate the ionic composition of intracellular compartments and thus their pH (Cooper, 2003). Elevated lysosomal pH has been detected in all NCL forms except CLN2 and CLN8 (Holopainen et al., 2001). Elevated pH most probably disturbs the catalytic activity of the lysosomes and is one important factor in explaining the accumulation of the ceroid and lipofuscin-like autofluorescent lipopigments characteristic of NCLs.

Cathepsin B null mice and Cathepsin L null mice show no obvious pathological phenotypes, but mice lacking both Cathepsin B and L manifest severe cerebral and cerebellar atrophy leading to death during the second to fourth week of life (Felbor et al., 2002). These mice only partly reproduce NCL-like features as the lysosomal accumulation material is distinct from the one found in NCL mice.

Experiments with larger animal models could be useful for assaying the biodistribution and efficacy of potential therapies, as these become available. Naturally occurring sheep strains that show recessive NCL-like disorders are the best characterized examples so far. NCL in New Hampshire sheep (OCL6) has been shown to be synthenic with the human CLN6 (Jolly and Walkley, 1999). This ovine model has a characteristic loss of GABAergic interneurons. A mutation in the ovine cathepsin D gene causes severe neurodegeneration and lysosomal storage of granular osmiophilic deposits (GROD) containing SAPs A and D, as in CLN1 patients with PPT1 deficiency (Tyynelä et al., 2000).

2.5 New Approaches to Characterize Molecular Pathogenesis

To clarify the potential molecular interactions and metabolic pathways of the CLN proteins, new genome-wide approaches can be applied to analyse the available animal models for NCLs (Brooks et al., 2003). Using oligonucleotide arrays, changes in gene expression in the brains of 10-week-old *Cln1*- and *Cln3*-knockout mice were detected compared to wild-type controls, and partly confirmed by immunoblotting (Elshatory et al., 2003). Despite the similarities in pathology, the two mutations affect the expression of different, non-overlapping sets of genes. Interestingly, the gene showing the highest degree of downregulation in the *Cln1*-knockout mouse (mouse EST, GenBank acc. no. AI505387) is the most highly up-regulated transcript in the *Cln3*-knockout. The significance of this observation awaits characterization of the function of this novel gene. In the *Cln3*-knockout mice, two genes associated with glutamate utilization were downregulated, namely glutamine synthetase and the glutamate receptor mGluR3. This might be caused by the presence of the autoantibody to GAD65, which inhibits conversion of glutamate to GABA and therefore leads to an accumulation of glutamate (Chattopadhyay et al., 2002).

The metabolic profile of cerebral cortex extracts of the *mnd* (Cln8) mouse has been analyzed by high-resolution NMR spectroscopy and principal component analysis (Griffin et al., 2002). A profound vitamin E deficiency was detected in both serum and brain. Dietary vitamin E supplementation significantly reversed some of the metabolic abnormalities, but had no effect upon brain pathology. An interesting finding was the increased level of glutamate in combination with decreased GABA in the cerebral

cortex tissue. This links the CLN8 protein deficiency with the abnormalities in the GABAergic neurons observed in CLN1, CLN3 and ovine CLN6.

In a screen for genes that control synapse development, *spinster* (*spin*), which encodes a multipass transmembrane protein, has been identified (Sweeney and Davis, 2002). The Spin protein has been localized to a late endosomal compartment and *Drosophila spin* mutants showed altered endosomal/lysosomal function and accumulation of ceroid lipid pigment and GM_2-ganglioside-like substance in the neurons. Synaptic overgrowth was detected in the *spin* mutants, and was shown to be caused by enhanced or misregulated TGF-β signaling. The parallel findings in *spin* mutants and NCLs is striking and these findings might provide important information about the connection between membrane traffic, lipid organization, synaptic development, growth factor signaling, and their defects resulting in neurodegenerative lysosomal storage diseases (Sanyal and Ramaswami, 2002).

2.6 Diagnostics

The identification of most of the mutations behind the NCLs has facilitated precise molecular diagnosis (Goebel et al., 1999; Mole et al., 2001b). However, more traditional methods and skillful clinicians are still needed to identify the cases with potential mutations in the *CLN* genes. When a neurodegenerative disease is suspected, the initial tests are MRI and electron microscopy of skin or of a rectal biopsy. If typical morphological ultrastructures for NCLs are detected, the next step is enzyme activity analysis (PPT and TTP) and DNA analysis. Analysis of the mutation has to be designed based on the clinical phenotype in combination with the local frequencies of the different NCL types and mutations.

Currently, electron microscopy is still the "gold standard" for determination of whether a patient with neurological abnormalities has an NCL disease or not. Based on the ultrastructure and clinical picture, the subtype is then predicted and mutation screening can start. Identification of the causative mutation is important for determination of the exact molecular diagnosis. In addition it allows carrier testing of relatives, and gives reliable information about the prognosis. Only in families with previously detected NCL mutations, is DNA analysis the primary diagnostic method.

Prenatal diagnosis is possible for families with previous NCL children (Rapola et al., 1999). Mutation analysis can be performed with DNA extracted from chorionic villus samples (CVS) or from cells cultured from amniocentecis samples. Prenatal diagnosis gives the parents the option to abort an affected foetus. Another possibility is preimplantation diagnosis which, however, is not in routine use for NCL families in Finland. Since prenatal diagnosis and mutation screening of the relatives of NCL

patients are associated with difficult ethical considerations, appropriate genetic counselling is crucial.

2.7 Therapies

At present, no curative treatment for NCLs is available and the therapeutic outlook remains bleak. Many ethical and technical challenges also remain to be solved before treatment can be offered to patients. With several animal models available, the systematic testing of therapeutic interventions is now possible. Replacement of the missing enzyme may be an option for CLN1 and CLN2, but the problems of enzyme production and delivery need to be overcome. Gene therapy could be an appropriate approach, and recent findings suggest that viral-mediated delivery may be feasible for CLN2 (Haskell et al., 2003).

A number of neurodegenerative diseases are characterized by oxidative damage. Thereby it has been suggested that treatment with antioxidants could delay or halt the progression of the NCLs. Long-term antioxidant supplementation with vitamin E and sodium selenite has been given to CLN3 patients, but with no significant response (Santavuori et al., 1985; Santavuori et al., 1988). Recently, metabolic profile analyses have been used to monitor the effect of dietary supplementation with vitamin E in the *Mnd* mouse model (*Cln8*) (Griffin et al., 2002). Although the vitamin E deficiency was reversed in the mutant mice, it had no effect upon brain pathology or the metabolic abnormalities caused by the disease.

Hematopoietic stem cell transplantation has been evaluated in three patients with CLN1 (Lönnqvist et al., 2001). The effect on the clinical progression of the disease was marginal and this treatment is currently not recommended. However, there is hope that transplanted neural stem cells may eventually be capable of replacing damaged neurons.

Accelerated apoptosis has been suggested to play a role in the neurodegeneration in CLN2 and CLN3 (Dhar et al., 2002). The antiapoptotic properties of the CLN3 protein were demonstrated on the basis of the growth rate of lymphoblasts homozygous for the common *CLN3* mutation (Persaud-Sawin et al., 2002). The drug flupirtine blocks apoptosis in lymphoblasts of CLN1, CLN2, CLN3, and CLN6 patients, as well as in CLN2 and CLN3 deficient neurons (Persaud-Sawin et al., 2002). Therefore flupirtine was suggested as a useful drug capable of halting the progression of neurodegenerative diseases caused by dysregulated apoptosis. As defects in regulation of apoptosis are involved in the development of cancer and as *CLN3* is overexpressed in several cancer cell lines, CLN3 has been proposed as a molecular target for cancer drug discovery (Rylova et al., 2002).

As the focus eventually shifts towards testing potential therapies on patients, it will be essential that rigorously defined landmarks exists for quantitative measurement of the disease progression and evaluation of therapeutic efficacy. Potential indicators for the clinical condition include motor, visual and verbal functions, somatosensory evoked responses and brain imaging. Although some protocols for follow up have been presented for specific forms of NCL, a standardized user friendly scale to score the clinical course of NCLs will prove particularly valuable in defining specific endpoints in the design of clinical trials (Steinfeld et al., 2002).

3. CLN5 (FINNISH VARIANT LATE INFANTILE NCL)

Neuronal ceroid lipofuscinoses (NCL) are inherited lysosomal storage disorders, composing the most common type of progressive encephalopathies in children. One of the six subtypes of NCL, the Finnish variant form of late infantile neuronal ceroid lipofuscinosis (CLN5), was originally described in 1982 (Santavuori et al., 1982). In that report, Santavuori presented 16 patients with similar histopathological findings but with clinical symptoms slightly different from those seen in the classical form of late infantile neuronal ceroid lipofuscinosis (LINCL; CLN2), previously also called Jansky-Bielschowsky disease. CLN5 later turned out to be a typical exampel of a disease belonging to the Finnish Disease Heritage.

The nomenclature and classification of the NCL diseases is complex and has been shifting over the years. Different names and abbreviations used for CLN5 are among others: vLINCL, fvLINCL, $vLINCL_{Finnish}$, $vLINCL_{Fin}$, variant CLN2, NCL5, variant Jansky-Bielschowsky disease and JBVD. In this thesis, I will use CLN5 as the name for the disease, *CLN5* for the defective gene, CLN5 for the defective protein, *Cln5* for the corresponding mouse gene and Cln5 for the mouse protein.

3.1 Clinical Features

In the original series of 16 CLN5 patients, the onset of the disease occurred between 5 and 7 years of age (Santavuori et al., 1982). The first symptoms were mental retardation and visual failure, followed by ataxia, myoclonia and epilepsy. In 1991 a series of 5 CLN5 patients was presented, showing that the spectrum of the disease was wider than had previously been assumed (Santavuori et al., 1991). The leading symptoms in these patients were mental and motor disturbances appearing between 4.5 and 7 years of age. Visual failure, ataxia, myoclonia and epilepsy developed in all the children before the age of 10 years (Figure 3). Independent walking ability was lost between 8-11 years. All the patients became blind by ten years of age and gradually lost their ability to speak and move, and became spastic and deeply retarded. These 21 patients have later been shown to be homozygous for the Finnish major mutation in the CLN5 gene.The life expectancy for CLN5 patients is about 20 years, with variation observed from 14 to 39 years (Norio, 2003c). Sleep disturbances are common in CLN5 patients, including an excess of nocturnal sleep and frequent daytime naps (Kirveskari et al., 2001).

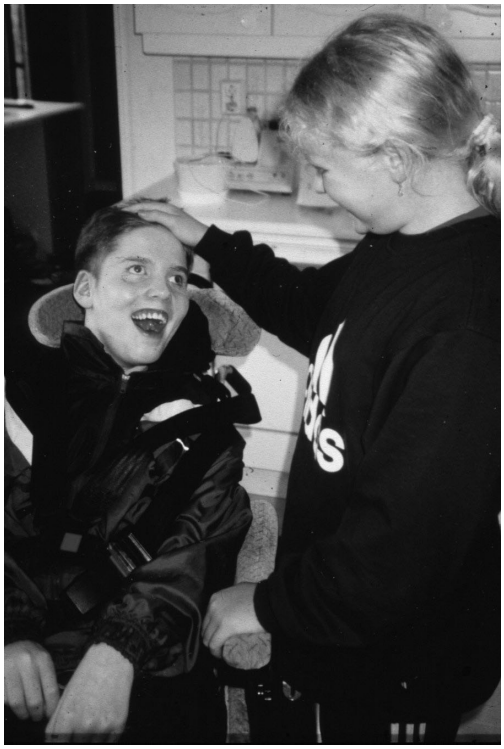

Figure 3. A 15-year-old Finnish CLN5 patient (left) together with his sister.

3.2 Genetic Background

The effort to clarify the genetic background of the CLN5 disease was initiated at the beginning of the 1990s. Carefully performed clinical diagnoses, high-quality population records kept by the Finnish Church since the 17th century, and active cooperation of the CLN5 families made this step possible. Initially the CLN1 and CLN3 loci were excluded for CLN5 by linkage analyses (Järvelä, 1991; Williams et al., 1994). Fourteen Finnish CLN5 families, with a total of 67 individuals, including 17 patients, were used in the linkage studies that resulted in mapping of the *CLN5* locus to a well-defined region on 13q21.1-q32 (Savukoski et al., 1994). Since one of the disease haplotypes ("4.5") generated with markers D13S162 and D13S160, located 3 cM apart, was present in 78% of the CLN5 chromosomes, it is likely that one ancient founder mutation was introduced into this haplotype and spread in the internal isolate of Finland.

To further investigate the origin and age of the founder mutation, genealogical and linkage disequilibrium analyses of 18 CLN5 families were performed (Varilo et al., 1996). Among the 36 parents of the CLN5 patients, 24 (67%) were found to be consanguineous with 1-7 other parents. The oldest traced common ancestor originates from the 17th century, 13 generations ago. All common ancestors in the large CLN5 pedigrees lived in five nearby parishes, Kauhajoki, Kauhava, Kuortane, Kurikka, and Lapua, which are situated in Southern Ostrobothnia in Western Finland (Figure 4). In

this region 12 affected children had been diagnosed during the last 20 years, revealing a local incidence of 1/1500. The major CLN5 mutation, homozygous in most Finnish patients, obviously occurred some 20-30 generations and 500-750 years ago. The combination of the results of the genealogical and linkage disequilibrium studies restricted the chromosomal region containing the *CLN5* gene from 4 cM (~ 4 Mb) to less than 400 kb.

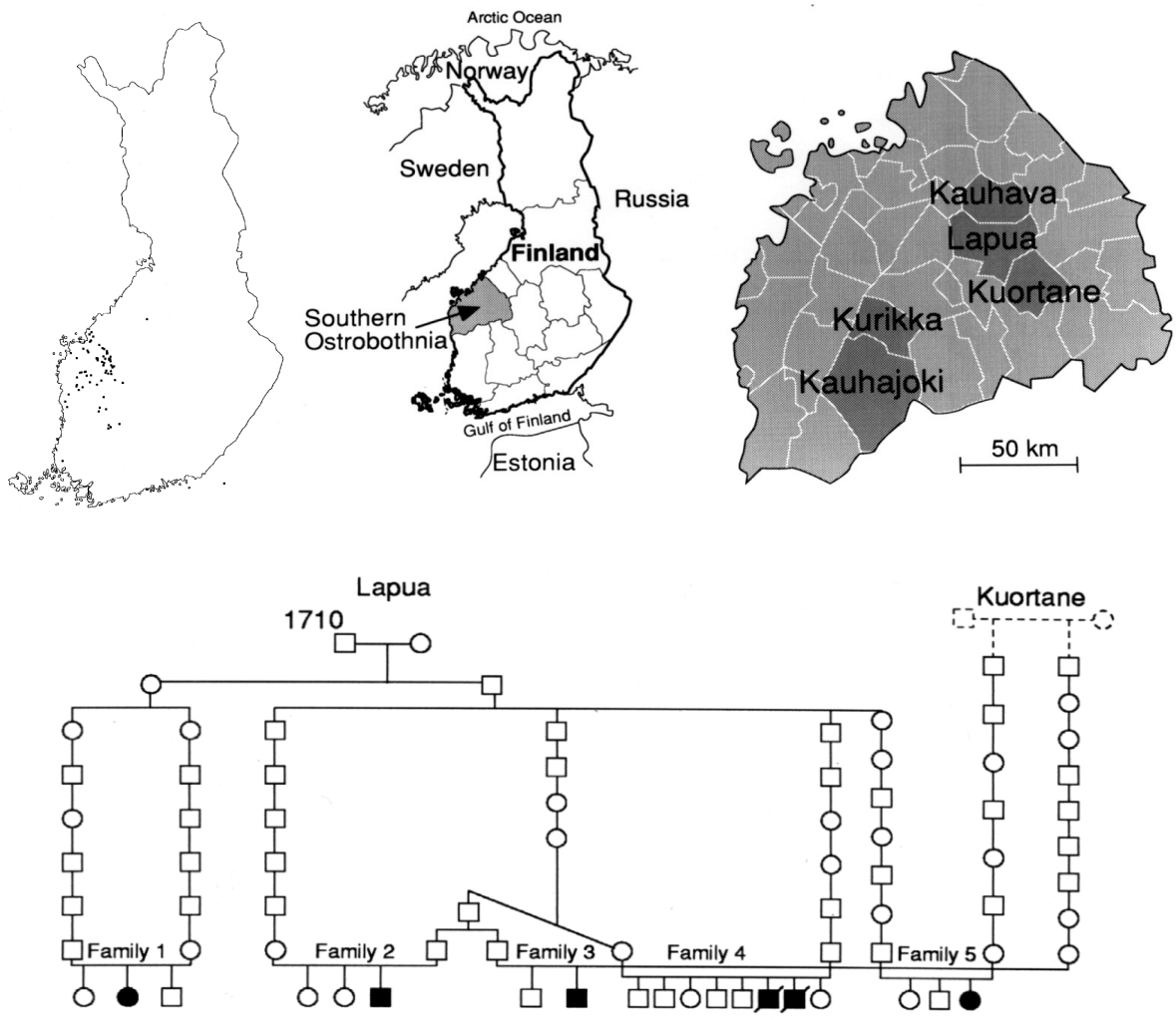

Figure 4. The map on the left shows the birthplaces (dots) of the grandparents of the Finnish CLN5 patients. All common ancestors of the CLN5 patients lived in five neighbouring parishes in Southern Ostrobothnia: Kauhava, Lapua, Kuortane, Kurikka and Kauhajoki. The pedigree shows the relationships between the five families with CLN5. Modified from Varilo et al., 1996 and Peltonen et al., 1999.

Ancient recombination events detected by isolation of two novel polymorphic markers restricted the critical region of the *CLN5* locus to 200 kb (Klockars et al., 1996). A physical map based on a contig of overlapping PAC clones over the CLN5 region was

constructed using the Fiber-FISH technique. This made it possible to initiate the last phase of the positional cloning of the *CLN5* gene, namely the mutation detection of the transcripts on the region. Three PAC clones covering the critical 200 kb containing the *CLN5* locus were sequenced as part of the Human Genome Project at the Whitehead Institute for Biomedical Research, USA. Three genes and two pseudogenes were identified by cDNA library screenings, EST database searches and using gene prediction programs (Klockars et al., 1999). Sequencing of these genes with DNA from patients and controls resulted in the detection of three mutations in one of the novel genes, which was therefore named *CLN5* (Savukoski et al., 1998). Identification of the disease causing mutations facilitates reliable prenatal diagnosis, which is now available for families with the major Finnish mutation (Rapola et al., 1999).

CLN5 is a glycoprotein predominantly targeted to lysosomes in transiently transfected BHK and COS cells. The pathogenesis has been suggested to be associated with defective lysosomal trafficking, which prevents the normal biological function of the CLN5 protein. (Isosomppi et al., 2002) The CLN5 protein has been reported to interact with CLN2 and CLN3, based on results from co-immunoprecipitation and *in vitro* binding assays (Vesa et al., 2002). mRNA and protein expression of CLN5 in the embryonic human brain has been characterized by *in situ* hybridization and immunohistochemical staining. Expression was evident in the developing cortex, hypothalamus and cerebellum. The most intense signal was detected in the precursors of the Purkinje cells of the cerebellum, correlating well with the severe cerebellar atrophy seen in the CLN5 patients. (Heinonen et al., 2000b)

3.3 Pathology

Macroscopically severe atrophy of the cerebellum combined with moderate cerebral atrophy has been observed in autopsies of CLN5 patients (Figure 5) (Tyynelä et al., 1997). The patients displayed a uniform and extreme cerebellar atrophy, with almost complete destruction of the granule cells and Purkinje cells. Advanced loss of cortical neurons was observed, the remaining cells displaying pronounced storage enlargements. The hippocampus was relatively spared, and there was extensive cortical astrocytosis. Electron microscopy showed that the predominant ultrastructure of the intraneuronal storage material was characterized by rectilinear (RL) or atypical curvilinear (CL) profiles, with occasional fingerprint bodies (FP). Previous analyses of biopsy samples had shown a combination of curvilinear (CL) and fingerprint (FP) structure (Santavuori et al., 1982).

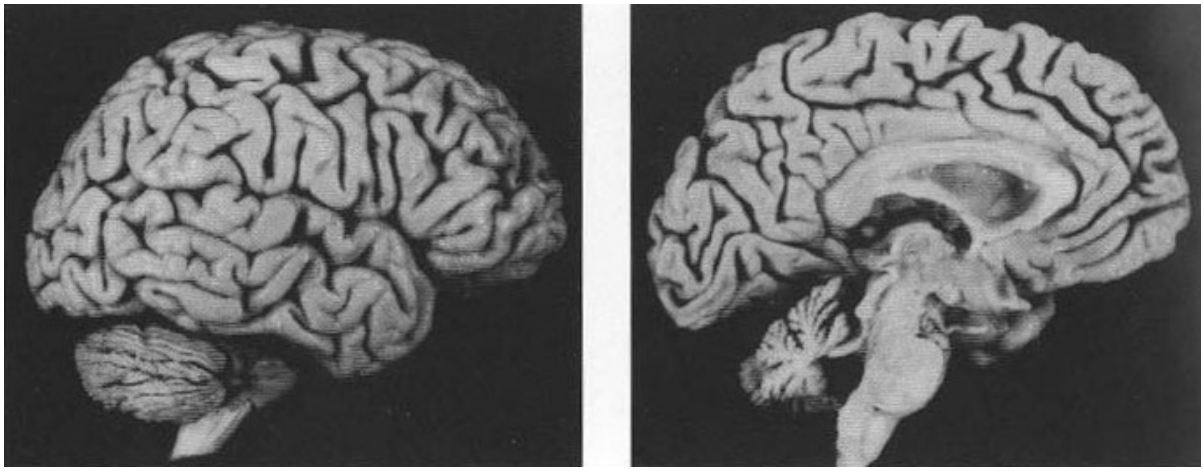

Figure 5. Whole brain (left) and sagittal section (right) of the brain of a 14-year-old female CLN5 patient. The severe atrophy of the cerebellum is the most striking macroscopic abnormality. Modfied from Tyynelä et al., 1997.

Pathomorphological and biochemical analyses showed that the main protein in the storage complexes in the lysosomes of CLN5 brain samples is the subunit c of the mitochondrial ATP synthase (Tyynelä et al., 1997). The storage complexes also contain minor amounts of sphingolipid activator proteins (SAPs). The immunohistological distribution of subunit c and SAPs in the central nervous system closely resembles those of classical LINCL.

3.4 Brain Imaging

The characteristic electrophysiological findings in CLN5 are posterior spikes in the EEG during a low-rate photic stimulation, and giant visual and somatosensory evoked potentials (Santavuori et al., 1982). Already at the time of diagnosis, magnetic resonance imaging (MRI) and computed tomography (CT) showed generalized brain atrophy, which was especially pronounced in the cerebellum (Autti et al., 1992; Autti et al., 1994; Autti et al., 1996; Autti et al., 1997). Magnetoencephaography (MEG) of CLN5 patients has shown a selective enhancement of the early cortical response to median nerve stimulation (Lauronen et al., 2002). This may reflect defective interneuronal inhibition in the cortex.

AIMS OF THE STUDY

When this study was initiated the clinical features of the CLN5 patients had been described and the autosomal recessive mode of inheritance had been established. The locus for *CLN5* had been mapped to chromosome 13q22 and the critical region had been restricted to less than 200 kilobases.

The initial aim of this study was to identify the *CLN5* gene mutated in this disease. During the project, the following additional goals were formulated:

- To characterize the novel *CLN5* gene and its disease-causing mutations.

- To find out whether there is a correlation between the clinical phenotypes and the genotypes of CLN5, caused by different mutations.

- To characterize the orthologous mouse *Cln5* gene, and to analyze the expression, processing and localization of the Cln5 protein it encodes.

MATERIALS AND METHODS

Table 4. The materials and methods used in this study are described in detail in the original publications (I-III) according to this table.

Materials and methods	Original publication
Patients	I, II
cDNA cloning	I, III
mRNA analysis by Northern blotting	I, III
PCR and sequencing	I, II
Bioinformatics	I, III
Brain imaging (CT and MRI)	II
mRNA analysis by *in situ* hybridization	III
Construction of expression plasmids	III
In vitro mutagenesis	III
Cell culture and transfection	III
Protein production by *in vitro* translation	III
Triton X-114 solubility assay	III
Antibodies	III
Western blotting	III
Metabolic labeling and immunoprecipitation	III
Immunofluorescence	III
Immunohistochemistry	III

Patients

The 18 CLN5 families, including 21 patients, we used for mutation screening (I) were the same as previously used for the linkage analyses (Savukoski et al., 1994; Klockars et al., 1996; Varilo et al., 1996). All patients were examined by the same physician and the diagnosis was based on the clinical course of the disease, neurophysiological and ophtalmological findings, and electron microscopy (EM) studies of biopsies from the rectum wall (Santavuori et al., 1982; Santavuori et al., 1991). DNA was extracted from frozen periferal blood samples according to standard protocols (Vandenplas et al., 1984).

The patient group used for genotype-phenotype correlations (II) consisted of six Finnish patients, one Dutch patient, and one Swedish patient with CLN5. Mutations in the CLN5 gene were found in both alleles of all eight patients. Three were homozygous for the Fin major mutation, two were homozygous for the Fin minor mutation, one was homozygous for the Dutch mutation and two were compound

heterozygotes. Information about the clinical courses of the patients were obtained from medical records and from discussions with the clinicians and parents of the patients. All patients received drug therapy for epilepsy.

Computed tomography (CT) and Magnet Resonance Imaging (MRI) of the patients were performed by experinced radiologists as described previously (Autti et al., 1992; Autti et al., 1994; Autti et al., 1996; Autti et al., 1997). Neurophysiological tests were done according to normal clinical practice (Santavuori et al., 1982; Santavuori et al., 1991).

Mutation detection and sequencing

The screening of mutations of positional candidate genes was performed by comparing PCR (Saiki et al., 1985; Saiki et al., 1988) amlified genomic DNA from CLN5 patients and controls using the manual solid-phase sequencing method (Syvänen et al., 1989). The mouse *Cln5* gene was characterized by cycle sequencing in accordance with the manufacturer's instructions using ABI PRISM BigDye Terminator sequencing kits (Perkin Elmer Corporation). The fragments were separated on an Applied Biosystems 377XL sequencer (Perkin Elmer) and the sequences were interpreted using the Sequencing Analysis 3.3 program (Perkin Elmer) and Sequencher 3.0 program (Gene Codes Corporation).

ETHICAL CONSIDERATIONS

Prenatal Diagnosis and Carrier Screening

All the blood and tissue samples used in this study have been collected with informed consent according to the World Medical Association's Declaration of Helsinki (http://www.wma.net). The declaration states that it is the duty of the physician in medical research to protect the life, health, privacy, and dignity of the human subjects.

As a result of this study, prenatal diagnosis is now available for parents with previous CLN5 children. Regarding the severity of this disease, abortion is an understandable choice if the foetus is tests positive - at least in Northern European societies. It is important to remember that the acceptance of prenatal diagnosis and abortion is greatly dependent on the local culture and the ideology and religion of the subjects. A difficult question is whether family members and relatives of CLN5 patients should be tested to find carriers. The situation is particularly problematic if the recently diagnosed patient has got younger, healthy siblings that could be diagnosed presymptomatically. Currently DNA tests are available only for the Finnish major CLN5 mutation at the laboratory of the Helsinki University Central Hospital.

As the carrier frequency of the Finnish major mutation locally in some communes in Southern Ostrobothnia is as high as 1:24-1:44 (Savukoski et al., 1998; Pastinen et al., 2001), screening of pregnant women for this mutation could be considered. However, carrier screening for rare recessive diseases has to be based on voluntariness and autonomy, and requires appropriate genetic counselling locally available, which has shown to be difficult and expensive to organize (Jallinoja, 2002). So far, no systematic carrier screening of any NCL mutation is offered.

Animal Experiments

The use of laboratory animals is a matter of great public concern. In the European Union there is a general need and commitment to reduce the number of animals used in laboratories for experimental and other scientific purposes to the minimum level that is scientifically necessary (EU Commission Communication to the Council and the Parliament COM (99)191; http://europa.eu.int/comm/environment/docum/99191_en.pdf).

In this study, rabbits and guinea pigs have been used for raising antibodies, and mouse tissues for *in situ* hybridization and immunohistochemical staining. Animal care and handling were consistent with the Guide for the Care and Use of Laboratory Animals (National Research Council, 1996)(http://bob.nap.edu/html/labrats/).

Criticism of the use of animals in medical research is commonly based on the rights of animals: (1) to live without pain, (2) to live in species-specific conditions, and (3) to live without having life involuntarily terminated. The first condition is widely accepted and efforts are made to minimize the suffering of laboratory animals. The right to live in species-specific conditions is not possible to fulfil in laboratories, but much can be done to enhance the living conditions of animals. The animals' right to life is more controversial. Peter Singers utilitarian theory is based on equal respect for the interests of each individual, but according to him the right to life is limited to self-conscious animals (Singer, 1975). Tom Regan, on the other hand defends deontological ethics and claims that even painless killing of an animal causes massive harm to the animal by depriving it of its chance to live (Regan, 1983). This discussion will continue and hopefully the principle of the three Rs - reduction, replacement and refinement - will prove efficient in reducing the suffering and killing of animals in biomedical research.

Social Impacts of Genetics

Since the beginning of genetic research, its social consequences have been widely discussed. This is not surprising, since today's genetics brings detailed information about the origin and nature of the human being and about the inherited variations between individuals. The responsible use of this new technology is a tremendous challenge for scientists and societies.

Can confidentiality always be guaranteed for genetic information about individuals and families? May genetic information be used for discrimination in health insurance and employment? Should products of genetic information be patented, and by whom - individual researchers, universities, funding organizations, private corporations, or the indivuals from whom the genetic information was obtained? What are the rights of indigenous peoples to their traditional knowledge? Is there a risk for stigmatization of patients with genetic diseases? What social consequences will the introduction of GMO crops have for the farmers in the South? Most of the ethical questions raised by genetics remain unsolved and continuous public discussion on these issues should be encouraged.

To meet the public concerns about genetics and medical research, new legislations have been introduced, ethical committees have been formed in many hospitals and research institutes, and ethical guidelines and recommendations have been developed. The National Human Genome Research Institute created the Ethical, Legal, and Social Implications (ELSI) Research Program in 1990 as an integral part of the Human Genome Project (Collins et al., 2003) (http://www.genome.gov/PolicyEthics/). In 1997, the General Conference of UNESCO adopted the Universal Declaration on the Human Genome and Human Rights, which states that the human genome is a heritage

of humanity and that "benefits from advances in biology, genetics and medicine, concerning the human genome, shall be made available to all, with due regard for the dignity and human rights of each individual". UNESCO's International Declaration on Human Genetic Data specifies in more detail the conditions that have to be considered in genetic research and genetic testing (http://www.unesco.org/). The European Society of Human Genetics has created professional guidelines concerning population screening programs, genetic testing for employment and insurances, genetic sevices and storage of DNA information (http://www.eshg.org).

The history of eugenics shows how dangerous the misuse of genetic knowledge can be. It was hoped that eugenics would "improve" whole populations by preventing certain individuals (i.e. disabled, homosexuals, psychiatric patients) from having children, while encouraging others. Even in recent years, genetics has been misused as Richard Herrnstein and Charles Murray claimed that IQ is largely genetically determined and explains social problems such as crime, welfare dependence and single parenting and thus programs to eliminate inequalities are doomed to fail (Herrnstein and Murray, 1994). Their controversial book, the Bell Curve, met an aggressive response i.e. by the ELSI working group of the Human Genome Project, arguing that "the lessons of genetics are not deterministic and genetic arguments cannot and should not be used to determine social policy" (Andrews and Nelkin, 1996).

RESULTS AND DISCUSSION

1. THE HUMAN *CLN5* GENE

1.1 Positional Cloning of the *CLN5* Gene (I)

When this study was started, the *CLN5* locus had recently been mapped to chromosome 13 by linkage analysis (Savukoski et al., 1994) and the critical region had subsequently been restricted to ~200 kb by linkage disequilibrium (Klockars et al., 1996). Three PAC clones covering the 200 kb containing the *CLN5* locus were sequenced as part of the Human Genome Project at the Whitehead Institute for Biomedical Research, USA (Figure 6).

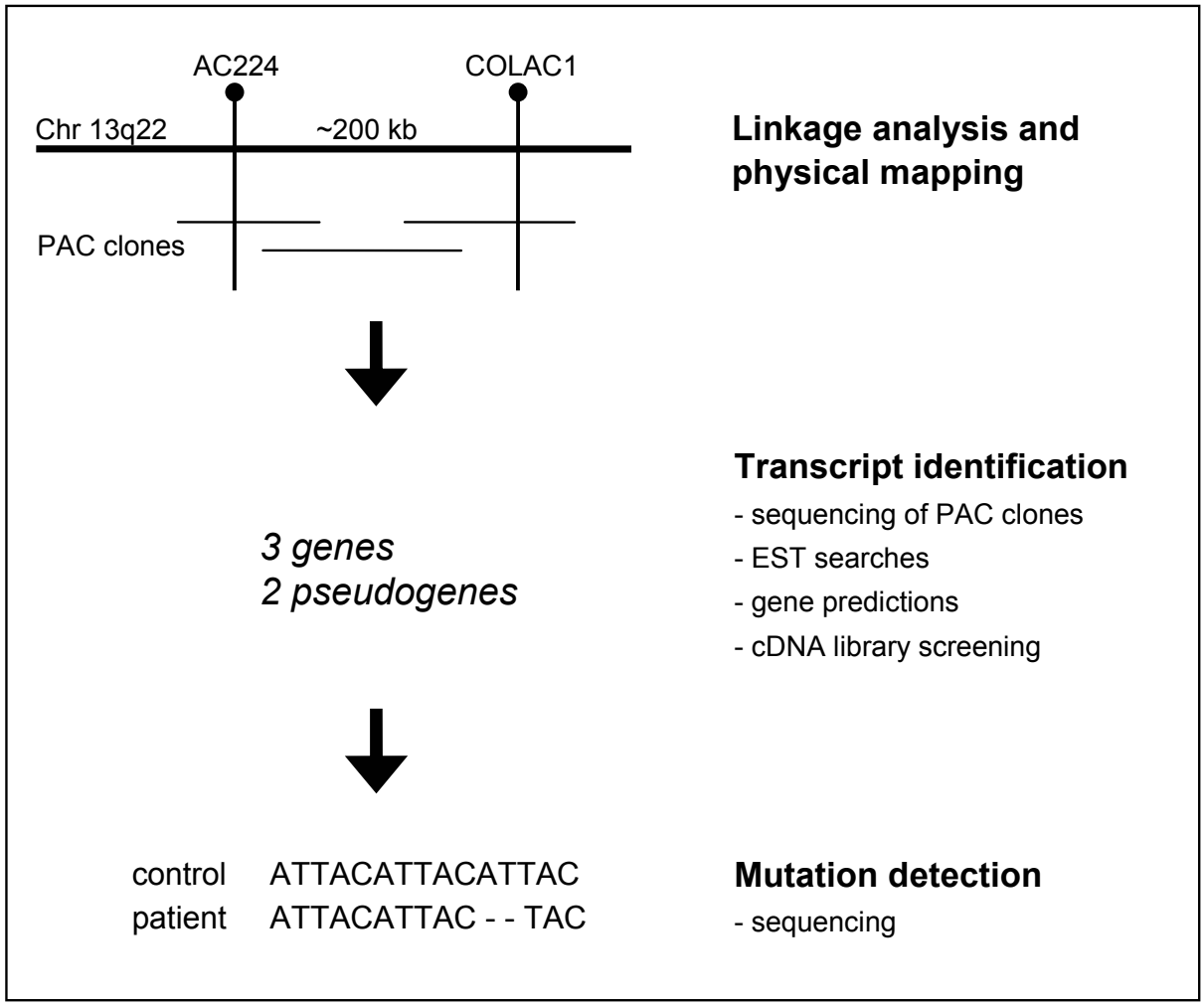

Figure 6. Strategy for the identification of the *CLN5* gene by positional cloning.

The complete genomic sequence of the region was essential for identification of the genes in the critical region (Klockars et al., 1999). The genomic sequence was used for gene predictions with the Genscan (http://genes.mit.edu/GENSCAN/) and MZEF (http://argon.cshl.org/genefinder/) softwares. The same sequence was queried against human and mouse EST databases using the BLAST algorithm (Altschul et al., 1997). In addition, traditional cDNA library screening was performed with the PAC clone located between the flanking markers. By these methods, three genes and two pseudogenes were identified on the critical *CLN5* region. Out of the 24 exons of the three genes, 15 were correctly predicted by MZEF, 20 by Genscan, and 21 by human EST homologies (Klockars et al., 1999). These three novel genes were sequenced using DNA of CLN5 patients, parents and non-related controls. Three mutations were initially identified in one of these three genes, which was therefore named the *CLN5* gene.

1.2 Characterization of the Novel *CLN5* Gene (I)

The human CLN5 gene spans a genomic region of 13 kb. The full length *CLN5* cDNA was determined by comparison and fusion of sequences from ESTs, a cDNA clone and 5'-RACE products. It resulted in a 4.1-kb cDNA sequence, consisting of 4 exons, and an open reading frame (ORF) of 1,380 bp. There are four ATG codons in the 5' end of the ORF. All of them can be used for translation initiation *in vitro* (Isosomppi et al., 2002; Vesa et al., 2002). However, only the fourth one is evolutionarily conserved (III), implying an important role for it in humans also. The 407 amino acid polypeptide translated from the first potential start methionine was predicted to contain two hydrophobic regions with the BCM Transmembrane Prediction Program and the Kyte-Doolittle hydrophobicity plot. On this basis, the CLN5 protein was considered to be a putative transmembrane protein, although no experimental evidence existed at that point to support this hypothesis. BLAST searches against the GenBank database did not show any homologies for the *CLN5* gene with any previously known genes or proteins.

The expression pattern of the *CLN5* mRNA was analyzed using a 1,291-bp PCR fragment to probe a human multiple-tissue Northern blot. Weak signals of 2.0 kb, 3.0 kb and 4.5 kb were detected in all the tissues. Despite the CNS-specific symptoms, this generalized expression is not surprising, since CLN5 patients, like all other NCL patients, show lysosomal accumulation of autofluorescent material in non-neural tissues also. Ubiquitous expression has also previously been demonstrated for *CLN1* (Camp et al., 1994) and *CLN2* (Sleat et al., 1997). However, the pattern of *CLN3* mRNA has some interesting trends. The highest expression level of *CLN3* is detected in the gastrointestinal tract, and in the glandular/secretory tissue expression is also apparently higher than in the CNS (Chattopadhyay and Pearce, 2000). CLN8 transcripts of three different sizes (1.4, 3.4, and 7.5 kb) can be detected by Northern

blotting, showing tissue-specific variation of expression levels (Ranta et al., 1999). None of the so far identified CLN proteins show higher expression levels in neural than in non-neural tissue. Thus, the CNS-specific symptoms of the patients cannot be explained by the expression patterns of the genes.

2. PHENOTYPE-GENOTYPE CORRELATION IN CLN5

2.1 Four Mutations in the CLN5 Gene (I, II)

So far four different mutations have been identified in the *CLN5* gene. The most common one is the Finnish major mutation, producing a premature stop codon that leaves the last 16 amino acids untranslated (Figure 7). This mutation is present in 88% (35/40) of the Finnish disease alleles, but has not been found outside Finland. Screening of 700 control subjects, originating from a high-risk area on the west coast of Finland, revealed a local carrier frequency of 1:24 (10/240) in a population sample from one community, where many of the ancestors of the patients have lived. The carrier frequency in the rest of the high-risk area was approximately 1:100 (4:460). A recent study gave similar results, showing a carrier frequency of 1:44 (9:396) with a sample from Kurikka, from Alajärvi and from Lapua, all belonging to the high-risk area (Pastinen et al., 2001). A sample from North Karelia in eastern Finland showed only one carrier, 1:385, and a sample from Oulu two carriers, 1:183. In a sample of 1000 people from Helsinki there was only one carrier, 1:1000. These findings correlate with the clustering of the ancestors' birthplaces and the linkage disequilibrium interval of ~11 cM suggesting that the major Finnish mutation originates from a single person who lived in southern Ostrobothnia some 20-30 generations and 500-750 years ago (Varilo et al., 1996; Peltonen et al., 1999).

The consequence of the Finnish minor mutation, found in 10% (4/40) of Finnish disease alleles, is an early stop codon in exon 1. The patients of one family in Northern Finland were homozygous for this mutation. A patient from Sweden was a compound heterozygote for this mutation in combination with a novel "Swedish" mutation. The missense mutation, found as homozygot only in one Dutch patient changes one negatively charged aspartic acid residue at position 279 to a polar but uncharged asparagine. The Swedish mutation is an insertion of a cytosine base, causing a frameshift leading to 29 wrong amino acids before a premature stop codon at position 253. This Swedish mutation is found in two compound heterozygotes, one from Finland and one from Sweden. Some of the ancestors of this Finnish patient were traced to Sweden, so, considering the population history of Finland, it seems likely that this rare mutation originates from Sweden.

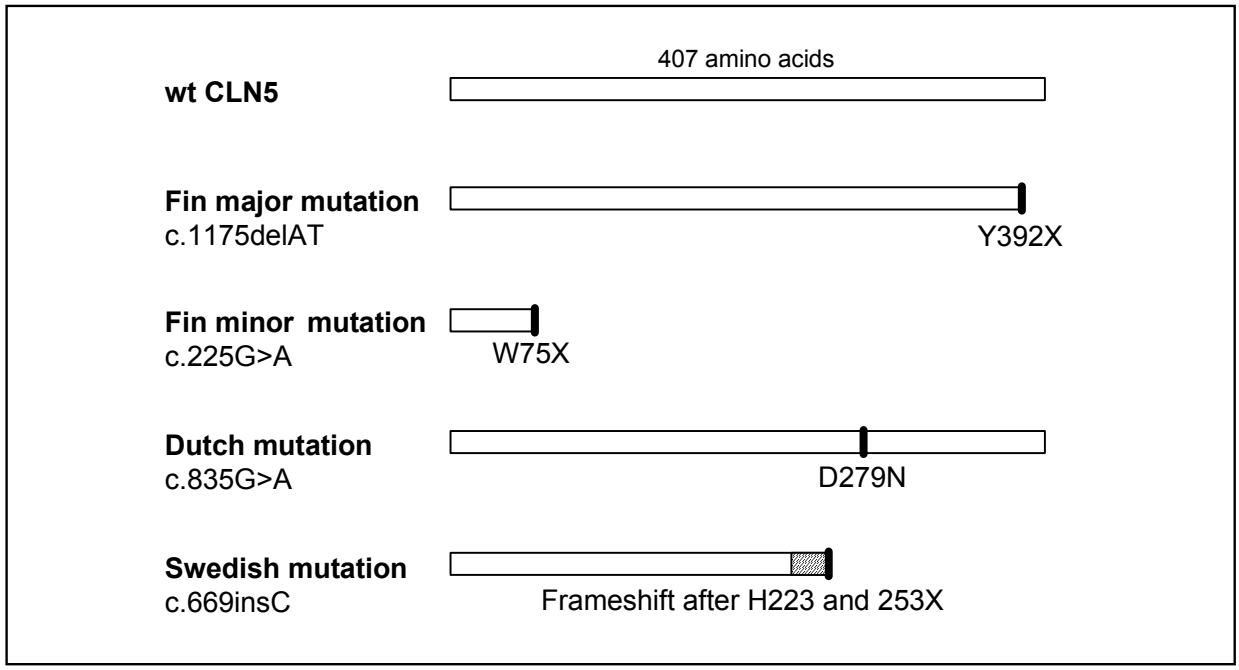

Figure 7. Predicted consequences of the four disease causing mutations on the CLN5 polypeptide. Modified from II. (Y, tyrosine; W, tryptophane; D, aspartic acid; N, asparagine; H, histidine; X, stop)

2.2 Clinical Features of Patients with Different Genotypes (II)

Clinical phenotypes of eight patients with different combinations of the four mutations were analyzed in detail: three Finnish major homozygotes, two Finnish minor homozygotes, the Dutch mutation homozygote, and two compound heterozygotes (Swedish + Fin major; Swedish + Fin minor). The detailed clinical data of the study is presented in the original publication (Table, II). The age at onset of the disease (2-4 y), as well as the ages of the initiation of: mental decline (3-7 y), visual deterioration (5-8 y), ataxia (7-11 y), epileptic seizures (7-11 y), lost walking ability (8-11 y), abolished ERG (6-7 y), and giant fVEP (6-10 y) were assessed. None of these indicators showed any clear correlation with any of the genotypes. However, the spectrum of clinical development was broad even between patients with identical mutations, with variation of 2-4 years in the onset of several signs.

Magnetic resonance images (MRIs) of six of the patients were also analyzed. All the patients had enlarged cerebellar sulci, an atrophic vermis, and a dilated IVth ventricle. The corpus callosum was abnormally thin in all but one patient. Moderate to severe cerebral atrophy was noticed in all three examinations of patients over 9.5 years of age. All patients had increased signal intensity of the periventricular white matter and the posterior limb of the capsula interna on T2-weighted images. In addition, all the patients had abnormally low signal intensity of the thalamus compared to the putamen and caudate nuclei at all ages. These results were in line with the previous brain

imaging reports of CLN5 patients, emphasizing cerebellar atrophy as the most striking structural abnormality (Santavuori et al., 1982; Autti et al., 1992).

2.3 Phenotype-Genotype Correlation in CLN5 (II)

The four CLN5 mutations have very different predicted consequences at the polypeptide level (Figure 7). However, no distinct difference was observed in the spectrum of symptoms, age of onset or in the speed of progression in the patients with different mutations. Thereby it appears likely that each of the four mutations severely affects the normal function of the protein, resulting in a functional null allele. Expression studies have now shown that the Fin major mutation severely interferes with the lysosomal targeting of the CLN5 protein (Isosomppi et al., 2002). It is likely that also the Fin minor and Swedish mutations, causing more severely truncated polypeptides, result in premature protein degradation, misfolding, or disrupted lysosomal targeting. The Dutch mutation changes one negatively charged aspartic acid residue to a polar but uncharged asparagines residue. The severity of this missense mutation suggests that the mutation affects a functionally or structurally important domain of the protein.

In contrast to CLN5, specific mutations causing protracted phenotypes have been identified for most of the other NCL disorders (Mole et al., 1999; Mole et al., 2001a). In addition to the infantile phenotype, *CLN1* mutations might cause late-infantile, juvenile and adult onset diseases. *CLN2* mutations cause both late-infantile and juvenile forms, and a *CLN3* mutation with protracted progression of the symptoms has been identified. Mutations in the *CLN8* gene cause two clearly distinct phenotypes: late-infantile NCL and progressive epilepsy with mental retardation. Detailed, updated information about all NCL mutations can be found in the NCL Mutation Database maintained by Dr. Sara Mole (http://www.ucl.ac.uk/ncl/).

3. THE MOUSE *Cln5* GENE AND PROTEIN

3.1 Expression and Localization of *Cln5* in the Mouse Brain (III)

The mouse *Cln5* gene consists of four exons and a coding sequence of 1026 bp, translated into a polypeptide of 341 amino acids. An expression construct was made by PCR amplification of the coding region from the mouse cDNA IMAGE clone #2135359 and insertion of this fragment into the expression vector pcDNA3.1.

The human *CLN5* gene was shown to provide transcripts of 2.0 kb, 3.0 kb and 4.5 kb in all the tissues studied (I). In the Northern blotting of mouse *Cln5* we detected, in all the tissues studied (heart, brain, spleen, lung, liver, skeletal muscle, kidney and testis) a

major transcript of 2.9 kb and a minor transcript of 2.3 kb (III, Fig. 1 A). Similar to human *CLN5*, some additional bands were seen in the skeletal muscle and lung tissues. This may indicate that *Cln5* is subjected to tissue-specific alternative splicing or polyadenylation signals. The expression level of *Cln5* is generally very low. In the brain at postnatal day 30 (P30) the steady state transcript level is higher than at P7 and embryonal day 13 (E13) (III, Fig. 1 B), suggesting that *Cln5* gene expression in the brain is under developmental control and increases during brain maturation. Previous studies have shown that the expression of *Ppt1* and *Cln3* are also developmentally regulated (Isosomppi et al., 1999; Suopanki et al., 1999a; Suopanki et al., 1999b; Ahtiainen et al., 2003). This could indicate that these *CLN* genes play an important, although so far unknown, role in the development and maturation of neurons.

Expression of *Cln5* mRNA in the CNS was further studied by non-radioactive *in situ* hybridization. Analyses of embryonic and postnatal mouse brains indicated that *Cln5* is expressed during early development but also in the mature brain. In agreement with the Northern hybridization results, *Cln5* gene expression could be detected in several developing brain regions as early as embryonic day 15 (E15). At E15, *Cln5* mRNA was distributed throughout the developing brain, e.g., in the ventricular zone of the developing cortical plate, in the ganglionic eminence, and in the hippocampus (III, Fig. 2).

In the adult mouse brain, *Cln5* expression was most intense in the Purkinje cell layer of the cerebellum, in the cerebral cortex, as well as in the hippocampal pyramidal cells (III, Fig. 2 G-L). The spatial expression pattern provides a striking parallel with the CLN5 patients' MRI findings which show severe cerebellar and cerebral atrophy, as well as a decreased thalamic signal intensity (Autti et al., 1992). The results from the mouse brain are also well in line with our earlier studies of the developing human brain, demonstrating *Cln5* expression especially in the cerebral cortex, the cerebellar Purkinje cells and the thalamic regions (Heinonen et al., 2000b). Interestingly, these findings show parallels with the *in situ* expression pattern reported for *Cln1* and *Cln3* (Isosomppi et al., 1999; Luiro et al., 2001).

Immunohistochemical stainings were performed on brain sections from P7 to P60 mice (III, Fig. 3 A-I) and the findings were well in line with the *Cln5* mRNA distribution pattern. Prominent immunostaining was demonstrated in the hippocampal interneurons occupying the stratum radiatum, stratum oriens, and the pyramidal cell layer (III, Fig. 3 G-I, arrows). Moreover, a specific labeling of CA3 pyramidal cells was found, whereas labeling in the CA1 region was much lower. In the large neurons the Cln5 staining had a granular appearance and was localized predominantly in the cell soma. However, some immunostaining could also be seen in the neuronal extensions (III, Fig. 3 B, D, E).

Immunohistochemical stainings corresponded well with the *in situ* hybridization results. Moreover, the Cln5 expression is found in diverse types of neurons, for example in the inhibitory Purkinje cells of the cerebellum and in the excitatory mitral cells of the olfactory bulb, rather than being restricted to any specific neuronal subpopulation. Subcellularly Cln5 localized predominantly to the neuronal cell soma, although some weaker immunostaining was also detectable in the neuronal extensions. Expression of the Cln5 in the hippocampal interneurons is interesting since earlier studies on NCL patients as well as animal models have supplied evidence for dysfunction of select inhibitory (GABAergic) circuits, resulting in neuronal loss and hypertrophy of the remaining neurons (Williams et al., 1977; Braak and Goebel, 1978; Braak and Goebel, 1979; Walkley and March, 1993; March et al., 1995; Cooper et al., 1999). The expression pattern of Cln5 in the brain and neurons is very similar to that reported previously for Cln1 and Cln3, suggesting a shared molecular pathway in different NCL disorders (Isosomppi et al., 1999; Luiro et al., 2001; Cotman et al., 2002).

3.2 Intracellular Targeting of the Mouse Cln5 Protein (III)

The subcellular localization of the mouse Cln5 protein was first studied with confocal immunofluorescence microscopy of transiently transfected HeLa and COS-1 cells. When cells were fixed with 4% paraformaldehyde followed by permeabilization with methanol at -20ºC, the rabbit anti-Cln5 antibody showed a vesicular pattern. In both HeLa and COS-1 cells these vesicles overlapped with the lysosomal Lamp1 marker H4A3 (III, Fig. 6 A-F). Even some ER-like staining was observed, but disappeared after treatment with cycloheximide.

The Finnish major mutation prevents lysosomal targeting of the prematurely terminated CLN5 polypeptide, implying that the pathogenesis is associated with disturbed intracellular trafficking (Isosomppi et al., 2002). Lysosomal targeting of soluble enzymes occurs mainly through the mannose-6-phosphate receptor mediated pathway (Rouille et al., 2000) but this has not been verified for CLN5. However, further studies are needed to describe the lysosomal targeting mechanism and putative enzymatic function of the CLN5 protein.

The intracellular localization was further studied in mouse primary neurons. Since no endogenous staining was detected with our antibody, we transfected the neurons with mouse *Cln5*. In the soma of transfected neuronal cells Cln5 was partially colocalized with the lysosomes, although ER-like staining was also observed, as in nonneuronal cells (III, Fig. 6 G-J). Interestingly, distinct Cln5 immunostaining was also detected in the neuronal extensions corresponding to the Cln5 immunohistochemistry in the brain. This staining did not colocalize with the lysosomal marker ID4B, indicating that in neurons, Cln5 may have an additional, extralysosomal function.

This neuronal staining pattern is not totally surprising, since a similar neuronal distribution has previously been demonstrated for PPT1 and CLN3 (Luiro et al., 2001; Ahtiainen et al., 2003). This finding may indicate extralysosomal functions in neurons for the Cln5 protein. To understand the reason for the CNS specificity of the symptoms and the cell degeneration in the patients, the intracellular targeting and the molecular interactions of CLN5 in neurons need to be clarified in detail. This will be of great interest since previous studies have described the transport of PPT1 and CLN3 into the neuronal synapses (Lehtovirta et al., 2001; Luiro et al., 2001; Ahtiainen et al., 2003). As CLN2 and CLN3 are potential interaction partners for CLN5 (Vesa et al., 2002), it will be interesting to see whether these three proteins follow the same novel intracellular transport pathway in neurons.

3.3 Comparison of the CLN5 Proteins of Different Species (I, III)

The CLN5 protein sequences of six different species were compared with each other. The human and mouse sequences are publicly available (GenBank accession nos. NP_006484 and XP_127882). The coding sequence of the *Dictyostelium discoideum CLN5* cDNA was obtained from a contig of five EST sequences (AU061512, BJ415600, BJ425060, C23678, C91571). The partial polypeptide sequences of pig (*Sus scrofa*; BI182240, BI304716), western clawed frog (*Silurana tropicalis*; BE468325), and channel catfish (*Ictalurus punctatus*; AL468325) were translated from EST sequences. Beginning from amino acid residue 102 of the human CLN5 polypeptide (WPVPY...) and reaching to the C-terminus, a striking homology is seen between the amino acid sequences of man, mouse, pig, catfish, and frog (Figure 8). The amino acid identity between man and mouse in this region is 80%, and the similarity 89%. For all the species studied, an N-terminal hydrophobic stretch is observed in the region corresponding to the amino acids 75-91 of the human polypeptide. The SignalP v2.0 program (http://www.cbs.dtu.dk/services/SignalP-2.0/) predicts this region to be a signal sequence cleaved by signal peptidase on the lumenal side of the ER membrane. BLAST comparisons of the genomic sequences 100 kb upstream of the human CLN5 and mouse Cln5 genes did not reveal any homologous promotor regions.

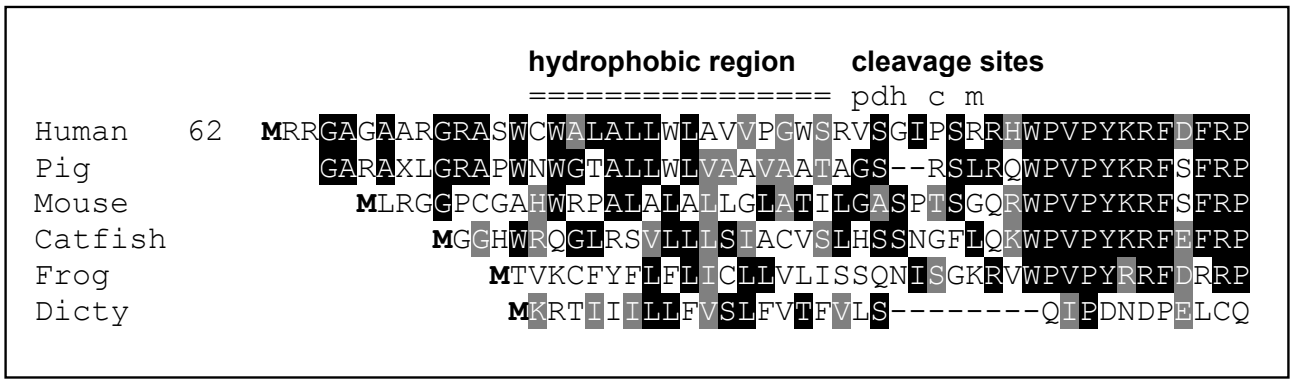

Figure 8. Comparison of the N-terminal CLN5 polypeptide sequences of six species. This sequence begins with the fourth potential start methionine of the human CLN5 peptide, located 62 amino acids downstream from the first one. The hydrophobic region, typically observed in signal pepties is marked by equal signs (=). The predicted signal peptide cleavage sites are marked by h (human), p (pig and frog), m (mouse), c (catfish) and d (dictyostelium). Black background indicates identical amino acid and grey background similar amino acid. Modified from III.

To analyze experimentally whether the N-terminal hydrophobic region of mouse Cln5 can act as a signal sequence, we compared the sizes of *in vitro* translated *Cln5* and deglycosylated Cln5 protein produced in COS-1 cells. Western analysis revealed that the apparent size of the deglycosylated Cln5 protein was 34 kDa, while the *in vitro* translation produced a band of 37 kDa (III, Fig. 5B). This would imply that mouse Cln5 undergoes signal sequence cleavage resulting in the removal of an approximately 3 kDa peptide.

To monitor the biosynthesis of the Cln5 protein, we transfected COS-1 cells with the mouse *Cln5* cDNA construct. The crude cell lysates were deglycosylated with Endo H or PNGase F to remove the high mannose oligosaccharides or all N-linked oligosaccharides. Western blot analysis revealed two bands of 48 and 50 kDa from the *Cln5* transfected cells. Treatment with Endo H produced one band of 35 kDa, and PNGase F treatment resulted in a 34 kDa band (III, Fig. 5C). Our results suggest that mouse Cln5 is a glycoprotein with high mannose-type sugars and additional complex-type sugars.

The human *CLN5* cDNA sequence has four in-frame AUG codons, previously shown to encode the polypeptides of 407, 378, 358 and 346 amino acids (Isosomppi et al., 2002). Only the fourth AUG is conserved in the other species (Figure 8). To analyze the significance of the fourth methionine for translation initiation, we transfected COS-1 cells with the complete human *CLN5* cDNA, and a cDNA with the first three potential start methionines inactivated (*CLN5*met4 construct). The *CLN5* and *CLN5*met4 constructs showed bands of identical size before and after deglycosylation

(III, Fig. 5B). This confirms that the fourth methionine of human *CLN5* can alone act as a translation initiator.

Table 5 shows the properties of the human CLN5 proteins translated from the first and fourth methionines (met1 and met4), compared to the mouse Cln5 protein. The similarities between the human CLN5met4 and the mouse Cln5 are striking. The lengths of the polypeptides, the molecular weights, and the glycosylation pattern are very similar, and both are predicted to undergo signal peptide cleavage, in contrast to CLN5met1. This would imply that the human CLN5 protein translated from the fourth methionine is carrying out the same function as the Cln5 protein in the mouse, and in all other species with identified CLN5 proteins. The use of downstream AUG codons for translation initiation is actually more common than is generally believed. A recent study shows that more than 40% of transcripts contain AUG codons upstream of the actual start codon (Peri and Pandey, 2001). The authors of that work suggest that for the purpose of identification of translation start sites from cDNA sequences, it is better to use homology-based alignments across species or across protein families, instead of relying solely on the most upstream AUG and its context.

Table 5. This table shows the basic features of the human CLN5 proteins translated from the first and fourth methionines (met1 and met4), compared to the mouse Cln5 protein. The 407 amino acid polypeptide translated from met1 is not predicted to undergo signal peptide (sp) cleavage. *The values indicated with a star are from Vesa et al., 2002.

	CLN5met1	CLN5met4	Mouse Cln5
Coding sequence (base pairs)	1224	1041	1026
Amino acids	407	346	341
Amino acids after sp cleavage	-	312	308
Molecular Weight (MW)	47*	50-60	50
Molecular Weight - Endo H	47*	38	35
Molecular Weight - PNGase F	47*	37	34
In vitro translation	47*	40	37
Predicted MW before sp-cleavage	46.3	40.3	39.1
Predicted MW after sp-cleavage	-	36.6	36.1
GenBank acc. no. (genomic)	NT_024524	NT_024524	NT_039609
GenBank acc. no. (mRNA)	NM_006493	NM_006493	XM_127882
GenBank acc. no. (protein)	NP_006484	NP_006484	XP_127882

However, the first initiation methionine of CLN5 has been shown to be capable of translation initiation *in vitro* (Isosomppi et al., 2002; Vesa et al., 2002). Using an N-terminal antibody (raised against amino acids 1-75), a 47-kDa band was detected in Western blotting (Vesa et al., 2002). Moreover, this protein was shown to be

membrane bound and to lack any N-linked glycans. This 47-kDa isoform can be produced *in vitro* in transient cDNA expression systems, but whether it also has a role *in vivo* remains to be clarified.

The 5' end of the human CLN5 gene was cloned by 5'RACE and so far no EST covering the first AUG codon has been published. This could indicate that the mRNA including that codon is an alternatively spliced variant, with tissue-specific or developmentally regulated expression. The fact that the 5' end of the coding sequence is not conserved in the mouse genome supports this hypothesis. Analyses of 9,434 orthologous genes in human and mouse have shown that exons that are only included in alternative splice forms are mostly not conserved and thus are products of recent exon creation (Modrek and Lee, 2003). As the total numbers of genes in human and mouse are almost identical, it seems likely that much of the recent evolutionary changes causing the differences between mammals are caused by alternative splicing. Thus, it is possible that human CLN5 encodes two different proteins, although mouse Cln5 encodes only one.

3.4 Is CLN5 a Transmembrane or a Soluble Protein? (I, III)

In publication I we speculated that CLN5 could be a transmembrane protein. This was based on computational analysis of the 407 amino acid polypeptide, as we assumed that the first AUG is the initiator of translation. This prediction was made using the TMpred transmembrane prediction program and Kyte-Doolittle hydrophobicity plot. However, the ability of these programs to distinguish between soluble and membranous proteins is poor (Moller et al., 2001). Currently the best performing programs are TMHMM (http://www.cbs.dtu.dk/services/TMHMM/)(Krogh et al., 2001) and SOSUI (http://www.tuat.ac.jp/mitaku/sosui/)(Hirokawa et al., 1998). Neither of them predicts any transmembrane regions for CLN5 in any species. The stronger of the two hydrophobic regions predicted, has been shown to belong to the signal sequence of the CLN5 polypeptide translated from methionine 4 (III).

To further pursue whether mouse Cln5 represents a membrane-bound or a soluble protein, Triton X-114 phase separation was performed. With this method soluble proteins are found exclusively in the aqueous phase, and membrane proteins recovered in the detergent phase (Bordier, 1981). Transfected COS-1 cells were lysed with Triton X-114 buffer, condensated, centrifuged, and the two phases were then separated and analyzed by Western blotting using the guinea pig anti-Cln5 antibody. Mouse Cln5 and also the 312 amino acid human CLN5 protein analyzed in parallel experiments were found exclusively in the aqueous phase (Figure 9). Protein Disulphide-Isomerase (PDI), which is a soluble protein of the endoplasmic reticulum, was endogenously detected from the *Cln5* transfected cells. It was recovered exclusively in the aqueous phase. The human transferrin receptor, which is a type II transmembrane protein with a

single 28 residue transmembrane region, was found predominantly in the detergent phase of the lysate of *Cln5* transfected cells.

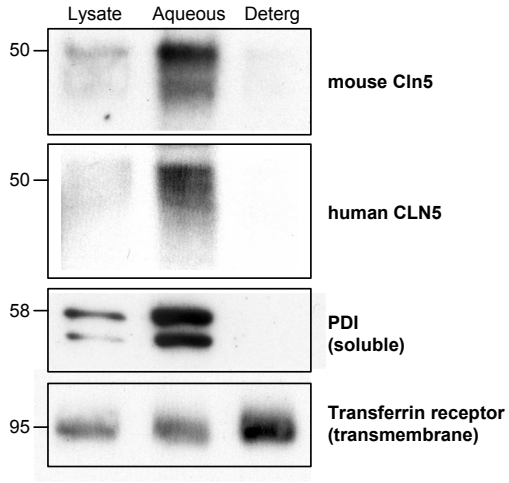

Figure 9. Triton X-114 phase separation of transfected COS-1 cells. The crude lysate, the aqueous phase and the detergent phase were analyzed by Western blotting. Modified from III.

These results clearly indicate that the CLN5 proteins of human and mouse, after signal peptide cleavage, are soluble proteins. Howerver, the 407 amino acid polypeptide translated from the first AUG codon has been detected from the detergent phase after Triton X-114 phase separation (Vesa et al., 2002). The location and function of this putative transmembrane isoform, possibly encoded by an alternatively spliced mRNA, remains elusive.

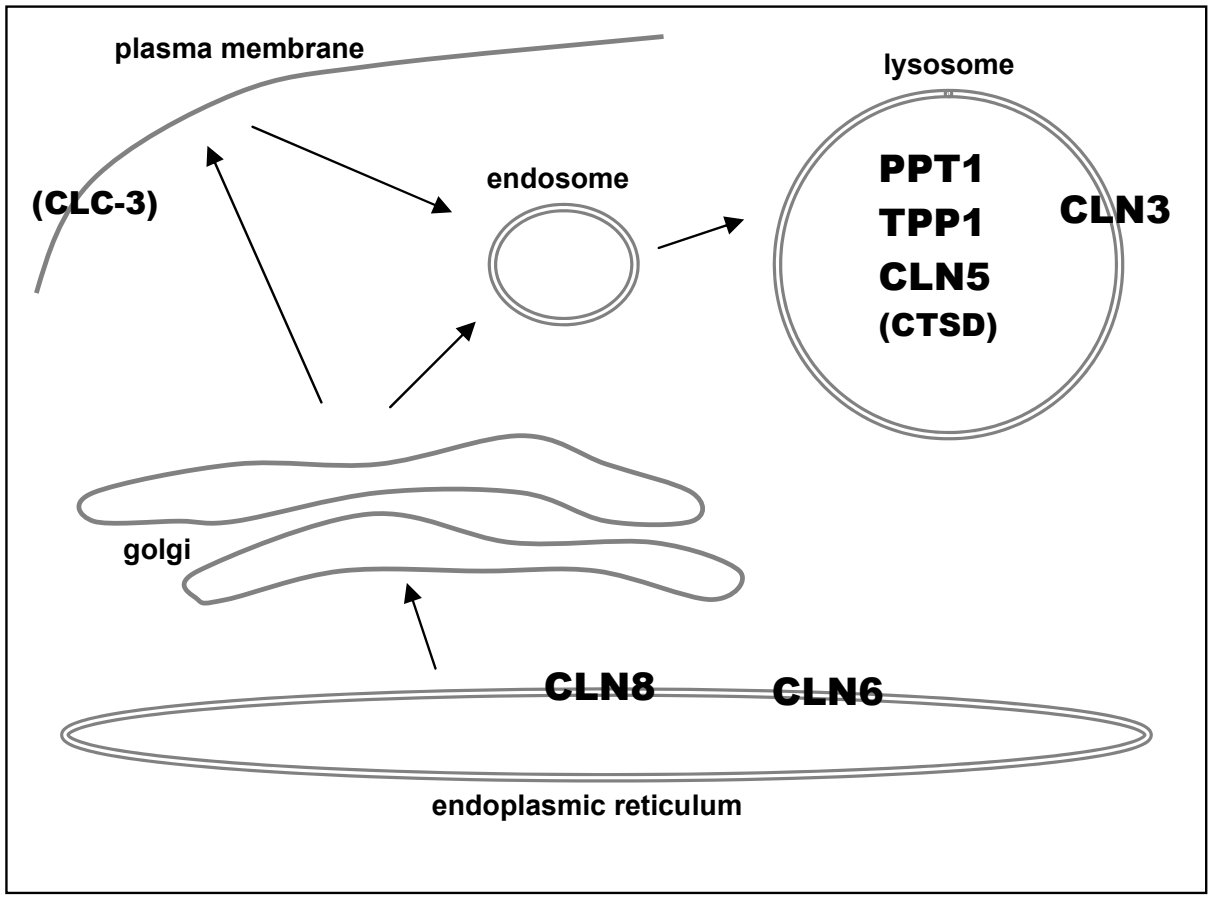

Figure 10. Schematic picture of the intracellular localization of the CLN proteins. Cathepsin D (CTSD) and Chloride channel-3 (CLC-3) deficient mice manifest NCL-like symptoms and lysosomal accumulation.

Among the six CLN proteins, PPT1, TPP1 and CLN5 have now been shown to represent soluble lysosomal proteins (Figure 10). CLN3 is a lysosomal membrane protein, and CLN6 and CLN8 are predicted membrane proteins localized to the endoplasmic reticulum (ER). Cathepsin D is a lysosomal enzyme, deficiency of which causes NCL in sheep and mouse (Koike et al., 2000; Tyynelä et al., 2000). Chloride channel-3 deficient mice also manifest NCL-like symptoms, which have been suggested to be associated with defective endosomal pH regulation (Yoshikawa et al., 2002). This implies that dysfunction of the ER-golgi-endosome-lysosome biosynthetic-secretory pathway might be essential for the molecular pathogenesis on the NCLs. However, extra-lysosomal localization has also been suggested for PPT1, CLN3 and CLN5 in neurons. The potential synaptic functions and intraneuronal targeting mechanisms of all the CLN proteins need to be clarified to better understand the pathogenesis of the corresponding NCL diseases.

CONCLUSIONS AND FUTURE PROSPECTS

The neuronal ceroid lipofuscinoses (NCLs) are globally the most important cause of childhood neurodegeneration. The clinical course is characterized by mental, motor and visual decline, further complicated by seizures. Utilizing the recent development of molecular genetics, six defective genes have been identified behind these diseases. The hallmark of the pathology is the accumulation of ceroid and lipofuscin resembling autofluorescent material in the lysosomes of both neuronal and non-neuronal cells. However, the molecular cascade and the reasons for the CNS specificity of the symptoms and for cell death are not understood. The NCL proteins can be roughly divided into two groups, lysosomal enzymes (PPT1, TPP1, Cathepsin D and CLN5) and membrane proteins (CLN3, CLN6 and CLN8). Currently, mouse models of all NCL diseases are available and will provide a useful tool for further studies of the pathological mechanisms and ultimately for therapeutic trials.

In this study we have identified the *CLN5* gene by positional cloning strategies. Four different mutations were detected in CLN5 patients, but, despite detailed analysis of the clinical development, we could not recognize any correlation between the phenotypes and the genotypes. To find out more about the general properties of the CLN5 proteins in different species, we characterized the mouse *Cln5* gene and showed that it encodes a soluble lysosomal glycoprotein, expressed throughout the developing and mature brain. We also demonstrated that mouse Cln5 undergoes signal peptide cleavage and closely resembles the human CLN5 protein translated from the fourth potential AUG codon.

So far, the reason for the neuron-specific cell degeneration seen in all NCL patients is not understood. Suggested explanations for the neuron-specific effects have been many: neurons may be more vulnerable to the storage materials, there might be neuron-specific substrates for the NCL enzymes, or neuron-specific metabolic pathways could be used for all NCL-proteins. Our results with Cln5 as well as previous studies with Cln1 and Cln3 suggest that these proteins might have an extra-lysosomal function in neurons, not observed in other cells. Further studies are needed to explain the potential synaptic function and the intracellular targeting mechanisms of CLN proteins in neurons.

An interesting finding is that most of the NCL disorders seem to be linked to dysfunction of inhibitory GABAergic hippocampal interneurons. In the serum of CLN3 patients and mice, an autoantibody to glutamic acid decarboxylase (GAD65) has been detected, which inhibits conversion of glutamate to GABA and so leads to an accumulation of glutamate (Chattopadhyay et al., 2002). Degeneration of interneurons has been shown in *Ppt1* mice (Cooper, 2003), *Cln3* mice (Mitchison et al., 1999), *mnd*

(Cln8) mice (Cooper et al., 1999), and ovine CLN6 (Jolly and Walkley, 1999). The *mnd (Cln8)* mice show increased levels of glutamate in combination with decreased GABA in the cerebral cortex tissue. (Elshatory et al., 2003) Microarray studies have suggested downregulation in the *Cln3* mice of two genes associated with glutamate utilization, namely glutamine synthetase and glutamate receptor mGluR3 (Griffin et al., 2002). Thus our prominent immunohistochemical staining of Cln5 in the hippocampal interneurons might be an important finding.

At present, very little quantitative data is available about which cellular populations are affected in NCLs and how the sequence of neurodegenerative events subsequently develops and ultimately leads to cell death. Discovery of the precise nature and order of these events will be crucial for understanding the neurobiology and the therapeutic opportunities for these devastating diseases. The mouse models will be useful in future studies on the mechanisms leading to cell death, and especially on the potential involvement of apoptosis. The role of the accumulation material in the lysosomes is still an essential question. Is it the primary cause of neuronal cell death, or just a secondary finding relating to the lost protein activity?

New approaches will be essential for analysing the animal models of NCLs, including microarrays for monitoring detailed mRNA expression profiles in different tissues and developmental stages. Identification of downstream genes involved in the same metabolic pathways as the NCL-proteins, would be important. Microarrays could also give new clues about the role of genes related to the metabolism of glutamate in the pathogenesis of different NCLs.

Comparative proteomics will provide a means for examining the genetic and environmental effects on protein expression in control and NCL mice. The role of potential molecular interactions between NCL proteins and emerged candidate proteins needs to be characterized to find out whether the NCL proteins are involved in a common metabolic pathway.

More generally, characterization of the molecular pathogenesis of the NCLs could give new insights into the mechanisms involved in neurodegeneration, neuron-specific apoptosis, brain development and survival of neurons. Despite the progress followed by the identification of the defective NCL genes, successful treatments for these fatal disorders remain a distant prospect. Accurate and reliable diagnostic methods, including prenatal and carrier diagnostics, and genetic counselling are the clinical benefits achieved so far.

ELECTRONIC INFORMATION

Internet addresses for databases, programs and organizations mentioned in the text are provided below.

World Wide Web Site	URL
Batten Disease Support & Research Association	http://www.bdsra.org/
BCM Search Launcher	http://searchlauncher.bcm.tmc.edu/
Bioinformatics@KTL	http://www.ktl.fi/bioinfo/
BLAST at NCBI	http://www.ncbi.nlm.nih.gov/BLAST/
CBS Prediction Servers	http://www.cbs.dtu.dk/services/
EMBnet	http://www.ch.embnet.org/
Ensembl	http://www.ensembl.org/
ExPASy	http://us.expasy.org/
GENSCAN Web Sever at MIT	http://genes.mit.edu/GENSCAN/
Guide for the Care and Use of Laboratory Animals	http://bob.nap.edu/html/labrats/
Human Genome Organisation (HUGO)	http://www.hugo-international.org/
Human Genome Project Information	http://doegenomes.org/
I.M.A.G.E. Consortium	http://image.llnl.gov/
Mouse Genome Informatics	http://www.informatics.jax.org/
MZEF Gene Finder	http://argon.cshl.org/genefinder/
National Center for Biotechnology Information	http://www.ncbi.nlm.nih.gov/
National Human Genome Research Institute	http://www.genome.gov/
NCL Mutation Database	http://www.ucl.ac.uk/ncl/
Online Mendelian Inheritance in Man	http://www.ncbi.nlm.nih.gov/Omim/
Policy&Ethics	http://www.genome.gov/PolicyEthics/
Swiss-Prot	http://www.ebi.ac.uk/swissprot/
The Institute for Genomic Research	http://www.tigr.org/
UCSC Genome Browser	http://genome.ucsc.edu/
UNESCO	http://www.unesco.org/
World Medical Association	http://www.wma.net/

ACKNOWLEDGEMENTS

This study was carried out at the Department of Molecular Medicine of the National Public Health Institute, Helsinki. The former head of the institute, Jussi Huttunen, and the new director, Pekka Puska, are acknowledged for providing excellent research facilities. The study was financially supported by Finska Läkaresällskapet, Suomen Lääketieteen Säätiö, Perkléns minnesfond, and the Helsinki Biomedical Graduate School.

For almost eight years I have had the privilege to be supervised by Leena Peltonen-Palotie. I am grateful for the opportunity to work in such a challenging, competitive, and creative environment. Anu Jalanko has been very helpful as co-supervisor and has always been available for guidance. I want to thank Outi Kopra for skilful assistance and practical help, especially regarding the manuscript of the last publication. I am grateful to Irma Järvelä for taking the initiative to the phenotype-genotype correlation work.

I want to thank Pirkko Santavuori for introducing me to the clinical viewpoint of the NCL-diseases. The opportunity to cooperate with an experienced clinician and the chance to meet some of the CLN5 patients and their parents has been invaluable. I want to thank all CLN5 patients, their families, and their clinicians for everything they have done to contribute to this project. Without their friendly and patient cooperation this study would not have been possible. I am particularly grateful to our collaborators Paul Uvebrant and Irene Hofman. Leena Lauronen and Taina Autti are acknowledged for excellent radiological expertise. Anna-Liisa Fabritius performed the immunohistochemical stainings for the last publication and Satu Kuure helped us with the in situ hybridizations. I am grateful to Jouni Vesa for providing us with cDNA constructs and Cln5-specific antibodies. Tuula Manninen is acknowledged for excellent technical assistance.

I wish to express my sincere gratitude to Helena Kääriäinen and Kari Majamaa for reviewing this thesis. Their comments and constructive criticism were important for improving the manuscript. Hannu Sariola is thanked for participating in my thesis committee. Jean Margaret Perttunen is acknowledged for revising the language of this thesis. I want to thank our secretaries, Sari Kivikko, Sari Mustala, Tuija Svahnbäck, and Helena Knuuttila for friendly assistance. Teemu Perheentupa, Juri Ahokas, Jari Raikko, and Martti Sariola have been of great help with a myriad of computer problems.

I will never forget the wonderful atmosphere of our CLN5 team. I want to thank Tuomas Klockars for introducing me to the field of genetics, for teaching me all the

basic methods, and for creating a relaxed and positive spirit in the lab. The discussions with Minna Savukoski about genetics and medicine have really taught me a lot. Juha Isosomppi has been invaluable in giving practical hints about different methods and thorough explanations about their theoretical backgrounds. Our hour-long discussions about the deepest essence of the CLN5 protein have been both fun and scientifically challenging.

For making the everyday-life in the lab most enjoyable I want to thank my present and former room mates Heidi, Elina, Aimee, Kaisu N, Henna, and Anna. For great company during the years I want to thank Joni, Jenny, Pekka, Tero H, Tintti, Will, Jani, Kai, Maria, Gilberto, Tanja, Taina, Nora, Laura, Kaisu L, Nina, Paula, Tarja, Anniina, Kristiina, Annika, Mira, Outi, Jenni, Emma, Ilona, Nabil, Mikko, Tero Y, Markus, Marjo, Tiina, Teppo, and Iski.

The countless coffee breaks with Jesper, flavoured with philosophical discussions about science, politics and life, created a friendship that helped me to survive even the toughest phases of this project. I have had the pleasure to spend the greater part of my time in the lab, as well as the first steps in the clinics, together with Niklas, mainly discussing tennis. Juha P has provided good humour lab bench company and shown a very sound attitude towards science. I want to thank Anil for nice company and many interesting reflections on neurology, cardiology, and fig alpha. Since the first moments in the Graduate School, Jonas has been an important friend for sharing the frustration about all madness in biomedical research and clinical medicine. PKU is also acknowledged.

I want to thank all my friends who have helped to keep my mind away from genetics. Especially all discussions with Marcus, Hamppe, Samuel, and Mika have been very therapeutic.

I also wish to thank all my cousins and their families for all joyful moments together. I am grateful to may aunt Britt, for her true caring and helpfulness. My grandparents, Ulla and Helge, have always been encouraging and shown great interest in my medical and scientific activities. My wonderful sisters, Maya and Hanna, are marvellous company for relaxing, chatting and enjoying life. I want to express my warmest thanks to my parents, Andrea and Kenneth, for awakening my interest for science and for teaching me to question prevailing truths. Your everlasting support and encouragement has been invaluable.

Most of all I want to thank Johanna for sharing her life with me. Your love and support mean everything to me.

Ville Holmberg
Helsinki, February 2004

REFERENCES

Adams, MD, Celniker, SE, Holt, RA, Evans, CA, Gocayne, JD, Amanatides, PG, Scherer, SE, Li, PW, Hoskins, RA, Galle, RF, George, RA, Lewis, SE, Richards, S, Ashburner, M, Henderson, SN, Sutton, GG, Wortman, JR, Yandell, MD, Zhang, Q, Chen, LX, Brandon, RC, Rogers, YH, Blazej, RG, Champe, M, Pfeiffer, BD, Wan, KH, Doyle, C, Baxter, EG, Helt, G, Nelson, CR, et al. (2000) The genome sequence of Drosophila melanogaster. Science 287:2185-2195.

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

Altschul, SF, Madden, TL, Schaffer, AA, Zhang, J, Zhang, Z, Miller, W and Lipman, DJ. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389-3402.

Andrews, LB and Nelkin, D. (1996) The bell curve: a statement. Science 271:13-14.

Autti, T, Raininko, R, Launes, J, Nuutila, A and Santavuori, P. (1992) Jansky-Bielschowsky variant disease: CT, MRI, and SPECT findings. Pediatric Neurology 8:121-126.

Autti, T, Raininko, R, Vanhanen, SL, Kallio, M and Santavuori, P. (1994) MRI of the normal brain from early childhood to middle age. I. Appearances on T2- and proton density-weighted images and occurrence of incidental high-signal foci. Neuroradiology 36:644-648.

Autti, T, Raininko, R, Vanhanen, SL and Santavuori, P. (1996) MRI of neuronal ceroid lipofuscinosis. I. Cranial MRI of 30 patients with juvenile neuronal ceroid lipofuscinosis. Neuroradiology 38:476-482.

Autti, T, Raininko, R, Santavuori, P, Vanhanen, SL, Poutanen, VP and Haltia, M. (1997) MRI of neuronal ceroid lipofuscinosis. II. Postmortem MRI and histopathological study of the brain in 16 cases of neuronal ceroid lipofuscinosis of juvenile or late infantile type. Neuroradiology 39:371-377.

Avery, OT, MacLead, CM and McCarty, M. (1944) Studies of the chemical nature of the substance inducing transformation of pneumococcal types. Induction of transformation by a desoxyribonucleic acid fraction isolated from Pneumococcus Type III. J Exp Med 79:137-158.

Batten, FE. (1903) Cerebral degeneration with symmetrical changes in the maculae in two members of a family. Trans Ophtalmol Soc UK 23:386-390.

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

Bielschowsky, M. (1913) Über spät-infantile familiäre amaurotische idiotie mit Kleinhirnsymptomen. Dtsch Z Nervenheilk 50:7-29.

Bolivar, VJ, Scott Ganus, J and Messer, A. (2002) The development of behavioral abnormalities in the motor neuron degeneration (mnd) mouse. Brain Res 937:74-82.

Bordier, C. (1981) Phase separation of integral membrane proteins in Triton X-114 solution. J Biol Chem 256:1604-1607.

Braak, H and Goebel, HH. (1978) Loss of pigment-laden stellate cells: a severe alteration of the isocortex in juvenile neuronal ceroid-lipofuscinosis. Acta Neuropathol (Berl) 42:53-57.

Braak, H and Goebel, HH. (1979) Pigmentoarchitectonic pathology of the isocortex in juvenile neuronal ceroid-lipofuscinosis: axonal enlargements in layer IIIab and cell loss in layer V. Acta Neuropathol (Berl) 46:79-83.

Bronson, RT, Lake, BD, Cook, S, Taylor, S and Davisson, MT. (1993) Motor neuron degeneration of mice is a model of neuronal ceroid lipofuscinosis (Batten's disease). Ann Neurol 33:381-385.

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

Brooks, AI, Chattopadhyay, S, Mitchison, HM, Nussbaum, RL and Pearce, DA. (2003) Functional categorization of gene expression changes in the cerebellum of a Cln3-knockout mouse model for Batten disease. Mol Genet Metab 78:17-30.

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

Camp, LA, Verkruyse, LA, Afendis, SJ, Slaughter, CA and Hofmann, SL. (1994) Molecular cloning and expression of palmitoyl-protein thioesterase. J Biol Chem 269:23212-23219.

Chattopadhyay, S and Pearce, DA. (2000) Neural and extraneural expression of the neuronal ceroid lipofuscinoses genes CLN1, CLN2, and CLN3: functional implications for CLN3. Mol Genet Metab 71:207-211.

Chattopadhyay, S, Ito, M, Cooper, JD, Brooks, AI, Curran, TM, Powers, JM and Pearce, DA. (2002) An autoantibody inhibitory to glutamic acid decarboxylase in the neurodegenerative disorder Batten disease. Hum Mol Genet 11:1421-1431.

Cohen, SN, Chang, AC, Boyer, HW and Helling, RB. (1973) Construction of biologically functional bacterial plasmids in vitro. Proc Natl Acad Sci U S A 70:3240-3244.

Collins, FS, Green, ED, Guttmacher, AE and Guyer, MS. (2003) A vision for the future of genomics research. Nature 422:835-847.

Cooper, JD, Messer, A, Feng, AK, Chua-Couzens, J and Mobley, WC. (1999) Apparent loss and hypertrophy of interneurons in a mouse model of neuronal ceroid lipofuscinosis: evidence for partial response to insulin-like growth factor-1 treatment. J Neurosci 19:2556-2567.

Cooper, JD. (2003) Progress towards understanding the neurobiology of Batten disease or neuronal ceroid lipofuscinosis. Curr Opin Neurol 16:121-128.

Cotman, SL, Vrbanac, V, Lebel, LA, Lee, RL, Johnson, KA, Donahue, LR, Teed, AM, Antonellis, K, Bronson, RT, Lerner, TJ and 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.

Darwin, C. (1859) On the origin of species by means of natural selection. John Murray, London.

de la Chapelle, A. (1993) Disease gene mapping in isolated human populations: the example of Finland. J Med Genet 30:857-865.

Dhar, S, Bitting, RL, Rylova, SN, Jansen, PJ, Lockhart, E, Koeberl, DD, Amalfitano, A and Boustany, RM. (2002) Flupirtine blocks apoptosis in batten patient lymphoblasts and in human postmitotic CLN3- and CLN2-deficient neurons. Ann Neurol 51:448-466.

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

Elshatory, Y, Brooks, AI, Chattopadhyay, S, Curran, TM, Gupta, P, Ramalingam, V, Hofmann, SL and Pearce, DA. (2003) Early changes in gene expression in two models of Batten disease. FEBS Lett 538:207-212.

Felbor, U, Kessler, B, Mothes, W, Goebel, HH, Ploegh, HL, Bronson, RT and 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.

Fleischmann, RD, Adams, MD, White, O, Clayton, RA, Kirkness, EF, Kerlavage, AR, Bult, CJ, Tomb, JF, Dougherty, BA, Merrick, JM and et al. (1995) Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. Science 269:496-512.

Franklin, RE and Gosling, RG. (1953) Molecular configuration in sodium thymonucleate. Nature 171:740-741.

Gabriel, SB, Schaffner, SF, Nguyen, H, Moore, JM, Roy, J, Blumenstiel, B, Higgins, J, DeFelice, M, Lochner, A, Faggart, M, Liu-Cordero, SN, Rotimi, C, Adeyemo, A, Cooper, R, Ward, R, Lander, ES, Daly, MJ and Altshuler, D. (2002) The structure of haplotype blocks in the human genome. Science 296:2225-2229.

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 and 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.

Gardiner, M, Sandford, A, Deadman, M, Poulton, J, Cookson, W, Reeders, S, Jokiaho, I, Peltonen, L, Eiberg, H and Julier, C. (1990) Batten disease (Spielmeyer-Vogt disease, juvenile onset neuronal ceroid-lipofuscinosis) gene (CLN3) maps to human chromosome 16. Genomics 8:387-390.

Gardiner, RM. (2000) The molecular genetic basis of the neuronal ceroid lipofuscinoses. Neurol Sci 21:S15-19.

Goebel, HH, Mole, SE and Lake, BD (1999) The Neuronal Ceroid Lipofuscinoses (Batten Disease). IOS Press. Amsterdam.

Goebel, HH. (2000) Morphological aspects of the neuronal ceroid lipofuscinoses. Neurol Sci 21:S27-33.

Golabek, AA, Kaczmarski, W, Kida, E, Kaczmarski, A, Michalewski, MP and Wisniewski, KE. (1999) Expression studies of CLN3 protein (battenin) in fusion with the green fluorescent protein in mammalian cells in vitro. Mol Genet Metab 66:277-282.

Golabek, AA, Kida, E, Walus, M, Wujek, P, Mehta, P and Wisniewski, KE. (2002) Human tripeptidyl-peptidase I: Biosynthesis, glycosylation and enzymatic processing in vivo. J Biol Chem 17:17.

Griffin, JL, Muller, D, Woograsingh, R, Jowatt, V, Hindmarsh, A, Nicholson, JK and Martin, JE. (2002) Vitamin E deficiency and metabolic deficits in neuronal ceroid lipofuscinosis described by bioinformatics. Physiol Genomics 11:195-203.

Guldberg, P, Henriksen, KF, Sipila, I, Guttler, F and de la Chapelle, A. (1995) Phenylketonuria in a low incidence population: molecular characterisation of mutations in Finland. J Med Genet 32:976-978.

Gupta, P, Soyombo, AA, Atashband, A, Wisniewski, KE, Shelton, JM, Richardson, JA, Hammer, RE and 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.

Haltia, M. (1975) Amauroottinen idiotia. Duodecim 91:312-316.

Haskell, RE, Hughes, SM, Chiorini, JA, Alisky, JM and Davidson, BL. (2003) Viral-mediated delivery of the late-infantile neuronal ceroid lipofuscinosis gene, TPP-I to the mouse central nervous system. Gene Ther 10:34-42.

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

Heinonen, O, Salonen, T, Jalanko, A, Peltonen, L and Copp, A. (2000b) CLN-1 and CLN-5, genes for infantile and variant late infantile neuronal ceroid lipofuscinoses, are expressed in the embryonic human brain. J Comp Neurol 426:406-412.

Hellsten, E, Vesa, J, Olkkonen, VM, Jalanko, A and 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.

Herrnstein, RJ and Murray, C (1994) The Bell Curve: The Reshaping of American Life by Differences in Intelligence. Free Press. New York.

Hirokawa, T, Boon-Chieng, S and Mitaku, S. (1998) SOSUI: classification and secondary structure prediction system for membrane proteins. Bioinformatics 14:378-379.

Holopainen, JM, Saarikoski, J, Kinnunen, PK and Jarvela, I. (2001) Elevated lysosomal pH in neuronal ceroid lipofuscinoses (NCLs). Eur J Biochem 268:5851-5856.

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

International Batten Disease Consortium. (1995) Isolation of a novel gene underlying Batten disease, CLN3. Cell 82:949-957.Isosomppi, J, Heinonen, O, Hiltunen, JO, Greene, ND, Vesa, J,

Uusitalo, A, Mitchison, HM, Saarma, M, Jalanko, A and Peltonen, L. (1999) Developmental expression of palmitoyl protein thioesterase in normal mice. Brain Res Dev Brain Res 118:1-11.

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

Jackson, DA, Symons, RH and Berg, P. (1972) Biochemical method for inserting new genetic information into DNA of Simian Virus 40: circular SV40 DNA molecules containing lambda phage genes and the galactose operon of Escherichia coli. Proc Natl Acad Sci U S A 69:2904-2909.

Jallinoja, P (2002) Genetics, negotiated ethics and the ambiguities of moral choices. National Public Health Institute. Helsinki.

Jansky, J. (1908) Dosud neposany pripad familiarni amauroticke idiotie komplikovane s hypoplasii mozeckovou. Sb Lek 13:165-196.

Jolly, RD and Walkley, SU. (1999) Ovine ceroid lipofuscinosis (OCL6): postulated mechanism of neurodegeneration. Mol Genet Metab 66:376-380.

Jutikkala, E and Pirinen, K (1996) A History of Finland. WSOY. Juva.

Järvelä, I. (1991) Infantile neuronal ceroid lipofuscinosis (CLN1): linkage disequilibrium in the Finnish population and evidence that variant late infantile form (variant CLN2) represents a nonallelic locus. Genomics 10:333-337.

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

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

Katz, ML, Gao, CL, Prabhakaram, M, Shibuya, H, Liu, PC and Johnson, GS. (1997) Immunochemical localization of the Batten disease (CLN3) protein in retina. Invest Ophthalmol Vis Sci 38:2375-2386.

Katz, ML, Liu, PC, Grob-Nunn, SE, Shibuya, H and Johnson, GS. (1999a) Characterization and chromosomal mapping of a mouse ortholog of the late-infantile ceroid-lipofuscinosis gene CLN2. Mamm Genome 10:1050-1053.

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

Kirveskari, E, Partinen, M and Santavuori, P. (2001) Sleep and its disturbance in a variant form of late infantile neuronal ceroid lipofuscinosis (CLN5). J Child Neurol 16:707-713.

Kittles, RA, Perola, M, Peltonen, L, Bergen, AW, Aragon, RA, Virkkunen, M, Linnoila, M, Goldman, D and Long, JC. (1998) Dual origins of Finns revealed by Y chromosome haplotype variation. Am J Hum Genet 62:1171-1179.

Klockars, T, Savukoski, M, Isosomppi, J, Laan, M, Järvelä, I, Petrukhin, K, Palotie, A and 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.

Klockars, T, Holmberg, V, Savukoski, M, Lander, ES and Peltonen, L. (1999) Transcript identification on the CLN5 region on chromosome 13q22. Human Genetics 105:51-56.

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 and Uchiyama, Y. (2000) Cathepsin D deficiency induces lysosomal storage with ceroid lipofuscin in mouse CNS neurons. J Neurosci 20:6898-6906.

Kremmidiotis, G, Lensink, IL, Bilton, RL, Woollatt, E, Chataway, TK, Sutherland, GR and Callen, DF. (1999) The Batten disease gene product (CLN3p) is a Golgi integral membrane protein. Hum Mol Genet 8:523-531.

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

Laitinen, T, Daly, MJ, Rioux, JD, Kauppi, P, Laprise, C, Petays, T, Green, T, Cargill, M, Haahtela, T, Lander, ES, Laitinen, LA, Hudson, TJ and Kere, J. (2001) A susceptibility locus for asthma-related traits on chromosome 7 revealed by genome-wide scan in a founder population. Nat Genet 28:87-91.

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, et al. (2001) Initial sequencing and analysis of the human genome. Nature 409:860-921.

Lauronen, L, Huttunen, J, Kirveskari, E, Wikstrom, H, Sainio, K, Autti, T and Santavuori, P. (2002) Enlarged SI and SII somatosensory evoked responses in the CLN5 form of neuronal ceroid lipofuscinosis. Clin Neurophysiol 113:1491-1500.

Lehtovirta, M, Kyttälä, A, Eskelinen, EL, Hess, M, Heinonen, O and 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.

Liu, CG, Sleat, DE, Donnelly, RJ and Lobel, P. (1998) Structural organization and sequence of CLN2, the defective gene in classical late infantile neuronal ceroid lipofuscinosis. Genomics 50:206-212.

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

Luiro, K, Kopra, O, Lehtovirta, M and 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.

Lönnqvist, T, Vanhanen, SL, Vettenranta, K, Autti, T, Rapola, J, Santavuori, P and Saarinen-Pihkala, UM. (2001) Hematopoietic stem cell transplantation in infantile neuronal ceroid lipofuscinosis. Neurology 57:1411-1416.

March, PA, Wurzelmann, S and Walkley, SU. (1995) Morphological alterations in neocortical and cerebellar GABAergic neurons in a canine model of juvenile Batten disease. Am J Med Genet 57:204-212.

Margraf, LR, Boriack, RL, Routheut, AA, Cuppen, I, Alhilali, L, Bennett, CJ and Bennett, MJ. (1999) Tissue expression and subcellular localization of CLN3, the Batten disease protein. Mol Genet Metab 66:283-289.

Maxam, AM and Gilbert, W. (1977) A new method for sequencing DNA. Proc Natl Acad Sci U S A 74:560-564.

Mendel, G. (1866) Versuche über Pflanzen-Hybriden. Verhandlungen des naturforschenden Vereines in Brünn 4:3-47.

Mitchison, HM, Hofmann, SL, Becerra, CH, Munroe, PB, Lake, BD, Crow, YJ, Stephenson, JB, Williams, RE, Hofman, IL, Taschner, PEM, Martin, JJ, Philippart, M, Andermann, E, Andermann, F, Mole, SE, Gardiner, RM and O'Rawe, AM. (1998) Mutations in the palmitoyl-protein thioesterase gene (PPT; CLN1) causing juvenile neuronal ceroid lipofuscinosis with granular osmiophilic deposits [published erratum appears in Hum Mol Genet 1998 Apr;7(4):765]. Human Molecular Genetics 7:291-297.

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 and 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.

Modrek, B and Lee, CJ. (2003) Alternative splicing in the human, mouse and rat genomes is associated with an increased frequency of exon creation and/or loss. Nat Genet 34:177-180.

Mole, SE, Mitchison, HM and Munroe, PB. (1999) Molecular basis of the neuronal ceroid lipofuscinoses: Mutations in CLN1, CLN2, CLN3, and CLN5. Human Mutation 14:199-215.

Mole, SE, Zhong, NA, Sarpong, A, Logan, WP, Hofmann, S, Yi, W, Franken, PF, van Diggelen, OP, Breuning, MH, Moroziewicz, D, Ju, W, Salonen, T, Holmberg, V, Jarvelä, I and Taschner, PE. (2001b) New mutations in the neuronal ceroid lipofuscinosis genes. Eur J Paediatr Neurol 5:7-10.

Mole, SE, Wheeler, RB, Sharp, JD and Cutler, D. (2003) Intracellular localisation of CLN6. Abstract at 9th International Congress on Neuronal Ceroid Lipofuscinoses, Chicago.

Moller, S, Croning, MD and Apweiler, R. (2001) Evaluation of methods for the prediction of membrane spanning regions. Bioinformatics 17:646-653.

National Research Council (1996) Guide for the Care and Use of Laboratory Animals. Washington D.C.

Nijssen, PC, Ceuterick, C, van Diggelen, OP, Elleder, M, Martin, JJ, Teepen, JL, Tyynela, J and 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.

Nirenberg, MW. (1963) The genetic code: II. Sci Am 208:80-94.

Norio, R. (2003a) Finnish Disease Heritage I: characteristics, causes, background. Hum Genet 112:441-456.

Norio, R. (2003b) Finnish Disease Heritage II: population prehistory and genetic roots of Finns. Hum Genet 112:457-469.

Norio, R. (2003c) The Finnish Disease Heritage III: the individual diseases. Hum Genet 112:470-526.

Pastinen, T, Perola, M, Ignatius, J, Sabatti, C, Tainola, P, Levander, M, Syvanen, AC and Peltonen, L. (2001) Dissecting a population genome for targeted screening of disease mutations. Hum Mol Genet 10:2961-2972.

Pearce, DA, Ferea, T, Nosel, SA, Das, B and Sherman, F. (1999) Action of BTN1, the yeast orthologue of the gene mutated in Batten disease. Nat Genet 22:55-58.

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

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

Peltonen, L and McKusick, VA. (2001) Genomics and medicine. Dissecting human disease in the postgenomic era. Science 291:1224-1229.

Peri, S and Pandey, A. (2001) A reassessment of the translation initiation codon in vertebrates. Trends Genet 17:685-687.

Persaud-Sawin, DA, VanDongen, A and Boustany, RM. (2002) Motifs within the CLN3 protein: modulation of cell growth rates and apoptosis. Hum Mol Genet 11:2129-2142.

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

Ranta, S, Topcu, M, Mole, SE and Lehesjoki, AE. (2003) Variant late infantile NCL in a subset of Turkish patients is allelic to Northern epilepsy. Abstract at 9th International Congress on Neuronal Ceroid Lipofuscinoses, Chicago.

Rapola, J. (1993) Neuronal ceroid-lipofuscinoses in childhood. Perspect Pediatr Pathol 17:7-44.

Rapola, J, Lähdetie, J, Isosomppi, J, Helminen, P, Penttinen, M and Järvelä, I. (1999) Prenatal diagnosis of variant late infantile neuronal ceroid lipofuscinosis (vLINCL[Finnish]; CLN5). Prenatal Diagnosis 19:685-688.

Regan, T (1983) The Case for Animal Rights. University of California Press. Berkeley and Los Angeles.
Rider, JA and Rider, DL. (1988) Batten disease: past, present, and future. Am J Med Genet Suppl 5:21-26.
Rouille, Y, Rohn, W and Hoflack, B. (2000) Targeting of lysosomal proteins. Semin Cell Dev Biol 11:165-171.
Rylova, SN, Amalfitano, A, Persaud-Sawin, DA, Guo, WX, Chang, J, Jansen, PJ, Proia, AD and Boustany, RM. (2002) The CLN3 gene is a novel molecular target for cancer drug discovery. Cancer Res 62:801-808.
Sachs, B. (1896) A family form of idiocy, generally fatal, and associated with early blindness (amaurotic family idiocy). NY Med J 63:697-703.
Saftig, P, Hetman, M, Schmahl, W, Weber, K, Heine, L, Mossmann, H, Koster, A, Hess, B, Evers, M, von Figura, K and 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.
Saiki, RK, Scharf, S, Faloona, F, Mullis, KB, Horn, GT, Erlich, HA and Arnheim, N. (1985) Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science 230:1350-1354.
Saiki, RK, Gelfand, DH, Stoffel, S, Scharf, SJ, Higuchi, R, Horn, GT, Mullis, KB and Erlich, HA. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487-491.
Sajantila, A and Pääbo, S. (1995) Language replacement in Scandinavia. Nat Genet 11:359-360.
Sajantila, A, Salem, AH, Savolainen, P, Bauer, K, Gierig, C and Paabo, S. (1996) Paternal and maternal DNA lineages reveal a bottleneck in the founding of the Finnish population. Proc Natl Acad Sci U S A 93:12035-12039.
Salonen, T, Hellsten, E, Horelli-Kuitunen, N, Peltonen, L and Jalanko, A. (1998) Mouse palmitoyl protein thioesterase: gene structure and expression of cDNA. Genome Res 8:724-730.
Sanger, F and Coulson, AR. (1975) A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J Mol Biol 94:441-448.
Santavuori, P, Rapola, J, Sainio, K and Raitta, C. (1982) A variant of Jansky-Bielschowsky disease. Neuropediatrics 13:135-141.
Santavuori, P, Westermarck, T, Rapola, J, Pohja, P, Moren, R, Lappi, M and Vuonnala, U. (1985) Antioxidant treatment in Spielmeyer-Sjögren's disease. Acta Neurol Scand 71:136-145.
Santavuori, P. (1988) Neuronal ceroid-lipofuscinoses in childhood. Brain & Development 10:80-83.
Santavuori, P, Heiskala, H, Westermarck, T, Sainio, K and Moren, R. (1988) Experience over 17 years with antioxidant treatment in Spielmeyer-Sjögren disease. Am J Med Genet Suppl 5:265-274.
Santavuori, P, Rapola, J, Nuutila, A, Raininko, R, Lappi, M, Launes, J, Herva, R and Sainio, K. (1991) The spectrum of Jansky-Bielschowsky disease. Neuropediatrics 22:92-96.
Santavuori, P, Lauronen, L, Kirveskari, E, Aberg, L, Sainio, K and Autti, T. (2000) Neuronal ceroid lipofuscinoses in childhood. Neurol Sci 21:S35-41.
Sanyal, S and Ramaswami, M. (2002) Spinsters, synaptic defects, and amaurotic idiocy. Neuron 36:335-338.
Savukoski, M, Kestilä, M, Williams, R, Järvelä, I, Sharp, J, Harris, J, Santavuori, P, Gardiner, M and Peltonen, L. (1994) Defined chromosomal assignment of CLN5 demonstrates that at least four genetic loci are involved in the pathogenesis of human ceroid lipofuscinoses. American Journal of Human Genetics 55:695-701.
Savukoski, M, Klockars, T, Holmberg, V, Santavuori, P, Lander, ES and Peltonen, L. (1998) CLN5, a novel gene encoding a putative transmembrane protein mutated in Finnish variant late infantile neuronal ceroid lipofuscinosis. Nature Genetics 19:286-288.
Seigel, GM, Lotery, A, Kummer, A, Bernard, DJ, Greene, ND, Turmaine, M, Derksen, T, Nussbaum, RL, Davidson, B, Wagner, J and Mitchison, HM. (2002) Retinal pathology and function in a Cln3

knockout mouse model of juvenile Neuronal Ceroid Lipofuscinosis (batten disease). Mol Cell Neurosci 19:515-527.

Sharp, JD, Wheeler, RB, Lake, BD, Savukoski, M, Jarvela, IE, Peltonen, L, Gardiner, RM and 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, Lake, BD, Fox, M, Gardiner, RM and Williams, RE. (1999) Genetic and physical mapping of the CLN6 gene on chromosome 15q21-23. Mol Genet Metab 66:329-331.

Sharp, JD, Wheeler, RB, Schultz, RA, Joslin, JM, Mole, SE, Williams, RE and Gardiner, RM. (2001) Analysis of candidate genes in the CLN6 critical region using in silico cloning. Eur J Paediatr Neurol 5:29-31.

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

Singer, P (1975) Animal Liberation: A New Ethics for our Treatment of Animals. New York Review/Random House. New York.

Sleat, DE, Donnelly, RJ, Lackland, H, Liu, CG, Sohar, I, Pullarkat, RK and 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 and Lobel, P. (1999) Mutational analysis of the defective protease in classic late-infantile neuronal ceroid lipofuscinosis, a neurodegenerative lysosomal storage disorder. American Journal of Human Genetics 64:1511-1523.

Sleat, DE, Wiseman, JA and Lobel, P. (2003) A mouse model for classical late-infantile neuronal ceroid lipofuscinosis (LINCL). Abstract at 9th International Congress on Neuronal Ceroid Lipofuscinoses, Chicago.

Smith, LM, Sanders, JZ, Kaiser, RJ, Hughes, P, Dodd, C, Connell, CR, Heiner, C, Kent, SB and Hood, LE. (1986) Fluorescence detection in automated DNA sequence analysis. Nature 321:674-679.

Sohar, I, Sleat, DE, Jadot, M and Lobel, P. (1999) Biochemical characterization of a lysosomal protease deficient in classical late infantile neuronal ceroid lipofuscinosis (LINCL) and development of an enzyme-based assay for diagnosis and exclusion of LINCL in human specimens and animal models. J Neurochem 73:700-711.

Soyombo, AA and Hofmann, SL. (1997) Molecular cloning and expression of palmitoyl-protein thioesterase 2 (PPT2), a homolog of lysosomal palmitoyl-protein thioesterase with a distinct substrate specificity. J Biol Chem 272:27456-27463.

Spielmeyer, W. (1905) Über familiäre amaurotische Idiotiee. Neurol Zentralbl 24:620-621.

Steinfeld, R, Heim, P, von Gregory, H, Meyer, K, Ullrich, K, Goebel, HH and 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.

Suopanki, J, Tyynelä, J, Baumann, M and Haltia, M. (1999a) Palmitoyl-protein thioesterase, an enzyme implicated in neurodegeneration, is localized in neurons and is developmentally regulated in rat brain. Neurosci Lett 265:53-56.

Suopanki, J, Tyynelä, J, Baumann, M and Haltia, M. (1999b) The expression of palmitoyl-protein thioesterase is developmentally regulated in neural tissues but not in nonneural tissues. Mol Genet Metab 66:290-293.

Suopanki, J, Lintunen, M, Lahtinen, H, Haltia, M, Panula, P, Baumann, M and Tyynelä, J. (2002) Status epilepticus induces changes in the expression and localization of endogenous palmitoyl-protein thioesterase 1. Neurobiol Dis 10:247-257.

Sweeney, ST and Davis, GW. (2002) Unrestricted synaptic growth in spinster-a late endosomal protein implicated in TGF-beta-mediated synaptic growth regulation. Neuron 36:403-416.

Svennerholm, L. (1962) The chemical structure of normal human brain and Tay Sachs gangliosidosis. Biochem Biophys Res Commun 9:436-441.

Syvänen, AC, Aalto-Setala, K, Kontula, K and Soderlund, H. (1989) Direct sequencing of affinity-captured amplified human DNA application to the detection of apolipoprotein E polymorphism. FEBS Lett 258:71-74.

Tahvanainen, E, Ranta, S, Hirvasniemi, A, Karila, E, Leisti, J, Sistonen, P, Weissenbach, J, Lehesjoki, AE and 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.

The C. elegans Sequencing Consortium. (1998) Genome sequence of the nematode C. elegans: a platform for investigating biology. Science 282:2012-2018.

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

Tyynelä, J, Suopanki, J, Santavuori, P, Baumann, M and Haltia, M. (1997) Variant late infantile neuronal ceroid-lipofuscinosis: pathology and biochemistry. Journal of Neuropathology & Experimental Neurology 56:369-375.

Tyynelä, J, Sohar, I, Sleat, DE, Gin, RM, Donnelly, RJ, Baumann, M, Haltia, M and 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 and Suopanki, J. (2000) Biochemical aspects of neuronal ceroid lipofuscinoses. Neurol Sci 21:S21-25.

Walkley, SU and March, PA. (1993) Biology of neuronal dysfunction in storage disorders. J Inherit Metab Dis 16:284-287.

Vandenplas, S, Wiid, I, Grobler-Rabie, A, Brebner, K, Ricketts, M, Wallis, G, Bester, A, Boyd, C and Mathew, C. (1984) Blot hybridisation analysis of genomic DNA. J Med Genet 21:164-172.

Varilo, T, Savukoski, M, Norio, R, Santavuori, P, Peltonen, L and Järvelä, I. (1996) The age of human mutation: genealogical and linkage disequilibrium analysis of the CLN5 mutation in the Finnish population. American Journal of Human Genetics 58:506-512.

Varilo, T, Laan, M, Hovatta, I, Wiebe, V, Terwilliger, JD and Peltonen, L. (2000) Linkage disequilibrium in isolated populations: Finland and a young sub-population of Kuusamo. Eur J Hum Genet 8:604-612.

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, et al. (2001) The sequence of the human genome. Science 291:1304-1351.

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

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

Vesa, J, Chin, MH, Oelgeschlager, K, Isosomppi, J, DellAngelica, EC, Jalanko, A and 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.

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

Vogt, H. (1905) Über familiäre amaurotische und verwandte Krankheitsbilder. Psychiatr Neurol 18:161-171.

Waterston, RH, Lindblad-Toh, K, Birney, E, Rogers, J, Abril, JF, Agarwal, P, Agarwala, R, Ainscough, R, Alexandersson, M, An, P, Antonarakis, SE, Attwood, J, Baertsch, R, Bailey, J, Barlow, K, Beck, S, Berry, E, Birren, B, Bloom, T, Bork, P, Botcherby, M, Bray, N, Brent, MR, Brown, DG, Brown, SD, Bult, C, Burton, J, Butler, J, Campbell, RD, Carninci, P, et al. (2002) Initial sequencing and comparative analysis of the mouse genome. Nature 420:520-562.

Watson, JD and Crick, FHC. (1953) A structure for deoxyribose nucleic acid. Nature 171:737-738.

Weissenbach, J, Gyapay, G, Dib, C, Vignal, A, Morissette, J, Millasseau, P, Vaysseix, G and Lathrop, M. (1992) A second-generation linkage map of the human genome. Nature 359:794-801.

Wheeler, RB, Sharp, JD, Schultz, RA, Joslin, JM, Williams, RE and 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, R, Santavuori, P, Peltonen, L, Gardiner, RM and Järvelä, I. (1994) A variant form of late infantile neuronal ceroid lipofuscinosis (CLN5) is not an allelic form of Batten (Spielmeyer-Vogt-Sjogren, CLN3) disease: exclusion of linkage to the CLN3 region of chromosome 16. Genomics 20:289-290.

Williams, RS, Lott, IT, Ferrante, RJ and Caviness, VS, Jr. (1977) The cellular pathology of neuronal ceroid-lipofuscinosis. A golgi- electronmicroscopic study. Arch Neurol 34:298-305.

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

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

Zeman, W and Donahue, S. (1963) Fine structure of the lipid bodies in juvenile amaurotic idiocy. Acta Neuropathol 3:144-149.

Zeman, W and Dyken, P. (1969) Neuronal ceroid-lipofuscinosis (Batten's disease). Pediatrics 44:570-583.