

CLN5, a novel gene encoding a putative transmembrane protein mutated in Finnish variant late infantile neuronal ceroid lipofuscinosis

Minna Savukoski¹, Tuomas Klockars¹, Ville Holmberg¹, Pirkko Santavuori², Eric S. Lander³ & Leena Peltonen¹

The neuronal ceroid lipofuscinoses (NCLs) represent a group of common recessive inherited neurodegenerative disorders of childhood, with an incidence of 1:12,500 live births¹. They are characterized by accumulation of autofluorescent lipopigments in various tissues. Several forms of NCLs have been identified, based on age at onset, progression of disease, neurophysiological and histopathological findings and separate genetic loci²⁻⁹. All types of NCL cause progressive visual and mental decline, motor disturbance, epilepsy and behavioral changes, and lead to premature death. One of the subtypes, Finnish variant late infantile neuronal ceroid lipofuscinosis (vLINCL; MIM256731) affects children at 4–7 years of age^{10,11}. The first symptom is motor clumsiness, followed by progressive visual failure, mental and motor deterioration and later by myoclonia and seizures. We have previously reported linkage for vLINCL on chromosome 13 (ref. 5) and constructed a long-range physical map over the region¹². Here, we report the positional cloning of a novel gene, *CLN5*, underlying this severe neurological disorder. The gene encodes a putative transmembrane protein which shows no homology to previously reported proteins. Sequence analysis of DNA samples from patients with three different haplotypes revealed three mutations; one deletion, one nonsense and one missense mutation, suggesting that mutations in this gene are responsible for vLINCL.

Three PAC clones covering the critical 200 kb containing the *CLN5* locus (Fig. 1) were sequenced as part of the Human Sequencing Project at Whitehead Institute. Transcripts from the region were identified based on searches of the EST database with the genomic sequence obtained from PACs and on screening of cDNA libraries with PAC 76n15 located between the flanking markers of the *CLN5* region¹² (Fig. 1). This resulted in numerous EST hits and cDNA clones obtained from a fetal brain cDNA library. Sequence analyses of these transcribed regions revealed one (a 1.4-kb cDNA clone) that contained a nucleotide difference between patients and controls.

The *CLN5* cDNA sequence was assembled from the fetal brain cDNA clone, several ESTs and a targeted 5'-RACE product from a fetal brain cDNA library. The 5' UTR sequence revealed 90% homology with a previously cloned CpG island, HS2F1F. At the 3' end of the first exon, the EST contig suggests an exon-intron boundary, although the cDNA clone does not, suggesting an alternatively spliced exon at this location. Due to the difficulties in RT-PCR of this region, only the splice site obtained from the EST contig could be confirmed (Fig. 1).

The *CLN5* cDNA was found to be approximately 4.1 kb, consisting of four exons (Fig. 1) with an open reading frame (ORF) of 1,380 bp. There are four ATG codons in the 5' end of the ORF. Based on consensus sequence analysis, the first ATG is considered

as the initiator of translation, resulting in a predicted polypeptide of 407 amino acids. The four exons of *CLN5* span a region of over 13 kb in genomic DNA. A 1,500-bp region upstream from the 5' UTR of *CLN5* was analysed using NNPP (ref. 13), suggesting three possible 50-bp promoter sequences, starting from bases 1026, 1091 and 1172 (scores 0.93, 0.97 and 0.81, respectively).

Northern-blot analysis of lymphoblast mRNA from *CLN5* Finnish major patients and control individuals revealed no detectable size difference or alteration in steady-state transcript level. RNA from Fin minor and Dutch patients was not available.

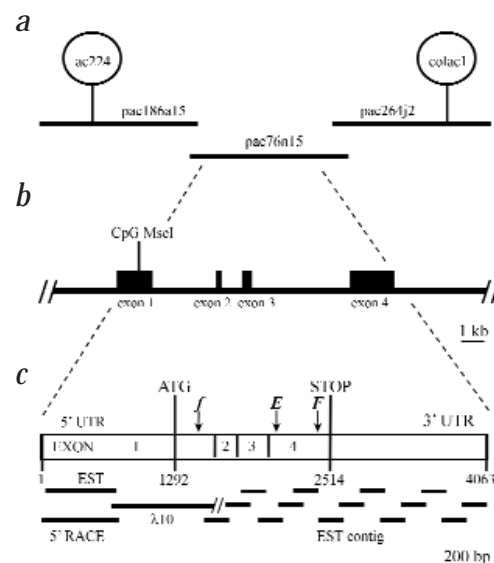


Fig. 1 Positional cloning of *CLN5*. **a**, Isolation of two novel polymorphic markers (*AC224*, *COLACT*) dramatically restricted the critical *CLN5* region and a 300-kb physical map was constructed over this chromosomal area. Transcripts were identified based on database searches with the genomic sequence and on screening of cDNA libraries with PAC 76n15 and analysed for mutations. **b**, *CLN5* has four exons spanning a region of over 13 kb in the genomic DNA. The first exon is 1,614 bp, the second 164 bp, the third 227 bp and the fourth 2,058 bp; the first intron is 2.8 kb, the second 0.7 kb and the third 4.3 kb. In addition, the intron between exons three and four shows high homology to two overlapping ESTs. These localize to a repetitive region and might represent an artifact, alternatively spliced exon of *CLN5* or another gene on the same region. With the first exon, a database hit to a previously cloned CpG island *MseI* fragment was found. **c**, A 4,063-bp cDNA was constructed from fetal brain cDNA clone, 5'-RACE product and an overlapping contig of numerous ESTs found in dbEST. All four exons recognized several dbEST hits and the exon-intron boundaries were confirmed by RT-PCR and by sequencing. More than 35 dbEST hits were found, covering more than 75% of the complete cDNA sequence (bases 16–603, 1,505–4,064). Length of the estimated coding region is 1,221 bp, corresponding to a polypeptide of 407 amino acids. Three mutations, *F*, *f* and *E*, were found from three different affected haplotypes (*CLN5* Fin major, *CLN5* Fin minor and *CLN5* European, respectively).

¹Department of Medical Genetics, University of Helsinki and Department of Human Molecular Genetics, National Public Health Institute, Mannerheimintie 166, 00300 Helsinki, Finland. ²Hospital for Children and Adolescents, Department of Neurology, University of Helsinki, Helsinki, Finland. ³Whitehead Institute for Biomedical Research, Cambridge, Massachusetts, USA. M.S. and T.K. contributed equally to this work. Correspondence should be addressed to L.P. e-mail: Leena.Peltonen@ktl.fi

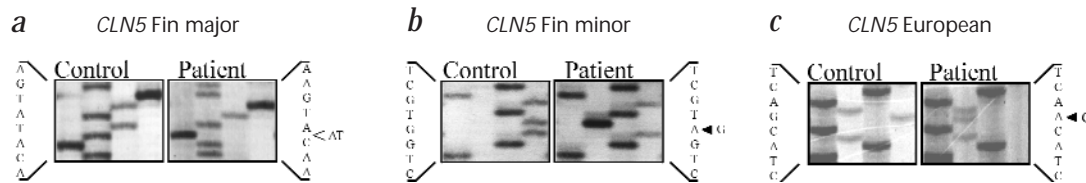


Fig. 2 Mutations in *CLN5*. The sequence lanes appear from left to right as C, A, T and G. **a**, Sequence of a patient carrying the *CLN5* Fin major mutation, (AT)₂₄₆₇ deletion. **b**, Sequence of *CLN5* Fin minor patient, with a homozygous haplotype different from *CLN5* Fin major patients, was found to carry a G₁₅₁₇→A transversion. **c**, Sequence of a Dutch patient with the *CLN5* European mutation, a G₂₁₂₇→A transversion.

All four exons were PCR amplified from genomic DNA of patients and controls, and analysed by direct sequencing. We identified three mutations in the *CLN5* cDNA in patients carrying three different homozygous haplotypes. The first, the Finnish major mutation (*CLN5* Fin major) was identified in seventeen families (34/36 Finnish disease chromosomes). It is a 2-bp deletion of (AT)₂₄₆₇ in exon 4, resulting in a change of Tyr(392)STOP (Fig. 2). This leads to a predicted protein of 391 amino acids instead of the 407 predicted in controls. The carrier frequency of the *CLN5* Fin major mutation is dependent on the population analysed. 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 of one community, where many of the ancestors of the Fin major families originate. Carrier frequency of the rest of the high-risk area was approximately 1:100 (4/460). No carriers were observed among 100 control individuals from elsewhere in Finland (Table 1).

The second disease mutation was identified in a single family with two patients carrying the homozygote *CLN5* Fin minor haplotype. A G₁₅₁₇→A transversion in exon 1 causes a substitution of Trp(75)STOP leading to a predicted protein of only 74 amino acids. No carriers were detected among 500 Finnish control subjects or in 100 control subjects from elsewhere in Europe (Fig. 2). Another alteration was also found in the same haplotype; an A₂₃₉₅→G transversion in exon 4 changing Lys(368)Arg. The observed carrier frequency of this change was nearly 20% in Fin-

land, and additionally, the father of the family was found to be homozygous for this variation. This change represents a common polymorphism, tightly linked to the *CLN5* Fin minor mutation.

The only mutation of non-Finnish origin (*CLN5* European) was detected in a Dutch patient, also with a homozygous haplotype, in exon 4. Transversion of G₂₁₂₇→A causes a change of Asp(279)Asn. No carriers of this alteration were observed among 100 European control individuals.

The tissue expression pattern of *CLN5* was analysed using a 1,291-bp PCR-derived fragment (nt 1,446–2,737) to probe a human multiple-tissue northern blot. Relatively weak hybridization signals of approximately 2.0 kb, 3.0 kb and 4.5 kb were detected in all tissues. In addition, a signal of approximately 5.5 kb could be seen in skeletal muscle. The assembled 4.1-kb *CLN5* cDNA corresponds best to the 4.5-kb transcript, the other signals possibly representing alternatively spliced transcripts.

Hybridization of an RNA dot blot using a 637-bp PCR derived fragment (nt 972–1,608) as a probe suggested the highest levels of expression in aorta, kidney, lung and pancreas of adult tissues. Fetal tissues (brain, heart, kidney, liver, spleen and lung) displayed fairly uniform expression levels, except for fetal thymus, showing approximately twofold higher signal intensity. The overall observation suggested expression of *CLN5* in all human tissues, with up to fivefold variation in expression levels.

No homologous genes or proteins were found in GenBank or other databases. Therefore, we sought to discern the function

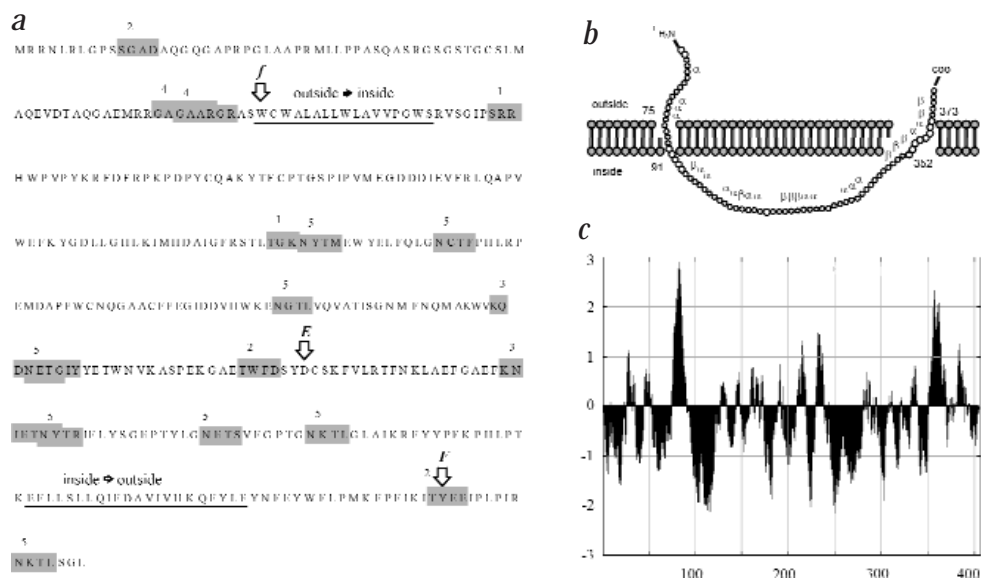


Fig. 3 Predicted features of the *CLN5* polypeptide. **a**, The amino acid sequence and predicted features of the *CLN5* protein. The three mutations are shown in the figure as F, f and E (*CLN5* Fin major, *CLN5* Fin minor and *CLN5* European, respectively). The structure of the estimated *CLN5* polypeptide of 407 amino acids was predicted using several computer programs. The PSORT predicted intracellular localization of the *CLN5* polypeptide to microbody (0.640), cytoplasm (0.450), lysosomal lumen (0.236) and mitochondrial matrix space (0.100). The polypeptide displays several PROSITE features, consisting of two protein kinase C phosphorylation sites (1), three casein kinase II phosphorylation sites (2), two tyrosine kinase phosphorylation sites (3), two overlapping N-myristoylation sites (4) and eight N-glycosylation sites (5). **b**, The BCM Transmembrane Prediction Program predicted two potential transmembrane regions (underlined) with the N terminus of the polypeptide outside. The α -helix and β -sheet predictions by PSSP prediction program are also shown. **c**, A Kyte & Doolittle hydrophobicity plot²⁴ of the predicted *CLN5* polypeptide reveals two strong hydrophobic regions in correlation with the predicted transmembrane regions.

Table 1 • CLN5 mutations, consequences and carrier frequencies

	FIN _{major}	FIN _{minor}	European
Mutation	del(AT) ₂₄₆₇	G ₁₅₁₇ →A	G ₂₁₂₇ →A
Consequence	Tyr ₃₉₂ →STOP	Tyr ₇₅ →STOP	Asp ₂₇₉ →Asn
Carrier frequency	1:20–1:200 (Finland)	No carriers in 500 samples (Finland)	No carriers in 100 samples (Finland)

from aspects of the sequence. The 407 amino-acid CLN5 polypeptide corresponds to a theoretical molecular weight of 46 kD and a calculated pI of 8.41. Structure and intracellular localization of the CLN5 polypeptide were predicted using several protein prediction programs^{15–17} (Fig. 3). Several potential modification sites, including protein kinase C phosphorylation sites, casein kinase II phosphorylation sites, tyrosine kinase phosphorylation sites, N-myristoylation sites and N-glycosylation sites were identified; all features shared with the predicted CLN3 polypeptide⁸. The PSORT (ref. 15) program provided no definitive intracellular localization of the predicted CLN5 polypeptide. Two hydrophobic regions were predicted to indicate transmembrane helices, which would suggest CLN5 to encode a transmembrane protein with an intraluminal loop. These features represent only predictions and firm conclusions on the polypeptide encoded by CLN5 must await experimental evidence.

Out of three known NCL genes, CLN1 and CLN2 represent genes encoding lysosomal enzymes^{9,18}, whereas CLN3 encodes a lysosomal membrane protein¹⁹, although this location was not initially predicted from the polypeptide sequence. It is tempting to speculate that NCL-disorders could be divided into two molecular subgroups: defects of lysosomal enzymes and defects of lysosomal membrane proteins, assuming that CLN5 would also fit in this category of proteins.

Identification of CLN5 provides a tool for understanding the pathology of vLINCL patients at the cellular and tissue level. This may, in turn, be a first step towards development of specific therapy and treatment of symptoms of this severe disease. It will also aid in understanding the possible biochemical function of CLN5 and other genes responsible for NCL disorders, which (based on their devastating phenotypes) are crucial for normal development and maturation of cortical neurons.

- Rider, J.A. & Rider, D.L. Batten disease: Past, present and future. *Am. J. Med. Genet. (Suppl.)* **5**, 21–26 (1988).
- Santavuori, P. Neuronal ceroid lipofuscinosis in childhood. *Brain Dev.* **10**, 80–83 (1988).
- Gardiner, R.M. *et al.* Batten disease (Spielmeyer-Vogt; juvenile onset neuronal ceroid lipofuscinosis) maps to human chromosome 16. *Genomics* **8**, 387–390 (1990).
- Järvelä, I. *et al.* Infantile neuronal ceroid lipofuscinosis (INCL, CLN1) maps to the short arm of chromosome 1. *Genomics* **8**, 170–173 (1991).
- Savukoski, M. *et al.* Defined chromosomal assignment of CLN5 demonstrates that at least four genetic loci are involved in the pathogenesis of human ceroid lipofuscinoses. *Am. J. Hum. Genet.* **55**, 695–701 (1994).
- Sharp, J. *et al.* Loci for classical and variant late infantile neuronal ceroid lipofuscinosis map to chromosomes 11p15 and 15q21–23. *Hum. Mol. Genet.* **6**, 591–595.
- Vesa, J. *et al.* Mutations in the palmitoyl protein thioesterase gene causing infantile neuronal ceroid lipofuscinosis. *Nature* **376**, 584–587 (1995).
- The International Batten Disease Consortium. Isolation of a novel gene underlying Batten disease, CLN3. *Cell* **82**, 949–957 (1995).
- Sleat, D.E. *et al.* Association of mutations in a lysosomal protein with classical late-infantile neuronal ceroid lipofuscinosis. *Science* **277**, 1802–1805 (1997).
- Santavuori, P., Rapola, J., Sainio, K. & Raitta, C. A variant of Jansky-Bielschowsky disease. *Neuropediatrics* **13**, 135–141 (1982).
- Santavuori, P. *et al.* Early juvenile neuronal ceroid lipofuscinosis or variant Jansky-Bielschowsky disease: diagnostic criteria and nomenclature. *J. Inher. Metab. Dis.* **16**, 230–232 (1993).
- Klockars, T. *et al.* 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 (1996).
- Reese, M.G., Harris, N.L. & Eeckman, F.H. Large scale sequencing specific neural

Methods

cDNA cloning. The PAC clone was digested with *EcoRI* and used as a probe in cDNA library screening. cDNA library screening was performed according to standard protocols²⁰ using a fetal brain Agt10 cDNA library (Clontech). Targeted 5'-RACE was performed using a fetal brain cDNA library (Clontech).

Northern-blot hybridizations. PolyA⁺ RNA was isolated from peripheral blood lymphocytes of vLINCL Fin major patients and healthy controls using the Fast Track 2.0 Kit (Invitrogen) for mRNA isolation. In each lane, polyA⁺ RNA (5 µg) was electrophoretically separated in a 0.8% agarose gel in the presence of formaldehyde and blotted onto a nylon membrane. For multiple tissue northern analysis we used human multiple tissue northern blot (MTN; heart, brain, placenta, lung, liver, skeletal muscle, kidney, pancreas; Clontech) and Human RNA Master Blot (Clontech). Northern hybridizations were performed at 65 °C using ExpressHyb (Clontech). PCR products were subcloned in a pBluescript-derived vector using pCR-Script Amp SK(+) Cloning Kit (Stratagene) and used as probes.

Mutation detection. Mutations were detected in vLINCL patients from PCR amplified genomic DNA with manual sequencing by solid-phase sequencing method²¹ using the following primers: CLN5 Fin major, 5'-CCCTTC-AAACACATTTGCCA-3'; CLN5 Fin minor, 5'-CAGGAGGTAGACACG-GCACA-3'; CLN5 European, 5'-AGCTTTGTTCTACTAGGTGACT-3'. Detected mutations were monitored against control panels of unrelated Finnish and/or European control subjects by size separation of radioactively labelled PCR products on denaturing PAGE (CLN5 Fin major mutation) or by minisequencing²² (CLN5 Fin minor and CLN5 European mutations).

Computational analyses. Nucleotide sequence was analysed with BlastN algorithms²³ against public databases. NNPP (ref. 13), PSORT (ref. 15), PSSP (ref. 16) and Tmpred (ref. 17) predictions were done through Internet. The Kyte-Doolittle hydrophobicity plot²⁴ was created interactively by ProtScale (EXPASY World Wide Web molecular biology server). Sequence alignments were performed using Sequencher 3.0 (Gene Codes Corporation).

GenBank accession numbers. CLN5 cDNA, AF068227; 76n15, AC001226; HS2F1F, Z58138; ESTs, N62724, N78448.

Acknowledgements

We are grateful to the vLINCL families. We thank P. Taschner for kindly providing the Dutch patient's DNA sample. P. Hakala, T. Airaksinen and J. Isosomppi are appreciated for expert technical assistance. This work was supported by grants from the Academy of Finland, Sigrid Juselius Foundation, the Ulla Hjeltn Fond for Foundation of Pediatric Research, the Rinnekoti Research Foundation and Finska Läkaresällskapet.

Received 13 April; accepted 2 June, 1998.

- networks for promoter and splice site recognition. in *Proceedings of the Pacific Symposium on Biocomputing*. (eds Hunter, L. & Klein, T., Kona, 1996).
- Varilo, T. *et al.* The age of human mutation: genealogical and linkage disequilibrium analysis of the CLN5 mutation in the Finnish population. *Am. J. Hum. Genet.* **58**, 506–512 (1996).
- Nakai, K. & Kanehisa, M. A knowledge base for predicting protein localization sites in eukaryotic cells. *Genomics* **14**, 897–911 (1992).
- Solovyev, V. V. & Salamov, A. A. Predicting α -helix and β -strand segments of globular proteins. *Comput. Appl. Biosci.* **10**, 661–669 (1994).
- Hofmann, K. & Stoffel, W. TMBASE - A database of membrane spanning protein segments. *Biol. Chem. Hoppe-Seyler* **374**, 166 (1993).
- Hellsten, E., Vesa, J., Olkkonen, V., Jalanko, A. & Peltonen, L. Human palmitoyl protein thioesterase: evidence for lysosomal targeting of the enzyme and disturbed cellular routing in infantile neuronal ceroid lipofuscinosis. *EMBO J.* **15**, 5240–5245 (1996).
- Järvelä, I. *et al.* Biosynthesis and intracellular targeting of the CLN3 protein defective in Batten disease. *Hum. Mol. Genet.* **7**, 85–70 (1998).
- Maniatis, T., Fritsch, E.S. & Sambrook, J. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1982).
- Syvänen, A.C., Aalto-Setälä, K., Kontula, K. & Söderlund, H. Direct sequencing of affinity captured amplified human DNA: application to the detection of apolipoprotein E polymorphisms. *FEBS Lett.* **258**, 71–74 (1989).
- Syvänen, A.C., Sajantila, A. & Lukka, M. Identification of individuals by analysis of biallelic DNA markers, using PCR and Solid-Phase minisequencing. *Am. J. Hum. Genet.* **52**, 46–59 (1993).
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–419 (1990).
- Kyte, J. & Doolittle, R.F. A simple method for displaying the hydrophobic character of a protein. *J. Mol. Biol.* **157**, 105–132 (1982).