Am. J. Hum. Genet. 64:1511–1523, 1999

Mutational Analysis of the Defective Protease in Classic Late-Infantile Neuronal Ceroid Lipofuscinosis, a Neurodegenerative Lysosomal Storage Disorder

David E. Sleat,[1] Rosalie M. Gin,[1] Istvan Sohar,[1] Krystyna Wisniewski,[3,6] Susan Sklower-Brooks,[4] Raju K. Pullarkat,[5] David N. Palmer,[7] Terry J. Lerner,[8,9] Rose-Mary Boustany,[10] Peter Uldall,[11] Aristotle N. Siakotos,[12] Robert J. Donnelly,[2] and Peter Lobel[1]

[1]Center for Advanced Biotechnology and Medicine and Department of Pharmacology, University of Medicine and Dentistry of New Jersey, Piscataway, and [2]Department of Pathology and Lab Medicine, New Jersey Medical School, Molecular Biocore Facility, Newark; Departments of [3]Pathological Neurobiology, [4]Human Genetics, and [5]Neurochemistry, New York State Institute for Basic Research in Developmental Disabilities, Staten Island; [6]State University of New York Health Science Center at Brooklyn, Brooklyn; [7]Animal and Food Sciences Division, Lincoln University, Canterbury, New Zealand; [8]Molecular Neurogenetics Unit, Massachusetts General Hospital, Charlestown, Massachusetts; [9]Department of Neurology, Harvard Medical School, Boston; [10]Duke University Medical Center, Department of Pediatrics and Neurobiology, Durham, North Carolina; [11]Juliane Marie Centret, Neuropediatric Clinic, Copenhagen; and [12]Department of Pathology, Indiana University School of Medicine, Indianapolis

Summary

The late-infantile form of neuronal ceroid lipofuscinosis (LINCL) is a progressive and ultimately fatal neurodegenerative disease of childhood. The defective gene in this hereditary disorder, *CLN2*, encodes a recently identified lysosomal pepstatin-insensitive acid protease. To better understand the molecular pathology of LINCL, we conducted a genetic survey of *CLN2* in 74 LINCL families. In 14 patients, CLN2 protease activities were normal and no mutations were identified, suggesting other forms of NCL. Both pathogenic alleles were identified in 57 of the other 60 LINCL families studied. In total, 24 mutations were associated with LINCL, comprising six splice-junction mutations, 11 missense mutations, 3 nonsense mutations, 3 small deletions, and 1 single-nucleotide insertion. Two mutations were particularly common: an intronic G→C transversion in the invariant AG of a 3′ splice junction, found in 38 of 115 alleles, and a C→T transition in 32 of 115 alleles, which prematurely terminates translation at amino acid 208 of 563. An Arg→His substitution was identified, which was associated with a late age at onset and protracted clinical phenotype, in a number of other patients originally diagnosed with juvenile NCL.

Received February 5, 1999; accepted for publication April 15, 1999; electronically published May 12, 1999.

Address for correspondence and reprints: Drs. Peter Lobel and David Sleat, Center for Advanced Biotechnology and Medicine, Piscataway, NJ 08854. E-mail: Lobel@waksman.rutgers.edu or Sleat@waksman.rutgers.edu

0002-9297/99/6406-0005$02.00

Introduction

The neuronal ceroid lipofuscinoses (NCLs) are a group of related hereditary neurodegenerative diseases that primarily affect children (Boustany 1996). All childhood forms are fatal and have similar clinical profiles, including epileptic seizures, loss of vision, ataxia and loss of motor coordination, and deterioration of mental capabilities. At the ultrastructural level, NCLs are characterized by the accumulation of an autofluorescent material, termed "ceroid-lipofuscin," in lysosome-like storage bodies in most cell types of patients.

Several NCLs have been classified, with distinct genetic bases. Infantile NCL (INCL; MIM 256730) is caused by defects in the gene *CLN1* (Vesa et al. 1995), which encodes a lysosomal enzyme, palmitoyl protein thioesterase (Camp and Hofmann 1993). The most common NCLs to occur in the United States are the juvenile (JNCL; MIM 204200) and classic late-infantile (LINCL; MIM 204500) forms. In LINCL, the defective gene, *CLN2*, encodes a recently identified lysosomal protease (Sleat et al. 1997). In JNCL, the defective gene, *CLN3* (The International Batten Disease Consortium 1995), is a transmembrane protein that is probably lysosomal (Jarvela et al. 1998). The normal function of the human CLN3 protein is not known, although there is evidence that a yeast homologue is involved in vacuolar pH homeostasis (Pearce and Sherman 1998). The defective gene, *CLN4*, in a rare adult form of NCL (MIM 204300) remains to be identified. There are also at least two variant forms of LINCL. In the Finnish-variant LINCL (MIM 256731), the defective gene, *CLN5*, has been identified (Savukoski et al. 1998) and is a putative transmembrane protein of currently unknown function. In another variant, the defective gene, *CLN6*, (MIM

Table 1

CLN2 Mutations Associated with LINCL

Family	Source (Identifier)[a]	Sample Type	Ethnicity	Age at Onset (years)	Lifespan (years)	EM	Presenting Symptom	CLN2 Protease[b]	Allele 1	Allele 2
1	1 (IGF032/85)	FB	Italian × Italian	3.0		CV	ES	Deficient	1917C→T (Gln66Stop)	1917C→T (Gln66Stop)
2	2 (GUS15820)	LB	Spanish-French-English-Irish-American Indian	3.0	11.5	CV	ES	Deficient	1946A→G (SJ)	3556G→C (SJ)
3	3 (KEW1)	Brain	Irish-American Indian-German	3.0	10.2	CV; G	ES; VD; CD	Deficient	1950G→A (Gly77Arg)	3556G→C (SJ)
4	4 (CABM006)	DNA, blood	Romany × Romany	2.2	>4.8	FP	ES	Deficient	3004A→G (Gln100Arg)/ 5171G→A (Gly389Glu)	3004A→G (Gln100Arg)/ 5171G→A (Gly389Glu)
5	1 (IGF087/79)	FB	Romany × Romany	3.0		CV	ES	Deficient	3004A→G (Gln100Arg)/ 5171G→A (Gly389Glu)	3004A→G (Gln100Arg)/ 5171G→A (Gly389Glu)
6	5 (BD00224)	LB	German-Bohemian × German-Bohemian	3.0	7.8	CV	CD	Deficient	3081-3091del	3670C→T (Arg208Stop)
7	13 (2740)	Brain						Deficient	3081-3091del	3670C→T (Arg208Stop)
8	6 (CABM011)	FB	Mexican × Mexican	3.0	>6.0	CV	ES	Deficient	3084C→T (Arg127Stop)	3670C→T (Arg208Stop)
9	3 (C10516); 2 (GUS-24180)	DNA LB	English-Scottish × English-Scottish	3.5	>8.7	CV; FP	ES; CD	Deficient	3556G→A (SJ)	3556G→C (SJ)
10	7 (DNP1)	Brain	English × Dutch-English	3.0	7.8	CV	ES	Deficient	3556G→A (SJ)	3556G→C (SJ)
11	2 (GUS26235); 3 (C11488)	LB DNA	German × English	3.5	>6.5	CV	VD	Deficient	3556G→A (SJ)	5271G→C (Gln422His)
12	5 (BD00173)	LB	German × German-English	3.0	8.5	CV	ES	Deficient	3556G→C (SJ)	3556G→C (SJ)
13	3 (C10557)	DNA	English × English-Irish	3.0	>8.5	CV	ES; CD	ND	3556G→C (SJ)	3556G→C (SJ)
14	3 (C8878); 5(BD00379)	DNA LB	Swedish-English-Irish × Scottish-Irish-Norwegian	2.5	>10.0	CV	ES	ND	3556G→C (SJ)	3556G→C (SJ)
15	8 (CABM001)	DNA	German × German	3.5	>4.5	CV	ES	ND	3556G→C (SJ)	3556G→C (SJ)
16	9 (CABM004)	Blood	Scottish-Irish × Scottish-Irish	3.0	>3.0	CV	ES	Deficient	3556G→C (SJ)	3556G→C (SJ)
17	9 (CABM009)	Blood	English-Irish × English-Swedish	3.3	>12.0	CV		Deficient	3556G→C (SJ)	3556G→C (SJ)
18	9 (CABM017)	Blood	English-Irish × English-German-Irish-Indian	3.5	>5.5	CV	ES	Deficient	3556G→C (SJ)	3556G→C (SJ)
19	5 (BD00219)	LB	English × English	3.0	10	CV	ES; CD	Deficient	3556G→C (SJ)	3670C→T (Arg208Stop)
20	5 (BD00431)	LB	Canadian × Canadian	2.0	>8.7	CV	ES	Deficient	3556G→C (SJ)	3670C→T (Arg208Stop)
21	3 (C7153); 5 (BD00020)	DNA LB	English × German	2.0	7.3	CV	ES	Deficient	3556G→C (SJ)	3670C→T (Arg208Stop)
22	3 (C9542); 5 (BD00056)	LB	German × Irish-Scottish	2.0	5.8	CV	ES	Deficient	3556G→C (SJ)	3670C→T (Arg208Stop)
23	9 (CABM0015)	Blood	German-Czech × Canadian-North American Indian	3.0	>4.8	CV	ES	Deficient	3556G→C (SJ)	3670C→T (Arg208Stop)
24	10 (CABM002)	DNA	Danish × Danish	2.9	>6.3	CV	ES	ND	3556G→C (SJ)	3670C→T (Arg208Stop)
25	11 (CABM003)	FB	northern European × northern European	3.8	>5.4	CV	ES	Deficient	3556G→C (SJ)	3670C→T (Arg208Stop)
26	12 (CABM005)	FB	Scottish-English × Scottish-English	2.3	>7.7	CV	ES	Deficient	3556G→C (SJ)	3670C→T (Arg208Stop)
27	9 (CABM016)	Blood	Scottish-Irish-English × Scottish-Irish-English	3.0	9.7	CV	ES	Deficient	3556G→C (SJ)	3670C→T (Arg208Stop)
28	2 (GUS18769)	LB	English-Irish-American Indian	3.0	>9.0	CV	ES; LMC	Deficient	3556G→C (SJ)	3670C→T (Arg208Stop)

29	2 (GUS26230)	LB	French-Irish-German					Deficient	3556G→C (SJ)	3670C→T (Arg208Stop)
30	3 (C7788)	LB	Irish-Scottish-English × Irish-Scottish-English	3.0	9.7	CV	ES; LMC	Deficient	3556G→C (SJ)	3670C→T (Arg208Stop)
31	5 (BD00397); 13(2577)	LB Brain	Scottish × English-Scottish	2.0	12.8	CV	CD	Deficient	3556G→C (SJ)	4188A→G (SJ)
32	3 (C9801)	DNA	French × Italian	3.0	>12	CV	ES	ND	3556G→C (SJ)	4288-4295del
33	9 (CABM007/8)[c]	Blood	Finnish-Swedish-Norwegian × English-Scottish-Norwegian	3.0	16		ES	ND	3556G→C (SJ)	4346G→A (Glu343Lys)
34	9 (CABM013/14)[c]	Blood						ND	3556G→C (SJ)	4655G→A (Cys365Tyr)
35	13 (1932)	Brain						Deficient	3556G→C (SJ)	5271G→C (Gln422His)
36	10 (CABM010)	DNA	Danish × Danish	2.8	>4.3	CV	ES	Deficient	3556G→C (SJ)	5271G→C (Gln422His)
37	3 (KEW2); 5 (BD00190)	Brain LB	French-German-Irish-Scottish-English	7.0	30.0	CV; FP	CD; LMC	Deficient	3556G→C (SJ)	5457G→A (Arg447His)
38	3 (F624/C6824)	DNA	Czech-Ukranian × Welsh	3.0	9.5	CV; FP	ES	ND	3556G→C (SJ)	5541C→T (Ser475Leu)
39	5 (BD00134)	LB	English	2.5	8.3	CV	ES; CD	Deficient	3670C→T (Arg208Stop)	3670C→T (Arg208Stop)
40	5 (BD00463)	LB	Polish × Polish-Swedish-French-Bohemian	2.9	7.7	CV	ES	Deficient	3670C→T (Arg208Stop)	3670C→T (Arg208Stop)
41	3 (C7149); 5 (BD00047)	DNA LB	Polish-German × Norwegian-German	3.0	>13.4	CV	ES	ND	3670C→T (Arg208Stop)	3670C→T (Arg208Stop)
42	3 (C8149)	DNA	Polish × Polish	3.0	12	CV; FP	ES; LMC	ND	3670C→T (Arg208Stop)	3670C→T (Arg208Stop)
43	9 (CABM012)	Blood	Norwegian-German-Polish	3.0	>5.5		ES	Deficient	3670C→T (Arg208Stop)	3670C→T (Arg208Stop)
44	3 (C11496)	DNA	Italian × Italian	2.0	>8.0	CV	CD	ND	3670C→T (Arg208Stop)	4396T→G (SJ)
45	5 (BD00385)	LB	English-German × Hungarian-Polish-Russian	3.0	11.8	CV; FP	ES	Deficient	3670C→T (Arg208Stop)	5159T→A (Val385Asp)
46	5 (BD00271)	LB	Polish × German	3.0	6.3	CV	CD	Deficient	3670C→T (Arg208Stop)	5271G→C (Gln422His)
47	3 (C11571)	DNA	Mexican × Mexican	3.0	>8.5	CV	ES	ND	3670C→T (Arg208Stop)	5271G→C (Gln422His)
48	3 (KEW3)	Brain	English-Irish	2.5	6	CV	ES	Deficient	3670C→T (Arg208Stop)	5271G→C (Gln422His)
49	14 (ANS1)	Brain	German-Irish-Polish-English	8.0	43	CV; FP		Deficient	3670C→T (Arg208Stop)	5457G→A (Arg447His)
50	2 (GUS26252)	LB						Deficient	3670C→T (Arg208Stop)	6025G→C (SJ)
51	15 (GM09404)	FB						Deficient	4023T→A (Ile287Asn)	Not identified
52	1 (IGF056/87)	FB	Italian × Italian	3.0		CV	ES	Deficient	4288-4295del	4288-4295del
53	2 (GUS23717)	LB	English × English	3.5	>7.3	CV	ES	Deficient	4346G→A (Glu343Lys)	4346G→A (Glu343Lys)
54	16 (WG308)	FB						Deficient	4654T→C (Cys365Arg)	6069InsA
55	16 (WG305)	FB						Deficient	4655G→A (Cys365Tyr)	4655G→A (Cys365Tyr)
56	7 (DNP2)	Brain						Deficient	4655G→A (Cys365Tyr)	5478C→A (Ala454Glu)
57	3 (C9538); 5 (BD00013)	DNA LB	Spanish × Spanish	2.5	10.3	CV	ES	Deficient	5271G→C (Gln422His)	5271G→C (Gln422His)
58	5 (BD00284)	LB	Romanian-Canadian × Romanian-Canadian	3.0	8	CV	VD; CD	Deficient	6151-6152delTC	6151-6152delTC
59	3 (KEW4)	Brain	Hispanic	0.8	8	CV; G	ES; CD	Deficient	Not identified	Not identified
60	2 (GUS16766)	LB	Middle Eastern					Deficient	Not identified	Not identified

NOTE.—LB = lymphoblasts; FB = fibroblasts; ES = epileptic seizures; VD = vision deterioration; CD = cognitive dysfunction; LMC = loss of motor coordination; CV = curvilinear bodies; FP = fingerprint bodies; SJ = splice-junction mutations; G = granular osmiophilic deposits; and nd = not determined.

[a] Sources: 1 =Instituto Gaslini; 2 = T. J. Lerner; 3 = Institute for Basic Research, Staten Island; 4= S. Goldwurm; 5 = Batten Disease Lymphocyte Cell Bank at the University of Indiana School of Medicine; 6 = F. Hisama; 7 = D. N. Palmer; 8 = P. Krebs; 9 = families; 10 = P. Uldall; 11 = E. Berry-Krais; 12 =C. Greenberg and L. Dilling; 13 = National Neurological Research Specimen Bank; 14 = A. N. Siakotos; 15 = Coriell Cell Collection; and 16 = McGill University Repository.

[b] Deficient CLN2 protease activities in LINCL cases ranged from 0% to 10% of normal control values, depending on sample source.

[c] The genotype of these patients with LINCL was inferred from sequence analysis of their respective parents.

601780) has not been identified but has been mapped to 15q21-23 by linkage analysis (Sharp et al. 1997; Haines et al. 1998). In addition, progressive epilepsy with mental retardation, or "northern epilepsy" (EPMR; MIM 600143), should probably be considered a form of NCL (Ranta et al. 1996).

In conjunction with genetic analysis and evaluation of clinical progression, examination of subcellular inclusion bodies by electron microscopy (EM) can help differentiate between the different forms of NCL. INCL is characterized by the presence of granular osmiophilic deposits, whereas inclusion bodies of patients with classic LINCL and JNCL have a curvilinear or fingerprint morphology, respectively. Inclusion bodies within cells of patients with variant (CLN6) LINCL tend to have a mixed curvilinear/fingerprint profile.

The CLN2 gene product—which is defective or missing in LINCL—was recently identified by means of an integrated biochemical genomic strategy (Sleat et al. 1997) rather than by traditional genetic methods. The deduced amino acid sequence from the cDNA sequence (GenBank accession number AF017456) of this previously unidentified protein revealed significant similarities with pepstatin-insensitive proteases from *Xanthomonas* and *Pseudomonas,* suggesting a similar catalytic function. This was confirmed when a pepstatin-insensitive protease activity was found to be missing in LINCL samples that were compared with normal controls. Recently, tripeptidyl peptidase I (TPPI) activity was found to be missing in LINCL fibroblasts (Vines and Warburton, 1999). This observation and a comparison of the murine TPPI and CLN2 protease sequences (GenBank accession numbers AJ011912 and AF111172, respectively) suggest that TPPI is the same as the CLN2 protease. It is not clear how the loss of this protease results in LINCL, but it does result in the accumulation of mitochondrial ATPase subunit c (Palmer et al. 1992). Whether this is a direct or an indirect consequence of CLN2 protease deficiency, and whether subunit c accumulation contributes directly to neuronal cell death, are questions that remain to be answered. LINCL is unusual among the lysosomal storage diseases in that it results from a deficiency in an acid protease activity. The other reported example is cathepsin K deficiency in pycnodysostosis (Gelb et al. 1996), which, unlike LINCL, is not neurodegenerative in nature.

A limited preliminary molecular analysis revealed four *CLN2* mutations associated with LINCL: a splice-junction mutation, a nonsense mutation, and two different missense mutations in the same codon (Sleat et al. 1997). The complete sequence of *CLN2* was subsequently determined (GenBank accession number AF039704; Liu et al. 1998), facilitating the design of a set of primers for the generation of overlapping PCR products spanning the entire coding sequence, the 5′ untranslated sequence (UTS), part of the 3′ UTS, most of the intronic sequence, and part of the 5′ untranscribed sequence, which presumably contains promoter elements. In the current study, our aim was to conduct a definitive genetic analysis of *CLN2,* using these primer sets, in a large population of patients with LINCL, to characterize the spectrum of molecular defects. We expected to shed light on the molecular pathology of LINCL and to aid in the development of an effective molecular genetic strategy for diagnosis and carrier testing. We identify 24 different *CLN2* mutations in patients with LINCL, of which 20 have not been previously described. One missense mutation, which presumably does not result in complete functional inactivation of the CLN2 protease, was associated with a late-onset and protracted form of LINCL in two families previously diagnosed with JNCL.

Patients and Methods

Patients

This study group encompasses 74 families with patients who were diagnosed with LINCL by classic clinical criteria (age at onset, clinical progression, and, when available, ultrastructural pathology [Kohlschutter et al. 1993]). Two other families represented patients originally diagnosed with JNCL. One (ANS1) was identified in the course of screening brain autopsy specimens for lysosomal enzyme deficiencies by means of mannose 6-phosphate glycoprotein blotting methods described elsewhere (Sleat et al. 1996; Valenzano et al. 1993). The other (KEW2/BD00190) was a patient with JNCL-like clinical progression but with no detectable mutations in CLN3 and with a mixed fingerprint/curvilinear inclusion morphology (Wisniewski et al. 1998*b*). Samples were obtained from cell and tissue banks, including the Batten Disease Support and Research Association and the Children's Brain Foundation Batten's Disease Lymphocyte Cell Bank (Indiana University School of Medicine), the Batten Disease Registry (Institute for Basic Research, Staten Island), Massachusetts General Hospital, the Istituto Giannina Gaslini (Genova, Italy), the National Neurological Research Specimen Bank (West Los Angeles Veterans Administration Medical Center), the Miami Brain Bank, the McGill University Cell Repository, and the Corriell Cell Collection (Camden, New Jersey). These samples can be accessed from their respective collections with the identification codes listed in the present study. Other samples were obtained from physicians and families directly, according to an institutional review board–approved protocol, and are prefixed with Center for Advanced Biotechnology codes. All samples and their respective sources are listed in table 1.

DNA samples were not available from two patients with LINCL. In these cases (which are noted in table 1),

Table 2

Patients Diagnosed with LINCL Who Are Not CLN2 Protease Deficient

Family	Source (Identifier)[a]	Sample Type	Age at Onset (years)	Lifespan (years)	EM	CLN2 Protease	PCR Products Sequenced
61	4 (BD00357)	LB	2.0	>7.7	CV; FP	Not deficient	A–I
62	4 (BD00043)	LB	4.0	17.5	CV; FP	Not deficient	D–H
63	4 (BD00388)	LB	4.0	>14.8	CV; FP	Not deficient	D–H
64	3 (C7097)	LB	2.5	8.5	CV	Not deficient	A–M
65	2 (GUS23572)	LB	3.5	>11.0	CV; FP; G	Not deficient	H
66	4 (BD00499)	LB	2.5	>8.0	CV; FP; G	Not deficient	None
67	3 (C6958)	FB	2.5	8.0	CV; FP	Not deficient	None
68	4 (BD00501)	LB	2	14.0	CV	Not deficient	A–M
69	1 (IGF075/86)	FB				Not deficient	A–M
70	1 (IGF117/85)	FB				Not deficient	A–M
71	1 (IGF026/90)	FB				Not deficient	A–M
72	1 (IGF078/84)	FB				Not deficient	A–M
73	3 (C10789)	FB	2	>7.0	CV; FP	Not deficient	A–M
74	3 (C7464)	LB	2	10.0	CV; FP	Not deficient	A–M

NOTE.—LB = lymphoblasts; FB = fibroblasts; CV = curvilinear bodies; FP = fingerprint bodies; and G = granular osmiophilic bodies.

[a] Sources: 1 =Instituto Gaslini; 2 = T. J. Lerner; 3 = Institute for Basic Research, Staten Island; and 4 = Batten Disease Lymphocyte Cell Bank at the University of Indiana School of Medicine

the mutant alleles associated with LINCL were inferred from sequence analysis of the parents.

We analyzed 60 families with classic LINCL. *CLN2* was also completely or partially sequenced in 8 normal individuals and in 19 patients affected with diseases other than LINCL.

Preparation of Genomic DNA

Primary human fibroblasts and transformed lymphoblasts were cultured in Roswell Park Memorial Institute media containing 10% or 15% heat-inactivated fetal bovine sera, respectively, with penicillin-streptomycin and nystatin (Gibco BRL). We prepared patient DNA from human brain autopsy samples (50 mg), cultured cells (~10^7 cells), or whole blood (200 μl) collected in the presence of citrate, using either the QiaAmp Tissue DNA purification kit (Qiagen Inc.) or standard phenol-extraction protocols. Genomic DNA was stored at −20°C, in water.

Sequence Analysis of CLN2

Mutational analysis of patients with LINCL was achieved with 13 PCR primer pairs, designated A–M (fig. 2), that allow amplification of the entire *CLN2* coding sequence, all of the intron-exon splice junctions, most of the intervening intronic sequence, and ~300 bp of the 5′ untranscribed sequence (Liu et al. 1998). In cases for which brain autopsy material, blood samples, or cell lines were available, we initially confirmed the diagnosis of LINCL by using the clinical CLN2 protease assay developed in our laboratory (Sohar et al., in press). Patients that lacked CLN2 protease activity were subjected to genetic analysis, until both of the mutant alleles were identified or all 13 amplification products were sequenced. The *CLN2* gene of most patients initially diagnosed with LINCL, but who were not CLN2 protease deficient, was subjected to a limited sequence analysis focusing on the two common mutant alleles that together comprise approximately two-thirds of all mutant alleles in LINCL (table 2). This was done to confirm that these samples did not represent false positives in the enzyme assay. Eleven samples were obtained only as DNA. These were subjected to genetic analysis, until either both of the mutant alleles were identified or all 13 amplification products were sequenced. In two of these cases, no mutations were identified after use of all 13 PCR primer sets, which suggests that these patients suffered from disorders other than LINCL. However, the possibility that pathogenic alleles are present, in regions of the gene excluded from our sequence strategy, cannot be discounted.

To allow dye-labeled primer-cycle sequencing, the forward primers contained the complement to the −21M13 primer, and the reverse primers contained the complement to the M13 reverse primer. PCR products were generated by use of 2 U AmpliTaq Gold (Perkin-Elmer) in 100-μl reactions comprising 200 ng genomic DNA, 0.2 mM of each deoxynucleotide, 200 pmol of each primer, 2 mM $MgCl_2$, 50 mM KCl, 10 mM Tris-HCl (pH 8.3), and 0.001% (w/v) gelatin. PCR cycling conditions were as follows: 94°C for 11 min, for polymerase activation and DNA denaturation; 10 cycles of 94°C for 1 min, 55°C for 1 min, and 72°C for 1 min; 30 cycles of 94°C for 1 min, 68°C for 1 min, and 72°C for 1 min;

followed by 72°C for 10 min, and then 4°C until products were purified. Amplified products were purified by use of QiaQuick PCR purification spin columns (Qiagen), eluted in 50 μl water at 25–100 ng/μl, and sequenced by ABI PRISM Dye Primer Cycle Sequencing kits (Perkin-Elmer) and an ABI 373 automated sequencer. The resulting sequence files were edited by FACTURA and assembled by AUTOASSEMBLER (Applied Biosystems).

We initially detected candidate mutations in *CLN2* by sequencing using the M13-21 dye-labeled primer, except for amplified products from exon 8, which we sequenced using the M13 reverse primer. After the first sequence analysis, we clarified regions of ambiguity, if present, by sequencing the opposite strand. PCR reactions yielding products containing potential mutations were then repeated, and the resulting products were resequenced on the strand opposite that sequenced initially—or, in most cases, in both directions—to verify the mutation.

Results

Identification of CLN2 Homologues

The first indication that the protein encoded by *CLN2* was related to a group of unusual bacterial pepstatin-insensitive proteases came from database searches, which revealed significant sequence similarity with proteases from *Xanthomonas* and *Pseudomonas* (GenBank accession numbers D83740 and D37970). In the present study, we conducted a further database search for other proteins with similarity to the CLN2 protease, to identify regions that are likely to be of structural or functional importance, thus aiding interpretation of *CLN2* mutations in LINCL. In addition to the two bacterial proteases, four other less but significantly ($P < .001$) similar potential CLN2 homologues have been identified. These are a *Dictyostelium discoideum* vegetative stage–specific V4-7 mRNA (GenBank accession number U27540), a gene encoded within a 100-bp fragment of the *Sulfolobus solfataricus* genome (GenBank accession number Y08256), the *Physarum polycephalum PHP* gene (GenBank accession number X64708), and an *Epichloe typhina* protease gene (GenBank accession number L76740). The murine *CLN2* protease (GenBank accession number AF111172; also listed as murine TPPI [accession number AJ011912]) has also been included in the sequence alignment presented in figure 1. It is clear that there are scattered regions of similarity throughout these proteins, but there are three blocks of high similarity, corresponding to amino acids Ala347–Gly362 (*box 1*), Pro379–Thr391 (*box 2*), and Gly473–Gly483 (*box 3*) of CLN2, which suggests that these regions may be of functional importance.

CLN2 Protease Activities in Patients with LINCL

In a parallel study, we have developed a rapid and sensitive assay for the CLN2 protease that is applicable to a wide range of clinical samples, including blood, fibroblasts, lymphoblasts, and brain autopsy material (Sohar et al. 1999). Throughout the present study, we have, when possible, used this assay for biochemical diagnosis of LINCL, limiting extensive sequence analysis only to samples that are CLN2 protease deficient (table 1). In validating the diagnostic CLN2 protease assay, we conducted limited sequence analysis on most samples that were not CLN2 protease deficient, using the H primer set that detects the presence of one or more mutant alleles in ~80% of patients with LINCL (table 2), and in six cases we analyzed the entire *CLN2* coding sequence. No mutations were identified in any patients who were not CLN2 protease deficient.

Forty-seven cell lines were obtained from patients diagnosed with LINCL. Fourteen of these were found to contain CLN2 protease activity and thus probably represent other NCLs (table 2). It is noteworthy that, where morphological or ultrastructural data were available, only 4 of the 35 genetically confirmed cases with LINCL had mixed curvilinear and fingerprint inclusions rather than the typically observed curvilinear bodies alone, whereas 8 of 10 patients who were not CLN2 protease deficient had mixed inclusions. A mixed storage-body morphology is typical of variant LINCL (Sharp et al. 1997), suggesting that, in some of these patients, disease may result from defects in CLN6.

Both mutant alleles were identified in all patients who lacked CLN2 protease activity, with the exception of three. In one case, CLN2 protease was measured in a frozen brain autopsy sample (KEW4), and it is possible that extrinsic factors involved in the removal or storage of the tissue may have resulted in a loss of activity. This is not the case with another patient, in whom no mutations were identified and yet whose lymphoblasts (GUS16776) lacked CLN2 protease activity. In another patient, represented by cell line GM09404, only one pathogenic allele could be found. It is possible that, in all three patients, pathogenic alleles lie outside the scope of our sequence analysis (i.e., >300 nucleotides upstream of the transcription start site, within intron 3, or within the 3′ UTS). Another possibility is that mutations in another protein, perhaps required for the stability or processing of the CLN2 protease, result in LINCL. A precedent for this already exists for the lysosomal storage disease galactosialidosis (D'Azzo et al. 1982). In this disease, mutations within lysosomal protective protein (cathepsin A) result in a combined β-galactosidase and α-neuraminidase deficiency.

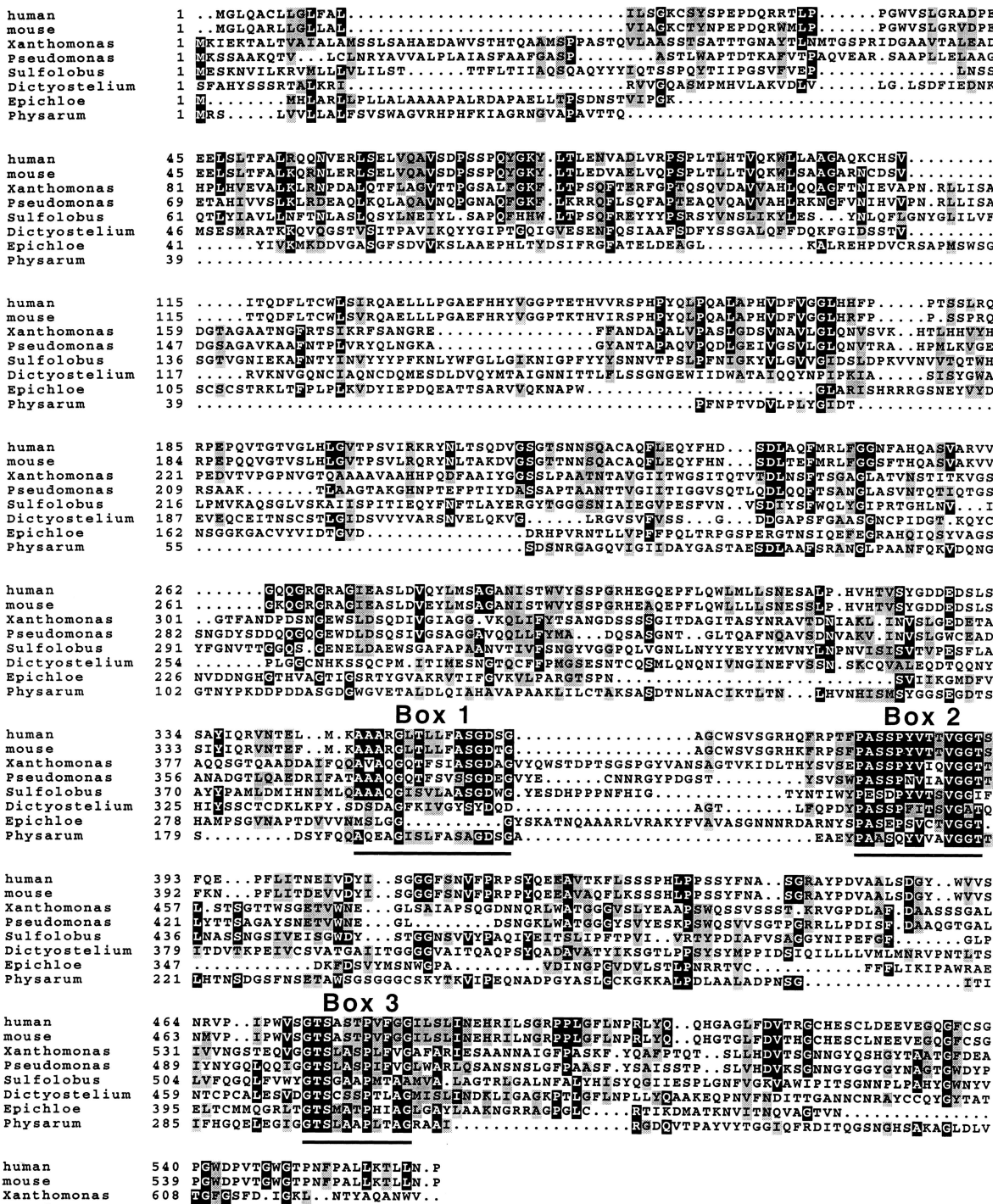

Figure 1 Sequence alignment of the human CLN2 protease, with eukaryotic and prokaryotic homologues. CLN2 protease homologues were identified by use of the program TBLASTN 2.0.6 (Sept-16-1998) to search the GenBank Non-Redundant database. Sequences with significant ($P < .001$) sequence similarity to the CLN2 protease were aligned by use of the GCG program PILEUP, with ENDWEIGHT (Genetics Computer Group 1991), and compiled for publication by use of BOXSHADE 3.21 (ISREC Bioinformatics Group [Swiss Institute for Experimental Cancer Research]). Regions of homology (*boxes 1–3*) are underlined.

Table 3

Frequency of *CLN2* Polymorphisms in an Unrelated Population Sample

Nucleotide	Position	Samples Analyzed	Allele		Percentage of Genotype		
			1	2	1[a]/1	1/2	2/2
1089	5′ untranscribed	46	T	C	54.4	30.4	15.2
1177	5′ untranscribed	47	G	A	91.5	2.2	6.4
1205	5′ untranscribed	46	T	C	89.1	4.4	6.5
1224	5′ untranscribed	45	G	A	97.8	2.2	.0
1767	Intron 2	58	C	T	98.3	1.7	.0
3004	Exon 4	56	A	G	89.2	5.4	5.4
3436	Intron 5	67	A	T	46.3	40.3	13.5
3767	Intron 6	126	T	C	62.7	30.2	7.1
5044	Intron 9	64	C	G	86.0	14.0	.0
5276	Intron 10	82	G	A	97.6	.0	2.4
5837	Exon 12	63	A	T	84.1	14.3	1.6
5644	Intron 11	43	T	C	72.0	23.3	4.7
5658	Intron 11	42	A	G	95.2	4.8	.0

[a] The relative frequency of each polymorphism was calculated from a large number of patients with LINCL and their immediate families, patients with unrelated disorders, and normal controls. To avoid redundancy, results from only one affected sibling are included here. In families in which the parents have been analyzed these data are included in place of data of the respective proband, to increase the scope of the survey among unrelated individuals.

CLN2 Polymorphisms

We identified 13 *CLN2* polymorphisms in the sequence analysis of the 60 LINCL families, 8 normal controls, and 19 patients with other genetic diseases. These polymorphisms are listed in table 3. Most were located in the 5′ untranscribed sequence or within introns, but two silent polymorphisms were identified within coding regions.

CLN2 Mutations

In the present study, we have identified the *CLN2* defects associated with disease in 58 LINCL families. Samples from these patients, their respective sources, and repository access numbers are listed in table 1. Twenty-four mutations were identified, 20 of which are previously unreported (table 1). These mutations, their predicted effects on the CLN2 gene product, and their frequency of incidence are summarized in tables 1, 4, and 5, and their respective locations within *CLN2* are indicated in figure 2. In 21 of the 58 families, the *CLN2* gene contained homozygous changes, and the remaining 37 were compound heterozygous. Half (12 of 24) of the different mutations were identified in only a single patient (private mutations), whereas two mutations were particularly common (3670C→T and 3556G→C) and accounted for 60% of all *CLN2* mutant alleles.

Eleven CLN2 missense mutations were identified within the LINCL families (table 4) but not in the healthy controls or patients affected with other genetic diseases. Most of the missense mutations do not result in amino acid substitutions at highly conserved residues except Val385 and Gly389, which are highly conserved or invariant components of box 1, between the CLN2 protease and its homologues (fig. 1 *box 1*).

Two different LINCL patients of Romany ethnicity

Table 4

Summary of CLN2 Mutations that Do Not Involve Intron-Exon Splice Junctions

Nucleotide Change	Mutation Type	Location	Amino Acid Change	No. of Alleles (%)
1917C→T	Nonsense	Exon 3	Gln66Stop	2 (1.7)
1950G→A	Missense	Exon 3	Gly77Arg	1 (.9)
3081-3091del	Deletion	Exon 4	Frameshift	2 (1.7)
3084C→T	Nonsense	Exon 4	Arg127Stop	1 (.9)
3670C→T	Nonsense	Exon 6	Arg208Stop	32 (27.8)
4023T→A	Missense	Exon 7	Ile287Asn	1 (.9)
4288-4295del	Deletion	Exon 8	Frameshift	3 (2.6)
4346G→A	Missense	Exon 8	Glu343Lys	3 (2.6)
4654T→C	Missense	Exon 9	Cys365Arg	1 (.9)
4655G→A	Missense	Exon 9	Cys365Tyr	4 (3.5)
5159T→A	Missense	Exon 10	Val385Asp	1 (.9)
5171G→A[a]	Missense	Exon 10	Gly389Glu	4 (3.5)
5271G→C	Missense	Exon 10	Gln422His	8 (7.0)
5457G→A	Missense	Exon 11	Arg447His	2 (1.7)
5478C→A	Missense	Exon 11	Ala454Glu	1 (.9)
5541C→T	Missense	Exon 11	Ser475Leu	1 (.9)
6069InsA	Insertion	Exon 13	Frameshift	1 (.9)
6151-6152delTC	Deletion	Exon 13	Frameshift	2 (1.7)

[a] This missense mutation was always found in association with A3004G (Gln100Arg), which is listed as a probable polymorphism in table 3 (for reasons, see Results).

Table 5

Summary of CLN2 Splice-Junction Mutations

Nucleotide Change	Location	Mutation Class[a]	No. of Alleles (%)	Splice-Site Consensus Sequence[b]	
				5′ (Donor): *NMAG*GTRAGTN *NMAG*GTRAGTN	3′ (Acceptor): YYNYAG*RNNNN* YYNYAG*RNNNN*
1946A→G	Exon 3	1	1 (.9)	*CTCAATACG*GT→*CTCAGTACG*GT	
4396T→G	Intron 8	2	1 (.9)	*TCAG*GTGACCT→*TCAG*GGGACCT	
3556G→A	Intron 5	2	3 (2.6)		CTACAG*TGGGG*→CTACAA*TGGGG*
3556G→C	Intron 5	2	38 (33.0)		CTACAG*TGGGG*→CTACAC*TGGGG*
4188A→G	Intron 7	1	1 (.9)		AAAAAAAAAAA→AAAAAGAAAAA
6025G→C	Intron 12	2	1 (.9)		CCCTAG*GTAAC*→CCCTAC*GTAAC*

[a] Mutation classes represent (1) generation of a cryptic splice junction and (2) changes in invariant splice-junction residues.
[b] Exonic sequences are in italics; underlined residues are invariant. Codes are M = A or C, R = A or G, N = any base, and Y = C or T.

were found to be homozygous for two missense mutations that were not observed in any other patients with LINCL (5171G→A [Gly389Glu] and 3007G→A [Gln100Arg]). It is probable that one of these changes represents a true mutation and the other a closely linked polymorphism. If this is the case, then Gly389Glu is the more likely candidate for a pathogenic allele, because this is a completely conserved residue (fig. 1, *box 1*). Another possibility is that each of these changes alone is not sufficient to result in an inactive CLN2 protease but their compound effect is. Familial genetic analysis does not shed light on this question, as both parents of one patient were heterozygous for both of the changes. Future structure-function analyses of the CLN2 protease should resolve this issue.

A number of mutations were observed within the invariant intronic elements of splice junctions, including 3556G→C, which accounts for ~32% of all LINCL mutant alleles. Two nucleotide changes were also identified, which potentially create cryptic splice junctions in close proximity to known sites. In one, 1946G→A, which is within exon 3, a potential splice site is generated that is just upstream of an established splice-donor site. Pre-mRNA processing at this cryptic site would result in a five-nucleotide deletion in the mature mRNA and a subsequent frameshift. Splicing at another cryptic site, resulting from mutation 4188A→G, would also cause a frameshift by adding 17 nucleotides to the final mRNA. In this case, the cryptic splice junction is in a more favorable context for splicing than the original because it is preceded by a TC-rich sequence, whereas the original is preceded by a run of 19 A residues, which is unusual and not generally considered typical of a splice junction. Neither of these cryptic splice sites has been proved by experiment; the possibility remains that they are polymorphisms and that the true mutations are yet to be identified. However, it should be noted that neither of these nucleotide changes was found in any of the normal controls or in any of the patients with LINCL, in whom both mutant alleles had been definitively identified, and where genetic analyses included the regions containing these two potential mutations (42 samples for nucleotide 1946, and 45 samples for nucleotide 4188).

Three nonsense mutations were identified, including 3670C→T (Arg208Stop), which accounted for ~28% of the mutant alleles. In addition, four frameshifts resulting from small deletions or insertions were also found, including a 2-nucleotide deletion located only 13 nucleotides upstream from the TGA stop codon. This frameshift mutation changes the four carboxy-terminal amino acids of CLN2 and is predicted to add 13 amino acids.

Two of the mutant alleles identified by Sleat et al. (1997) (3556G→C and 3670C→T) account for 60% of all CLN2 mutations detected in the present study. Both of these alleles were found in 40% of patients (genotypes 3556[C/C]:3670[C/C], 3556[G/C]:3670[C/T], or 3556[G/G]:3670[T/T]), whereas 38% of patients contained one of these two alleles (genotypes 3556[G/C]:3670[C/C] or 3556[G/G]:3670[C/T]); thus, a total of 78% of patients with LINCL contain one or two of these alleles. The other mutations (Cys365Arg and Cys365Tyr) identified by Sleat et al. (1997) were not commonly found in patients with LINCL, representing together only 4.2% of all mutant alleles.

Clinical Parameters and CLN2 Mutations

One of the aims of the present study was to determine whether the clinical progression of LINCL could be predicted from the CLN2 genotype. However, it is clear from table 1 that, with the exception of two patients (ANS1 and BD00190), the clinical progression of LINCL in our study population was relatively homogeneous and thus uninfluenced by the type of mutation involved.

Patients ANS1 and BD00190 represent two apparently unrelated individuals who were originally diagnosed with JNCL, both with onset at age ~8 years and death at ages 43 years and 30 years, respectively. Both of these patients are compound heterozygous. Patient

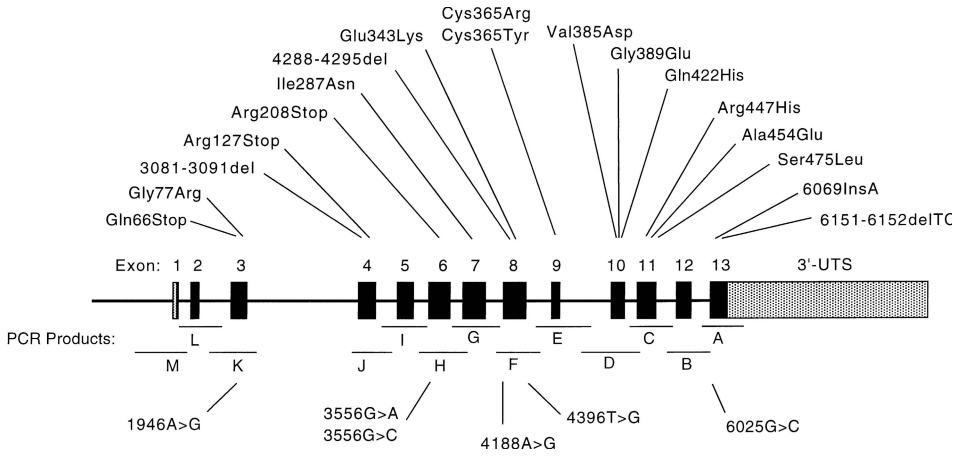

Figure 2 Schematic drawing of the *CLN2* gene, showing the positions of mutations associated with LINCL. Deletions, insertions, nonsense, and missense mutations are indicated above the schematic, and mutations involved in splicing are indicated below.

ANS1 is heterozygous for the common Arg208Stop mutation (3670C/T), and patient BD00190 is heterozygous for the common splice-junction mutation (3556G/C), but both share a unique monoallelic CLN2 mutation that is not observed in the other patients with LINCL. This mutation (5457G→A) results in a missense mutation, Arg447His. A limited familial analysis of these patients indicated that the Arg447His mutation is associated with protracted LINCL. In the case of patient ANS1, one parent of the proband was heterozygous for the Arg208Stop mutation (3670[C/T]:5457[G/G]), whereas the other parent was heterozygous for the Arg447His change (3670[C/C]:5457[G/A]). An unaffected sibling carrier was identical to the first parent at these alleles, being heterozygous for only the Arg208Stop mutation. In the case of patient BD00190, an affected sibling (BD00196) was genotypically identical (3556[G/C]:5457[G/A]) to the proband. Clinical progression of the typical and atypical LINCL cases, including two affected siblings of patient BD00190, is summarized in table 6.

When analyzing our study population with respect to mutation incidence, we also considered the ethnic origin/race of the patients, when known. Incidence of LINCL was panethnic, but a disproportionate number (24/50) of patients were completely or largely of English-Scottish-Irish extraction. There was little correspondence between the types of mutation and ethnic background, although the two common mutations (3670C→T and 3556G→C) were identified more frequently in patients of predominantly northern European origin. For example, pathogenic alleles in 58% of patients of predominantly northern European origin are restricted to the two common mutations, compared with 24% of patients of eastern, central, or southern European origin. Ethnic linkage was observed with two mutations. A double homozygous missense mutation (Gln100Arg:Gly389Glu) was identified only in two patients of Romany origin residing in Italy. In addition, a homozygous deletion (4288-4295del) was identified in a patient of Italian descent and was also one of the pathogenic alleles in a compound-heterozygous genotype, in a patient of mixed Italian-French descent.

Discussion

In the present study, we have located pathogenic *CLN2* alleles in 58 patients with LINCL, identifying 24 mutations, 20 of which were previously undescribed. Two *CLN2* mutations that were originally identified in LINCL—a splice-junction mutation and a nonsense mutation—were found to be particularly common, together accounting for 60% of mutant alleles (an observation made also by Zhong et al. 1998). At least one of either of these two common mutations is detected in 78% of all patients with LINCL, and both are amplified on the same exon 6 fragment with the H primer set (Liu et al. 1998), indicating that molecular genetic diagnosis of LINCL should be a practical proposition. However, the rest of the observed LINCL mutations, which each individually account for ~1%–7% of the total number of observed mutant alleles, are scattered throughout the *CLN2* coding sequence. In the absence of clues from ethnicity (see Results), these alleles are potentially difficult and time consuming to identify. Therefore, a sequential sequencing strategy is presented here, which allows for the most rapid identification of mutant alleles, with the smallest number of PCR/sequencing reactions. In this optimized strategy, the sequential order of exon sequencing is dictated by a function of both the relative

frequency of each mutation contained within each exon and the number of different mutations contained within each exon. The value of this approach in the identification of mutant alleles, as a percentage of total and for identification of both mutant alleles in patients with LINCL, is demonstrated in figure 3. The probability of identifying at least one mutant allele as the exons are sequentially analyzed is also presented, as this may be sufficient for the initial diagnosis. (The efficiency of diagnosis, defined by the detection of at least one mutant allele, can be improved slightly by altering the order of sequencing to PCR products H, D, F, E, A, G, C, and J [data not shown]). Additional sequencing may be required for patients in whom both alleles are not identified (e.g., patients KEW4, GUS16776, and GM09404).

Little is currently known about the synthesis or processing of the CLN2 protease, other than that the primary translation product contains an ~25–kD light chain or propiece preceding the N-terminus of the 46-kD form that was originally purified (Sleat et al. 1997). In this study, 10 of 11 missense mutations identified were located within the 46-kD C-terminal portion of the predicted protein. This clustering of missense mutations suggests that the 25-kD N-terminal portion of the primary translation product actually represents a propiece that is not required for the catalytic function of the *CLN2* gene. This is typical of lysosomal proteases, which are synthesized as zymogens and activated within acidified compartments by removal of propieces (reviewed by Gabel 1993).

Two unrelated patients (ANS1 and BD00190) were originally diagnosed with JNCL on the basis of age at onset, age at death, and inclusion morphology. EM analysis of a postmortem brain sample from one patient (ANS1) was reported to have found fingerprint inclusions, which are typical of JNCL, although analysis of buffy-coat leukocytes revealed both curvilinear and fingerprint profiles. EM analysis of buffy-coat leukocytes from patient BD00190 and siblings also revealed a mixed inclusion morphology. We analyzed CLN2 protease activities in samples (brain autopsy samples and/or cell lines) derived from these patients, as potential atypical LINCL cases, and confirmed a lack of activity. Sequence analysis determined that the CLN2 defects in both probands and an affected sibling of one were compound heterozygous, with either monoallelic common splice-junction or Arg208Stop mutations. However, both of these atypical patients shared a rare monoallelic mutation that resulted in the substitution of arginine 447 for histidine, which was not observed in other patients with LINCL. It is likely that this mutation causes an incomplete loss of function of the CLN2 protease, resulting in the protracted phenotype. CLN2 protease activity was not detectable in samples from these patients, suggesting that any residual activity was very low compared with normal controls. Such observations are not uncommon with other lysosomal storage diseases, in which missense mutations result in a partial dysfunction and subsequent mild or protracted phenotype (reviewed by Gieselmann 1995).

The Arg447His mutation appears to be rare, on the

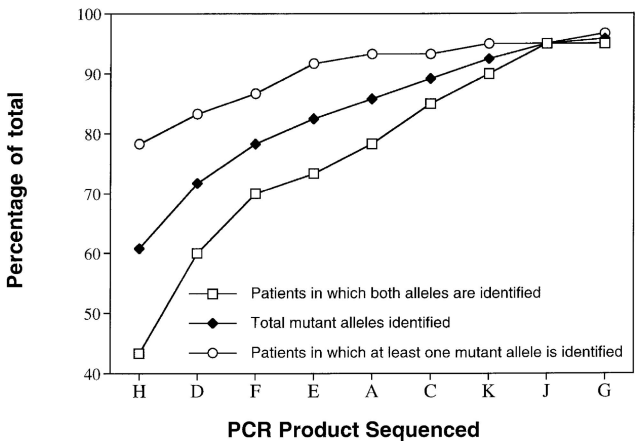

Figure 3 An optimized sequencing strategy for efficient identification of CLN2 mutations associated with LINCL.

Table 6

Clinical Synopsis of Typical and Atypical Cases of Classic LINCL

Classic LINCL Subtype	Age at Onset (years)	Lifespan (years)
Late-onset, protracted LINCL:		
Family 50 (ANS1)	8	43
Family 39 (BD00190 and 2 affected siblings)[a]	6–8	24, 30, >29
Overall	7.3 ± .5	>31.5 ± 8.1
Typical LINCL	2.9 ± 0.4[b]	>8.3 ± 2.8[c]

[a] BD00190, BD00196, and BD00190sib represent three affected siblings, the latter of which was not included in this study. ANS1 is unrelated.

[b] $n = 46$; range 2.0–3.8.

[c] $n = 43$; range 3.0–16.0.

basis of the results of this study, but it should be kept in mind that we analyzed, for the most part, only patients diagnosed with LINCL, expressing typical clinical progression. Thus one intriguing possibility is that there may exist individuals who are homozygous for Arg447His. The likely phenotype of such an individual remains open to speculation. It is possible that homozygosity for Arg447His may have no apparent pathogenic effects or, alternatively, that this genotype could result in a highly attenuated form of LINCL that might not be diagnosed or that might be confused with other late-onset neurodegenerative disorders. CLN2 defects do not, however, appear to be responsible for adult NCL (Kufs disease), as CLN2 protease activities are the same or higher than controls in adult NCL lymphoblasts (Sohar et al. 1999).

With the identification of the defective genes in the three most common forms of NCL and subsequent molecular genetic analyses of atypical forms, it is becoming increasingly apparent that there exist atypical late-onset or protracted variants of each. In this study, we have identified a *CLN2* mutation that results in a phenotype more characteristic of JNCL. Other recent studies highlight *CLN1* mutations that result in clinical expression more typical of *CLN2* or *CLN3* defects (Das et al. 1998; Mitchison et al. 1998; Wisniewski et al. 1998*a,* 1998*b*) and *CLN3* mutations that result in a greatly protracted clinical course (Munroe et al. 1997). However, in conjunction with traditional parameters (clinical history and subcellular morphology) and rapid biochemical tests, molecular genetic analyses of the type described here now allow for the unequivocal diagnosis of each of type of NCL, allowing prenatal and presymptomatic testing in addition to heterozygote identification in affected families. In addition, the ability to determine precisely the defective gene in affected children may also allow for the development of targeted therapeutic approaches that are specifically tailored to each NCL subtype.

Acknowledgments

We gratefully acknowledge the many parents and families and their respective physicians who have participated in this study. We would particularly like to thank Lance Johnston of the Batten Disease Support and Research Association for his help in enrolling subjects in this study. Other collaborators whom we wish to thank are Dr. Stefano Goldwurm (Clinica Pediatrica, Universita' di Milano, Fondazione Tettamanti, Ospedale San Gerardo Monza, Milan), Dr. Petra Krebs (Institut fuer Humangenetik, Magdeburg, Germany), Dr. Fuki Hisama (Yale University School of Medicine, New Haven, Connecticut), Drs. Cheryl Greenberg and Louise Dilling (Health Sciences Center, Winnipeg, Canada), and Dr. Elizabeth Berry-Kravis (Rush-Presbyterian-St. Luke's Medical Center, Chicago). We thank the Laboratorio di diagnosi pre e postnatale malattie metaboliche (Istituto G. Gaslini, Genova, Italy) for providing us with specimens from the collection "Cell lines and DNA bank from patients affected by genetic disease," supported by Telethon grants (project C. 20); the National Institute of General Medical Sciences Human Genetic Mutant Cell Repository at the Corriell Institute for Medical Research (Camden, New Jersey); the McGill University Repository for Mutant Human Cell Strains (Toronto, Canada); and the Batten Disease Lymphocyte Cell Bank at the Indiana University School of Medicine. Some brain samples were obtained from the National Neurological Research Specimen Bank, Veterans Administration Greater Los Angeles Health Care System (West Los Angeles, California), which is sponsored by NINDS/NIMH, the National Multiple Sclerosis Society, the Veterans Administration Greater Los Angeles Healthcare System, and the Veterans Health Services and Research Administration of the Department of Veterans Affairs. This work was supported by the Children's Brain Disease Foundation and by National Institutes of Health grants DK45992 and NS37918 (to P.L.).

Electronic-Database Information

Accession numbers and URLs for data in this article are as follows:

GenBank, http://www.ncbi.nlm.nih.gov/Web/Genbank/index.html (for CLN2 protease sequence [AF11172], murine TPPI protease sequence [AJ011912], CLN2 protein [AF017456], *Xanthomonas* [D83740], *Pseudomonas* [D37970], *D. discoideum* [U27540], *S. solfataricus* [Y08256], *P. polycephalum PHP* [X64708], *E. typhina* protease [L76740], and murine CLN2 [AF111172])

ISREC Bioinformatics Group, http://www.isrec.isb-sib.ch/software/BOX_form.html (for BOXSHADE 3.21)

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim (for NCL, *CLN1* [MIM 256730]; classic LINCL, *CLN2* [MIM 204500]; JNCL, *CLN3* [MIM 204200]; Finnish variant LINCL, *CLN5* [MIM 256731]; variant LINCL, *CLN6* [MIM 601780]; and EPMR [MIM 600143]).

References

Boustany R-M (1996) Neurodystrophies and neurolipidoses. In: Moser HW (ed.) Handbook of clinical neurology, Vol 22. Elsevier Science, Amsterdam, pp 671–700

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

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

D'Azzo A, Hoogeveen A, Reuser A, Robinson D, Galjaard H (1982) Molecular defect in combined beta-galactosidase and neuraminidase deficiency in man. Proc Natl Acad Sci USA 79:4535–4539

Gabel CA (1993) Post-translational processing and intracellular transport of newly synthesized lysosomal enzymes. In: Loh YP (ed) Mechanisms of intracellular trafficking and pro-

cessing of proproteins. CRC Press, Boca Raton, Florida, pp 103–130

Gelb BD, Shi GP, Chapman HA, Desnick RJ (1996) Pycnodysostosis, a lysosomal disease caused by cathepsin K deficiency. Science 273:1236–1238

Genetics Computer Group (1991) Program manual for the GCG package, version 7. University of Wisconsin, Madison

Gieselmann V (1995) Lysosomal storage diseases. Biochim Biophys Acta 1270:103–136

Haines JL, Boustany RMN, Alroy J, Auger KJ, Shook KS, Terwedow H, Lerner TJ (1998) Chromosomal localization of two genes underlying late-infantile neuronal ceroid lipofuscinosis. Neurogenetics 1:217–222

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

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

Kohlschutter A, Gardiner RM, Goebel HH (1993) Human forms of neuronal-ceroid lipofuscinosis (Batten disease): consensus on diagnostic criteria, Hamburg 1992. J Inherit Metab Dis 16:241–244

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

Mitchison HM, Hofmann SL, Becerra CH, Munroe PB, Lake BD, Crow YJ, Stephenson JB, et al (1998) Mutations in the palmitoyl-protein thioesterase gene (PPT; CLN1) causing juvenile neuronal ceroid lipofuscinosis with granular osmiophilic deposits. Hum Mol Genet 7:291–297

Munroe PB, Mitchison HM, O'Rawe AM, Anderson JW, Boustany R-M, Lerner TJ, Taschner PEM, et al (1997) Spectrum of mutations in the Batten disease gene, *CLN3*. Am J Hum Genet 61:310–316

Palmer DN, Fearnley IM, Medd SM (1992) Mitochondrial ATP synthase subunit c storage in the ceroid-lipofuscinoses (Batten disease). Am J Med Genet 42:561–567

Pearce DA, Sherman F (1998) A yeast model for the study of Batten disease. Proc Natl Acad Sci USA 95:6915–6918

Ranta S, Lehesjoki AE, Hirvasniemi A, Weissenbach J, Ross B, Leal SM, de la Chapelle A, et al (1996) Genetic and physical mapping of the progressive epilepsy with mental retardation (EPMR) locus on chromosome 8p. Genome Res 6:351–360

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

Sharp JD, Wheeler RB, Lake BD, Savukoski M, Jarvela IE, Peltonen L, Gardiner RM, et al (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

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

Sleat DE, Sohar I, Lackland H, Majercak J, Lobel P (1996) Rat brain contains high levels of mannose-6-phosphorylated glycoproteins including lysosomal enzymes and palmitoyl-protein thioesterase, an enzyme implicated in infantile neuronal lipofuscinosis. J Biol Chem 271:19191–19198

Sohar I, Sleat DE, Jadot M, 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 and animal specimens. J Neurochem 78, in press

Valenzano KJ, Kallay LM, Lobel P (1993) An assay to detect glycoproteins that contain mannose 6-phosphate. Anal Biochem 209:156–162

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

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

Wisniewski KE, Connell F, Kaczmarski W, Kaczmarski A, Siakotos A, Becerra CR, Hofmann SL (1998*a*) Palmitoyl-protein thioesterase deficiency in a novel granular variant of LINCL. Pediatr Neurol 18:119–123

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

Zhong N, Wisniewski KE, Hartikainen J, Ju W, Moroziewicz DN, McLendon M, Brooks SS, et al (1998) Two common mutations in the CLN2 gene underlie late infantile neuronal ceroid lipofuscinosis. Clin Genet 54:234–238